MOSFLM tutorial for the new Interface

1. Introduction

1.1 Background

MOSFLM can process diffraction images from a wide range of detectors and produces, as output, an MTZ file of reflection indices with their intensities and standard deviations (and other parameters). This MTZ file is passed onto other programs of the CCP4 program suite (POINTLESS, SORTMTZ, SCALA, TRUNCATE) for further data reduction.

The MOSFLM program was originally written to process data collected on film. It was then modified to process data collected using the image plate detector developed at the EMBL outstation in Hamburg by Jules Hendrix and Arno Lentfer, and the name was changed to ipmosflm. This is the current version of the program, which will also process data from CCD and pixel detectors.

1.2 Installation

The new GUI (imosflm) is currently available for Windows, Mac OSX and Linux platforms. A For details of installation visit http://www.mrc-lmb.cam.ac.uk/harry/mosflm and follow the link to the New GUI imosflm.

1.3 Documentation

There are two distinct sources of documentation for MOSFLM, although neither of these currently makes any reference to the new GUI. At present, this document is the only documentation available for the new GUI.

1. The MOSFLM user guide. This is available as a plain text file (mosflm_user_guide.txt) or on the web (www.mrc-lmb.cam.ac.uk/harry/mosflm/) as a PDF or HTML document. It is a very good idea to look through this guide before starting serious data processing with MOSFLM, although you do not need it for this tutorial.

2. The "on-line" help. If you type "help" at the "MOSFLM =>" prompt (after starting the program) all possible keywords are listed, with information on each keyword. This information is stored in an ascii file (mosflm.hlp) which can also be read (and searched) with an editor. This relies on having the environment variable "CCP4_HELPDIR" set to the directory containing this file. This is also available on the MOSFLM web pages under "keyword synopses"

1.4 Aims of the tutorial

Your task is to process 84 images, hg_001.mar1600 to hg_084.mar1600, collected on a Mar345 image plate detector at a synchrotron beamline. These are crystals of a small domain (91 aa) that have been soaked in a mercury compound, resulting in a dataset with a strong anomalous signal which can easily be used to solve the structure. These images have kindly been provided by Camillo Rosano.

2. Overview of the new GUI

Start the program by typing "imosflm". After a brief delay, the following window will appear:
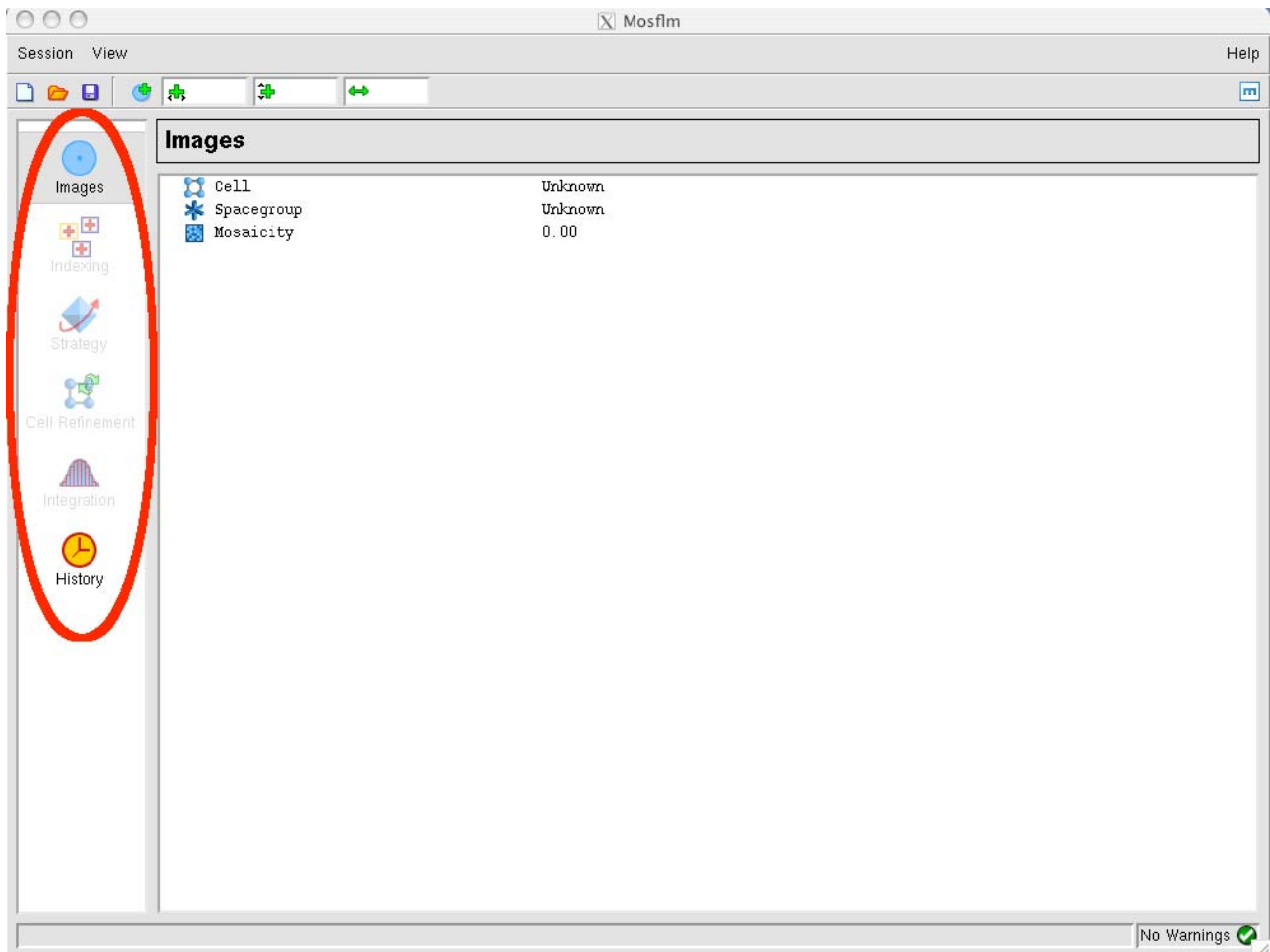
The basic operations (Images, Indexing, Strategy, Cell Refinement, Integration, History) can be selected by clicking on the appropriate icon, but those that are not appropriate will be greyed out and cannot be selected.

