

Data-collection strategies

Zbigniew Dauter

National Cancer Institute, Frederick and Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA

Correspondence e-mail: dauter@bnl.gov

The optimal strategy for collecting X-ray diffraction data from macromolecular crystals is discussed. Two kinds of factors influencing the completeness of data are considered. The first are geometric, arising from the symmetry of the reciprocal lattice and from the experimental setup; they affect quantitatively the completeness of the measured set of reflections. The second concern the quality, or information content, of the recorded intensities of these measured reflections.

Received 28 January 1999

Accepted 22 June 1999

1. Introduction

Owing to technological advances in both hardware and software in recent years, collection of diffraction data from macromolecular crystals becomes ever easier and faster. Parallel advances have occurred in the subsequent steps of the analysis, such as phasing, refinement and model building, where more powerful programs accelerate the process and make difficult cases more tractable. In the light of such general progress, the importance of the quality of the intensities should be emphasized. As all subsequent, mainly computational, steps of structure analysis become faster and easier, the primary data quality plays a more important role, since data collection is the last experimental stage and is often difficult to repeat. High data quality makes all subsequent steps easier and leads to more precise atomic models.

X-ray data collection is the last experimental step of the analysis, but it is not a mere technicality and should be treated as an important scientific process. The factors involved are complex. Some can be treated in an automatic manner by the controlling software. Others require decisions to be made by the experimenter. The present detector-controlling and data-processing programs often allow the use of some built-in default parameters. Most of them can be recommended; however, some parameters involved in the setting up of the data collection should be considered for each individual application. Crystals of macromolecules differ widely in their characteristics, as do the various detectors. Appropriate values of the parameters involved in setting up the experiment should be selected to ensure the best possible data quality. The choice of strategy for collecting data using the rotation method has been discussed previously (Arndt, 1968; Klinger & Kretsinger, 1989; Arndt & Wonacott, 1977; Vickovic *et al.*, 1994; Leslie, 1996; Ravelli *et al.*, 1997; Dauter, 1997 and numerous data-collection workshops). The presentation of the basic concepts relevant for this subject can be found in several compendia and textbooks (*e.g.* Giacovazzo, 1992; Helliwell, 1992; *International Tables for X-ray Crystallography*, 1992).

The most important factor in assessing the X-ray data is the completeness. X-ray data consist of a set of indices and their

associated intensities, with their standard uncertainties. Both should be complete; indices in terms of numbers and intensities in terms of the information content. The quantitative completeness of indices is mainly dependent on factors governed by the geometry of the crystal lattice and of the detector setup.

The qualitative completeness of the measured intensities depends in a somewhat more complicated way on other factors, such as exposure time, crystal diffracting power and characteristics of the detector and X-ray source. The intensities should be complete, but obviously not all of them are strong. However, the weak intensities also contain information. Some direct-methods applications make use of the fact that certain reflections have very weak intensities. Neglecting the weak reflections in refinement introduces bias and removes part of the information. It is not good practice to reject all reflections weaker than, for example, 1σ at the data-processing stage, particularly if the estimation of errors may also be dubious. The uncertainties (σ s) of the measured intensities have often been treated lightly in macromolecular crystallography. In part, this reflects difficulties involved in their proper estimation. However, contemporary detectors and processing programs allow the user to obtain proper statistically estimated uncertainties of all measured intensities. Many sophisticated algorithms, notably phasing and refinement programs based on the maximum-likelihood principle, depend on properly estimated standard deviations of the structure amplitudes.

Factors influencing diffraction data collected using two-dimensional detectors will be discussed, firstly in the quantitative or geometrical context and secondly in the qualitative context.

2. Quantitative completeness of indices

2.1. The rotation method

All geometrical considerations of diffraction can be rationalized using the concept of the Ewald sphere, which illustrates Bragg's law of diffraction in three dimensions. For geometric considerations, the diffraction from a crystal can be treated as a reflection of X-rays from planes in the crystal. In reality, this process is based on the interference of X-rays scattered from atoms (or rather their electrons) positioned in the crystal in an ordered fashion.

The Ewald construction is shown in Fig. 1. The radiation of wavelength λ is represented by a sphere of radius $1/\lambda$ centered on the X-ray beam. The crystal is represented by the reciprocal lattice, with its origin at the point on the Ewald sphere where the direct beam leaves it. Each reciprocal-lattice point lies at the end of a vector perpendicular to the corresponding family of crystal planes and with a length inversely proportional to the interplanar spacing d . If the reciprocal-lattice point lies on the surface of the Ewald sphere, 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 at 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 sphere centre with the reciprocal-lattice point at the surface.

For any particular crystal orientation, only a few reflections can be in the diffracting position, but most of them will not lie on the surface of the Ewald sphere. The number depends on the density of the reciprocal lattice and hence on the unit-cell dimensions. A small-molecule crystal with short unit-cell dimensions and a sparsely populated reciprocal lattice may not give rise to any diffraction in some orientations. Crystals of macromolecules have unit-cell dimensions much larger than the wavelength of the radiation used, and several reciprocal-lattice points (reflections) will lie on the surface of the Ewald sphere in any crystal orientation.

In general, to observe the diffraction from a number of reflections, the reciprocal-lattice points have to be moved to the surface of the Ewald sphere or the sphere radius has to be changed so that different reflections will lie on its surface. The first approach, using a constant Ewald sphere and therefore a selected wavelength (monochromatic radiation), requires that the crystal be rotated to bring successive reflections into diffraction (Fig. 2). If the crystal is only rotated about a single axis, this is called the rotation method; this is the most common procedure used for recording diffraction data in macromolecular crystallography and is discussed below. The other approach, with a stationary crystal and radiation of continuous-wavelength spectrum (white radiation), is called the Laue method. It is only used in special applications when diffraction data have to be collected rapidly. This technique will not be discussed here; its requirements are quite different from those of the commonly used rotation method.

The reciprocal lattice consists of points arranged in planes. Owing to the large unit-cell dimensions of macromolecular crystals compared with the X-ray wavelengths, these planes

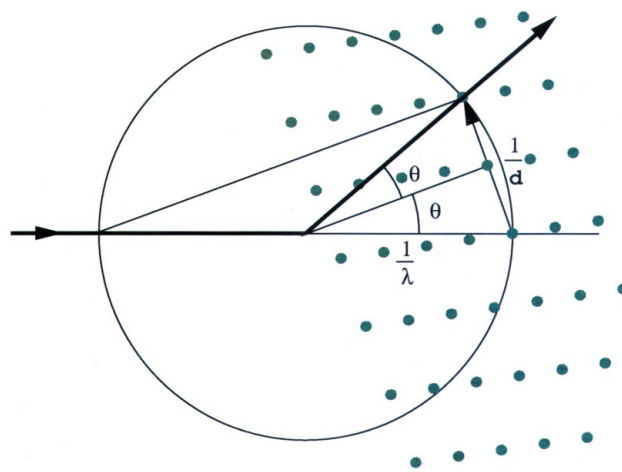


Figure 1
The Ewald construction. When the reciprocal-lattice point crosses the surface of the sphere, the trigonometric condition $1/d = (2/\lambda)\sin\theta$ is fulfilled. This is the three-dimensional illustration of Bragg's law $\lambda = 2d\sin\theta$.

are densely populated in relation to the size of the Ewald sphere. If the plane of reflections in the reciprocal lattice is perpendicular to one of the real-space crystal axes, all reflections have one common index. It is instructive to inspect how the arrangement of reflections into reciprocal-lattice planes defines the diffraction pattern on a two-dimensional detector.

The plane intersects with the sphere, giving a small circle which projects onto the detector as an ellipse, since all rays diffracted by reflections from the same plane form a cone. When the crystal is not rotated during the X-ray exposure or rotates only very little, as in so called 'still' or 'pseudo-still' photographs (Fig. 3), the diffraction pattern will consist of spots arranged in a set of concentric ellipses originating from one family of parallel planes in the reciprocal lattice. However, if the crystal is rotated, the start and end orientations of the plane form two intersecting ellipses with all reflections recorded between them in the form of a lune, as in Fig. 4. All reflections within the same lune originate from the same reciprocal-lattice layer and represent direct lattice planes parallel to one common axis. Because reflections are arranged in families of parallel planes, there will be a family of

concentric lunes on the detector. The lunes will be more pronounced if the reciprocal-lattice planes are oriented nearly perpendicular to the X-ray beam or parallel to the detector plane. Crystals with large unit-cell dimensions will produce more pronounced lunes, while in the diffraction patterns of small structures the lunes are not distinguishable.

The width of each lune varies around its circumference. They are widest in the direction perpendicular to the rotation axis, when the width is proportional to the rotation range per exposure. In the direction along the rotation axis, the width is very small, since the intersection of the plane with the Ewald sphere does not vary significantly. This is illustrated in Fig. 4.

