Background theory of data collection

William Clegg

School of Chemistry, Newcastle University, United Kingdom

In order to determine a crystal structure, we essentially need to perform a Fourier transformation of an X-ray diffraction pattern. The required experimental data are the directions and intensities of the diffracted X-ray beams (reflections). The reflection amplitudes needed for the Fourier transform are readily calculated from the measured intensities, while the corresponding reflection phases are inaccessible experimentally, but are derived from the intensities by a variety of computational techniques. The purpose of an X-ray single-crystal diffraction experiment is thus to measure, as precisely and accurately as possible, the geometry (directions, recorded as positions on a detector correlated with crystal orientation) and relative intensities of usually thousands of discrete Bragg reflections in a diffraction pattern from a small single crystal rotated about one or more axes while immersed in an X-ray beam. The measured data should be free of systematic errors that may affect the directions and/or intensities, or should be corrected as far as possible for their effects. The data set should also be complete (minimising missing parts of the diffraction pattern), extend to a sufficiently high diffraction angle (Bragg angle) to offer good resolution of features in the electron density of the crystal structure, and preferably include symmetry-equivalent and other 'redundant' measurements to provide evidence of data quality and self-consistency.

A full set of diffraction data normally consists of a long list of individual reflections, each identified by its three indices *hkl* (these, together with the unit cell parameters, define the reflection geometrically), and a value for the intensity, together with an estimate of its precision or reliability (a standard uncertainty, or estimated standard deviation). Obtaining this information involves the application of basic diffraction theory, some of which has already been developed in previous sessions, in the operation of an X-ray diffractometer, the extraction of useful data from the raw recorded images, and the application of various corrections. Here we outline how the theory applies to these experimental procedures.

Diffraction geometry

An understanding of the diffraction of X-rays by a single crystal can be built up from the fundamental scattering interaction of X-rays with electrons, through atoms, molecules, and assemblies of molecules (such as the symmetry-generated contents of an entire unit cell), to a complete crystal structure. This leads to the Fourier transform in the form of the structure factor equation:

$$F(hkl) = \sum_{j=1}^{N} f_j \exp[2\pi i(hx_j + ky_j + lz_j)]$$

where each atom j at a position (x_j, y_j, z_j) in the unit cell has atomic scattering factor $f_j{}'$ incorporating atomic displacements.

As part of this treatment, the concept of the reciprocal lattice is introduced, as a convenient representation of the geometrical basis of diffraction:

$\mathbf{a}^* = (\mathbf{b} \times \mathbf{c}) / V \quad \mathbf{b}^* = (\mathbf{c} \times \mathbf{a}) / V \quad \mathbf{c}^* = (\mathbf{a} \times \mathbf{b}) / V$

$V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$

with the useful simplifying consequences that

$\mathbf{a} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1$

and $\mathbf{a} \cdot \mathbf{b}^* = \mathbf{a} \cdot \mathbf{c}^* = \mathbf{b} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{c}^* = \mathbf{c} \cdot \mathbf{a}^* = \mathbf{c} \cdot \mathbf{b}^* = 0$

At the same time Bragg's law emerges as a single equation for the geometry governing each individual diffracted beam:

$$\lambda = 2d(hkl)\sin\theta(hkl)$$

with its alternative interpretation as an apparent reflection of X-rays by a set of parallel lattice planes having interplanar spacing $d(hkl)$ and a plane normal bisecting the incident and diffracted X-rays and coplanar with them.

A further model that helps visualise the geometry of diffraction, particularly in showing how the Bragg equation is satisfied sequentially for individual sets of lattice planes as a crystal is rotated in an X-ray beam, is the Ewald construction. In this, a sphere of radius $1/\lambda$ is centred on the crystal, and the incident X-ray beam, together with its continuation beyond the crystal and beam-stop, forms one diameter. As the crystal is rotated while the sphere remains fixed, the reciprocal lattice rotates with the crystal. If the origin of the reciprocal lattice is placed on the surface of the sphere at the exit point of the incident beam, rather than at the crystal (this is a purely geometrical construction, not a physical reality), then rotation takes individual reciprocal lattice points through the sphere surface at different times.
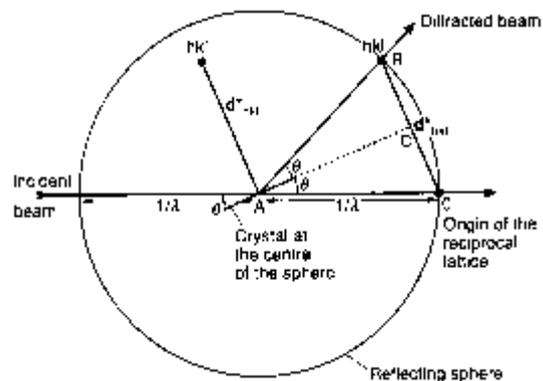
Using standard trigonometric relationships, we can show that the Bragg equation is satisfied for each point as it intersects the sphere, so this construction demonstrates not only the diffracted beam direction but also the point (or time) in the rotation at which the diffraction occurs. It also shows that there is a limit to the volume of reciprocal space that can give rise to diffraction, which is enclosed within a larger sphere of radius $2/\lambda$, centred on the reciprocal lattice origin. Calculations equivalent to this construction are important in setting up data collection strategies and in analysing the resultant diffraction images.

