X-RAY CRYSTALLOGRAPHIC STUDIES OF GLYCOGEN PHOSPHORYLASE **b**

A Thesis submitted for the degree of Doctor of Philosophy

University of Oxford

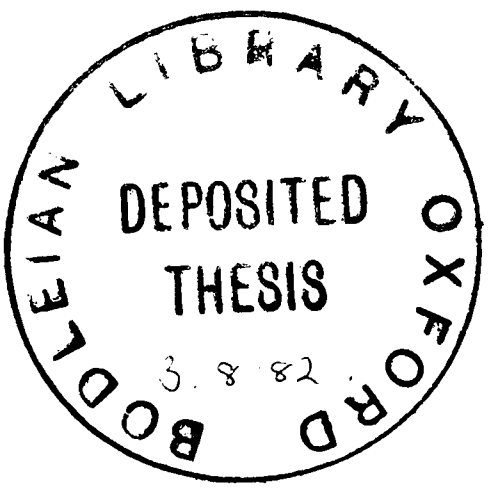
November 1981

David Leslie Wild

Laboratory of Molecular Biophysics

Department of Zoology

Wolfson College



ABSTRACT

X-Ray Crystallographic
Studies of Glycogen phosphorylase b

D.L. Wild
Wolfson College

D.Phil.
Michaelmas Term, 1981

The structure of rabbit muscle glycogen phosphorylase *b*, an important regulatory enzyme in glycogen metabolism, has been studied by X-ray crystallographic techniques. This work was carried out as part of a group project, and the crystal structure of the enzyme had already been solved to 3 Å resolution, using the technique of Multiple Isomorphous Replacement.

A search for additional heavy atom isomorphous derivatives was carried out, and photographic data to 3 Å resolution were collected for a further ethylmercurythiosalicylate derivative, using a screenless oscillation camera. The collection and reduction of this data, and the refinement of the heavy atom positions is described. The inclusion of this data allowed a new electron density map (with figure of merit = 0.63) to be calculated, which enabled previously ambiguous areas in the electron density to be interpreted.

Data to 2 Å resolution have been collected on an oscillation camera, using a synchrotron radiation source and cylindrical film cassettes. An intensity gain of up to 130 times, compared to a GX6 rotating anode source, was obtained with the synchrotron radiation source. A reduction in radiation damage was also observed. The collection and reduction of the 2 Å data is described. The final overall merging R-factor was 15%. Some systematic errors remain in the data, and possible sources of these errors are discussed, and improvements to the data processing procedure suggested. The 2 Å data were empirically scaled to the 3 Å data and have been used in the first stages of the refinement of the phosphorylase *b* structure.

The contribution of the crystallographic results towards an understanding of phosphorylase *b* as an allosteric protein is discussed.

Acknowledgements

I would like to thank my supervisors, Dr. Keith Wilson and Dr. Louise Johnson, for their help and advice throughout this project. This work was done as part of a group project, and I would like to thank my colleagues Irene Weber, David Yeates, John Jenkins, Enrico Stura and David Stuart for their help, advice and c tion.

I would also like to thank Mike Pickford, Mike Richert, John Marsh and Philip Keighley for technical assistance, Lance Mangold for the photographs, and Mrs. Caron Crisp for her excellent typing of this thesis.

This work was supported by a Studentship from the Medical Research Council, and laboratory facilities were kindly provided by Professor Sir David Phillips.

Finally I would also like to thank my parents for all their help.

CONTENTS

		Page
Symbols and Abbreviations		i
Chapter 1	A Brief Outline of Protein Crystallography	1
	1. Introduction	1
	2. Basic Diffraction Theory	2
	3. Isomorphous Replacement	3
	4. Calculation of the Primary Phase Set	10
	5. The Difference Fourier Synthesis	12
Chapter 2	Glycogen Phosphorylase	14
	1. Properties of Glycogen Phosphorylase	14
	2. Crystallographic Analysis of Glycogen Phosphorylase	20
Chapter 3	The Oscillation Method and the Processing of Photographic Data	26
	Introduction	26
	Oscillation Geometry	32
	The Processing of Oscillation Photographs	34
	1. Film Densitometry	34
	2. Crystal Orientation	35
	3. Prediction of Reflections in an Oscillation Range	38
	4. Determination of the Optronics Matrix	40
	5. Integration	41
	6. Inter-film Scale Determination	43
	7. Radiation Damage	49
	8. Reduction to Unique Data	51
	9. The Treatment of Negative Intensity Observations	53
	10. Scaling of Data to the Native Intensities	54

		Page
Chapter 4	The Collection, Processing and Refinement of Heavy Atom Derivative Data	56
	1. Heavy Metal Search	56
	2. Collection and Processing of 3 Å Data for EMTS7	72
	3. Reprocessing of Previous Data	85
	4. Scaling to Native	88
	5. Refinement of Heavy Atom Positions	90
	6. New Phase Calculation	95
	7. Location of Heavy Metal Binding Sites on the Molecule	98
Chapter 5	Collection of 2 Å Native Data	104
	1. The Problem of Data Collection	104
	2. The Design and Construction of the Cylindrical Cassette	106
	3. The Use of Synchrotron Radiation in Protein Crystallography	108
	4. Experimental Conditions	113
	5. Collection of 2 Å data for phosphorylase *b*	115
	6. Diffuse Scattering	119
Chapter 6	Processing of the Native 2 Å Data	123
	1. Determination of Crystal Orientation Matrices	123
	2. Prediction of Reflections in an Oscillation Range	124
	3. Determination of the Optronics Matrix	126
	4. Integration	127
	5. Inter-film Scale Determination	132
	6. Reduction to Unique Data	135
	7. Scaling and Merging of the Complete Data Set	144
	8. Scaling to 3 Å Native Data	146
	9. Conclusions	147

		Page
Chapter 7	Discussion and Conclusion	151
	1. Contribution to Improved 3Å Map	151
	2. Phosphorylase as an Allosteric Protein	154
	3. Contribution of High Resolution Data Collection.	161
Appendix 1	The Computer Programmes	166
Appendix 2	Errors in the Alignment of the Cassette	168
References		175

Symbols and Abbreviations

a, b, c	Unit cell dimensions in direct space
a*, b*, c*	Reciprocal cell dimensions
ADP	Adenosine monophosphate
AMP	Adenosine diphosphate
ATP	Adenosine triphosphate
AU	$KAu(CN)_2$ isomorphous derivative
B	Atomic temperature factor
BES	N,N-bis(2-hydroxyethyl)-2-aminoethane sulphonic acid
cAMP	Cyclic 3',5'-adenosine monophosphate
D	Observed optical density
E, E"	Standard lack of closure errors in the isomorphous and anomalous phase triangles
EDTA	Ethylene diamino tetraacetic acid
EMTS	Ethyl mercuri thiosalicylate
$\underline{\ }_H$	The heavy atom contribution to the structure factor of the derivative
$\underline{f}_H$	The calculated heavy atom structure factor
$\lvert\underline{F}_H\rvert$ or F_H	The modulus of $\underline{F}_H$
F_{HLE}	The lower estimate of F_H from isomorphous and anomalous measurements
F_{HUE}	The upper estimate of F_H from isomorphous and anomalous measurements
$\underline{F}_P$	The structure factor of the native protein
$\lvert\underline{F}_P\rvert$ or F_P	The modulus of $\underline{F}_P$
$\underline{F}_{PH}$	The structure factor of the isomorphous derivative
$\lvert\underline{F}_{PH}\rvert$ or F_{PH}	The modulus of $\underline{F}_{PH}$
$\underline{F}_{PI}$	The structure factor of the protein-inhibitor complex
G3	Maltotriose
G1P	Glucose-1-phosphate
G6P	Glucose-6-phosphate

h, k, l	Reciprocal lattice indices
$\underline{h}$	Reciprocal lattice vector (h, k, l)
I	Intensity of a reflection
IMP	Inosine monophosphate
k	Theoretical ratio of isomorphous to anomalous observation
k_{emp}	Empirical ratio of isomorphous to anomalous observation
m	Figure of merit of the phases
M.W.	Molecular weight in daltons
N_{obs}	Number of observations
NAD	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
O.D.	Standard unit of optical density
P	phosphate
Pi	Inorganic phosphate
PLP	Pyridoxal phosphate
Pt	The $Pt(NH_3)_2Cl_2$ isomorphous derivative
R	The merging R factor: $\frac{\sum_{\underline{h}} \sum_{i} \lvert \bar{I}(\underline{h}) - I_i(\underline{h}) \rvert}{\sum_{\underline{h}} \sum_{i} I_i(\underline{h})}$
R_c	$\frac{\sum \lvert F_{PH} - \lvert F_P + f_H \rvert \rvert}{\sum \lvert F_{PH} - F_P \rvert}$ for centric observations
rms	Root mean square
s	$4\sin^2\theta/\lambda^2$, θ is the Bragg angle and λ the wavelength
S.R.	Synchrotron radiation
V	The volume of the unit cell
$\underline{x}$	Positional vector within unit cell (x, y, z)
α_P, α_{iso}	The protein phase and the isomorpnous protein phase
γ	The reflecting range of the crystal
ΔF	The isomorphous difference, $\left\lvert \lvert \underline{F}_{PH} \rvert - \lvert \underline{F}_P \rvert \right\rvert$

ξ, ζ, τ	The cylindrical polar coordinates
$\rho(\underline{r})$	The electron density
$\phi, \Delta\phi$	The rotation angle and rotation range
σ	The standard deviation of an observation
$\Delta F''$	The anomalous difference $\left\lvert \lvert \underline{F}_{PH}^{+} \rvert - \lvert \underline{F}_{PH}^{-} \rvert \right\rvert$

CHAPTER 1

A BRIEF OUTLINE OF PROTEIN CRYSTALLOGRAPHY

1) INTRODUCTION

Protein crystallography is the application of X-ray diffractic techniques to the analysis of the structures of crystalline proteins. Protein molecules are often difficult to crystallise and must be kept in equilibrium with their mother liquor, since the crystals contain a significant percentage of solvent. Proteins crystallise with large unit cells and therefore a large number of diffraction intensities must be measured to achieve a given resolution. These intensities are weaker than for small molecules, necessitating higher intensity X-ray sources or longer exposure times. However, with longer exposure times radiation damage is often a severe problem for protein crystals, because of their large water content.

What follows is an outline of protein crystallographic theory relevant to the work presented. A more detailed account of the methods of structure determination may be found in many textbooks (*e.g.* Stout and Jensen, 1968; Blundell and Johnson, 1976).

In general, the scheme for the X-ray analysis of a crystalline protein is as follows:-

i) crystallisation
ii) preparation of heavy atom derivatives
iii) data collection
iv) data processing
v) determination of heavy atom positions
vi) calculation of phases
vii) interpretation of electron density maps
viii) refinement of the structure.

2) BASIC DIFFRACTION THEORY

X-rays are diffracted by the electrons in a molecule because their wavelength is of the order of atomic dimensions. The diffraction pattern of the molecule is described by the Fourier transform of the electron density:

$$F(\underline{h}) \propto \int \rho(x)\exp(2\pi i\underline{h}.\underline{x})dx \qquad 1.1$$

In a crystal the electron density of a molecule is convoluted with the crystal lattice and so the diffraction pattern of a crystal is described by the product of the molecular density transform and the lattice transform (the reciprocal lattice). This gives rise to a discrete diffraction pattern which represents the continuous Fourier transform of the contents of the unit cell sampled at the reciprocal lattice points. Thus the electron density distribution of the unit cell may be calculated from the inverse Fourier transform of the "structure factors" $F(\underline{h})$:

$$\rho(\underline{x}) = \sum_{h} F(\underline{h})\exp(-2\pi i\underline{h}.\underline{x}) \qquad 1.2$$

This is analogous to the formation of an image by a glass lens at optical wavelengths and the use of an incomplete set of structure factors will result in a loss of resolution of the image due to "series termination" errors. The structure factor $F(\underline{h})$ is a complex number, possessing amplitude and phase. The amplitude of the structure factor is proportional to the experimentally measured intensity, $F(\underline{h})^2 \propto I$. The phase is, however, not directly observable and its determination constitutes the main problem of crystallography.

3) ISOMORPHOUS REPLACEMENT

a) Method

The phase problem has been successfully solved for protein crystals by the method of Multiple Isomorphous Replacement (M.I.R.) (Green *et al.* 1954). This method involves the introduction of a small number of intense scattering centres into the crystal. These are usually heavy atoms which bind at specific sites without disturbing the protein structure. The requirement is that the heavy atom replaces another group or solvent molecule on the protein without causing conformational changes. If the derivative is not isomorphous, changes in, for example, the cell dimensions will also give rise to differences in the diffraction pattern. These effects increase with Bragg angle. Crick and Magdoff (1956) calculated that a change of all cell dimensions by 0.5% will give changes in intensity averaging approximately 15% in the 3 Å resolution sphere.

Heavy atoms may be introduced into the protein crystal by:-

i) the diffusion of a heavy atom reagent into the protein crystal through the large channels of mother liquor in the crystal,

ii) chemical modification of the protein, followed by crystallisation,

iii) co-crystallisation of protein and heavy atom reagent.

b) Location of Heavy Atom Positions

If $\underline{F}_P$ is the structure factor of the protein, $\underline{F}_{PH}$ the structure factor of the isomorphous derivative and $\underline{F}_H$ the contribution of the heavy atom to the structure factor of the derivative, then

$$\underline{F}_{PH} = \underline{F}_P + \underline{F}_H \qquad 1.3$$

Before the method of M.I.R. can be used to determine the phases of the native protein structure factors, the heavy atom contribution $\underline{F}_H$ must be determined. If the positions of the heavy atoms within the unit cell are known, $\underline{F}_H$ may be calculated from the following expression:-

$$\underline{F}_H(h) = \sum_n f_n \exp(-2\pi i(\underline{h}.\underline{x}_n) \qquad 1.4$$

where f_n is the heavy atom scattering factor

x_n is the position of site n

and the summation extends over the whole unit cell.

In general the heavy atom positions are found from a difference Patterson synthesis with coefficients:-

$$|\underline{F}_H|^2 \simeq \left| |\underline{F}_{PH}| - |\underline{F}_P| \right|^2 \qquad 1.5$$

(Perutz, 1956; Blow, 1958). Because the Patterson function is the Fourier transform of $|\underline{F}_P|^2$ it may always be calculated from a set of intensities. The Patterson function represents a map of interatomic vectors. There are N^2 peaks for a structure of N atoms, N of which are self-self vectors which superimpose at the origin and the remaining N(N-1) are distributed throughout the unit cell. For two atoms i and j containing Z_i and Z_j electrons respectively, a peak height of $Z_i Z_j$ is produced. Thus the heavier atoms give rise to higher peaks. The difference Patterson as defined above is only approximately equal to $|\underline{F}_H|^2$. It was at first thought that the heavy atom positions could best be derived from the 'proper' difference Patterson with coefficients ($|\underline{F}_{PH}|^2 - |\underline{F}_P|^2$) (Phillips 1966). The resulting vector map would include peaks corresponding not only to heavy atom-heavy atom vectors but also to heavy atom-protein interactions. Both Patterson functions will be sensitive

to lack of isomorphism. However, the 'proper' difference Patterson, as defined above, would be more sensitive to errors of scaling and measurement than the one defined in equation 1.5.

Direct methods have also been used to locate heavy atom positions. Wilson (1978) describes the use of the tangent formula in the analysis of the isomorphous derivatives of four proteins. The method led to the correct structure for almost all the derivatives studied, but was limited by the experimental accuracy of the isomorphous derivatives.

c) Estimation of $\underline{F}_H$

In centric zones the vectors $\underline{F}_P$ and $\underline{F}_H$ are colinear, so that:-

$$|\underline{F}_H| = \left| |\underline{F}_{PH}| \pm |F_P| \right| \qquad 1.6$$

If the degree of heavy atom substitution is small then $|\underline{F}_H|$ will usually be correctly estimated from the difference of $|\underline{F}_{PH}|$ and $|\underline{F}_P|$. If both $|\underline{F}_{PH}|$ and $|\underline{F}_P|$ are very small it may be that the correct estimate is $\left| |\underline{F}_{PH}| + |\underline{F}_P| \right|$. This is known as "crossing-over" For acentric zones, $|\underline{F}_{PH}| - |\underline{F}_P|$ will not be an accurate estimate of $|\underline{F}_H|$ for the purposes of refinement and a different expression must be used. An estimate of $|\underline{F}_H|$ may be obtained from a combination of the isomorphous and anomalous scattering information. Anomalous scattering occurs when the frequency of the incident X-rays is close to an absorption frequency of the scattering atom. A phase shift occurs which is described by a complex structure factor $\underline{F}_H = f'+if''$ and Friedel's Law ($|\underline{F}_H| = |\underline{F}_{-H}|$) breaks down. $|\underline{F}_H|$ may be estimated from the expression:-

$$F_H^{\,2} = F_P^{\,2} + F_{PH}^{\,2} \pm 2F_P F_H \left(1-(k\Delta/2F_P)^2\right)^{\frac{1}{2}}$$

(Matthews, 1966a; Harding, 1962).

Taking the negative sign gives F_{HLE} (Heavy Atom Lower Estimate), whilst the positive sign gives F_{HUE} (Heavy Atom Upper Estimate). This is analogous to "crossing over" in the centric case. k is the anomalous ratio f'/f" and Δ is the anomalous difference $F_{PH}^{+}-F_{PH}^{-}$. Δ is often overestimated because of random errors in measurement and various methods have been proposed to reduce this overestimation.

1) Matthews (1966a) and Kartha and Parthasarathy (1965a,b) proposed the use of an empirical value for k:-

$$k_{emp} = \frac{2<|F_{PH}-F_{P}|>}{<|\Delta|>} \qquad 1.8$$

2) Dodson *et al.* (1975) proposed that a 'bias' correction, calculated from statistical theory be subtracted from the observed F_{HLE} and the theoretical value of k used.

$$F^{2}_{HLEB} = F^{2}_{HLE} - \text{BIAS} \qquad 1.9$$

where BIAS = $(k^2/4)\sigma^2_{ANO}$.

This method has the advantage that each term is treated independently, rather than in blocks of reciprocal space.

d) Scaling of Derivative to Native Data

The scaling of heavy atom derivative data to the native set requires careful analysis to avoid the occurence of 'ghosts' (holes or peaks) at the heavy atom sites in the final electron density map. An approximate value of the derivative scale factor, obtained from setting

$$K_{PH}\sum F_{PH}^{2}/\sum F_{P}^{2} = 1$$

may prove inadequate for the later stages of an analysis. For centric data F_H may be replaced by $K_{PH}F_{PH}-F_P$ which leads to the expression:-

$$K_{PH}\sum F_{PH}F_P = \sum F_P^2$$

where K_{PH} is the derivative scale factor and summation is over all reflections in the centric zone (Blundell and Johnson, 1976).

For acentric data F_H may be replaced by F_{HLE}. The method of Arnone *et al.* (1971), based on a concept suggested by Kraut *et al.* (1962) uses the expression:-

$$\sum F_{HLE}^2 = \sum (K_{PH}F_{PH})^2 - \sum F_P^2 \qquad 1.10$$

where F_{HLE} is the lower estimate of equation 1.7. An approximate scale for F_{PH} needs to be known before F_{HLE} is calculated. This may be obtained from the expression given for centric terms. An improved value for K_{PH} (known as the 'Kraut' scale factor) is then obtained using equation 1.10 and the process iterated until convergence is reacned. The 'Kraut' scale factor may be subsequently refined during the least squares refinement of the heavy atom parameters (see Section 3e) (Evans, 1973) but experience suggests that the least squares minimum is not sharply defined.

e) Refinement of Heavy Atom Positions

In solving the phase problem by isomorphous replacement, the parameters characterizing a heavy atom site need to be accurately refined. These paramters are the occupancy, positional and thermal parameters of the site, and the scale factors relating derivative to native structure factor amplitudes. They are refined by a least-squares minimization of the discrepancy between the observed heavy

atom structure factors $F_H(obs)$ and the calculated structure factors $F_H(calc)$:-

$$\sum w(F_H(obs)-F_H(calc))^2 \qquad 1.11$$

where w is a weight.

The refinement scheme used depends on whether independent phase information is available.

i) Centric Refinement

In centric zones $F_H(obs)$ is simply the isomorphous difference given by equation 1.6:-

$$F_H = F_{PH} \pm F_P$$

the positive sign giving the case of "cross-over". Such "cross-over" terms are usually rejected from the refinement procedure.

After refinement, difference Fourier syntheses with coefficients $|F_H(obs)-F_H(calc)|\exp i\alpha_{calc}$ (where αcalc are calculated phases) may be calculated to check the refined parameters and locate additional binding sites.

ii) F_{HLE} Refinement

For acentric zones where no independent phase information is available, F_{HLE} values may be used for $F_H(obs)$. These are obtained from a combination of anomalous scattering and isomorphous replacement information, and are given by the lower estimate of expression 1.7. F_{HLE} values tend to have rather large errors and the value of F_{HLE} may be estimated using k_{emp} (equation 1.8) or the bias corrected F_{HLE} described earlier (equation 1.9). The weights may then be taken as $w = 1/\sigma^2$ where σ is the estimated

standard deviation of the F_{HLE}. Alternatively, a weighting scheme may be used which weights down the smaller and larger F_{HLE} values, since these are most likely to suffer from large errors. A suitable weighting scheme is:-

$$= \left[1 + \frac{(F_{HLE}-b)^2}{a}\right]^{\frac{1}{2}} \qquad 1.12$$

where a and b are chosen to make $w(F_{HLE}-F_H(calc))^2$ reasonably constant in different ranges of F_{HLE} (Adams *et al.*, 1969; Blundell *et al.*, 1971).

Difference Fourier syntheses with coefficients $|F_{HLE}-F_H(calc)|\exp(i\alpha_H(calc))$ may be used to check the refined parameters and locate additional sites.

The quality of the heavy atom derivative may also be assessed by calculating the following error expressions:-

$$E(F_{HLE}) = \frac{\sum_h (|F_{HLE}-F_H(calc)|)^2}{n} \qquad 1.13a$$

and

$$R(F_{HLE}) = \frac{\sum|F_{HLE}-F_H(calc)|}{\sum F_{HLE}} \qquad 1.13b$$

where E is the mean square error and the 'R factor' is an index of agreement between the observed and calculated structure. A completely random structure would give R = 59% for acentric terms and R = 83% for centric terms (Wilson 1949, 1950).

iii) "Phase" Refinement

When independent phase information is available phase refinement can be used for centric or acentric data. This involves minimizing the expression:-

$$\sum w(|\underline{F}_{PH}|-|\underline{F}_{P}+\underline{F}_{H}|)^2 \qquad 1.14$$

and the weight w is the figure of merit (defined in Section 4). The derivative under refinement should be excluded from the phase calculation to avoid making any contribution to the phase information. Phase refinement is only useful when there are at least two independent derivatives without common sites. The course of the refinement can be followed by comparing the initial and final 'Kraut' R values:-

$$\text{Initial } R_k = \frac{\sum \left| F_{PH}(obs)-|\underline{F}_{P}| \right|}{\sum F_{PH}(obs)} \qquad 1.15$$

$$\text{Final } R_k = \frac{\sum |F_{PH}(obs)-F_{PH}(calc)|}{\sum F_{PH}(obs)} \qquad 1.16$$

where $F_{PH}(obs) = |\underline{F}_{PH}|$ and $F_{PH}(calc) = |\underline{F}_{P}+\underline{F}_{H}|$.

'Double' difference Fourier syntheses with coefficients $\left(F_{PH}(obs)-|\underline{F}_{P}+\underline{F}_{H}|\right)\exp(i\alpha_P)$ may be used to identify minor sites and check the validity of the refined parameters.

4) CALCULATION OF THE PRIMARY PHASE SET

For centric zones $\underline{F}_P$, $\underline{F}_{PH}$ and $\underline{F}_H$ are colinear and the phases take on special values such as 0 and π. Thus equation 1.6 can give an unambiguous determination of the phase, using only one isomorphous derivative for such terms. However, for acentric reflections, the use of one isomorphous derivative leaves a phase ambiguity. This is illustrated in Figure 1.1 using the Harker construction (Harker, 1956). This ambiguity can be resolved by including two derivatives (Figure 1.2) or the anomalous scattering information

The Harker construction for perfectly isomorphous derivatives with no experimental errors.

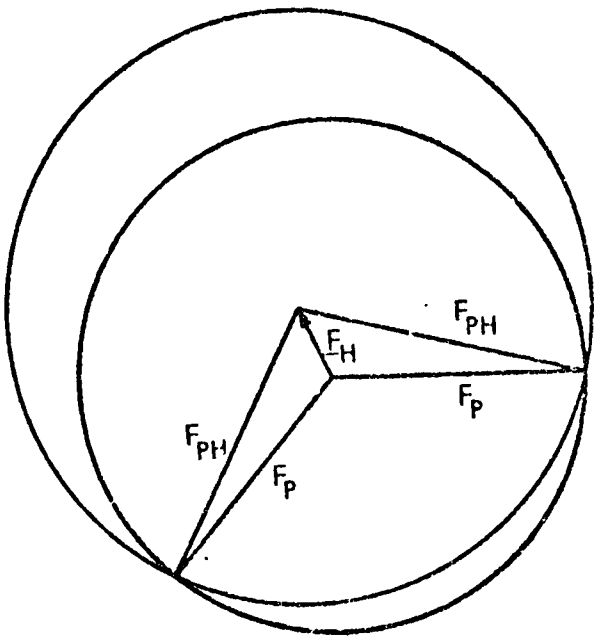


FIGURE 1.1 A single isomorphous derivative: the vector sum $\underline{F}_{PH} = \underline{F}_P + \underline{F}_H$ is a closed triangle, the magnitudes of all 3 sides are known, but only the direction of $\underline{F}_H$. This has two possible solutions for the phase of $\underline{F}_P$.

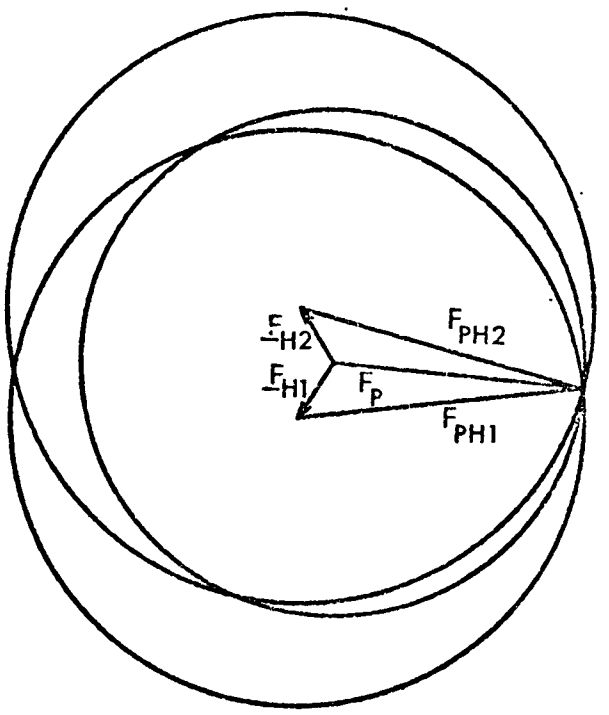
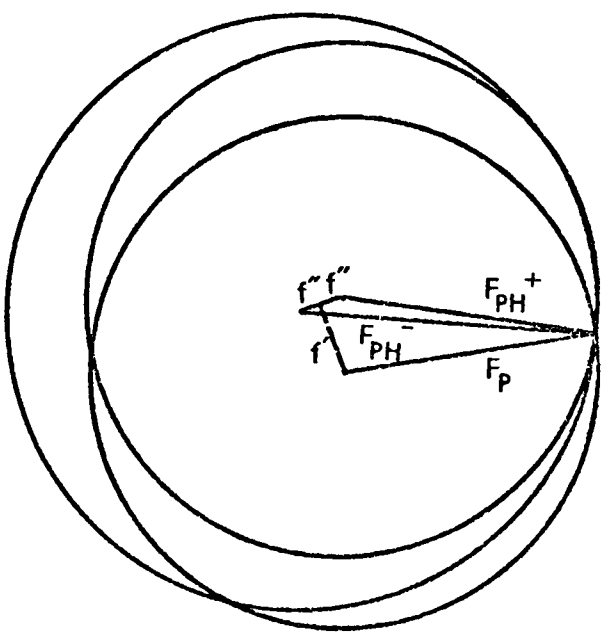


FIGURE 1.2 Two isomorphous derivatives: the second derivative also has 2 possible solutions for the protein phase, one of these coincides with the true solution from the first derivative so that there is no phase ambiguity.



$\underline{F}_H = f' + if''$

FIGURE 1.3 One derivative with anomalous scattering information: The structure amplitudes of the Bijvoet pairs, $\underline{F}_{PH}^+$ and $\underline{F}_{PH}^-$, behave like 2 separate derivatives and again the true protein phase is at the point of intersection of all 3 circles.

(Figure 1.3). When anomalous scattering occurs, the structure factor can be written as $\underline{F}_H = f'+if''$ (Section 3c) and Figure 1.3 shows how the combined information from isomorphous replacement and anomalous scattering allows an unambiguous determination of the phase (Peerdeman *et al.*, 1951).

In practice errors generally prevent the structure factor circles in the Harker construction from intersecting at a unique point. The probability analysis of Blow and Crick (1959) takes into account such errors of measurement. For acentric reflections all errors are assumed to lie in $\underline{F}_{PH}$. The probability that the phase angle is correct is

$$P(\alpha) = \exp(-\varepsilon(\alpha)^2/2E^2) \qquad 1.17$$

where $\varepsilon(\alpha) = |\underline{F}_{PH}|-|\underline{F}_P+\underline{F}_H|$, the "lack of closure" error, and E is the r.m.s. error in $\underline{F}_H$:-

$$\langle E^2\rangle = \langle(|\underline{F}_{PH}|-|\underline{F}_P+\underline{F}_H|)^2\rangle$$

Anomalous scattering information is combined in a similar manner (North,1965; Matthews,1966b). For several sources of phase information the probabilities may be multiplied. Blow and Crick (1959) showed that an electron density map calculated with phases corresponding to the centroid of the probability distribution, and amplitudes weighted by the figure of merit (see below), has the least r.m.s. error. These phases are termed the "best phases", α_B. The Fourier coefficients are weighted by the figure of merit, m, calculated from:-

$$m\cos\alpha_B = \frac{\sum P(\alpha)\cos\alpha}{\sum P(\alpha)} \qquad m\sin\alpha_B = \frac{\sum P(\alpha)\sin\alpha}{\sum P(\alpha)} \qquad 1.18$$

The figure of merit is an indication of the reliability of the phase determination. The total r.m.s. error in the electron density in the unit cell is then:-

$$\langle \Delta\rho^2 \rangle = \frac{1}{V^2} \sum_{\underline{h}} F_P^2(\underline{h})(1-m^2) \qquad 1.19$$

(Blundell and Johnson, 1976)

5) THE DIFFERENCE FOURIER SYNTHESIS

The difference Fourier synthesis provides an effective method for locating the binding sites of small molecules such as substrates or effectors to the protein. It can also be used to analyse new heavy atom derivatives when estimates of the protein phases are available. The difference Fourier synthesis, $\Delta\rho$, is defined by:-

$$\Delta\rho = \frac{1}{V} \sum_{h} m(F_{PI}-F_P)\exp i\alpha_P \exp(-2\pi i \underline{h}.\underline{x}) \qquad 1.20$$

where I represents the bound molecule.

The true electron density for the bound molecule has Fourier coefficients $F_I \exp i\alpha_I$. In the case of a centrosymmetric structure $(F_{PI}-F_P)\exp i\alpha_P$ is equivalent to $F_I \exp i\alpha_I$ except for rare cases of "cross-over" and so the bound molecule will be represented at its full height in the difference Fourier map. For the non-centrosymmetric case, a synthesis based on coefficients $(F_{PI}-F_P)\exp i\alpha_P$ will contain features which approximately represent additional atoms not included in the phasing, but at half their true heights. The relative contribution of acentric and centric terms is partially compensated by using coefficients $2(F_{PI}-F_P)$ in the difference Fourier synthesis (Luzzati, 1953).

Errors in the native protein phases will also cause the peak heights in the difference Fourier map to be reduced. If these

errors are reflected in the figure of merit, m, then the mean contribution of the structure factor $m(F_{PI}-F_P)\exp i\alpha_P$ to the unknown structure is $m^2(F_{PI}-F_P)$ (Henderson and Moffat, 1971). Errors in the difference Fourier synthesis are attributable to three main sources:-

i) errors in the protein phases

ii) errors in the measurement of F_{PI} and F_P

iii) the use of coefficients $2m(F_{PI}-F_P)\exp i\alpha_P$ instead of $F_I\exp i\alpha_I$.

The overall r.m.s. error in the difference Fourier synthesis is:-

i) for centric terms

$$\langle\Delta\rho^2\rangle = \frac{\delta^2}{V^2} \qquad 1.21$$

ii) for acentric terms

$$\langle\Delta\rho^2\rangle = \frac{1}{V^2}\sum_{\underline{h}}(F_{PI}-F_P)^2 + \frac{\delta^2}{V^2} \qquad 1.22$$

(Weber 1977 from Blundell and Johnson, 1976)

where δ is the r.m.s. error in F_{PH} and F_P.

CHAPTER 2

GLYCOGEN PHOSPHORYLASE

1) PROPERTIES OF GLYCOGEN PHOSPHORYLASE

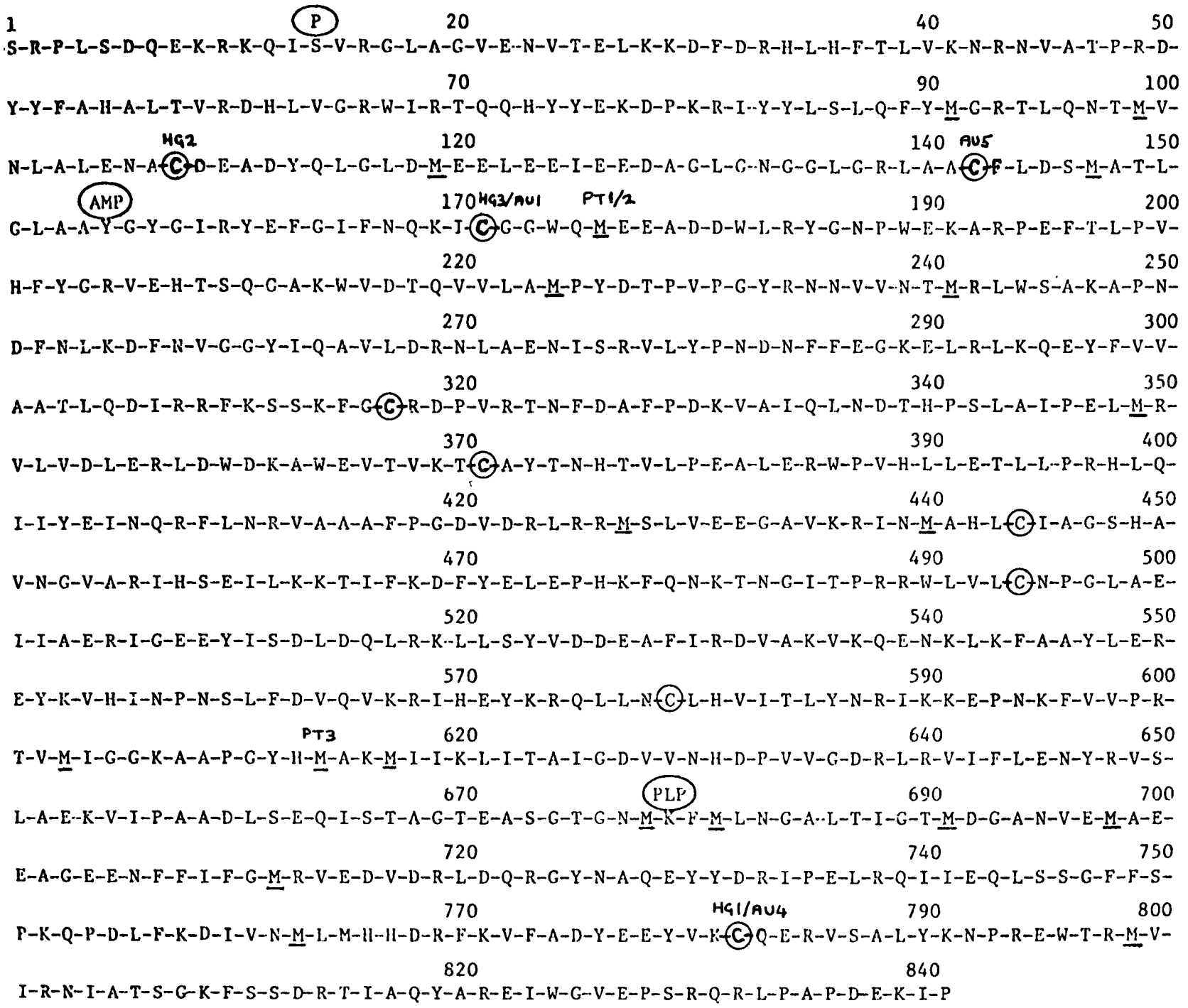
Glycogen phosphorylase (E.C. 2.4.1.1) occupies a vital position in the control of energy metabolism of living systems as it catalyses the first step in the anaerobic degradation of the carbohydrate energy store, glycogen, to pyruvate (the Embden-Meyerhof-Parnas glycolytic pathway). The reaction catalysed is:-

$$(\text{glycogen})_{n+1} + P_i \rightleftharpoons (\text{glycogen})_n + \alpha\text{-D-glucose-1-phosphate};$$

i.e. the sequential phosphorylisis of the $\alpha(1\text{-}4)$ linked glucose units until four sugars remain from an $\alpha(1\text{-}6)$ branch point.

Glycogen phosphorylases have been isolated from a wide spectrum of living organisms, but the enzyme derived from rabbit skeletal muscle has been the most extensively studied and is the subject of the present work. The enzyme has a subunit molecular weight of 97, 412 daltons. The sequence of 841 amino acids has been determined by Titani *et al.* (1977) and is shown in Figure 2.1. Each subunit contains the cofactor pyridoxal-5'-phosphate covalently linked to the amino group of lysine 679. The properties of the enzyme have been extensively reviewed (Graves and Wang, 1972; Fischer *et al.*, 1970, 1971). Phosphorylase exists in two interconvertible forms; phosphorylase *b*, found in resting muscle, which requires AMP, IMP or certain analogues of AMP for activity, and phosphorylase *a* which contains phosphate and is active in the absence of any effectors. The enzyme exhibits an equilibrium between dimer and tetramer; for phosphorylase *b* it favours the dimer and for phosphorylase *a* the tetramer, although the work of Metzger *et al.* (1967) shows that only dimeric phosphorylase *a* binds to glycogen. The monomer of

FIGURE 2.1



A	B	C	D	E	F	G	H	I	K	L	M	N
Ala	Asx	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn

P	Q	R	S	T	V	W	X	Y	Z
Pro	Gln	Arg	Ser	Thr	Val	Trp	X	Tyr	Glx

The amino acid sequence of rabbit muscle phosphorylase (from Titani et al., 1977). The sequence shows the positions of the coenzyme, PLP, the phosphate P, which is bound to phosphorylase a, and the binding site for the activator, AMP. Cysteine residues are in circles and methionine residues underlined. The probable binding sites for the heavy atom derivatives described in Chapter 4 are also marked.

phosphorylase is inactive, and is produced under rather harsh chemical conditions which often denature the protein. Reversible dissociation can be carried out with reagents such as para-chloro-mercuribenzoate (Madsen and Cori, 1956).

The economics of intermediary metabolism dictate that a particular pathway should actively function only as required. It is logical to control a pathway at the initial step to prevent the build up of intermediates. The glycolytic pathway is primarily controlled at two points. Firstly at the initial step of the degradation of glycogen by phosphorylase. Secondly at the important branch point from fructose-6-phosphate, where phosphofructokinase causes metabolites to follow the glycolytic path rather than one of the alternatives for hexose metabolism. In addition to its importance as a key glycolytic control enzyme, phosphorylase exhibits two of the three main mechanisms for the control of protein function. These are:-

i) long term control performed by activating or repressing the synthesis of the enzymes of the pathway at the genetic or m-RNA level,

ii) medium or short term control can be exercised by the covalent modification of a protein to produce an active or inactive form as required: the time dependence of this type of control varies with the details of modif-ication process,

iii) short term control can be expressed by a response of the protein to the changing levels of metabolites within the local environment (usually the cell).

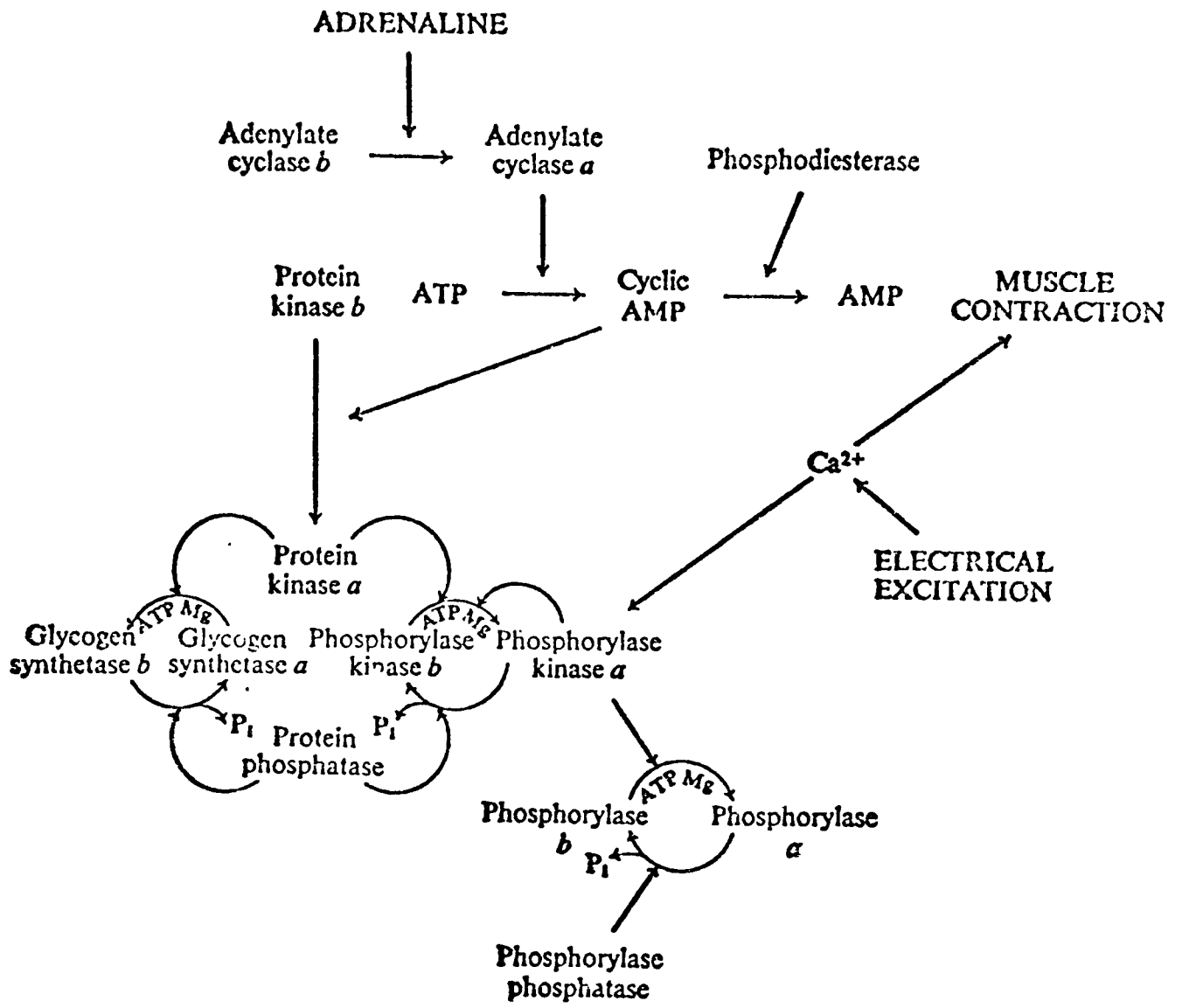
The control of phosphorylase is exercised by systems of type two and three.

i) The interconversion of phosphorylase *b* and active phosphorylase *a* by the phosphorylation of a single amino acid (Serine 14) per monomer represents a control system of type two. This phosphorylation and dephosphorylation is catalysed by the enzymes phosphorylase kinase and phosphorylase phosphatase which are in turn activated by a complex cascade process under hormonal or nervous control (Figure 2.2). Under hormonal stimulation, adrenaline binds to a receptor in the cell membrane and activates the production of cyclic AMP, which in turn activates protein kinase. This enzyme then activates phosphorylase kinase and inactivates glycogen synthetase. Nervous stimulation of muscle produces Ca^{2+} ions which activate the Ca^{2+} dependent phosphorylase kinase.

ii) Changes in the intracellular levels of various metabolites not implicitly involved in the catalytic mechanism also affect the activity of phosphorylase (type 3 above). The principal allosteric activator is AMP, which can produce 80% of the full activity of phosphorylase *b*. This activation is opposed by the inhibitors ATP, ADP and G-6-P and by glucose. IMP is a less effective activator than AMP (Black and Wang, 1968). The 5'-phosphate group of AMP is essential for its action and the NH_2 group of the adenine ring is involved in the binding of AMP (Okazaki *et al.*, 1968). This may account for the differences between AMP and IMP. However, recent work by Rahim *et al.* (1976, 1980) suggests that in the absence of phosphorylase *a*, nucleotide activation by IMP becomes the important mechanism for phosphorylase activation. They show that in the muscles of phosphorylase kinase deficient I-strain mice, phosphorylase *b* activation is probably mediated by IMP formation, rather than AMP. IMP is present at a higher concentration than in normal muscles and its concentration rises during muscle work,

FIGURE 2.2

The Nervous and Hormonal Regulation of Glycogen Metabolism



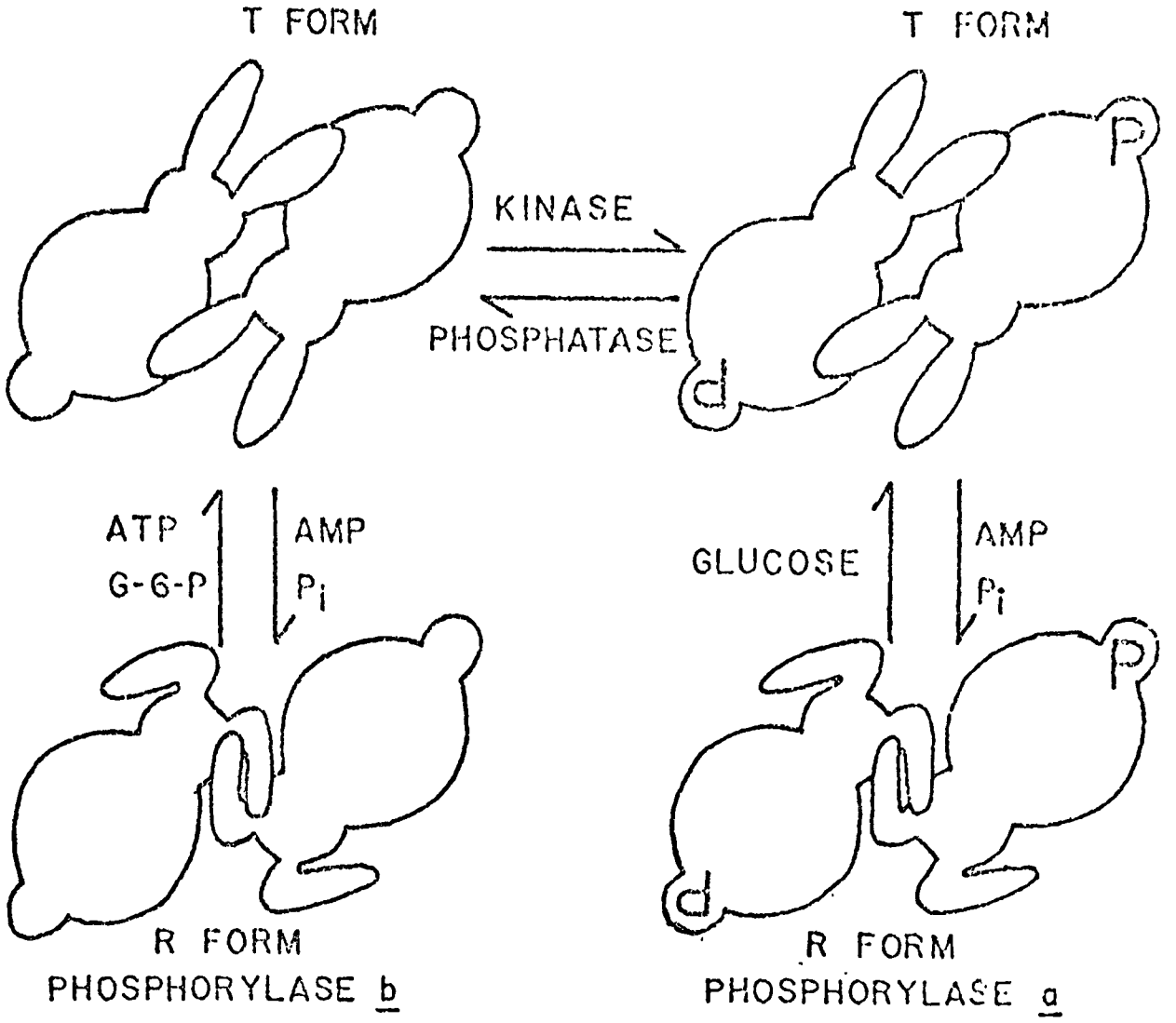
a represents the active enzyme and b the inactive enzyme. Phosphorylase is activated by the hormone, adrenaline, or by the electrical stimulation of muscle, and breaks down glycogen. The hormonal stimulation also causes inactivation of glycogen synthetase so that no new glycogen is formed.

attaining a concentration comparable to its activation constant for phosphorylase *b*.

The binding of the substrates G-1-P or P_i to phosphorylase *b*, cooperatively enhances the affinity for AMP. Phosphorylase *a* is active in the absence of metabolites and does not exhibit allosteric inhibition by ATP, ADP or G-6-P. However, the feedback inhibition by glucose remains. Although this has no physiological significance in muscle, which does not contain free glucose, it may have some significance in liver, which is almost freely permeable to glucose. According to Hers (1976), glucose inhibits liver phosphorylase *a*, and promotes its subsequent conversion to inactive phosphorylase *b*. This observation has led to the suggestion that phosphorylase may also act as a glucose receptor in the liver, in addition to its role in glycogen utilization. The interconversions of the various forms of the enzyme are shown schematically in Figure 2.3. The activation process can be considered in two stages; the addition of phosphate to phosphorylase *b* and the conversion of the inactive T form of phosphorylase *a* to the active R form by metabolites (in the terminology of Monod *et al*. 1965).

The glycolytic pathway assumed a prominent role at an early stage in evolution as a mechanism for obtaining energy from carbohydrates, and so phosphorylases from a wide range of species, which diverged from their common ancestors many millions of years ago, exhibit similar properties. The properties of phosphorylases from various species such as E. Coli, yeast, potato, rabbit and human have been compared by Fischer *et al*., (1970). Other phosphorylases which have been described are from slime mould (Thomas and Wright, 1976) and blowfly flight muscle (Childress and Sacktor, 1970). All these enzymes have a subunit molecular weight around 100,000. The enzymes from yeast, fungus and all the animals studied, exist in two forms; an inactive *b* form and a phosphorylated, active *a* form.

FIGURE 2.3



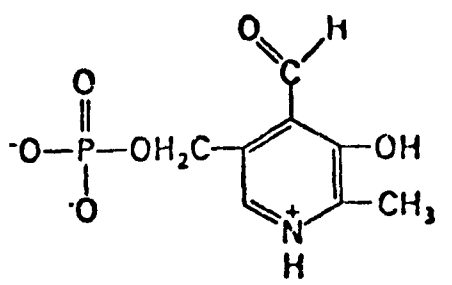
Schematic illustration of the relationship of the metabolically and allosterically interconvertible forms of the enzyme. The equilibrium between the tense (T) and relaxed (R) forms is delicately balanced by the concentrations of substrate molecules as well as positive and negative effectors.
(From Fletterick and Madsen, 1977).

Those from E. Coli, slime mould and potato exist in only one form. The PLP cofactor is present in all glycogen phosphorylases studied and is essential for activity. The homology of the amino acid sequences around the PLP site and serine-phosphate have been investigated. Rabbit and dog fish muscle phosphorylase show homology for both sites and yeast phosphorylase is also homologous for the PLP site (Lerch and Fischer, 1975). This conservation suggests that the cofactor does not play a purely structural role, which could be performed quite well by amino acids, but is directly or indirectly involved in the catalytic mechanism. PLP at acid pH is bound as a Schiff base to the ε-amino group of Lysine 679 (Figure 2.4) and, unlike other PLP-containing enzymes, remains active when the Schiff base is reduced by sodium borohydride (Fischer *et al.*, 1958). The reconstitution of phosphorylase with analogues of PLP has shown that the groups at positions 2,3 and 6 of the pyridoxal ring are not essential for activity and that the CHO group at position 4 is required for binding (Fischer *et al.*, 1970). The 5'-phosphate group is essential for activity and may act as a proton or glycosyl donor-acceptor (Feldman and Helmreich, 1976; Feldman *et al.*, 1978). Fluorescence studies suggest that the PLP is buried in the interior of the molecule in a hydrophobic environment (Shaltiel and Cortijo, 1970; Jones and Cowgill, 1971). Honikel and Madsen (1973) suggest that the PLP is in a partially open hydrophobic pocket, which is closed by substrates and activators to give a fully active state. As a result of such reconstitution studies with pyridoxal and modified pyridoxal phosphates an important catalytic role for the phosphate moiety of the coenzyme has been established (Graves and Wang, 1972; Graves *et al.*, 1978). NMR studies using P^{31} resonance have shown that in the active form of the enzyme the phosphate is in its dianionic form and that the transition from the inactive to the active form involves

FIGURE 2.4

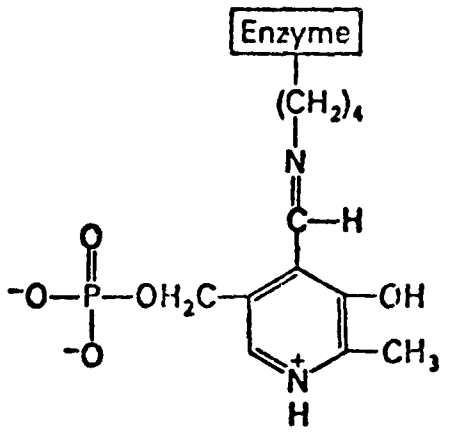
The Coenzyme Pyridoxal-Phosphate

a)



Pyridoxal phosphate
(PLP)

b)



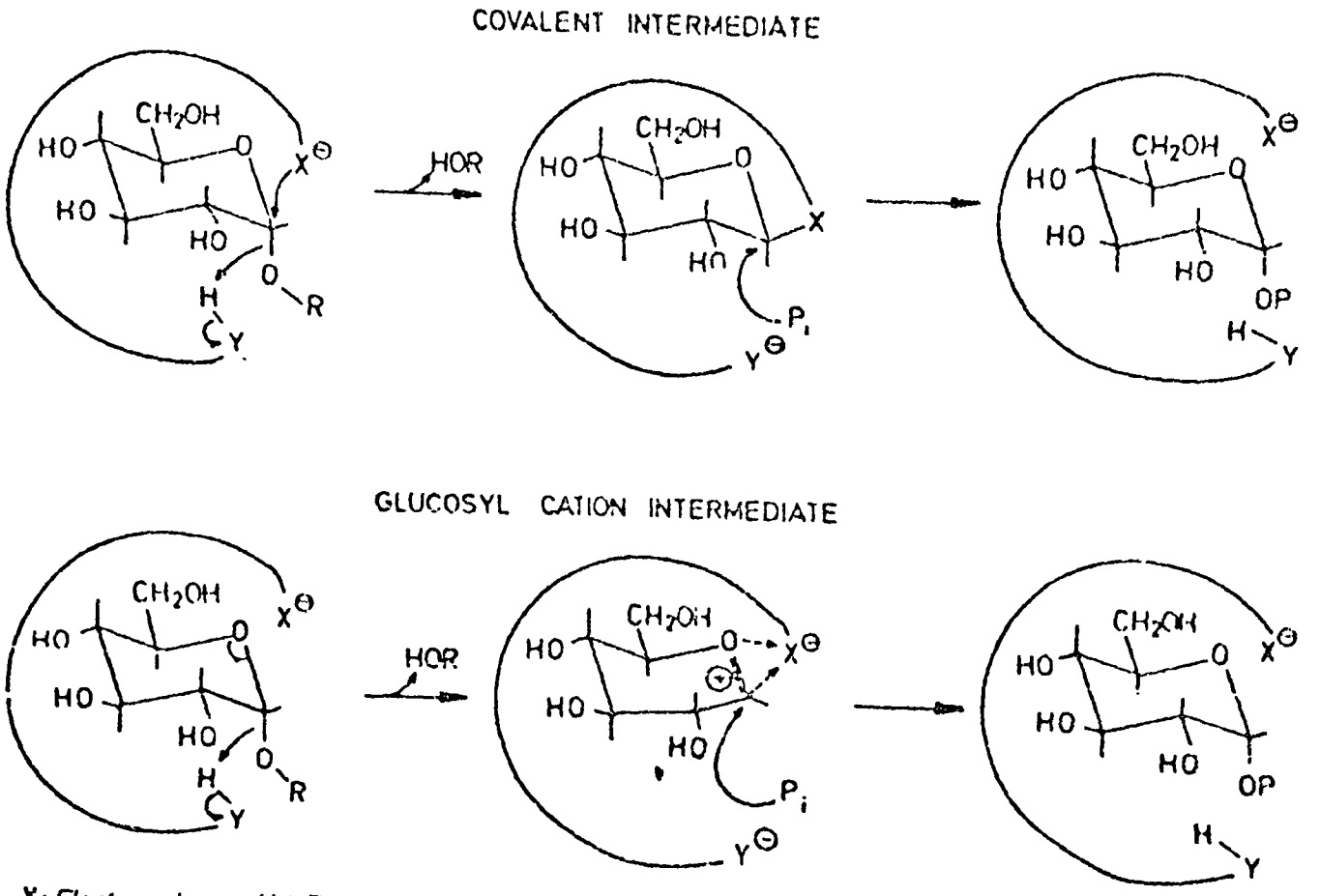
Schiff base of PLP
and enzyme

The coenzyme, pyridoxal-phosphate, is covalently bound to the ε-amino group of lysine 679 in phosphorylase. It is probably bound as a Schiff base as shown above.

deprotonation of the phosphate group (Feldmann and Helmreich, 1976; Feldmann and Hull, 1977; Feldmann *et al.*, 1978; Palm *et al.*, 1979). These studies have led to definite suggestions for the role of the phosphate in the reaction mechanism (Feldmann *et al.*, 1978). Possible mechanisms are shown in Figure 2.5. The double displacement reactions involve general acid catalysis and either the formation of a β-glucosyl-enzyme covalent intermediate (*cf.* Koshland, 1953) or a carbonium-oxonium ion intermediate stabilized by a nucleophilic group on the protein (*i.e*: analogous to lysozyme (Blake *et al.*, 1967). The dianionic form of the phosphate could thus function either as a base or a nucleophile. Results from X-ray crystallographic studies of glucose-1-phosphate binding to phosphorylase *b* (Johnson *et al.*, 1980) show that the closest approach of the phosphate group of PLP to the glycosidic oxygen of G-1-P is 5.3 Å. This is sufficiently close to suggest an important role for the PLP, but the relative orientations of the cofactor and substrate are inconsistent with the role of the PLP phosphate in its dianionic form. In addition, a glycogen substrate could not be accommodated at the active site in this conformation. Johnson *et al.*, (1980) point out that it is not easy to envisage conformational changes of the enzyme which overcome these difficulties. However, they have shown by model building studies that an alternative binding mode for G-1-P is possible, which allows a glycogen substrate to fit easily. This alternative orientation leads to a proposed mechanism in which the phosphate group of the PLP acts as a nucleophile and the imidazole of Histidine 376 functions as a general acid to protonate the glycosidic oxygen.

Phosphorylase is found in particles isolated from muscle cells which are rich in glycogen (Meyer *et al.*, 1970, Haschke *et al.*, 1970). In addition these particles also contain other enzymes involved in glycogen metabolism. Glycogen is a branched polymer of α(1-4)

FIGURE 2.5



X: Electron donor; HY. Proton donor; R: $(Glucosyl)_n$; P_i: Orthophosphate;

Possible double displacement mechanisms for phosphorylase. The reactions involve general acid catalysis and either a covalent β-glucosyl-enzyme intermediate (upper diagram) or a carbonium-oxonium ion intermediate (lower diagram)(from Feldmann et al.,1978).

linked glucose units with α(1-6) linked branch points. Phosphoryl-ase degrades glycogen by removal of single residues from the non-reducing end and stops four residues before an α(1-6) branch point. Phosphorylase *in vitro* also catalyses the synthesis of glycogen from a primer of at least 4 glucose units.

2) CRYSTALLOGRAPHIC ANALYSIS OF GLYCOGEN PHOSPHORYLASE

Independent crystallographic studies are being carried out on phosphorylase *b* in Oxford and phosphorylase *a* in Edmonton with the aim of providing a structural explanation for the catalytic and control properties of the enzyme. A favourable tetragonal crystal form of phosphorylase *b* was discovered in 1972 (Eagles *et al.*, 1972). The crystals used for the analysis described here were grown in the presence of the activator IMP although phosphoryl-ase *b* crystals may also be grown without IMP. The crystallographic parameters are given in Table 2.1. The asymmetric unit contains one subunit of phosphorylase and the two subunits of the dimer are related by the crystallographic twofold axis at z = ½. Phosphoryl-ase *a* has been found to crystallise in the same crystal form in the presence of the inhibitor 50 m*M* glucose (Fletterick *et al.*, 1976a). A comparison of the electron density maps for phosphorylase *a* and *b* at 6 Å resolution showed that the overall structures of the molecules were remarkably similar but that a large shift in the position of the first 20 amino-terminal amino acids occurred with the addition of the phosphate group to Serine-14 (Fletterick *et al.*, 1976a). The structures of both forms have been solved at 3 Å resolution (Fletterick *et al.*, 1976b; Johnson *et al.*, 1978) and more recently at 2.5 Å for phosphorylase *a*. (Sprang and Fletterick, 1979) The crystallographic work on phosphorylase *a* has been recently reviewed (Fletterick and Madsen, 1980) and the binding of various metabolites to both forms has also been extensively studied (Weber

TABLE 2.1

The Tetragonal Crystals of Phosphorylase b

Space Group: $P4_32_12$

Unit cell parameters: $a = b = 128.5\text{Å}$, $c = 115.9\text{Å}$

Volume of unit cell: $V = 1.94 \times 10^6\ \text{Å}^3$

Asymmetric unit: one subunit of phosphorylase b

Molecular weight 97,000 daltons

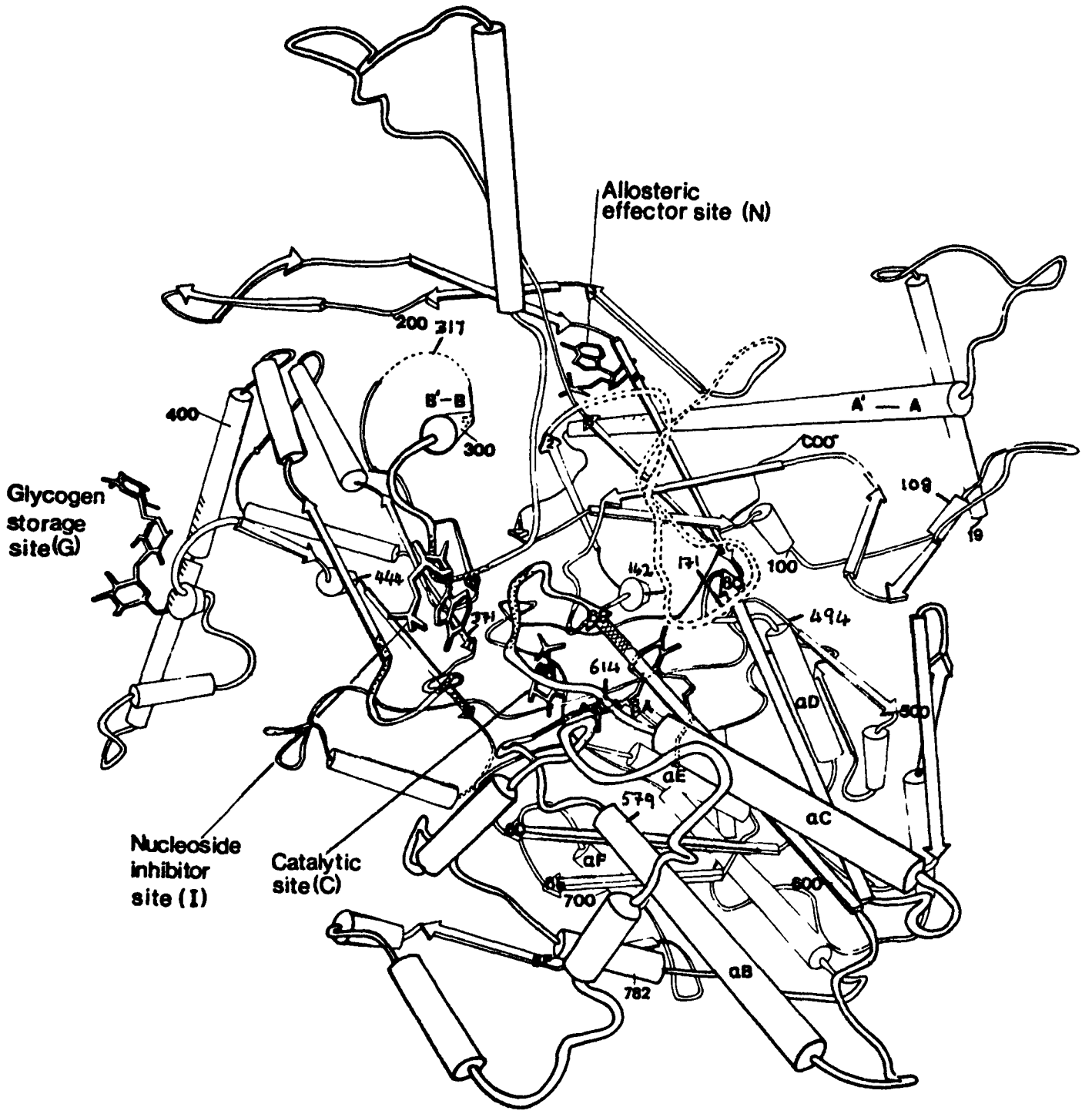
Equivalent positions in the unit cell:

x , y , z	$y , x , \bar{z}$
$\bar{x} , \bar{y} , \frac{1}{2} + z$	$\bar{y} , \bar{x} , \frac{1}{2} - z$
$\frac{1}{2} - y , \frac{1}{2} + x , \frac{3}{4} + z$	$\frac{1}{2} - x , \frac{1}{2} + y , \frac{3}{4} - z$
$\frac{1}{2} + y , \frac{1}{2} - x , \frac{1}{4} + z$	$\frac{1}{2} + x , \frac{1}{2} - y , \frac{1}{4} - z$

et al., 1978; Fletterick *et al.*, 1976b; Johnson *et al.*, 1979).

A schematic drawing of the phosphorylase *b* molecule, based on the 3 Å electron density map is shown in Figure 2.6. This map was interpreted with the aid of the amino acid sequence of 841 residues determined by Titani *et al.* (1977). The first 19 residues appear to be flexible in phosphorylase *b* (Johnson *et al.*, 1979). The overall shape of the subunit is compact, except for two loops (the "cap", residues 36-45 and the "tower", residues 244-273) which cross the twofold axis z = ½ and extend into the symmetry related subunit of the dimer. The structure may be described in terms of three domains. The first 311 residues form a self-consistent fold termed the N-terminal domain. This domain has as its core a 5-stranded β-sheet and contains two long α-helices AA' and BB' which flank the nucleotide effector site (shown binding AMP). The β-sheet continues with strands 6 and 7 and a few hydrogen bonds from strand 8 of the next domain, the glycogen binding domain (residues 322-490). This domain contains a strong glycogen binding site (Weber *et al.*, 1978, Johnson *et al.*, 1980) on the surface of the molecule (shown binding maltotriose), and probably represents the site at which the molecule binds to glycogen in the glycogen particle. Both the N-terminal domain and the glycogen binding domain form self-consistent folds, but in view of the close binding between them, they could be considered as two sub-domains of a large 490-residue domain, and Sprang and Fletterick (1979) describe the similar structure of phosphorylase *a* in this fashion. The remaining part of the molecule (residues 490-841) forms the C-terminal domain, which contains a region of 6 parallel strands of sheet and helices similar to that found in lactate dehydrogenase (Rossmann *et al.*, 1975). A similar region has been found in phosphorylase *a* and a least-squares comparison of the best orientation of this domain relative to the NAD domain in lactate dehydrogenase shows that 99 of the α-carbons are within

FIGURE 2.6



A schematic perspective drawing of the phosphorylase b molecule viewed down the crystallographic y axis. α-helices and β-strands are represented by cylinders and arrows, respectively. The fold divides into 3 parts, the N-terminal domain (top right); the glycogen binding domain (left); and the C-terminal domain (bottom right). The strands and helices of the "nucleotide binding domain" which is located at the centre of the C-terminal domain are lettered A to F. AMP is shown bound at sites N and I, and G-1-P at the catalytic site. Maltotriose is shown bound at the glycogen storage site. The pyridoxal phosphate is approximately 8 Å from the G-1-P and linked to Lys 679 at the start of the E-helix. The positions of the cysteine residues are also marked. From Johnson et al., (1979).

3.4 Å of their mates, although the two adenine sites are not identical (Fletterick and Madsen 1977). In Figure 2.6 the strands and helices are labelled A to F according to the convention used for the dehydrogenases. The overall folding of phosphorylase *a* shows close similarity to phosphorylase *b*, with the same domain structure. Sprang and Fletterick (1979) point out that although, in a large protein, a larger and more representative sample of folding possibilities might be expected, phosphorylase exhibits a similar folding pattern to smaller proteins. It exhibits a topology and packing typical of the class of proteins referred to as α/β by Levitt and Chothia (1976). The cores of the domains are β-sheet, surrounded by α-helices on the "external" parts of the molecule.

Several metabolite binding sites have been located on both phosphorylase *a* and *b* and are shown in Figure 2.6.

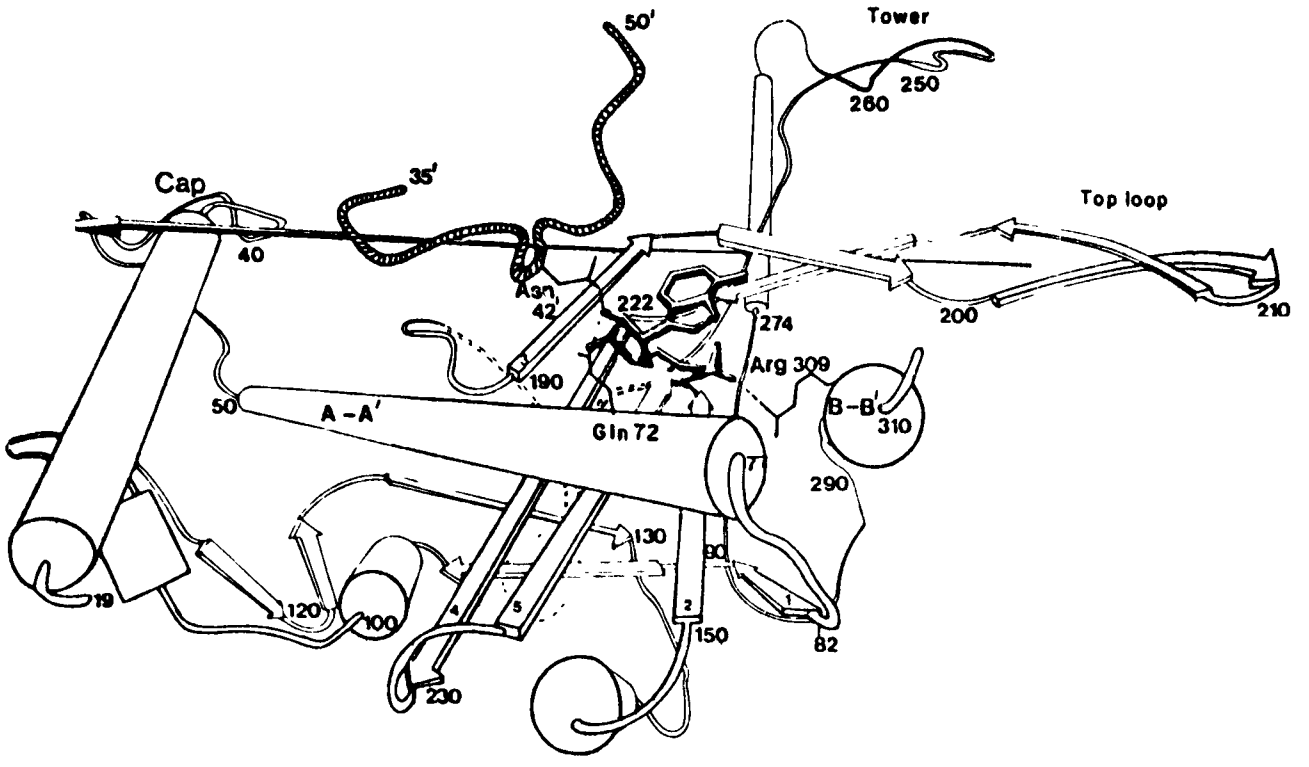
The catalytic site (shown binding G-1-P) has been located at the centre of the molecule at the junction of the three domains (Weber *et al.*, 1978). It is approximately 8 Å from the essential pyridoxal phosphate, which is linked *via* a Schiff base to Lysine 69 at the beginning of the αE helix. The catalytic site is buried in a hydrophobic region at the bottom of a deep crevasse approximately 20 Å long and is only accessible to the solvent *via* this channel. The glycogen storage site described earlier is located on the surface of the molecule approximately 33 Å from the catalytic site and 40 Å from the allosteric effector site (Site N) which is situated in the N-terminal domain close to the subunit-subunit interface. Nucleotides bound at site N make several interactions with the enzyme (Johnson *et al.*, 1979). These contacts involve the N10 position of the base, the phosphate, which makes an ionic contact with Arg 309, and the 2' hydroxyl of the ribose, which appears to make contact with Gln 72 on the AA' helix of its own subunit and a residue on the cap of the symmetry related subunit

(Figure 2.7). A nucleoside inhibitor site (site I) has also been located at the entrance to the crevasse which leads to the catalytic site. This site was first identified as the inhibitor site from studies on phosphorylase *a* (Fletterick *et al.*, 1976b; Kasvinsky *et al.*, 1978a,b). Binding at this site blocks access to the catalytic site.

The phosphates of AMP and ATP bound at site I in phosphorylase *b* do not make any specific contacts with the enzyme.

There is general agreement on the location of these four binding sites for phosphorylase *a* and phosphorylase *b* but with some differences. In particular, although G-1-P binds readily to the catalytic site in phosphorylase *b*, in phosphorylase *a* binding is only achieved at high concentrations and with pronounced conformational changes. Conformational changes have also been reported when phosphorylase *a* crystals, grown in the presence of the inhibitor glucose, are cross-linked and then reacted with substrates and activators (Madsen *et al.*, 1978). The crystals crack and then reanneal, a phenomenon not observed with phosphorylase *b*. The conformational changes which accompany the activation of phosphorylase *a* mainly involve residues in the N-terminal half of the molecule, and the differentiation of structure between the domains is clearly relevent. These changes involve the whole of the N-terminus including the helix AA' and the allosteric effector site (site N). Changes also occur in the tower region which penetrates the dimer interface. For phosphorylase *b* (at 3 Å resolution) no well defined conformational changes are observed on binding the activator AMP, and it is suggested by Johnson *et al.* (1979) that the forces which stabilise the crystal lattice of phosphorylase *b* prevent the allosteric response of the enzyme in the crystal. High resolution structures for both phosphorylase *a* and *b* should help to elucidate the molecular basis for the changes in enzyme properties brought

FIGURE 2.7



A schematic perspective drawing of the N-terminal domain of phosphorylase b as viewed down the crystallographic y axis (i.e. the opposite direction to that shown in Figure 2.6). The allosteric activator AMP is shown bound at the nucleotide effector site, together with some of its most important contacts to the enzyme. The 2-fold rotation axis which relates the two subunits of the dimer is shown by the darkened arrow and is inclined 45° to the plane of the page. The cap of the symmetry-related subunit is shown shaded. From Johnson et al., (1979).

about by the interconversion of the two forms; the homotropic and heterotropic cooperativity observed with substrates and allosteric effectors and the subunit interactions within the dimer. In addition, the catalytic mechanism and the role of the cofactor PLP should also become more apparent. The work described here represents the first step towards this goal, with the collection and processing of 2 Å data for phosphorylase *b*.

