

PART 9. X-RAY DATA COLLECTION

Chapter 9.1. Principles of monochromatic data collection

Z. DAUTER AND K. S. WILSON

9.1.1. Introduction

X-ray data collection is the central experiment in a crystal structure analysis. For small-molecule structures, the availability of intensity data to atomic, usually around 0.8 Å, resolution means that the phase problem can be solved directly and the atomic positions refined with a full anisotropic model. This results in a truly automatic structure solution for most small molecules.

Macromolecular crystals pose much greater problems with regard to data collection. The first arise from the size of the unit cell, resulting in lower average intensities of individual reflections coupled with a much greater number of reflections (Table 9.1.1.1). Secondly, the crystals usually contain considerable proportions of disordered aqueous solvent, giving further reduction in intensity at high resolution and, in the majority of cases, restricting the resolution to be much less than atomic. Thirdly, the crystals are sensitive to radiation damage (see Section 9.1.12). Such problems have severe implications for all subsequent steps in a structure analysis. Solution of the phase problem is generally not possible through direct methods, except for a small number of exceptionally well diffracting proteins. The refined models require the imposition of stereochemical constraints or restraints to maintain an acceptable geometry.

At modern synchrotron beamlines, cryogenic cooling and high-efficiency two-dimensional (2D) detectors have made data collection technically easier, but it remains a fundamental scientific procedure underpinning the whole structural analysis. Therefore, it is essential to take the greatest care over this key step. The aim of this chapter is to indicate procedures for optimizing data acquisition. Overviews on several issues related to this topic have been published (Carter & Sweet, 1997; Evans & Walsh, 2005; Dauter, 2005).

9.1.2. The components of a monochromatic X-ray experiment

To collect X-ray data from single crystals, the following elements are required:

- (1) a source of X-rays;
- (2) optical elements to focus the X-rays onto the sample;
- (3) a monochromator to select a single wavelength;
- (4) a collimator to produce a beam of defined dimension;
- (5) a shutter to limit the exposure of the sample to X-rays;
- (6) a goniostat with associated sample holder to allow rotation of the crystal;
- (7) the crystalline sample itself;
- (8) a cryogenic cooling device for vitrified crystals;
- (9) an efficient, generally 2D, detector system;
- (10) software to control the experiment and store and display the X-ray images;
- (11) data-processing software to extract intensities and associated standard uncertainties for the Bragg reflections in the images.

On a number of beamlines, automated procedures have been implemented for sample changing, automatic crystal centring, evaluating the diffraction and proposing a strategy for data collection. These allow rapid and more effective selection of the best sample and optimal parameters.

Many of these are discussed elsewhere in this volume. This chapter aims to provide guidance in those areas where choices are to be made by the experimenter and is concerned with the interrelations between parameters and how they conspire for or against different strategies of data collection.

9.1.3. Data completeness

The advantage of diffraction methods over spectroscopy is that they provide a full 3D view of the object. Diffraction methods are theoretically limited by the wavelength of the radiation used, but, in practice, every diffraction experiment is further limited by the aperture and quality of the lens. In the X-ray experiment, the aperture corresponds to the resolution limit and the quality of the 'lens' to the completeness and accuracy of the measured Bragg reflection intensities.

In this context, completeness has two components, the first of which is geometric and hence quantitative. It is necessary to rotate the crystal so that all unique reciprocal-lattice points pass through the Ewald sphere and the associated intensities are recorded on the detector. Ideally, the intensities of 100% of the unique Bragg reflections should be measured. The second component is qualitative and statistical: for each hkl , the intensity, I_{hkl} , should be significant, with its accuracy correctly estimated in the form of an associated standard uncertainty, $\sigma(I)$. The data should be significant in terms of the $I/\sigma(I)$ ratio throughout the resolution range. This point will be returned to below, but it is especially important that the data at low resolution are complete and not overloaded on the detector, and that there is not a significant fraction of essentially zero-level intensities in the higher-resolution shells.

9.1.4. X-ray sources

There are two principal sources of X-rays appropriate for macromolecular data collection: rotating anodes and synchrotron storage rings. These are not discussed in detail here as they are described in Chapters 6.1 and 8.1.

Table 9.1.1.1
Size of the unit cell and number of reflections

Compound	Unit cell		Reflections	Average intensity
	Edge (Å)	Volume (Å ³)		
Small organic	10	1000	2000	1
Supramolecule	30	25000	30000	1/25000
Protein	100	1000000	100000	1/1000000
Virus	400	100000000	1000000	1/100000000

9. X-RAY DATA COLLECTION

9.1.4.1. Conventional sources

Rotating anodes were initially developed for biological scattering experiments on muscle samples and have the advantage of higher intensity compared with sealed-tube generators. They usually have a copper target providing radiation at a fixed wavelength of 1.542 Å. Alternative targets, such as silver or molybdenum, provide lower intensities at short wavelengths, but have not found general applications to macromolecules. It is also possible to use a chromium target, giving a longer wavelength of 2.29 Å. Historically, rotating anodes were first used with nickel filters to give monochromatic Cu $K\alpha$ radiation. Current systems are equipped with either graphite monochromators, a focusing mirror or multilayer optics. The latter provide substantially enhanced intensity. Rotating anodes remain the source of choice in most structural biology laboratories. An important choice for the user is in the selection of optimal collimator aperture: this should roughly match the crystal sample dimensions. For large crystals, especially if the cell dimensions are also large, it may be preferable to use collimator settings smaller than the crystal in order to resolve the diffraction spots on the detector. Considerable progress has been made in the technologies for rotating anodes in recent years, as well as in the development of high-intensity sealed-tube sources for home laboratories (Bloomer & Arndt, 1999).

9.1.4.2. Synchrotron storage rings

The radiation intensity available from rotating anodes is limited by the heat load per unit area on the target. In the early 1970s, it was realized that synchrotron storage rings produced X-radiation in the necessary spectral range for studies in structural molecular biology (Rosenbaum *et al.*, 1971), and the last four decades have seen great advances in their application to macromolecular crystallography (Helliwell, 2004; Hendrickson, 2000). Synchrotron radiation (SR) is now used for the great majority of newly determined protein-crystal structures.

The general advantages of SR are:

- (1) High intensity: third-generation sources provide more than 1000 times the intensity of a conventional source.
- (2) A highly parallel beam allowing the resolution of closely spaced spots from large unit cells.
- (3) Short wavelengths, less than 1 Å, essentially eliminating the problems of correcting for absorption.
- (4) Tunability of the wavelength, allowing its optimization for single- or multiple-wavelength applications; this is simply not possible with a conventional source.
- (5) The ability to use a white, non-monochromated beam, the so-called Laue technique discussed in Chapter 8.2.

SR beamlines take a number of forms. The source may be a bending magnet or an insertion device, such as a wiggler or an undulator. The properties of different beamlines thus vary considerably and it is vital to choose an appropriate beamline for any particular application. The beamline capabilities are, of course, affected by the detector as well as the source itself. As far as the user is concerned, the primary questions regard the intensity, the size of the focal spot, the wavelength tunability and the detector system.

The present consensus for new synchrotron beamlines for macromolecular crystallography is that they should be on sources with an energy of at least 3 GeV and should receive radiation from tunable undulators. Together, these provide high and tunable intensity over the range required for most crystal-

lographic experiments, including multiwavelength anomalous dispersion (MAD). The impact of free-electron lasers, which are currently under construction at a number of sites, is not yet possible to assess.

Present beamlines produce radiation of extremely high quality for macromolecular data collection. At third-generation sources complete data sets can be collected from cryogenically vitrified single crystals in minutes.

9.1.5. Goniostat geometry

9.1.5.1. Overview

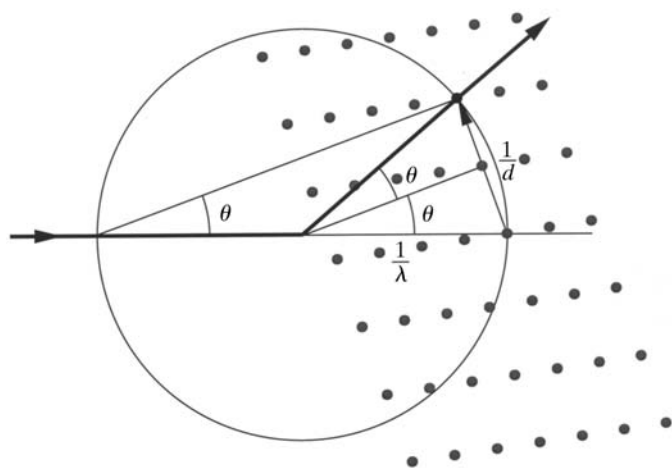
The diffraction condition for a particular reflection is fulfilled when the corresponding reciprocal-lattice point lies on the surface of the Ewald sphere. If a stationary crystal is irradiated by the X-ray beam, only a few reflections will lie in the diffracting position. To record intensities of a larger number of reflections, either the size of the Ewald sphere or the crystal orientation has to be changed. The first option, with the use of non-monochromatic, or 'white', radiation, is the basis of the Laue method (Chapter 8.2). If the radiation is monochromatic, *i.e.* single-wavelength, the crystal has to be rotated during exposure to bring successive reflections into the diffraction condition.

9.1.5.2. The screenless rotation method and 2D detectors

In the early days of protein crystallography, a number of geometries were used for X-ray cameras, notably the Weissenberg and precession methods. In addition, single-counter diffractometers were used with four-circle goniostats. However, in practice only the screenless rotation geometry (Arndt & Wonacott, 1977) survives today. This requires a 2D detector, which was initially in the form of photographic film. It is of significance that typical film sizes were of the order of 10 × 10 cm with up to 2000 × 2000 scanned pixels; a similar effective area has proved effective for image plates and charge-coupled devices (CCDs).

However, automation of protein-data collection needed efficient 2D detectors (Part 7). The first were multiwire proportional counters, which found widespread use in the early 1980s (Hamlin, 1985). These finally proved to be limited by a combination of spatial resolution and dead time of the read-out. A major advance occurred in the late 1980s with the widespread introduction of imaging plates (Amemiya, 1995), scanned on-line both at synchrotron beamlines and on laboratory rotating-anode sources. This represented a revolution in macromolecular data collection, making it technically straightforward to save full 2D images with sufficient positional resolution and dynamic range to computer disk automatically. The limiting factor of the imaging plate proved to be the slow read-out time of the order of several seconds to minutes. At high-intensity sources in particular, *e.g.* third-generation SR sites, exposure times per image can fall to one second or less, and with an imaging plate the bulk of the time is spent reading the detector image rather than collecting data. Typical data-collection times with imaging plates remained of the order of several hours, even with the use of SR. This is a much smaller problem with rotating-anode sources, where exposure times dominate the duty cycle.

For high-intensity SR sites, the detector of choice is the CCD (Gruner & Ealick, 1995). The spatial resolution is comparable with that of imaging plates, but the read-out time can be as low as one to two seconds. This means that complete data can be

**Figure 9.1.6.1**

The Ewald-sphere construction. A reciprocal-lattice point lies on the surface of the sphere if the following trigonometric condition is fulfilled: $1/2d = (1/\lambda) \sin \theta$. After a simple rearrangement, it takes the form of Bragg's law: $\lambda = 2d \sin \theta$. Therefore, when a reciprocal-lattice point with indices hkl lies on the surface of the Ewald sphere, the interference condition for that particular reflection is fulfilled and it gives rise to a diffracted beam directed along the line joining the centre of the sphere to the reciprocal-lattice point on the surface.

recorded in minutes rather than hours and has transformed approaches to data collection.

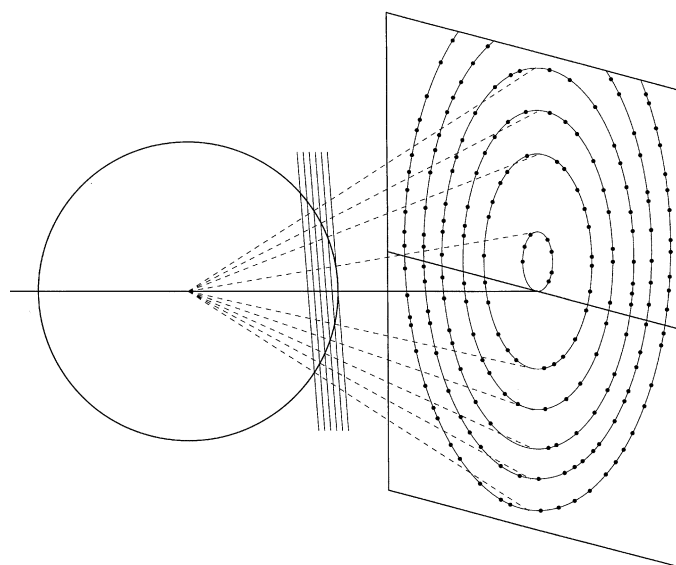
While the CCD has revolutionized data-collection times, further advances are expected from the use of solid-state pixel detectors. Such detectors record individual X-ray quanta and have essentially zero read-out time. The most advanced of these, the PILATUS 1M device, is a hybrid pixel array detector (Broennimann *et al.*, 2006; Hülsen *et al.*, 2006), first installed at the Swiss Light Source.

Almost all current 2D detectors are used in conjunction with a goniostat, providing rotation of the crystal about a single axis during exposure. Indeed, the majority of instruments have only a single rotation axis. The remainder are based on the kappa (ω , κ , φ) cradle to select different initial orientations of the sample in the beam; the sample is nevertheless subsequently rotated about a single axis for data collection.

9.1.6. Basis of the rotation method

9.1.6.1. Rotation geometry

The physical process of diffraction from a crystal involves the interference of X-rays scattered from the electron clouds around the atomic centres. The ordered repetition of atomic positions in all unit cells leads to discrete peaks in the diffraction pattern. The geometry of this process can alternatively be described as resulting from the reflection of X-rays from a set of hypothetical planes in the crystal. This is explained by the Ewald construction (Fig. 9.1.6.1), which provides a visualization of Bragg's law. Monochromatic radiation is represented by a sphere of radius $1/\lambda$, and the crystal by a reciprocal lattice. The lattice consists of points lying at the end of vectors normal to reflecting planes, with a length inversely proportional to the interplanar spacing, $1/d$. In the rotation method, the crystal is rotated about a single axis, with the rotation angle defined as φ . A seminal work, giving an excellent background to this field by a number of contributors, was edited by Arndt & Wonacott (1977). A more recent review of the method is provided by Dauter (2005).

**Figure 9.1.6.2**

The plane of reflections in the reciprocal sphere that is approximately perpendicular to the X-ray beam gives rise to an ellipse of reflections on the detector.

