

## 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

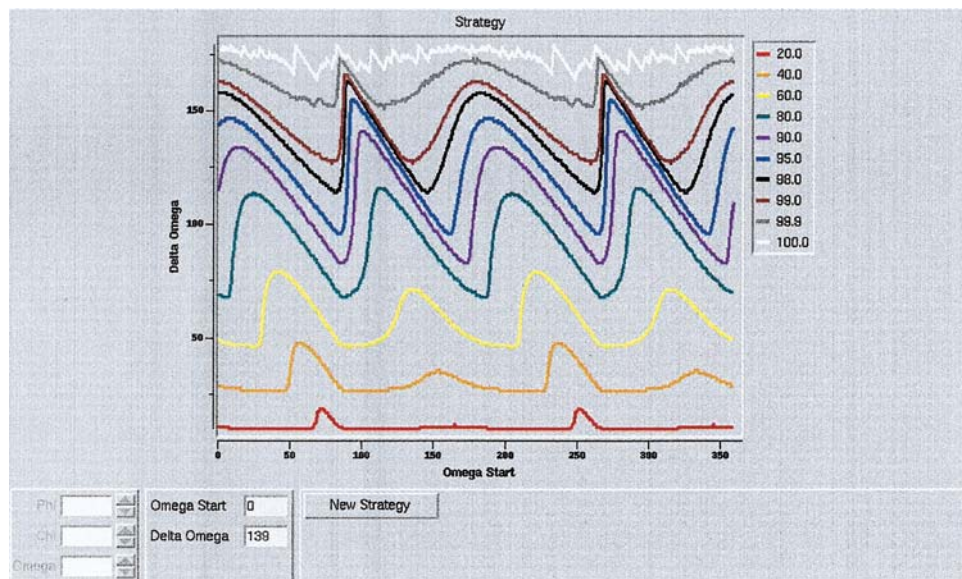


Fig. 11.4.10.1. The strategy program as implemented in the *HKL* package. The colours of the lines reflect different data-completeness goals. The y axis is the total oscillation range necessary to achieve a particular completeness goal. The x axis shows the starting position of the spindle axis.

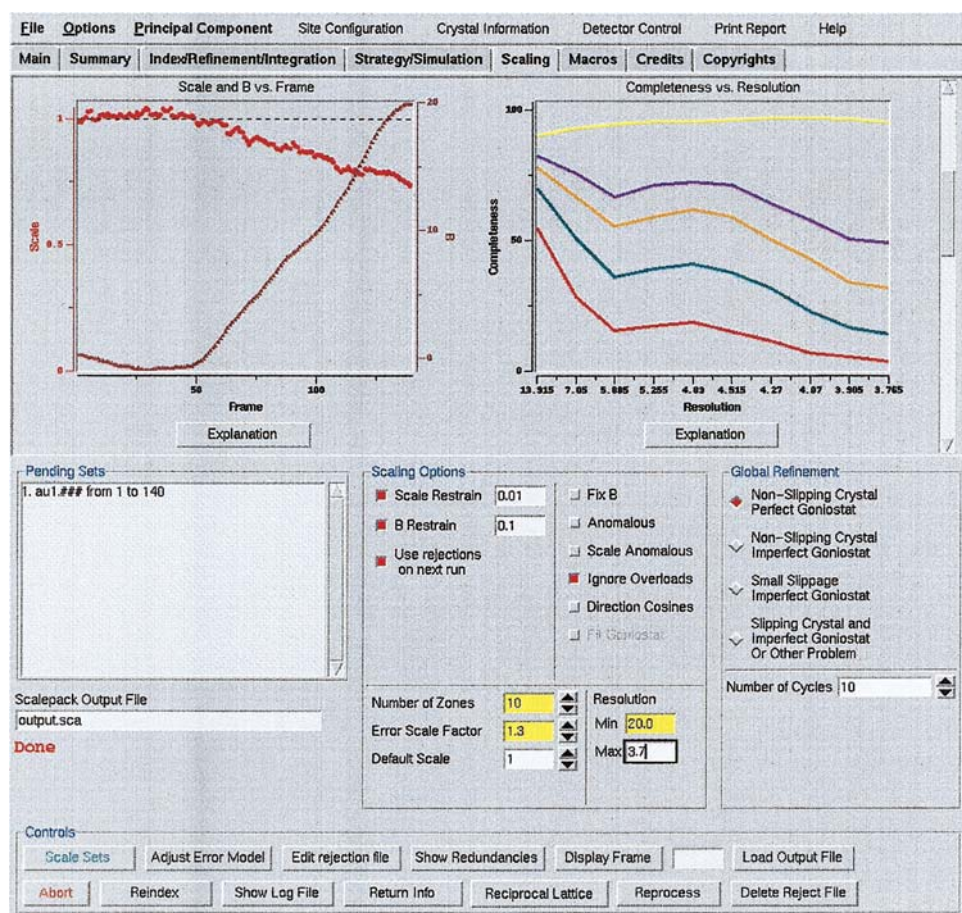


Fig. 11.4.10.2. The scaling window of the graphical user interface. The scaling can be done during the integration and data-collection process. In particular, the experimenter can observe the crystal decay in the plots showing the scale and  $B$  factors (on the left). The completeness for different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