Within each lune, diffraction spots are arranged along lines, reflecting the regularity of the reciprocal lattice. Their pattern

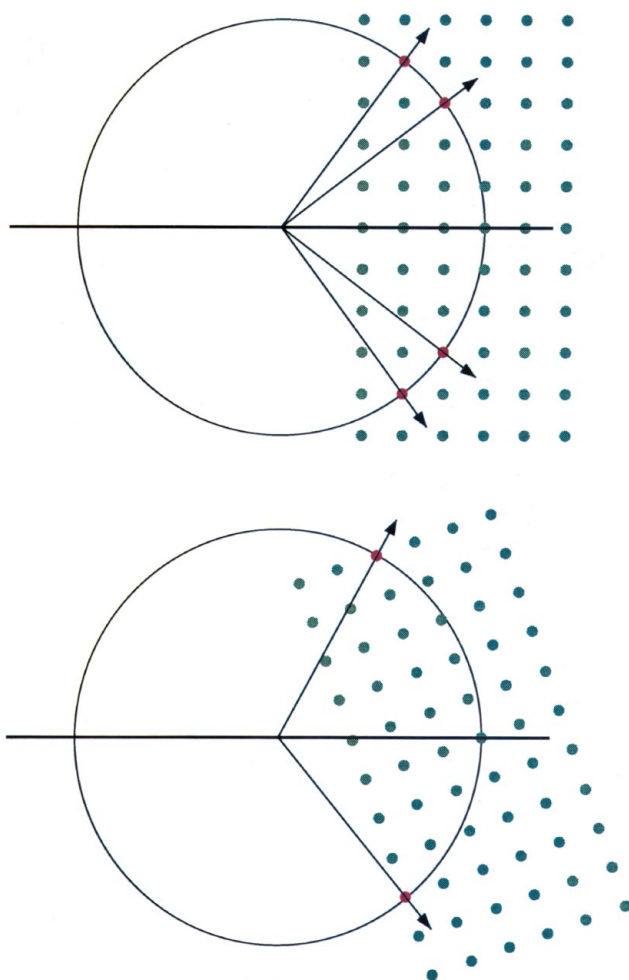


Figure 2

To bring more reflection into diffraction, the crystal represented by the reciprocal lattice has to rotate.

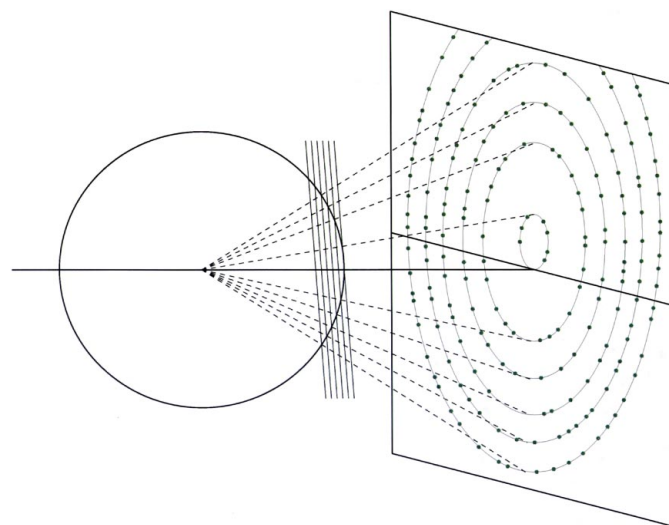


Figure 3

A still exposure with a stationary crystal contains only a small number of reflections arranged in a set of narrow ellipses.

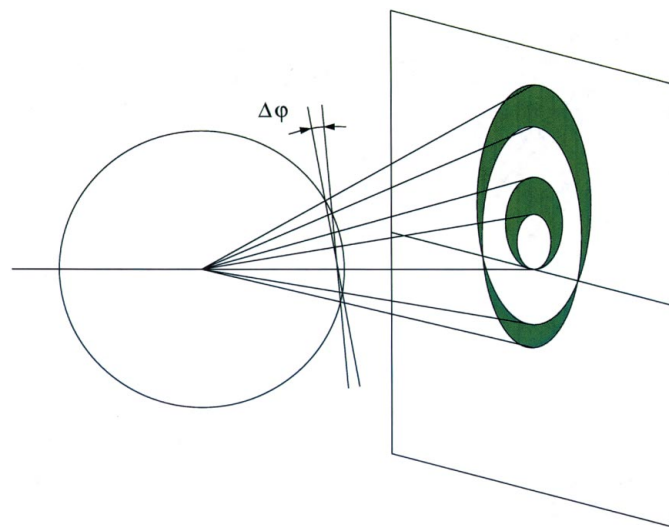


Figure 4

When the crystal is rotated, reflections from the same plane in the reciprocal lattice form a lune, limited by two ellipses corresponding to the start and end positions.

is distorted to a different extent as a consequence of the mapping of the curved Ewald sphere on the flat (or sometimes cylindrical) detector surface. The straight lines of reflections become hyperbolas. The degree of distortion depends on the diffraction angle, *i.e.* resolution. At low angles, the surface of the Ewald sphere can be approximated by a plane, and at low resolution the lunes look like precession photographs and are easy to interpret and index even by eye.

2.2. Crystal mosaicity and beam divergence

The Ewald construction represents the radiation as a sphere of radius $1/\lambda$ attached rigidly to the beam, and the crystal in a particular orientation as a reciprocal lattice consisting of mathematical dimensionless points. In practice, the incident radiation is not directed precisely along one line and all parts of the crystal are not in the same unique orientation. The X-ray beam can be focused and collimated to be parallel within a small angle, about 0.2 or 0.4° on a rotating-anode source with or without mirror optics, respectively, and to somewhat smaller values on synchrotron beamlines, where the horizontal and vertical beam divergence may differ. Crystals are composed of small mosaic blocks slightly misoriented with respect to one another, which adds some divergence to the total rocking curve, that is to the amount of rotation during which an individual reflection diffracts. This is schematically illustrated in Fig. 5. In addition, the X-radiation is monochromated to a defined narrow wavelength window and has a bandpass $\delta\lambda/\lambda$ of the order 0.0002 – 0.001 at synchrotron beamlines; this is considerably wider on laboratory sources. The Ewald sphere has two limiting orientations which results in a defined active width, and reciprocal-lattice points can be represented by disks extended angularly; mosaicity does not extend them radially since the diffraction angle θ remains constant (Fig. 6). The wavelength bandpass effectively broadens the Ewald sphere. These effects cause the diffraction by a particular reflection to be spread over a range of crystal rotation and therefore a period of time.

2.3. Partially and fully recorded reflections

The finite value of the rocking curve (the total effect of beam divergence and crystal mosaicity) has consequences for the diffraction pattern. In the rotation method, images are

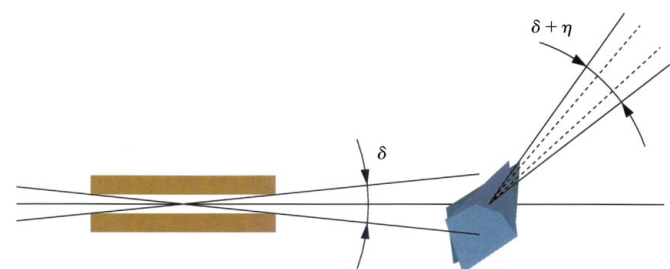


Figure 5
Schematic illustration of how beam divergence and crystal mosaicity combine to give the total rocking curve of the diffracted rays.

exposed in a continuous series of narrow crystal rotations and each reflection diffracts during a defined interval of crystal rotation. Some reflections come into the diffracting position during one exposure and finish during the next. Consequently, part of their intensity will be recorded on one image and another part on the next. If the rotation range per image is small compared with the rocking curve, individual reflections can be spread over several images. Such reflections are termed partially as opposed to fully recorded, the latter having all their intensity present on a single image. Inspection of Fig. 7, which schematically represents a lune on two consecutive images, illustrates how partials are present at the edges of every lune. The lower edge of each lune contains the remaining intensity of those partials which started diffracting on the previous image, and the upper edge contains those partials which will have the rest of their intensity on the next exposure. Thus, comparison of two successive exposures shows that some spots are common to both images. Fig. 8 illustrates the effect of mosaicity on the diffraction pattern. If the mosaicity increases, the lunes become wider because there are more partial reflections. When the mosaicity reaches the value of the rotation range, there are no fully recorded reflections at all.

It is easy to judge by visual inspection if the mosaicity is low or high. If it is low, all lunes have sharply defined edges. If it is high, the intensities of reflections fade out gradually and there are no clearly visible borders. 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 crystal mosaicity.

The effect of mosaicity should not be confused with the effect of crystal splitting. This effect, sometimes incorrectly termed ‘twinning’, produces diffraction patterns with over-

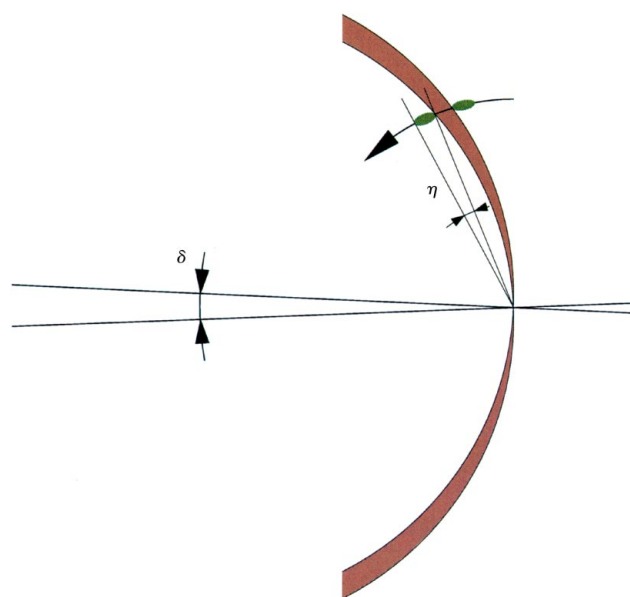


Figure 6
Representation of beam divergence and crystal mosaicity in reciprocal space, which cause the diffraction by a particular reflection to take place in a finite time and therefore during a defined crystal rotation.

lapped multiple lattices. Depending on the degree of splitting, separate regular lattices can be identified or reflection profiles are elongated or smeared out. The effect on the diffraction image differs depending on the crystal orientation. In the simplest case, the crystal consists of two parts slightly rotated with respect to one another around a particular axis. When such a rotation axis lies along the X-ray beam, the reflection profiles will be elongated or doubled in the plane of the detector, and when it lies parallel to the detector plane, the lunes will be broadened. The latter effect is similar to that of high mosaicity, but the reflection profiles will not be significantly affected. It is therefore good practice to judge the crystal quality from inspection of two initial test exposures separated by 90° rotation.

