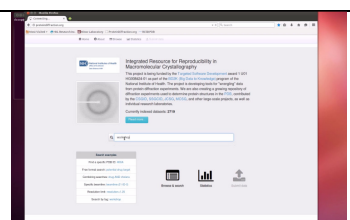


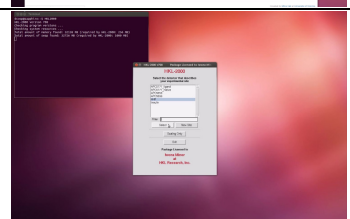


Introduction to Indexing and Scaling

Welcome to HKL Research's introduction to indexing and scaling of crystallographic data using the HKL suite. The data for this tutorial is available at proteindiffraction.org. Search using the keyword "Workshop" to find several excellent training datasets. The one we will work with today is that of 1WQ6, the NerveHomology2 domain of ETO.



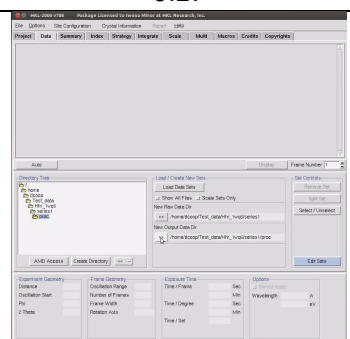
When you start HKL, you are first presented with a list of experimental sites. The site definition contains information about the X-ray detector and other parameters that describe the experimental setup. The most important parameters are the type of detector and the position of the direct beam. For this tutorial, we will use the site called HHR, which corresponds to an ADSC-Q4 detector.



0:27

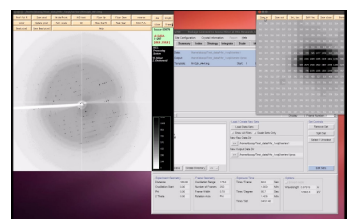
The HKL interface has multiple tabs across the top of the screen, and is in the Data tab when the program launches.

The first step of processing data, is to select where the data is. Use the Directory Tree on the left to navigate to and select the directory where the diffraction images are and click the double arrow button in the "New Raw Data Dir" field. It is a good idea to create a new directory for the program output to keep the original images separate from the files that result from processing the images, so we will click the "Create Directory" button. After creating a directory, select it and click the double arrow button in the "New Output Data Dir" field.



0:52

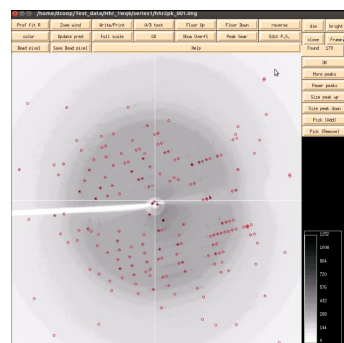
To see the first image of the dataset, click the Display button, which is under the data window. The bright and dim buttons in the upper right allow you to adjust the contrast for the images. The "Frame" button allows you to scroll through images. Middle clicking on the Frame button will display the next image in the direction of the arrow. Left click to change the direction of the arrow to down, right click to change the arrow to up. A Zoom window can be opened to get a better look at individual reflections. Middle clicking on the main image or in the zoom window will re-center the zoom window. Zoom in far enough and you can see the photon counts for each pixel.



1:35

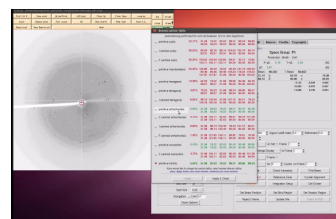
The first step of indexing the data is to “Peak Search” the data. There are two ways to do this. The first is to select the “Peak Search” button at the top of the image display. This opens a dialog on the right that provides ways to alter the number of peaks selected. The Frame/Up button can be used to select peaks on more than one image. The Pick Add and Remove buttons are very useful for weak data or if you need to select one crystal lattice from an image that clearly contains reflections from more than one crystal. Click OK when you are done.

Because the default parameters for peak selection are usually good enough, the alternate way perform a peak search is to switch to the “Index” tab and select “Peak search”. This will automatically select peaks from the first frame. It will also display the image if you have not previously done so.



