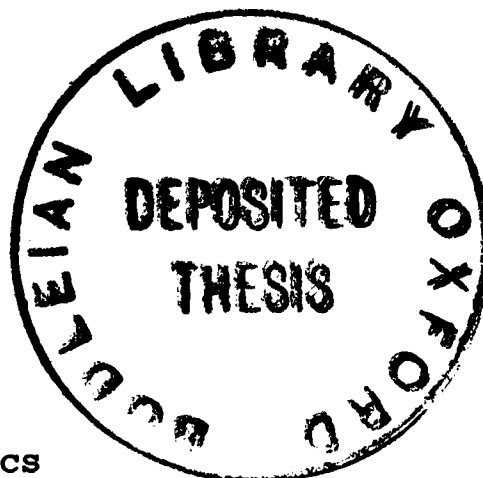


STRUCTURAL STUDIES OF IMMUNOGLOBULIN MOLECULES
THE FV FRAGMENT OF MOUSE MYELOMA PROTEIN M315

DAVID RICHARD ROSE

Thesis submitted for the degree of
Doctor of Philosophy
University of Oxford



Laboratory of Molecular Biophysics
Department of Zoology

Exeter College
Michaelmas Term, 1981

ABSTRACT

Structural Studies of Immunoglobulin Molecules

The Fv Fragment of Mouse Myeloma Protein M315

David R. Rose
Exeter College

D.Phil. Thesis
Michaelmas Term, 1981

The antibody is the main weapon in the body's defence against invading foreign materials. The different functions of the antibody molecule are performed by different parts of the structure. Some of these regions can be isolated in the form of enzymically-cleaved fragments. The Fv fragment, consisting of the variable regions of the light and heavy chains of the molecule, is the site of binding of the foreign antigen to which the antibody is specific.

Protein M315, a mouse myeloma protein, has an affinity for small molecules containing dinitrophenyl groups, comparable to that of a true antibody for its antigen. It is thus a particularly appropriate tool for examining antigen-antibody interactions.

The genetic mechanism of how the very large repertoire of antigen-binding specificities of antibodies arises in an organism has puzzled immunologists in the past. Recently, some progress on discovering this mechanism has been made, and some of the observations which have been reported and theories which have been proposed are reviewed.

The Fv fragment, which contains the antigen-binding site of the immunoglobulin, has been isolated from M315, and crystallized. The 6Å resolution data were measured on the four-circle diffractometer, and those to 2.7Å resolution by rotation photography. Phases were found by the method of isomorphous replacement with two heavy-atom compounds. The determination and refinement of the positions and extent of heavy-atom binding, are described. Electron density maps were calculated initially to 6Å and finally to 4.5Å resolution.

The 4.5Å resolution electron density map shows two complete Fv fragments in the asymmetric unit. The interactions between the V_H and V_L subunits are very close. However, the two molecules are oriented quite differently in the unit cell.

An initial examination of the binding of a small hapten, dinitroaniline, has been made. Crystals which were isomorphous with the native ones, were grown of Fv in the presence of the hapten. Preliminary indications are that the two crystallographically independent molecules may have different binding properties.

The directions in which the project may proceed, including the extension of the resolution and the assignment of atomic coordinates, are discussed.

ACKNOWLEDGEMENTS

The work described in this thesis was done under the supervision of Professor Sir David Phillips. I am very grateful for all that he has taught me not only about the technique of crystallography, but about how to conduct research, as well.

Nearly every member of the laboratory, from lecturer down to student, has at one time or another offered me help or advice. I thank them all for making the laboratory such a marvellous place for learning.

All the Fv crystals used in this project were grown by Dr. R. Aschaffenburg. I wish to thank him for his generosity in providing crystals and solutions on request, and for his friendship which I greatly value.

I am indebted to Dr. B.J. Sutton for his collaboration in some aspects of the project, for his advice and instruction throughout, for many useful discussions, and for his friendship.

I would like to acknowledge the interest of Dr. R. Dwek in the Fv study and the help provided by him and his group in the preparation of the fragment for crystallization.

All the occupants of F12 and F9 have provided great companionship, encouragement and (invariably conflicting) advice. I am especially grateful to Drs. P. Artymiuk and W. Pulford for their help with computing.

The excellent technical staff of the laboratory has contributed to this project in the never-ending struggle to keep the necessary equipment in working order. My thanks go to Lance Mangold for his help with the photographs.

Financial support during this work was provided by the Rector and Fellows of Exeter College through the generous award of an Usher Cunningham Studentship.

I am grateful to my parents for their support and encouragement

throughout my education.

Finally, I must acknowledge the contribution made by my wife, Tina, to the presentation of this thesis. She spent many weeks typing the text, drawing diagrams, and putting up with the stresses that accompany such an undertaking.

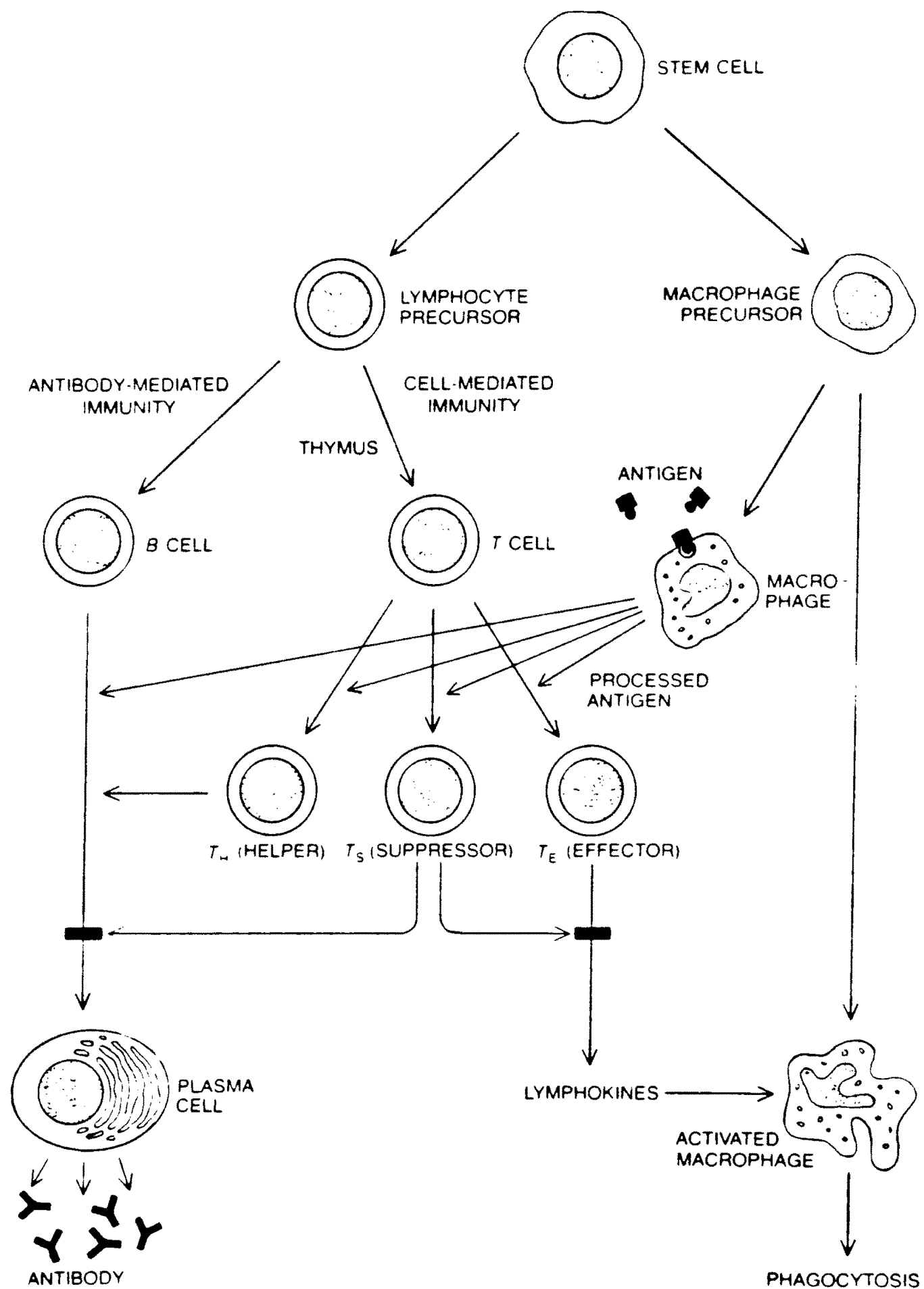
TABLE OF CONTENTS

		<u>PAGE</u>
CHAPTER I	THE STRUCTURE OF IMMUNOGLOBULINS AND THE GENERATION OF DIVERSITY	
I.A.	The Immunoglobulin Molecule	1
I.B.	Three-Dimensional Structure	2
I.C.	Haptens and Myeloma Proteins	3
I.D.	The Generation of Diversity	4
CHAPTER II	IMMUNOGLOBULIN M315	
II.A.	Production and Purification of M315	15
II.B.	The Hapten-Binding Activity of M315	18
CHAPTER III	PROTEIN STRUCTURE DETERMINATION BY X-RAY DIFFRACTION	
III.A.	Diffraction from a Crystal	25
III.B.	Isomorphous Replacement	31
III.C.	Bragg's Law and the Ewald Sphere	33
III.D.	Methods of Data Collection	34
III.E.	The Determination and Refinement of Heavy Atom Positions	40
III.F.	The Calculation of Phases	43
CHAPTER IV	THE STRUCTURE DETERMINATION TO 6Å RESOLUTION	
IV.A.	Preliminary Crystallography	46
IV.B.	The Search for Isomorphous Heavy-Atom Derivatives	50
IV.C.	Native Data to 6Å Resolution	52
IV.D.	Data Collection from Platinum and Uranyl Derivatives	55
IV.E.	The Isomorphous Difference Patterson Functions	57
IV.F.	Preliminary Refinement of the Heavy-Atom Sites	58
CHAPTER V	THE EXTENSION OF THE RESOLUTION	
V.A.	Data Collection	65
V.B.	Data Processing	67
V.C.	Refinement of the Heavy-Atom Sites	68

CHAPTER VI	THE ELECTRON DENSITY MAPS AND THEIR INTERPRETATION	
VI.A.	Hapten-Binding Studies	72
VI.B.	The Presentation of the Maps	73
VI.C.	The 6Å Resolution Map	77
VI.D.	The 4.5Å Resolution Map	79
CHAPTER VII	CONCLUSION	
VII.A.	The Improvement of the 4.5Å Resolution Result	85
VII.B.	The Generation of a Starting Model	87
VII.C.	The Extension of the Resolution	88
VII.D.	Hapten-Binding Studies	88
APPENDIX I	COMPUTER PROGRAMS	90
APPENDIX II	REPRINT Aschaffenburg <u>et al</u> , 1979	92
APPENDIX III	DATA TO 6Å RESOLUTION WITH PHASES	93
APPENDIX IV	DATA TO 4.5Å RESOLUTION WITH PHASES	94
APPENDIX V	DATA TO 2.7Å RESOLUTION WITH PHASES	95
REFERENCES		96

ABBREVIATIONS AND SYMBOLS

$\text{\AA}$	Angstrom Unit (10^{-10} m)
a,b,c	Real cell dimensions
a*,b*,c*	Reciprocal cell dimensions
B	Isotropic temperature factor
E	Lack-of-closure error
$\bar{F}$	Structure factor
F,Fs	Structure factor amplitude
$\Delta f', \Delta f''$	Real and imaginary parts of the anomalous scattering
h,k,l	Reflection indices
H (subscript)	Heavy-atom
m	Figure of merit
P (subscript)	Protein
PH (subscript)	Protein - Heavy-atom complex
R	Residual index (defined as used)
u,v,w	Vector space coordinates
w	Weight
Xo,Yo,Zo	Rotation camera axes
α	Phase
β	Monoclinic angle
$\Delta(\text{iso})$	Isomorphous difference
$\Delta, \Delta(\text{anom})$	Anomalous difference
θ	Bragg angle
λ	Wavelength
ρ	Electron density
σ	Standard deviation
$\omega, \phi, \chi, \theta$	Diffractometer angles
C	Constant region
D	Diversity segment
Dnp	2,4 dinitrophenyl
Fab	Antigen-binding fragment
Fc	Crystallizable fragment (C_H2 and C_H3 domains)
Fv	Variable fragment (V_H and V_L)
H (subscript)	Heavy chain
Ig	Immunoglobulin
J	Joining region
L (subscript)	Light chain
M315	Myeloma Protein from tumour MOPC-315
PEG	Polyethylene glycol
Tnp	2,4,6 trinitrophenyl
V	Variable region



DEVELOPMENTAL PATTERNS of *T* and *B* lymphocytes from their origin in stem cells of the bone marrow are depicted. The *B* cells evolve directly from the stem cells; the *T* cells develop under the influence of the thymus gland. On stimulation by an antigen *B* cells differentiate into plasma cells secreting the antibodies that attack bacteria and viruses before they enter the cells of the host; this is antibody-mediated immunity. *T* lymphocytes exert their effects directly; this is cell-mediated immunity. There are also subpopulations of *T* cells: helper cells, which interact with *B* cells to amplify the production of antibody; effector cells, which carry out the direct cell-killing functions of *T* cells and make the lymphokines that are responsible for delayed hypersensitivity, and suppressor cells, which regulate both parts of the immunological response. Some macrophage cells present antigens to *T* and *B* cells in the proper orientation; others are activated by lymphokines to destroy invading microorganisms (phagocytosis).

Figure I.1. From Rose (1981).

CHAPTER I

The Structure of Immunoglobulins and the Generation of Diversity

The key weapon in the defence of vertebrates against the invasion of foreign materials from the environment is the production and activity of antibodies. Antibodies are released by mature B lymphocytes stimulated by exposure to the foreign antigen as presented by the macrophage. Production of antibodies can be either enhanced or inhibited by T lymphocytes (figure I.1). Many reference books on the complex processes of the immune system are available (for example, Hobart and McConnell, 1978, Roitt, 1974, and Rose, Milgrom and van Oss, 1979). The following discussion is concerned only with the structure and specificity of the antibody molecule.

I.A. The Immunoglobulin Molecule

Antibodies comprise a group of serum proteins called immunoglobulins. They have a characteristic four-chain structure (figure I.2) which consists of one pair of identical heavy chains (molecular weight 50000 to 75000) and one pair of identical light chains (molecular weight about 25000) (Edelman and Gally, 1964).

Each polypeptide chain is made up of units of about 110 residues in length, which have largely similar amino-acid sequences. A heavy chain contains four (or occasionally five) units and a light chain two. Each unit associates closely with the adjacent unit of another chain to form a domain, but only through the connecting polypeptide chain with the preceding or following unit of the same chain (figure I.2).

The class of an immunoglobulin is determined by the amino-acid sequence of the three (or four) carboxy-terminal domains of the heavy chain, termed the constant (C) domains. There have been five classes observed of which two (IgG and IgA) are divided into subclasses (table I.1).

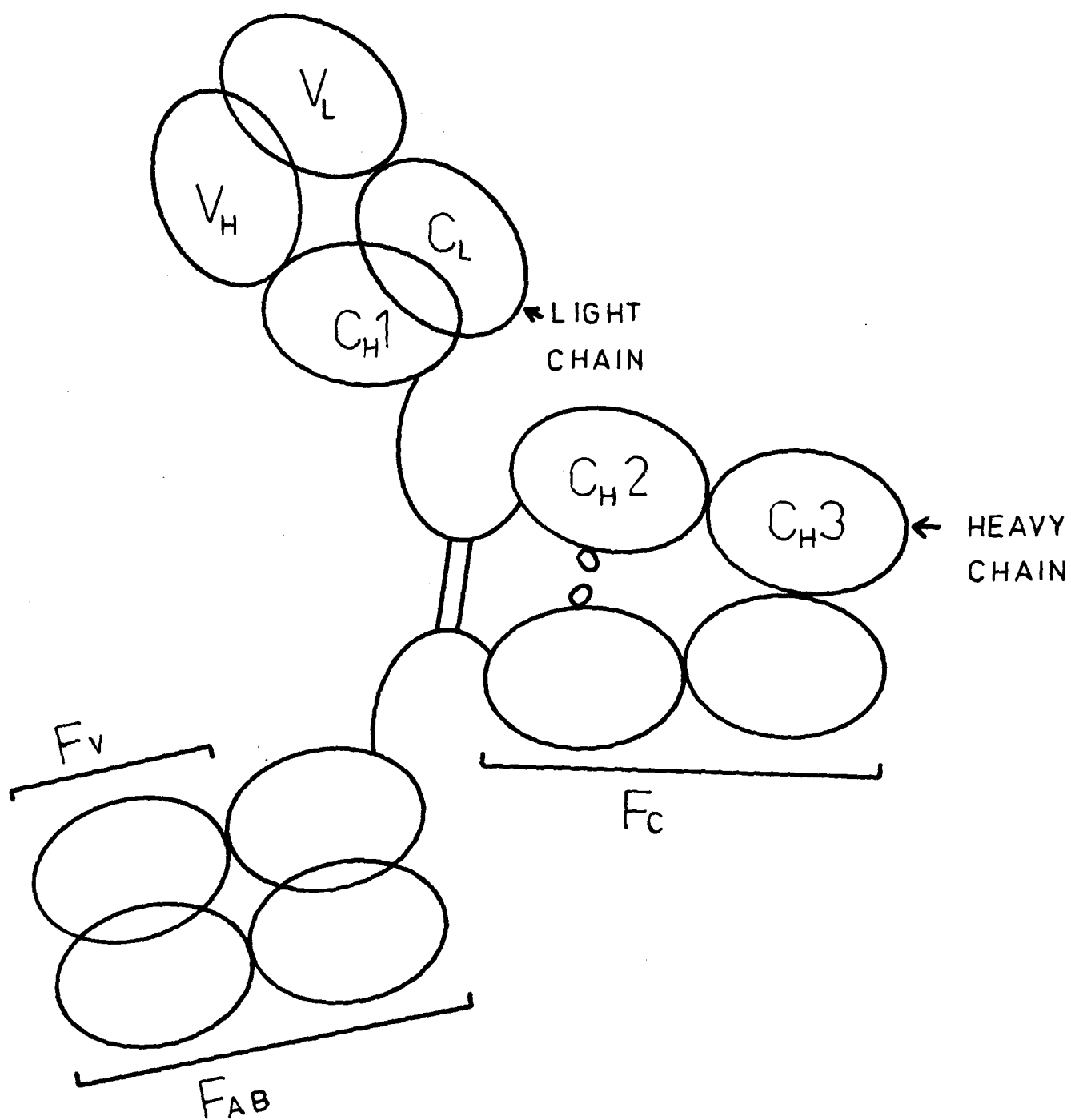


Figure I.2. A representation of the two-chain structure of the immunoglobulin molecule. Three fragments obtained by enzymic cleavage, the hinge disulphide bond (double line) and the inter-heavy chain carbohydrate are shown. Abbreviations: H, heavy chain; L, light chain; C, constant region; V, variable region.

<u>Class</u>	<u>H-Chain</u>	<u>No. of CH</u> <u>Regions</u>	<u>(H₂L₂)</u>	<u>Structure</u>	<u>Concentration(mg/ml)</u> <u>in Normal Serum</u>
IgG1	γ1	3		monomer	9
IgG2	γ2	3		monomer	3
IgG3	γ3	3		monomer	1
IgG4	γ4	3		monomer	0.5
IgM	μ	4		pentamer	1.5
IgA1	α1	3		monomer	3
IgA2	α2	3		monomer	0.5
sIgA ⁽¹⁾	α1 or α2	3		dimer	0.05
IgD	δ	3 or 4		monomer	0.03
IgE	ε	4		monomer	0.0003

(1) Secretory IgA

Table I.1. Properties of Human Immunoglobulins
after Nissonoff et al (1975).

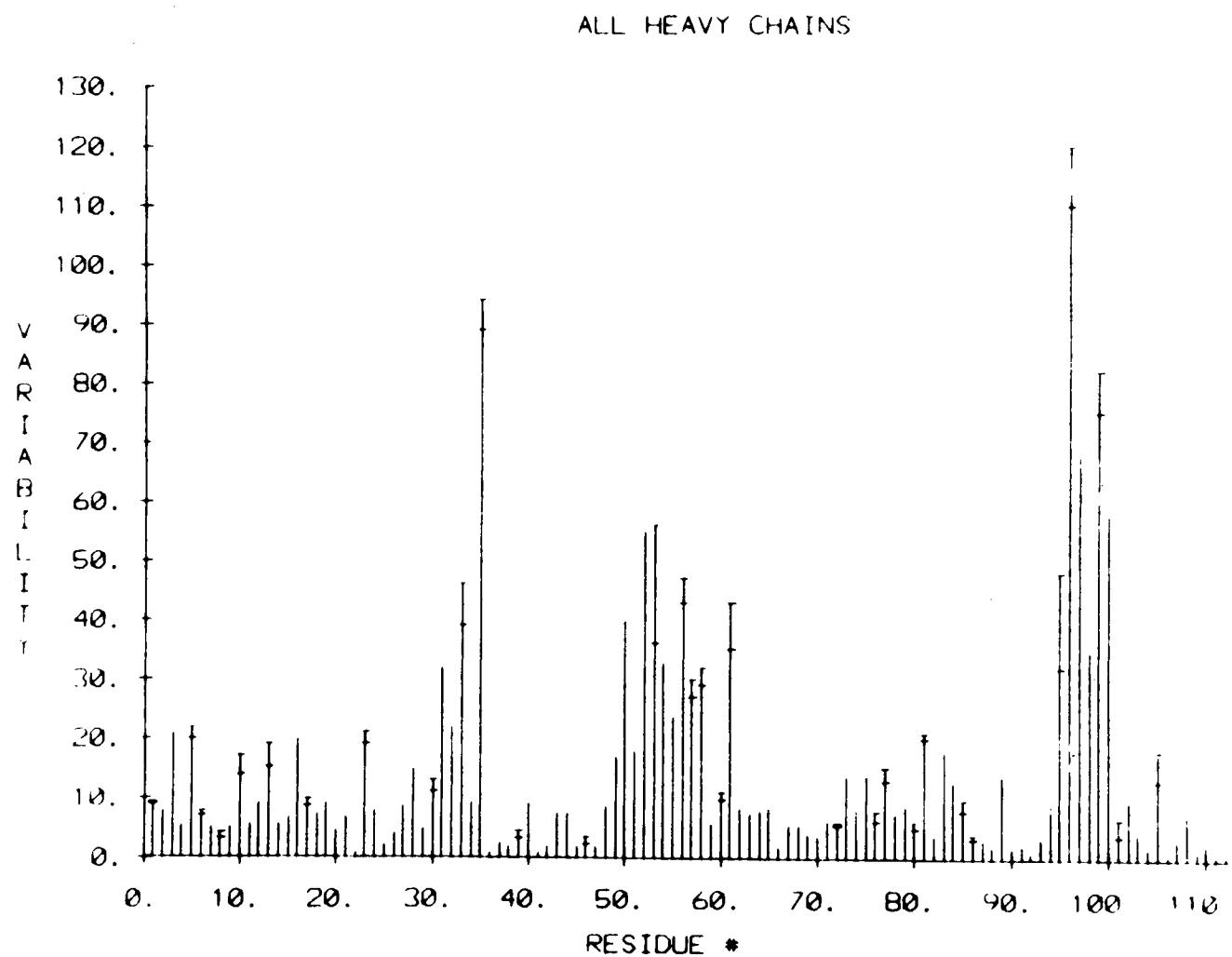
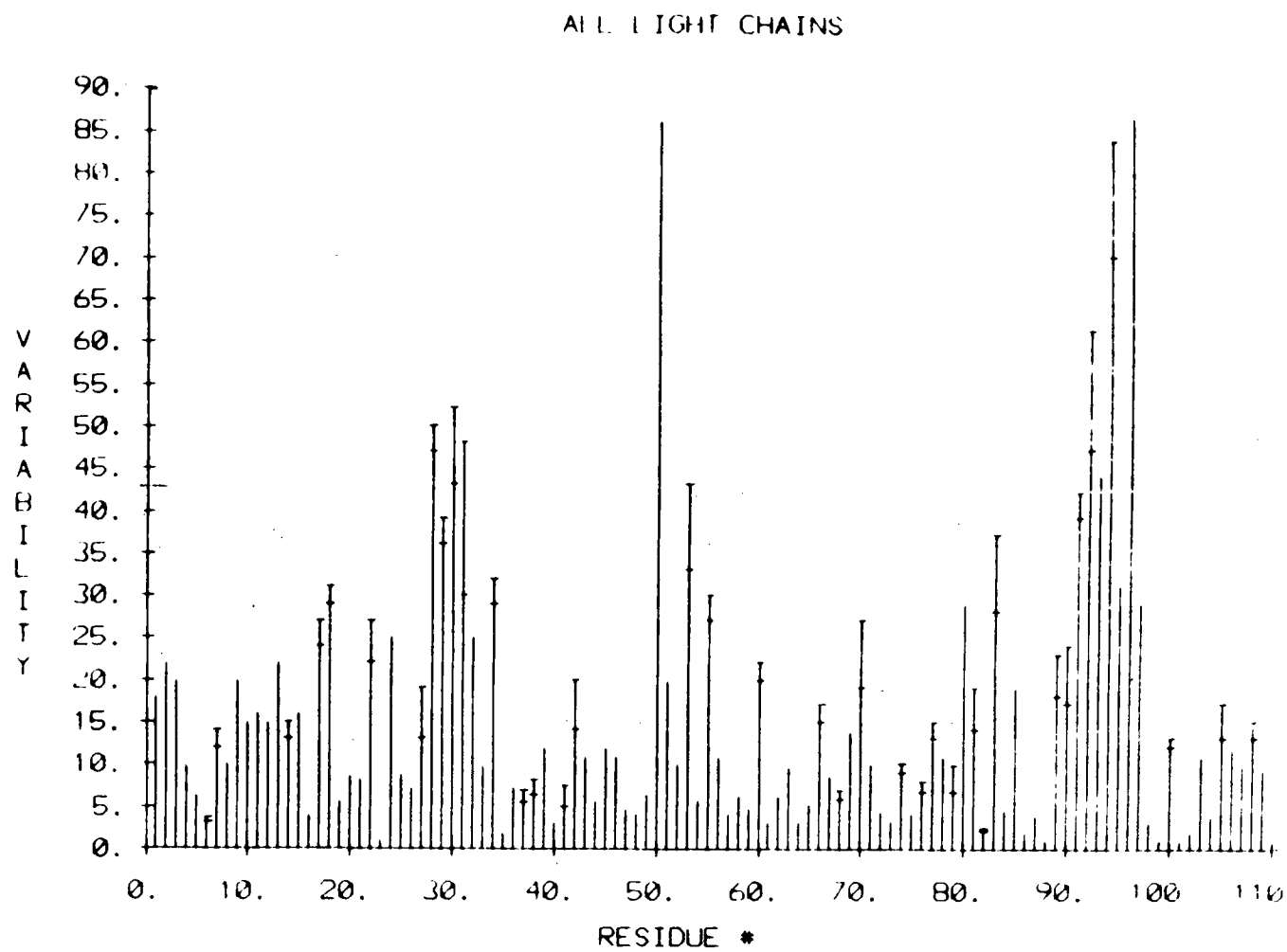


Figure I.3. Plots of the variability of each amino-acid position in immunoglobulin V-regions. All known light and heavy chain sequences were used. From Kabat, Wu and Bilofsky (1976).

The carboxy-terminal unit of the light chain also has a common amino-acid sequence in all immunoglobulins. The light chains fall into two genetically determined types, κ and λ , both with sub-types, either one of which can associate with a heavy chain of any class.

The amino-terminal units of both heavy and light chains are termed variable (V) because their amino-acid sequences differ in immunoglobulins with different antigen-binding specificities. Wu and Kabat (1970) defined a quantitative measure of the variability of a position in the sequence; the variability of a position is the number of different amino-acid residues observed at that position divided by the frequency of occurrence of the most commonly observed residue at that position. Thus an invariable (conserved) position has a variability of 1.0. Plots of the variability of V-regions against the sequence number (figure I.3) show that within a unit, three short segments of four to six amino-acid residues in length, termed hypervariable regions (Kabat and Wu, 1971), show exceptionally high variabilities.

I.B. Three-Dimensional Structure

Because of the weak interaction of units along a chain described in the previous section, immunoglobulins can easily be cleaved by enzymes. Some of the resulting fragments are named in figure I.2. Porter (1959) first used papain to obtain two fragments and named one Fab because it contained the antigen-binding properties of the antibody. As the experiments were performed on pooled IgG of varied antigenic specificity from rabbit serum, the Fab fraction was inhomogenous. The second fragment was named Fc because it was readily crystallizable and, therefore, uniform within the IgG class. The Fc fragment contains recognition sites for many of the functions mediated by the antibody regardless of antigenic specificity, the activation of the complement system and the binding to

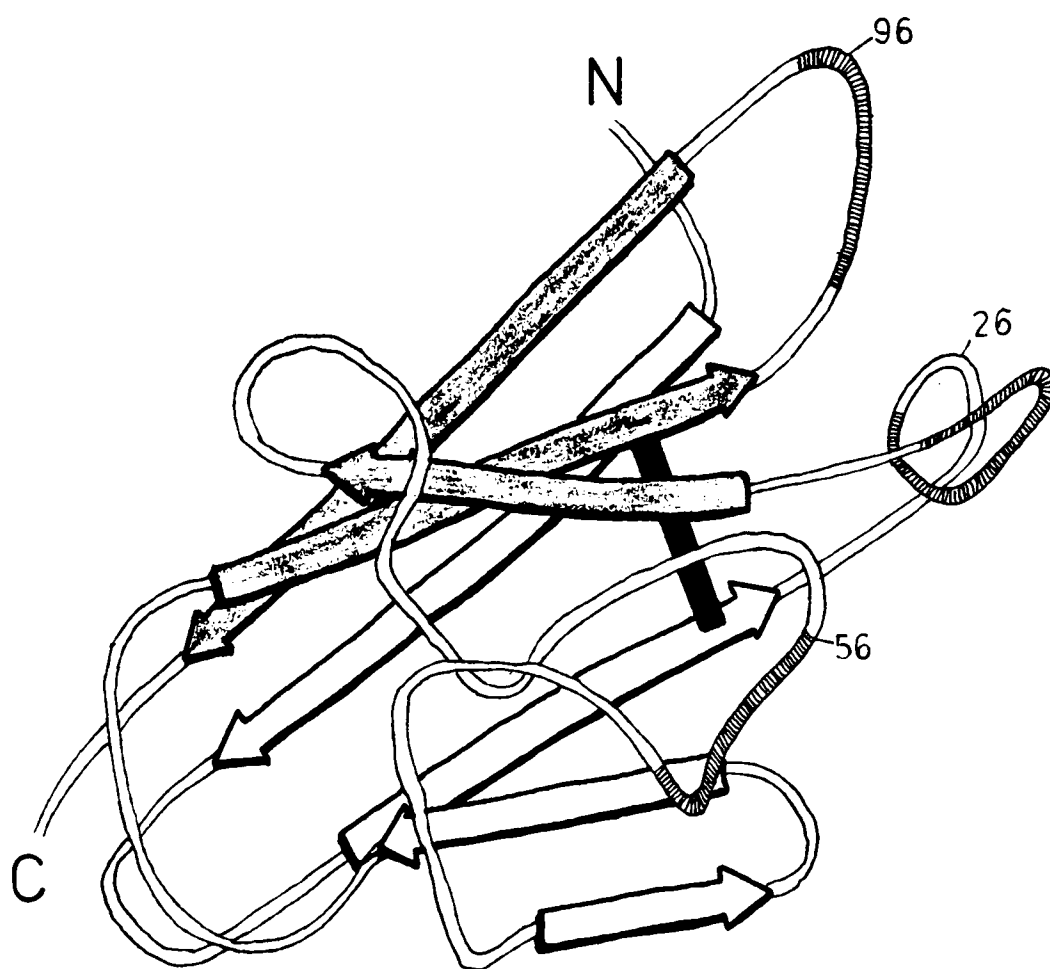


Figure I.4. The tertiary fold of a variable region. β -strands are represented by arrows, the disulphide bridge by a dark bar. The hypervariable loops are featured. After Schiffer et al (1973).

receptors on cells, for example.

X-ray crystallographic studies on a number of immunoglobulin fragments have been carried out (see reviews by Amzel and Poljak, 1979, Edmundson et al, 1978, and Kabat, 1978). The most striking feature which has emerged is the tertiary fold of the 110-residue unit referred to above. Each unit folds independently into a β -sheet sandwich (Cohen et al, 1980); that is, two anti-parallel β -sheets, one of three and the other of four strands, forming parallel faces connected by an inter-sheet disulphide bond (figure I.4). In general, this feature, the immunoglobulin fold, is common to all half-domains, constant and variable, though variations do exist. In particular, the V-region folding units have longer loops connecting the strands, and the faces formed by the sheets are less planar.

The hypervariable regions found in variable segments lie on the inter-strand loops of the fold. In the variable domain, three loops from the light chain lie close to three loops from the heavy chain. Residues lying within these hypervariable loops form the antigen-combining site. Thus, the amino-acid sequence of the framework region (the strands making up the β -sheet sandwich) is largely conserved, even in the variable units.

As a result, the entire antigen-binding site is contained within the variable domain. Indeed, an Fv fragment, (see figure I.1) obtained by enzymic cleavage (Chapter II) shows the full antigen-binding activity of the intact immunoglobulin (Inbar et al, 1972).

I.C. Haptens and Myeloma Proteins

Because of the diverse specificity of antibodies in the blood stream, it is impossible to purify in large amounts a particular antibody molecule. Even if it were feasible, structural studies of antigen-antibody complexes would be difficult because antigens normally occurring in the organism are

usually larger even than the antibody itself.

A hapten is a much smaller molecule, which binds to the antibody in a similar way to an antigen. Unlike antigens, haptens cannot independently generate an immune response. However, antibodies specific to the hapten can be elicited by binding it to a large carrier molecule. Once formed, these antibodies will bind to the hapten even in the absence of the carrier (van Oss and Grossberg, 1979). The hapten can also stimulate a clone of cells that make anti-hapten antibody, to reproduce.

The technique of Kohler and Milstein (1975) has facilitated the production of large amounts of monoclonal (homogenous) antibodies of a known specificity. However, before the advent of this method, structural studies depended to a large extent on the cancerous disease, multiple myeloma. A clone of uncontrollably reproducing B cells creates a large amount of specific antibody in the blood, from which whole immunoglobulin molecules, the myeloma proteins, can be obtained. The light chains are sometimes secreted in the urine and called Bence Jones proteins. Unfortunately, the true antigen corresponding to these homogenous antibodies is not known, if indeed one ever existed. Screening by a large number of small molecules, however, often leads to one which binds to the myeloma protein to an extent comparable to a true antigen-antibody combination. Most of the structural studies accomplished on immunoglobulin binding properties have been on myeloma proteins and strongly-binding small molecules commonly (though not necessarily correctly) called haptens.

I.D. The Generation of Diversity

The specific nature of the antibody-antigen interaction implies that the number of different combining sites and idiotypes that an organism can produce is very large. The problem of how such a sizable repertoire of specificities arises remained a mystery for a long time. However, modern

techniques for the production of large amounts of a particular sequence of DNA by molecular cloning, for identifying specific nucleotide sequences in an unknown stretch of DNA, and for rapidly sequencing DNA and RNA molecules, have brought an understanding of the generation of diversity within reach.

1. Historical Theories of Antibody Formation (see Lachmann, 1975 for primary references).

In 1900, Ehrlich first proposed a selective theory by which an invading molecule combined with one of a number of pre-existing binding sites on a cell and stimulated the cell to produce and release antibodies containing this site. The theory, however, was inconsistent with the presence of free antibodies to non-biological antigens to which the organism would not normally come into contact.

For much of the first half of the twentieth century, a neo-Lamarckian theory was popular, according to which an antibody template was moulded around an invading antigen and then used by the organism to produce specific antibodies. However, it was shown in the 1950's and the 1960's that the specificity of antibodies was embedded in their primary amino-acid sequence. The only way to reconcile the theory to this observation is to propose that the template protein can control the synthesis or translation of the DNA coding for the antibody, a process that has not been detected in nature.

In the late 1950's, due mainly to the ideas of Jerne, Burnet and Lederberg, the clonal selection theory, for which there is now much evidence, was proposed. This was, in effect, a return to the selective theory with the important difference that each antibody-producing cell carried only one specificity. An invading antigen bound to a cell with the appropriate combining site displayed on its surface, stimulating the cell to proliferate into a clone of specific antibody-producing cells.

The number of specific combining sites required by an organism during its lifetime has been estimated to be of the order of 10^6 to 10^7 (Lachmann, 1975). Clearly, however, as the organism does not know at its inception to which antigens it will be exposed, a much larger repertoire must be present. The problem is how this diversity of antibody specificities arises.

The germ-line hypothesis states that the DNA coding for each antibody exists as a separate gene on the germ-line of the organism. Thus, once the immune system is fully developed, the complete antibody repertoire exists and no further evolutionary processes are required. This concept implies the existence of a large amount of redundant DNA, that which codes for the constant regions shared by all immunoglobulins of a certain class.

A second view is that a standard set of genes codes for each class of immunoglobulin and the variety of combining sites is generated somatically by mutations in the V-region-coding DNA sequences. The mechanism by which mutations arise only in the proper regions of DNA is mysterious. A structural requirement within the V-regions, the conservation of the framework β -sandwich, may permit changes only in the loop regions (Novotny, 1973, 1975), but this does not explain the constancy of the C-regions. The two viewpoints are not mutually exclusive; somatic mutations occurring in a large number of germ-line genes could account for the generation of diversity.

A key to the resolution of this dilemma has been the observation of the heavy-chain switch (Nossal et al, 1964). Although a single clone of antibody-producing cells expresses only one specificity, it can produce simultaneously or consecutively more than one class of immunoglobulin (for example, Sledge et al, 1976). In other words, identical variable region sequences of both light and heavy chains can associate with different

constant region sequences, which determine the class of the immunoglobulin.

2. The Genetics of Antibody Production.

Dreyer and Bennett (1965) first introduced the idea that an immunoglobulin molecule is coded for by more than one gene. They proposed that one portion of nucleic acid codes for the constant region of the light chain, while the gene segments for the variable region lie in rings of nucleic acid stacked along the chromosomes. Any one of this large number of rings can be combined with the constant region gene segment to complete the gene for the entire chain.

Gally and Edelman (1970) reviewed the evidence supporting the somatic recombination theory. By assuming that about 10 diverse V-region gene segments were available per subgroup, they estimated that at least 10^6 light chains could be formed by this mechanism alone. If a similar process occurs for the heavy chain, a large repertoire of idiotypes can be generated. Another attractive feature of this theory is that it also accounts for the heavy-chain switch. Gally and Edelman proposed that a VC gene derived by recombination might be excised in some way releasing the V-region gene segment to recombine again with another C-region gene segment, a proposal supported by Wang et al (1970).

Rabbitts (1979) refined this mechanism into a two-step model for the integration of heavy chain V- and C-region gene segments. In the first step, the germ-line DNA is translocated into a continuous stretch of DNA containing only one V-region gene segment, a short interruption sequence, and the gene segments for the constant region types. The second step is the transcription to the mRNA precursor, still retaining the codes for the C-regions. In the processing of this precursor to mRNA, the V-region gene segment is brought contiguous to that for one of the C-regions, the intervening RNA being excised. At a later time, the V-region gene segment

can be re-transposed in the 3' direction, adjacent to another C-region segment. This model implies that the heavy-chain switch is sequential; because the sequences intervening the V- and C-region gene segments are excised, a cell can only express constant regions in the order in which their gene segments lie on the mRNA. Further evidence (see section c below) has verified this prediction. First, the formation of the genes for the variable regions of the light and heavy chains will be considered.

a. Light Chains.

The technique of Southern (1975) for detecting known sequences of DNA, particularly applied by Tonegawa and his colleagues (Hozumi and Tonegawa, 1976) has provided clear evidence in support of the multiple gene segment-one polypeptide hypothesis. DNA from an early mouse embryonic cell and from a mature B lymphocyte was hybridized with mRNA coding for the whole κ light chain. From the different patterns of hybridization, Tonegawa concluded that the V- and C-region genes were separate in the immature but adjoining in the differentiated mature cell.

An even more interesting result came from this group when a DNA transcript of mRNA coding for a λ chain was used as a probe (Brack et al, 1978). They found that the light chain is coded for by not two, but three genes. Approximately the last 10 residues of the variable region, termed the J (joining) segment, are encoded separately from the first 95 residues. The existence of J-segments in the κ chains was demonstrated by Weigert et al (1978).

Soon afterwards, the sequencing techniques developed by Maxam and Gilbert (1977) and Sanger and Coulson (1975,1978) were applied to DNA coding for immunoglobulin chains. Seidman et al (1978) found a number of related V_{κ} genes, each coding for residues 1 to 97 of the light chain, in an uninterrupted sequence of germ-line DNA. Seidman and Leder (1978) also

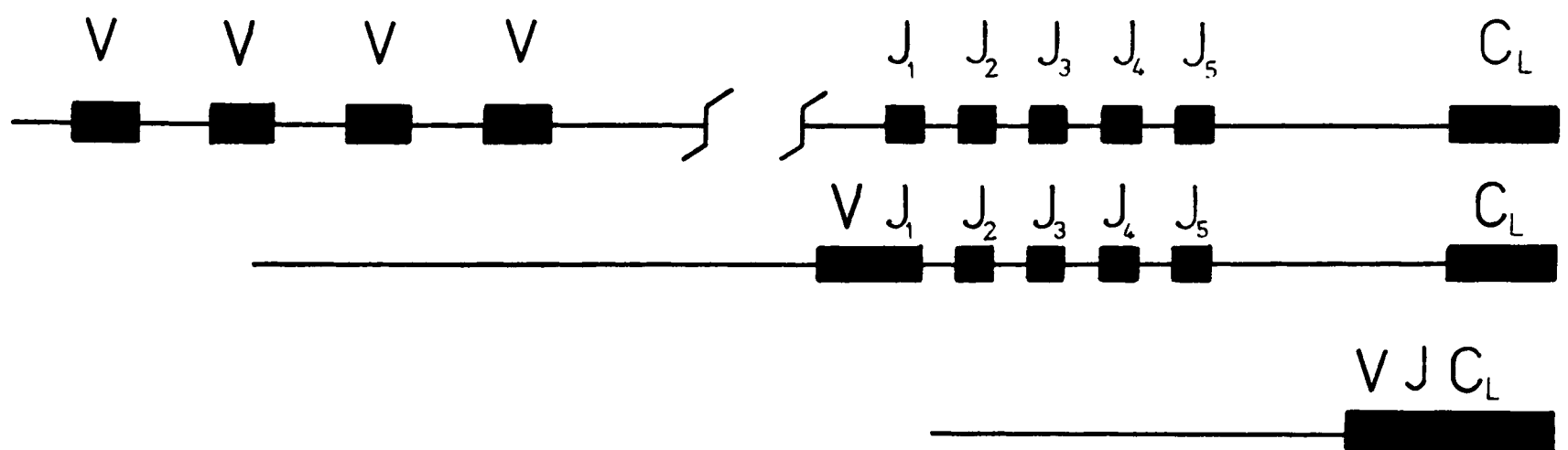


Figure I.5. The recombination of gene segments encoding the variable, joining and constant regions of a light chain. After Dunnick (1979).

VK21 5'		95	96	97		
		C	C	T	-	C C C - A C A - G T G
J1		<u>C A - C T G - T G</u> G - T G G - A C G 3'				
<u>Immunoglobulins</u>	<u>Recombined 95-96</u>		<u>Dipeptide Sequence</u>			
	95	96	95	96		
M63,M70,M41,etc.	C	C	G	-	T	G G
	C	C	T	-	T	G G
M173,6684	C	C	T	-	C	G G
M11, 7940	C	C	T	-	C	C G

Figure I.6. An example of flexibility at a V-J recombination site generating variability at residue 96. Recognition heptamers (see Table I.2) at the joining site are indicated.
 After Sakano, Rogers,et al (1979).

published some electron micrographs of the mRNA coding for a light chain, in a double-stranded complex with the germ-line DNA. Loops of uncoded DNA where the V-J-C junctions occur are clearly visible.

With the determination of more sequences of germ-line DNA, it has become clear that there are, at least in the κ -chain system, a number of J-segment gene sequences, as well as numerous V-region genes. Thus, as Seidman et al (1979) point out, antibody diversity could result from a somatic recombination of one of the many ("hundreds") of V-region genes with the several J-segment genes (figure I.5).

Sakano, Maki, et al (1979) have specified the number of J_{κ} gene segments in the mouse germ-line as five, one of which is non-functional. They have also suggested that the site of V-J joining may have some flexibility. This suggestion is in agreement with the observation that the point of V-J joining is in the third hypervariable region and the residue at the junction, 96 in mouse κ chains, shows more variability than any other residue (Weigert et al, 1980) (see figure I.3). Figure I.6 shows an example of the flexibility of V-J joining. The first non-coding triplet after the $V_{\kappa}21$ gene is CCT and the initial codon of the $J_{\kappa}1$ gene is GTG. The amino-acid codes which can be generated by combining these triplets without upsetting the "phase" of the DNA are shown, with examples in which they have been observed.

b. Heavy Chains.

Tsuzukida et al (1979) have examined an unusual human IgA which appears to have an identical C_{H3} region to the α_1 subclass, but C_{H1} , hinge and C_{H2} regions identical to the α_2 subclass. Study of several heavy chain deletion mutants (Frangione and Franklin, 1979) and sequencing studies (Sakano, Rogers et al, 1979) have confirmed that there are separate genes encoding the C_{H1} , hinge, C_{H2} and C_{H3} regions.

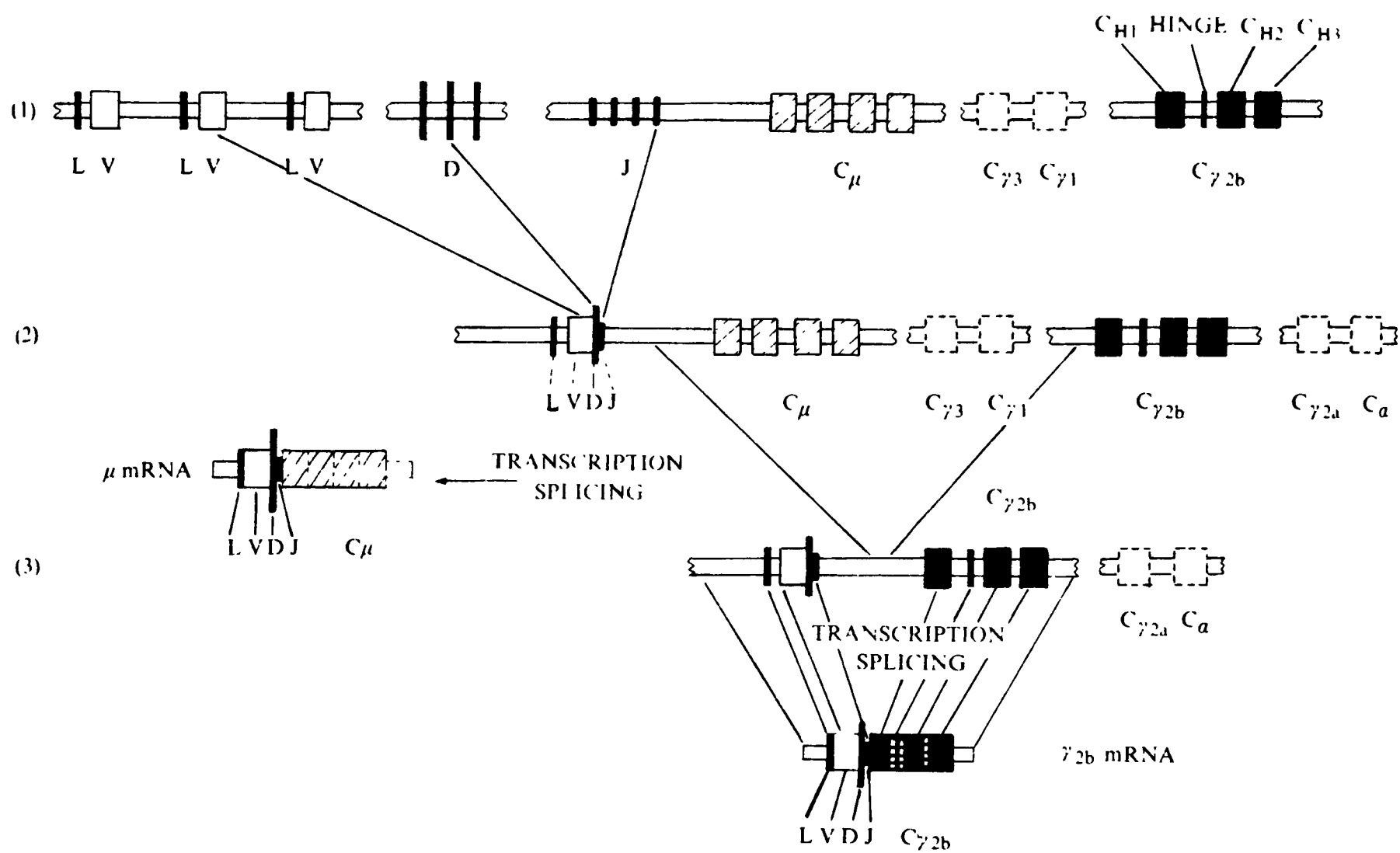


Figure I.7. The series of events in the formation of the mRNA encoding an IgG2b molecule. From Molgaard (1980).

Heavy chain V-regions are encoded separately from the C-regions. Kemp et al (1979) have searched for sequences homologous to mRNA coding for a μ chain in mouse embryo DNA. They concluded that a string of closely similar V-region genes lies along the germ-line DNA. The joining of V- and C-region gene segments occurs at the level of the nuclear RNA of mature B cells (Rabbitts, 1978).

Rabbitts et al (1980) have estimated that there are 40 to 50 V_H genes in mouse and about 80 in human germ-line DNA. This compares with about 300 V_K genes in the mouse and, presumably, a similar number in man. On the assumption that both species have similar numbers of J-segment genes for the various chains, these observations could be interpreted as an indication that somatic variation (mutation) plays a more important role in humans than in mice, and in V_H gene production than that of V_L .

Subsequent studies by Hood and his colleagues (Schilling et al, 1980, Early et al, 1980) have provided an alternative interpretation. Comparison of the amino-acid sequences of immunoglobulins which bind similar molecules, has shown that in addition to V- and J-gene segments, heavy chain genes include a third component, the D (diversity) segment. Germ-line DNA and heavy-chain mRNA sequences confirm that the D-gene segment encodes the third hypervariable region, amino-acid residues 102 to 106, of the heavy chain. Because of the larger number of diverse variable regions which can be generated by the inclusion of one of a number of D-segment genes, fewer V_H than V_L gene segments are required.

Thus, in the formation of a heavy chain gene, a V-region gene (for residues 1 to 101) recombines with one of a number of D-gene segments (102 to 106), and, by a similar mechanism, to a J-segment gene (107 to 115). The gene is completed by a second type of recombination, that of the VDJ-gene with the C-region gene (figure I.7). The nature of the non-coding

The two conserved blocks of sequences near the V-J or V-D-J joining sites			
J DNA segments	Nanomers		Heptamers
J _{K1}	GGT <u>TTTT</u> TGT	23	CACTGTG
J _{K2}	<u>AG</u> TTTTTGT	23	CAGTGTG
J _{K3}	GGGTTTTGT	21	CACTGTA
J _{K4}	GGTTTTTGT	24	CACTGTG
J _{K5}	GGTTTTTGT	23	CACTGTG
J _{λ1}	GGTTTTTGC	12	CACAGTG
J _{H1}	<u>AG</u> TTTTAGT	22	GACTGTG
J _{H2}	GGTTTTTGT	23	TAGTGTG
J _{H3}	<u>ATTTATT</u> TGT	21	CACAATG
		23	CAATGTG
J _{H4}	GGTTTTTGT	22	TATTGTG
		31	TACTATG
Basic sequence	GGTTTTTGT		CACTGTG
V DNA segments	Heptamers		Nanomers
V _{K21C}	CACAGTG	11	ACAAAAACC
V _{K21B}	CACAGTG	12	ACAAAAACC
V _{K41}	CACAGTG	12	ACATAAACC
V _{K2}	CACAGTG	12	ACATAAACC
V _{λ1}	CACAATG	22	TCAAGAACA
V _{λ11}	CACAATG	23	ACAAGAACA
V _{H141}	CACAGTG	23	ACAAATACC
Basic sequence	CACAGTG		ACAAAAACC

Two blocks of conserved sequences found in the 5'-flanking region of J DNA segments and in the 3'-flanking region of embryonic V DNA segments are compared. The numbers between the two types of sequences indicate the distance between them in base pairs. The bases different from those of the basic sequences in the corresponding positions are underlined.

Table 1.2. From Sakano et al (1980).

DNA sequences around the joining sites indicates that the two recombination mechanisms are different (Davis et al,1980, Sakano et al,1980).

Early et al (1980) have proposed a scheme for the first type of recombination, the VDJ joining. In the germ-line DNA on the 3' side of the V-genes and on the 5' side of the D- and J-genes there are short blocks of seven and nine base pairs in length of conserved sequence. Table I.2 shows that these blocks are arranged in a complementary manner; that is, 3' to the V_H -gene lie first a heptamer then a nanomer, whereas 5' to the J_H -gene lie a complementary nanomer followed by a complementary heptamer. The number of nucleotides separating each nanomer and heptamer is, almost without exception, either 11 or 22. As there are 10 nucleotides in one turn of the Watson-Crick double helix, the spacer sequences correspond to either one or two turns of DNA. The proposed mechanism envisions that a protein or complex of proteins recognizes the conserved blocks and the lengths of the spacer sequences. Joining only occurs between spacers of different length; a one turn spaced set of blocks joins with a set separated by two turns.

A similar process apparently occurs in light chain genes (Sakano et al,1980). Table I.2 shows that the "one turn-two turn" theory could prevent the joining of, for example, V_K and J_λ gene segments, because the spacer lengths between conserved blocks are reversed in λ - and κ -chain genes; V_K is followed by a one turn and J_K preceded by a two turn spacer, while the V_λ spacer is two turns and that of J_λ , one turn.

