Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants

By Wolfgang Kabsch

Max-Planck-Institut für medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, 69028 Heidelberg, Germany

Abstract

An algorithm has been developed for the automatic interpretation of a given set of observed reciprocal-lattice points. It extracts a reduced cell and assigns indices to each reflection by a graph-theoretical implementation of the local indexing method. All possible symmetries of the observed lattice compatible with the metric of the reduced cell are recognized and reported, together with the unit-cell constants and the linear index transformation relating the conventional to the reduced cell. This algorithm has been incorporated into the program *XDS* [Kabsch (1988). *J. Appl. Cryst.* **21**, 916–924], which is now able to process single-crystal area-detector data without prior knowledge of the symmetry and the unit-cell constants.

Introduction

Determination of space-group symmetry and unit-cell geometry is one of the first steps in the crystal-lattice structure analysis of macromolecules and is usually done before any diffraction data sets are collected. Often, this information is obtained by using a precession camera, which permits the recording of undistorted pictures of planes in the reciprocal lattice of the crystal. A great advantage of this method is that symmetry and metrical parameters of the reciprocal lattice are directly visible on the pictures and can be easily measured.

Occasionally, only very few crystals are available and in this situation it would be highly desirable to use all crystals for data collection rather than for precession photographs. Duisenberg (1992) has described a general indexing method that is applicable even to twin lattices. It determines, from the observed reciprocal-lattice points, a number of direct-lattice vectors from which primitive cells are extracted in a subsequent step. His method has been implemented in the semiautomatic program *DIRAX* and successfully used for data collection.

Unlike the method of Duisenberg (1992), the approach described in this paper cannot handle twin crystals. It has been developed for the fully automatic processing of rotation diffraction data from protein single crystals of initially unknown symmetry and cell constants. This is carried out in three steps. In the first step, a reduced cell is extracted from a number of diffraction spots, which are located automatically in the first few rotation pictures of the data set. In the second step, indices are assigned to these spots with respect to the reduced cell, using a graph-theoretical implementation of the local indexing method. In the third step, information about possible space-group symmetries is derived from the metric of the observed reduced cell and reported together with the unit-cell constants and the linear index transformation relating the conventional to the reduced cell in each case. This information may be used by the crystallographer for deriving the correct space group after all rotation pictures have been processed and integrated intensities are available. Depending on the completeness and redundancy of the collected data, in most cases a unique conclusion is obtained.

Basis extraction from a given lattice

It is assumed that a list is available containing the detector coordinates and spindle positions from a number of strong reflections recorded by the rotation method. As described in an earlier paper (Kabsch, 1988*a*), reciprocal-lattice vectors corresponding to each spot can be computed from knowledge of the detector position and the incident-beam wave vector. The problem is to determine a set of basis vectors of the lattice underlying the observed diffraction spots. A robust solution has been found that takes into account the moderate accuracy of the automatically determined lattice points and tolerates a small number of artefacts among them. The basis extraction from the given lattice is carried out in three steps.

The first step consists of reducing the list of given reciprocal-lattice points, which usually contains several hundred entries, to a small number of low-resolution difference-vector clusters (Kabsch, 1988*a*). From the given list p_i^* ($i = 1, \ldots, n$) of observed reciprocal-lattice vectors, all low-resolution

differences $\mathbf{p}_i^* - \mathbf{p}_j^*$ are formed and accumulated in a three-dimensional histogram. As expected, the low-resolution difference vectors cluster nicely in the histogram. The most densely populated clusters are located by an automatic spot-finding procedure and a list of m difference-vector clusters $\mathbf{v}_\mu^*$ ($\mu = 1, \ldots, m$) is obtained. The population f_μ of a difference-vector cluster $\mathbf{v}_\mu^*$ is defined as the number of times the difference between any two reciprocal-lattice vectors $\mathbf{p}_i^* - \mathbf{p}_j^*$ happens to be approximately equal to $\mathbf{v}_\mu^*$.