9.1.6.2. Diffraction pattern at a single orientation: the 'still' image

For a stationary crystal in any particular orientation (a so-called 'still' exposure), only a fraction of the total number of Bragg reflections will satisfy the diffracting condition. The number of reflections will be very limited for a small-molecule crystal, possibly zero in some orientations. Macromolecules have large unit cells, of the order of 100 Å, compared with the wavelength of the radiation, which is about 1.0 Å. In geometric terms, the reciprocal space is densely populated with points in relation to the size of the Ewald sphere. Thus, more reflections diffract simultaneously but at different angles, since many reciprocal-lattice points (reflections) lie simultaneously on the surface of the Ewald sphere in any crystal orientation. This is the great advantage of 2D detectors for large cell dimensions.

The real crystal is a regular and ordered array of unit cells. This means that reciprocal space is made up of a set of points organized in regular planes. For a still exposure, any particular plane of points in the reciprocal lattice intersects the surface of the Ewald sphere in the form of a circle. The corresponding diffracted rays, originating from the centre of the Ewald sphere, form a cone that intersects the sphere on the circle formed by the set of points. In most experiments, the detector is placed perpendicular to the direct beam and the cone of diffracted rays forms an ellipse of spots on its surface (Fig. 9.1.6.2). If a major axis of the crystal lies nearly parallel to the beam, then the ellipses will approximate a set of circles around the centre of the detector. All reflections within each circle will have one index in common, corresponding to the unit-cell axis lying along the beam. For non-centred unit cells, this index will increase by one in successive circles. The gaps between the circles depend on the spacing between the members of the set of reciprocal-lattice planes and are inversely proportional to the real cell dimension related to these planes.

Still exposures were used extensively in the early applications of the rotation method for estimation of crystal alignment. The geometric location of the spots with respect to the origin allows accurate determination of the unit-cell parameters and the crystal orientation. This approach has been superseded in modern

9. X-RAY DATA COLLECTION

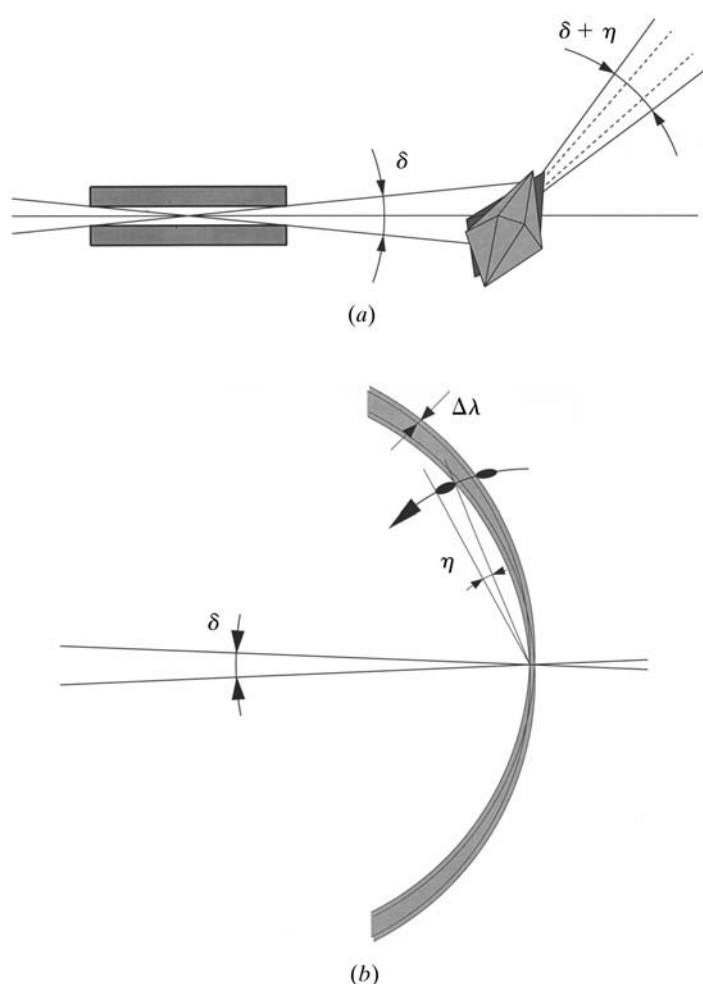


Figure 9.1.6.3

Schematic representation of beam divergence (δ) and crystal mosaicity (η). (a) In direct space, (b) in reciprocal space, where the additional thickness of the Ewald sphere results from the finite wavelength bandpass, $\delta\lambda/\lambda$.

software packages by autoindexing algorithms using real rotation images.

9.1.6.3. Rocking curve: crystal mosaicity and beam divergence

The Ewald-sphere construction assumes an ideal source with a totally parallel X-ray beam and an ideal crystal with all unit cells having identical relative orientation, resulting in infinitely sharp Bragg reflections. These assumptions lead to a sphere of radius $1/\lambda$ attached rigidly to the beam and with the crystal in a particular orientation as a reciprocal lattice consisting of mathematical points. A real experiment deviates from this in three respects. Firstly, the incident beam is not strictly parallel. On a conventional rotating-anode source the beam can only be focused and collimated to be parallel within a small angle, with a divergence of about 0.2° (with mirror optics) and 0.4° (with a monochromator). On SR sources, a much smaller beam divergence can be achieved, and, indeed, beamlines on third-generation SR sources approach the ideal ever more closely. The horizontal and vertical beam divergence may differ, and this must be taken into account. The Ewald sphere now has two limiting orientations which result in a nonzero active width. Secondly, the X-radiation is only monochromatic within a defined wavelength bandpass, $\delta\lambda/\lambda$, of the order 0.0002–0.001 at synchrotron beamlines, but considerably more for laboratory sources. The wavelength bandpass, in effect, thickens the surface of the Ewald

sphere. Thirdly, real crystals are made up of small mosaic blocks imperfectly oriented relative to one another, increasing the total rocking curve. At room temperature, protein crystals often show a mosaic spread less than 0.05° , but for some samples this may be much larger. However, vitrification of crystals in many cases leads to substantial increase of mosaicity to sometimes more than 1° . In the reciprocal lattice, the effect of this is to give a finite dimension to each of the lattice points.

These effects are schematically illustrated in Fig. 9.1.6.3. The combined result is that the diffraction of a particular reflection is spread over a range of crystal rotation.

9.1.6.4. Rotation images and lunes

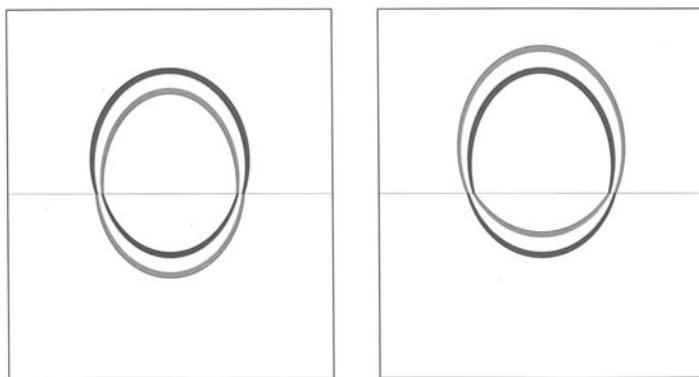
Using monochromatic radiation, in order to measure the remaining reflections that do not lie on the surface of the sphere, the crystal must be rotated to bring the reflections into the diffracting condition. If the crystal is rotated about a single axis during sequential exposures, this is known as the rotation method. The rotation axis is, in practice, chosen to be perpendicular to the beam to preserve the symmetry between the two halves of the complete pattern. This is the most commonly applied method of data collection for macromolecular crystals (Arndt & Wonacott, 1977).

If the crystal is rotated during exposure, the ellipses observed on a still image change their position on the detector. In effect, all reflections diffracting during one exposure will be contained within lunes formed between the two limiting positions of each ellipse at the start and end of the given rotation. The width of the lunes in the direction of the crystal rotation, perpendicular to the rotation axis, is proportional to the rotation range per exposure. In contrast, along the rotation axis the width of the lunes is very small, since the intersection of the reciprocal-lattice plane with the Ewald sphere does not change significantly. For crystals of small molecules, the lunes are not pronounced, owing to the sparse population of reciprocal space, but for crystals with large cell dimensions, the lunes are densely populated by diffraction spots and often exhibit clear and well pronounced edges. At high resolution, the mapping of the reciprocal lattice within each lune is distorted, and rows of reflections form hyperbolas. At low diffraction angles, where the surface of the Ewald sphere is approximately flat, this distortion is minimal, and the lunes look like fragments of precession photographs.

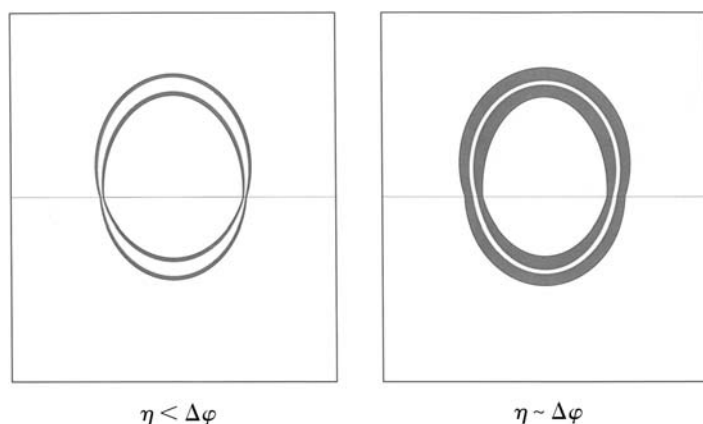
9.1.6.5. Partially and fully recorded reflections

The rotation method gives rise to lunes of data between the ellipses that relate to the start and the end of the rotation range used for the exposure. The data are complete if the Ewald sphere has been crossed by all reflections in the asymmetric part of the reciprocal lattice, which means that the crystal has to be rotated in total by a substantial angle. However, it is impossible to record all the data in a single exposure with such a wide rotation, owing to overlapping of the diffraction spots.

In practical applications to macromolecules, the total rotation is divided into a series of narrow individual rotations of width $\Delta\varphi$. In each of these, the crystal is exposed for a specified time or X-ray dose per angular unit. Each reflection diffracts over a defined crystal rotation and hence time interval, owing to the finite value of the rocking curve or angular spread, here referred to as ξ , the combined effect of beam divergence (δ) and crystal mosaicity (η). Provided ξ is less than $\Delta\varphi$, some reflections will start and finish crossing the Ewald sphere and hence diffract within one exposure. Their full intensity will be recorded on a

**Figure 9.1.6.4**

A single lune on two consecutive exposures. The partial reflections appear on both images and their intensity is distributed over both.

**Figure 9.1.6.5**

Appearance of a lune for (left) a crystal of low mosaicity and (right) a highly mosaic crystal. Characteristically, the width of the lune along the rotation axis is wider if the mosaicity is high.

single image, and these are referred to as fully recorded reflections, or fullys.

Other reflections will start to diffract during one exposure, but will still be diffracting at the end of the $\Delta\varphi$ rotation range. The remaining intensity of these reflections will be recorded on subsequent images. There will of course be corresponding reflections at the start of the present image. These reflections are termed partially recorded, or partials. Fig. 9.1.6.4 shows schematically how a lune appears on two consecutive exposures, with partials at each edge. The partials at the bottom edge of each lune contain the rest of the intensity of the partials from the previous exposure. The rest of the intensity of the partials at the top of the lune will appear on the next exposure. Superposition of two successive images will reveal some spots common to both: they are the partials shared between the two. If the angular spread ξ is small compared with the rotation range $\Delta\varphi$ then most reflections will be fully recorded. As ξ increases, the proportion of partials will rise, and when it reaches or exceeds $\Delta\varphi$ in magnitude there will be no fully recorded reflections. If the rotation range per image is small compared with the rocking curve, individual reflections can be spread over several images.

As ξ increases, the lunes become wider (Fig. 9.1.6.5), since there are more partial reflections crossing the Ewald sphere at any one time. The appearance of the lunes can be used to estimate the mosaicity of the crystal. If the edges are sharply defined, then the mosaicity is low. In contrast, if the intensities at the edges gradually fade away, then the mosaicity must be high. Indeed, this phenomenon can be exploited by the integration

software to provide accurate definition of the orientation parameters and of the mosaicity.

A key characteristic of high mosaicity is that all lunes are wide in the region along the rotation axis. On still exposures, the width of the rings is proportional to the angular spread. The width of lunes is expected to be very small along the rotation axis. If they are wide in this region, this is especially indicative of high mosaic spread. While highly ordered crystals with low mosaicity are preferable and often lead to data of the highest quality, high mosaic spread is not a prohibitive factor in accurate intensity estimation, provided it is properly taken into account in estimating the data collection and integration parameters, such as individual rotation ranges.

9.1.6.6. The width of the rotation range per image: fine φ slicing

An important variable in the rotation method is the width of the rotation ranges per individual exposure. The two basic approaches can be termed wide and fine φ slicing, and differ in the relation between the angular spread and the rotation range per exposure. The two methods are applicable under different experimental constraints.

Fine φ slicing requires that the individual intensities are divided over several consecutive images, *i.e.* $\Delta\varphi$ should be substantially less than ξ (Kabsch, 1988). This approach possesses two very positive features. Firstly, it minimizes the background by integrating intensities only over a φ range equivalent to the rocking curve of the crystal. Secondly, it allows the fitting of 3D profiles to the pixels that compose a reflection, the first two dimensions being the xy plane of the detector and the third the φ rotation. In combination, these should provide an optimum signal-to-noise ratio for the measured intensities and would appear to be the method of choice for data collection.

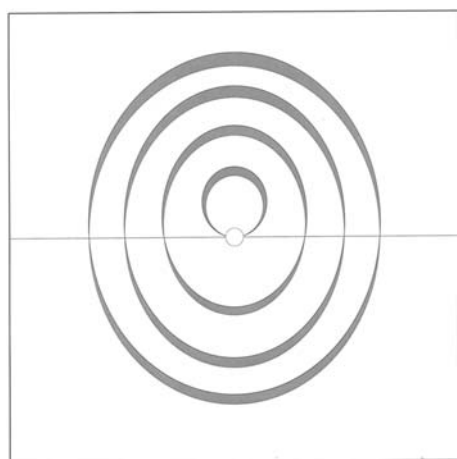
However, this involves a very large number of images, which can pose logistical problems in terms of data handling. Only if the read-out time is negligible in comparison with the exposure time can fine slicing be applied. If the detector read-out is slow, fine slicing becomes totally impractical. Multiwire chambers allow fine φ slicing, but unfortunately their disadvantages in terms of effective dynamic range preclude their use on high-intensity sources. Imaging plates are generally too slow for this approach.

