

The Rossmann Fourier autoindexing algorithm in *MOSFLM*

Harold R. Powell

MRC-LMB, MRC Centre, Hills Road, Cambridge
CB2 2QU, England

Correspondence e-mail:
harry@mrc-lmb.cam.ac.uk

The fast Fourier transform (FFT) autoindexing routines written by the Rossmann group at Purdue University have been incorporated in *MOSFLM*, providing a rapid and reliable method of indexing oscillation images. This is a procedure which extracts direct-space information about the unit cell from the FFT. The method and its implementation in *MOSFLM* are discussed.

Received 8 February 1999

Accepted 9 July 1999

1. Introduction

Most newcomers to crystallography and diffraction-image processing probably view indexing as a 'black-box process' (Fig. 1): a diffraction image is presented to a program (often only available without the source code), which performs the process automatically in a matter of seconds; some programs are more powerful and reliable than others, and it has become the habit of many workers to use one program for the auto-indexing step before using another for peak integration. Usually, the result of autoindexing is presented as a choice of unit cells corresponding to solutions for different Bravais lattices, each with a discrepancy index from a basis triclinic cell. The combined orientation matrix [**UB**] or [**A**], which contains the unit-cell dimensions and the information regarding the mis-setting of the unit cell with respect to a set of machine coordinates, is produced as a matter of course. The mechanics of some of these black boxes have been discussed at length elsewhere (Duisenberg, 1992; Higashi, 1990; Kim, 1989; Macíček & Yordanov, 1992); for example, the existing autoindexing method distributed with *MOSFLM* (Leslie, 1992) relies on the calculation of many difference vectors between diffraction maxima in reciprocal space (Kabsch, 1993). A brief explanation of the new method which has been incorporated into *MOSFLM* will be presented here, together with a short discussion of its implementation. The principal advantage of recent autoindexing techniques over those used formerly (*e.g.* Messerschmidt & Pflugrath, 1987) is that prior knowledge of unit-cell dimensions and crystal orientation is not required.

The black box is doing a number of things: searching for suitable reflections, converting the two-dimensional detector coordinates to three-dimensional reciprocal-space coordinates, performing the indexing itself, choosing a likely triclinic cell, applying transformations to the initial triclinic cell in order to generate the 44 lattice characters and calculating their discrepancy indices.

2. Fourier indexing

The use of a three-dimensional Fourier transform around the origin for indexing a diffraction pattern was suggested over a

decade ago (Bricogne, 1986); a similar approach appears to have been incorporated in the program *DENZO*, which has been distributed as part of the diffraction-image processing suite *HKL* for some years (Otwinowski & Minor, 1997). This program presents a particularly robust implementation and has been widely used in many laboratories worldwide in recent years. The underlying processes involved in the method have been hinted at, but no details of the program code are currently available. A three-dimensional FFT has been used to index diffraction images by calculating a Patterson function from a set of reflections which have all been assigned unit intensity (Campbell, 1998). A single FFT is calculated, but even with a coarse grid for the reflection binning, this three-dimensional approach involves the calculation of many thousands of data points. The method makes extensive use of code which is pre-existing in programs of the *CCP4* suite (Collaborative Computational Project, Number 4, 1994).

In order to produce an autoindexing routine which compares well with that used in *HKL* (in the sense that a similar subset of test images can be indexed by each program), a set of subroutines using a large number of one-dimensional FFTs has been written (Steller *et al.*, 1997) and discussed at this meeting (Rossmann & Van Beek, 1999); these have been implemented in the *DPS* image-processing suite, which is distributed with the ADSC CCD area detector. The underlying code (the '*DPS* indexing routine') has been made available to the author and has been implemented in *MOSFLM*. The method will be discussed below, but has been explained rigorously elsewhere (Steller *et al.*, 1997). The use of one-dimensional Fourier transforms to provide three-dimensional information about a diffraction experiment is hardly a novel idea; it was pre-empted in the early days of X-ray crystallography before the development of fast electronic computers and became formalized in the production and development of tools such as Beevers–Lipson strips in 1936, which were in common use in the science until the late 1950s and early 1960s. Computationally, it is much less expensive to compute a large number of one-dimensional FFTs with a fine grid than even small three-dimensional FFTs with a much coarser grid.

