International Tables for Crystallography (2006). Vol. F, Section 11.1.3, pp. 209–210.

11. DATA PROCESSING

11.1. Automatic indexing of oscillation images

BY M. G. ROSSMANN

11.1.1. Introduction

Auto-indexing routines have been used extensively for initiating diffraction data collection with a single-point-detector device (Sparks, 1976, 1982). These methods depend upon the precise knowledge of the reciprocal-lattice vectors for a few selected reflections. Greater difficulty has been encountered for automatic indexing of oscillation images recorded on two-dimensional detectors using randomly oriented crystals, as is frequently the case for macromolecular crystal samples. In the past, the practice was to orient crystals relative to the camera axes with an accuracy of at least 1°. In this case, the indexing procedure required only refinement of the crystal orientation matrix (Wonacott, 1977; Rossmann, 1979). The 'American method' (Rossmann & Erickson, 1983), where crystals are oriented more or less randomly, is currently used because of the need for optimizing available synchrotron time and because of the deterioration in radiation-sensitive crystals during the setting process.

A variety of techniques were suggested to determine the crystal orientation, some of which required initial knowledge of the cell dimensions (Vriend & Rossmann, 1987; Kabsch, 1988), while more advanced techniques (Kim, 1989; Higashi, 1990; Kabsch, 1993) determined both cell dimensions and crystal orientation. All these methods start with the determination of the reciprocal-lattice vectors assuming that the oscillation photographs are 'stills'. The methods of Higashi and Kabsch, as well as, in part, Kim's, analyse the distribution of the difference vectors generated from the reciprocal-lattice vectors. The most frequent difference vectors are taken as the basis vectors defining the reciprocal-lattice unit cell and its orientation. In addition, Kim's technique requires the input of the orientation of a likely zone-axis direction onto which the reciprocal-lattice vectors are then projected. The projections will have a periodicity distribution consistent with the reciprocal-lattice planes perpendicular to the zone axis. Duisenberg (1992) used a similar approach for single-point-detector data, although he did not rely on prior knowledge of the zone-axis direction. Instead, he defined possible zone axes as being perpendicular to a reciprocal-lattice plane by combining three, suitably chosen, reciprocal-lattice points.

None of the above techniques were entirely satisfactory as they sometimes failed to find a suitable crystal orientation matrix. A major advance was made in the program *DENZO*, a part of the *HKL* package (Otwinowski & Minor, 1997), which not only has a robust indexing procedure but also has a useful graphical interface. Unfortunately, the indexing technique used in the procedure has never been described, except for a few hints in the manual on the use of an FFT (fast Fourier transform). Indeed, Bricogne (1986) suggested that a three-dimensional Fourier transformation might be a powerful indexing tool, and Strouse (1996) developed such a procedure for single-point-detector data. However, for large unit cells this procedure requires an excessive amount of memory and time (Campbell, 1997).

11.1.2. The crystal orientation matrix

The position $\mathbf{x}$ (x, y, z) of a reciprocal-lattice point can be given as

$$\mathbf{x} = [\Phi][A]\mathbf{h}. \qquad (11.1.2.1)$$

The matrix $[\Phi]$ is a rotation matrix around the camera's spindle axis for a rotation of φ. The vector $\mathbf{h}$ represents the Miller indices (h, k, l) and $[A]$ defines the reciprocal unit-cell dimensions and the orientation of the crystal lattice with respect to the camera axes when $\varphi = 0$. Thus,

$$[A] = \begin{pmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{pmatrix}, \qquad (11.1.2.2)$$

where a_x^*, a_y^* and a_z^* are the components of the crystal $\mathbf{a}^*$ axis with respect to the orthogonal camera axes. When an oscillation image is recorded, the position of a reciprocal-lattice point is moved from $\mathbf{x}_1$ to $\mathbf{x}_2$, corresponding to a rotation of the crystal from φ_1 to φ_2. The recorded position of the reflection on the detector corresponds to the point $\mathbf{x}$ when it is on the Ewald sphere somewhere between $\mathbf{x}_1$ and $\mathbf{x}_2$. The actual value of φ at which this crossing occurs cannot be retrieved from the oscillation image. We shall therefore assume here, as is the case in all other procedures, that $[\Phi][A]$ defines the crystal orientation in the centre of the oscillation range. Defining the camera axes as in Rossmann (1979), it is easy to show that a reflection recorded at the position (X, Y) on a flat detector normal to the X-ray beam, at a distance D from the crystal, corresponds to

$$\begin{aligned} x &= \frac{X}{\lambda(X^2 + Y^2 + D^2)^{1/2}} \\ y &= \frac{Y}{\lambda(X^2 + Y^2 + D^2)^{1/2}} \\ z &= \frac{D}{\lambda(X^2 + Y^2 + D^2)^{1/2}}, \end{aligned} \qquad (11.1.2.3)$$

where λ is the X-ray wavelength.

