



Pilkington Library

Author/Filing Title Brown, K

Accession/Copy No. 040152649

Vol. No. Class Mark

LOAN COPY

0401526496




The Application of Multiple Wavelength Anomalous Dispersion Methods to Protein Crystallography

by
Kieron Brown

A Doctoral Thesis submitted in partial fulfilment of
the requirements for the award of Doctor of Philosophy of
Loughborough University

June 1997

© by Kieron Brown 1997

 Loughborough University Library
Date Feb 98
Class
Acc No 040152649

99102676

Abstract

In order to collect crystallographic data rapidly and efficiently from proteins, an off-line image plate detector was designed and installed on the beamline BM14, at the European Synchrotron Radiation Facility. The principal component of the detector was a rotating support frame onto which six image plates could be mounted. A barcode system identified each plate and differences in plate orientation were corrected by fiducial spots recorded using a xenon arc-lamp with fibre optic light guides. A suite of computer software was written to control the detector and manage all aspects of the data collection.

As a test of the detector, a multiple wavelength anomalous dispersion experiment was performed, on the ribonucleotide reductase free radical protein R2 (RNR). The structure of this 43.5kDa protein was successfully solved using the anomalous scattering from mercury atoms. The results from the data analysis are presented, in which the anomalous scattering information was treated in a manner analogous to multiple isomorphous replacement. Based upon these results, various aspects of the detector design are discussed.

In an attempt to gain a greater understanding of the mechanism by which the signal recognition particle (SRP) arrests the translocation of nascent protein molecules by the ribosome, the crystal structure of the 9kDa protein from the SRP (SRP9) was studied by the technique of X-ray diffraction.

Original experiments were performed on crystals of the murine protein, grown at pH 5.0. Attempts to solve the structure by molecular replacement proved unsuccessful. After a comprehensive search for heavy atom derivatives, a selenomethionine incorporated protein was prepared and experiments performed using the techniques of isomorphous replacement and multiple wavelength anomalous dispersion.

The structure of the murine SRP9 protein could not be solved by these techniques. Twinning by merohedry is proposed as the principal explanation for this result. The discussion proposes that the crystals are composed of two domains whose orientations

are related by a 60° rotation. X-ray diffraction is observed from both regions simultaneously, giving two diffraction patterns which superimpose exactly. A program was written to deconvolute the observed intensities, but neither the techniques of molecular replacement nor multi-wavelength anomalous dispersion produced a successful structure solution.

During the final stages of this thesis however, the selenomethionine incorporated human-SRP9 protein was crystallised. Molecular replacement techniques were not successful so multi-wavelength anomalous dispersion methods were used and the structure solved. Initial views of the structure are presented, and tentative comparisons made with the homologous structure of the heterodimer SRP9/14 which show that the SRP9 molecule undergoes two principal conformational changes; the formation of an N-terminal β -strand and a change in the position of the C-terminal α -helix.

The discussion suggests that these differences are significant enough to explain why the SRP9 structure could not be solved by molecular replacement techniques.

Future work will address the effect of pH on the SRP9 conformation, using a new crystal form, grown under different conditions at pH 8.0. In addition, attempts will be made to solve murine SRP9 using the de-twinned data to ascertain whether the molecular conformation is conserved between the murine and human SRP9 structures, as well as confirming the validity of the de-twinning procedure.

Acknowledgements

I would like to express my sincere thanks to my supervisors, Dr. Clive Wilkinson and Prof. Kurt Ziebeck, for their supervision and guidance during this thesis, and for giving me the opportunity to work in Grenoble, at the EMBL and the ESRF.

I would also like to thank Dr. Stephen Cusack, for allowing me to work in his group, and on the SRP9 project.

Seán McSweeney, Anders Åberg, Gordon Leonard and Elspeth Gordon, for stimulating discussions, and being willing to spend numerous sleepless nights, helping me perform experiments on the beamline. Thanks particularly to Gordon and Seán, without whose help and advice, the human SRP9 would not have been finished in time to appear in this thesis.

Andy Thompson, who over the last four years has trusted me to do over fifty-two data collections on BM14, and who wisely didn't trust me when I foolishly said, "It's the last one."

Ulrike Kapp for preparing and crystallising the SRP9 proteins, and sharing the numerous ups-and-downs, as the SRP9 project leapt between periods of excitement and despair. I'll be prepared, next time you say "I've got some nice looking crystals".

Franck Felisaz, whose ability and foresight, ensured that the fast image plate detector worked so efficiently.

Dr. Par Nordlund, for allowing me to contribute to the RNR project and for always being prepared to supply crystals, often at only twenty-four hours notice!

Darcy Birse for sharing numerous occasions when crystals exploded in cryo-solutions. And finally, Carmen Berthet for her support and encouragement, which persuaded me not to abandon the SRP9 project.



Symbols used in the text

A	the amplitude of displacement of a damped harmonic oscillator
A	the absorption factor
a	distance between two atoms in a linear array
a, b, c	elementary translations
a^*, b^*, c^*	basis vectors of the reciprocal lattice
$a(\lambda), b(\lambda), c(\lambda)$	simple scattering factor ratios
B, B_j	the temperature factor
$C(\theta)$	function of polarisation
c	the speed of light
d	the distance between atomic planes
$\delta(r^2)$	squared mean displacement of an atom from equilibrium
ε	the lack of closure error
ε_o	permittivity of medium
E	root mean square error associated with a measurement
E	the extinction factor
E_o	amplitude of incoming electromagnetic wave
e	charge of electron
F_{hkl}	the structure factor
F_{ph}	the scattering factor of the structure containing the native protein with the addition of heavy atoms
F_p	the native protein structure factor
F_h	the heavy atom structure factor
$ F_{ph} $	the amplitude of the structure factor for reflection hkl , with no phase implied
F_p, F_{ph}	the structure factors for the reflection hkl for the protein (p), heavy atom derivative (ph)
F^+, F^-	the value of $F(hkl)$ and $F(\overline{hkl})$ in the presence of anomalous scattering
F', F''	real and imaginary parts of the anomalous structure factor

$\Delta_{ano}, \Delta_{disp}$	anomalous and dispersive differences
$ \lambda F_{obs}^{\pm} $	the wavelength dependent observed structure factor for (hkl) and $(\bar{h}\bar{k}\bar{l})$ with no phase implied
F_T^o	the wavelength independent structure factor for all atoms in the absence of anomalous scattering
F_A^o	the wavelength independent structure factor for the normal scattering component of the anomalous scatterers
\bar{F}_{obs}	the mean Bijvoet difference
$f_j, f(\underline{k})$	atomic scattering factor
f_o	the normal scattering factor
f'	the real component of the anomalous scattering factor
f''	the imaginary part of the anomalous scattering factor
g	damping constant
g	gain of a detector
h, k, l	integers
I_P	intensity at the observation point P
I_{hkl}	intensity of the reflection with Millar indices hkl
I_{pixel}	the intensity in one pixel
K_z, K_y	polarisation constants
\underline{k}	conjugate parameter of \underline{r}
$\underline{\kappa}$	scattering vector
k_r	the restoring force
L	the Lorentz factor
λ	wavelength of radiation
μ	the linear absorption coefficient
m	mass of electron
M_W	the molecular weight
N	an integer number
N_f	a normalisation factor
N_e	the number of electrons per unit volume
N_A	the number of anomalous scatterers

n	an integer
ω	angular frequency of electromagnetic radiation
P	the polarisation factor
$P(uvw)$	the Patterson function
ψ	angle between direction of acceleration of an electron, and the direction of observation
ϕ	phase angle
ϕ_h	the phase of the heavy atom partial structure
ϕ_p	the phase of the native protein
ϕ_T^o, ϕ_A^o	the phase associated with F_T^o and F_A^o , respectively
$[R]$	a rotation matrix
R_{sym}, R_{fac}	reliability factors
$^{ano}R_{Cullis}$	the anomalous Cullis R factor
$\rho(\underline{r})$	electron density
\underline{r}	a vector between two points
σ	the error associated with a measurement
$\hat{\underline{S}}_o$	unit vector of incident wave
$\hat{\underline{S}}$	unit vector of scattered wave
\underline{t}	a translation vector
2θ	scattering angle
u, v, w	vectors separating atoms in Patterson space
V	the volume of the crystal
V_c	volume of the unit cell
X	magnitude of F_{hkl} projected along the real axis
X_1, X_2	set of vectors representing the atoms in object (1) or (2)
x, y, z	Cartesian coordinate axes
x	displacement of electron along x -axis
\ddot{x}	acceleration in the x -direction
Y	magnitude of F_{hkl} projected along the imaginary axis
Z	atomic number

Contents

1	Introduction	1
2	Principles of X-ray Diffraction	5
2.1	Thomson Scattering	5
2.2	Scattering by Atoms	7
2.3	Bragg's Law and the Ewald Sphere	8
2.4	Structure Factor	11
3	The Phase Problem	16
3.1	Direct Methods	17
3.2	Molecular Replacement	18
3.3	Isomorphous Replacement	18
3.4	Multiple Isomorphous Replacement	20
3.5	Anomalous Scattering	24
3.6	Single Isomorphous Replacement with Anomalous Scattering	28
3.7	Multiwavelength Anomalous Dispersion	32

4	Detectors for X-Ray Crystallography	38
4.1	Detector Types	39
4.1.1	Characteristics of Detector Systems	40
4.1.2	Film	41
4.1.3	Image Plates	42
4.1.4	Charged-coupled Device	43
4.1.5	Multi-wire Proportional Counter	43
4.2	Requirements for Protein Crystallography	44
4.3	The Fast Image Plate Changer	46
5	Performance of the Fast Image Plate Detector	63
5.1	Background to the Experiment	63
5.2	Experiment and Results	66
5.3	Discussion	75
6	Crystal Structure of SRP9	80
6.1	Biological Significance	80
6.2	Experiment and Results	83
6.3	Discussion	116
6.4	Structure of Human SRP9	118
6.5	Discussion	132
7	Conclusion	133

A	Software to Analyse murine-SRP9 Data	135
A.1	Program getSign	135
A.2	Program getAlpha	138
A.3	Program getRatio	140
A.4	Program deTwin	142
B	The Fast Image Plate Detector Software	151
B.1	Program carousel	151
B.2	Program carousel_menu	155
B.3	Program washing	162
B.4	Program DataBase	165
B.5	Program msqL_strings.h	169
B.6	Program experiment_table_viewer.phtml	169

Chapter 1

Introduction

X-ray crystallography provides a means of determining the structure of macromolecules such as proteins, nucleic acids and carbohydrates. As with all methods of crystallography, the phase problem must be overcome if a structure is to be solved. In macromolecular crystallography, this is achieved by three principal methods; molecular replacement (MR), multiple isomorphous replacement (MIR) and multiwavelength anomalous dispersion (MAD). The technique of molecular replacement was developed by Rossmann and Blow [1], and draws on the fact that different proteins can have structures that resemble one another. The technique is therefore restricted to proteins for which a homologous structure is already known. In the method of multiple isomorphous replacement heavy atoms are attached to the original, native, protein molecule, to give so-called derivatives. The differences between the diffraction patterns from the native and derivative molecules, are then exploited. Unfortunately, in addition to requiring several derivatives, the technique is dependent upon the conformation of the molecule in the crystals being close to identical. Multiwavelength anomalous dispersion techniques require only one heavy metal derivative and have the advantage that the phase problem can be solved without reference to the native structure factors.

Bijvoet [2] was the first to suggest that the phenomenon of anomalous scattering could be used to solve the phase problem. In 1975 Hoppe and Jakubowski [3] performed an experiment using a conventional X-ray source, where they collected data at two different wavelengths near the iron K-absorption edge. The first multi-wavelength experiment performed using synchrotron radiation was that of Phillips et al. [4], in

which data from many wavelengths near the iron-edge were collected.

An algebraic formulation of the wavelength dependence of the observed structure factor was first proposed by Karle [5] [6], which was later modified and implemented by Hendrickson [7] [8]. Since then, the technique of using the anomalous scattering at multiple wavelengths has been considered in an approach analogous to MIR [9] [10] [11].

The first MAD experiments [12] [13] [14] demonstrated that the success of the technique was dependent upon measuring very accurately, the small differences in intensity arising from anomalous signals. This placed well-defined requirements on the experimental instrumentation, in particular, the X-ray source and the detector.

The need for tunable sources of X-rays means that experiments are restricted to synchrotrons. In order that the anomalous signal can be fully exploited, the wavelength needs to be selected with great accuracy and remain stable during the entire length of the experiment. With regard to the detector, area detectors are most well-suited, because of their ability to measure many diffraction spots simultaneously. However, if these detectors are to be useful for MAD data collection, the observed intensities should be determined reliably, thus imposing the need for uniform spatial response across the entire active surface, as well as a uniform energy response throughout the energy range of interest. In addition, the detector should also be able to collect data rapidly, since biological samples are sensitive to X-rays and can quickly deteriorate during the course of the experiment.

Two of the original experiments [13] [14] used multi-wire proportional detectors (MWPC) [15] and film [16], which fulfilled many of the requirements, including rapidity, uniformity of spatial response and sensitivity to X-rays photons. However, these detectors suffered from the drawback of having limited spatial resolution. Image plates have also been used as detectors [17]. This type of detector was seen to offer many advantages over both MWPC and film, including good spatial resolution, uniformity of spatial response and a large dynamic range. In addition, they avoided the inconvenient processing procedure required with film.

The intense sources of X-ray radiation, found at third generation synchrotrons, increase the demands placed on the detecting system. The advantages offered by the image plate medium mean that it is still well suited to crystallographic experiments. However, a significant drawback has begun to be recognised, and that is of the instrumental dead-time.

The first part of this thesis presents the work undertaken to build a detector for use in protein crystallography, with particular application to multiwavelength anomalous dispersion experiments.

In chapter 2, the background to the fundamentals of X-ray theory is presented, including Bragg's law, the Ewald construction, the structure factor and Friedel's law.

In chapter 3, the phase problem of crystallography is described, as well as a more detailed presentation of the principal techniques used to overcome it, including molecular replacement, multiple isomorphous replacement and multiwavelength anomalous scattering.

Chapter 4 describes the fast image plate changer, a detector designed and installed on the beam-line BM14 at the European Synchrotron Radiation Facility. This detector uses image plates, and was designed for the collection of diffraction data from MAD experiments, from a third generation synchrotron. By basing the design on an off-line scanning principle, it was hoped to show that an image plate medium offered a reliable and time efficient detector for collecting MAD data.

As a test of the fast image plate detector, a multiple wavelength anomalous dispersion experiment was performed, on the ribonucleotide reductase free radical protein R2 (RNR). The structure of this 43.5kDa protein was successfully solved treating the anomalous scattering from mercury atoms in a manner analogous to multiple isomorphous replacement. The experiment and results are presented in chapter 5. Based on the results, various aspects of the detector design are evaluated.

Chapter 6 presents the attempts to solve the crystal structure of the 9 kDa protein of the signal recognition particle (SRP9). X-ray crystallographic experiments

were performed on the native and selenomethionine-incorporated proteins, after an exhaustive search for heavy atom derivatives proved unsuccessful. The techniques of MR, MIR and MAD were then employed, but without success, a result attributed to the phenomenon of twinning by merohedry. During the final stages of this thesis, however, a MAD experiment was performed on the selenomethionine-incorporated human SRP9 protein, from which the crystal structure was obtained.

Chapter 2

This chapter describes the *fundamentals* of X-ray scattering theory; *Bragg's law*, the *Ewald construction*, the *structure factor* and *Friedel's law*.

Principles of X-ray Diffraction

2.1 Thomson Scattering

Due to the charge and mass of an electron, its interaction with an X-ray photon produces readily observable effects. Thomson was the first to consider the effect of an electromagnetic wave incident upon an isolated electron [18].

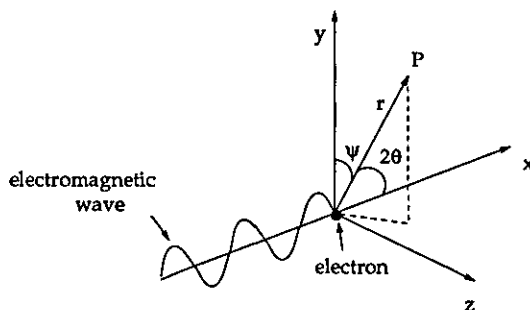


Figure 2.1: Plane-polarised electromagnetic wave incident on an electron, centered on the origin of a Cartesian coordinate system

Figure 2.1 shows a plane-polarised X-ray wave moving along the x-axis with its electric vector in the y direction, incident upon an electron situated at the origin. ψ is the angle between the direction of acceleration and the direction of observation, 2θ

the angle between the direction of the incident wave and the direction of observation, P the observation point at a distance r from the origin.

The phase of the plane wave is given by

$$\phi(x, t) = (kx - \omega t) \quad (2.1)$$

where

$$k = \frac{2\pi}{\lambda}$$

and ω is the angular frequency. The wave can then be described in complex form as in equation (2.2)

$$E_o(x, t) = E_o e^{i(kx - \omega t)} \quad (2.2)$$

The electric field exerts a force on the electron, which according to classical theory, is driven to perform simple harmonic motion [19, p. 142]. The instantaneous acceleration is obtained as in equation (2.3)

$$a(t) = \frac{q}{m} E_o(0, t) = -\left(\frac{e}{m}\right) E_o e^{-i\omega t} \quad (2.3)$$

where $q = -e$ is the charge of the electron, and m the electron mass. The displacement of the electron, x , can then be obtained

$$x = -\frac{E_o e}{\omega^2 m} e^{i\omega t} = \frac{E_o e}{\omega^2 m} e^{i(\omega t + \pi)} \quad (2.4)$$

As it oscillates, the electron radiates electromagnetic radiation [20]. In c.g.s units, the far-field radiation at a point P, distance r from the electron, has the electric field strength $E_{rad}(r, t)$

$$\begin{aligned} E_{rad}(r, t) &= -\frac{-ea(t - r/c)}{rc^2} \sin\psi \\ &= -E_o \frac{e^2}{mc^2 e^{-i\omega(t-r/c)}} \sin\psi \end{aligned} \quad (2.5)$$

It can be seen that the incident and scattered wave have the same angular frequency, but that there is a phase shift of π radians, as indicated by the negative sign.

Polarisation of the incident radiation is taken into account by decomposing the electric vector in two mutually perpendicular directions; along the y and z axes [19]. The amplitude of the electric vector at point P thus becomes

$$E_{rad} = -E_o \frac{e^2}{mc^2 r} (K_z + K_y \cos 2\theta) \quad (2.6)$$

where K_z and K_y are simple constants. When the radiation is completely unpolarised, $K_z = K_y = \frac{1}{2}$ and

$$E_{rad} = -E_o \frac{e^2}{mc^2 r} \left(\frac{1 + \cos 2\theta}{2} \right) \quad (2.7)$$

The polarisation state of the radiated wave is then unpolarised at $\theta = 0^\circ$ and 180° and plane-polarised at $\theta = 90^\circ$.

In a vacuum, the intensity is equal to the square of the amplitude [21], then equation (2.7) can be rewritten to give the resultant intensity at the observation point P

$$I_{rad} = I_o C(\theta) \frac{e^4}{m^2 c^4 r^2} \left(\frac{1 + \cos^2 2\theta}{2} \right) \quad (2.8)$$

where $C(\theta)$ is the polarisation function, dependent upon θ .

2.2 Scattering by Atoms

Within the atom, electrons are distributed over a finite volume. If the electron density is $\rho(\underline{r})$, then when integrated over the whole volume occupied by the electrons, the atomic number, Z , is obtained, as in equation (2.9). In addition, the Fourier transform of $\rho(\underline{r})$ gives the atomic scattering factor, $f(\underline{k})$, in equation (2.10) [22], where \underline{k} is the conjugate parameter to \underline{r} .

$$Z = \int \rho(\underline{r}) d^3r \quad (2.9)$$

$$f(\underline{k}) = \int \rho(\underline{r}) e^{i\underline{k}\cdot\underline{r}} d^3r \quad (2.10)$$

As a first approximation, the distribution of the electrons can be assumed to be spherically symmetric. In this case, equation (2.10) reduces to a one-dimensional integral as in equation (2.11)

$$f(k) = \int_0^\infty 4\pi r^2 \rho(r) \frac{\sin kr}{kr} dr \quad (2.11)$$

Here, $4\pi r^2 \rho(r)$ is the radial distribution function of the atom. Figure 2.2 shows the scattering factor for carbon, where it can be seen that the maximum value taken by $f(k)$ is the atomic number, Z . In addition, figure 2.2 shows that $f(k)$ decreases with increasing $\sin\theta$.

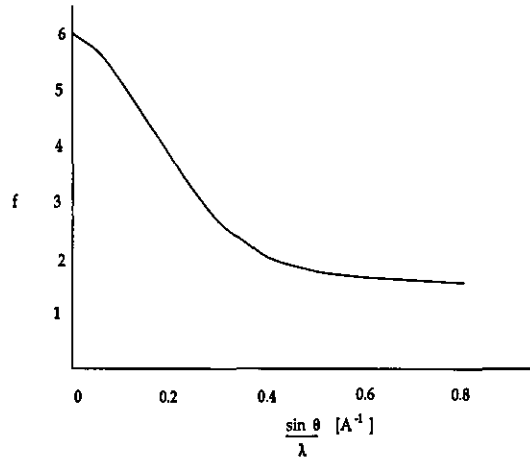


Figure 2.2: Atomic scattering factor for carbon, plotted against scattering angle, $\frac{\sin\theta}{\lambda}$

2.3 Bragg's Law and the Ewald Sphere

Figure 2.3 illustrates a periodic array of atoms, illuminated by a plane wave, with wavelength λ . Atoms A and B will be sources of secondary emission. The two waves

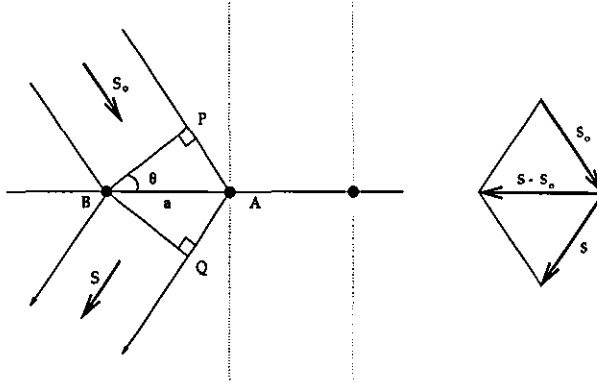


Figure 2.3: A 1-dimensional periodic array of atoms, separated by distance a

emitted will interfere constructively whenever the path difference between them is equal to an integer number of wavelengths

$$PA + AQ = h\lambda \quad (2.12)$$

where h is an integer. Bragg expressed this result as a geometric equation, equation (2.13) [23].

$$2d\sin\theta = n\lambda \quad (2.13)$$

where a and h have been rewritten as d and n , respectively, and θ is the scattering angle.

Equation (2.12) can be written in vector form as

$$\begin{aligned} -\underline{a} \cdot \hat{\underline{S}}_0 + \underline{a} \cdot \hat{\underline{S}} &= h\lambda \\ \underline{a} \cdot (\hat{\underline{S}} - \hat{\underline{S}}_0) &= h\lambda \end{aligned} \quad (2.14)$$

where $\hat{\underline{S}}_0$ is the unit vector of the incident wave, and $\hat{\underline{S}}$ the unit vector of the scattered wave. By defining the scattering vector as $\underline{\kappa}$

$$\underline{\kappa} = \frac{1}{\lambda} (\hat{\underline{S}} - \hat{\underline{S}}_0) \quad (2.15)$$

then

$$\frac{\underline{a}}{h} \cdot \underline{\kappa} = 1 \quad (2.16)$$

Extending this result to a 3-dimensional periodic array with elementary translations \underline{a} , \underline{b} and \underline{c} , leads to the Laue equations [24].

$$\frac{\underline{a}}{h} \cdot \underline{\kappa} = 1 \quad \frac{\underline{b}}{k} \cdot \underline{\kappa} = 1 \quad \frac{\underline{c}}{l} \cdot \underline{\kappa} = 1 \quad (2.17)$$

These equations are satisfied if the scattering vector takes the form

$$\underline{\kappa} = h\underline{a}^* + k\underline{b}^* + l\underline{c}^* \quad (2.18)$$

where h , k and l are integers and \underline{a}^* , \underline{b}^* and \underline{c}^* satisfy the following conditions;

$$\begin{aligned} \underline{a}^* \cdot \underline{a} &= 1 & \underline{a}^* \cdot \underline{b} &= 0 & \underline{a}^* \cdot \underline{c} &= 0 \\ \underline{b}^* \cdot \underline{a} &= 0 & \underline{b}^* \cdot \underline{b} &= 1 & \underline{b}^* \cdot \underline{c} &= 0 \\ \underline{c}^* \cdot \underline{a} &= 0 & \underline{c}^* \cdot \underline{b} &= 0 & \underline{c}^* \cdot \underline{c} &= 1 \end{aligned} \quad (2.19)$$

this will be the case when

$$\underline{a}^* = \frac{(\underline{b} \times \underline{c})}{V_c} \quad \underline{b}^* = \frac{(\underline{c} \times \underline{a})}{V_c} \quad \underline{c}^* = \frac{(\underline{a} \times \underline{b})}{V_c} \quad (2.20)$$

where V_c is the volume of the unit cell, given by

$$V_c = \underline{a} \cdot (\underline{b} \times \underline{c}) \quad (2.21)$$

\underline{a}^* , \underline{b}^* and \underline{c}^* are the basis vectors of the reciprocal lattice, which can be used in the Ewald sphere construction, shown in figure 2.4.

Ewald showed [25] that there will be a diffracted beam if the Ewald sphere passes through both the origin and a reciprocal lattice point. Hence, when elastic diffraction

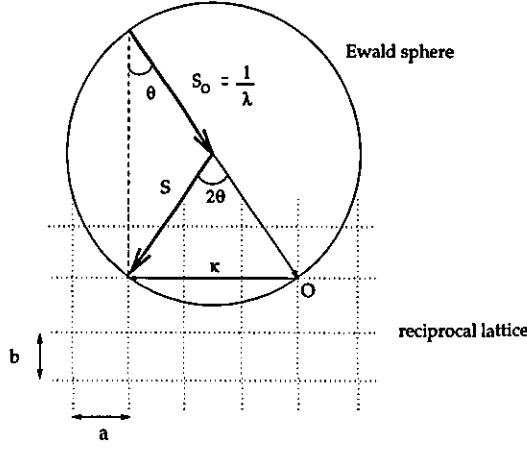


Figure 2.4: Ewald sphere in reciprocal space.

occurs, the change in wavevector between incoming and outgoing waves is equal to the reciprocal lattice vector. Figure 2.4 can be formalised into equation (2.22).

$$|\underline{\kappa}| = \frac{2}{\lambda} \sin\theta = \frac{1}{d} \quad (2.22)$$

where the right-hand side can be re-arranged to give Bragg's equation, equation (2.13).

$\underline{\kappa}$ can now be called the reciprocal lattice vector, which has two important properties; firstly, that it is perpendicular to the real space lattice planes with Miller indices (hkl) , and secondly, that its modulus, $|\underline{\kappa}|$, is the reciprocal of the distance between successive (hkl) lattice planes, $|\underline{\kappa}| = \frac{1}{d_{hkl}}$.

2.4 Structure Factor

A crystal is a 3-dimensional array of atoms, the coordinates of each atom can be described by \underline{r} where

$$\underline{r} = x\underline{a} + y\underline{b} + z\underline{c} \quad (2.23)$$

in which \underline{a} , \underline{b} and \underline{c} are the basis vectors of the lattice, and x , y and z are fractional coordinates.

The scattering of X-rays by crystals can be described either by the kinematical or the dynamical approximation [26]. The kinematic approximation is a simplistic treatment, since it assumes that the crystal specimen is small and that the amplitude of the incoming X-ray wave is the same at all points. The dynamical theory offers a more rigorous treatment, since it takes into account multiple diffraction, in which a diffracted wave is allowed to re-interfere with an incident wave prior to emerging from the specimen. It is however only relevant to diffraction from perfect crystals, in which the 3-dimensional atomic array suffers no distortions or imperfections. For the purposes of the present discussion, aimed at describing the diffraction of X-rays from non-perfect protein crystals, the assumptions made in the kinematical approximation are acceptable.

The scattering from any one atom, A , with respect to an electron at the origin will be

$$f_A \exp(2\pi i \underline{r}_A \cdot \underline{\kappa}) \quad (2.24)$$

Therefore, in the kinematical approximation, where the amplitude of the incident wave is the same at all points within the lattice, the amplitude of the resultant wave is simply the sum of the amplitudes from each individual atom, equation (2.25).

$$\begin{aligned} F_{hkl} &= \sum_{j=1}^N f_j \exp(2\pi i \underline{r}_j \cdot \underline{\kappa}) \\ &= \sum_{j=1}^N f_j \exp[2\pi i (hx + ky + lz)] \end{aligned} \quad (2.25)$$

since

$$\underline{r} \cdot \underline{\kappa} = hx + ky + lz \quad (2.26)$$

where f_j is the atomic scattering factor as given in equation (2.11), $\underline{\kappa}$ the reciprocal lattice vector as given in equation (2.22), and N is the number of atoms within the unit cell.

F_{hkl} is called the structure factor. The diffraction pattern from a crystal lattice is thus represented by a reciprocal lattice, in which each lattice point is identified by a

set of Miller indices and each lattice point is weighted by the corresponding value of the structure factor, F_{hkl} .

The structure factor can be separated into real and imaginary parts and represented on an Argand diagram, see figure 2.5.

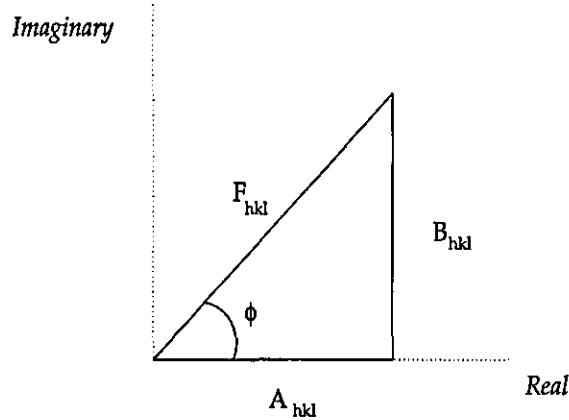


Figure 2.5: The representation of the structure factor on an Argand diagram.

This can be written

$$F_{hkl} = A_{hkl} + B_{hkl} \quad (2.27)$$

where

$$A_{hkl} = \sum_j f_j \cos 2\pi(hx + ky + lz) \quad (2.28)$$

$$B_{hkl} = \sum_j f_j \sin 2\pi(hx + ky + lz) \quad (2.29)$$

and

$$\tan \phi = \frac{B_{hkl}}{A_{hkl}} \quad (2.30)$$

In the absence of anomalous scattering, the weighted reciprocal lattice always has a centre of symmetry, a property referred to as Friedel's law. The centrosymmetric nature of the diffraction pattern is described mathematically as

$$I_{hkl} = I_{-h-k-l} \quad (2.31)$$

From equation (2.25), the structure factor can be written

$$F_{hkl} = \sum_{j=1}^N f_j \exp[2\pi i(hx + ky + lz)] \quad (2.32)$$

$$F_{-h-k-l} = \sum_{j=1}^N f_j \exp[2\pi i(-hx - ky - lz)] \quad (2.33)$$

$$(2.34)$$

Now

$$F_{-h-k-l} = A_{-h-k-l} + iB_{-h-k-l} \quad (2.35)$$

deMoivre's theorem [27] states that

$$e^{\pm i\phi} = \cos\phi \pm i\sin\phi \quad (2.36)$$

from which

$$F_{-h-k-l} = A_{hkl} + iB_{hkl} \quad (2.37)$$

This is shown in figure 2.6

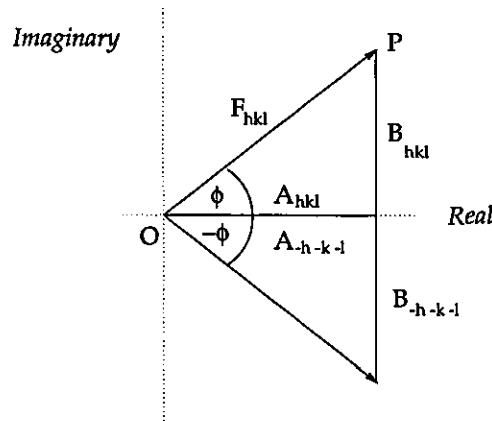


Figure 2.6: The relationship between F_{hkl} and F_{-h-k-l} in the Argand representation of Friedel's law.

Since the intensity is proportional to the square of the structure factor amplitude [28], then

$$I_{hkl} = I_{-h-k-l} \quad (2.38)$$

Equation (2.38) is a mathematical representation of Friedel's law, and shows that reflections related by a centre of symmetry have the same intensity.

The theory presented in this chapter provides a means of putting into context the problem of determining the phase of the resultant scattered radiation. In the next chapter, this problem is considered.

Chapter 3

This chapter presents the *phase problem of crystallography* and describes the principal techniques used in *macro-molecular crystallography* to overcome it; *molecular replacement, isomorphous replacement and anomalous scattering*

The Phase Problem

The diffraction pattern from a crystal can be considered to be the Fourier transform of the electron density distribution corresponding to the real structure [29, p. 112].

The electron density at a position x, y, z is thus given by

$$\rho(x, y, z) = \frac{1}{V_c} \sum_{hkl} F_{hkl} \exp - 2\pi i [hx + ky + lz] \quad (3.1)$$

where V_c is the volume of the unit cell and F_{hkl} is the structure factor for an observation, hkl . In order to calculate $\rho(x, y, z)$ it is therefore necessary to have a knowledge of F_{hkl} . As shown in equation (2.25), F_{hkl} consists of both an amplitude and a phase.

It was shown by Bragg [30] that the formation of an image can be broken down into two stages. Firstly the diffraction stage, and secondly a recombination stage. Recombination of the diffracted radiation produces the required image, but requires a lens. Such a lens does not exist for radiation with a wavelength corresponding to that of X-rays. As a consequence, in crystallographic experiments, the imaging process is truncated at the first stage, where it is only amplitudes that can be measured.

Equation (3.1) cannot therefore be solved since the phase associated with the structure factor is unknown. This is the so-called phase problem of crystallography.

Methods of Overcoming the Phase Problem

There are four principal methods of overcoming the phase problem in protein crystallography; direct methods, molecular replacement, multiple isomorphous replacement and multiple wavelength anomalous dispersion. With the exception of direct methods, each technique uses the function proposed by Patterson [31]. The so-called Patterson function, shown in equation (3.2), is based upon experimentally derivable quantities, and is used to determine the coordinates of atoms within the unit cell.

$$P(uvw) = \frac{1}{V_c} \sum_{hkl} |F_{hkl}|^2 \cos 2\pi (hu + kv + lw) \quad (3.2)$$

V_c is the volume of the unit cell, u , v and w are dummy variables, h , k and l are the Miller indices of the reciprocal lattice point and F_{hkl} the associated structure factor.

The essential feature of the Patterson function is that there are no phase dependent terms, since the structure factor appears as a squared quantity. The function is evaluated at each point on a 3-dimensional grid defined within the unit cell by the values of u , v and w . It is the convolution of the electron density at all points x, y, z within the unit cell, with the electron density found at points $x + u, y + v, z + w$. If any two atoms are separated by the vector (u, v, w) , then there will be a peak in the Patterson map at coordinates (u, v, w) . The height of the peak is directly proportional to the product of the atomic numbers of the atoms and the number of contributing vectors.

3.1 Direct Methods

Direct methods require near atomic resolution data in order to overcome the phase problem. Protein crystals rarely diffract to such high resolution (about 1Å), and

therefore direct methods are not routinely used in macromolecular crystallography.

However, for the methods of isomorphous replacement and anomalous scattering described below, direct methods may be used as a means of solving the heavy atom partial structure.

3.2 Molecular Replacement

Molecular replacement draws on the feature that different proteins can have three-dimensional structures which closely resemble one another. Such a case might arise, for example, if more than half the primary, amino-acid sequence for the two proteins is the same. Molecular replacement can also be applied to a protein which has crystallised in different crystal forms, or for which a structure has been determined by nuclear magnetic resonance.

The technique was pioneered by Rossmann and Blow in 1962 [1] and relies on the fact that two similar objects can be related to each other by rotation and translation operations, with three degrees of freedom in each case. If X_1 is the set of vectors representing the atoms of one object, and X_2 those of the second, then the transformation can be expressed as in equation (3.3) [19, p. 553].

$$\underline{X}_2 = [R]\underline{X}_1 + \underline{t} \quad (3.3)$$

where $[R]$ is the matrix which rotates the coordinates X_1 into the new orientation, and \underline{t} is a translation vector. Once these two functions have been determined, the phases from the known structure can be taken as reliable estimates for those of the unknown structure.

3.3 Isomorphous Replacement

In the method of isomorphous replacement, heavy atoms are attached to the original protein molecule (referred to as the native) to give a so-called derivative. This can

be achieved either at the crystallisation stage, or afterwards, by soaking the native crystal in a solution containing the desired heavy atom. The intensity differences between the native and derivative diffraction patterns can then be used to overcome the phase problem.

For a structure containing many light atoms and a few heavy atoms, the structure factor for a given reflection can be decomposed into two parts;

$$F_{ph} = F_p + F_h \quad (3.4)$$

where F_p is the structure factor for the light atoms, and F_h that for the heavy atoms. This equation can be represented on an Argand diagram, as shown in figure 3.1.

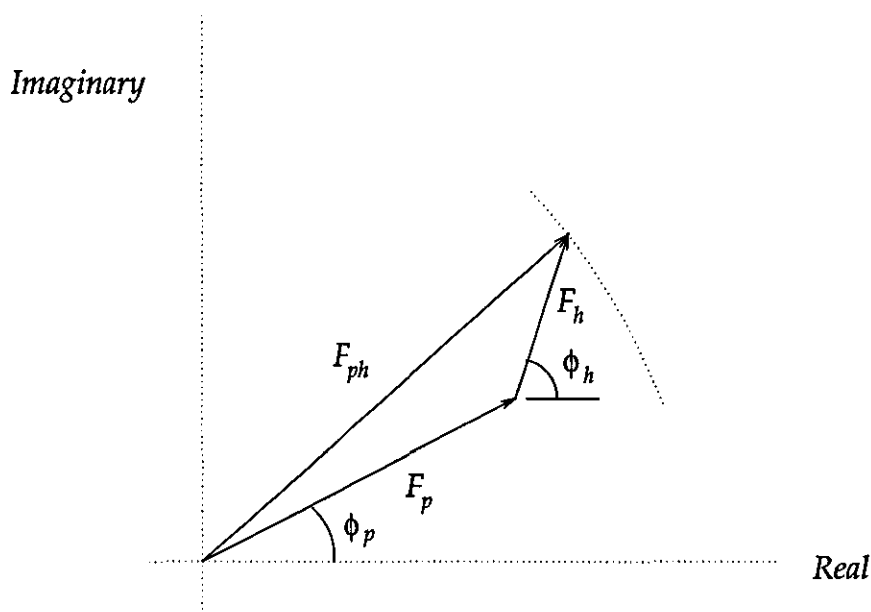


Figure 3.1: The derivative structure factor, F_{ph} , decomposed into the structure factor of the protein molecule, F_p , and that of the heavy atoms, F_h

Here the length of F_h is exaggerated for clarity. The success of the isomorphous replacement method is dependent upon the length of F_h being small relative to F_p . This will be the case when there are many more light atoms than heavy atoms in the unit cell.

In the case of centric reflections, F_{ph} , F_p and F_h are collinear, the phase of F_p is 0 or π [32, p. 237].

Since the magnitude of F_h is small compared with that of F_p , the observed structure factor from a derivative can be approximated to the difference between the structure factor amplitudes of the native and derivative molecules. This is expressed in equation (3.5).

$$|F_h| \approx |F_{ph} - F_p| \quad (3.5)$$

If $|F_{ph} - F_p|$ is then used as the coefficient in a Patterson synthesis, as shown in equation (3.6), then the coordinates of the heavy atoms may be determined.

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F_{ph} - F_p|^2 \cos 2\pi (hu + kv + lw) \quad (3.6)$$

With a knowledge of F_h , the protein phase, ϕ_p , can be found via a graphical method, shown in figure 3.2. AB is the heavy atom vector with known magnitude $|F_h|$ and phase angle ϕ_h , OA the native protein vector, with known magnitude $|F_p|$ derived from experiment, and unknown phase ϕ_p , and OB the derivative vector with known magnitude $|F_{ph}|$, again derived from experiment.

A circle of radius $|F_p|$ is centred on point A, a second of radius $|F_{ph}|$ is centred on point B. The points of intersection of these two circles, O and O', represent two possible solutions for the protein phase, ϕ_p .

3.4 Multiple Isomorphous Replacement

The phase ambiguity presented in the heavy-atom method can be overcome by attaching a second heavy atom of a different species. This technique was originally used by Perutz et al. [33] in the solution of haemoglobin, and is called multiple isomorphous replacement (MIR).

In figure 3.3 the phase triangle from figure 3.2 is shown as triangle OAB. In addition, a second triangle OAC is shown representing the solution from a second derivative. The protein phase is unambiguously given by ϕ_p .

The multiple isomorphous replacement method therefore requires data from the

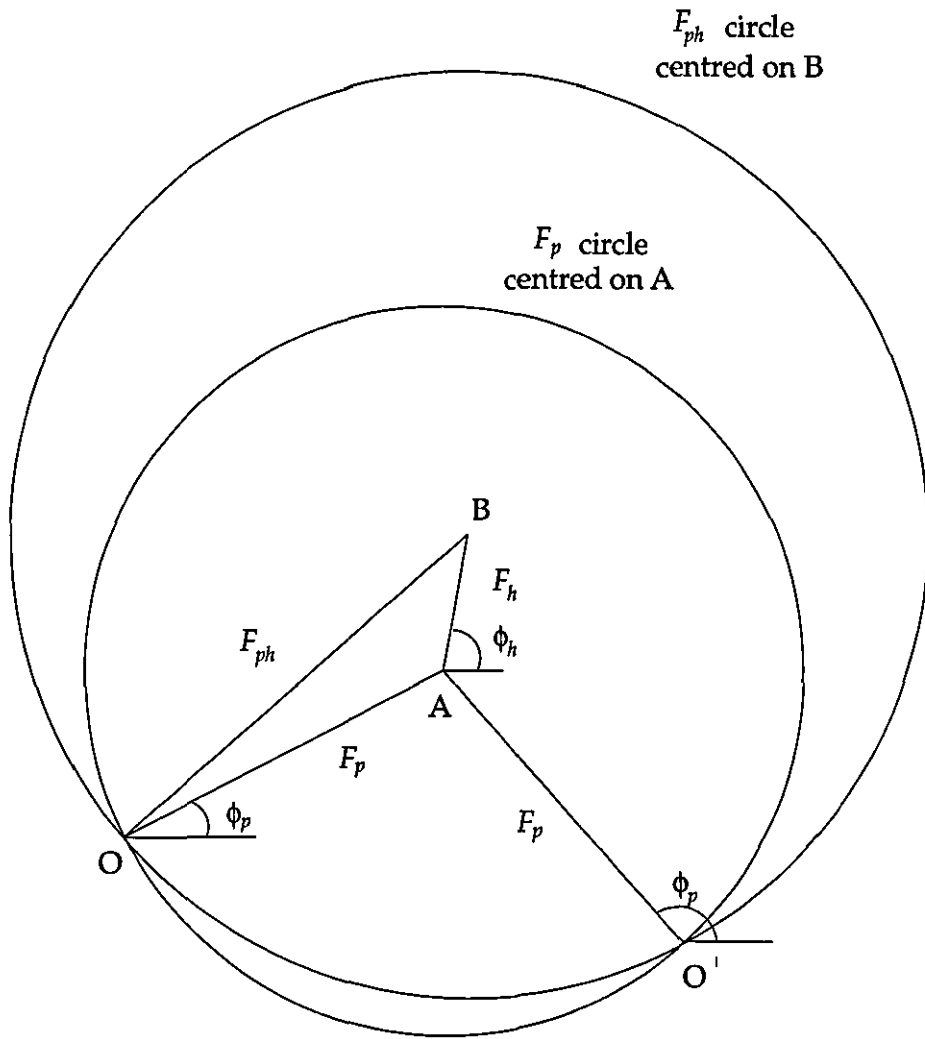


Figure 3.2: Graphical solution for the protein phase using a single derivative, F_{ph}

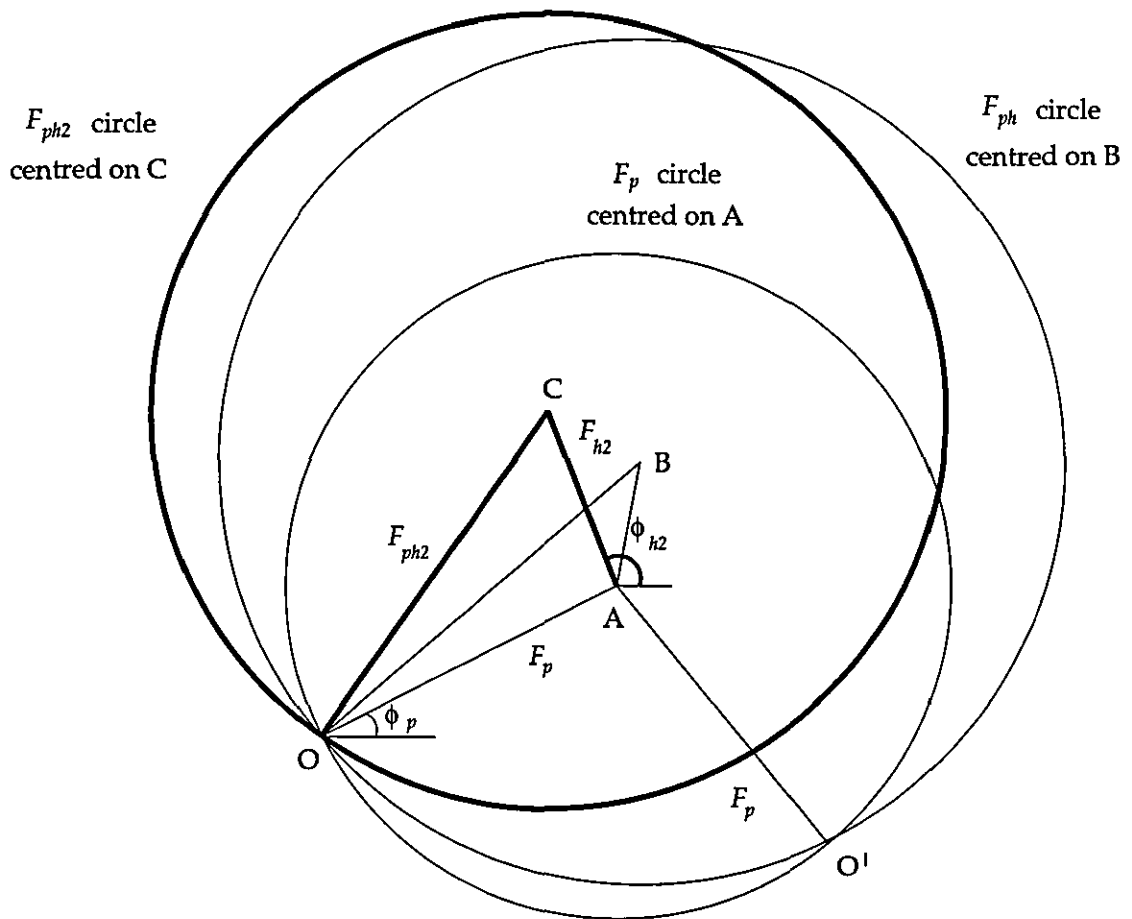


Figure 3.3: Graphical solution for the protein phase using multiple derivatives.

native protein and two heavy atom derivatives. The requirement thereby arises that the conformation of the molecules of each crystal must be close to identical.

The Effect of Errors in the Heavy Atom Method

It is only when ignoring errors that the structure factors F_p , F_h and F_{ph} form a closed triangle, as in figure 3.2. In practice, there is an uncertainty in the value of F_h and F_{ph} , which results in an uncertainty in the value of the phase angle. Blow and Crick [34] addressed this problem by assuming that the distribution of errors follows a Gaussian distribution, and introduced a phase probability distribution.