CHAPTER 3

THE OSCILLATION METHOD AND THE PROCESSING OF PHOTOGRAPHIC DATA

Introduction

A major problem in the determination of the crystal structure of proteins which crystallise in large unit cells is the large number of reflections, many of weak intensity, which must be collected. The main types of data collection used are photographic methods, or counter methods using a diffractometer. The main advantage of photographic methods is that a large number of diffracted intensities may be recorded simultaneously, whilst counter methods have the advantage of accurate and automatic measurement of the integrated intensities. However, with a diffractometer each diffracted beam must be measured separately and this leads to long exposure times and subsequent radiation damage problems.

Screenless photographic methods have been assessed by Arndt (1968). These methods do not use a layer-line screen since it limits the reflections recorded to one plane of the reciprocal lattice. The crystal oscillates or precesses through a small angle and all reflections which pass through the Ewald sphere in this range are recorded on film. Arndt (1968) calculates that a 1° rotation of a crystal with a 100 Å unit cell axis enables one reflection to be measured on a four-circle diffractometer, approximately 17 reflections to be recorded on a screened precession photograph or 1560 reflections to be recorded with screenless photography at 2.25 Å. The main disadvantage of screenless photography is that the films must be digitized and the intensities integrated by computer. The recent development of electronic area detector methods (see below) should overcome this problem. Data collection methods have been described using both screenless

precession (Xuong *et al.*, 1968; Xuong and Freer, 1971) and screenless oscillation geometry (Arndt 1968; Arndt *et al.*, 1973). Initially the precession method was favoured because existing instruments could be used. However, the oscillation camera offers distinct advantages.

a) Lorentz Factor

The Lorentz factors for the oscillation and precession cameras are given in Table 3.1 and the variation of the Lorentz factor with ξ is shown in Figure 3.1 (both from Weber, 1977). This graph shows that the Lorentz factor for the oscillation camera varies slowly across the film (except near the rotation axis). However, for small angle precession photographs, the Lorentz factor varies rapidly near the blind region (ξ_{min}) and the outer limit (ξ_{max}). Thus the Lorentz factor varies over each reflection in these regions and each density point within a reflection must be corrected separately.

b) Partial Reflections

Partially recorded reflections are recorded at the beginning and end of the precession or oscillation range. In the precession method the partially recorded reflections cannot be used because the Lorentz factor is very large and varies rapidly at the edges of the range, and precessing about adjacent angular ranges on the spindle axis does not sweep out adjacent volumes of reciprocal space. With the oscillation method the Lorentz factor is easy to apply, adjacent rotation ranges sweep out adjacent volumes of reciprocal space, and the rotation angle ϕ is known accurately. This means that the partially recorded reflections at the end and beginning of adjacent ranges may be summed to give effectively more intensity measurements. Crystals with large unit cells require a

TABLE 3.1

The Blind Region

The limiting radius of the blind region is $F\xi_{min}$, where F is the crystal to film distance and ξ_{min} is:

a) Oscillation Camera $\xi_{min} = 1 - (1-\zeta^2)^{\frac{1}{2}}$

b) Precession Camera $\xi_{min} = \sin\bar{\nu}_n - \sin\mu$

The precession camera also has an outer limiting radius for diffraction from upper levels of $F\xi_{max}$, where $\xi_{max} = \sin\bar{\nu}_n + \sin\mu$

The Lorentz Factor

Rotation Camera $L^{-1} = (\sin^2 2\theta - \zeta^2)^{1/2}$

Precession Camera $L^{-1} = A^{-1}B^{-1}$

where $A^{-1} = \sin\Upsilon \sin\bar{\nu}_n \cos\bar{\mu} \sin\bar{\mu}$

$= 4\cos\bar{\mu}[4\xi^2 \sin^2\bar{\mu} - (\sin^2\bar{\nu}_n - \sin^2\bar{\mu} - \xi^2)^2]^{1/2}$

$$B^{-1} = \left[\frac{1}{1+\tan^2\bar{\mu}\sin^2(\tau+\alpha)} + \frac{1}{1+\tan^2\bar{\mu}\sin^2(\tau-\alpha)}\right]^{-1}$$

where $\bar{\mu}$ is the precession angle

$\pi/2 - \bar{\nu}_n$ is the semi-angle of the cone of diffraction

$\cos\bar{\nu}_n = \cos\bar{\mu} - \zeta_n$

Υ is the projection of 2θ on to the zero layer

$$\cos\Upsilon = \frac{2 - \zeta_n^2 - \xi^2}{2(1-\zeta_n^2)^{1/2}}$$

τ is the angle between the horizontal axis of the universal joint and the projection of the reciprocal lattice vector h on to the zero layer.

α is an angle which describes the passage of the reciprocal lattice point through the reflecting sphere.

$$\cos\alpha = \frac{\xi^2 + \sin^2\bar{\mu} - \sin^2\bar{\nu}_n}{2\xi\sin\bar{\mu}}$$

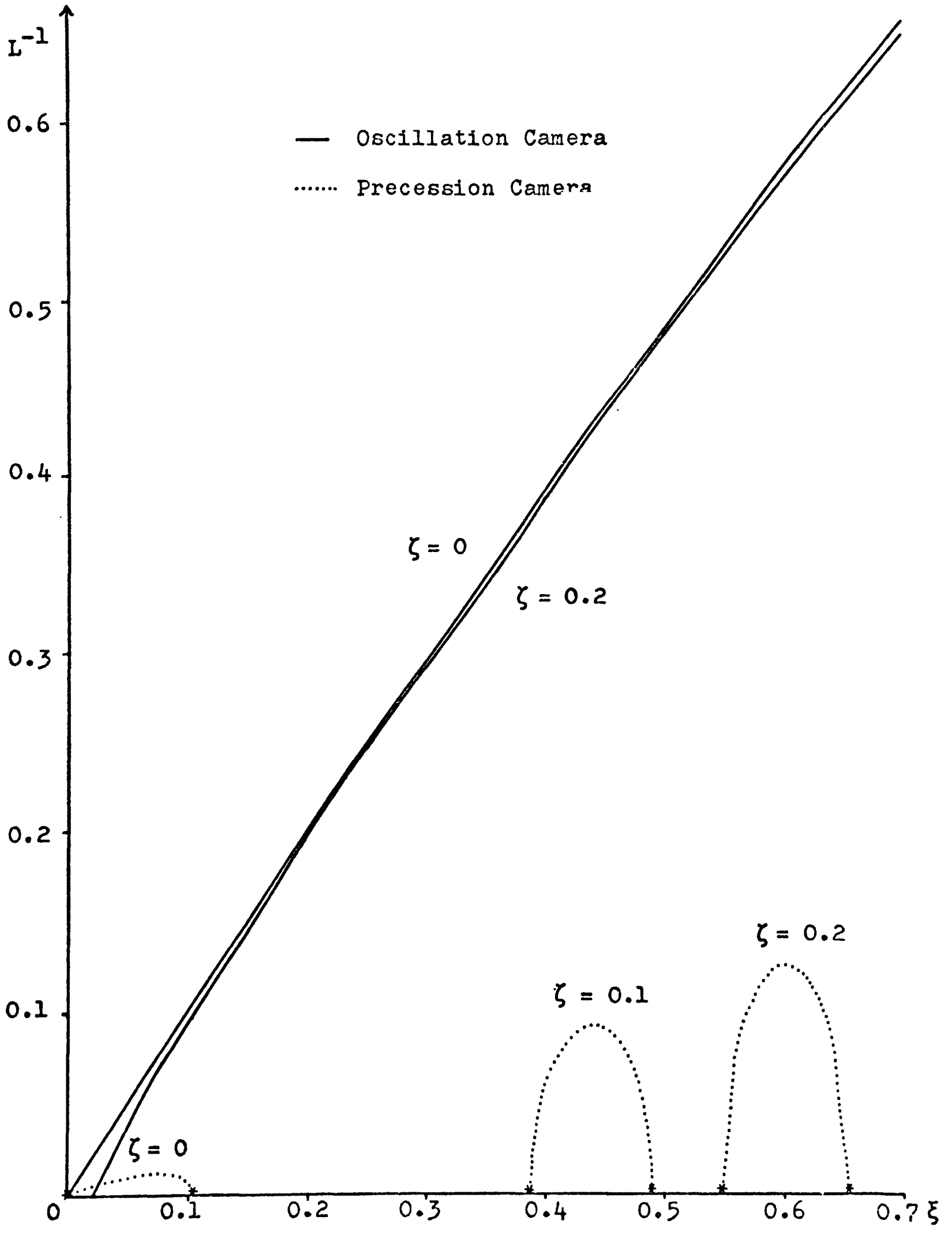
The Polarization Correction

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

FIGURE 3.1

The Variation of the Lorentz Factor with ξ

The Lorentz factors are given in Table 4.2, the values plotted for the precession camera have angle $\tau = 0^{o}$ and precession angle $\mu = 3^{o}$. The Lorentz correction for the precession camera varies rapidly near ξ_{min} and ξ_{max} (*).



smaller rotation range to minimize overlapping reflections, but since the reflection size is effectively constant, more partially recorded reflections will be produced. The summation of partially recorded reflections is thus especially useful for large unit cells.

The advantages of the oscillation method are also discussed by Milledge (1966) and Arndt (1968). A recent collection of papers (Arndt and Wonacott, 1977) provides an excellent review.

The screenless oscillation method requires a good area detector which records the intensities and positions of reflections efficiently. Although film is an accurate and sensitive detector of X-ray intensities, available at relatively low cost, it has some disadvantages. The diffraction intensities must be digitized by scanning the film on a microdensitometer and integrated by computer. Another drawback is that all temporal information in the diffraction pattern is lost. In order to keep to a minimum the total number of photographs required, the rotation range for any one photograph is usually chosen to be several times greater than the reflecting range of the crystal. Hence an area of film containing a diffraction spot accummulates "background" quanta during the whole exposure period, whilst "signal" quanta only arrive during the time corresponding to the reflecting range of the crystal. This means that the signal to noise ratio is lower than would be the case in an "ideal" diffraction experiment. The use of electronic area detectors is an attempt to perform such an "ideal" experiment. The main requirements for an area detector (Arndt and Faruqi in Ardnt and Wonacott, 1977) are:-

1) sensitivity to X-ray wavelengths
2) spatial resolution adequate for the pattern to be recorded
3) absorption efficiency close to unity
4) large dynamic range - capable of dealing with wide range of counting rates

5) low non-X-ray background

The two types of area detectors which have shown greatest promise are multi-wire proportional chambers (Cork *et al.* 1975) and scintillation screens with television-type read out (Arndt *et al.*, (1972); Hashizume *et al.*, (1975); Minor *et al.*, (1974)). Thomas and Hamilton (1972) have used an array of semiconductor detectors but at the present time the construction of two-dimensional arrays of adequate resolution would be too expensive.

Weber (1977) compared diffractometer and photographic methods for the collection of 3 Å data for phosphorylase *b*, and concluded that screenless photographic methods produce the four equivalent sets of reflections faster than a five-circle diffractometer measures one set of unique reflections. However, the diffractometer has the advantage that the intensities are measured and output in digital form immediately. Wilson and Yeates (1979) also compared the actual times taken for the collection and processing of 3 Å phosphorylase *b* data using the Enraf-Nonius oscillation camera with estimated times for a single-counter diffractometer. This comparison is shown in Table 3.2. The estimated time for the diffractometer is the time taken to collect two equivalent observations of each intensity, whilst the camera produces of necessity four equivalent terms. The oscillation camera is thus almost an order of magnitude faster than the diffractometer for the collection of data. The subsequent processing of the data is considerably slower for the photographic system but the combined collection and processing time is halved when compared to the diffractometer. The method of data collection chosen for phosphorylase *b* at 3 Å and 2 Å resolution was the Enraf-Nonius oscillation camera using conventional photographic film as detector. A number of other protein structures have also been determined by this method, including, *inter alia*; tyrosyl t-RNA

TABLE 3.2

The time required for collection and processing of a 3 Å data set for phosphorylase *b* using the Arndt-Wonacott oscillation camera compared with that for a single counter diffractometer.

	Camera	Diffractometer
Data collection	1 week	10 weeks
Scanning of films	24 h	-
Integration		
Real time	4 weeks	1 week
C.P.U. time	4 h	30 min
Reduction to unique data set		
Real time	1 week	1 week
C.P.U. time	2 h	2 h
Total real time	6 weeks	12 weeks

The diffractometer times are estimated by extrapolation of experience of 6 Å resolution (Johnson *et al.*, 1974). For both methods the times quoted are optimal and ignore problems such as unexpected deterioration of crystals during irradiation (from Wilson and Yeates, 1979).

synthetase (Irwin *et al.*, 1976), glyceraldehyde 3-phosphate dehydrogenase (Biesecker *et al.*, 1977) and pyruvate kinase (Levine *et al.*, 1977). These studies suggest that the quality of the data is at least comparable to that obtained using a diffractometer. A comparison has also been made for phosphorylase *b* (Wilson and Yeates, 1979). The merging R factor for 6 Å data collected by the oscillation method was typically 5.0% compared to 5.2% for data collected on a four-circle diffractometer (Johnson *et al.*, 1974).

Oscillation Geometry

The geometry of reflection is shown in Figure 3.2. The crystal is rotated about a horizontal rotation axis which is normal to the direct X-ray beam. This can be envisaged as the reciprocal lattice being rotated through the Ewald sphere so that individual reciprocal lattice points are brought into the reflecting position and the reflections recorded on film. The recorded diffraction patterns are interpreted using the programme suite OSCAR, described in the following sections. The crystal must be rotated through a total of $180^{\circ}+2\theta_{max}$ (where θ_{max} is the Bragg angle corresponding to the maximum resolution required) to record all of reciprocal space (Figure 3.3). However, this angle may be reduced by the symmetry of the space group. For phosphorylase *b* with space group $P4_32_12$, if the crystal is mounted about the 4-fold axis a total rotation of 45° is sufficient to provide four symmetry related equivalents of the unique terms. This is illustrated in Figure 3.4 for the case of data to 3 Å resolution. After rotation about one axis there remains a blind region of reciprocal space in which no reflections have been recorded. This is shown in Figure 3.5.

Figure 3.6 shows the percentage of the resolution sphere which is lost in the blind region. For data to 3 Å resolution about 2% of the reflections are lost, for 2 Å resolution about 5%. Data in

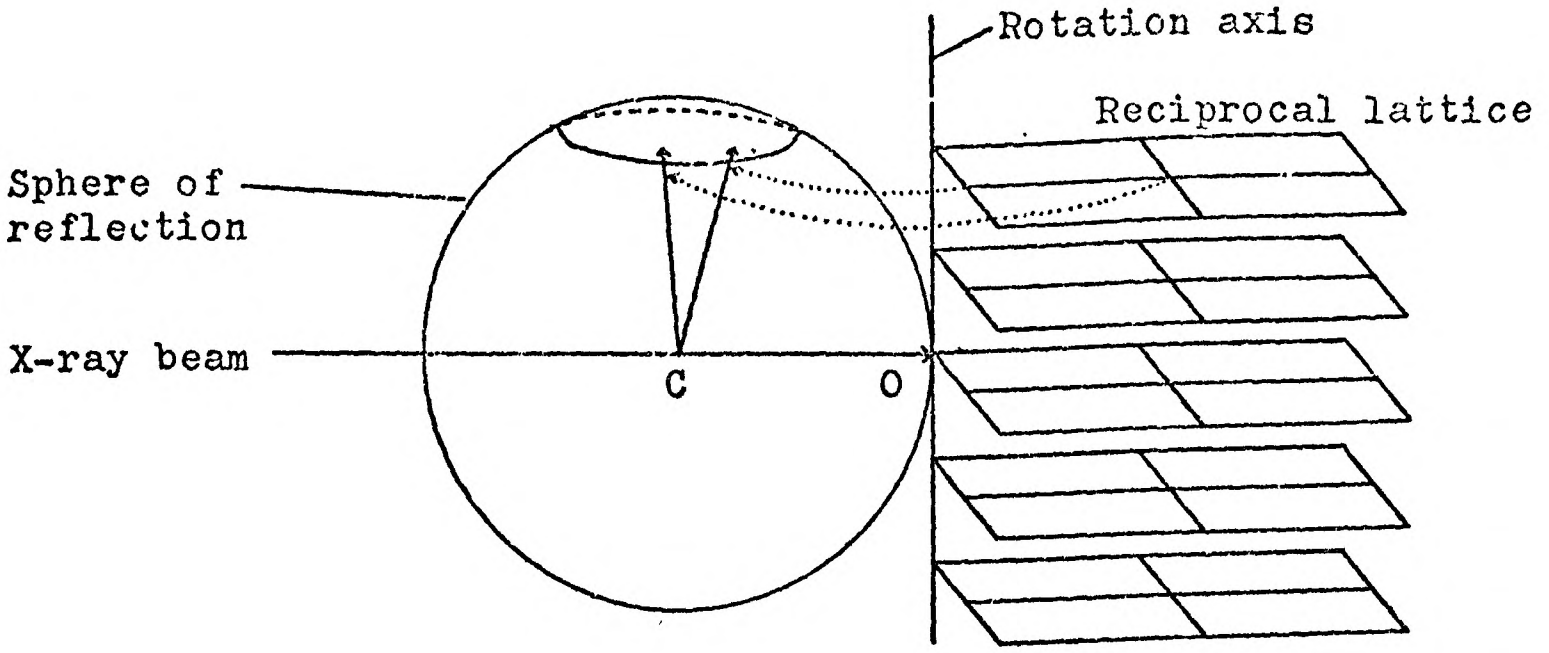


FIGURE 3.2 The Geometry of the Oscillation Camera

This shows part of the reciprocal lattice and indicates the manner in which reciprocal lattice points are brought into the reflecting position by a rotation about the rotation axis.

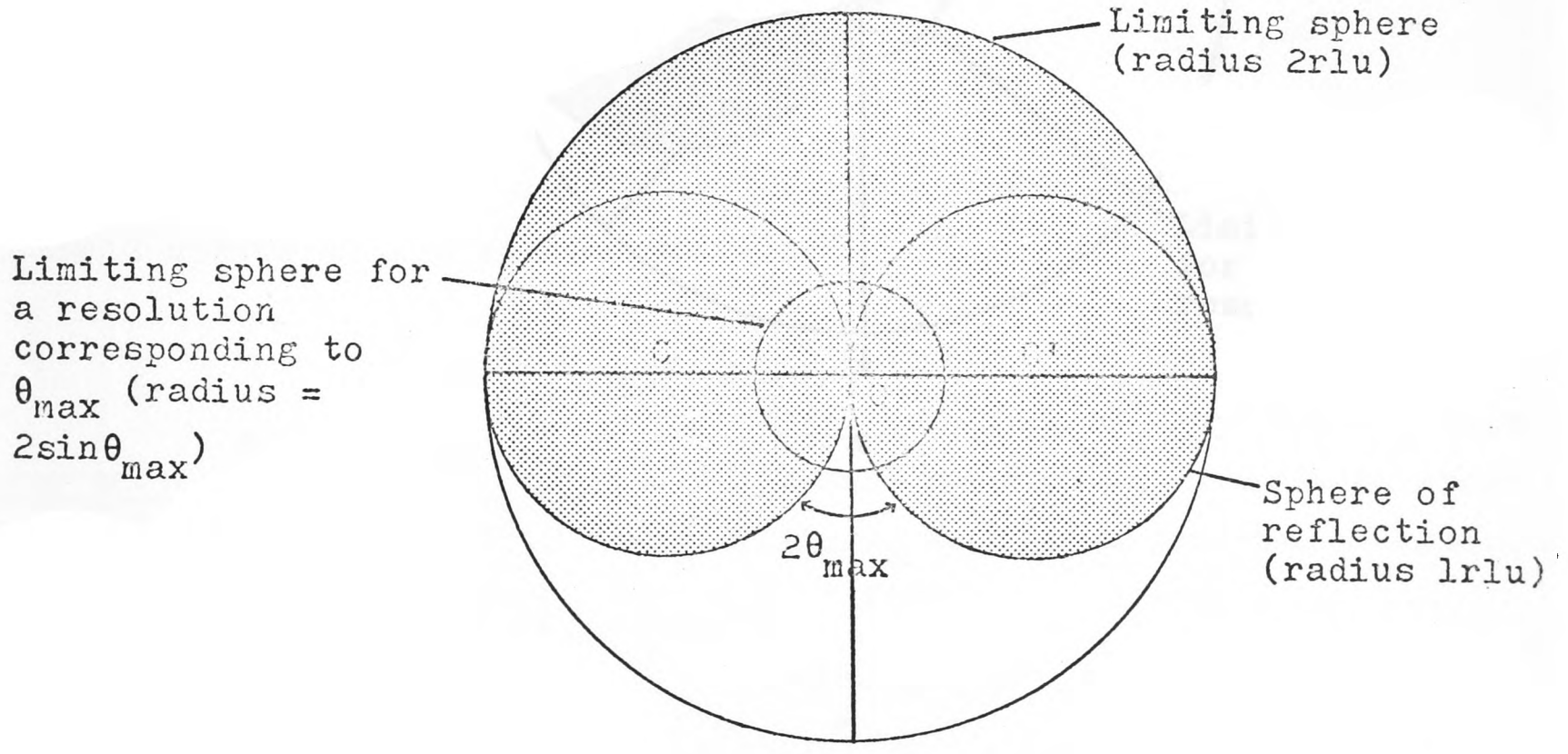
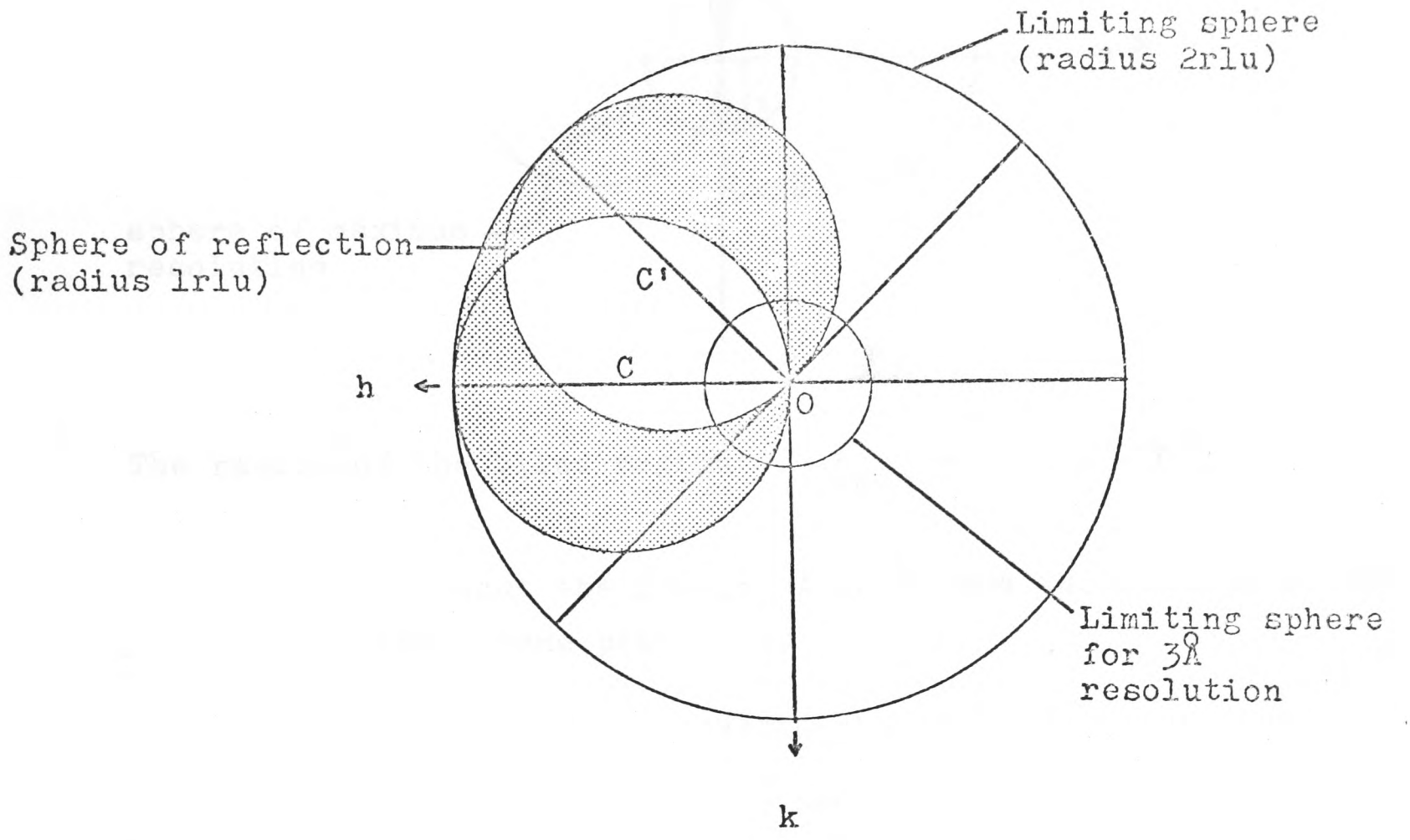


FIGURE 3.3 Projection down rotation axis showing the area of reciprocal space swept out as the crystal is rotated through 180^{o} (shaded area). The reciprocal lattice is fixed in the diagram and the centre of the sphere of reflection is rotated from C to C'. It is necessary to rotate the crystal through $180^{o} + 2\theta_{max}$ in order to record data to a resolution corresponding to θ_{max}.
The small circle corresponds to 3Å resolution.

FIGURE 3.4

The Total Rotation Required to Collect 3Å Data

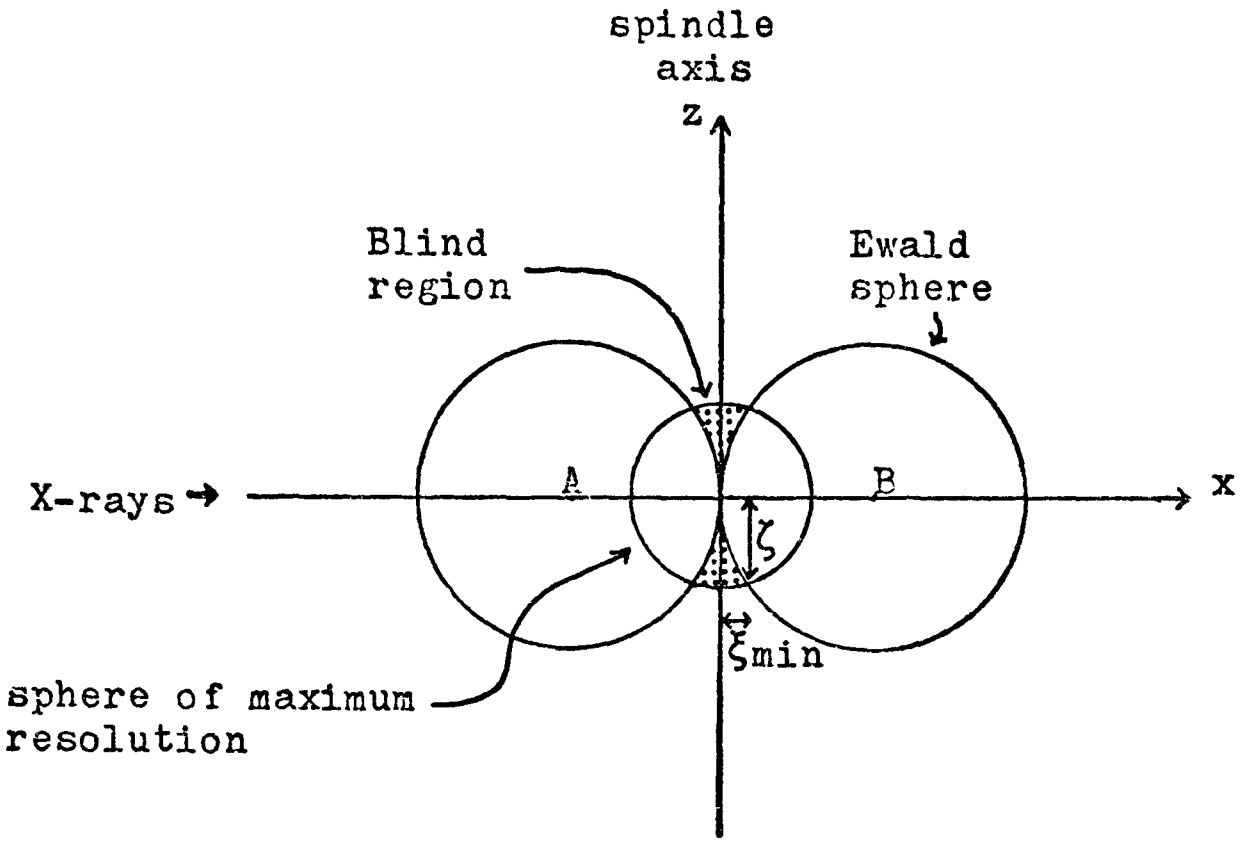
A projection down the rotation axis showing the area of reciprocal space swept out as the crystal is rotated through 45° (shaded area), or as the centre of the sphere of reflection is rotated from C to C'. The small circle corresponds to 3Å resolution. Data are recorded to 3Å resolution in a 45° rotation of a tetragonal crystal such as phosphorylase b of space group $P4_32_12$ and mounted with c parallel to the rotation axis.



Figures 3.2, 3.3 and 3.4 and part of Table 3.1 are taken from Blundell and Johnson (1976).

FIGURE 3.5

The Blind Region for Screenless Oscillation Photography



The radius of the blind region: $\xi_{min} = 1 - (1 - \zeta^2)^{\frac{1}{2}}$

The crystal rotates about the z axis; A and B are the centres of the Ewald sphere at its extreme positions.

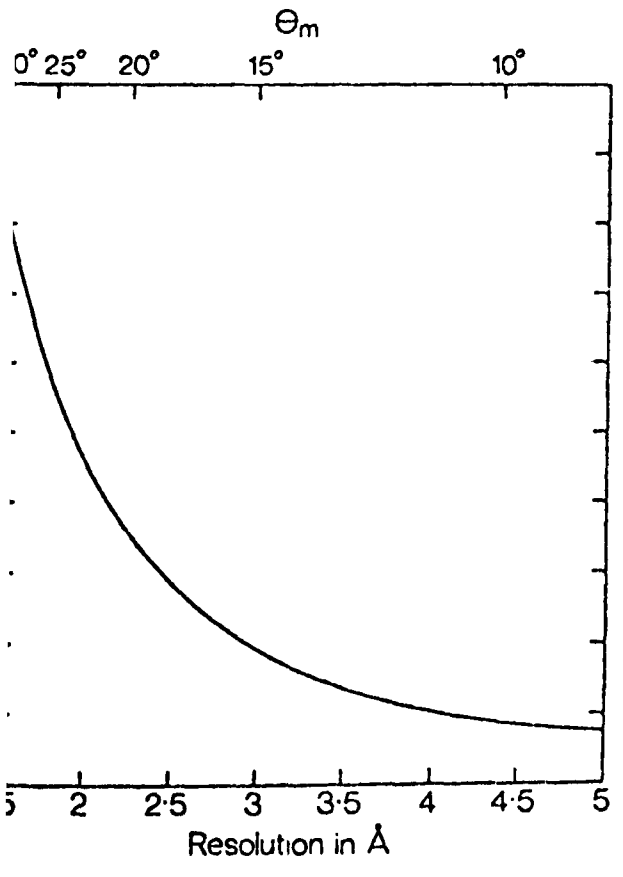


FIGURE 3.6

The percentage of reciprocal space which cannot be measured by rotating about a single axis, as a function of resolution. The fraction lost is:

$$-\tfrac{5}{4}\cos\theta_m + \tfrac{1}{2}\cos^3\theta_m + \tfrac{1}{2}\cos\theta_m \sin^2\theta_m + \tfrac{3}{8}\frac{\cos\theta_m}{\sin^2\theta_m} - \tfrac{3}{8}\frac{\theta_m}{\sin^3\theta_m}$$

where θ_m is the maximum Bragg angle and CuKα radiation is assumed (from Arndt and Wonacott, 1977.).

the blind region are collected by rotating the crystal through θ_{max} about another axis.

Recorded reflections fall into one of three categories: (i) fully recorded, (ii) partially recorded at the start or end of rotation, (iii) overlapping terms. An optimum rotation range for each photograph ($\Delta\phi$) is chosen, which is a compromise between maximising the number of fully recorded reflections per film and minimising the number of overlapping reflections. The maximum rotation range which avoids overlapping reflections is

$$\Delta\phi_{max} = \frac{p^*}{d^*} - \gamma \qquad \text{(Arndt } et\ al.\text{, 1973)}$$

where p* is the reciprocal lattice spacing, d* is the maximum resolution in reciprocal space and γ is the reflecting range of the crystal.

The reflecting range (γ) is dependent on the divergence of the X-ray beam and the crystal mosaic spread. It is estimated empirically by measuring the width (x) of a reflection of average intensity.

Then $\gamma = \frac{(x-a)}{F}$ where a is the crystal size

F is the crystal-film distance.

The limits on $\Delta\phi$ are then

$$\gamma < \Delta\phi \leq \frac{p^*}{d^*} - \gamma \qquad 3.1$$

The optimum rotation range within these limits is determined by calculating the proportions of fully recorded, partially recorded and overlapping reflections as a function of rotation range for a given crystal to film distance, reciprocal cell dimensions, mosaic

spread and ideal crystal orientation. Figure 3.7 shows results obtained for phosphorylase *b* for data to 3 Å and 2 Å resolution with a reflecting range of 0.35° and 0.25° respectively. The optimum rotation range for 2 Å data collection was chosen to be 0.5°. Data were recorded with sequential, non-overlapping oscillation ranges which allowed the summation of partially recorded terms at the end of one range to those at the beginning of the next range.

The Processing of Oscillation Photographs

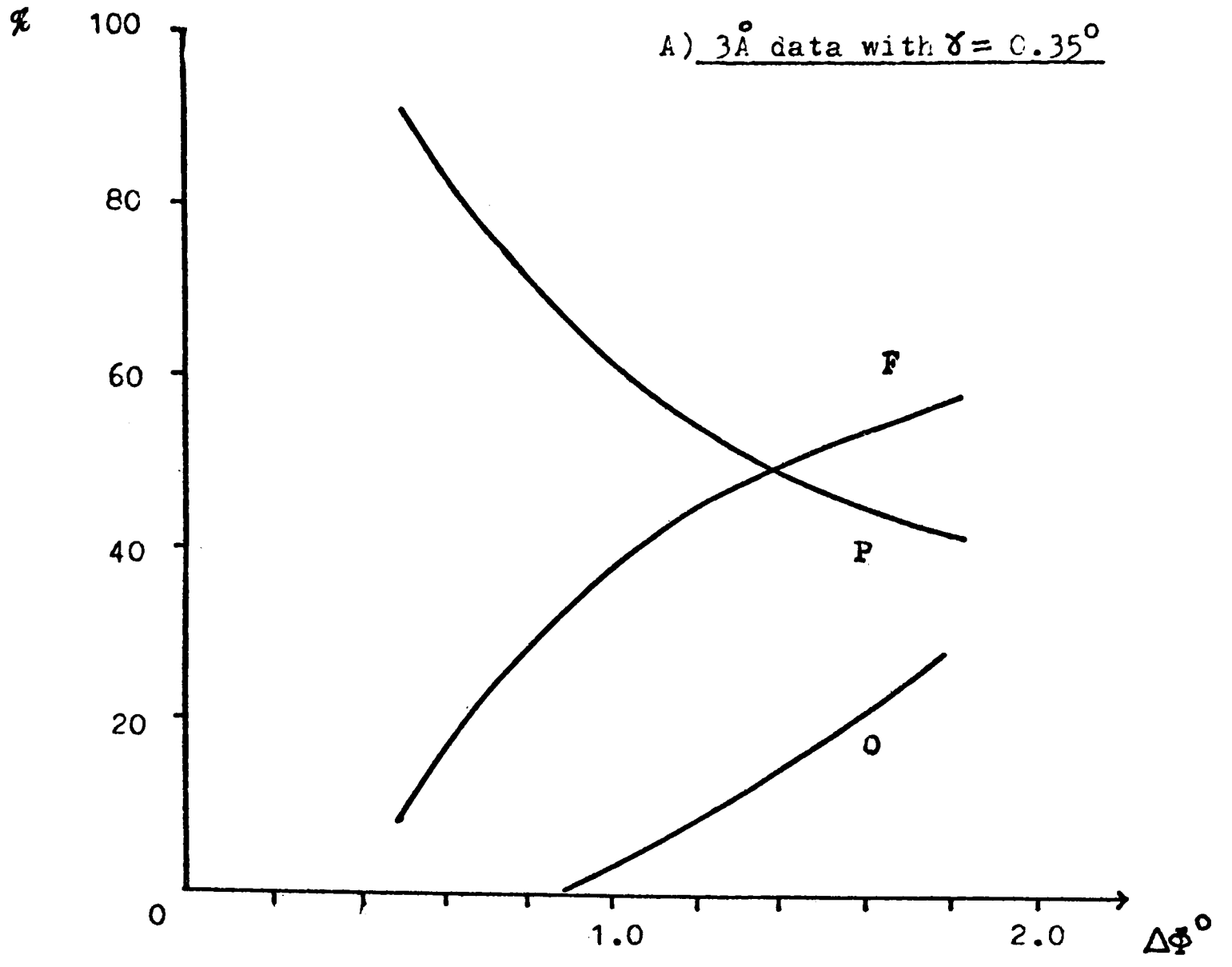
The processing system used to produce a set of scaled, merged intensities for oscillation photographs is described. The method followed is essentially that reported by Wilson and Yeates (1979). The computer programmes denoted by capital letters are listed in Appendix 1.

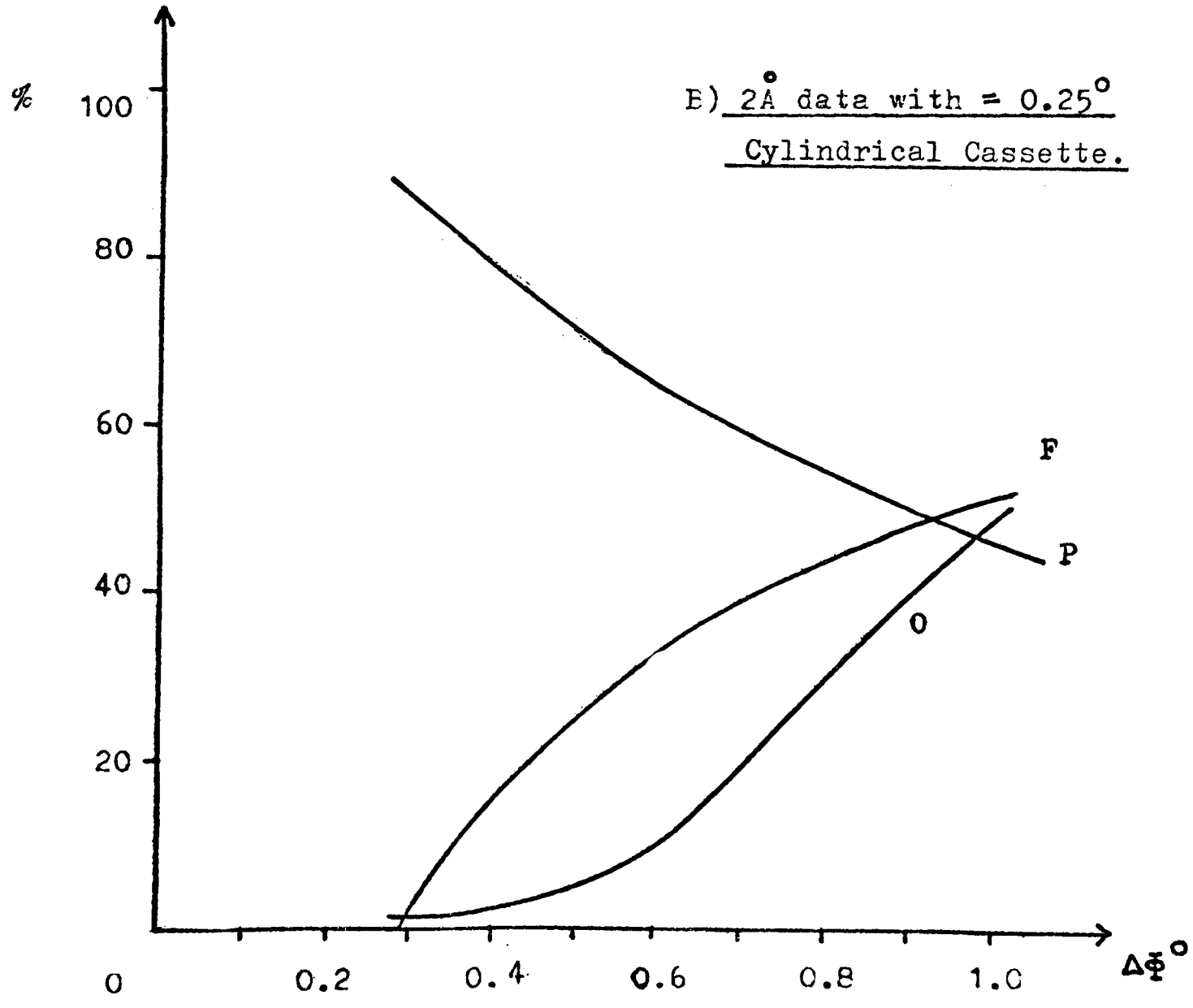
1) FILM DENSITOMETRY

Oscillation photographs are digitized using an Optronics Photoscan P-1000 densitometer, operating in an off-line mode and writing directly to magnetic tape. Subsequent steps in the integration process are carried out on an ICL 1906A or 2980 computer. The reflections are roughly 1 mm in diameter and are scanned on a 100 μm grid with optical density (O.D.), units of 0.0-2.0 output as integers in the range 0-255. The 100 μm grid used is large enough to avoid fluctuations due to grain size but small enough to minimize the Wooster effect (Matthews *et al.*, 1972). Wooster (1964) describes the errors introduced into logarithmic O.D. measurements which arise when large density changes occur within the area illuminated by the light spot. Matthews *et al.* (1972) calculated that, for a reflection of size 0.8 x 0.6 mm^2 and peak O.D. 2.0 measured with 100 μm raster and spot size, a maximum error of 3% is expected. A similar

FIGURE 3.7

The Reflections Recorded at Different Rotation Ranges





The number of reflections classified as fully recorded (F), partially recorded (P) or overlapping (O) is plotted as a percentage of the total number of reflections on the photograph for various rotation ranges ($\Delta\Phi$) for 3Å and 2Å data with different crystal reflecting range (γ).

calculation for phosphorylase *b* with peak size of 1.1 mm x 1.1 mm gives an error of less than 1% (Weber, 1977). The combined non-linear response of the densitometer and X-ray film is calibrated periodically. An empirical scale (Figure 3.8) to correct each O.D. point for non-linearity has been produced by recording a series of exposures of the direct beam using a tube collimator.

2) CRYSTAL ORIENTATION

In order to be able to index the films, five different coordinate systems are used:-

i) the unit cell axes (a^*, b^*, c^*)

ii) a right handed orthogonal crystal coordinate system fixed relative to the crystal ($x_o y_o z_o$)

iii) a right handed orthogonal coordinate system fixed relative to the camera (xyz)

iv) the film coordinate system (X_F, Y_F)

v) the Optronics coordinate system (X_O, Y_O).

Figure 3.9 shows the relationship between the unit cell axes and the crystal coordinate system. The coordinates of a reciprocal lattice point p($\underline{H}$) in the orthogonal camera system are given by the matrix equation:

$$\underline{X} = [\phi]\,[S]\,[W]\,\underline{H}^T$$

where [W] transforms from the crystallographic to the orthogonal crystal coordinate system

[S] describes the missetting of the crystal by three orthogonal angles about the axes (x_o, y_o, z_o)

[ϕ] describes the rotation relative to some datum. The origin $\phi = 0^\circ$ is defined when one of the reciprocal lattice axes is parallel to the X-ray beam.

The matrix W has the form:-

FIGURE 3.8 Empirical Correction for Non-Linearity of Film Response

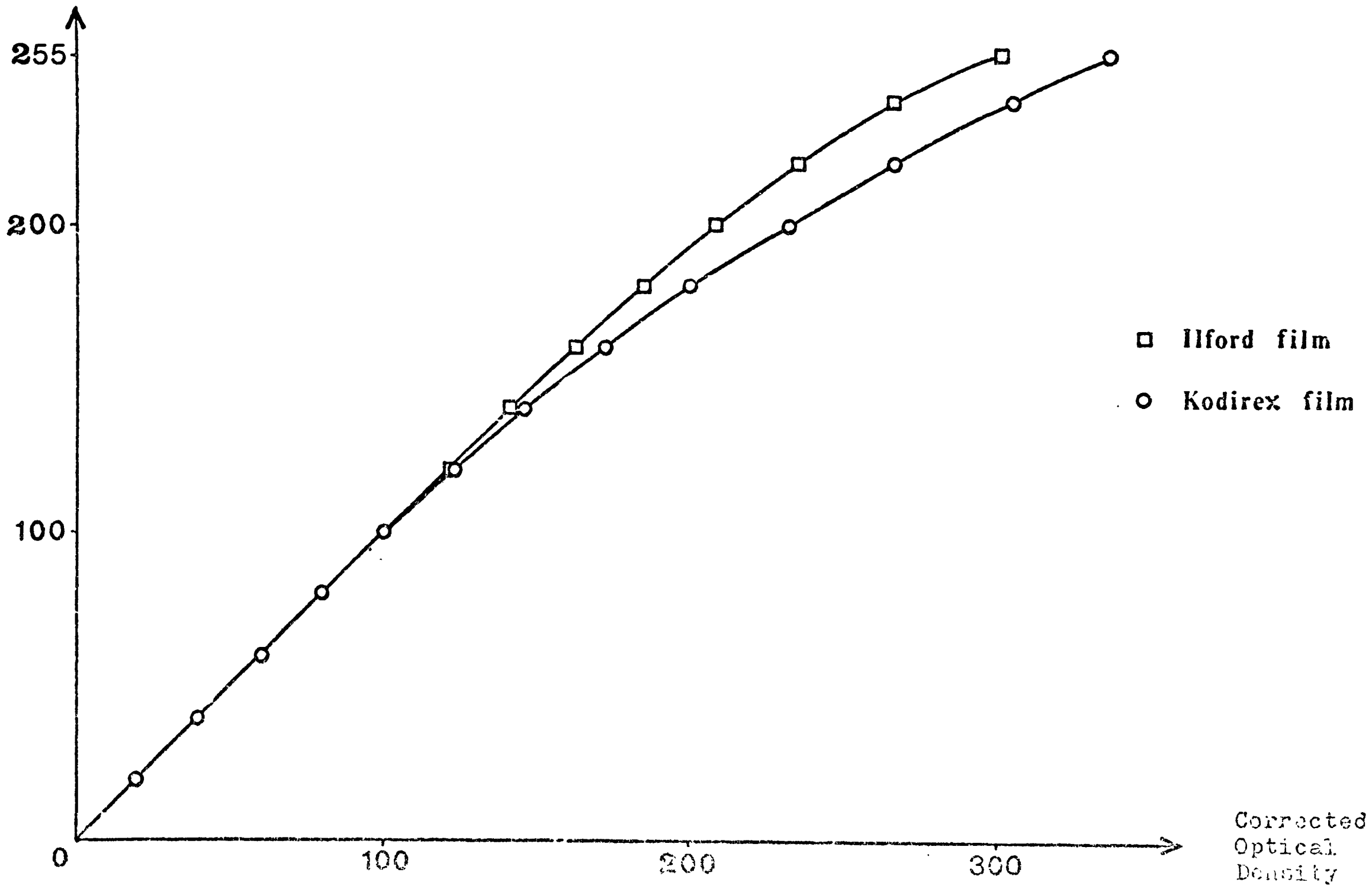
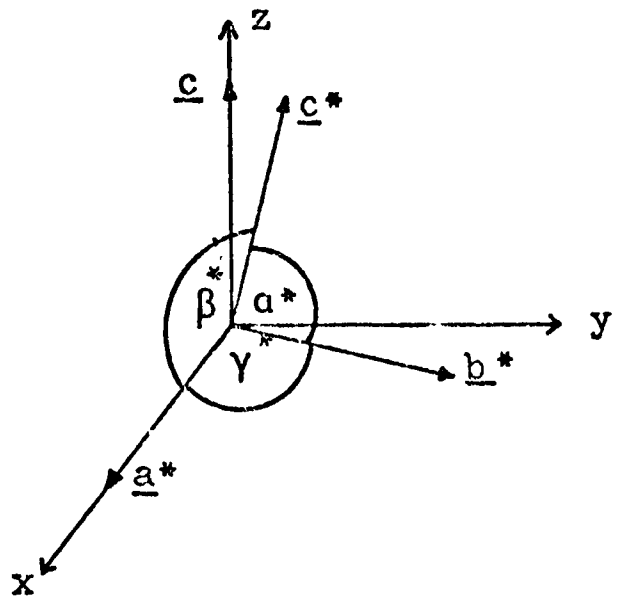


FIGURE 3.9

Relation Between the Unit Cell Axes and the Crystal Coordinate

:em



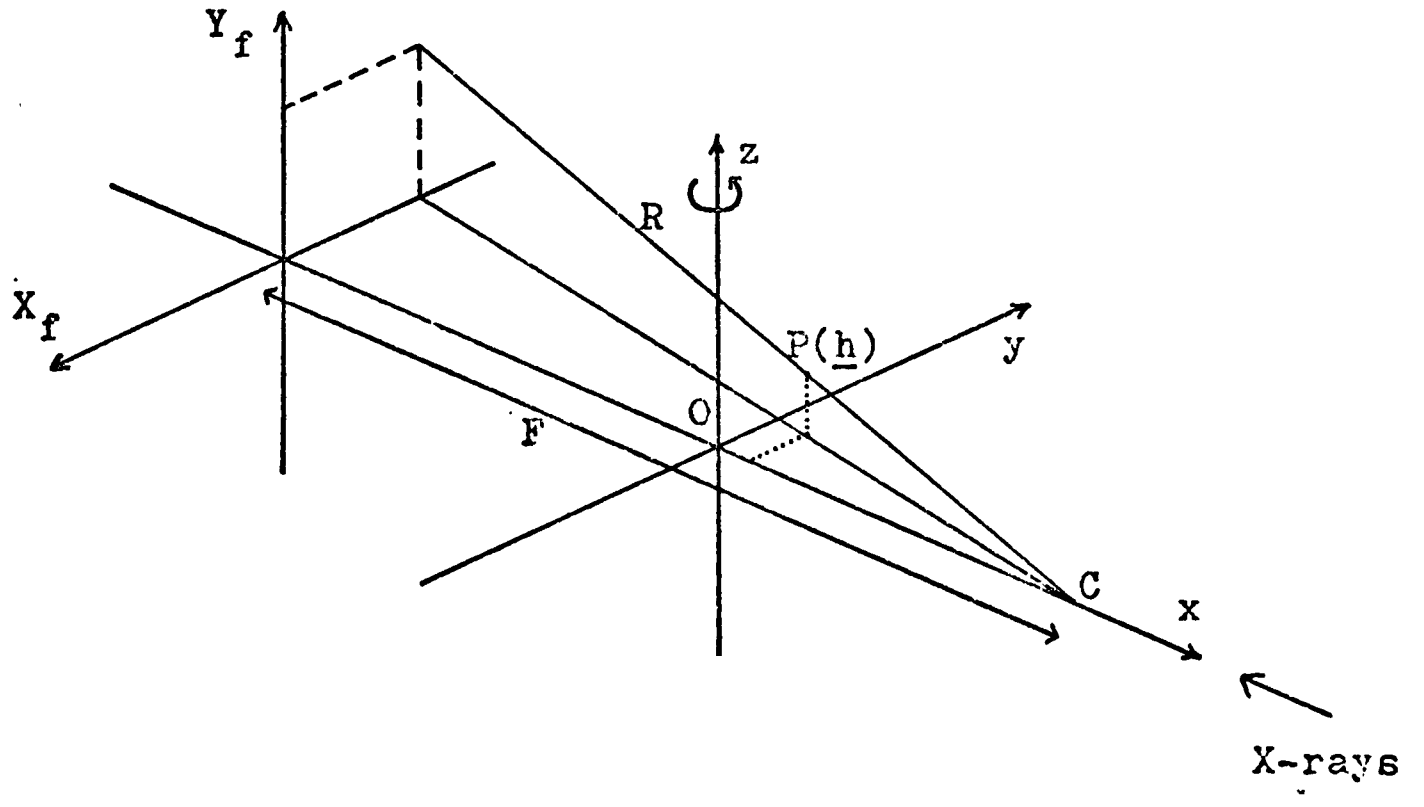
crystal is set to rotate about c.

FIGURE 3.10

The Relation Between Crystal and Film Coordinate Systems For Flat Plate Cassette.

Crystal coordinates (x, y, z)

Film coordinates (X_f, Y_f)



The crystal axes are centred at O, and C is the centre of the Ewald sphere. The crystal rotates about the z axis.

$$[W] = \begin{pmatrix} a^*\cos\beta^* & b^*\cos\alpha^* & c^* \\ a^*\sin\beta^* & \frac{b^*(\cos\gamma^*-\cos\alpha^*\cos\beta^*)}{\sin\beta^*} & 0 \\ 0 & b^*\cos\psi & 0 \end{pmatrix} \qquad 3.3$$

where $\cos\psi = \frac{\lambda}{|\underset{\sim}{b}|.|\underset{\sim}{b}^*|}$ *i.e*: ψ is the angle between $\underset{\sim}{b}$ and $\underset{\sim}{b}^*$.

The rows or columns of [W] vary according to which axis is the rotation axis. The [W] shown corresponds to a rotation about $\underline{b}$.

The matrix [S] has the form:-

$$[S] = \begin{pmatrix} \cos\Delta Y\cos\Delta Z & -\cos\Delta X\sin\Delta Z-\sin\Delta X\sin\Delta Y\cos\Delta Z & \sin\Delta X\sin\Delta Z-\cos\Delta X\sin\Delta Y\cos\Delta Z \\ \cos\Delta Y\sin\Delta Z & \cos\Delta X\cos\Delta Z-\sin\Delta X\sin\Delta Y\sin\Delta Z & -\cos\Delta X\sin\Delta Y\sin\Delta Z-\sin\Delta X\cos\Delta Z \\ \sin\Delta Y & \sin\Delta X\cos\Delta Y & \cos\Delta X\cos\Delta Y \end{pmatrix} \qquad 3.4$$

The determination of reliable integrated intensities relies upon precise values for the unit-cell parameters and crystal orientation. The position of a reflection and its degree of partial recording (*i.e*: whether the reciprocal lattice "point" has completely or partially passed through the Ewald sphere) must be determined. The distribution of partially recorded reflections, and their degree of recording, is particularly sensitive to the orientation of the crystal and is used to determine the crystal missetting angles and the unit cell parameters. A set of partially recorded reflections are obtained from 'setting' photographs taken at $\phi = 0^\circ$ and $\phi = 90^\circ$. These may be still photographs or small angle oscillation photographs. For small angle oscillation photographs the degree of partial recording must be carefully estimated, and experience has shown that it is better to use 'still' photographs and assume that all reflections are 50% partially recorded. This

method has the advantage of not requiring a subjective estimation of the degree of partial recording of a reflection, which was often found to be in error. Initial estimates of the missetting angles are made from the 'setting' photographs and the coordinates of 40-50 partially recorded reflections are measured by hand or determined from a digitized still photograph using the programme CORDOBA. Precise values of the crystal orientation and unit cell parameters are calculated by Segment 1 of the programme OSCAR by an iterative least squares minimization process of:

$$(D_{obs}-D_{calc})^2 \qquad 3.5$$

where D_{obs} and D_{calc} are the observed and calculated distances of a partially recorded reciprocal lattice point (P(H)) from the centre of the Ewald sphere. The calculated distance

$$D_{calc} = (1-2x+x^2+y^2+z^2)^{\frac{1}{2}}$$

where the coordinates (x,y,z) are determined from the observed coordinates of the reflection on the film (X_F,Y_F). The reciprocal lattice is considered as a lattice of small spheres with radii

$$e(\underline{H}) = \frac{\gamma d^*(\underline{H})}{2} \qquad 3.6$$

where γ is the reflecting range of the crystal

$d^*(\underline{H})$ is the distance from the origin of reciprocal space. The observed distance from the centre of the Ewald sphere is estimated as

$$D_{obs} = 1+ge(\underline{H}) \qquad 3.7$$

where g is a factor depending on the observed degree of recording.

3) PREDICTION OF REFLECTIONS IN AN OSCILLATION RANGE

The refined values of the crystal orientation and unit cell parameters are then used to generate a predicted set of indices and film coordinates of all reflections appearing on the film for a given rotation, using Segment 2 of OSCAR. A search for reflections is made, through a volume of reciprocal space which is defined by the Ewald sphere at the start (ϕ_i) and end (ϕ_f) of the rotation range, and limited by the sphere of resolution. The indices are determined from equation 3.2 and the film coordinates are generated as

$$(X_F, Y_F) = \mathcal{F}(T, \chi, S)_{x,y} \qquad 3.8$$

where S is the crystal to film distance and $\mathcal{F}$ depends on the geometry of the film detector. Figure 3.10 shows the relationship between the film coordinates and the orthogonal camera coordinate system (x,y,z) for the case of a flat plate film cassette. From similar Δ's it can be shown that:

$$X_F = -Sy/1-x \qquad Y_F = z(S^2+X_F^2)^{\frac{1}{2}}/(1-z^2)^{\frac{1}{2}} \qquad 3.9$$

Figure 3.11a shows the geometry of the cylindrical cassette and the definition of the film coordinate system (X_F, Y_F) and Figure 3.11b the definition of the angles T, χ in the crystal coordinate system.

From Figure 3.11b

$$\tan\chi = z/(1-z^2)^{\frac{1}{2}} \quad \text{and} \quad \sin\chi = z \quad \text{and} \quad \tan T = y/1-x$$

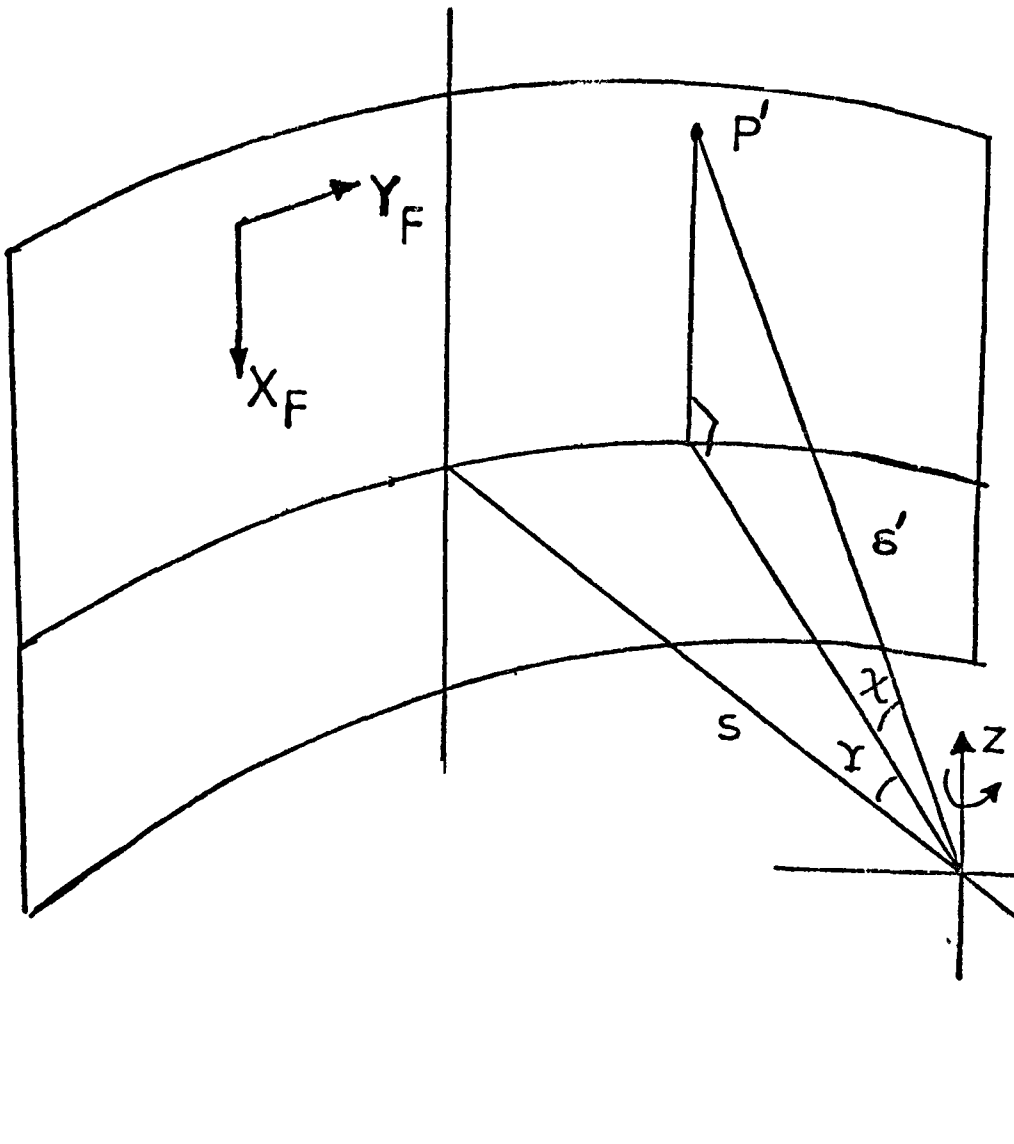
The film coordinates are given by:

FIGURE 3.11

a) <u>The Relationship Between Crystal and Film Coordinate Systems for Cylindrical Cassette.</u>

Crystal coordinates (x,y,z)

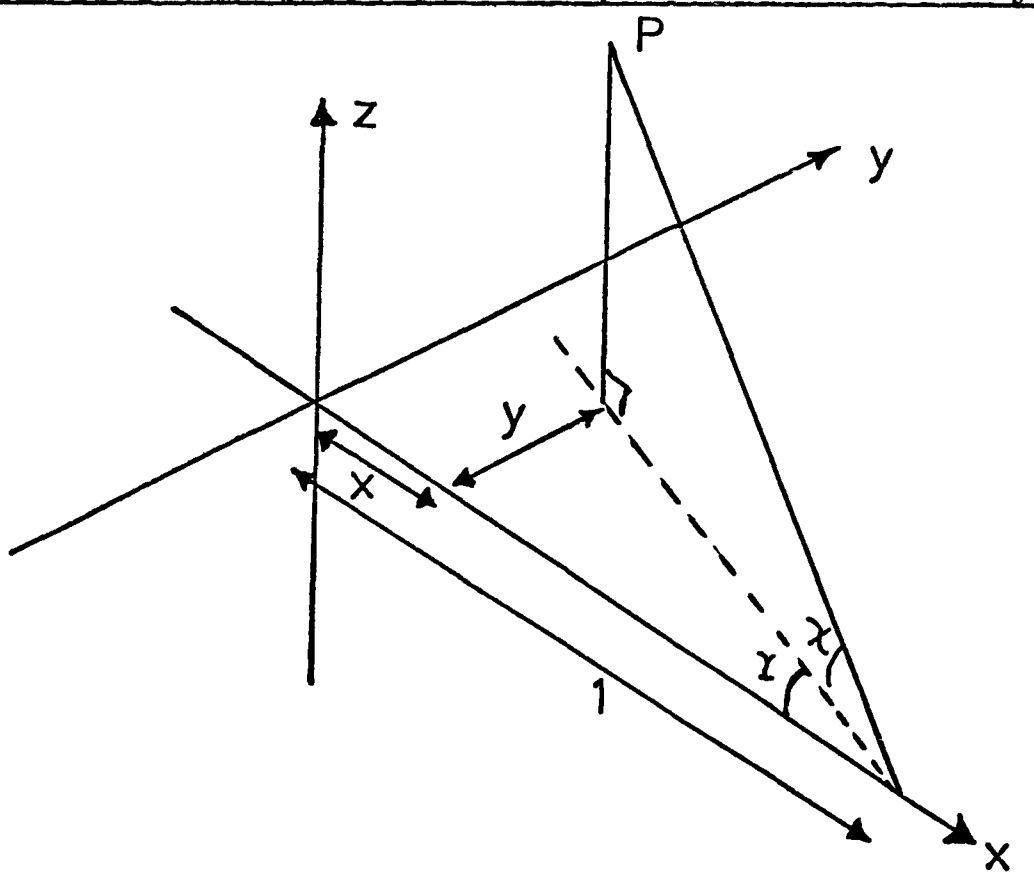
Film coordinates (X_F, Y_F)



$X_F = -s \tan\chi$

$Y_F = s\gamma$

b) <u>Definition of angles γ, χ in crystal coordinate system.</u>



$\sin\chi = z$

$\tan\gamma = \frac{y}{1 - x}$

$$X_F = -S\tan\chi = -Sz/(1-z^2)^{\frac{1}{2}} \qquad 3.10$$

$$Y_F = S.T = S.\tan^{-1}(y/1-x)$$

Predicted reflections are rejected if they are

i) outside the resolution sphere

ii) too close to the direct beam (within 5 mm), since they would be behind the backstop and high background region

iii) partially in contact with the Ewald sphere during the complete rotation

iv) partially within the blind region.

The remaining reflections are then flagged as

i) fully recorded, if the reflection completely cuts the Ewald sphere during the rotation

ii) partially recorded at ϕ_i or ϕ_f, if the distance of the reflection from the Ewald sphere is less than $e(\underline{H})$ (equation 3.6)

iii) overlapping, if a box, equal to the peak size and centred on the reflection, overlaps with the box centred on a neighbouring reflection. These reflections are rejected.

A plot of the positions of the predicted reflections is then made, using different coloured inks for fully and partially recorded reflections, to compare with the actual photograph. This is similar to the procedure described in Arndt and Wonacott (1977, p. 90). Individual reflections may be compared and this appears to be a sensitive method for detecting errors in the crystal orientation and unit cell parameters. If there is bad agreement between the predicted plots and the film, a new crystal orientation matrix may

be determined by rerunning Segment 1 of OSCAR or by manual adjustments, or the unit cell parameters may be varied slightly and a new plot produced. The programme MATIS may be used to calculate the orthogonalisation and rotation matrix for a tetragonal or cubic crystallographic system. The programme calculates [W] (equation 3.3) and [S] (equation 3.4) for given cell parameters and missetting angles. Any errors are best detected at this stage, since they may not otherwise become apparent until the agreement between equivalent reflections are analysed later. Other possible procedures for crystal orientation are discussed in Arndt and Wonacott (pp. 105-117 and 139-151).

4) DETERMINATION OF THE OPTRONICS MATRIX

The magnetic tape containing the digitized record of the film is scanned by Segment 3 of OSCAR to find the positions on the tape of optical density values corresonding to the three fiducial marks (spots or holes). These are at known positions (X_F, Y_F) in the film coordinate system and this allows an approximate transformation from the film coordinate system to the scanner reference frame to be determined. The relationship between the film coordinate system and the Optronics frame is shown in Figure 3.12 for the flat plate and cylindrical cassettes. The transformation is defined as

$$\underline{X}_0 = [R]\underline{X}_F + [T]$$

where [R] and [T] are rotational and translational matrices respectively. The observed positions of the fiducial marks may be checked using the plots described in Section 3, since they may have been displaced due to film slippage in the cassette. This film slippage may be minimised by fixing the back of the cassette with locking screws, or by tightly fixing the front and back together

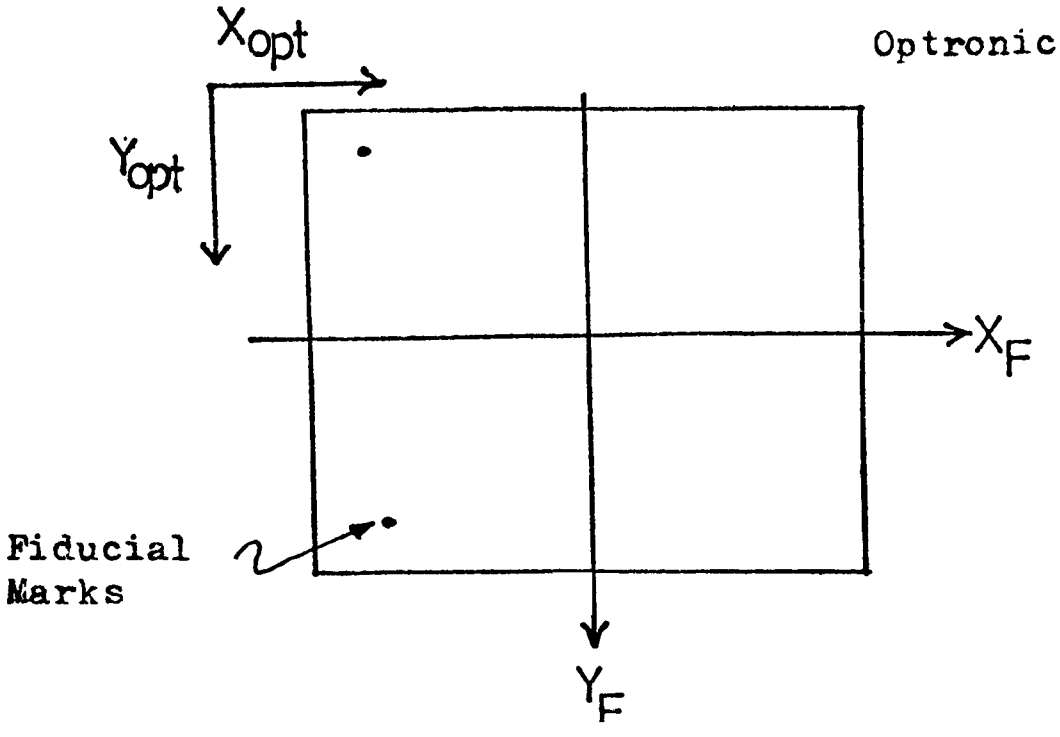
FIGURE 3.12

The Relationship Between the Film and Optronics Coordinate Systems.

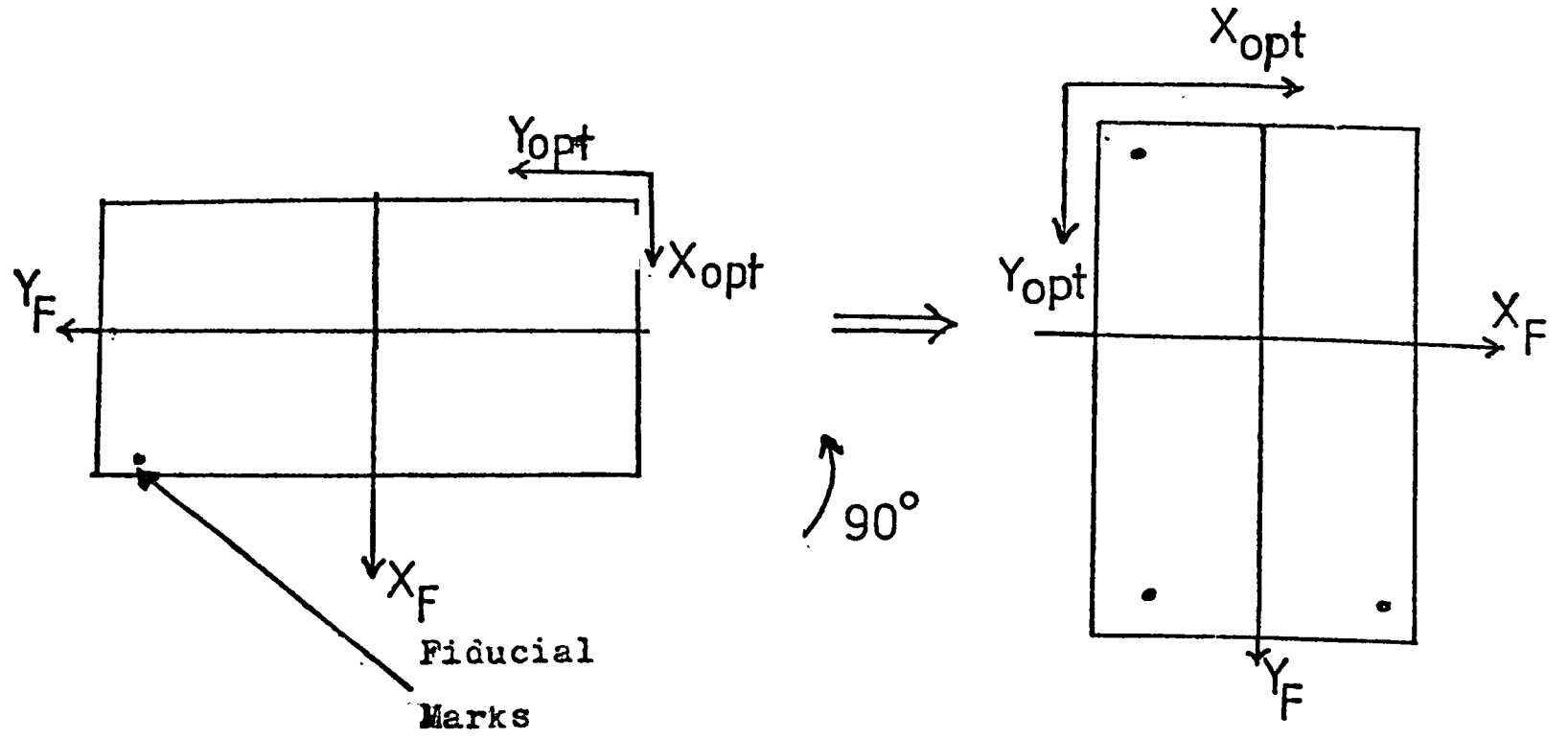
a) Flat Plate Cassette.

Film coordinates (X_F, Y_F)

Optronics coordinates (X_{opt}, Y_{opt})



b) Cylindrical Cassette.



with insulation tape at the corners. The latter method was used here. The approximate transformation is used to define the approximate positions of a set of fully recorded reflections. The programme automatically defines a mean background density level and searches for 5 or 6 reflections above this level in each octant of the film. Alternatively the mean background level may be set manually, or the coordinates of a set of fully recorded reflections may be used. The precise positions of these peaks are obtained by a centre of gravity search over the density points above a bias level. The centroid of the peak is calculated as

$$X = \frac{\sum X_i \bar{O}_i}{\sum \bar{O}_i} \qquad Y = \frac{\sum Y_i \bar{O}_i}{\sum \bar{O}_i}$$

where $\bar{O}_i$ is the optical density of point (X_i, Y_i) above the bias level. The coordinates of the centroids are then used to define a precise transformation:-

$$(X_O, Y_O) = \underset{\sim}{O} (X_F, Y_F)$$

where the Optronics matrix $\underset{\sim}{O}$ is a correction of the second degree in X_F to account for imprecise crystal to film distance and misseting of the film cassette.

5) INTEGRATION

Each reflection is then intrgrated using Segment 4 of OSCAR. The integrated intensity is determined by summing the optical density values in a box centred on its predicted position and subtracting a background correction:

$$I(\underline{H}) = \sum_{i=1}^{N_p} O_i' - k \sum_{j=1}^{4} \sum_{k=1}^{N_b} O_{jk}'$$

where O_i' and O_{jk}' are the contributions from the peak and background sampling areas N_p and N_b respectively, and k is the background to

peak scaling factor $N_p/4N_b$. The integration box used is shown in Figure 3.13 where the peak size (n_x,n_y), the background size (b_x,b_y) and the displacement of the backgrounds relative to the peak $(\Delta x,\Delta y)$ are defined. O' represents an optical density value corrected for non-linearity in the response of the film scanner system as described in section 1.

The size of the diffraction maxima on the film will vary for different values of (T,χ) and the size of the integration box is varied to accommodate this effect. This oblique incidence expansion is allowed for by using a variable box size:

$$n = n_o f(T,\chi)$$

The background sampling areas are kept constant and the background to peak scale factor k is varied as

$$k(T,\chi) = N_p(T,\chi)/4N_b$$

The oblique incidence expansion factor $f(T,\chi)$ depends on the geometry of the film. Figure 3.14 shows the oblique incidence correction for flat plate and cylindrical geometry.