V_H genes are followed by conserved blocks separated by two turns and J_H genes are preceded by two turn spacers (Matthyssens and Rabbitts,1980). One might predict, therefore, that D-segment genes are both preceded and followed by one turn spacers (Early et al,1980). By the "one turn-two turn"

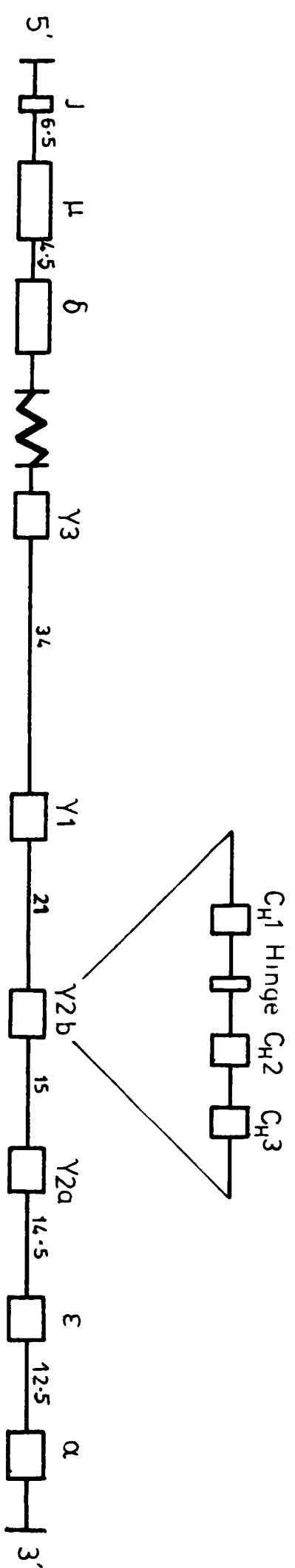


Figure I.8. The sequence of genes encoding the heavy chain C-regions in the mouse germ-line. Distances are in kilobases.

theory, a V-J joining would be prevented in the heavy chain. However, direct evidence in the form of nucleotide sequences surrounding D-segment genes has not yet appeared.

Bernard and Gough (1980) have described a myeloma protein in which a V_H-J_H joining occurs. Analysis of the DNA preceding the J-segment concerned shows that two conserved heptameric sequences occur, separated by distances of 11 and 22 nucleotides from the same nanomer. The first of these heptamers is presumably similar to the prototype heptamer by coincidence. As a result, however, a V_H-J_H recombination is possible without breaking the "one turn-two turn" rule.

c. The Heavy Chain Switch.

Rabbitts et al (1980) used the Southern hybridization technique to examine the nuclear DNA of B cells producing different classes of immunoglobulin. The DNA of an IgG2b producing cell line contains a gene for a $\gamma 2b$ chain but not one for a $\gamma 1$ chain. IgG2a cell lines contain neither $\gamma 2b$ nor $\gamma 1$ genes and an IgA-producing cell shows no detectable genes for any γ chain. This is evidence of the deletion of the C_H -region genes between the complete V-region gene and that of the C-region to which it combines. Furthermore, the sequence of C-region genes along the germ-line DNA can be determined (Cory et al, 1980).

The sequence of commonly expressed heavy-chain constant-region genes (figure I.8) has been elucidated by Shimizu et al (1981). They have also described a conserved sequence of 49 base pairs in length, the S (switch) regions, located 5' to each C-region gene. However, a mechanism for the heavy-chain switch has yet to be proposed.

It is apparent that the constant region rearrangement and deletion events occur after transcription of the chromosomal DNA (Wabl et al, 1980). In contrast, the rearrangement of V-, D- and J-gene segments occurs during

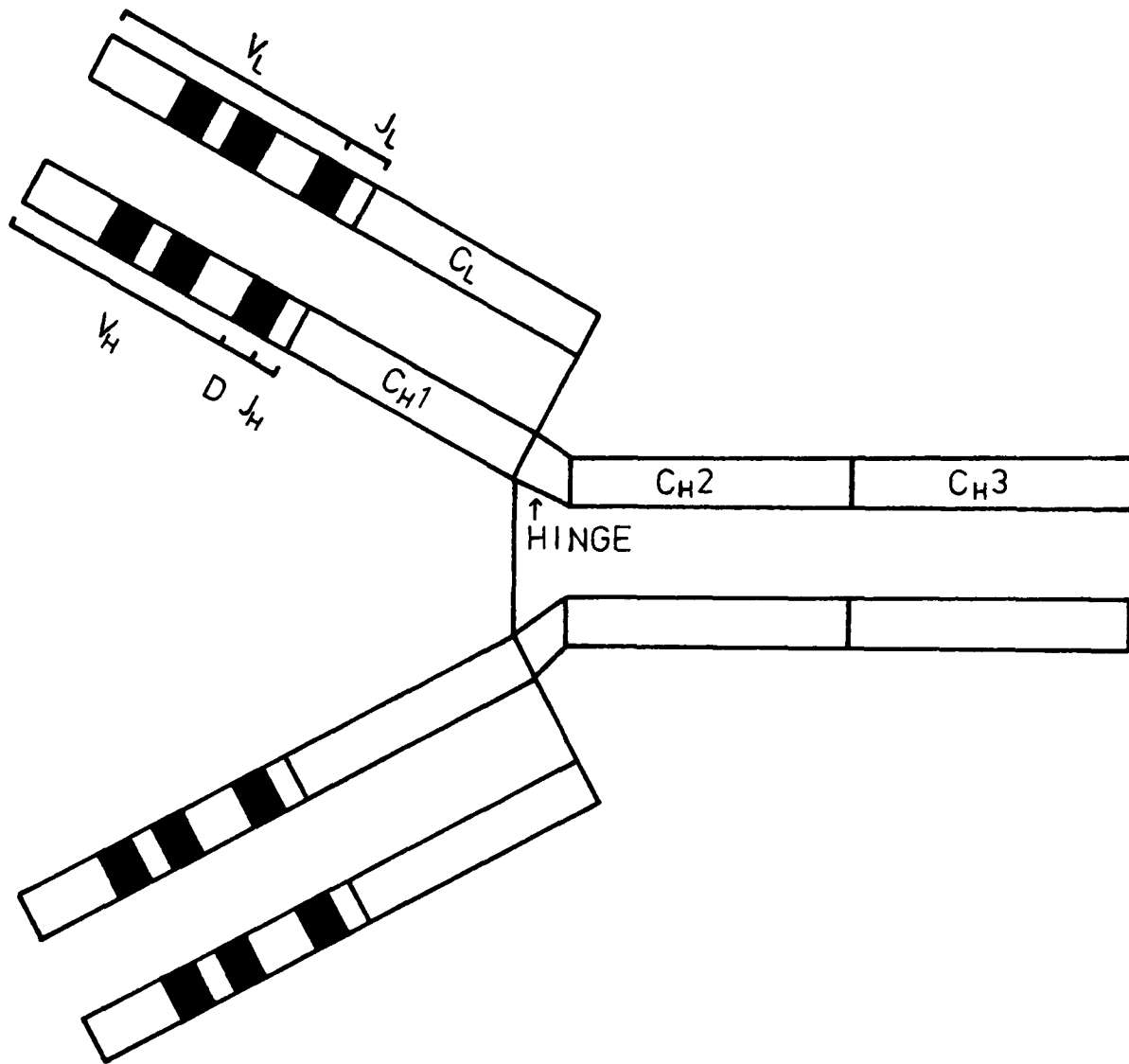


Figure I.9. A representation of an immunoglobulin showing the portions encoded in separate gene segments. After Molgaard (1980).

B cell differentiation and maturation.

d. Conclusion.

Figure I.9 delineates the portions of the immunoglobulin molecule which, by present knowledge, are encoded in separate gene segments. The formation of the gene of a whole immunoglobulin chain occurs in at least two stages, probably by different mechanisms. During B cell differentiation, the cell commits itself to a particular specificity by the recombination of the V-, D- and J-segments. If the cell is stimulated to form a clone, all its progeny will exhibit the same antigen complementarity. In the mature B cell, the complete V-region gene can switch from one C-region gene to another, producing antibodies of different classes. The motivation and mechanism of this event are, as yet, unknown.

The process of the generation of diversity outlined above is advantageous, not least because of its economical use of DNA; a very large number of diverse variable regions can be produced from a relatively small length of DNA. The theory accounts, without the need for somatic mutation, for much of the hypervariability of the third complementarity-determining region, although evidence exists that some somatic mutation is still required to account for all the observed amino-acid sequences (Pech et al, 1981). It is interesting to speculate whether the first two hypervariable regions might be generated in a similar way. Perhaps the V-region genes are further sub-divided into as yet unobserved gene segments which join together at approximately residues 30 and 50 (35 and 55 in heavy chains) by a similar mechanism to V-J joining.

There is some evidence to support this speculation. Wu et al (1979) have reported several regions of closely conserved nucleotide sequence at homologous positions in DNA coding for λ and κ light chains. The regions all occur in the code for the amino-acid sequence responsible for

maintaining the tertiary structure of the folding unit (the framework region) but some are adjacent to hypervariable regions. It is proposed that these conserved sequences might serve as recognition sites for the insertion or recombination of "mini-genes" (Kabat et al, 1978).

Recombination by a mechanism like that described above, which allows flexibility of the joining site, could account for local hypervariability.

A second piece of evidence arises from an abnormal λ chain mRNA reported by Bothwell et al (1981). The myeloma MOPC-315 normally produces a specific IgA with a $\lambda 2$ light chain (see Chapter II). The abnormal mRNA apparently corresponds to the sequence of a $\lambda 1$ light chain from the N-terminal to residue 31 and immediately thereafter proceeds into the C_L region sequence. This might result from an unusual rearrangement of the initial $\lambda 1$ "mini-gene" directly with the C_L gene.

It is clear that an ingenious mechanism for the generation of a large diversity of antibody specificities from a set of germ-line genes has been evolved in vertebrates. The segment of the immunoglobulin in which antigen complementarity is manifested, the Fv domain, is the subject of the rest of this thesis. The following chapter describes a method for the isolation of the Fv fragment from the whole immunoglobulin molecule and the studies which have been carried out on its antigen-binding properties. Subsequent chapters are concerned with the elucidation of the three-dimensional atomic arrangement of the Fv fragment by the method of crystallography.

CHAPTER II

Immunoglobulin M315

The myeloma protein M315 typifies the phenomenon of hapten-binding to immunoglobulins. Dower (1979) has reviewed many of the early experiments that were done to characterize the protein. This chapter is concerned with the methods by which M315 and its fragments are produced and isolated, and the studies that have been carried out on its hapten-combining site.

II.A. Production and Purification of M315

1. The Whole Immunoglobulin

A Balb/c-2 mouse is the result of crossing a Balb/c mouse with one from the C57Bl/6 strain, backcrossing a member of the F1 generation with a Balb/c mouse, choosing an offspring which carries the C57Bl/6 heavy chain linkage group, and repeating the backcross in the same way seven times (Potter and Lieberman, 1967). The MOPC-315 tumour was induced in a Balb/c-2 mouse by giving it three intraperitoneal injections of 0.5 ml of mineral oil (Bayol F), one each at 2, 4 and 6 months of age (Potter and Boyce, 1962). Once induced, the plasmacytoma can be propagated in Balb/c mice by appropriate transfers (Goetzl and Metzger, 1970). A whole myeloma protein, M315, but no Bence Jones protein, is produced by the tumour.

By the original method of Eisen et al (1968), M315 was precipitated from the serum with saturated ammonium sulphate in the presence of buffered saline, and isolated by chromatography on DEAE-sephadex. The fraction with anti-dinitrophenyl (Dnp) activity was mildly reduced and alkylated to release IgA monomers of sedimentation coefficient 6.6S.

An alternative method, based on that of Goetzl and Metzger (1970), was used to isolate the protein in this study (figure II.1)(Aschaffenburg et al, 1979). The mouse serum was first reduced and alkylated and then applied to an

Figure II.1. Method for Isolating Protein M315 and its Fv Fragment

Serum from mouse with myeloma MOPC-315
dialysed against 0.2M tris/HCl buffer pH 8.2.

Add Na₂EDTA to concentration of 0.002M.

Reduction and Alkylation:

- a. Add dithiothreitol to concentration of 0.01M.
Incubate 1 hour at room temperature.
- b. Add solid iodoacetamide to concentration of 0.03M.
Incubate 1/2 hour at room temperature.
- c. Dialyse against 0.05M tris/NaCl buffer pH 7.2.

First Affinity Column

- a. Run protein onto Dnp-sepharose column.
- b. Wash rapidly with 0.05M tris buffer pH 7.2 until
optical density (OD) of effluent at 280nm less than 0.05.
- c. Elute IgA with 0.1M ammonia.
If necessary, lower pH to 8.2 with solid ammonium bicarbonate.
Collect fractions with OD(280)>1.0.
- d. Dialyse against 0.05M sodium acetate/0.15M sodium chloride pH 4.7.

Digestion

- a. Concentrate protein to about 10mg/ml on a PM30 Diaflo.
Add sodium azide to a concentration of 0.005M.
- b. Add solution of 10mg pepsin/1ml 0.005M sodium acetate pH 4.6
to a concentration of 1mg pepsin/50mg IgA.
Digest at 37°C for 3 hours.
- c. Lower pH to 3.5 with acetic acid.
Digest at 37°C for 3 hours.
- d. Raise pH to 7.0 with 2M tris.
Dialyse against 0.2M tris/NaCl pH 7.2.
Spin to remove precipitate.

Isolation of Fv

- a. Concentrate digest to 20ml on PM10 Diaflo.
Apply to sephadex G-75 column.
- b. Elute with 0.05M tris/0.15M NaCl pH 7.2.
- c. Pool Fv fractions as determined by OD(280).

Purification - Second Affinity Column

- a. Apply Fv to Dnp-sepharose column.
- b. Wash with tris buffer as above.
- c. Elute with 0.005M Dnp-glycine in tris/NaCl pH 7.2.
- d. Run fractions down Dowex column.
Pool all with OD(280)>0.05.

Concentrate Fv on PM-10 Diaflo to about 20mg/ml.

Lyophilize.

Check purity by SDS gel electrophoresis.

Yields about 70mg of Fv from 55ml of serum.

110
115*
120
 V_H : Gly-Thr-Thr-Leu-Thr-Val-Ser-Ser-Glu-Ser-Ala
 V_L : Leu-Gly-Gln-Pro-Lys-Ser-Thr-Pro-Thr-Leu-Thr

	<u>Number of Residues to 110</u>			<u>Fv Composition</u> ⁽¹⁾
	<u>in V_H</u>	<u>in V_L</u>	<u>Total</u>	
Ser	12	8	20	19.3
Thr	9	15	24	27.6
Lys	5	3	8	8.6
Pro	4	4	8	9.5
Val	5	9	14	14.9

(1) From Hochman et al (1973).

Figure II.2. Elucidation of the cleavage points of Fv from its amino-acid composition and the sequences of the heavy and light chains. The proposed cleavage points are marked by stars. See figure II.4 for V_H and V_L sequences to position 110.

ϵ -N-dinitrophenyllysine sepharose column, for separation by affinity chromatography. M315 was eluted from the column either with 0.1 M ammonia or with 0.05 M Dnp-glycine. Elution with the former resulted in a high pH, which was immediately lowered to 8.2 with solid ammonium bicarbonate, in some of the protein-containing fractions. The latter treatment might therefore be expected to be less damaging to the protein, although the hapten must be removed by passage down a column of Dowex 1-X8. However, successful crystallizations of the Fv fragment have only been achieved with M315 eluted with ammonia.

2. Digestion and Isolation of Fragments

The first fragmentation of M315 was effected with papain at pH 7.4 (Eisen et al, 1968). As expected, this yielded Fab and Fc fragments in the ratio 2 to 1 (see Chapter I for nomenclature of fragments). Pepsin digestion at pH 4.5 yielded the Fab' fragment, crystals of which were grown from 0.002 M sodium acetate in the presence of hapten (Inbar et al, 1972), but were unsuitable for x-ray analysis.

When the pH was lowered to 3.6, further pepsin digestion released the Fv fragment (molecular weight 25250)(Inbar et al, 1972). In the scheme outlined in figure II.1, the Fv cleavage was accomplished by a two-step pepsin treatment. After three hours at 4.5, the pH was lowered to 3.5 by direct addition of acetic acid, and the digestion was continued for a further three hours. This is a very unusual pattern of cleavage and it is not known why M315 is susceptible in contrast to other immunoglobulins.

The points at which the cleavage occurs in the light and heavy chains were deduced from the amino-acid sequences of V_H and V_L (section B2 below) and the compositions of Fv, V_H and V_L (Hochman et al, 1973). The light chain contains 8 serine residues and the heavy chain 12, in the sequences to position 110. As Fv contains 19 or 20 serine residues (figure II.2), the chains

<u>Compound</u>	(1) <u>Ka (M⁻¹) of</u> <u>M315</u>	<u>Ka (M⁻¹) of anti-</u> <u>Dnp-antibody</u>	(2) <u>References</u>
ε-N-Dnp-L-lysine	3.3 x10 ⁶ (3)	23 x10 ⁶ (3)	1,2
Dnp-glycine	0.08x10 ⁶	-	3
Tnp-glycine	0.13x10 ⁶	-	3
2,4 dinitroaniline	0.25x10 ⁶	0.3x10 ⁶ (3)	3,2
Dnp-2aminoethylphosphate	1.11x10 ⁶	-	4
Dnp-2aminoethyldiphosphate	1.61x10 ⁶	-	4
Dnp-2aminoethyltriphosphate	2.00x10 ⁶	-	4

(1) Hapten + Ig = H-Ig , Ka=[H-Ig]/[H][Ig] at pH 7.0,30°C.

- (2)
1. Dower et al, 1978
 2. van Oss and Grossberg, 1979
 3. Dower and Dwek, 1978
 4. Dower and Dwek, 1979

(3) Measured at 25°C.

Figure II.3. Affinity of M315 for several haptens.
Dnp - 2,4 dinitrophenyl, Tnp - 2,4,6 trinitrophenyl.

are apparently cleaved before the next serine residue, position 115 in V_L and 116 in V_H . If the residues in positions 110 to 120 are considered in turn, it can be deduced that the light chain is cleaved after residue 114 and the heavy chain after 115. This is not a definitive method because of the uncertainty in the number of some amino-acid residues, for example Thr, Lys and Pro, in the Fv fragment. However, chemical determinations of the carboxy-terminal groups have not been made.

Two isolations of other Fv fragments have been published. Sharon and Givol (1976) prepared an Fv fragment of an immunoglobulin similar to M315. The protein, from mouse myeloma XRPC-25, has less anti-Dnp activity than M315 by about a factor of 10 M^{-1} in its association constant (see figure II.3). The Fv was made by an indirect method. The heavy and light chains of the IgA were dissociated, and the light chain cleaved with trypsin at pH 8.2. The variable region was recombined with the heavy chain, and digestion with papain at pH 5.7 yielded an Fv fragment of molecular weight 23400.

Another Fv was obtained directly by a peptic digestion of the Waldenstrom protein (IgM, κ) Hi at pH 4.5 (Kakimoto and Onoue, 1974). The molecular weight of the fragment, 30100, indicates that the point of cleavage was further from the N-terminal than the cleavage point of Fv315. No strongly-binding hapten to protein Hi has been reported.

Some immunoglobulin heavy chains have been cleaved by enzymic digestion to yield the variable region (for example, Roseblatt and Haber, 1978) but light chains have generally been resistant to such treatment. One report (Ehrlich et al, 1979) has been made of a successful V_L isolation after which the half-domain was recombined with the corresponding heavy chain to restore the antigen-binding properties of the intact light chain-heavy chain recombinant.

II.B. The Hapten-Binding Activity of M315

1. The Affinity of M315 for Dnp compounds.

The reason that M315 has been of particular interest in the study of antigen-binding to antibodies, is the very high affinity it shows for small molecules containing 2,4-dinitrophenyl (Dnp) or 2,4,6-trinitrophenyl (Tnp) groups. Eisen et al (1968) reported that the intrinsic association constant of M315, $1.6 \times 10^7 \text{ M}^{-1}$ at 4°C for ϵ -Dnp-L-lysine, was essentially the same as that of antibodies raised against Dnp and Tnp ligands by conventional immunization procedures. The binding affinities of several compounds to M315 and, where known, to anti-Dnp antibodies, are shown in figure II.3.

The hapten-binding activity of M315 was shown to lie entirely in the Fv fragment by Inbar et al (1972). They examined the quenching of tryptophan fluorescence upon addition of hapten to the fractions obtained on the cleavage of Fab. The fragment that showed quenching (the Fv) demonstrated the same effect as the intact Fab. In addition, they determined that the first eight amino-acid residues from the amino terminus of the heavy chain of Fv were identical to those of M315.

Wells et al (1973) prepared anti-idiotypic antibodies to M315 and showed that Fv (labelled with ^{125}I) was as effective as the intact protein in inhibiting the reaction of these antibodies and M315. The binding of the Fv fragment to the antiserum could be inhibited by the presence of a Dnp hapten. This was conclusive evidence that the Fv holds the idiotypic determinants of the immunoglobulin.

Blaser and Eisen (1978) searched the immunoglobulin repertoire of normal mice for the presence of light chains similar to that of M315. They estimated that 1% of free serum immunoglobulins have light chains of the same sub-type (see section B.2 below). Cotner and Eisen (1978) found a

(a) Alignment of V _H Sequences										
	0	2	0	4	0	2	0	4	0	0
McPC-603	1	Glu	-Val	-Lys	-Leu	-Val	-Glu	-Ser	-Gly	-Gly
MOPC-315	1	Asp	-Val	-Gln	-Leu	-Gln	-Glu	-Ser	-Gly	-Pro
New	1	Pca	-Val	-Gln	-Leu	-Pro	-Glu	-Ser	-Gly	-Pro
	0	4	0	0	0	0	0	4	0	4
11	Leu	-Val	-Gln	-Pro	-Gly	-Gly	-Ser	-Leu	-Arg	-Leu
11	Leu	-Val	-Lys	-Pro	-Ser	-Gln	-Ser	-Leu	-Ser	-Leu
11	Leu	-Val	-Ser	-Pro	-Gly	-Glx	-Thr	-Leu	-Ser	-Leu
	0	4	0	4	0	1	4	0	4	0
21	Ser	-Cys	-Ala	-Thr	-Ser	-Gly	-Phe	-Thr	-Phe	-Ser
21	Thr	-Cys	-Ser	-Val	-Thr	-Gly	-Tyr	-Ser	-Ile	-Thr
21	Thr	-Cys	-Thr	-Gly	-Ser	-Thr	-Val	-Ser	-Thr	-Phe
	0	4	2	4	2	4	C	3	C	
31	Asp	-	-Phe	-Tyr	-Met	-Glu	-Trp	-Val	-Arg	-Gln
31	Ser	-Gly	-Tyr	-Phe	-Trp	-Asn	-Trp	-Ile	-Arg	-Gln
31	Ala	-	-Val	-Tyr	-Ile	-Val	-Trp	-Val	-Arg	-Gln
	2	0	0	0	0	C	2	C	4	4
40	Pro	-Pro	-Gly	-Lys	-Arg	-Leu	-Glu	-Trp	-Ile	-Ala
40	Phe	-Pro	-Gly	-Asn	-Lys	-Leu	-Glu	-Trp	-Leu	-Gly
40	Pro	-Pro	-Gly	-Arg	-Gly	-Leu	-Glu	-Trp	-Ile	-Ala
	1	4	2	0	0	0	0	0	2	0
50	Ala	-Ser	-Arg	-Asn	-Lys	-Gly	-Asn	-Lys	-Tyr	-Thr
50	Phe	-Ile	-Lys	-Tyr	-Asp	-Gly	-	-	-	-Ser
50	Tyr	-Val	-Phe	-Tyr	-His	-Gly	-	-	-	-Thr
	2	0	3	1	0	0	4	0	0	1
58b	Thr	-Glu	-Tyr	-Ser	-Ala	-Ser	-Val	-Lys	-Gly	-Arg
57	Asx	-(Tyr, Gly)	Asx	-Pro	-Ser	-Leu	-Lys	-Asn	-Arg	
57	Ser	-Asp	-Thr	-Asp	-Thr	-Pro	-Leu	-Arg	-Ser	-Arg
	4	0	4	0	2	0	0	0	0	3
68	Phe	-Ile	-Val	-Ser	-Arg	-Asp	-Thr	-Ser	-Gln	-Ser
68	Val	-Ser	-Ile	-Thr	-Arg	-Asp	-Thr	-Ser	-Gln	-Asn
68	Val	-Thr	-Met	-Leu	-Val	-Asn	-Thr	-Ser	-Lys	-Asn
	0	4	1	4	0	4	0	0	4	0
78	Ile	-Leu	-Tyr	-Leu	-Gln	-Met	-Asn	-Ala	-Leu	-Arg
78	Gln	-Phe	-Phe	-Leu	-Lys	-Leu	-Asp	-Ser	-Val	-Thr
78	Gln	-Phe	-Ser	-Leu	-Arg	-Leu	-Ser	-Ser	-Val	-Thr
	0	0	2	3	4	C	4	C	4	C
88	Ala	-Glu	-Asp	-Thr	-Ala	-Ile	-Tyr	-Tyr	-Cys	-Ala
88	Thr	-Glx	-Asx	-Thr	-Ala	-Thr	-Tyr	-Tyr	-Cys	-Ala
88	Ala	-Ala	-Asp	-Thr	-Ala	-Val	-Tyr	-Tyr	-Cys	-Ala
	3	3	2	C	0	0	C	C	C	C
98	Arg	-Asn	-Tyr	-Tyr	-Gly	-Ser	-Thr	-Trp	-Tyr	-Phe
98	Gly	-Asp	-Asn	-Asp	-His	-	-	-Leu	-Tyr	-Phe
98	Arg	-Asx	-Leu	-Ile	-Ala	-	-	-Gly	-Cys	-Ile
	1	4	C	4	0	4	2	0	4	0
106	Asp	-Val	-Trp	-Gly	-Ala	-Gly	-Thr	-Thr	-Val	-Thr
106	Asp	-Tyr	-Trp	-Gly	-Gln	-Gly	-Thr	-Thr	-Leu	-Thr
106	Asx	-Val	-Trp	-Gly	-Gln	-Gly	-Ser	-Leu	-Val	-Thr
	4	0	0							
116	Val	-Ser	-Ser							
116	Val	-Ser	-Ser							
116	Val	-Ser	-Ser							

(b) Alignment of V _L Sequences										
	0	3	0	4	0	4	1	0	0	0
McPC-603	1	Asp	-Ile	-Val	-Met	-Thr	-Gln	-Ser	-Pro	-Ser
REI	1	Asp	-Ile	-Gln	-Met	-Thr	-Gln	-Ser	-Pro	-Ser
MOPC-315	1	Pca	-Ala	-Val	-Val	-Thr	-Glu	-Glu	-	-Ser
Mcg	1	Pca	-Ser	-Ala	-Leu	-Thr	-Gln	-Pro	-Pro	-
New	1	Pca	-Ser	-Val	-Leu	-Thr	-Gln	-Pro	-Pro	-
	4	0	4	0	0	0	0	0	4	0
11	Leu	-Ser	-Val	-Ser	-Ala	-Gly	-Glu	-Arg	-Val	-Thr
11	Leu	-Ser	-Ala	-Ser	-Val	-Gly	-Asp	-Arg	-Val	-Thr
10	Leu	-Thr	-Thr	-Ser	-Pro	-Gly	-Gly	-Thr	-(Val, Ile)	
10	Ala	-Ser	-Gly	-Ser	-Leu	-Gly	-Gln	-Ser	-Val	-Thr
10	Val	-Ser	-Gly	-Ala	-Pro	-Gly	-Gln	-Arg	-Val	-Thr
	4	0	4	0	4	0	0	0	3	0
21	Met	-Ser	-Cys	-Lys	-Ser	-Ser	-Glx	-Ser	-Leu	-Leu
21	Ile	-Thr	-Cys	-Gln	-Ala	-Ser	-Gln	-	-	-
20	Leu	-Thr	-Cys	-Arg	-Ser	-Ser	-Thr	-Gly	-Ala	-Val
20	Ile	-Ser	-Cys	-Thr	-Gly	-Thr	-Ser	-Ser	-Asp	-Val
20	Ile	-Ser	-Cys	-Thr	-Gly	-Ser	-Ser	-Ser	-Asn	-Ile
	0	0	C	C	C	1	C	3	4	C
27	Asx	-Ser	-Gly	-Asx	-Glx	-Lys	-Asx	-Phe	-Leu	-Ala
28	-	-	-	-Asp	-Ile	-Ile	-Lys	-Tyr	-Leu	-Asn
30	-	-	-	-Thr	-Thr	-Ser	-Ser	-Asn	-Tyr	-Ala
30	-	-	-	-Gly	-Gly	-Tyr	-Asn	-Tyr	-Val	-Ser
27c	-	-	-	-Gly	-Ala	-Gly	-Asn	-His	-Val	-Lys
	4	C	3	C	0	0	0	1	C	C
35	Trp	-Tyr	-Glx	-(Glx)	-Lys	-Pro	-Gly	-Glx	-Pro	-Pro
35	Trp	-Tyr	-Gln	-Gln	-Thr	-Pro	-Gly	-Lys	-Ala	-Pro
37	Trp	-Ile	-Glx	-Glx	-Lys	-Pro	-Asx	-His	-Leu	-Phe
37	Trp	-Tyr	-Gln	-Gln	-His	-Ala	-Gly	-Lys	-Ala	-Pro
34	Trp	-Tyr	-Gln	-Gln	-Leu	-Pro	-Gly	-Thr	-Ala	-Pro
	0	C	4	4	C	1	1	0	1	0
45	Lys	-Leu	-Leu	-Ile	-Tyr	-X	-X	-X	-X	-X
45	Lys	-Leu	-Leu	-Ile	-Tyr	-Glu	-Ala	-Ser	-Asn	-Leu
47	Thr	-Gly	-Leu	-Ile	-Gly	-Gly	-Thr	-Ser	-Asp	-Arg
47	Lys	-Val	-Ile	-Ile	-Tyr	-Glu	-Val	-Asn	-Lys	-Arg
44	Lys	-Leu	-Leu	-Ile	-Phe	-His	-Asn	-Asn	-Ala	-Arg
	0	0	0	4	1	0	1	4	0	4
55	X	-X	-Gly	-Val	-Pro	-Ala	-Arg	-Phe	-Ser	-Gly
55	Gln	-Ala	-Gly	-Val	-Pro	-Ser	-Arg	-Phe	-Ser	-Gly
57	Ala	-Pro	-Gly	-Val	-Pro	-Val	-Arg	-Phe	-Ser	-Gly
57	Pro	-Ser	-Gly	-Val	-Pro	-Asp	-Arg	-Phe	-Ser	-Gly
61	-	-	-	-	-	-	-	-Phe	-Ser	-Val
	0	4	0	0	2	0	4	0	4	0
65	Ser	-Gly	-Ser	-Arg	-Thr	-Asp	-Phe	-Thr	-Leu	-Thr
65	Ser	-Gly	-Ser	-Gly	-Thr	-Asp	-Tyr	-Thr	-Phe	-Thr
67	Ser	-Leu	-Ile	-Gly	-Asp	-Lys	-Ala	-Ala	-Leu	-Thr
67	Ser	-Lys	-Ser	-Gly	-Asn	-Thr	-Ala	-Ser	-Leu	-Thr
64	Ser	-Lys	-Ser	-Gly	-Ser	-Ser	-Ala	-Thr	-Leu	-Ala
	4	0	0	2	0	0	1	2	4	
75	Ile	-Asx	-Pro	-Val	-Glx	-Ala	-Asx	-Asp	-Val	-Ala
75	Ile	-Ser	-Ser	-Leu	-Gln	-Pro	-Glu	-Asp	-Ile	-Ala
77	Ile	-Thr	-Gly	-Ala	-Glx	-Thr	-Glx	-Asp	-Asp	-Ala
77	Val	-Ser	-Gly	-Leu	-Gln	-Ala	-Glu	-Asp	-Glu	-Ala
74	Ile	-Thr	-Gly	-Leu	-Gln	-Ala	-Glu	-Asp	-Glu	-Ala
	0	4	C	4	C	4	C	2	0	0
85	Thr	-Tyr	-Phe	-Cys	-X	-X	-X	-X	-X	-X
85	Thr	-Tyr	-Tyr	-Cys	-Gln	-Gln	-Tyr	-Gln	-Ser	-Leu
87	Met	-Tyr	-Phe	-Cys	-Ala	-Leu	-Trp	-Phe	-Arg	-Asx
87	Asp	-Tyr	-Tyr	-Cys	-Ser	-Ser	-Tyr	-Glu	-Gly	-Ser
84	Asp	-Tyr	-Tyr	-Cys	-Gln	-Ser	-Tyr	-Asp	-Arg	-Ser
	0	C	3	C	4	C	0	4	0	
95	-X	-X	-X	-Phe	-Gly	-Gly	-Gly	-Thr	-Lys	
95	-Pro	-Tyr	-Thr	-Phe	-Gly	-Gln	-Gly	-Thr	-Lys	
97	-His	-Phe	-Val	-Phe	-Gly	-Gly	-Gly	-Thr	-Lys	
97	Asp	-Asn	-Phe	-Val	-Phe	-Gly	-Thr	-Gly	-Thr	-Lys
94	-Leu	-Arg	-Val	-Phe	-Gly	-Gly	-Gly	-Thr	-Lys	
	4	0	3	0	0					
104	Leu	-Glu	-Ile	-Lys	-Arg					
104	Leu	-Gln	-Ile	-Thr	-Gly					
106	Val	-Thr	-Val	-Leu	-Gly					
107	Val	-Thr	-Val	-Leu	-Gly					
105	Leu	-Thr	-Val	-Leu	-Arg					

Figure II.4. Amino-acid sequences of the (a) heavy and (b) light chain variable regions of M315 aligned with those of other immunoglobulins. The numbers above the M603 sequence signify: (0)completely exposed to solvent, (1)mainly exposed, (2)partly buried, (3)mainly buried, (4)completely buried, (C)in contact with the homologous half-domain.
From Padlan et al, 1976.

higher fraction of M315-like light chains in Balb/c strain mice than in those of most other strains. They also found that this fraction increased considerably on immunization with a Dnp ligand. Thus, cells producing immunoglobulins very similar to M315 are preferentially stimulated on exposure to Dnp. This is further indication that M315 forms a true hapten-antibody combination with Dnp compounds.

2. Sequence Studies on M315.

The amino-acid sequences of the light (Schulenberg et al,1971) and heavy chains (Francis et al,1974) of the Fv domain of M315 are shown in figure II.4. The protein is of the IgA1 class. The sequence of the light chain showed enough differences from the λ -type chains observed previously for it to be given a distinctive sub-type, λ 2. The sequence of the entire light chain (Dugan et al,1973) showed that λ 2 differs from the more common λ chains (λ 1) at 29 positions in the constant region and 7 in the framework of the variable region.

Tonegawa et al (1978) determined the nucleotide sequence of the mouse germ-line gene coding for the λ 2 variable region. They confirmed six framework substitutions from λ 1 (amino-acid positions 16, 19, 62, 71, 85 and 87). The seventh framework difference (Ile 38 in λ 2) does not appear in the DNA code (Val). This was interpreted as a mutation in the expression of the gene. There are also two examples of codons within the hypervariable regions of M315 (amino-acid positions 55 and 94-96) which do not agree with the λ 2 sequence. This could be due to somatic mutation or to the joining of gene segments as discussed in Chapter I.

A precursor sequence of 19 amino-acid residues in length is cleaved from the amino terminal of the light chain gene product. Most of the DNA for this sequence is separated from the λ 2 gene by 93 base pairs of untranslated DNA (Tonegawa et al,1978). The heavy chain amino terminal

is preceded by an 18 residue leader sequence which is similar only in its hydrophobic nature to the light chain leader (Jilka and Petska,1979).

3. Studies on the Structure of the Hapten-Combining Site.

The antigen-binding site of an antibody can be regarded as being moulded on the framework of the folding unit by the six hypervariable loops (see Chapter I). The prediction of the tertiary structures of proteins ab initio from their primary sequences has been an intensively studied field in recent years, but has been of very limited use to the detail required for examining antigen-antibody interactions. However, once several immunoglobulins had been observed by x-ray crystallography, it was hoped that the short hypervariable sequences of a molecule whose structure was unknown could be inserted into a standard framework to predict the structure of the combining site. The variable domain of M315 was submitted to this type of model building by Padlan et al (1976).

The model was based on the crystal structures of two Fab fragments, that of a mouse IgA with a κ light chain, M603 (Segal et al,1974) and that of a human IgG with a λ 1 light chain, (Poljak et al,1974) (see figure II.4 for sequence alignments with M315). The framework of the variable regions of M315 was built using M603 as the standard folding unit. Each hypervariable segment of M315 was added in a configuration similar to the corresponding segment of either New or M603, depending on which had the closest sequence homology to M315. The model was adjusted to account for the known interaction of the Dnp hapten with Trp 93L^{*} (Eisen et al,1968) and the availability of Tyr 33L and Lys 52H for affinity labeling (Haimovich et al,1972).

R.A.Dwek and his co-workers used the model described above as a

* A letter following an amino-acid position indicates that it is in the heavy (H) or light (L) chain.

starting point for a series of experiments to investigate the local interactions made by the hapten with the side chains of the combining site. Using magnetic resonance techniques, Dower et al (1977) examined the effect of the aromatic Dnp ring on the surrounding aromatic residues. The geometry of the combining site could be defined on the basis of ring current calculations. The Dnp group forms a stacking interaction with Trp 93L and lies in an "aromatic box" made up of Trp 93L, Phe 34H and Tyr 34L. The stability of the complex is increased by hydrogen bonds between the NO₂ groups of Dnp and Asp 36L and Tyr 34L.

Sutton et al (1977) used a series of nitroxide spin label compounds of Dnp to determine the dimensions of the combining site. The box is about 12Å deep and has an entrance of 8Å by 11Å. The Dnp ring appears to be held rigidly in the site. A positively charged amino-acid residue, probably Arg 95L, sits at the entrance to the site. There appears to be a binding site for lanthanides, particularly gadolinium (Gd) (III) about 6Å from the Dnp ring, as the electron spin resonance signal of the spin label is quenched on the binding of Gd³⁺.

There appears to be no simple relationship between the size of the site and its binding affinity. Two mouse myeloma IgA's that bind Dnp haptens to a lesser affinity than M315 have combining sites of similar depth but one (X25) has a smaller and the other (M460) a larger entrance than M315 (Willan et al, 1977). Both M460 and X25 have lanthanide binding sites which do not appear to affect hapten binding, but only X25 has a positively charged residue at the entrance to the combining site. However, X25 has no residue corresponding to Tyr 34L, which, in M315, interacts with one of the NO₂ groups on the hapten. There is, therefore, a smaller effect on binding to X25 than to M315 (Hardy and Richards, 1978).

The three histidine residues in the Fv fragment of M315 were assigned

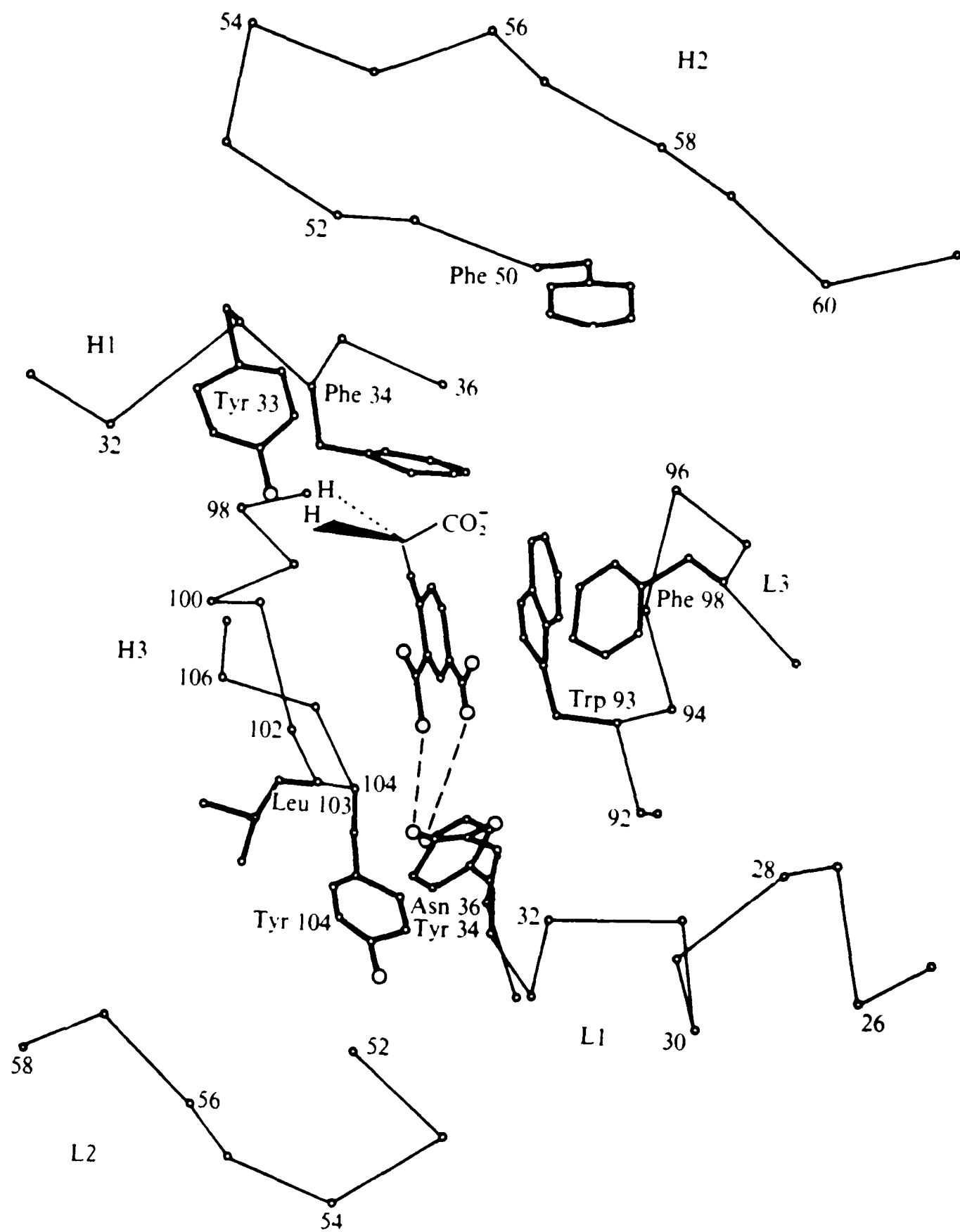


Figure II.5. The refined combining site of M315 with a dinitrophenyl ring in the bound position. The hypervariable loops are abbreviated L (light chain) and H (heavy chain). The side chains of only the residues involved in the binding site are shown.
From Dwek et al, 1977.

pKa values by use of high resolution nuclear magnetic resonance (Wain-Hobson et al,1977). Two of these residues, His 102H and His 97L, are in hypervariable regions and, therefore, lie near the hapten-combining site. It was shown that His 102H is on the surface of the protein bordering the combining site, while His 97L is further away from the site in an hydrophobic environment.

In addition, the binding of several Dnp and Tnp haptens with PO_3^{2-} groups (for example, a series of oligophosphate derivatives of the form N-Dnp-2 aminoethyl (phosphate)_n (n=1 to 3)) was observed using ^{31}P nuclear magnetic resonance (Dower and Dwek,1979). The results could not distinguish between Arg 95L and Lys 52H as the positive charge at the entrance to the combining site. However, Arg 95L was suggested as the more likely, a conclusion supported by the chemical modification studies (Klostergaard et al,1977).

On the basis of these experiments, Dwek et al (1977) presented a refined combining site of M315, a representation of which is shown in figure II.5. The properties of hapten binding appear to be identical in the Fv and Fab fragments and in the intact IgA monomer (Morris et al,1978). The binding of Tnp derivatives (Dower et al,1978) and chlorinated phenyl-glycines and anilines (Gettins et al,1978) to M315 is very similar to that of Dnp compounds; stacking interactions with Trp 93L are again formed.

Many of the results outlined above are based on the assumption that the Dnp rings of different haptens are bound to the M315 combining site rigidly and in the same way. This assumption may not be strictly valid. Examination by Raman spectroscopy of the binding of several Dnp derivatives to M315 and M460 (Kumar et al,1978, Gettins et al,1981) indicated that, for example, the binding of Dnp-L-lysine shows different effects from that of Dnp-NHCH₃. This finding was interpreted to mean that at the level of

atomic contacts with neighbouring residues, there may not be one unique scheme for the binding of the Dnp moiety.

Finally, a prediction of the structure of the Fv domain has been made by a method different from that described above (Stanford and Wu, 1981). Again a framework was built, this time based on the structure of the IgG New. The hypervariable loops were fitted to the framework and their structure was established by predicting ϕ and ψ angles for short peptides. Three residues at a time were considered. Possible choices of ϕ and ψ for a tripeptide were made from a library of tripeptides found in proteins made up primarily of β -sheets. Most of the one to two thousand structures which could be generated for each loop were eliminated by the restrictions imposed by the framework and by steric hindrance. For each of the remaining structures, about five, each ϕ and ψ was varied in a range of plus or minus 30° to minimize the sum of the squared distances between the predicted C α positions of the tripeptide and the corresponding C α positions in New. In general, the experimental observations described earlier could be accommodated by this model. However, four of the six hypervariable loops were significantly different from those of the model of Padlan et al (1976).

The crystal structure of the Fv fragment of M315 will be useful for several reasons. It will add to the evidence for the conservation (or non-conservation) of the framework of immunoglobulin folding units. It will demonstrate whether one of the proposed models of the combining site is accurate and, therefore, whether it is possible to predict the structure of an antibody combining site of known amino-acid sequence, with a prototype framework. In addition, it will provide a test for the theory on which the n.m.r. and e.s.r. distance calculations described above, are based. In combination with hapten-binding studies, it may disclose minor

conformational changes of areas outside the immediate combining site when a hapten is bound, and provide a clue to how a signal can be sent over a long stretch of protein to trigger the complement cascade, for example, in the Fc region of an antibody. It will provide an insight into how the residues at the joining sites of the gene segments in which the V-regions are encoded contribute to the combining site. Finally, an examination can be made of the idiotype (the antigenic properties of the antibody). Anti-idiotypic antibodies are presumably specific only to the variable residues in the V-region of an immunoglobulin. The Fv structure may allow consideration of the question of how the amino-acid residues that make up the specific antigen-combining site participate in forming the idiotype of the antibody.

The final chapters of this thesis describe the structure determination to $6\text{\AA}$ resolution, the extension of the resolution to $4.5\text{\AA}$, the collection of data which will further improve the resolution, and the initiation of studies of the complex formed by Fv with a Dnp hapten.

CHAPTER III

Protein Structure Determination by X-Ray Diffraction

The objective of this chapter is to describe in very simple terms the theory of x-ray crystallography. There have been numerous reviews of protein crystallography (for example, Phillips, 1966, Blundell and Johnson, 1976 and Sherwood, 1976) and this chapter is in no way meant to be as comprehensive or detailed as these.

III.A. Diffraction from a Crystal

1. The Protein Crystal.

A crystal can be thought of as being made up of two components. The first is one or more protein molecules arranged in a unit cell. The second is a three dimensional lattice at every point of which the unit cell is repeated. The diffraction pattern of an incident beam of x-rays is contributed to by both these constituents.

a. The Molecular Transform.

X-rays are diffracted by electrons. The intensity of x-ray diffraction, F_M , is mathematically related to the electron density, ρ , at every point, $\bar{x}$, in the molecule. The Fourier transform of a function $f(x')$ at a point $\bar{x}^*$ is (Ramachandran and Srinasan, 1970),

$$T(\bar{x}^*) = \int_{-\infty}^{\infty} f(\bar{x}') \exp (2\pi i \bar{x}' \cdot \bar{x}^*) dx' \quad (A1).$$

The Fourier integral theorem defines a class of functions $f(x)$, where

$$f(\bar{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\bar{x}') \exp [2\pi i (\bar{x}' - \bar{x}) \cdot \bar{x}^*] dx' dx^* \quad (A2).$$

By substituting (A1) into (A2), we find that

$$f(\bar{x}) = \int_{-\infty}^{\infty} T(\bar{x}^*) \exp (-2\pi i \bar{x} \cdot \bar{x}^*) d\bar{x}^* \quad (A3).$$

This is the inverse Fourier transform of T , differing from the Fourier transform in the sign of the exponential. Expressions (A1) and (A2) describe the relationship of F_M and ρ , where

$$F_M(\bar{x}^*) = \int_V \rho(\bar{x}) \exp (2\pi i \bar{x} \cdot \bar{x}^*) d\bar{x} \quad (A4),$$

the integral being over the volume of the molecule. As the electron density is continuous throughout the molecule, the molecular transform is continuous as well.

b. The Diffraction of X-Rays by a Lattice.

Let us describe a three-dimensional lattice in terms of three axes $\bar{a}$, $\bar{b}$ and $\bar{c}$. A general lattice point at the end of a vector $\bar{r}$ from the origin can be defined as $n_1 \bar{a} + n_2 \bar{b} + n_3 \bar{c}$, where n_1 , n_2 and n_3 are integers. The Fourier transform at $\bar{r}^*$ of the lattice can be written,

$$\begin{aligned} F_L(\bar{r}^*) &= \sum_{n_1, n_2, n_3} V_L \exp (2\pi i \bar{r}^* \cdot \bar{r}) \\ &= \sum_{n_1, n_2, n_3} V_L \exp [2\pi i \bar{r}^* \cdot (n_1 \bar{a} + n_2 \bar{b} + n_3 \bar{c})] \end{aligned} \quad (A5),$$

where V_L is the volume of the unit cell ($|abc|$). As n_1 , n_2 and n_3 are integers, the integrals defined above are replaced by sums over all values of n_1 , n_2 and n_3 .

Equation (A5) can be subdivided into factors of the form

$$\sum_{n_1} \exp (2\pi i n_1 \bar{r}^* \cdot \bar{a}) \text{ etc.,}$$

which also have values only at discrete points. Therefore, the result of (A5) is another three dimensional lattice which is the subject of the next

section.

First, let us examine a mathematical expression of the crystal. The convolution of two functions $g(x)$ and $h(x)$ is

$$C(x) = g(x) * h(x) = \int_{-\infty}^{\infty} g(X) h(x-X) dX.$$

Suppose g is a lattice of value 1 at discrete values of X . $C(x)$ is then $h(x)$ with its origin displaced to each discrete X . If we imagine that h is the electron density contained in one unit cell, C represents the crystal. Now, let us calculate the Fourier transform of $C(x)$ at x^* ,

$$\int_{-\infty}^{\infty} C(x) \exp(2\pi i x x^*) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(X) h(x-X) dX [\exp(2\pi i x x^*)] dx.$$

Next, let $(x-X) = y$. Hence,

$$T_C(x^*) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(X) h(y) \exp[2\pi i x^*(X+y)] dX dy$$

$$= \int_{-\infty}^{\infty} g(X) \exp(2\pi i x^* X) dX \int_{-\infty}^{\infty} h(y) \exp(2\pi i x^* y) dy$$

$$T[g(x)*h(x)] = T_g(x^*) T_h(x^*).$$

Therefore, the Fourier transform of the convolution of two functions is the product of their Fourier transforms. In terms of the crystal, then, the continuous function $F_M(\bar{x}^*)$ is evaluated only at a point where $F_L(\bar{x}^*)$ has a value. In other words, the molecular transform is sampled by the diffraction lattice.

Let us now consider briefly the nature of the diffraction lattice.

c. The Reciprocal Lattice.

If we describe the diffraction lattice by a coordinate system with axes $\bar{a}^*$, $\bar{b}^*$ and $\bar{c}^*$, the general position $\bar{r}^*$ can be expressed as,

$$\bar{r}^* = h\bar{a}^* + k\bar{b}^* + l\bar{c}^* \quad (\text{A6}).$$

Since these axes are arbitrary, we can conveniently require that,

$$\bar{a}^* \cdot \bar{a} = \bar{b}^* \cdot \bar{b} = \bar{c}^* \cdot \bar{c} = 1 \quad (\text{A7a}),$$

$$\text{while,} \quad \bar{a}^* \cdot \bar{b} = \bar{a}^* \cdot \bar{c} = \bar{b}^* \cdot \bar{c} = 0 \quad (\text{A7b}).$$

Therefore, $\bar{a}^*$ is perpendicular to both $\bar{b}$ and $\bar{c}$,

$$\bar{a}^* = s(\bar{b} \times \bar{c}) \quad (\text{A8}).$$

To find the scalar quantity s , we can combine equations (A7a) and (A8),

$$\bar{a}^* \cdot \bar{a} = s(\bar{a} \cdot \bar{b} \times \bar{c}) = 1$$

$$s = (\bar{a} \cdot \bar{b} \times \bar{c})^{-1}.$$

$$\text{So,} \quad \bar{a}^* = \frac{\bar{b} \times \bar{c}}{\bar{a} \cdot \bar{b} \times \bar{c}}$$

and analogously,

$$\bar{b}^* = \frac{\bar{c} \times \bar{a}}{\bar{b} \cdot \bar{c} \times \bar{a}} \quad \text{and} \quad \bar{c}^* = \frac{\bar{a} \times \bar{b}}{\bar{c} \cdot \bar{a} \times \bar{b}} \quad (\text{A9}).$$

Clearly, the direction of each expression (A9) is determined by the numerator. The magnitude of, for example, $\bar{a}^*$ is $(|\bar{a}| \cos \gamma^*)^{-1}$, where γ^* is the angle $\bar{a}$ makes with $\bar{a}^*$ (or $\bar{b} \times \bar{c}$, see figure III.1). It is clear from the figure that γ , the angle $\bar{a}$ makes with the $\bar{bc}$ plane, is $90 + \gamma^*$.