When the number of heavy atoms is small compared with the total number of atoms present within the unit cell, the error associated with the measurement of F_h can be ignored. The closed triangle can be modified, see figure 3.4, where ϵ is the error associated with the measurement of F_{ph} , and is called the lack of closure error.

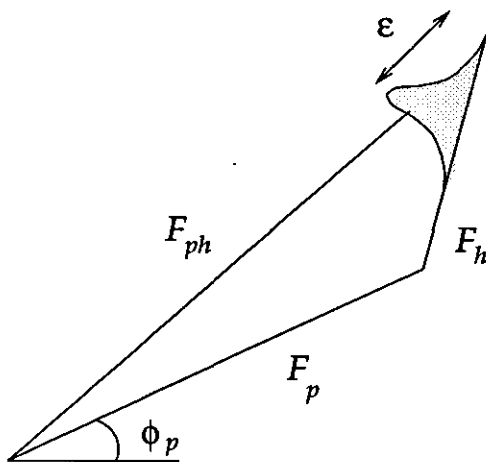


Figure 3.4: The effect of errors in the determination of the protein phase angle, ϕ_p .

Mathematically, ϵ can be written as in equation (3.7) where k is a scale factor. The lack of closure is the difference between the observed and theoretical derivative structure factor amplitudes, $|F_{ph}|_{observed}$ and $|F_{ph}|_{theoretical}$, respectively [35, p. 170].

$$\epsilon = k |F_{ph}|_{observed} - |F_{ph}|_{theoretical} \quad (3.7)$$

The smaller the lack of closure error, the greater the probability that the protein phase is correct, a relationship which Blow and Crick expressed as a probability function, equation (3.8)

$$P(\phi_{p1}) = P(\varepsilon) = N_f \exp - \left(\frac{\varepsilon^2}{2E^2} \right) \quad (3.8)$$

where N_f is a normalisation factor, given that the phase angle must lie somewhere between 0 and 2π . E is the root mean square error associated with the measurements.

The lack of closure offers a means of reducing the error in the phase angle allocated to each hkl observation. This procedure, called refinement, takes into account many parameters associated with the heavy atoms attached to the protein molecule. Firstly, their coordinates, secondly the degree to which the atoms 'vibrate' due to thermal energy (the so-called temperature factor, defined in equation (4.1)), thirdly the occupancy, defined as the percentage of a particular site, on all the molecules present in the crystal, to which a heavy metal ion is bound, and finally, the scale factor k . By modifying these parameters, the lack-of-closure error can be reduced, with the consequence that the phase angle ϕ_p can be obtained with greater certainty.

3.5 Anomalous Scattering

A second method for overcoming the phase ambiguity present in the method of single isomorphous replacement, makes use of anomalous scattering.

The Fundamentals of Anomalous Scattering

Anomalous scattering arises from the bound nature of atomic electrons. In the presence of an oscillating electric field, the equation of motion of a bound atomic electron is given by equation (3.9), where m is the electron mass, e the electron charge, k_r the restoring force per unit displacement, g a damping constant and E_o the electric field strength of the incoming radiation.

$$m\ddot{x} + g\dot{x} + k_r x = -E_o e \exp - i\omega t \quad (3.9)$$

Due to the presence of damping, the solution of this equation has the form

$$x = Ae^{i\omega t} \quad (3.10)$$

with A given by [29, p. 196]

$$\begin{aligned} A &= \frac{E_0 e}{(k_r - \omega^2 m) + ig\omega} \\ &= \frac{E_0 e[(k_r - \omega^2 m) - ig\omega]}{(k_r - \omega^2 m)^2 + g^2 \omega^2} \end{aligned} \quad (3.11)$$

For an unrestrained and undamped electron $k_r = 0$ and $g = 0$ whereupon

$$x = -\frac{E_0 e}{\omega^2 m} \exp(i\omega t) \quad (3.12)$$

which is the same as equation (2.4).

When $k_r \neq 0$, and $g = 0$, there is a resonance frequency corresponding to $\omega = \omega_0$ when

$$k_r - \omega_0^2 m = 0 \quad (3.13)$$

hence

$$k_r = \omega_0^2 m \quad (3.14)$$

and equation (3.11) can be rewritten as

$$A = \left(\frac{(\omega_0^2 - \omega^2)m E_0 e}{(\omega_0^2 - \omega^2)^2 m^2 + g^2 \omega^2} \right) - \left(\frac{E_0 e g \omega}{(\omega_0^2 - \omega^2)^2 m^2 + g^2 \omega^2} \right) i \quad (3.15)$$

The amplitude of the scattered wave at resonance can be compared with the amplitude scattered from a free electron. This is achieved by taking the real and imaginary components of equation (3.15) and dividing by the amplitude in equation (3.12). The real part is then

$$f' = \frac{-\omega^2 (\omega_0^2 - \omega^2) m^2}{(\omega_0^2 - \omega^2)^2 m^2 + g^2 \omega^2} \quad (3.16)$$

and the imaginary part,

$$f'' = \frac{g\omega^3 m}{(\omega_0^2 - \omega^2)^2 m^2 + g^2 \omega^2} \quad (3.17)$$

In figure 3.5, f' and f'' are plotted as functions of wavelength [29, p. 201]

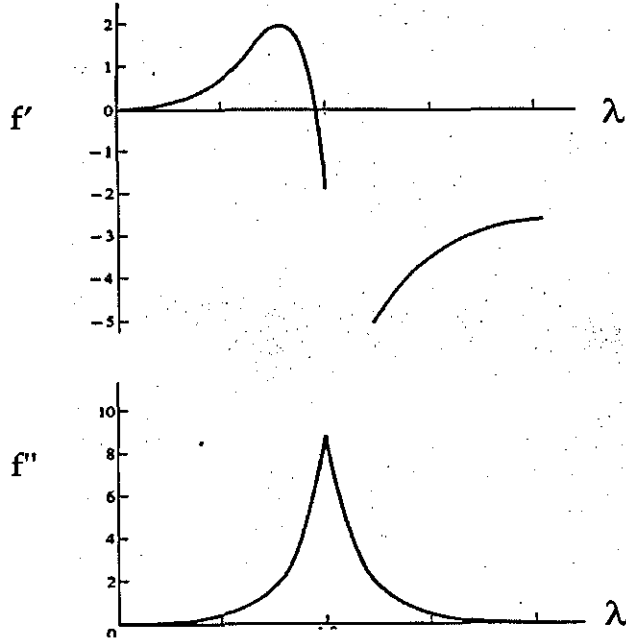


Figure 3.5: The real and imaginary fractional scattering amplitudes f' and f'' as functions of wavelength.

The effect of resonance on the atomic scattering factor is expressed by means of an anomalous increment, $\Delta f_a = f' + if''$, as suggested by James [36, p. 149] hence

$$\begin{aligned} f &= f_o + \Delta f_a \\ &= f_o + f' + if'' \end{aligned} \quad (3.18)$$

f_o represents the normal scattering as if the electron were free, f' is the real component of the anomalous scattering and is in-phase with f_o , f'' is the imaginary component of the anomalous scattering and has a phase which is $\frac{\pi}{2}$ in advance of f_o .

In the quantum mechanical description of the atom, the electron energy is restricted to certain discrete values. The energy associated with the electron when it oscillates at the resonant frequency ω_o is given as $E_o = \hbar\omega_o$, where \hbar is Planck's constant. If this energy is greater than the binding energy, the electron is ejected from the atom. As a consequence, the form of the f'' curve above is modified, an example of which is shown in figure 3.6, along with the f' curve.

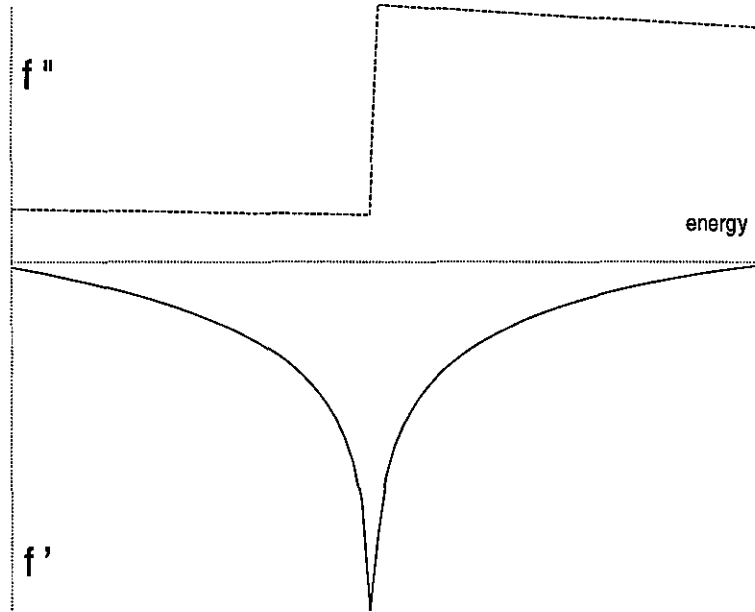


Figure 3.6: The real and imaginary fractional scattering amplitudes, in the quantum mechanical case.

The value of the damping constant, g can be determined experimentally from the linear absorption coefficient, μ , where μ is given by

$$\mu = \frac{N_e e^2}{2gc\epsilon_o} \quad (3.19)$$

where N_e is the number of electrons per unit volume, e the charge of the electron, c the speed of light and ϵ_o the permittivity of the medium. The absorption can be measured near the 'edge' of f'' in figure 3.6, where $\omega \simeq \omega_o$. Substituting for ω and g in equation (3.17) gives

$$f'' = \frac{2mc\omega\epsilon_o}{e^2} \mu \quad (3.20)$$

From equation (3.16), as $\omega \rightarrow \omega_o$, $f' \rightarrow \infty$, however, f' can be calculated using the Kramers-Kronig transform [37]

$$f' = \frac{2}{\pi} \int \frac{\omega f''(\omega)}{(\omega_o^2 - \omega^2)} d\omega \quad (3.21)$$

Substituting the atomic scattering factor from equation (3.18) into the structure factor equation, gives

$$F_{hkl} = \sum_{j=1}^N (f_o + f' + if'') \exp 2\pi i [hx + ky + lz] \quad (3.22)$$

The intensity of the diffraction pattern can now be calculated for any energy of incident X-rays.

The Breakdown of Friedel's Law

A consequence of the anomalous component f'' being $\frac{\pi}{2}$ in advance of f_o is the breakdown of Friedel's law. The graphical representation of this effect is shown in figure 3.7, where F_{ph}^+ and F_{ph}^- represent reflections related by a centre of symmetry, as was the case in figure 2.6. Here it can be seen that F_{ph}^+ no longer equals F_{ph}^- .

3.6 Single Isomorphous Replacement with Anomalous Scattering

Single isomorphous replacement with anomalous scattering (SIRAS) takes advantage of the breakdown of Friedel's law. Figure 3.8 shows a modified version of figure 3.7, in which F_{ph}^- has been reflected across the real axis. As in the SIR technique, data are collected from a native and a derivative crystal, however in the SIRAS technique, the heavy atoms in the derivative are exploited as sources of anomalous scattering. As a consequence, it is the anomalous structure factor amplitude that is obtained from the derivative. With a knowledge of $|F_{ph}^+|$ and $|F_{ph}^-|$, and the native structure factor amplitude $|F_p|$, SIRAS leads to a value of the protein phase, ϕ_p , as shown in figure 3.8.

OA represents the native structure factor, AB the normal heavy atom structure factor, BC and BD the imaginary part of the anomalous structure factor. A unique

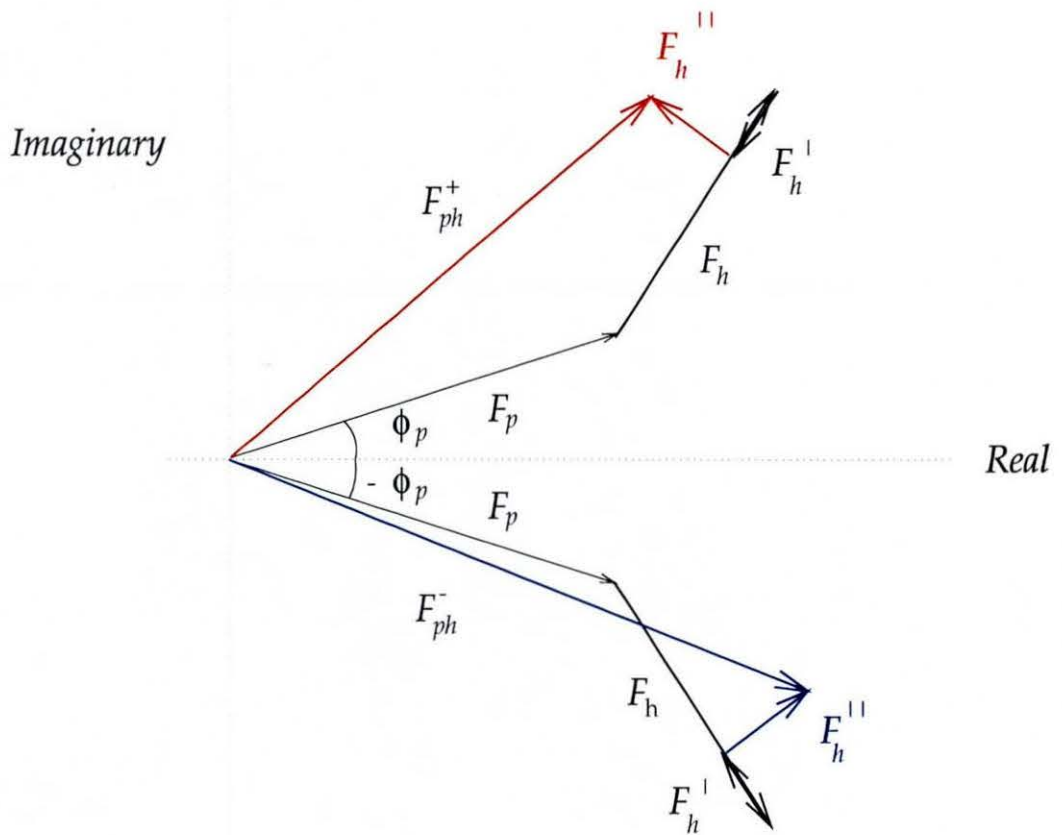


Figure 3.7: Graphical representation of the violation of Friedel's law due to anomalous scattering

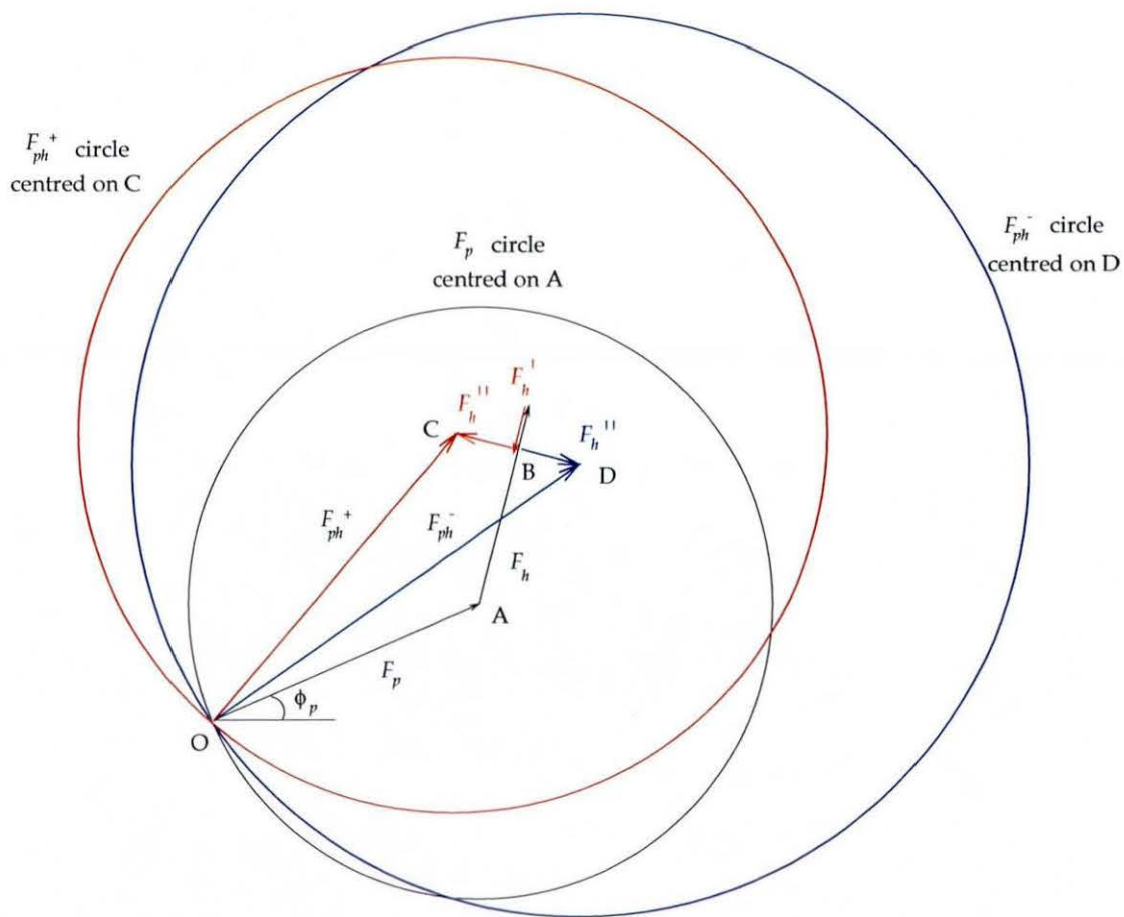


Figure 3.8: Solution of the protein phase using a single heavy atom derivative with anomalous scattering, SIRAS

solution for the protein phase is obtained at the point of intersection of the three circles; the first of radius F_p centred on point A, the second of radius F_{ph}^- centred on D and the third of radius F_{ph}^+ centred on C.

The effect of experimental errors is to make the identification of the unique solution more difficult. In this case, the normal and anomalous scattering from a second derivative can be included in the analysis, a method called multiple isomorphous replacement with anomalous scattering (MIRAS).

Bijvoet in his original paper of 1954 [2] suggested that anomalous scattering could be used in this way to resolve the phase ambiguity. This was later demonstrated by Blow [38], in the detailed study of the phasing of the reflections of the non-centrosymmetric zone $[1\ 0\ 0]$ of horse haemoglobin. The anomalous scattering was derived from mercury atoms. However, as pointed out by Blow, the combination of the anomalous scattering results with those from isomorphous replacement was difficult to achieve in a rigorous manner, leading to rather arbitrary results. Ramamchandran and Raman [39] also discussed how the incorporation of the anomalous scattering information in the technique of multiple isomorphous replacement would resolve the phase ambiguity. The earlier treatment of Blow was later made more rigorous by Blow and Rossmann [40], with a procedure closely analogous to figure 3.8, in which the breakdown of Friedel's law was considered to give two independent, isomorphous derivatives. The technique was successfully demonstrated by Blake et al. [41], but was found to be sensitive to errors in the scaling of the native and anomalous derivative datasets. As an alternative, North [42] considered the intensity difference between the Bijvoet pairs, Δ_{ano} , given in equation (3.23).

$$\Delta_{ano} = |F_{ph}^+| - |F_{ph}^-| \quad (3.23)$$

The advantage of this approach was that Δ_{ano} would be derived from the anomalous scattering from one specimen, thereby avoiding scaling errors inherent in the method proposed by Blow and Rossmann.

3.7 Multiwavelength Anomalous Dispersion

In the technique of multi-wavelength anomalous dispersion (MAD) the phase problem can be solved from a derivative protein, without reference to the native. The MAD technique exploits both the intensity difference between Bijvoet pairs, so-called anomalous difference, and the intensity difference between the same reflection measured at different wavelengths, the so-called dispersive difference. The effect of changing wavelength on the measured intensity is described schematically in figure 3.9

A formulation of the wavelength dependence of the observed structure factor was first proposed by Karle [6], and later modified by Hendrickson [7], [8]. Unlike SIRAS, the MAD technique does not lead to a value of the native protein phase, ϕ_p . In the Hendrickson approach, it is the phase of the structure factor associated with all atoms in the absence of anomalous scattering, which is calculated. This is denoted ϕ_T , and is shown in figure 3.10, where F_A^o is the structure factor of the normal component of scattering from the anomalous scatterers, and ϕ_A its associated phase.

Karle's original phase equation then could be re-written as in equation (3.24) below [8].

$$\begin{aligned}
 |\lambda F_{obs}^{\pm}|^2 &= |F_T^o|^2 \\
 &+ a(\lambda) |F_A^o|^2 \\
 &+ b(\lambda) |F_T^o| |F_A^o| \cos(\phi_T^o - \phi_A^o) \\
 &\pm c(\lambda) |F_T^o| |F_A^o| \sin(\phi_T^o - \phi_A^o)
 \end{aligned} \tag{3.24}$$

where

$$a(\lambda) = \frac{f'^2 + f''^2}{(f^o)^2} \quad b(\lambda) = 2 \left(\frac{f'}{f^o} \right) \quad c(\lambda) = \pm 2 \left(\frac{f''}{f^o} \right) \tag{3.25}$$

where f^o , f' and f'' are the scattering factors as previously defined. All wavelength dependent quantities therefore became simple scattering factor ratios, $a(\lambda)$, $b(\lambda)$, $c(\lambda)$,

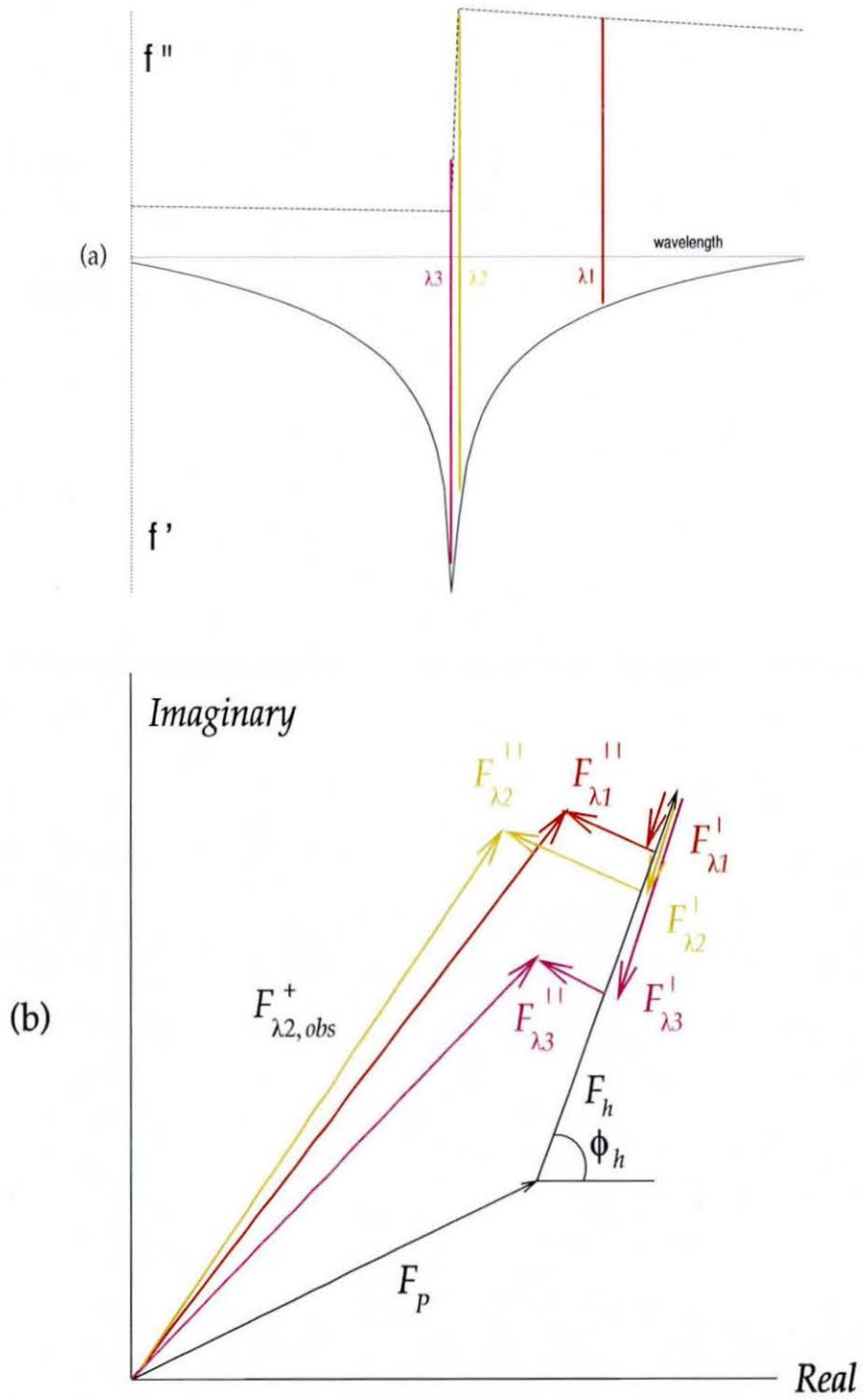


Figure 3.9: The wavelength dependence of the observed structure factor, F_{obs} . F'_{λ_1} , F'_{λ_2} and F'_{λ_3} represent the structure factors for the dispersive differences at λ_1 , λ_2 and λ_3 , respectively. F''_{λ_1} , F''_{λ_2} and F''_{λ_3} , the structure factors of the anomalous differences at the respective wavelengths.

Imaginary

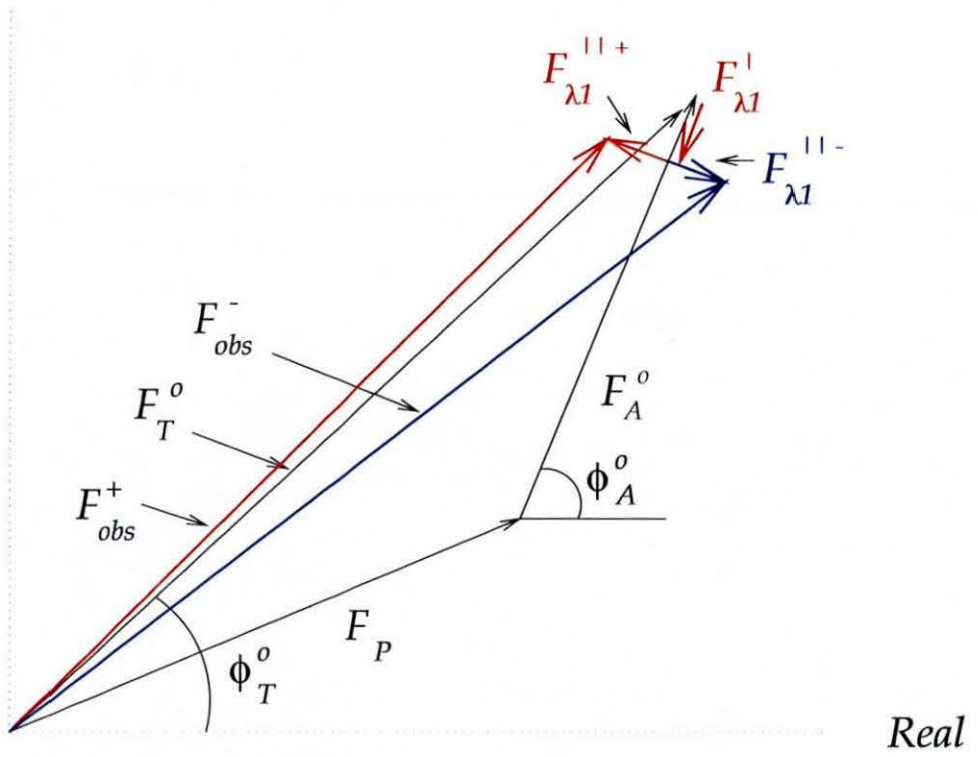


Figure 3.10: Argand diagram according to Hendrickson's formulation

and were independent of structure. In addition, the quantities $|F_T^o|$, ϕ_T^o and $|F_A^o|$, ϕ_A^o , were only dependent on normal scattering, and therefore independent of wavelength.

In the procedure developed by Hendrickson, it is the parameters $|F_T^o|$, $|F_A^o|$ and $(\phi_T^o - \phi_A^o)$ which are derived by least squares fit of the phase equation to the multiple measurements of each reflection. It is necessary to have estimates of the real and anomalous scattering factors for this calculation to be performed. The parameter $|F_A^o|^2$ can then be used as a coefficient in a Patterson synthesis or direct methods, in order to determine the partial structure of the anomalous scatterers, and thereby ϕ_A^o . With this knowledge, ϕ_T^o can be obtained.

In an alternative method, MAD has been considered in a manner analogous to multiple isomorphous replacement [9]. The graphical description of this approach is shown in figure 3.11, where only one of the Bijvoet reflections is shown, in the two wavelength case. \bar{F}_{obs,λ_3} is the structure factor of the average of the Bijvoet difference at the wavelength λ_3 , F_{obs,λ_1}^+ is a Bijvoet measurement at the second wavelength λ_1 and ϕ_H is the phase angle obtained from the heavy atom substructure. It can be seen that the triangle OAB , is analogous to the triangle OAB in figure 3.3.

In the MIR approach, it is the phase angle ϕ_N which is calculated by choosing one of the datasets to be a 'pseudo-native'. The remaining datasets are then considered to be 'pseudo-derivatives' with the addition of anomalous scattering. The phase is obtained using the anomalous differences of each 'derivative' and the dispersive ('pseudo-isomorphous') differences between the 'native' and 'derivatives'.

The graphical solution for the phase ϕ_N is shown in figure 3.12, where the colours represent the measurements of the Bijvoet reflections at each wavelength. Six circles are drawn, which coincide giving a unique value for the phase.

Imaginary

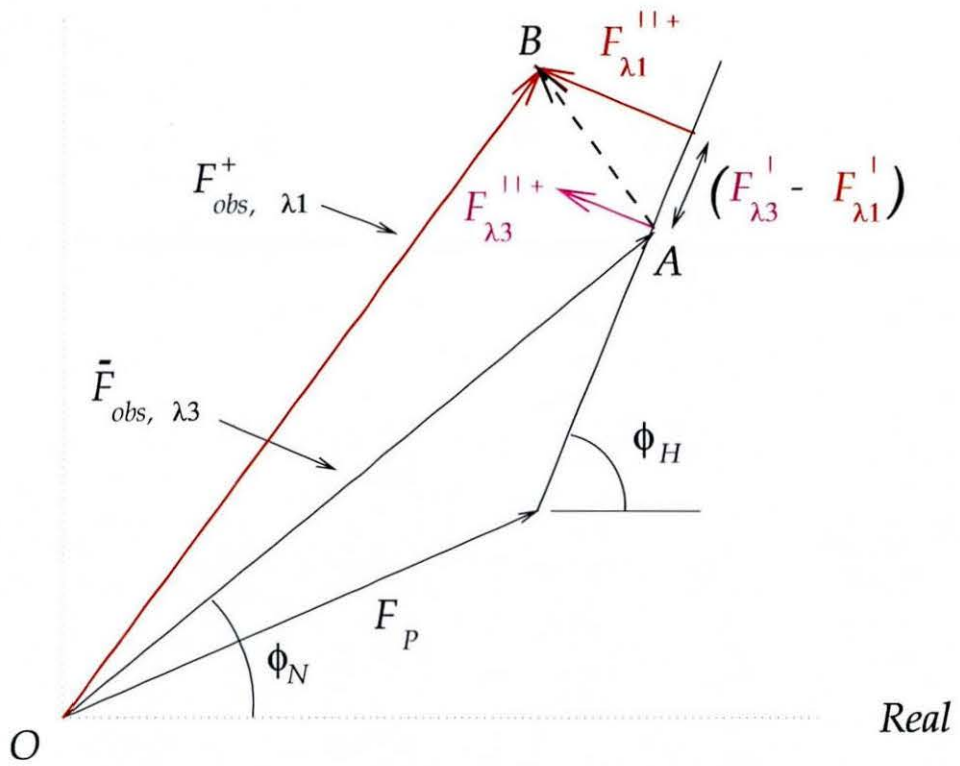


Figure 3.11: Argand diagram description of MAD in a manner analogous to isomorphous replacement. The wavelengths λ_1 and λ_3 are chosen for consistency.

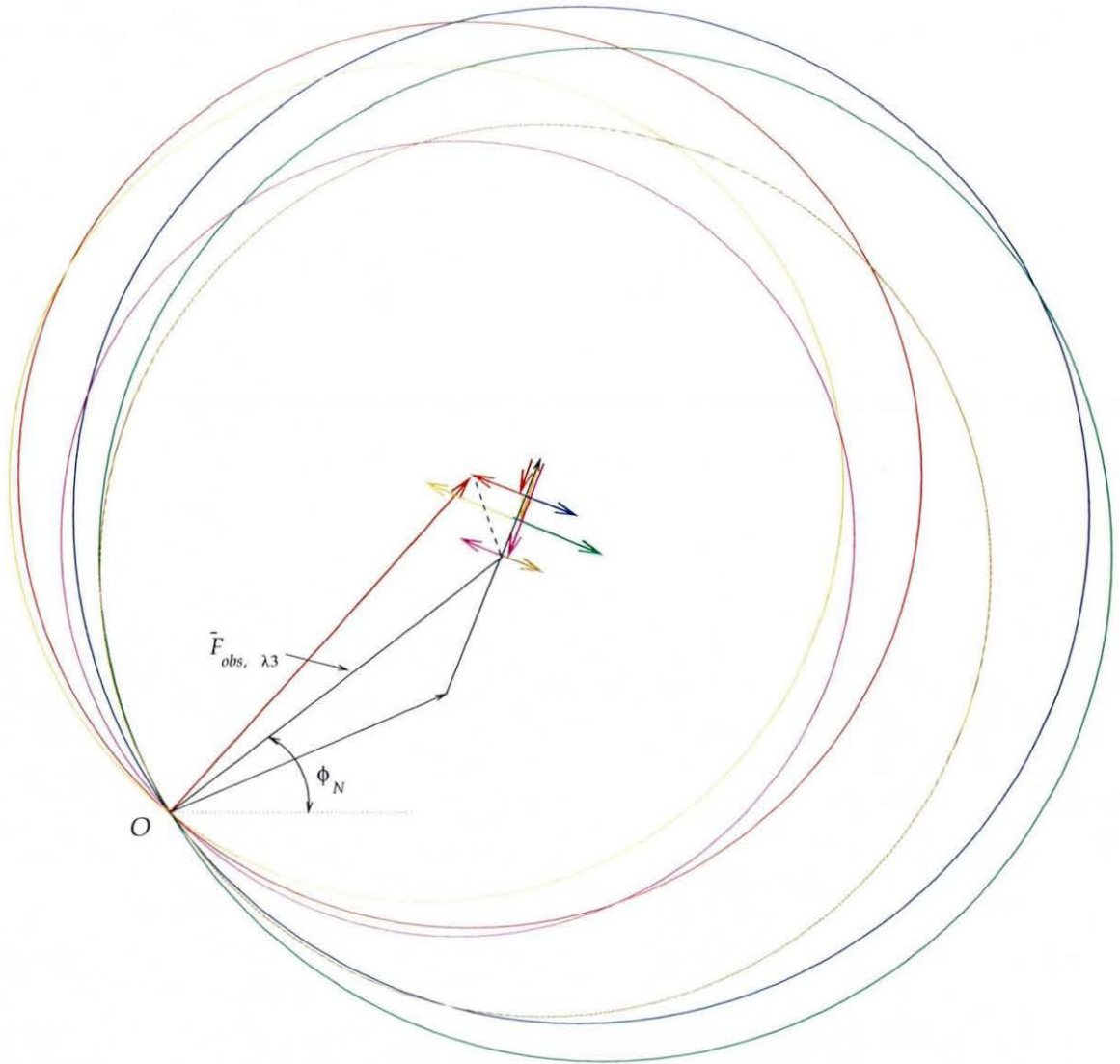


Figure 3.12: The graphical solution for the phase angle ϕ_N in the MIR approach to MAD.

Chapter 4

After summarising the characteristics of different detector systems, this chapter describes the *fast image plate detector*; the *design criteria* and the *control software*. This detector was conceived for collecting diffraction data rapidly from protein crystals, using the technique of multi-wavelength anomalous dispersion, and was installed on the beam-line BM14 at the European Synchrotron Radiation Facility.

Detectors for X-Ray Crystallography

As previously stated in chapter 3, the imaging process in X-ray crystallography is only partially completed, with the formation of the diffraction pattern. Detectors are only able to measure the intensity of features within this diffraction pattern. This intensity is proportional to the squared structure factor magnitude where the coefficient of proportionality is composed of many factors, represented algebraically in equation (4.1) [44].

$$I_{hkl} = \left(\frac{\lambda^3 r_o^2 V}{V_c} \right) \left(\frac{1}{\omega_s} \right) LPABE |F_{hkl}|^2 \quad (4.1)$$

where λ is the wavelength, r_o the classical electron radius, V the crystal volume, V_c the unit cell volume, ω_s the angular scanning velocity, L the Lorentz factor, P the polarisation factor, as described in equation (2.7), A the absorption factor, B the temperature factor, E the extinction factor.

The Lorentz factor accounts for the fact that reciprocal lattice points in different regions of space pass through the Ewald sphere at different speeds. It is a purely

geometric factor and can be expressed as in equation (4.2) [45, p. 179].

$$L = \frac{1}{\sin 2\theta} \quad (4.2)$$

The dominant absorption process of X-rays within a specimen is the photoelectric effect. It can be quantified using the expression given in equation (4.3), where the incident X-ray intensity is given by I_i and the outgoing intensity by I_o . μ is the linear absorption coefficient, and x the thickness of the specimen.

$$I_o = I_i e^{-\mu x} \quad (4.3)$$

The temperature factor describes the effect of the thermal motion of the atoms present in the sample. These atoms vibrate with a frequency which is much lower than the frequency of the X-ray radiation, as a result only the time-averaged position of the atoms is recorded in a diffraction pattern. The intensity of scattered radiation is reduced by a factor shown in equation (4.4) [45, p. 189].

$$B = \exp\left(\frac{-2B_j \sin^2\theta}{\lambda^2}\right) \quad (4.4)$$

where

$$B_j = 8\pi^2 \langle \delta(r)^2 \rangle$$

$\delta(r)^2$ being the squared mean-displacement of the atom from its equilibrium position.

The extinction correction arises from the effects of multiple scattering of X-ray waves within the sample. In the kinematic theory of diffraction, its effects are ignored.

4.1 Detector Types

The detection of X-rays relies on the process of photoemission, where an incident X-ray photon interacts with a bound atomic electron. The energy deposited by the photon then promotes the electron to an unbound state, where it may interact with neighbouring atoms, resulting in further ionisation.

In crystallographic experiments, X-ray photons have an energy in the range 5 - 20 keV. This is many orders of magnitude greater than the ionisation energy of an atom, hence one photon is likely to ionise many atoms.

Area X-ray detectors fall into two categories; photon-counting detectors and integrating detectors.

Photon counting detectors, such as multi-wire proportional counters, record an event each time an X-ray photon passes into the detector. There is no means of internal storage, the detector is read out many times per second. Such detectors are count-rate limited.

Integrating detectors, such as films, image plates and CCD detectors, accumulate some physical event each time an X-ray photon impinges upon the surface. During an exposure, information is stored within the detector. Such detectors are exposure-time limited.

4.1.1 Characteristics of Detector Systems

The performance of a detector system can be described using five principal quantities; the dynamic range, the spatial resolution, the detective quantum efficiency, the linearity of response and the uniformity of spatial response. According to Allinson [46], a detector should be able to measure diffraction intensities within an uncertainty of a few percent, and determine the spot position with an uncertainty of one pixel.

The *dynamic range* is defined as the ratio between the largest and smallest detectable signal per pixel, without saturation.

The *spatial resolution* is the minimum distance between two neighbouring spots, such that they may still be resolved. It is expressed in terms of the point spread function (PSF). The PSF is the image generated on the detector by an input having the dimensions of only one pixel. It represents the minimum achievable spot size which, in general, will be larger than one pixel as a result of noise introduced by the detector.

The *detective quantum efficiency* (DQE) is a parameter used to describe the efficiency of the detector. It is defined as the ratio of the mean number of detected photons to the mean number of incident photons per pixel [46].

The *linearity of response* describes the relationship between the X-ray dose and the magnitude of the signal output by the detector. These two parameters should be linearly proportional.

The *spatial response* describes the sensitivity of local regions of the active surface. The signal output when the detector is uniformly illuminated by X-rays should be uniform for all regions.

4.1.2 Film

A film emulsion contains crystalline grains of silver bromide as a colloidal solution in gelatin. An incident X-ray photon creates electron/hole pairs from the AgBr ionic molecule. The electrons may become trapped by defects in the crystal, while the holes are trapped at interstitial silver atoms to form silver ions. During development, it is the density of these silver ions which is recorded.

Film has a limited dynamic range, which necessitates the stacking of up to three pieces of film behind each other for each exposure. X-rays will be subject to absorption as they pass through each film, therefore an intense diffraction spot which saturates the first film, might be attenuated enough to be measured on the successive film.

A second inconvenience of film is the non-linearity of energy response due to the presence of bromine in the emulsion. Absorption by the bromine ions is observed when the X-ray wavelength approaches about 0.9\AA .

A third drawback of film is the development procedure. A dataset can easily comprise several hundred pieces of film; each must be chemically developed before being electronically digitised, in order that the diffraction data can be processed. This is an inconvenient and time-consuming procedure, which prevents data from being processed whilst the experiment is being performed.

4.1.3 Image Plates

An image plate is composed of a layer of organic binder mixed with BaFBr:Eu^{2+} . The ionic compound BaFBr:Eu^{2+} has the PbFCl structure [47] and is a photo-stimulable X-ray storage phosphor.

Electron/hole pairs are created in BaFBr:Eu^{2+} when ionising radiation impinges upon it. The hole is trapped by the Eu^{2+} ion, forming Eu^{3+} . The electron is attracted to the positive charge environment created by a bromine ion vacancy, thereby forming an F-centre [48].

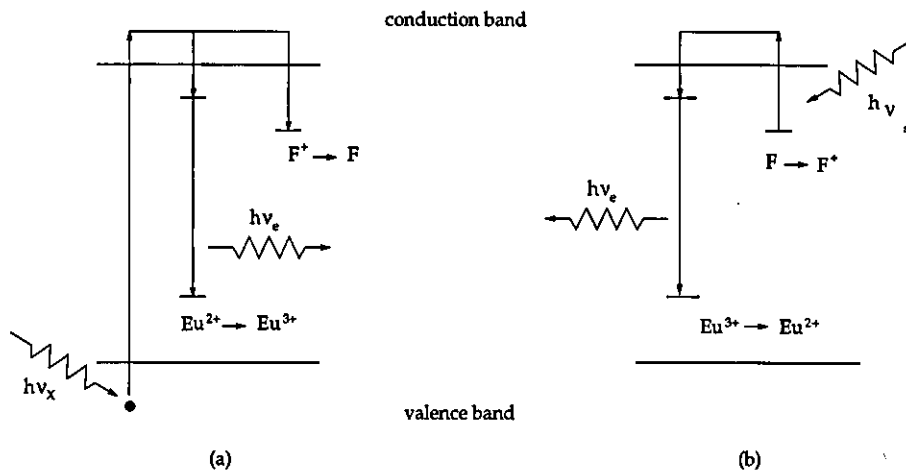


Figure 4.1: Energy level diagram for irradiation/photostimulation of an image plate

The F-centre is a storage site of the energy of the incident X-ray photon. This energy is recovered by optical stimulation.

The F-centres are destroyed by the application of infra-red light, whereupon the electron is sent into the conduction band. Being free, this electron is available for capture by the Eu^{3+} ion. This capture is accompanied by the emission of blue light, a process called photo-stimulated luminescence.

In contrast to film, image plates offer a large dynamic range, and do not involve any chemical development procedure prior to electronic digitisation. However, the reading and erasing cycle generally takes a few minutes, and can render on-line image plate detectors unsuitable for synchrotron sources of high brilliance. Image plates do have

the advantage of being re-usable, and having a large active area. One disadvantage of image plates however, is that the intensity of an image stored in the phosphor, is subject to decay.

4.1.4 Charged-coupled Device

A charged-coupled device (CCD) is a silicon semi-conductor chip on which is defined a grid of potential wells [49]. The incident X-ray photon creates electron/hole pairs, and the electrons become trapped in these potential wells. Each row and column of wells is connected separately, allowing the position and intensity of the incident X-ray photon to be determined.

The CCD-based detector has a large dynamic range, and is very sensitive to X-rays. In addition, images can be read very rapidly. However, one disadvantage of this type of detection system is that of dark-current. Dark-current is a signal output by the detector, arising from the accumulation of electronic noise, when no X-rays are incident on the detecting surface. It results in the systematic increase of the observed intensities. Being a temperature related effect, the dark-current can be reduced by cooling the electronic circuits.

4.1.5 Multi-wire Proportional Counter

The multi-wire proportional counter (MWPC) is a gas-filled detector containing grids of fine wires [50]. It is a photon-counting device. The incident X-ray photon is absorbed by a gas atom, which then emits either a photoelectron or an Auger electron. Since the energy of the X-ray photon is much greater than the ionisation energy of the atoms, each photon creates a cloud of ions and electrons. These charged particles drift in opposite directions due to the presence of an electric field, towards a fine grid of wires constituting anode and cathode electrodes. The intensity of the radiation impinging on the active surface is determined by measuring the magnitude of the current which flows between these electrodes. The position of the interaction is obtained by recording which wires receive the current pulse.

Type	Characteristics							Drawbacks
	PSF 2% of peak intensity	Spatial resolution (FWHM)	Linearity of response	Dynamic range	Uniformity of spatial response	Energy range	DQE	
film	$\approx 5\mu\text{m}$	$<25\mu\text{m}$	✓	$10^2 : 1$	✓	0 - 10keV	$>10\%$	<ul style="list-style-type: none"> • chemical processing procedure • not reusable
image plate	$400\mu\text{m}$	$\sim 100\mu\text{m}$	✓	$10^5 : 1$	✓	0 - 30keV	$<80\%$	<ul style="list-style-type: none"> • memory effects • long read-out time • latent image decay
CCD	$40\mu\text{m}$	$<60\mu\text{m}$	✓	$10^5 : 1$	✓	0 - 30keV	$>20\%$	<ul style="list-style-type: none"> • dark-current • damaged by high dose
MWPC	$500\mu\text{m}$	100 - 200 μm	✓	$10^5 : 1$	✓	0 - 85keV	$<70\%$	<ul style="list-style-type: none"> • count-rate limited • parallax errors • damaged by high dose

Table 4.1: Summary of the characteristics of detector systems. The values of the DQE are taken from the following references; IP [51], MWPC [52], CCD [46]. The point spread function (PSF) values are from [53, p. 253].

MWPC are count-rate limited, which restricts their use to low flux X-ray sources. Their spatial resolution is lower than that of other detecting media, and makes the MWPC more suitable for small-angle diffraction studies. However, the fast read-out time offers the advantage that this type of detector is well adapted to time-resolved experiments.

The characteristics of each of the detectors described above is summarised in table 4.1.

4.2 Requirements for Protein Crystallography

Protein crystallography imposes well-defined requirements on a detector system.

Protein crystals have large unit cells, containing many hundreds of atoms, as a consequence, diffraction patterns contain a large number of spots, closely spaced together. It is necessary to have *good spatial resolution* if these spots are to be resolved. Protein crystals may diffract from low resolution to very high resolution. In order to record large regions of reciprocal space at the same time, the detector must have a

large active area and be *compact* to allow the crystal to detector distance to be changed easily. X-ray crystallographic experiments are performed within an energy range of 5 to 20 keV, any detector must be *sensitive* to X-rays within this range. In addition, the response of the detector at all energies within this range must be *uniform*. At any particular energy chosen, the detector's response must be *linear* with respect to X-ray dose, so as to represent relative intensities of spots realistically. In addition, the detector must have a *wide dynamic range* if it is to record both very intense and very weak reflections simultaneously. Within diffraction patterns, the intensities of symmetry related reflections are often compared. Such reflections are recorded on different regions of the detector, for the successful treatment of the data, it is therefore important that the detector has a *uniform spatial response*. Since protein molecules are sensitive to radiation damage, it is important that the detector is sensitive to those X-rays diffracted by the crystal, thus having a *high DQE*, and that the diffraction data are collected *rapidly*.