In the second step, a best set of three linear independent vectors $\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*$ is selected, from the list $\mathbf{v}_\mu^*$ ($\mu = 1, \ldots, m$), that maximizes the function Q defined by equations (1)–(5).

$$Q(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*) = \max_{\substack{1 \le \mu_1 < \mu_2 < \mu_3 \le m \\ \mathbf{v}_{\mu_1}^* \cdot (\mathbf{v}_{\mu_2}^* \times \mathbf{v}_{\mu_3}^*) \ne 0}} Q(\mathbf{v}_{\mu_1}^*, \mathbf{v}_{\mu_2}^*, \mathbf{v}_{\mu_3}^*) \quad (1)$$

$$Q(\mathbf{v}_{\mu_1}^*, \mathbf{v}_{\mu_2}^*, \mathbf{v}_{\mu_3}^*) = \sum_{\mu=1}^{m} f_\mu q(\xi_1^\mu, \xi_2^\mu, \xi_3^\mu) \quad (2)$$

$$q(\xi_1^\mu, \xi_2^\mu, \xi_3^\mu) = \exp\left(-2 \sum_{k=1}^{3} \{ [\max\,(|\xi_k^\mu - h_k^\mu| - \varepsilon, 0)/\varepsilon]^2 \right.$$
$$\left. + [\max\,(|h_k^\mu| - \delta, 0)]^2 \} \right) \quad (3)$$

$$\mathbf{v}_{\mu_k} \cdot \mathbf{v}_{\mu_l}^* = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$\xi_k^\mu = \mathbf{v}_\mu^* \cdot \mathbf{v}_{\mu_k} \qquad \mathbf{v}_\mu^* = \sum_{k=1}^{3} \xi_k^\mu \mathbf{v}_{\mu_k}^*,$$

$$h_k^\mu = \text{nearest integer to } \xi_k^\mu. \quad (5)$$

The purpose of this algorithm is to determine a reciprocal-lattice basis that explains as many as possible of the differences between the observed reciprocal-lattice vectors by nearly integral indices of small magnitude with respect to this basis. The absolute maximum of Q is assumed if all difference vectors can be expressed as integral multiples of the best triplet. Unfortunately, this ideal situation never occurs in practice and, for this reason, a quality measure q for the assigned indices h_k^μ is used as defined in (3). The continuous function q, which contains the two positive parameters ε and δ, was constructed such that the value of q sharply drops from its maximum of one if the expansion coefficients ξ_k^μ deviate by more than ε from their nearest integers h_k^μ or if the indices are absolutely larger than δ. The constraint on the allowed range of indices prevents the selection of a spurious triplet of very short difference-vector clusters, which might be present in the set. In cases of rather skew axes $\mathbf{v}_{\mu_k}^*$ and large deviations of $\mathbf{v}_\mu^*$ from the lattice, the indices h_k^μ as defined by (4) may not represent the nearest lattice point. Even then, the difference-vector cluster $\mathbf{v}_\mu^*$ will be correctly treated as unreliable by a small value of the quality measure q

as long as a small value for ε is used. Excellent results have been obtained with $\varepsilon = 0.05$ and $\delta = 5$. The specific functional form of q as defined by (3) is arbitrary to a certain extent and could probably be replaced by any other one with similar properties.

The final solution of the problem is obtained in the third step. The best vector triplet found in the previous step is refined by minimizing the residual

$$E(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*) = \sum_{\mu=1}^{m} q_\mu \left(\mathbf{v}_\mu^* - \sum_{k=1}^{3} h_k^\mu \mathbf{b}_k^* \right)^2, \quad (6)$$

where q_μ and h_k^μ are computed according to (3), (4) and (5) using the initial best vector triplet. A reduced cell as defined by Buerger (1957) is derived from the refined vector triplet. It consists of the shortest three linear independent real-space lattice vectors forming a right-handed set. In most cases, this reduced-vector triplet leads to nearly integral indices for all difference-vector clusters. However, occasionally it happens that some indices turn out to be very close to half-integers. Such a difference-vector cluster is used to modify the reduced-vector triplet. Let Δ_μ^* be the residual vector after subtraction of the integral indices from the difference-vector cluster $\mathbf{v}_\mu^*$ and $\mathbf{b}^*$ the ideal residual vector with exact half-integral indices as defined in (7):

$$\Delta_\mu^* = \sum_{k=1}^{3} \delta_k^\mu \mathbf{b}_k^*, \quad \mathbf{b}^* = \sum_{k=1}^{3} (\tfrac{1}{2}) d_k^\mu \mathbf{b}_k^*, \quad (7)$$

$$\delta_k^\mu = \xi_k^\mu - h_k^\mu, \quad d_k^\mu = \text{nearest integer to } 2\delta_k^\mu.$$

A new vector triplet is formed by replacing exactly one of the vectors $\mathbf{b}_k^*$ by $\mathbf{b}^*$ if $|d_k^\mu| = 1$ and Buerger-reduced in a subsequent step. All difference-vector clusters with integral indices with respect to the old base and the vector with half-integral indices now have integral indices with respect to the new reduced base.