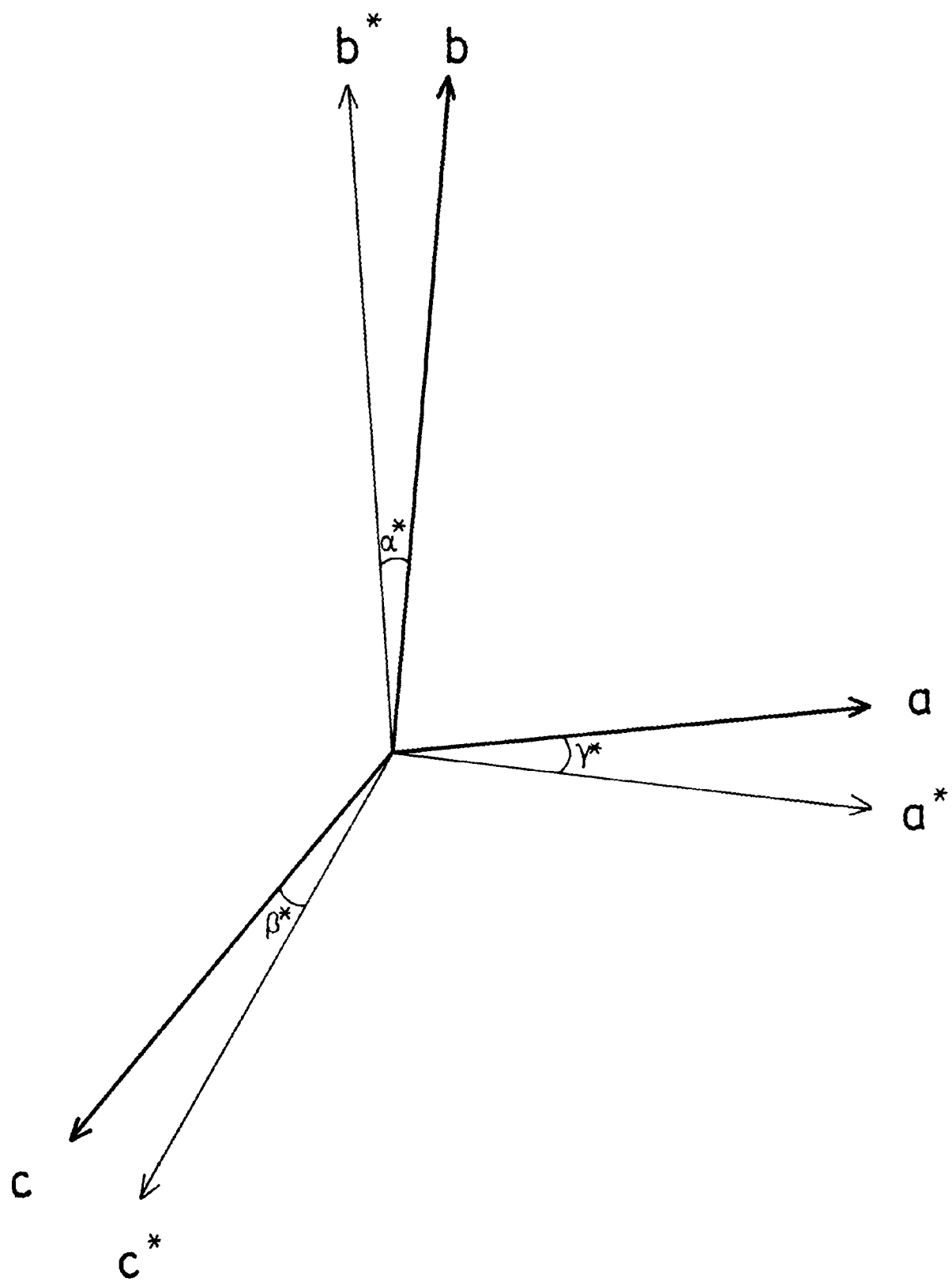


Figure III.1. The relationship between the real and reciprocal axes.

So, the definition of the diffraction lattice in terms of the reciprocal lattice is,

$$\begin{aligned} |a^*| &= (|a|\sin\gamma)^{-1}, \\ |b^*| &= (|b|\sin\alpha)^{-1}, \\ |c^*| &= (|c|\sin\beta)^{-1} \end{aligned} \quad (A10).$$

The inverse nature of these terms has led to the diffraction lattice being called the reciprocal lattice.

d. The Diffraction from a Crystal.

The vector $\bar{x}$ (equation A4) described relative to the same co-ordinate system as $\bar{r}$ (A5), is,

$$\bar{x} = x_1 \bar{a} + x_2 \bar{b} + x_3 \bar{c}.$$

Performing the convolution, we find the general position for an atom is,

$$\bar{r}_a = x\bar{a} + y\bar{b} + z\bar{c},$$

where $x = x_1 + n_1$, $y = x_2 + n_2$ and $z = x_3 + n_3$.

By applying the convolution theorem, we can evaluate the diffraction at a point r^* (A6) in reciprocal space,

$$\begin{aligned} \bar{F}(hkl) &= \int_V \rho(xyz) \exp 2\pi i \bar{x} \cdot \bar{r}^* dv \cdot \int V_L \exp 2\pi i \bar{r} \cdot \bar{r}^* dr \\ &= V_L \int_V \rho(xyz) \exp 2\pi i \bar{r}^* \cdot (\bar{x} + \bar{r}) dv \\ &= V_L \int_V \rho(xyz) \exp 2\pi i \bar{r}^* \cdot \bar{r}_a dv \end{aligned} \quad (A11).$$

The product $\bar{r}^* \cdot \bar{r}_a$ is easily evaluated,

$$\bar{r}^* \cdot \bar{r}_a = (h\bar{a}^* + k\bar{b}^* + l\bar{c}^*) \cdot (x\bar{a} + y\bar{b} + z\bar{c}).$$

The first term of the product is,

$$hx(\bar{a}^* \cdot \bar{a}) + hy(\bar{a}^* \cdot \bar{b}) + hz(\bar{a}^* \cdot \bar{c}).$$

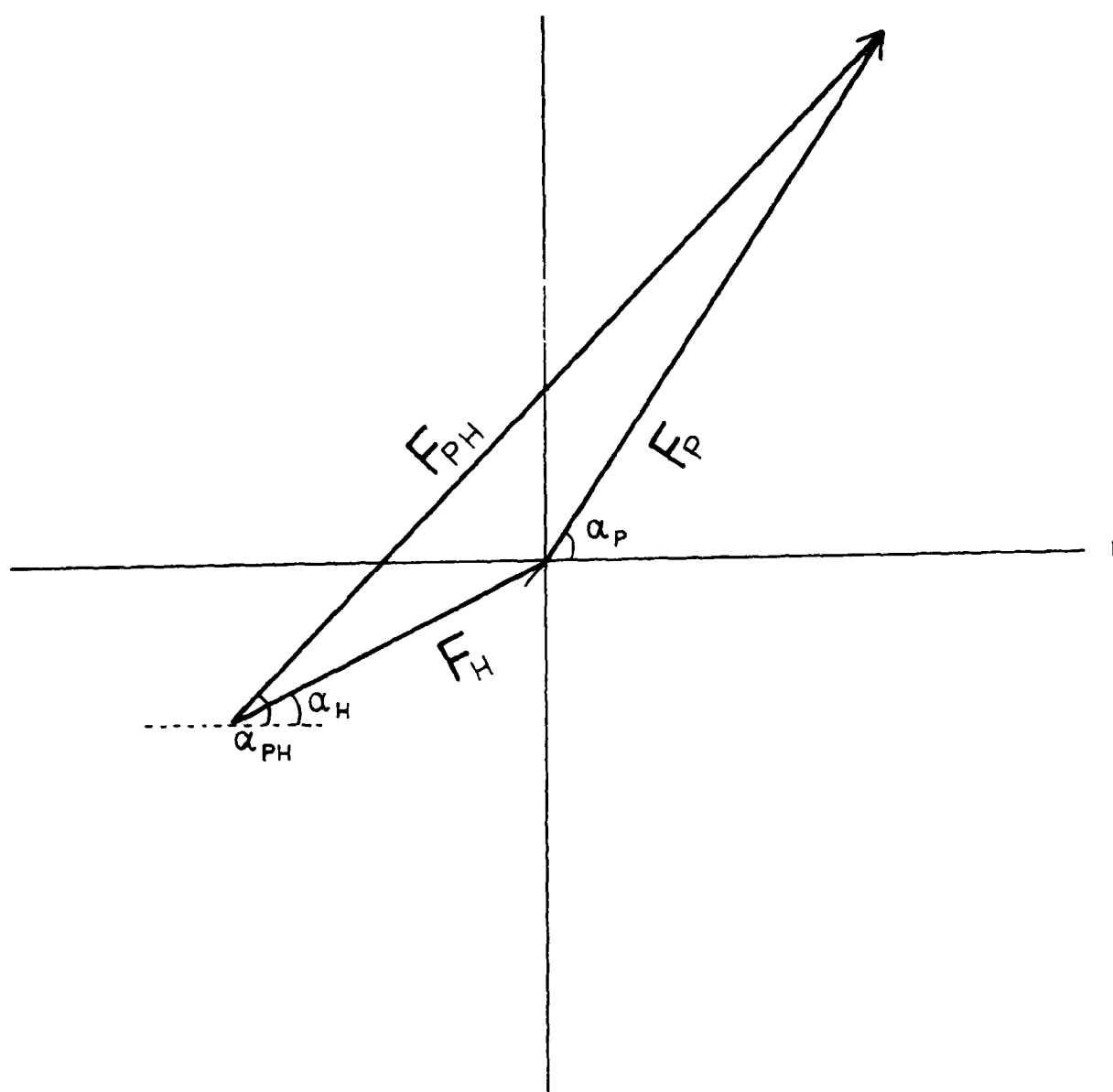


Figure III.2. Single Isomorphous Replacement. An Argand diagram showing the complex structure factors of the protein, the heavy atom, and the protein-heavy atom complex.

By conditions (A7), this reduces to hx . By similar reasoning with the other terms,

$$\bar{r}^* \cdot \bar{r}_a = hx + ky + lz$$

and the structure factor is,

$$\bar{F}(hkl) = V_L \int_V \rho(xyz) \exp [2\pi i(hx+ky+lz)] dv \quad (A12),$$

the integral being over all points (xyz) (all atoms) in the molecule.

The goal of the crystallographic investigation is to find the electron density at (xyz) from the diffraction pattern. By Fourier inversion,

$$\rho(xyz) = \frac{1}{V_L} \sum_{hkl} \bar{F}(hkl) \exp [-2\pi i(hx+ky+lz)] \quad (A13).$$

The complex structure factor $\bar{F}$ can be expressed as,

$$\bar{F}(hkl) = F_s(hkl) \exp i\alpha(hkl)$$

where F_s is the structure factor amplitude and α the phase of a reflection point (hkl) .

The measurable intensity, I , is proportional to the product of $\bar{F}(hkl)$ and $\bar{F}^*(hkl)$, where $\bar{F}^*$ is the complex conjugate of $\bar{F}$. So,

$$\begin{aligned} I(hkl) &\propto F_s \exp i\alpha \cdot F_s \exp -i\alpha \\ &= kF_s(hkl)^2 \end{aligned} \quad (A14).$$

The constant of proportionality depends on the method of measurement (section D).

Thus, the intensity yields the structure factor amplitude, but not the phase, so all the required elements of (A13) are not known. The determination of the phase is the major problem in a crystal structure analysis.

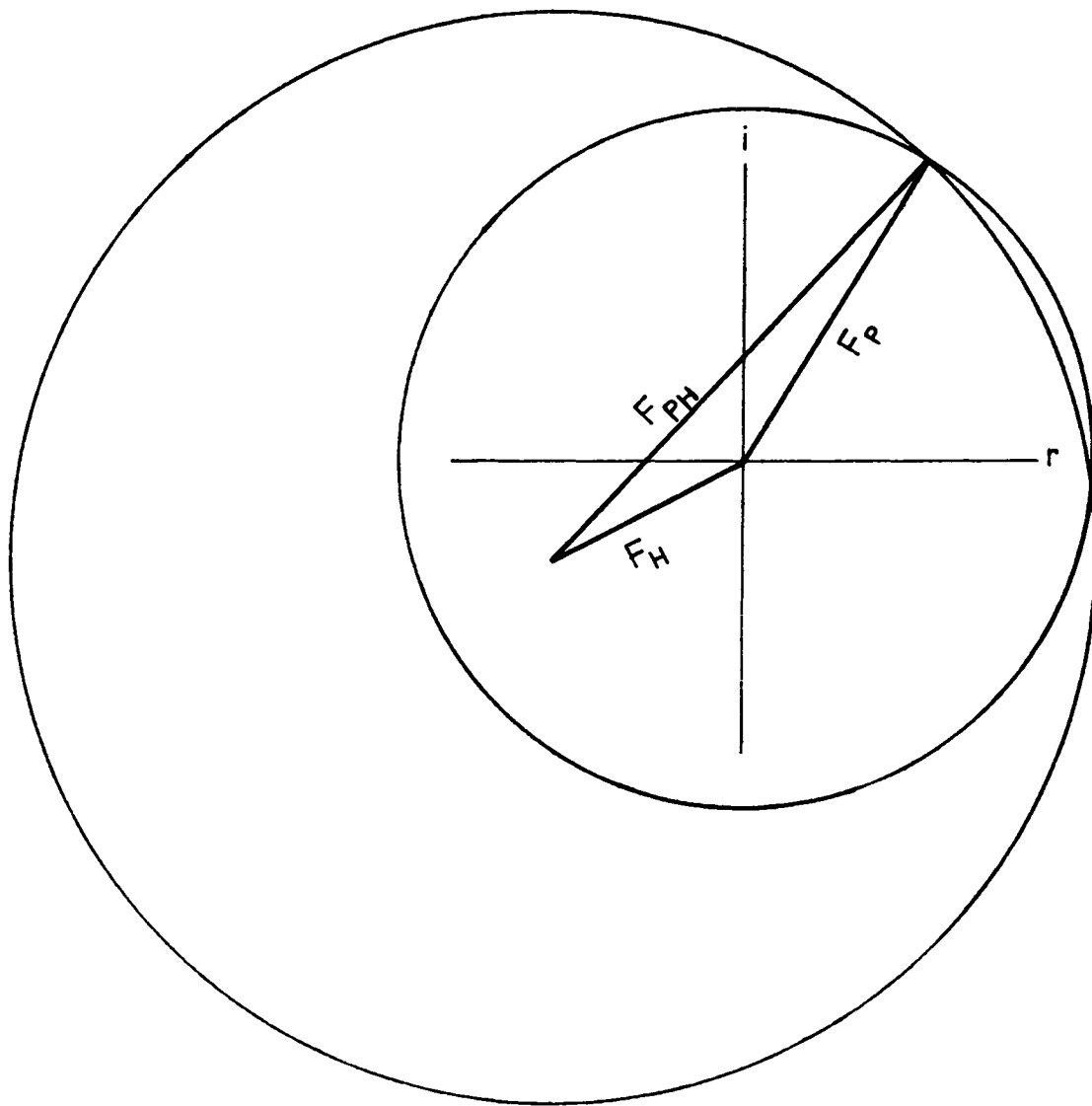


Figure III.3. The Harker Construction. Single isomorphous replacement leaves an ambiguity in the protein phase determination as there are two points at which the circles intersect.

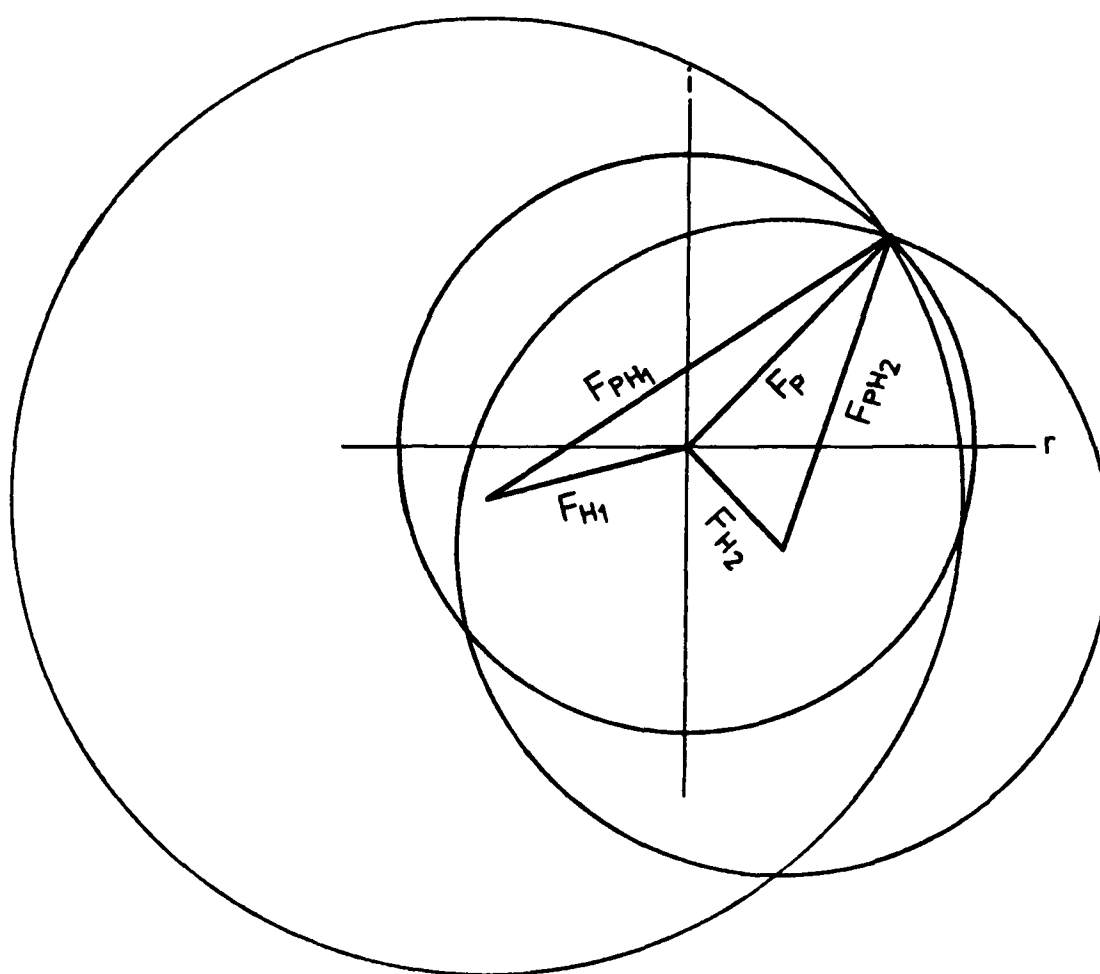


Figure III.4. Double Isomorphous Replacement. The ambiguity can be resolved with information from a second heavy-atom derivative.

III.B. Isomorphous Replacement

The only independent method for determining the phase in a protein crystal structure analysis is isomorphous replacement, or, more accurately, isomorphous addition (Green et al, 1954). In addition to the native protein, diffracted intensities are measured from an isomorphous crystal of a complex of the protein and one or more strongly scattering units such as an atom of high electron density (a heavy atom). Figure III.2 represents the complex quantities $\bar{F}_P$, $\bar{F}_{PH}$ and $\bar{F}_H$, the structure factors of the protein, complex and heavy atoms alone, on an Argand diagram, where

$$\bar{F}_{PH} = \bar{F}_P + \bar{F}_H \quad (B1).$$

The only directly measurable values, however, are F_{sP} and F_{sPH} (see equation (A14) above). As we shall see later, $\bar{F}_H$ can be calculated if we know the positions in the unit cell and the x-ray scattering behaviour of the heavy atoms. $\bar{F}_P$ and $\bar{F}_{PH}$ must be represented as circles about the tail and head of the $\bar{F}_H$ vector, since their phases are not known (figure III.3), (Harker, 1956). The circles, of radius F_{sP} and F_{sPH} , intersect only where equation (B1) is satisfied. Figure III.3 shows that two possible values of α_P result from this construction. The ambiguity can be resolved if intensities from a second heavy-atom isomorphous derivative are measured. The resulting information, comprised by the vector $\bar{F}_{H2}$ and the circle F_{PH2} is superimposed on the Harker construction to give figure III.4. There is one common point of intersection of the three circles, which indicates the true value of α_P .

2. Anomalous Scattering.

Friedel's law relates the diffraction at a reciprocal lattice point (hkl) (equation A12), to that at the point (-h,-k,-l),

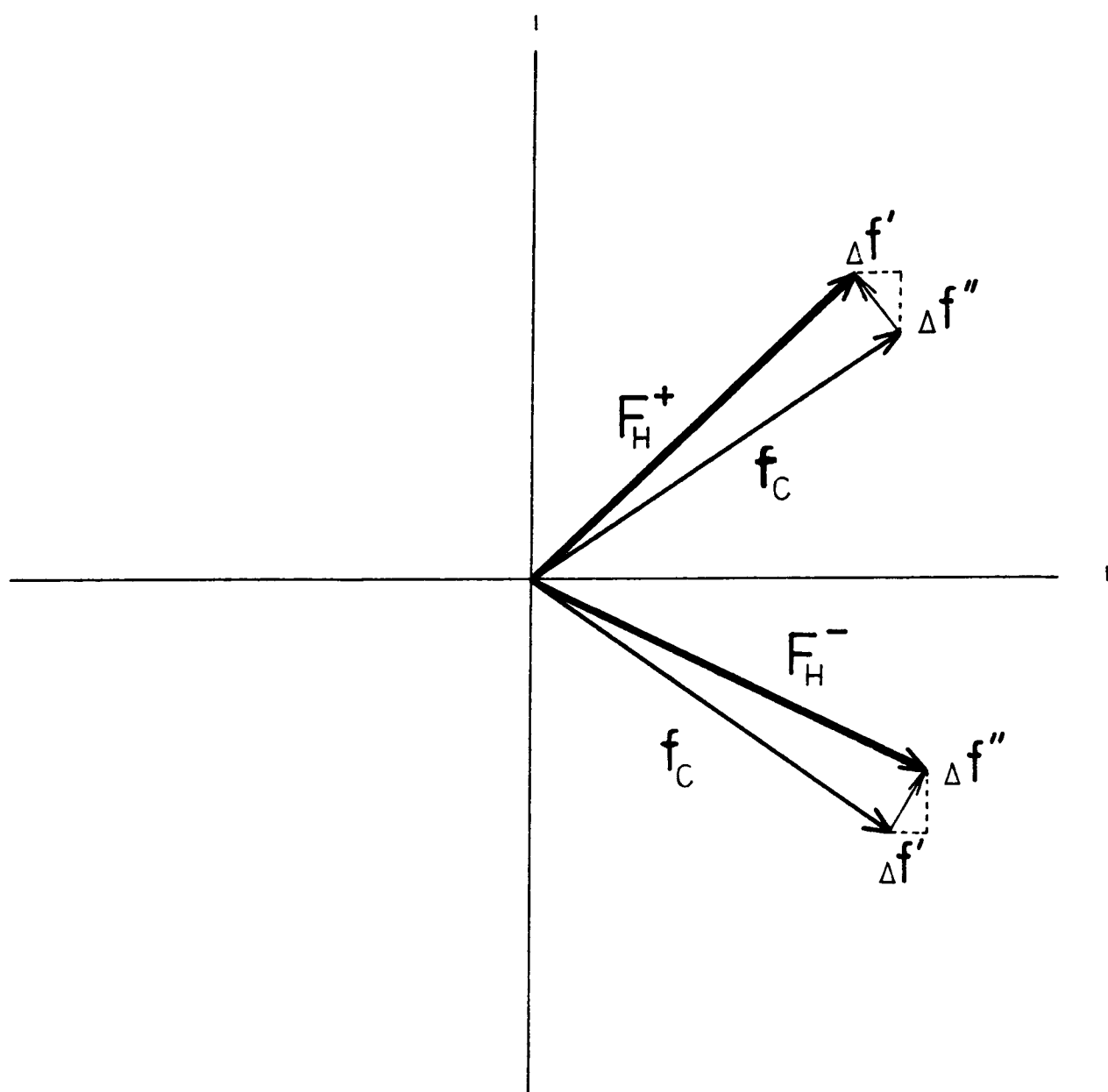


Figure III.5. Anomalous Scattering. The addition of the anomalous component ($\Delta f' + \Delta f''$) to the structure factors (f_C) of Friedel mates leads to the breakdown of Friedel's law.

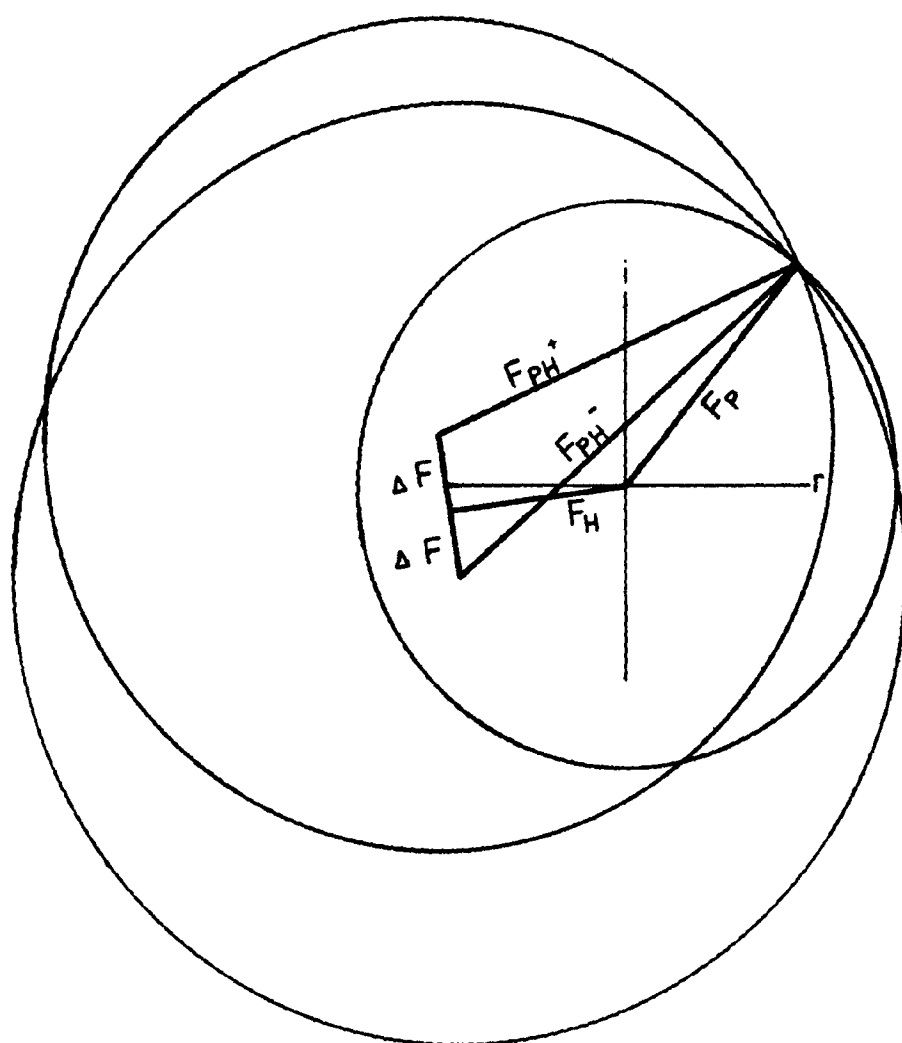


Figure III.6. The use of information from anomalous scattering to relieve the protein phase ambiguity.

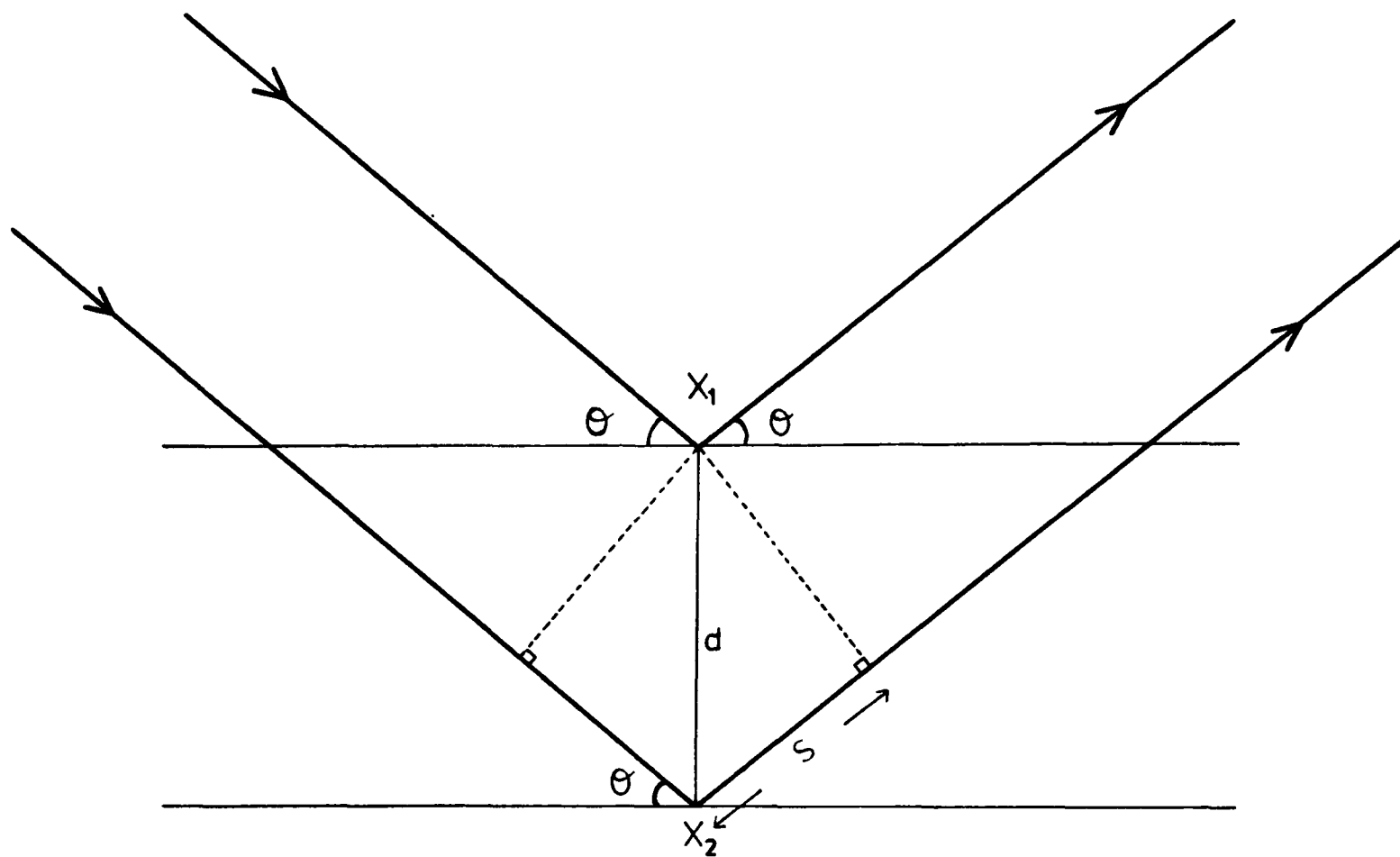


Figure III.7. An Illustration of Bragg's Law. The reflection of the incident beam by two points X_1 and X_2 separated by d .

$$\begin{aligned}\bar{F}(\overline{hkl}) &= v_L \int_V \rho(xyz) \exp [2\pi i(-hx-ky-lz)] dv \\ &= v_L \int_V \rho(xyz) \exp [-2\pi i(hx+ky+lz)] dv\end{aligned}\quad (B2).$$

Equations (A12) and (B2) are complex conjugates. Therefore, from expression (A14),

$$I(hkl) = I(\overline{hkl}).$$

We have assumed to this point that the scattering properties of an atom are the same in all directions. This is not true, particularly in heavy atoms where electrons commonly occupy the spherically unsymmetrical higher electronic orbitals. The structure factor amplitude can be corrected for this effect by the addition of two terms, one real and one imaginary, to the expected magnitude for a spherically symmetric atom, f_c ,

$$F_s = f_c + \Delta f' + i\Delta f'' \quad (B3).$$

For heavy atoms, this can significantly alter the values of $F_s(hkl)$ and $F_s(\overline{hkl})$ (figure III.5), resulting in the breakdown of Friedel's Law. In the diffraction pattern, this manifests itself in a difference between $I(hkl)$ and $I(\overline{hkl})$, and is called anomalous scattering. Its use in the phase determination is shown in figure III.6 in the form of a Harker construction. The circles of radii F_{s_p} , $F_{s_{PH}}(hkl)$ and $F_{s_{PH}}(\overline{hkl})$ intersect at a single point, relieving the ambiguity on α_p .

In practice, errors in measurements and crystal imperfections result in the three circles of figures III.4 and III.6 not intersecting at a point (see section F below). To improve the estimate of α_p , the methods of multiple isomorphous replacement and anomalous scattering are often used in conjunction.

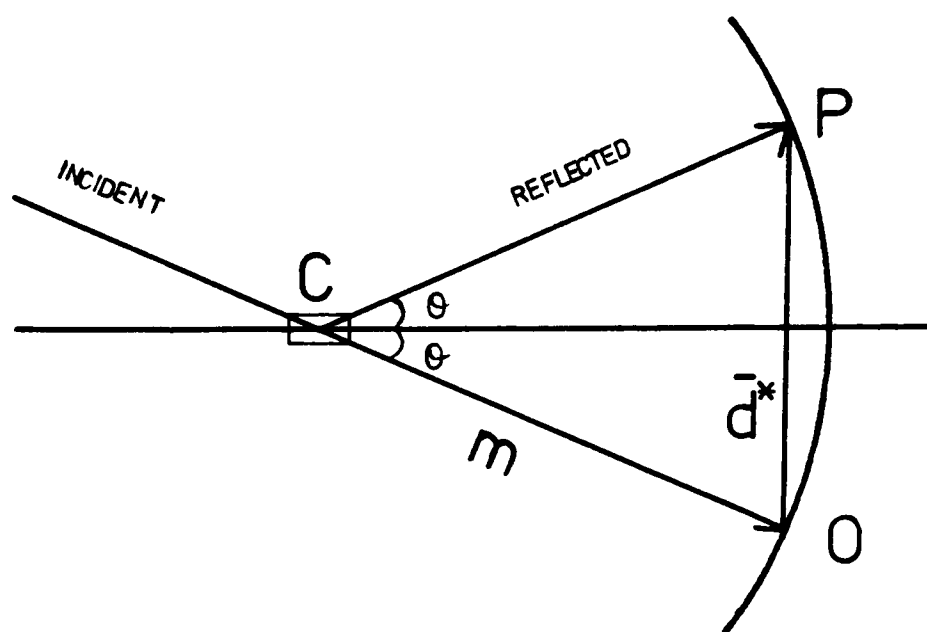


Figure III.8. The origin of reciprocal space, O , and the reflection position, P , lie on a circle. In three dimensions, this corresponds to the Ewald sphere or the sphere of reflection.

III.C. Bragg's Law and the Ewald Sphere

All the x-rays diffracted to a point (hkl) have the same phase.

Consider an incident beam of x-rays on two points X_1 and X_2 , making an angle θ with imaginary planes of reflection (the hkl planes) (figure III.7). The reflected rays are parallel but have a path difference of,

$$2s = 2d \cos (90-\theta) = 2d \sin\theta,$$

where d is the separation of the points. As the reflected beams are in phase, this path difference is an integral multiple of the wavelength of the x-rays. That is,

$$n\lambda = 2d \sin\theta \quad (C1),$$

as proposed by W.L.Bragg (1913).

The vector magnitude from the origin of the reciprocal lattice to a point (hkl) at which there is diffraction from atoms separated by the spacing d , is

$$d^* = \frac{1}{d} = \frac{2 \sin\theta}{\lambda} \quad (C2).$$

Therefore, atoms closer together diffract at a higher angle θ to a point further from the origin of reciprocal space (defined by the undiffracted x-ray beam). As figure III.8 shows, since the magnitude of the incident and reflected rays, m , is the same, the reciprocal lattice point d^* lies on a circle of radius m centred at the crystal, C . The value of m is easily found, since,

$$\sin\theta = \frac{1/2 d^*}{m},$$
$$m = \frac{d^*}{2 \sin\theta} = \frac{2 \sin\theta}{\lambda \times 2 \sin\theta} = \frac{1}{\lambda} \quad (C3).$$

In three dimensions, this corresponds to a sphere of radius $1/\lambda$, called the Ewald sphere or the sphere of reflection.

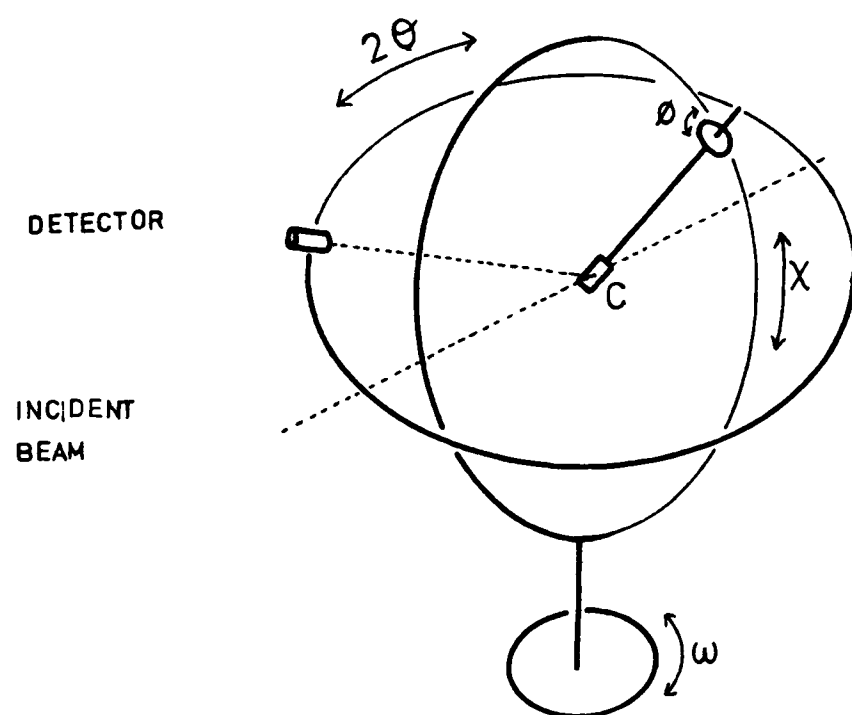


Figure III.9. The four-circle diffractometer normal beam geometry.
 $\omega=0$ when the χ -circle plane is perpendicular to the incident beam.
 $\chi=0$ when the ϕ -circle is parallel to the ω -circle.
 $2\theta=0$ where the incident beam intersects the 2θ -circle after passing through the crystal.
 $\phi=0$ is defined by the crystal axes. (After Arndt and Willis, 1966).

Conversely, the intensities of diffraction at all reciprocal lattice points within a sphere of radius d^* result from reflection at angles less than or equal to θ (equation C2). The resolution of a crystallographic study is the minimum interplanar spacing $d_{\min}$ for a scattering angle θ ,

$$R = d_{\min} = \frac{\lambda}{2 \sin \theta}.$$

Clearly, R decreases (the resolution increases) as the amount of data (the radius d^*) increases. In general, from a low resolution ($R > 6\text{\AA}$) study, the molecular boundaries and, perhaps, some features of the tertiary structure can be distinguished while at high resolution ($R < 2.5\text{\AA}$), the larger atoms can be seen and the nature of an amino-acid side chain can often be determined.

III.D. Methods of Data Collection

This section will describe briefly the methods used for data collection in the structure determination of Fv. Details of the data measurement and processing will be considered in Chapters IV and V.

1. The Four Circle Diffractometer.

The low resolution ($R \geq 6\text{\AA}$) data on Fv were collected on the four circle diffractometer in the normal beam setting.

a. Geometry.

The crystal orientation is defined by three angles, ω , ρ and χ , and the detector is moved from one diffracting position to another by rotation about the 2θ circle (figure III.9). To measure the intensity of a reflection the entire diffracted x-ray beam must pass through the detector. This is achieved by rotating the diffracting crystal about the axis ω which is always normal to the incident beam in the normal beam geometry.

b. The Measurement of Intensities.

The crystal is stepped through its diffracting position about the

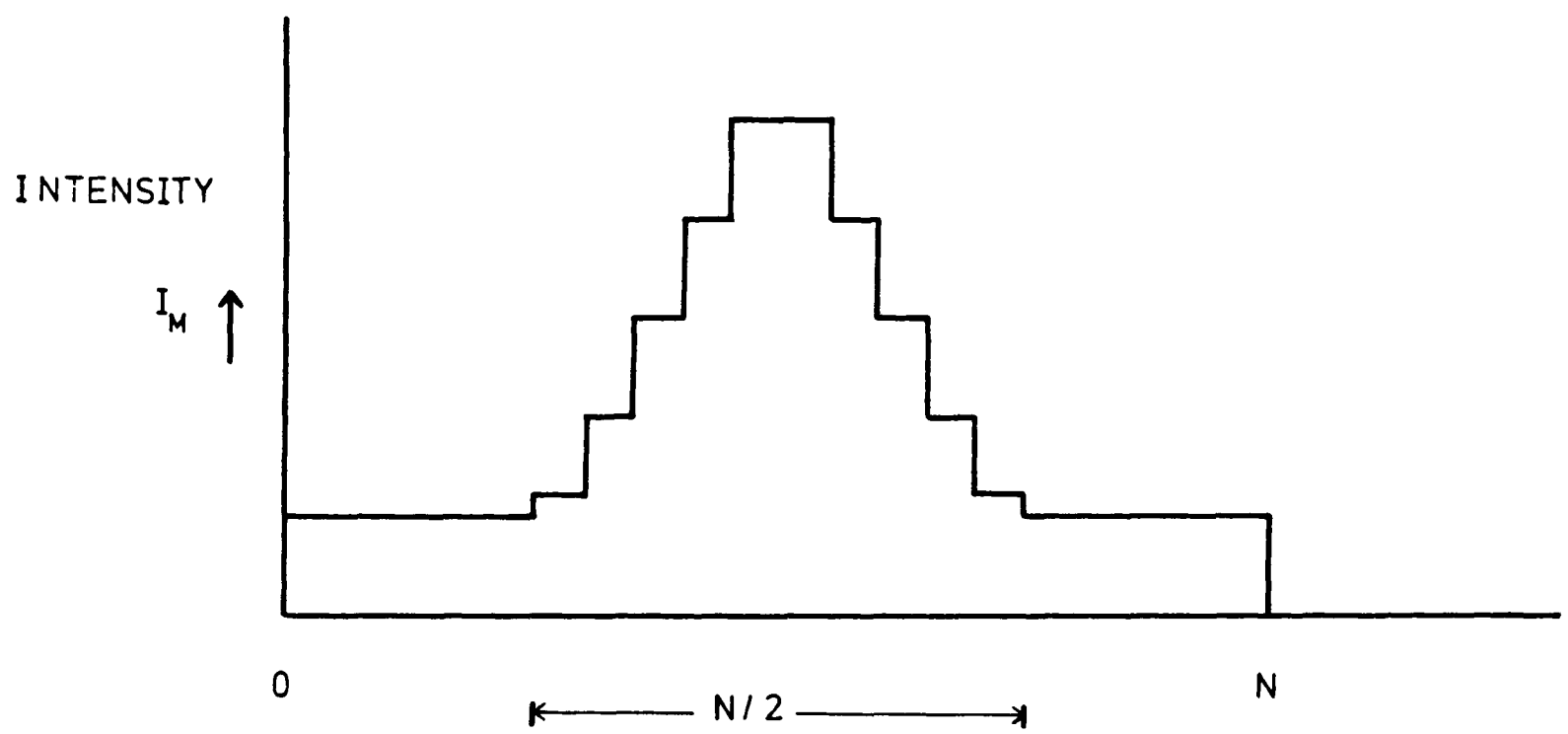


Figure III.10. The Ordinate Analysis Method of Peak Integration.
After Watson et al, 1970.

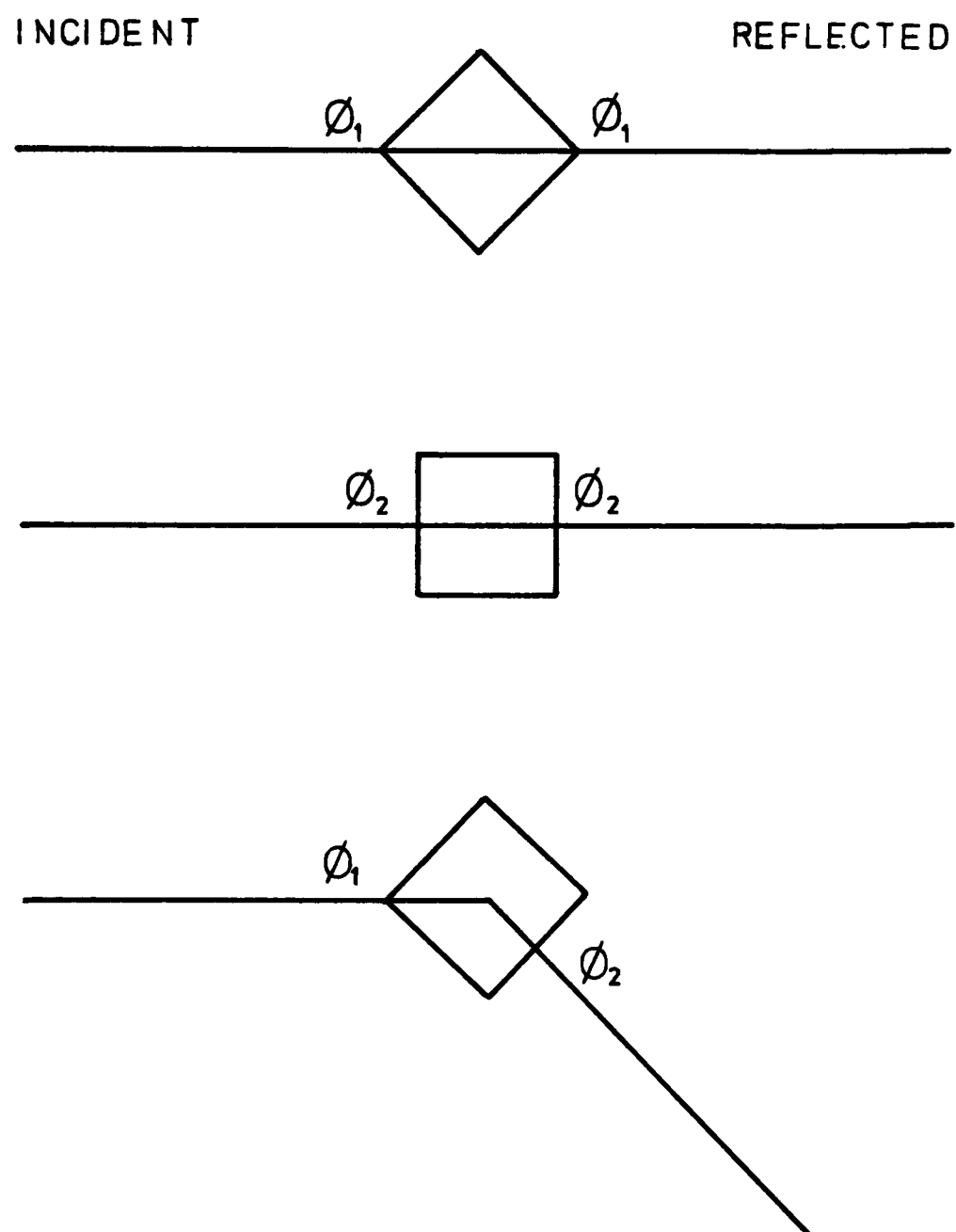


Figure III.11. The Absorption Correction. The intensities of the first two reflected beams can be used to correct the third.

ω axis. The peak of diffracted x-rays is superimposed on a background of scattered radiation due in part to inhomogeneity of the x-rays and imperfections in the crystal. As the background differs in different regions of reciprocal space, background measurements are taken on either side of the peak. Measurements are taken for N steps in ω , where $N/2$ is the width of the broadest peak to be measured (figure III.10). The $N/2$ largest consecutive intensities are added together and the $N/2$ remaining intensities are subtracted as background, yielding the peak intensity, I_p (Watson et al, 1970). An estimate of the error attached to the measurement due to variations in x-ray production and counter accuracy, can be made in the form of a standard deviation σ , the square root of the intensities of both peak and background,

$$\sigma = \left(\sum_{M=1}^N I_M \right)^{1/2}.$$

This does not take into account such errors as crystal imperfection and mis-setting, but the standard deviation is a useful monitor of the significance of a measurement.

The integration procedure outlined above was carried out by the program PRIMROSE (see appendix I). The program applies two further corrections to each integrated intensity, the polarization and Lorentz corrections.

The incident x-ray beam is unpolarized but a certain amount of polarization occurs on diffraction. This effect varies with the angle of reflection, θ , and the correction applied to the intensity is,

$$I_1 = I_p \left(\frac{2}{1 + \cos^2 2\theta} \right).$$

The Lorentz correction takes into account the different speeds at which reciprocal lattice points pass through their diffracting positions. Points closer to the origin of reciprocal space spend more time in the

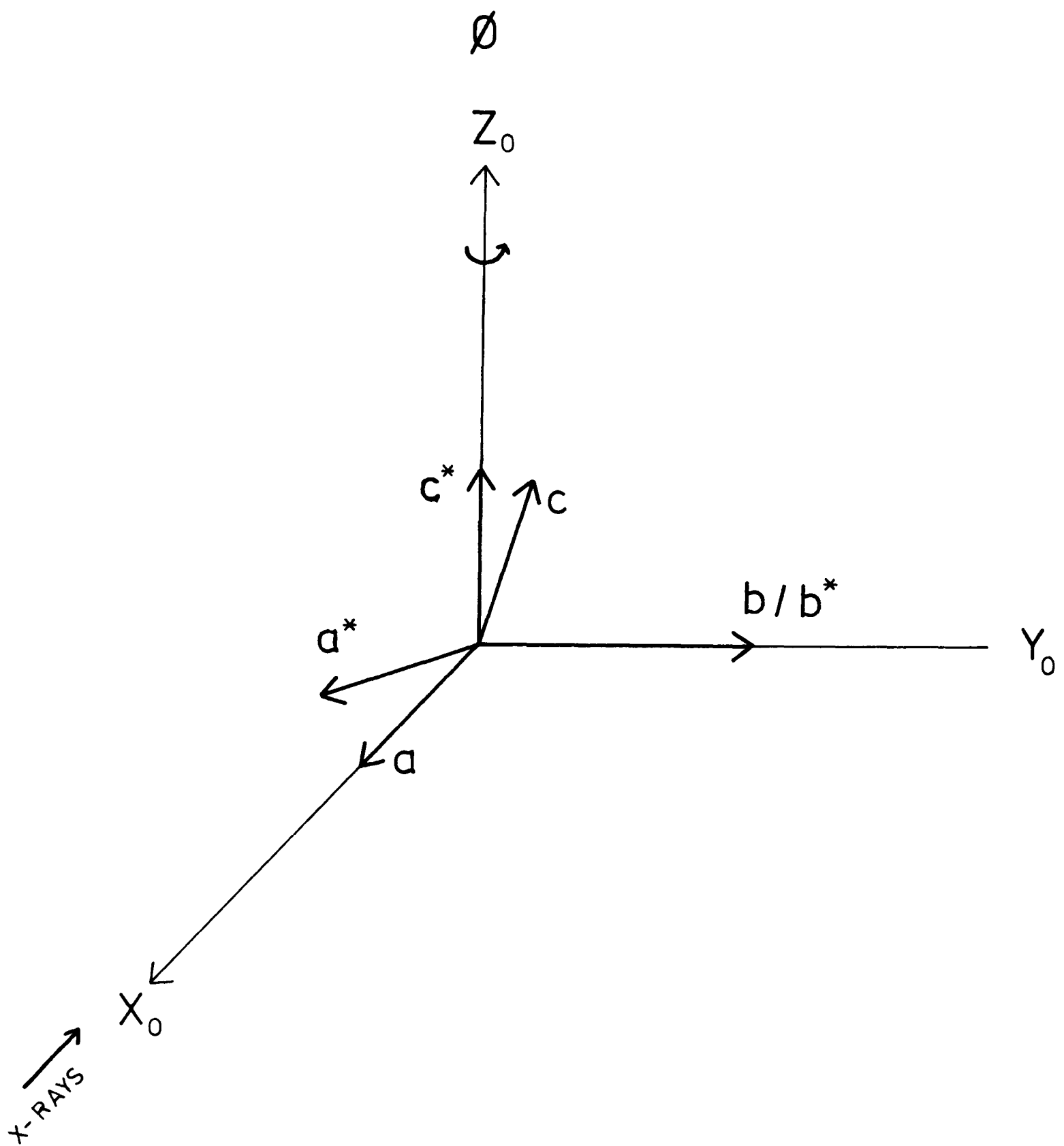


Figure III.12. The Rotation Camera. The position of the crystal axes in relation to the camera axes (X_0, Y_0, Z_0) for a c^* mount.

x-ray beam than points further away. In the symmetrical-A setting, where the χ circle bisects the incident and reflected beams, this varies only with resolution (i.e. d^*), and the correction applied is,

$$I = I_1 (\sin 2\theta).$$

c. The Absorption Correction.

X-ray beams being diffracted in different directions follow paths of different length through the crystal. The intensity of an x-ray beam passing through a thickness, t , of crystal decreases because of absorption by the crystal, according to the expression,

$$I = I_0 e^{-\mu t},$$

where I_0 is the incident and I the emergent intensity, and μ the absorption coefficient of the crystal. There is a simple method for correcting this effect on the four-circle diffractometer (North et al, 1968). The crystal is oriented so that the incident and reflected beams are equally inclined to the rotation axis of the crystal, ρ . As the crystal rotates, the reflecting condition is continuously satisfied, but the intensity of the diffracted beam varies according to its path through the crystal. A non-empirical absorption curve, examples of which can be seen in Chapter IV, can be plotted as a function of ρ . The correction applied to an intensity is derived from the absorption at the ρ values of the incident and reflected beams (see figure III.11). To account for the variation of absorption with resolution, absorption curves can be measured at several values of θ .

2. The Rotation Camera.

The higher resolution data were collected on the rotation (or oscillation) camera. Again, the crystal is in the normal beam setting, the oscillation axis being normal to the incident x-ray beam. An important difference is that the oscillation and rotation axes (ω and ρ on the four

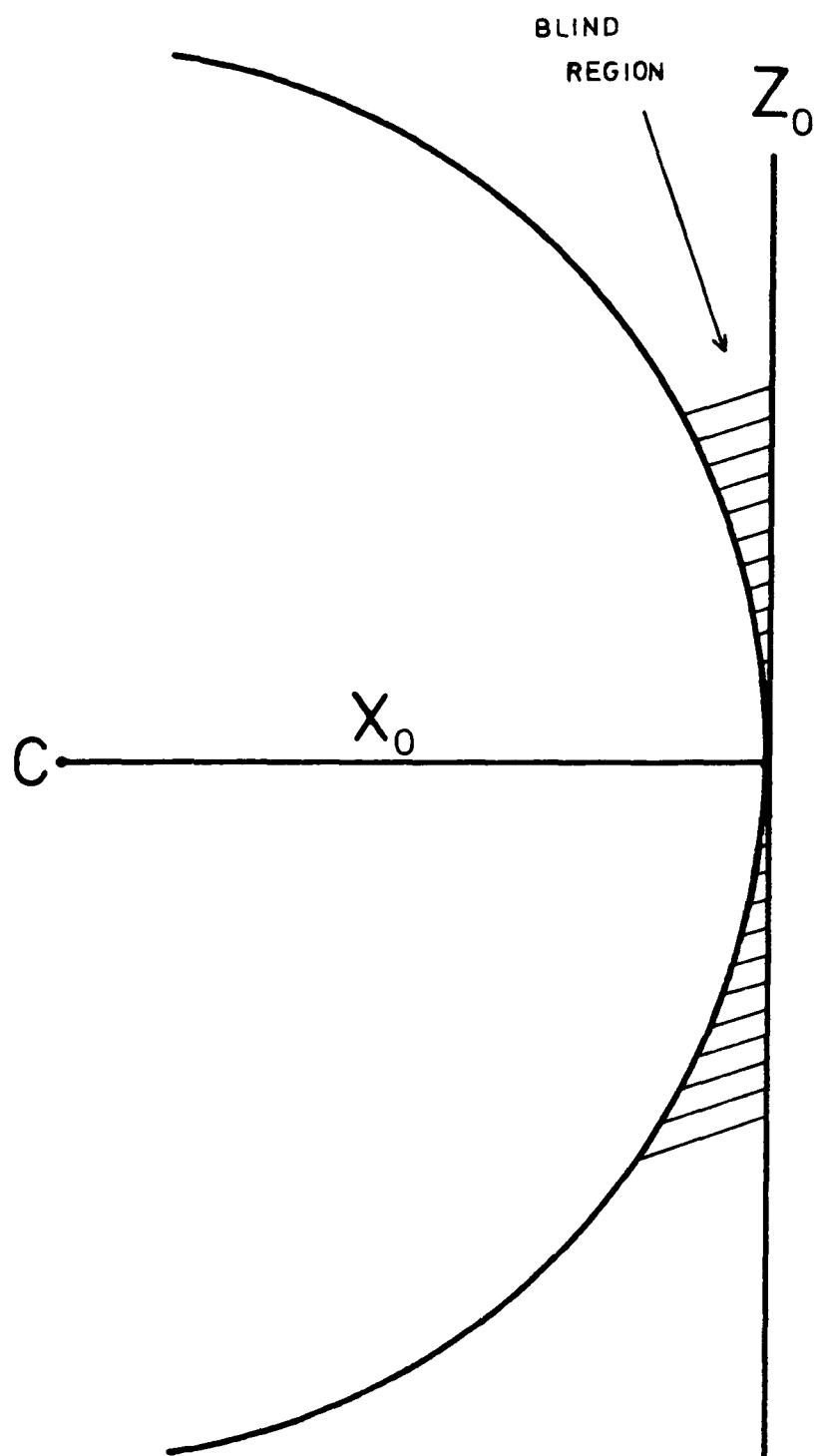


Figure III.13. The blind region which is not sampled by the sphere of reflection centred at the crystal, C , on rotation about Z_0 .

circle diffractometer) are coincident.

