MOLBIOCH 840 Practical Protein Crystallography
Use of CrystalClear Software to Collect and Process X-ray diffraction data
Chuck Bell

Overview of X-ray Diffraction Data Collection. In data collection, a crystal is mounted on the Phi axis (or spindle), which leaves the crystal upside down on our setup. The crystal can either be mounted in a capillary tube for room temperature data collection, or more typically in a mounted cryoloop pin for data collection at about -180°C. After the crystal is mounted on the instrument, the video screen is used to align the crystal to the presumed center position of the x-ray beam (the rotation center of Phi and the horizontal line on the video screen). After alignment, the detector distance is set, the Phi axis is locked, the enclosure doors are closed, and the shutter switch is moved to the "EXT" position. The crystal is then ready to collect a few initial "oscillation" images, one at 0° and a second at 90°. During collection of the oscillation images, the crystal is rotated about the Phi axis for about 1° as the crystal is exposed, usually for 2-5 minutes per image. The initial images are inspected for crystal quality, and the software then automatically indexes the crystal. The indexing procedure determines the unit cell dimensions and lattice type of the crystal, as well as its orientation relative to the x-ray beam. After indexing, the software determines the best starting and ending rotation angels for the data collection, and a sweep of several (~100) oscillation images is collected, usually overnight.

Overview of X-ray Diffraction Data Processing. In data processing, the set of oscillation images is converted into a file that contains the intensities and sigmas of all of the symmetrically unique hkl reflections that were measured. There are essentially four steps involved in data processing (1) collection of initial images, crystal evaluation and indexing, (2) collection of the full set of oscillation images, (3) integration of reflection intensities, and (4) scaling, merging, and averaging of reflection intensity data. These steps are described in detail below. The output is a file called "ScaleAveraged.ref", which has header information (space group, cell dimension) and then h,k,l, Intensity, and sigma for all of the unique reflections (usually about 5-20 thousand) that were measured.

CrystalClear Software. The CrystalClear software package comes with the R-AxisIV++ detector purchased from Rigaku-MSC. The software runs on Windows 2000 or XP machines, but requires a "dongle" (thing you plug into the back for licensing purposes) to run it, and there are only a few of these available. So really the software is normally run on the Windows 2000 computers in the x-ray lab that control the R-AxisIV++ detectors. The CrystalClear package is a GUI that is based on a program called d*Trek, which is the main program that does the data processing (crystal indexing, integration of reflections, scaling and merging of data). In general the software is very user friendly, and guides you through the necessary steps in data collection and processing. In most of the steps the default parameters are usually good as starting points, and the user can usually just simply click "OK" or "Run" to successfully accomplish all of the steps.

There is a CrystalClear manual in the x-ray lab that is a detailed reference for you. This handout is intended as condensed version of the essentials needed to go through the process with a basic understanding of what is going on. In this handout, a one-page quick flow-chart for the steps involved in data collection is given, followed by a more detailed explanation of each step.

Quick Guide to Collecting and Processing an X-ray Diffraction Data Set

1. Log on to computer (Administrator/raxis4++).
2. Double-click CrystalClear icon on Desktop, Log on with your Username or "guest" (no password needed).
3. Select a Project and type in a new Sample name to create a directory for storing images and processing files.
4. At "Do you wish to initialize" click NO.
5. In Setup, change crystal to detector distance to what you plan to use for data collection.

6. Mount crystal and click OK (you can actually mount crystal ahead of time).
7. Lock Phi axis and click OK. Set detector distance, close all doors, make sure shutter will open, and move shutter switch to "EXT" position.
8. In Initial Images, select "Screen Schedules Two: 0,90". Choose exposure time and oscillation width. Click Run. The images will be collected. Make sure shutter opens. The images will automatically open in the image display window. Examine images for single crystal and resolution limit.

9. In Find Spots, select resolution limits (in Advanced) and click Run. You will see blue circles appear over spots that are identified. Check that they are real.
10. In Index Spots, select "Unknown", and click Run. The auto-indexing procedure will identify possible solutions, and highlight the highest symmetry one with an acceptable residuals score. Click OK if you are happy with this choice.
11. In Refine Cell, click OK. The indexed crystal and detector parameters will be refined, check that the detector distance doesn't change too much. Note the residuals at the top right. Click Run a few times and then OK when they stop decreasing.
12. In Predict Spots, select the images (usually 1-2), resolution limits and click Run. You will see blue circles where spots are predicted to occur based on the refined indexing solution. Check that the pattern of blue circles matches the pattern of observed spots. Adjust the mosaicity to improve the match.
13. In the "Processing State Display", select the correct space group symmetry if you know it, and click Save. Otherwise, leave at default, which should be lowest symmetry within the indexed lattice type.

14. In Strategy, click Run. This will select the best data collection parameters (start/end angle) based on the indexing solution and symmetry, and feed them into Collect Images.

15. In Collect Images, select exposure time, oscillation width, expand start/end angle as desired, and click Run. This will start the image collection. Images will be launched in display window as they are collected.
16. Use mask features of image display window to mask out beamstop shadow region.

17. In Integrate Reflections, choose appropriate resolution limits, and click Run. The images will be integrated as they are collected. You will see blue circles over the spots. Usually for a full data set this will go overnight.

18. In Laue Check, select OK at chosen Laue symmetry group (highest symmetry with acceptable R-merge). Do the same for Centricity Check and Space Group Check. Change space group if necessary.

19. In Scale and Average, adjust parameters if needed, and click Run. You will see a log file being written that gives you statistics on the intensity measurements. Choose final resolution limits, and Run again. Your output files will be "ScaleAveraged.ref" (hkl data), "dtscaleaveraged.log" (log file with statistical information on data quality), and "dtintegrate.log" (logfile with integration statistics, including systematic absences information).

20. Backup your data using the DVD drive. Delete your images from the computer.

SETUP

CRYSTAL EVALUATION
- Mount Crystal
- Initial Images

ASSIGN UNIT CELL
- Find Spots
- Index Spots
- Refine Cell
- Predict Spots

STRATEGY

COLLECT IMAGES

INTEGRATE REFLECTNS

ANALYZE DATA
- Laue Check
- Centricity Check
- Space Group Check

SCALE AND AVERAGE

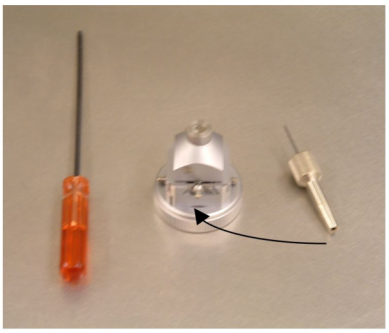
Details of Data Collection Procedure

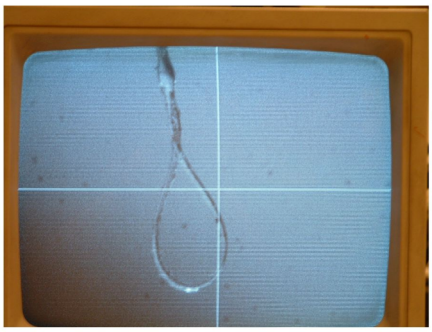
Mounting and aligning the crystal. The first step in data collection is mounting and aligning the crystal. There are two ways of mounting the crystal, capillary tube and nylon loop.

