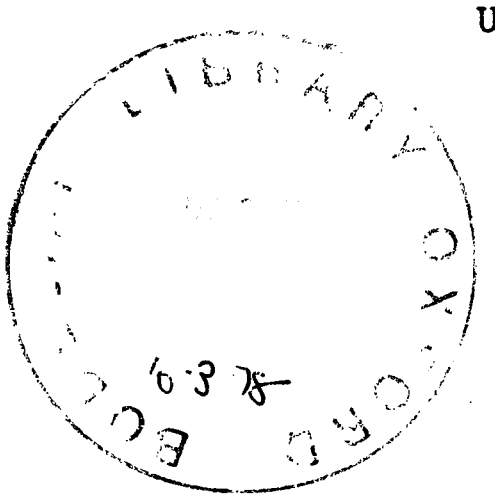
X-RAY STUDIES CONCERNING THE STRUCTURE OF 6-PHOSPHOGLUCONATE DEHYDROGENASE

Thesis submitted for the degree of

Doctor of Philosophy in the

University of Oxford

December 1977



John Richard Helliwell

Laboratory of Molecular Biophysics

Department of Zoology

Balliol College

To my Mother and Father

and Madeleine

Abstract

X-ray studies concerning the structure of 6-phosphogluconate dehydrogenase
John Richard Helliwell, Balliol College,
D.Phil., Michaelmas 1977.

This thesis is concerned with the structure determination of 6-phosphogluconate dehydrogenase (6-phospho-D-gluconate:$NADP^+$ oxido-reductase (decarboxylating), E.C.1.1.1.44) from sheep liver. This protein is the second enzyme in the pentose phosphate pathway and catalyses the conversion of 6-phosphogluconate to ribulose-5-phosphate and CO_2.

Chapter 1 gives an introduction to the chemical properties of 6-phospho-gluconate dehydrogenase (6PGDH) and compares these properties for the enzyme from a variety of sources. The common structural features of the NAD^+-dependent dehydrogenases are discussed in terms of divergent and convergent evolution. Details of an $NADP^+$-dependent enzyme, E.coli dihydrofolate reductase are also given. This is the only published high resolution structure of an $NADP^+$-dependent enzyme to date (Matthews et al. (1977)).

The crystallographic techniques used in the course of this work are discussed in Chapter 2. Particular attention is paid to the application of synchrotron radiation in protein crystallography in optimising heavy atom anomalous scattering. Appendix I also gives some experimental details relevant to this.

Chapters 3, 4, and 5 give details of the structural results. In Chapter 3 the low resolution (6 Å) structure is presented with a full analysis of the derivatives ($KAu(CN)_2$ and $K_2Pt(CN)_4$) used in the double derivative phase determination. The 6 Å electron density map could be interpreted. It showed the overall shape of the monomer and enabled the dimer axis to be identified. It also yielded the structurally interesting features of several long columns of density of the correct cross section for α-helix and also a cleft region. Chapter 4 is concerned with the determination of the coenzyme ($NADP^+$) binding position at 6 Å resolution. Difference electron density maps calculated from X-ray measurements on crystals soaked in the oxidised coenzyme indicated the binding of $NADP^+$ in the cleft region. Also, by using an iodinated derivative of the coenzyme, iodopyridine adenine dinucleotide phosphate ($IADP^+$), it was possible to identify the nicotinamide moiety of the coenzyme and thus orient the coenzyme with respect to the enzyme. Chapter 5 describes the screenless rotation method and its application to the 2.6 Å crystallographic study of 6PGDH. In particular, the collection of the data for the native enzyme crystals and the $K_2Pt(CN)_4$ derivative. The phases calculated at 6 Å with two derivatives (as described in § 3) were extended to 2.7 Å by calculating phases, using the photographic data, by the method of single isomorphous replacement with anomalous scattering. This enabled an electron density map to be calculated to 2.7 Å. Preliminary details of the quality of this electron density map are given.

CONTENTS

		PAGE
Chapter 1	INTRODUCTION	1
	1.1 Pentose phosphate pathway	1
	1.2 Preparation of the sheep liver enzyme	2
	1.3 Properties of the enzyme	2
	1.4 Reasons for studying the enzyme	5
	1.5 Common features in protein three-dimensional structure with particular reference to the dehydrogenases	6
Chapter 2	X-RAY STRUCTURAL STUDIES OF PROTEINS	13
	2.1 Isomorphous replacement	16
	2.2 Methods of intensity measurement	17
	2.3 Derivation of structure factors from intensities	28
	2.4 Location of heavy atom sites	32
	2.5 Heavy atom parameter refinement	37
	2.6 Calculation of the primary phase set	41
	2.7 The difference Fourier technique in protein crystallography	43
Chapter 3	PRELIMINARY CRYSTALLOGRAPHIC WORK AND 6 Å STRUCTURE DETERMINATION OF 6PGDH FROM SHEEP LIVER	45
	3.1 Crystallization	45
	3.2 Preliminary crystallography and derivative search	45
	3.3 Data collection and processing	48
	3.4 Location of the heavy atom positions	50
	3.5 Refinement of the heavy atom parameters	56
	3.6 Phase calculation	60
	3.7 The electron density map	62
Chapter 4	STUDIES OF COENZYME AND SUBSTRATE BINDING TO 6PGDH AT 6 Å RESOLUTION	66
	4.1 Soaking conditions	67
	4.2 Data collection and processing	68

	4.3 Difference Fourier maps	74
	4.4 Conclusions and further work	77
Chapter 5	THE STRUCTURE DETERMINATION OF 6PGDH TO 2.6 Å RESOLUTION	79
	5.1 The screenless rotation method	80
	5.2 Film processing	84
	5.3 High resolution data collection from 6PGDH crystals	97
	5.4 Processing the 6PGDH data films	101
	5.5 Analysis of the derivative	108
	5.6 Phasing to 2.7 Å	117
	5.7 The electron density map to a resolution of 2.7 Å	119
	5.8 Conclusion	120
Appendix I	APPLICATION OF SYNCHROTRON RADIATION TO PROTEIN CRYSTALLOGRAPHY. PRELIMINARY EXPERIMENTS ON 6PGDH CRYSTALS USING THE NINA SYNCHROTRON, DARESBURY, U.K.	122
Appendix II	Adams, M.J., Helliwell J.R., Bugg C.E. (1977).	125
Appendix III	Adams M.J., Archibald I.G., Helliwell J.R. (1977)	126
Appendix IV	Computer programs	127
Appendix V	Structure factor tables	130
References		132

Acknowledgments

Foremost I would like to thank my supervisor Dr. Margaret Adams for her help, encouragement and friendship and for passing on the benefit of her experience during the past three years.

A special thank you to the other members of the 6PGDH group: Ian Archibald and Sue Jenkins (neé Clift) for their friendship and the enjoyment of working with them. Also, the pleasure of working with Dr. Charlie Bugg of the University of Alabama during 1974-5 and Dr. Don Voet of the University of Pennsylvania during 1976, who also spent time on the project.

I would also like to mention Professor David Phillips for the pleasure of working in his laboratory. Keith Wilson for very helpful computing advice and many stimulating discussions. Also, David Yeates and Stuart Oatley for further computing advice, David Rice and Mike Sternberg for useful discussions. Guy and Eleanor Dodson for keeping alive my enthusiasm for the subject. Dr. Peter Main of the University of York for stimulating my interest in Biophysics in the first place. Dr. Juan Bordas of the Daresbury Laboratory for help with the synchrotron work. Mr. Ray Harris of the Oxford Enzyme Preparation Laboratory for preparing the enzyme. Dr. C.E. Phelps of the University Computing Laboratory for help in scheduling jobs in the last three months. Also my contemporaries, Diana Grace, Irene Weber, Christine Spencer, Steve White and Mike Sternberg for their friendship during these three years in Oxford. The members of the laboratory cricket team with whom I enjoyed many hours on the field this last summer. Good luck, next year in the Science Area league.

Also, Mike Pickford, John Marsh and Phil Keighley for expert technical assistance and Chris Blackmore for his excellent photography.

Mrs. Felicity Jenkins for turning "a seemingly illegible script" into something hopefully meaningful.

Dr. John Jones, my college supervisor, for his advice.

The Science Research Council for a studentship for the period 1974-77; Balliol College for a Graduate Scholarship for the sessions 1975-76 and 1976-77 as well as financial help towards the cost of conferences and my thesis; the Graduate Studies Committee of this University for financial assistance for one extra term.

Finally, I would like to thank my fiancee Miss Madeleine Berry for making this last year in Oxford such a wonderful time.

This project has involved several people and certain pieces of work have been utilised by myself and presented in this thesis. I would like to acknowledge the work of Dr. C.E.Bugg on the data collection and analysis of the Au derivative of 6PGDH at 6 Å resolution, described in §3. The balsa wood model was built by Miss Janice Baines.

Several people were involved in the data collection and processing of the high resolution data, though most of the work was done by me. A detailed breakdown of the work done by myself (JRH), Dr. Margaret Adams (MJA), Ian G. Archibald (IGA) and Dr. D. Voet (DV) is given in Tables 5.2 and 5.4.

Unless otherwise stated the work described in this thesis is my own and has not been submitted previously for a degree in this or any other university.

15-12-77

Symbols and abbreviations

h, k, l	Miller indices
$\underline{a}$, $\underline{b}$, $\underline{c}$	direct cell vectors
$\underline{a}^*$, $\underline{b}^*$, $\underline{c}^*$	reciprocal cell vectors
V	unit cell volume
V_x	crystal volume
V''	empirical cell volume used in FFT
I or J	intensity of a reflection
$F_{\underline{h}}$	structure factor
$\underline{F}_P$	structure factor for native protein
$\underline{F}_{PH}$	structure factor for a derivative
$\underline{F}_{PH}^+$, $\underline{F}_{PH}^-$	structure factors for Friedel related reflections
$\Delta_{ANO} = \lvert\underline{F}_{PH}^+\rvert - \lvert\underline{F}_{PH}^-\rvert$	anomalous difference
$\Delta_{ISO} = \lvert\underline{F}_{PH}\rvert - \lvert\underline{F}_P\rvert$	isomorphous difference
$\underline{F}_{HCALC}$	calculated heavy atom structure factor
F_{HLE}	estimate of $\lvert\underline{F}_H\rvert$, the heavy atom structure amplitude
F_{HLEB}	bias corrected F_{HLE}
K_K	Kraut scale between derivative and native (N_V) data
$E_{\underline{h}}$	normalized structure factor
α_P	native phase angle
m	figure of merit
α_{PH}	derivative phase angle
α_H	heavy atom phase angle
ε	lack of closure error
E (E")	average error used in phasing
$\rho(\underline{x})$ or $\rho(\underline{r})$	electron density at $\underline{x}$ or $\underline{r}$ in the unit cell
$\langle\Delta\rho^2\rangle^{\frac{1}{2}}$	root mean square error in electron density
θ	Bragg angle
$f = f(\theta)$	atomic scattering factor

f', f''	real and imaginary part of the anomalous scattering
o_j'	heavy atom occupancy for site j
λ	wavelength
B	temperature factor
P	polarization factor
L	Lorentz factor
T_{hkl}	transmission factor for a reflection hkl
A_{hkl}	absorption factor for a reflection hkl
$\bar{\mu}$	precession angle
$\Theta, \omega, \chi, \varphi$	4 circle diffractometer angles
ξ, ζ, τ	cylindrical coordinates of a reciprocal lattice point
φ	rotation angle in the screenless rotation method
γ	reflecting range of the crystal

6PGDH	6-phosphogluconate dehydrogenase
6PG	6-phosphogluconate
R-5-P	ribulose-5-phosphate
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidized form)
NADPH	ditto (reduced form)
NAD	nicotinamide adenine dinucleotide
ADP	adenosine diphosphate
ATP	adenosine triphosphate
PGK	phosphoglycerate kinase
LDH	lactate dehydrogenase
MDH	malate dehydrogenase
LADH	liver alcohol dehydrogenase
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
DH	a dehydrogenase
RNA	ribonucleic acid
DNA	deoxyribonucleic acid

Ⓟ	phosphate
M	figure of topological relatedness
M.W. or m.w.	molecular weight
K_m	Michaelis constant
K_I	inhibition constant

All other symbols are as defined, in context, in the text.

CHAPTER I

INTRODUCTION

6-phosphogluconate dehydrogenase (6-phospho-D-gluconate: $NADP^+$ oxido-reductase (decarboxylating), EC 1.1.1.44) is the second enzyme in the pentose phosphate pathway. 6PGDH catalyzes the conversion of 6-phosphogluconate (6PG) to ribulose-5-phosphate (R-5-P) and CO_2 utilizing the coenzyme nicotinamide adenine dinucleotide phosphate ($NADP^+$ oxidized, NADPH reduced form),see Figure 1.1. Two orientations of the hydrogen of the NADPH transferred in the oxidation-reduction reaction are possible and commonly called H_A (or H(R)) and H_B (or H(S)) orientations (Figure 1.2). Different dehydrogenases are either A or B side specific. For instance, LDH, LADH and MDH, which all utilize NAD^+, are H_A specific whereas GAPDH and 6PGDH are H_B specific. GAPDH binds NAD^+ and 6PGDH, $NADP^+$.

Other oxidative decarboxylating enzymes do exist; malic enzyme catalyzing the conversion of S-malate to pyruvate and CO_2; isocitrate dehydrogenase catalyzing the conversion of 2S-3R-isocitrate to α-keto-glutarate and CO_2. Both the latter are, like 6PGDH, $NADP^+$ dependent but unlike 6PGDH, A side specific (Popjak (1970) and references cited therein).

1.1 Pentose phosphate pathway

The pentose phosphate pathway occurs in the soluble extra mitochondrial cytoplasm of most organisms except very simple bacteria and uses as starting material glucose-6-phosphate, which is also produced in the metabolism of glycogen in the glycolytic pathway.

The pentose phosphate pathway has two major roles:

1. The formation of pentose sugars, in particular ribose-5-phosphate, a constituent of nucleotides and hence of RNA, DNA and coenzymes such as NAD^+ and $NADP^+$.

2. The generation of NADPH used in reductive synthesis e.g. fatty acids, mono-nucleotides, amino acids and urea. The pathway is most important in the

Figure 1.1 The reaction catalysed by 6-phosphogluconate dehydrogenase (6PGDH).

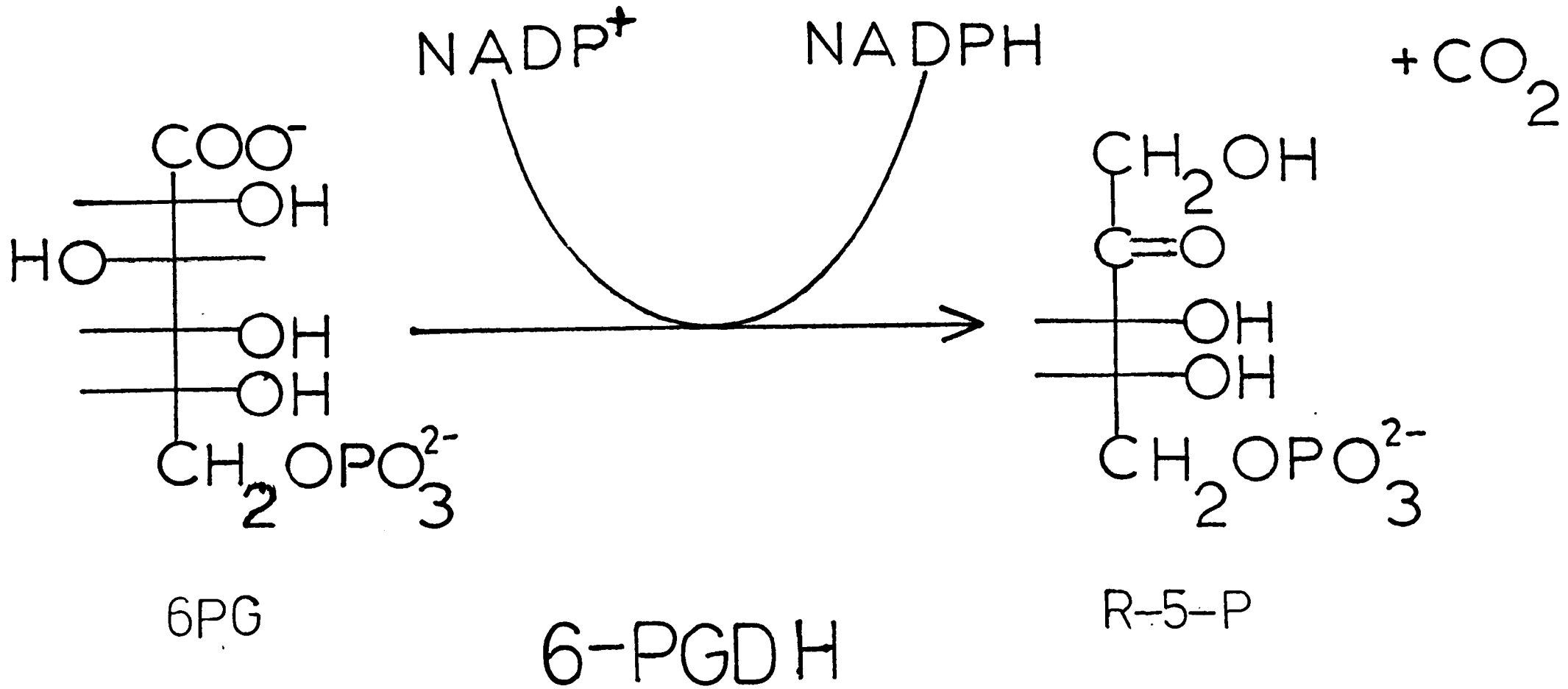


Figure 1.2 The coenzyme, nicotinamide adenine dinucleotide phosphate (i.e. with R ≡ phosphate), NADP. When R = H the molecule shown is NAD. ADPR ≡ adenosine diphosphate ribose.

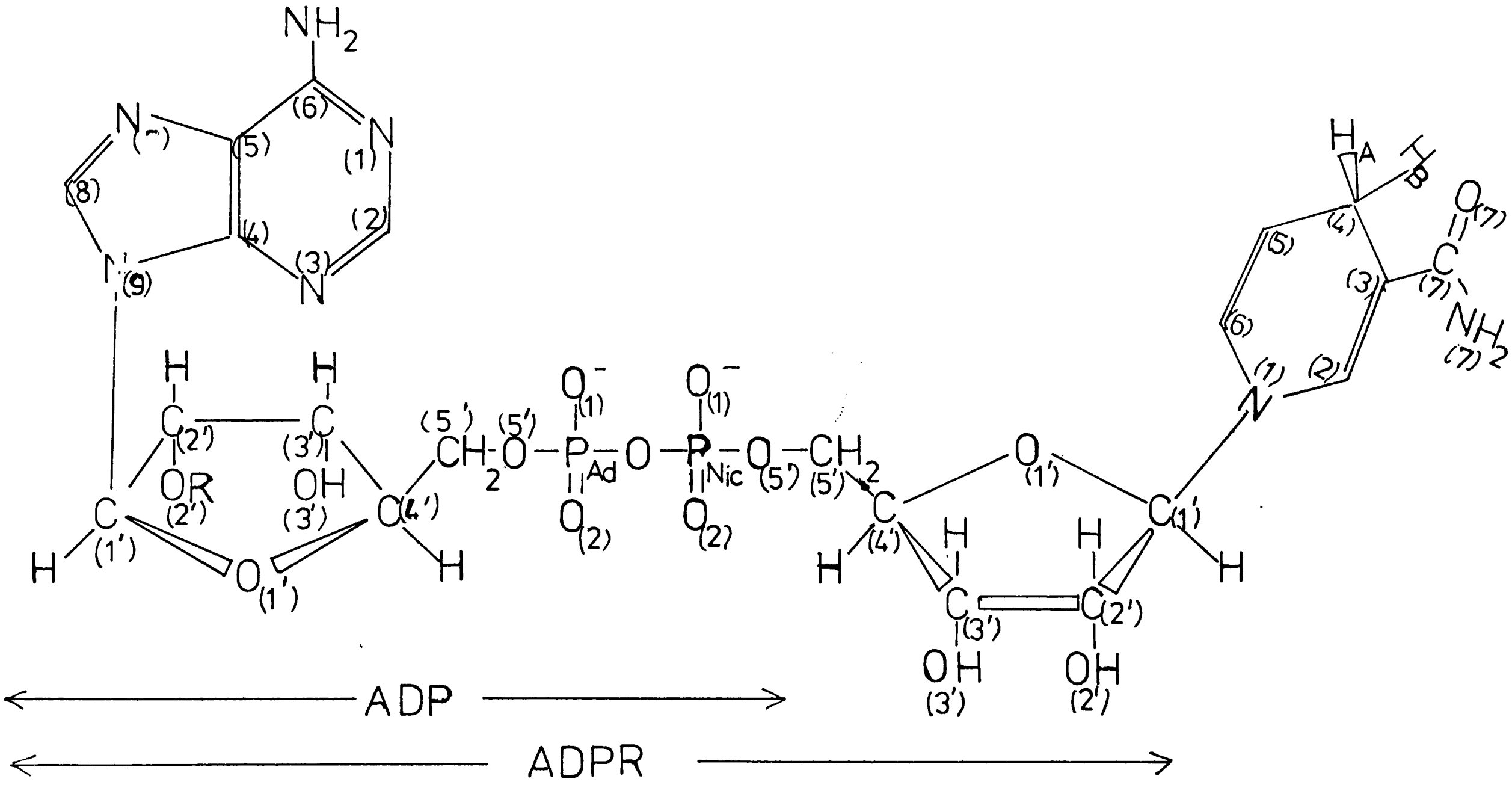
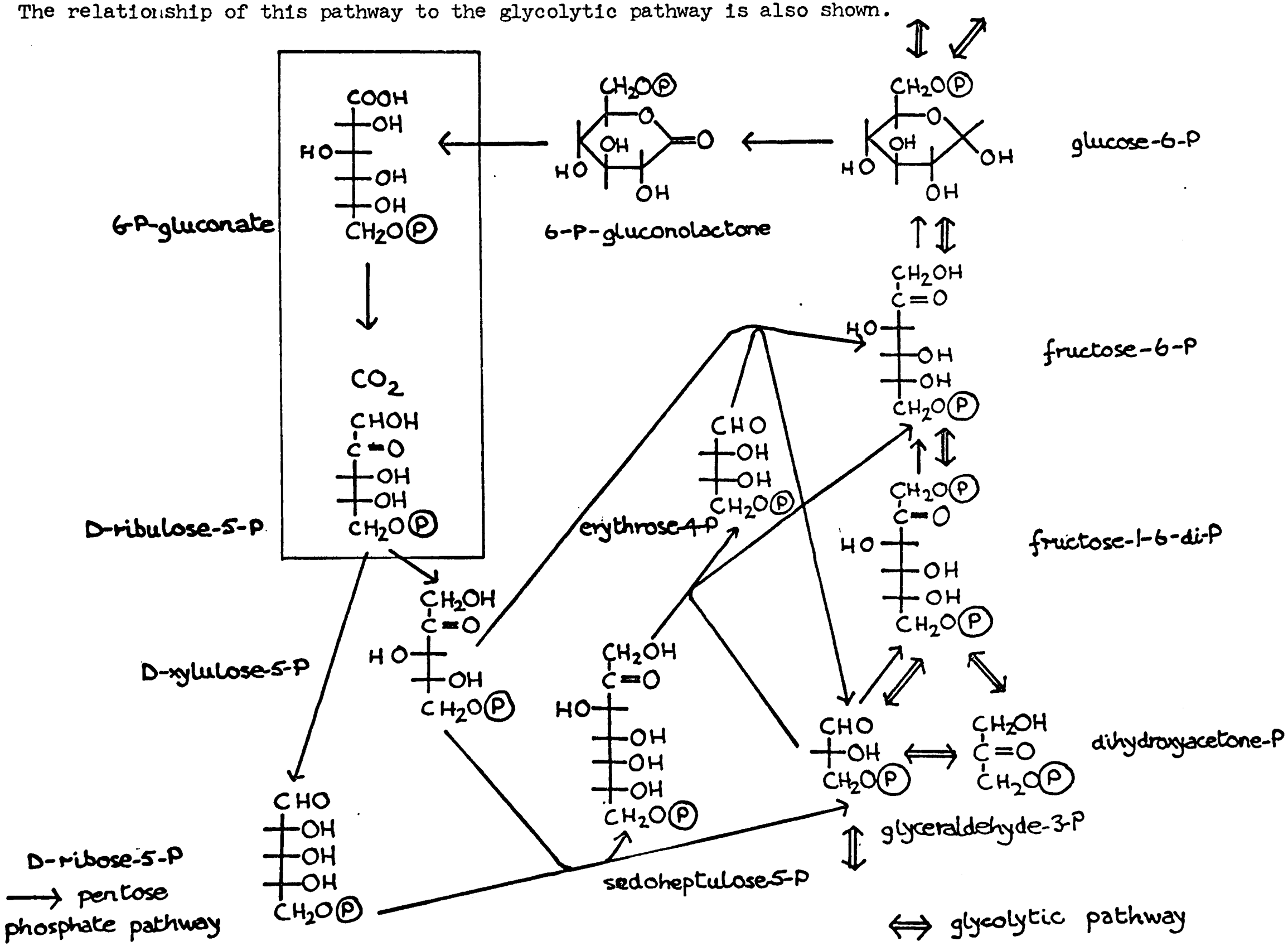


Figure 1.3 The pentose phosphate pathway starts with the conversion of glucose-6-P to 6-P-gluconolactone. The relationship of this pathway to the glycolytic pathway is also shown.



liver, mammary glands, adipose tissue and adrenal cortex.

The reactions in the pentose pathway and the relationship of this metabolic pathway to the glycolytic pathway are shown in Figure 1.3.

1.2 Preparation of the sheep liver enzyme

The work described in this thesis used crystals of the enzyme from sheep liver. The enzyme used has been purified by the Oxford Enzyme Preparation Laboratory by a procedure modified from that described by Silverberg and Dalziel (1973).

The isolation procedure involves acid treatment (at pH 5.0) of an extract of minced sheep liver in phosphate buffer, ammonium sulphate fractionation followed by chromatography on CM Sephadex and QAE Sephadex. The enzyme is then eluted, using a potassium chloride gradient, as a single symmetrical peak of uniform specific activity. The criteria used for establishing the purity are firstly, that electrophoresis should yield a single protein band on a polyacrylamide gel and secondly, that the specific activity of the enzyme should be in the range 15-22 U/mg (Silverberg and Dalziel (1973)). Enzyme which satisfies these tests is usually crystallizable. More extensive tests of purity were performed by Silverberg and Dalziel (1973) based on enzyme with the same specific activity and a similar preparation procedure. These involved assay tests using glucose-6-phosphate and ribose-5-phosphate and indicated no glucose-6-phosphate dehydrogenase or phosphoribose isomerase in the extract. These enzymes are adjacent to 6PGDH in the pentose phosphate pathway and may have been expected to contaminate the extract.

From 10 kg of sheep liver up to 300 mg of 6PGDH are obtained by the present preparation procedure.

1.3 Properties of the enzyme

The procedure of Silverberg (1973) yielded an enzyme with an estimated molecular weight of 94000 $\pm$ 2000 Daltons based on four different methods: SDS gel electrophoresis (47000), disc gel electrophoresis (92000), gel filtration (95000), and sedimentation equilibrium (96000). The SDS gel electrophoresis method

involved adding the protein to a hot (80-90°C) solution which was kept hot for 5-10 minutes before being put onto the gel, thus denaturing the enzyme. The molecular weight estimate by this method was therefore of one subunit and indicated that the enzyme was dimeric. The molecular weight estimate of 94000 obtained for the sheep liver enzyme was consistent with molecular weight estimates and subunit composition of the enzyme from other sources eg. 101000 from Candida utilis (Rippa, Signorini, Pontremoli (1967)), 102000 from rat liver (Procsal and Holten (1972)), 100000 from Bacillus stearothermophilus (Pearse and Harris (1973)), 108000 from Streptococcus faecalis (Bridges et al. (1975)), 104000 from rabbit mammary gland (Betts and Mayer (1975)). See Table 1.1.

A value of 83000 has been obtained for the molecular weight of 6PGDH from pig liver (Toews, Kanzi and Carper (1976)).

The amino acid composition of the sheep liver enzyme has been established (Silverberg (1973)) and is compared with the amino acid compositions of the enzyme from other sources in Table 1.2. The enzymes from the micro-organisms are low in cysteine residues but high in tyrosine and glutamic acid when compared with the mammalian enzymes. The distribution of residues between acidic, basic, hydrophobic and aromatic classes are very similar in the 6PGDH's from all five sources.

Reaction of the urea-denatured sheep liver enzyme with a large excess of 5-5'-dithiobis-(2-nitrobenzoic acid) shows that there are seven thiol groups in the subunit, two of which are especially reactive and essential for activity and protected by substrate or coenzyme. This was supported by studies of the reaction of the native enzyme with p-hydroxymercuri benzoate. At pH 7.0 only two thiol groups reacted with a loss of 85% of the enzymic activity (Silverberg, Dalziel and Adams (1973)). This sensitivity of the sheep liver enzyme to reagents that react with sulphydryl groups is shared by the enzyme from other sources e.g. the enzyme from S. faecalis (Bridges et al. (1975)) C. utilis (Rippa, Signorini and Pontremoli (1967)) and N. crassa (Scott and

Table 1.1 Summary of some physical and kinetic properties of pure 6-phosphogluconate dehydrogenase from different sources.

Table adapted from Bridges et al (1975).

Source of enzyme	M.W.	No subunits	K_m(NADP) μM	K_m(6PG) μM	Ref.
C. utilis	101000	2	20	52	1
Neurospora crassa	110000-120000	2	30	10	2
Streptococcus faecalis	108000±3600	2	15	24	3
Bacillus stearothermophilus	100000	2	-	-	4
rat liver	102000±3000	2	13	71	5
sheep liver	94000±2000	2	-	-	6
sheep liver	—	—	6.8(7.0)	6.9(15)	7(8)
human erythrocyte	104000	2	30	20	9
rabbit mammary gland	104000	2	23	54	10

References

1. Rippa, Signorini and Pontremoli (1967)
2. Scott and Abramsky (1973)
3. Bridges, Palumbo and Wittenberger (1975)
4. Pearse and Harris (1973)
5. Procsal and Holten (1972)
6. Silverberg and Dalziel (1973)
7. Villet and Dalziel (1972)
8. Dyson, D'Orazio and Hanson (1973)
9. Pearse and Rosemeyer (1974)
10. Betts and Mayer (1975)

Table 1.2 Amino acid composition of 6PGDH from various sources
Table adapted from Silverberg (1973).

Amino acid	1 sheep liver	2 *Candida utilis*	3 *B. Stearothermophilus*	4 Human red blood cell	5 *Streptococcus faecalis*
lysine	32	31	31	38	35
histidine	10	6	9	9	6
arginine	18	15	21	21	19
aspartate	41	46	37	50	43
threonine	16	19	19	20	25
serine	20	21	16	26	26
glutamate	33	45	57	45	69
proline	16	18	18	14	18
glycine	44	47	43	48	43
alanine	37	50	53	43	47
valine	21	26	25	29	17
methionine	11	8	11	12	13
isoleucine	30	15	40	28	33
leucine	41	38	38	42	40
tyrosine	11	18	19	12	22
phenylalanine	20	20	18	24	18
tryptophan	7	6	5	8	9
half-cystine	9	4	3	11	3
Total	417	433	463	480	486

1. Silverberg (1973)
2. Rippa *et al*. (1971)
3. B.M.F.Pearse (personal communication)
4. B.M.F.Pearse (Ph.D. thesis, University of London)
5. Bridges *et al*. (1975).

Abramsky (1973)).

Chemical modification studies with the C. utilis enzyme using active site-specific reagents have indicated a role of histidine in the binding of both 6PG and NADP, and an involvement of tyrosine as well in the binding of coenzyme (Rippa and Pontremoli (1969)), Rippa et al.(1971) and Rippa et al. (1972)).

The amino acid sequencing of sheep liver 6PGDH has recently been commenced by Dr. A. Carne and Dr. J.I. Harris, M.R.C. Laboratory in Cambridge, in conjunction with the X-ray structure determination in Oxford. The chemical modification results should provide important marker residues when fitting the sequence to the enzyme structure.

The mechanism of the reaction is random order in addition of coenzyme and substrate, in tris-acetate buffer (Dyson, D'Orazio and Hanson (1973)). However, studies of the sheep liver enzyme in phosphate buffer indicate a compulsory order mechanism in which $NADP^+$ binds first to the enzyme (Silverberg (1973)). Phosphate is a competitive inhibitor of 6PGDH.

A large number of compounds are known to act as inhibitors of the enzyme. Pyridoxal-5-phosphate binds to 6PGDH from C. utilis forming an inactive complex, it only binds to the free enzyme and inhibits non-competitively with respect to 6PG (Rippa, Spanio and Pontremoli (1967)). A variety of metabolic intermediates of the glycolytic pathway and tricarboxylic acid cycle have been found to be inhibitors of sheep liver 6PGDH. All nucleoside 5'-triphosphates and 5'-diphosphates are inhibitors with K_I values averaging 0.8 mM. nucleoside 5'-monophosphates are also inhibitory with K_I values averaging 6 mM. The following compounds are also inhibitory with K_I values averaging 6 mM: oxalacetate, oxalate, fructose-1-P, fructose-6-P, glucose-6-P, P_i, PP_i, citrate and sulphate. Fructose-1, 6-diphosphate is a potent inhibitor of the enzyme (K_I = 71 μM) (Dyson and D'Orazio (1971), (1973)). The presence of EDTA prevents this inhibition (but not that of the other inhibitors) suggesting that a metal is necessary for fructose-1,6-diphosphate

to bind to the enzyme.

Dyson and D'Orazio (1973) suggest that the K_m for 6PG (15 μM) and the K_I for fructose-1,6-diphosphate (71 μM) are such that inhibition of 6PGDH by fructose-1,6-diphosphate may be of regulatory significance.

The coenzyme NAD^+ does not bind to the enzyme either in solution (Silverberg (1973)) or in the crystal (M.J.Adams, personal comm., Fig.4.1e).

1.4 Reasons for studying the enzyme

The interest in 6PGDH has several aspects:

1. The reaction catalysed by 6PGDH is a major source of cellular NADPH and the metabolic function and control of 6PGDH is a potentially fruitful area of study.

2. 6PGDH like isocitrate dehydrogenase and malic enzyme combines an oxidation-reduction reaction with a decarboxylation and the interrelation of these two catalytic functions may be of importance in the understanding of enzyme catalysis.

3. 6PGDH utilizes $NADP^+$ instead of NAD^+. The NAD^+ linked dehydrogenases represent a class of dinucleotide binding proteins which have been extensively studied by X-ray crystallographic techniques (LDH (Adams et al. (1973), MDH (Webb et al. (1973)), LADH (Eklund et al. (1974)), lobster GAPDH (Buehner et al. (1974), Bacillus stearothermophilus GAPDH (Biesecker et al. (1977)). The only difference between NADH and NADPH is the replacement of the 2' hydroxyl on the ribose of the adenosine moiety by a phosphate. Whereas the function of NADH-linked enzymes is energy production, NADPH-linked enzymes are biosynthetic and it is not clear what role the 2' phosphate group in NADPH may play in this distinction. A comparison of the protein structures of NADPH linked enzymes and the NAD linked enzymes should shed light on this and the related question of whether they evolved independently or from a common ancestor. One high resolution study of an NADPH linked enzyme has been published recently (Matthews et al. (1977)).

1.5 Common features in protein three-dimensional structure with particular reference to the dehydrogenases

Similarities in the structures of proteins published to date has attracted considerable attention both in terms of the protein folding problem and in terms of evolutionary relationships between proteins.

It has been found that in about 80% of the globular proteins of known three-dimensional structure, β-pleated sheets are important features. Since β-sheets inherently require the close interaction of regions separated in the amino acid sequence, an understanding of β-sheets is central to the understanding of the organisation and folding of proteins and also essential in any successful method of predicting three-dimensional structure from amino acid sequence (Richardson (1977)).

The similarities in structure and/or sequences between different proteins (and within different parts of the same protein) could be the result of either divergent or convergent evolution.

Three different forms of divergent evolution can be distinguished and all three may have occurred in the evolution of a protein.

Firstly, divergent evolution occurring in parallel with the evolution of different species.

Secondly, divergent evolution resulting from internal gene replication i.e. a single gene evolved from two or more copies of a smaller gene.

Thirdly, divergent evolution resulting from external gene replication. Different copies of a primitive gene evolved by replication to code for proteins with different, although probably related but more specific functions. Thus the descendants of the ancestral protein might have similar structures , and possibly sequences, to each other and to the ancestral protein.

In convergent evolution, the different genes that coded for different proteins arose independently. Owing to stereochemical or functional requirements or both, these proteins either had or converged during subsequent

evolution to a similar structure (Sternberg (1977)).

When considering relationships between different proteins, interest will centre, for instance, on the structures expected as a result of the third type of divergent evolution or convergent evolution.

For example, the structural and sequence similarity of myoglobin to both the α and β chains of haemoglobin (Perutz et al. (1968)) suggested that the genes that code for the α and the β chains of haemoglobin and for myoglobin have evolved from a common ancestor.

The three pancreatic serine proteinases: α-chymotrypsin, tosyl elastase and trypsin (Stroud et al. (1971)) possess a very similar geometric arrangement for the side chains of Ser, His, and Asp that are involved in catalysis and similar tertiary structures. The similarity in tertiary structures together with the similarly positioned disulphide bridges strongly suggests that these enzymes have diverged from a common ancestor.

Subtilisin BPN' and the trypsin family of proteins have a similar proteinase function and a virtually identical geometric arrangement for the side chains of Ser, His and Asp that form the active site (Kraut et al. (1971) ; Delbaere et al. (1975)). But the amino acid sequences, the three-dimensional structures and the order of the three active site residues along the chain are quite different in subtilisin and in the trypsin family. Thus it is very likely that the similar geometries of the active sites are a result of convergent evolution to a functionally favourable geometry.

It has also been found that the structures of the NAD^+ linked dehydrogenases (LDH, MDH, LADH, GAPDH) possess similar structural domains. That is, the structures of those parts of the polypeptide chains whose function is to bind the NAD^+ coenzyme are remarkably similar although the complete subunit conformations are not alike, since the remainder of the polypeptide chains which are required for substrate binding, catalysis and formation of oligomeric structure, are different in each case.

For these proteins the dinucleotide binding domain includes a 6-stranded parallel sheet (βA, βB, βC, BD, βE and βF) and 3 or 4 connecting helices

(αB, αC, αE and α1F in LDH and MDH), two on each side of the sheet. There is also a general left handed twist from one strand to the next amounting to about 100° between the extreme ends of the sheet (Figure 1.4). In the LDH and MDH cases the whole subunit structure, coenzyme binding and catalytic, are very similar. Comparison of LDH, LADH and GAPDH shows detailed differences in the coenzyme domain (for instance in GAPDH there are three additional β-strands) and totally different catalytic domains and different subunit-subunit contacts. The folding pattern of these proteins is shown in Figure 1.5.

The biological function of the domain is to bind the coenzyme NAD^+ to the enzyme in such a way as to bring the C-4 atom of the nicotinamide moiety into a proper position with respect to the substrate and catalytic residues of the enzyme. The conformations of the bound coenzyme have been studied in each of the four dehydrogenases mentioned (LDH: Chandrasekhar et al. (1973); LADH: Eklund et al. (1974); MDH: Webb et al. (1973); GAPDH: from lobster, Olsen et al. (1976), from Bacillus stearothermophilus, Biesecker et al. (1977)). The conformations are all very similar except for a 180° rotation about the glycosidic bond linking the ribose to the nicotinamide moiety for NAD^+ bound to GAPDH. The adenine and nicotinamide moieties are about 14 Å apart (centre to centre of the aromatic rings and about 17 Å end to end) and are nearly perpendicular to each other (Figure 1.6). The similarities of the coenzyme conformation when bound to the dehydrogenases reflect the similarity of the fold in the dinucleotide binding domains and of detailed similarities in those regions that are actually involved in binding. A functionally important conserved residue is an aspartate (LDH, 53; LADH, 223; and GAPDH, 32; no sequence exists for MDH), the last residue of βB. It forms a hydrogen bond with the 2' hydroxyl of the adenine ribose and is thus a significant constraint on coenzyme binding.

Rao and Rossmann (1973) made a quantitative measure of the similarities in structure. They rotated and translated one protein into the other in such a way as to minimize the sum of the squared distances between all

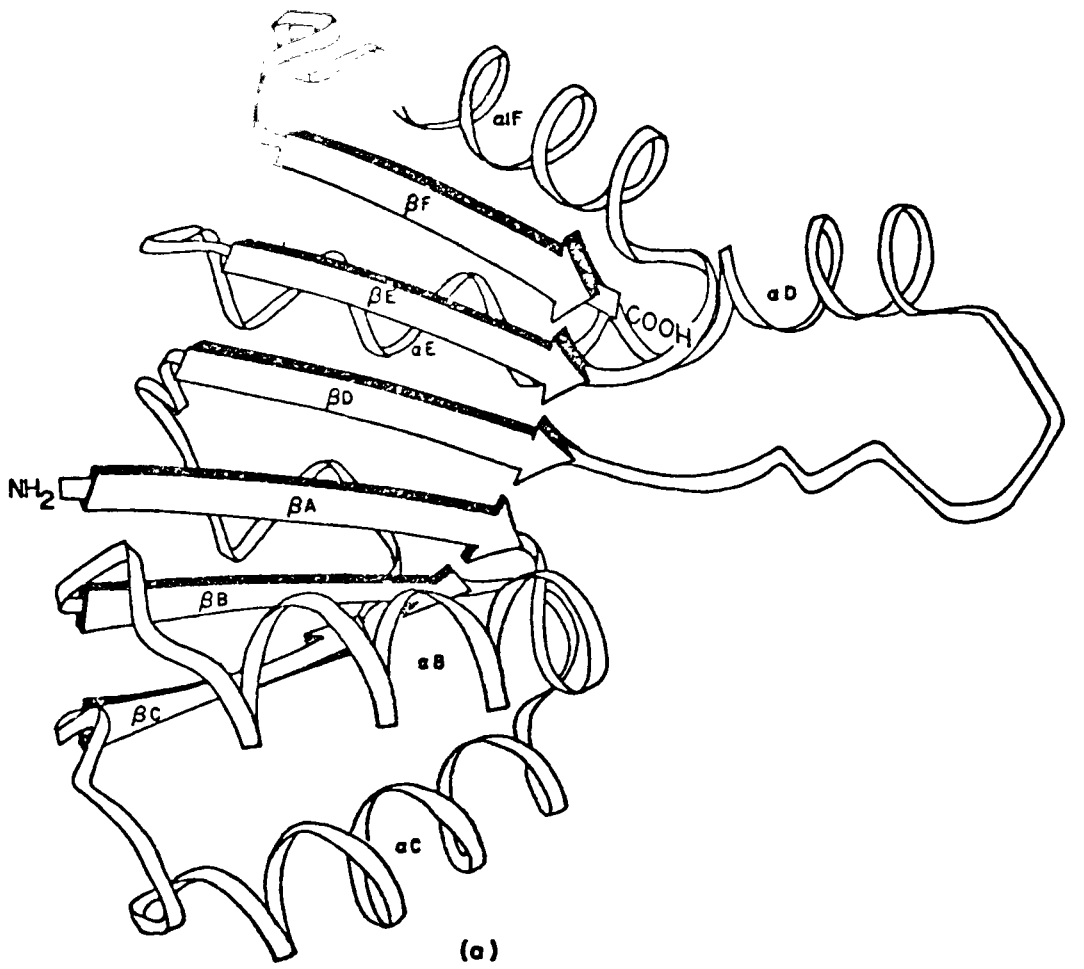


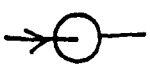
Figure 1.4 Schematic drawings of the NAD^+ binding region in dehydrogenases; (a) viewed from top, perpendicular to the six-stranded parallel sheet and (b) viewed from end, looking along the extended polypeptide chains in the sheet, with their amino terminal end closest to the observer. Alpha helices are labelled "α" and strands of the β-sheet "β". From Rossmann et al. (1975).

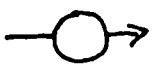
Legend to Figure 1.5

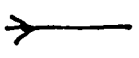

β-strand viewed from N-terminus.

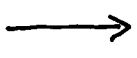

β-strand viewed from C-terminus.

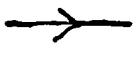

Helix with axis approximately parallel to strands viewed from C-terminus.


Helix with axis approximately parallel to strands viewed from N-terminus.


Helix whose axis is not parallel to strands.


N-terminus of polypeptide chain.


C-terminus of polypeptide chain.

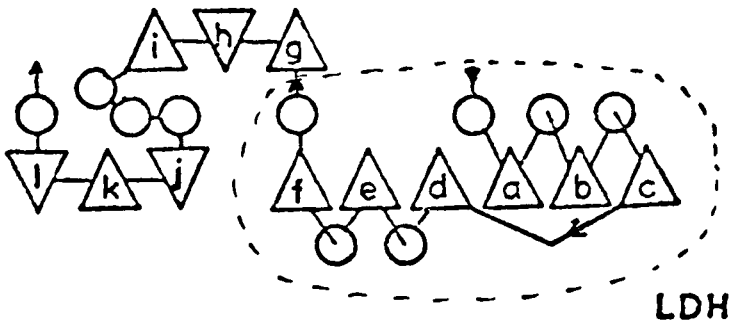

Connection.

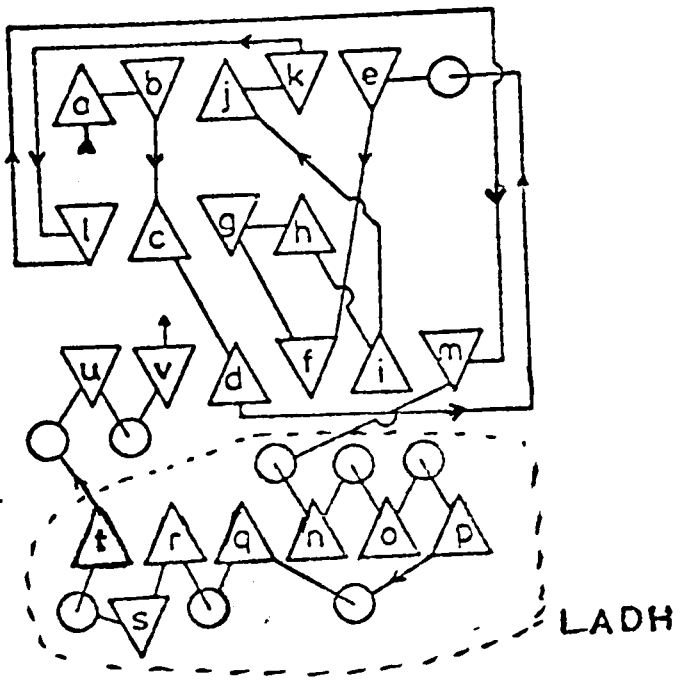
LADH Horse liver alcohol dehydrogenase.

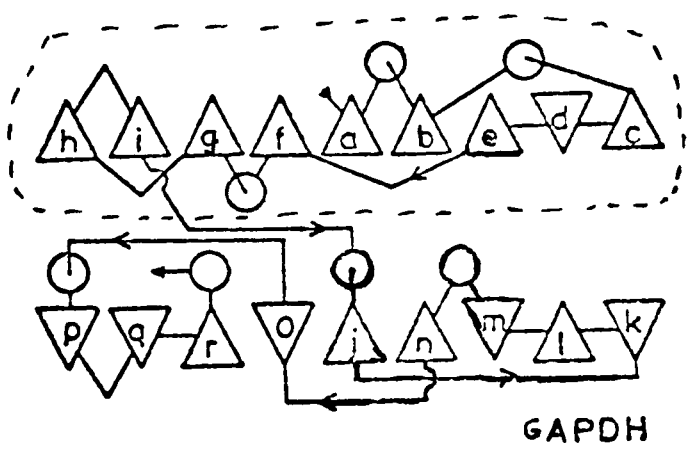
GAPDH Lobster glyceraldehyde-3-phosphate dehydrogenase.

LDH Dogfish lactate dehydrogenase.

PGK Bovine phosphoglycerate kinase.







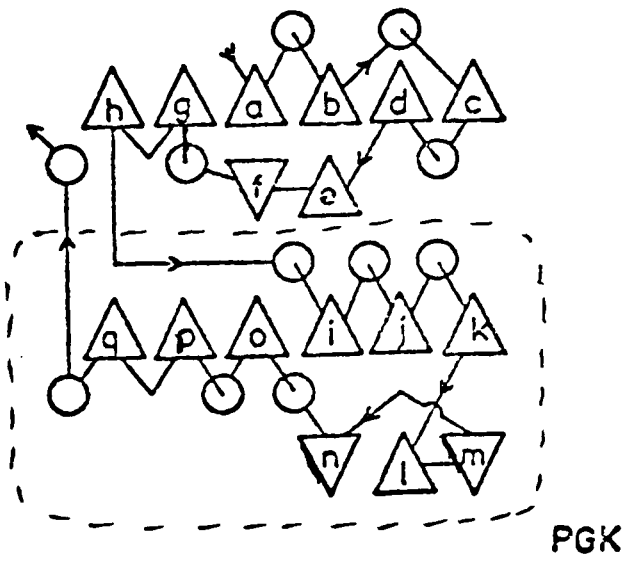
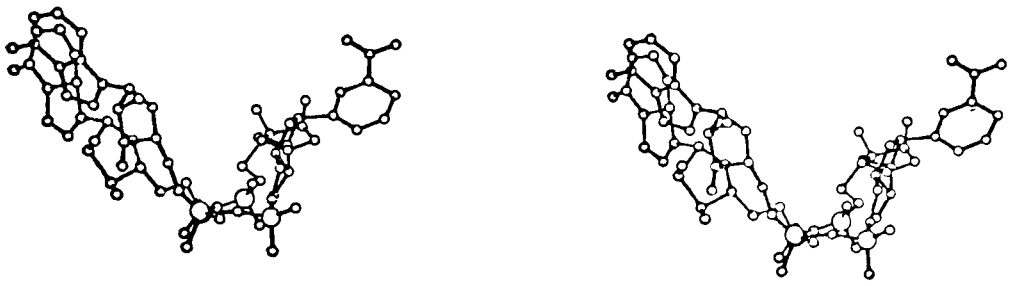


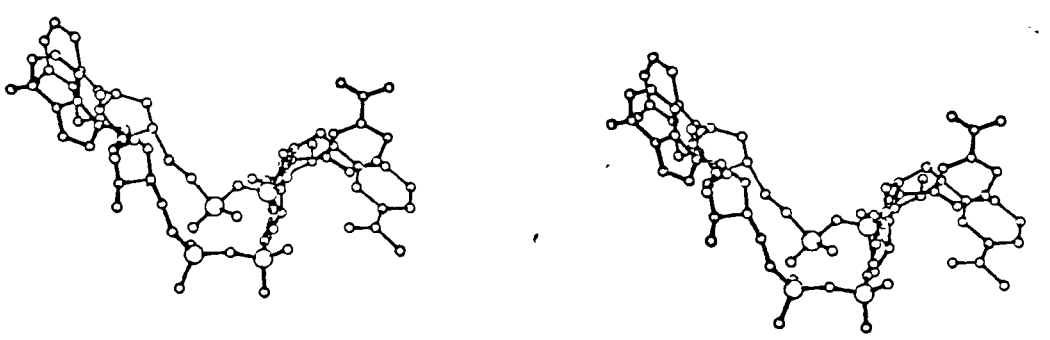
Figure 1.5 Schematic diagrams of the chain folds of LDH, LADH, GAPDH and PGK. The latter binds ADP (or ATP) as coenzyme and the remainder are NAD^+ dependent. The very similar coenzyme domains are enclosed within a dotted boundary. The marked differences in chain fold between the remainders of each structure are readily apparent. For a key to the notation used see opposite.(From Sternberg (1977)). Note: the structures of the LDH and MDH subunits are very similar, apart from the first 20 residues at the N-terminal end of LDH which is absent in MDH.



(a)



(b)



(c)

Figure 1.6 Comparison of the NAD^+ conformation and position between (a) s-MDH and LDH, (b) LADH and LDH (only ADPR is shown for LADH), and (c) GAPDH and LDH. The structure for LDH-bound NAD is in double outline in every case (From Rossman et al. (1975)).

C_α atoms assumed to be equivalent. The root mean square deviation, σ, of these distances was taken as a quantitative measure of similarity between the structures. For instance $\sigma = 3.1$ Å for 34 equivalent atoms comparing the mononucleotide binding units of LDH.

The differences that do exist in the relative positions of the various structural elements among these four dehydrogenases led Rossmann et al. (1975) to conclude that the four central strands and two connecting helices have been more strongly conserved during the course of evolution than have been the extremities of the structure. Similarly folded tertiary structures are found in a number of other proteins (for instance PGK (Blake and Evans (1974)), see Figure 1.5). Rossmann et al. (1975) proposed that the βαβαβ structure may be a mononucleotide binding unit of widespread occurrence as it has a common evolutionary origin.

Schultz and Shirmer (1974) suggested that structural relationships be assessed on the basis of topology, i.e. chain folding without reference to the exact geometry. The topology of a β sheet is fully described by the strand sequence in the sheet and the above-below pattern of the connections. There are finite, calculable numbers of possible strand sequences and possible connection patterns and the number of possible topologies (M) is also calculable and finite.

For each of two proteins containing an n-stranded parallel sheet with identical topology then since such a sheet can exist in M different topologies, the probability that by chance both sheet topologies are identical is 1/M i.e. the significance of their relationship or "figure of topological relatedness" is M:1.

Following the analysis of Schultz and Schirmer (1974):if we have an n-stranded parallel sheet then the number of possible strand sequences in an n-stranded parallel sheet is $n!/2$. That is the number of permutations of n items reduced by a factor of 2 since symmetrically related strands such as ABCD and DCBA can be superimposed by rotation. Also, there are (n-1) connections between the strands which can run either above or below the sheet

giving 2^{n-1} possible above-below connection patterns. Combining strand sequences and connection patterns, $M = 2^{n-1}\frac{n!}{2}$, other things being equal. Comparing LDH and LADH, for example, both have n = 6, therefore, M = 11520, indicating that it is unlikely that the LDH and LADH topologies were very similar purely by chance.

Sternberg and Thornton (1976) pointed out however that the βαβ fold showed a strong preference for folding in a right-handed fashion (Figure 1.7). Only 1 out of 58 proteins containing βαβ units (where the strands considered could be either adjacent or non-adjacent) was left-handed. The exception found was in subtilisin. The probability that such a biased distribution would occur by chance was very small (< 1 in 10^{8}, the exact value depending on the grouping of the proteins involved).

The calculations of Schultz and Schirmer (1974) assumed that the left- and right-handed connections were equally probable. By taking into account that the probability of obtaining a right-handed βαβ in practice is 0.96 (instead of 0.5), then we have

$$M = 2^{n-1}\frac{n!}{2}\frac{1}{\left(\frac{0.96}{0.5}\right)^{(n-1)}}$$

there being (n-1) βαβ in an n-stranded sheet, i.e. there are less possible ways of folding the sheet than in the absence of this effect. Table 1.3 shows the corrected figure of topological relatedness for a variety of structures. Sternberg and Thornton (1976) conclude that the topological similarities between some structures, especially adenyl kinase and flavodoxin (3:1), adenyl kinase and any DH (2:1) and individual examples of the Rossmann fold (βαβαβ) (3:1) may be the result of convergent evolution since it is not unlikely that the folds could occur by chance. However, the value of 442:1 for the DH:DH is high and the similarity is most probably explained by divergence from a common ancestor.

It is interesting to point out that both convergent and divergent theories of evolution indicate that the right-handed βαβ unit is particularly stable. Such super-secondary structures have probably arisen both by

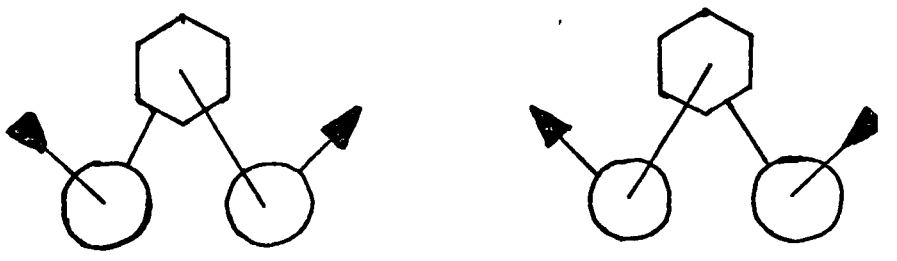


Figure 1.7 A schematic view of a right-handed (a) and a left-handed (b) βαβ unit, where ◯ ≡ a β strand and ⬡ ≡ α helical connecting region. In (a) as one proceeds along the chain from the N-terminus (indicated by ⟶) a highly distorted right-handed helix is formed from the strand, the axis of the α-helix and the next strand. In (b) proceeding from the N terminus in similar fashion, a left-handed helix is formed (From Sternberg and Thornton (1976)).

Table 1.3

The corrected figures of topological relatedness calculated for the probability (p) of a right-handed βαβ unit of 0.50 and 0.96.

Compared sheet topologies	Figure of topological relatedness	
	p = 0.50	p = 0.96
adenyl kinase/subtilisin	240:1	34:1
adenyl kinase/flavodoxin	24:1	3:1
adenyl kinase/ any DH	16:1	2:1
subtilisin/flavodoxin	480:1	35:1
subtilisin/any DH	240:1	18:1
flavodoxin/any DH	480:1	35:1
A DH/ another DH	11520:1	442:1
Rossmann fold/Rossmann fold	12:1	3:1

(From Sternberg and Thornton (1976))

convergence and divergence .

This analysis rests on the assumption that the present sample of known structures is a representative sample of all proteins containing β-sheet.

Richardson (1977) in an analysis of β structure also concludes that many of the protein pairs having identical or closely similar β-sheet topologies merely represent coincidentally close choices among a few highly probable topologies. However, evolutionary relationships almost certainly underlie the similarity of the nucleotide binding domains of the dehydrogenases.

A recent study of an $NADP^+$ dependent enzyme, dihydrofolate reductase (DHFR) at 2.5 Å has been published (Matthews et al. (1977)). The overall folding of this enzyme is dominated by an 8-stranded sheet beginning at the amino terminus and ending with a single antiparallel strand at the carboxy terminus. Thirty percent of the polypeptide backbone is β-sheet. There is the usual twist to the sheet amounting to 130° (Figure 1.8). All the βcβ cross over connections are right handed. The first 6 strands of the β-sheet are parallel and are buried within the molecule.

The β-sheet topology (Figure 1.9) of DHFR, Matthews et al. (1977) point out, is unlike any other known protein structure. Only in PGM the first four parallel strands in a six stranded mixed β-sheet are topologically similar to the first five parallel strands in DHFR, if the shortest strand D is considered an excursion. This similarity is most probably a coincidence, they conclude, due to the requirements of protein folding.

However, inspection of the connectivity in the sheets of DHFR and the NAD^+ binding regions of the dehydrogenases does show similarity if the type of connectivity is ignored (α-helix or extended chain). Cutting the connection between strands A and B of DHFR and comparing D, C, B, E with C, B, A, D, of DHFR and the NAD^+ binding domain respectively, the connectivity is identical. It can be concluded from the analysis of Sternberg and Thornton (1976) that, since , the similarity in connectivity extends only one strand more than the Rossmann fold, the figure of topological relatedness

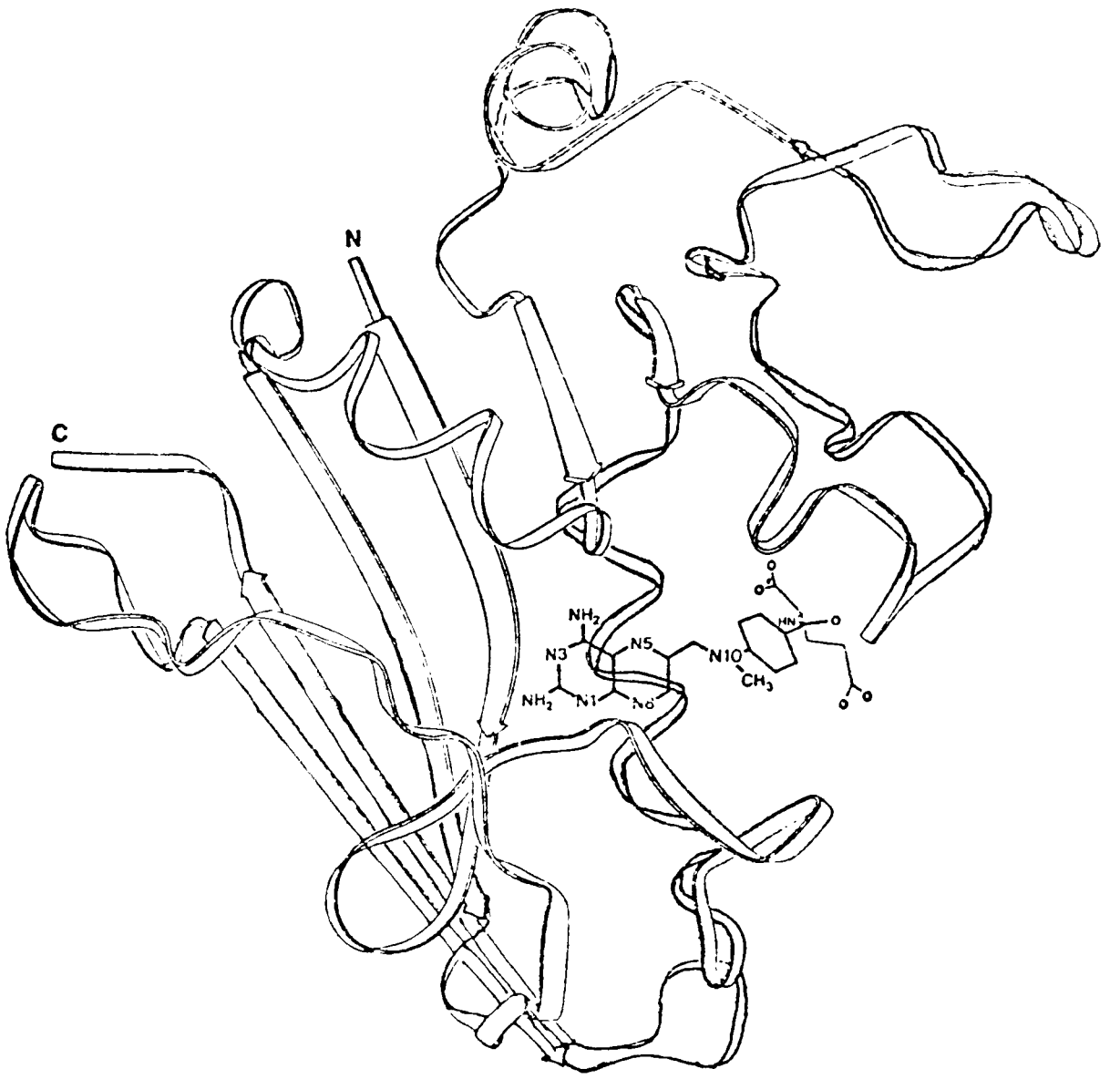
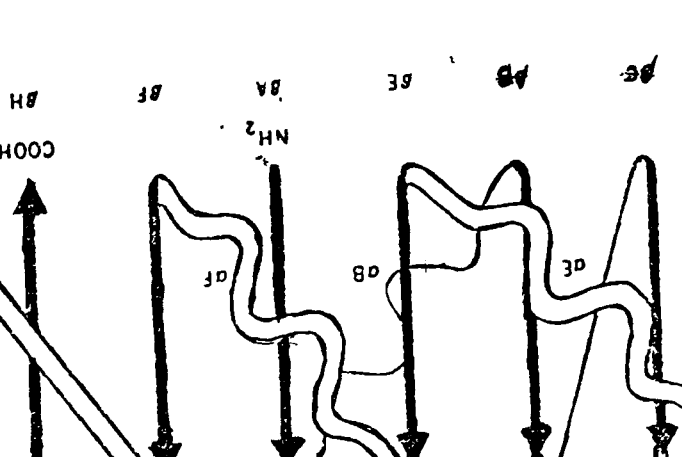


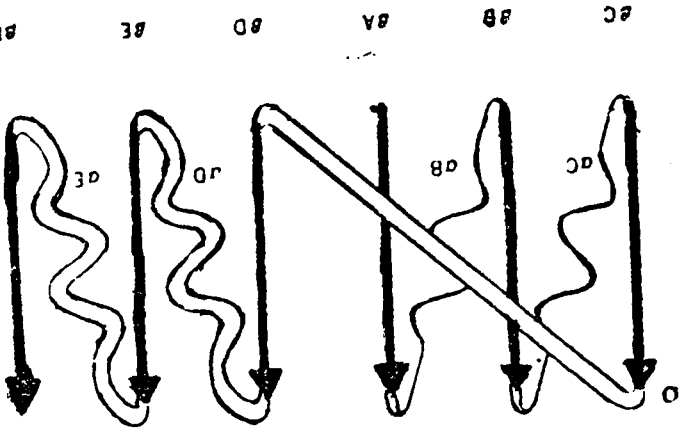
Figure 1.8 Representation of the backbone chain folding of E. coli dihydrofolate reductase containing a bound methotrexate molecule (a strong inhibitor of the enzyme). The drawing is derived from a computer generated plot of all atoms in the drug and all α-carbon atoms of the enzyme. Strands of the central pleated sheet are shown as heavy arrows (from Matthews et al. (1977)).

Figure 1.9

(a) Topological diagram showing how individual strands are connected to form the central β-pleated sheet in DHFR.
(b) Topological diagram for the NAD^+ binding regions in dehydrogenases. No attempt is made to describe the lengths of the individual strands or the conformation of the connections, except that α-helical connections are represented by wavy lines. (From Matthews et al. (1977)). (c) Topological diagram for PGM (α-helical connections not shown), from Richardson (1977). Connections drawn with thin lines are below and those drawn with thick lines are above the sheet.



1.9(c)



1.9(b)

1.9(a)

is 14:1 and the similarity is probably unlikely to have occurred other than by chance, i.e. not by divergent evolution from a common ancestor.

On consideration of NADPH binding Matthews et al. (1977) point out that the 2' phosphate of NADPH is likely to interact with Arg 44 of DHFR, precisely what would be expected i.e. the positive charge on the Arg interacts with the negative charge of the phosphate. This is in contrast to the hydrogen bond between the 2' hydroxyl of NAD^+ and the aspartate residue in the NAD^+ binding domain of the dehydrogenases.

The question whether 6PGDH, as an $NADP^+$ dependent dehydrogenase, can be expected to bear a sufficiently close resemblance in structure to that of the NAD^+ dependent dehydrogenases and possess a common ancestry is thrown wide open with the determination of the DHFR structure. Perhaps, instead, a better question to ask is whether 6PGDH has a very similar structure to DHFR, so establishing the family of $NADP^+$ dependent enzymes with a common ancestor.

CHAPTER 2

X-RAY STRUCTURAL STUDIES OF PROTEINS

The diffraction pattern is the continuous Fourier transform of the contents of the unit cell sampled at the reciprocal lattice points. The convolution of the unit cell with the nodes of the crystal lattice generates the crystal.

The function C(x) representing a one dimensional crystal is given by:

$$C(x) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{n=+\infty} \delta(u-na)\, f(x-u)\, du$$

where $\sum_{n=-\infty}^{n=+\infty} \delta(u-na)$ is the infinite series of Dirac δ functions which each possess the property that $\int_{-\infty}^{+\infty} \delta(x)\, dx = 1$ and $f(x)$ represents the unit cell.

The diffraction pattern in 3-D is the product of the integrals:

$$D = \int_V \rho(\underline{r})\ \exp(2\pi i \underline{r}.\underline{S})\, dV \ \cdot \int_{V_x} \sum_{n=-\infty}^{n=+\infty} \delta(\underline{x}-\underline{r}_n) \exp(2\pi i \underline{x}.\underline{S})\, dV_x$$

where V = the unit cell volume;

V_x = the crystal volume;

$\underline{S} = \underline{s} - \underline{s}_o$ and $\underline{s}$ = the wave vector of the scattered X-ray beam and $\underline{s}_o$ is the wave vector of the incident beam.

$\underline{r}$ = a vector from the origin of the unit cell.

$\underline{r}_n = n_1\underline{a} + n_2\underline{b} + n_3\underline{c}$; n_1, n_2, n_3 are integral,

$\underline{a}$, $\underline{b}$, $\underline{c}$, are vectors parallel to the unit cell and of dimensions $|\underline{a}|$, $|\underline{b}|$, $|\underline{c}|$; and is a vector to the origin of all the unit cells.

$$D = \int_V \rho(\underline{r}) \exp(2\pi i \underline{r}.\underline{S})\, dV. \sum_n \exp(2\pi i \underline{r}_n.\underline{S})$$

which is non zero for $\underline{r}_n.\underline{S}$ integer which yields the equations:

$$\underline{a}.\underline{S} = h$$

$$\underline{b}.\underline{S} = k$$

$$\underline{c}.\underline{S} = l; \qquad h, k, l \text{ integers.}$$

These are the vector equations of planes and are known as the Laue equations.

Redefining vectors put $\underline{X} = \underline{r} = x\underline{a} + y\underline{b} + z\underline{c}$ where x, y, z are fractions of the unit cell edge and $\underline{h} = \underline{S} = h\underline{a}^* + k\underline{b}^* + l\underline{c}^*$ where $\underline{a}^*$, $\underline{b}^*$, $\underline{c}^*$ are reciprocal lattice vectors defined by the equations:

$$\underline{a}.\underline{a}^* = 1 \qquad \underline{a}.\underline{b}^* = 0 \qquad \underline{a}.\underline{c}^* = 0 \qquad \text{etc.}$$

The diffraction pattern is made up of discrete points (defined by the reciprocal lattice) and each point is weighted by the amplitude squared of the structure factor, $F_{\underline{h}}$:

$$F_{\underline{h}} = \int_V \rho(\underline{x}) \exp(2\pi i \underline{h}.\underline{x}) dV.$$

Let $\underline{x} = \underline{x}_j + \underline{r}$ where $\underline{r}$ is now a vector from the centre of each atom and $\underline{x}_j$ is a vector from the unit cell origin to each atom centre.

Then,

$$\rho(\underline{x}) = \sum_{\substack{N \text{ atoms of} \\ \text{unit cell}}} \rho_j(\underline{r})$$

$$F_{\underline{h}} = \int_V \sum_{j=1}^{N} \rho_j(\underline{r}) \exp(2\pi i \underline{h}.(\underline{x}_j + \underline{r})) \, dV.$$

$$= \sum_{j=1}^{N} \left\{ \int_V \rho_j(\underline{r}) \exp(2\pi i \underline{h}.\underline{r}) \, dV \right\} \exp(2\pi i \underline{h}.\underline{x}_j)$$

The integral is the Fourier transform of the electron density of the finite jth atom and is the atomic scattering factor for an atom at rest. This needs to be multiplied by $\exp-(B\sin^2\theta/\lambda^2)$ to allow for thermal vibration, and is written f_j (in the notation used here).

Therefore, $$F_{\underline{h}} = \sum_{j=1}^{N} f_j \exp(2\pi i \underline{h}.\underline{x}_j) \qquad (2.1)$$

units: electrons

and $$\rho(\underline{x}) = \frac{1}{V} \sum_{\underline{h}} F_{\underline{h}} \exp(2\pi i \underline{h}.\underline{x}) \qquad (2.2)$$

units: electrons Å^{-3}

(This treatment is similar to that given by Woolfson (1970)).

The structure factor $F_{\underline{h}}$ is a complex number, possessing amplitude and phase. Experimentally, only the intensity can be measured giving the amplitude $|F_{\underline{h}}|$ of the structure factor; the phase of the diffracted beam with respect to the incident beam is lost. It is the determination of the phase which is the main problem tackled by X-ray structure analysis.

The combination of the structure factors in the Fourier series (equation 2.2) is a mathematical method equivalent to the method of image formation by a glass lens at optical wavelengths. The best image of an object illuminated by visible light in a microscope is obtained if the objective has an aperture wide enough to admit all the light diffracted by the object (after Abbé, (1887)). Leaving out scattered light from the object results in a deterioration of the quality (resolution) of the image in a way analogous to the series termination effects in an electron density map calculated from an incomplete set of structure factors. The mathematical method of X-ray crystallography is needed because the refractive index of materials at X-ray wavelengths is almost exactly unity thus giving no refractive, and therefore focussing, effect.

No attempt will be made to cover methods of structure determination in general as these topics are well covered by text books (Stout and Jensen (1968), Woolfson (1970)). However, discussion will be given on methods relevant to the work presented. This discussion will necessarily include isomorphous replacement and direct methods. An extensive review of protein crystallography has been published recently (Blundell and Johnson (1976)).

2.1 Isomorphous replacement

The method of isomorphous replacement is central to the determination of the primary phase set in protein structure analysis (Green, Ingram and Perutz (1954)). With protein crystals, isomorphous derivatives can be prepared by either reacting heavy atom substituted molecules with the protein in solution, followed by crystallisation, or by diffusion of such reagents through the large channels of mother liquor in the crystal, which contains a large percentage of solvent. Occasionally a metal ion already present may be replaced by another heavier one (e.g. zinc by lead in insulin crystals). The requirement is that the heavy atom merely replaces another group or solvent molecule on the protein without causing conformational changes, which can lead to changes in unit cell dimensions and/or crystal deterioration. Discernible changes in the intensities of reflections can be produced by using the substitution of a heavy atom on the protein, itself a molecule of large molecular weight.

For acentric reflections the use of one isomorphous derivative leaves a phase ambiguity (Figure 2.1) which can only be resolved by using anomalous scattering information or a second derivative. In the absence of errors this would be sufficient. Several derivatives (with anomalous scattering information) are used wherever possible.

For centric reflections only one derivative is needed for a unique determination of the sign of the structure factor (Figure 2.2) which accounts for the better quality of structure determinations of protein crystals with a large percentage of centric reflections.

In both cases the position and occupancy of the heavy atoms need to be known. These positions can be determined using the Patterson function or by direct methods (see § 2.4).

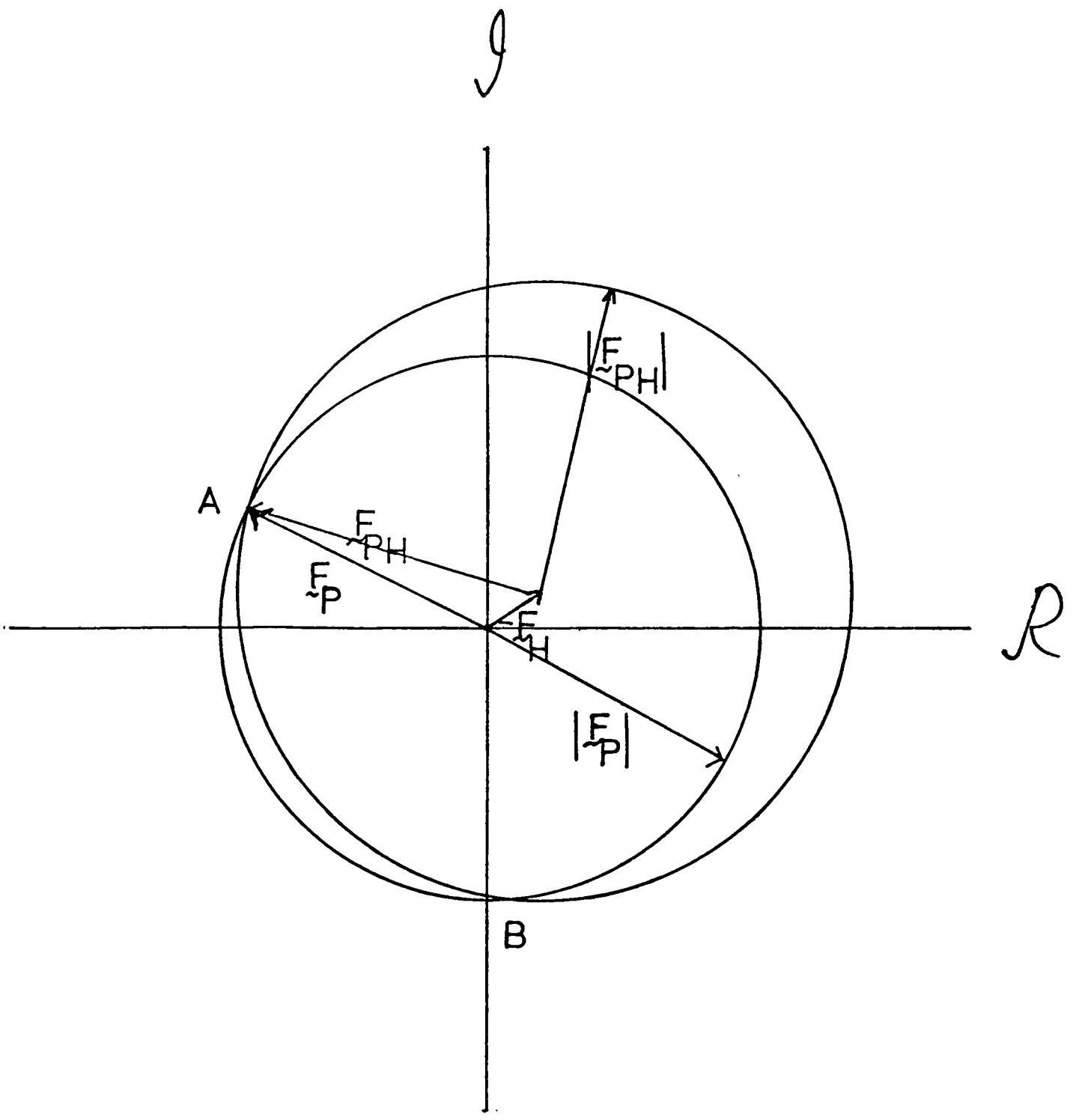
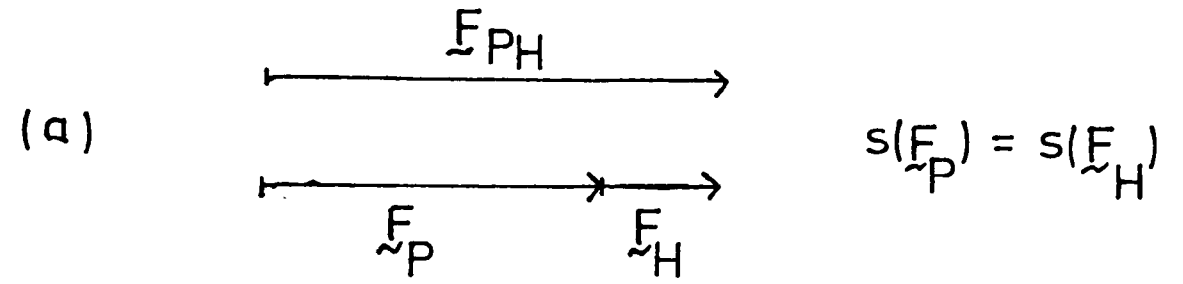


Figure 2.1 The Harker construction for phase calculation by the method of single isomorphous replacement. Note the two possible values for α_P at A and B (see Figure 2.9).

Figure 2.2 Phase determination in a centric zone by the method of single isomorphous replacement. The situation corresponding to "cross over" is depicted in (c).



For the protein crystallographer, once the derivatives have been prepared, the crystallographic procedure of isomorphous replacement has four stages.

(i) The measurement of intensities of the X-ray diffraction patterns for native and derivative crystals.

(ii) The reduction and correction of the intensity data.

(iii) The determination of the heavy atom positions.

(iv) The determination of the phases.

2.2 Methods of intensity measurement

Data collection from protein crystals faces three major problems.

(i) the inherent weakness of the individual reflections;

(ii) the voluminous amount of data to be collected;

(iii) radiation damage (see § 2.3).

(i) Sources of radiation

When considering the inherent weakness of the individual reflections consideration needs to be given to the sources of radiation available.

(a) Conventional sources of X-rays

These sources utilize the emission of a line spectrum from a metal target when bombarded by an accelerated beam of electrons from a heated filament.

The target can be stationary and enclosed in a sealed tube giving a source which has a long lifetime (for example the Phillips tube, focus 8×0.4 mm^2 run at 1.2 kW has a lifetime of 1000 hours). X-ray output is limited by the heating of the target, but by rotating the target the heating effect is distributed over a greater surface area thus enabling increased powers to be utilized (eg. Elliot G X 6 rotating anode, focal spot 2×0.2 mm^2, run at powers up to 2 kW). These X-ray sources, involving working parts within a high vacuum system, can present maintenance

problems and are less reliable than the conventional fine focus sealed tube. The typical X-ray spectrum emitted by a tube with a Cu target is shown in Figure 2.3. The monochromatisation of the X-ray beam from a Cu target is often accomplished by means of a Ni filter which has an absorption edge coincident with the Cu K_β emission line. The attenuation for various thicknesses of Ni filter is given in Table 2.1. A higher degree of monochromatisation is achieved by reflecting the direct X-ray beam from a single crystal (e.g. of graphite) set at the appropriate Bragg angle for the characteristic radiation. This gives a better peak to background ratio than is obtained with a Ni filter and a reduction in radiation damage for sensitive protein crystal specimens.

The energy in the diffracted beam is given by Darwin's formula:

$$E_n(\underline{h}) = I_o \lambda^3 e^4 P.L.A.V_x. |F(\underline{h})|^2 \quad \overline{\omega} \quad \overline{m^2c^4} \quad \overline{V^2}$$

where ω = angular velocity of the rotating crystal,

λ = the wavelength of the X-rays used,

e = the electronic charge,

m = the mass of the electron,

c = the velocity of light,

P = the polarization factor which for a Bragg angle θ and unpolarized incident radiation is $\frac{1}{2}(1 + \cos^2 2\theta)$.

L = the Lorentz factor. This accounts for the relative time each reflection spends in the reflecting position and depends on the geometry used in data collection.

A = absorption factor,

V_x = crystal volume,

V = unit cell volume,

$|F_{\underline{h}}|$ = structure factor amplitude for reflection, $\underline{h}$.

Figure 2.3

The typical X-ray spectrum emitted by a tube with a Cu target operating at about 35 kV before (——) and after (---) passage through a Ni filter. The characteristic wavelengths are $K_\beta = 1.3922$, $K_\alpha = 1.5418$ (Å).

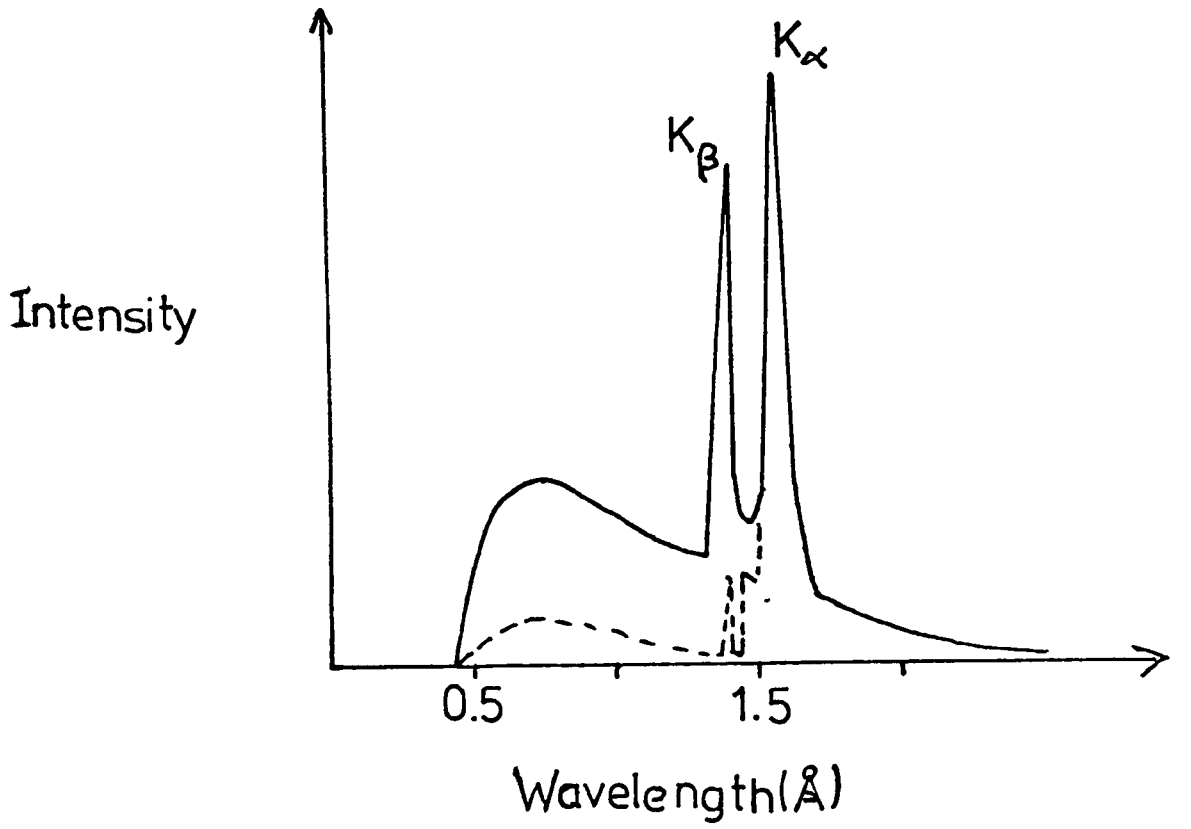


Table 2.1

Attenuation of the X-ray beam for various thicknesses of Ni filter.

Thickness of filter	% Kα intensity	% Kβ intensity	100.Kβ/Kα
0	100	21	21
0.015	60	1.3	2.2
0.025	41	0.1	0.2

Because $E_n(\underline{h})$ is inversely proportional to V^2, the larger the unit cell studied, the weaker each $E_n(\underline{h})$. The situation can be improved by increasing V_x, but many interesting proteins yield only small crystals. In this situation, I_o, the intensity of the incident beam,needs to be increased. With a conventional source, careful selection of the Ni filter thickness needs to be made (see Table 2.1). More intense sources of X-rays in the form of synchrotrons and storage rings are available. These sources offer other useful characteristics.

(b) The synchrotron

This type of machine was developed primarily for use in high energy physics research, where charged particles are maintained in circular orbits by bending magnets and is the culmination of the cyclotron and synchro-cyclotron (Littlefield and Thorley (1970)). Periodically the beam is extracted and allowed to bombard a target from which elementary particles are emitted and studied.

An accelerating charge radiates in a direction tangential to the instantaneous path of the orbiting charged particles. The spectrum of the radiation emitted is a slowly varying continuum extending from the infra-red into the X-ray region (Figure 2.4), and represents an energy loss process for the orbiting charged particles. The lost energy is restored by radio frequency accelerator cavities. The emission is confined by relativistic effects to a narrow cone with low angular divergence in the plane of the ring (Figure 2.5).

Because a wide range of experiments can be performed using this radiation source, dedicated storage rings have been constructed, from which the electron beam is not extracted for high energy physics bombardment experiments.

The literature on synchrotron radiation is growing rapidly. For a general review see Godwin (1969) and of its applications, Codling (1973),

Figure 2.4

Spectral output of the storage rings DORIS and SPEAR and the NINA synchrotron, in terms of the number of photons $N(\lambda)$ emitted for a given wavelength λ.

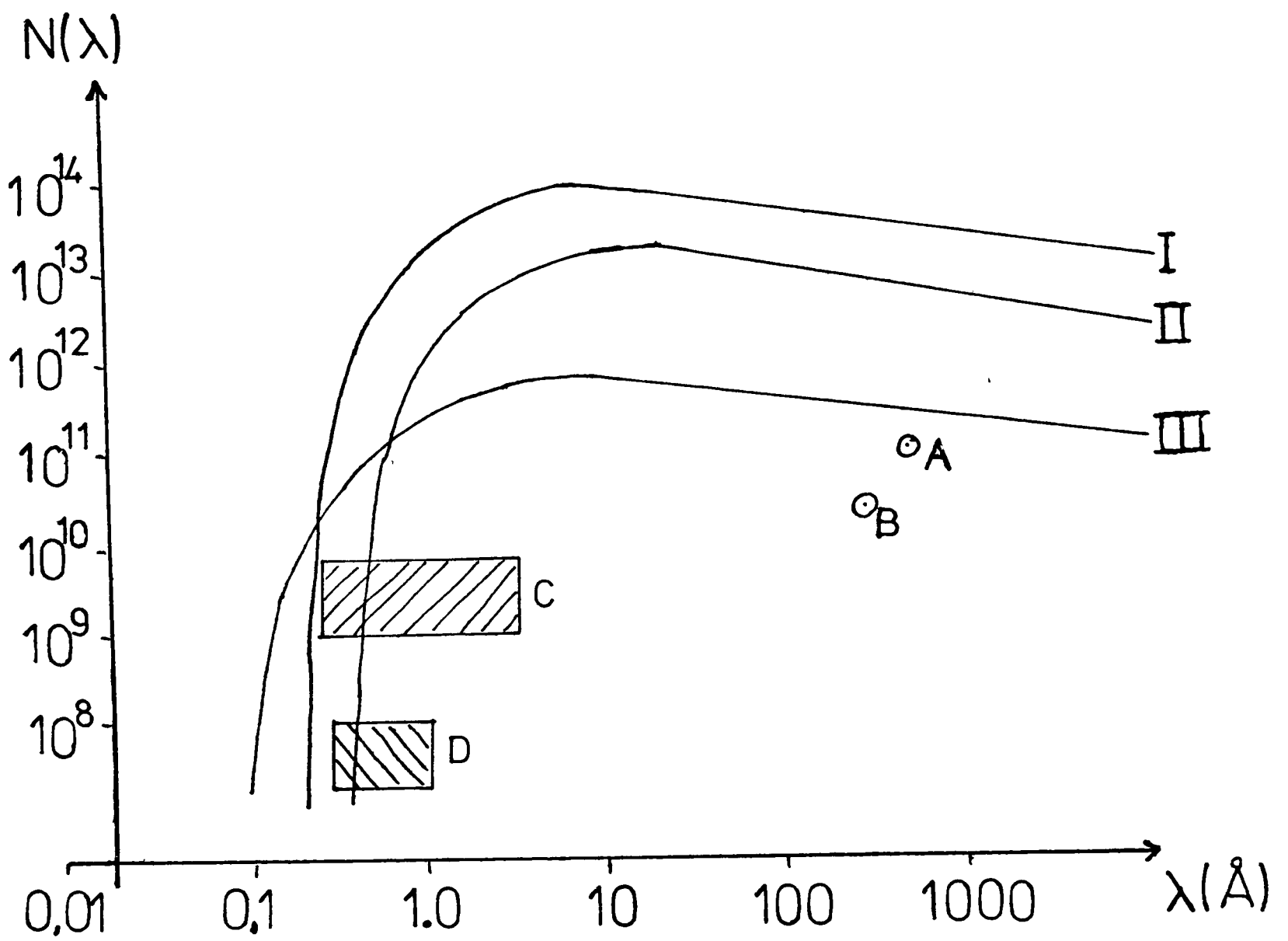
Curve I: corresponds to the performance of DORIS at Hamburg (3GeV, 1000 mA).

Curve II: to the storage ring SPEAR(Stanford, U.S.A.) (2.5GeV, 250 mA).

Curve III: to the characteristics of the NINA synchrotron (Daresbury, U.K.) (5GeV, 20 mA).

Point A marks the performance of a Helium resonance lamp and B the He^+ resonance line. The rectangle C covers the range within which characteristic X-ray sources are available at certain wavelengths. The rectangle D covers the range of Bremstrahlen from a 1.6 kW rotating anode X-ray tube.

(From the SRC report: "A plan for future research with synchrotron radiation based on a Dedicated Storage Ring").


N(λ)
10^14
10^13
10^12
10^11
10^10
10^9
10^8
I
II
III
A
B
C
D
0,01
0,1
1.0
10
100
1000
λ(Å)

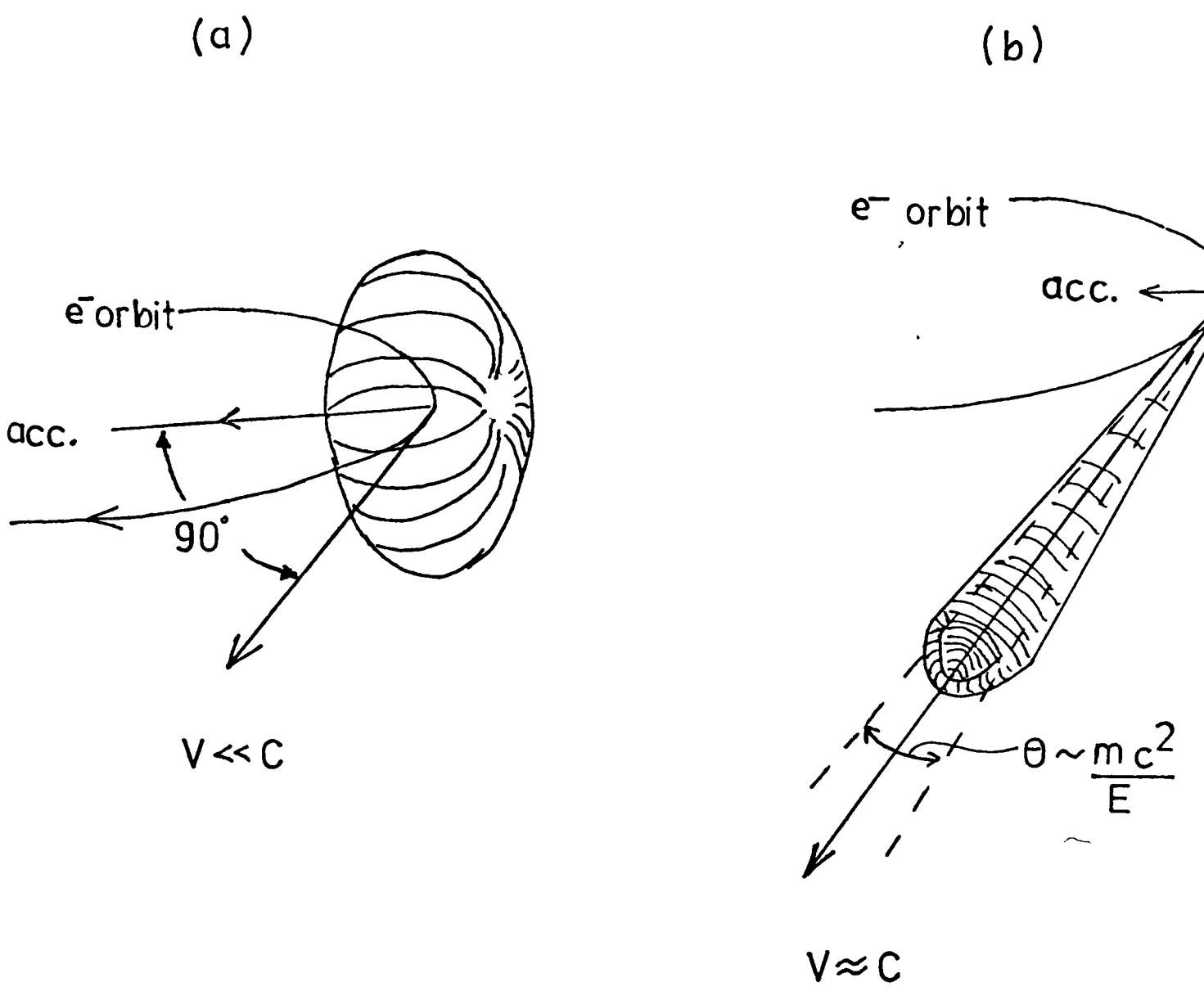


Figure 2.5

Radiation pattern for the power radiated by a radially accelerated electron for (a) the non-relativistic case and (b) the relativistic case (from Haensel (1973)).

more specifically its use in molecular biology, Holmes (1974), Barrington-Leigh and Rosenbaum (1976) and in protein single crystal diffraction, Wyckoff (1973), Harrison (1973), Phillips et al. (1976), Harmsen, Leberman and Schultz (1976) and Phillips et al. (1977).

In protein single crystal work two main properties of these sources are of use:

(i) In the case of the storage ring, the increased intensity over conventional sources.

According to Phillips et al. (1977) the storage ring SPEAR (Stanford, U.S.A.) provides an intensity gain of 7.5 over a Ni filtered, Elliot G X 6 rotating anode whereas the synchrotron has a poorer intensity (x 0.32). For both the synchrotron and storage ring the beam was focussed with a mirror/monochromator system (Webb et al. (1977)) so the intensity comparisons are made for working protein single crystal diffraction experiments.

The protein crystals studied by Phillips et al. (1976) did not suffer increased radiation damage when intercepting the very high photon flux in the diffraction experiment. For example, with azurin crystals, the radiation damage (to 2.7 Å) was minimal after 28 hours exposure in the X-ray beam from the storage ring. This compares with a decrease of 25% after 100 hours of exposure with Ni filtered radiation from a Cu tube operated at 700 W. Thus, at least with these crystals, a time-dependent decay process is operative, distinctly non-linear in dose for X-rays and even though decomposition could be proceeding at the same rate, data are being collected much faster than with a sealed tube.

(ii) The wavelength tunability of the radiation over the range 0.5 to 3.0 Å is useful in the optimisation of anomalous scattering and therefore maximising the size of the Bijvoet difference $(I_{\underline{h}} - I_{-\underline{h}})$.

X-rays are absorbed by matter according to the law:

$$I = I_o \exp(-\mu x)$$

where I_o is the incident X-ray intensity, I is the transmitted intensity,

μ is the absorption coefficient and x the thickness of the absorber. The X-ray absorption spectrum of a substance is characterized by several absorption edges which result from abrupt increases in μ at photon energies just sufficient to liberate an inner shell electron from an atom (Figure 2.6). The inner shell electron is said to scatter anomalously i.e. other than as a free electron.

The atomic scattering factor has to be rewritten to include this effect:

$$f = f_o + f'(\omega) + i\, f''(\omega)$$

(the nomenclature is that recommended by the Proceedings of the Inter Congress Conference on Anomalous Scattering, Ramaseshan and Abrahams (1975)). f' and f" are respectively the real and imaginary parts of the anomalous component, ω is the frequency of the incident wave. The variation of f' and f" with frequency is shown in Figure 2.7.

For crystal structures containing only light atoms (C, N, O, H) f' and f" are negligible at X-ray wavelengths. From equation 2.1:

$$|F_{\underline{h}}| = |F_{-\underline{h}}| \qquad \text{(Friedel's Law)}.$$

However, when a heavy atom is added, yielding an (N+1) atom structure:

$$F_{\underline{h}} = \sum_{j=1}^{N} f_j \exp(2\pi i \underline{h}.\underline{x}) + (f_{N+1} + f'_{N+1} + if''_{N+1}) \exp(2\pi i \underline{h}.\underline{x})$$

Then $|F_{\underline{h}}| \neq |F_{-\underline{h}}|$ (see Figure 2.8).

The anomalous difference $(|F_{\underline{h}}| - |F_{-\underline{h}}|)$, written as $(|\underline{F}_{PH}^{+}| - |\underline{F}_{PH}^{-}|)$ in §2.5, can be utilized along with the isomorphous difference to solve the phase problem (Figure 2.9). The larger f", the more accurate will be the determination of the phase. The wavelength of the radiation has to be chosen carefully in order to maximize f".

For wavelengths greater than the absorption edge f" is small and constant but f' varies with λ. In this case the changes in the diffraction pattern are equivalent to the changes due to isomorphous replacement but the crystals are perfectly isomorphous.

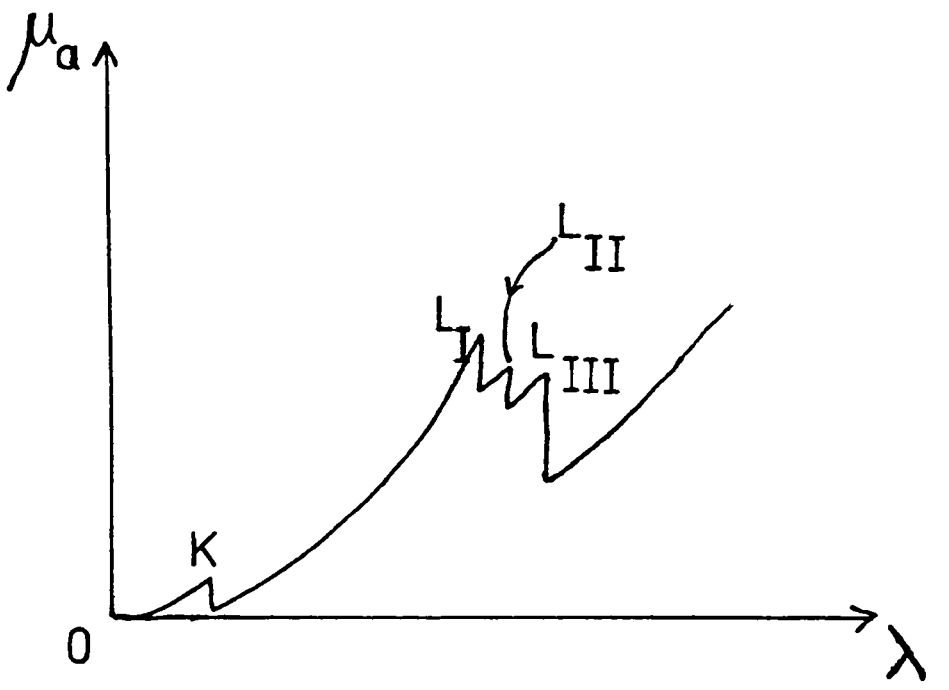


Figure 2.6

The variation of the mass absorption coefficient with λ for a typical element (from Woolfson (1970)).

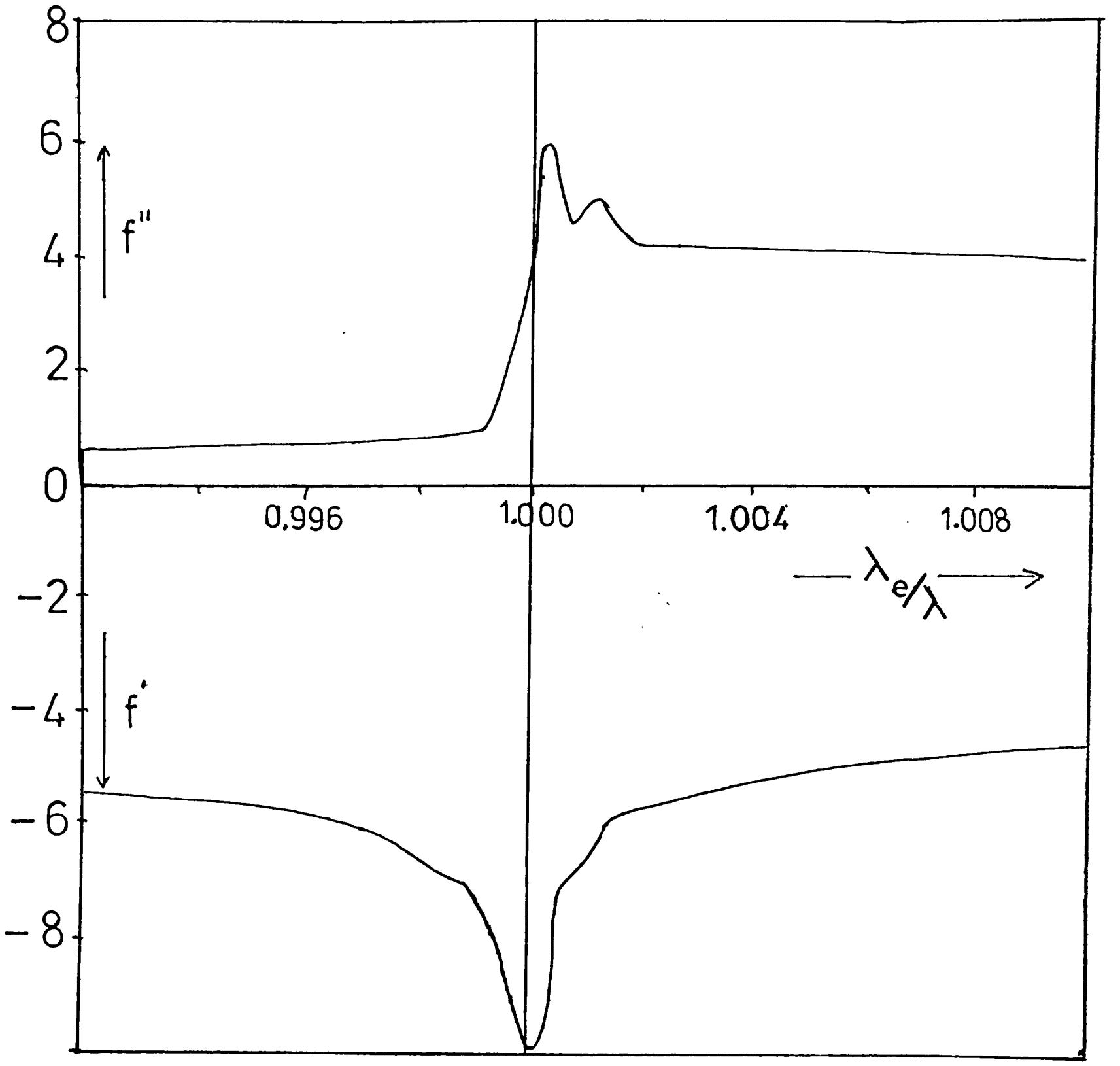


Figure 2.7

The corrections, f' and f", to the atomic scattering factors for the anomalous scattering due to Ga K electrons. Note that for wavelengths, $\lambda >$ the wavelength of the edge, λ_e, f" $\approx$ 0 and constant (from Phillips et al. (1977) after Fukamachi and Hosaya (1975)).