If an approximate $[A]$ matrix is available, the Miller indices of an observed peak at (X, Y) can be roughly determined using (11.1.2.3) and (11.1.2.1), where

$$\mathbf{h} = [A]^{-1}[\Phi]^{-1}\mathbf{x}, \qquad (11.1.2.4)$$

with the error being dependent upon the width of the oscillation range, the error in the detector parameters and errors in determining the coordinates of the centres of the recorded reflections.

11.1.3. Fourier analysis of the reciprocal-lattice vector distribution when projected onto a chosen direction

If the members of a set of reciprocal-lattice planes perpendicular to a chosen direction are well separated, then the projections of the reciprocal-lattice vectors onto this direction will have an easily recognizable periodic distribution (Fig. 11.1.3.1). Unlike the procedure of Kim (1989), which requires the input of a likely zone-axis direction, the present procedure tests all possible directions and analyses the frequency distribution of the projected reciprocal-lattice vectors in each case. Also, unlike the procedure of Kim, the periodicity is determined using an FFT.

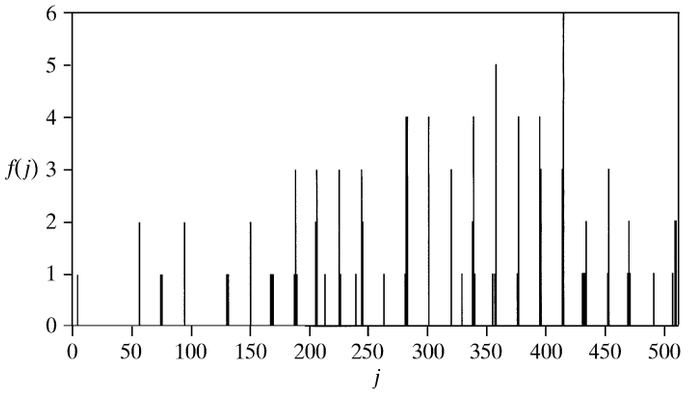

Fig. 11.1.3.1. Frequency distribution of the projected reciprocal-lattice vectors for a suitably chosen direction of a diffraction pattern from a fibritin crystal (Tao *et al.*, 1997). Reproduced with permission from Steller *et al.* (1997). Copyright (1997) International Union of Crystallography.

Let **t** represent a dimensionless unit vector of a chosen direction. Then, the projection p of a reciprocal-lattice point **x** onto the chosen vector **t** is given by

$$p = \mathbf{x} \cdot \mathbf{t}. \qquad (11.1.3.1)$$

To apply a discrete FFT algorithm, all such projections of the reciprocal-lattice points onto the chosen direction **t** are sampled in small increments of p. For the given direction, the values of the projections are in a range between the endpoints $p_{\min}$ and $p_{\max}$. If the maximum real cell dimension is assumed to be $a_{\max}$, then the maximum number of reciprocal-lattice planes between the observed limits of p is $(p_{\max} - p_{\min})/(1/a_{\max})$. Hence, the number of useful grid points along the direction **t** should be

$$m = (p_{\max} - p_{\min}) n a_{\max}, \qquad (11.1.3.2)$$

where n represents the number of grid points between successive reciprocal-lattice planes and is normally set to 5. Then, the frequency $f(p)$ in the range $p < \mathbf{x} \cdot \mathbf{t} < p + \Delta p$ can be given as $f(p)\Delta p = f(j)$, where j is the closest integer to $(p - p_{\min})/\Delta p$ and $\Delta p = n a_{\max}$. Thus, the discrete Fourier transform of this frequency distribution will be given by the summation