The orientation of the crystal is defined by an orthogonal set of axes X_0 , Y_0 and Z_0 (see figure III.12), where Z_0 is parallel to the rotation axis, φ , and X_0 is anti-parallel to the incident x-ray beam at $\varphi = 0$. Intensities of diffraction are measured on a photographic film held in a flat cassette perpendicular to the incident beam. As the crystal rotates, many reciprocal lattice points intersect the sphere of reflection at any time. In practice, only the reflections occurring in a limited rotation can be recorded before they begin to overlap on the film.

Data are collected on several films in angular steps about φ of a fixed size. To minimize the effect of a variable x-ray beam, the crystal is oscillated through the required angle instead of simply stepped through it once. When the end of the rotation for a particular film is reached, some lattice points will be sitting in their diffracting positions. These are referred to as partial reflections since the peaks are only partially recorded on the film. However, if the following rotation step starts at precisely this value of φ , the unrecorded parts of the partial reflections will appear on the next film. The contributions of the two films are added together to obtain the full intensities of the reflections. Wooster (1964) pointed out that in practice, some systematic error may be introduced because of the sharp gradient of optical density at the edge of partially recorded reflections.

Figure III.13 shows a crystal in the rotation camera geometry with a point P in its diffracting position. As the crystal rotates about φ , the sphere of reflection rotates about the axis Z_0 through the origin of the reciprocal lattice. Clearly there is a region of reciprocal space which will never intersect the sphere of reflection. This blind region can only be measured by reorienting the crystal so that a different crystal axis

lies parallel to the rotation axis. In fact, at the highest resolution achieved in this analysis, $R = 2.7\text{\AA}$, only about 2% of the reciprocal lattice points lie in this region (Arndt and Wonacott, 1977, p.81).

The major disadvantage of the rotation camera technique is the greater number of steps involved in the processing procedure, each increasing the potential for error. After chemical development of the films, the degree of exposure to x-rays of every point on the film is measured by an automatic microdensitometer. Each film is scanned in steps of $100\mu\text{m}$ and the optical density of each point, determined by the degree of attenuation of a beam of light passed through the film, is transferred to a magnetic tape to be read by the film processing program OSCAR (Wilson and Yeates, 1979 and appendix I).

First, the precise orientation of the crystal relative to the X_0 , Y_0 and Z_0 axes is determined by the location of reflection spots on still photographs taken at, in general, three values of ρ . Integrated intensities were then determined by summing the optical densities of all points on the data film defined as being within the diffraction spot, and subtracting a background estimated from the optical densities of four boxes, one at each corner of the peak.

Occasionally a diffracted beam saturates the x-ray film; that is, it blackens the film to such an extent that it cannot be blackened further. These intensities are measured on a second film placed behind the first in the same cassette. The diffracted beam is usually sufficiently weakened by absorption by the first film that it does not saturate the second. The second films are processed in the same way as the first and the intensities from the two are merged by the program FILMPACK (appendix I).

Lorentz and polarization corrections are applied as described in section III.1b and a correction is made for non-linearity of the increase in

optical density of the film with increasing x-ray exposure. To account for absorption by the first film, a function of θ is applied to intensities on the second film because a high angle reflection passes through the film at a more oblique angle than one at low angle. A standard deviation is estimated for each reflection and, finally, contributions of reflections partially recorded on different films are added and the data are merged into a unique set.

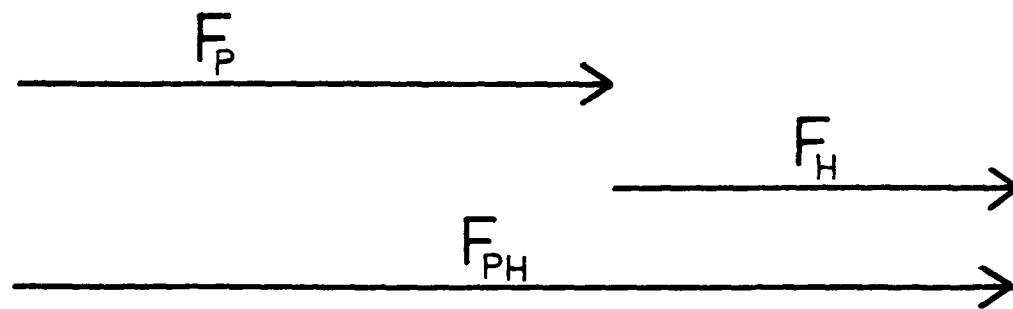
No directly measurable absorption correction like that described in section III.D1c is feasible on the rotation camera data. Empirical approaches to the absorption problem (most recently Stuart and Walker, 1979) have been attempted, but such a method was not applied to the Fv data. Instead, a crude correction for absorption was made by scaling the low resolution terms collected on the rotation camera to the absorption corrected four circle diffractometer set. The scale, calculated as a function of ρ , was then applied to the whole rotation camera data set (see discussion of local scaling in the following section).

3. Scaling Data Sets Together.

In the following discussion, the integrated intensities from the two data collection methods are treated in the same way.

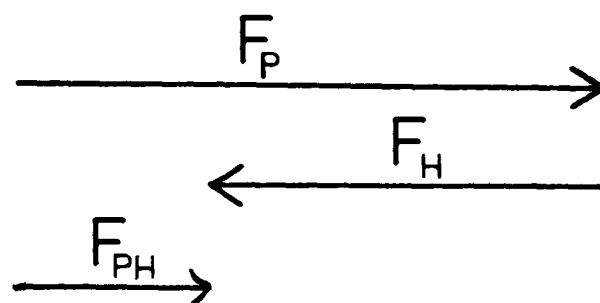
The derivative and native data sets must be placed on the same relative scale. The program ANSC (appendix I) analyses the intensities of equivalent reflections in the different data sets. An overall scale factor is calculated and applied (Fox and Holmes, 1966). In addition, systematic errors can cause variations in scale factor over regions of reciprocal space. These are analysed and a procedure of local scaling (Matthews and Czerwinski, 1975) can be carried out. This procedure involves the examination of small equivalent areas of reciprocal space, for example that within a cube of fixed size, and calculation of the scale factor for the reflections

a.



$$|F_{PH}| = |F_P| + |F_H|$$

b.



$$|F_{PH}| = |F_P| - |F_H|$$

Figure III.14. F_H can be obtained directly from F_P and F_{PH} for centric reflections where the phases of all three structure factors are 0 or π . Part b, where $F_{PH} < F_P$, is rare.

within the cube as it moves in a direction depending on the index being varied. Commonly, variations are examined as functions of the lattice indices h , k and l , resolution and the angle ρ . The latter is particularly important in checking for residual absorption effects.

III.E. The Determination and Refinement of Heavy Atom Positions

In part B of this chapter, methods for determining the phase of a reflection were described. These procedures depended on the knowledge of the quantity $\bar{F}_H$, the structure factor of the heavy atoms, which can be estimated if the positions of the heavy atoms in the crystal unit cell are known.

The Patterson function, P , (Patterson, 1934) is defined,

$$P(uvw) = \int_{V_C} \rho(xyz) \rho(x+u, y+v, z+w) dv \quad (E1),$$

where V_C is the volume of the unit cell. Clearly, P will have a large value when the electron densities of both (xyz) and $(x+u, y+v, z+w)$ are large, that is, when (uvw) is an interatomic vector. Another form of the Patterson function can be derived from the equation (A13) (see, for example, Blundell and Johnson, 1976, p.137),

$$P(uvw) = \frac{1}{V_L} \sum_{hkl} F(hkl)^2 \exp [-2\pi i(hu + kv + lw)] \quad (E2).$$

As this is a function of the structure factor amplitudes alone, it can be calculated directly from intensity data. We could not expect to be able to interpret such a function for a protein which has many atoms of similar size. However, in a Patterson function calculated with the structure factor amplitudes of the heavy atoms, F_H , the inter-heavy atom vectors will appear as large peaks. If there are only a few heavy atoms in the unit cell, their positions can be deduced from these vectors.

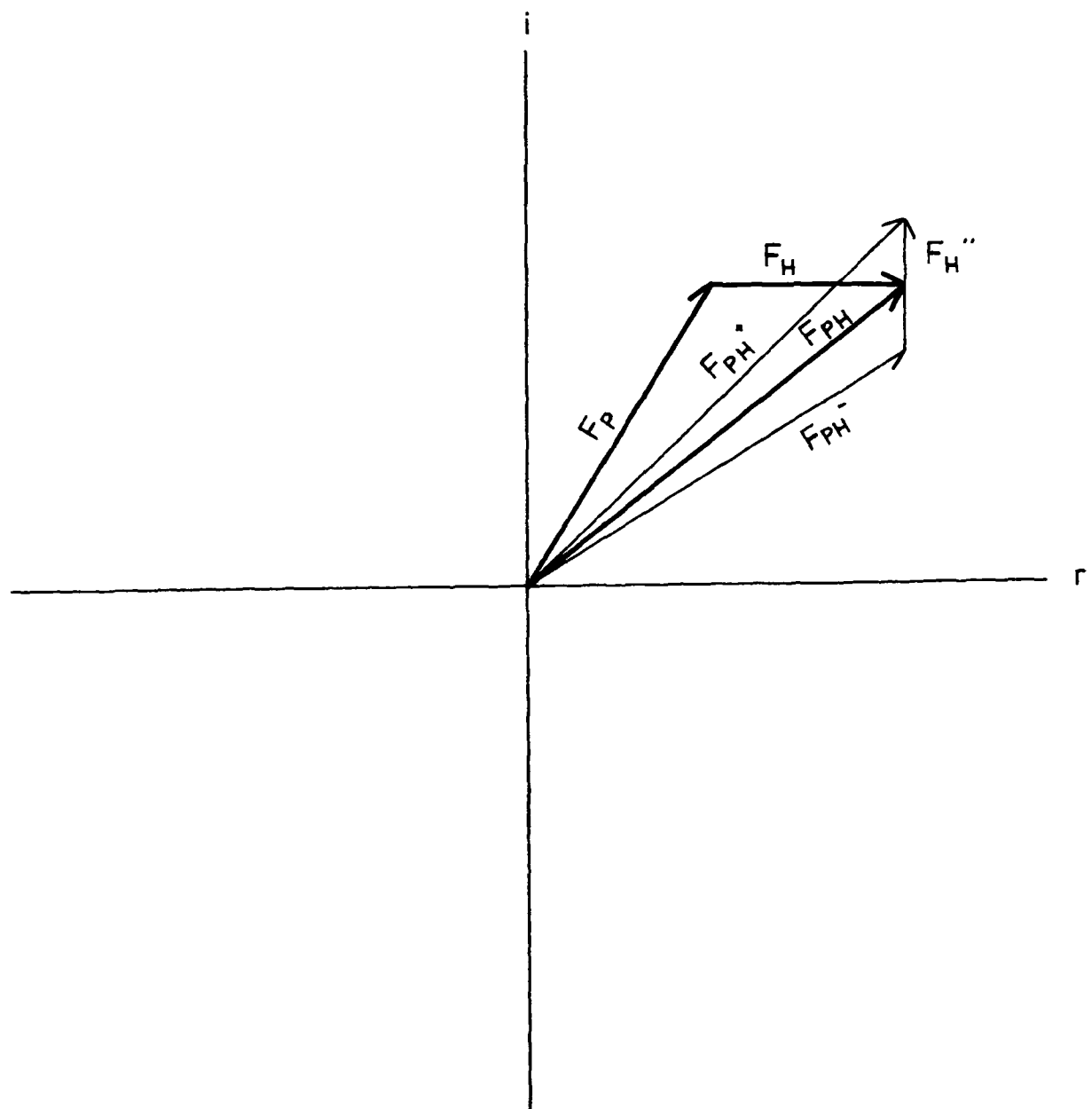


Figure III.15. The inclusion of the anomalous scattering information in the estimation of F_H . Compare with figure III.5.

It was stated earlier that $F_H = F_{PH} - F_P$ is a useful expression only when the phases of the structure factors are known. In some instances, though, F_H can be estimated from the structure factor amplitudes F_{PH} and F_P alone, since $|F| \propto I^{1/2}$.

1. The Calculation of F_H .

a. Centric Reflections.

In some projections, a noncentrosymmetric structure (like a protein) can appear centrosymmetric. Reflections in centric zones have phases of either 0 or π (figure III.14). For most centric reflections,

$$F_H = F_{PH} - F_P \quad (E3).$$

Unfortunately, very few reflections from an Fv crystal lie in a centric zone (only those with $k = 0$). Equation (E3) is only useful for other reflections when $\alpha_H - \alpha_{PH}$ is very small.

b. Acentric Reflections.

The most common situation is illustrated in figure III.15, where the anomalous contribution is included as a resultant vector ($\bar{F}_H''$) perpendicular to $\bar{F}_H$, and the vectors representing the ($\overline{hkl}$) reflection (see figure III.5) are reflected across the real axis. By means of the cosine rule, it can be derived that (for example, Blundell and Johnson, 1976, p.177),

$$F_{PH(+)} - F_{PH(-)} = (2F_H/k) \sin(\alpha_{PH} - \alpha_H) \quad (E4),$$

where $k = F_H/F_H''$ and is derived theoretically $((fc + \Delta f')/\Delta f'')$, in the notation of figure III.5). This anomalous difference provides a good estimate of F_H only when $\alpha_{PH} - \alpha_H$ is close to $\pi/2$. In general, the isomorphous and anomalous differences are used in conjunction to estimate F_H

$$F_H^2 = F_P^2 + F_M^2 \pm 2|F_M|^2 F_P^2 - (k/4)^2 (\Delta I)^2 |^{1/2} \quad (E5),$$

where $F_M^2 = 1/2 (F_{PH(+)}^2 + F_{PH(-)}^2)$

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

The uncertain sign in (E5) is negative when $\alpha_{PH} - \alpha_P$ is small ($\cos(\alpha_{PH} - \alpha_P)$ is positive), as it is for most reflections. The lower estimate of F_H , F_{HLE} , is, therefore, usually calculated.

Once F_H has been estimated in this way, an iterative refinement procedure is commonly performed, minimizing the expression

$$\sum_{hkl} w[F_{HLE} - F_{H(calc)}]^2,$$

where $F_{H(calc)}$ is derived from the current best estimates of the heavy-atom positions and occupancies. The weighting term, $w = (\sigma_{F_P}^2 + \sigma_{F_H}^2)^{-1}$, decreases the contribution of badly measured reflections. In addition, terms are left out if it is possible that the sign in (E5) is positive; that is, if $F_{PH} - F_P$ is large.

As a monitor of refinement,

$$R_{(F_{HLE})} = \frac{\sum (F_{HLE} - F_{H(calc)})}{\sum F_{HLE}}$$

is calculated. A random calculation of $F_{H(calc)}$ gives an $R_{(F_{HLE})}$ of 0.586 (acentric) or 0.828 (centric terms) (Wilson, 1950). The progress of refinement is also followed by examining a plot of $F_{H(calc)}$ against F_{HLE} the slope of which should approach unity.

Finally, a difference Fourier $(|F_{HLE} - F_{H(calc)}| \exp(i\alpha_{calc}))$ will show maxima or minima at points where heavy-atom positions are wrongly left out or included in the calculation of $F_{H(calc)}$.

2. Phased Refinement.

Once a solution is obtained for one derivative, a set of protein

phases α_p can be calculated using anomalous scattering to relieve the phase ambiguity, where possible (section III.B2). The heavy atom positions of a second derivative will show as peaks in a difference Fourier $|F_{PH} - F_P| \exp(i\alpha_p)$. These can be used to calculate F_{H_2} .

An alternative refinement scheme to that described in the last section is available, since F_{PH_2} can be calculated,

$$F_{PH_2(\text{calc})} = |F_P \exp(i\alpha_p) + F_{H_2(\text{calc})} \exp(i\alpha_{\text{calc}})|,$$

and the expression to be minimized is (Dodson and Vijayan, 1971 and Blow and Matthews, 1973),

$$\sum_{hkl} w (F_{PH_2(\text{obs})} - F_{PH_2(\text{calc})})^2.$$

The Kraut R-factor is described as,

$$R_K = \frac{\sum (kF_{PH(\text{obs})} - F_{PH(\text{calc})})}{\sum F_{PH(\text{obs})}},$$

where k, the scale factor between $F_{PH(\text{obs})}$ and F_P , can also be refined.

III.F. The Calculation of Phases

Ideally, the solution of two heavy atom derivatives with anomalous scattering information from each provides more than the minimum requirement to relieve the phase ambiguity. In practice, in a Harker construction for a reflection, all the circles do not intersect at a common point and often leave a confusing picture of what the protein phase should be. This is a result of errors in the method which can be dealt with in two ways.

Firstly, an overall error, E, for a reflection can be calculated from a combination of errors in the determination of F_{PH} and F_P , δ , and in the calculation of $\bar{F}_H$, e, as a result of positional or occupancy inaccuracies or non-isomorphism in the derivative crystals. Thus $E^2 = \delta^2 + e^2$, and can be estimated directly for centric terms, where,

$$E^2 = (|F_{PH} \pm (F_P + \bar{F}_H)|)^2.$$

Secondly, the errors manifest themselves in a lack of closure of the triangle formed by F_P , F_H and F_{PH} , and can be expressed directly as

$$\epsilon_j = (F_{PH(obs)} - F_{PH(calc)})$$

for the derivative j , where $F_{PH(calc)}$ closes the phase triangle.

By the method of Blow and Crick (1959), the probability of a reflection having a protein phase α is,

$$P_j(\alpha) = \exp \left[-\frac{\epsilon_j(\alpha)^2}{2E_j^2} \right].$$

The probabilities from several derivatives are multiplied:

$$P(\alpha) = \prod_j P_j = \exp \left\{ -\sum_j \frac{\epsilon_j(\alpha)^2}{2E_j^2} \right\} \quad (F1).$$

Blow and Crick have shown that the error in electron density is smallest when the phase at the centroid of the probability distribution, α_{best} , not the most probable phase, is used in the electron density calculation.

To weight down the terms where α is badly determined, F_P is multiplied by a figure of merit, m , where,

$$m = \frac{\int_{\alpha=0}^{2\pi} P(\alpha) \exp(i\alpha) d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha) d\alpha} \quad (F2).$$

From equation (F1) it can be seen that m is dependent on the estimate of E .

Information from anomalous scattering is included in equation (F1) as a separate derivative. The probability distribution is calculated differently, though, because anomalous scattering results are usually more accurate than those from isomorphous differences (North, 1965 and Matthews, 1966). Thus, the anomalous scattering lack of closure,

$$\epsilon'_j(\alpha) = F_{PH(-)} - F_{PH(+)} - \frac{2F_P F''_H}{F_{PH}} \sin(\alpha_{PH} - \alpha_H),$$

and E'_j is derived from the standard deviations of the anomalous scattering measurements.

Our best estimate of the electron density is, then, a Fourier synthesis with coefficients,

$$mF_P \exp(i\alpha_{best}).$$

CHAPTER IV

The Structure Determination to 6Å Resolution

IV.A. Preliminary Crystallography

1. Crystallization.

The conditions under which the Fv fragment crystallizes were ascertained by Dr. R. Aschaffenburg. Crystals formed in solutions containing approximately 13% polyethylene glycol (PEG) $[\text{OH}-(\text{CH}_2-\text{CH}_2-\text{O})_N\text{H}]$ of average molecular weight either 4000 or 6000. Fv prepared as described in section II.A2, was dissolved in 0.1M imidazole/HCl buffer (pH 7.0) to a clear solution containing 6.7 mg/ml. To a sample of 100 μl of protein solution, 35 μl of a stock solution consisting of equal weights of PEG and water, was added. The sample turned distinctly turbid and after some time fine floccules formed on the walls of the glass crystallization tube. In one to two weeks, crystals developed between the floccules. The size and number of crystals increased over several weeks. Figure IV.A1 is a photograph of an Fv crystal grown by this method.

2. Preliminary Crystallography.

A single crystal was mounted in a glass capillary tube, the mother liquor immediately around the crystal removed, and the ends of the tube sealed with wax. The tube was then fixed onto a standard goniometer head and placed into the diffracting position on a Buerger precession camera. Initially, x-ray photographs were taken of the stationary crystal with the incident beam perpendicular to the film, and the goniometer head adjusted such that a crystal axis lay parallel to the incident beam. The axis and the plane of the film were then offset from the incident beam by a fixed angle and rotated about the incident beam. If the upper levels of diffraction are blocked by a layer-line screen, this gives an

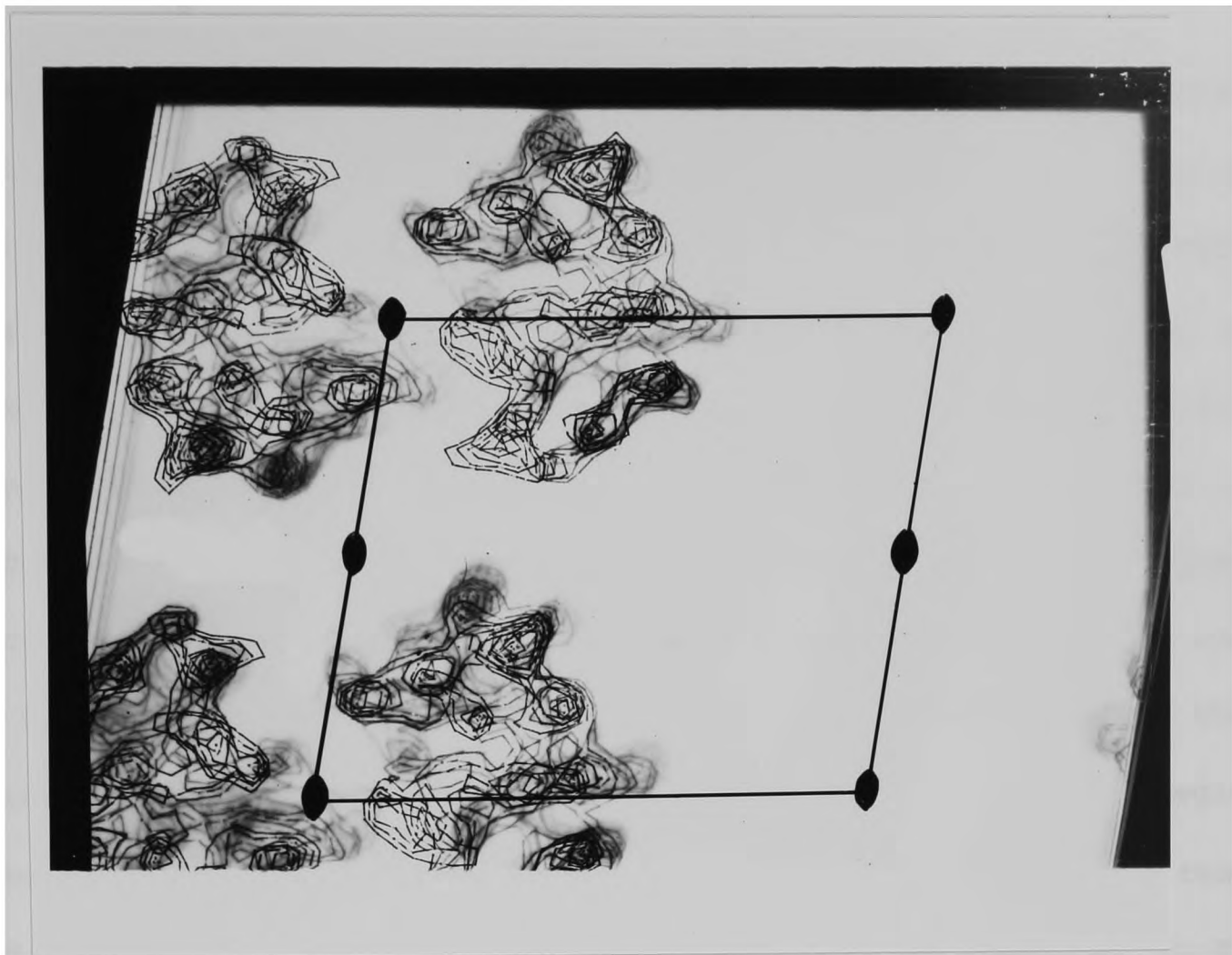


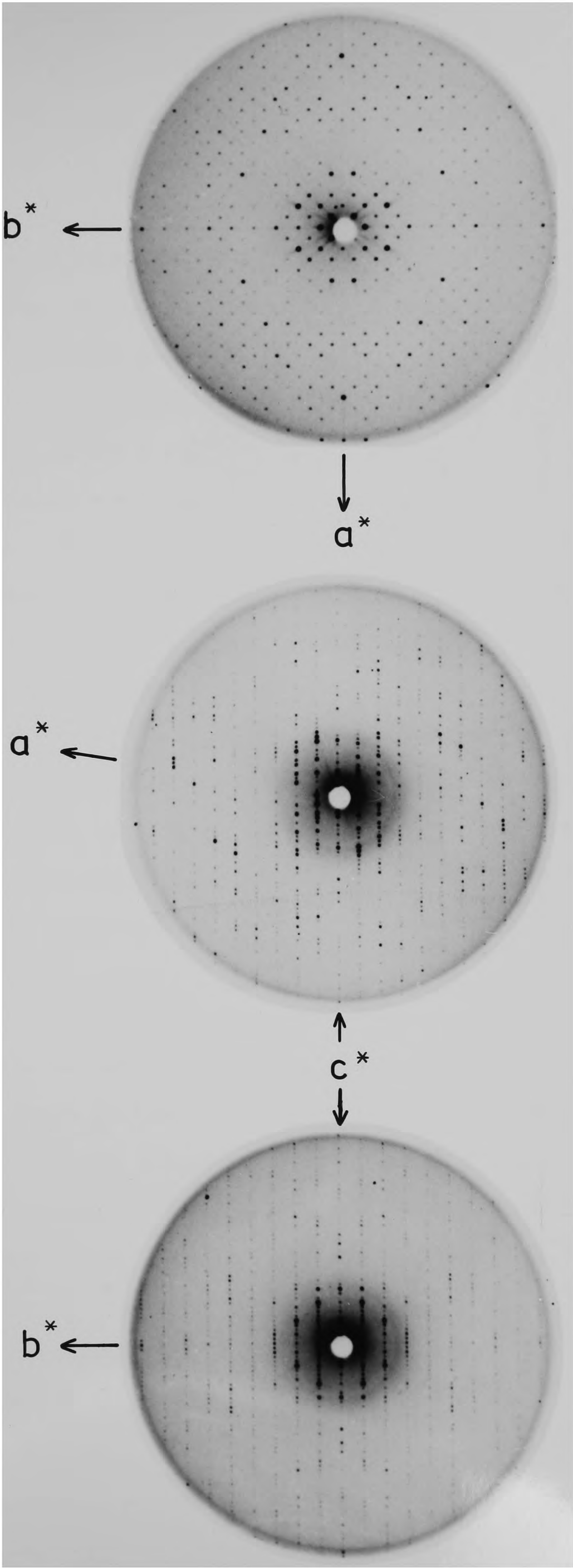
Figure VI.B1. A sample of the calculated electron density from the Fv domain of Fab New (Poljak et al, 1974).

hk0 Zone

h0l Zone

0kl Zone

Figure IV.A2. Precession photographs of the three major zones of an Fv crystal. The edge of the photograph corresponds to about 3Å resolution.



undistorted image of the reciprocal lattice plane perpendicular to the crystal axis, and the dimensions of the reciprocal lattice can be measured. Precession photographs, taken with $\text{CuK}\alpha$ radiation, of the three major zones of a native Fv crystal, are shown in figure IV.A2.

The symmetry of these zones indicates that the crystals are of space group C2 with unit-cell dimensions,

$$a = 59.8 \pm 0.5 \text{ \AA}$$

$$b = 56.5 \pm 0.5 \text{ \AA}$$

$$c = 138.8 \pm 1.0 \text{ \AA}$$

$$\alpha = 90.0^\circ, \quad \beta = 99.6 \pm 0.1^\circ, \quad \gamma = 90.0^\circ.$$

The relationship of the crystal morphology to the real and reciprocal axes is given in figure IV.A3.

3. The Use of Polyethylene Glycol as a Crystallizing Agent.

The chief obstacle in the study of a protein structure by x-ray diffraction is the procurement of the protein in a crystalline form, still very much a process of trial-and-error. The use of PEG to obtain protein crystals has arisen from the observation that a large fraction (13 out of 22) of attempts to crystallize different proteins with PEG succeeded (McPherson, 1976).

The mode of action of PEG as a crystallizing agent and its effects on proteins are not clearly understood. Ogston (1958) described the mechanism as mutual exclusion of proteins and large polymers in solution; the PEG, on dissolving, physically forces the protein out of solution. However, the concentration of PEG required for crystallization is low enough to indicate that volume exclusion alone probably does not account for this event.

Pittz and Timasheff (1978) studied solutions containing protein and another organic solvent, 2-methyl-2,4-pentanediol (MPD), and provided a

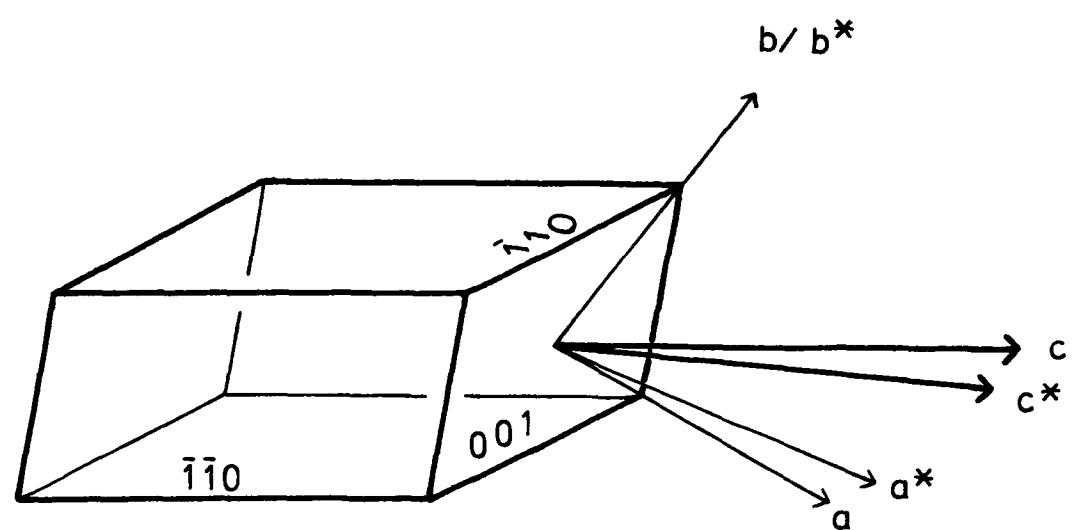


Figure IV.A3. The relationship of the Fv crystal morphology to the real and reciprocal axes.

more chemical explanation. They proposed that in solutions containing up to 50% MPD, protein is preferentially hydrated.

The addition of protein increases the chemical potential of the MPD and creates a thermodynamically unstable situation from which phase separation occurs. In addition, because the protein is hydrated, MPD does not have access to hydrophobic amino-acid residues in the protein and, therefore, cannot induce conformational changes. Thus, crystals of the protein in its native conformation, result.

Lee and Lee (1981) extended this view to PEG solutions. The destabilization of PEG solutions is greater by hydrophilic than by hydrophobic proteins and by negatively-charged than by positively-charged ones.

Phase separation of PEG-protein solutions has been reported by Alber et al (1981) who observed an "oil phase" in which protein crystals grew. However, although the "oil" phase contained all the protein, it appeared to have virtually the same PEG concentration as the "aqueous" phase. A similar "oiling out" phenomenon was observed by Dr.R.Aschaffenburg (personal communication) in solutions of PEG and β -lactamase II (see below).

The Fv crystals also provided a confusing observation. The density of the crystals was measured on a gradient formed by xylene and bromobenzene (Low and Richards,1952). A volume of xylene was layered on to an equal volume of bromobenzene in a graduated cylinder. The interface was gently mixed by a plunger motion with a Pasteur pipette, and two interfaces formed with the gradient in between. Solutions of various concentrations of CsCl were used to calibrate the gradient. A crystal was blotted as dry as possible and was applied to the column. The point at which the crystal equilibrated was noted.

The results for Fv and two well-characterized protein crystals,

Protein	Crystal Density (g/ml)±0.03	Partial Specific Volume (ml/g)	Density of Solvent(g/ml) Calculated Measured	Solvent	"Thickness" of Solid PEG Coat (5)
Fv	1.27	0.724	1.13 1.04	13%(w/v)PEG(6000)	2.3x10 ⁻³ mm
β2(PEG)	1.30	0.74 ⁽¹⁾	1.23 1.04	10%(w/v)PEG(6000)	8.3x10 ⁻⁴ mm
β2(salt)	1.12	0.74 ⁽¹⁾	1.01 1.01	1mM ZnSO ₄	-
HEW Lysozyme	1.25	0.721 ⁽²⁾ 0.688 ⁽³⁾	1.03 0.96	- 5%(w/v) NaCl	-
Haemoglobin ⁽⁴⁾	1.16	0.752	1.01 -	2.8M (NH ₄) ₂ SO ₄	-
(1) Assumed					
(2) from CRC Handbook of Biochemistry and Molecular Biology					
(3) from Tanford (1961)					
(4) all values from Perutz (1946)					
(5) see text					

Figure IV.A4. Crystal density measurements and subsequent calculations.

hen-egg-white lysozyme and haemoglobin, are presented in figure IV.A4. Also shown are results from two crystal forms of β -lactamase II from Bacillus cereus, one grown in the presence of PEG, the other in a dilute salt solution.

From the crystal density, the volume of the unit cell, and the molecular weight of the protein, the density of the solvent within the crystal can be calculated in the following way. The asymmetric unit in space group C2 occupies one quarter of the unit cell. Hence, the mass of the asymmetric unit contents can be determined directly from the crystal density. The mass of protein in the asymmetric unit depends on the number of molecules present and the calculation was performed with values of one, two and three Fv fragments. The mass not accounted for by the protein was assigned to the solvent. The partial specific volume of Fv (Hochman et al, 1973) was used to calculate the volume which the mass of protein would be expected to occupy, and the volume of the asymmetric unit available to solvent was found. After consideration of the fraction occupied by protein, it was concluded that the asymmetric unit contains two Fv fragments (see figure IV.A5). As can be seen in figure IV.A4, the calculated solvent densities of the two crystal forms grown from PEG were considerably greater than the measured densities of the solvents from which the crystals were grown. The solvent densities for salt-grown crystals calculated in the same way were, as expected, close to 1.0 g/ml.

The origin of this unexpected result remains a mystery. In contrast to the theory of Lee and Lee described above, it appears that PEG is associated with the crystals. This can be envisioned in two ways, small polymers of ethylene glycol entering the solvent channels of the crystal, or a coat of PEG surrounding the crystal. The mass of PEG which accounts for the density anomaly can be calculated, and the thickness of solid

Contents of Assymetric Unit Number of Fv Fragments	Protein Molecular Weight	% of Crystal Weight	Occupied Volume	$V_m(\text{\AA}^3/\text{dalton})^{(1)}$	Calculated Solvent Density within Crystal(g/ml)
1	25250	28 ± 1	26.0±0.2	4.71±0.01	1.24±0.05
2	50500	55 ± 1	52.0±0.2	2.36±0.01	1.13±0.05
3	75750	83 ± 1	78.0±0.2	1.57±0.01	0.98±0.05

(1) V_m =Volume of assymetric unit / protein molecular weight in assymetric unit
(Matthews,1968). Expected range for protein crystals, 1.68 to 3.53.

Figure IV.A5. Examination of the amount of protein in the Fv assymetric unit.

PEG coating an average sized Fv crystal ($0.5 \times 0.2 \times 0.2 \text{ mm}^3$) estimated. The results, included in figure IV.A4, indicate that the coat would not be visible under the microscope.

Another observation related to this mystery was made on crystals of both Fv and β -lactamase II. In six to eight weeks after a crystallization tube was unsealed, the crystals therein became disordered and resistant to dissolution in water. The latter behaviour is similar to that of crystals cross-linked in the traditional way (Quiocho and Richards, 1964), though an increasingly thick deposit of PEG on the crystal surface also provides an explanation.

It is clear that the effects of PEG on protein solutions and crystals are not understood. An examination of this interaction would be worthwhile if it led to a more definite approach to the growing of protein crystals.

IV.B. The Search for Isomorphous Heavy-Atom Derivatives

The method for examination of a heavy-atom compound was to dissolve it initially to a concentration of 5 mM in the PEG buffer from which the native crystals were grown. A small crystal or fragment was immersed in the solution and observed for physical changes. If it was stable, several larger crystals were added to the soak. Crystals were removed and mounted for x-ray photography after suitable periods, for example, 7, 10 and 14 days. The diffraction patterns were examined for intensity changes of some reflections, which indicate a change in the diffracting material (the addition of a heavy atom), and for cell dimension changes which indicate non-isomorphism. The results sometimes led to a repetition of the experiment under different conditions, for instance shorter or longer soak times or a change in heavy-atom concentration. The main problem encountered by this method was the insolubility of many heavy-atom compounds in the PEG solution. A list of several of the soluble compounds used is given in

<u>Compound</u>	<u>Observations</u>
$K_3UO_2F_5$	good intensity changes, isomorphous
$Pt(NH_3)_2Cl_2$	good intensity changes, isomorphous
$CdCl_2$	intensity changes, tendency for lack of isomorphism especially at 10mM or more
$Gd(NO_3)_3$	low concentration(1mM) no changes high concentration, crystal quickly disintegrates
$La(NO_3)_3$	similar behaviour to $Gd(NO_3)_3$
Ethyl mercuric phosphate	no clear intensity changes
$HgCl_2$	no clear intensity changes (barely soluble)
K_2PtCl_4	no clear changes except at low angle
$K_2Pt(CN)_4$	no clear changes
$CsCl$	no clear changes
$MnCl_2$	no clear changes
K_2OsO_4	no clear changes
$Th(NO_3)_4$	crystal breaks up after about 3 weeks
Haptens:	
Dnp-L-aspartic acid	no clear changes
Dnp-2aminoethyltriphosphate	crystal unstable, no changes
	+ $MnCl_2$ possible small intensity changes not reproducible

Figure IV.B.1 Some of the soluble Heavy-atom compounds used in the search for isomorphous derivatives.

figure IV.B1.

Fv is known to have a binding site for lanthanides (see section II.B3), so the effects of gadolinium nitrate were investigated. At concentrations of greater than about 1 mM the crystals rapidly broke up, while at 1 mM they seemed to be indefinitely stable but showed no intensity changes. The behaviour of crystals in the presence of lanthanum nitrate was very similar. Therefore, it is likely that either the lanthanide binding site is very close to a point of contact between Fv molecules in the crystal, or that there are conformational changes in the Fv on binding which corrupt the crystal lattice.

The binding of Dnp haptens was also investigated by this method. The compound Dnp-2-aminoethyl-triphosphate (see section II.B1) was used in nuclear magnetic resonance studies to create a binding site for manganese atoms. In general, Dnp compounds are bright yellow, so their integration into the crystal can be followed to some extent by eye. Neither the hapten nor MnCl_2 alone resulted in noticeable intensity changes, but together they gave promising results. Soaking studies of dinitroaniline, a smaller hapten which is expected to lie completely within the combining site, were also made. However, it was decided not to use haptens as mediators for heavy-atom binding for the purpose of deriving phase information, because this might influence hapten-binding studies by difference Fourier techniques.

Initially two, $\text{Pt}(\text{NH}_3)_2\text{Cl}_2$ and $\text{K}_3\text{UO}_2\text{F}_5$, and later a third, CdCl_2 , derivatives were found (figure IV.B2). The soak times are in general rather long, presumably because of the viscous nature of the PEG solutions.

It is difficult to guess where these heavy atoms have bound to Fv, although all three compounds are quite different in their reactivities with proteins. The uncharged $\text{Pt}(\text{NH}_3)_2\text{Cl}_2$ can penetrate hydrophobic regions

<u>Compound</u>	<u>Soaking Conditions</u>
$\text{Pt}(\text{NH}_3)_2\text{Cl}_2$	10mM, more than 7 days.
$\text{K}_3\text{UO}_2\text{F}_5$	10mM, more than 7 days.
$\text{CdCl}_2(1)$	7.5mM or less, at least 13 days.

(1) Later found to be non-isomorphous (section IV.D1).

Figure IV.B2. Soaking conditions for the heavy-atom derivatives used in data collection. The crystals were stable in the heavy-atom solutions, and no increase in intensity changes was observed, after the times shown.

of the protein and attack equally well anions and neutral groups on amino-acid residues (Petsko et al,1978). The $(\text{UO}_2)^{2+}$ cation generally binds to carboxylate groups and hydroxyl side-chains. Uranyl compounds have an affinity to sites which also bind lanthanides (Blundell and Johnson,1976). Therefore, it is possible that if uranyl binding does not induce the conformational change observed in the presence of gadolinium, it may have interacted with the lanthanide binding site. The Cd^{2+} ion has not commonly been used to form heavy-atom derivatives but it is likely to behave in a similar way to the Pt^{2+} ion, the binding affinity of which is very sensitive to the soaking conditions (Petsko et al,1978).

IV.C. Native Data to 6Å Resolution

1. Collection.

The procedure described here for the collection of three-dimensional diffraction data from native Fv crystals was carried out in essentially the same way for crystals of the three derivatives. Minor variations in the procedure will be dealt with later.

A Philips fine focus x-ray tube operating at 40 kV, 30 mA produced $\text{CuK}\alpha$ radiation (wavelength 1.5418Å) which was passed through a nickel filter and a 1.0 mm collimator before reaching the crystal. The specimen was mounted in a sealed glass capillary tube as described earlier, placed on a standard goniometer head, and attached to a Hilger and Watts four-circle diffractometer controlled by a Digital Equipment Corporation PDP-8 computer. The diffracted x-rays passed through a second collimator, the diameter of which was chosen to maximize the signal-to-noise ratio of the peak profile. The geometry of the four-circle diffractometer was reviewed in section III.D.

Diffracted rays were detected in a proportional counter and the amplified pulse was then passed through a pulse-height analyser which

filtered out unwanted noise.

The arcs of the goniometer head were adjusted so that a reciprocal axis, in general c^* , was co-incident with the ϕ -circle axis (see figure III.9). This was accomplished roughly by centring the peak from a very low angle reflection (0 0 4) about the expected diffraction position (calculated from the cell dimension), and then improved with a higher angle reflection (0 0 16). In the course of this procedure, the value of the c^* cell dimension was refined. The arc adjustment was checked by examination of the (0 0 -16) reflection.

Low angle reflections along the other two axes ((0 2 0) and (4 0 0)) were located by rotating the diffractometer circles to the appropriate positions, and an initial orientation matrix was calculated. As the circle settings for higher angle reflections were determined, the cell dimensions and, hence, the orientation matrix were refined.

To examine a reflection, the ω -circle was stepped through the diffracting position in, typically, 16 or 20 steps of 0.04° . The number of steps depended on the width of the broadest reflection; in general, an equal number of measurements were made of background and peak. The profile of each peak was recorded initially on paper tape and then transferred to magnetic tape on another computer (Ferranti Argus 500). The derivation of intensities by the ordinate analysis method was described in section III.D and was performed on the Oxford University 1906A computer.

The $6\text{\AA}$ resolution data were collected in two shells, the inner, $7.5\text{\AA}$ to the minimum accessible angle (about $55\text{\AA}$ resolution), and the outer, 6.0 to $8.0\text{\AA}$ resolution. In space group $C2$, each reflection hkl has three equivalent reflections, $\bar{h}k\bar{l}$, $\bar{h}\bar{k}l$ and $h\bar{k}l$. Therefore, one unique set of data includes only one quarter of the diffraction sphere. In addition, any reflection of which $h + k$ is an odd number is a systematic absence and

Key
Native Data x
pt Data o
Ur Data ▲

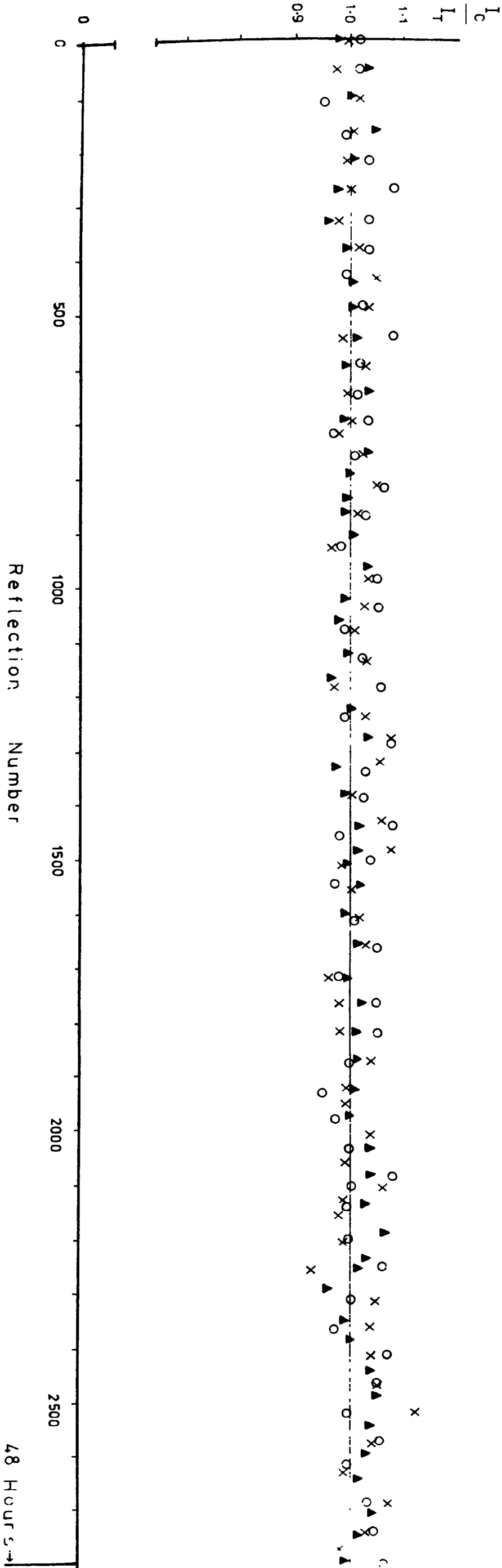


Figure IV.C1. The intensity of the (0 12 0) reflections throughout the course of the native and derivative data collection. $I(0)$, initial intensity, $I(T)$, intensity at time T .

was not measured. In the native data collection, a unique set of reflections (those with $+h, +k, \pm l$) was measured in both shells and, as no radiation damage was observed (see figure IV.C1), a set of equivalent reflections $+h, -k, \pm l$ was recorded.

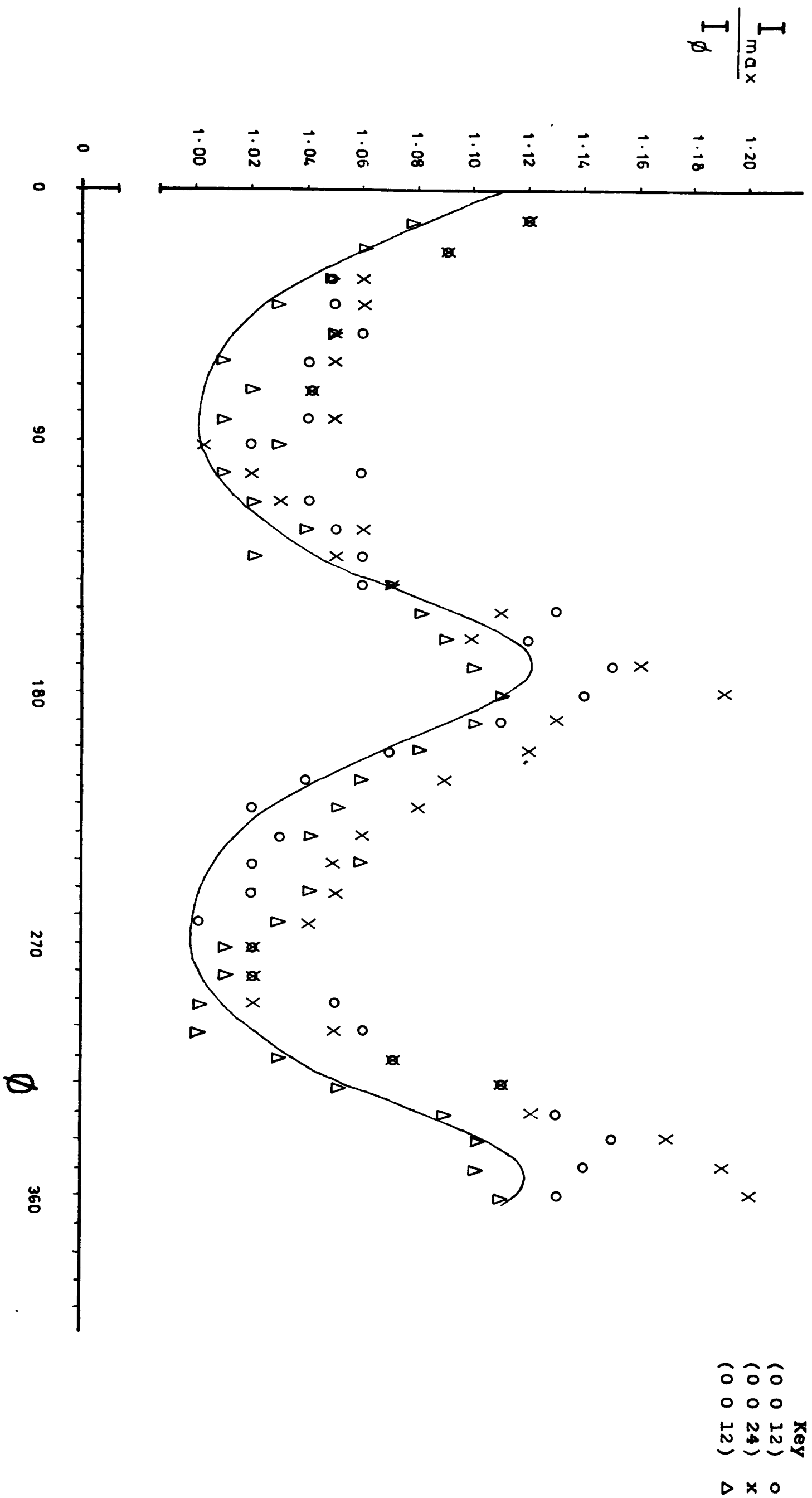
During data collection, radiation damage was monitored by the re-measurement of three standard reflections, $(16\ 0\ 0)$, $(0\ 12\ 0)$ and $(0\ 0\ 12)$, after approximately every fifty data measurements. Figure IV.C1 shows that there was no detectable decrease in the intensity of the $(0\ 12\ 0)$ reflection throughout the course of data collection. The other standard reflections showed similar behaviour and no correction was made to the data for radiation damage.

The effects of absorption by the crystal were measured by the method of North, Philips and Mathews (1968) described in section III.D. An absorption curve was taken at 10° intervals of ρ for three axial reflections $(0\ 0\ 12)$, $(0\ 0\ -12)$ and $(0\ 0\ 24)$, ($11.5\text{\AA}$, $11.5\text{\AA}$ and $5.75\text{\AA}$ resolution, respectively). The measurements are shown in figure IV.C2 with the curve corresponding to the applied correction, obtained by averaging the absorption at values of ρ separated by 180° . Thus, reflections of the form hkl and $h\bar{k}l$ received the same absorption correction.

2. Native Data Processing.

The ordinate analysis and absorption correction were performed on each resolution shell separately by the program PRIMROSE (appendix I). Account was also taken of the Lorentz and Polarization effects (section III.D). Finally, a standard deviation was assigned to each reflection based on the total number of counts.

The program SSM (appendix I) was used to calculate scale factors between the shells by the comparison of common equivalent measurements and the data were sorted and merged into a unique set. In the process of this



comparison, the standard deviation for each term was modified to be a better estimate of the error (French and Wilson, 1978). Figure IV.C3 shows that the agreement between multiple measurements of equivalent reflections was good.

IV.D. Data Collection from Platinum and Uranyl Derivatives

Because anomalous scattering information was being recorded from derivative crystals, twice as many reflections were required to complete one unique set of data. Strictly, Friedel equivalents (which show anomalous differences) are hkl and $\bar{h}\bar{k}\bar{l}$. In space group $C2$, however, a reflection $\bar{h}k\bar{l}$ has the same phase as hkl , whereas $h\bar{k}\bar{l}$ has the same phase as $\bar{h}\bar{k}\bar{l}$. Therefore, reflections with indices $(h \pm k \pm l)$ were measured and the effect of anomalous scattering was retained by specifying that the reflections hkl and $\bar{h}\bar{k}\bar{l}$ were not equivalent. The anomalous difference in intensity is relatively small compared to the intensity itself. It is therefore essential to make the conditions under which Friedel mates are measured, as similar as possible.

Any unequal effects due to radiation damage were minimized by altering the native data collection strategy such that Friedel equivalents were measured close together in time. Measurements of reflections on a level of k with the h index varying between the resolution limits of the shell, was followed immediately by that of the equivalent level $\bar{k}$. The unique data set was thus divided into four parts, two halves, $(h \pm k \pm l)$ and $(h \pm k \bar{l})$, within each of the two resolution shells. Both halves of the outer $(6.0 - 8.0\text{\AA})$ shell were collected before proceeding to the inner $(7.5\text{\AA} - \infty)$ shell.

