11.5. The use of partially recorded reflections for post refinement, scaling and averaging X-ray diffraction data

BY C. G. VAN BEEK, R. BOLOTOVSKY AND M. G. ROSSMANN

11.5.1. Introduction

Recent advances in the use of frozen crystals of biological samples for X-ray diffraction data collection (Rodgers, 1994) often result in data for which most of the observed reflections on each frame are partially observed. This might be avoided by increasing the oscillation ranges, but this would cause many reflections to overlap with their neighbours. Hence, it is necessary to develop scaling procedures that are independent of the exclusive use of fully recorded reflections.

A set of measured Bragg intensities is dependent on the properties of the crystal, radiation source and detector. Usually, these factors cannot be kept constant throughout the data collection. The crystal may decay, weakening the Bragg intensities, or even 'die', which requires the use of several crystals for a full data set. The intensity and position of the primary X-ray beam may vary, especially at synchrotron-radiation sources. Finally, the detector response may change when, for example, different films or imaging plates are used during the data collection.

Most data sets can be divided into series of subsets, or frames, collected under more-or-less constant conditions. These frames need to be placed on a common arbitrary scale. The scaling can be performed by comparing the intensities of multiply measured reflections or symmetry-equivalent reflections on different frames.

A least-squares procedure frequently used for scaling frames of data is the Hamilton, Rollett and Sparks (HRS) method (Hamilton *et al.*, 1965). The target for the HRS least-squares minimization is

$$\psi = \sum_h \sum_i W_{hi}(I_{hi} - G_m I_h)^2, \qquad (11.5.1.1)$$

where I_h is the best estimate of the intensity of a reflection with reduced Miller indices h, I_{hi} is the intensity of the ith measurement of reflection h, W_{hi} is a weight for reflection h_i and G_m is the inverse linear scale factor for frame m on which reflection h_i is recorded. The reduced Miller indices are those corresponding to an arbitrarily defined asymmetric unit of reciprocal space. The HRS expression (11.5.1.1) assumes that all reflections h_i are full, that is, their reciprocal-lattice points have completely passed through the Ewald sphere.

For all unique reflections h, the values of I_h must correspond to a minimum in ψ. Thus,

$$\partial\psi/\partial I_h = 0. \qquad (11.5.1.2)$$

Therefore, the best least-squares estimate of the intensity of a reflection is

$$I_h = \sum_i W_{hi} G_m I_{hi} \Big/ \sum_i W_{hi} G_m^2. \qquad (11.5.1.3)$$

Since ψ is not linear with respect to the scale factors G_m, the values of the scale factors have to be determined by an iterative nonlinear least-squares procedure. As the scale factors are relative to each other, the HRS procedure requires that one of them be fixed.

Fox & Holmes (1966) describe an improved method of solving the HRS normal equations. Their approach is based on the singular value decomposition of the normal equations matrix. The advantage of the Fox and Holmes method, apart from the accelerated convergence of the least-squares procedure, is that no *ad hoc* decision needs to be made as to which scale factor should be fixed. Furthermore, 'troublesome' frames of data can be identified as causing negligibly small eigenvalues in the normal equations matrix.

11.5.2. Generalization of the Hamilton, Rollett and Sparks equations to take into account partial reflections

When a Bragg reflection is completely exposed within the oscillation range of one frame, a so-called 'full reflection', it gives rise to the 'full intensity'. In general, a Bragg reflection will occur on a number of consecutive frames as a series of partial reflections, and the full intensity can only be estimated from the measured intensities of the partial reflections. Let I_{him} represent the intensity contribution of reflection h_i recorded on frame m; if all the parts of h_i are available in the data set, then

$$I_{hi} = \sum_m (I_{him}/G_m). \qquad (11.5.2.1)$$

In practice, there will always be reflections that do not have all their parts available. In such cases, the only way to estimate the full intensity of a reflection is to apply an estimated value of the partiality to the measured reflection intensities.

Various models have been proposed to calculate the reflection partiality. Here we use Rossmann's model (Rossmann, 1979; Rossmann *et al.*, 1979) with Greenhough & Helliwell's (1982) correction. This model treats partiality as a fraction of a spherical volume swept through the Ewald sphere. The coordinates of the reciprocal-lattice point are defined by the Miller indices of the reflection, the crystal orientation matrix and the rotation angle. The volume of the sphere around the reciprocal-lattice point accounts for crystal mosaicity and beam divergence. Alternative geometrical descriptions of a reciprocal-lattice point passing through the Ewald sphere have been given by Winkler *et al.* (1979) and Bolotovsky & Coppens (1997).

Provided the reflection partiality, p_{him}, is known, the full intensity is estimated by

$$I_{hi} = I_{him} \big/ p_{him} G_m. \qquad (11.5.2.2)$$

This expression can produce as many estimates of I_{hi} as there are parts of reflection h_i, while expression (11.5.2.1) produces only one estimate of I_{hi} when all parts of reflection h_i are recorded. Having defined the relationships between measured intensities of partial reflections and estimated full reflections by expressions (11.5.2.1) and (11.5.2.2), two methods of generalizing the HRS equations can be considered.

Method 1. If a reflection h_i occurs on a number of consecutive frames and all parts of I_{him} are available in the data set, the generalized HRS target equation takes the form

$$\psi = \sum_h \sum_i \sum_m W_{him} \left\{ I_{him} - G_m \left[I_h - \sum_{m' \neq m} (I_{him'}/G_{m'}) \right] \right\}^2. \qquad (11.5.2.3)$$

Using expression (11.5.1.2), the best least-squares estimate of I_h will be

$$I_h = \frac{\sum_i \left[\sum_m (I_{him}/G_m)\right] \left(\sum_m W_{him} G_m^2\right)}{\sum_i \sum_m W_{him} G_m^2} = \frac{\sum_i I_{hi} \sum_m W_{him} G_m^2}{\sum_i \sum_m W_{him} G_m^2}. \qquad (11.5.2.4)$$

Method 2. If the theoretical partiality, p_{him}, of the partially recorded reflection h_{im} can be estimated, the generalized HRS target equation takes the form

references