The fine-slicing method is likely to receive a resurgence of interest with the introduction of fast read-out detectors such as the PILATUS, where diffraction images can be acquired continuously without the use of a beam shutter (Hülshen *et al.*, 2006).

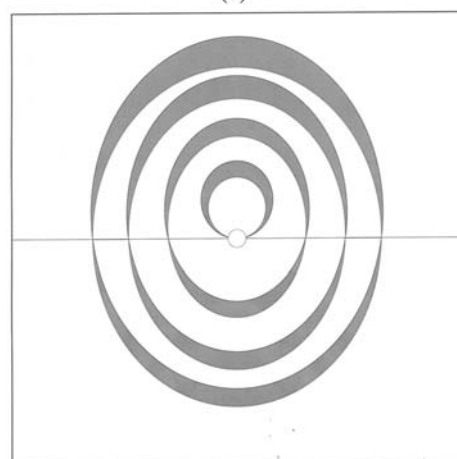
9.1.6.7. Wide slicing

The object of the wide-slicing approach is to acquire the data on as small a number of individual exposures as possible. It involves large $\Delta\varphi$ values per image, usually of the order of 0.5° or more, which exceed the angular spread. Each image contains a considerable proportion of fully recorded reflections. Originally, wide slicing was used to minimize the large numbers of X-ray films to be processed. Only the wide-slicing approach is tractable for detector systems where the read-out time is relatively slow in relation to exposure, *e.g.* imaging plates with read-out times of 20 seconds to minutes.

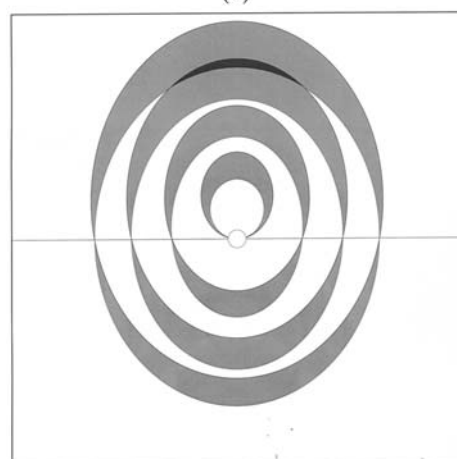
Wide slicing has two drawbacks. Firstly, during integration of the intensity data, only 2D profiles are fitted for each individual spot in the wide slicing. Secondly, each reflection profile overlaps a background which accumulates throughout the whole time and



(a)



(b)



(c)

Figure 9.1.6.6

The width of the lunes is proportional to the rotation range per image, $\Delta\varphi$, which increases from (a) to (c). If the rotation range is large, the lunes overlap at high resolution.

angular range of the exposure, even when the reflection concerned is not diffracting.

The aim is to use the maximum acceptable rotation range per image. The lunes on an image have a finite width proportional to the rotation range. This width restricts the allowed angular range per image, as overlap of spots resulting from overlap of adjacent lunes must be avoided if the intensities are to be successfully integrated (Fig. 9.1.6.6). Several factors affect the degree of overlap and will be discussed in the rest of this section. A simple formula (Fig. 9.1.6.7) can be used to estimate the maximum permitted rotation range per image:

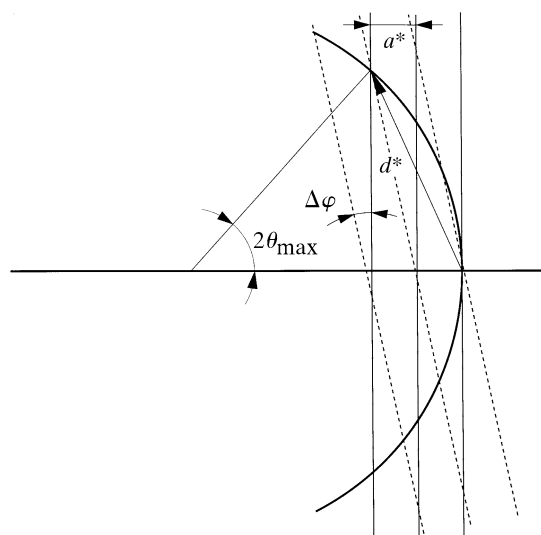


Figure 9.1.6.7

The largest allowed rotation range per exposure depends on the dimension of the primitive unit cell oriented along the X-ray beam; this is diminished by high mosaicity.

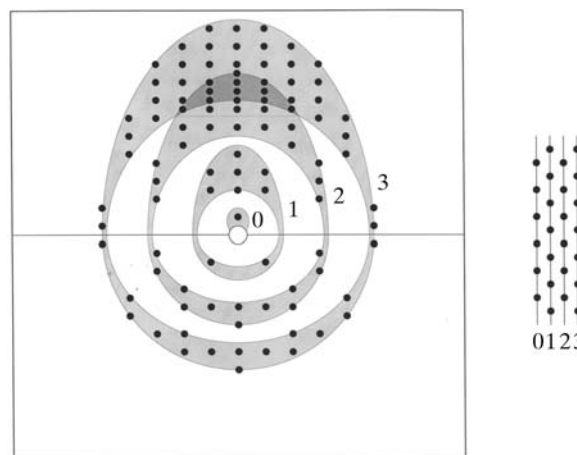


Figure 9.1.6.8

If the crystal lattice is centred or if its orientation is non-axial, the reflections do not overlap in spite of overlapping lunes, as illustrated to the right with consecutive layers of reflections viewed from the side.

$$\Delta\varphi = 180d/\pi a - \xi,$$

where the factor $180/\pi$ converts radians to degrees, ξ is the angular spread of the reflection, d is the high-resolution limit and a is the length of the primitive cell dimension along the direction of the X-ray beam. However, this simplistic equation can be somewhat misleading. It most strictly applies when the lunes are densely packed with reflections, for an orthogonal cell rotated about a major axis. If this is not the case, then often rows of reflections from one lune fit between rows in the adjacent lune without overlap. For example, for a trigonal crystal with its a axis along the beam and rotating about its c axis, even and odd lunes contain rows of reflections that lie between one another on the detector (Fig. 9.1.6.8).

It can be extremely hard to record data from samples with a very long cell dimension. If the long axis lies along the X-ray beam, then it will restrict $\Delta\varphi$ considerably to very low values. This is exacerbated if the mosaicity is substantial. It is therefore beneficial to have the longest axis oriented roughly along the

9.1. PRINCIPLES OF MONOCHROMATIC DATA COLLECTION

spindle axis, as it can then never lie parallel to the beam. This can be a problem with cryogenic samples mounted in loops, where the preferred orientation is hard to dictate, and this is an example where a κ -goniostat is an advantage, allowing reorientation of the crystal.

Simplistic application of the above equation for estimating $\Delta\varphi$ is not recommended. The degree of overlap also depends on pixel size, beam cross section, crystal size and mosaicity, and crystal-to-detector distance. Given the influence of these additional parameters, it is in practice better to employ the integration software first to interpret the diffraction pattern and to simulate predicted patterns by adjusting the data-collection parameters. Several modern packages have such strategy features, and it is vital to employ them before collecting data. In fact, from a theoretical point of view, the highest signal-to-noise ratio is obtained when $\Delta\varphi$ corresponds to the rocking curve of the diffraction, *i.e.* the mosaic spread plus the beam divergence.

9.1.7. Rotation method: geometrical completeness

This topic has been reviewed by Dauter (2005).

9.1.7.1. Total rotation range for non-anomalous data

The total set of structure-factor amplitudes from a crystal is a sphere of points in reciprocal space, with a radius defined by the maximum resolution. The intensities of the two hemispheres of data show a centrosymmetric relationship based on Friedel's law, which only breaks down if anomalous scatterers are present. However, the diffraction pattern possesses internal symmetry related to that of the real-space unit cell. This means that for all space groups an asymmetric unit of reciprocal space can be defined. Provided the intensities of all reflections in this asymmetric unit have been measured, those of all others can be generated by the symmetry operations and the Fourier transform for the complete structure computed.

The asymmetric unit has the shape of a wedge extending from the origin at the centre of the reciprocal sphere with a cutoff at a maximum radius corresponding to the limiting diffraction angle (resolution). Once the Laue symmetry group of the crystal has been determined (*International Tables for Crystallography*, Volume A, 2005), it is straightforward to define the shape of this wedge and establish which data must be recorded to make up a complete unique set. For macromolecular crystals, where there can be no centre of symmetry, the possibilities are further simplified to the point group rather than the Laue group. All space groups belonging to the same point group have the same asymmetric unit. The only differences lie in the presence or absence of screw axes or centring. Thus, space groups $P2_12_12_1$, $P2_12_12$, $P222_1$, $P222$, $I222$ and $I2_12_12_1$ all belong to point group (symmetry class) 222 and have the same asymmetric unit in reciprocal space. The only consequence of the presence of screw axes or lattice centring is to introduce systematic absences for some classes of reflection within this asymmetric unit of the point group.

It is usual to define the limits of the asymmetric unit by placing restrictions on the indices. For point group 222, the common conventional choice of limits on the reflection indices hkl is

$$0 \leq h \leq h_{\max}, \quad 0 \leq k \leq k_{\max}, \quad 0 \leq l \leq l_{\max},$$

where h_{\max} , k_{\max} and l_{\max} are defined by the maximum resolution. In all point groups, there are multiple but equivalent ways of defining the asymmetric unit, but a default definition is generally chosen by the data-reduction software. For example, in triclinic

symmetry, any hemisphere constitutes an asymmetric unit, and there are three typical choices of index limits:

$$0 \leq h \leq h_{\max}, \quad \bar{k}_{\min} \leq k \leq k_{\max}, \quad \bar{l}_{\min} \leq l \leq l_{\max},$$

or

$$\bar{h}_{\min} \leq h \leq h_{\max}, \quad 0 \leq k \leq k_{\max}, \quad \bar{l}_{\min} \leq l \leq l_{\max},$$

or

$$\bar{h}_{\min} \leq h \leq h_{\max}, \quad \bar{k}_{\min} \leq k \leq k_{\max}, \quad 0 \leq l \leq l_{\max}.$$

The standard choices of asymmetric unit taken from the CCP4 program suite (Collaborative Computational Project Number 4, 1994) are shown in Table 9.1.7.1.

The data are complete if the Ewald sphere has been crossed by all reflections in the asymmetric part of the reciprocal lattice. During data acquisition and reduction, all measured indices are conventionally transformed to this asymmetric unit of reciprocal space. Firstly, this allows merging of symmetry-equivalent measurements as appropriate. Secondly, it allows the completeness of the data to be assessed efficiently, using contributions from the whole sphere.

For all point groups, rotation of the crystal by 180° from any starting angle on the φ spindle axis is sufficient to provide a complete set of data (this is not sufficient if anomalous-dispersion measurements are required; see Section 9.1.7.2). Given such a total rotation, the redundancy of the measurements will increase with higher crystal symmetry. Thus, for a triclinic space group, the unique data will be measured almost twice on average (see the blind region below); for orthorhombic, eight times; for hexagonal class 6, 12 times; and for 622, 24 times. Redundancy is, in principle, advantageous, giving improved data quality (again see below), but it is generally possible to record complete unique data with a minimal overall rotation and correctly chosen starting angle on the spindle. It is of course necessary to determine the crystal orientation matrix, and this remains a vital part of the data-collection strategy. To minimize the effects of radiation

Table 9.1.7.1

Standard choice of asymmetric unit in reciprocal space for different point groups from the CCP4 program suite

Point group	Index limits	
1	hkl : $hk0$: $0k0$:	$l \geq 0$ $h \geq 0$ $k \geq 0$
2	hkl : $hk0$:	$k \geq 0, l \geq 0$ $h \geq 0$
222	hkl :	$h \geq 0, k \geq 0, l \geq 0$
4	hkl : $0kl$:	$h \geq 0, k > 0, l \geq 0$ $k \geq 0$
422	hkl :	$h \geq k, k \geq 0, l \geq 0$
3	hkl : $00l$:	$h \geq 0, k > 0$ $l > 0$
321	hkl : hhl :	$h \geq k, k \geq 0$ $l \geq 0$
312	hkl : $h0l$:	$h \geq k, k \geq 0$ $l \geq 0$
6	hkl : $0kl$:	$h \geq 0, k > 0, l \geq 0$ $k \geq 0$
622	hkl :	$h \geq k, k \geq 0, l \geq 0$
23	hkl : hkh :	$h \geq 0, k > h, l > h$ $k \geq h$
432	hkl :	$h \geq 0, k \geq l, l \geq h$

9. X-RAY DATA COLLECTION

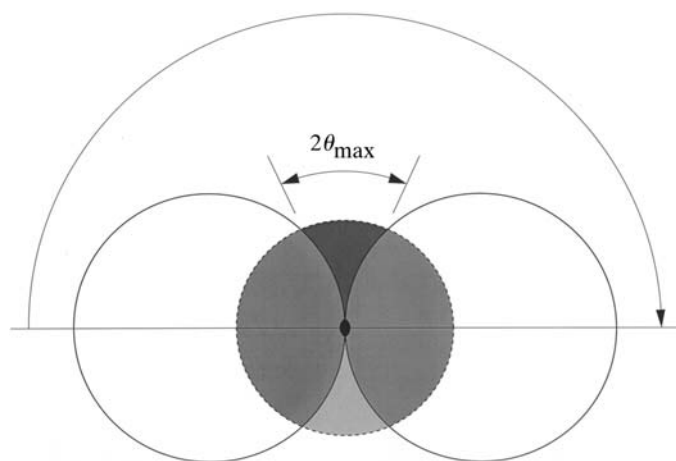


Figure 9.1.7.1

Rotation of a triclinic crystal by 180° in the X-ray beam, represented as rotating the Ewald sphere with a stationary crystal, projected along the rotation axis. For the purpose of analysing the relation of data completeness to crystal symmetry and orientation the two representations are equivalent.

damage, it is often beneficial to collect complete data with the minimal rotation range. This may well change with the advent of extremely fast detectors, when the decision-making process may take longer than data collection.

Thus, the crystal point-group symmetry has a profound effect on the total rotation range and the optimal starting spindle and crystal orientation for the most efficient recording of complete unique data. The rest of this section suggests strategies for the collection of complete data with minimal total rotation when anomalous-dispersion measurements are not required.