The procedure described here has, thus, reduced the long list of n given reciprocal-lattice vectors $\mathbf{p}_i^*$ to three linear independent basis vectors. Despite its simplicity, it has been found to produce reliable results automatically as long as the diffraction pattern is based on one dominant lattice. It tolerates a small number of artefacts among the input list of observed reflections. However, it cannot handle twin lattices and in these cases the method of Duisenberg (1992) seems to be more appropriate.

Lattice indexing

Once a basis of the lattice is available, integral indices $\mathbf{h}^i$ must be assigned to each reciprocal-lattice vector $\mathbf{p}_i^*$ ($i = 1, \ldots, n$). The simple method of assigning the nearest integers of $\mathbf{p}_i^* \cdot \mathbf{b}_k$ ($k = 1, 2, 3$) as indices of $\mathbf{p}_i^*$ could easily lead to a misindexing of the larger

reciprocal-lattice vectors because of inaccuracies in the basis vectors. A solution of this problem is provided by the local indexing method described in an earlier paper (Kabsch, 1988a). The idea, summarized in (8) and (9),

$$H(\mathbf{h}^1, \ldots, \mathbf{h}^n) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(\mathbf{h}^i - \mathbf{h}^j - \mathbf{h}^{ij})^2,$$

$$w_{ij} = \begin{cases} 1 & \text{if } |h_k^{ij}| < \delta; \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

$$\xi_k^{ij} = (\mathbf{p}_i^* - \mathbf{p}_j^*) \cdot \mathbf{b}_k,$$

$$h_k^{ij} = \text{nearest integer to } \xi_k^{ij}, k = 1, 2, 3, \quad (9)$$

is to use only small index differences $\mathbf{h}^{ij}$ between pairs of neighbouring lattice vectors. The indices $\mathbf{h}^i$ of the lattice vectors are found by the solution of a linear system of equations obtained by minimizing the expression H. For large n, this earlier implementation of the idea of local indexing is clearly impractical.

Fortunately, an alternative implementation has been found that is highly efficient, with respect to both storage utilization and computing time. The reciprocal-lattice points can be considered as nodes of a tree. The tree connects the n points to each other with the connections as its branches. The length of a possible branch between nodes i and j is given as

$$l_{ij} = 1 - \exp\left(-2 \sum_{k=1}^{3} \{[\max(|\xi_k^{ij} - h_k^{ij}| - \varepsilon, 0)/\varepsilon]^2 \right.$$

$$\left. + [\max(|h_k^{ij}| - \delta, 0)]^2\} \right), \quad (10)$$

which is zero if none of the indices h_k^{ij} is absolutely larger than δ and the ξ_k^{ij} are integer values to within ε. If the length of a tree is defined as the sum of the lengths of its branches, the shortest of all possible n^{n-2} trees is determined by the elegant algorithm described by Dijkstra (1976). Starting with arbitrary indices 0, 0, 0 for the root node, the local indexing method consists of then traversing the shortest tree, thereby assigning to each node visited the indices of its predecessor plus the small index differences between the two corresponding reciprocal-lattice points. Bad points are recognized by large values of the lengths of their connecting branches and removed from the list. Finally, a constant offset is determined and added to the indices such that the centroids of the good observed reciprocal-lattice points $\mathbf{p}_i^*$ and their corresponding grid vectors $\sum_{k=1}^{3} h_k^i \mathbf{b}_k^*$ are as close as possible.

