

11.4. DENZO and SCALEPACK

BY Z. OTWINOWSKI AND W. MINOR

11.4.1. Introduction

X-ray diffraction data analysis, performed by the *HKL* package (Otwinowski, 1993; Otwinowski & Minor, 1997) or similar programs (Rossmann, 1979; Howard *et al.*, 1985; Blum *et al.*, 1987; Bricogne, 1987; Howard *et al.*, 1987; Leslie, 1987; Messerschmidt & Pflugrath, 1987; Kabsch, 1988; Higashi, 1990; Sakabe, 1991), is used to obtain the following results:

- (1) estimates of structure factors and determination of the crystal symmetry;
- (2) estimates of the crystal unit-cell parameters;
- (3) error estimates of the structure factors and unit cell;
- (4) detector calibration; and
- (5) detection of hardware malfunctions.

Other results, like indexing of the diffraction pattern, are in most cases only intermediate steps to achieve the above goals. The *HKL* system and other programs also have tools to validate the results by self-consistency checks.

The fundamental stages of data analysis are:

- (1) visual inspection of the diffraction images;
- (2) (auto)indexing;
- (3) diffraction geometry refinement;
- (4) integration of the diffraction peaks;
- (5) conversion of the data to a common scale;
- (6) symmetry determination and merging of symmetry-related reflections; and
- (7) statistical summary and estimation of errors.

This order represents the natural flow of data reduction, but quite often these steps are repeated based on information obtained at a later stage.

The three basic questions in collecting diffraction data are:

- (1) *whether to collect*;
- (2) *what to collect*; and
- (3) *how to collect and analyse the data*.

These questions and steps (1)–(7) of data analysis are intimately intertwined.

Data analysis makes specific assumptions which the collected data must, or at least should, satisfy. However, the experimenter can verify whether the data satisfy those assumptions only by data analysis. This circular logic can be broken by an iterative process. On-line data analysis provides immediate feedback during data collection and can remove the guesswork about *whether*, *what* and *how* from the process. The description of data analysis and algorithms that follows will make frequent references to the assumptions about the data and offer guidelines on how to make the experiment fulfil these assumptions.

This article uses the *HKL* package coordinate system to describe data algorithms and analysis. However, as most equations are written in vector notation, they can be easily adapted to conventions used in other programs.

11.4.2. Diffraction from a perfect crystal lattice

X-ray photons can scatter from individual electrons by inelastic and incoherent processes. The coherent scattering by the whole crystal is called diffraction.* Energy conservation, when expressed in photon momentum vectors, is equivalent to

$$\mathbf{S} \cdot \mathbf{S}_0 = \frac{1}{2} \mathbf{S} \cdot \mathbf{S}, \quad (11.4.2.1)$$

where \mathbf{S} is the diffraction vector, defined as the change of photon momentum in the scattering process, and \mathbf{S}_0 is the vector which has beam direction and length $1/\lambda$. Diffraction from a perfect crystal lattice occurs when diffraction from all repeating crystal elements is in phase, which can be stated in vector algebra as

$$\mathbf{S} \cdot \mathbf{a} = h \quad (11.4.2.2)$$

$$\mathbf{S} \cdot \mathbf{b} = k \quad (11.4.2.3)$$

$$\mathbf{S} \cdot \mathbf{c} = l. \quad (11.4.2.4)$$

In shorter notation, these may be written as $\mathbf{h} = [A]\mathbf{S}$, which is equivalent to $\mathbf{S} = [A]^{-1}\mathbf{h}$, where $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are the real-space crystal periodicity vectors,

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

and h, k, l are the integer Miller indices. Often, the orientation matrix is defined in reciprocal space as the inverse of $[A]$.

The condition for crystal diffraction with Miller indices h, k, l is the existence of a (unique) vector \mathbf{S} which is a solution to equations (11.4.2.1)–(11.4.2.4). Equation (11.4.2.1) states the diffraction condition for vector \mathbf{S} . Mathematically speaking, the space of the solutions to equations (11.4.2.2)–(11.4.2.4) is called reciprocal space, and vector \mathbf{S} belongs to this space. However, the following presentation does not depend on the properties of reciprocal space. The laboratory coordinate system used has its origin at the position of the crystal. A diffraction peak at the detector position in three-dimensional laboratory space $\mathbf{X} = \{x, y, z\}$ corresponds to vector \mathbf{S} :

$$\mathbf{S} = \mathbf{X}/\lambda|\mathbf{X}| + \mathbf{S}_0. \quad (11.4.2.5)$$

Rotation of the crystal around the goniostat axes can be described by vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in equations (11.4.2.2)–(11.4.2.4) as a function of the goniostat angles, and vectors $\mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0$ represent the crystal orientation at the zero position of the goniostat. These rotations are described by Bricogne (1987):

$$\mathbf{a} = [R_1(\varphi_1)][R_2(\varphi_2)][R_3(\varphi_3)]\mathbf{a}_0, \quad (11.4.2.6)$$

where

$$[R(\varphi)] = \cos(\varphi) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + [1 - \cos(\varphi)] \begin{pmatrix} e_x e_x & e_x e_y & e_x e_z \\ e_x e_y & e_y e_y & e_y e_z \\ e_x e_z & e_y e_z & e_z e_z \end{pmatrix} \\ + \sin(\varphi) \begin{pmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{pmatrix}, \quad (11.4.2.7)$$

where e_x, e_y, e_z represent the direction cosines of a rotation axis. To complete the description of diffraction geometry, we need a function $X(p, q)$, describing the position in experimental space of each pixel with integer coordinates $\{p, q\}$. This function is detector-specific and describes the detector geometry and distortion. For a planar detector,

$$\mathbf{X}(p, q) = [R_x][R_y][R_z]([R_{2\theta}]([L](K[D(p, q)] - \mathbf{B}) + \mathbf{T}_D) - \mathbf{T}_D) + \mathbf{T}_D, \quad (11.4.2.8)$$

where R_x, R_y, R_z represent the detector misorientation, $R_{2\theta}$ represents rotation around the 2θ (swing) axis, \mathbf{T}_D is the detector

* Owing to the large difference in mass between the crystal and the photon, the energy of the photon is virtually unchanged.

translation from the crystal, operation L represents the axis naming/direction convention used by the detector manufacturer (eight possibilities), K is an operation scaling pixels to millimetres, D is a detector distortion function and \mathbf{B} represents the beam position on the detector surface.

Equations (11.4.2.1)–(11.4.2.8) fully describe the existence and position of the diffraction peaks, which is all that is needed for the autoindexing procedure.

11.4.3. Autoindexing

Among the number of autoindexing algorithms proposed (Vriend & Rossmann, 1987; Kabsch, 1988; Kim, 1989; Higashi, 1990; Leslie, 1993), the method based on periodicity of the reciprocal lattice tends to be the most reliable (Otwinowski & Minor, 1997; Steller *et al.*, 1997).

Autoindexing starts with a peak search, which results in the set of $\{p, q, i\}$ triplets, where i is the number of the image in which the peak with position $\{p, q\}$ was found. The program takes advantage of the fact that for any rotation matrix

$$([R]\mathbf{S}) \cdot ([R]\mathbf{a}) = \mathbf{S} \cdot \mathbf{a}. \quad (11.4.3.1)$$