As stated above, for all crystals, rotation by 180° is sufficient to cover fully both sides of the Ewald sphere with intensity measurements. This is necessary for a triclinic crystal rotated around any arbitrary axis and also for a monoclinic crystal rotated around its unique b axis (Fig. 9.1.7.1). A twofold redundancy of unique data results; fourfold for the monoclinic case. Now consider a rotation of less than 180° (Fig. 9.1.7.2). Owing to the curvature of the Ewald sphere and the centre of symmetry arising from Friedel's law, the region of the sphere with reflections measured twice is diminished, and for part of the sphere there are no measurements. Most importantly, the proportions are resolution dependent. With a limited rotation, the high-resolution intensities reach a higher completeness than those at low resolution: data 90% complete at high resolution may be missing 20% of the low-resolution shells. Indeed, the low-resolution terms only become complete when a full 180° has been achieved.

The major data-processing software packages provide estimates of overall completeness as a function of total rotation range and starting point. However, they tend to neglect this variation with resolution. The fundamental importance of completeness at low resolution will be returned to later.

For total rotation by a given percentage of the angle needed to provide complete data, the resulting percentage completeness will be higher, again as a consequence of the curvature of the Ewald sphere. Consider again the triclinic case, when complete data require rotation by 180° . A single continuous range of 90° gives a completeness of about 65% (Fig. 9.1.7.3). Splitting the rotation range is advantageous; for example, if the crystal is rotated over two ranges of 45° , separated by a gap of 45° , the

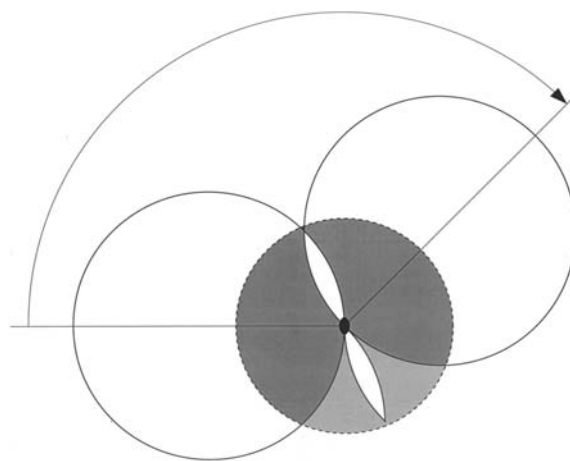


Figure 9.1.7.2

Rotation of a triclinic crystal by 135° is not sufficient to obtain totally complete data. At high resolution the completeness is higher than at low resolution, where a full 180° rotation is required.

completeness typically rises to about 80%. In summary, for the triclinic case, the starting point and crystal orientation are irrelevant, but if it is impossible to cover 180° in the time available, it is better to use two or more sets of ranges. The software can again often provide advice on such strategies.

When the crystal has symmetry elements, the situation is more complex. Now the completeness is sensitive to the starting point of rotation and the crystal orientation, as well as the total rotation range used. All three must be considered in defining an optimum strategy for minimal rotation to give complete data. Consider an orthorhombic unit cell where the asymmetric unit comprises any octant of the reciprocal lattice. Minimal complete data require a total rotation of 90° between any twofold axis and the plane perpendicular to it (Fig. 9.1.7.4). This requires that one of the major axes must lie along the direction of the beam, either at the start or end of the 90° rotation, when the other two axes will lie in the plane of the detector. It is not necessary to rotate around one of the major axes, but the rotation axis should lie in one of the three major planes. If these conditions on crystal orientation or starting point are not satisfied, then more than a 90° rotation will be required. The proper selection of starting point is vital. A 90°

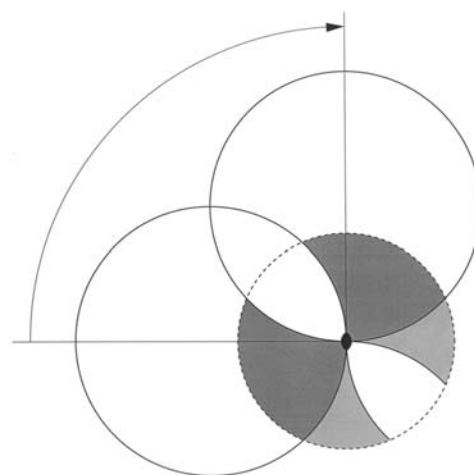


Figure 9.1.7.3

After a 90° rotation out of a required 180° , the overall completeness is higher than 50%.

9.1. PRINCIPLES OF MONOCHROMATIC DATA COLLECTION

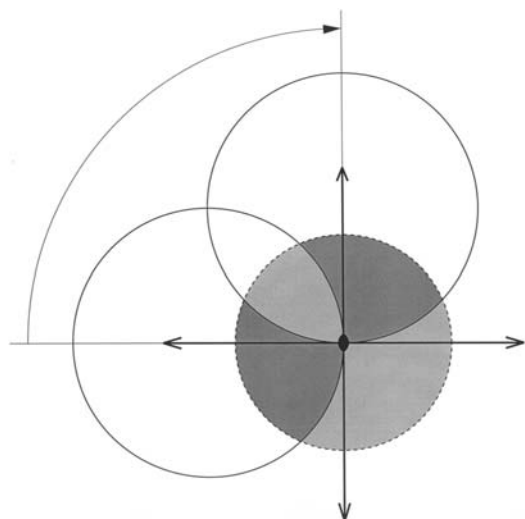


Figure 9.1.7.4

For an orthorhombic crystal, a 90° rotation is sufficient provided the starting or final orientation is along the major axis.

rotation starting midway between two axial positions, when the major axis only lies along the beam after 45°, will reduce the completeness after 90° to about 65%, since in essence the same 45° of unique data will be measured twice, albeit with high redundancy (Fig. 9.1.7.5). This emphasizes the need to define the crystal symmetry and orientation properly before data collection if minimalist protocols are to be employed.

In general, the higher the crystal symmetry, the more the completeness depends on the crystal orientation. In point groups 321 or 312, the asymmetric unit may be defined as a 30°-wide wedge that spans the space between the positive and negative direction of the threefold *c* axis. The index limits are

$$0 \leq h \leq h_{\max}, \quad 0 \leq k \leq h, \quad \bar{l}_{\max} \leq l \leq l_{\max}.$$

If the crystal is mounted with the threefold axis along the rotation spindle, it is sufficient to rotate by 30°, but only if the *a* or *b* axis lies along the beam at the start or end of the range. In contrast, if the crystal is rotated around **a** or **b**, then it is necessary to cover 90°. The second procedure will lead to a threefold increase in redundancy, but at the expense of a longer time.

The total rotation requirements for various crystal symmetries and orientations are given in Table 9.1.7.2. It is difficult to give

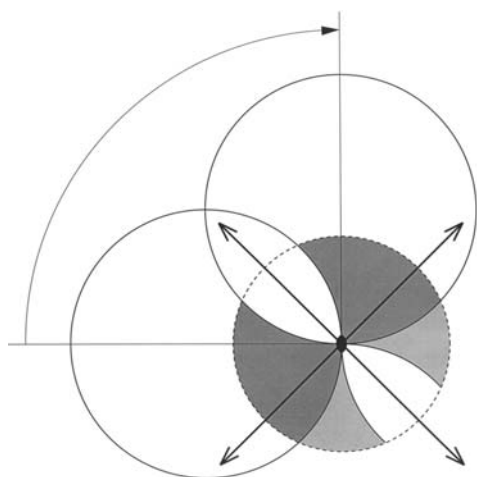


Figure 9.1.7.5

Rotation of an orthorhombic crystal by 90° between two diagonal orientations leaves a part of the reciprocal space unmeasured.

Table 9.1.7.2

Rotation range (°) required in different crystal classes

The direction of the spindle axis is given in parentheses; *ac* means any vector in the *ac* plane.

Point group	Native data	Anomalous data
1	180 (any)	180 + 2θ _{max} (any)
2	180 (<i>b</i>); 90 (<i>ac</i>)	180 (<i>b</i>); 180 + 2θ _{max} (<i>ac</i>)
222	90 (<i>ab</i> or <i>ac</i> or <i>bc</i>)	90 (<i>ab</i> or <i>ac</i> or <i>bc</i>)
4	90 (<i>c</i> or <i>ab</i>)	90 (<i>c</i>); 90 + θ _{max} (<i>ab</i>)
422	45 (<i>c</i>); 90 (<i>ab</i>)	45 (<i>c</i>); 90 (<i>ab</i>)
3	60 (<i>c</i>); 90 (<i>ab</i>)	60 + 2θ _{max} (<i>c</i>); 90 + θ _{max} (<i>ab</i>)
32	30 (<i>c</i>); 90 (<i>ab</i>)	30 + θ _{max} (<i>c</i>); 90 (<i>ab</i>)
6	60 (<i>c</i>); 90 (<i>ab</i>)	60 (<i>c</i>); 90 + θ _{max} (<i>ab</i>)
622	30 (<i>c</i>); 90 (<i>ab</i>)	30 (<i>c</i>); 90 (<i>ab</i>)
23	~60	~70
432	~35	~45

reliable estimations for cubic crystals, since then the requirements vary dramatically with the crystal orientation.

In the above, it was assumed that the detector was mounted centrally with respect to the incident X-ray beam. If it is offset either by a 2θ arm or by a translation, then the completeness for any total rotation range will be reduced. Software will generally be required to estimate the effective completeness and derive optimum strategies. For minimalist approaches to obtaining a high completeness, the importance of selecting the total rotation range, the optimal starting point and indeed the crystal orientation must be stressed. This means that the crystal orientation must be defined at the start of the experiment from the initial exposures.

9.1.7.2. Total rotation range for anomalous-dispersion data

In the presence of anomalous-scattering centres, Friedel's law breaks down and the intensities of the two halves of the reciprocal sphere are no longer equivalent. Strictly speaking, reflections related by a centre of symmetry or mirror relation cease to have equal intensities, but those related by pure rotation preserve their equivalence. The non-equivalent pairs of reflections are known as Bijvoet pairs. In macromolecular crystallography, it is often highly desirable to record the intensity differences between the Bijvoet mates to provide information on the position of anomalous scatterers, usually to be exploited in phasing procedures (Part 14). The anomalous signal should also be retained for so-called native data, for example in the discrimination between water and ions in the surface solvent shell.

This implies that the intensities of the unique reflections have to be measured for both hemispheres of reciprocal space. In the general (triclinic) case, this requires the rotation of the crystal by a wider rotation range. At very low resolution, the surface of the Ewald sphere can be approximated by a plane. In this case, rotation of the lower half of the Ewald sphere will cover a full hemisphere of data, and the upper half will cover the remaining centrosymmetrically related hemisphere. At high resolution, the surface of the Ewald sphere increasingly deviates from planarity by the Bragg angle θ on each side (Fig. 9.1.7.6). To record complete anomalous data for such a triclinic crystal therefore requires it to be rotated by 180° + 2θ_{max} from a random starting position. This will measure each Bijvoet mate at least once. However, only after a total rotation of 360° will the average multiplicity reach a value of two.

Similar reasoning applies to higher-symmetry space groups. Intensity data for two asymmetric units related by a centre of

9. X-RAY DATA COLLECTION

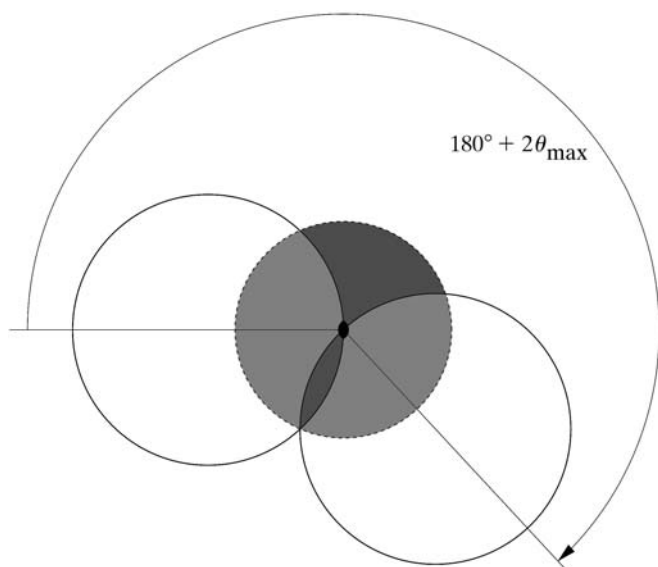


Figure 9.1.7.6

For data containing an anomalous signal, when both Bijvoet mates have to be measured, 180° rotation of a triclinic crystal is not sufficient and at least an additional $2\theta_{\max}$ is required.

symmetry or a mirror need to be recorded. For some cases, the total range remains the same for completeness of anomalous data as for native. However, in several symmetries or orientations, the total range must again be increased by either θ_{\max} or $2\theta_{\max}$ (Table 9.1.7.2).

9.1.7.3. Blind region

Even after rotation of the crystal about a single axis by 360° , some reflections do not cross the surface of the Ewald sphere and cannot be measured. These lie in a cusp around the rotation axis which is referred to as the blind region. This is in principle a disadvantage of the single-rotation method, but for most systems the problems are easily overcome. Owing to the curvature of the Ewald sphere, the width of the blind region increases with the resolution and directly depends on a single parameter, the diffraction angle θ (Fig. 9.1.7.7). The fraction, B_θ , of unrecordable reflections lying in the blind region at a particular resolution with

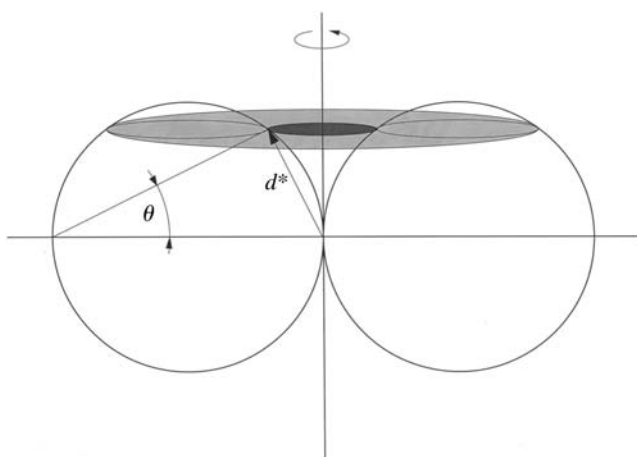


Figure 9.1.7.7

Rotation by 360° leaves the part of the reciprocal space in the blind region unmeasured, since the reflections near the rotation axis do not cross the surface of the Ewald sphere. The rotation axis in this projection lies vertically in the plane of the figure.