Independent of the area detector being used, three items which are used in calculating the reciprocal-space coordinates of the reflections *via* equations (1), (2) and (3) must be known about the diffraction image, namely, the direct-beam position, the crystal-to-detector distance and the radiation wavelength.

$$x = X/[\lambda(X^2 + Y^2 + D^2)^{1/2}], \quad (1)$$

$$y = Y/[\lambda(X^2 + Y^2 + D^2)^{1/2}], \quad (2)$$

$$z = D/[\lambda(X^2 + Y^2 + D^2)^{1/2}], \quad (3)$$

where x , y and z are reciprocal-space coordinates, X and Y are detector coordinates in some distance unit, *e.g.* millimetres, D is the crystal-to-detector distance (at X, Y) in the same units as X and Y , and λ is the radiation wavelength.

An error of more than half the inter-reflection distance for the direct-beam position must inevitably lead to failure of indexing, in the sense that although the unit-cell dimensions

may be more-or-less correct, the reflections themselves will have the wrong indices (Otwinowski & Minor, 1997). A number of other parameters are required, about which assumptions are made, *e.g.* detector planarity, orientation of the detector coordinate system with respect to a program coordinate system, crystal rotation axis, orthogonality of the detector to the X-ray beam *etc.*, but these are detector dependent and so can be regarded as secondary.

With regard to indexing an oscillation image, the assumption is made that the diffraction maxima are taken from a still image; provided that the oscillation angle and crystal mosaicity are both small, this is close enough to being valid that successful indexing can be performed. For an oscillation image, the φ angle of the crystal can be taken as the mean of its starting and finishing orientation angles for the image concerned. By including the rotation of the crystal in the

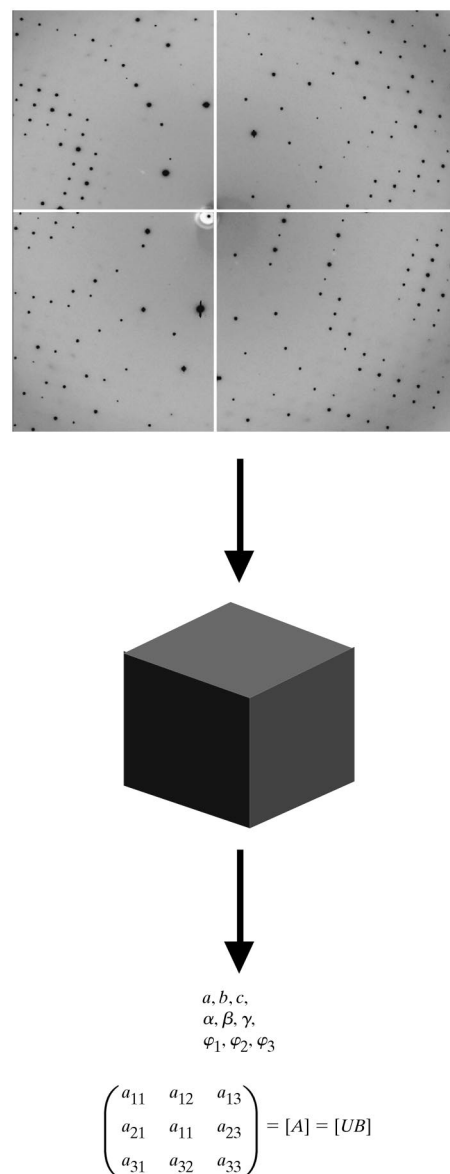


Figure 1
Autoindexing viewed as a black-box process.

calculation of the reciprocal-lattice coordinates, data from multiple images can be included in the matrix determination.

