Reduce Intensity Data

This paragraph describes the general steps followed by data reduction programs for all area detector instruments.

1. The background in different regions of reciprocal space is determined.

2. From the orientation matrix(matrices), the locations of all peaks on each frame are determined.

3. From the selected box around each peak, the peak is integrated in that 2-dimensional slice and the backgrounds are subtracted.

4. The peaks from successive slices are summed for each (*hkl*) to get a total intensity for each peak.

5. The peak intensities are corrected for Lorentz and polarization effects and numerous anomalies in the detector and incident beam.

6. The peaks from all frames are sorted based on (*hkl*) values.

7. Selected statistics about the data are prepared, and the cell parameters are optimized from a selected set of peaks.

Data reduction may be started any time after about 1/3 of the unique intensity data are collected to verify the Laue symmetry of the data and to solve and partially refine the model of atoms.

Best results can be achieved by performing multiple data reduction runs, each time optimizing the orientation matrix and the spot size. After the first data reduction run(s), that is(are) used to solve and partially refine the model, import a better orientation matrix by reading in the *project*_0m.p4p file in the Sample > Import menu. Also examine the *project*_0*n*._ls (where *n* is the run number) files to get an idea of better X, Y, and Z values for the integration box.

The IUCr requires that at least 98% of all unique data be collected to at least 0.82Å resolution for small-molecule crystal structures.

Specific Steps to Run the Data Reduction Program in APEX - Saint

1. In the *Reduce Data > Integrate Images* routine select the data runs using the *Find Runs* command. Select those runs that have a significant number (> 25) of frames. A *Fast scan* run may be used to initially solve and refine the structure to check for correct Laue symmetry. However, after all data are collected, omit all *Fast scan* runs until the very last run as described below.

2. In the *Refinement Options* menu, uncheck the box for *Enable box size refinement*. Adjust the size of the box to resemble the average spot size. If the values for X and Y are < 0.5, change the X and Y values to 0.95 and to not change the Z value. You may improve these in later data reduction iterations. In the *Periodic Refinement* panel, ***always*** set the symmetry to "Triclinic". In the *Global Refinement* panel, set the symmetry of the cell to "Triclinic" for initial runs. *Close* this menu.

3. Select the *Integration Options* menu. In the *Background Subtraction*" panel select *Use Best Plane Background*. Set the *Active Image Queue Half-Width* (*Images*) to at least "12". For twinned samples select *Additional Parameters* and change the overlap factor to "15.0 – 20.00". Close this menu.

4. Begin data reduction with *Execute*. Once integration has completed, check any statistics desired, then close this integration run with the *Close* button.

5. New integration runs are improved by the following. Use *Sample* > *Import* to update the cell parameters using the *project*_0m.p4p. Also examine the *project*_0*n*._ls files to determine better values for the X, Y, and Z parameters for the integration box, adjusting these parameters in the *Refinement Options* menu. If the average $I/\sigma > 20$, then check the *Enable box size refinement* box. *Execute* data reduction. Repeat this step until no large shifts are seen in the parameters.

6. In the penultimate data reduction pass, set the cell symmetry to the value expected in the *Global Refinement* panel. After executing this pass, *rename* the *project*_0m.p4p file to the *project*.p4p.

7. In the final data reduction pass, use the *Find Runs* option to add the *Fast scan* run(s) to the end of the list of runs. Execute data reduction without making any changes to the other options. At the end of this data reduction run, **copy** the *project*_0m._ls file to *project*._ls and **copy** the *project*_0m.raw to *project*.raw . Note these files will be *project*_0m.mul and *project*.mul for integrations of a twinned sample.

The penultimate run, which does not include fast scan data, gives the best unit cell information. The final run gives the best *.raw or *.mul file for the absorption correction program. Fast scan data is usually needed to correct the data for absorption/scaling, but should ***not*** be used to determine the best cell parameters.

Indexing Faces of a Crystal

For strongly-absorbing, untwinned samples, a routine in APEX3 is available to help index the faces of a crystal. The *Reduce Data* > *Index Crystal Faces* > *Acquire New* menu will allow you to collect 360 images of the crystal, 1 image per degree as the sample is rotated by 360° in ϕ. The images are stored in a zipped file with the name *project*.vzs. After the images have been collected, an image of the crystal is displayed. Below the image set the *Maximum Miller index* to 6. Rotate the crystal by moving the small inner circle around the outer circle below the crystal image. Select faces by moving the T-tool until the top of the "T" is along the edge of the crystal.

By hitting enter, the indices of the face and the distance from the face to the center of the crystal are added to the list at the right. Continue rotating the sample and identifying faces until the black outline around the crystal is a reasonable approximation of the crystal's shape and the "Closed" box in the lower right of the screen says "Yes". Hit the *Remove Invisible Faces* button below the list of faces. Save the faces to the *project*f.p4p file using *Export > P4P File*.

Absorption Correction or Scaling

There are two programs that were written to provide these corrections, SADABS and TWINABS. The program SADABS is used for regular (untwinned) data sets; the program TWINABS is used for data sets showing non-merohedric twinning or splitting that required more than one orientation matrix for integration. Both programs follow a similar path to determine appropriate absorption correction. Note that this correction is generally not needed to solve untwinned data for structure solution; however, TWINABS must always be run to work with twinned/split data sets. The final data set ***must*** be corrected for absorption before publication.

One of the first questions asked by the program is the name for the new listing file describing all steps taken by the program. Typically, this file should be named "*project*.abs". The program will ask for the symmetry of the crystal – on the first pass through this program, be sure to select the proper Laue symmetry. If the Laue symmetry is not the point group of the space group, and most of the data has a redundancy > 6 then rerun the absorption correction later using the correct point group. The name of the data files is typically input as "*project*". From this point forward, default responses can typically be chosen. If faces of the crystal were indexed, be sure to include the file with the faces and have the program determine the best value for the linear absorption coefficient of the sample.

In the SADABS program, the output file is usually *project*.hkl that is in SHELX hklf-4 format. In the TWINABS program, it is possible to output both an "hklf-4" format file and an "hkl-5" format file (see the refinement program for details). Typically the hklf-4 file is saved as *project*.hkl and the hklf-5 file should be saved as *project*.hkl5. Note that the hklf-4 file from TWINABS will typically only use the strongest twin domain of the crystal, and the hklf-5 file should be created from all domains. Twinned data sets are typically solved and partially refined with the hklf-4 format file; and then the refinement is finished using the hklf-5 file by renaming *project*.hkl5 to *project*.hkl and adding the appropriate number of dummy variables on a BASF instruction.