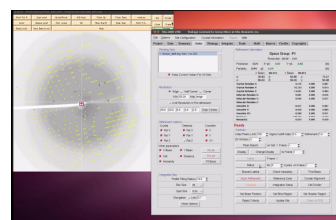
3:01

The Index button will try to auto-index the peaks using into all of the Bravais Lattices and presents a table with the Bravais Lattices types ordered from higher to lower symmetry, from top to bottom. Each crystal class is presented with two unit cell values—the top is a primitive triclinic unit cell that fits the data and is closest to the a unit cell that compatible with that crystal class and then second row is the unit cell that would be necessary for that crystal class. For example, primitive tetragonal crystals have $a = b$ and all angles = 90. Each crystal class is presented with a correlation coefficient that represents how well these two unit cells agree. HKL color codes the fit from dark green to red. Often you may select the crystal with the highest symmetry with a good correlation coefficient, but a better practice is to initially select the P1 unit cell and refine the unit cell before selecting the crystal class.



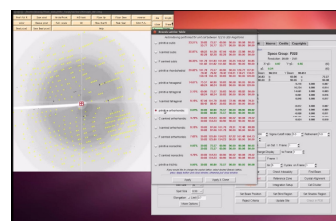
3:53

To do this, “Apply and close” this window, and use the “Refine” button. Initially, HKL will only refine a small set of parameters. The overall measurements of how well the selected crystal class fits the data are the Positional and Partiality Chi Square values. These are also color coded from dark green to red for good to bad. The values that are being refined are presented with their current values, how much they have changed during the refinement cycle, and the error associated with the value. Click “Refine” several times and the values should become stable. You can then add more parameters by selecting “Fit All” in the “Refinement Options” box. Click Refine again until all the parameters become stable.



4:51

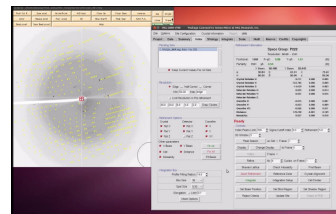
At this point, revisit the Bravais Lattice table, and notice that the Primitive Orthorhombic class is the highest symmetry lattice type with a low correlation coefficient. Select this lattice and “Apply and Close” this window. Click refine a few more times to ensure that the fit is still good. If the chi square values shoot up dramatically during refinement, you should abort the refinement and re-index the data. The peak search data is saved in a file, so the peak search does not have to be performed again unless you want to change the peaks you are working with.



5:37

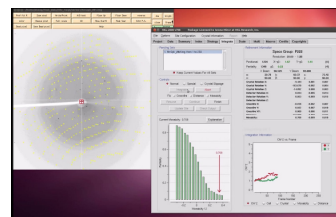
At this point, the peaks displayed on the image are color coded. Reflections that are recorded completely on this image will be green, whereas reflections that are split between more than one image will be yellow. These reflections will be combined during the scaling process. Reflections that are red have been rejected, perhaps because the background of the spot is not uniform, the selected spot size is too small for the reflection, or the reflection overlaps with another reflection. Keep an eye on the reflections during the refinement process. Too many rejected reflections is an indication that your indexing could be wrong, and will probably be accompanied by a spike in the chi square values.

This data set indexes easily, but allow me to mention some things that can help index more difficult data sets. You can increase the number of peaks that are used for the indexing, or increase the sigma level used for indexing. Increase the number of frames used for indexing by performing a manual peak pick and picking peaks from several frames. See the HKL Research video on advanced indexing for more tips.



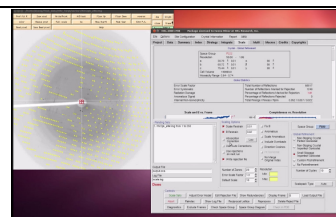
6:11

Once you are satisfied with the unit cell, use these parameters to “integrate” the whole dataset, which will index each individual frames and record the intensity and error associated with each peak. The process can be started with the Integrate button on the index tab, which will automatically switch the display to the integrate tab. Alternatively, use the integrate button on the Integrate tab to start the process. HKL will work through the available images. The image will be displayed; peaks will be predicted and then refined for each image. The display will show a mosaicity histogram for the image being indexed and a cumulative graph on the right. The buttons on the bottom let you examine various parameters. Once integration is complete, it is now time to scale the data.