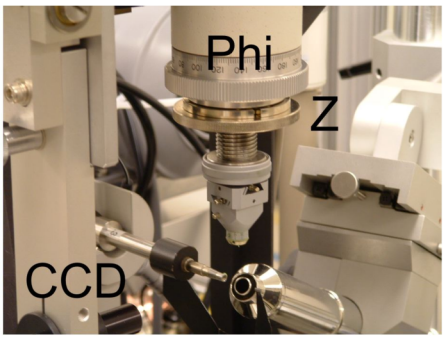
Capillary tube mounting. This is for room temperature data collection (not using the cold stream). In this case the capillary tube is secured onto the goniometer head with the mounting clay, and the goniometer head is then screwed onto the phi axis.

Nylon loop mounting. The second way of mounting the crystal is with the copper pins containing the nylon loops used for cryo-crystallography. In this case, the crystal pin (usually in a plastic vial) is retrieved from the Dewar containing liquid nitrogen using the clamps, and placed hanging upside down onto the magnetic base of the goniometer head, which is positioned on the Phi axis such that after removing the vial the crystal will be in the cold stream. This will be demonstrated in class. The idea is that the crystal is kept at cryogenic temperatures as it is transferred from the vial to goniometer head.

Crystal alignment. After the crystal is mounted it is necessary to align it to the beam center. This is done using the video screen that shows the image of the crystal from the CCD camera. First, the Z translation of the crystal is adjusted to center the crystal vertically, so that the crystal is centered on the horizontal axis on the video screen. Next, the crystal is centered to the rotation center of the phi axis, which should be close (but not always directly on) the vertical axis on the video screen. This is done by using the alignment tool to adjust the two perpendicular translations on the stage of the goniometer head. With a square face of the goniometer pointed towards the CCD camera, adjust the goniometer stage so that the crystal is centered in that direction. Then rotate the crystal 90° and make the similar alignment for the perpendicular direction. In principle the crystal should now be aligned, but this should be repeated a few times for fine adjustment. In the end, the crystal should stay in position as the phi axis is rotated (i.e. it should be superimposed on the rotation center) and should be centered in the z direction. Note that the user is relying on the instructor to properly align the video camera and phi axis to the x-ray beam.





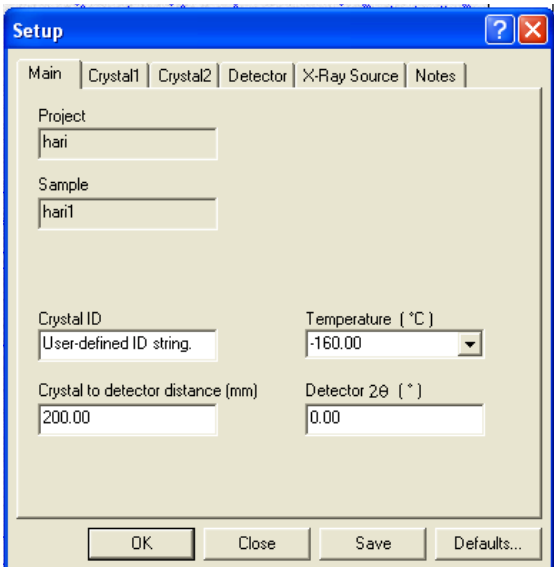


Setting up for data collection. After aligning the crystal, you need to lock down the spindle axis, set the correct detector distance, and close all of the plexiglass doors to the port you are using (right or left). Then you can test to make sure that the shutter opens manually (press switch upward to "Open" position and verify that the light comes on), and finally flip the shutter switch to the "EXT" (down) position, so that the shutter can be externally controlled by the control computer. At this point you should also check to make sure that the beam stop is correctly in place, and check to make sure the helium is flowing through the mirrors.

Choosing the detector distance. The optimal detector distance depends on how far you expect the crystal to diffract. A crystal that diffracts to higher angle (higher resolution) will require a shorter detector distance. A distance of about 200 mm means that the edge of the detector corresponds to about

2.3Å. In general, it is best to have the detector distance set so that the very edge of the visible diffraction pattern is just inside the edge of the detector. The larger detector distance (while still observing all of the visible data) is preferable because the spots will be spread out (and thus easier to integrate), and the signal to noise will be improved because the noise falls off faster than the signal with increasing distance. This is offset somewhat however, by increased absorption of the diffracted x-rays by the air as the detector distance is increased.

SETUP. Now that your crystal is mounted and aligned, you can begin the data collection process. Double-click on the CrystalClear icon on the desktop of the control computer, and login using your own account if you have one, or the "guest" account if you don't. There is no password for the "guest" username. The software will ask you to select a "project" and a "sample". You can either select an old project (one previously created) or a new one. The sample is usually new; generally each crystal that is mounted will have its own sample name. The "project" and "sample" are simply names that create directories for storing the images and the processing files. You will also select "Screen, Collect, and Process" meaning that you want to collect data on a new crystal, rather than simply process data that was collected previously. Clicking OK will launch you into the Setup window, and the software will ask you if you wish to initialize the detector. Click NO. In Setup, you can enter lots of data parameters, but most these are just for your own bookkeeping purposes. The one parameter you will want to be sure to enter is the distance, as that will affect image processing. The software will tell you to mount the crystal and click OK and then lock the Phi and click OK. You should have already done these earlier, so click OK. Note: if the Phi axis is not locked, then the instrument will attempt to rotate the Phi axis to the zero position, but it will not be able to, and it will spin it for several minutes and then report an error (this gets annoying, so try to remember to lock down Phi).



CRYSTAL EVALUATION. Once the Setup is complete, you will go to the "Initial Images" window, and under "Screen Schedules" select "Two: 0,90" This will set the instrument to collect two oscillation images at 0° and 90° on the Phi axis. A 1° oscillation image starting at 0° means that the instrument will oscillate Phi back and forth from 0-1° as the image is being collected. There are several parameters you can change, for example exposure time (default 1 min) and oscillation width (default 0.5°). For most cases it is OK to leave these at the default values, though often for small crystals it is appropriate to increase the exposure time to 2-10 minutes. For crystals with small unit cells (if you happen to know what it is) choose an oscillation width of 1-2°. Click OK, and the software will control the detector to collect the initial images. First the detector will initialize itself, then you will see the yellow box on the lower right that will indicate that the shutter is open, and then you know your image is being collected. You should check to make sure that the light above the shutter (inside the enclosure) actually goes on.

Viewing the initial images. The instrument should automatically launch the images in the image display window as they are collected. This window has several tools for viewing the images. You can click on the image and read out the resolution of a particular spot in the diffraction pattern. You can zoom in and out to look at a particular area in detail. You can adjust the contrast for a particular region of the pattern you are interested in. All of the tools have squares with symbols at the top of the image

display window. Placing the cursor over a particular square (without pressing on it) and waiting a few seconds will launch a description of what that tool is for. There is an image number field on the top right in which you can change the image number (1 or 2 at this point) to switch images.

