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

$$\mathbf{S} \cdot \mathbf{S}_0 = \tfrac{1}{2}\mathbf{S} \cdot \mathbf{S}, \qquad (11.4.2.1)$$

---

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

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

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

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

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

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 $\mathbf{S}$:

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

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):

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

where

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

$$+ \sin(\varphi)\begin{pmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{pmatrix}, \qquad (11.4.2.7)$$

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,

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

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

226