2. 1 Drop Down menus

Clicking on "Session" will result in a drop down menu that allows you to save the current session or reload a previously saved session (or add images):
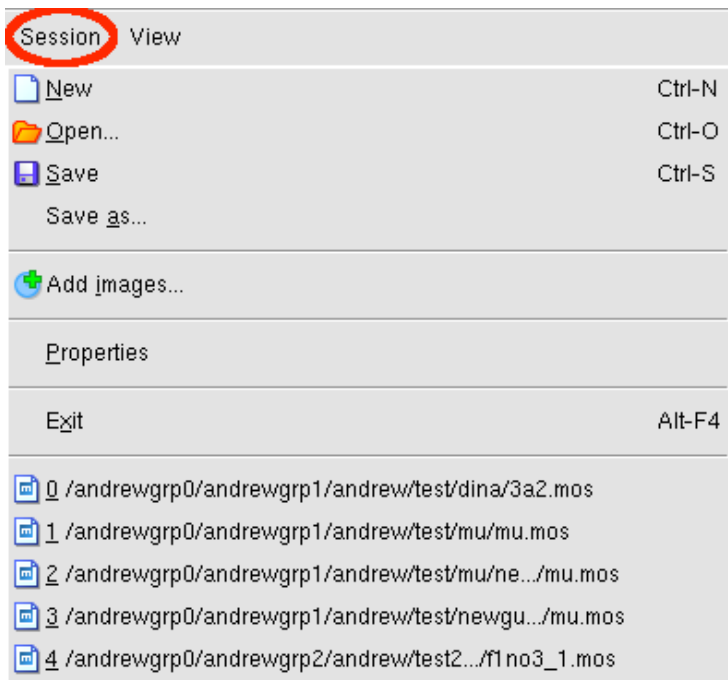
Clicking on "View" will allow you see (and modify) Experiment Settings or Processing options. These will be described later.
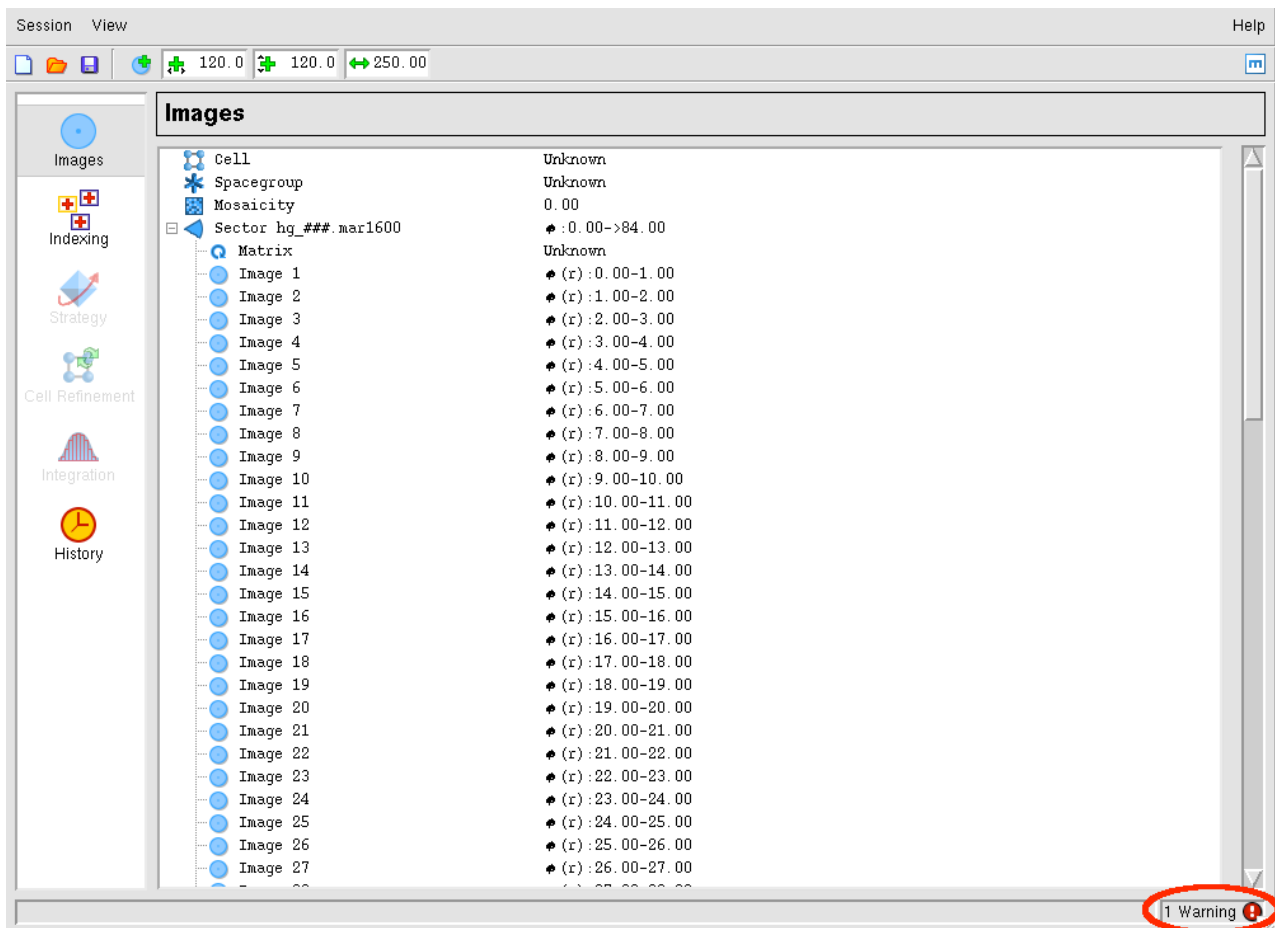

The three small icons below "Session" allow you to start a new session, open a saved session or save the current session. Moving the mouse over these icons will result in display of a tooltip describing the action taken if the icon is clicked.

3. Adding images to a session

To add images to a session, use the "Add images..." icon:
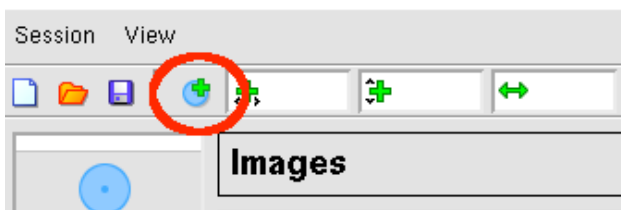
Select the correct directory (the default is the directory in which imosflm was launched) from the popup Add Images window. All files with an appropriate extension (which can be selected) will be displayed. Double clicking on any file will result in all images with the same template being added to the session (the template is the whole part of the filename except the number field that specifies the image number). (An alternative is to single click and then click on Open)

|		
Cell	Unknown	
Spacegroup	Unknown	
Mosaicity	0.00	
Sector hq_###.mar1600	:0.00->84.00	
Matrix	Unknown	
Image 1	(r):0.00-1.00	
Image 2	(r):1.00-2.00	
Image 3	(r):2.00-3.00	
Image 4	(r):3.00-4.00	
Image 5	(r):4.00-5.00	
Image 6	(r):5.00-6.00	
Image 7	(r):6.00-7.00	
Image 8	(r):7.00-8.00	
Image 9	(r):8.00-9.00	
Image 10	(r):9.00-10.00	
Image 11	(r):10.00-11.00	
Image 12	(r):11.00-12.00	
Image 13	(r):12.00-13.00	
Image 14	(r):13.00-14.00	
Image 15	(r):14.00-15.00	
Image 16	(r):15.00-16.00	
Image 17	(r):16.00-17.00	
Image 18	(r):17.00-18.00	
Image 19	(r):18.00-19.00	
Image 20	(r):19.00-20.00	
Image 21	(r):20.00-21.00	
Image 22	(r):21.00-22.00	
Image 23	(r):22.00-23.00	
Image 24	(r):23.00-24.00	
Image 25	(r):24.00-25.00	
Image 26	(r):25.00-26.00	
Image 27	(r):26.00-27.00	

The phi values for all the images (read from the image header) will be displayed in the Image window.

All images with the same template belong to the same "Sector" of data.

Multiple sectors can be read into the same session. Each sector can have a different crystal orientation.

Note that a "Warning" has appeared. Click anywhere in the warnings box to get a brief description of the warning. In this case, it is because the direct beam coordinates stored in Mar image plate images are not "trusted" by MOSFLM, and the direct beam coordinates have been set to the physical centre of the image. Click on the cross on the right hand side to turn off this warning.
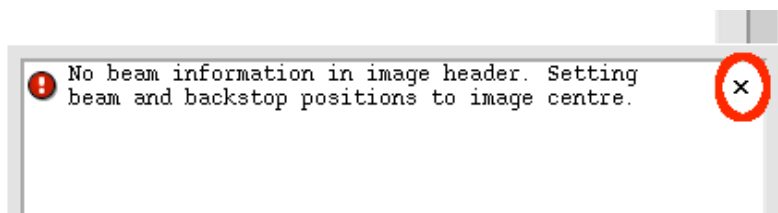
The direct beam coordinates (read from the image header or set by MOSFLM) and crystal to detector distance are displayed. These values can be edited if they are not correct.
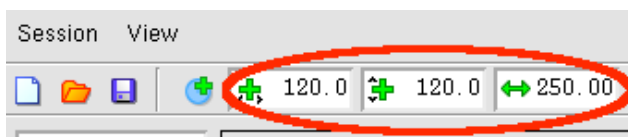

Advanced Usage

1. The phi values of images can be edited in the Image window. Double click with the mouse over the phi values to make them active, allowing new values to be given. These new values are propagated for all following images in the same sector.
2. To delete a sector, double click on the sector name to select it, then use the right mouse button on the sector to bring up a "delete" button. Move the mouse over the delete button and release the right mouse button to delete the images.
3. If one sector is added and used for indexing, and then a second sector from the same crystal is added, the matrix for the second sector will not be defined. To define it, double click on the matrix name for the first sector, save it to a file, double click on the matrix name (Unknown) for the new sector and read the matrix file written for the first sector.