For the flat plate

$$f(T,\chi) = \mathrm{Sec}T\mathrm{Sec}\chi = \mathrm{Sec}2\theta \qquad 13.11a$$

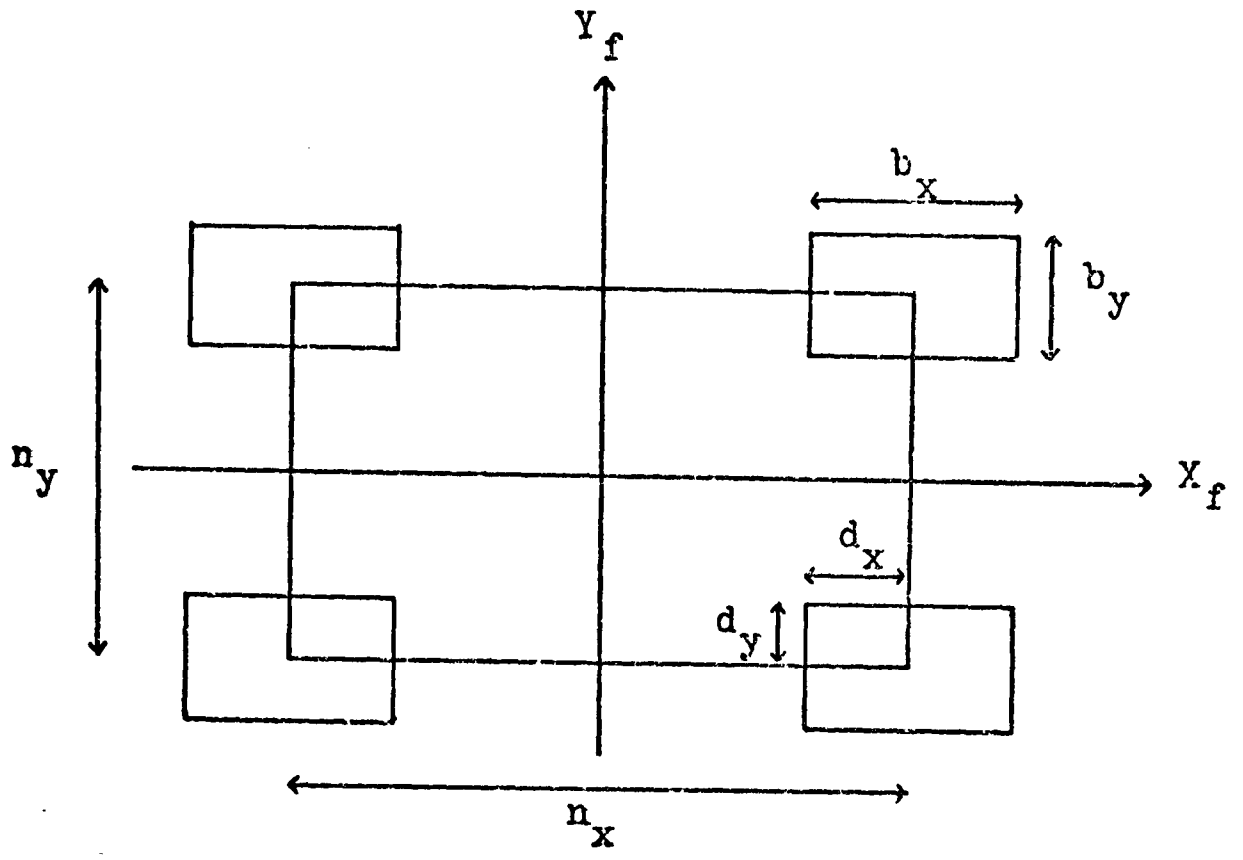
For the cylindrical cassette

$$f(T,\chi) = \mathrm{Sec}\chi \qquad 13.11b$$

where χ is the angle between the generator for the nth layer line and the normal to the rotation axis. With cylindrical geometry, for

FIGURE 3.13

The Box for Integration



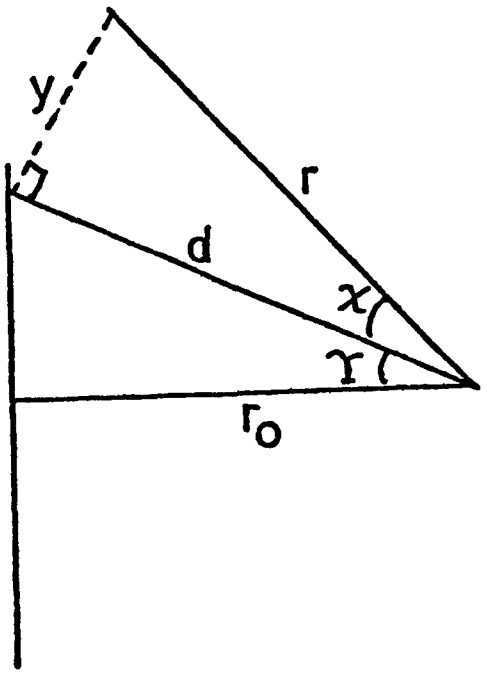
The peak size is $n_x \times n_y$

The total background size is $4(b_x \times b_y)$

FIGURE 3.14

Oblique Incidence Factor.

a) For Flat Plate Cassette.



$$x = r_o \tan\chi$$

$$y = d\tan\chi$$

where $d = r_o \sec\gamma$

$$\therefore y = r_o \sec\gamma \tan\chi$$

Oblique incidence factor is given by:-

$$\frac{r}{r_o} = f$$

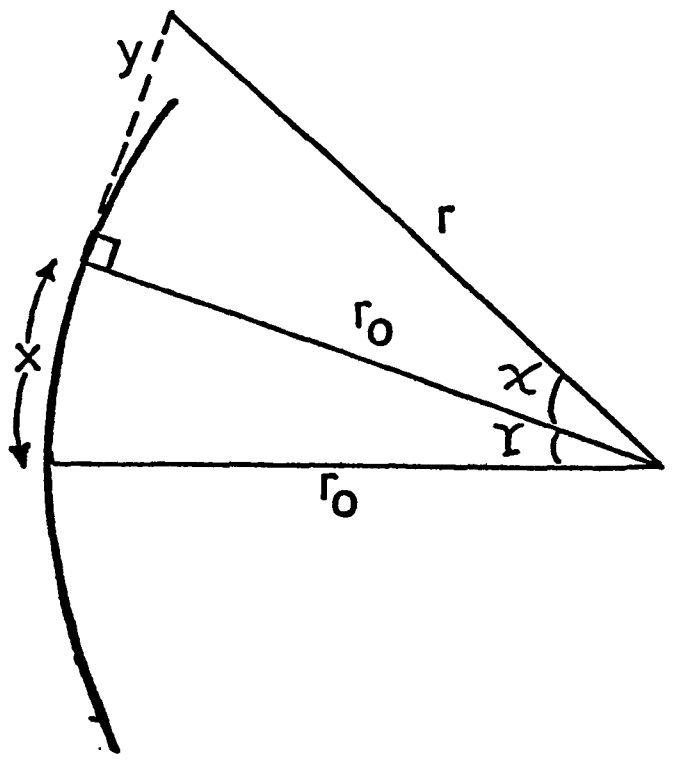
where $r = d\sec\chi$

$$= r_o \sec\gamma \sec\chi$$

$$f = \sec\chi \sec\gamma$$

$$f = \sec 2\theta$$

b) For Cylindrical Cassette.



$$x = r_o\gamma$$

$$y = r_o \tan\chi$$

Oblique incidence factor is given by :-

$$f = \frac{r}{r_o}$$

$$f = \sec\chi$$

the Zeroth level reciprocal lattice orthogonal to the rotation axis, the diffracted beams are normal to the film, and an expansion for oblique incidence is only necessary in the direction of X_F (as defined in Figure 3.11a).

Reflections are rejected at this stage if they exceed the saturation level of the film, or if any of the 4 backgrounds differs significantly (say by 10%) from their mean value. This situation may arise from radiation streaks, dust, damage, excess background scattering, reflections too close together or inaccurately predicted. Reflections having negative intensities and not rejected as background failures are retained and their treatment is described in Section 9. The overall scheme for integration is summarised in Figure 3.15.

6) INTER-FILM SCALE DETERMINATION (FILMPACK)

The programme FILMPACK estimates the standard deviations of reflections within each photograph, calculates the film-to-film scale factor, applies the scale factor, Lorentz-polarization and oblique incidence corrections, and merges equivalent terms from the weak and strong films. All equivalent terms are sorted together and the standard deviations of the intensities are estimated independently for the weak and strong film from the expression:

$$\sigma_i = \left[\sum_{i=1}^{N} \frac{(I_i - \bar{I})^2}{N-1} \right]^{\frac{1}{2}}$$

where I_i is the observed intensity for the ith member of a set of N fully recorded equivalent terms

$\bar{I}$ is the unweighted mean intensity of the set

σ_i is the estimated standard deviation of the ith term.

FIGURE 3.15

A Scheme for the Integration of Reflections Using OSCAR

1. Refine the crystal orientation, the reciprocal cell parameters and the crystal to film distance using a set of 50 to 80 partially recorded reflections from the orientation photographs.

CRYSTAL ORIENTATION MATRIX

2. Generate film coordinates for the predicted reflections in a given Φ range. Each reflection is flagged as fully recorded, partially recorded or overlapping. Reflections are ignored if they overlap with others or are partially in the blind region. A plot of the predicted reflections is compared with the original data film to test the matrix.

Films are scanned on an Optronics densitometer to produce a set of density values.

3. The fiducial marks generate a transformation from the theoretical film coordinates to coordinates in the scanner frame of reference. This is refined by centring on 40 to 50 fully-recorded reflections.

4. Individual optical density points are corrected for the non-linear response of film and scanner. An integrated intensity is calculated for each reflection. Reflections are rejected if they are too strong (with 4 density points above 250), or have unequal background densities.

A SET OF INTEGRATED INTENSITIES

Sections 2, 3, and 4 are repeated for each data film of the crystal.

Both truly equivalent and Friedel-related terms are used, and the estimates for each set are summed in 10 ranges of mean intensity. The estimated standard deviation for each reflection is calculated from the following empirical expression:

$$\sigma_{est} = A+BI \qquad 3.12$$

The constants A and B are determined by a least squares minimization of the sum of the squares of the differences between the left and right hand sides of the equation for the ten ranges of I (mean intensity of the set). This empirical expression apparently gives reasonable agreement with the observed errors (Wilson and Yeates, 1979). When there are insufficient numbers of equivalent terms default values for the estimated standard deviations are assumed from experience with previous photographs. The estimated standard deviations are used to provide weights for use in the calculation of the film-to-film scale factor. A factor $f(T,\chi)$ defined as in equations 3.11 (Figure 3.14) allows for oblique incidence absorption by the first film (Cox and Shaw, 1930; Whittaker, 1953) and so the intensities on the strong and weak films are related by:

$$I_s = I_w \exp(Atf(T,\chi)) \qquad 3.13$$

The film constant At is then calculated from

$$At = \sum_{i=1}^{N} \left\{ \frac{1}{\sigma_r^2} \left[\log_e \left(\frac{\langle I_i \rangle}{\langle I_{2i} \rangle} \right) \right] / f(T,\chi) \right\} / \sum \frac{1}{\sigma_r^2} \qquad 3.14$$

Here $\langle I_i \rangle$ and $\langle I_{2i} \rangle$ are the mean intensities for a set of equivalent terms recorded on the first and second films respectively, σ_r the estimated standard deviation of $\log_e(\langle I_i \rangle / \langle I_{2i} \rangle)$, and N the total

number of sets of equivalents (since for the cylindrical cassette secχ is different for each of the members of a set of equivalents, each reflection appears individually in the sum). Terms are included in this summation if the mean intensity is greater than 6 standard deviations for both the first and second films. The intensities for the weak film are then corrected for oblique incidence absorption using equation 3.13. Fully recorded equivalents are summed to give mean weighted intensities and observations more than 4 standard deviations from this mean are rejected. The number of rejected terms is monitored and provides a useful check for poorly determined crystal orientation matrices. A mean weighted intensity for terms with identical observed indices on both films is calculated. Equivalent terms within each film are not averaged. Finally the Lorentz-polarization and the oblique incidence correction are applied to the data. A second oblique incidence correction (Cox and Shaw, 1930; Whittaker, 1953) is required since the more obliquely scattered beams pass through a thicker layer of emulsion and a larger number of grains are exposed. This correction is of the type:

$$I_{corr} = \frac{I(2\theta)(1-e^{-A})(1+e^{-(A+B)})}{(1-e^{-A\sec 2\theta})(1+e^{-(A+B)\sec 2\theta})} \qquad 3.15$$

for the flat plate cassette (and for the cylindrical cassette a similar expression with secχ replacing sec2θ). $I(2\theta)$ is the observed intensity at 2θ and e^{-A}, e^{-B} are absorption factors for a normal beam in a single emulsion and in the film base respectively.

The Lorentz factor for the oscillation camera is:

$$L^{-1} = (\sin^2 2\theta - \zeta^2)^{\frac{1}{2}} \qquad 3.16$$

and the polarization correction for Ni filtered CuKα radiation is

$$P^{-1} = \frac{2}{1+\cos^2 2\theta} \qquad 3.17$$

where ζ is the cylindrical polar coordinate of the reflection.

Azaroff (1955) has derived a polarization correction for crystal monochromated radiation by considering the intensity of an X-ray beam to be the sum of two independent components in two orthogonal directions of polarization. For the usual arrangement with a monochromator crystal in the vertical plane, and normal-beam geometry, the expression reduces to:-

$$P = \frac{\zeta^2(1-|\cos 2\theta_1|)+\cos^2 2\theta_2+|\cos 2\theta_1|}{1+|\cos 2\theta_1|}$$

where θ_1 is the Bragg angle for the monochromator and θ_2 that for the specimen (Arndt and Wonacott, 1977). For CuKα radiation the expression equals:-

$$P = 0.528\cos^2 2\theta + 0.0560\zeta^2 + 0.471$$

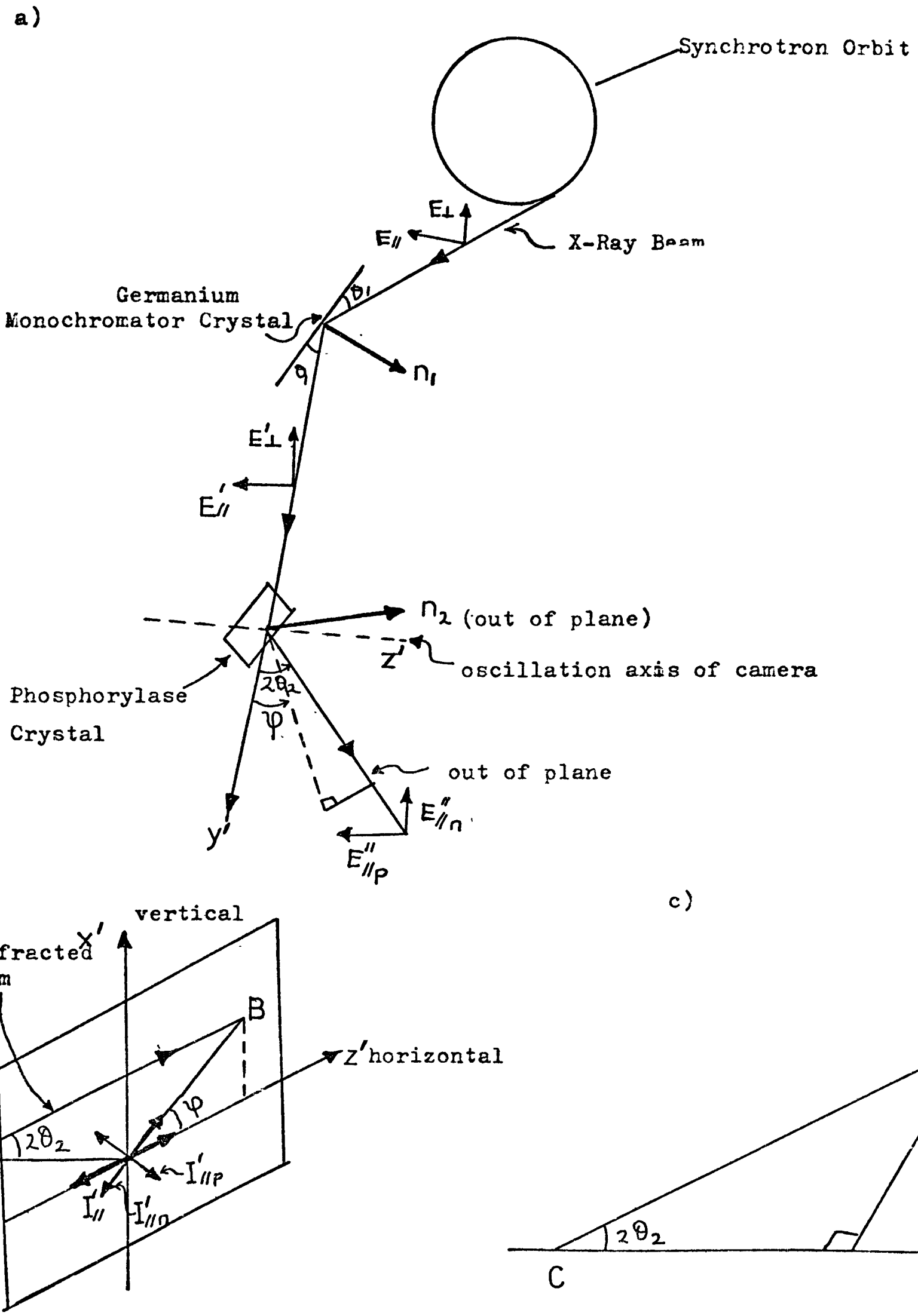
and is thus not very different from that for an unpolarized incident beam (equation 3.17).

Using a similar approach to that of Azaroff, a polarization correction may be derived for synchrotron radiation after reflection from a crystal monochromator (see also Chapter 5).

Assume that the normal to the reflecting planes of the monochromator crystal lie in the plane of orbit (y-z plane), so that the beam reflected by the monochromator crystal also lies in the yz plane along y' (see Figure 3.16a). The intensity may be resolved into two components ($I_\perp$ and $I_\parallel$). The synchrotron radiation is highly polarized in the plane or orbit (see Chapter 5) and so, in general, $I_\parallel >> I_\perp$.

FIGURE 3.16

Synchrotron Radiation Polarization Factor.



a) Considering the ∥ component

Polarization by the monochromator gives

$$I'_{\parallel} = I_{\parallel}\cos^2 2\theta_1$$ where θ_1 is the Bragg angle for the monochromator

This component is then reflected by the crystal (Figure 3.16b) and may be resolved further into two components $I'_{\parallel n}$ and $I'_{\parallel p}$.

along normal $\quad I'_{\parallel n} = I'_{\parallel}\cos^2\varphi$

along plane $\quad I'_{\parallel p} = I'_{\parallel}\sin^2\varphi$

where φ is the angle the projection of the diffracted beam makes with the horizontal z'.

Now consider the diffracted beam.$I'_{\parallel n}$ will be attenuated by $\cos^2 2\theta_2$ to give $I''_{\parallel n}$ (Figure 3.16c).

$$I''_{\parallel n} = I'_{\parallel}\cos^2\varphi\cos^2 2\theta_2$$

However, component $I'_{\parallel p}$ is already in the plane perpendicular to the diffracted beam and will not be attenuated.

$$I''_{\parallel p} = I'\sin^2\varphi$$

b) Considering the ⊥'r component

The perpendicular component is not affected by the first reflection at the monochromator.

$$I'_{\perp} = I$$

Resolving $I'_{\perp}$ along two perpendicular directions at the second reflection:-

$$\text{along normal} \qquad I'_{\perp n} = I_{\perp}\sin^2\varphi$$

$$\text{along plane} \qquad I'_{\perp p} = I_{\perp}\cos^2\varphi$$

Again the normal component is attenuated by $\cos^2 2\theta_2$ and the plane component is not:-

$$I''_{\perp n} = I_{\perp}\sin^2\varphi\cos^2 2\theta_2$$

$$I''_{\perp p} = I_{\perp}\cos^2\varphi$$

The final intensity is the sum of these components:-

$$I'' = I''_{\parallel p} + I''_{\parallel n} + I''_{\perp n} + I''_{\perp p}$$

i.e:

$$I'' = I_{\parallel}(\cos^2 2\theta_1\cos^2\varphi\cos^2 2\theta_2 + \cos^2 2\theta_1\sin^2\varphi) + I_{\perp}(\sin^2\varphi\cos^2 2\theta_2 + \cos^2\varphi)$$

$$I'' = I_{\parallel}(\cos^2 2\theta_1(1-\cos^2\varphi\sin^2 2\theta_2) + I_{\perp}(1-\sin^2\varphi\sin^2 2\theta_2) \qquad 3.18a$$

The actual polarization correction used is an expression derived empirically by Fourme and Kahn (1981) and is given by:-

$$P = \tfrac{1}{2}(1+\cos^2 2\theta_2 - 0.89\sin^2 2\theta_2\cos 2\varphi)$$

$$= \frac{1+\cos^2 2\theta_2}{2}\left(1 - \frac{0.89\sin^2 2\theta_2\cos 2\varphi}{1+\cos^2 2\theta_2}\right)$$

$$P = P_0(1-\tau'\alpha) \qquad 3.18b$$

where $P_0 = \dfrac{1+\cos^2 2\theta_2}{2}$ the normal polarization correction

$$\alpha = \frac{\sin^2 2\theta_2\cos 2\varphi}{1+\cos^2 2\theta_2}$$

$\tau' = 0.89$. This term is calculated from the source characteristics and assuming that the monochromator crystal is perfect.

The polarization correction, P^{-1} is shown as a function of ψ for different values of θ for the two expressions in Figure 3.17. They agree to within 3% for the region of interest.

7) RADIATION DAMAGE (FILMRAD)

The observations from each crystal can be analysed for radiation damage as a function of $s=(\sin\theta/\lambda)^2$ and time t using one of two models:

i) The 'Blake-Phillips' correction:

$$I_0 = I_t \exp\left(\frac{-\mu_2+\mu_3}{t}\right)+\exp(-\mu_3 t)[1.0-\exp(-\mu_2 t)\exp(-Ds)]^{-1} \qquad 3.20$$

ii) a 'linear quadratic' correction:

$$I_0 = I_t(1.0+t(A+Bs+Cs^2)) \qquad 3.21$$

where I_0 is the intensity at zero time

I_t is the intensity at time t

μ_2, μ_3, D are physical constants for the decay process

A,B,C are empirical constants

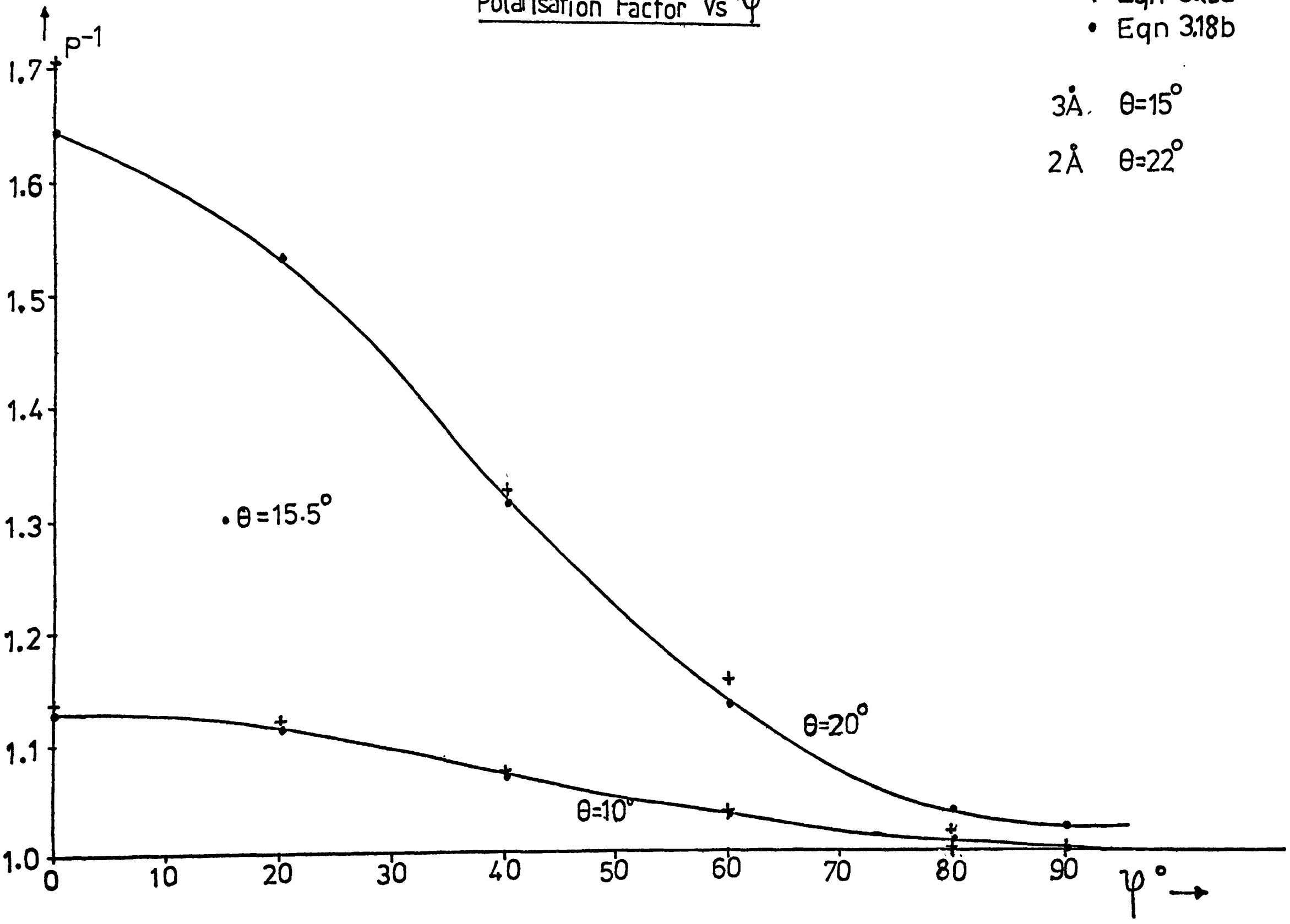
s is $\sin^2\theta/\lambda^2$ in (i) and $\sin\theta/\lambda$ in (ii).

Expression (i) is a formula devised by Hendrickson *et al.* (1973) from the model proposed by Blake and Phillips (1962) for radiation damage in protein crystals. This model considers the decaying crystal to contain

a) undamaged molecules which provide most of the diffracted intensity

Figure 3.17

Polarisation Factor vs φ



b) disordered molecules which diffract only at low angles

c) amorphous molecules which do not diffract.

μ_2 and μ_3 are the rate constants for disordering and destruction and D is a measure of the disorder. In this model the intensity of a reflection decreases with increasing time and scattering angle. Expression (ii) is a purely empirical model based on general observations of the form of decay of protein crystals.

The constants are determined, after sorting equivalent terms from the photographs, by a least-squares minimization of the expression:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\sigma_j^2}(I_{tj}-\bar{I}_0/C_{tj})^2 \qquad 3.22$$

where N is the total number of sets of equivalent terms

I_{tj} is the intensity of term j in a set of n equivalent terms at time t

σ_j is the standard deviation of term j

C_{tj} is the current estimate of the correction calculated from 3.20 or 3.21

$\bar{I}_0$ is the current estimate of the mean intensity for this set corrected to time t = 0.

The observed intensities for this process may be obtained from a series of data photographs taken from one crystal or from the first photograph and a photograph over the same rotation range repeated at the end of the data set. Wilson and Yeates (1979) have reported that this method is not totally satisfactory. For the phosphorylase *b* 3 Å data both models gave a satisfactory fit to the decay observed for many crystals, but for more than 25% of the

crystals no satisfactory fit was obtained. They suggested that phenomena other than radiation damage were contributing to the variation of intensity with time. Such phenomena might include; evaporation or condensation of liquid around the crystal in the capilliary tube, fluctuation in the intensity of the X-ray source, variations in the developing conditions of sequential films. Such effects would be difficult to model and Wilson and Yeates suggest that the determination of temperature factors as well as relative scales by the method of Fox and Holmes (1966) might provide a solution to the problem. The determination of relative scales is described in the next section.

8) REDUCTION TO UNIQUE DATA (SSM)

The programme SSM (scale-sort-merge) determines inter-film scales, sums the separate parts of partially recorded reflections and merges equivalent terms to give a unique data set. Data recorded by any method may be included to fill in the blind region or increase the correlation between films. If a satisfactory radiation damage correction has been applied all of the photographs from one crystal should require approximately the same scale factor and ideally should show the same variation of intensity with scattering angle. Owing to difficulties in modelling radiation damage and to other possible phenomena which affect the variation of diffracted intensity with time (see previous section), a residual trend in the scale factors may be apparent (Wilson and Yeates, 1979). A possible solution to this problem lies in the calculation of both a relative scale and temperature factor for data from each photograph. The relative temperature factor should correct for residual effects of radiation damage, crystal disorder and those differences in absorption which are dependent on scattering angle. The programme SSM uses a method implemented by

Helliwell and Wilson (unpublished work) which is based directly on that of Fox and Holmes (1966). The algorithm of Fox and Holmes is extended to calculate both the scale C_j and temperature factor B_j of the form:

$$K_j = C_j \exp(-2B_j \sin^2\theta/\lambda^2)$$

where K_j is the scale to be applied to the intensity of a reflection at angle θ on the jth film. Fox and Holmes calculate the reciprocal correction:

$$G_j = 1/C_j . \exp(2B_j \sin^2\theta/\lambda^2)$$

by minimizing:

$$\sum_h \sum_j w_{hj} (I_{hj} - G_j I_h)^2 \qquad \text{(taking } B_j = 0\text{)}$$

where I_h is the mean intensity of a set of equivalent reflections given by:

$$I_h = \frac{\sum w_{hk} G_k I_{hk}}{\sum w_{hk} G_k^2}$$

Helliwell and Wilson have modified this algorithm (as proposed by various members of the MRC Laboratory of Molecular Biology, Cambridge) to calculate shifts to the real scale and temperature factors. The two sets of parameters are refined independently in a cyclic manner, starting with scale factors. The scale factors are weighted by the inverse of the variance of the observations (estimated as in section 6). Wierenga (1978) has described a method

for refining both scale and temperature factors simultaneously. However, the available computing facilities have meant that this approach could not be used owing to the large amount of store required. After application of the inter-film scale factors the partially recorded terms are summed to give complete intensities. The following analysis is performed to give improved estimates of the standard deviations. The distribution of the expression:

$$(F_i^2 - \langle F^2 \rangle)/\sigma(F^2) \qquad 3.23$$

in ranges of mean F^2 is tabulated for all sets of equivalent terms. F_i^2 is the modulus squared of the ith member of a set of equivalent terms, $\bar{F}^2$ the mean-squared modulus for the set and σ_F the standard deviation of F_i^2. This distribution should be normal, with a mean of 0.0 and a standard deviation of 1.0, if σ_F is a true estimate of the error. The standard deviation is modified according to Dodson (1976), using an expression originally derived for diffractometer data:-

$$\sigma_{new} = A(\sigma_{old}^2 + B^2\bar{F}^2)^{\frac{1}{2}} \qquad 3.24$$

and the constants A and B are varied empirically to obtain a distribution with the required properties. The mean weighted intensity and its standard deviation are then calculated for the unique terms. Data from different crystals and different films are averaged for each data set. Negative mean intensities are not rejected and their treatment is described in the next section.

9) THE TREATMENT OF NEGATIVE INTENSITY OBSERVATIONS (TRUNCATE)

Although the true intensities of all reflections from a structure are positive, counting errors may result in negative

values being measured. This is particularly true for reflections with small structure factor moduli. The treatment of such negative terms is a problem, since omitting them or setting them equal to zero will result in bias in the distribution of structure factor amplitudes. French and Wilson (1978) propose a method based on Bayesian statistics which aims to provide more reliable estimates for structure factor moduli, whether positive or negative. It is assumed that each measured intensity (I) with its estimated standard deviation (σ) may be represented as a Gaussian distribution. The probability distributions of X-ray intensities are described by Wilson (1949, 1950). These probability distributions are:

$$P(I) = S^{-1}\exp(-I/S) \quad \text{for centric reflections}$$
$$P(I) = (2\pi SI)^{-\frac{1}{2}}\exp(-I/2S) \quad \text{for acentric reflections} \qquad 3.25$$

where $S = \langle I \rangle$, the mean intensity.

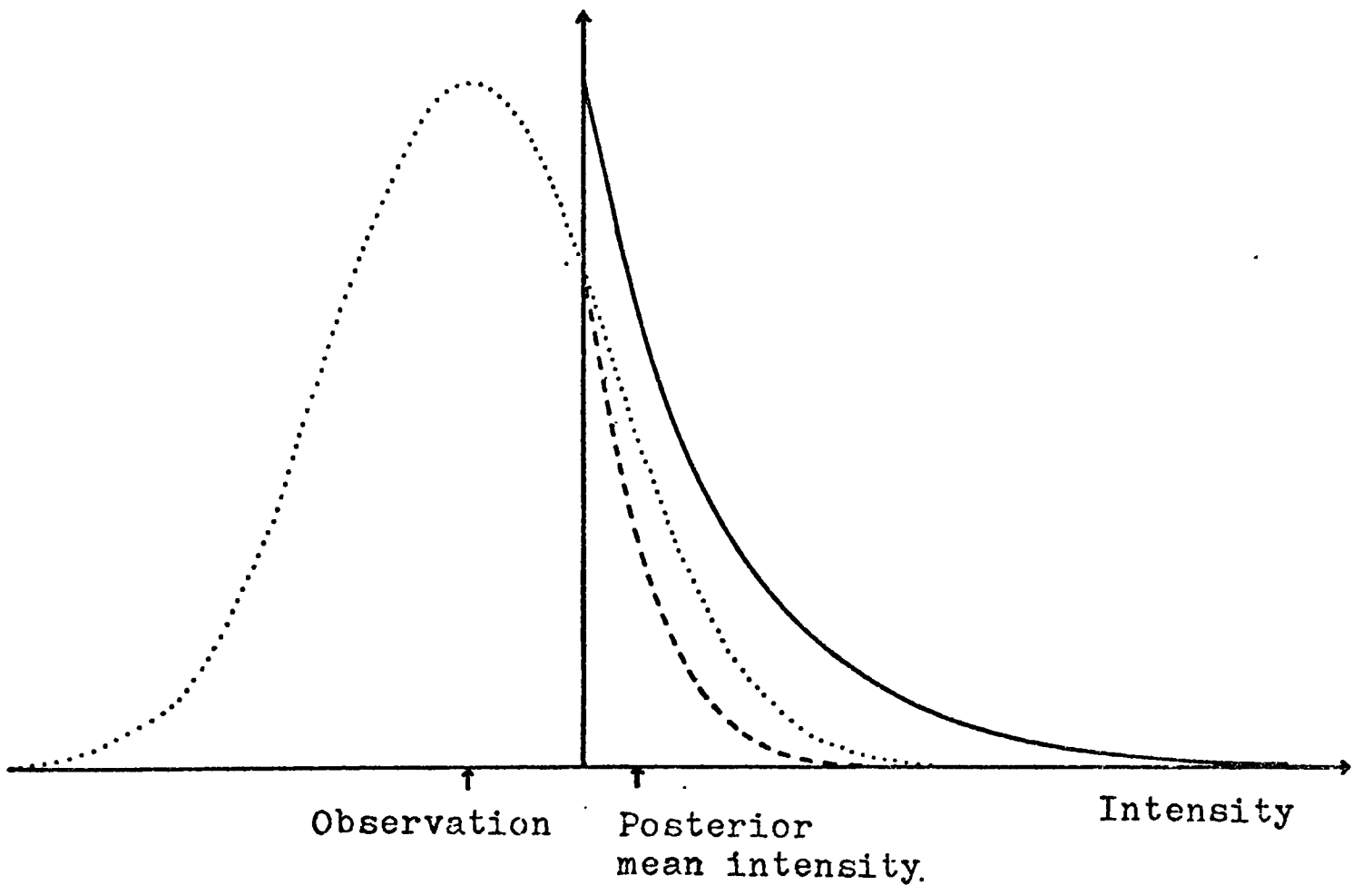
These prior distributions are combined with the observed Gaussian distribution for each intensity to produce, by Bayes' theorem, a posterior distribution which is positive and truncated (posterior $\propto$ prior x likelihood). This is illustrated in Figure 3.18. The programme TRUNCATE performs this operation on the data in ranges of $s = 4\sin^2\theta/\lambda^2$ using the standard deviations output by SSM. All observations, positive or negative, are treated in the same manner. The effect of the procedure is most evident on the smallest observations and is negligible for observations greater than about 3 standard deviations in the acentric case and about 6 standard deviations in the centric case.

10) SCALING OF DATA TO THE NATIVE INTENSITIES

Native, derivative and substrate data sets are collected together by the programme CAD and the derivative and substrate sets

FIGURE 3.18

The Truncation Procedure for an Acentric Reflection



— The prior distribution, expression (5.5) with S calculated in ranges of $4\sin^2\theta/\lambda^2$

--- The posterior distribution which is positive and truncated.

..... The observation, represented by a Gaussian curve of mean, I, and standard deviation, σ, from expression (5.3).

The posterior is proportional to prior x likelihood according to Bayes' Theorem as described by French & Wilson (1977), where the likelihood is the Gaussian distribution representing the observation.

are scaled to the native intensities by the programme ANSC with a scale factor K defined as:

$$K \frac{\sum F_{PH}^2}{\sum F_P^2} = 1 \qquad 3.26$$

where F_{PH} is the structure factor amplitude for the derivative

F_P is the structure factor amplitude for the native protein.

The scale factor is analysed for variation with h,k,l, $4\sin^2\theta/\lambda^2$ and F^2. It is also analysed in volumes of reciprocal space, taking groups of index level of the rotation axis together and within each set of levels divided into ranges of ϕ and ξ^2 (cylindrical polar coordinates). A second routine then applies the scale factors. These may be a linear function of any of the reciprocal space variables analysed (*i.e*: K = a+b$\underline{h}$ *etc*., where the constants a and b are obtained from the previous analysis), or applied to the data in blocks of reciprocal space. The latter method is a 'local scaling' technique, similar to that described by Matthews and Czerwinski (1975), who calculate a scale factor K to minimize the expression:

$$\sum_h w\left(K^{-1} - \frac{F_2(h)}{F_1(h)}\right)^2$$

where the summation is only over the reflections in the 'local' region of reciprocal space.

CHAPTER 4

THE COLLECTION, PROCESSING AND REFINEMENT OF HEAVY ATOM DERIVATIVE DATA

1) HEAVY METAL DERIVATIVE SEARCH

A previous search for heavy metal derivatives of phosphorylase *b* had yielded two compounds which produced isomorphous derivatives; ethylmercurythiosalicylate (EMTS) and platinum diaminodichloride (Pt) (Wilson, 1974). It was discovered that there were several mercury binding sites and that variations in the soaking conditions produced varying occupancies of these sites. Table 4.1 shows the refined parameters and soaking conditions for the derivatives used to calculate 3 Å phases in hol projection as described by Johnson *et al.*, (1978a). Three almost identical EMTS photographs were included, to reduce observational errors. Since the sites have different occupancies, the photographs were treated independently.

Table 4.2 gives the refined parameters used for the calculation of the 3 Å map described by Weber *et al.* (1978). Ambiguities in the interpretation of this electron density map made it desirable to have additional heavy atom derivatives and so a new search was undertaken. Because of the possible variations in occupancy of the mercury sites which could be obtained, a variety of different soaking conditions using EMTS were tried. A number of other heavy atom compounds which had been successful with other proteins were also screened. The main results of this search are summarised in Table 4.3. Unless otherwise stated, the diffusion of heavy atoms into the crystal took place by soaking in a solution of the heavy atom salt in a buffer of 10 m*M*-N BES (N-bis(2-hydroxymethyl)-2-aminoethane sulphonic acid), 10 m*M* magnesium acetate, 2 m*M*-IMP, and 0.1 m*M* EDTA at pH 6.8. The preparation of the crystals has been described in detail by Weber (1977). Because the signs of

TABLE 4.1

Derivatives used for refinement of 3 Å phases in projection

Site		Occ.	x	y	z	B	N_{obs}	f_H	E	R_c %
EMTS1	Hg 1	0·76	0·195	0·226	0·037	46	741	169	94	50·1
	2	0·85	0·029	0·088	0·384	44				
	3	0·37	0·161	0·383	0·322	38				
	4	0·18	0·198	−0·028	0·473	40				
EMTS2	Hg 1	0·69	0·195	0·223	0·038	42	739	165	106	53·8
	2	0·86	0·031	0·089	0·386	39				
	3	0·24	0·154	0·379	0·330	40				
	4	0·16	0·211	−0·018	0·469	40				
EMTS3	Hg 1	0·67	0·196	0·226	0·038	42	726	164	100	51·8
	2	0·86	0·030	0·088	0·386	38				
	3	0·18	0·147	0·379	0·330	40				
	4	0·15	0·196	−0·013	0·475	40				
EMTS	Hg 1	0·88	0·202	0·227	0·034	73	717	151	109	57·3
1MP	2	0·63	0·030	0·089	0·387	21				
	3	0·15	0·150	0·378	0·322	40				
	4	0·16	0·190	0·016	0·470	40				
PtEMTS	Hg 1	0·28	0·202	0·227	0·031	53	722	116	84	63·0
	2	0·33	0·026	0·093	0·384	44				
	Pt 1	0·53	0·147	0·370	0·333	25				
	2	0·27	0·171	0·358	0·347	25				
Pt	Pt 1	0·73	0·150	0·362	0·338	25	712	160	107	62·8
	2	0·50	0·172	0·355	0·351	25				
	3	0·40	0·254	0·384	0·261	25				
	4	0·21	0·280	0·496	0·294	25				

EMTS1, 2 and 3, 1 mM-EMTS, 1 day soak.
EMTS 1MP, 1 mM-EMTS in 1MP-free crystals, 1 day soak.
PtEMTS, 0·5 mM-EMTS for 1 day, then 0·5 mM-Pt for 1 day.
Pt, 1 mM-Pt $(NH_3)_2Cl_2$ for 1 day.
Mean figure of merit 0·88.
N_{obs}, Number of observations.
f_H, Root-mean-square value of calculated structure factor.
E, Root-mean-square error between observed and calculated structure factors.
R_c Centric R value: $\frac{\sum_h ||F_{PH} - F_P| - f_H|}{\sum_h |F_{PH} - F_P|}$ where summation over all centric terms.

check on the noise level in these syntheses, a difference map was also calculated for the 2 native photographs.

(from Johnson et al, 1978a)

TABLE 4.2

The Final Refinement Parameters for the two Derivatives

(from Weber et al.,1978).

Derivative	Site	O	x	y	z	B	rmsE	rms f_H	No of reflections	R_c
EMTS	Hg 1	.46	.1943	.2244	.0392	32.7	85	118	17999	0.55
	Hg 2	.67	.0288	.0885	.3838	21.3				
	Hg 3	.16	.1597	.3796	.3246	25.0				
$Pt(NH_3)_2Cl_2$	Pt 1	.53	.1496	.3634	.3385	10.0	91	108	18199	0.66
	Pt 2	.30	.1721	.3589	.3514	10.0				
	Pt 3	.18	.2517	.3830	.2530	25.0				

O is occupancy, E is the rms lack of closure error, f_H is the rms calculated structure factor,

$$R_c = \frac{\Sigma \left| F_{PH} - |F_P + f_H| \right|}{\Sigma |F_{PH} - F_P|} \quad \text{for centric data}$$

The overall figure of merit is 0.54

Table 4.3

Heavy atom reagent	MW	Concentration	Soaking time	Intensity changes (hol)	Comments
EMTS	404	1 m*M*	2 days	√	EMTS4
		2 m*M*	1 day		
EMTS	404	1 m*M*	2 weeks	√	EMTS5
EMTS	404	1 m*M*	1 week	√	EMTS7
Iodoacetamide + EMTS		10 m*M*	5 hrs	√	EMTSIA
		1 m*M*	1 day		
$Pt(NH_3)_2Cl_2$	299	1 m*M*	6 days	√	PT1
Aurothioglucose*	392	1 m*M*	2 days	some changes	AUG
Aurothioglucose*		100 m*M*	2 days	very large changes	
Aurothioglucose*		10 m*M*	2 days	-	crystals cracked
$KAu(CN)_2$	288	1 m*M*	2 days	√	AU1
		1 m*M*	7 days	√	AU2
		1 m*M*	5 weeks	√	AU3

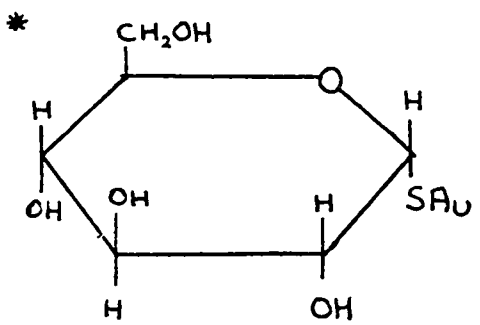
* 

Table 4.3 (cont.)

Heavy atom reagent	MW	Concentration	Soaking time	Intensity changes (hol)	Comments
$K_2Pt(CN)_4$	377	1 m*M*	2 days	-	crystals cracked
$Pt(NH_3)_4Cl_2$	333	1 m*M*	2 days	no changes	
$Pt(Cl)_4$		1 m*M*	-	-	crystals cracked
$Pb(NO_3)_2$	331	1 m*M*	-		in 10 m*M* cacodyate buffer pH 6.05 - crystals cracked
		1 m*M*	1 day		10 m*M* cacodyate buffer + 10 m*M* magnesium acetate + 2 m*M* IMP pH 6.8 -crystals cracked

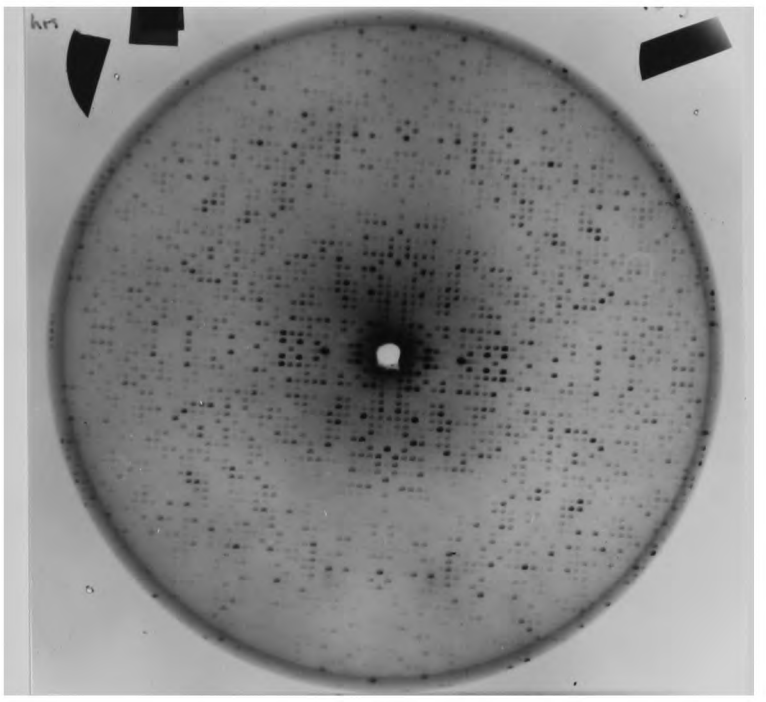
hol reflections were known, the powerful difference Fourier technique described in Chapter 1 could be used. Screened precession photographs of the hol zone were taken, using a precession angle of 15° and a crystal to film distance of 100 mm, corresponding to a 3 Å resolution limit. The photographs were recorded using a Supper precession camera with Ni-filtered CuKα radiation from an Elliot GX6 rotating anode run at 40KV, 40mA. The collimator cross-section was 0.45 mm and the exposure time for a two-film pack of Ilford Industrial G film was 20 hours. Photographs were digitized and output to magnetic tape using the Optronics P-1000 densitometer with 100 μm raster and light spot, and 0-2 O.D. range. The following photographs (see Table 4.3) were processed in this manner:-

EMTS4, EMTS5, EMTS7, PT1, AU1, AU2, AU3, AUG

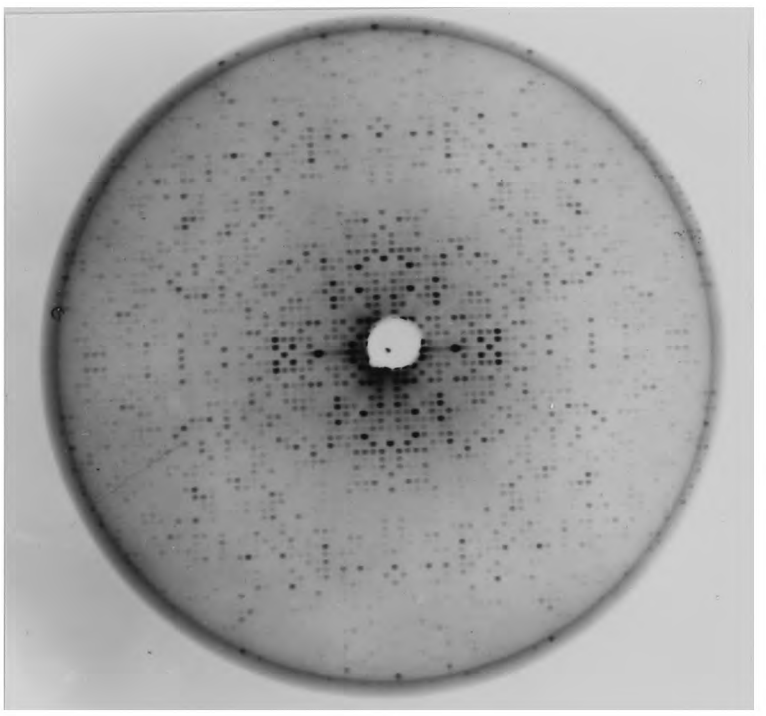
Figure 4.1 shows photographs for the hol zone of native phosphorylase *b*, EMTS7 and AU3. The reflections were integrated using the programme PERCY (described in Appendix 1) on the ICL 1906A computer. The operation of the programme is very similar to OSCAR, described in Chapter 3. Three fiducial holes were punched at known positions parallel to the crystal axes, and subsequently used to determine an approximate orientation matrix to define the orientation of the crystal lattice relative to the Optronics coordinate system. This matrix was refined using the positions of about 40 strong reflections. Each optical density point was corrected for the non-linear response of the film and densitometer from an empirical scale (Figure 3.8). The reflections were integrated by summing the optical density values in a box of size 1.1 x 1.3 mm^2, and subtracting four background corrections (defined in Figure 3.13). A subsequent test for residual non-linearity was made by

FIGURE 4.1

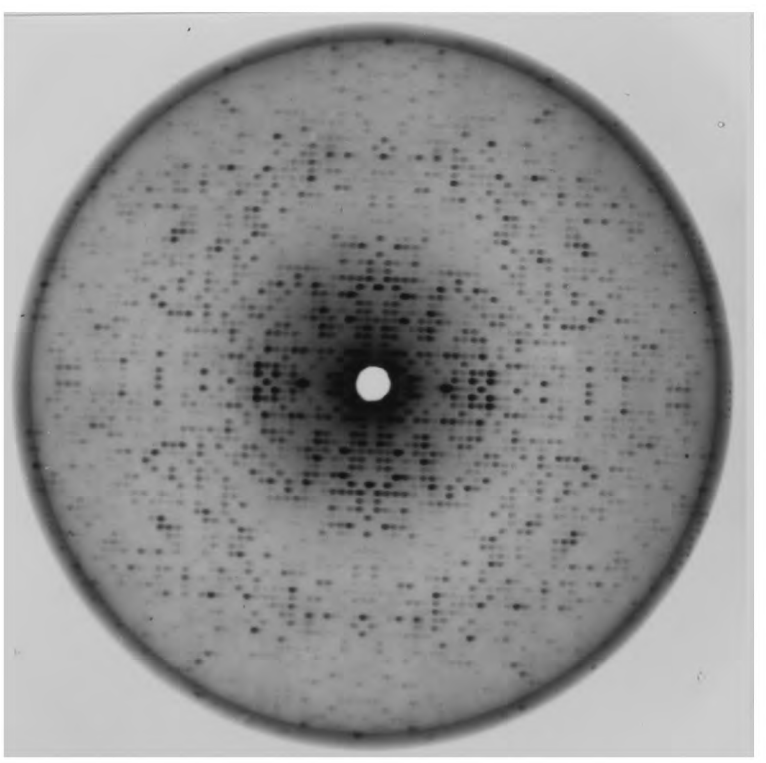
hOl Precession Photographs of Heavy Atom Derivatives of Phosphorylase b.



Phosphorylase b crystal soaked in 1mM EMTS for 7 days (EMTS7). Shows intensity changes when compared to native, such as reflections (0 0 12) and (18 0 22)



Native phosphorylase b crystal.



Phosphorylase b crystal soaked in 1mM $KAu(CN)_2$ for 5 weeks (AU3). Shows intensity changes when compared to negative, such as (10 0 2) and (0 0 28).

the programme FILMPACK, and a parabolic scale of the form;

$$I(\text{corrected}) = k*I+b*I^2$$

was applied to the intensities (Matthews *et al.*, 1972). The parabolic component corrects for residual non-linearity of film response at high optical densities, which would otherwise result in underestimation of the strong reflections. The weak and strong intensities were merged according to the expression:-

$$I = (I_w/k_w+I_s/k_s)/(1/k_w^2+1/k_s^2)$$

where I_s and I_w are the observed intensities on the strong and weak films and the k's are the calculated scale factors, given by the scaled intensity (including the parabolic component) divided by the observed intensity. The equation assumes that the error in intensity is independent of the absolute optical density.

No absorption correction was applied. Lorentz and polarization corrections were applied by FILMPACK and the negative intensities retained and processed by TRUNCATE (see Chapter 3). The data were scaled to the native intensities using a scale factor K, where

$$K \frac{\sum F^2_{PH}}{\sum F^2_{P}} = 1$$

as described in Chapter 3, Section 10. The scale factors were analysed for variation with h,k,l or $\sin^2\theta/\lambda^2$ and corrected for any systematic trends. The corrections applied are shown in Table 4.4. The agreements between equivalent reflections and fractional isomorphous changes are summarized in Table 4.5.

The data were then processed by the programme FHLEKR which

Table 4.4

Corrections applied to Scale factor by ANSC

Derivative	Correction	Slope (b)	Intercept (a)
EMTS4	s	2.8679	0.9496
EMTS5	s	1.5165	0.9530
PT1	s	2.6263	0.9755
AU1	overall scale	0.8283	-
AU2	overall scale	0.8536	-
AU3	l	0.00501	0.6970
AUG	l	-0.00215	0.9183

Linear corrections of the form a + b$\underline{h}$ *etc.* are applied to the scale factor for variations with $s = 4\sin^2\theta/\lambda^2$, h, k or l. The slope and intercept are listed.

Table 4.5

Derivative	Merging R* Strong	Merging R* Weak	Scale applied to weak film	Fractional isomorphous change	Mean isomorphous difference
EMTS4	10.2	23.4	3.1	23.9	145.9
EMTS5	9.2	21.9	3.0	24.1	146.5
EMTS7	15.7	21.1	3.1	23.9	145.7
EMTSIA	12.3	29.1	3.0	-	-
PT1	6.9	13.5	3.1	19.1	116.2
AU1	7.0	12.3	2.9	14.3	87.1
AU2	9.1	19.0	2.9	16.1	97.7
AU3	9.3	13.0	2.7	17.2	104.7
AUG	6.7	12.7	3.1	13.8	84.0

* $$R = \frac{\sum_{h}^{N} \sum_{i=1}^{N} |\bar{I}(h) - I_i(h)|}{\sum_{h}^{N} \sum_{i=1}^{N} \bar{I}_i(h)}$$

calculates and refines the scale factor K by the method of Kraut (Arnone *et al.*, 1971) as described in Chapter 1, Section 3d. This involves the iterative application of the equation:

$$\sum F_H^2 = \sum K^2 F_{PH}^2 - \sum F_P^2$$

where $F_H = |\underline{F}_{PH}| - |\underline{F}_P|$ for centric data.

The following parameters for the heavy atoms were refined using the programme REFINE; occupancy, positional parameters (x,y,z) and temperature factor B. The programme minimizes the expression:-

$$w^2 (F_H - F_{Hcalc})^2$$

where $F_H = |\underline{F}_{PH}| - |\underline{F}_P|$ for centric data.

Here the weights $w = 1/\sigma^2$, where σ is the modified estimated standard deviation of F_H. If the Kraut scale factors were refined together with the heavy atom parameters, they dropped below unity. Since this would be physically unrealistic, they were kept fixed during the refinement. The values of the Kraut scales are given in Table 4.6. The calculation of Kraut scales and centric refinement was carried out for the new derivatives and also for the earlier derivatives, since these had not previously been processed using TRUNCATE and Kraut scales.

The results of these refinements are summarized in Table 4.6. All the EMTS derivatives were refined with two major and two minor sites, the Pt with one double and two minor sites (Weber, 1977), AU1 using a three site model and AU2 and AU3 with five sites. AUG was refined with 2 binding sites. The approximate coordinates of these sites had been obtained by calculating difference Fourier

Table 4.6

Derivative	Site	Occupancy	x	y	z	B	Kraut Scale	N_{obs}	rms FH	rms E	R%
EMTS 1	Hg 1	0.7707	0.1949	0.2248	0.0386	46.51	1.02839*	732	182	100	49.0
	2	0.8875	0.0281	0.0870	0.3834	34.48					
	3	0.3115	0.1608	0.3807	0.3209	24.17					
	4	0.1297	0.2012	0.0038	0.4763	40.00*					
EMTS 2	Hg 1	0.7787	0.1949	0.2230	0.0398	46.38	1.03747*	722	181	116	53.5
	2	0.8958	0.0305	0.0880	0.3837	32.13					
	3	0.1928	0.1592	0.3756	0.3206	40.00*					
	4	0.1457	0.2050	-0.0005	0.4789	40.00*					
EMTS 3	Hg 1	0.7190	0.1960	0.2252	0.0402	31.79	1.03594*	714	177	114	52.8
	2	0.9259	0.0304	0.0881	0.3841	43.68					
	3	0.1735	0.1538	0.3769	0.3196	40.00*					
	4	0.1607	0.2019	-0.0046	0.4799	40.00*					

* not refined.

Table 4.6 (cont.)

Derivative	Site	Occupancy	x	y	z	B	Kraut Scale	N_{obs}	rms FH	rms E	R%
EMTS4	Hg 1	0.7490	0.1961	0.2243	0.0384	49.61	1.03157*	733	175	114	54.0
	2	0.8595	0.0299	0.0886	0.3826	31.69					
	3	0.2426	0.1582	0.3820	0.3235	40.00*					
	4	0.1237	0.1976	-0.0213	0.4733	40.00*					
PT	Pt 1	0.9245	0.1516	0.3623	0.3385	50.00*	1.02942*	728	114	106	61.3
	2	0.4536	0.1748	0.3576	0.3535	50.00*					
	3	0.2931	0.2558	0.3872	0.2575	50.00*					
PT1	Pt 1	0.8914	0.1507	0.3630	0.3382	50.00*	1.02204*	724	141	105	61.3
	2	0.4677	0.1740	0.3586	0.3526	50.00*					
	3	0.2794	0.2557	0.3858	0.2572	50.00*					

Table 4.6 (cont.)

Derivative	Site	Occupancy	x	y	z	B	Kraut Scale	N_{obs}	rms FH	rms E	R%
PTEMTS	Pt 1	0.9270	0.1519	0.3576	0.3392	50.00*					
	2	0.2854	0.1778	0.3576	0.3551	50.00*					
	3	0.3514	0.2565	0.3842	0.2562	50.00*		698	148	114	61.1
	Hg 1	0.4518	0.1928	0.2236	0.0265	40.00*					
	2	0.1108	0.0363	0.0907	0.3964	40.00*					
AU 1	Au 1	0.3791	0.1522	0.3857	0.3223	25.00*					
	2A	0.2065	0.2370	0.9987	0.4540	25.00*	1.01078*	726	89	81	66.0
	2B	0.3158	0.2147	0.9706	0.4656	25.00*					
AU 2	Au 1	0.4176	0.1515	0.3836	0.3199	25.00*					
	2A	0.1843	0.2445	1.0037	0.4480	25.00*					
	2B	0.2332	0.2060	0.9726	0.4682	25.00*	1.01365*	726	112	90	63.7
	4	0.3567	0.1988	0.2209	0.0368	25.00*					
	5	0.2606	0.1899	0.1282	0.2559	25.00*					

Table 4.6 (cont.)

Derivative	Site		Occupancy	x	y	z	B	Kraut Scale	N_{obs}	rms FH	rms E	R%
AU3	AU	1	0.3402	0.1514	0.3852	0.3179	25.00*					
		2A	0.1906	0.2523	0.9973	0.4500	25.00*					
		2B	0.1690	0.2015	0.9776	0.4653	25.00*	1.02155*	704	125	90	59.7
		4	0.4775	0.1959	0.2232	0.0374	25.00*					
		5	0.4167	0.1884	0.1341	0.2542	25.00*					
AUG	AU	1	0.4564	0.1609	0.3804	0.3225	25.00*	1.01184*	718	86	80	71.0
		2	0.2428	0.2630	0.4971	0.2926	25.00*					

syntheses with coefficients:

$$(|F_{PH}|-|F_{P}|)\exp i\alpha_{iso}$$

where α_{iso} are the current isomorphous phases. After refinement of the atomic parameters, new difference Fourier syntheses with coefficients:

$$(|F_{H}|-|F_{Hcalc}|)\exp i\alpha_{calc}$$

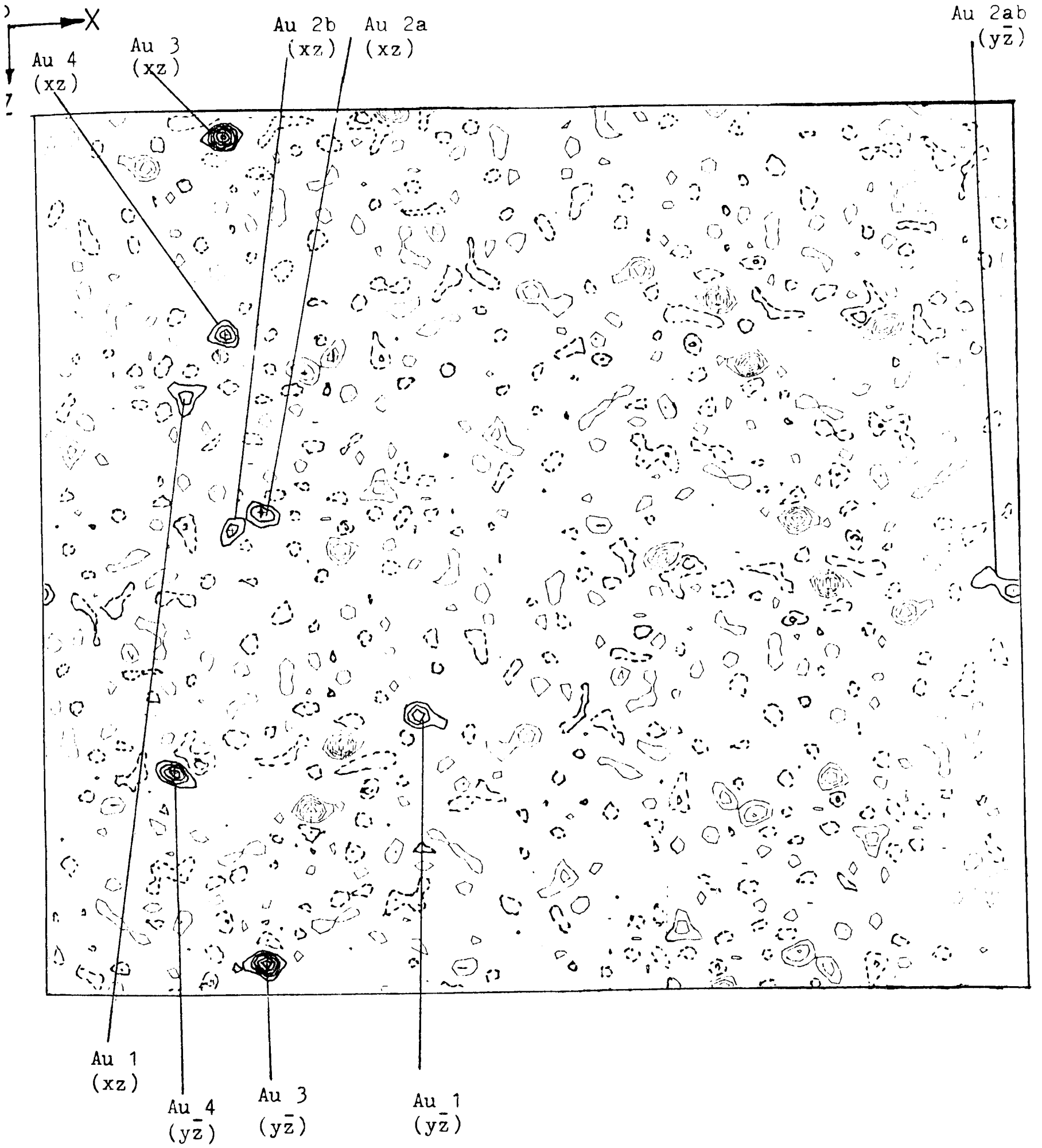
were calculated. These difference maps were used to search for the existence of minor sites and to check for errors in the refined values of position, occupancy and temperature factor for the heavy atoms. The quality of the refinements were monitored by three criteria; the R factor (equation 1.13b), the ratio F_H/E (where E is defined in equation 1.13a), and the difference Fourier map after refinement.

The best refinements for the gold ($KAu(CN)_2$) derivative were obtained by modelling the binding in the vicinity of the nucleotide binding site (site N) as a split site, as shown in Table 4.6 (Sites 2A and B). Figure 4.2 shows the difference Fourier synthesis for AU3 contoured at sixteen levels of -40 -30 -20 -10 10 20 30...120 in arbitrary units. The 3 Å difference Fourier in projection must be solved for the eight equivalent positions in the unit cell of space group $P4_32_12$, given in Table 2.1. These give rise to two sets of identical peaks, four at coordinates (x,z) and four at $(y,\bar{z})$ and symmetry related positions. The positions of the various heavy atom binding sites are discussed with reference to the folding of the molecule in Section 7.

A new phase calculation was carried out using all three derivatives. AUG was excluded since it had the lowest fractional

FIGURE 4.2

3 Å Projection Difference Fourier Map for Derivative AU 3



isomorphous change of all the gold derivatives. Hence the derivative data included in the phase calculation comprised the following:-

EMTS1,2,3,4; PT, PT1; AU1,2,3. (see Table 4.3)

The programme PHASE requires the following parameters for each derivative; occupancy, coordinates (x,y,z) and temperature factor for each heavy atom site, an overall scale (the Kraut scale was used) and an estimate of the rms error (E) in the isomorphous contribution to the heavy atom structure factor (from centric terms only). The rms errors are shown as functions of $\sin^2\theta/\lambda^2$ in Figure 4.3a. The parameters from the refinements given in Table 4.6 were used. For the 1108 reflections used in the phase calculations the mean figure of merit (m) was 0.9071, compared to 0.88 using the derivatives listed in Table 4.1. The variation of m with $\sin^2\theta/\lambda^2$ is shown in Figure 4.3b. After phasing 'double' difference Fourier syntheses with coefficients

$$(|\underline{F}_{PH}|-|\underline{F}_{P}+\underline{f}_{H}|)\exp i\alpha_{calc}$$

were calculated but these did not show any significant features. The new hol phases were used to calculate projection difference Fourier maps for the substrates and effectors (Weber, 1977). As an example, Figure 4.4 shows the difference Fourier maps for glucose-1-phosphate calculated with both sets of phases. The substrate peaks on the new map are clearer than before, although the map appears more noisy, owing to the inclusion of a lower contour level. These new hol phases have since been used for the screening of many more substrates.

FIGURE 4.3a

Estimated rms errors in the isomorphous contribution of the Heavy Atom Derivatives used in the centric phase calculation.

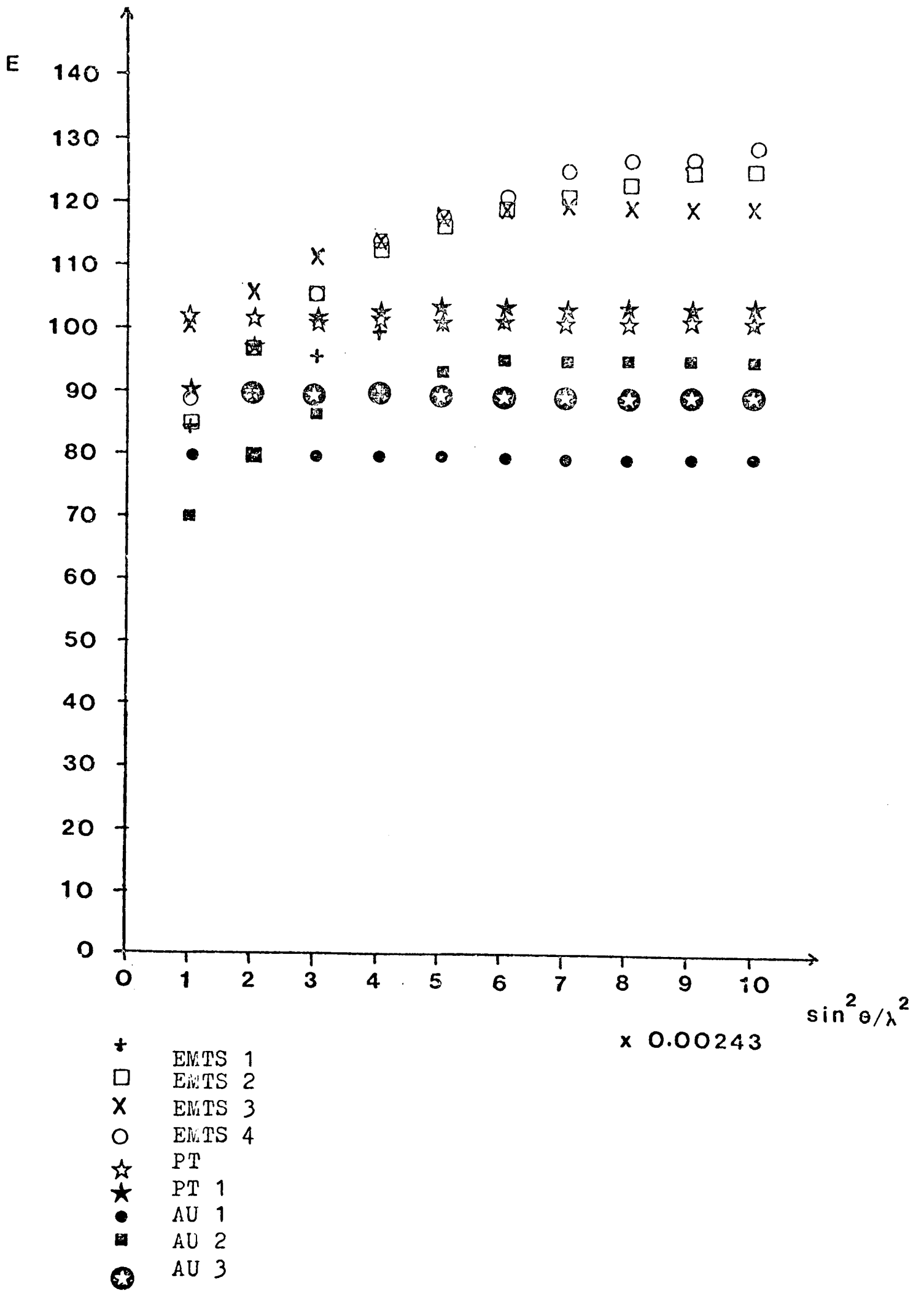


FIGURE 4.3b

Variation of figure of merit with $\sin^2\theta/\lambda^2$ for centric phase calculation.

The overall figure of merit was 0.9071 for 1108 reflections.

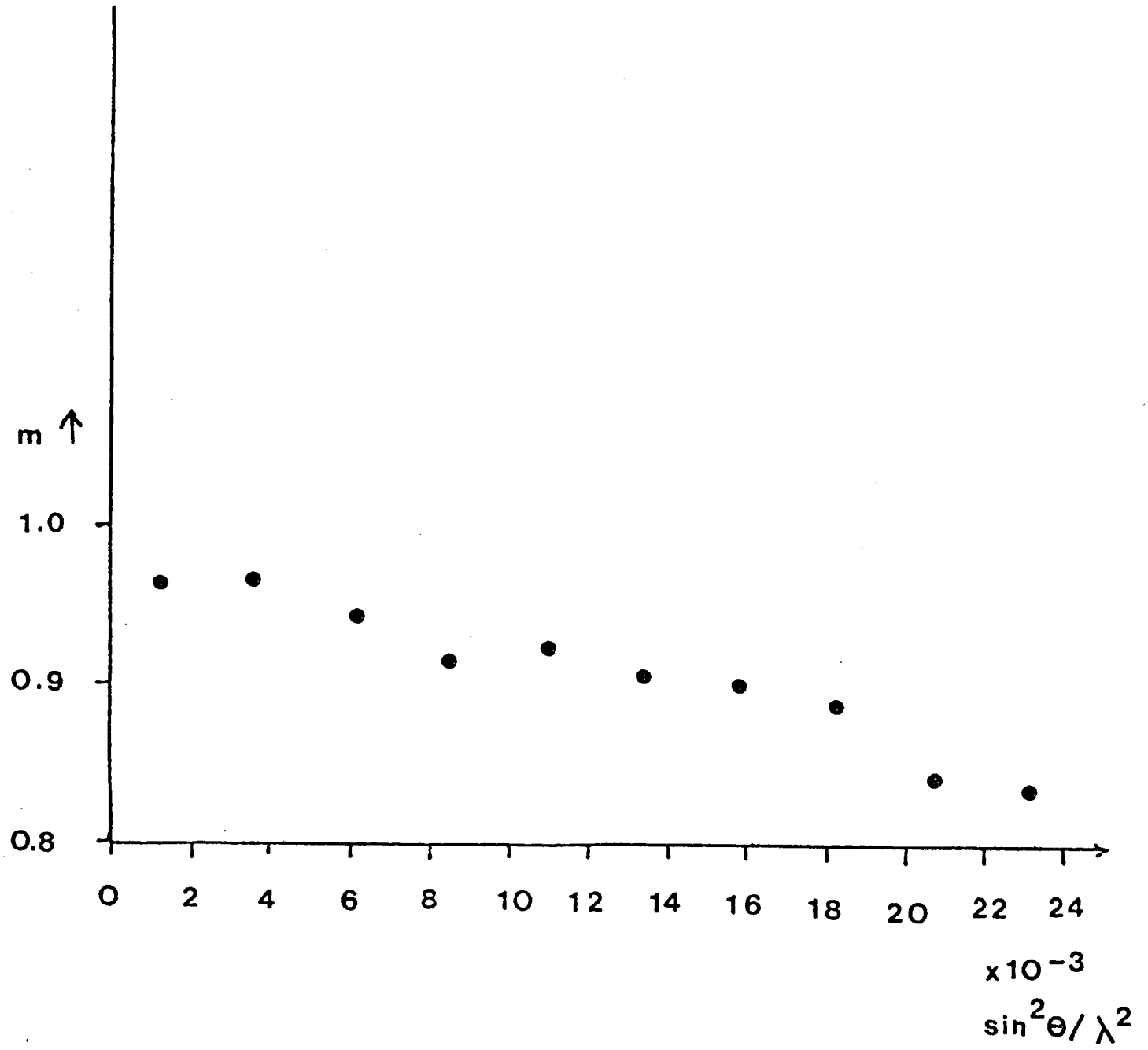
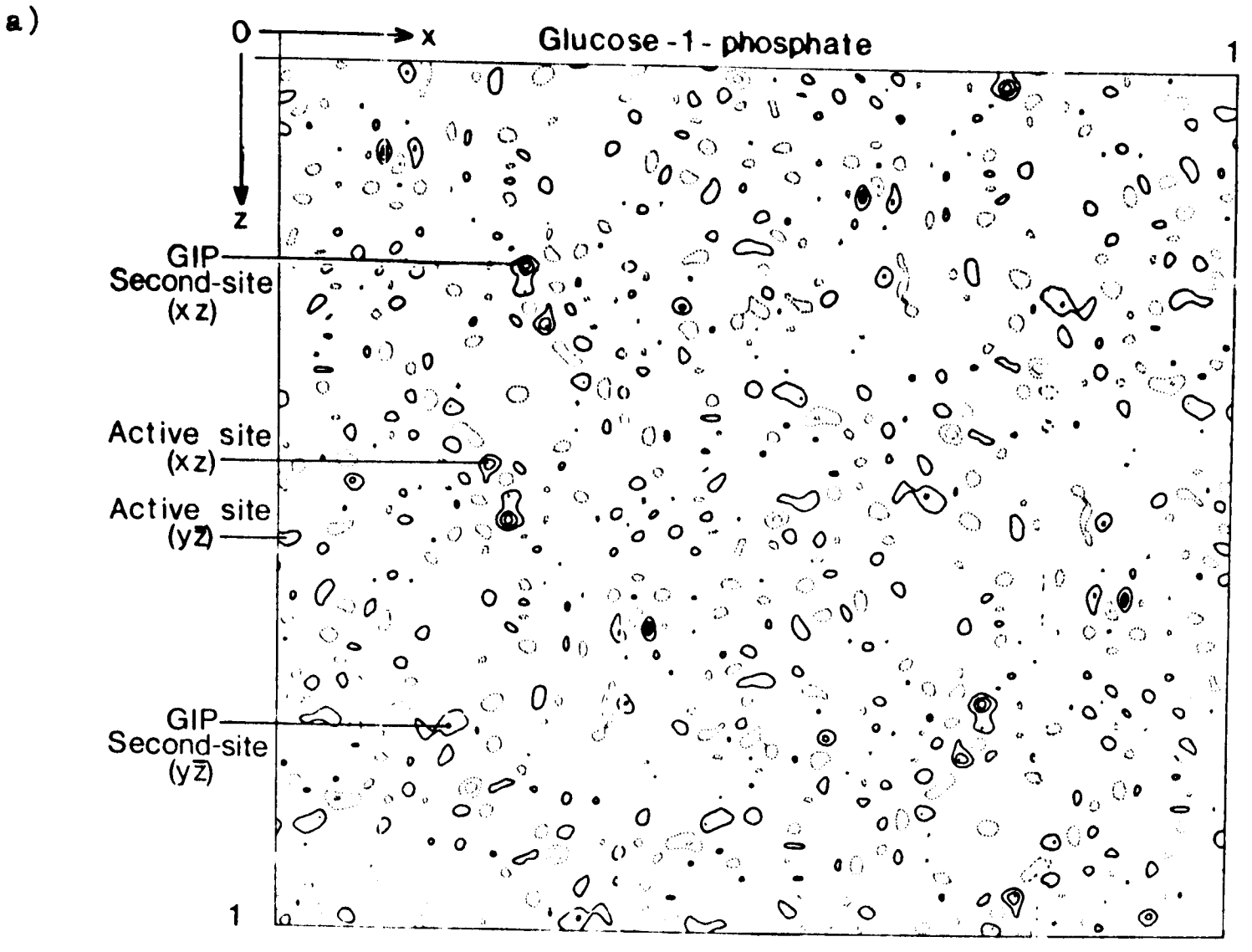


FIGURE 4.4



3 Å projection difference Fourier maps for Glucose-1-phosphate calculated with a) phases from Johnson et al,(1978), b) new phases.