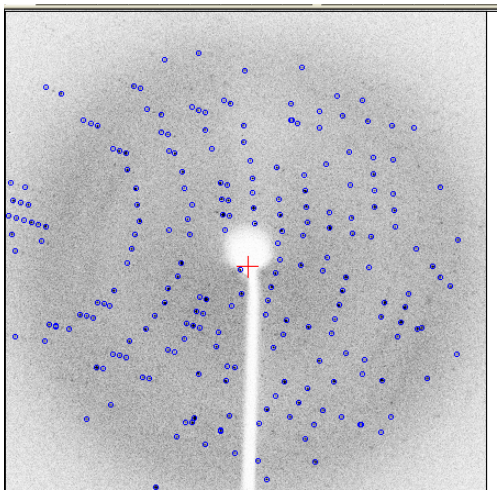
Assessing the initial images: from the initial images you will want to get an idea of how the crystal diffracts, whether it is salt or protein, and whether or not it is single.

Salt or protein. Because salt crystals have small unit cells, they will have only a few (~5) widely spaced spots that will be very intense and only be at high resolution (5Å or so and higher). Salt crystals will NOT give diffraction spots near the center of the detector (low angle, low resolution spots). Protein crystals on the other hand will have a pattern of closely spaced spots, the most intense of which will be at low resolution, near the center of the detector.

Resolution limit. All good protein crystals will have a diffraction pattern with strong, intense spots at low angle (in the middle of the detector), with the intensity of the spots fading out at higher angles. Clicking the cursor at the radius of the image where you just barely see spots (where the pattern fades out) will tell you the approximate resolution limit of the crystal. In general, if you are going to go ahead and collect a full data set on a particular crystal, it will be best to reset the detector distance so that the edge of the diffraction pattern is just inside the edge of the plate.

Single crystal or not. The diffraction pattern of a crystal should look like a lattice of spots, where only certain regions of the lattice (usually rings or "lunes" of spots) are visible. Occasionally crystals may be twinned or cracked. A common example is crystals that grow as thin plates stacked together. This can lead to two or more diffraction patterns superimposed, where the spots will fall on different lattices, or spots that are on the same lattice but doubled or streaked. It really just takes experience in distinguishing non-single from single crystal diffraction patterns. Ultimately the indexing (see below) will tell you if the crystal is single or not. In some cases it is possible to get usable data from a split crystal, but this is not ideal- you should try another crystal if you have one.

ASSIGN UNIT CELL. The objective here is to determine the unit cell dimensions (a,b,c and α,β,γ) and lattice type of your crystal, as well as the three rotation angles that define how the crystal is oriented. This can be determined (by a very fancy algorithm!) from the positions of the spots in the observed diffraction pattern. Briefly, the software is trying to construct a lattice in reciprocal space from the positions of the spots observed on the images (known detector x,y position and angle in Phi). Once a satisfactory reciprocal lattice is identified, the crystal lattice in real space can be easily calculated. Note that it is essential that a reliable beam center position be known in order for the indexing to work. The beam center position (seen as a red cross in the center of the image) is measured by the instructor after each beam alignment, stored in a particular system file, and read in automatically during the indexing routine.

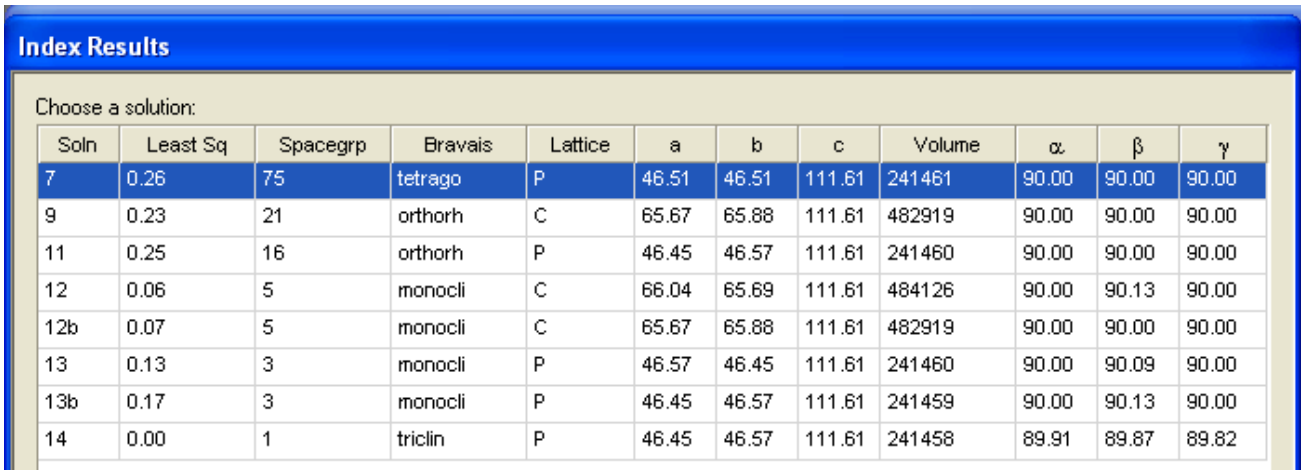


Find Spots. After collecting the initial images, the CrystalClear software will automatically open the "Find Spots" window. Clicking OK will start the Find Spots routine. Briefly, the software is searching the image for clusters of pixels that have high intensity values. If you zoom all the way into a particular spot you can see what the actual pixel intensity values are for that spot. The software will indicate the spots that are found on each image by drawing a blue circle around that spot. For a crystal that is likely to be successfully indexed, you should have a minimum of about 20 spots per image to

work with. For crystals that diffract to low resolution, you may want to go into the "Advanced" Window, and set the resolution limits to something that corresponds to the actual observed diffraction limits (for example, set resolution to 50.0 and 6.0). This will help to ensure that the spots that are picked are from the protein diffraction pattern, and not noise from ice ring diffraction or whatever.

Index Spots. After the Find Spots routine, the software will automatically launch the AutoIndexing window. Choose "Unknown Spacegroup" and click Run. You will see a screen showing a logfile that is being created, and then an "Index Results" window containing a table of possible solutions will appear, with the suggested solution highlighted in blue. The objective here is to choose the solution with the highest symmetry that has a reasonable (low, say below 2.0) "Least Sq" fitness score (the software does this for you when it highlights a particular solution). The rationale in choosing the highest symmetry lattice is that if the lattice constructed from the observed spots has angles of 90, 90, 90, then the true lattice is likely to have at least orthorhombic symmetry. In other words, the angles are not likely to be 89.5 (just by chance close to 90) and not be exactly 90. In any case, choose the solution with the highest symmetry and click OK.