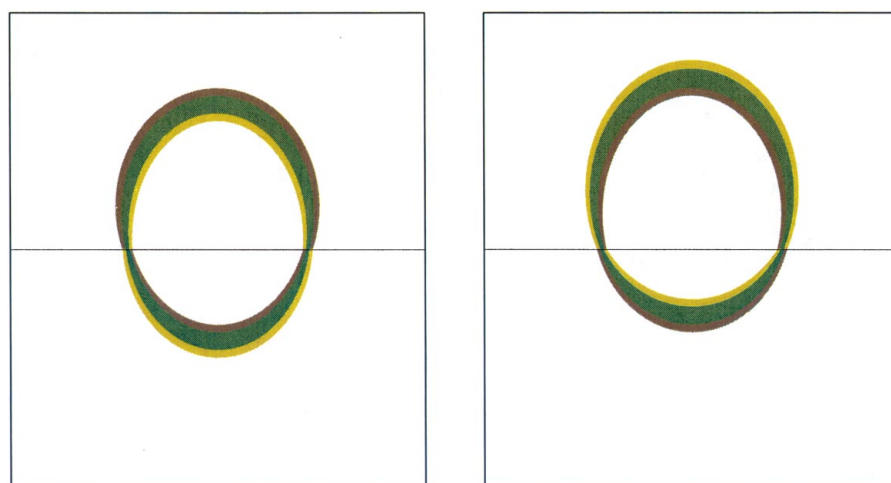


Figure 7

Schematic representation of a lune on two consecutive exposures. The first image, on the left, contains the remaining intensity of partially recorded reflections from the previous image (yellow), fully recorded reflections (green) and a fraction of the intensity of reflections which still diffract at the end of exposure (brown). On the next exposure, on the right, the remaining intensity of the latter reflections is present (brown) as well as further fully and partially recorded reflections.

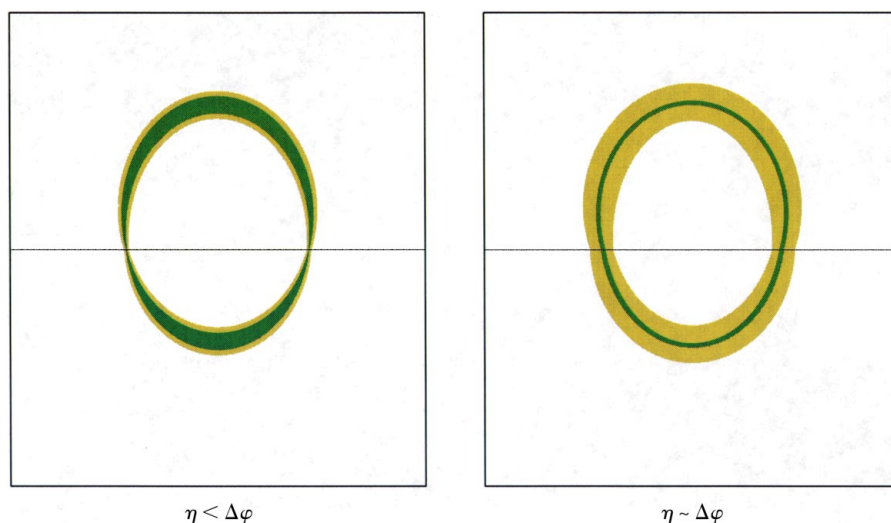


Figure 8

The difference between analogous lunes for low (left) and high mosaicity (right). With increased mosaicity the width of the lune widens, most characteristically along the rotation axis.

2.4. Wide and fine slicing

In the context of the angular width of an individual reflection, two approaches within the rotation method can be considered. So-called 'wide slicing' is based on collecting images wider than the rocking curve, usually of the order of 0.5° or more. The images contain mainly fully recorded reflections, with some partials. In the 'fine-slicing' method, images are much narrower than the reflection width, 0.1° or less, so that each reflection is spread over several images. The two methods require a different approach to the integration of intensities. In the wide-slicing method, each reflection has a two-dimensional profile. In the fine-slicing approach, three-dimensional profiles can be constructed, with the φ axis of rotation as a third dimension.

The disadvantage of wide slicing relates to the fact that the rotation range is greater than the rocking curve. As a consequence, each reflection profile is overlapped on the background which accumulates during the whole image exposure, even when reflections do not diffract. In this context, there is no advantage in cutting the rotation range further than the crystal rocking width. However, finer slicing allows the construction of the three-dimensional profiles, which may provide more accurate intensity integration.

The main factor for or against wide or fine slicing is the read-out time of the detector. If this is negligible in comparison with the exposure time (as for ionization chambers or some CCDs), then fine slicing can be recommended. If the detector dead-time is relatively large (as for imaging plates), wide slicing is usually the method of choice.

2.5. Rotation range

In the fine-slicing approach, there are no practical limitations resulting from the geometry of the rotation method. For wide slicing, a few factors must be taken into account for selection of the rotation range per single exposure. In principle, it should be small enough to avoid overlap of neighbouring lunes, Figs. 9(a)–9(c). A simple formula can be derived (Fig. 10) and used to estimate the maximum permitted rotation range:

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

where the factor $180/\pi$ converts radians to degrees, η is the angular width of the reflection (mosaicity and beam diver-

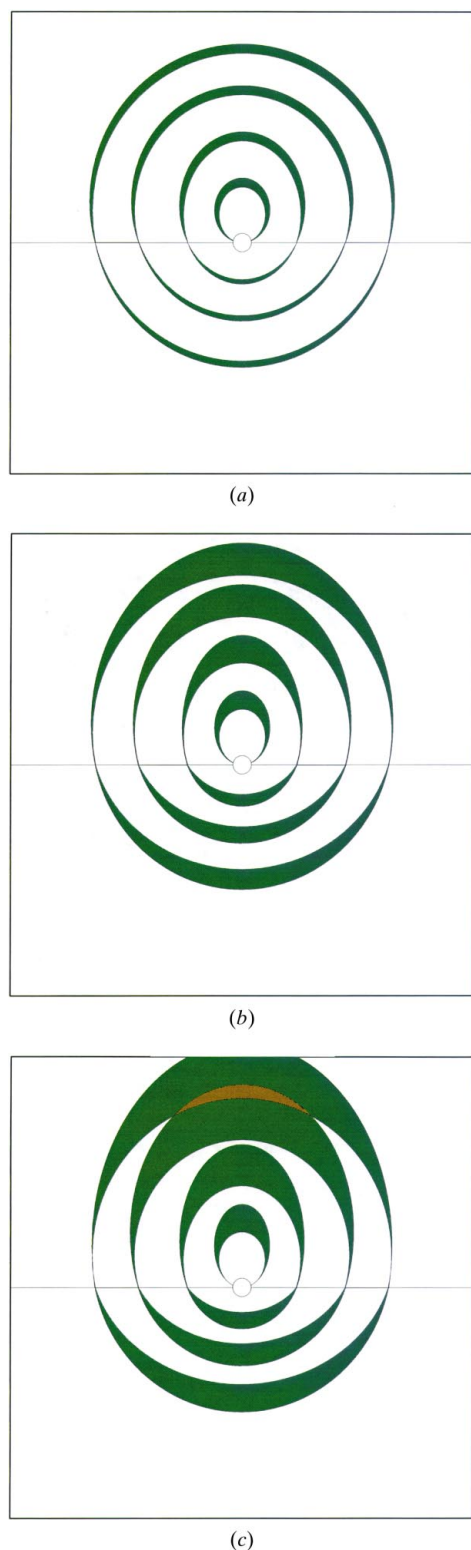


Figure 9

A series of luns resulting from the family of parallel planes in the reciprocal lattice. The gap between consecutive luns depends on the spacing between planes or the unit-cell dimension in the direction perpendicular to the planes. If the rotation range is small (*a*), the luns are narrow and gaps wide. With increased rotation range (*b*), the luns are wider and gaps smaller. If the rotation range is increased further (*c*), the luns start overlapping and the reflection profiles from two consecutive luns may also overlap.

gence), d is the high-resolution limit and a is the length of the primitive unit-cell dimension along the direction of the X-ray beam.

This is not a very strict requirement and applies mainly when reflections are dense within each layer, *i.e.* the unit-cell dimensions are large and the crystal orientation is axial. Otherwise, reflections from successive layers project onto the detector in different positions. If a hexagonal crystal is oriented with its a axis along the beam, then even and odd layers contain lines of reflections which project between each other (Fig. 11). A similar situation occurs when a tetragonal crystal is oriented along its 110 direction. The degree of overlap of individual reflections on the detector will in addition depend on the size of their profiles, which in turn is a result of crystal size and mosaicity, beam divergence and cross-section, detector pixel size and crystal-to-detector distance. It is best to decide on the optimal rotation range after interpreting the first diffraction image or, preferably, two images exposed 90° apart. Most popular integration packages allow the user to rapidly index and interpret individual images and such a procedure is highly recommended. The diffraction pattern can then be generated for different crystal orientations and checked for overlap of reflection profiles, already adjusted in size to real diffraction spots.

It is very difficult to collect data from crystals which have one very large unit-cell dimension if the latter lies along the X-ray beam. Particularly if the crystal is mosaic, it may be impossible to avoid reflection overlap. It is much better if the longest axis is aligned close to the spindle axis of crystal rotation, because it will then never lie parallel to the beam.