4. Image Display

When images are added to a session, the first image of the sector is displayed in a separate Display window.
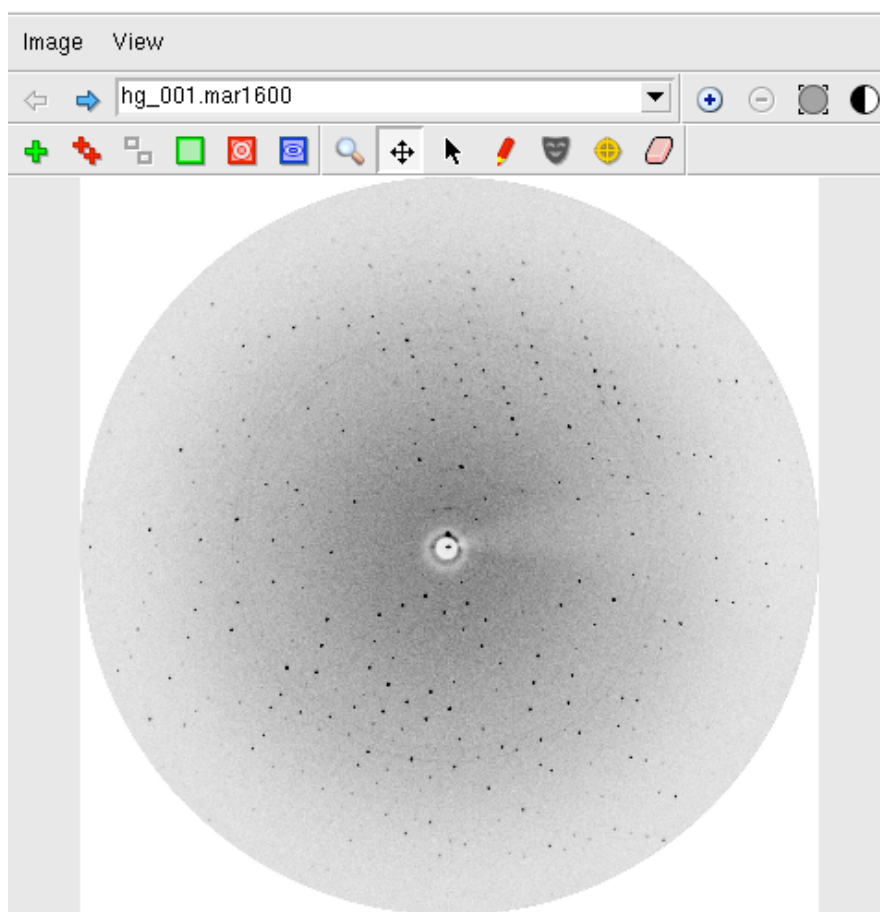

The "Image" drop down menu allows display of the previous or next image in the series. The "View" drop down menu allows the image to be displayed in different sizes (related by scale factors of two), based on the image size and the resolution of the monitor.

The line below allows selection of different images, either using right and left arrow or selecting one from the drop down list of all images in that sector. (The image being displayed can also be changed by double-clicking on an image name in the main "Images" window.)

"+" and "-" will zoom the image without changing the centre. The "Fit image" icon will restore the image

to its original size (right mouse button will have the same effect). The "Contrast" icon will give a histogram of pixel values. Use the mouse to drag the vertical dotted line, to the right to lighten the image, to the left to darken it. Try adjusting the contrast.

4.1 Display Icons
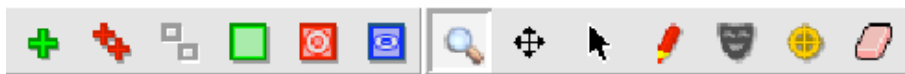
The six icons on the left, control the display of the direct beam position, spots found for indexing, predicted spots, masked areas, spot-finding search area and resolution limits respectively.

These are followed by icons for Zoom, Pan and Selection Tools, and tools for adding spots manually (for indexing), editing masks, circle fitting and erasing spots or masks.

4.1.1 Masked areas .... circular beamstop shadow

Select the masked area icon. A green circle will be display showing the default position and size of the backstop shadow.

Make sure that the Zoom icon (magnifying glass) is selected and use the left-mouse-button (abbreviated to LMB in following text) to drag out a rectangle around the centre of the image. The inner dotted yellow rectangle will show the part of the image that will actually appear in the zoom.
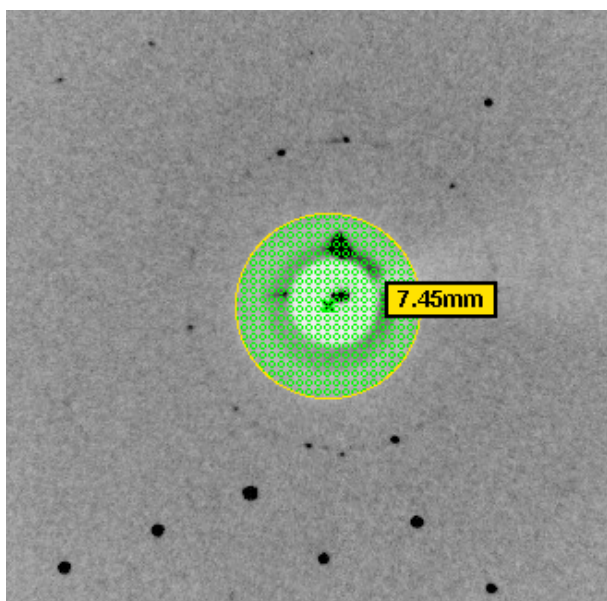
Choose the Selection Tool. When placed over the perimeter of the circle, the radius of the circular backstop shadow will be displayed. *Use the LMB to drag the circle to increase its diameter to that of the actual shadow on the image.* The position of the circle can be adjusted with LMB placed on the cross that appears in the centre of the green circle. *Adjust the size and position of the circle so that it matches the shadow.*

4.1.2 Masked areas... general exclusions

Choose the Masking tool. Any existing masked areas will automatically be displayed. *Use LMB to define the four corners of the region to be masked.* When the fourth position is given, the masked region will be shaded. This region will be excluded from spot finding and integration. This provides a powerful way of dealing with backstop shadows. To edit an existing mask, choose the Selection tool and use the LMB to drag any of the four vertices to a new position.


To delete a mask, choose the Spot and mask eraser tool. Place the mouse anywhere within the shaded masked area, and use LMB to delete the mask. *Delete any masks that you have created.*

4.1.3 Spot search area

Select the Show spotfinding search area icon. The inner and outer radii for the spot search will be displayed as shown below. If the images are very weak, the spot finding radius will automatically be reduced, but this provides additional control.