Requirements specific to MAD Experiments

The success of MAD experiments is dependent upon accurately measuring the intensity difference between reflections. These reflections are recorded at different positions on the detector and at different X-ray energies. The detector must therefore offer a *uniform spatial response* as well a *uniform energy response*, if the intensities are to be determined reliably.

Rapid data collection in MAD experiments offers the advantage that complete datasets at many wavelengths can be collected from one crystal. In this manner, potential problems arising from crystal-to-crystal differences can be avoided.

4.3 The Fast Image Plate Changer

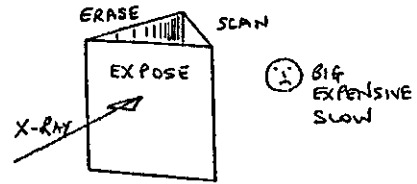
Conception

The conception drawings are shown in figure 4.2, where the development of the cylindrical form of the detector can be seen.

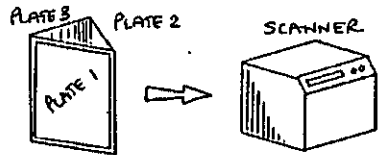
An on-line detector (*drawing 1*) was not considered to be suitable, since it would not offer any advantages over commercially available detectors. An off-line detector, where many plates are first exposed and later scanned (*drawing 2*), would fulfill the requirement of rapid data collection. Unfortunately, off-line scanning introduces the problem of plate identification (*drawing 3*), since the order in which plates were exposed must be known, so that diffraction images are contiguous. This task is particularly difficult when transporting the image plates from the detector to the scanner. A barcode system was considered to be a reliable and efficient solution (*drawing 3c*), where each image plate can be identified both at the detector and later at the scanner. An additional potential problem of an off-line system concerned the reproducibility of the plate positions. Variations might potentially arise from translational or rotational movement (*drawing 4*). Fiducial spots were considered the best solution. However, these spots would need to be recorded quickly, since the exposure times during data collection could be in the order of 5 to 10 secs. The proposed solution draws on the fact that image plates are sensitive to ultra-violet light. Using a xenon arc lamp, with a suitable low band-pass filter to remove the infra-red component, ultra-violet light passes through two fibre optic cables, and is focused onto the image plate. If the cables and lenses are made from fused silica, the ultra-violet light would suffer very little attenuation. In order to protect the image plates from exposure to daylight, they are inserted into a cassette, having a cover which is transparent to X-rays (*drawing 5*). The additional advantage of the barcode identification system would be that the data collection procedure could be controlled by software, running on the control computer (*drawing 6*).

Figure 4.2: Original conception drawing of the fast image plate detector

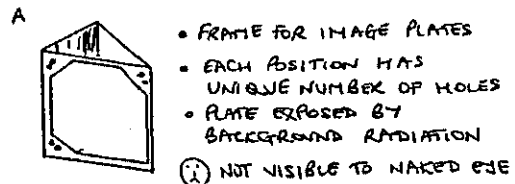
① ROTATING TRIANGLE - ON LINE



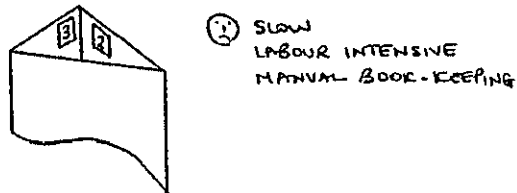
② ROTATING TRIANGLE - OFF LINE



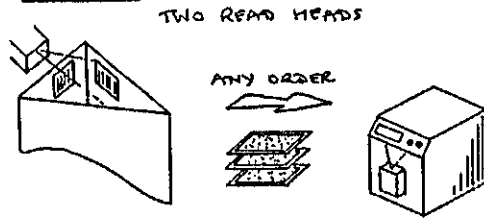
③ PLATE IDENTIFICATION



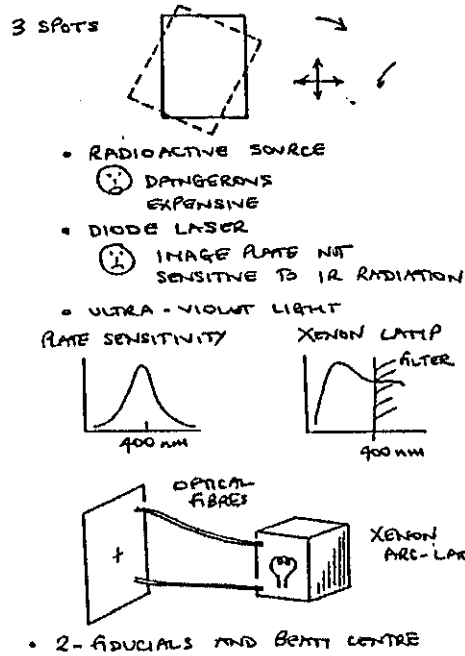
B PLATE NUMBERS



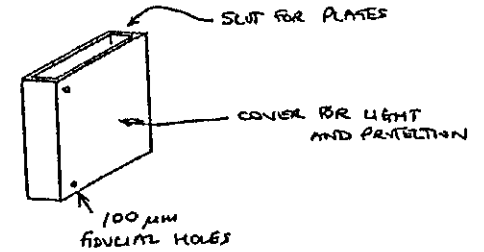
C BARCODE



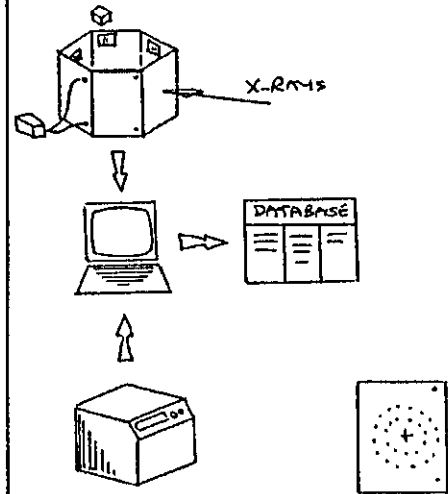
④ FIDUCIAL SPOTS



⑤ IMAGE PLATE CASSETTE



⑥ SYSTEM



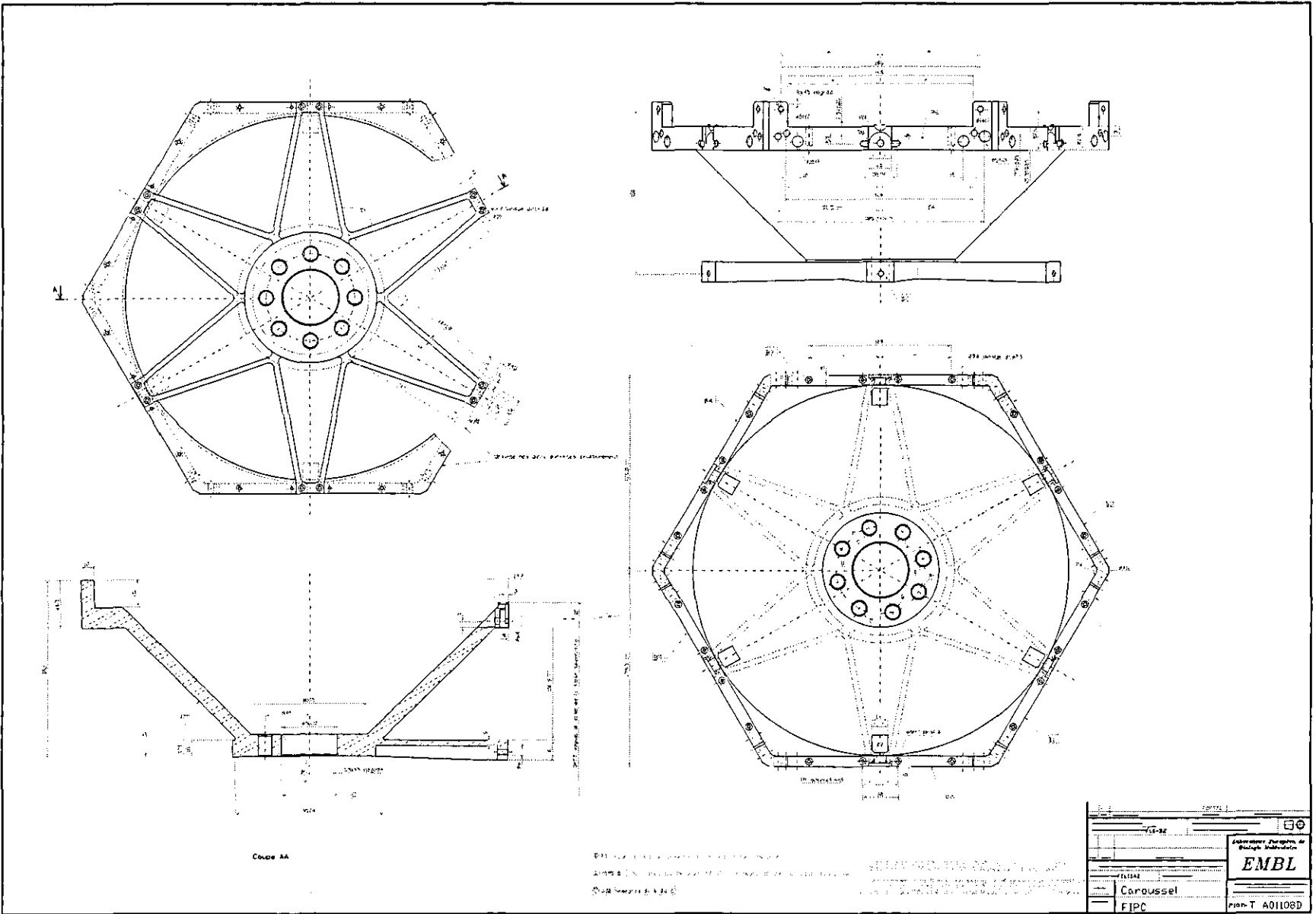


Figure 4.3: Technical drawing of the carousel frame [54]

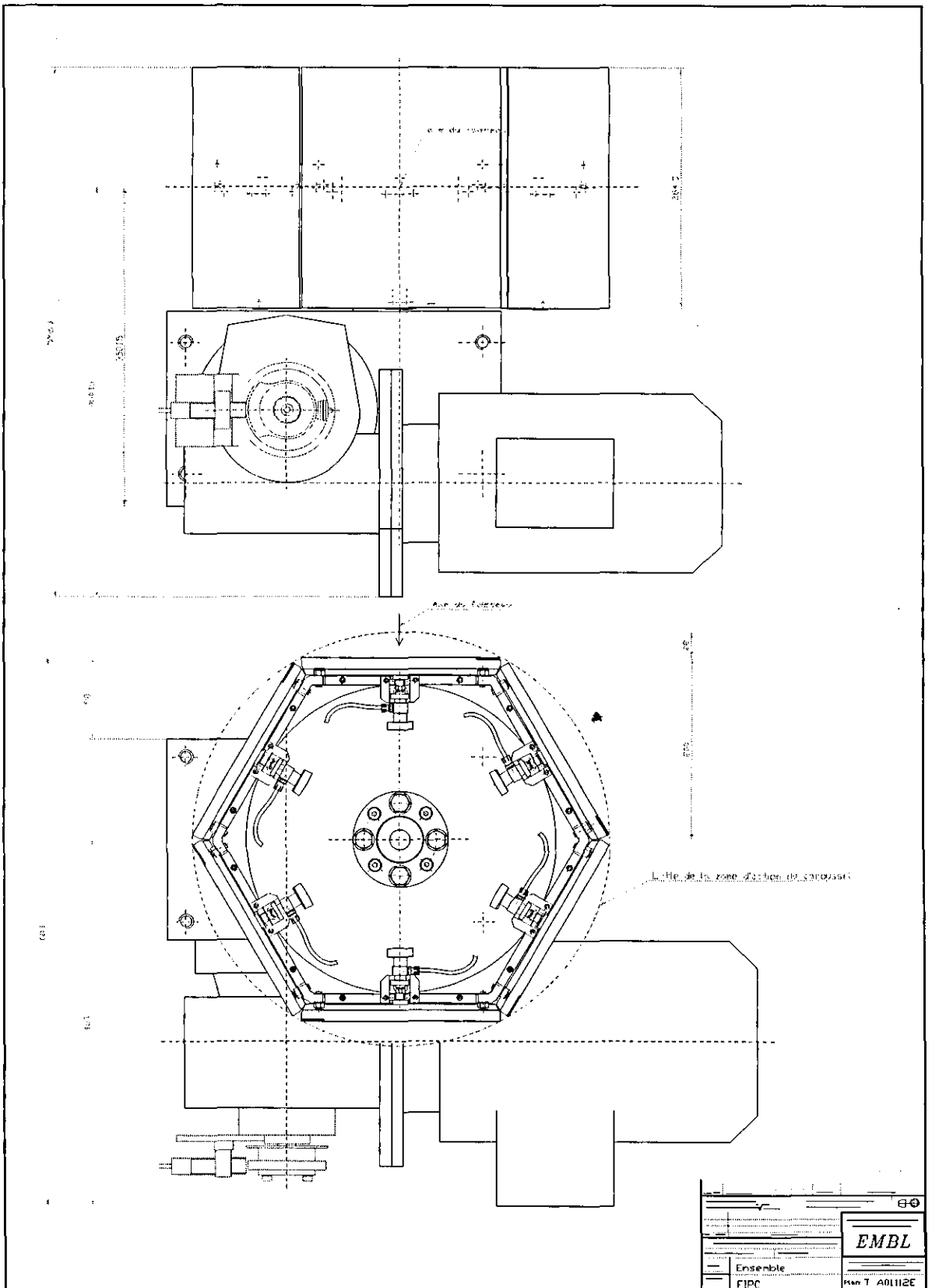


Figure 4.4: Technical drawing of the detector, showing the cassettes mounted, and the indexed rotation table [54]

Detailed Description

The principal component of the detector is the cassette support frame, called the carousel, illustrated in figure 4.3. It is a rigid frame constructed from aluminium, having the form of a cone. Six cassettes can be mounted on this carousel, via a mounting bracket and hand operated locking nut. The carousel itself is mounted on a table which rotates through six, indexed positions. Figure 4.4 shows how the cassettes and carousel are mounted on this table. The rotation between successive positions takes a period of $150\mu\text{s}$, the positional accuracy being achieved by a light sensor, integrated into the drive shaft of the motor.

Figure 4.5 shows how the cassettes are mounted onto the carousel.

Image plates are held in individual cassettes which protect against daylight and damage by a carbon fibre cover, see figure 4.5. In order to reduce uniformity of response errors to a minimum, the image plate is held in position within the cassette by a vacuum system. The basis of this system is a grid of channels present on the internal surface of the cassette. The vacuum is maintained whilst the cassette is fixed to the carousel by the opening of a valve situated in the cassette mounting bracket. Upon removing the cassette from the carousel, (see photograph), the valve automatically closes, and the vacuum is released.

The position of the image plate cassettes in the six mounting positions, is reproducible to within $100\mu\text{m}$ in translation and 0.02° vertical and horizontal angle. These values were calculated from an experiment in which the carousel was placed on an optical bench, with a mirror attached to the face of each of the six cassettes. The cassettes were mounted on the carousel, and a laser beam arranged to be at normal incidence to one of them. A screen with a pin-hole at its centre was placed in front of the laser at a distance of 10m from the carousel, and position of the reflected spot recorded. After rotation, the displacement of the reflected beam was measured. By geometry, the variation of cassette orientation and position could be calculated.

For the purposes of producing the fiducial spots, a compact xenon arc-lamp, with

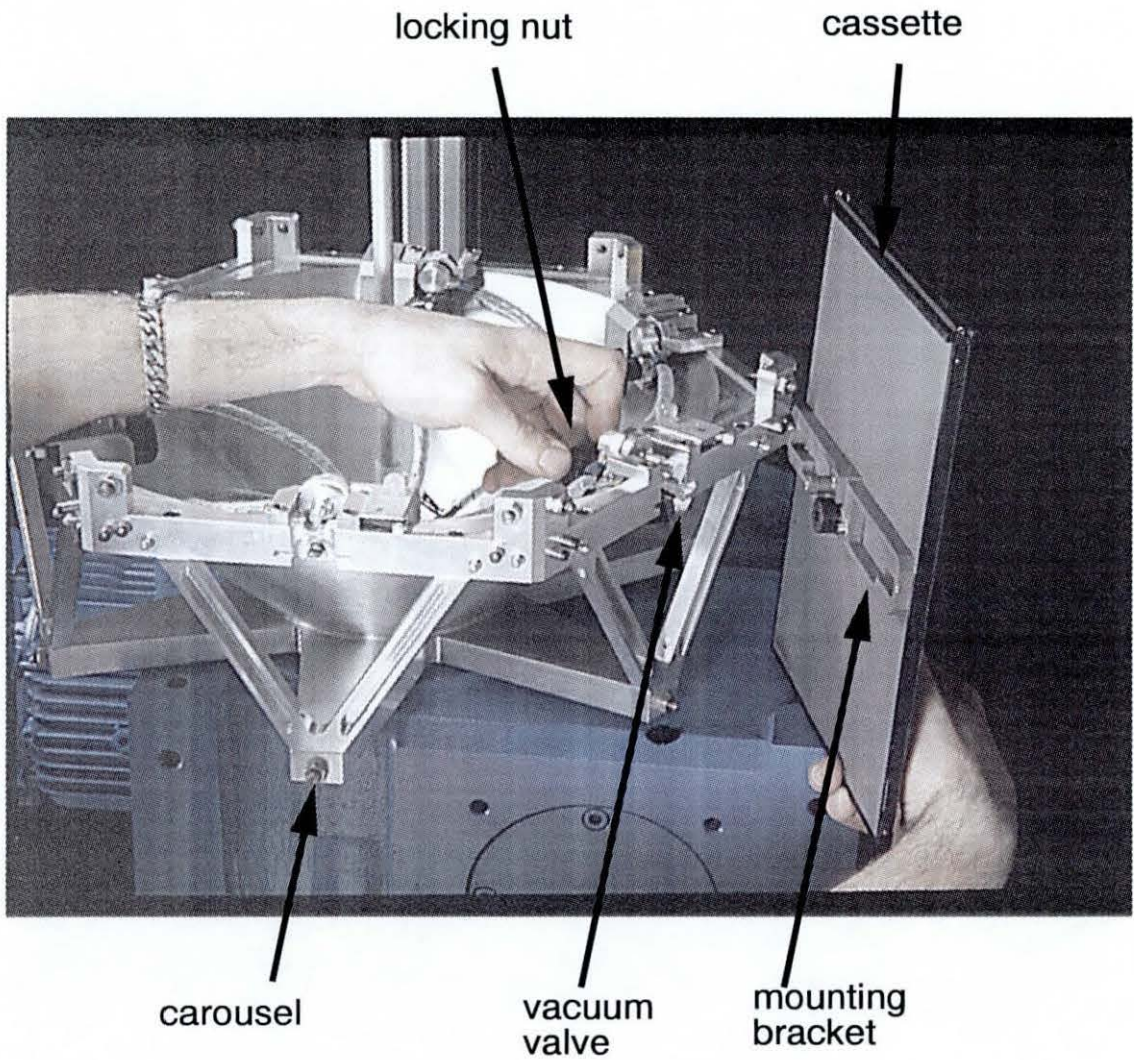


Figure 4.5: Photograph showing the carousel, vacuum system, cassette and mounting bracket

Problem	Solution
uniform response	flatness of cassette surface $< 50\mu\text{m}$ vacuum support system
reproducible plate position	carousel frame and cassette reproducible $< \pm 100\mu\text{m}, \pm 0.02^\circ$ fiducial spot system
rapid data collection	carousel rotation $\sim 150\mu\text{s}/60^\circ$ 2 sets of 6 plates, off-line scanning
latent image decay	software records elapsed time between exposure and scan

Table 4.2: Design criteria of the fast image plate changer

integral electronically operated shutter is placed next to the detector. A bifurcated optical fibre made from fused silica is connected to the shutter. Small lenses, also made from silica, are attached to the ends of the optical fibre, and focus the ultra-violet light to a spot of about 3mm diameter. The end of each fibre is aligned with a $100\mu\text{m}$ hole drilled near two corners of the cover of each cassette.

A barcode sticker is placed in the same position on the back of each of the image plates. These barcodes are visible through a small window in the cassette, so that, at the time of the exposure, the image plate can be identified by the reading head.

A photograph of the detector is shown in figure 4.6.

A summary of the design criteria is given in table 4.2.

The detector uses image plates of size $20 \times 25\text{cm}$, and can be translated to a minimum crystal to film distance of 150mm. When using an X-ray wavelength of 1\AA data of about 1.4\AA resolution can be recorded.

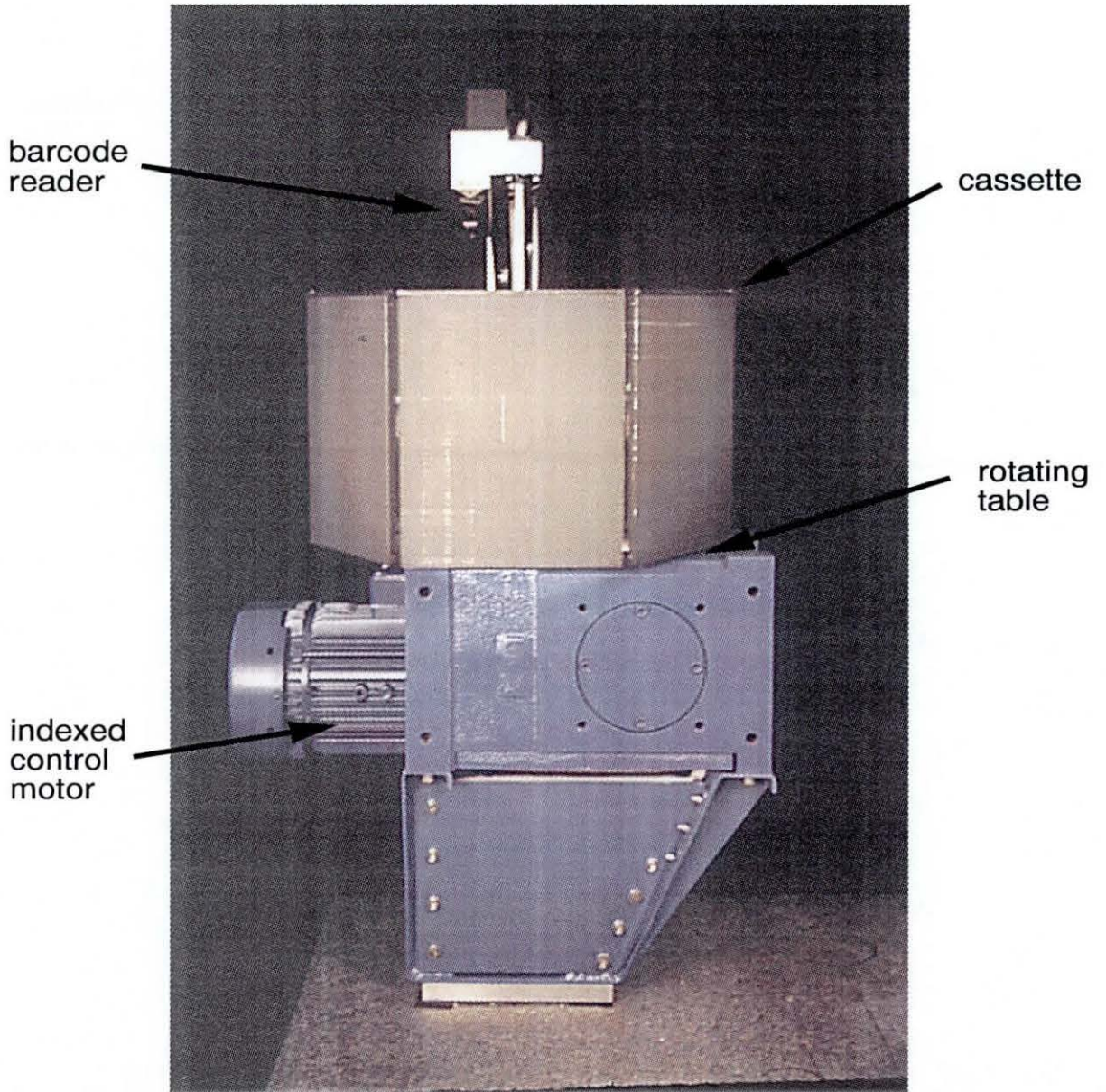


Figure 4.6: Photograph of the detector assembly, showing the cassettes mounted on the carousel, and the indexed rotating table

Control Software

The detector is controlled by computer software, which performs the following principal tasks; the identification of the image plate to be exposed, the control of the X-ray shutter and spindle axis, and the rotation of the indexed table. Furthermore, the software is responsible for identifying the image plates prior to the off-line scanning. The schematic representation of the tasks performed by the software is shown in figure 4.7, where the definition between detector and scanner can clearly be seen.

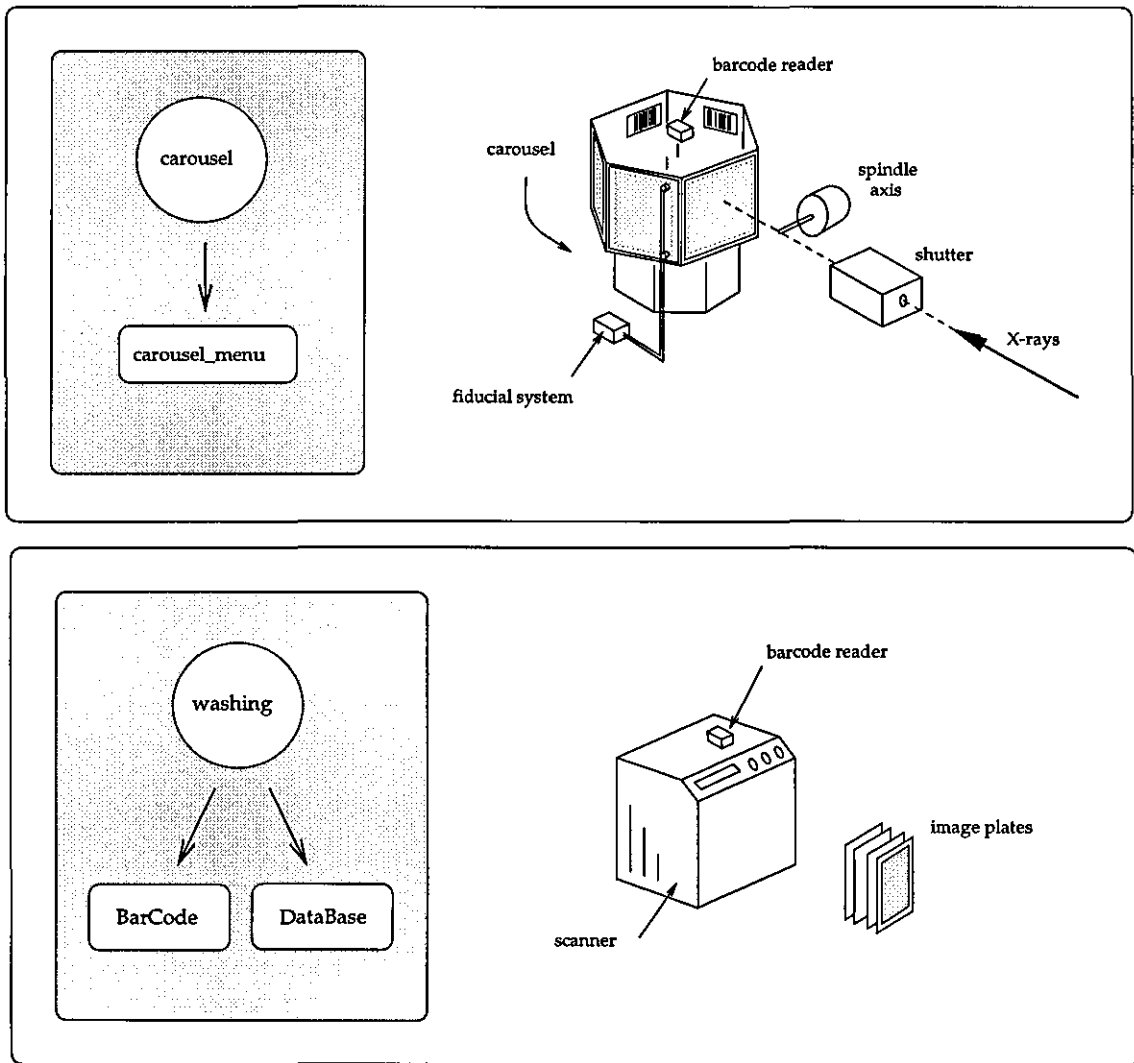


Figure 4.7: A schematic representation of the tasks performed by the control software

The foundation of the control software is a database. The database stores information about each image plate; whether it has been exposed or scanned. In addition,

information concerning experimental parameters, such as X-ray wavelength, crystal to film distance and oscillation range, are recorded for each image of the dataset. The database is an essential component, allowing the software at the detector and scanner to operate in a coordinated manner.

Within the database there are two tables; the status table and the experiment table. These are shown in figures 4.8 and 4.9.

plNumber	eState	sState	eTime
1	0	1	-
2	1	0	847204124
3	0	1	-
4	0	1	-
5	1	0	847203978
6	1	0	847203865

Figure 4.8: The status table, containing four columns. The plNumber is the value of the barcode, the flags in the eState and sState columns take the value of 0 or 1. eTime is the exposure date.

imNumber	plNumber	headerFlag	filePrefix	startAngle	oscRange	expTime	wavelength	distance	decayT	date
1	2	1	tintin	0	1.25	45	0.9792	250	864	Tue Nov 12 14:35:11 1996
2	4	1	tintin	1.25	1.25	45	0.9792	250	786	Tue Nov 12 14:36:37 1996
3	1	1	tintin	2.5	1.25	45	0.9792	250	712	Tue Nov 12 14:37:51 1996
4	3	1	tintin	2.75	1.25	45	0.9792	250	645	Tue Nov 12 14:38:58 1996

Figure 4.9: Experiment table containing all important experimental parameters; file name, oscillation range and starting angle, exposure time, X-ray wavelength, crystal-to-detector distance and date.

The exposure state of an image plate is called the eState, and describes whether the plate has been exposed to X-rays. The exposure state of every plate is recorded in the status table, in the eState column, see figure 4.8. Prior to any exposure, the eState takes the value of *zero*, signifying that the plate is blank. Once the image plate has been exposed, the exposure state assumes the value *one*. In a similar way, the scan state of every image plate describes whether the plate has been read by the off-line scanner. Again, the scan state of every plate is recorded in the status table,

in the sState column. A value of *zero* in the sState column indicates that the image plate has not been read.

The values of the exposure state and scan state are of primary importance, since they determine the actions taken by the control software. This is clearly shown in the block diagram in figure 4.10.

If, prior to an exposure, the software discovers that the image plate has not been scanned, then the exposure is aborted, the carousel rotated and the next plate checked. If, on the other hand, when preparing to scan an image plate, the software finds that the plate has not been exposed, it prints a warning message.

At the same time as setting the exposure state to *one*, the date of the exposure is written in the column labelled eTime. This information is expressed in terms of the number of seconds elapsed since the 1st January 1970, the American Standards Institute (ANSI) standard for the C-programming language. This information is stored temporarily, and is re-read when the image plate is identified at the off-line scanner. In this way, an elapsed time between the exposure and scanning of the image can be calculated, and used in a correction of intensity decay.

The roles played by each program within the software suite can be described as follows;

Program	Description
<code>carousel</code>	The principal program which manages the detector. <code>carousel</code> interfaces directly with the beamline hardware, in order to control the carousel rotation, the spindle axis and the X-ray shutter. <code>carousel</code> also calls routines from the slave program <code>carousel_menu</code> via the intrinsic SPEC function <code>data_pipe</code> . <code>carousel</code> is written in SPEC-language

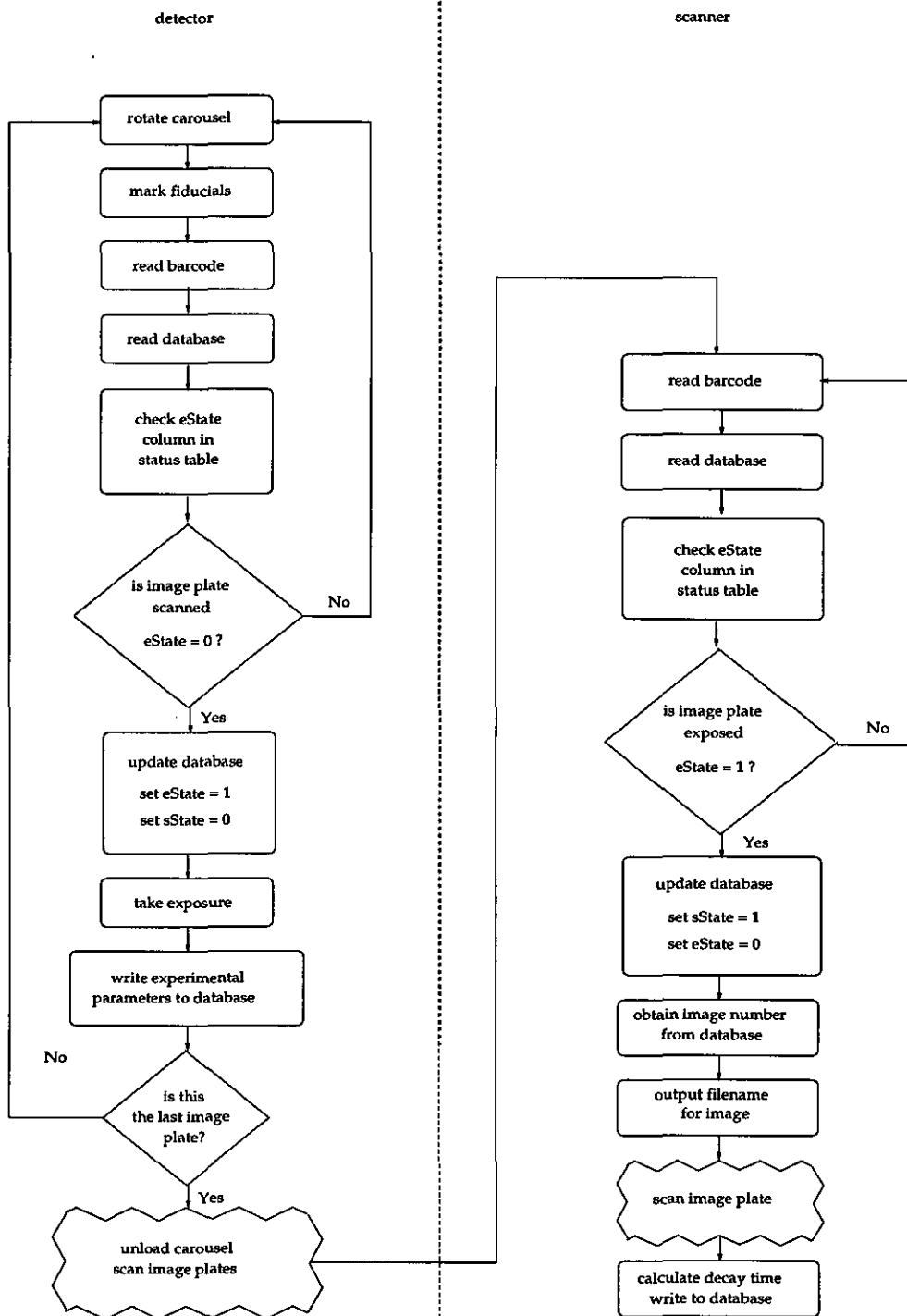


Figure 4.10: Sequence block diagram. The rectangular boxes represent actions, diamond boxes queries and the jagged boxes human intervention

carousel_menu This program is divided into six, self-contained routines, which are selected via a switch/case logic. These routines perform the following tasks; establishes a connection to the database, imports the serial line, reads the barcode, checks whether the plate has been erased, allocates the image number and writes all the experimental parameters to the database. **carousel_menu** is written in C-language.

washing The principal program which manages the activities at the off-line scanner. **washing** is responsible for defining the filename of the image files and calculating the elapsed time between exposure and scanning. The program performs the following tasks; imports the serial line, reads the barcode, checks that the image plate has been exposed, calculates the elapsed time, and defines the filename. This is achieved by calling external functions found within two secondary programs, **BarCode** and **DataBase**. **washing** is written in C-language

BarCode Contains the functions which import the serial line, and communicates with the barcode reader. **BarCode** is written in C-language

DataBase Contains the functions which communicate with the two tables in the database. **DataBase** is written in C-language.

Communication with the database is achieved from the C-programs using an application programming interface (API). The API functions are accessed by including the `mysql.h` header file in the C-program and then linking against the `mSQL` [55] library. The actual language used within the database is the structured query language, SQL. The source code of the programs is given in appendix B.

Hardware Control

The installation of the detector requires six connections, which are shown in figure 4.11.

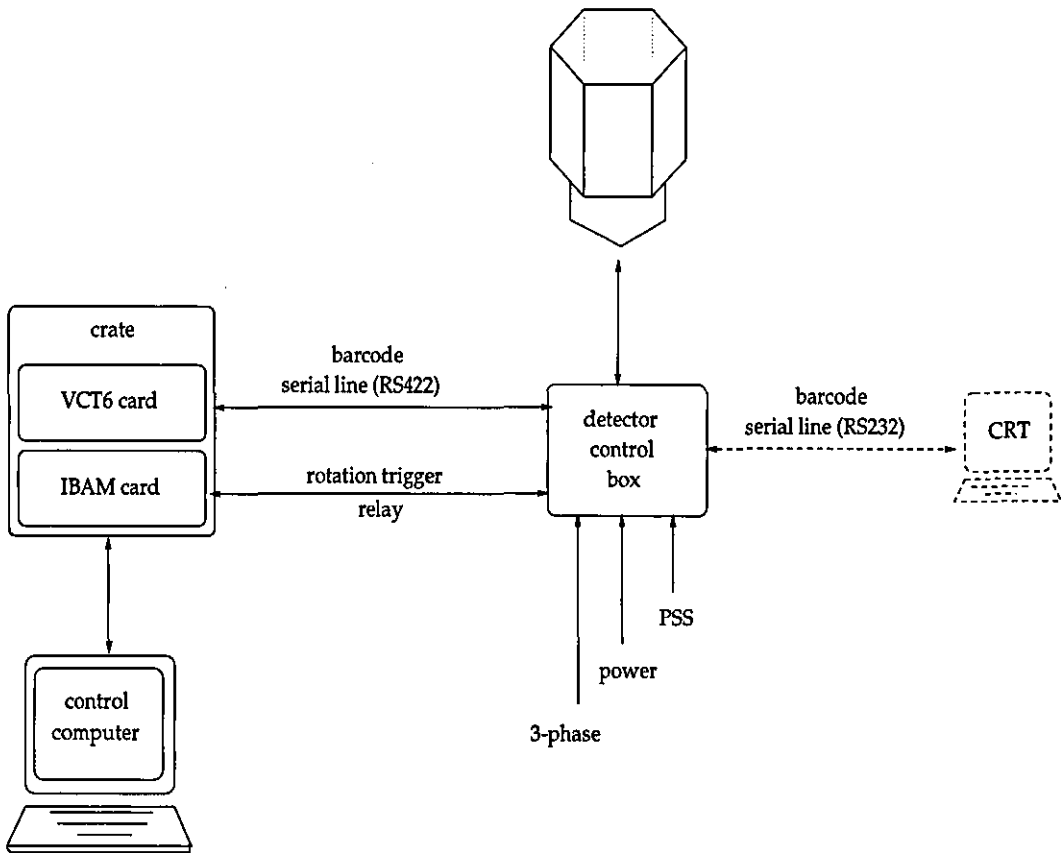


Figure 4.11: A schematic diagram showing the connections required by the detector; 3-phase electrical supply for the indexed motor, 220V electrical supply for the vacuum system, compressed air for the vacuum system, relay cable to trigger the rotation of the carousel, RS422 serial line for the barcode reader, personal safety system (PSS) for the security in the experimental environment.

The barcode reader can be configured manually, if required, by using a CRT via a RS232 serial line.

Information sent via serial lines can easily become corrupted when the serial line is

many metres in length. Given the large distance between the detector and the crate, it was considered necessary to use the serial line protocol RS422, in which the devices cable is more elaborately shielded.

The barcode readers are set up with the following parameters

Parameter	value
communication	RS422
baud rate	9600
data bits	7
stop bits	2
parity	even

Sample Data Collection

There are two interfaces to the control software; the first via the program `carousel`, the second via the program `washing`.

`carousel` requests certain information before beginning the data collection. This information defines all the experimental parameters, and is shown in figure 4.12.

The `washing` program interface is shown in figure 4.13.

The Database Browser

In order that the progress of an experiment can be readily checked, a program was written to display the information contained in the experiment table. This program, called `experiment_table_viewer`, is written in the scripting language PHP [56], thus allowing visualisation via a World Wide Web browser, as illustrated in figure 4.14.


```

SPEC> carousel

Database name: (dbase1)? dbase10
Filename prefix: (tintin)? tonton
Starting frame number: (1)? 1
Number of plates presently on carousel: (3)?
Total number of images: (3)?
Number of passes per image: (1)? 1
Exposure time per pass (s): (2)? 60
Starting angle (degrees): (21.25)? 0
Oscillation range (degrees): (1.25)? 1.25
Oscillation overlap (degrees):(0)? 0

-----

Start of data collection

current plate is 1
current image is 1
barcode is 30
image plate is clean
taking exposure
Pass 1:
Moving phi to 21.250000 (- 0.262000)
Moving phi to 22.500000 (+ 0.012000)
Tue Nov 12 14:35:11 1996

==> experimental parameters written to database

```

Figure 4.12: Example of the interface to the software, via the program carousel

```

opd14@b119 > washing

-----

washing program : reads barcode at the Fuji scanner

-----

Database name : dbase10
Give filename prefix : tonton
Introduce barcode

-----

the filename for this image is : tintin_001.IMG

-----

Ready for next plate
Hit c to continue, q to quit c

Program complete

```

Figure 4.13: Example of the interface to the software, via the program washing

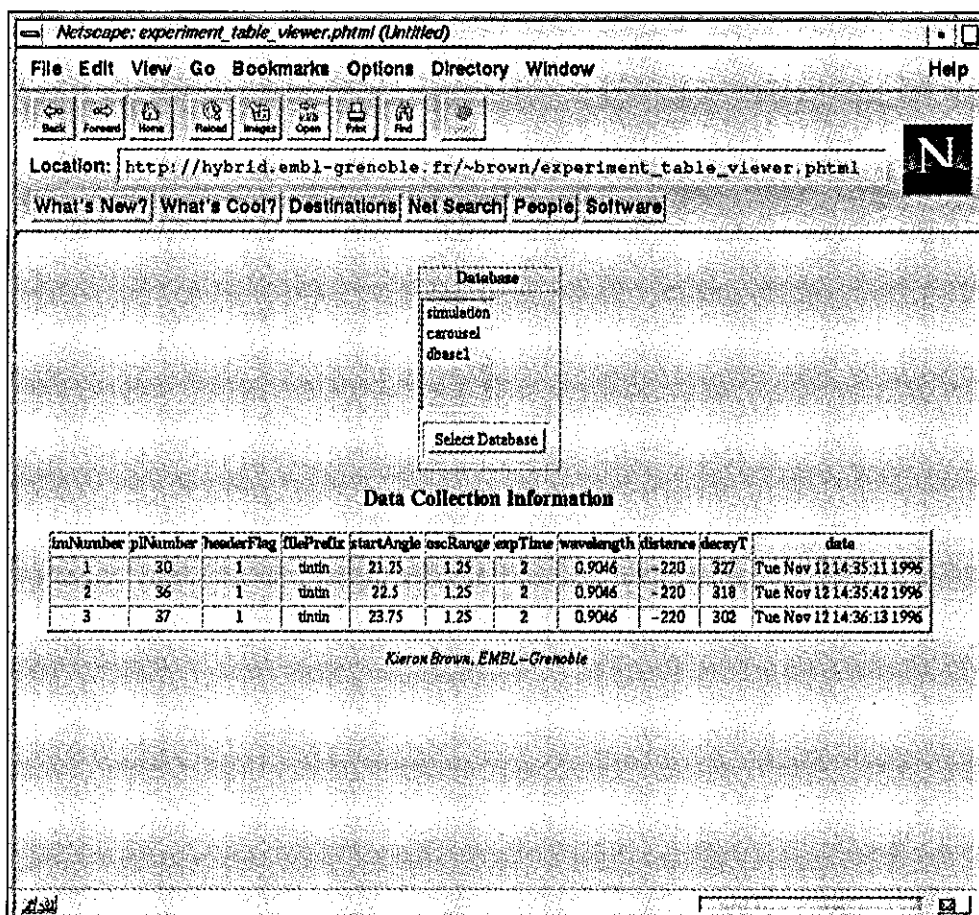


Figure 4.14: Database viewer, using a World Wide Web browser. Many different databases can be selected using the “Select Database” command. The contents of the experiment table are then displayed automatically.

Chapter 5

As a test of the fast image plate detector, a *multiple wavelength anomalous dispersion* experiment was performed, on the *ribonucleotide reductase free radical protein R2* (RNR). The structure of this 43.5kDa protein was successfully solved using the anomalous scattering from *mercury atoms* which were introduced into the native crystals. The results from the data analysis are presented, in which the anomalous scattering information was treated in a manner analogous to multiple isomorphous replacement. Based upon the results, various *aspects of the detector design* are discussed.

Performance of the Fast Image Plate Detector

5.1 Background to the Experiment

The observed intensity of radiation scattered from a sample is dependent upon both crystal and detector properties. By carefully selecting the crystalline sample, conclusions can be drawn about the performance of the detector system.

The intensity of the i th independent measurement of reflection hkl can be compared to the average intensity of all symmetry related reflections via the R_{sym} coefficient given in equation (5.1) [35, p. 285].

$$R_{sym} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)} \quad (5.1)$$

Some of the symmetry related reflections will fall on different regions of the same image plate. The remaining reflections will be recorded on different plates. The R_{sym} coefficient therefore offers a means of comparing the response of different components of the detector, and thereby determining the significance of the following potential errors

- uniformity of response
- reproducibility of the position of each image plate
- intensity decay of the latent image

However, the R_{sym} only gives reliable results if the gain of the scanning system has been correctly estimated. The scanner gain is a factor which converts the signal recorded by the detector into an equivalent number of absorbed X-ray photons. With a knowledge of its value the standard deviation of the observed intensities can be better estimated. Furthermore, outlying reflections can be identified and rejected.

The gain of the detector, g , is estimated by considering the variance of the background intensity within a region of the image. This region is assumed to contain a large number of pixels but be free of diffraction spots. The variance, σ^2 , is then compared to the mean intensity of the pixels within this region, \bar{I}_{pixel} , as shown in equation (5.2).

$$g = \frac{\sigma^2}{\bar{I}_{pixel}} \quad (5.2)$$

The calculation assumes that there is no correlation between pixels and that the observed intensities have a Poisson distribution. Miyahara [57] has made measurements which indicate that there is, in fact, significant pixel correlation when using image plates.

One potential problem associated with an off-line scanning system will arise from the variation in the position of successive image plates relative to the incident X-ray beam. The effect of this error will be most significant when considering partially

recorded reflections, since these can appear on many successive images. Ideally, the partial fractions from the different images should be added together, a procedure which is dependent upon the position of each fraction being the same. The significance of any positional errors introduced by the detector can therefore be judged through a successful partial summation procedure.

In order that the experimental results could be used to draw conclusions about the performance of the detector, care was taken over the choice of suitable sample. This choice was based upon five principal requirements;

- presence of heavy atom ions
- diffraction to high resolution
- medium unit cell dimensions
- long lifetime in the X-ray beam
- known structure

Heavy metal ions would be needed if a multiwavelength anomalous dispersion experiment was to be performed. High resolution diffraction and medium unit cell dimensions were seen as a requirement because this would take advantage of the large active surface area of the image plate. The lifetime of the sample in the X-ray beam would need to be long, since comparisons would be made of time-dependent parameters such as the latent image decay. Such comparisons would be invalidated by significant sample deterioration. Finally, selecting a sample with a known structure, would enable the electron density maps from a successful experiment to be compared with known results.