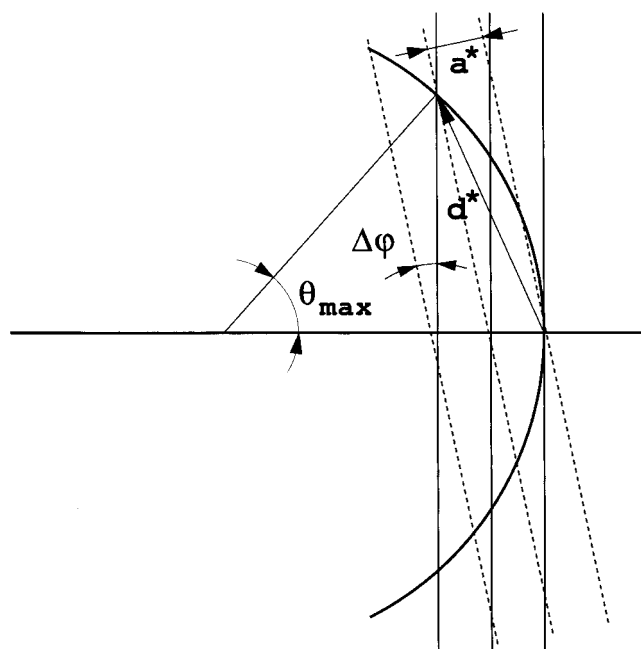


Figure 10

To avoid the overlap of luns at maximum resolution ($d^* = 1/d$), the rotation range ($\Delta\varphi$) cannot be wider than the spacing between planes ($a^* = 1/a$), leading to the condition $\Delta\varphi < 180d/\pi a - \eta$, if the mosaicity η is also taken into account.

Unfortunately, plate-like crystals often have the very long cell edge perpendicular to the flat face. It is difficult to mount such crystals across the spindle axis. A goniostat with κ geometry may be used to reorient the crystal or the mounting loop can be bent (Fig. 12) to accommodate a flat crystal in a skewed orientation.

2.6. Crystal-to-detector distance

The longer the crystal-to-detector distance, the better the signal-to-noise ratio in the recorded diffraction pattern, since the background area increases with the square of the distance, whereas reflection profiles increase less. The distance should therefore be adjusted to match the maximum resolution of the diffraction. It is advisable to inspect two images 90° apart, as some crystals display anisotropy and diffract further in one direction than another. A key and difficult decision is to judge how far meaningful intensities extend, and initial images should be carefully inspected visually with maximum display contrast. It is advisable to apply some safety margin, *i.e.* set the distance a little closer than results from such an inspection.

In some cases, additional factors must be taken into account. If one unit-cell dimension is so large that setting the detector distance to maximum diffraction resolution would lead to significant overlap of reflection profiles, it is better to sacrifice the resolution for full completeness of the data and set the distance so that reflection profiles separate. If the detector setup permits, it can be shifted from the central position using either the 2θ arm or a simple vertical displacement. With such an offset, higher diffraction angles and higher resolution data can be collected. However, a larger total rotation may be necessary to achieve complete data. This only applies when the reflection overlap is a consequence of the long axis being oriented in the plane of the detector. If it is caused by the overlap of lunes (discussed in the previous section), increase of the distance and detector offset will not help.

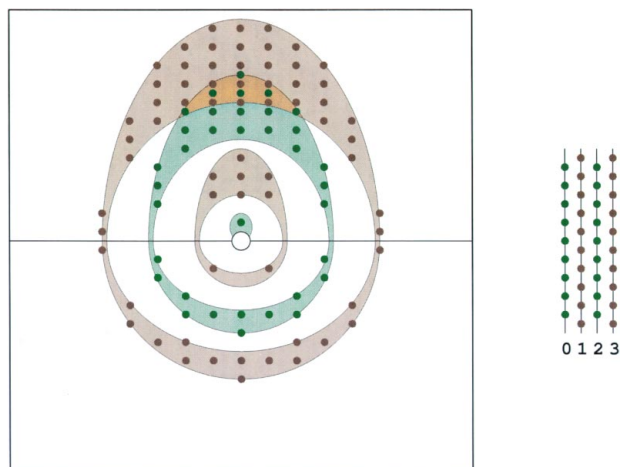


Figure 11

When the cell is centered or if the cell angles differ from 90° , the reflections from neighbouring lunes will not overlap, as in the case of a hexagonal crystal exposed along its a axis.

2.7. Wavelength

The wavelength of X-radiation produced by a rotating-anode source is fixed at the value characteristic for the anode metal, usually copper with $\lambda = 1.542 \text{ \AA}$. In contrast, the user of a synchrotron beamline often has the freedom of choosing the radiation wavelength.

If data are collected with the aim of recording the anomalous diffraction signal, the wavelength must be appropriately optimized. The requirements of multiwavelength anomalous dispersion experiments are particularly strict and are discussed in a separate article. For single-wavelength anomalous dispersion data, it is usually sufficient to adjust the wavelength to be a little shorter than the absorption edge of the anomalous scatterer present in the crystal. If possible, it is also instructive to record a fluorescence spectrum from the crystal or at least from a standard sample containing the desired element or its salt. In the latter case, some safety margin should be adopted, setting the wavelength about $0.001\text{--}0.002 \text{ \AA}$ shorter (or the energy $10\text{--}20 \text{ eV}$ higher) than the observed edge of the standard, allowing for the possible chemical shift of the signal.

For native data collection at a synchrotron, any value of the wavelength can be used, ensuring the high intensity of the beam, which may vary depending on the characteristics of the source and beamline optics. At most synchrotrons, wavelengths below 1 \AA are used, as this minimizes the absorption of radiation by the crystal and its mother liquor and the air scatter. The prolonged lifetime of crystals owing to avoidance of secondary damage is not a significant factor today since cryogenically frozen samples are generally used. Short wavelength is advantageous for collecting very high resolution data, since it decreases the maximum recordable 2θ angle and minimizes the blind region (see below). The advantage of longer wavelength is the stronger interaction with crystals, leading to enhanced intensity of diffracted rays.

2.8. Blind region

In the rotation method, the crystal is rotated around a single axis. Using the reciprocal-space construction, the X-radiation

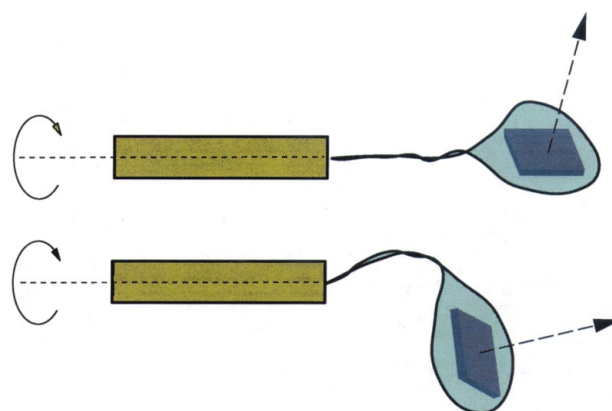


Figure 12

It is advantageous to orient the longest crystal unit-cell dimension along the spindle axis. If the crystal is a thin plate, a bent loop can be used to achieve this.

is represented by the Ewald sphere and the crystal by the lattice of points rotating around an axis tangential to the sphere. Reflections diffract when the corresponding lattice points cross the surface of the sphere. For analysis of the mutual relationship between the radiation and the crystal, disregarding detector and radiation source position, it is convenient to treat the crystal as stationary and the radiation sphere as rotating, which is easier to visualize graphically. Fig. 13 shows that not all the reflections can diffract, since some reciprocal-lattice points lying close to the rotation axis will never cross the Ewald sphere, even after 360° rotation. This part of the reciprocal lattice, on both sides of the spindle axis, is called the 'blind region' or 'cusp'. Following the curvature of the sphere, the width of the blind region varies: at low resolution it is narrow and it broadens at high resolution. Its width depends only on the relationship between the resolution and the wavelength or, in other words, on the value of the diffraction angle θ . The fraction of the reciprocal space within the blind region, equivalent to the fraction of unrecordable reflections at a particular angle B_θ is

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

The total fraction of reflections lost in the blind region up to a certain limit of the θ angle, B_{tot} , is

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

A graph showing the proportion of data contained in the blind region B_{tot} as a function of resolution for selected wavelengths is shown in Fig. 14. At a particular resolution, the blind region is narrower if the wavelength is short, since the surface of the Ewald sphere is flatter (Fig. 15). As mentioned previously, this is an advantage of using short-wavelength radiation.

When the crystal has symmetry axes, it is possible to record reflections which are symmetry-equivalent to those in the blind region if the unique axis itself does not lie in it. Skewing

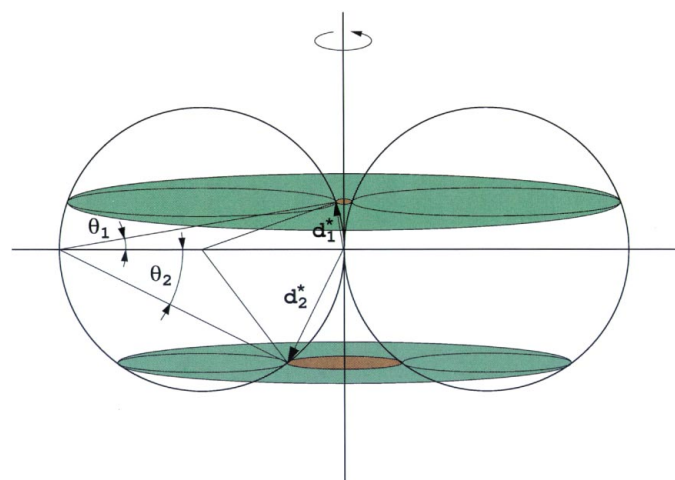


Figure 13

A full 360° rotation of a crystal is here represented as the equivalent rotation of the Ewald sphere with the crystal stationary. Reflections in the blind region, close to the rotation axis, will never cross the surface of the sphere. The blind region is narrow at low resolution and wide at high resolution. Its half-width equals the diffraction angle θ at a given resolution.

the symmetry axis by at least θ_{max} from the spindle direction ensures that there will be no loss of completeness owing to the blind region (Fig. 16). Monoclinic crystals should be skewed away from the ac plane as well as from the b axis.