Note:- the sites marked "Active Site" and "G-1-P Second Site" were later renamed Site N and Site C respectively (Weber et al.,(1978)).

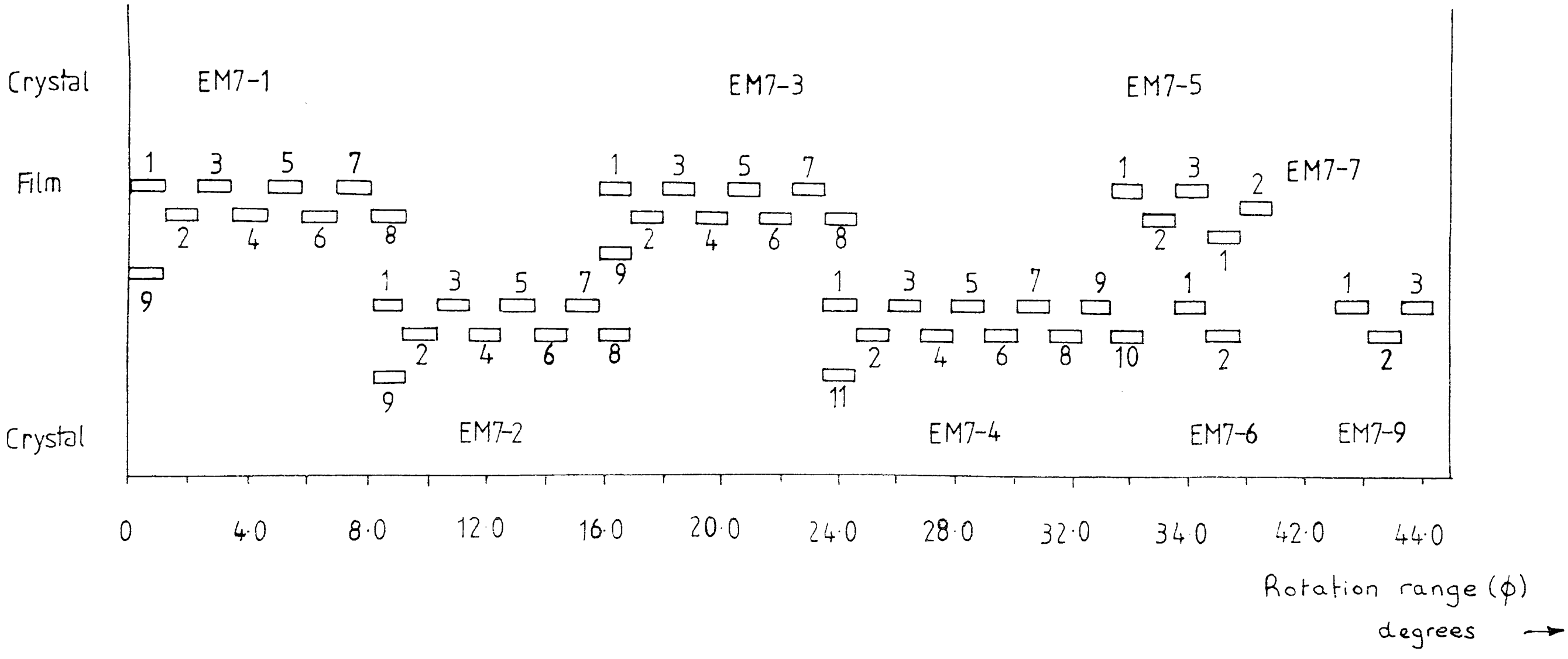
2) COLLECTION AND PROCESSING OF 3 Å DATA FOR EMTS7

a) Data Collection

From Table 4.6 it can be seen that the AU derivative gave, in general, poorer refinement parameters than EMTS, and so it was decided to produce a second EMTS derivative by soaking crystals in 1 m*M* EMTS for seven days to increase the occupancy of site HG3. Longer soaking times tended to crack the crystals. In fact, great difficulty was experienced in obtaining reproducible results from a seven day soak. Often the crystals would crack, or, if intact when mounted, would rapidly become disordered during data collection. Because of these difficulties a full data set (45° rotation) was not obtained. Figure 4.5 shows the range of data collected. This accounts for 93.8% of the complete data set to 3 Å resolution. Data were collected on the Enraf-Nonius oscillation camera by the method described in Chapter 3. An Elliot GX6 rotating anode generator run at 40KV, 40mA was used, with Ni filtered CuKα radiation for data sets 1-3. For data sets 4-9 a graphite monochromator was used. The crystal to film distance of 96 mm allowed data to 3 Å resolution to be recorded. Small angle oscillation or still photographs at 0° and 90° were taken to determine the precise crystal orientation. For crystals 1-4, 1° oscillation photographs were taken at -0.5° to 0.5° and 89.5° to 90.5° with an exposure time of 1.1 hours. Later, after experience with using still photographs for crystal orientation (Chapter 3), this method was adopted for crystals 5-9. The data films were exposed for between 3.1 and 4.6 hours, with a rotation angle of 1.1°. Large crystals (*i.e*: ~0.8 x 0.8 x 1.5 mm^3) could be moved along in the beam after irradiation and another area of crystal exposed. However, due to radiation damage problems this was often impossible. The complete data set comprised 48 film packs and is shown schematically in Figure 4.5. Data for crystals 1-4 were collected on Ilford

Figure 4.5

Range of EMTS-7 data



Industrial G film and for crystals 5-9 on Kodirex film, after experience with previous data sets had shown that data of comparable quality could be obtained using the faster Kodirex film (Weber, 1977).

b) Data Processing

The data were processed by the method described in Chapter 3. Reflections were integrated using a box size of 11 x 11 raster units and four background areas of 4 x 4 raster units displaced by 2 raster units (Figure 3.12). Each density point was corrected for non-linearity using the scale appropriate to Kodirex or Ilford Industrial G. After integration the Lorentz, polarization and oblique incidence corrections were applied as described previously, and intensities from the strong and weak films scaled together and merged in FILMPACK. The results are listed in Table 4.7. Values of mean intensity on the strong film $\langle I_s \rangle$, merging R factors for the strong (R_s) and weak (R_w) films and At, the absorption term in the weak to strong scale factor, are averaged over the film packs for each crystal. The theoretical values of At are 1.075 for Ilford Industrial G and 1.061 for Kodirex (Morimoto and Uyeda, 1963). Differences between the measured and theoretical values may be due to non-standard developing conditions. No radiation damage or absorption corrections were applied to the EMTS7 data. Various methods have been used to correct protein crystal diffraction data for absorption effects. North *et al.* (1968) and Kopfmann and Huber (1968) describe empirical methods for the correction of diffractometer data. North *et al.* proposed to correct the intensity of a reflection by the arithmetic mean of two transmissions measured in the directions of the incident and reflected beams, whilst Kopfmann and Huber utilised the mathematical separation of the transmissions along the paths of the incident and reflected

Table 4.7

EMTS7: Integration and Merging

Data Set	No. of Filmpacks	$\langle I_s \rangle$	R_s%	R_w%	At
1	9	550	9.2	19.3	1.019
2	9	879	5.1	9.2	1.037
3	9	980	4.6	7.3	1.029
4	11	853	5.4	11.4	1.071
5	3	654	4.0	5.7	0.712
6	2	582	5.2	7.2	0.759
7	2	748	7.8	9.4	0.704
9	3	578	4.5	8.4	0.805

Parameters tabulated are:

$\langle I_s \rangle$ mean intensity on strong film

R_s merging R factor for strong film

R_w merging R factor for weak film

At scale factor, where corrected intensity on weak film is

$I' = I_w \exp(-At\sec 2\theta)$.

beams. However, both these methods utilise a diffractometer and on the rotation camera the crystal cannot be rotated to any desired position for the necessary measurements. Whilst the crystal could be transferred from the camera to a diffractometer this is unlikely to reproduce exactly the same collimation geometry. Another limitation of both the above methods is that the crystal must be small enough to see all of the X-ray source all of the time. This is not the case for phosphorylase *b* crystals.

Schwager, Bartels and Huber (1973) have described a method for exploring part of the transmission surface for a crystal mounted on a precession or rotation camera. The method relies upon measuring the relative attenuation of the direct X-ray beam for different positions of the crystal. On the precession camera the crystal is tilted by the precession angle (μ) and the motion yields transmission values on cones around the primary beam direction. On the rotation camera the spindle axis may be inclined and the crystal rotated about this new axis. The reflected beams then lie on cones around the spindle axis and for the transmitted intensity to be monitored on the film, the film cassette must move with a constant speed as the crystal is rotated. A device has been marketed by Enraf-Nonius which allows the film to be translated horizontally (K. Bartels in Arndt and Wonacott, 1977). This method can only be used when the beam is both monochromatic and entirely contained within the crystal. For the collection of the phosphorylase *b* 3 Å and 2 Å data the equipment mentioned above was not available and so the method of Schwager *et al*. could not be applied.

A different approach to the absorption problem was proposed by Katayama, Sakabe and Sakabe (1972). The transmission is approximated by a Fourier series in the polar angles of both the incident and diffracted beams. The coefficients of this series are obtained by the analysis of the intensity differences between symmetry-

related reflections within a single data set. This requires high lattice symmetry and also that equivalent reflections do not have the same absorption factors. Stuart and Walker (1979) have described a method for correcting existing structure factors for absorption and decay effects which uses some aspects of the approach of Katayama *et al*. Where an absorption-corrected subset of the data (for instance, a low-resolution data set) is available, this may be used as a reference set to obtain absorption factors for the high resolution data set by a least-squares fitting process. When this cannot be done, the method allows a relative correction between native and isomorphous derivative sets to be calculated. For phosphorylase *b* data to 3 Å resolution no absorption correction was applied. The crystals used were consistently larger than the beam in all directions. Hence random errors in the observation of structure factor amplitudes have been reduced at the expense of the introduction of some systematic errors due to absorption and decay. We justify this approach on two grounds. The morphology of the crystals is reasonably constant and the absorption might be expected to be similar for all such crystals. The systematic errors introduced should roughly be the same for all data. Secondly, the phases of the structure factors have been determined by the technique of multiple isomorphous replacement. In this method the data for the isomorphous derivatives are scaled to the structure factor moduli of the native protein before calculation of the phases. As described in Chapter 3, Section 10, a "local scaling" technique was applied to the phosphorylase *b* data, in which the data are scaled in blocks of reciprocal space to maximise the significance of the isomorphous differences throughout the data. This "local scaling" should effectively minimise systematic differences between the native and derivative moduli, and leave random errors as the most significant. This approach will not

remove systematic errors in the native moduli, but the removal of systematic errors in ΔF allows refinement and phasing to be carried out successfully. Figure 4.6 shows the scale factors F^2_{PH}/F^2_P for the EMTS7 derivative before and after "local" scaling. The scale as a function of h is presented as it showed the largest variation. The improvement after "local" scaling is evident from the figure. The actual relationship between absorption and scale factor is complex as the unique native data to which the unique derivative data are scaled are a merged set of observations from several crystals.

The derivative data were then scaled and merged using the programme SSM as described in Chapter 3, Section 8. Initial estimates of the scale factors were determined from the mean intensities of each photograph. This helped to reduce the number of iterations required for convergence. The first round of least-squares iterations refined the scale factors using only those terms corresponding to spacings greater than about 6 Å. This was done since the low-angle intensities are almost independent of the temperature factor and hence should give the best unbiased estimate of the true scale factors. Intensities which differed by more than four standard deviations from the current weighted estimate of the mean intensity were excluded from the scale determination. This prevented the possibility of a small number of terms with large systematic errors from dominating the scaling. The percentage of terms rejected is shown in Table 4.8. An odd number of rounds was performed, the procedure ending with the determination of scale factors. A convergence criterion for each round of 0.00001 for scale factors and 0.0005 for temperature factors was used.

The results for the EMTS7 derivative are shown in Table 4.9. The scale and temperature factors vary appreciably between

FIGURE 4.6

Variation of scale factor for EMTS 7 with h, before and after scaling.

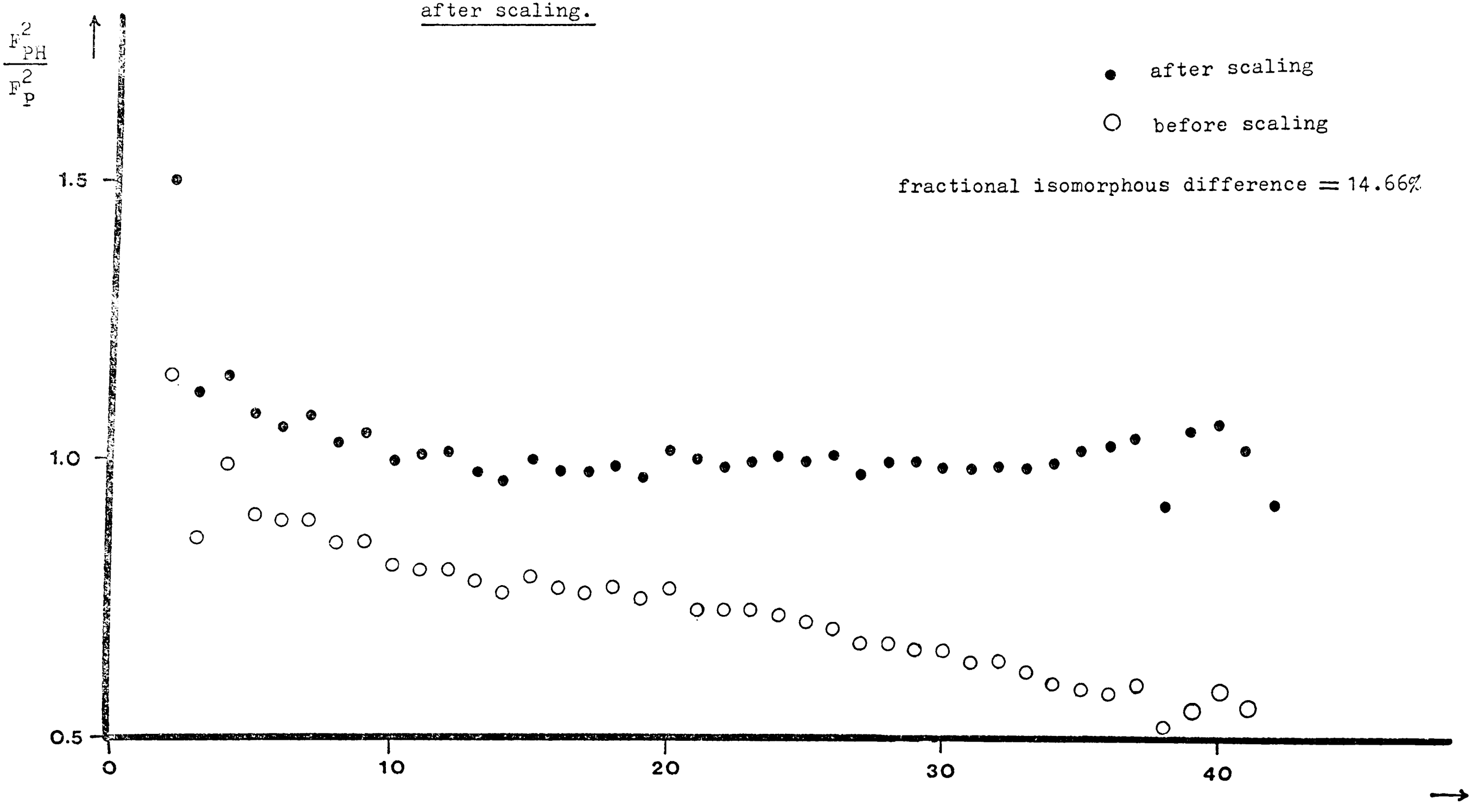


Table 4.8

	Native	EMTS7	EMTS	Pt (1&2)
Total No. of terms	67192	63067	80415	49384
No. of unique terms	18447	33838	36637	29104
% terms rejected	0.51	0.17	0.41	0.18
overall relative scale (Kraut)		1.02957	1.03069	1.02948
Merging R* Factor (%)	8.89	10.98	9.36	10.34
Merging R* Factor (%)**	10.24	-	9.76	11.80
Mean Isomorphous difference	-	98.4	107.9	101.9
Fractional Isomorphous difference (%)	-	14.70	16.05	15.10

$$^{*}\text{Merging R Factor} = \frac{\sum_h \sum_i |\bar{I}(h) - I_i(\underline{h})|}{\sum_h \sum_i I_i(\underline{h})}$$

$$\text{Fractional isomorphous difference} = \frac{\sum |F_{PH} - F_P|}{\sum F_P} \times 100$$

** From Weber (1977)

Table 4.9

Results of SSM for EMTS7 data

Photograph	Scale (K)	Temperature Factor (B)	R Factor (%)
EM7-1-1	1.00000	0.00000	11.4
1-2	1.11100	1.89668	11.7
1-3	1.18684	-0.25816	11.8
1-4	1.06514	-0.47719	12.8
1-5	1.20343	-0.64086	13.2
1-6	1.22309	-0.66420	10.5
1-7	1.28945	-0.75983	11.5
1-8	1.25817	-3.49570	12.3
1-9	1.02756	-5.17742	11.3
EM7-2-1	0.63959	-4.34027	9.5
2-2	0.67920	-3.63981	9.0
2-3	0.67401	-4.74964	8.1
2-4	0.67272	-4.21737	8.5
2-5	0.68107	-4.83957	8.3
2-6	0.69002	-5.87475	8.5
2-7	0.71028	-6.71251	7.7
2-8	0.73483	-7.20287	7.4
2-9	0.66987	-9.99772	9.8
EM7-3-1	0.52500	-3.84747	7.7
3-2	0.51763	-3.95779	8.1
3-3	0.50614	-6.19005	9.8
3-4	0.52163	-6.13734	8.2
3-5	0.54194	-7.45209	8.6

Table 4.9 (cont.)

Photograph	Scale (K)	Temperature Factor (B)	R Factor (%)
3-6	0.52339	-9.67170	10.0
3-7	0.58435	-7.49953	9.7
3-8	0.55438	-10.30268	10.3
3-9	0.60911	-12.63696	8.5
EM7-4-1	0.67903	-3.30114	9.9
4-2	0.64017	-4.85816	11.6
4-3	0.65319	-4.35972	8.7
4-4	0.60541	-6.40608	11.3
4-5	0.70721	-4.63594	10.3
4-6	0.73511	-9.44660	10.5
4-7	0.53669	-10.25794	12.3
4-8	0.58725	-8.66239	10.0
4-9	0.59628	-9.25549	10.7
4-10	0.61410	-10.16087	11.6
4-11	0.74449	-11.35510	11.5
EM7-5-1	1.12822	-1.89586	10.7
5-2	1.02883	-4.01678	9.6
5-3	1.05652	-3.80519	7.2
EM7-6-1	1.08215	-3.43493	7.4
6-2	1.22148	-7.99539	8.8
EM7-7-1	0.77632	-2.25768	8.1
7-2	1.04645	-5.40147	12.8

Table 4.9 (cont.)

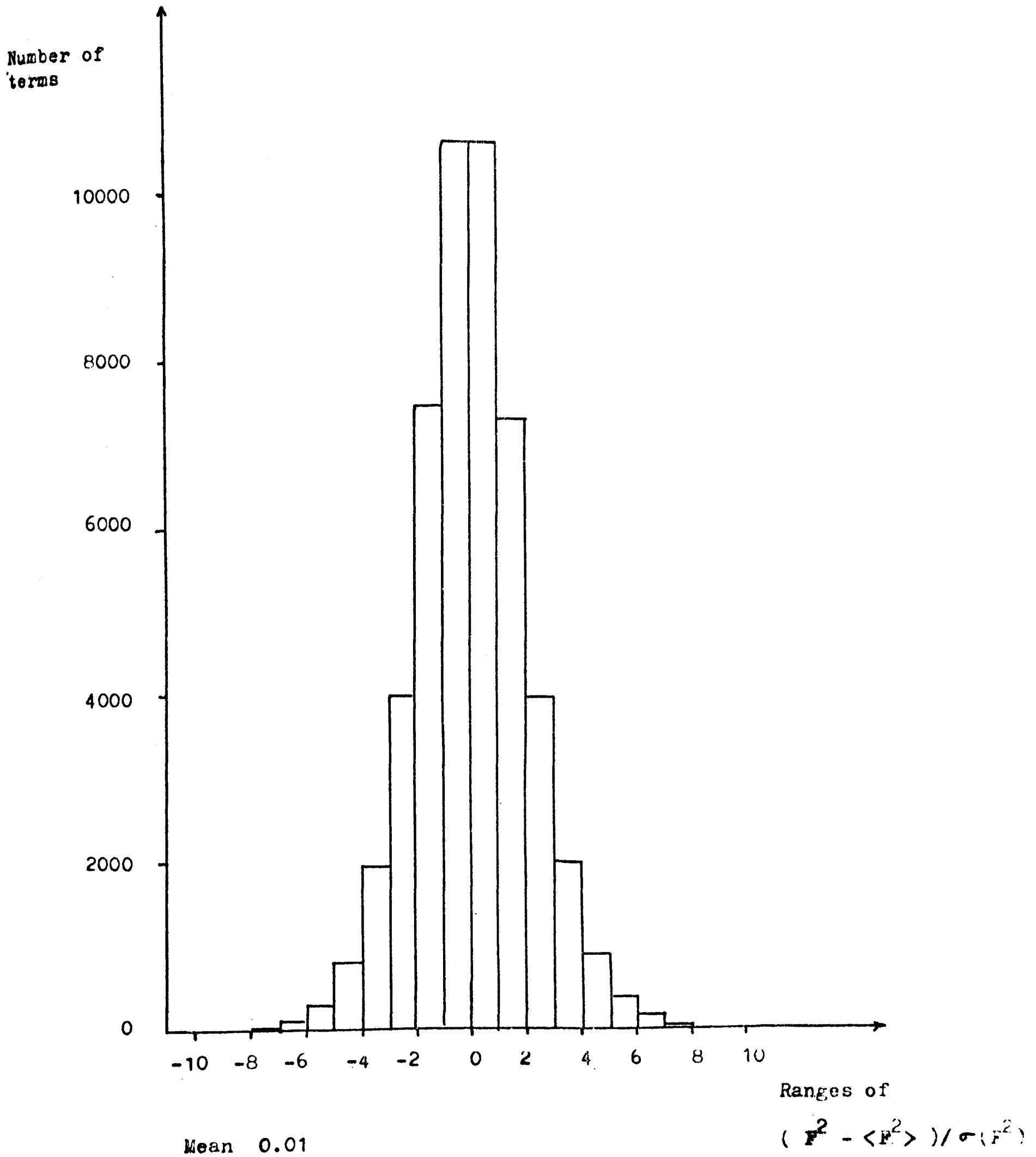
Photograph	Scale (K)	Temperature Factor (B)	R Factor (%)
EM7-9-1	0.93398	-4.80371	10.5
9-2	0.96432	-6.02724	9.1
9-3	1.21206	-4.71844	10.8

crystals. This emphasises the desirability of the present method of refinement of both temperature factor and scale. Within each crystal there is a general tendency for the temperature factor to increase (*i.e*: to large negative values) with photograph number. This is to be expected on account of radiation damage. However, this trend is not completely smooth and fluctuations in scale and temperature factor occur within crystals. This may be due to effects such as variation in X-ray flux during the exposure, variation in film developing conditions or in the accuracy of the film orientation matrix during integration. For a number of films (*i.e*: EM7-3-8, EM7-3-9, EM7-4-7, EM7-4-10,11) the B factor has a large negative value. This may be due to insufficiently accurate crystal orientation matrices for integration (Stura and Wilson, unpublished results). At this stage films which scaled badly (*i.e*: have a large number of rejected reflections) could have been removed, but since there were no very bad agreements all the EMTS7 data were kept. An R factor for each film (defined in Table 4.8) was calculated and is shown in Table 4.7, and the overall R factor for each derivative in Table 4.8. The standard deviation for each term is modified by equation 3.24 and the agreement between symmetry related observations is tested by plotting the distribution of $(F^2-\langle F^2\rangle)/\sigma(F^2)$ as shown in Figure 4.7.

The separate parts of partially recorded reflections are summed and these intensities analysed with respect to symmetry related terms which were fully recorded. The agreement between fully recorded and summed partially recorded terms is shown in Table 4.10. The intensities from the sum of partially recorded reflections appear overestimated compared to the intensities of the fully recorded terms. A similar phenomenon was observed by Irwin *et al.* (1976) who presented an analysis based on F rather than F^2. Wilson and Yeates (1978) point out that it is more

FIGURE 4.7

't-plot' for EMTS 7.



Mean 0.01

Standard deviation 1.02

Table 4.10

Fully versus partial analysis for EMTS7

Number of terms	Range of $4\sin^2\theta/\lambda^2$ Min.	Max.	$\langle F_F^2 \rangle$	$\langle F_P^2 \rangle$	% Difference
73	0.00000	0.00599	3030	2961	2.29
184	0.00599	0.01198	4088	4222	-3.23
339	0.01199	0.01798	3203	3294	-2.80
393	0.01798	0.02397	1954	2012	-2.93
380	0.02397	0.02997	1732	1820	-4.91
476	0.02997	0.03596	2100	2179	-3.69
497	0.03596	0.04195	2869	3058	-6.41
502	0.04195	0.04795	3599	3817	-5.86
516	0.04795	0.05394	3643	3867	-5.85
553	0.05394	0.05993	2753	3004	-8.36
522	0.05993	0.06593	2684	2925	-8.61
538	0.06593	0.07192	2315	2531	-8.93
541	0.07192	0.07791	2177	2377	-8.77
541	0.07791	0.08391	1643	1833	-10.73
574	0.08391	0.08990	1256	1363	-8.16
468	0.08990	0.09589	1204	1297	-7.43
496	0.09589	0.10189	964	1077	-11.11
455	0.10189	0.10788	1016	1073	-5.38
352	0.10788	0.11387	850	971	-13.29

rational to base the analysis on F^2, since the squared moduli are directly related to the experimental observations, *i.e*: the intensities. Wilson and Yeates discuss four possible causes for this apparent overestimation:-

i) Inaccuracy in the determination of the crystal orientation matrix. For terms flagged as fully recorded but which are in fact partially recorded, this will lead to an underestimate of the 'fully' recorded terms, and hence an apparent overestimate of the partially recorded ones.

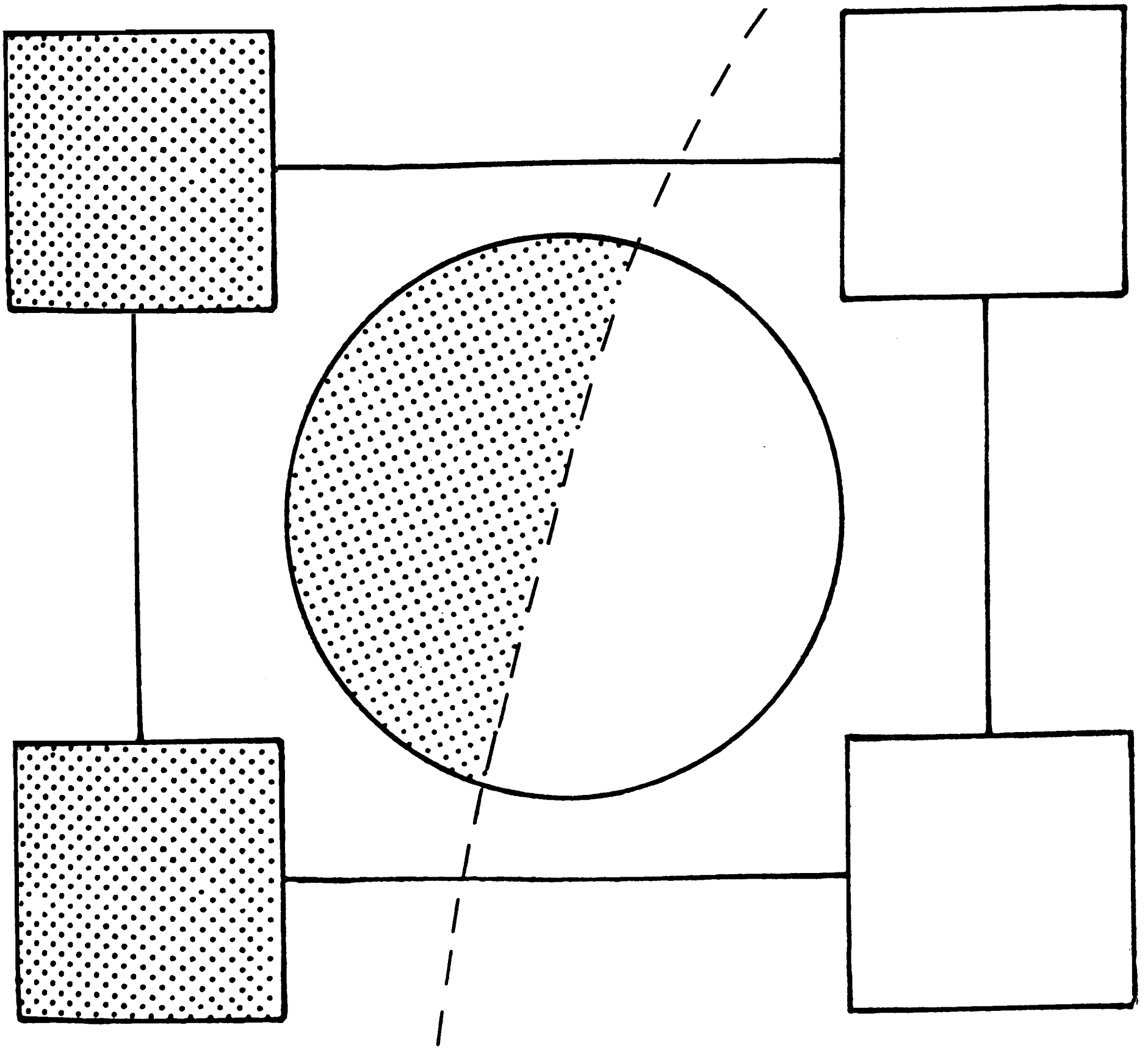
ii) Underestimation of the crystal reflecting range γ. This will lead to inaccurate prediction of fully and partially recorded terms in a given oscillation range and will have a similar effect to (i).

iii) The use of Ni filtered CuKα rather than monochromatic radiation, which may give a higher background within the lunes of data.

iv) The effect of the type of integration box used for partially recorded reflections. Figure 4.8 shows the box for a term partially recorded by 50%. The dashed line represents the edge of a lune. Half the box and two background areas lie within the reflection and half the box and two background areas lie outside. Taken together the four backgrounds are expected to give a reasonable approximation to the true background. This will obviously not be the case, but Wilson and Yeates point out that such error should not systematically affect the integrated intensities if a statistically significant number of terms is used.

Effect (iii) is discounted by Wilson and Yeates. From Table

FIGURE 4.8



The integration box for a term partially recorded by 50%. The dashed line represents the edge of the 'lune'. The dotted half of the term is recorded and lies inside the lune as do the two dotted backgrounds. The remaining two backgrounds and the 'unrecorded' half of the spot lie outside the lune. From Wilson and Yeates (1979).

4.10 it can be seen that the degree of apparent overestimation increases with $4\sin^2\theta/\lambda^2$. This is consistent with explanations (i) and (ii) above since both are likely to affect high angle more than low-angle data. These two effects have been investigated by Stura and Wilson (unpublished results) who have shown that improved crystal orientation matrices and estimates of γ do produce a reduction in the degree of overestimation of the partially recorded reflections. The effect of γ is shown in Table 4.11 for native c* data to 3 Å resolution, where the data were reintegrated with two values of γ, both higher than the original estimate. Table 4.12 shows the effect of improved crystal orientation matrices on the degree of overestimation of partially recorded reflections. The data presented are from Pt crystal 4. Photographs 1-7 were reintegrated with improved orientation matrices. Photographs 9-19 were not reprocessed. The reduction in the degree of overestimation is evident from this table (E.A. Stura, unpublished results).

After scaling and merging, the data were truncated in the manner described in Chapter 3, Section 9, using the programme TRUNCATE.

3) REPROCESSING OF PREVIOUS DATA

This work was carried out in direct collaboration with L.N. Johnson, E.A. Stura and K.S. Wilson. My major contribution was in processing the data from the SSM stage onwards.

The data sets used in the calculation of the new 3 Å map comprised the following:-

i) Reprocessed native c* data to 3 Å resolution and a weak 6 Å set to fill in strong terms

ii) Reprocessed EMTS data to 3 Å resolution

iii) New EMTS7 data to 3 Å resolution

Table 4.11

Overestimation of partially recorded reflections as a function of γ:-

Results from SSM for Edmonton c* data

a) Summary

	old data	low γ	high γ
Total number reflections	70,334	64,878	62,166
Unique reflections	17,994	17,759	17,553
Rejected reflections	326	95	101
Largest % overestimation of partials	-10.24%	-9.43%	-8.09%
Merging R	9.5%	8.75%	8.87%

b) % Overestimation of partially recorded reflections as function of $4\sin^2\theta/\lambda^2$

Range of $4\sin^2\theta/\lambda^2$	No. reflections	old γ (0.25)	% overestimation low γ*	% overestimation high γ*
1	63	-4.67	-0.88	-0.09
2	160	-1.06	0.31	-0.19
3	203	-0.91	0.56	1.97
4	254	-2.68	-1.81	-2.62
5	293	-1.84	-0.90	-0.55
6	313	-5.26	-3.40	-3.42
7	340	-4.76	-1.93	-2.37
8	363	-5.16	-3.58	-3.81
9	380	-5.67	-3.81	-4.06
10	402	-7.44	-4.49	-3.96
11	392	-6.79	-4.37	-3.72

Table 4.11 (cont.)

Range of $4\sin^2\theta/\lambda^2$	No. reflections	old γ (0.25)	low γ*	high γ*
12	418	-9.18	-6.84	-6.17
13	420	-7.28	-5.24	-5.39
14	453	-9.44	-7.14	-7.71
15	434	-9.41	-7.62	-6.41
16	432	-10.17	-5.53	-7.13
17	383	-10.24	-5.34	-4.71
18	412	-6.25	-6.22	-4.94
19	317	-8.60	-7.00	-7.62
20	316	-8.06	-9.43	-8.69

The data were reintegrated with a low and high value of γ which varied according to the data set:-

Data set	low γ	high γ
1	0.32	0.37
2	0.39	0.43
3	0.39	0.43
4	0.35	0.38
5	0.32	0.37
6	0.32	0.37
7	0.32	0.37

iv) Reprocessed PT data to 3 Å resolution (crystals 1 and 2)

v) Bakers dimercurial data to 4 Å resolution.

The Bakers dimercurial data were collected and processed by J.A. Jenkins. The native, EMTS and PT data were reprocessed in the following manner:-

i) The photographs were reintegrated after obtaining better plots from Segment 2 of OSCAR, using the methods described in Chapter 3 to improve the crystal orientation matrices and estimates of γ. The effect of a better estimate of γ is shown for the native c* data in Table 4.11, and the effect of improved orientation matrices and better plots is shown for the PT data in Table 4.12.

ii) Both the scale factor and temperature factor were refined in SSM, as described in Section 2.

iii) For the PT data set, crystals 1 and 2 (collected in Edmonton) and crystal 4 (collected in Oxford) were treated as separate crystals. After merging, scaling and truncation, crystals 1 and 2 were refined together, crystal 4 was refined separately and crystals 1,2 and 4 refined together. No successful refinement was obtained with crystal 4 alone or with crystals 1,2 and 4 combined. The data from crystal 4 were, therefore, discarded.

4) SCALING TO NATIVE

All the derivative data were scaled to the native data using the programme ANSC as described in Chapter 3, Section 10, applying the technique of 'local' scaling (Matthews and Czerwinski, 1975). After analysis of any systematic trends which occurred, corrections were applied to the scale factor in blocks of reciprocal space defined by levels of ξ^2 and ϕ ranges (see Chapter 3, Section 10).

Table 4.12

Effect of improved orientation matrices on overestimation of partially recorded reflections: Results from SSM

Pt data from crystal 4

Photographs 1-7 reprocessed by E.A. Stura

Photographs 9-19 as before.

Range of $4\sin^2\theta/\lambda^2$	% overestimation	
	old data	Reprocessed data
1	4.96	4.77
2	-4.98	-0.84
3	-6.45	-7.69
4	-3.25	-4.85
5	-8.83	1.23
6	-5.86	-7.97
7	-5.09	-2.88
8	-12.61	-5.30
9	-15.86	-12.77
10	-27.02	-7.70
11	-17.48	-5.40
12	-32.16	-6.58
13	-7.86	-11.22
14	-16.25	-14.26
15	-18.86	-16.50
16	-21.71	-15.88
17	-15.66	-14.20
18	-13.54	-14.65
19	-22.93	-4.51
20	-31.06	-10.33

Some areas of reciprocal space still showed unusually high or low scale factors after correction, and these trends were removed empirically by applying individual scale factors to data in these regions of reciprocal space. Figure 4.6 shows the scale factors as a function of h for the EMTS7 derivative, before and after local scaling, and the improvement is evident. For all the derivatives the final analysis was reasonably flat.

5) REFINEMENT OF HEAVY ATOM POSITIONS

The scale factor used in F_{HLE} refinement was the Kraut scale factor, which was calculated by the programme FHLEKR as described in Chapter 1, Section 3d, using the bias correction of Dodson *et al.* (1975). The Kraut scale was then kept constant during subsequent refinement. The values used are shown in Table 4.8. F_{HLE} refinement, using bias corrected F_{HLE} values (Chapter 1, Sections 3c and 3e) was then carried out for all the derivatives by the programme REFINE with option FHLESC. The initial refinement used only the centric terms and provided approximate values for the heavy atom parameters (occupancy, positional parameters (x,y,z) and temperature factor (B)). These parameters were then used as initial values for an acentric refinement using all the data for each derivative, as described in Chapter 1, Section 3. Terms with large values of F_{HLE} (greater than 400) were rejected, since these were assumed to be spurious. Terms with F_{HUE} less than 260 were also rejected, corresponding to the possibility of "cross-over" in the centric case. The quality of the refinement was monitored by the value of the R factor (equation 1.13b), the error ratio F_{calc}/E (where E is defined by equation 1.13a) and by the calculation of a double difference Fourier map with coefficients

$$(F_{HLE}-F_{calc})\exp i\alpha_{calc}$$

after refinement. Figure 4.9 shows sections of the difference Fourier for EMTS7 after refinement.

For each derivative the protein phases were calculated using the other derivatives, and used to produce a difference Fourier synthesis with coefficients

$$(|\underline{F}_{PH2}|-|\underline{F}_{P}|)\exp i\alpha_{P1}$$

where F_{PH2} is the structure factor for the derivative under investigation and α_{P1} is the protein phase determined from the other derivatives. These difference maps were checked for minor sites. Finally, after calculating a set of protein phases using all derivatives, double difference Fourier syntheses with coefficients

$$(|\underline{F}_{PH}|-|\underline{F}_{P}+\underline{f}_{H}|)\exp i\alpha_{P}$$

were calculated.

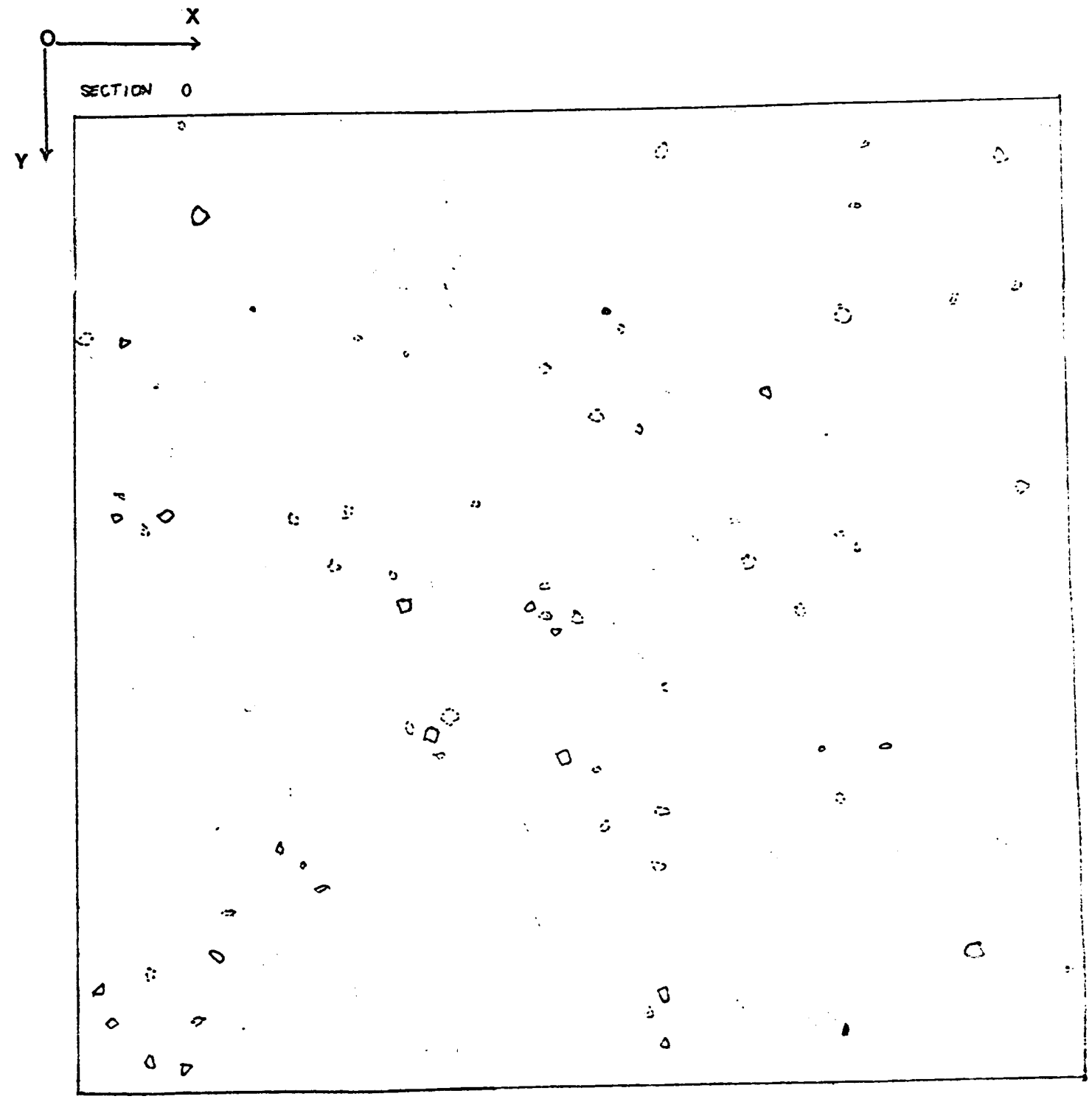
The first set of refinements are summarised in Table 4.13. For EMTS7 it was necessary to fix the temperature factor (B) for site HG3. The occupancies for all the sites of EMTS were higher than for EMTS7, with the exception of HG3, which emphasises the difficulty of obtaining reproducible behaviour from a longer soak. The first refinement for the PT12 data used a three site model as described in Weber (1977). After refinement some noise features were observed on the difference Fourier map around the site PT1. The data were subsequently refined in shells of reciprocals space and it was observed that the occupancies of site PT1 increased at high angle, whilst those for site PT2 decreased at high angle. An attempt was made to model this trend using modified form factors, and subsequent refinement gave an improved agreement between the observations and model. The final parameters of this refinement

FIGURE 4.9

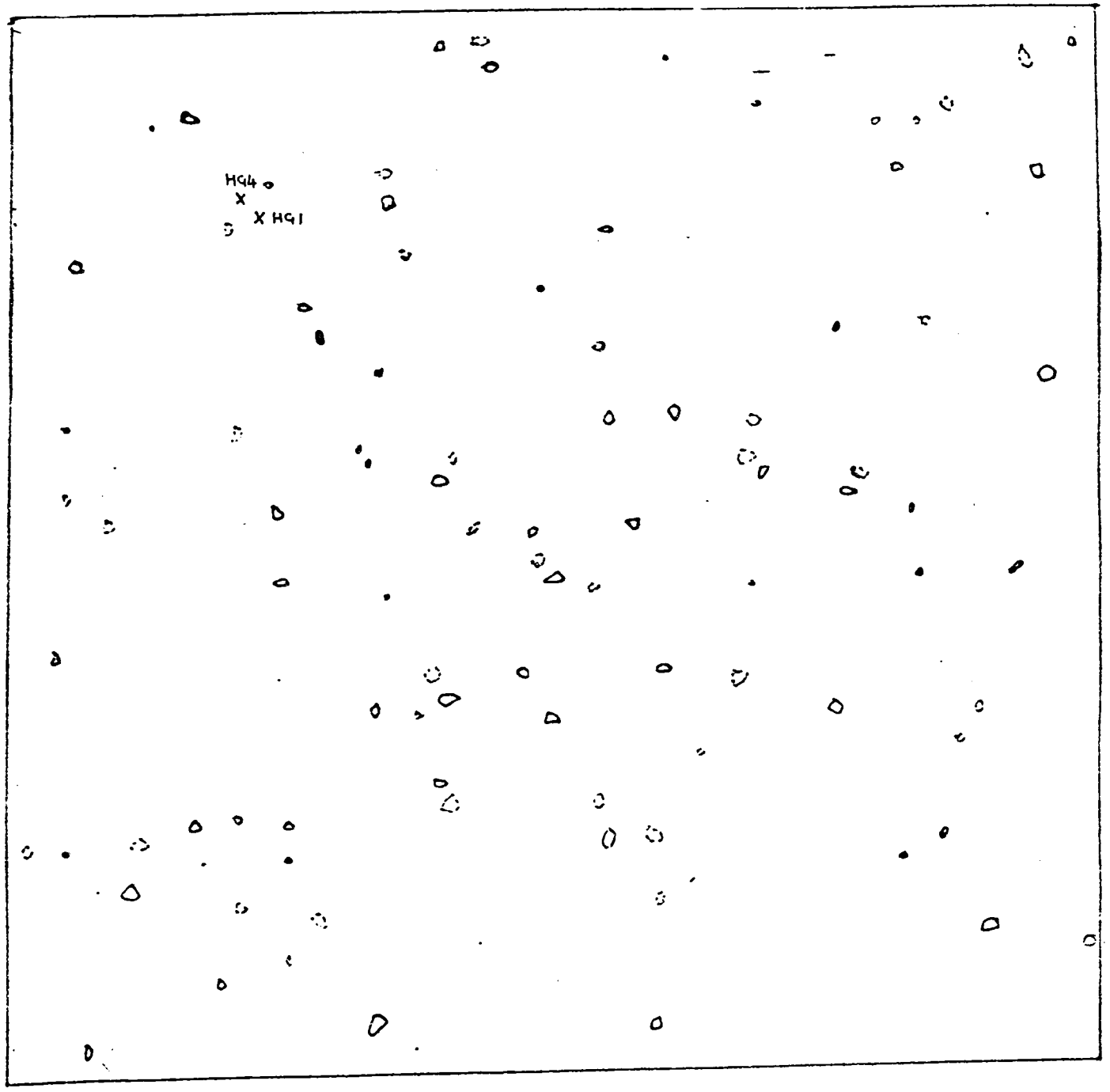
Sections of EMTS 7 Double Difference Fourier Synthesis after Refinement.

The following 8 pages show the sections of the EMTS 7 Double Difference Fourier synthesis after refinement. The equivalent positions in the unit cell are given in Table 2.1. The contour level is ± 200 arbitary units and negative contours are shown dotted. The expected positions of Hg binding sites are marked.

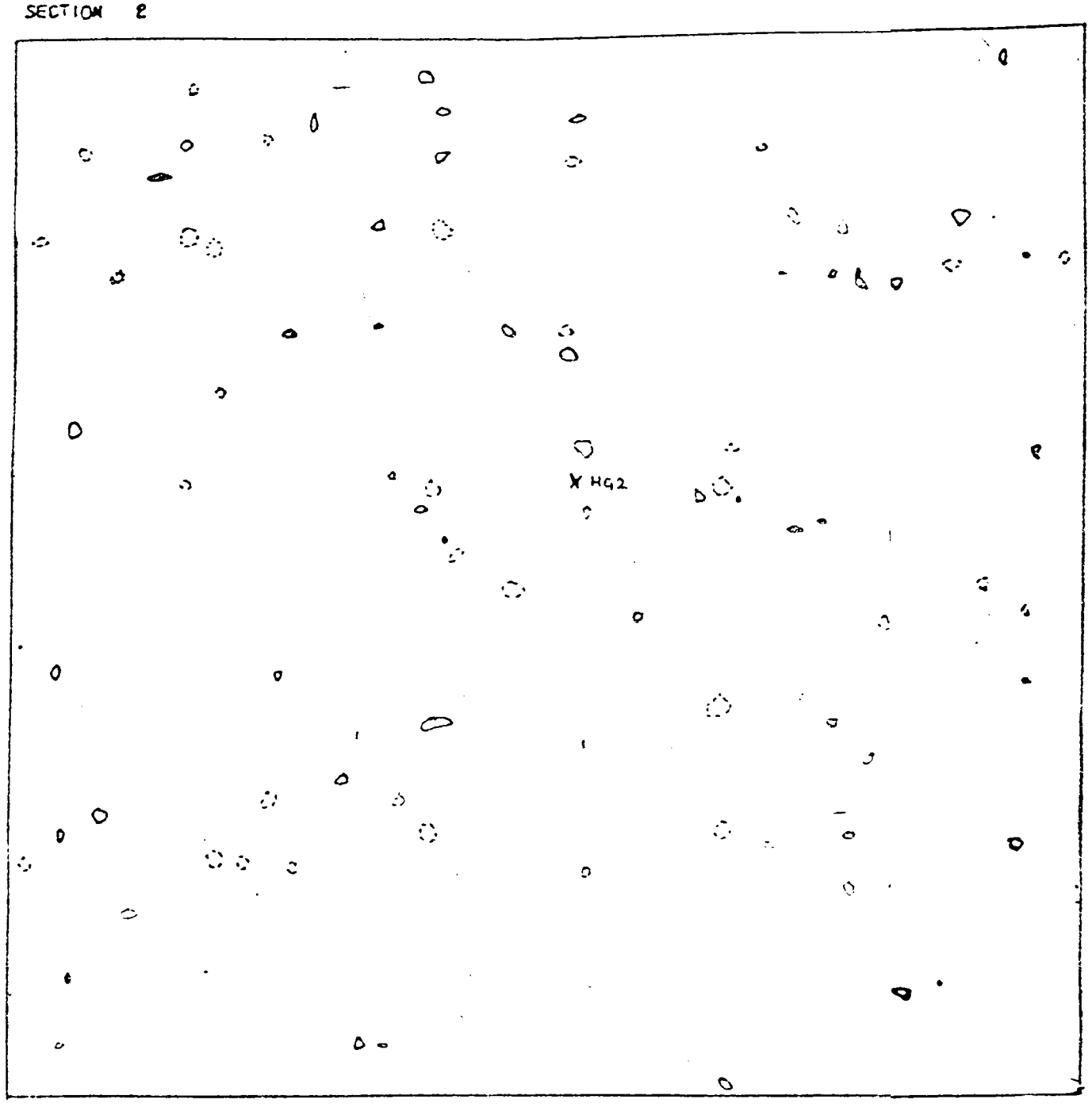
Sections z = 0 - 16/128 are shown.