Either can be changed by dragging with the LMB. *Do not change the radii for these images.*
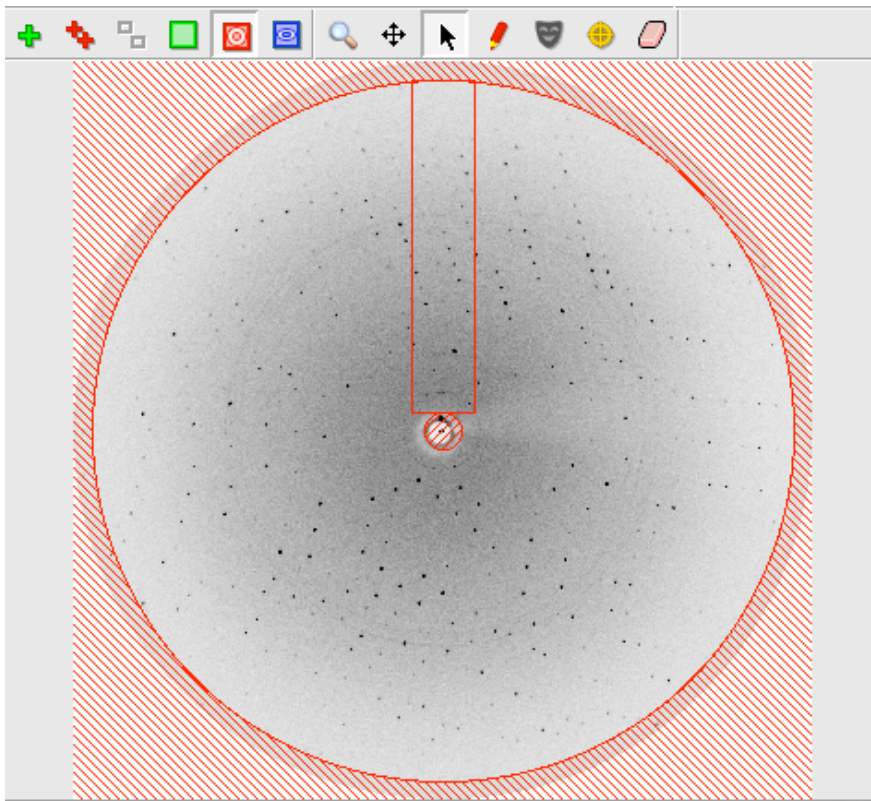
4.1.4 Resolution limits

Select the Show resolution limits icon. The low and high resolution limits will be displayed. The resolution limits can be changed by dragging the perimeter of the circle with LMB (make sure that the Selection Tool has been chosen). The resolution limits will affect Strategy, Cell refinement and Integration, but not spot finding or indexing. The low resolution is not strictly correct (it falls within the backstop shadow) but does not need to be changed because spots within the backstop shadow will be rejected.

4.1.5 Zooming and Panning

First select a region of the image to be zoomed with the Zoom tool.

Select the Pan tool and pan the displayed area by holding down LMB and moving the mouse. This is rapid on a dedicated machine, but may be slow if run over a network.

4.1.6 Circle fitting

The circle fitting tool can be used to determine the direct beam position by fitting a circle to a set of points on a powder diffraction ring on the image (typically due to icing) or to fit a circular backstop shadow (although using the masking tool is probably easier).

Select the circle fitting tool. Three new icons will appear in the image display area. There are two (faint) ice rings visible on the image at 3.91Å and 3.67Å. Click with LMB on several positions (6-8) on the outer ring (as it is slightly stronger). Then click on the top circular icon.

A circle that best fits the selected points (displayed as yellow crosses) will be drawn, and the direct beam position at the centre of this circle will be indicated with a green cross. The direct beam coordinates will be updated to reflect this new position.
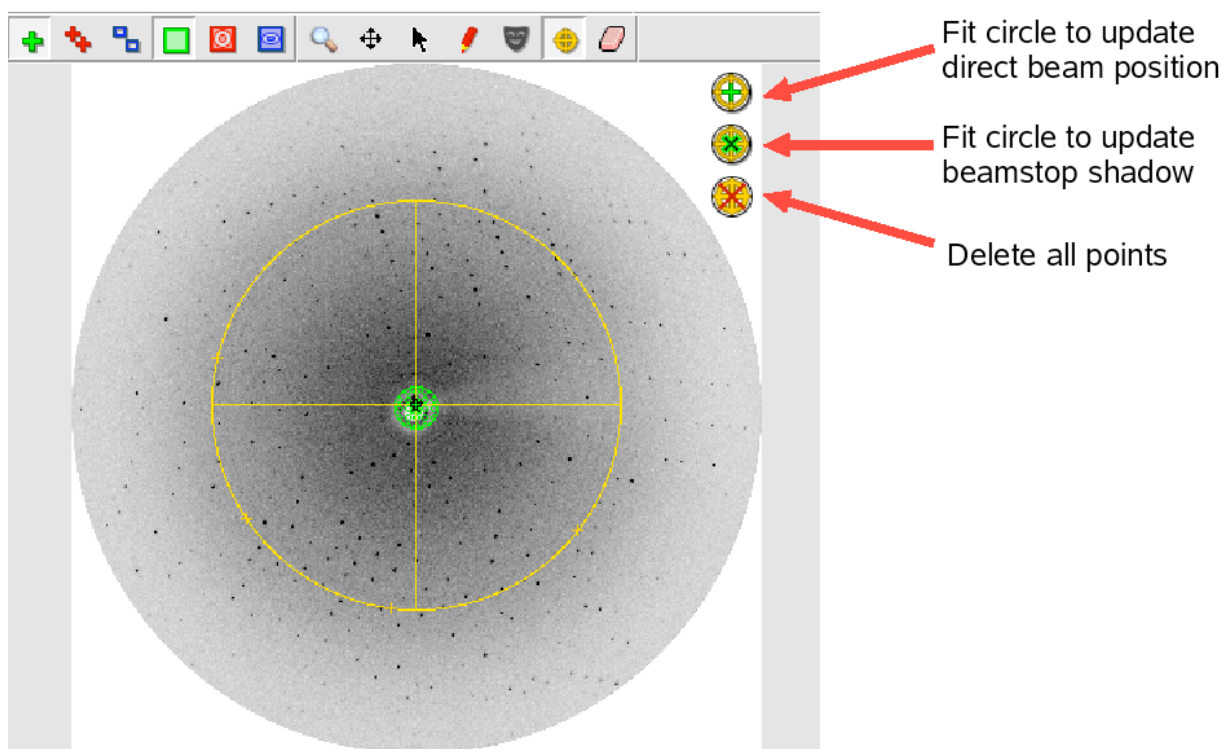
4.2 Other functionalities

Right mouse button will return the display to the full size image, if it is zoomed.

To get a small zoom window that can be moved over the image, hold down "Shift" with the Zoom tool selected. The area within the dotted square will be zoomed within the solid square.

Holding down "Alt" (or "Command" ⌘ on Macs) will display the resolution at the current mouse position. If positioned over a found spot, the spot coordinates and I/σ(I) will be given.

If positioned over a predicted spot, the indices of that spot will be given.

5. Spot finding, indexing and mosaicity estimation

When images have been added, the "Indexing" operation becomes accessible (it is no longer greyed out).

*Click on Indexing.* This will bring up the major Indexing window in place of the Images window.

5.1 Spot Finding

By default, two images 90 apart in phi (or as close to 90 as possible) will be selected and a spot search carried out on both images.

Found spots will be displayed as crosses (red for those above the intensity threshold, yellow for those below). They will also be shown in the Image Display window. The intensity threshold normally defaults to 20, but will be automatically reduced to 10 or 5 for weak images. The threshold is determined by the last image to be processed.

The number of spots found, together with any manual additions or deletions, are also given.

Images to be searched for spots can be specified in several ways:

Indexing

Images: 1, 84
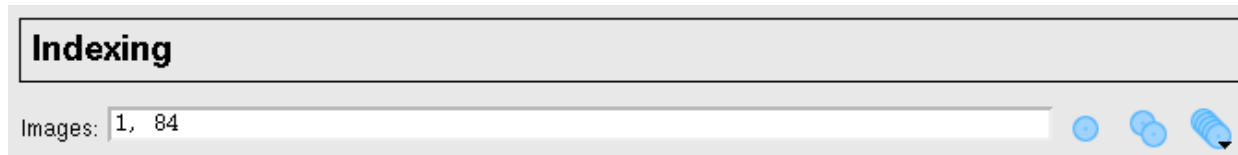
1) Simply type in the numbers of the images (eg 1,84 above).

2) Use the "Pick first image" icon

3) Use the "Pick two images ~90 apart" icon (default behaviour)

4) Use the "Select images ..." icon. If selected, all images in the sector are displayed. Click on a image to select it, then double click on the search icon for that image to run the spot search. The image will move to the top of the list (together with other images that have been searched).

Images to be used for indexing can be selected from those that have been searched by clicking on the "Use" button. If this box was previously checked, then clicking will remove this image from those to be used for indexing. It can be added again by clicking the "Select images ..." icon and clicking on the "Use" box.