When

$$[R] = [R_3(-\varphi_3)][R_2(-\varphi_2)][R_1(-\varphi_1)], \quad (11.4.3.2)$$

equation (11.4.3.1) applied to equation (11.4.2.2) becomes:

$$([R]\mathbf{S}) \cdot \mathbf{a}_o = h, \quad (11.4.3.3)$$

where \mathbf{a}_o is a three-dimensional vector with as yet unknown components. Note that the matrix $[R]$ represents crystal rotation when the crystal is in the diffraction condition defined by the existence of the solution to equations (11.4.2.1)–(11.4.2.4), described by vector \mathbf{S} . For data collected in the wide oscillation mode* the angle at which diffraction occurs is not known *a priori*; however, it can be approximated by the middle of the oscillation range of the image. Combining the peak position $\{p, q\}$ with equations (11.4.2.5) and (11.4.2.8) provides an estimate of the vector \mathbf{S} . So, we expect that equation (11.4.3.3) and similar equations for k and l are approximately (owing to approximation and experimental errors) satisfied. The purpose of autoindexing is to determine the unknown vectors $\mathbf{a}_o, \mathbf{b}_o, \mathbf{c}_o$ and the $\{h, k, l\}$ triplet for each peak. To accomplish this, three equations (11.4.3.3) for each peak must be solved. *DENZO* introduced a method based on the observation that the maxima of the function

$$\sum_i \cos[2\pi([R]\mathbf{S}_i) \cdot \mathbf{a}_o] \quad (11.4.3.4)$$

are the approximate solutions to this set of equations (11.4.3.3). To speed up the search for all significant maxima, a two-step process is used. The first step is the search for maxima of function (11.4.3.3) on a three-dimensional uniform grid, made very fast owing to the use of a fast Fourier transform (FFT) to evaluate function (11.4.3.4). Function (11.4.3.4) is identical to structure-factor calculations in the space group $P1$, which allows the use of the crystallographic FFT. Because the maxima at the grid points (HKL uses a $96 \times 96 \times 96$ grid) only approximate the maxima of function (11.4.3.4), the vectors resulting from a grid search are optimized by the Newton method. Function (11.4.3.4) has maxima not only for basic periodic vectors $\mathbf{a}_o, \mathbf{b}_o$ and \mathbf{c}_o , but also for any integer linear combination of them. Any set of three such vectors with a minimal nonzero determinant can be used to describe the crystal lattice. Steller *et al.*

(1997) describe the algorithm that finds the most reliable set of three vectors. This set needs to be converted to the one conventionally used by crystallographers, as defined in *IT A* (1995).

To generate the conventional solution, two steps are used. Step 1 finds the reduced primitive triclinic cell. *IT A* provides the algorithm for this step. Subsequently, step 2 finds conventional cells in Bravais lattices of higher symmetry.

11.4.3.1. Lattice symmetry

The relationship between a higher-symmetry cell and the reduced primitive triclinic cell can be described by

$$[A] = [M][P], \quad (11.4.3.5)$$

where $[A]$ and $[P]$ are matrices of the type $\{a_0, b_0, c_0\}$, with $[P]$ representing the reduced triclinic primitive cell, and $[M]$ is one of the 44 matrices listed in *IT A*.† If $[A]$ is generated using equation (11.4.3.5) from an experimentally determined $[P]$, owing to experimental errors it will not exactly satisfy the symmetry restraints. *DENZO* introduced a novel index that helps evaluate the significance of this violation of symmetry. This index is based on the observation that from $[A]$ one can deduce the value of the unit cell, apply symmetry restraints to the unit cell and calculate any matrix $[A']$ for the unit cell that satisfies these symmetry restraints. If $[A]$ satisfies symmetry restraints, the matrix $[U]$, where

$$[U] = [A][A']^{-1}, \quad (11.4.3.6)$$

will be unitary and

$$[U]^T - [U]^{-1} = 0. \quad (11.4.3.7)$$

The index of distortion printed by *DENZO* is

$$\left\{ \sum_i \sum_j ([U]_{ij}^T - [U]_{ij}^{-1})^2 \right\}^{1/2} / 6, \quad (11.4.3.8)$$

where i and j are indices of the 3×3 matrix $[U]$.

The value of this index increases as additional symmetry restraints are imposed, starting from zero for a triclinic cell. Autoindexing in *DENZO* always finishes with a table of distortion indices for 14 possible Bravais lattices, but does not automatically make any lattice choice.

11.4.3.2. Lattice pseudosymmetry

The cell-reduction procedure cannot determine lattice symmetry, since it cannot distinguish true lattice symmetry from a lattice accidentally having higher symmetry within experimental error (e.g. a monoclinic lattice with $\beta \simeq 90^\circ$ is approximately orthorhombic). If one is not certain about the lattice symmetry, the safe choice is to assume space group $P1$, with a primitive triclinic lattice for the crystal, and to check the table again after the refinement of diffraction-geometry parameters. A reliable symmetry analysis can be done only by comparing intensities of symmetry-related reflections, which is done later in *SCALEPACK* or another scaling program.

11.4.3.3. Data-collection requirements

The total oscillation range has to cover a sufficient number of spots to establish periodicity of the diffraction pattern in three dimensions. It is important that the oscillation range of each image is small enough so that the lunes (rings of spots, all from one reciprocal plane) are resolved. One should note that the requirement

* This is when the crystal oscillation angle during the measurement of a single diffraction pattern is larger than the angular reflection width.

† It should be noted that No. 17 contains an error (Kabsch, 1993).

11. DATA PROCESSING

for lune separation is distinct from the requirement for spot separation. If lunes overlap, spots may have more than one index consistent with a particular position on the detector.

The autoindexing procedure described above is not dependent on prior knowledge of the crystal unit cell; however, for efficiency reasons, the search is restricted to a reasonable range of unit-cell dimensions, obtained, for example, from the requirement of spot separation. In *DENZO*, this default can be overridden by the keyword 'longest vector', but the need to use this keyword is a sign of a problem that should be fixed. Either the defined spot size should be decreased or data should be recollected with the detector further away from the crystal.

11.4.3.4. Misindexing

Autoindexing is sensitive to inaccuracy in the description of the detector geometry. The specified position of the beam on the detector should correspond to the origin of the Bragg-peaks lattice (Miller index 000). Autoindexing will shift the origin of the lattice to the nearest Bragg lattice point. An incorrect beam position will result in the nearest Bragg lattice point not having the index 000. In such a situation, all reflections will have incorrectly determined indices. Such misindexing can be totally self-consistent until the intensities of symmetry-related reflections are compared. This dependence of the indexing correctness on the assumed beam position is the main source of difficulties in indexing (Gewirth, 1996; Otwinowski & Minor, 1997). The beam position has to be precise, as the largest acceptable error is one half of the shortest distance between spots.

Indexing limited to determining h , k , l triplets is not very sensitive to other detector parameters. Errors by a degree or two in rotation or by 10% in distance are unlikely to produce wrong values of h , k and l . Sometimes even a very large error, such as the distance being too large by a factor of 5, will still produce the correct h , k , l triplets. The detector position error will be compensated by an error in the lattice determined by autoindexing. For this reason, the accuracy of the lattice is not a function of the autoindexing procedure, but depends mainly on the accuracy of the detector description. By the same token, the distortion of the lattice also depends on the accuracy of the detector parameters.

11.4.3.5. Twins

Special care has to be taken if more than one crystal contributes to the diffraction image. When there is a large disproportion between volumes (e.g. the presence of a satellite crystal), autoindexing may work without any modifications. In the case of similar volumes, the manual editing of weaker reflections and resolution cuts can make the proportion of reflections from one crystal in the peak-search list large enough for the autoindexing method to succeed. If the crystals have a similar orientation, using only very low resolution data may be the right method. In the case of twinned crystals, autoindexing sometimes finds a superlattice that results in integer indices simultaneously for both crystals. In such a case, *DENZO* solves the problem of finding the best three-dimensional lattice that incorporates all of the observed peaks. Unfortunately, for a twinned crystal, this is a mathematically correct solution to an incorrectly posed problem.

11.4.4. Coordinate systems

There are four natural coordinate systems used to describe a diffraction experiment, defined by the order in which the data are stored, the beam and gravity, or the beam and the goniostat axes

(spindle or 2θ). These coordinate systems will be called, respectively, *data*, *beam-gravity*, *beam-spindle* and *beam- 2θ* .

11.4.4.1. Beam-gravity

To visualize a diffraction pattern, beam-gravity is the coordinate system clearly preferred by human physiology. The universal preference to relate to the gravity direction is revealed by the observation that people generally perceive an image in a mirror as inverted left-right rather than top-down. Hence *XdisplayF* uses the beam-gravity coordinate system, except when diffraction data cannot be related to gravity.*

11.4.4.2. Data

The first (1983) *DENZO* implementation used the data coordinate system to describe the beam position on the detector and to define the integration box. This is still the case in order to keep backward compatibility.