If the crystal is triclinic, there is no way to avoid loss of completeness arising from the blind region in a single rotation pass. To collect missing reflections, the crystal has to be re-oriented by at least $2\theta_{\text{max}}$ from the previous spindle-axis direction, e.g. using κ -goniostat arcs. The second pass of data collection should cover the missing $2\theta_{\text{max}}$ width of reciprocal space.

In summary, the detrimental effect of the blind region on the completeness of data is significant only if the crystal is aligned along the unique symmetry axis or if it is in space group $P1$. At low resolution it can be neglected altogether!

2.9. Total rotation range

Selection of the total rotation range appropriate for the crystal symmetry is the most important factor influencing the completeness of data. In principle, collecting 180 or even 360° (with anomalous signal in low symmetry) of data will always ensure maximum completeness. As discussed below, it would also result in multiple measurements of equivalent intensities, leading to more accurate data. However, the available beam time can often be limited, especially at synchrotron sites, and minimization of the time of the experiment is a factor to be taken into account in the normal practice of data collection. The analysis of the crystal symmetry in relation to the geometry of the rotation method allows one to specify conditions leading to the minimal complete data set when all unique reflections are measured at least once. Such consid-

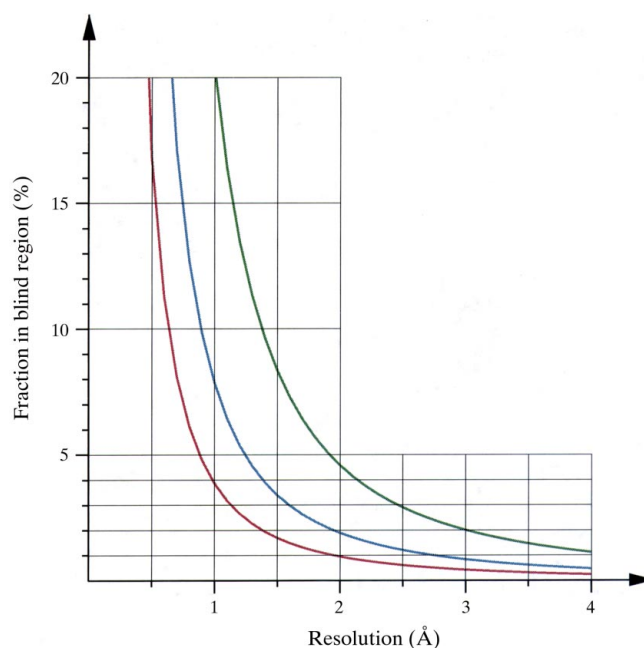


Figure 14

A graph showing the total fraction of reflections located in the blind region for different wavelengths: 1.54 Å (green), 1 Å (blue) and 0.71 Å (red). Only at very high resolution is there the possibility of a significant loss of completeness because of the blind region.

erations can be expected to be less important at third-generation synchrotron sources. On the other hand, some crystals, even if frozen, may not survive the exceedingly intense radiation from third- or fourth-generation synchrotron sources and in this case it would be beneficial to reach high completeness as soon as possible, following the optimal strategy.

The data are complete if the Ewald sphere has been crossed by all reflections (or their symmetry mates) in the asymmetric part of the reciprocal lattice, which always has the shape of a wedge with the apex at the origin and is limited by the maximum-resolution sphere. Its shape and volume is characteristic for the particular Laue symmetry group. Restricting the analysis to macromolecular crystals with the centre of symmetry excluded, it is sufficient to consider the point-group symmetry (*i.e.* crystal class). The presence of screw axes is irrelevant for these considerations; for example, $P4_12_12$, $P4_32_2$

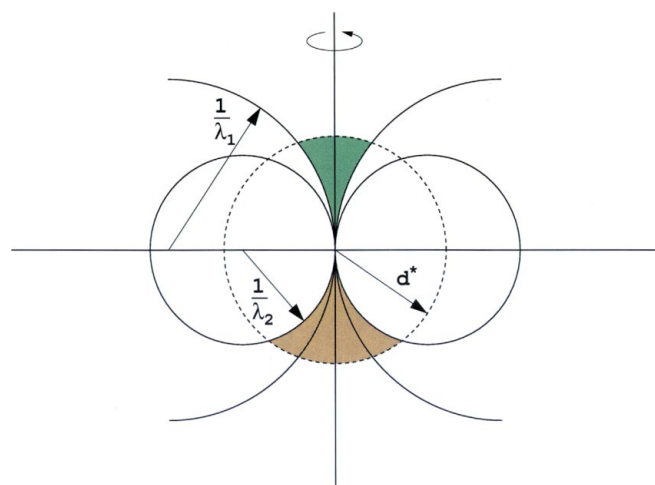


Figure 15

The blind region is narrower with short-wavelength radiation (green) than long-wavelength radiation (brown), since the corresponding Ewald sphere is flatter.

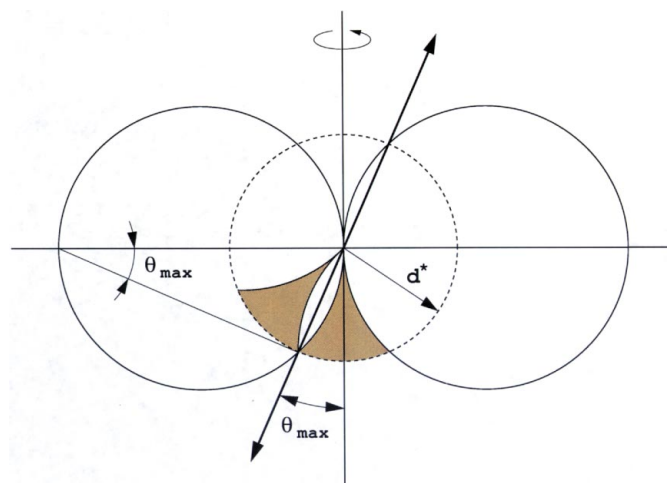


Figure 16

To avoid loss of completeness arising from the blind region, it is sufficient to skew the crystal from the axial orientation by θ_{\max} , the diffraction angle at highest resolution.

and $P4_21_2$ belonging to point group 422 have identical asymmetric units in reciprocal space. In some point groups, the asymmetric unit can be specified in more than one way; for example, in triclinic symmetry any hemisphere constitutes an asymmetric unit.

When a crystal is rotated by 180° , both sides of the Ewald sphere cover 180° of reciprocal space. Fig. 17 illustrates the case when a monoclinic crystal is rotated around the unique axis b . This also applies to a triclinic crystal rotated around any arbitrary axis. After 180° rotation, the lower side of the Ewald sphere covers the region marked in green and the upper side covers the region marked in brown. Reflections in the dark-brown region will be measured twice, but the centrosymmetrically related blue region will not be covered at all. When anomalous differences are not required, it is sufficient to collect 180° of data to achieve full completeness (except for the blind region present in such an orientation). When both Friedel mates must be collected, a wider rotation range of $180^\circ + 2\theta_{\max}$ is necessary. If the crystal symmetry is monoclinic, each Bijvoet mate is then measured twice owing to the symmetric relation of the volumes above and below the plane of the graph. If there is no anomalous signal, each unique reflection is measured four times.

The situation after 135° of rotation is shown in Fig. 18. There is only a small region with reflections measured twice, but there are some reflections not covered at all. Characteristically, the high-resolution data are completed first and the missing region at lowest resolution is only filled when the rotation approaches 180° . This should be taken into account in calculations of predicted completeness using integration software. Such programs usually give the overall value, but

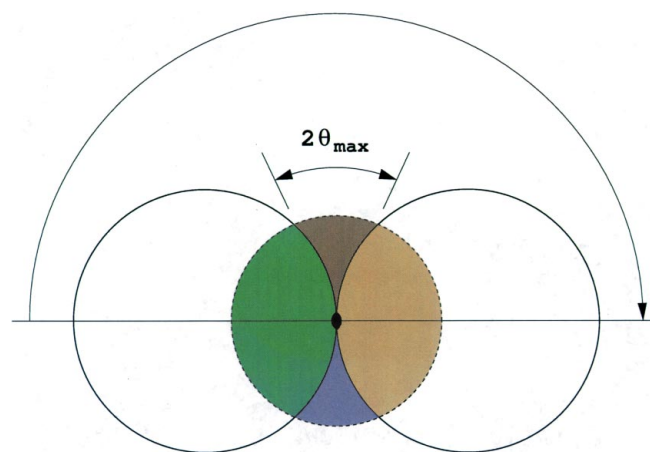


Figure 17

The diffraction sphere (dashed line) corresponding to the highest resolution limit of diffraction and the Ewald sphere at the start and end of a 180° rotation. The lower side of the Ewald sphere covers one part of the reciprocal space (green) and the upper side another part (brown). They overlap over the $2\theta_{\max}$ range (dark brown). If a monoclinic crystal is rotated around its twofold axis, 180° is sufficient to achieve full completeness, even if individual Bijvoet mates have to be recorded separately for anomalous data. If the crystal is triclinic, 180° is sufficient for the native data, owing to the centrosymmetric relation between the non-covered region (blue) and the covered part (dark brown). However, for anomalous triclinic data $180^\circ + 2\theta_{\max}$ have to be covered.

data 95% complete in total may lack 20% of the reflections in low-resolution shells. This effect results from the curvature of the Ewald sphere and is more pronounced at very high resolution. If one collects atomic resolution data in several passes with different exposures and resolution limits, it is not necessary to cover all the theoretically required rotation range in the highest resolution pass, but the lowest resolution pass must be complete.