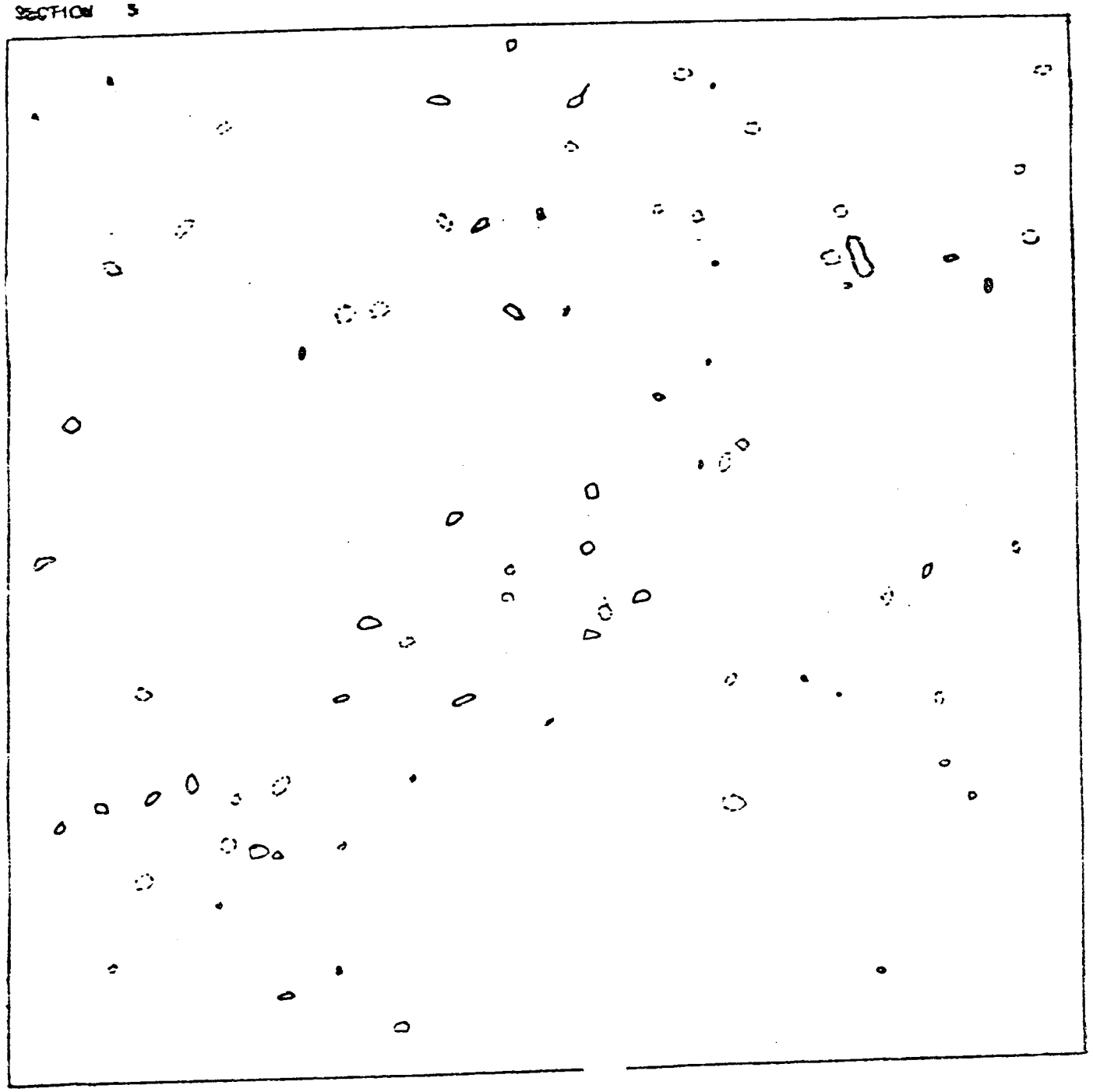
SECTION 1



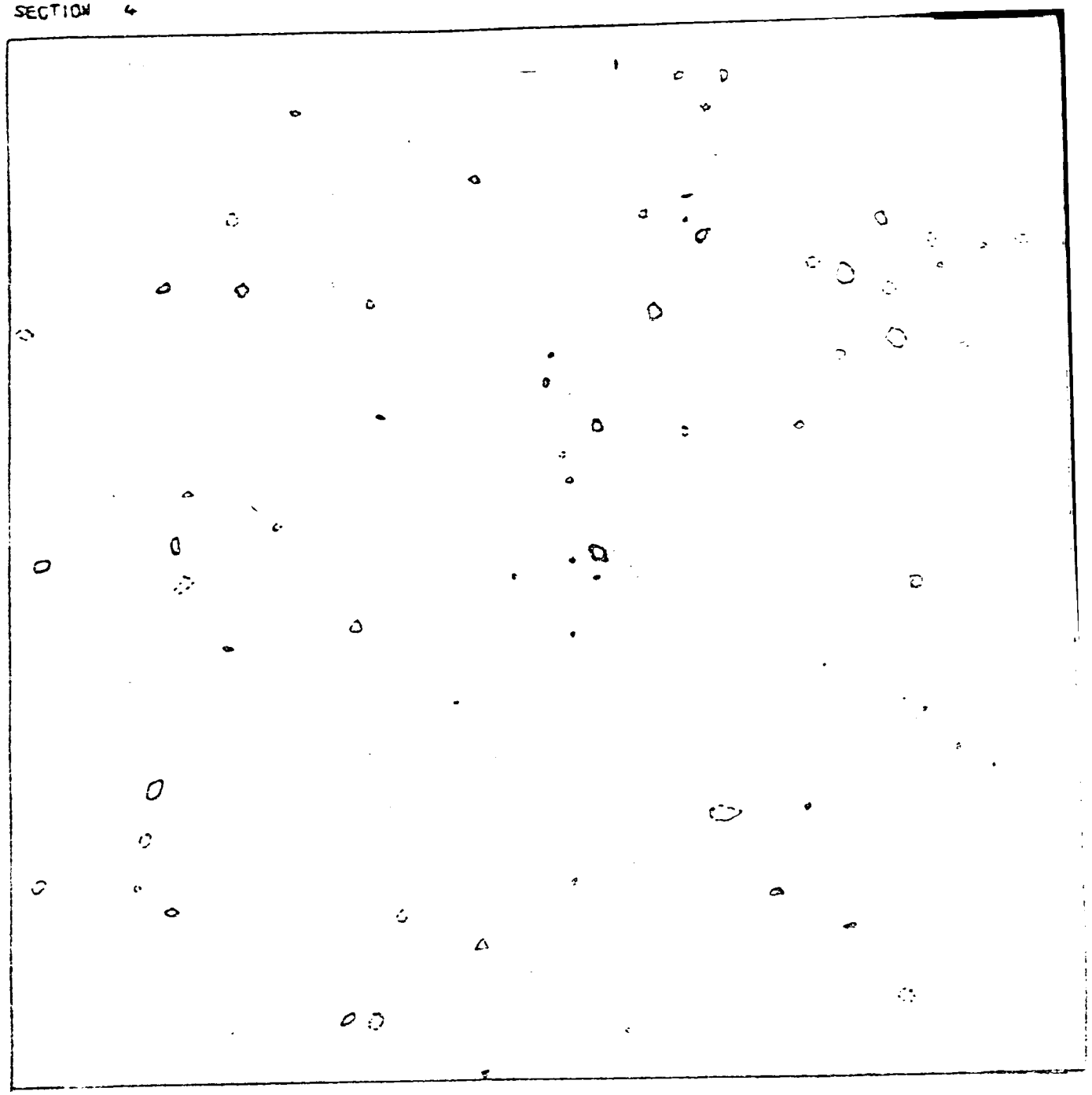
SECTION 2



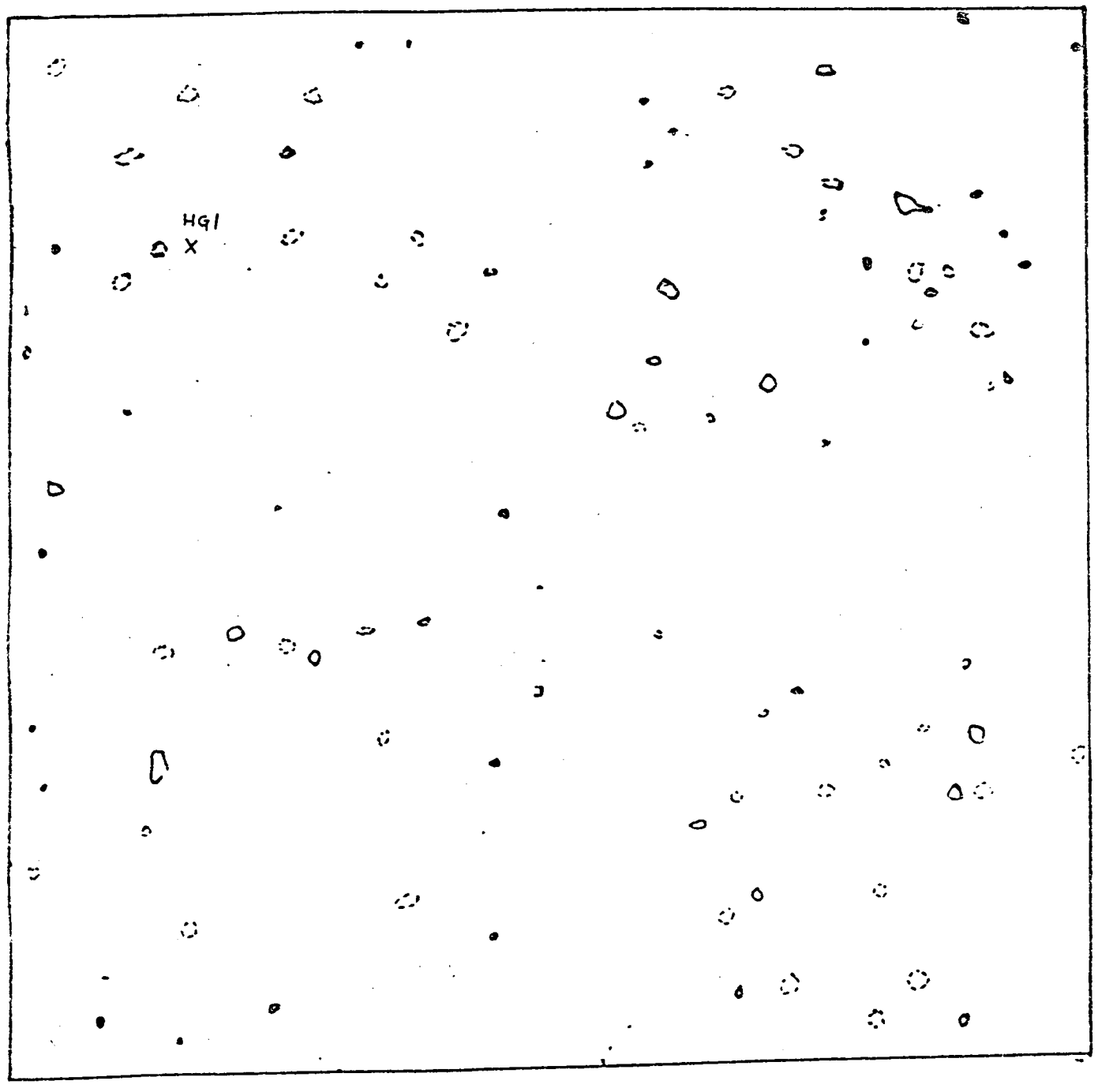
SECTION 3



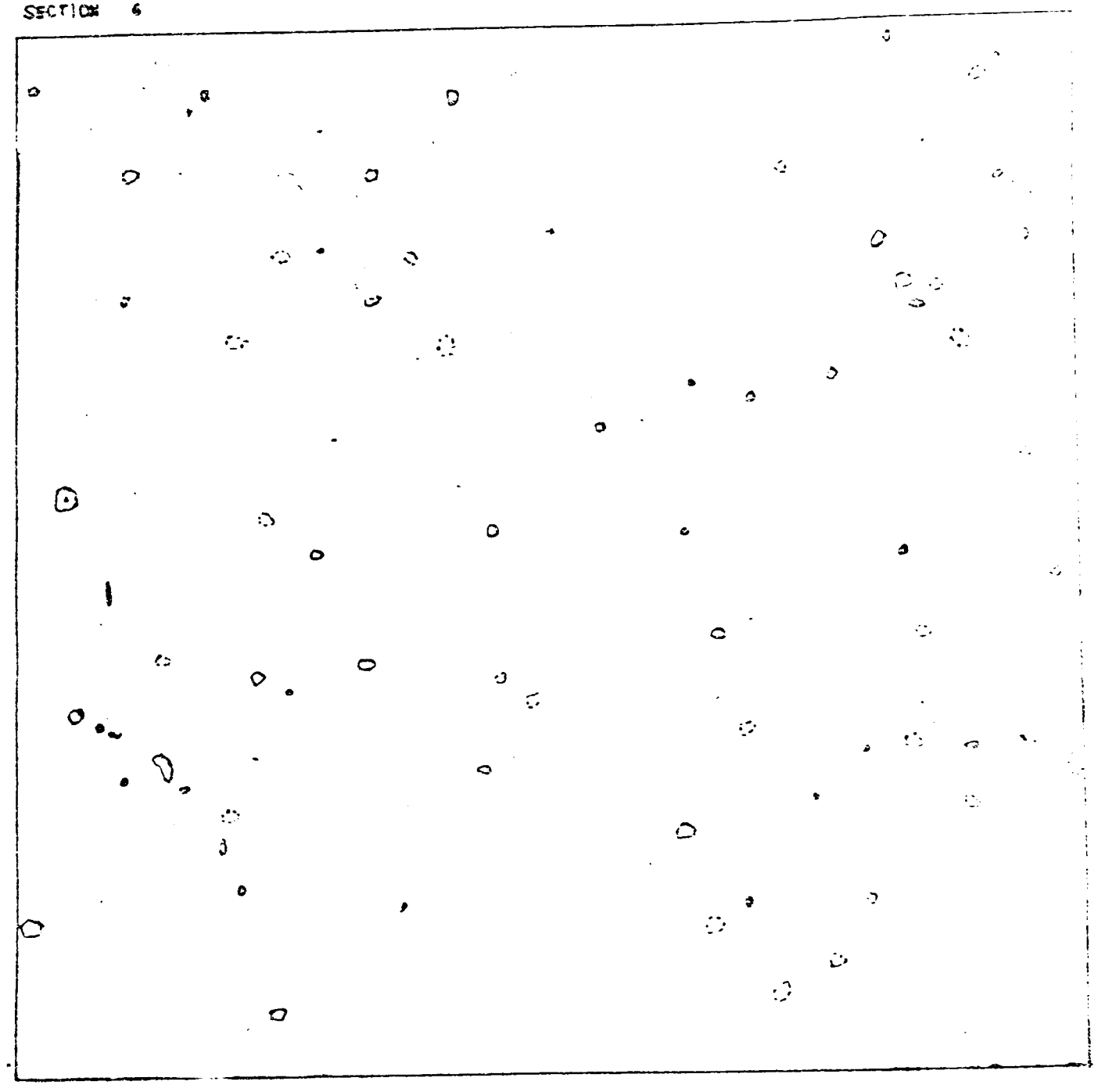
SECTION 4



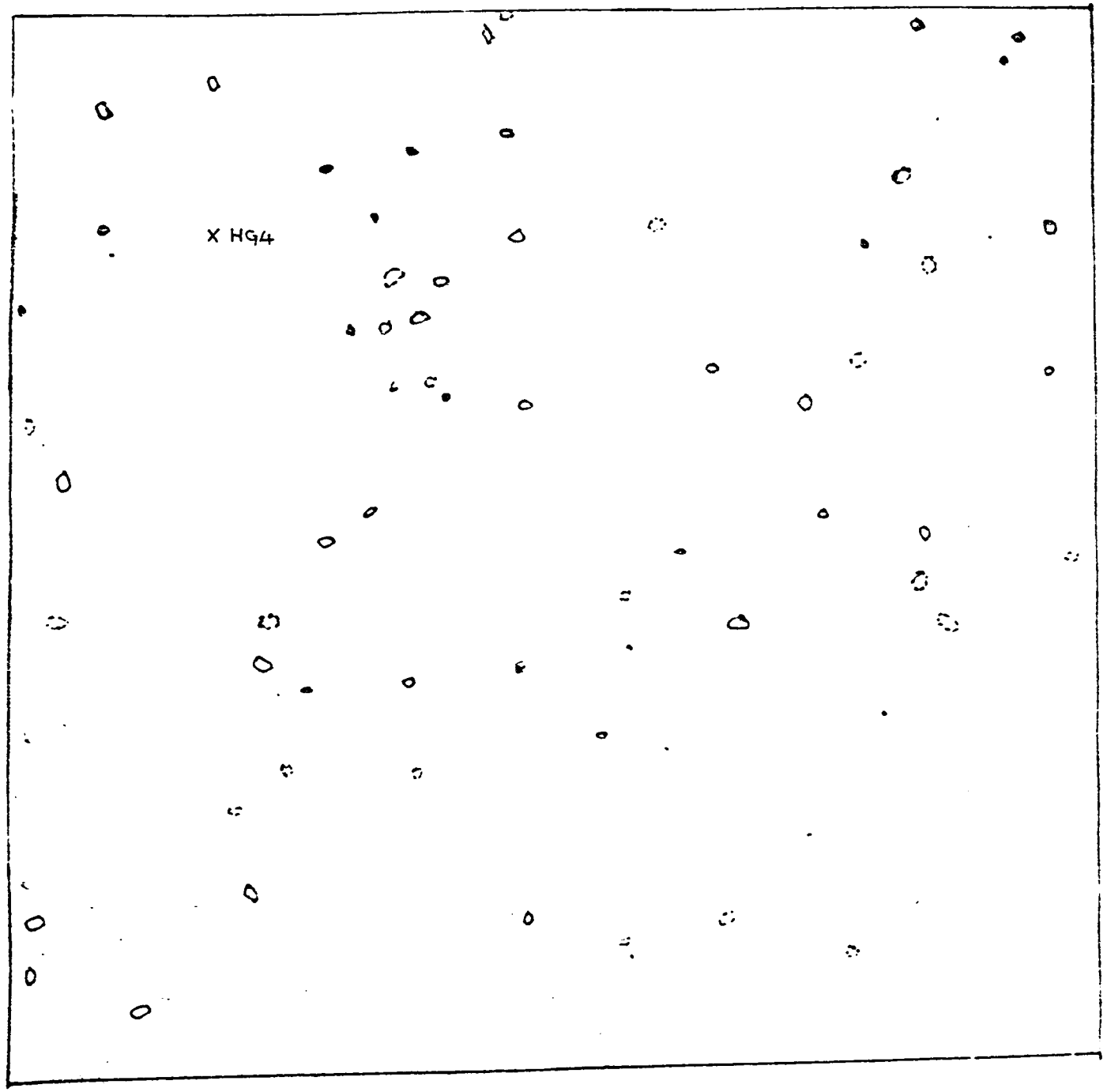
SECTION 5



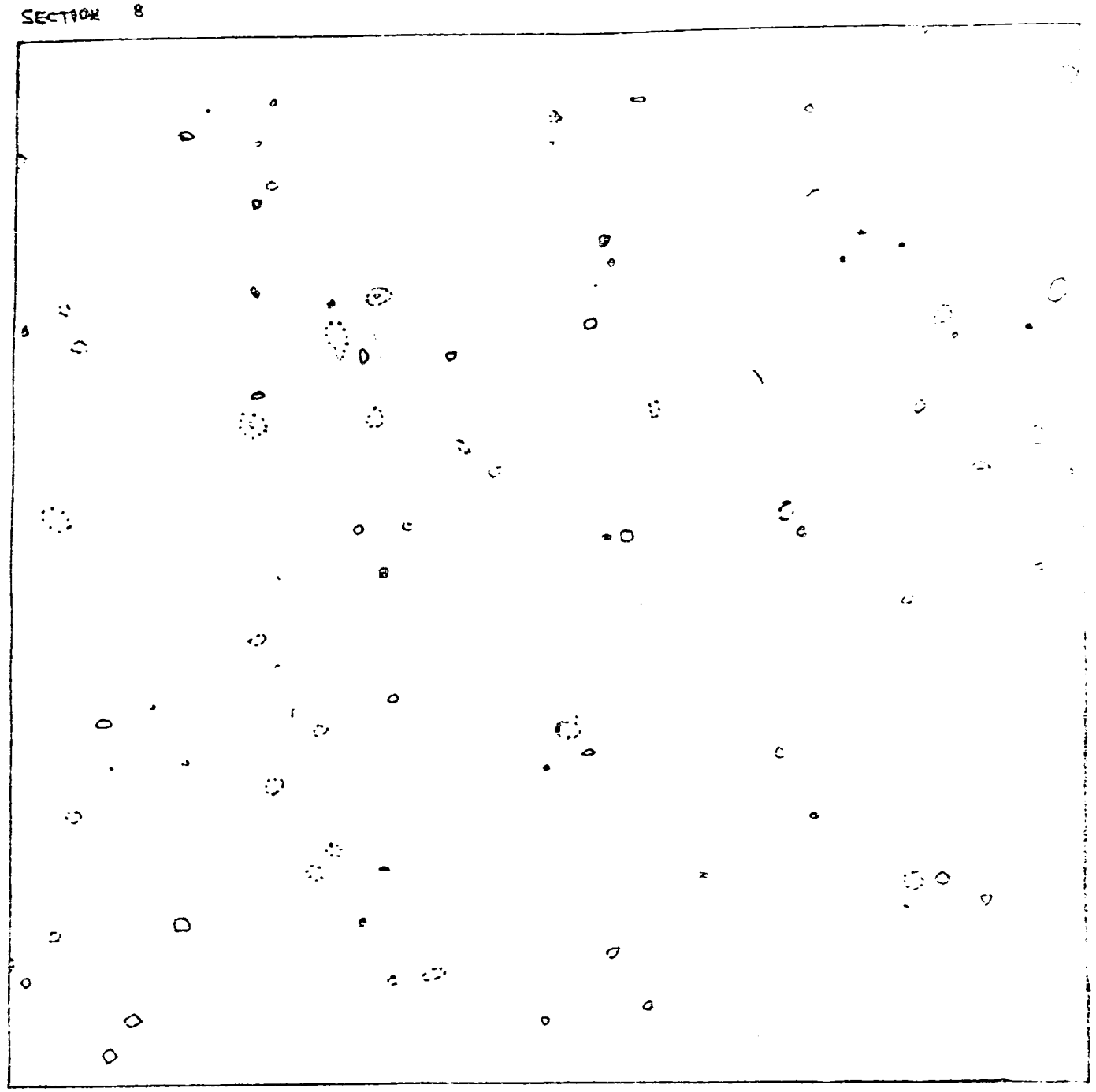
SECTION 6



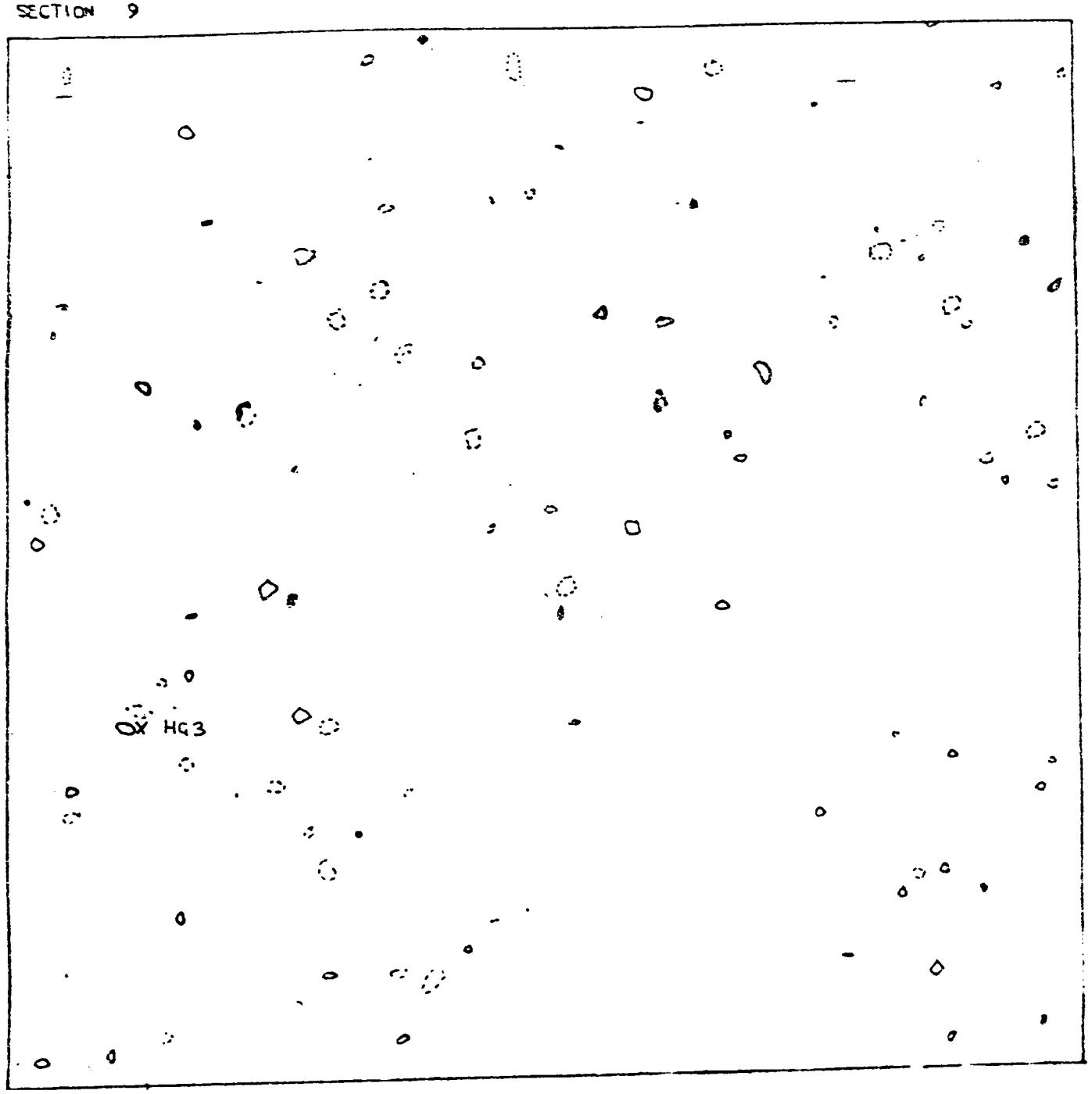
SECTION 7



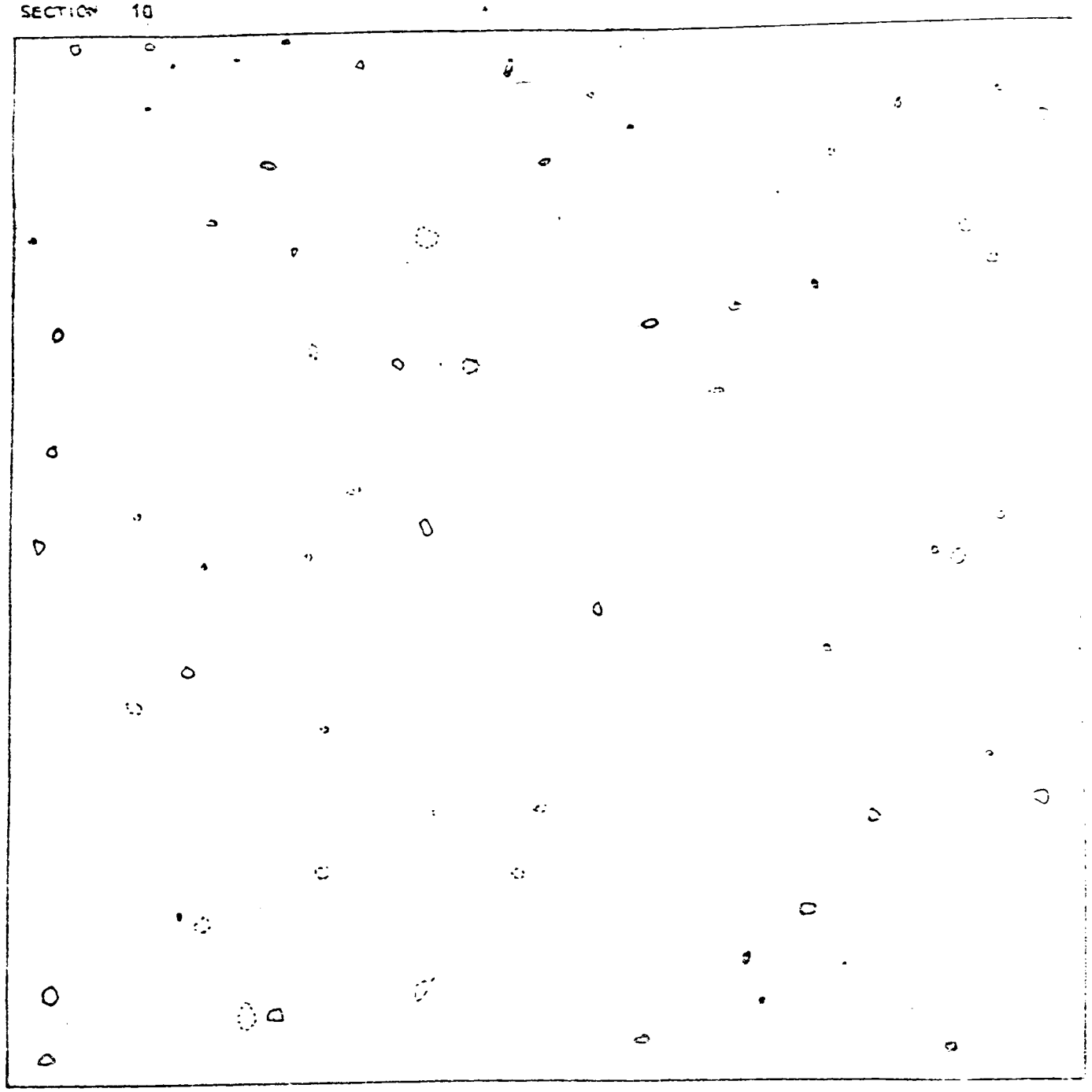
SECTION 8



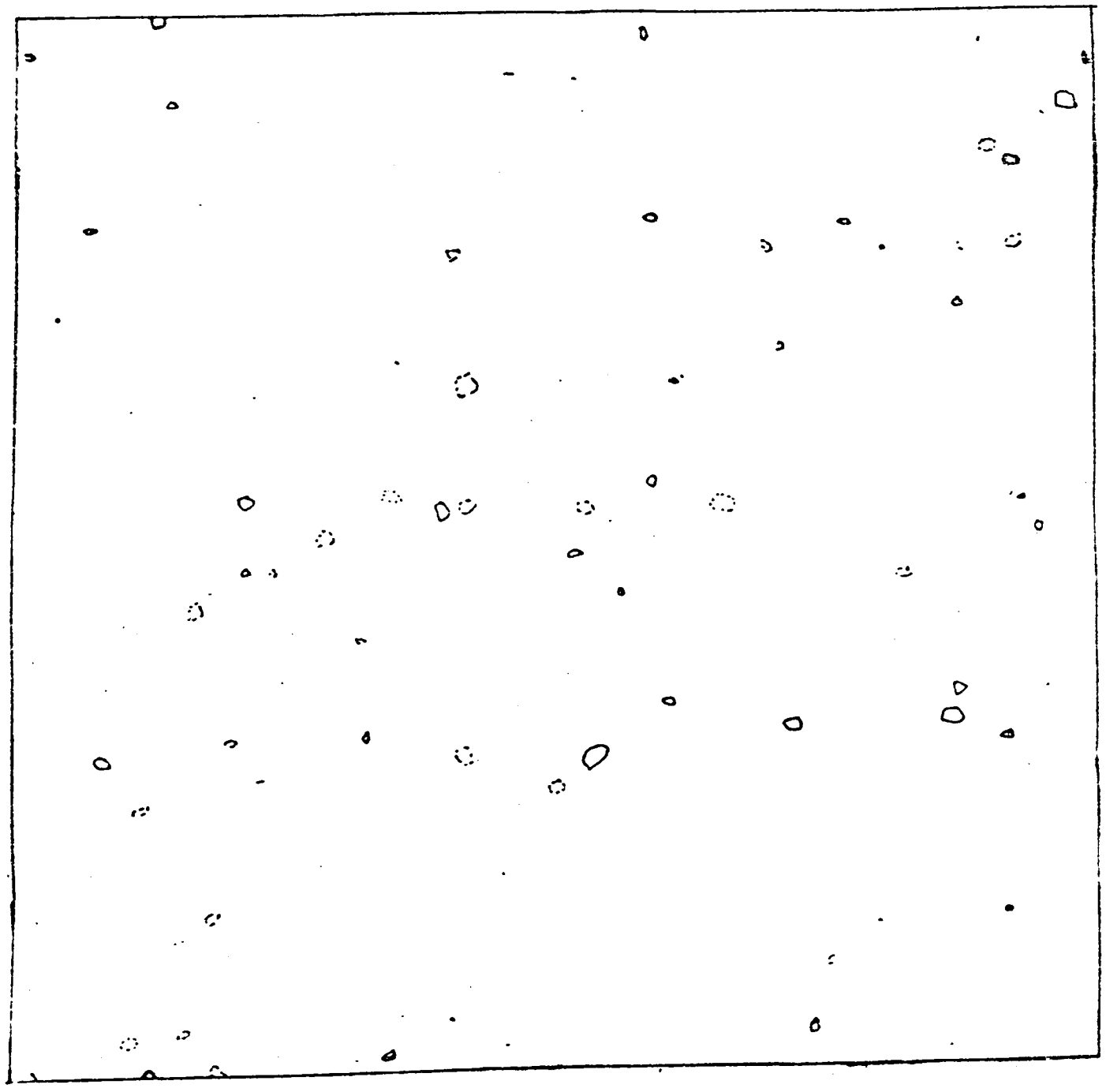
SECTION 9



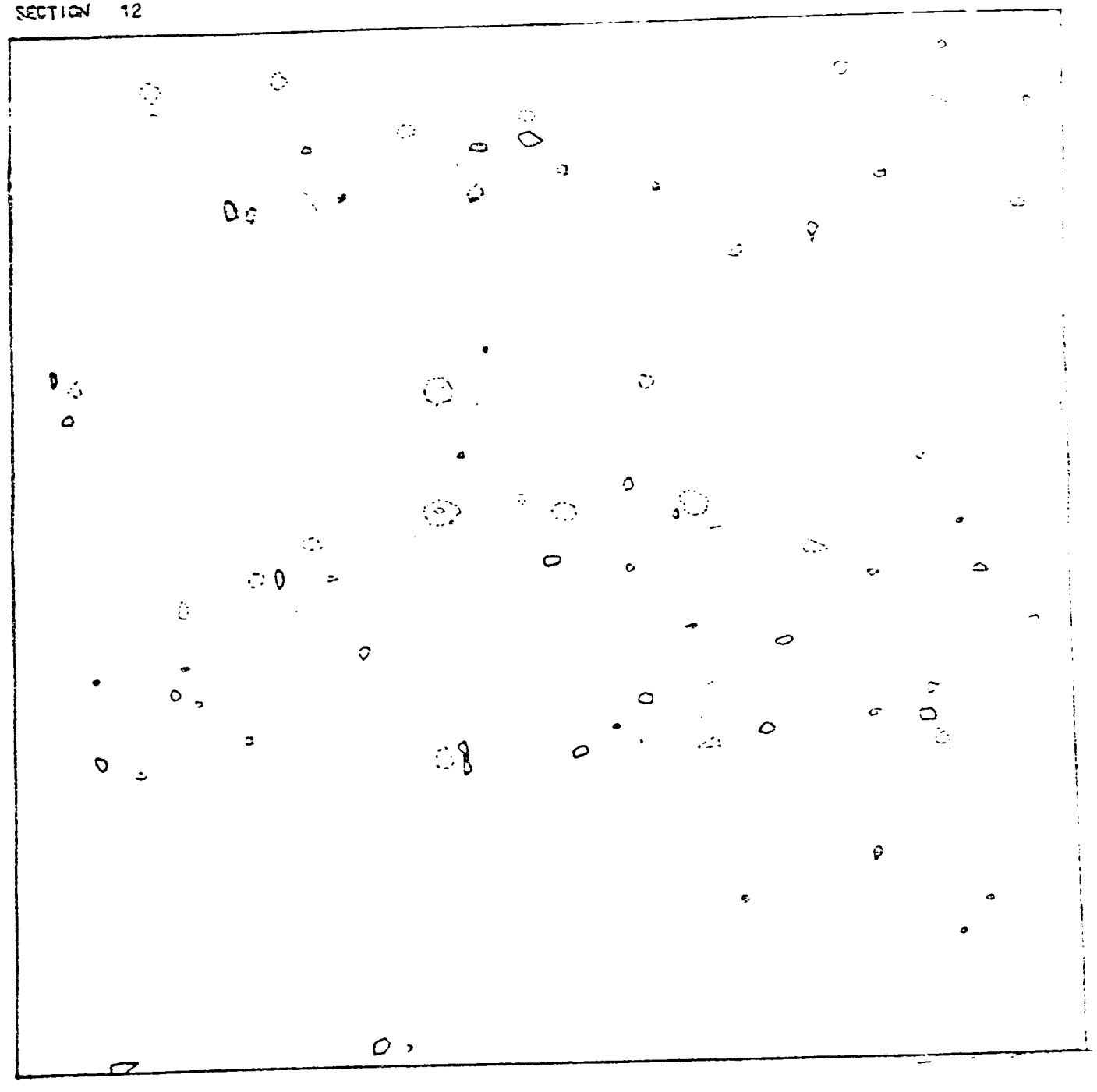
SECTION 10



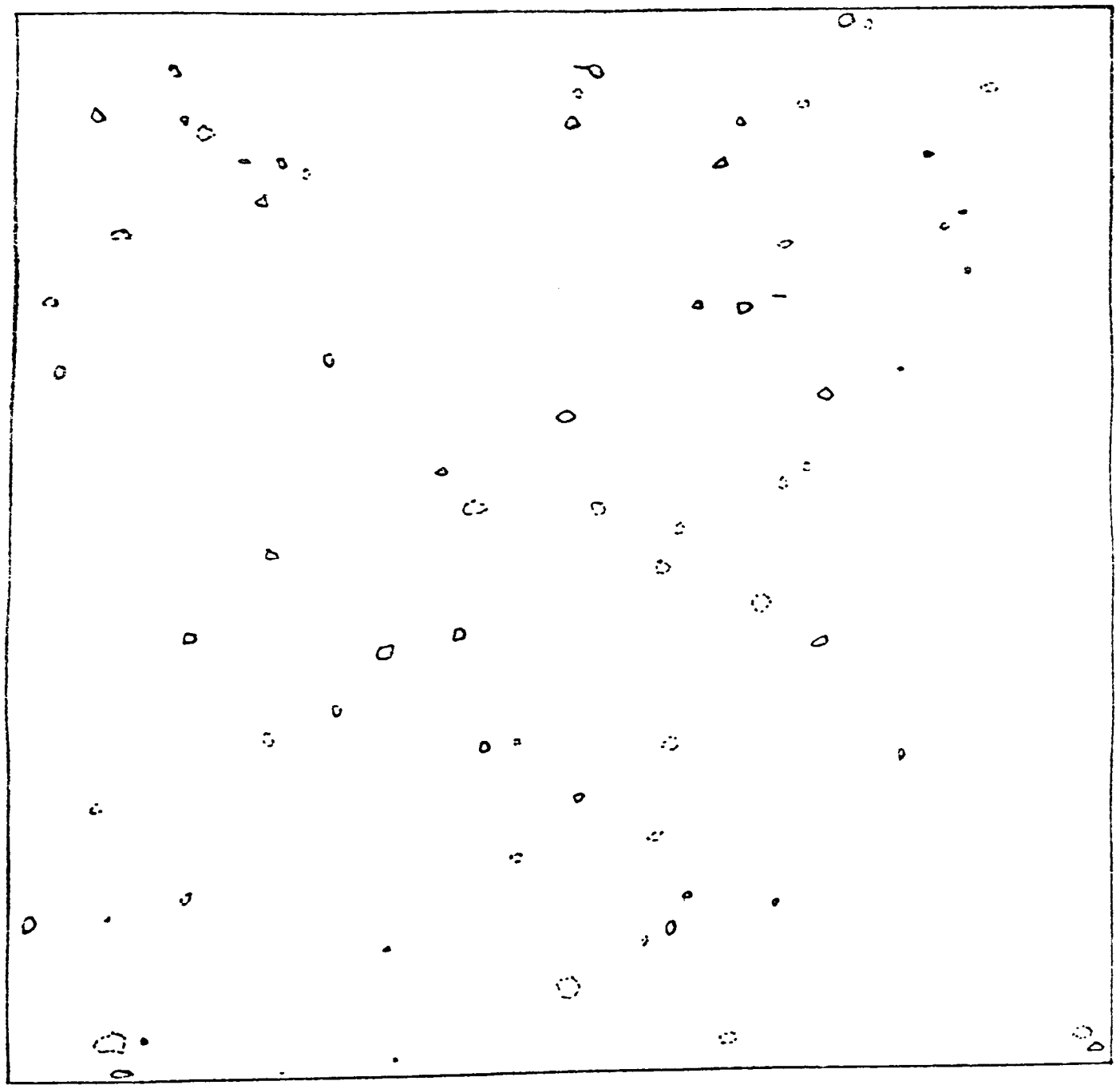
SECTION 11



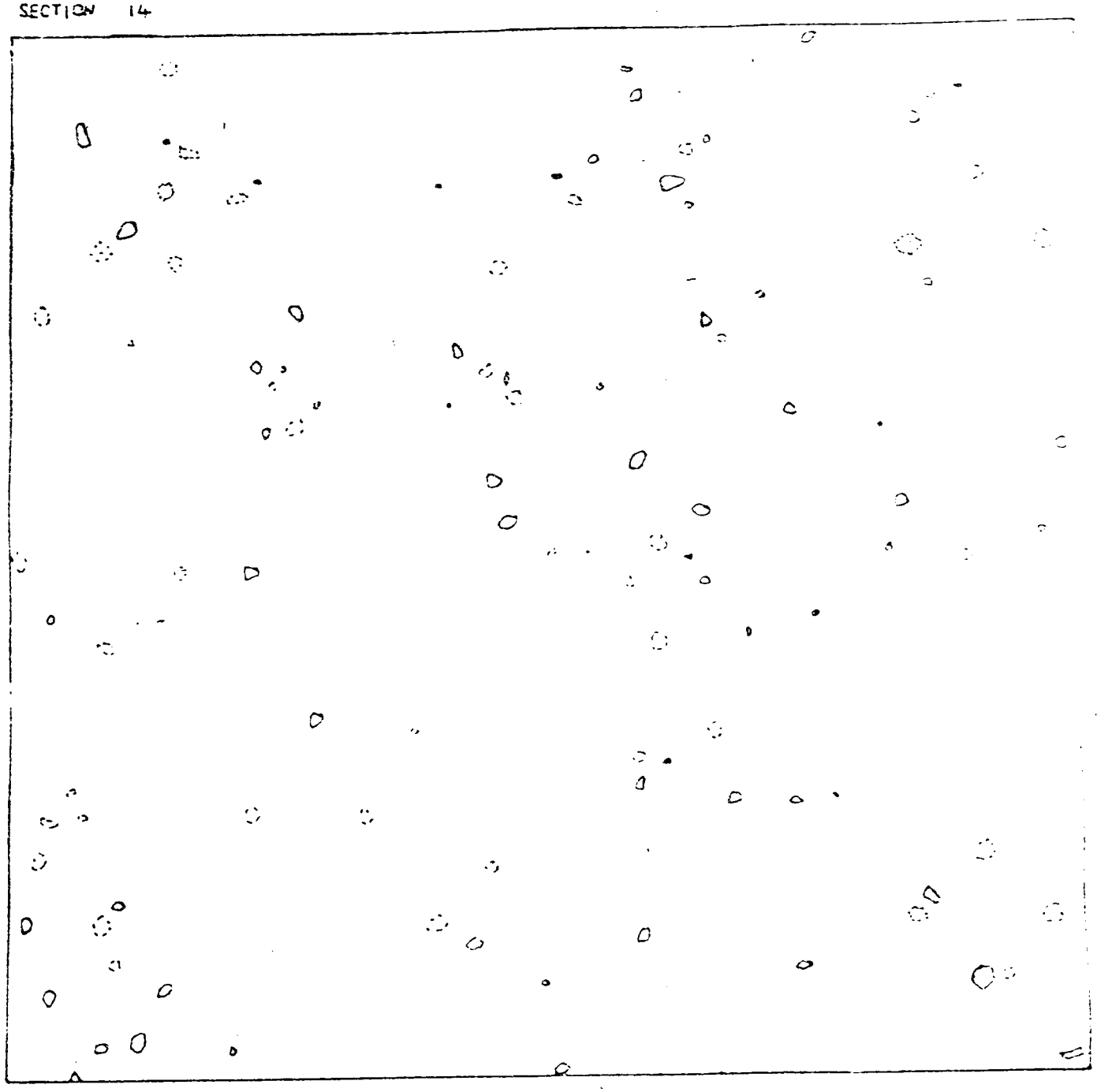
SECTION 12



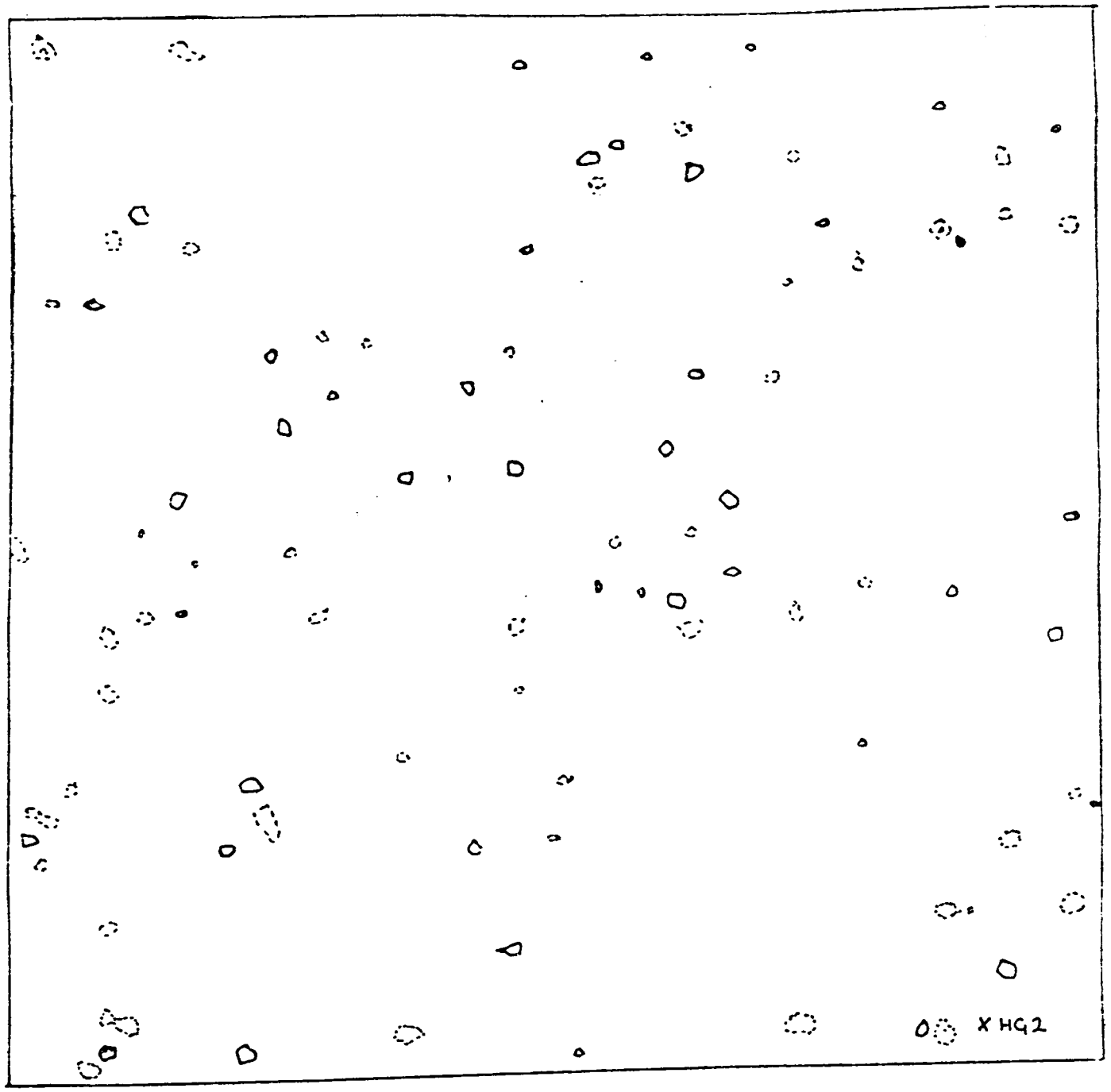
SECTION 13



SECTION 14



SECTION 15



SECTION 16

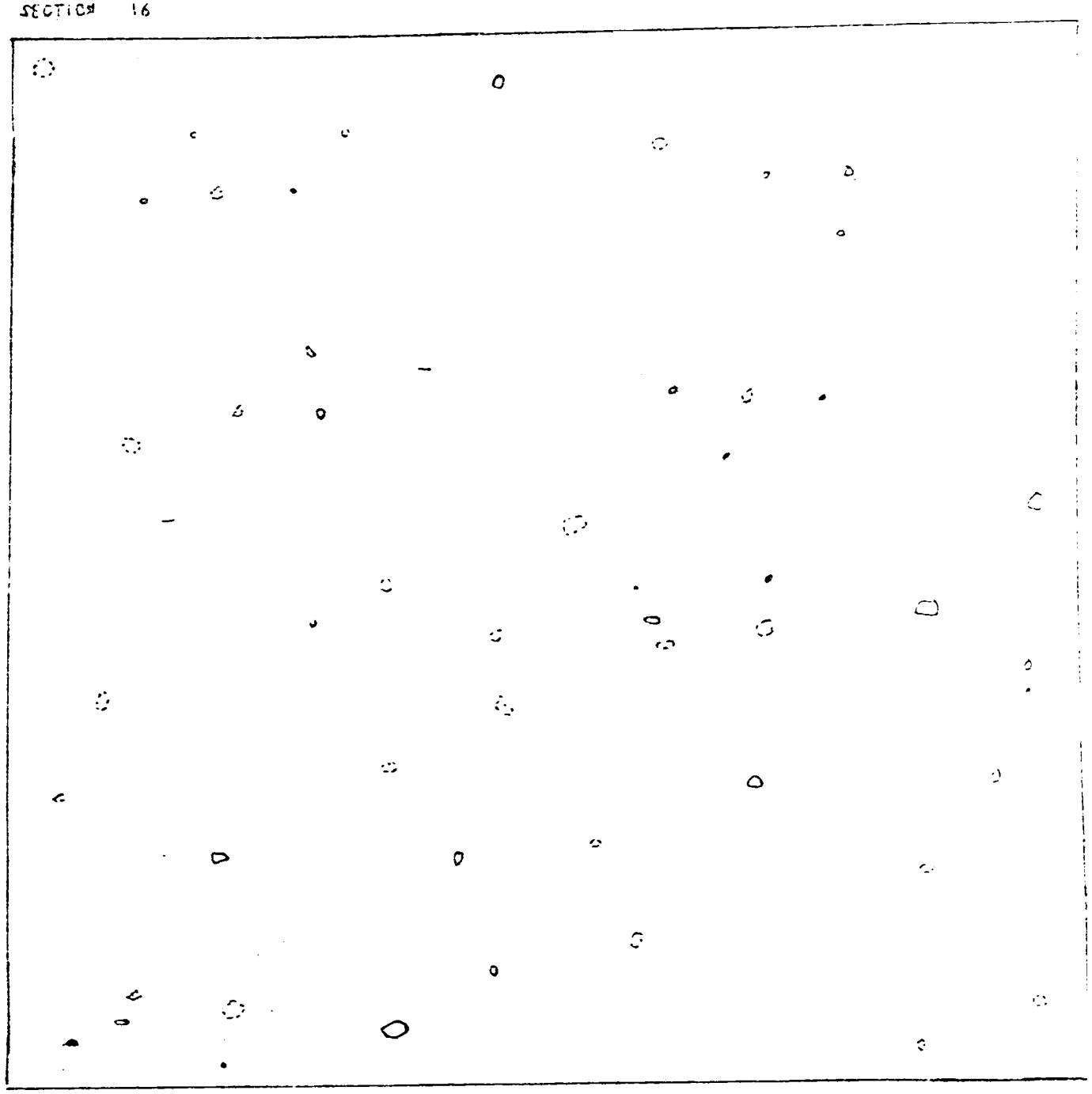


Table 4.13

Summary of first refinements

Site	Occupancy	x	y	z	B	centric R_c	centric F_c	centric E_c	centric N_{obs}	acentric R_a	acentric F_a	acentric E_a	acentric N_{obs}
EMTS													
1	0.5754	0.1948	0.2240	0.0393	28.3								
2	0.7654	0.0286	0.0887	0.3839	15.9								
3	0.1707	0.1600	0.3786	0.3239	25.0	58.3	156.5	111.0	1988	54.9	153	112	9957
4	0.0824	0.1800	0.2104	0.0532	25.0								
EMTS 7													
1	0.4922	0.1941	0.2213	0.0401	32.8								
2	0.6369	0.0303	0.0884	0.3864	13.1	59.4	137.6	102.4	1675	58.5	132.9	107.9	9404
3	0.2443	0.1655	0.3821	0.3205	25.0								

Table 4.13 (cont.)

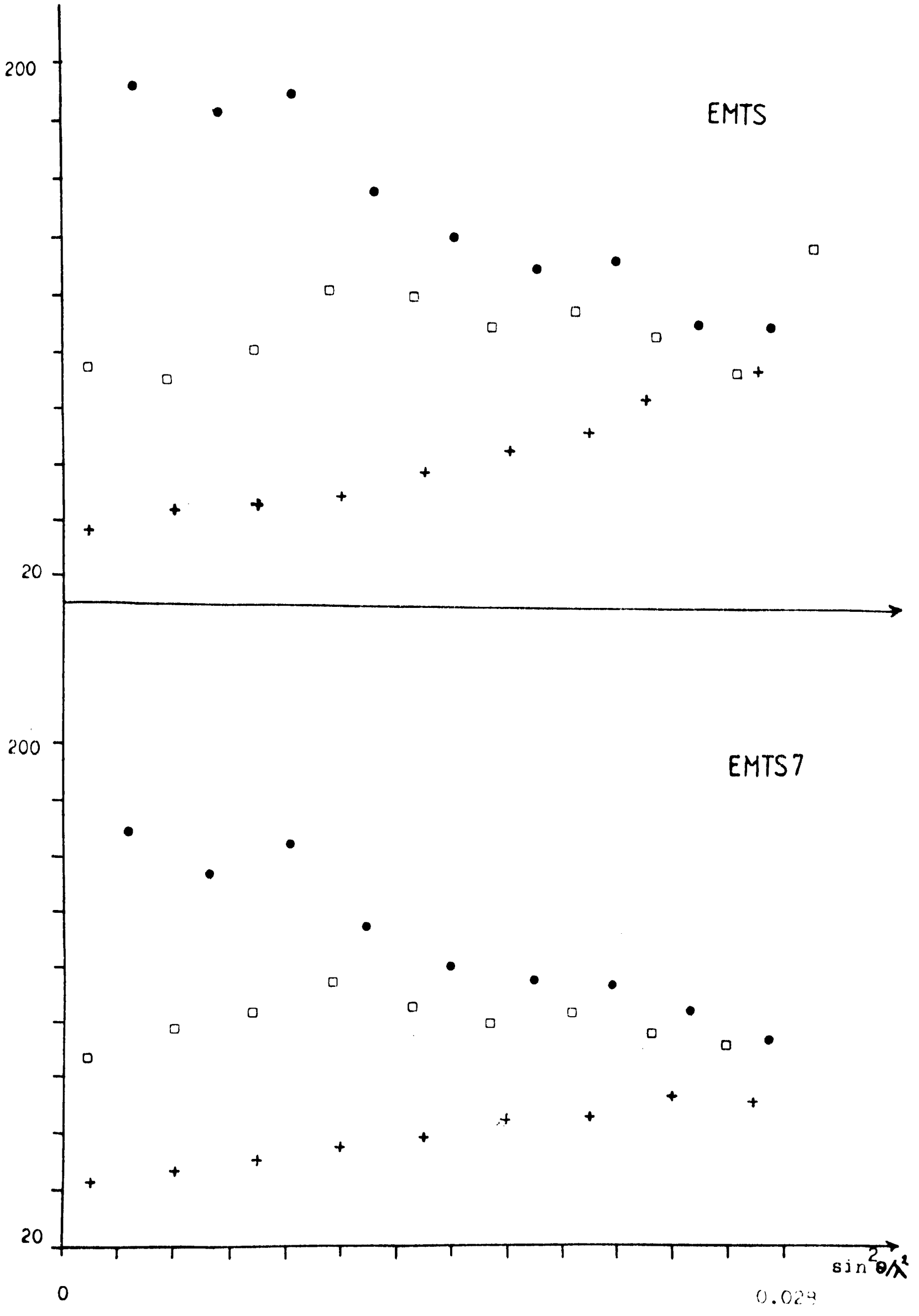
Site	Occupancy	x	y	z	B	R_c	F_c	E_c	N_{obs}	R_a	F_a	E_a	N_{obs}
PT12													
1	0.6061	0.1468	0.3645	0.3385	5.0								
2	0.2726	0.1703	0.3616	0.3486	5.0	65.6	137.4	110.4	1670	59.7	135.7	110.8	7944
3	0.2594	0.2556	0.3838	0.2574	30.0								
4	0.1800	0.1530	0.0547	0.0818	10.0								
Bakers	3.4 Å												
1	0.826	0.1927	0.2175	0.0401	28.4								
5	0.190	0.1934	0.1938	0.0081	30.0	65.1	143.4	115.1	927	60.6	139.2	117.1	4398
3	0.171	0.1603	0.3801	0.3207	25.0								

are given in Table 4.13.

The calculation of the phases of the native protein structure factors by the programme PHASE required the following parameters for each derivative; the occupancy, positional parameters and temperature factor for each heavy atom site, the overall scale (the Kraut scale factor was used) and an estimate of the rms error in the isomorphous and anomalous contributions to the heavy atom structure factors. The rms error in the isomorphous contribution (E) was obtained from a centric refinement, fixing all the parameters at the values obtained from the acentric refinement. The rms errors in the anomalous contributions (E") were obtained by running the programme PHASE with initial approximate values of E" = 1/3E (where E is the isomorphous rms error). This method was proposed by North (1965), who pointed out that the error should be lower for anomalous measurements, as these do not involve such systematic errors as those involved in crystal to crystal scaling. The mean values of the anomalous E output by the programme PHASE for the 'best' phase were used in a second calculation. The phasing procedure involves the minimization of E with respect to protein phase and might therefore be expected to give an underestimate of the E" values. However, for anomalous terms the degree of minimization is very small. The values calculated in this manner were considered satisfactory. The rms errors in the isomorphous and anomalous contributions to the heavy atom structure factors are shown as functions of $\sin^2\theta/\lambda^2$ in Figure 4.10a and b, and the error in deriving the anomalous from the isomorphous E can be seen in these figures. The overall figure of merit (m) obtained in this phase calculation was 0.6397 for 18,555 reflections. The variation of m with $\sin^2\theta/\lambda^2$ is shown in Figure 4.11 and represents an improvement for high angle terms over the previous phase calculation (Weber *et al.*, 1978) (Table 4.2).

FIGURE 4.10a

Calculated contributions of the Heavy Atom derivatives.



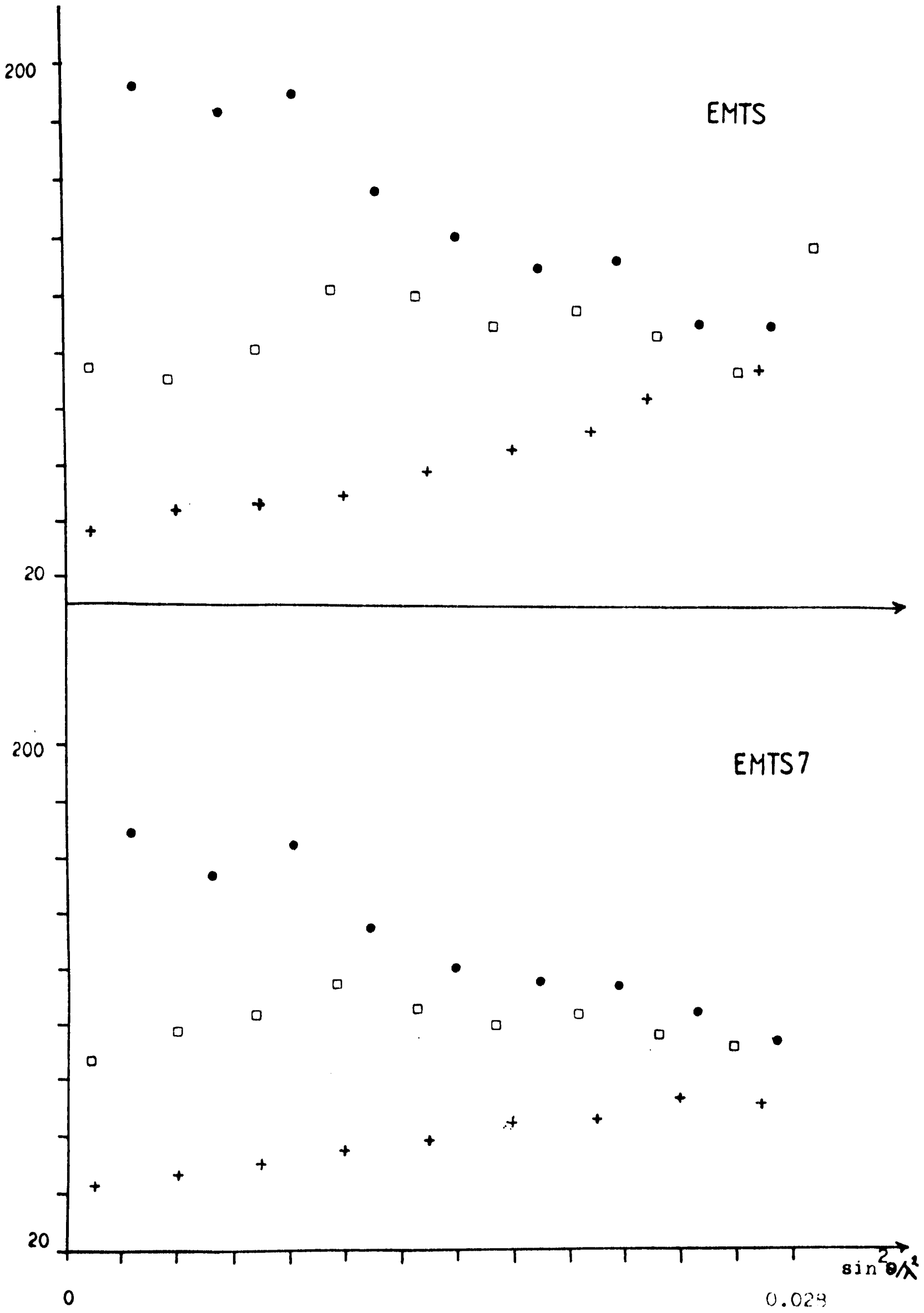
● The mean calculated heavy atom contribution, F_c

□ The rms error, E, in the isomorphous contribution

♦ The rms error, E", in the calculated anomalous contribution

FIGURE 4.10a

Calculated contributions of the Heavy Atom derivatives.



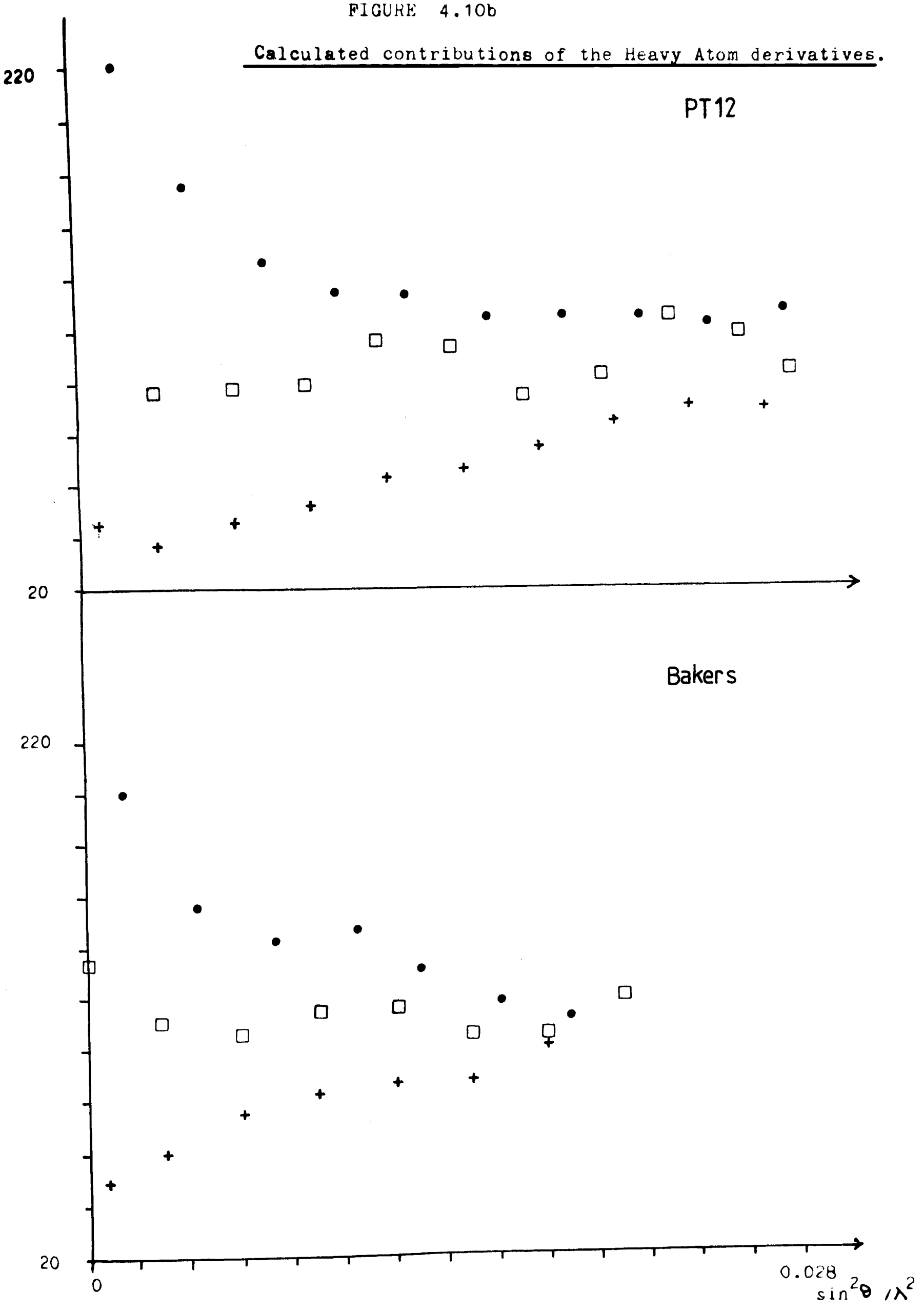
● The mean calculated heavy atom contribution, F_c

□ The rms error, E, in the isomorphous contribution

♦ The rms error, E", in the calculated anomalous contribution

FIGURE 4.10b

Calculated contributions of the Heavy Atom derivatives.



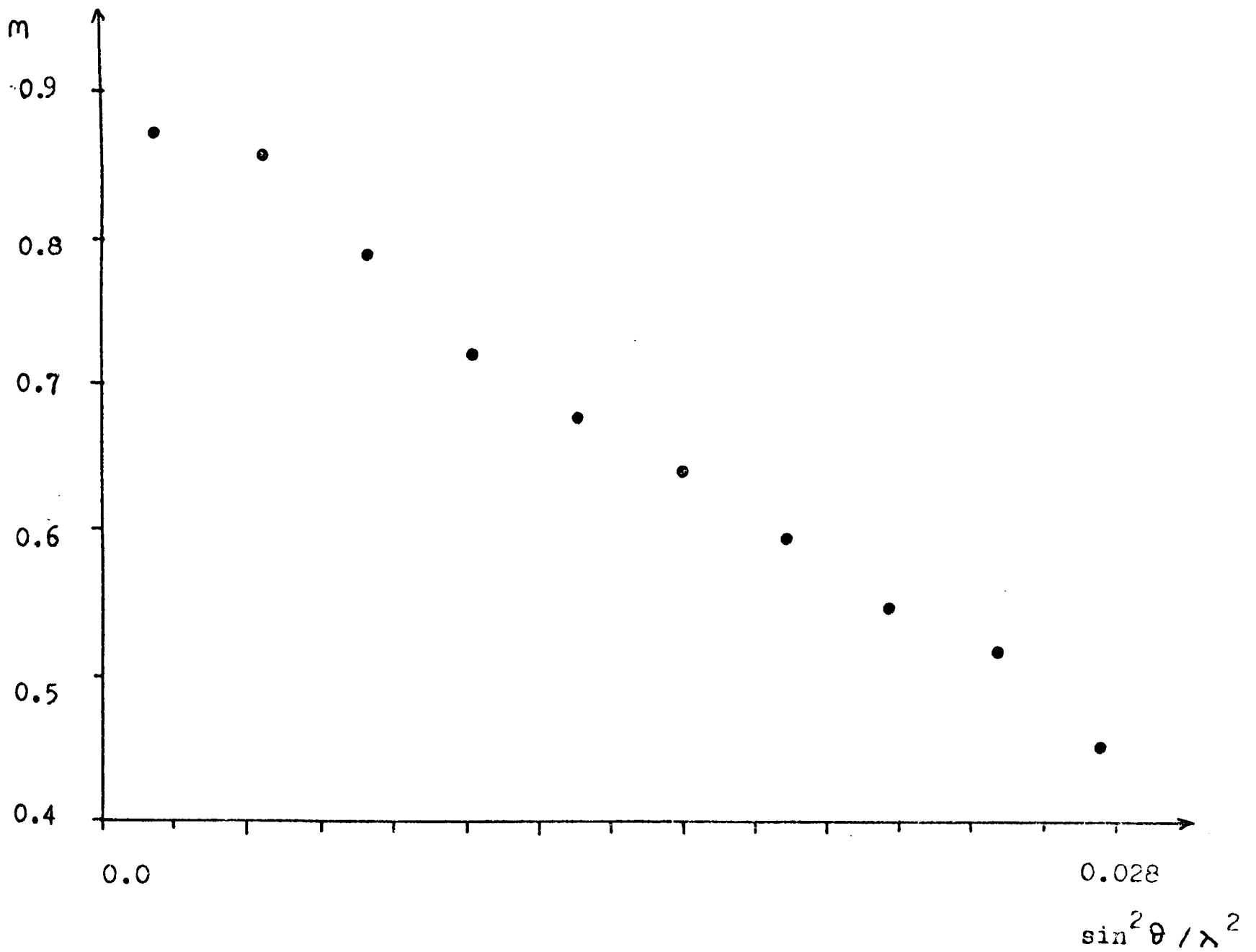
● The mean calculated heavy atom contribution, F_c

□ The rms error, E, in the isomorphous contribution

+ The rms error, E", in the calculated anomalous contribution

FIGURE 4.11

Variation in figure of merit after first phase calculation.



6) NEW PHASE CALCULATION

A second phase calculation was carried out after a number of modifications to the data and further refinements.

(i) The low angle data on the 3 Å photographs were often over-exposed, and to overcome this a set of oscillation camera data to 6 Å resolution was collected by J.A. Jenkins and E.A. Stura. These data were included to provide the low angle terms.

The standard deviation (σ) for both 6 Å and 3 Å terms had previously been calculated as

$$\sigma^2 = \sum_{i=1}^{N} (\bar{x}-x_i)^2 \ /n^2 .$$

This could give misleading values if only two measurements were present and $x_1 = x_2$ (*i.e*: $\sigma = 0$). These terms would then be included in the refinement with the wrong weights. The standard deviation was changed to

$$\sigma^2 = (\sigma_1{}^2+\sigma_2{}^2)/4 \qquad (cf.\ \text{Topping, 1965})$$

and the 3 Å and 6 Å photographic data were merged using this standard deviation. 7967 out of 18698 terms were affected by this change (including both 6 Å and 3 Å data).

(ii) The 6 Å diffractometer data (Johnson *et al.*, 1974) were included as follows:-

a) Terms with F_P(diff)>200 were selected and scaled to the new data.

b) If F_P and α_P from 3 Å data were present, 3 Å values were chosen.

c) If F_P but not α_P from 3 Å data were present, α_P from diffractometer data was used.

d) If neither F_P nor α_P from 3 Å data were present, F_P and α_P from diffractometer data were used.

The inclusion of these data also helped to fill in the blind region to 6 Å.

(iii) A further site HG4 was included in the refinement of EMTS7, since the final 3-D Fourier from F_{HLE} refinement showed a single contour level, about the same as the noise level, at this position. However, since centric refinement gave a large B value for HG1 (which is approximately 3.3 Å from HG4), this may have been an attempt by the refinement to account for a weak site at HG4. Firstly HG4 was introduced but fixed whilst the other three sites were refined. Then the other sites were fixed whilst HG4 was refined. The same procedure was carried out for the EMTS derivative.

(iv) Although the best results (*i.e*: lowest R factor and highest F_C/E) for PT12 were obtained from the refinement with modified form factors, it was decided to use the results from refinement with normal form factors. The analysis of occupancy with $\sin^2\theta/\lambda^2$ gave no indication of the extreme modification of form factors suggested by F_{HLE} refinement. Instead the temperature factors of sites PT1 and PT3 were modified and normal form factors were used. The parameters from this refinement are given in Table 4.14.

(v) Since the main differences between the centric and acentric refinements were slight increases in occupancy and temperature factors, it was decided to use the parameters obtained from a centric refinement, using 3 Å data only, for the phase calculation. Higher limits of F_{HLE} and F_{HUE} (*i.e*: $F_{HLE}<475$ and $F_{HUE}>300$) were used to take into account the new scale of the data. The final parameters used in the phase calculation are given in Table 4.14. The rms errors in the isomorphous and anomalous contributions to the heavy atom structure factor are shown in Figure 4.12a and b.

The overall figure of merit from this phase calculation was 0.63 for 18,555 reflections. The variation of m with $\sin^2\theta/\lambda^2$ is

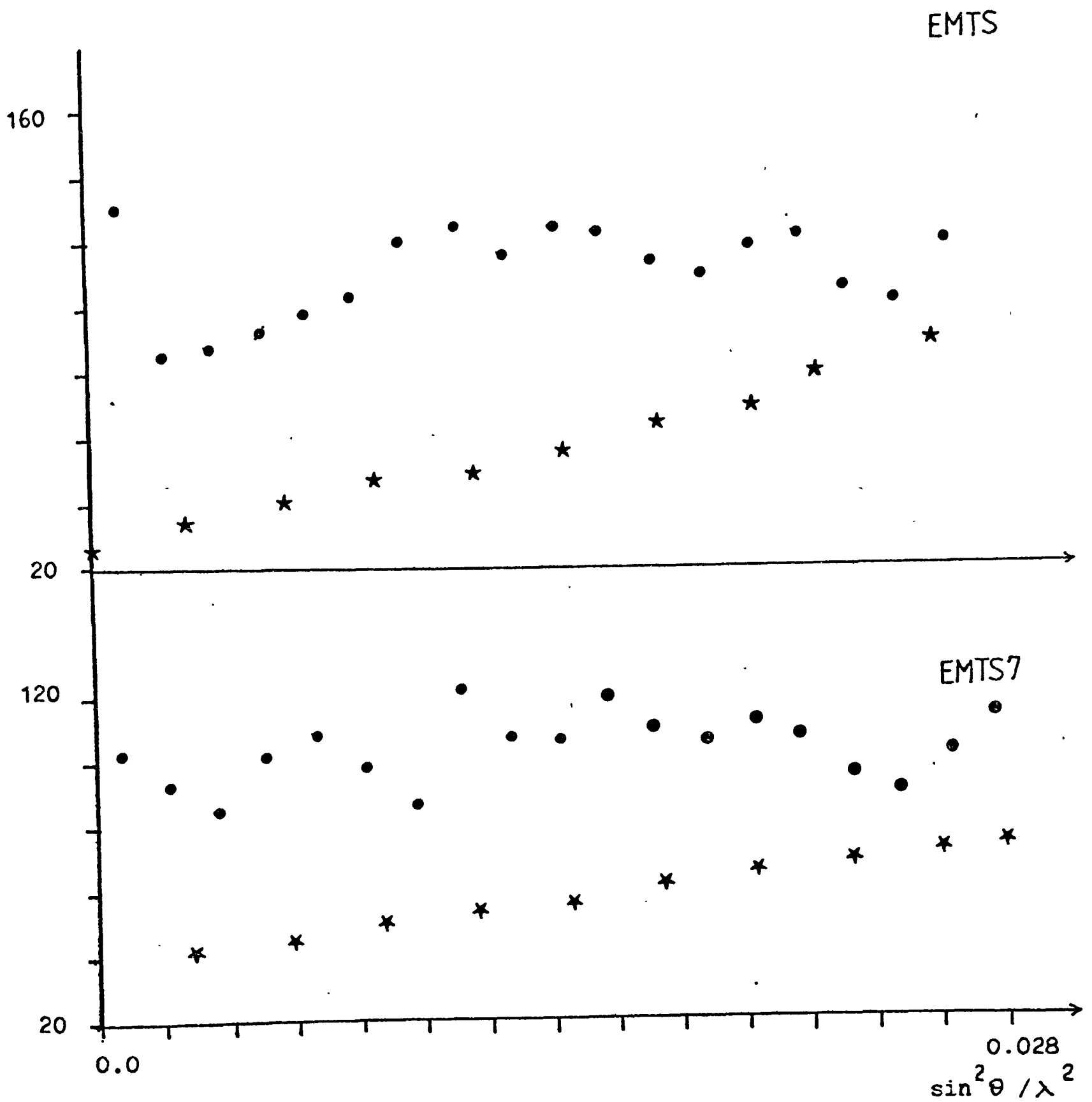
Table 4.14

Summary of second refinement - centric
3 Å only: F_{HLE}<475 F_{HUE}>300

Site	Occupancy	x	y	z	B	R_c	F_c	E_c	N_{obs}
EMTS									
	1	0.6084	0.1960	0.2243	0.0387	30.57			
	2	0.8244	0.0282	0.0890	0.3840	24.09			
	3	0.2070	0.1588	0.3791	0.3238	25.0	55.1	163.6	112.4	1948
	4	0.1043	0.1800	0.2105	0.0513	25.0			
EMTS7									
	1	0.5740	0.1946	0.2222	0.0389	50.17			
	2	0.7564	0.0299	0.0884	0.3867	21.19			
	3	0.2399	0.1661	0.3819	0.3213	25.0	57.1	149.2	105.6	1645
	4	0.1547	0.1782	0.2081	0.0540	25.0			
PT12									
	1	0.7077	0.1484	0.3662	0.3386	10.0			
	2	0.3722	0.1720	0.3609	0.3520	5.0			
	3	0.2547	0.2544	0.3849	0.2572	5.0	65.6	153.9	119.3	1631
	4	0.1800	0.1518	0.0570	0.0838	10.0			
Bakers									
	1	0.8027	0.1920	0.2199	0.0403	23.66			
	5	0.2672	0.1977	0.1969	0.0052	30.0	62.5	155.4	120.7	900
	3	0.1842	0.1678	0.3819	0.3156	25.0			

FIGURE 4.12a

RMS errors in the isomorphous and anomalous contributions to the heavy atom structure factors after second phase calculation.

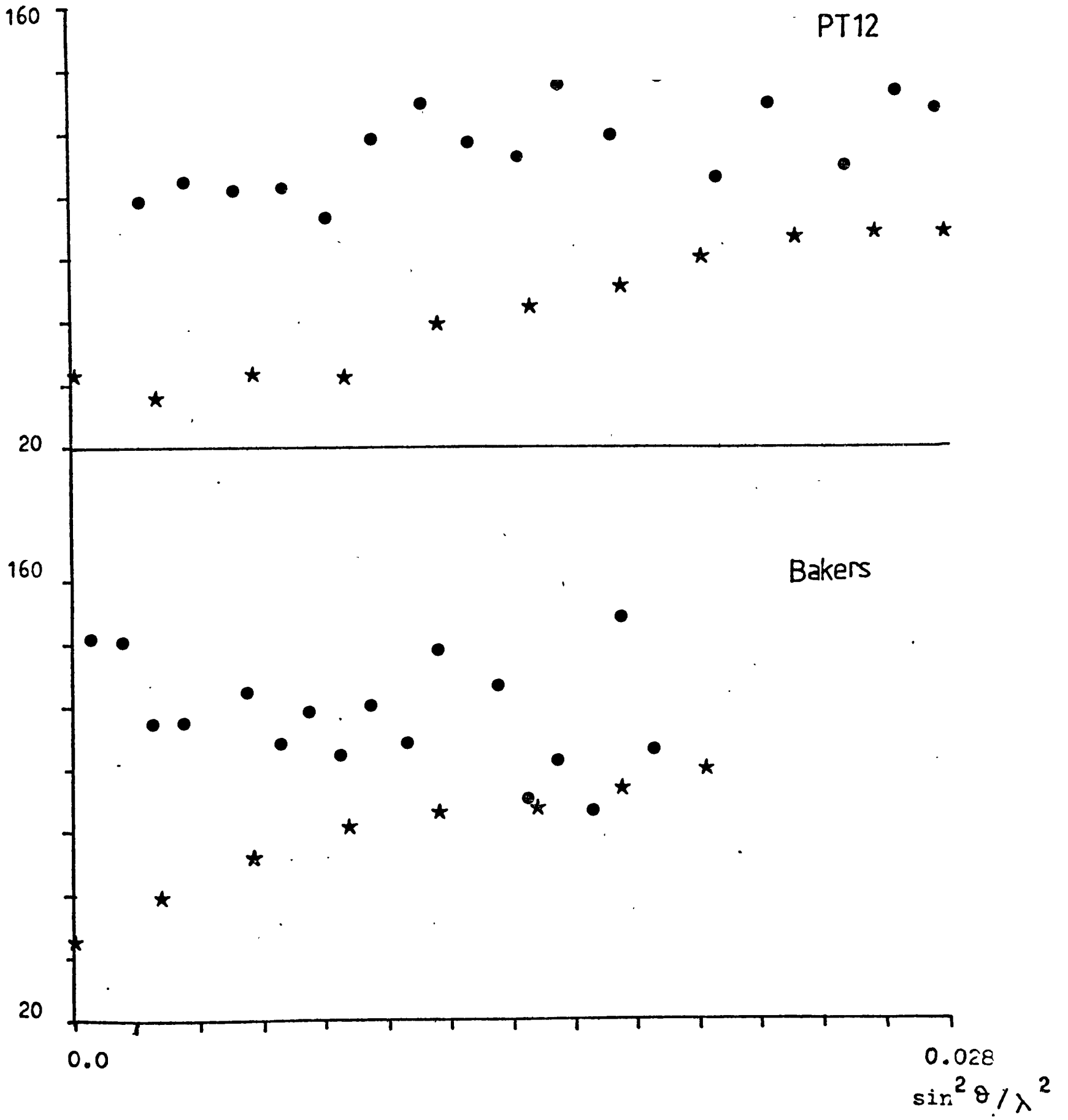


● E, the rms error in the isomorphous contribution

★ E", the rms error in the anomalous contribution

FIGURE 4.12b

RMS errors in the isomorphous and anomalous contributions to the heavy atom structure factors after second phase calculation.

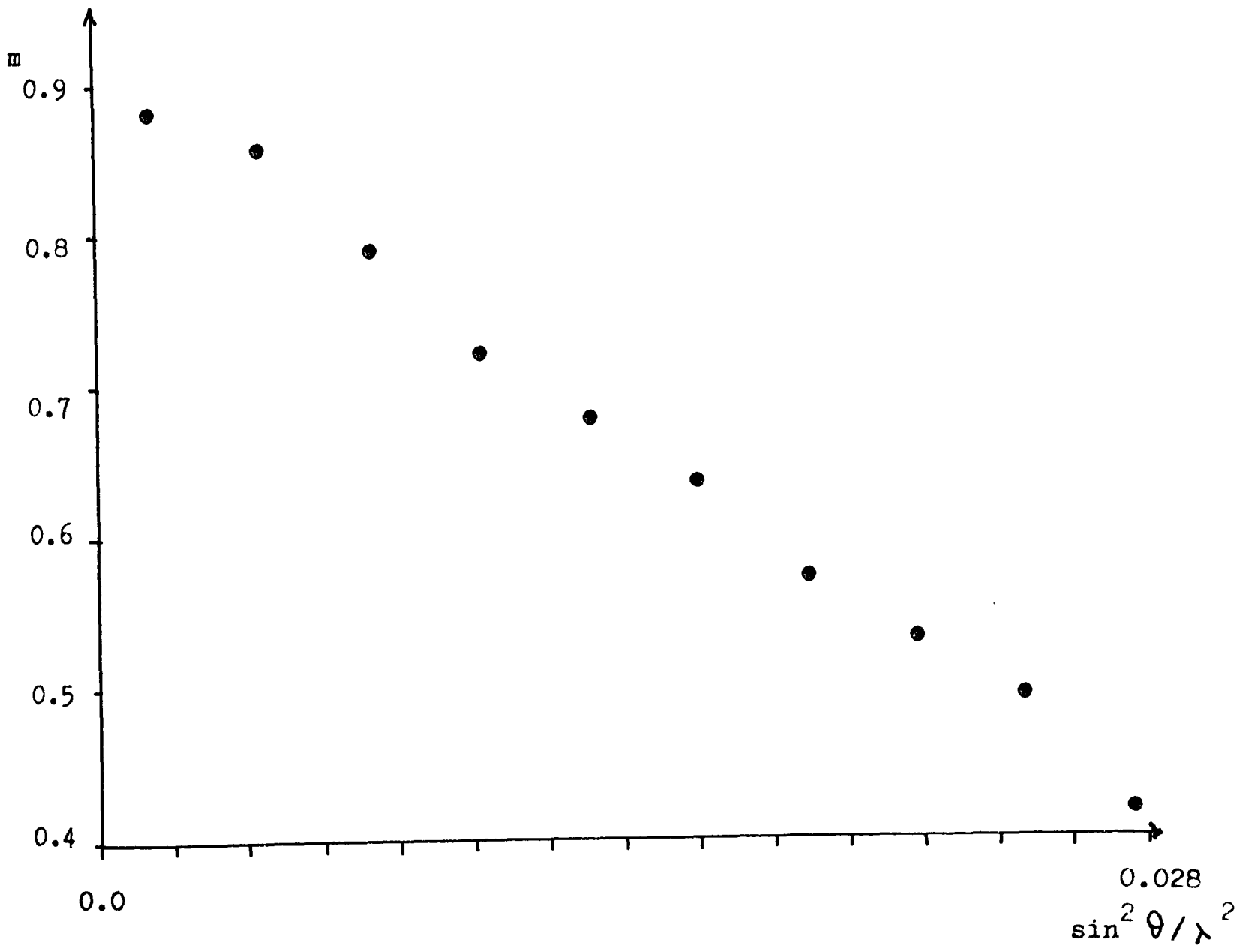


● E, the rms error in the isomorphous contribution

★ E", the rms error in the anomalous contribution

FIGURE 4.13

Variation in figure of merit after second phase calculation.



Overall figure of merit 0.63

shown in Figure 4.13.

Both phase calculations gave an improved electron density map, which allowed easier interpretation of some areas of electron density which had previously been ambiguous (Weber *et al.*, 1978). The new maps showed greater connectivity for the polypeptide chain and better electron density for the amino acid side-chains. A molecular model on the scale of 2 cm/Å has been constructed using a Richards Box (Richards, 1968) and co-ordinates have been measured using an electronic co-ordinate measuring device (Johnson *et al.*, 1980). This model has allowed some preliminary analysis to be made of the heavy metal binding sites, described in the next section.

7) LOCATION OF HEAVY METAL BINDING SITES ON THE MOLECULE

The way in which the heavy metal reagents react with phosphorylase *b* depends on the accessibility of reactive amino-acid side chains as well as the charged, polar or hydrophilic nature of the reagent itself. In general the heavy metal binding sites appear near the surface of the molecule and are summarized in Table 4.15 and shown in Figure 2.6.

The mercury binding sites were all associated with cysteine residues, which is consistent with the high binding constants of mercury compounds used in isomorphous replacement to sulphydryl groups. It is interesting to compare the results with those obtained from studies on the reactivity of the sulphydryl groups in solution (Akramovic-Kikic *et al.*, 1970; Battell *et al.*, 1968). In these studies Cys 171 and 317 could be reacted without the loss of enzyme activity, whilst Cys 108 and 142 reacted much more slowly with loss of activity and dissociation of the protein into monomers. The alkylation of Cys 108 is prevented by the presence of AMP (Battell *et al.*, 1968). The remaining sulphydryl groups did

Table 4.15

Heavy Metal Binding Sites

Derivative	Site	Residue probably involved in binding	
EMTS	HG1	Cys 782	
	HG2	Cys 108	
	HG3	Cys 171	
$Pt(NH_3)_2Cl_2$	PT1	Met 176 or Cys 171	
	PT2		
	PT3	Met 614 or His 613	
$KAu(CN)_2$	AU1	Cys 171	
	AU2A	Arg 309	(Site N)
	AU2B	+?	
	AU4	Cys 782	
	AU5	Cys 142	

not react with iodoacetamide unless the protein was fully denatured.

The major mercury binding sites (HG1 and HG2) are Cys 782 and 108 respectively. Cys 108 occurs at the N-terminus of a short helix, close to the dimer-dimer interface but not in contact with it. The low reactivity of Cys 108 with iodoacetamide is explained by its relative inaccessibility, in a small pocket exposed to the solvent. The protection of Cys 108 by AMP binding is more difficult to explain, since it is approximately 27 Å from the AMP site (site N). However, site N lies at one end of the AA' helix and Cys 108 lies close to the other end (Figure 2.6) and it is possible to speculate about some interaction *via* the helix. Cys 782 has not previously been reported as a reactive sulphydryl and its lack of reactivity in solution remains unexplained, since its position is relatively exposed and appears no less accessible than that of Cys 108 (Figure 2.6). HG3 is associated with the partially disordered loop (shown dotted in Figure 2.6) containing Cys 171, which, together with Cys 317, is one of the fast reacting sulphydryls described by Akramovic-Kicic *et al*. (1970). The actual binding site is difficult to locate accurately because of this disorder, and by the fact that, due to the crystallographic symmetry, the other disordered loop between residues 311-323 lies close to this loop in the electron density map. Since this second disordered loop contains Cys 317, it is possible that this residue may be associated with HG3. The disorder of the loop containing Cys 171 may be due to mobility. Alternatively, the disturbance may be caused by the binding of the platinum derivative close to Cys 171 (see below).

Previous treatment of the crystals with 10 m*M* iodoacetamide for five hours before soaking in 1 m*M* EMTS lead to the protection of site HG3, but no change in the occupancies of HG1 and HG2. This result suggested that Cys 782 and Cys 108 are not readily modified

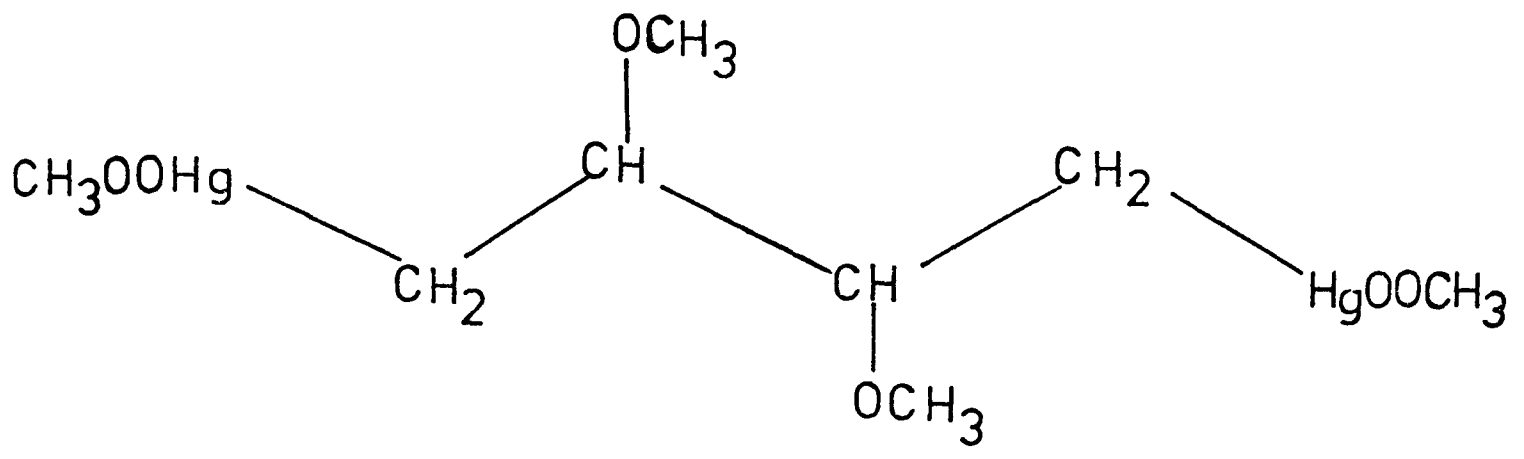
by iodoacetamide, which agreed with solution studies and supported the assignment of HG3 to Cys 171 (Weber *et al.*, 1978). HG3 has the lowest occupancy of the three mercury binding sites, and the difference in reactivity between the crystal and solution studies may be due to the fact that, in the crystal, Cys 171 lies close to Cys 317. Site HG4, which was included in the second refinement, may reflect an attempt to take account of the large temperature factor for HG1, since this site is approximately 3.3 Å away from HG1. A residual positive peak in the electron density at Cys 782 suggests that there may still be some error in the modelling of this site. Site HG5 which occurs in the Bakers dimercurial derivative at 4.95 Å from HG1 almost certainly corresponds to the second mercury in the Bakers dimercurial molecule (Figure 4.14).

The reactions of protein crystals to platinochloride have been discussed by Petsko (1973). At acid pH's the platinochloride ion ($PtCl_4^{2-}$) reacts with methionines, cysteine disulphides, N-termini and histine imidazole, which all form stable complexes. These are good nucleophiles which can displace chloride from platinum complexes. Cysteine sulphydryl groups are less nucleophilic. The neutral $Pt(NH_3)_2Cl_2$ molecule (which was the compound used to produce a derivative of phosphorylase *b*) can penetrate into hydrophobic areas of the protein. It will require a stronger nucleophile but will be reactive to anionic nucleophiles such as $R\text{-}S^-$. Three platinum binding sites have been observed in phosphorylase *b*. PT1 and PT2 are a double site, associated with the disordered loop which contains Cys 171 and Met 176. Because of this disorder the actual binding site is difficult to determine. PT3 is probably associated with Met 614 or His 613.

The gold ligand $Au(CN)_2^-$ will tend to bind to anionic sites, but may also give tetrahedral complexes with 'soft' ligands such as cysteine (Blundell and Johnson, 1976). ('Soft' ligands are

FIGURE 4.14

Baker's Dimercurial



The observed separation of Hg1 - Hg5 is 4.95 Å. If the molecule is extended a separation of ~ 7 Å would be expected.

polarizeable and form covalent bonds (Pearson, 1963)). Phosphate may also bind in similar anionic pockets and interfere with the binding of $Au(CN)_2^-$. For instance, in chironomus haemoglobin (Huber *et al.*, 1969a,b), $Au(CN)_2^-$ does not bind in the presence of phosphate. $Au(CN)_2^-$ may also act as an inhibitor. In liver alcohol dehydrogenase (Branden *et al.*, 1973) it binds in strict competition with the coenzyme NAD, at both the adenosine and phosphate parts of the binding site. In phosphorylase *b* the binding sites for $KAu(CN)_2$ are almost all associated with the mercury binding sites, and hence cysteine residues. AU1 is close to HG3 (Cys 171), AU4 is close to HG1 (Cys 782) and AU5 is close to Cys 142. The last result is rather surprising, since it is one of the slow-reacting sulphydryls and occurs in a rather inaccessible position in the interior of the molecule. However, Arg 138 and Arg 488 which occur in the vicinity may provide a positively-charged environment favourable to the binding of $Au(CN)_2^-$. However, why the ligand should bind to Cys 142 in preference to any of the other, more exposed sulphydryls, remains unexplained. $Au(CN)_2^-$ is also observed to bind in the nucleotide binding site N, at both the adenosine and phosphate parts of the site. The coordinates of site 2A are close to the phosphate binding site, and those of 2B close to the adenosine site. The binding of AMP in the crystal has been studied crystallographically and by model building (Johnson *et al.*, 1979), and the results are shown in Figure 2.7. The phosphate group is located between the α-helices AA' and BB' and makes an ionic interaction with Arg 309. Tyr 75 (AA' helix) is in contact with the base and the phosphate binding pocket. Arg 309 is also a likely candidate for binding $Au(CN)_2^-$. The gold ligand thus shows the same type of behaviour as that observed in liver alcohol dehydrogenase, where Arginine residues are also involved in binding $Au(CN)_2^-$ at the nucleotide site (Branden *et al.*, 1973).

No binding of $Au(CN)_2$ was observed at the second nucleotide site (Site I). In the studies of nucleotide binding (Johnson *et al.*, 1979) it was observed that the phosphates of AMP and ATP bound at this site did not make any specific contacts with the enzyme, and so this does not appear to be a high affinity binding site for anions.

Thus most of the heavy metal binding sites in phosphorylase *b* appear plausible when compared with those observed in other proteins. However, a number of questions, mainly related to the reactivity of the sulphydryl groups, remain unanswered. These may be elucidated when an atomic model at higher resolution is obtained.