The orientation matrix

Geometrical calculations for the operation of diffractometers and interpretation of diffraction patterns are based on three-dimensional sets of axes, which need to be defined and related to each other. One set of axes is provided by the crystal itself: the axes of the reciprocal lattice, **a***, **b***, **c*** (these being uniquely related to the direct lattice vectors **a**, **b**, **c** by the definition given earlier). Note that these are generally of different lengths and not necessarily orthogonal. When a crystal is mounted in a random orientation on a goniometer head, these axes are initially unknown and need to be established.

The second set of axes is an orthogonal one (for convenience), also centred on the crystal, with unit vectors in each direction (the units are arbitrary, since it is directions rather than distances that matter). One axis, usually denoted **z**, is along the rotation axis of the goniometer head (the diffractometer ϕ axis); the other two, **x** and **y**, are perpendicular to this, but the axis convention may be different for different machines – one example would be to take **x** along the incident X-ray beam when all diffractometer angles are zero, with **y** completing a right-handed axis set.

The third set of axes, also orthogonal, is laboratory-fixed, so it does not rotate with the crystal, as do the first two. Typically, one axis is the (normally vertical) ω axis and another is the incident X-ray beam; this and the second set of axes are coincident when all diffractometer angles are zero. On modern area-detector diffractometers, the geometrical characterisation of each X-ray reflection referred to these laboratory axes will usually be expressed in terms of a set of diffractometer angles (ϕ, χ, ω, ϕ, κ, ω) together with detector settings (offset angle and crystal–detector distance) and the positions (two coordinates) of the reflection as recorded on the detector face.

Conversions between the second (crystal-fixed orthogonal) and third (laboratory-fixed) axis sets is machine-specific and essentially hidden from the user; it involves standard algebraic and trigonometric relationships related to the mathematics of rotation matrices.

The relationship between the two crystal-fixed axis sets is expressed in the crystal orientation matrix **A**:

$$\mathbf{x} = \mathbf{A}\,\mathbf{h}$$

where the vector **x** denotes the three coordinates *xyz* in the second axis set, and **h** is the same vector represented by the indices *hkl*, which are coordinates of the reciprocal lattice for a Bragg reflection. Finding the orientation matrix **A** is an essential step in measuring and interpreting a diffraction pattern. It is usually done on the basis of a preliminary restricted data collection before a full data set is measured, though it can be left until later, with the complete data collected 'blind' for subsequent analysis.

The nine elements of the 3×3 matrix **A** are the components (projections) of the three reciprocal cell axes on the three axes of the **x** set:

$$\mathbf{A} = \begin{pmatrix} a^*_x & b^*_x & c^*_x \\ a^*_y & b^*_y & c^*_y \\ a^*_z & b^*_z & c^*_z \end{pmatrix}$$

and encapsulate information on the six (reciprocal) lattice parameters and the orientation of the crystal and its (reciprocal) unit cell axes on the goniometer head (requiring three parameters). The unit cell parameters are easily extracted from the orientation matrix once it is known:

$$(\mathbf{A'A})^{-1} = \begin{pmatrix} \mathbf{a.a} & \mathbf{a.b} & \mathbf{a.c} \\ \mathbf{b.a} & \mathbf{b.b} & \mathbf{b.c} \\ \mathbf{c.a} & \mathbf{c.b} & \mathbf{c.c} \end{pmatrix}$$

where **A′** is the transpose of the matrix **A**.

Any observed reflection, for which the *xyz* coordinates are available from the diffractometer and detector parameters, can be indexed by the inverse relationship

$$\mathbf{h} = \mathbf{A}^{-1}\mathbf{x}$$

to give the reflection indices *hkl*.