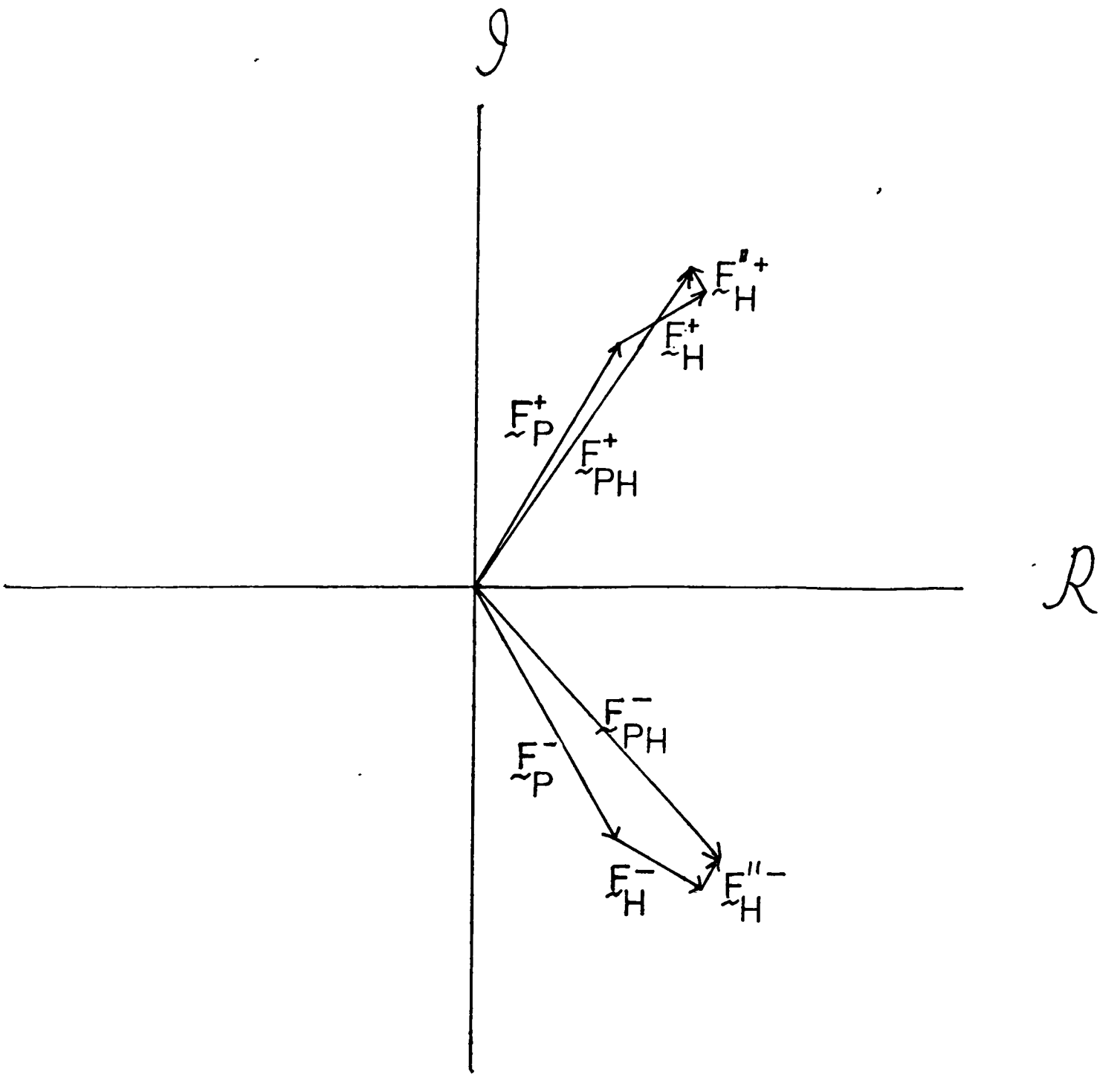


Figure 2.8

Vector diagrams for the Friedel related reflections for the native ($\underset{\sim}{F}_P^+$ and $\underset{\sim}{F}_P^-$) and derivative ($\underset{\sim}{F}_{PH}^+$ and $\underset{\sim}{F}_{PH}^-$), where the derivative contains an anomalously scattering atom. Note $\underset{\sim}{F}_H''$ breaks the symmetry about the real axis.

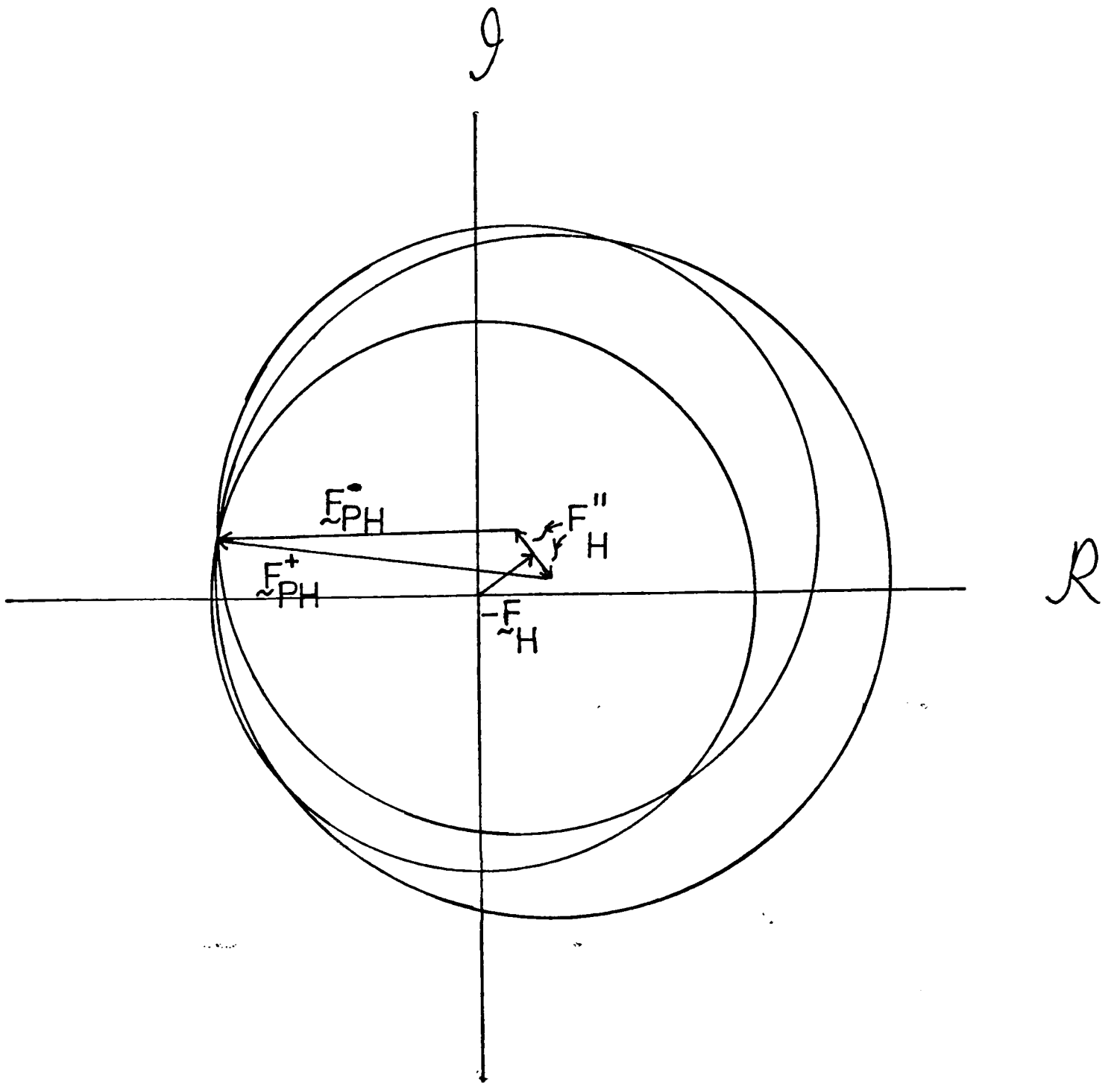


Figure 2.9

Harker diagram to illustrate phase determination using anomalous scattering combined with single isomorphous replacement.

Phillips et al. (1977) performed experiments on the iron containing protein rubredoxin at the SPEAR storage ring using seven discrete wavelengths of radiation above and below the iron K edge (1.74 Å). The iron site could be correctly located from difference Patterson and Fourier maps based either on f' or f". The phases of the Bragg reflections were also obtained from three films collected at different wavelengths. When compared with the known phases for rubredoxin, the differences averaged to 60° indicating that some phasing information was available even from the relatively poor data.

The continuous spectrum of radiation from the synchrotron is a smoothly varying function of the wavelength of high intensity. Previous attempts at optimising the anomalous scattering involved the use of the characteristic lines of various targets, e.g. copper, chromium, iron, cobalt, molybdenum and silver (see Table AI.1; e.g. Blow (1957)) and the use of the continuous white radiation background has been suggested (Herzenberg and Lau (1967)). The characteristic lines do not lead to full optimisation of the anomalous scattering and are not as conveniently obtained as with the automatic wavelength selection achieved with the synchrotron. As for the background radiation this is not sufficiently intense (see Figure 2.4).

An alternative method of obtaining more significant anomalous differences is to select a heavy metal derivative with an absorption edge near Cu K_α. For instance Dy (Z=66) has an L edge at 1.579 Å which being on the long wavelength side of Cu K_α (1.5418) should have f" optimised (Figure 2.7). Possible soluble compounds are $DyCl_3$ (used on insulin. Blundell (1970)), $Dy(Ac)_3.H_2O$ (used on 6PGDH, see Table 3.1).

Appendix I contains details of the work done on 6PGDH crystals at the NINA synchrotron, Daresbury, U.K.

(ii) Data Collection

(a) Diffractometry

Automatic diffractometers were developed to meet the need of collection of large quantities of data (Arndt and Willis (1966)). Once a crystal has been aligned, computer instructions set the crystal and counter for the sequential measurement of reflections. In the case of the four circle diffractometer single reflections are measured in turn.

Because of the large asymmetric units associated with protein crystals and consequently the densely packed reciprocal lattice, several reflections are quasi-simultaneously in the reflecting position. This property is utilized with 3 and 5 counter machines to increase the rate of data collection (Phillips (1964), Arndt, North and Phillips (1964), Banner (1972), Evans (1973)).

A wide range of molecular weights exist for the proteins studied so far crystallographically, e.g. rubredoxin (6300 Daltons), lysozyme (14,600), 6PGDH (100,000), phosphorylase b (200,000), TMV disc protein (600,000). The quantity of diffraction data to be collected increases as the amount of protein in the crystallographic asymmetric unit increases (see Table 2.2).

Efficient use of the diffractometers available would mean feasible high resolution studies of small molecular weight protein molecules and low resolution studies of high molecular weight proteins on the 4 circle, single counter diffractometer and high resolution studies of medium molecular weight proteins (e.g. prealbumin, 28000 Daltons in asymmetric unit) on the 5 circle, 5 counter diffractometer. For high resolution studies of high molecular weight molecules (e.g. 6PGDH, ρPb, TMV disc protein) screenless film techniques are used which are the most efficient means of data collection.

Generally, an ω scan is made on a diffractometer through a given diffraction peak and the intensity computed directly (e.g. the ordinate analysis method (Watson et al (1970)), or the peak centroid method

Table 2.2 Illustration of the increased amount of data to be collected when the size of the crystallographic asymmetric unit increases as the molecule under study becomes more complex.

Compound	Number of atoms in asymmetric unit	Crystal planar spacing (Å)	Number of independent reflections
cephalosporin C	28	0.7	2580
vitamin B_{12}	95	1.0	3200
pig insulin	900	6 2.8 1.9 1.5	210 2100 7000 13154
6PGDH	4000*	6 2.6	1500 16000
PP_b	8000**	6 3.0	2500 20000
TMV disc	600,000	2.8	200000

*6PGDH is a dimer of M.W. = 100,000 with a monomer in the asymmetric unit.

**PP_b is a dimer of M.W. = 200,000 with a monomer in the asymmetric unit.

Tickle (1975)) or processed off line from the peak profile (O/P onto paper tape) by more sophisticated software (Diamond (1969), French (1975)).

In the ordinate analysis method the crystal is moved through the reflecting position in a series of 2N steps. The largest consecutive N counts are taken as the peak and the remainder as the background. This method allows for slight mis-setting of the crystal without the need to widen the peak scan range during automatic data collection. Unfortunately the weakest reflections (which can be less than the noise fluctuations in the scan) will be overestimated since it is a requirement that a peak be found. With multiple counter diffractometers this is not so much of a problem since the precise position of a weak reflection can be defined from stronger reflections in the quasi-simultaneous set, taking into account the quasi-simultaneous separation (Evans 1973, p. 47). Even for five weak reflections satisfactory results can be obtained. Occasionally, however, through extreme weakness of all five reflections or as a result of disturbance of the reflection profile by a noise pulse on one of the counters, an incorrect estimate is made,so even with the five counter machine off line profile fitting is considered an attractive proposition (Oatley (1976)).

The peak centroid method uses

$$\frac{\sum_{i=1}^{2N} (C_i - B)x_i}{\sum_{i=1}^{2N} (C_i - B)}$$

to compute the position of the peak, where C_i is the count and x_i the angular position for the ith step. B is the background level which is taken as the mean of the C_i and the summation is for $C_i > B$ only.

The peak is then taken as the sum of the N steps centred on this position and the sum of the remaining N steps is taken as background. This method overcomes the tendency for overestimation in the ordinate

analysis procedure.

Where extensive data handling facilities are available, accurate intensities and standard deviations can be determined by methods designed to analyse the peak profile. French (1975) considers five sources of information which can be used in the derivation of integrated intensities from a measured profile:

(a) the local behaviour of the background expected from the collection geometry;

(b) the properties expected in the peak shape, eg. continuity;

(c) the shapes of the peak already analysed. Diamond (1969) has shown that the peak shape varies slowly through reciprocal space;

(d) the position within the scan of the peak last measured and the reliability of the diffractometer in moving from one reflection to the next;

(e) for multiple counter diffractometers the positions of the peaks within the simultaneously collected scans will be defined by the collection geometry and, since peak shape varies little over small areas of reciprocal space, the shape is approximately the same for all counters.

Use of the ordinate analysis method is discussed in § 3 and of French's profile fitting method in § 4, as applied to the 6 Å studies of 6PGDH.

Other methods of estimating intensities on a diffractometer can be used, e.g. the method of Wyckoff designed originally to overcome the problems associated with non-uniformity of a broad source and high background encountered using a flow cell (Wycoff et al. (1967)). In this method a limited number of intensity measurements is made in the vicinity of the peak of a reflection at small intervals of ω. A fraction of these, such as the lowest 3 out of 6, is ignored and the sum of the rest taken as the peak count. The background for each peak is estimated from a background curve measured separately as a function of $(2\theta, \chi, \varphi)$ in regions of reciprocal space not including diffraction maxima.

This method has also been used successfully for more rapid data collection (e.g. Zappe, Krohne-Ehrich and Schultz (1977)).

(b) Photographic methods

For maximum speed, screenless photographic methods offer considerable gains (Arndt (1968)), for which all the reflections to a given resolution (see § 5) are recorded. This is especially useful for larger unit cells.

The two methods in common use are the screenless precession method (Xuong et al. (1968)) and the screenless rotation method (Arndt et al. (1973), Mosley (1973), Arndt and Wonacott (1977)).

The screenless precession method has the advantage that precession cameras are widely available in protein crystallography laboratories, but suffers from two disadvantages.

(a) for a small angle precession the Lorentz factor varies considerably, even within the area of each spot (Blundell and Johnson (1976)).

We have:

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

where $A^{-1} = 4 \cos\bar{\mu} \left[4\xi^2 \sin^2\bar{\mu} - (\sin^2\bar{\nu}_n - \sin^2\bar{\mu} - \xi^2)^2 \right]^{\frac{1}{2}}$

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

where:

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

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

$\cos\bar{\nu}_n = \cos\bar{\mu} - nd^*$ for the nth upper level of a set of reciprocal lattice planes separated by the distance d^* parallel to the X-ray beam;

τ is the angle between the horizontal axis of the universal joint and the projection of the reciprocal lattice vector $\underline{h}$ on to the zero layer;

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

ξ is a parameter defining the film radius (Blundell and Johnson (1976)).

There will be a minimum and maximum of ξ for each n within which spots

are recorded, i.e.

$$\xi_{max} = \sin\bar{\nu}_n + \sin\bar{\mu}$$

$$\xi_{min} = \sin\bar{\nu}_n - \sin\bar{\mu}$$

At the values $\xi = \xi_{min}$ or $\xi = \xi_{max}$ for a given $\bar{\mu}$, $A^{-1} = 0$ and $B > 0$. Hence, as $\xi \rightarrow \xi_{min}$ or ξ_{max}, L diverges rapidly, i.e. reflections on the edge of a recorded annulus spend a relatively longer period of time crossing the Ewald sphere than those in the centre of the annulus.

(b) the films contain large numbers of partially recorded reflections (the exact number being determined by the size of the precession angle) which are not contiguous from one film to the adjoining one. In fact, the film to film dial setting increment on the camera has to be chosen such that reflections partially recorded on one film are fully recorded on the next. Also, upper level reflections are split, though this splitting can be minimized by reducing the crystal to film distance by ~ 0.1 of its value (Xuong and Freer (1971)).

The screenless rotation method (see § 5) complements the advantages and disadvantages of the screenless precession method. Rotation cameras are not widely available and are expensive to purchase. But,

(a) the Lorentz factor for this geometry varies slowly except for a small number of reflections near the blind region (see § 5);

(b) contiguous partial reflections can be added to give structure amplitudes provided the camera motors are sufficiently free from backlash (Arndt et al.(1973), Mosley (1973)).

In both methods a restricted rotation/precession angle is used to reduce the number of overlapping reflections recorded on each exposure (see § 5).

With the increasing number of large proteins being studied in individual laboratories, existing precession cameras are being converted to work with both precession and rotation geometry (Huber (personal communication), Pickford (personal communication)). The disadvantage of

these cameras is that only one cassette can be exposed at a time and so the data collection is not an automatic or painless affair. The data collection on one crystal needs to be continuous because even on removal from the X-ray source some protein crystals continue to suffer radiation damage.

The application of the screenless rotation method to the 2.6 Å study of 6PGDH crystals is discussed in § 5.

2.3 Derivation of structure factors from intensities

Several corrections need to be applied to correct the intensity sets from various crystals into an internally scaled subset of a unique asymmetric unit for a particular space group (Dodson (1975)).

$$|F_{\underline{h}}|^2 = \frac{K.R.A.I_{\underline{h}}}{L\ P}$$

where K is a crystal scale factor;

R is a radiation damage correction;

LP is a Lorentz and polarization correction;

A is an absorption correction.

When several different crystals are used, the flexibility of the polypeptide chain can mean that different crystals may have slightly different structures (see § 3.1) and certainly those containing heavy atoms may be differently substituted so it is important to collect sufficient common data to establish that they are essentially the same. An accurate scale can be calculated using the method of Hamilton, Rollett and Sparks (1965) weighting each intensity with its standard deviation and excluding the very weak and very strong reflections. An alternative method is that of Fox and Holmes (1966) (see § 5).

The LP corrections for various geometries are readily available from International Tables (see § 2.2, § 5).

(a) Radiation damage

Radiation damage is an effect suffered to a greater or lesser extent by almost all protein crystals on exposure to X-rays and results in a decrease in intensity of the reflections during the course of the diffraction experiment. The effect may not be isotropic in reciprocal space, but invariably high angle reflections are affected most severely. Occasionally an initial increase in the intensities of low angle reflections occurs, presumably because of an initial temperature rise of the crystal on irradiation causing evaporation of mother liquor around the crystal and thus a lowering of the absorption. Radiation damage is wavelength dependent as can be expected from the increase in the absorption coefficient μ with increase in λ (Figure 2.6).

The means adopted to deal with radiation damage differ according to the instrument used to collect the data and according to the resolution of the data collected.

For diffractometer data standard reflections are monitored every 50 reflections or so to chart the radiation effects (and to detect crystal missetting) and a correction inversely proportional to $\sum_{\text{standards}} |F_h|^2$ is adequate for low resolution data. The effects of irradiation are more severe at higher angles, therefore for high resolution data collection a plane of reflections is collected before and after data collection and the variation of I with time and $\sin\theta$ calculated.

For high resolution screenless rotation data the first film of a data set is recollected at the end of data collection and an analysis performed of the strong fully recorded intensities in ranges of $\sin\theta$ in an analogous fashion to that for corresponding diffractometer data.

Blake and Phillips (1962) showed that the diffracted intensity I(t) after an exposure time t is related to the initial intensity I_o by the law:

$$I(t) = I_o\,(A_1 + A_2 \exp(-B_2 \frac{\sin^2\theta}{\lambda^2}))\quad .$$

The notation is explained in Figure 2.10.

Hendrickson et al. (1973) found that the data for the crystal structure analysis of sea lamprey haemaglobin at 2 Å conformed nicely to this law yielding a radiation damage factor:

$$R(t,\theta) = \exp-(k_1+k_3)t + \exp-k_3 t(1 - \exp-k_1 t)\exp-D\frac{\sin^2\theta}{\lambda^2}$$

Fletterick et al. (1976) used a slightly different scheme which was incorporated into a general formulation of the problem by Hendrickson (1976) where he considered all possibilities in the irreversible paths of transition of the irradiated protein (Figure 2.10).

The radiation damage corrections used in this study of 6PGDH crystals are discussed in § 3 and § 5.

(b) Absorption

For an isolated crystal the absorption factor $A(\underline{p},\underline{s})$ can be calculated using the formula:

$$A(\underline{p},\underline{s}) = \sum_{V_x} \exp(-\mu(p_{V_i} + s_{V_i}))\,\Delta V_i$$

where V_x and ΔV_i are respectively the volume and volume element of the crystal; p_{V_i} is the primary beam path length to the ith volume element; and s_{V_i} is the secondary beam path length from the ith volume element; μ is the crystal absorption coefficient.

Because of the glass and mother liquor present with a wet protein crystal mount, it is not possible to use this correction for protein crystal diffraction data. Even without the effect of the glass and mother liquor the number of reflections (and the correspondingly large number of primary and secondary path lengths) would make the calculation of all the $A(\underline{p},\underline{s})$ prohibitive.

On the diffractometer, where the crystal is smaller than the beam, a rapid correction can be obtained with the method of North et al. (1968).

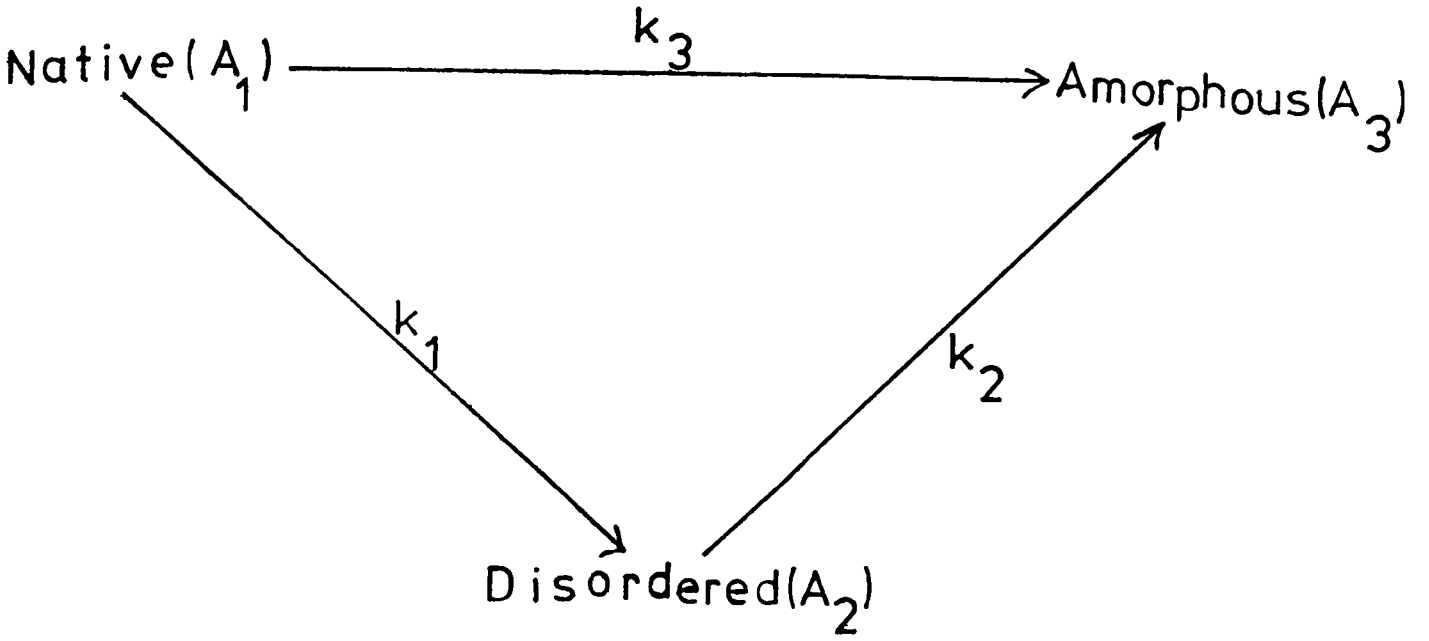


Figure 2.10

The irreversible paths of transition of the irradiated protein (after Hendrickson (1976)), where k_1, k_2, k_3, are rate constants; A_1, A_2, A_3 represent the respective fractions of each state, at time t, subject to the conditions: $A_1(t)+A_2(t)+A_3(t) = 1$; $A_1(0) = 1$, $A_2(0) = 0$, $A_3(0) = 0$. Hendrickson et al (1973) used $k_2 = k_3$. Fletterick et al. (1976) used $k_3 = 0$ and $k_1 = k_2$.

The absorption can be quickly checked at the end of data collection. Allowance can then be made for liquor movement with a time dependent North et al. (1968) correction if necessary (Wilson, personal communication).

North et al. (1968) mounted crystals with a reciprocal axis parallel to the spindle, φ, axis of the diffractometer and recorded the intensity variation of a strong axial reflection as the crystal was turned through φ .

Now $A(\underline{p},\underline{s}) = A(\varphi_p, \mu_p, \varphi_s, \mu_s)$ (see Figure 2.11), and the following approximations are made.

$$A(\underline{p},\underline{s}) \approx \tfrac{1}{2}(A(\underline{p}) + A(\underline{s}))$$

$$A(\underline{p}) = \frac{I_{max}}{I_{\varphi}}$$

i.e. the μ dependence of the absorption correction is ignored.

Clearly, for any crystal an axial reflection with low sin θ will have quite a different path length from one with high sin θ from any-particular V_i and although the ratio of I_{max}/I_{φ} will be equal for both, $A(\underline{p},\underline{s})$ should be quite different.

Also, $A(\underline{p}) = A(\underline{s})$ and $A(\underline{p}) \equiv A(-\underline{s})$ is assumed which means that equivalent reflections will have identical $A(\underline{p},\underline{s})$ which will not be the case for an asymmetric crystal or mother liquor distribution (Figure 2.12).

Kopfmann and Huber (1968) and Huber and Kopfmann (1969) suggested a method which avoids these two errors. The approximations are made:

$$A(\underline{p},\underline{s}) \approx A(\underline{p}).A(\underline{s})$$

where $A(\underline{p}) \equiv A(-\underline{s})$

but $A(\underline{p}) \neq A(\underline{s})$.

This approach is more accurate but lengthy since it involves the monitoring of sufficient intersecting curves, by rotating a crystal about the normal to a given general diffraction plane, and finding values for $A(\varphi, \mu)$ at intervals of φ and μ of 12° to 15°. From these results a transmission surface can be generated. This gives a better correction at the cost of more irradiation time of the crystals.

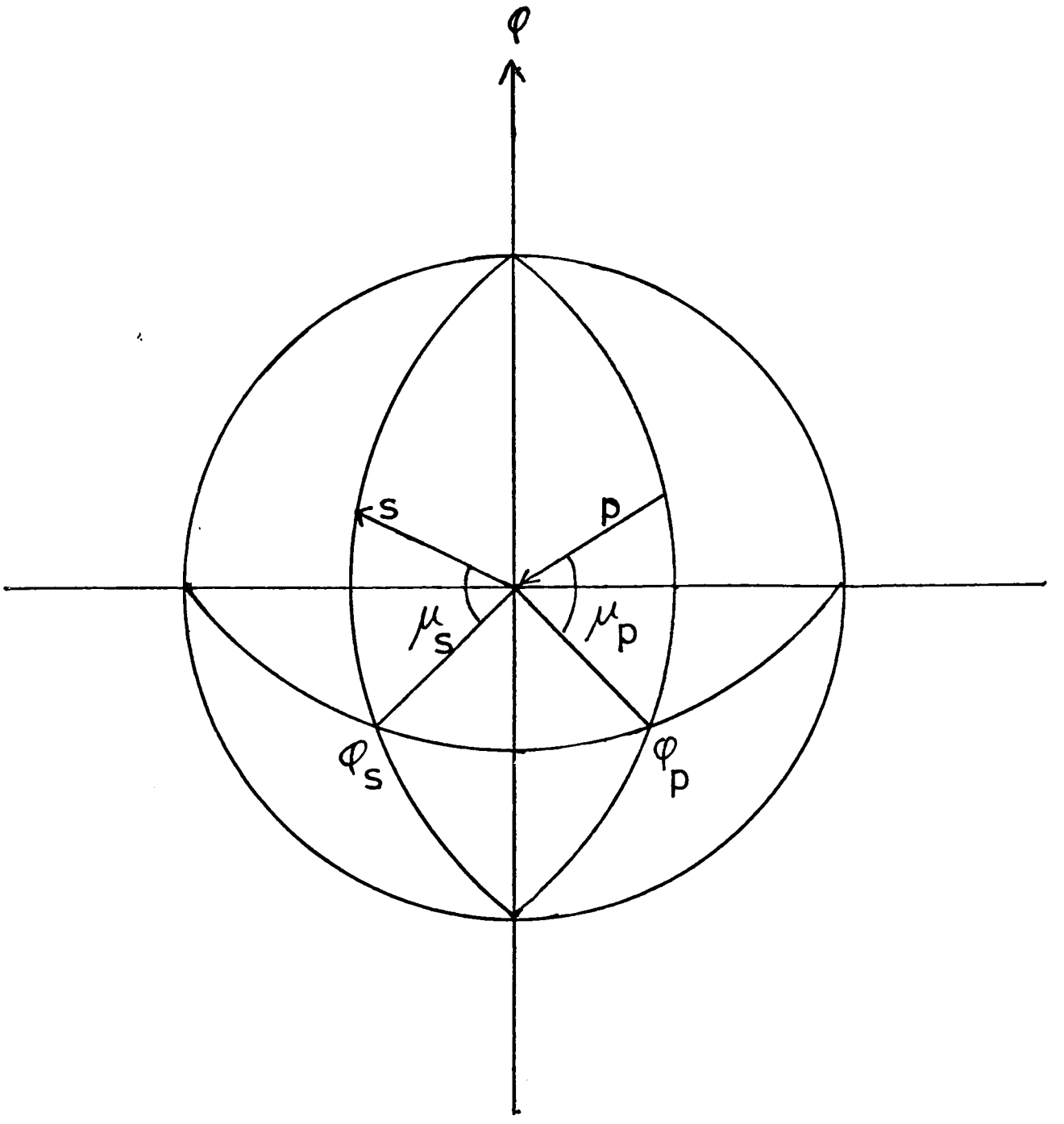


Figure 2.11

Schematic diagram showing the angles $\varphi_{\underline{P}}, \varphi_{\underline{S}}$, $\mu_{\underline{P}}$ and $\mu_{\underline{S}}$, where φ marks the φ rotation axis of the diffractometer and the subscripts P and S refer to the primary and secondary beam path lengths respectively.

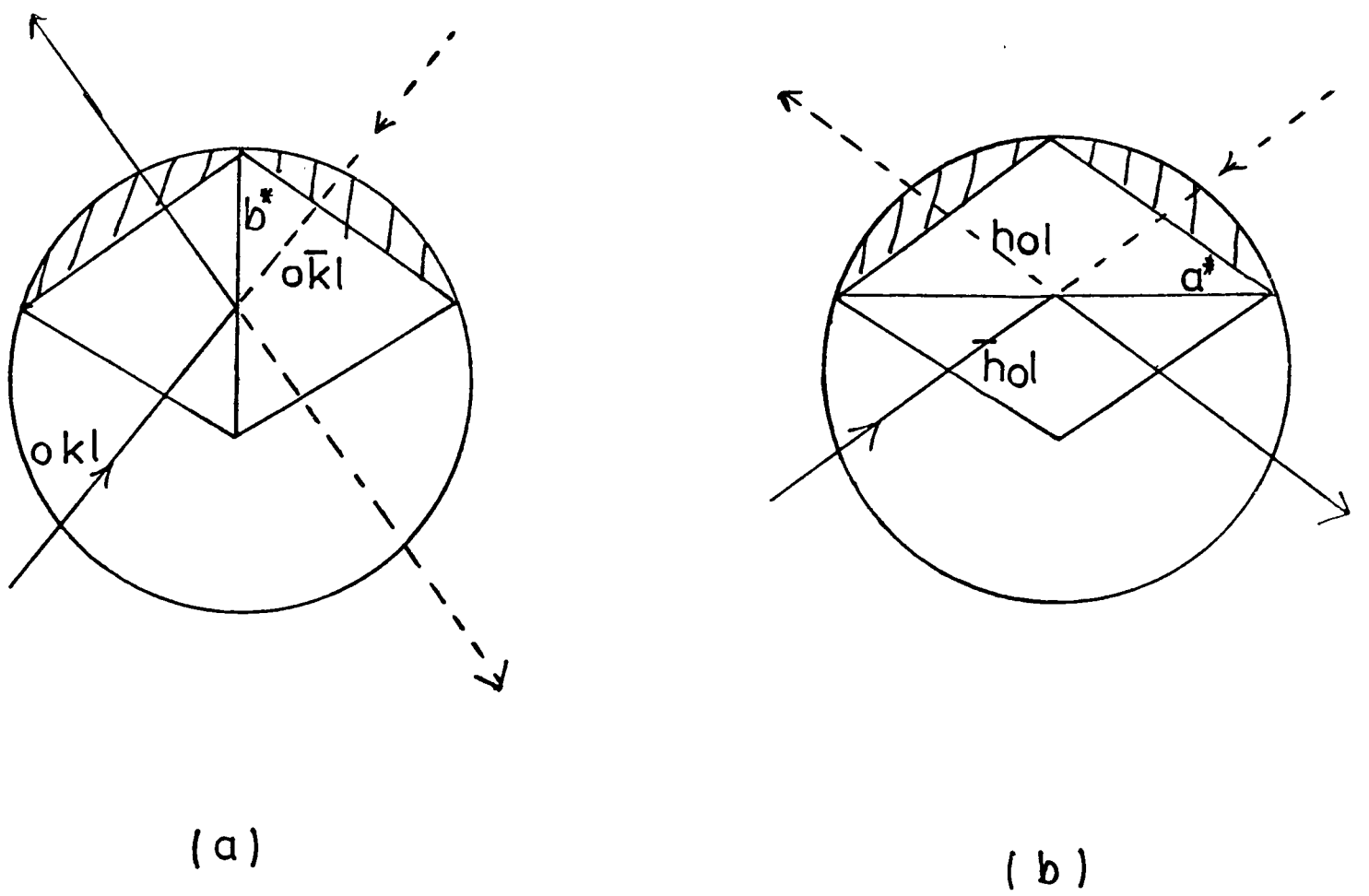


Figure 2.12

Asymmetric mount due to the mother liquor distribution in the capillary. Anomalous scattering differences would be seriously affected in (b) but not in (a). From North et al. (1968).

With the screenless rotation method the strategy adopted to deal with absorption depends on whether the crystal is bigger or smaller than the incident beam and is more fully discussed in § 5.

2.4 Location of heavy atom sites

The position of the heavy atom sites in the unit cell are required in determining the phase. In the absence of any phase information (as is the case with the first derivative of a particular protein crystal studied) use can be made of the Patterson function or direct methods. In the later stages of a structural study, a difference Fourier map can be calculated using structure amplitudes of an unsolved derivative and the best set of phases available (see § 2.7).

(a) the difference Patterson

$$P(uvw) = \frac{2}{V} \sum_{hkl} |\underline{F}_H|^2 \cos 2\pi(hu+kv+lw)$$

where $|\underline{F}_H| = \big||\underline{F}_{PH}| - |\underline{F}_P|\big|$ for a given hkl, contains peaks corresponding to inter-heavy atom vectors between symmetrically related positions in the unit cell as well as peaks corresponding to cross vectors between different heavy atom sites.

The estimate $\big||\underline{F}_{PH}| - |\underline{F}_P|\big|$ for $|\underline{F}_H|$ is exact only for centric reflections but the Patterson synthesis with terms $\big||F_{PH}| - |F_P|\big|^2$ is a good approximation to the true $|\underline{F}_H|^2$ map even for an acentric structure. Anomalous scattering information if significant can be used to calculate F_{HLE}'s (see § 2.5), (Dodson and Vijayan (1971), Dodson, Evans and French (1975)) and can give an improved estimate of acentric $|\underline{F}_H|$ over the simpler $\big||\underline{F}_{PH}| - |\underline{F}_P|\big|$.

An alternative method that can be applied in solving the heavy atom positions without phase information is to use direct methods.

(b) Direct methods

The theoretical basis of direct methods was laid down in the papers

of Karle and Hauptmann (1950), Sayre (1952), Cochran (1952) and Zachariasen (1952) and reviewed recently by Woolfson (1971). The implementation of the popular direct methods computer programme for structure determination, MULTAN (Germain, Main and Woolfson (1971)) for use in locating heavy atoms in protein crystals has been fully discussed by Wilson (1974). Other reports on the usefulness of this method in protein crystallography have been given (Steitz (1968), Neidle (1973), Navia and Sigler (1974)).

Direct methods can be used for ab initio phase determination of crystal structures with a small number of atoms (less than $\sim$ 200 atoms). The isomorphous differences correspond to such a structure, i.e. the satellite configuration of heavy atom positions in the unit cell and so this configuration too can be solved ab initio. The use of MULTAN in the solution of the $KAu(CN)_2$ and $K_2Pt(CN)_4$ derivatives of 6PGDH is discussed in § 3.

Sayre (1952) derived relationships between the signs of centro-symmetric structure factors, which if given a starting set of phases enable the phases of related reflections to be determined leading to a solution of the structure.

Consider a structure containing equal resolved atoms. The square of equation 2.2 is

$$\rho^2(\underline{x}) = \frac{1}{V^2} \sum_{\underline{h}_1}' \sum_{\underline{h}_2}' F_{\underline{h}_1} F_{\underline{h}_2} \exp\left[-2\pi i(\underline{h}_1 + \underline{h}_2).\underline{x}\right] \qquad 2.3$$

Putting $\underline{h}_1 + \underline{h}_2 = \underline{h}$ gives

$$\rho^2(\underline{x}) = \frac{1}{V} \sum_{\underline{h}}' \left(\frac{1}{V} \sum_{\underline{h}'}' F_{\underline{h}'} F_{\underline{h}-\underline{h}'} \right) \exp(-2\pi i \underline{h}.\underline{x}) \qquad 2.4$$

By comparison of 2.2 and 2.4 we have:

$$G_{\underline{h}} = \frac{1}{V} \sum_{\underline{h}'}' F_{\underline{h}'} F_{\underline{h}-\underline{h}'}$$

as the hth Fourier coefficient of $\rho^2(\underline{x})$.

If $\rho(\underline{x})$ shows a number of equal resolved atoms then $\rho^2(\underline{x})$ shows the same atoms but of different shape. If f and g are the scattering factors for each $\underline{h}$ in each case then:

$$F_{\underline{h}} = f \sum_{j=1}^{N} \exp(2\pi i \underline{h}.\underline{x}_j)$$

and

$$G_{\underline{h}} = g \sum_{j=1}^{N} \exp(2\pi i \underline{h}.\underline{x}_j)$$

$$G_{\underline{h}} = \frac{g}{f} \cdot F_{\underline{h}}$$

Therefore,

$$F_{\underline{h}} = \theta_h \sum_{\underline{h}'} F_{\underline{h}'} \; F_{\underline{h}-\underline{h}'} \qquad 2.5$$

$\theta_h = f/gV$ allows for the sharpening of the squaring process.

From 2.5, Sayre's equation, for a centrosymmetric structure , if one of the products on the right hand side is large, then it may so bias the summation that there is a high probability that $F_{\underline{h}}$ has the same sign, i.e.,

$$s(\underline{h}).s(\underline{h}').s(\underline{h}-\underline{h}') \approx +1$$

Even when the structure factors are too small to give inequality information a probability relationship exists. A general relationship between phases rather than just signs can be derived (Karle and Karle (1966)) and is known as the tangent formula, i.e.:

$$\tan \varphi_{\underline{h}} = \frac{\sum_{\underline{h}'} E_{\underline{h}'} E_{\underline{h}-\underline{h}'} \sin(\varphi_{\underline{h}'} + \varphi_{\underline{h}-\underline{h}'})}{\sum_{\underline{h}'} E_{\underline{h}'} E_{\underline{h}-\underline{h}'} \cos(\varphi_{\underline{h}'} + \varphi_{\underline{h}-\underline{h}'})} \qquad 2.6$$

where $E_{\underline{h}} = \dfrac{F_{\underline{h}}}{\langle |F_{\underline{h}}|^2 \rangle^{\frac{1}{2}}}$

and is the normalized structure factor for reflection $\underline{h}$.

Inequality relationships between structure factors are used to obtain the phases of a small number of structure factors (Harker and Kasper (1948)).

For ab initio phasing the starting phase set contains those phases chosen to specify the origin and enantiomorph, those determined with a high degree of reliability from the inequality relationships and a small number (2 or 3) of general phases which are either given symbols or allowed to take on a number of different values giving rise to several sets of phases. From a careful choice of the starting set (see §3), phases can be developed for the remaining structure factors by use of the tangent formula.

A discriminating figure of merit criterion is required to distinguish the sets of phases developed, which though it may not enable the correct solution to be picked out, will at least eliminate the majority of the incorrect ones. The indicators used in judging the phasing (given by MULTAN) are:

(i) the absolute figure of merit ABSFOM (Germain et al.(1971) which is based on the value of

$$\chi = \sum_{\underline{h}} E_{\underline{h}} E_{\underline{h}'} E_{\underline{h}-\underline{h}'}$$

which for a correct set of signs should be large.

(ii) minimum γ_o, PSIZERO (Cochran and Douglas (1955))

$$\gamma_o = \sum_{\underline{h}} \left| \sum_{\underline{h}'} E_{\underline{h}'} E_{\underline{h}-\underline{h}'} \right|$$

is calculated for each set of signs. The $E_{\underline{h}}$ selected are zero or very small and are not involved in the sign determining process. The γ_o figure of merit is related to Sayre's equation for, if the summations over $\underline{h}'$ contained all the observed data, each of them should be zero. PSIZERO should therefore be a minimum for the correct set.

(iii) the residual R_K (Karle and Karle (1966))

$$R_K = \frac{\sum_{\underline{h}} \left| |E_{\underline{h}}| - k \left| \langle E_{\underline{h}} E_{\underline{h}-\underline{h}'} \rangle_{\underline{h}'} \right| \right|}{\sum_{\underline{h}} |E_{\underline{h}}|}$$

and k is computed from:

$$k = \frac{\sum_{\underline{h}} |E_{\underline{h}}|}{\sum_{\underline{h}} |\langle E_{\underline{h}}, E_{\underline{h}-\underline{h}'} \rangle_{\underline{h}'}|}$$

The value of R_K (RESID) should be a minimum for the correct phase set.

(iv) A combined figure of merit COMFOM:

$$\text{COMFOM} = \frac{\text{ABSFOM-MINABSFOM}}{\text{MAXABSFOM-MINABSFOM}} + \frac{\text{MAXPSIZERO-PSIZERO}}{\text{MAXPSIZERO-MINPSIZERO}} + \frac{\text{MAXRESID-RESID}}{\text{MAXRESID-MINRESID}}$$

where the MAX and MIN refer to the maximum and minimum values for the criteria in all the phase sets.

On the basis of these figures of merit, the best phase set can be selected and an EMAP calculated giving a possible solution of the heavy atom structure (see § 3).

The solution obtained needs to be cross checked with the Patterson synthesis, though this needs to be done carefully since Patterson and direct methods are closely related (through the squaring process) and incorrect E maps will contain atomic positions related by vectors occurring in the Patterson map (Blundell and Johnson (1976) p. 351). This was confirmed in the case of 6PGDH with the $K_2Pt(CN)_4$ derivative (see § 3). It may be thought that the problem can be circumvented, however, by comparison of F_{HLE}^2, $\left||\underline{F}_{PH}| - |\underline{F}_P|\right|^2$ Patterson syntheses and E maps calculated using F_{HLE}'s and $|\underline{F}_{PH}| - |\underline{F}_P|$. Unfortunately, experience shows that E's calculated using F_{HLE}'s seem to be much less successful than if isomorphous differences are used (Wilson (1974) p. 62). This was also the case for the $K_2Pt(CN)_4$ derivative of 6PGDH (§ 3).

2.5 Heavy atom parameter refinement

In solving the phase problem by isomorphous replacement the parameters characterizing the heavy atom substitution need to be accurately defined as does the scale factor relating native to derivative measurements.

Conventional crystallographic refinement minimizes

$$M = \sum \omega (|\underset{\sim}{F}_{OBS}| - |\underset{\sim}{F}_{CALC}|)^2$$

by the method of linear least squares (Woolfson (1970) p. 347).

The strategy adopted depends on whether independent phase information is available or not. The situation where no such information is available will be considered first.

Anomalous scattering information can be used to calculated F_{HLE}'s (Dodson and Vijayan (1971)), Figure 2.13, which can be used as the F_{OBS} in a full 3-D least squares refinement.

By the cosine rule:

$$|\underline{F}_H|^2 = |\underline{F}_P|^2 + |\underline{F}_{PH}|^2 - 2|\underline{F}_P||\underline{F}_{PH}| \cos\alpha$$

where $|\underline{F}_{PH}| = K e^{-Bs^2} |\underline{F}_{PH}'|$, K and B are relative scale and temperature factor of derivative with respect to native, $|\underline{F}_{PH}'|$ is the measured derivative structure amplitude.

$$\Delta_{ANO} = |\underline{F}_{PH}^{+}| - |\underline{F}_{PH}^{-}| \approx 2 |\underline{F}_H''| \sin\beta$$

$$\Delta_{ISO} = |\underline{F}_{PH}| - |\underline{F}_P| \approx |\underline{F}_H|$$

By the sine rule:

$$\frac{\sin' \alpha}{|\underline{F}_H|} = \frac{\sin\beta}{|\underline{F}_P|} = \frac{\Delta_{ANO}}{2|\underline{F}_H''||\underline{F}_P|}$$

Define $k = k(\theta)$

$$= \frac{|\underline{F}_H|}{|\underline{F}_H''|} = \frac{f_o + f'}{f''} ,$$

the ratio of the real to anomalous scattering of the heavy atom in the derivative crystal. This equation is based on the assumption that the

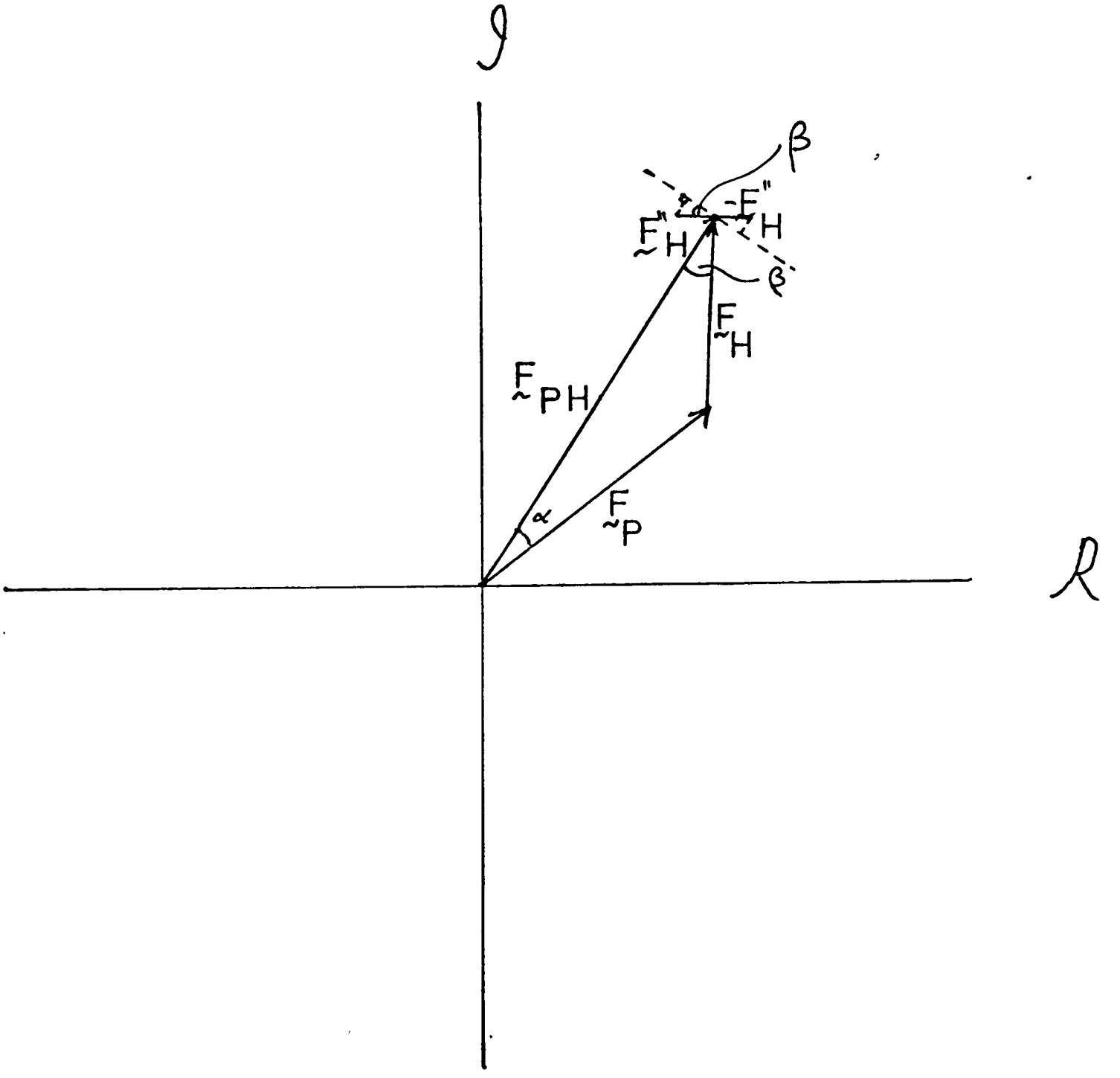


Figure 2.13

Vector diagram showing the relation between $\underset{\sim}{F}_P$, $\underset{\sim}{F}_{PH}$, $\underset{\sim}{F}_H$ and $\underset{\sim}{F}_H''$, where $\alpha = \alpha_{PH} - \alpha_P$ and $\beta = \alpha_H - \alpha_{PH}$.

heavy atom is not part of a complex ion (see Matthews (1966)).

$$\cos\alpha = \pm \sqrt{1 - \sin^2\alpha}$$

$$|\underline{F}_H|^2 \approx |\underline{F}_P|^2 + |\underline{F}_{PH}|^2 \mp 2|\underline{F}_P||\underline{F}_{PH}| \left[1 - \left(\frac{k\Delta_{ANO}}{2|\underline{F}_P|} \right)^2 \right]^{\frac{1}{2}}$$

Using the approximation $(1 - x)^{\frac{1}{2}} \approx 1 - \frac{1}{2}x$, for x small,

and taking the "-" sign and writing $|\underline{F}_H| = F_{HLE}$, we get:

$$F_{HLE}^2 \approx \Delta_{ISO}^2 + \left(\frac{k\Delta_{ANO}}{2} \right)^2 \qquad 2.7$$

The few reflections where the "+" sign should have been taken, for which $|\underline{F}_H| = F_{HUE}$ can be excluded. For centric reflections, $\Delta_{ANO} = 0$ and the equalities follow:

$$F_{HLE} = (|\underline{F}_{PH}| - |\underline{F}_P|)$$
$$F_{HUE} = (|\underline{F}_{PH}| + |\underline{F}_P|) \text{ , corresponding to "cross over".}$$

According to Dodson, Evans and French (1975):

$$E(\Delta^2) = (E(\Delta))^2 + Var(\Delta)$$

Therefore from 2.7:

$$E(F_{HLE}^2) \approx (E(\Delta_{ISO}))^2 + \frac{k^2}{4}(E(\Delta_{ANO}))^2 + \sigma^2_{ISO} + \frac{k^2}{4}\sigma^2_{ANO}$$

i.e. F_{HLE}^2 and F_{HLE} are overestimated. Generally, σ^2_{ISO} can be neglected compared with $\frac{k^2}{4}\sigma^2_{ANO}$ so that for acentric reflections a bias correction needs to be applied replacing F_{HLE} by $F_{HLEB} = F_{HLE} - BIAS$ where $BIAS = \frac{k^2}{4}\sigma^2_{ANO}$.

An alternative method for reducing this bias is to use an estimate for k less than $\frac{f_o + f'}{f''}$ $(= k_{theor})$ and known as k_{emp} $(= 2\langle\Delta_{ISO}\rangle / \langle\Delta_{ANO}\rangle)$, (Matthews (1966)). Because the errors associated with Δ_{ANO} are large k_{emp} is less than k_{theor} in practice. This method has the weakness of treating all reflections in a given sin θ annulus (remember k = k(θ)) similarly and disregards the different standard deviations of the Δ's for different reflections. However, the calculation of BIAS needs standard deviations which truly reflect the error distribution (see § 3, § 5).

Because of the problem of getting a reliable estimate of $|\underline{F}_H|$ for acentric reflections, crystal structure analyses where a sufficiently high percentage of the data is centric use the centric data only and Δ_{ISO} as $|\underline{F}_H|$ (i.e. F_{HOBS}) in the refinement. A centric least squares refinement was preferred with 6PGDH crystals (space group $C222_1$, 23% of the data being centric at 6 Å resolution), see § 3.

For both centric and acentric refinement, progress can be monitored by inspecting the following quantities:

$$\text{(a)}\quad R = \sum \left(\frac{(F_{HOBS} - F_{HCALC})}{F_{HOBS}} \right),$$

where F_{HCALC} is calculated from the occupancies and coordinates of the heavy atoms according to equation 2.1;

$$\text{(b)}\quad wR = \sum \frac{w(F_{HOBS} - F_{HCALC})^2}{F_{HOBS}^2} \quad ;$$

(c) slope, g, of the plot of r.m.s. F_{HCALC} for different ranges of F_{HOBS} against the median F_{HOBS} of that range;

(d) goodness of fit, G.F. $= \left(\sum w \Delta^2/(r-s) \right)^{\frac{1}{2}}$, where r is the number of reflections and s the number of parameters refined and $\Delta = (F_{HOBS} - F_{HCALC})$;

(e) $w\Delta^2$ as a function of $\sin^2\theta/\lambda^2$;

(f) the difference Fourier synthesis, $\Delta\rho$, with terms $\Delta \exp i\alpha_H$.

Note: $w = 1/\sigma^2 (F_{HOBS})$; $\sigma^2(F_{HOBS}) = \sigma^2(F_P) + \sigma^2(F_{PH})$; $\sigma'(F) \approx \frac{\sigma(I)}{2F}$; $\sigma(I) = (I + b)^{\frac{1}{2}}$ where $(I + b)$ is the total count in a diffractometer scan; $\sigma(F) = \sigma'(F) + k' F$ where k' is an instability constant compensating for the under-estimation by counting statistics of the intensity errors especially of stronger reflections (Stout and Jensen (1968)) and is an approximation to a formula suggested by McCandlish et al. (1975). For film data, of course, counting statistics are not applicable and the $\sigma(I)$ are derived from the agreement between

equivalent reflections.

The standard deviation of Δ_{ISO} is generally $\approx \Delta_{ISO}$ so R is usually high. A good solution gives R = 40% but quite useable ones have R = 60%. A totally incorrect solution gives R = 83%. A wrong solution will give ω R $\approx$ R. The variation of $\langle \omega \Delta^2 \rangle$ with sinθ should be constant for each range and if it is much higher for one than the others the refinement process will try to fit heavy atom parameters to minimize that range at the expense of the others. Sometimes low angle reflections or those with the highest Δ_{ISO} dominate M completely. Check (e) should yield a slope of 1.0 but a gradient of more than 0.3 usually shows correct placement of most of the heavy atom scattering material. The difference Fourier can show minor binding sites or disturbances perhaps associated with two binding sites close together or incorrect temperature factors or perhaps lack of isomorphism in the $\underline{F}_P$ and $\underline{F}_{PH}$ series (Dodson (1975)).

When independent phase information is available a 3-D refinement using the minimisation function:

$$M = \sum m \left(|\underline{F}_{PH}| - |\underline{F}_P + \underline{F}_H| \right)^2$$

can be performed, where m = figure of merit (see §2.6). Usually, only those terms with high m are used otherwise serious underestimation of occupancies tends to occur (Dodson (1975)). It is better to exclude the derivative to be refined from the phasing so that the phases and the parameters to be refined are independent. This precludes the use of alternating cycles of phasing and refinement which are sometimes used leading to unusually high figures of merit. Phase refinement is most suitable when at least,and preferably more than,two good independent derivatives are available which do not possess common sites.

The criteria for judging this refinement are simpler than those for refinement on $(F_{HOBS} - F_{HCALC})^2$ with $F_{HOBS} = F_{HLE}$ or Δ_{ISO}, since it is unlikely that the heavy atoms have been wrongly placed. They are:

(a) The difference between the "initial Kraut" R and the

"final Kraut" R;

(b) the difference Fourier with the terms

$$\sum_{hkl} W\,(F_{PHOBS} - F_{PHCALC})\,\exp(-2\pi i\alpha_{PHCALC})$$

for all reflections, not just those used for refinement.

The Kraut R factor is defined as:

$$\frac{\sum_{hkl} |F_{PHOBS} - F_{PHCALC}|}{\sum_{hkl} F_{PHOBS}}$$

Since this is always a small number it is only informative in comparison with the initial Kraut R which is:

$$\frac{\sum_{hkl} |F_{PHOBS} - |\underline{F}_P||}{\sum_{hkl} F_{PHOBS}}$$

Note: $F_{PHOBS} = |\underline{F}_{PH}|$; $F_{PHCALC} = |\underline{F}_P + \underline{F}_H|$.

These R factors are analysed in ranges of sin θ and $|\underline{F}_{PH}|$ and a check made that the final values are always significantly less than the initial ones.

The difference Fourier is valuable in much the same way as for F_{HLE} refinement but also can highlight errors in the relative scale and temperature factor between $|\underline{F}_P|$ and $|\underline{F}_{PH}|$; these are suspect if the value of the Fourier at all the sites is of the same sign, and above background.

2.6 Calculation of the primary phase set

In § 2.1 it was shown that in the absence of errors the phase of the native structure factor can be known exactly for the centric reflections using a single heavy atom derivative, the acentric reflections with a single derivative and anomalous scattering information or with two derivatives in the absence of the anomalous differences. In practice, errors are a

serious consideration and all possible sources of information need to be incorporated for an accurate evaluation of the phase. The effect of the errors is to prevent the structure amplitude circles intersecting at unique points.

Blow and Crick (1959) made the assumption that for acentric reflections all the errors (experimental and those from non-isomorphism) lie in the magnitude of $\underline{F}_{PH}$.

Then for any phase angle α, the probability of that angle being correct is:

$$P(\alpha) = N \exp\left(\frac{-\xi^2}{2E^2}\right)$$

which is a maximum when $\xi = 0$ and

$$\xi(\alpha) = |\underline{F}_{PH}| - |\underline{F}_P + \underline{F}_H|$$

is the lack of closure error.

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

is the average error.

Anomalous scattering information can be incorporated similarly (North (1965), Matthews (1966)).

$$P_{an}(\alpha) = \exp\left(\frac{-\varepsilon''^2}{2E''^2}\right)$$

$$\varepsilon''(\alpha) = \Delta_{ANO} - \Delta_{CALC} :$$

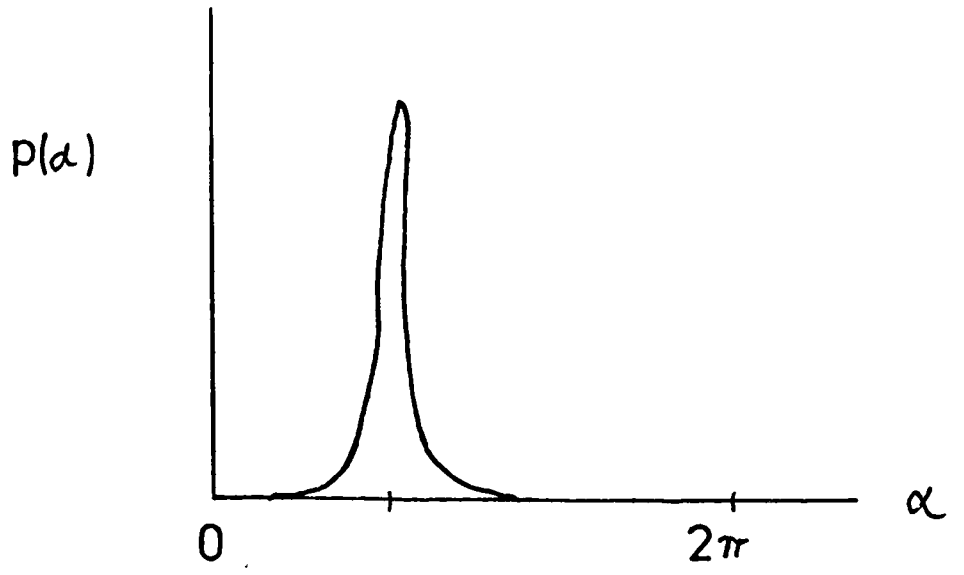
and $\Delta_{CALC} = \dfrac{2|\underline{F}_P||\underline{F}_H''| \sin(\alpha_P - \alpha_H)}{|\underline{F}_{PH}|}$; here $\Delta_{ANO} = F_{PH}^{+} - F_{PH}^{-}$

For several derivatives the probabilities may be multiplied. A discussion of the estimation of E^2 and E'^2 is given in § 3.

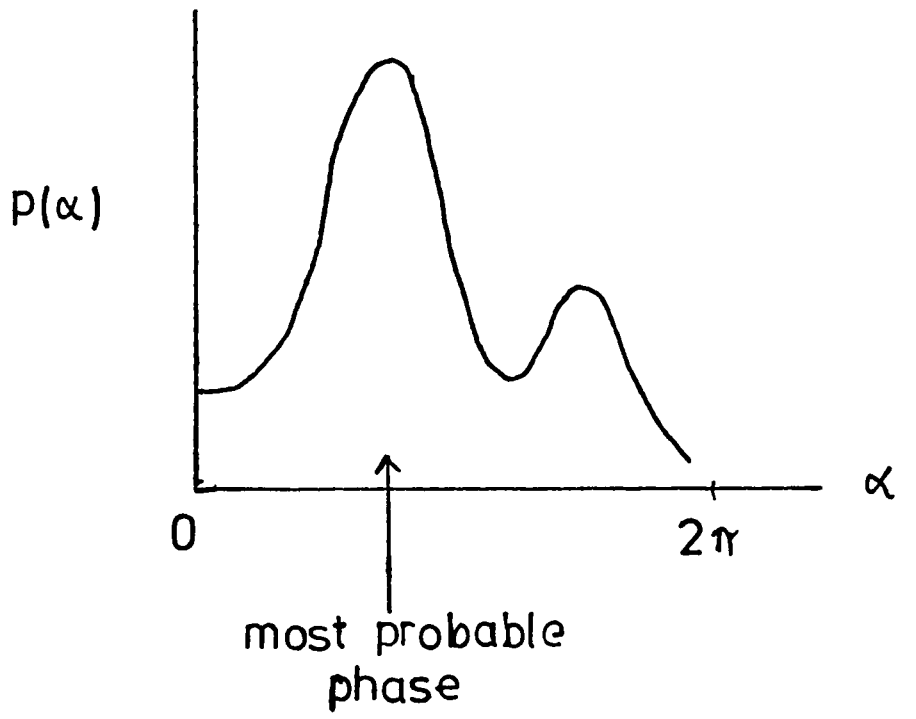
Blow and Crick (1959) showed that if the electron density map is calculated using phases corresponding to the centroid of the probability of the phase distribution the map has the smallest r.m.s. error.

The centroid coordinates (x,y) are calculated as:

$$x = \frac{\int_0^{2\pi} \rho(\alpha) \cos\alpha \, d\alpha}{\int_0^{2\pi} \rho(\alpha) \, d\alpha} \quad ; \quad y = \frac{\int_0^{2\pi} \rho(\alpha) \sin\alpha \, d\alpha}{\int_0^{2\pi} \rho(\alpha) \, d\alpha}$$



a good case



a bad case

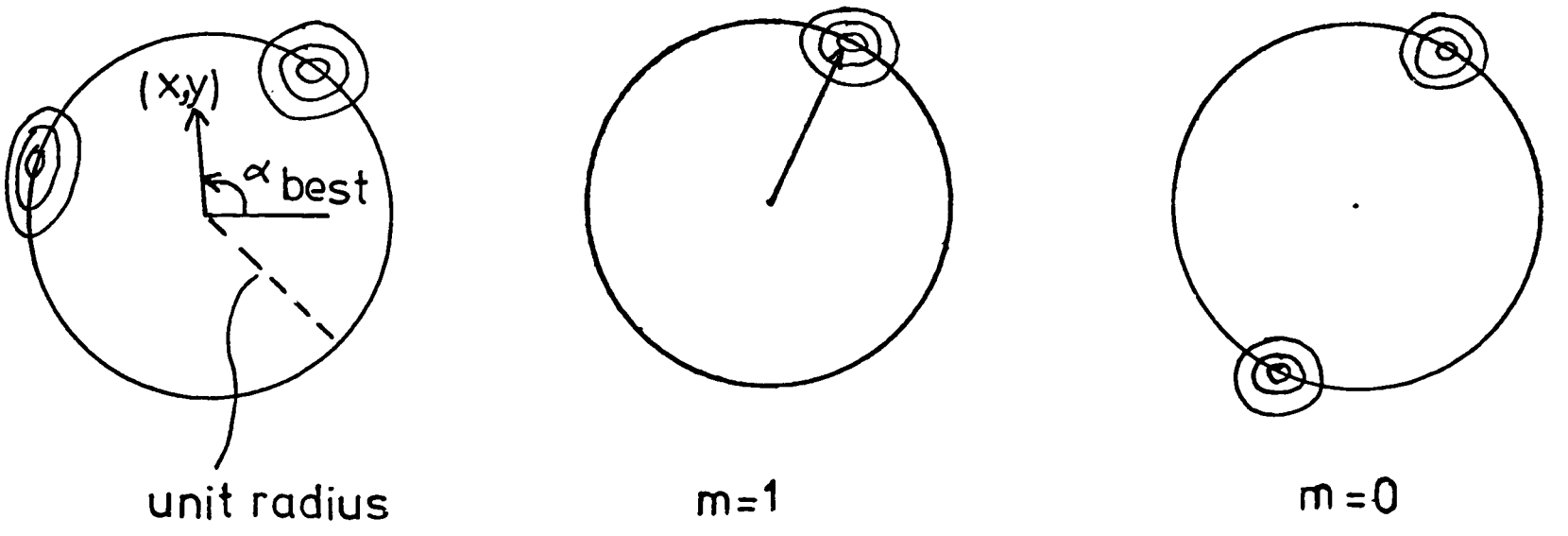


Figure 2.14 Some examples of phasing and the meaning of figure of merit.

$\tan \alpha_{best} = y/x$; $m = (x^2 + y^2)^{\frac{1}{2}}$ and is the magnitude of the vector from the origin and is indicative of the quality of the phase determination for that reflection and is known as the figure of merit (see Figure 2.14).

The calculated electron density map for the native protein is:

$$\rho(\underline{x}) = \frac{1}{V} \sum_{\underline{h}} |\underline{F}_{\underline{h}}| \, m \, e^{i\alpha_{best}} \exp(-2\pi i \underline{h}.\underline{x}) \qquad 2.8$$

2.7. The difference Fourier technique in protein crystallography

Once the structure of the protein has been established, even at low resolution, it is of interest to study the binding of small molecules to that protein crystal form. In practice, crystals are soaked in a solution containing a concentration of the small molecule that is higher than its equilibrium binding constant. There is no guarantee that the small molecule will bind at the same site as in the dissolved protein, in fact, it may not bind at all if the binding site is inaccessible in the crystal. In the case of enzymes where the mechanism of enzyme catalysis is of primary interest, these studies can involve the binding of substrates, coenzymes and inhibitors (see Appendix III). The latter may be required in studies aimed at designing drugs and/or antibacterial agents (Helliwell (1977)). The difference Fourier technique is also used in the solving of new heavy atom derivatives when the α_p are available.

The difference Fourier $\Delta\rho$ is defined by:

$$\Delta\rho_{obs} = \frac{1}{V} \sum_{\underline{h}} m(|\underline{F}_{PI}| - |\underline{F}_P|) \exp i\alpha_p . \exp(-2\pi i \underline{h}.\underline{x}) \qquad 2.9$$

(where I $\equiv$ the bound molecule) which is an approximation to the true structure:-

$$\Delta\rho_{true} = \frac{1}{V} \sum_{\underline{h}} |\underline{F}_I| \exp i\alpha_I . \exp(-2\pi i \underline{h}.\underline{x})$$

In centric projections ($\alpha = 0$ or π), $(|\underline{F}_{PI}| - |\underline{F}_P|) \exp i\alpha_p$ is equivalent to $|\underline{F}_I| \exp i\alpha_I$ except in rare cases of "cross over". Hence, for the centric terms the difference Fourier contains an accurate representation of the

bound molecule with the atoms at full height.

The acentric situation is more complicated but it can be shown (Blundell and Johnson (1976)) that a difference Fourier synthesis based on coefficients ($|\underline{F}_{PI}| - |\underline{F}_P|$) exp $i\alpha_P$ contains features accurately representing the additional atoms not included in the phasing but these atoms will appear at half their true heights and there will be some background noise. This is corrected for by multiplying the coefficients ($|\underline{F}_{PI}| - |\underline{F}_P|$) of the acentric terms by 2.

Because of further uncertainties in the native phases, the peak heights will be further reduced. If these uncertainties are monitored correctly by the figure of merit m, then the mean contribution of the structure factor m ($|\underline{F}_{PI}| - |\underline{F}_P|$) exp$i\alpha_P$ to the unknown correct structure has been shown to be $m^2(|\underline{F}_{PI}| - |\underline{F}_P|)$ (Blundell and Johnson (1976)).

It is important that the measurements are as precise as possible so that small changes in structure factor amplitude are greater than the standard deivations of their measurements.

Because the structures of the active sites of enzymes are extremely responsive to the binding of substrates and substrate analogues, conformational changes may take place which give rise to cracking or disorder of the crystal. It is important,however, that the crystals of protein plus inhibitor be closely isomorphous with the native protein crystals, at least with respect to the cell dimensions so that the molecular transform is sampled at the same reciprocal lattice points.

With 6PGDH crystals, which have had diffused coenzyme and a coenzyme analogue bound to the enzyme molecules, the cell dimensions are very close to those of the native. This work is discussed in § 4.

In those instances where large conformational changes occur, de novo structure determination has to be attempted.

CHAPTER 3

PRELIMINARY CRYSTALLOGRAPHIC WORK AND 6 Å STRUCTURE DETERMINATION OF 6PGDH FROM SHEEP LIVER

3.1 Crystallization

Crystals were first reported by Silverberg, Dalziel and Adams (1973) and are long needles. The largest crystals have a volume of 0.6 x 0.25 x 2.0 mm^3. They grow from a gel which forms a 5 mg enzyme/ml solution in 50 mM potassium phosphate, 55% saturated with ammonium sulphate (pH 6.5). The variation of pH from 6 to 8.5 affects the detailed crystal morphology but does not affect the space group or cell dimensions. The crystals have been grown on occasion from ammonium sulphate concentrations ranging from 45% to 60% saturation. The crystals can be grown in 8/32 dialysis tubing but large crystals are grown most reliably from 0.25 ml of solution in small glass tubes 5 mm in diameter. Figure 3.1 shows a photomicrograph of the 6PGDH crystals.

3.2 Preliminary crystallography and derivative search

X-ray precession photographs show the crystals to be orthorhombic, space group C_{222_1}, with eight general positions in the unit cell and cell dimensions in Å:

a = 72.72; b = 148.15; c = 102.91.

For a unit cell volume of 1.11 x 10^6 $Å^3$ and assuming a molecular weight of 47,000 for the enzyme subunit and a single subunit/asymmetric unit, the volume occupied per Dalton is V_m = 2.95 $Å^3$ which compares well with the mean value of 2.4 $Å^3$ obtained by Matthews (1968) for 116 different protein crystals. Therefore, the two subunits of the dimer are related by one of the two sets of the crystallographic twofold axes.



Figure 3.1 Photomicrograph of sheep liver 6PGDH crystals at pH = 6.5

Figure 3.2 shows the h k 0 diffraction zone where the C face centring condition (spots with (h+k) odd are absent) is evident. The effect of the 2_1 screw axis parallel to c can be seen from the systematic absence l odd on the h 0 l and 0 k l pictures (Figure 3.3).

Crystals of the enzyme have been reported from Bacillus stearothermophilus (Pearse and Harris (1973)). These crystals have been presented to Dr. M.J.Adams by Dr. D.Blow for a parallel crystallographic study with the enzyme from sheep liver. These crystals are hexagonal plates of maximum radius 0.1 mm and space group $P3_12$. The structural work is not advanced.

A modified crystal form of the sheep liver 6PGDH crystals does exist, however, and occurs on spontaneous reversible transformation of the sheep liver crystals described above. The space group is still $C222_1$, but the cell dimensions (in Å) are now:

a = 71.9; b = 146.8; c = 100.6.

This effect was first noticed (November, 1974) on a h 0 l zone of a $K_2Pt(CN)_4$ derivative precession photograph where the intensity changes obtained were too excessive for just heavy atom substitution. Figure 3.4 compares the h 0 l zones of the crystal Type A and Type B (transformed) $K_2Pt(CN)_4$ derivative precession photographs. The transformation causes a rotational disordering of the crystals giving a mosaic spread of approximately 0.8° about c as axis and seems more prevalent with the platinum derivative than for the native, for which the transformation also occurs.

Transfer of Type A crystals from ammonium sulphate mother liquor into 2.5 M potassium tartrate at pH 7.2 or 1.8 M ammonium citrate at pH 6.5 reliably gives the Type B crystals (Archibald, personal communication). This is a very interesting phenomenon but a frustration in data collection when the transformation occurs spontaneously from Type A to Type B (see § 5). Data collected from the Type B crystals to 6 Å resolution enabled a difference map to be evaluated with respect to

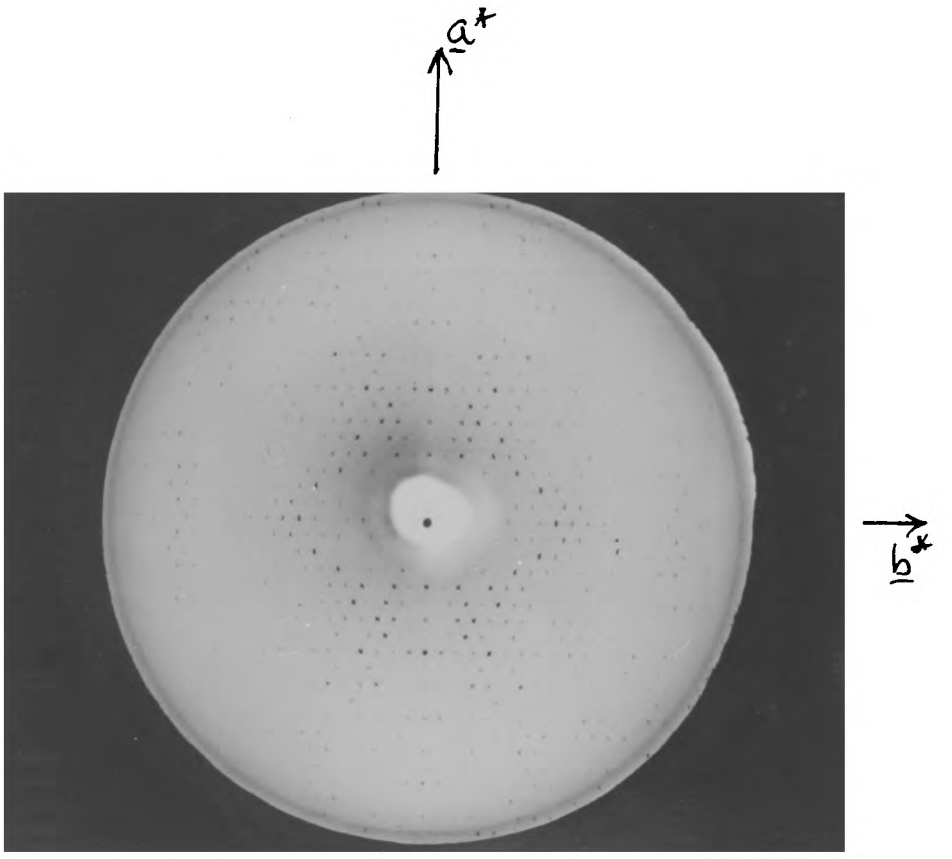


Figure 3.2 A 12° precession photograph of the **hk0** zone of the 6PGDH $C222_1$ crystals. The C face centring is evident since (h+k) odd reflections are absent (Taken by Susan Clift):

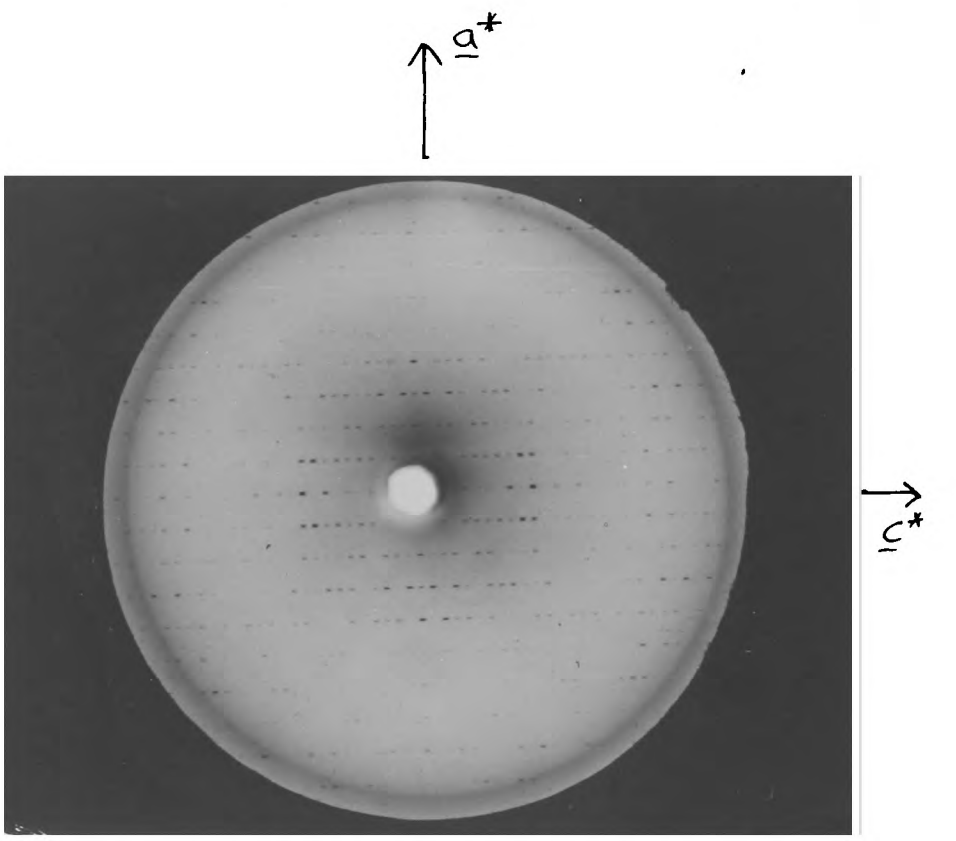


Figure 3.3a A 12° precession photograph of the h0l zone of the 6PGDH $C222_1$ crystals; the 2-fold screw axis is evident since 00l reflections with l odd are absent (Taken by Susan Clift).

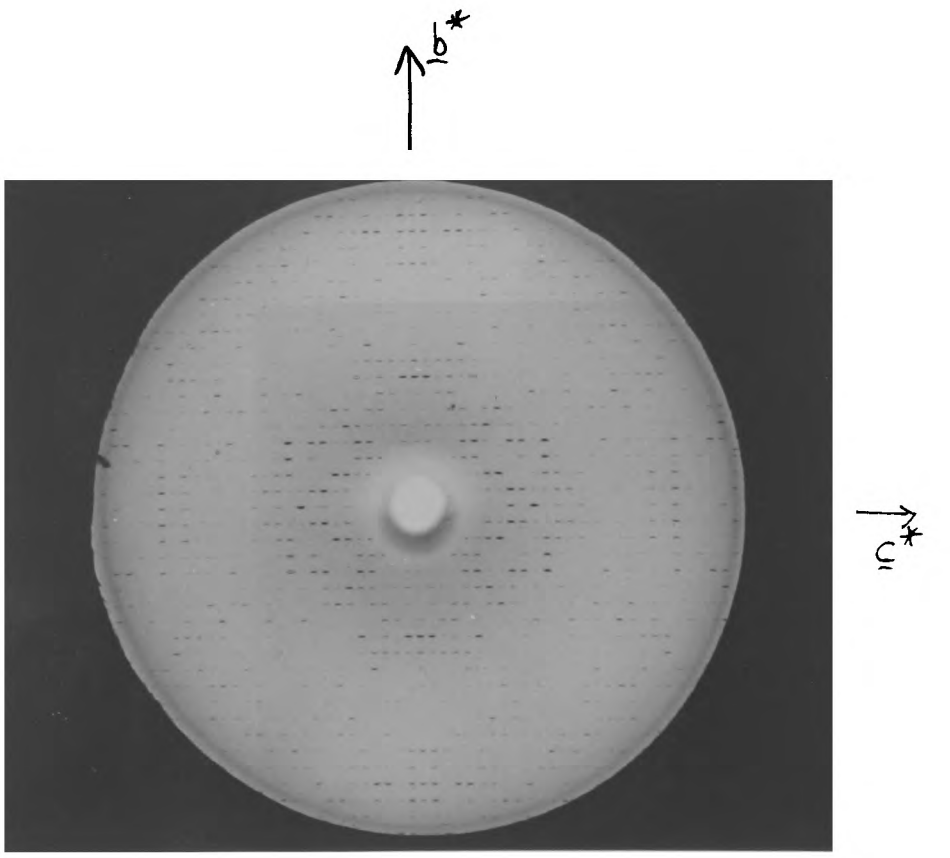
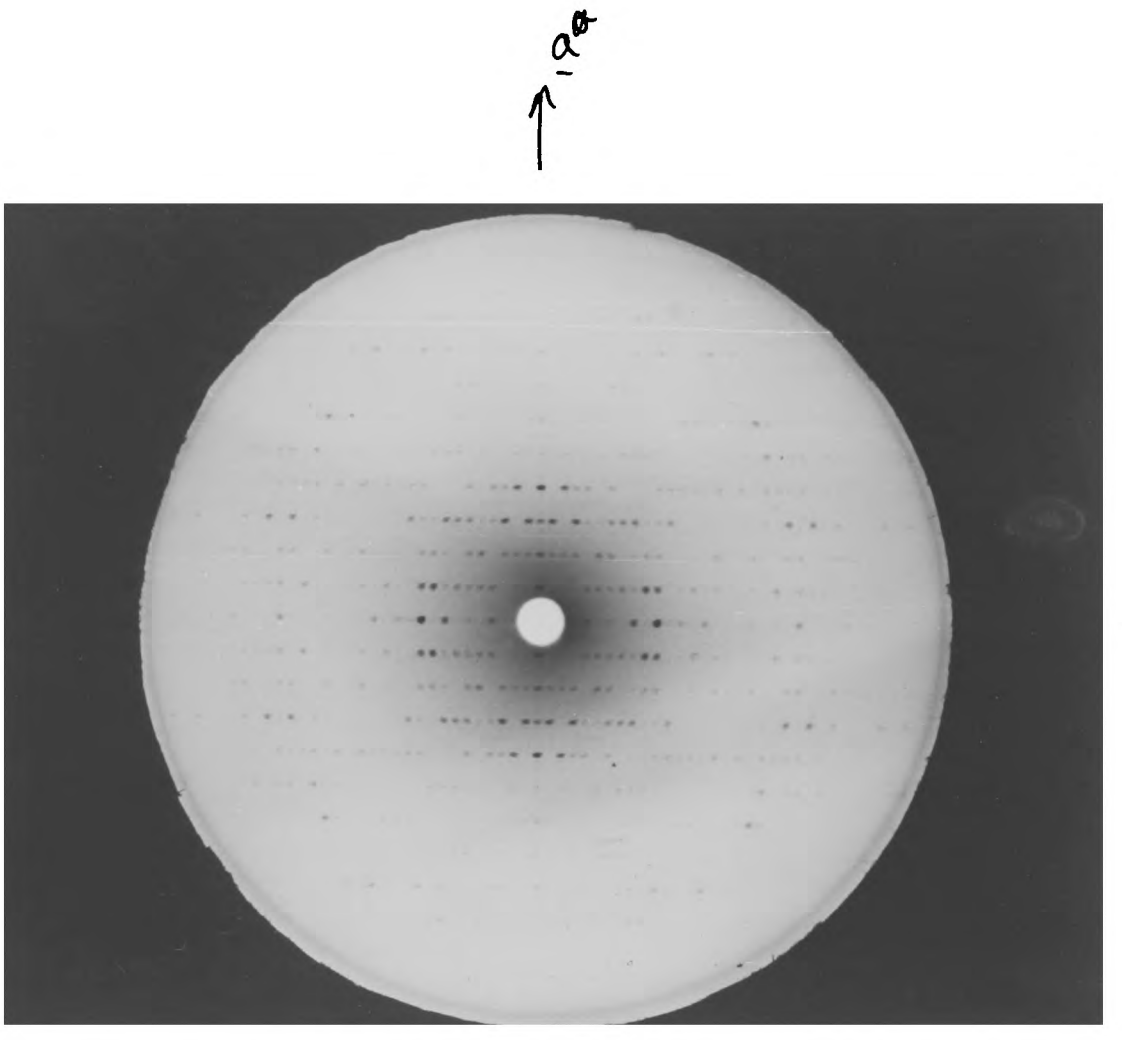
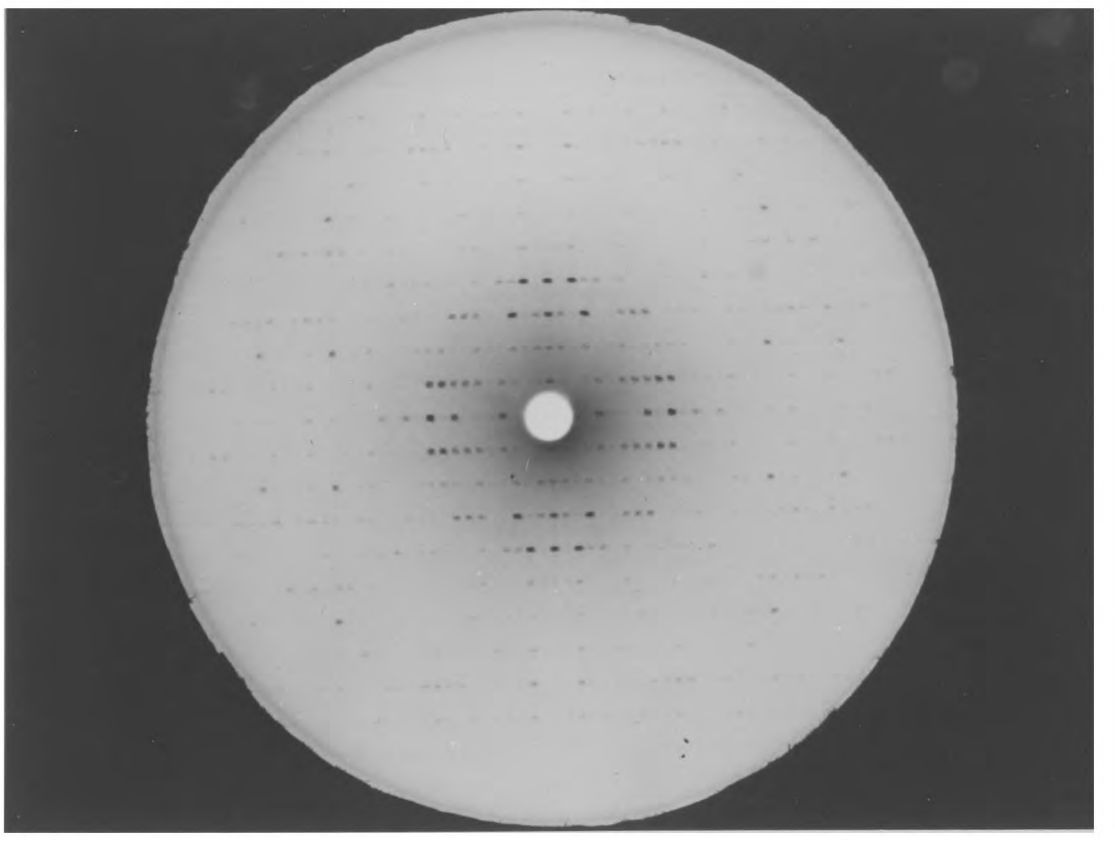


Figure 3.3b A 12° precession photograph of the 0kl zone of the 6PGDH $C222_1$ crystals (Taken by CEB).



(a)



(b)

Figure 3.4 16° precession photographs of the h0l zone $K_2Pt(CN)_4$ derivatives of crystal types (a) Type A, 2.7 mM, 24 hour soak; (b) Type B, 2.7 mM, 48 hr soak (taken by JRH). Attention is drawn to the different intensity distribution along the $\underline{a}^*$ axis, particularly the 600 and 601 reflections. Compare with Figures 3.5(a) and (b).

Type A, using the phases for Type A crystals (see § 3.6). The phenomenon seems to involve a rotation of the two monomers with respect to each other in the dimer about an axis parallel to c (Archibald, personal communication).

The crystal structure analysis presented in this thesis is entirely concerned with the Type A, sheep liver 6PGDH $C222_1$ crystals unless otherwise stated and these will be referred to as 6PGDH crystals. The work presented in this chapter has been published (Adams, Helliwell and Bugg (1977), see Appendix II).

Heavy atom derivatives

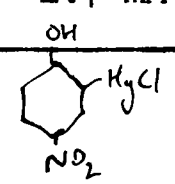
A preliminary survey of heavy atom derivatives was made by Dr. P. Engels and the survey extended by Dr.M.J.Adams, I.G.Archibald, Dr.C.E.Bugg and J.R.Helliwell, the results of which are summarized in Table 3.1. Heavy atom compounds were tried which had been successful with other proteins and involved the diffusion of salts of heavy atoms into the crystals. This search indicated that the $K_2Pt(CN)_4$ and $KAu(CN)_2$ derivatives showed the best intensity changes consistent with preserving the isomorphism of the crystals. The results with mercurials were disappointing. Silverberg et al. (1973) reported that there are seven thiol groups in the subunit of which two are especially reactive and essential for enzymatic activity. This seemed to imply that mercurial derivatives could be prepared. Crystals of the enzyme soaked in solutions of low concentration of p-hydroxy-mercuri benzoate (PHMB) showed very small changes in the diffraction pattern whereas larger concentrations caused cracking of the crystals followed by disintegration.

The soaking conditions of the $K_2Pt(CN)_4$ and $KAu(CN)_2$ derivatives were optimised by varying the concentration and soaking time. Precession photographs of the h 0 l zone were used to monitor the effect of these experiments. The crystals used for data collection were soaked in 22 μM $KAu(CN)_2$ for 16 hours for the Au derivative and 2.7 mM $K_2Pt(CN)_4$

Table 3.1 Derivative search

Heavy atom reagent	M.W.	Concentration	Soaking time	Intensity changes (zone)	Comments
PHMB(Na salt)	379.0	1.25 μM 6.67 μM 66.7 μM	1 day 1 day 1 day	small (hhl)	very fragile crystals crazed, no pattern } (MJA)
$Hg(OAc)_2$	318.7	2 μM	2 days	none (h0l)	crystal crazed, diffuse spots (PE)
2,4 mr -NP*	374	5 μM	2 days	very small (h0l)	poor quality photo (PE)
$NH_4(ReO_4)$	286.2	1 mM 10 mM	2 days 2 days	very small (h0l)	(MJA) crystals fragile and crazed (MJA)
$UO_2(NO_3)_2.6H_2O$	502.2	1mg/ml	3 days	none (0hl)	(PE)
$UO_2Ac_2.2H_2O$	424.2	1 mg/ml	2 days	none (h0l)	(PE)
$Pb(NO_3)_2$	330	1mg/ml	2 days	none (h0l)	(PE)
$KAu(CN)_2$	288.14	22 μM	1 day	good, reproducible (h0l)	(PE, CEB, MJA)
$K_2Pt(CN)_4.3H_2O$	431.41	2.7 mM	1 day	good, reproducible (h0l)	(PE, JRH)
$Th(NO_3)_4$	544.05	0.22 mM	1 day	small (h0l)	(IGA)
$Dy(Ac)_3.H_2O$	411.64	5 mM	1 day	v.poor diffraction	pH = 5.7 to improve solubility(JRH)
$KAu(CN)_2$ + $K_2Pt(CN)_4$**	288.14 431.41	22 μM 2.7 mM	48 hrs 24 hrs	changes attributable to both Pt and Au (h0l)	crystals v. fragile; several attempts failed before this (JRH)

* 2,4 mono-nitro phenol



** double derivative soak

for 24 hours for the Pt derivative. Under these conditions the intensity changes were reproducible. Figure 3.5 shows the h 0 l zones for the native, Pt, Au and PHMB (i.e. Hg) derivatives. Particular attention is drawn to reflections 401, 600 and 0020 for the Pt; 204, 206 for the Au and 204 for the Hg derivatives respectively. The increase in intensity of the 401 reflection for the Pt derivative was also observed for crystals containing bound coenzyme (see § 4).

3.3 Data collection and processing

Three dimensional sets of X-ray data were collected to 6 Å resolution for the native and the Au, Pt and Hg derivatives with a c axis mount using CuK_α radiation from a Phillips fine focus X-ray tube operated at 40 kV and 30 mA. Intensities were measured at room temperature on a Hilger and Watts Y290 4 circle diffractometer controlled by a Digital equipment PDP8 computer. Data were collected to 6 Å rather than higher resolution in order to check that the crystals and derivatives were adequate for a full structural analysis. Also the data beyond 6 Å were weaker. The 4 circle diffractometer was the most suitable instrument at low resolution for these crystals (see § 2.2).

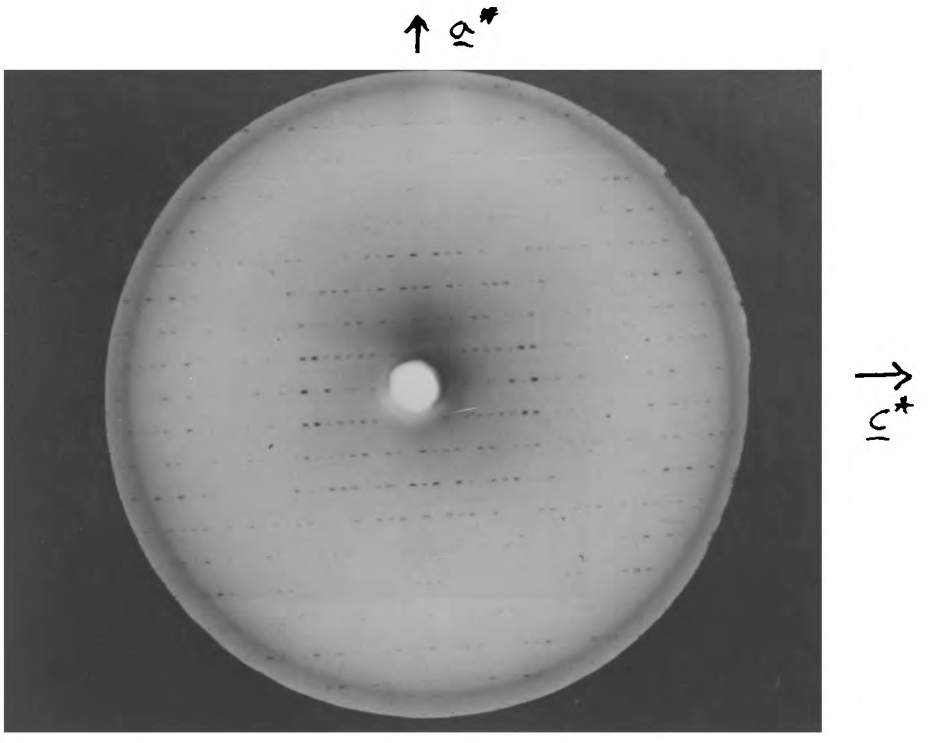
The data were collected in layers of constant h for reflections hkl, $hk\bar{l}$, $\bar{h}kl$ and $\bar{h}k\bar{l}$, the general absences with indices (h+k) odd were not monitored. Two Friedel pairs were thus measured for each heavy atom derivative and four equivalents for the native enzyme. With l being the fastest moving index the Friedel pairs hkl, $hk\bar{l}$ and $\bar{h}kl$, $\bar{h}k\bar{l}$ were collected within 30 minutes of each other. This minimised any differences due to errors in correcting for radiation damage, thus helping to obtain an accurate estimate of the anomalous difference.

The crystals were cut from large needles and were typically of volume 0.6 x 0.25 x 0.6 mm^3. Intensities were measured by ordinate analysis using 50 steps of 0.02° in ω and counting for 1 second each step. The

Figure 3.5 12° precession photographs of the h0l zones for (a) the native (N_V); (b) the platinum derivative (2.7 mM, 16 hrs); (c) the gold derivative (20μM, 24 hrs); (d) the mercury derivative (1.5 μM, 24 hrs). Attention is drawn to reflections 401, 600 and 0020 for the Pt; 204, 206 for the Au and 204 for the Hg. The cell dimensions ($|\underline{a}|$ and $|\underline{c}|$ estimated from these photographs show them to be closely isomorphous crystals.

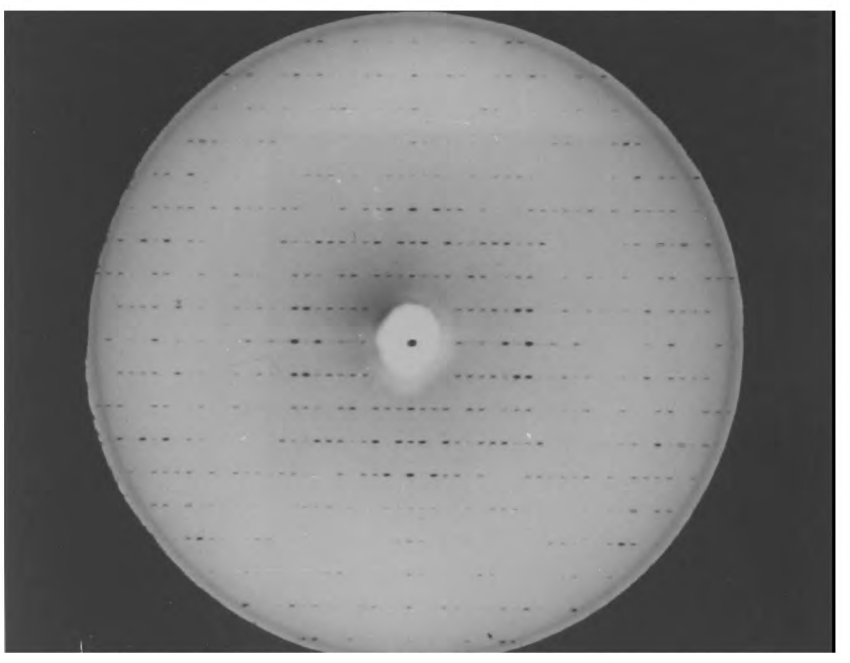
3.5(a) N_V

(Taken by Susan Clift)



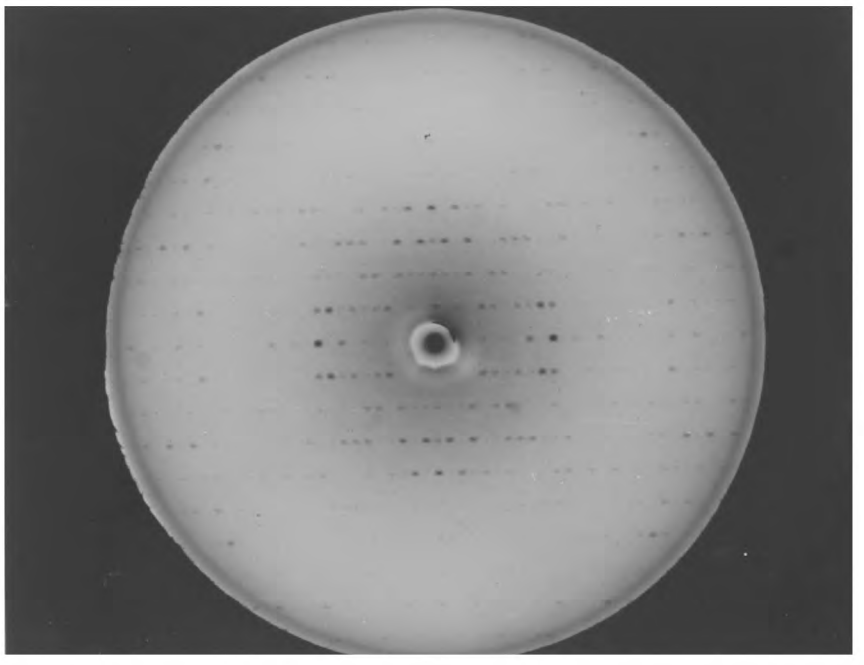
3.5(b) Pt

(PE)



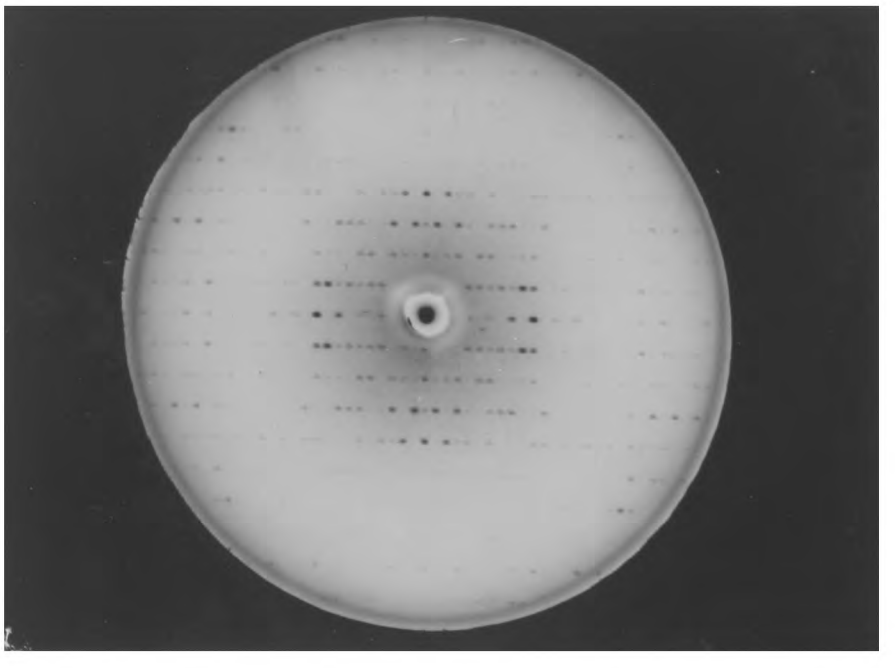
3.5(c) Au

(MJA)



3.5(d) Hg

(MJA)



whole of the 6 Å data set, 6000 measurements, 1500 independent terms, could then be obtained from a single crystal in 5 days. The average radiation damage estimated from the intensities of three standard reflections in orthogonal directions of reciprocal space (the 0010, 0160, and 800 reflections) collected every 50 reflections, was 14.5% in intensity. Intensities for the heavy atom derivatives showed no more decay than those for the native protein. Absorption curves were measured on the 0010 reflection in 10° intervals of φ both before and after data collection to detect any change in the absorption of the crystal during data collection, perhaps due to movement of mother liquor around the crystal.

The absorption curves for the $00\bar{1}0$ reflection were also measured to detect any variation, owing perhaps to the miscentring of the crystal in the X-ray beam.

The raw data were corrected for:

(a) LP factors (for normal beam geometry with $\omega = \theta$ and the χ circle bisecting the diffracted and main beams);

(b) radiation damage using a correction inversely proportional to $\sum_{\text{standards}} I$;

(c) absorption effects using the method of North et al.(1968).

The parameters relevant to the data collection are summarized in Table 3.2.

Before merging, 70% of the intensities, I, were greater than $4\sigma(I)$ and 77% greater than $3\sigma(I)$. These are representative figures for each data set (native, Pt, Au and Hg).

The data from each crystal were separately merged using the computer program SSM (see Appendix IV).

Derivative to native scaling was performed by evaluating:

$$k.\frac{\sum |\underline{F}_{PH}|^2}{\sum |\underline{F}_{P}|^2} = C$$

Table 3.2 Parameters relevant to data collection in the 6 Å study of 6PGDH crystals

Crystal description	Soaking conditions conc.(mM)	time(h)	Absorption I_{max}/I_{min}	Radiation damage (<>% decrease in intensity of standards)	R_{symm}	Mean fractional isomorphous change	Overall k_{emp} (from ANSC)
native	—	—	2.83	17.4	6.54	—	—
$K_2Pt(CN)_4 \cdot 3H_2O$	2.7	24	1.99	14.7	5.40	8.34	2.49
$KAu(CN)_2$	0.022	16	2.08	10.7	4.52	8.30	2.95
PHMB	0.001	22	2.25	15.3	5.53	6.10	1.94

Note

I_{max} = maximum transmitted intensity for the axial reflection 0 0 10 (φ=90°); I_{min} = minimum transmitted intensity for the same reflection (φ= 0°).

$$R_{symm} = \sum_{nref} \left(\sum_{i=1}^{n} (I_i - \bar{I}) \right) / \sum_{i=1}^{n} \bar{I}$$

where $\bar{I}$ = average intensity for the n equivalent indices. For the native data, Friedel opposites are treated as equivalent (n=4) and for the heavy atom derivative, treated separately (n=2).

Overall $k_{emp} = 2\langle \Delta_{ISO} \rangle / \langle | \Delta_{ANOM} | \rangle$

Mean fractional isomorphous change = $$\frac{\sum \| |F_{PH}| - |F_P| \|}{\sum |F_P|}$$

where C was taken as unity.

Using the programme ANSC (Appendix IV) it was found that K was independent of sin θ for each data set, presumably because at this resolution the differences in radiation damage and ordering of the crystals were slight. ANSC outputs the value of K as well as the ratio of the isomorphous to the anomalous component of the scattering (the empirical k of Dodson and Vijayan (1971)) and the mean fractional change in intensity for the derivative (see Table 3.2). Figure 3.6 shows k_{emp} versus $4\sin^2\theta/\lambda^2$ for the Pt, Au and Hg derivatives; k_{theor} is included for comparison.