5.1.1 Difficult images

Parameters affecting the spot search can be modified by selecting the "View" drop down window and selecting "Processing options". The resulting new window contains five tabs relating to Spot finding, Indexing, Processing, Advanced refinement and Advanced integration.

The Spot finding window allows the Search area, Spot discrimination parameters, Spot size parameters, Minimum spot separation and Maximum peak separation within spots (to deal with split spots) to be reset. It also allows the choice between a local background determination (preferred) and a radial background determination. The local background method also uses an improved procedure for recognising closely spaced spots. The only parameters commonly changed are:

1. Minimum spot separation. This should be the size (in mm) of an average spot (not a very strong spot). Change to values estimated by manual inspection of spots if there are difficulties due to badly split spots. This separation parameter is very important when spots are very close, but usually the program will determine a suitable value.
2. Minimum pixels per spot. Default value 6, but this will be reduced automatically to 4 if spots are very small.
3. Local background box size. Reducing this from 50 to (say) 20 can reduce the number of "false" spots found near any sharp shadow on the image.

For data collected in house, spots can be quite large and rather weak. In such cases the spot finding can be

greatly improved by the following:

1. Reduce the spot finding threshold, eg to 2.
2. Increase the minumum number of pixels per spot to 20-40.
3. Reduce the spot rms variation to 1.0
4. Set the minimum spot separation to a sensible value, eg 1.5mm

5.2 Indexing

Providing there are no errors during spot finding, indexing will be carried out automatically after spot finding. If the image selection or indexing parameters are changed, the "Index" button must be used to carry out the indexing.

Session   View   Help

Indexing

Images: 1, 84   Index

Image	Phi	Auto	Manual	Deleted	Total	Search	Use
1	0.0-1.0	297	0	0	297	⊕	☑
84	83.0-84.0	245	0	0	245	⊕	☑

| Total | | 0 | 542 | 0 | 0 | 542 | | |

Solutions:

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	58.5	58.6	62.1	90.2	118.0	120.0	0.21	0.37	0.44 (0.1)
2 (ref)	aP	0	58.5	58.6	62.1	61.9	62.0	60.0	0.21	0.37	0.44 (0.1)
3 (ref)	mC	1	101.4	58.6	62.2	90.0	123.0	90.0	0.42	0.44	0.44 (0.1)
4 (ref)	mC	1	101.5	58.5	62.1	90.0	123.0	90.0	0.20	0.40	0.44 (0.2)
5 (ref)	mC	2	101.5	58.5	62.1	90.0	122.9	90.0	0.20	0.38	0.45 (0.2)
6 (ref)	mC	2	101.5	58.5	62.1	90.0	123.0	90.0	0.20	0.40	0.44 (0.2)
7 (ref)	hR	3	58.6	58.6	156.4	90.0	90.0	120.0	0.20	0.53	0.45 (0.3)
8 (ref)	mC	59	84.3	85.5	58.5	90.0	132.9	90.0	1.41	0.61	0.05 (0.0)
9 (ref)	oI	60	58.5	61.5	85.4	90.0	90.0	90.0	1.38	0.59	0.14 (0.0)
10 (ref)	mC	60	103.2	62.1	58.6	90.0	124.7	90.0	0.72	0.75	0.39 (0.1)
11 (ref)	mC	60	101.5	58.5	62.1	90.0	123.0	90.0	0.20	0.40	0.44 (0.2)
12 (ref)	oI	60	58.5	62.4	85.2	90.0	90.0	90.0	1.03	0.86	0.33 (0.1)
13 (ref)	hR	61	60.0	60.0	145.4	90.0	90.0	120.0	1.88	0.72	0.27 (0.2)
14 (ref)	tI	61	58.9	58.9	85.2	90.0	90.0	90.0	1.16	0.90	0.27 (0.1)
15 (ref)	mC	61	101.5	58.5	62.1	90.0	123.0	90.0	0.20	0.40	0.44 (0.2)

Spacegroup: h3

Mosaicity: 0.00   Estimate

                                                                              No Warnings

Autoindexing will be carried out by MOSFLM using spots (above threshold) from the selected images. The threshold is set by MOSFLM but can be changed using the icon in the toolbar. The list of solutions, sorted by increasing penalty score, will appear in the lower part of the window. The preferred solution will be highlighted in blue. There will usually be a set of solutions with low penalties (0-20) followed by other solutions with significantly higher penalties. The preferred solution is that with the highest symmetry from the group with low penalty values.

> Note that *all* these solutions are really the same P1 solution transformed to the 44 characteristic lattices, with lattice symmetry constraints applied. Therefore, if the P1 solution is wrong, then all the others are wrong as well.

For solutions with a penalty less than 200, the refined cell parameters ("ref") will be shown. This can be expanded (click on the + sign) to show the "reg" unrefined but regularised cell (symmetry constraints applied) and the "raw" cell (no symmetry constraints applied).

The rms error (rmsd) in predicted spots positions ($\sigma(x,y)$ in mm) and the rms error in ($\sigma(\phi)$ in degrees) are given for each solution.

Usually the penalty will be less than 20 for the correct solution, although it could be higher if there is an error in the direct beam coordinates (or distance/wavelength). The rmsd (error in spot positions) will typically be 0.1-0.2mm for a correct solution, but if the spots are split or very elongated it can be as high as 1mm or higher.

The predicted pattern for the highlighted solution will be shown on the image display with the following colour codes:

Blue: Fully recorded reflection
Yellow: Partially recorded reflection
Red: Spatially overlapped reflection... these will NOT be integrated
Green: Reflection width too large (more than 5 degrees)... not integrated.

Providing there are no errors in the indexing, MOSFLM will automatically estimate the mosaic spread based on the preferred solution.

Select other solutions with a higher penalty and see how well the predicted patterns match the diffraction image.

The rmsd and visual inspection of the predicted pattern are the best ways of checking if a solution is correct. If the agreement is not good, then the autoindexing has probably failed.
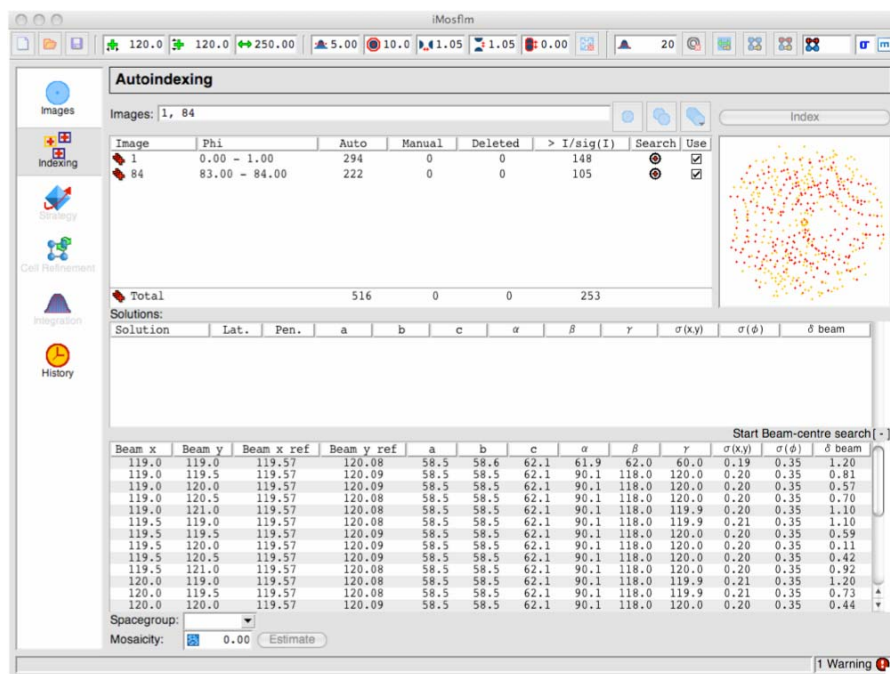
5.2.1 If the indexing fails: Direct beam search

The indexing is very sensitive to errors in the direct beam coordinates. For a correct indexing solution, these should be correct to better than half the minimum spot separation. Check, for example, that the current direct beam position is behind the backstop. If there are any ice rings, these can be used to determine the direct beam position.