11.4.4.3. Beam-spindle

Until 1998, *DENZO* supported only a single-axis goniostat and used a beam-spindle coordinate system to define crystal and detector orientation and polarization. Initially, the goniostat spindle axis was assumed to be horizontal, so the direction perpendicular to the beam and spindle was described by the keyword *vertical*, which in reality may not relate to the gravity direction for some goniostats. The keyword *rotx* relates to rotation around the spindle axis, *roty* around the *vertical axis* and *rotz* around the beam axis. The definition of the orientation matrix in the communication file between *DENZO* and *SCALEPACK* uses an unintuitive convention: the letter y in *roty* relates to the first element of the vector, x in *rotx* relates to the second and z in *rotz* to the third. However, the matrix always has a positive determinant, so this convention has no impact on the handedness of the coordinate system. This unfortunate choice of convention, preserved for backward compatibility reasons, appears only in the communication file and has no significance for anybody who does not inspect the matrix.

11.4.4.4. Beam- 2θ

The recent addition of a general goniostat introduced a conceptual change in the *DENZO* coordinate system. The data-collection axis can be oriented in any direction, so in principle *rotx*, *roty* and *rotz* no longer need to be defined relative to the data-collection axis. However, to keep the useful correlation between refinable parameters (*crystal rotz* and *detector rotz* being close to 100% correlated), one real and two virtual goniostats are used simultaneously in *DENZO*. Refinable crystal parameters (*crystal rotx*, *roty*, *rotz*) are still defined, as in the past, by the data-collection axis and the beam. This means that the directions of rotations defined by *fit crystal rotx*, *roty* and *rotz* do not rotate around the data-collection axis as the program advances from one image to another. This coordinate system changes with the change in direction of the data-collection axis. Crystal orientation is defined by three constant, perpendicular axes, which, in the current version, no longer have to be aligned with the physical crystal goniostat. However, the so-called *2 theta* rotation has a fixed axis, and, if it exists, it defines the *DENZO* coordinate system together with the beam axis. Thus the current coordinate system in *DENZO* should be called beam- 2θ . Fortunately for the user, the conversions between different coordinate systems are handled transparently. For example, the refined change in the crystal orientation is converted from the refined goniostat to the crystal-orientation goniostat. The

* There may occasionally be an exception to this when the experimental system is not known.

movements of the physical goniostat are converted into appropriate changes in the diffraction pattern. The physical goniostat appears only to describe the data collection and, optionally, to calculate the physical goniostat angles that achieve particular crystal alignments.

The *DENZO* coordinate system (Gewirth, 1996) is used in the definition of crystal goniostats, 2θ goniostat, Weissenberg coupling and polarization.

This discussion of the coordinate systems shows that the conceptual complexity of the program description does not result in complexity of the actual use of the program. The success of data analysis does not require a full understanding of the relations between internal *DENZO* goniostats and the coordinate systems. The reason for this complexity was to create a simple pattern of correlations between crystal and detector parameters in *DENZO* refinement. This in turn allows for simple and easy-to-understand control of the refinement process and simplifies problem diagnostics. For example: the definition of refined *crystal rotx* as rotation around the data-collection axis makes hardware problems when driving the spindle and shutter result only in fluctuations of *crystal rotx*. The constant nonzero value of the refined shifts between frames of *crystal roty* and *rotz* is a sign of misalignment of the data-collection axis. Although the program compensates for this misalignment with changes in crystal orientation, this introduces a small error in the Lorentz factor. The nature of these problems is such that they do not result in a complete failure of the experiment, but they do have an impact on the quality of the result. It is up to the experimenter and the instrument manager to assess the significance of these indications.

11.4.5. Experimental assumptions

To achieve the main target of a diffraction experiment – the estimation of structure factors – three components need to be determined, with maximum possible precision:

(1) the crystal response function (the relationship between the crystal structure factor and the number of diffracted X-ray photons, which depends also on the X-ray source characteristics);

(2) the detector response function; and

(3) the geometrical description of the detector relative to the directions of the X-ray beam and crystal goniostat axes.

The main difficulty of data analysis in protein crystallography is the complexity of the process that determines these components. *HKL* can determine all three directly from the data produced by the analogue-to-digital converter (ADC). The only extra program needed is one that sends the raw ADC signal to the computer disk. For charge-coupled-device (CCD) detectors, spatial detector distortion and sensitivity per pixel functions need to be established in a separate experiment. Usually it is worthwhile to establish a geometrical description of the detector in a separate diffraction experiment. A precise determination requires a well diffracting, high symmetry, non-slipping crystal and a special data-collection procedure.

11.4.5.1. Crystal diffraction

The crystal response function consists of two types of factors included in the analysis: additive factors, which are represented by the background, and a number of multiplicative factors, such as exposed crystal volume, overall and resolution-dependent decay, Lorentz factor, flux variation, polarization, *etc.* Other factors, like extinction and non-decay radiation damage (radiation damage can result not only in decay, but also in a change in the crystal lattice, often a main source of error in an experiment), are ignored by *HKL*, except for their contribution to error estimates.

11.4.5.2. Data model

The detector response function is the main component for the data model. *HKL* supports

(1) data stored in 8 or 16 bit fields;

(2) overflow table;

(3) linear, bilinear, polynomial and exponential response, with the error model represented by an arbitrary scale;

(4) saturation limit;

(5) value representing lack of data;

(6) constant offsets per read-out channel;

(7) pattern noise;

(8) lossless compression;

(9) flood-field response; and

(10) sensitivity response.

HKL supports most data formats, which represent particular combinations of the above features. The formats define the coordinate system, the pixel size, the detector size, the active area and the fundamental shape (cylindrical, spherical, flat rectangular or circular, single or multi-module) of the detector.

The main complexity of the data-analysis program and the difficulties in using it are not in application of the data model but rather in the determination of the unknown data-model parameters. The refinement of the data-model parameters is an order of magnitude more complex (in terms of the computer code) than the integration of the Bragg peaks when the parameters are known.

The data model is a compromise between an attempt to describe the measurement process precisely and the ability to find parameters describing this process. For example, the overlap between the Bragg peaks is typically ignored due to the complexity of spot-shape determination when reflections overlap. The issue is not only to implement the parameterization, but also to do it with acceptable speed and stability of the numerical algorithms. A more complex data model can be more precise (realistic) under specific circumstances, but can result in a less stable refinement and produce less precise final results in most cases. An apparently more realistic (complex) data model may end up being inferior to a simpler and more robust approach. The complexity of model-quality analysis is due to the fact that some types of errors may be much less significant than others. In particular, an error that changes the intensities of all reflections by the same factor only changes the overall scale factor between the data and the atomic model. Truncation of the integration area results in a systematic reduction of calculated reflection intensities. A variable integration area may result in a different fraction of a reflection being omitted for different reflections. The goal of an integration method is to minimize the variation in the omitted fraction, rather than its magnitude. Similarly, if there is an error in predicting reflection-profile shape, this constant error has a smaller impact than a variable error of the same magnitude.

The magnitude and types of errors are very different in different experiments. The compensation of errors also differs between experiments, making it hard to generalize about an optimal approach to data analysis when the data do not fully satisfy the assumptions of the data model. For intense reflections, when counting statistics are not a limiting factor, none of the current data models accounts for all reproducible errors in experiments. This issue is critical in measuring small differences originating from dispersive effects.

11.4.5.3. Data-model refinement

The parameters of the data model can be classified into four groups:

(1) Those refinable from self-consistency of the data by a (nonlinear) least-squares method.

11. DATA PROCESSING

(2) Parameters that can be determined from internal self-consistency of the data, but for which least squares is not implemented. For example, error-estimate parameters are in this category.

(3) Parameters that have to be established in a separate experiment, *e.g.* pixel sensitivity from flood-field exposure.

(4) Parameters that are obtained from hardware description.

The least-squares method is based on minimization of a function that is a sum of contributors of the following type:

$$(\text{pred} - \text{obs})^2 / \sigma^2 = \chi^2, \quad (11.4.5.1)$$

where *pred* is a prediction based on some parameterized model, *obs* is the value of this prediction's measurement and σ^2 is an estimate of the measurement and the prediction uncertainty. *DENZO* has the following least-squares refinements:

- (1) refinement of unit-cell vectors in autoindexing;
- (2) refinement of background and background slope; and
- (3) refinement of crystal orientation, unit cell, mosaicity, beam focus and position, detector orientation and position, and geometrical distortions that are parameterized differently for different detectors.