The scale K can also be calculated using the program FHLEKR (Appendix IV), which can calculate FHLE's using a k_{emp}, by iteration of the equality:

$$K_K \; |\underline{F}_{PH}|^2 = |\underline{F}_P|^2 + \text{FHLE}^2$$

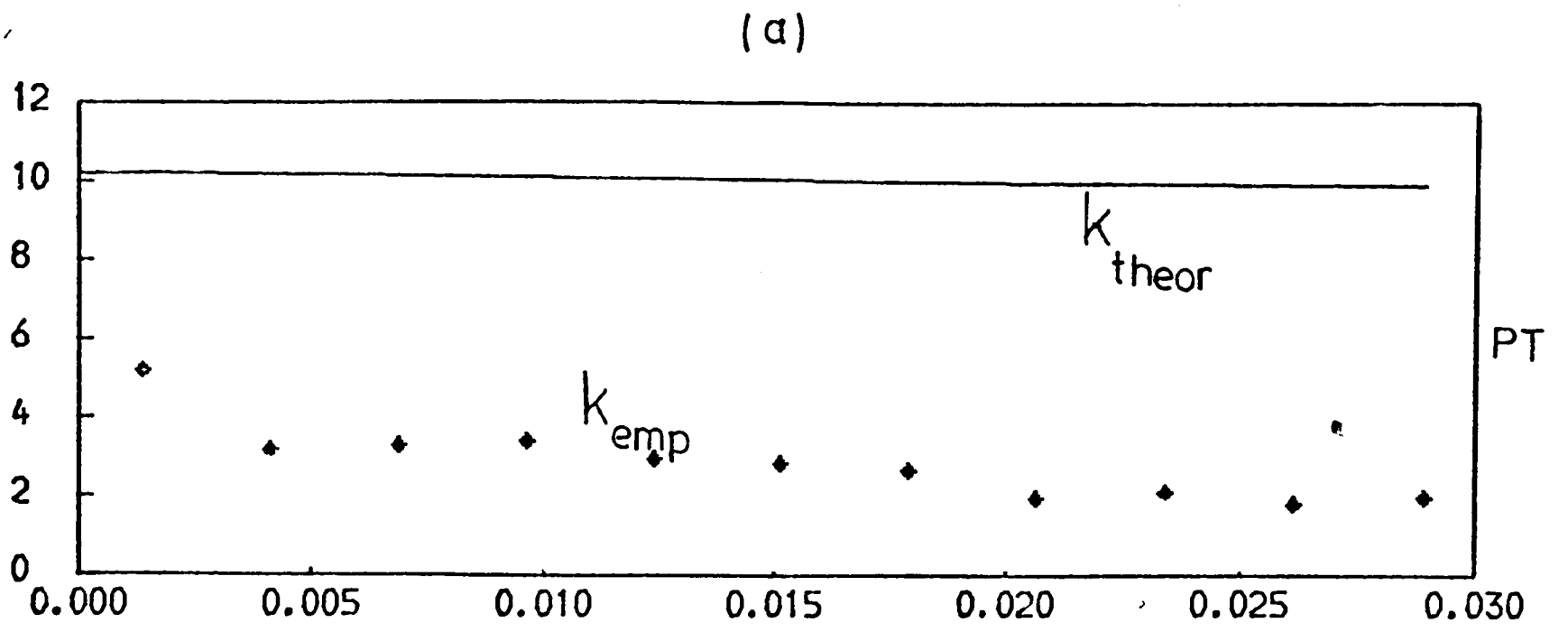
where the subscript K is used to denote Kraut scale. The iteration is needed since K_K is involved in the calculation of FHLE. The program also outputs an analysis of the data, some details of which are given in Table 3.3. This Table shows that the ratio of $<\Delta_{ISO}>/<\sigma(\Delta_{ISO})>$ decreases to $\lesssim$ 1 at the higher angles of the data sets. In retrospect this provides support for collecting the data in shells, starting with the outer shell first before radiation damage is too significant. Perhaps even spending more crystal lifetime on this range than on the low angle data.

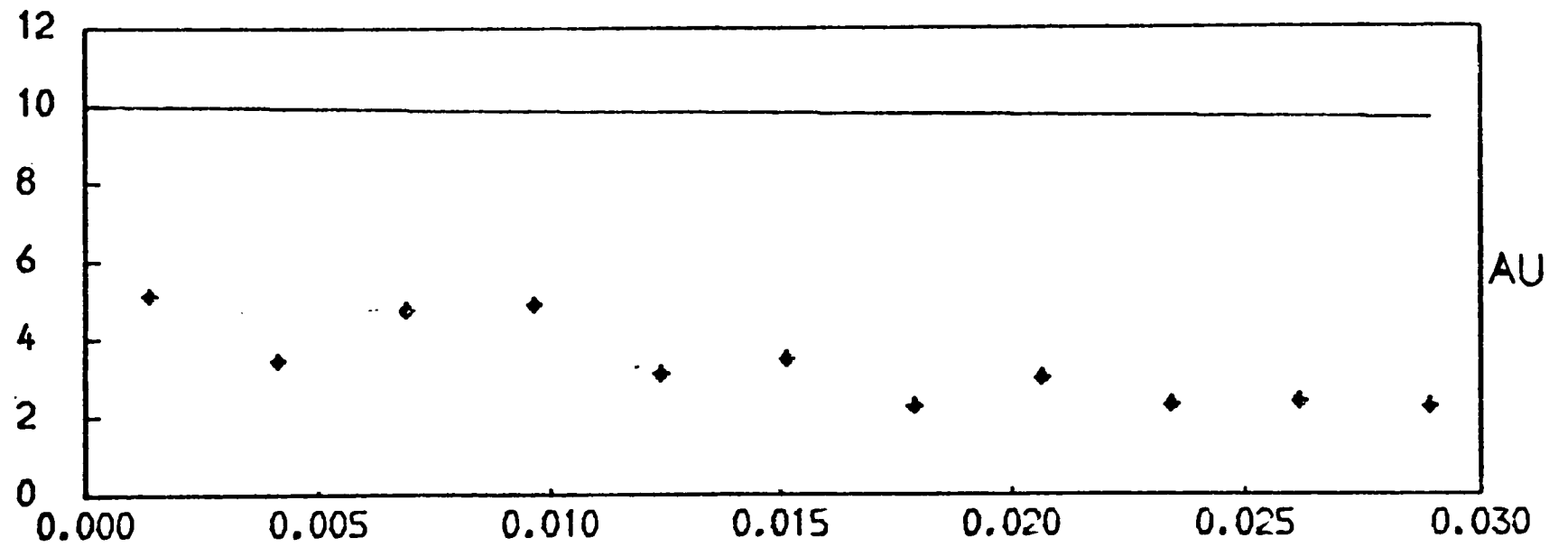
3.4 Location of the heavy atom positions

In the absence of any starting phase information the methods available for solving the heavy atom structure involve the calculation of isomorphous difference Pattersons, FHLE^2 Pattersons and direct methods (see discussion in § 2.4).

Table 3.4 gives the equivalent positions of the space group $C222_1$

Figure 3.6 Plot showing the variation of k_{emp} with $4\sin^2\theta/\lambda^2$ for (a) Pt, (b) Au, (c) Hg derivatives. The value $k_{theoretical} = (f + f')/f''$ is included for comparison.





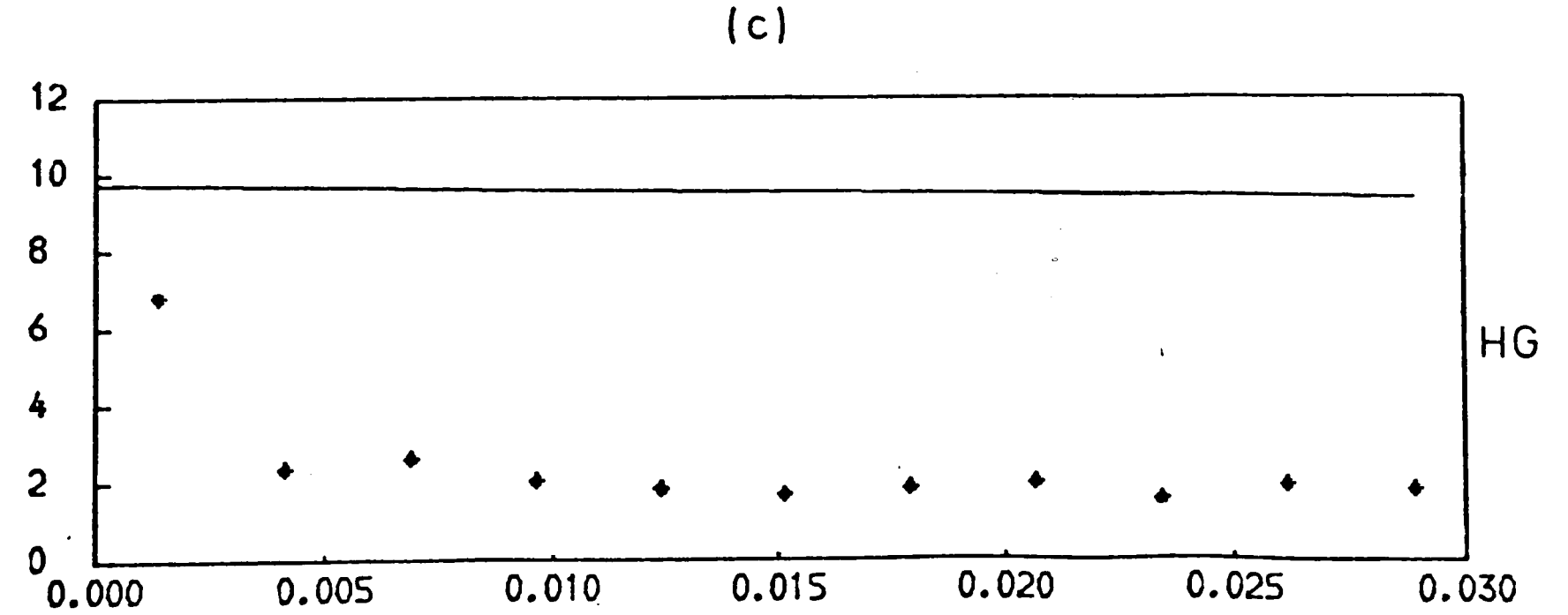


Table 3.3 (a) Analysis of the native 6PGDH and Au and Pt derivative data to 6 Å resolution in ranges of $4\sin^2\theta/\lambda^2$ for the acentric and centric reflections.

$4\sin^2\theta/\lambda^2$	Number	$\langle\lvert F_P\rvert\rangle$	σ	$\langle\lvert F_{PH}\rvert\rangle$	σ	$\langle\Delta_{ano}\rangle$		$\langle\Delta_{iso}\rangle$		σ $\langle F_{HLE}\rangle$		σ $\langle F_{HLEB}\rangle$		$\langle\lvert F_{PH}\rvert\rangle$	σ	$\langle\Delta_{ano}\rangle$		$\langle\Delta_{iso}\rangle$		σ $\langle F_{HLE}\rangle$		σ $\langle F_{HLEB}\rangle$	
						Au derivative										Pt derivative							
0.0-0.005	48(c) 67(a)	308	9	320	8	- 12	- 11	44 24	17 9	44 42	17 33	44 54		321	9	- 13	- 12	35 27	18 10	35* 44	18 31	36* 60	
0.005-0.01	60(c) 158(a)	399	10	399	8	- 11	- 14	34 24	19 11	34 37	19 54	34 48		400	9	- 14	- 16	36 21	19 12	36* 39	19 44	37* 52	
0.01-0.015	65(c) 205(a)	369	12	369	10	- 14	- 16	30 24	24 13	30 42	24 68	30 32		371	12	- 16	- 21	36 23	26 15	36 40	26 48	36 56	
0.015-0.02	58(c) 263(a)	281	15	279	13	- 15	- 22	31 21	31 18	31 40	31 72	31 44		285	15	- 16	- 25	29 24	33 19	29 40	33 70	29 47	
0.02-0.025	53(c) 282(a)	247	18	245	15	- 17	- 26	37 23	35 22	37 42	35 83	37 41		249	18	- 20	- 31	31 21	39 23	31 39	39 89	31 44	
0.025-0.03	35(c) 183(a)	253	18	255	16	- 18	- 27	37 21	34 22	37 39	34 85	37 41		254	19	- 23	- 33	34 23	38 24	34 40	38 72	35 48	

c = centric; a = acentric.

* These values are slightly different because different runs of the program FHLEKR were required to obtain the F_{HLE}, k_{emp} and the F_{HLEB}, k_{theor} analysis. These mean values quoted involved slightly different numbers of reflections.

Table 3.3b Comparison of the F_{HLE}, F_{HLEB} estimates made for the heavy atom structure scattering amplitude $|\underline{F}_H|$ using a k_{theor} and a k_{emp}. The centric and acentric estimates should be approximately equal (achieved most effectively with an F_{HLE}, k_{emp}).

A		$\langle F_{HLE} \rangle$ k_{emp}	$\langle F_{HLEB} \rangle$ k_{emp}	$\langle F_{HLE} \rangle$ k_{theor}	$\langle F_{HLEB} \rangle$ k_{theor}
Au	c	37	37	37	37
	a	40	29	73	45
Pt	c	34	34	34	34
	a	40	28	85	50

Note: The Kraut scales used were:

	F_{HLE} k_{emp}	F_{HLEB} k_{theor}
Au	0.8762	0.8710
Pt	1.03541	1.03189

Table 3.4

a) Equivalent positions in the space group $C222_1$

$(0, 0, 0)$; $(\frac{1}{2}, \frac{1}{2}, 0)$

X, Y, Z

$X, \bar{Y}, \bar{Z}$

$\bar{X}, \bar{Y}, \frac{1}{2} + Z$

$\bar{X}, Y, \frac{1}{2} - Z$

b) Equivalent positions in the Patterson space group C_{mmm}

$(0, 0, 0)$; $(\frac{1}{2}, \frac{1}{2}, 0)$

X, Y, Z

$\bar{X}, \bar{Y}, Z$

$X, \bar{Y}, \bar{Z}$

$\bar{X}, Y, \bar{Z}$

$\bar{X}, \bar{Y}, \overline{Z}$

$X, Y, \bar{Z}$

$\bar{X}, Y, Z$

$X, \bar{Y}, Z$

c) Harker vectors in the Patterson space group C_{mmm}

$0, 0, 0$

$0, 2Y, 2Z$

$2X, 2Y, \frac{1}{2}$

$2X, 0, \frac{1}{2}-2Z$

Note: The space group of the Patterson can be derived from that of the real structure by the addition of a centre of symmetry and loss of the translational symmetry elements along screw axes (or glide planes, this latter not being relevant to protein crystallography).

(number 20 in International Tables) and the equivalent positions and Harker vectors in the Patterson space group C_{mmm} (number 65) respectively.

(i) Au derivative

The interpretation of the isomorphous difference, Δ_{ISO}^2, Patterson for the Au derivative was straightforward and the major site of substitution was readily determined. Figure 3.7 shows the Harker sections and selected general sections for the Au derivative Δ_{ISO}^2 Patterson with the Harker and cross vectors marked corresponding to the refined positions. Figure 3.8 shows the Harker sections for the F_{HLE}^2 Patterson synthesis. The peaks on these sections are less well pronounced than those on the isomorphous difference Patterson but the positions are similar.

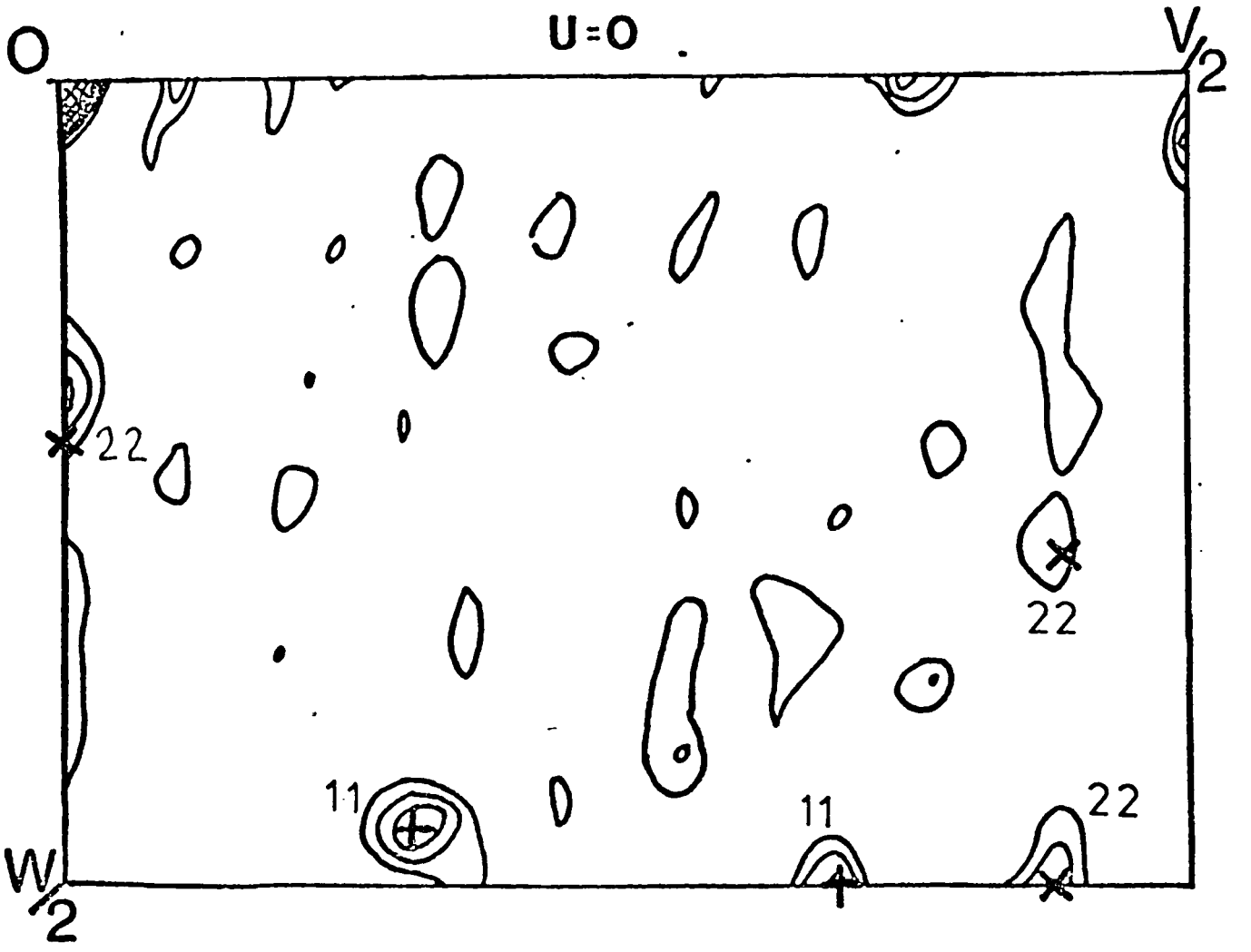
The site of major substitution for this derivative was confirmed by using the direct methods program MULTAN.

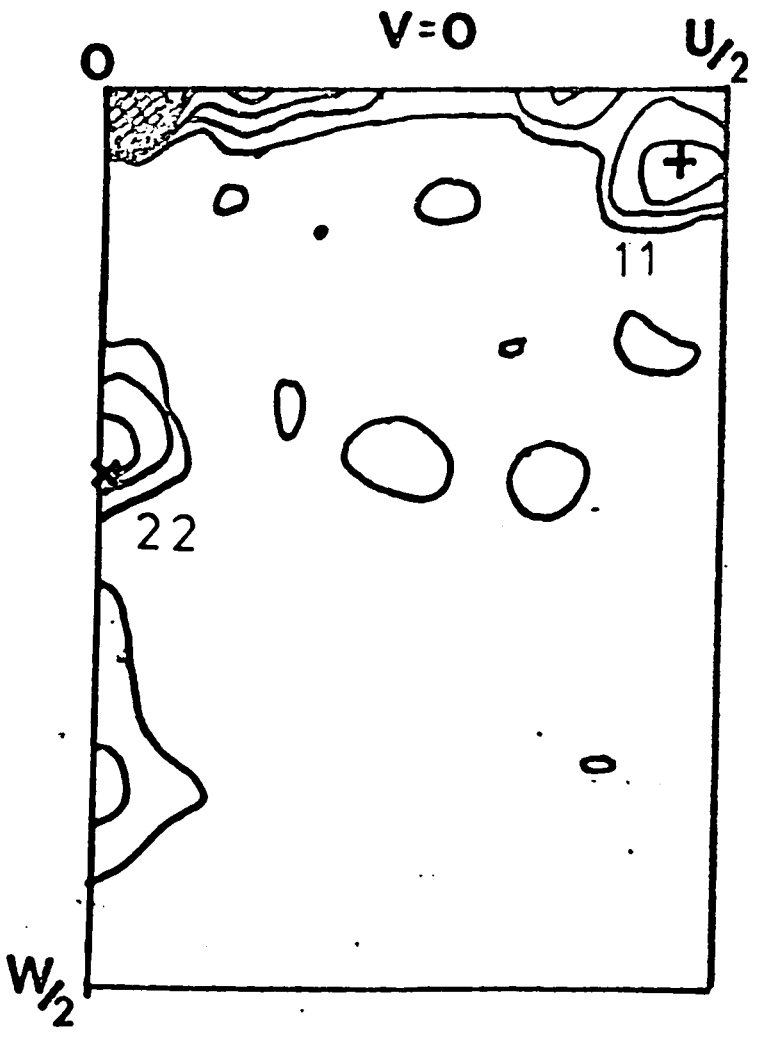
The isomorphous differences to 6 Å resolution were normalized and the statistical distribution of the resultant E_h (Table 3.5) corresponded to a centrosymmetric structure as expected since ΔISO involves a projection of the vector $\underline{F}_{PH}$ onto $\underline{F}_P$. The two hundred largest E_h down to 1.42 were chosen for input to MULTAN together with the fifty smallest terms for the PSIZERO reliability test (see § 2.4). The two thousand triple product relations with the highest associated reliabilities were used for phase determination. The starting phase set included two reflections to fix the origin and enantiomorph, two general reflections whose phase was varied and ten reflections having a probability of greater than 0.9 of being correct (Table 3.6). The figures of merit for the sixteen phase sets generated are shown in Table 3.7. Phase set 10 yielded two heavy atom sites, the Harker and cross vectors for these sites agreed with the difference Patterson. The major site had a peak height of 416 units, the second site 203 units and the next highest 95 units (Table 3.8). Figure 3.9 shows sections from EMAP10

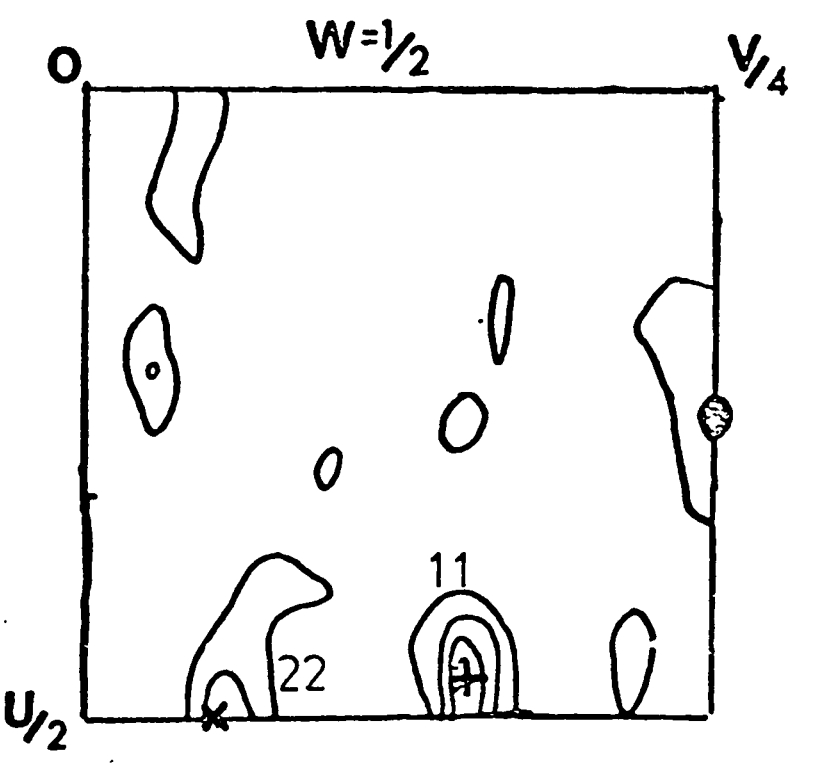
Figure 3.7

The three Harker sections and selected general sections for the Au derivative isomorphous difference Pattersons. The Harker and cross vectors corresponding to the final refined positions of Au1 and Au2 are marked. The origin peak is 1000, and contours start at 25 and the contour interval is 25.

Figure 3.7



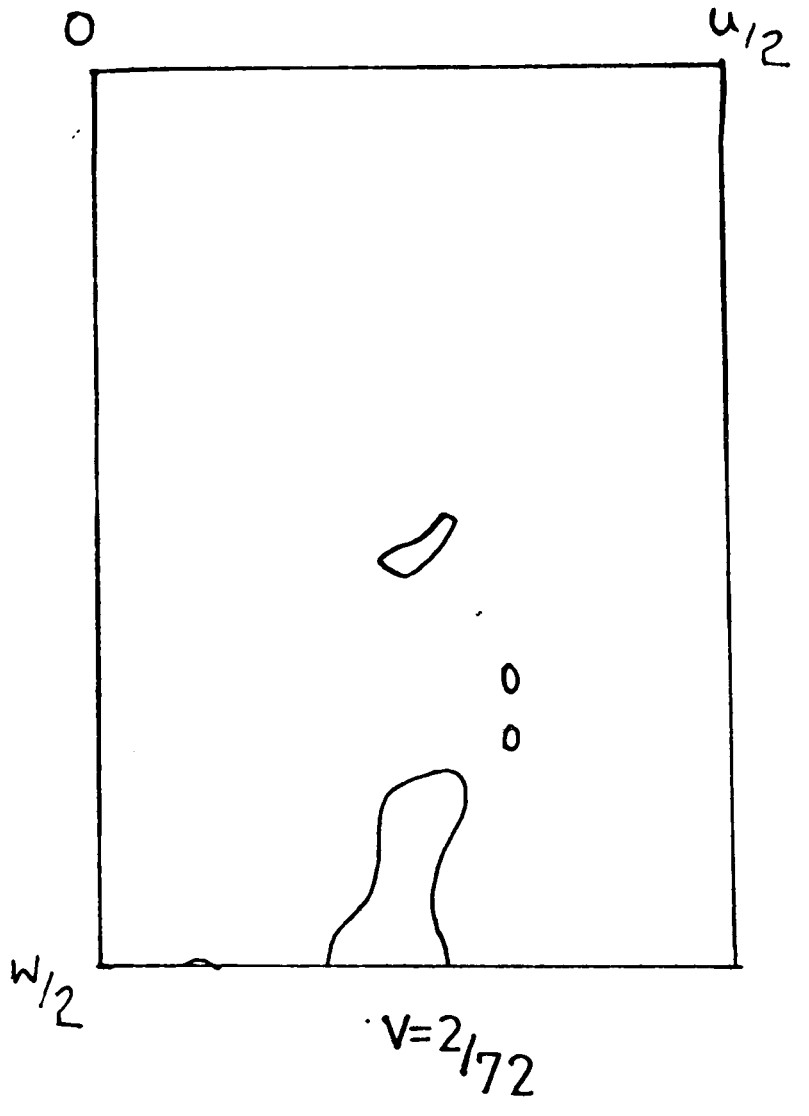


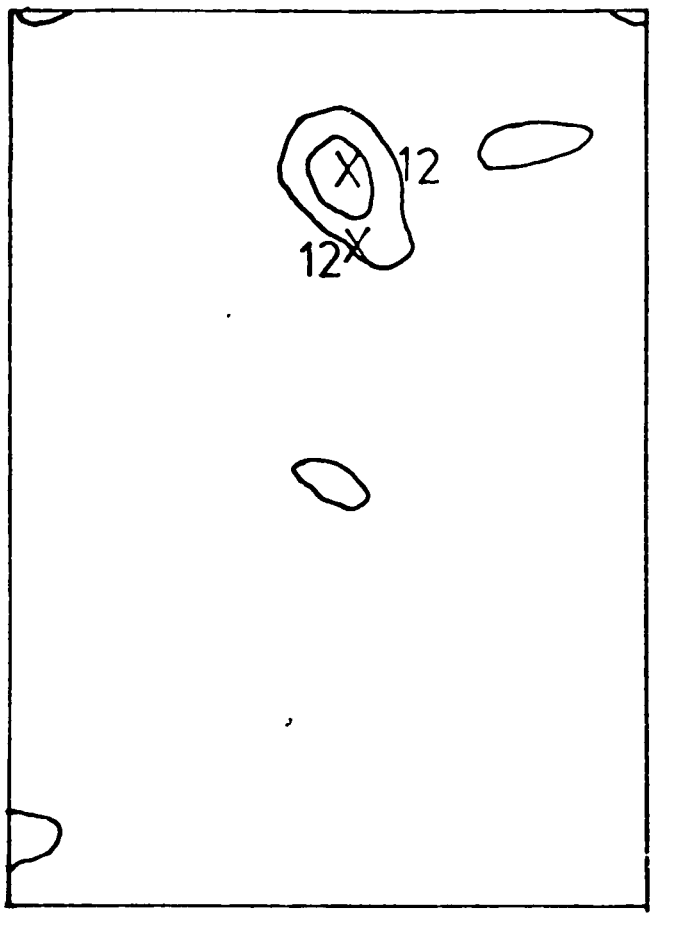


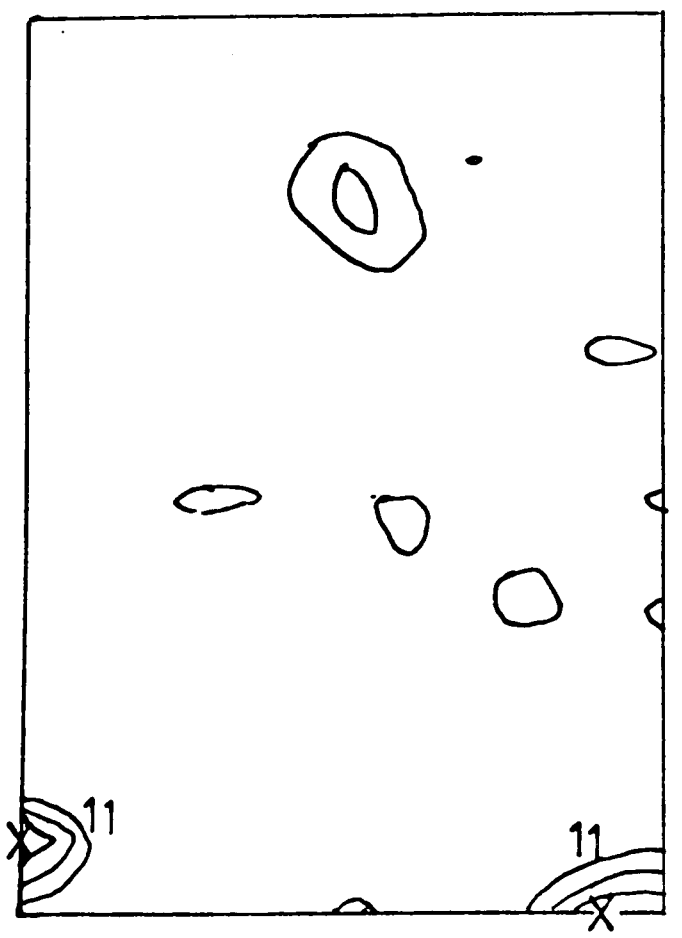
+ 1-1

× 2-2

Figure 3.7







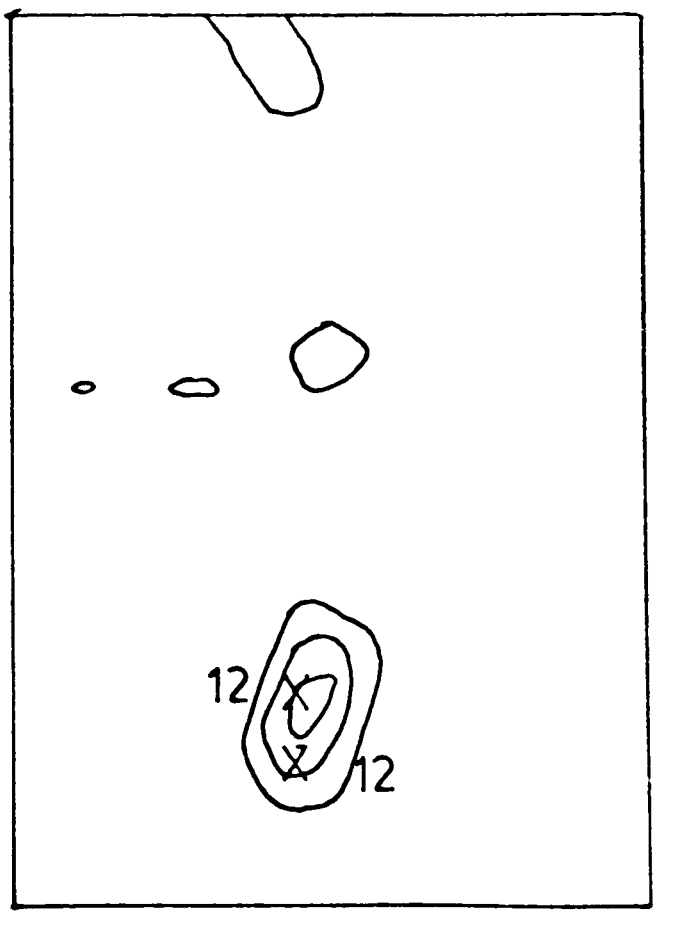
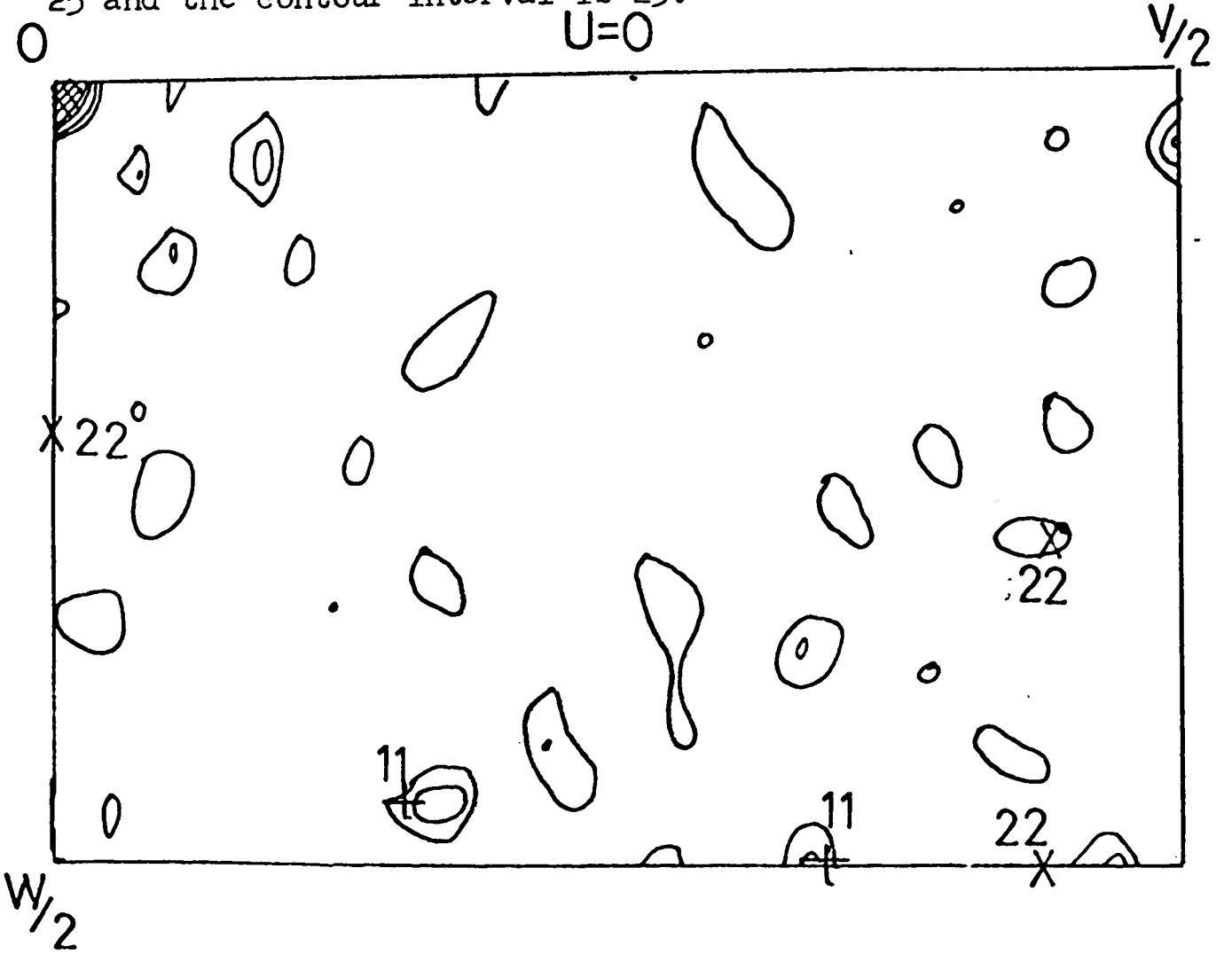
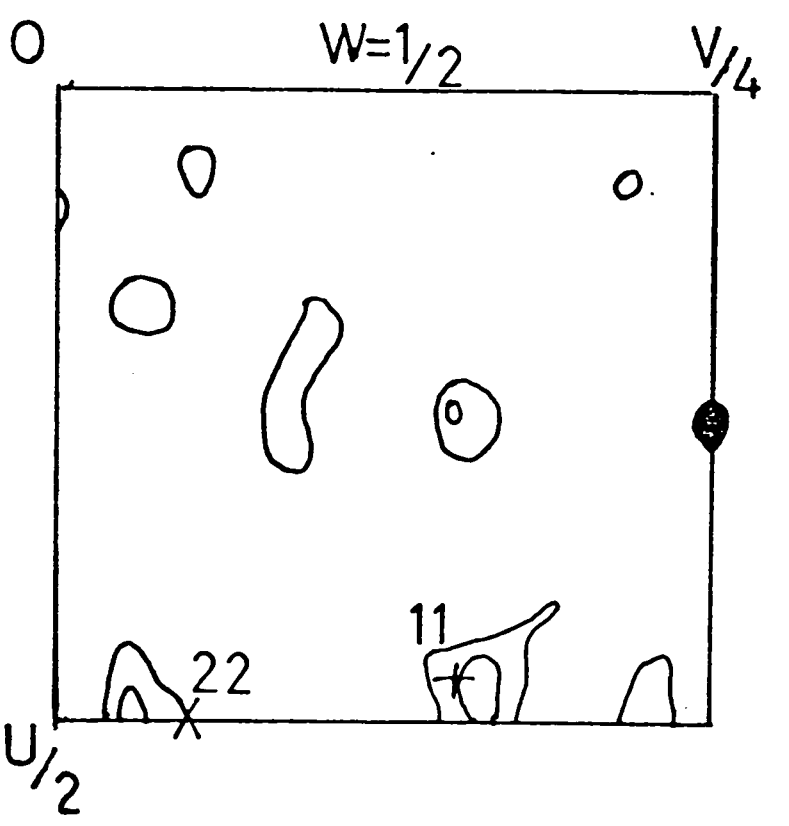


Figure 3.8 The three Harker sections for the Au derivative F_{HLE}^2

Patterson synthesis. The origin peak is 1000, and contours start at 25 and the contour interval is 25.







+ 1-1

X 2-2

Table 3.5 Statistical distribution of the $E_{\underline{h}}$ values calculated from the isomorphous differences for the 6PGDH Au derivative to 6 Å.

	Experimental	Theoretical Centro-symmetric	Theoretical non-centro-symmetric
$\langle E_{\underline{h}}^2 \rangle$	1.0	1.0	1.0
$\langle E_{\underline{h}}^2 - 1 \rangle$	1.0161	0.9680	0.7360
$\langle E_{\underline{h}} \rangle$	0.7725	0.7980	0.8860

Table 3.6 The starting set selected by MULTAN for the Au derivative phase generation to 6 Å with 200 $E_{\underline{h}}$ as the basic set.

Reflection type	Indices	$\lvert E_{\underline{h}} \rvert$	$\lvert F_{PH} \rvert$	Permitted phase	Initial weight
Origin fixing	0 14 1	3.73	105	0	1.0
	1 19 5	2.78	165	45	1.0
Enantiomorph	1 19 5	2.78	165	45	1.0
Σ_1	2 22 0	3.72	141	0	0.99
	0 0 10	2.64	1494	180	1.0
	4 0 0	2.60	464	0	0.99
	0 0 16	2.43	154	0	0.93
	0 0 8	2.22	692	0	0.99
	0 0 14	1.97	534	180	0.80
	8 0 0	1.85	1154	0	0.99
	0 20 6	1.63	198	0	0.98
	8 0 10	1.54	500	180	0.84
	4 0 14	1.54	272	180	0.97
General	2 6 6	2.40	276	$\pm 45^\circ$, $\pm 135^\circ$	1.0
	2 8 5	2.83	518	$\pm 45^\circ$, $\pm 135^\circ$	1.0

Table 3.7 The figures of merit for the sixteen phase sets generated by MULTAN for the 6PGDH Au derivative with 200 E_h at 6 Å. EMAP 10 yielded the correct heavy atom structure. The relative weights for the combined figure of merit were 1.0 in each case. ABSFOM, PSIZERO, RESID and COMBINED FOM are defined in § 2.

NUMSET	ABSFOM	PSIZERO	RESID	COMBINED FOM
1	0.5137	0.7516E 04	41.86	0.8670
2	0.4173	0.6216E 04	38.69	1.3044
3	0.4323	0.6027E 04	39.14	1.3382
4	0.4806	0.7046E 04	39.66	1.1509
5	0.7301	0.9276E 04	37.34	1.5198
6	0.7302	0.9278E 04	37.33	1.5203
7	0.7317	0.9275E 04	36.61	1.6083
8	0.7311	0.9276E 04	36.73	1.5925
9	0.7277	0.9247E 04	37.52	1.4991
10	0.6465	0.7049E 04	33.22	2.3294
11	0.4203	0.5311E 04	36.52	1.7874
12	0.4724	0.5746E 04	36.76	1.7876
13	0.3489	0.5242E 04	40.16	1.1968
14	0.4208	0.6008E 04	38.48	1.3892
15	0.4249	0.5890E_n04	37.07	1.6667
16	0.4237	0.5673E_n04	38.48	1.4799

Figure 3.9 Selected sections of EMAP10 for the Au derivative with the positions corresponding to the final refined positions of Au1 and Au2 marked. (a) $y = 6/72\ b$ (b) $y = 8/72\ b$ (c) $y = 17/72\ b$

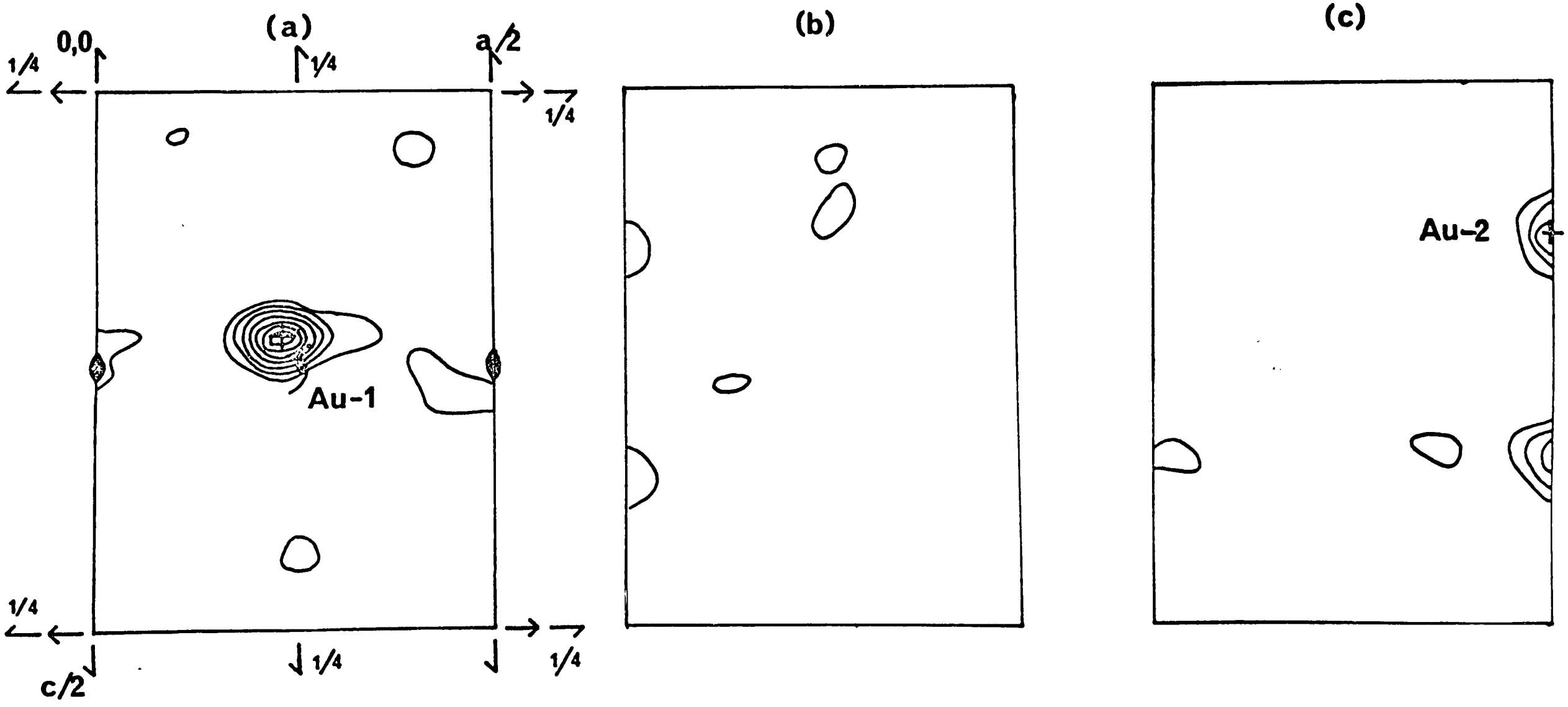


Table 3.8 Absolute and relative peak heights for the largest peaks obtained from EMAP 10 derived from MULTAN for the Au derivative. The refined occupancy is given for comparison.

Peak label	Peak height	Relative peak wrt highest peak	Refined occupancy	Refined occupancy wrt site 1
Au 1	416	1.0	0.7550	1.0
Au 2	203	0.4880	0.3770	0.4993
next highest	95	0.2284	-	-

with the final refined positions marked. Both the occupancies and positions agree very well with the final refined parameters. Though phase set 10 had the highest combined figure of merit, it did not have the highest ABSFOM (possessed by phase set 7). EMAP7 was calculated and showed a major site related to that on EMAP10 by a shift in x and y coordinates by one quarter of the unit cell (not equivalent to an origin shift).

(ii) Pt derivative

The Harker sections of the difference Patterson of the Pt derivative were less readily interpretable. This map (Figure 3.10) showed a larger number of peaks compared with the corresponding map for the Au and the Harker sections for the Pt $FHLE^2$ Patterson contained only weak peaks (Figure 3.11).

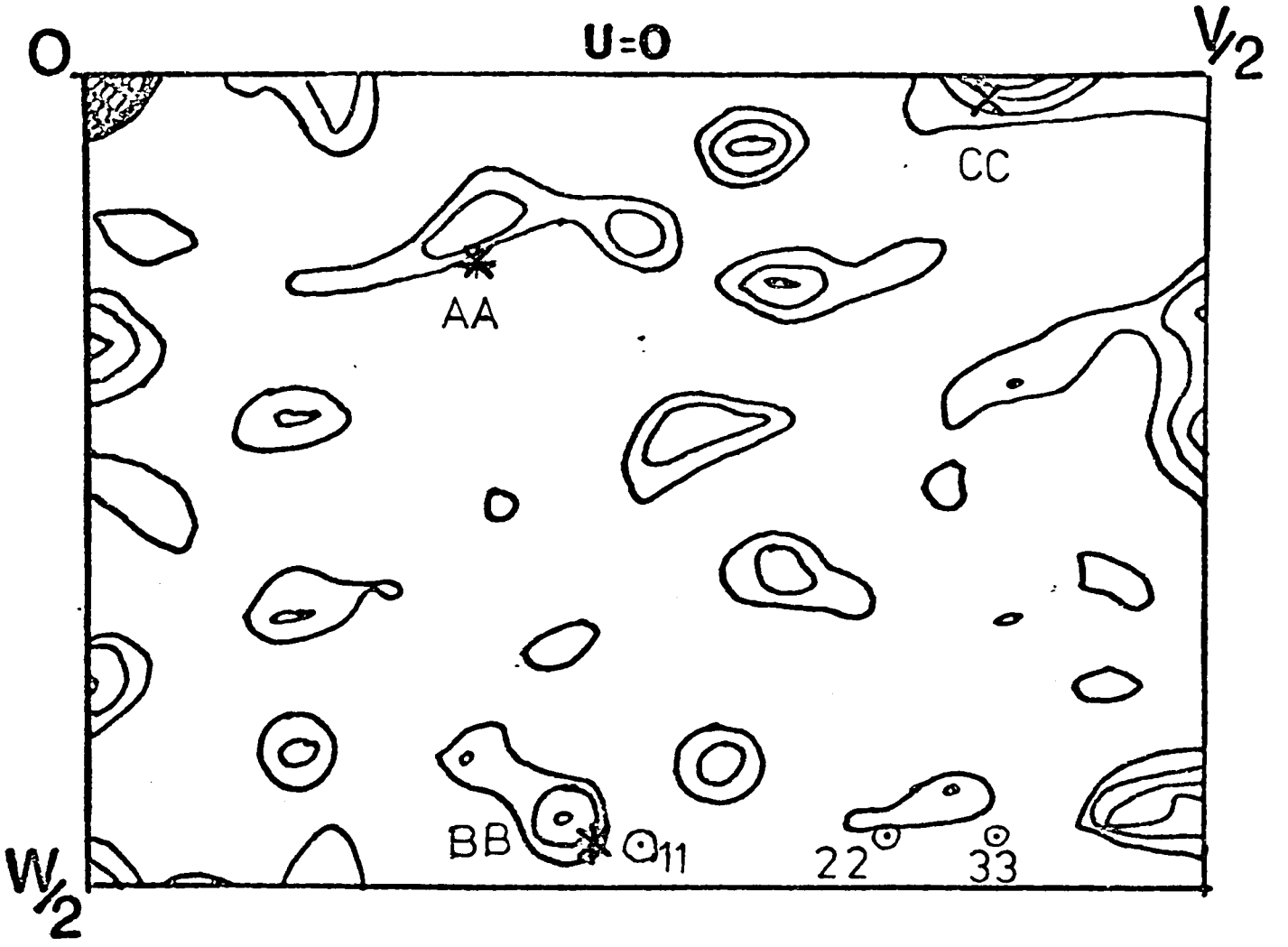
It was hoped that use of MULTAN would yield the heavy atom structure since no headway was made interpreting the Pt Patterson syntheses mentioned above. A similar procedure was followed as for the Au. Two hundred of the largest $E_{\underline{h}}$'s down to 1.44 were chosen for input to MULTAN together with the 50 smallest terms for the PSIZERO test. Table 3.9 shows the reflections used in the starting phase set. One of the general reflections was also used to fix the enantiomorph and hence could only assume two, instead of four possible values. Hence, eight phase sets were generated, the figures of merit of which are shown in Table 3.10. An EMAP corresponding to phase set 1 (highest ABSFOM and combined FOM) was calculated yielding three large peaks (Table 3.11). The Harker vectors corresponding to these three positions showed poor agreement with the difference Patterson sections (see Figure 3.10).

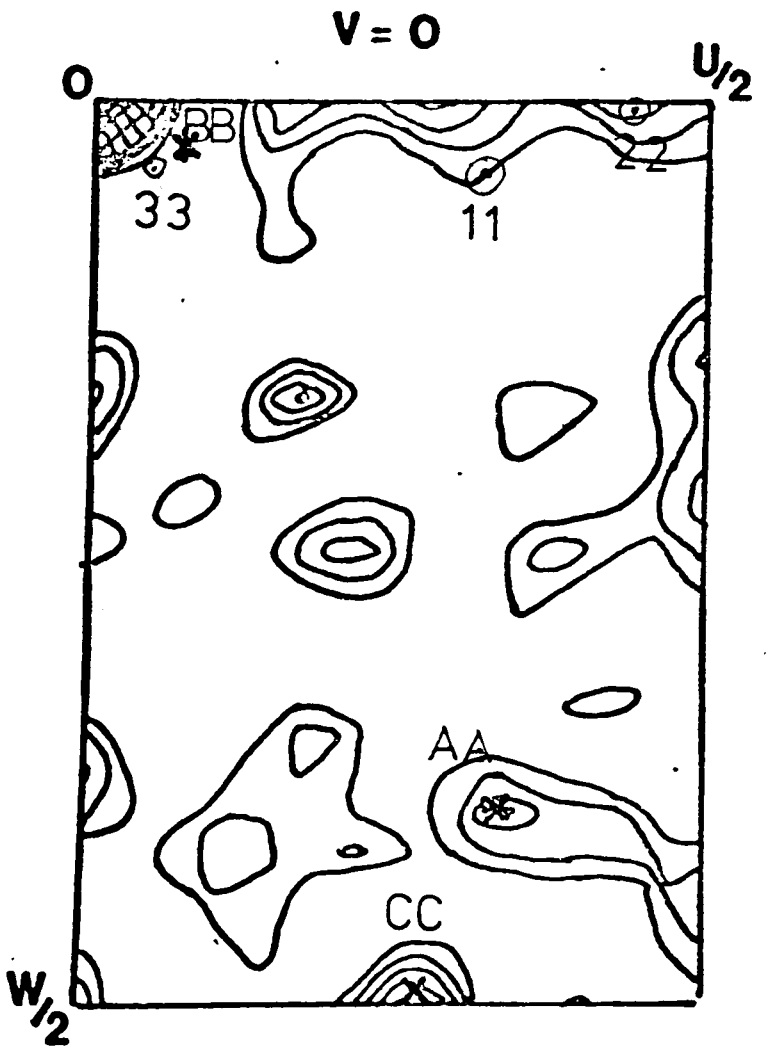
At this point in time phases were calculated using the anomalous differences for the gold derivative (see § 3.6) which enabled a difference Fourier map to be calculated with terms $(\ |\underline{F}_{PT}| - |\underline{F}_{P}|\ )e^{i\alpha_{Au}}$. The map indicated a 3 site solution (Figure 3.12, Table 3.12) different from that yielded by MULTAN as described above. The Harker and cross vectors

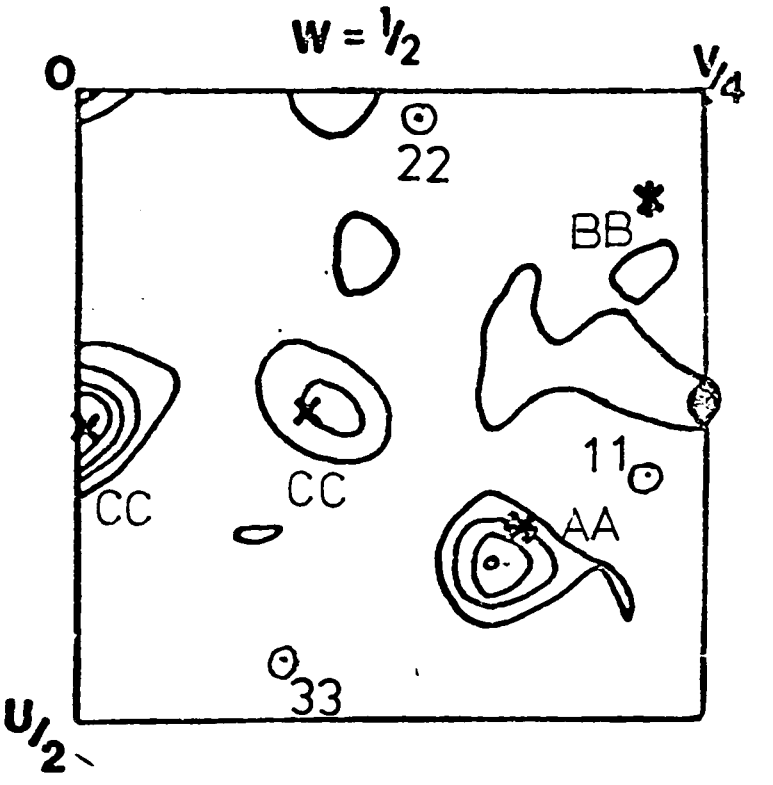
Figure 3.10

The three Harker and selected general sections for the Pt derivative isomorphous difference Patterson showing the Harker and cross vectors. The numbered labels (1-1 etc.) are vectors for sites from the MULTAN solution run 1, and the alphabetic labels (A-A etc.) are vectors for sites from the difference Fourier calculated using Au phases and MULTAN run 2. The origin peak is 1000, contours start at 25 and the contour interval is 25.

Figure 3.10







AA
BB
CC

Figure 3.10

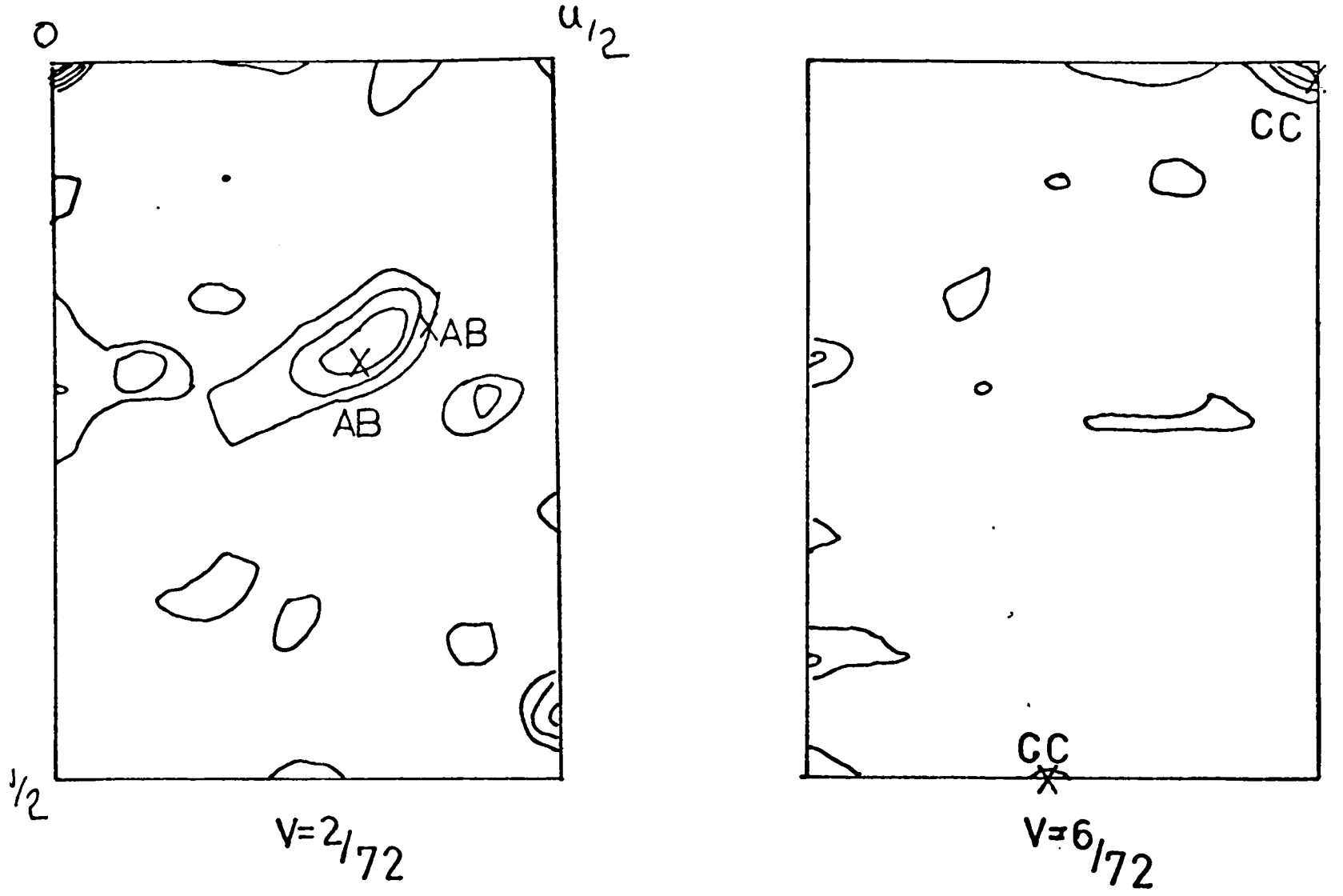
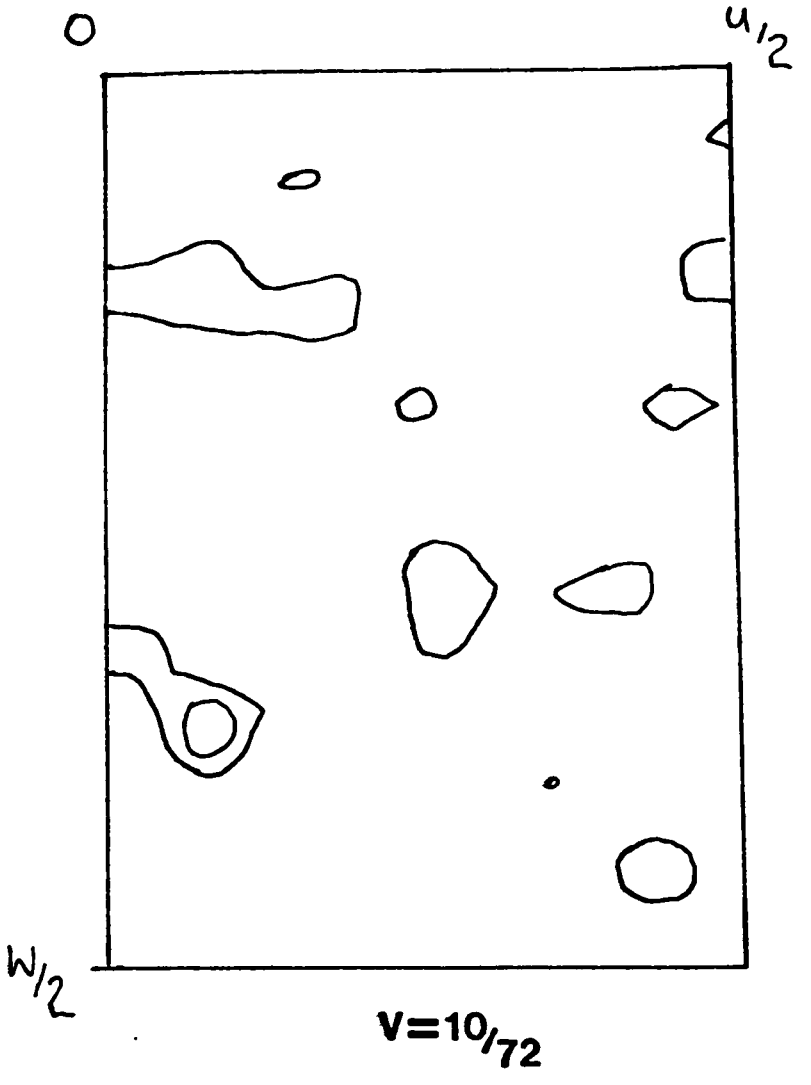
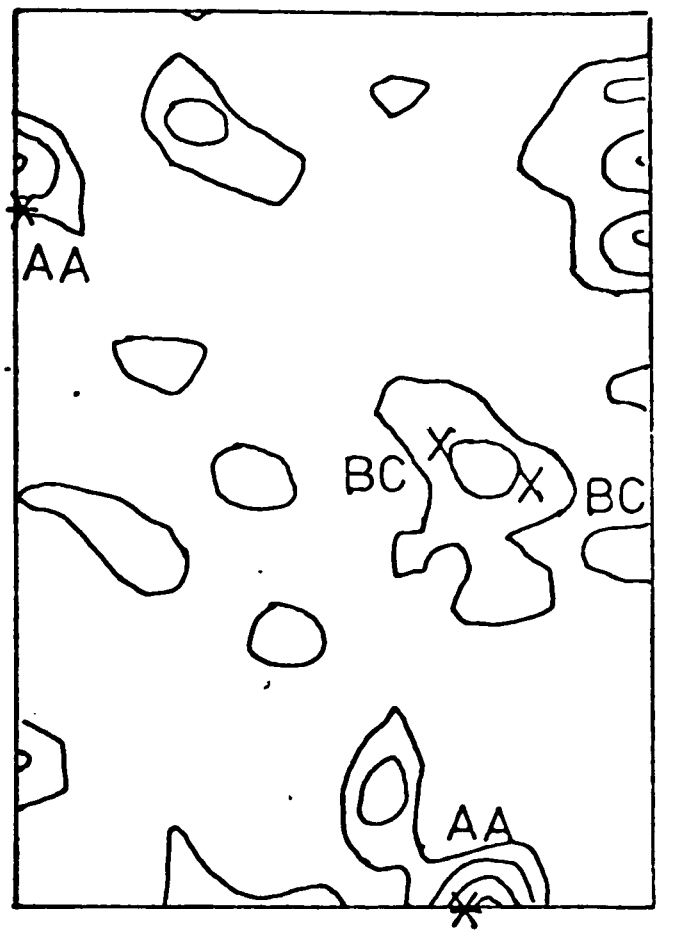


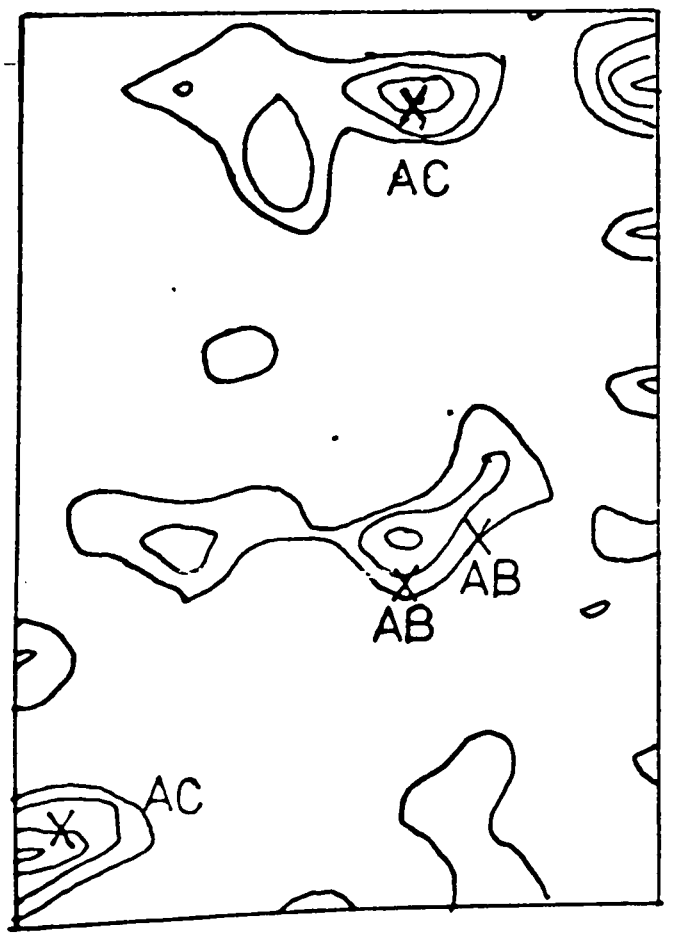
Figure 3.10



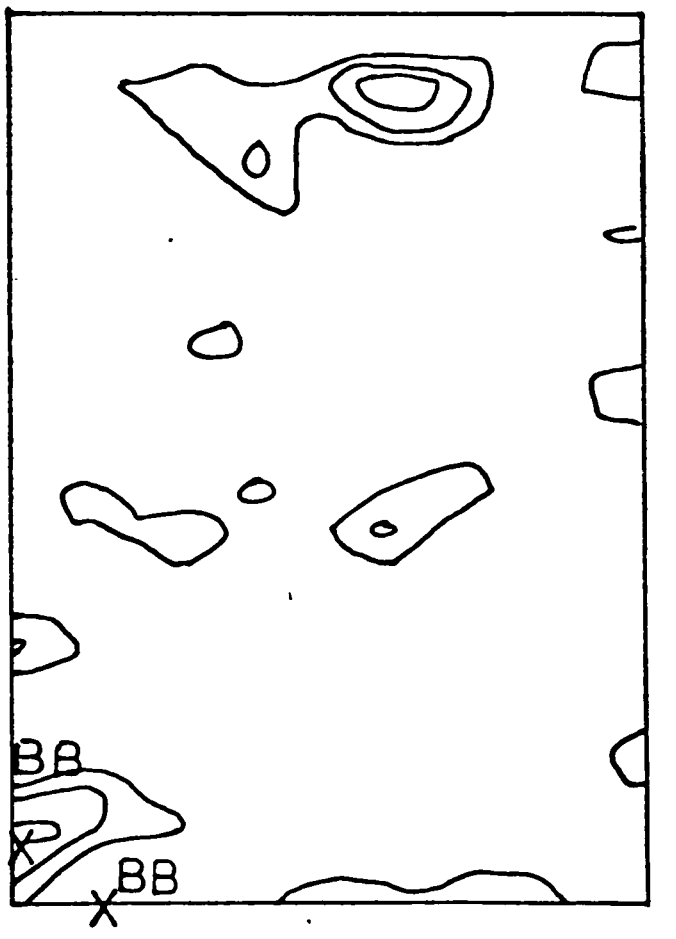
V=10/72



V=13/72

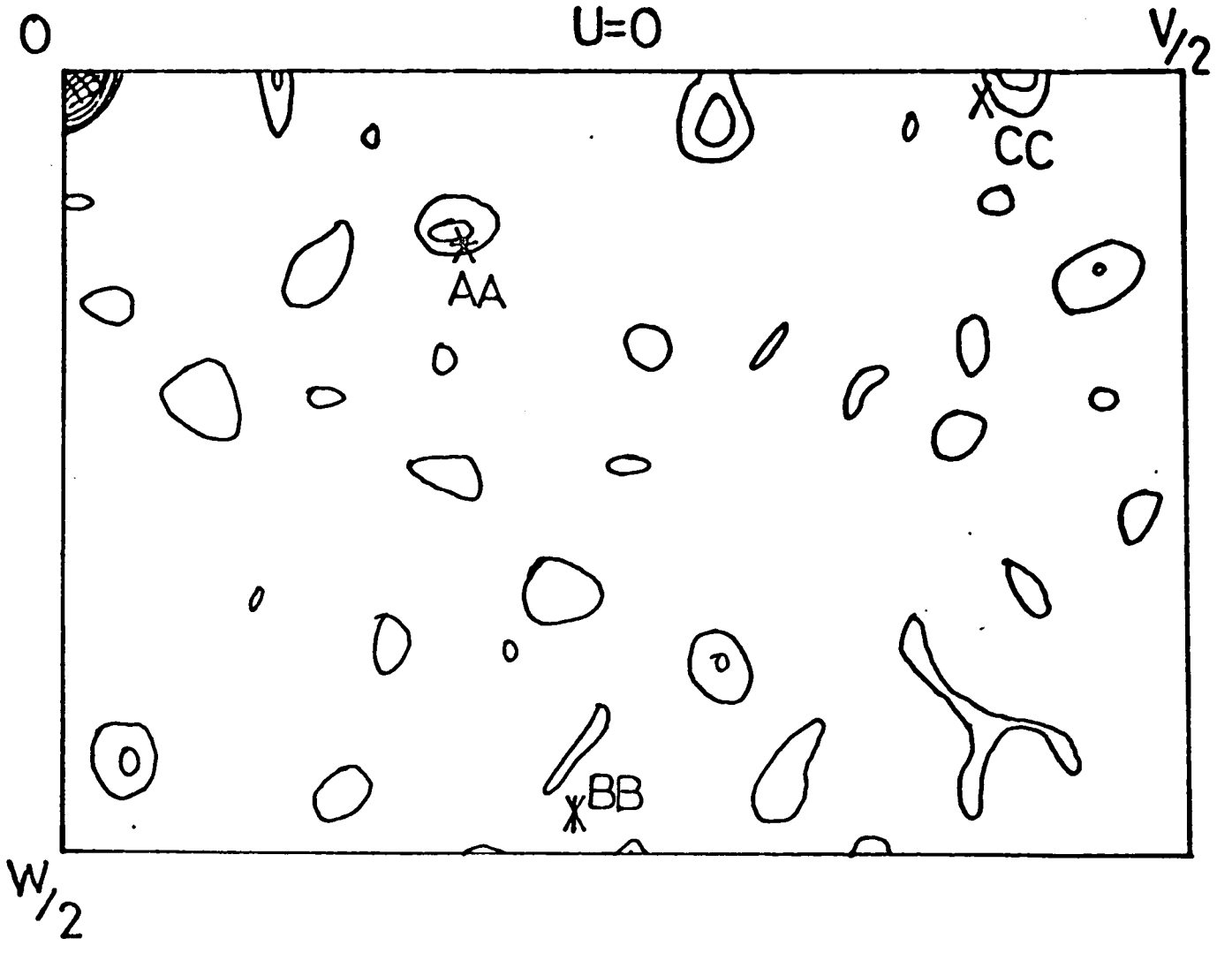


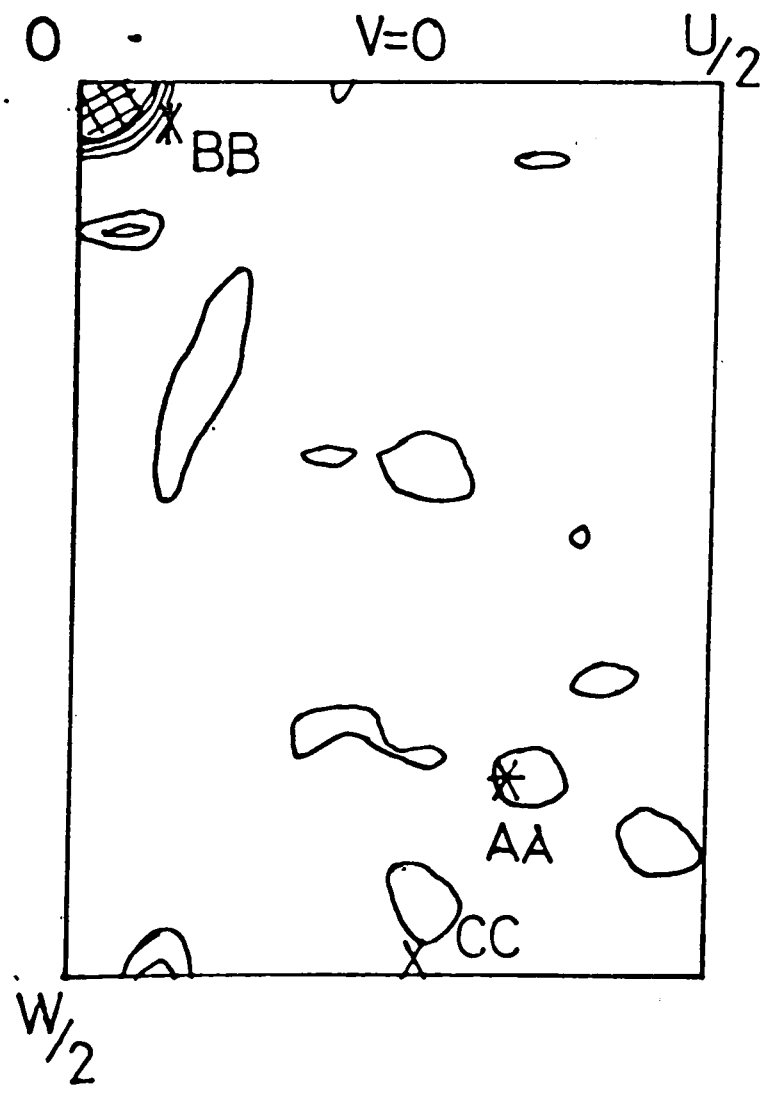
V=15/72

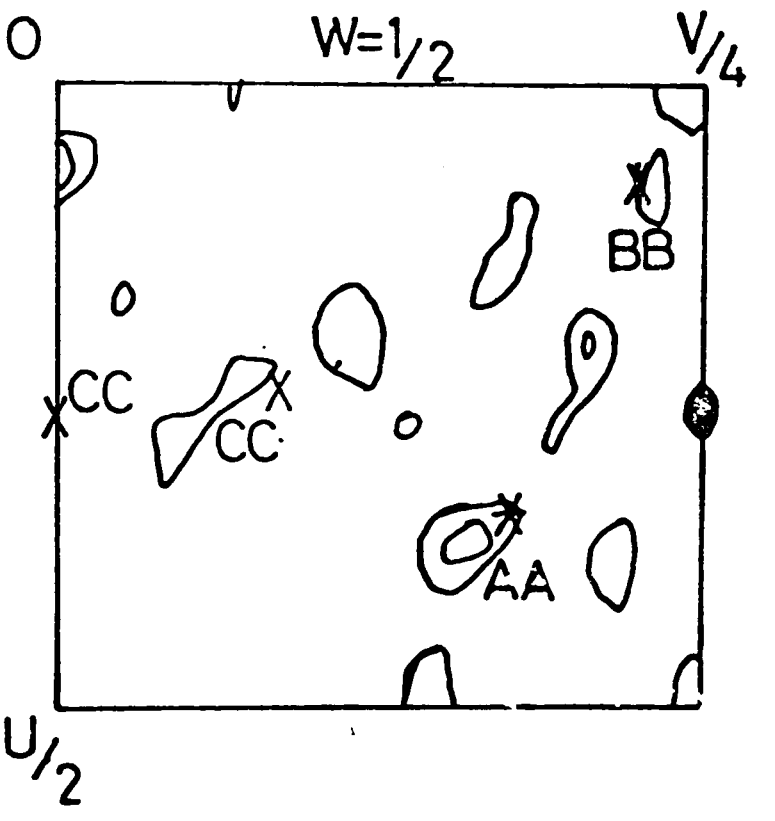


V=16/72

Figure 3.11 The three Harker sections for the Pt derivative F_{HLE}^2 Patterson synthesis. The origin peak is 1000, contours start at 25 and the contour interval is 25.







* A-A
* B-B
X C-C

Table 3.9 The starting set selected by MULTAN for the Pt derivative, run 1 with no $|\underline{F}_P|$, $|\underline{F}_{PH}|$ rejection, and with 200 $E_{\underline{h}}$.

Reflection type	Indices	$E_{\underline{h}}$	$\|\underline{F}_{PH}\|$	Permitted phase	Initial weight
Origin fixing	0 2 5	3.04	343	0	1.0
	1 23 0	2.73	194	0	1.0
Enantiomorph Σ_1	1 7 1	1.94	393	45°, 135°	1.0
	0 10 0	3.11	328	0	0.98
	0 20 10	2.84	244	0	0.96
	0 0 14	2.80	466	180	0.99
	0 14 6	2.80	218	0	0.84
	2 14 0	2.76	1011	0	0.94
	0 20 6	2.61	193	0	0.86
	8 0 12	2.59	259	0	0.98
	0 20 8	2.41	134	0	0.95
	0 0 10	2.40	1225	180	1.0
	0 16 0	2.28	1016	0	0.94
	6 0 0	1.96	458	0	0.97
	0 14 14	1.93	221	0	0.93
	6 6 0	1.45	704	180	0.94
General	1 7 1	1.94	393	45°, 135°	1.0
	1 5 16	1.81	319	± 45°, ± 135°	1.0

Table 3.10 The figures of merit for the phase sets generated by MULTAN for the 6PGDH Pt derivative with 200 E_h at 6 Å. EMAP 1 gave a solution with 3 sites.

NUMSET	ABSFOM	PSIZERO	RESID	COMBINED FOM
1	0.4657	0.5059E 04	33.88	2.3546
2	0.2728	0.3421E 04	33.60	2.0213
3	0.2390	0.3010E 04	33.54	2.0000
4	0.4624	0.6236E 04	42.46	1.0299
5	0.4610	0.6385E 04	40.33	1.2185
6	0.4627	0.6337E 04	41.69	1.0880
7	0.3850	0.4346E 04	33.99	2.1977
8	0.4620	0.6341E 04	40.42	1.2254

Figure 3.12 Selected sections of the difference Fourier calculated with terms$(FP_{Pt} - FP)\ e^{i\alpha_{Au}}$.

Sites PTA, PTB and PTC are marked as is the ghost peak corresponding to Au1. (a) $y = \frac{6}{72}\ b$, (b) $y = \frac{8}{72}\ b$, (c) $y = \frac{15}{72}\ b$.

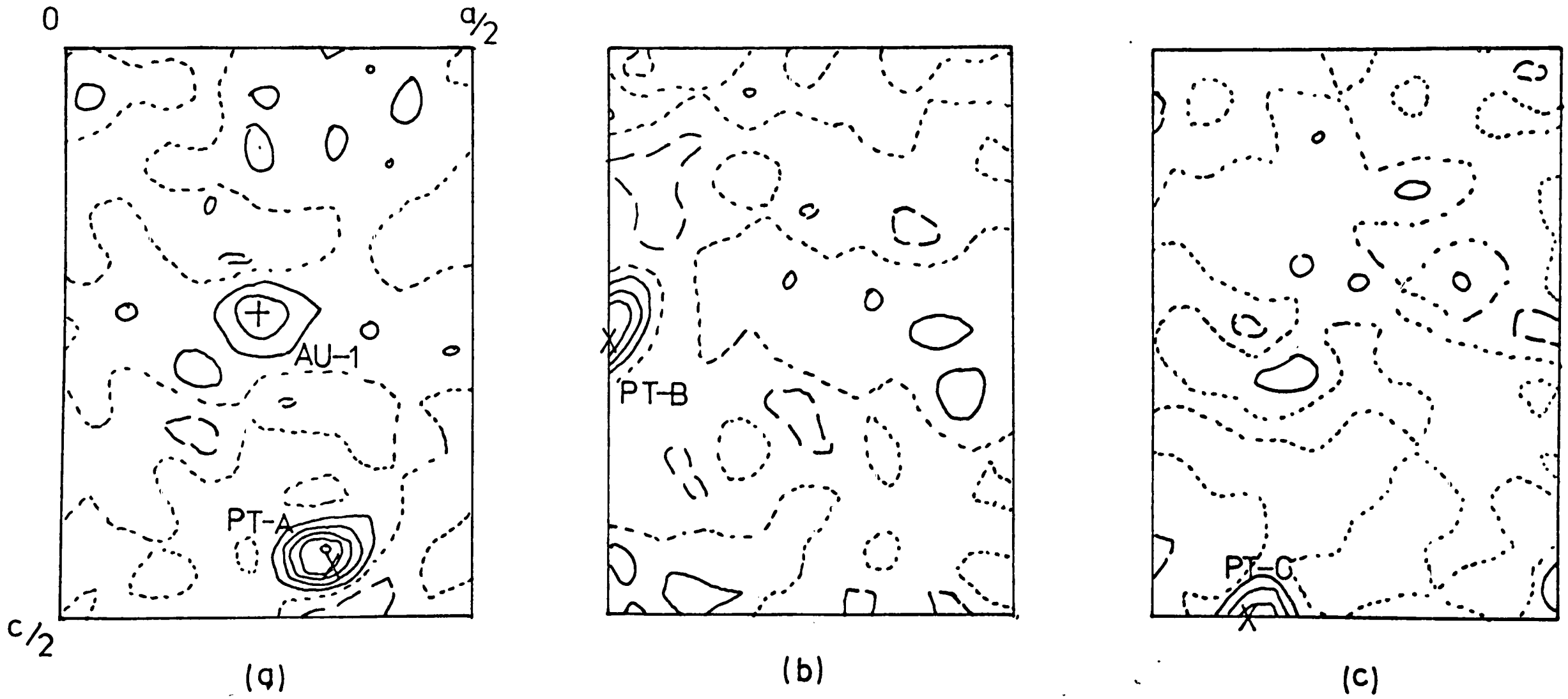


Table 3.11 The absolute and relative peak heights for the largest peaks obtained from EMAP #1 derived from MULTAN run 1 for the Pt derivative. The peak heights should be x 1.021 to bring them onto the same scale as in Au EMAP#10.

Peak label	Sectional coordinates X, Y, Z	Peak height	Relative peak height wrt highest peak
Pt 1	14, 28, 13	361	1.0
Pt 2	8, 22, 12	133	0.3684
Pt 3	20, 20, 12	180	0.4986
next highest		115	0.3186

Table 3.12 The absolute and relative peak heights for the largest peaks obtained from the Pt difference Fourier using Au phases on the correct hand.

Peak label	Sectional coordinates X, Y, Z	Peak height	Relative peak height wrt highest peak	Refined occupancy	Rel. refined occupancy
PtA	13, 6, 21	257	1.0000	0.5377	1.0000
B	1 , 8, 11	205	0.7977	0.4099	0.7582
C	5 , 15, 24	186	0.7237	0.4782	0.8893
Au1(ghost)		140	0.5448		
next highest		105	0.4086		
$\langle b_g \rangle$		86	0.3346		

Note In Tables 3.11 and 3.12 the sectional coordinates are in units of 1/40, 1/72, 1/48 along $|\underline{a}|$, $|\underline{b}|$, $|\underline{c}|$ respectively

for these newly obtained sites are shown in Figure 3.10 and agree better with the difference Patterson.

Attention was directed to the reason why MULTAN had failed to yield a correct solution since it had worked so well for the Au derivative. Since 30% of the data were less than $4\sigma(I)$ (before merging) it was desirable to check whether the weak and less significant data had unduly affected the phasing procedure. A simple test rejecting all data below 200 (on $|\underline{F}_P|$ and $|\underline{F}_{PH}|$, where the maximum $|\underline{F}_P|$, for example, was approximately 1200) was used which amounted to 477 reflections being left out of the total of 1447. This accounted for 33% of the data and hence approximated to a $4\sigma(I)$ cut off (assuming of course that the poorest $I/\sigma(I)$'s were associated with the weakest data). Two hundred largest $E_{\underline{h}}$'s down to 1.26 were selected, a lower value for $E_{\underline{h}}$ being tolerated since less reflections were now available. The reflections used in the starting phase set are shown in Table 3.13, the constitution of which is different from that shown in Table 3.9. The figures of merit for the phase sets generated are shown in Table 3.14. Phase set 13 had the highest ABSFOM, the next to the lowest RESID and the highest combined FOM. An EMAP calculated using this phase set yielded a heavy atom structure of three sites (Figure 3.13), the coordinates of which after suitable rotation agreed with those predicted using the gold derivative phase information. Table 3.15 shows the peak heights of these sites.

In Run 1, the badly estimated data had a significant effect on the phasing procedure both in the starting phase set and in the basic set of 200 $E_{\underline{h}}$'s used. This illustrates how crucial accurate data can be and in this case the weak data was the source of the problem. It is this data that ordinate analysis tends to overestimate (see § 2.2) and yielded spuriously large isomorphous differences.

For the Pt derivative, an alternative attack was tried using F_{HLE}'s

Table 3.13 The starting set selected by MULTAN for the Pt derivative, run 2 with a rejection cut off of 200 for $|\underline{F}_P|$ and $|\underline{F}_{PH}|$.

Reflection type	Indices	$E_{\underline{h}}$	$\|\underline{F}_{PH}\|$	Permitted phase	Initial weight
Origin fixing	0 2 5	3.15	343	0	1.0
	1 7 1	1.83	393	45	1.0
Enantiomorph $\sum_1$	1 7 1	1.83	393	45	1.0
	0 0 14	2.61	466	180	0.94
	2 22 0	1.47	194*	0	0.93
	4 20 0	1.36	327	180	0.89
	0 0 10	1.36**	1225	180	0.99
	0 8 10	1.30	266	0	0.81
General reflections	1 7 8	1.76	440	$\pm$45, $\pm$135	1.0
	4 2 4	1.71	408	$\pm$45, $\pm$135	1.0

*The rejection test was applied after the $|\underline{F}_{PH}|$ / $|\underline{F}_P|$ scale of 1.03 was used.

** This compares with a value of 2.40 for the 0 0 10 reflection in Table 3.9 for the same derivative. MULTAN run 1 used a scale between $|\underline{F}_{PH}|$ and $|\underline{F}_P|$ of 1.07, and run 2 a scale of 1.03. A third MULTAN run performed on all the data with a scale of 1.03, yielded a heavy atom structure which was different from run 1 and run 2, i.e. it was still incorrect.

Table 3.14 The figures of merit for the 16 phase sets generated by MULTAN run 2 with amplitude rejection of 200 for the 6PGDH Pt derivative with 200 E_h of 6 Å. EMAP 13 gave a solution of 3 sites agreeing with those predicted independently by the Au phase set.

NUMSET	ABSFOM	PSIZERO	RESID	COMBINED FOM
1	0.5957	0.4593E 04	34.95	2.0270
2	0.3780	0.3656E 04	38.24	1.1261
3	0.4128	0.3674E 04	37.74	1.3412
4	0.5412	0.4666E 04	35.33	1.7129
5	0.4653	0.3924E 04	38.91	1.1498
6	0.4830	0.3933E 04	37.12	1.5680
7	0.4468	0.3768E 04	36.71	1.6151
8	0.5039	0.4280E 04	38.48	1.1710
9	0.5382	0.4287E 04	37.08	1.5692
10	0.4738	0.3963E 04	39.37	1.0676
11	0.4846	0.4031E 04	39.94	0.9526
12	0.4954	0.4352E 04	39.67	0.8595
13	0.6014	0.4546E 04	35.10	2.0449
14	0.4539	0.4022E 04	37.51	1.3341
15	0.3227	0.2940E 04	37.12	1.5644
16	0.5400	0.4676E 04	35.95	1.5792

Figure 3.13 Selected sections of EMAP13 from MULTAN run 2 for the Pt derivative with the positions corresponding to the final refined positions of PTA, PTB and PTC marked.

(a) $y = \frac{6}{72}$ b, (b) $y = \frac{8}{72}$ b, (c) $y = \frac{15}{72}$ b .

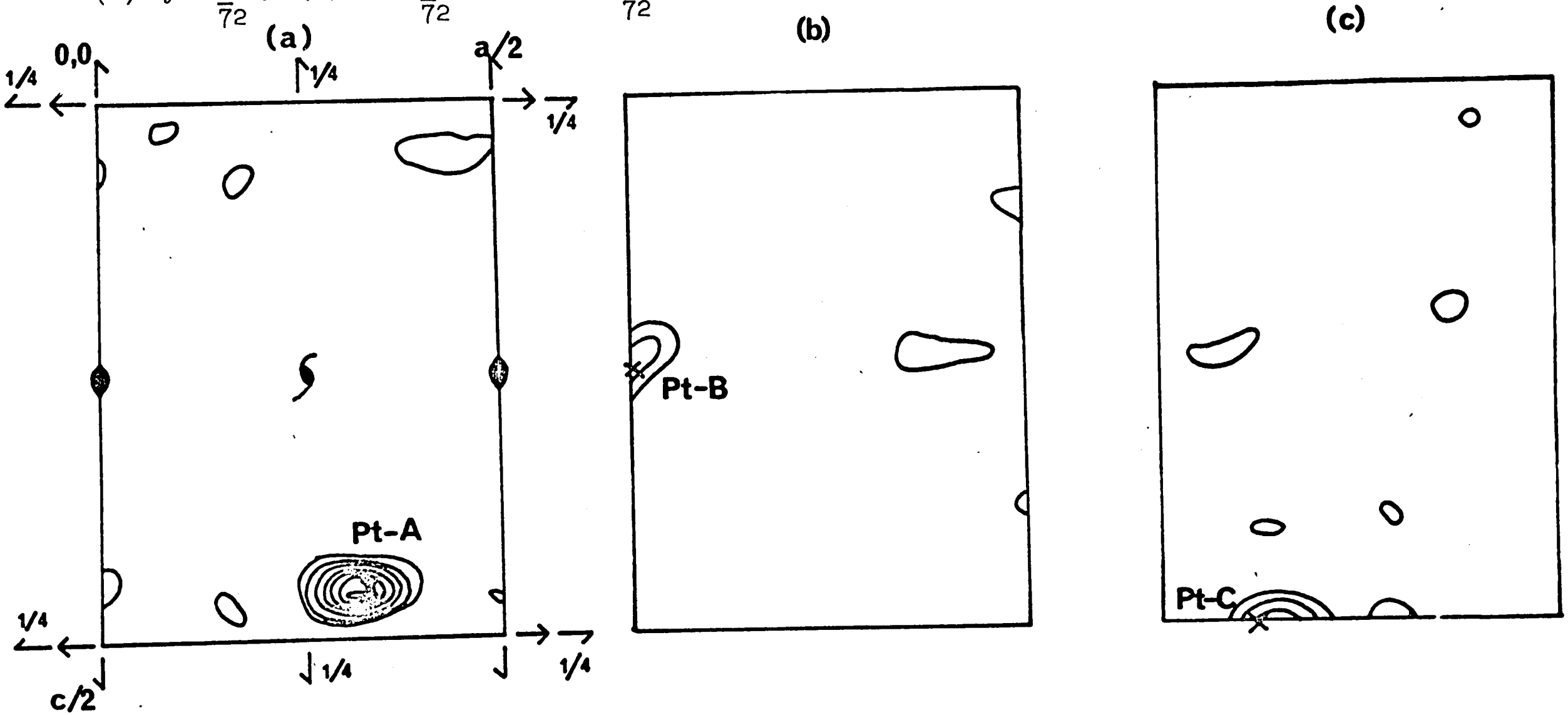


Table 3.15 The absolute and relative peak heights for the largest peaks obtained from EMAP 13 derived from MULTAN run 2 with $|\underline{F}_P|$ & $|\underline{F}_{PH}|$ cut off of 200 for the Pt derivative. The peak heights should be multiplied by 0.962 to bring them onto the same scale as EMAP 10 for the Au derivative.

Peak label	Peak height	Relative peak height wrt highest peak	Refined occupancy	Relative occupancy
PtA	373	1.0000	0.5377	1.0000
PtB	121	0.3244	0.4099	0.7582
PtC	188	0.5040	0.4782	0.8893
Next highest	93	0.2493		

Table 3.16 The statistical distribution of the $E_{\underline{h}}$ values calculated from F_{HLE}'s for the Pt derivative to 6 Å resolution.

	Experimental	Theoretical	
		Centrosymmetric	Non-centrosymmetric
$\langle E_{\underline{h}}^2 \rangle$	1.0000	1.0000	1.0000
$\langle E_{\underline{h}}^2 - 1 \rangle$	0.8536	0.9680	0.7630
$\langle E_{\underline{h}} \rangle$	0.8385	0.7980	0.8860

as the estimate of $|\underline{F}_H|$ in the input to MULTAN. Expectations were not high since Wilson (1974) had obtained negative results using F_{HLE}'s to solve the heavy atom structures of other protein derivatives using MULTAN. The F_{HLE}'s were calculated using a k_{emp} (from ANSC) by the program FHLEKR. No cut off based on $|\underline{F}_P|$, $|\underline{F}_{PH}|$ was used in the input to the program. The procedure adopted in MULTAN was identical to that used previously using isomorphous differences. However, the statistical distribution of the $E_{\underline{h}}$ was closer to being non-centrosymmetric than for the $E_{\underline{h}}$'s derived from the isomorphous differences (see Table 3.16 and compare with Table 3.5). EMAPs corresponding to phase set 8 (highest combined FOM) and phase set 13 (highest ABSFOM) were calculated neither of which showed peaks corresponding to the correct positions. Because the use of F_{HLE}'s in the I/P to MULTAN did not yield a correct solution the problem encountered with the Pt derivative cannot be due to making the approximation $\Delta_{ISO} = |\underline{F}_H|$ but lies with the inaccurate estimation of Δ_{ISO}.

Having obtained the Pt 3 site solution the scale factors and heavy atom positions of this derivative were refined (see § 3.5) and phases calculated. A difference Fourier calculated with terms $(|\underline{F}_{Au}| - |\underline{F}_P|)\, e^{i\alpha_{Pt}}$ showed the Au first and second sites clearly (see Figure 3.14, Table 3.17) and formed a cross check on the correctness of the heavy atom structures. Details of the single derivative phasing for the Pt and Au are given in § 3.6.

It should be made clear how the enantiomorph was determined. With only one isomorphous derivative at low resolution it is not possible to choose the correct hand of the space group. Inclusion of the anomalous scattering information for the gold allowed the computation of two sets of phases corresponding to two possible observed signs for the anomalous differences. The reversal corresponds to a change of enantiomorph since

Figure 3.14 Selected sections of the difference Fourier calculated with terms $(FP_{Au}-FP)e^{i\alpha_{Pt}}$. Sites Au1 and Au2 are marked as is the ghost peak corresponding to PTA. (a) $y = \frac{6}{72}$ b, (b) $y = \frac{8}{72}$ b, (c) $y = \frac{17}{72}$ b.

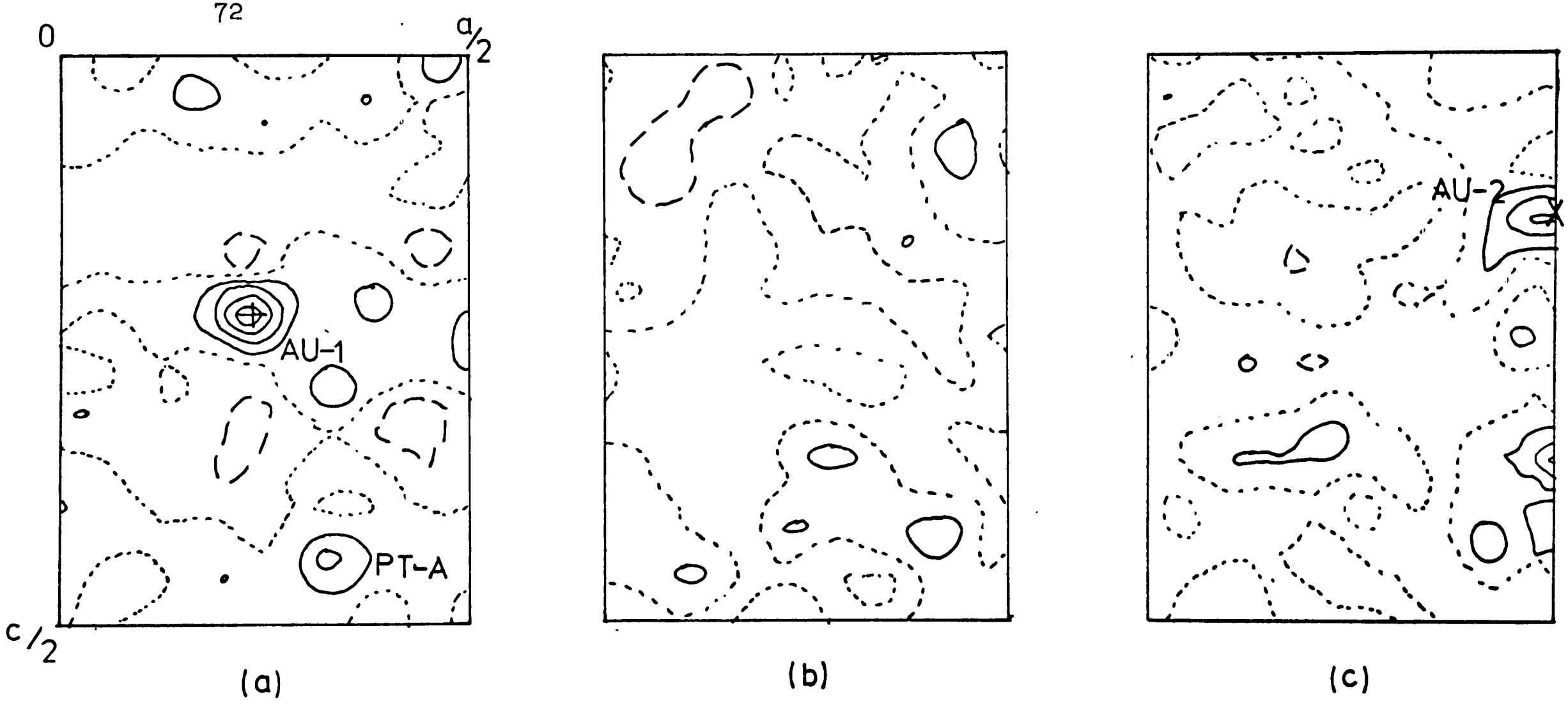


Table 3.17 Absolute and relative peak heights of the Au as derived from the Au difference Fourier using Pt phases. This map is on the same scale as the Pt difference Fourier using Au phases (Table 3.12).

Peak label	Peak heights	Relative peak height wrt highest peak	Refined occupancy	Relative refined occupancy
Au 1	225	1.0000	0.7550	1.0
Au 2	150	0.6667	0.3770	0.4993
next highest	93	0.4133		
<bg>	75	0.3333		

the sign of the measured anomalous difference is absolute. The anomalous information can be left out altogether in which case the phase ambiguity ofthe acentric terms is not resolved (see § 2.1). The calculation of difference Fourier maps for an independent derivative using these three different phase sets should yield peaks of different heights, the one containing the highest peak corresponds to the phases on the correct hand.

Sections from three difference Fouriers calculated with terms $\left(|\underline{F}_{PT}| - |\underline{F}_{P}|\right) e^{i\alpha_{Au}}$ with the three different estimates of α_{Au}, as outlined above, are illustrated in Figure 3.15 and clearly indicate the phase set on the correct hand. This phase set yielded the highest heavy atom peaks with respect to the Au ghost peak. The phase set calculated on the wrong, i.e. other, hand, showed much reduced Pt peaks. The phase set calculated with no anomalous differences yielded Pt peak heights in between those based on the other two phase sets.

(iii) Hg derivative

The analysis of the Hg derivative was made considerably simpler since phases could now be calculated using the solved heavy atom structures of the Au and Pt derivatives. However, because of the relatively low value of the mean fractional isomorphous difference and k_{emp} output by ANSC for the Hg derivative, hopes did not run high. A difference Fourier map was calculated using terms $(|\underline{F}_{Hg}| - |\underline{F}_{P}|) e^{i\alpha(Au+Pt)_2}$, where $\alpha(Au+Pt)_2$ was the final, best double derivative phase set. This map yielded only small peaks at positions corresponding to the gold heavy atoms (i.e. Au1 and Au2) which were presumably ghost peaks. However, no ghost peaks were observed at the positions corresponding to the Pt binding sites (Table 3.18a). In order to check the possibility that the mercury had substituted weakly at the same sites as the Au a difference Fourier map was calculated using terms $(|\underline{F}_{Hg}| - |\underline{F}_{P}|) e^{i\alpha_{Pt}}$, where the α_{Pt} phases were calculated using the isomorphous and anomalous differences of the Pt only and on the correct hand. This map showed small peaks corresponding to the three Pt binding sites (i.e. PtA, PtB, PtC) which were the largest

Figure 3.15

Determination of the enantiomorph. Difference Fourier for the platinum derivative with phases obtained from the gold derivative alone, (a), (c) and (e) are section Y = 6/72b; (b), (d) and (f) are section Y = 8/72b. The anomalous component is included with the right hand in (a) and (b) and with the wrong hand in (e) and (f). In (c) and (d) no anomalous component has been included. The major platinum site (A), one of the smaller sites (B), and the ghost of the major gold site (1), are indicated.

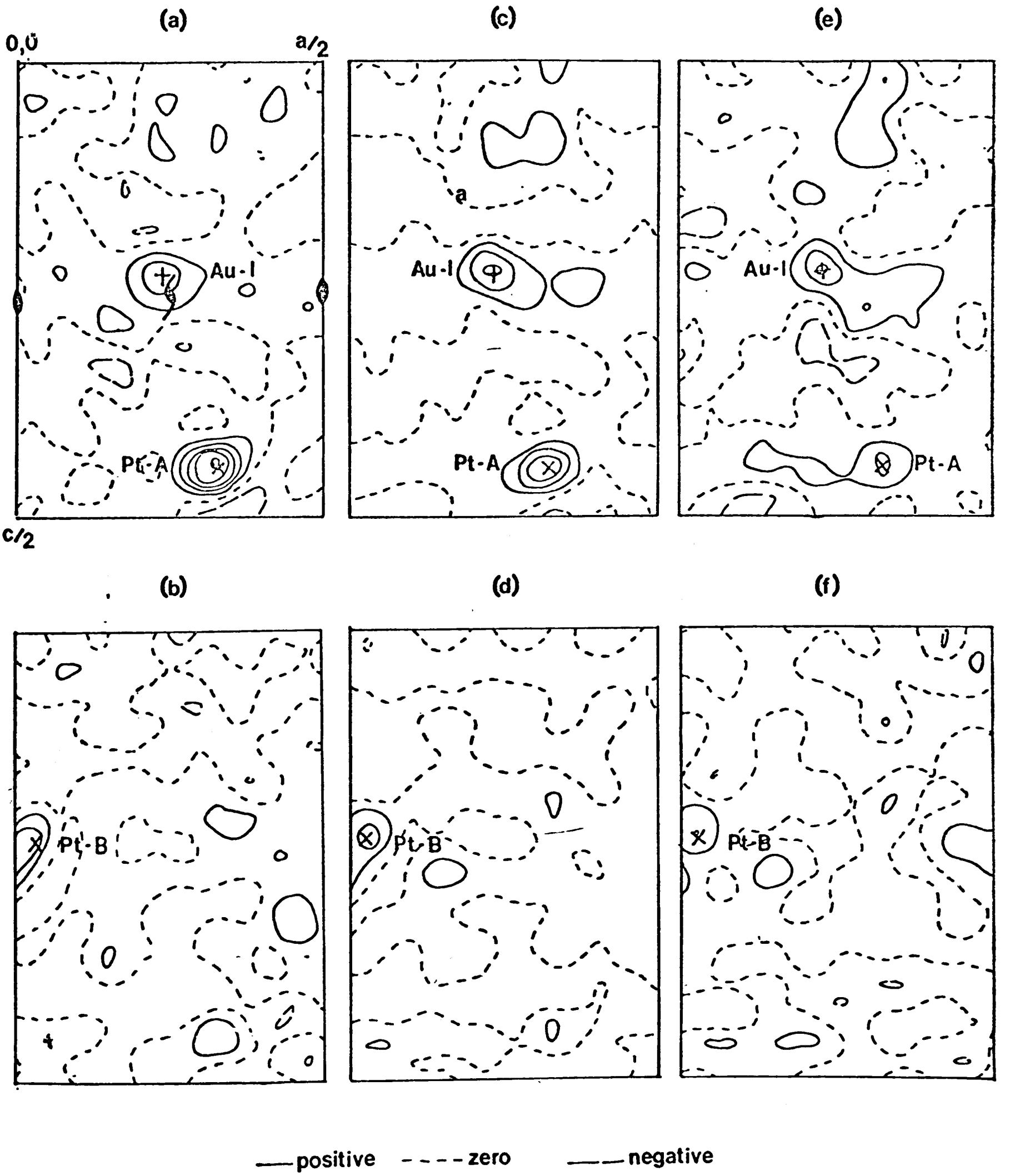

(a)
(c)
(e)
0,0
a/2
c/2
Au-I
Pt-A
a
(b)
(d)
(f)
Pt-B
positive
zero
negative

Table 3.18 Absolute and relative peak height for the largest peaks on the Hg difference map as derived using a) double derivative phase set $(Au+Pt)_2$ and b) the Pt single derivative phase set. The peaks are on the same scale as the peaks on the Pt, Au difference maps detailed in Tables 3.12, 3.17, respectively.