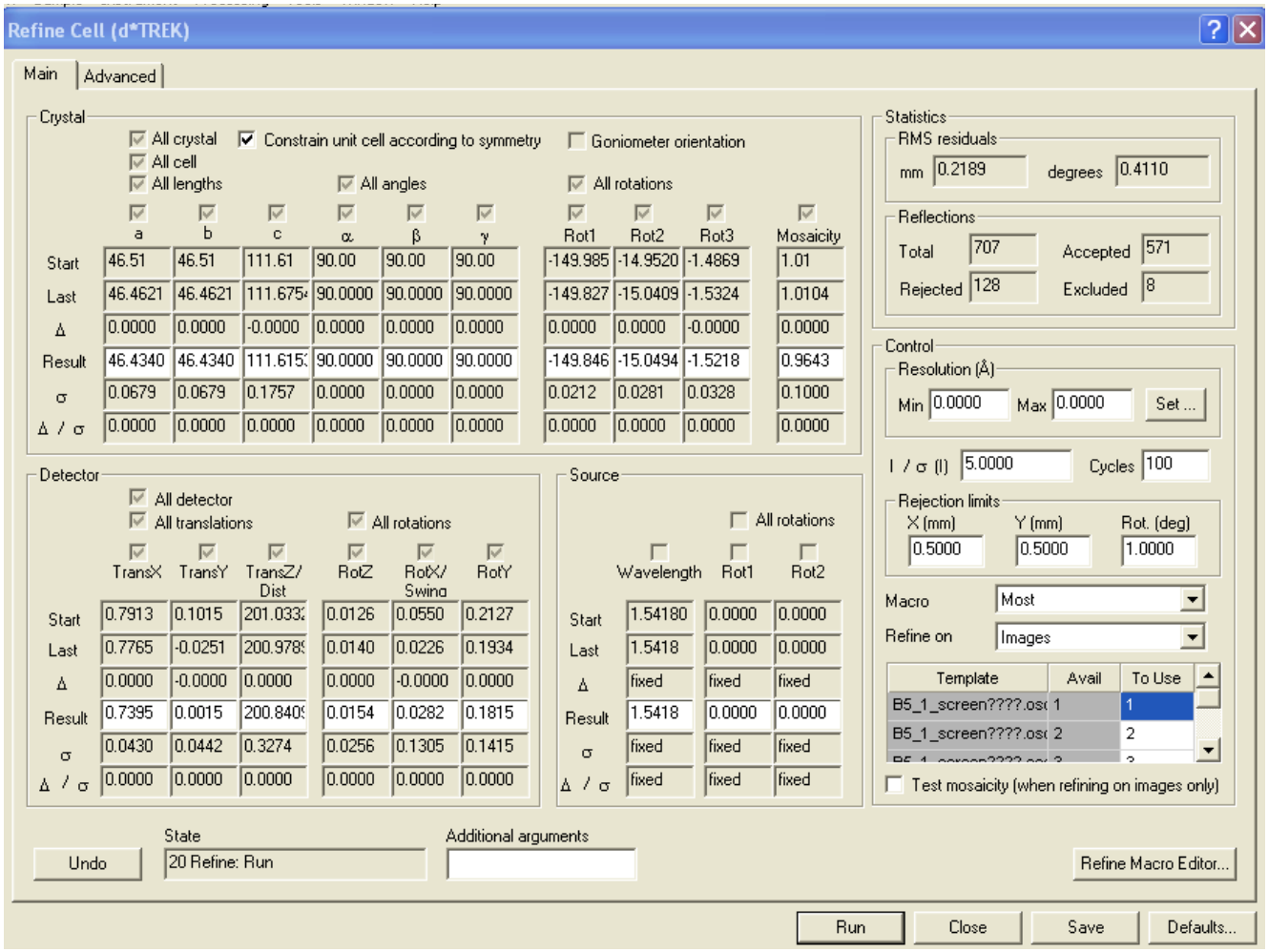
For a good crystal, the indexing should work on the first try every time. If indexing fails, it is usually because your crystal is not single, or there are not enough strong spots to work with. The best thing to try is to make sure you are picking spots only in the resolution range of the observed diffraction pattern. You can also adjust the spot picking criteria in the Advanced section of the Find Spots window, hope to get more spots found, and retry the indexing. Normally for a decent crystal, even if it only diffractions to only 7Å or so, indexing should work easily.



Soln	Least Sq	Spacegrp	Bravais	Lattice	a	b	c	Volume	α	β	γ
7	0.26	75	tetrago	P	46.51	46.51	111.61	241461	90.00	90.00	90.00
9	0.23	21	orthorh	C	65.67	65.88	111.61	482919	90.00	90.00	90.00
11	0.25	16	orthorh	P	46.45	46.57	111.61	241460	90.00	90.00	90.00
12	0.06	5	monocli	C	66.04	65.69	111.61	484126	90.00	90.13	90.00
12b	0.07	5	monocli	C	65.67	65.88	111.61	482919	90.00	90.00	90.00
13	0.13	3	monocli	P	46.57	46.45	111.61	241460	90.00	90.09	90.00
13b	0.17	3	monocli	P	46.45	46.57	111.61	241459	90.00	90.13	90.00
14	0.00	1	triclin	P	46.45	46.57	111.61	241458	89.91	89.87	89.82

Refine Cell. After an initial indexing solution is chosen, a Refinement window is automatically launched. The initial indexing solution consists of the crystal parameters (a,b,c, and α,β,γ) as well as the crystal orientation parameters (Rot1,Rot2,Rot3). The solution can be further refined if the detector parameters (distance, beam center, tilt, twist, etc.) are allowed to vary. Basically, the software simultaneously varies all of these parameters (crystal parameters, crystal orientation, detector parameters, beam center) to optimize the intensity/sigma for as many of the observed spots as possible (especially the most intense spots (say I/sigI > 5), which are the most reliable to use as a basis for refinement). After clicking OK to run a round of refinement, and repeating a time or two, you will see updated (refined) values for all of these parameters. You will also see two numbers at the top right of the Refinement window that give you some idea of the fitness of the solution. These are the refinement residuals expressed in mm (x,y position of spot on detector) as well as degrees (position of spot in image rotation, i.e. the third dimension). Scores of about 0.2 mm and 0.5° are pretty good. Also check to see that only a small fraction of the picked spots are getting rejected (based on comparing observed

and predicted positions). This can be seen under “Reflections” which gives total # of picked spots, # accepted, # rejected (too far out of position), and # excluded (based on I/sigI criteria), and also on the image display window, which shows blue circles for accepted, green for excluded, and red circles for rejected reflections.

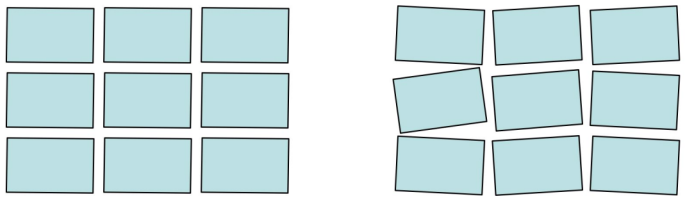


Processing State Display. After refinement, a Processing State Display window is launched that allows you to enter further space group information about you crystal. The indexing can determine which lattice group your crystal belongs to, but usually not what the Space Group is. This can only be done after collecting a complete data set. For example, the indexing might determine Primitive trigonal/hexagonal, but can’t distinguish between P6, P622, P61, P3, etc. If however, you know what the actual space group is from previous data collections, then you should choose it from the pull-down menus. This will be useful for the Strategy and Integration routines, described below.