If the accuracy of the direct beam coordinates (generally read from the image header) is uncertain, the program can perform a grid search around the input coordinates. The number and the size of the steps can be set from the Indexing tab of the Processing options menu (View->Processing options, see §2.1) but defaults to two steps of 0.5mm on each side of the input coordinates. The direct beam search is started by clicking on the "Start beam-centre search" text in the indexing pane. While in progress, it can be stopped by clicking on "Abort bean-centre search".
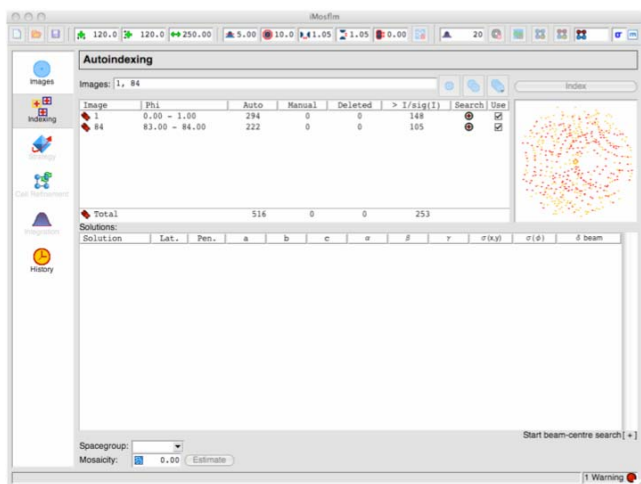
The indexing will be carried out for each set of starting coordinates (Beam x and Beam y in the table), and if a soltuion is found the refined beam coordinates (Beam x ref, Beam y ref), the unit cell parameters of the triclinic solution and the rmsd error in spot coordinates and in phi will be listed. The correct solution will generally be the one with the smallest rms error in spot positions ($\sigma(x,y)$). Note that **only** the triclinic (P1) solution will be listed. To complete the indexing, double click on the chosen solution, and the full indexing will carried out for these beam coordinates, with the solutions appearing in the upper window. The results of the direct beam search can be hidden (collapsed) by clicking on the [-] symbol next to "Start beam centre search" and recovered by clicking on [+]

Beam x	Beam y	Beam x ref	Beam y ref	a	b	c	α	β	γ	σ(x.y)	σ(φ)	δ beam
119.0	119.0	119.57	120.08	58.5	58.6	62.1	61.9	62.0	60.0	0.19	0.35	1.20
119.0	119.5	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.81
119.0	120.0	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.57
119.0	120.5	119.57	120.08	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.70
119.0	121.0	119.57	120.08	58.5	58.5	62.1	90.1	118.0	119.9	0.20	0.35	1.10
119.5	119.0	119.57	120.08	58.5	58.5	62.1	90.1	118.0	119.9	0.21	0.35	1.10
119.5	119.5	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.59
119.5	120.0	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.11
119.5	120.5	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.42
119.5	121.0	119.57	120.08	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.92
120.0	119.0	119.57	120.08	58.5	58.5	62.1	90.1	118.0	119.9	0.21	0.35	1.20
120.0	119.5	119.57	120.08	58.5	58.5	62.1	90.1	118.0	119.9	0.21	0.35	0.73
120.0	120.0	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.44

5.2.2 If the indexing fails: Other parameters.

Errors in other physical parameters (wavelength, crystal to detector distance) can also result in failure. All these parameters should be checked.

Several parameters used in autoindexing can also be adjusted using icons that appear above the Indexing bar.
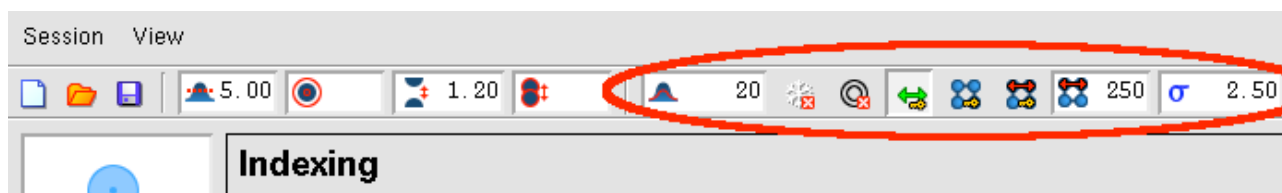
Weak images

1. MOSFLM automatically reduces the I/σ(I) threshold for weak images, and it may also reduce the resolution to 4Å, but lower values can be tried. It is important not to include spots that are not "real", a small number of false spots can prevent the indexing from working.
2. Try changing parameters for spot finding (see 5.1.1)

Multiple lattices

1. Try increasing the I/σ(I) threshold (default 20), for example to 40 or 60, so that only spots from the stronger lattice are selected. This can be done by entering a new value in the leftmost of the circled

icons above (and also via the using the Indexing tab of Processing Options).

All cases

1. Include more images in the indexing.
2. In case the crystal orientation has changed, try indexing using only one image.
3. If there are ice rings/spots, use the ice/ring exclusion option.
4. If the cell parameters are known, reduce the maximum allowed cell edge to the known maximum cell edge. This can sometimes help filter incorrect solutions.
5. If the detector distance is uncertain, and the images are high resolution (eg 2Å), allow the detector distance to refine during cell refinement.

5.2.2 Space group selection

Note that the indexing is based solely on information about the unit cell parameters. It will therefore be very difficult (or impossible) to determine the correct Laue group in the presence of pseudosymmetry, for example a monoclinic space group with $\beta \sim 90$ will appear to be orthorhombic, an orthorhombic space group with very similar a and b cell parameters will appear to be tetragonal. These can only be distinguished when intensities are available (ie after integration) by running POINTLESS.

In addition, it is not possible to distinguish between Laue Groups 4/m and 4/mmm, 3 and 3/m, 6/m and 6/mmm, m3 and m3m. This will not affect integration of the images, but it will affect the strategy calculation. **In the absence of additional information, the lower symmetry should be chosen to ensure that a complete dataset is collected.**

The presence of screw axes also cannot be detected, so there is no basis on which to distinguish $P2_1$ from P2 etc. This does not affect any aspect of data collection or processing, and can be chosen (on the basis of systematic absences) after integration by running POINTLESS.

In this example, there is no way of knowing at this stage if the space group is H3 or H32, so leave it as H3.

5.3 Mosaicity estimation
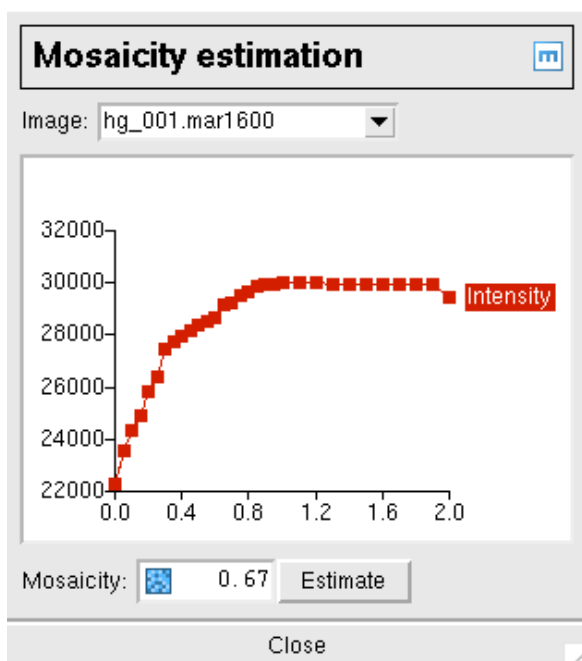
The mosaicity will be estimated automatically for the preferred solution. However, if another solution is chosen, the mosaicity should be estimated again.

Click on the "Estimate" button to estimate the mosaicity.

A window will appear which plots the total predicted intensity as a function of mosaic spread, from which an estimate is determined.

The prediction will be updated. Typing different values of mosaic spread in the box will allow a visual estimate of the effect of changing the mosaic spread.

The mosaic spread can also be defined in the Images pane.