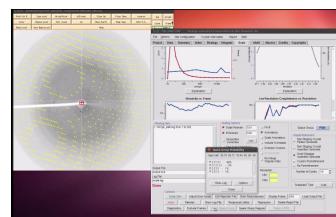
7:14

To do this, switch to the “Scale” tab and click the “Scale Sets” button. This process normalizes the intensity of all of the frames, combines the data from the partial reflections, and merges symmetry related reflections. A good approach to determine the spacegroup is to initially scale the data in the lowest symmetry spacegroup for the crystal class and then select use the “Check Space Group” button.



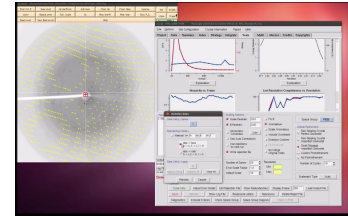
8:03

In this example, we know that we intend to use the anomalous signal from the data to solve the structure, so we will select the “anomalous” button. Leaving the rest of the parameters at their default settings, click “Scale Sets.” When it says it is done, you can get an idea of the quality of the data by looking at the Global Statistics at the top, as well as the various graphs that are presented. The final values will change, however, once we have the correct spacegroup, so click on the “Check Space Group” Button.



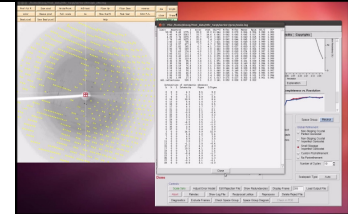
8:27

In this case, it suggests that our Space Group is $P 2_1 2_1 2_1$, which is a permutation of the $P 2_1 2_1 2$ space group. In order to transform this into a standard spacegroup setting, select the "Reindex" button. In this case, we need the axis without the screw axis to be on the C axis, so we will select abc->bca and select "Reindex". Now we can select $P 2_1 2_1 2$ from the Space Group selector, and scale the data again. I like to check the log file to verify the re-indexing process. At the end of the log file is a list of all reflections that should be systematically absent for the selected space group. In this case, all of the listed reflections are indeed absent.



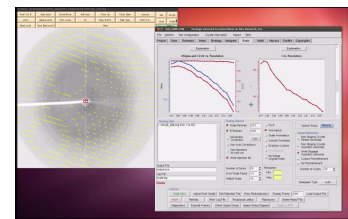
9:00

If your version of HKL is not linked to CCP4, the Check Space Group will not work. In this case, scale the Data in $P 2_1 2_1 2_1$ and examine the log file. You should see that one axis clearly does not have a screw axis. If this axis is not on the I axis, reindex the data.



9:54

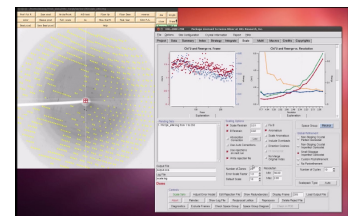
Scaling the data is an iterative process. The first pass can be used to determine resolution limits and the overall Error Scale Factor. To set the resolution limit, look at the graph of I/σ vs Resolution. This also has a graph of the $CC1/2$, which is another parameter that some people use for setting the resolution limit. If your data is weaker than 2 sigma, a line will appear on the graph suggesting a resolution to truncate your data. This is strong data, so we will use all of it. Set the upper and lower limits of the data based on the strength of diffraction and the size of the beam stop. The Error Scale Factor can be increased if too many reflections are being rejected, but that is not a problem in this example.



10:18

Once the data has been scaled once, some reflections will probably be flagged for rejection. Select "Use rejections on next run" to exclude these reflections from the scaling process. Scale the data a few times until the number of reflections marked for rejection stabilized and the global statistics are also stable.

You can click on the "Explanation" button under each graph to see what each line represents. In the graph of Chi square vs resolution, the blue line represents the chi square values when Friedel pairs are kept separate, and the orange line represents the chi square values if when the Friedel pairs are merged. The fact that the chi squares are higher when they are merged indicates that there is a measurable difference between the Friedel pairs, which indicates an anomalous signal. This example has one of the strongest anomalous signals you may ever see. You can usually solve the structure if there is even a modest difference between these two lines at lower resolutions.



11:02