In general, a given fraction of the rotation range yields a larger fraction of data. For example, after 90° rotation when 180° is required, the completeness may reach about 65%. However, it is possible to obtain higher completeness without increasing the total rotation range covered by splitting the whole range into smaller parts. 45° of data collected twice but separated by a 45° gap, as shown in Fig. 20, will give much higher completeness than a single 90° pass, again as a result of the curvature of the sphere.

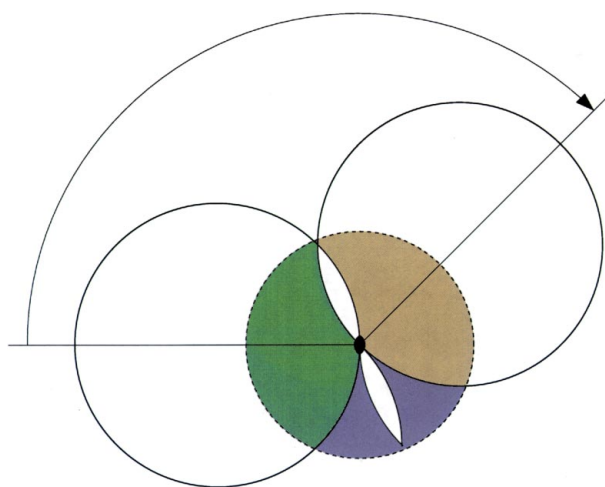


Figure 18

After 135° of rotation when 180° is required, the high-resolution shell may be filled, but the low-resolution region will not be complete.

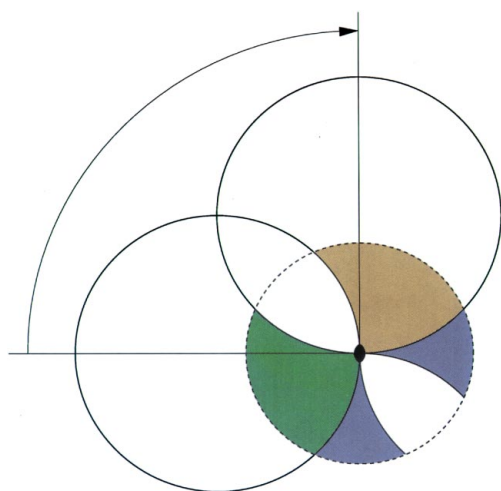


Figure 19

The fractional completeness is higher than the fraction of the required rotation range; 90° rotation out of 180° gives about 65% of unique data.

When an orthorhombic crystal is rotated around any of its twofold axes, the required rotation range is 90° , as illustrated in Fig. 21. In fact, this also applies if the crystal is rotated around any vector lying in one of the axial planes, since the asymmetric unit corresponds to 90° of data between one of the axes and a plane perpendicular to it. It is advantageous to have the crystal oriented around *e.g.* the 110 vector, since in the exactly axial orientation there will be some reflections lost in the blind region. However, the 90° must be between the axis and the plane of symmetry. If the rotation range starts in the diagonal orientation, as in Fig. 22, the same 45° of data will be collected twice, giving a $\sim 65\%$ complete set, similar to Fig. 19. When the orthorhombic crystal is oriented around an arbitrary axis not in the symmetry plane, more than 90° of rotation is required. In such cases, it is advisable to estimate the necessary rotation range and start using the strategy programs available within some data-processing packages.

In general, the required rotation range depends on the crystal orientation. For example, in 622 symmetry the asymmetric unit is a wedge, 30° wide but spanning the space between the sixfold axis along *c* and the *ab* plane. Therefore, if rotated around the *c* axis, the 622 crystal requires only 30° of data, but if rotated around a vector in the *ab* plane 90° are necessary.

In the above, it is assumed that the detector position is symmetrical with respect to the X-ray beam. If it is offset by a 2θ angle, then only one side of the Ewald sphere is relevant and the required rotation ranges will be different, *e.g.* an orthorhombic crystal will require $90^\circ + 2\theta_{\max}$ for completeness. It is then better to rely on software predictions of required rotation.

Table 1 lists the required rotation range for crystals of different classes in various typical orientations. A central position of the detector is assumed. For cubic crystals, it is difficult to give reliable estimations, since they vary dramatically with the crystal orientation.

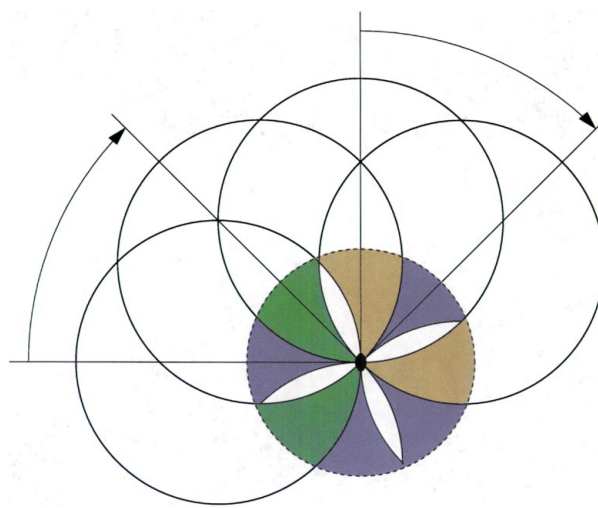


Figure 20

If 90° is split into two 45° ranges separated by 45° gaps, the total completeness is considerably higher than for one continuous range.

Table 1

Rotation range (°) required to collect a complete data set 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 + $2\theta_{\max}$ (any)
2	180 (<i>b</i>); 90 (<i>ac</i>)	180 (<i>b</i>); 180 + $2\theta_{\max}$ (<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 + $2\theta_{\max}$ (<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

Taking into account the importance of selecting the optimal rotation range, it should again be stressed that it is highly advantageous to interpret the first trial images from a newly mounted crystal, establish its orientation and symmetry and then decide where to start and how wide a rotation to cover. Tools available within most popular integration packages allow the quick and reliable making of such estimates.

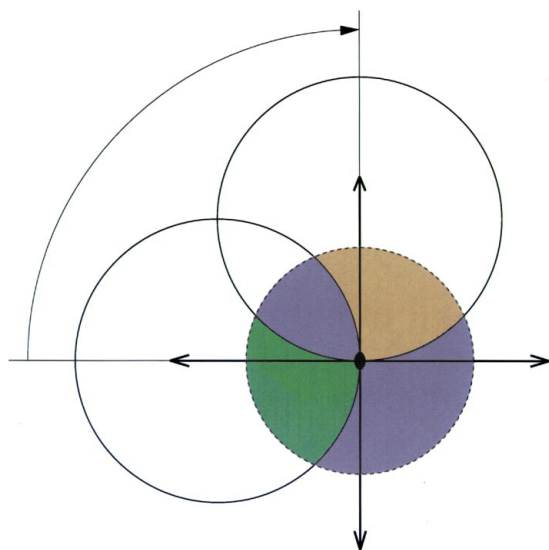
2.10. Equivalent indexing

In certain point groups, reflections can be indexed in multiple ways, all permitted but not equivalent, so that intensities of reflections indexed according to different schemes do not agree. This is possible in point groups which have symmetry lower than the symmetry of their Bravais lattice. In other words, the symmetry of reflection positions is higher than the symmetry of the distribution of their intensities which defines the true symmetry of the crystal. Groups

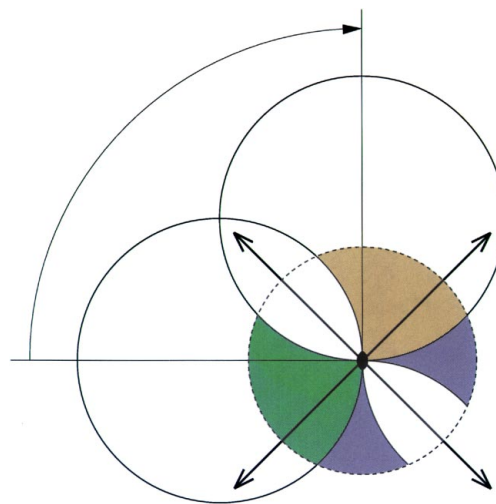
with polar axes, such as 3, 4 or 6, can be indexed with the *c* axis up or down the same direction. The directionality of polar axes is not defined by the lattice if its two other dimensions are equivalent. In monoclinic symmetry, the twofold axis is polar, but its direction is specified by the non-equivalence of the two remaining axes perpendicular to it. Fig. 23 illustrates the case of crystal class 4 with the two possible indexing schemes. Reflections defined by the same indices in both schemes have different intensities.

Another example of the multiple-indexing choice is in cubic symmetry 23 with twofold axes placed along the lattice four-folds. Rotation by 90° leads to alternative, although perfectly permitted, indexing of reflections.

Alternative, non-equivalent indexing schemes are not important if all data are collected from one crystal. However, when they are merged from several crystals or intensities are compared between native and derivative data it has to be taken into account. It does not matter how reflections are indexed during intensity integration, since all possibilities will perfectly match the crystal lattice, but for scaling and merging of intensities all reflections must be indexed in the same way. To re-index all reflections to the alternative scheme, it is necessary to apply the symmetry operation which is included in the (higher) symmetry of the lattice but does not occur in the (lower) crystal point-group symmetry. For example, reflections in symmetry 3 can be indexed in four non-equivalent ways, since there are four ways of locating a threefold axis in the hexagonal lattice of 622 rotational symmetry. The operations required for re-indexing are either sixfold rotation or one of the alternative twofold rotations which are not present in the point group 3. Instead of sixfold rotation, twofold rotation around the *c* axis can be applied, since it is included in 6 and absent in 3. Table 2 lists point groups with alternative indexing possibilities, with the symmetry operations required for re-indexing.