SCALEPACK can refine the following parameters by least-squares methods:

- (1) unit cell, crystal orientation and mosaicity, including changes of these parameters during an experiment;
- (2) goniostat internal alignment angles;
- (3) crystal absorption, using spherical harmonics (Katayama, 1986; Blessing, 1995) expansion of the absorption surface;
- (4) uniformity of exposure, including shutter timing error;
- (5) correction to the Lorentz factor resulting from a misalignment of the spindle axis;
- (6) reproducible wobble of the rotation axis resulting from a misalignment of gears in a spindle assembly;
- (7) non-uniform smooth detector response, for example, resulting from decay of the image-plate signal during scanning; and
- (8) other factors contributing to scaling resulting from a slow fluctuation of beam intensity, change in exposed volume, overall crystal decay and resolution-dependent crystal decay.

11.4.5.4. Correlation between parameters

Occasionally, the refinement can be unstable due to high correlation between some parameters. High correlation results in the errors in one parameter compensating for the errors in other parameters. In the case where compensation is 100%, the parameter would be undefined, but the error compensation by other parameters would make the predicted pattern correct. In such cases, eigenvalue filtering [related to singular value decomposition, described by Press *et al.* (1989) in *Numerical Recipes*] is employed to remove the most correlated components from the refinement to make it more stable. Eigenvalue filtering works reliably when starting parameters are close to the correct values, but may fail to correct large errors in the input parameters if the correlation is close to, but not exactly, 100%. Once the whole data set is integrated, global refinement [also called post refinement: Rossmann *et al.* (1979); Winkler *et al.* (1979); Evans (1987); Greenhough (1987); Evans (1993); Kabsch (1993)] can refine crystal parameters (unit cell and orientation) more precisely and without correlation with detector parameters. The unit cell used in structure-determination calculations should come from the global refinement (in *SCALEPACK*) and not from *DENZO* refinement.

11.4.5.5. Single- and multiframe refinement

The crystal and detector orientation parameters can be refined for each group of images or for each processed image separately.

Refinement performed separately for each image allows for robust data processing, even when the crystal slips considerably during data collection.

11.4.5.6. Active area

Not every pixel represents a valid measurement. Specification of the active detector area in *DENZO* is derived from the format and the definition of the detector size. Detector calibration with flood-field exposure will calculate the sensitivity for each pixel and will also determine which pixels should be ignored. The input command can additionally label some areas of the detector to be ignored, most frequently the shadow caused by the beam stop and its support. To define the shape of the area shadowed by the beam stop, the useful commands are *ignore circle* and *ignore quadrilateral*. There are also commands to ignore triangular shapes, margins of the detector and a particular line or pixel.

11.4.5.7. Flood field

The basic method for calibration of the spatial dependence of detector sensitivity is to measure the response to a flood-field exposure. The amount of relative exposure per pixel needs to be known. *DENZO* allows for either a uniform or an isotropic source. If the source is at the crystal position, *DENZO* refinement (with a separate crystal exposure) can be used to define the geometry of the source relative to the detector. To calculate the flood-field response, an earlier determination of the detector distortion is required. The flood-field response is converted to a sensitivity function. Large deviations from the local average are used to define inactive pixels. The edge of the active area needs special treatment, depending on the method of phosphorus deposition.

11.4.5.8. Absolute configuration

Absolute configuration is defined relative to the data-coordinate system and is only affected by the sign of the parameter *y scale*. A mirror transformation of the data does not affect the self-consistency of the data. Thus, the correctness of the absolute configuration cannot be verified by data-reduction programs.

11.4.5.9. Correcting diffraction images

HKL can also generate data corrected for the above factors and/or for geometrical conversion and distortion in uncompressed, lossless compressed and lossy (non-reversible to the last digit) compressed modes in linear or 16 bit floating-point encoded format. Fig. 11.4.5.1 shows data from the APS-1 detector in (a) uncorrected mode, (b) transformed to an ideal rectangular detector and (c) transformed to a spherical detector.

11.4.5.10. Detector goniostat

The detector goniostat in *DENZO* can have only one rotation axis – 2θ . In the complex transformations described in equation (11.4.2.8), the geometrical scale is affected by pixel-to-millimetre conversion and distortion. For different instruments, the scale is defined differently. For detectors without distortion, the scale is defined by the value of the pixel size in the 'slow' direction. For detectors with distortion characterized by polynomials (*e.g.* CCD detectors), the scale is also defined by the way the distortion is determined. In such a case, the source of scale is the separation between holes in the reference grid mask or, alternatively, the goniostat translation. As the distance of the detector active surface from the crystal cannot be measured precisely, the difference between the two distances is the ultimate source of the scale reference. The angle between the detector distance translation and

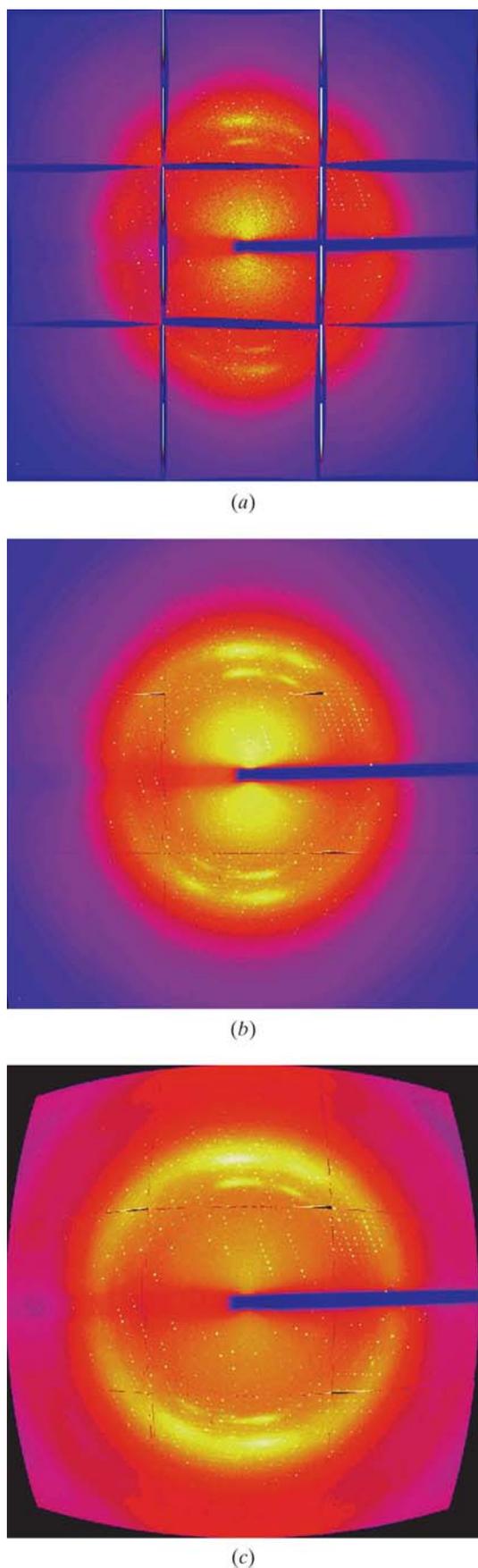


Fig. 11.4.5.1. The transformations in *DENZO* applied to APS-1 detector data. (a) Raw data are affected by geometrical distortion introduced by nine fibre-optic tapers; (b) the same image converted to planar Cartesian space; (c) the same data converted to a virtual spherical detector.

the X-ray beam completes the definition of the detector goniostat in *HKL*.

11.4.5.11. Crystal goniostat

The physical goniostat is defined by six angles. Two angles define the direction of the main axis (ω) in the *DENZO* coordinate system. The third angle defines the zero position of the ω axis. The fourth is the angle between ω and the second axis (κ or χ). The fifth defines the zero position of the second axis. The sixth is the angle between the second and the third axes. This type of goniostat definition allows for the specification of any three-axis goniostat (*EEC Cooperative Workshop on Position-Sensitive Detector Software*, 1986). Each type of goniostat is represented by six angles. Misalignment of the goniostat is represented as an adjustment to these angles, which can be refined by the *HKL* system.

11.4.5.12. Crystal orthogonalization convention

Crystal orientation specified by the three angles needs a definition of a zero point. Any crystal axis, or its equivalent reciprocal-space zone perpendicular to it, can be used as a reference. The definition of zero point aligns the crystal axis with the beam direction and one of the reciprocal axes with the x direction. The user can specify both axes.