Bravais-lattice determination

In the context of data collection by a four-circle diffractometer, Katayama (1986) has described a

method to find the Bravais lattice that uses Delaunay's classification of a primitive cell into 24 types. The algorithm of Le Page (1982), which locates all twofold axes of the lattice and derives the crystal system from the number of these axes and their mutual angles, has been encoded in the program *LEPAGE* (Spek, 1988) and is also used in the area-detector software package *MADNES* (Messerschmidt & Pflugrath, 1987). For various reasons, these approaches have not been adopted here. The determination of the possible Bravais lattices is based upon the concept of the reduced cell whose metrical parameters characterize 44 lattice types as described in *International Tables for Crystallography* (1989). The reduced cell is defined there by a number of conditions (inequalities) that must be satisfied by the components of its metric tensor. The main conditions express the facts that the reduced cell consists of the shortest three lattice vectors sorted in increasing lengths with either all acute or all nonacute angles between them. The special conditions are imposed to resolve remaining ambiguities. The 44 lattice types are then characterized by additional equality relations among the six components of the reduced-cell metric tensor. Andrews & Bernstein (1988) describe unit cells as vectors in a Euclidean six-dimensional space and define each lattice type by its characteristic linear subspace. This formulation has been used to measure the similarity of a given six-dimensional cell vector with each lattice type: it is the length of the component perpendicular to the subspace corresponding to each lattice type (Paciorek & Bonin, 1992; Andrews & Bernstein, 1988).

Any primitive triclinic cell describing a given lattice can be converted into such a reduced cell (Křivý & Gruber, 1976; Andrews & Bernstein, 1988). It is well known, however, that the reduced cell derived is sensitive to experimental error and slightly different input vectors may result in a quite different reduced cell. Hence, the direct approach of first deriving the correct reduced cell and then finding the lattice type is unstable and may, in certain cases, even prevent the identification of the correct Bravais lattice.

Despite these difficulties, a stable procedure has been developed that identifies all lattice types compatible with an observed lattice basis. Stability is obtained by avoiding any decision as to what the 'true' reduced cell is. The essential ingredients of this procedure are: (a) a database of possible reduced cells; (b) a backward-search strategy that finds the best-fitting cell in the database for each lattice type.

The database is derived from a seed cell consisting of the three shortest linear independent observed lattice vectors sorted in order of increasing length. All cells of the same volume as the seed cell are included in the database, whose basis vectors can be linearly expressed in terms of the seed vectors by the

indices -1, 0 or $+1$. In addition, each of the three basis vectors of a cell must be of the same length – within an estimate of the experimental error – as the corresponding seed vector. Now, each unit cell in the database is considered as a potential reduced cell, although some of the defining conditions as given in ch. 9 in *International Tables for Crystallography* (1989) may be violated. These violations are treated as being due to experimental error.

The backward-search strategy starts with the hypothesis that the lattice type is already known and identifies the best-fitting unit cell in the database of possible reduced cells. Contrary to a forward-directed search, it can now always be decided which conditions for the components of the metric tensor of the reduced cell have to be satisfied. The total amount by which all these equality and inequality conditions are violated is used as a quality index. This measure is defined in (11) and (12) for lattice type 20 mC testing a potential reduced cell $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ from the database for agreement. If none of the conditions are violated, a quality index $p_{20} = 0$ will result. Positive values indicate that some conditions are not satisfied.

$$A = \mathbf{b}_1 \cdot \mathbf{b}_1, \quad B = \mathbf{b}_2 \cdot \mathbf{b}_2, \quad C = \mathbf{b}_3 \cdot \mathbf{b}_3,$$
$$D = \mathbf{b}_2 \cdot \mathbf{b}_3, \quad E = \mathbf{b}_1 \cdot \mathbf{b}_3, \quad F = \mathbf{b}_1 \cdot \mathbf{b}_2, \quad (11)$$

$$\begin{aligned}
p_{20}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = {} & \max(0, A - B) + \max(0, B - C) \\
& + \max(0, 2|D| - B) + \max(0, 2|E| - A) \\
& + \max(0, 2|F| - A) + \max(0, -D) \\
& + \max(0, -E) + \max(0, -F) \\
& + |B - C| + |E - F|. \quad (12)
\end{aligned}$$

All potential reduced cells in the database are tested and the smallest value for p_{20} is assigned to lattice type 20. This test is carried out for all 44 possible lattice types, using quality indices derived in a similar way from the defining conditions as listed in ch. 9 of *International Tables for Crystallography* (1989). For each of the 44 lattice types thus tested, the procedure described here returns the quality index, the conventional cell parameters and a transformation matrix relating the original indices of the observed lattice points to the new indices with respect to the conventional cell. A typical example is shown in Table 1. These index transformation matrices are derived from those given in Table 9.3.1 in *International Tables for Crystallography* (1989). (Note that the matrix for lattice type 17 mC given there is wrong: instead of $1\bar{1}0/110/\bar{1}01$ it should be $1\bar{1}0/\bar{1}10/\bar{1}01$.)