Predict Spots. After an indexing solution has been refined, it must be tested against the observed diffraction patterns on the initial images. In the Predictions window, select the images you wish to show predictions for, and then click OK. The program will now calculate the predicted positions of spots on the initial images based on the refined indexing solution and detector parameters. The predictions are shown as blue circles. You will also see some red circles that are rejected for various reasons. Basically, you want to check to see that the general pattern of the blue circles matches the observed diffraction pattern. Check this for both of the initial screen images taken at 0 and 90. Of course you may see some blue circles that do not coincide with observed spots, especially at higher

resolution (higher angle), but there will always be some weaker reflections. There may be some observed spots that do not coincide with blue circles, especially at low resolution, but don't worry about this if there are only a few of these. If there seems to be a systematic difference between the patterns of predicted and observed spots, then there may be a problem, and the refined indexing solution may not be correct.

The Predictions window is a good place to get a feel for your crystal **mosaicity** (also called mosaic spread). In the input parameters to the Predictions window, you can change the mosaicity. The mosaiciy has units of degrees and basically reflects how tightly arranged the unit cells of the crystal are to one another. A low value for the mosaic spread (say 0.2°) generally means a tight arrangement of unit cells, whereas a higher value (say 1.0°) indicates a higher degree of slop. Imagine a mosaic in the New York subway where the squares are not perfectly arranged relative to one another- the higher the degree of imperfection, the higher the mosaicity. A crystal with a low mosaicity will have small, tight spots, and thus with all other things being equal there will be fewer spots observed per image at a given oscillation width (think of the Ewald sphere of reflection). A crystal with a higher mosaicity will have larger spots, and hence a greater number of spots will be observed per image (the rings or "lunes" of spots will be wider). Thus, in running the Predictions, you can adjust the mosaicity to optimize how well the predictions fit the observed pattern. Note that although the spots are observed on the 2D detector, they are actually three dimensional (roughly spherical), such that a given spot may be observed on multiple successive images. This will be explained in more detail below under integration.



Low mosaicity (0.2°) High mosaicity (1.0°)

STRATEGY. Once you have settled on an indexing solution, you are ready to collect a full sweep of oscillation images in an attempt to get a complete data set, which is usually no problem if your crystal is frozen. The successive images will be collected by rotating around the Phi axis. For example image #1 might cover 0-1°, image #2 covers 1-2°, and so on for typically 100 or so images. For each image, the crystal is oscillated back and forth several times as the image is exposed. In Strategy, the software determines the minimal number of degrees that need to be covered to get a complete data set (based on the lattice symmetry or the space group symmetry if you have chosen it), and what value of Phi to start at. That is, for a chosen oscillation width, say 0.5°, the software will determine, starting angle and number of images. For high symmetry space groups (e.g. cubic, hexagonal) you will generally need to collect only about 60° or so of data (the actual amount will depend on how the crystal axes are oriented relative to the beam). For low symmetry space groups (triclinic, monoclinic) you may have to collect up to 90-180° or so of data. Collecting more than enough data, to get a high redundancy of measurements for each hkl reflection, is always a good thing. It is simply a matter of how much time you have to collect your data.

COLLECT IMAGES. Running Strategy will automatically feed the optimal image collection parameters into the image collection window. The two parameters you might want to change are oscillation width and exposure time. For oscillation width, usually 0.5° is optimal, though if you have a small cell, or if you are doing room temperature data collection, you may wish to change this to 1-2°. Typically 5 minutes is a reasonable amount of exposure time, though anywhere from 2-20 minutes is

reasonable. What you choose may depend on how well your crystal diffracts, how many images you have to collect, and when you wish the data collection to end. The software will indicate how much disk space the data set will take, and how much disk space is available. It will also indicate the expected time of completion. Clicking Start will begin the process. Your images will appear in an image display window and be updated as they are collected.

Masking out the beam stop shadow. The beam stop leaves a shadow on the detector consisting of a circle in the center and an arm extending from the center down to the bottom of the image. You will want to box out this region of the detector so the software knows not to look in the shadow region for spots to integrate. Reflections that would normally be in those regions are blocked by the beam stop, and will be lost in the shadow. To do this, use the circle and square tools to draw a mask covering the shadow region, then write the mask to a file to be used during image processing.

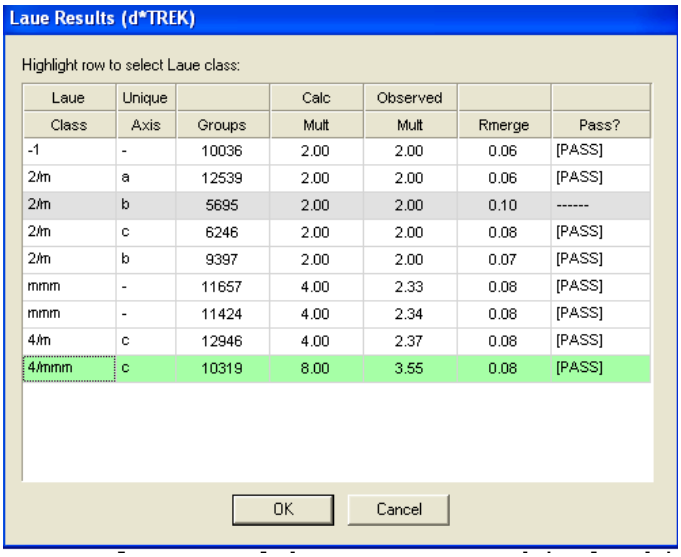
INTEGRATE. Open the Integration window, select the images you wish to integrate (usually the default is all of those scheduled), select reasonable resolution limits for your observed diffraction pattern, and click OK. The software will begin integrating reflection intensities from the images as they are collected. It is OK to set Integration up at the beginning of your data collection, let it run overnight as your data set is being colleted, and come in the next morning to continue processing with Scale/Average.

Integration extracts intensity values for all of the reflections from each image by summing up pixel values for each spot. From the indexing parameters that were determined and refined previously, the software knows where to look on each image for all of the hkl reflections that are observed. This is exactly what was being done in the Predictions routine described above. The software draws a blue circle around each predicted spot location, and integrates (sums up) the pixel values that fall within the boundary of a particular spot. The software also measures the background intensity in the region just outside the spot, and this is used to determine the sigma value associated with each spot. As the software goes along integrating all of the spot intensities, it actually optimizes the crystal and detector parameters to obtain the highest intensity/sigma ratios for the strongest spots- this will in principle lead to optimized integration of the weaker spots as well. It does this refinement/integration procedure in groups of about 4 images. Many reflections actually extend across one or more successive images, i.e. the spots are three-dimensional. For these “partial” observations, the intensities extracted from successive images on which the reflection is observed are summed to determine the complete intensity for that reflection.

The software also does something called **profile fitting** to define the spot shape. Basically this allows for the spot to be non-spherical. Different regions of the detector (corresponding to different angles of diffraction) tend to have similar spot shapes that may deviate from spherical, especially at high angles. Thus profile fitting finds the optimal spot shape for a given region of the detector, and this is continuously refined as the software goes along processing all of the images. In the end, the Integration routine writes a file containing all of the observed hkl reflection intensities and their associated sigmas. Note that some of the reflection intensities are based on integrations of spots from multiple images, whereas others may have been observed on a single image. Assuming you have collected enough data, a given hkl reflection should have had its intensity measured multiple times.