3. Indexing in practice

If a high-quality diffraction image is considered (*e.g.* Fig. 1), it is a simple matter to see that in this case there are two reciprocal-lattice dimensions more or less perpendicular to each other. The extraction of periodic information from a diffraction image is trivial for the human eye. With experience and the knowledge that there is more information contained in the separation of the lunes of reflections, it would be possible to determine approximate values for the third linear dimension

and the remaining inter-axial angles. If this image were to be reproduced life-size on paper, it would be straightforward to measure the spot separations with a ruler and the angles between the rows of spots with a protractor to obtain an adequate estimate of the unit-cell dimensions.

However, extracting the periodic information is far from trivial for a computer program, and any program has to do a lot of work to determine that there is a periodicity along any one direction. The process of analysing periodic patterns is simplified by the use of Fourier transforms, but there is still considerable calculation to be performed.

The indexing process itself can be presented by way of a flow chart (Fig. 2). The first step is to provide a selection of the reflections (N) on the image for the indexing process; the

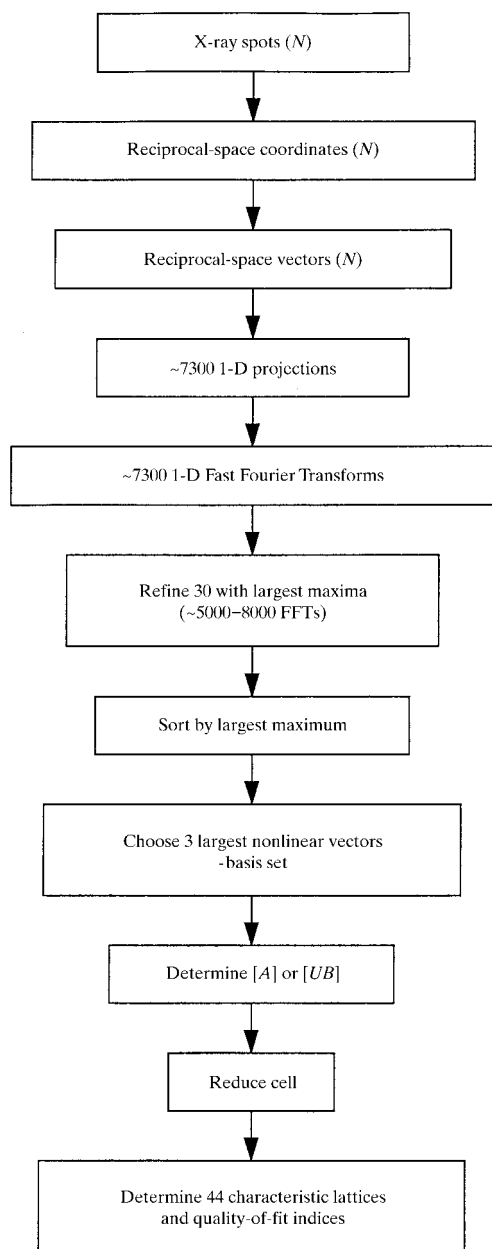


Figure 2
Flow chart showing a simplified representation of the autoindexing process. The diffraction image is the central region of the test image of ribonucleotide synthase as supplied with the *DPS* suite.

