J. Appl. Cryst. (1997). **30**, 1036–1040

An Algorithm for Automatic Indexing of Oscillation Images using Fourier Analysis

INGO STELLER,† ROBERT BOLOTOVSKY AND MICHAEL G. ROSSMANN*

Department of Biological Sciences, 1392 Lilly Hall of Life Sciences, Purdue University, West Lafayette, IN 47907-1392, USA. E-mail: mgr@indiana.bio.purdue.edu

(Received 7 March 1997; accepted 13 June 1997)

Abstract

A fast and reliable algorithm has been developed for auto-indexing oscillation images. The observed reflection positions are used to compute the corresponding reciprocal-lattice vectors assuming a stationary crystal. These vectors are then projected onto a chosen direction. The projections are subjected to a one-dimensional Fourier analysis. A large Fourier term will be found that has a periodicity corresponding to the interplanar distance if the chosen direction happens to be perpendicular to a set of widely separated reciprocal-lattice planes. Exploration of such directions over a hemisphere establishes the best potential basis vectors of the real cell. The program has successfully determined the lengths and directions of basis vectors for numerous unit cells having cell dimensions ranging from 16 to 668 Å and oscillation ranges from 0.2 to 2.0°.

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 on 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, analyze 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, 1996), which not only has a robust indexing procedure but also has a useful graphical interface. Unfortunately, the indexing technique used in the procedure was never 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).

In creating a modern processing package, it is essential to have an indexing procedure that is at least equal to the rapid and robust process used by *DENZO*. Here we describe a procedure that has both these properties. Unlike all the other procedures mentioned above, with the possible exception of *DENZO*, the technique described here combines a simple scan of reciprocal space with a one-dimensional Fourier analysis.

† Present address: CompuNet Computer AG, Severinstrasse 42, 45127 Essen, Germany.

2. The crystal orientation matrix

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

$$\mathbf{x} = [\Phi][A]\mathbf{h}. \tag{1}$$

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,

$$[A] = \begin{pmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{pmatrix} \tag{2}$$

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 center 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

$$\begin{aligned} x &= \frac{X}{\lambda(X^2 + Y^2 + D^2)^{1/2}} \\ y &= \frac{Y}{\lambda(X^2 + Y^2 + D^2)^{1/2}} \\ z &= \frac{D}{\lambda(X^2 + Y^2 + D^2)^{1/2}} \end{aligned} \tag{3}$$

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 (3) and (1), where

$$\mathbf{h} = [A]^{-1}[\Phi]^{-1}\mathbf{x} \tag{4}$$

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

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. 1). Unlike the procedure of Kim (1989), which requires the input of a likely zone-axis direction, or the method of Duisenberg (1992), which relies on the arbitrary selection of three reciprocal-lattice points, the present procedure tests all possible directions and analyzes the frequency distribution of the projected reciprocal-lattice vectors in each case. Also, unlike the procedures of Kim or Duisenberg, the periodicity is determined using an FFT.

Let $\mathbf{t}$ represent a dimensionless unit vector of a chosen direction. Then the projection p of a reciprocal-lattice point $\mathbf{x}$ onto the chosen vector $\mathbf{t}$ is given by

$$p = \mathbf{x} \cdot \mathbf{t}. \tag{5}$$

To apply a discrete FFT algorithm, all such projections of the reciprocal-lattice points onto the chosen direction $\mathbf{t}$ 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 $\mathbf{t}$ should be

$$m = (p_{\max} - p_{\min})na_{\max}, \tag{6}$$

where n is 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 $\Delta p = na_{\max}$. Thus, the discrete Fourier transform of this frequency distribution will be given by the summation

$$F(k) = \sum_{j=0}^{m} f(j)\exp(2\pi ikj). \tag{7}$$

The transform is then calculated using a fast Fourier

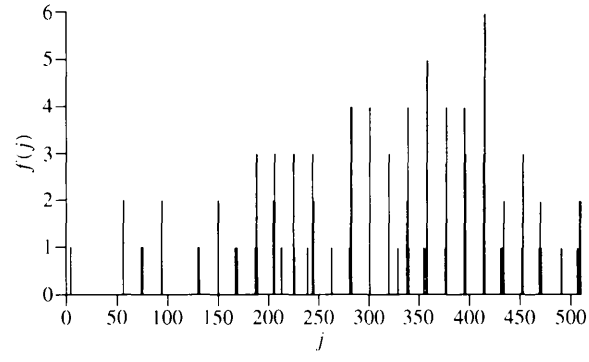

Fig. 1. Frequency distribution of the projected reciprocal-lattice vectors for a suitably chosen direction of a fibritin image in data-set number 6 of Table 2.