**Figure 21**

An orthorhombic crystal requires 90° rotation between two axial orientations.

**Figure 22**

If an orthorhombic crystal is exposed starting at diagonal orientation, 90° is equivalent to collecting twice the same 45° and is not sufficient for complete data.

When collecting multiple data sets from the same crystal, as in two-exposure passes or in a MAD experiment, it is advisable to adopt a common orientation matrix from the first indexing to all data. If other parameters differ, they can be modified and different rotation starting positions can be easily related. Independent autoindexing for each data pass may lead to confusion resulting from non-equivalence of indexing schemes.

2.11. Interpretation of the example images

The series of images presented in Figs. 24(a)–24(f) were recorded from a crystal of lysozyme, space group $P4_32_12$, unit-cell dimensions $a = b = 78.6$, $c = 37.2$ Å, crystal-to-detector distance 243 mm, wavelength 0.92 Å, resolution 2.7 Å, oscillation range 1.5° and crystal mosaicity $\sim 0.5^\circ$. These images illustrate some of the points discussed above. Each lune in Figs. 24(a)–24(d) consists of reflections arranged as squares reflecting the tetragonal symmetry with $a = b$. The crystal was rotated around the axis diagonal between a and b , which is evident from the way the squares of reflections are arranged.

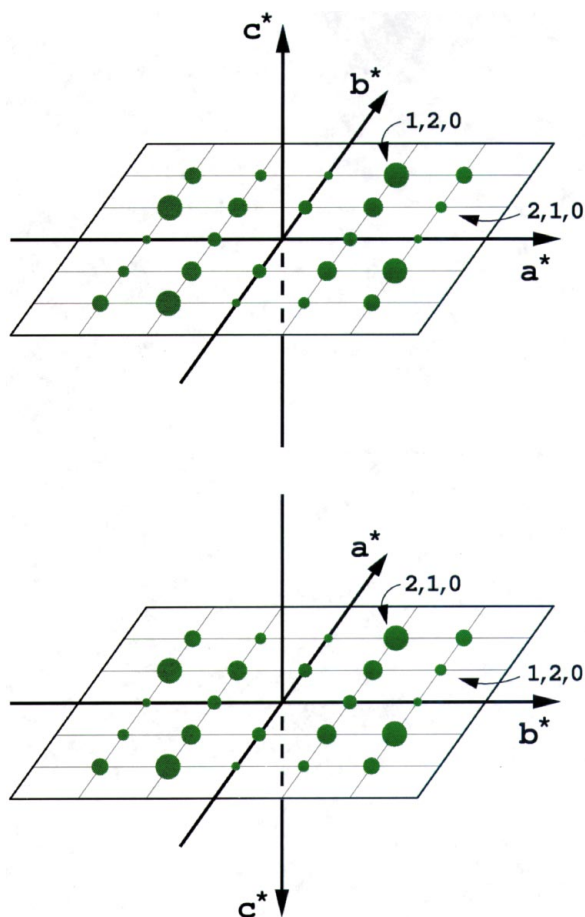


Figure 23

Two ways of indexing the tetragonal lattice in point group 4, with the fourfold axis directed 'up' or 'down'. These two ways are not equivalent, since reflections with the same indices will have different intensities. In this case the symmetry of reflection positions (lattice) is higher than the symmetry of their intensities.

Table 2

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	
$P4$, ($P4_1$, $P4_3$), $P4_2$, $I4$, $I4_1$	$hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
$P3$, ($P3_1$, $P3_2$)	$hkl \rightarrow \bar{h}\bar{k}l$ or $hkl \rightarrow kh\bar{l}$ or $hkl \rightarrow \bar{k}\bar{h}l$	$\bar{1}00/0\bar{1}0/001$ $010/100/00\bar{1}$ $0\bar{1}0/\bar{1}00/00\bar{1}$
$R3$	$hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
$P321$, ($P3_121$, $P3_221$)	$hkl \rightarrow \bar{h}\bar{k}l$	$\bar{1}00/0\bar{1}0/001$
$P312$, ($P3_112$, $P3_212$)	$hkl \rightarrow \bar{h}\bar{k}l$	$\bar{1}00/0\bar{1}0/001$
$P6$, ($P6_1$, $P6_5$), ($P6_2$, $P6_4$), $P6_3$	$hkl \rightarrow kh\bar{l}$	$010/100/00\bar{1}$
$P23$, $P2_13$, ($I23$, $I2_13$), $F23$	$hkl \rightarrow kh\bar{l}$	$010/\bar{1}00/001$

The gaps between lunes are large, a consequence of the relatively short third unit-cell dimension c , which is oriented along the beam perpendicular to the detector plane. This axis was almost perfectly perpendicular to the detector at the point between the images in Figs. 24(a) and 24(b), and the corresponding zero-layer almost vanishes behind the shadow of the beam-stop. In such an orientation, the $\{hk0\}$ plane in the reciprocal lattice is tangential to the Ewald sphere at the origin. The images 90° away (Figs. 24e and 24f) look quite different. There are more lunes with smaller gaps between them, but they are less densely populated by reflections, consistent with the orientation of the reciprocal lattice. Now the lunes are parallel to the hhl family of planes normal to the 110 vector. These planes have finer spacing, but the distances between reflections within each plane are longer. The average density of spots is constant over the whole reciprocal lattice and therefore the number of reflections present in each image is approximately equal and does not depend on the crystal orientation. The hhl lunes even partially overlap at higher resolution close to the detector edge but, owing to the diagonal orientation, reflections on each successive lune fit between those on the previous one.

Closer inspection of the reflections within the zero-layer $hk0$, in particular the 'pseudo-precession' patterns in the images in Figs. 24(c) and 24(d), proves that only every second reflection is observed along the lines of spots passing through the origin containing reflections $h00$ or $0k0$. This reflects the presence of the 2_1 screw axes within space group $P4_32_12$. The presence of the 4_3 screw axis can be confirmed on the images exposed after 90° rotation (Figs. 24e and 24f), when the c axis lies vertically in the plane of the detector and the $00l$ reflections can be seen. Another confirmation of the crystal Laue symmetry group being $P4/mmm$ and not $P4/m$ is the presence of diagonal (in this case vertical) mirror correspondence of reflection intensities on the left and right halves of the lunes. It must be stressed that the positions of the reflections define only the Bravais lattice and it is the symmetry of the intensity pattern which reflects the point-group symmetry and the arrangement of molecules (or, more correctly, structural motifs) in the crystal.

A detailed comparison shows that some reflections at the edges of every lune are present in pairs of images. The crystal mosaicity was about 0.5° , which is one-third of the rotation range. Therefore, about one-sixth of the reflections from each lune is expected to show up on both images. The substantial mosaicity of the crystal can be judged from the fact that the lune edges are not clearly defined, but the intensities gradually fade out. However, it is not necessary to compare two images to realise how the crystal was rotated. It is clear where the lunes are narrow (along the spindle axis, close to the blind region) and where they are wide (direction of rotation, perpendicular to the spindle axis).

3. Qualitative factors

3.1. Reflection profiles

The first and easiest to inspect visually are the reflection profiles, which can be checked on the initially exposed images. They should be regular with a single peak. Their shape should reflect the size and shape of the crystal: if the crystal is needle-like, reflection profiles will be elongated; otherwise, elongation of spot profiles, especially in the direction perpendicular to the detector radius, is a bad sign. When the profiles are irregular, it is vital to expose the crystal in another orientation and compare the profiles, since crystal splitting may not be equally obvious in all orientations. After indexing the diffraction pattern, the integration profiles should be matched with the size and shape of diffraction spots. The spots should not extend into the area defined as background. Selection of too small integration profiles will lead to incorrect integration of intensities; when profiles are too large, the estimation of standard uncertainties will be biased.

3.2. Exposure time

Exposure time is the factor which most strongly influences the reflection intensities. In principle, the higher the intensities, the higher the signal-to noise-ratio and therefore the higher the data quality. This is a simple consequence of counting statistics. Doubling the intensity enhances the signal-to-noise ratio by $2^{1/2}$. In practice, other factors also play an important role.

There are always limitations on the available beam time. In the time available for the experiment, all images necessary to achieve complete data have to be recorded, even at the cost of underexposing them. It is better to have maximally complete data of somewhat lower intensity than only a part of theoretically superior data. This consideration has often been of special importance at synchrotron stations, but may diminish with third-generation sources and fast detectors such as CCDs or pixel devices.

3.3. Overloads

The dynamic range of the detector is another factor to be taken into account. Each detector has a certain count level which saturates its pixels, resulting mainly from the limitations of its electronics. Pixels which accumulate more counts are

overloaded and cannot be used for accurate evaluation of reflection intensity. There is a method of approximate evaluation of intensities affected by overloaded pixels which have a 'top-hat' profile and are only overloaded at the central pixels. This is based on the overlap of a standard reflection profile on the pixels in the shoulder of such a spot, but gives a less reliable measurement. For some applications, especially those based on the Patterson function (like molecular replacement) or direct methods, it may be preferable to accept such estimations, since complete absence of the strongest reflections will seriously bias the results, but for the final refinement of the model they should perhaps be excluded.