The quality index as defined in equations (11) and (12) for lattice type 20 is not particularly good because its contributing terms are not independent of each other. Despite these shortcomings, it clearly indicates lattice types that are in agreement with the observed

lattice points. At this stage, there is no automatic decision making; rather, decisions are made by the crystallographer after all rotation pictures have been processed and integrated intensities are available. A better measure has recently been proposed by Paciorek & Bonin (1992) but ultimately – as described in the next section – one would like to determine the best metrical parameters for each lattice type by minimizing the deviations between the calculated and observed diffraction geometry and use the residual as quality index.

Automatic processing of rotation diffraction records

The performance of the algorithm described above has been tested under a wide variety of conditions over a period of 18 months and has been found to produce reliable results. It has been added to two existing computer programs as described below.

The new version of the *REFIX* program (Kabsch, 1988a) processes a given list of reflection coordinates observed on one or several rotation pictures. It extracts a reduced cell, assigns indices to the given reflections and refines all parameters controlling the observed X-ray diffraction pattern. All 44 possible lattice types are tested for agreement with the metric of the reduced cell and the result is reported together with the estimated conventional cell constants. Table 1 shows the results obtained from a 1.5° oscillation data film, the same one that was used in the example described in the earlier paper (Kabsch, 1988a). The space group of the crystal is $C222_1$ and the cell constants are $a = 72.9$, $b = 100.1$ and $c = 92.6$ Å but this information was not available to the program. The results shown in Table 1 were deduced from a list of 1313 diffraction-spot coordinates with the spindle positions at diffraction set equal to the centre of the oscillation range of the film. The parameters of the reduced cell extracted from the given spots are reported in the entry for lattice type 44. Apparently, Bravais lattices mC (lattice types 10, 14), mP (type 34) and oC (type 13) and the triclinic types 31 and 44 are in agreement with the data. Note that this includes the correct orthorhombic Bravais lattice oC with derived cell constants close to the true ones. Obviously, a decision between orthorhombic and monoclinic space groups cannot be based on the geometry of the observed diffraction pattern alone; it requires a comparison of the integrated intensities of symmetry-equivalent reflections. According to Table 1, the correct reflection indices h_{13}, k_{13}, l_{13} for lattice type 13 oC are related to the reduced-cell indices h, k, l by the linear reindexing transformation $h_{13} = 1 \times h + 1 \times k + 0 \times l + 0$, $k_{13} = -1 \times h + 1 \times k + 0 \times l + 0$, $l_{13} = 0 \times h + 0 \times k + 1 \times l + 0$. The origin of the indices can be moved from its default $0, 0, 0$ to correct a possible misindexing of the reflections.

Table 1. *Results from the automatic processing of a 1.5° oscillation data film.*

The film was digitized at 100 μm raster. The space group is $C222_1$ and the cell constants are $a = 72.9$, $b = 100.1$ and $c = 92.6$ Å but this information was not made available to the program. The results shown were derived from 1313 spots located automatically.