Unfortunately, the natural mount of Fv crystals is with the long axis of the crystal, c , parallel to that of the capillary tube (c^* along the axis of rotation, ϕ). The reflections hkl and $\bar{h}\bar{k}\bar{l}$ therefore suffer

Data Set	1	2	3	4	5
Equivalent	hk±1	h $\bar{k}$ ±1	hk±1	hk±1	h $\bar{k}$ ±1
Resolution Shell(Å)	6.0-8.0	6.0-8.0	7.5-∞	7.5-∞	7.5-∞
Maximum Absorption ⁽¹⁾	1.20	1.20	1.15	1.15	1.15
Number of Terms	655	774	434	187	816
Number Rejected ⁽²⁾	0	0	0	3	1
Fox and Holmes Scale ⁽³⁾	1.00	1.04	1.05	1.24	1.06
Merging R-factor ⁽⁴⁾	0.032	0.027	0.024	0.040	0.025
Total Reflections			2862		
Independent			1216		
Overall Merging R-factor			0.0282		

- (1) At Value of ρ where $I_{\max}/I_{\rho}$ is largest.
- (2) A term is rejected when the difference between the measurements of two (or more) equivalent reflections is greater than twice the standard deviation. In this instance, the larger measurement is accepted rather than the mean.
- (3) Fox and Holmes (1966).
- (4) Merging $R = \sum (\sum \langle I \rangle_n - I_N / \sum I_N)$
 where n is the number of sets of N equivalent measurements (I_N) of mean intensity $\langle I \rangle$. This is the mean of the merging R-factors of a data set with each of the other four sets.

Figure IV.C3. Statistics of the Merging of the Native 6Å Resolution Data into a Unique Set.

different absorption properties and a careful absorption correction was important. The trend of absorption with ρ was the same as that of the native crystals (see figure IV.C3), however the maximum absorption was markedly higher in the derivative crystals (figure IV.D1).

The procedure for the reduction of the measurements to F^2 and merging into a unique set was described above. Merging statistics for the two derivatives are presented in figure IV.D1.

The program CAD (appendix I) was used to amalgamate the two derivative and the native data sets into one list of reflections. The square roots of the intensities were taken to relate them directly to the structure factor amplitudes and anomalous information was included by taking the difference between the F's of Friedel mates.

The derivative data were then placed on the same relative scale as the native, and a detailed analysis of the structure factor amplitudes was performed as a function of each of the indices, the resolution, and the angle of rotation, ρ (ANSC, appendix I). The final analyses after the application of local scale factors where appropriate, are shown in figure IV.D2.

Once the similarity between the data sets had been maximized, the remaining differences, presumably due to the effects of the bound heavy atoms, could be observed in terms of the mean fractional isomorphous difference ($\sum (F_{PH} - F_P) / \sum F_P$). The analysis as a function of resolution (figure IV.D3) shows that, in general, the isomorphous differences of both derivatives were larger at low angle, and no increasing trend at higher resolution, which would indicate a lack of isomorphism, is present.

1. Data Collection from the Cd Derivative.

Data from the $CdCl_2$ were collected from a crystal mounted about the

a. Platinum Derivative				
Data Set	1	2	3	4
Equivalent	$h\pm k l$	$h\pm k \bar{l}$	$h\pm k l$	$h\pm k \bar{l}$
Resolution Shell($\text{\AA}$)	6.0–8.0	6.0–8.0	7.5– ∞	7.5– ∞
Maximum Absorption ⁽¹⁾	1.53	1.53	1.53	1.53
Number of Terms	799	662	729	603
Number Rejected ⁽¹⁾	0	0	1	0
Fox and Holmes Scale ⁽¹⁾	1.00	0.98	1.04	1.04
Merging R-Factor	0.035	0.036	0.024	0.025
(2) Total Reflections			2792	
Independent			2300	
Overall Merging R-factor			0.0278	

b. Uranyl Derivative				
Data Set	1	2	3	4
Equivalent	$h\pm k l$	$h\pm k l$	$h\pm k \bar{l}$	$h\pm k \bar{l}$
Resolution Shell($\text{\AA}$)	6.0–8.0	7.5– ∞	6.0–8.0	7.5– ∞
Maximum Absorption ⁽¹⁾	2.16	2.16	2.16	2.16
Number of Terms	790	664	724	608
Number Rejected ⁽¹⁾	0	0	0	0
Fox and Holmes Scale ⁽¹⁾	1.00	0.99	0.94	0.98
Merging R-factor ⁽¹⁾	0.028	0.025	0.029	0.028
(2) Total Reflections			2786	
Independent			2290	
Overall Merging R-factor			0.0278	

(1) See Figure IV.C3.

(2) Includes Friedel Equivalents as Separate Reflections.

Figure IV.D1. Merging Statistics of the Platinum and Uranyl 6 $\text{\AA}$ Data.

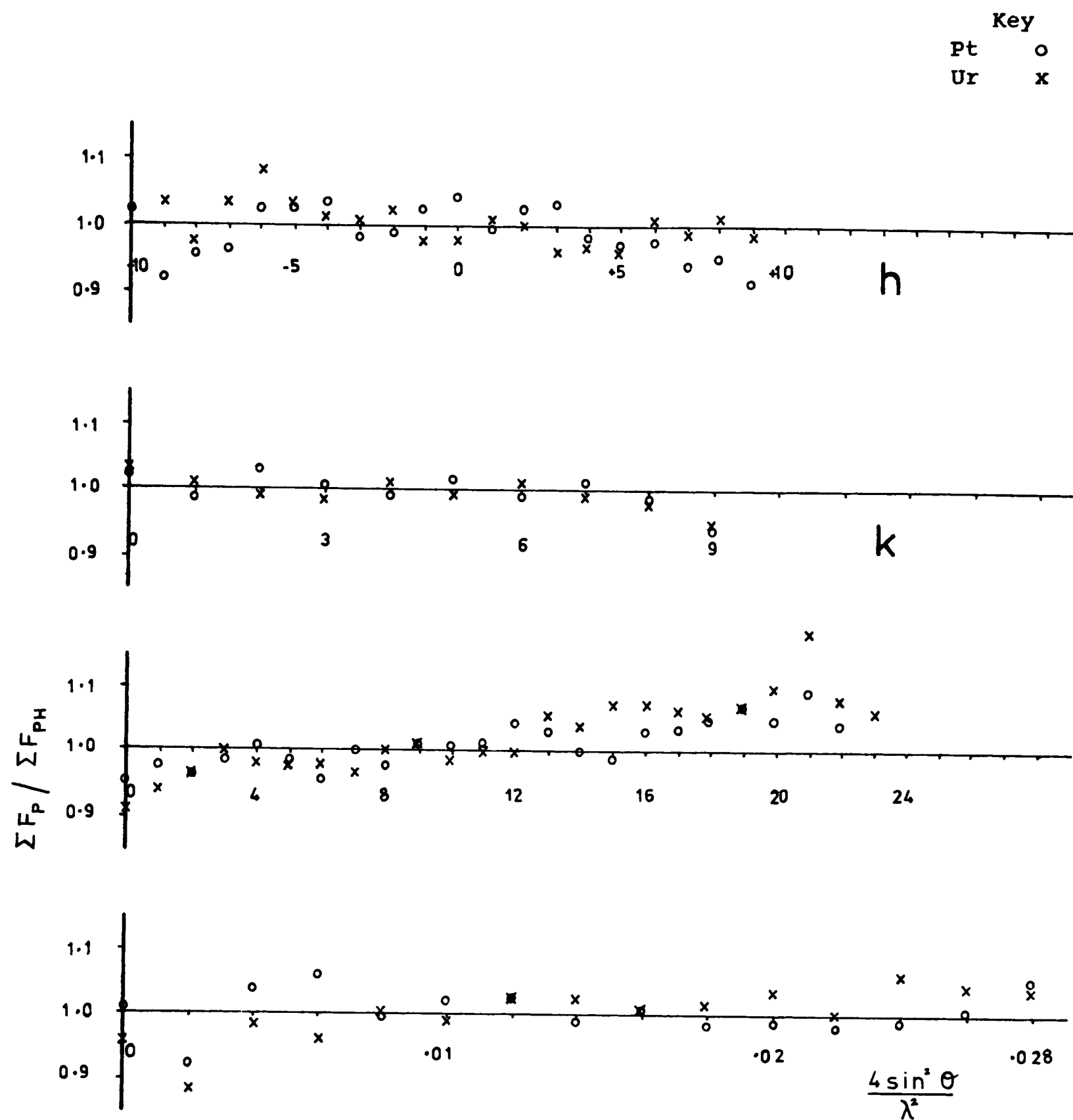


Figure IV.D2. Analysis of the platinum and uranyl derivative F's with respect to the native as functions of the h, k and l indices and resolution.

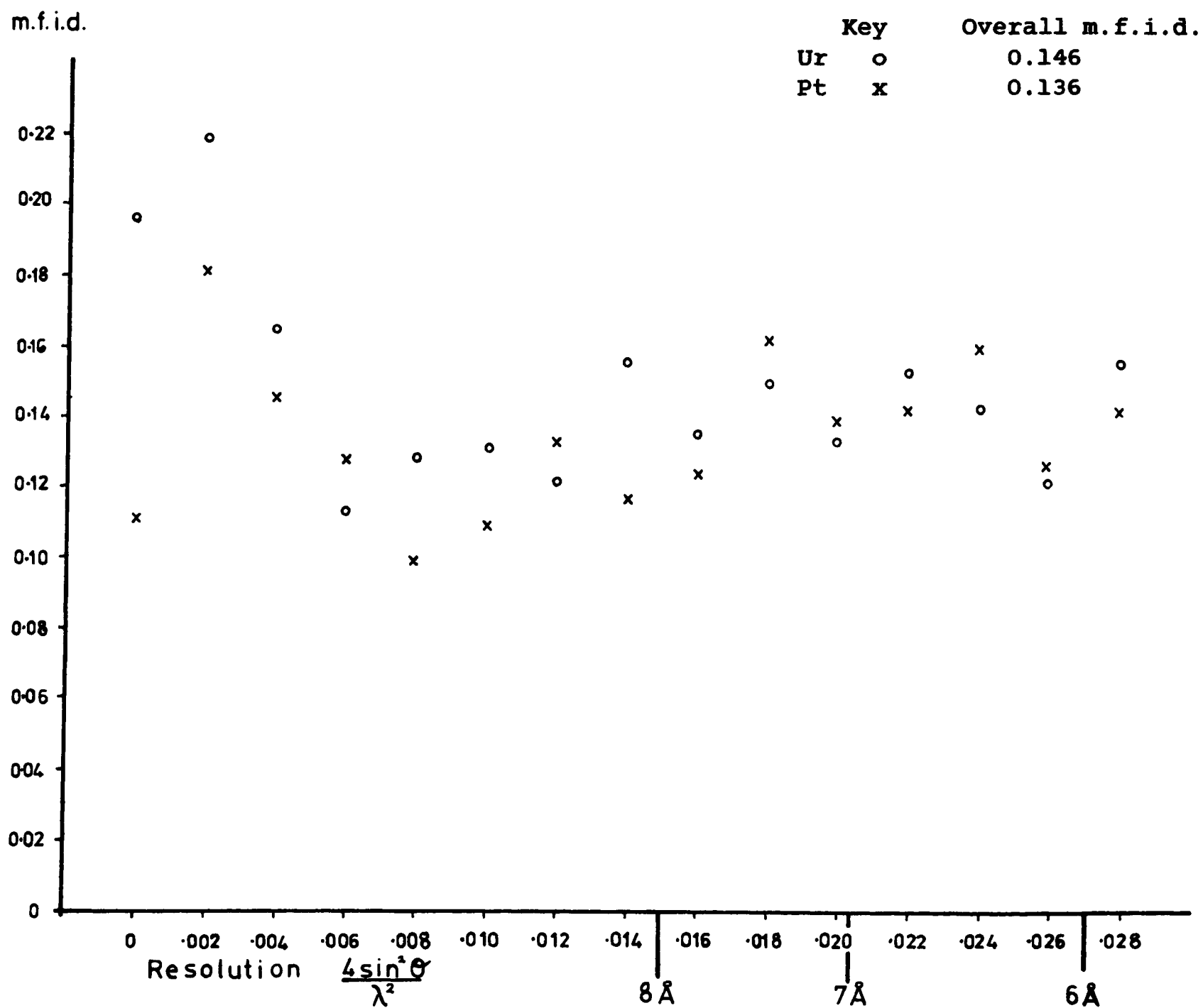


Figure IV.D3. The variation of the mean fractional isomorphous difference (m.f.i.d.) $[\sum(F_{PH} - F_P)/\sum F_P]$ of the platinum and uranyl derivatives with resolution.

b/b* axis. In this orientation, the absorption corrections applied to the hkl and $h\bar{k}l$ reflections were the same. Otherwise, the data were collected and processed in the same way as the those from the other two derivatives.

It soon became clear, however, that the quality of the cadmium derivative was much poorer than that of the platinum and uranyl ones. Figure IV.D4 shows that the mean fractional isomorphous difference rises sharply with resolution.

The difference Patterson synthesis was very noisy with no outstanding peaks. A difference Fourier synthesis calculated with platinum and uranyl-derived phases also showed no clear sites of cadmium binding. Thus, no useful phase information could be gained from the cadmium derivative.

IV.E. The Isomorphous Difference Patterson Functions

In section III.E, the Patterson function was defined and the use of difference Patterson techniques in the determination of the sites of heavy-atom binding was discussed. The syntheses for the platinum and uranyl derivatives are presented in figures IV.E1 and IV.E2. All the self-vectors (cross-vectors between a site and its symmetry-related position) occur on the Harker section, $v = 0$. As expected, the largest peak occurs at the origin, where the cross-vectors between each site and itself are superimposed.

Both the platinum and uranyl Pattersons were interpreted initially in terms of two heavy-atom sites. Only the y-coordinate difference between two sites can be found from the Patterson. Therefore, the y-coordinate of one was arbitrarily set to 0.0 and the other site was defined with respect to it. A third site for each arose from the refinement procedures outlined in the following section, and the cross-vectors between the sites are indicated in figures IV.E1 and IV.E2.

m.f.i.d.

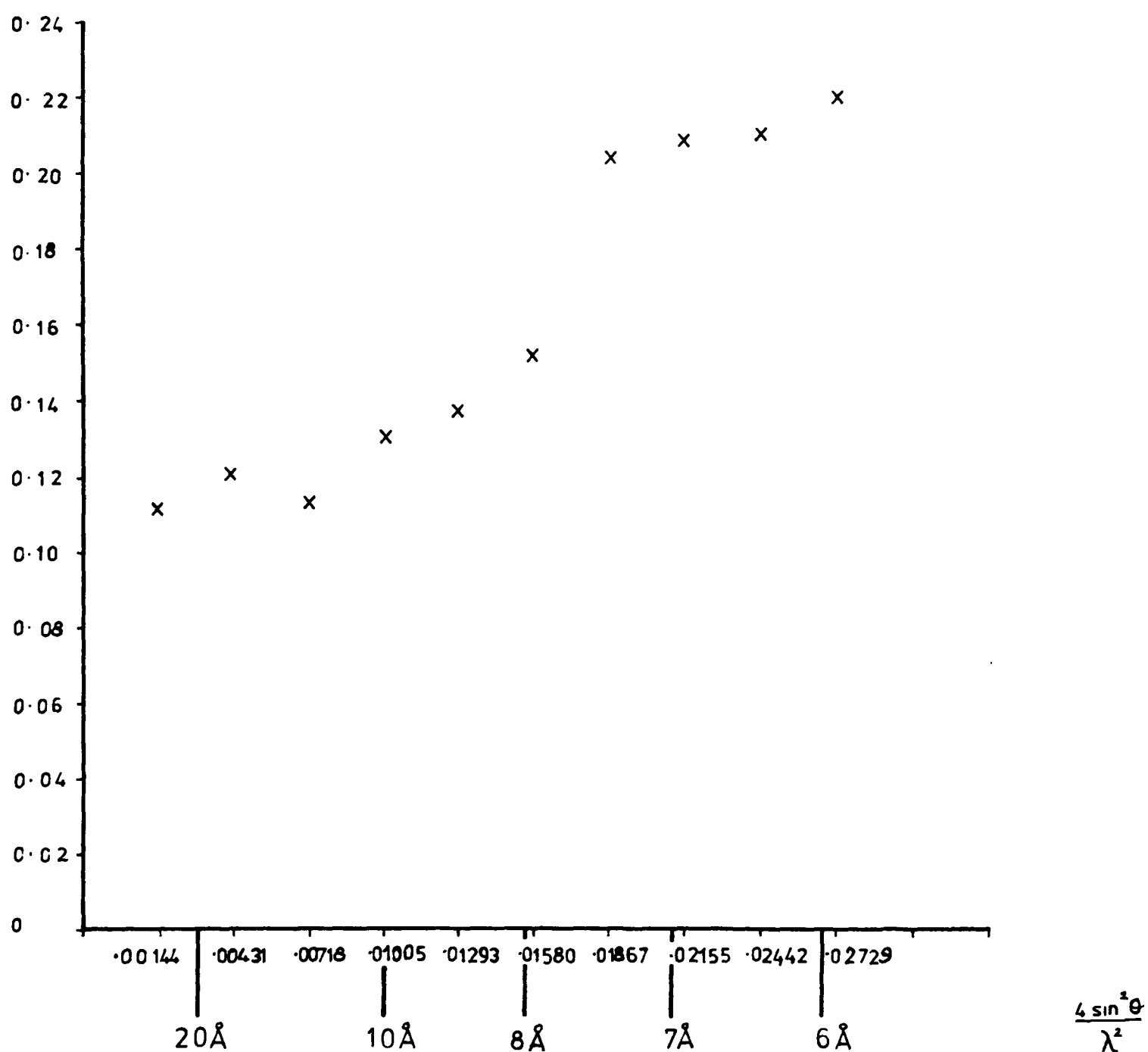


Figure IV.D4. The variation of the mean fractional isomorphous difference (see figure IV.D3) of the cadmium derivative with resolution. Overall m.f.i.d. 0.1616.

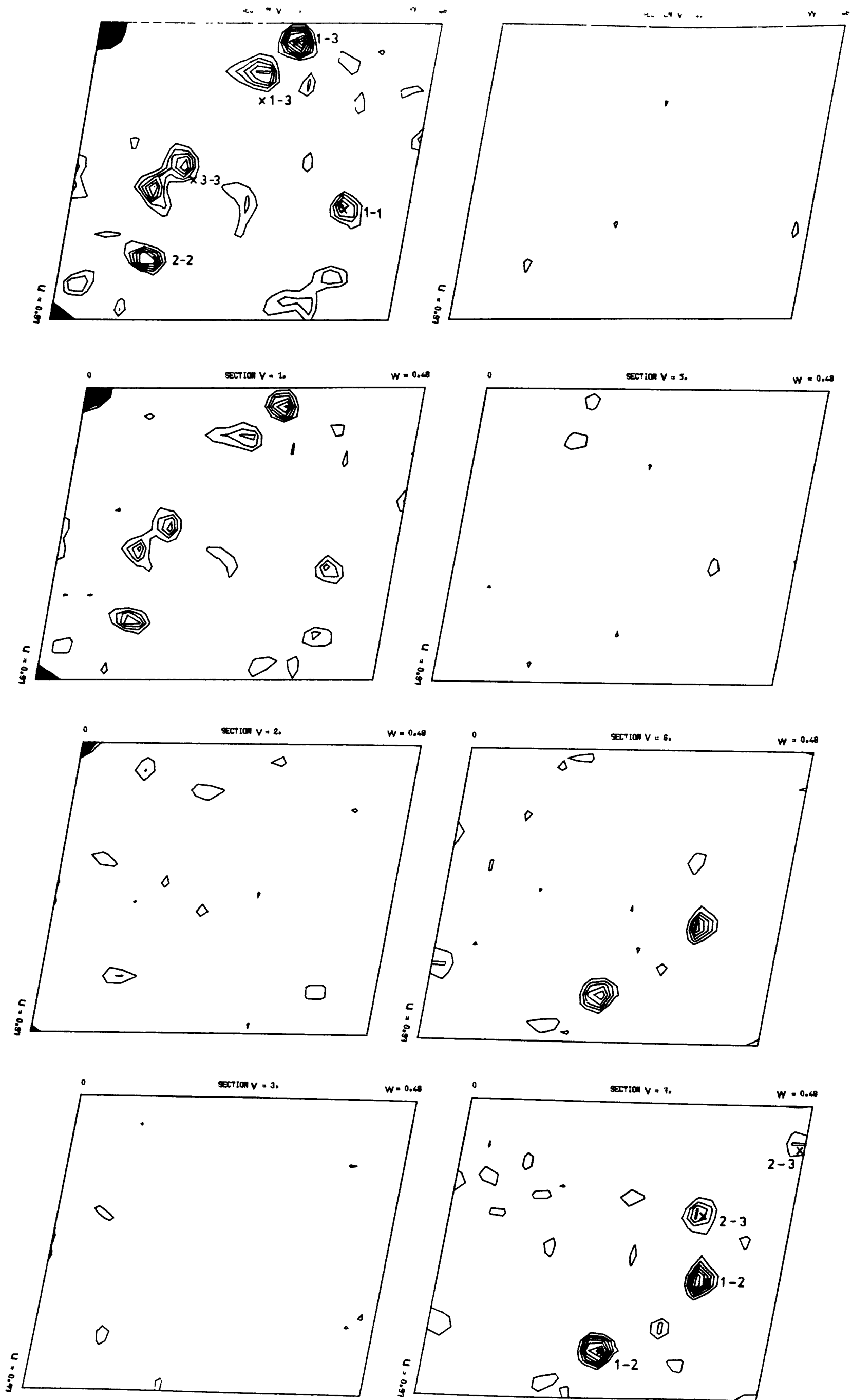


Figure IV.E1. The Platinum Difference Patterson.
 Origin Peak=[99], contour interval [1] starting at [6]

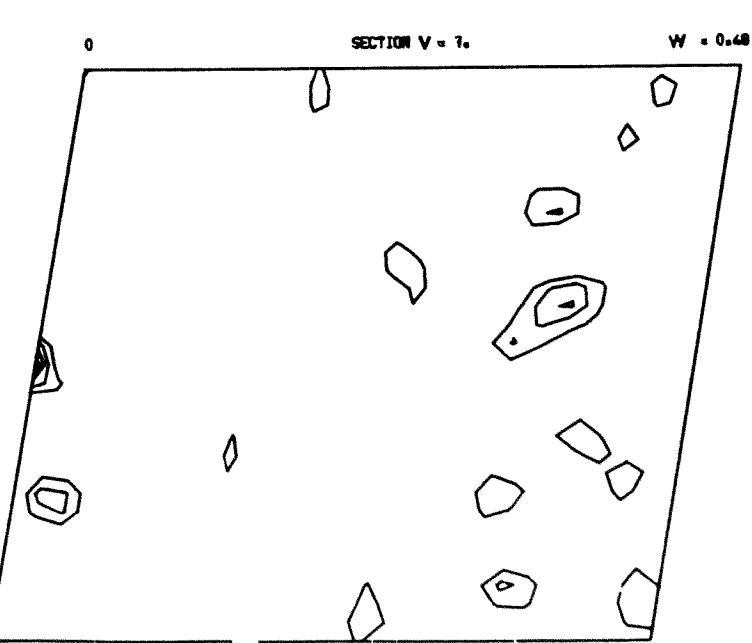
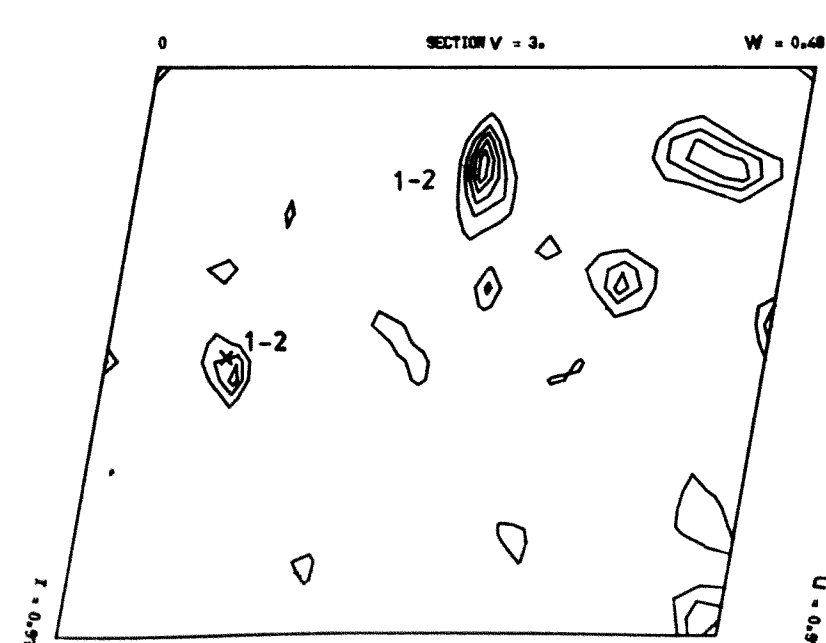
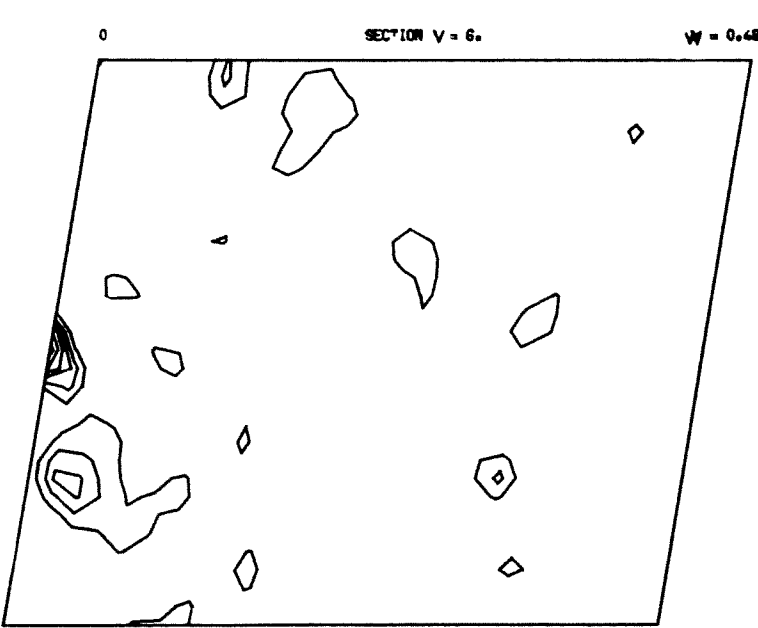
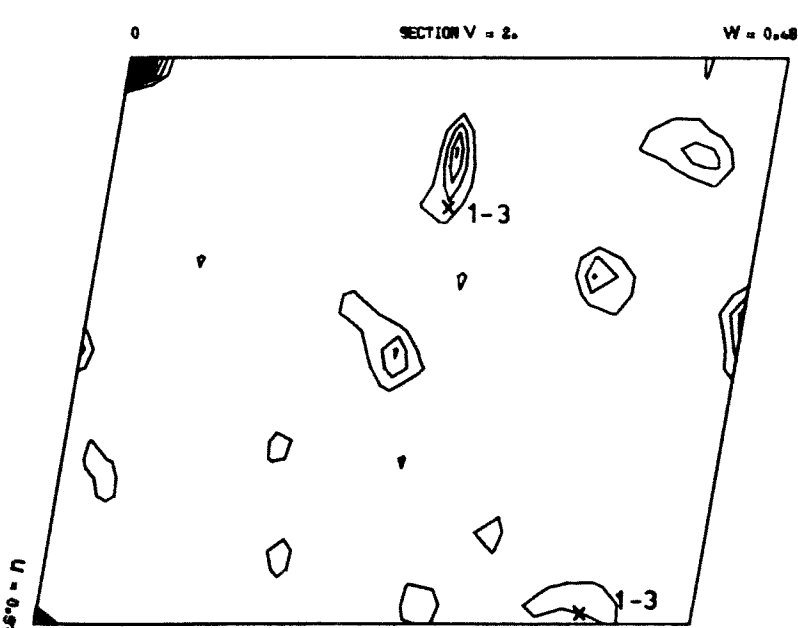
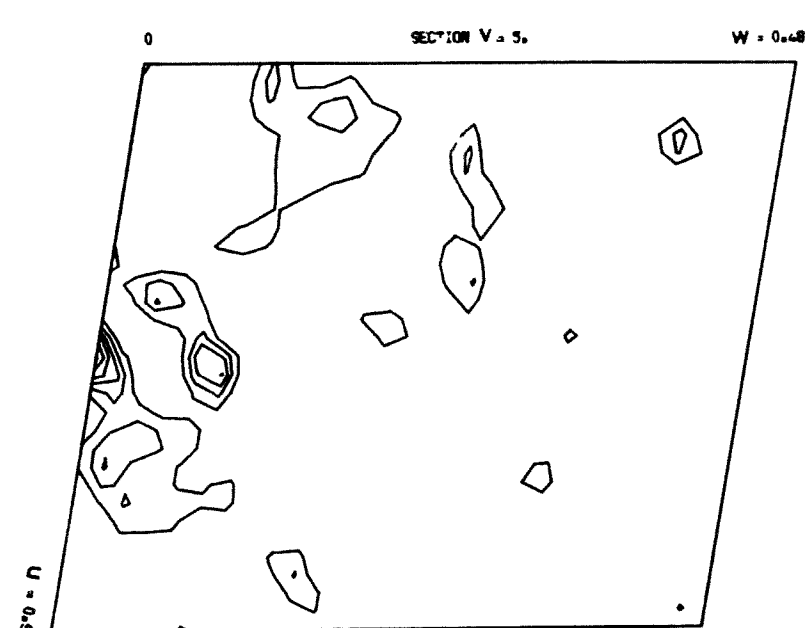
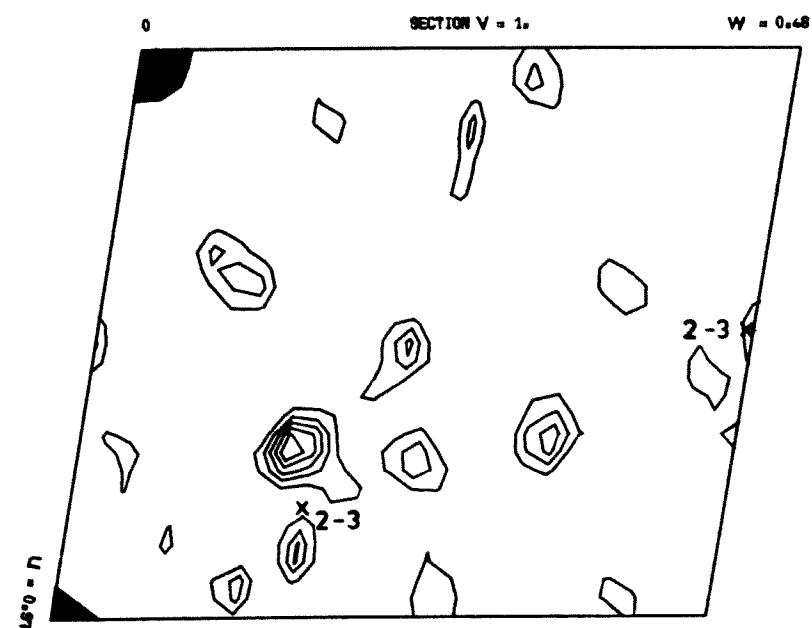
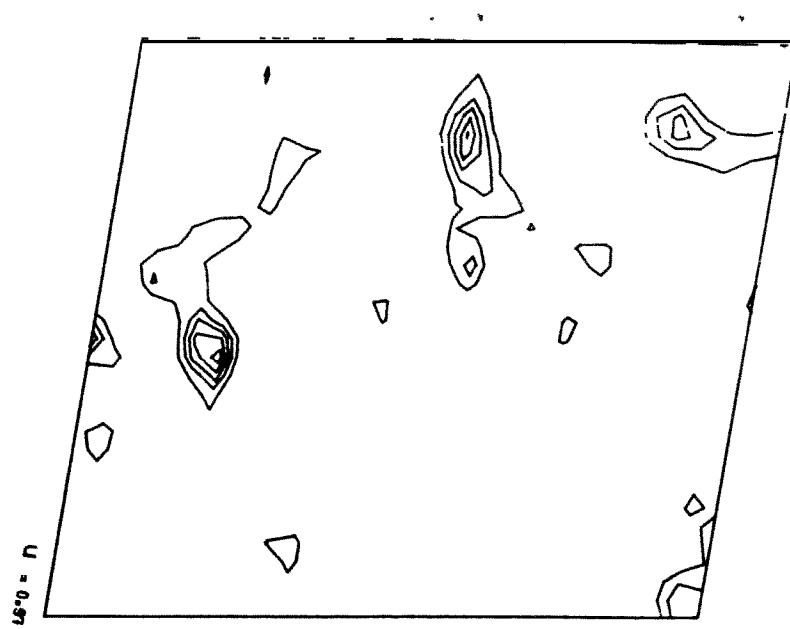
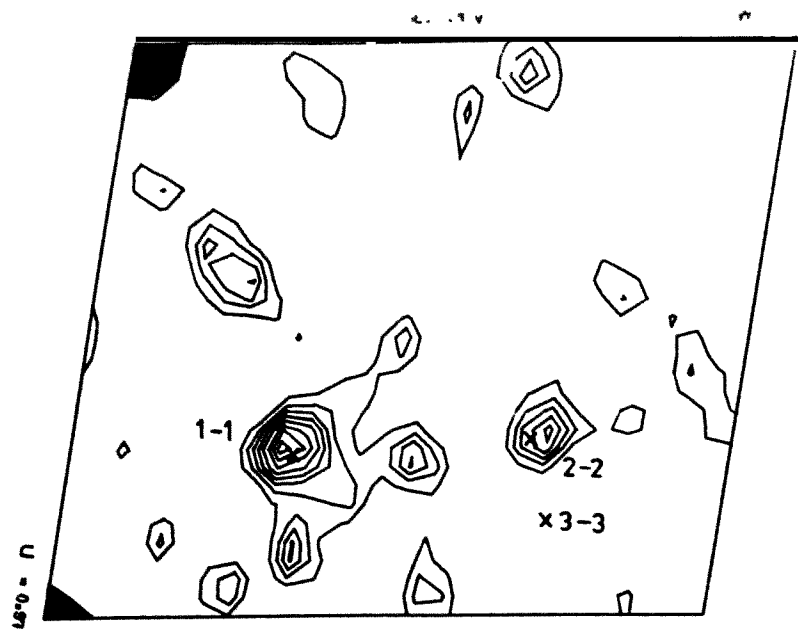


Figure IV.E2. The Uranyl Difference Patterson.
Origin Peak=[99], contour interval [1] starting at [6].

IV.F. Preliminary Refinement of the Heavy-Atom Sites

The starting point for the refinement was an estimate of the heavy-atom structure factor, F_{HLE} (section III.E), for each reflection. Figure IV.F1 shows some details of the F_{HLE} calculation. This procedure tends to overestimate F_H for a number of reasons (Dodson et al, 1975), not least of which is the comparatively small size of the anomalous differences. A "bias" correction based on the standard deviations (σ) of the isomorphous and anomalous differences ($\sigma_{iso}^2 + (k^2/4)\sigma_{anom}^2$, k as in equation III.E4) compensates to some extent for this error. In addition, reflections were rejected for which the upper estimate of F_H was more likely to be correct than F_{HLE} , although it is likely that some incorrect estimates remained. Because of this, the refined parameters may overestimate the heavy-atom presence.

In addition to the atomic coordinates, each heavy-atom position was assigned an occupancy factor to compensate for the fact that every protein molecule in the crystal may not have a bound heavy atom at that position, and that some of the native protein or solvent atoms around the position may have been displaced. As the former contribution is dominant, a fully-substituted position on an absolute scale, would have an occupancy about 1.0. However, because the data are not on an absolute scale, only relative occupancies of positions can be refined; an occupancy of 1.0 or larger does not indicate full substitution.

In the scaling procedure followed during the data processing, the derivative and native data were placed on the same relative scale. In fact, however, there is slightly more scattering material in the derivative due to the heavy-atom presence. A scale factor of slightly greater than 1.0 to be applied to each derivative was calculated simultaneously with the F_{HLE} values and iteratively refined.

	<u>Uranyl</u>		<u>Platinum</u>	
	F_{HLE}	$F_{HLE} - BIAS^{(1)}$	F_{HLE}	$F_{HLE} - BIAS$
(2) Scale factor	1.16	1.083	1.034	1.030
(3) $\Delta(iso)$ acentric	92.32		164.3	
centric	138.90		199.1	
(4) $\Delta(anom)$	27.5		52.2	
(5) $K(emp)$	6.72		6.30	
mean F_{HLE} :				
acentric	233.3	195.7	261.76	226.79
centric	138.9	138.9	199.12	199.12
(6) weighted mean F_{HLE} :				
acentric	214.3	252.8	298.25	272.48
centric	148.7	148.7	195.95	195.95
(7) number accepted		802		1054
% of reflections		67.4		88.6

(1) F_{HLE} corrected by the bias term (see text).

(2) Derivative-to-native.

(3) $\Delta(iso) = |F_{PH} - F_P|$, averaged over all reflections.

(4) $\Delta(anom) = |F_{PH(+)} - F_{PH(-)}|$, averaged over all reflections.

(5) $K(emp) = 2\Delta(iso)/\Delta anom$.

Theoretical $K((f + f')/f'')$ is 13.4 (Ur) and 5.96 (Pt).

(6) Weight applied to each term = $1/\text{Variance}(F_{HLE})$ (Dodson et al, 1975)

(7) Rejection criteria: $K(|F_{PH(+)} - F_{PH(-)}|)/2F_P > 1.10$ (Ur)
 > 1.20 (Pt)

$F_{HLE} > 400$ (Ur) or 500 (Pt)
or $F_{HLE} < 400$ (Ur) or 500 (Pt)

Figure IV.F1 The F_{HLE} calculations for the Uranyl and Platinum derivatives. The F 's are on an arbitrary scale on which average F_P is about 1500.

If the overestimation of the heavy-atom contribution is significant, the occupancies and scale factor from F_{HLE} refinement will be too large. A comparison of these results with those from phased refinement will be made below.

1. Uranyl Derivative - F_{HLE} Refinement.

The first step in the heavy-atom refinement was the F_{HLE} refinement of the two initial uranyl sites. The F_H estimates are more likely to be accurate for centric terms. However, because the only centrosymmetric projection in space group C2 is h0l, the refinement of y-coordinates with only centric reflections is impossible. A preliminary examination of the other variables was made in this way.

The progress of the uranyl F_{HLE} refinement is given in figure IV.F2. At steps B and C, each site was refined independently and the other site located in a Fourier synthesis with coefficients $(F_{H(obs)} - F_{H(calc)})\alpha_{calc}$ derived from the single-site refinement. No minor sites were revealed at this point from a similar synthesis following refinement of both sites. Nonetheless, the R-factors which monitor the course of refinement reduced to below the expected random expected values of 0.83 (centric) and 0.59 (acentric) mentioned in section III.E.

A preliminary set of protein phases based on the uranyl derivative alone was calculated at this point, the ambiguity being relieved as well as possible with information from anomalous scattering. When a derivative has only one site of binding, it gives rise to a centrosymmetric array of sites in space group C2 and, if the native and derivative data are indexed on consistent right-handed sets of axes, the anomalous contribution can be added to the structure factor calculation to advance the phase without ambiguity. But, when a second site is included (with a different y-coordinate), the choice of the hand of the array of heavy-atom sites is

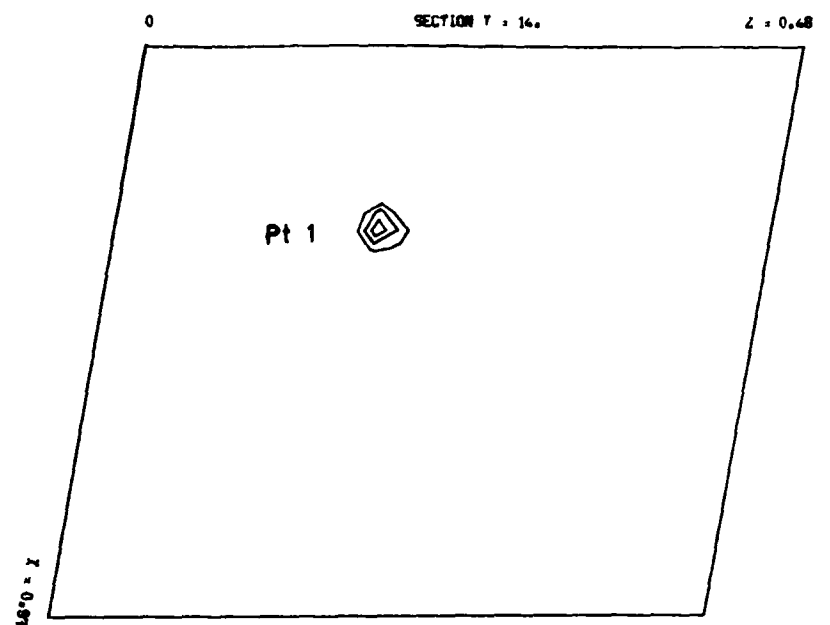
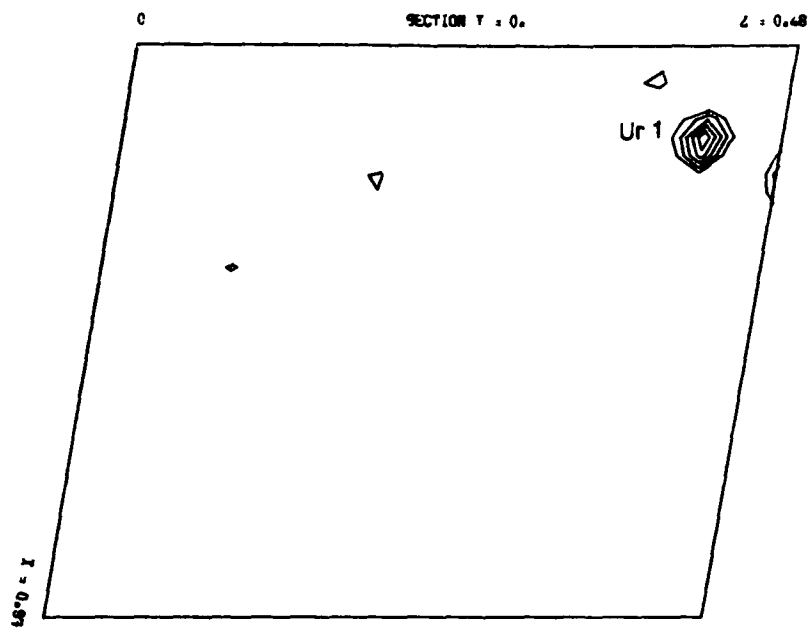
Initial Parameters												
(1)Scale Factor	1.030	0.978	(3)	B	(3)	C	(3)	D	(3)	E	F	G
Ur1 Occupancy	1.0	0.518				1.003	1.001	0.986	0.985	1.021	1.095	0.797
(2)x	0.1563	0.1583				1.148	1.148	1.183	1.162	1.162	0.1655	0.1626
Y	0.0000					0.1538	0.1538	0.1631	0.1673	0.1673	0.1655	0.1626
Z	0.4200	0.4204				0.4212	0.4212	0.4219	0.4204	0.4204	0.4208	0.4230
Ur2 Occupancy	1.0	0.769				1.108		0.398	0.483	0.483	0.398	0.327
X	0.1641	0.1653				0.1799		0.1640	0.1595	0.1595	0.1572	0.1990
Y	0.4219	0.4151									0.3741	0.3567
Z	0.3300	0.3290				0.3395		0.3267	0.3725	0.3725	0.3292	0.3397
(4)n(obs)		909				132	132	134	134	134	148	802
(5)gradient		0.10				0.01	0.09	0.19	0.19	0.19	0.15	0.05
(6)R	0.771	0.503									0.617	0.550
(7)Rc						0.650	0.613	0.584	0.587			
(1) Derivative-to native scale factor												
(2) Fractional unit cell coordinates												
(3) F _{HLE} refinement with centric reflections only												
(4) Number of observations used in refinement												
(5) Gradient of the plot F _{H(calc)} against F _{HLE} (ideally 1.0)												
(6) $R = \sum F_{HLE} - F_{H(calc)} / \sum F_{HLE}$												
(7) R evaluated for centric terms only												

Figure IV.F2. F_{HLE} refinement of Uranyl derivative sites 1 and 2. Each step represents 5 cycles of refinement.

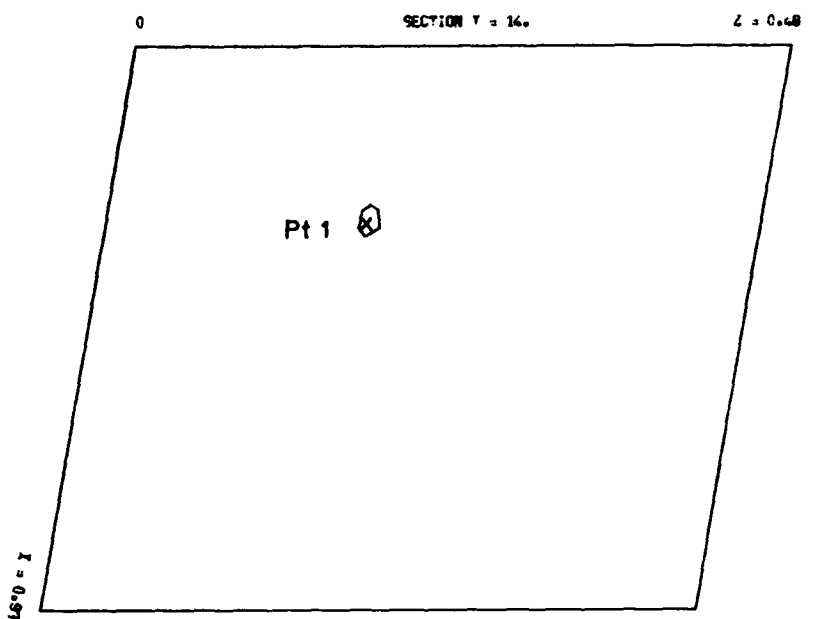
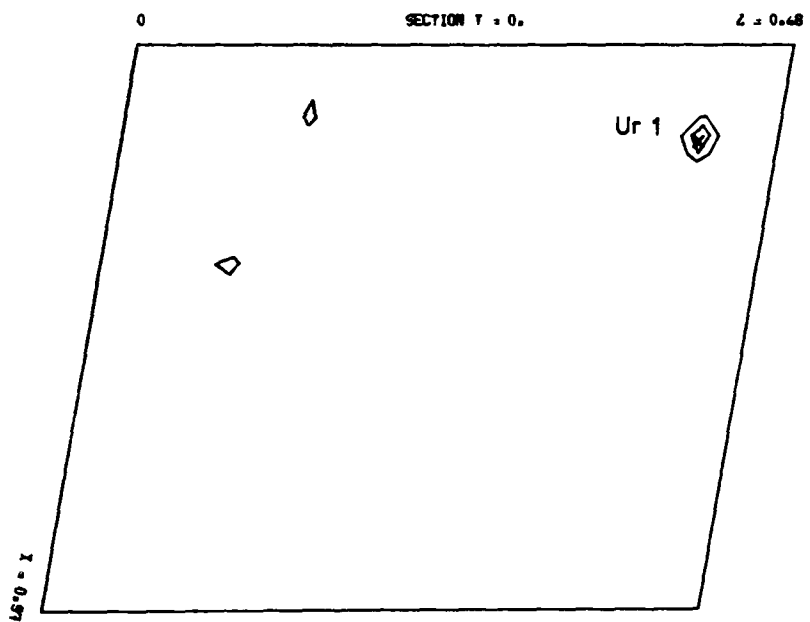
Uranyl

Platinum

Phase retarded by anomalous:



No anomalous:



Phase advanced by anomalous:

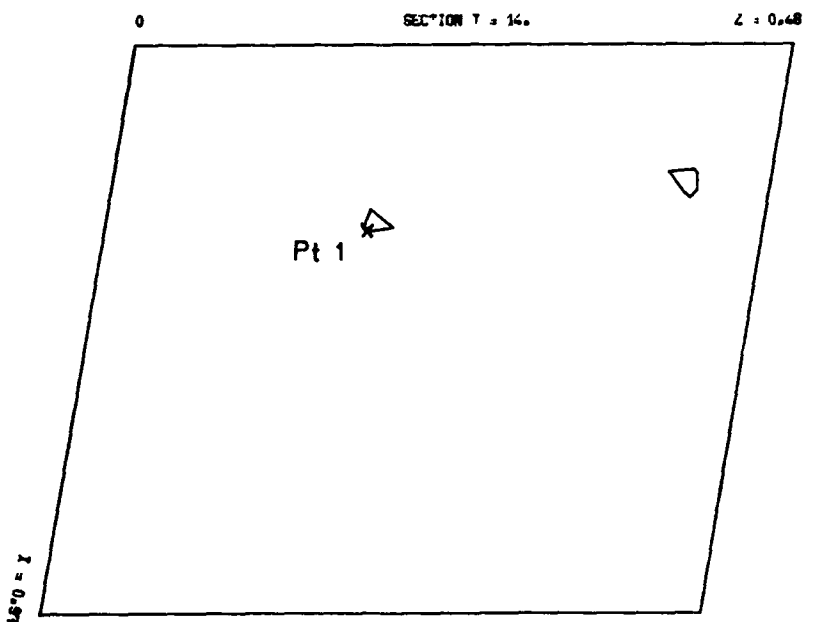
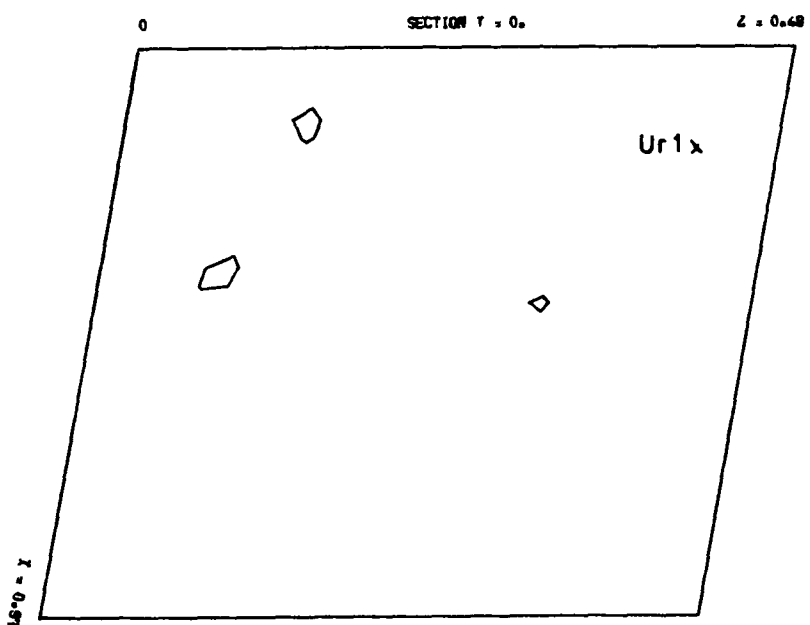


Figure IV.F3. Sections from the $F_{PH} - F_P$ difference Fourier syntheses with the anomalous contribution included to advance and retard the phase. Contoured at intervals of 10 (the estimated background level) starting at 20.

arbitrary. A simple check whether or not this choice was correct is to calculate a set of phases with the anomalous contribution advancing, and another with it retarding the phase. Figure IV.F3 shows that the features at the platinum site in a difference Fourier (see next paragraph) were enhanced when the anomalous contribution was applied to retard the phase. Therefore, the hand of the heavy-atom axis system had to be reversed (for example, by negating the y-coordinates), before the calculation of phases to be used in a protein Fourier.

The difference Fourier, of coefficients $(F_{PH(Pt)} - F_P)\alpha_{Ur}$, confirmed the positions of the proposed platinum sites and allowed them to be defined with respect to the same y-axis origin as the uranyl sites.

2. Platinum Derivative - F_{HLE} Refinement.

The course of the platinum derivative refinement is presented in figure IV.F4. Again, centric terms were used in the early steps to define the x- and z-coordinates and to check difference maps of the type described above. The R-factor decreased, and the agreement between $F_{H(obs)}$ and $F_{H(calc)}$ (the gradient in figure IV.F4) improved through these steps.

From step E, several rounds of F_{HLE} refinement were performed which included all the terms. Refinement of the y-coordinates of the two sites alternated, one being held constant during each round, in order to minimize the tendency to drift from the coordinate system defined by the uranyl derivative.

After this refinement, a second set of protein phases was calculated, this time based only on the platinum derivative. Besides showing that the hand of the solution again should be changed (figure IV.F3) and confirming the original two uranyl sites, the $(F_{PH(Ur)} - F_P)\alpha_{Pt}$ difference Fourier revealed a third minor site.

A refinement of the uranyl parameters that made use of these

Initial Parameters										
	A	B	C	D	E	F	G	H	I	J
Scale factor(1)	1.01	1.010		1.010	1.009					
Ptl occupancy	1.0	0.803	1.125	0.276	0.297	0.599	0.547	0.669	0.737	1.079
x(2)	0.3750	0.3834	0.3862	0.3917	0.3964	0.3935	0.3919	0.3927	0.3900	0.3855
y	0.2813					0.2738		0.2781		0.2774
z	0.5600	0.5611	0.5630	0.5775	0.5802	0.5713	0.5718	0.5712	0.5702	0.5659
Pt2 occupancy	1.0	0.748								
x	0.1875	0.1893	1.195	1.180	1.179	1.055	1.081	1.016	0.972	0.632
y	0.0938	0.0846	0.2077	0.2047	0.2039	0.2016	0.2014	0.2010	0.2002	0.1976
z	0.8000	0.8020	0.8164	0.8184	0.8176	0.8153	0.8150	0.8151	0.8148	0.8070
n(obs)(4)	790	143	143	143	143	1054	1054	1054	1054	1054
gradient(5)	0.09	0.05	0.21	0.24	0.26	0.13	0.12	0.13	0.14	0.17
R (6)	0.694	0.546				0.518	0.520	0.517	0.514	0.488
Rc (7)		0.658	0.538	0.522	0.520					

See Figure IV.F2 for notes.

Figure IV.F4. **F_{HLE}** Refinement of Platinum Derivative Sites 1 and 2.