Obtaining the orientation matrix and unit cell

A preliminary restricted set of diffraction images measured quickly yields a list of observed reflections, each with orthogonal crystal-fixed coordinates *xyz* and an intensity. Somehow indices *hkl* must be assigned to these and the orientation matrix established. The problem is aggravated by the limited data available (usually a relatively small section or sections of the reciprocal lattice) and the possible inclusion in the list of non-fitting reflections, usually of relatively low intensity, which may arise from measurement errors, non-diffraction effects, or a non-ideal single crystal such as a composite or twin. One approach is to present the complete list of reflections as a rotatable graphical reciprocal space display, in which the user may visually discern a lattice arrangement and select reciprocal lattice axes

manually. However, automatic computer indexing routines are generally used as a first attempt. The various methods can be summarised and explained briefly. In each case, the initial set of reflections, which are vectors in reciprocal space, can be augmented by adding to the list all their vector differences, many of which will be found to be essentially identical and can be averaged in that case.

The simplest approach is to select the three shortest non-coplanar vectors and make these the reciprocal cell axes; their nine xyz coordinates are the elements of $\mathbf{A}$. Now the indices of all the reflections can be calculated. If any of these are simple fractions, one or more of the reciprocal axes is too long and the necessary correction is made. Other indices not close to integers indicate non-fitting reflections. This method necessarily gives a primitive unit cell, and this must be checked to see if any alternative centred cell reveals higher metric symmetry.

A different method works primarily in real space rather than reciprocal space. Again the three shortest non-coplanar vectors are taken, and they are arbitrarily assigned indices $100, 010,$ and 001 (ensuring that they form a right-handed set). This gives an orientation matrix and unit cell, $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{c}'$. Although this cell is probably not correct, it must be a sub-cell, *i.e.* all true lattice vectors are consistent with it, but not all of its lattice vectors are consistent with the correct larger cell. Now all vectors $\mathbf{t} = u\mathbf{a}' + v\mathbf{b}' + w\mathbf{c}'$ are generated with integral u, v, w up to a specified maximum length and these are tested against every reflection: for a genuine lattice vector, $\mathbf{t.x}$ is an integer within a suitable small tolerance. The correct unit cell axes can be selected by hand or automatically from the list of genuine lattice vectors.

Both of these methods are adversely affected by the presence of significant numbers of non-fitting reflections, and will certainly fail if the initial set of three vectors do not all belong to a single diffraction pattern from one crystal component. A more exhaustive method circumventing such problems takes every combination of three vectors in the list in turn, their end-points defining a plane in reciprocal space, perpendicular to which lies a potential real space lattice vector; the length of this vector is obtained by projecting all the reciprocal lattice vectors on to it and finding the reciprocal space repeat from the pattern produced, ignoring non-fitting reflections. The complete list of real space vectors is examined to find a self-consistent set that defines a lattice, and a unit cell is chosen. Multiple-component samples such as twins (in which the same lattice is found in two or more different orientations) can be recognised.

A similar approach, requiring considerable computing power, generates real space vectors not from the list of observed reflections, but entirely at random within specified length limits, and tests them against all the reflections for consistency as described above. For each vector, badly fitting reflections are ignored and the rest are used to refine the three vector components to give the best (least-squares) fit. Most random vectors will be discarded, but those giving an acceptable fit to a reasonable proportion of the reflections are retained, and a unit cell is selected from vectors that fit the same (maximum sized) subset of reflections. This method also provides an effective solution for multiple-component samples, with individual components identified and their relative orientations indicated.

Whatever method is used, the elements of the orientation matrix should be refined to give the best fit to as many of the recorded reflections as possible. This will then enable the complete diffraction pattern to be indexed and optimum quality reflection intensities to be extracted. A further step, usually performed at the end of data collection, is to refine the unit cell parameters using all reflections of significant intensity, applying symmetry constraints appropriate to the crystal system.

Measuring reflection intensities

On a modern area-detector diffractometer, full data sets are usually collected by rotating one axis over a wide range (typically 180° for ϕ, ω, depending on mechanical restrictions) while holding other angles and the detector fixed. In order to give complete coverage of the unique portion of reciprocal space, especially in lower-symmetry crystal systems, it is often necessary to carry out several such scans with different values for the stationary axes. Depending on the detector size, crystal–detector distance, and X-ray wavelength, the detector may also need to be set to more than one swing angle relative to the incident X-ray beam. The detector distance must be sufficient to resolve adjacent reflections along a short reciprocal axis (long unit cell axis), and the maximum Bragg angle should be large enough to give a minimum d spacing of about 0.8 Å or less, in order to provide good resolution of electron density features in the structure. In general, it is advantageous to exceed the minimum symmetry-unique coverage of reciprocal space by a substantial margin, as the resultant data redundancy (also known as multiplicity of observations) improves data precision, provides an indication of data consistency, and enables absorption and other corrections to make use of symmetry-equivalent measurements. For chiral materials for which an absolute configuration is desirable, the coverage must also include an extensive set of Friedel pairs or their symmetry-equivalents; success in this also depends on choosing an appropriate X-ray wavelength for the chemical composition, so that there is significant resonant scattering.