	Peak label	Sectional coordinates*			Peak height
a)	Au 1 (ghost)	9	5.5	11	37
	Au 2 "	19.5	15.75	6.5	25
	Pt C "	5	14.5	24	15
	next highest	10.5	26.5	11.5	11
	<background>				8
b)	PTA (ghost)	13	6	21.5	12
	PTB "	12	8.5	0.5	13
	PTC "	5	14.5	24	17
	next highest (similar site as Au 1)	8	5.5	11	12
	" " (similar site as Au 2)	18	15.5	7	11
	" "	9.5	26.5	16	11
	<background>				9

* In units of 1/40, 1/72, 1/48 along a , b , c respectively.

on the map (see Table 3.18b). The next highest peaks were however at similar positions as the Au1 and Au2 sites. This indicated that perhaps the mercurial was bound very weakly at the same sites as the Au and agrees with the observation of the character of the intensity differences in the h0l precession zones of the Au and Hg derivatives with respect to native (Figure 3.5). If this is taken as evidence of binding of the mercurial, expected to bind at one of the seven thiol groups of the enzyme subunit (see § 3.2), then the $Au(CN)_2^-$ has bound at two of these sites.

Because of the poor binding of the mercurial and the cracking of the crystals at higher concentrations of the reagent it was decided not to pursue this derivative further.

3.5 Refinement of the heavy atom parameters

In § 2.5 the means available for refining the heavy atom positions and occupancies and relative scale factors between the native and derivative data were discussed. In the situation, which existed with the 6PGDH crystals, the relevant factors were:

(i) enough centric data were present (23% of the data set to 6 Å) for the data to parameter ratio to be sufficiently overdetermined without using the acentric data (for which $\Delta_{ANO} \leq \sigma(\Delta_{ANO})$, see Table 3.3);

(ii) the single derivative phases would not be of sufficient quality to enable phase refinement of one derivative with the independent phase information of the other.

It was decided to use centric least squares refinement to give the relevant parameters for phasing and for comparison perform acentric, $(F_{HLE} - F_{HCALC})^2$ and 3D phase refinements.

These refinements were performed using the program REFINE (see Appendix IV) which has the options FHLESC, FHLEKR and PHASED.

Centric FHLESC refinement corresponds to minimising the difference

Table 3.19 Indicators of heavy atom refinement

Refinement index	gold		platinum		
	1 site	2 sites	1 site	2 sites	3 sites
R_C	73.9	61.4	73.7	65.5	60.7
wR	35.4	22.8	38.5	42.4	34.3
g	0.24	0.51	0.09	0.25	0.29
G.F.	1.92	1.25	1.65	1.41	1.22
Number of reflections	319	319	322	319	323

Terms are defined in § 2.5.

between the isomorphous differences, Δ_{ISO}'s and the F_{HCALCS} since $\Delta_{ANO} = 0$ and $F_{HLE} = |\underline{F}_{PH}| - |\underline{F}_{P}|$ (see § 2.5). Generally, in FHLESC bias corrected F_{HLE}'s estimated using a k_{theor} are used and by calculating the F_{HLEB}'s between each refinement cycle, the scale factor k between the native and derivative data used in the calculation of F_{HLE}, can be refined also. This is an improvement on the option F_{HLEKR} where the I/P data is a set of F_{HLE}'s calculated with a k_{emp} and a native to derivative Kraut scale, K_K (see § 3.3). The value of K_K cannot be refined in the FHLEKR option. However, because the bias correction did not ensure that the estimation of $|\underline{F}_H|$ for centric and acentric reflections was approximately equal, acentric FHLESC refinement is not discussed (see further on in this section).

The option PHASED performs refinement of one derivative with the independent phase information of one or more other derivatives and allows m weighting and an m cut-off to be used.

In the centric FHLESC refinement an initial run was performed with a wide $4\sin^2\theta/\lambda^2$ range of 0.0011 - 0.028 and $|\underline{F}_P|$ and $|\underline{F}_{PH}| < 2\sigma(F)$ rejection to establish the native to derivative scale factor k. This scale was then used in the final refinements with a narrower $4\sin^2\theta/\lambda^2$ range of 0.006 - 0.028 and not refined. In these latter refinements the use of the smaller $4\sin^2\theta/\lambda^2$ range and the amplitude rejection aimed at improving the reliability of the refined parameters since a few large differences between F_{HOBS} and F_{HCALC} from poorly measured terms could determine the calculated parameter shifts in the refinement. This procedure had the effect of underestimating the average error, obtained from this refinement, used in the phasing (see § 3.6).

Table 3.19 shows the improvement of the reliability parameters of the centric FHLESC refinement as sites are added to the postulated heavy atom

structure. The values of R_c are rather high but this may be explained since $\sigma(F_{HOBS})$ is of the same order of magnitude as F_{HOBS}; wR is lower as might be expected for a correct solution and refines. Both the gradient (g) and goodness of fit (G.F.) move toward their ideal values of unity. For the correct weighting scheme ($w\Delta^2$) should be independent of sin θ. For the Au derivative this was achieved with a value of k' = 0.03, where k' is used to increase the standard deviation of stronger reflections (§ 2.5). Figure 3.16 shows a plot of $w\Delta^2$ for the Au, for k' = 0.0, 0.03 and unit weights refinement and Figure 3.17 for the Pt, for the k' = 0.0 and unit weights refinements. Using k' = 0.0 means the refinement uses a weight based simply on counting statistics.

In Table 3.20 details are given of the refined parameters for the Au derivative centric k' = 0.03, 0.0, unit weights and acentric FHLEKR refinements. The positional parameters agree closely. The scale factor used in the k' = 0.0, 0.03 centric refinements agrees well with the Kraut scale, whereas the scale from the unit weights refinement is a bit low. The occupancies agree reasonably well.

The parameters used for phasing were taken from the centric FHLESC (k' = 0.03) refinement. The results of the acentric FHLEKR refinement agree very well with this.

Table 3.21 shows the refined parameters for the Pt derivative centric k' = 0.0 and unit weights and acentric FHLEKR refinements. The positional parameters agree closely. The scale factor in the k' = 0.0 centric refinement is a bit smaller than the Kraut scale but still greater than 1.0 which it needed to be. The occupancies agree less well than they did for the Au. The occupancies of sites B and C are reversed in the k' = 0.0 centric FHLESC and acentric FHLEKR refinements.

In retrospect it seems difficult to choose between the centric k' = 0.0

Figure 3.16

Plot of $w\Delta^2$ v.$4\sin^2\theta/\lambda^2$ for the k' = 0.03, 0.00 and unit weights refinements for the Au. The values of $w\Delta^2$ have been normalized such that the maximum is 1.0 in each case and involved dividing by 7.2, 2.7 and 1180.9 respectively.

Figure 3.17

Plot of $w\Delta^2$ v.$4\sin^2\theta/\lambda^2$ for the k' = 0.0 and unit weights refinements for the Pt, with the values of $w\Delta^2$ normalized to 1.0 by dividing by 3.3 and 922.0 respectively.

Figure 3.16

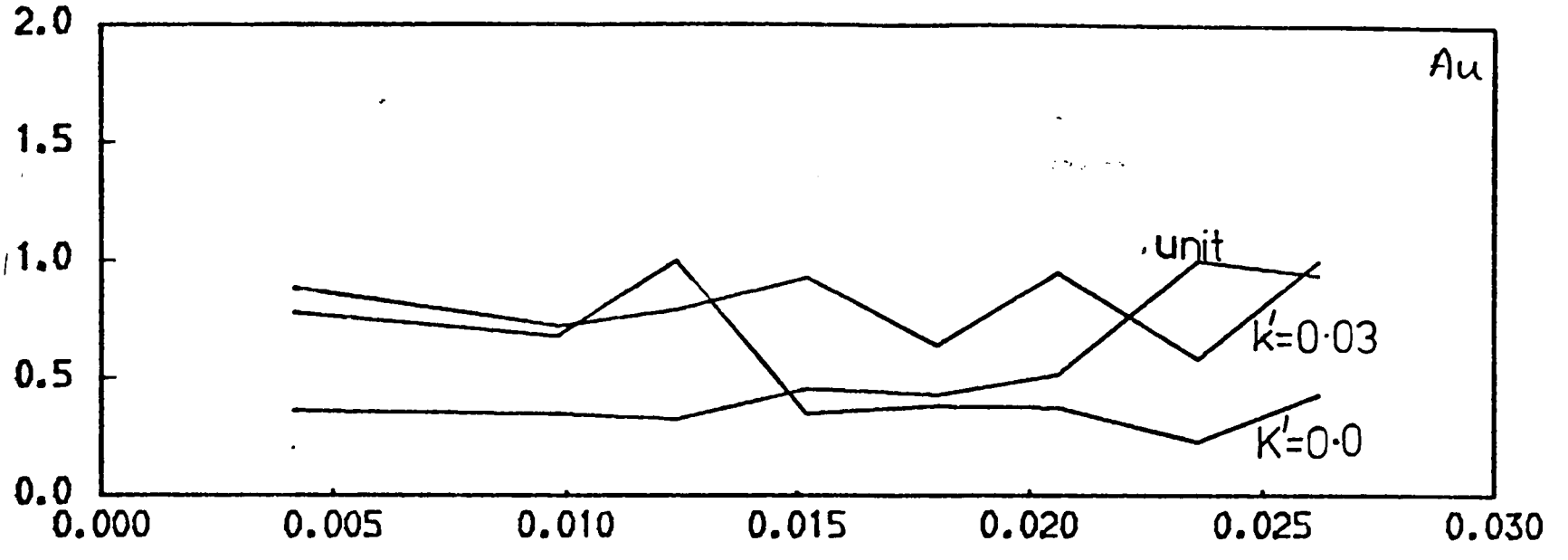


Figure 3.17

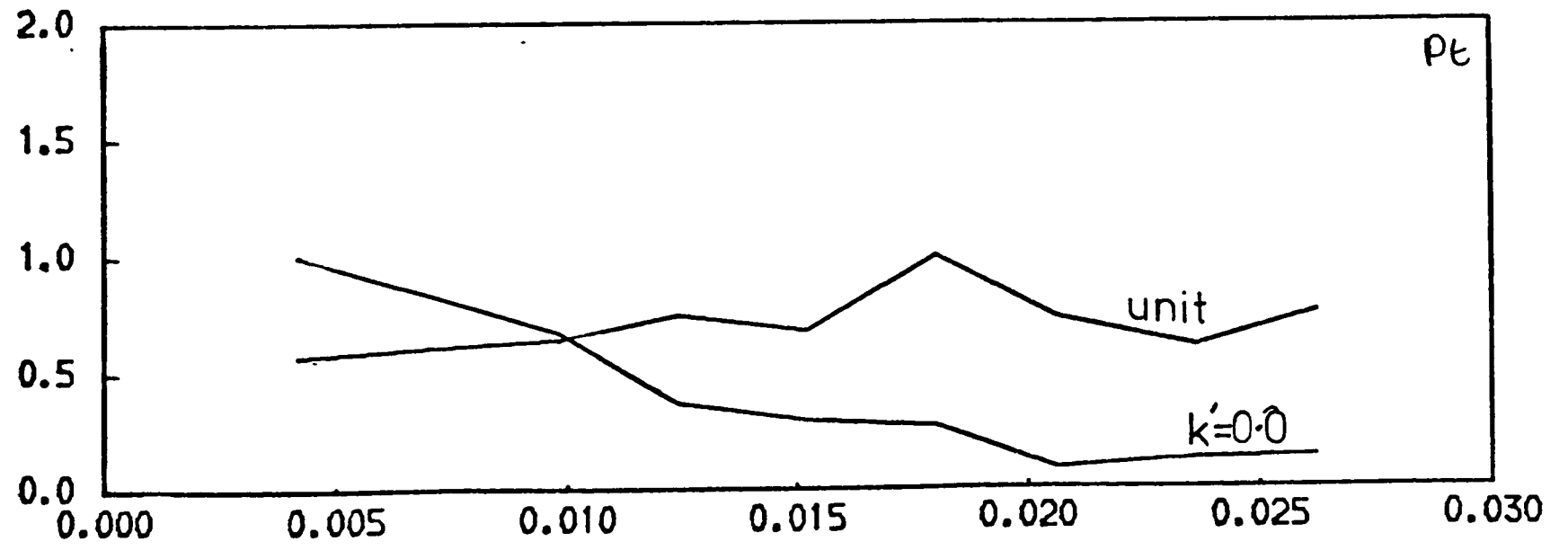


Table 3.20 Refinements performed on the Au derivative

Parameter	Centric $k' \doteq 0.0$	Centric $k' = 0.03$	Centric unit weights	Acentric FHLEKR, FHLE's, k_{emp}
Scale factor	0.8712*	0.8712*	0.8543	0.8762*
Site 1				
Occupancy	0.8393	0.7551	0.6850	0.7436
x	0.2224	0.2293	0.2317	0.2306
y	0.0764	0.0792	0.0796	0.0772
z	0.2367	0.2291	0.2243	0.2261
Site 2				
Occupancy	0.4177	0.3768	0.4166	0.3839
x	0.4996	0.4969	0.4943	0.4904
y	0.2285	0.2228	0.2206	0.2214
z	0.1486	0.1428	0.1442	0.1431
nref	237	257	256	1024
R	72.5	61.1	55.1	47.5
$\mathcal{N}$R	21.6	43.7	58.6	40.5
gradient(g)	0.42	0.33	0.39	0.22
g.F.	1.2082	0.7605	25.5864	0.8601

*unrefined

In each refinement a relative temperature factor of -3.5 was applied to $|\underline{F}_{PH}|$. The scale factor of 0.8712 was selected on the basis of a $k' = 0.0$ centric refinement with a $4\sin^2\theta/\lambda^2$ range of 0.0011 - 0.028, nref = 323. It agrees well with the Kraut scale.

Table 3.21 Refinements performed on the Pt derivative: FHLESC and FHLEKR

Parameter	Centric k' = 0.0	Centric unit weights	Acentric FHLEKR, FHLE's, k_{emp}
Scale factor	1.0189*	1.0189*	1.03541*
Site A			
Occupancy	0.5377	0.5373	0.5976
X	0.3330	0.3354	0.3201
Y	0.0875	0.0859	0.0870
Z	0.4453	0.4494	0.4483
Site B			
Occupancy	0.4099	0.3531	0.3882
X	0.0301	0.0246	0.0254
Y	0.1153	0.1172	0.1172
Z	0.2368	0.2323	0.2374
Site C			
Occupancy	0.4782	0.3908	0.3312
X	0.1276	0.1277	0.1341
Y	0.2065	0.2076	0.2026
Z	0.5036	0.5027	0.5033
nref	264	204	1027
R	62.0	59.2	49.9
wR	34.0	61.3	46.7
gradient(g)	0.23	0.22	0.16
g.F.	1.1138	26.0856	0.9106

* The comments at the bottom of Table 3.20 apply here also.

and unit weights Pt refinements, the former has a poorer R but better g and the latter a better $w\Delta^2$ v. $4\sin^2\theta/\lambda^2$ plot. The parameters used in phasing were from the centric k' = 0.0 FHLESC refinement.

The refinement of the Au and Pt derivatives have been described elsewhere (Dodson (1975)), to illustrate the use of FHLEB in acentric least squares refinement. In this report, a bias correction was used along with a k_{theor} but it was stated that since $<$ FHLEB $>$ was higher than $<\Delta_{ISO}>$ a value of k smaller than k_{theor} was used, i.e. the bias correction had not been effective in ensuring that $0.9 <$ FHLEB$> = <\Delta_{ISO}>$ in accord with Wilson statistics (Wilson (1949)). This meant, of course, that the overestimation of $|\underline{F}_H|$ had been corrected for twice, once by use of the bias correction and again by use of a reduced k_{theor}. Table 3.3b shows the mean values of the estimates made for the heavy atom structure factor amplitude, using a k_{theor} and k_{emp}, in terms of F_{HLE} and FHLEB. The centric and acentric estimates for $|\underline{F}_H|$ should be approximately equal and this is achieved the most admirably for F_{HLE}'s calculated with a k_{emp}. The bias corrected F_{HLE} calculated with a k_{theor} for acentric reflections is still significantly larger than for the centrics and in refinement this would lead to overestimation of the heavy atom occupancies. Hence, in this work, it appears that the bias correction was not as effective as the use of k_{emp} for calculating $|\underline{F}_H|$ for acentric reflections. It does mean, however, that the use of a k_{emp} reduces all F_{HLE}'s irrespective of the relative accuracy of $|\underline{F}_P|$ and $|\underline{F}_{PH}|$.

For comparison, Table 3.22 shows the refined parameters for the Pt derivative using 3-D data in the PHASED option. Three different m cut offs were used and the refinement then m weighted. There is good internal consistency of the positional parameters and with the positions

Table 3.22 Phase refinement of Pt derivative (3 sites) at different m cutoffs and m weighted

Parameter	m > 0.0	m > 0.50	m > 0.85
Scale factor	1.0510	1.0510	1.0546
Site A			
Occupancy	0.3506	0.3693	0.3881
X	0.3181	0.3176	0.3152
Y	0.0863	0.0864	0.0866
Z	0.4466	0.4462	0.4474
Site B			
Occupancy	0.2285	0.2485	0.2720
X	0.0191	0.0194	0.0187
Y	0.1210	0.1210	0.1211
Z	0.2201	0.2279	0.2272
Site C			
Occupancy	0.2146	0.2186	0.2454
X	0.1359	0.1342	0.1398
Y	0.2029	0.2030	0.1990
Z	0.5014	0.5014	0.4990
nref	1434	793	237
initial R	8.2	8.4	9.2
final R	7.4	7.0	7.7

estimated in the FHLESC, FHLEKR refinements. The scale factors agree with each other but are higher than the Kraut scale. The occupancies are lower than the FHLESC, FHLEKR refinements but are seen to increase gradually as the m cut off is increased. Above m = 0.85, the number of reflections available for refinement decreases markedly.

Double difference maps calculated using $(F_{HLE} - F_{OBS})\ e^{i\alpha_c}$ from the acentric FHLEKR refinements for the Pt and the Au showed no evidence of disturbance around the binding sites which would have been indicative of incorrect estimation of occupancies or temperature factors.

It was felt that at this resolution any further investment of time in the refinement would not yield an improvement of the quality of the final electron density map. From the relative ease of interpretability of the Patterson and EMAPs and the agreement factors for the refinements for the Pt and Au derivatives, it is fair to say that the quality of the Pt derivative was not as high as the Au derivative.

3.6. Phase calculation

Phases were calculated using the statistical procedure of Blow and Crick (1959) using the program PHASE written by Dr. H. Muirhead (modified for use in Oxford by Dr. G. Bentley). Blow and Crick (1959) showed that a native or difference Fourier electron density map computed with the best phase gives the lowest root mean square error in the electron density (see § 2.6). This phase is estimated by minimising the lack of closure error, ε, for each source of information at discrete intervals (5-10°) around the phase circle and taking the centroid of the probability distribution. Associated with this probability distribution is a figure of merit indicating the reliability of the phase estimate. The program calculates the best phase, the most probable phase and figure of merit for each reflection and the required input includes the occupancy, temperature factor and positional parameters for each heavy atom, the

overall scale and temperature factor to be applied to $|\underline{F}_{PH}|$ and an estimate of the average error E for each source of phase information. Table 3.23 shows the final parameters for the heavy atom sites used for phasing. A useful analysis in ranges of $4\sin^2\theta/\lambda^2$ of $|\underline{F}_P|$, $|\underline{F}_{PH}|$, FHOBS, FHCALC and $\langle \varepsilon^2 \rangle^{\frac{1}{2}}$ for the best and most probable phase as well as the $\langle \alpha_P - \alpha_H \rangle$ for each derivative is output.

The root mean square error $\langle \Delta\rho^2 \rangle^{\frac{1}{2}}$ in the electron density can be calculated using:

$$\langle \Delta\rho^2 \rangle = \frac{1}{V^2} \sum_{\underline{h}}{}' |\underline{F}_P|^2 \ (1 - m^2) \qquad 3.1$$

(Dickerson, Kendrew and Strandberg (1961)), where V is the volume of the unit cell and the $|\underline{F}_P|$ have to be on an absolute scale.

Provided that m is accurately estimated, $\langle \Delta\rho^2 \rangle$ gives a feel for the quality of the map. The value of m depends on the value of E (see § 2.6) input to the program and for the final phase set should equal the average root mean square lack of closure, in ranges of $4\sin^2\theta/\lambda^2$.

An initial estimate of E can be had from the refinement of the centric zones using the expression

$$\langle E^2 \rangle = \langle \ (\text{FHOBS} - \text{FHCALC})^2 \ \rangle$$

Hence the value calculated in the centric FHLESC refinement described in the previous section (§3.5) can be used. In this refinement though, the use of $4\sin^2\theta/\lambda^2$ and $|\underline{F}_P|$ (or $|\underline{F}_{PH}|$) cut offs causes an underestimation of $\langle E^2 \rangle$ on I/P to PHASE.

The anomalous average error E" cannot be estimated in the same manner as the isomorphous E since no comparable method to centric refinement is possible. North (1965) suggests that an initial approximation can be made putting E" = E/3 since the error for the anomalous measurements should be relatively lower, as these do not involve such systematic errors as those incurred by crystal to crystal scaling.

Table 3.23 Final parameters used for phasing

	B_{ov}[a]	Site	Z[b]	Positional parameters x	y	z	B_i[c]	r.m.s. E[d] centric	all	r.m.s. F_H	R_C	wR[e]
$K_2Pt(CN)_4$	-3.5	A	54 -	0.3330	0.0875	0.4453	15	277	200	314	62.0	34.0
		B	41	0.0310	0.1153	0.2368	15					
		C	48	0.1276	0.2065	0.5036	15					
$KAu(CN)_2$	-3.5	1	78	0.2293	0.0792	0.2291	15	241	174	328	61.1	43.7
		2	39	0.4969	0.2228	0.1428	15					

[a] B_{ov} is the overall temperature factor to be applied to the derivative data for best fit with the native data.

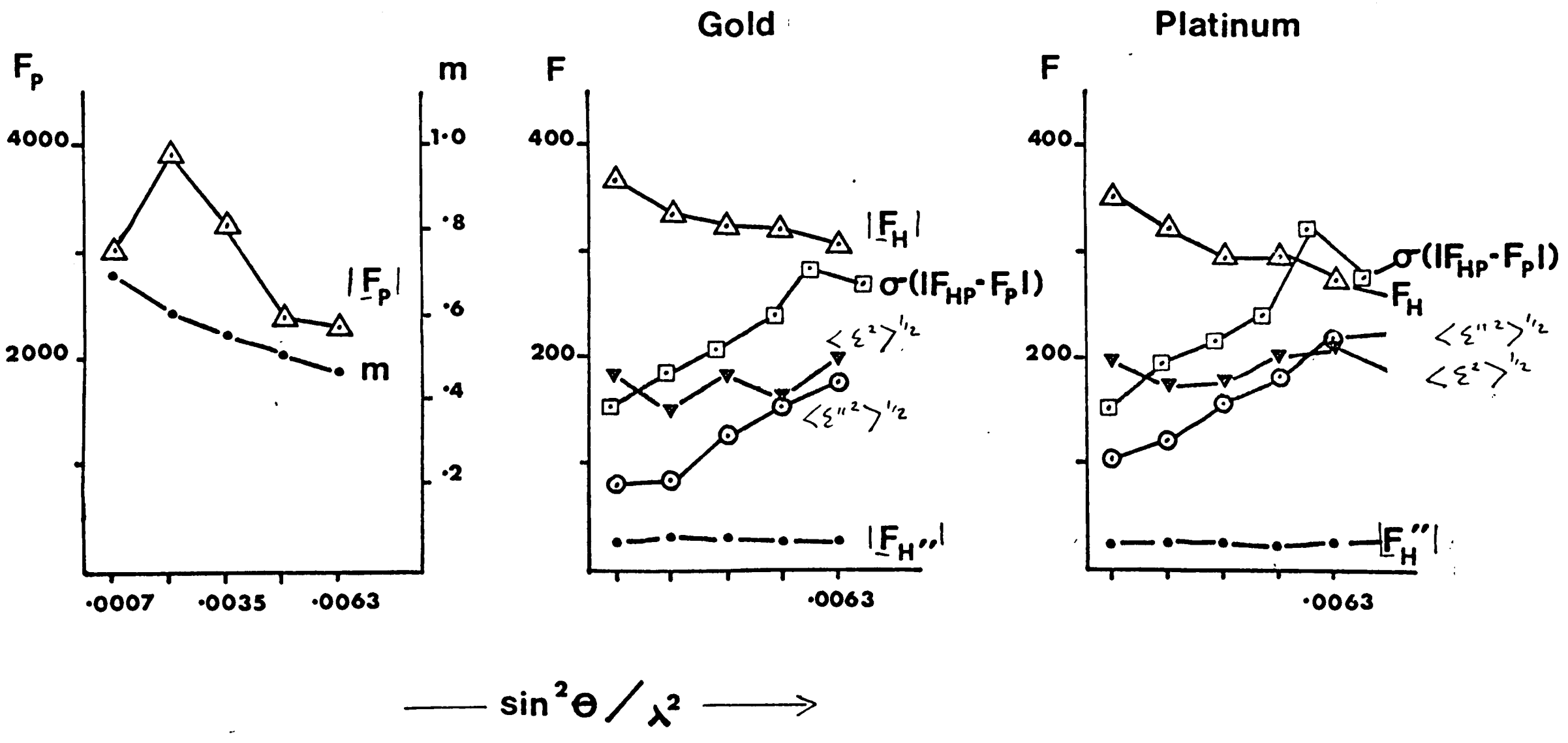
[b] Z is given in electrons, assuming full occupancy of site 1 of the Au derivative.

[c] B_i, the individual heavy atom temperature factors, were not refined.

[d] E is the lack of closure error, $|\underline{F}_{PH}| - |\underline{F}_H + \underline{F}_P|$.

[e] R_C and wR are defined in the text (and are given for centric terms only). These are not directly comparable with those of Table 3.19 since different numbers of reflections were used, and, for the Au derivative, the weights have been modified.

Figure 3.18 Variation of $|\underline{F}_p|$ and m with $\sin^2\theta/\lambda^2$ and of calculated $|\underline{F}_H|$, r.m.s. lack of closure error E, calculated anomalous component and r.m.s. E" for each derivative with $\sin^2\theta/\lambda^2$.



The initial estimates of E and E" were used in a primary phase calculation yielding phase set $(Au+Pt)_1$ and the mean values in ranges of $4\sin^2\theta/\lambda^2$ of these parameters for the best phase, output by the PHASE program were used in a second calculation yielding phase set $(Au+Pt)_2$. Phases were calculated as well for the single derivatives using the anomalous scattering information (and E" = E/3). These phase sets and an additional one calculated for the Pt derivative on the wrong hand (equivalent to the signs of the Pt anomalous differences being reversed) are compared in Table 3.24). The phase sets $(Au+Pt)_1$ and $(Au+Pt)_2$ can be seen to agree closely whereas the phases (on the correct hand) for the Au and Pt phased separately can be seen to agree less well.

The quality of the phase set $(Au+Pt)_2$ is indicated in Figure 3.18 which shows the variation of m and $|\underline{F}_P|$ with $\sin^2\theta/\lambda^2$ as well as $|\underline{F}_H|$, $\langle\varepsilon^2\rangle^{\frac{1}{2}}$, $\sigma(\Delta_{ISO})$, $|\underline{F}_H''|$ and $\langle\varepsilon''^2\rangle^{\frac{1}{2}}$ with $\sin^2\theta/\lambda^2$. Figure 3.19 shows a plot of m versus $|\underline{F}_P|$ for $(Au+Pt)_2$ indicating that the weaker reflections, constituting a large fraction of the data set, are less well phased than the stronger ones and must account to some extent for the rather low figure of merit of $(Au+Pt)_2$.

3.7. The electron density map

The phase set $(Au+Pt)_2$ was used to calculated a native electron density map, weighted by m and contoured on the 1906A graph plotter using the program CONTOUR (see Appendix IV). The contour level and interval were chosen empirically to yield columns of density of suitable cross section. The corresponding values for the electron density in electrons Å^{-3} and root mean square error were calculated subsequently. The former from the equation:

$$\rho(r) = \frac{(K\rho(\text{calc})\ V'' + F_{000})}{V} \qquad 3.2$$

Table 3.24 Comparison of various phase sets

Phase set		$(Au+Pt)_1$[a]	$(Au+Pt)_2$[b]	Au 'correct hand'[c]	Pt 'correct hand'[c]	Pt 'wrong hand'[d]
	$\langle m \rangle$[e]	0.69	0.55	0.52	0.52	0.52
$(Au+Pt)_1$	acentric[f]		26.6°			
	centric[g]		99.4%			
$(Au+Pt)_2$	acentric			40.0°	51.0°	
	centric			83%	77.4%	
Au 'correct hand'	acentric				76°	
	centric				60%	
Pt 'correct hand'	acentric					89.9°
	centric					89.9%[h]

a Phase set $(Au+Pt)_1$ has E estimated from the centric refinement and E" = 1/3E. In fact E = 227 and 229 respectively for the Au and Pt derivatives as estimated from the centric refinements.

b Phase set $(Au+Pt)_2$ has E and E" (see Table 3.23) set equal to the lack of closure errors of the phasing $(Au+Pt)_1$.

c Phase sets (Au 'correct hand') and (Pt 'correct hand') have E estimated from the centric refinement and E" = 1/3E and use single isomorphous replacement and anomalous scattering with the correct hand.

d Phase set (Pt 'wrong hand') is similar except that it has the wrong hand of the heavy atom solution.

e $\langle m \rangle$ is the mean figure of merit.

f 'Acentric' refers to the mean difference in phase angle between the acentric reflections of the phase sets concerned.

g 'Centric' refers to the percent of centric reflections with the same sign in the phase sets concerned.

h The centric reflections with changed phases are those with phase $\pm\pi/2$.

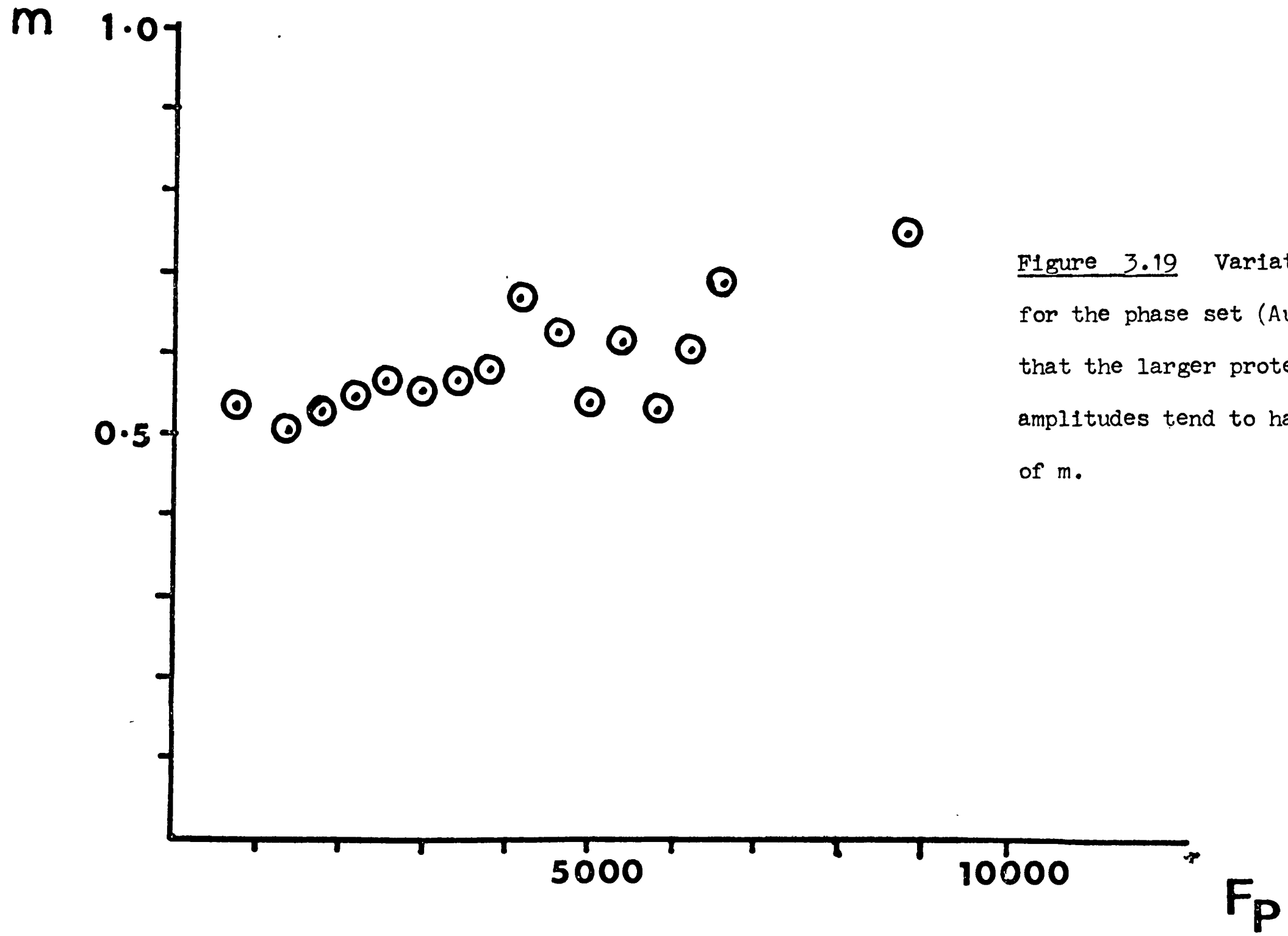


Figure 3.19 Variation of m with $|F_p|$ for the phase set $(Au+Pt)_2$ indicating that the larger protein structure amplitudes tend to have a larger value of m.

where ρ (r) is in e Å^{-3}; V" is the empirical volume used in the fast Fourier transform program FFT (see Appendix IV) for the unit cell; ρ (calc) is the electron density calculated in FFT; V is the volume of the unit cell; K is the scale required to put the $|\underline{F}_p|$ onto an absolute scale; $F_{000} = \sum_j f_j(o) = \frac{400000}{14} \times 7$ assuming that 6GPDH is made up of nitrogen atoms and with m.w. in the unit cell of 8 x 50000. The absolute scale can be calculated using a Wilson plot. However, at 6 Å resolution the shape is such that an inaccurate estimate of this scale will be made.* Instead, an absolute scale was established by assuming that the occupancy of Au1 was 100%. In the refinement the scale between FHOBS and FHCALC was such that the occupancy of the gold major site was 0.7551. If this occupancy is now assumed to be 100% a further scale of 1/occ(Au1) needs to be applied where occ (Au1) = 0.7551.

The electron density map calculated used a V" = 365.0 and had a maximum peak of 1593 units and was contoured in intervals of 250 from 250.

The 250 level corresponds to a ρ (r) given by:

$$\rho(r) = \frac{((1/0.7551) \times 250 \times 365 + 200000)}{1.11 \times 10^6}$$

$$= 0.2891 \text{ e } \text{Å}^{-3}$$

The maximum peak corresponds to a ρ (r) of 0.8739 e Å^{-3}.

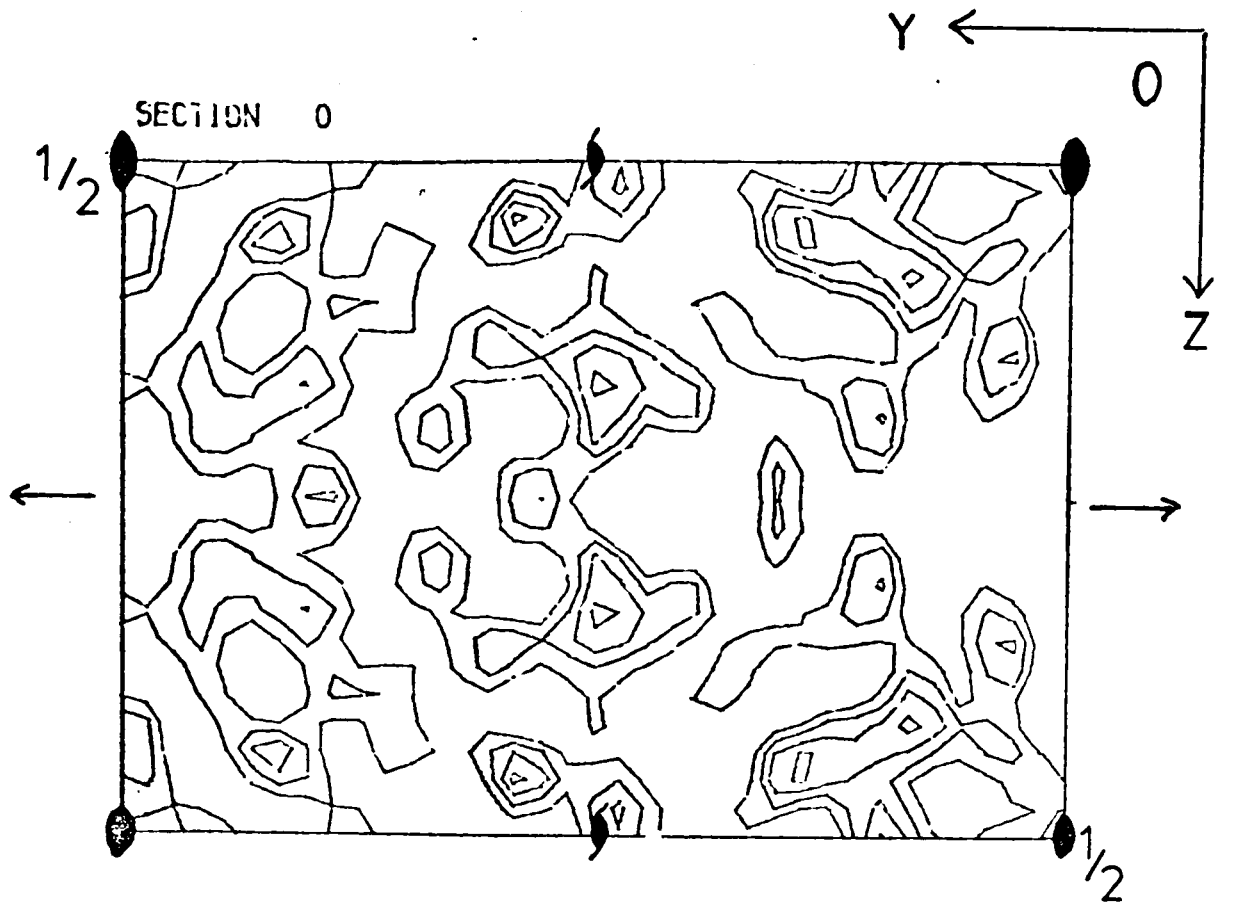
The r.m.s. error in the electron density $\langle \Delta \rho^2 \rangle^{\frac{1}{2}}$ evaluated using equation 3.1, was 0.1556 e Å^{-3}.

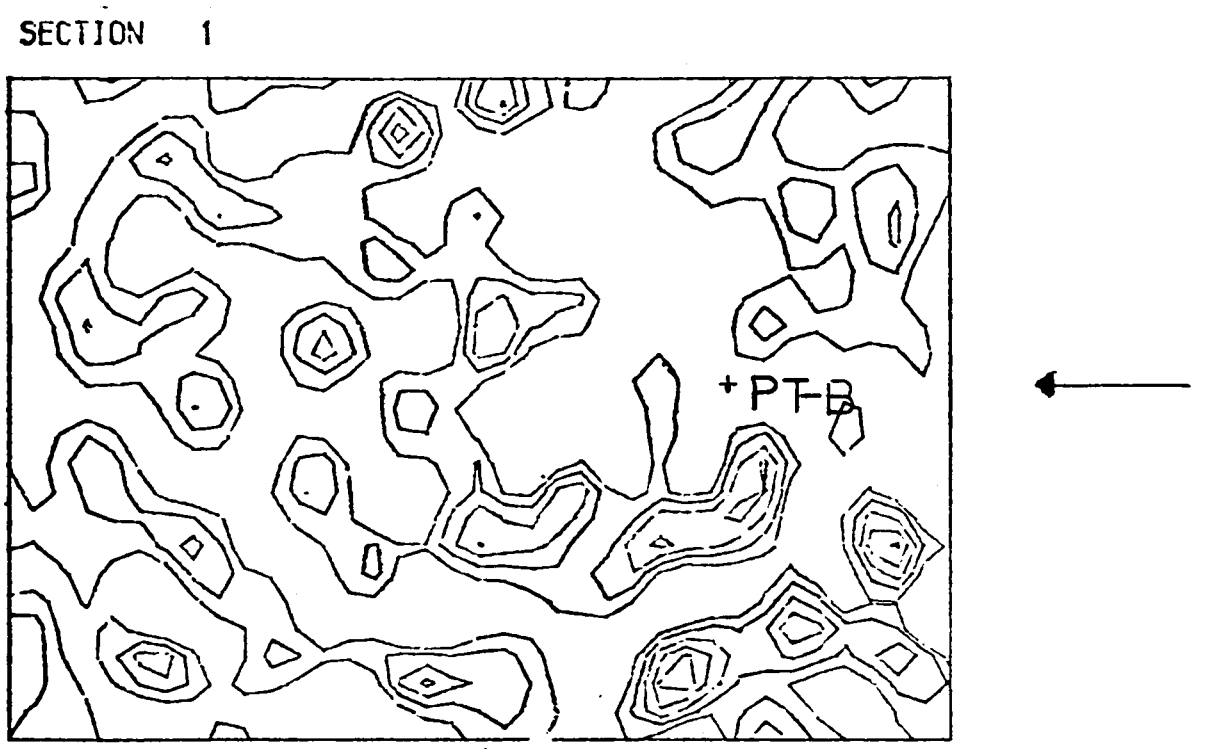
Figure 3.20 shows half the unit cell containing the protein , the electron density having been calculated at 2 Å intervals and contoured in X sections on a scale of 8 Å/ cm. The second contour level (250) was taken as protein. A similar map calculated on a scale of 4 Å/cm was used to build a balsa wood model which is illustrated in Figure 3.21.

* see Fig. 5.23.

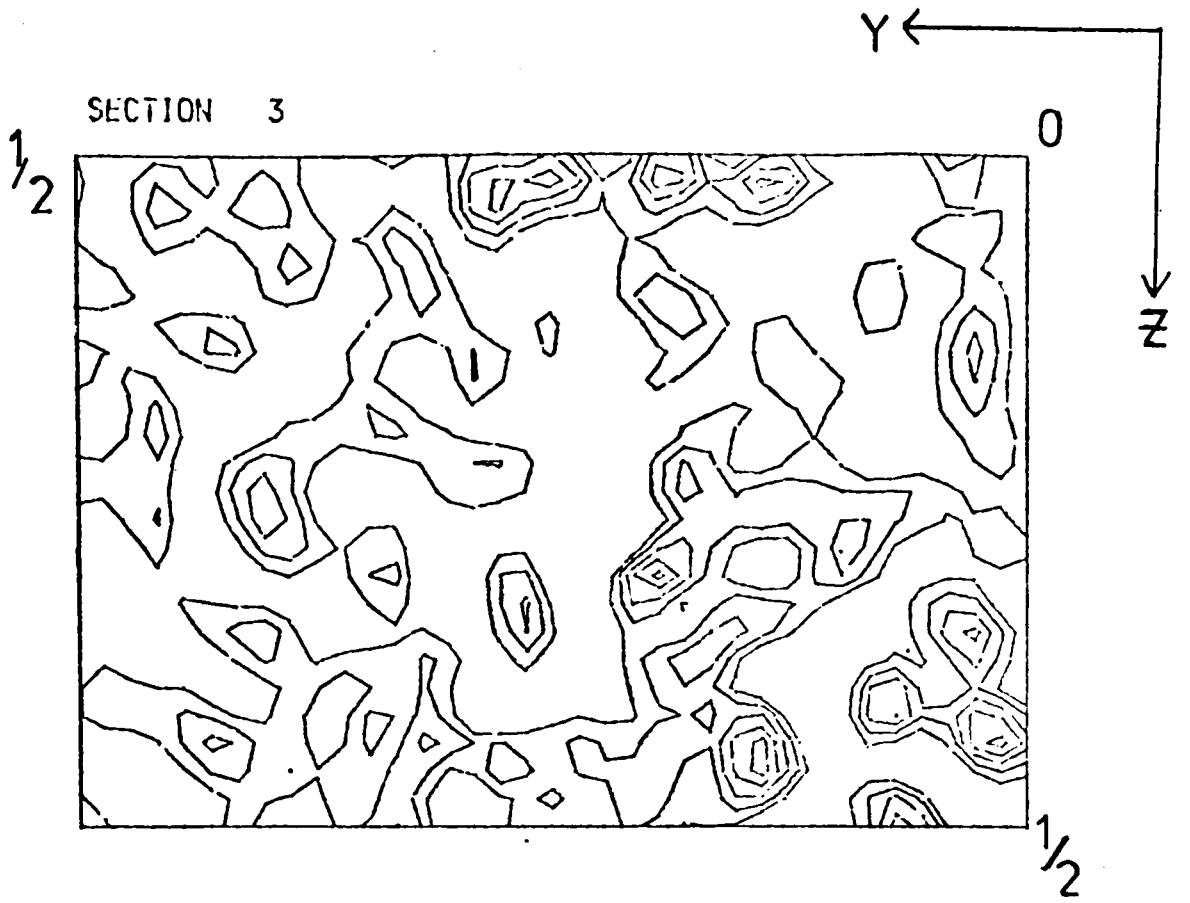
Figure 3.20

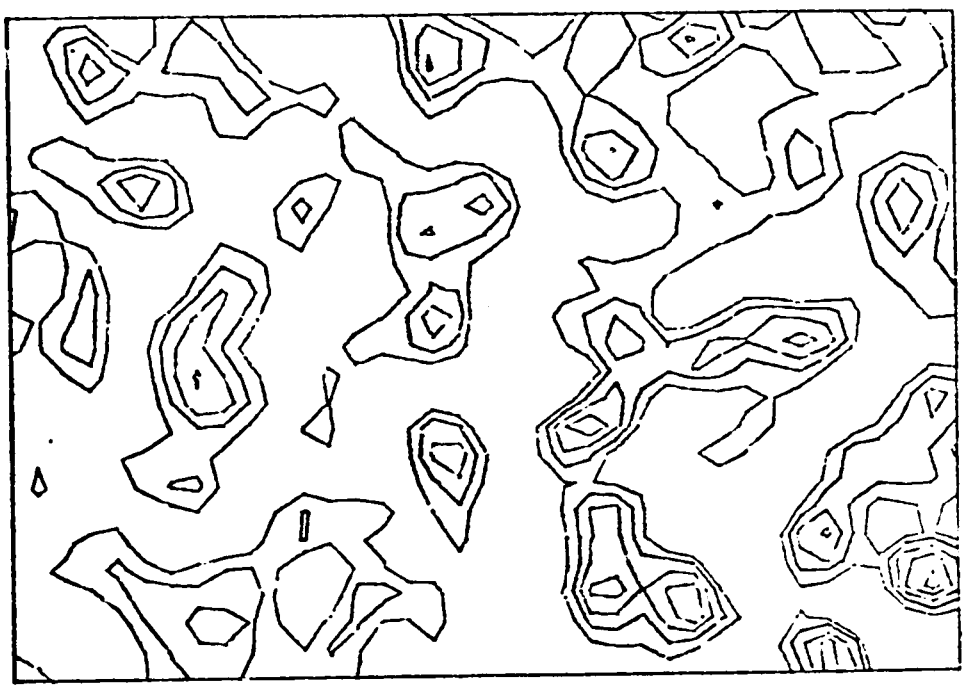
X sections running in 2 Å intervals from 0/40 to 21/40, with $0 < Y < \frac{1}{2}$ and $0 < Z < \frac{1}{2}$, showing the electron density map which was calculated at 2 Å intervals and each section contoured on a scale of 8 Å/cm. The positions of the heavy atom binding sites are marked, as is the strong column of density on section 13, the density which is not simply connected to either subunit on section 16 and a cut piece of density connecting both subunits of a dimer in section 17. An effect of the two-fold axis parallel to Y at $X = \frac{1}{2}$ can be seen in sections X = 19, 20 and 21. Those sections containing particularly interesting features are indicated by ←.

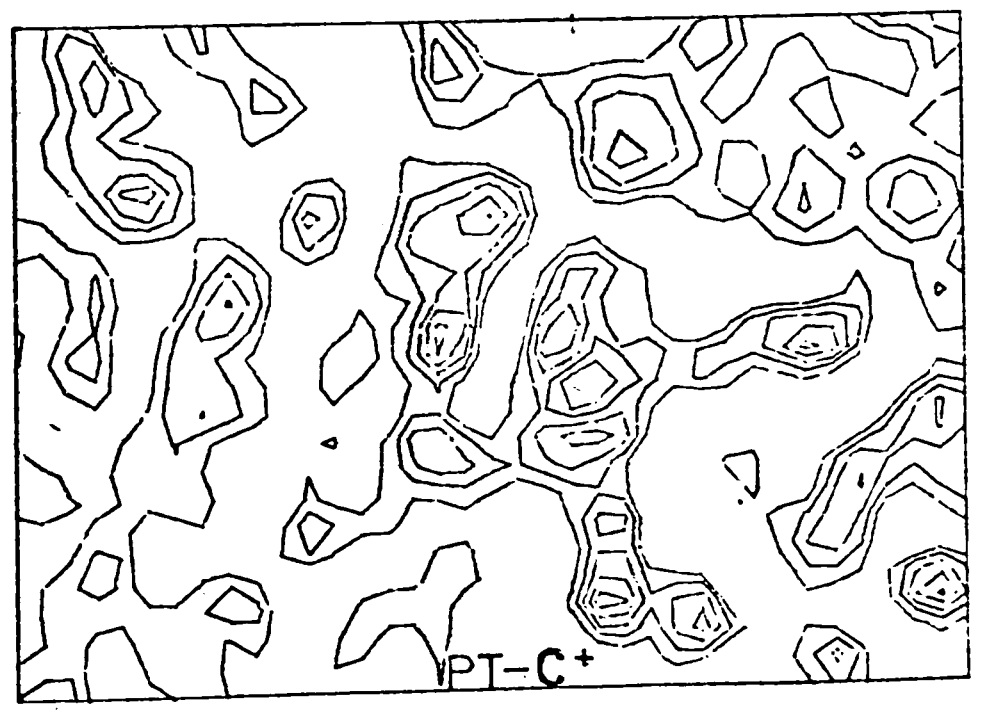

Y
0
Z
SECTION 0
1/2
1/2


SECTION 1
+PT-B

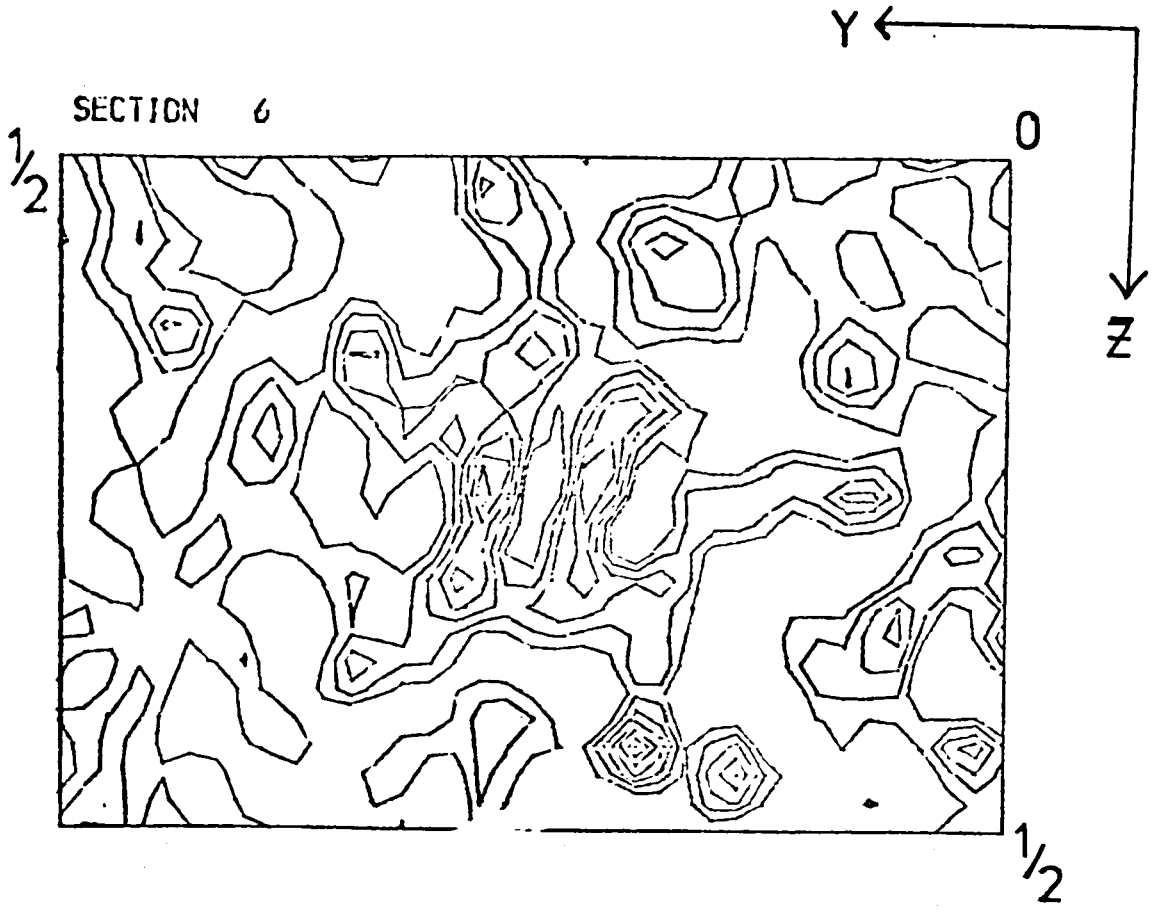

SECTION 2


Y
Z
SECTION 3
0
1/2
1/2

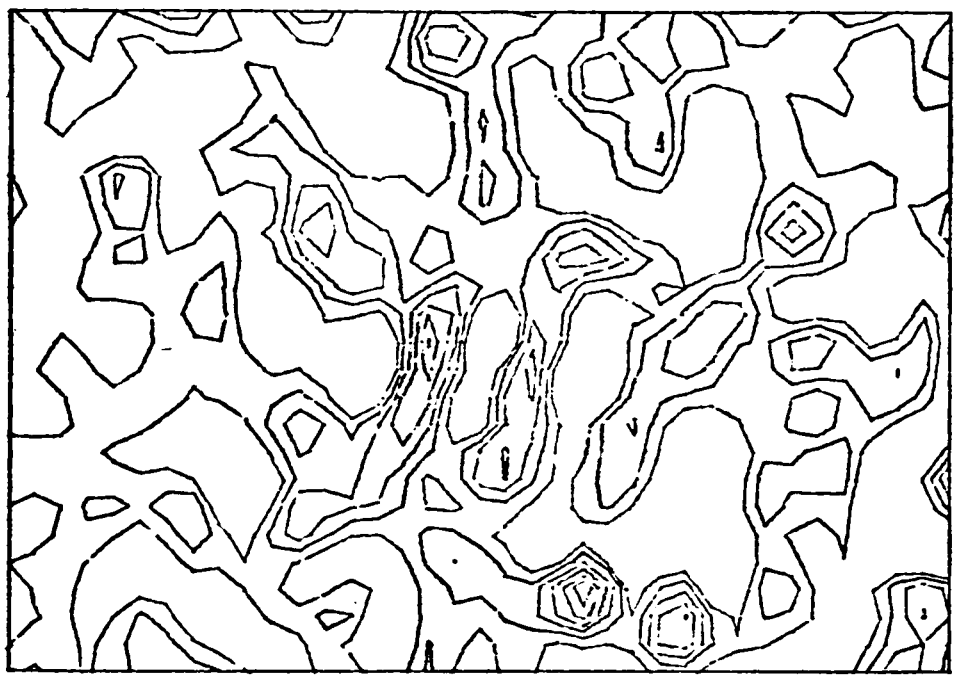

SECTION 4


SECTION 5
PT-C+

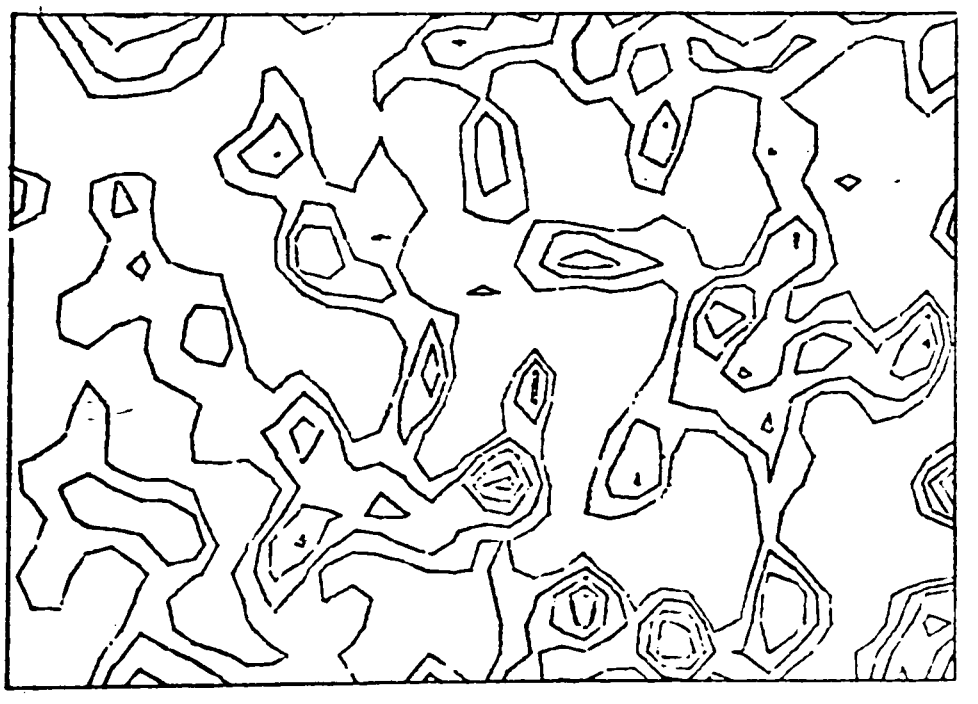
SECTION 6



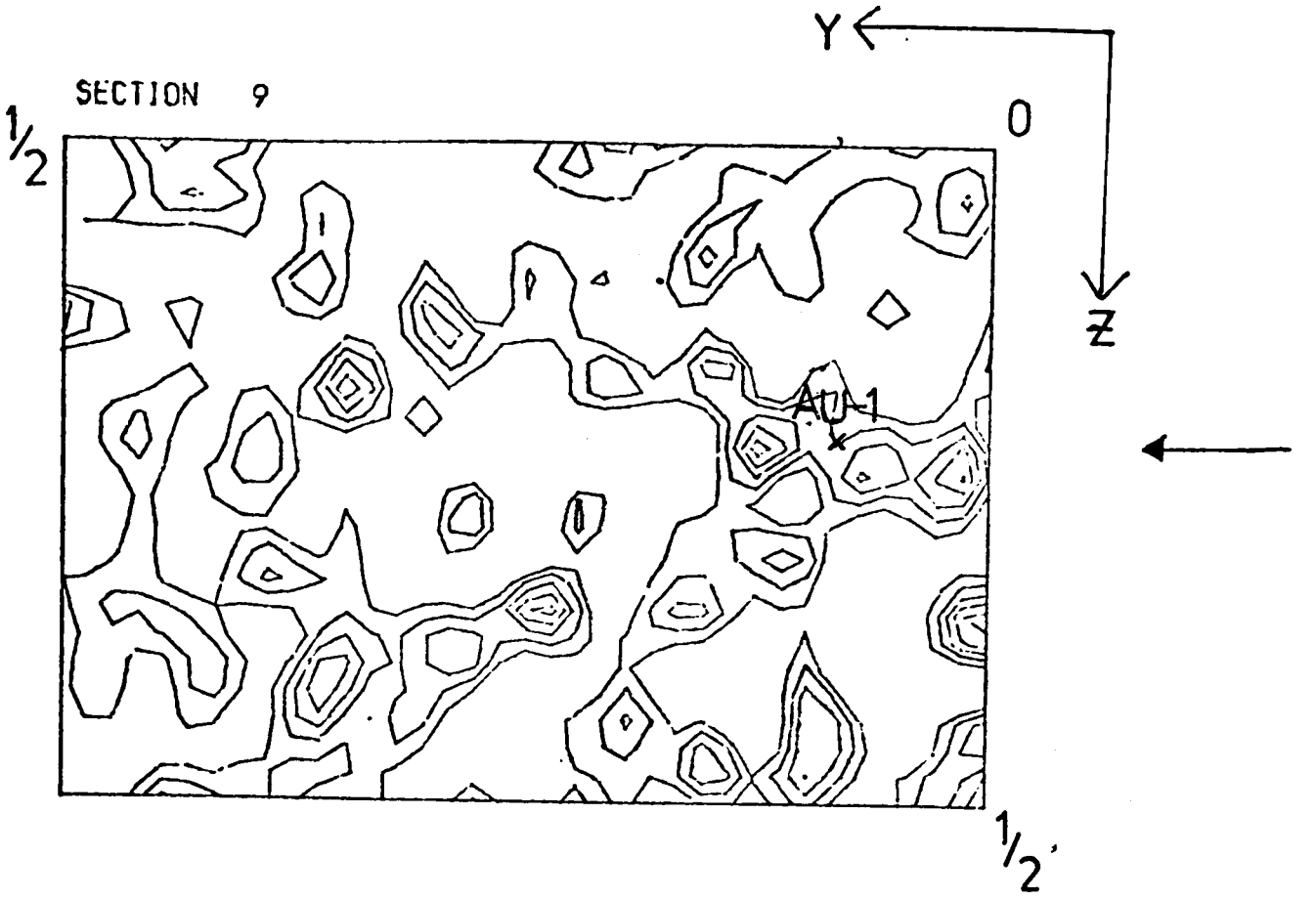
SECTION 7



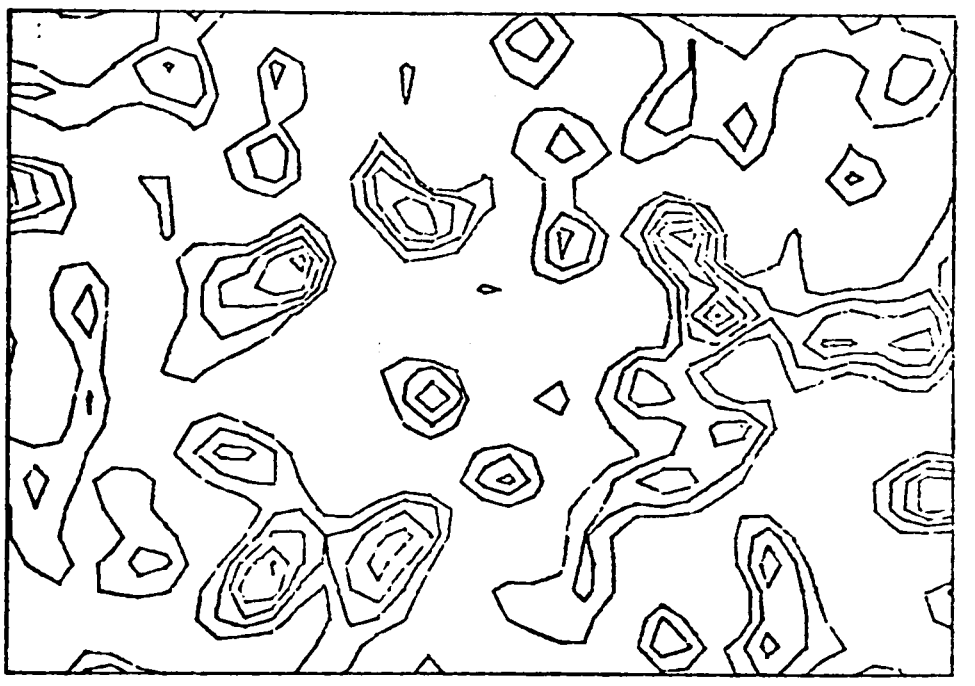
SECTION 8



SECTION 9



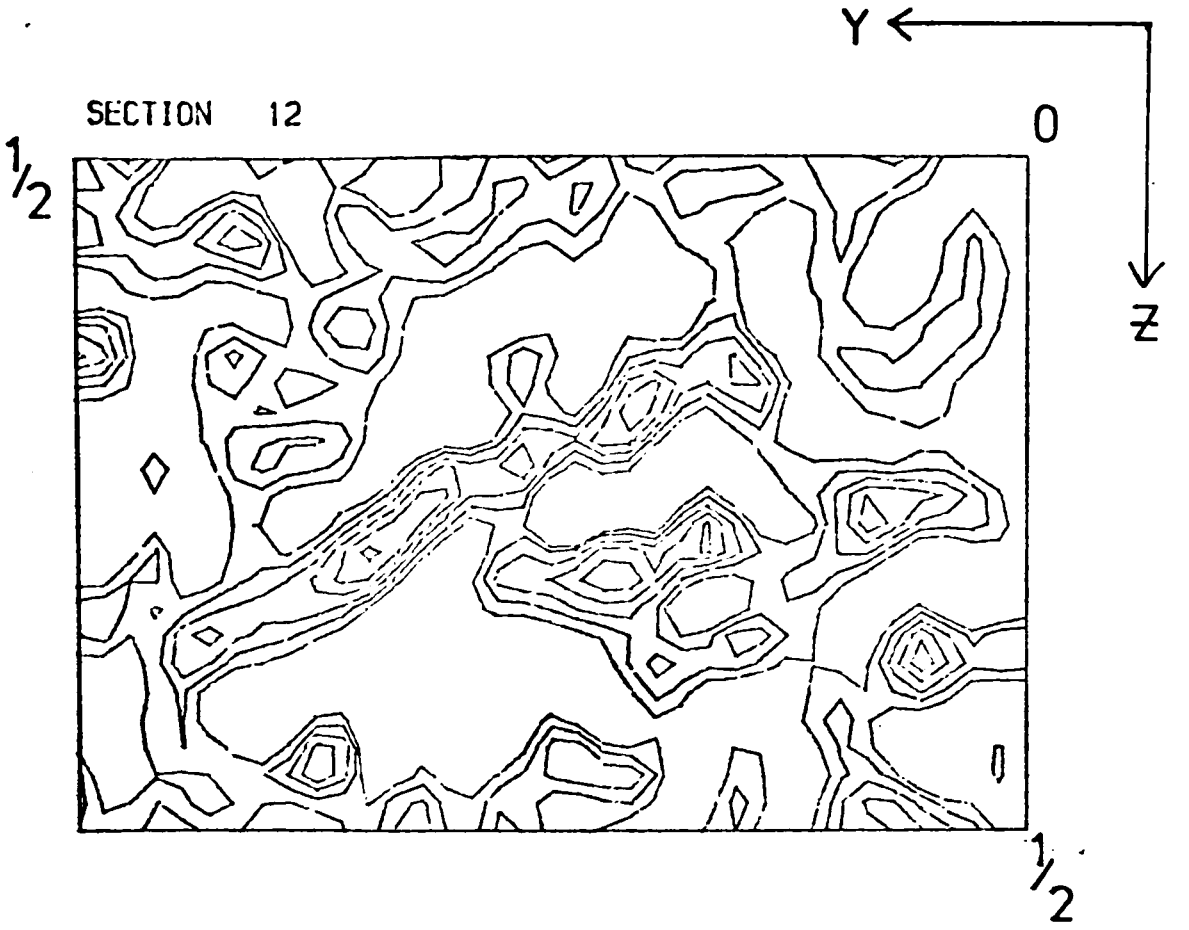
SECTION 10



SECTION 11



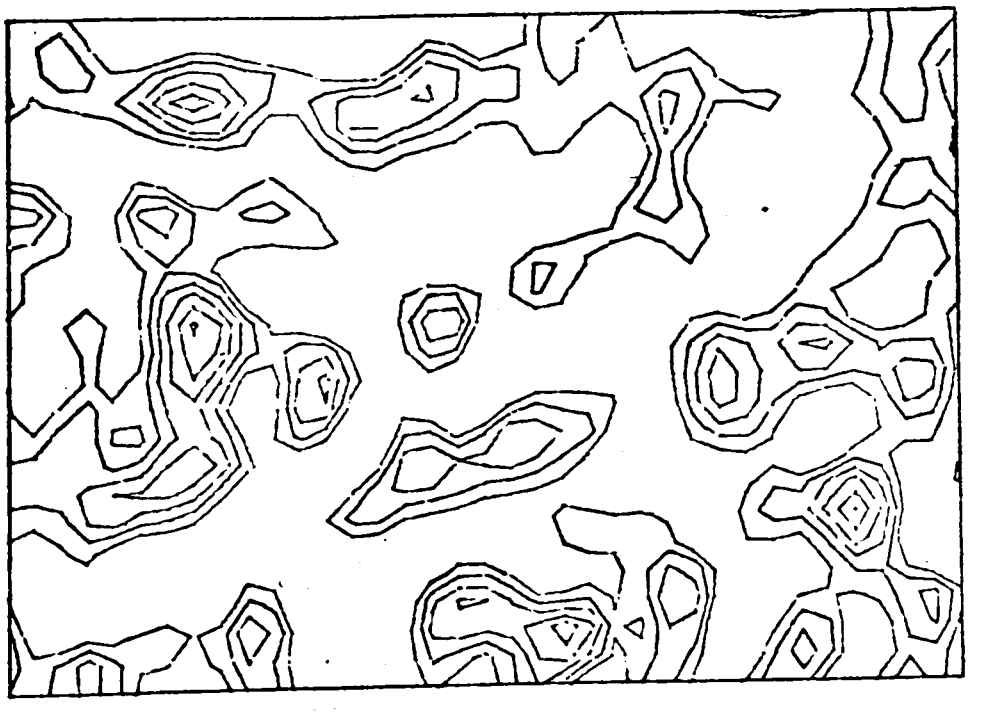
SECTION 12

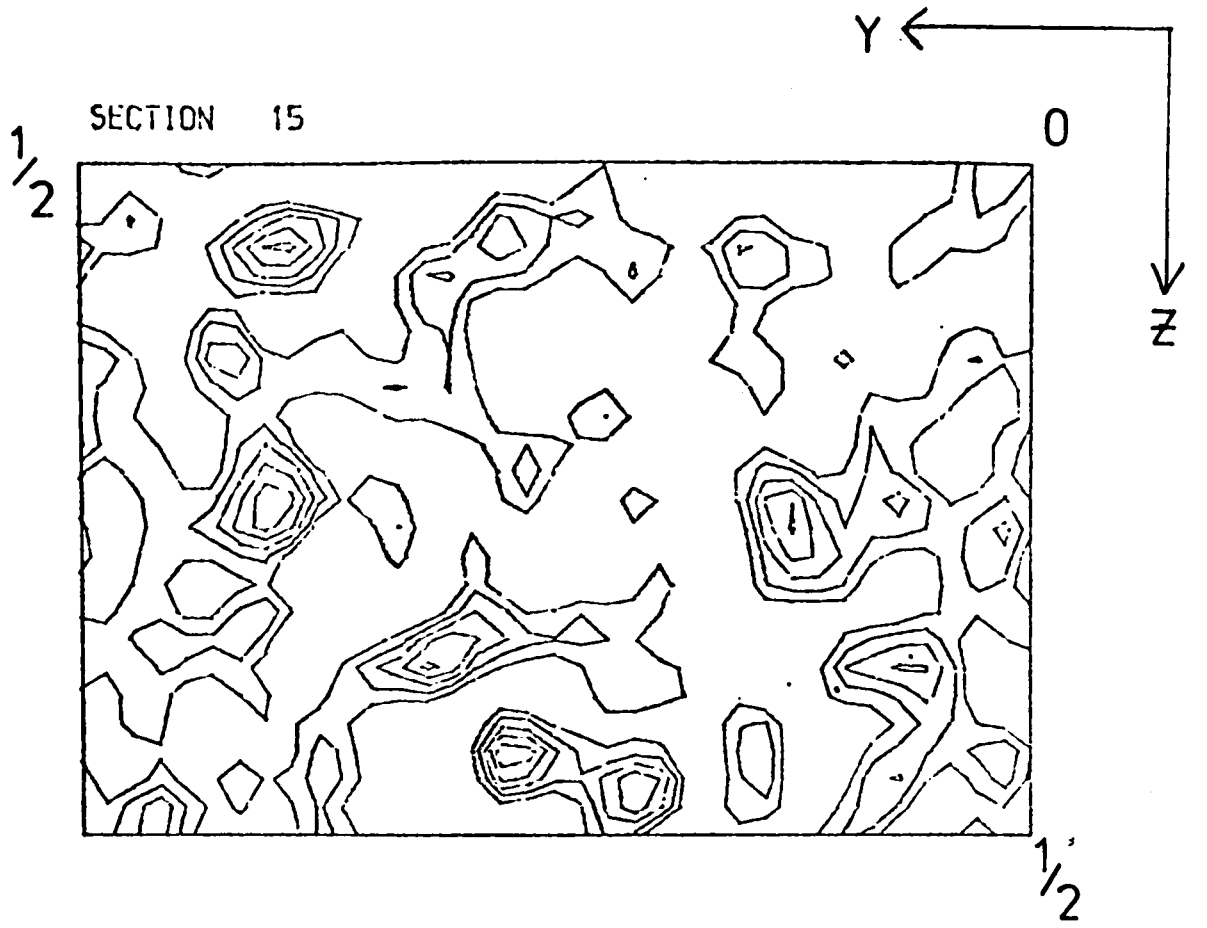


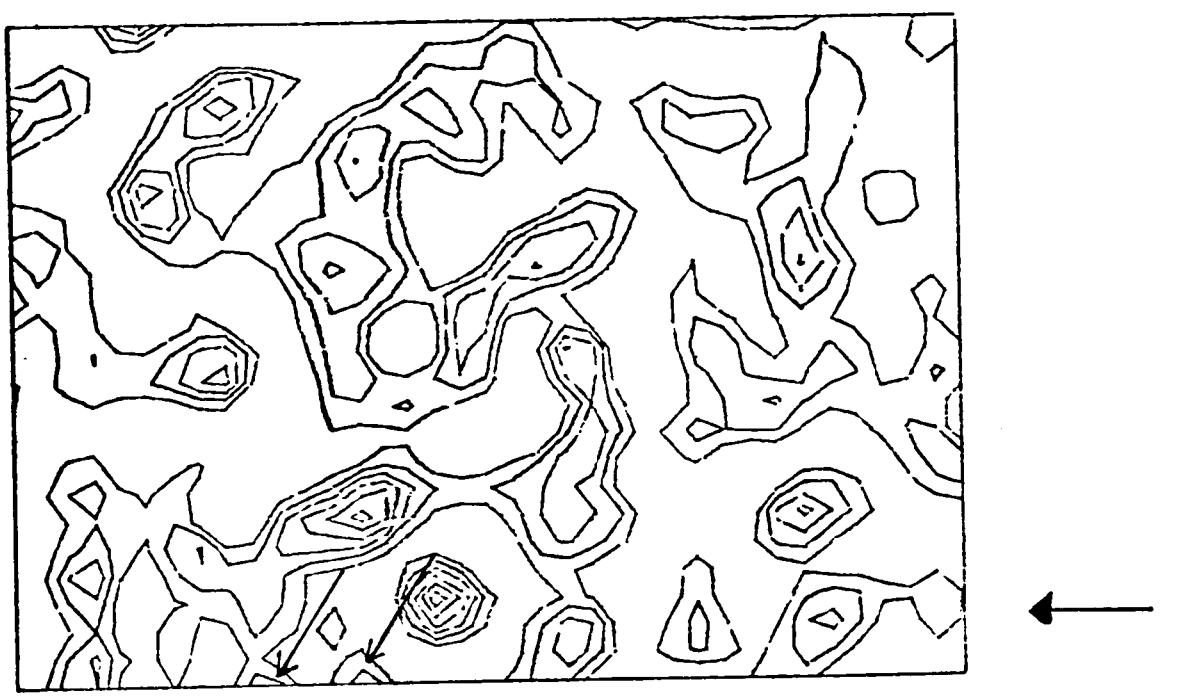
SECTION 13

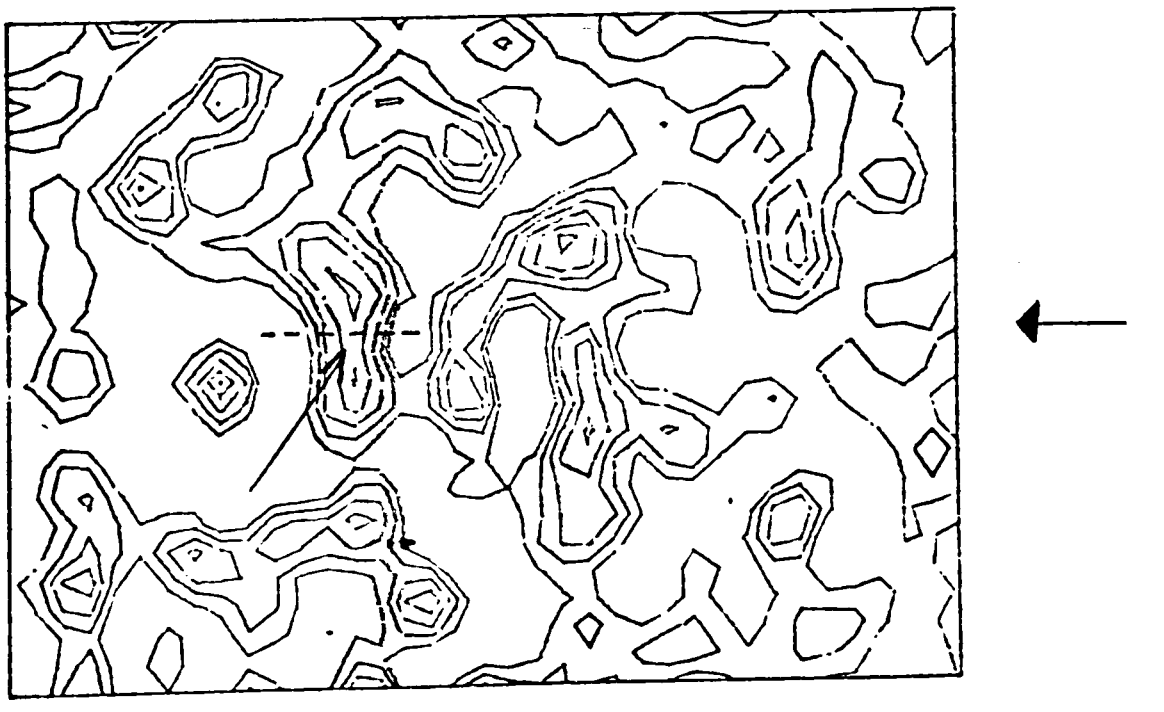


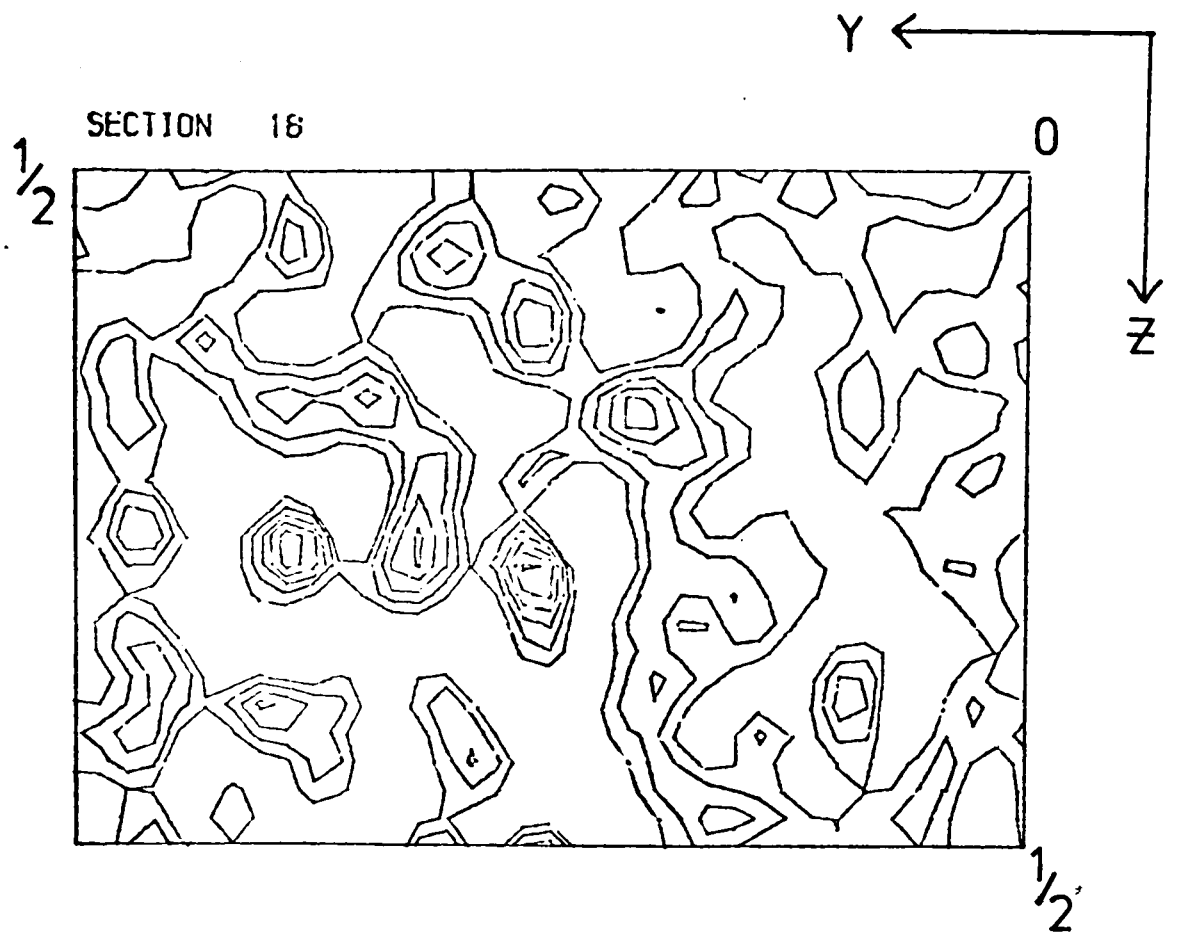
SECTION 14



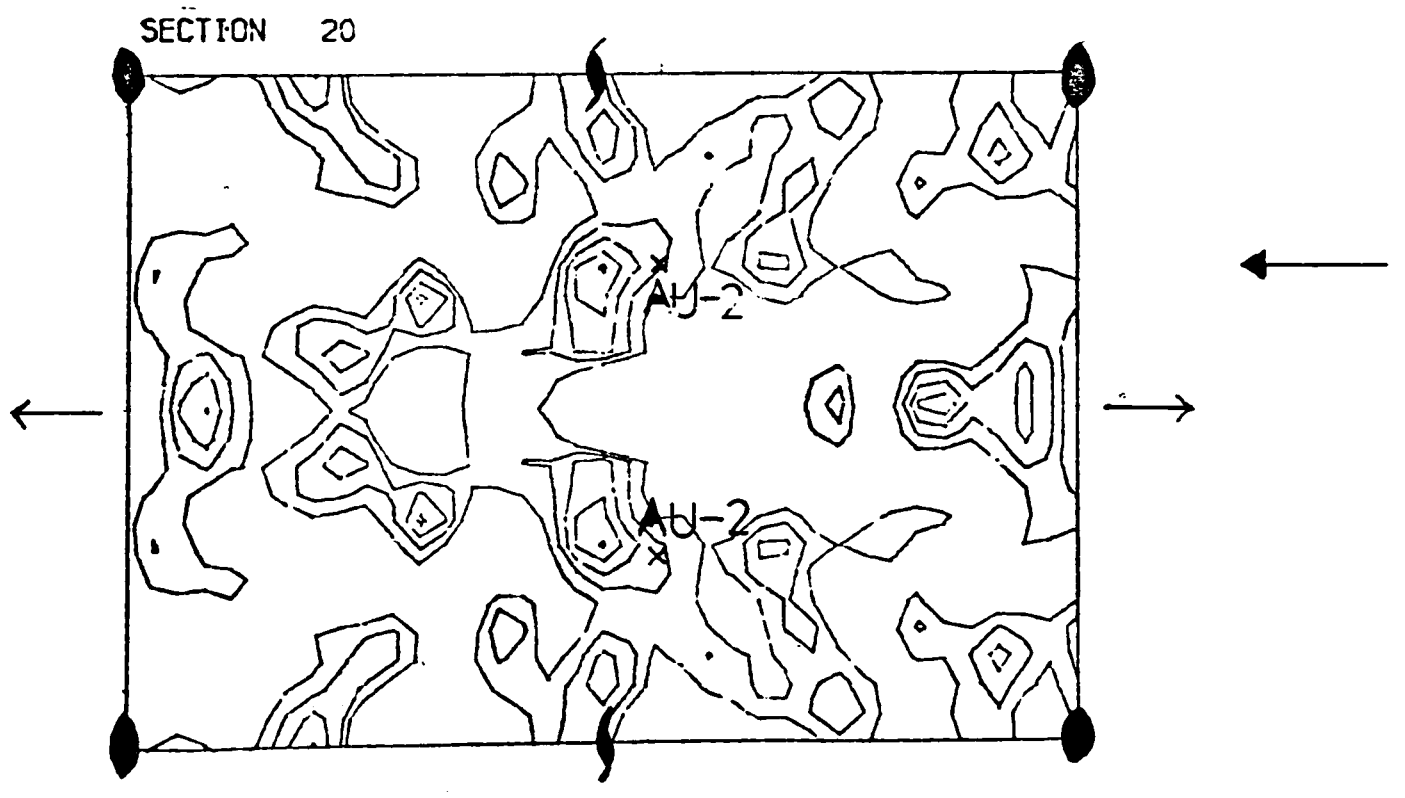

Y
Z
SECTION 15
0
1/2
1/2

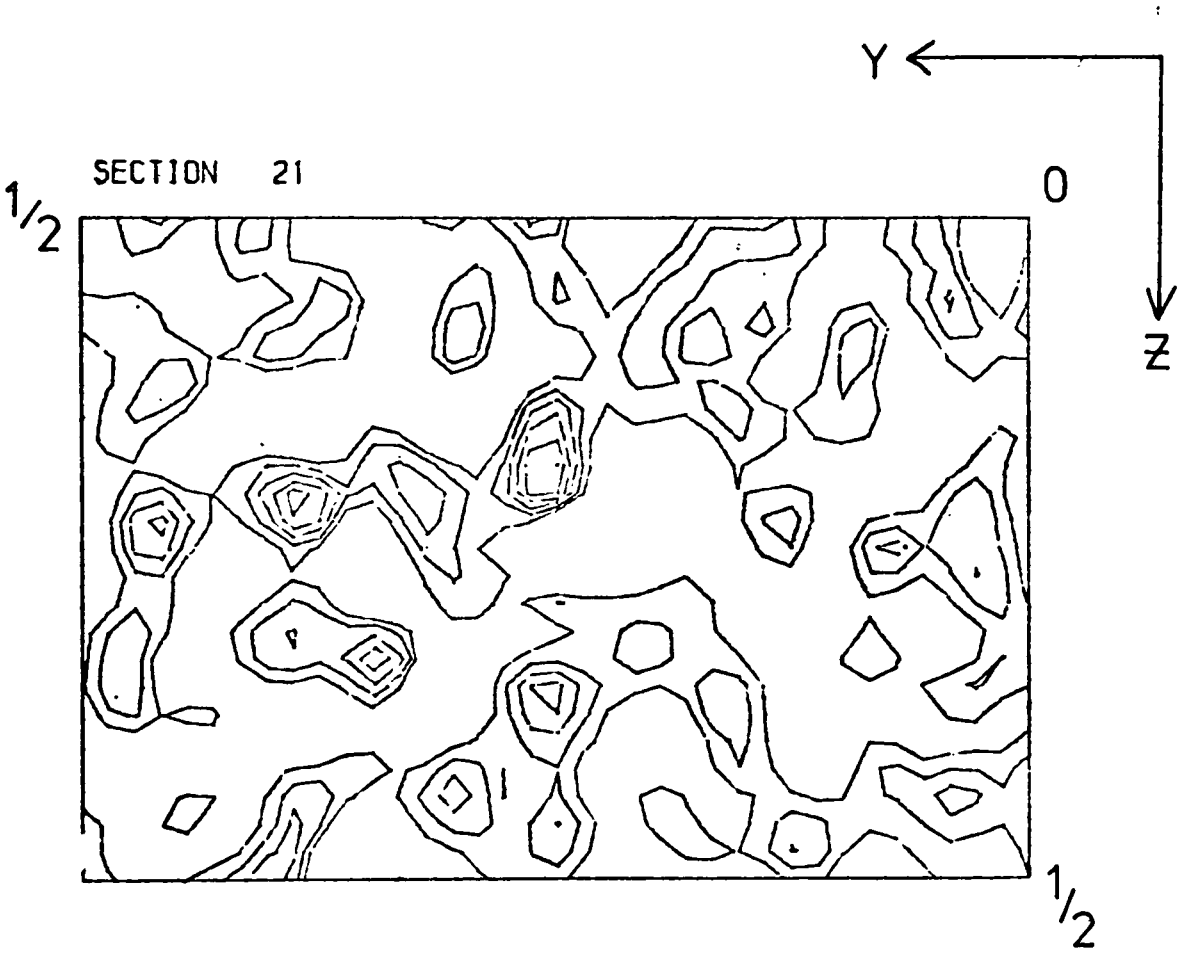

SECTION 16


SECTION 17


Y
Z
SECTION 18
0
1/2
1/2


SECTION 19


SECTION 20
AU-2
AU-2


Y
SECTION 21
0
1/2
Z
1/2

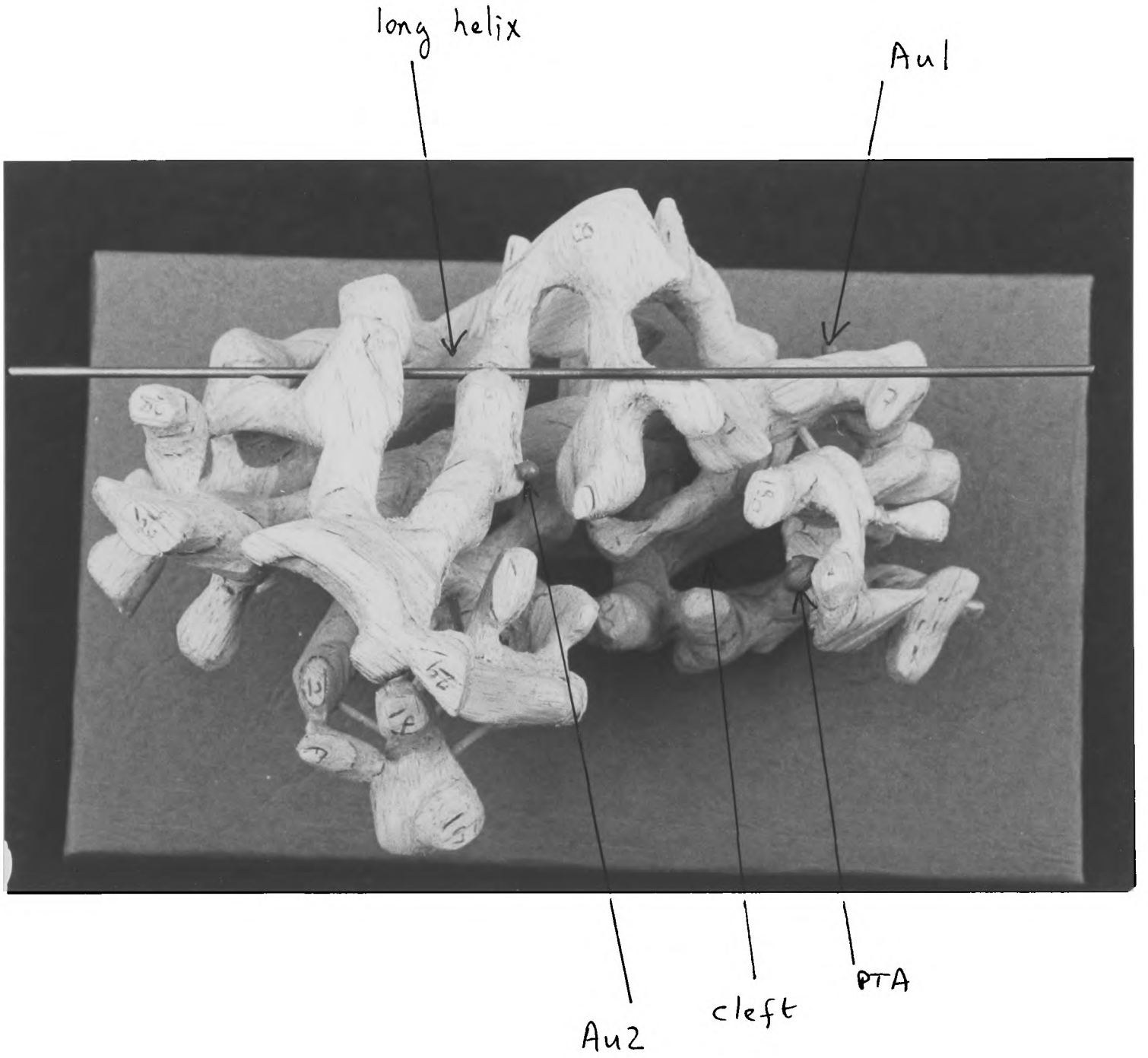


Figure 3.21 A view of the balsa wood model down a of . 6PGDH at 6 Å resolution. The molecular 2-fold axis is indicated by a metal rod. Gold sites (1,2) and platinum site (A) are arrowed. The cleft and the long helix are also indicated. For further views of the model see Figure 10 of Appendix III. Compare with Figure 4.8.

The space group $C222_1$ contains two 2-fold axes parallel to a and b and either might have been the dimer axis. The map shows that the choice of the axis parallel to b with the dimer centred at approximately 0, 3/4, 1/4 is the appropriate one, which is illustrated in Figure 3.22. This shows the section containing the two-fold axis and the two sections below it. There is very little difficulty in recognising the dimer boundary. The one ambiguity which exists in interpreting which density to associate with which subunit is a small piece of density that is not clearly connected and could be associated with the molecule in one of two ways (see section X = 16 in Figure 3.20). The molecular 2-fold axis is empty and the individual subunits can be identified, strong density is cut only twice in the contact area (see section X = 17 for eg. in Figure 3.20).

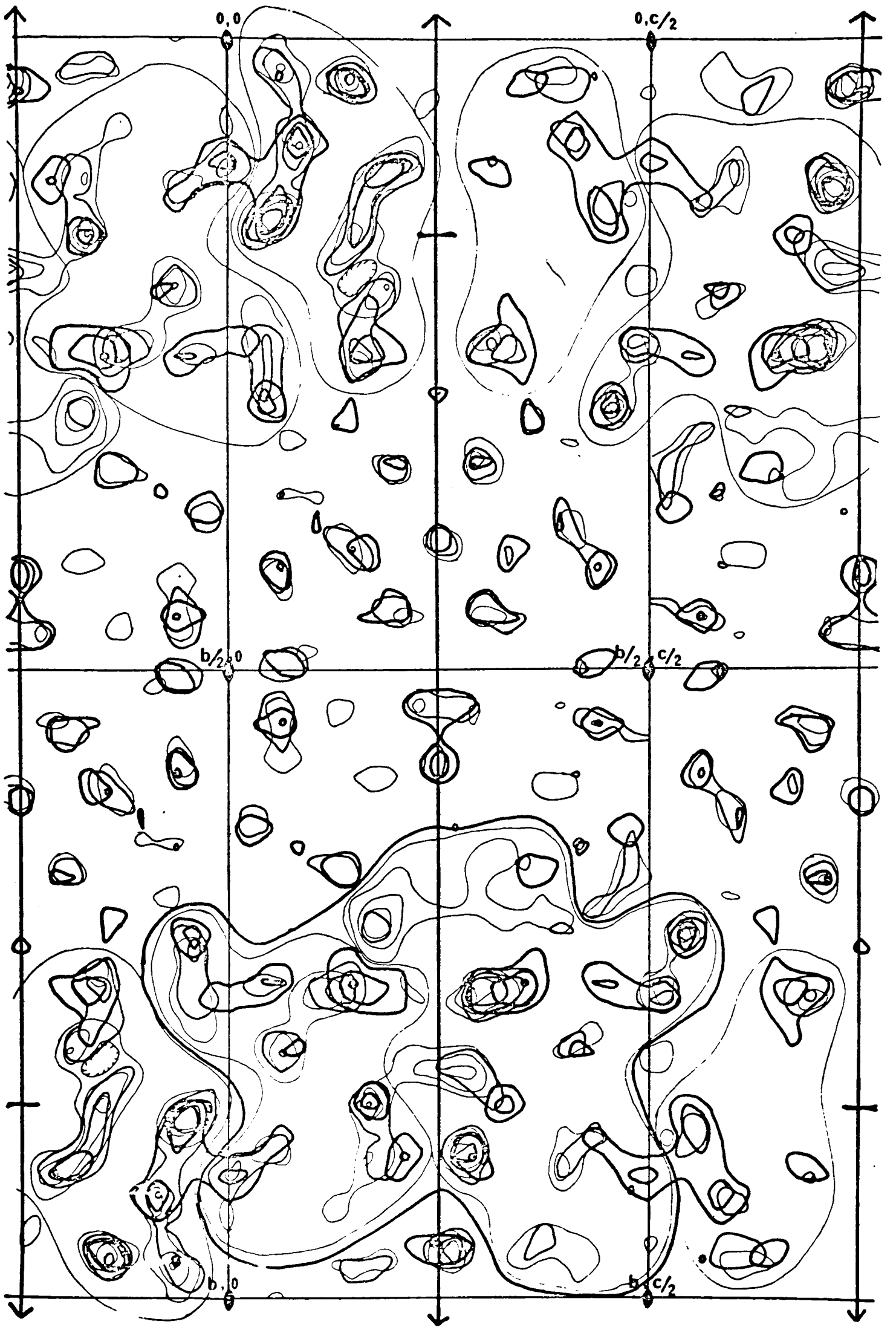
The balsa wood model shows the single subunit, the pins in the model representing the two Au sites and the three Pt sites, all of which are on or close to the surface of the dimer. The major site of Pt substitution can be seen to be at the edge of a rather large cleft region.

An interesting feature of the map is the presence of a number of long columns of density, some forming the backbone of the subunit and some at the subunit interface (these are indicated by arrows in the picture of the balsa model). These columns of density are of the correct dimensions to be helices and one of them is the strongest feature of the map (section X = 13).

In summary, the subunit is about 90 Å long and is ellipsoidal with the major axis of the ellipsoid in the ab plane, making an angle of 20° to the b axis. It extends through 60 Å parallel to a, 80 Å parallel to b, and 60 Å parallel to c. There is a noticeable indentation on the subunit distant from the 2 fold axis. The subunit does not appear to be bilobal at this resolution. Two columns of density run for more than 60 Å almost the entire length of the subunit and have cross sections compatible with

Figure 3.22

Sections of the 6 Å resolution electron density map from phase set $(Au+Pt)_2$ close to the molecular 2-fold axis, the crystallographic axis parallel to b. The dimer is outlined with a bold line and the individual subunits with faint lines.



Sectioned in a

x = 0/40 ——————

-1/40 ——————

-2/40 ——————

that for α-helix. The dimer extends over 90 Å in the a and c directions and 80 Å in b. The monomer-monomer contact area is a surface 25 Å by 50 Å (a by b) and is lined with some of the columns of density which may be helix.

In a low resolution study only a limited amount of structural information can be expected to be obtained.

Since approximately half a typical protein crystal is composed of solvent, the protein molecule should stand out as a high density region in sharp contrast to the essentially randomised solvent structure. The molecules should be clearly distinguished from one another apart from a limited number of contacts vital for the 3-D stability of the crystals where inter and intra molecular spacings are comparable. The binding site of the heavy atoms should lie on or close to the hydrophilic surface* (see Figure 3.21). Within the molecule there should be chains of density corresponding to the secondary structures and separated by the order of 10 Å. It will not be possible to follow the main chain continuously since the electron density will branch at various points because the data does not extend to a high enough resolution to resolve the chain uniquely. The course of the chain is usually most difficult to follow in regions of pleated sheet where at this resolution the electron density tends to extend across the sheet rather than along the chains which make up the sheet. It is interesting to note that the position of the 6PGDH molecule where there is a cleft contains density that is not simply connected unlike other areas of the molecule featuring well resolved columns of density.

The electron density map calculated for sheep liver 6PGDH at 6 Å has enabled the molecular boundary to be followed and has enabled the molecular 2 fold axis to be determined. It has also yielded such interesting features as the long columns of density, which if α-helix are longer than in any other protein structure yet published (followed by hexokinase , with 50 Å (Steitz et al. (1976)), and also the cleft region (see § 4).

* not always true eg HgI_3 in Mb.

CHAPTER 4

STUDIES OF COENZYME AND SUBSTRATE BINDING TO 6PGDH AT 6 Å RESOLUTION

The enzyme crystallizes in the absence of coenzyme or substrate and so attempts to locate the active site were centred on diffusion experiments involving the substrates and coenzymes.

The stability of the coenzyme is dependent on pH; the oxidized form ($NADP^+$) having to be kept in slightly acidic conditions and the reduced form (NADPH) in alkali (Lowry *et al.* (1961)). The crystals are stable through the pH range 6.0 to 8.5 and hence binding studies involving both $NADP^+$ and NADPH are possible.

Of the substrates, the reduced substrate 6 phosphogluconate (6PG) is far cheaper than ribulose-5-phosphate (R.5.P) and so likely binding studies involve 6PG. Studies of the binding to the enzyme, in the crystal, of CO_2 involved in the formation of R.5.P are not feasible.

The orientation of the substrate with respect to coenzyme can be obtained by studying the abortive ternary complex of enzyme, reduced or oxidized substrate and reduced or oxidized coenzyme, respectively, the situations where the enzyme does not catalyze a reaction but does bind these molecules. This gives two possible combinations 6PG/NADPH and R.5.P/$NADP^+$ for study.

It was decided that since the structure determination had been performed on crystals at pH 6.5 the binding of $NADP^+$ (stable at pH 6.5) would be studied first, followed by 6PG, NADPH and NADPH/6PG. An additional experiment was made possible by the gift of a sample of 3-iodopyridine adenine dinucleotide phosphate ($IADP^+$) by Dr. J.-F. Biellmann of the University of Strasbourg. This molecule has the amide replaced by an iodine atom on the pyridine ring of the nicotinamide moiety of the coenzyme.

The experiments described in this chapter include the taking of 12° precession photographs of crystals with bound $NADP^+$, $IADP^+$ or 6PG as well as the 6 Å data collection of the $NADP^+$ and $IADP^+$ bound crystals, details of which have been published (Adams, Archibald and Helliwell (1977), Appendix III). Details of the experiments involving the NADPH and NADPH/6PG soaked crystals are given in Clift (1977).

4.1. Soaking conditions

Suitable conditions for the separate binding of $NADP^+$, $IADP^+$ and 6PG to 6PGDH crystals were studied by monitoring the effect on the native h0l diffraction zone by taking 12° precession photographs. This involved performing soaking experiments in the buffer used for growing crystals, 65% ammonium sulphate, 50 mM potassium phosphate at pH 6.5.

The results of the soaking experiments are given in Table 4.1. At concentrations of $NADP^+$ comparable with its Michaelis constant (20 μM, Silverberg (1973)), no discernible effect on the diffraction pattern was obtained. However, at an increased concentration of 5 mM, reproducible intensity changes occurred some of which were similar to those for the platinum derivative (see Figure 3.5). At the time it was felt that increasing the soaking time from 1 day to 3 weeks enhanced the intensity changes, but in retrospect this seems difficult to assess. Crystals soaked in 5 mM NAD^+ for 1 day and for 3 weeks showed no intensity changes (Figure 4.1) indicating that NAD^+ does not bind to the enzyme crystals, in agreement with the fluorescence studies of Silverberg (1973).

Enzyme crystals transferred to solutions containing 5 mM $IADP^+$ showed reproducible changes which were maximal after 24 hours and similar to those given by $NADP^+$.

For 6PG small intensity changes were obtained for a solution

Table 4.1 Some coenzyme and substrate soaking experiments

Compound	M.W.	Concentration	Soaking time	Intensity changes	Comments
[a]$NADP^+$	833.4	200 μM	1 day	no (h01)	MJA
		5 mM	1 day	yes (")	401,402;JRH, MJA
		5 mM	3 wks	" (")	stronger changes? JRH, MJA
[b]NAD^+	735.5	5 mM	3 wks	no (")	MJA
[c]$IADP^+$	836.23	50 μM	2 wks	no (")	MJA
		5 mM	2 days	yes (")	401,402 again;MJA
		5 mM	2 wks	"	not stronger;MJA
[d]6PG	342.1	43 mM	1 wk	"	401,402;JRH

[a] Trisodium salt. The molecular weight given is for the anhydrous form.

[b] Hydrated free acid molecular weight.

[c] Anhydrous molecular weight of free acid.

[d] Trisodium salt. Anhydrous molecular weight.

The suppliers of these compounds at various times have been:

a: Boehringer or Sigma, b: Sigma or Koch light; c: Dr.J.F.Biellman, University of Strasbourg; d: Sigma.

containing 43 mM 6PG and soaking for 1 week. The intensity reversal of the 401 and 402 reflection observed for the $NADP^+$, $IADP^+$ and Pt is not as marked. These conditions have not been repeated since attention was primarily directed at the coenzyme binding studies where the intensity differences were larger.

Figure 4.1 shows h0l zones for 12° precession photographs for crystals with bound $NADP^+$, $IADP^+$, 6PG and $K_2Pt(CN)_4$ as well as a native photograph. Cell dimensions (|a| and |c|) measured from these photographs show that no cell dimension differs from the native by more than 1% in |a| or 0.5% in |c|.

For a 1 ml soak solution, a concentration of 5 mM $NADP^+$ means 10^7 molecules of $NADP^+$ per enzyme subunit in a crystal of volume $0.4 \times 0.2 \times 0.4$ mm^3.

4.2. Data collection and processing

The strategy adopted here towards data processing was different from that used in processing the native and heavy atom derivative data described in § 3, because of the problems associated with the ordinate analysis method of estimating intensities from a profile scan of a reflection on the diffractometer (see § 2, § 3).

One method which avoided the systematic overestimation of weak reflections was that of French (1975) which analyses the peak profile (see § 2).

French's method was used in these crystal binding studies and necessitates the re-collection of the native data to avoid any complications associated with the different error structures of ordinate analysis and peak profile fitted data sets.