11.4.5.13. Refinement and calibration

Both the refinement and calibration procedures determine the properties of the instrument. The principal difference between refinement and calibration is that calibration is performed with data obtained outside the current diffraction experiment, and refinement uses data obtained during the current diffraction experiment. *DENZO* performs both refinement and calibration, and in some cases the difference between calibration and refinement is a question of semantics, as the refined data from one experiment can be used as a reference for another experiment, or even as a reference for a subsequent refinement cycle or for another part of the same experiment.

11.4.6. Prediction of the diffraction pattern

The autoindexing procedure assigns Miller indices only to strong spots, ones that can be found through a peak search. The target of the experiment is to estimate structure factors for all reflections captured by the detector. Therefore, positions of all spots need to be predicted by applying the following equations to all possible triplets \mathbf{h} . Using

$$\mathbf{S} = [\mathbf{A}]^{-1}\mathbf{h}, \quad (11.4.6.1)$$

we have to find the matrix $[\mathbf{A}]$ that generates the vector \mathbf{S} , which satisfies the diffraction condition [equation (11.4.2.1)], knowing that the matrix $[\mathbf{A}]$ is a function of the crystal orientation [equation (11.4.2.6)]. The rotation of the crystal during the experiment creates a straightforward algebraic problem that results in a complex equation defining the angle at which the reflection occurs. This angle also defines the image at which the reflection appears. Knowing this angle, the vector \mathbf{S} can be calculated, and, from equation (11.4.2.5), the direction of the vector \mathbf{X} can be found:

$$\mathbf{X}/|\mathbf{X}| = \lambda(\mathbf{S} - \mathbf{S}_0). \quad (11.4.6.2)$$

Calculation of the length of vector \mathbf{X} requires knowledge of the detector orientation, which, for flat detectors, is described here by vector \mathbf{G} , perpendicular to the detector and with length equal to the crystal-to-detector distance:

11. DATA PROCESSING

$$\mathbf{Y} = [R_{2\theta}]^{-1} \lambda (\mathbf{S} - \mathbf{S}_0), \quad (11.4.6.3)$$

$$\mathbf{H} = [R_z]^{-1} [R_y]^{-1} [R_x]^{-1} \mathbf{G}, \quad (11.4.6.4)$$

$$\mathbf{X} = [R_{2\theta}] \frac{\mathbf{H} \cdot \mathbf{H}}{\mathbf{Y} \cdot \mathbf{H}} \mathbf{Y}. \quad (11.4.6.5)$$

Then, by inverting equation (11.4.2.8), the position in pixels, $\{p, q\}$, of the reflection can be calculated:

$$\{p, q\} = [D]^{-1} [K]^{-1} ([L]^{-1} ([R_{2\theta}]^{-1} ([R_z]^{-1} [R_y]^{-1} [R_x]^{-1} (\mathbf{X} - T_D) + T_D) - T_D) + B). \quad (11.4.6.6)$$

11.4.6.1. Refinement of crystal and detector parameters

The precision of the integration step depends on precise knowledge of the peak position. The autoindexing step provides only the approximate orientation of the crystal, and the result of that step is imprecise if the initial values of the detector parameters are poorly known. A nonlinear least-squares refinement process is used to improve the prediction (*EEC Cooperative Workshop on Position-Sensitive Detector Software*, 1986). Depending on the particulars of the experiment, the same parameters (*e.g.* crystal-to-detector distance) are more precisely known *a priori*, or are better estimated from the diffraction data. *DENZO* allows for the choice of fixing or refining each of the parameters separately. This flexibility is important to characterize a detector, but when detector parameters are already known, the *fit all* option and detector-specific default values are quite reliable.

DENZO can refine the position and orientation of the detector (six parameters). It can also refine internal parameters of the detector including:

- (1) curvature radius and rotation of active surface – *film rotation*, for cylindrical detectors;
- (2) x to y scale and x to y skew;
- (3) radial/angular distortion for spiral scanners; and
- (4) polynomial distortion, separate for each module of multi-module CCD detectors (Naday *et al.*, 1998).

Detector- and crystal-parameter refinement in *DENZO* is achieved by minimizing the sum of the three functions of the type in equation (11.4.5.1). The contribution resulting from the measurement of position p of the reflection is

$$\chi_p^2 = \sum_{hkl} (p_{\text{pred}} - p_{\text{cent}})^2 / \sigma_p^2. \quad (11.4.6.7)$$

The measurement of position q contributes a similar term.

11.4.6.2. Bragg's law for non-ideal conditions: mosaicity

The Bragg condition [equation (11.4.2.1)] assumes ideal crystals and a parallel X-ray beam. In reality, crystals are mosaic and the beam has some angular spread. The value of *mosaicity* describes the range of orientations of the crystal lattice within a sample. As the impacts of mosaicity and the beam's angular spread on the angular width of reflections are equivalent, the keyword *mosaicity* describes the sum of both effects.

DENZO assumes the following model of angular shape of diffraction peaks:

$$M = \frac{\text{mos}}{\pi} \left[1 + \cos \frac{\pi(\varphi - \varphi_c)}{\text{mos}} \right] \quad (11.4.6.8)$$

for φ_c in the range $(\varphi_c - \text{mos}/2; \varphi_c + \text{mos}/2)$, otherwise $M = 0$,

$$P_{\varphi_1 \varphi_2} = \int_{\varphi_1}^{\varphi_2} M(\varphi) d\varphi, \quad (11.4.6.9)$$

where *mos* is mosaicity, φ_c is the predicted angle and P is the predicted partiality of data collected by oscillating from φ_1 to φ_2 .

Partiality is a number that represents what fraction of the reflection intensity is present in one image. If partiality is 1, such reflections are called fully recorded; otherwise they are called partials. For partials, predictions of partiality can be compared with the observed fraction P_0 of the reflection intensity present in one image. The partiality model contributes the following term to the refinement:

$$\chi_P^2 = (P_{\varphi_1 \varphi_2} - P_0)^2 / \sigma_{P_0}^2. \quad (11.4.6.10)$$

The combined positional and partiality refinement used in *DENZO* is both stable and very accurate. The power of this method is in proper weighting (by estimated error) of two very different terms – one describing positional differences and the other describing intensity differences. Both detector and crystal variables are uniformly treated in the refinement process.

11.4.6.3. Detector distortions

The design of detectors results in pixels not being positioned on an exact square or rectangular grid. A correct understanding of the detector distortions is essential to accurate positional refinement. The types of distortions are detector-specific. The primary sources of error include misalignment of the detector position sensors and optical or magnetic distortion in CCD-based detectors. If the detector distortion can be parameterized, then these parameters should be added to the refinement. For example, in the case of spiral scanners, there are two parameters describing the end position of the scanning head. In a perfectly adjusted scanner, these parameters would be zero. In practice, however, they may deviate from zero by as much as 1 mm. Such misalignment parameters can correlate very strongly with other detector and crystal parameters, particularly for low-symmetry lattices or in the case of low-resolution data. If the distortions are stable, it is better to determine them in a separate experiment optimized for that task.

Fibre-optic tapers used in many CCD detectors have distortion that has to be individually determined for each instrument. The distortion is stable over time and its spatial characteristics are dominated by a smooth component and a small local shear. In high-quality tapers used in X-ray instruments, the small local shear can be ignored. The smooth component can be parameterized in a number of ways, for example by splines (Hammersley, 1998) or polynomials (Messerschmidt & Pflugrath, 1987). *DENZO* uses two-dimensional Chebyshev polynomials (Press *et al.*, 1989) in $\{x, y\}$ or $\{p, q\}$ coordinates, normalized to the range $\{[-1, +1], [-1, +1]\}$. Typically, fifth- or seventh-order polynomials result in a positional error (r.m.s.) lower than 7 μm (about one tenth of the detector pixel). *DENZO* can use either a grid mask pattern or the X-ray diffraction pattern to refine the coefficients of the Chebyshev polynomials. If a grid mask is used, it has to be precisely made and positioned. The use of crystallographic data requires precise knowledge of detector and crystal parameters that are not known *a priori* with the required precision. The crystal and detector parameters can be determined in the same experiment as detector distortion. However, this experiment needs to be designed to minimize the impact of correlations between the parameters involved. The data analysis requires the description of the distortion function and its inverse. In *DENZO*, both are approximated in terms of Chebyshev polynomials. The magnitude of the approximation error is the same for the distortion function and its inverse.