CHAPTER 5

COLLECTION OF 2 Å NATIVE DATA

1) THE PROBLEM OF DATA COLLECTION

Difficulties often arise in protein crystallography from a combination of large unit cells (due to high molecular weights or complex packing arrangements) and small crystals. The intensity of the X-ray diffraction pattern depends on the number of unit cells which are scattering. For a protein of molecular weight ~40,000 Daltons a crystal of dimensions of 0.3 mm in each direction will often give a good diffraction pattern, but a protein of 300,000 molecular weight would require all crystal dimensions to be approximately twice as large to give the same intensity.

Although large crystals can be grown routinely, phosphorylase *b* presents problems for high resolution (2 Å) data collection because of its large subunit molecular weight of 97,412 (asymmetric unit) and its large unit cell dimensions. The complete 2 Å data set comprises approximately 64,000 reflections, each of which must be measured four times using the oscillation camera in order to record the unique data.

The method chosen for data collection was the screenless oscillation method using the Enraf-Nonius oscillation camera. The advantages of this method have been discussed in Chapter 3. However, the standard flat-plate film cassettes supplied with the camera impose some limitations on the collection of high resolution data. Whilst flat films are satisfactory for maximum Bragg angles of up to about 45°, there are restrictions on their use for the collection of high resolution data for large molecules. The factors involved are resolution, cell dimensions and spot size. In general, the crystal to film distance is restricted by the maximum resolution required, and the theoretical maximum resolution

obtainable is determined by the dimensions of the camera. The flat plate cassettes expose a circular area of 59 mm radius and the shortest possible crystal to film distance is 55 mm. This would give a theoretical maximum of 1.9 Å using CuKα radiation. Another factor which must be taken into account, however, is the separation of the spots on the film, and, for a given crystal to film distance, this will be determined by the longest unit cell axis. For phosphorylase *b* the longest unit cell axis is 128.5 Å and with a crystal to film distance of 55 mm (and λ = 1.54 Å) the separation will be approximately 0.7 mm. Thus spots of 1 mm diameter (a typical spot size for phosphorylase *b*) would not be resolved.

Yet another factor which must be taken into account is that, for high resolution data, the angle of incidence of the diffracted rays may be too oblique to resolve reflections on the film. The area of the spot on the film will be increased by a factor of $Sec^2 2\theta$ (where θ is the Bragg angle for the reflection). For the example given above (maximum radius = 59 mm, minimum crystal to film distance = 55 mm) this factor will be approximately 2.2.

To overcome these problems a cylindrical cassette was built, specifically for the collection of high resolution data for phosphorylase *b*. The design and construction are described in Section 2.

A second problem concerned the exposure times necessary to record the 2 Å data using conventional X-ray sources. Using an Eliot GX6 rotating anode run at 40KV, 40mA, a 0.5° oscillation photograph with data extending to 2 Å required 13 hours exposure. Only one such oscillation photograph could be recorded from each crystal before the onset of unacceptable radiation damage. This probably meant that the crystal was becoming disordered during the latter part of the exposure. Furthermore, one of the main advantages of the oscillation method is lost if only one photograph

per crystal can be recorded. This is the ability to sum the separate parts of partially recorded reflections at the beginning and end of adjacent rotation ranges (*i.e*: on adjacent photographs taken from the same crystal). With only one photograph, the degree of partial recording must be estimated, which is always subject to error, or a large rotation range must be used to provide an overlap with the next photograph. In this case a large number of overlapping reflections will be produced.

Thus the collection of the complete phosphorylase *b* 2 Å data set, requiring a rotation of 45°, proved virtually impossible with conventional X-ray sources. This problem was overcome by the use of the brilliant, highly focussed beam provided by synchrotron radiation, as described in Sections 3-5.

2) THE DESIGN AND CONSTRUCTION OF THE CYLINDRICAL CASSETTE

The cylindrical cassette was originally designed by D.G.R. Yeates and built by M.G. Pickford in the laboratory workshop. The experimental implementation and the modification of computer programmes were carried out by myself.

The design of the cassette is such that the reflections in the plane of reciprocal space orthogonal to the axis of rotation impinge on the film at an angle of 90°. The optimum crystal to film distance chosen was a compromise between the best resolution (short distance) and the resolution of spots on the film (long distance). For phosphorylase *b* (a* = b* = 128.5 Å, c* = 115.9 Å) the crystal to film distance chosen was 96 mm, which had also been used for the collection of the 3 Å data. The radius of the cylinder is thus 96 mm and the arc length of 160 mm enabled data extending to 2 Å to be recorded. The axis of the cylinder is coincident with the oscillation axis of the crystal. For the zeroth level reciprocal lattice orthogonal to the oscillation axis,

the diffracted beams will be everywhere normal to the film, if the crystal is perfectly set. For other levels the peak size will vary as $P \approx P_o \, Sec\chi$, where χ is the angle between the generator for the nth layer line and the normal to the rotation axis (Figure 3.11). Correction for oblique incidence is thus only necessary in this direction. The modifications incorporated into the processing programme OSCAR have described in Chapter 3.

The overall construction of the cassette is shown in Figure 5.1 and consists of four sections:-

i) back of the cassette - identical to the Nonius assembly (Figure 5.2),

ii) two cradles of $\frac{3}{8}$" (1 cm) thick aluminium alloy (Figures 5.1 and 5.3),

iii) segment of large diameter tube ($\frac{5}{16}$", 8 mm wall),

iv) film holder (Figures 5.7, 5.8).

The body of the cassette was made from a segment of large diameter aluminium alloy tube, with suitably thick walls, so that the required curvature could be machined into it. The curve of the cradles was rough cut. The base of the cradles was fine machined and then fixed to the back of cassette. The curvature of the cradles was then machined, square to the axes of the cassette, to the outside curvature of the segment of large diameter tube. This was then fixed to the cradles with six screws from the inside (Figure 5.4.). The inside curve of the segment of tube was then machined to the radius of a cylinder concentric to the Ewald sphere + the thickness of the back plate of the film holder + the thickness of the number of films required. The film holder runners were made from an aluminium alloy strip, step-machined and then bent to the curvature of the cassette (Figure 5.5). The depth of the step was equal to the

FIGURE 5.1

Cylindrical Cassette:- overall view.

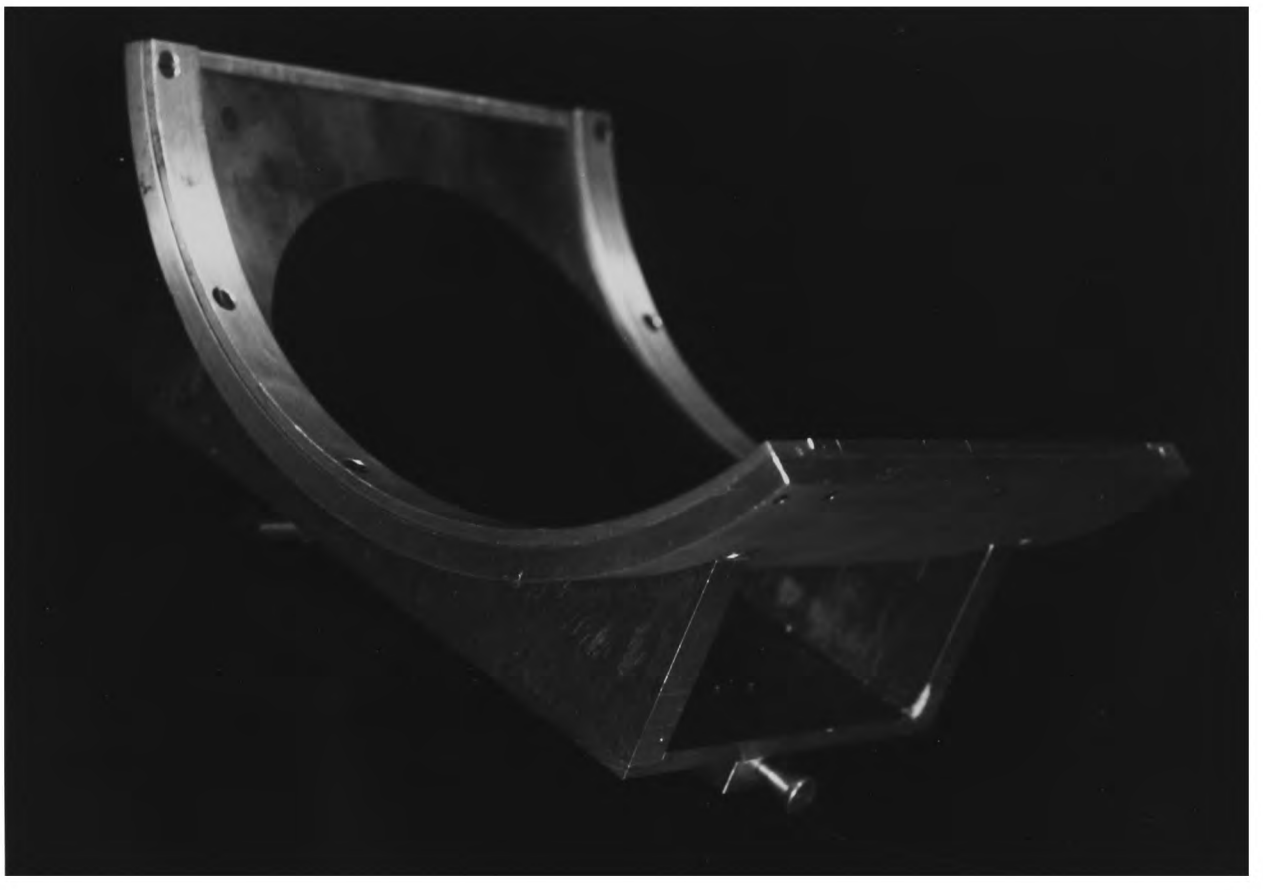


FIGURE 5.2

Cylindrical Cassette:- back view.

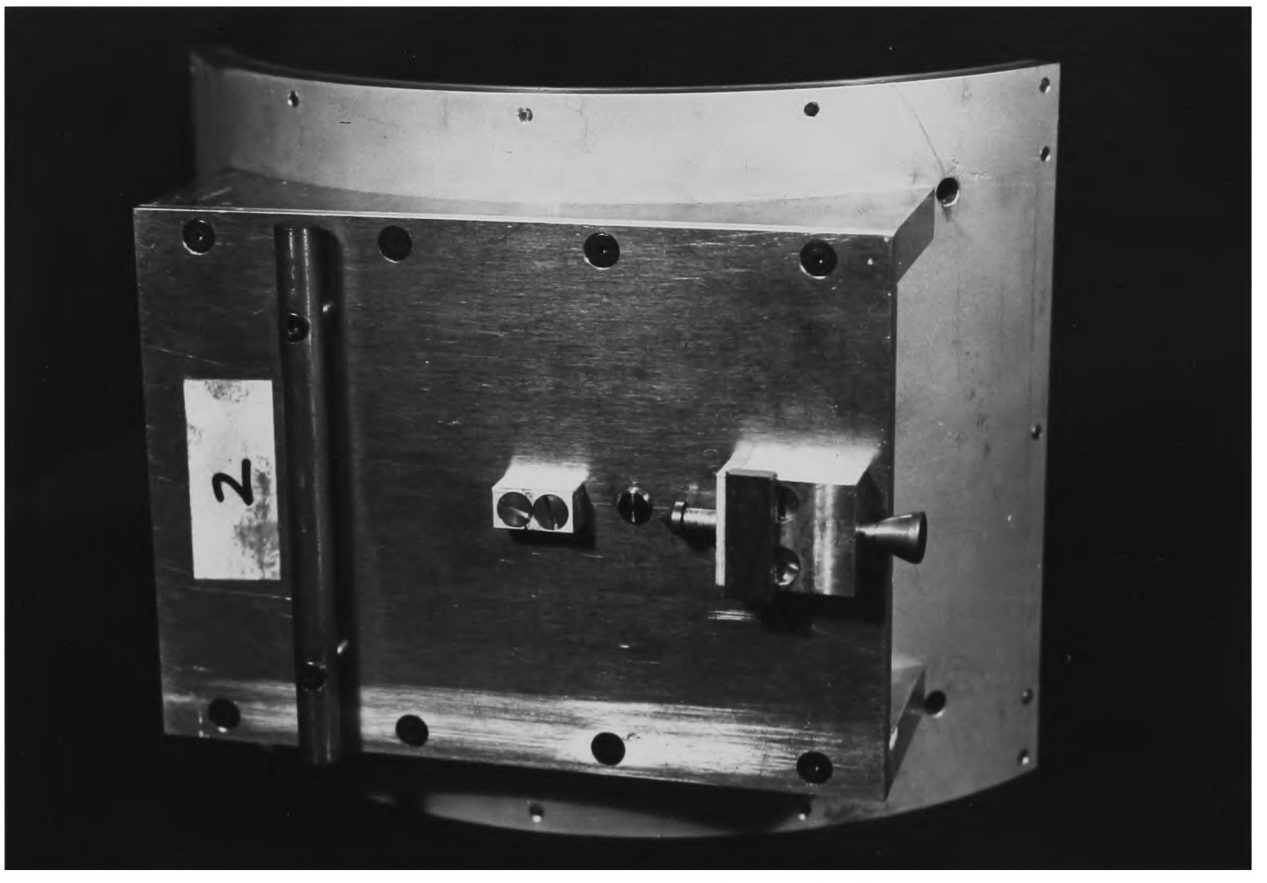


FIGURE 5.3

Cylindrical Cassette:- end view, showing cradles between back assembly and body of cassette.

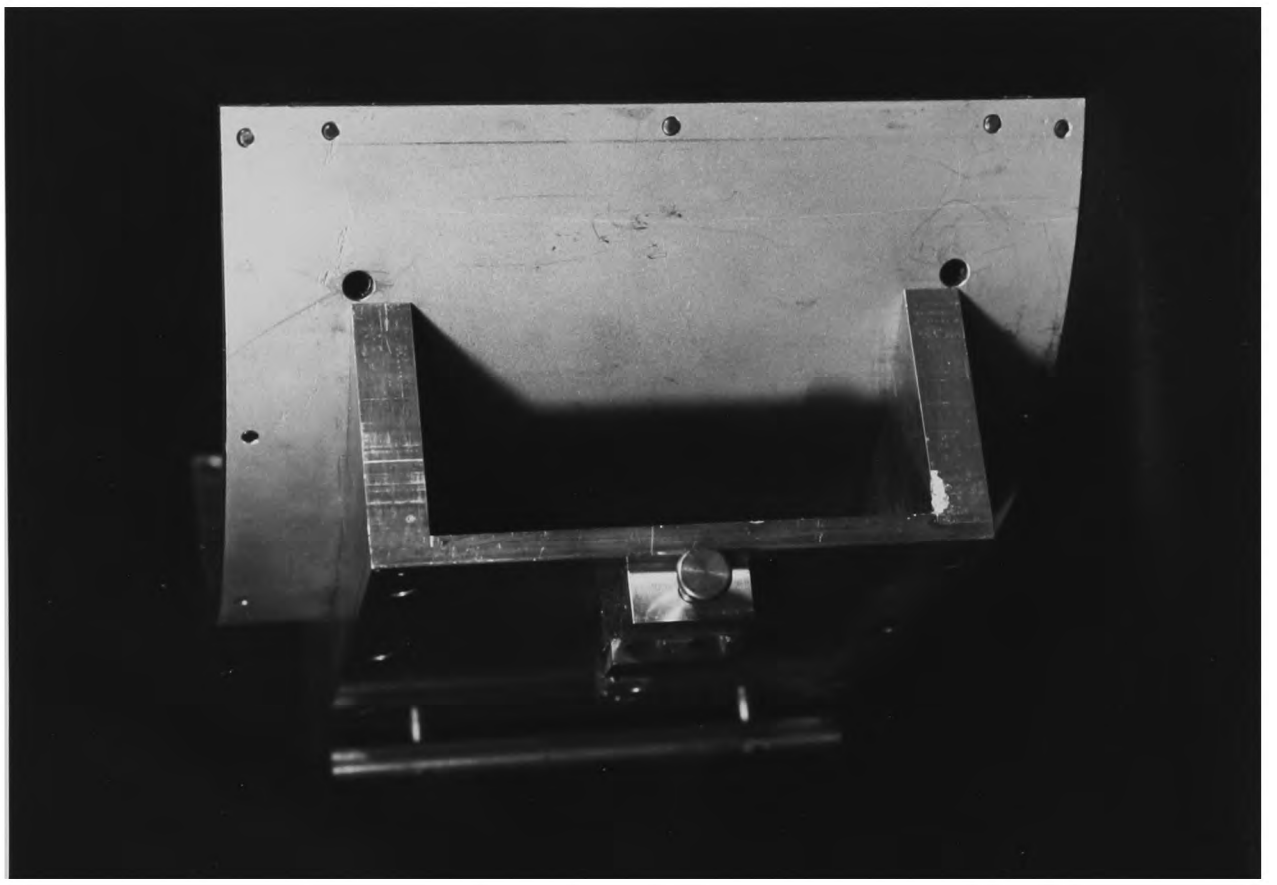


FIGURE 5.4

Cylindrical Cassette:- inside of body of cassette, showing screws which fix it to the cradles.

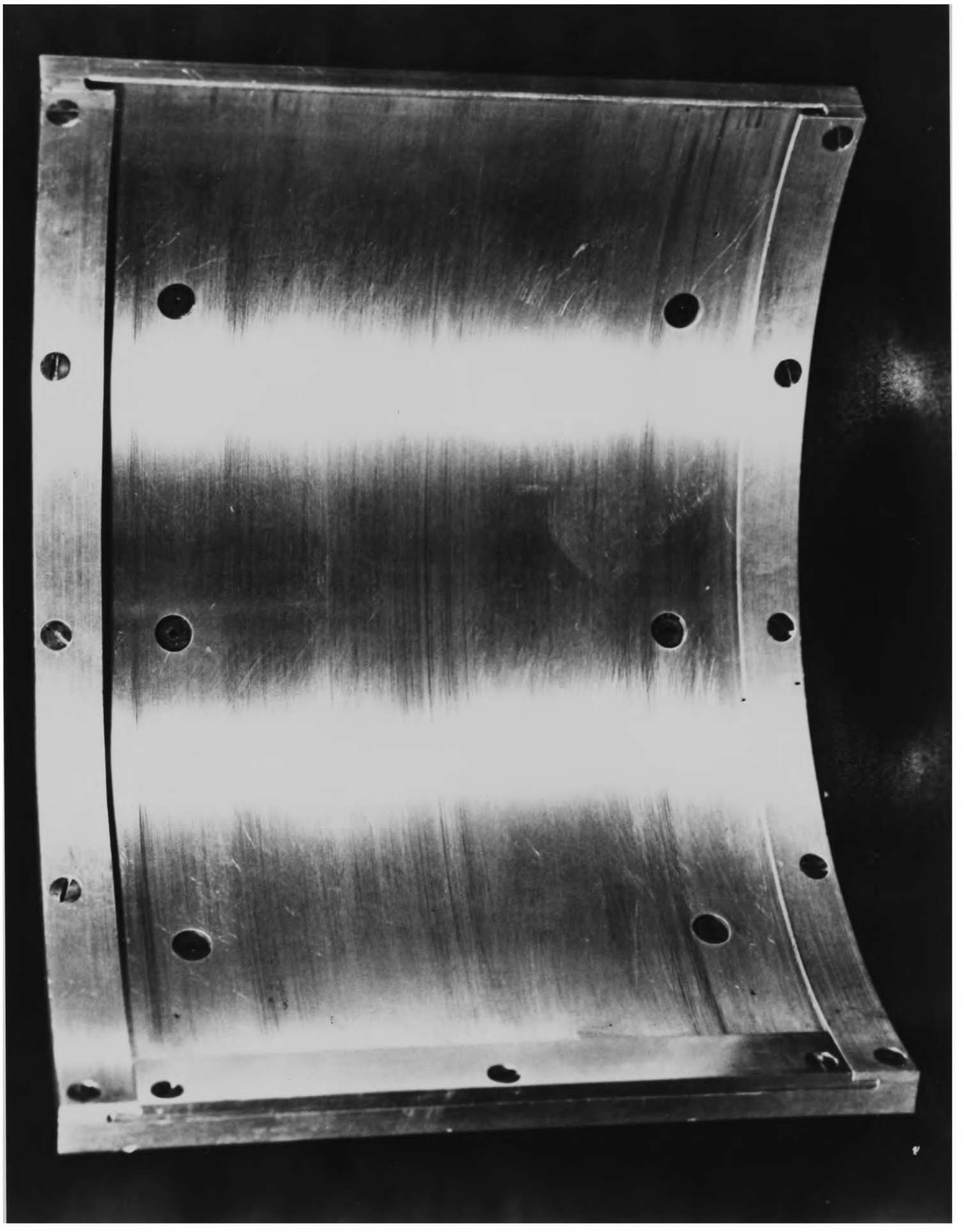


FIGURE 5.5

Cylindrical Cassette:- insertion of film holder between runners.

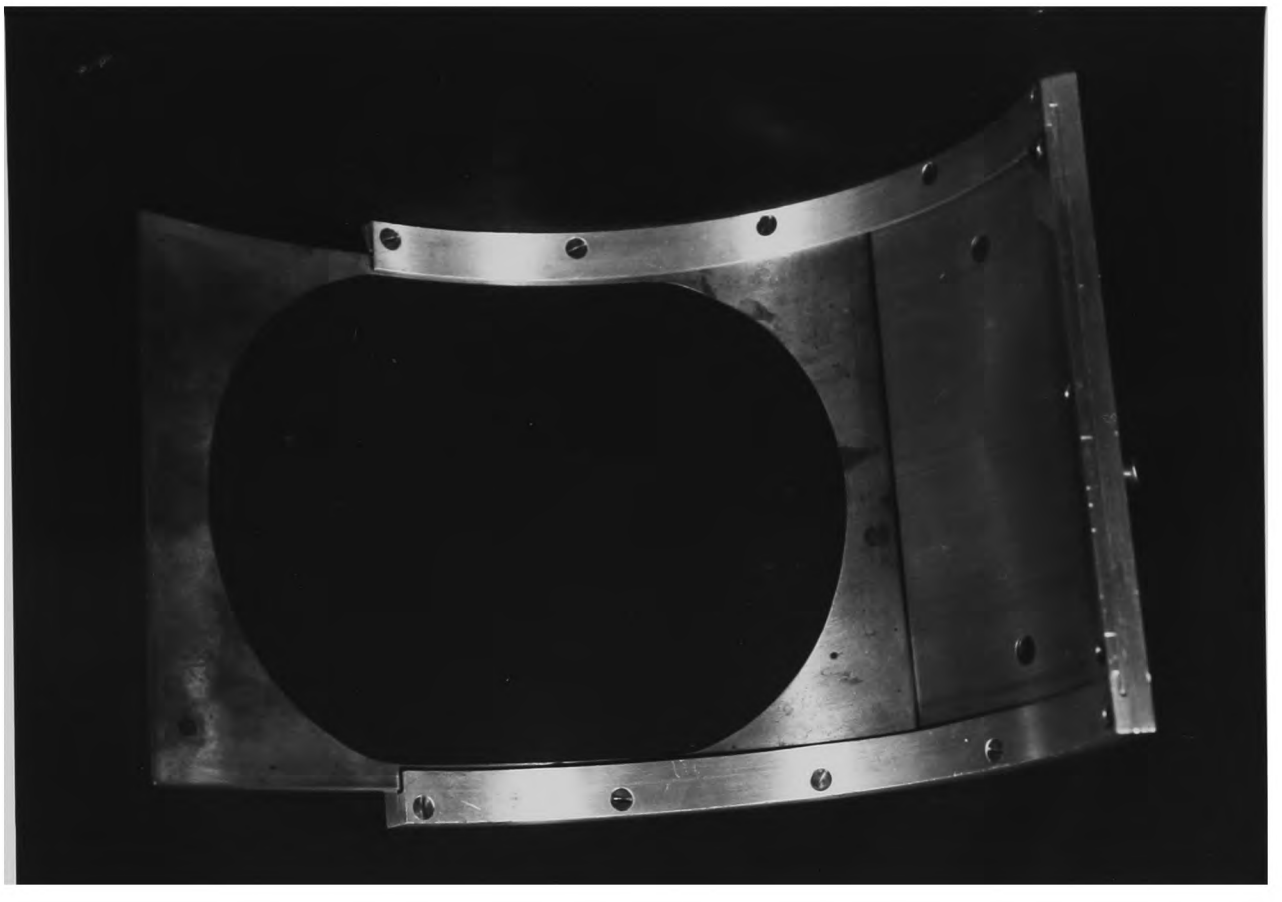


FIGURE 5.6

Cylindrical Cassette:- front view, showing film holder in position.

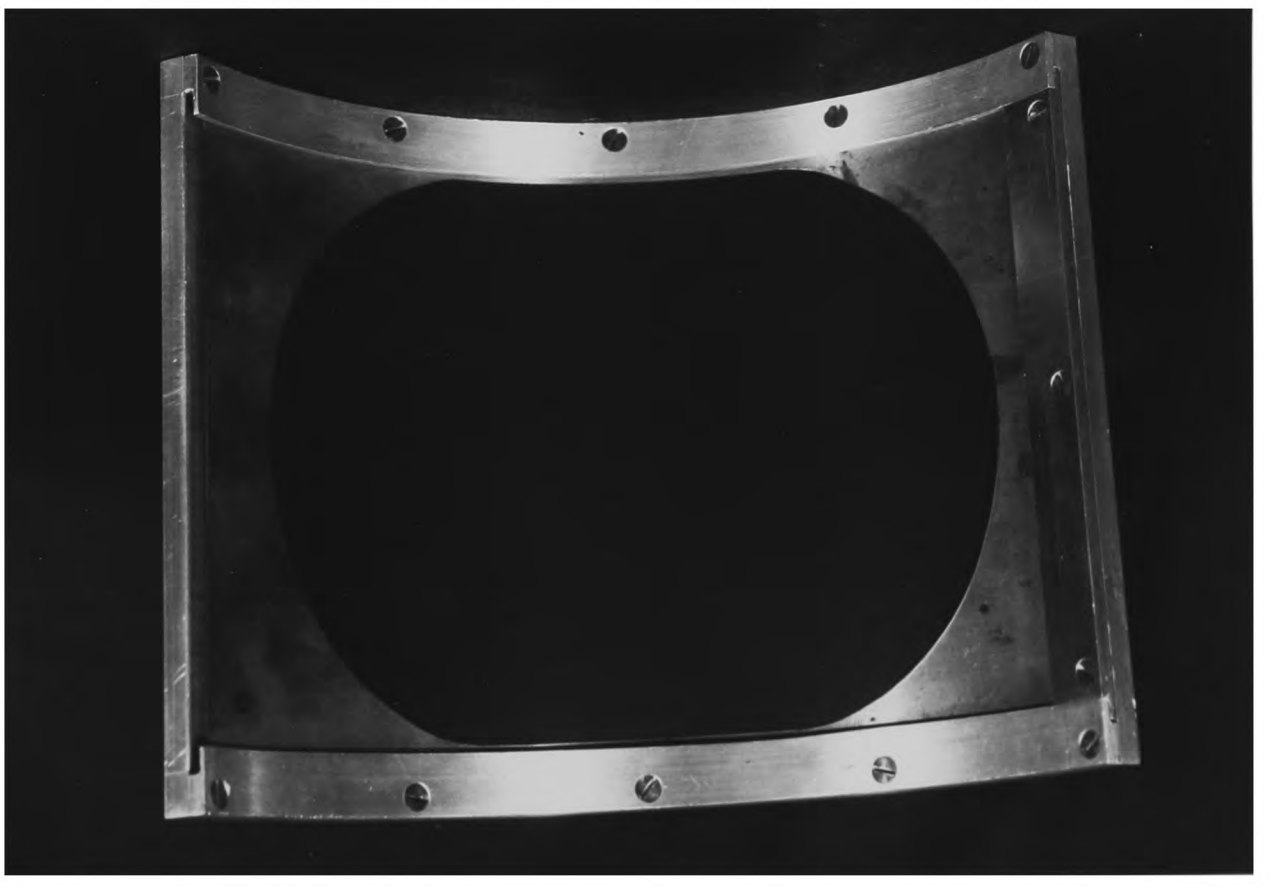
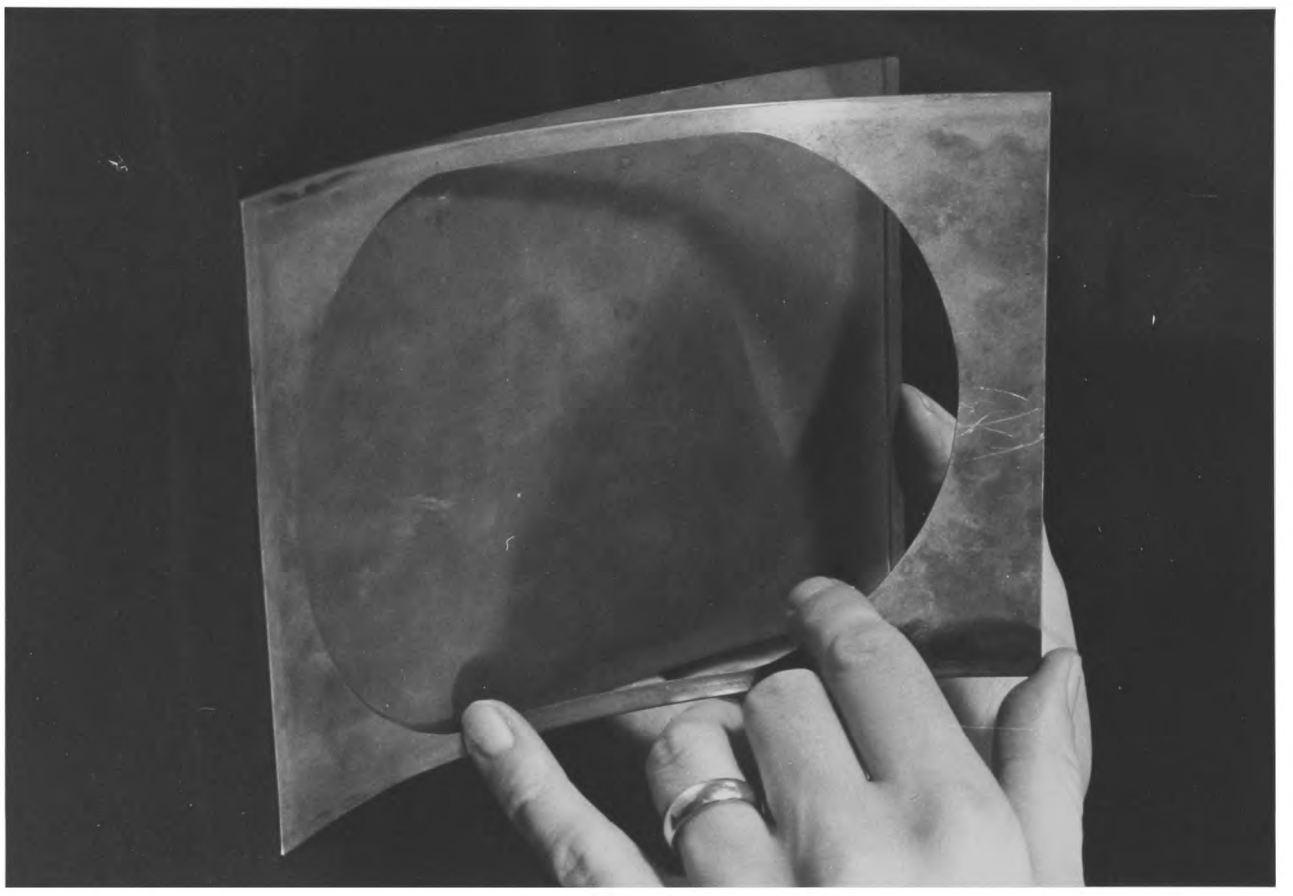
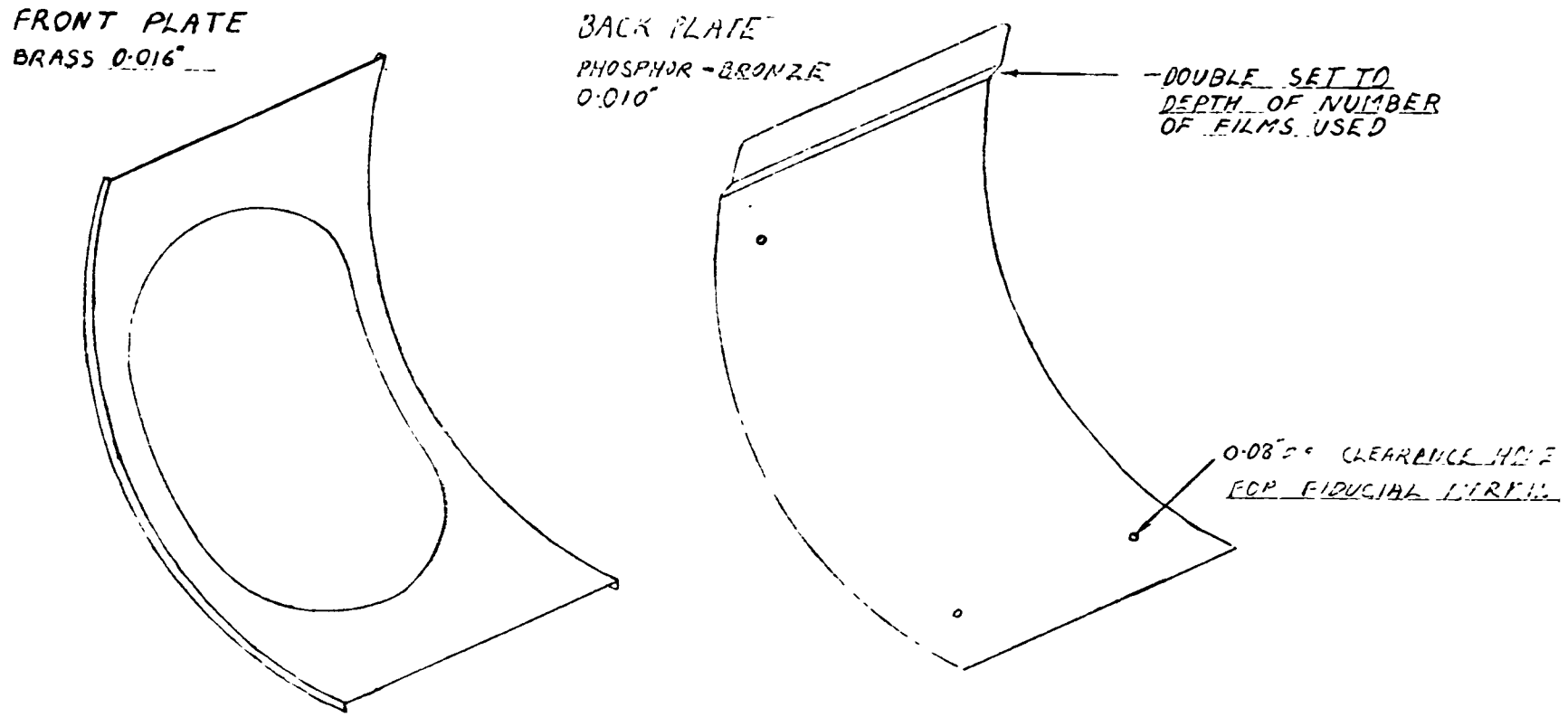


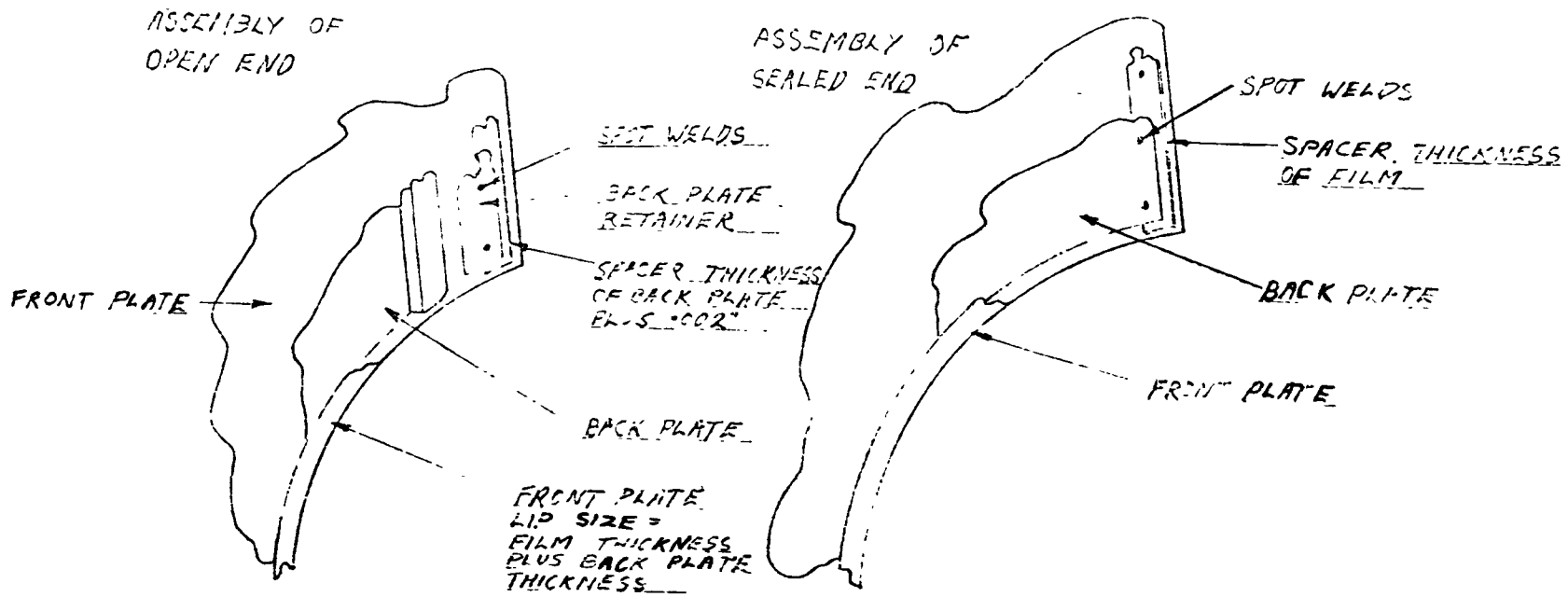
FIGURE 5.7

Cylindrical Cassette:-Film Holder.



FILM HOLDER FIGURE 5.8





thickness of the film holder + suitable clearance. The bottom step was step-machined to produce a suitable stop for the film holder, to position the film accurately relative to the beam axis (Figure 5.6). The film holder was designed to hold two 175 x 125 mm films, which could be digitized on the Optronics Photoscan P-1000 densitometer. The construction of the film holder is shown in detail in Figure 5.8.

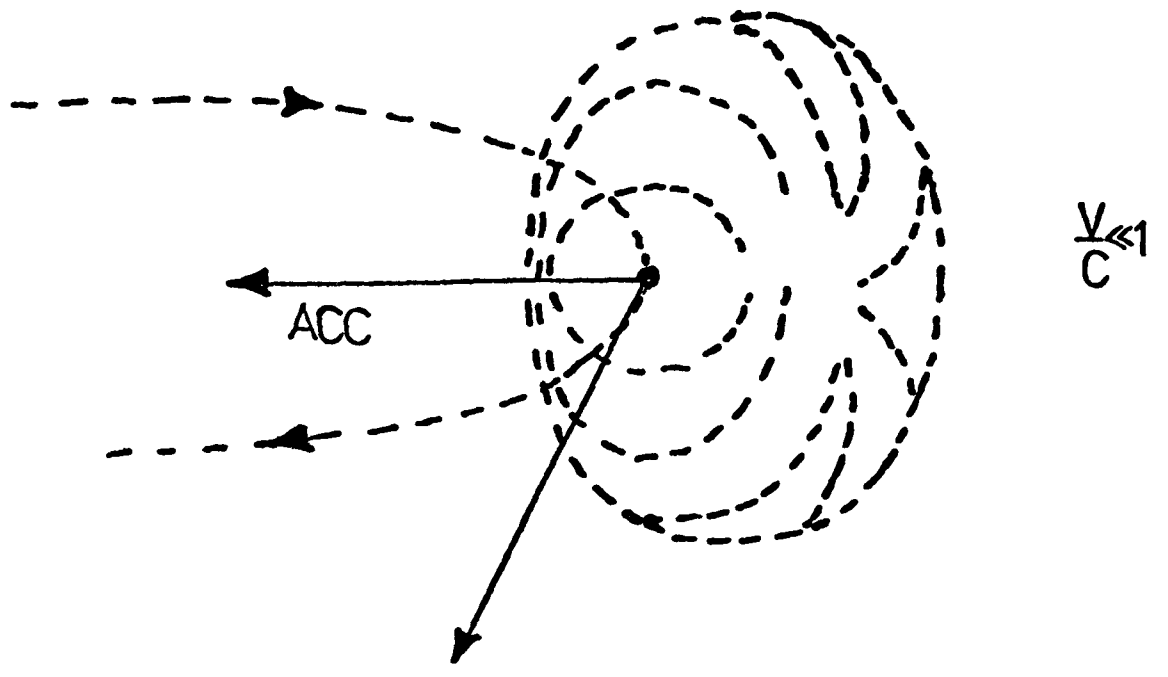
3) THE USE OF SYNCHROTRON RADIATION IN PROTEIN CRYSTALLOGRAPHY

Synchrotron radiation (S.R.) is the intense electromagnetic radiation emitted by orbiting electrons in synchrotrons and storage rings; essentially the same effect as the emission of radio waves from an aerial. However, the difference between a S.R. source and a conventional dipole lies in the speed with which the electrons are moving (Figure 5.9). An electron possesses a small rest energy (about 0.5 MeV) and for kinetic energies above the MeV range will be travelling close to the speed of light. In most high energy accelerators the charged particles are magnetically confined to a roughly circular path and they experience a radial acceleration towards the centre of the circular orbit. S.R. is the dominant mechanism of energy loss in a circular electron accelerator. Relativistic effects cause a sharp forward peaking of the radiation and a large increase in the total energy radiated. Thus at any instant of time the energy radiated is projected into a very small cone of $\sim 10^{-3}$ radians in the direction of electron motion. Secondly, the high velocity of the electron towards the observer causes a very large blue Doppler shift in frequency. For an electron accelerator in the GeV range, the peak power radiated will be at a frequency of about 10^{18} Hz (equivalent to a photon energy and wavelength of approximately 10 KeV and 1 Å) rather than at the fundamental radio frequency of the machine (say 10^{6} Hz).

FIGURE 5.9

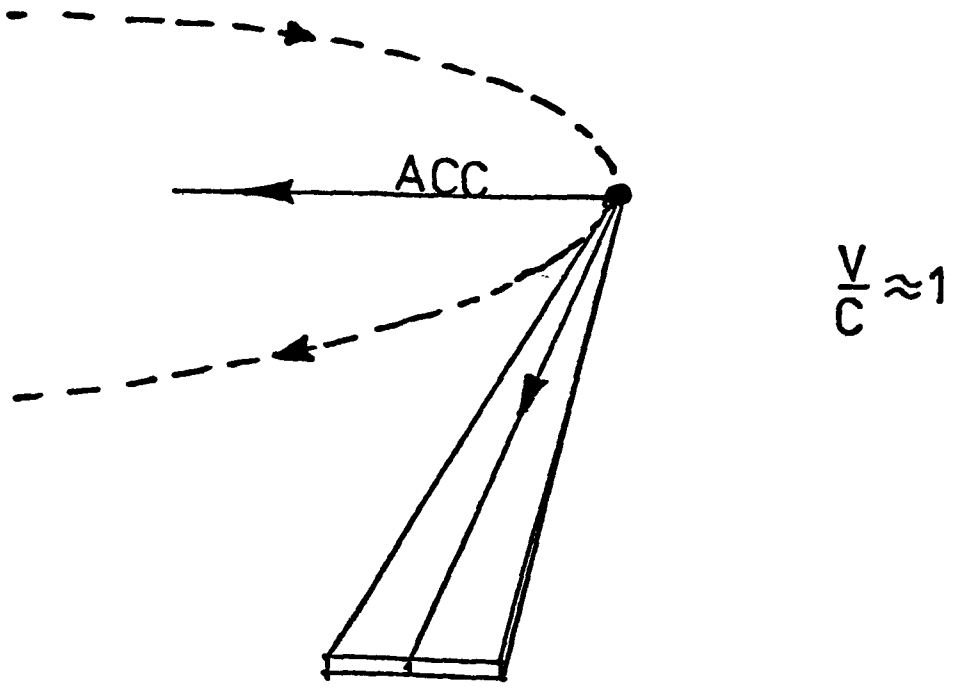
Radiation Emission Patterns of Electrons in Circular Motion.

a)



At low energy the radiation is emitted in a non-directional dipole manner.

b)



At high energy, when the velocity of the electron (v) is close to the velocity of light (c), and the energy is larger than its rest mass energy, relativistic effects cause a sharp forward peaking of the radiation. At the same time a large increase in the total energy radiated occurs. This is the case of synchrotron radiation. From Stuhrmann (1978).

The radiation emitted forms a continuous spectrum from the infra-red to the X-ray region. The properties of S.R. which are most interesting for protein crystallography can be summarised as follows (Stuhrmann, 1978):-

i) intense continuous spectrum from I.R. to hard X-rays,
ii) strong collimation in the instantaneous direction of flight (typically 0.1 mrad for X-rays),
iii) highly polarized radiation with the electric field vector in the orbital plane,
iv) elliptical polarization above and below the plane of orbit,
v) pronounced time structure which is a copy of the pulse structure of the electron beam (pulses as short as 100 psec),
vi) absolute calculability of all the properties of the source once the parameters of the particle beam are given.

The intensity of the radiation in the X-ray region is very high when compared with a conventional X-ray source. Quantitative comparison of intensity between synchrotron and conventional sources depends on the particular application. Because S.R. is highly collimated, its relative advantage over conventional sources increases as the specimen size decreases. Since the energy in conventional source is confined to a small number of well defined spectral lines, S.R. also has an advantage where a wider wavelength range is required. In summary, then, the main advantages of S.R., are the brilliant, highly focussed beam and the ability to tune the wavelength. In the collection of phosphorylase *b* data only the brilliance of the beam was exploited, giving greatly

reduced exposure times and increased crystal lifetime. So far the tuneability of the source has been neglected in these studies.

A general review of S.R. in biological structure research is given by Stuhrmann (1978) and its application to single crystal diffraction experiments has been described by a number of groups, *inter alia* Phillips *et al.*, (1976, 1977), Harmsen *et al.*, (1976). Previous studies have tended to concentrate on the theory of the use of S.R. in the study of biological molecules (Wyckoff (1973), Harrison (1973)) and the investigation of specialised techniques, such as wavelength variation for anomalous scattering observations (Phillips, *et al.*, 1977). The latter technique utilises the tuneability of the source to optimise the anomalous contributions to the scattering factors. When a single wavelength is used, anomalous scattering from heavy atoms produces the Bijvoet differences described in Chapter 1, Section 3c (F = f+f'+if"). Blow (1958) and Blow and Rossmann (1961) proposed the use of these differences to assist in the phasing of protein diffraction data. A number of structures have been solved from one isomorphous derivative plus anomalous differences, *inter alia* rubredoxin (Herriot *et al.*, 1970), ferrodoxin (Sieker *et al.*, 1977), and phosphoglycerate kinase (Blake and Evans, 1974). It had also been shown by Hoppe and Jakubowski (1975) that it was possible to obtain phases for the protein erythrocruorin using two wavelengths from two conventional X-ray tubes. The anomalous dispersion techniques are more easily accomplished with S.R. because the wavelength of the X-ray beam can be tuned around the wavelength of the K or L absorption edge for many heavy atoms. The real and imaginary components of the anomalous scattering contribution can be selectively altered while other contributions to the diffraction pattern remain the same. Phillips *et al.*, (1977) described the application of this technique to the Fe-containing protein rubredoxin. A series of precession photographs was

obtained at seven discrete wavelengths above and below the FeK absorption edge. The data showed intensity changes due to f' varying with wavelength and from Bijvoet differences. These changes were used to obtain information about the location of the Fe atom and about the phases of the reflections.

A report has been published by Harmsen *et al.* (1976) which compares the intensity of conventional rotating anode sources with the DESY synchrotron for high-angle diffraction experiments. They concluded that for crystals with cell dimensions below 100 Å the synchrotron source could be compared to collimated Ni-filtered CuKα radiation since the X-ray diffraction photographs had similar quality. Machine parameters of 6.5 GeV and 3.1 mA gave an absolute intensity of 20 x 10^6 photons/sec which was about three times weaker than the corresponding beam from the rotating anode tube (64 x 10^6 photons/sec). The intensities would be equal for the synchrotron running at its then maximum electron current of 10 mA. For crystals with cell dimensions above 100 Å, Harmsen *et al.* stated that the synchrotron source should be compared to a doubly-focussed beam from a conventional source, since the S.R. beam is convergent. They calculated that a double focussing quartz mirror arrangement on an Elliot GX13 rotating anode tube gave an absolute intensity of 4 x 10^6 photons/sec compared to 20 x 10^6 photons/sec for the synchrotron parameters used in the study and 65 x 10^6 photons/sec if the synchrotron ran at 10 mA. This corresponds to an intensity gain of 5 times and 16 times respectively. Phillips *et al.* (1976) reported a gain of at least 60 times in intensity over a sealed X-ray tube operating at 40 KV, 30 mA, with the SPEAR S.R. source (4.2 GeV, 100 mA). Experiments using the S.R. source at L.U.R.E. (see below) will be described by Wilson *et al.*, (1981). This paper describes studies on two other proteins (in addition to phosphorylase *b*) which represented different problems often

Table 5.1

Protein	Oscillation Angle	Resolution	Exposure time on GX6	Ex ti L.
Phosphorylase b	0.5°	2.0Å	13 hours	15 6
β-lactamase I	2.0°	2.5Å	16 hours	20
Troponin-C	2.0°	6.0Å	16 hours	20

Exposure times and crystal lifetimes for the proteins stu

* The second figure refers to the use of a larger diamet

† The lifetime of the β-lactamase I crystals is not gove source.

(from Wilson et al, 1981)

encountered in protein studies; β-lactamase I and Troponin C. For β-lactamase I only small crystals were available, and high resolution data collection was virtually impossible with conventional X-ray sources. For Troponin C, the unit cell was large and the crystals small, and data collection to even 6 Å was impractical with a rotating anode source. The improvements using S.R. in terms of reduced exposure time and effective crystal lifetime reported by Wilson *et al.*, (1981) are given in Table 5.1. For all these proteins a reduction in exposure times of approximately 50 times compared to a rotating anode source was obtained.

4) EXPERIMENTAL CONDITIONS

The collection of the 2 Å data for phosphorylase *b* was carried out at the Laboratoire pour l'Utilisation du Rayonnement Electromagnetique (L.U.R.E.), Orsay, France, using the S.R. source of the D.C.I. storage ring. Other synchrotron radiation sources (synchrotrons or storage rings) exist at Stanford USA (SPEAR), Hamburg, West Germany (DESY and DORIS), and Novosibirsk, USSR (VEPP3) and a dedicated source is presently under construction at Darsebury, UK.

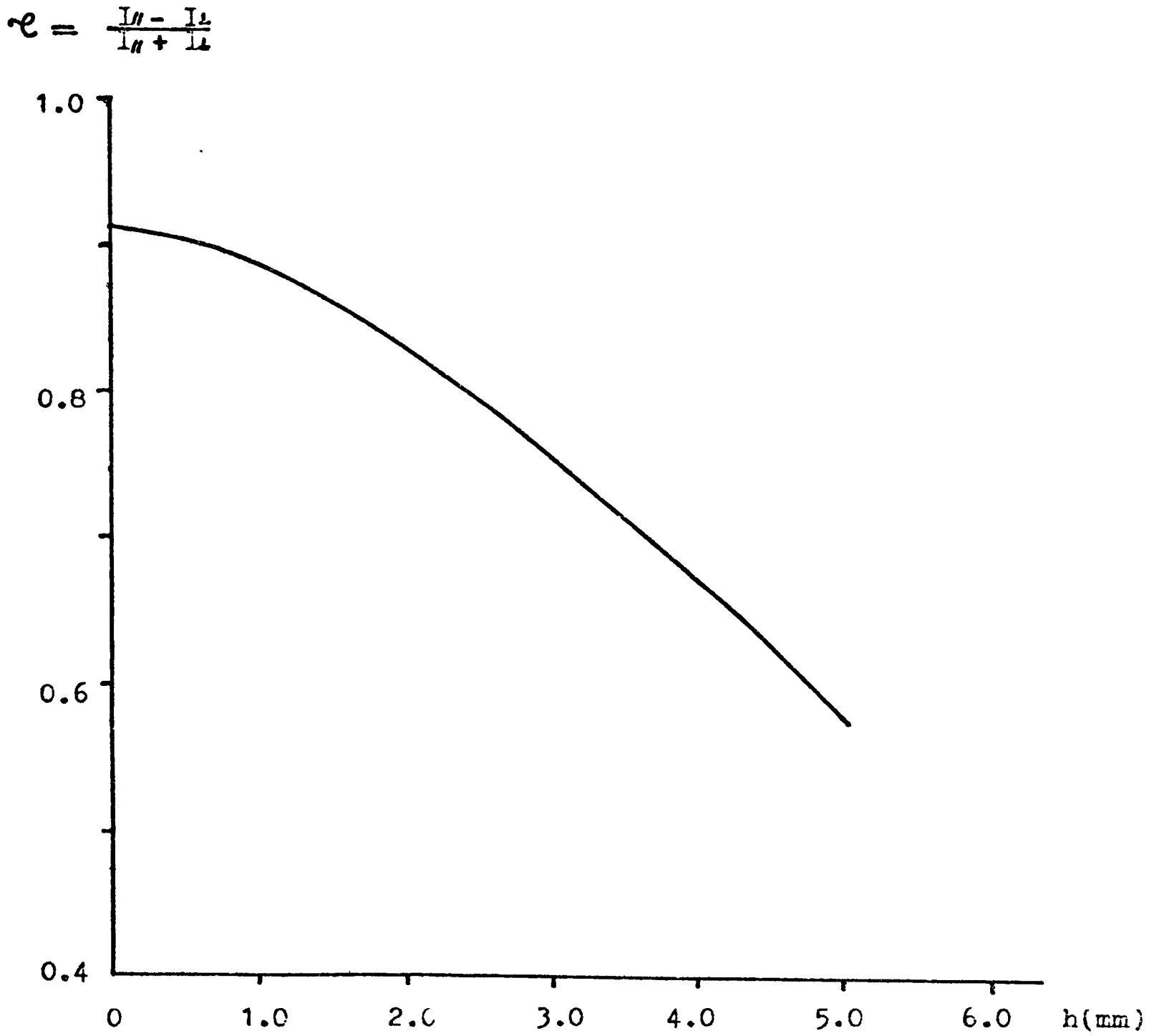
The DCI storage ring was designed to achieve a high luminosity in colliding electron beam experiments of high energy physics. With a stored beam energy of ~1.87 GeV, a beam current of 200-500 mA and a short magnetic radius (4 m), DCI is also a powerful S.R. source. The radiation spectrum is a continuum characterized by the critical wavelength λ_c = 3.4 Å at 1.87 GeV. The emission is peaked in the 0.8-4 Å region, which is the most useful range for X-ray diffraction experiments. The flux of hard radiations (λ_c<0.4 Å) is low, so that shielding problems are reduced. The S.R. from 18 mrads of orbit is shared between 5 stations at present (Fourme *et al.* 1977). After injection of electrons into the storage ring and

subsequent energy ramping, the electrons circulate in the ring emitting useful S.R. The electron current subsequently decays at variable rates. After injection the current is generally around 200 mA and reinjection is carried out when it drops below about 80 mA. The time between injections varies between 2-12 hours, averaging at about 7 hours. Typical operating conditions during the dedicated shifts are 1.72 GeV and 120 mA in one bunch, giving a total flux of approximately 2×10^{10} photons.sec^{-1}.mm^{-1} through a 0.3 mm diameter collimator at 1.54 Å wavelength. Details of the experimental conditions and the single crystal protein crystallography station are described by Fourme and Kahn (1981).

S.R. emitted in the plane of the orbit is linearly polarized in this plane. In general, at any out of plane angle, the radiation is elliptically polarized. Figure 5.10 shows how the difference in intensity of radiation polarized parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) to the plane of orbit varies with distance from the plane of orbit, after monochromatisation. The polarization correction to be applied to the X-ray diffraction data is given in Fourme and Kahn (1981) and Chapter 3. Several factors have to be introduced into the calculation to describe the actual source; the distribution of particles in the bunch and the average particle trajectories as well as the angular spread in these trajectories.

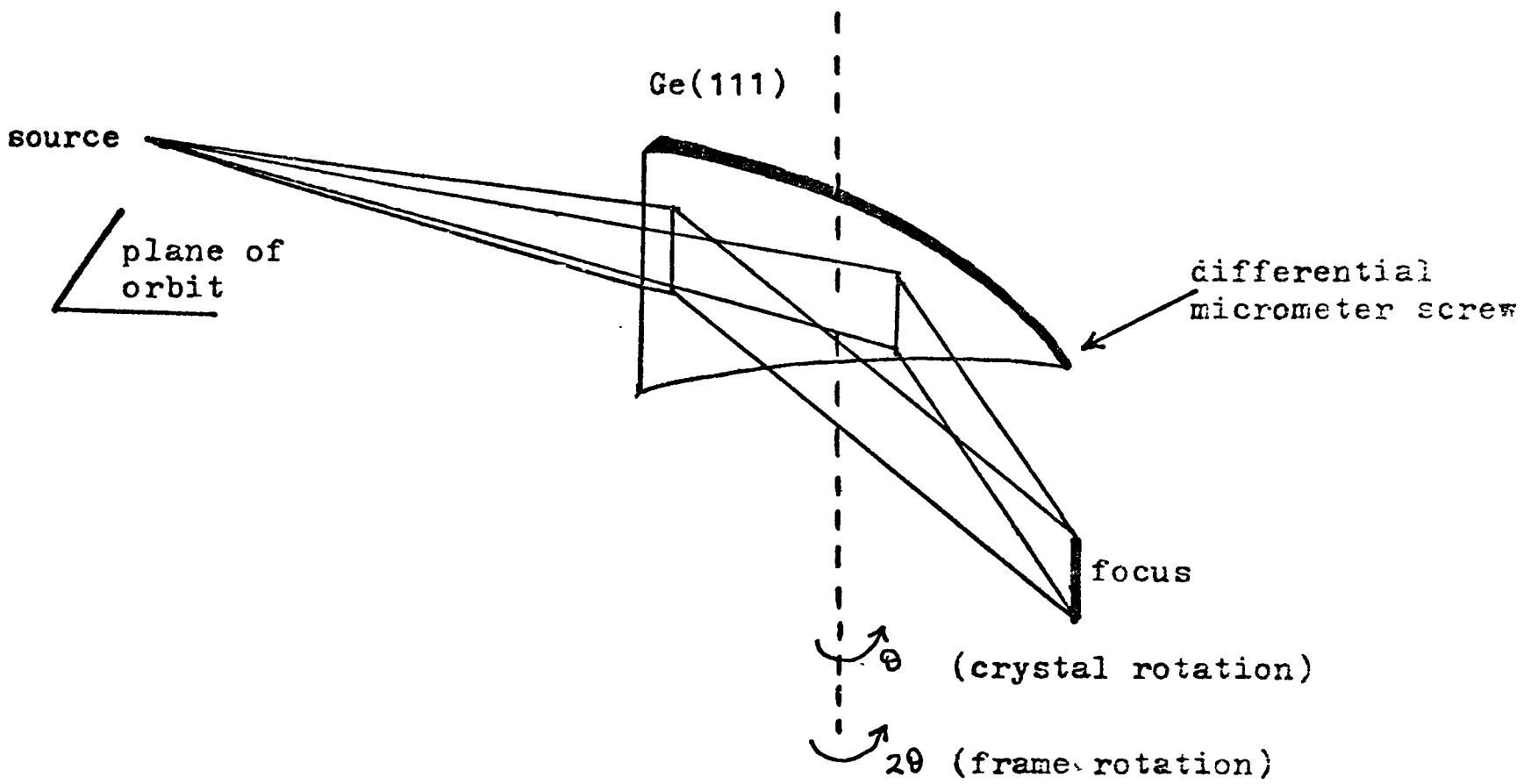
The design of the monochromator is described by Lemonnier *et al.*, (1978) and Fourme and Kahn (1981). A germanium beam of constant thickness is shaped to a triangle and mounted as a cantilever. The base of the triangle is fixed and the free end can be displaced by a differential micrometer screw so that a cylindrical curvature is obtained. Horizontal focussing is thus obtained by reflection from the curved crystal (Figure 5.11). The focus is a narrow vertical line and the focus distance can be adjusted. In the horizontal plane the beam is convergent and in the vertical

FIGURE 5.10



Rate of polarization ($\mathcal{C}$) plotted versus h (distance to the average plane of orbit) after the monochromator, at the sample location (from Lemonnier et al.,1978).

FIGURE 5.11



Principle of the curved crystal monochromator (from Lemonnier et al.,1978).

plane the natural small divergence of the beam (~0.3 mrad) is preserved. The wavelength can be tuned by rotating the crystal (θ angle). With a crystal asymmetrically cut to (111) planes, the wavelength is tuneable from 0.7 to 3.5 Å and the higher order harmonics at the detector level are negligible for $\lambda<1.6$ Å (Lemonnier *et al.*, 1978). An Enraf-Nonius oscillation camera is mounted on a heavy frame with 2θ rotation. The standard control unit has been modified to provide the more rapid rotations required for work with S.R. The beam current is monitored and displayed on a digital readout, which allows the user to adjust exposure times accordingly. The camera is aligned by visual observation of the direct beam, after collimation, on a fluorescent screen, and manual adjustment of the vertical and horizontal positioning of the camera. Only rarely does the camera require realignment after a new injection of electrons; usually the new beam emits radiation directly collinear with the previous one.

5) COLLECTION OF 2 Å DATA FOR PHOSPHORYLASE *b*

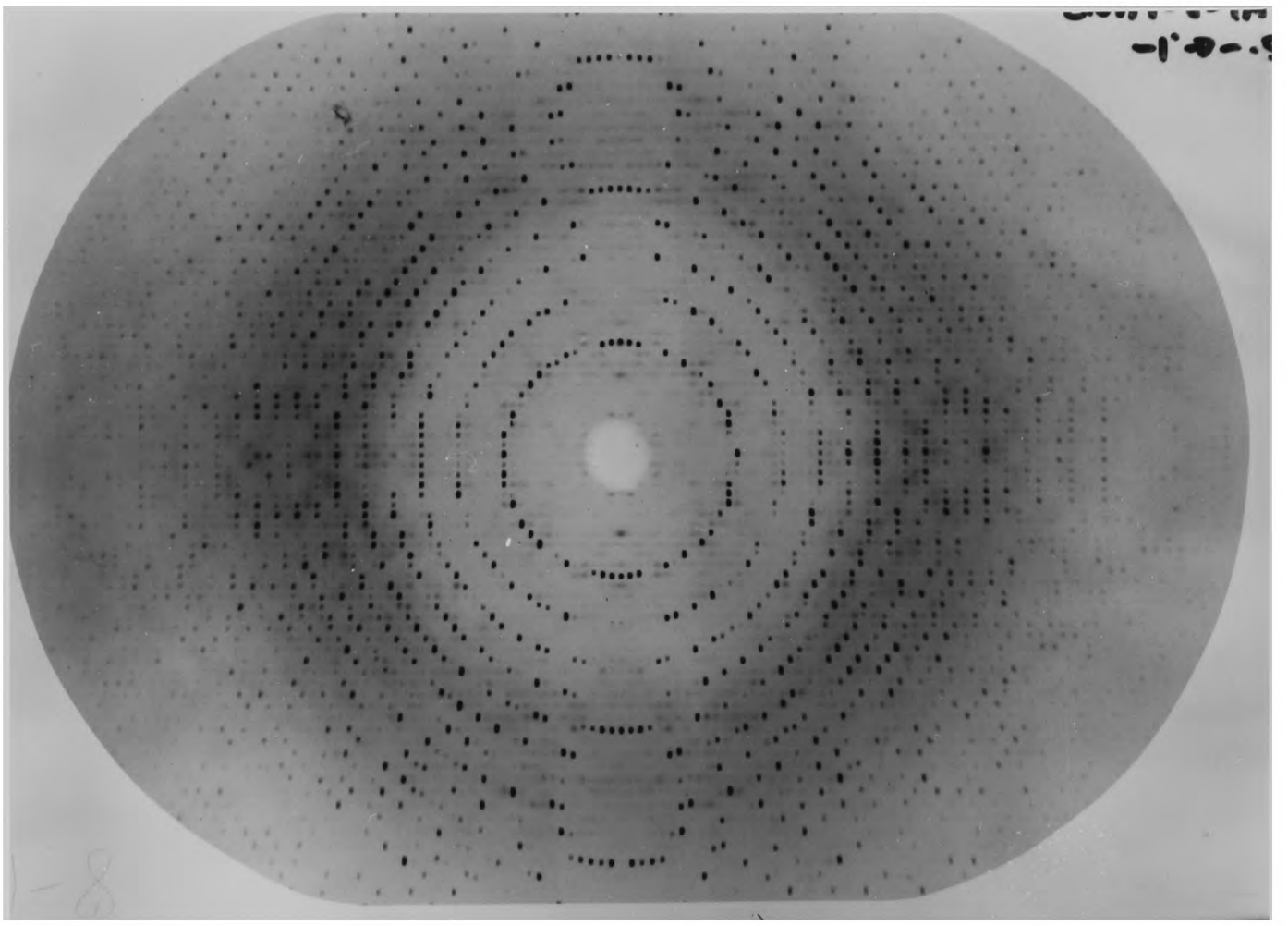
Collection of data to 2 Å resolution was begun in Oxford on an Elliot GX6 rotating anode (run at 40 KV, 40 mA with graphite monochromated radiation). An oscillation angle of 0.5° was chosen as the best compromise between maximising the angle used whilst minimising the number of overlapping reflections (Figure 3.7). A 0.5° oscillation photograph with data extending to 2 Å required 13 hours exposure on the GX6 rotating anode with a 0.3 mm diameter collimator, giving a beam size smaller than the crystal in all dimensions. A comparable photograph at L.U.R.E. required an exposure time of 15 minutes (with machine parameters of 1.72 GeV, 120 mA beam current, 1.55 Å wavelength) with a 0.3 mm collimator. This is an improvement in exposure time of approximately 50 times, comparable to the estimated difference in flux between Oxford and

L.U.R.E. A much greater gain proved possible with the large (1.5 x 0.8 x 0.8 mm^3) phosphorylase *b* crystals, due to the highly convergent S.R. beam from the germanium monochromator. The convergent beam significantly reduced the spot size for a given collimator, and hence allowed a collimator of 1.0 mm diameter to be used without increasing the proportion of overlapping reflections. This resulted in a significantly larger crystal volume being irradiated and exposure times dropped to about 6 minutes on average, a reduction of 130 times compared to the GX6 rotating anode. A gain in exploitation of crystal volume could probably also be achieved by the use of a focussing monochromator on a conventional rotating anode. A wavelength of 1.55 Å for the S.R. was used, to keep all parameters as close as possible to those used in the collection of 3 Å data for phosphorylase *b*. Photographs from L.U.R.E. and Oxford are shown in Figure 5.12. It is clear from this figure that not only could photographs be recorded with a 130-fold reduction in exposure times at L.U.R.E., but also that the data on the L.U.R.E. photographs extends to higher resolution.

Another improvement was obtained with regard to crystal lifetime. Although quantitative estimates of radiation damage were difficult, since individual crystals vary greatly, an improvement of at least 5 times in effective crystal lifetime was noted at L.U.R.E. Only one 0.5° oscillation photograph with data extending to 2 Å could be recorded from each crystal in 13 hours on the rotating anode source before unacceptable radiation damage occurred. At least five such exposures could be obtained at L.U.R.E., although the high resolution (2 Å) data disappeared in general after about 1½ hours total exposure. Macroscopic cracks were also observed to occur in the cylindrical region of the crystal which had been irradiated, at the same time as the disappearance of the 2 Å data. It is unclear at present whether these cracks were due to primary

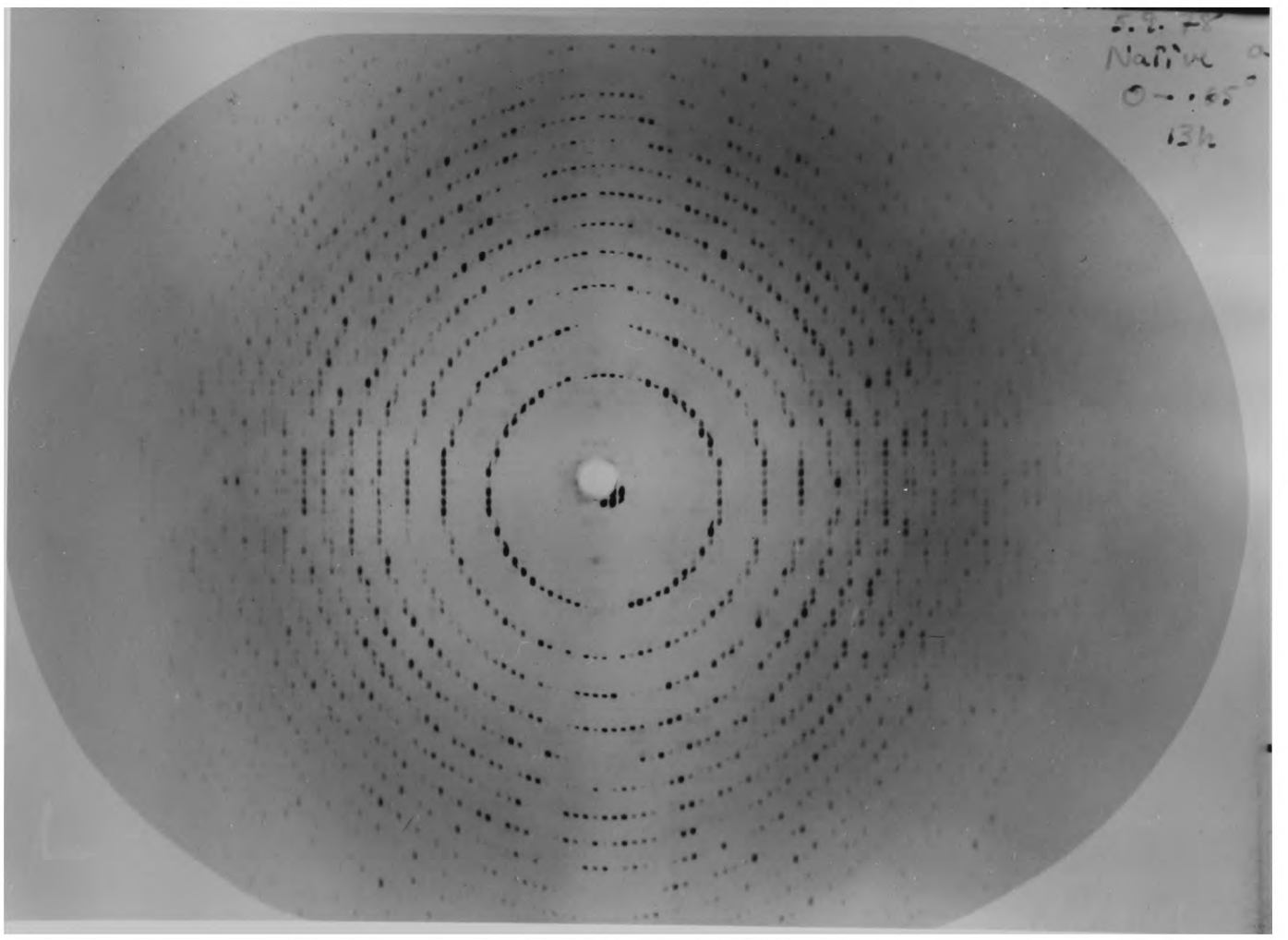
FIGURE 5.12

Rotation Photographs of Phosphorylase b taken with the Cylindrical Cassettes, at L.U.R.E. and Oxford.



L.U.R.E.

6 minutes exposure.



OXFORD (Elliot GX6).

13 hours exposure.

or secondary effects of X-ray absorption, or to the heating effect of the beam. The latter possibility will be investigated in the future by flowing air at room temperature over the crystal during the collection of data.

However, the fact that greater diffracted intensities could be obtained with a higher flux for phosphorylase *b* crystals suggests that secondary effects may be more important in the mechanism of radiation damage. Protein crystals commonly contain about 50% solvent and the effect of the ionizing X-radiation on water, resulting in the production of free radicals, is thought to be the most significant primary step in the radiation damage process for protein crystals. A chain reaction initiated by free radical formation may account for the onset of the radiation damage effects described above after 1½ hours exposure.

Data collection at L.U.R.E. used 8 x 24 hours sessions, during which the laboratory was dedicated to the production of S.R.

This allowed the collection of the complete native 2 Å data set, comprising 45° of data on crystals mounted about c* and 30° of data on crystals mounted about a* to fill in the 'blind region'. 45° of data on crystals soaked in 1 m*M* EMTS and mounted about the c* axis were also collected; these data generally extended little beyond 2.3 Å (Figure 5.13). In all roughly 30 crystals were used. The data were collected by myself, E.A. Stura and J.A. Jenkins. In theory this amount of data collection would have required several months on a conventional source, and the use of some 250 crystals. In practice the task would have been impossible due to the inability of collecting more than one photograph per crystal, since the oscillation method is more effective if more than one exposure can be recorded per crystal, enabling partially recorded terms to be summed and used. A comparison of the recording conditions using the rotating anode and S.R. sources

FIGURE 5.13

Rotation Photograph of Phosphorylase b crystal soaked in 1mM EMTS for 2 days, taken at L.U.R.E.

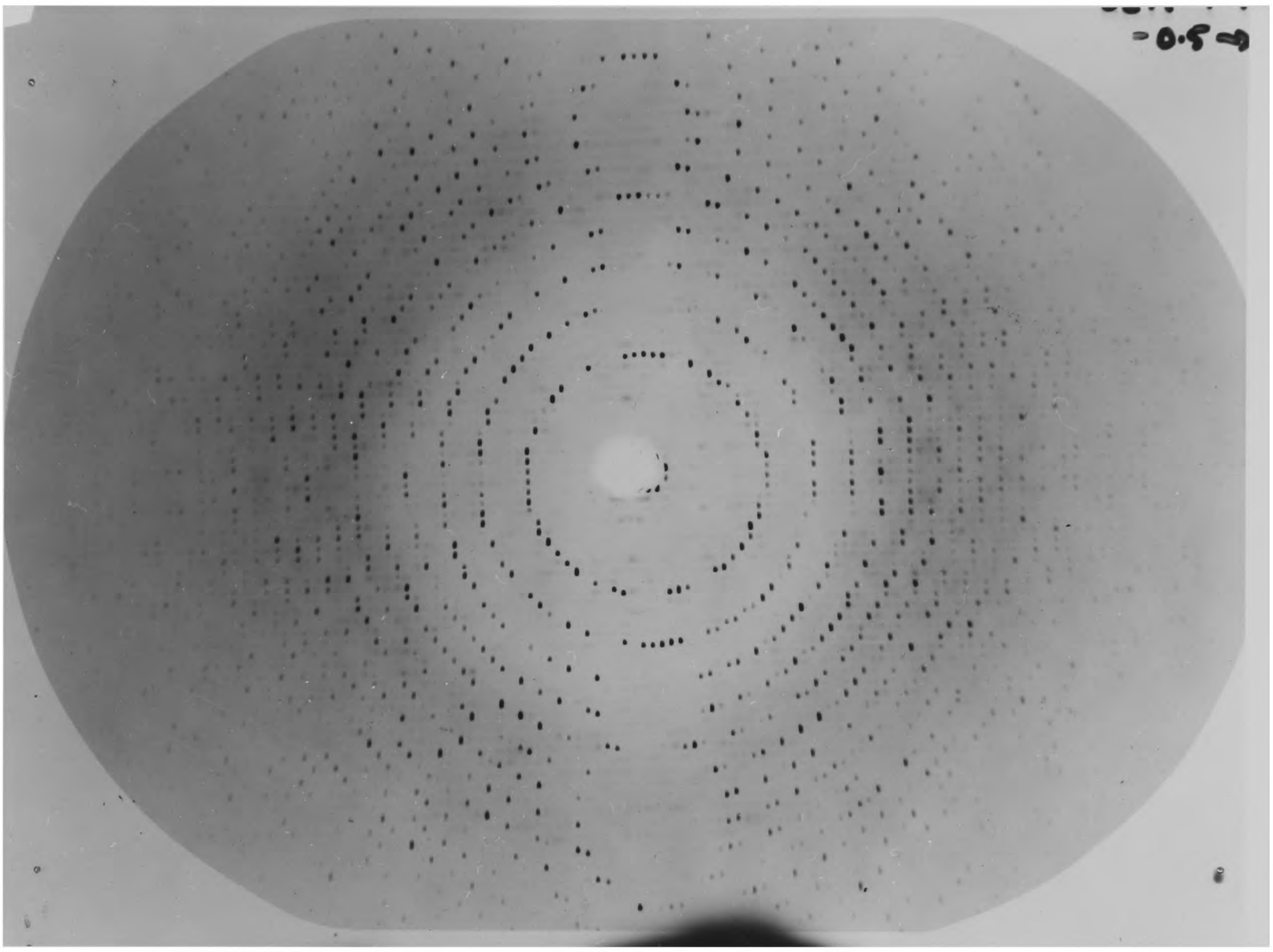


TABLE 5.2

Recording Conditions

	3 Å data	2 Å data
Camera	: Nonius screenless oscillation	Nonius screenless oscillation
X-ray source	: Elliot GX6 rotating anode at 1.6 KW	DCI storage ring at LURE : 1.72 GeV, 200 mA
Radiation	: CuKα, λ = 1.5418 Å, Ni filter	Synchrotron radiation : λ ~ 1.55 Å
Collimation	: Standard, 0.3 mm beam at crystal	focussing monochromator and standard collimation, 1.0 mm beam at crystal
Detector	: Ilford Industrial G x-ray film	Kodirex x-ray film
Cassettes	: Flat plate, maximum radius 60 mm	cylindrical cassette : maximum radius 80 mm
Crystal to film distance	: 96.0 mm	96.0 mm
Maximum Resolution	: 2.80 Å	1.9 Å
Total ϕ for data set	: 45 degrees (4 equivalents)	45 degrees (4 equivalents)
$\Delta\phi$ for each exposure	: 1.25 degrees	0.5 degrees
Number of films per data set	: 37 two film packs	91 two film packs
Exposure time	: 7500 sec/deg, 2.6 hours	500-1000 sec/deg, 4-8 minutes
Total time	: for 80,000 reflections: 96.4 hours	for 280,000 reflections: 9.6 hours

is shown in Table 5.2.