	Lattice type	Quality index	Conventional cell constants						Reindexing transformation
			a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	
1	cF	999.0	119.3	137.3	119.1	121.1	77.5	122.6	11$\bar{1}$0/1$\bar{1}$10/$\bar{1}$$\bar{1}$$\bar{1}$0
2	hR	770.1	74.6	111.7	137.4	103.6	89.1	108.8	1100/$\bar{1}$0$\bar{1}$0/$\bar{1}$110
3	cP	769.7	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
5	cI	936.0	111.7	74.6	112.6	70.1	53.6	71.2	1010/1100/0110
4	hR	769.5	101.1	111.9	119.1	116.5	89.1	116.4	1$\bar{1}$00/$\bar{1}$0$\bar{1}$0/$\bar{1}$$\bar{1}$$\bar{1}$0
6	tI	999.0	112.6	111.7	74.6	71.2	70.1	53.6	0110/1010/1100
7	tI	999.0	111.7	74.6	112.6	70.1	53.6	71.2	1010/1100/0110
8	oI	999.0	74.6	111.7	112.6	53.6	70.1	71.2	$\bar{1}$100/$\bar{1}$0$\bar{1}$0/0$\bar{1}$$\bar{1}$0
9	hR	772.7	62.1	74.6	296.6	90.5	105.8	125.5	1000/$\bar{1}$$\bar{1}$00/$\bar{1}1\bar{3}$0
10	mC	24.0	101.1	74.6	92.9	90.1	90.0	91.3	1$\bar{1}$00/1100/0010
11	tP	174.8	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
12	hP	122.8	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
13	oC	23.8	74.6	101.1	92.9	90.0	90.1	88.7	1100/$\bar{1}$100/0010
15	tI	672.7	62.1	63.5	200.2	77.0	77.6	107.2	1000/0100/1120
16	oF	999.0	74.6	101.1	200.2	90.5	111.8	88.7	$\bar{1}$$\bar{1}$00/1$\bar{1}$00/1120
14	mC	23.4	74.6	101.1	92.9	90.0	90.1	88.7	1100/$\bar{1}$100/0010
17	mC	999.0	101.1	74.6	111.7	71.2	116.4	88.7	1$\bar{1}$00/$\bar{1}$$\bar{1}$00/$\bar{1}0\bar{1}$0
18	tI	999.0	112.6	119.1	62.1	68.7	99.5	115.4	01$\bar{1}$0/1110/1000
19	oI	999.0	62.1	112.6	119.1	64.6	68.7	80.5	$\bar{1}$000/01$\bar{1}$0/$\bar{1}$$\bar{1}$$\bar{1}$0
20	mC	746.3	112.6	112.6	62.1	99.5	99.6	111.3	0$\bar{1}$$\bar{1}$0/0$\bar{1}$10/$\bar{1}$000
21	tP	748.0	63.5	92.9	62.1	90.1	107.2	90.0	0100/0010/1000
22	hP	999.0	63.5	92.9	62.1	90.1	107.2	90.0	0100/0010/1000
23	oC	747.8	112.6	112.6	62.1	80.5	99.6	68.7	0110/0$\bar{1}$10/1000
24	hR	999.0	154.8	112.6	62.1	80.5	80.9	84.3	1210/0$\bar{1}$10/1000
25	mC	746.1	112.6	112.6	62.1	80.5	99.6	68.7	0110/0$\bar{1}$10/1000
26	oF	624.9	62.1	123.9	195.9	86.4	108.4	101.5	1000/$\bar{1}$$\bar{2}$00/$\bar{1}0\bar{2}$0
27	mC	499.7	123.9	62.1	112.6	80.5	119.7	78.5	$\bar{1}$$\bar{2}$00/$\bar{1}$000/01$\bar{1}$0
28	mC	325.0	62.1	195.9	63.5	95.4	107.2	71.6	$\bar{1}$000/$\bar{1}$0$\bar{2}$0/0$\bar{1}$00
29	mC	99.8	62.1	123.9	92.9	90.0	90.1	78.5	1000/1200/0010
30	mC	336.4	63.5	196.5	62.1	95.4	107.2	71.1	0$\bar{1}$00/0$\bar{1}$20/$\bar{1}$000
31	aP	0.2	62.1	63.5	92.9	90.0	89.9	72.8	1000/0$\bar{1}$00/00$\bar{1}$0
32	oP	152.0	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
40	oC	413.0	63.5	196.4	62.1	84.5	107.2	108.9	0$\bar{1}$00/0120/$\bar{1}$000
35	mP	151.8	63.5	62.1	92.9	90.1	90.0	107.2	0$\bar{1}$00/$\bar{1}$000/00$\bar{1}$0
36	oC	400.3	62.1	195.9	63.5	84.6	107.2	108.4	1000/$\bar{1}$0$\bar{2}$0/0100
33	mP	151.2	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
38	oC	100.1	62.1	123.9	92.9	90.0	90.1	101.5	$\bar{1}$000/1200/00$\bar{1}$0
34	mP	1.0	62.1	92.9	63.5	90.0	107.2	90.1	$\bar{1}$000/00$\bar{1}$0/0$\bar{1}$00
42	oI	661.3	62.1	63.5	200.2	103.0	102.4	107.2	$\bar{1}$000/0$\bar{1}$00/1120
41	mC	412.2	196.4	63.5	62.1	107.2	95.5	71.1	0$\bar{1}$20/0$\bar{1}$00/$\bar{1}$000
37	mC	400.1	195.9	62.1	63.5	107.2	95.4	71.6	1020/1000/0100
39	mC	99.9	123.9	62.1	92.9	90.1	90.0	78.5	$\bar{1}$$\bar{2}$00/$\bar{1}$000/00$\bar{1}$0
43	mI	999.0	74.6	200.2	63.5	103.0	127.3	68.2	1100/1120/0$\bar{1}$00
44	aP	0.0	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010