All these factors, together with the overall level of diffraction intensity and its variation with Bragg angle (affected mainly by atomic displacements, disorder, and data collection temperature among other properties), feed into decisions about the data collection parameters, chiefly angle ranges and measurement times/speeds. These decisions can be assisted by relatively simple or sophisticated strategy calculations in which targets are set by the user for data precision (signal-to-noise), completeness, redundancy, and overall time available. These choices are important if data of optimum quality are to be obtained, on which will depend the outcome of the whole structure solution enterprise.

Whether the diffractometer scans are made with continuous motion or in separate steps, the diffraction pattern will be recorded as a sequence of discrete images covering a particular scan angle range. The choice of image angular interval may be dictated, to some extent, by the detector technology: for example, image plates and CCD detectors are integrating devices that need to be read out in a separate stage, enforcing a step-scan approach, while hybrid pixel detectors can be exposed and read continuously. In general, thin-slice operation is to be preferred, as it spreads each reflection over a number of consecutive narrow images, giving a more precise centroid of the reflection in angular terms and allowing

some kind of profile fitting to improve the intensity estimation for weak reflections that are not clearly distinguished from the background.

The process (known as integration) of locating each reflection on the appropriate image(s) on the basis of the orientation matrix, and of estimating the intensity and its standard uncertainty from the relevant sets of pixels, often with analysis and similarity-fitting of the reflection profiles in three dimensions, is usually carried out in parallel with the data collection itself, the integration running a few images behind the X-ray exposure. This allows some monitoring of the diffraction in real time, including a recognition of any small mechanical slippage of the crystal, which can usually be compensated by continual adjustment to the orientation matrix to follow the movement.

Processing and correcting the diffraction data

The data from integration are a set of indexed raw intensities. A number of adjustments need to be made before these can adequately represent $|F(hkl)|^2$ squared amplitudes suitable for crystal structure solution and refinement through applications of the Fourier transform equations.

1. Corrections for diffraction-specific effects

Reciprocal lattice points pass through the Ewald sphere surface at different speeds and therefore reflections have raw intensities that are not properly on the same scale. Correction for this is diffractometer-specific and purely dependent on geometry, so is easily made and always necessary. It is usually applied together with a correction for the fact that X-rays are partially polarised when they are diffracted; this effect at the sample crystal is a simple function of Bragg angle, but it is further influenced by the source of X-rays (laboratory source or synchrotron) and optical components used to select a single wavelength.

2. Frame scaling

Various factors can affect the relative overall intensity of individual diffraction images, particularly any variations in incident X-ray intensity, X-ray-induced or chemical deterioration in the crystal, as well as imperfections in the crystal mounting and centring in the incident beam and the size and profile of the X-ray beam itself. The impact of these factors, and corrections for them, can be made by comparing reflections that are symmetry-equivalent and should therefore have the same intensity. Other factors can also affect the intensities of symmetry-equivalent reflections, particularly absorption (see below), so corrections for these need to be made in parallel, either based on measured crystal dimensions or combined in a semi-empirical approach with the other frame-scaling corrections.

3. Absorption corrections

When X-rays are diffracted, they are also absorbed by the crystal through which they pass. The amount of absorption depends on the material's absorption coefficient (itself dependent on the chemical composition and on the X-ray wavelength) and on the path length of the X-rays through the crystal for each individual reflection. In some cases the absorption is small and can be neglected, but it must be corrected if the effect is significant. If the crystal shape can be reasonably accurately described as a set of indexed faces and their distances from a central point in the crystal, then corrections can be calculated with a high degree of accuracy. Often corrections are made by modelling the absorption effect with mathematical functions such as spherical harmonics and refining the parameters of these functions to minimise the variation of intensity of symmetry-equivalent reflections that have different X-ray path lengths in the crystal; such corrections are called empirical (based on the diffraction data rather than physical dimensions) or semi-empirical (when an overall average crystal dimension is included in the calculations).