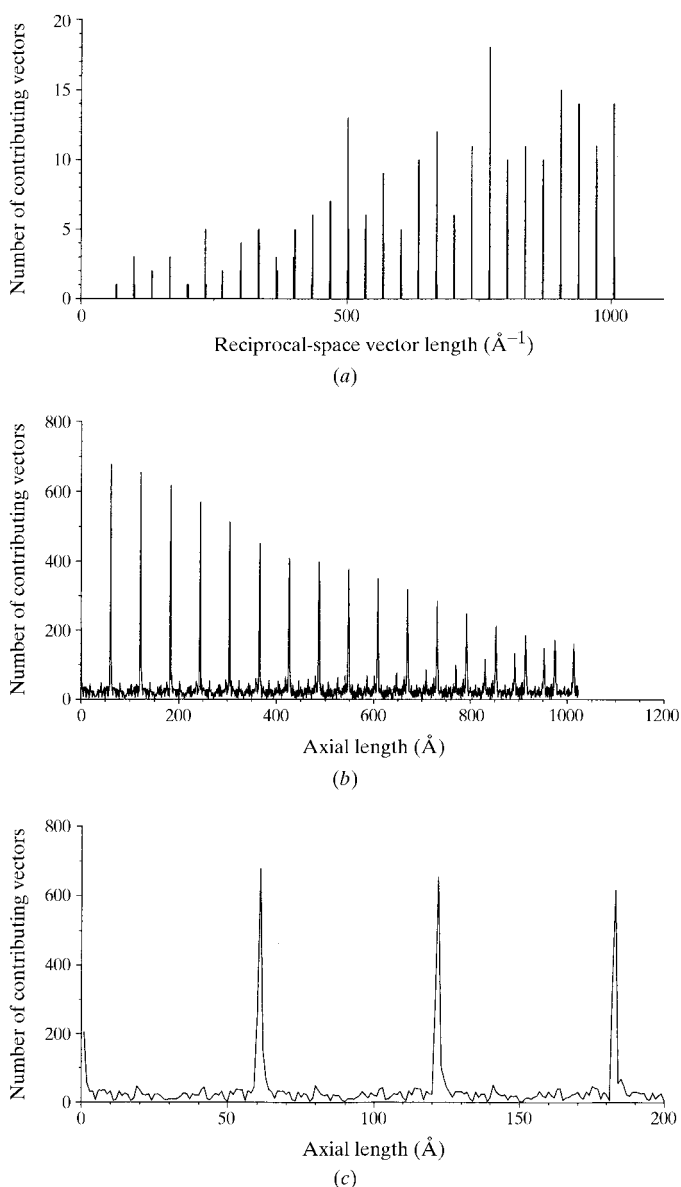


Figure 3
(a) Projection of three-dimensional reciprocal-space coordinates onto a direction corresponding to a crystal principal axis, calculated from the image in Fig. 2. (b) FFT calculated from this projection. (c) Expanded view of (b) showing the largest maximum at $\sim 62 \text{\AA}$ and subsequent repeats.

existing peak-search routine in *MOSFLM* is used to find reflections on either a single image or on several images; a subset of these is selected *via* an $I/\sigma(I)$ cutoff. In most cases, the program makes sensible choices about the values of parameters. The *DPS* method works most reliably with 200+ spots, but can index in some favourable cases with as few as 25.

The reciprocal-lattice coordinates can be considered to be vectors between the origin and each of the reflections, and these are projected onto ~ 7300 directions around a hemisphere of reciprocal space (Figs. 3*a* and 4*a*). Using spherical polar coordinates, going from the 'north pole' to the 'equator', ψ ranges from 0 to $\pi/2$ in steps of 0.03 rad, and around the equator φ ranges from 0 to 2π , keeping the increment on the line of latitude roughly equal; the number of steps is calculated from $(2\pi\sin\psi)/0.03$, *i.e.* no projections at the 'north pole' to ~ 212 increments at the 'equator' (depending on whether one rounds up or down to the nearest integer), increasing gradually as one moves further 'south' (Fig. 5). These numbers were chosen to provide a compromise between the time taken to calculate the projections and FFTs and the time taken in refining the vectors. It can plainly be seen that more useful information is contained in the projection onto a reciprocal-

axis vector than for a randomly chosen direction (Figs. 3*b* and 4*b*); similarly, it is obvious without close inspection of the FFTs that one corresponding to a real-space axis contains information that can be extracted without difficulty, while another calculated from the randomly chosen direction appears to be essentially random noise. The calculation of the FFT takes us from a frequency to a distance domain in real space.

Although to the human eye there is a prominent period in the FFT (Fig. 3*b*), it is important to remember that this indexing method does not use the periodic information directly. Only the largest vector shorter than the longest axial length chosen by the user in each FFT is used. The 30 FFTs with the largest maxima are chosen and sorted into decreasing order of the maxima and each is refined to its maximum value. Large peaks which are separated by a low background level only occur in the FFTs when the reciprocal-lattice vector is projected onto a reciprocal-lattice axis.