ANALYZE. After integration, you can now run a few checks on your observed reflection intensities, these are Laue, Centricity, and Space Group. You can let the program go through the routines, click OK, and then go into Scale and Average. Here is an explanation of what is going on for these routines.

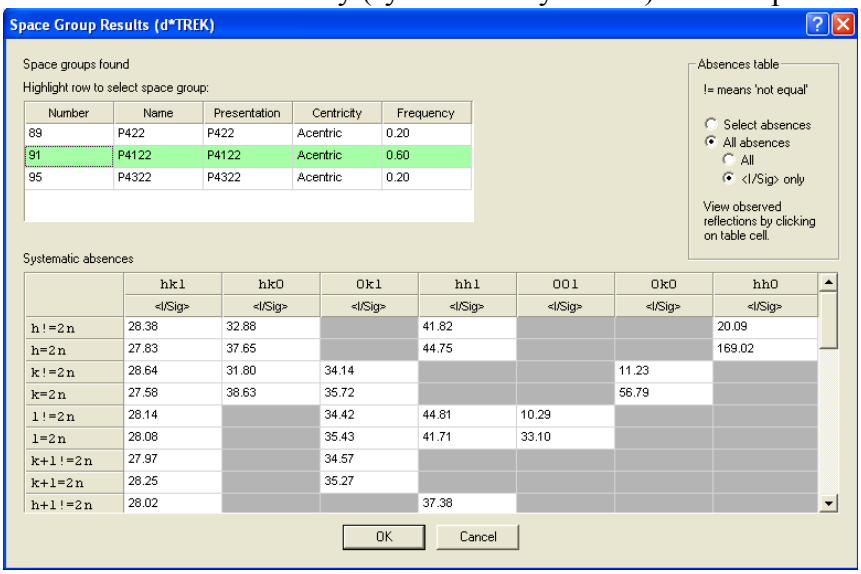
Laue Check. This determines the Laue symmetry of your diffraction pattern. The diffraction pattern has certain symmetries that are related to the symmetry of crystal. For example, a 2-fold crystallographic symmetry axis leads to mirror (m) symmetry in the diffraction pattern. In short, the diffraction pattern has all of the symmetry of the crystal, plus a center of inversion. Combining inversion symmetry with a 2-fold axis leads to mirror symmetry. Reflections related by Laue symmetry will have the same intensity values. The indexing from the initial two images determines the spot positions, which allows the lattice type to be determined. However, during indexing the intensities of possible symmetry related reflections are not measured, and thus the full symmetry of the crystal is not known at the indexing stage. After collecting all of the images and extracting the intensities of all of the reflections, the intensities of possible symmetry related reflections can be compared, and thus Laue symmetry (symmetry of the diffraction pattern) can be determined. The software highlights the highest-symmetry Laue class that has a reasonable R-merge score (below about 15%). R-merge will be elaborated on later, but briefly it indicates how well intensities of symmetry related reflections agree with one another. A value of 5-10% is the typical error in measurement of identical or symmetry related reflections.



Laue Class	Unique Axis	Groups	Calc Mult	Observed Mult	Rmerge	Pass?
-1	-	10036	2.00	2.00	0.06	[PASS]
2/m	a	12539	2.00	2.00	0.06	[PASS]
2/m	b	5695	2.00	2.00	0.10	------
2/m	c	6246	2.00	2.00	0.08	[PASS]
2/m	b	9397	2.00	2.00	0.07	[PASS]
mmm	-	11657	4.00	2.33	0.08	[PASS]
mmm	-	11424	4.00	2.34	0.08	[PASS]
4/m	c	12946	4.00	2.37	0.08	[PASS]
4/mmm	c	10319	8.00	3.55	0.08	[PASS]

Centricity Check. The software will determine the centricity.

Space Group Check. This makes a final determination of the space group for your crystal based on the indexing solution (lattice type), Laue symmetry, and one final consideration related to systematic absences. Often it is necessary to distinguish a pure rotation axis (say 2-fold) from a screw axis (say 2_1). These two will have very different packing, but will have exactly the same lattice and Laue symmetry. The only way to distinguish between these two is by systematic absences along a particular axis of the reciprocal lattice (h, k, or l). For example, a 2_1 axis along the *a* crystal axis will lead to h not equal to 2n (i.e., odd) reflections with zero intensity (systematically absent). In the Space Group routine, the software goes through and searches along the reciprocal axes for systematic absences to determine if the symmetry axes of the crystal are pure rotation axes or screw axes. The table with the measured intensities of the axial reflections (h,0,0; 0,k,0; 0,0,l) is listed near the end of the dtintegrate.log file (the log file that was written by the Integration routine). After running Space Group, select the space group (if you know it) or otherwise take the default, and hit OK.



Number	Name	Presentation	Centricity	Frequency
89	P422	P422	Acentric	0.20
91	P4122	P4122	Acentric	0.60
95	P4322	P4322	Acentric	0.20

	hkl <I/Sig>	hk0 <I/Sig>	0kl <I/Sig>	hhl <I/Sig>	00l <I/Sig>	0k0 <I/Sig>	hh0 <I/Sig>
h!=2n	28.38	32.88		41.82			20.09
h=2n	27.83	37.65		44.75			169.02
k!=2n	28.64	31.80	34.14			11.23	
k=2n	27.58	38.63	35.72			56.79	
l!=2n	28.14		34.42	44.81	10.29		
l=2n	28.08		35.43	41.71	33.10		
k+l!=2n	27.97		34.57				
k+l=2n	28.25		35.27				
h+l!=2n	28.02			37.38			