Scale factor(1)	1.0	A	B	C	D	E	F	G
Ur1 occupancy	1.09	0.746	0.519	0.517	0.518	0.519	0.679	1.042
x(2)	0.1655	0.1507	0.1520	0.1523	0.1521	0.1520		0.880
y	0.0000	-0.0022	-0.0137	-0.0116	-0.0120	-0.0137		
z	0.4208	0.4232	0.4235	0.4237	0.4238	0.4235		
Ur2 occupancy	0.40	0.402	0.330	0.333	0.333	0.330	0.315	0.255
x	0.1572	0.1542	0.1648	0.1727	0.1716	0.1677		
y	0.3741	0.3971	0.3863	0.3852	0.3854	0.3867		
z	0.3292	0.3329	0.3353	0.3347	0.3346	0.3352		
Ur3 occupancy	0.75		0.298			0.298	0.327	0.397
x	0.4063		0.4015			0.4017		
y	0.4063		0.4203			0.4198		
z	0.1800		0.1737			0.1736		
n(obs)(3)		135	460	460	460	460	235	97
R _K (4)	0.165	0.147	0.133	0.136	0.136	0.133	0.138	0.153
wR _K (5)	0.173	0.158	0.152	0.155	0.155	0.152	0.157	0.169
F.M. cutoff(6)	0.0	0.0	0.65	0.65	0.65	0.65	0.80	0.90

- (1),(2),(3) As in Figure IV.F2.
- (4) $R_K = \sum |F_{PH(obs)} - F_{PH(calc)}| / \sum F_{PH(obs)}$ (Kraut R-factor).
- (5) Weighted by the inverse of the lack-of-closure error, E.
- (6) Only reflections with figures of merit above the cutoff were used in the refinement.

Figure IV.F5. Platinum-phased refinement of uranyl derivative parameters.

platinum-derived phases (section III.E) was initiated.

3. Uranyl Derivative - Phased Refinement.

Figure IV.F5 describes the course of this refinement. The quantity being minimized in this scheme is the difference between the observed F_{PH} and that calculated ($|F_P + F_{H(calc)}|$) from the current heavy-atom parameters. So, the appropriate R-factor to use is the Kraut R (R_K , see note 4, figure IV.F5).

Several steps were carried out both including and excluding the new minor site. It is important to have a guide to whether the decrease in R-factor on inclusion of the new site is significant in terms of the number of parameters being varied. A convenient R-factor ratio introduced by Hamilton (1965) was used for this purpose throughout this series of refinements.

For an illustration, let us consider the addition of the third uranyl site between steps D and E. The number of degrees of freedom (observations - parameters varied) is 448 and the increase in the number of parameters is 4. With reference to the tables computed by Hamilton (1965), we find that an R-factor ratio of 1.011 would be significant at the 5% level, and one of 1.017 at the 0.5% level. In fact the ratio (0.136/0.133) is 1.023 and indicates that we would be incorrect in fewer than 0.5% of analogous instances if we accepted the improvement as significant. The minor site was therefore retained.

4. A Comparison of the Refinement Schemes.

Figure IV.F6 shows the final heavy-atom parameters derived from the above refinements and those from a uranyl-phased platinum refinement. The positions resulting from different refinement schemes are the same within the resolution of the data. The only large discrepancy is in the occupancy of the first platinum site. Because the agreement within the uranyl

Refinement Scheme	F_{HLE}	Phased	Difference($\text{\AA}$)	F_{HLE}	Phased	Difference($\text{\AA}$)
	Ur	Ur		Pt	Pt	
Scale factor	1.083(1)	1.042		1.030(1)	1.033	
Site 1 Occ.	0.327	0.255		1.079	0.529	
(2) x	0.1990	0.1677	1.9	0.3855	0.3903	0.29
y	0.3567	0.3867	1.7	0.2774	0.2620	0.87
z	0.3397	0.3352	0.62	0.5659	0.5619	0.56
Site 2 Occ.	0.797	0.880		0.632	0.554	
x	0.1626	0.1520	0.63	0.1976	0.1927	0.29
y	0.0204	-0.0137	1.9	0.0535	0.0632	0.55
z	0.4230	0.4235	0.07	0.8070	0.8051	0.26
Site 3 Occ.		0.397				
x		0.4017				
y		0.4198				
z		0.1736				

(1) Derived from F_{HLE} calculation (see figure IV.F1).

(2) Fractional cell coordinates.

Figure IV.F6 A comparison of the results from the two refinement schemes.

derivative is good, these values can be accepted with some confidence.

As discussed earlier, the F_{HLE} -derived platinum values are possibly overestimates. At the same time, because of the inherent inaccuracy of one-derivative phases, the phased refinement result could conceivably be underestimating the platinum presence.

One way to resolve this dilemma is to perform a refinement which, on the surface, is biased, a two-derivative phased refinement. This is theoretically unsound because it consists of refining parameters which were used to calculate the phases in the first place. However, by using a high figure-of-merit cut-off, thereby choosing to include only reflections with well-determined, presumably unbiased phases, we can cautiously use this method.

5. Final Phased Refinements of Occupancies.

The following procedure was performed on a PDP 11/70 computer with a different series of programs (Refine(11) and Phase(11), appendix I).

Before the refinement was started, a third platinum site derived from the difference Patterson, was added. Although a comparatively minor site, its self- and cross-vectors accounted for virtually all the remaining peaks in the Patterson.

Figure IV.F7 shows that the uranyl results agree well with those from the initial refinements. As expected, the platinum occupancies showed a compromise between the earlier, contradictory results. The closeness of the occupancies of the first two platinum sites is in agreement with the sizes of the self-vectors in the difference Patterson synthesis.

The results of this final phased refinement were accepted with some confidence because the uranyl parameters, in particular, showed no evidence of bias; convergence was very fast and the occupancies did not increase significantly beyond their values expected from previous refinements.

	<u>Pt Derivative</u>		<u>Ur Derivative</u>	
Scale factor	1.029	1.035	1.042	1.025
Occupancy:				
Site 1	0.530	0.878	0.880	0.907
Site 2	0.550	0.868	0.260	0.385
Site 3	0.370	0.482	0.400	0.398
$R_K(1)$	0.095	0.078	0.103	0.103
$\langle FH \rangle(2)$	199.5	312.2	255.7	272.6
$\langle E \rangle(3)$	187.0	171.4	216.1	215.9
$R_X(4)$	0.531	0.436	0.663	0.653
$R_{X(cen)}(5)$	0.533	0.507	0.639	0.622
cycles for convergence		3		3
observations		441		427

(1) Defined in Figure IV.F5. note 4.

(2) $F_{H(calc)}$ averaged over centric reflections.

(3) Lack-of-closure error averaged over centric reflections.

(4) $R_X = \sum [(F_{PH(obs)} - F_{P(obs)}) - F_{H(calc)}] / \sum (F_{PH(obs)} - F_{P(obs)})$,
summed over all reflections.

(5) R_X summed over centric reflections only.

Figure IV.F7. The final (two derivative-) phased refinement of the occupancies and scale factors. The figure of merit cutoff was 0.80.

6. Analysis of the Final Two Derivative Phases.

Several checks were made that the final protein phases were not biased towards either of the derivatives. The first was a search for "ghosts" in difference and double-difference Fourier syntheses (figure IV.F8). No features corresponding to the uranyl sites were detected in either of the platinum syntheses, and only one platinum site showed slight ghosting in the uranyl difference Fourier. In both instances, the heavy-atom peaks in the difference syntheses were brought down very nearly to background level in the double-difference maps. The only slight residual, positive features at the first two uranyl positions, indicated that these occupancies may be slightly underestimated. However, in light of the fact that they had already been refined against biased phases, the occupancies were not increased further.

A second check was the calculation of histograms analyzing the number of reflections against the difference between α_p and α_H (figure IV.F9). A bias, if present, would manifest itself in a build up of the number of reflections at each end of the scale, with phase differences of 0 or 180°.

Thirdly, a native protein Fourier synthesis was printed as a number field and the areas around heavy-atom positions examined. No evidence of ghosting was found.

The role of the isomorphous and anomalous lack-of-closure errors in the calculation of the phase probability was discussed in section III.F (see especially equation III.F1). Ideally, these errors should be small compared to the heavy-atom structure factor amplitude. Furthermore, the errors from centric reflections should reflect more accurately the true errors because the full length of the lack-of-closure, rather than a projection onto the real axis (refer to figures III.14 and III.15), is measured.

Synthesis	<u>Pt Derivative</u>		<u>Ur Derivative</u>	
	$(F_{PH}-F_P)$	$(F_{PH}-(F_P+F_H))$	$(F_{PH}-F_P)$	$(F_{PH}-(F_P+F_H))$
Site:				
Pt1	[101] ⁽¹⁾	[2]	[-2]	[2]
Pt2	[91]	[7]	[19]	[1]
Pt3	[49]	[-2]	[4]	[1]
Ur1	[5]	[4]	[103]	[18]
Ur2	[7]	[3]	[52]	[16]
Ur3	[1]	[5]	[43]	[-6]
Background ⁽²⁾	[14]	[9]	[17]	[12]

(1) Peak heights are on an arbitrary, but equal, scale.

(2) Most sections have positive and negative features of this size.

Figure IV.F8. Peak heights on difference $[(F_{PH}-F_P)\alpha_P]$ and double-difference $[(F_{PH}-(F_P+F_H))\alpha_P]$ Fourier syntheses calculated with the final two-derivative protein phases.

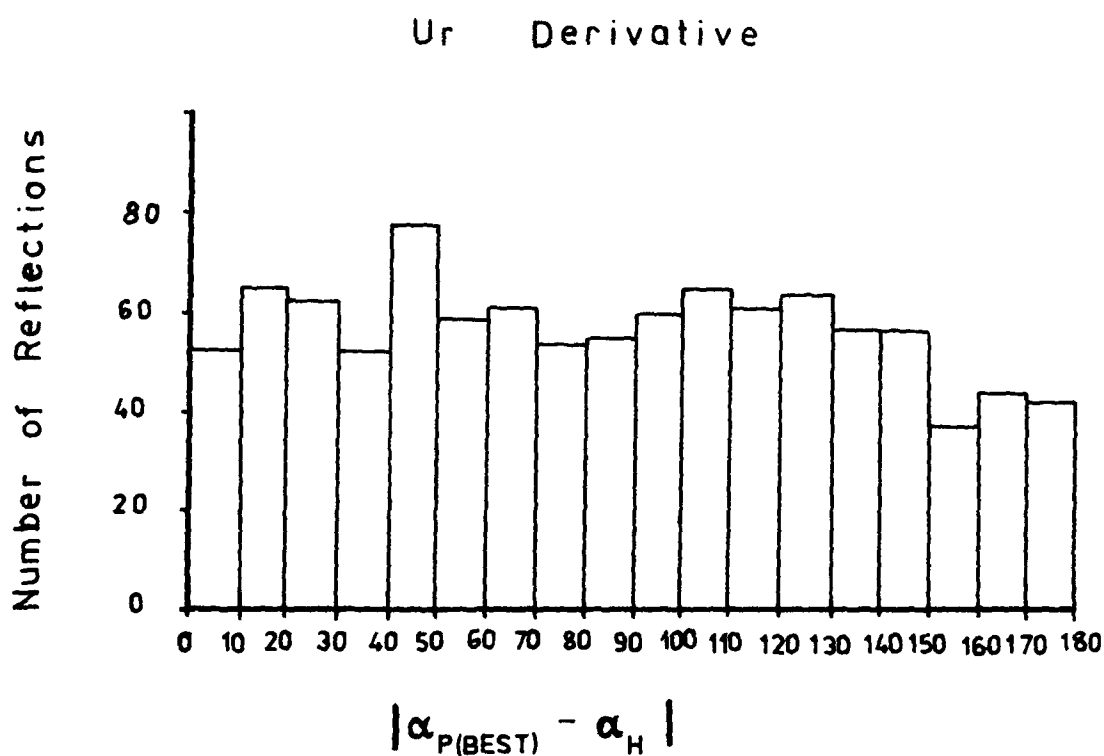
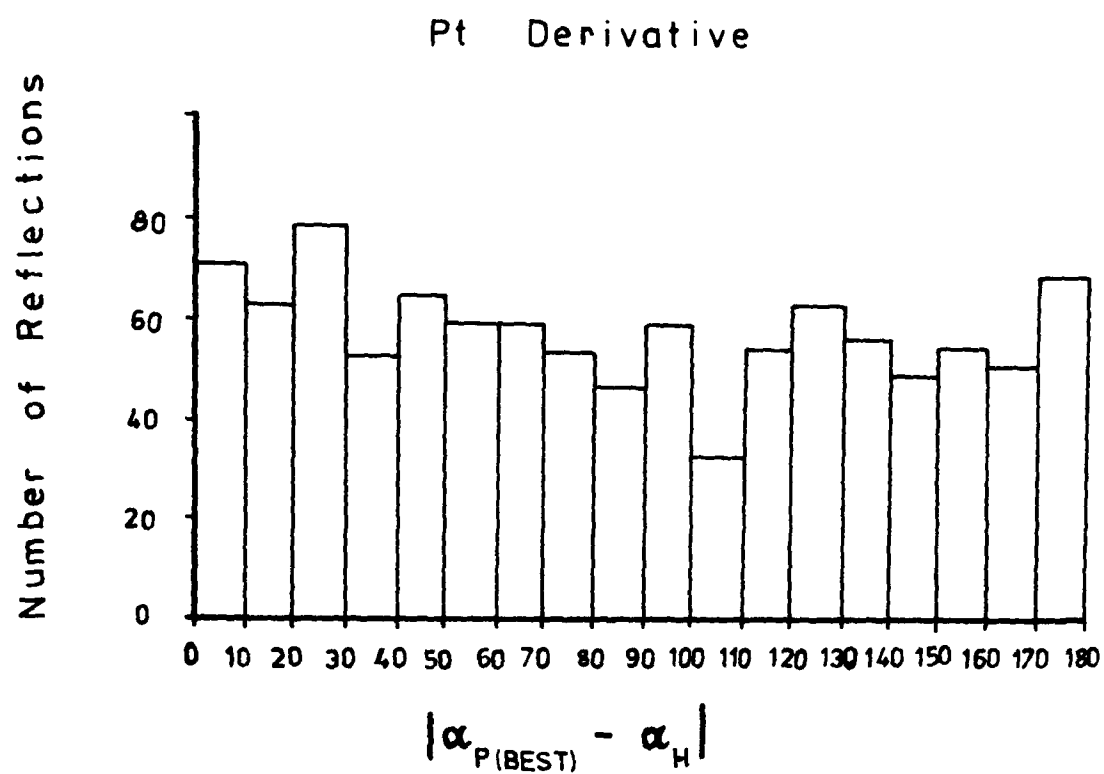


Figure IV.F9. The correlation between $\alpha_{P(BEST)}$ and α_H for the acentric terms.

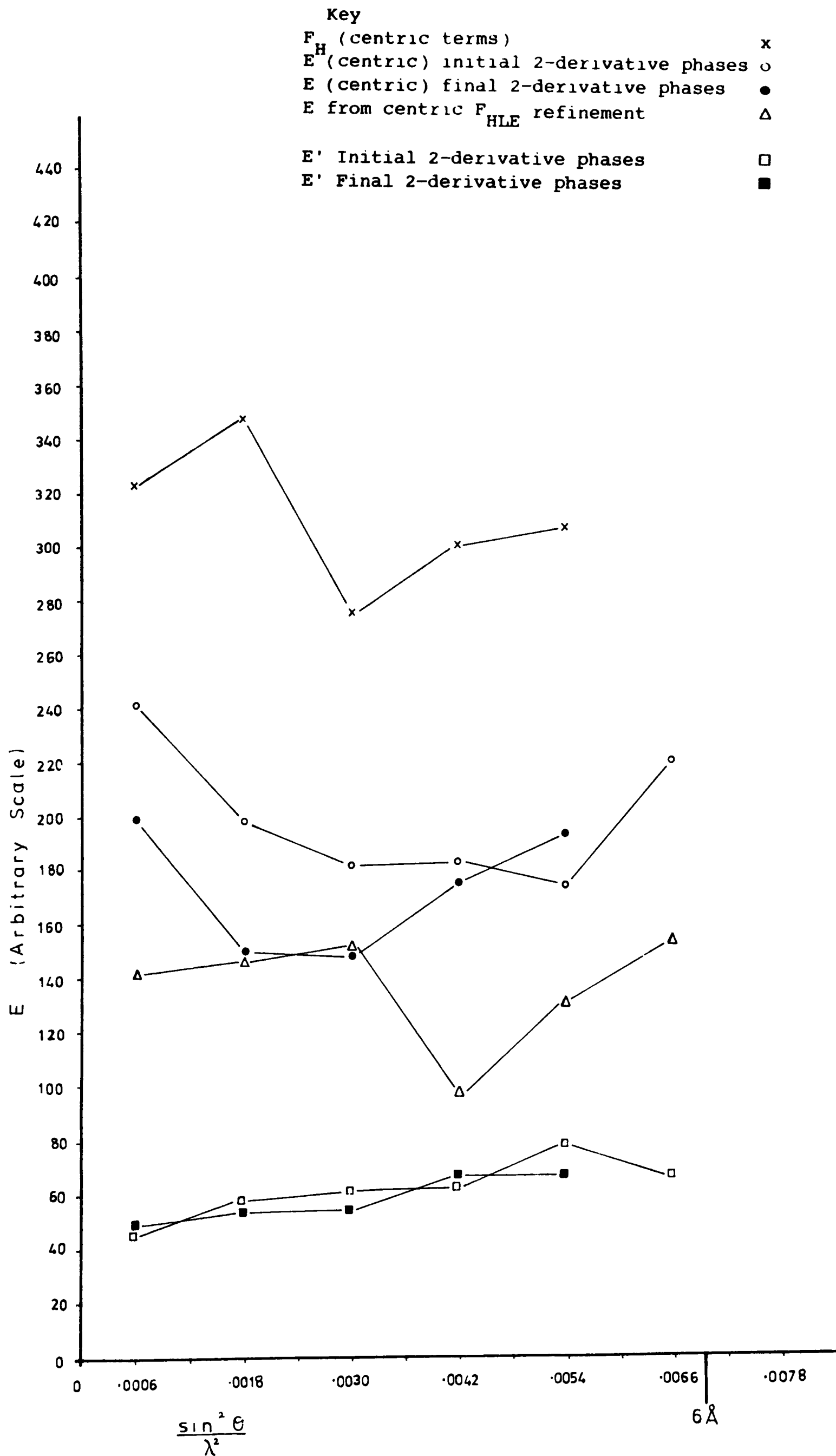


Figure IV.F10. Pt Derivative. Comparison of the isomorphous (E) and anomalous (E') lack-of-closure errors with F_H derived from the final parameters.

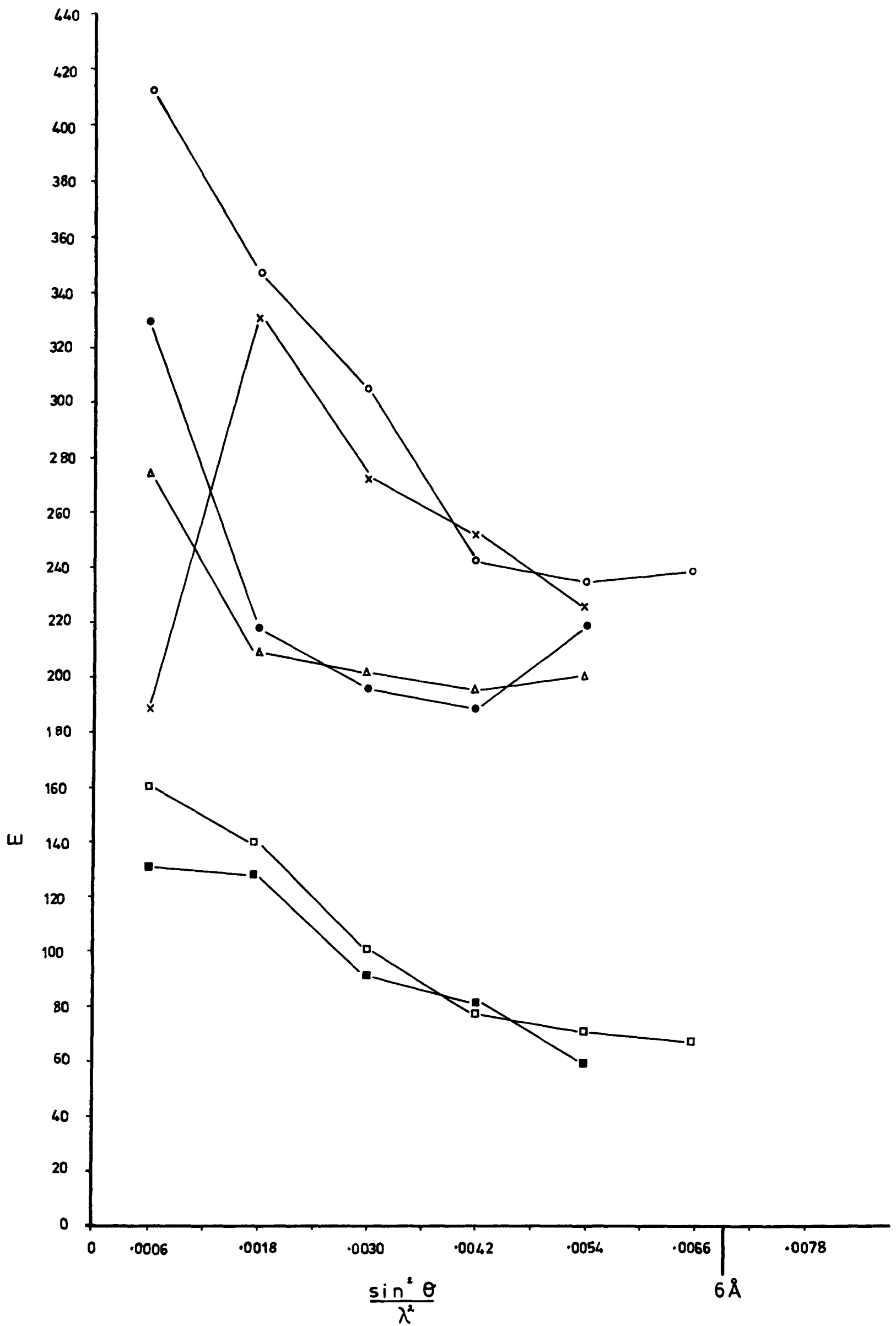


Figure IV.F11. Uranyl derivative. Comparison of E and E' with F_H derived from the final parameters. See figure IV.F10 for key.

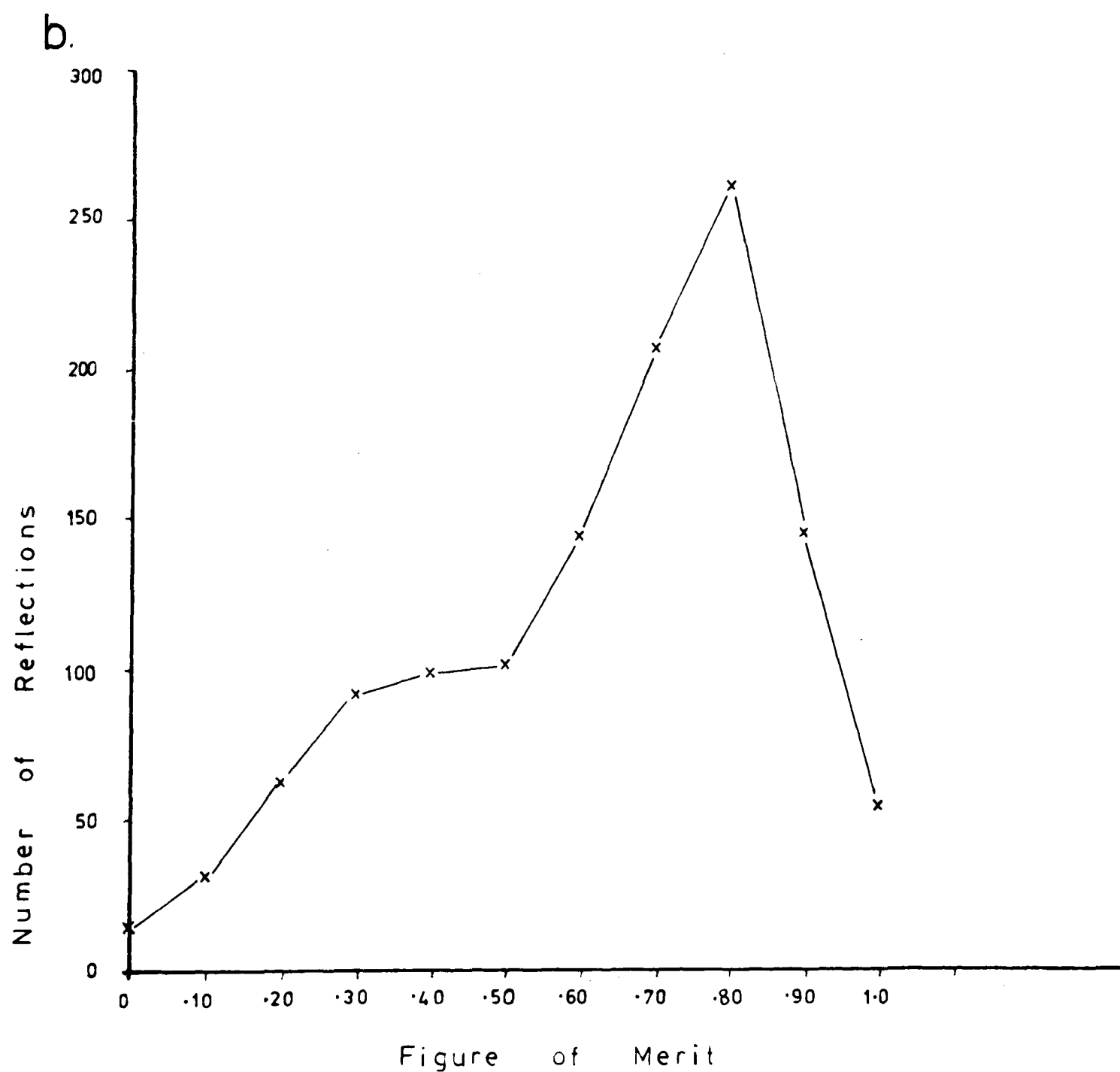
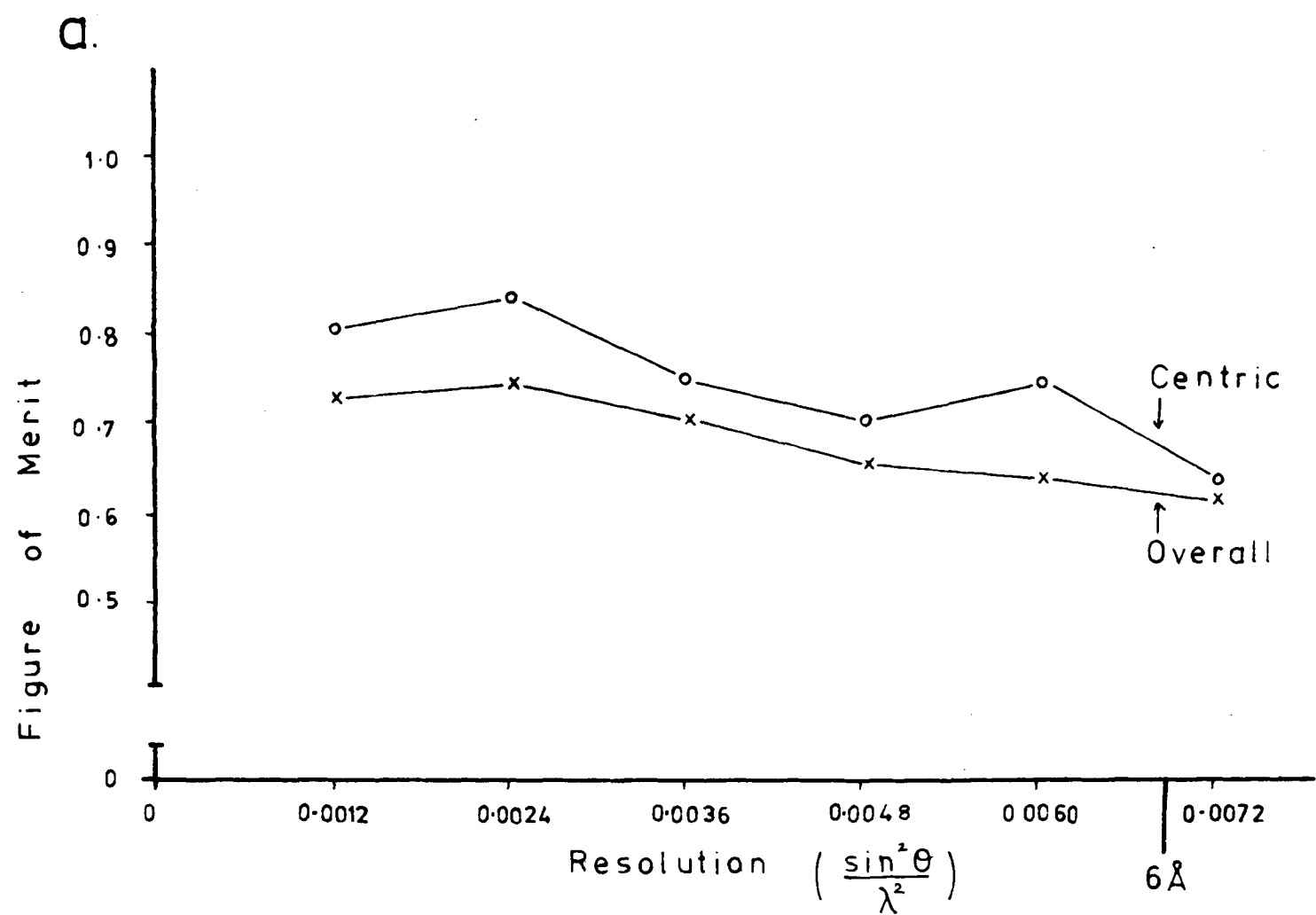


Figure IV.F12. Analysis of the figures of merit of the final two-derivative phases. (a) Figure of Merit vs. Resolution. (b) Number of Reflections vs. Figure of Merit. The mean overall figure of merit was 0.67.

Figures IV.F10 and IV.F11 compare the lack-of-closure errors of centric reflections at various stages of the refinement, with F_H . Clearly, the figures-of-merit of the phases are very dependant on the estimates of the lack-of-closure errors (equation III.F2) and, thus, the choice estimates to be used in the phasing is important. The values used in the final phase calculation (and shown in figures IV.F10 and IV.F11) came from the final phased refinements, and resulted in an overall figure-of-merit of 0.67 (see figure IV.F12).

CHAPTER V

The Extension of the Resolution

The rotation camera method was briefly reviewed in section III.D2 with reference to the more comprehensive account by Arndt and Wonacott (1977).

V.A. Data Collection

Photographic data were collected on an Enraf-Nonius rotation camera (Arndt et al, 1973) with radiation from a GEC-Elliott GX 6 rotating anode x-ray tube operated at 40 kV, 40 mA. The effective focal spot of the beam was 0.2 mm × 0.2 mm. CuK α x-rays were selected by reflection from a graphite crystal. This results in an x-ray beam which is weaker but less contaminated with unwanted wavelengths than one simply filtered through nickel. Thus, although longer exposure was required for each photograph, the resulting signal-to-noise ratio was better. Because x-rays with wavelengths higher than the characteristic CuK α radiation are strongly absorbed by, and particularly damaging to protein crystals, cleaner radiation often results in improved crystal lifetime.

To facilitate the data processing, the crystal was rotated about a reciprocal axis. All the Fv data were collected by rotation about the c* axis. As mentioned in chapter IV, the natural mounting axis of the crystal is c. Although possible, mounting about a or b is difficult to achieve and usually results in an unstable situation in which the crystal may slip in the capillary tube. In addition, because reflections with indices h+k odd are systematically absent in space group C2, the density of reflections in the already longer a* and b* directions is reduced even further. As a result, the crystal can be rotated about c* through the maximum range allowable by the instrument, 4°, before a significant number of reflections overlap on the film. On the other hand, because of the long c cell

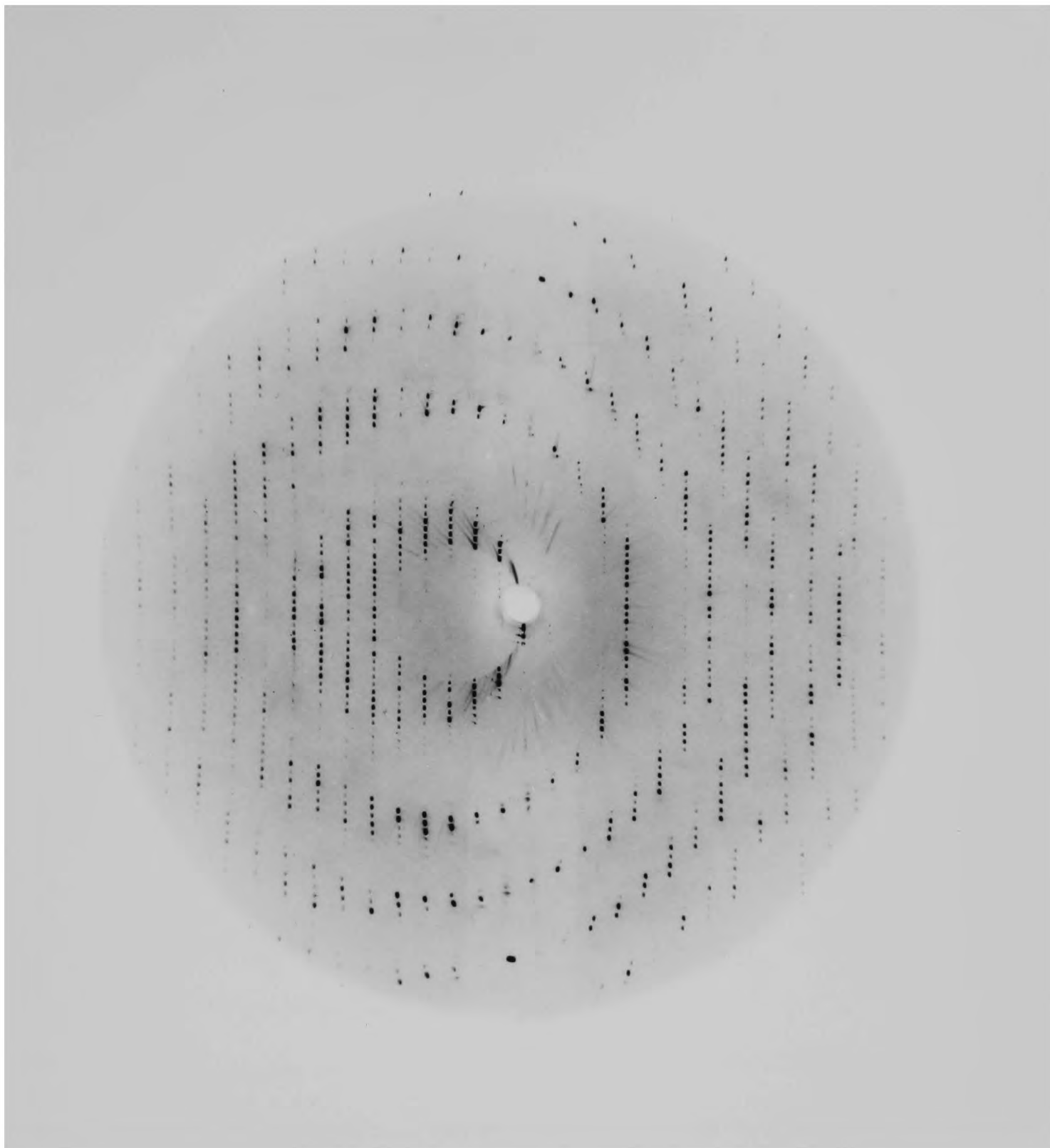


Figure V.1. X-ray Photograph of a 4° Rotation of an Fv Crystal about the c^* Axis. The spots at the edge of the photograph correspond to $2.8\text{\AA}$ resolution.

dimension, the maximum allowable rotation range about a^* or b^* is only $1/2$ to 1° . In fact, when a b^* mount was achieved, even a rotation of $1/2^\circ$ led to an uninterpretable photograph.

The only serious disadvantage to a c^* mount arises in the measurement of anomalous scattering effects from derivative crystals. In this instance, a b^* mount would be preferable because it would allow the measurement of the equivalents ($h \pm k \ 1$) on the same film and, therefore, minimize any radiation damage effects. However these equivalents would suffer different absorption effects, for which there is no simple correction on the rotation camera. There is a case even here, therefore, for mounting about another axis (say c^*) and collecting, in subsequent photographs, the hkl and $\bar{h}\bar{k}\bar{l}$ reflections.

As all the data were measured in rotations about the c^* axis, reflections in the blind region (see section III.D2) were not recorded. However, at $2.7\text{\AA}$ resolution, only 2.5% of the complete data set is lost for this reason (Arndt and Wonacott, 1977, p.81).

1. Data Collection Strategy.

First, a unique set of native data was collected by a total rotation about c^* of 90° . Figure V.1 shows one of the data photographs, each of which recorded a 4° rotation.

Figure V.2 shows the region of reciprocal space which is measured during a 90° rotation. In fact, nearly 1.5 times the number of reflections necessary for a unique set of data, was measured (see figure V.3). Although only one crystal was used, the data set was divided into three parts. As the crystal was larger than the incident beam, when one section of the crystal began to show signs of radiation damage, it was translated along the rotation axis and data collection continued from another section. Each segment of the data was treated separately in the processing procedure.

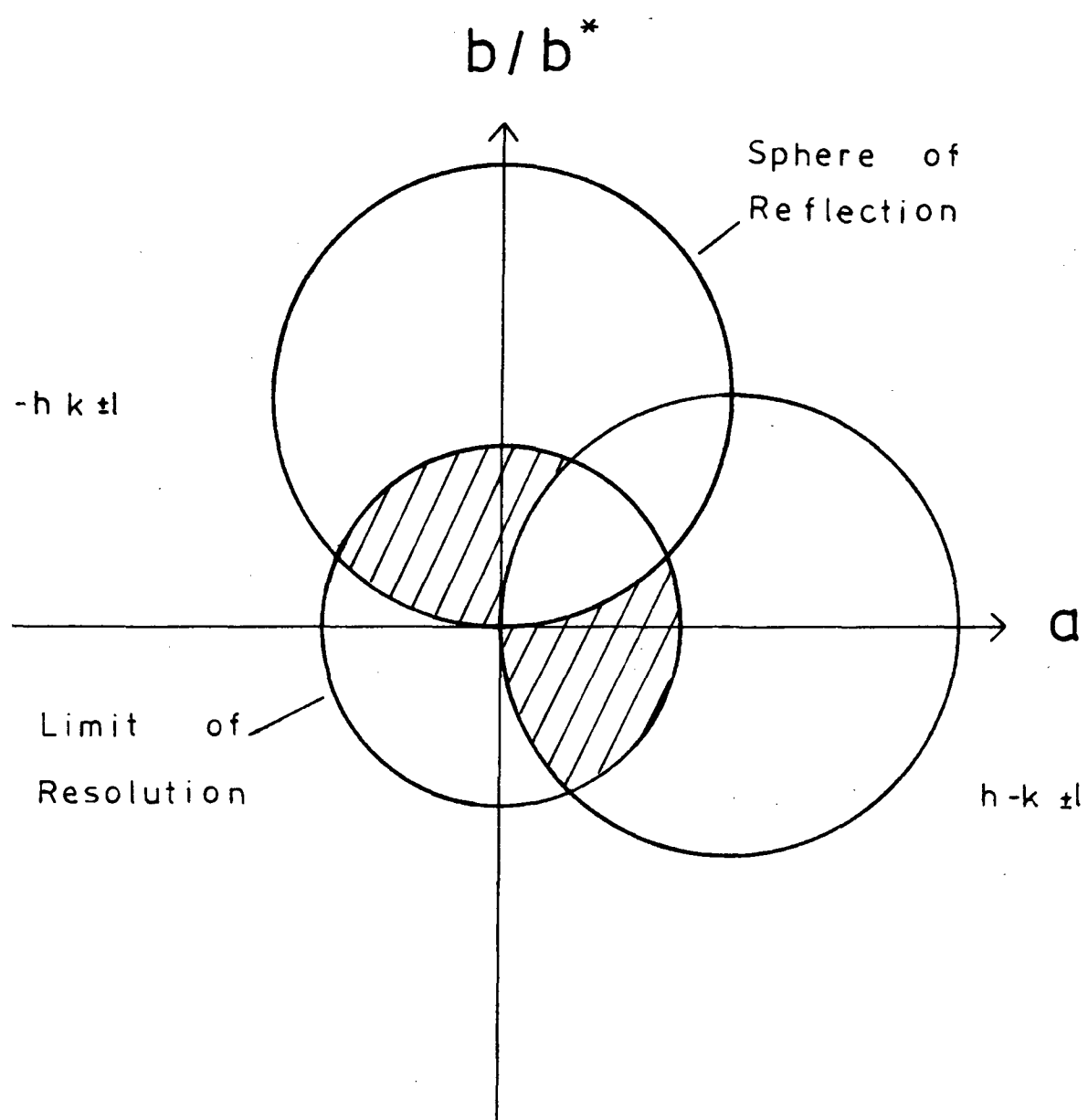


Figure V.2. The volume of reciprocal space (shaded) sampled by a 90° rotation about the c^* axis (out of the plane of the diagram).

Subset	Fv7	Fv10	Fv14
Number of film packs	8	6	8
Rotation range(°)	-1.0-31.0	34.0-58.0	57.0-93.0
Fox and Holmes scale(1)	1.17	0.31	0.19
Number of reflections(2)	9056	6243	7697
Number rejected(3)	45	115	162
Merging R-factor(3)	0.123	0.133	0.102
Total reflections(4) 16102			
Independent 9797			
Overall merging R 0.119			
Number of negative terms 460			
Maximum resolution(Å) 2.79			

- (1) Fox and Holmes (1966), averaged over all the film packs.
(2) Before merging strong and weak films.
(3) See figure IV.C3.
(4) After merging strong and weak films.

Figure V.3. Merging statistics for the consolidation of the three native photographic data subsets into one unique data set.

Subset	Pt4	Pt5	Ur5
Number of film packs	12	9	23
Rotation range(°)	20.0-68.0	67.0-91.0, 184.0-196.0	0.0-92.0
Fox and Holmes Scale(1)	1.27	0.88	2.01
Number of reflections(2)	11480	8640	26918
Number rejected(3)	468	148	444
Merging R-factor(3)	0.25	0.16	0.27
Total reflections(4)	13758		18874
Independent		8997	10839
Overall merging R		0.142	0.267
Number of negative terms		275	729
Maximum resolution($\text{\AA}$)		2.91	2.74

Figure V.4. Statistics of the merging of data from the platinum and uranyl derivatives into unique sets. Notes as in figure V.3.

The objective of the high resolution data collection was to extend the resolution of the 6Å protein map. It was therefore of primary importance to obtain a complete set of data from each of the two heavy-atom derivatives as efficiently as possible. The anomalous scattering information was therefore disregarded and unique sets of data from platinum and uranyl-soaked crystals (as described in figure IV.B2) were measured. The intention was to attempt to collect anomalous information after completion of the unique set. However, neither crystal survived long enough for a significant number of Friedel equivalents to be measured.

Again, the platinum derivative crystal had to be reset during the course of data collection, although, remarkably, all the uranyl data were taken from one setting of the crystal.

V.B. Data Processing

The transformation of the spots on the films into integrated intensities was performed as described in section III.D2. The contributions from reflections partially recorded on different films were added together. Then, each film was assigned a scale factor based on the multiple recording of equivalent reflections. Figure V.3 shows the merging statistics for the native data, the values being averaged within each of the three subsets. An advantage to scaling each film individually is that most of the radiation damage and some of the absorption effects were corrected. The discussion of a further treatment of the absorption problem follows in section V.B2. The merging statistics for the two derivative data sets are shown in figure V.4.

Unlike the ordinate analysis method used in processing the diffractometer data, the integration of film data can yield negative intensities. Wilson (1949) derived an expression for the distribution of intensities (in terms of the probability of a reflection having a particular intensity) based on the assumption that the scattering occurs

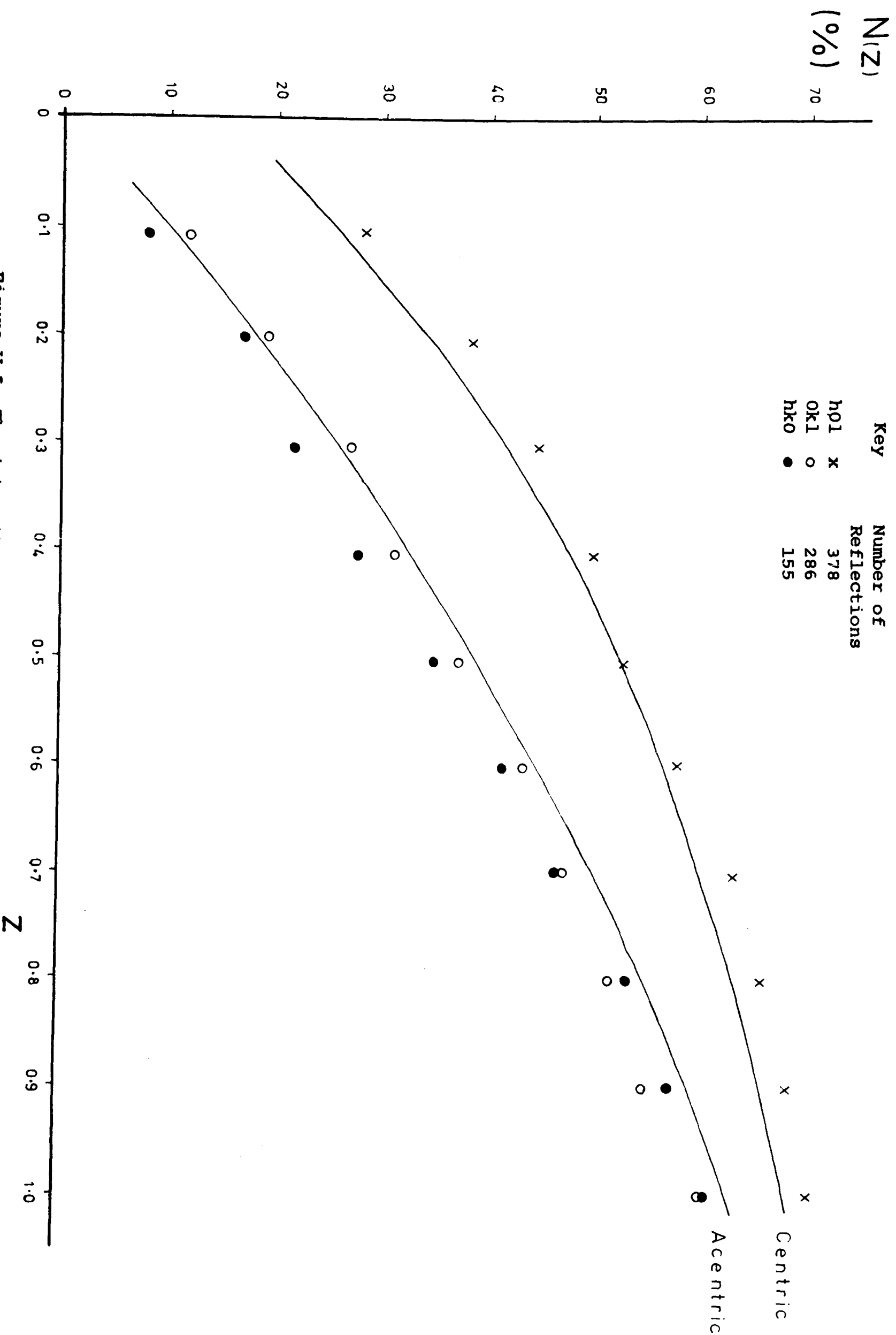


Figure V.5. The intensity distributions of the photographic native data in the three major zones compared to the theoretical probability distributions (Howells et al, 1950). The maximum resolution was 2.8Å.

Key
 Before Scaling o
 After Scaling x

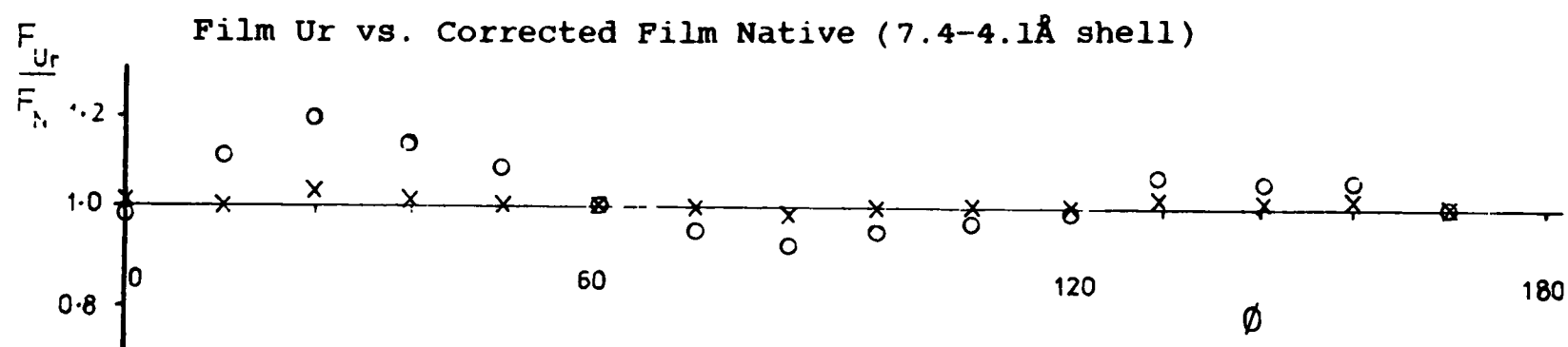
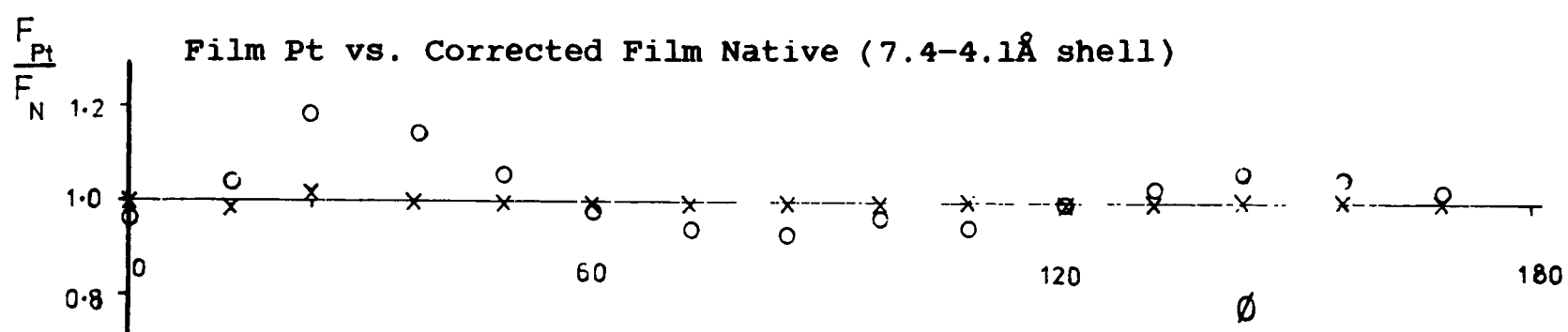
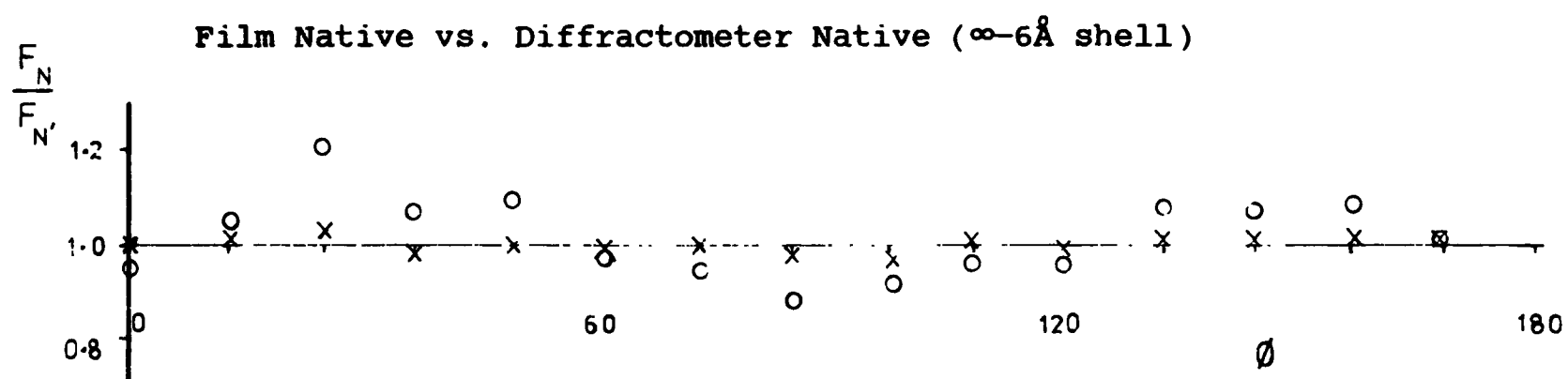


Figure V.6. The absorption correction with a local scale factor as a function of ϕ (Program ANSC, appendix I).