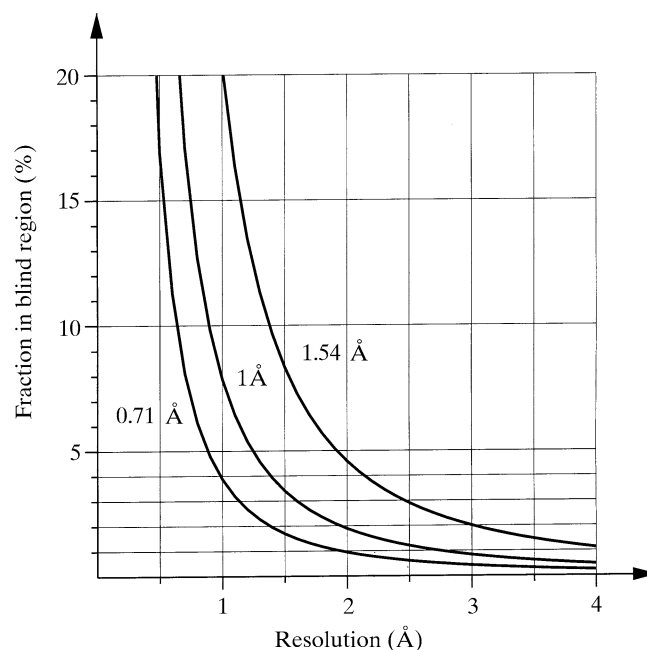


Figure 9.1.7.8

Dependence of the total fraction of reflections in the blind region on the resolution for three different wavelengths: 1.54, 1 and 0.71 Å.

Bragg angle θ is given by

$$B_\theta = 1 - \cos \theta.$$

The cumulative fraction, B_{tot} , of reflections in the blind region up to a certain resolution is given by

$$B_{\text{tot}} = 1 - 3(4\theta - \sin 4\theta)/(32 \sin^3 \theta).$$

B_{tot} is shown graphically as a function of resolution for selected wavelengths in Fig. 9.1.7.8.

For a particular resolution limit, the blind region is narrower if the wavelength is short, since the surface of the Ewald sphere is flatter (Fig. 9.1.7.9). This is an advantage of using short-wavelength radiation. For Cu $K\alpha$ radiation at 2.0 Å resolution, the blind region amounts to less than 5%. With shorter wavelengths, it falls below 2%.

The two halves of the blind region on either side of the Ewald sphere are related by the centre of symmetry. In the triclinic case, the blind region is therefore unavoidable with a single mount of

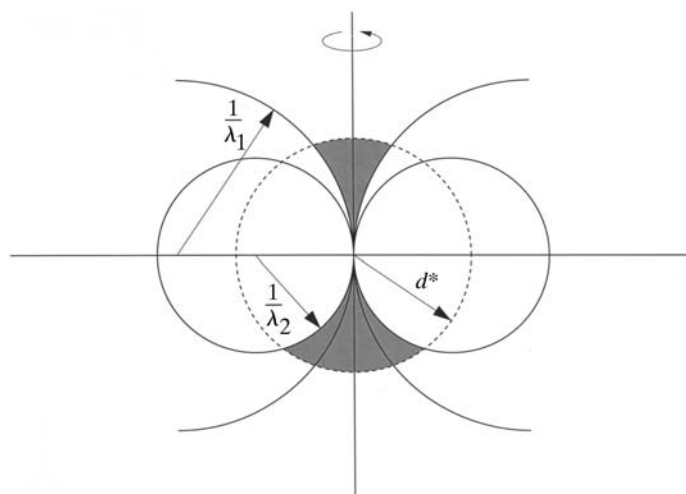
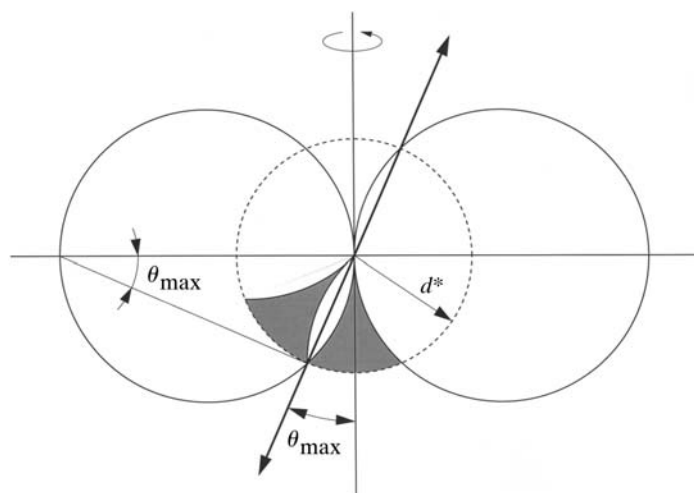


Figure 9.1.7.9

For shorter wavelengths the blind region is narrower, since the Ewald sphere is flatter.

**Figure 9.1.7.10**

If the crystal has a symmetry axis, it should be skewed from the rotation axis by at least θ_{\max} to be able to collect the reflections equivalent to those in the blind region.

the crystal. The only solutions are to use a second mount of the crystal offset by at least 2θ from the first, easily achievable with a κ -goniostat, or to measure from a second sample.

For crystals with symmetry higher than $P1$, reflections that are symmetry-equivalent to those in the blind region may be recorded, and there will be no loss of unique reflections. Only if the unique axis passes through the blind region approximately parallel to the spindle axis will the reflections lying close to it not be repeated by symmetry in another region of reciprocal space. To avoid the blind region, it is sufficient to misorient the unique symmetry axis by at least θ_{\max} from the rotation axis (Fig. 9.1.7.10). To achieve full completeness, monoclinic crystals should not be oriented along the unique twofold axis or along any vector in the ac plane.

The reciprocal-lattice points on the border of the blind region cross the surface of the Ewald sphere at a very acute angle or fail to cross it completely, staying in the diffracting position for a considerable time. Their intensity cannot be measured accurately, because the Lorentz factor is large and its magnitude is very sensitive to minor errors in the orientation matrix. These reflections are located on the detector window along the line parallel to the spindle axis and should not be integrated.

The detrimental effect of the blind region on the completeness of data is negligible at medium and low resolution or if the crystal is non-axially oriented. This means that a simple single rotation axis is sufficient for the majority of applications.

9.1.7.4. Alternative indexing

If the crystal point-group symmetry is lower than the symmetry of its Bravais lattice, then the reflections can be indexed in more than one way. In other words, the symmetry of the reflection positions is higher than the symmetry of the distribution of their intensities. This situation typically arises for point groups with polar axes, such as groups 3, 4 or 6, which can be indexed with the c axis pointing in either one of two directions. The lattice does not define the directionality of such axes if its two remaining cell dimensions are equivalent. This problem does not occur in the monoclinic system, despite the polar twofold axis, as the two other axes are not equivalent. The most complex case is point group 3, which can be indexed in the 622 lattice in four non-equivalent ways. The other such groups have only two alternatives.

Table 9.1.7.3

Space groups with alternative, non-equivalent indexing schemes

Symmetry operations required for re-indexing are given as relations of indices and in the matrix form. In brackets are the chiral pairs of space groups indistinguishable by diffraction. These space groups may also display the effect of merohedral twinning, with the twinning symmetry operators the same as those required for re-indexing.

Space group	Re-indexing transformation	
$P4, (P4_1, P4_3), P4_2, I4, I4_1$	$hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
$P3, (P3_1, P3_2)$	$hkl \rightarrow \bar{h}\bar{k}l$	$\bar{1}00/0\bar{1}0/001$
	or $hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
	or $hkl \rightarrow \bar{k}\bar{h}l$	$0\bar{1}0/\bar{1}00/00\bar{1}$
$R3$	$hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
$P321, (P3_121, P3_212)$	$hkl \rightarrow \bar{h}\bar{k}l$	$\bar{1}00/0\bar{1}0/001$
$P312, (P3_112, P3_212)$	$hkl \rightarrow \bar{h}\bar{k}l$	$\bar{1}00/0\bar{1}0/001$
$P6, (P6_1, P6_5), (P6_2, P6_4), P6_3$	$hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
$P23, P2_13, (I23, I2_13), F23$	$hkl \rightarrow \bar{k}\bar{h}l$	$010/\bar{1}00/00\bar{1}$

There is an analogous problem for cubic space groups within point group 23. Here the lattice possesses fourfold symmetry, but the intensity distribution has only twofold symmetry. Rotation by 90° leads to alternative, although perfectly permitted, indexing of reflections.

Each allowed scheme is permitted and self-consistent for a single crystal, since all possibilities will perfectly match the crystal lattice. However, under alternative indexing schemes, the same reflection will be given different indices, which can pose problems when data from more than one crystal are to be merged or compared. Merging is needed when more than one sample is required to record a complete data set. Comparison is needed when looking for heavy-atom derivatives or for ligand complexes with isomorphous crystals. For these, the reflections of one crystal must be selected as a standard, and it is easy to make other crystals consistent with this standard either by changing the orientation matrix at the time of intensity integration or by applying re-indexing to the integrated intensity set. The alternative indexing schemes are related by those symmetry operations present within the higher symmetry of the Bravais lattice but absent from the point-group symmetry. The point groups with alternative indexing systems are shown in Table 9.1.7.3, together with the necessary symmetry operations for re-indexing.

Several experiments require the recording of multiple data sets from the same crystal. One example is the collection of more than one pass with different exposure times (see below), and a second is in multiwavelength anomalous dispersion (MAD) experiments. In these experiments, the software systems may independently choose any of the alternative systems for different sets, which may then be incompatible and need re-indexing. It is much simpler to ensure a common orientation matrix modified as appropriate for all sets at the time of intensity integration.

9.1.8. Crystal-to-detector distance

The crystal-to-detector distance (CTDD) should be selected so that the whole area of the detector is usefully exploited. The shorter the CTDD, the higher the resolution of the indexed reflections at the edge of the image; but if the CTDD is too short, then the outer regions of the detector window record only indices with attached noise rather than intensities. A longer CTDD spreads the background radiation over a larger area of the detector as the background level diminishes in proportion to the square of the CTDD. In contrast, owing to collimation and focusing, the profiles of the Bragg reflections do not broaden so

9. X-RAY DATA COLLECTION

much, and the signal-to-noise ratio is enhanced at longer distances. It is advantageous to use the largest possible CTDD under the condition that meaningful data extend to, but not beyond, the active edge of the detector.

It is not straightforward to judge the resolution limit of meaningful diffraction. The most scientific approach involves recording, processing and merging a small number of images and making a decision on the basis of the resulting intensity statistics. However, this does require time, which should only pose a problem on ultra-high-intensity sources with very rapid data collection. A more pragmatic approach relies on visual inspection of the initial exposures using a graphical display at various contrast levels. Normally, if reflections are not visible by eye at the highest display contrast, their intensities are not meaningful. Some safety margin can be applied by setting the CTDD to a slightly shorter value than that estimated from visual inspection. Naturally, the resolution limit to which meaningful intensities extend depends on the exposure time, and the decision concerning the CTDD should follow the selection of the appropriate exposure (Section 9.1.11.2).

In addition to the significance of the reflection intensities, another important factor is the spatial resolution of spot profiles on the detector. If the crystal cell dimensions are large, the profiles may superimpose and the reflections may be impossible to integrate. At longer CTDD, the diffraction pattern spreads out and the profile overlap diminishes. If necessary, the detector can be offset from the central position to measure high-resolution data at long CTDD, but a larger total rotation is required to reach full data completeness. This applies only if the overlap of profiles belonging to the same lune results from a long axis lying parallel to the detector plane. The superposition of reflection profiles resulting from overlapping lunes will not be alleviated by increasing the CTDD; the only remedy for this is to reduce the rotation range $\Delta\varphi$ per exposure.

In addition to the proper selection of the CTDD, attention should be paid to the proper positioning of the beam stop. It should be centred with respect to the direct beam and cover the beam cross section completely. No part of the direct beam should reach the detector, and there should be no indirect scatter by the beam stop. The optimal reduction of air scatter is to have the smallest beam stop consistent with the dimensions of the beam, placed as close to the crystal as possible. For a given size of beam stop, the crystal-to-beam stop distance should be matched to the CTDD, sufficiently far from the crystal to minimize its shadow and concomitant obstruction of the valuable lowest-resolution reflections. If the beam stop is mounted on a metal wire, it is better to position the wire along the spindle axis where it will only interfere with those reflections around the blind region.

9.1.9. Wavelength

The wavelength of X-radiation can be tuned only at synchrotron sources. Rotating-anode generators produce radiation at a fixed wavelength which is characteristic of the metal of the anode, usually copper with $\lambda = 1.542 \text{ \AA}$.

The proper selection of the wavelength is most important for collecting data containing an anomalous-scattering signal. In general, the imaginary component $\Delta f''$ of the anomalous-dispersion signal is high on the short-wavelength side of the absorption edge of the anomalous scatterer present in the crystal. Near the absorption edge, both components, real $\Delta f'$ and imaginary $\Delta f''$, vary significantly. This variation is utilized in the

MAD technique, the strict requirements of which are discussed in Chapters 14.2 and 14.3.

If the data are collected using a single wavelength with the aim of measuring Bijvoet differences, $\Delta F_{\text{anom}} = F^+ - F^-$, the requirements are not as strict as for MAD. However, it may be advisable to record the fluorescence spectrum around the region of the expected absorption edge. If the fluorescence signal from the crystalline sample is too weak, the appropriate metal or salt standard can be used. When using anomalous scatterers displaying large white lines within their spectra, the wavelength should be accurately adjusted on the basis of the spectrum measured from the actual sample.