Three dimensional sets of X-ray data were collected to 6 Å resolution for a native crystal and $NADP^+$ and $IADP^+$ bound crystals

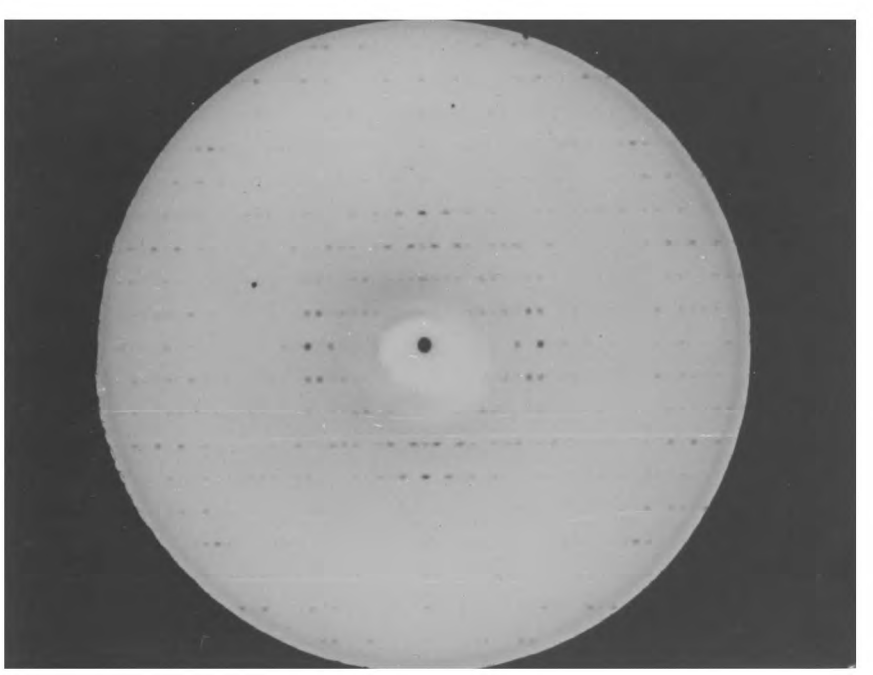


$\vec{c}^*$

4.1(a) $NADP^+$

5 mM 3 week soak

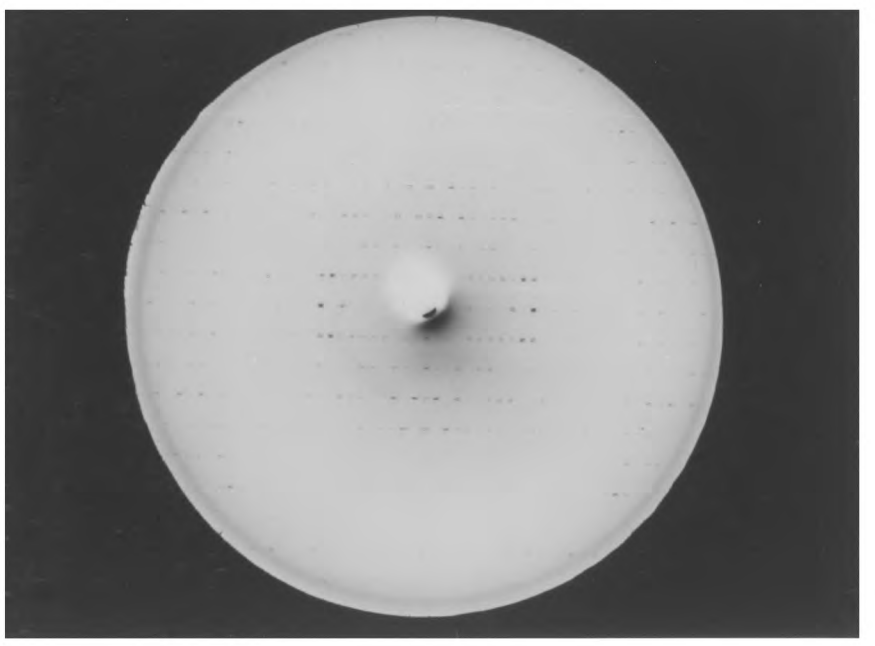
(Taken by JRH)



4.1(b) $IADP^+$

5 mM 2 day soak

(MJA)



4.1(c) 6PG

43 mM 7 days

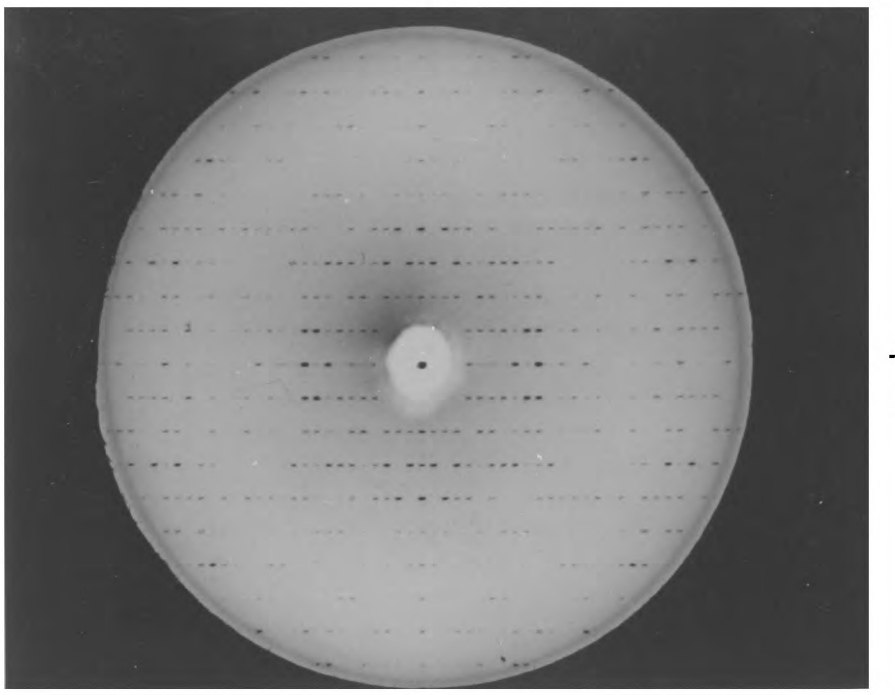
(JRH)

Figure 4.1 12° precession photographs of the h0l zone for (a) $NADP^+$; (b) $IADP^+$; (c) 6PG; (d) $K_2Pt(CN)_4$ bound and (e) an NAD^+ soaked and (f) native (N_V) crystals. Attention is drawn to the 401, 402 doublet. The soaking conditions corresponding to each photograph are shown.

4.1(d) Pt

2.7 mM 16 hrs

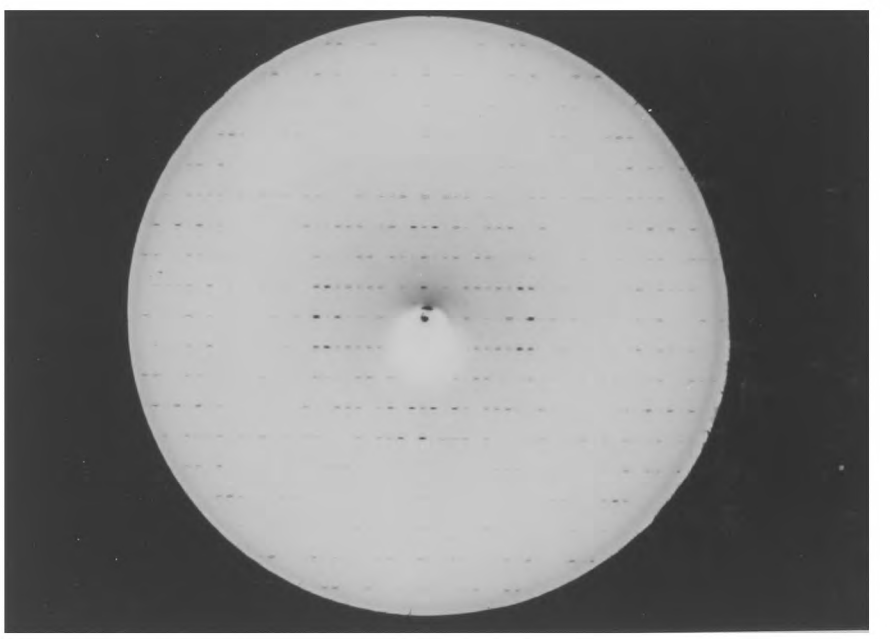
(MJA)



4.1(e) NAD^+

5 mM 3 weeks

(MJA)



4.1(f) N_V

(Taken by Susan Clift)

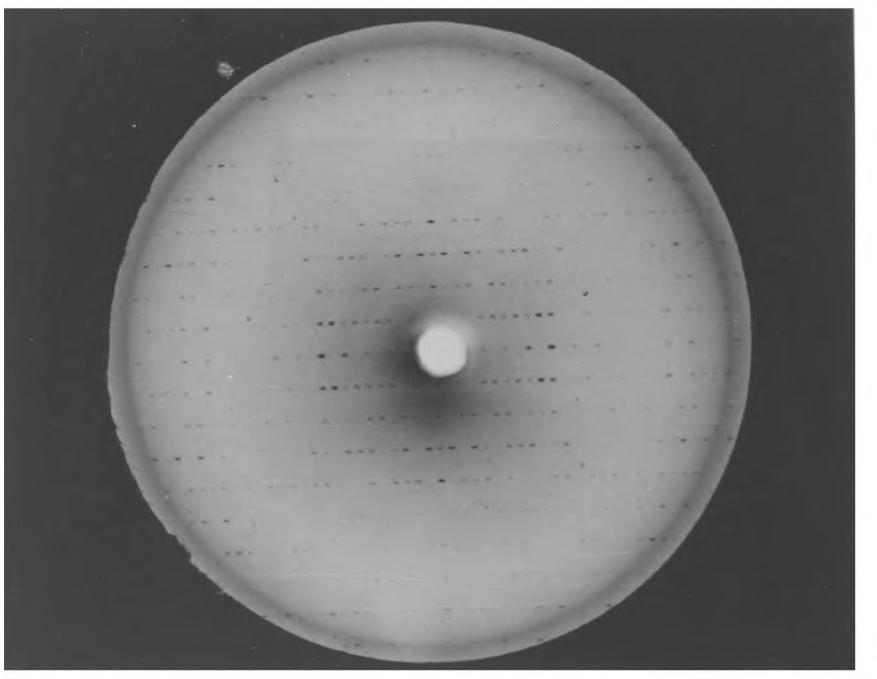


Table 4.2 Some parameters relevant to the data collection from N_V, $NADP^+$ and $IADP^+$ bound crystals.

Compare Table 3.2.

	Soaking conditions concn.(mM)	time(dys)	No. of steps in scan	ωstep increment ($\Delta\omega$) in°	Time for each $\Delta\omega$ in secs.	Diffracted beam collimator (mm)	Absorption I_{max}/I_{min}	Radiation damage (% in I)
N_V	-	-	12	0.05	6	2.5	4.14	11.2
$NADP^+$ shell	1	5	21	12	0.06	3.5	2.5	1.96	11.8
" "	2					4			
" "	3					6			
" "	4					4			
" "	5					6			
$IADP^+$ batch	1	5	2	12	0.04	6.5	3.5	2.00	19.1
"	2				0.05				

with a c axis mount using CuK_α radiation. Peak profiles were measured at room temperature on a Hilger and Watts Y290 4 circle diffractometer controlled by a Digital Equipment PDP8 computer with a Phillips fine focus X-ray tube operated at 40 kV 30 mA. The data were collected in layers of constant h for reflections hkl and $hk\bar{l}$, the general absences with indices (h+k) odd were not monitored. A complete hkl block of reflections was collected, before starting on the equivalent $hk\bar{l}$ block. The data were not collected in shells except for the $NADP^+$ bound crystals where more time was spent on the high angle reflections. However, the high angle shell was not collected first which it should have been if serious attempts were being made to obtain better quality high angle data. This would have prevented the radiation damage being as noticeable, since this effect predominates at high angles of diffraction.

The crystals were typically 0.4 x 0.2 x 0.4 mm^3 in volume, which was smaller than those used for the initial native and derivative data collection. An ω scan was used, the full scan width per reflection being determined by the angular range of reflection of the crystal and the necessity that the widest peak of the data set (typically a strong reflection, eg. 0 0 10) was approximately 2/3 of the full scan width. From the 3-D diffractometer data collection described in § 3, it was expected that a crystal would last approximately 100 hours and so the crystal lifetime was divided between each reflection appropriately. This determined the step increment in ω and the counting time for each one of these steps for the 3400 reflections to be measured, which included the standards. Radiation damage was monitored by remeasuring 3 strong reflections in orthogonal directions of reciprocal space (the 0 0 10, 0 16 0 and 8 0 0 reflections) each 50 reflections. Absorption curves were measured on the 0 0 10 reflection in 10° intervals of φ both before and after data collection to detect any change in the absorption of the crystal during data collection. Table 4.2 gives some parameters

relevant to the data collection of each data set. The substituted crystals were closely isomorphous with the native crystal as was indicated by the measurements from the precession photographs in Figure 4.1.

The computer program PRIMROSE (Appendix IV) was used to fit profiles to the measured step counts. This method has been fully described by French (1975) and summarized by Oatley (1977). Details necessary for understanding the data processing will be given here.

The i^{th} count, C_i in a scan through a reflection of N steps obeys Poisson statistics, i.e.

$$C_i \sim P(\lambda_i)$$

where λ_i is the mean of the i^{th} count and λ_i is given by

$$\lambda_i = J\pi(x_i) + b(x_i) + \eta_i \quad ; \qquad i = 1(1), N$$

where J = true intensity, $\pi(x_i)$ is the peak shape function, $b(x)$ the background function and η_i is a small random error known as the instrument instability error and is due to small variations in the strength of the incident beam, slight breakdowns in the counting chain, etc.

The background function is given by:

$$b(x, \beta) = \beta_1 + \beta_2 x$$

and the peak function by:

$$\pi(x, \underline{\alpha}) = \left(\frac{\alpha_4}{\alpha_2\sqrt{2\pi}}\right)\frac{\partial g.}{\partial y}\exp\left[\frac{-(\alpha_3+\alpha_4 g(y))^2}{2}\right]$$

where $y = \dfrac{(x - \alpha_1)}{\alpha_2}$

α_1 = mean of the peak position and α_2, α_3, α_4 are the second, third and fourth moments about the mean respectively.

$g(y) = \sinh^{\xi}(y)$.

The curve type, $\xi = -1$ gives peaks with flatter tops than curve type $\xi = +1$. The peak shapes from a single crystal fall into one or other class, so the curve type does not have to be chosen for each reflection. The choice for the entire data set may be made by examination of the first few peaks processed.

The principle of the method is simple; that given sufficiently reasonable starting information about the peaks, a profile can be fitted successively to each measured step count and each profile fit gives information which can be used to fit subsequent reflections (see § 2.2).

In effecting the reduction of a complete set of data several steps are followed.

(i) An informed guess is made at the peak parameters (including $\underline{\alpha}, \xi$) of the first peak. The movements allowed for the parameters $\underline{\alpha}$ (the free variances) are set to generous values and the first 25 reflections or so are fitted.

(ii) Using the monitoring output of these fits, it is simple to improve the initial guess at the peak parameters $\underline{\alpha}$. The variation from one peak to the next in the processing run suggests better values for the free variances and also for the transition variances, which are used to update the knowledge of the prior distribution for $\underline{\alpha}$. On reading a measured step count, if the previous reflection was in the same region of reciprocal space and was fitted successfully, the prior distribution for $\underline{\alpha}$ is weakened by the predefined values of the transition variances (French (1975)).

The relative instability and modelling error, e^2, can be estimated from this initial output and has to have a value such that the standardised error between measured and fitted profiles:

$$\frac{(c_i - \nu_i)}{((1 + e^2\nu_i)\nu_i)^{\frac{1}{2}}}$$

(where ν_i is the predicted count) had mean, standard deviation, skewness and kurtosis of 0.0, 1.0, 0.0 and 0.0, respectively. Setting e^2 too large makes this standard deviation less than one and too small makes it larger than one.

(iii) The initial run is repeated with the improved guess at all these parameters until the fitted profiles explain the data adequately. A reasonable set of prior parameters and variances are determined in two or three attempts.

The algorithm in PRIMROSE converges well providing that the initial approximation to the peak parameters are reasonable (especially α_1) and the prior variances do not allow the peak parameters too much freedom. Usually convergence in the fitting is obtained in one cycle and if the current peak remains in the same region of reciprocal space then the prior moments ($\underline{\alpha}$) are weakened using predefined values of S_s, $\Delta\underline{\alpha}_s$ and var($\Delta\alpha_s$). Here S_s, for the Sth reflection, is a shift matrix allowing for changes in scale and predictable variation in peak shape, e.g. width; $\Delta\underline{\alpha}_s$ are the free variances and allow for the predictable shifts in the peak parameters and var ($\Delta\underline{\alpha}_s$) are the transition variances which weaken the distribution of belief according to the trust held in the predicted shift. If more than one cycle is needed the algorithm continues, when if still unsuccessful the prior distribution is reset to the free values and fitting reattempted. When a convergence failure has occurred on the previous reflection or processing has moved to a new region of reciprocal space, the free variances are used with the profile parameters set to either the last fitted values or those at the beginning of the previous row, whichever is nearer in reciprocal space. The fitting is then allowed

to run for the preset number of cycles. Occasionally, the parameters converge to ridiculous values and so a check is made on the quality of fit using a goodness of fit parameter G_f where:

$$G_f = \max_{n=1}^{N} \left| \sum_{i=1}^{n} (c_i - y_i) \right| \Big/ \sum_{i=1}^{N} c_i$$

where N is the number of steps in the scan.

If G_f is less than some specified value the fit is accepted. If not, or the fitting fails to converge, the ordinate analysis results are taken for the reflection intensity.

The parameters used in the fitting of the data are summarized in Table 4.3. The process worked smoothly for the native and $NADP^+$ data sets. Problems were encountered with the $IADP^+$ data, however. This crystal had slipped slightly causing the peak of the measured step count to move in position in the scan on passing from one reflection to the next. However even when one edge of the reflection was to the extreme end of the scan, the profile fitting provided sensible answers. The free variances var(α_1) etc. for the peak parameters had to be set at higher values than for the native and $NADP^+$ for this reason. Some examples of the fitting for each data set are given in Figure 4.2. Figure 4.2d shows a very weak reflection which ordinate analysis places between points 2 and 7 of the scan since this has the maximum sum count over 6 consecutive counts in the scan. The profile analysis places thepeak at position 6.85 in the scan in agreement with previously placed stronger reflections. Figure 4.2j shows a reflection to the right hand edge of the scan for the $IADP^+$. Figure 4.2k shows a weak reflection which ordinate analysis manages to place between points 4 and 10 and is close to the peak position of 7.87 for the profile fitted result.

Following the profile fitting stage the program applies to the intensity

	Native		NADP+										IADP+			
shell | batch			1		2		3		4		5		1		2	
ε, for curvetype	+1		+1		+1		+1		+1		+1		+1		+1	
"free" peak parameter distribution																
$E(\alpha_1)$ $Var(\alpha_1)$	7.0	0.06	7.0	0.06	7.75	0.06	7.0	0.06	6.5	0.06	7.0	0.06	7.0	0.1	9.0	0.04
$E(\alpha_2)$ $Var(\alpha_2)$	4.0	0.04	4.0	0.04	3.5	0.04	4.0	0.04	3.0	0.04	3.25	0.04	4.5	0.4	3.5	0.4
$E(\alpha_3)$ $Var(\alpha_3)$	0.0	0.01	0.0	0.01	0.0	0.01	0.0	0.01	0.0	0.01	0.0	0.01	0.0	0.1	0.0	0.1
$E(\alpha_4)$ $Var(\alpha_4)$	3.25	0.01	3.25	0.01	3.0	0.01	3.25	0.01	3.5	0.01	3.4	0.01	3.5	0.01	3.65	0.01
Prior for intensities and background																
E(J) Var(J)	?	10^{20}	→													
$E(\beta_1)$ $Var(\beta_1)$	?	10^{20}	→													
$E(\beta_2)$ $Var(\beta_2)$	0	10^{-20}	→													
Transition variances																
$Var(\Delta\alpha_{s1})$	0.006		→										0.1			
$Var(\Delta\alpha_{s2})$	0.004		→										0.04			
$Var(\Delta\alpha_{s3})$	0.001		→										0.01			
$Var(\Delta\alpha_{s4})$	0.001		→										0.001			
e^2	0.00005		→										0.0001		0.00002	
Convergence criteria:																
max. cycles	6	→														
G_f less than	0.04		0.03	→												
Number of reflns. fitted successfully	3263		519		189		970		705		964		2623		692	
Number where O/A taken	0		7		0		0		5		3		2		0	

Figure 4.2 Some profile fitting examples for each data set.

(a) and (b) native; (c) - (h) $NADP^+$ and (i) - (l) $IADP^+$. The ordinate analysis estimate is included for comparison.

(a) 3 5 2 reflection, native , nref = 600.

Ordinate analysis : I = 1959, σ = 92

peak position = 8 ; peak width = 7 steps.

Profile fitting : J = 2011, σ = 83

α_1 = 7.55; α_2 = 4.16; α_3 = -0.21; α_4 = 2.91.

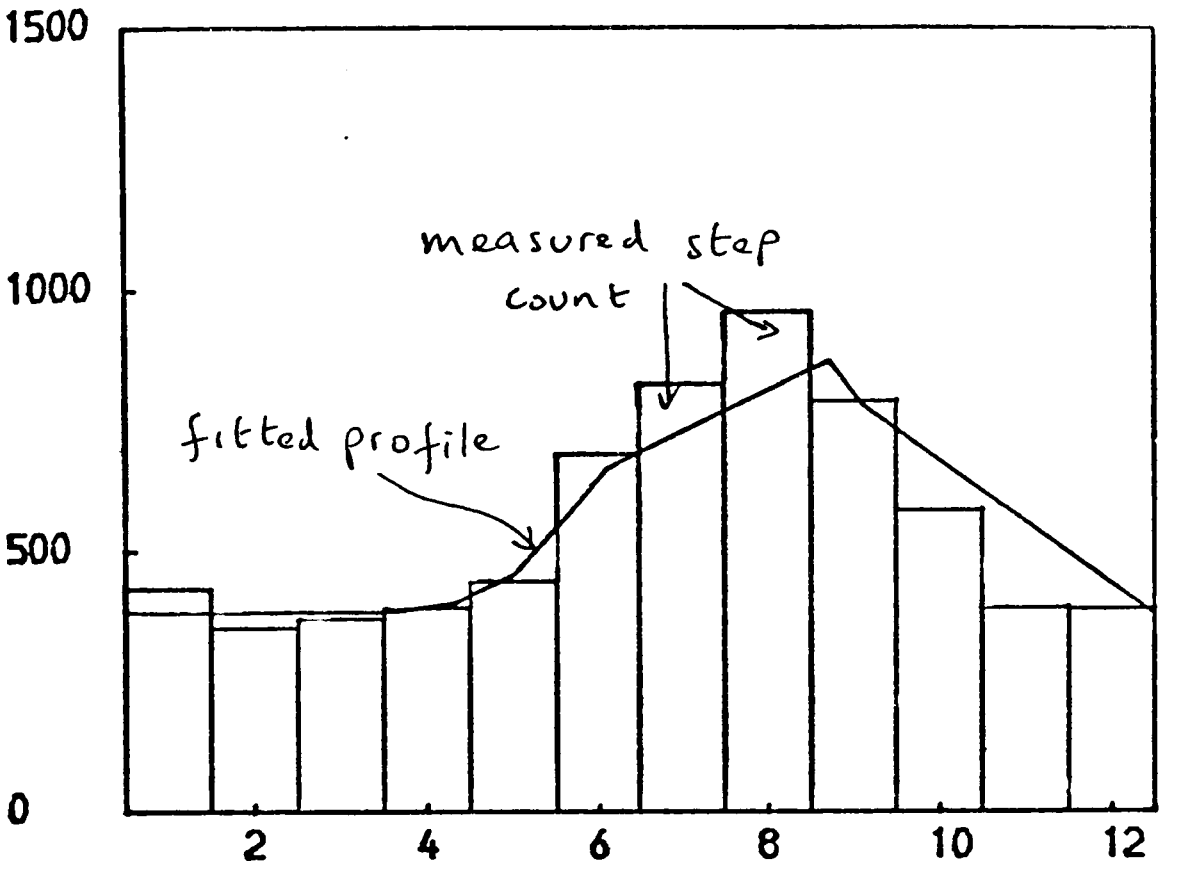
(b) 10 8 -8 reflection native, nref = 3200.

Ordinate analysis : I = 649 , σ = 61

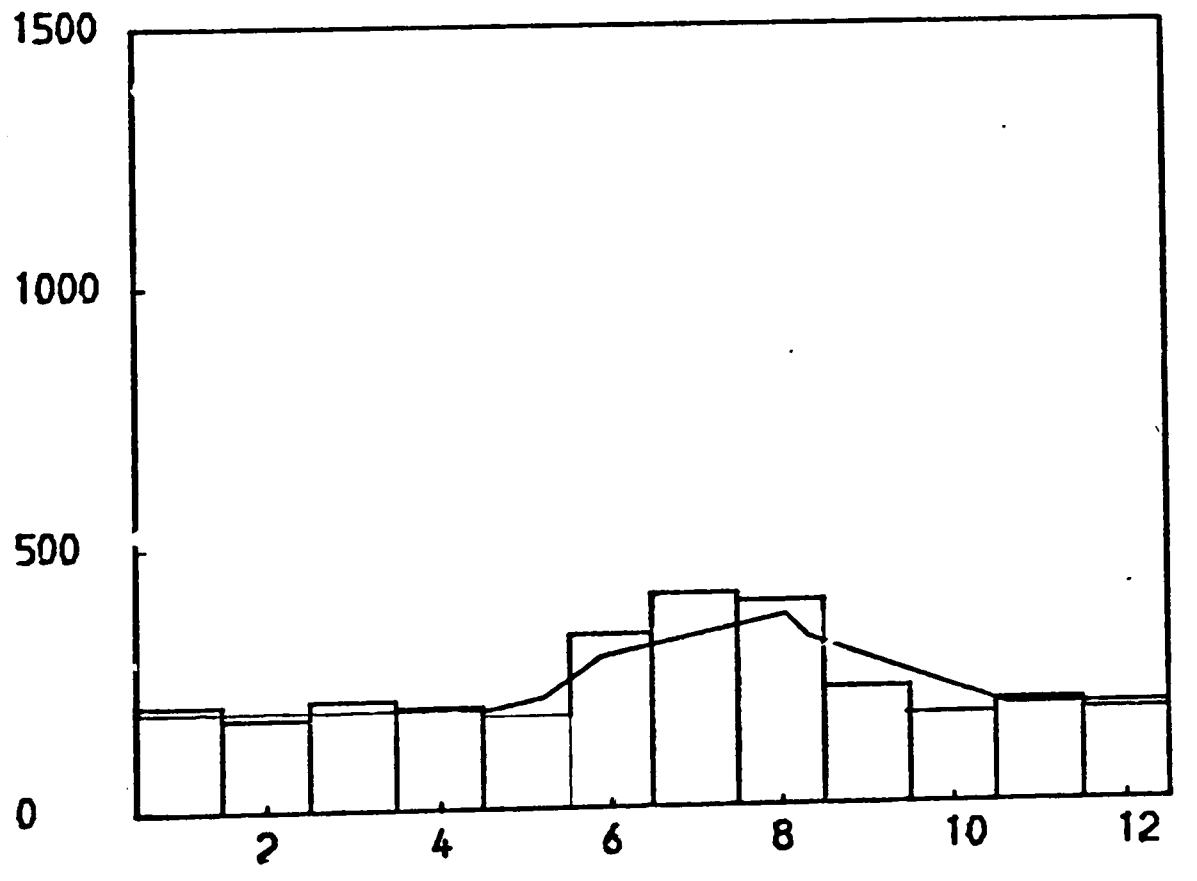
peak position = 7, peak width = 7 steps.

Profile fitting : J = 607 , σ = 46

α_1 = 7.10; α_2 = 2.81; α_3 = -0.11; α_4 = 2.94.


1500
1000
500
0
measured step count
fitted profile
2
4
6
8
10
12

(a)


1500
1000
500
0
2
4
6
8
10
12

(b)

(c) 1 15 2 reflection $NADP^+$, nref = 200 in Batch #1 (256 reflections in total).

Ordinate analysis: I = 1057 σ = 52 peak position = 6.5

peak width = 6 steps

Profile fitting: J = 1040 σ = 51 α_1 = 6.72

α_2 = 4.11

α_3 = -0.01

α_4 = 3.15

(d) 1 11 9 reflection $NADP^+$, nref = 187 in Batch #1

Ordinate analysis: I = 114 σ = 39 peak position = 4.5

peak width = 6 steps

Profile fitting: J = 11, σ = 31 α_1 = 6.85

α_2 = 4.07

α_3 = 0.01

α_4 = 3.18

Notice how ordinate analysis centres on the noise peak at the left hand end of the scan.

(e) 7 1 5 reflection $NADP^+$, n ref = 120 in Batch # 2

(189 reflections in total).

Ordinate analysis: I = 2645 σ = 67 peak position = 6.5

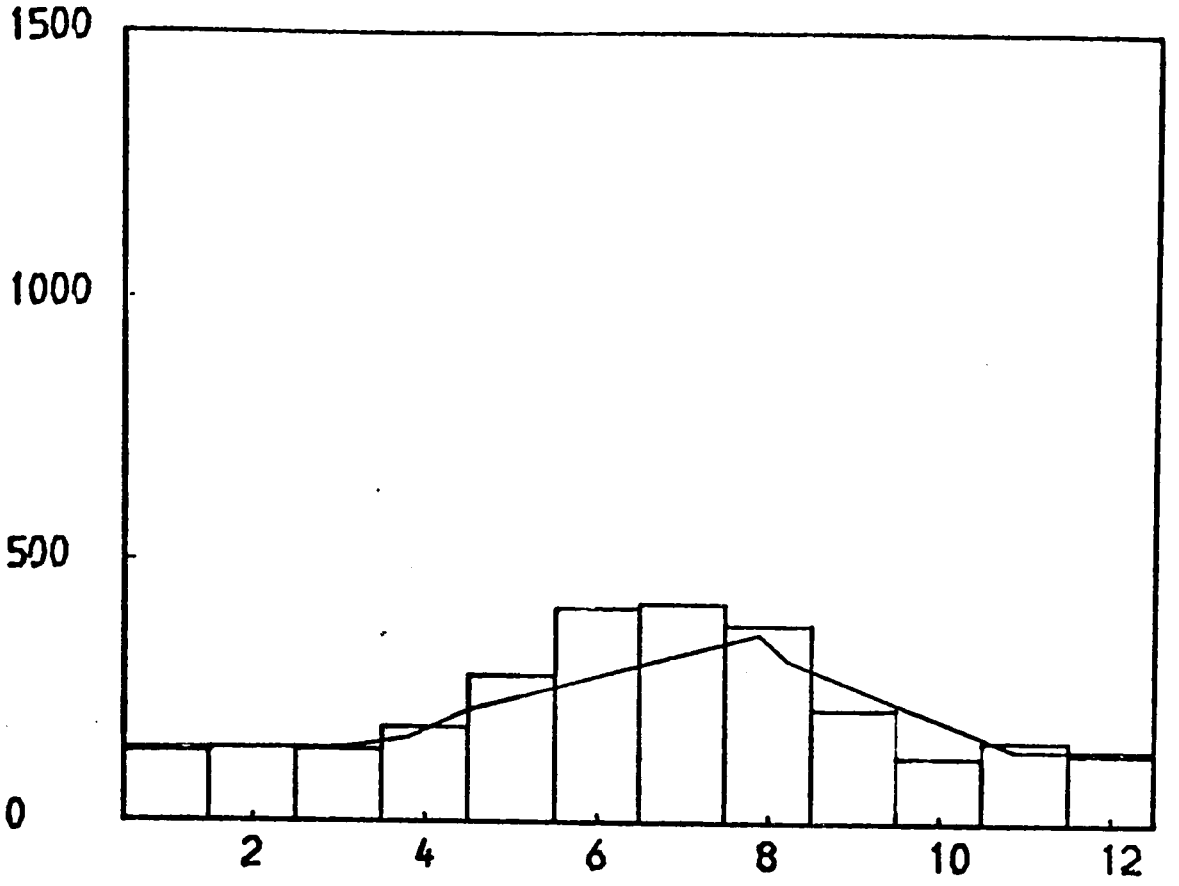
peak width = 6 steps

Profile fitting: J = 2716 σ = 71 α_1 = 7.09

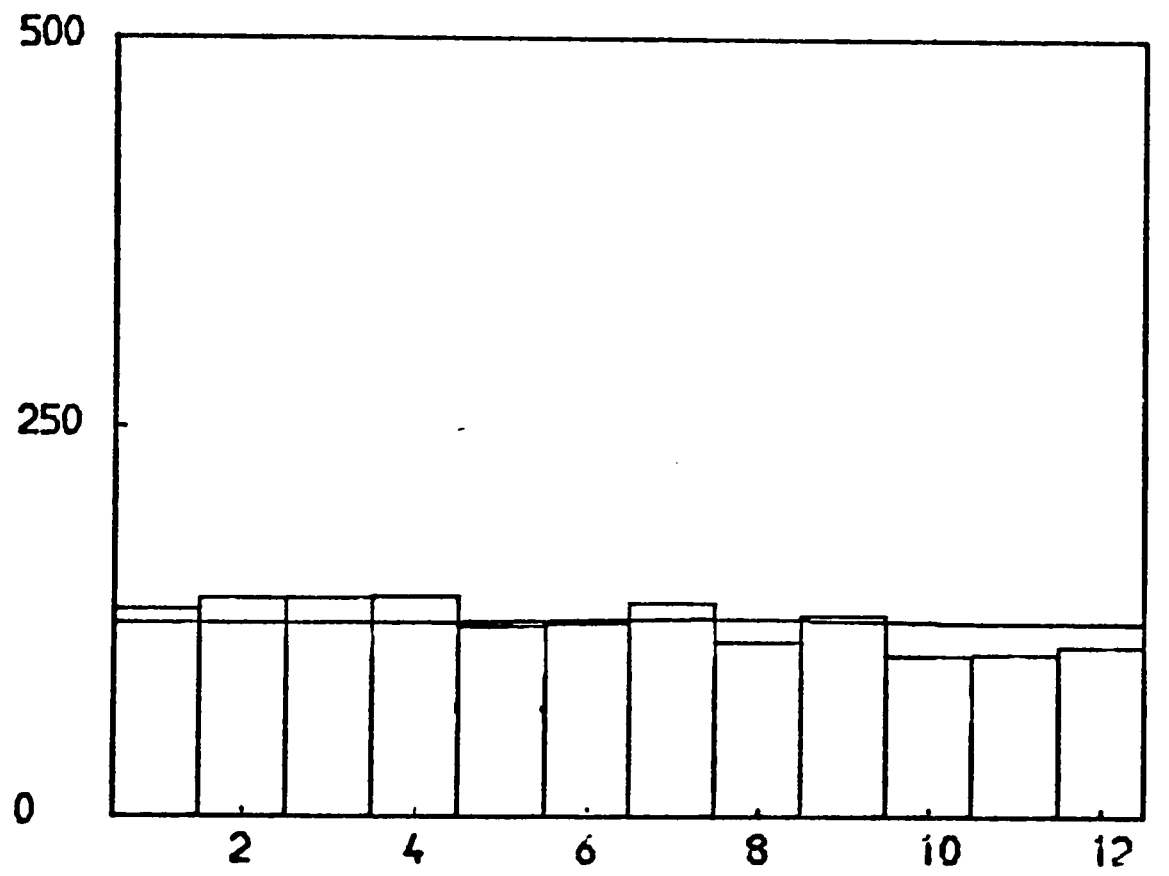
α_2 = 3.60

α_3 = 0.12

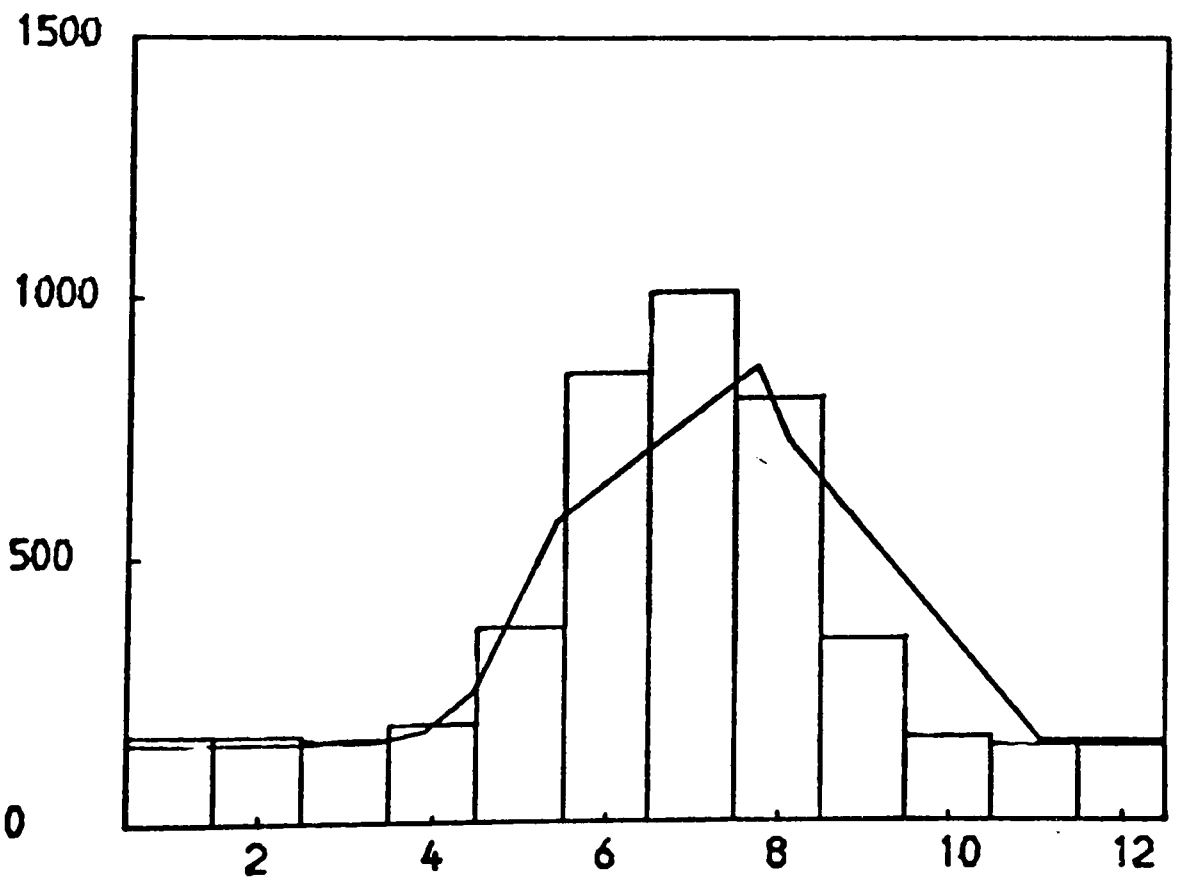
α_4 = 2.92


1500
1000
500
0
2
4
6
8
10
12

(c)


500
250
0
2
4
6
8
10
12

(d)


1500
1000
500
0
2
4
6
8
10
12

(e)

(f) 5 5 12 reflection $NADP^+$ nref = 450 in Batch # 3 (970 reflections in total)

Ordinate analysis: I = 909 σ = 55 peak position = 8.5

peak width = 6 steps

Profile fitting: J = 924 σ = 50 α_1 = 7.38

α_2 = 3.44

α_3 = 0.05

α_4 = 3.30

(g) 4 4 -8 reflection $NADP^+$ nref = 440 in Batch # 4

(710 reflections in total)

Ordinate analysis: I = 2164 σ = 64 peak position = 6.5

peak width = 6 steps

Profile fitting: J = 2096 σ = 64 α_1 = 7.13

α_2 = 2.99

α_3 = -0.18

α_4 = 3.45

(h) 7 1 -13 reflection $NADP^+$ nref = 600 in Batch # 5

(967 reflections in total)

Ordinate analysis: I = 494 σ = 48 peak position = 8.5

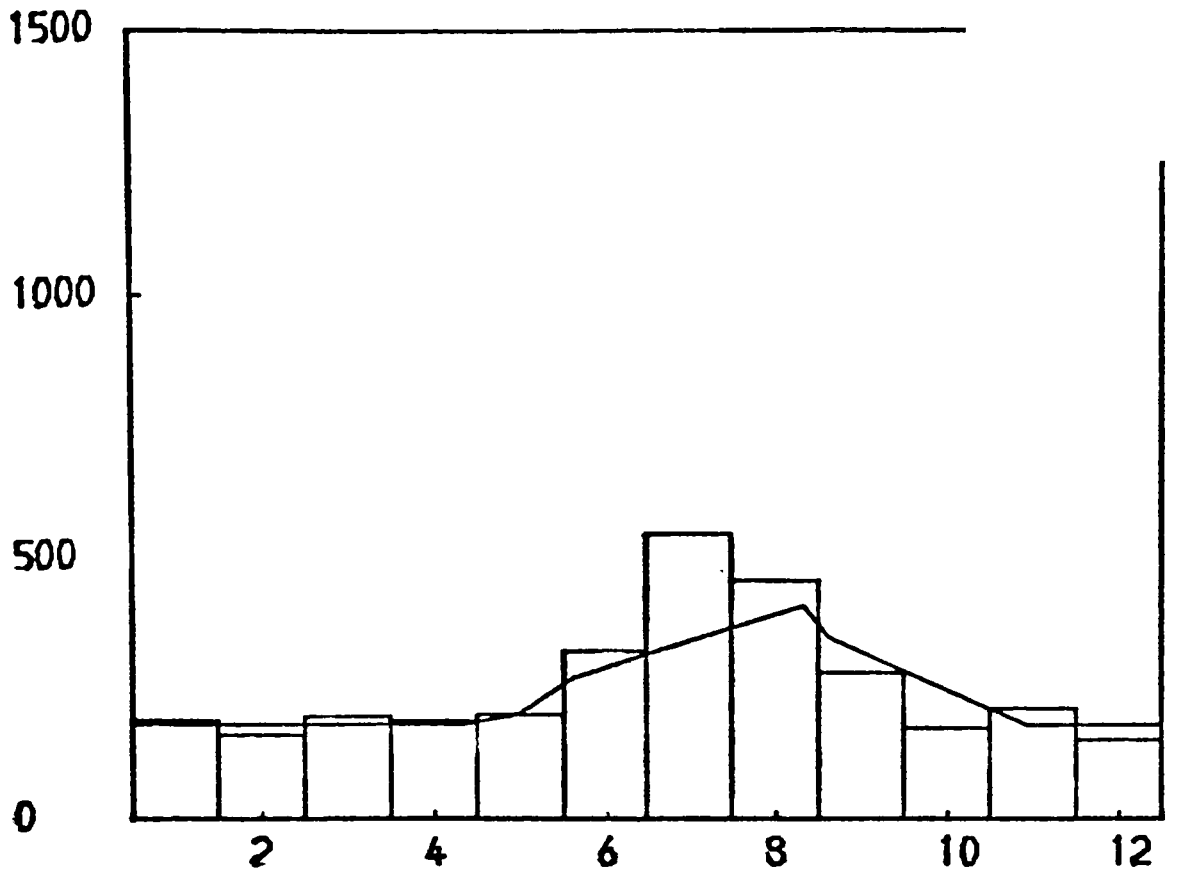
peak width = 6 steps

Profile fitting: J = 494 σ = 41 α_1 = 7.62

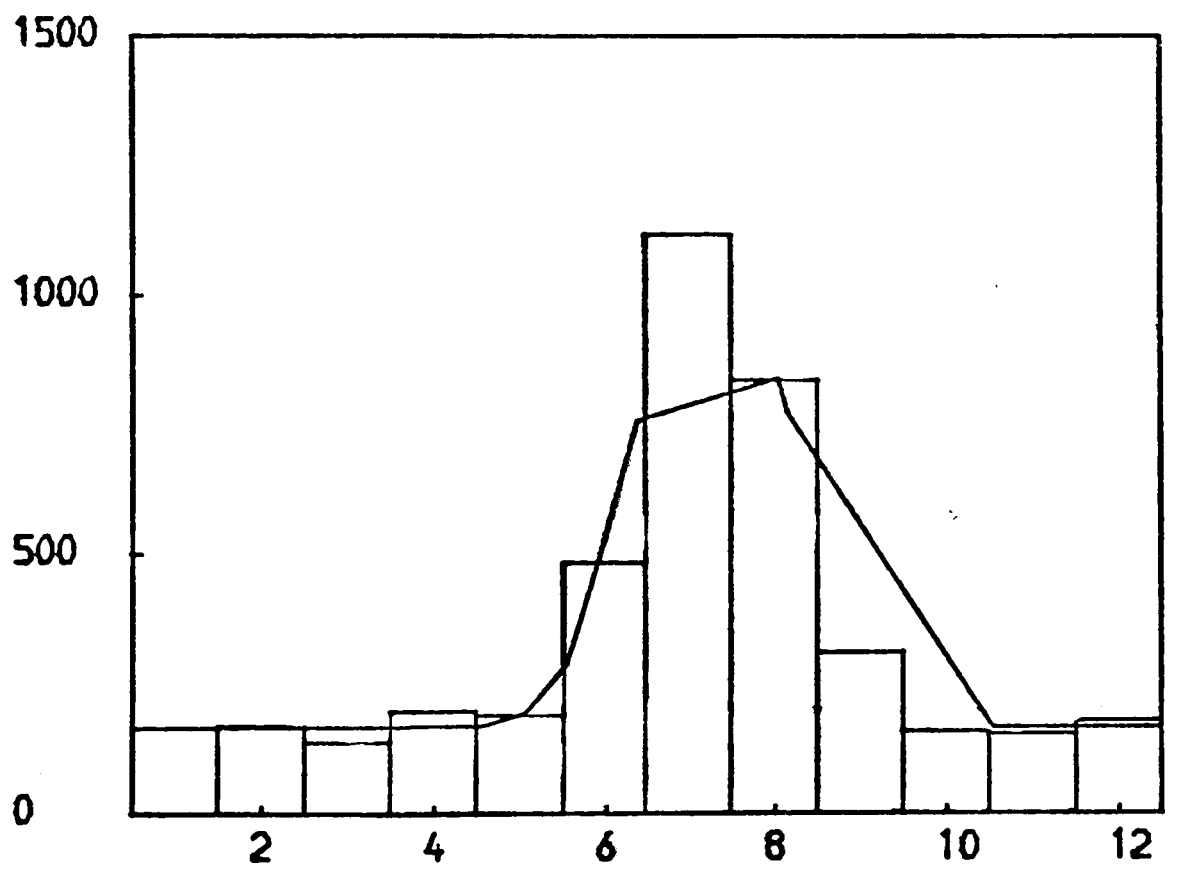
α_2 = 3.43

α_3 = -0.09

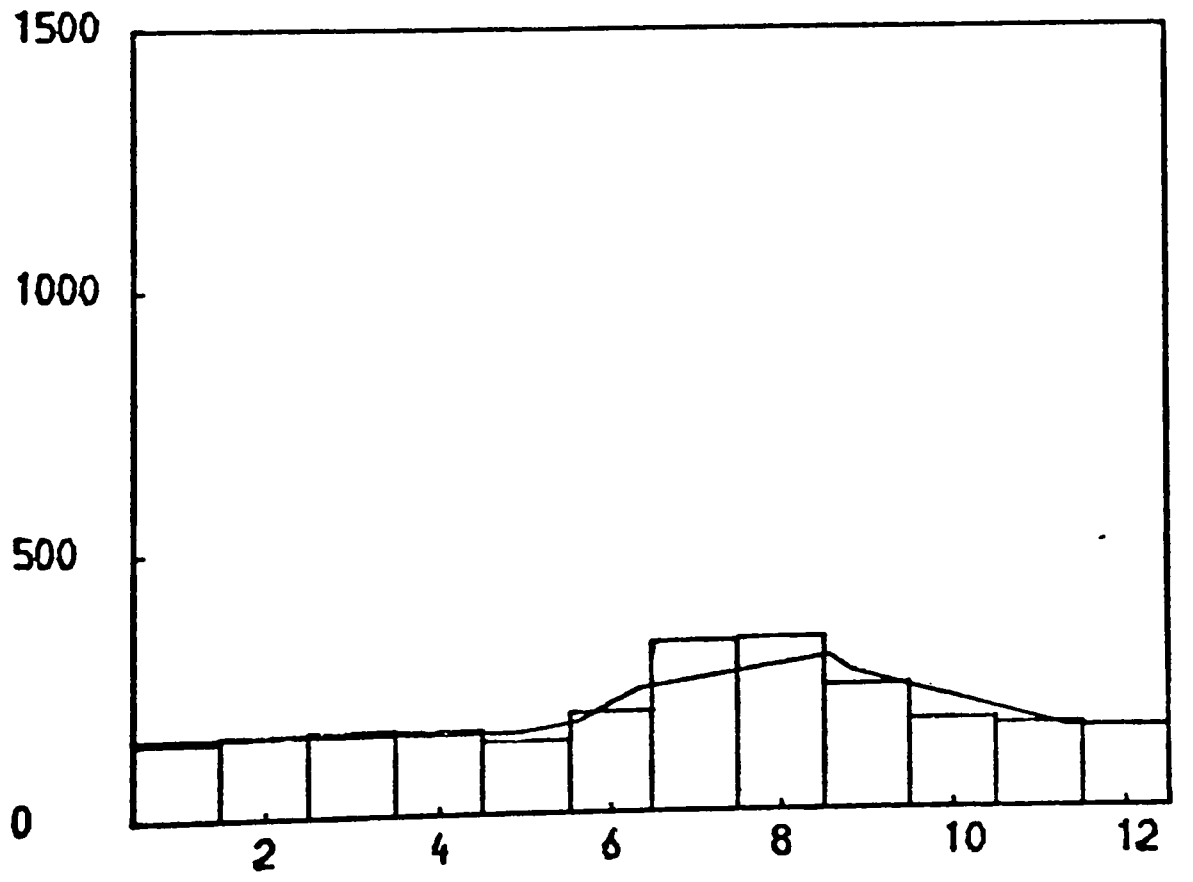
α_4 = 3.32


1500
1000
500
0
2
4
6
8
10
12

(f)


1500
1000
500
0
2
4
6
8
10
12

(g)


1500
1000
500
0
2
4
6
8
10
12

(h)

(i) 2 2 8 reflection IADP$^+$ nref = 400 in Batch #1 (2625 reflections in total)

Ordinate analysis: I = 1969 σ = 117 peak position = 8

peak width = 7 steps

Profile fitting: J = 2030 σ = 103 α_1 = 7.44 ,

α_2 = 4.54

α_3 = 0.06

α_4 = 3.54

(j) 3 11 0 reflection IADP$^+$ nref = 650 in Batch #1

(2625 reflections in total)

Ordinate analysis: I = 2207, σ = 119 peak position = 9

peak width = 7 steps

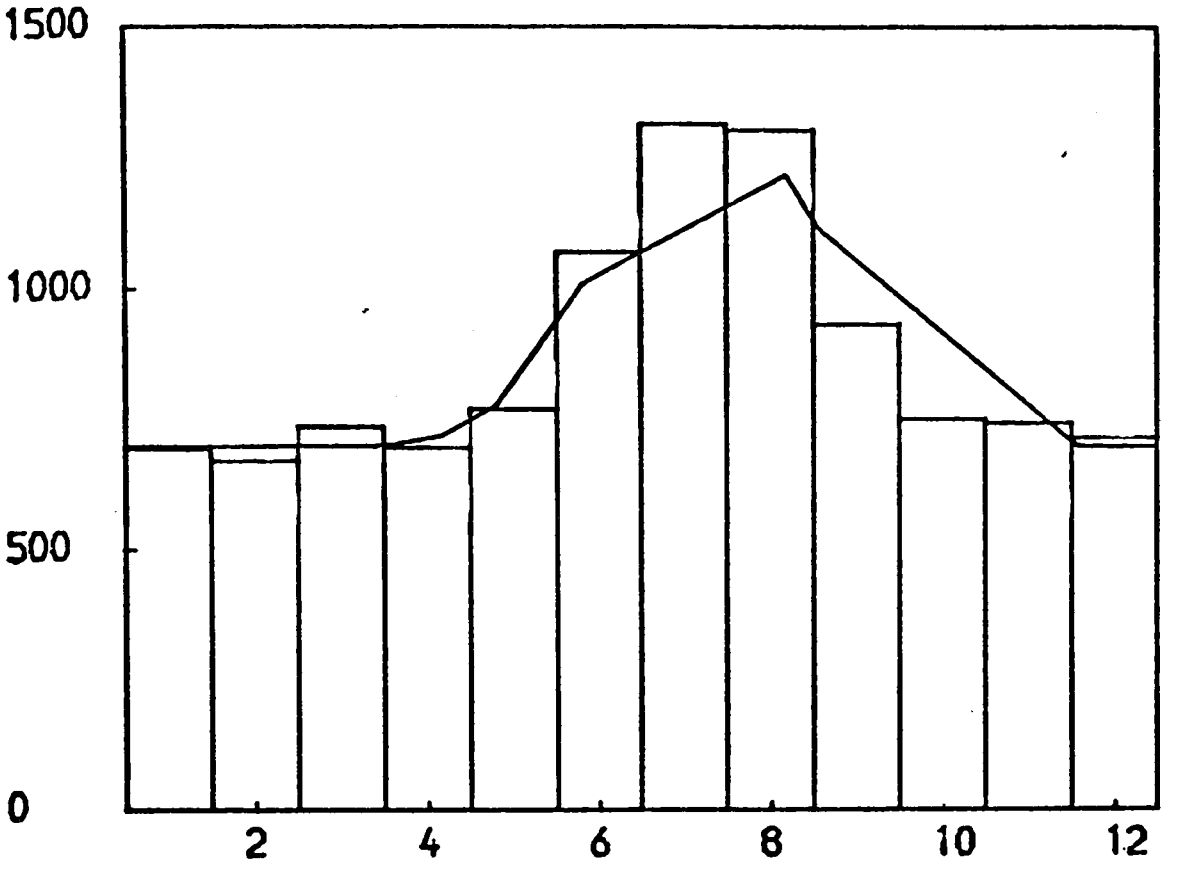
Profile fitting: J = 2136, σ = 104 α_1 = 8.35

α_2 = 4.42

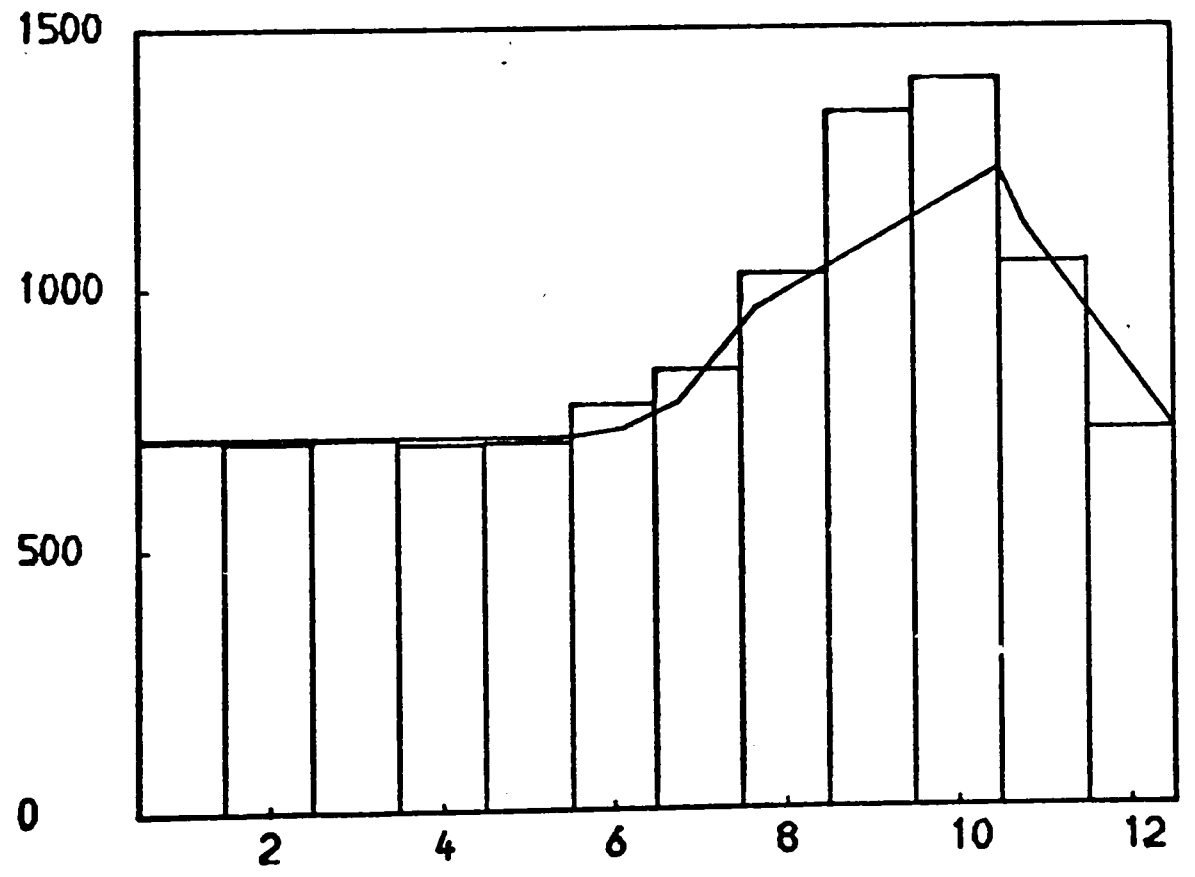
α_3 = -0.89

α_4 = 3.52

Notice that since the reflection depicted in (i) the crystal has slipped and that profile analysis has adjusted to the peak movement.



(i)



(j)

(k) 6 4 10 reflection I_{ADP}^{+} nref = 1100 in Batch # 1

(2625 reflections in total)

Ordinate analysis: I = 114 σ = 92

peak position = 7

peak width = 7

Profile fitting: J = 50 σ = 64

α_1 = 7.87

α_2 = 4.66

α_3 = -0.16

α_4 = 3.54

(l) 7 1 -3 reflection I_{ADP}^{+} nref = 280 in Batch # 2

(692 reflections in total)

Ordinate analysis: I = 2532, σ = 113

peak position = 9

peak width = 7

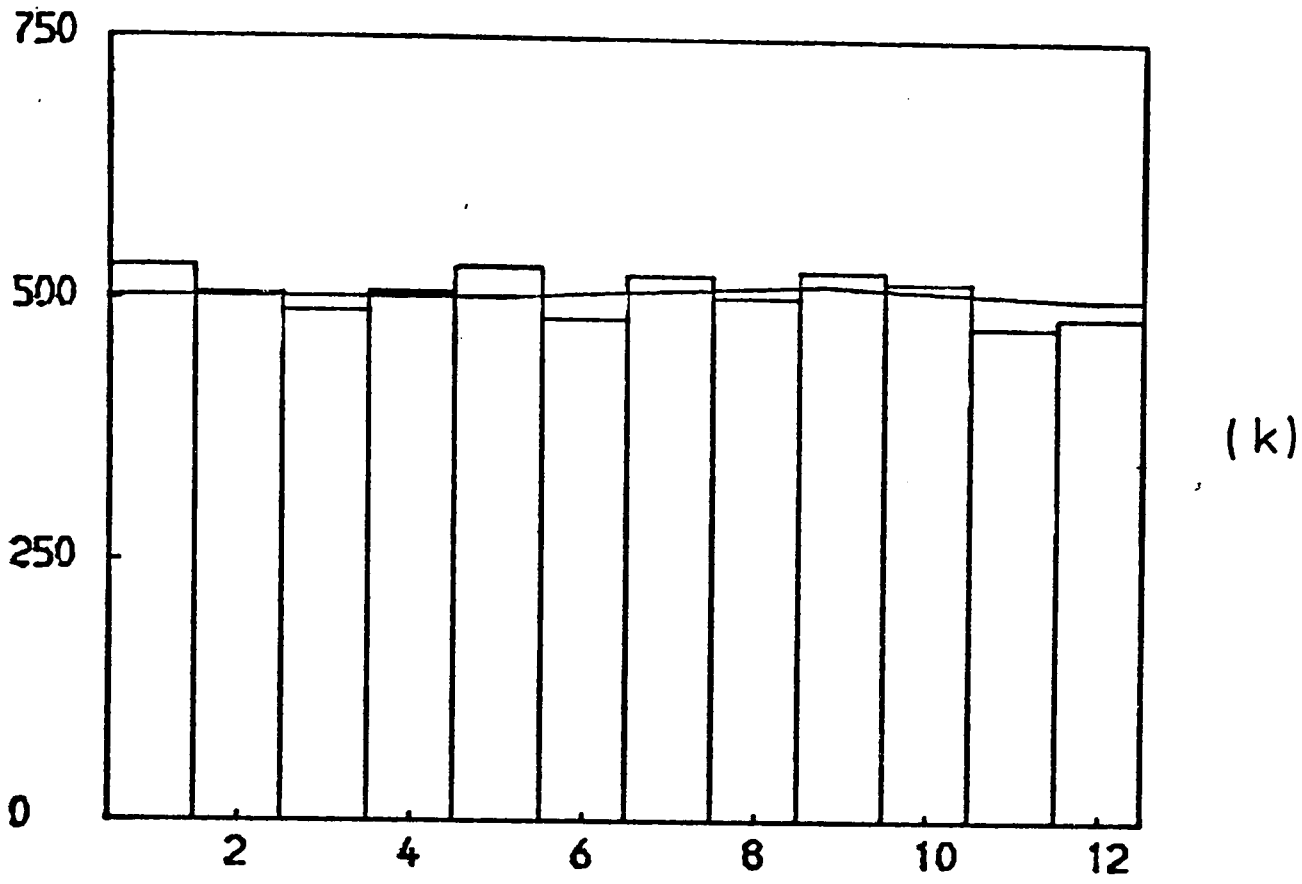
Profile fitting: J = 2441 σ = 91

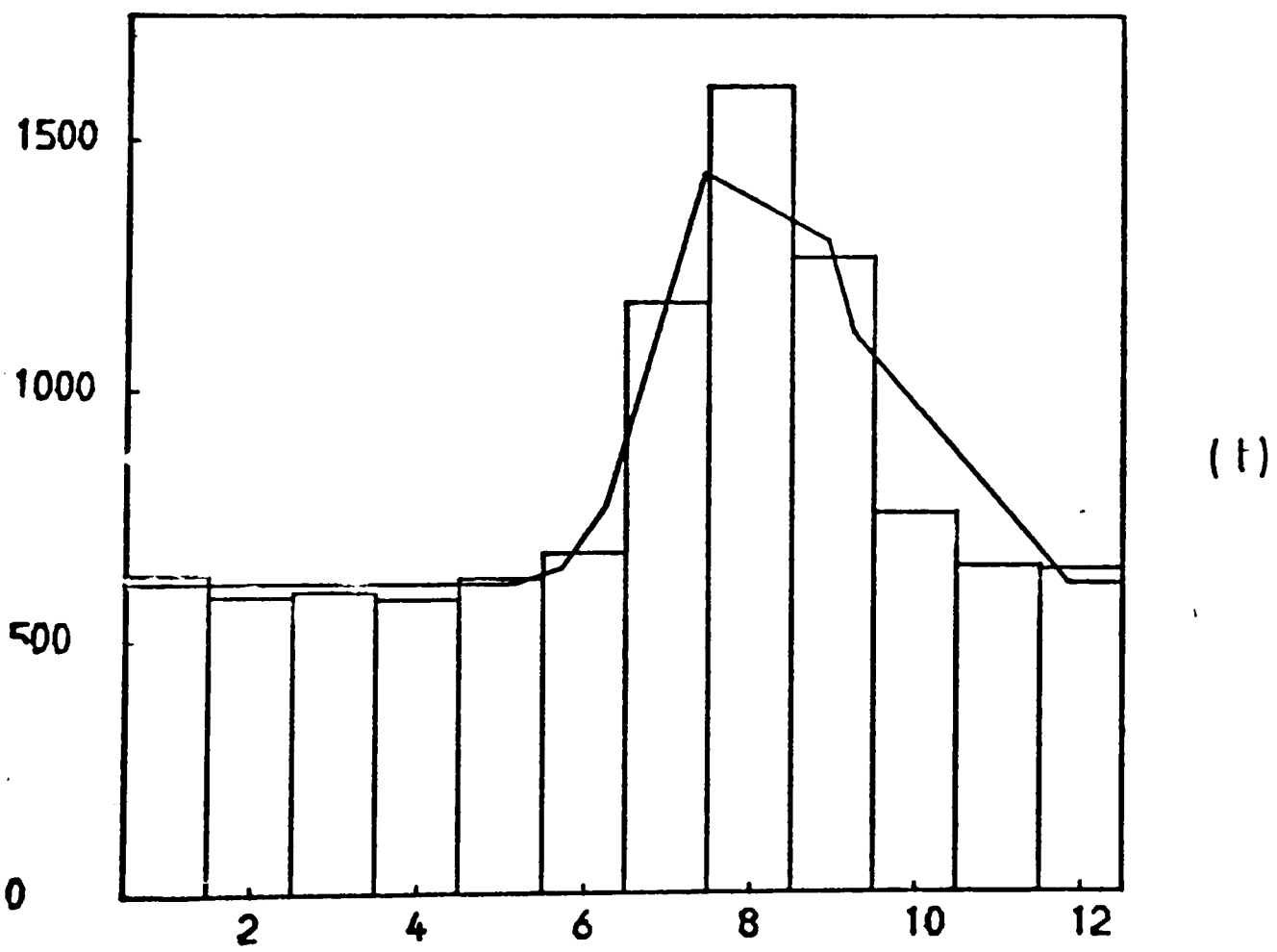
α_1 = 8.56

α_2 = 3.60

α_3 = 0.47

α_4 = 3.65


750
500
250
0
2
4
6
8
10
12
(k)


1500
1000
500
0
2
4
6
8
10
12
(t)

the LP correction, a correction for absorption effects using the method of North et al. (1968) and calculates a correction for radiation damage using a scale inversely proportional to $\sum_{standards} I$. Table 4.4 gives intensity statistics output by PRIMROSE and shows that the quality of the N_V and $NADP^+$ data are comparable but that of the $IADP^+$ is poorer.

Each data set was merged using the program SSM and the derivative to native scale factors analysed by ANSC. Table 4.5 gives the merging R factors, scales used between the data sets and mean fractional changes in $|\underline{F}|$. The merging R factors are higher than those given in Table 3.2 but this is in fact due to the large fraction of weak data for which the overestimation by ordinate analysis of the intensities causes an unfair reduction in the merging R relative to the profile fitted data set (see French (1975) § 5.5.x).

4.3. Difference Fourier maps

Difference Fourier maps were calculated with coefficients $(F_{NADP^+} - F_{N_V})$, $(F_{IADP^+} - F_{N_V})$ and $(F_{IADP^+} - F_{NADP^+})$ using the best phase set $(Au+Pt)_2$ and the single derivative phase set calculated using the isomorphous and anomalous differences of the Au derivative.

Difference maps computed for the $NADP^+$ and $IADP^+$ series of differences using best phases showed a large peak which was in a similar position in each. The peak was 14 Å long and one end was adjacent to the major Pt site. There was no correspondingly large negative density or alternate negative and positive density which would indicate conformational changes of the enzyme, for the $NADP^+$ map. For the $IADP^+$ map, however, there is a positive/negative ripple in the region close to the binding site on section X=11 (see Figure 4.4). The peak sizes are -99, the biggest negative peak on section 11 and one of the biggest on the map, and +74 the highest positive peak on section 10 (see Table 4.6 to compare with

Table 4.4 Intensity statistics output by PRIMROSE for the Native, $NADP^+$ and $IADP^+$ data sets.

		Number of reflections processed	Negative	Number(%) of reflections measured $0<\frac{I}{\sigma}<1$	$1<\frac{I}{\sigma}<2$	$2<\frac{I}{\sigma}<3$	$3<\frac{I}{\sigma}$
Native		3263	194(5.9)	246(7.5)	274(8.4)	492(15.1)	2057(63.0)
$NADP^+$	1	526	20(3.8)	32(6.1)	25(4.8)	57(10.9)	392(74.5)
	2	189	8(4.2)	9(4.8)	14(7.4)	17(9.0)	141(74.6)
	3	970	55(5.7)	104(10.7)	85(8.8)	180(18.6)	546(56.3)
	4	710	23(3.2)	39(5.5)	39(5.5)	75(10.6)	534(75.2)
	5	967	64(6.6)	96(9.9)	113(11.7)	167(17.3)	527(54.5)
Totals		3362	170(5.1)	280(8.3)	276(8.2)	496(14.8)	2140(63.7)
$IADP^+$	1	2625	219(8.3)	273(10.4)	324(12.3)	489(18.6)	1319(50.3)
	2	692	54(7.8)	81(11.7)	102(14.7)	149(21.5)	306(44.2)
Totals		3317	273(8.2)	354(10.7)	426(12.8)	635(19.1)	1625(49.0)

Table 4.5 Information output by SSM and ANSC giving the merging R factors between the 2 equivalents of a data set, the scales applied between two data sets and the mean fractional isomorphous change in $|\underline{F}|$ between the native and the given data sets.

	R_{symm}	Mean fractional isomorphous change	a	b
N_V	7.3	-	-	-
$NADP^+$	6.8	11.81	0.91	-2.9873
λADP^+	10.4	14.60	0.99	-0.1433
$\lambda ADP^+ : NADP^+$	-	12.29	1.1	2.9031

where the scale k applied to the derivative was of the form $a+b(4\sin^2\theta/\lambda^2)$.

the other peaks on the map). On the ($IADP^+$ - $NADP^+$) map calculated with best phases there exists a positive/negative ripple in the same position in Y and Z but displaced in X from 10 to 9.5 and from 11 to 8 for the positive and negative peak respectively. The movement from the negative to positive peak is 4 Å bringing the positive peak to within 4 Å of the Pt end of the site. If these peaks are significant it could be inferred that the binding of the $IADP^+$ induces a conformational change in the enzyme which the $NADP^+$ does not.

Because the coenzyme density is close to the Pt site of major substitution it was considered that it might be sensitive to errors in the phases causing ripple around the heavy atom sites. A second set of difference maps was thus calculated using phases derived only from the gold derivative. Although these maps would be noisier since the phases are less reliable there should be no bias at the platinum site. The major peak on each of these maps was in the same position and of the same size as before and there was no other feature above background. The end of the density adjacent to the Pt site was however reduced in magnitude relative to the other end indicating perhaps that the peak at this end of the density in the best phase map was a composite of coenzyme and a Pt ghost peak. However, comparison of this map with that calculated using the series of differences $IADP^+$ - $NADP^+$ and best phases showed no disturbance of the Pt site indicating that there was in fact no ghosting of this site. This is supported by the work presented in § 3 where in the Hg difference maps the major ghosting was associated with the Au derivative sites and very little with the Pt sites. The reduction observed at the Pt end of the coenzyme density is most probably an effect therefore of the increased errors of the Au phases.

For the $IADP^+$ - N_V map the effect of using the Au instead of best phases is to reduce the magnitude of the peak by one contour level uniformly over the coenzyme peak (see Figure 4.7) rather than just at one

Table 4.6 Comparison of the signal to noise ratios for the difference maps shown.

	Map		V''	Maximum peak height	⟨Background⟩over sections not containing the signal	Next highest positive peak	Highest negative peak
$(NADP^+-N_V)$	α_{best}	Fig.4.3	94.9	155	78	99	-102
"	α_{Au}	Fig.4.7	100.0	134	78	88	-87
$(IADP^+-N_V)$	α_{best}	Fig.4.4	118.0	143	78	86	-99
	α_{Au}	Fig.4.7	119.6	115	78	88	-97
$(IADP^+-NADP^+)\alpha_{best}$		Fig.4.5	109.2	94	78	87	-100

These difference maps were contoured in 7 intervals (-90, -60, 60, 90, 120, 150, 180) where the negative contours are shown dashed.

end.

On the ($IADP^+$ - $NADP^+$) difference map there was a small positive peak at a distance 10 Å removed from the Pt binding site.

Figure 4.3 shows composite views in X and Z for the ($NADP^+$ - N_V) difference map using best phases. Figures 4.4 and 4.5 show similar views for the ($IADP^+$ - N_V) and ($IADP^+$- $NADP^+$) difference syntheses, respectively. Figure 4.6 shows similar maps calculated with Au phases and superimposed on the final native map. Figure 4.7 shows section X=13 for the best and Au phased maps for the two series of terms. A view of the balsa wood model of the protein subunit is shown in Figure 4.8 with the density corresponding to $NADP^+$ arrowed. Table 4.6 compares the signal to noise ratios in all these maps as well as the relative scales. The positions of the platinum and iodine sites are marked on the maps where appropriate by a * and a label.

The interpretation of these maps is that the significant positive density in the ($NADP^+$- N_V) and ($IADP^+$-N_V) map corresponds to the coenzyme and coenzyme derivative respectively. The small positive region in the ($IADP^+$ - $NADP^+$) map is assigned to the difference between the substituent iodine and the replaced amide group of the pyridine ring of the nicotinamide moiety and serves to orient the coenzyme with respect to the enzyme. The adenine part of the coenzyme should thus be close to the platinum binding site. The binding of platinum compounds at the coenzyme binding sites has been observed for LDH, MDH and LADH. In LADH the site is a general anion binding site and is at the position of the pyrophosphate of ADP-ribose. The position of the Pt site in 6PGDH seems to correspond to the adenine moiety however. Also, in LADH, $Au(CN)_2$ acts in strict competition with the coenzyme NAD^+, binding at two sites. One site is normally occupied by the pyrophosphate of ADP-ribose and the other by the adenine part (Eklund et al. 1976) whereas in 6PGDH the $Au(CN)_2^-$

Figure 4.3

Composite views for (a) sections Z = 19/48⇒24/48 with the Z-axis coming out of the paper and (b) sections X=7/40⇒16/40 , with the X-axis coming out of the paper, for the difference synthesis calculated with terms in ($NADP^+$ - native)exp$i\alpha_{best}$ where m is the figure of merit, on a scale of 4 Å/cm. The positions corresponding to the Pt binding site and the iodine atom are marked. By holding (a) vertical and (b) horizontal, the axes form an orthogonal system.

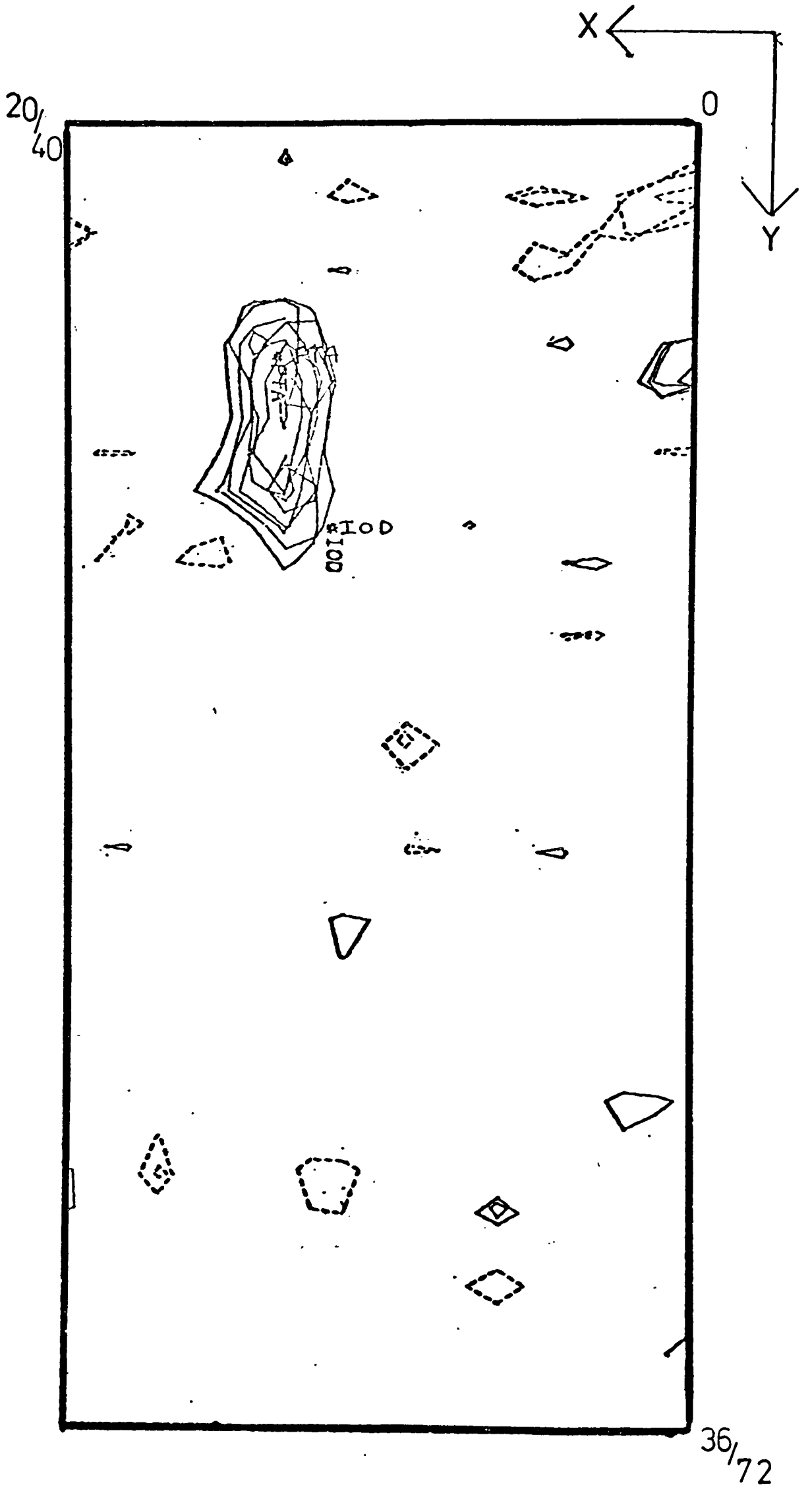


Figure 4.3a

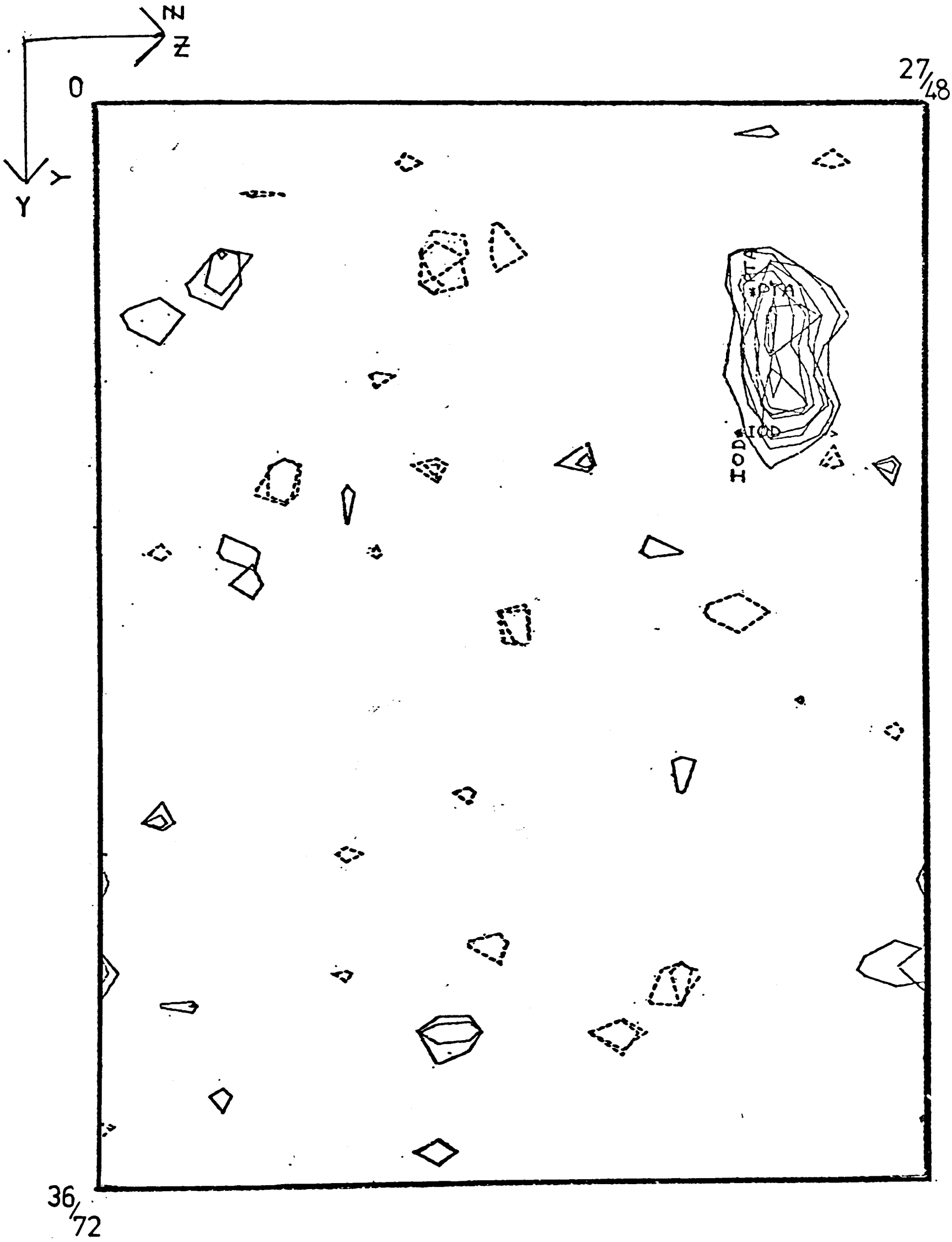


Figure 4.3b

Figure 4.4

Composite views for (a) sections $z=19/48 \rightarrow 24/48$, Z-axis out of the paper and (b) sections $X=7/40 \rightarrow 16/40$, X-axis out of the paper, for the difference syntheses calculated with terms in $(IADP^{+} - N_{V}) \exp i\alpha_{best}$, on a scale of 4 Å/cm.

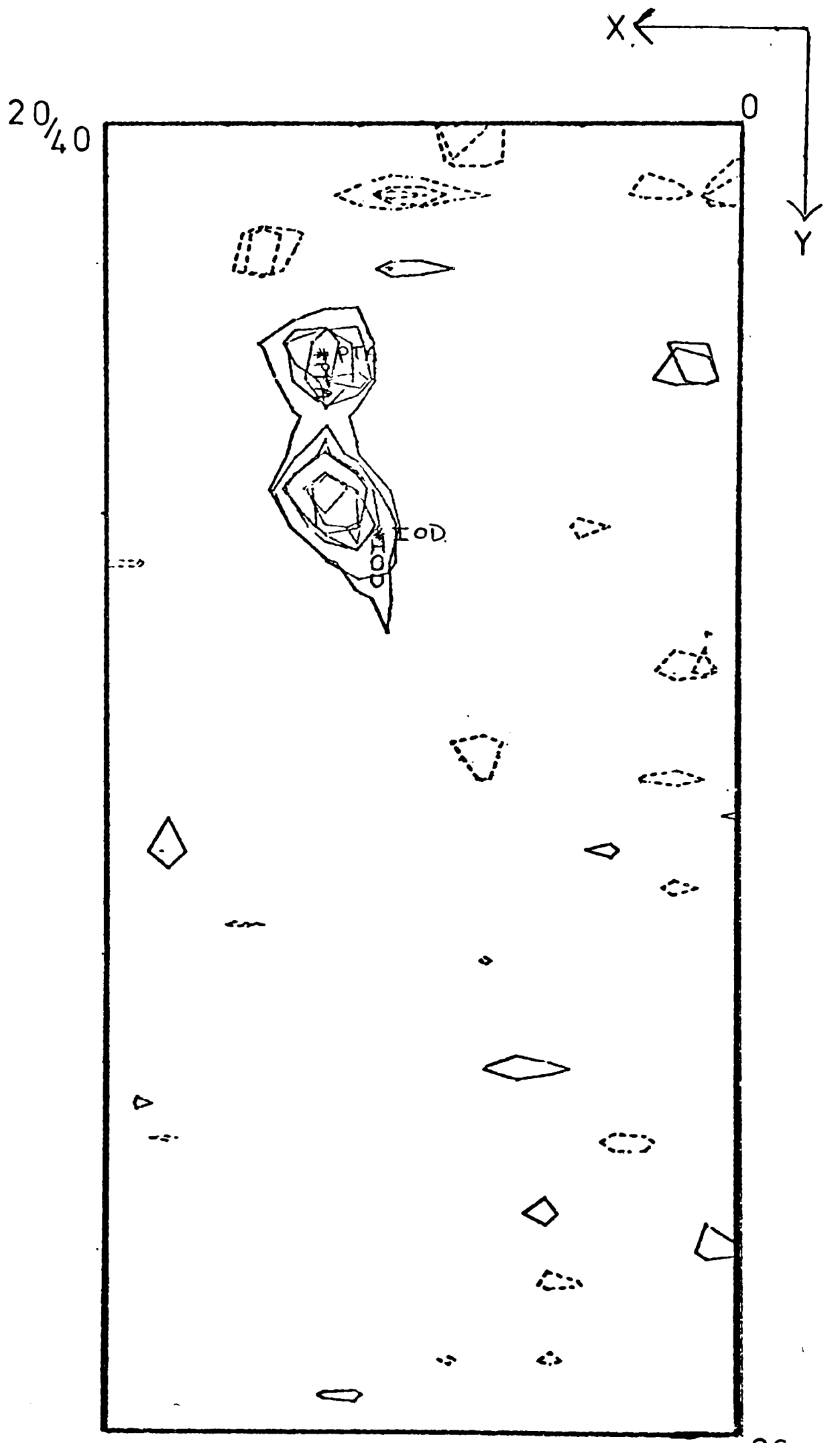


Figure 4.4a

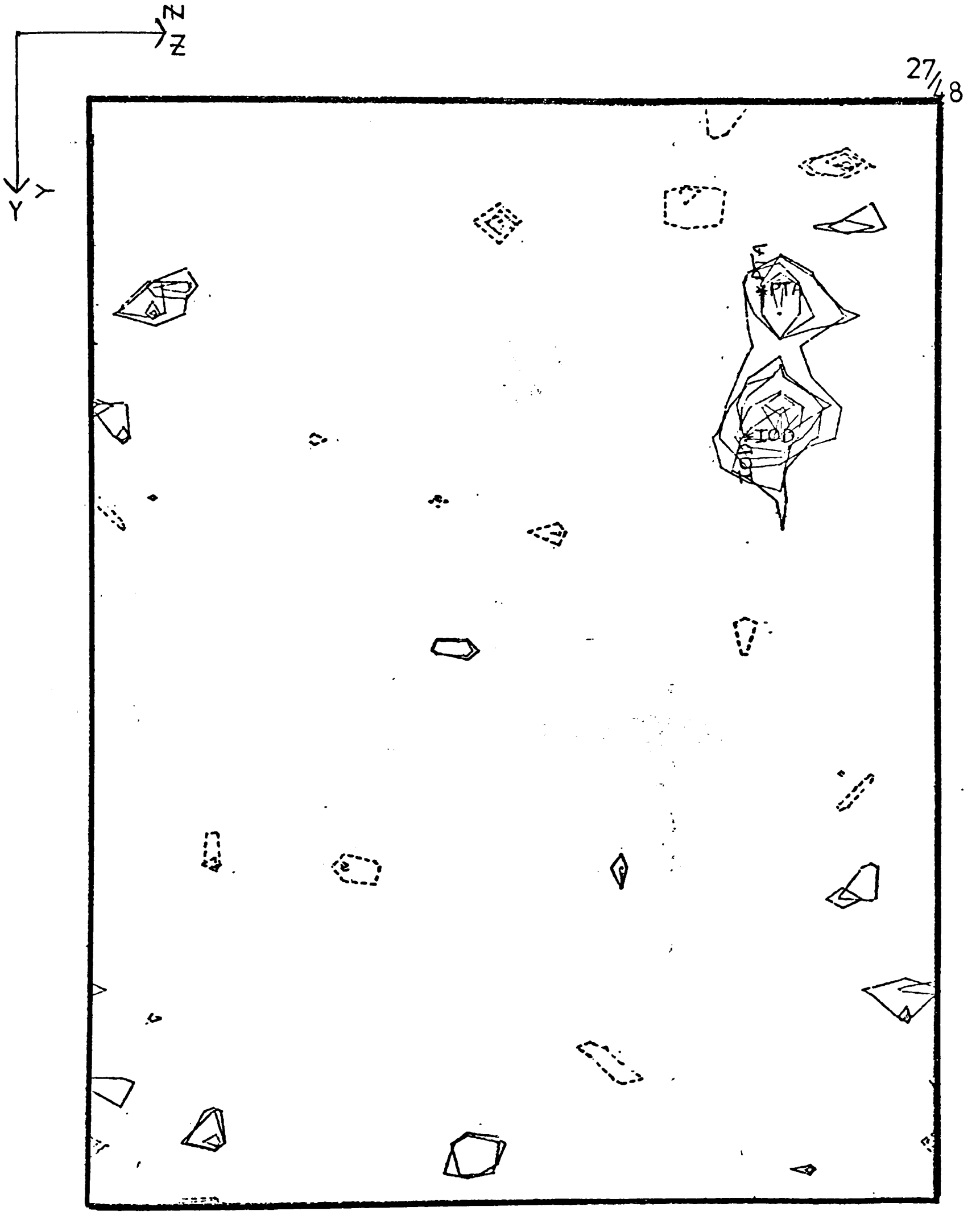


Figure 4.4b

Figure 4.5

Composite views for (a) sections Z=19/48→24/48 , Z-axis out of the paper and (b) sections X=7/40→16/40, Xaxis out of the paper , for the difference synthesis calculated with terms in ($IADP^+$- $NADP^+$) $expi\alpha_{best}$, on a scale of 4 Å/cm.

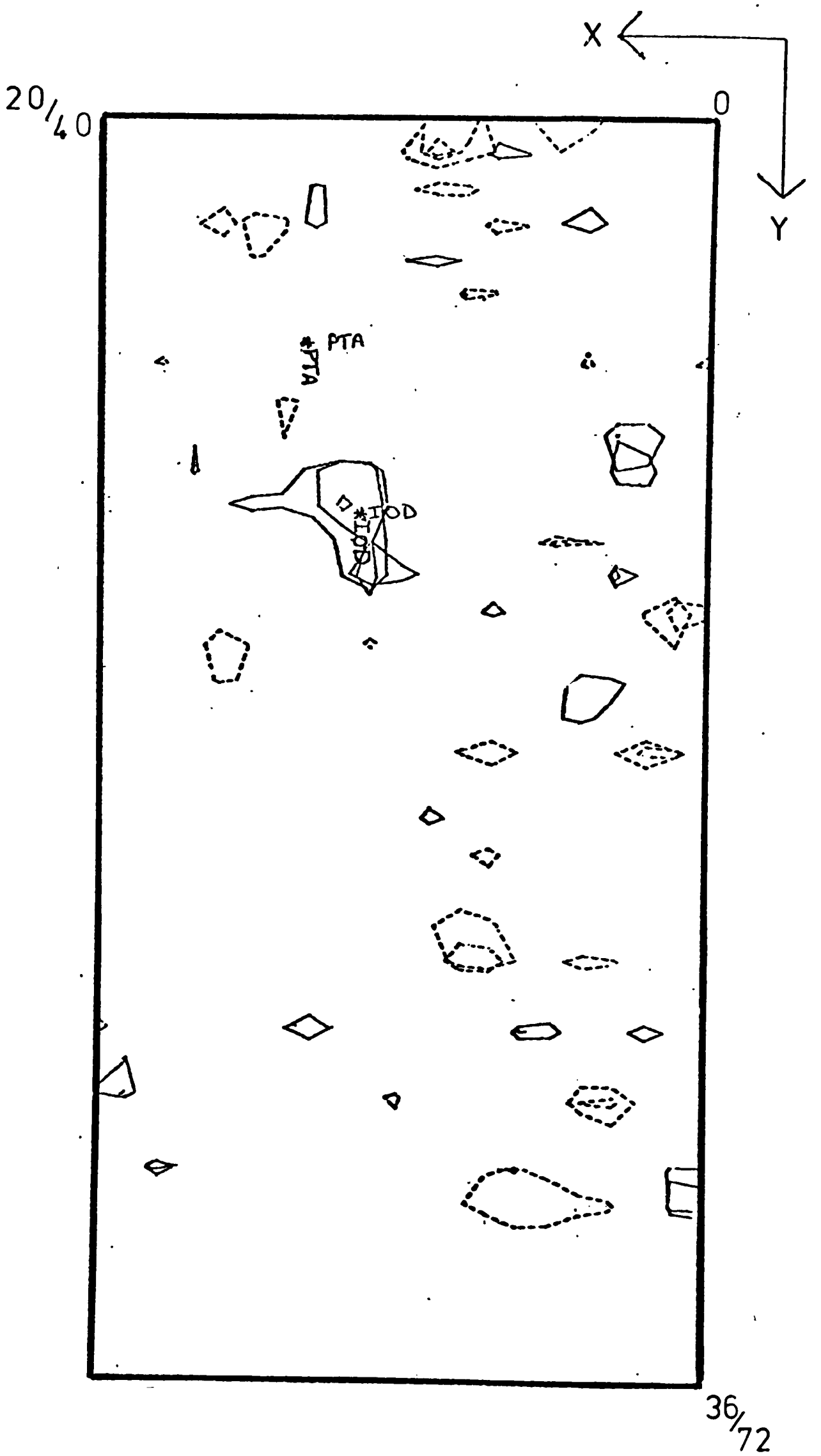


Figure 4.5a

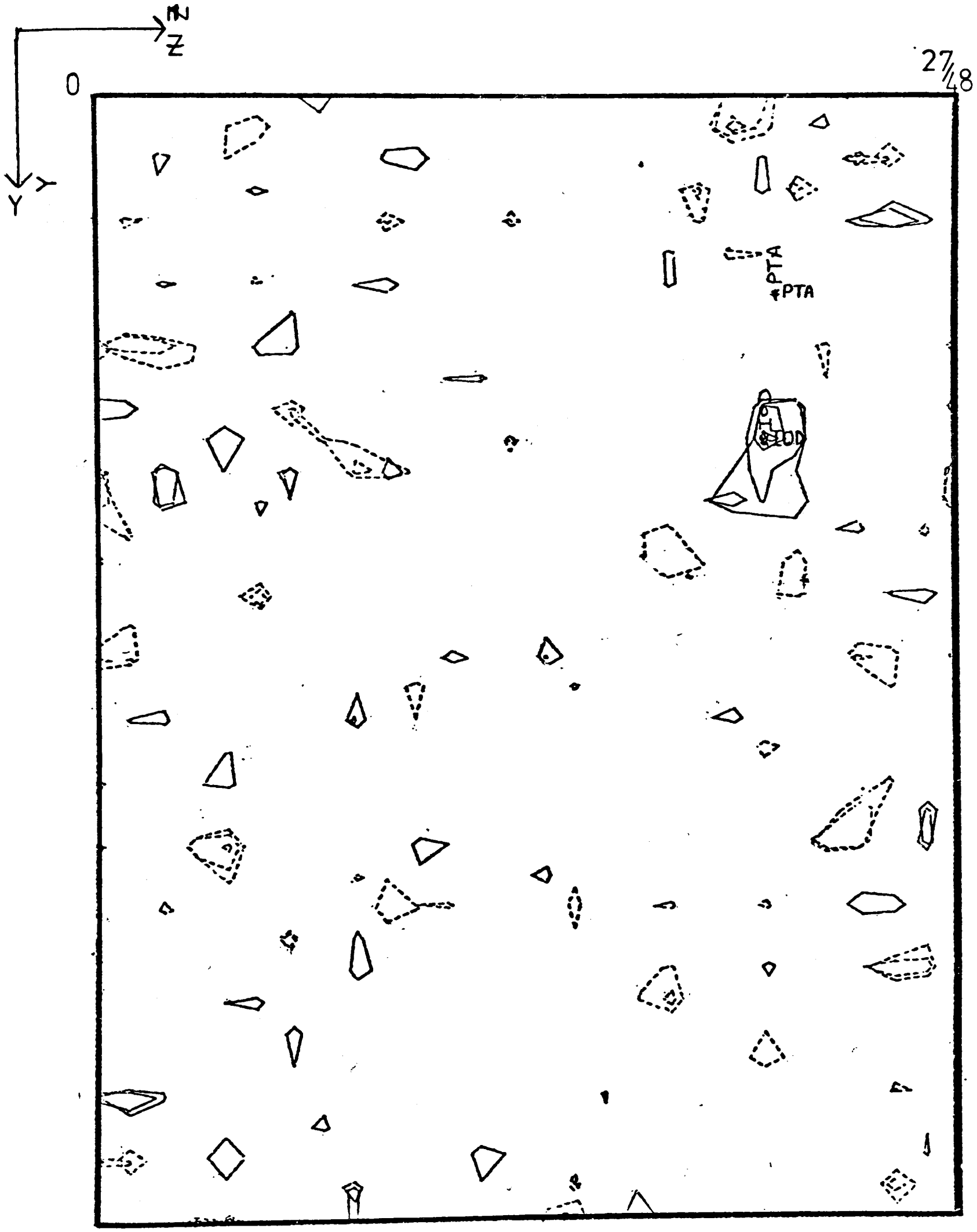


Figure 4.5b

Figure 4.6

Composites of difference Fouriers for $NADP^+$ and $IADP^+$. Protein density (in views down a only) is shaded.

(a) $NADP^+$ - native protein, α_{Au}

(b) $IADP^+$ - native protein, α_{Au}

(c) $IADP^+$ - $NADP^+$, α_{best}

Figure 4.6

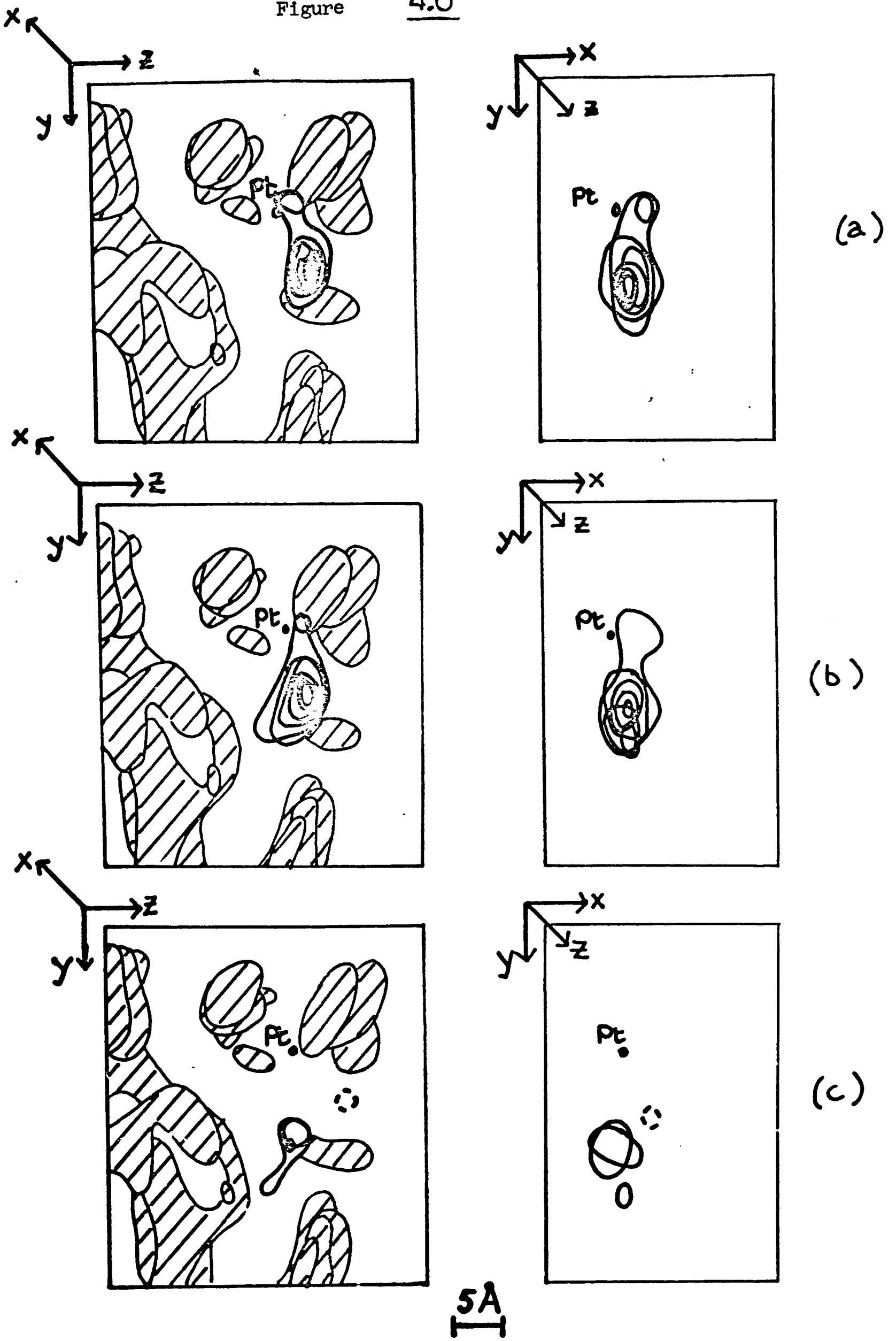


Figure 4.7

Section X = 13/40 compared for the best and Au phased difference maps for the series.

(a) $NADP^+$ - native, α_{best};

(b) $NADP^+$ - native, α_{Au};

(c) $IADP^+$ - native, α_{best};

(d) $IADP^+$ - native, α_{Au};

to illustrate the improvement obtained in the difference electron density by using the double derivative "best" phase set.