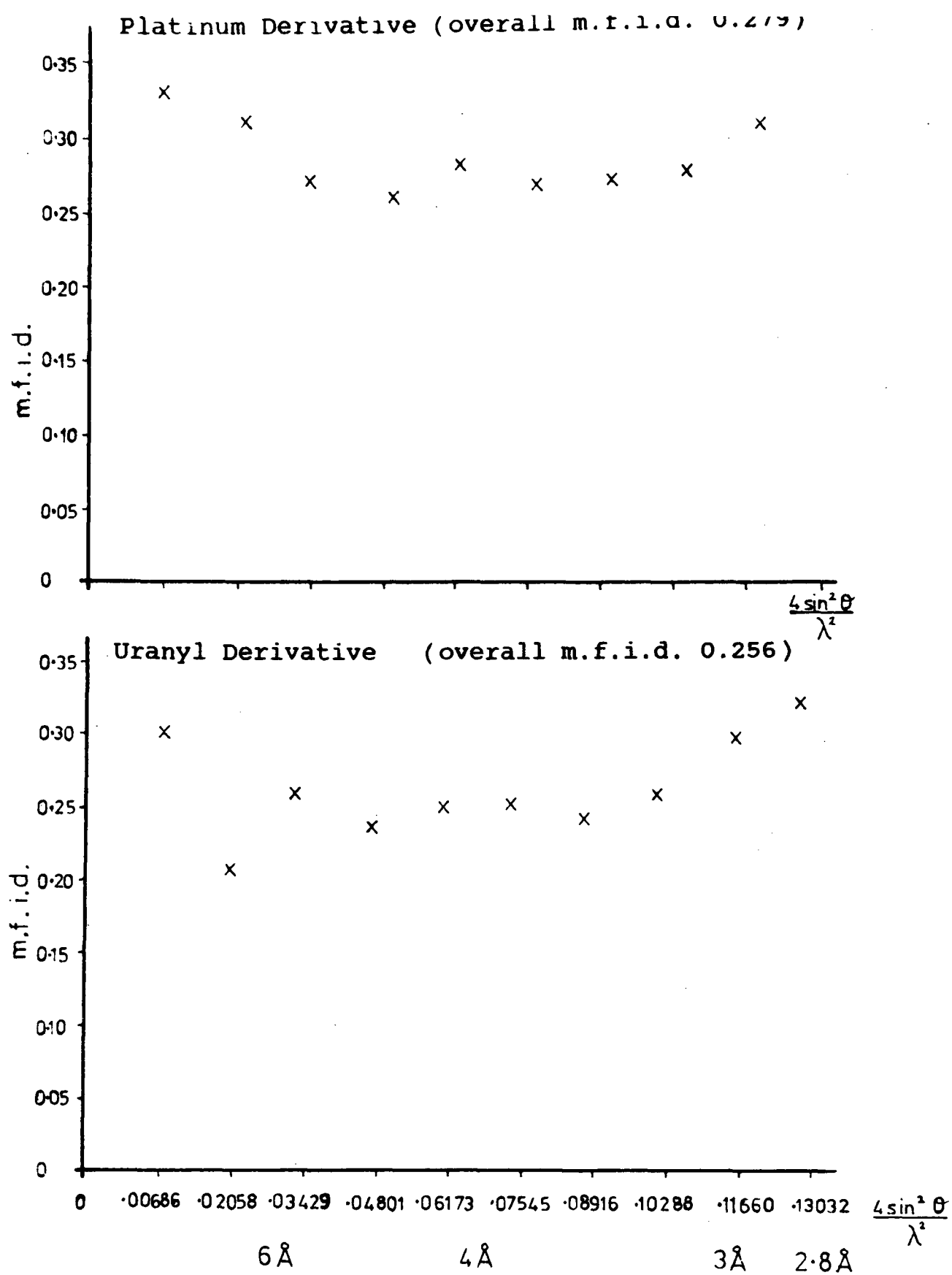


Figure V.7. The variation of the mean fractional isomorphous difference (see figure IV.D3) with resolution. Platinum and Uranyl Derivative Film Data after ρ Scaling.

from a large number of randomly-placed atoms. This holds to some extent for high-resolution protein data and a program TRUNCATE (appendix I) was used to model the photographic data to this all-positive distribution.

1. Native Intensity Distributions.

Howells et al (1950) described a method of examining the distribution of intensities. The fraction of reflections, $N(z)$, which have intensities less than or equal to z times the average intensity, is plotted as a function of z (figure V.5). The film data within the three major zones were treated in this way and follow the theoretical probability distributions for centric (h0l) and acentric structures.

2. Absorption Correction.

In order to take account of absorption effects, the photographic data were compared to those which had been corrected by the empirical method (section III.D1) on the diffractometer, and a scale factor derived as a function of ρ (section III.D3). Although the comparison could only be made within $6\text{\AA}$ resolution, the agreement between film and diffractometer data was significantly improved within this limit after scaling (figure V.6), and the correction was applied to all the native data.

The data from the two derivatives were then scaled to the corrected native data as a function of ρ within resolution shells. The improvement in agreement after scaling (figure V.6) was evident throughout the resolution limits.

As an indication of the quality of the heavy-atom derivatives, the mean fractional isomorphous difference is shown in figure V.7 as a function of resolution (see section IV.D). The isomorphism appears to be good for most of the data although a slight upturn may be present beyond $3\text{\AA}$ resolution.

V.C. Refinement of the Heavy-Atom Sites

Platinum derivative	A	B	C
Scale factor	1.035	1.038	1.056
Overall B	0.0		1.142
Site 1			
occupancy	0.88	1.02	0.83
B-factor	20.0	17.1	8.32
Site 2			
occupancy	0.87	1.06	1.04
B-factor	20.0	19.0	18.06
Site 3			
occupancy	0.48	0.72	0.69
B-factor	20.0	34.2	34.61
Uranyl derivative			
Scale factor	1.025	1.089	1.013
Overall B	0.0		-4.068
Site 1			
occupancy	0.91	0.85	0.90
B-factor	20.0	8.1	8.54
Site 2			
occupancy	0.38	0.32	0.33
B-factor	20.0	17.5	13.64
Site 3			
occupancy	0.40	0.49	0.51
B-factor	20.0	31.1	25.36

Figure V.8. Results of the phased refinement of the heavy-atom parameters using reflections within the resolution range 6.0–2.7Å. The figure of merit cutoff was 0.80 (about 1700 observations).
Column A – Final 6Å refined parameters (Chapter IV)
Column B – After refinement of individual B-factors
Column C – After refinement of overall B-factor,
Parameters used in high resolution phasing.

The following analysis was performed with the REFIN(11) and PHASE(11) computer programs (appendix I).

The goal of this part of the structure determination was to extend the resolution of the 6Å result. Thus, in order not to lose the structural information obtained at low resolution, and because the 6Å data sets included anomalous scattering information, a set of data was constructed which consisted of the diffractometer data to 6Å resolution and photographic data beyond this point. Hence, the refinements described here were carried out on data to the 6Å to 2.7Å resolution shell. Because of the possibility that the mode of binding of the heavy-atoms may have differed slightly in the crystals used for the film data collection, the refined parameters for the 6Å diffractometer data (Chapter IV) were not changed.

As a starting point for the refinement, a set of two-derivative phases was calculated using the parameters resulting from the refinements described in Chapter IV (column A in figure V.8). The figure of merit cutoff for reflections used in the refinements was 0.80.

At 6Å resolution, the scattering from the individual heavy atoms could be modelled fairly well to the theoretical curve by a simple scale factor. Any variation in scattering behaviour with resolution was taken up in the overall scale factor, the individual occupancies, or both. The B-factor (the resolution component of the scale factor k ; $k = k'e^{-B\sin^2\theta/\lambda^2}$) was therefore held at a value of 0.0 throughout the refinement procedure and the individual B-factors for each site (applied to the occupancy in the same way) were held at 20.0.

At higher resolution, however, the scattering of the heavy atoms may not show a simple relationship to the theoretical scattering, independent of resolution. In general, this can be largely taken account of by a B-factor applied to the overall scale factor. Each binding site may show

peculiar characteristics which are modelled by an individual B-factor. For example, a low-occupancy site may also be a weak binding site and therefore have a larger B-factor (a sharper decrease with increasing resolution) than a high-occupancy site.

The first stage in the heavy-atom refinement was the addition of an individual B-factor for each site. It was clear from the start that there was no significant change in the positions of any of the sites on refinement.

In general, the occupancy and the B-factor are interactive. Ideally, appropriate values for each can be determined by refining them together. Although convergence was relatively slow (15 cycles), this was possible with the platinum derivative. The two main sites behaved alike and refined to similar values. The minor platinum site increased in occupancy approximately in the same proportion as the first two sites, but showed a large increase in B-factor (figure V.8).

The uranyl derivative did not permit simultaneous refinement of occupancies and B-factors, and the results in column B (figure V.8) resulted from iterative cycles of refinement of each individually. There was no large change in the occupancy of any of the sites, although the major site showed a significant decrease in B-factor. The minor sites behaved differently in terms of B-factor, the second site yielding a smaller value than that of the third.

The final stage of the high-resolution heavy-atom refinement was the addition of the overall B-factors for each derivative (column C, figure V.8). The refinement yielded a small positive value for the platinum derivative, corresponding to a slight decrease in the overall scale factor with resolution (by a factor of 0.96 at $2.8\text{\AA}$). Further refinement of the individual parameters affected only the first site. Because its occupancy

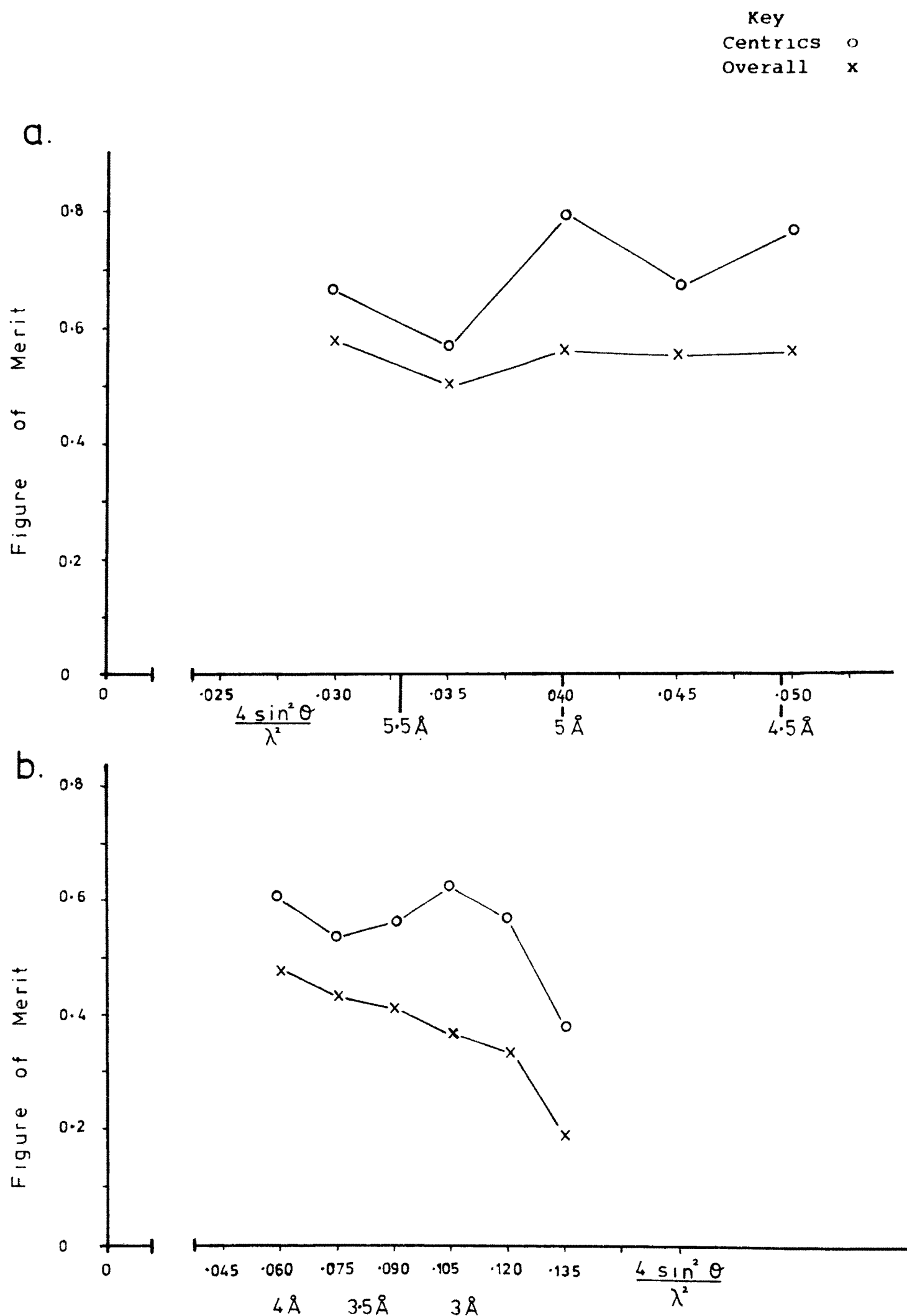


Figure V.9. Variations of the figures of merit of the (a) 4.5Å and (b) 2.7Å resolution phases with resolution. The overall figure of merit was 0.6019 to 4.5Å and 0.4104 to 2.7Å resolution. See figure IV.F12 for the results of the phasing to 6Å resolution.

and B-factor both decreased, however, the contribution from the site to the heavy-atom scattering is unlikely to have changed significantly.

The uranyl derivative gave an overall B-factor of about -4.1 . This corresponds to an increase in the scale factor with resolution by about a factor of 1.14 at $2.8\text{\AA}$, and may be an indication of an increasing lack of isomorphism at high resolution. There was very little effect on the refinement of the individual occupancies and B-factors.

The values in column C (figure V.8) were used in to calculate phases for the data to $2.7\text{\AA}$ resolution. The figures of merit were based upon the estimates of the lack-of-closure errors of centric reflections. Figure V.9 shows that the figures of merit were generally satisfactory to about a resolution of $4.5\text{\AA}$, beyond which they fell sharply with resolution.

A protein Fourier synthesis was calculated with all data to a resolution of $4.5\text{\AA}$ (overall figure of merit 0.60). Difference ($F_{\text{PH}} - F_{\text{P}}$) and double-difference ($F_{\text{PH}} - F_{\text{P}} - F_{\text{H}}$) maps showed only small residual features at the heavy-atom sites and no significant ghosting was found in the native protein map.

The electron density maps will be presented and their interpretations discussed in the next chapter.

CHAPTER VI

The Electron Density Maps and their Interpretation

On the basis of existing crystallographic evidence, certain features of the structure of Fv can be predicted. It is very likely, for example, that the basic folding pattern of each half-domain will be similar to that suggested in figure I.4.

On the other hand, by their nature, the variable regions show more variation in this folding pattern than is seen in, for instance, the constant regions. Most of the structural differences between variable regions are found in the inter-strand loops, especially those which make up the antigen-combining site. No specific features of secondary structure can be expected in these loops, although regular features have been observed in other V-regions. In the V_H region of Fab New, one of these loops is quite extensive and forms a two-stranded β -structure which can be regarded as an extension of the three-stranded sheet of the basic structure (Amzel and Poljak, 1979). This extra loop is not present in the V_L subunit. In addition, some inter-strand loops have been observed which contain short stretches of α -helical structure (seldom more than one turn).

The configuration of these portions of the Fv cannot be foreseen and, in fact, there may be some danger in approaching the electron density map with too firm a pre-conceived picture of the structure.

VI.A. Hapten-Binding Studies

With a view to yielding information which might be of some use in locating the antigen-combining site in the map, hapten-binding experiments were initiated relatively early in the project. Some of the soaking trials were described in Chapter IV.

In order to insure that the structure of the hapten-bound complex was

Site	Coordinates			Peak Size(1)
	x/32	y/32	z/52	
1	16	4	14	[48]
2	4	14	5	[46]
3	3	7	36	[45]
4	2	9	33	[43]
5	11	6	38	[43]
6	1	5	29	[42]
7	12	2	25	[41]

(1) Arbitrary Scale

Figure VI.A1. The peaks arising from the hapten co-crystal - native data difference Fourier synthesis at 6Å resolution. The background level was about [35].

as close as possible to that found in solution, an effort was made to crystallize the protein in the presence of the small molecule, dinitroaniline. As the binding constant of this molecule to Fv is very high (see figure II.3), it was envisioned that, if hapten-binding disrupted the crystals, crystallization would not occur in the same form; the binding of hapten would be the dominant process. It is known that crystals of deoxy-haemoglobin are disrupted in the process of oxygenation. The interaction energy that stabilizes the deoxy- structure is of the order of 12 kcal/mole (Perutz,1970), as is that of hapten-antibody interactions (Mukkur,1980).

After the addition of 10 μ l of a 2 mg/ml dinitroaniline solution to 100 μ l of the 6.7 mg/ml Fv solution (resulting in the presence of about 4 moles of hapten to 1 of Fv), 4.5 μ l of PEG stock solution was added to induce crystallization (see section IV.A). This yielded a number of relatively small, bright yellow crystals which appeared to be isomorphous with the native. A set of 6 $\text{\AA}$ resolution data, consisting of one unique section of the reflection sphere, was collected from one of these rather weakly-diffracting crystals on the four-circle diffractometer. Several peaks above the noise level of the difference Fourier synthesis with the native data, calculated with the final 6 $\text{\AA}$ resolution phases, were identified (figure VI.A1) and their positions on the 4.5 $\text{\AA}$ resolution electron density map will be discussed below.

VI.B. The Presentation of the Maps

The Fourier synthesis was made in sections along the y-axis (32 sections in a unit cell). Each section is one unit cell in z by two unit cells in x. Because the c cell-dimension is just greater than twice a, this leads to conveniently shaped sections. Therefore, the dimensions of a section are 138 $\text{\AA}$ by 119 $\text{\AA}$. The β -angle appears between the x and z axes and the sectioning axis is perpendicular to the plane of the section.

The largest calculated value of electron density in the Fourier map was 0.89. However, because the structure factors are not on an absolute scale, this does not represent a real electron density. In addition, no base-line term for the x-ray scattering (the $F(000)$ term) has been included. As a result, the smallest value for the electron density on the protein Fourier map is negative.

In order to relate these calculated electron density values (ρ_{calc}) to true electron densities (ρ_{abs}), we have to estimate the scale factor required to put the F 's on an absolute scale (K_{abs}) and $F(000)$, for use in the equation

$$\rho_{\text{abs}} = K_{\text{abs}} \rho_{\text{calc}} + F(000) / V \quad (\text{VI.1}),$$

where V is the volume of the unit cell (4.59×10^5 , from the cell dimensions). $F(000)$ can be estimated if we approximate that the protein is made up entirely of nitrogen atoms. As there are 8 Fv molecules, each of molecular weight 25250, in the unit cell, then, the scattering contribution from the protein is, $[8 \times (25250/14) \times 7] = 1.0 \times 10^5$ electrons.

The scattering effects of the solvent in the unit cell can be estimated, as it occupies about 1/2 the unit cell, that is, about $2.3 \times 10^5 \text{ \AA}^3$ ($2.3 \times 10^{-19} \text{ cm}^3$). If the solvent is assumed to be water, this volume would contain $2.3 \times 10^{-19} \text{ g}$, or

$$(2.3 \times 10^{-19} \times 6.02 \times 10^{23} / 18) = 7.69 \times 10^3$$

water molecules, that is, 7.69×10^4 electrons. Our estimate of $F(000)$, then, is $(1.0+0.77) \times 10^5 = 1.77 \times 10^5$ electrons, and

$$F(000)/V = 0.39 \text{ electrons/\AA}^3.$$

We can find K_{abs} in equation VI.1 if we propose that the electron density of the solvent (estimated from the map to be -0.10) is slightly more than that of bulk water, say $0.35 \text{ electrons/\AA}^3$. Then,

$$0.35 = K_{\text{abs}}(-0.10) + 0.39 ,$$

$$K_{\text{abs}} = 0.40.$$

Hence, the maximum calculated electron density corresponds to

$$\rho_{\text{abs}} = 0.40 \times 0.89 + 0.39 = 0.61 \text{ electrons/\AA}^3.$$

If this estimate for K_{abs} is correct, the maximum occupancy shown by any of the heavy-atom sites, 1.04 (Pt2), is, on an absolute scale, only $(1.04 \times 0.40) = 0.42$.

The problem of contouring the electron density is one of compromising between too low a contour level, where nearly all regions of the map will be connected by contours and molecular boundaries will be difficult to discern, and too high a level, where it will be impossible to follow a continuous line of density throughout a single molecule. The root mean square error of the map can sometimes be a guide to the appropriate contour level.

The total error in the Fourier calculated with the "best" phase (see section III.F) can be calculated by summing $F^2/V^2 (1-m^2)$ for each reflection (Blundell and Johnson, 1976, p.369), where V is the volume of the unit cell and m the figure of merit of the phase. On the (non-absolute) scale described above, this calculation yielded values of about $0.057 \text{ e-/\AA}^3$ for the root mean square error in electron density for both the $4.5\text{\AA}$ and the $6\text{\AA}$ maps. Inspection of the background levels of each section indicated that this may be a slight underestimate and the final contour levels that were chosen, intervals of 0.076 starting at 0.152, probably correspond quite well to intervals of the r.m.s. error starting at twice the r.m.s. error.

1. The Criteria of Interpretation.

On the basis of the experience of other structure determinations, some criteria can be described for the interpretation of low-resolution maps (North and Phillips, 1968 and Blundell and Johnson, 1976). At $6\text{\AA}$

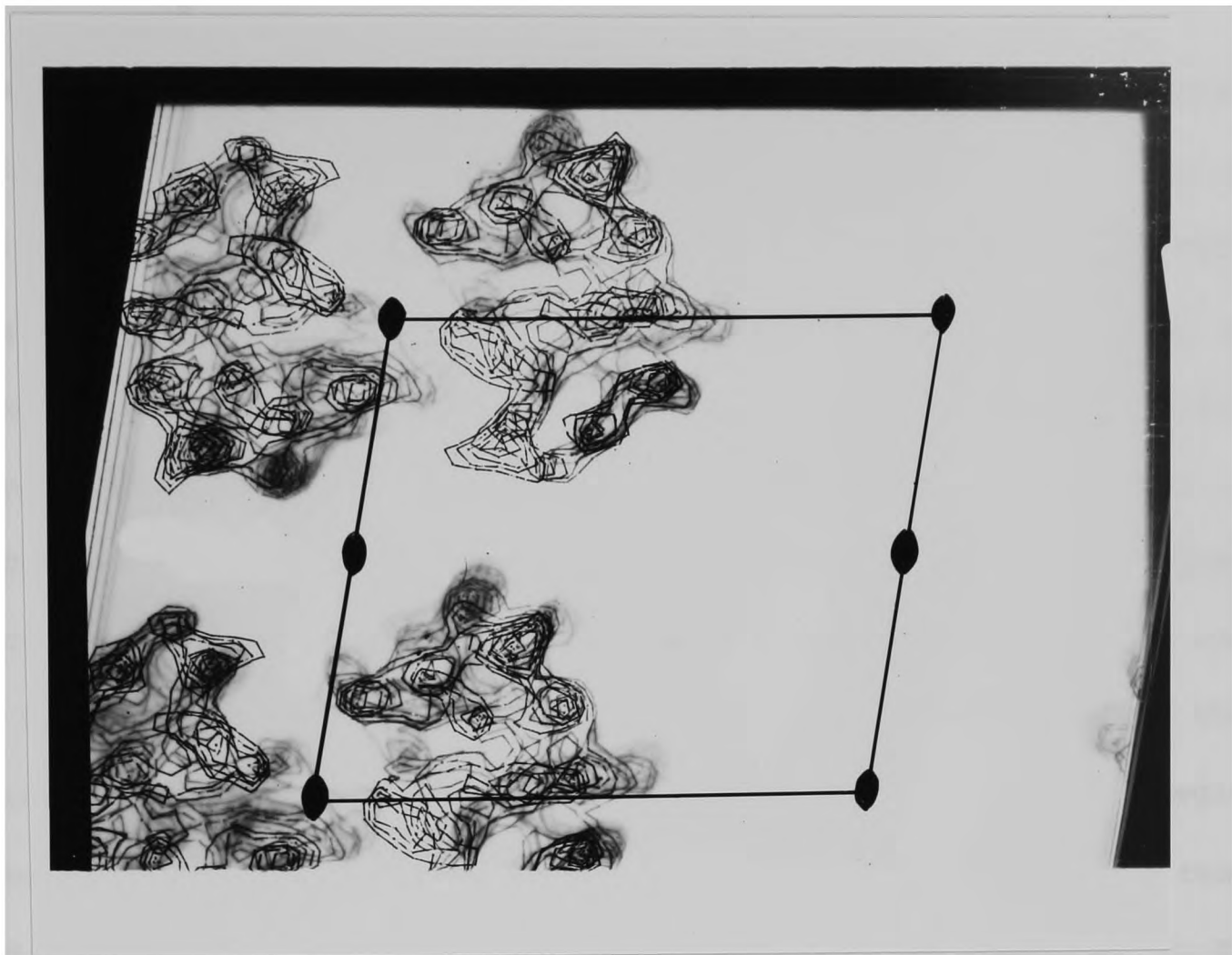


Figure VI.B1. A sample of the calculated electron density from the Fv domain of Fab New (Poljak et al, 1974).

resolution, some parts of a polypeptide chain are visible, but there are often breaks in the chain, especially near the surface of the molecule. Helical structures are usually apparent at this resolution. However, in regions of β -sheet, which we would expect to find in the Fv map, density often runs in the spaces which lie between strands. Thus, while we might hope to discern the orientation of a β -sheet, we would certainly not expect to see individual strands. This expectation is substantiated by the section of the Fourier map calculated to $6\text{\AA}$ resolution from the coordinates of the Fv part of Fab New (figure VI.B1). Here, a suggestion of the direction in which the strands run may be found, but even the region between the two β -sheets which make up the folding unit, is confused, and the disulphide bond is not distinguishable. The two subunits, V_H and V_L , are evident, however, as separate masses of density.

One would expect, though, to see clear regions of solvent of considerably lower mean electron density than the protein. This is not to say that points at which two individual molecules closely approach each other will be clear. If the separation is small enough, density at these points may be as continuous as within a single molecule.

Thus, at $6\text{\AA}$ resolution, one would normally aim not to trace unambiguously the polypeptide chain, but to gain an impression of the shape and dimensions of the molecule and, possibly, to define its boundaries with respect to others in the unit cell.

The improvement in this picture at $4.5\text{\AA}$ resolution is not, in general, significant in terms of viewing individual features of the structure such as side-chains or disulphide bonds. Especially where β -structure is concerned, however, one might gain a clearer impression of individual sheets. In addition, the gaps between molecules should be clearer and except in the most extreme examples, points of close approach should be

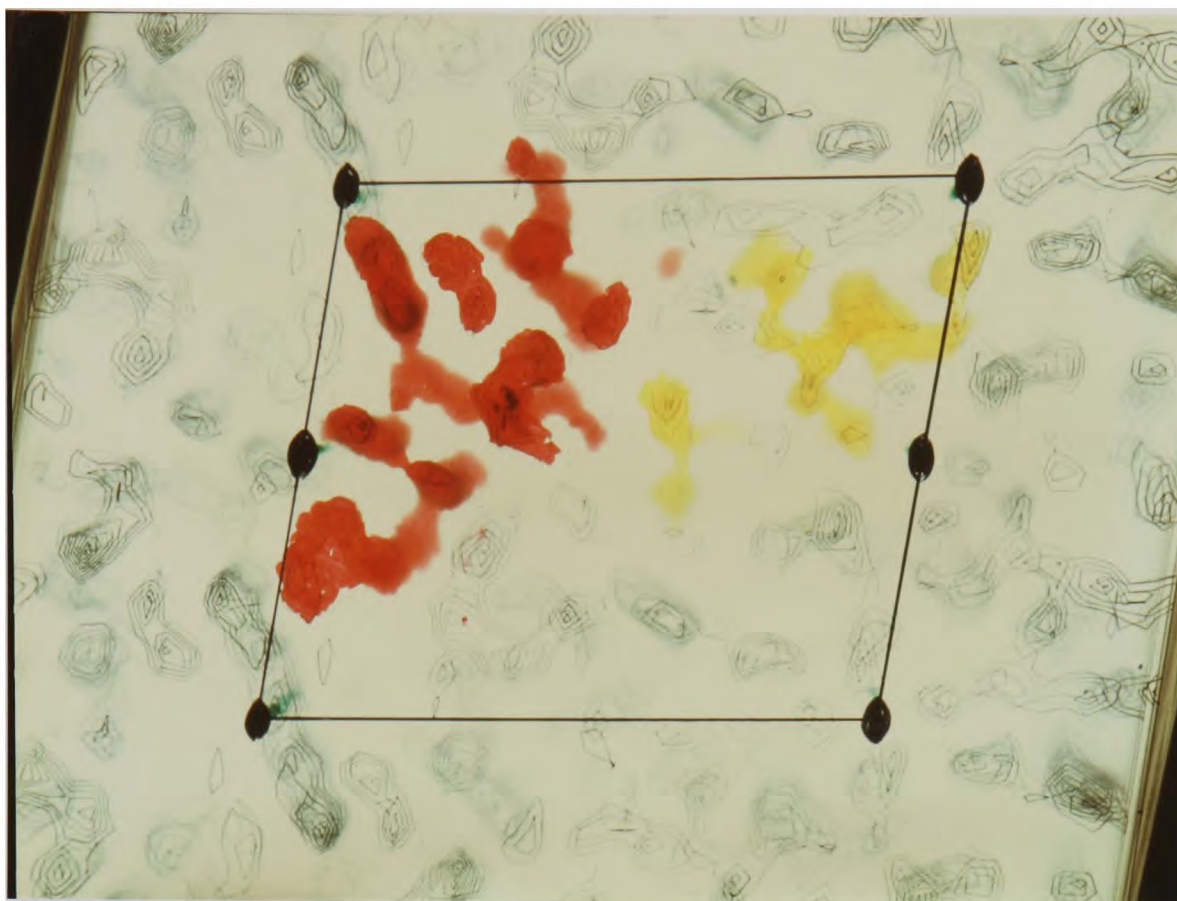
distinguishable.

North and Phillips (1968) presented a series of photographs of a section of the hen-egg-white lysozyme structure at increasing resolutions. Comparing the examples at 5.5 and 4.5Å, one can clearly see regions which appeared continuous at the lower resolution and have become discrete at the higher one. In addition, what was at low resolution a single mass of density has become in the improved map distinctly divisible into smaller units. Comparison with the higher resolution examples of the same region indicates that these smaller units still represent large groups of atoms, but the improvement is clear.

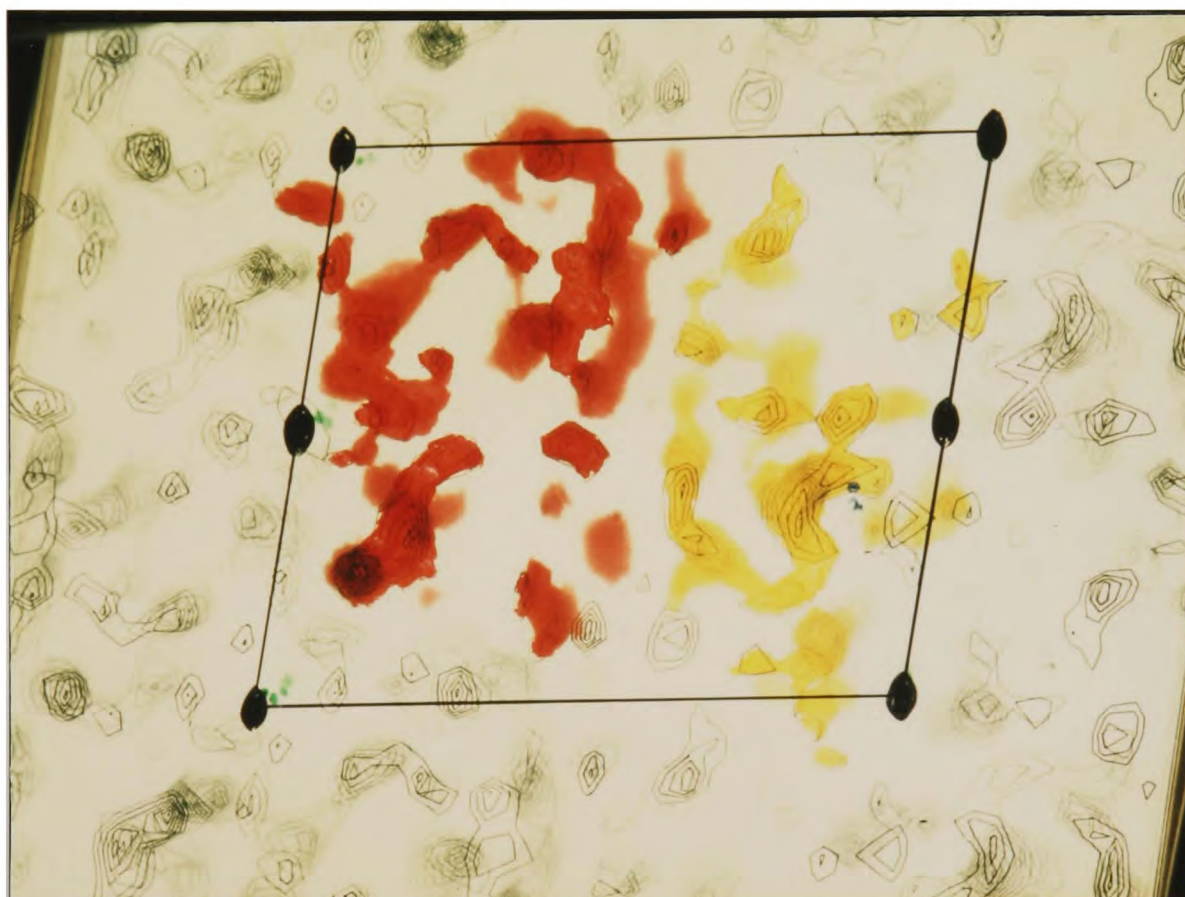
VI.C. The 6Å Resolution Map

Photographs of the 6Å resolution map are presented in figure VI.C1. The positions of the two-fold axes are marked and, for reference, an area of dimensions 60Å by 69Å (a by 1/2 c) shown. The starting point in the interpretation was to identify a large area of continuous density by shading it red and connecting to it all the continuous regions on neighbouring sections. Regions where the continuity was tenuous were left unassigned, initially. As discussed above, points of close approach of molecules may appear continuous at this resolution.

When this procedure had progressed as far as possible, it was repeated with yellow shading, starting with a piece of density as remote as possible from that already assigned. On completion of this stage, most of the density had been coloured either red or yellow. That which remained was clearly part of one region or the other but it was not obvious to which of two crystallographically-related molecules it belonged. For example, the lower area coloured red on photograph a (figure VI.C1) was later shown (with reference to the 4.5Å map) to belong to a symmetry-related molecule; the related density belonging to the rest of the red region is displaced in

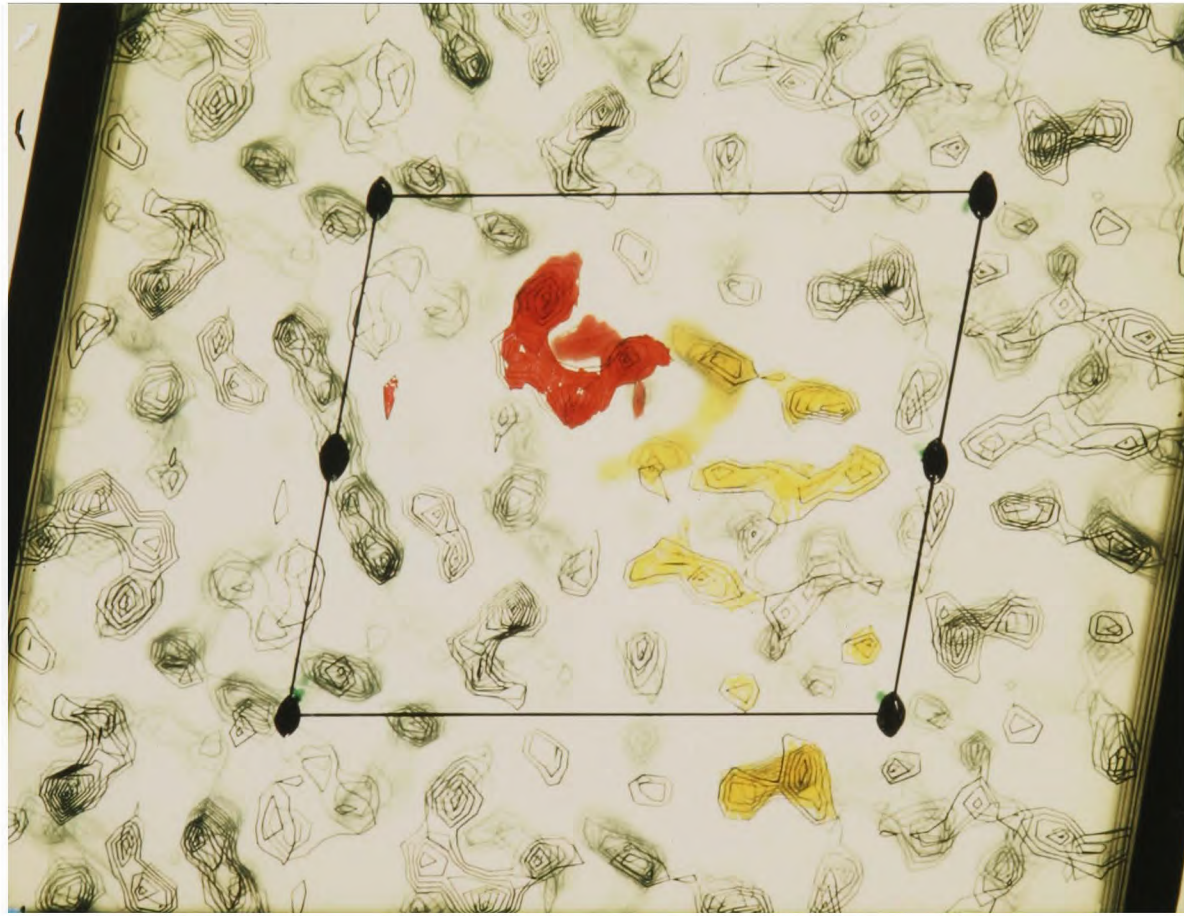


(a) sections 1 to 5

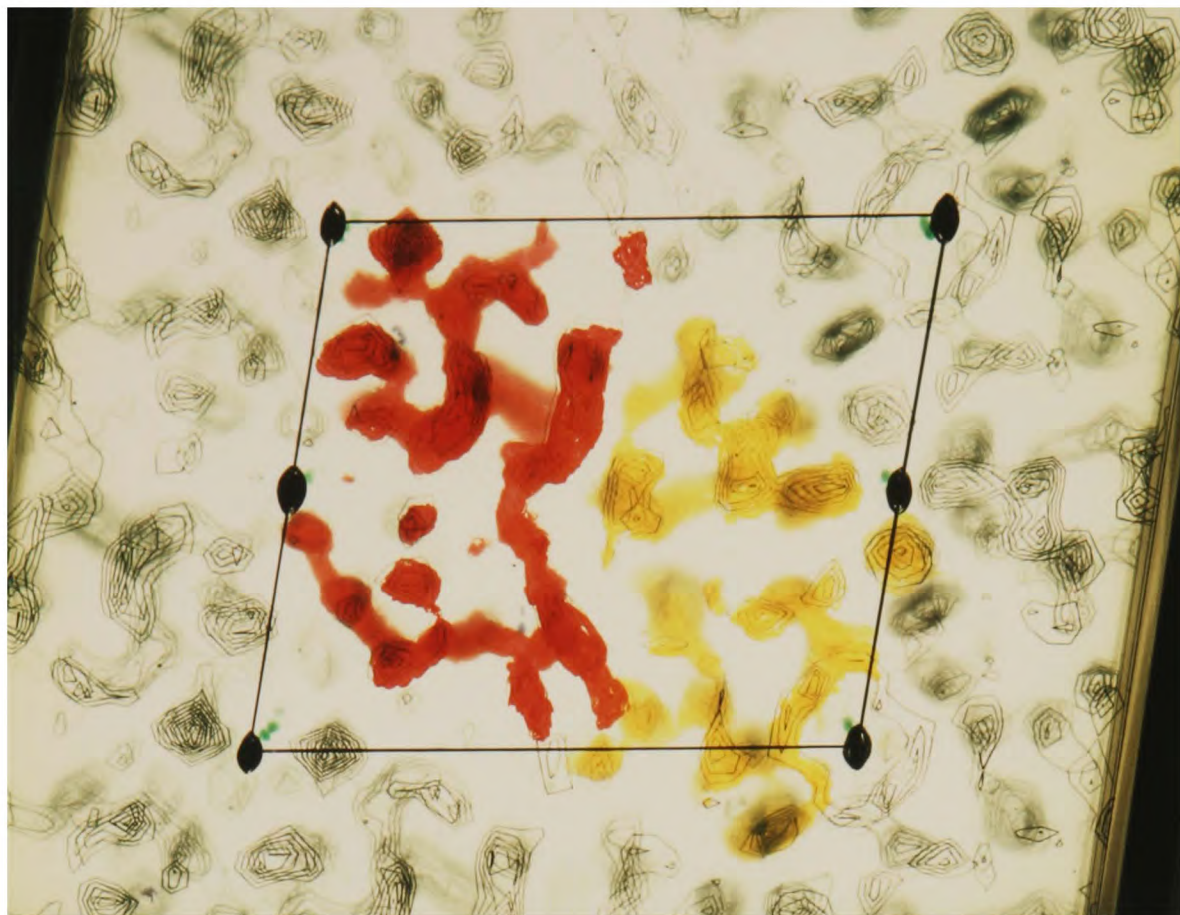


(b) sections 6 to 10

Figure VI.C1. The 6Å Resolution Electron Density Map.
 Contour interval (relative scale) 0.076 starting at 0.152 e-/Å³.



(c) sections 11 to 16



(d) sections 17 to 22

the y direction by $1/2 b$, and can be seen most clearly in photograph d (figure VI.C1) surrounding the two-fold axis at the top left. The continuity is maintained with the same region in photograph VI.C1c.

1. Discussion.

The heavy-atom sites were marked on the map, the platinum in blue and the uranyl in black. In general, they lie close to stretches of electron density, but no outstanding peaks which may have resulted from bias in the phasing, are present.

a. The Red Molecule.

The red-coloured density is a rather large region of density which decreases into a less dense region and then expands again into a larger one, in progressing down the sectioning axis. This is consistent with the two folding units of the Fv fragment and has been interpreted in this way.

Photograph VI.C1a gives an impression of a concentrated area of red density in which a column of density may indicate the direction in which the β -strands run in this folding unit.

Photograph VI.C1b shows more of the first ("upper") folding unit of the red molecule. Note that the direction of the β -strands has changed, marking the lower (three-stranded) sheet of the fold. As described above, the red density in the lower half of the region has been mis-assigned.

The lower folding unit can be seen in photograph VI.C1c, although the picture is complicated by the error in interpretation. A clearer picture of this part of the molecule will emerge in the $4.5\text{\AA}$ map.

b. The Yellow Molecule.

The region of density coloured yellow appears to cover a wider area in the x,z plane but extends over fewer sections. In the main body of the region in photograph VI.C1c, two distinct regions, presumably corresponding to the two folding units, are visible. The direction of the

Platinum Sites (blue disks)

Site	Disk
1	1
2	2
3	4

Uranyl Sites (red disks)

Site	Disk
1	4
2	2
3	3

Hapten Sites (green disks)

Site	Disk
1	3
2	5
3	35
4	34
5	36
6	38
7	73

Figure VI.D1. A key to the coloured disks on the 4.5Å resolution Fv map.

strands is not as obvious as in some parts of the red density. Again, details of the structure will become much more apparent in the 4.5Å resolution map.

c. General Remarks.

Now that we have observed the general layout of the molecules, we can consider some of the characteristics of the structure in terms of the relationship between the subunits.

The interactions between subunits of the same molecule ("trans" interactions in the nomenclature of Amzel and Poljak (1979)), although not covalent, are extensive and stabilizing. In V-region pairs, the contact face of each unit is the three-stranded β -sheet.

In the x,z projection of the Fv map, this interaction is clearer in the yellow molecule, where the homology subunits lie adjacent to one another, than in the red molecule where the subunits lie one on top of the other and are, therefore, superimposed in this position. However, bearing in mind the nature of the interaction, we can get some idea which strands belong to which β -sheet; the three-stranded sheets of the two subunits of each molecule should form the interface.

Although trans interactions are strong, it is not clear how much flexibility can be induced in the structure from forces such as crystal packing. As we shall see, the 4.5Å map appears to provide some evidence that the relationships between the subunits of the two Fv molecules are different, to some extent.

VI.D. The 4.5Å Resolution Map

The dimensions of the 4.5Å map are identical to those of the 6Å one. Again the area 60Å by 69Å is indicated on each photograph.

The heavy-atom sites and the peaks in the hapten-bound data - native data difference Fourier are shown by coloured circles. A key is given in figure VI.D1. In general, higher numbers represent less significant sites.

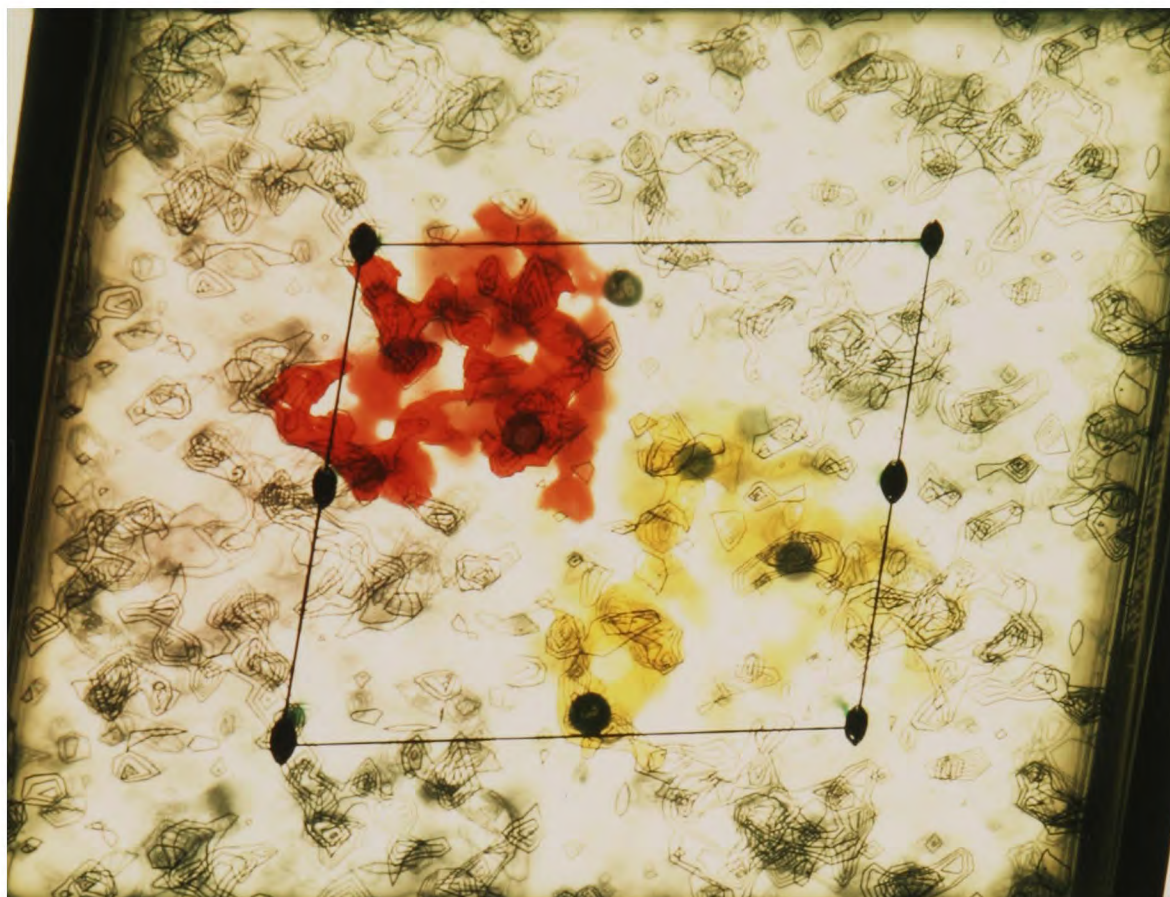


Figure VI.D2. The 4.5Å Electron Density Map.
Overall View - sections -1 to 14.
Contour interval (relative scale) 0.076 starting at 0.152 e-/Å³.

The procedure for interpreting the map in terms of continuous regions of density was the same as that for the 6Å map. Figure VI.D2 is an overall view of the electron density map. Again, the yellow density covers a more extensive area and the two lobes representing the two subunits are visible. The red density is very compact but covers many more sections.

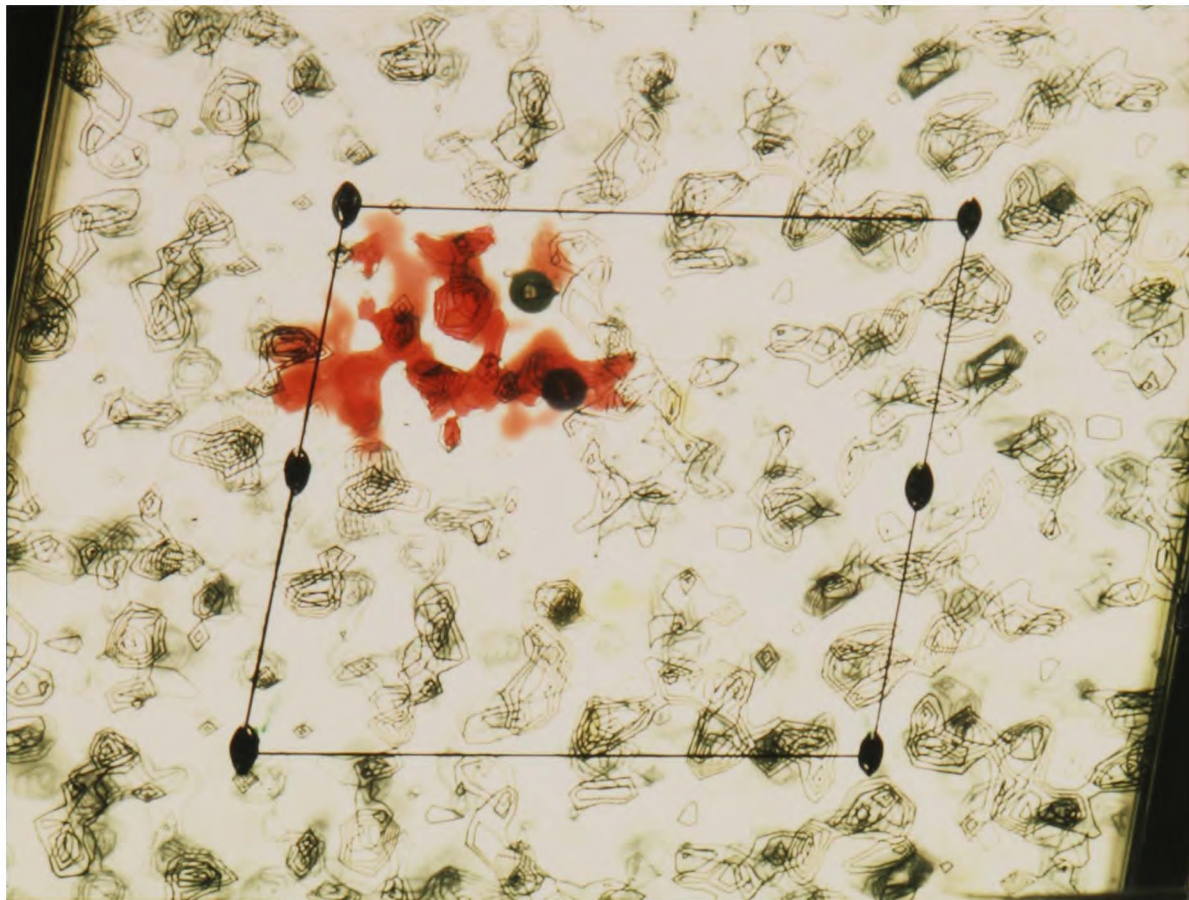
In comparison with the 6Å map, the overall layout of the coloured densities is the same. The gaps due to solvent are in the same places between the red and yellow regions. Within the red region, the interpretation of the continuous part which appeared to be mis-assigned above, has been changed. Although a single contour had to be broken in this process, the new connection which has been found is more convincing (see the extreme left of photograph VI.D3b, joining with the top of the red density on photograph VI.D3d).

1. The Red Molecule.

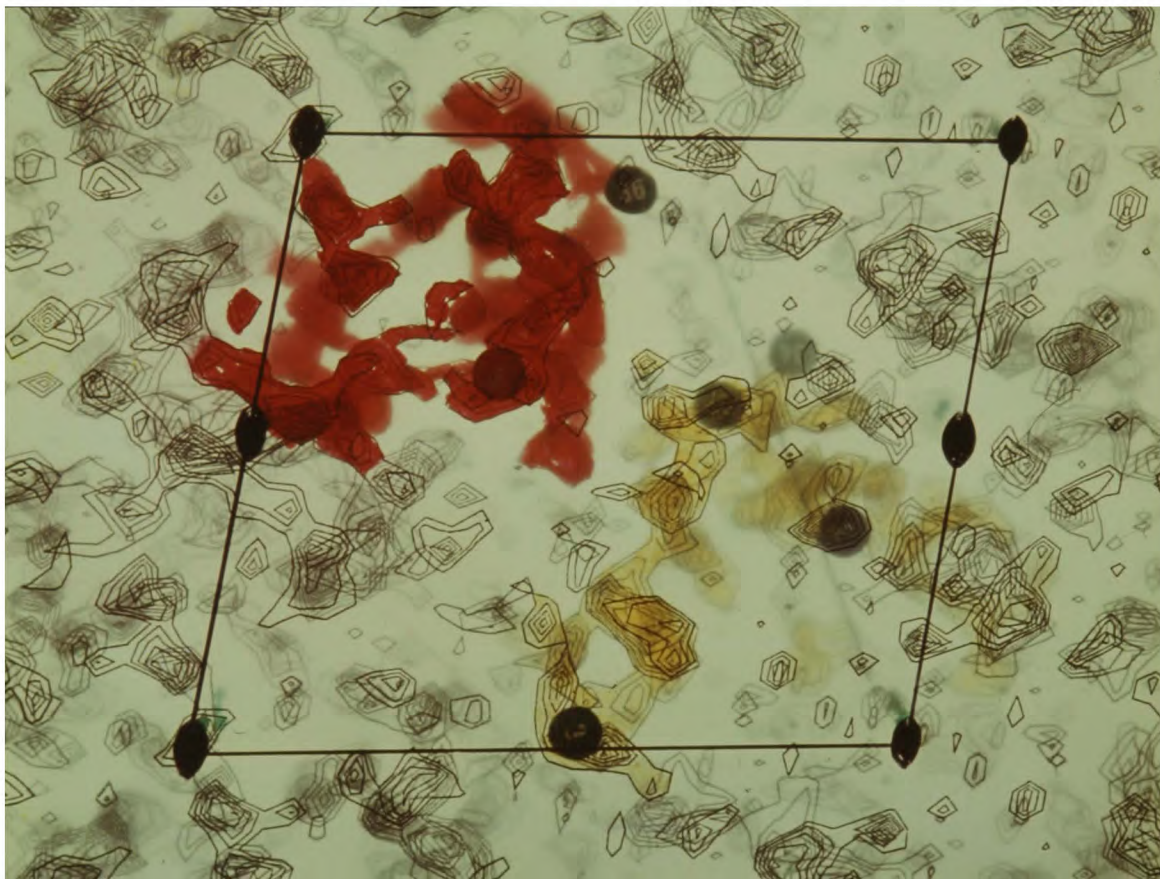
In photograph VI.D3a nearly all the visible red density is part of the upper subunit. Photograph VI.D3b probably gives the best view of the relationship between the two red subunits. This shows the lower half of the upper subunit with the strands of the three-stranded sheet seemingly running almost parallel to the sectioning axis. The lower subunit is just coming into view to the left of the upper, and its three-stranded sheet, which should form the contact with the upper subunit appears to be oriented from lower left to upper right.

Thus, the view of this molecule seems to be approximately down an axis along which the molecule is most extended. It therefore, as a whole, appears long and of small cross-section.