Three non-linear vectors at a time are chosen from this group of 30 to give a basis set of a direct-space primitive unit cell; the corresponding projections in reciprocal space give the nine components of the $[\mathbf{A}]$ or $[\mathbf{UB}]$ matrix. This choice is repeated for different combinations of the real-space vectors and a test is applied to determine the deviation from integer values of the indices calculated from (4); the solution with the smallest number of reflections which have deviations on any one axis greater than a chosen cut-off (0.2 in this case) is chosen and passed back to *MOSFLM*.

$$\mathbf{h} = [\mathbf{A}]^{-1}[\Phi]^{-1}\mathbf{x}, \text{ where} \quad (4)$$

\mathbf{h} = approximate Miller indices of a reflection with reciprocal lattice constants \mathbf{x} ,

$[\Phi]$ = matrix for rotation of φ around the crystal oscillation axis.

The Bravais lattice routines are then called from *MOSFLM*; solutions for each of the 44 lattice characters are calculated from the basis triclinic lattice in order to provide a list of possible solutions of higher symmetry together with their discrepancy indices.

4. *DPS* indexing in *MOSFLM*

In the implementation of the *DPS* routines in *MOSFLM*, reflections from several images may be included in the indexing process, which was not an option in the original ADSC implementation of the *DPS* code; it was found that with some test data sets, any image which contributed a disproportionately large number of reflections tended to dominate the indexing process. In the case of *MOSFLM*, it was decided that this choice is best left to the user; experience has shown that the inclusion of multiple image data can lead to successful indexing where other methods have failed (Williams, 1999). No facility is provided at present in the *DPS* routines for the crystallographer to supply a known unit cell.

The autoindexing is run from the X-windows interface to *MOSFLM* (Campbell, 1995; Meyer, 1998). The user is given a

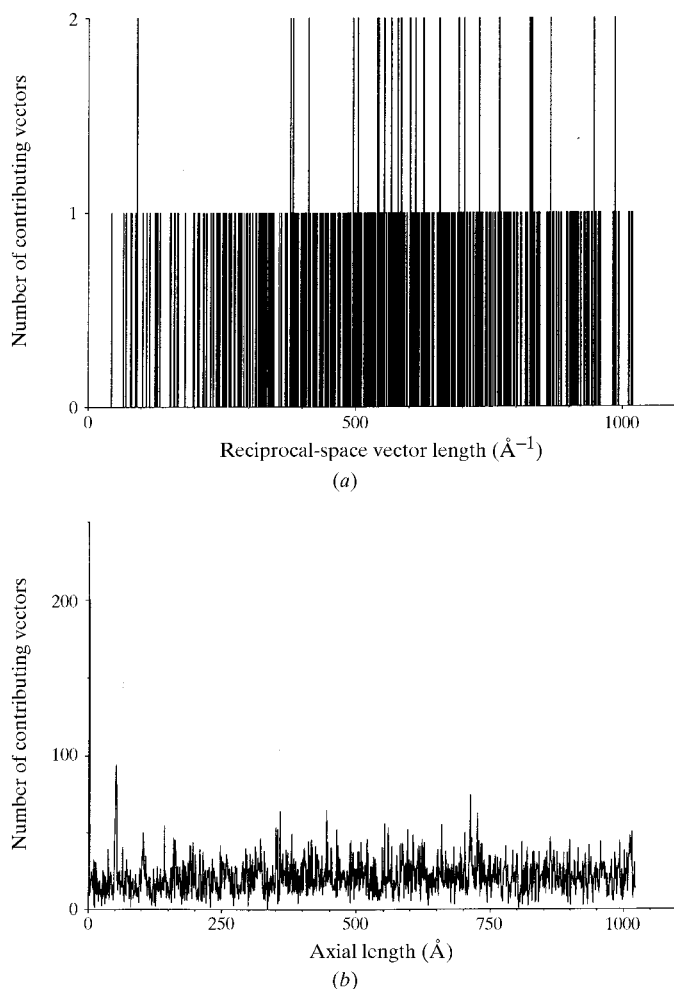


Figure 4

(*a*) Projection of three-dimensional reciprocal-space coordinates onto a randomly chosen direction. (*b*) FFT calculated from this projection.

choice of whether to use the new *DPS* or the original *REFIX* method, and in the case of the *DPS* method the user has to suggest a maximum likely unit-cell edge (a value of 200 Å is given as a default). The program prints the number of reflections used in the indexing to the output 'popup' window and indexing proceeds.