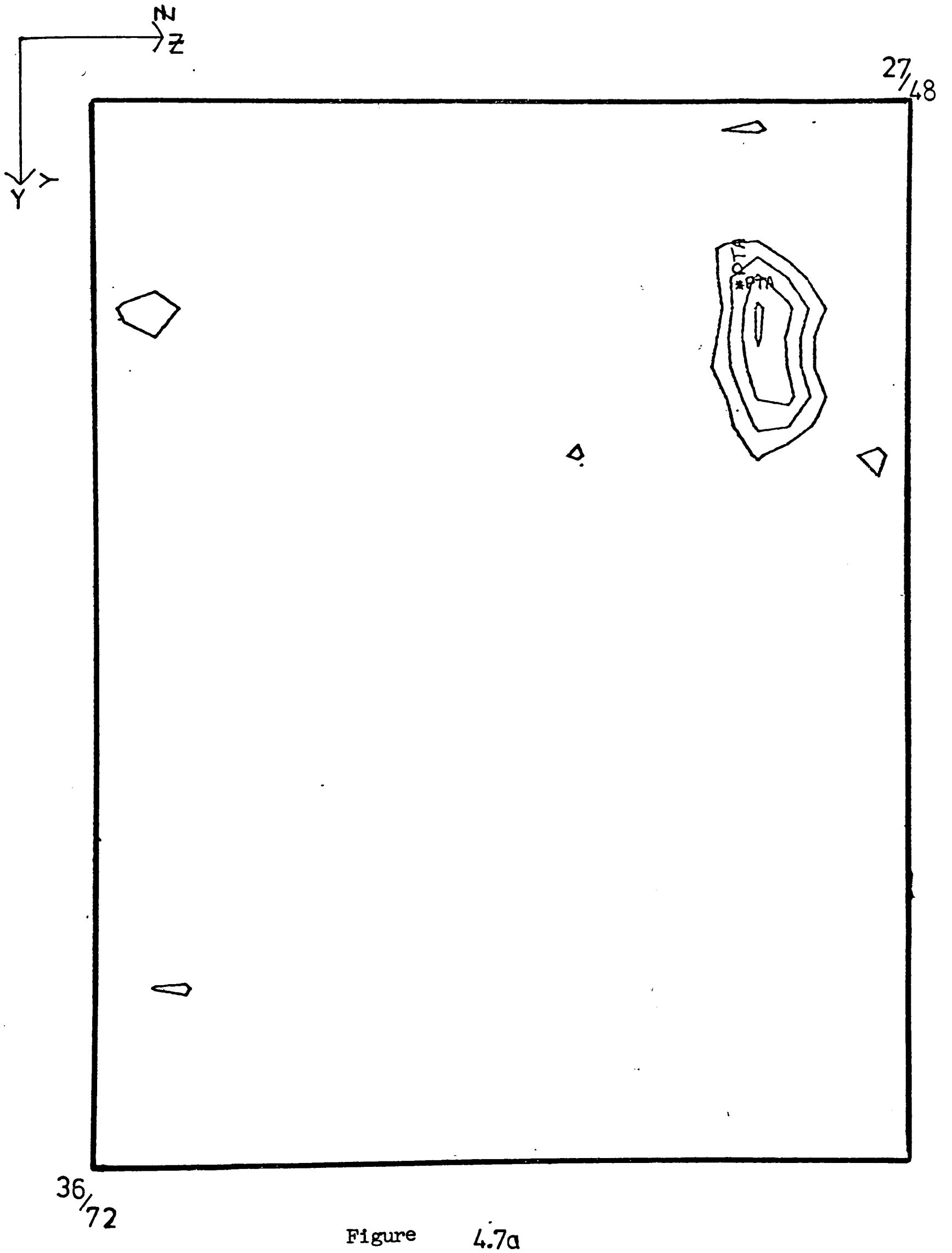


Figure 4.7a

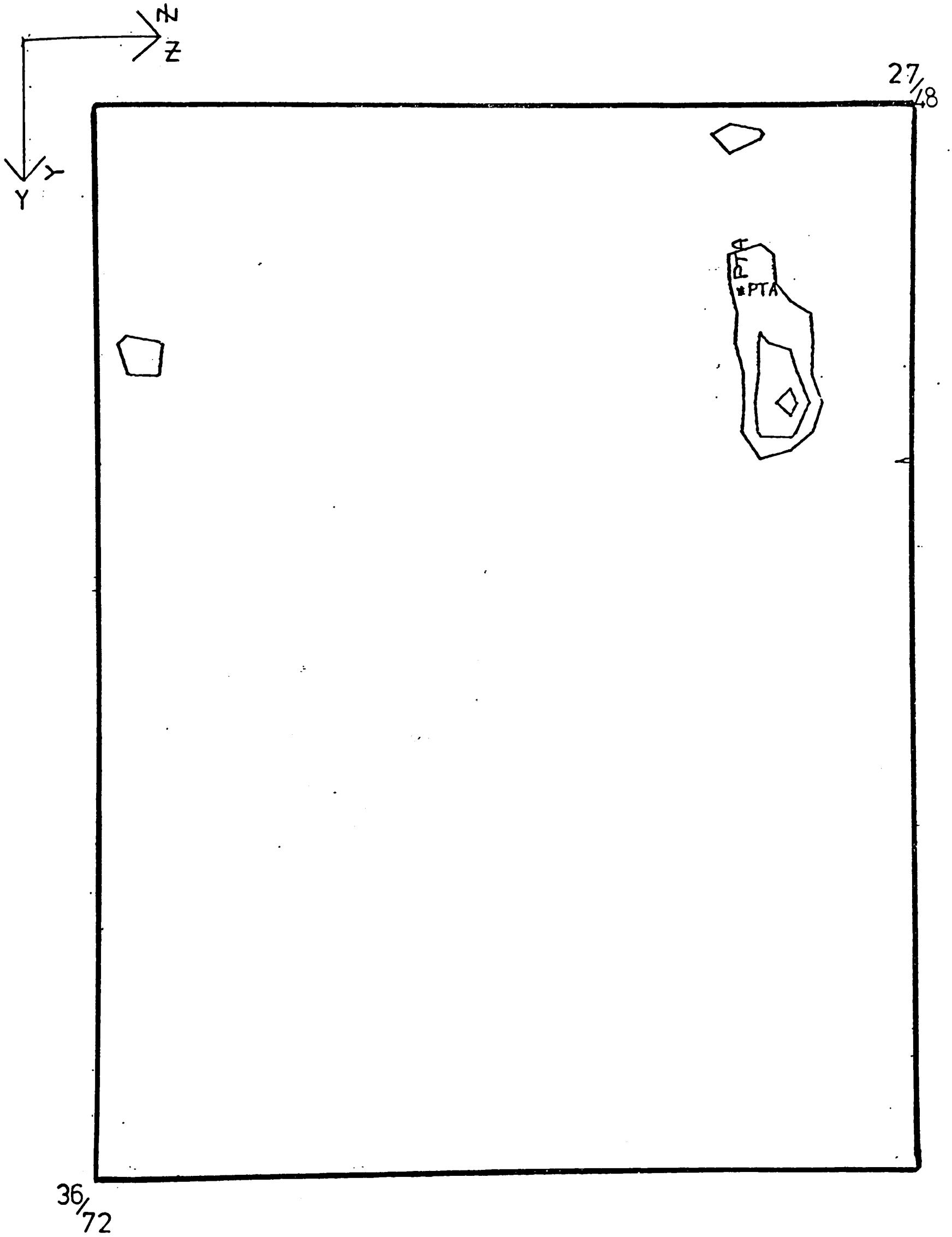


Figure 4.7b

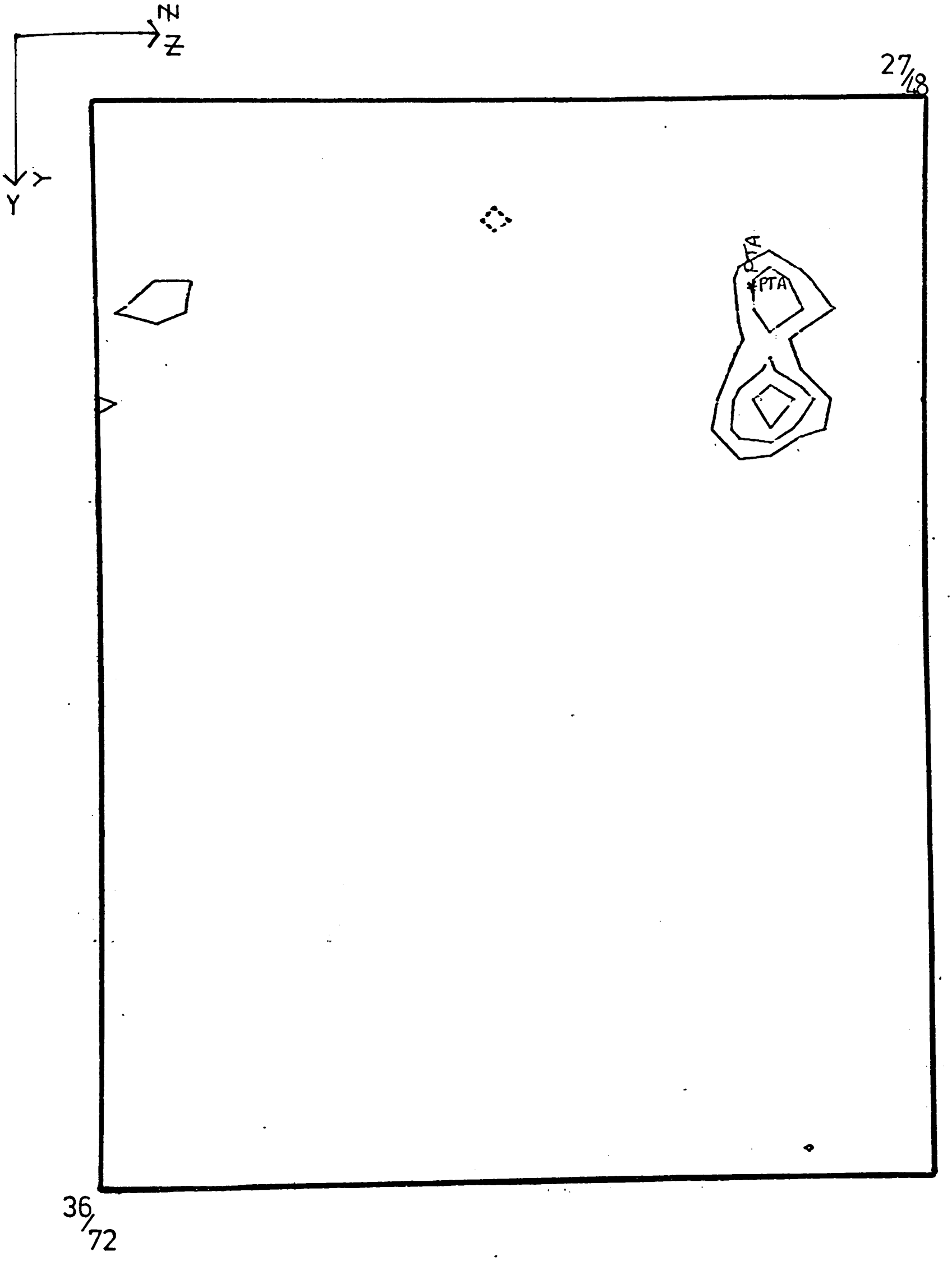


Figure 4.7c

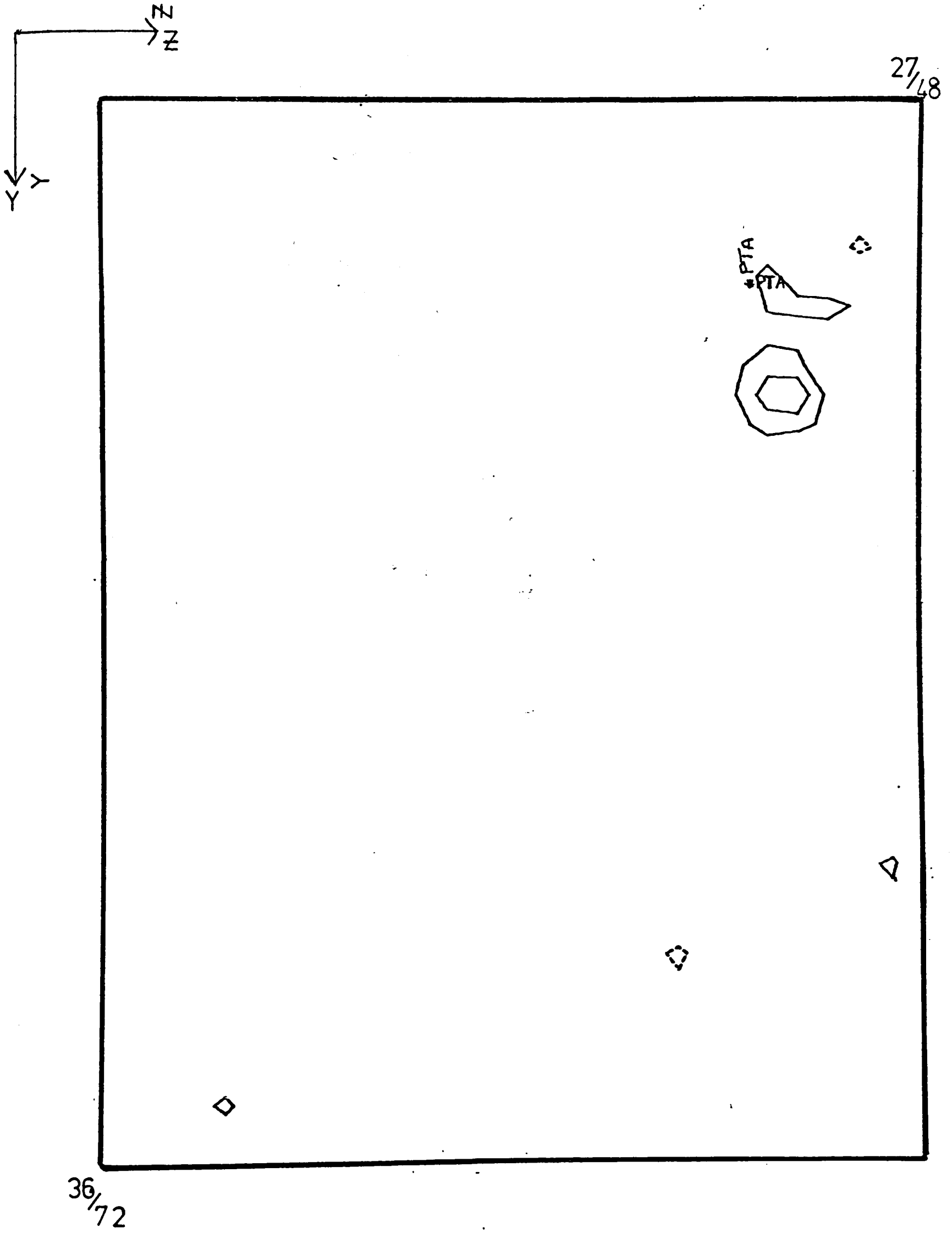


Figure 4.7d



Figure 4.8 A view of the balsa wood model of the protein subunit looking into the cleft, with the darker coloured density corresponding to the $NADP^+$ arrowed. Compare with Figure 3.21.

binds at sites distinct from the coenzyme binding site.

The size and shape of the peak corresponding to $NADP^+$ on binding to 6PGDH indicates that while the coenzyme certainly does not have the bases stacked, it is probably in a slightly less open conformation than that chosen by NAD^+ in LDH (Chandrasekhar et al. (1973), MDH (Webb et al. (1973), GAPDH (Olsen et al. (1976)). The conformation of the coenzyme bound to 6PGDH will only become clear at high resolution.

The orientation of the molecule at one edge of and pointing into the cleft is interesting. It suggests an area of the subunit which might provide a substrate site at the other side of the subunit cleft on the larger part of the subunit. A later study involving the binding of 6PG and NADPH to a 6PGDH crystal forming a ternary complex showed a slightly extended peak beyond that observed for NADPH at the nicotinamide end of the molecule, giving firmer evidence for the position of the substrate site (Clift, 1977).

A centric least squares refinement performed using $|F_{IADP^+} - F_{NADP^+}|$ as FOBS and using ($I\text{-}CONH_2$) electrons to calculate FCALC gave similar results for the proposed iodine site and for a random site, i.e. convergence to low positive occupancy. The random site did, however, have a poorer R factor. The conclusion here was that the scattering of the ($I\text{-}CONH_2$) was not sufficiently strong to give meaningful isomorphous differences for use as FOBS in such a refinement.

4.4 Conclusions and further work

In summary, the following information has been obtained. The coenzyme binding site has been identified and the orientation of the coenzyme with respect to the enzyme determined by studying the binding of the iodinated derivative, $IADP^+$. This served to suggest an area of the subunit which might provide a substrate site. A tentative statement can be made that at this resolution the conformation of the $NADP^+$ in

6PGDH is different to the conformations observed for NAD^+ elsewhere. Conformational changes of the enzyme on binding $NADP^+$ appear to be absent, though curiously there is an indication of such a change in binding the $IADP^+$.

Since the binding at 6 Å resolution of NADPH and 6PG/NADPH to 6PGDH has been studied (Clift (1977)) further low resolution work would involve the binding of 6PG, which yielded smaller intensity differences on precession photographs than the bound coenzyme, and *N*-8-bromo-ADP^+ this molecule has a bromine atom substituted at position 8 of the adenine ring of $NADP^+$ - which would serve to accurately position the adenine moiety of the coenzyme. A sample of N-8-bromo-ADP has been given by Dr. J.-F. Biellmann. Though the conformation of the coenzyme and substrate can only be accurately discerned at higher resolution, the value of this work is two fold.

(i) It has shown that high resolution studies of the coenzyme and substrate binding should be successful since the degree of isomorphism between the crystals is high and that the difference maps should be relatively uncomplicated by large conformational changes of the enzyme.

(ii) It has not been possible to compare the nature of the coenzyme binding site with the "nucleotide binding site" common to the four NAD^+ dependent dehydrogenases and perhaps also to other enzymes (Rossmann et al. 1975). However, using the low resolution difference maps in combination with the high resolution native electron density map, a good deal can be said about the relative position of the $NADP^+$ with respect to the polypeptide chain and a similar comparison made as that by Rossmann et al. (1975) with various other enzymes.

CHAPTER 5

THE STRUCTURE DETERMINATION OF 6PGDH TO 2.6 Å RESOLUTION

This crystal form of 6PGDH possesses one monomer of molecular weight 50000 Daltons in the crystallographic asymmetric unit. There are at 6 Å resolution a total of 1500 and at 2.6 Å 16,000 (approx.) independent reflections, respectively. A structural study of an enzyme can involve the data collection of at least one native set, and heavy atom derivatives and bound coenzyme, substrate and inhibitor data sets. This means the collection and processing of a very large number of reflections. In this regime the data collection method has to be efficient to enable a reasonable amount of data to be collected from one crystal.

In the work on this enzyme at high resolution these considerations precluded the use of the single counter 4 circle diffractometer and of the 5 counter 5 circle diffractometer. The most efficient means of data collection available in terms of the amount of data that could be collected per crystal were the screenless photographic methods discussed in § 2. There are unsatisfactory features associated with the screenless precession method: namely that adjacent films are not contiguous in reciprocal space and upper level spots are split. Several structures have been solved successfully with this method, however, for example, chymotrypsinogen, cytochrome c_2, trypsin-trypsin inhibitor complex, but in each case the agreement factors between equivalent reflections seemed considerably poorer than for diffractometer data (7-9% compared to 4-6% respectively), (Blundell and Johnson (1976)).

The screenless rotation method suffers none of these problems and typically the agreement factors between equivalent reflections are as

good, if not better, than diffractometer data. Since the data from 6PGDH crystals were expected to be weak because of the thin plate shape of the crystals, screenless rotation photography was the most attractive method, the basic principles of which are given in the next section.

5.1 The screenless rotation method

The mode of data collection is simple: the crystal is rotated about a direct axis which is in the horizontal plane and perpendicular to the X-ray beam. The reciprocal lattice points passing through the Ewald sphere can be recorded to a resolution specified by the distance from the film to the crystal (Figure 5.1, Table 5.1).

The total range of rotation which it is necessary to move the crystal in order to record the whole of reciprocal space is $(180^{\circ}+2\theta_{max})$ (see Figure 5.2). When the diffraction pattern possesses high symmetry the range required to record a complete unique data set is reduced (for 6PGDH $C222_1$ crystals the diffraction pattern has mmm symmetry and hence a rotation of 90° is sufficient to yield a data set of four equivalents, see Figure 5.3). Such data sets are in fact not quite complete since rotation about a single axis leaves a small region of reciprocal space unaccessed by the surface of the Ewald sphere. This "blind region" has a volume V_B which increases with the resolution of the data to be collected.

$$V_B = 2\pi \left[\frac{2}{3} \int'^{3}_{max} + 2 \int'_{max} - \int'^{2}_{max} \int'_{max} - \frac{1}{2} \sin\left(2\sin^{-1}\left(\int'_{max} \right)\right) - \sin^{-1}\left(\int'_{max} \right) \right]$$

where $$\int'^{2}_{max} = \left[1 - \frac{(2 - \int'^{2}_{max})^2}{4} \right]$$

(see Figs. 5.4, 5.5), (from Mosley (1973)).

The fraction of the data set which is inaccessible because of the blind region is $f = \frac{V_B}{V_S}$

Figure 5.1 The screenless rotation method - relationship between the film and coordinate system and the basic geometry (Z is parallel to the rotation axis) for a flat plate cassette. P' has been selected as a spot at the limiting edge of the film in the cassette (note: $\tan 2\theta_{max} = R_{max}/D$). The cylindrical polar coordinates (ξ, ζ, τ) of P are also shown.

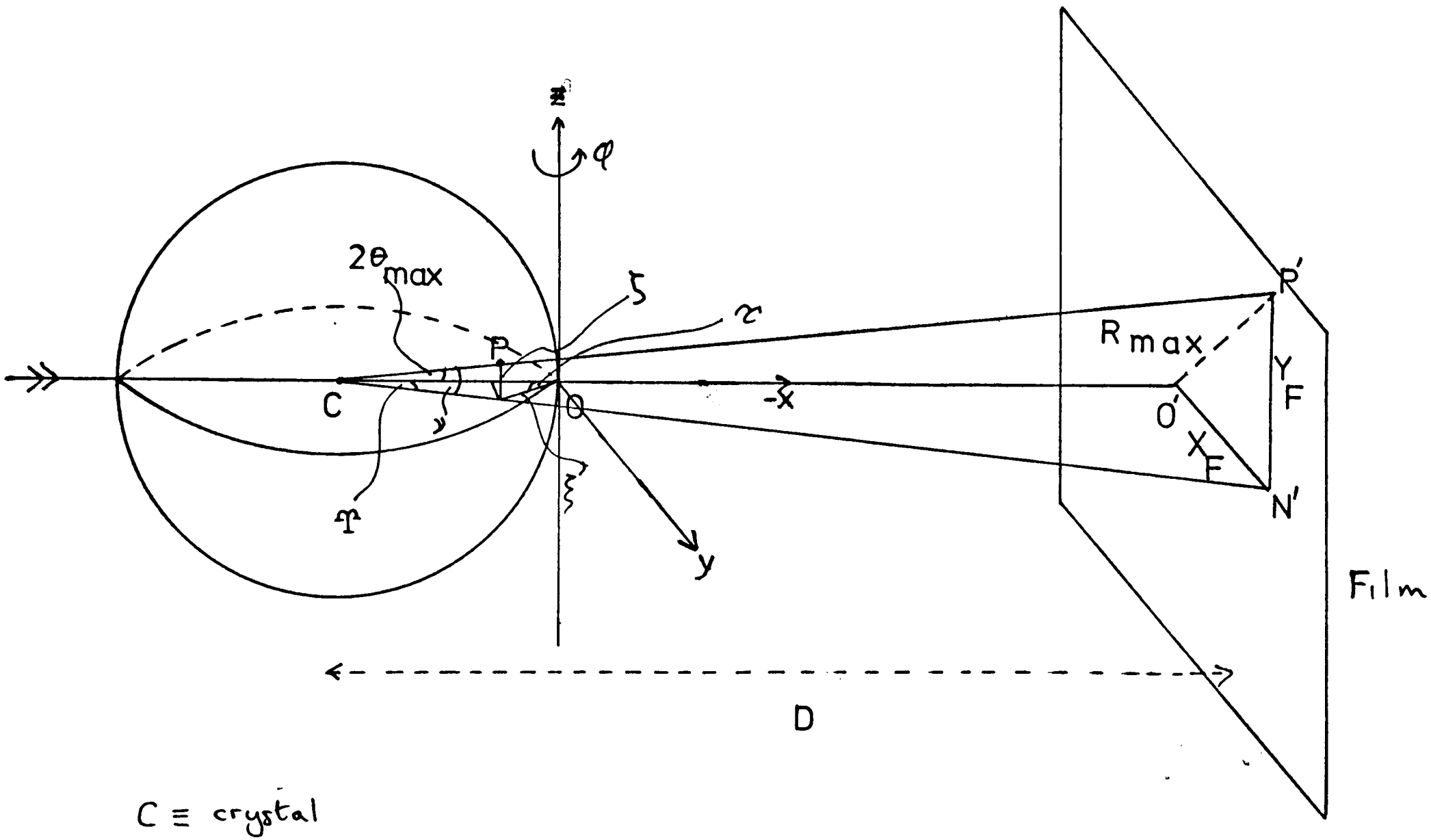


Table 5.1

Resolution of data intercepted by flat plate cassette for different crystal-film distances

D (mm)	$2\theta_{max}$ (°)	$d_{resn.}$ (Å)
60	44.5	2.0
80	36.4	2.5
85	34.8	2.6
96	31.6	2.8
100	30.5	2.9
120	26.2	3.4
140	22.9	3.9
150	21.5	4.1

Note: $2\theta_{max} = \tan^{-1} R_{max}/D$

$d_{resn.} = \lambda/2\sin\theta_{max}$

$R_{max} = 59$ mm.

Figure 5.2 Rotation of the Ewald sphere through 180° about the rotation axis c. A further rotation of $2\theta_{max}$ is required to complete the sampling of reciprocal space within the resolution sphere.

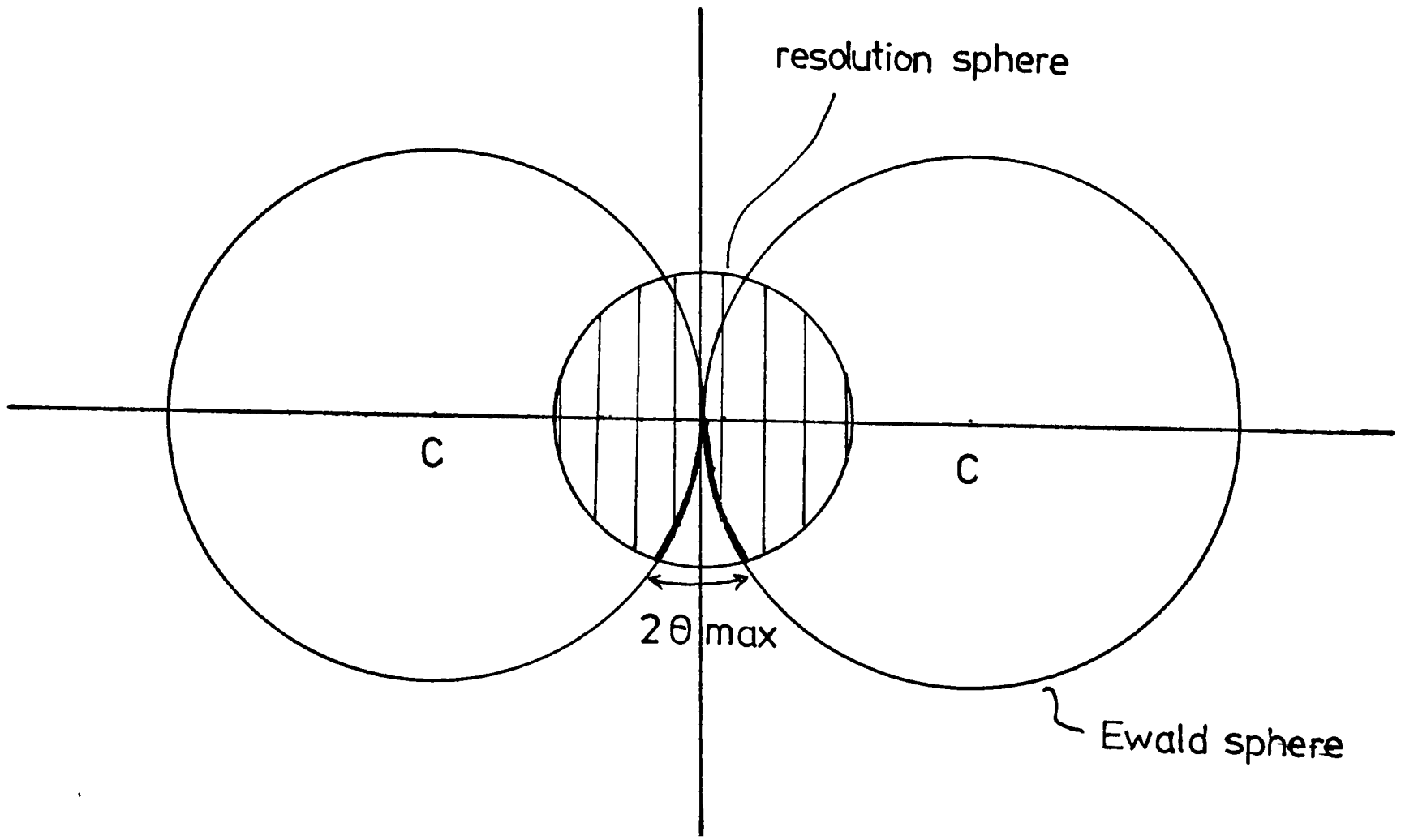
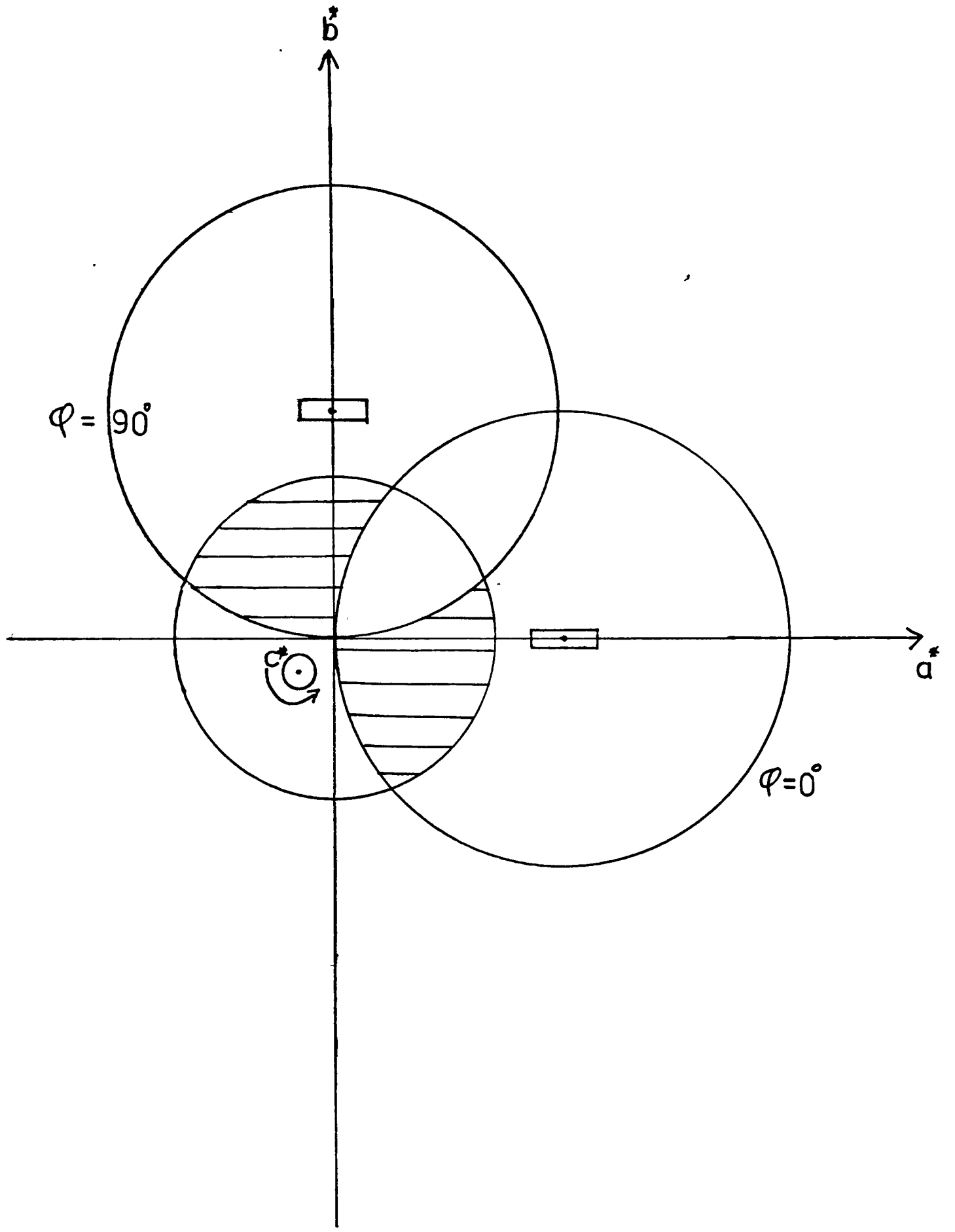


Figure 5.3 For 6PGDH C222$_1$ crystals a total rotation of 90° yields a complete set of four equivalent reflections (except for the blind region). The plate shape of the crystals is illustrated.



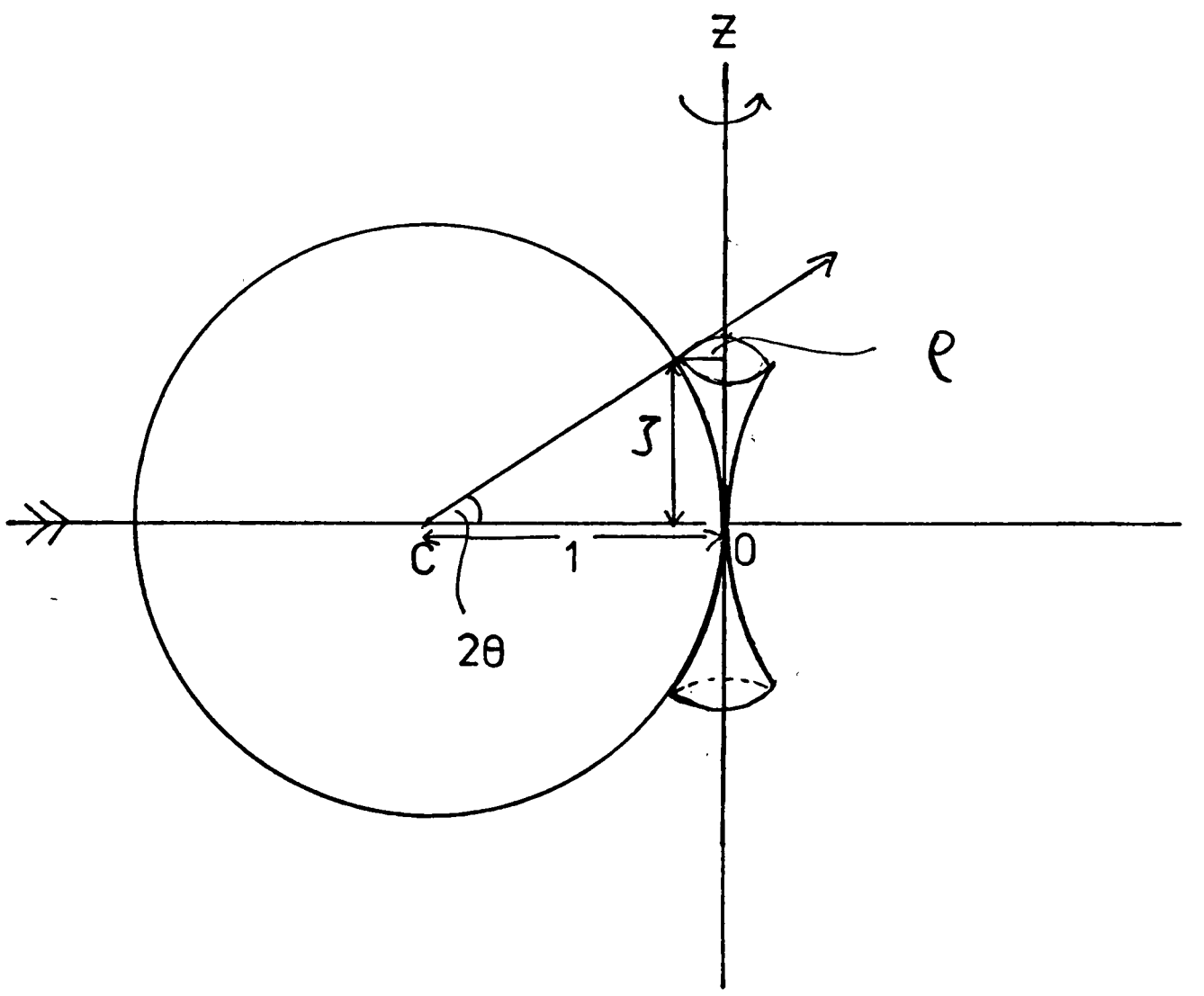


Figure 5.4

The blind region, about the rotation axis Z, of radius:

$$\rho = 1 - (1-J^2)^{\frac{1}{2}}$$

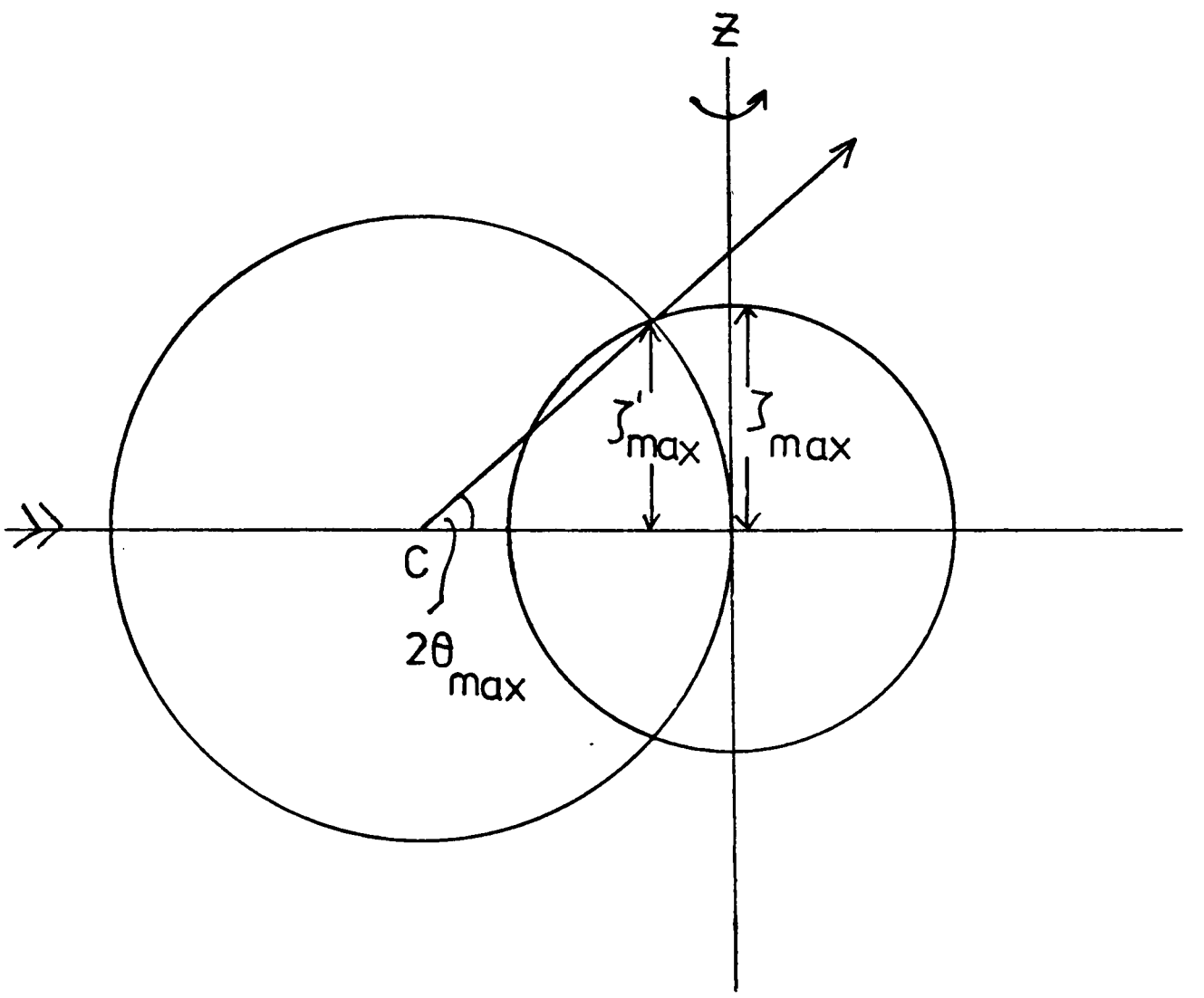


Figure 5.5

The limit of diffraction :definition of ζ'_{max}, ζ_{max}.

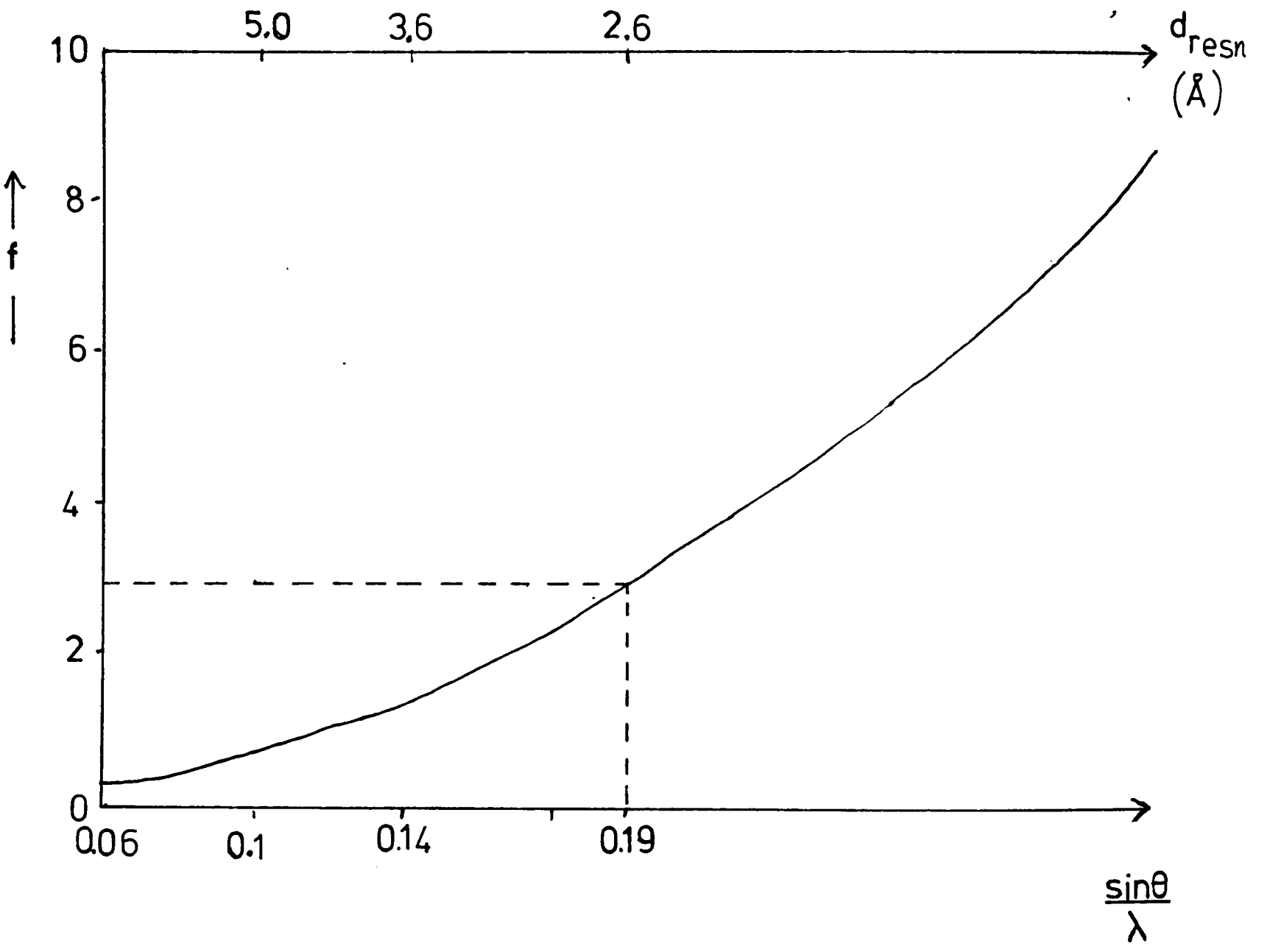


Figure 5.6. Variation of f (% of blind region volume to Ewald sphere volume) with sin θ/λ and dresn (from Mosley (1973)).

where $V_S = \frac{4}{3}\pi \int_{max}^{3}$.

The variation of f is shown in Figure 5.6 (from Mosley (1973)).

The reflections in the blind region are recorded by rotating the crystal about a different real axis or by supplanting the data set with screened precession photographs of zones parallel to the axis of rotation.

For the rotation geometry the Lorentzfactor is given by:

$L = 1/(\sin^2 2\theta - \int^2)^{\frac{1}{2}}$

and for reflections with $\sin 2\theta \approx \int$, $L \rightarrow \infty$ i.e. reflections lying on the blind region surface remain in the diffracting position for all rotation about z. These reflections are discarded in subsequent data processing.

The size of rotation of the crystal for each data film (the rotation range) is determined by the density of reciprocal lattice points and the angular range of reflection for the specimen. The angular range of reflection is the sum of the cross fire in the incident beam and the mosaic spread of the crystal. If the rotation range is too large reflections will overlap (Figure 5.7) and data lost (§ 5.3), and if it is too small, very few reflections will be fully recorded, i.e. pass completely through the Ewald sphere. When a reciprocal lattice point (considered to be of finite size) does not pass through the sphere completely, it is said to be partially recorded. Provided the motor incrementing the rotation angle (the stepping motor) is sufficiently accurate and free from backlash, partial reflections (partials) from one film to the adjacent one are additive because the corresponding part of the reciprocal space is contiguous. Mosley (1973, p. 116) describes a prototype single cassette rotation camera where the backlash of the stepping motor was as much as 15' of arc. Since the reflecting range of a crystal is typically 20', partials would not be additive. A rotation camera with better tolerances on the motors is described by Arndt et al.(1973). With this camera every reverse movement of the crystal

(* for pure monochromatic radiation)

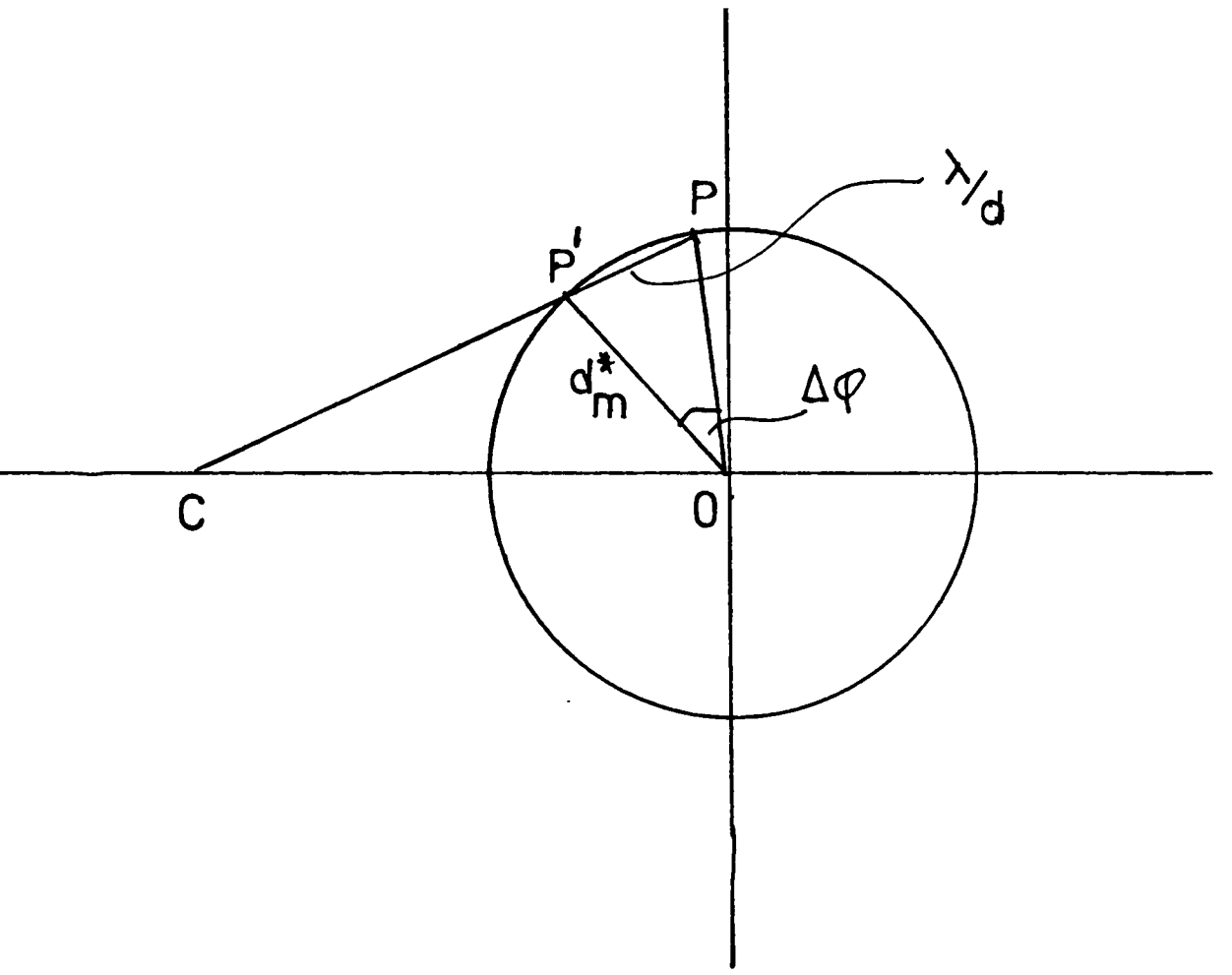


Figure 5.7

Geometric construction to determine the maximum permissible rotation angle ($\Delta\varphi$) for a given crystal orientation (φ). P and P' represent the closest reciprocal lattice points in reciprocal space, for a given axial mount, that could overlap and are separated by λ/d. The radius of the resolution sphere is d^*_m and $d^*_m = \lambda/d_{resn}$.

shaft is followed by an overshoot and a slow forward approach to the starting position to eliminate backlash effects. The camera incorporates a rotating cassette changer mounted on a separate optical carriage for exposing in series up to 8 different cassettes, making data collection essentially automatic. This instrument is marketed by Enraf-Nonius Ltd. and is known as the Arndt-Wonacott rotation/oscillation camera. The camera is controlled by an electronic sequencing circuit constructed from standard digital logic integrated circuits. Parameters such as the exposure time (including the number of oscillations per exposure), the rotation range and angle, plus the degree of overlap between successive films can be selected.

With screenless photography the collimation of the X-ray beam is very critical since any scattering arising from the collimator will serve to increase the general background level on the film. Calculations seeking to optimise collimation conditions have been dealt with by Huxley (1953) and Arndt and Wonacott (1977) in which the following parameters are considered as variables: a, f, s, D, where a is the dimension of a cubic crystal, f is the focus size, s is the source to crystal distance, and D the crystal to film distance, (Figure 5.8). The source collimator is designed so that all points of the crystal may see all points of the X-ray source, whilst avoiding any radiation scattered within the collimator from reaching the film.

Of these four variables, a and f are usually relatively invariant, D is determined by the resolution of data required, and so s is the only parameter that can be optimised. When the source to crystal distance on the laboratory rotation camera was changed from 139mm to 80mm the spot intensities were darker though the spot size was greater.

It is a requirement of the method that the orientation of the crystal be accurately determined for the prediction of reflections on the

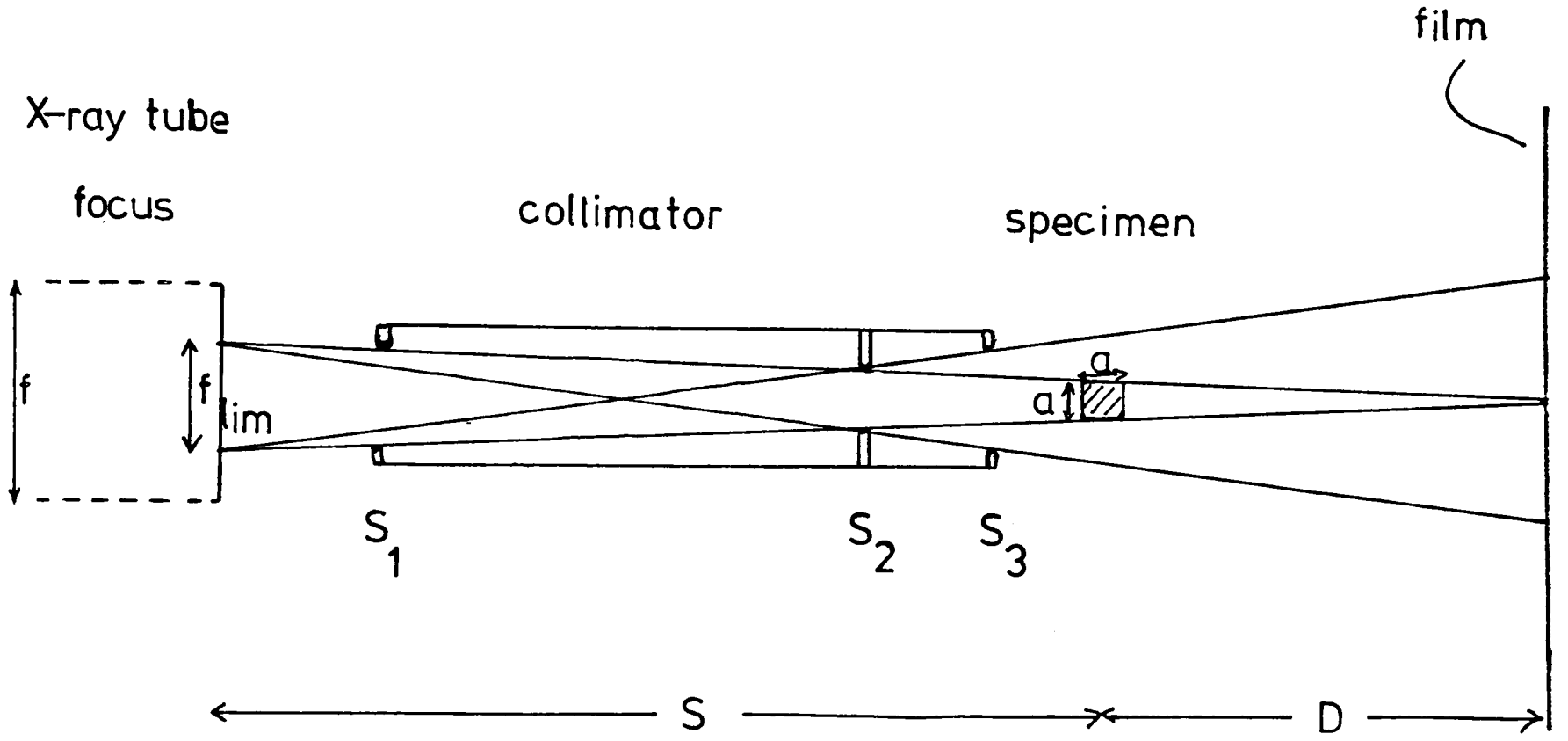


Figure 5.8 Collimation geometry (after Huxley 1953) where f_{lim} is the useable size of the focus and is determined by S_1; S_2 forms the limiting aperture which restricts the beam to the crystal and S_3 is a guard aperture which prevents radiation scattered from S_2 from reaching the crystal. The other symbols are defined in the text.

film for data processing (see § 5.2). This can be achieved by taking setting photographs 90° apart and using the degree of recording of a set of partial reflections on each photograph as an accurate monitor of the crystal orientation. Together with the coordinates of these partials on a film, a least squares refinement optimising the setting angles, crystal to film distance and reciprocal cell parameters can be performed. The setting photographs are generally taken at the beginning and end of data collection. A radiation damage photograph has also to be recorded and is a retake of the first data photograph for the crystal. This photograph serves as a check for crystal slippage. The repeat of the setting photographs at the very end of data collection serves to check that the initial setting photographs correspond to the crystal orientation during data collection, i.e. that the crystal did not slip between the setting photographs and the first data photograph.

A correction for absorption can be estimated for each reflection. If the crystal is bigger than the beam in all dimensions, the direct beam, suitably attenuated, can be monitored as a function of φ and μ (where μ is the angle of rotation about an axis perpendicular to the X-rays and the rotation axis, φ). Such a rotation is possible in increments of 15° for μ from 0° to 45° on the Arndt-Wonacott camera. From such measurements it is possible to evaluate a transmission surface for a particular crystal and calculate an absorption correction for each reflection by interpolation on that surface (Schwager, Bartels, and Huber (1973)).

When the entire crystal is bathed in the X-ray beam the transmission is best determined on a 4 circle diffractometer by selecting a series of suitable low angle reflections and measuring their intensities as a function of azimuthal angle (see § 5.2f).

5.2 Film processing

The information on a given film has to be reduced to digital form and this is achieved by means of a densitometer which measures the optical density of the blackened film given by:

$$O.D. = \log \frac{I_o'}{I_t'}$$

where I_o' is the intensity of the incident light and I_t' the intensity of light transmitted by the blackened spot on the X-ray film. The revival of interest in photographic techniques in recent years has been due partly to the introduction of fully automatic microdensitometers. The densitometer in use in this laboratory is a Photoscan P1000 (Optronics Inc.), a drum type densitometer, which is interfaced to a 7 track Kennedy magnetic tape deck and operates in an off line mode.

The film is placed over a hole in a cylindrical drum which can rotate. Light is transmitted through the film with a light beam of diameter 50, 100 or 200 μ focussed at the film and the optical density recorded on a scale from 0 to 255 where 255 corresponds to an optical density of 2 or 3 O.D. units.

Nockolds and Kretsinger (1970) found that a similar densitometer was linear up to a value of O.D. 1.9 on Ilford Industrial G X-ray film with the 200μ raster. In this work the non linearity was accounted for by applying a correction factor to the O.D. values derived from an empirical curve. This was evaluated from measurements of a film in which the main X-ray beam is recorded for a series of fixed exposure times.

The integrated intensity for a reflection on a film is obtained by summing the O.D. readings within a small region containing the reflection and subtracting from this sum the estimate of the background intensity in the same region.

For rotation photographs there are several other systematic corrections other than the non-linearity correction that need to be applied (see Figure 5.9):

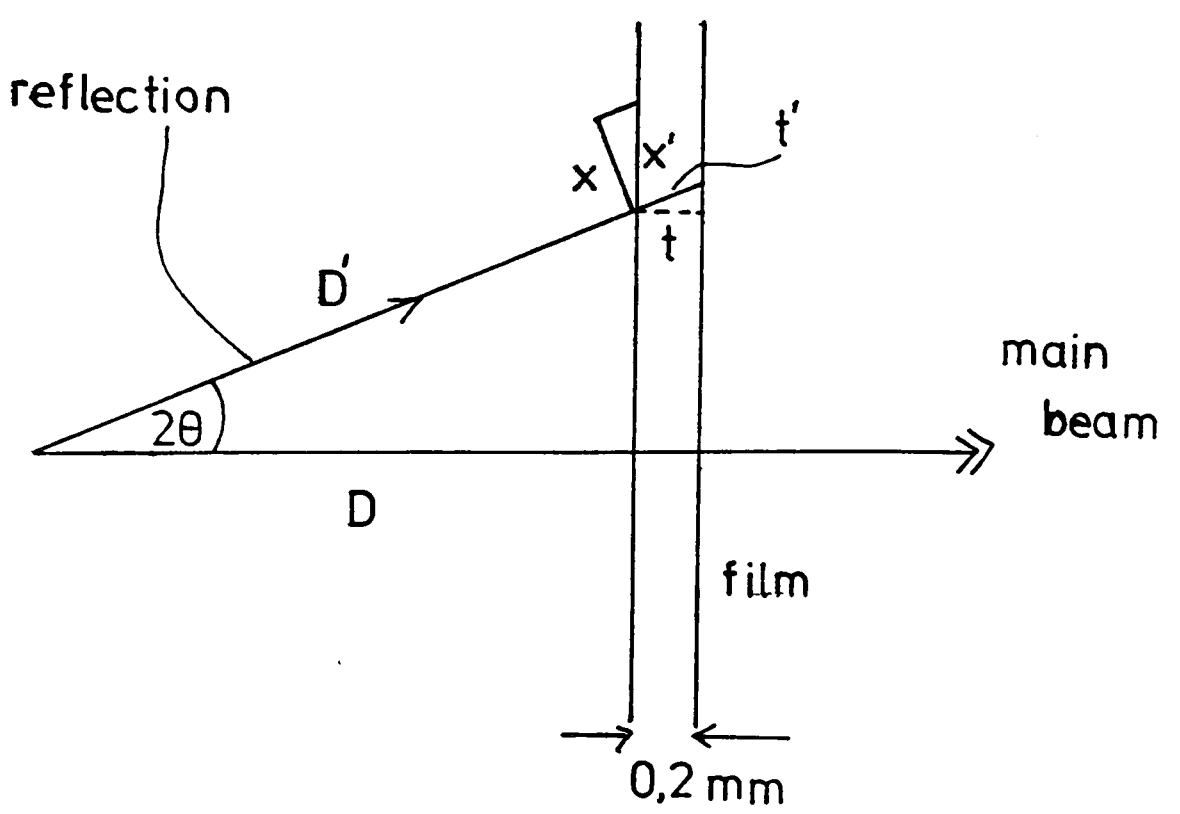


Figure 5.9 Oblique incidence absorption in X-ray film; $t' = t \sec 2\theta$. Note that (i) $D' = D \sec 2\theta$, implying increased air absorption as 2θ increases; (ii) $x' = x \sec 2\theta$, implying spot size increases as 2θ increases.

a) Oblique incidence absorption: The more widely scattered X-rays suffer greater absorption due to the increase in path length through the film with 2θ and hence intensities recorded on the second film need to be multiplied by $\exp(\mu t')$ where $t' = t \sec 2\theta$ and t is the film thickness.

b) Oblique incidence: The more widely scattered X-rays strike the film obliquely so that a thicker layer of the X-ray film is presented to the beam. The number of grains exposed is sec2θ times as great for oblique incidence as for normal incidence. A correction factor of the type:

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

where $I(2\theta)$ is the observed intensity at 2θ and I_{corr} is the corrected value and e^{-A} and e^{-B} are the absorption factors for a normal beam in a single emulsion and in the film base respectively (Cox and Shaw, 1930; Whittaker, 1953).

c) Oblique incidence air absorption: The path length of the X-rays from the crystal to the film through the air varies as $D' = D \sec 2\theta$. The more widely scattered X-rays suffer greater absorption and need to be multiplied by $\exp(\mu_{air} D')$.

d) The increase in D' with 2θ results in an increase in spot size with 2θ which has to be allowed for.

To obtain a set of corrected intensities a software suite incorporating the following features is required.

a) A program which refines the crystal lattice parameters, the missetting angles and the crystal-film distance.

b) A program which generates a set of reflection coordinates in the film coordinate system for a given rotation range and mounting axis.

c) A program which will obtain a transformation matrix relating the predicted film coordinates and the actual coordinates of the reflections

on the film as placed on the Optronics drum. The program calculates an integrated intensity for each reflection and makes various checks for radiation streaks, excessive background scatter by the collimator and for reflections occurring too close together and also applies a correction for the non-linear response of the film and densitometer.

d) A program to scale the two films in a film pack together and apply corrections due to oblique incidence absorption, oblique incidence and LP factor.

e) A program to calculate an appropriate radiation damage correction.

f) A program to calculate an appropriate absorption correction.

In this laboratory, the program package (a-e above) has been developed by Mosley (1973), Yeates and Wilson (1977) (see Appendix IV). Program f has not been implemented but an outline of the calculations that need to be performed is given in this chapter. All the major computer programs used in the course of this work and the input advice to each are on microfiche in Appendix IV.

The computer programs of Mosley were implemented on the Ferranti Argus 500 computer within this laboratory. The advantages claimed for this (Mosley (1973) p. 119) being that:

a) The "hands on" approach was useful when developing the system;

b) the on-line cathode ray tube of the Argus system was very useful for film orientation (equivalent to SEG 3 of the Yeates software package).

The disadvantage of that system lay with the relatively small amount of core store available (12K) which made programming difficult. Also, this program suite was specific to orthogonal space groups and no means were available for refining the crystal lattice and misorientation parameters.

The Yeates software package (OSCAR and AUTOOSCAR) contains all the necessary features for derivation of intensities and is implemented on the Oxford University ICL 1906A computer.

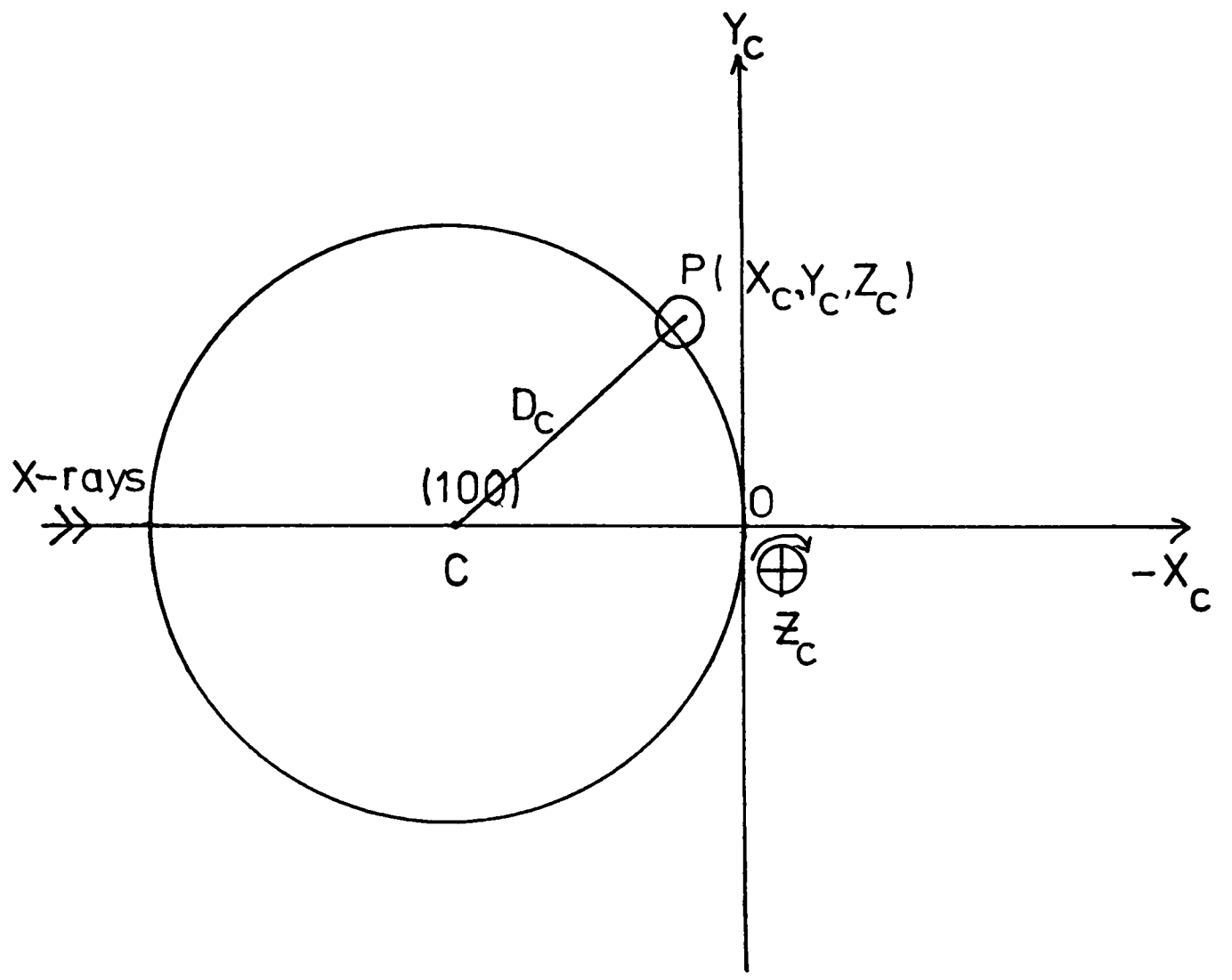


Figure 5.10

Calculation of D_c for a reciprocal lattice "sphere" at (X_C, Y_C, Z_C) where $D_c^2 = (1-X_C)^2+Y_C^2+Z_C^2$.

The program is divided into 4 segments:

(i) SEG1

This section of the program obtains best estimates of the crystal misorientation angles and reciprocal cell parameters by performing a least squares minimisation procedure on D_C and D_O where:

$$D_C = (1 - 2X_C + X_C^2 + Y_C^2 + Z_C^2)^{\frac{1}{2}},$$

see Figure 5.10, and D_O is obtained from the degree of partial recording for each spot. The input required for the program includes a list of partials for a given rotation angle with their respective film coordinates and an estimate of the degree of partial recording, f.

By definition, a partially recorded reflection will lie close to the surface of the Ewald sphere at points corresponding to the ends of the rotation range, the degree to which partial recording occurs, depends on its distance from the surface and upon its size. If γ is the mosaic spread, the radius of each reflection (assumed spherical) is:

$$e(\underline{h}) = \gamma \frac{d^*(h)}{2}$$

and therefore $D_O = 1 + (\frac{1}{2} - a)e$,

where a is calculated from f and accounts for the reciprocal lattice "sphere" being inside or outside the surface of the sphere at one extreme of the rotation range.

The refined misorientation angles are used to generate a misorientation matrix which is used to calculate the coordinates of a reflection $\underline{X}_c$ where

$$\underline{X}_c = [\Phi][S_Z][S_Y][S_X][W]\,\underline{h}$$

$[\Phi]$ is a rotation matrix for a rotation φ° relative to datum.

$[S_Z][S_Y][S_X]$ describes the missetting of the crystal by angular errors $\Delta\varphi_Z, \Delta\varphi_Y, \Delta\varphi_X$, about three orthogonal axes fixed in the crystal and are given by:

$$\begin{pmatrix} \cos\Delta\varphi_z & \sin\Delta\varphi_z & 0 \\ -\sin\Delta\varphi_z & \cos\Delta\varphi_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\Delta\varphi_y & 0 & -\sin\Delta\varphi_y \\ 0 & 1 & 0 \\ \sin\Delta\varphi_y & 0 & \cos\Delta\varphi_y \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Delta\varphi_x & \sin\Delta\varphi_x \\ 0 & -\sin\Delta\varphi_x & \cos\Delta\varphi_x \end{pmatrix}$$

and since, for $\Delta\varphi$ small $\cos\Delta\varphi \approx 1$,

$$[S] = [S_Z][S_Y][S_X] \approx \begin{pmatrix} 1 & \sin\Delta\varphi_Z & -\sin\Delta\varphi_Y \\ -\sin\Delta\varphi_Z & 1 & \sin\Delta\varphi_X \\ \sin\Delta\varphi_Y & -\sin\Delta\varphi_X & 1 \end{pmatrix}$$

which can be used as a rough hand calculation check for the correctness of the computed matrix. $[W]$ transforms from the crystallographic axial system to an orthogonal one defined as in Figure 5.11.

Hence $[W]$ is

$$\begin{pmatrix} a^* \cos\beta^* & b^* \cos\alpha^* & c^* \\ a^* \sin\beta^* & \dfrac{b^*(\cos\gamma^* - \cos\alpha^* \cdot \cos\beta^*)}{\sin\beta^*} & 0 \\ 0 & b^* \cos\gamma & 0 \end{pmatrix}$$

where $\cos\gamma$ is $\lambda / \underline{b}.\underline{b}^*$. The matrix is permuted to allow for the three possible rotation axes.

This section of the program also refines the crystal to film distance between each cycle of the main refinement.

(ii) SEG 2

This section of the program predicts the possible reflections for a given rotation range within the resolution sphere. The reflections generated have to be searched to exclude classes of reflections that are not required:

(a) Those lying too close to the direct beam spot on the film which may have been interfered with by the beam stop.

(b) Those in contact with the Ewald sphere for the whole of the rotation φ_i to φ_f (i.e. whose L factors are effectively ∞).

(c) Those that lie within the blind region.

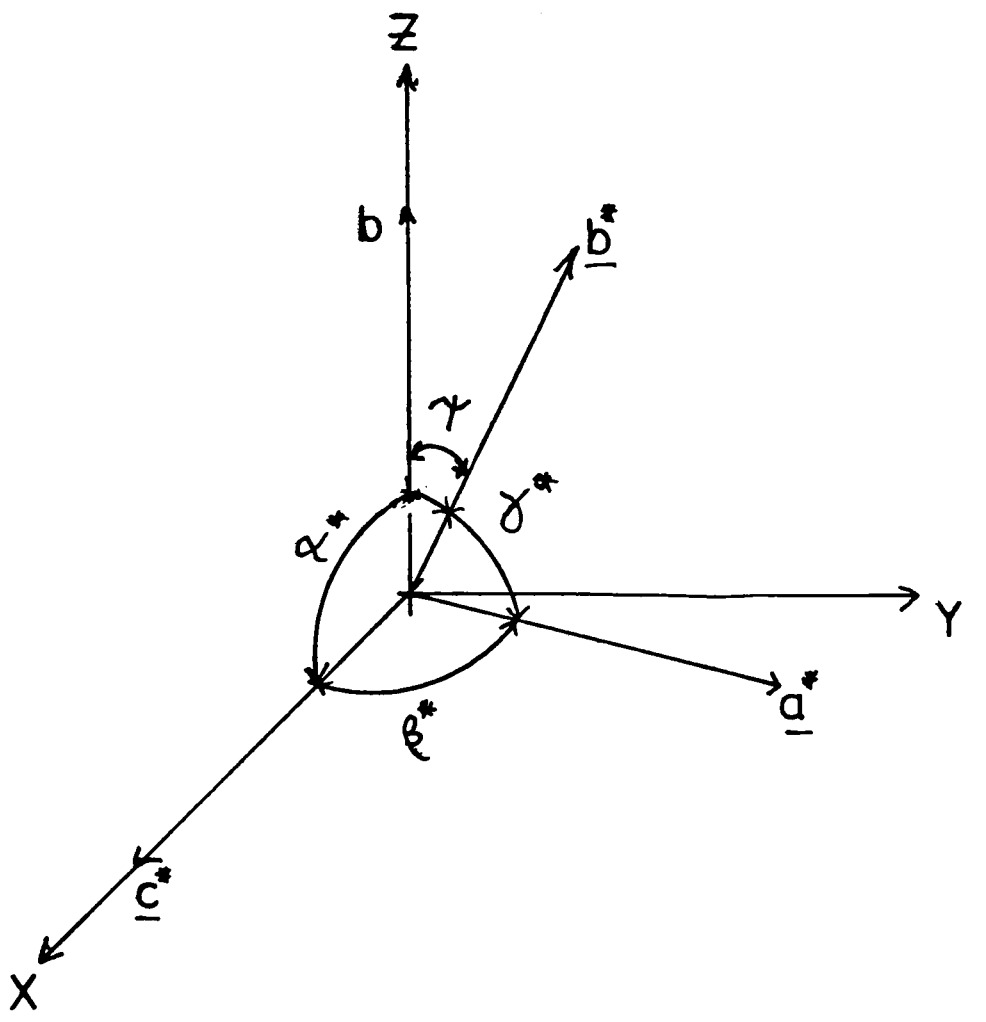


Figure 5.11

Relationship between the crystallographic axial system ($\underline{a}^*$, $\underline{b}^*$, $\underline{c}^*$), and the orthogonal axial system (X, Y, Z) where for example $\underline{X} = ha_x^* + kb_x^* + lc_x^*$ and a_x^* is the projection of $\underline{a}^*$ onto axis X.

The reflections remaining are classified as:

(a) fully recorded;

(b) partially recorded at φ_i;

(c) partially recorded at φ_f;

(d) overlapping.

A list of predicted coordinates X_F for the given hkl, suitably flagged, are then output. A plot of what the film is expected to be like can be output on the 1906A graph plotter.

(iii) SEG 3

The coordinates generated by SEG 2 when transformed into Optronics space will probably not agree with the coordinates of reflections on the film. This is due to the following errors:

(i) Displacement of the film in the $X_F Y_F$ plane. A translation vector $\underline{T}$ needs to be added to the $\underline{X}_F$ to correct for this.

(ii) Rotation of the film in the $X_F Y_F$ plane. A rotation matrix $[R]$ needs to be applied to the $\underline{X}_F$.

(iii) Imprecise crystal-film distance.

(iv) Film not perpendicular to the incident X-ray beam.

(Mosley (1973)).

Errors (i) and (ii) arise from the film as mounted in the Optronics not coinciding with the ideal orientation (defined by the set of $\underline{X}_F$ output from SEG 2), see Figure 5.12.

The true reflection coordinates $X_{OPT} Y_{OPT}$ are given by:

$$X_{OPT} = aX_F^2 + bX_F Y_F + cY_F^2 + dX_F + eY_F + f$$

$$Y_{OPT} = gX_F^2 + iX_F Y_F + jY_F^2 + mX_F + nY_F + p$$

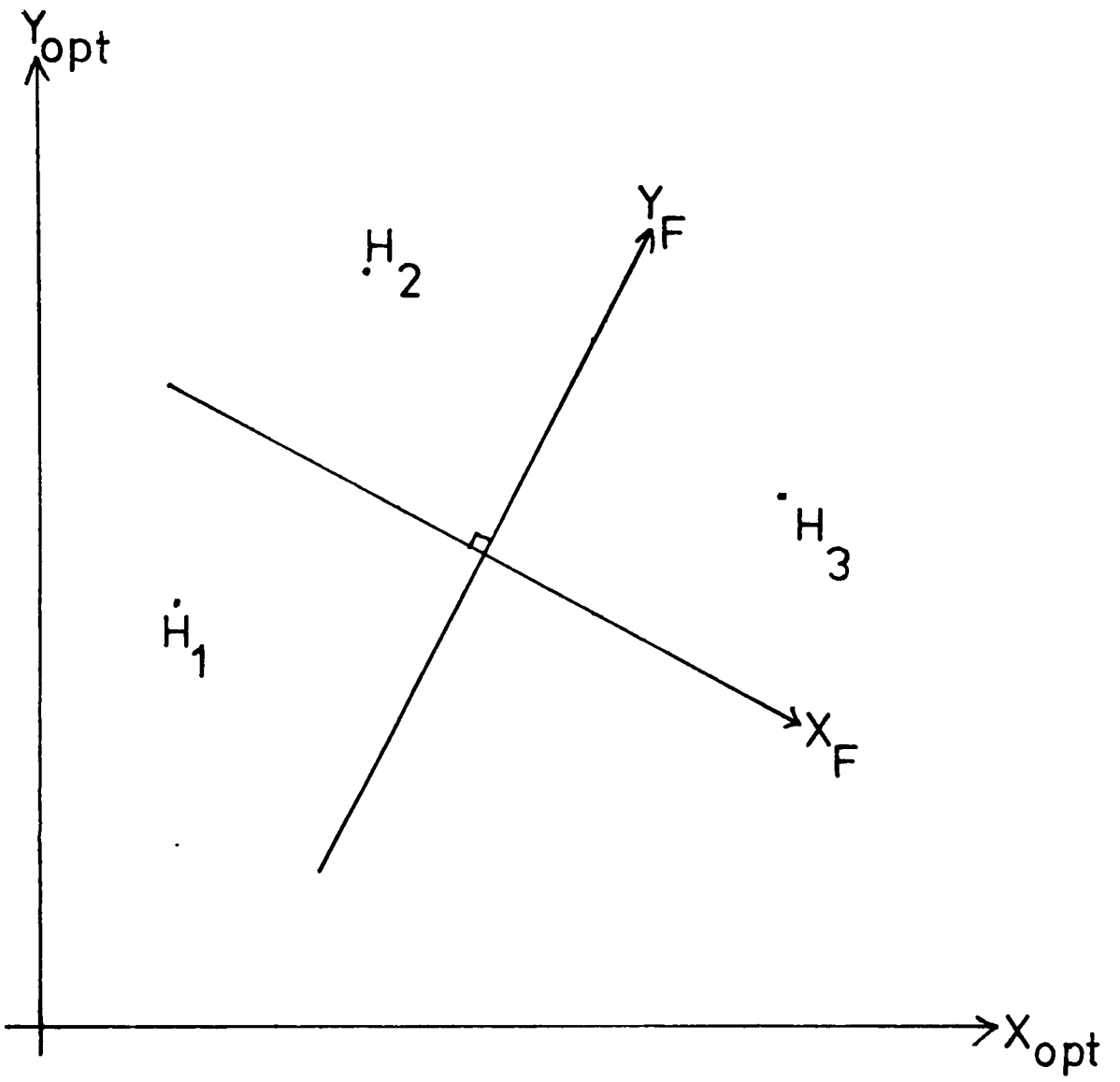
where a - p are constants for a given film, and can be found if the X_{OPT}, Y_{OPT}, X_F, Y_F are known for a minimum of six reflections. For more than six a least squares approach can be used to calculate the constants.

This section of the program requires the digital image of the film on the magnetic tape (MT) produced by the densitometer. Given the

Figure 5.12 Relationship between the film coordinate system (X_F, Y_F) and the optronics coordinate system where:

$$\underline{X}_{opt} = [R]\,\underline{X}_F + \underline{T} \ .$$

X_{opt} is parallel to the axis of rotation of the drum densitometer. H_1, H_2, H_3 are fiducial marks which define the centre of the film and the film axes and are produced by passing an X-ray beam through the rear of the cassette.



coordinates in $\underline{X}_F$ of H_1 and H_2 the program scans the MT to obtain the corresponding coordinates in $\underline{X}_{OPT}$. An initial approximate rotation from $\underline{X}_F$ to $\underline{X}_{OPT}$ can be derived.

Given a set of fully recorded reflections, the equation:

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

can be used to generate their approximate $\underline{X}_{OPT}$. A centre of gravity search is used to centre each reflection given as:

$$\underline{X}_i = \frac{\sum_i \underline{x}_i \bar{D}_i}{\sum_i \underline{x}_i}$$

where $\bar{D}_i$ signifies that only optical densities above some bias level are considered. The respective coordinates for each reflection can be used in a least squares procedure to generate a transformation matrix such that:

$$\underline{X}_{OPT} = [C]\ \underline{X}_F \ .$$

The film is scanned in two halves and after the first half (left hand side) the approximation based on the fiducials is updated to enable initial errors to be corrected for the remainder (right hand side).

The early version of this section of OSCAR required the input of 30 to 40 intense reflection coordinates measured by eye for each film. A later version generated a set of reflections automatically (AUTOOSCAR). This program uses the initial transformation and by scanning the film (left hand side first) selects suitably intense reflections to generate the improved transformation matrix. For both OSCAR and AUTOOSCAR, SEG 3 these reflections needed to be evenly distributed around the film and as close to the resolution limit as possible since in this region inaccuracies in determining the matrix would produce the largest effect. AUTOOSCAR substantially reduced the labour of the process of integrating a film. The set of reflections used to generate the optronics transformation matrix can be listed to line printer or microfiche (Figure 5.13). At this point of the procedure the optimum box size for integration can be selected.

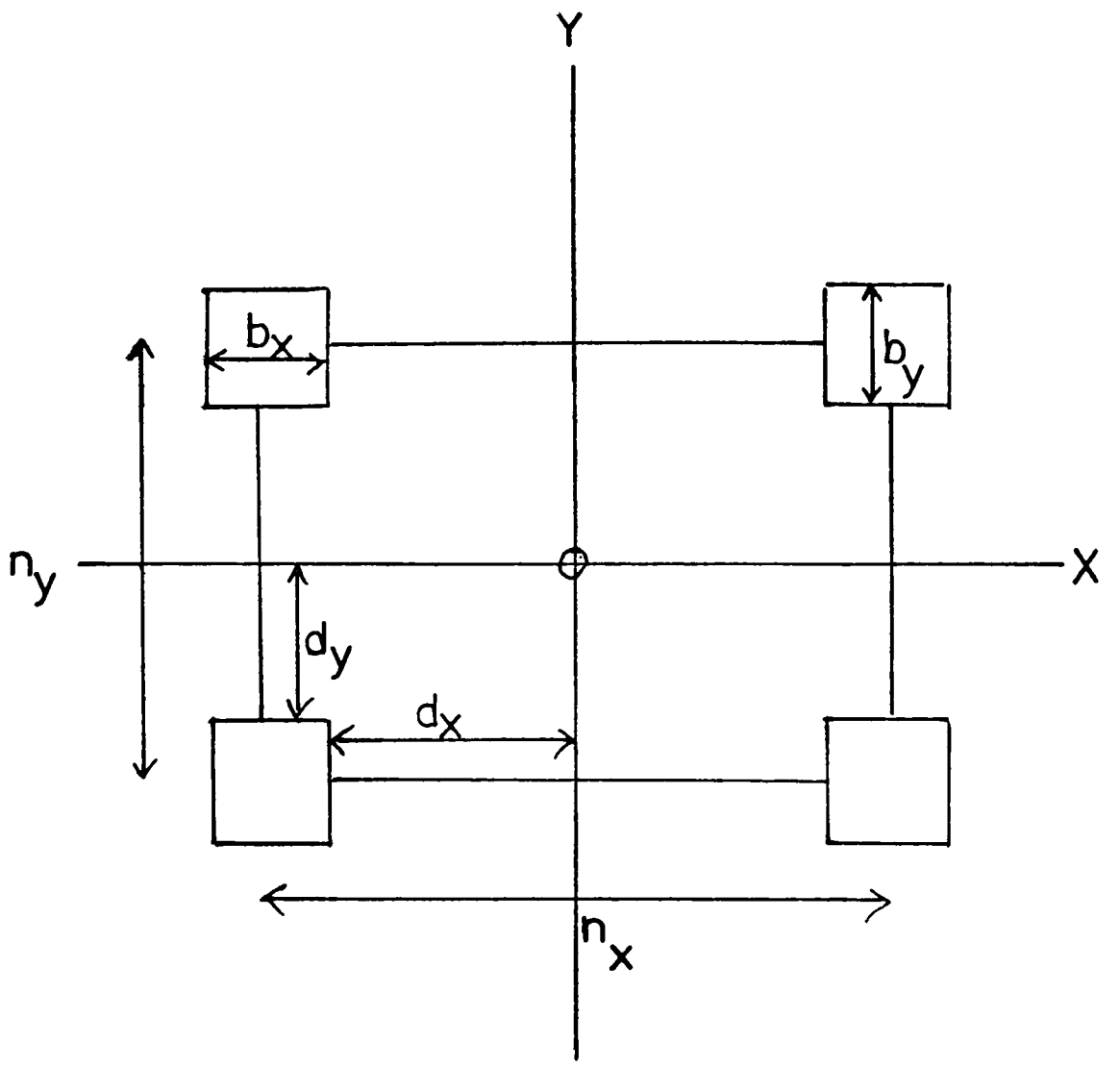
(iv) SEG 4

The transformation derived in SEG 3 is applied to the set of predicted



Figure 5.13 Microfiche of the set of reflections used to generate the Optronics transformation matrix in SEG3 for PT25C, 64-66° data film(A). Without any magnification the spots can be seen centred in a rectangular box. It is at this point of the data processing that the integration box size can be selected.

Figure 5.14 The integration box for SEG4 in OSCAR, the film processing program; n_x, n_y, b_x, b_y, d_x, d_y, are all specifiable parameters.



reflection film coordinates, the resulting list sorted and a single pass made through the scanner data calculating the integrated intensity for each reflection within a specified integration box (Figure 5.14). A correction can be applied for unequal peak and background areas if necessary. Account is also taken of the increase in spot size with 2θ and of the non-linearity associated with the densitometer and film response. After box expansion a check for overlapping peaks is also made and if they do overlap the reflections are rejected.

This software package has recently been further updated for processing data from a cylindrical cassette. The standard detector supplied with the Enraf-Nonius Arndt/Wonacott oscillation camera is a flat plate film cassette and the physical size of the area available for exposure imposes restrictions on the crystal to film distance for data to a specified resolution. In addition the size and shape of the diffraction maxima recorded vary due to oblique incidence as

$$x' = x \sec 2\theta$$

$$y' = y \sec 2\theta$$

(Figure 5.9), where x, y are the dimensions of the peak, assumed rectangular.

Thus it may not be possible to record and resolve high angle data. In order to overcome these problems a set of 4 cylindrical cassettes (total weight equal to eight flat plate cassettes) have been designed and built that are interchangeable with the flat plate cassettes (the radius of curvature is maintained to better than 0.05 mm over an arc of 100°). It is possible to record intensities within a larger region of reciprocal space for the same crystal to film distance (see Figure 5.15).

With this geometry, the peak size varies as

$$x = x$$

$$y = y \sec \chi \quad , \qquad \text{where}$$

χ is the angle between the generator for the n^{th} layer line and the normal to the rotation axis , i.e. the spot shape only changes in the

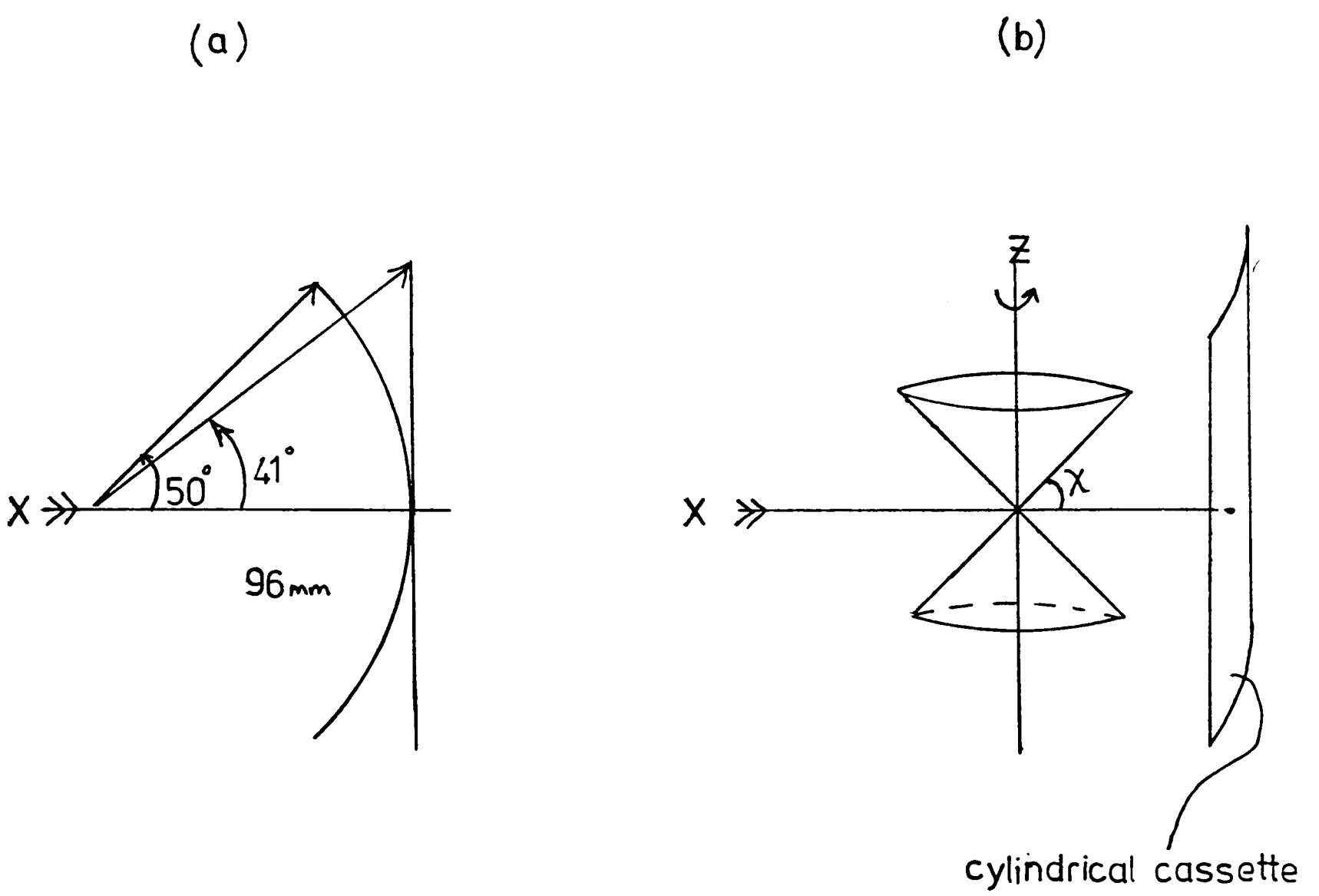


Figure 5.15 The geometry of the cylindrical cassette associated with the oscillation camera. (a) illustrates how the curvature of the cylindrical cassette enables higher angle data to be collected compared with a flat plate cassette of the same film radius. (b) illustrates the meaning of χ (see text for definition).

direction parallel to the rotation axis and not perpendicular to it.

(v) Filmpack

This program (due to Wilson) reduces the integrated data from packs of two films collected on the precession or oscillation camera and performs several operations including the rejection of certain reflections.

(i) systematic absences;

(ii) terms where one of the four backgrounds differs from the mean by more than 10% for precession photographs (performed in SEG 4 for oscillation photographs);

(iii) partially recorded terms on oscillation photographs if required.

Standard deviations are estimated from the agreement between equivalent and Friedel related reflections as a function of intensity within the strong and the weak film, i.e.

$$E(\sigma) = A + B\,I$$

where A and B are constants to be determined.

The scale factor between the weak and strong film is computed the form of which is:

(i) for precession photographs:

$I\,W' = k\,.\,(IW + BIW^2)$ where IW is an intensity on the weak film,

$IS' = IS + BIS^2$ where IS is an intensity on the strong film.

after Matthews et al. (1972).

(ii) for oscillation photographs:

$$IW' = IW\,.\,\exp(At \sec 2\theta)$$

At is to be determined. For sec $2\theta = 1$, exp (At) is the overall strong to weak scale factor. The dependence of this scale on sec 2θ corrects for oblique incidence absorption in the weak film.

The data are corrected for LP factors. Also the oscillation photographs are corrected for the oblique incidence effect. Bad agreements between equivalents are monitored and rejected.

The O/P analysis from this program is a useful test for the quality of the integrations performed by OSCAR on a given crystal in terms of the

number of negative terms, which is a test of the box size and optronics transformation matrix used in integration. Also, the merging R between equivalents is output where

$$R_{symm} = \sum_{nref} \sum_{i=1}^{n} (I_i - \bar{I}) / \sum_{i=1}^{n} \bar{I}$$

(notation defined in Table 3.2),

as well as the values of $\langle I/\sigma(I) \rangle$ in ranges of $\sin\theta/\lambda$. This also forms a test of the quality of the crystal.

(vi) Filmrad

This program (due to Wilson) estimates radiation damage parameters by comparing the data from different photographs. In general there is a radiation damage photograph which is a repeat of the first data photograph. In the cases of high symmetry, different data photographs for a given crystal will contain equivalent reflections collected at different times which also contain the information necessary for calculating a radiation damage correction.

Two possible corrections can be applied to the data:

(a) an empirically based correction which is quadratic in sin θ and linear in time, t:

$$I1 = I2 \cdot (1.0 + (a + bs + cs^2)\, t),$$

$s = \frac{\sin\theta}{\lambda}$; a, b, c, are constants to be determined;

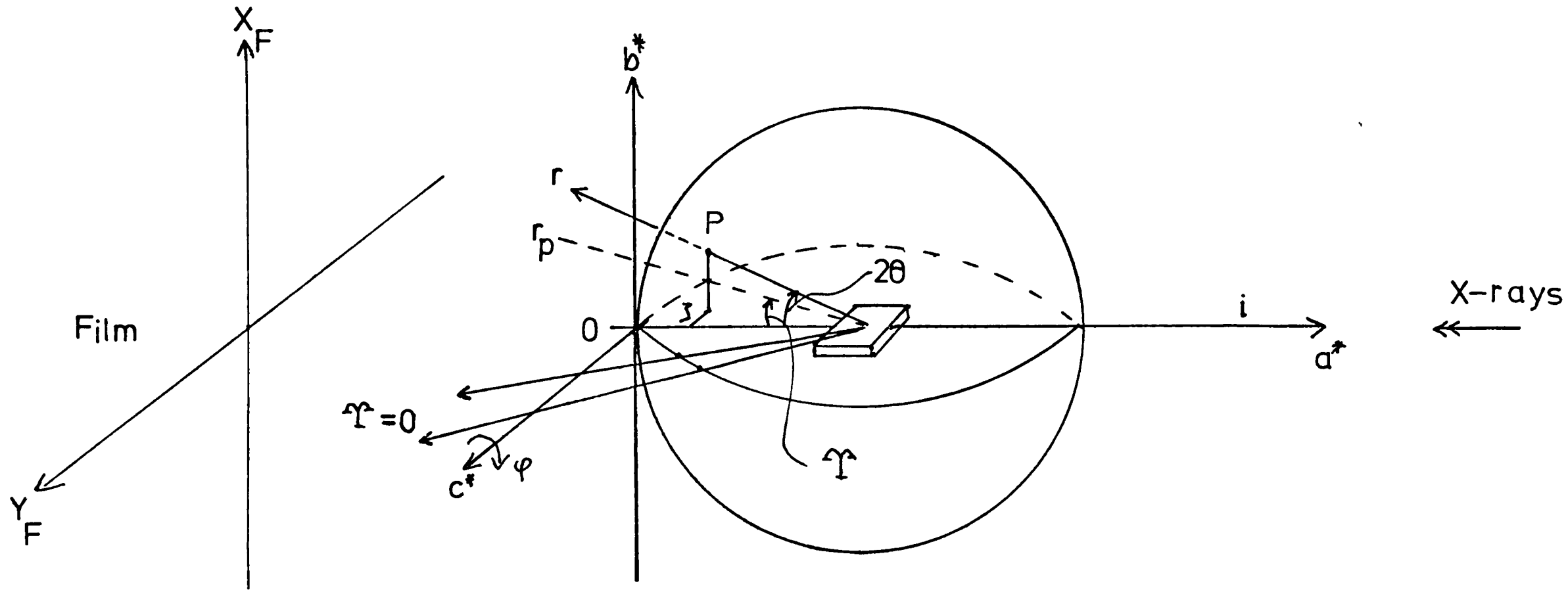
(b) an expression of the type used by Hendrickson et al. (1973) (see § 2.3),

$$I1 = I2/(\exp-(k_1+k_3)t + (\exp(-k_3 t).(1-\exp(-k_1 t))\exp(-\frac{D\sin^2\theta}{\lambda^2})))$$

For both expressions the best parameters are obtained by a least squares minimisation of:

$$\sum w(I1 - I2 \cdot CORR)^2$$

Figure 5.16 Diagram showing the relationship between 2θ, Υ and J. When $J = 0$, $\Upsilon = 2\theta$, when $\Upsilon = 0$, i and r are in the equatorial plane (the plane defined by the rotation axis and the incident beam).



where i is the incident beam, r is the reflected beam and r_p is the projection of $\dot{r}$ onto a plane normal to the equatorial plane.

where $w = \frac{1}{\sigma^2(I)}$ and is optional and CORR is the correction applied.

The correction applied to the data for any given crystal is selected on the basis of the agreement between the predicted I1/I2 and the observed I1/I2 in ranges of $\sin\theta/\lambda$. For some crystals correction (b) converges slowly,if at all, in which case (a) is useful.

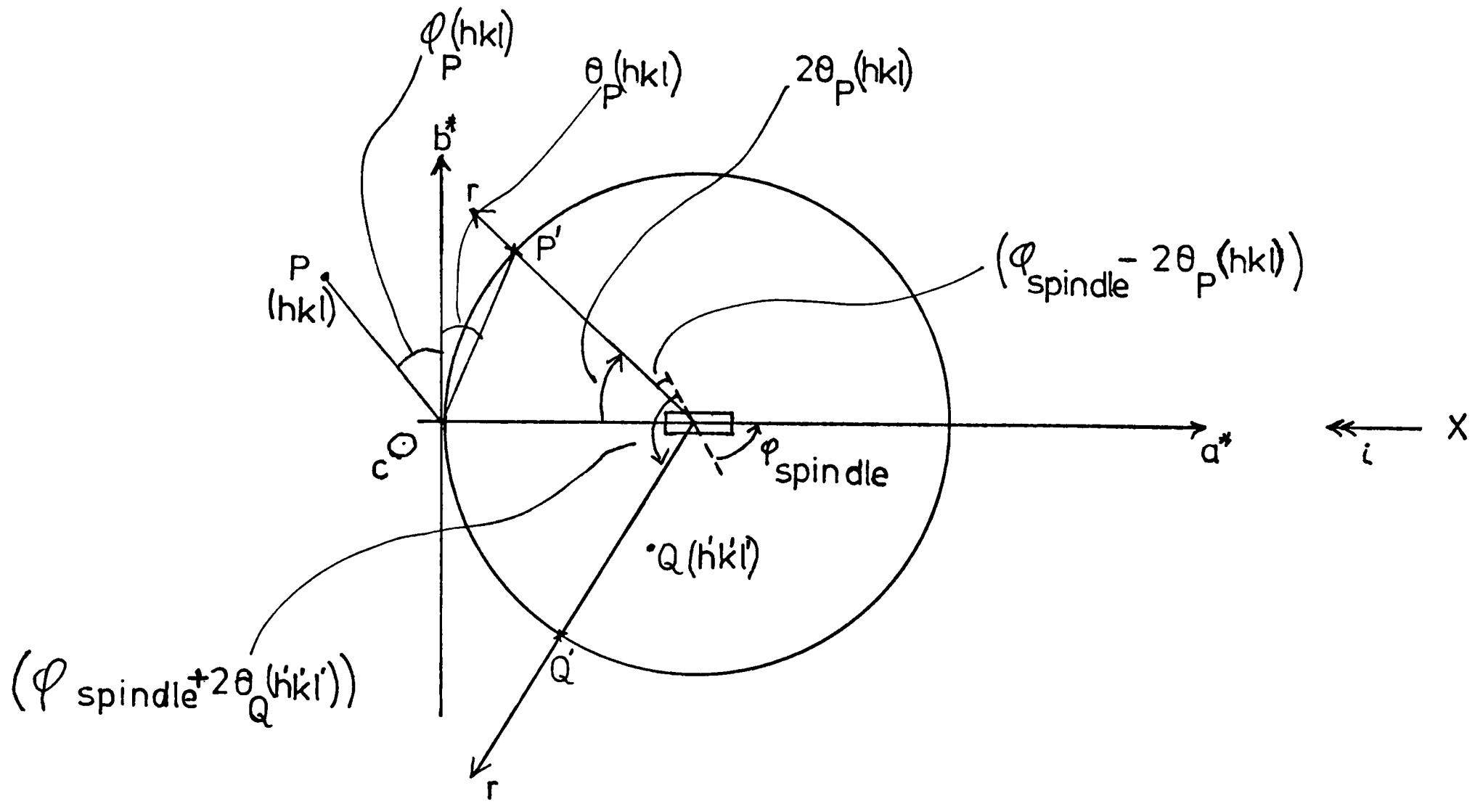
(vii) Absorption

No program exists at present for dealing with absorption. Details will be presented of the calculations that need to be made for applying an absorption correction to the data collected,described in this chapter, from 6PGDH $C222_1$ crystals.

For plate crystals the effect of absorption is significant. A typical 6PGDH crystal used on the rotation camera (dimensions 0.3 x 0.2 x 0.3 mm^3) had $I_{max}(0\ 0\ 10)/I_{min}(0\ 0\ 10) \approx 2.0$ for a C mount, when monitoring the 0 0 10 reflection on the 4 circle diffractometer to estimate a correction of the North et al. (1968) type. On any one rotation photograph, there is a variation in absorption as one moves from the rotation axis to the axis perpendicular to this (Figure 5.16). Schwager, Bartels and Huber (1973) suggest that in situations where the cross section of the crystal reflected beam is smaller than that of the crystal, the direct beam, suitably attenuated, can be monitored as a function of φ and μ. The irradiated crystal volume remains essentially constant over a small range of φ; it is thus only necessary to make relative absorption corrections within any one film.

Enraf-Nonius supply collimators down to a size giving a beam of 0.2 mm at the crystal which would be smaller than the plate dimensions of 6PGDH $C222_1$ crystals only for the largest 6PGDH crystals available. Since not enough of these were available and because it was thought that such a small collimator would be detrimental to obtaining reasonable spot intensities, a collimator yielding a 0.3 mm beam was chosen. In this situation the entire crystal is bathed in the X-ray beam and the

Figure 5.17 Situation where $\zeta = 0$, incident (i) and reflected (r) beams lie in a plane perpendicular to the rotation axis (c) (see Fig. 5.16). $\varphi_p(hkl) = \tan^{-1} \frac{ha^*}{kb^*}$. $\varphi_{spindle} = 0$ corresponds to minimum transmission. The dotted line indicates the reference line through the crystal after a rotation of $\varphi_{spindle}$.



transmission is best determined on a 4 circle diffractometer by selecting a series of suitable low angle reflections and measuring their intensities as a function of aximuthal angle. It is because this was the method adopted to deal with absorption in the 6PGDH study that more attention is being given to evaluating the absorption correction by the North et al. (1968) method than the Schwager et al. (1973) method.

According to North, Phillips and Mathews (1968) (see §2.3) the approximation is made that the absorption factor for a given hkl is the mean of the absorption corrections for beams passing through the crystal in the directions of the incident and reflected rays i.e. in terms of the transmission factor:

$$T(hkl) = \frac{\left(T(\varphi_i) + T(\varphi_r)\right)}{2}$$

where $T(hkl) = 1/A(hkl)$.

A is the absorption correction and φ_i and φ_r are the azimuthal angles of the incident and reflected beams. These values need to be calculated for each reflection and the value of the absorption at this angle can be determined from the empirical I_{max}/I_{φ} curve derived from the 4 circle diffractometer measurements of I_{φ} for a given axial reflection. Because of the uncertainty of the φ origin on removal of the crystal from the rotation camera to the diffractometer it will be necessary to assume that $\varphi_{spindle} = 90^{\circ}$ (say) corresponds to the maximum transmission for the diffractometer absorption curve.

Referring to Figure 5.17, where $\mu = 0$, the angles φ_i and φ_r for each reflection can be calculated as follows.

To bring reciprocal lattice point P to P' (lying in the plane normal to the rotation axis) onto the Ewald sphere involves a rotation of $\varphi_P(hkl) + \theta_P(hkl)$ which is identical to the spindle axis rotation φ_{spin}. The trace of the incident beam makes an angle φ_{spin} with the

line through the crystal corresponding to minimum transmission (the reference line). The trace of the reflected beam forms an angle

$$\varphi_P(hkl) + \theta_P(hkl) - 2\theta_p(hkl)$$

Therefore:

$$\varphi_{iP} = \varphi_{spindle}$$

$$\varphi_{rP} = \varphi_{spindle} - 2\theta_p(hkl)$$

For Q to Q'

$$\varphi_{iQ} = \varphi_Q(h'k'l') - \theta_Q(h'k'l')$$

$$\varphi_{rQ} = \varphi_Q(h'k'l') - \theta_Q(h'k'l') + 2\theta_Q(h'k'l')$$

i.e. $\varphi_{iQ} = \varphi_{spindle}$;

$$\varphi_{rQ} = \varphi_{spindle} + 2\theta_Q(h'k'l').$$

In the case where $\mu \neq 0$, we replace 2θ by Υ, the projection of 2θ onto the plane normal to the rotation axis (Figure 5.16), and so for a general reflection

$$\varphi_i = \varphi_{spindle}$$

$$\varphi_r = \varphi_{spindle} \mp \Upsilon$$

The sign associated with Υ in the equation is dependent on whether the reciprocal lattice point generates a reflection with $\pm X_F$ on the film.

When $\Upsilon = 0$, i and r are in the equatorial plane and $\varphi_i = \varphi_r$. It is easier to see from this situation that all reflections with the same $\varphi_{spindle}$ have an identical absorption correction i.e. the variation with 2θ of the absorption is ignored (see § 2.3).

5.3 High resolution data collection from 6PGDH crystals

Choice of mount and rotation range

In selecting these parameters use was made of the C face centring property of the crystals. This imposes the condition that reflections with (h+k) odd are absent and makes a c axis mount the best for maximising the rotation range $\Delta\varphi$. In this case (from Figure 5.7):

$$d_m^* \Delta\varphi \approx \frac{\lambda}{d} \qquad \text{and} \qquad d = |\underline{b}| / 2$$

Therefore:

$$\Delta\varphi = \frac{2.6}{148.15} \times 2$$

$$= 2.0^\circ.$$

This can be decreased by an amount equal to the reflecting range of the crystal (i.e. the sum of the cross fire in the incident beam and the mosaic spread of the crystal specimen) to prevent any overlapping of spots.

In the absence of the C face centring $\Delta\varphi$ reduces to 1.0° for a c axis mount and in which case a b axis mount would be more suitable than a c axis mount. For a b axis mount:

$$\Delta\varphi = \frac{2.6}{|\underline{c}|}$$

$$= \frac{2.6}{102.9} = 1.44^\circ.$$

It is worth tolerating a small percentage of overlapping reflections because this means an increased $\Delta\varphi$ can be used, resulting in a smaller number of films to be processed and a reduced percentage of partial reflections.

In summary, for 6PGDH 2.6 Å data collection a c axis mount, a rotation range of 2.0° and a crystal to film distance of 85 mm were selected. Generally, for native crystals, data extended to 2.6 Å and the extra labour involved in using $\Delta\varphi = 2^\circ$ seemed worthwhile. $\Delta\varphi$ increases to 2.3° for 3.0 Å data. For the Pt derivative crystals the data usually extended at best to 2.6 Å for only the first photograph of the data set and the use of a

2° rotation range seems not worthwhile in retrospect. Furthermore, a considerable labour of data processing could be saved by varying $\Delta\varphi$ as φ changed from 0° to 90° because of the marked difference in $|\underline{a}|$ and $|\underline{b}|$.

At $\varphi = 0$, ($\underline{a}^{*}$ anti-parallel to the direct beam)

$$\Delta\varphi = \frac{2.6}{|\underline{a}|/2} = 4.1^{\circ}.$$

At $\varphi = 90^{\circ}$, $\Delta\varphi = 2^{\circ}$.

The camera and data collection

The camera was mounted on an Elliott GX6 rotating anode run at 40 kV, 40 mA. The radiation was filtered using a Ni foil of thickness 0.015 cm (see Table 2.2) and the beam collimated to a size of 0.3 mm at the crystal. Camera alignment checks were made to ensure normal beam geometry, accurate perpendicularity of the cassettes with respect to the X-ray beam and maximum X-ray flux through the collimator. Several reflections were sufficiently intense to saturate one film and so two films were placed in a cassette, constituting a film pack. Spots that are too intense to be monitored on the A(strong) or B(weak) films are lost in the data processing. An "A" photograph from native crystal 8 is shown in Figure 5.18. This is one of the strongest native photographs. Several diffuse features can be seen on the film which occur as haloes around an intense spot. These haloes disappear when the Ni filter is replaced by a monochromator (Archibald, personal communication).