When the results of the hapten-binding experiment were examined, it was clear that nearly all the peaks occurred in the yellow region. There certainly are no indications that a hapten has bound in the region between

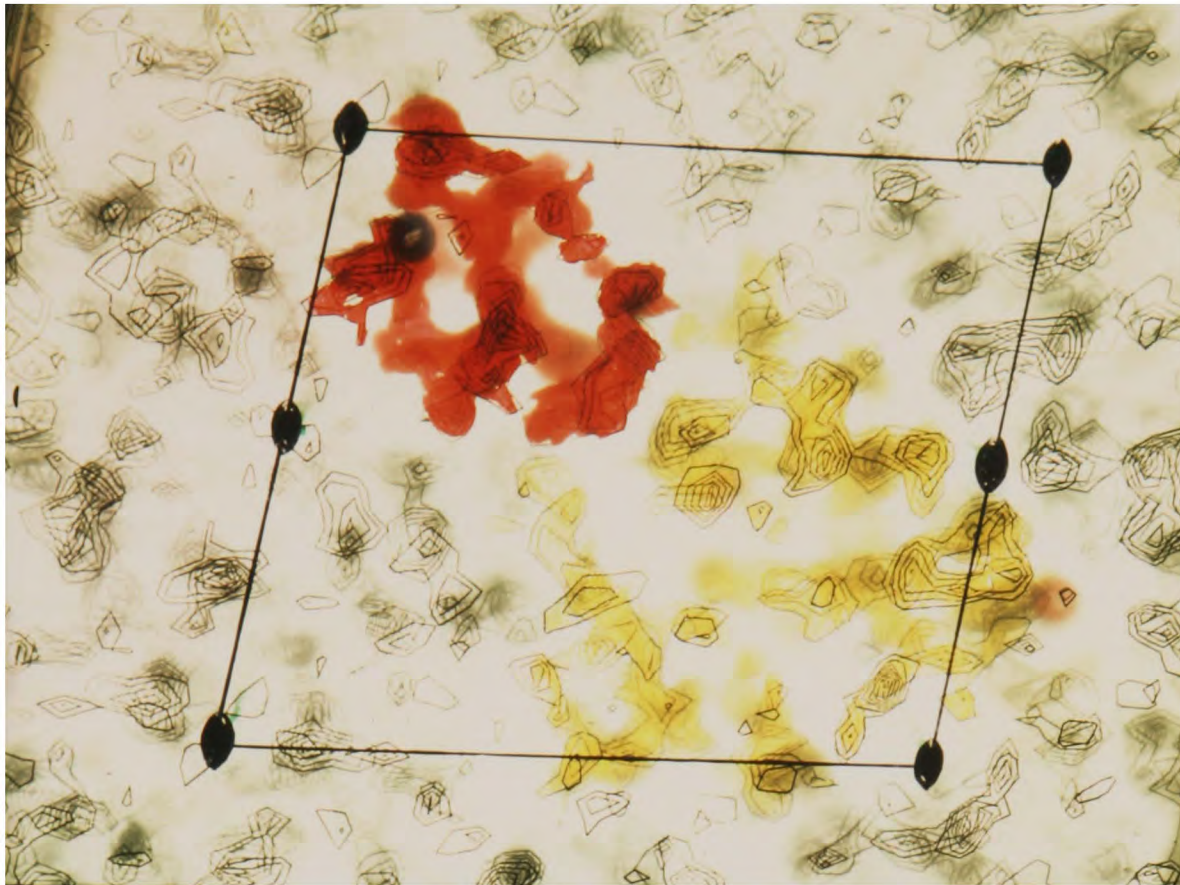


(a) sections -6 to 2

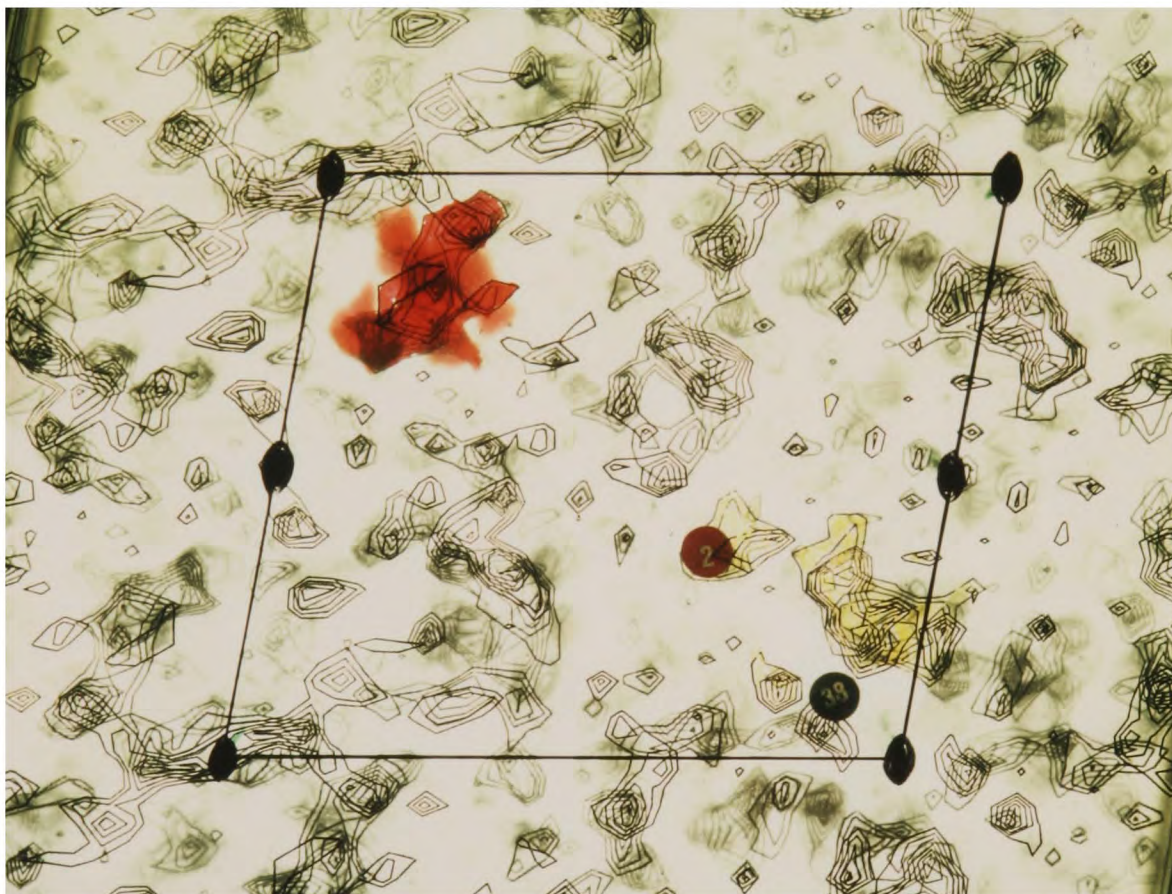


(b) sections 3 to 9

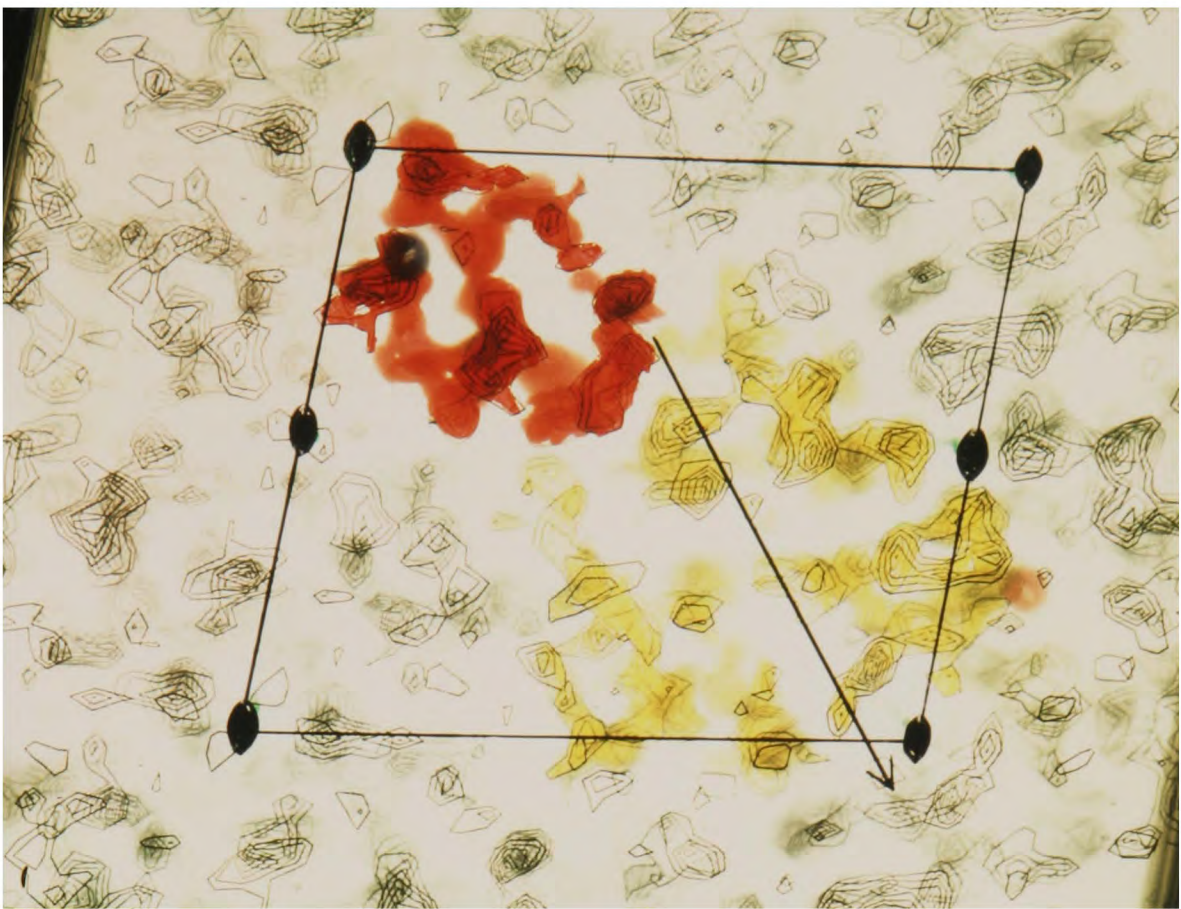
Figure VI.D3. The 4.5 Å Electron Density Map.
Contour interval (relative scale) 0.076 starting at 0.152 e-/Å³.



(c) sections 10 to 20



(d) sections 20 to 27



(e) sections 10 to 16



(f) sections 17 to 22

the two red subunits. The red molecule does seem to be much more closely packed than the yellow one with respect to analagous molecules from neighbouring asymmetric units. It may be that the crystal-forming forces in some way disrupt the conformation of the binding site of this molecule. This conclusion would cast doubt on the assumption made earlier that the hapten-binding forces are stronger than the crystal-forming ones. It is true, however, that in solution, the hapten and protein molecules are in equilibrium. Consequently, there are always protein molecules present to which hapten is not bound. Perhaps these molecules have been incorporated into the crystals. Clearly, a resolution of this problem will have to await a more definitive set of hapten-bound data.

2. The Yellow Molecule.

Although the volume occupied by the yellow density is virtually identical to the red, it is much shallower along the sectioning axis. The two subunits are seen clearly in photograph VI.D3b, the area of interaction between them being closest at the top left corner of the yellow density. The proposed position for the local two-fold axis relating the two subunits of the yellow molecule is shown in the background of photograph VI.D3b and is more clearly visible in photograph VI.D3e.

The assignment of a small region of density in the $6\text{\AA}$ map (figure VI.C1b, the uppermost yellow-coloured area) has been reassigned to a symmetry-related yellow molecule. The density in question can be seen in photograph VI.D3f surrounding the red disk (number 2). The area in the $4.5\text{\AA}$ map which corresponds to that originally coloured in the $6\text{\AA}$ map, is on photograph VI.D3b to the right of the green disk (number 36, in the red density).

Strands of density running from upper left to lower right (approximately parallel to the two-fold axis) in the lower part of

photograph VI.D3e, and from left to right above the two-fold axis, are visible. Because of their positions away from the closest area of contact between the subunits, these are both probably parts of the four-stranded β -sheets.

There are two peaks on the hapten-bound difference Fourier map which occur in the region of interaction of the two yellow subunits. These features are visible as green disks (numbers 34 and 35) on photograph VI.D3b at the top of the yellow density, and provide some encouragement for future hapten-binding studies.

The layout of the yellow molecule is more obvious in terms of picturing the positions of the folding units, because the two subunits are clearly visible separately. If one were to describe the orientation of the red molecule with respect to the yellow one, one might propose that the view of the red molecule is similar to that of the yellow viewed down an axis approximately parallel or perpendicular to the x-axis (compare the two molecules in photograph VI.D3c, for example). The two possibilities cannot be distinguished at this resolution, as they are based on the overall shapes of the molecules.

3. Final Remarks.

We have seen that the 4,5Å resolution map can be interpreted in terms of two Fv molecules, each of which is divisible into two clear subunits. Some of the features of the subunits, for instance the direction of the β -strands, can be discerned in places. The largest dimensions of the two molecules in the x, y and z directions are roughly 33Å × 57Å × 37Å (red) and 53Å × 37Å × 47Å (yellow).

The orientations of the two molecules in the asymmetric unit are quite different, as are their packing properties with neighbouring molecules. It is not yet clear whether or not there are local differences

in the structures of the two molecules, but some preliminary evidence from hapten-binding studies implies that their binding properties may not be identical. The situation may be analagous to one observed in crystals of triose phosphate isomerase from chicken muscle (Rivers,1977). The crystals were grown in the presence of sulphate, an inhibitor to the activity of the enzyme. On transfer to a phosphate solution one subunit shows a conformational change into the enzymically active form, while a second crystallographically unrelated one maintains its inactive conformation, stabilized, apparently by inter-molecular contacts in the crystal. It is important to recall, however, that for Fv, this evidence is based on a comparatively weak data set and is certainly not conclusive.

In any case, the interactions between the subunits of both molecules are very strong, as expected from previous structural studies. It is also encouraging that the solvent boundaries between the molecules could be maintained throughout the interpretation and that a single contour had to be broken at only one point (as described earlier) despite the relatively low resolution of the map.

The implications of the Fv structure at this resolution and the directions in which the project could progress are discussed in the final chapter.

CHAPTER VII

Conclusion

This thesis has described the steps leading to the calculation of an electron density map of the Fv fragment of mouse myeloma protein M315 at a resolution of $4.5\text{\AA}$. It is clear from the discussion in the preceding chapter that the unambiguous interpretation of a map at this resolution in terms of a protein structure is impossible. The interpretation described in the previous chapter was made with a pre-conceived notion of the structure of Fv and of how this structure would appear at $4.5\text{\AA}$ resolution.

This notion was based largely on the examination of low resolution electron density maps of two similar molecules, the variable domain of Fab New (Poljak et al, 1974) and the structure of the rabbit Fc fragment (Sutton, 1980). A picture of a sample of the former was presented in figure VI.B1, in the discussion of the criteria for examining low resolution maps. This map was also useful in identifying specific features of electron density characteristic of an immunoglobulin which were also found in the Fv map. For example, one feature visible in figure VI.B1 which occurs more than once in both the Fab New and the Fv maps is one in which a stretch of electron density suddenly doubles back into another stretch almost parallel to the first. It is features like this which provide confidence, not in every detail of the interpretation, but that the density in the map represents a structure in agreement with the expectations of that of Fv.

Because of the inherent differences between constant and variable region structures, the Fc map is not expected to be as good a basis for comparison as that discussed above. However, it is a result arrived at by very similar methods and in very similar conditions to the Fv result. The

Fc map is of higher quality than the Fv one. The protein phases were based on three heavy-atom derivatives, two of which contributed anomalous scattering information to 4.5Å resolution.

One striking characteristic in the 4.5Å resolution Fc map can be clearly seen at the interface between the analogous subunits (such as the two C_H³ regions), which is formed by the four-stranded β-sheet of the basic fold. The strands in the adjoining sheets run in different directions and can be seen forming a cross in the electron density. This feature is not obvious in the Fc map (nor is it evident in the calculated Fab New map). Certainly the Fc molecule lies in an ideal orientation in the unit cell for this feature to be viewed, and the orientation of the Fv is quite different. Furthermore, this cross may be a product of the relatively planar nature of the Fc β-sheets in comparison to those of the Fv. However, the Fc map does have local features, such as the density reversal described above, in common with the Fv map.

The point of this discussion, then, is that we can see features in the Fv electron density which provide some assurance that it represents the β-sandwich structure we would expect. We cannot, however, put the interpretation of Chapter VI beyond doubt because the larger-scale interaction between subunits is not as clear as it was in, for example, the Fc map at the same resolution, whether this be due to a poorer quality map or to the inherent nature of the Fv molecule.

This interpretation does, on the other hand, provide a base from which the study of the Fv structure can proceed in a number of directions, some of which will now be considered.

VII.A. The Improvement of the 4.5Å Resolution Result

The most obvious approach to such an improvement is to find additional heavy-atom derivatives. The uranyl and platinum derivatives, though useful

at low resolution, were clearly not of the highest quality and showed signs of decreasing value beyond 4.5Å. Some of the compounds examined in the initial search (Chapter IV) yielded results suggesting that, under appropriate conditions, they may be of some use. In fact, a number of compounds not previously tried are at present being investigated (B.J.Sutton and V.Yadava, personal communication).

In the absence of a new derivative, those used previously may be exploited further. For example, no anomalous scattering measurements have been made beyond 6Å resolution. In view of the decreasing quality of the derivatives beyond 4.5Å resolution, a sensible way to proceed might be to return to the diffractometer, where the measurement of small intensity differences is more practical and more accurate. Based on the statistics resulting from the data processing (Chapters IV and V), the 6Å data were of a higher quality than those to 4.5Å resolution. The result at 6Å resolution suggests that the 4.5Å electron density map might be improved considerably by the improvement of the data and the inclusion of anomalous scattering information.

A third alternative is to take advantage of the non-crystallographic symmetry that may exist between the two Fv molecules in the asymmetric unit. Bricogne (1976) described a method for deriving phase information from the presence of two or more crystallographically independent molecules. A basic assumption that would have to be made by this method is that the two Fv fragments in the asymmetric unit are in identical conformations. Another is that there are, in fact, two complete Fv domains present in the asymmetric unit. Professor D. Davies (personal communication) found that on crystallization, another Fv fragment dissociated into separate V_H and V_L units.

Although they have the same amino-acid sequences and, presumably,

basically the same structure, the two Fv molecules are fragments of a larger molecule and the folding units within each Fv are not held together by any covalent forces. It may be that in different crystallographic environments, local variations in the conformations may occur. However, the discussion in Chapter VI indicated that the inter-subunit contacts were tight, and the powerful technique of phase averaging is one which may be useful in clarifying the structure.

VII.B. The Generation of a Starting Model

There are many procedures for the refinement of a model structure with respect to an electron density map but all of them depend on a close initial approximation to the map. Clearly, the next major step in the interpretation of the Fv map is the assignment of atomic coordinates to the electron density. This may be facilitated by the existence of coordinates of Fab structures from previous crystallographic investigations (see Chapter I). The danger of such an approach was mentioned earlier. Furthermore, these coordinates must be superimposed on the Fv density, by no means a straightforward step at this resolution.

This might be attempted visually by using suitable apparatus such as a Richards box (Richards, 1968), in which the electron density map and an atomic model on a suitable scale appear superimposed in a half-silvered mirror held at 45° to both. The model could then be rotated manually until it agrees as well as possible with the map.

Alternatively, computational methods may be used to maximize the agreement of the two structures either in real or reciprocal space. The rotation function of Rossmann and Blow (1962), as formulated by Crowther (1971), has been calculated to compare the Fv 6Å structure with that of the Fv domain of the Fab of IgG New (M. Lewis, personal communication). Two peaks of maximum superposition were found but it has been impossible

so far to reconcile these results with the electron density maps. It is perhaps not surprising that this method should not yield clear results. There are four folding units in the asymmetric unit, two in each Fv. Therefore, especially since the folding units of different Fv molecules are close together at several points in the unit cell, there are a number of ways in which the model structure could be rotated to yield good overall agreement with the electron density. However, a reassessment of this approach, in particular with an improved electron density map, might be helpful.

VII.C. The Extension of the Resolution

Once a good model of the protein is obtained, higher angle native data can be included even without good heavy-atom derivatives. The phases of reflections for which there is no independent information can be derived from the model, and those which have been found from isomorphous replacement can be improved by combining them with calculated phases.

Native data have already been collected photographically to a resolution of $2.7\text{\AA}$. Some preliminary experiments have been performed on the DESY synchrotron facility at Hamburg, and have shown that useful native data can be measured to $2.0\text{\AA}$ resolution. This result has recently been confirmed at the Science and Engineering Research Council synchrotron facility in Daresbury (B.Sutton and V.Yadava, personal communication). The path to a high resolution structure from a starting model will most likely, therefore, be relatively straightforward.

VII.D. Hapten-Binding Studies

Most of the interesting results from this project in terms of structural immunology will come from the examination of the complex of Fv with several dinitrophenyl haptens. Some of the issues which might be addressed were mentioned at the end of Chapter II.

The preliminary co-crystallization experiments of Fv with dinitroaniline were encouraging, especially in the apparent isomorphism of the resulting crystals. As data of better quality and higher resolution become available, it should be possible to see the hapten more clearly.

It will also be very interesting to observe in the native structure, the location of those amino-acid residues which have been identified as being responsible for the idiotypic characteristics of M315. In addition to increasing the understanding of the antibody molecule, studies of this kind are important in discovering how the antigenic properties of macromolecules are formed, knowledge which may be important medically if methods can be developed for selectively stimulating the immune system against a specific antigen. The application of such a method would be widespread in the treatment of disease caused by factors with identifiable antigenic determinants.

In conclusion, then, this study has provided the foundation for a detailed analysis of the atomic structure of the Fv fragment. In the near future, the interpretation described in Chapter VI will be confirmed or modified by the addition of information from more heavy-atom derivatives and higher resolution data. Once the path is clear to the unambiguous assignment of atomic coordinates, the incorporation of the 2.7Å native data which already exist and the addition of further data could lead to a structure to 2.0Å or beyond, and make possible all the immunologically-interesting investigations that have been discussed.

APPENDIX I

Computer Programs

Oxford University ICL 1906A and 2980

PRIMROSE (S.French and P.R.Evans)

Integrates the four-circle diffractometer output by the ordinate analysis method, applies the absorption, Lorentz and polarization corrections.

OSCAR (D.G.R.Yeates, modified for the 2980 by E.A.Stura)

Used to determine the crystal orientation on the rotation camera, the indexing of the reflections, and the derivation of numerical intensities from the spots on the photograph.

FILMPACK (K.S.Wilson, modified for the 2980 by E.A.Stura)

Analyses and merges together data collected on the superimposed films within a single filmpack.

SSM (E.J.Dodson and K.S.Wilson, modified for the 2980 by E.Stura)

Sorts equivalent reflections, calculates and applies scale factors to subsets of data based on multiple measurements of equivalent reflections, merges subsets into a single unique data set, adjusts standard deviations.

TRUNCATE (K.S.Wilson)

Used to correct negative intensities from photographic data according to an all-positive Wilson distribution.

CAD (E.J.Dodson and K.S.Wilson, modified for the 2980 by R.Pickersgill)

Collects together several data sets and produces a single list of reflections.

ANSC (P.R.Evans, modified for the 2980 by S.Gower)

Analyses and scales data sets as a function of index, intensity, resolution and volumes of reciprocal space. Used in the absorption correction of the film data.

FHLEKR (G.Bentley, P.R.Evans, E.J.Dodson, K.S.Wilson)

Calculates and analyses F_{HLE} coefficients and refines the derivative-to-native scale factor.

REFINE (Busing, Levy and Martin, P.Main, modified by E.Dodson and K.Wilson)

F_{HLE} or phased refinement of heavy-atom parameters.

PHASE (G.Bentley, M.J.Geisow, E.J.Dodson, modified for the 2980 by S.Gower)

Calculates Blow and Crick 'best' and most probable phases, assigns figures of merit, analyses protein phases and lack-of-closure errors in terms of the heavy-atom phases and structure factor amplitudes.

FFT (L.Ten-Eyck, modified by E.Dodson, P.Evans and, for the 2980, W.Pulford and J.Fail)

Fast Fourier Transform used to calculate Fourier, difference Fourier, and Patterson syntheses.

CONTOUR (S.H.Banyard, P.E.Nixon, P.R.Evans, K.W.Snape, S.J.Oatley, modified for the 2980 by R.Pickersgill and J.Fail)

Draws three-dimensional contour maps of Fourier syntheses.

PDP 11/70

MSCALE (V.Yadava)

Local scaling of two sets of data in volumes of reciprocal space.

REFINE(11) (P.Shaw)

Phased refinement of heavy-atom parameters.

PHASE(11) (P.Shaw)

Calculates Blow and Crick 'best' phases and assigns figures of merit.

RMSERROR (D.R.Rose)

Calculates the root mean square error in a Fourier synthesis by the equation in section VI.B.

C2FOR (J.Campbell)

C2 space group specific Fourier calcution.

C2VERT (D.Stuart)

Prepares Fourier output for contouring.

MAO (D.Stuart)

Extends boundaries of the Fourier from the asymmetric unit.

KONTOR (J.Campbell)

Contours the Fourier output.

ALTEXT (A.Turner)

Prints text on matrix plotter.

APPENDIX II

Aschaffenburg, R., Phillips, D.C., Rose, D.R., Sutton, B.J., Dower, S.K., Dwek, R.A.
(1979) Biochemical Journal 181, 497-499.

Crystallization of the Fv Fragment of Mouse Myeloma Protein M315

Rudolf ASCHAFFENBURG,* David C. PHILLIPS,* David R. ROSE,*
Brian J. SUTTON,* Steven K. DOWER† and Raymond A. DWEK†

*Laboratory of Molecular Biophysics, Department of Zoology, University of Oxford,
South Parks Road, Oxford OX1 3PS, U.K., and

†Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(Received 17 April 1979)

The Fv fragment of mouse myeloma protein M315 was crystallized from poly(ethylene glycol) solution in the form of monoclinic crystals, space group *C*2 and unit cell dimensions $a = 5.96\text{ nm}$ ($59.6\text{ \AA}$), $b = 5.66\text{ nm}$ ($56.6\text{ \AA}$), $c = 13.79\text{ nm}$ ($137.9\text{ \AA}$) and $\beta = 99.7^\circ$. Some unusual effects of poly(ethylene glycol) on protein crystals were noted and are discussed.

The Balb/c mouse myeloma (immunoglobulin A, λ 2) protein M315, first isolated by Eisen *et al.* (1968), has been of particular interest in structural studies of immunoglobulins because of two unusual features: first, its light chain, designated λ 2, differs in the amino acid sequence of its constant region from that of the prototype λ 1 chain (Dugan *et al.*, 1973; Tonegawa *et al.*, 1978); secondly, the protein is susceptible to an unusual cleavage by pepsin that allows its Fab (antigen-binding) fragment to be further digested into a smaller fragment, called Fv, which comprises the variable regions of the light chain and heavy chain and shows hapten-binding behaviour identical with that of the intact immunoglobulin A molecule (Inbar *et al.*, 1972).

Hochman *et al.* (1973) have shown by isoelectric focusing that the reaction yields a homogeneous product with a molecular weight, by sedimentation equilibrium, of 25250 ± 1200 . By comparing the amino acid composition of the light-chain variable region with the sequence determined by Dugan *et al.* (1973), they deduced that the primary point of cleavage was between residues 114 and 117 of the light chain. Similar reasoning applied to the heavy chain, which has since been sequenced by Francis *et al.* (1974), suggested that here the main cleavage point was probably between residues 114 and 116.

The Fv fragment has been shown to combine with various haptens containing 2,4-dinitrophenyl and 2,4,6-trinitrophenyl groups, and a model of the combining site was built by using structural information from other Fab fragments (Padlan *et al.*, 1976). This model was modified and refined by using data from n.m.r., e.s.r. and chemical modification studies (Dwek *et al.*, 1977; Sutton *et al.*, 1977), but some details are still being called into question (Kumar *et al.*, 1978).

In the structural studies of immunoglobulins, X-ray crystallography has clearly been an important tool (for recent reviews see Poljak, 1978; Kabat,

1978), but a previous attempt to obtain suitable crystals of the Fab fragment of myeloma protein M315 was unsuccessful (Inbar *et al.*, 1971). We now report the crystallization of the Fv fragment of myeloma protein M315.

Materials and Methods

Preparation

Reduced and alkylated myeloma protein M315 immunoglobulin A monomer was extracted from mouse serum by using an ϵ -N-dinitrophenyl-lysine-Sepharose column, as in the previously published procedure (Goetzl & Metzger, 1970). The immunoglobulin A was eluted with 0.1 M-ammonia, according to a slight modification suggested by D. Givol (personal communication). The pH of the fractions was lowered immediately to 8.2 when necessary, and the fractions with $A_{280} > 0.05$ were pooled and dialysed overnight against 10 litres of 0.05 M-sodium acetate buffer containing 0.15 M-NaCl at pH 4.7. The protein solution was then concentrated to 10 mg/ml by ultrafiltration (Amicon PM 30), its pH lowered to 4.5 with acetic acid, and sufficient pepsin added to give a concentration of 0.2 mg/ml. After incubation at 37°C for 3 h to release the Fab fragment, the solution was further acidified to pH 3.5 with acetic acid and again incubated for 3 h to release the Fv fragment (Inbar *et al.*, 1972). The digest was neutralized with 2 M-Tris, dialysed overnight against 10 vol. of 0.05 M-Tris buffer containing 0.15 M-NaCl at pH 7.2, and centrifuged at 15 000g for 20 min. The Fv fragment was then isolated by applying the concentrated (Amicon PM 10) supernatant to a Sephadex G-75 column (80 cm $\times$ 3 cm), and elution with the Tris-buffered NaCl at pH 7.2. The fractions containing Fv fragment (elution volume approx. 200 ml) were pooled, dialysed overnight against 5 litres of 0.05 M-NH₄HCO₃, pH 7.0, and freeze-dried.

After reduction and alkylation, the product gave two merging bands on sodium dodecyl sulphate/polyacrylamide-gel electrophoresis, corresponding to the variable domains of the light chain and heavy chain.

Crystallization

Single crystals suitable for X-ray analysis were prepared by using poly(ethylene glycol) (mol.wt. 6000) (BDH Chemicals, Poole, Dorset, U.K.). Freeze-dried Fv fragment was dissolved in 0.1 M-imidazole/HCl buffer, pH 7.0, to a clear solution containing 1 mg/150 μ l of buffer. Samples (100 μ l) were pipetted into narrow glass tubes, and 35 μ l of poly(ethylene glycol) stock solution was added. The stock solution consisted of equal weights of poly(ethylene glycol) and water; the rather viscous mixture was stirred before each use to enhance its fluidity. On addition of the poly(ethylene glycol), the protein solution goes distinctly turbid, and on standing fine floccules form on the walls of the tubes. Changes occur, usually after 1–2 weeks, when isolated crystals are seen to develop between the floccules. The size and number of crystals tend to increase subsequently over a period of several weeks. Large crystals are usually distinctly yellow, probably owing to traces of the dinitrophenyl hapten used in the isolation procedure.

X-ray studies

Precession and oscillation X-ray photographs were taken by using nickel-filtered Cu K_α radiation, generated at 40 kV and 40 mA.

Results

The crystals are prisms with rectangular cross-section, elongated along the z-axis, the best developed form being {110}. The space group is C2, with unit cell dimensions $a = 5.96 \pm 0.01$ nm (59.6 ± 0.1 Å), $b = 5.66 \pm 0.01$ nm (56.6 ± 0.1 Å), $c = 13.79 \pm 0.02$ nm (137.9 ± 0.2 Å) and $\beta = 99.7 \pm 0.2^\circ$.

The diffraction pattern extends to a resolution of better than 0.28 nm (2.8 Å), and crystals mounted in sealed glass capillary tubes at room temperature sustain over 100 h of exposure to X-rays before showing noticeable radiation damage.

In a survey of 116 different crystal forms of globular proteins, Matthews (1968) observed that the fraction of the crystal volume occupied by solvent is most commonly near 43% and always between 27 and 65%. Only if two molecules of Fv fragment of molecular weight 25250 lie in the asymmetric unit (8 molecules in the unit cell) is the calculated fractional solvent content, 47%, within this range. However, the density of the crystals was measured on xylene/

bromobenzene density gradients (Low & Richards, 1952) and found to be 1.27 ± 0.03 g/ml. Despite this unexpectedly high value (see the Discussion section), packing considerations and the extremely low fractional solvent content of 22% resulting from the analogous calculation lead us to believe it unlikely that there are three molecules in the asymmetric unit.

Discussion

The use of poly(ethylene glycol) as a crystallizing agent (McPherson, 1976) has become more common in recent years, though its mode of action is not clearly understood. In examining crystals of Fv fragment and others grown from poly(ethylene glycol) in this laboratory, we have found that the measured density of the crystals is greater than expected. One would expect, for essentially salt-free crystals, that the volume of the unit cell not occupied by protein would contain solvent of density close to 1.0 g/ml. If it can be assumed that the partial specific volume can be used to estimate the volume occupied by a certain mass of crystalline protein, the solvent densities of hen's-egg-white lysozyme (Imoto *et al.*, 1972) and horse methaemoglobin (Perutz, 1946) are calculated to be 1.025 and 1.006 g/ml respectively. However, the corresponding calculation for the Fv fragment, using a partial specific volume of 0.723 ml/g (Hochman *et al.*, 1973), gives a solvent density of 1.15 ± 0.05 g/ml.

One explanation might be that polymers of ethylene glycol very much smaller than the average molecular weight of 6000 can enter the crystal lattice and perhaps cross-link the protein molecules. Alternatively, if one assumes that no poly(ethylene glycol) can enter the crystals and that the solvent within the crystal has a density close to 1.0 g/ml, the high overall density could be explained by a coating of poly(ethylene glycol) around the crystal. The thickness of solid poly(ethylene glycol) covering an average-sized crystal required by this explanation would be of the order of 10^{-3} mm. Either of these alternatives could explain the observation that, after a period of weeks, the crystals did not dissolve in poly(ethylene glycol)-free buffer solutions, or even in pure water.

We are indebted to the Medical Research Council for supporting this work. D. R. R. holds an Usher Cunningham Studentship of Exeter College and B. J. S. an E. P. Abraham Cephalosporin Studentship of Linacre College. This is a contribution from the Oxford Enzyme Group.

References

- Dugan, E. S., Bradshaw, R. A., Simms, E. S. & Eisen, H. N. (1973) *Biochemistry* **12**, 5400–5416

- Dwek, R. A., Wain-Hobson, S., Dower, S., Gettins, P., Sutton, B., Perkins, S. J. & Givol, D. (1977) *Nature (London)* **266**, 31-37
- Eisen, H. N., Simms, E. S. & Potter, M. (1968) *Biochemistry* **7**, 4126-4134
- Francis, S. H., Leslie, R. G. Q., Hood, L. & Eisen, H. N. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1123-1127
- Goetzl, E. J. & Metzger, H. (1970) *Biochemistry* **9**, 1267-1278
- Hochman, J., Inbar, D. & Givol, D. (1973) *Biochemistry* **12**, 1130-1135
- Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. (1972) *Enzymes 3rd Ed.* **7**, 665-870
- Inbar, D., Rotman, M. & Givol, D. (1971) *J. Biol. Chem.* **246**, 6272-6275
- Inbar, D., Hochman, J. & Givol, D. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2659-2662
- Kabat, E. A. (1978) *Adv. Protein Chem.* **32**, 1-75
- Kumar, K., Phelps, D. J., Carey, P. R. & Young, N. M. (1978) *Biochem. J.* **175**, 727-735
- Low, B. W. & Richards, F. M. (1952) *J. Am. Chem. Soc.* **74**, 1660-1666
- Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491-497
- McPherson, A. (1976) *J. Biol. Chem.* **251**, 6300-6303
- Padlan, E. A., Davies, D. R., Pecht, I., Givol, D. & Wright, C. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 627-637
- Perutz, M. F. (1946) *Trans. Faraday Soc.* **42B**, 187-195
- Poljak, R. J. (1978) *CRC Crit. Rev. Biochem.* **5**, 45-84
- Sutton, B. J., Gettins, P., Givol, D., Marsh, D., Wain-Hobson, S., Willan, K. J. & Dwek, R. A. (1977) *Biochem. J.* **165**, 177-197
- Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1485-1489

APPENDIX III

Data to 6Å Resolution with Phases

Each F and Δ is followed by a standard deviation.

$$s = 4\sin^2\theta / \lambda^2$$

m = figure of merit

m.p. = most probable

Format:

h k l s F(Nat) F(Pt) $\Delta_{\text{anom}}(\text{Pt})$ F(Ur) $\Delta_{\text{anom}}(\text{Ur})$ $\alpha(\text{best})$ m $\alpha(\text{m.p.})$.

6Å DATA NATIVE + TWO DERIVATIVES PHASED

APPENDIX IV

Data to 4.5Å Resolution with Phases

Each F and Δ is followed by a standard deviation.

$$s = 4\sin^2 \theta / \lambda^2$$

m = figure of merit

m.p. = most probable

Format:

h k l s F(Nat) F(Pt) $\Delta_{\text{anom}}(\text{Pt})$ F(Ur) $\Delta_{\text{anom}}(\text{Ur})$
F(Pt)_{film} F(Ur)_{film} $\alpha(\text{best})$ m $\alpha(\text{m.p.})$.



APPENDIX V

Data to 2.7Å Resolution with Phases

Each F and Δ is followed by a standard deviation.

$$s = 4\sin^2 \theta / \lambda^2$$

m = figure of merit

m.p. = most probable

Format:

h k l s F(Nat) F(Pt) Δ_{anom}(Pt) F(Ur) Δ_{anom}(Ur)
F(Pt)_{film} F(Ur)_{film} α(best) m α(m.p.).



REFERENCES

- Alber,T., Hartman,F.C., Johnson,R.M., Petsko,G.A., Tsernoglou,D. (1981)
J. Biol. Chem. 256, 1356-1361.
- Amzel,L.M. and Poljak,R.J. (1979) Ann. Rev. Biochem. 48, 961-997.
- Arndt,U.W., Champness,J.M., Phizackerley,R.P., Wonacott,A.J. (1973)
J. Appl. Cryst. 6, 457-463.
- Arndt,U.W. and Willis,B.T.M. (1966) 'Single Crystal Diffractometry',
Cambridge University Press.
- Arndt,U.W. and Wonacott,A.J. (1977) 'The Rotation Method in Crystallography',
North-Holland Publishing Co., Amsterdam.
- Aschaffenburg,R., Phillips,D.C., Rose,D.R., Sutton,B.J., Dower,S.K., Dwek,R.A.
(1979) Biochem. J. 181, 497-499.
- Bernard,O. and Gough,N.M. (1980) Proc. Nat. Acad. Sci. 77, 3630-3634.
- Blaser,K. and Eisen,H.N. (1978) Proc. Nat. Acad. Sci. 75, 1495-1499.
- Blow,D.M. and Crick,F.H.C. (1959) Acta Cryst. 12, 794-802.
- Blow,D.M. and Matthews,B.W. (1973) Acta Cryst. A29, 56-62.
- Blundell,T.L. and Johnson,L.N. (1976) 'Protein Crystallography', Academic
Press, London.
- Bothwell,A.L.M., Paskind,M., Schwartz,R.C., Sonenshein,G.E., Gefter,M.L.,
Baltimore,D. (1981) Nature 290 65-67.
- Brack,C., Hiramama,M., Lenhard-Schuller,R., Tonegawa,S. (1978) Cell 15,
1-14.
- Bragg,W.L. (1913) Proc. Camb. Phil. Soc. 17, 43-57.
- Bricogne,G. (1976) Acta Cryst. A32, 832-847.
- Cohen,F.C., Sternberg,M.J.E., Taylor,W.R. (1980) Nature 285, 378-382.
- Cory,S., Jackson,J., Adams,J.M. (1980) Nature 285, 450-456.
- Cotner,T. and Eisen,H.N. (1978) J. Exp. Med. 148, 1388-1399.
- Crowther,R.A. (1971) in 'The Molecular Replacement Method' (M.G. Rossmann,ed.)
Gordon and Breach, New York.
- Davis,M.M., Calame,K., Early,P.W., Livant,D.L., Joho,R., Weissman,I.L.,
Hood,L. (1980) Nature 283, 733-739.
- Dodson,E., Evans,P., French,S. (1975) in 'Anomalous Scattering' (Ramaseshan,
S. and Abrahams,S.C.,eds.), Munksgaard,Ltd., Copenhagen, 423-436.
- Dodson,E. and Vijayan,M. (1971) Acta Cryst. B27, 2402-2411.

- Dower, S.K. (1979) D. Phil. Thesis, Univ. of Oxford.
- Dower, S.K. and Dwek, R.A. (1978) *Biomolecular Structure and Function*, 295-328.
- Dower, S.K. and Dwek, R.A. (1979) *Biochemistry* 18, 3668-3674.
- Dower, S.K., Gettins, P., Jackson, R., Dwek, R.A., Givol, D. (1978) *Biochem. J.* 169, 179-188.
- Dower, S.K., Wain-Hobson, S., Gettins, P., Givol, D., Jackson, W.R.C., Perkins, S.J., Sunderland, C.A., Sutton, B.J., Wright, C.E., Dwek, R.A. (1977) *Biochem. J.* 165, 207-225.
- Dreyer, W.J. and Bennett, J.C. (1965) *Proc. Nat. Acad. Sci.* 54, 864-869.
- Dugan, E.S., Bradshaw, R.A., Simms, E.S., Eisen, H.N. (1973) *Biochemistry* 12, 5400-5416.
- Dunnick, W. (1979) *Nature* 281, 253-254.
- Dwek, R.A., Wain-Hobson, S., Dower, S., Gettins, P., Sutton, B.J., Perkins, S.J. (1977) *Nature* 266, 31-37.
- Early, P., Huang, H., Davis, M., Calame, K., Hood, L. (1980) *Cell* 19, 981-992.
- Edelman, G.M. and Gally, J.A. (1964) *Proc. Nat. Acad. Sci.* 51, 846-853.
- Edmundson, A.B., Ely, K.R., Abola, E.E. (1978) *Contemp. Topics in Molec. Immun.* 7, 95-118.
- Ehrlich, P.H., Matsueda, G.R., Haber, E., Margolies, M.N. (1979) *Biochemistry* 18, 4411-4418.
- Eisen, H.N., Simms, E.S., Potter, M. (1968) *Biochemistry* 7, 4126-4134.
- Fox, G.C. and Holmes, K.C. (1966) *Acta Cryst.* 20, 886-891.
- Francis, S.H., Leslie, R.G.Q., Hood, L., Eisen, H.N. (1974) *Proc. Nat. Acad. Sci.* 71, 1123-1127.
- Frangione, B. and Franklin, E.C. (1979) *Nature* 281, 600-602.
- French, S. and Wilson, K.S. (1978) *Acta Cryst.* A34, 517-525.
- Gally, J.A. and Edelman, G.M. (1970) *Nature* 227, 341-348.
- Gettins, P., Dwek, R.A., Perutz, R.N. (1981) *Biochem. J.* 197, 119-125.
- Gettins, P., Givol, D., Dwek, R.A. (1978) *Biochem. J.* 173, 713-722.
- Goetzl, E.J. and Metzger, H. (1970) *Biochemistry* 9, 1267-1278.
- Green, D.W., Ingram, V.M., Perutz, M.F. (1954) *Proc. Roy. Soc.* A225, 287-307.
- Haimovich, J., Eisen, H.N., Hurwitz, E., Givol, D. (1972) *Biochemistry* 11, 2389-2398.

- Hamilton, W.C. (1965) *Acta Cryst.* 18, 502-510.
- Hardy, R.R. and Richards, J.H. (1978) *Biochemistry* 17, 3867-3871.
- Harker, D. (1956) *Acta Cryst.* 9, 1-9.
- Hieter, P.A., Korsmeyer, S.J., Waldmann, F.A., Leder, P. (1981) *Nature* 290, 368-372.
- Hobart, M.J. and McConnell, J. (1978) 'The Immune System', Blackwell Scientific Publications, Oxford.
- Hochman, J., Inbar, D., Givol, D. (1973) *Biochemistry* 12, 1130-1135.
- Howells, E.R., Phillips, D.C., Rogers, D. (1950) *Acta Cryst.* 3, 210-214.
- Hozumi, N. and Tonegawa, S. (1976) *Proc. Nat. Acad. Sci.* 73, 3628-3632.
- Inbar, D., Hochman, J., Givol, D. (1972) *Proc. Nat. Acad. Sci.* 69, 2659-2662.
- Inbar, D., Rotman, M., Givol, D. (1971) *J. Biol. Chem.* 246, 6272-6275.
- Jilka, R.L., and Petska, S. (1979) *J. Biol. Chem.* 254, 9270-9276.
- Kabat, E.A. (1978) *Adv. in Protein Chem.* 32, 1-75.
- Kabat, E.A. and Wu, T.T. (1971) *Ann. N.Y. Acad. Sci.* 190, 382-393.
- Kabat, E.A., Wu, T.T., Bilofsky, H. (1976) 'Variable Regions of Immunoglobulin Chains', Bolt, Beranek and Newman, Cambridge, Mass.
- Kabat, E.A., Wu, T.T., Bilofsky, H. (1978) *Proc. Nat. Acad. Sci.* 75, 2429-2433.
- Kakimoto, K. and Onoue, K. (1974) *J. Immunol.* 112, 1373-1382.
- Kemp, D.J., Cory, S., Adams, J.M. (1979) *Proc. Nat. Acad. Sci.* 76, 4627-4631.
- Klostergaard, J., Krausz, L.M., Grossberg, A.L., Pressman, D. (1977) *Immunochemistry* 14, 107-110.
- Kohler, G. and Milstein, C. (1975) *Nature* 256, 495-497.
- Kumar, K., Phelps, D.J., Carey, P.R., Young, N.M. (1978) *Biochem. J.* 175, 727-735.
- Lachmann, P.J. (1975) in Hobart and McConnell (1975), Chapter 8.
- Lee, J.C. and Lee, L.L.Y (1981) *J. Biol. Chem.* 256, 625-631.
- Low, B. and Richards, F.M. (1952) *J. Am. Chem. Soc.* 74, 1660-1666.
- Matthews, B.W. (1966) *Acta Cryst.* 20, 230-239.
- Matthews, B.W. (1968) *J. Mol. Biol.* 33, 491-497.
- Matthews, B.W. and Czerwinski, E.W. (1975) *Acta Cryst.* A31, 480-487.

- Matthyssens, G. and Rabbitts, T.H. (1980) Proc. Nat. Acad. Sci. 77, 6561-6565.
- Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. 74, 560-564.
- McPherson, A. (1976) J. Biol. Chem. 251, 6300-6303.
- Molgaard, H.V. (1980) Nature 286, 657-659.
- Morris, A.T., Givol, D., Dwek, R.A. (1978) Immunochemistry 15, 519-522.
- Mukkur, T.K.S. (1980) T.I.B.S. 5, 72-74.
- Nisonoff, A., Hopper, J.E., Spring, S.B. (1975) 'The Antibody Molecule', Academic Press, New York.
- North, A.C.T. (1965) Acta Cryst. 18, 212-216.
- North, A.C.T. and Phillips, D.C. (1968) Prog. Biophys. 1, 1-132.
- North, A.C.T., Phillips, D.C., Mathews, F.S. (1968) Acta Cryst. A24, 351-359.
- Nossal, G.J.V., Szenberg, A., Ada, G.L., Austin, C.M. (1964) J. Exp. Med. 119, 484-502.
- Novotny, J. (1973) J. Theor. Biol. 41, 171-180.
- Novotny, J. (1975) J. Theor. Biol. 53, 67-82.
- Ogston, A.G. (1958) Faraday Soc. Trans. 54, 1754-1757.
- Padlan, E.A., Davies, D.R., Pecht, I., Givol, D., Wright, C. (1976) Cold Spring Harbor Symp. Quant. Biol. 41, 627-637.
- Patterson, A.L. (1934) Phys. Rev. 46, 372-376.
- Perutz, M.F. (1946) Trans. Faraday Soc. 42B, 187-195.
- Perutz, M.F. (1970) Nature 228, 726-739.
- Petsko, G.A., Phillips, D.C., Williams, R.J.P., Wilson, I.A. (1978) J. Mol. Biol. 120, 345-359.
- Phillips, D.C. (1966) Advances in Structural Research by Diffraction Methods 2, 75-140.
- Pittz, E.P. and Timasheff, S.N. (1978) Biochemistry 17, 615-623.
- Poljak, R.J., Amzel, L.M., Chen, B.L., Phizackerly, R.P., Saul, F. (1974) Proc. Nat. Acad. Sci. 71, 3440-3444.
- Porter, R.R. (1959) Biochem. J. 73, 119-126.
- Potter, M. and Boyer, C.R. (1962) Nature 193, 1086-1087.
- Potter, M. and Lieberman, R. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 187-202.

- Quiococho, F.A. and Richards, F.M. (1964) *Proc. Nat. Acad. Sci.* 52, 833-839.
- Rabbitts, T.H. (1978) *Nature* 275, 291-296.
- Rabbitts, T.H. (1979) *Int. Rev. Biochem.* 23, 49-69.
- Rabbitts, T.H., Forster, A., Dunnick, W., Bentley, D.L. (1980) *Nature* 283, 351-356.
- Rabbitts, T.H., Matthyssens, G., Hamlyn, P.H. (1980) *Nature* 284, 238-243.
- Ramachandran, G.N. and Srinivasan, R. (1970) 'Fourier Methods in Crystallography', John Wiley and Sons, Inc., New York.
- Richards, F.M. (1968) *J. Mol. Biol.* 37, 225-230.
- Rivers, P.S. (1977) D. Phil. Thesis, Univ. of Oxford.
- Roitt, I. (1974) 'Essential Immunology', Blackwell Scientific Publications.
- Rose, N.R. (1981) *Scientific American* 244, 80-103.
- Rose, N.R., Milgrom, F., van Oss, C.J. (1979) 'Principles of Immunology', Macmillan Publishing Co., Inc., New York.
- Roseblatt, M.S. and Haber, E. (1978) *Biochemistry* 17, 3877-3882.
- Rossmann, M.G. and Blow, D.M. (1962) *Acta Cryst.* 15, 24-31.
- Sakano, H., Huppi, K., Heinrich, G., Tonegawa, S. (1979) *Nature* 280, 288-294.
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., Tonegawa, S. (1980) *Nature* 286, 676-683.
- Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., Tonegawa, S. (1979) *Nature* 277, 627-633.
- Sanger, F. and Coulson, A.R. (1975) *J. Mol. Biol.* 94, 441-448.
- Sanger, F. and Coulson, A.R. (1978) *FEBS Letters* 87, 107-110.
- Schiffer, M., Girling, R.L., Ely, K.R., Edmundson, A.B. (1973) *Biochemistry* 12, 4620-4631.
- Schilling, J., Clevinger, B., Davie, J.M., Hood, L. (1980) *Nature* 283, 35-40.
- Schulenberg, E.P., Simms, E.S., Lynch, R.G., Bradshaw, R.A., Eisen, H.N. (1971) *Proc. Nat. Acad. Sci.* 68, 2623-2626.
- Segal, D.M., Padlan, E.A., Cohen, G.H., Rudikoff, S., Potter, M., Davies, D.R. (1974) *Proc. Nat. Acad. Sci.* 71, 4298-4302.
- Seidman, J.G. and Leder, P. (1978) *Nature* 276, 790-795.
- Seidman, J.G., Leder, A., Edgell, M.H., Polsky, F., Tilghman, S.M., Tiemeier, D.C., Leder, P. (1978) *Proc. Nat. Acad. Sci.* 75, 3881-3885.

- Seidman, J.G., Max, E.E., Leder, P. (1979) *Nature* 280, 370-375.
- Sharon, J. and Givol, D. (1976) *Biochemistry* 15, 1591-1594.
- Sherwood, D. (1976) 'Crystals, X-rays and Proteins', Longman, London.
- Shimizu, A., Takahashi, N., Yamawaki-Kataoka, Y., Nishida, Y., Kataoka, T., Honjo, T. (1981) *Nature* 289, 149-153.
- Sledge, C., Fair, D.S., Black, B., Krueger, R.G., Hood, L. (1976) *Proc. Nat. Acad. Sci.* 73, 923-927.
- Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
- Stanford, J.M. and Wu, T.T. (1981) *J. Theor. Biol.* 88, 421-439.
- Stuart, D. and Walker, N. (1979) *Acta Cryst.* A35, 925-933.
- Sutton, B.J. (1980) D. Phil. Thesis, Univ. of Oxford.
- Sutton, B.J., Gettins, P., Givol, D., Marsh, D., Wain-Hobson, S., Willan, K.J., Dwek, R.A. (1977) *Biochem. J.* 165, 177-197.
- Tanford, C. (1961) 'Physical Chemistry of Macromolecules', John Wiley and Sons, Inc., New York.
- Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O., Gilbert, W. (1978) *Proc. Nat. Acad. Sci.* 75, 1485-1489.
- Tsuzukida, Y., Wang, C., Putnam, F.W. (1979) *Proc. Nat. Acad. Sci.* 76, 1104-1108.
- van Oss, C.J. and Grossberg, A. (1979) in Rose *et al.*, (1979), Chapter 5.
- Wabl, M.R., Johnson, J.P., Haas, I.G., Tenkhoff, M., Meo, T., Inan, R. (1980) *Proc. Nat. Acad. Sci.* 77, 6793-6796.
- Wain-Hobson, S., Dower, S.K., Gettins, P., Givol, D., McLaughlin, A.C., Pecht, I., Sunderland, C.A., Dwek, R.A. (1977) *Biochem. J.* 165, 227-235.
- Wang, A.C., Wilson, S.K., Hopper, J.E., Fudenberg, H.H., Nisonoff, A. (1970) *Proc. Nat. Acad. Sci.* 66, 337-343.
- Watson, H.C., Shotton, D.M., Cox, J.M., Muirhead, H. (1970) *Nature* 225, 806-816.
- Weigert, M., Gatmaitan, L., Loh, E., Schilling, J., Hood, L. (1978) *Nature* 276, 785-790.
- Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J., Hood, L. (1980) *Nature* 283, 497-499.
- Wells, J.V., Fudenberg, H.H., Givol, D. (1973) *Proc. Nat. Acad. Sci.* 70, 1585-1587.
- Willan, K., Marsh, D., Sunderland, C.A., Sutton, B.J., Wain-Hobson, S., Dwek, R.A., Givol, D. (1977) *Biochem. J.* 165, 199-206.

Wilson,A.J.C. (1949) Acta Cryst. 2, 318-321.

Wilson,A.J.C. (1950) Acta Cryst. 3, 397-398.

Wilson,K. and Yeates,D. (1978) Acta Cryst. A35, 146-157.

Wooster,W.A. (1964) Acta Cryst. 17, 878-882.

Wu,T.T. and Kabat,E.A. (1970) J. Exp. Med. 132, 211-249.

Wu,T.T., Kabat,E.A., Bilofsky,H. (1979) Proc. Nat. Acad. Sci. 76, 4617-4621.