The contrast between the intensities of the strongest and weakest reflections is very large. It is therefore inevitable that if the exposure is adjusted to adequately measure the weak high-resolution reflections, some of the strongest ones will be overloaded. They should be measured in a separate rotation pass, with shorter exposures adjusted to adequately cover the strong reflections below the overload limit. In some cases, the required speed of spindle-motor rotation may be exceedingly fast and beyond its limit of reliability. Instead of increasing the rotation speed, it is then better to attenuate the beam intensity, *e.g.* using aluminium foil(s) of appropriate thickness. The second pass does not need to extend to the same resolution limit: it is sufficient to cover the region containing overloads, and the distance can be increased as well as the rotation range per image. The difference in effective exposure should not exceed 10–20 for successful scaling of all data.

3.4. R factor, I/σ and estimated uncertainties

The data quality is usually judged by the global R_{merge} factor, based on intensities or rather F^2 . This gives the average ratio of the spread of intensities of the multiply measured symmetry-equivalent reflections to the estimated value of the reflection intensity

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

This global value is not a proper statistical quantifier and is calculated in different ways in different programs. The value of R_{merge} is highly influenced by data multiplicity. As a consequence, it is always higher for data in high-symmetry space groups than those in low symmetry. Higher multiplicity always leads to improved data quality, although it increases the R_{merge} factor.

Merging of equivalent intensities provides an opportunity to identify and reject outliers, *i.e.* intensities wrongly measured and not agreeing with their equivalents. A small number of outliers may result from erroneous classification of partially and fully recorded reflections, particularly those lying close to the blind region, from individual CCD detector pixels affected by 'zingers', *i.e.* sparks from trace radioactivity of the taper glass, from shadowed or inactive regions of the detector window *etc.* However, the number of outliers rejected from the merging process should be small, at most 1%. It is possible to 'improve' the R_{merge} by rejecting a large number of measurements until the multiplicity is low. This is an extremely bad

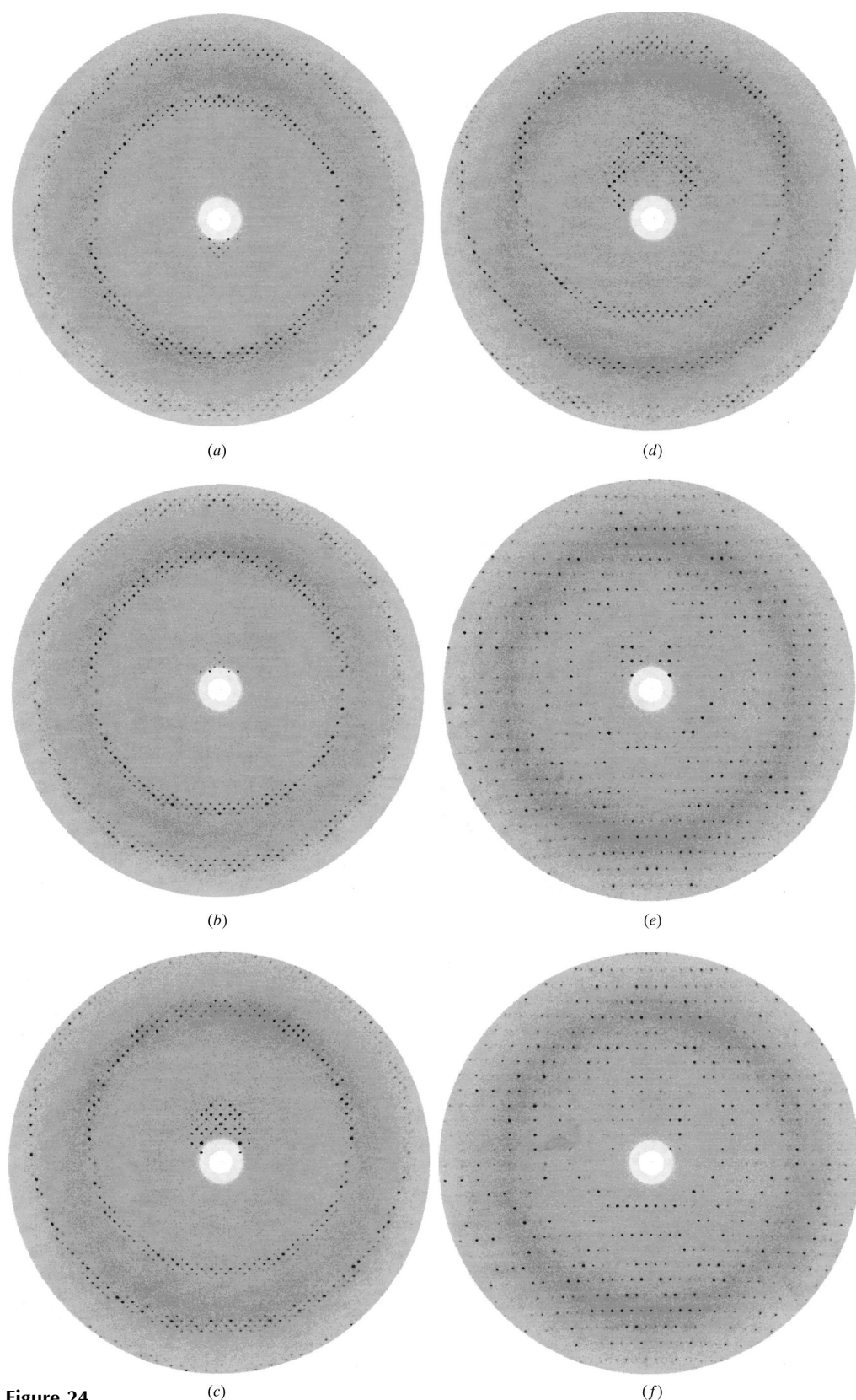


Figure 24

A series of images from a crystal of lysozyme with unit-cell parameters $a = 78.6$, $c = 37.2$ Å in space group $P4_32_12$ at 2.7 Å resolution with 1.5° rotation per exposure. (a)–(d) Four consecutive images exposed with the crystal oriented along its fourfold axis. Large gaps between lunes result from a short c axis. (e, f) Two images exposed 90° away, along the 110 direction. The lunes are much wider, overlapping at high resolution. A close inspection of the zero-layers reveals the presence of systematic absences resulting from the 2_1 axes (c and d) and the 4_3 axis (e and f).

practice which adversely influences the overall quality of data. There should always be a physical reason for rejecting outliers, other than just bad agreement with symmetry-equivalent intensities.

Similarly, it is not advisable to reject reflections weaker than a certain limit, say 1σ , at the stage of data merging. As pointed out before, weak intensities also carry information and their neglect introduces bias into intensity distribution affecting, for example, the refined overall temperature factor.

Complementary information about the data quality to R_{merge} is given by the ratio of intensities to their uncertainties, $\sum I_{hkl} / \sum \sigma(I_{hkl})$, provided that σ s are estimated properly. This is not trivial, since most detectors (such as imaging plates or CCDs) do not count individual X-ray quanta directly and counting statistics may be biased. The detector-gain factor specifying the detector response to a single quantum of radiation should be taken into account in the evaluation of intensity uncertainties during integration. Usually, data-processing packages provide means for checking and correcting the level of intensity uncertainties based on the χ^2 test or on the t plot, which requires that the ratio of the spread of estimated intensities to the associated uncertainties:

$t = (\langle I \rangle - I_i) / \sigma(I)$ should have a normal distribution with an average of 0.0 and standard deviation of 1.0. Correct estimation of intensity standard uncertainties is important in all successive applications based on statistical or probabilistic treatments, such as maximum-

likelihood phasing, refinement and direct methods of solving heavy- or anomalous-atom positions.

In principle, data contain some information if I/σ is higher than 1.0. However, there will be few meaningful intensities among a majority of unreliable estimations if the ratio is close to 1.0. For certain applications, it may be advisable to accept weak data. For example, direct methods use only the subset of most meaningful reflections but extending to as high a resolution as possible. In the standard applications, the accepted resolution limit is where the I/σ falls below about 2.0. R_{merge} may then reach 20–40%, depending on the symmetry and redundancy.

4. Final remarks

Optimal strategies for data collection are influenced by several factors. Some are general and others depend on the characteristics of a particular crystal or detector. The selection of data-acquisition parameters is not trivial and is often the result of a compromise between several requirements. It is difficult to obtain very high multiplicity of measurements in a time-limited experiment. It is essential to know the relative importance of particular parameters in the whole process and make appropriate decisions. As synchrotron beamlines become brighter, detectors become faster and data-processing

software becomes more sophisticated, the whole process of data collection becomes easier from the technical point of view, yet the crucial scientific decisions still have to be made by the experimenter. It must be stressed that it is always beneficial to sacrifice some time and interpret the initial diffraction images thoroughly in order to avoid mistakes which may have an adverse effect on data quality and the whole of the subsequent structural analysis.

References

- Arndt, U. W. (1968). *Acta Cryst.* **B24**, 1355–1357.
 Arndt, U. W. & Wonacott, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North Holland.
 Dauter, Z. (1997). *Methods Enzymol.* **276**, 326–344.
 Giacovazzo, C. (1992). Editor. *Fundamentals of Crystallography*. IUCr/Oxford Science Publications.
 Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation*. Cambridge University Press.
International Tables for X-ray Crystallography (1992). Vol. C, edited by A. J. C. Wilson. Dordrecht: Kluwer Academic Publishers.
 Klinger, A. L. & Kretsinger, R. H. (1989). *J. Appl. Cryst.* **22**, 287–293.
 Leslie, A. G. W. (1996). *CCP4 Newsl. Protein Crystallogr.* **32**, 7–8.
 Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M. & Kroon, J. (1997). *J. Appl. Cryst.* **30**, 551–554.
 Vickovic, I., Kalk, K. H., Drenth, J. & Dijkstra, B. W. (1994). *J. Appl. Cryst.* **27**, 791–793.