Here is the table in dtintegrate.log that gives the information on the systematic absences:

```
h00, 0k0, 00l reflections
-------------------------------------------------------------
    h     k     l Intensity SigmaInt I/sigI  Pix0  Pix1  Angle
-------------------------------------------------------------
   -3     0     0        54        2   22.4  1554  1702  97.91
   -4     0     0      5733       24  238.2  1570  1768 100.08
   -5     0     0        43        3   15.1  1587  1834 105.59
   -6     0     0       511        9   54.4  1600  1903 108.18
   -7     0     0        84        4   21.7  1615  1970 113.17
   -8     0     0       102        4   24.2  1628  2041 117.41
   -9     0     0        84        4   22.1  1638  2113 121.62
  -11     0     0        65        4   16.9  1650  2265 131.30
  -12     0     0      1998       24   84.2  1648  2346 135.44
  -13     0     0       109        5   20.6  1643  2429 143.55
  -14     0     0       190        6   30.4  1627  2516 150.28
    0    19     0         0        0    0.0    14  1687 132.75
    0    18     0        87       14    6.4   119  1673 133.74
    0    17     0       464       26   17.8   219  1660 134.76
    0    16     0       826       39   21.2   312  1650 135.43
    0    15     0       204       16   13.1   405  1637 136.75
    0    14     0       239       19   12.9   492  1627 137.45
    0    13     0       557       33   16.8   577  1616 138.73
    0    12     0      1658       56   29.4   659  1606 139.46
    0    11     0       255       15   17.0   736  1596 140.51
    0    10     0         0        0    0.0   812  1587 141.75
    0     9     0         0        0    0.0   886  1579 142.75
    0     8     0       244       16   15.6   958  1570 143.20
    0     7     0         0        0    0.0  1029  1563 144.43
    0     6     0       315       16   20.3  1098  1554 145.09
    0     5     0        91       10    8.9  1168  1547 146.75
    0     4     0      5103       46  109.9  1235  1539 147.14
    0     3     0         0        0    0.0  1303  1531 148.49
    0     2     0      3292       26  128.4  1369  1523 149.04
    0     1     0         0        0    0.0  1436  1516 150.57
    0    -1     0         0        0    0.0  1568  1501 152.23
    0    -2     0      3444       26  131.1  1635  1493 152.92
    0    -3     0       105        7   15.2  1701  1486 153.77
    0    -4     0      5608       48  116.8  1770  1477 154.82
    0    -5     0        85        9    9.3  1836  1470 156.25
    0    -6     0       524       21   24.4  1904  1462 156.93
    0    -7     0         0        0    0.0  1974  1454 158.25
    0    -8     0       107       11    9.6  2046  1446 159.25
    0    -9     0         0        0    0.0  2118  1437 159.97
    0   -10     0       905       30   29.7  2190  1429 160.70
    0   -11     0         0        0    0.0  2268  1420 161.70
    0   -12     0      1849       58   31.9  2346  1409 162.56
    0   -13     0         0        0    0.0  2427  1402 164.25
    0     0   -39       248       25   10.0  2685  1207  75.71
    0     0   -40      4610       92   49.9  2724  1196  75.99
    0     0   -41       141       19    7.3  2763  1185  77.25
    0     0   -42       120       21    5.8  2803  1174  77.75
    0     0   -43       220       27    8.2  2844  1163  77.71
    0     0   -45       345       31   11.3  2929  1139  79.03
    0     0   -44      8396      136   61.7  2887  1152  77.70
    0     0   -46       357       29   12.2  2973  1127  79.25
-------------------------------------------------------------
Number of reflections written in 'zone.ref': 51
```

In this case, the data indicate a 4_1 axis along *c*, because of the 00l reflections that were measured only the 4n ones are intense (0,0,-40 and 0,0,-44). Similarly, the data indicate a 2_1 axis along *b*, because only the even 0l0 reflections are intense. This information (combined with the previously determined lattice and Laue symmetry) indicates that the space group is either $P4_12_12$, or $P4_32_12$. These two cannot be distinguished from one another at this point, but fortunately $P4_12_12$ occurs far more frequently and is thus more likely.

SCALE AND AVERAGE. In Integrate, you have extracted the intensities of a total of about 50,000 or so hkl reflections, but there may only be 10,000 or so of these that are unique (the exact number will depend on your crystal symmetry, resolution limit, # of images, etc.). In other words, many of these will be separate measurements of the same hkl reflection or of symmetry related hkl reflections (in either case the intensity values should be exactly the same and the values can be merged and averaged). The final step in data processing is to merge all of the redundant intensity measurements of the hkl reflections into single, averaged values for the intensity of each symmetrically unique reflection hkl. Scale and Average will output the data into a file called **ScaleAveraged.ref**.

Before Averaging, the software automatically scales all of the images relative to a particular image (usually the first image). There are several reasons why the images might not be on the same scale. For one, as the crystal is being rotated the number of unit cells within the x-ray beam may change. There may also be a slight deterioration in the crystal quality due to radiation damage, differences in absorption of the diffracted x-rays at different angles, or a change in the intensity ("drift") of the incident x-ray beam over time. For all of these reasons, the reflections on image #1 might on average be stronger (or possibly weaker) than the reflections on later images. Thus a scale factor is applied to each image, based on comparisons of the multiple measurements of the same hkl reflection (or symmetry related reflections) that are observed on different images. After scaling, the reflections are then merged, averaged, and output to ScaleAveraged.ref.

A logfile called **dtscaleaverage.log** is written, which contains the statistical analysis that is necessary to evaluate the quality of the data set. Most of the important information is summarized in a table at the end of the logfile, such as R-merge, Intensity/Sigma, % completeness, and Redundancy. The tables above the summary information at the very end of the logfile give the relevant statistics as a function of resolution. These tables will help you decide what resolution limit to cut the data off at. Generally, where the un-averaged intensity/sigma gets below about 2.0 and the R-merge gets above 30% is about the resolution cutoff. Once you have decided where to cut the data off, enter the resolution limits and run Scale and Average one final time. This will give you a data set and a final table of statistics that will be convenient to use for structure determination and publication.

Here are the tables and summary information at the very end of the "dtscaleaverage.log" file:

```
Completeness vs Resolution
-------------------------------------------------------------------------------
 Resolution      Calc      Num      Num      Num      Num      Num      Avg   %Comp   %Comp
    range      unique      obs     rejs    mults   single unique    mult  shell   cumul
-------------------------------------------------------------------------------
 46.31 - 5.16     592     3456       96      539       26      565    5.95    95.4    95.4
  5.16 - 4.09     541     3473       24      513       21      534    6.46    98.7    97.0
  4.09 - 3.58     523     3413       13      503       19      522    6.51    99.8    97.9
  3.58 - 3.25     525     3431       24      506       18      524    6.50    99.8    98.3
  3.25 - 3.02     510     3362       15      498       11      509    6.58    99.8    98.6
  3.02 - 2.84     516     3378       23      510        5      515    6.51    99.8    98.8
  2.84 - 2.70     501     3326        8      498        3      501    6.62   100.0    99.0
  2.70 - 2.58     492     3290       12      491        1      492    6.66   100.0    99.1
```

2.58 - 2.48	507	3335	16	504	3	507	6.55	100.0	99.2
2.48 - 2.39	515	2811	19	469	27	496	5.63	96.3	98.9
46.31 - 2.39	5222	33275	250	5031	134	5165	6.39	98.9	98.9

Redundancy vs Resolution

Resolution range	Calc unique	Percent of reflections measured N times, N = 0	1	2	3	4	5-8	9-12	>12	%Comp shell	%Comp cumul
46.31 - 5.16	592	4.6	4.4	13.5	7.3	9.6	40.7	19.6	0.3	95.4	95.4
5.16 - 4.09	541	1.3	3.9	5.9	8.5	9.6	43.8	27.0	0.0	98.7	97.0
4.09 - 3.58	523	0.2	3.6	4.4	5.7	14.1	45.5	26.4	0.0	99.8	97.9
3.58 - 3.25	525	0.2	3.4	5.0	4.6	15.8	44.0	27.0	0.0	99.8	98.3
3.25 - 3.02	510	0.2	2.2	5.1	2.7	17.6	46.1	26.1	0.0	99.8	98.6
3.02 - 2.84	516	0.2	1.0	6.2	2.7	18.8	46.7	24.4	0.0	99.8	98.8
2.84 - 2.70	501	0.0	0.6	4.2	4.4	18.8	45.3	26.7	0.0	100.0	99.0
2.70 - 2.58	492	0.0	0.2	3.0	4.5	19.3	47.8	25.2	0.0	100.0	99.1
2.58 - 2.48	507	0.0	0.6	4.5	4.9	18.5	45.2	26.2	0.0	100.0	99.2
2.48 - 2.39	515	3.7	5.2	11.1	8.9	14.2	39.0	17.9	0.0	96.3	98.9
46.31 - 2.39	5222	1.1	2.6	6.4	5.5	15.5	44.3	24.6	0.0	98.9	98.9