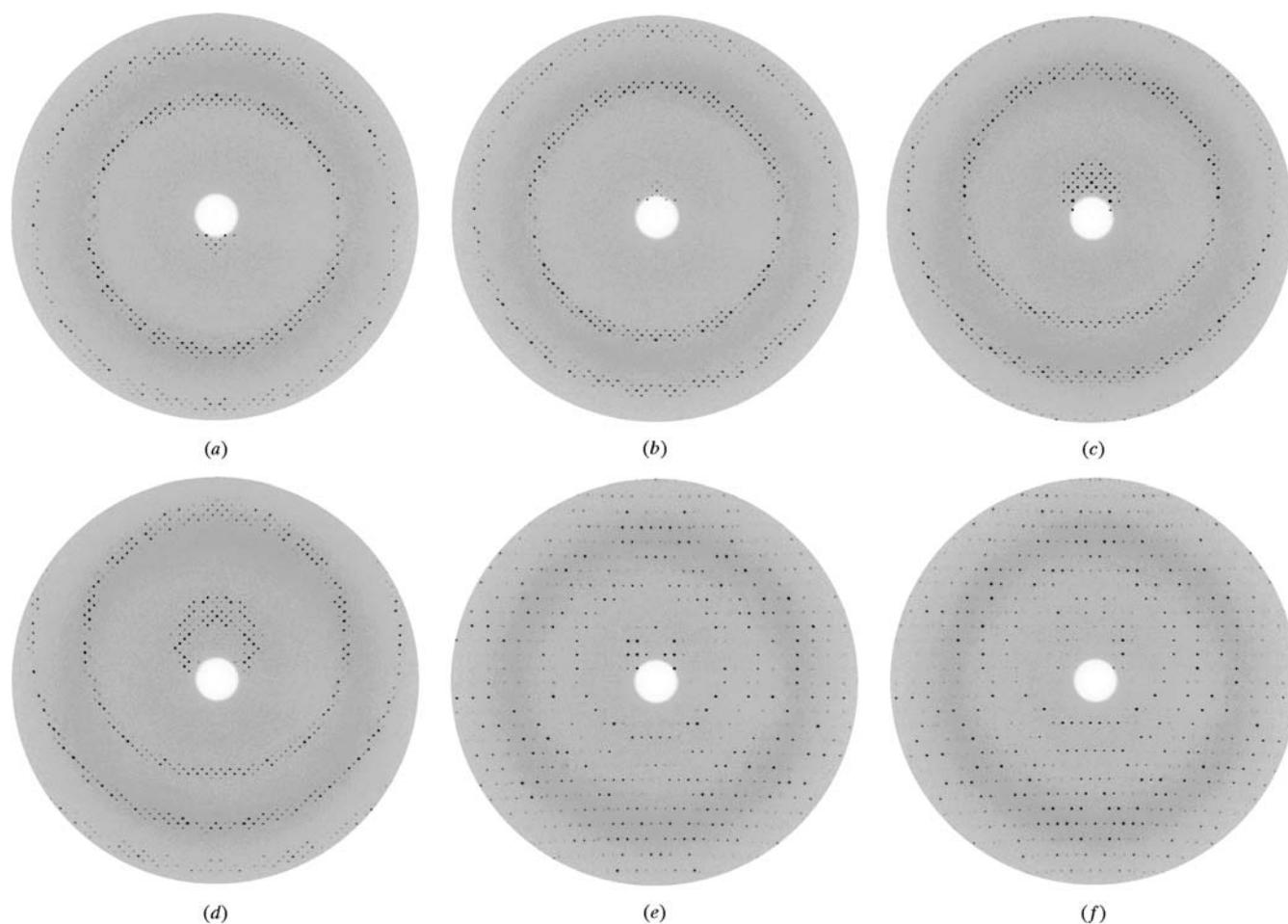
For collecting data without an anomalous signal, there are no strict requirements concerning the wavelength. The maximum intensity provided by the beamline depends on the energy of particles in the synchrotron storage ring and on the beamline optics. Typically, wavelengths around 1 \AA or shorter are used at most synchrotrons, assuring high beam intensity and low absorption of X-rays by the sample and air, thus reducing the radiation damage of the crystal. This is of particular importance at the very bright beamlines at third-generation synchrotrons. To diminish the effect of air absorption further, it is possible to fill the space between the crystal and the detector with helium. Short wavelengths are advantageous for collecting high-resolution data, since the diffraction angles are smaller and there is no need to use a very short CTDD. The effect of profile elongation owing to the oblique incidence of diffracted X-ray beams on the detector is then smaller, and the blind region is narrower.

9.1.10. Lysozyme as an example

Tetragonal hen egg-white lysozyme (Chapter 25.1 and Blake *et al.*, 1967), crystallizing in the space group $P4_32_12$ with cell dimensions $a = b = 78.6$ and $c = 37.2 \text{ \AA}$, is used here as a model system to illustrate some of the points made above, based on Dauter (1999). The example involves a set of two consecutive blocks of images with a crystal-to-detector distance of 243 mm, a wavelength of 0.92 \AA , a resolution of 2.7 \AA , an oscillation range of 1.5° and a crystal mosaicity around 0.5° . These images are shown in Fig. 9.1.10.1(a-f).

The first four images, (a-d), were exposed with the tetragonal fourfold c axis lying approximately along the direction of the beam. On these images, the reflections within each lune are arranged in a square grid, reflecting the tetragonal symmetry with $a = b$. The squares are oriented with their diagonals in the horizontal and vertical directions of the image, as the crystal was mounted with its [110] direction along the spindle rotation axis. Indeed, at the end of image (a) and the start of image (b), the c axis lay almost perfectly along the beam, and the zero-layer lune almost disappears behind the beam-stop shadow, since the corresponding ($hk0$) plane in reciprocal space is tangential to the Ewald sphere at the origin of the reciprocal lattice.

The lunes are widely spaced with clear gaps between them, because the third cell dimension, c , which is perpendicular to the detector plane, is relatively short, 37.2 \AA . Images (e-f), exposed at an angle on the rotation spindle roughly 90° away from (a-d), have a quite different appearance, despite the rotation range per image being the same. Each lune is less densely populated by reflections, but the number of lunes is larger and the gaps between them much smaller. This arises from the lunes now being parallel to the (hhl) family of planes, as the $[\bar{1}\bar{1}0]$ vector is now parallel to the beam. The interplanar spacing within this family is less than for those on images (a-d), hence at high resolution,

**Figure 9.1.10.1**

Images recorded from a crystal of lysozyme. (a–d) Four consecutive exposures with the crystal fourfold axis parallel to the X-ray beam. (e–f) Two successive exposures 90° away, when the fourfold axis lies vertically in the plane of the image. The crystal [110] direction is parallel to the rotation axis, horizontal in the plane of the images.

close to the edge of the detector window, the lunes overlap on images (e–f). The reflections, however, do not overlap, as the crystal orientation is diagonal; the lunes are sparsely populated, with large separation between adjacent spots, so the reflections on successive lunes fit between one another. It should be noted that the density of reflections in different regions of the reciprocal lattice is constant, and that the total number of reflections recorded on an image depends only on the rotation range, not on the crystal orientation.

The zero-layer lune containing reflections with indices $hk0$ is especially evident on exposures (c–d) directly above the centre of the image. With such a lune close to the centre, the reciprocal lattice shows minimal distortion owing to its projection onto the detector plane, and the lune appears as a ‘pseudo-precession’ pattern. The systematic absence of every second reflection, with odd index, along the $h00$ and $0k0$ lines indicates the presence of twofold screw axes of symmetry along the crystal axes a and b . Images (e–f), 90° away, have the hhl lune at the centre and, although it is less well separated from higher lunes, the presence of a fourfold screw axis along c is confirmed by the presence of only every fourth reflection on the $00l$ line. This allows the identification of the space group as $P4_12_12$ or its enantiomorph, $P4_32_12$. In general, the positions of the reflections define only the Bravais lattice, and it is the symmetry of the intensity pattern which reflects the point group. Thus, further confirmation that the symmetry belongs to point group $P422$ rather than $P4$ comes from the symmetric relation of the intensity distribution on either side of each lune in images

(a–d). This is equivalent to the earlier use of precession photography for space-group elucidation.

Close inspection shows that the reflections at the edges of the lune are also present on the adjacent image. The rotation range was 1.5° , and the mosaicity was estimated at 0.5° , and thus about one-sixth of the reflections are partially recorded at each edge of the lune, giving one-third partially recorded terms in total. The lack of sharpness at the edge of the lunes confirms a substantial level of mosaicity.

9.1.11. Rotation method: qualitative factors

9.1.11.1. Inspection of reflection profiles

Reflection profiles should be checked on the first recorded images. Very often a quick inspection of the profiles can disqualify a bad crystal without further loss of time. The profiles should have a single maximum and smooth shoulders. If the crystal shape is irregular, it may be reflected in the spot profile. Profiles should not have double maxima or be substantially elongated or smeared out, which usually arises from crystal splitting. The profiles should certainly be inspected if initial autoindexing of the diffraction pattern is unsuccessful.

Even if the spot profiles appear to be regular on the first image, it is good practice to inspect a second image at a substantially different φ rotation angle, preferably 90° away, since crystal splitting may have a similar effect on the appearance of the lunes and profiles as does high mosaicity on a single image (Section

9. X-RAY DATA COLLECTION

9.1.6.3). High mosaicity and splitting (often incorrectly referred to as twinning) must not be confused. If two parts of a split crystal are slightly rotated with respect to one another around a certain axis, the diffraction patterns will look different depending on the orientation. When such an axis is perpendicular to the detector plane, the spots will be doubled or smeared out. When the axis is parallel to the detector plane, the profiles resulting from the two parts of the crystal will overlap almost perfectly, but the lunes will be broadened, similar to the effect of high mosaicity.

After indexing the diffraction pattern, the integration profiles should be matched with the size and shape of the diffraction spots. The spots should not extend into the area defined as background. Selection of integration profiles that are too small will lead to incorrect integration of intensities. In contrast, if the profile areas are too large then the standard uncertainties will be wrongly estimated.

9.1.11.2. Exposure time

According to the principles of counting statistics, the longer the exposure, the better the signal in the data. The standard uncertainty of the measurement is equal to the square root of the number of counts, and the signal-to-noise ratio increases with the accumulated counts. In practice there are limitations to this rule.

The dynamic range and saturation limit of the detector is one limiting factor. It may be impossible to measure adequately the strongest as well as the weakest reflection simultaneously, since their intensities differ by several orders of magnitude. If the exposure time is long enough to record the weakest intensities, then in general at low resolution the most intense reflections may saturate some pixels within their profile on the detector. Such reflections are termed ‘overloads’ and this problem will be addressed in Section 9.1.11.3.

Exposure time can be limited by the total time available for the experiment. This is often a particularly acute problem for synchrotron-data collection, with high oversubscription of beamlines. The decisions concerning exposure time depend on the expected application of the data, since different applications have different requirements, as addressed in Section 9.1.13. Within the given time constraints, the first priority should be data completeness, even at the expense of underexposure. In this context, it is useful to recall that to increase the statistical signal-to-noise ratio by a factor of two, it is necessary to prolong the exposure time by at least a factor of four.

9.1.11.3. Overloads

Some detectors, or their associated read-out systems, are limited in the number of counts they can accumulate in one pixel. The number recorded reaches a maximum number which cannot be further increased, *i.e.* the pixels can become saturated. This means that these pixels retain the same maximum value on longer exposure whilst other, non-saturated, pixels continue to accumulate counts. The intensity in saturated pixels will hence be underestimated compared to the others and any intensities estimated from profiles including such pixels will be biased towards low values. It is essential that pixels that are saturated are flagged and recognized by the processing software. There are several ways to deal with the problem of saturation.

(1) Reject all reflections that contain saturated pixels. These will tend to be at low resolution. If more than a very few are rejected, this can be a truly disastrous choice, especially if the data are to be used for molecular replacement. In addition, missing the largest terms degrades the continuity and infor-

mation content of all electron-density maps derived therefrom. This point is relevant to several applications (Section 9.1.13).

- (2) Reject only those pixels that are saturated, and fit average standard profiles estimated from the non-saturated spots. This gives a poorer estimate than if the pixels were not saturated, but for applications such as molecular replacement or direct methods where the high-intensity data are essential it is certainly better than option (1).
- (3) Reduce the exposure time to ensure that there are no overloaded pixels. This is a trade-off, because if there is a large contrast between the intensity of the weakest and the strongest terms in the pattern, then the weaker terms will have a low and possibly unacceptable signal-to-noise ratio under this regime.
- (4) Use more than one pass through the rotation range, with different exposure times. The longest exposures should be sufficient to ensure that the intensities of the data at the high-resolution limit of the pattern are statistically significant. The shortest should ensure that the number of saturated pixels in the ‘low-resolution’ pass is minimized. If the contrast between the low- and high-resolution passes is too great, differing by a factor of much more than about ten, then additional passes with intermediate exposure times should be used to allow satisfactory scaling of the data from these images. The CTDD for each pass with shorter exposure should be increased only so as to cover the resolution to which reflections were saturated on the previous pass. The rotation range on individual images can then be increased accordingly, in the wide φ -slicing option. On bright synchrotron beamlines, if the second pass requires exceedingly fast rotation of the spindle-axis motor and rapid opening and closing of the beam shutter beyond the limit of reliability, it may be better to attenuate the beam, for example with a series of aluminium foils. As discussed in Section 9.1.7.1, if high-resolution data are collected in several passes with different exposures and resolution limits, it may not be necessary to cover all of the theoretically required rotation range in the highest-resolution pass. The curvature of the Ewald sphere results in the high-resolution data being completed with a smaller total rotation range than the low. It is vital that the lowest-resolution pass covers the total rotation range required for complete data.

Clearly, the optimum solution is to have a detector with a sufficient dynamic range to cover pixels of both weak and strong reflections. The dynamic range has already been increased with recent imaging plates and CCDs. Enhanced dynamic range may prove to be the most important advance of solid-state pixel detectors.

An additional advantage of the fine-slicing approach is that it leads to fewer overloads. Each reflection profile is divided between several separate images and as a result the effective dynamic range of the detector is increased.

9.1.11.4. R factor, $I/\sigma(I)$ ratio and estimated uncertainties

It is customary to judge data quality by the overall R_{merge} , calculated using the squares of the structure-factor amplitudes (often erroneously called intensities):

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle|}{\sum_{hkl} \langle I_{hkl} \rangle}.$$

R_{merge} provides a measure of the distribution of symmetry-

9.1. PRINCIPLES OF MONOCHROMATIC DATA COLLECTION

equivalent observed intensities around the average value. However, the most popular form of R_{merge} given above is not a proper, statistically valid quantifier. It does not take into account the multiplicity of the measurements and, as a consequence, it actually rises with increased multiplicity, falsely indicating degradation of the data quality when in reality they have a higher accuracy. Modifications of R_{merge} have been proposed to include the effect of multiple measurements properly (Diederichs & Karplus, 1997; Weiss, 2001) (see Chapter 2.2).

A better quantity for assessing the quality of the X-ray data is the $\sum_{hkl} I_{hkl} / \sum_{hkl} \sigma(I_{hkl})$ ratio, provided the standard uncertainties, $\sigma(I)$, are correctly estimated. Detectors such as imaging plates or CCDs do not measure individual X-ray quanta directly, having a gain factor dependent on the response of the individual detector pixel to a single X-ray photon. If the gain factor is not known accurately for a particular detector, the resulting standard uncertainties of the measured intensities will be estimated at an incorrect level. If the multiplicity of the reflections is higher than unity, it is possible to correct the uncertainties *a posteriori*. This can be done either from a comparison with the expected values using the χ^2 test, or by using the *t*-plot. The latter requires that the ratio of the differences between equivalent intensity measurements to their standard uncertainties, $t = (I_i - \langle I \rangle) / \sigma(I_i)$, follows a normal distribution with a mean of 0.0 and standard deviation of 1.0. Both of these methods assume the errors have a normal distribution, and that only the mean and width have been incorrectly estimated and should be appropriately adjusted. They cannot take into account systematic errors of measurement.