11.4.7. Detector diagnostics

The *HKL* package has a number of tools that can detect possible detector or experimental setup problems (Minor & Otwinowski, 1997). Visual inspection of the image may provide only a very rough estimate of data quality. A check of the analogue-to-digital converter can provide rough diagnostics of detector electronics. Examination of the background can provide information about detector noise, especially when uncorrected images can be examined in the areas exposed to X-rays and areas where pure read-out noise can be observed. *DENZO* provides several diagnostic tools during the integration stage, as the crystallographer may observe crystal slippage, a change of unit-cell parameters or a change of the values of positional and angular χ^2 during the refinement. Even more tools are provided at the data-scaling stage. By observing scale factors, poor crystal alignment can be detected. Other tools may help diagnose X-ray shutter malfunction, spindle axis alignment and internal detector alignment problems. The final inspection of outliers may again provide valuable information about detector quality. The clustering of outliers in one area of the detector may indicate a damaged surface; if most outliers are partials, it may indicate a problem with spindle backlash or shutter control. The zoom mode may be used to display the area around the outliers to identify the source of a problem: for example, the existence of a satellite crystal or single pixel spikes due to electronic failure. Sometimes, even for very strong data, a histogram of the pixel intensities may stop below the maximum valid pixel value, indicating saturation of the data-acquisition hardware or software.

11.4.8. Multiplicative corrections (scaling)

Proper error estimation requires the use of Bayesian reasoning and a multi-component error model (Schwarzenbach *et al.*, 1989; Evans, 1993). In *SCALEPACK*, the estimated error of the measurement is enlarged by a fraction of the expected, rather than the observed, intensity. This algorithm reduces the bias towards reflections with an integrated intensity below the average.

The scaling model allows for a large number of diverse components to contribute to the multiplicative correction factor s for each observation,

$$s = \exp \left[\sum_i p_i f_i(hkl) \right], \quad (11.4.8.1)$$

where p_i are *a priori* unknown coefficients of the scaling components and f_i represent different functional dependence of the scale factor for each observation. The simplest scaling model has a separate scale factor for each group (batch) of data, for example, one scale factor per image. In such a case,

$$f_i = \delta_{ij}, \quad (11.4.8.2)$$

where j is the batch index for a particular reflection. For resolution-dependent decay, represented by one temperature factor per batch of data,

$$f_{i+n} = [(\mathbf{S} \cdot \mathbf{S})/2] \delta_{ij}, \quad (11.4.8.3)$$

where n is the number of batches needed to make p_i represent the logarithm of the overall i th batch scale factor and p_{i+n} represents the temperature factor of batch i . \mathbf{S} is the scattering vector for each reflection.

Coefficients of crystal absorption are much more complex. Scaling coefficients are associated with spherical harmonics (Katayama, 1986; Blessing, 1995) as a function of the direction of vector \mathbf{S} , expressed in the coordinate system of the rotating crystal. Each spherical harmonic index lm has two (or, in the case of $m = 0$, one) p_g coefficients. One of these spherical harmonic

functions is given by

$$f_g = \left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!} \right]^{1/2} P_{lm}(\cos \Psi) \cos(m\Phi), \quad (11.4.8.4)$$

where Ψ and Φ are polar-coordinate angles of vector \mathbf{S} in the crystal coordinate system, and P_{lm} is a Legendre polynomial (Press *et al.*, 1989). The other spherical harmonic of index lm has a sine instead of a cosine as the last term.

The multiplicative factor is applied to each observation and its σ to obtain the corrected intensity I_{corr} and associated σ . The averaged intensity over all symmetry-related reflections $\langle I_{\text{corr}} \rangle$ is obtained by solving the two following equations:

$$W = 1 / [(\sigma E_1)^2 + (\langle I_{\text{corr}} \rangle E_2)^2], \quad (11.4.8.5)$$

where E_1 and E_2 are the user-specified *error scale factor* and *estimated error*, respectively, and

$$\langle I_{\text{corr}} \rangle = \sum I_{\text{corr}} W / \sum W. \quad (11.4.8.6)$$

Thus,

$$\sigma(I) = I / [\sum W]^{1/2}. \quad (11.4.8.7)$$

During parameter refinement, the scale (and B , if requested) for all scaled batches are refined against the difference between the $\langle I_{\text{corr}} \rangle$'s and I_{corr} 's for individual measurements, summed over all reflections (Fox & Holmes, 1966; Arndt & Wonacott, 1977; Stuart & Walker, 1979; Leslie & Tsukihara, 1980; Rossmann & Erickson, 1983; Walker & Stuart, 1983; Rossmann, 1984; Schutt & Evans, 1985; Stuart, 1987; Takusagawa, 1987; Tanaka *et al.*, 1990). $\langle I_{\text{corr}} \rangle$'s are recalculated in each cycle of refinement. There is full flexibility in the treatment of anomalous pairs. They can be assumed to be equivalent (or not) and they may be merged (or not). This approach allows the crystallographer to choose the best scaling and merging strategy.

11.4.8.1. Polarization

A polarization correction may be applied during *DENZO* calculations. Sometimes the exact value of polarization is not known. This error may be corrected during the scaling procedure. This feature can be used to refine the polarization at synchrotron beamlines. Very high resolution data should be used for this purpose.

11.4.9. Global refinement or post refinement

The process of refining crystal parameters using the combined reflection intensity measurements is known as global refinement or post refinement (Rossmann, 1979; Evans, 1993). The implementation of this method in *SCALEPACK* allows for separate refinement of the orientation of each image, but with the same unit-cell value for the whole data set. In each batch of data (a batch is typically one image), different unit-cell parameters may be poorly determined. However, in a typical data set, there are enough orientations to determine all unit-cell lengths and angles precisely. Global refinement is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image.

11.4.10. Graphical command centre

The goal of the command centre is to coordinate all phases of the experiment and to facilitate interactive experiments in which data

11. DATA PROCESSING

analysis is done on-line, where results are automatically updated when new data are collected. In such experiments, it is possible to adjust the data-collection strategy to guarantee the desired result,

particularly with regard to data completeness (Fig. 11.4.10.1). The strategy should take into account limitations arising from radiation damage or shortage of allocated time. The radiation damage (Fig. 11.4.10.2) can be estimated both from

experience of the beamline with similar crystals (with all frozen crystals being rather similar, since they have a limited range of sensitivity to a particular radiation dose) and by evaluating real-time changes in scale and B factors.

Another goal of the command centre is to enable efficient use of high-speed, high-intensity synchrotron beamlines, where the rate of data flow is enormous. The traditional approach to data processing and management [graphical user interface (GUI)-based or not] is to execute data collection and processing steps serially. This approach is well tuned to the human style of thinking 'one task at a time', but does not allow for efficient use of synchrotron time. Manual methods of coordinating data backup, file transfer between computers or allocating disk space or other resources needed to complete an experiment considerably slow the work at fast synchrotron beamlines. Since all of the tasks can be organized from the command centre, the experimenter is free to concentrate on data collection and assessment of data quality rather than data management.

The command centre consists of three components: a database, a transition-state engine (a set of rules that define possible atomic changes of the database) and a GUI. It is based on the idea of a single database that stores all the information about data processing and data collection. The database is a dynamic one; it can describe not only the data already collected, but also those being collected and even those planned or considered to be collected. Each data-entry step or program-execution step, including the data-collection program, induces a change in the database. One of the main functions of the GUI is to provide for user input and editing of the database. The other major function of the GUI is to provide reports from the database (to visualize the status of the database). The complexity of the database results in the need to create hierarchical access to the information.

