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.

references