In general,the steps followed for data collection off one crystal were:

(a) The crystal was aligned on the camera by taking still photographs at $\varphi = 0^{\circ}$ and 90°. Accurate corrections to the crystal alignment were made by comparison of the intensities of equivalent reflections on the stills. It is an advantage to have the crystal sufficiently well set that equivalent reflections across the mounting axis are recorded on the same film, especially to obtain an accurate Bijvoet difference $(I^{+}-I^{-})$. This approach minimizes errors in the anomalous difference due to variations in film response and

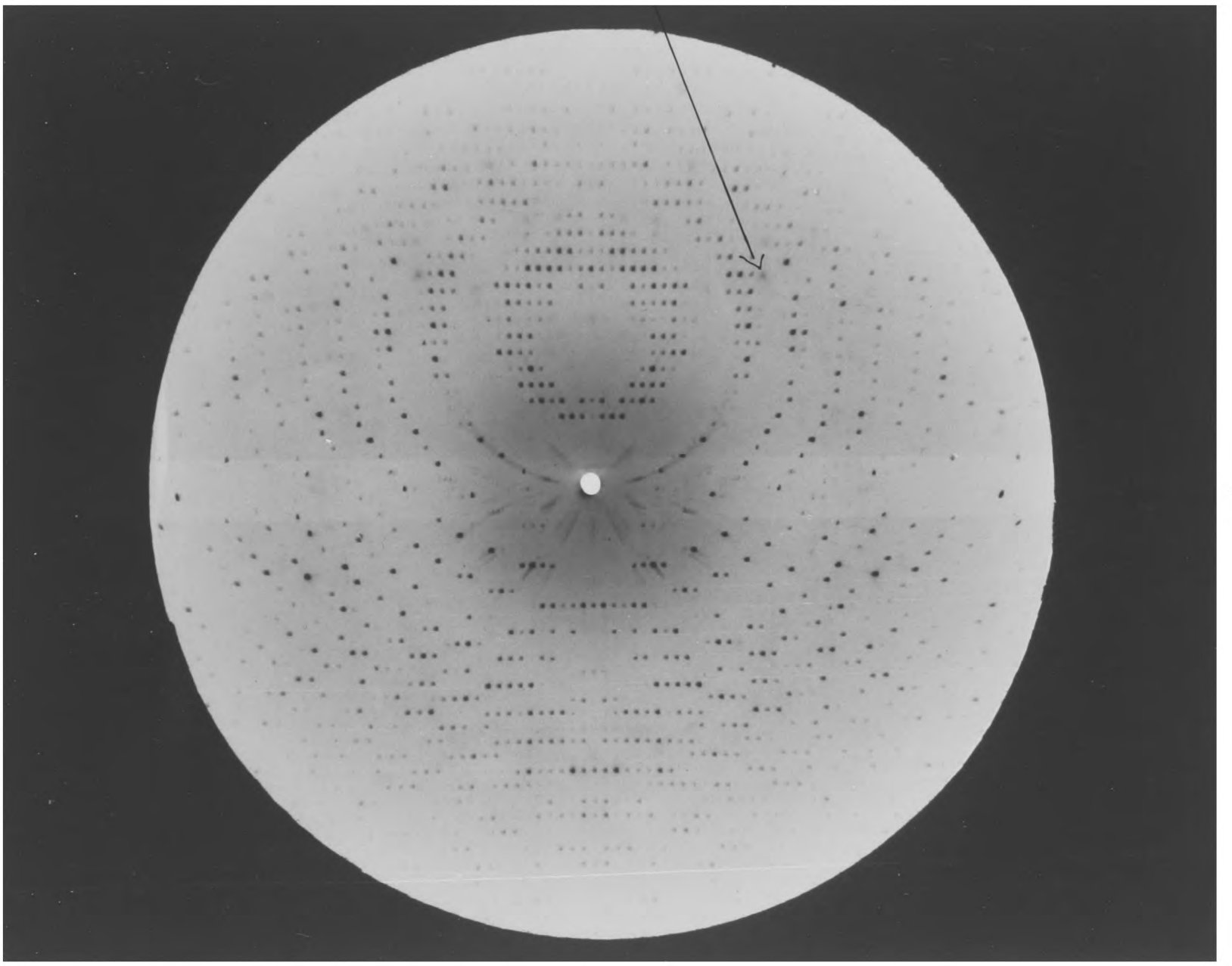


Figure 5.18 A native oscillation data photograph; crystal 8, $\varphi = 50\text{-}52^{\circ}$, $\underline{c}$ axis mount, $\underline{a}^{*}$ antiparallel to the X-ray beam at $\varphi = 0^{\circ}$.

thickness.

(b) Orientation photographs were taken at $\varphi = 0^{\circ}$ and 90°.

(c) The exposure rate (seconds/$^{\circ}$) was set according to the spot intensities on these photos and automatic data collection commenced. The first photograph overlapped by 2° with the last photograph of the previous crystal for crystal to crystal scaling purposes. The final photograph was a repeat of the first and was used in the radiation damage analysis. As outlined previously the setting photographs were also repeated.

Room temperature was held constant at about 18°C. Above 25°C the crystals decayed rapidly. The use of a stream of cold air produced by the device of Marsh and Petsko (1973) produced no improvement in crystal lifetime. In fact, a few crystals dried out presumably because the capillary was not air tight.

Films were developed according to the following standard procedure for Ilford Industrial G film:

Developer: Ilford Phenisol; 5 minutes at 20° C.

Stop Bath: 2% acetic acid; 30 seconds.

Fixer: May and Baker Super Amfix; 10 minutes.

Wash: Bath of running tap water; 30 minutes.

Drying: Oven at 37°C; 45 minutes.

Derivative data collection was very disheartening since the platinum soaked crystals were quite fragile. Initially crystals were cut to a suitable size before soaking but were found to be hypersensitive on mounting. It was found however that by leaving the needle crystal uncut, after soaking,the crystals retained more of their natural strength. The use of long needles enabled up to a maximum of three data sets to be collected off one crystal simply by advancing the crystal down the rotation axis. However, this did prevent the collection of an absorption curve on the diffractometer because these crystals ($\approx$ 1 mm in length) were bigger than the X-ray beam. This invalidates the assumptions of the North et al. (1968) method of absorption

Table 5.2 Details of data collection

(a) native crystals

Crystal	Crystal size* (mm^3)	sec/°	mount	Range of φ	I_{max}/I_{min}	Data collected by
1	0.36x0.16x0.36	7500	c	0-18°	2.52	IGA
2	0.4x0.16x0.32	7500	c	16-30°	2.18	IGA
3	0.4x0.2x0.3	7500	c	28-44°	1.99	JRH
26	0.33x0.15x0.5	7500	c	46-52°	**	JRH
8	0.4x0.25x0.4	7500	c	50-62°	2.17	JRH/IGA
9	**	7500	c	60-72°	2.13	JRH
10	**	7500	c	70-82°	2.09	IGA
11	0.46x0.2x0.3	7500	c	80-92°	2.54	IGA/JRH
13	0.45x0.25x0.35	7500	a	88-100°	**	JRH
14	0.5x0.2x0.27	7500	a	98-108°	**	JRH

* estimated using a 2 circle goniometer and microscope (x40) or a stereo microscope (x 50).

** no record kept.

1. I_{max} and I_{min} are defined in Table 3.2 and refer to absorption measurements on the 0 0 10 reflection on the 4 circle diffractometer.

2. For a c axis mount for $\varphi = 0°$ a^* is antiparallel to the direct beam. For an a axis mount, for $\varphi = 90°$ b^* is anti parallel to the direct beam.

(b) derivative crystals

Crystal	Crystal size (mm^3)	sec/^{o}	mount	Range	soaking time (hrs)	Data collected by
PT2	0.32x0.15x0.32	7500	c	0-10^o	17	JRH
3	0.35x0.15x0.4	7500	c	8-22^o	18	JRH
8	0.4x0.15x0.33	7500	c	20-30^o	17	JRH
10	0.43x0.18x0.38	7500	c	28-36^o	17	JRH
30C	0.4x0.15x1.33	5000	c	34-42^o	22	JRH
29B	**	5000	c	40-48^o	20	JRH
31	0.3x0.15x0.3	5000	c	46-54^o	20	JRH
30B	0.4x0.15x1.33	5000	c	52-60^o	22	JRH
24B	0.28x0.13x1.1	7500	c	58-66^o	17	JRH
25C	0.35x0.16x0.55	5000	c	64-74^o	18	JRH
30A	0.4x0.15x1.33	5000	c	72-82^o	22	JRH
26B	0.4x0.2x0.45	5000	c	80-90^o	17	JRH

Average 18.9 $\pm$ 2.2 hrs

correction since the volume of diffracting material varies with (φ, μ).

Table 5.2 shows information relevant to the data collection off each native and platinum crystal. Attention is drawn to the fact that for Table 5.2b, the crystals labelled PT30A, PT30B, PT30C are in fact the same needle, but irradiated in different places. PT29B, PT25C and PT26B were cut fragments from a given needle where PT29A, PT25A, PT25B, and PT26A were not useable in data collection. PT24B was a portion of a needle. Data collected off the other end of the needle, PT24A was not used because of crystal slippage.

The data collection on the native crystals was started in March 1976 and the Pt data collection completed in April 1977. This is a slow rate of progress and occurred because:

(a) only one camera exists in the laboratory and was run as a shared facility between 3 main users. On average 2 out of 8 weeks were allocated to each group.

(b) The type A $\longrightarrow$ type B transformation delayed progress especially for the derivative data collection. Because of a shortage of crystals only two crystals would be set to soak for any given 19 hour soak. If both transformed or died suddenly a full day was lost. It was not feasible to mount several crystals at once because the crystals diffracted reliably only if mounted fresh from the soak solution.

(c) Until the use of full needle crystals in the derivative data collection, less than 50% success rate was obtained in terms of useable data from a given crystal.

(d) The native and derivative crystals were radiation sensitive, particularly the latter. This accounts for the larger number of derivative crystals that needed to be used for a "complete" data set of four equivalent reflections.

5.4 Processing the 6PGDH data films

Refinement of crystal misorientation parameters

This was performed using SEG1 of AUTOOSCAR as described earlier (§ 5.2). The input to the program includes the film coordinates and degree of partial recording for a set of partials taken from setting photographs 90° apart and sometimes including partials from data photographs. The misorientation matrix output had to predict spots correctly in position and flag them appropriately as a partial or a full.

The refinement is sensitive to incorrect estimates of the degree of partial recording and so the following procedure was adopted.

(i) A single cycle refinement was run varying the three misorientation angles and crystal to film distance. With reference to the agreement between D_o and D_c obviously bad data was removed. The observed film coordinates were compared with those calculated from the present set of parameters and disagreements checked by remeasurement of these coordinates. Reflections with truly observed differences between the observed and calculated coordinates were not removed from the refinement.

(ii) A refinement was run for sufficient cycles to produce convergence. Reciprocal cell parameters were not varied. The checks described in (i) above were made. Only when the input data passed these checks was a full parameter refinement attempted. The use of only four variables in the early stages of the procedure produced quicker and more reliable convergence.

(iii) The misorientation matrix generated was used in the calculation of film coordinates for a given rotation range. The plot output from the 1906A graph plotter was compared with the appropriate photograph. Spots that were not predicted were input into the refinement and the procedure (i) to (iii) reiterated.

Three types of orientation photograph were investigated for estimating the degree of partial recording.

(a) The photographs $-1 \rightarrow 0^\circ$ and $89 \rightarrow 90^\circ$. The degree of recording was estimated as the geometrical fraction of the spot recorded.

Table 5.3 Details of the refinement of cell parameters, misorientation angles and crystal to film distance

(a) native crystals

Crystal	$\|\underline{a}\|$	$\|\underline{b}\|$	$\|\underline{c}\|$	x	y	z	D	γ	r.m.s. residual	number of reflections, nref.	orientation photos used to estimate partials
1	72.97	148.73	102.72	0.20	-0.30	0.35	84.95	0.35	0.59×10^{-3}	40	$0 \rightarrow 2^{\circ}$ $89 \rightarrow 90^{\circ}$
2	72.67	148.28	102.27	0.05	-0.05	0.05	84.62	0.35	0.46×10^{-3}	33	$-1 \rightarrow 0^{\circ}$ "
3	72.49	148.22	102.67	0.01	0.06	-0.06	84.41	0.4	0.69×10^{-3}	49	" "
26	72.83	148.79	102.90	-0.02	-0.05	0.06	84.76	0.4	0.72×10^{-3}	72	" "
8	72.80	148.60	102.70	0.01	0.05	-0.01	84.84	0.4	0.71×10^{-3}	80	" "
9	72.57	148.23	102.86	-0.15	0.02	0.03	84.48	0.4	0.60×10^{-3}	97	" "
10	72.73	148.36	102.72	0.01	0.09	-0.03	84.80	0.4	0.63×10^{-3}	85	" "
11	72.89	148.43	102.79	0.0	0.05	-0.02	84.99	0.4	0.68×10^{-3}	90	" "
average=	72.74 $\pm$ 0.16	148.46 $\pm$0.23	102.70 $\pm$ 0.19						average=$(0.64 \pm 0.1) \times 10^{-3}$		

Note

r.m.s. residual $\sqrt{\frac{(D_o - D_c)^2}{nref}}$; D = crystal to film distance ; γ = crystal reflecting range

(b) derivative crystals

Crystal	|a|	|b|	|c|	x	y	z	D	γ	r.m.s. residual	number of reflections nref.	orientation photos used to estimate partials
PT2	72.67	148.94	102.88	-0.11	-0.03	0.14	84.26	0.4	0.64×10^{-3}	38	$0 \to 2^{\circ}$, $89 \to 90^{\circ}$
3	73.01	149.09	103.34	-0.08	-0.46	-0.22	84.00	0.4	0.56×10^{-3}	53	$0 \to 1$, $89 \to 90^{\circ}$
8	73.07	148.98	102.88	-0.23	0.30	0.57	84.54	0.4	0.71×10^{-3}	68	$0 \to 1$, $89 \to 90$
10	73.12	149.60	103.19	-0.02	0.14	0.11	84.61	0.4	0.66×10^{-3}	61	89.0→89.5; 89.5→90.0 0.0→0.5; 0.5→1.0
30C	72.85	148.99	103.24	0.25	0.09	-0.22	84.90	0.4	0.68×10^{-3}	59	89.5→90.0; 90.0→90.5 -1→0.5; -0.5→0.0
29B	72.58	148.92	103.08	0.18	-0.08	-0.18	84.85	0.5	0.67×10^{-3}	76	"
31	72.46	148.53	102.87	-0.05	-0.30	0.12	84.62	0.4	0.61×10^{-3}	61	"
30B	72.71	148.68	103.05	0.27	0.11	-0.09	84.83	0.4	0.69×10^{-3}	76	"
24B	72.52	148.80	103.02	0.29	0.40	-0.24	84.53	0.5	0.86×10^{-3}	87	-1→0, 89→90
25C	72.46	148.71	102.87	-0.14	-0.17	-0.14	84.74	0.5	0.46×10^{-3}	100	stills 0°, 90°
30A	72.74	148.41	102.98	0.23	0.12	0.11	84.84	0.5	0.77×10^{-3}	92	-1→0.5, -0.5→0.0 80→82 -1→0, 88→90

26B did not refine satisfactorily

average $\overline{72.75}$ $\overline{148.88}$ $\overline{103.04}$
$\pm$ 0.24 $\pm$0.32 $\pm$0.24

average $\overline{= (0.67 \pm 0.1) \times 10^{-3}}$

Notice the small value of the r.m.s. residual for PT25C for which stills were used.

(b) The photographs $-1 \rightarrow -0.5^{\circ}$, $-0.5^{\circ} \rightarrow 0^{\circ}$ and $89.0 \Rightarrow 89.5^{\circ}$, $89.5 \rightarrow 90^{\circ}$. This involved no more crystal exposure than (a). The intensities of partials recorded at -0.5° and at 89.5° could be compared directly on the appropriate adjacent films.

(c) Stills at 0° and 90°. This reduced the exposure time devoted to obtaining a misorientation matrix. Here, each spot, which is by definition a partial, is arbitrarily assigned a degree of recording of 0.5 and the error associated with taking the centre of the partial as the centre of the spot was ignored.

In each method the spots selected were evenly distributed around the film.

For 6PGDH crystals where the spot shape is not spherical and is at best rectangular, method (a) was very inefficient and the cycle (i) to (iii) overleaf had to be iterated many times. Method (b) was better than (a) but not as convenient as (c). Method (c) removes the subjective eye estimate of the degree of partial recording. Using (c) a correct matrix was obtained for crystal PT25C on the second iteration of the cycle (i) to (iii) above. Also, the r.m.s. residual, $\langle (D_o - D_c)^2 \rangle^{\frac{1}{2}}$, for this refinement is one of the best of all the refinements performed (Table 5.3). Stills have been used successfully in the data processing of the Au derivative of 6PGDH (Archibald, personal communication). The r.m.s. residuals for methods (a) and (b) are on average equal.

The integrated data

The data films were integrated according to the scheme outlined in § 5.2.

In assessing the quality of each crystal data set several factors need to be considered: the agreement between the equivalent reflections and the percentage of weak reflections (both of which get worse as the crystal is increasingly irradiated) and the amount of radiation damage suffered by the crystal as observed in the diffraction pattern.

Table 5.4 gives details of the average merging R factors between

Table 5.4 Details of the agreement factors between equivalents based on information output by FILMPACK[c].

The merging R factors are averages over all the films from a given crystal.

(a) native

Crystal	Rotation range	$\langle R_A \rangle$	$\langle R_B \rangle$	$\langle R_{AB} \rangle$	$\langle I_A \rangle$	$\langle I_B \rangle$	% of reflns. with $I<3\sigma$ for the first A film of crystal data set	Processed by[a]
Native 1	0-18	9.8±1.5	22.2±4.7	18.2±3.0	440±61	168±31	31	JRH
2	16-30	11.5±1.6	23.7±3.8	21.9±3.0	366±49	135±22	46	DV
3	28-44	10.2±1.9	23.2±2.4	21.4±1.0	421±63	147±21	40	JRH
26	46-52	8.9±1.3	12.1±1.8	13.4±1.8	745±84	348±59	26	JRH
8	50-62	7.3±0.7	10.8±3.3	12.8±0.9	1005±132	494±64	23	JRH
9	60-72	10.3±1.3	17.8±2.0	17.4±2.0	693±99	291±48	27	JRH
10	70-82	8.5±0.8	13.9±1.4	14.1±1.1	781±62	350±30	22	JRH
11	80-92	7.4±0.9	12.8±2.8	13.2±1.6	920±90	398±47	19	JRH
h0l precession[b]		17.2	25.5	18.5	252	120	61	JRH

[a] See "Acknowledgments" for explanation of initials.

[b] Two A films were used from separate packs - no reflections were lost due to saturation on the strong film.

[c] Note LP factors are applied by this program.

(b) Pt derivative

Crystal	Rotation range	$\langle R_A \rangle$	$\langle R_B \rangle$	$\langle R_{AB} \rangle$	$\langle I_A \rangle$	$\langle I_B \rangle$	% of reflns. with $I<3\sigma$ for the first A film of crystal data set	Processed by[a]
PT2	0-10	14.3±4.8	26.2±7.3	25.8±6.5	352±98	137±41	53	MJA
PT3	8-22	11.5±2.4	18.1±3.3	18.3±3.2	366±68	151±29	51	MJA
PT8	20-30	9.8±1.1	11.6±1.6	14.5±1.9	498±65	258±34	42	MJA
PT10	28-36	11.4±1.8	15.5±2.9	17.0±1.8	441±56	201±40	53	MJA
PT30C	34-42	10.6±1.6	16.9±4.0	15.2±2.6	529±73	220±39	55	JRH
PT29B	40-48	11.6±1.7	19.0±3.1	19.1±1.9	517±62	218±33	43	JRH
PT31	46-54	9.3±1.3	16.3±2.5	15.5±1.3	292±27	116±13	47	JRH
PT30B	52-60	7.4±0.5	12.8±1.6	11.5±0.5	915±80	392±59	40	JRH
PT24B	58-66	11.7±2.9	19.7±5.2	16.4±3.3	653±166	311±93	47	JRH
PT25C	64-74	7.9±1.7	12.8±1.3	12.2±1.8	936±131	408±52	41	JRH
PT30A	72-82	7.9±1.3	12.6±2.7	11.6±1.6	955±112	416±48	37	JRH
h0l[b] precession		17.4	19.9	15.8	214	148	65	JRH

equivalent fully recorded reflections on both the strong (A) and weak (B) films (after LP factors have been applied) and between the strong and scaled up weak films (AB) where the average is taken overall the films from the given crystal. The standard deviations of these $\langle R \rangle$'s indicate the consistency of the quality of processing of each film and the degree to which R increases with exposure time. The percentage of reflections with $I < 3 \times \sigma(I)$ is given for the first film off each crystal. On comparing Table 5.4 a and b, this percentage is considerably greater for the platinum crystals than the native and indicates that in the high angle annuli of the film there are few significantly measured reflections for the derivative data. The level of significance of the measurements at high angles is considered in §5.5. It must also be remembered that the large number of weak reflections at the outer edge of the film (2.6 Å limit) tend to increase the merging R factors giving an unfair indication of the data quality. Table 5.5 gives details of the radiation damage corrections applied to the crystals.

To give an indication of the variation of merging R and $\langle I1/I2 \rangle$ with $\frac{\sin\theta}{\lambda}$ where I1 is the intensity of a fully recorded reflection in a given $\sin\theta/\lambda$ shell on the first data photograph and I2 for the same reflection on the radiation damage photograph, Figure 5.19 shows these quantities for 3 native crystals. These crystals are the best, average and worst of those from which native data have been collected, assessed in terms of $\langle R_{AB} \rangle$ and $\%I < 3\sigma(I)$ and are therefore representative of the native crystals studied.

It should be pointed out that there are certain gaps in the data sets. The blind region data for the native was collected on two crystals by rotating about a as axis, but not processed. The platinum blind region data was not collected. One crystal each for the native and platinum did not behave well in refining misorientation parameters and accounts for the gap 44°-46° for the native and 82°-90° for the platinum data sets. These gaps are offset by the inclusion of h0l precession photographs ($\bar{\mu}=15^{\circ}$) and of the low resolution diffractometer data (suitably scaled). The latter data were

Table 5.5 Radiation damage correction applied

(a) native crystals

Crystal	Type of correction applied	a k_1	b k_3	c D
1	LQ[1]	0.041	-0.731	7.30
2	LQ	-0.045	0.295	2.26
3	BP[2]	3.719	0.029	11.59
26	LQ	0.01	0.446	8.09
8	BP	0.137	0.033	86.05
9	BP	0.637	0.040	19.33
10	BP	0.148	0.015	38.89
11	BP	0.093	0.007	54.48

1 See section § 5.2 (vi); method (a) = LQ

2 See section § 5.2 (vi); method (b) = BP

The remaining symbols are also explained in § 5.2 (vi).

Table 5.5

(b) derivative crystals

Crystal	Type of correction	a k_1	b k_3	c D
2	BP	-0.203	0.012	236.81
3	LQ	-0.008	0.413	6.59
8	LQ	0.195	-3.31	23.18
10	LQ	0.309	-5.63	36.39
30C	BP	0.120	0.030	69.24
29B	BP	0.687	0.027	20.20
31	BP	0.172	0.000	44.44
30B	BP	0.178	0.018	60.37
24B	BP	0.243	0.148	61.55
25C	LQ	-0.046	0.783	4.00
30A	LQ	0.010	0.10	6.31

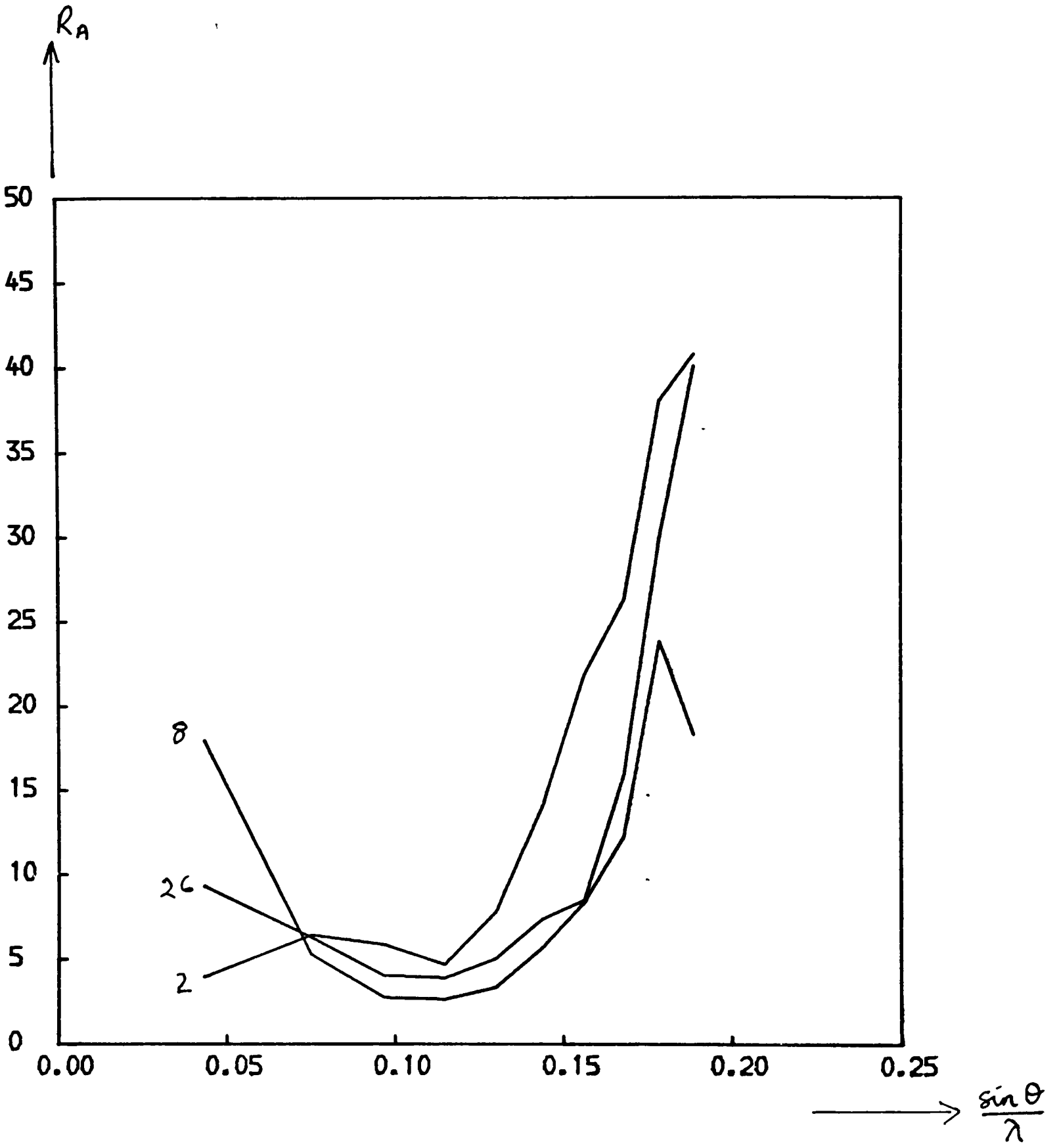


Figure 5.19a The variation of the merging R factor between equivalents in the A film of a film pack, R_A, with $\sin\theta/\lambda$ for the native crystals 8, 26, 2.

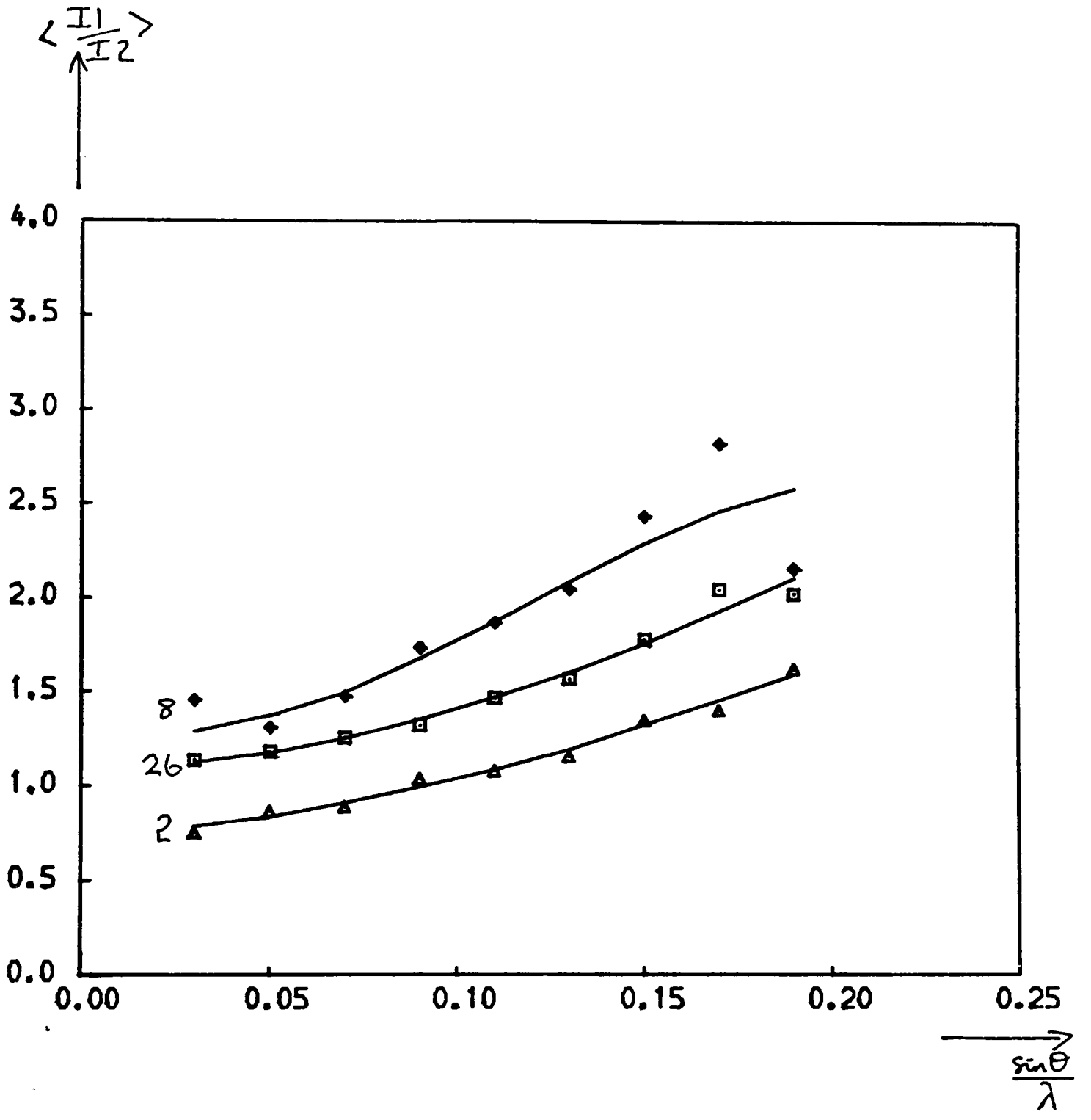


Figure 5.19b Variation of $\langle I1/I2 \rangle$ with $\sin\theta/\lambda$ for native crystals 8 (◆), 26 (□) and 2 (△). I1 is the intensity of a fully recorded reflection in a given $\sin\theta/\lambda$ shell on the first data photograph of a crystal data set and I2 is the same quantity for the last data photograph, which is a repeat of the first. This plot gives an indication of the radiation damage suffered by three crystals which are representative of all the native data.

added prior to the calculation of the electron density map (§ 5.6).

Merging the data

The photographic data were merged using the program SSM. Friedel equivalents hkl, hk$\bar{l}$ and $\bar{h}$kl, $\bar{h}$k$\bar{l}$ were not merged for the platinum derivative data. Each photograph data file was treated separately so that file to file scaling would compensate for errors in film developing and radiation damage correction between films from the same crystal. An initial file to file scale was established by comparing the mean intensity from fully recorded reflections for each film. These scales were then refined using the method of Fox and Holmes (1966) in preference to that of Hamilton, Rollett and Sparks (1965) because it offers more rapid convergence when dealing with very large numbers of reflections. Partial reflections from contiguous films were then added and the equivalents merged. Particular attention was paid to making sure that the distribution of $\sigma(I)$ with I was correct. This was assessed in terms of the distribution of $\langle (I-\bar{I})/\sigma(I) \rangle$ in ranges of $\bar{I}$ which should have a mean of 0.0 and standard deviation of 1.0. If necessary, the $\sigma(I)$'s were modified to SN6(σ^2 +(SN5*I)^2); use of a small positive SN5 reduced the standard deviation of this distribution. The value of the $\sigma(I)$'s was important in the treatment of weak reflections. Table 5.6 gives details of the data merging for the native and platinum data sets, including merging R, SN5, SN6, mean and standard deviations of the distribution $\left\langle \frac{(I-\bar{I})}{\sigma(I)} \right\rangle$. The merging R to 6 Å is included for the native data and this value (6.18%) compares well with the diffractometer data merging R (6.54%). (Compare Tables 3.2 and 4.5, 5.6.).

An analysis is output of the $\langle I \rangle$ based on partially and fully recorded reflections of an equivalent set in ranges of I (Table 5.7). For both data sets the intensities based on summed partials are greater than those from fulls, on average. The relative overestimation is approximately constant with I and is bigger for the native than for the derivative data. A possible source of this relative overestimation or bias is due to background scatter. The white radiation that remains after Ni filtering contributes generally

Table 5.6 Details relevant to the merging of the native and platinum photographic data

Type	Soaking conditions	R_{symm}(%)	Unique no. of reflns.	SN5	SN6	μ_e	σ_e
native	-	16.56	16203	0.04	1.02	0.01	0.97
platinum	2.7 mM, 19± 2 hrs	16.18	30223	0.04	1.1	0.01	1.05

Notes

1. $\sigma(I)$ is reset after merging to SN6. $(\sigma^2(I) + (SN5.I)^2)^{\frac{1}{2}}$; μ_e and σ_e are the mean and standard deviation respectively of the error distribution $(I - \bar{I}) / \sigma(I)$, which should be equal to 0.0 and 1.0, respectively, if the $\sigma(I)$ have been estimated correctly.

2. R_{symm} is defined in Table 3.2. The values quoted here are for data to the resolution limit of the film (2.6 Å). R_{symm} was also calculated for the native film data to 6 Å and was = 6.18% (over 1416 unique terms).

Table 5.7 The overestimation of intensities for partially compared with fully recorded reflections

(a) native to 2.6 Å

nref	$\langle I \rangle^a$	$\frac{\langle I_P - I_F \rangle}{\langle I \rangle} \times 100$
1765	67	-34
992	294	14
576	488	14
384	698	9
262	893	9
212	1101	9
154	1290	8
99	1497	12
93	1689	7
61	1911	7
52	2084	9
36	2301	9
30	2496	14
26	2713	9
20	2907	11
8	3112	11
10	3325	6
9	3465	16
12	3691	6
41	4844	5
	average[b]	9.7± 3.1

(b) native to 6 Å

nref	$\langle I \rangle^a$	$\frac{\langle I_P - I_F \rangle}{\langle I \rangle} \times 100$
98	88	9
55	290	4
46	478	5
32	677	2
27	911	6
24	1119	5
21	1277	2
15	1509	2
14	1701	6
7	1910	6
12	2097	8
6	2304	4
7	2448	9
2	2669	10
5	3090	7
2	3521	-9
10	4417	5
	average[c]	4.8± 4.3

(c) Pt to 2.6 Å

nref	$\langle I \rangle^a$	$\frac{\langle I_P - I_F \rangle}{\langle I \rangle} \times 100$
3333	53	-59
1436	290	10
836	493	8
489	693	8
345	892	7
254	1089	7
208	1303	6
123	1496	5
127	1701	5
95	1893	5
77	2085	6
42	2315	8
34	2485	4
28	2694	5
43	2889	6
22	3110	11
17	3299	7
8	3504	1
10	3686	3
61	4876	6
	average[b]	6.2± 2.3

[a] The average intensity is calculated over all reflections (fulls and partials) in the given intensity range.

[b] The average overestimation leaving out the weakest range of intensities.

[c] Over all intensity ranges.

to the background noise on the film and may also be expected to produce a local background (or halo) about a spot, proportional to the normal spot intensity. The average background around a partially recorded reflection will be underestimated compared to an equivalent fully recorded reflection. This implies that an intensity derived from combining two fragments will be greater than the corresponding intensity recorded in one pass (Irwin (1976)).

Another effect which contributes to the overestimation of partials is the angular backlash of the camera stepping motor which leads to the volume of each spot, when recorded as a partial, being oversampled. The Arndt-Wonacott camera has been built so that this is minimized. The backlash is approximately 0.002° which should be compared with the angular spread of reflection of about 0.4°.

The relative discrepancy between partially and fully recorded intensities can be explained alternatively by the underestimation of fully recorded reflections. This can occur either because of errors in the misorientation matrix or underestimation of the reflecting range, γ, of the crystal.

For a poorly determined matrix or too small a, γ, partials will be flagged as fulls with the result that these fully recorded reflections will have too small an intensity. In the case of a poor matrix, any fulls flagged as partials will have the correct intensity but a larger error associated with them since an unnecessarily large area of the film has been sampled for that reflection. Both the effects of the errors in the misorientation matrix and the underestimation of γ are more severe at higher angles. It is interesting that the average overestimation of partials with respect to fulls is about 5 and 10% for the native data to 6 Å and 2.6 Å respectively i.e. much larger at higher angles.

Table 5.7 also shows that the overestimation for the platinum data is less marked. It may be thought initially that this was due to more accurately determined matrices. However, figures forthe average r.m.s. residuals between D_o and D_c for the platinum and native crystal refinements (Table 5.3) are comparable and therefore do not support this conclusion. An alternative

explanation lies with the fact that the platinum data does not extend to the edges of the film, unlike the native. Hence, errors in the matrix could not be expected to affect the partial estimates of the intensity, at higher angles, as much as in the native data. Another possible explanation lies with the estimates of γ ($\approx 0.3°$ for the native crystals and $\geq 0.4°$ for the derivative). The higher value was necessary for the derivative because the soaking of the crystals in the heavy atom solution tended to disorder them . It could be argued, however, that γ had been underestimated for the native crystals. This could also explain the different results obtained for the native and platinum data.

A very curious feature, in Table 5.7, is the observation that for the 2.6 Å data for the lowest <I> range the partials are underestimated significantly. No explanation for this effect can be given.

Treatment of weak reflections

This has been according to the method of French and K.S. Wilson (1977) using the program TRUNCATE (Appendix 4).

The true intensities of weak reflections are, of course, non-negative, but because of background fluctuations this may be the recorded value. Leaving out negative intensity measurements can cause sampling bias. There is more knowledge of the intensity than just that it should be positive. Namely that any moderate or large data set obeys Wilson's statistics (A.J.C.Wilson (1949)), i.e.:

for acentric reflections: $$_{(1)}P(I) = \Sigma^{-1}\exp\left(\frac{-I}{\Sigma}\right) \quad \text{if } I \geq 0$$
$$= 0 \quad \text{if } I < 0$$

for centric reflections $$_{(\bar{1})}P(I) = (2\pi \Sigma I)^{-1/2}\exp\left(\frac{-I}{2\Sigma}\right) \quad \text{if } I \geq 0$$
$$= 0 \quad \text{if } I < 0$$

where Σ = mean intensity in the appropriate shell of reciprocal space.

Wilson (1949) also suggested that $\Sigma(\theta) = \sum_{j=1}^{N} f_j^2(\theta)$ which is however not true for protein crystal diffraction data (see later in this chapter).

If we assume that the observation I for a reflection with true intensity $\bar{I}$ satisfies the normal distribution i.e.:

$$I \approx N(\bar{I}, \sigma^2)$$

Then we have the posterior distribution for e.g.:

$$P(I|\bar{I}) \propto \exp(-\tfrac{1}{2}(I-\bar{I})^2/\sigma^2).\exp(-I/\Sigma) \quad \text{if } I \geqslant 0$$

$$= 0 \quad \text{if } I < 0$$

(see Figure 5.20).

This is a truncated positive distribution. A similar result follows for the centric case.

The program works as follows:

1. The maximum value of $4\sin^2\theta/\lambda^2$ (S) is found and the data is then split up into equal ranges of S with a minimum of 20 terms in each range.

2. The mean intensity is computed as a function of S and an interpolation table set up for the mean intensity.

3. The observed intensity (possibly negative) is read in along with its standard deviation. The truncated values are calculated according to the truncated Wilson distribution.

4. Wilson's distribution is output for the centric and acentric data in ranges of S to ensure that the assumption of the distribution as a prior is valid. This is done in terms of the cumulative distribution i.e.:
If we express each intensity in terms of its local average $Z = I/\Sigma_I$
we have:

$$(i)P(Z)dZ = e^{-Z}dZ$$

and $$(\bar{I})P(Z)dZ = (2\pi Z)^{-\frac{1}{2}}\exp(-Z/2)$$

Early attempts to use the P(Z) frequency plots for distinguishing the two distributions showed them to be unduly susceptible to small sample errors. Instead, the cumulative distributions

$$N(Z) = \int_0^Z P(Z)dZ$$ were preferred (Rogers (1965)).

These are

$$(1)N(Z) = 1 - \exp(-Z)$$

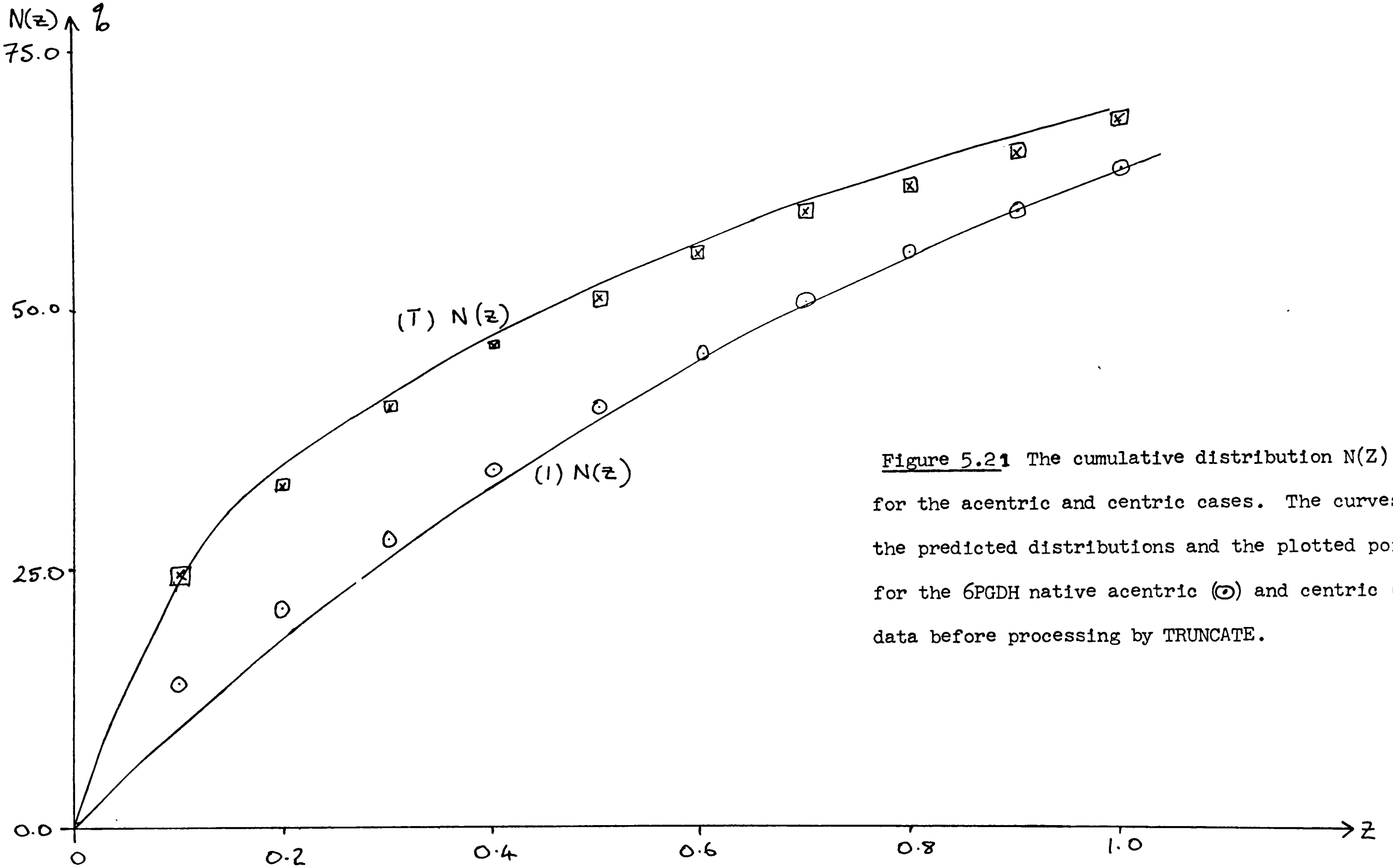


Figure 5.21 The cumulative distribution N(Z) for the acentric and centric cases. The curves are the predicted distributions and the plotted points are for the 6PGDH native acentric (⊙) and centric (⊠) data before processing by TRUNCATE.

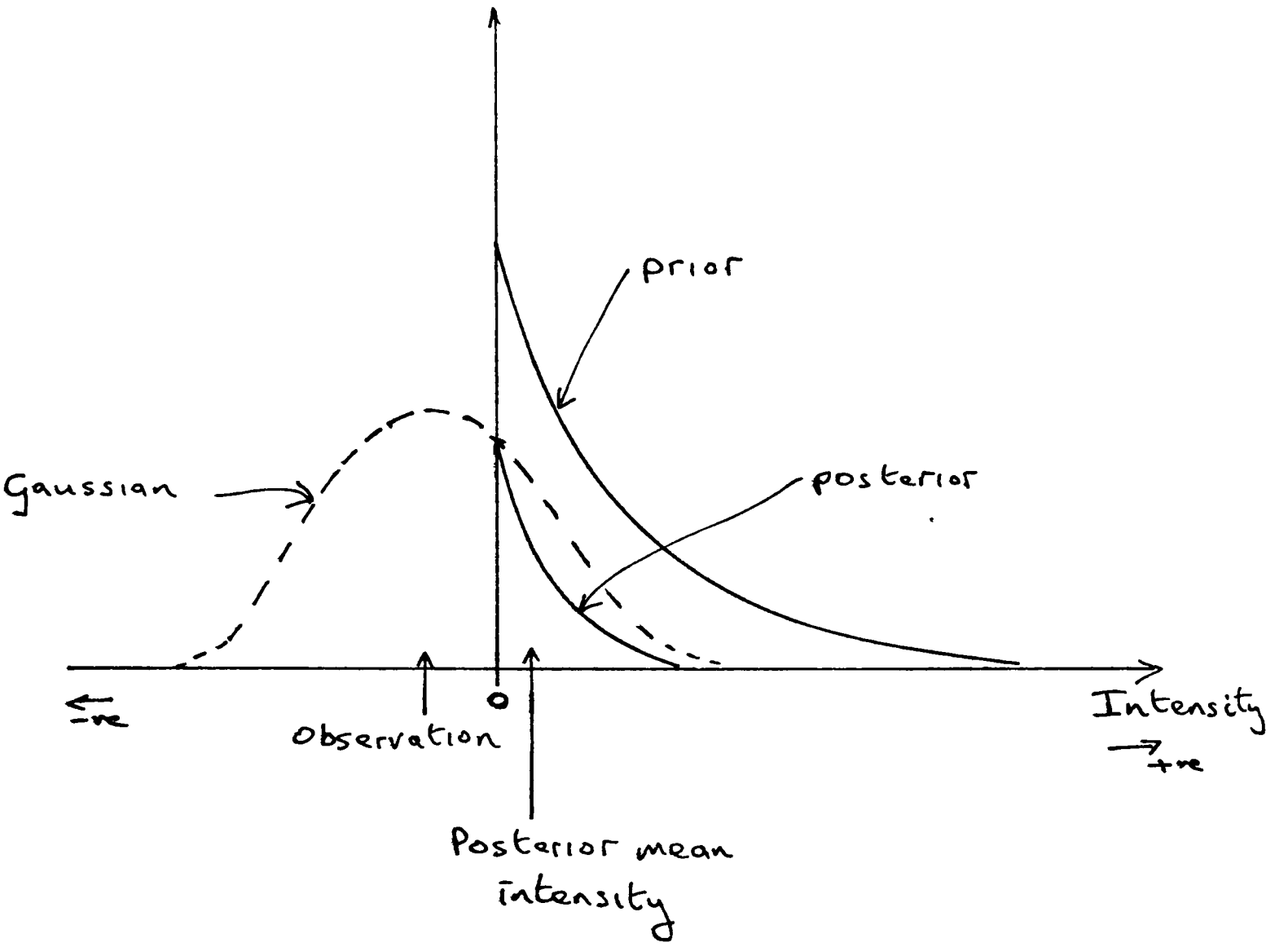


Figure 5.20 Calculation of the intensity for a negatively observed reflection. The observation is represented by a normal distribution with the mean value indicated. The prior distribution is the truncated positive Wilson distribution for an acentric reflection in this case (i.e. at I=0, $P(I) = \sum^{-1}$). The posterior distribution is proportional to the prior x observed distribution. The posterior mean intensity is now the output intensity for that observation.

$$(\bar{I})^{N(Z)} = \text{erf}\left(\sqrt{\frac{Z}{2}}\right)$$

Figure 5.21 shows a plot of N(Z) versus Z for the centric and acentric cases. The plotted curves correspond to the theoretical values estimated from Wilson's statistics and the plotted points the actual values for the 6PGDH native data set. The curve fits the points well and shows that the use of Wilson's probability distribution for the intensities is valid at least for those intensities with $0 \lesssim Z \lesssim 1$. Similar agreement is found for the derivative data.

In summary, the effect of the method is to make the negative intensity observations positive in a way consistent with the expected intensity distribution. The correct estimation of standard deviations for input to the program is important since the normal distribution $N(\bar{I}, \sigma^2)$ for each reflection and consequently the posterior distribution $P(I|\bar{I})$ depends fairly critically on the estimated $\sigma(I)$'s. This was why particular care was exercised in estimating the σ's in SSM. For strong reflections the posterior distribution is barely different from normal and the estimated intensity unchanged.

5.5 Analysis of the derivative

Scaling of the data

This involved the analysis of the Friedel equivalents for the derivative data set and the scaling of the derivative to the native camera data sets as well as the scaling of the native 6 Å 4 circle profile fitted data to the native 6 Å camera data.

Of particular importance in this work is the usefulness of the anomalous differences. When comparing Friedel related reflections from the same crystal it could be observed that in one region of reciprocal space the reflections $F(\underline{h})$ tend to be consistently larger than $F(\bar{\underline{h}})$ reflections, while in another region the opposite may be true. Such systematic differences could be due to a number of causes such as absorption by the crystal and capillary, non-uniformity of the X-ray beam or errors in the misorientation matrix.

To check for such effects the ratio:

$$\sum F(\underline{h})/F(\bar{\underline{h}})$$

was analysed using the program FRIEDANAL (Appendix 4) in blocks of φ , ξ and groups of levels about the mounting axis, c, for significant deviation from unity and perhaps trends in the ratio. This analysis was performed prior to the scaling of the derivative to native data and after treating weak reflections. The interval values for φ , ξ and levels about c were selected empirically to ensure that sufficient reflections were contained within each block. Hopefully local variations were not damped down by using blocks containing too many reflections. This information on the useful interval size for analysis was utilized subsequently for derivative to native scaling.

The ratios in each block were remarkably close to 1.0 even for reflections collected at the edges of the film though deviations from unity certainly increased towards the outer limit of the film. It was decided not to apply the calculated corrections as there seemed to be no systematic deviations from unity.

The derivative and native camera data sets were scaled together by making the ratio

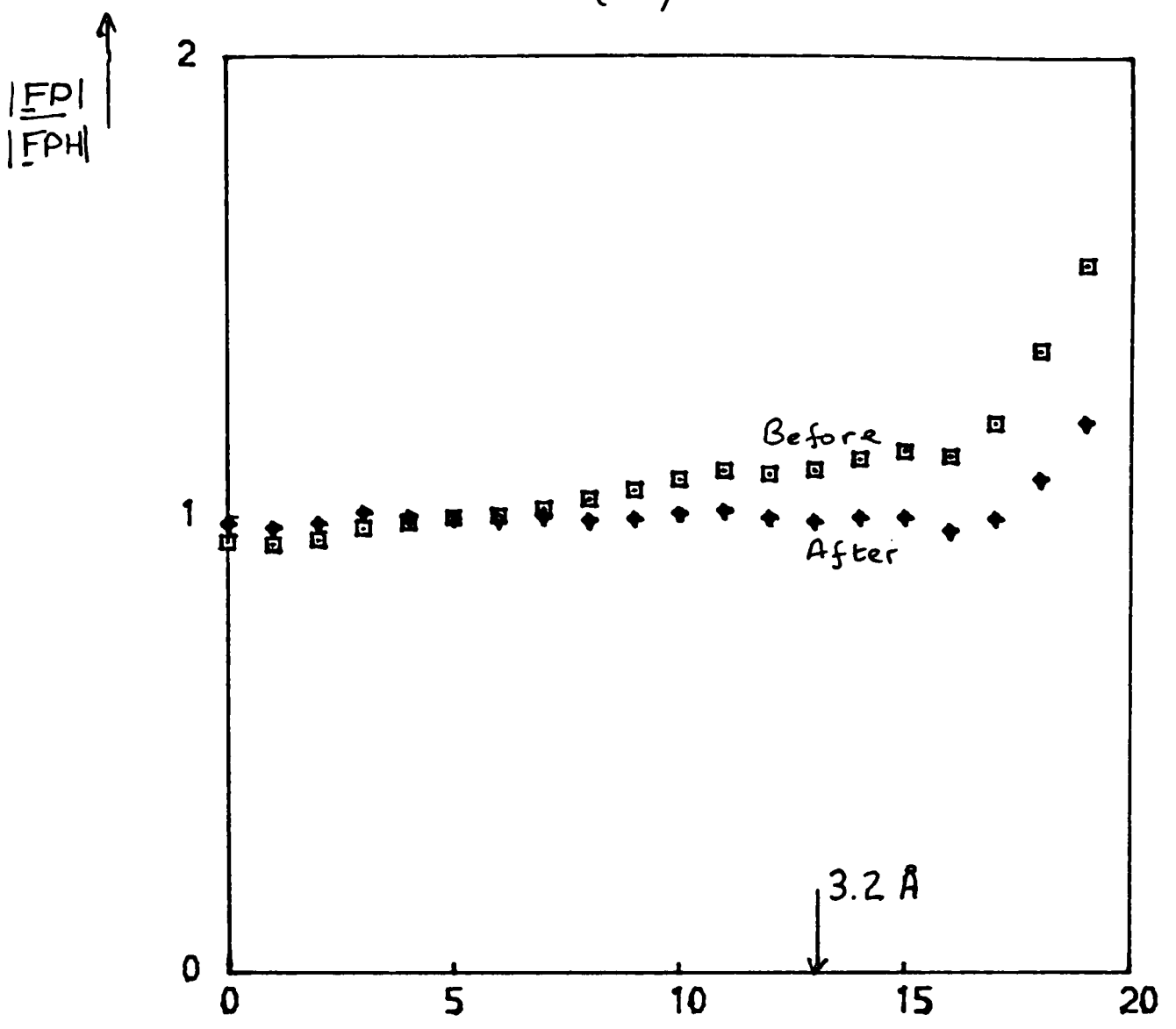
$$\frac{\sum |\underline{F}_P|}{\sum |\underline{F}_{PH}|} = 1.0$$ using the program ANSC.

This was achieved by analysing this ratio in blocks of reciprocal space as a function of φ , ξ and levels about c. In nearly all blocks there were at least 50 reflections. In Figure 5.2a the derivative to native scale factor is shown plotted against $4\sin^2\theta/\lambda^2$ both before and after scaling.

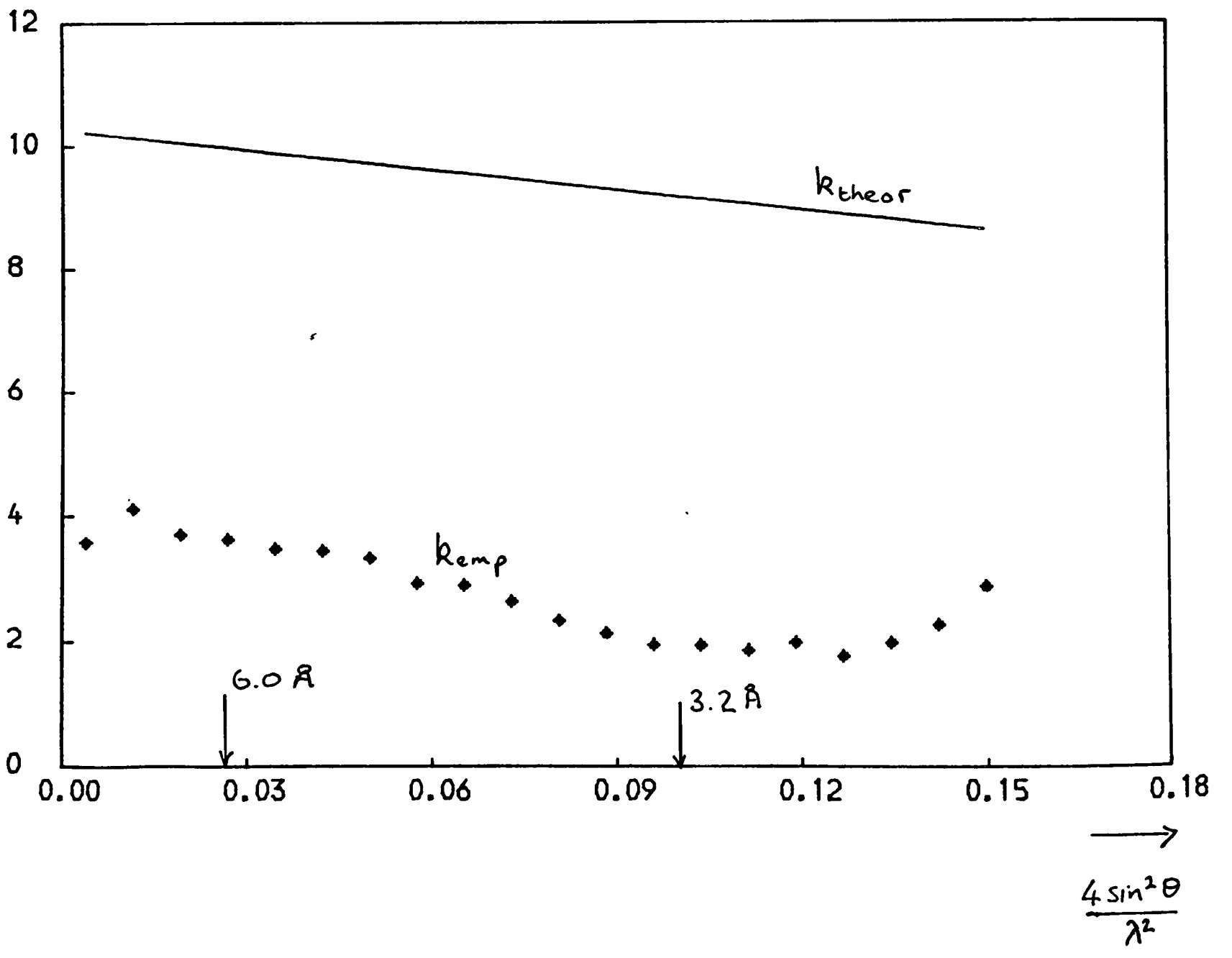
The mean fractional isomorphous difference $\frac{\langle |\underline{F}_{PH}| - |\underline{F}_P| \rangle}{\langle |\underline{F}_P| \rangle}$

was 14.3% (10.3% to 6 Å and compares with a value of 8.34% for the 4 circle 6 Å data) and the overall k_{emp} = 2.29 as output by ANSC. Figures 5.2a b and c show k_{emp} and the mean fractional isomorphous difference plotted in intervals of $4\sin^2\theta/\lambda^2$, the latter increases substantially beyond a resolution of

(a)


|F_P|
|F_PH|
2
1
0
Before
After
3.2 Å
0
5
10
15
20

(b)


12
10
8
6
4
2
0
k_theor
k_emp
6.0 Å
3.2 Å
0.00
0.03
0.06
0.09
0.12
0.15
0.18
4 sin²θ / λ²

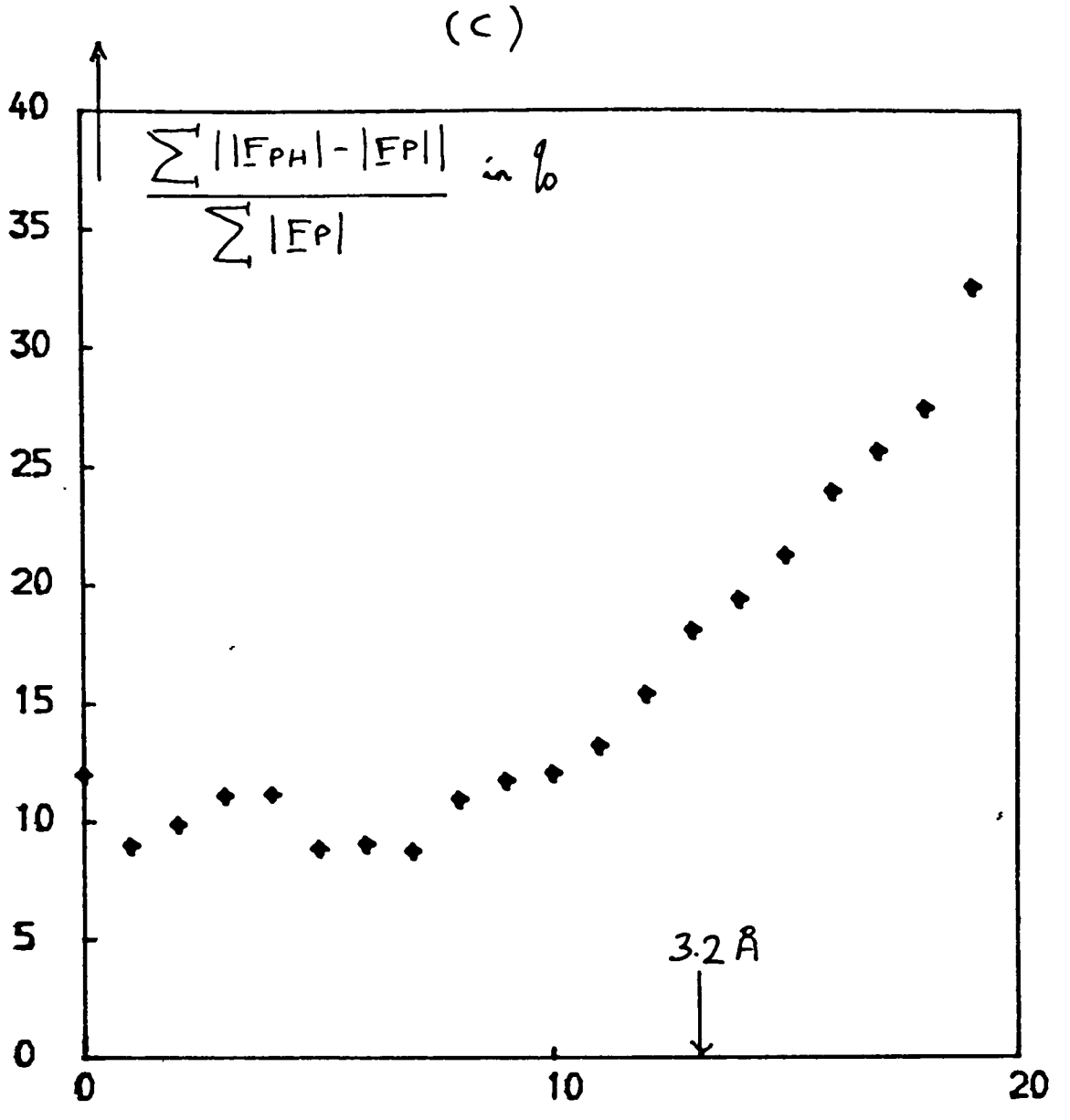


Figure 5.22

(a) The derivative to native scale factor before (⊡) and after (⊕) the scaling of derivative to native in blocks of reciprocal space.

(b) k_{emp} and k_{theor} for the Pt derivative. Compare with Fig. 3.6.

(c) The mean fractional isomorphous difference.

All plotted against $4 \sin^2\theta/\lambda^2$; (a) and (c) in intervals = 0.0077.

3.2 Å indicating perhaps that at high angles errors in the data are substantial, so increasing the isomorphous differences. The close agreement of the cell parameters indicates that this effect is not due to lack of isomorphism of the native and derivative series.

The scale between the camera F_P's and the four circle F_P's showed no trends in k or l but a significant trend in h. The extreme of the scale values in each case were $\{1.87, 2.45\}_h$, $\{2.07, 2.26\}_k$, and $\{2.08, 2.21\}_l$. This can be understood by reference to Figure 5.3 which shows the relationship of the crystal morphology to reciprocal space. At $\varphi = 0^\circ$ the absorption experienced by the diffracted beam will decrease markedly as h increases i.e. $\langle F_P \text{ camera} \rangle$ will increase as a function of h, all other things being equal. The F_P's from the four circle having been corrected for absorption will therefore have to have an increasingly large scale factor applied to compensate for the increase of "F_P camera" with h. The scaled four circle data (on h) compared well with the camera data in terms of h, k, l, $4\sin^2\theta/\lambda^2$ and $|\underline{F}|$, though for the latter the scaling had not been effective for the weak reflections The new scale between derivative and native in intervals of $|\underline{F}|$ was 1.1868 for the weak reflections compared with a variation of 1.0186 to 0.9329 for the remainder, more intense reflections.

This scaled data was used finally to fill in the blind region to 6 Å and other missing reflections for the final native electron density map. It should be emphasized that the data used in the refinement of the heavy atom parameters was entirely collected on the camera and not "contaminated" with data from the diffractometer until the calculation of the final electron density map.

Wilson plot

A.J.C. Wilson (1949) showed that if the asymmetric unit of the crystal contained a relatively large number of randomly distributed atoms of approximately the same scattering factor f_j then the mean intensity of the reflections in a given range of $\sin^2\theta/\lambda^2$ is given by:

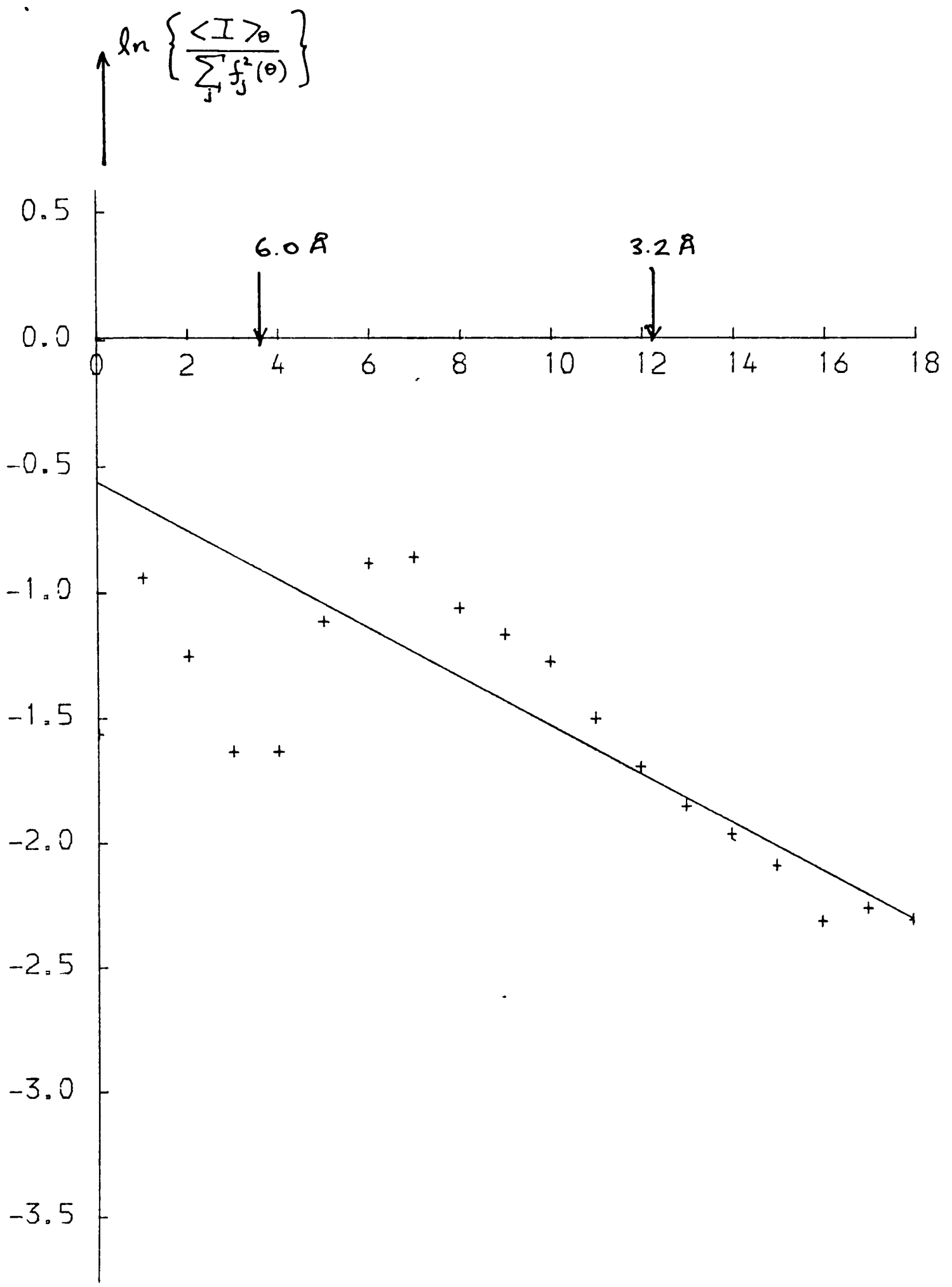


Figure 5.23 Wilson plot for the 6PGDH native data to 2.6 Å resolution in intervals of $\sin^2\theta/\lambda^2 = 0.002$.

$$\langle I \rangle_\theta \propto \sum_{j=1}^{N} f_j^2(\theta)$$

Now, allowing for thermal vibration of the protein atoms and for a scale factor K to put the data onto an absolute scale we have:

$$\langle I \rangle_\theta = K \left\{ \sum_{j=1}^{N} f_j^2(\theta) \right\} \exp(-2B\sin^2\theta/\lambda^2)$$

Rearranging:

$$\ln\left\{\frac{\langle I \rangle_\theta}{\sum_j f_j^2(\theta)}\right\} = \frac{-2B\sin^2\theta}{\lambda^2} + \ln K$$

Hence a plot of $\ln\left\{\frac{\langle I \rangle_\theta}{\sum_j f_j^2(\theta)}\right\}$ v. $\sin^2\theta/\lambda^2$ should be a straight line with slope -2B and intercept ln K.

Figure 5.23 shows the Wilson plot for the 6PGDH native data to 2.6 Å resolution, calculated assuming that the 6PGDH crystal unit cell contained 28571 nitrogen atoms giving a total molecular weight = 400,000 Daltons in the unit cell. It is seen that there is a significant deviation from linearity especially in the low angle region, which indicates a breakdown in the assumptions underlying the Wilson equation. The 6PGDH structure like other proteins is not made up of randomly distributed atoms because of the secondary structural elements associated with the protein. The shape of the curve means that there is some flexibility therefore in the calculation of the absolute scale and temperature factor. It was for this reason that an absolute scale for the 6 Å diffractometer data was not calculated in this manner. The values estimated from the plot, for the photographic data, of the temperature factor and absolute scale were 24 and 1.8 respectively. A similar plot for the derivative (before derivative to native scaling) gave B ≃ 33 and an absolute scale of 1.5.

Check on heavy atom substitution

At this point in the analysis a difference Fourier map was calculated with terms $(|\underline{F}P_{PT}| - |\underline{F}P|)\exp^{i\alpha}Au$, where the structure amplitudes were measured photographically and the α_{Au} were the single derivative Au phases (see § 3) estimated originally from data collected on the diffractometer. Several sections of the map are shown in Figure 5.24 and clearly show the

Table 5.8 Peak heights on the difference Fourier maps with coefficients (FP_{PT}-FP) calculated using α_{Au} and $\alpha_{(Au+Pt)_2}$. The Pt isomorphous differences were calculated using the photographic data.

Peak	Sectional coordinates[a]			Peak heights	
	x	y	z	α_{Au} map	$\alpha_{(Au+Pt)_2}$ map
PTA	13	6	21.5	59	121
PTB	0.5	8.5	11	52	94
PTC	5	14.5	24	45	89
⟨bg⟩				24	24

Compare with Table 3.12.

[a] In units of 1/40, 1/72, 1/48 in x, y, z, respectively.

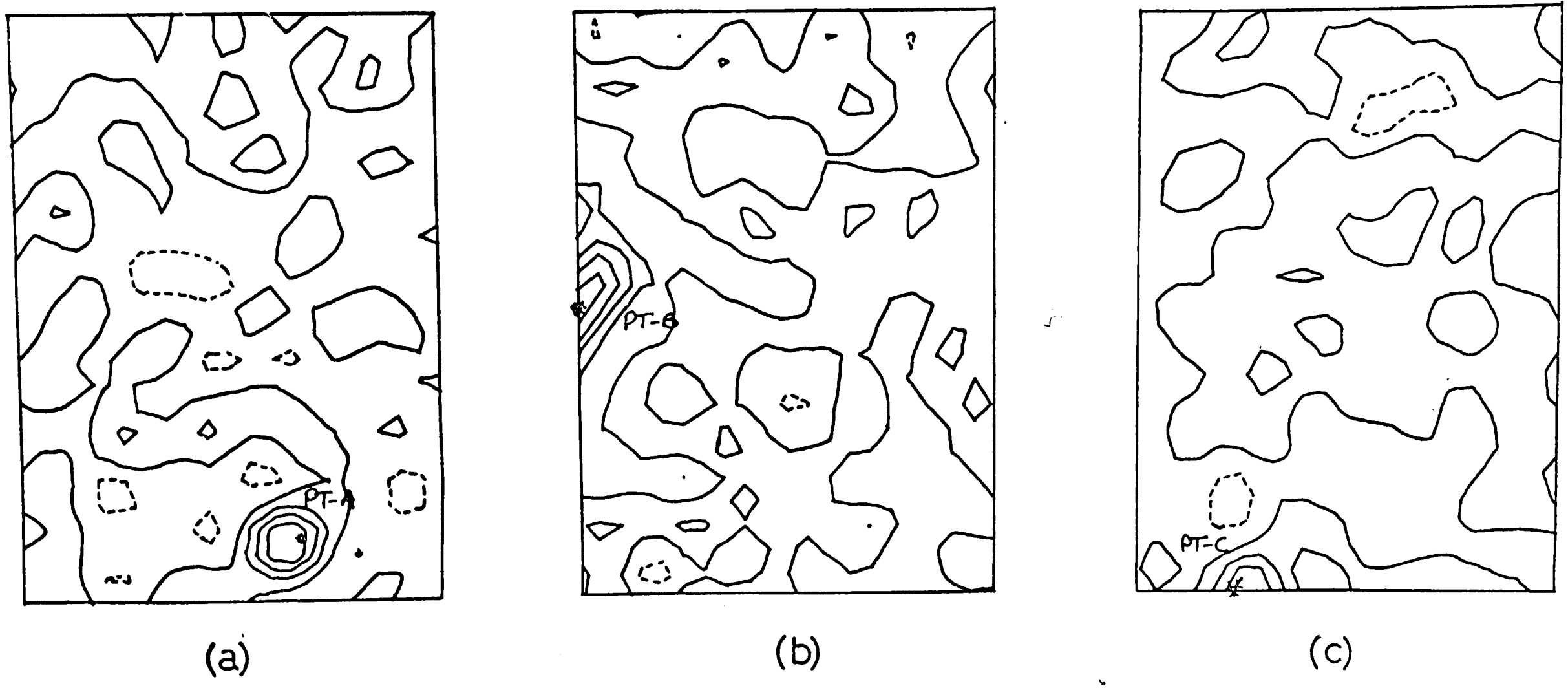


Figure 5.24 Selected sections of the difference Fourier calculated with coefficients (FP_{PT}-FP) from the photographic data and α_{Au} (based on diffractometer data). Sites PTA, PTB and PTC are marked. See Table 5.8 opposite. (a) y = 6/72 b; (b) y = 9/72 b; (c) y = 15/72 b.

---- negative, —— positive. Contours start at 0 and the contour interval is ± 15.

three heavy atom sites PTA, PTB and PTC (see Table 5.8).

Estimation of $|\underline{F}_H|$

The program FHLEKR was used to calculate F_{HLE}'s using the k_{emp} output by ANSC and F_{HLEB}'s using a k_{theor} (see § 3). Table 5.9 shows the average values of $|\underline{F}_H|$ for the centric and acentric terms in each case. The centric average estimate of $|\underline{F}_H|$ should be 0.9 the acentric average estimate which is achieved best with F_{HLE}'s calculated using k_{emp}. The program also estimates a derivative to native scale factor (see Table 5.9) and outputs useful data statistics. Structure amplitudes under 200 (where the maximum $|\underline{F}p| = 2310$) had $F \lesssim 3\sigma(F)$ and $\Delta_{ISO} \lesssim 1\sigma(\Delta_{ISO})$. Also, $\Delta_{ISO} \lesssim 1\sigma(\Delta_{ISO})$ at a resolution of ≈ 3.2 Å. At this resolution $\langle \Delta_{ANO} \rangle \lesssim 0.75\,\sigma(\Delta_{ANO})$. The output file of F_{HLE}'s, calculated with a k_{emp}, was used in the refinement of the heavy atom parameters.

Patterson maps

Isomorphous difference Patterson maps were calculated at three different resolutions 6 Å, 3.2 Å and 2.7 Å involving 1371, 9204 and 14866 reflections, respectively. All spuriously high isomorphous differences (>400) were rejected. Each had an origin peak of 1000 units. Figure 5.25 shows the sections containing the Harker vectors for the 6 Å map and Figure 5.26 sections containing Harker and some cross vectors for the 3.2 Å map. The 2.7 Å map contained few vectors and is not shown. These maps should be compared with Figure 3.10.

Refinement

Having established that the three major sites were located on the protein consistent with the low resolution study (see § 3) it was necessary to refine not only the occupancies and positions but also the temperature factors.

The refinements that are described are performed only on data collected from the camera which has been merged from all the derivative crystals. No investigation has been made of possible differences in the heavy atom occupancy from crystal to crystal due to variations in soaking time or errors

Table 5.9 Comparison of the F_{HLE}, F_{HLEB} estimates made for the heavy atom structure scattering amplitude $|\underline{F}_H|$ using a k_{theor} and a k_{emp}. The centric estimate should= 0.9 the acentric estimate on average for the whole data set. (This is achieved most effectively with an F_{HLE}, k_{emp}.)

	$\langle F_{HLE} \rangle k_{emp}$	$\langle F_{HLEB} \rangle k_{emp}$	$\langle F_{HLE} \rangle k_{theor}$	$\langle F_{HLEB} \rangle k_{theor}$
centric	72.6[a]	72.6[a]	73.5[b]	73.5[b]
acentric	83.0[a]	57.1[a]	142.9[b]	72.3[b]

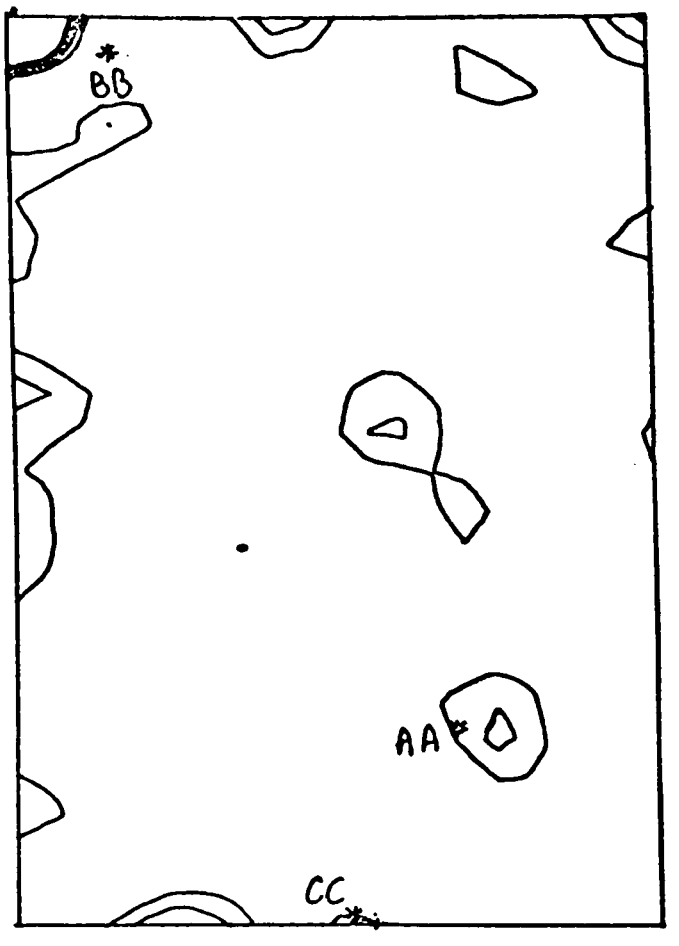
Note the Kraut scales used were:

	F_{HLE} k_{emp}	F_{HLEB} k_{theor}
Pt	1.025	1.031

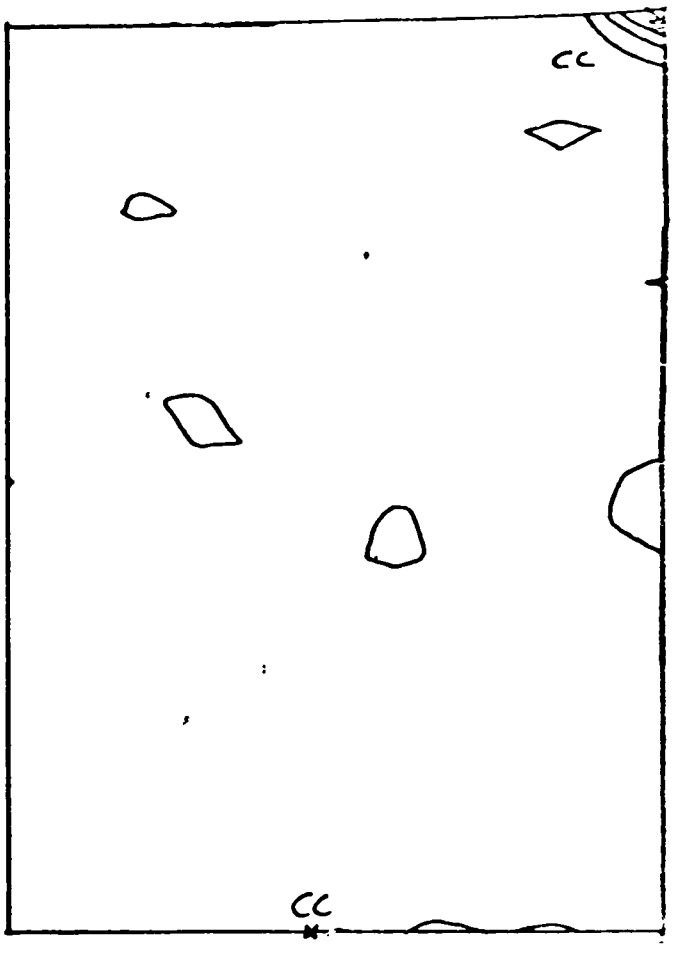
[a] In this calculation there were 1554 centric and 13817 acentric terms.

[b] In this calculation there were 1555 centric and 10193 acentric terms.

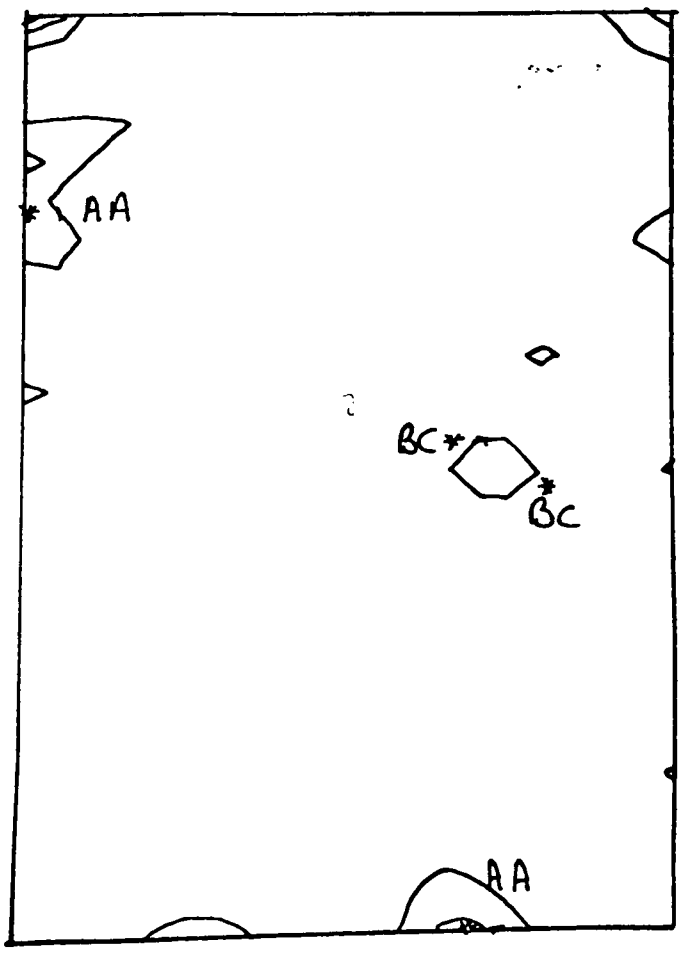
Figure 5.25 Isomorphous difference Patterson to 6 Å using the photographic data, showing the Harker vectors AA, BB, CC. Compare with Figure 3.10. The origin peak is 1000 units and the contour interval is 30 starting at 30.



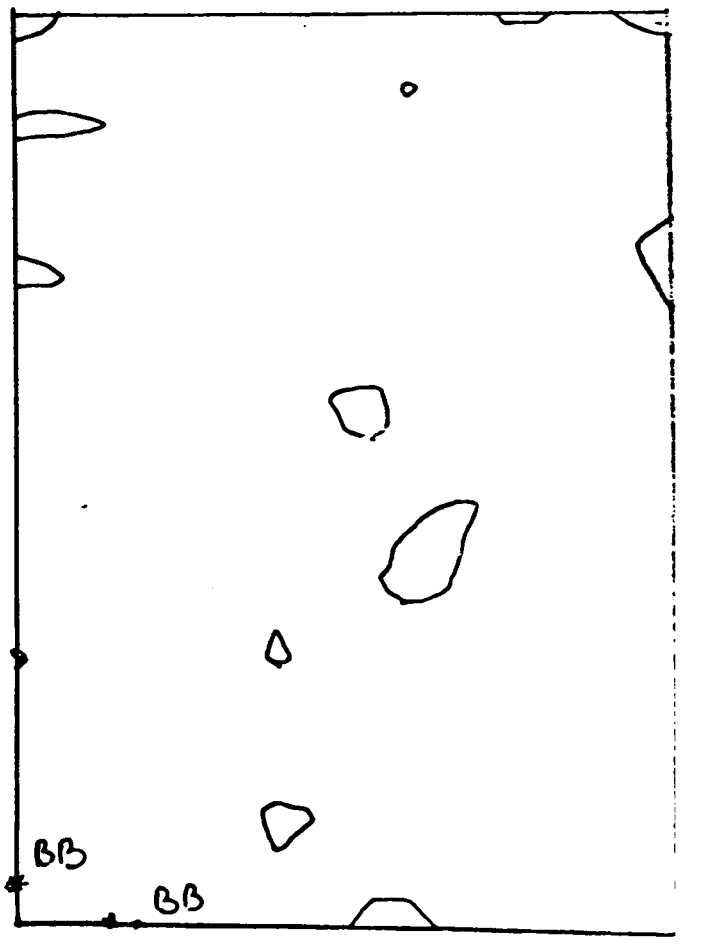
V=0/72



V=6/72

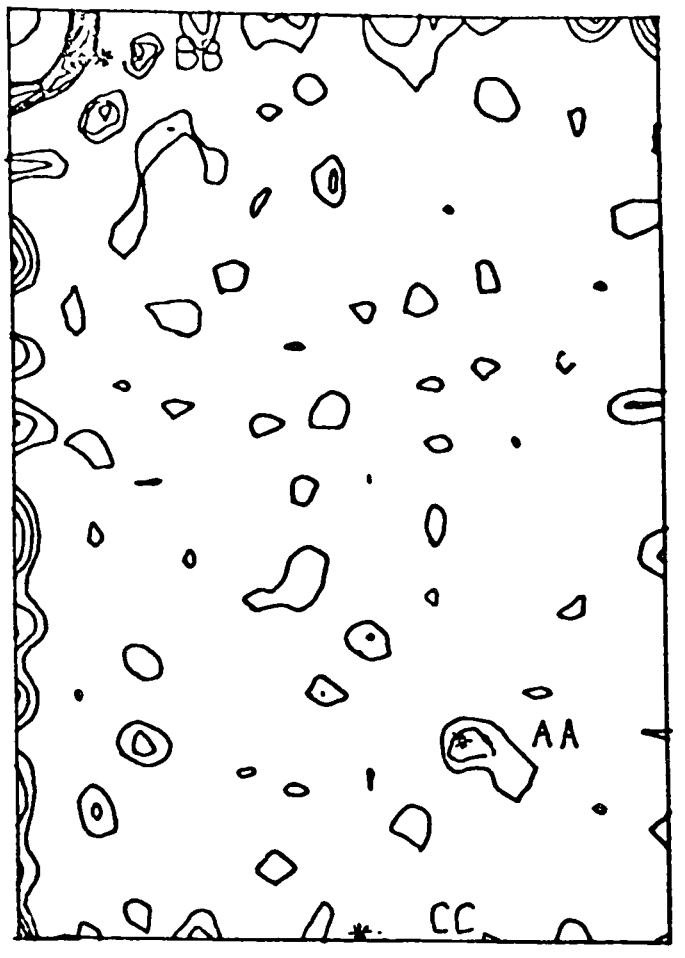


V=13/72

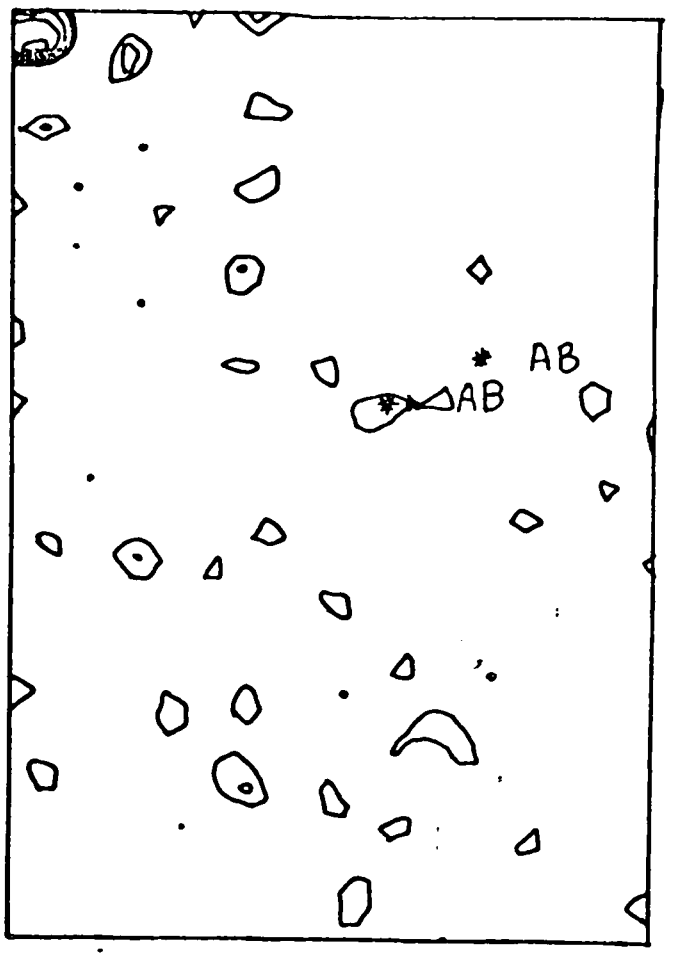


V=16/72

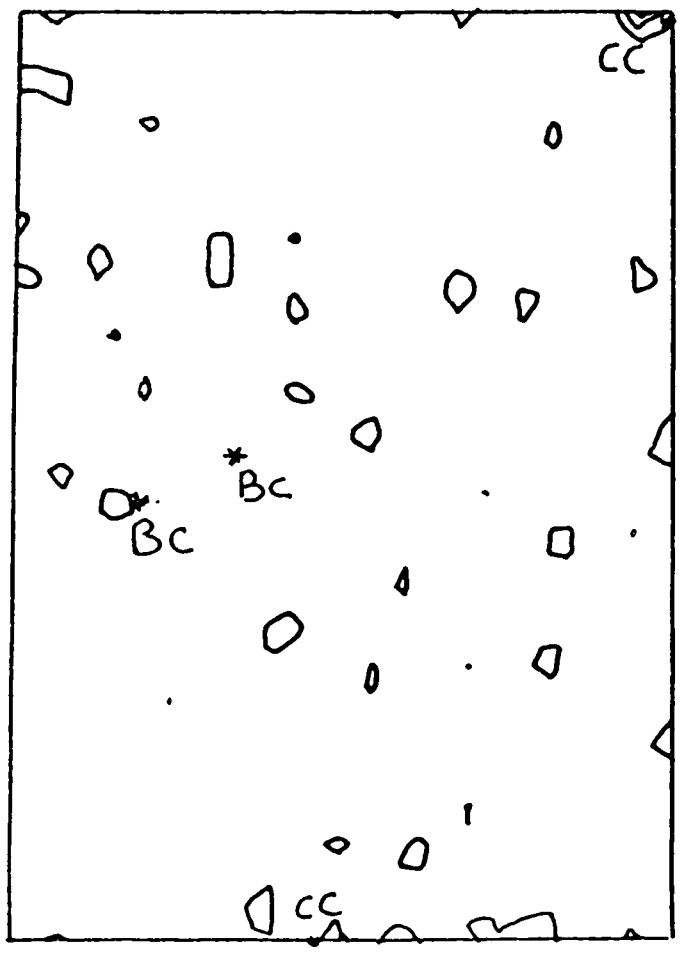
Figure 5.26 Isomorphous difference Patterson to 3.2 Å using the photographic data. The Harker and some of the cross vectors are marked. Compare with Figures 5.25 and 3.10. The origin peak is 1000 units and the contour interval is 10 starting at 10.



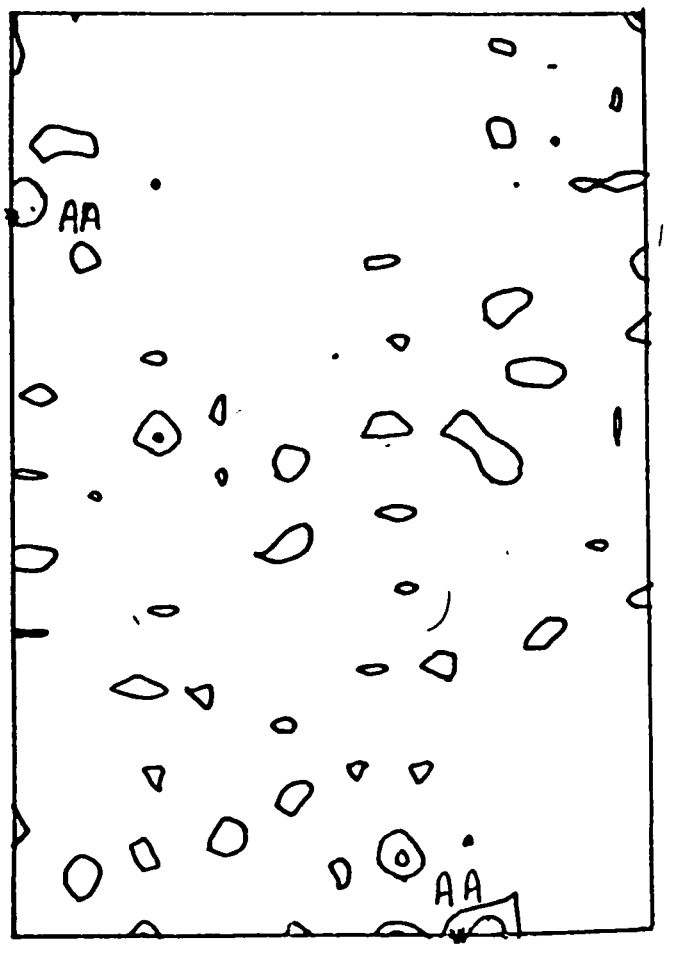
V=0/144



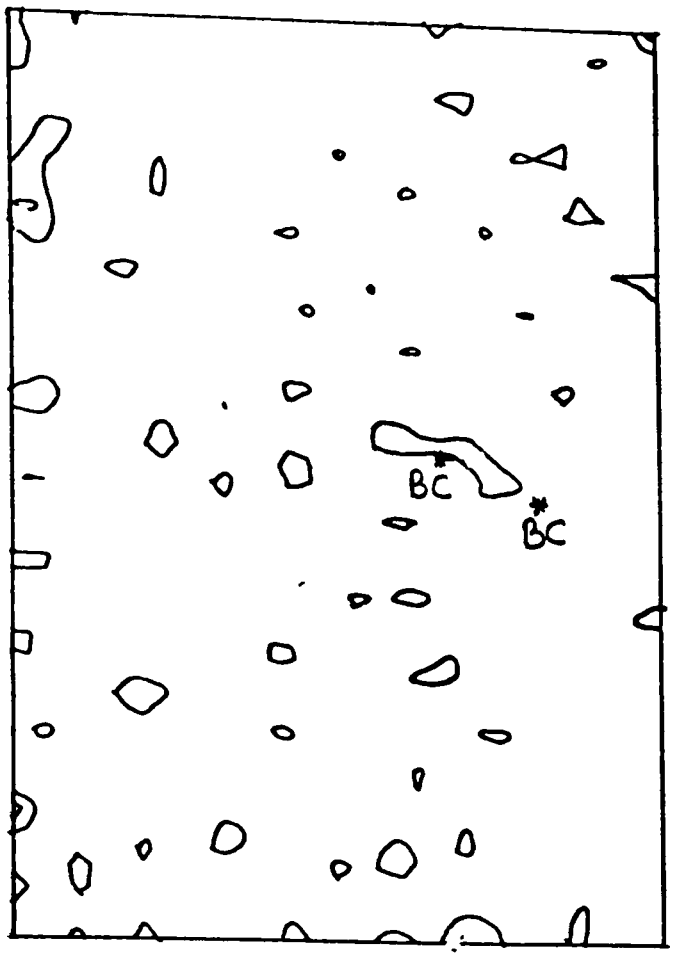
V=4/144



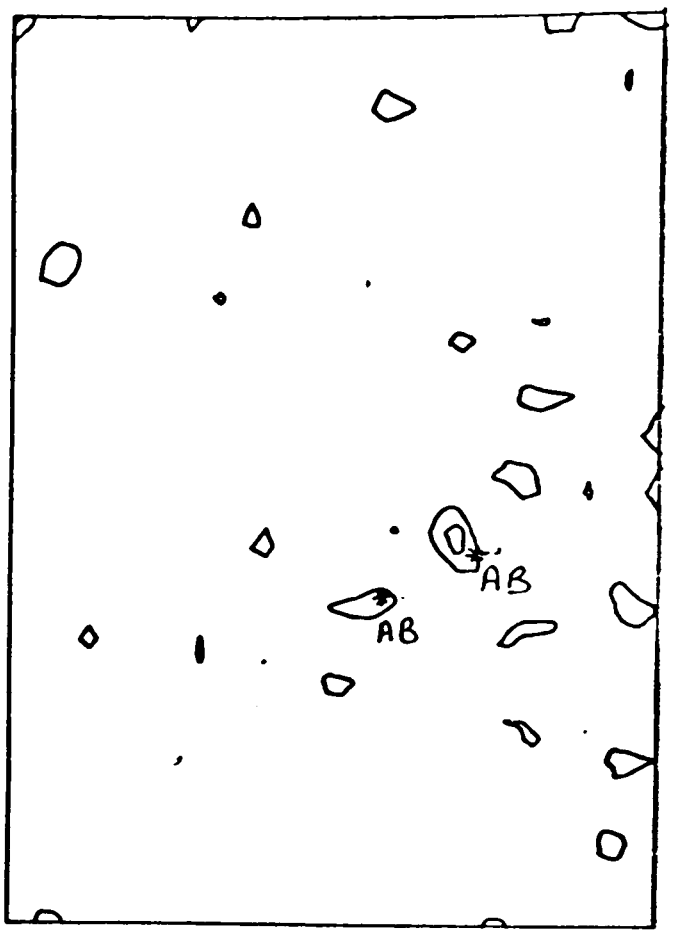
V=13/144



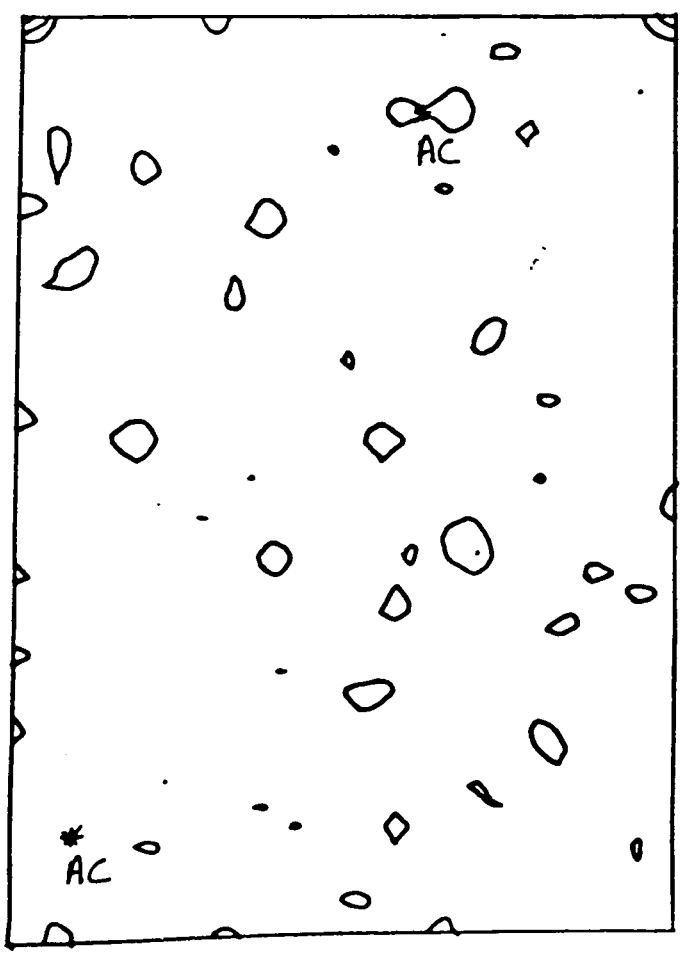
V=25/144



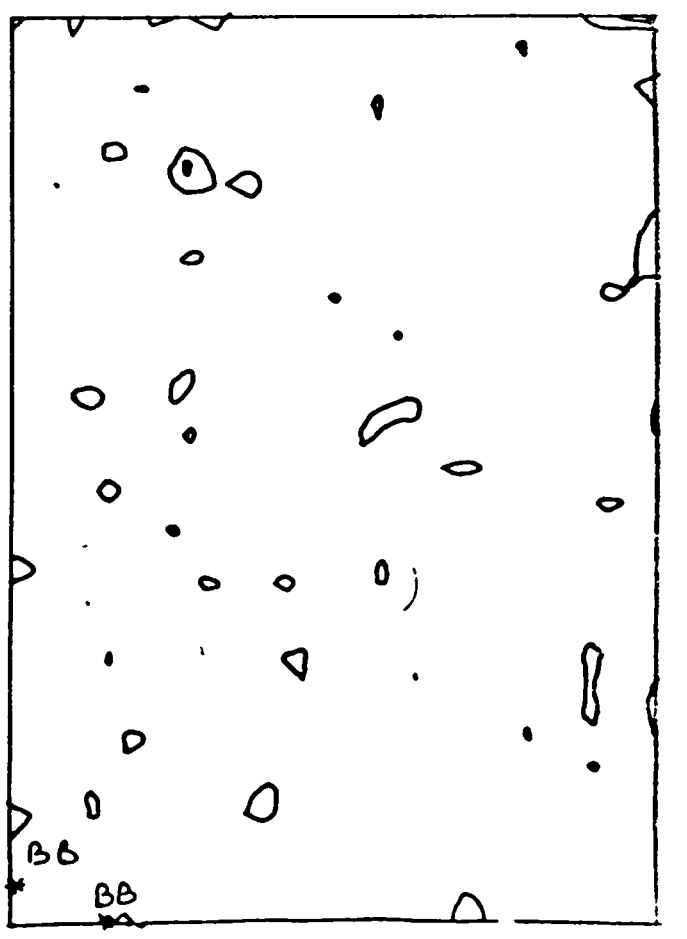
V=26/144



V=29/144



V=30/144



V=33/144

in the concentration of the heavy atom reagent in each soaking solution.

The rationale adopted for the refinement was based on the fact that sufficient centric terms existed (~1500 to 2.7 Å) for the data to parameter ratio to be sufficiently overdetermined without using acentric data.

Except where stated otherwise the derivative to native scale factor was the Kraut scale (= 1.025) and was not refined and weights of $1/\sigma^2(F_{NOBS})$ were used, and also positional parameters were refined in each case. Those reflections with spuriously high isomorphous differences (>400) were rejected, amounting to 16 out of approximately 15000 reflections. Reflections with $F < \sigma(F)$ were also rejected.

Three groups of refinements were performed.

Set 1. Alternating refinements of occupancies and temperature factors using the centric data to 2.7 Å were performed. An initial three dimensional 6 Å refinement of the occupancies with the temperature factors set at 15.0 (see §3.5) provided starting values for the heavy atom parameters for this series of refinements. The final values obtained here for the heavy atom parameters are given in Table 5.10.

Set 2. This involved the refinement of the occupancies only, in shells of $4\sin^2\theta/\lambda^2$, with the temperature factors arbitrarily set to 15.0. The starting values were obtained from a 2.7 Å centric refinement of the occupancies and positions with the B's set to 15.0. Table 5.11 summarizes the results of these refinements. For each site the occupancy increases with $4\sin^2\theta/\lambda^2$ indicating that the temperature factors are too high. At high angles ($4\sin^2\theta/\lambda^2 > 0.10$) the occupancies vary oddly, for instance, in shell 9, the occupancy for site B = 0.0013 and in shell 10, for the same site, the occupancy = 1.0594. When the width of the $4\sin^2\theta/\lambda^2$ annulus was increased at high angles the refinement stabilized (shells 11 and 12).