The command-centre database abstraction is based on the following descending hierarchy: instrument type; site; experiment; crystal; three-dimensional (3D) group; image. Each lower level of the hierarchy inherits the properties of the higher levels. When a program finishes analysing an instance of a particular level, the higher-level instance is updated, so that instances of the same level communicate only through the change of state of their common parent. The site instances are

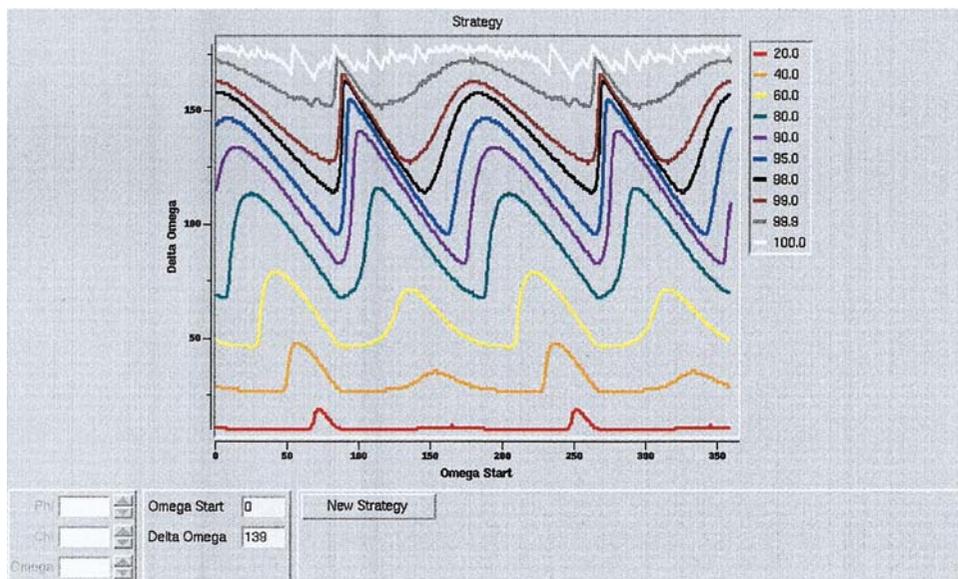


Fig. 11.4.10.1. The strategy program as implemented in the *HKL* package. The colours of the lines reflect different data-completeness goals. The y axis is the total oscillation range necessary to achieve a particular completeness goal. The x axis shows the starting position of the spindle axis.

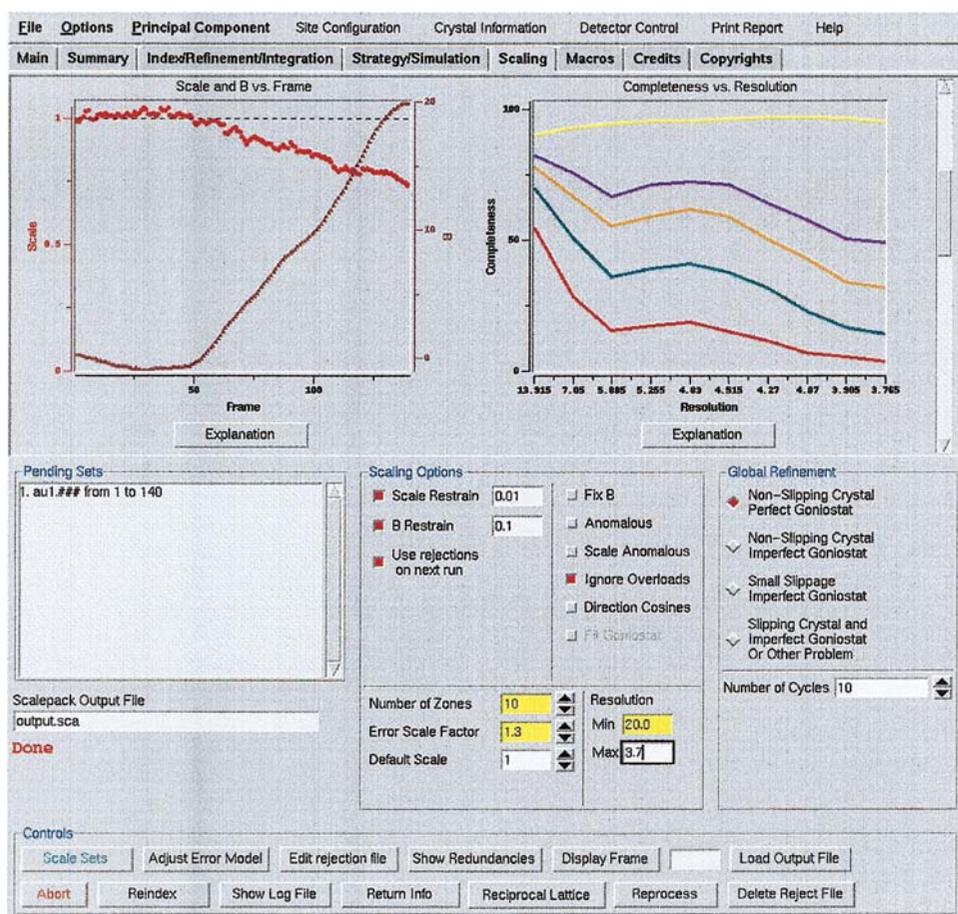


Fig. 11.4.10.2. The scaling window of the graphical user interface. The scaling can be done during the integration and data-collection process. In particular, the experimenter can observe the crystal decay in the plots showing the scale and B factors (on the left). The completeness for different $I/\sigma(I)$ is shown on the right. The scrolled widget contains many plots presenting the statistics calculated by merging all available data.

11.4. DENZO AND SCALEPACK

created when data from a new detector appear or when the detector is rebuilt, which is done rarely, typically by the X-ray equipment administrator. The instance of the experiment allows for data from more than one crystal of the same space group. The uniform series of diffraction images form 3D groups. There is no limit to the number of 3D groups and, in the case of non-uniformity in the series (*e.g.* found during data analysis), the 3D group can be split into two or more smaller 3D groups. The smallest 3D group can consist of one image. The crystal instance contains a set of 3D groups with a relative orientation and exposure level known *a priori*. In practice, this means that data contained in a single crystal instance were collected from one sample at one site with potentially different settings of goniostat, data-collection axis, crystal translation, detector position, detector mode (*e.g.* binned/unbinned) or exposure level.

11.4.11. Final note

The methods presented here have been used to solve a great variety of problems, from inorganic molecules with 3 Å unit-cell

parameters to a virus of 700 Å diameter which crystallized in a $700 \times 1000 \times 1400$ Å cell. The most important test, stressing the precision and robustness of the method, is the successful application of the programs to many multiple-wavelength anomalous dispersion structure determinations.

Acknowledgements

This work was supported by NIH grant GM-53163. We would like to acknowledge the contributions of the following people who developed other data-analysis programs and interactions with whom contributed to the *HKL* program development and to the ideas presented here: M. G. Rossmann, W. Kabsch, J. Pflugrath, G. Bricogne, P. Evans, G. Sheldrick, A. Howard and A. Leslie. We would also like to thank H. Czarnocka, D. Tomchick, A. Pertsemlidis and W. Majewski for help in preparing this manuscript.

11. DATA PROCESSING

11.3 (cont.)

- Schutt, C. & Winkler, F. K. (1977). *The oscillation method for very large unit cells*. In *The rotation method in crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 173–186. Amsterdam, New York, Oxford: North-Holland.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *An algorithm for automatic indexing of oscillation images using Fourier analysis*. *J. Appl. Cryst.* **30**, 1036–1040.
- Wirth, N. (1976). *Algorithms + data structures = programs*, pp. 264–274. New York: Prentice-Hall.