The data-merging procedure in addition allows the identification of statistical ‘outliers’ and their exclusion from the data (Read, 1999). Outliers are defined as those observations that lie sufficiently far from the mean of a set, and assumption of a normal distribution suggests they suffer from substantial systematic errors of measurement. In a crystallographic experiment, outliers are those intensity measurements that deviate unexpectedly from the mean intensity of a set of symmetry-equivalent reflections. In the recording of rotation data, one typical source of such systematic errors is erroneous classification of reflections predicted as partially or fully recorded. This is a severe problem for those reflections lying close to the blind region. A second example is the presence of so-called ‘zingers’ in individual CCD detector pixels caused by scintillations from trace radioactivity of the taper glass. Other problems such as shadowed or inactive regions of the detector window give rise to a range of such systematic errors.

A small number of outliers may be expected from such causes. However, the total fraction of reflections flagged as outliers and rejected from the merging process should be small, certainly much less than 1%. Larger fractions indicate serious deficiencies in the hardware or the software and suggest something is very wrong with the experiment. There should always be a physical reason for rejecting outliers, other than just a need to reject those agreeing poorly with their symmetry-equivalent intensities in order to drive down R_{merge} . It is always possible to reduce R_{merge} and to provide an apparent ‘improvement’ in the data by rejecting a large percentage of measurements, but this is extremely bad practice.

Good crystallographic data depend strongly on an appropriate statistical procedure. It is also inappropriate to exclude those reflections with intensities lower than a cutoff limit, such as 1σ , before or during the process of data merging. Weak intensities also carry information and their neglect introduces bias into the

measured intensity distribution, affecting, for example, the overall or individual atomic temperature factors.

The true outer resolution limit of the diffraction pattern is not trivial to define and indeed depends to some extent on the application. If $I/\sigma(I)$ is higher than 1.0, then a resolution shell of data indeed contains some information in a statistical sense – provided of course that $\sigma(I)$ has been correctly estimated. However, as $I/\sigma(I)$ falls close to unity there will in practice be very few significant observations amongst a great deal of noise. It is necessary to make some decision about where to cut the effective resolution. For the application of direct methods, for example using *SHELXD* (Sheldrick, 2008), the cutoff is often defined as the resolution shell where $I/\sigma(I)$ falls to 2.0, when R_{merge} usually reaches 20–40% depending on the symmetry and redundancy. Cruickshank (1999*a,b*) has provided a formula for a data precision indicator (DPI) which includes the effect of falling $I/\sigma(I)$ ratio (see Chapter 2.2).

For other applications it may be advisable to accept even very weak data. Direct methods use only a subset of the most meaningful reflections, but which should extend to as high a resolution as possible. In addition, when the data are sparse from crystals that only diffract to very limited resolution, perhaps around 3 Å, then it is essential to retain all the experimental data, even if they are weak.

9.1.12. Radiation damage

9.1.12.1. Historical perspective

All crystals irradiated with X-rays absorb at least a fraction of the radiation, resulting in damage to the sample (Henderson, 1990). The energy from the absorbed photons may initially result in the disruption of chemical bonds, before being eventually dissipated as thermal energy. For well ordered small-molecule crystals the lattice is close-packed and the effects arising from the absorbed photons are restricted to the immediate environment of the absorption event, so-called primary damage. Only when a substantial fraction of the crystal has been affected do cooperative effects set in.

In contrast, roughly 50% of a macromolecular crystal is disordered aqueous solvent (Matthews, 1968). At room temperature this allows a secondary mechanism of radiation damage, resulting from diffusion of radicals and ions produced at the primary absorption site which affect chemical moieties at positions remote from this site. The details of this process remain poorly understood but are related to the extremely damaging effects of X-rays on biological tissue. A consequence of this damage is that degradation of the crystal order continues even after the irradiation is stopped or interrupted. For collection of data at room temperature from protein crystals mounted in capillaries, secondary damage contributes significantly to the rate of deterioration of the diffraction pattern. One of the gains of the early applications of SR was that it allowed recording of data to proceed ahead of the effects of secondary damage, increasing the effective, if not the absolute, lifetime of the crystal in the X-ray beam. An experiment often required several crystals, all of which showed the effects of temporal decay in their recorded intensities, which needed to be merged to provide complete data.

9.1.12.2. Cryogenic vitrification

In the early 1990s, the introduction of protein-data collection at cryogenic temperatures, using so-called flash cooling, was a major breakthrough (Garman & Schneider, 1997; Rodgers, 1997;

9. X-RAY DATA COLLECTION

Garman & Owen, 2006). Such vitrification of crystals largely prevented the effects of secondary damage. On the X-ray sources then available, it was in most cases possible to record complete data from a single sample without significant degradation of the diffraction, enormously simplifying the strategy of data collection and merging.

Almost all data are currently collected from vitrified samples (see Part 10). The prolonged life of the sample and modest rates of data acquisition, even at second-generation SR sources with imaging plates, allow enough time for careful analysis of the initial images and optimization of the strategy.

A second major advantage of cryogenic data collection is that it allows crystals to be reused after initial data have been recorded. Two examples show the usefulness of this approach. Firstly, when screening the binding of heavy atoms for phase determination or ligands for complex formation, data can first be recorded to the minimum resolution needed to determine whether the binding is successful. Secondly, a series of vitrified crystals can be screened for their degree of order in the home laboratory, and the best stored and retained for subsequent improved collection either in the home laboratory or at a synchrotron site. The ability to transport vitrified crystals has proved invaluable in this respect, and leads to optimal use of synchrotron resources.

9.1.12.3. High-intensity third-generation SR sources

The advent of third-generation SR sources and insertion devices has led to X-ray beams of unprecedented intensity. The speed of data collection can be of the order of 1 second per 1° rotation. In association with CCD detectors able to read out images within a few seconds, this means that a complete data set can be obtained in a few minutes. At first sight, this would seem to have solved the problem of macromolecular data collection, as such speeds should allow recording of highly redundant accurate data to the highest resolution in a tractable time.

However, with such high intensities it appears that the effects of radiation damage are significant and result in specific effects on susceptible parts of the structure. The useful active exposure lifetime of typical crystals seems to be around five minutes, with substantial degradation of the diffraction pattern ensuing even for vitrified crystals. The first manifestation of radiation damage is the disruption of disulfide bridges and decarboxylation of aspartates and glutamates. This effect means appropriate strategies for selecting the optimal starting point of rotation in order to minimize the total rotation required for collection of complete data are once more essential. Several strategy programs, such as *BEST* (Popov & Bourenkov, 2003; Bourenkov & Popov, 2006), now permit this to be done effectively.

9.1.12.4. Correcting data for the effects of radiation damage

The overall effect of radiation damage is that the higher-resolution intensities decrease faster than those at low resolution. This effect is largely taken into account by the relative *B* factors applied to individual images during data scaling and merging by the major data-reduction programs.

However, such scaling does not allow for the effect of specific structural damage (*e.g.* the S–S bridges and carboxylic groups) on individual reflection intensities. A method to deal with this has been proposed by Diederichs *et al.* (2003). This is based on a zero-dose extrapolation of intensities and requires that a timestamp be attached to each individual intensity measurement.

Such a timestamp has also been used to assist in estimation of phases by *SHARP* (Schiltz *et al.*, 2004).

9.1.13. Relating data collection to the problem in hand

The data-collection protocol should be matched to the purposes for which the data are to be used. Different applications present a range of different needs, requiring the intensities (or structure-factor amplitudes) to be exploited in different ways. In this section a representative set of applications is outlined in terms of how the tactics and strategies of data collection can vary.

9.1.13.1. Isomorphous-anomalous derivatives

The phasing of proteins by isomorphous replacement requires the collection of data from crystals of one or more heavy-atom derivatives of the protein that are isomorphous to the parent native crystal. Preparation of derivatives involves either soaking of native crystals in the heavy-atom solution or co-crystallization with the heavy-atom reagent (Part 12). Data collection can be split into two parts. The first step is to establish whether a potential derivative is isomorphous and contains the expected heavy atoms. The second is to collect the data on this derivative to provide the necessary phase information for the native structure factors. The problems of how to utilize the phase information are addressed in Part 12. Here, strategies applicable to the two steps are described.

Screening of derivatives can be carried out by collecting data to the resolution limits of the crystals. This can consume substantial data-collection resources and lead to irrelevant data that are not from isomorphous crystals or do not contain the anticipated heavy-atom signal. It is preferable to record the minimum data sufficient to identify a potential derivative in order to save time and resources, as many samples may need to be screened. A minimal strategy can exploit some or all of the following protocols:

- (1) An essentially complete native-data reference set should be available, although not necessarily to the ultimate resolution limit.
- (2) Preparation of a set of crystals with a selected set of potential heavy atoms, the number depending on crystal availability.
- (3) Collection of a small number of images from each potential derivative crystal, ideally on the home-laboratory rotating-anode source or an SR beamline if necessary. These data can be recorded to a low resolution: in principle 4 \AA or less should be enough. The resulting partial derivative data are scaled with the complete native set. The fractional isomorphous difference can be evaluated easily and compared with the expected agreement with the native data. In general, values less than 10% suggest that the heavy atom is not bound. Values higher than about 30% suggest an unacceptable level of non-isomorphism. Intermediate values suggest, but do not guarantee, that the derivative is worth pursuing. Normal probability plots can be helpful in this respect (Howell & Smith, 1992).
- (4) Given a positive result from point (3), complete data may be recorded on the same or an equivalent derivative crystal. Again, it may be useful to record data to low resolution in the first instance. 4 \AA resolution is again quite sufficient to solve the structure of a heavy-atom constellation using direct or Patterson methods, allowing the more complete characterization of the potential derivative.

9.1. PRINCIPLES OF MONOCHROMATIC DATA COLLECTION

- (5) If the compound proves to be a useful derivative, data can then be recorded to higher resolution for the computation of phase information. It may not be appropriate to record data to the highest resolution as for the native protein. In this context, the strength of the data is of primary importance, and relatively weak data at high resolution may be less relevant.

Some practical points are highly relevant here. The ability to store and reuse vitrified crystals means that potential derivatives can first be screened at the lowest possible resolution, and the crystal can be preserved and used later only if the derivative proves to provide useful phase information. The final resolution for data collection will then depend on the degree of isomorphism. The wavelength, if tunable, should be set to a value just below the absorption edge in order to maximize the anomalous signal. The redundancy can play an important role, as it is useful to have a large number of independent measurements so that outliers in the native or derivative data can be excluded, as these can cause major problems in either the Patterson or direct-methods approaches for locating the heavy atom (Part 12).

9.1.13.2. Anomalous scattering, MAD and SAD

The requirements for collecting data with an intrinsically weak anomalous signal are several. The highest possible resolution should not be the primary consideration. The emphasis is on data quality, as it is necessary to measure very small differences in structure-factor amplitudes, which are already in themselves relatively weak. Important considerations include the following.

- (1) Optimization of the wavelength, particularly for MAD experiments.
- (2) Ensuring that the anomalous data are complete in terms of all possible Bijvoet pairs. This is not always addressed by the currently available data-processing software.
- (3) High redundancy of measurements significantly enhances the quality of the signal, as this provides effective averaging of errors and allows the rejection of statistical outliers. The latter is especially important for direct-methods solution of the anomalous-scattering constellation.
- (4) However, the crystal lifetime is finite owing to the effects of radiation damage, which can introduce changes in intensities of the same order as the anomalous signal. For SAD (single-wavelength anomalous dispersion)/MAD data, the exposures should be limited to ensure the data are complete before the onset of substantial damage. This may well mean that the resolution limit should be set more modestly than for native data.

For MAD experiments (Hendrickson, 1999; Smith, 1991), which can only be carried out at SR sites, the optimum number of wavelengths at which data should be recorded remains unclear. Given finite beam time, the trade-off may be between measuring with limited redundancy at several wavelengths as against higher redundancy at a smaller number of wavelengths, or even at one wavelength.

SAD represents the limiting case. All data are recorded at one wavelength, reducing the requirement for fine monochromatization and for fine tunability and stability. Now quality, especially in the form of redundancy, is the dominating factor since all phasing is based purely on a single anomalous difference for each reflection. In recent years, SAD phasing has come to predominate in the number of novel structures deposited in the PDB.

9.1.13.3. Molecular replacement

For the initial data required for molecular replacement (MR), high resolution is not essential. Firstly, the method depends on homologous models that are usually only an imperfect representation of the structure under investigation, hence high-resolution data cannot be accurately modelled and will only introduce noise into the analysis. Secondly, the rotation function, the first step in MR, is based on the representation of the Patterson function in terms of spherical harmonics, which is limited in its accuracy.

In contrast, it is vital for MR applications that the most intense low-resolution terms are measured. The lack of such reflections strongly affects the rotation- and translation-function computations, as the functions are based on Patterson syntheses involving the square of the structure-factor amplitudes, and are dominated by the largest terms. Elimination of the strongest few per cent of the low-resolution data may well prevent a successful solution by MR.

However, for refinement of structures solved by MR, it is important that data be recorded to a resolution sufficient to allow escape from the phase bias introduced by the model. This is a key point. There are many examples where collection of data to a higher resolution has enabled the refinement of an MR solution which would not refine at the lower resolution.

9.1.13.4. Definitive data for refinement of protein models

All structures benefit from the highest accuracy in their atomic coordinates to shed light on the details of their biological function. These may include substrate or inhibitor complexes and mutants as well as native proteins where the analysis requires the full potential of X-ray crystallography. Many of these crystals will not diffract to atomic resolution; nevertheless, all steps in a detailed crystal structure analysis are made simpler as the resolution and quality of the data are increased. This includes solution of the phase problem, interpretation of the electron-density maps and refinement of the model.

The most appropriate strategy for data collection involves decisions based on a complex and mutually dependent set of parameters including:

- (1) Crystal quality and availability. If only one crystal is available, the choices are limited. If many are available, then some experimentation is recommended to select a high-quality sample. This is greatly aided by the recent introduction of automated sample changers and strategy software on a number of beamlines.
- (2) Cryogenic vitrification. In many cases, this allows collection of data from a single crystal. If appropriate cryogenic conditions cannot be established, making it necessary to record room-temperature data, this can affect strategy dramatically, in that several crystals might well be required to record the target resolution and completeness.
- (3) X-ray source and detector. The availability of these again places restrictions on the experiments that are tractable. An SR source will always provide better data, but has logistical problems of availability and access. For some problems, SR becomes *sine qua non* and a rotating anode is just insufficient. These include the use of MAD techniques, very small crystals, large and complex structures with large unit cells such as viruses, and where atomic resolution data are needed.
- (4) Overall data-collection time allocated. This has an obvious overlap with point (3). In particular, if SR is to be used later, then the resolution limit on the home source may be modest.