algorithm for all integer values between 0 and $m/2$ (Fig. 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. 2). The largest $F(k)$ (when $k = l$), other then $F(0)$, will, therefore, correspond to an interval of d^* between reciprocal-lattice planes in the direction of **t** where $d^* = l/(na_{max})$.

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 (Table 1). 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 $[A]$ (2).

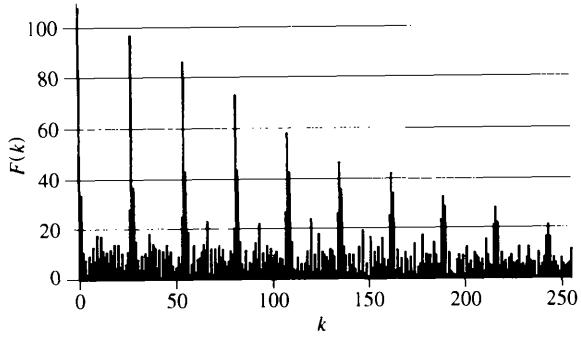

Fig. 2. Fourier analysis of the distribution shown in Fig. 1. The first maximum, other than $F(0)$, is at $k = 27$, corresponding to $(1/d^*) = 41.9$ Å and a value of $F(27) = 97.0$.

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 that 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 (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 analyzed in terms of the 44 lattice characters (Burzlaff, Zimmermann & de Wolff, 1992; Kabsch, 1993) in order to evaluate the most likely Bravais lattice and crystal system.

5. The effect of errors on indexing

When the vectors **x** are calculated from peak coordinates, it is necessary to make assumptions about the camera parameters such as crystal-to-film distance, pixel size, orthogonality of detector to X-ray beam, *etc.* Errors in these parameters affect the determination of the reciprocal-lattice vectors and, hence, the crystal orientation matrix $[A]$ determined by the procedure. However, the definition of the $[A]$ matrix is dependent only on the crystal cell parameters and the crystal orientation. Thus, in principle, the $[A]$ matrix should be completely independent of camera parameters; in practice, however, that is not the case.

The components of the basis vectors parallel to the X-ray beam are necessarily rather inaccurate when applying any auto-indexing procedure. This is because the usual flat detector records data only in a forward direction and because the normal oscillation angle is small, resulting in a lack of information about the extent of the reciprocal-lattice along the X-ray beam. Thus, it would be an advantage to combine images of one crystal taken at different rotation angles, or best, separated by a 90° rotation. In principle, this is not difficult as the vectors **x** from different orientations of the crystal can be combined with different oscillation angles $\delta\varphi$ using equation (1). However, in practice, the errors in the values of camera parameters used for calculating the

Table 1. *Sorted list of refined directions of the largest F(l) terms for the fibritin data-set*†

Three nonlinear vectors were selected from this list to generate an orientation matrix.

Polar coordinates		Components of real space vector **t**				
ψ (rad)	φ (rad)	x (Å)	y (Å)	z (Å)	$1/d^*$ (Å)	$F(l)$‡
1.714	−1.318	39.197	−10.165	5.821	40.91	103.1
1.420	−0.307	−12.876	40.721	6.285	43.17	101.6
1.278	0.711	26.194	30.405	12.134	41.93	97.1
1.321	0.187	13.419	70.942	18.398	74.51	89.7
1.530	−0.616	−64.592	91.251	4.609	111.89	88.0
1.345	1.472	104.591	10.329	24.076	107.82	87.6
1.552	−0.796	−52.301	51.178	1.406	73.19	87.3
1.282	1.266	65.109	20.486	20.255	71.20	86.4
1.496	2.159	90.908	−60.680	8.189	109.61	83.6
1.360	0.004	0.402	111.661	23.890	114.19	80.8
1.277	1.063	91.279	50.803	31.557	109.13	79.4
1.069	−2.421	−70.541	−80.315	58.608	121.91	71.9
0.958	−1.561	−96.026	0.930	67.467	117.36	61.1
0.639	−1.723	−57.155	−8.789	77.768	96.91	60.6
0.754	−2.416	−44.303	−49.928	71.115	97.53	60.5
0.401	−0.971	−30.927	21.251	88.302	95.94	54.9
0.309	−2.378	−18.495	−19.330	83.832	88.00	54.4
0.384	2.551	20.990	−29.103	87.755	94.81	51.7
0.144	0.618	8.073	11.360	96.331	97.33	51.5
0.329	2.602	16.840	−28.096	95.851	101.29	45.0