$$F(k) = \sum_{j=0}^{m} f(j) \exp(2\pi i k j). \qquad (11.1.3.3)$$

The transform is then calculated using a fast Fourier algorithm for all integer values between 0 and $m/2$ (Fig. 11.1.3.2). The Fourier coefficients that best represent the periodicity of the frequency distribution will be large. The largest coefficient will occur at $k = 0$ and correspond to the number of vectors used in establishing the frequency distribution. The next set of large coefficients will correspond to the periodicity that represents every reciprocal-lattice plane. The ratio of this maximum to $F(0)$ will be a measure of the tightness of the frequency distribution around each lattice plane. Subsequent maxima will be due to periodicities spanning every second, third *etc.* frequency maximum and will thus be progressively smaller (Fig. 11.1.3.2). The largest $F(k)$ (when $k = l$), other than $F(0)$, will, therefore, correspond to an interval of d^* between reciprocal-lattice planes in the direction of **t** where $d^* = l/(n a_{\max})$.

11.1.4. Exploring all possible directions to find a good set of basis vectors

The polar coordinates ψ, φ will be used to define the direction **t**, where ψ defines the angle between the X-ray beam and the chosen direction **t**. The Fourier analysis is performed for each direction **t** in the range $0 < \psi \leq \pi/2, 0 < \varphi \leq 2\pi$. A suitable angular increment in ψ was determined empirically to be about 0.03 rad (1.7°). For each value of ψ, the increment in φ is taken to be the closest integral value to $(2\pi \sin\psi)/0.03$. This procedure results in ~7300 separate, roughly equally spaced, directions.

For each direction **t**, the distribution of the corresponding $F(k)$ coefficients is surveyed to locate the largest local maximum at $k = l$. The ψ and φ values associated with the 30 largest maxima are selected for refinement by a local search procedure to obtain an accuracy of 10^{-4} rad (~0.006°). If the initial angular increment (0.03 rad) used for the hemisphere search was reduced, then it would not be necessary to refine quite as many local maxima. However, to increase the efficiency of the search procedure, the ratio of angular increments to the number of refined positions was chosen to minimize the total computing time. The $F(l)$ values of the refined positions are then sorted by size. Directions are chosen from these vectors to give a linearly independent set of three basis vectors of a primitive real-space unit cell. These are then converted to the basis vectors of the reciprocal cell. The components of the three reciprocal-cell axes along the three camera axes are the nine components of the crystal orientation matrix [A] (11.1.2.2).

The final step in the selection of the best [A] matrix is to choose various nonlinear combinations of the refined vectors that have the biggest $F(l)$ values. That set of three vectors which gives the best indexing results is then chosen to represent the crystal orientation matrix [A]. A useful criterion is to determine the nonintegral Miller indices h' from (11.1.2.4) using the [A] matrix and the known reciprocal-lattice vectors **x**. Any reflection for which any component $|\mathbf{h} - \mathbf{h}'|$ is bigger than, say, 0.2 is rejected. The best [A] matrix is chosen as the one with the least number of rejections. In most cases, the best combination corresponds to taking the three largest $F(l)$ values.

The program goes on to determine a reduced cell from the cell obtained by the above indexing procedure (Kim, 1989). The reduced cell is then analysed in terms of the 44 lattice characters (Burzlaff *et al.*, 1992; Kabsch, 1993) in order to evaluate the most likely Bravais lattice and crystal system.

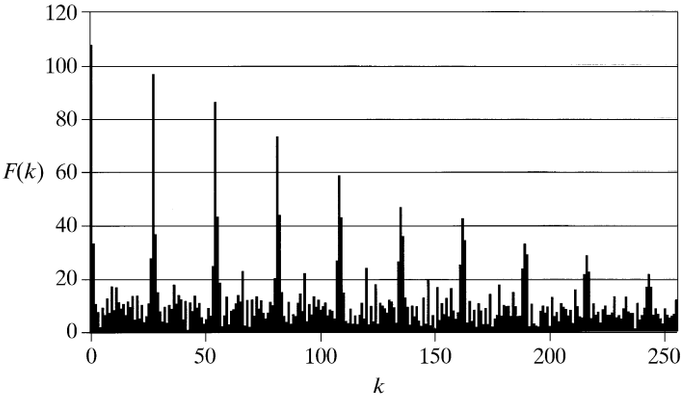

Fig. 11.1.3.2. Fourier analysis of the distribution shown in Fig. 11.1.3.1. The first maximum, other than $F(0)$, is at k = 27, corresponding to $(1/d^*) = 41.9$ Å and a value of $F(27)$ = 97.0. Reproduced with permission from Steller *et al.* (1997). Copyright (1997) International Union of Crystallography.

references