11.4

- Arndt, U. W. & Wonacott, A. J. (1977). *The rotation method in crystallography*. Amsterdam: North Holland.
- Blessing, R. H. (1995). *An empirical correction for absorption anisotropy*. *Acta Cryst.* **A51**, 33–38.
- Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. (1987). *A system for collection and on-line integration of X-ray diffraction data from a multiwire area detector*. *J. Appl. Cryst.* **20**, 235–242.
- Bricogne, G. (1987). *The EEC cooperative programming workshop on position-sensitive detector software*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 120–146. Warrington: Daresbury Laboratory.
- EEC Cooperative Workshop on Position-Sensitive Detector Software (1986). Phase I and II, LURE, Paris, 16 May–7 June; Phase III, LURE, Paris, 12–19 November.
- Evans, P. (1993). *Data reduction: data collection and processing*. In *Proceedings of the CCP4 study weekend. Data collection and processing*, 29–30 January, edited by L. Sawyer, N. Isaac & S. Bailey, pp. 114–123. Warrington: Daresbury Laboratory.
- Evans, P. R. (1987). *Postrefinement of oscillation camera data*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 58–66. Warrington: Daresbury Laboratory.
- Fox, G. C. & Holmes, K. C. (1966). *An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks*. *Acta Cryst.* **20**, 886–891.
- Gewirth, D. (1996). *HKL manual*. 5th ed. Yale University, New Haven, USA.
- Greenhough, A. G. W. (1987). *Partials and partiality*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 51–57. Warrington: Daresbury Laboratory.
- Hammersley, A. P. (1998). *The FIT2D home page*. <http://www.ccp14.ac.uk/ccp/web-mirrors/fit2d/computing/expg/subgroups/dataanalysis/FIT2D/index.html>
- Higashi, T. (1990). *Auto-indexing of oscillation images*. *J. Appl. Cryst.* **23**, 253–257.
- Howard, A., Nielsen, C. & Xuong, Ng. H. (1985). *Software for a diffractometer with multiwire area detector*. *Methods Enzymol.* **114**, 452–472.
- Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salemme, F. R. (1987). *The use of an imaging proportional counter in macromolecular crystallography*. *J. Appl. Cryst.* **20**, 383–387.
- International Tables for Crystallography* (1995). Vol. A. *Space-group symmetry*, edited by Th. Hahn. Dordrecht: Kluwer Academic Publishers.
- Kabsch, W. (1988). *Evaluation of single-crystal X-ray diffraction data from a position sensitive detector*. *J. Appl. Cryst.* **21**, 916–924.
- Kabsch, W. (1993). *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants*. *J. Appl. Cryst.* **26**, 795–800.
- Katayama, C. (1986). *An analytical function for absorption correction*. *Acta Cryst.* **A42**, 19–23.
- Kim, S. (1989). *Auto-indexing oscillation photographs*. *J. Appl. Cryst.* **22**, 53–60.
- Leslie, A. (1993). *Autoindexing of rotation diffraction images and parameter refinement*. In *Proceedings of the CCP4 study weekend. Data collection and processing*, 29–30 January, edited by L. Sawyer, N. Isaac & S. Bailey, pp. 44–51. Warrington: Daresbury Laboratory.
- Leslie, A. G. W. (1987). *Profile fitting*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 39–50. Warrington: Daresbury Laboratory.
- Leslie, A. G. W. & Tsukihara, T. (1980). *A strategy for collecting isomorphous derivative data with the oscillation method*. *J. Appl. Cryst.* **13**, 304–305.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *Crystal orientation and X-ray pattern prediction routines for area-detector diffraction systems in macromolecular crystallography*. *J. Appl. Cryst.* **20**, 306–315.
- Minor, W. & Otwinowski, Z. (1997). *Advances in accuracy and automation of data collection and processing*. In *Proceedings of the IUCr computing school*, Bellingham, 1996.
- Naday, I., Ross, S., Westbrook, E. M. & Zentai, G. (1998). *Charge-coupled device/fiber optic taper array X-ray detector for protein crystallography*. *Opt. Eng.* **37**, 1235–1244.
- Otwinowski, Z. (1993). *Oscillation data reduction program*. In *Proceedings of the Daresbury CCP4 study weekend. Data reduction and processing*, edited by L. Sawyer, N. Isaacs and S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Processing of X-ray diffraction data collected in oscillation mode*. *Methods Enzymol.* **276**, 307–326.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1989). *Numerical recipes – the art of scientific computing*. Cambridge University Press.
- Rossmann, M. G. (1979). *Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting*. *J. Appl. Cryst.* **12**, 225–238.
- Rossmann, M. G. (1984). *Synchrotron radiation studies of large proteins and supramolecular structures*. In *Proceedings of the study weekend at Daresbury Laboratory. Biological systems: structure and analysis*, 24–25 March, edited by G. P. Diakun & C. D. Garner, pp. 28–40. Daresbury: Science and Engineering Research Council.
- Rossmann, M. G. & Erickson, J. W. (1983). *Oscillation photography of radiation-sensitive crystals using a synchrotron source*. *J. Appl. Cryst.* **16**, 629–636.
- Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). *Processing and post-refinement of oscillation camera data*. *J. Appl. Cryst.* **12**, 570–581.
- Sakabe, N. (1991). *X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation*. *Nucl. Instrum. Methods A*, **303**, 448–463.
- Schutt, C. E. & Evans, P. R. (1985). *Relative absorption correction for rotation film data*. *Acta Cryst.* **A41**, 568–570.
- Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Gonschorek, W., Hahn, Th., Huml, K., Marsh, R. E., Prince, E., Robertson, B. E., Rollett, J. S. & Wilson, A. J. C. (1989). *Statistical descriptors in crystallography*. *Acta Cryst.* **A45**, 63–75.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *An algorithm for automatic indexing of oscillation images using Fourier analysis*. *J. Appl. Cryst.* **30**, 1036–1040.
- Stuart, D. (1987). *Absorption correction*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin & M. Z. Papiz, pp. 25–38. Warrington: Daresbury Laboratory.
- Stuart, D. & Walker, N. (1979). *An empirical method for correcting rotation-camera data for absorption and decay effects*. *Acta Cryst.* **A35**, 925–933.
- Takusagawa, F. (1987). *A simple method of absorption and decay correction in intensities measured by area-detector X-ray diffractometer*. *J. Appl. Cryst.* **20**, 243–245.
- Tanaka, I., Yao, M., Suzuki, M., Hikichi, K., Matsumoto, T., Kozasa, M. & Katayama, C. (1990). *An automatic diffraction data*

REFERENCES

- collection system with an imaging plate. *J. Appl. Cryst.* **23**, 334–339.
- Vriend, G. & Rossmann, M. G. (1987). Determination of the orientation of a randomly placed crystal from a single oscillation photograph. *J. Appl. Cryst.* **20**, 338–343.
- Walker, N. & Stuart, D. (1983). An empirical method for correcting diffractometer data for absorption effects. *Acta Cryst.* **A35**, 158–166.
- Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). The oscillation method for crystals with very large unit cells. *Acta Cryst.* **A35**, 901–911.
- 11.5**
- Bernal, R., Burch, A., Fane, B. & Rossmann, M. G. (1998). Unpublished results.
- Blessing, R. H. (1997). Outlier treatment in data merging. *J. Appl. Cryst.* **30**, 421–426.
- Bolotovskiy, R. & Coppens, P. (1997). The φ extent of the reflection range in the oscillation method according to the mosaicity-cap model. *J. Appl. Cryst.* **30**, 65–70.
- Bolotovskiy, R., Steller, I. & Rossmann, M. G. (1998). The use of partial reflections for scaling and averaging X-ray area detector data. *J. Appl. Cryst.* **31**, 708–717.
- Choi, H. K., Lee, S., Zhang, Y. P., McKinney, B. R., Wengler, G., Rossmann, M. G. & Kuhn, R. J. (1996). Structural analysis of Sindbis virus capsid mutants involving assembly and catalysis. *J. Mol. Biol.* **262**, 151–167.
- Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature (London)*, **354**, 37–43.
- Colbert, C. & Bolin, J. (1999). Unpublished results.
- Dokland, T., McKenna, R., Ilag, L. L., Bowman, B. R., Incardona, N. L., Fane, B. A. & Rossmann, M. G. (1997). Structure of a viral procapsid with molecular scaffolding. *Nature (London)*, **389**, 308–313.
- Fox, G. C. & Holmes, K. C. (1966). An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks. *Acta Cryst.* **20**, 886–891.
- Gewirth, D. (1996). *The HKL manual. A description of the programs DENZO, XDSPLAYF and SCALEPACK*, 5th ed., pp. 87–90. Yale University, New Haven, USA.
- Greenhough, T. J. & Helliwell, J. R. (1982). Oscillation camera data processing: reflecting range and prediction of partiality. I. Conventional X-ray sources. *J. Appl. Cryst.* **15**, 338–351.
- Hamilton, W. C., Rollett, J. S. & Sparks, R. A. (1965). On the relative scaling of X-ray photographs. *Acta Cryst.* **18**, 129–130.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Rodgers, D. W. (1994). Cryocrystallography. *Structure*, **2**, 1135–1140.
- Rossmann, M. G. (1979). Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *J. Appl. Cryst.* **12**, 225–238.
- Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature (London)*, **317**, 145–153.
- Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). Processing and post-refinement of oscillation camera data. *J. Appl. Cryst.* **12**, 570–581.
- Rossmann, M. G., Momany, C. A., Cheng, B. & Chakravarty, S. (1997). Unpublished results.
- Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). The oscillation method for crystals with very large unit cells. *Acta Cryst.* **A35**, 901–911.
- Wonacott, A. J. (1977). *Geometry of the oscillation method*. In *The rotation method in crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 75–103. Amsterdam: North Holland.