† See data-set number 6 in Table 2. ‡ Maximum value is equal to 108, the number of reflections used.

positions **x**, and the assumption that the crystal is stationary for any given image, introduces different errors into the calculation of the position **x** for widely separated images.

An attempt was made to combine the reciprocal-lattice vectors derived from three separate images, taken at φ = 0, 14.8 and 37.8°, recorded on a CCD (charge-coupled device) detector using a frozen human rhinovirus 16 (HRV16) crystal (data-set number 2 in Table 2) at beamline SBC-19ID at the Advanced Photon Source (Argonne National Laboratory, Chicago). Each image was indexed successfully when analyzed by itself. However, on combining the information from the three images, the FFT systematically determined the [*A*] matrix for one of the images that contained about 30% more useful reflections than the other two images. This showed that the FFT found the dominant periodicity and that the positions of the reciprocal-lattice points for the other images did not mesh precisely with those of the dominant image. Although unsuccessful for the purpose initially proposed, this result is particularly interesting to show that split crystals containing a dominant fragment would be readily indexable with the auto-indexing procedure described here. Omission of the indexed reflections would then allow indexing of the minor component of the crystal.

The 'best' [*A*] matrix can be found by post-refinement procedures (Schutt & Winkler, 1977; Rossmann *et al.*, 1979; Winkler *et al.*, 1979), which depend only on reflection intensities. The camera parameters can then be refined with [*A*] fixed.

6. Applications

The above method has been tested on a variety of data-sets with primitive cell dimensions ranging from about 16 to 668 Å and oscillation ranges from 0.2 to 2.0° (Table 2). The diffraction images were surveyed for peaks using the programs *SCANDUMP* (Kim, 1989) or *HKL* (Otwinowski & Minor, 1996). The major differences in the peak picking procedures were in their speed and versatility.

The present indexing program requires input of crystal-to-film distance, wavelength, raster size, direct beam position and the image rotation parameter that relates the raster grid orientation to the camera axes. In addition, the program requires an estimate of the maximum cell dimension, a_{max}. If the input value of a_{max} is very much larger than the actual largest cell length, then the program uses an excessive number of grid points in sampling the projections p, requiring more time than is really necessary in the FFT. If the maximum cell dimension is set substantially too small, the program is likely to fail.

The [*A*] matrix found in each of the examples given in Table 2 was able to predict correctly the positions of observed reflections. Furthermore, the [*A*] matrix found for data-set numbers 1 and 2 was used for successful data reduction with the Purdue data processing package (Rossmann, 1979; Rossmann *et al.*, 1979). The auto-indexing was sufficiently accurate to proceed with intensity estimation in all cases. When the input parameters (especially direct beam position) were set

Table 2. *Indexing of different images*

(1) Oscillation range in °. (2) Resolution limit in Å. (3) Number of reflections used for indexing. (4) Maximum expected cell dimension in Å. (5) Time used for indexing in s on an SGI O^2 R5000SC workstation. (6) Intercellular adhesion molecule 1.