Table 5.1 shows the gains in intensity and crystal lifetime for phosphorylase *b* and the other proteins studied by Wilson *et al.* (1981). The large reduction in exposure time for phosphorylase *b* (130 times) was rather exceptional in that large crystals were available and a larger diameter collimator could be used. However, for all the proteins studied, a reduction in exposure times of at least 50 times was observed. This is in marked contrast to the conclusions of Harmsen *et al.* (1976), based on their early experiments at DESY, that the S.R. source gave no appreciable advantage for normal sized crystals (with cell dimensions around and below 100 Å), and an intensity gain of up to 16 times for crystals with larger cell dimensions. However, the intensity gain is comparable to that reported by Phillips *et al.* (1976), who found a gain of 60 times using the S.R. source at SPEAR.

A further advantage of the focussing geometry reported by Wilson *et al.* was the possibility of resolving reflections which lay on closely spaced reciprocal lattices for β-lactamase I.

The S.R. source at L.U.R.E. has made possible the investigatior of the crystal structures of biological molecules which would have remained inaccessible before the availability of such a brilliant X-ray source.

6) DIFFUSE SCATTERING

From Figure 5.12 it can be seen that considerable subsidiary scattering occurs. A 'ghost' lattice of relatively diffuse spots is clearly visible between and around the Bragg diffraction maxima. Such diffuse scattering generally occurs when random irregularities are introduced into the crystal lattice. These may be due to imperfections such as dislocations, vacancies, isomorphous substitution, atomic disorder, or due to the intrinsic constant disorder

caused by thermal vibrations of the atoms. Such effects have been discussed in detail by James (1948), Amorós and Amorós (1968) and Willis and Pryor (1975). The effect of thermal vibrations on the scattering of X-rays from a crystal was first investigated by Debye (1913, 1914). In a real crystal the molecules or atoms associated with the lattice points are linked to one another by interatomic forces. The displacements of one atom must affect those of its neighbours, and this interdependence of the lattice points has an effect on the scattered intensity. Broad and diffuse maxima are produced in the directions of the lattice spectra. The actual motions of the lattice may be thought of as a large number (3N, where N is the number of atoms) of lattice waves, and to each of these there corresponds a pair of maxima in the scattering function. Taken together, these independent maxima produce a continuous distribution about each reciprocal lattice point (James, 1948).

It remains necessary to investigate the nature of the diffuse scattering observed with phosphorylase *b*, to identify it as due to thermal diffuse scattering (TDS) or to some other cause. James (1948) describes the total scattered intensity by the equation:-

$$\bar{I}(S/\lambda) \approx |f_0|^2 e^{-2M} I_0(S/\lambda) + I_2(S/\lambda)$$

where $\bar{I}$ is the average scattered intensity, I_0 is the intensity scattered by an undisturbed lattice of point atoms, I_2 is the thermal diffuse scattered intensity, and $|f_0|$ is the atomic scattering factor for the atom at rest. Both components of the scattered intensity are modified by the factor e^{-2M}, which is the normal 'temperature factor' (*i.e*: $M \propto T.\sin^2\theta/\lambda^2$).

However,

$$I_2 \propto |f_0|^2 e^{-2M}.T$$

Hence, the thermal diffuse scatter is proportional to the temperature (T), and as the temperature decreases the diffuse scatter also decreases with respect to Bragg angle. Assuming this linear dependence on temperature, cooling the crystals to -50 °C should give a reduction of 25% in TDS:-

$$\frac{I(223\ ^{\circ}K)}{I(298\ ^{\circ}K)} = 0.75$$

Hence, a reduction of the diffuse scattering by this amount on lowering the temperature of phosphorylase *b* crystals should positively identify it as TDS. Alternatively, if the diffuse scattering is due to static disorder (due, for instance, to different molecules in the crystal adopting different conformations or occupying different positions and orientations in the various unit cells), the reduction in temperature should have much less effect.

If the diffuse scattering is positively identified as TDS, it remains unclear whether useful structural information concerning the modes of molecular vibration could be obtained from it. There are considerable problems in such an analysis even for small molecules (James, 1948). However, if such information could be extracted, it might prove useful for correlation with patterns of temperature factors in macromolecules. Frauenfelder *et al.* (1979) have shown that X-ray diffraction intensities measured at different temperatures, coupled with crystallographic refinement, can yield information about the mean square displacements and conformational potentials of protein molecules. Artymiuk *et al.*, (1979) and Sternberg *et al.*, (1979) have also analysed the temperature factors obtained from crystallographic refinements to make deductions about atomic displacements in protein molecules.

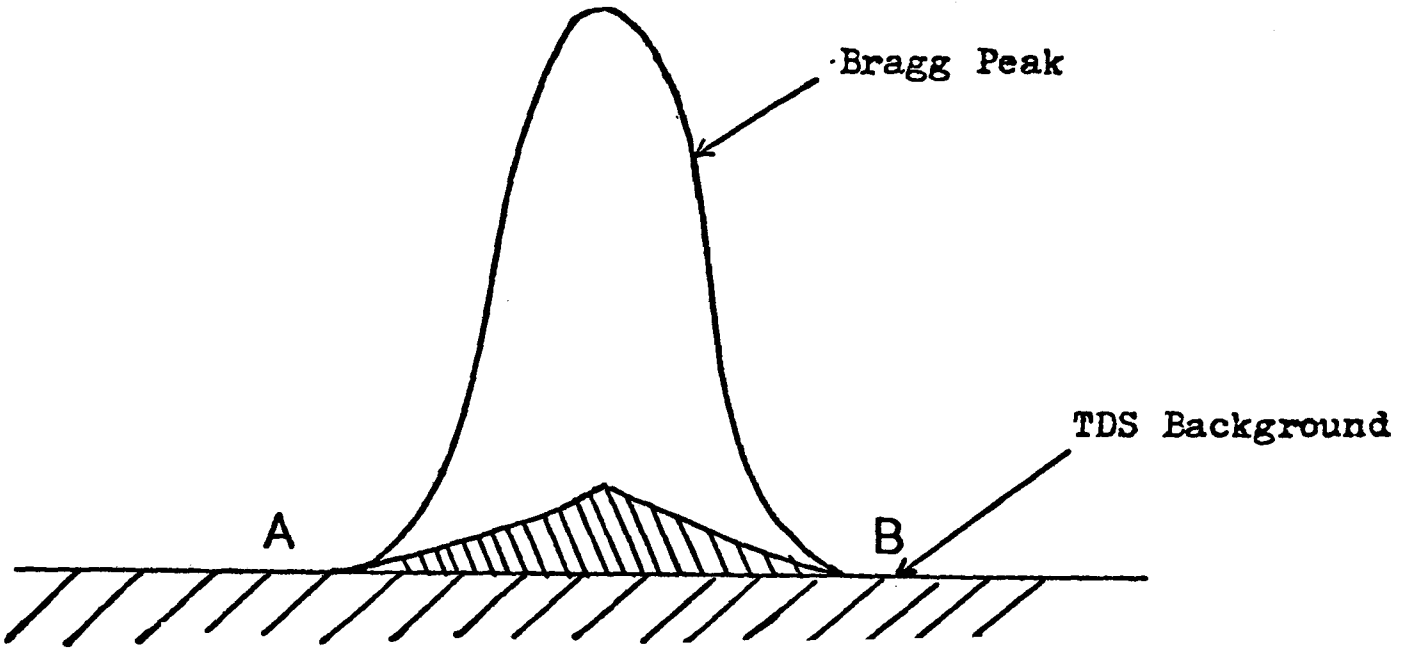
The diffuse scattering may also have a negative influence

during data reduction. It contributes to the background intensity around the Bragg reflection (Figure 5.14), in a manner for which it may be difficult to apply satisfactory correction. This is discussed further in Chapter 6. Hence a reduction in the diffuse scattering on lowering the temperature should also give an improvement in the quality of the Bragg intensity data.

The collection of data at low temperatures would also have advantages in the study of enzyme-substrate complexes. The high rate of data collection using S.R. would reduce the necessary lifetime of the complex and hence the time for which a stable low temperature need be maintained. It is intended to carry out such experiments for phosphorylase *b* at L.U.R.E. in the range -50 °C to -30 °C, to investigate the nature of the diffuse scattering and to study enzyme-substrate complexes.

FIGURE 5.14

The effect of Thermal Diffuse Scattering on the measurement of Bragg intensities.



If the integrated Bragg intensity is taken to be the area lying above the straight line AB, the intensity will be over-estimated by the TDS contribution shown by the shaded region above AB. (From Willis and Pryor, 1975).

CHAPTER 6

PROCESSING OF THE NATIVE 2 Å DATA

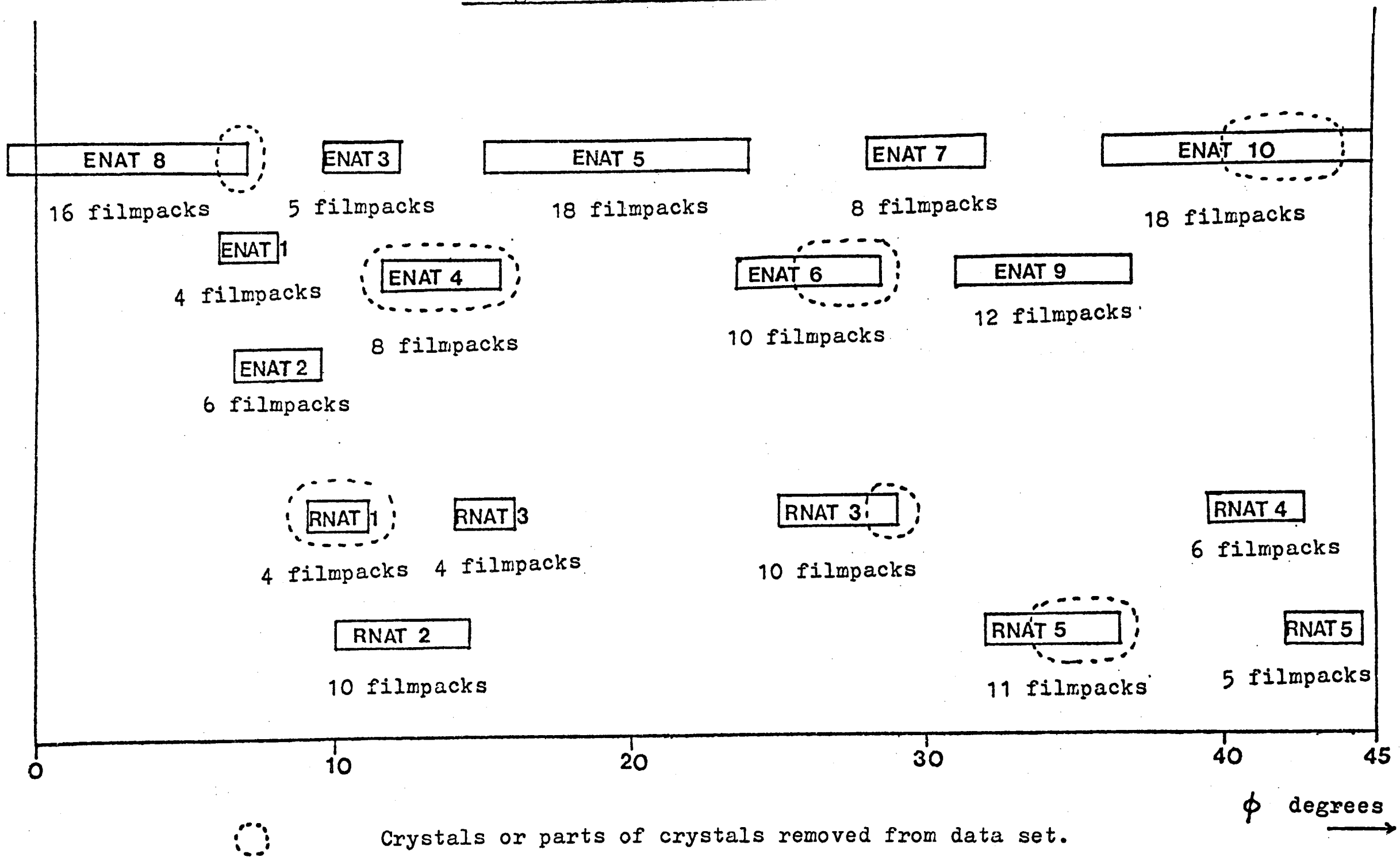
This chapter describes the processing and reduction of the native 2 Å c* data, from the measured intensities to the set of unique structure factor amplitudes. The 2 Å native a* data have been processed separately by E.A. Stura and this will be described elsewhere (Stura, 1981). The data were processed essentially by the same procedure as that described in Chapters 3 and 4, with appropriate modifications to accommodate the cylindrical cassettes. The complete (45°) c* data set collected is shown schematically in Figure 6.1. It was collected from 15 crystals and comprised 205 film packs in total. The data were collected in two stages; RNAT being the recollection of weak ENAT data.

1) DETERMINATION OF CRYSTAL ORIENTATION MATRICES

Determination of crystal orientation was carried out using Segment 1 of OSCAR and 'still' orientation photographs recorded using the standard flat plate cassettes. The crystal to film distance used was such that, when the cylindrical cassettes were in place, a crystal to film distance of approximately 96 mm was obtained. The distance between the film surfaces in the cylindrical and flat plate cassettes was 13.157 mm, which gave a crystal to film distance of 109 mm for the flat plate cassettes. This meant that data to a theoretical maximum resolution of about 3.1 Å would be recorded on the orientation photographs (with λ = 1.54 Å). A factor of 13.157 mm was subtracted from the refined crystal to film distance obtained from the Segment 1 orientation using the flat plate photographs. Orientation photographs were taken at 0° and 90° and the coordinates of about 60 reflections from each were measured by hand and the missetting angles estimated

FIGURE 6.1

Range of 2 Å c* native data.



The orientation matrix was then refined by Segment 1 of OSCAR. The procedure adopted was first to refine the crystal to film distance and missetting angles and then, keeping these fixed, to refine the cell parameters. If convergence of the crystal to film distance was not obtained within about 10 cycles, it would be fixed at its original value and only the missetting angles refined. The procedure appeared to be quite sensitive to the initial estimates of the missetting angles. When good estimates of these were obtained, convergence usually occurred within 10 cycles of refinement. Orientation matrices for all the crystals were produced in this manner. However, the use of flat plate photographs, with data to 3 Å resolution, to refine orientation matrices for the 2 Å cylindrical data photographs, may give rise to systematic errors. This is discussed further in Section 6.

2) PREDICTION OF REFLECTIONS IN AN OSCILLATION RANGE

The refined crystal to film distance and orientation matrices obtained from Segment 1 of OSCAR were then used to predict the reflections occurring on the data films. Segment 2 of OSCAR produced a list of the coordinates of predicted reflections for each film, which were plotted by the CALCOMP plotter on the ICL 1906A or 2980 computers. Fully recorded reflections were plotted in black, those partially recorded at the start of a rotation range in red, and those partially recorded at the end in green. Each plot was carefully checked to see that each reflection on the corresponding photograph was represented by a spot on the plot, particularly for the high angle data. This was often difficult to see, on account of the very close packing of spots on the photographs, and this may have been another source of error in the measurement of the

integrated intensities. If the agreement was obviously bad, especially for partially recorded reflections at the edges of the lunes, an attempt was made to obtain improved orientation matrices. This was done either by rerunning Segment 1 of OSCAR, or, more often by "manual" adjustments to the crystal to film distance or missetting angles. This was often a time-consuming process of trial-and-error. If the spots on a photograph were displaced towards or away from the centre, relative to the spots on the plots, then a new crystal to film distance was estimated from the radial displacement between the spots.

Similarly, if the centres of the main zone of the photograph and plot were displaced by an amount which was too small to measure, then new estimates of the missetting angles ($\Delta x, \Delta y$ or Δz) had to be obtained. This was usually done by finding partially recorded reflections which were missing from the plot, or fully recorded reflections labelled as partially recorded (or vice-versa) and estimating the error in the degree of partial recording.

Since the errors were usually small, several estimates were made and the correct value determined by a trial-and-error comparison with the plots (Stura, 1981). When new estimates of the crystal-to-film distance or missetting angles were obtained, Segment 1 of OSCAR could be rerun, or new orthogonalisation and rotation matrices calculated by the programme MATIS (see Chapter 3). Large errors in the missetting angles between consecutive photographs were often observed when the crystal had slipped in the capilliary tube. For some crystals, the slippage was so bad that new matrices were needed for each successive photograph. This would have meant that it would not not have been possible to use partially recorded reflections on sequential photographs, and it was decided to remove these

crystals from the data set. In total it was not possible to obtain satisfactory plots, using these methods, for crystals ENAT-4, RNAT-1 and parts of crystals ENAT-6, ENAT-10, RNAT-3 and RNAT-5 (see Figure 6.1). However, these data were not unique, since the oscillation ranges which they covered were also covered by other crystals.

Another variable which can affect the plot of predicted reflections is the crystal reflecting range, γ.

The value of γ was originally estimated empirically, as described in Chapter 3. Plots with slightly different values of γ were then produced for a typical crystal by Segment 2 of OSCAR, and compared to the data photographs. If partially recorded reflections at both the start and end of the rotation range were missing from either the plot or the corresponding photograph, then this would imply an incorrect estimate of γ. Plots which were judged to be acceptable were obtained for all the data photographs using a γ of 0.25.

However, as described above, it was often difficult to distinguish partially recorded reflections at large diffraction angles, because of the close packing of the spots. This may have led to an underestimate of γ, which in turn may have contributed to the systematic apparent overestimation of the intensities of partially recorded reflections described in Section 6.

3) DETERMINATION OF THE OPTRONICS MATRIX

The films were digitized on the Optronics Photoscan P-1000 densitometer, using a 100 μm raster and light spot size, since the spots on the films were approximately 1.1 mm in diameter. The digitized record was then read by Segment 3 of OSCAR. Fiducial holes were used to determine an approximate transform-

ation from the film coordinate system to the Optronics reference frame as described in Chapter 3, Section 4. The programme then carried out a search for 5-6 reflections above a mean background density in each octant of the film. Some difficulty was experienced in using this method, since the background density varies radially across the cylindrical films, as can be seen in Figure 5.12. A modification was made to the programme (Stura, 1981) which allowed the background density to be evaluated radially. This led to an improved selection of reflections in each area of the film. The precise positions of these reflections, obtained by a centre of gravity search, were used to determine a precise transformation between the film and Optronics frames.

4) INTEGRATION

Each reflection was then integrated using Segment 4 of OSCAR in the manner described in Chapter 3, Section 5. The optical density values in a box centred on the predicted position of a reflection were summed, and a background correction was subtracted. The type of integration box used was that shown in Figure 3.13. The size of the integration box was varied by the oblique incidence expansion factor, $f(T,\chi)$, defined in equation 3.11b. Reflections were rejected at this stage if they exceeded the saturation level of the film, or if any of the four backgrounds differed by more than 10% of their mean value. Many low angle reflections on both strong and weak films were too dark to measure. Tables 6.1 and 6.2 show the results of the integrations of a typical ENAT and RNAT crystal respectively and the results summed over each crystal are shown in Table 6.3. One problem which may affect the integrated intensities and for which no correction was applied, is the contribution made by what

TABLE 6.1

Results of OSCAR integrations for a typical ENAT crystal

Film	Number Rejected*	Number Integrated	Number of background fails	Number too strong
ENAT-7-1A	678	2209	1	449
7-1B	664	2481	9	176
7-2A	677	2246	5	342
7-2B	652	2490	4	125
7-3A	691	2045	2	602
7-3B	663	2476	2	204
7-4A	678	2105	5	494
7-4B	677	2409	5	189
7-5A	783	2061	2	571
7-5B	679	2390	1	218
7-6A	705	2126	2	552
7-6B	698	2459	3	217
7-7A	667	2111	3	492
7-7B	666	2417	6	178
7-8A	710	1980	0	610
7-8B	702	2389	1	208

* Number Rejected as being:-

a) classified as overlapping, or
b) outside the Optronics frame, or
c) outside the resolution limit.

TABLE 6.2

Results of OSCAR integrations
for a typical RNAT crystal

Film	Number Rejected	Number Integrated	Number of background fails	Number too strong
RNAT-4-1A	434	2542	34	194
4-1B	423	2694	29	61
4-2A	414	2562	44	151
4-2B	388	2346	22	32
4-3A	428	2625	45	127
4-3B	484	2707	37	9
4-4A	395	2370	12	346
4-4B	376	2606	34	95
4-5A	437	2475	27	266
4-5B	413	2717	21	60
4-6A	426	2486	21	212
4-6B	417	2623	36	52

TABLE 6.3

Results of OSCAR integrations:- Summary*

Crystal	Number of Filmpacks	Number Rejected	Number Integrated	Number of background fails	Number too strong
ENAT-1	4	1219	19258	187	6185
ENAT-2	6	2036	35388	185	2050
ENAT-3	4	1812	23647	175	1237
ENAT-5	18	13993	96209	174	8781
ENAT-6	4	3858	21108	24	1241
ENAT-7	8	10990	36394	51	5627
ENAT-8	13	5586	82925	110	2910
ENAT-9	12	14579	59680	240	2782
ENAT-10	10	12192	54824	147	3046
RNAT-2	8	2643	46010	431	682
RNAT-3	11	9160	56433	395	3322
RNAT-4	6	5035	30753	362	1605
RNAT-5	7	9177	33870	263	1310

* These results are for both A and B films, summed over each crystal.

appears to be Thermal Diffuse Scattering (TDS). As was discussed in Chapter 5, TDS contributes to the Bragg intensity and the background around it, and the method which OSCAR uses to evaluate these cannot distinguish between the contributions. Willis and Pryor (1975) discuss the correction of X-ray diffraction intensities for TDS. If E_{meas} is the total integrated intensity, and E_0 the contribution from Bragg scattering (Figure 6.2), then

$$E_{meas} = E_0(1+\alpha)$$

where $E_0\alpha$ is the contribution from the first-order TDS. α is known as the 'TDS correction factor' and

$$\alpha = \frac{E_1^p - E_1^b}{E_0}$$

where E_1^p is the TDS intensity associated with the Bragg peak and E_1^b is the TDS intensity associated with the background (Figure 6.2).

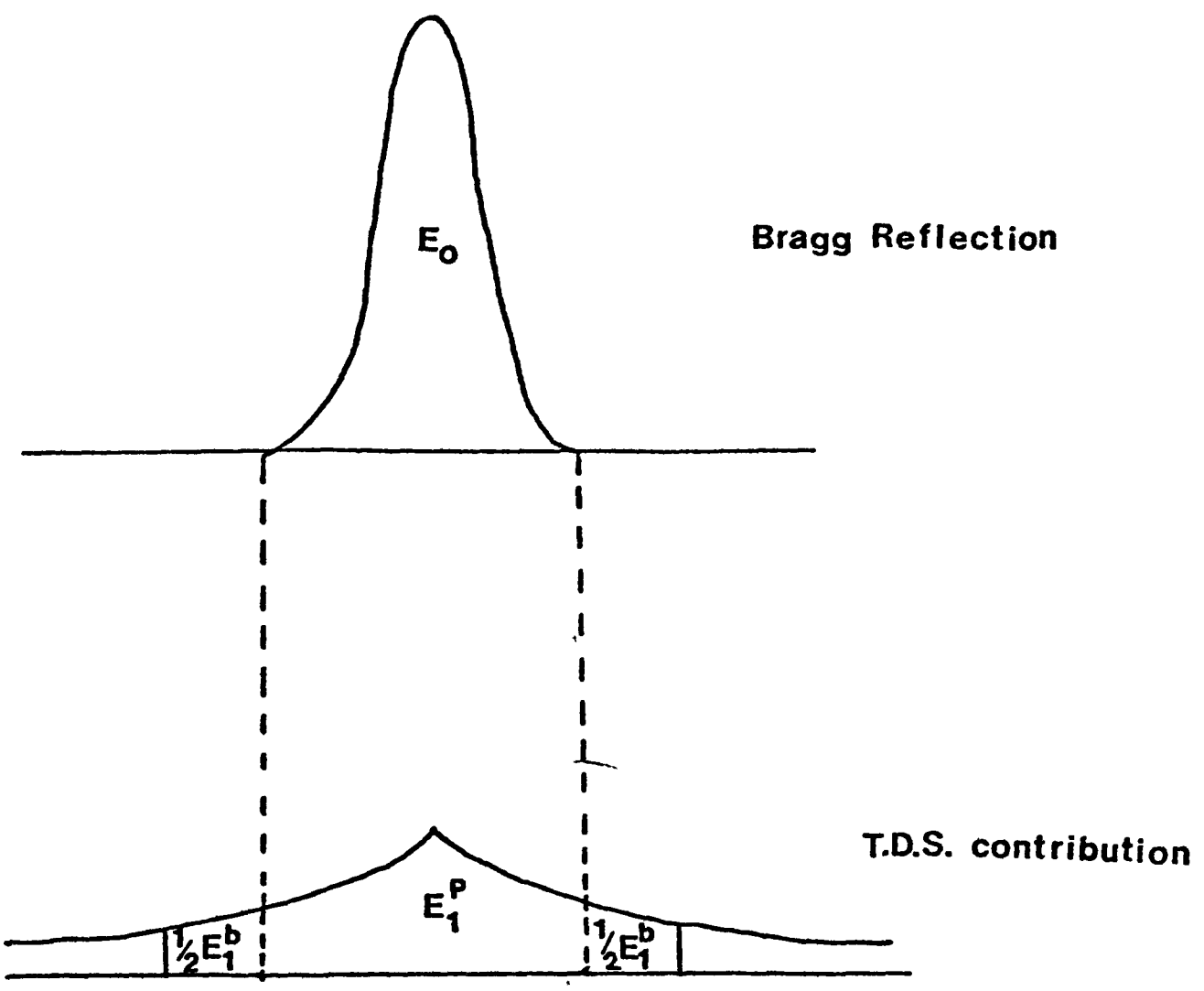
α may be calculated numerically if the elastic constants of the crystal are known. However, Willis and Pryor show that α is a function of $\sin^2\theta/\lambda^2$ and that, if α is small

$$(1+\alpha) \approx \exp\alpha \approx \exp(\text{constant}.(\sin^2\theta/\lambda^2)).$$

Thus the result of not applying a TDS correction to the measured intensities is equivalent to a reduction of the overall B-factor. This systematic error may, therefore, be alleviated by the technique of refining relative B-factors during the scaling of the data by SSM (Section 6), which eliminates *relative* errors between the films in a data set.

FIGURE 6.2

The T.D.S. contribution to a Bragg Reflection and background intensity.



(from Willis and Pryor, 1975).

5) INTER-FILM SCALE DETERMINATION (FILMPACK)

After integration the intensities from the strong and weak films were scaled together using the programme FILMPACK. This programme applied the Lorentz, polarization and oblique incidence corrections as described in Chapter 3, Section 6, for synchrotron radiation and cylindrical cassettes. Results for typical ENAT and RNAT crystals are shown in Tables 6.4 and 6.5, and the results for all crystals are summarized in Table 6.6. In this table the values of mean intensity on the strong film. ($<I_s>$), the merging R factor between films, and At, the absorption term in the weak to strong scale factor, are averaged over the filmpacks for each crystal. Since there were very few fully recorded equivalent terms, it was often not possible to calculate the standard deviation (equation 3.12), and default values were assumed, based on experience with previous 3 Å resolution oscillation photographs (σ = 50+0.01 I for strong films and 40+0.01 I for weak films).

The value of the mean merging R-factor between films for the total data set is 12.1%, which is comparable to those obtained for 3 Å oscillation data (Table 4.8). The R-factors for the RNAT crystals are somewhat lower than those for the ENAT crystals, which may reflect improvements in the determination of the crystal orientation matrices. However, the merging R-factors quoted in the table are somewhat misleading, since they tend to be dominated by the low resolution data. Figure 6.3 shows the variation of R with $4\sin^2\theta/\lambda^2$ for a typical 2 Å and 3 Å photograph.

The values of the film constant At are slightly lower than the theoretical value of 1.061 for Kodirex (Morimoto and Uyeda, 1963) which may reflect non-standard developing conditions (*e.g*: lack of nitrogen burst agitation and thermostatic temperature control). At this stage in the data reduction fully recorded

TABLE 6.4

Merging of a typical ENAT crystal:- results output by FILMPACK

Filmpack	$\langle I_s \rangle$	R_s%	R_w%	R%	At	Number of Reflections output	Number of negative terms	Number of terms rejected
ENAT-7-1	902	7.2	6.1	11.5	0.942	1550	152	4
7-2	850	6.4	5.2	9.6	0.911	2491	235	2
7-3	803	8.5	3.8	11.2	0.947	2478	239	3
7-4	930	6.4	5.9	10.8	0.945	2421	213	1
7-5	988	9.1	5.4	11.0	0.989	2555	292	4
7-6	854	9.7	6.0	12.8	0.923	2476	223	6
7-7	802	8.7	5.2	12.1	0.990	2476	251	1
7-8	729	12.5	7.9	14.5	0.949	1457	130	3

TABLE 6.5

Merging of a typical RNAT crystal:- results output by FILMPACK

Filmpack	$\langle I_s \rangle$	R_s%	R_w%	R%	At	Number of Reflections output	Number of negative terms	Number of terms rejected
RNAT-4-1	770	6.0	7.1	8.3	0.847	1672	218	1
4-2	528	7.6	5.8	7.9	0.744	2733	505	5
4-3	596	6.6	7.5	9.1	1.073	2779	454	3
4-4	686	10.1	7.1	11.4	1.128	2624	465	1
4-5	573	10.6	7.2	12.3	0.925	2762	457	2
4-6	624	7.5	7.1	9.7	0.950	1607	284	4

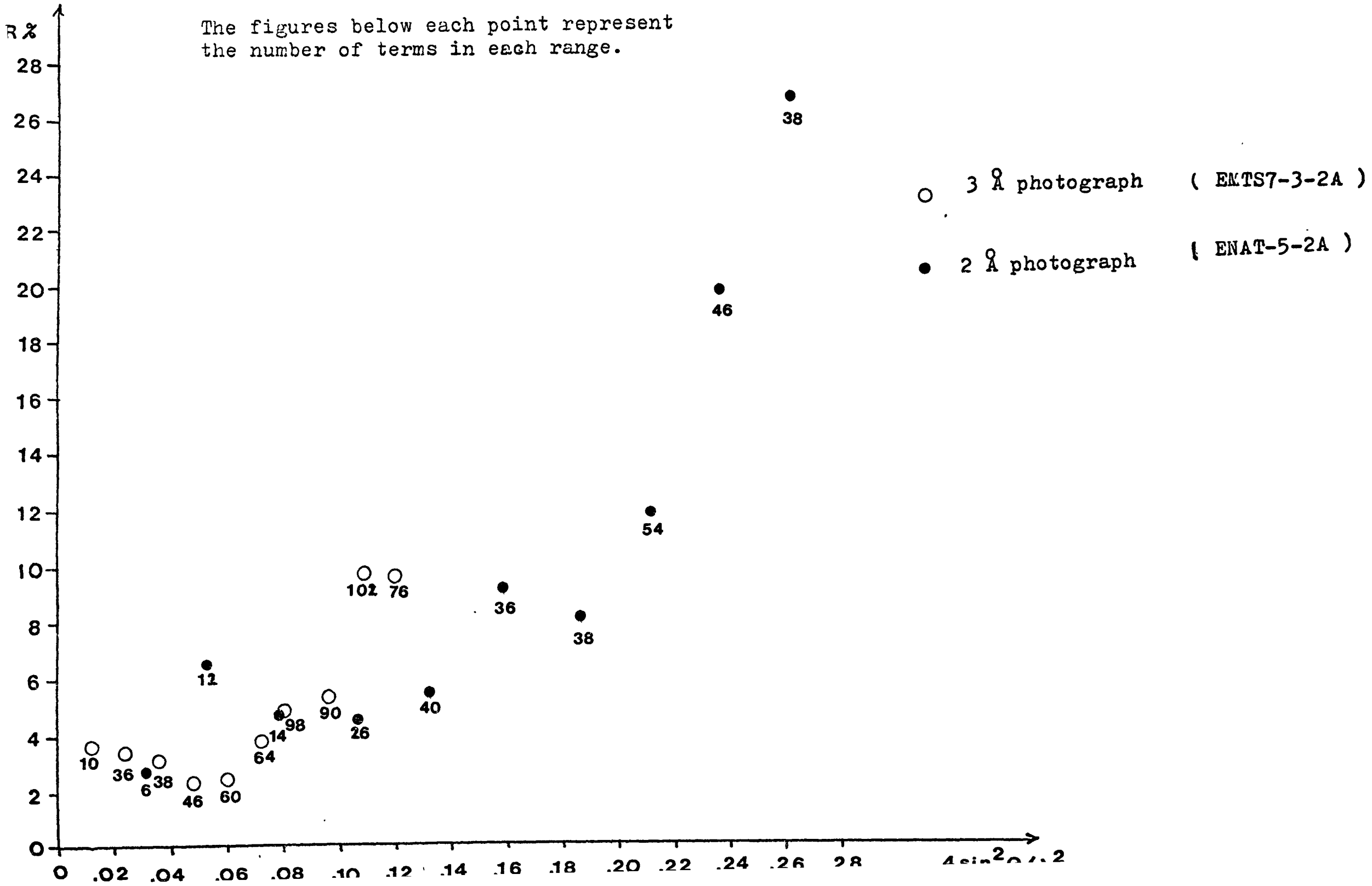
TABLE 6.6

Merging of ENAT and RNAT Filmpacks:- Summary

Data Set	Number of filmpacks	$\langle I_s \rangle$	Mean values of R%	At	Number of unique reflections	% Reflections rejected
ENAT-1	4	576	19.6	0.945	9666	0.84
ENAT-2	6	577	17.0	1.143	16290	0.35
ENAT-3	4	581	14.4	0.969	11019	0.21
ENAT-5	18	740	13.6	0.882	47515	0.28
ENAT-6	4	645	11.9	0.947	9942	0.0
ENAT-7	8	857	11.7	0.950	17904	0.13
ENAT-8	13	644	12.5	0.872	39580	0.12
ENAT-9	12	695	9.5	0.888	29203	0.14
ENAT-10	10	667	10.9	0.912	23734	0.08
RNAT-2	7	592	9.6	0.8413	18169	0.24
RNAT-3	9	823	8.6	0.8127	21093	0.46
RNAT-4	6	630	9.8	0.9444	14177	0.11
RNAT-5	7	881	8.1	0.8483	12337	0.73

FIGURE 6.3

Variation of R-factor with $4\sin^2\theta/\lambda^2$ for a 3Å and 2Å photograph



reflections were summed to give mean weighted intensities and observations more than 4 standard deviations from this mean were rejected. The number of rejected terms normally gives some indication of the quality of the crystal orientation matrix determination, but allowance must be made for the fact that there were often insufficient fully recorded terms to enable the standard deviation to be estimated, and default values were used. At this stage film packs which scaled badly (*e.g*: ENAT-5-3 with merging R = 48.2%) were removed from the data set after several attempts to improve the scaling by reintegration of the films had failed. No radiation damage or absorption corrections have yet been applied to the data. The systematic errors introduced by this omission have been discussed in Chapter 4.

6) REDUCTION TO UNIQUE DATA (SSM)

The data were then processed by SSM in two halves, owing to limitations on the maximum computer job time. Crystals ENAT 1,2,3,5,6,8 formed data set ENAT and ENAT 7,9,10 and RNAT 2,3,4,5 formed data set RNAT. SSM determines inter-film scales, sums the separate parts of partially recorded reflections and merges equivalent terms to give a unique data set. The programme refined both relative scale and temperature factors using an extension to the method of Fox and Holmes (1966). As was described in Chapter 3, this should compensate for residual relative effects of radiation damage, crystal disorder, those differences in absorption which are dependent on scattering angle, and possibly the lack of a TDS correction (Section 4). An odd number of refinement cycles was performed, ending with the determination of the scale factors. Convergence criteria of 0.00001 for scale factors of 0.0005 for temperature factors were used. Reflections whose intensities differed by more than 4 standard deviations from

the current weighted estimate of the mean intensity were rejected from the scale determination. The results of this refinement are shown in Tables 6.7 and 6.8. As was observed with the EMTS7 data (Table 4.9) the scale factor varied appreciably between crystals and for different photographs taken from the same crystal. There is also the same general tendency for the relative B value to increase (to a large negative value) with photograph number, although considerable fluctuations occur within each crystal. Possible causes of this variation have been discussed in Chapter 3. The t-plots (distributions of $(F^2-\langle F^2\rangle)/\sigma(F^2)$) for the two data sets ENAT and RNAT are shown in Figures 6.4 and 6.5. This distribution should be normal, with a mean of 0.0 and a standard deviation of 1.0 if σ_F is a true estimate of the error in F^2. The t-plots in Figures 6.4 and 6.5 were judged to be acceptable.

The separate parts of partially recorded reflections were summed and the intensities analysed with respect to symmetry related terms which were fully recorded. Tables 6.9 and 6.10 show the agreement between fully recorded and partially summed terms in ranges of $4\sin^2\theta/\lambda^2$. For data to about 3 Å resolution ($4\sin^2\theta/\lambda^2 = 0.1089$) the intensities for the partially recorded terms show the same degree of apparent overestimation, compared to the fully recorded terms, as was observed with the 3 Å data (Table 4.10). However, for data beyond 3 Å resolution, there is a large increase in the degree of apparent overestimation, which reaches approximately 50% for data at the 2 Å resolution limit. As was discussed in Chapter 4, Section 2b, the most likely causes of this systematic error are

(i) underestimation of the crystal reflecting range and
(ii) inaccuracy in the determination of the crystal orientation matrix.

TABLE 6.7

Data Set ENAT; Results of Fox and Holmes Scaling

Crystal	File Number	Scale	B-value
ENAT-1	1	1.00	0.00
	2	1.11	1.15
	3	1.26	-1.06
	4	1.03	-2.45
ENAT-2	5	2.21	-4.48
	6	2.78	-4.95
	7	3.83	-6.34
	8	3.58	-5.04
	9	3.59	-4.24
	10	3.06	-5.97
ENAT-3	11	3.50	-5.20
	12	4.34	-4.79
	13	4.41	-4.17
	14	3.60	-4.97
ENAT-5	15	1.90	-2.99
	16	1.87	-3.06
	17	1.50	-3.03
	18	1.91	-3.04
	19	2.11	-4.34
	20	2.24	-5.01
	21	2.39	-2.82
	22	0.98	-4.16
	23	2.13	-4.46
	24	1.96	-4.94
	25	2.95	-4.01
	26	1.95	-4.64

TABLE 6.7 (cont.)

Crystal	File Number	Scale	B-value
ENAT-5	27	3.33	-2.91
	28	3.46	-2.97
	29	3.65	-2.38
	30	2.04	-4.58
	31	2.04	-4.58
ENAT-6	32	3.32	-3.49
	33	3.04	-3.99
	34	3.87	-4.05
	35	3.99	-4.36
ENAT-8	36	2.53	-3.97
	37	3.30	-3.96
	38	2.14	-3.59
	39	3.00	-4.71
	40	3.41	-4.84
	41	3.63	-4.83
	42	2.61	-6.70
	43	3.56	-7.14
	44	3.86	-8.15
	45	4.50	-6.27
	46	5.86	-1.46
	47	5.35	-2.28
	48	6.27	-1.02

TABLE 6.8

Data Set RNAT: Results of Fox and Holmes Scaling

Crystal	File Number	Scale	B-value
ENAT-7	1	1.00	0.00
	2	1.35	0.84
	3	0.74	1.66
	4	0.83	0.45
	5	0.72	0.40
	6	0.83	-0.49
	7	0.85	0.17
	8	0.77	0.99
ENAT-9	9	2.31	-2.49
	10	2.52	-1.44
	11	3.04	0.04
	12	2.90	0.24
	13	1.80	-2.27
	14	2.51	-1.86
	15	2.76	-1.91
	16	2.34	-2.19
	17	2.72	-1.52
	18	3.12	-1.86
	19	3.21	-2.69
	20	2.40	-3.50
ENAT-10	21	1.71	-1.47
	22	1.99	-0.56
	23	2.18	-2.67
	24	1.56	-1.11
	25	2.13	-0.82
	26	2.44	-1.90

TABLE 6.8 (cont.)

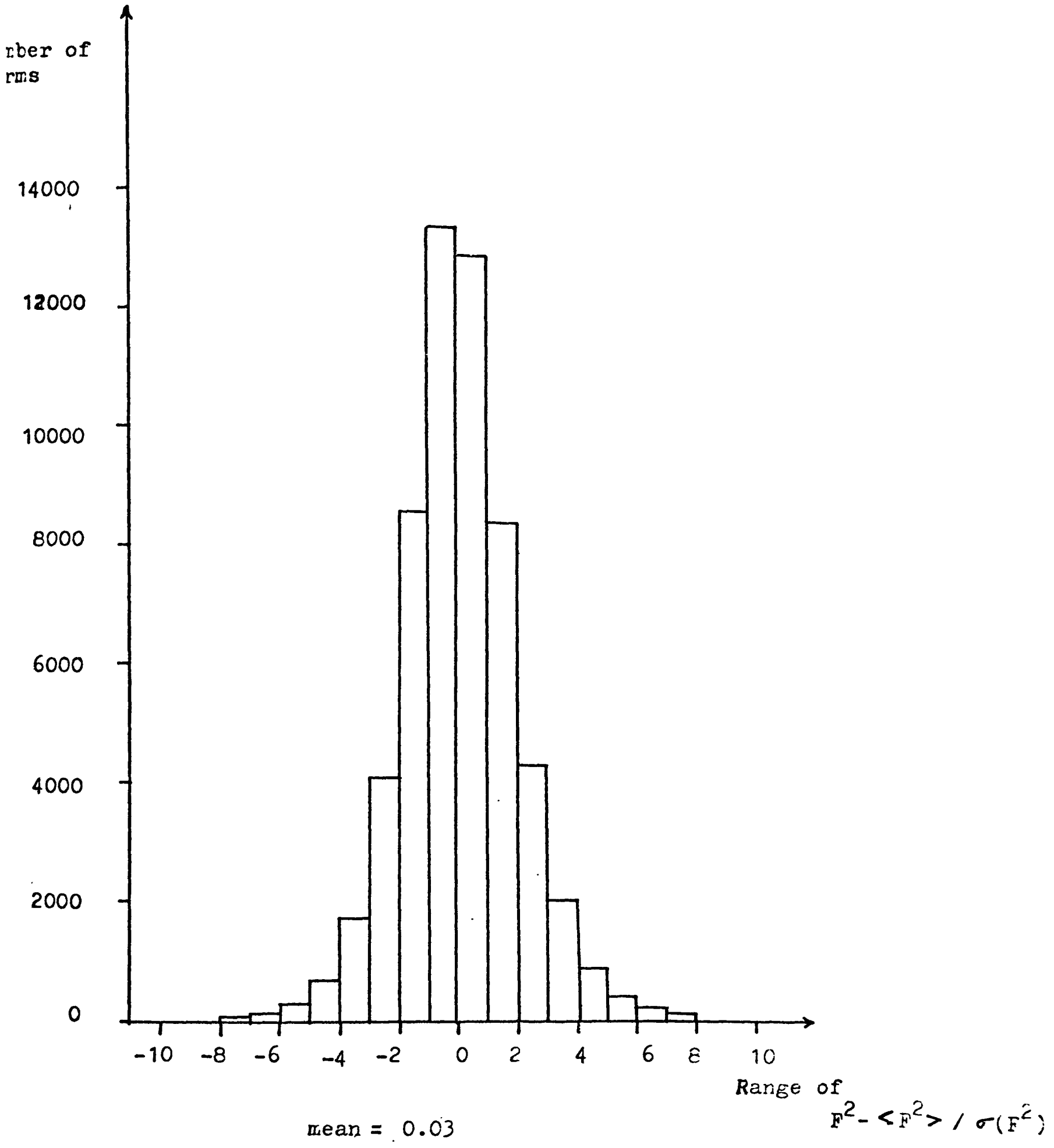
Crystal	File Number	Scale	B-value
ENAT-10	27	2.16	-7.28
	28	2.03	0.21
	29	1.84	-2.25
	30	1.42	-0.05
RNAT-2	31	6.02	-0.18
	32	3.42	-1.07
	33	3.52	-1.76
	34	4.14	-2.75
	35	4.57	-0.24
	36	4.05	-0.34
	37	4.06	-0.45
RNAT-3	38	1.65	0.59
	39	1.98	-1.19
	40	1.50	0.26
	41	1.34	-1.61
	42	1.74	-1.46
	43	2.15	-0.22
	44	2.50	-1.49
	45	2.42	-4.18
	46	1.96	-4.76
RNAT-4	47	2.14	-0.67
	48	2.50	-0.65
	49	3.17	-2.06
	50	3.08	-0.74
	51	1.51	0.68
	52	2.51	-2.80
	53	2.38	-3.13

TABLE 6.8 (cont.)

Crystal	File Number	Scale	B-value
	54	2.47	-1.08
RNAT-5	55	2.66	-0.78
	56	1.23	0.38

FIGURE 6.4

't-plot' for ENAT data set .

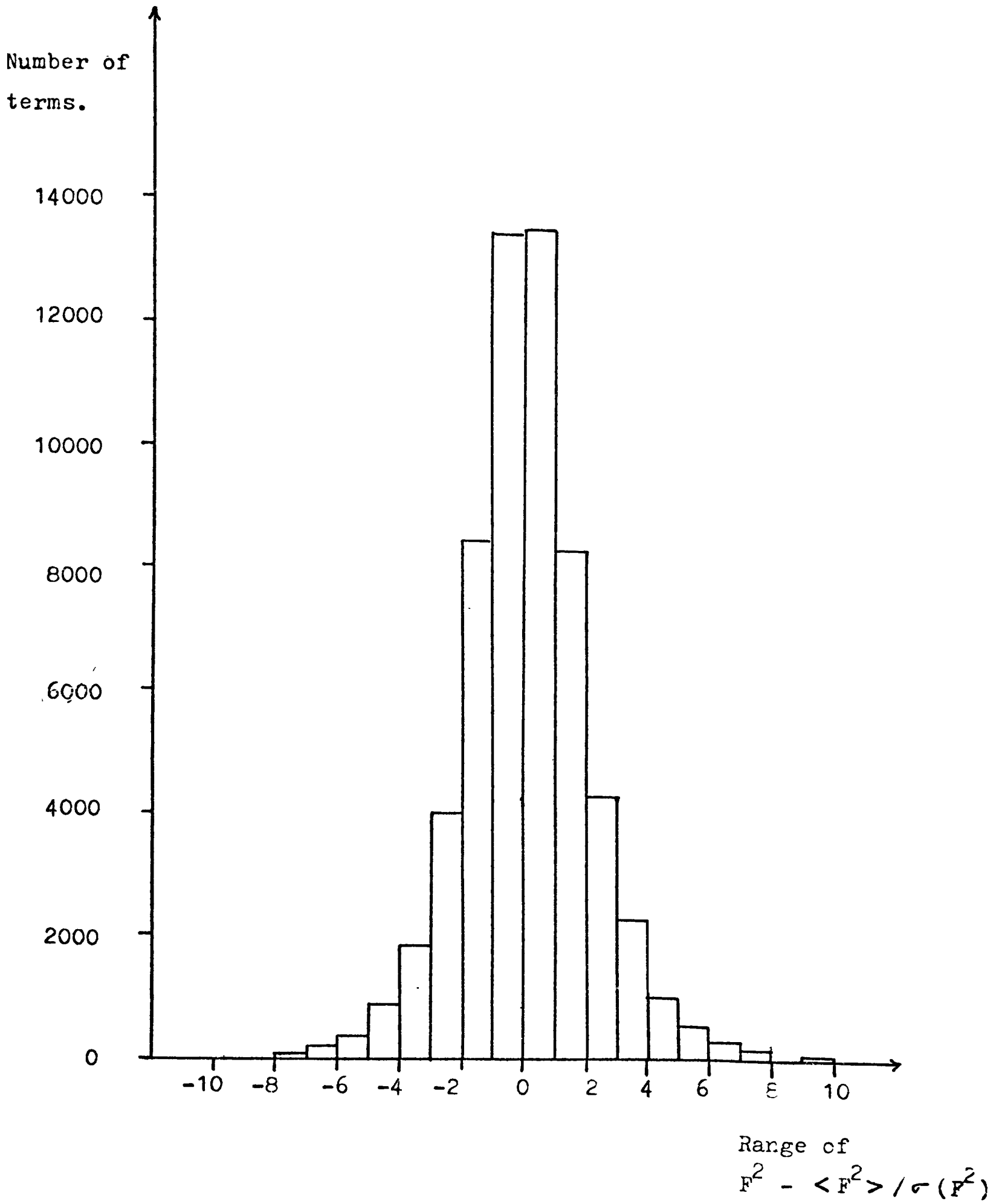


mean = 0.03

standard deviation = 1.0

FIGURE 6.5

't-plot' for RNAT data set.



mean = 0.05

standard deviation=1.08

TABLE 6.9

Fully versus partial analysis for ENAT data

Number of terms	Range of $4\sin^2\theta/\lambda^2$ min	max	$\langle F_F^2 \rangle$	$\langle F_P^2 \rangle$	% difference
85	0.0000	0.0132	12876	13270	- 3.01
325	0.0132	0.0265	14577	14693	- 0.79
442	0.0265	0.0397	15785	16036	- 1.57
479	0.0397	0.0530	23538	24681	- 4.74
545	0.0530	0.0662	25147	26700	- 5.99
680	0.0662	0.0794	22348	23962	- 6.97
792	0.0794	0.0927	15541	16851	- 8.09
839	0.0927	0.1059	11011	12235	-10.52
906	0.1059	0.1192	8227	9274	-11.96
935	0.1192	0.1324	6353	7148	-11.79
957	0.1324	0.1456	4729	5432	-13.84
955	0.1456	0.1589	4095	4712	-14.06
935	0.1589	0.1721	3295	3902	-16.87
876	0.1721	0.1854	2650	3067	-14.59
807	0.1854	0.1986	2338	2846	-19.59
730	0.1986	0.2118	1995	2576	-25.42
656	0.2118	0.2251	1645	2082	-23.44
522	0.2251	0.2383	1176	1569	-28.58
420	0.2383	0.2516	864	1215	-33.80
117	0.2516	0.2648	612	973	-45.60

TABLE 6.10

Fully versus partial analysis for RNAT data

Number of reflections	$4\sin^2\theta/\lambda^2$ min	$4\sin^2\theta/\lambda^2$ max	$\langle F_F^2 \rangle$	$\langle F_P^2 \rangle$	% difference
92	0.00000	0.01323	10230	10251	-0.20
327	0.01323	0.02646	11207	11638	- 3.77
435	0.02646	0.03969	12255	12779	- 4.19
490	0.03969	0.05292	16751	17861	- 6.41
563	0.05292	0.06615	15868	17433	- 9.39
656	0.06615	0.07938	12631	14315	-12.50
701	0.07938	0.09261	9103	10374	-13.05
751	0.09261	0.10584	6106	7178	-16.15
751	0.10584	0.11907	4282	5212	-19.60
766	0.11907	0.13230	3012	3752	-21.86
715	0.13230	0.14553	2293	2841	-21.35
740	0.14553	0.15876	1900	2560	-29.60
672	0.15876	0.17199	1569	2152	-31.36
698	0.17199	0.18522	1154	1614	-33.30
682	0.18572	0.19845	1031	1463	-34.60
572	0.19845	0.21168	775	1182	-41.60
499	0.21168	0.22491	622	989	-45.50
366	0.22491	0.23814	374	651	-53.90
294	0.23814	0.25137	329	554	-50.80
102	0.25137	0.26465	357	582	-48.00

It should be possible to test the effect of (i) by reintegrating the data with different values of γ. The effect of (ii) is more difficult to test. Since the crystal orientation matrices were determined using only 3 Å data from the flat plate cassettes, it is possible that they are not sufficiently accurate for the 2 Å data obtained with the cylindrical cassettes. By making appropriate modifications to Segment 1 of OSCAR, it should, in future, be possible to use data from 2 Å orientation photographs taken with the cylindrical cassettes to refine the crystal orientation parameters.

It is also important to improve the criteria used to judge whether or not a plot is acceptable. As was described earlier, the close packing of high angle data means that agreement between plots and films is often difficult to check accurately. An improvement could probably be made by first predicting the 'setting still' photographs. These have dark, well-separated spots and accurate agreement between plots and films should be easier to check. Data films could then be predicted and checked. This should result in an improvement in the estimation of γ and in checking the accuracy of the orientation matrix. The results of the scaling and merging are summarized in Table 6.11. The merging R-factor for ENAT was 9.3% and for RNAT, 10.3%. Both these figures are comparable to those obtained for 3 Å oscillation camera data (Table 4.8).

7) SCALING AND MERGING OF THE COMPLETE DATA SET

The two halves of the c* data set, ENAT and RNAT, were then scaled together and merged by the programme SSM. The results are summarized in Table 6.12, and represent the reduction of 70530 reflections to 46,563 unique terms. The first run gave a final relative scale and temperature factor of 1.44361 and -8.02

TABLE 6.11

Results from SSM for ENAT and RNAT data separately

	ENAT	RNAT
Number of reflections used	71,772	70,265
Number of unique terms	36,766	33,925
Number rejected	215	675
Merging R-factor	9.28%	10.27%

TABLE 6.12

Results from SSM for merging of ENAT and RNAT data

(1*b* represents a re-run of 1*a* with modified t-plot)

	1*a*	1*b*
Number reflections used	70,444	70,530
Number of unique terms	46,563	46,563
Number rejected	242	156
Merging R-factor	15.57%	15.01%
Relative scale	1.44361	1.43696
Relative B-value	-8.02	-7.68
t-plot	mean=0.15 s.d.=1.40	mean=0.11 s.d.=1.19

respectively. Since the B value relative to the first photograph should be 0 this large negative value is rather worrying, as it suggests that the two halves of the data set show a different variation of intensity with Bragg angle, θ. In retrospect, the two halves of the data set should both have been scaled to the *same* first photograph (*i.e*: the first photograph of ENAT). The t-plot had a mean of 0.15 and a standard deviation of 1.40 and was judged unsatisfactory. The coefficients A and B (equation 3.24) were modified to improve the t-plot. The second run then gave a distribution with a mean of 0.11 and a standard deviation of 1.19 (Figure 6.6), which is still not completely satisfactory and suggests that the standard deviations used are still not good estimates of the errors in the intensities. The refined relative B value was -7.68 which is still quite high. The merging R-factor was 15.01%, which is an increase in the values obtained for the two halves of the data set separately.

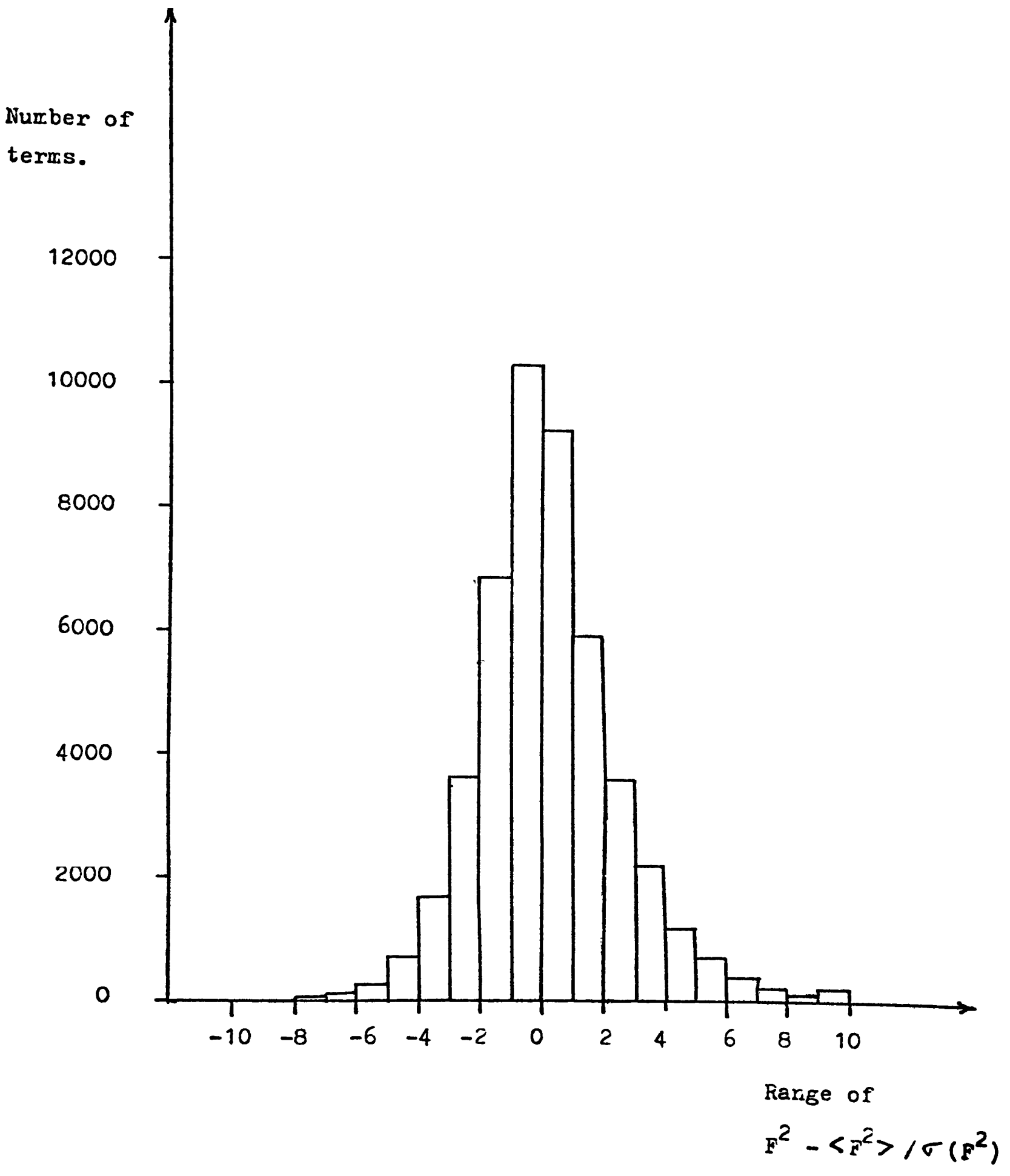
After scaling and merging, the complete data set was truncated, as discussed in Chapter 3.

8) SCALING TO 3 Å NATIVE DATA

The complete 2 Å c* data set was then scaled to the 3 Å native data using the programme ANSC. As described in Chapters 3 and 4, scaling as a function of h,k, and l was used, which should minimize systematic differences between the intensities, and leave random errors as the most significant. It is here that the systematic errors in the 2 Å data become most apparent. The scale factors were analysed for variation with h,k,l, $\sin^2\theta$ and F^2. The greatest variation was observed with h and l. Scales were applied to the 2 Å and 3 Å data separately. Figure 6.7 shows the variation of the scale factors with h before and after scaling of the 3 Å data, and Figure 6.8 before and after scaling the 2 Å data

FIGURE 6.6

't-plot for combined ERNAT data set.



mean = 0.11

standard deviation = 1.19

FIGURE 6.7

Variation of scale factor with h, before and after applying scale to 3 Å data.

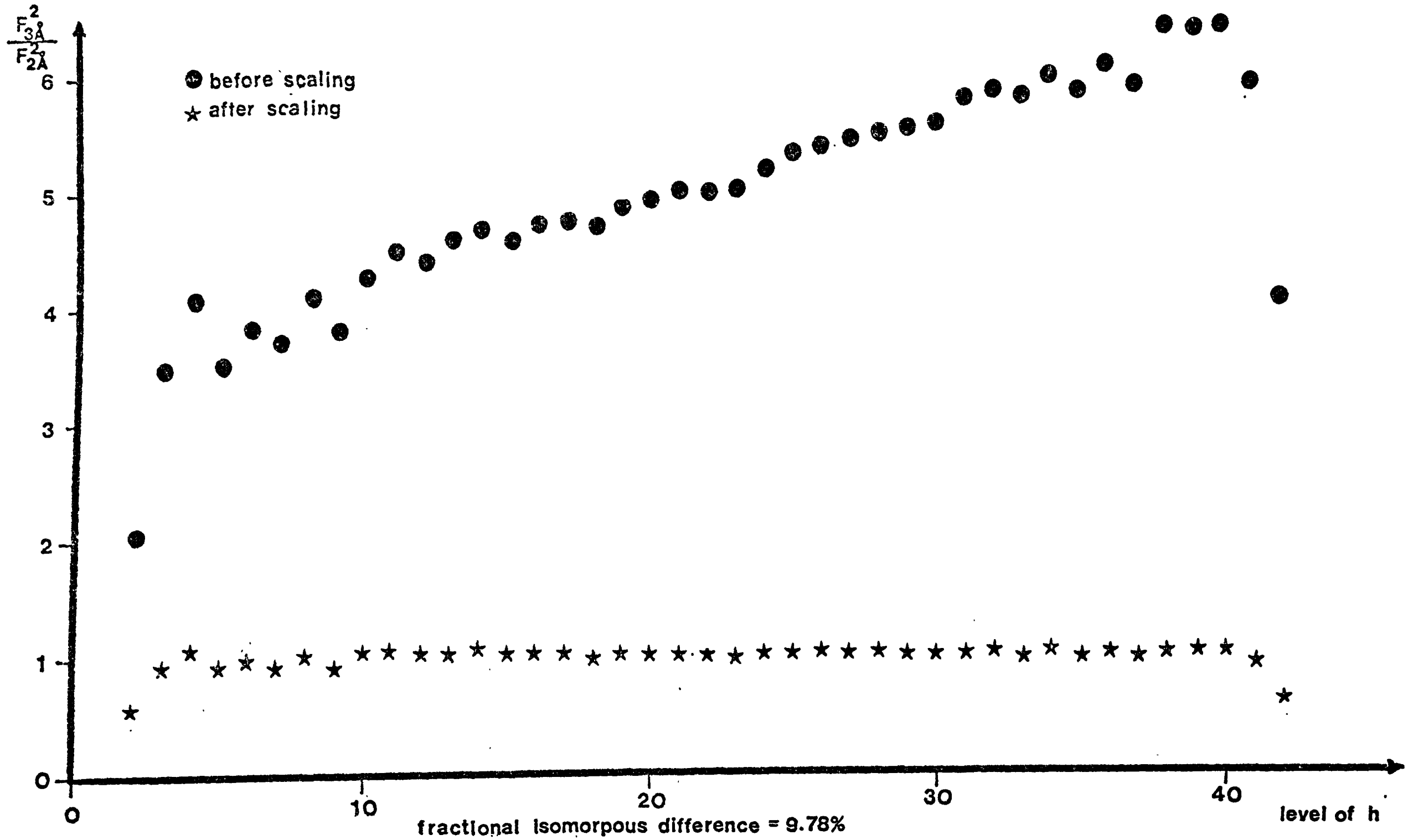


FIGURE 6.8

Variation of scale factor with h before and after applying scale to 2 Å data.

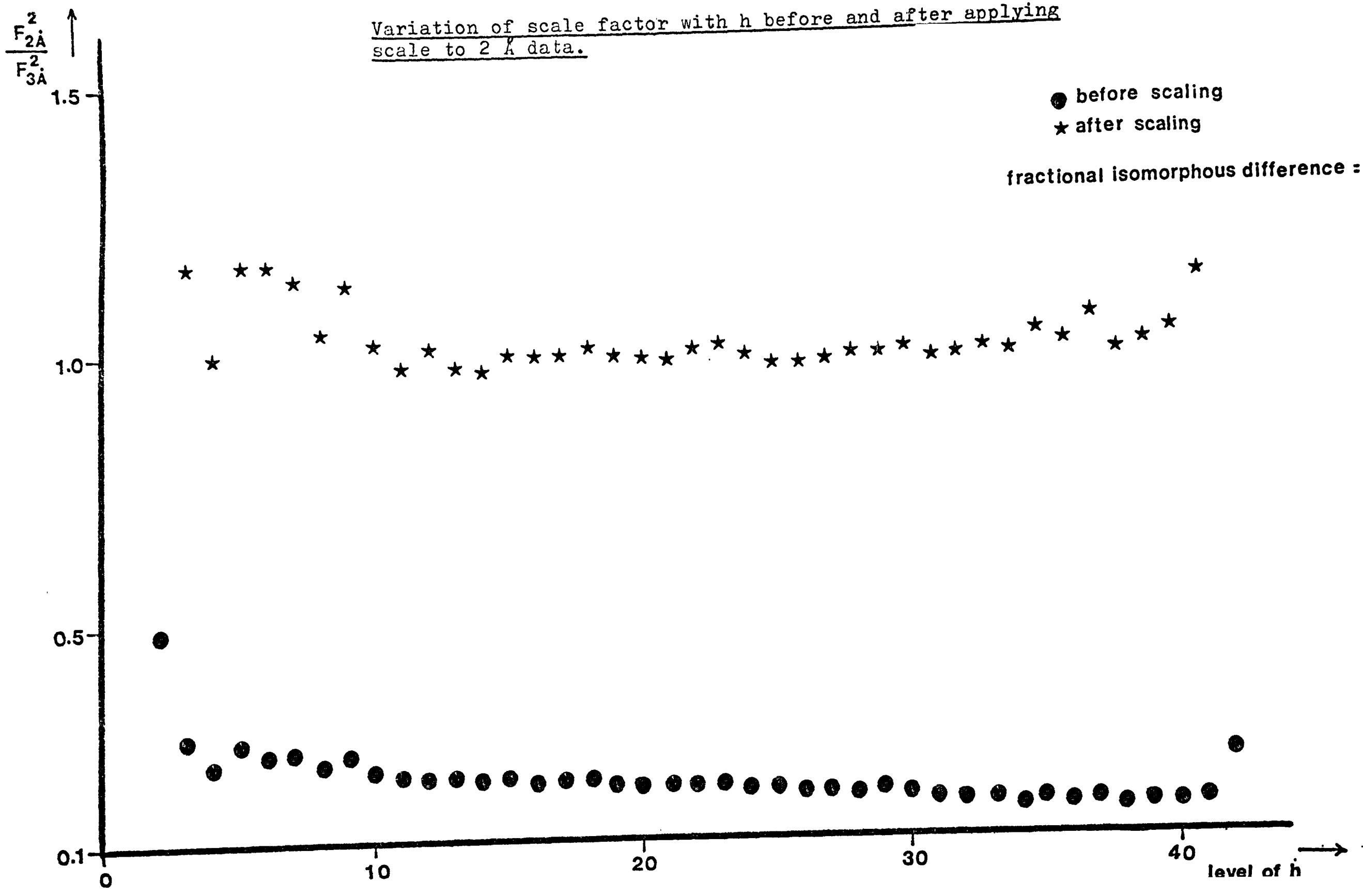


FIGURE 6.9

Variation of scale factor with l before and after applying scale to 3 Å data.

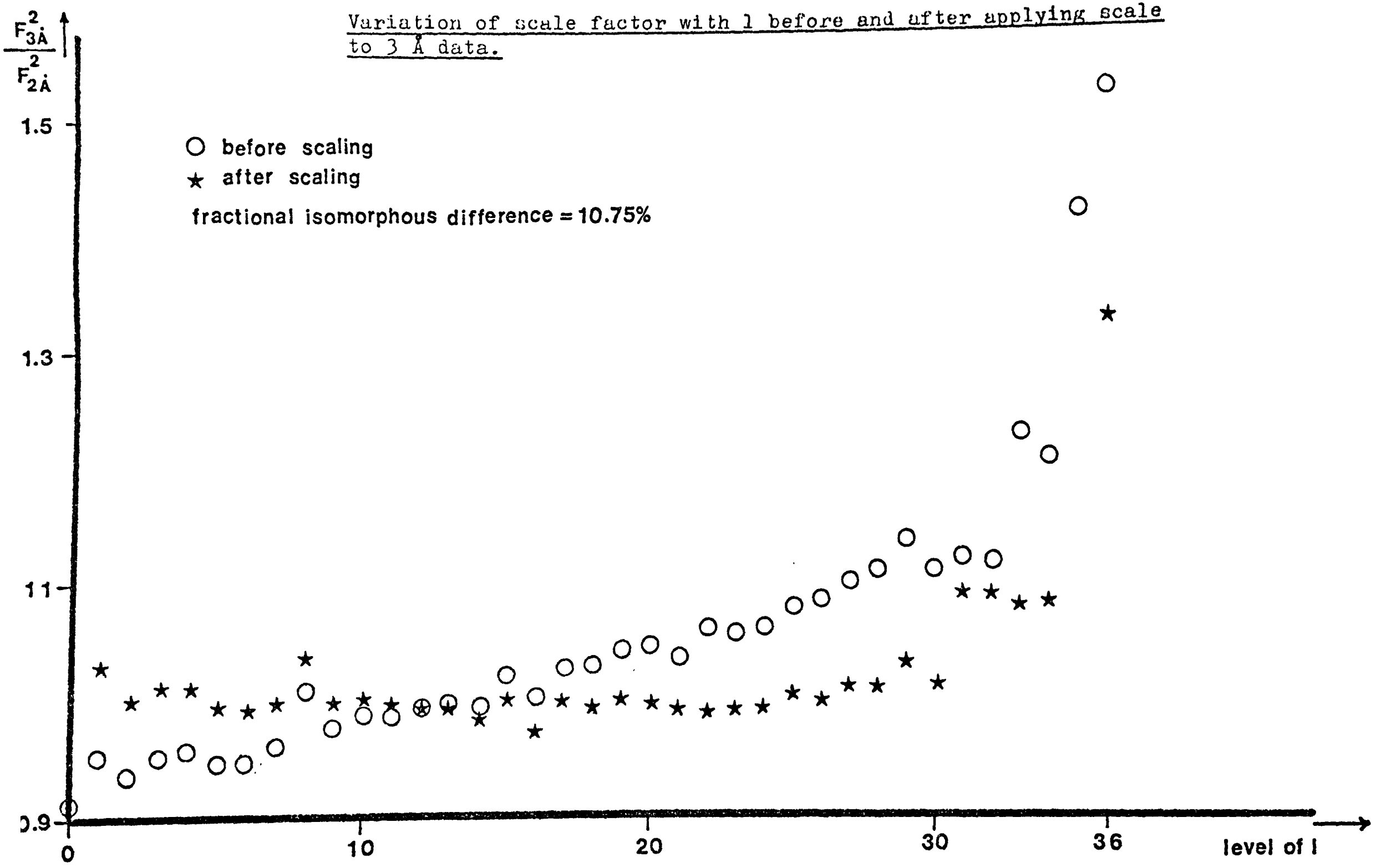


Figure 6.9 shows the variation with l before and after applying a scale to the 3 Å data. It is unclear as to which method of scaling is best. It was decided to scale the 2 Å to the 3 Å data, to leave the 3 Å data unchanged for the refinement procedure described below.

9) CONCLUSIONS

The quality of the 2 Å data was not as good as had been expected, and this is probably due to systematic errors introduced during the data processing procedure. The following sources of error have been discussed above:-

i) possible under estimation of γ; to correct for this it would be necessary to reprocess the data using different values of γ. Improved estimates of γ could probably be obtained by initially predicting and checking 'setting still' photographs,

ii) the use of 3 Å flat plate orientation photographs to refine the orientation matrices for 2 Å cylindrical data photographs; to correct for this it would be necessary to recollect the data using 2 Å cylindrical orientation photographs,

iii) difficulties in assessing the 'quality' of the plots of predicted reflections due to the close packing of spots on the 2 Å data photographs, and consequently choosing an inaccurate orientation matrix. An improvement could probably be obtained by predicting plots for 'setting still' photographs, as in (i) above,

iv) incomplete modelling of the random errors in the intensities,

v) possible errors in the measurement of the integrated intensities introduced by diffuse scattering.

To some extent the relative systematic errors between photographs have been allowed for by refining the relative scale and temperature factors, and the systematic differences between the 2 Å and 3 Å data have been allowed for by scaling the 2 Å to the 3 Å data set. However, there are still some systematic errors in the absolute data set.

Two possible solutions to the problem of systematic errors due to absorption are as follows:-

i) the highest few percent of the data at each resolution could be collected on a diffractometer, and an absorption surface calculated (North *et al.*, 1968; Kopfman and Huber, 1968). This might pose problems for phosphorylase *b*, since the high resolution data are very weak, and long exposure times on a diffractometer would lead to readiation damage problems,

ii) since absorption increases with λ^3 (Blundell and Johnson, 1976), data could be collected on a synchrotron radiation source with a lower wavelength (say 1 Å), to reduce absorption problems.

However, despite the systematic errors remaining in the 2 Å data, they have been used in the initial stages of refinement of the phosphorylase *b* structure.

This work has been carried out by David Stuart on the Cray 1 computer at the S.R.C. laboratory, Daresbury, using the restrained parameters structure factor least-squares refinement method of Konnert (1976). The progress of the refinement to date is shown

in Figure 6.10. The initial stages were carried out using data from 10 to 3 Å resolution, with initial atomic coordinates obtained from the 3 Å resolution model described earlier. Data from 10 to 5 Å resolution were then excluded. At the end of this stage, the R-factor had dropped to 33.8%. The high resolution data from 3 to 2.5 Å were then included.

The R-factor initially increased to 44.2% and then refined down to 39.1%, with an overall r.m.s. error in bond lengths of 0.06 Å. Table 6.13 shows the variation in R-factor with the resolution of the data.

The next stages in the process will be the merging of the a* data, to fill in the 'blind region', the refinement of temperature factors and the combination of isomorphous and calculated phases.