Ribonucleotide reductase (RNR) was considered to be a suitable choice of sample for three reasons; firstly, the crystals were known to be easily derivatised with mercury derivatives [58], secondly, reliable cryo-protectant conditions have been determined and the crystals are known to be resistant to X-rays, and finally, diffraction from the crystals of RNR has been observed beyond a resolution of 2Å [59].

The crystals of RNR were supplied by Par Nordlund (Stockholm University) and

are of the space group $P2_12_12_1$, with cell dimensions $a = 73\text{\AA}$ $b = 84\text{\AA}$ $c = 112\text{\AA}$.

5.2 Experiment and Results

An experiment was performed on BM14 at the European Synchrotron Radiation Facility (ESRF). A single crystal of a mercury derivative of RNR of dimensions $300\mu\text{m} \times 50\mu\text{m} \times 50\mu\text{m}$ was soaked for two minutes in cryo-protectant (comprising mother liquor and glycerol at a concentration of 30%). The crystal was then carefully mounted in a cryo-loop and flash frozen to 115K, using an Oxford Cryo-system.

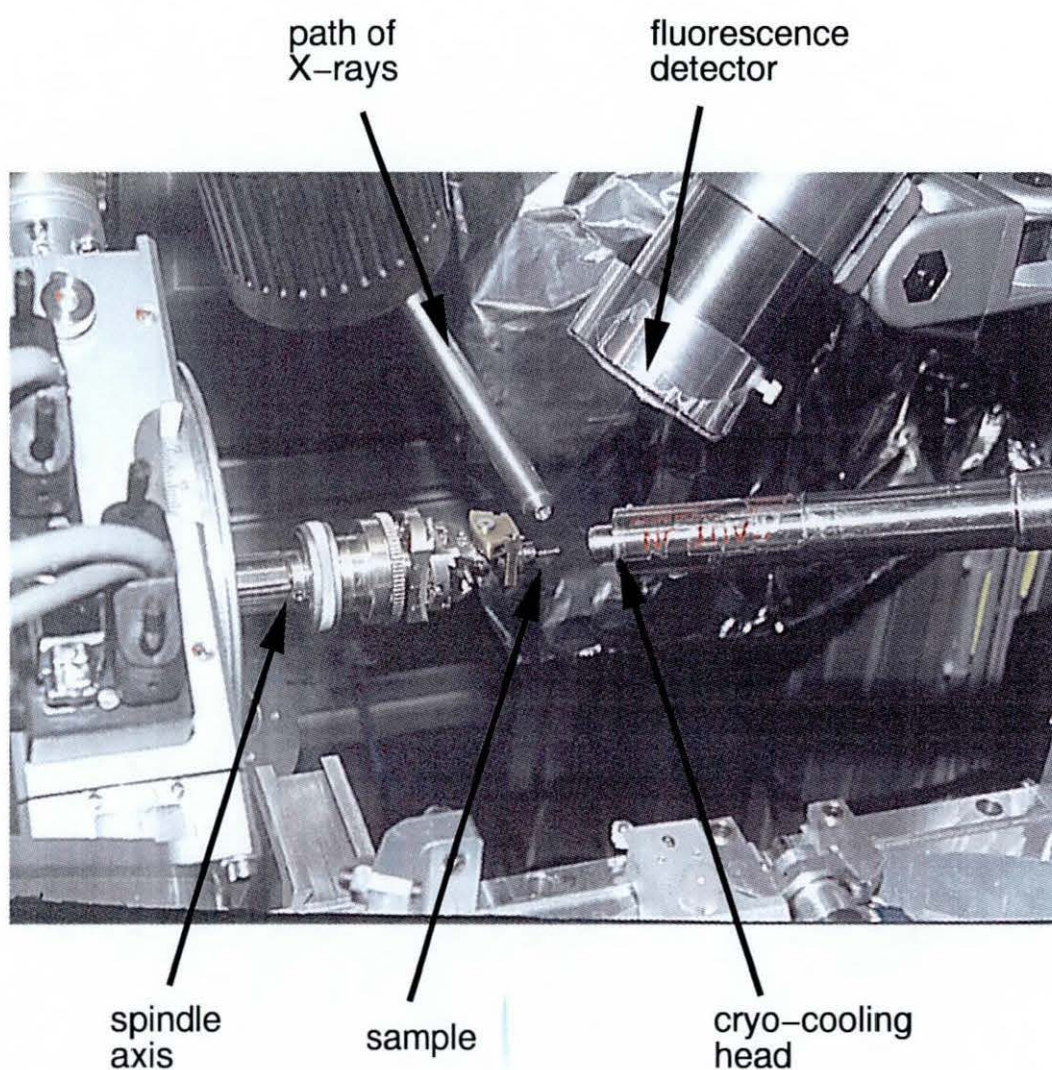


Figure 5.1: The experimental environment, showing fluorescence detector, cryo-cooling head and X-ray collimator. The carousel is in the foreground

The fluorescence from the crystal was measured as a function of incident X-ray energy around the mercury L_{III} absorption edge in order to determine the wavelengths to be used in the data collection. The fluorescence from the crystal is shown in figure 5.2.

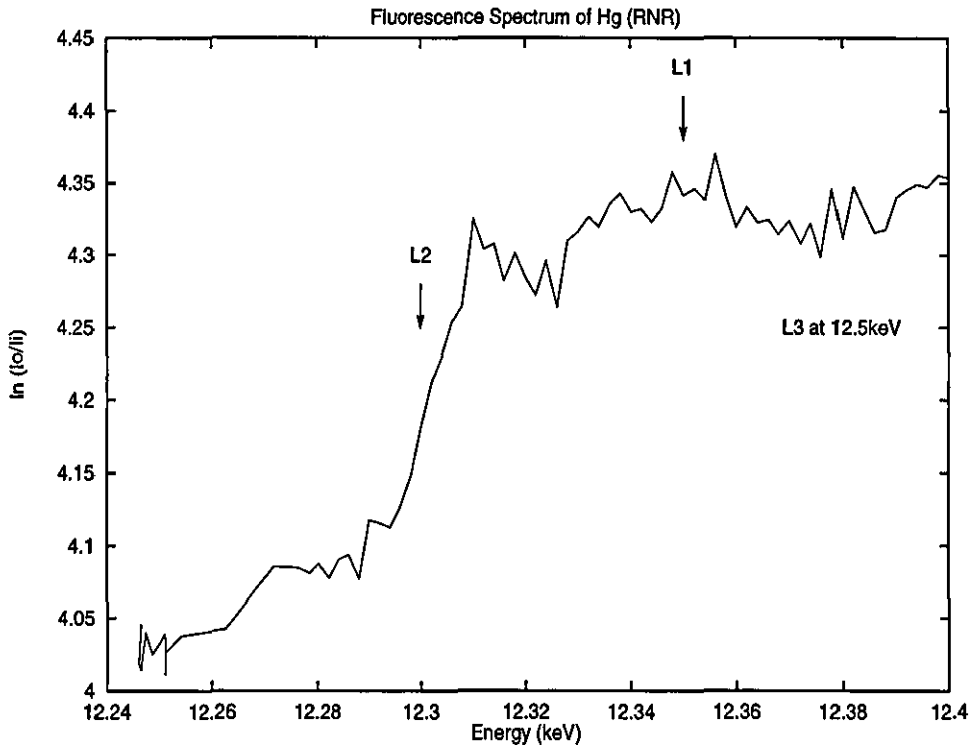


Figure 5.2: EXAFS scan Hg edge. The energies at which data was collected are marked.

Three wavelengths were chosen;

$$\lambda_1 = 1.0047\text{\AA} \quad \text{near, above edge}$$

$$\lambda_2 = 1.0080\text{\AA} \quad \text{inflection point}$$

$$\lambda_3 = 0.9916\text{\AA} \quad \text{remote, above edge}$$

These were chosen with the aim of maximising the anomalous difference between Bijvoet pairs at any one wavelength, as well as maximising the dispersive difference between identical reflections at different wavelengths.

Data were collected with the crystal in an arbitrary setting. Complete datasets were obtained prior to changing the incident X-ray energy. The datasets comprised 82.5° of data using oscillation range 1.25° , and exposure time of 30s.

Dataset	λ (Å)	Resolution (Å)	R_{sym} (%)	R_{anom} (%)	Completeness (%)	Redundancy
L1	1.0047	1.9	6.1 (20.9)	7.3 (16.6)	84	3.2
L2	1.0080	1.9	6.3 (24.7)	5.8 (17.8)	91	3.2
L3	0.9916	1.9	6.7 (24.9)	7.5 (19.7)	90	3.2

Table 5.1: MAD dataset processing statistics. $R_{anom} = \frac{\sum |<I+> - <I->|}{\sum <I>}$. $I+$, $I-$ are the Bijvoet intensities. The bracketed values are for 2Å resolution.

In order to give an estimate of the gain of the scanner a region containing about 1000 pixels was selected from an arbitrary image. This gave a value of the gain as 0.17. When repeating the measurement in a box of equal size translated by one pixel, the value obtained was 0.18.

Data Treatment

The data was indexed and integrated using the program MOSFLM [60], where it was seen that partial fractions from successive images could not be reliably added.

Data were scaled and reduced by the programs SCALA, AGROVATA and TRUNCATE from the program suite CCP4 [61]. In order to avoid losing the partial reflections from the analysis, they were scaled to give an equivalent 'full' intensity. The merging statistics are shown in table 5.1.

The data was treated in a manner analogous to multiple isomorphous replacement, see section 3.7. For this the data measured at λ_2 were treated as the 'pseudo-native', while the datasets at λ_1 and λ_3 , were treated as 'pseudo-derivatives' with anomalous scattering. This choice meant that the dispersive (pseudo-isomorphous) differences between native and the derivatives were always positive, see figure 3.11. Inter-wavelength scaling was then performed by SCALEIT. Likely limits for the maxi-

mum anomalous and dispersive differences were also calculated and the results used in subsequent calculations. The anomalous differences between Bijvoet pairs measured at λ_1 were then used as coefficients in a Patterson calculation. In addition, dispersive difference Pattersons were calculated between the different wavelengths, and attempts made to identify peaks present on both the anomalous and dispersive Pattersons. The anomalous difference Patterson maps for the Harker sections $u = \frac{1}{2}$, $v = \frac{1}{2}$ and $w = \frac{1}{2}$ are shown in figure 5.3.

From the general equivalent positions for this space group [62, p. 363], the Harker vectors were calculated to be as follows;

$$\begin{aligned} & \frac{1}{2} - 2x, -2y, \frac{1}{2} \\ & -2x, \frac{1}{2}, \frac{1}{2} - 2z \\ & \frac{1}{2}, \frac{1}{2} - 2y, -2z \end{aligned}$$

The grid position of each peak in the Patterson, consistent for site 1, see figure 5.3, was then used to obtain the real-space coordinates of the site. The co-ordinates and occupancy of the site were then refined against the Patterson map using program VECREF. Additional positional refinement was performed by MLPHARE. The one-electron model was used, as previously described, in order that the refined dispersive occupancy reflected the value of $\Delta f'$ between wavelengths, and the anomalous occupancy the value of f'' for each wavelength. The λ_2 dataset was taken as native in order that all the dispersive occupancies were positive. MLPHARE was run in two passes; during the first, only the centric reflections were used in order to obtain estimates of the 'pseudo-isomorphous' lack-of-closure errors, and dispersive occupancies. These occupancies were then fixed, in a second run in which the acentric reflections were included.

Using these initial phases, an anomalous Fourier was calculated, taking as coefficients the intensity difference between Bijvoet pairs at λ_1 . Additional sites found were refined in VECREF prior to being included in a new round of phase refinement.

A total of four mercury sites were found, whose coordinates are given in table 5.2.

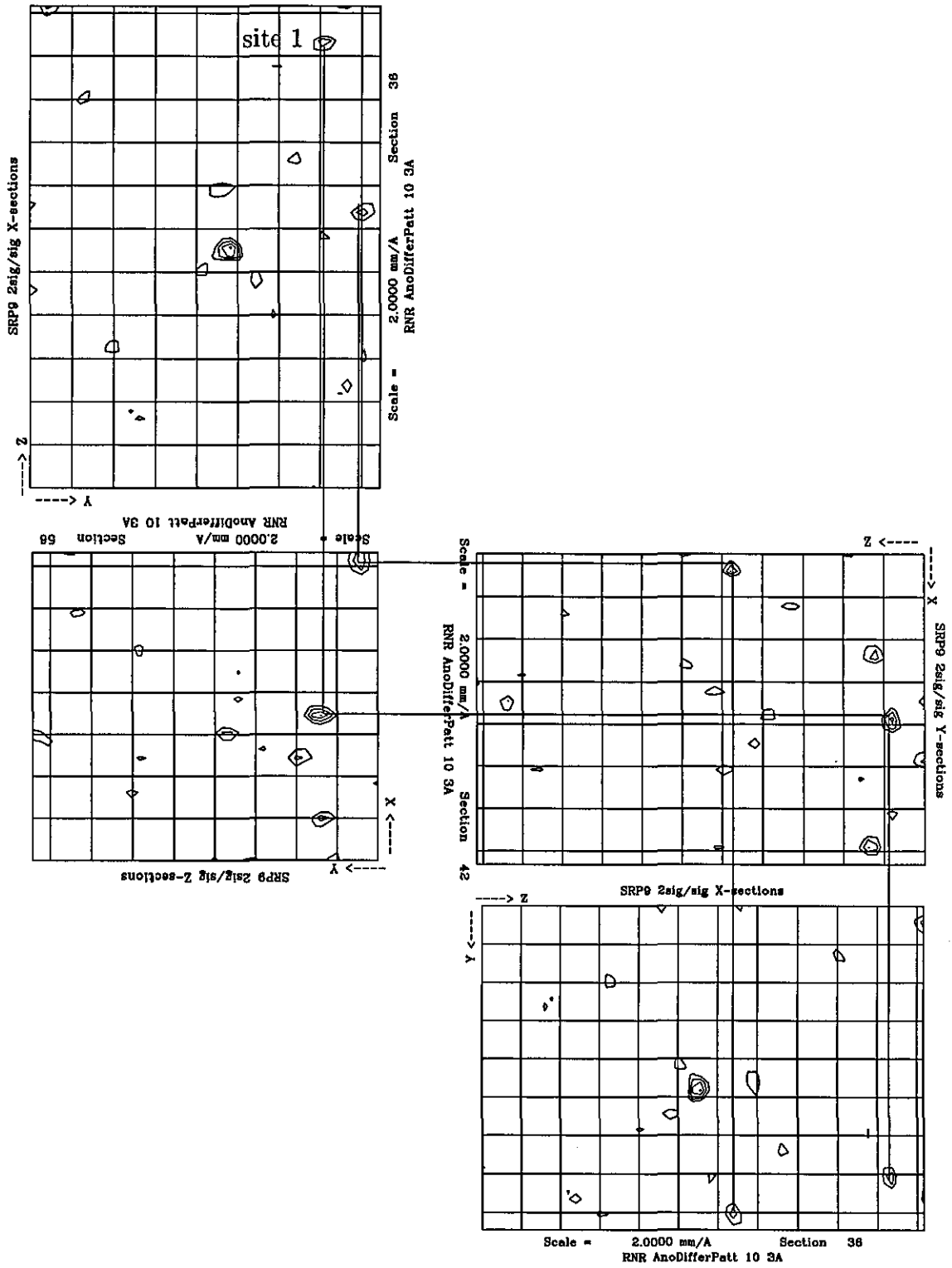


Figure 5.3: The $u = \frac{1}{2}$, $v = \frac{1}{2}$ and $w = \frac{1}{2}$ Harker sections of the anomalous difference Patterson, contoured with base level at 2σ in 0.5σ steps. The lines join consistent peaks, giving two possible sites.

site	coordinate			occupancy(λ_1)	
	x	y	z	disp.	anom.
1	0.3611	0.0476	0.2321	12.807	11.096
2	0.3760	0.1709	0.0159	7.536	10.408
3	0.1468	0.3884	0.3436	8.044	8.986
4	0.5000	0.4861	0.3566	6.509	9.355

Table 5.2: Coordinates of mercury sites found, with their anomalous and dispersive occupancies as calculated by MLPHARE for λ_1 dataset.

The figure of merit, anomalous Cullis R-factor, and phasing power [35, p. 287-288], were compared after the inclusion of each site in MLPHARE.

The improvement in the phasing statistics as additional sites are added is shown in table 5.3.

The phases were further improved by the technique of solvent flattening [63] and histogram matching [64] as implemented in the program DM. The Matthews number [65] indicated that there were 8 molecules in the unit cell, corresponding to a solvent content of approximately 38%.

The final phases calculated by DM were used in a Fourier calculation by FFT, taking the structure factor amplitudes from the λ_2 dataset. The electron density map of half the unit cell is shown in figure 5.4.

The electron density was viewed using the graphics program O [66], in which clear evidence was seen of secondary structure, as well as side-chain features. The polypeptide chain derived from the known structure was superimposed on the electron density, as a visual aid to judge the quality of the maps.

Figure 5.5 shows a typical view of the electron density, with an α -helix superim-

sites	FOM	$acenP_{power}$		$R_{Cullis}(\lambda_1)$		
		$\lambda_1 : \lambda_2$	$\lambda_3 : \lambda_2$	centric	acentric	anom
1	0.38	0.56	0.30	0.90	0.97	0.82
1, 2	0.42	0.63	0.34	0.88	0.95	0.77
1, 2, 3	0.44	0.73	0.36	0.83	0.93	0.75
1, 2, 3, 4	0.46	0.84	0.40	0.78	0.90	0.72

Table 5.3: The improvement of the MLPHARE phasing statistics at 2.5\AA , after the inclusion of additional sites.

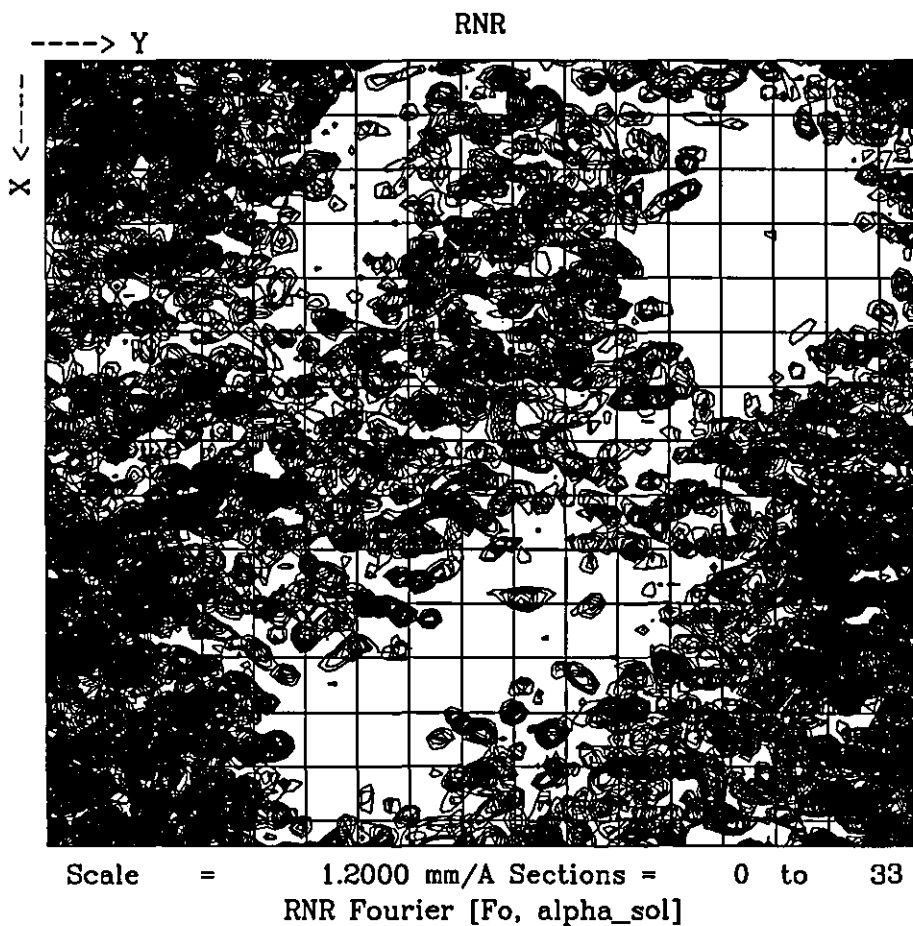


Figure 5.4: Electron density map of asymmetric unit, at 2.5\AA after solvent flattening.

posed as an aid to the eye. Figure 5.6 shows the di-iron site, with clear density around the two iron ions, whose coordinates are taken from the solved structure.

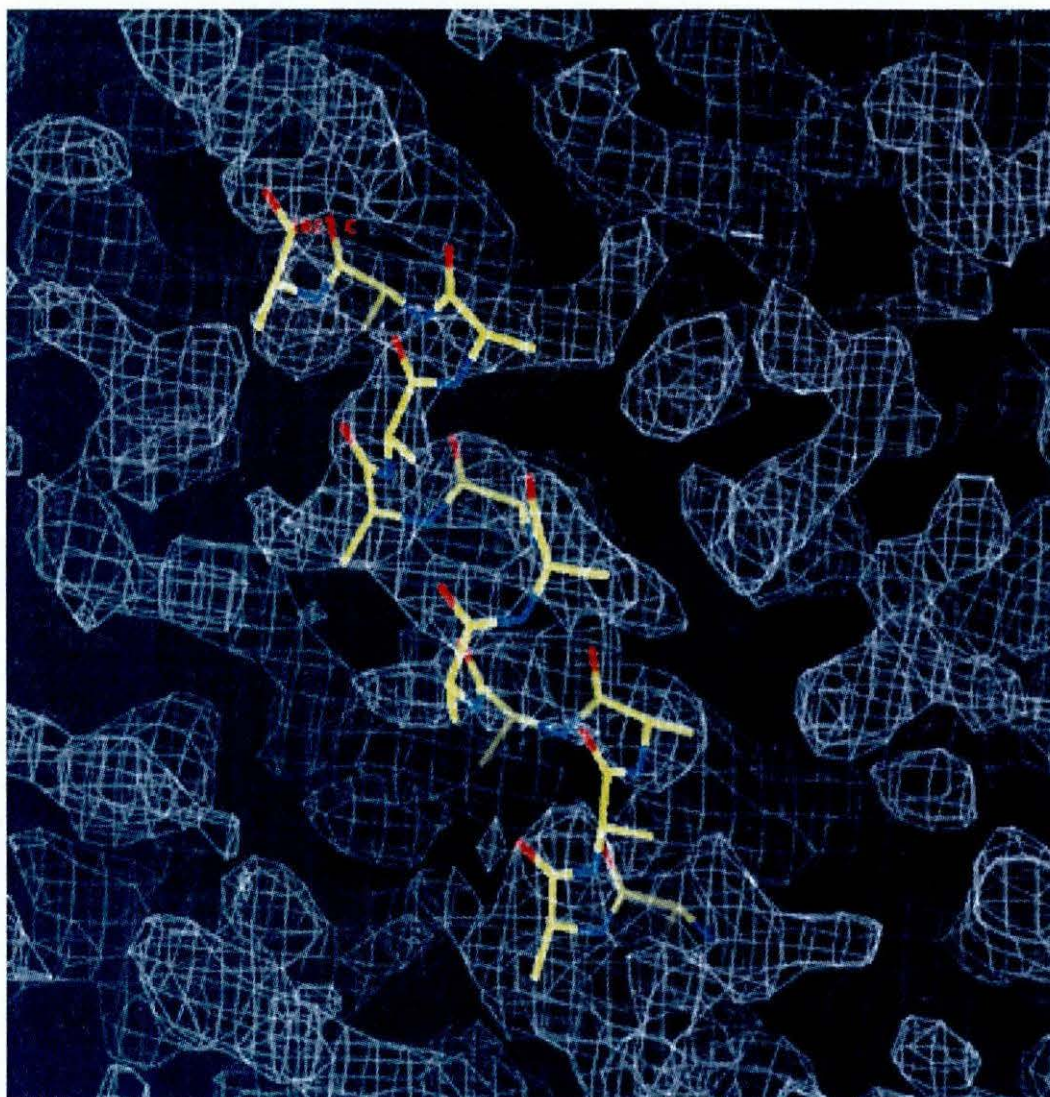


Figure 5.5: General view of electron density map, with an α -helix from the MIR structure superimposed

The map correlation coefficient calculated by OVERLAPMAP [61] between the solvent flattened MAD map, and the map calculated from the refined structure was 0.336. Figure 5.7 shows a plot of correlation coefficient against section number, and shows that some regions of the maps are more correlated than others.

The mean phase error, calculated by PHISTATS [61], was 69.127° .

Despite the low overall correlation coefficient, the experimental and solvent flat-

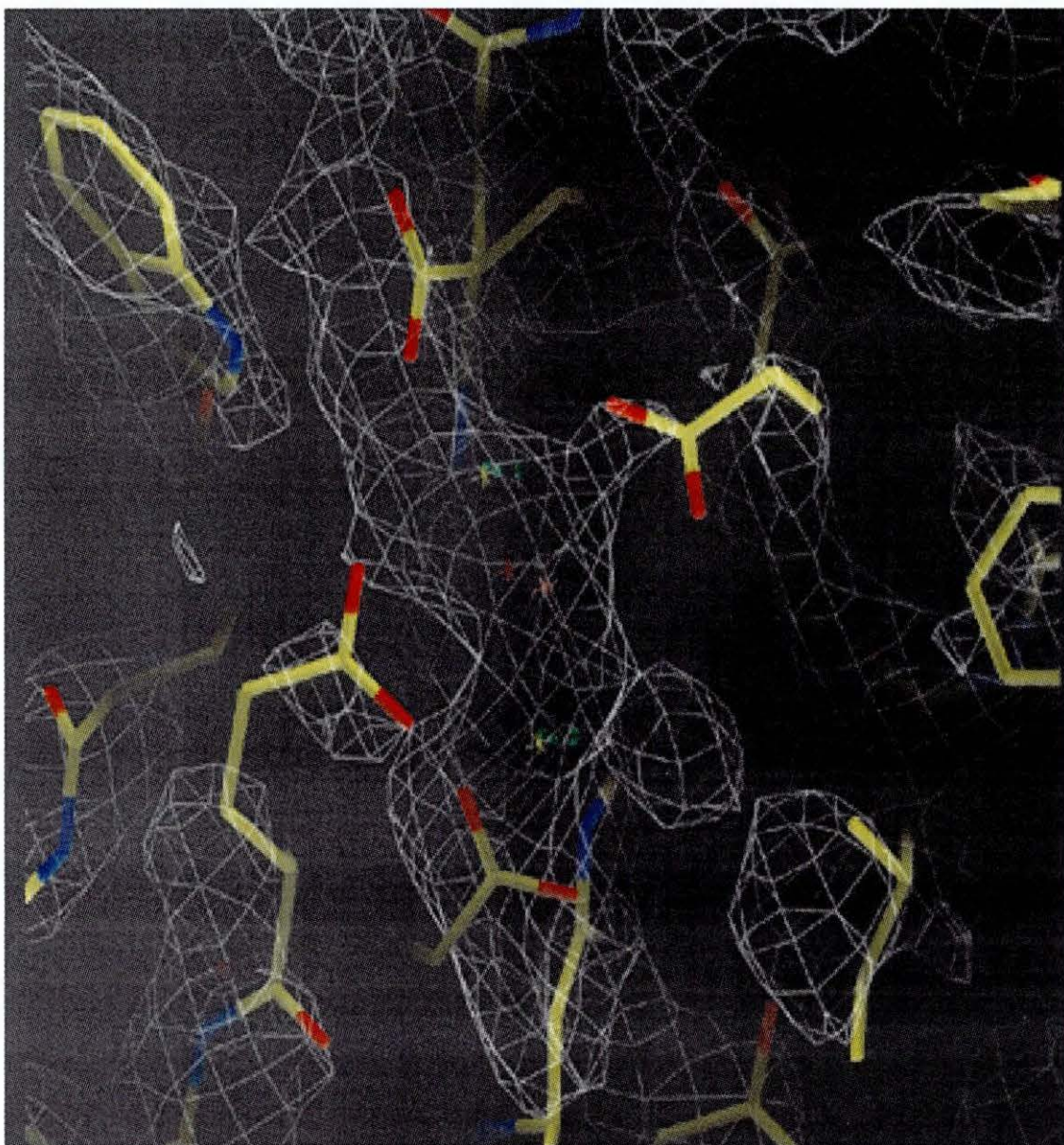


Figure 5.6: Electron density map for the region around the di-iron site. Density for the iron ions is visible. The Fe coordinates are taken from the MIR structure.

Correlation Coefficient

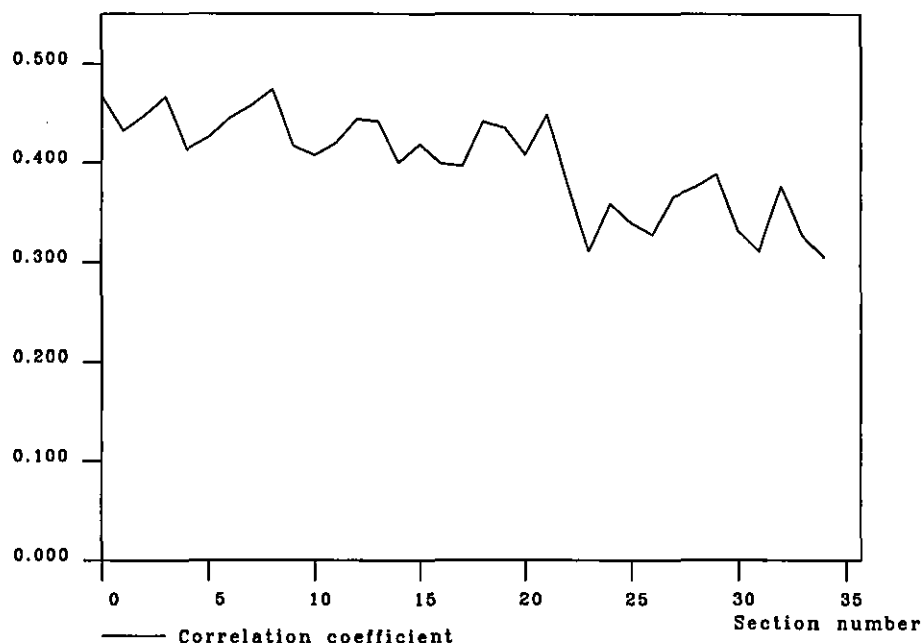


Figure 5.7: The map correlation coefficient between the solvent flattened MAD map, and the refined structure

tened MAD maps were readily interpretable.

5.3 Discussion

Previous experiments performed on the RNR crystals, using a MAR image plate detector have shown, that the overall R_{sym} of RNR data collected on a rotating anode to a resolution of about 1.8\AA were in the region of 6% [59]. The R_{sym} obtained from the present experiment has a similar value. Therefore, any errors introduced by the system appear not to significantly reduce data quality.

An important parameter requiring attention was the reproducibility of the image plate position. MOSFLM refines the horizontal and vertical offsets of each image, from which it was seen that there is a significant fluctuation of the position of successive image plates. The root mean square vertical offset was about $360\mu\text{m}$, and the horizontal offset $490\mu\text{m}$. This is shown in figure 5.8, where the dashed line represents the horizontal offset, and the solid line the vertical offset. From these results, it is

clear that the image plate position varies more in the horizontal direction than in the vertical direction.

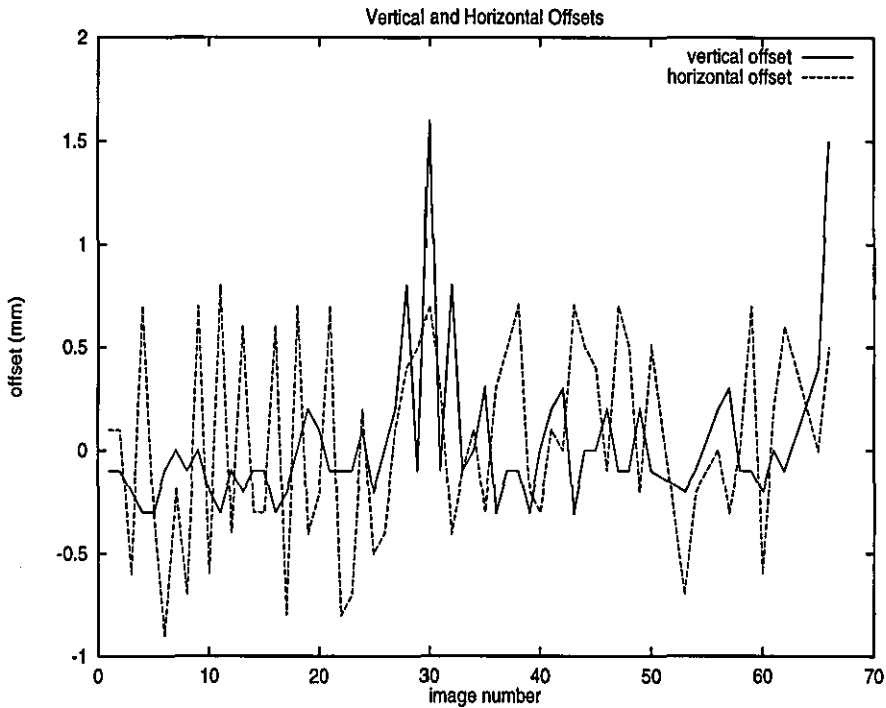


Figure 5.8: Vertical and horizontal offsets

An additional parameter of interest is the vertical angle of the image plate, arising from a rotation in the plane of the plate. This was determined using the position of the fiducial spots, and is shown in figure 5.9. The value can be seen to fluctuate by up to two degrees in an irregular fashion.

The explanation for these fluctuations in plate positions and angle could lie with the design of the image plate cassette, or the off-line scanner. However, there are several reasons why the cassettes are unlikely to be the cause of the error. Firstly, the cassettes are designed with guide rails at the sides, between which the image plate can slide freely up and down. This design allows the image plate to fall readily to the bottom of the cassette where it is supported by a third rail. The position of this support rail is known to vary by only $30\mu\text{m}$ [67]. Secondly, the position of the fiducial holes on each cassette is known to be the same to within $50\mu\text{m}$, hence the value of the plate angle obtained using the fiducial spot will represent only the image plate angle. Thirdly, the positional reproducibility of the cassette position when mounted

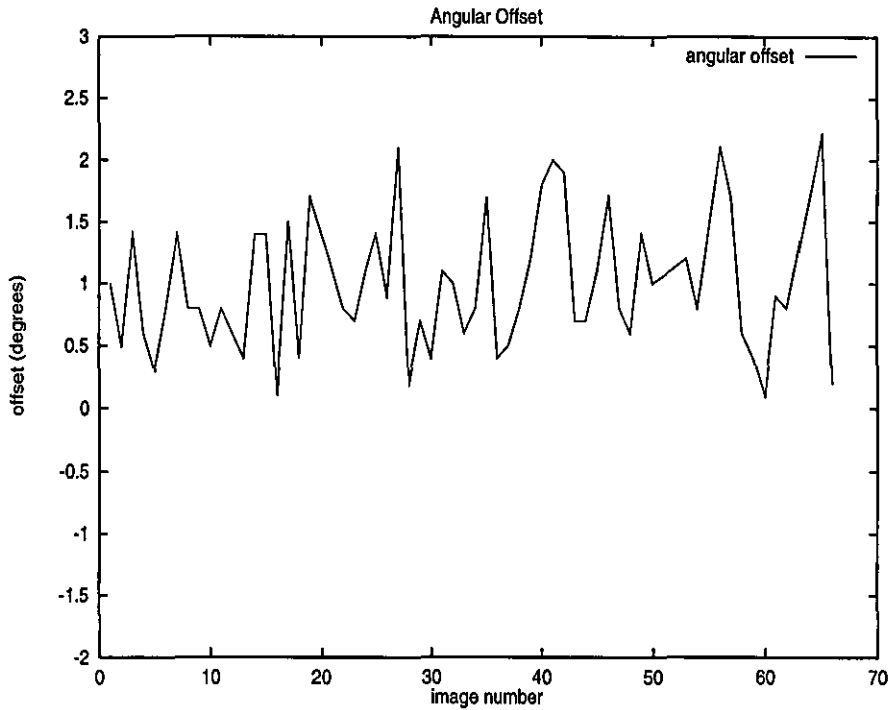


Figure 5.9: Angular offset

on each of the six faces of the carousel has been measured to be less than $50\mu\text{m}$. Taken together, it seems unlikely that the variation in the position arises from deficiencies in the cassette or detector design. Therefore, suspicion falls on the off-line scanner. After an investigation of the control mechanism, it appears that the scanner mounts the image plates in a non-reproducible manner during the scanning procedure.

Due to the absence of suitable software, the fiducial marks could not be used to correct for the angular mis-orientation of the image plate. The problems encountered when attempting to add partial reflections from successive images suggests that the use of the fiducial spots would have significantly improved the data quality, by avoiding the use of 'scaled partials'.

One potential disadvantage of an off-line scanning system is that the order in which the image plates are exposed can easily be confused while the plates are carried between the detector and the scanner. The experiment demonstrated that this problem has been eliminated by the use of the barcode identification system. The plates could be exposed and scanned in arbitrary order.

The barcode sticker played a second role, which was to insure that the image plates were not introduced into the scanner up-side-down; this simple feature meant that all the diffraction images were contiguous and correctly oriented.

The use of the image decay correction using the time period recorded in the database, was considered to be unnecessary. The batch-to-batch scales as applied by SCALA are plotted in figure 5.10, and show no evidence that the decay of the latent image is a significant problem. However, there is a smooth drop in scale factor for the first 15 images. This feature can be observed in the data from all three wavelengths, and might therefore be attributable to a systematic error, such as a variation in the absorption coefficient arising from nonuniform crystal morphology. An alternative explanation might be that of the decay of the incident X-ray beam.

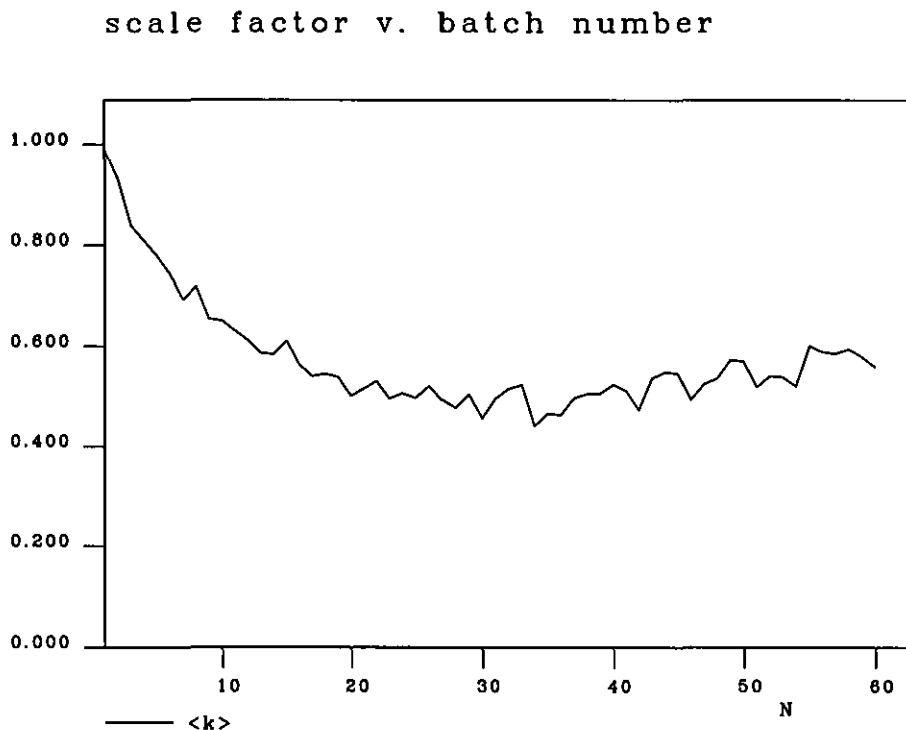


Figure 5.10: Scale factor for each batch

The detector was able to collect data up to a resolution of 1.7\AA at the X-ray wavelength of about 1\AA corresponding to the mercury L_{III} edge. This compares with the commercially available MAR detector, when used in 'large format'. The speed of data collection when using the fast image plate detector was higher than the MAR. Table 5.4 compares the carousel with the small and big MAR formats for a collection

detector	data collection time
carousel	3 hrs 40 mins
small MAR	4 hrs 24 mins
big MAR	5 hrs 50 mins

Table 5.4: Comparison of data collection times of the present experiment using the carousel and MAR in small and large formats. The cycle time of small MAR taken as 90s, large MAR 170s.

comprising 66 images with an exposure time of 90 secs.

When considering the map correlation coefficient, it is worth pointing out that the solvent flattened MAD map is being compared with the calculated MIR map from the final refined structure. For a fairer comparison, it would have perhaps been better to have used the original observed MIR map.

However, given the present result, despite the low overall correlation coefficient, the clarity of the electron density maps indicates that the fast image plate detector can be successfully used to perform multiwavelength anomalous dispersion experiments.

Chapter 6

So easy it seemed
Once found, which yet unfound most
would have thought impossible

John Milton

Attempts to solve the *murine-SRP9 molecular structure* are presented. X-ray crystallographic experiments were performed on the *native and selenomethionine-incorporated proteins*, after an *exhaustive search for heavy atom derivatives* proved unsuccessful. The *techniques of MR, MIR and MAD* were employed without success. This was attributed to the presence in the crystals of *twinning by merohedry*. During the late stages of this thesis however, the *selenomethionine incorporated human-SRP9 protein* was crystallised. Molecular replacement techniques were not successful so *multiwavelength anomalous dispersion* methods were used and the *structure solved*. *Initial views* of the structure are presented, and *tentative comparisons* made with the homologous structure of the heterodimer SRP9/14. Possible reasons for the difficulties encountered with the murine-derived protein are discussed.

Crystal Structure of SRP9

6.1 Biological Significance

The signal recognition particle (SRP) is found within the cytoplasm of eukaryotic cells, see figure 6.1 (after [68]), and plays an essential role in the sorting of proteins.

The roles performed by SRP are described schematically in figure 6.2, where the letters A to D highlight the principal stages.

SRP recognises a signal sequence [69] [70] at the N-terminal of the nascent protein

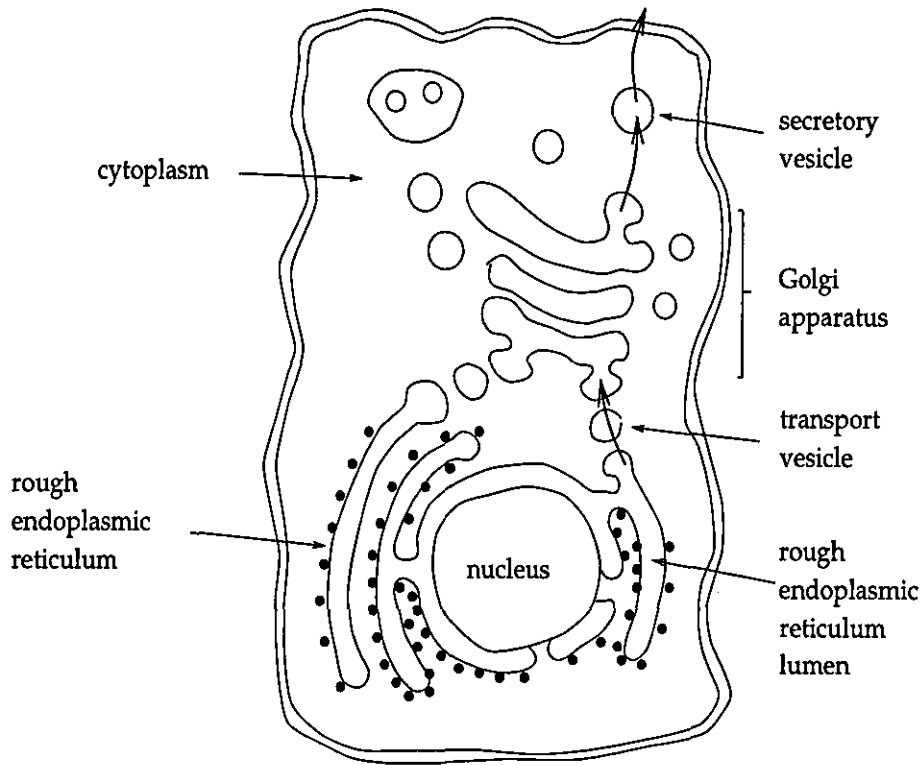


Figure 6.1: Schematic representation of the cell, showing the nucleus, cytoplasm, the rough endoplasmic reticulum (RER) and Golgi apparatus. The transport pathways of material from interior of the RER to the exterior are also shown.

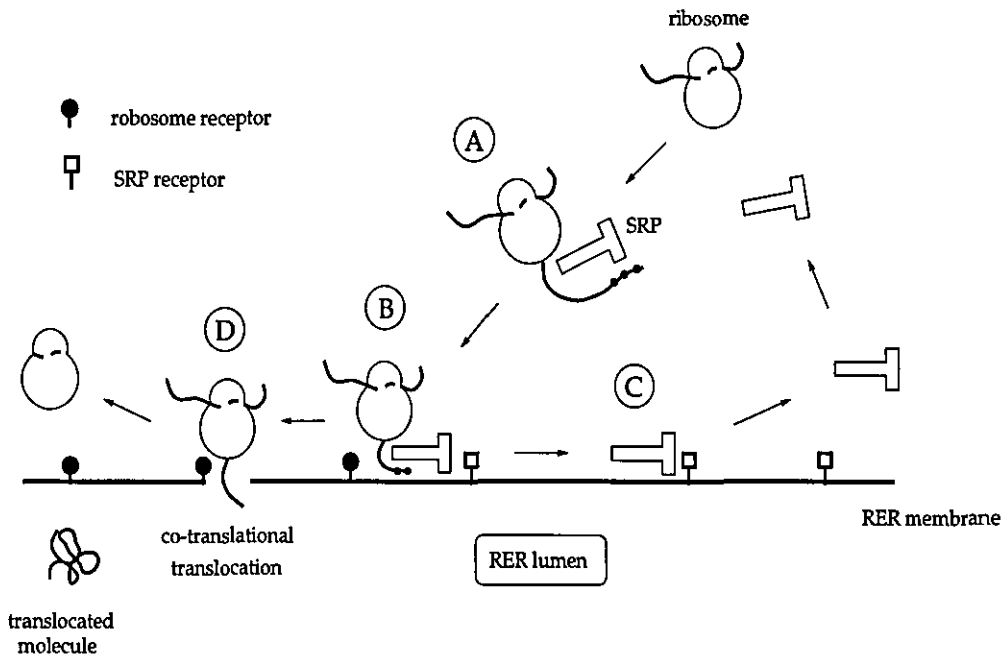


Figure 6.2: The SRP cycle, adapted from [70]

chains as they emerge from the ribosome, stage A in the figure. This signal sequence contains between 7 and 20 hydrophobic residues [71], and offers a means of identifying proteins which need to be translocated out of the cytoplasm. The interaction of the SRP with the nascent protein, triggers a pause in the synthesis by the ribosome (elongation arrest) [72]. The complex of ribosome, SRP and nascent chain then makes its way to the rough endoplasmic reticulum [73] (stage B). Here, it binds to the membrane surface via an interaction with the SRP receptor [74], which then releases the SRP from the ribosome (stage C). Protein synthesis resumes and the nascent protein is translocated to the RNR lumen (stage D). From here, the protein follows the secretory pathway, passing via the Golgi apparatus to the cell membrane.

Mammalian SRP consists of 6 proteins with molecular weights of approximately 72, 68, 54, 19, 14 and 9kD [75] and one molecule of RNA of about 300 nucleotides [76], referred to as the SRP RNA.

Electron microscopy has shown [77] that the signal recognition particle is rod-shaped, with a width of about 6nm and length of 24nm. The proteins are distributed in the particle in a manner shown in figure 6.3 [78] [79].

