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,

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

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,

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

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

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

where n is the number of batches needed to make p_i represent the logarithm of the overall ith 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

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

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 $I_{\rm corr}$ and associated σ. The averaged intensity over all symmetry-related reflections $\langle I_{\rm corr}\rangle$ is obtained by solving the two following equations:

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

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

$$\langle I_{\rm corr}\rangle = \sum I_{\rm corr} W \Big/ \sum W. \qquad (11.4.8.6)$$

Thus,

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

During parameter refinement, the scale (and B, if requested) for all scaled batches are refined against the difference between the $\langle I_{\rm corr}\rangle$'s and $I_{\rm corr}$'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). $\langle I_{\rm corr}\rangle$'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

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.

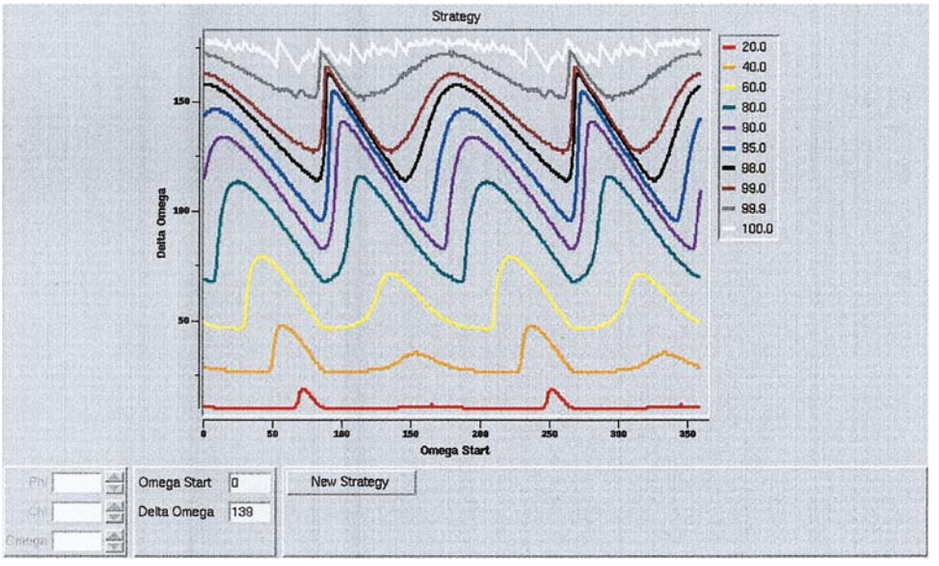

Fig. 11.4.10.1. The strategy program as implemented in the *HKL* package. The colours of the lines reflect different data-completeness goals. The y axis is the total oscillation range necessary to achieve a particular completeness goal. The x axis shows the starting position of the spindle axis.

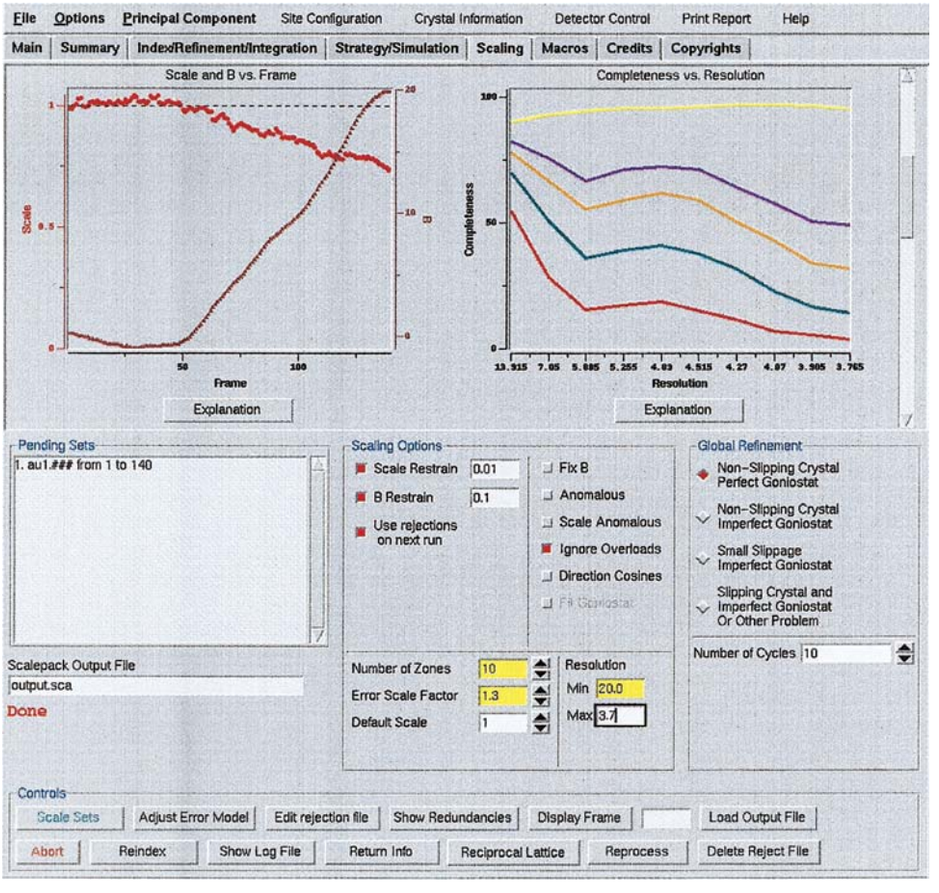

Fig. 11.4.10.2. The scaling window of the graphical user interface. The scaling can be done during the integration and data-collection process. In particular, the experimenter can observe the crystal decay in the plots showing the scale and B factors (on the left). The completeness for different $I/\sigma(I)$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

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

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.