Data-set no.	Data-set	Image format	(1)	(2)	(3)	(4)	(5)	Unit-cell dimensions a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)
1	Frozen HRV14	FUJI	0.25	4.0	2454	500	30.9	432.3	437.0	437.2	89.5	89.8	89.8
2	Frozen HRV16	APS-1 (3 × 3 CCD)	0.20	4.0	247	500	10.3	329.7	339.8	351.6	89.7	89.8	90.0
3	ICAM-1 (6)	RAXIS II	2.0	2.1	48	200	4.2	42.3	63.4	88.5	90.0	90.0	90.4
4	Collagen	RAXIS IV	0.25	1.7	118	100	3.5	15.6	39.7	58.2	75.2	83.6	89.7
5	φX174	FUJI	0.30	2.1	2064	1000	42.5	664.1	668.8	668.2	71.2	71.2	69.4
6	Fibritin	RAXIS II	0.80	2.1	108	130	3.4	40.9	41.8	87.1	89.0	89.5	62.6

correctly, the indexing procedure was successful for all tested data-sets.

A limit to the indexing procedure will be reached when the approximations implied by equation (3) reach an error of about half a reciprocal-lattice unit or when there is a significant amount of overlap between successive 'lunes'. As the experimenter will wish to avoid spot overlap, the chosen oscillation angle, $\delta\varphi$, will be suitably small. The actual size will depend on the size of the unit-cell dimensions. The larger the unit-cell dimensions, the smaller will be the oscillation angle. Thus, the conditions that might disrupt successful indexing will normally be outside the likely experimental range. It is easy to show that, in general, for a cubic crystal of cell dimension a, it will be necessary for $\delta\varphi$ (°) to be approximately less than $(R/a)(90/\pi)$, where the limit of resolution is R. If $\delta\varphi$ exceeds this value, then the approximation of equation (3) will be rather poor and considerable reflection overlap will occur beyond a resolution of R.

7. The program

The auto-indexing program has been written in C and implemented on an SGI O^2 workstation. It will become a component of a general data-processing system (DPS). The auto-indexing component of DPS is available over the WWW including the source code. The run-time for auto-indexing is given in Table 2 and is sufficiently short for the procedure to be run interactively. The program described here will also be included (Leslie, 1996) in the *MOSFLM* data-processing package (Leslie, 1992).

We thank Charles E. Strouse (University of California, Los Angeles) for supplying us with his Fourier transform indexing program, and Dan Marinescu and Robert E. Lynch (Purdue University) for valuable discussions and encouragement. We also thank the members of the Purdue Structural Biology group for supplying data-sets. We thank Cheryl Towell and Sharon Wilder for help in preparation of the manuscript. The work was supported by a grant from the National Science Foundation to MGR. IS would like to thank the Deutsche Forschungsgemeinschaft for a postdoctoral fellowship.

References

Bricogne, G. (1986). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software* (Phase III), pp. 28. LURE, Paris, 12–19 November.

Burzlaff, H., Zimmermann, H. & de Wolff, P. M. (1992). *International Tables for Crystallography*, Vol. A, edited by T. Hahn, pp. 737–749. Dordrecht: Kluwer Academic Publishers.

Campbell, J. W. (1997). *CCP4 Newsletter* 33, edited by M. Winn, Warrington, England: Daresbury Laboratory.

Duisenberg, A. J. M. (1992). *J. Appl. Cryst.* **25**, 92–96.

Higashi, T. (1990). *J. Appl. Cryst.* **23**, 253–257.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 67–71.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

Kim, S. (1989). *J. Appl. Cryst.* **22**, 53–60.

Leslie, A. G. W. (1992). *CCP4 and ESF-EACMB Newsletter on Protein Crystallography* No. 26. Warrington, England: Daresbury Laboratory.

Leslie, A. G. W. (1996). Personal communication.

Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.

Rossmann, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.

Rossmann, M. G. & Erickson, J. W. (1983). *J. Appl. Cryst.* **16**, 629–636.

Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

Schutt, C. & Winkler, F. K. (1977). *The Rotation Method in Crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 173–186. Amsterdam: North-Holland.

Sparks, R. A. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlácek, pp. 452–467. Copenhagen: Munksgaard.

Sparks, R. A. (1982). *Computational Crystallography*, edited by D. Sayre, pp. 1–18. New York: Oxford University Press.

Strouse, C. E. (1996). Personal communication.

Vriend, G. & Rossmann, M. G. (1987). *J. Appl. Cryst.* **20**, 338–343.

Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). *Acta Cryst.* A**35**, 901–911.

Wonacott, A. J. (1977). *The Rotation Method in Crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 75–103. Amsterdam: North-Holland.