9. X-RAY DATA COLLECTION

If SR is not likely to be employed, then a higher resolution may be aimed for, requiring more time, and again dependent on the pressure on local resources.

Whatever the resource, it is good to define a strategy that will provide high completeness of the unique amplitudes at the highest resolution, with the realization that there may be some conflict between these two requirements owing to radiation damage.

9.1.13.5. A series of mutant or complex structures

The detailed geometry of the molecule is already known and the rather general effects of ligand binding or mutation can be initially identified at a relatively modest resolution and completeness. As with heavy-atom screening, it is often advisable to check that the desired complex or structural modification has been achieved by first recording data at low resolution.

However, if the analysis then proves to be of real chemical interest, with a need for accurate definition of structural features, the data should be subsequently extended in resolution and quality. As with the identification of isomorphous derivatives, this approach has benefited greatly from cryogenic vitrification, where the sample can be screened at low resolution and then preserved for subsequent use.

9.1.13.6. Atomic resolution applications

As for MAD data, the needs for atomic resolution data are extreme, but rather different in nature. Atomic resolution refinement is addressed in Chapter 18.4. Suffice it to say that by atomic resolution it is meant that meaningful experimental data extend close to 1 Å resolution. There are two principal reasons for recording such data. Firstly, they allow the refinement of a full anisotropic atomic model, leading to a more complete description of subtle structural features. Secondly, direct methods of phasing are dependent upon the principle of atomicity.

The problems to be faced include:

- (1) The high contrast in intensities between the low- and high-angle reflections. This may be much larger than the dynamic range of the detector. If exposure times are long enough to give good counting statistics at high resolution, then the low-resolution spots will be saturated. The solution is to use more than one pass with different effective exposure times.
- (2) The overall exposure time is often considerable and substantial radiation damage may finally result. The completeness of the low-resolution data is crucial, and it is strongly recommended to collect the low-resolution pass first as the time taken for this is relatively small.
- (3) The close spacing between adjacent spots within the lunes on the detector, dependent on the cell dimensions. The only aid is to use fine collimation.
- (4) The overlap of adjacent lunes at high diffraction angle, especially if a long cell axis lies along the beam direction. Using an alternative mount of the crystal is the simplest solution. Otherwise, the rotation range per image must be reduced, increasing the number of exposures. This was a problem with slow read-out detectors, but is largely alleviated with CCDs.
- (5) For direct-methods applications, a liberal judgement of resolution limit should be adopted. Even a small percentage of meaningful reflections in the outer shells can assist the phasing. These weak shells can be rejected or given appro-

priate low weights in the refinement. The strong, low-resolution terms are vital for direct methods.

9.1.14. The importance of low-resolution data

The low-resolution terms define the overall shape of the object irradiated in the diffraction experiment. Omission of the low-resolution reflections, especially those with high amplitude, considerably degrades the contrast between the major features of the object and its surroundings. For a macromolecule, this means that the contrast between it and the envelope of the disordered aqueous solvent is diminished and, furthermore, the continuity of structural features along the polymeric chain may be lost. Refinement and analysis of macromolecules at all resolutions, be they high or low, involves the inspection of electron-density syntheses. These can be interpreted visually, on a graphics station, or automatically with a variety of software. In all of these, at all resolutions, the importance of the low-resolution terms is crucial. A special problem is in the interpretation of the partially ordered solvent interface. The biological activity of most enzymes and ligand-binding proteins is located precisely at this interface, and for a true structural understanding of how they function this region should be optimally defined. This is seriously impaired by the absence of the strong, low-resolution terms. The problems become more severe as the upper resolution limit of the analysis becomes poorer. Thus, at 1 Å resolution the omission of the 7 Å data shell will have less effect compared with a 3 Å analysis – but remember that ideally, no low-resolution data should be omitted!

In some phasing procedures, the presence of complete, especially high-intensity, low-resolution, data is even more crucial. The big, low-resolution amplitudes dominate the Patterson function, and methods based on the Patterson function are therefore especially sensitive. This encompasses one of the major techniques of phase determination for macromolecules: molecular replacement. Direct methods of phase determination utilize normalized structure factors and predominantly exploit those of high amplitude. The relations between the phases of those reflections with high amplitudes, such as the classical triple-product relationship, are strongest and most abundant for reflections with low Miller indices, hence at low resolution.

The importance of the low-resolution reflections in terms of geometric and qualitative context cannot be overemphasized.

9.1.15. Data quality over the whole resolution range

It is not possible to judge data quality from a single global parameter, especially R_{merge} , not even from the overall $I/\sigma(I)$ ratio. Such a parameter may totally neglect problems such as the omission of all low-resolution terms due to detector saturation. A set of key parameters including $I/\sigma(I)$, R_{merge} , percentage completeness, redundancy of measurements and number of overloaded high-intensity measurements must be tabulated in a series of resolution shells. This information should be assessed during data collection to guide the experimenter in the optimization of the choice of such parameters as exposure time, attainable resolution and required redundancy. As stated in Section 9.1.13, the requirements will vary with the application.

The effect of sample decay also requires such tables. The X-ray intensities decay more rapidly at high angle than at low, and consideration of this effect requires knowledge of the relative B values that need to be applied to the individual images during data scaling. An often subjective decision will need to be made regarding at what stage the decay is sufficiently high that further

9.1. PRINCIPLES OF MONOCHROMATIC DATA COLLECTION

images should be ignored. The effects of damage are likely to be systematic rather than just random, and cannot be totally compensated for by scaling. This remains true even for cryogenically vitrified crystals, especially with ultra-bright synchrotron sources.

Following an earlier recommendation by the IUCr Commission on Biological Molecules (Baker *et al.*, 1996), this tabulated information, as a function of resolution, should be deposited with the data and the final model coordinates in the Protein Data Bank. Only then is it possible to have a true record of the experiment and for users of the database to judge the correctness and information content of a structural analysis.

9.1.16. Strategies for automated data acquisition

Progress in crystal-handling hardware has resulted in the development of sample changers both for synchrotron beamlines and the home laboratory. A sequence of samples contained in a Dewar can be mounted on the goniostat, centred in the beam and exposed to X-rays without user intervention. The gain in efficiency arises from the fact that manual intervention is no longer needed between samples, and at SR sites access to the hutch is avoided.

While this provides the potential for greatly increased throughput, it still requires intelligent decision-making software for evaluation of crystals and for optimal strategies of data collection to be achieved with minimal (ideally zero) user input. The steps required in such a system are:

- (1) Automated mounting and dismounting of samples from a Dewar.
- (2) Automated centring of the sample (or at least loop) in the beam.
- (3) Recording and interpretation of two images preferably 90° apart in crystal orientation.
- (4) Repetition of steps (1)–(3) for a number of samples of the same protein and ranking of these samples in terms of diffraction quality.
- (5) Selection of the best sample and definition of the optimum strategy for data collection, taking account of information provided by the user with regard to the minimum resolution *etc.*
- (6) Collection and integration of complete and ideally redundant data.

Robotic sample changers with at least some elements of the above are now operational at many sites [for example, see Leslie *et al.* (2002); McPhillips *et al.* (2002) and Cipriani *et al.* (2006), and also Chapter 9.2]. Considerable advances are expected in the near future, allowing routine automated screening of samples at major synchrotrons. In addition, remote access to synchrotrons, by submission of crystals in Dewars for collection by SR staff (the so-called Fedex procedures) or for direct control by the user from their home laboratory through the Internet, is now possible at a number of synchrotron facilities. All of these moves towards automation require electronic databases for the tracking and transfer of samples and their associated data and parameters.

9.1.17. Final remarks

Optimal strategies for data collection are dependent on a number of factors. The alternative data-collection facilities to which access is potentially available, how long it takes to gain access and the overall time allocated all place restraints on the planning of

the experiment. In view of this, it is not possible to provide absolute rules for optimal strategies.

Even after the source and overall time have been allocated or planned, the strategy is still the result of a compromise between several competing requirements. Some are general, others depend on the characteristics of a particular crystal or detector. As seen in the previous section, it is not possible to define protocols relevant for all applications. Rather, it is important to consider the relative importance of the parameters that can be varied to the problem in question and make the appropriate decisions.

Thus, data collection may have become easier from a technical point of view, but several crucial scientific decisions still have to be made by the experimenter. It is always beneficial to sacrifice some beam time and interpret the initial diffraction images, so as to avoid mistakes which may have an adverse effect on data quality and the whole of the subsequent structural analysis.

References

- Amemiya, Y. (1995). *Imaging plates for use with synchrotron radiation*. *J. Synchrotron Rad.* **2**, 13–21.
- Arndt, U. W. & Wonacott, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North Holland.
- Baker, E. N., Blundell, T. L., Vijayan, M., Dodson, E., Dodson, G., Gilliland, G. L. & Sussman, J. L. (1996). *Deposition of macromolecular data*. *Acta Cryst.* **D52**, 609.
- Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *On the conformation of the hen egg-white lysozyme molecule*. *Proc. R. Soc. London Ser. B*, **167**, 365–377.
- Bloomer, A. C. & Arndt, U. W. (1999). *Experiences and expectations of a novel X-ray microsource with focusing mirror. I*. *Acta Cryst.* **D55**, 1672–1680.
- Bourenkov, G. P. & Popov, A. N. (2006). *A quantitative approach to data-collection strategies*. *Acta Cryst.* **D62**, 58–64.
- Broennimann, Ch., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006). *The PILATUS 1M detector*. *J. Synchrotron Rad.* **13**, 120–130.
- Carter, C. W. Jr & Sweet, R. M. (1997). Editors. *Methods in Enzymology*, Vol. 276, pp. 183–358. San Diego: Academic Press.
- Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.-S., Caserotto, H., Cusack, S., Dallery, M., di-Chiaro, F., Guijarro, M., Huet, J., Larsen, S., Lentini, M., McCarthy, J., McSweeney, S., Ravelli, R., Renier, M., Taffut, C., Thompson, A., Leonard, G. A. & Walsh, M. A. (2006). *Automation of sample mounting for macromolecular crystallography*. *Acta Cryst.* **D62**, 1251–1259.
- Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography*. *Acta Cryst.* **D50**, 760–763.
- Cruickshank, D. W. J. (1999a). *Remarks about protein structure precision*. *Acta Cryst.* **D55**, 583–601.
- Cruickshank, D. W. J. (1999b). *Remarks about protein structure precision. Erratum*. *Acta Cryst.* **D55**, 1108.
- Dauter, Z. (1999). *Data-collection strategies*. *Acta Cryst.* **D55**, 1703–1717.
- Dauter, Z. (2005). *Efficient use of synchrotron radiation for macromolecular diffraction data collection*. *Prog. Biophys. Mol. Biol.* **89**, 153–172.
- Diederichs, K. & Karplus, P. A. (1997). *Improved R-factor for diffraction data analysis in macromolecular crystallography*. *Nat. Struct. Biol.* **4**, 269–275.
- Diederichs, K., McSweeney, S. & Ravelli, R. B. G. (2003). *Zero-dose extrapolation as part of macromolecular synchrotron data reduction*. *Acta Cryst.* **D59**, 903–909.
- Evans, G. & Walsh, M. (2005). Editors. *Data Collection and Analysis. Proceedings of the CCP4 Study Weekend*. *Acta Cryst.* **D62**, 1–124.
- Garman, E. F. & Owen, R. L. (2006). *Cryocooling and radiation damage in macromolecular crystallography*. *Acta Cryst.* **D62**, 32–47.
- Garman, E. F. & Schneider, T. R. (1997). *Macromolecular cryocrystallography*. *J. Appl. Cryst.* **30**, 211–237.

9. X-RAY DATA COLLECTION

- Gruner, S. M. & Ealick, S. E. (1995). *Charge coupled device X-ray detectors for macromolecular crystallography*. *Structure*, **3**, 13–15.
- Hamlin, R. (1985). *Multiwire area X-ray diffractometers*. *Methods Enzymol.* **114**, 416–452.
- Helliwell, J. R. (2004). *Macromolecular Crystallography with Synchrotron Radiation*. Cambridge University Press.
- Henderson, R. (1990). *Cryo protection of protein crystals against radiation damage in electron and X-ray diffraction*. *Proc. R. Soc. London Ser. B*, **241**, 6–8.
- Hendrickson, W. A. (1999). *Maturation of MAD phasing for the determination of macromolecular structures*. *J. Synchrotron Rad.* **6**, 845–851.
- Hendrickson, W. A. (2000). *Synchrotron crystallography*. *Trends Biochem. Sci.* **25**, 637–643.
- Howell, P. L. & Smith, G. D. (1992). *Identification of heavy-atom derivatives by normal probability methods*. *J. Appl. Cryst.* **25**, 81–86.
- Hülsen, G., Broennimann, C., Eikenberry, E. F. & Wagner, A. (2006). *Protein crystallography with a novel large-area pixel detector*. *J. Appl. Cryst.* **39**, 550–557.
- International Tables for Crystallography* (2005). Vol. A. *Space-Group Symmetry*, edited by Th. Hahn. Heidelberg: Springer.
- Kabsch, W. (1988). *Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector*. *J. Appl. Cryst.* **21**, 916–924.
- Leslie, A. G. W., Powell, H. R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. & Nave, C. (2002). *Automation of the collection and processing of X-ray diffraction data – a generic approach*. *Acta Cryst.* **D58**, 1924–1928.
- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines*. *J. Synchrotron Rad.* **9**, 401–406.
- Matthews, B. W. (1968). *Solvent content in protein crystals*. *J. Mol. Biol.* **33**, 491–497.
- Popov, A. N. & Bourenkov, G. P. (2003). *Choice of data-collection parameters based on statistic modelling*. *Acta Cryst.* **D59**, 1145–1153.
- Read, R. J. (1999). *Detecting outliers in non-redundant diffraction data*. *Acta Cryst.* **D55**, 1759–1764.
- Rodgers, D. W. (1997). *Practical cryocrystallography*. *Methods Enzymol.* **276**, 183–203.
- Rosenbaum, G., Holmes, K. C. & Witz, J. (1971). *Synchrotron radiation as a source for X-ray diffraction*. *Nature (London)*, **230**, 434–437.
- Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Phasing in the presence of severe site-specific radiation damage through dose-dependent modelling of heavy atoms*. *Acta Cryst.* **D60**, 1024–1031.
- Sheldrick, G. M. (2008). *A short history of SHELX*. *Acta Cryst.* **A64**, 112–122.
- Smith, J. L. (1991). *Determination of three-dimensional structure by multiwavelength anomalous diffraction*. *Curr. Opin. Struct. Biol.* **1**, 1002–1011.
- Weiss, M. S. (2001). *Global indicators of X-ray data quality*. *J. Appl. Cryst.* **34**, 130–135.