The program *XDS* (Kabsch, 1988b) has been developed for the reduction of single-crystal diffraction data from a sequence of adjacent rotation pictures recorded at a fixed X-ray wavelength by an electronic area detector. The new version is now able to process area-detector data sets automatically without knowledge of the space-group symmetry and the cell constants of the crystal. The necessary modifications fit smoothly into the original program structure. *XDS* is organized into eight steps (major subroutines), which are sequentially called by the main program. Information between the steps is exchanged by files that allow the repetition of selected steps with different sets of input parameters, without rerunning the whole program. A brief functional description of these eight individual steps is given below and includes the new program features.

XYCORR calculates a look-up table of spatial correlations at each detector pixel.

INIT estimates the initial background at each pixel and determines the trusted region of the detector surface.

COLSPOT locates strong diffraction spots occurring in the first few rotation pictures covering a total rotation range of 5°. When the space group and cell constants are unknown, a larger number of rotation

pictures may be included in the search and up to 3000 strong diffraction spots collected.

IDXREF extracts a reduced cell and indexes all spots found in the previous step by the method described in this paper. If the space group and cell constants are provided by the user, the reduced-cell vectors are reinterpreted accordingly; otherwise the crystal is described by its reduced-cell basis vectors and triclinic symmetry. In both cases, all 44 possible lattice types are rated and the results are reported for inspection by the crystallographer. A typical example is shown in Table 1. However, no automatic decisions are made by the program. Finally, all parameters controlling the diffraction experiment are refined to match the observed reflection positions as closely as possible.

COLPROF extracts from the rotation pictures individual three-dimensional profiles of all reflections, updates the background at each pixel and refines, in periodic intervals, the parameters controlling the diffraction experiment.

PROFIT estimates the intensity of each reflection from its three-dimensional profile.

CORRECT applies correction factors to the reflection intensities that partially compensate for radiation damage and absorption effects. After correction, *R*-factor statistics between symmetry-related reflections are printed, as well as the intensities of all reflections with indices of types $h00$, $0k0$ and $00l$, to help the crystallographer to find the correct space group. Different space groups can easily be tested by repeating *CORRECT* and *GLOREF* with the appropriate reindexing transformation and conventional cell constants found in the table printed in the *IDXREF* step. All reflections are reindexed and their corrected intensities and standard deviations are saved in a file.

GLOREF refines the diffraction parameters by using the observed positions of all strong spots found in the rotation pictures. It reports the root-mean-square error between calculated and observed positions along with the refined unit-cell constants. Again, for trying out possible space groups, the crystallographer consults the table printed in the *IDXREF* step and selects the appropriate reindexing transformation and starting values for the conventional cell constants.

References

ANDREWS, L. C. & BERNSTEIN, H. J. (1988). *Acta Cryst.* A44, 1009–1018.
BUERGER, M. J. (1957). *Z. Kristallogr.* 109, 42–60.
DIJKSTRA, E. W. (1976). *A Discipline of Programming*, pp. 154–167. New Jersey: Prentice-Hall.
DUISENBERG, A. J. M. (1992). *J. Appl. Cryst.* 25, 92–96.
International Tables for Crystallography (1989). Vol. A, pp. 738–749. Dordrecht: Kluwer Academic Publishers.
KABSCH, W. (1988*a*). *J. Appl. Cryst.* 21, 67–71.
KABSCH, W. (1988*b*). *J. Appl. Cryst.* 21, 916–924.
KATAYAMA, C. (1986). *J. Appl. Cryst.* 19, 69–72.
KŘIVÝ, I. & GRUBER, B. (1976). *Acta Cryst.* A32, 297–298.
LE PAGE, Y. (1982). *J. Appl. Cryst.* 15, 255–259.
MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* 20, 306–315.
PACIOREK, W. A. & BONIN, M. (1992). *J. Appl. Cryst.* 25, 632–637.
SPEK, A. L. (1988). *J. Appl. Cryst.* 21, 578–579.