On a 233 MHz Pentium PC running Linux, the auto-indexing process takes around 20 s with ~500 reflections. The results are sorted with respect to increasing penalty and those with penalties less than 200 are written to a new popup window (all the solutions are output to the launch window and session log file); experience indicates that values greater than ~20 are unlikely to be correct and there is normally a sharp cutoff between the largest 'correct' and the smallest 'incorrect' result. It is usually good practice in the first instance to choose a result which balances the highest possible symmetry with the lowest possible penalty, unless prior knowledge has indicated otherwise.

In normal circumstances, single images from data sets result in an [A] matrix which is sufficiently correct to proceed with data integration. Occasionally, there are one or two images which do not index immediately and these are the ones which would be expected to be difficult for a single-image method, *i.e.* those that have one shorter crystal axis virtually parallel with the X-ray beam. In this case, there is a paucity of information about this third axis and it should not be surprising that difficulty arises in indexing (although images which fit this pattern can occasionally be indexed, *e.g.* Fig. 6). Indexing from another image or images from the data set may well overcome this problem.

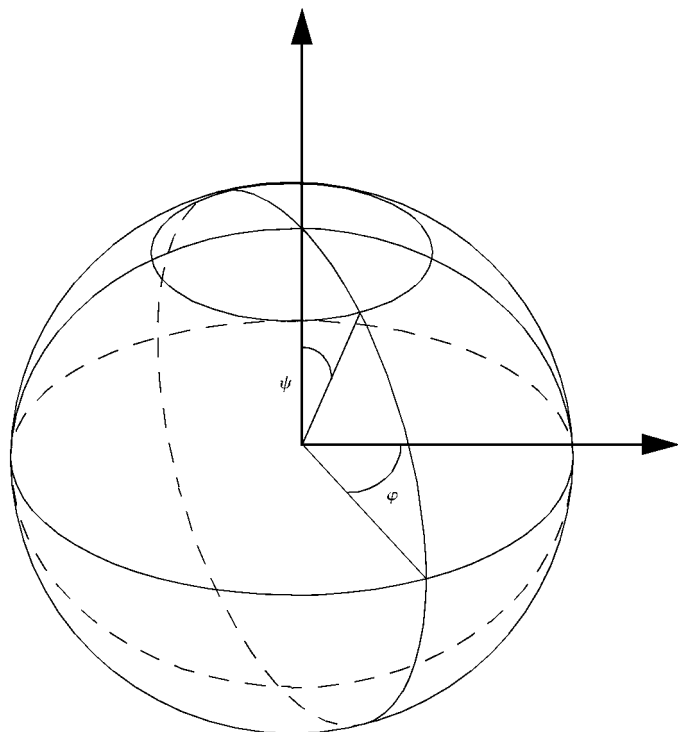


Figure 5
Spherical polar coordinates used in the *DPS* projection generation. $\Delta\psi = 0.03 \text{ rad} \approx 1.7^\circ$, $N\Delta\phi = 2\pi \sin\psi/0.03$.

For most images, if >100 reflections have been chosen and the maximum unit-cell edge has been chosen sensibly, failure is very rare. However, if too few reflections are used the routine may fail. Fortunately, there are several ways to overcome this problem: (i) reduce the $I/\sigma(I)$ cutoff at the indexing stage, so that more reflections which have been found are passed to indexing, (ii) modify the peak-search criteria so that there are more reflections available from the current image and (iii) include reflections from other image(s).

Typically, a wildly aberrant estimate in the longest unit-cell edge does not result in failure (a fourfold overestimate, *e.g.* 800 Å for a real 200 Å axis, has allowed the test image to be indexed), but if the unit-cell dimensions are very anisotropic, the short edge(s) may not index correctly.