From the intercept and gradient of a plot of ln(occupancy) v. $\sin^2\theta/\lambda^2$ it is possible to obtain a better estimate of the occupancy and a correction to the temperature factor respectively, i.e.

Table 5.10 Comparison of the refinements performed on the Pt derivative

(a)

Parameter	Centric 2.7 Å $1/\sigma^2$ (set 1)	Centric 3.2 Å $1/\sigma^2$ (set 3)[a]	Centric 3.2 Å unit weights (set 3)	Centric 3.2 Å FHLESC $1/\sigma^2$ (set 3)	Centric 6 Å $1/\sigma^2$
scale factor	1.025	1.025	1.025	0.9891	1.025
Site A					
occupancy	0.3930	0.3781	0.3111	0.3458	0.4022
x	0.3258	0.3251	0.3260	0.3233	0.3232
y	0.0878	0.0876	0.0895	0.0865	0.0862
z	0.4489	0.4490	0.4486	0.4490	0.4504
B	2.63	3.30	2.08	-0.21	15.0
Site B					
occupancy	0.2477	0.2436	0.1745	0.2437	0.2397
x	0.0241	0.0239	0.0206	0.0227	0.0296
y	0.1164	0.1163	0.1180	0.1180	0.1228
z	0.2357	0.2359	0.2376	0.2366	0.2365
B	-17.49	-23.69	-16.88	-24.66	15.0
Site C					
occupancy	0.2271	0.2240	0.2276	0.2091	0.2470
x	0.1177	0.1192	0.1205	0.1207	0.1297
y	0.2037	0.2034	0.2039	0.2028	0.2046
z	0.4989	0.4997	0.4995	0.4998	0.5000
B	-6.27	-9.36	-11.84	-7.64	15.0
nref	1522	1117	1117	1109	289
R	73.2	72.6	70.1	71.4	62.1
wR	60.2	59.7	72.3	60.4	50.7
gradient	0.14	0.19	0.12	0.20	0.47
g.F.	1.3387	1.4955	66.9715	1.4988	1.7758
E[b]	71.03	68.47	66.61	65.91	62.43

[a] Final parameters used in the phase calculation. See Table 5.13.

[b]
$$E = \left(\sum \frac{(F_{HOBS} - F_{HCALC})^2}{nref} \right)^{\frac{1}{2}}$$

For the definition of the other symbols see § 3.5.

Table 5.11 Results of the shell refinements

Shell number	$4\sin^2\theta/\lambda^2$	Occupancy PTA	PTB	PTC	R	wR	E	g	nref
1	0.0068	0.4787	0.2696	0.2242	62.2	51.7	70.3	0.32	122
2	0.0204	0.3130	0.2438	0.2515	64.7	50.0	57.0	0.28	167
3	0.0348	0.3597	0.2394	0.2071	71.2	61.6	56.1	0.30	166
4	0.0488	0.4553	0.1044	0.2128	78.7	69.6	77.0	-0.07	163
5	0.0624	0.3717	0.3942	0.3575	77.4	63.0	67.9	0.05	166
6	0.0768	0.5834	0.3426	0.3178	74.7	56.2	68.6	0.18	162
7	0.0908	0.6393	0.4635	0.3612	77.7	64.6	79.9	-0.01	153
8*	0.105	0.7446	0.0013	0.2758	79.1	63.8	74.5	0.12	153
9*	0.119	0.8405	0.3865	0.3363	75.8	64.3	82.4	0.07	134
10*	0.133	0.2794	1.0594	0.1374	64.2	52.7	69.2	0.35	136
11	0.098	0.7102	0.2525	0.3118	79.8	66.1	78.2	0,06	306
12	0.119	0.7604	0.2788	0.3086	75.8	66.6	78.5	0.09	423

* The occupancies obtained in these shell refinements were not plotted in Fig. 5.27 or used in the estimates given in Table 5.12. Note The heavy atom temperature factors have been arbitrarily set to 15.0 in these refinements.

Table 5.12 Estimates of heavy atom B's and occupancies from the ln O_j' v. $\sin^2\theta/\lambda^2$ plots shown below.

	3.5 Å		2.7 Å	
	B	occupancy	B	occupancy
PTA	0	0.3549	-13	0.3211
PTB	1	0.2146	2	0.2209
PTC	-9	0.1987	-2	0.2116

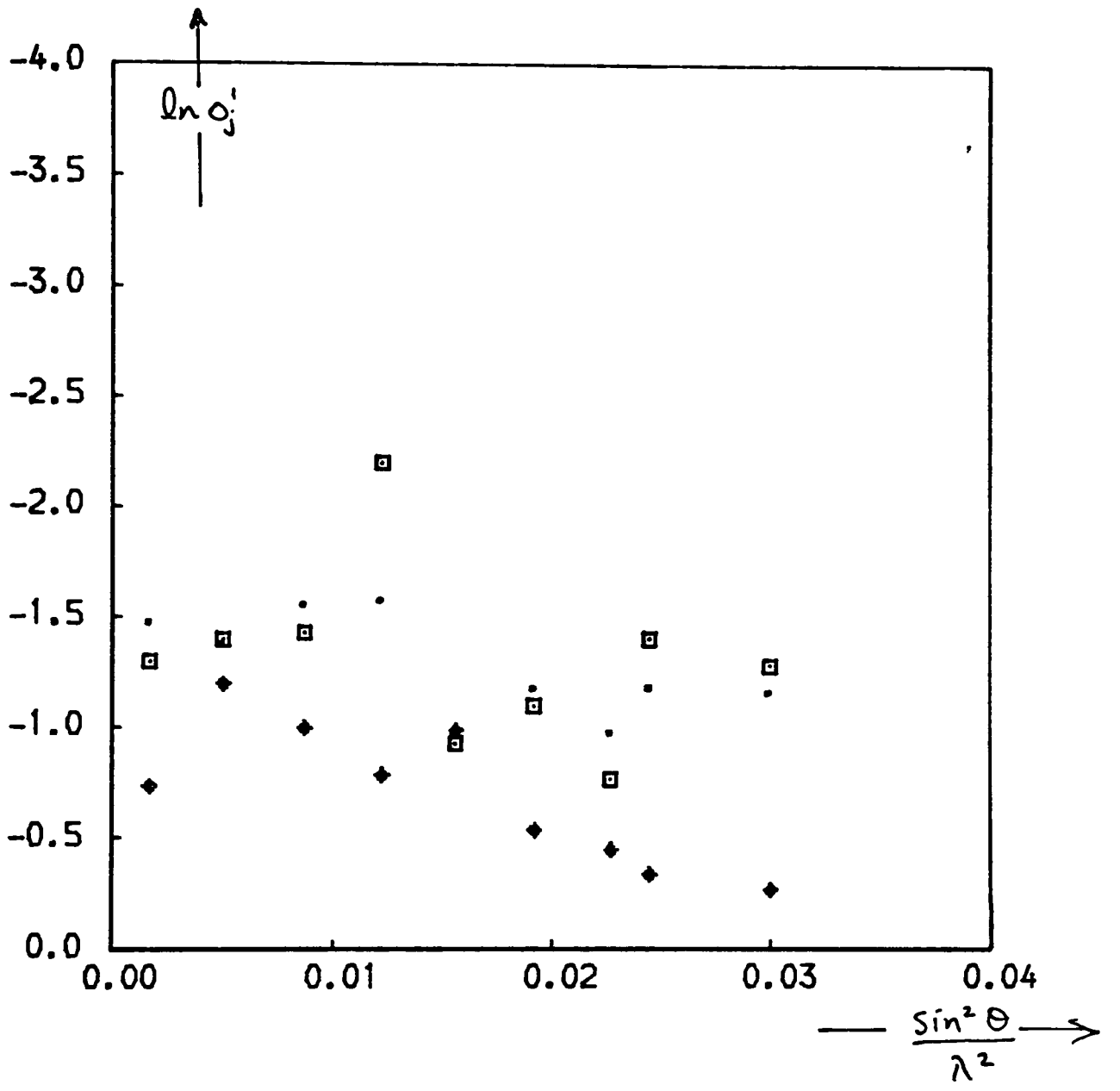


Figure 5.27 Plot of ln O_j' v. $\sin^2\theta/\lambda^2$ for heavy atom sites PTA (*), PTB (⊡) and PTC (.) where O_j' is the refined occupancy for a given shell shown in Table 5.11 opposite.

$$\underline{F}_{HCALC} = \sum_{j=1}^{N} f_j \, O_j' \exp(-B_j \frac{\sin^2\theta}{\lambda^2}) \exp(2\pi i \underline{h}.\underline{x})$$

where the summation is over N heavy atom sites.

Then if $O_j' = O_j \exp-(\Delta B_j \frac{\sin^2\theta}{\lambda^2})$

where O_j' is the jth site occupancy which varies with $\sin^2\theta/\lambda^2$ because of an incorrect estimate B_j of the true temperature factor $(B_j+\Delta B_j)$.

We have

$$\ln O_j' = \ln O_j - \Delta B_j \sin^2\theta/\lambda^2$$

and a plot of $\ln O_j'$ v. $\sin^2\theta/\lambda^2$ yields the improved estimate for the occupancy (from the intercept) and the correction to the temperature factor (from the gradient) for each site j. Figure 5.27 shows a plot of $\ln O_j'$ v. $\sin^2\theta/\lambda^2$ for sites PTA, PTB, PTC. The errors on each point are large and make the positioning of a straight line through the points difficult. The gradient depends on whether the plotted points corresponding to shells 7, 11 and 12 are taken as meaningful. A best least squares straight line was fitted to 3.5 Å and to 2.7 Å. The new estimates for the site occupancies and temperature factors are compared in Table 5.12 and depending on the resolution these estimates are significantly different. One thing is particularly clear, the high angle isomorphous differences($>$ 3.2 Å) give unreliable results.

Set 3. From the experience gained with the shell refinement it was decided to repeat the set 1 refinements but to a resolution of 3.2 Å instead of 2.7 Å. The same data rejection tests were applied (Δ_{ISO}'s $>$ 400 and reflections with $F < \sigma(F)$ were rejected). The occupancies and temperature factors were refined in alternate cycles (see Table 5.10).

For comparison a centric 3.2 Å unit weights refinement was run to check how much the low angle reflections were dominating the refinement since the higher angle reflections had lower weights. Also, in all the previous refinements it has been assumed that the Kraut scale was a reliable estimate of the derivative to native scale factor. Again, for comparison an FHLESC refinement was run (see § 3.5) on the centric data to 3.2 Å alternately

refining the occupancies and temperature factors as well as the scale factor. The same data rejection tests were applied.

Table 5.10 compares the results of these refinements. The heavy atom parameters are similar in the weighted and unit weight Kraut scale refinements except for the occupancies for sites A and B. The scale factor in the FHLESC refinement decreased markedly and gave a better E value.

Table 5.10 also includes the results of a centric 6 Å $1/\sigma^2$ refinement using the Kraut scale for comparison with the results shown in Table 3.21. The photographic data refinement compares very favourably with the diffractometer refinement showing very good correlation of FHOBS and FHCALC (ρ = 0.47 compared with 0.23). Occupancies of the 6 Å refinement shown in Tables 3.21 and 5.10 are on the same absolute scale if for the diffractometer refinement the absolute occupancy of Au1 is 0.755 and if for the photographic data the absolute scale is approximately 1.8.

In assessing which set of refined parameters to use for phasing, several results were important.

1. The high angle data (beyond 3.2 Å) were not considered meaningful enough to yield acceptable parameters because of the results of the shell refinement. This was anticipated earlier with the marked increase in the mean fractional isomorphous difference at about $\approx$ 3.2 Å (see Figure 5.21c).

2. The parameters from the unit and $1/\sigma^2$ weighted refinements were similar. It was felt that a weighting scheme was prefer able to reduce the influence on the refinement of the poorly measured terms.

3. The F_{HLESC} refinement involving variation of the derivative to native scale factor and alternating cycles of refinement of occupancy and positions gave better agreement parameters than the corresponding FHLEKL refinement using the Kraut scale. For instance, compare the E values (Table 5.10).

The FHLESC refinement parameters were used for phasing but in fact yielded substantial holes (several times the protein electron density) at

Table 5.13 Final parameters used for phasing to 2.7 Å resolution

Derivative to native scale	B_{ov} [a]	Site	Z [b]	x	y	z	B_j [c]	r.m.s. centric	E [d] all	r.m.s. F_{HCALC}
1.025	0.0	PTA	29	0.3251	0.0876	0.4490	3.3	137	96	168
		PTB	19	0.0239	0.1163	0.2359	-23.7			
		PTC	17	0.1192	0.2034	0.4997	-9.4			

[a] B_{ov} is the overall temperature factor to be applied to the derivative data for best fit with the native data.

[b] Z is given in electrons. In comparing these with the values given in Table 3.23 it should be remembered that the absolute scales for the data have been estimated in different ways.

[c] B_j, the individual heavy atom temperature factors.

[d] r.m.s. E is the r.m.s. lack of closure error $\langle (|\underline{F}_{PH}| - |\underline{F}_H + \underline{F}_P|)^2 \rangle^{\frac{1}{2}}$

each heavy atom site indicating that the scale factor was too low. So, in fact, the parameters finally used for phasing were from the FHLEKR refinement with alternating cycles of variation of temperature factors and occupancies.

Using this set of parameters an $F_{HOBS}e^{i\alpha HCALC}$ map and an $(F_{HOBS}-F_{HCALC})e^{i\alpha HCALC}$ map were calculated (to 2.7 Å) on the same scale and used 14515 reflections.

The $F_{HOBS}e^{i\alpha HCALC}$ map gave peaks at the three heavy atom sites PTA, PTB, PTC, of heights 1956, 1103, and 1055 units respectively. In the $(F_{HOBS}-F_{HCALC})e^{i\alpha HCALC}$ map there was a peak of 213 units (the largest positive peak on the map) at site A (X=26/80, Y=13/144, Z=43/96) with evidence of rippling around the site. That is, a peak of -134 at X=24/80, Y= 12/144, Z=43/96 and a peak of 100 at X=22/80, Y=14/144, Z=43/96. These results indicate that the refined occupancy of site A is too low and the temperature factor is too low (for the X-direction). The latter suggests that anisotropic temperature factors should be used. At site B there was a peak of 84 units and no discernible rippling. For site C there was no disturbance at the binding site. The largest positive peak not attributable to any of the sites was 102 units, which at about 10% of the peak heights of sites B and C was not considered to be a minor binding site.

Though this set of parameters (Table 5.13) was used eventually other possibilities need to be explored.

1. Calculation of phases in shells of $4\sin^2\theta/\lambda^2$ using the parameters from the shell refinement. This is useful where the variation of the heavy atom scattering with $\sin^2\theta/\lambda^2$ cannot be adequately explained with an exponential function.

2. Anisotropic temperature factor refinement needs to be attempted.

3. The most suitable refinement performed on the whole data set should be performed on the data from each crystal to check for wide variations in heavy atom occupancies and temperature factors from crystal to crystal.

A disturbing result is the occurrence of negative temperature factors for sites B and C. The effect of these is to compensate for the fall off of the atomic scattering factor for Pt giving an F_{HCALC} which is approximately constant in reciprocal space (see Figure 5.28). It would be interesting to see how the refinements changed after the application of an absorption correction to the data.

5.6 Phasing to 2.7 Å

Phases were calculated according to the method of Blow and Crick (1959) using the program PHASE as described in § 3.6.

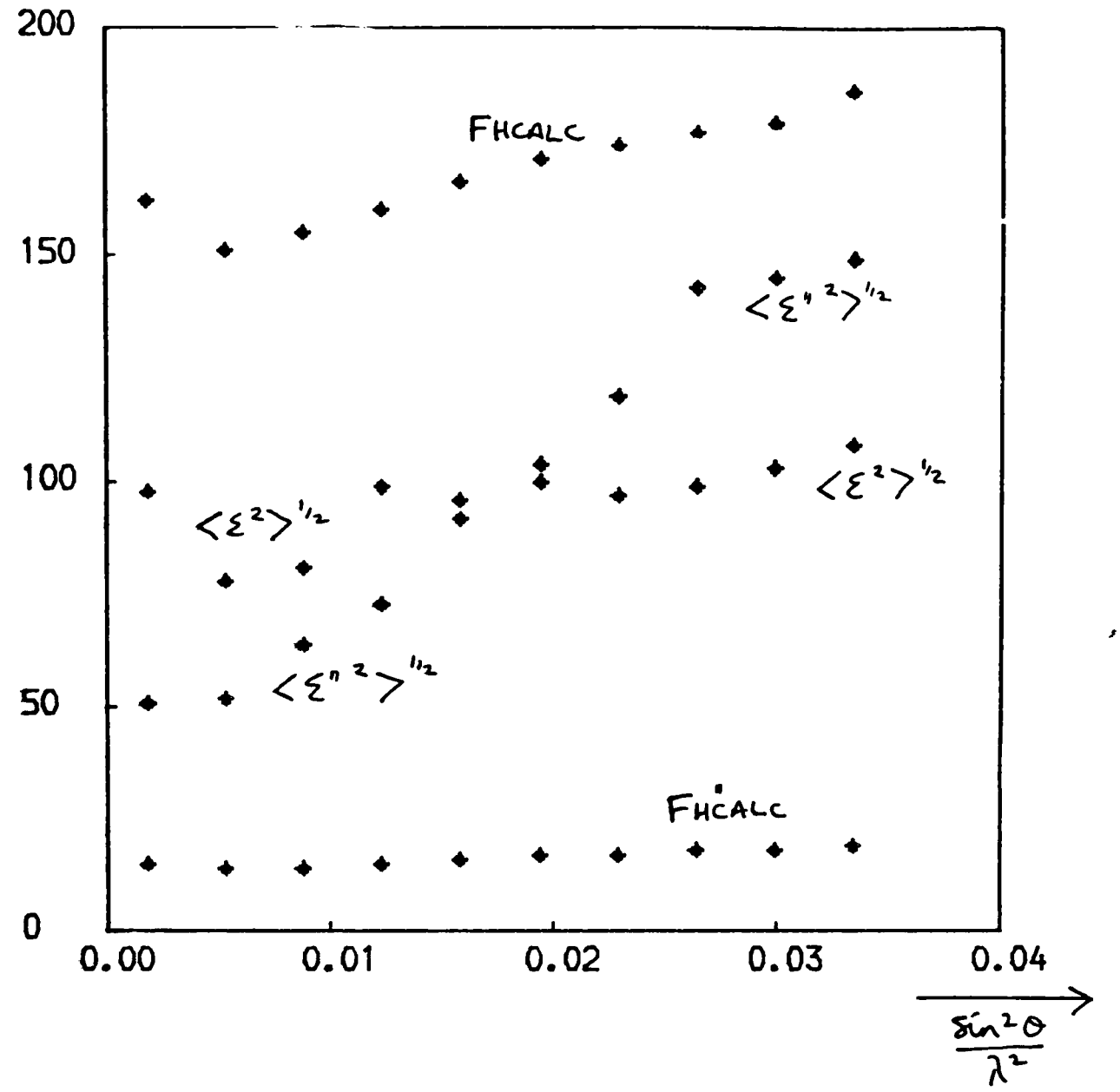
The isomorphous average errors, in ranges of $4\sin^2\theta/\lambda^2$, were estimated initially from the centric refinement. The anomalous E" were set initially to 1/3 E and a preliminary phase set calculated. The output average root mean square lack of closure errors from this phasing run were used in the input to a second phase calculation. The average figure of merit to 2.7 Å for 14922 reflections phased by single isomorphous replacement with the anomalous scattering of the Pt derivative was 34.42%.

Figure 5.28 shows a plot of r.m.s. F_{HCALC}, $F_{H"CALC}$, $\langle \varepsilon^2 \rangle^{\frac{1}{2}}$, $\langle \varepsilon"^2 \rangle^{\frac{1}{2}}$, m and F_P v. $\sin^2\theta/\lambda^2$. The $\langle \alpha_P - \alpha_H \rangle = 85.7^\circ$.

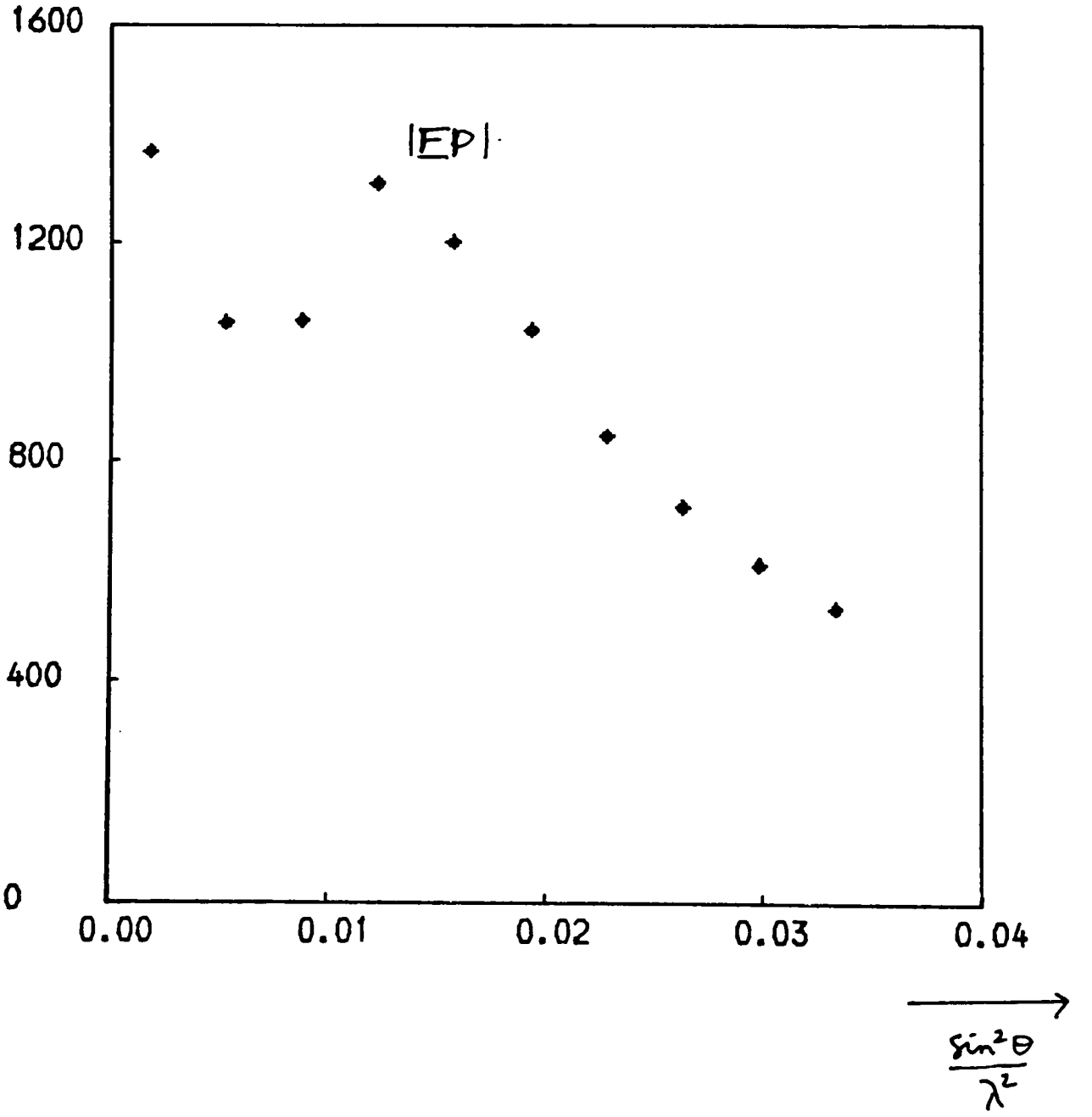
The figure of merit for the centric reflections can be seen to be higher on average than for the acentrics. The quality of phasing on the basis of the anomalous differences is poor. Figure 5.29 is a histogram of the number of reflections for a given 10° interval of $(\alpha_P - \alpha_H)$. This shows a substantial number of acentric reflections with $(\alpha_P - \alpha_H) = 0$ or π. Single isomorphous phases without anomalous scattering would give phases, α_P, which are always equal to or opposite to α_H. The histogram shows that the anomalous differences do not resolve the phase ambiguity for a substantial number of reflections.

A third phase set was calculated on the opposite hand (i.e. x y z of the heavy atoms changed to $\bar{x}\ \bar{y}\ \bar{z}$) which is equivalent to changing the sign of the anomalous differences. Two difference Fourier maps were calculated to 6 Å using the Au derivative isomorphous differences (diffractometer data) using the "photographic" phases on the different hands. Figure 5.30 shows the two

(a)


200
150
100
50
0
FHCALC
<ε''²>^1/2
<ε²>^1/2
<ε²>^1/2
<ε''²>^1/2
FHCALC
0.00
0.01
0.02
0.03
0.04
sin²θ/λ²

(b)


1600
1200
800
400
0
|EP|
0.00
0.01
0.02
0.03
0.04
sin²θ/λ²

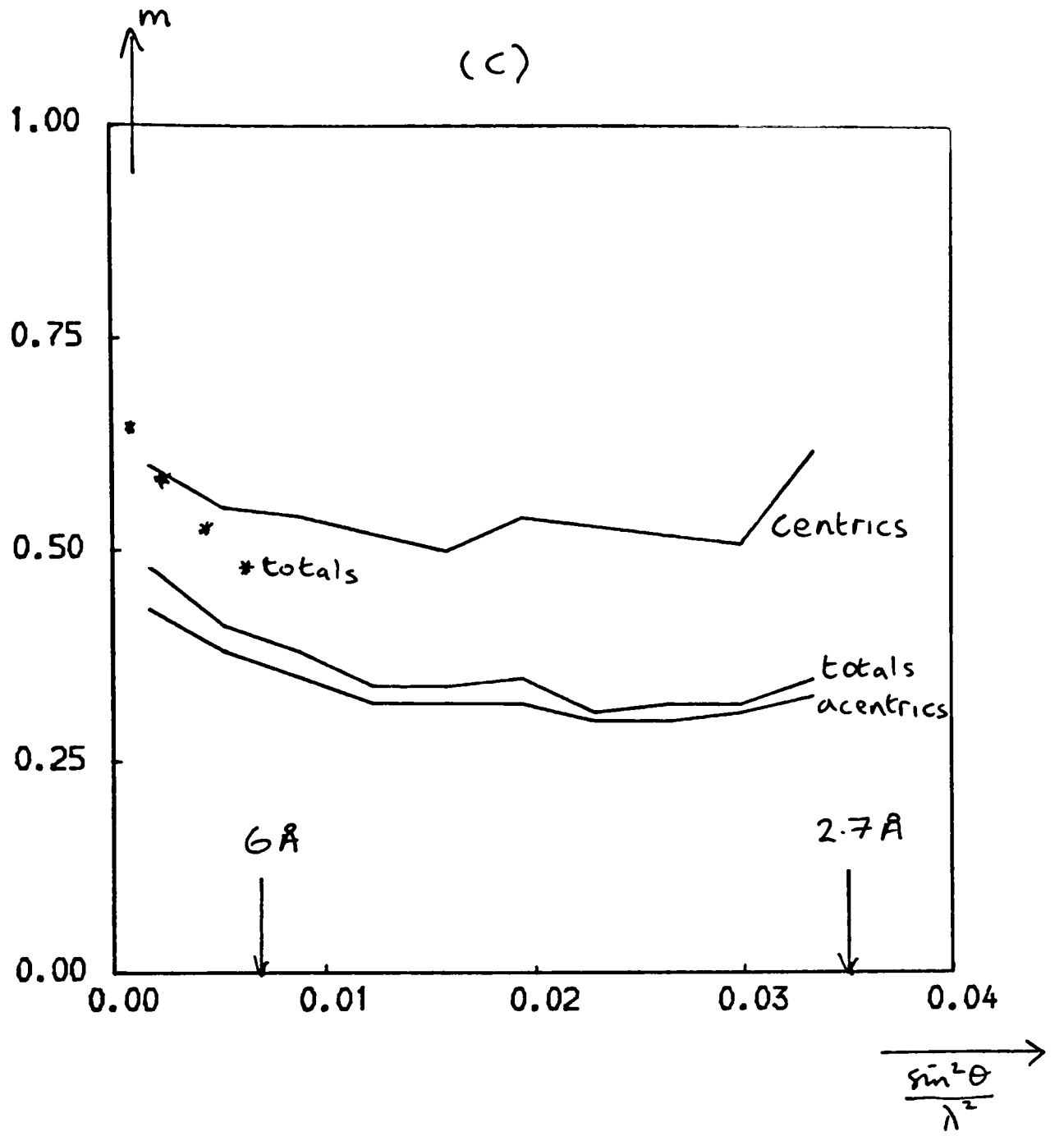


Figure 5.28 Details from the phase calculation. Variation of (a) $F_{HCALC}, \langle \varepsilon^2 \rangle^{\frac{1}{2}}$, $F_{H''CALC}, \langle \varepsilon''^2 \rangle^{\frac{1}{2}}$; (b) $|\underline{F}_p|$; and (c) figure of merit v. $\sin^2\theta/\lambda^2$. In (c) the curves are for the phase calculation using the photographic data only and the plotted points (*) are the values of the average figure of merit for the double derivative phases $(Au+Pt)_2$ which are combined with the photographic phases immediately prior to calculating the final electron density map.

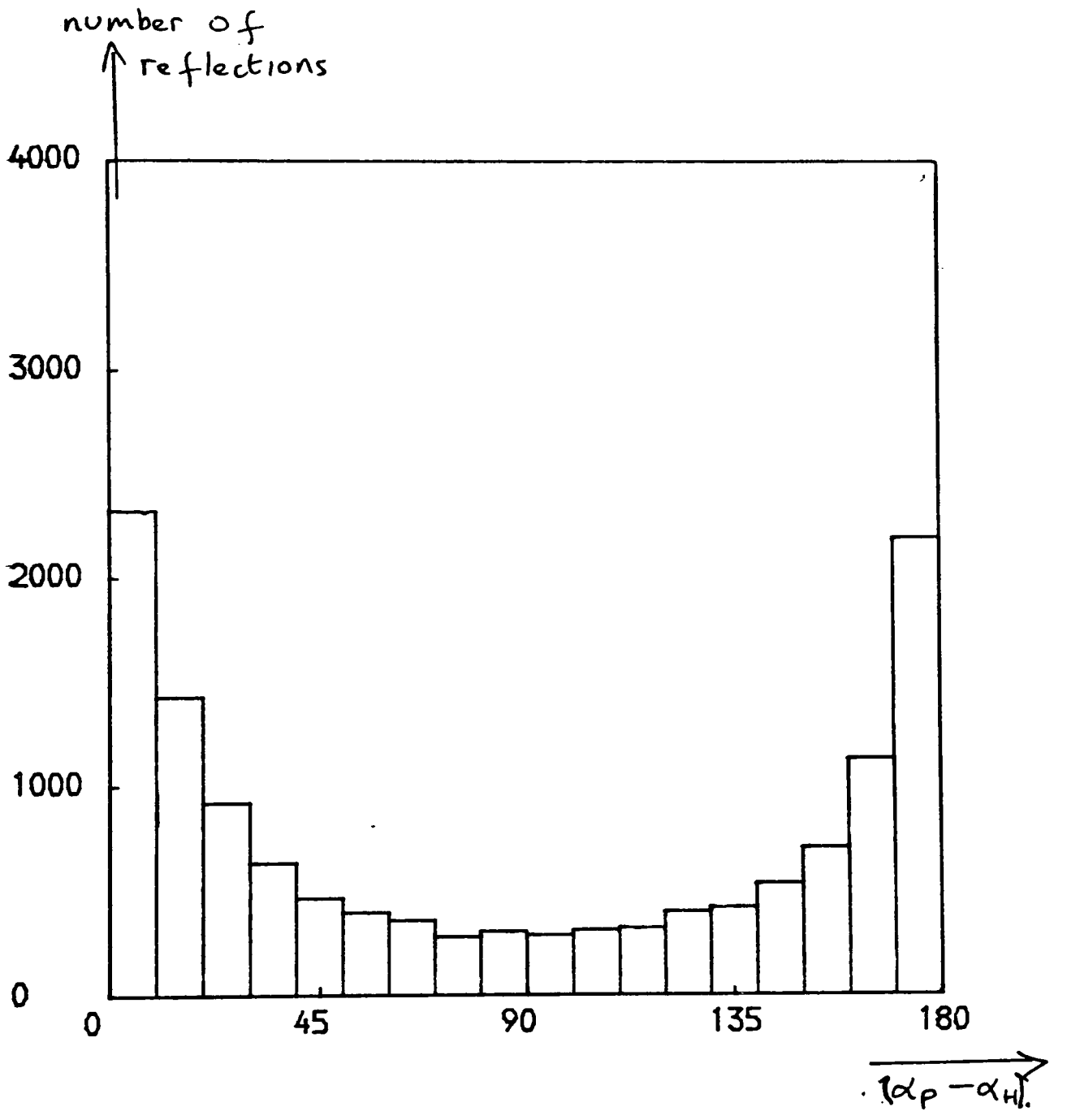


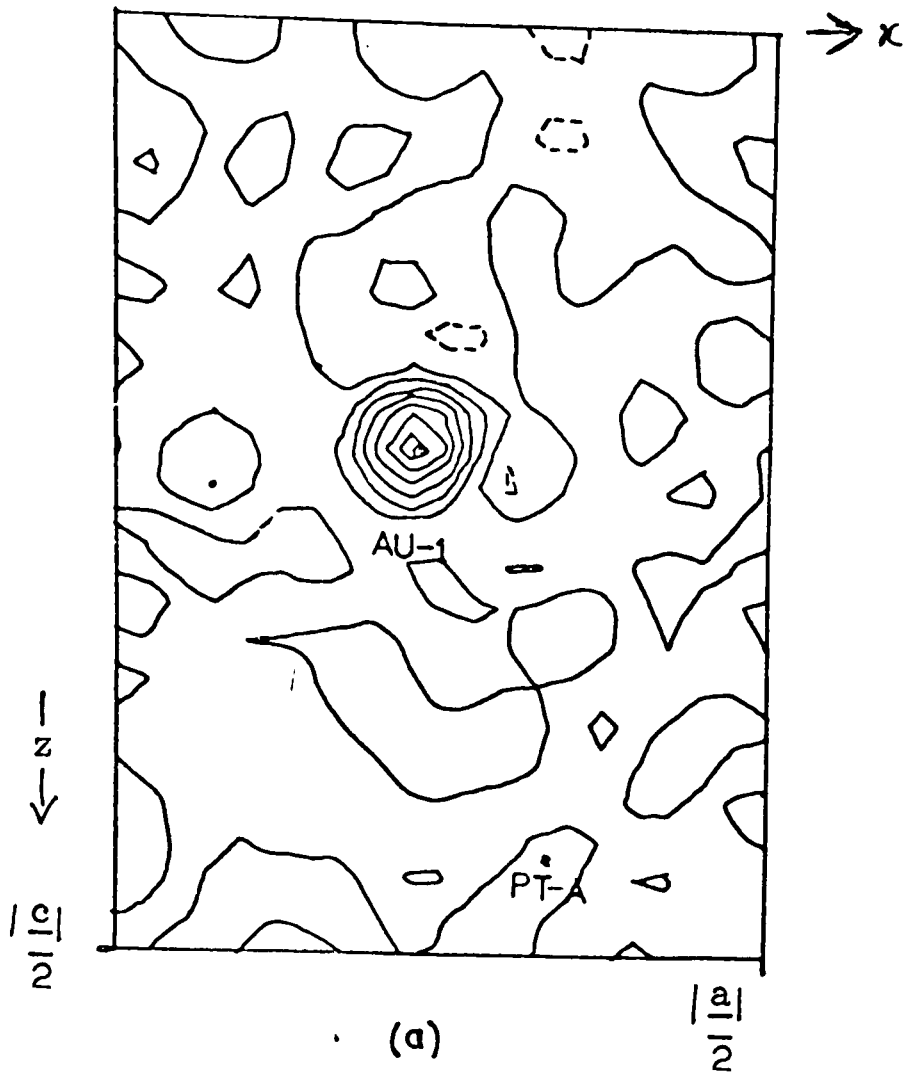
Figure 5.29 Histogram showing the correlation of α_P and α_H.

Table 5.14 Peak heights on the difference Fourier map with coefficients (FP_{Au} - FP) calculated using the "photographic" phases on the right hand (α_{PTRH}) and on the wrong hand (α_{PTWH}). The Au isomorphous differences were calculated using diffractometer data (compare with Table 3.17).

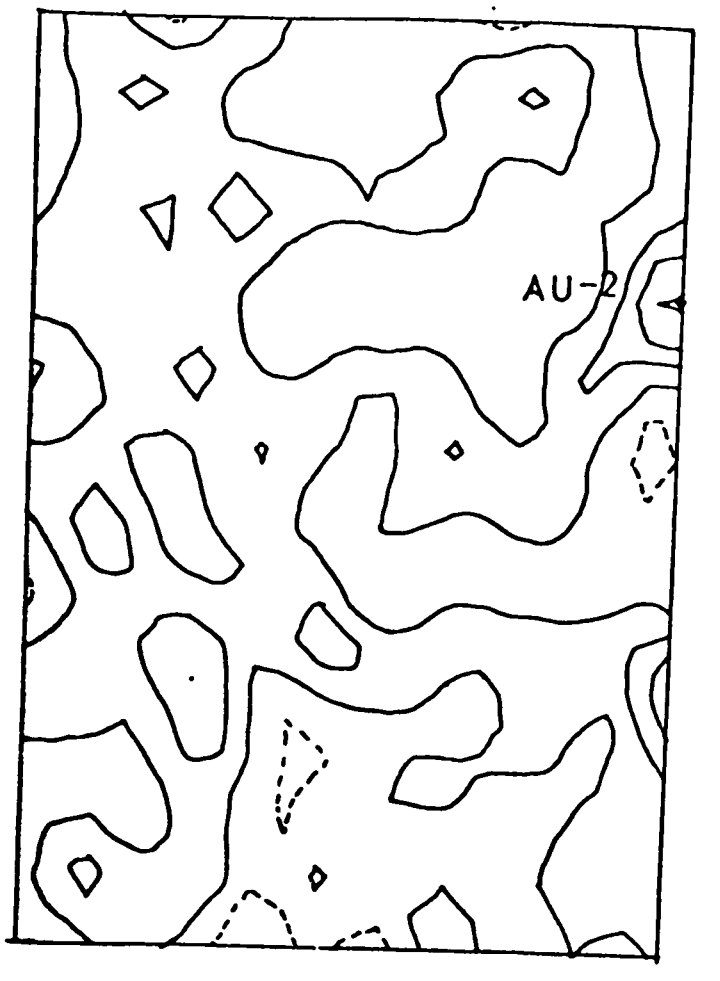
	α_{PTRH}				α_{PTWH}			
	Sectional coordinates[a]				Sectional coordinates[a]			
Peak	x	y	z	peak height	x	y	z	peak height
Au1	9	6	11	34	9	6	13	26
Au2	20	16	7	16	20	16	7	6
<bg>				9				9

a in units of 1/40, 1/72, 1/48 in x, y, z, respectively.

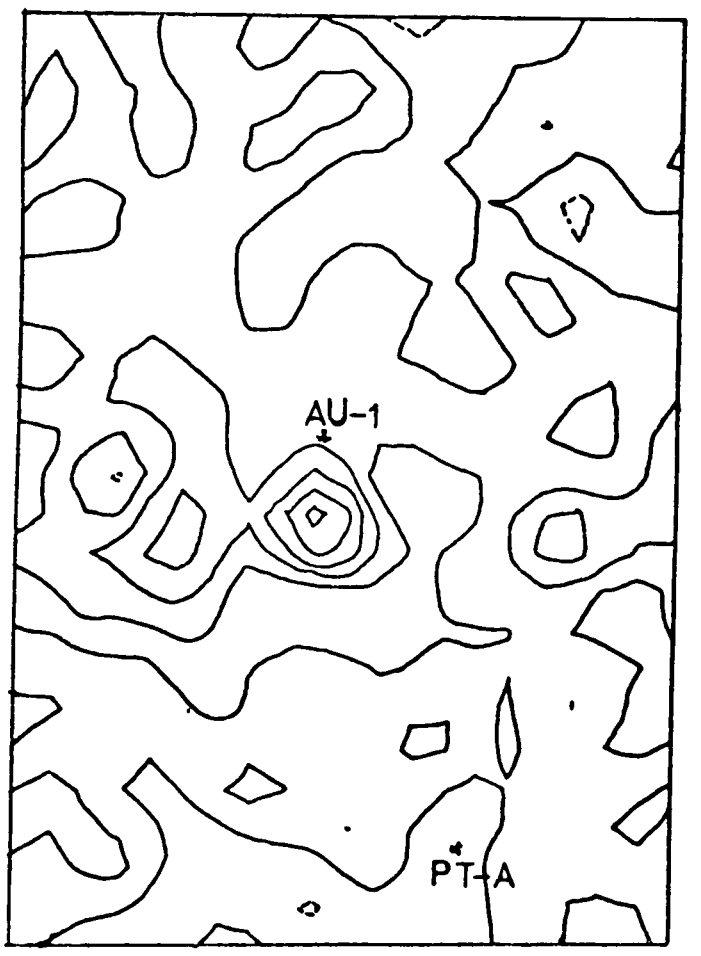
Note Tables 5.8 and 5.14 are for difference maps on the same scale as the heavy atom difference maps shown in § 3, details of which are tabulated for instance in Table 3.17.



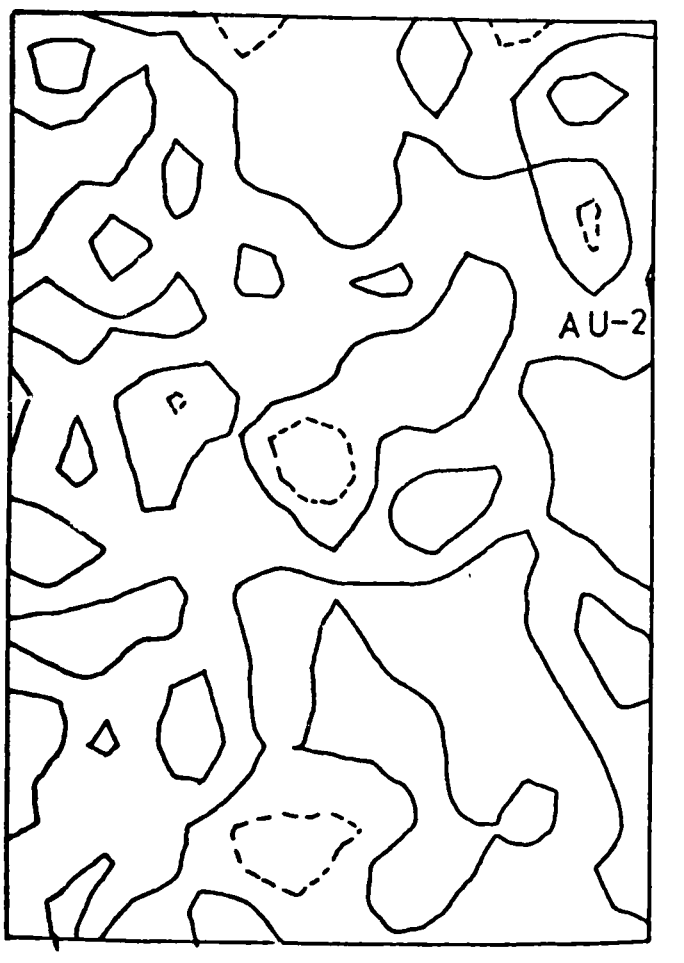
(a)



(b)



(c)



(d)

Figure 5.30 Selected sections of the difference Fourier map with coefficients (FP_{Au}-FP) calculated from diffractometer data and two different sets of phases: (a) and (b) single derivative Pt phases with anomalous dispersion on the correct hand; (c) and (d) single derivative Pt phases with the sign of the anomalous reversed. (a) and (c) y = 6/72 b, (b) and (d) y = 16/72 b. The major (Au1) and minor (Au2) gold sites are marked.

---- negative

—— positive

contours start at 0 and the contour interval = ± 5

sections containing the Au major and minor site for each map. On the correct hand Au1 and Au2 give peaks at 3.8 x and 1.6 x background respectively (cf. Table 3.17 and 5.14). On the wrong hand Au1 is shifted about 6 Å from its correct position and is 2.9 x background and Au2 is not visible. The photographic and diffractometer derived difference maps (Figures 3.14, 3.15 and 5.30) are not strictly comparable. For the diffractometer single derivative phase calculation E" was set to 1/3 E causing the average figure of merit for the acentric and centric data to be comparable. For instance the single derivative diffractometer Pt phase set calculated with E" = 1/3 E has a ratio of $\langle m \rangle_{centric} : \langle m \rangle_{acentric} = 1.24$. For the single derivative photographic Pt phase set calculated with appropriate E" (see Figure 5.28) where E" > 1/3 E $\langle m \rangle_{centric} : \langle m \rangle_{acentric} = 1.47$, to 6 Å. In the latter case, therefore, centric terms dominate more in an m weighted difference Fourier map. The centric reflections are unaffected by a change in sign of the anomalous and hence the differences between the maps on the right and wrong hands shown in Figure 5.30 are not as marked as they were in Figure 3.15.

Before calculating a native electron density map the phase set could be improved by using the low resolution double derivative phase set $(Au+Pt)_2$ (described in Table 3.24 for instance). Also, any reflections missing in the photographic data set, to 6 Å, could be included by using the suitably scaled low resolution data. These missing reflections occur because they are simply not integrated satisfactorily in processing (background failures or simply too intense even for a B film) or they lie in the blind region.

So, to 6 Å the photographic $|\underline{F}_P|$'s were combined with the $(Au+Pt)_2$ phases. Only when m for an $(Au+Pt)_2$ phase was zero was the single derivative photographic phase kept. Missing reflections had a $|\underline{F}_P|$ included from the diffractometer profile fitted data as well as the $(Au+Pt)_2$ phase. Figure 5.28 shows a plot of m v. $\sin^2\theta/\lambda^2$ for the final phase set to 6 Å (beyond 6 Å the phases are from the single derivative photographic phasing calculation) and shows how the figure of merit changes discontinuously at 6 Å.

Table 5.15 Comparison of diffractometer and photographic phase sets to 6 Å

Phase set		$(Au+Pt)_2$[a]	Pt[b]
	<m>	0.55	0.43
$(Au+Pt)_2$	acentric[c]		54.7°
	centric[d]		67.0%

[a] phase set $(Au+Pt)_2$ is identical to that quoted in Table 3.24 and is the final double derivative diffractometer phase set.

[b] Pt is the final single derivative photographic phase set (before merging with $(Au+Pt)_2$) calculated to 2.7 Å using the parameters shown in Table 5.13 .

[c] 'Acentric' refers to the mean difference in phase angle between the acentric reflections of the phase sets compared.

[d] 'Centric' refers to the percent of centric reflections with the same sign in the phase sets.

Since these phases to 6 Å are more accurate than those beyond 6 Å it seems fair to allow the low resolution data to be weighted up in the native electron density map. Table 5.15 compares the 6 Å diffractometer double derivative phase set $(Au+Pt)_2$ and the single derivative photographic phases to 6 Å.

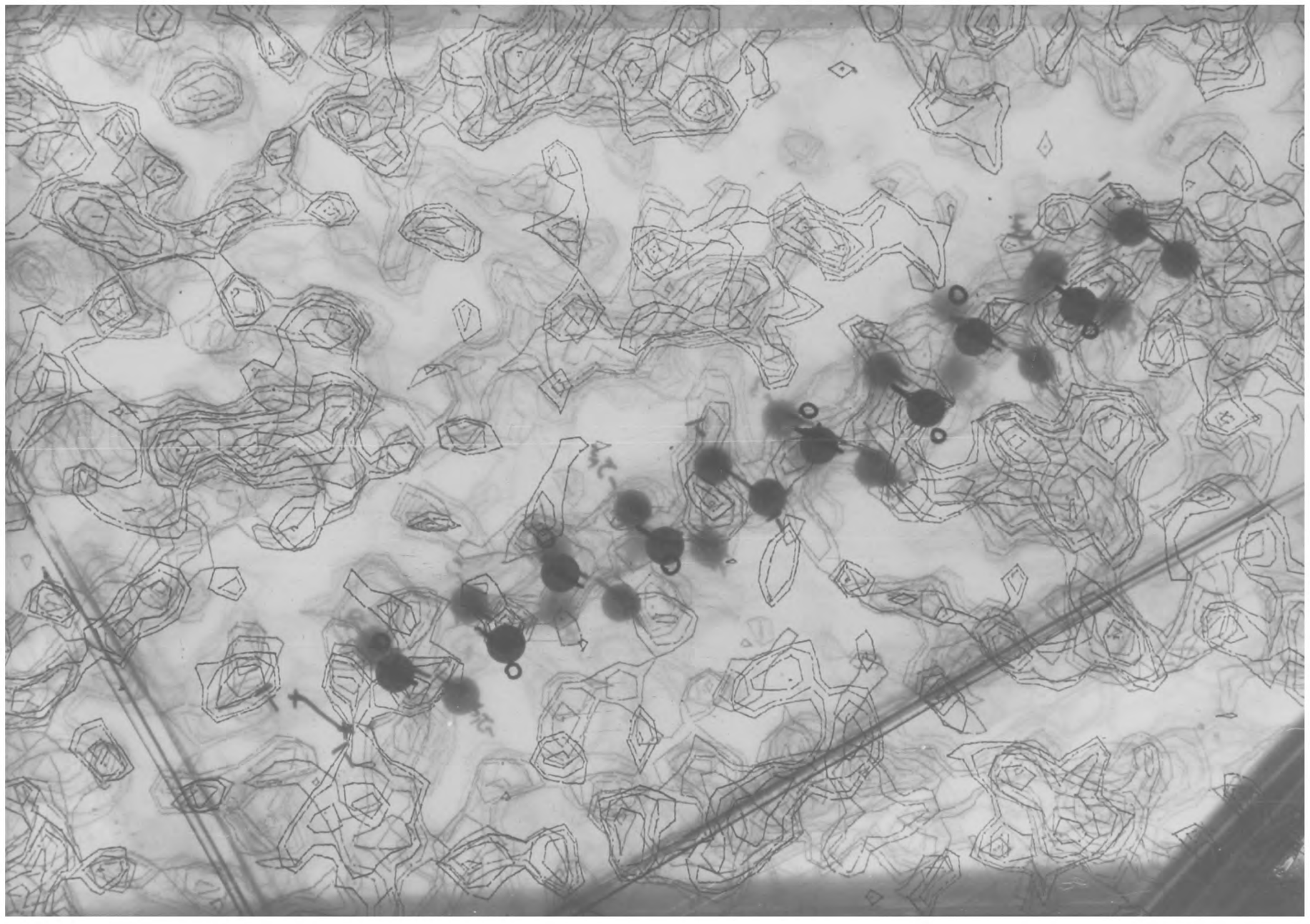
5.7 The electron density map to a resolution of 2.7 Å

The final phase set, comprising the double derivative phases to 6 Å and the single derivative phases from 6 to 2.7 Å, was used to calculate a native electron density map in intervals 1/80, 1/144 and 1/96 along x, y, z, respectively. The map was figure of merit weighted and contoured on a scale of 4 Å/cm for easy comparison with the low resolution map transparancies. The two maps were therefore identical in scale and orientation (X sections). The only difference between the maps being that there were twice as many sections for the 2.7 Å map compared with the 6 Å map. The contour level and interval were chosen empirically to yield sensible continuity and columns of density of suitable cross section.

The electron density map calculated used a V" = 365.0 (see equation 3.2) and was contoured at 125, 250, 500, 750, 1000, 1250, 1500 and 1750 units. Assuming that the data are on an absolute scale (consistent with the native Wilson plot described earlier) and using equation 3.2: the 125 level corresponds to a $\rho(\underline{r}) = 0.22$ e Å^{-3} and the 500 level to a $\rho(\underline{r}) = 0.35$ e Å^{-3}. The r.m.s. error in the electron density, $\langle \Delta\rho^2 \rangle^{\frac{1}{2}}$ calculated using equation 3.1 was 0.19 e Å^{-3}. The contour level at 125 is thus approximately the error level and one would hope to trace the polypeptide chain at the second (or possibly the third) contour level.

There is a lot of detail in the map. For instance, features of the correct dimension for large aromatic side chains extend from the main chain. However, it is too soon to have interpreted the whole map in terms of a polypeptide chain of about 400 residues. The relatively high noise level means the chain cannot be traced unambiguously by rapid casual inspection. A feel for the quality of this map can be obtained by looking

(a)



Z

Y

Figure 5.31 Stacked sections of the 2.7 Å electron density map illustrating the long helix in the yz plane. (a) X = 23/80 → 29/80; and (b) X = 19/80 → 24/80. The marker discs represent C_α positions for an α-helix. The numbering is an internal section numbering system starting at 0 (X = 28/80). In (b) at the C terminal end of the helix the chain turns strongly into another extended column of density. Compare with X = 13/40, Figure 3.20.



at one of the long helices seen at low resolution. The one in the plane of the sections was chosen since it facilitated the fitting of a standard set of C_α positions of an α-helix. Figure 5.31 shows stacked sections of this region. The α-helical conformation extends over 34 residues and several side chains, by pointing in the same direction away from the axis of the helix, indicates the N to C terminal direction of the chain. At the C-terminal end the helix turns sharply into another extended column of density which runs at an angle of 20° to the one fitted. The marker discs show the C_α atoms of the helix. The actual α-helical conformation extends some 51 Å which may be compared with the 60 Å estimate from the 6 Å electron density map.

So far no detailed interpretation of the coenzyme binding region has been attempted. The map is confused in this region by some disturbance around the major heavy atom site PTA and a great deal of care will be required in this area.

Further work on the map is in progress.

5.8 Conclusion

The work described in this chapter (and § 3) has brought the structure determination to the exciting stage of interpreting a high resolution electron density map. Some preliminary results fitting an α-helix to an especially long piece of connected density have been given. This illustrates that a good deal of detail does exist in this high resolution map. Careful comparison of the results obtained in § 4 and this map should provide information regarding the coenzyme-enzyme interaction. This was shown to be of particular interest in § 1 for comparison with NAD^+ binding to the dehydrogenases and $NADP^+$ binding to *E.coli* DHFR.

Further work on the enzyme lies in improving the quality of the electron density map (by including the Au derivative), studying the coenzyme and substrate binding to high resolution and understanding the Type A to Type B transition (described in § 3.2). This work will be aided a great deal when the amino acid sequence has been determined.

Improvements in the techniques of protein crystallography, for instance using synchrotron radiation to optimise heavy atom anomalous scattering described in § 2 and Appendix 1, will lead to an improvement in the quality of electron density maps and the rate at which these are determined.

APPENDIX I

APPLICATION OF SYNCHROTRON RADIATION TO PROTEIN CRYSTALLOGRAPHY: PRELIMINARY EXPERIMENTS ON 6PGDH CRYSTALS USING THE NINA SYNCHROTRON, DARESBURY, U.K.

In § 2 the important application of synchrotron radiation to protein crystallography in optimising the anomalous scattering of heavy atoms was discussed. The work of Phillips et al. (1976) involved crystals, of various sizes, of rubredoxin, azurin, nerve growth factor and L-glutaminase asparaginase to test the beam from the storage ring SPEAR (Stanford, U.S.A.) as a possible X-ray source in crystal structure analysis. The data obtained at several different wavelengths above and below the Fe K edge (1.74 Å) from the iron containing protein, rubredoxin (M.W. = 6000) were analysed in detail (Phillips et al. (1977)).

The work described here involved similar experiments directed at optimising the anomalous scattering of the heavy atoms in the $KAu(CN)_2$ and $K_2Pt(CN)_4$ derivatives of 6PGDH, respectively. The values of f' and f" at the wavelengths of the Ag, Mo, Cu, Fe and Cr Kα lines are shown in Table AI.1. In Table AI.2, the values of the wavelengths for the L and K absorption edges for Au and Pt are given. The aim of this work was to test the usefulness of the synchrotron as an X-ray source in the structural analysis of a large protein molecule (6PGDH has a molecular weight of 100,000 Daltons).

One 24 hour shift was available (December 1976) in one of the NINA machine cycles; the closure of the source in March 1977 prevented continuation of the experiments.

The experiments proposed were the recording of precession photographs on a Stoe precession camera of:

(1) the h 0 l centric zone of the Au and Pt derivatives of 6PGDH respectively on the long wavelength side of the appropriate absorption edge;

Table AI.1 Variation of f' and f" of Au and Pt at 5 different wavelengths. (From International Tables for X-ray Crystallography.)

	CrKα		FeKα		CuKα		MoKα		AgKα	
Element	f'	f"	f'	f"	f'	f"	f'	f"	f'	f"
Pt	-4.6	12.8	-4.5	9.9	-5.2	6.9	-2.4	8.4	-1.3	5.7
Au	-4.7	13.5	-4.5	10.4	-5.1	7.3	-2.7	8.8	-1.4	6.0

The wavelengths of the $K\alpha_1$ emission lines for Cr, Fe, Cu, Mo, Ag, are respectively: 2.29, 1.94, 1.54, 0.71, 0.56 Å.

Table AI.2 Wavelengths of the K and L absorption edges of Au and Pt in Å. (From International Tables for X-ray Crystallography.)

Element	K	L
Pt	0.15816	1.07239
Au	0.15344	1.03994

(ii) the h h l acentric zone of both derivatives above and below the edge.

Photographs of type (i) were aimed at observing the variation of f' with λ and of type (ii) maximizing f".

The apparatus was located 47 m from the electron orbit and the machine parameters at the time of the experiment were 4GeV, 18 mA (average), 25 mA (peak). The beam was monochromated using a Si [1 1 1] channel cut computer controlled monochromator with a band pass $\delta\lambda/\lambda = 10^{-4}$ and which then entered a pin hole collimator on the Stoe camera. The camera had been previously roughly aligned with a laser beam reflected onto beam axis by a plane mirror. Final adjustment was achieved by observing the spot shape of the main beam on polaroid film. Themain advantage of the Stoe camera was the optical system for centering the crystal specimen involving a telescope, whose axis was perpendicular to the collimator axis.

All the crystal specimens were mounted in glass capillaries in Oxford and the diffraction tested on a GX 6 rotating anode run at 40 kV, 40 mA. The specimens were then transported by car to Daresbury. No facilities existed at NINA for mounting crystal specimens.

The positions of the L absorption edges of the Pt and Au atoms were located in the azimuthal coordinate space of the monochromator stepping motor by monitoring the transmitted intensity of the beam through the appropriate foil with an Argon ionisation chamber.

No useable precession photographs of the crystals were obtained on this run. In the absence of focussing of the beam in the horizontal and vertical planes the crystal intercepts only a small solid angle of the available beam intensity. Since the divergence of the beam in the vertical plane is 0.1 mrad the vertical dimension of the beam 47 m from the source is 4.7 mm. A crystal that presents a face of size 0.2 x 0.2 mm^2 to the beam intercepts only a small fraction of the available intensity. A mirror/ monochromator system in a working protein crystal diffraction experiment

has been described subsequently by Webb et al. (1977).

The X-ray diffraction of the crystals was re-tested on return to Oxford and found to be weak, though present. Transport of the samples in this manner was not ideal.

APPENDIX II

Adams, Helliwell, Bugg (1977) J. Mol. Biol. 112, 183-197.

Redacted journal article "Structure of 6-phosphogluconate dehydrogenase from sheep liver at 6 Å resolution" can be found at http://dx.doi.org/10.1016/S0022-2836(77)80138-1

APPENDIX III

Adams, Archibald, Helliwell (1977) 2nd Int.Symp. of Pyridine Nucleotide Dependent Dehydrogenases, Konstanz, April 1977.

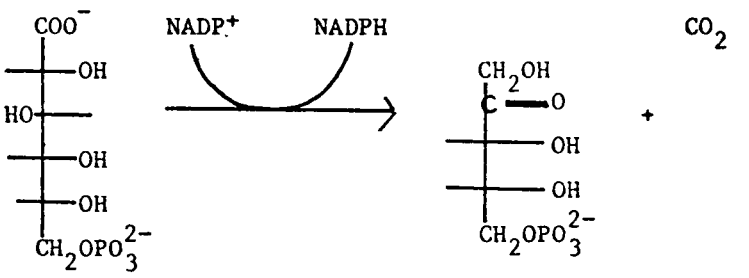
PREPRINT
2nd International
Symposium on
Pyridine Nucleotide-Dependent
Dehydrogenases
KONSTANZ (GERMANY)
March 27 – April 1, 1977

X-ray Diffraction Studies on Sheep Liver 6-Phosphogluconate Dehydrogenase at 6Å Resolution

M.J. Adams, I.G. Archibald, J.R. Helliwell.
Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford, U.K.

INTRODUCTION

The enzyme 6-phosphogluconate dehydrogenase (6-PGDH), is an $NADP^+$ dependent dimer of molecular weight 94,000 Daltons. (Silverberg and Dalziel, 1973). It is a member of the pentose phosphate pathway and acts as an oxidative decarboxylase yielding ribulose-5-phosphate and NADPH.



The two subunits of the dimer act independently and there is no specific requirement for Mg^{2+} ions. (Villet and Dalziel, 1972). The mechanism of the oxidation has been studied under various conditions with different results. A random order mechanism has been suggested in buffers such as tris-acetate (Dyson, D'Orazio and Hanson, 1973) while an ordered binding of co-enzyme followed by substrate has been indicated in phosphate buffer at pH 7. (Silverberg, 1973). Inhibition by various nucleotide phosphates, co-enzyme fragments and some other metabolic intermediates has been studied. (Dyson and D'Orazio, 1973). The enzyme has been crystallised from ammonium sulphate and crystals were first reported by Silverberg, Dalziel and Adams (1973).

THE STRUCTURE AT 6Å

The structure determination by the method of isomorphous replacement has been described (Adams, Helliwell and Bugg, 1977) and will only be reviewed briefly here. The enzyme crystallises in the spacegroup $C222_1$ with $\underline{a}$ = 72.7Å, b = 148.2Å, $\underline{c}$ = 102.9Å with a single subunit per

asymmetric unit. Soaking crystals in solutions of salts containing heavy atoms yielded two suitable heavy atom derivatives. The most satisfactory conditions for substitution were 22μM potassium dicyanoaurate(I),$KAu(CN)_2$, for 22 hours and 2.7mM potassium tetracyanoplatinate (II), $K_2Pt(CN)_4$, for 24 hours. Heavy atom positions were found by use of both Patterson maps and direct methods. For each derivative in turn, measurements of anomalous dispersion were combined with isomorphous differences to resolve the phase ambiguity. The resultant phase sets, each calculated using only one derivative were combined with the isomorphous differences of the second derivative and used to calculate difference Fouriers. The platinum positions were the largest peaks on the platinum difference Fourier using gold phases as were the gold sites on the gold difference Fourier with platinum phases. One of the two ways of combining isomorphous and anomalous differences gave a noticeably superior difference map for each derivative, thus enabling the 'hand' to be obtained readily. Figure I shows one section from each map which includes the major site of each derivative.

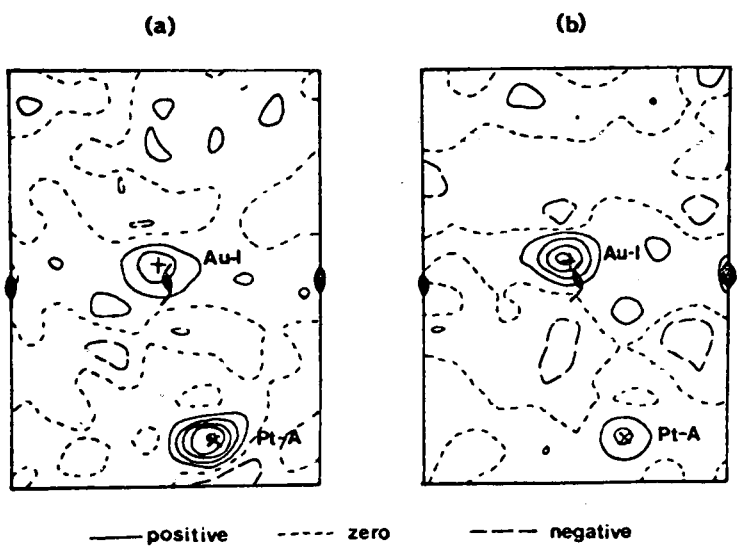


Fig. I

Comparison of section y = 6/72 b of difference Fouriers.
(a) Platinum difference Fourier with gold phases showing site Pt(A) and ghost of Au(1).
(b) Gold difference Fourier with platinum phases showing site Au(1) and ghost of Pt(A).

The heavy atom parameters were refined by the method of Least Squares using the centric data only, some 23% of the terms. The function $M = \sum\omega(|F_{HOBS}|-|F_{HCALC}|)^2$ was minimised where ω is an experimental weighting factor and $|F_{HOBS}|$ and $|F_{HCALC}|$ are the 'observed' and calculated structure amplitudes for the heavy atom components alone. The final heavy atom parameters are given in Table I and were used to calculate phases for the native structure factors. A 'best' electron density map (Blow and Crick, 1959) with terms $m|F_p|\exp(i\alpha_{best})$ was calculated.

The electron density map shows the dimer to be centered at approximately $0,\frac{3}{4},\frac{1}{4}$ with the dimer axis the crystallographic 2-fold axis parallel to b. The dimer was readily isolated and the individual subunits could be distinguished with ease. Two views of a balsa wood model of the subunit

TABLE I FINAL PARAMETERS USED FOR PHASING

				Positional Parameters				r.m.s.E*		r.m.s.
	B_{ov}**	Site	Z+	x	y	z	B_i^x	centric	all	F_H
$K_2Pt(CN)_4$	-3.5	A	54	0.3330	0.0875	0.4453	15	277	200	314
		B	41	0.0301	0.1153	0.2368	15			
		C	48	0.276	0.2065	0.5036	15			
$KAu(CN)_2$	-3.5	1	78	0.2293	-0.0792	0.2291	15	241	174	328
		2	39	0.4969	0.2228	0.1428	15			

+Z is given in electrons, assuming full occupancy of site 1 of the gold derivative.

xB_i, the individual heavy atom temperature factors, were not refined.

*r.m.s. E is the lack of closure error, $|F_{PH}| - |F_H+F_P|$. $|F_{PH}|$ is the heavy atom derivative structure amplitude, $|F_P|$ the protein structure amplitude, F_H the calculated heavy atom component.

**B_{ov} is the overall temperature factor to be applied to the derivative data for best fit with the native data.

are shown in Figure II. The protein subunit is ellipsoidal with a large cleft on one surface distant from the 2-fold axis. Although the subunit is not bilobal, the cleft divides it into two unequal parts. The major site of platinum substitution is on one edge of the cleft on the smaller part of the subunit, and the minor gold site is not far distant from the other edge on the larger part of the subunit. The other most interesting feature of the subunit is that a large part of it is made up

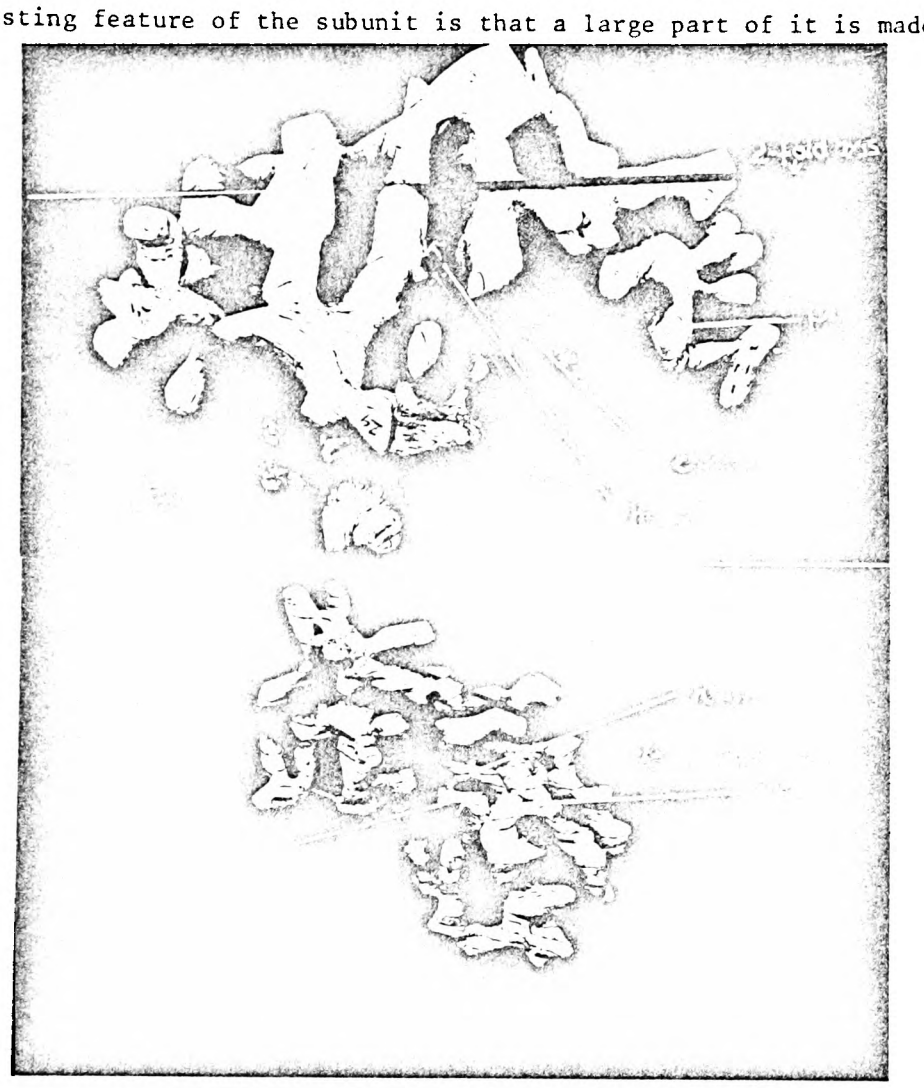


Fig. II

Views of the balsa wood model. The molecular 2-fold axis is indicated by the metal rod.
(a) Viewed from above, looking down (a).
(b) View showing outside of the dimer.
The minor gold site, the major platinum site, the cleft and a part of one long helix are arrowed.

of columns of density of the right dimensions for α-helix. Many of these columns run for up to 20Å but two are more than 60Å long, almost the length of the subunit, and form its backbone. They underlie the cleft region and one approaches the subunit interface.

BINDING OF THE CO-ENZYME $NADP^+$ AND THE ANALOGUE 3-IODOPYRIDINE ADENINE DINUCLEOTIDE PHOSPHATE TO CRYSTALS OF 6-PGDH

Attempts have been made to find suitable conditions for binding of $NADP^+$ to crystals of 6-PGDH. Experiments were carried out in the buffer used for growing crystals, 65% ammonium sulphate, 50mM potassium phosphate at pH 6.5. In these conditions, low concentrations of $NADP^+$ (5μM) had no effect on the intensity distribution. Use of 5mM $NADP^+$ gave small intensity changes which were maximal after 3 weeks soaking. These differences were reproducible and some of the intensity changes were similar to those for the platinum derivative.

A sample of 3-iodopyridine adenine dinucleotide phosphate ($IADP^+$) was kindly prepared for us by Dr. J-F Biellmann and his colleagues. Enzyme crystals, transferred to solutions containing 5mM $IADP^+$ showed reproducible intensity changes which were maximal after 24 hours. These changes were similar to those given by $NADP^+$ after 3 weeks. With $IADP^+$, longer soaking did not enhance the differences.

Diffractometer data sets to 6Å resolution were collected for both $NADP^+$ and $IADP^+$ soaked crystals. Data were processed by a technique which analysed the peak profile and is particularly suitable for small crystals with large numbers of weak reflections. 3000 measurements, 1500 independent reflections were collected from a single crystal for each derivative. The data were corrected for Lorentz-polarisation factors, absorption and radiation damage in the normal way.

Difference electron density maps were computed for the two series of differences using 'best' phases. Each of the two difference maps showed a large peak which was in a similar position in each. The peak was 12Å long and one end was adjacent to the major platinum site. This peak was twice as high as any other feature on each map. There was no very large negative density. Since the peak was close to the platinum site, it was considered that it might be sensitive to errors in the phases causing

ripple around heavy atom sites. A second set of difference maps was thus calculated using phases derived only from the gold derivative making use of anomalous and isomorphous differences. Although these maps would be noisier since the phase sets were less reliable, there should be no bias at the platinum site. The major peak on each of these maps was in the same position and of the same size as before and again there was no other feature above background level. The two peaks must therefore correspond to $NADP^+$ and $IADP^+$ density. Composites of the difference maps viewed down both a and c are shown in Figure III(a) and (b). Figure IV shows a view of the balsa wood model of the protein subunit with the density corresponding to $NADP^+$ arrowed.

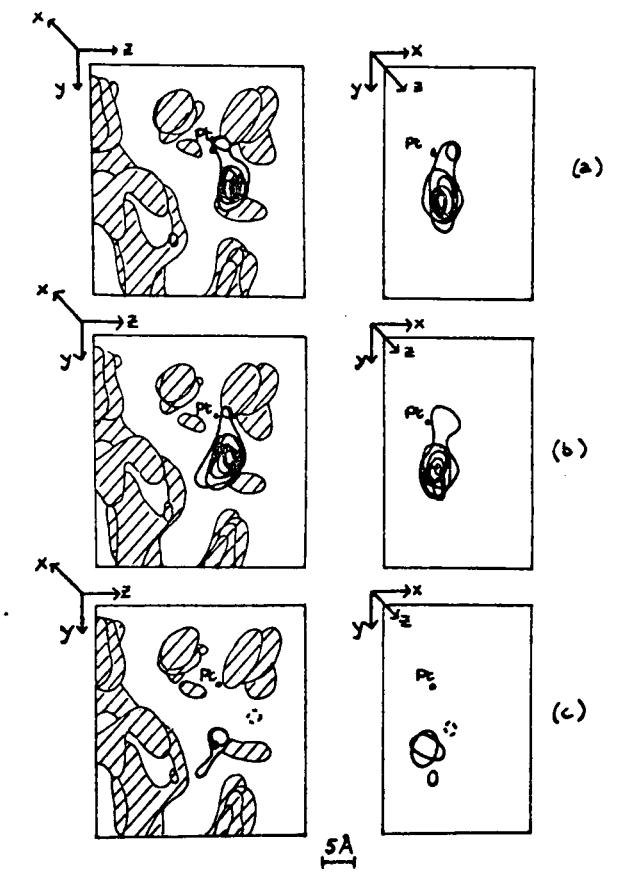


Fig. III

Composites of difference Fouriers for $NADP^+$ and $IADP^+$. Protein density (in views down a only) is shaded.
(a) $NADP^+$ - native protein.
(b) $IADP^+$ - native protein.
(c) $IADP^+$ - $NADP^+$

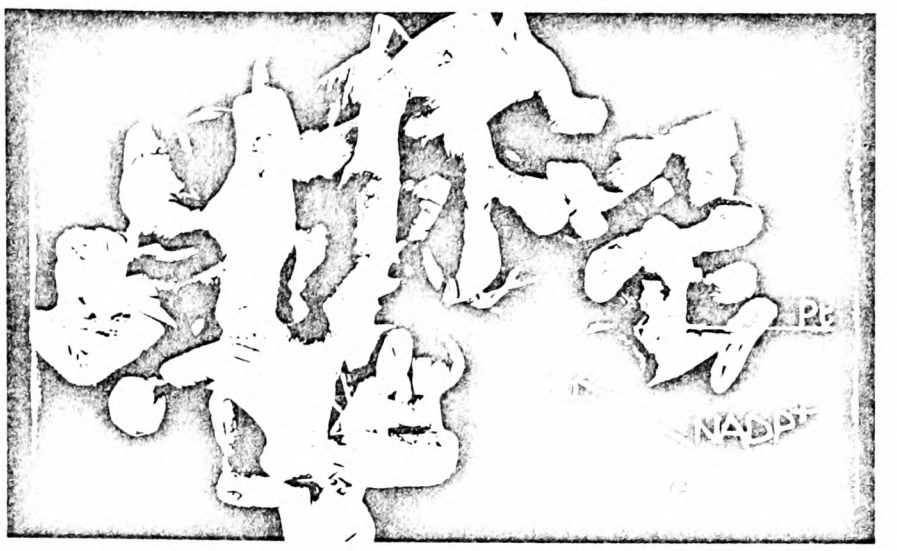


Fig. IV

View of the balsa wood model from above as in Fig. II(a) with $NADP^+$ density arrowed.

A further difference map with terms $m(F_{IADP^+} - F_{NADP^+})\exp(i\alpha_{best})$ showed a smaller positive region at one end of the co-enzyme density. This region is shown in Figure III(c). The peak, which should correspond to the difference between the iodine atom and the amide group at position 3 of the pyridine ring, is about 10Å from the site of platinum substitution and serves to orient the co-enzyme. The adenine part of the co-enzyme should thus be close to the platinum binding site. The binding of platinum compounds at the co-enzyme binding sites has been observed for lactate dehydrogenase (LDH), malate dehydrogenase (MDH) and liver alcohol dehydrogenase (LADH). In LADH the site is a general anion binding site and is at the position of the pyrophosphate. (Brändén et al. 1975).

The size and shape of the peak corresponding to $NADP^+$ indicates that while the co-enzyme certainly does not have the bases stacked, it is probably in a slightly less open conformation than that chosen by NAD^+ in LDH, MDH, LADH and glyceraldehyde phosphate dehydrogenase. The orientation of the molecule at one edge of and pointing into the cleft is interesting, it suggests an area of the subunit which might provide a substrate site, at the other side of the cleft on the larger part of the subunit. Any conclusion of this nature at this stage of the crystal structure analysis must, however, be extremely tentative.

It is even more difficult to compare the nature of the co-enzyme binding site with the 'nucleotide binding site' common to the four NAD^+ dependent dehydrogenases and maybe also to other enzymes (Rossmann et al., 1974) and the conclusions of this paragraph are entirely speculative. The smaller part of the enzyme contains four short columns of density which are within 60° of being co-axial. It is in principle difficult to see pleated sheet at 6Å resolution, and indeed, this map provides neither evidence for nor against its existence in this area. It may be argued that the elongation of the co-enzyme density should be in a rather different direction from that observed if the helices have been correctly identified as those of a 'nucleotide binding site'. The matter cannot be settled, however, without more data. A high resolution structure, in which the main chain can be traced is required before any informed decision can be made about the nature of the $NADP^+$ binding site.

EXPERIMENTS WITH DIFFERENT SUPPORTING MEDIA

The experiments described in this section arose from a chance observation that occasional crystals show a quite different intensity pattern but very similar cell dimensions and the same spacegroup. The occurrence of this form was unusual and unpredictable.

When the ammonium sulphate medium was changed to 3M potassium tartrate at pH 8 for 12 hours, crystals became very fragile, but those which could be handled showed the transformed pattern though they were less well ordered than the normal native enzyme crystals. The same transformed pattern could be produced reliably with 1.8M sodium citrate at a measured pH of 7.2, and small crystals give diffraction patterns which were well ordered. Figure V compares the intensity patterns of the normal 'sulphate' type (a) and the 'citrate' type (b) in the (hOℓ) zone. The cell dimensions of the transformed crystals changed by a maximum of 3% and are $\underline{a}$ = 71.9Å, $\underline{b}$ = 146.8Å, $\underline{c}$ = 100.6Å. The nature of the transformation is of considerable interest. The intensity changes are much too large to be accounted for without some movement of the molecule but the molecular 2-fold axis must still be exact. Unfortunately there has been rotational disorder in every crystal investigated which is large enough for collection of diffractometer data and so the nature of the change has not been further elucidated yet.

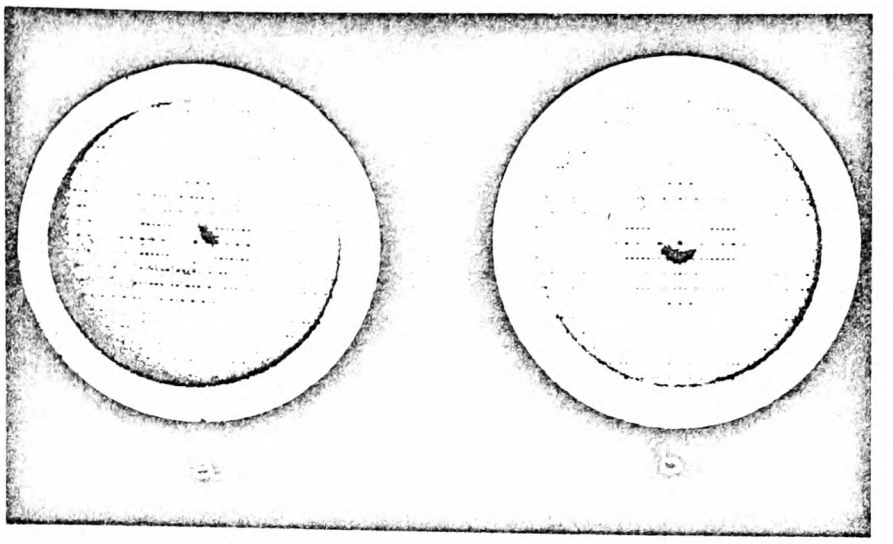


Fig. V

Comparison of hOℓ photographs of (a) native 6-PGDH in ammonium sulphate, (b) crystals transferred to citrate.

When crystals were left in a supporting medium of 3M sodium/potassium phosphate buffer for seven days at pH 6.5, much smaller intenstiy changes were observed, not suggestive of any large-scale movement of the protein. The maximum change in cell dimensions was 0.5Å with a = 73.1Å, b = 147.7Å, c = 102.9Å. Diffractometer data have been collected to 6Å for a crystal in phosphate and a difference map between the phosphate and the sulphate forms has been calculated.

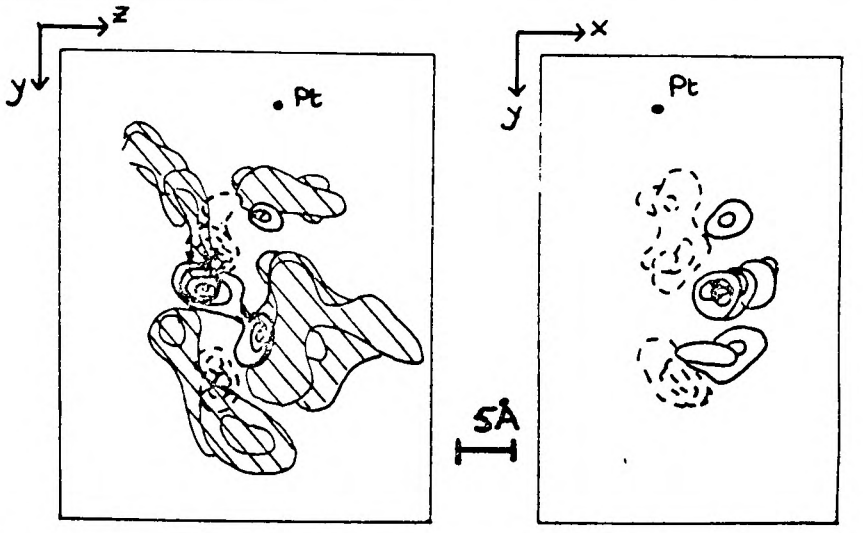


Fig. VI

Composite of difference Fouriers, phosphate crystal - sulphate crystal. Protein density (in view down a only), is shaded.

The major features of this map are all in one area 10 x 20 x 12Å^3 which is illustrated in Figure VI. There are two peaks and two holes which are up to 7Å long in their longest dimensions and a third small peak. The holes are on pieces of density at the edge of the cleft, the one closer to the major platinum site lies 15Å from it and is a little beyond the end of the co-enzyme site attributed to the nicotinamide. The second hole lies about 26Å from the platinum site. The peaks lie between the holes and may indicate movement of electron density. The further small peak is 10Å from the platinum site at a position where a part of the co-enzyme binds. The positions of the peaks and holes are indicated on a view of the balsa wood model on Figure VII.

It is extremely interesting to note that the changes take place in this part of the cleft. It would seem that either protein movement has occurred in this area or that bound sulphate ions have been removed and phosphate ions have bound in slightly different places. In view of the changed kinetics in the presence of phosphate and the fact that it is a competitive inhibitor (as in sulphate), it is tempting to suggest that some of the changes label the substrate site. The possibility seems strengthened since the peaks are not far from the nicotinamide portion of the co-enzyme. The substrate site would, of course, be best marked by

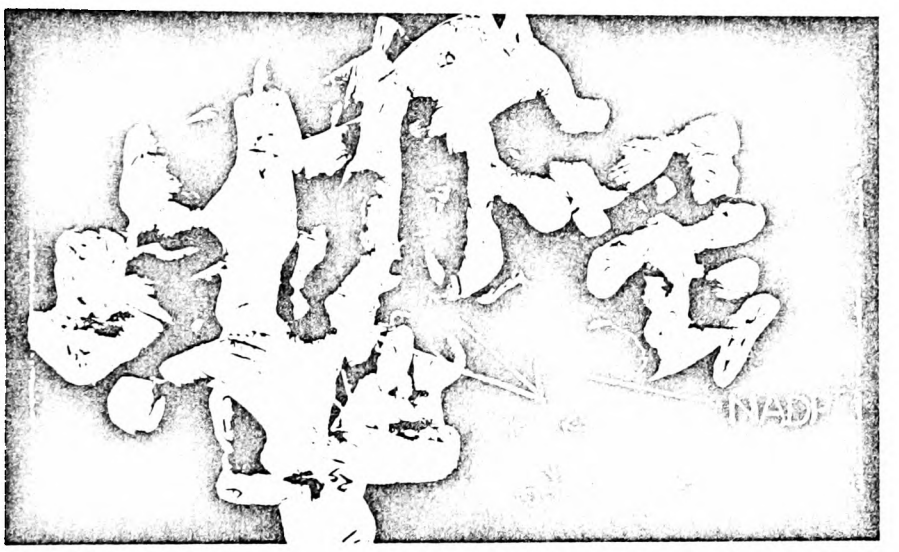


Fig. VII

View of the balsa wood model from above as in Fig. II(a) with position of phosphate-sulphate differences arrowed. The co-enzyme is in place and is indicated.

binding 6-phosphogluconate, but it has proved difficult to do this in the crystals in ammonium sulphate and in the absence of co-enzyme. Attempts are being made to bind 6-phosphogluconate in presence of NADPH and it is hoped that this will give a clearer indication of the substrate binding site. Extension of this work to higher resolution will give more definitive results.

ACKNOWLEDGEMENTS

We thank Professor D.C. Phillips and Professor Dorothy Hodgkin for their help and encouragement and Professor Don Voet for assistance and for helpful discussions. We are indebted to the Oxford Enzyme Group Preparation Laboratory for enzyme preparations. We acknowledge financial support from the Medical Research Council. One of the authors (IGA) holds an M.R.C. research scholarship, a second (JRH) a Science Research Council research scholarship. The third author (MJA) is a fellow of Somerville College, Oxford and a member of the Oxford Enzyme Group.

REFERENCES

Adams, M.J., Helliwell, J.R., Bugg, C.E.: The structure of 6-phosphogluconate dehydrogenase from sheep liver at 6Å resolution. Manuscript submitted to J. Mol. Biol. (1977).

Blow, D.M., Crick, F.H.: The treatment of errors in the isomorphous replacement method. Acta Cryst. 12, 794-801 (1959).

Brändén, C-I., Jörnvall, H., Eklund, M., Burugren, B.: The Enzymes 3rd edn. ed. P. Boyer (Academic Press) Vol. XI. Oxidation-Reduction Chap. 3 Alcohol Dehydrogenases p. 103-190 (1975).

Dyson, J.E.D., D'Orazio, R.E.: Sheep liver 6-phosphogluconate dehydrogenase - inhibition by nucleoside phosphates and by other metabolic inhibitors. J. Biol. Chem. 248, 5428-5435 (1973).

Dyson, J.E.D., D'Orazio, R.E., Hanson, W.H.: Sheep liver 6-phosphogluconate dehydrogenase, isolation procedure & effect of pH, ionic strength and metal ions on the kinetic parameters. Arch. Biochem. Biophys., 154, 623-635 (1973).

Rossmann, M.G., Moras, D., Olsen, K.W.: Chemical and biological evolution of a nucleotide binding protein. Nature 250, 194 (1974).

Silverberg, M.: Studies of 6-phosphogluconate dehydrogenase. D.Phil. Thesis, Oxford University.

Silverberg, M., Dalziel, K.: Crystalline 6-phosphogluconate dehydrogenase from sheep liver. Eur. J. Biochem., 38 229-238 (1973).

Silverberg, M., Dalziel, K., Adams, M.J.: Crystalline 6-phosphogluconate dehydrogenase from sheep liver. Biochem. Soc. Trans. 1, 1132-1134 (1973).

Villet, R.H., Dalziel, K.: Studies of 6-phosphogluconate dehydrogenase from sheep liver II Kinetics of the oxidative-decarboxylation reaction, co-enzyme binding and analyses for metals. Eur. J. Biochem. 27 251-258 (1972).

Received January 31, 1977

APPENDIX IV

COMPUTER PROGRAMS

PRECESS (Dr. K.S. Wilson): Calculation of screen settings for precession photographs.

OSCAR (Dr. D.G.R. Yeates): Primary processing of screenless oscillation photographs on the 1906A.

FILMPACK (Dr. K.S. Wilson): Merging of two film packs of oscillation or screened precession photographs.

FILMRAD (Dr. K.S. Wilson): Radiation damage analysis for photographic data.

PRIMROSE (Dr. S. French): Primary data reduction of diffractometer data (including reduction of peak profile data). Also applies Lorentz polarisation and absorption correction.

SSM (Mrs. E. Dodson and Dr. K.S. Wilson): Scales, sorts and merges sets of data from different sources. Uses Hamilton, Rollett and Sparks (1965) scaling procedure, or Fox and Holmes (1966) method.

TRUNCATE (Dr. K.S. Wilson): Truncation of reduced intensity data to all positive Wilson's distribution (1949) according to the method of Wilson and French (1977).

FRIEDANAL (Dr. K.S. Wilson): Analysis of Friedel equivalents in different ranges of reciprocal space.

CAD (Mrs. E. Dodson and Dr. K.S. Wilson): Collects data for native and derivatives and outputs to a single disc file.

ANSC (Dr. P.R. Evans): Analyses and scales derivatives to the native data set.

STAT (Dr. J. Mosley): Calculates Wilson and Crick Magdoff plots for the derivatives.

NORMALIZE, MULTAN, SOLUTION (Main, Germain and Woolfson: modified for use in Oxford by Dr. K.S. Wilson): Direct methods for automated interpretation of derivatives: normalization, phasing by tangent formula, selection of phase set.

SORTFFT (Mrs. E. Dodson): Sorts data for large Fouriers before input to FFT.

FFT (L .Ten-Eyck: modified for use in Oxford by J. Kirz, Mrs. E. Dodson and P.R. Evans): Cooley-Tukey fast Fourier program.

MAPERROR (Dr. K.S. Wilson): Calculates r.m.s. errors in Fourier synthesis weighted on figure of merit.

PEAKS (Dr. K.S. Wilson): Peak search for binary output of FFT.

PATTPKS (Dr. K.S. Wilson): Calculates self and cross vectors for heavy atom difference Patterson maps.

CONTOUR (Dr. S.H.Banyard, Dr. P.R. Evans and Dr. S.J. Oatley): Contours Fourier output.

FHLEKR (Dr. G. Bentley, Dr. P.R. Evans, Mrs. E. Dodson and Dr. K.S. Wilson): Uses procedure of Arnone et al. (1971) to scale derivative to native and outputs FHLE coefficients (Dodson and Vijayan (1971)) and FHLEB coefficients (Dodson, Evans and French (1975)).

REFINE (Busing, Martin and Levy: ORLS; from Dr. P. Main and subsequently much modified by Mrs. E. Dodson and Dr. K.S. Wilson) Full matrix least squares refinement of heavy atom parameters for the derivatives.

PHASE (Dr. H. Muirhead, Dr. G. Bentley, Dr. M.Geisow): Computes protein phases using Blow and Crick's (1959) treatment of errors.

ANALYSE (Dr. P.R. Evans): Analyses phases produced by PHASE.

PHASEANAL (Mrs. E. Dodson): Analyses agreement between any two columns

in a binary data file.

BINED (Mrs. E. Dodson): Edits a binary data file.

GRAPH (Dr. S.J.Oatley): Graphs two columns X and Y.

Copies of the programs and the input instructions are enclosed on microfiche below.



Appendix V Structure factor tables

The fiche marked

(a) "6 A DATA" contain the 6 Å data used in § 3 and § 4 and the columns of data are in the order:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

FP_1 σ FP_{Au} σ ΔAu σ FP_{PT} σ ΔPT σ FP_2 σ FP_{NADP} σ FP_{IADP} σ $\alpha(Au+Pt)_2$ m

FP_1 corresponds to the native data measured using ordinate analysis in § 3. FP_2 corresponds to the native data measured as profiles and fitted by a calculated profile according to the method of French (1975) in § 4 .

(b) "2.7 A DATA" contain the photographic native and platinum data used in § 5. Also included are FP_{2Sc}, σ, $\alpha(Au+Pt)_2$, m as well as the final combined data set (double derivative phase to 6 Å and single derivative phases from 6 Å to 2.7 Å) used to calculate the 2.7 Å electron density map. The columns of data are in the order:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

FP σ FP_{PT} σ ΔPT σ αPT m FP_{2Sc} σ $\alpha(Au+Pt)_2$ m FP_F σ α_F m

αPT are the single derivative phases calculated in § 5.

FP_{2sc} are the FP_2 (as in (a) above) scaled to the photographic native data.

FP_F and α_F refer to the final combined data set (see § 5).

REFERENCES

Abbé Archiv.F.Mikroskopische Anat. (1887) 9, 413.

Adams M.J., Archibald I.G. and Helliwell J.R. (1977) 2nd International Symposium of pyridine dependent dehydrogenases, Konstanz.

Adams M.J., Buehner M., Chandrasekhar K., Ford G.C., Hackert M.L., Liljas A., Rossmann M.G., Smiley I.E., Allison W.S., Everse J., Kaplan N.O. and Taylor S.S. (1973) PNAS USA 70 1968.

Adams M.J., Helliwell J.R., Bugg C.E. (1977) J.Mol.Biol. 112, 183-197.

Arndt U.W. (1968) Acta Cryst. B24, 1355.

Arndt U.W., Champness J.N., Phizackerley R.P. and Wonacott A.J. (1973) J.Appl.Cryst. 6, 457.

Arndt U.W., North A.C.T., Phillips D.C. (1964) J.Sci.Inst. 41, 421.

Arndt U.W. and Willis B.T.M. (1966) "Single crystal diffractometry" C.U.P.

Arndt U.W. and Wonacott A.J. (1977) "The rotation method in crystallography", North-Holland.

Banner D.W. (1972) D.Phil. thesis, University of Oxford.

Barrington Leigh J. and Rosenbaum G. (1976) Ann Rev.Biophys. and Bioeng. 239.

Betts S.A. and Mayer R.J. (1975) Biochem.J. 151, 263-270.

Biesecker G., Harris J.I., Thierry J.-C., Walker J.E. and Wonacott A.J. (1977) Nature 266, 328.

Blake C.C.F. and Evans P.R. (1974) J.Mol.Biol. 84, 585.

Blake C.C.F. and Phillips D.C. (1962) "Biological effects of ionising radiation at the molecular level" in IAEA Symp. Vienna p. 183.

Blow D.M. (1957) D. Phil., Cambridge.

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

Blundell T.S. (1970) unpublished results.

Blundell T.L. and Johnson L.N. (1976) "Protein crystallography", Academic press.

Bridges R.B., Palumbo M.P. and Wittenberger C.L. (1975) J.Biol.Chem. 250, 6093-6100.

Buehner M., Ford G.C., Moras D., Olsen K.W. and Rossmann M.G. (1976)

J.Mol.Biol. 90, 25.

Chandrasekhar K., McPherson Jr. A., Adams M.J., Rossmann M.G. (1973) J.Mol.Biol. 76, 503-518.

Clift S.E. (1977) Part II thesis, Oxford.

Cochran W. (1952) Acta Cryst. 5, 65-67.

Cochran W. and Douglas A.S. (1955) Proc.Roy.Soc.Ser.A. 227, 486-500.

Codling K. (1973) Reports on progress in Physics, 36, 541.

Cox E.G. and Shaw W.F.B. (1930) Proc.Roy.Soc. 127, 71.

Delbaere L.T.J., Hutcheon W.L.B., James M.N.G. and Thiessen W.E. (1975) Nature 257, 758-763.

Diamond R. (1969) Acta Cryst. A25, 43.

Dickerson R.E., Kendrew J.C. and Strandberg B.E. (1961) Acta cryst. 14, 1188.

Dodson E.J. (1975) International Summer School of Crystallographic computing (Int. Union of Crystallography), Prague on "Refinement of heavy Atom Parameters".

Dodson E.J. (1975) ibid. "Handling protein native and derivative data".

Dodson E.J., Evans P.R. and French S. (1975) in "Anomalous scattering", p. 423 Eds. Ramaseshan S. and Abrahams S.C.

Dodson E.J. and Vijayan M. (1971) Acta cryst. B27, 2402.

Dyson J.E.D. and D'Orazio R. E. (1971) Biochem. Biophys.Res.Comm. 43, 183-188.

Dyson J.E.D. and D'Orazio R.E. (1973) J.Biol.Chem. 248, 5428-5435.

Dyson J.E.D., D'Orazio R.E. and Hanson W.H. (1973) Arch.Biochem. Biophys. 154, 623-635.

Eklund H., Nordström B., Zeppezauer E., Söderlund G., Ohlsson I., Boiwe .T. and Brändén C.I. (1974) Febs letts. 44, 200.

Eklund H., Nordström B., Zeppezauer E., Söderlund G., Ohlsson I., Boiwe T., Söderburg B.O., Tapi A.O., Brändén C.I. and Åkeson Å. (1976) J.Mol.Biol. 102, 27-59.

Evans P.R. (1973) D.Phil., Oxford.

Fletterick R.J., Sygusch J., Murray N., Madsen N.B. and Johnson L.N. (1976) J.Mol.Biol. 103, 1-13.

Fox G.C. and Holmes K.C. (1966) Acta Cryst. 20, 886.

French S. (1975) D.Phil., Oxford.

French S. and Wilson K.S. (1977) Manuscript in preparation.

Fukamachi T. and Hosoya S. (1975) Acta Cryst. A31, 215.

Germain G., Main P., and Woolfson M.M. (1971) Acta Cryst. A27, 368-376.

Godwin R.P. (1969) Springer Tracts in Modern Physics (Berlin) 51 1-73.

Green D.W., Ingram V.M. and Perutz M.F. (1954) Proc.Roy.Soc. Ser A 225, 287.

Haensel R. (1973) in Int. Symp. for Synchrotron Radiation users, Daresbury, DNPL/R26 ed Marr G.V. and Munro I.H.

Hamilton W.C., Rollett J.S. and Sparks R.A. (1965) Acta Cryst. 18, 129.

Harker D. and Kasper J.S. (1948) Acta Cryst 1, 70-75.

Harmsen A, Lebermann K and Schultz G.E. (1976) J.Mol.Biol. 104, 311-314.

Harrison S.W. (1973) in "Research applications of synchrotron radiation" p. 103-5, Brookhaven BNL 50381 eds. Watson and Perlman.

Helliwell J.R. (1977) New Scientist 76, 646-648.

Hendrickson W.A. (1976) J.Mol.Biol. 106, 889-893.

Hendrickson W.A., Love W.E. and Karle J. (1973) J.Mol.Biol. 74, 331.

Herzenberg A. and Lau H.S.M. (1967) Acta Cryst. 22, 24-28.

Holmes K.C. (1974) Proc. Int.Conf.Vac.U.V.Radiation physics 4th pp. 809-822, Hamburg.

Huber R. and Kopfmann G. (1969) Acta Cryst. A25, 143.

Huxley H.E. (1953) Acta Cryst 6, 457.

International Tables for X-ray Crystallography eds. Henry N.F.M. and Lonsdale K.

Irwin M.J. (1976) D.Phil. Cambridge.

Karle J. and Hauptmann H. (1950) Acta Cryst. 3, 181-187.

Karle J. and Karle I.L. (1966) Acta Cryst. 21, 849.

Kopfmann G. and Huber R. (1968) Acta Cryst. A24, 348.

Kraut J., Robertus J.D., Birktoft J.J., Alden R.A., Wilcox, P.E. and Powers J.C. (1971) Cold Spring Harbor Symp. 36, 117-123.

Littlefield T.A. and Thorley N. "Atomic and Nuclear Physics" Modern University Physics series (revised 1970).

Lowry O.H., Passoneau J.V. and Rock M.K. (1961) J.Biol.Chem. 236, 2756.

Marsh D.J. and Petsko G.A. (1973) J.Appl.Cryst. 6, 76.

Matthews B.W. (1966) Acta Cryst 20, 230.

Matthews B.W. (1966) Acta Cryst. 20, 82.

Matthews B.W. (1968) J.Mol.Biol. 33, 491.

Matthews B.W., Klopfenstein C.E. and Colman P.M. (1972) J.Phys.Sci.Inst. 5, 353.

Matthews K.A., Alden R.A., Bolin J.T., Freer S.T., Hamlin R., Xuong N., Kraut J., Poe M., Williams M., Hoogsteen K. (1977) Science 197, 452-455.

McCandlish L.E., Stout G.H. and Andrews L.C. (1975) Acta Cryst. A31, 245-249.

Mosley J., (1973) D.Phil. Oxford.

Navia M. and Sigler P. (1974) Acta Cryst. A30, 706.

Neidle S. (1973) Acta Cryst B29, 2645-2647.

Nockolds C.E. and Kretsinger R.H. (1970) J.Phys.Sci.Inst. 3, 842.

North A.C.T. (1965) Acta Cryst. 18, 212.

North A.C.T., Phillips D.C., Scott Mathews F. (1968) Acta Cryst., A24, 351.

Oatley S.J. (1976) D.Phil., Oxford.

Olsen K.W., Michael Garavito R., Sabesan M.N., Rossmann M.G. (1976) J.Mol.Biol. 107, 577-584.

Pearse B.M.F. Ph.D. thesis, London.

Pearse B.M.F. and Harris J.I. (1973) FEBS letts. 38, 49-52.

Pearse B.M.F. and Rosemeyer M.A. (1974) Eur.J.Biochem. 42, 213-223.

Perutz M.F., Muirhead H., Cox J.M. and Goaman L.G.G. (1968) Nature 219, 131-139.

Phillips D.C. (1964) J.Sci.Instr. 41, 123.

Phillips J.C., Wlodawer A., Goodfellow J.M., Watenpaugh K.D., Sieker, L.C., Jensen L.H., Hodgson K.O. (1977) Acta Cryst. A33 445-455.

Phillips J.C., Wlodawer A., Yevitz M, Hodgson K.O. (1976) PNAS USA 73, 128-132.

Popjak G. (1970) in The Enzymes ed. P.D.Boyer vol II. p. 116-215, New York, Academic Press.

Procsal D. and Holten D. (1972) Biochemistry 11, 1310.

Ramaseshan S. and Abrahams S.C. (1975) Eds. Inter. Congress Conference on Anomalous Scattering IUCR.

Rao S.T. and Rossmann M.G. J.Mol.Biol. 76, 241 (1973).

Richardson J.S. (1977) Nature 268, 495-500.

Rippa M., Picco C., Signorini M., Pontremoli S. (1971) Arch.Biochem.Biophys. 147, 487-492.

Rippa M. and Pontremoli S. (1969) Arch.Biochem.Biophys. 133, 112-118.

Rippa M., Signorini M., Pontremoli S. (1967) Eur.J.Biochem., 1, 170-178.

Rippa M., Signorini M., Pontremoli S. (1972) Arch.Biochem.Biophys. 150, 503-510.

Rippa M., Spanio L. and Pontremoli S. (1967) Arch.Biochem.Biophys. 118, 48-57.

Rogers D. (1965) in "Computing methods in crystallography" (ed. J.S. Rollett) Pergamon press.

Rossmann M.G., Liljas A., Brändén C.I., Banaszak L.J. (1975) in The Enzymes, Ed., Boyer P.D. p. 61-102, Academic Press, New York, 3rd. Ed.

Sayre D. (1952) Acta Cryst. 5, 60.

Schultz G.E. and Schirmer R.H. (1974) Nature 250, 142-144.

Schwager P., Bartels K., Huber R. (1973) Acta Cryst. A29, 291.

Scott W.A. and Abramsky T. (1973) J.Biol.Chem. 248, 3535-3541.

Silverberg M. (1973) D.Phil., Oxford.

Silverberg M. and Dalziel K. (1973) Eur.J.Biochem. 38, 229-238.

Silverberg M., Dalziel K. and Adams M.J. (1973) Biochem.Soc Trans. 1, 1132-1134

Steitz T.A. (1968) Acta Cryst, B24, 504.

Steitz T.A., Fletterick, R.J., Anderson W.F. and Anderson C.M. (1976) J.Mol.Biol. 104, 197-222.

Sternberg M.J.E. (1977) D.Phil. , Oxford.

Sternberg M.J.E. and Thronton J.M. (1976) J.Mol.Biol. 105, 367-382.

Stout G.H. and Jensen L.H. (1968) "X-ray structure determination", Macmillan.

Tickle I.J. (1975) Acta Cryst. B31, 329.

Toews M.L., Kanji M.I. and Carper W.R. (1976) J.Biol.Chem. 251, 7127-7131.

Villet R.H. and Dalziel K. (1972) Eur.J.Biochem. 27, 251-258.

Watson H.C., Shotton D.M., Cox J.M., Muirhead H. (1970) Nature 225, 806.

Webb L.E., Hill E.J., Banaszak L.J. (1973) Biochemistry 12, 5101.

Webb N.G., Samson S., Stroud R.M., Gamble R.C. and Baldeschwieler J.D. (1977) J.Appl.Cryst. 10, 104-110.

Whittaker E.J.W. (1953) Acta Cryst. 6, 218.

Wilson A.J.C. (1949) Acta Cryst. 2, 318.

Wilson K.S. (1974) D.Phil., Oxford.

Woolfson M.M. (1970) "An Introduction to X-ray crystallography" CUP.

Woolfson M.M.(1971) Rep. on Progress in Physics , 34, No. 5, p. 369.

Wycoff H.W. (1973) in "Research Applications of Synchrotron Radiation", eds. Watson and Perlman BNL 50381 pp. 133-138.

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

Xuong N. and Freer S.T. (1971) Acta Cryst B27, 2380.

Xuong N., Kraut J., Sely O., Freer S.T. and Wright C.S. (1968) Acta Cryst. B24, 289.

Zachariasen W.H. (1952) Acta Cryst. 5, 68-73.

Zappe H.A., Krohne-Ehrick G. and Schultz G.E. (1977) J.Mol.Biol. 113, 141-152.

Errata