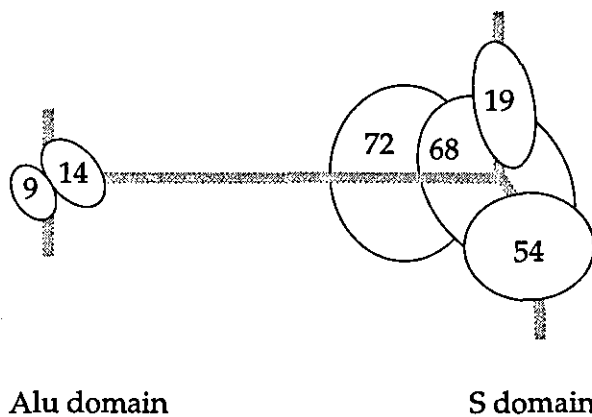


Figure 6.3: The signal recognition particle, comprising SRP9/14 and SRP68/72 heterodimers, and SRP54 and SRP19 monomers

In vitro studies have shown that the SRP9 and SRP14 proteins form a heterodimer [79], that binds tightly to the SRP RNA at sequences near the 3' and 5' ends (Alu domain). It is the SRP9/14 which is responsible for the pause in the synthesis of the nascent

protein by the ribosome. As yet, no explanation has been proposed for this interaction. The group of S. Cusack (EMBL Grenoble) is working in collaboration with the group of K. Strub (University of Geneva) with the ultimate aim of providing a structurally based description of this translation arrest mechanism. Crystallographic studies using proteins cloned and expressed by K Strub's group [80] are currently underway on the heterodimer SRP9/14, as well as the SRP9. During the course of this work, the structure of SRP9/14 was determined. A knowledge of the SRP9 structure will enable a more complete description of the SRP-ribosome interaction to be proposed.

6.2 Experiment and Results

Preparation of Native Crystals

The native protein of murine SRP9 (murine-SRP9) was purified by ion-exchange chromatography, by Ulrike Kapp at EMBL-Grenoble, using an expression clone constructed by K. Strub (University of Geneva) [81]. N-terminal sequencing [82] confirmed that murine-SRP9 contained 86 residues, giving a predicted molecular weight of 10167.

Crystals of the protein were grown by the hanging drop method in a precipitant solution of 28% ammonium sulphate, with 100mM Na-citrate pH 5.5 as buffer at 20°C. When using freshly purified protein stock solution, crystals appeared after a period of 3 months. It was seen however, that if crystallisation drops were setup using 3-month old protein stock solution, crystals appeared after only 3 days. In both cases, the crystals showed hexagonal-based, bipyramidal morphology, with typical dimensions $150\mu m \times 150\mu m \times 300\mu m$, see figure 6.4.

SDS-PAGE Gel Results

There is evidence that the molecule of SRP9 only crystallises in the dimeric form. An SDS-PAGE gel [83] of dissolved crystals, in the absence of reducing agent [84],



Figure 6.4: Native murine SRP9 crystals, showing hexagonal-based bipyramidal morphology

shows a band with a molecular weight close to that expected for the dimeric form, see figure 6.5.

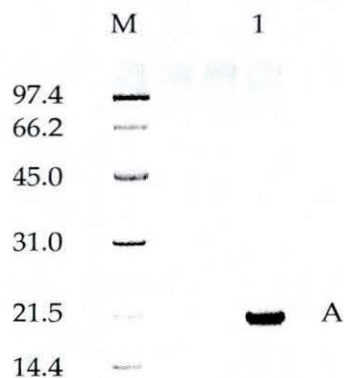


Figure 6.5: Silver stained SDS-PAGE, showing a clear band at molecular weight corresponding to twice that of the monomer. The marker weights are expressed in kDalton.

SDS-PAGE gel performed on the protein stock solution under non-reducing conditions, show that the freshly purified stock solution contains only the monomeric form, see lane 1, figure 6.6. However, after one month, a significant number of dimer molecules have formed, see lane 3. After 3 months, dimer molecules are the majority form, see lane 2. The crystallisation trials showed that crystals only appeared from this older stock solution.

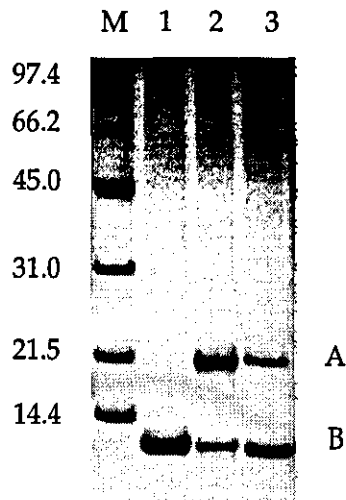


Figure 6.6: Bands marked with the letter A represent the dimer molecule, bands B the monomer. The marker weights are expressed in kDalton. Lane 1: fresh protein stock solution. Lane 3: 1-month old stock solution. Lane 2: 3-month old stock solution.

Mass Spectrometry Results

The native protein stock solution was analysed by mass spectrometry [85], a section of which is shown in figure 6.7.

The spectrum shows three clear peaks. The left-hand peak corresponds to a molecular weight of 20121, which is lower than predicted for the dimer molecule from the results from the sequence. The difference appears to arise from the fact that the molecule lacks the N-terminal methionine residue. The right-hand peaks correspond to a molecular weight of a dimer with the addition of a one and two molecules of molecular weight 77, respectively. This corresponds to the molecular weight of the reducing agent β -mercapto-ethanol (C_2H_6OS , hereafter referred to as β -ME), which was used during protein purification. β -ME binds covalently to cysteine residues [86], of which there are two (see figure 6.7). It therefore appears likely that up to two β -ME molecules have become bound to the protein.

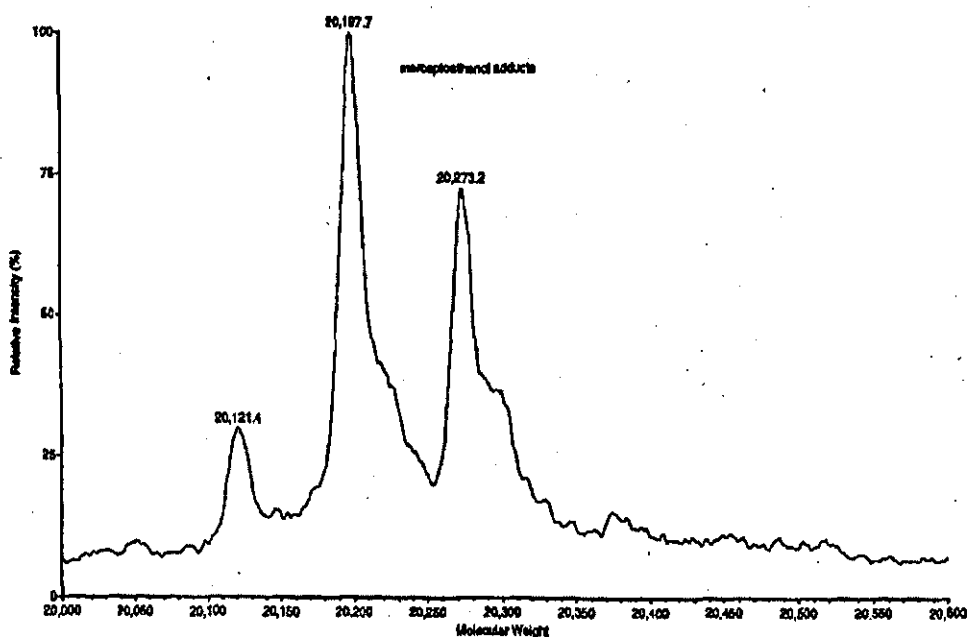


Figure 6.7: Section of the mass spectrum performed on 3-month old protein stock solution.

Preparation of Heavy Atom Derivatives

The native protein was crystallised in the presence of heavy metal ions under a variety of conditions as summarised in table 6.1.

As an alternative method of binding the heavy atoms to the native molecule, crystals were soaked in solutions containing the heavy metal ions. A summary of the soaking conditions used is shown in table 6.2.

Experimental Search for Derivatives

Experiments were performed on beamline BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble. Data were collected from both co-crystallised and soaked crystals, as well as from the native protein. All crystals were flash frozen to 110K, using a cryo-protectant solution comprising mother liquor and glycerol at a concentration of 30%.

The Laue class of the crystals was determined by consideration of the symme-

Metal Ion	Compound	Trials	pH
Ag	AgNO ₃	1/3/5x protein conc.	5.5
Au	AuCl ₄	1/10x protein conc.	5.5
	KAu(CN) ₂	1/5x protein conc.	5.0/5.5
Bi	BiNO ₃	1/5x protein conc.	5.0
		0.5/1/3/5x protein conc.	5.5
Cd	Cd-Acetate	1/5x protein conc.	5.0/5.5
Cd	CdCl ₂	1/10x protein conc.	5.5
Eu	Eu(NO ₃) ₃	1/5x protein conc.	5.0/5.5
Hg	HgCl ₂	0.5/1/3/5/10/20x protein conc.	5.0/5.5
	Thimerosal	1/3/5x protein conc.	5.0/5.5
	PCMB	1/10x protein conc.	5.5
	EMTS	1/10x protein conc.	5.5
I	KI/I ₂	0.5/1/5/10x protein conc.	5.5
In	InCl	1/10x protein conc.	5.5
Ir	NaIrCl ₆	1/5x protein conc.	5.0/5.5
Os	OsO ₄	1/10x protein conc.	5.5
Pb	(CH ₃)Pb-Acetate	1/5x protein conc.	5.0
		1/5/10/100x protein conc.	5.5
	Pb(NO ₃) ₂	1/10x protein conc.	5.5
Pd	KPdCl ₄	1/10x protein conc.	5.5
Pt	KPt(CN) ₄	1/10x protein conc.	5.0/5.5
	KPtCl ₄	1/10x protein conc.	5.5
Re	Re ₂ O ₇	1/10x protein conc.	5.5
Ru	Ru ₂ (OH) ₂ Cl ₄ .7NH ₃	1/10x protein conc.	5.5
	RuCl ₃	1/10x protein conc.	5.5
Sr	SrCl ₂	1/10x protein conc.	5.5
Tb	TbCl ₃	1/5x protein conc.	5.0/5.5
Tl	Tl-Acetate	1/5x protein conc.	5.0/5.5
U	UO ₂ (NO ₃) ₂	1/5x protein conc.	5.0/5.5
	K ₃ UO ₂ F ₅	1/10x protein conc.	5.5
W	Na ₂ WO ₄	1/10x protein conc.	5.5
Yb	YbNO ₃	1/3/5x protein conc.	5.0/5.5

Table 6.1: Summary of the heavy metal compound co-crystallisation trials. protein conc. represents protein concentration.

Metal Ion	Compound	Trials
Ag	AgNO ₃	0.1 mM, 1 hr
Au	KAu(CN) ₂	0.2 mM, 2-24 hrs
Bi	BiNO ₃	1 mM, 2-24 hrs
Cd	Cd-Acetate	1 mM, 2-24 hrs
Eu	Eu(NO ₃) ₃	1 mM, 2-24 hrs
Hg	HgCl ₂	1 mM, 1 hr
	Thimerosal	0.1/1 mM, 45mins
Ir	NaIrCl ₆	1 mM, 1 hr
La	LaCl ₃	100 mM, 40 hrs
Pb	(CH ₃)Pb-Acetate	10 mM, 1/15 hrs
Pt	KPt(CN) ₄	0.5 mM, ~2 hrs
U	UO ₂ (NO ₃) ₂	0.2 mM, 1 hr

Table 6.2: Summary of the heavy metal compound soaking trials

try of their diffraction patterns. The hexagonal arrangement of the diffraction spots indicated that the crystal lattice belonged to either the trigonal, hexagonal or rhombohedral systems. In order to identify the space group more precisely, systematically absent reflections were identified. This was achieved by processing the data in the rhombohedral space group R32, using MOSFLM [60]. It was seen that all missing reflections occurred only along the zone $[0\ 0\ 1]$, where every third reflection was observed.

A list of possible space groups taken from the International Tables [62, p. 46] is given in table 6.3, where the reflection criteria describing which reflections are systematically absent in each space group are shown.

The space groups from the rhombohedral system were eliminated since all reflections with the conditions $-h + k + l = 3n$, $h + l = 3n$ were observed.

The data was reprocessed in one of the remaining space groups from each of the point groups, and the intensity of symmetry related reflections compared using the R_{sym} reliability factor, as given in equation (6.1) [35, p. 285].

Laue class	Point group	Space group	Reflection condition
$\bar{3}$	3	P3 ₁	l = 3n
		P3 ₂	l = 3n
		R3	-h + k + l = 3n, h + l = 3n, l = 3n
$\bar{3}m1$	321	P3 ₁ 21	l = 3n
		P3 ₂ 21	l = 3n
		R32	-h + k + l = 3n, h + l = 3n, l = 3n
$\bar{3}1m$	312	P3 ₁ 12	l = 3n
6/m	6	P6 ₁	l = 3n
		P6 ₄	l = 3n
6/mmm	622	P6 ₂ 22	l = 3n
		P6 ₄ 22	l = 3n

Table 6.3: Possible space group choices relevant to proteins, based upon diffraction pattern symmetry and systematic absences.

$$R_{sym} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)} \quad (6.1)$$

Prior to the calculation of R_{sym} , a rejection criterion was imposed on the reflections such that those with an intensity differing greatly from the mean were discarded. This is formulated in equation (6.2), where $I_i(hkl)$ is the intensity of the i th independent reflection, $\overline{I(hkl)}$ the average intensity of all symmetry related reflections and $\sigma(I(hkl))$ its associated standard deviation. The rejection criterion is thus applied in terms of the number of allowable standard deviations between the intensity of the i th reflection and the mean.

$$\frac{|I_i(hkl) - \overline{I(hkl)}|}{\sigma(I(hkl))} > \sigma_{rejection} \quad (6.2)$$

where for the present case

$$\sigma_{rejection} = 4 \quad (6.3)$$

The number of reflections rejected in this manner was recorded. Since the designation of reflections considered as symmetry related differs in each point group, the number of rejected reflections offered a means of distinguishing possible point groups. Using the above criterion, the choice of point group was reduced to either 321 or 622.

A summary of the processing statistics for the native and potential derivative crystals is given in figure 6.8. It can be seen that the R_{sym} value for the native processed in space group $P3_221$ is 4.5%, whereas when processed in space group $P6_222$, R_{sym} becomes 19%. This significant deterioration in R_{sym} indicating clearly that the crystal belongs to the point group 321 and not 622

There were many crystals for which the R_{sym} does not differ appreciably for the data processed in the two point groups; examples are those crystals grown in the presence of $PbNO_3$, $InCl$, RhO_7 , $YbNO_3$ and $K_3UO_2F_5$. Attention is drawn to the $HgCl_2$ crystal, in which the c-axis length doubled, $c' = 2c$. It is therefore unclear which point group best describes the crystal symmetry.

In order to decide whether any of the crystals were derivatives, each 'derivative dataset' was compared to the native, individually. With both datasets processed in the same space group, the 'derivative' was scaled to the native using the program SCALEIT from the program suite CCP4 [61]. The R_{merge} reliability factor was then used to compare the intensity of reflections from the two datasets. R_{merge} is defined in equation (6.4) [35, p. 286], where I_{hkl}^{native} is the average intensity of all symmetry related reflections, hkl , in the native, and $I_{hkl}^{derivative}$ the same set of symmetry related reflections from the derivative.

$$R_{merge} = \frac{\sum_{hkl} | I_{hkl}^{native} - I_{hkl}^{derivative} |}{\sum_{hkl} I_{hkl}^{native}} \quad (6.4)$$

Care was taken to insure that the diffraction patterns from the two crystals had been consistently indexed. An indexing ambiguity arises in trigonal space groups because the a and b axes of the unit cell are equal in length, and cannot be distinguished. As a result, an observation may be indexed either as hkl or $kh-l$. It was seen that

Compound	Space Group	R_{sym} (%)	R_{anom} (%)	Complete (%)	Multiplicity	Resol ⁿ (Å)	Cell (Å)	R_{merge} (%)	Anom signal	Comments
native	P3 ₂ 21 P6 ₂ 22	4.5 19	- -	95 96	5 7	2.3	64.4, 110.3	-	-	sym. ops. → P3
BiNO ₃	P3 ₂ 21 P6 ₂ 22	8 18	6	56	3	3.6	63.9, 110.8	22	✓	
PbNO ₃	P3 ₂ 21 P6 ₂ 22	14 19	14	76	2	3.3	64.4, 110.4	26	✓	sym. ops. → P3 but SD distrib ⁿ bad
(CH ₃) ₃ Pb-Ac	P3 ₂ 21 P6 ₂ 22	4.9 18	3.7	79	2.3	2.4	64.4, 110.3	10	✓	sym. op./SD → P3
InCl	P3 ₂ 21 P6 ₂ 22	17 17	-	76	1.5	4	64.4, 110.2	42	-	poor data quality I/sigI ≈ 3
RhO ₇	P3 ₂ 21 P6 ₂ 22	7.8 9	-	87	2.4	3.2	64.2, 110.6	38	-	sym. ops. → P3
YbNO ₃	P3 ₂ 21 P6 ₂ 22	7.7 9.6	5.5 4.2	92 94	3.4 6	3.6	64.4, 110.4	29	✓	
SrCl ₂	P3 ₂ 21 P6 ₂ 22	3 13	-	64	2.8	2.3	64.2, 110.4	7	-	not P6
Na ₂ WO ₄	P3 ₂ 21 P6 ₂ 22	3.2 12	2.9	84	2.3	2.3	64.3, 110.3	10	✓	sym. op./ SD → P3
Ru ₂ Cl ₄	P3 ₂ 21 P6 ₂ 22	4.1 17	-	90	3.3	2.4	64.2, 110.4	9	-	sym. op./SD → P3
Hg-thim	P3 ₂ 21 P6 ₂ 22	3.2 12	2.0	91	4.2	2.8	64.3, 110.1	5	✓	not P6
PCMB	P3 ₂ 21 P6 ₂ 22	6.4 15.2	4.4 11.3	81 83	3.8 5.7	2.6	64.4, 110.9	14	✓	sym. ops. → P3
K ₃ UO ₂ F ₅	P3 ₂ 21 P6 ₂ 22	6.9 7.9	8.5 6.6	99 99	3.5 6.5	3.7	64.2, 110.5	52 81 *	✓	convincing P6
Th-Ac	P3 ₂ 21 P6 ₂ 22	7.3 12.5	7.0 11.6	85 98	2.2 3.1	3	66.3, 114.3	11	✓	cell dimensions!
KPt(CN) ₄	P3 ₂ 21 P6 ₂ 22	5.9 16	4.6	58	1.7	3.6	65.5, 112.5	7	✓	cell dimensions! not P6
KPtCl ₄	P3 ₂ 21 P6 ₂ 22	4.1 13	4.0 12	97 99	3.6 5	2.9	64.9, 110.2	20	✓	sym. ops. → P3
Eu(NO ₃) ₃	P3 ₂ 21 P6 ₂ 22	5.5 16.1	-	85 86	3.1 4.2	2.9	64.2, 110.1	11	-	cell dimensions! not P6
Cd-Ac	P3 ₂ 21 P6 ₂ 22	7.0 10.0	-	89 89	3.2 6	2.9	64.0, 110.4	28	-	probably P6
CdCl ₂	P3 ₂ 21 P6 ₂ 22	5.2 5.7	-	96	2	3.2	64.1, 110.7	34 12	-	probably P6
HgCl ₂	P6122	6.4	7.3	94	2.5	4.5	67.8, 219.6	-	✓	cell dimensions!

Figure 6.8: Summary of heavy atom derivative search

when comparing two datasets which had been indexed in an inconsistent manner, the R_{merge} assumed values greater than 50%. Upon re-indexing one dataset via the above transformation, using the program REINDEX [61], the R_{merge} fell significantly.

Consideration was given to the average value of R_{merge} in bins extending from low to high resolution. Comparisons were then made with 'typical' plots from known derivatives [87, p. 106]. The average R_{merge} for all the data from each crystal is given in figure 6.8. Only those crystals grown in the presence of BiNO_3 and $\text{KPt}(\text{CN})_4$ showed evidence of being derivatives, and were therefore chosen as candidates for solving the protein structure by multiple isomorphous replacement.

The squared isomorphous differences between the native and 'derivative' datasets were taken as coefficients in a Patterson function calculation, as described in section 3.3. This was performed by the program FFT [61]. Several Patterson maps were calculated using data from different resolution ranges, but no peaks were found using data from either crystal. This result indicated either that no heavy atoms had bound, or that the crystals were not isomorphous with the native.

Incorporation of Selenomethionine in the Native Protein

The incorporation of selenium into selected residues of a protein has been shown to be a reliable means of obtaining a heavy atom derivative [12] [88]. A selenomethionine incorporated protein was therefore chosen as a means of overcoming the difficulties encountered when preparing heavy atom derivatives.

A selenomethionine incorporated murine-SRP9 protein was prepared by Ulrike Kapp (EMBL-Grenoble), following a protocol suggested by Van Duyne [89]. Due to the increased sensitivity of the incorporated protein to oxidation, β -ME was used during the purification as suggested by Doublé [91].

Crystals were grown by the hanging drop method using the same conditions as those of the native protein. Crystallisation trials showed that the reducing agent dithiothreitol (DTT), prevented crystal growth.

As with the native protein, the length of time required for crystal growth was dependent upon the age of the protein stock solution. Crystals appeared after 3-months if the stock solution was fresh, and only 3 days if the solution was more than 3-months old. The crystal morphology was the same as that of the native.

Mass Spectrometry Results

The selenomethionine incorporated protein stock solution was analysed by mass spectrometry [90], the results of which are shown in figure 6.9.

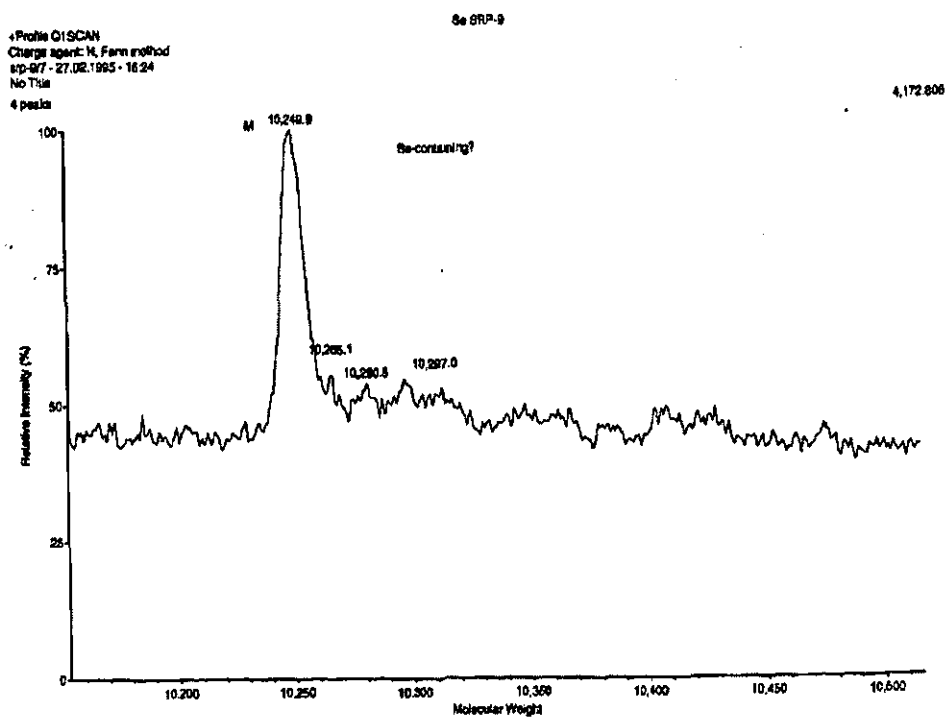


Figure 6.9: Section of the mass spectrum performed on freshly purified, selenomethionine incorporated murine-SRP9 protein.

The peak has a molecular weight corresponding to the native protein with four sulphur atoms replaced with four selenium atoms.

This result is in agreement with the results of the N-terminal sequencing [82], in which the protein molecule was identified as having four methionine residues.

Preliminary Experiments

Preliminary experiments were performed on BM14 at the ESRF. Data was collected from crystals at room temperature and at low temperature, where the crystals were flash frozen to 110K using as cryo-protectant a solution comprising mother liquor and glycerol at a concentration of 30%.

In both cases, anisotropic diffraction was observed. An example of a diffraction pattern is shown figure 6.10, where the resolution is seen to extend to about 4Å in the direction corresponding to the c^* reciprocal axis, but only about 8Å in the direction, corresponding to the a^* or b^* axes.

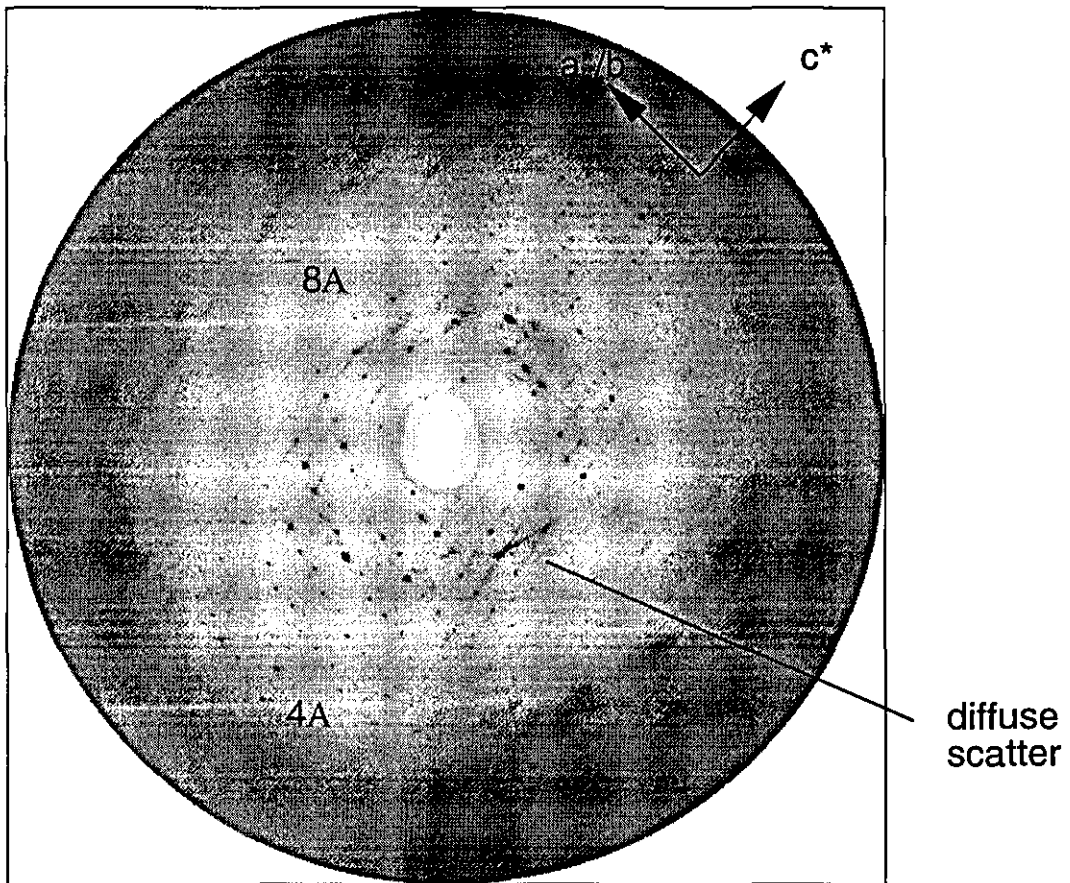
In addition, diffuse scattering is seen at low resolution, extending between some of the Bragg spots, and which takes the form of a crossed pattern.

Ytterbium Derivative of Selenomethionine Incorporated Protein

In addition to the above crystals, the selenomethionine incorporated protein was co-crystallised in the presence of ytterbium ions; 6.7mM and 67mM YbNO_3 was used, corresponding to 1 times and 10 times the protein concentration, respectively. This heavy metal ion was chosen since it has been shown that the L_{III} absorption edge of ytterbium leads to significant anomalous scattering effects [93].

A preliminary experiment was performed on BM14 (ESRF), in order to determine whether the ytterbium ions had bound to the protein. The crystals were flash frozen to 110K, using a cryo-protectant containing the mother liquor, 30% glycerol and heavy metal solution at the same concentration as in the crystallisation drop.

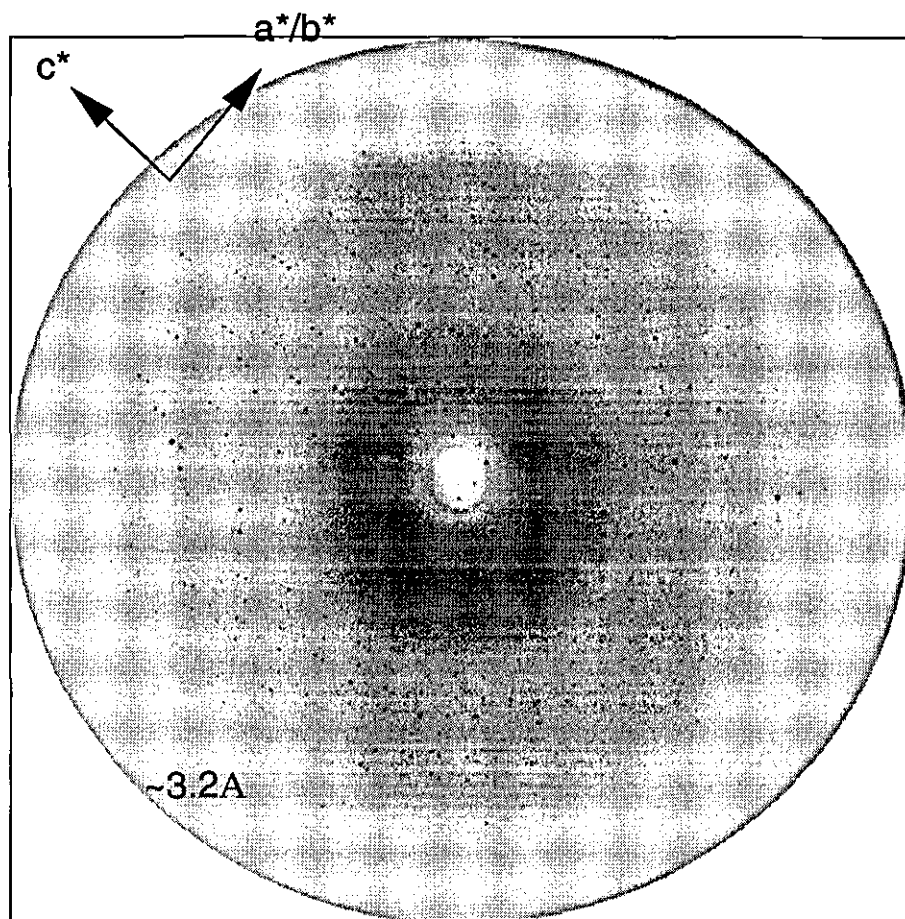
The X-ray fluorescence from a 100mM solution of YbNO_3 was first measured as a function incident X-ray energy in the region of the L_{III} absorption edge. Having thereby obtained an accurate estimation of the energy of the edge, the solution was removed and replaced with the crystal, and the fluorescence measured as a function of incident X-ray energy.



SRP9 SeMet (AS) BL19 ESRF

Figure 6.10: Diffraction pattern from frozen selenomethionine incorporated murine-SRP9 crystal, showing features of anisotropy and diffuse scattering.

No absorption edge could be observed, suggesting that the ytterbium ions had not bound to the molecule. However, upon taking an oscillation photograph of the crystal the diffraction was seen to be isotropic, as shown in figure 6.11.



SRP9 SeMet+Yb BL19 3/11/95

Figure 6.11: Isotropic diffraction pattern from a crystal of selenomethionine incorporated protein, grown in the presence of ytterbium.

In the absence of ytterbium, the selenomethionine incorporated protein appears to be less ordered in the a^*/b^* plane than the native protein molecule. However, when crystallising in the presence of ytterbium ions, the chemical environment was somehow altered with the consequence that the molecules were able to re-establish an ordered arrangement.

murine-SRP9 MAD Experiment

A multi-wavelength anomalous dispersion experiment was performed on BM14 at the European Synchrotron Radiation Facility, ESRF. A single crystal of the selenomethionine incorporated murine SRP9, co-crystallised with YbNO_3 was soaked for 30 seconds in cryo-protectant comprising mother liquor and 30% glycerol, before being flash frozen to 115K using an Oxford Cryo-system.

The X-ray fluorescence from the crystal was measured as a function incident X-ray energy from the selenium K-edge, with the aim of determining the wavelengths to be used in the data collection. The fluorescence spectrum from the crystal is shown in figure 6.12.

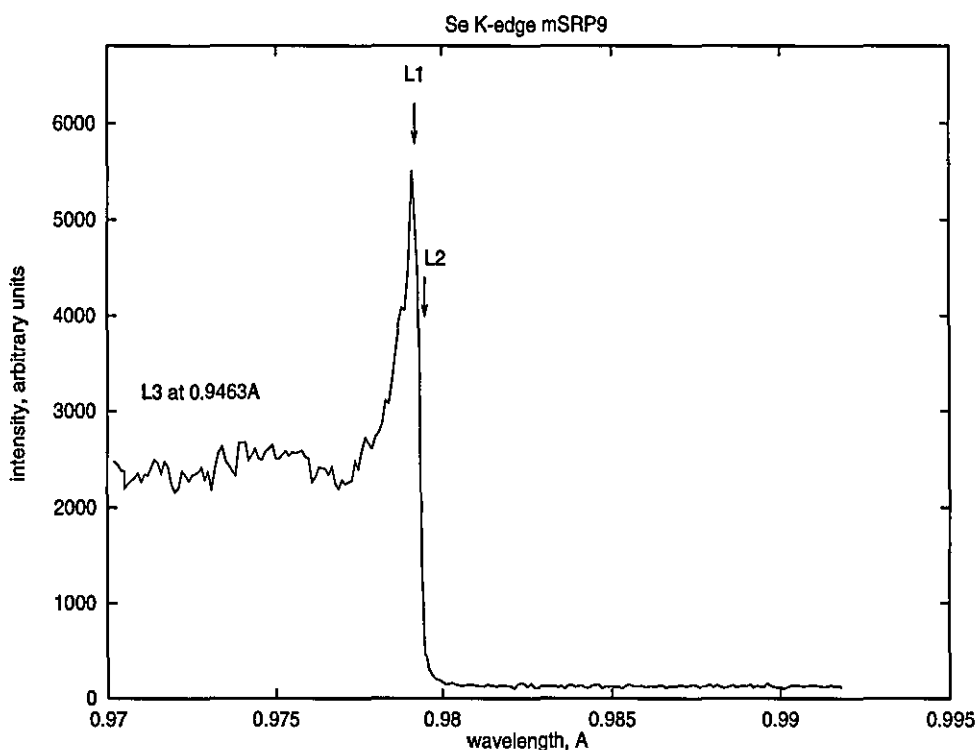


Figure 6.12: Fluorescence spectrum from Se K-edge. The wavelengths at which data was collected are marked.

Three wavelengths were chosen;

$$\lambda_1 = 0.9792\text{\AA} \quad \text{white line}$$

$$\lambda_2 = 0.9795\text{\AA} \quad \text{inflection point}$$

	λ_1	λ_2	λ_3
λ_1	9.5	2.0	4.4
λ_2	-	4.8	2.4
λ_3	-	-	5.6

Table 6.4: Theoretical anomalous and dispersive differences, as given by equations (6.5) and (6.6). The values are expressed in percent, taking theoretical values of f' and f'' from [53].

$$\lambda_3 = 0.9463\text{\AA} \quad \text{remote, above edge}$$

These wavelengths maximised the anomalous differences between Bijvoet pairs at any one wavelength, and the dispersive difference between identical reflections at the different wavelengths. The expected anomalous and dispersive differences were calculated using the equations (6.5) and (6.6), below [43], in which N_A is the number of anomalous scatterers, and M_W the molecular weight of the protein.

$$\Delta_{ano} = \sqrt{\frac{N_A}{2}} \frac{2f''}{\sqrt{(3.14 \times M_W)}} \quad (6.5)$$

$$\Delta_{disp} = \sqrt{\frac{N_A}{2}} \frac{|f'_{\lambda_1} - f'_{\lambda_2}|}{\sqrt{(3.14 \times M_W)}} \quad (6.6)$$

The results are given in table 6.4.

Data were collected with the crystal in an arbitrary setting, using the MAR Image Plate detector. Complete data sets were obtained prior to changing incident X-ray energy. The data sets comprised 80° of data using an oscillation range of 2.0° , and exposure time of 90s.

Dataset	λ (Å)	Resolution (Å)	Space Group	R_{sym} (%)	R_{anom} (%)	Completeness (%)	Redundancy
L1	0.9792	2.75	P3 ₁ 21	5.1 (19.9)	5.4 (11.8)	93	4.0
			P6 ₂ 22	8.3 (25.5)	5.3 (9.6)	95	6.8
L2	0.9795	2.75	P3 ₁ 21	4.3 (17.0)	4.2 (11.8)	89	2.4
			P6 ₂ 22	7.5 (20.7)	5.6 (13.3)	94	4.2
L3	0.9919	2.9	P3 ₁ 21	5.0 (16.6)	4.0 (10.4)	92	3.4
			P6 ₂ 22	7.6 (19.9)	4.4 (7.9)	94	6.2

Table 6.5: MAD dataset processing statistics, in space groups P3₁21 and P6₂22. R_{anom} is defined as $\frac{\sum |<I+> - <I->|}{\sum <I>}$

Data Analysis

After following the same protocol as described in section 6.2, the data were processed in space groups P3₁21 and P6₂22 using MOSFLM [60]. In an attempt to overcome potential indexing ambiguity in the trigonal space group, the orientation matrix of the crystal was obtained from the first dataset collected and imposed on the other data sets.

Scale factors within each wavelength were calculated by SCALA [61], and applied by AGROVATA [61], before averaging symmetry equivalent observations.

The merging statistics are shown in table 6.5 for the two space groups, where it can be seen that the R_{sym} factor is not significantly different between the two space groups. However, when processing in P6₂22, a large increase in the number of reflections failing the rejection criterion was seen. For this reason the space group was chosen to be either P3₁21 or P3₂21.

Structure factor calculations were performed by TRUNCATE [61].

By comparing the sign of the anomalous difference between a Bijvoet pair measured at different wavelengths, it was possible to verify that the data at each wavelength had been consistently indexed. This was implemented in the program GETSIGN, given

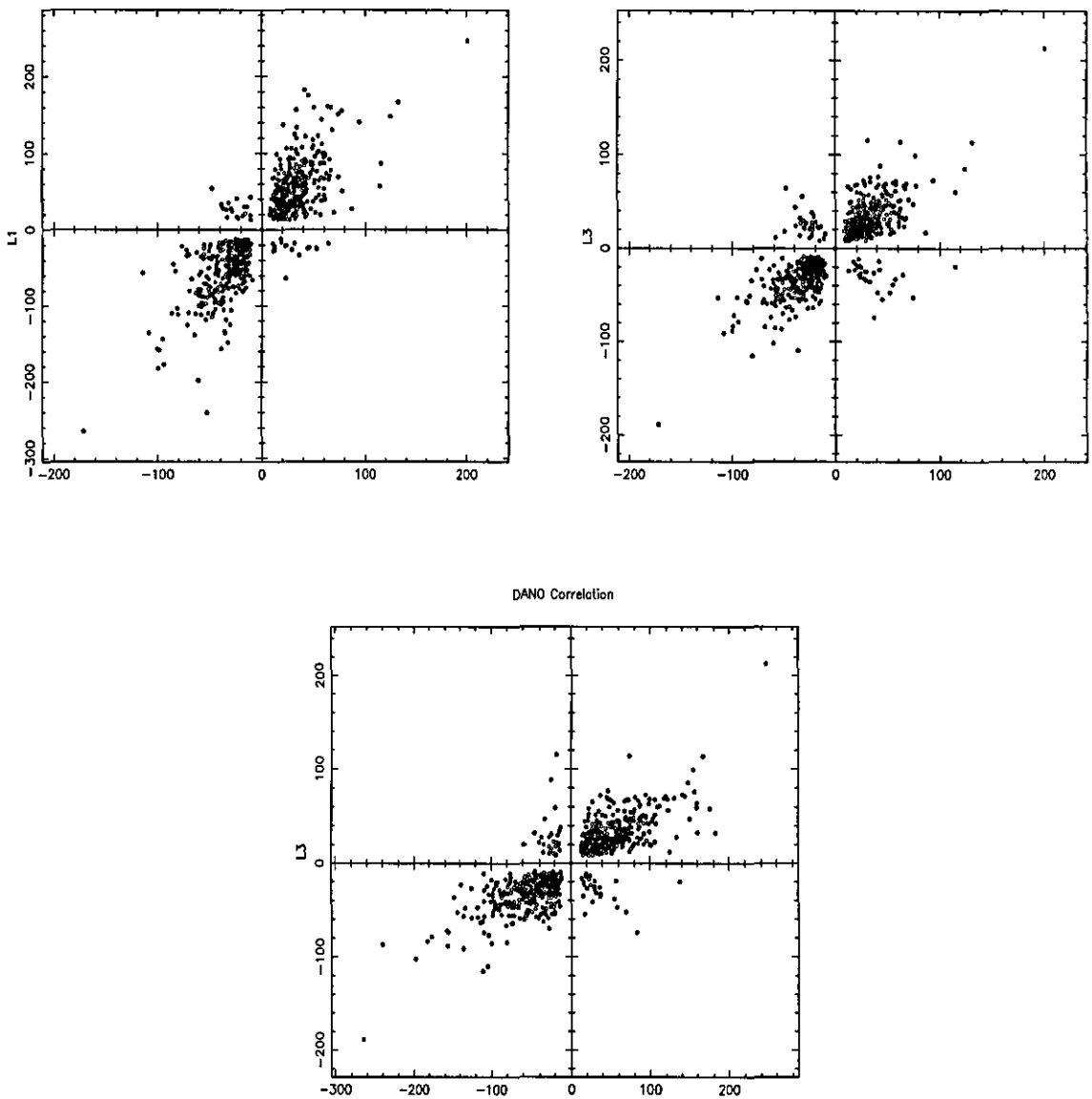


Figure 6.13: DANO correlation for each combination of wavelengths. Top left; $\lambda_1 : \lambda_2$, top right; $\lambda_3 : \lambda_2$, bottom; $\lambda_3 : \lambda_1$

in appendix A. The sign of the Bijvoet difference is conserved in datasets indexed in a consistent manner. The program represents this information in a graphical form, which is shown in figure 6.13.

The data was treated in a manner analogous to multiple isomorphous replacement, see section 3.7. For this the data measured at λ_2 were treated as the 'pseudo-native', while the datasets at λ_1 and λ_3 , were treated as 'pseudo-derivatives' with anomalous scattering. This choice meant that the dispersive (pseudo-isomorphous) differences between native and the derivatives were always positive, see figure 3.11.

Dataset	λ (Å)	Resolution (Å)	Space Group	R_{sym} (%)	R_{anom} (%)	Completeness (%)	Redundancy
L1	0.9792	2.75	P3 ₁	4.8 (16.3)	5.7 (16.8)	89.8	2.3

Table 6.6: Merging statistics for λ_1 dataset processing in the space group P3₁. R_{anom} is defined as
$$\frac{\sum |<I+> - <I->|}{\sum <I>}$$

Inter-wavelength scaling was then performed by SCALEIT. Likely limits for the maximum anomalous and dispersive differences were also calculated and the results used in subsequent calculations. The anomalous differences between Bijvoet pairs measured at λ_1 were then used as coefficients in a Patterson calculation. In addition, dispersive difference Pattersons were calculated between the different wavelengths, and attempts made to identify peaks present on both the anomalous and dispersive Pattersons. Typical examples of the Patterson map at the Harker section $w = \frac{1}{3}$ are shown in figure 6.14.

There were few strong features apparent in the Patterson functions, and no peaks conserved between the maps drawn from the anomalous or dispersive differences. Attempts were made to solve the Pattersons using automated methods as implemented in RSPS [61]. However, when refining the co-ordinates and occupancies of these sites against the information present in the Patterson, as implemented in program VECREF, it was clear that the chosen solutions were incorrect.

The data were scaled and merged in space group P3₁, the lowest symmetry space group consistent with the diffraction pattern. The merging statistics for the white-line wavelength, λ_1 are shown in table 6.6.

A self-rotation function [94] was calculated in order to check the space group assignment. Figure 6.15 shows the $\kappa = 180^\circ$, within resolution range of 8 - 4Å and using a radius of 15Å.