5.3.1 The mosaic block size

In some cases, the apparent mosaic spread is much large at low resolution than at high resolution, so that if it is

set to a value that results in all the low resolution spots being predicted, many more spots are predicted than are visible at high resolution. This effect is best dealt with by adjusting the "Mosaic block size" which is also displayed in the Images pane. This is modelling the size, in microns, of the individual mosaic blocks in the crystal. Decreasing this to values of 0.5-5 microns (say) will result in predicting more spots at low resolution but have virtually no effect at high resolution.

6. Saving a session
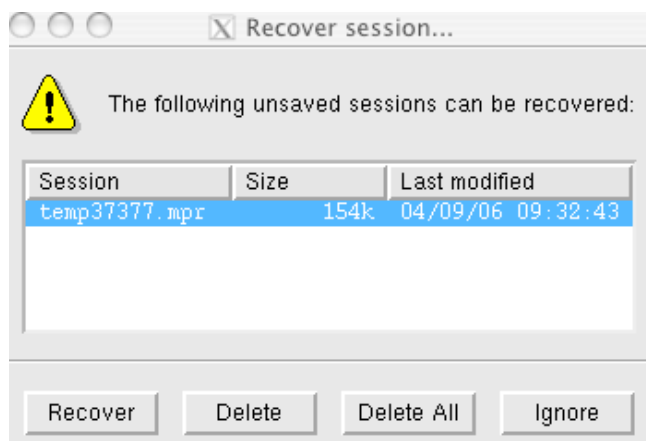
The session can be saved at any time, using the Session drop down menu referred to in 2.1. The first time that a session is saved, you will be prompted for a filename. The filename convention for saved session files is that the extension is ".mos".

Save the current session. Then exit from imosflm (using the Session drop down menu), restart imosflm and read in the saved session.

If the program crashes, it should be possible to recover all but the latest actions. Restart imosflm and it will pop up a "Recover session ..." window.

Selecting the Recover button will restore the session as far as possible. This file is written to the directory ".mosflm" in the users home directory.

7. Data collection strategy

Once the crystal orientation and (probable) Laue group have been determined, it is possible to calculate a data collection strategy and the Strategy icon is no longer greyed out (in fact, all other operations, Strategy, Cell Refinement and Integration become possible at this point).

Select the Strategy icon. This will open the Strategy window.
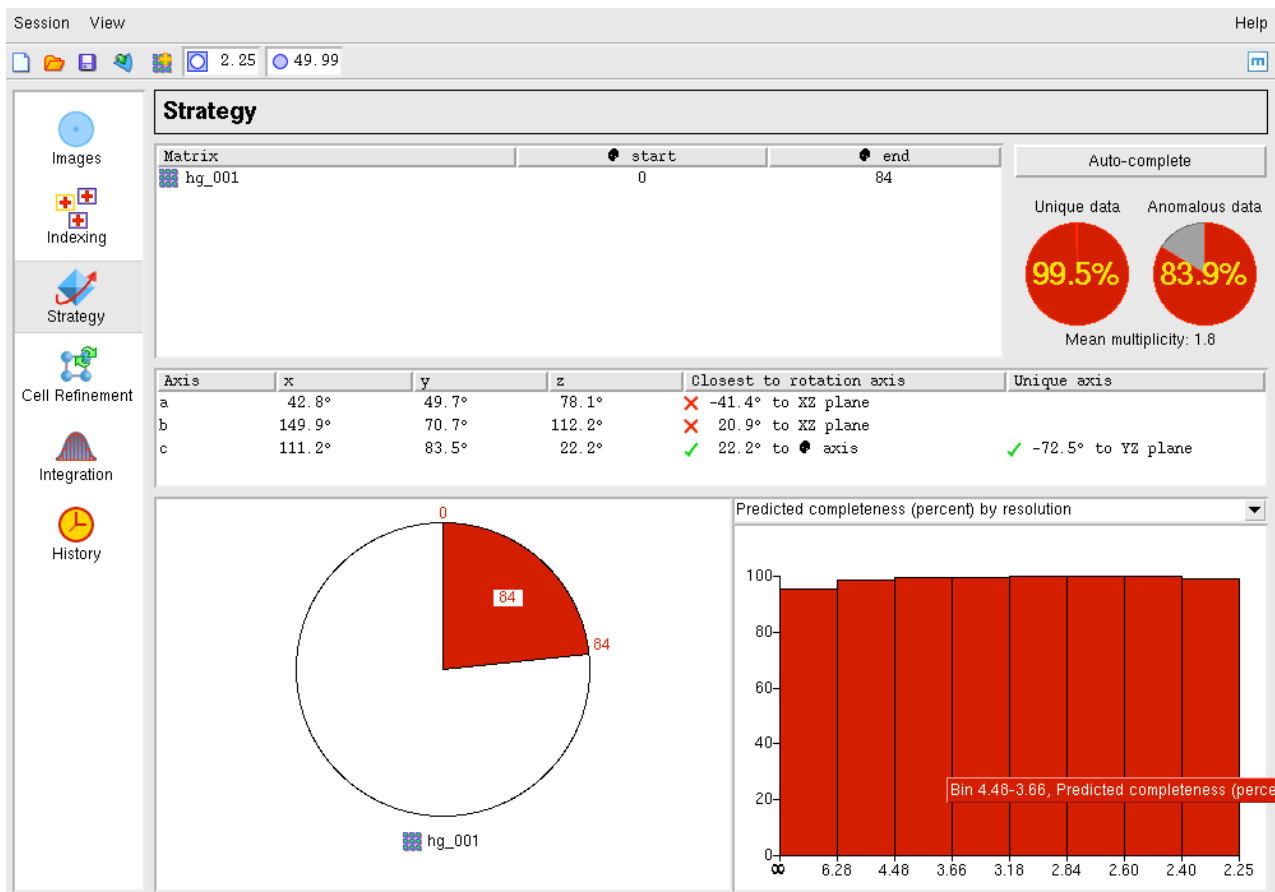
The statistics initially presented in the window are based on the assumption that all images between the smallest phi value and the largest phi value in the current sector(s) have been collected (the phi ranges are listed). If processing a dataset that has already been collected (as in this tutorial) this is appropriate.

More usually, the Strategy option will be used after two initial images, 90° apart in φ, have been collected, and the statistics will represent the completeness of that 90° of data.

The orientation of the crystal, expressed as the angles between the a,b,c unit cell axes and the X,Y,Z coordinate frame, is given. X is along the X-ray beam, Z is the rotation axis. A warning will be given if the unique axis is so close to the rotation axis that there will be missing cusp data.

7.1 Evaluating completeness manually

The completeness of any given segment of data can be determined manually using the circle at the bottom left of the window. *Click on the red segment shown in this circle*.
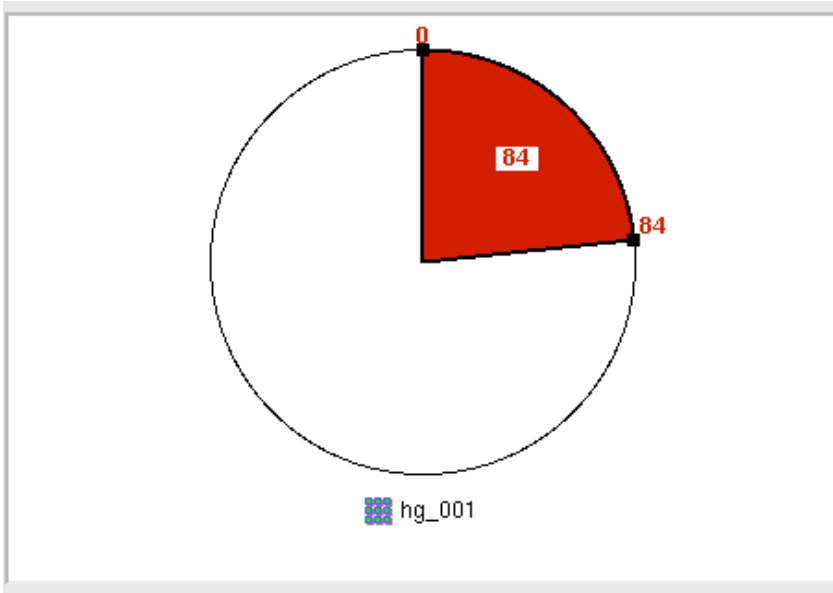
This will result in small black boxes being drawn on the circumference at the start and end of the red sector.