Resolution range	Percent of reflections measured AT LEAST N times, N = 13	9	5	4	3	2	1
46.31 - 2.39	0.0	24.6	69.0	84.5	89.9	96.3	98.9

Rmerge vs Resolution

Resolution range	Average counts	Num rejs	Num mults	I/sig unavg	I/sig avg	Rducd ChiSq	Model Eadd*	Rmerge shell	Rmerge cumul
46.31 - 5.16	6599	96	539	15.7	39.0	0.87	0.02	0.031	0.031
5.16 - 4.09	5101	24	513	9.7	24.7	0.84	0.04	0.050	0.039
4.09 - 3.58	3686	13	503	7.2	17.9	0.90	0.05	0.070	0.046
3.58 - 3.25	2382	24	506	5.6	13.7	0.93	0.05	0.092	0.052
3.25 - 3.02	1227	15	498	3.9	9.1	0.99	0.09	0.150	0.059
3.02 - 2.84	798	23	510	3.2	7.4	1.02	0.09	0.196	0.064
2.84 - 2.70	608	8	498	2.9	6.6	1.07	0.10	0.229	0.069
2.70 - 2.58	510	12	491	2.6	6.0	1.06	0.12	0.266	0.074
2.58 - 2.48	490	16	504	2.5	5.7	1.13	0.14	0.281	0.078
2.48 - 2.39	424	19	469	2.3	5.1	1.25	0.15	0.330	0.082
46.31 - 2.39	2232	250	5031	5.6	13.7	1.00	0.05	0.082	0.082

I/sig unavg is the mean I/sig for the unaveraged reflections in the input file.
I/sig avg is the mean I/sig for the unique reflections in the output file.
* When EMul == 4.80

Summary of data collection statistics

Spacegroup	P4122	
Unit cell dimensions	46.31 46.31 111.14	
	90.00 90.00 90.00	
Resolution range	46.31 - 2.39	(2.48 - 2.39)
Total number of reflections	33025	
Number of unique reflections	5165	
Average redundancy	6.39	(5.63)
% completeness	98.9	(96.3)

```
 Rmerge                          0.082           (0.330)
 Reduced ChiSquared              1.00            (1.25)
 Output <I/sigI>                 13.7            (5.1)
-------------------------------------------------------------
  Note: Values in () are for the last resolution shell.

   33275 reflections in data set
       0 reflections rejected (|ChiSq| >  50.00)
     250 reflections total rejected (  0.75%  |Deviation|/sigma >   7.73)
    1516 reflections excluded from scaling/absorption (I/sig <= 3.00)

Writing ScalAveraged.ref ...
Number of reflections written in 'ScalAveraged.ref': 5165
... done writing.
Number of reflections written in 'dtscaleaverage_rejects.ref': 250

dtscaleaverage: Done.
```

A note on some of the parameters in the dtscaleaveraged.log file.

R-merge. This is the most useful parameter in assessing the quality of the data. It is essentially the average % error of the intensity measurements:

$$\text{R-merge} = \Sigma_h \Sigma_i \mid I_{hi} - \langle I_h \rangle \mid / \Sigma_h \Sigma_i \langle I_h \rangle$$

where I_{hi} is the i^{th} observation for reflection h, and $\langle I_h \rangle$ is the mean intensity value for a reflection that was measured multiple times. The sum is over all i measurements and over all hkl reflections. An R-merge value of about 5% (0.05) or below is considered excellent, whereas a value greater than 10% is poor. To get a better feel for the meaning of R-merge, if the mean value for the intensity measurement was 100, then two measurements of 95 and 105 would give an R-merge of about 5% ([5 + 5]/[100 + 100]). R-merge is given in the tables as a function of resolution. You will see as the resolution gets higher, R-merge increases. Usually when R-merge gets above 30% is a good place to cut off the data. Also in the logfile is a value of R-merge for each image, and sometimes bad images (perhaps due to crystal decay at the end) will have high R-merge values, and these can be omitted from the process if necessary.

%Completeness. For a given crystal with a unit cell within a particular resolution shell, there will be a theoretical total number of reflections for the reciprocal lattice. % completeness simply is an indication of what % of the possible reflections were actually measured at least once. Note that if the edge of the detector is at say 2.6Å, then data from 2.3-2.6 Å might be less complete because for that resolution shell only data in the corners of the detector was measured. Usually you want data to be at least 90% complete, and > 50% complete for the highest resolution shell. % completeness can be increased simply by collecting more data or moving the detector forward, which shouldn't be a problem if the crystal is frozen.

I/sigI. This is the average measured intensity divided by the associated sigma value for both the raw data (before averaging for individual measurements) as well as the data after averaging. The sigma value is related to the background intensity in the region outside of the spot. A value of 1 for I/sigI would indicate data that is not above the background. Usually a good place to cutoff the data is the resolution at which I/sigI for unaveraged reflections gets below 2.0.

Redundancy. This indicates on average how many times each reflection was measured. As noted above, higher redundancy is more favorable- the more independent measurements you have for each reflection, the more confidence you have in the mean value and the associated sigma. Higher redundancy also leads to slightly higher R-merge values, but the confidence in the mean measurement is also high.

ChiSq. This is a statistical number related to the error models used in scaling and averaging. A value close to 1.0 for all resolution ranges indicates that things are going well. Values much higher than 1.0 (say 2.0) indicate that the scaling/averaging could be improved by changing the input parameters.

Where to cut off the data. It is best to integrate at a resolution slightly higher than the crystal actually diffracts to, and then use the statistical information in the dtscaleaverage.log file to choose an appropriate resolution to cutoff the data. Then rerun Scale and Average at the cutoff resolution to get a final data set with final statistics. The appropriate resolution for cutting off the data is a matter of some debate, but in general this is the resolution shell at which R-merge is > 30%, and I/sigI (unaveraged) gets lower than 2.

Here are a few lines of the final output file, "ScaleAveraged.ref":

```
3 2 0 5
CRYSTAL_MOSAICITY= 1.4545 0.0000 0.0000;
CRYSTAL_ORIENT_ANGLES= 29.9644  -15.1562  -1.2253;
CRYSTAL_SPACEGROUP=91;
CRYSTAL_UNIT_CELL= 46.3118  46.3118  111.1354  90.0000  90.0000  90.0000;
SOURCE_WAVELENGTH= 1.0000  1.5418;
nH
nK
nL
fIntensity
fSigmaI
     0     0    40 6043.79 581.814
     0     0    44 10626.6 3771.72
     1     0     0 -30.4871 12.7317
     1     0     1 -16.8526 12.2327
     1     0     2 -17.2999 13.6181
     1     0     4 6895.02 165.237
     1     0     5 2265.90 92.8565
     1     0     6 4364.60 145.194
     1     0     7 104.651 26.9161
     1     0     8 1958.94 147.902
     1     0     9 18131.5 410.524
     1     0    10 641.225 101.757
```

Note that the reflections are sorted, beginning with low resolution. Also note that some of the intensity measurements result in negative values, which is not real (they can't actually be lower than background). These measurements where the error is higher than the intensity value.