The presence of peaks with height corresponding to 4σ at $\omega = 90.0^\circ$, $\kappa = 180.0^\circ$,

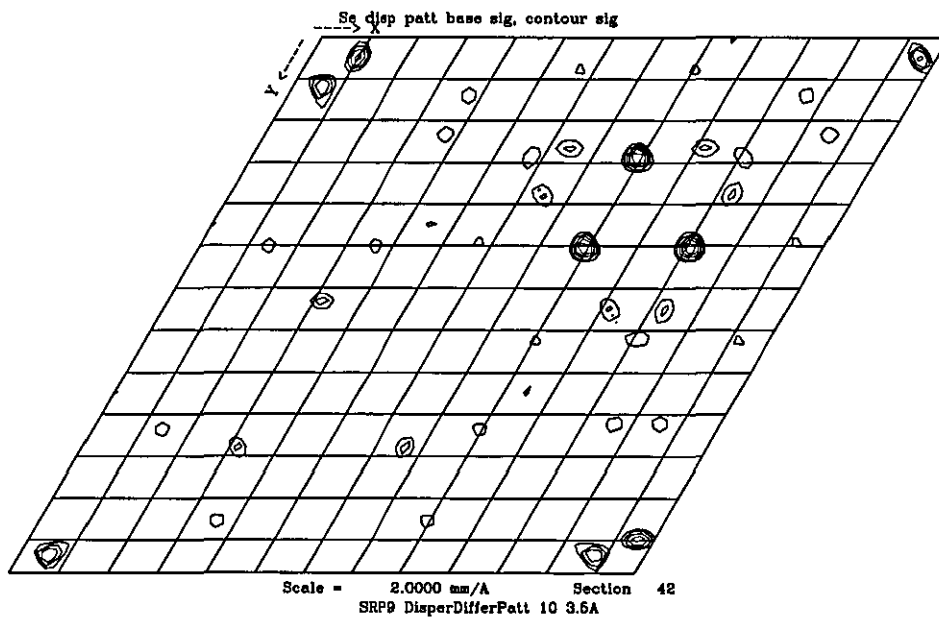
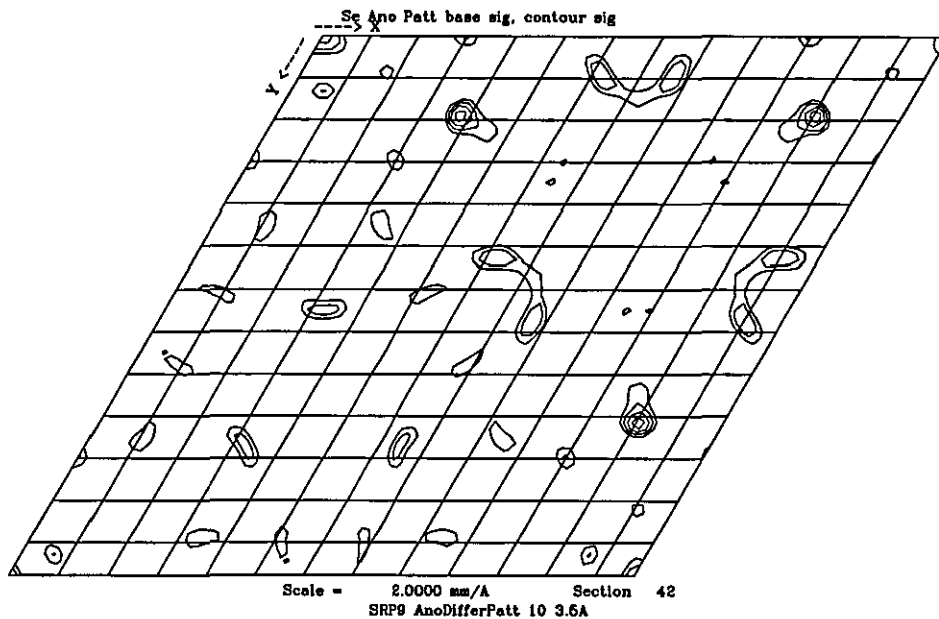
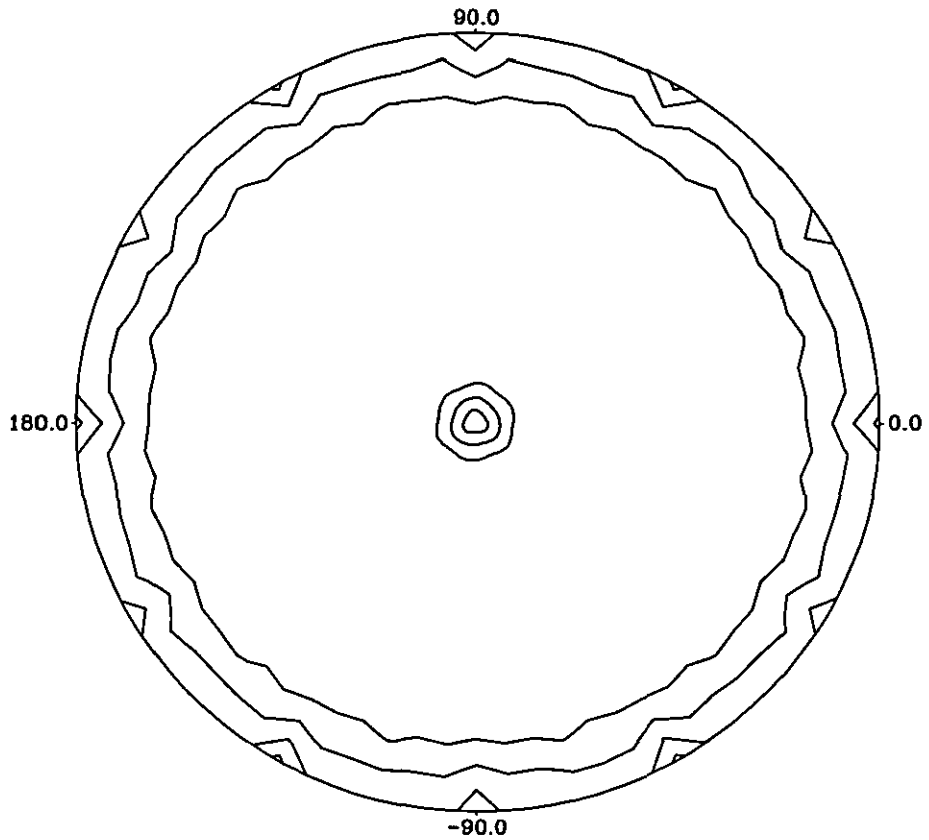


Figure 6.14: The $w=1/3$ sections of the anomalous and dispersive Pattersons. Contouring is set to base= 1.5σ , step= 0.5σ . Resolution limits are 10 - 3.5Å.



self rotn for native RNA as P321
Section kappa = 180

Figure 6.15: Self rotation function using all the data. Contouring is set to base= 1σ , step= 1σ

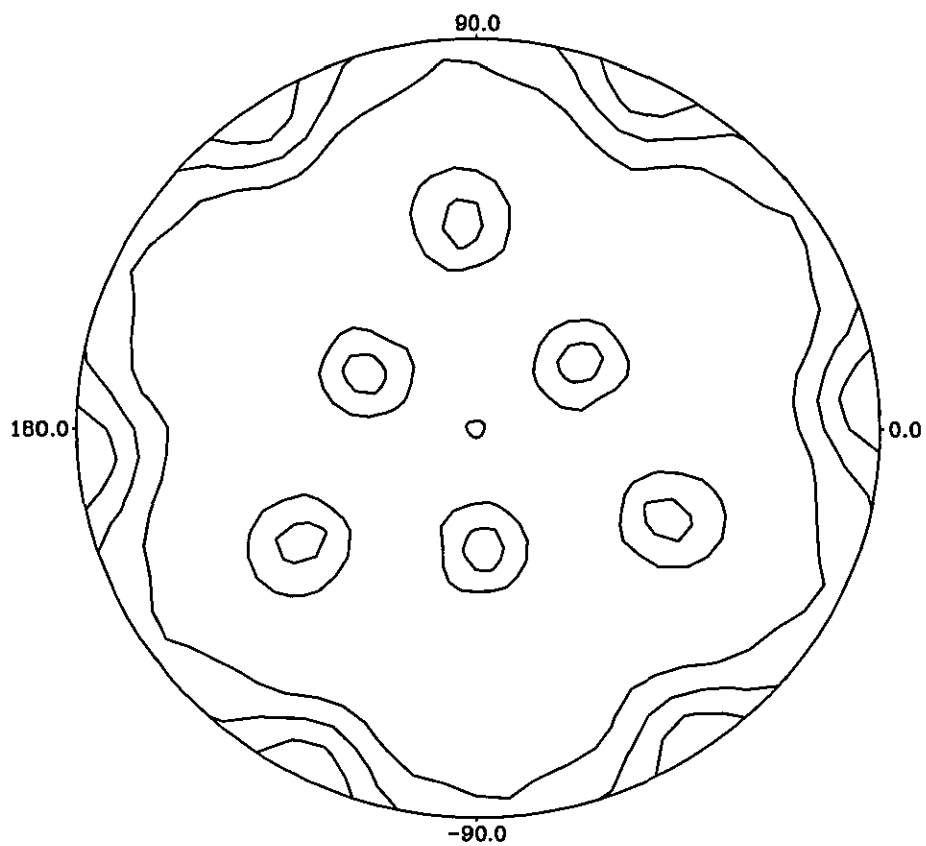
and $\phi = 0.0^\circ/60.0^\circ/120.0^\circ$ suggests that the space group belongs to the point group 321 [62, p. 758]. However, there are also peaks, with height 3σ , at $\phi = 30.0^\circ/90.0^\circ/150.0^\circ$, which if taken together with the original peaks, is indicative of the point group 622 [62, p. 763]. Attempts were made to reduce the level of the background by changing the resolution limits of the data and modifying the molecular radius, but this did not improve the quality of the plots. Taking the 4σ peaks to be the most significant, the self-rotation function suggests that the point group is 321, but with the presence of significant non-crystallographic symmetry.

However, when removing all data with a ratio $\frac{F}{\sigma} < 45$, so that only the strongest 10% of the data remained, there was evidence of peaks at $\omega = 90.0^\circ, \phi = 4.1^\circ, \kappa = 180.0^\circ$ and $\omega = 90.0^\circ, \phi = 64.1^\circ, \kappa = 180.0^\circ$ and $\omega = 34.6^\circ, \phi = 33.6^\circ, \kappa = 180.0^\circ$, see figure 6.16.

These results placed doubt over the choice of space group. In the point group 321, there is a 2-fold axis at 60° to the crystallographic a and b axes. The self-rotation function using only 10% of the data suggests that the point group is in fact 3, where the 2-fold crystallographic axis for point group 321, is in reality a non-crystallographic axis 4° away. Furthermore, the peak at $\omega = 34.6^\circ, \phi = 33.6^\circ, \kappa = 180.0^\circ$ provides evidence of a second non-crystallographic 2-fold axis.

Taking the data processed in the space group $P3_1$, a cumulative intensity distribution [95] was plotted in which the squared intensity of each merged reflection was compared with the mean squared intensity of all reflections. The result from the dataset λ_1 is shown in figure 6.17, and was seen to be the same for λ_2 and λ_3 .

The experimental distribution is seen to differ significantly from the theoretical distribution predicted from Wilson's calculation. Upon comparing the results with those of Rees [96] and Stanley [97], the presence of twinning by merohedry was suspected. Based on this result, the program GETRATIO (appendix A) was written to calculate the so-called Wilson ratio [95], given in equation (6.7).



SRP9 SeMet (tw) 45 sig resol 8 4 radius 15
Section kappa = 180

Figure 6.16: Self rotation function with 10% best measured data, contours: base sigma, step sigma.

Cumulative intensity distribution²

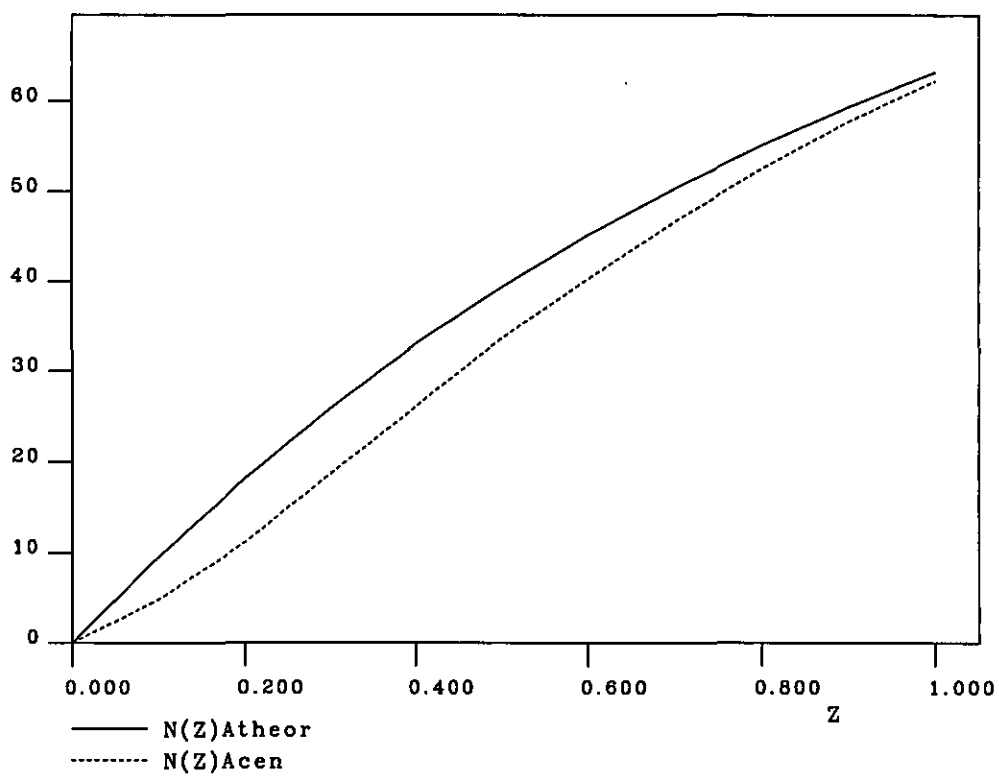


Figure 6.17: Cumulative Intensity Distribution murine-SRP9-SeMet-Yb TWINNED. There are no centric zones in this space group.

$$\frac{\overline{I_{hkl}^2}}{(\overline{I_{hkl}})^2} \quad (6.7)$$

According to Redinbo and Yeates [98] this ratio takes a value of 1.5 for perfectly twinned crystals, and 2.0 for non-twinned crystals. The value obtained from the observed data gave a value of the Wilson ratio of 1.65, suggesting that twinning by merohedry was indeed present, but not to the point of being perfect.

A twinned crystal contains two or more domains, whose relative orientations can be described by symmetry operations called twinning operators. Twinning by merohedry arises if the twinning operator belongs to the point group of the crystal lattice [99, p. 12]. Hemihedral twinning is a special case of merohedral twinning arising when there are only two domains. The diffraction pattern from a hemihedrally twinned crystal is not visibly altered because the diffraction patterns associated with each domain exactly overlap; each spot therefore includes a contribution from both domains of the crystal.

The crystal morphology (see figure 6.4) indicated that hemihedral twinning was a likely possibility, with the bipyramidal crystals being composed of two single crystal pyramids sharing the same base. Figure 6.18 shows the proposed twinning operation, and the possible arrangement of the molecules in the crystal, when the twinning operator corresponds to a 60° rotation about a 3-fold crystallographic axis.

With the twin domains defined as in figure 6.18, the twinning phenomenon might arise from the physical position of the crystal in the X-ray beam, as shown in figure 6.19. During experiments, the X-ray beam was collimated so that its cross-section was $150\mu\text{m} \times 150\mu\text{m}$, which was smaller than the dimensions of the crystals. It would seem likely that the beam then illuminates both twin domains at the same time, but in a proportion dependent upon crystal orientation.

In order to deconvolute the observed intensities, it was necessary to identify the reflections related by the twinning operation, labelled as A and B in figure 6.19. This was achieved using the construction shown in figure 6.20.

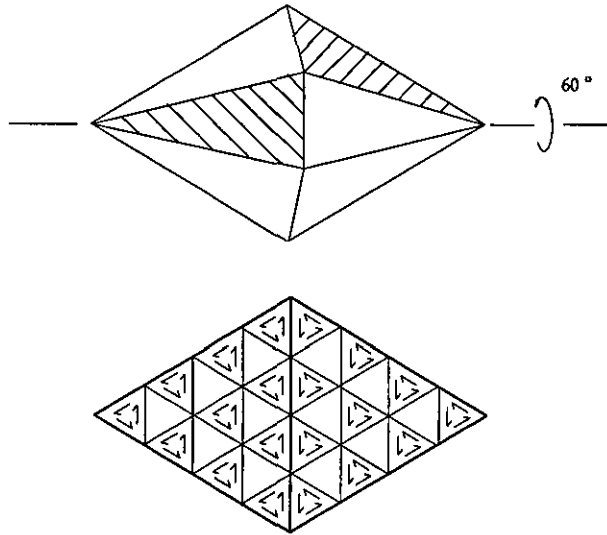


Figure 6.18: The proposed twinning operation corresponding to a 60° rotation about c^* . Adapted from ref. [100]

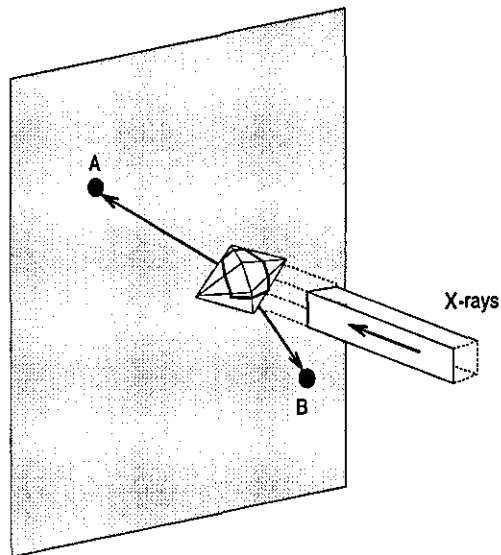


Figure 6.19: The physical position of the specimen in the X-ray beam might determine the twinning fraction

P31

twinning operation 1 : +60 deg.rotation about l

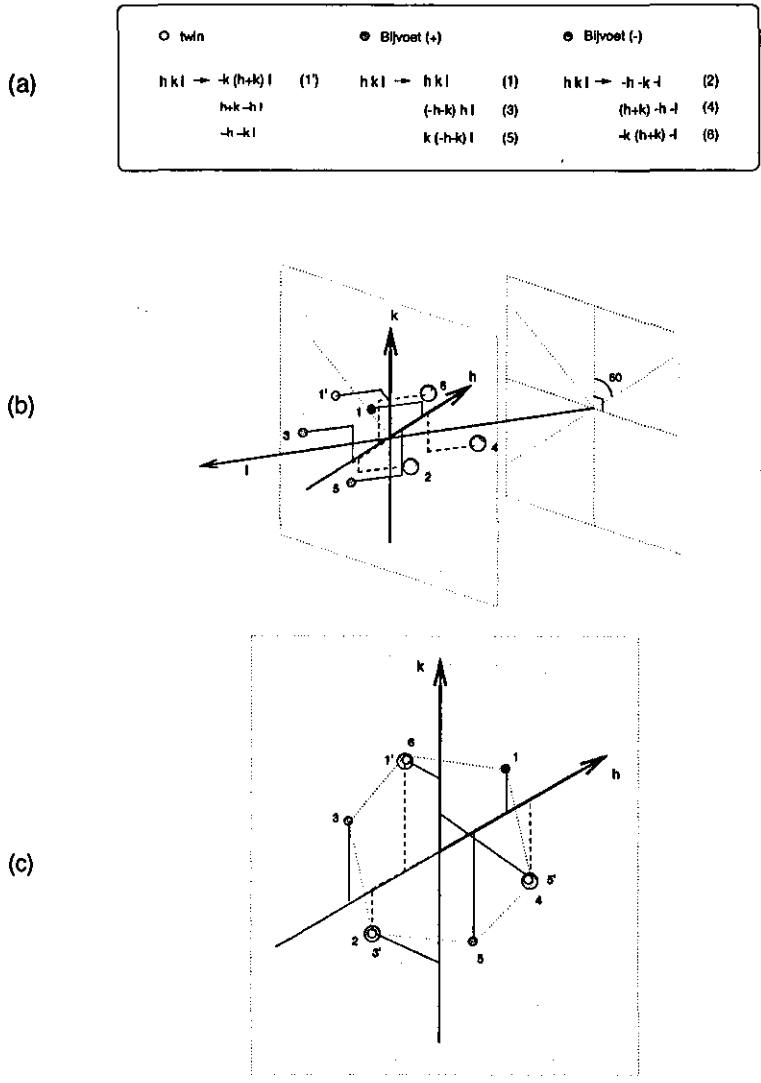


Figure 6.20: The identification of the twin pairs generated by a 60° rotation about the c* axis

This figure is a reciprocal space representation, where in part (b) the reflection hkl is represented as a red sphere. The action of the symmetry operators, listed in part (a) as (1), (3) and (5), is to generate the reflections represented by the blue spheres. The centre of symmetry of the diffraction pattern is represented here by the yellow spheres labelled (2), (4) and (6), generated by the inversion centre from spheres (1), (3) and (5), respectively. The reflection (1') is generated from (1) by a 60° rotation about the c^* axis. Assuming that the c^* axis lies parallel to a line joining the two tips of the crystal, reflections (1') and (1) will be related by the proposed twinning operator. In a similar way, the twinning operation generates (3') from (3) and (5') from (5).

The twinning operation therefore relates reflection hkl to $-k(h+k)l$, $(h+k)-hl$ and $-h-k$ (listed as (1'), (3') and (5'), respectively in part (a)). This result was later seen to be in agreement with the twinning operation proposed by Yeates [100].

The observed intensity of a twinned reflection can be written as the weighted sum of the intensities of the two reflections related by the twinning operation, where the weight is the twinning fraction, α , thus

$$I_{obs(hkl)} = \alpha I_{hkl} + (1 - \alpha) I_{-h-kl} \quad (6.8)$$

$$I_{obs(-h-kl)} = (1 - \alpha) I_{hkl} + \alpha I_{-h-kl} \quad (6.9)$$

With a knowledge of which reflections were twin related, the program GETALPHA (appendix A) was written to calculate the twinning fraction. The twinning fraction quantifies the percentage of each domain contributing to the overall diffraction pattern. The algorithm used in the program was based on an equation proposed by Yeates [101], which is shown in equation (6.10). α is the twinning fraction and $|\overline{H}|$ the overall mean value of H for all twin pairs within the dataset, where H is defined in equation (6.11).

$$\alpha = \frac{1 - 2|\overline{H}|}{2} \quad (6.10)$$

where

$$H = \frac{q - p}{q + p} \quad (6.11)$$

where p and q are the intensity of the reflections related by the twinning operation. GETALPHA gave a twinning fraction of 32%.

One problem with the hypothesis, is that it appears to be possible that the twinning fraction will change as the crystal rotates during the experiment. For this reason it was judged useful to use merged data, thereby averaging out any differences.

In the space group $P3_1$, AGROVATA was used to merge symmetry equivalent reflections whilst keeping Bijvoet pairs separate. The intensity of each merged reflection, measured at a particular wavelength, was deconvoluted by program DETWIN (appendix A). The program searches for pairs of reflections related by the twinning relation, and uses the intensities of each to solve the pair of simultaneous equations presented in equation (6.2). In this way, the program is able to deconvolute the contributions of the twin regions to each observation, thereby effectively de-twinning the data.

The output from program DETWIN for one of the wavelengths is shown below.

```
-----  
Program deTwin : de-twinning program  
-----  
Twinning fraction [0 < a <= 0.5] : 0.325  
19196 TOTAL number of individual Bijvoets  
Beginning algorithm  
3461 B(+):Twin pairs  
4090 B(-):Twin pairs  
-----  
4838 B(+):B(-) pairs
```

5055 Twins with no Bijvoet partner

15102 TOTAL number observations de-twinned

Completed calculation

It can be seen that there are 19196 reflections input to the program, but only 15102 reflections output. The completeness of the data therefore falls during the deconvolution procedure. This is a consequence of the fact that when only one member of a twin pair is present, deconvolution can not be performed; the program then removes this reflection from the dataset.

As a means of verifying the success of the de-twinning procedure, the cumulative intensity distribution was plotted, and is shown in figure 6.21

Cumulative intensity distribution²

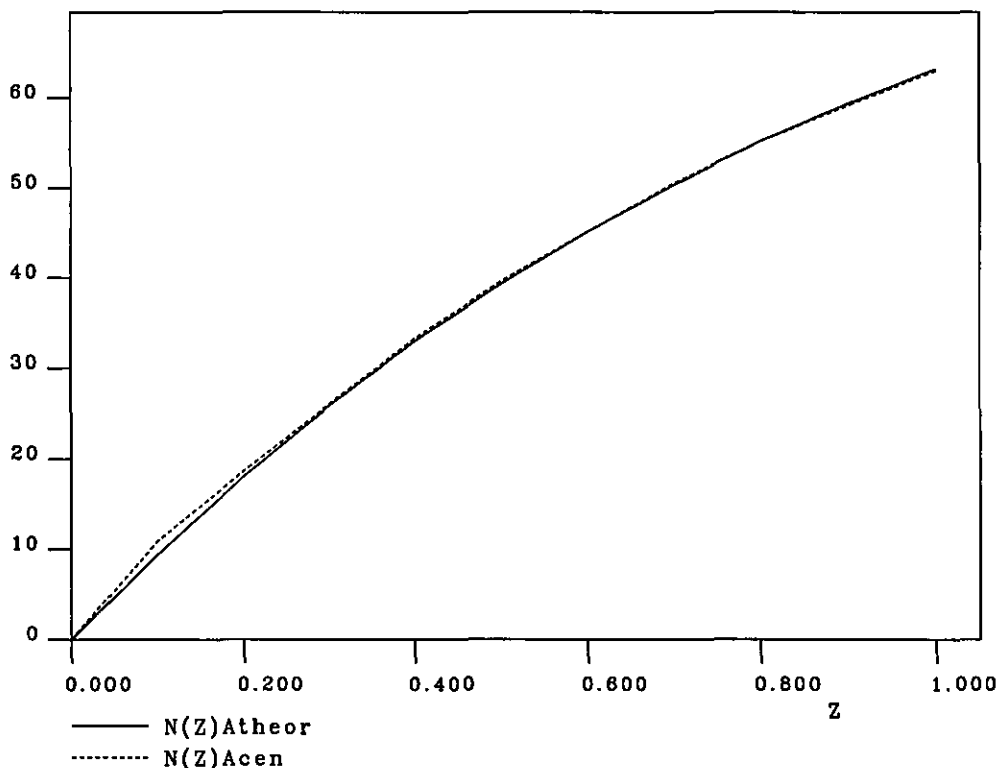


Figure 6.21: The cumulative intensity distribution for de-twinned data from λ_1

The de-twinned intensities now appear to follow the theoretically predicted distribution. This suggests that the de-twinning operation has correctly deconvoluted the observed intensities. Upon repeating the calculation performed by GETALPHA, the twinning ratio appeared to be less than 1%, providing additional evidence that the de-twinning procedure had been successful.

In addition, a self-rotation function was calculated using the de-twinned data, taking the 10% best measured data, as previously described. There does not appear to have been any changes introduced as a result of de-twinning, see figure 6.22.

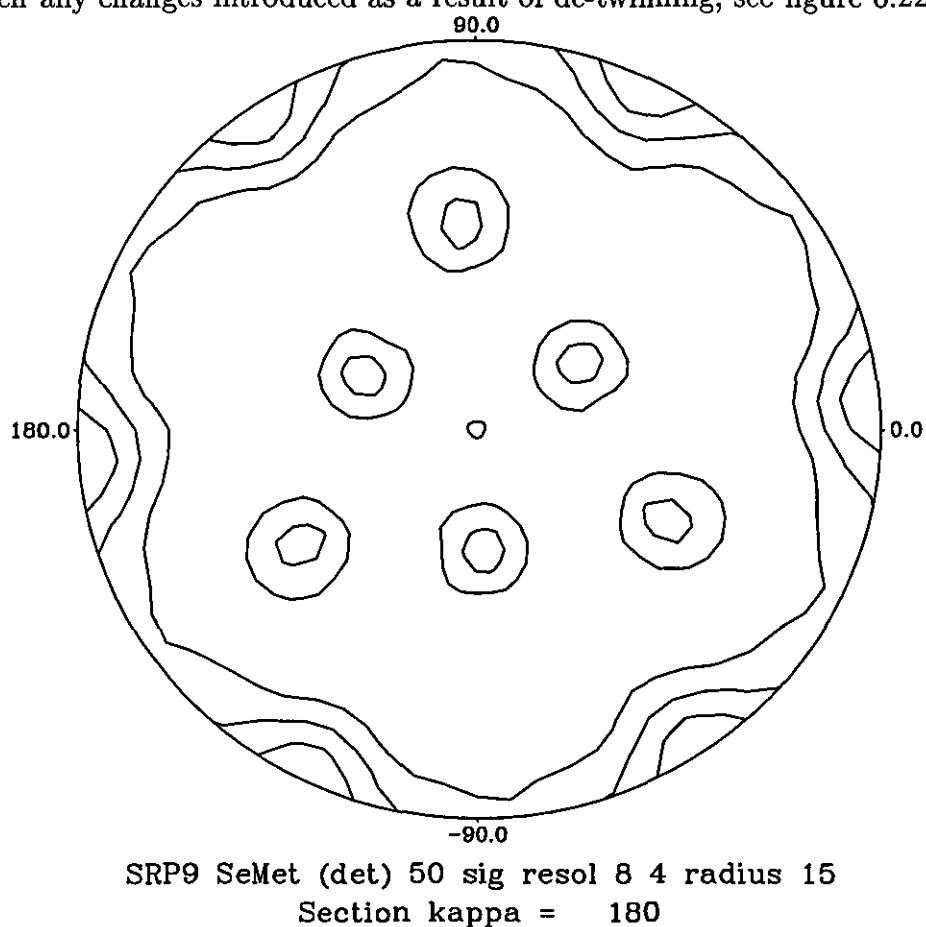


Figure 6.22: Self rotation function with 10% best measured de-twinned data, contours: base sigma, step sigma.

The de-twinned data for each wavelength was placed on a unique scale by program SCALEIT, and likely limits for the maximum anomalous and dispersive differences calculated. With the data from the three wavelengths combined into one file (by

program CAD), the program MADINP (appendix A) was written to reformat and calculate the resolution for each reflection. The data was then in a suitable format to be used for the program MADLSQ [102]. Inter-wavelength scale factors were kept fixed, while the values of the f' and f'' were refined. Convergence was achieved when assuming that there were about 20 selenium selenium atoms within the unit cell. From the Matthews coefficient [65], it appeared likely that there were 18 molecules in the unit cell. The MADLSQ result therefore suggests that there was only one ordered selenium atom per molecule.

The square of the heavy atom structure factors thus obtained, were taken as coefficients in a Patterson calculation using program FFT. Various different resolution ranges of data were compared, in order to identify conserved peaks. The Harker section $w = \frac{1}{3}$ shown in figure 6.23 was chosen as the most representative.

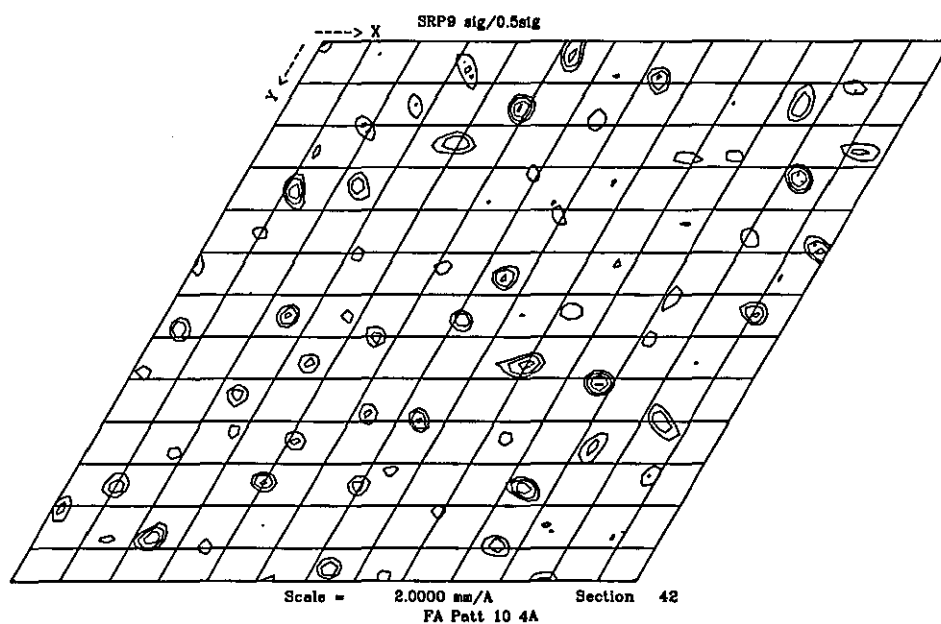


Figure 6.23: Harker section, $w = \frac{1}{3}$, for the Patterson function with square of the heavy atom structure factors, F_A^2 , as coefficients. Contours: base sigma, step sigma. Resolution limits are 10 - 4Å.

The Patterson was interpreted in $P3_1$ and $P3_2$, using the program RSPS. The coordinates and occupancies were refined by VECREF; the coordinates, dispersive and anomalous occupancies of the first site were refined by MLPHARE, before the calcu-

lation of phases. A one-electron model was used in order that the refined dispersive occupancy reflected the value of $\Delta f'$ between the wavelengths, and the anomalous occupancy the value of f'' for each wavelength. Convergence of the values of these parameters could be obtained after 25 cycles of refinement. The phases obtained were used to calculate an anomalous residual Fourier (using program FFT) with the aim of identifying additional heavy atom sites, taking as coefficients the Bijvoet differences at λ_1 . The only residual peaks obtained appeared to be 'shadows' of the original site, giving identical coordinates.

The reason for the lack of success at identifying additional peaks could be attributed to the poor quality of the initial phases. This would indicate that the heavy atom site identified from the Patterson was in fact not correct. Selecting different sites to calculate the initial phases lead to the same result.

Attempts to Solve mouse-SRP9 by Molecular Replacement

Attempts were made to solve the structure by molecular replacement methods, taking the SRP9 moiety of the murine-SRP14/9 structure [86] with both native and selenomethionine incorporated proteins.

Various modifications were made to the search model, including the removal of loop regions, and truncation to include only the helix or β -sheet regions. In addition, data from many different resolution ranges was used, with the aim of eliminating the effects of either gross or detailed differences between the structures. However, no clear solution could be obtained.

Since the technique of molecular replacement is based upon the Patterson function calculated from the observed intensities, its success will depend upon whether these intensities are reliably measured. The data completeness is also an important criterion, if the technique is to be able to find a valid solution. The lack of success of this method could be attributed to a reduction of data completeness after de-twinning, or to the deterioration in the quality of the corrected intensities. Alternatively, another possible explanation is that the model structure is not homologous to the murine-

SRP9, and therefore that the conformation of the SRP9 molecule is different in the free form compared with the heterodimeric form.

6.3 Discussion

The results from experiments performed during the search for heavy atom derivatives, indicated that different crystals of the mouse-SRP9 did not conserve space group. This result is consistent with the hypothesis that the crystals are twinned by hemihedry. The proposed twinning operation, where the two pyramids of the crystal belong to different domains, modifies the observed intensities, with the consequence that distinguishing between the trigonal and hexagonal space groups becomes difficult. Alternative twinning operations were considered, but were rejected since they would not lead to hexagonal symmetry. The degree to which the intensities are modified is governed by the twinning fraction, whose value will be different for each crystal as a result of changes in the specimen position with respect to the X-ray beam.

If the apparent Laue symmetry of the diffraction pattern is $6/mmm$, then the space group could be deduced as $P6_222$ or $P6_422$. In the presence of twinning by merohedry, these apparent space groups can arise from any one of the following non-twinned space groups [99, p. 13]; $P3_1$, $P3_2$, $P3_121$ or $P3_221$. The self-rotation function for $\kappa = 180^\circ$ shows no evidence for the presence of 2-fold symmetry axes, thus suggesting that the space group is either $P3_1$ or $P3_2$.

Since the emergence of 2-fold non-crystallographic axes in the self-rotation function calculated with only 10% of the data, was observed with both the twinned and de-twinned data, it seems likely that these axes relate molecules within a dimer. There is clear evidence from the SDS-PAGE and mass spectroscopy results that murine-SRP9 crystallises in the dimeric form. The mass spectroscopy results also showed the presence of up to two β -ME molecules per dimer, a result which may offer an explanation of the actual mechanism by which the dimer forms. β -ME has been seen to bind to cysteine residues [86], of which there are four on the murine-SRP9 molecule. The presence of only two β -ME molecules per dimer, indicates that two

cysteine residues are unavailable, perhaps as a consequence of the formation of a disulphide bridge [103].

Additional evidence for this hypothesis is given from the crystallisation trials, where the presence of the reducing agent DTT was seen to prevent crystal formation. If a disulphide bridge is to form, the terminal hydrogen from the cysteine residue must be removed by oxidation. This reaction would be prevented by the DTT, thereby preventing dimerisation and subsequent crystal growth. The same effect is perhaps seen with the reducing agent β -ME, used during the purification. β -ME reacts with any oxygen ions present in the environment of the stock solution. After a period of 3 months, the supply of β ME is depleted; oxidation of the protein molecule can no longer be prevented, with the consequence that the cysteine residues become available for disulphide bridge formation.

The cumulative intensity distribution for the de-twinned data offers evidence that the deconvolution procedure had been successful. However the subsequent phase calculation clearly does not provide phases of high enough quality to allow the structure to be solved. A possible explanation might be that the coordinates of the heavy atom sites obtained are incorrect. It has been shown that the Patterson maps calculated with heavy atom structure factors, F_A , as coefficients give clearer peaks than those calculated from the anomalous and dispersive differences [105]. However, this result is dependent upon the reliable calculation of F_A by MADLSQ. It is essential to have a correct estimation of the errors, if the calculation of structure factors from observed integrated intensities is to be reliable. In the de-twinning algorithm, two observed intensities are used to solve a pair of simultaneous equations. It is possible that the problem of the propagation of errors during this procedure has not been correctly addressed, thereby effectively leading to a deterioration in quality of the data.

Within the 86 residues of the murine-SRP9, four residues are methionine; a ratio which is higher than that observed for most soluble proteins [104]. Given the presence of six monomers per asymmetric unit, the crystals of selenomethionated protein will contain 24 selenium atoms per asymmetric unit. Recent results from experiments performed using the MAD technique have shown that Patterson maps calculated from

data containing many sites is a non-trivial task [106] [107]. The success in such cases would be particularly dependent upon the data being complete and well-measured with reliable error estimates. In the present case, few of these criteria are fulfilled; it therefore appears unlikely that the structure can be solved.

6.4 Structure of Human SRP9

During the writing of this thesis, experiments were performed on the human SRP9 (human-SRP9). These experiments lead to the successful structure solution. Since this result was obtained only a few days before the submission of this thesis, only a brief summary of the crystallisation and experimental procedure can be given.

Preparation of Selenomethionine Incorporated human-SRP9

A selenomethionine incorporated human-SRP9 protein was prepared by Ulrike Kapp (EMBL-Grenoble), using an expression clone provided by K. Strub (University of Geneva). β -mercapto-ethanol (β -ME) was used during the purification in an attempt to prevent the oxidation of the protein Doublé [91]. Mass spectrometry results [92] confirmed that four selenium atoms were bound to each molecule.

Crystals were grown by the hanging drop method in 15% PEG 1500, with 100 mM Na-acetate at pH4.5 as buffer, 50 mM ammonium sulphate and 2mM DTT reducing agent at a temperature of 4°.

After 2 days, crystals had grown to about 100 μ m in length, having square-based bipyramidal morphology.

A non-reducing SDS-PAGE gel, performed on a solution containing dissolved crystals, indicated that the molecule was in monomeric form [108]. There is clearly a significant difference in behaviour between the human and murine SRP9 molecules. A sequence alignment between the two sources, shows that there are six non-conserved residues, see table 6.7.

human	M P Q Y Q T W E E F S R A A E K L Y L A D P M K A R V V L K	30
murine	M P Q F Q T W E E F S R A A E K L Y L A D P M K V R V V L K	30
human	Y R H S D G N L C V K V T D D L V C L V Y K T D Q A Q D V K	60
murine	Y R H V D G N L C I K V T D D L V C L V Y R T D Q A Q D V K	60
human	K I E K F H S Q L M R L M V A K E A R N V T M E T E	86
murine	K I E K F H S Q L M R L M V A K E S R N V T M E T E	86

Table 6.7: Sequence alignment of murine and human SRP9. The six, non-conserved residues are highlighted in red.

human-SRP9 MAD Experiment

A multi-wavelength anomalous dispersion experiment was performed on BM14 at the European Synchrotron Radiation Facility, ESRF. A single crystal of the selenomethionine incorporated human SRP9, was flash frozen to 115K. As cryo-protectant three separate solutions were used, comprising mother liquor and glycerol at 10%, 15% and 20%. The crystal was soaked for 20 seconds in the 10% and 15% solutions, then 40 seconds in the 20% solution before being flash-frozen in the cold-stream.

The X-ray fluorescence from the crystal was measured as a function incident X-ray energy from the selenium K-edge, in order to determine the wavelengths to be used in the data collection. The fluorescence spectrum from the crystal is shown in figure 6.24.

Four wavelengths were chosen;

$$\lambda_1 = 0.9785\text{\AA} \quad \text{white line}$$

$$\lambda_2 = 0.9787\text{\AA} \quad \text{inflection point}$$

$$\lambda_3 = 0.8850\text{\AA} \quad \text{remote, above edge}$$

$$\lambda_4 = 1.1270\text{\AA} \quad \text{remote, below edge}$$

Data were collected with the crystal in an arbitrary setting, using the ESRF Image Intensified CCD detector. Complete data sets were obtained prior to changing the incident X-ray energy. The data sets comprised 75° of data using an oscillation range of 0.5°, and exposure time of 30s.

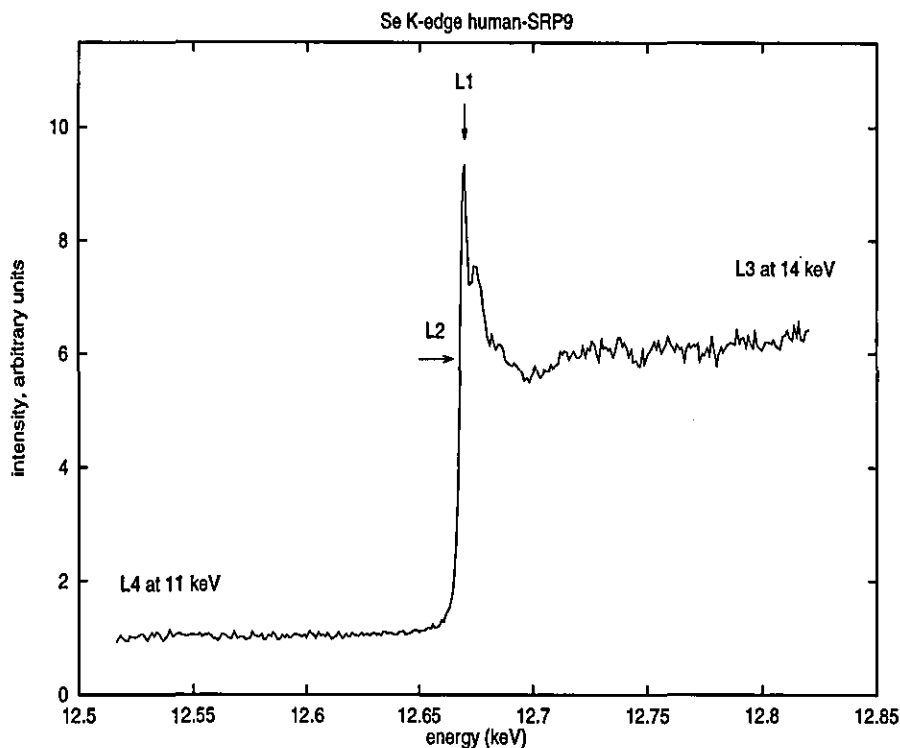


Figure 6.24: Fluorescence spectrum from Se K-edge. The energies at which data was collected are marked.

Data Analysis

Data was indexed and integrated using DENZO [109], from which the crystal system was deduced to be tetragonal. Scale factors were calculated and applied by SCALEPACK [109]; using R_{sym} and based on the systematic absences observed, the space group was chosen to be $P4_12_12$ or $P4_32_12$. The data at each wavelength was merged by AGROVATA [61], and structure factors obtained using TRUNCATE. The merging statistics for each wavelength are shown in table 6.8

The data was treated in a manner analogous to multiple isomorphous replacement, see section 3.7. For this the data measured at λ_2 were treated as the 'pseudo-native', while the datasets at λ_1 and λ_3 , were treated as 'pseudo-derivatives' with anomalous scattering. Inter-wavelength scaling was then performed by SCALEIT, and likely limits for the maximum anomalous and dispersive differences calculated as previously described. The anomalous differences measured at λ_1 were then used in a Patterson

Dataset	λ (Å)	Resolution (Å)	R_{sym} (%)	R_{anom} (%)	Completeness (%)	Redundancy
L1	0.9785	2.7	5.2 (9.5)	5.7 (7.7)	93	5.4
L2	0.9787	2.7	5.3 (9.9)	2.9 (5.3)	93	5.3
L3	0.8850	2.7	5.3 (9.1)	3.0 (5.3)	91	5.5
L4	1.1270	2.9	5.0 (13.6)	2.1 (5.0)	91	4.9

Table 6.8: MAD dataset processing statistics, in space groups $P4_12_12$. R_{anom} is defined as $\frac{\sum |<I+>-<I->|}{\sum <I>}$. Values of R_{sym} and R_{anom} in brackets are for the outer resolution shell.

calculation. In addition, dispersive difference Pattersons were calculated between the different wavelengths, and attempts made to identify consistent peaks. The anomalous and dispersive Patterson maps for the Harker sections $w = \frac{1}{4}$, $w = \frac{1}{2}$ and $u = \frac{1}{2}$ are shown in figures 6.25 6.26 and 6.27, respectively.

Solving the Patterson by RSPS in space group $P4_12_12$, gave one potential site. The coordinates and occupancy of this site were refined in VECREF, before being further refined in MLPHARE. The one-electron model was used, as previously described, in order that the refined dispersive occupancy reflected the value of $\Delta f'$ between wavelengths, and the anomalous occupancy reflected the value of f'' for each wavelength. The λ_2 dataset was taken as native in order that all the dispersive occupancies were positive. MLPHARE was run in two passes; during the first, only the centric reflections were used in order to obtain estimates of the pseudo-isomorphous lack-of-closure errors. These values were used in a second run in which the acentric reflections were included.

The phases obtained from MLPHARE were improved by the techniques of solvent flattening [63] and histogram matching [64] as implemented in the program DM. The Matthews number [65] indicated that there were 2 molecules per asymmetric unit, corresponding to a solvent content of approximately 33%. The DM phases were used in a residual Fourier calculation, taking the Bijvoet differences from the λ_1 dataset as

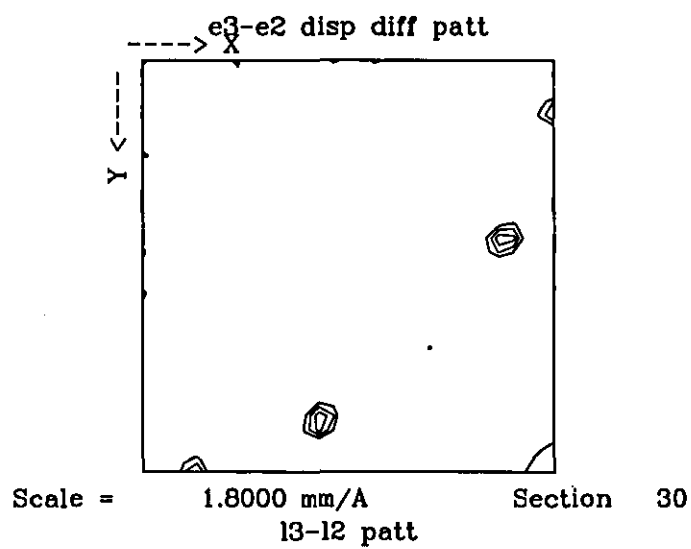
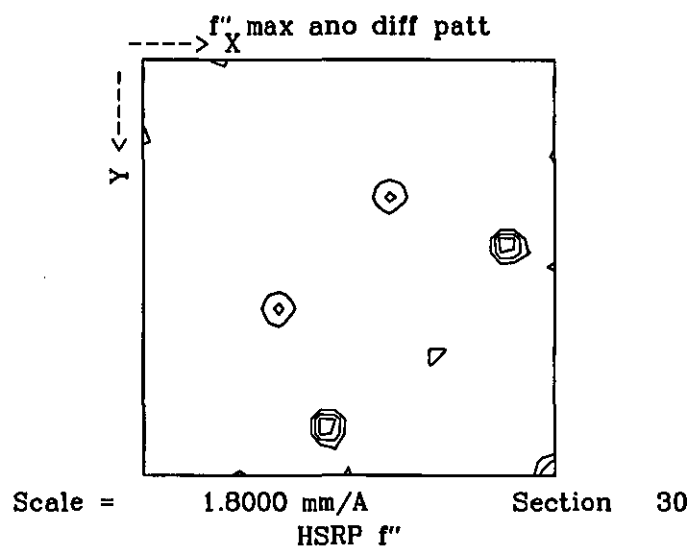


Figure 6.25: The $w=1/4$ sections of the anomalous and dispersive Pattersons. Contouring is set to $\text{base}=1.5\sigma$, $\text{step}=0.5\sigma$. The resolution range is 15-2.7Å.

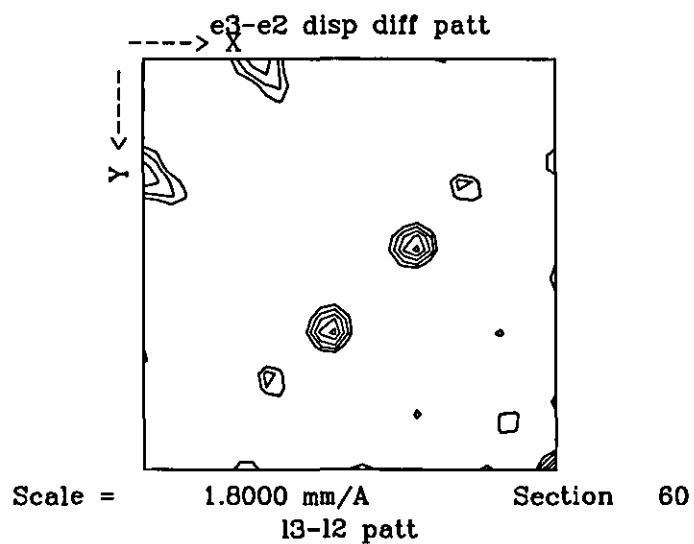
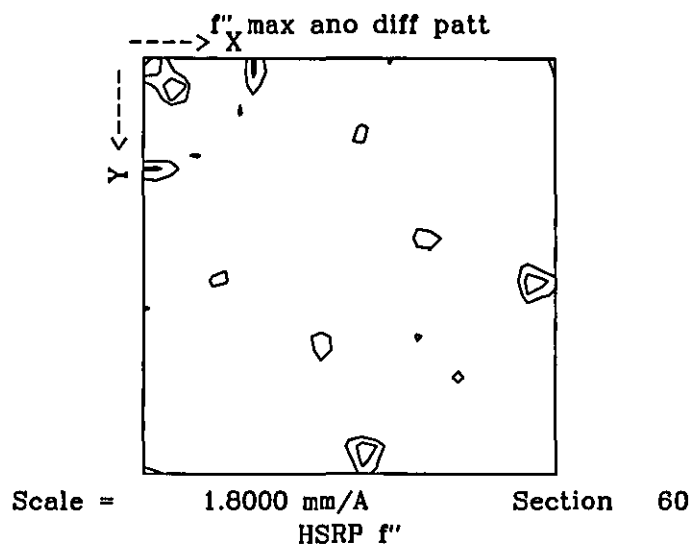


Figure 6.26: The $w=1/2$ sections of the anomalous and dispersive Pattersons. Contouring is set to base= 1.5σ , step= 0.5σ . The resolution range is 15-2.7Å.