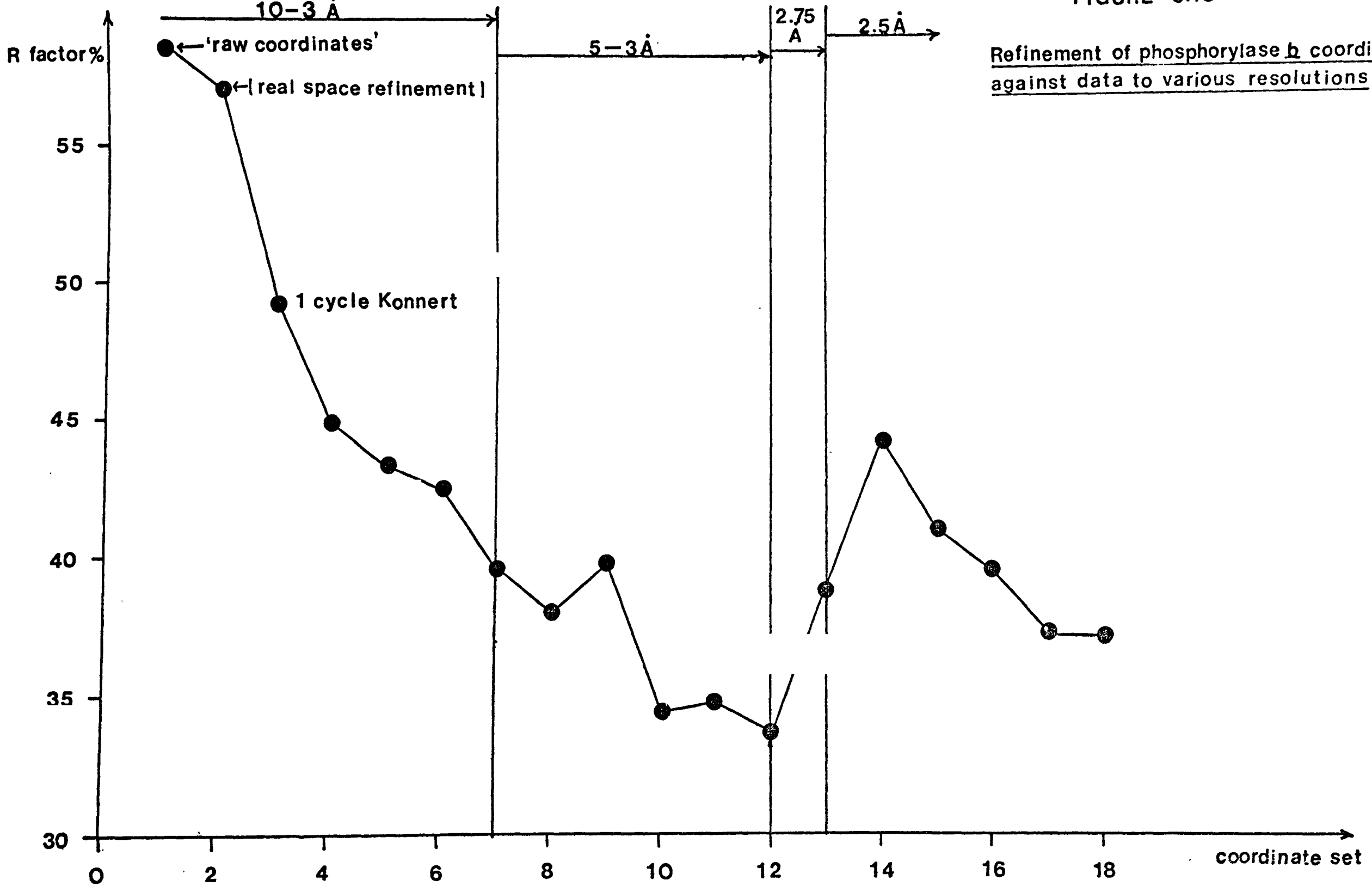


FIGURE 6.10

Refinement of phosphorylase b coordinates against data to various resolutions.

TABLE 6.13

Results from Konnert refinement
using data to various resolutions

Minimum resolution (Å)	Shell R-factor	Overall R-factor
4.50	0.308	0.308
4.00	0.311	0.310
3.50	0.345	0.325
3.10	0.381	0.342
2.80	0.466	0.364
2.60	0.540	0.383
2.50	0.549	0.391

CHAPTER 7

DISCUSSION AND CONCLUSIONS

Glycogen phosphorylase presents an interesting subject for X-ray crystallographic study, both for its key role in the control of energy metabolism in eukaryotic cells and as an example of a large and complex allosteric protein. The analysis of its structure at high resolution should provide many new insights into the functions and metabolic control of allosteric enzymes. The work presented here has contributed to the solution of the phosphorylase *b* structure at both 3 Å and 2 Å resolution.

1) CONTRIBUTION TO IMPROVED 3 Å MAP

At the time crystallographic studies of phosphorylse *b* were begun (Eagles *et al.*,1972), it was not clear whether the method of heavy atom isomorphous replacement would work for a protein with such a large molecular weight. Crick and Magdoff (1956) had calculated the average fractional change in intensity ($\phi\,\Delta I$) due to the addition of heavy atoms to be:-

$$\phi\,\Delta I \approx 2\left[\frac{N_H}{N_P}\right]^{\frac{1}{2}}\frac{f_H}{f_P} \qquad \text{for centric terms}$$

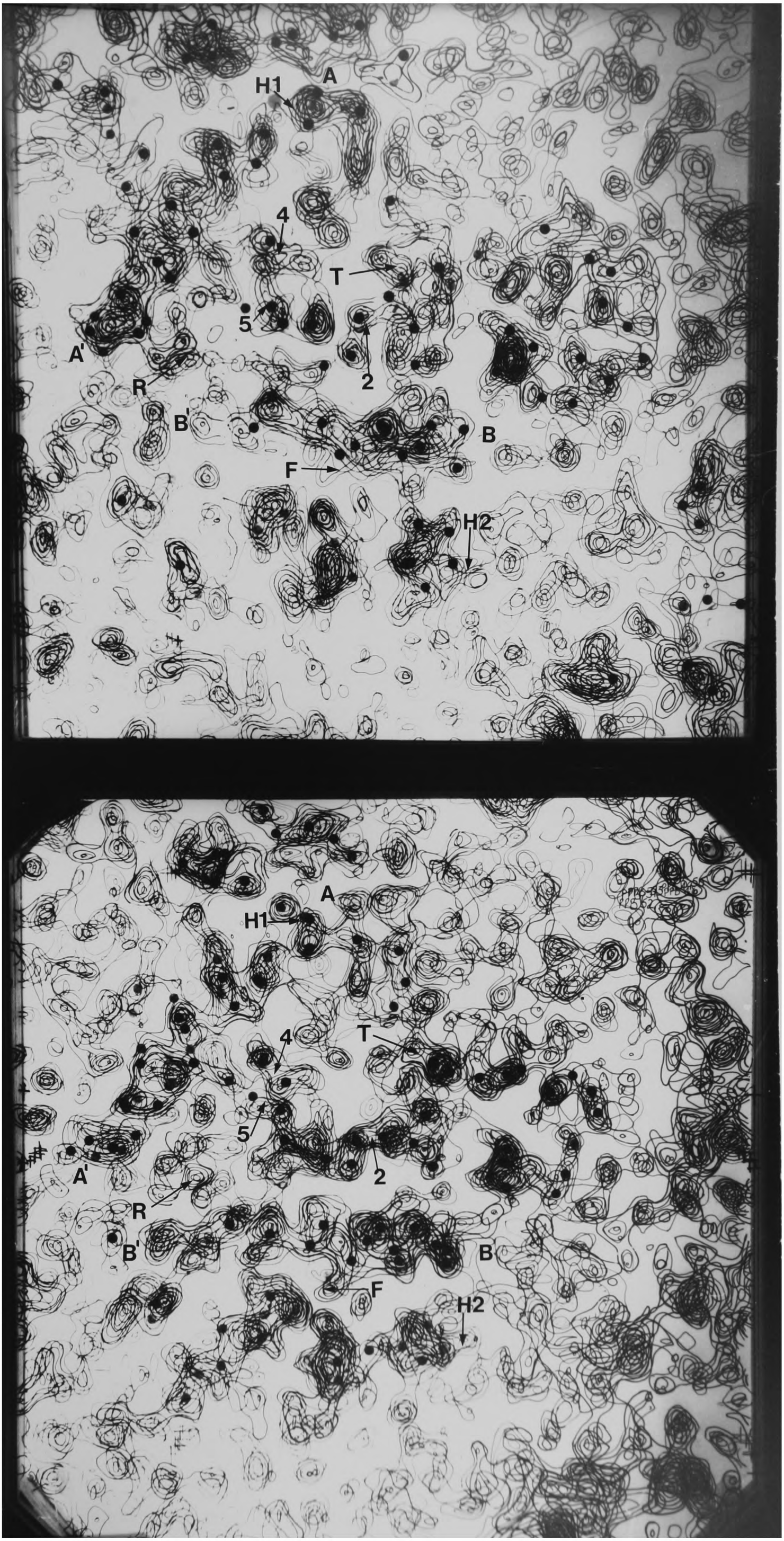
$$\approx\sqrt{2}\left[\frac{N_H}{N_P}\right]^{\frac{1}{2}}\frac{f_H}{f_P} \qquad \text{for acentric terms}$$

provided Wilson statistics are obeyed. Here N_H and N_P are the number of heavy atoms and protein atoms in the unit cell, and f_H and f_P are estimates of the scattering factors for the heavy atoms and protein atoms respectively. The addition of one mercury atom to phosphorylase *b* (MW 97,412) should thus give a fractional change in intensity of approximately 19% for acentric reflections and 27% for centric reflections. Evidently measurements of intensities must be made with considerable precision to detect such changes.

However, the solution of the phosphorylase *b* structure to 6 Å resolution (Johnson *et al.*,1974), using two heavy atom isomorphous derivatives and their anomalous contributions, showed that the isomorphous replacement method would work. The merging R-factor for both derivatives was approximately 5%. The 6 Å electron density map allowed the molecular boundary to be clearly distinguished and allowed the crystallographic two-fold axis at z=1/2 to be identified as the symmetry axis of the dimer. Later work on phosphorylase *a* in Edmonton showed that the structure was also solveable to 3 Å resolution (Fletterick *et al.*,1976b). Shortly afterwards the structure of phosphorylase *b* was also solved to 3 Å resolution, using two heavy atom isomorphous derivatives and their anomalous contributions (Johnson *et al.*,1978;Weber *et al.*,1978). The electron density map had an overall figure of merit of 0.54, and was of sufficient quality to allow the fold of the polypeptide chain to be traced by placing stickers corresponding to α-carbon positions 4 Å apart. The availability of the amino-acid sequence (Titani *et al.*, 1977) also enabled the PLP and metabolite binding sites to be identified, as described in Chapter 2. However, there were severe limitations to the map. Although, in several areas (particularly α-helical regions), the chain connectivity could be followed with confidence for long stretches, in others it was often tenuous and difficult to follow. This was paticularly the case in a number of areas at the edge of the molecule where the chain looped back on itself, and in areas of β-pleated sheet, where the connectivity across the sheet often appeared as good as the connectivity along the strands. Many amino-acid side chains were also not clearly resolved (Johnson *et al.*,1978). For a more meaningful interpretation of the structure at 3 Å resolution, a clear improvement in phasing was required, and my contribution to the phosphorylase project began

with a search for additional heavy atom derivatives, described in Chapter 4. The behaviour shown by most of the heavy atom reagents described in Chapter 4 was fairly predictable, with the noteable exception of aurothioglucose. This reagent bound at two sites, Cys 171 and a site close to site N (possibly Arg 193). It is interesting to note that the glucose moeity did not bind to the catalytic site, although glucose itself binds to this site. These results were obtained from the crystal soaked in 1mM aurothioglucose for two days and further study, at higher concentrations (the K_m for the substrate G1P is 37mM), and possibly with cross-linked crystals (to prevent allosteric transitions from taking place) might provide more detailed information about this interaction.

The addition of data from two new derivatives (EMTS7 and Baker's dimercurial), together with the reprocessing of the old data, described in Chapter 4, enabled a new electron density map, with a figure of merit of 0.63, to be calculated. The new map showed better connectivity for the polypeptide chain and better electron density for the amino-acid side chains. In particular, the areas of β-sheet showed improved connectivity along the strands. As an example, Figure 7.1 shows sections z=48/128 to 53/128 of the two versions of the 3 Å electron density map. The two α-helices AA' and BB' which line the nucleotide binding site are clearly visible in both maps. Three strands of β-sheet are visible; one strand of parallel β-sheet between residues 154-159, and two strands of anti-parallel β-sheet between residues 227-229 and 241-248. These strands are labelled 2,4 and 5, respectively, as in Figure 2.7.
The improvement in connectivity in the new map is evident. Examples of amino-acid side chains which show improved electron density in the new map are His 55 (AA' helix), Phe 298 (BB' helix), Trp 182, His 389 and Arg 309. The new map enabled a molecular model on


H1
A
4
T
5
A'
2
R
B'
B
F
H2
A
H1
4
T
5
A'
2
R
B'
B
F
H2

a scale of 2cm/Å to be built by L.N.Johnson, J.A.Jenkins and G. Zannotti. This model, together with model-building studies for AMP, ATP and G1P bound to the molecule, allowed a detailed description to be made of the amino-acid residues involved in the subunit-subunit interface, nucleotide and G1P binding, and led to proposals for a catalytic mechanism (Johnson *et al.*,1979,1980). These results have already been described in detail in Chapter 2.

2) PHOSPHORYLASE AS AN ALLOSTERIC PROTEIN

This section summarises the contribution of the crystallographic results towards an understanding of phosphorylase *b* as an allosteric protein.

The two classic theories of allosteric proteins are those of Monod, Wyman and Changeux (1965) (MWC), and Koshland, Némethy and Filmer (1966) (KNF). According to the MWC model, allosteric proteins are oligomers which exist in at least two conformational states; T (=tense) or R (=relaxed), which are in equilibrium. Differences in the number and energies of the bonds between the subunits result in the T-state being constrained, relative to the R-state. The T-state has a lower affinity for ligands and is the predominant form when no ligands are bound. The model assumes symmetry between the binding sites, i.e. all the sites in each state are equivalent and have the same binding constants for ligands. Control by allosteric activators or inhibitors is explained by the effector altering the T-R ratio by preferentially stabilizing one of the two forms. An activator binds to the R form and increases its concentration, whilst an inhibitor binds preferentially to the T-state and makes a transition to the R-state more difficult. The model excludes a mixed T-R state, i.e. all the subunits of the enzyme must be in one

state or the other. The phenomenon of positive cooperativity is thus explained by the fact that the conformation of all the subunits must change simultaneously. The assumption of the MWC model which is most open to criticism is that of symmetry. The KNF model avoids this assumption but uses another simplifying feature. It assumes that the transition from the T-R state is a sequential process, and that there is no dramatic switch from one state to another. In the absence of ligands, the protein exists in one conformation. When a ligand binds, it induces a conformational change in the subunit to which it is bound. The conformation of each subunit then changes in turn as it binds effectors. Conformational changes may be transmitted to neighbouring vacant subunits via the subunit interfaces, providing an explanation for cooperativity. Positive cooperativity occurs when a conformational change in one subunit increases the ligand binding affinities of the remaining subunits, and negative cooperativity occurs when the ligand binding affinities of the remaining subunits are decreased. Negative cooperativity cannot be explained by the MWC model, since the binding of the first ligand molecule can only stablize the high affinity state and cannot increase the proportion of the T-state.

From a structural point of view, then, allostery occurs through conformational changes which link ligand binding sites to subunit interfaces. When the MWC and KNF models were first proposed, structural information concerning such changes was not available, and the models were based on equilibrium binding and kinetic studies. Although other physical and chemical techniques may provide information about the characteristics of interactions between ligand binding sites and subunits, only X-ray crystallography is capable, in principle, of providing a picture of the

enzyme "frozen" in its various conformational states. Such pictures should allow proposals to be made about the movement of individual residues and atoms during allosteric transitions. These hypotheses could then be tested by solution techniques. The only allosteric protein for which a detailed structural interpretation exists is haemoglobin (Perutz, 1979). In the crystal, haemoglobin exists in only two states; one characteristic of deoxyhaemoglobin and the other characteristic of all liganded haemoglobins. The deoxy (T) state is stablized by four pairs of salt bridges between the interfaces of the two α chains and two salt bridges between the α and β chains. On oxygenation the salt bridges between each subunit are broken and there is a decrease in the hydrophobic area which is in contact at the subunit interfaces. The 'constraints' between the subunits in the T-state are thus the salt bridges and hydrophobic interactions, and these might also be expected to play an important role in other allosteric proteins. Apart from phosphorylase, other allosteric proteins whose crystallographic studies are well advanced are aspartate transcarbamylase (Monaco *et al.* 1978), phosphofructokinase (Evans and Hudson,1979) and hexokinase (Bennet and Steitz,1978).

The allosteric properties of phosphorylase *b* have been investigated extensively by kinetic and equilibrium binding studies, which suggest that the enzyme may exist in several conformations. Although most of the studies have been analysed on the basis of the MWC model, several discrepancies between the observed allosteric properties and the predictions of the model have been noted, particularly relating to the heterotropic interactions of AMP and P_i (see review by Graves and Wang,1972). Differences between IMP and AMP activation (see Chapter 2) led Black and Wang (1968) to suggest that the AMP induced allosteric transition in phosphorylase *b* consists of two stages; that of enhancing enzyme affinity towards

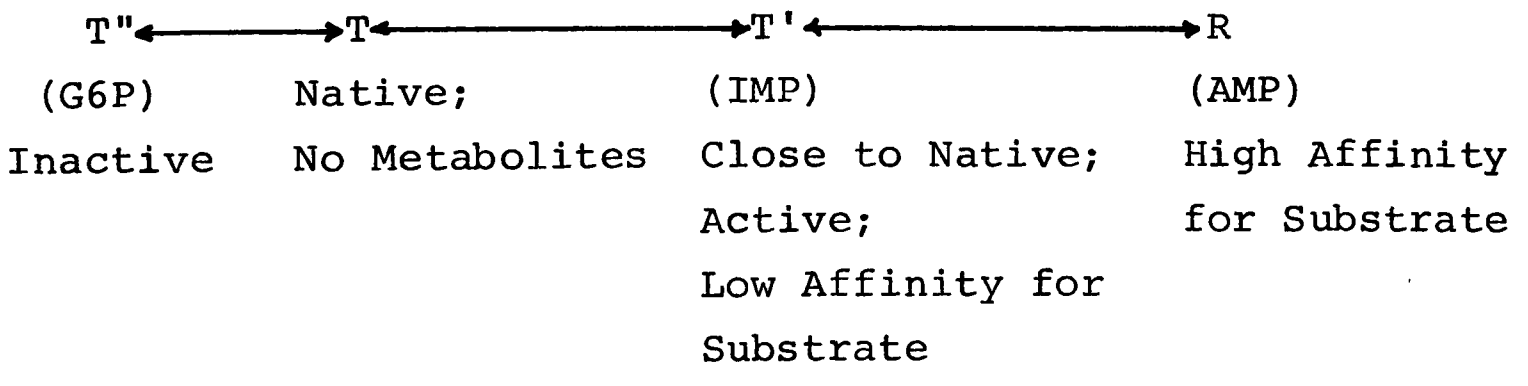
the substrate and that of enhancing the catalytic efficiency of the enzyme. Binding of the weak activator, IMP, may only affect the second stage of this process. Based on these observations, Black and Wang proposed a scheme involving four different conformational states. Madsen and Shechosky (1967) also analysed the inhibition of phosphorylase *b* by ATP, ADP and G6P according to a two-state concerted model. The results are in qualitative agreement with the model except for the effect of P_i concentration on the heterotrophic cooperativityof ATP. Helmreich *et al.* (1967) have also studied the allosteric interactions of phosphorylase *a* with AMP, GlP and glucose and concluded that they agree with the basic premises of the MWC model, although other interpretations are possible. A model involving a minimum of four conformational states has also been proposed to account for the allosteric interactions of glucose and AMP with phosphorylase *a* (Wang and Black,1968).

Crystals of phosphorylase *b* have been shown to be active (Kasvinsky and Madsen,1976), but the activity in the crystal, with 2mM IMP, is very much less than that in solution in the presence of the strong activator,AMP. Since AMP, but not the weaker activator IMP, causes dispruption of the crystal lattice, Johnson *et al.*(1979) concluded that the form of the enzyme in the crystal is closer to the T-state than the R-state. No large conformational changes were observed either in the presence of 100mM AMP or 5mM AMP+100mM arsenate, which suggests that in the tetragonal crystal form phosphorylase *b* is constrained so that it cannot respond to the allosteric activators and the T→R transition cannot take place (Johnson *et al.*, 1978). However, as described in Chapter 2, large conformational changes have been observed for phosphorylase *a*, when crystals, grown in the presence of the inhibitor glucose (and thus presumed to be in the T-conformation) are cross-linked and then

soaked in solutions of substrates (Madsen *et al.*, 1978).

As mentioned above, the MWC model makes an assumption of symmetry. The phosphorylase *b* dimer in the crystal form is symmetrical, since the 2-fold crystallographic axis is the symmetry axis of the dimer. When the crystals crack or become disordered on addition of ligands there is no evidence for a change in space group, i.e. no evidence for a change in the symmetry of the molecule. However, such evidence is far from definitive since no gross conformational changes have been observed in phosphorylase *b*. Conformational changes in phosphorylase *a* were not accompanied by a change in space group, despite a slight change in cell dimensions which suggested that the subunits of the dimer moved apart by 2Å (Madsen *et al.*, 1978).

Further evidence to suggest that a two-state model is too simple to describe the allosteric behaviour of phosphorylase *b* comes from the observation that the activator AMP and the inhibitor G6P tend to crack crystals, whilst the enzyme shows some activity in the crystal form in the presence of the weaker activator, IMP. This tends to support a model similar to that of Black and Wang in which IMP stabilizes a conformation somewhere between the 'pure' T and R states, i.e.:-



Analysis of the conformational changes occuring in the cross-linked phosphorylase *a* crystals showed that residues 1-75, 103-124, 240-287 and 383-397 exhibited extensive structural changes (Madsen

et al., 1978). Although all these residues are in the N-terminal half of the molecule, all the major ligand binding sites are linked by them. In particular the AA' helix (residues 49-75) moved inward by about 5Å along its entire length, together with the Ser 14-P, which is hydrogen-bonded to Arg 69. Johnson *et al.*(1978) have suggested a possible explanation for the different effects of AMP and ATP binding to phosphorylase, based on this movement of the AA' helix observed in phosphorylase *a*. If a similar mechanism exists for phosphorylase *b*, then the binding of the allosteric activator AMP may encourage a shift of the AA'-helix towards the BB'-helix through the interactions it makes with residues on both these helices, especially Tyr 75, Arg 309 and the main chain of Asn 44 from the symmetry-related subunit (Figure 2.7). These effects could then be transmitted to the AA' helix of the symmetry related subunit through contacts in the cap region (42 to 44), which would undoubtedly affect the AA' helix (starting at residue 50). This would provide an explanation for the homotropic cooperativity of the allosteric effector sites. On binding the inhibitor ATP, the helices might be kept apart by the presence of additional phosphate groups. Johnson *et al.* point out that although this mechanism appears plausible in the light of present results, detailed structural studies of phosphorylase *b* in its active conformation would be necessary to confirm it.

A crucial step towards the complete description of the allosteric mechanism of phosphorylase is, therefore, to obtain crystals of both phosphorylase *a* and *b* in the R-conformation. However, it has not been possible to obtain crystals of either form of the enzyme in the R-conformation by soaking in solutions of substrates and allosteric activators, presumably because the

conformational changes leading to the R-state cannot be accomodated within the existing crystal lattice. As described previously, phosphorylase *a* crystals crack when soaked in solutions of substrates (Madsen *et al.*,1978). Although phosphorylase *b* can bind the substrate GlP and the allosteric activator AMP without disruption of the crystal, no well-defined conformational changes are observed. Johnson *et al.*(1979) suggest that, when the crystals are soaked in solutions of ligands which favour a conformational change, the free energy of fracture may be related to the size and morphology of the crystals. Some crystals will become disrupted and some will remain intact without conformational changes. In the crystallographic study of protein-ligand interactions and associated conformational changes, Blake(1978) and Steitz *et al.*(1977) have emphasized the importance of co-crystallization of binary or ternary complexes, in order to overcome the interference caused by crystal lattice forces. This technique was successfully used with the allosteric protein yeast hexokinase, which was crystallized in two different crystal forms, one grown in the absence of glucose and the other grown as a glucose complex. A comparison of the two crystal structures established the existence of large conformational changes in the enzyme (Bennet and Steitz,1978). It would appear, then, that the only way to fully establish the conformational changes which take place in the phosphorylase molecule is to produce crystals of the R-state of the enzyme separately. It has not yet been possible to crystallize phosphorylase *a* from rabbit muscle in the R-conformation (i.e. in the presence of AMP and/or substrates (Fletterick and Madsen,1980) However, crystals from shark muscle phosphorylase *a* have been obtained in the presence of 30mM phosphite, 30mM maltotriose and 5% polyethylene glycol, both with and without AMP (Munn *et al.*,1979). It is likely that the enzyme is in the R-conformation since the crystals crack when transferred to a buffer

containing 50 mM glucose, an inhibitor which stabilizes the T-conformation (Fletterick and Madsen,1980). Extensive efforts to obtain crystals of phosphorylase *b* in the R-conformation are also being made (M.Sampson, personal communication).

3) CONTRIBUTION OF HIGH RESOLUTION DATA COLLECTION

Complementary to these efforts is the solution and refinement of the structure of phosphorylase *b* at 2 Å resolution, which should provide a description at the atomic level of the T-conformation of the enzyme. This will be necessary, since at 3 Å resolution, the carbonyl oxygens of peptides are not resolved and aromatic groups appear round rather than flat, preventing a complete description of the conformations of amino-acid side chains. In fact, to achieve a complete understanding of the catalytic mechanism and allosteric response, the atomic positions would need to be determined to better than 0.1 Å (Blundell and Johnson,1976,pg.420).

The second part of the work presented here has contributed to this goal through the collection and processing of the native 2 Å resolution data. In addition, this data collection has involved the use of novel techniques; synchrotron radiation and a cylindrical film cassette, which may have applications in other protein structure determinations. As described in Chapter 5, the collection of high resolution data for phosphorylase *b* presented a potentially formidible problem because of its large subunit molecular weight and large unit cell dimensions. Although the advantages of using S.R. in protein crystallography have been recognized for many years, there have been relatively few applications in structure determination to date (see Chapter 5). In this study, by solely utilizing the brightness and high degree of collimation of the S.R. beam, it was possible to record high resolution data which would have been too

weak to record using a conventional X-ray source. This work also represents the first use of a cylindrical cassette to record high resolution oscillation data for a protein with a large unit cell.

The data obtained using these techniques have been integrated and reduced to a set of unique structure factors, and the overall quality of the data appears reasonable. The major limitation of a cylindrical cassette is that it is designed for use with a particular crystal to film distance and must be positioned accurately at this distance to maintain the cylindrical geometry. The most significant errors which are likely to occur are those due to missetting of the cassette resulting in the centre of curvature no longer being coincident with the centre of the Ewald sphere. Appendix 2 gives a treatment of these errors. Refinement of the crystal to film distances in Segment 1 of OSCAR gave a maximum deviation of 0.6% from the expected value. If we consider only errors in the crystal to film distance this would give an error in the predicted film coordinates of 0.4-1%(up to 5 raster units (Appendix 2)). Such errors due to imprecise crystal to film distance, together with angular missetting of the cassette and rotation or displacement of the film in the cassette are accommodated by the Optronics orientation matrix. At present the programme uses the empirical approach described in Appendix 2, Section 4, to determine the constants of quadratic polynomials by a linear least-squares process. As described in Chapter 3, Section 4, an approximate transformation is first used to define the approximate positions of a set of fully recorded reflections.. The precise positions of about 20-40 reflections are then obtained by a centre of gravity search and the coordinates of the centroids are used as the observations in the solution of the quadratic polynomials. After transformation to the Optronics coordinate system, reflections whose observed and predicted positions differed

by more than 1 raster unit were rejected from the refinement procedure. The r.m.s. errors in the predicted X and Y coordinates of the set of reflections were never greater than 0.5 raster units.

It is possible that an improvement might be obtained by solving the non-linear equations for the error terms derived in Appendix 2. However, the present procedure appears capable of predicting the positions of reflections to better than 1 raster unit. Clearly further improvements to the quality of the reduced data set could be obtained by scaling and merging all the filmpacks together, rather than in two parts. However, this was not possible when the data were first processed, owing to limitations of computer filestore size and job-time. Further improvements could also be made in the scaling of the 2Å and 3Å data sets to remove the residual trends in the scale factor which remain for the low angle data (Figure 6.8). The most significant errors in the data remain the discrepancies between fully and partially recorded reflections, particularly for the high angle data. In addition to the methods proposed in Chapter 5, it is possible that a number of other modifications to the integration procedure could improve the quality of these data. The method of 'post-refinement', which was originally developed for the processing of data from crystalline viruses with very large unit cells, could be used to obtain improved refinements of crystal orientations and unit cell dimensions (Schutt and Winkler,1977; Winkler, Schutt and Harrison,1979; Rossmann *et al.*, 1979). The diffracted intensities from such very large unit cells are often sufficiently weak to require very long exposure times for small oscillation ranges. Radiation damage problems often mean that only one photograph can be obtained during the lifetime of the crystal. Since photographs are not contiguous, partially recorded reflections cannot be summed and must either be

discarded, or corrected to their fully recorded equivalents by a calculated degree of partiality, p_{calc}. This obviously requires a precise knowledge of the crystal orientation parameters, the unit cell dimensions and the mosaic spread. 'Post-refinement' refers to the procedure of refining these parameters by the comparison of a partial reflection on one film with a fully recorded observation of the same reflection on the same or another film. An observed degree of partiality, p_{obs}, can thus be obtained and compared to the calculated value, p_{calc}, which is derived from estimates of the crystal orientation parameters. A least-squares minimization of $\sum w(p_{obs}-p_{calc})^2$ leads to the refinement of the crystal orientation, cell-dimensions and mosaic spread. Although the post-refinement method was developed for use at an earlier stage in the data processing procedure, it could usefully be used to obtain improved orientation matrices for the phosphorylase *b* data. The observed degrees of partiality could be obtained by a comparison of the integrated intensities of a number of partially recorded reflections and their fully recorded equivalents, or the ratio of intensities of partially recorded reflections on adjacent films, and the calculated degrees of partiality obtained from the present orientation matrices.

A further improvement could also be made in the method used to integrate the weak, closely packed, high angle reflections. Rather than use a rectangular integration box, a previously determined profile could be fitted to the optical densities (Ford, 1974; Rossmann, 1979). This would increase the accuracy of integration by including information from neighbouring reflections. Since the spot shapes vary over the surface of the film, a continuously variable profile would be required; Rossmann(1979) has described a technique which assumes that the profile of each reflec-

tion is the same over limited area of the film, except for a single scale factor. Variation of the profile over the film is determined by examining the mean profile within different areas. Integration takes place in two passes over the film. During the first pass, the profiles in five areas of the film are determined and are then used to obtain a variable profile for integration in a second pass. Partial reflections cannot be determined by this profile fitting technique, since their profile will depend on how far the point has passed through the Ewald sphere, and must be integrated in the usual manner. However, if the technique of profile fitting is used in conjunction with post-refinement, it should be possible to obtain accurate estimates of the intensities of partial reflections. Rossmann's procedure involves modelling the background of a reflection as a plane. This procedure would be valuable for the phosphorylase *b* data collected with the S.R. source, in view of the contribution of diffuse scatter to the backgrounds. If profile fitting were used, more points could be included in the estimation of the background than are used in the present procedure. After calculating the background plane, each optical density point in the reflection could be compared to the model and points with significant deviation removed. The r.m.s. deviation of points from the plane would also give an improved estimate of the standard deviation of the reflection.

Computer Programmes

OSCAR — This programme, written by Dr D. G. R. Yeates, computes the integrated intensities of reflections from densitometered rotation photographs. It was modified for the cylindrical cassettes by D.L.Wild and E.A.Stura.

PERCY — This was written by Dr D. G. R. Yeates to calculate the integrated intensities of reflections from densitometered precession photographs.

FILMPACK — This was written by Dr K. S. Wilson and applies Lorentz, polarization and oblique incidence corrections, then scales and merges measurements from strong and weak films.

FILMRAD — This was written by Dr K. S. Wilson to analyse and correct for radiation damage in photographic data.

SSM — This was written by Mrs E. Dodson and scales, sorts and merges intensities from different sources. The Hamilton, Rollett and Sparks (1965) scaling procedure was used. It has been modified by Dr K. S. Wilson to sum the partially recorded terms from rotation photographs and now uses the scaling method of Fox and Holmes (1966).

TRUNCATE — This was written by Drs S. French and K. S. Wilson and produces a positive, truncated distribution of intensities.

CAD — This was written by Mrs E. Dodson and modified by Dr K. S. Wilson. It collects together in one file the data from the native, inhibitors or derivatives.

ANSC — This programme, written by Dr P. R. Evans, analyses and scales the derivative or inhibitor data to the native intensities.

FHLEKR — This was written by Dr G. Bentley, Mrs E. Dodson and Dr P. R. Evans; it scales the derivative to the native data and calculates F_{HLE} values (Arnone *et al.*, 1971).

REFINE — This version of ORFLS by Busing, Martin and Levy was modified by Mrs E. Dodson, Drs P. R. Evans and K. S. Wilson and computes a least squares refinement of heavy atom parameters for derivatives.

APPENDIX 1 (cont.)

CORDOBA — This programme, written by E.A. Stura, determines the coordinates of a number of partially recorded reflections from a digitized 'still' orientation photograph for use with Segment 1 of OSCAR.

MATIS — This programme, written by E.A. Stura, calculates orthogonalization and rotation matrices for a tetragonal or cubic crystallographic system for use in OSCAR.

APPENDIX 2

ERRORS IN THE ALIGNMENT OF THE CASSETTE

The misalignment errors considered will be:-

a) three orthogonal angular errors (δx, δy, δz)

b) translation of the origin (Δx, Δy)

c) error in the crystal to film distance, c.

1) FOR ALL TYPES OF CASSETTE (see Figure A2.1)

First consider a rotation δz which transforms the characteristic angles (γ, χ) to (γ', χ') (Figure 3.11)

$$x' = x\cos\delta z - y\sin\delta z$$

$$y' = y\cos\delta z + x\sin\delta z$$

From Figure A2.1

$$\tan\gamma' = \frac{x'}{c} = \frac{x}{c}\cos\delta z - \frac{y}{c}\sin\delta z$$

but

$$x/c = \tan\gamma \quad \text{and} \quad y = c\sec\gamma\tan\chi$$

$$\therefore \underline{\tan\gamma' = \tan\gamma\cos\delta z - \sec\gamma\tan\chi\sin\delta z} \qquad \text{A2.1}$$

similarly

$$\tan\chi' = \frac{y'}{d} = \frac{y}{d}\cos\delta z + \frac{x}{d}\sin\delta z$$

$$\underline{\tan\chi' = \tan\chi\cos\delta z + \sin\gamma\sin\delta z} \qquad \text{A2.2}$$

2) FOR FLAT PLATE CASSETTE (Figure A2.2)

Now consider the angular errors $\delta x, \delta y$ and the translation of the origin ($\Delta x, \Delta y$):-

FIGURE A2.1

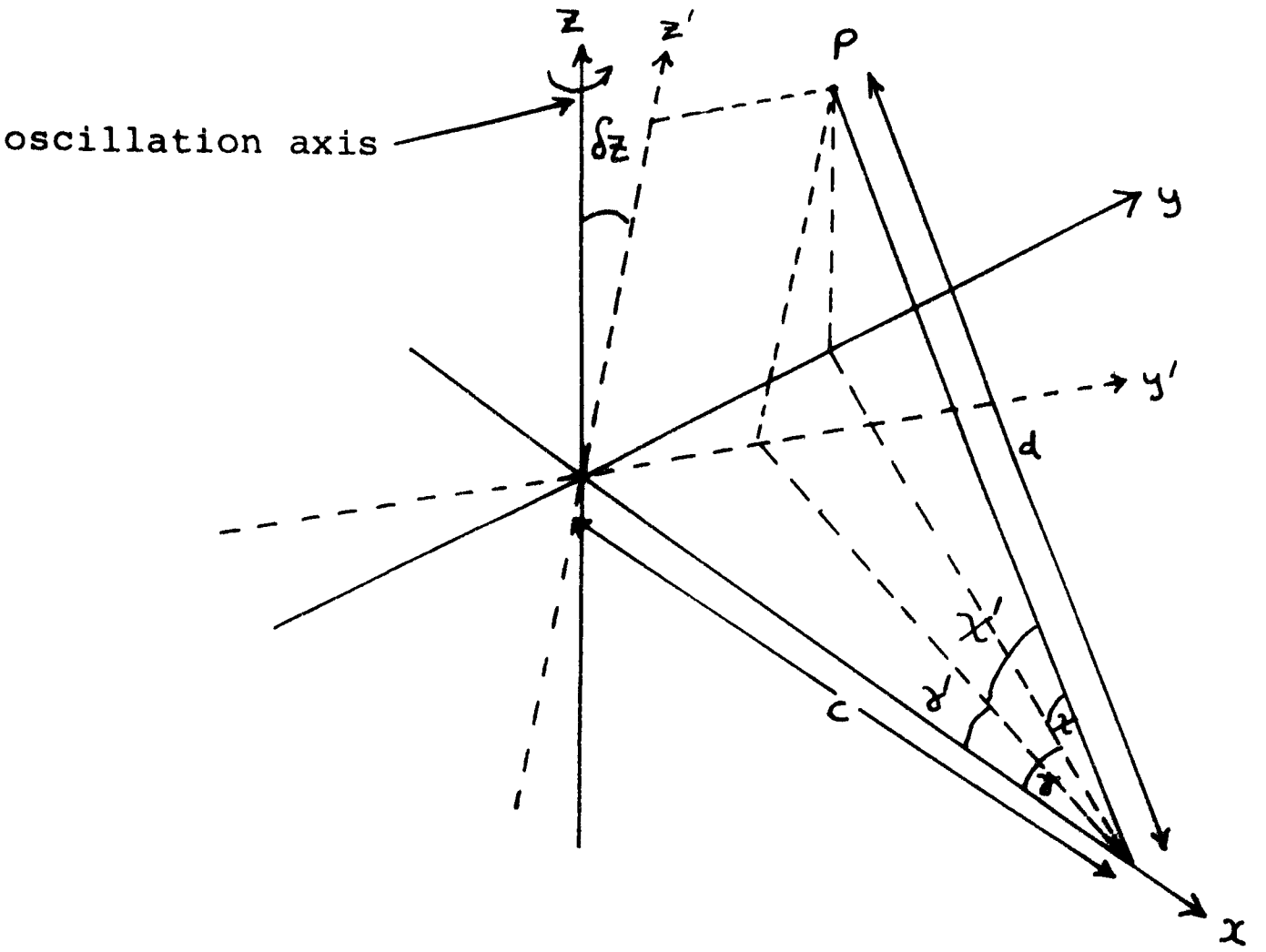
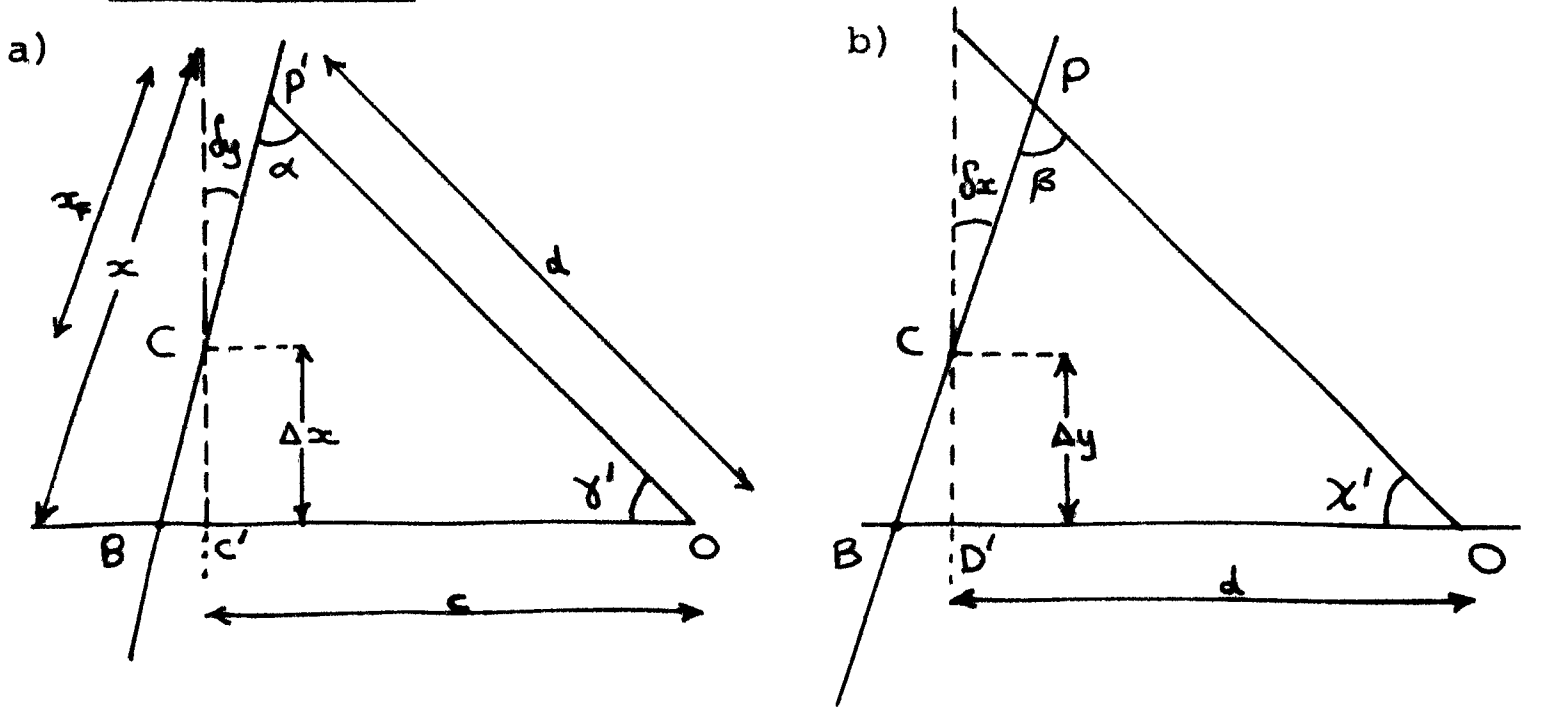


FIGURE A2.2



O = Origin of Ewald Sphere

B = Direct Beam Spot

C = Centre as defined by Fiducials

From Figure A2.2a:

$$BC' = \Delta x \tan\delta y$$

Then from $\Delta OP'B$

$$(c + \Delta x \tan\delta y) \sin\gamma' = x \sin\alpha$$

But

$$\alpha = 90 - (\gamma' - \delta y)$$

$$\Rightarrow x = (c + \Delta x \tan\delta y) \sin\gamma' \sec (\gamma' - \delta y)$$

and

$$x = x_F + CB$$

$$\therefore x_F = x - \Delta x \sec\delta y$$

$$\therefore \underline{x_F = (c + \Delta x \tan\delta y) \sin\gamma' \sec (\gamma' - \delta y) - \Delta x \sec\delta y}$$

also

$$(c - \Delta x \cot\gamma') \sin\gamma' = x_F \sin\alpha$$

$$\therefore \underline{x_F = (c - \Delta x \cot\gamma') \sin\gamma' \sec (\gamma' - \delta y)} \qquad \text{A2.3}$$

From $\Delta OBP'$

$$(c + \Delta x \tan\delta y) \sin (90 - \delta y) = d \sin\alpha$$

$$\therefore \underline{d = (c + \Delta x \tan\delta y) \cos\delta y \sec (\gamma' - \delta y)} \qquad \text{A2.4}$$

From ΔBOP (Figure A2.2b)

$$(d + \Delta y \tan\delta x) \sin\chi' = y \sin\beta$$

where

$$\beta = 180 - (\chi' + 90 - \delta x) = 90 - (\chi' - \delta x)$$

$$\therefore y = (d + \Delta y \tan\delta x) \sin\chi' \sec (\chi' - \delta x)$$

$$\therefore \underline{y_F = (d + \Delta y \tan\delta x) \sin\chi' \sec (\chi' - \delta x) - \Delta y \sec\delta x}$$

also

$$y_F \sin\beta = \sin\chi' (d - \Delta y \cot\chi')$$

$$\therefore \underline{y_F = (d - \Delta y \cot\chi') \sin\chi' \sec (\chi' - \delta x)} \qquad \text{A2.5}$$

If there are no errors in alignment,

$$\delta x = \delta y = \delta z = \Delta x = \Delta y = 0$$

$$\Rightarrow \left.\begin{array}{l} \underline{x_F = c\tan\gamma} \qquad d = c\sec\gamma \\ \underline{y_F = c\sec\gamma\tan\chi} \end{array}\right\} \text{equivalent to equations 3.9}$$

3) FOR CYLINDRICAL CASSETTE

Consider the case when the centre of curvature of the cassette is not coincident with the centre of the Ewald sphere.

(i) For case c>r (Figure A2.3a)

Film coordinate,

$$x_F = r\beta$$

now

$$\beta = \gamma' + \alpha + \delta y$$

$$\therefore x_F = r\,(\gamma' + \alpha + \delta y) \qquad \text{A2.6}$$

but

$$r\sin\alpha = QS\sin\gamma'$$

and

$$QS = Q'S' = c - S'O - RC$$

$$\therefore \alpha = \sin^{-1}\left\{\frac{\sin\gamma'}{r}\,(c - r\cos\delta y - (r\sin\delta y + \Delta x)\cot\gamma')\right\}$$

$$\Rightarrow \underline{\sin\alpha = \frac{c}{r}\sin\gamma' - \cos\delta y\sin\gamma' - (\sin\delta y + \frac{\Delta x}{r})\cos\gamma'}$$

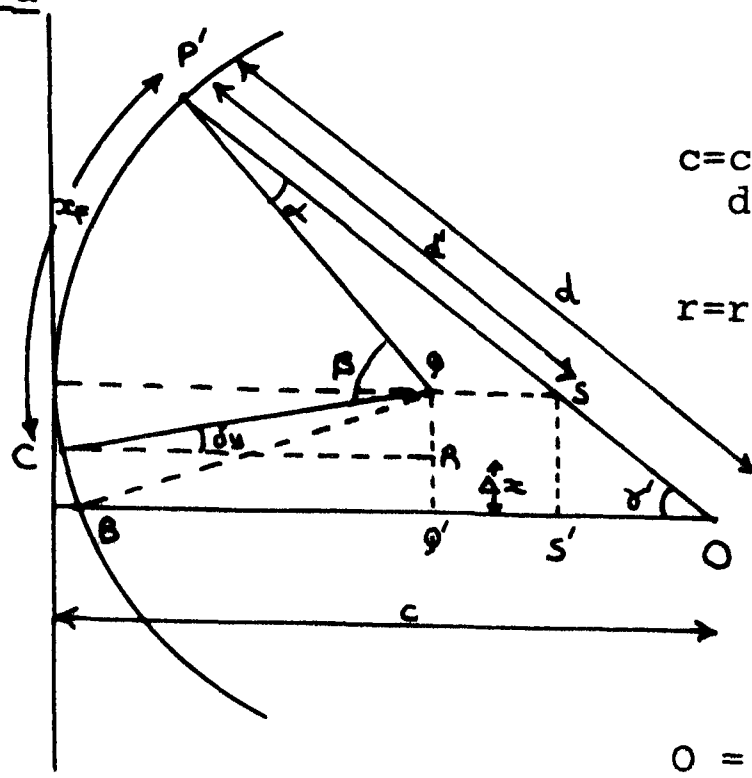
From $\Delta QSP'$

$$\frac{d}{\sin(\alpha+\gamma')} = \frac{r}{\sin\gamma'}$$

$$\therefore d = r\,\mathrm{cosec}\gamma'\,\sin(\alpha+\gamma') + SO$$

$$\therefore \underline{d = \mathrm{cosec}\gamma'\,\{r\sin(\alpha+\gamma') + r\sin\delta y + \Delta x\}} \qquad \text{A2.7}$$

FIGURE A2.3a



c=crystal to film distance

r=radius of cylinder

⊙ oscillation axis

O = Centre of Ewald Sphere

Q = Centre of Curvature of Cassette

C = Centre defined by Fiducials

B = Direct Beam Spot

FIGURE A2.3b

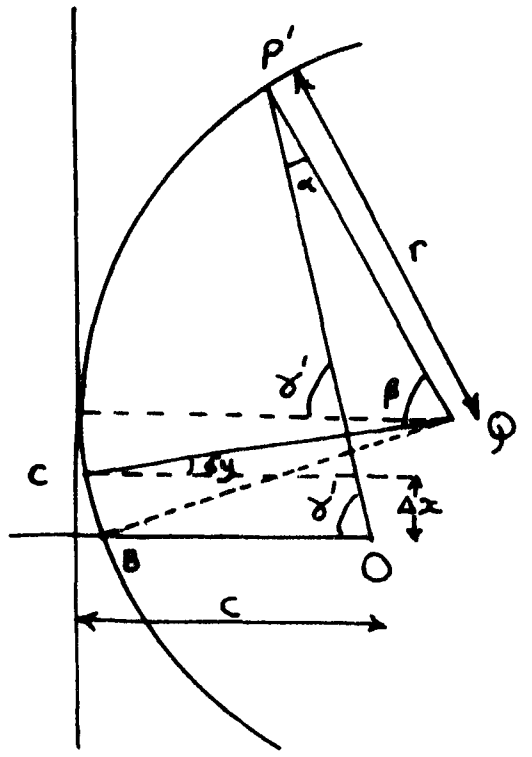
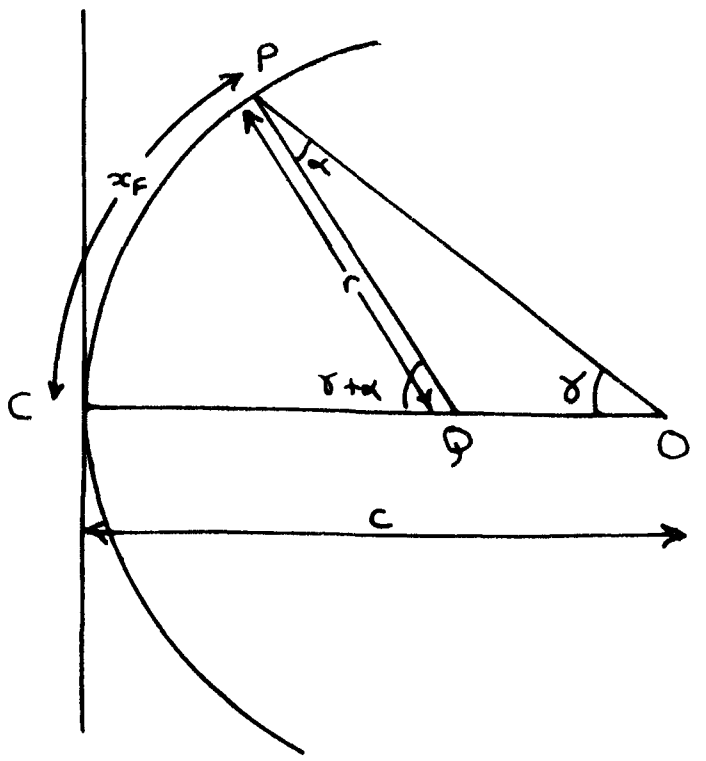


FIGURE A2.3c



We also have, as for flat plate:-

$$\underline{y_F = (d - \Delta y \cot\chi') \sin\chi' \sec (\chi' - \delta x)} \qquad \text{A2.8}$$

(ii) <u>For case c < r</u> (Figure A2.3b)

$$x_F = r\beta$$

$$\beta = \delta y + \gamma' - \alpha$$

$$\underline{x_F = r (\delta y + \gamma' - \alpha)}$$

and

$$\frac{r}{\sin\gamma'} = \frac{r\cos\delta y - c + (r\sin\delta y + \Delta x) \cot\gamma'}{\sin\alpha}$$

$$\sin\alpha = \frac{\sin\gamma'}{r} \left\{ - c + r\cos\delta y + (r\sin\delta y + \Delta x) \cot\gamma' \right\}$$

$$\underline{\equiv - \sin\alpha \text{ (for } c > r \text{)}}$$

and

$$\sin\gamma' = (r\sin(\alpha + \gamma') + r\sin\delta y + \Delta x) / d$$

hence d as before.

Thus collecting all the expressions we have:-

a) <u>Flat plate</u>

$$x_F = (c - \Delta x \cot\gamma') \sin\gamma' \sec (\gamma' - \delta y)$$

$$y_F = (d - \Delta y \cot\chi') \sin\chi' \sec (\chi' - \delta x)$$

$$d = (\Delta x + x_F \cos\delta y) \operatorname{cosec}\gamma'$$

b) <u>Cylindrical</u>

$$x_F = r (\gamma' + \delta y + \alpha)$$

$$y_F = (d - \Delta y \cot\chi') \sin\chi' \sec (\chi' - \delta x)$$

$$d = (r\sin(\alpha + \gamma') + r\sin\delta y + \Delta x) \operatorname{cosec}\gamma'$$

and

$$\sin(\alpha) = (c - r\cos\delta y - (r\sin\delta y + \Delta x) \cot\gamma') \frac{\sin\gamma'}{r}$$

C) For all cassettes

$$\tan\gamma' = \tan\gamma\cos\delta z - \sec\gamma\tan\chi\sin\delta z$$

$$\tan\chi' = \tan\chi\cos\delta z + \sin\gamma\sin\delta z$$

(iii) If we consider only errors in the crystal to film distance, and ignore translation of the origin and angular missetting errors, i.e.

$$\Delta x = \Delta y = \delta y = \delta z = 0 \qquad \text{(Figure A2.3c)}$$

$$\gamma' = \gamma$$

$$\chi' = \chi$$

then

$$x_F = r\ (\ \gamma + \propto\ ) \qquad y_F = d\tan\chi$$

$$\sin\propto = \pm\left\{\frac{c}{r}\sin\gamma - \sin\gamma\right\} = \pm\ \sin\gamma\ (\ \frac{c}{r} - 1\ )$$

$$d = r\,\mathrm{cosec}\,\gamma\,\sin\ (\propto + \gamma\ ) = \frac{r\sin\ (\propto + \gamma\ )}{\sin\gamma}$$

If $\propto \rightarrow 0$, $c \rightarrow r$

$$\Rightarrow \underline{x_F = r\gamma \qquad y_F = r\tan\chi} \qquad \text{(as in equations 3.10)}$$

The maximum error in refined crystal to film distance observed was $\pm$ 0.6% (Chapter 7),

i.e. $\frac{c}{r} = (\ 1 \pm 0.006\ )$.

This would give an error in the predicted film coordinates x_F, y_F of 0.4 → 1% (up to 5 raster units). In fact such errors are compensated for by applying correction factors (which are determined by the empirical approach described in Section 4) to the predicted reflection coordinates. This gives an Optronics orientation matrix which correctly predicts the spot positions.

4) **Mosley** (1973) has described an empirical approach in which these errors may be accommodated by correction factors applied to the predicted reflection positions. The four types of error considered for a flat plate cassette are:-

(i) displacement of the film in the x_F y_F plane, resulting in additive correction factors being applied to the theoretical reflection positions,

(ii) imprecise crystal to film distance, resulting in multiplicative correction factors,

(iii) rotation of the film in the x_F y_F plane, accommodated by the application of a rotation matrix to the theoretical positions,

(iv) film not perpendicular to the incident x-ray beam;

e.g. from equation A2.3

$$x_F = (c - \Delta x \cot\gamma')\sin\gamma' \sec (\gamma' - \delta y)$$

and if $\delta x = \Delta x = \Delta y = 0$ and $\delta z = 0$ (i.e. only errors in δy)

then $\gamma' = \gamma$

and

$$x_F = c\sin\gamma \sec (\gamma - \delta y)$$

and if $\delta y = 0$, $x_F = x_o = c\tan\gamma$

$$x_F = \frac{x_o\cos\gamma}{\cos(\gamma-\delta y)} = \frac{x_o\cos\gamma}{\cos\gamma\cos\delta y + \sin\gamma\sin\delta y}$$

$$x_F = \frac{x_o}{\cos\gamma(\cos\delta y + \tan\gamma\sin\delta y)}$$

Let

$$A = \frac{1}{\cos\delta y} \text{ and } B = \tan\delta y$$

then

$$X_F = \frac{Ax_o}{1+B\tan\gamma}$$

Assuming δy is small

$$x_F = Ax_o (1 - B\tan\gamma)$$

and

$$\tan\gamma = \frac{x_o}{c}$$

$$x_F = Ax_o (1 - \frac{Bx_o}{c})$$

$$\underline{x_F = Ax_o - Dx_o^2} \qquad \text{where } D = \frac{Bx_o}{c}$$

Similar corrections will apply for y_F.

These four types of error may, therefore, be accommodated by the use of a quadratic polynomial in the predicted reflection coordinates (X_F, Y_F in the notation of Chapter 3.4). The true coordinates in the Optronics reference frame (X_O, Y_O) are given by:-

$$X_O = aX_F^2 + bX_FY_F + cY_F^2 + dY_F + eY_F + f$$

$$Y_O = gX_F^2 + hX_FY_F + iY_F^2 + jX_F + kY_F + 1$$

where a-1 are constants for a given film. These constants may be evaluated by a least-squares approach if the values of X_O, Y_O, X_F and Y_F are known for more than six reflections.

In matrix notation:-

$$\underset{\sim}{X}_O = \underset{\sim}{O}\underset{\sim}{X}_F$$

where $\underset{\sim}{O}$ is the 2x6 Optronics orientation matrix containing the constants a-1(Chapter 3.4).

REFERENCES

1. Adams, M.J., Blundell, T.L., Dodson, E., Dodson, G.G., Vijayan, M., Baker, E.N., Harding, M.M., Hodgkin, D.C., Rimmer, B. and Sheet, S. (1969). Nature (Lond.), 224, 491.

2. Akramovic-Kicic, O., Smillie, L.B. and Madson, N.B. (1970). J. Biol. Chem., 245, 1358.

3. Arndt, U.W. (1968). Acta Cryst., B24, 1355.

4. Arndt, U.W., Champness, J.M., Phizackerley, R.P. and Wonacott, A.J. (1973). J. Appl. Cryst., 6, 457.

5. Arndt, U.W., Gilmore, D.J. and Boutle, S.H. (1972). Adv. Electron. Electron. Phys., 33B, 1069.

6. Arndt, U.W. and Wonacott, A.J. (1977). Editors. The Rotation Method in Crystallography, North-Holland, Amsterdam.

7. Arnone, A., Bier, C.J., Cotton, F.A., Day, V.W., Hazen, E.E., Richardson, D.C., Richardson, J.S., Yonath, A. (1971). J. Biol. Chem., 246, 2302.

8. Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C. and Sternberg, M.J.E. (1979). Nature, 280, 563.

9. Azaroff, L.V. (1955). Acta Cryst., 8, 701.

10. Battell, M.L., Sarkadas, C.G., Smillie, L.B. and Madsen, N.B. (1968). J. Biol. Chem., 243, 6202.

11. Biesecker, G., Harris, J.I., Thierry, J.-C., Walker, J.E. and Wonacott, A.J. (1977). Nature (London), 266, 328.

12. Black, W.J. and Wang, J.H. (1968). J. Biol. Chem., 243, 5892.

13. Blake, C.C.F. and Evans, P.R. (1974). J. Mol. Biol. 84, 585.

14. Blake, C.C.F., Johnson, L.N., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1967). Proc. Roy. Soc. Ser. B, 167, 378-388.

15. Blake, C.C.F. and Phillips, D.C. (1962). in: Biological Effects of Ionizing Radiation at the Molecular Level, pp. 183-191. IAEA Symposium, Vienna.

16. Blow, D.M. (1958). Proc. Roy. Soc., A247, 302.

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

18. Blow, D.M. and Rossmann, M.G. (1961). Acta Cryst., 14, 1195.

19. Blundell, T.L., Dodson, E.J., Dodson, G.G., Hodgkin, D.C. and Vijayan, M. (1971). Contemporary Phys., 12, 209.

20. Blundell, T. and Johnson, L.N. (1976). Protein Crystallography, Academic Press (London).

21. Branden, C.I., Eklund, H., Nordstrom, B., Boiwe, T., Soderlund, G., Zeppezauer, E., Ohlssan, I. and Akeson, A. (1973). Proc. Nat. Acad. Sci. U.S.A., 70, 2439.

22. Childress, C.C. and Sacktor, B. (1970). J. Biol. Chem., 245, 2927.

23. Cork, C., Hamlin, R., Vernon, W.; and Xuong, N. (1975). Acta Cryst., A31, 702.

24. Cox, E.G. and Shaw, W.F.B. (1930). Proc. Roy. Soc. London, 127, 71.

25. Crick, F.H.C. and Magdoff, B. (1956). Acta Cryst., 9, 901.

26. Debye, P. (1913). Verh. d. Deutsch. Phys. Ges., 15, 678, 738, 857.

27. Debye, P. (1914). Ann. d. Physik, 43, 49.

28. Dodson, E.J., Evans, P.R. and French, S. (1975). in: Anomalous Scattering, p. 423 (Eds: Ramaseshan, S. and Abramsons, S.C.), Munksgaard, Copenhagen.

29. Eagles, P.A.M., Iqbal, M., Johnson, C.N., Mosley, J. and Wilson, K.S. (1972). J. Mol. Biol., 71, 803.

30. Evans, P.R. (1973). D. Phil. Thesis, University of Oxford.

31. Feldman, K. and Helmreich, E.J. (1976). Biochemistry, 15, 2394-2401.

32. Feldman, K., Hörl, M., Klein, H.W. and Helreich, E.J.M. (1978). Proc. 11th FEBS meeting, Copenhagen, 1977, Vol. 42 (ed. Esmann, V.), 205-218 (Pergamon, Oxford 1978).

33. Feldman, K. and Hull, W.E. (1977). Proc. Nat. Acad. Sci. (U.S.A.), 74, 856-860.

34. Fischer, E.H., Keut, A.B., Snyder, E.R. and Krebs, E.G. (1958). J. Am. Chem. Soc., 80, 2906.

35. Fischer, E.H., Pocker, A., and Saari, J.C. (1970). Essays in Biochemistry, 6, 23.

36. Fischer, E.H., Heilmeyer, L.M.G. and Haschke, R.H. (1971). Current Topics in Cellular Regulation, 4, 211.

37. Fletterick, R.J., and Madsen, N.B. (1977). Trends in Biochemicals Sciences, 2, 145-148.

38. Fletterick, R.J. and Madsen, N.B. (1980). Ann. Rev. Biochem., 49, 31-61.

39. Fletterick, R.J., Sygusch, J., Murray, N., Madsen, N.B. and Johnson, L.N. (1976a). J. Mol. Biol., 103, 1-13.

40. Fletterick, R.J., Sygusch, J., Semple, M. and Madsen, N.B. (1976b). J. Biol. Chem. 251, 6142-6146.

41. Fourme, R., Kahn, R., Gadet, A., André, D., Janin, J. and Risler, J.L. (1977). Paper PI.10, 4th European Crystallographic Meeting, Oxford.

42. **Fourme**, R. and Kahn, R. (1981). In preparation.

43. **Fox**, G.C. and Holmes, K.C. (1966). Acta Cryst., 20, 886.

44. Frauenfelder, H., Petsko, G.A. and Tsernoglou, D. (1979). Nature, 280, 558.

45. French, S. and Wilson, K.S. (1978). Acta Cryst., A34, 517.

46. Graves, D.J., Parrish, R.F., Ulring, R.J. and Korytnyk, W. (1978). Proc. 11th FEBS Meeting, Copenhagen, 1977, Vol. 42 (ed. Esmann, V.), (195-204) (Pergamon, Oxford 1978).

47. Graves, D.J. and Wang, J.H. (1972). in The Enzymes, Vol. VII, 3rd edition, Academic Press, P435.

48. Green, D.W., Ingram, V.M. and Perutz, M.F. (1954). Proc. Roy. Soc., A225, 287.

49. Harding, M.M. (1962). D. Phil. Thesis, Oxford University.

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

51. Harmsen, A., Leberman, R. and Schultz, G.E. (1976). J. Mol. Biol., 104, 311.

52. Harrison, S.C. (1973). Research Applications of Synchrotron Radiation, pp. 105-108. Eds. Watson, R.W. and Perlman, M.L. (Publication BNL 50381, Brookhaven National Laboratory, Upton, N.Y.).

53. Haschke, R.H., Heilmeyer, L.M.G., Meyer, F. and Fischer, E.H. (1970). J. Biol. Chem., 245, 6657.

54. Hashizume, H., Kohra, K. and Oguro, Y. (1975). J. Appl. Cryst., 8, 249.

55. Henderson, R.C. and Moffat, J.K. (1971). Acta Cryst., B27, 1414.

56. Hendrickson, W.A., Love, W.E. and Karle, J. (1973). J. Mol. Biol., 74, 331.

57. Herriott, J.R., Sieker, L.C., Jensen, L.H. and Lovenberg, W. (1970). J. Mol. Biol. 50, 391.

58. Hers, H.G. (1976). Ann. Rev. Biochem., 45, 167-189.

59. Honikel, K.O. and Madsen, N.B. (1973). Can. J. Biochem., 51, 344-356.

60. Hoppe, W. and Jakubowski, U. (1975). Anomalous Scattering, ed: Ramaseshan, S. and Abrahams, S.C., pp. 437-461. Munksgaard, Copenhagen.

61. Huber, R., Epp, O. and Formanek, H. (1969a). J. Mol. Biol., 42, 591.

62. Huber, R., Epp, O. and Formanek, H. (1969b). Naturwissenschaften, 56, 362.

63. Irwin, M.J., Nyborg, U., Reid, B.R. and Blow, D.M. (1976). J. Mol. Biol., 105, 577.

64. James, R.W. (1948). The Optical Principles of the Diffraction of X-rays, Chapter 5. G. Bell and Sons, London.

65. Johnson, L.N., Jenkins, J.A., Wilson, K.S., Stura, E.A. and Zanotti, G. (1980). J. Mol. Biol., 140, 565-580.

66. Johnson, L.N., Madsen, N.B., Mosley, J. and Wilson, K.S. (1974). J. Mol. Biol., 90, 703.

67. Johnson, L.N., Stura, E.A., Wilson, K.S., Sansom, M.S.P. and Weber, I.T. (1979). J. Mol. Biol., 134, 639-653.

68. Johnson, L.N., Weber, I.T., Wild, D.L., Wilson, K.S. and Yeates, D.G.R. (1978a). J. Mol. Biol., 118, 579-591.

69. Johnson, L.N., Weber, I.T., Wild, D.L., Wilson, K.S. and Yeates, D.G.R. (1978b). Proc. 11th FEBS meeting, Copenhagen, 1977, Vol. 42 (ed. Esmann, V.), 185-194 (Pergamon, Oxford).

70. Jones, D.C. and Cowgill, R.W. (1971). Biochemistry, 10, 4276-4282.

71. Kartha, G. and Parthasarthy, R. (1965a). Acta Cryst., 18, 745.

72. Kartha, G. and Parthasarthy, R. (1965b). Acta Cryst., 18, 749.

73. Kasvinsky, P.J., Madsen, N.B., Fletterick, R.J. and Sygusch, J. (1978a). J. Biol. Chem., 253, 1290-1296.

74. Kasvinsky, P.J., Madsen, N.B. Sygusch, J. and Fletterick, R.J. (1978b). J. Biol. Chem., 253, 3443-3351.

75. Katayama, C., Sakabe, N. and Sakabe, K. (1972). Acta Cryst., A28, 293.

76. Konnert, J.H. (1976). Acta Cryst., A32, 614.

77. Kopfmann, G. and Huber, R. (1968). Acta Cryst., A24, 348.

78. Koshland, D.E. (1953). Biol. Rev., 28, 416-436.

79. Kraut, J., Sieker, L.C., High, D.F. and Freer, S.T. (1962). Proc. Nat. Acad. Sci. U.S.A., 48, 1417.

80. Lemonnier, M., Fourme, R., Rousseaux, F. and Kahn, R. (1978). Nuclear Instruments and Methods, 152, 173.

81. Lerch, K. and Fischer, E.H. (1975). Biochemistry, 14, 2009-2014.

82. Levine, M., Muirhead, H., Stammers, D.K. and Stuart, D.I. (1977). Nature (London), 271, 626.

83. Levitt, M. and Chothia, C. (1976). Nature, 261, 552-558.

84. Amorós, J. and Amorós, M. (1968). Molecular Crystals: their transforms and diffuse scattering, John Wiley, New York.

85. Luzzati, V. (1953). Acta Cryst., 6, 142.

86. Madsen, N.B. and Cori, C.F. (1956). J. Biol. Chem., 223, 1055.

87. Madsen, N.B., Kasvinsky, P.J. and Fletterick, R.J. (1978). J. Biol. Chem., 253, 9097-9101.

88. Matthews, B.W. (1966a). Acta Cryst., 20, 230.

89. Matthews, B.W. (1966b). Acta Cryst., 20, 82.

90. Matthews, B.W. and Czerwinski, E.W. (1975). Acta Cryst., A31, 480.

91. Matthews, B.W., Klopfenstein, C.E. and Colman, P.M. (1972). J. Sci. Inst., 5, 353.

92. Metzger, B., Helmreich, E., and Glaser, L. (1967). Proc. Nat. Acad. Sci. (U.S.A.), 57, 994.

93. Meyer, F., Heilmeyer, L.M.G., Haschke, R.H. and Fischer, E.H. (1970). J. Biol. Chem., 245, 6642.

94. Milledge, H.J. (1966). Acta Cryst., 21, A220.

95. Minor, T.C., Milch, J.R. and Reynolds, G.T. (1974). J. Appl. Cryst., 7, 323.

96. Monod, J., Wyman, J., and Changeux, J.-P. (1965). J. Mol. Biol., 12, 88-118.

97. Morimoto, H. and Uyeda, R. (1963). Acta Cryst., 16, 1107.

98. North, A.C.T. (1965). Acta Cryst., 18, 212.

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

100. Okazaki, T., Nakazawa, A. and Hayaishi, O. (1968). J. Biol. Chem., 243, 5266.

101. Palm, D., Schachtele, K.-H., Feldmann, K. and Helmreich, E.J.H. (1979). FEBS Letters, 101, 403-406.

102. Pearson, R.J. (1963). J. Am. Chem. Soc., 85, 3533.

103. Peerdeman, A.F., Bommel, A.J.V. and Bijvoet, J.M. (1951). Proc. Acad. Sci. Amst., 54, 16; Nature, London, 165, 271.

104. Perutz, M.F. (1956). Acta Cryst., 9, 867.

105. Petsko, G.A. (1973). D. Phil. Thesis, Oxford University.

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

107. Phillips, J.C., Wlodawer, A., Yevitz, M.M. and Hodgson, K.O. (1976). Proc. Natl. Acad. Sci. U.S.A., 73, 128.

108. Phillips, J.C., Wlodawer, A., Goodfellow, J.M., Watenpaugh, K.D., Sieker, L.C., Jensen, L.H. and Hodgson, K.O. (1977). Acta Cryst., A33, 445.

109. Rahim, Z.H.A., Perrett, D. and Griffiths, J.R. (1976). FEBS Lett., 69, 203.

110. Rahim, Z.H.A., Perrett, D., Lutaya, G. and Griffiths, J.R. (1980). Biochem. J., 186, 331.

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

112. Rossmann, M.G., Liljas, A., Branden, C.-I. and Banaszah, L.J. (1975). in The Enzymes (ed. Boyer, P.D.), 3rd Ed., Vol. II, pp. 62-94. Academic Press, New York.

113. Schwager, P., Bartels, K. and Huber, R. (1973). Acta Cryst., A29, 291.

114. Shaltiel, S. and Cortijo, M. (1970). Biochem. Biophys. Res. Commun., 41, 594-600.

115. Sieker, L.C., Adman, E. and Jensen, L.H. (1972). Nature, London, 235, 40.

116. Sprang, S. and Fletterick, R.J. (1979). J. Mol. Biol., 131, 523-551.

117. Sternberg, M.J.E., Grace, D.E.P. and Phillips, D.C. (1979). J. Mol. Biol., 130, 231.

118. Stout, G.H. and Jensen, L.H. (1968). X-ray Structure Determination, Macmillan, New York.

119. Stuart, D. and Walker, N. (1979). Acta Cryst., A35, 925.

120. Stuhrmann, H.B. (1978). Quarterly Reviews of Biophysics, 11, 71.

121. Stura, E.A. (1981). D. Phil. Thesis, University of Oxford, (to be submitted).

122. Thomas, D.A. and Wright, B.E. (1976). J. Biol. Chem., 251, 1253.

123. Thomas, R. and Hamilton, W.C. (1972). Acta Cryst., A28, Paper XXIV-17.

124. Titani, K., Koide, A., Hermann, J., Ericsson, L.H., Kumar, S., Wade, R.D., Walsh, K.A., Neurath, H. and Fischer, E.H. (1977). Proc. Nat. Acad. Sci. (U.S.A.), 74, 4762-4766.

125. Topping, J. (1965). Errors of observation and their treatment, Chapman and Hall, London.

126. Weber, I.T. (1977). D. Phil. Thesis, University of Oxford.

127. Weber, I.T., Johnson, L.N., Wilson, K.S., Yeates, D.G.R., Wild, D.L. and Jenkins, J.A. (1978), Nature, 274, 433-437.

128. Whittaker, E.J.W. (1953). Acta Cryst., 6, 218.

129. Wierenga, R.K. (1978). Ph.D. Thesis, University of Groningen, Netherlands.

130. Willis, B.T.M. and Pryor, A.W. (1975). Thermal Vibrations in Crystallography, Cambridge.

131. Wilson, A.J.C. (1949). Acta Cryst., 2, 318.

132. Wilson, A.J.C. (1950). Acta Cryst., 3, 397.

133. Wilson, K.S. (1974). D. Phil. Thesis, University of Oxford.

134. Wilson, K.S. (1978). Acta Cryst., B34, 1599-1608.

135. Wilson, K.S., Stura, E.A., Wild, D.L., Todd, R.J., Babu, Y.S., Jenkins, J.A., Mercola, D.A., Johnson, L.N., Fourme, R. and Kahn, R. (1981). In preparation.

136. Wilson, K.S. and Yeates, D. (1979). Acta Cryst., A35, 146.

137. Wooster, W.A. (1964). Acta Cryst., 17, 878.

138. Wyckoff, H.W. (1973). Research Applications of Synchrotron Radiation, pp. 133-138. Eds. Watson, R.W. and Perlman, M.L. (publication BNL 50381, Brookhaven National Laboratory, Upton, N.Y.).

139. Xuong, N. and Freer, S.T. (1971). Acta Cryst., B27, 2380.

140. Xuong, N., Kraut, J., Seely, O., Freer, W.T. and Wright, C.S. (1968). Acta Cryst., B24, 289.

ADDITIONAL REFERENCES FOR CHAPTER 7

141. Bennett, W.S. and Steitz, T.A. (1978). Proc.Nat. Acad. Sci. (U.S.A.), 75, 4848.

142. Blake, C.C.F. (1978). Nature, 271, 707.

143. Evans, P.R. and Hudson, P.J. (1979). Nature, 279, 500.

144. Ford, G.C. (1974). J.Appl. Cryst. 7, 555.

145. Helmreich, E., Michaelides, M.C. and Cori, C.F. (1967). Biochemistry, 6, 3695.

146. Kasvinsky, P.J. and Madsen, N.B. (1976). J.Biol. Chem., 251, 6852.

147. Koshland, D.E., Némethy , Jr.G. and Filmer, D.(1966). Biochemistry, 5, 365.

148. Madsen, N.B. and Schechosky, S. (1967). J.Biol. Chem., 242, 3301.

149. Monaco, H.L., Crawford, J.L. and Lipscomb, W.N. (1978). Proc. Nat. Acad. Sci. (U.S.A.), 75, 5276.

150. Mosley, J. (1973). D.Phil. Thesis, University of Oxford.

151. Munn, M., Sprang, S. and Fletterick, R.J. (1979). Am. Crystallogr. Assoc. Meet. Hawaii, 6, NO.2, 63 (Abstr.).

152. Perutz, M.F. (1979). Ann. Rev. Biochem., 48, 327.

153. Rossmann, M.G. (1979). J. Appl. Cryst., 12, 225.

154. Rossmann, M.G., Leslie, A.G.W., Abdel-Meguid S.S. and Tsukihara, T. (1979). J. Appl. Cryst., 12, 570.

155. Schutt, C. and Winkler, F.K. (1977). In The Rotation Method in Crystallography. (eds. Ardnt, U.W. and Wonacott, A.J.), pp 173-186. North-Holland, Amsterdam.

156. Steitz, T.A., Anderson, W.F., Fletterick, R.J., and Anderson, C.M. (1977). J. Biol. Chem., 252, 4494.

157. Wang, J.H. and Black, W.J. (1968). J. Biol. Chem.,243, 5266.

158. Winkler, F.K., Schutt, C.E. and Harrison, S.C. (1979). Acta. Cryst., A35, 901.