A more difficult case arises if the crystal is excessively mosaic: if the mosaicity and beam-divergence parameters sum to several degrees, then the assumption that a still image is being analysed, with its true φ value close to the mean of the starting and finishing oscillation angles, becomes invalid. It may happen that the unit-cell dimensions can be determined with some accuracy, but it is increasingly likely with rising mosaicity that the orientation of the crystal cannot be determined. This would be true of any indexing method which relies on this initial assumption.

5. Distribution

The current version of *MOSFLM* with the *DPS* indexing routine is available *via* anonymous ftp from <ftp://ftp.mrc-lmb.cam.ac.uk/pub/mosflm>. The routines for building *MOSFLM*

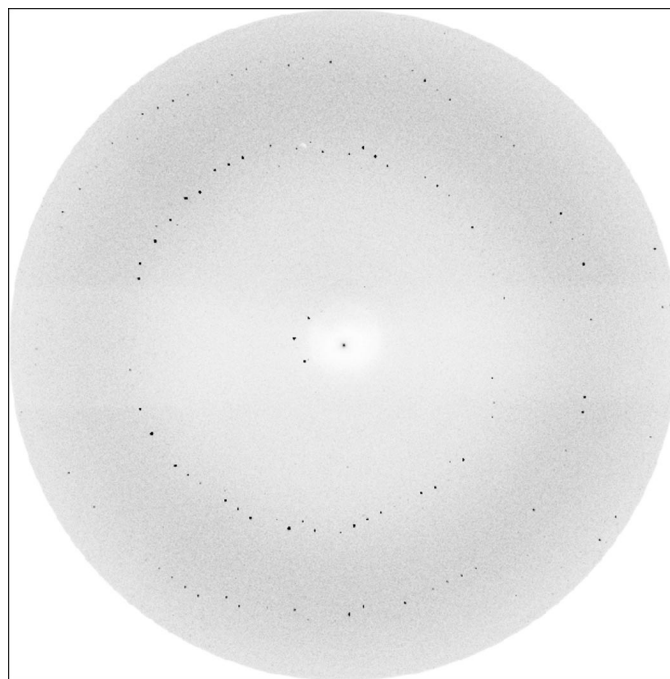


Figure 6
Diffraction image of a crystal of amphiphysin 2 SH3 domain. A short real-space axis is nearly parallel with the X-ray beam, so there is very little information about this axis. The image was indexed successfully with only 51 reflections.

have been streamlined to allow a straightforward build using standard Fortran and C compilers on Intel PCs (RedHat Linux 5.1), SGI (Irix 5.3, 6.2, 6.4 and 6.5) and Digital Alpha (Digital UNIX 4.0). Pre-built executable images have also been prepared for these platforms as well as for PowerPC computers running LinuxPPC Release 4 (e.g. Apple Macintosh PPC); these include both the new *DPS* and the existing *REFIX* (Kabsch, 1993) indexing methods.

My thanks go to Professor Michael Rossmann and Dr Ingo Steller for producing the *DPS* code and making it available, and to CCP4 for funding this work.

References

- Bricogne, G. (1986). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software (Phase III)*, p. 28. Paris: LURE.
- Campbell, J. W. (1995). *J. Appl. Cryst.* **28**, 236–242.
- Campbell, J. W. (1998). *J. Appl. Cryst.* **31**, 407–413 (see also <http://www.dl.ac.uk/CCP/CCP4/newsletter33/campbell.html>).
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Duisenberg, A. J. M. (1992). *J. Appl. Cryst.* **25**, 92–96.
- Higashi, T. (1990). *J. Appl. Cryst.* **23**, 253–257.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kim, S. (1989). *J. Appl. Cryst.* **22**, 53–60.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACMB Newslett. Protein Crystallogr.* **26**.
- Macíček, J. & Yordanov, A. (1992). *J. Appl. Cryst.* **25**, 73–80.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- Meyer, J. (1998). Personal communication.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rossmann, M. G. & Van Beek, C. G. (1999). *Acta Cryst.* **D55**, 1631–1640.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *J. Appl. Cryst.* **30**, 1036–1040.
- Williams, R. L. (1999). Personal communication.