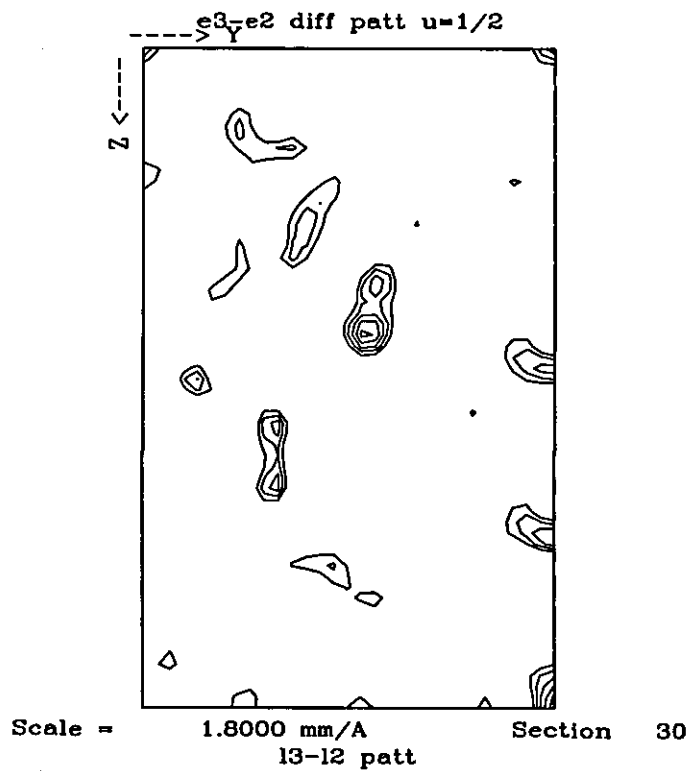
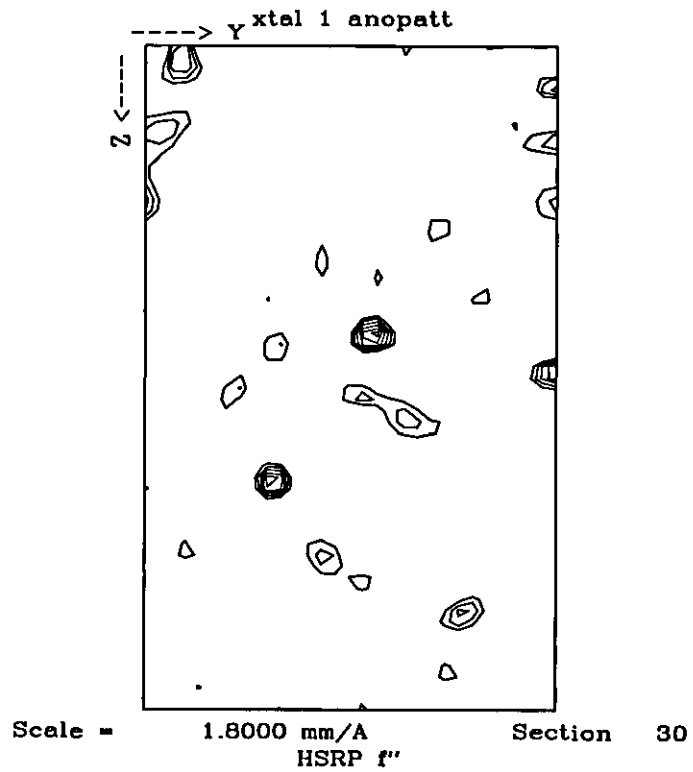


Figure 6.27: The $x=1/2$ sections of the anomalous and dispersive Pattersons. Contouring is set to base= 1.5σ , step= 0.5σ . The resolution range is 15-2.7Å.

site	coordinate			occupancy λ_1	
	x	y	z	disp.	anom.
1	0.110	0.830	0.040	1.638	2.638
2	0.832	0.113	0.005	1.432	2.281
3	0.715	0.269	0.065	0.830	1.839
4	0.129	0.137	0.0711	0.708	0.986
5	0.750	0.371	0.068	0.843	1.335
6	0.317	0.496	0.000	0.948	1.275

Table 6.9: Coordinates of selenium sites found, with their anomalous and dispersive occupancies as calculated by MLPHARE for λ_1 dataset.

coefficients, and additional selenium sites were identified. These were first refined in VECREF, before being included in MLPHARE for a new round of phase calculation.

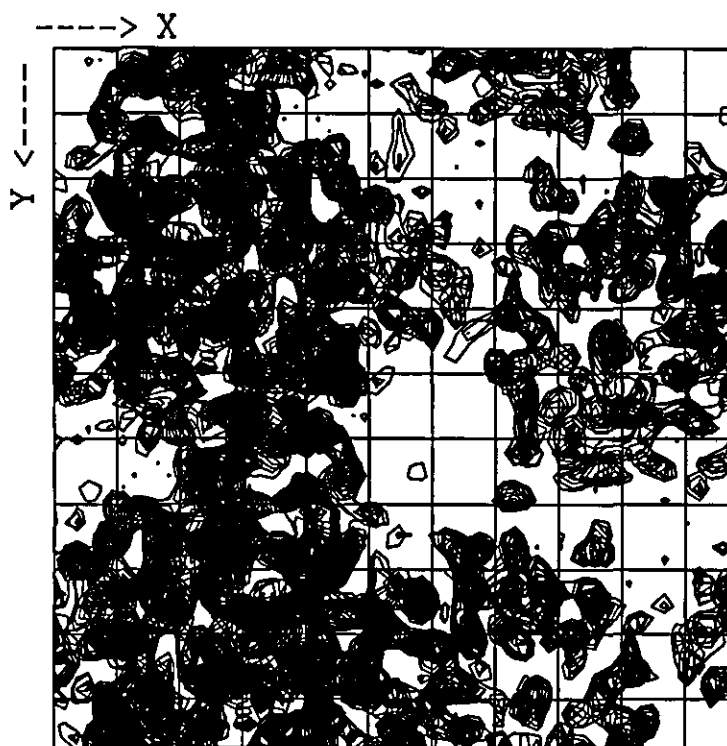
In total, six sites were found whose coordinates are given in table 6.9.

As previously described (section 5.2), the figure of merit, centric and acentric Cullis R-factor, anomalous Cullis R-factor and phasing power were monitored after the inclusion of each site in MLPHARE. A new residual Fourier was calculated and the procedure repeated until no additional valid sites could be identified.

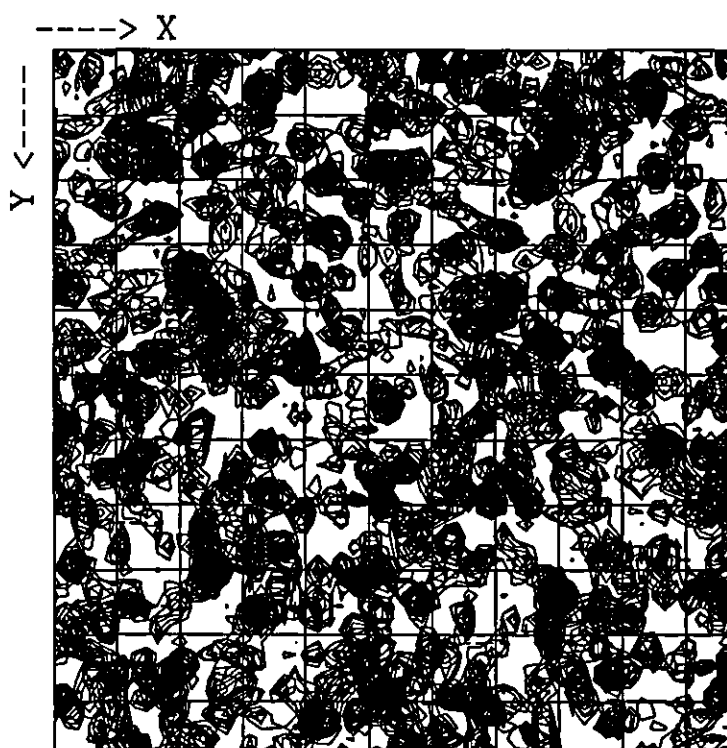
The improvement in the phasing statistics as additional sites are added is shown in table 6.10.

The final MAD phases modified by DM were used in a Fourier calculation by FFT, taking the structure factor amplitudes from the λ_2 dataset. The electron density maps for one asymmetric unit are shown in figure 6.28, for the correct hand and the inverted hand. From this result, the space group was identified as $P4_12_12$.

The electron density was viewed using the graphics program O [66], from which



Scale = 1.2000 mm/A Sections = 0 to 29
SRP9 [Fo, alpha_sol]



Scale = 1.2000 mm/A Sections = 0 to 29
SRP9 [Fo, alpha_sol]

Figure 6.28: Electron density maps from 20 to 2.7Å after solvent flattening. Top: the correct hand. Bottom: the inverted hand.

sites	FOM	$^{acen}P_{power}$		$R_{Cullis}(\lambda_1)$		
		$\lambda_1 : \lambda_2$	$\lambda_3 : \lambda_2$	centric	acentric	anom
1	0.42	0.69	0.86	0.85	0.89	0.78
1, 2	0.46	0.86	1.10	0.87	0.90	0.75
1, 2, 3	0.48	0.94	1.17	0.84	0.88	0.70
1, 2, 3, 4	0.49	0.97	1.21	0.84	0.87	0.69
1, 2, 3, 4, 5	0.50	1.05	1.27	0.82	0.86	0.67
1, 2, 3, 4, 5, 6	0.52	1.11	1.33	0.79	0.84	0.65

Table 6.10: The improvement of the MLPHARE phasing statistics at 2.7Å, after the inclusion of additional sites.

clear evidence could be seen of secondary structure, as well as side-chain features. Figure 6.29 shows a portion of the electron density map, in the vicinity of an α -helix.

An initial skeleton was calculated from the electron density [110], which was used to trace the SRP9 molecule. In addition, the selenium sites were plotted, in order that the position of methionine residues could be checked with reference to the sequence.

The molecule is composed of a four-stranded β -sheet with two helices present on one side, forming an β - α - β - β - β - α fold, see figure 6.30.

It appears that the SRP9 forms a dimer via an interaction of the C-terminal α helices from the two molecules, see figures 6.31, 6.32. This result is confirmed by the symmetrical distribution of the selenium sites.

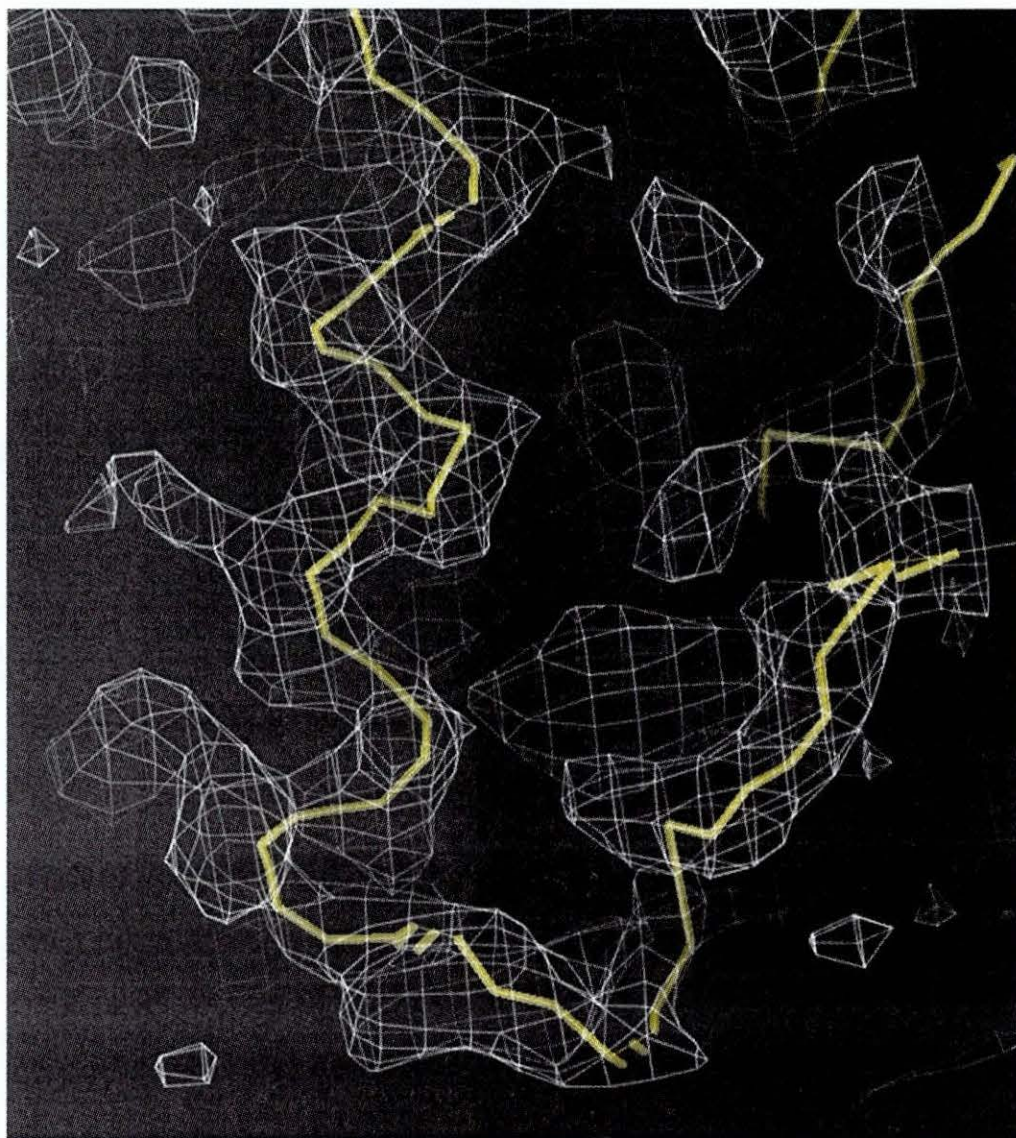


Figure 6.29: Electron density map, showing an α -helix and loop region. The bones from the initial skeleton are shown in yellow.

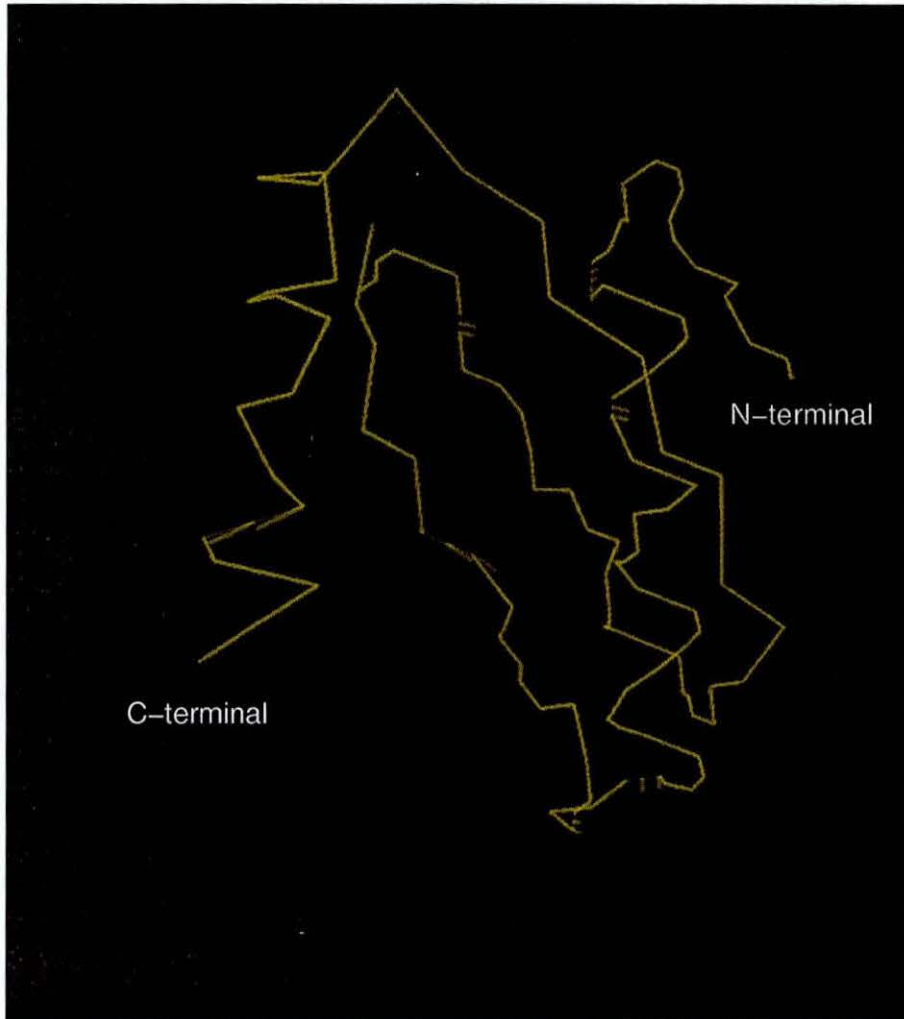


Figure 6.30: The skeleton representation of the human-SRP9 molecule, showing the β - α - β - β - β - α fold.

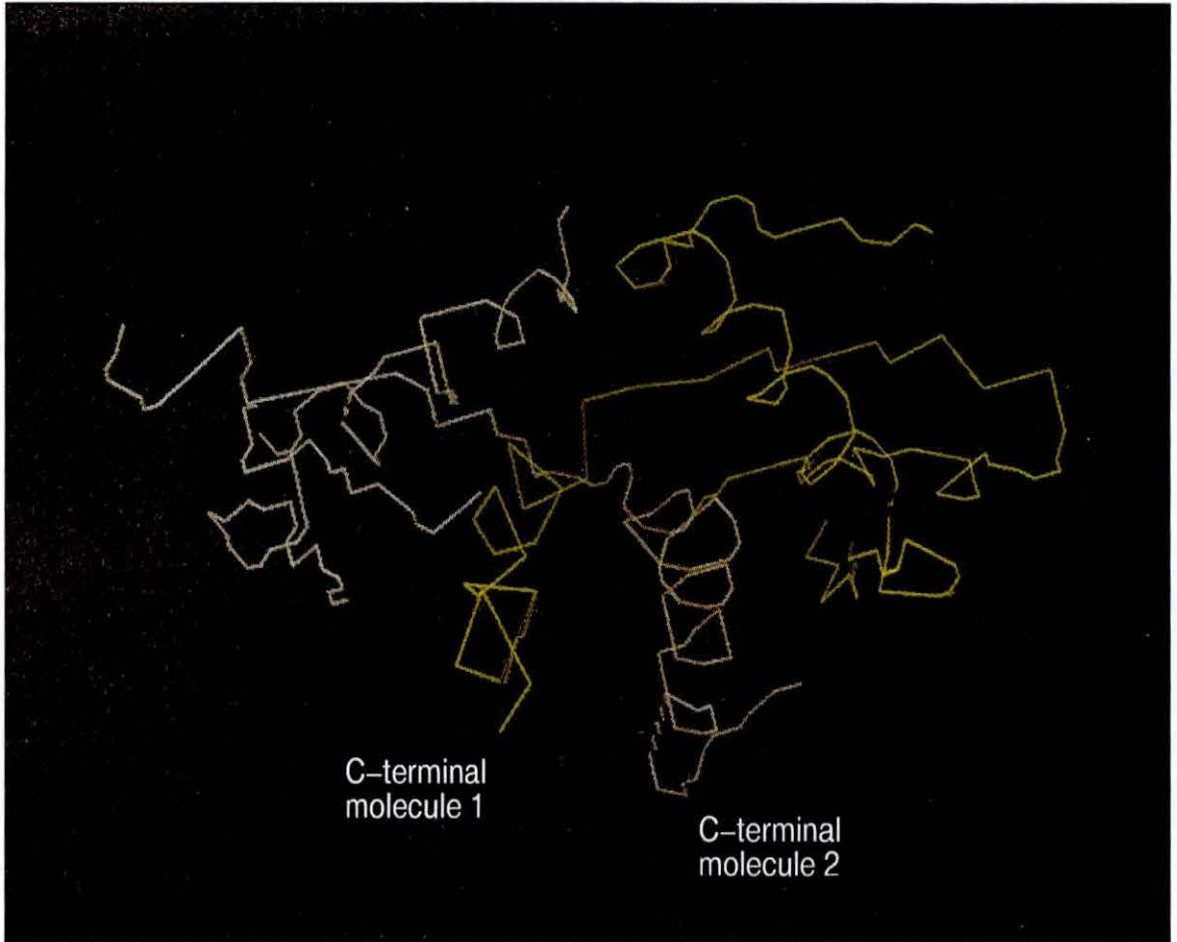


Figure 6.31: Dimeric form of the human-SRP9 molecule, appears to take place via an interaction between the two C-terminal helices. Viewed from above.

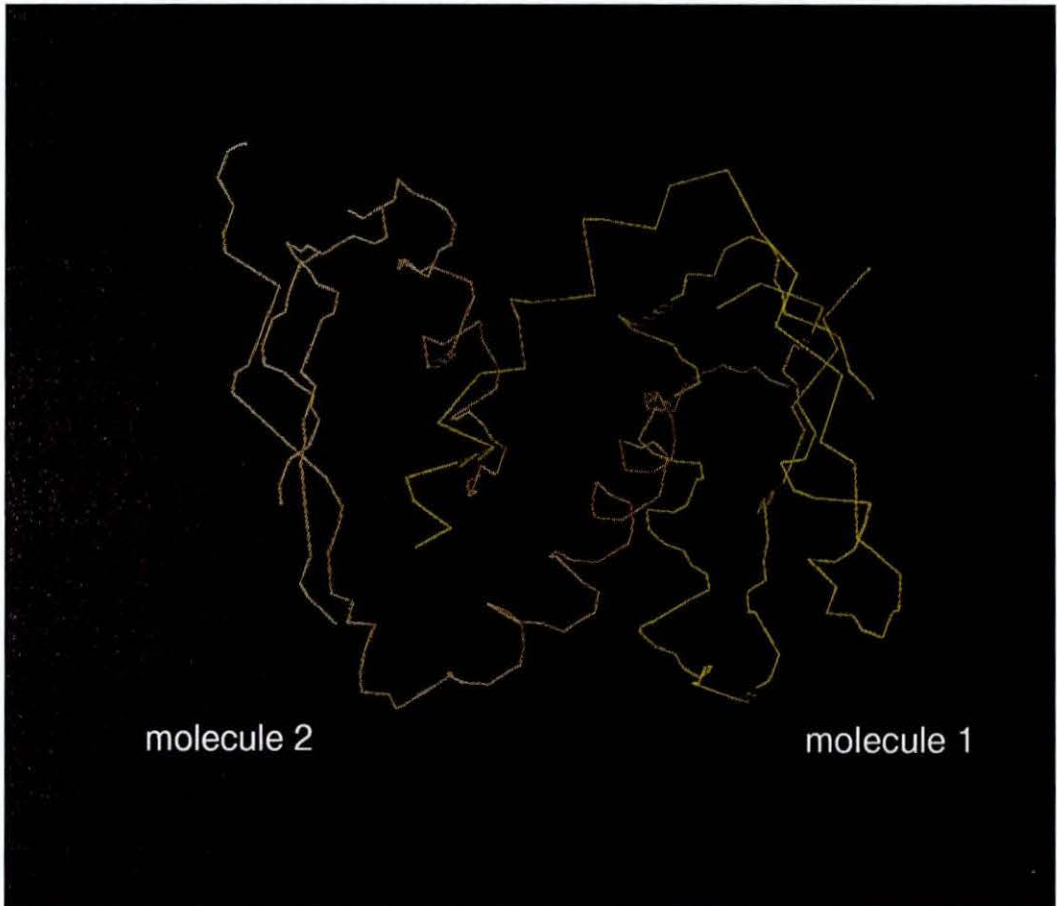


Figure 6.32: A second view of the dimeric form, from the side.

6.5 Discussion

When comparing the structure to that of the SRP9 moiety of the murine-SRP14/9 heterodimer, a significant difference is seen in the orientation of the C-terminal α -helix, with respect to the β -sheet. Furthermore, the N-terminal strand present in the human-SRP9, is not seen in the murine-SRP14/9 fusion protein.

If the conformation of the SRP9 molecule is conserved in the murine and human species, the displacement of the C-terminal α -helix would provide an explanation for the lack of success when using the SRP9 moiety of the SRP14/9 in molecular replacement searches, for reasons as previously discussed.

Further work is required before any explanation can be proposed of the movement of the C-terminal helix. This will include an investigation of the effect of the change of pH of the crystallisation conditions.

Chapter 7

Conclusion

An off-line image plate detector was designed and installed on the beam-line BM14, at the European Synchrotron Radiation Facility. The detector was conceived for collecting diffraction data from protein crystals, using the technique of multiwavelength anomalous dispersion. The requirements of such experiments were identified and used to define the design criteria. Image plates were held in individual cassettes by a vacuum system, which reduced non-uniformity of response errors to a minimum. Fiducial spots were recorded using a Xenon arc-lamp and fibre optic light guides, in order that any differences in plate position could be corrected. Systematic errors introduced by the effects of latent image decay could also be corrected from a record of the elapsed time between the exposure and scan. A barcode identification system, meant that the image plates could be exposed and scanned in arbitrary order; all experimental parameters were conveniently recorded in a database.

A demonstration MAD experiment was performed using the detector, on the ribonucleotide reductase free radical protein R2 (RNR). The structure of this 43.5kDa protein was successfully solved using the scattering from mercury atoms which were introduced into the native crystals. From the results of this experiment, aspects of the detector design were evaluated.

Since this experiment, the detector has been used on several occasions, (William Hunter (Dundee University, UK), Janos Hajdu (Uppsala University, Sweden), which have confirmed that it is a reliable means of collecting protein crystallographic data

using the MAD method.

Multiwavelength anomalous dispersion experiments were performed on the murine and human 9kDa protein of the signal recognition particle (SRP9). Original experiments were performed on crystals of the murine protein, which later turned out to be twinned by merohedry. An attempt was made to de-twin the observed intensities, but after this, neither the techniques of molecular replacement nor multiwavelength anomalous dispersion produced a successful solution. In the final stages of the thesis however, the structure of the human-SRP9 was solved from crystals of the selenomethionine-incorporated protein. Initial views of the structure are presented, and tentative comparisons made with the homologous structure of the heterodimer SRP9/14.

Future work will involve building and refining the present structure, and then to try and establish whether the conformational changes observed are an artifact of the acidic crystallisation conditions or represent genuine differences between the SRP9 molecule in the free form compared with the heterodimeric form.

In addition, attempts will be made to solve murine SRP9 using the de-twinned data to ascertain whether the molecular conformation is conserved between the murine and human SRP9 structures, as well as confirming the validity of the de-twinning procedure.

Appendix A

Software to Analyse murine-SRP9 Data

The programs written for the SRP9 project.

- getSign to check for consistent indexing in trigonal space groups
- getAlphato calculate the twinning fraction
- getRatio to calculate the Wilson ratio
- deTwin to deconvolute the twinned intensities

A.1 Program getSign

```
/*
*****
*
*           Program getSign
*
* Program to find DANO relationship between two datasets
* in order to show consistent Bijvoet choice
*
* Kieron BROWN, Jan 1996
*
*****
it is necessary to set environmental variables

    setenv PGPLOT_FONT /user1/sean/lib/grfont.dat
    setenv PGPLOT_DEV "/xwin"
    setenv PGPLOT_DIR /user1/sean/pgplot

prior to running the program

*/

#include "cpgplot.h"

#include <stdlib.h>
#include <stdio.h>
#include <math.h>

main()
```

```

{
  int h1, k1, l1;
  int j, n;
  float d_l1, sigd_l1, d_l2, sigd_l2;
  float dcut;
  float d1[10000], d2[10000];
  float xmin, xmax, ymin, ymax, tmp, tmp2;

  char ps;
  char xlab[40], ylab[40];
  char ifn[50], record[76];

  FILE *ifp1;

  printf("Enter the input filename: ");
  scanf("%s", &ifn);

  /* open the reflection file */

  ifp1 = fopen(ifn, "r");

  /* define the selection criterion */

  printf("Sigma cutoff on DANO: ");
  scanf("%f", &dcut);

  /* label axes */

  printf("Give x-axis label: ");
  scanf("%s", &xlab);

  printf("Give y-axis label: ");
  scanf("%s", &ylab);

  /* read in the reflection file */

  n = 1;

  while (fgets(record, 76, ifp1) != NULL) {

    sscanf(record, "%d%d%d%f%f%f", &h1, &k1, &l1, &d_l1, &sigd_l1,
    &d_l2, &sigd_l2);

  /* apply selection criterion */

    if ((fabs(d_l1) > (dcut * sigd_l1)) && (fabs(d_l2) > (dcut * sigd_l2))) {

  /* write out to a new array (for the plotting) */

      d1[n] = d_l1;
      d2[n] = d_l2;
      n += 1;
    }
  }

  /* Calculate xmax and xmin */

  xmin = d1[1];
  xmax = d1[2];
  for (j = 1; j <= n; j++) {
    if (xmin > d1[j]) xmin = d1[j];
    if (xmax < d1[j]) xmax = d1[j];
  }
}

```

```

}
if (xmin > xmax) {
    tmp = xmax;
    tmp2 = xmin;
    xmin = tmp;
    xmax = tmp2;
}
xmin = xmin - 0.1*xmin;
xmax = xmax + 0.1*xmax;

/* Calculate ymax ymin */

ymin = d1[1];
ymax = d1[2];
for (j = 1; j <= n; j++) {
    if (ymin > d1[j]) ymin = d1[j];
    if (ymax < d1[j]) ymax = d1[j];
}
if (ymin > ymax) {
    tmp = ymax;
    tmp2 = ymin;
    ymin = tmp;
    ymax = tmp2;
}
ymin = ymin - 0.1*ymin;
ymax = ymax + 0.1*ymax;

/* Call cpgbeg to initiate PGPLOT */

cpgbeg(0, "/xwin", 1, 1);
cpgask(1);
cpgpap(7.0,0.861);

/* Call cpgenv */

cpgenv(xmin, xmax, ymin, ymax, 0, 1);
cpglab(xlab, ylab, "DANO Correlation");
cpgbox("bcts", 0.0, 0, "bcts", 0.0, 0);

/* Mark the points */

cpgpt(n, d1, d2, 21);

/* make a hard copy or quit? */

printf("Print or quit? [p/q]: ");
scanf("%s", &ps);

if (ps == 'p') {
    printf("Making postscript file\n\n");
}

/* Make plot to postscript file */

cpgbeg(0, "/ps", 1, 1);
cpgask(1);
cpgpap(0.0,0.861);

/* Call PGENV */

cpgenv(xmin, xmax, ymin, ymax, 0, 1);
cpglab(xlab, ylab, "DANO Correlation");
cpgbox("bcts", 0.0, 0, "bcts", 0.0, 0);

```

```

/* Mark the points */
    cpgpt(n, d1, d2, 21);
    sleep(2);
}

printf("end of plotting\n\n");
sleep(1);

/* End the plotting */

cpgend();

fflush(ifp1);
fclose(ifp1);
}

```

A.2 Program getAlpha

```

/*****
*
*           Program getAlpha
*
* Program to determine the twinning fraction given the
* defined twinning operation (Twin Op. I)
*
* Kieron BROWN, Aug 1996
*
*****/

#include <stdio.h>
#include <math.h>
#include <stdlib.h>

#define number 40000

main()
{
    int    m, n, nmax, q, u;
    int    misym[number], batch[number];
    double h[number], k[number], l[number];
    int    h1, k1, l1;
    int    h2, k2, l2;

    double z[number][13];
    double alphaacen;
    double modh;
    double high, high2, high2total, high2mean;
    double i[number], sigi[number];
    double imean1, sigimean1;
    double imean2, sigimean2;

    char    record1[90], record2[90];
    char    ifn[50];

    FILE    *ifp1;

```

```

FILE    *ofp1;

printf("\n -----\n\n");
printf("    Program getAlpha : calculate twinning fractions\n");
printf("\n -----\n\n");

/* define the input filename */

printf("Enter the input filename: ");
scanf("%s", &ifn);
fflush(stdin);

/* open the reflection files */

ifp1 = fopen(ifn, "r");

/* read reflections into an array */

printf("\nReading reflection file %s\n\n", ifn);
n = 0;
u = 0;

while (fgets(record1, 90, ifp1) != NULL) {

/* format H K L M/ISYM BATCH I SIGI DELI SIGDELI
    0 1 2 3          5 6
*/
    sscanf(record1, "%d%d%d%d%lf%lf%*lf%*lf", &h[n], &k[n],
        &l[n], &misym[n], &batch[n], &i[n], &sigi[n]);

    if (misym[n] == 1.0 || misym[n] == 2.0) u += 1;

/* for blank line */

    fgets(record2, 90, ifp1);

    n += 1;
}

nmax = n;
printf("%d TOTAL number of valid records in FILE\n", nmax);
printf("Beginning algorithm\n\n");

for (n = 0; n < nmax; n++) {

/* only treat observations with m/isym = 0 */

    if (misym[n] == 0) {

        for (m = 0; m < nmax; m++) {

/*****
the TWINNING CRITERION is defined here for Tw Op I

            h, k, l -> -h, -k, l

        look for the twin */

if (h[n] == (-1*h[m]) && k[n] == (-1*k[m]) && l[n] == l[m]) {

    h1 = h[n];

```

```

k1 = k[n];
l1 = l[n];
imean1 = i[n];
sigimean1 = sigi[n];

h2 = h[m];
k2 = k[m];
l2 = l[m];
imean2 = i[m];
sigimean2 = sigi[m];

/* calculate H */

bigh = (imean1 - imean2)/(imean1 + imean2);
bigh2 = bigh*bigh;
modh = sqrt(bigh2);

/* ignore observations with negative intensities */

if (modh < 1.0) {
    bigh2total += bigh2;
    q += 1;
}
}
}
}

/* calculate mean of H^2 */

bigh2mean = bigh2total/(q - 1);

/* acentrics */

alphaacen = (1 - sqrt(3.0*bigh2mean))/2;

printf("twinning fraction = %4.2lf\n\n", alphaacen);
fclose(ifp1);
}

```

A.3 Program getRatio

```

/*
                                Program getRatio

to calculate the Wilson ratio, to check for
twinning.
The resolution is calculated before determining
the value of  $\langle I^2(h) \rangle / \langle I(h) \rangle^2$ 
for acentrics
equation is given in Acta Cryst D49 375

Kieron BROWN
July 1996

*/

#include <stdio.h>

```



```

#include <math.h>

#define acell 64.8
#define bcell 64.8
#define ccell 110.7
#define alpha 90.0
#define beta 90.0
#define gamma 120.0

main()
{
    int        h, k, l, n, nmax;

    double     pi, degtor;
    double     cosa, cosb, cosg, sina, sinb, sing;
    double     offset, vol;
    double     astar, bstar, cstar;
    double     ast2, bst2, cst2;
    double     csas, csbs, csgs, abgs, cabs, bcas;
    double     cell;
    double     stol2;
    double     s;

    float      ratio, low, high, resmax, resmin;
    float      i, sigi, i2, itotal, i2total;
    float      i2mean, imean, imean2;

    char       record[50], string[20];

    FILE       *ifp1;

    ifp1 = fopen("hkli.dat", "r");

/* define resolution range */

    printf("Resolution range? [low/high] ");
    gets(string);
    sscanf(string, "%f%f", &low, &high);

/* convert to sin(theta)/lambda */

    resmin = 1/(2*low);
    resmax = 1/(2*high);

/* set up for computation of s from h,k,l */

    pi = 3.1415927;
    degtor = pi / 180.0;
    cosa = cos(alpha*degtor);
    cosb = cos(beta*degtor);
    cosg = cos(gamma*degtor);
    sina = sin(alpha*degtor);
    sinb = sin(beta*degtor);
    sing = sin(gamma*degtor);
    offset = sqrt(1.-(cosa*cosa)-(cosb)*(cosb)-(cosg)*(cosg) + 2*cosa*cosb*cosg);
    vol = acell * bcell * ccell * offset;
    astar = (bcell * ccell * sina) / vol;
    bstar = (ccell * acell * sinb) / vol;
    cstar = (acell * bcell * sing) / vol;
    ast2 = astar * astar;
    bst2 = bstar * bstar;
    cst2 = cstar * cstar;

```

```

csas = (cosb*cosg-cosa) / (sinb*sing);
csbs = (cosg*cosa-cosb) / (sing*sina);
csgs = (cosa*cosb-cosg) / (sina*sinb);
abgs = 2.0 * astar * bstar * csgs;
cabs = 2.0 * cstar * astar * csbs;
bcas = 2.0 * bstar * cstar * csas;

/* Reflection loop */

n = 1;
itotal = 0.0;
i2total = 0.0;
while (fgets(record, 50, ifp1) != NULL) {

    sscanf(record, "%d%d%d%f%f", &h, &k, &l, &i, &sigi);

/* select acentrics */

    if (h != 0) {
        if (k != 0) {
if (k != (-1*h)) {

/* Calculate sin(theta)/lambda = s from h,k,l */

        stol2 = h*h*ast2 + k*k*bst2 + l*l*cst2 + h*k*abgs + l*h*cabs + k*l*bcas;
        s = 0.5 * sqrt(stol2);

        if (s < resmax && s > resmin) {
            i2 = i*i;
            itotal += i;
            i2total += i2;
            n += 1;
        }
    }
}
}
}
nmax = (n - 1);
printf("\nNumber of measurements = %d\n", nmax);
i2mean = i2total/nmax;
imean = itotal/nmax;
imean2 = imean*imean;

/* calculate the required ratio */

ratio = i2mean/imean2;

printf("<I^2>/<I>^2 (ACENTRICS) = %3.2f (2 -> 1.5)\n\n", ratio);

exit();
}

```

A.4 Program deTwin

```

/*
                                Program deTwinB

    general detwinning program

```

INPUT FILE:

~~~~~

mtzdump hklin agro.mtz

with records remerge as output

working with a single array defined as follows

| h | k | l | m/isym | batch | i | sigi | flag1 | icorr | sigicorr | delicorr | sigdelicorr | flag2 |    |
|---|---|---|--------|-------|---|------|-------|-------|----------|----------|-------------|-------|----|
| 0 | 1 | 2 | 3      | 4     | 5 | 6    |       | 7     | 8        | 9        | 10          | 11    | 12 |

OUTPUT FILE:

~~~~~

an ascii file either in format suitable for agrovata or truncate, requestde by program

NOTE

alpha must be greater than 0.00001 and less than or equal to 0.5 for HP machines!

Kieron BROWN
October 1996

*/

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
```

```
/* change this number depending upon the
   size of the reflection file */
```

```
#define number 100000
```

```
main()
```

```
{
```

```
  int    isym, m, n, nmax;
  int    q, p, r, s, t, u;
  float  a, b, c, d;
```

```
  double e, f, g, j, alpha;
  double z[number][13];
  double deli, sigdeli;
  double coef1, coef2, coef10, coef20;
  double wp, wn, iwmean, sigiwmean;
  double ibpos, sigibpos, ibneg, sigibneg;
  char   record1[90], record2[90], twstring[5], opstring[5];
  char   ifn[50], ofn[50];
```

```
  FILE   *ifp1;
```

```
  FILE   *ofp1;
```

```
  printf("\n ----- \n\n");
  printf("   Program deTwinB : general detwinning program \n");
  printf("\n ----- \n\n");
```

```
/* define the input filename */
```

```
  printf("Enter the input filename: ");
```

```

scanf("%s", &ifn);
fflush(stdin);

/* define the output filename */

printf("Enter the output filename: ");
scanf("%s", &ofn);
fflush(stdin);

/* set the twinning fraction */

printf("\nTwinning fraction [0 < a <= 0.5] : ");
gets(twstring);
sscanf(twstring, "%lf", &alpha);

/* define output format */

printf("Output format required [remerge/truncate] : ");
gets(opstring);

/* open the reflection file */

ifp1 = fopen(ifn, "r");

/* open the output file */

ofp1 = fopen(ofn, "w");

/* coefficients for the intensity algorithm */

a = (alpha/(2*alpha-1));
b = (1-alpha)/(2*alpha-1);
c = (1/alpha)*(1 + ((1-alpha)*(1-alpha)/(2*alpha-1)));
d = (1-alpha)/(2*alpha-1);

/* coefficients for the sigma algorithm (NOT IMPLEMENTED)

e = sqrt(alpha)/(2*alpha-1);
f = sqrt((1-alpha))/(2*alpha-1);
g = sqrt((1-alpha))*sqrt(alpha)/(2*alpha-1);
j = 1/sqrt(alpha)*((1-alpha)/(2*alpha-1)); */

/* the counters */

p = 1; /* B(+):B(-) pair */
q = 1; /* B(-):B(+) pair */
r = 1; /* Twin with no Bijvoet partner */
s = 1; /* B(+):T pairs */
t = 1; /* B(-):T pairs */
u = 1; /* */

/* read reflections */

printf("\nReading reflection file %s\n\n", ifn);
n = 0;

while (fgets(record1, 90, ifp1) != NULL) {

/* format H K L M/ISYM BATCH I SIGI DELI SIGDELI
      0 1 2 3 4 5 6
*/

```

```

    sscanf(record1, "%lf%lf%lf%lf%lf%lf%lf%*lf%*lf", &z[n][0],
    &z[n][1], &z[n][2], &z[n][3], &z[n][4], &z[n][5], &z[n][6]);

    if (z[n][3] == 1.0 || z[n][3] == 2.0) u += 1;

/* for blank line */

    fgets(record2, 90, ifp1);

/* initialise the detwin and o/p flags */

    z[n][12] = 0;
    z[n][7] = 0;
    n += 1;
}

nmax = n;
printf("%d TOTAL number of valid records in FILE\n", nmax);
printf("%d TOTAL number of individual Bijvoets\n\n", u-1);
printf("Beginning algorithm\n\n");

/* starting from first reflection */
/* here, n and m are twins */

    for (n = 0; n < nmax; n++) {

/* read detwin flag */

        if (z[n][7] == 0.0) {

/* read the Bijvoet flag : case 1 */

            if (z[n][3] == 1.0) {

/*****
            the TWINNING CRITERION is defined here for Tw Op I

                h, k, l -> -h, -k, l

            look for the twin */

for (m = 0; m < nmax; m++) {

            if (z[m][0] == z[n][0] && z[m][1] == z[n][1] && -1*z[m][2]
            == z[n][2] && z[m][3] == 1.0 && z[m][7] == 0.0) {

/*****/

/* calculate the new intensities */

                z[n][8] = a*z[m][5] - b*z[n][5];
                z[m][8] = c*z[n][5] - d*z[m][5];

/* the sigma algorithm */

                coef1 = sqrt((z[n][8]/z[n][5])*(z[n][8]/z[n][5]));
                coef2 = sqrt((z[m][8]/z[m][5])*(z[m][8]/z[m][5]));

                coef10 = sqrt(coef1);
                coef20 = sqrt(coef2);

                z[n][9] = coef10*z[n][6];

```

```

    z[m][9] = coef20*z[m][6];
/* update the del columns */
    z[n][10] = z[n][8];
    z[n][11] = z[n][9];
    z[m][10] = z[m][8];
    z[m][11] = z[m][9];
/* update the detwin flag */
    z[n][7] = 1.0;
    z[m][7] = 1.0;
    s += 1;
}
}

/* read the Bijvoet flag : case 2 */
    if (z[n][3] == 2.0) {
/*****
    the TWINNING CRITERION is defined here for Tw Op I
        h, k, l -> -h, -k, l
    look for the twin */
for (m = 0; m < nmax; m++) {
    if (z[m][0] == z[n][0] && z[m][1] == z[n][1] && -1*z[m][2]
        == z[n][2] && z[m][3] == 2.0 && z[m][7] == 0.0) {
/*****/
/* calculate the new intensities */
        z[n][8] = a*z[m][5] - b*z[n][5];
        z[m][8] = c*z[n][5] - d*z[m][5];
/* the sigma algorithm */
        coef1 = sqrt((z[n][8]/z[n][5])*(z[n][8]/z[n][5]));
        coef2 = sqrt((z[m][8]/z[m][5])*(z[m][8]/z[m][5]));
        coef10 = sqrt(coef1);
        coef20 = sqrt(coef2);
        z[n][9] = coef10*z[n][6];
        z[m][9] = coef20*z[m][6];
/* update the del columns */
        z[n][10] = z[n][8];
        z[n][11] = z[n][9];
        z[m][10] = z[m][8];
        z[m][11] = z[m][9];
/* update the detwin flag */
        z[n][7] = 1.0;

```

```

        z[m][7] = 1.0;
        t += 1;
    }
}
    }
}

/*****

/* calculate the weighted mean and sigma for B(+) and B(-) pairs */
/* here, n and m are Bijvoets */

    for (n = 0; n < nmax; n++){
        if (z[n][7] == 1.0 && z[n][3] == 1.0 && z[n][12] == 0.0) {

/* this is a Bivjoet(+) */

            for (m = 0; m < nmax; m++) {

if ((z[m][0] == z[n][0]) && (z[m][1] == z[n][1]) && (z[m][2]
== z[n][2]) && (z[m][3] == 2.0) && (z[m][7] == 1.0)) {

    wp = (1/(z[n][9]*z[n][9]));
    wn = (1/(z[m][9]*z[m][9]));

    iwmean = (z[n][8]*wp + z[m][8]*wn)/(wp + wn);
    sigiwmean = 1/(sqrt(wp + wn));

    isym = 0;
    deli = 0.0;
    sigdeli = 0.0;

/* output to file */

    if (strcmp(opstring, "r", 1) == 0) {
/* for reinput into AGROVATA = records remerge */
        fprintf(ofp1, "
%4d%4d%4d%5d%6d%10.2f%10.2f%10.2f%10.2f\n\n", (int)z[n][0],
%(int)z[n][1], (int)z[n][2], isym, (int)z[n][4], iwmean, sigiwmean,
%deli, sigdeli);
        fprintf(ofp1,
"%4d%4d%4d%5d%6d%10.2f%10.2f%10.2f%10.2f\n\n",
(int)z[n][0], (int)z[n][1], (int)z[n][2], (int)z[n][3],
(int)z[n][4], z[n][8], z[n][9], z[n][10], z[n][11]);
        fprintf(ofp1,
"%4d%4d%4d%5d%6d%10.2f%10.2f%10.2f%10.2f\n\n",
(int)z[m][0], (int)z[m][1], (int)z[m][2], (int)z[m][3],
(int)z[m][4], z[m][8], z[m][9], z[m][10], z[m][11]);
    }

    if (strcmp(opstring, "t", 1) == 0) {
/* for input into TRUNCATE = records truncate */
        fprintf(ofp1,
"%4d%4d%4d%12.2lf%12.2lf%12.2lf%12.2lf%12.2lf%12.2lf%12.2lf%12.2lf%12.2lf%12.2lf\n\n",
(int)z[n][0], (int)z[n][1],
(int)z[n][2], iwmean, sigiwmean, z[n][8], z[n][9],
z[n][8], z[n][9], z[m][8], z[m][9], z[m][8], z[m][9]);
    }

/* update o/p flags */

```

```

    z[n][12] = 1.0;
    z[m][12] = 1.0;
    p += 1;
}
}
}

    if (z[n][7] == 1.0 && z[n][3] == 2.0 && z[n][12] == 0.0) {
/* this is a Bijvoet(-) */

    for (m = 1; m < nmax; m++) {
if ((z[m][0] == z[n][0]) && (z[m][1] == z[n][1]) && (z[m][2]
== z[n][2]) && (z[m][3] == 1.0) && (z[m][7] == 1.0)) {

    wn = (1/(z[n][9]*z[n][9]));
    wp = (1/(z[m][9]*z[m][9]));

    iwmean = (z[n][8]*wn + z[m][8]*wp)/(wn + wp);
    sigiwmean = 1/(sqrt(wn + wp));

    isym = 0;
    deli = 0.0;
    sigdeli = 0.0;

/* output to file */

    if (strcmp(opstring, "r", 1) == 0) {
/* for reinput into AGROVATA = records remerge */
        fprintf(ofp1,
            "%4d%4d%4d%5d%6d%10.21f%10.21f%10.21f%10.21f\n\n",
            (int)z[n][0], (int)z[n][1], (int)z[n][2], (int)isym,
            (int)z[n][4], iwmean, sigiwmean, deli, sigdeli);
        fprintf(ofp1, " %4d%4d%4d%5d%6d%10.21f%10.21f%10.21f
            %10.21f\n\n", (int)z[m][0], (int)z[m][1], (int)z[m][2],
            (int)z[m][3], (int)z[m][4], z[m][8], z[m][9], z[m][10],
            z[m][11]);
        fprintf(ofp1,
            "%4d%4d%4d%5d%6d%10.21f%10.21f%10.21f%10.21f\n\n",
            (int)z[n][0], (int)z[n][1], (int)z[n][2], (int)z[n][3],
            (int)z[n][4], z[n][8], z[n][9], z[n][10], z[n][11]);
    }

    if (strcmp(opstring, "t", 1) == 0) {
/* for input into TRUNCATE = records truncate */
        fprintf(ofp1,
            "%4d%4d%4d%12.21f%12.21f%12.21f%12.21f%12.21f%12.21f
            %12.21f%12.21f%12.21f%12.21f\n", (int)z[n][0],
            (int)z[n][1], (int)z[n][2], iwmean, sigiwmean, z[m][8],
            z[m][9], z[m][8], z[m][9], z[n][8], z[n][9], z[n][8],
            z[n][9]);
    }

/* update o/p flags */

    z[n][12] = 1.0;
    z[m][12] = 1.0;
    q += 1;
}
}
}
}
}

```



```

}

for (n = 0; n < nmax; n++){
  if (z[n][7] == 1.0 && z[n][12] == 0.0) {
    isym = 0;
    deli = 0.0;
    sigdeli = 0.0;

    if (strcmp(opstring, "r", 1) == 0) {
/* for reinput into AGROVATA = records remerge */
fprintf(ofp1, "%4d%4d%4d%5d%6d%10.2lf%10.2lf%10.2lf%10.2lf\n",
          (int)z[n][0], (int)z[n][1], (int)z[n][2],
          (int)isym, (int)z[n][4], z[n][8], z[n][9], deli, sigdeli);

fprintf(ofp1,
        "%4d%4d%4d%5d%6d%10.2lf%10.2lf%10.2lf%10.2lf\n",
        (int)z[n][0], (int)z[n][1], (int)z[n][2], (int)z[n][3],
        (int)z[n][4], z[n][8], z[n][9], z[n][10], z[n][11]);
        r += 1;
    }

    if (strcmp(opstring, "t", 1) == 0) {
/* for input into TRUNCATE = records truncate */
if (z[n][3] == 1.0) {
  ibneg = 0.0;
  sigibneg = 0.0;
  fprintf(ofp1, "%4d%4d%4d%12.2lf%12.2lf%12.2lf%12.2lf\n",
              (int)z[n][0],
              (int)z[n][1], (int)z[n][2], z[n][8], z[n][9], z[n][8],
              z[n][9], ibneg, sigibneg, ibneg, sigibneg, ibneg,
              sigibneg);
  r += 1;
}
if (z[n][3] == 2.0) {
  ibpos = 0.0;
  sigibpos = 0.0;
  fprintf(ofp1, "%4d%4d%4d%12.2lf%12.2lf%12.2lf%12.2lf\n",
              (int)z[n][0],
              (int)z[n][1], (int)z[n][2], z[n][8], z[n][9], ibpos,
              sigibpos, ibpos, sigibpos, z[n][8], z[n][9], ibpos,
              sigibpos);
  r += 1;
}
}
}
}

/* print statistics */

printf("%d B(+):T pairs\n", (s-1));
printf("%d B(-):T pairs\n", (t-1));
printf("-----\n");
printf("%d B(+):B(-) pairs\n", (p-1));
/* printf("%d B(-):B(+) pairs\n", (q-1)); */
printf("%d Twins with no Bijvoet partner\n", (r-1));
printf("-----\n");
printf("%d TOTAL number observations detwinned\n", 2*(s+t-2));
printf("\nCompleted calculation\n\n");

fclose(ifp1);
fclose(ofp1);
}

```

Appendix B

The Fast Image Plate Detector Software

The fast image plate detector control programs.

- carousel.mac
the principal program which manages the detector
- carousel_menu.c
slave program to carousel.mac
- washing.c
allocates image number to the image file
- BarCode.c
communicates with the barcode reader at the detector
- DataBase.c
communicates with the database
- mysql_strings.h
a library of regularly used strings
- experiment_table_viewer.phtml
the database display program

B.1 Program carousel

```
# initial : 5.96
# author: Kieron Brown
#####
#%TITLE% Macros for BL19 carousel.
# %NAME% CAROUSEL.MAC - BL19 carousel
#####
# %DESCRIPTION%
#%OVERVIEW%
#%END%
#####
#
# list of modifications 23/8/96 KB
#
# line 73: changed cur_image=0 to cur_image=CAR_STARTIMAGENUM
```

```

# line 90: eState no longer updated at beginning of exposure
#           now at end, line 126
# line 126: exposure time written at end of exposure
#

global CAR_DATABASE CAR_IMAGEPREFIX CAR_STARTIMAGENUM CAR_PLATESNUM
global CAR_IMAGESNUM CAR_PASSESNUM CAR_OSCILLATION
global CAR_OVERLAP CAR_EXPOSTIME CAR_STARTANGLE CAR_IMAGEDIR

def carousel '
{

if ($#) {
CAR_DATABASE           = "$1"
CAR_IMAGEPREFIX       = "$2"
CAR_STARTIMAGENUM     = $3
CAR_PLATESNUM         = $4
CAR_IMAGESNUM         = $5
CAR_PASSESNUM         = $6
CAR_EXPOSTIME         = $7
CAR_STARTANGLE        = $8
CAR_OSCILLATION       = $9
CAR_OVERLAP           = $10
CAR_IMAGEDIR          = "$11"
}
else {
CAR_DATABASE = sprintf("%s", getval("Database name: ", CAR_DATABASE))
CAR_IMAGEPREFIX = sprintf("%s", getval("Filename prefix: ",
CAR_IMAGEPREFIX))
CAR_STARTIMAGENUM = getval("Starting frame number: ", CAR_STARTIMAGENUM)
CAR_PLATESNUM = getval("Number of plates presently on carousel: ",
CAR_PLATESNUM)
CAR_IMAGESNUM = getval("Total number of images: ", CAR_IMAGESNUM)
CAR_PASSESNUM = getval("Number of passes per image: ", CAR_PASSESNUM)
CAR_EXPOSTIME = getval("Exposure time per pass (s): ", CAR_EXPOSTIME)
CAR_STARTANGLE= getval("Starting angle (degrees): ", CAR_STARTANGLE)
CAR_OSCILLATION = getval("Oscillation range (degrees): ",
CAR_OSCILLATION)
CAR_OVERLAP = getval("Oscillation overlap (degrees): ", CAR_OVERLAP)
}

printf("Database name: %s\n", CAR_DATABASE)
printf("Filename prefix: %s\n", CAR_IMAGEPREFIX)
printf("Starting frame number: %d\n", CAR_STARTIMAGENUM)
printf("Number of plates: %d\n", CAR_PLATESNUM)
printf("Total number of images: %d\n", CAR_IMAGESNUM)
printf("Number of passes per image: %d\n", CAR_PASSESNUM)
printf("Exposure time per pass (s): %f\n", CAR_EXPOSTIME)
printf("Starting angle (degrees): %f\n", CAR_STARTANGLE)
printf("Oscillation range (degrees): %f\n", CAR_OSCILLATION)
printf("Oscillation overlap (degrees): %f\n", CAR_OVERLAP)

#
# SELECT a database case(1)
#

str1 = sprintf("1 %s", CAR_DATABASE)
data_pipe("carousel_menu", str1)

#
# IMPORT the serial line case(2)
#

```

```

data_pipe("carousel_menu", "2")

#
# Undef the "mono" motor
#
pseudodel mono
rdef cleanup1 \
    undef cleanup1; mthpseudomotor mono
\

#
# begin image number loop
#
image_cnt = 1
for (cur_image = CAR_STARTIMAGENUM; cur_image < (CAR_STARTIMAGENUM +
    CAR_IMAGESNUM);) {
    for (cur_plate = 1; cur_plate <= CAR_PLATESNUM; cur_plate++) {
        printf("current plate is %d\n", cur_plate)
        printf("current image is %d\n", cur_image)

        n = 0
        while (1) {
            n += 1
#
# READ barcode case(3)
#

            barcode = data_pipe("carousel_menu", "3")
            printf("barcode is %d\n", barcode)
            if (barcode == 0) {
                printf("EXITING\n")
                undef cleanup1
                mthpseudomotor mono
                exit
            }
#
# CHECK scan_state case(4)
#

            str4 = sprintf("4 %d", barcode)

            scanned = data_pipe("carousel_menu", str4)

#
# scanned is 2 => not scanned
#
#           1 => scanned

            if (scanned == 2) {

# rotate to next plate if this one
# is not scanned

                shutdown 14
                printf("Rotating carousel\n")
                sleep(3)
                shutdown 14

                if (n == 7) {
                    printf("No suitable plates found\nExit\n")
                    undef cleanup1

```

```

        mthpseudomotor mono
        exit
    }
}
if (scanned == 1) break
}

start_angle = CAR_STARTANGLE + ((image_cnt-1)*CAR_OSCILLATION)
stop_angle = start_angle + CAR_OSCILLATION
expo_time = CAR_EXPOSTIME

#
# While statement to make the passes for each frame
#
    printf("taking exposure\n")
    curr_pass = 0
    while (curr_pass < CAR_PASSESNUM) {
        printf("Pass %d:\n", curr_pass+1)

#
# take the exposure (phi.mac)
#

        if (curr_pass == 0) {
# first pass
            phi_oscillation start_angle stop_angle expo_time 1
        }
        else {
            phi_oscillation start_angle stop_angle expo_time 0
        }

        curr_pass++
    }

#
# WRITE imNumber/barcode/headerFlag/startAngle
# and eState case(5)
#

    str5 = sprintf("5 %d %d %6.2f", barcode, cur_image, start_angle)
    data_pipe("carousel_menu", str5)

#
# WRITE parameters case(6)
#
# filePrefix CAR_IMAGEPREFIX
# oscRange CAR_OSCILLATION
# expTime totalexptime
# wavelength
# distance
# date
#

    waitmove ; get_angles
    distance = A[taux]
# mono not connected to SPEC
#    waitmove ; get_angles ; calcE
#    wavelength = LAMBDA
    LAMBDA = 0.0

    totalexptime = CAR_EXPOSTIME * CAR_PASSESNUM
    str6 = sprintf("6 %d %s %f %f %f %f", barcode, CAR_IMAGEPREFIX, \
CAR_OSCILLATION, totalexptime, LAMBDA, distance)

```

```

data_pipe("carousel_menu",str6)

if (cur_plate < CAR_PLATESNUM) {
# ROTATE carousel, the carousel relay is on address 14

    shutope 14
    printf("Rotating carousel...\n")
    sleep(3)
    shutclose 14
}
# increase image number and image counter

    cur_image++
    image_cnt += 1
}

# all the plates have been exposed but not all the images

if (cur_image < CAR_IMAGESNUM) {
    todo = CAR_IMAGESNUM - cur_image
    printf("\nimages remaining : %d\nRELOAD carousel\n", todo)
    CAR_PLATESNUM = getval("Number of plates on the carousel? : ",
    CAR_PLATESNUM)

    if (CAR_PLATESNUM > todo) {
        printf("only using %d plates\n", todo)
        CAR_PLATESNUM = todo
    }
    if (CAR_PLATESNUM == 0) {
        undef cleanup1
        mthpseudomotor mono
        printf("\n--- Data collection complete ---\n\n")
        exit
    }
}
}

#
# Rodef "mono" motor
#
    undef cleanup1
    mthpseudomotor mono

# the data collection has been completed

    printf("\n--- Data collection complete ---\n\n")
}'

```

B.2 Program carousel_menu

```

/*
        carousel_menu.c

```

functions called by SPEC via the data_pipe() function

```
case(1) : select database (db_select)
case(2) : import serial line (config_carousel)
case(3) : read barcode (Car_trigger)
case(4) : check scan state, write exp state (f1)
case(5) : write image_num/barcode/header-flag (image_number)
case(6) : write parameters
```

NOTE ON SPEC

the arguments passed from SPEC in the data_pipe() function are

arg[0] first argument NOT function name

.

so this is different from C

20 May 1996
Kieron BROWN

Modifications 23/8/96 KB

Line 251: eState now updated when parameters bc/startAngle etc written to database

*/

```
#include <Admin.h>
#include <API.h>
#include <DevServer.h>
#include <SerialLine.h>
#include <stdio.h>
#include <mysql.h>
#include <mysql_strings.h>
#include <user_pipe.h>
#include <time.h>

int          DB, SC, Q1;
int          cmd, up = 1, down = 0, socket;
int          bc, n, image, e_time;
float        startAngle, oscRange, exposureTime, wavelength;
float        distance;
double       r;
char         str10[70], str11[70], str12[70];
char         str13[70], str14[150], str20[100];
char         filePrefix[20], sState[];
char         curDate[40], database[20];

time_t       timer;
m_row        row;
m_result     *res;

long         readwrite = 0;
devserver    ds1, ds2;
long         error;
short        devstatus;
```

```

DevLong    readstring_argin1, readstring_argin2;
DevString  readstring_argout1, readstring_argout2;
DevString  writestring_argin;
DevLong    writestring_argout;
DevString  retstr1;

user_code(argc, argv)
/* int argc;*/
char **argv;
{
    sscanf(argv[0], "%d", &cmd);

    switch(cmd)
    {
        case(1) :

/* connect to local host computer, where the mSQL daemon is running */

        printf("\n-----\n\nStart of data
collection\n");

        if (argc < 2) {
printf("case(1) :: Usage : <case> <database> \n");
do_quit_return;
        }

        sscanf(argv[1], "%s", &database);

        /* printf("\n==> carousel_menu : msqlConnect\n"); */
        if ((socket = msqlConnect("")) == -1) {
printf("%s\n", msqlErrMsg);
        }

/* select the database */

        /* printf("==> carousel_menu : msqlSelectDB\n"); */
        if ((DB = msqlSelectDB(socket, database)) == -1 ) {
printf("%s\n", msqlErrMsg);
        }
        break;

        case(2) :

/* import a serial line for the carousel reader */

        /* printf("==> carousel_menu : dev_import serial line\n");
        */
        devstatus = dev_import("D14/d142ser/carbcode",
readwrite,
&ds1,
&error);
        /* printf("\ndev_import(%s) returned %d\n\n",
        "d14/d142ser/carbcode",devstatus); */
        if(devstatus != 0) {
printf("Error message: %s\n",dev_error_str(error));
        printf("Possible reason : problem with SERIAL LINE to carousel\n");
exit(1);
        }
        break;
    }
}

```



```

    case(3) :

/* flush the barcode buffer */

    /*      printf("==> carousel_menu : flush barcode buffer\n"); */
    readstring_argin1 = SL_RAW;
    devstatus = dev_putget(ds1, DevSerReadString,
&readstring_argin1,
D_LONG_TYPE,
&readstring_argout1,
D_STRING_TYPE,
&error);
    if(devstatus < 0) {
/* printf("\nDevSerReadString dev_putget() returned %d\n",
devstatus);
printf("Error message: %s\n", dev_error_str(error)); */
    }
    free(readstring_argout1);

/* trigger the barcode reader */

    /*      printf("==> carousel_menu : triggering barcode
reader\n"); */
    writestring_argin = "S";
    devstatus = dev_putget(ds1, DevSerWriteString,
&writestring_argin,
D_STRING_TYPE,
&writestring_argout,
D_LONG_TYPE,
&error);
    if (devstatus < 0) {
printf("\nDevSerWriteString dev_putget() returned %d\n",
devstatus);
printf("Error message: %s\n", dev_error_str(error));
    }
    else {
/* printf("\n%d\n", writestring_argout); */
    }

/* allow enough time to scan barcode */

    sleep(1);

/* read the buffer take only the first 2 bytes */

    /*      printf("==> carousel_menu : read barcode buffer\n"); */
    readstring_argin2 = SL_NCHAR | (3 << 8);
    devstatus = dev_putget(ds1, DevSerReadString,
&readstring_argin2,
D_LONG_TYPE,
&readstring_argout2,
D_STRING_TYPE,
&error);
    if (devstatus < 0) {
/* printf("\nDevSerReadString dev_putget() returned %d\n",
devstatus);
printf("Error message: %s\n", dev_error_str(error)); */
printf("\nNO PLATE in exposure position\n");
    }

/* stop the barcode reader */

/* printf("==> carousel_menu : stop barcode reader\n"); */

```

```

writestring_argin = "E";
devstatus = dev_putget(ds1, DevSerWriteString,
    &writestring_argin,
    D_STRING_TYPE,
    &writestring_argout,
    D_LONG_TYPE,
    &error);
if (devstatus < 0) {
    printf("\nDevSerWriteString dev_putget() returned %d\n",
        devstatus);
    printf("Error message: %s\n", dev_error_str(error));
}
else {
    /* printf("\n%d\n", writestring_argout); */
}

/* again allow enough time to send the character */

sleep(1);

set_return_value(0.0);
do_quit_return();
}

/* return the barcode value */

    /*      printf("==> carousel_menu : readstring_argout2
    (barcode) is %s\n", readstring_argout2); */
    /*      printf("==> carousel_menu : ret str is %d\n",
    set_return_string(readstring_argout2)); */
    if (set_return_string(readstring_argout2) != 0) {
free(readstring_argout2);
printf("Unable to return string\n");
do_quit_return();
}
/* next line PROBLEM */
    readstring_argout2 = NULL;
    free(readstring_argout2);
    break;

    case(4) :

/* confirm that the plate has been scanned */

    if (argc < 2) {
printf("case(4)::Usage: <case> <barcode>\nUnable to confirm \
scan state\nContinue\n");
do_quit_return();
}

    sscanf(argv[1], "%d", &bc);

    sprintf(str10, "%s %d", SEL_S_STATE_STR, bc);

/* str10 :
SELECT sState
FROM status
WHERE plNumber = 23
*/

    if ((Q1 = mysqlQuery(socket, str10)) == -1) {

```

```

printf("%s\n", mysqlErrMsg);
}
res = mysqlStoreResult();
row = mysqlFetchRow(res);

/*      printf("==> carousel_menu : scan confirmation\n"); */

sprintf(sState, "%c", *row[0]);
/*      printf("sState = %s\n", sState); */
if (strcmp(sState,"0",1) == 0) { /* scan flag=0 */
mysqlFreeResult(res);
/* mysqlClose(); */
printf("plate not scanned\n");

/* pass back value to SPEC */

r = 2.0;
set_return_value(r);
}
else if (strcmp(sState,"0",1) == 1) {
printf("image plate is clean\n");
mysqlFreeResult(res);

/* pass back value=1 (true) to SPEC */

r = 1.0;
set_return_value(r);
}
break;

case(5) :

/* put image number/barcode/headerFlag=0/startAngle into experiment
table and update eState and write eTime in status table */

if (argc < 4) {
printf("case(5)::Usage: <case> <barcode> <image> <startAngle>\
\nUnable to write image number\nExit");
exit();
}

sscanf(argv[1], "%d", &bc);
sscanf(argv[2], "%d", &image);
sscanf(argv[3], "%f", &startAngle);

sprintf(str13, "%s%d, %d, %d, %7.3f%s", INS_STRA, image, bc, down, \
startAngle, END_STR);
/* str13 :
INSERT INTO images (image, implNumber, flag, startAngle)
VALUES (1, 23, 0, 12)
*/

if ((Q1 = mysqlQuery(socket, str13)) == -1) {
printf("%s\n", mysqlErrMsg);
}

/*      mysqlFreeResult(res);*/
/*      printf("==> carousel_menu : image/barcode/flag written
to database\n"); */

/* set eState flag up, sState flag down */

```

```

        sprintf(str11, "%s%s%d, %s%d %s%d" , UP_STRA, UP_STRB, up,
        %UP_STRD, down, UP_STRC, bc);
/* str11 :
    UPDATE status
    SET eState = 1,
        sState = 0
    WHERE imNumber = 23
*/

        if ((Q1 = mysqlQuery(socket, str11)) == -1) {
printf("%s\n", mysqlErrMsg);
        }

/* put eTime into status table */

        e_time = time(NULL);

        sprintf(str12, "%s%s%d %s%d" , UP_STRA, UP_STRE, e_time, UP_STRC, bc);
/* str12 :
    UPDATE status
    SET eTime = 8300000
    WHERE plNumber = 23
*/

        if ((Q1 = mysqlQuery(socket, str12)) == -1) {
printf("%s\n", mysqlErrMsg);
        }
        break;

        case(6) :
/* put parameters into the experiment table */

        if (argc < 7) {
printf("case(6)::Usage: <case> <barcode> <filePrefix> \
<oscRange> <exposureTime> <wavelength> <distance> \
\nUnable to write parameters\nExit\n");
exit();
        }

        sscanf(argv[1], "%d", &bc);
        sscanf(argv[2], "%s", &filePrefix);
        sscanf(argv[3], "%f", &oscRange);
        sscanf(argv[4], "%f", &exposureTime);
        sscanf(argv[5], "%f", &wavelength);
        sscanf(argv[6], "%f", &distance);

        time(&timer);
        sprintf(curDate, "%s", ctime(&timer));
        printf ("%s\n", curDate);

        sprintf(str14, "%sfilePrefix = '%s', oscRange = %5.3f, expTime =
        %3.1f, wavelength = %6.4f, distance = %5.1f, date = '%s' %s%d
        %s%d", UP_STRH, filePrefix, oscRange, exposureTime, wavelength,
        distance, curDate, UP_STRJ, bc, UP_STRK, down);

        /* printf("str14 = %s\n", str14); */

/* str14 :
    UPDATE experiment
    SET filePrefix = toto,

```

```

        startAngle = 0,
        oscRange = 2,
        expTime = 30,
        wavelength = 0.99,
        distance = 250,
        curDate = Mon
WHERE plNumber = 23 AND headerFlag = 0
*/

        if ((Q1 = mysqlQuery(socket, str14)) == -1) {
printf("ERROR, Q1 = %d\n", Q1);
printf("%s\n", mysqlErrMsg);
mysqlFreeResult(res);
        }
        else {
/* mysqlFreeResult(res);*/
        }

        printf("\n==> experimental parameters written to database\n");
        printf("\n-----\n");
        break;

        default :
        break;
    }
return;
}

```

B.3 Program washing

```

/*
        washing

Image is written to the filename given in the Fuji software.
This file is stored in a temporary directory.

Program scanner will rename each file according to the information
in the 'experiment' table in the database, and move the file to a
new directory

communication with the database

status table          -> READ exposure time

experiment table      -> FIND image number
                       WRITE header flag (1)
                       WRITE decay-time

Kieron BROWN
EMBL-Grenoble
*/

#include <stdio.h>
#include <time.h>

```

```

main()
{
    extern int    import_scanner();
    extern int    confirm_expo();
    extern int    getetime();
    extern int    decaytime();
    extern int    getimage();
    extern int    headerflag();
    extern int    Sc_trigger();
    extern int    db_select();
    extern void   cleanup_db();

    int          bc;
    int          v = 1;
    int          e_time, s_time, decay_t;
    int          im_number;
    char          temp_file[20], new_file[20], cmd[30], filePrefix[20];
    char          database_name[20];

    printf("\n
-----\n\n");
    printf("      washing program : reads barcode at the Fuji
scanner\n");
    printf("\n
-----\n\n");

    /* select a database */

    /* printf("==> selecting the database\n"); */
    printf("Database name : ");
    scanf("%s", &database_name);

    if (db_select(database_name) == -1) {
        printf("ERROR in function db_select.\nExiting...\n");
        cleanup_db();
        exit();
    }

    /* file prefix */

    printf("Give filename prefix : ");
    scanf("%s", &filePrefix);

    /* import the serial line */

    /* printf("==> importing serial line\n"); */
    if (import_scanner() != 1) {
        printf("cannot import the serial line\nExit\n");
        exit();
    }

    while (1) {

    /* read barcode at the scanner */

        /* printf("==> triggering barcode\n"); */
        Sc_trigger(&bc);

        /* if (Sc_trigger(&bc) != 1) {

```

```

    printf("ERROR:  problem in Sc_trigger.\n");
    cleanup_db();
    exit();
}
*/
/* check if this is the killer barcode */

    /*      printf("==> checking for killer barcode\n"); */

    if (bc == 55) {
printf("---- Program terminated by killer barcode ----\n\n");
exit();
    }

/* confirm that plate has been exposed */

    v = confirm_expo(bc);

    /*      printf("==> confirming_expo\n"); */
    if (v == -1) {
printf("Check database setup.\nExiting...\n");
cleanup_db();
exit();
    }
    if (v == 1) {

printf("ERROR:  plate has not been exposed.\n
\nIntroduce next plate.\n");
    }
    if (v == 2) {

/* read eTime from status table */

/* printf("==> getting eTime\n"); */
if (getetime(bc, &e_time) == -1) {
    printf("Check database setup.\nExiting...\n");
    cleanup_db();
    exit();
}

/* read scan time */

s_time = time(NULL);

/* calculate delta_time */

decay_t = (s_time) - (e_time);

/* get the image number for this barcode */

/* printf("==> getting image number\n"); */
if (getimage(bc, &im_number, filePrefix) == -1) {
    printf("ERROR in g4.  Unable to read image number.\nExiting...\n");
    cleanup_db();
    exit();
}

/* print the filename for the user */

printf("\n
-----\n\n");
if (im_number < 10) {

```

```

    printf("    the filename for this image is :
%s_00%d.IMG\n", filePrefix,
im_number);
}
if (im_number >= 10 && im_number < 100) {
    printf("    the filename for this image is : %s_0%d.IMG\n",
filePrefix,
im_number);
}
if (im_number >= 100) {
    printf("    the filename for this image is : %s_%d.IMG\n",
filePrefix,
im_number);
}
printf("\n
-----\n");

/* put decayT into experiment table */

/* printf("=> putting decayT into exp. table\n"); */
if (decaytime(decay_t, bc, im_number, filePrefix) == -1) {
    printf("Check database setup.\nExiting...\n");
    cleanup_db();
    exit();
}

/* change header_flag to up */

/* printf("=> changing headerFlag\n"); */
if (headerflag(bc, im_number, filePrefix) != 0) {
    printf("Check database setup.\nExiting...\n");
    cleanup_db();
    exit();
}
    } /* expo_state confirmed */

/* return for new barcode */

    printf("Ready for next plate\nHit c to continue, q to quit ");
    fflush(stdin);
    if (getc(stdin) == 'q') break;
    fflush(stdin);
}

printf("\n\nProgram complete\n");
exit();
}

```

B.4 Program DataBase

```

/*
        DataBase

    functions which communicate with the database

```



```

    Kieron BROWN
    EMBL-Grenoble
*/

#include <stdio.h>
#include <mysql.h>
#include <mysql_strings.h>

int          socket;
int          up = 1, down = 0;
int          Q1;
m_row       row;
m_result    *res;
/* res is a result handle.  It is a pointer to an m_result structure */

void         cleanup_db();

int db_select(char database_name[20])
{
    int DB;

    /* connect to local host computer, where the mSQL daemon is running */

    if ((socket = mysqlConnect("")) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }

    /* select the database */

    if ((DB = mysqlSelectDB(socket, database_name)) == -1 ) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }
    return;
}

int confirm_expo(int bc)
/* confirm plate has been exposed */
{
    char str16[70], str17[70];
    int SC;

    sprintf(str16, "%s%d", SEL_E_STATE_STR, bc);
    /* str16 :  SELECT eState FROM experiment
       WHERE plNumber = 23 */

    if ((Q1 = mysqlQuery(socket, str16)) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }
    res = mysqlStoreResult();
    row = mysqlFetchRow(res);
    mysqlDataSeek(res, 1);

    if (SC = strcmp(row[0], "0", 1) == 0) {    /* expose flag=0 */
        return(1);
    }
    else {
        mysqlFreeResult(res);
    }

    /* set e_state down, s_state up */
}

```

```

        sprintf(str17, "%s%s%d, %s%d %s%d", UP_STRA, UP_STRB, down,
        %UP_STRD, up, UP_STRC, bc);
/* str17 : UPDATE status
        SET eState = 0, sState = 1
        WHERE plNumber = 23
*/
    if ((Q1 = mysqlQuery(socket, str17)) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }

}
return(2);
}

int getetime(int bc, int *pe_time)
/* read eTime from status table */
{
    char str18[70];

    sprintf(str18, "%s%d", SEL_ETIME_STR, bc);
/* str18 : SELECT eTime
        FROM status
        WHERE plNumber = 23
*/

    if ((Q1 = mysqlQuery(socket, str18)) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }
    res = mysqlStoreResult();
    row = mysqlFetchRow(res);

    sscanf(row[0], "%d", pe_time);
    return;
}

/*****/

int decaytime(int decay_t, int bc, int image, char filePrefix[20])
/* put decayT into experiment table */
{
    char str19[100];
    int n;

    sprintf(str19, "%s%s%d %s%d %s%d %s'%s'", UP_STRH, UP_STRN, decay_t,
    UP_STRJ, bc, UP_STRL, image, UP_STRM, filePrefix);

/* str19 : UPDATE experiment
        SET decayT = 180
        WHERE plNumber = 23
        AND imNumber = 1
        AND filePrefix = 'tonton'
*/
    if ((Q1 = mysqlQuery(socket, str19)) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }
}
}

```

```

/*****/

int getimage(int bc, int *pimage, char filePrefix[20])
/* read image number from experiment table */
{
    char str20[120];

    sprintf(str20, "%s %s%d %s%d %s'%s'", SEL_IMAGE_STR, UP_STRC, bc,
        UP_STRK, down, UP_STRM, filePrefix);

/* str20 : SELECT imNumber
            FROM experiment
            WHERE plNumber = 23
            AND headerFlag = 0
            AND filePrefix = 'toton' */

    if ((Q1 = mysqlQuery(socket, str20)) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }

    res = mysqlStoreResult();
    row = mysqlFetchRow(res);

    sscanf(row[0], "%d", pimage);
    return;
}

/*****/

int headerflag(int bc, int image, char filePrefix[20])
/* change headerFlag to up */
{
    char str21[100];
    int n;

    sprintf(str21, "%s%s%d %s%d %s%d %s'%s'", UP_STRH, UP_STRI, up,
        UP_STRJ, bc, UP_STRL, image, UP_STRM, filePrefix);

/* str21 : UPDATE experiment
            SET headerFlag = 1
            WHERE plNumber = 23
            AND imNumber = 1
            AND filePrefix = 'tonton'
*/
    if ((Q1 = mysqlQuery(socket, str21)) == -1) {
        printf("%s\n", mysqlErrMsg);
        return(-1);
    }
    return(0);
}

/*****/

void cleanup_db()
/* close the connection to the database */
{
    mysqlFreeResult(res);
    mysqlClose(socket);
    return;
}

```

B.5 Program mysql_strings.h

```
/* NOTE this has been written for latest BL19 version */
#define INS_STRA "INSERT INTO experiment (imNumber, plNumber,
headerFlag, startAngle) VALUES ("
#define UP_STRA "UPDATE status SET "
#define UP_STRB "eState = "
#define UP_STRC "WHERE plNumber = "
#define UP_STRD "sState = "
#define UP_STRE "eTime = "
#define UP_STRH "UPDATE experiment SET "
#define UP_STRJ "headerFlag = "
#define UP_STRK "WHERE plNumber = "
#define UP_STRL "AND headerFlag = "
#define UP_STRM "AND imNumber = "
#define UP_STRN "AND filePrefix = "
#define END_STR ")"
#define SEL_S_STATE_STR "SELECT sState FROM status WHERE plNumber = "
#define SEL_E_STATE_STR "SELECT eState FROM status WHERE plNumber = "
#define SEL_IMAGE_STR "SELECT imNumber FROM experiment"
#define SEL_ETIME_STR "SELECT eTime FROM status WHERE plNumber = "

/*
List of strings, for reference
NOTE we take the following parameters in these examples
                image number      = 1
                barcode            = 23
eTime          = 8300000
deltaT         = 300
number exposures = 30
up             = 1
down          = 0

str10 : SELECT sState FROM status WHERE plNumber = 23
str11 : UPDATE status SET eState = 1, sState = 0 WHERE plNumber = 23
str12 : UPDATE status SET eTime = 8300000 WHERE plNumber = 23
str13 : INSERT INTO experiment (image, plNumber, headerFlag,
startAngle) VALUES (1, 23, 0, 12)
str14 : UPDATE experiment SET filePrefix = toto, startAngle = 0,
oscRange = 2, expTime = 30, wavelength = 0.99, distance = 250, date = Mon
WHERE plNumber = 23 AND headerFlag = 0
str16 : SELECT eState FROM status WHERE plNumber = 23
str17 : UPDATE status SET eState = 0, sState = 1 WHERE plNumber = 23
str18 : SELECT eTime FROM status WHERE plNumber = 23
str19 : UPDATE experiment SET headerFlag = 1 WHERE plNumber = 23
AND imNumber = 1 AND filePrefix = tonton
str20 : SELECT imNumber FROM experiment WHERE plNumber = 23
AND headerFlag = 0 AND filePrefix = tonton
*/
```

B.6 Program experiment_table_viewer.phtml

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 3.0//EN">
<html> <head>
<title></title>
</head>
```

```

<body>
<h1></h1>

<?
  $dbhost = "bl19.esrf.fr";
  $table = "experiment";
>

<?setlogging(0)>
<?setshowinfo(0)>

<center>

<table border>
<tr>
  <th>Database</th>
</tr>

  <form action="<?echo $PHP_SELF"> TARGET="_self" method="POST">
<td>
<select name="dbname" size=5>
<?
  if ($dbhost);
  mysql_connect($dbhost);
  $dbs_res = mysql_listdbs();
  $num_dbs = mysql_numrows($dbs_res);
  $i = 0;
  while($i < $num_dbs);
    echo "<option>";
    echo mysql_dbname($dbs_res,$i);
    echo "\n";
    $i++;
  endwhile;
  endif;
>
</select>
<br>
<input type="submit" value=" Select Database ">
</td>
</form>

</table>

<? if ($dbname); >

<center>
<h2>Data Collection Information</h2>
<table border="2">

<?
mysql_connect($dbhost);
$result = mysql("$dbname", "select * from $table");
echo "<tr>";
$fields = mysql_numfields($result);
$i = 0;
while($i < $fields);
  echo "<th>";
  $field_names[$i] = mysql_fieldname($result,$i);
  echo $field_names[$i];
  echo "</th>";
  $i++;

```

```

endwhile;
echo "</tr>\n";

$rows = mysql_numrows($result);
$i = 0;
while($i < $rows);
  echo "<tr>";
  $f = 0;
  while($f < $nfields);
    echo "<td align=center>";
    $dummy = mysql_result($result,$i,$field_names[$f]);
    if (!strval($dummy));
echo "-";
    else;
echo $dummy;
    endif;
    echo "</td>";
    $f++;
  endwhile;
  echo "</tr>\n";
  $i++;
endwhile;

endif;
>

</table>
</center>

<hr><address>Kieron Brown, EMBL-Grenoble
</address>

</body> </html>

```

Bibliography

- [1] Rossmann M.G., and Blow D.M. (1962) *Acta Cryst.*, **15**, 24
- [2] Bijvoet J.M. (1954) *Nature*, **173**, 888
- [3] Hoppe H., Jakubowski V., (1975) in *Anomalous Scattering*, ed. Ramaseshan S., and Abrahams S.C., Munksgaard, p. 437
- [4] Phillips J., Wlodawer A., Goodfellow J., Watenpaugh K., Sieker L., Jensen L., Hodgson K. (1977) *Acta. Cryst.*, **A33**, 445
- [5] Karle J. (1967) *Applied Optics*, **6**, 12, 2132
- [6] Karle J. (1980) *Int. J. Quantum Chem.: Quantum Biol. Symp.*, **7**, 357
- [7] Hendrickson W.A. (1985) *Trans. American Crystallographic Association*, **21**, 11
- [8] Hendrickson W.A., Smith J.L., Sheriff S. (1985) in *Methods in Enzymology*, Academic Press, **115**, 41
- [9] Terwilliger T. (1994) *Acta Cryst.*, **D50**, 17
- [10] Ramakrishnan V., Finch J.T., Graziano V., Lee P.L., Sweet R.M. (1993) *Nature*, **362**, 219
- [11] Glover I.D., Denny R.C., Nguti N.D., McSweeney S.M., Kinder S.H., Thompson A.W., Dodson E.J., Wilkinson A.J., Tame J.R.H. (1995) *Acta Cryst.*, **D51**, 39
- [12] Hendrickson W.A., Pähler A., Smith J.L., Satow Y., Merritt E.A. and Phizackerley R.P. (1989) *PNAS*, **86**, 2190
- [13] Kahn R., Fourme R., Bosshard R., Chiadmi M., Risler J.L., Diderberg O. and Wery J.P. (1985) *FEBS Lett.*, **179**, 133

- [14] Guss J.M., Merritt E.A., Phizackerley R.P., Hedman B., Murata M., Hodgson K.O., Freeman H.C. (1988) *Science*, **241**, 806
- [15] Phizackerley R.P., Cork C.W. and Merritt E.A. (1986) *Nuclear. Inst. Methods in Physics*, **A246**, 579
- [16] Kahn R., Fourme R., Gadet A., Janin J., Dumas C., André, D. (1982) *J. Appl. Crystallogr.*, **15**, 330
- [17] Ogata C.M., Hendrickson W.A., Gao X., Patel D.J., (1989) *Abstracts of the American Crystallographic Association Meeting, series 2*, **17**, 53
- [18] *Introduction to Modern Physics*, ed. Richtmyer F.K., Kennard E.H., Cooper J.N., Tata McGraw-Hill, p. 188, (1987)
- [19] *Fundamentals of Crystallography*, ed. Giacovazzo C., International Union of Crystallography - Oxford University Press, (1994)
- [20] Jackson J.D., *Classical Electrodynamics*, Wiley & Sons, p. 679, (1975)
- [21] Hecht E., *Optics*, Addison Wesley, p. 44, (1987)
- [22] Als-Nielsen J., in *Neutron and Synchrotron Radiation for Condensed Matter Studies (HERCULES)*, Vol 1., ed. Baruchel J., Hodeau J.L., Lehmann M.S., Regnard J.R., Schlenker C., Springer-Verlag, p. 19, (1993)
- [23] Bragg W.L. (1912) *Nature*, **90**, 410
- [24] Beurger, *X-Ray Crystallography*, Wiley & Sons, p 40, (1966)
- [25] Ewald P. P. (1921) *Z. Kristallogr. Miner.*, **56**, 129
- [26] Schlenker M., in *Neutron and Synchrotron Radiation for Condensed Matter Studies (HERCULES)*, Vol 1., ed. Baruchel J., Hodeau J.L., Lehmann M.S., Regnard J.R., Schlenker C., Springer-Verlag, p. 147, (1993)
- [27] Arfken G., *Mathematical Methods for Physicists*, Academic Press, p. 356, (1985)
- [28] Glusker J.P., Trueblood K.N., *Crystal Structure Analysis - A Primer*, Oxford University Press, p. 74, (1985)

- [29] Woolfson M.M., An Introduction to X-ray Crystallography, Cambridge University Press, (1978)
- [30] Bragg W.L. (1939) Nature, **143**, 678
- [31] Patterson A.L. (1934) Phys Rev, **46**, 372
- [32] International Tables for Crystallography, Volume B, ed Shmueli U., The International Union of Crystallography - Kluwer Academic Press, 1993
- [33] Perutz M.F., Rossmann M.G., Cullis A.F., Muirhead H., Will G., North A.C.T. (1960) Nature Lond., **185**, 416
- [34] Blow D.M., Crick F.H.C. (1959) Acta Cryst., **12**, 794
- [35] Drenth J., Principles of Protein X-ray Crystallography, Springer-Verlag, (1994)
- [36] James R.W., The Optical Principles of the Diffraction of X-rays, Oxbow Press, (1962)
- [37] Kronig R. de L. and Kramers H.A. (1928) Z. Phys., **48**, 174
- [38] Blow D.M. (1958) Proc. Roy. Soc., **A247**, 302
- [39] Ramachandran G.N., Raman S. (1956) Current Science, **25**, 348
- [40] Blow D.M., Rossmann M. (1961) Acta. Cryst., **14**, 1195
- [41] Blake C.C.F., Fenn R.H., North A.C.T., Phillips D.C., Poljak R.J. (1962) Nature (London), **196**, 1173
- [42] North A.C.T. (1965) Acta. Cryst., **18**, 212
- [43] Smith J.L. (1991) Current Opinion in Structural Biology, **1**, 1002
- [44] McIntyre G.J., in Neutron and Synchrotron Radiation for Condensed Matter Studies (HERCULES), Vol 1., ed. Baruchel J., Hodeau J.L., Lehmann M.S., Regnard J.R., Schlenker C., Springer-Verlag, p. 179, (1993)
- [45] Stout G.H., Jensen L.H., X-Ray Structure Determination, Wiley & Sons, (1989)

- [46] Allinson N.M. (1994) *J. Sync. Rad.*, **1**, 54
- [47] Liebich B. (1977) *Acta Cryst*, **B33**, 2790
- [48] von Seggern H. and Voigt T., Knüpfer W. and Lange G. (1988) *J Appl. Phys.*, **64**, (3), 1405
- [49] Allinson N.M. (1982) *Nuclear Inst. Methods*, **201**, 53
- [50] Borkowski C.J., Kopp M.K. (1968) *Rev. Sci.* **39**, 1515
- [51] Amemiya Y., in *Synchrotron Radiation and Biophysics*, ed. Hasnain S.S., Wiley, (1990)
- [52] Morse J., in *Neutron and Synchrotron Radiation for Condensed Matter Studies (HERCULES)*, Vol 1., ed. Baruchel J., Hodeau J.L., Lehmann M.S., Regnard J.R., Schlenker C., Springer-Verlag, p. 108, (1993)
- [53] Helliwell J., *Macromolecular Crystallography with Synchrotron Radiation*, Cambridge University Press, (1992)
- [54] Felisaz F. (1995) personal communication.
- [55] mSQL : Hughes Technologies, <http://hughes.com.au/>
- [56] PHP programing script: <http://www.vex.net/php/>
- [57] Miyahara J. and Takahashi K., Amemiya Y., Kamiya N. and Satow Y. (1986) *Nuclear Inst. Methods*, **A246**, 572
- [58] Nordlund P., Eklund H. (1993) *J. Mol. Biol.*, **232**, 123
- [59] Nordlund P. (1996) personal communication
- [60] Leslie A.G.W. (1992) Recent Changes to the MOSFLM Package for Processing Films and Image Plate Data. A. G. W. Leslie, in *CCP4 and ESF-EACMB Newsletter on Protein Crystallography No. 26*. Daresbury Laboratory, Warrington, U.K.

- [61] Collaborative Computing Project No. 4. (1994) The CCP4 Suite: Programs for Crystallography. *Acta Cryst.* **D50**, 760
- [62] International Tables for Crystallography, Volume A (1993) ed. Shmueli U., The International Union of Crystallography - Kluwer Academic Press
- [63] Wang. B.C.(1985) in *Methods in Enzymology*, ed. Wyckoff H.W., Hirs C.H.W., Timasheff S.N., Academic Press, **115**, 90
- [64] Zhang, K.Y.J., Main P. (1990) *Acta Cryst* **A46**, 377
- [65] Matthews B.W. (1968) *J. Mol. Biol.*, **33**, 491
- [66] Jones, T.A. (1991) *Acta Cryst.*, **A47**, 110
- [67] Felisaz F. (1996) personal communication.
- [68] *Molecular biology of the Cell*, ed. Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J.D., Garland Publishing, p. 352, (1983)
- [69] Stoffel W, Blobel P. (1981) *Eur J Biochem*, **120**, 519
- [70] Walter P., Gilmore R., Blobel G. (1984) *Cell*, **38**, 5
- [71] Lütcke H. (1995) *Eur. J. Biochem.*, **228**, 531
- [72] Walter P. and Blobel G. (1981) *Journal of Cell Biology*, **91**, 557
- [73] Anderson D., Walter P. Blobel G. (1982) *Journal of Cell Biology*, **93**, 502
- [74] Gilmore R. (1982) *Journal of Cell Biology*, **95** , 470
- [75] Walter P. and Blobel G. (1980) *PNAS*, **77**, 7112
- [76] Walter P. and Blobel G. (1982) *Nature*, **299**, 691
- [77] Andrews D., Walter P., Ottensmeyer F.P. (1985) *PNAS*, **82**, 785
- [78] Siegel V., Walter P. (1988) *PNAS*, **85**, 1801
- [79] Strub K., Moss J., Walter P. (1991) *Mol. Cell Biol.*, **11**, 3949

- [80] Bovia F., Bui N., Strub K. (1994) *Nucleic Acids Res.*, **22**, 2028
- [81] Studier F.W., Rosenberf A.H., Dunn J.J., Dubendorff J.W. (1990) in *Methods in Enzymology*, **185**, 1074
- [82] Ashman K. (1994) personal communication
- [83] Laemmli U.K. (1970) *Nature* **227**, 680
- [84] Kapp U. (1996) unpublished results
- [85] Jensen O. (1996) personal communication
- [86] Birse D., Kapp U., Strub K., Cusack S. and Aberg Å. (1997) *EMBO Journal*, **16**, 13, 3757
- [87] McRee D.E., *Practical Protein Crystallography*, Academic Press, (1993)
- [88] Hendrickson W.A., Horton J.R., and LeMaster D.M. (1990) *EMBO Journal*, **9**, no. 5, 1672
- [89] Van Duyne G.D., Standaert R.F., Karplus P.A., Schreiber S.L. and Clardy J. (1993) *J. Mol. Biol.*, **229**, 105
- [90] Jensen O. (1995) personal communication
- [91] Doublé S., in *Crystallisation of Nucleic Acids and Proteins*, ed. Ducruix A., Giegé R., Oxford Univeristy, p. 315, (1992)
- [92] Jensen O. (1995) personal communication
- [93] Phillips J.C. and Hodgson K.O. (1980) *Acta Cryst.*, **A36**, 856
- [94] Blow D.M., in *Proceedings of the Daresbury Study Weekend on Molecular Replacement*, Daresbury, UK, p. 2, (1985)
- [95] Wilson A.J.C. (1949) *Acta Cryst.*, **2**, 318
- [96] Rees D.C. (1980) *Acta Cryst.*, **A36**, 578
- [97] Stanley E. (1972) *J. Appl. Cryst.*, **5**, 191

- [98] Redinbo M.R., Yeates T.O. (1993) *Acta. Cryst.*, **D49**, 375
- [99] International Tables for Crystallography, Volume C, ed. Shmueli U., The International Union of Crystallography - Kluwer Academic Press (1993)
- [100] Yeates T.O., in *Methods in Enzymology*, ed. Carter C.W., Sweet R., Academic Press, **276**, pp. 344, (1997)
- [101] Yeates T.O. (1988) *Acta. Cryst.*, **A44**, 142
- [102] Hendrickson W.A. (1991) *Science*, **254**, 51
- [103] Branden C., Tooze J., *Introduction to Protein Structure*, Garland Publishing, (1991)
- [104] Klapper M.H.(1977) *Biochem. Biophys. Res. Commun.*, **78**, 1018
- [105] Fanchon E. (1996) personal communication
- [106] Bertrand J., Auger G., Fanchon E., Martin L., Blanot D., van Heijenoort J. and Diderberg O. (1997) *EMBO*, **16**, 3416
- [107] Doublé S., Long A., Tabor S., Richardson C. and Ellenberger T.E. In preparation.
- [108] Kapp U. (1997) unpublished results
- [109] Otwinowski Z., *Oscillation Data Reduction Program*, in *Data Collection and Processing*, ed. Sawyer L. Isaacs N.W., Bailey S., DL/SCI/R34, Daresbury Laboratory, UK, p56, (1993)
- [110] Kleywegt G.J. and Jones T.A. Halloween ... Masks and Bones, in "From First to Final Model" (CCP4), pp. 59-66 (1994)