Use the LMB to click on the black box at "0", and drag this around the circle.

The sector size will change in increments of 5 degrees. When the LMB is released, the statistics for that sector will be displayed. The statistics are presented in two ways. Near the top right of the window, the overall completeness is shown as red circles for unique data and anomalous pairs. The average multiplicity is given beneath these circles.
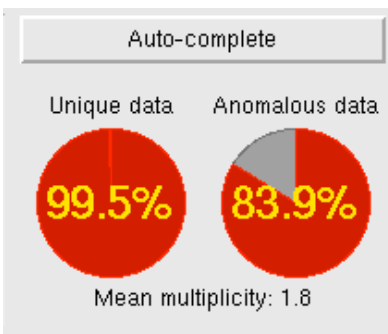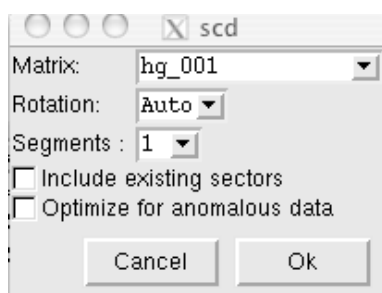

More detailed statistics (such as completeness as a function of resolution) are given in a series of histograms that can be selected by a drop down menu.
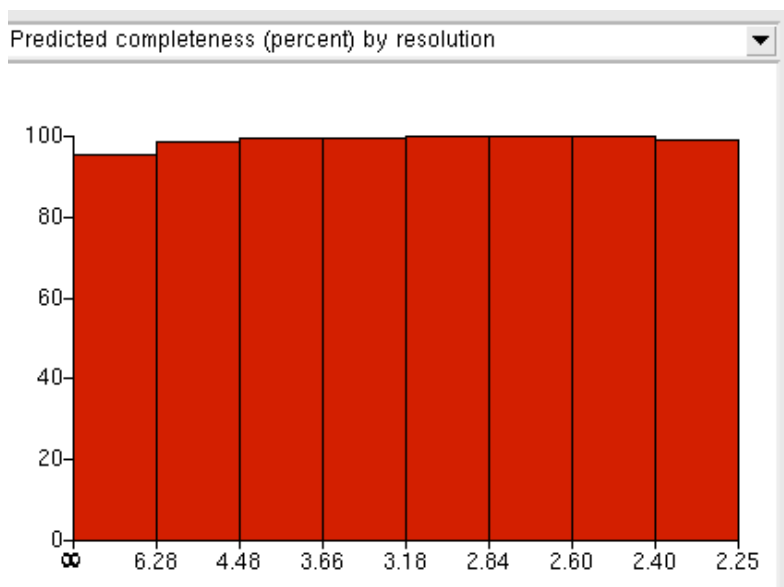
7.2 Calculating a strategy automatically

Select the Auto-complete button to calculate a strategy automatically.

An "scd" (Strategy calculation data) popup window will appear:

If multiple matrices have been defined (for different sectors of data) then the appropriate one can be selected (if there is only one sector, this has no effect).

The total ϕ rotation to be used is normally calculated by MOSFLM based on the Laue group and the crystal orientation (Auto). However, it is sometimes possible to achieve a high completeness with a significantly smaller total rotation (eg 60° in two 30° segments will typically give >94% completeness for orthorhombic space groups) and this can be useful if radiation damage is a serious problem.

Total rotation angles of between 30° and 90° can be selected from the drop down list. The number of segments to be used can be set between 1 and 3.

Two check boxes follow. If the first is checked, then it is assumed that data corresponding to the ϕ range listed have already been collected, and a collection scheme to complete the data will be calculated. Only check this box if these data have indeed been collected.

The second box should be checked if the anomalous signal is to be used.

THIS IS IMPORTANT. The optimum strategy is often different when maximising completeness of the anomalous data.

Choose the default values (Auto, 1 segment) but check the anomalous data box, then click on "Ok". Look at the various statistics presented as bar charts.

In space group H3, if rotating around the c axis, a total rotation of 120° would be required to collect a complete dataset (ignoring any data lost in the cusp). Because the c axis is 22° away from the rotation axis in this case, it is possible to collect very high completeness (for unique data) with a rotation much smaller than this.

Try to find the minimum total rotation that will give a dataset that is >95% complete for the unique data. Now do the same, but requiring >95% completeness for the anomalous data.

IGNORE SECTION 7.3 IF TIME IS SHORT

7.3 Calculating a strategy using multiple crystals

When collecting data from crystals that are very radiation sensitive, it may be necessary to use several crystals to collect a complete dataset. If data have already been collected from several crystals, the GUI can be used to calculate the best strategy for data collection from the current crystal. This is not implemented in the simplest way at present, but is usable. Assume that you have just indexed images from the latest crystal. Select the Strategy window.

Step 1. Set the phi range for the current crystal to zero (the current phi start, end are indicated in the top window). To do this, click anywhere in the red sector corresponding to the current phi range. Black squares will appear at the phi start/end positions. Click LMB on either of these and drag it onto the other one so that the phi range becomes zero. (You cannot enter phi values in the top window).

Known bug. It is only possible to move the phi values in steps of 5°. If the initial value is not a multiple of 5° degrees, the first new value caused by dragging the mouse will be a multiple of 5° and further steps will be in multiples of 5°. This means that in neither the start nor the final phi values are multiples of 5° then it will not be possible to reset the phi range to zero, just reset it to the smallest possible value.

Step 2. Use the "Add Matrix" icon to read in the matrix file for the segment of data that has already been collected. Make sure that this matrix has a unique name. When accepted, a new circle, with the matrix name underneath, will be displayed in the lower central window. In order to set the phi range for the data collected from this crystal, select that matrix (by clicking on the circle above it) and set the phi values graphically by holding down the Ctrl key and positioning the mouse anywhere within the central lower window. A solid line with a phi value attached will appear in the circle. Press the LMB to define the starting phi (as a multiple of 5°) and drag the mouse to sweep out the phi range for this segment of data. Again, this can currently only be set in multiples of 5°.

Repeat Step 2 for any other segments of data that have already been collected.

The top window will display the matrix name and the phi start/end values for all segments of data entered.

Step 3. When all the data have been entered, click on the matrix name of the current crystal in the lower central window. Then select "Auto-complete". In the window that appears (see 7.2) select a total rotation range for the current crystal. Normally the number of segments will be left at one, but if a rotation range of above 40° is chosen it might be worth trying 2 segments.

Then check the box "Include existing sectors". Then select OK. The best sector (of the specified phi width) will be chosen for the current crystal to get maximum completeness. Other sectors can be tried by dragging the phi start/end values in the graphical phi range display.

Known bug. The information on data collected already is not saved when the Strategy window is left, and so if multiple crystals are being used, this information will have to be entered separately for each crystal.

7.4 Calculating a suitable oscillation angle

To calculate the maximum possible oscillation angle while avoiding spatial overlaps, click on the "Check for overlaps" button. A popup window will appear, from which the option to calculate the maximum oscillation angle (as a function of φ) or check how many overlaps will occur for different (user selected) oscillation angles. The results are plotted in the same pane as the histograms for the strategy calculations. Note that the overlap calculation is based on the current values for the mosaic spread and spot separation, and can be very sensitive to these values.

8. Cell refinement

It is important to determine the cell parameters accurately before integrating the images. Although the unit cell is refined as part of the autoindexing, providing the diffraction extends beyond ~3.5Å resolution, it is possible to obtain more accurate cell parameters using a procedure known as post-refinement. This procedure requires the integration of a series of images in ideally two or more separate segments at widely different φ values. The distribution of the intensity of partially recorded reflections over the images on which they occur is used to refine the unit cell, crystal orientation and mosaic spread.

Select the "Cell Refinement" icon
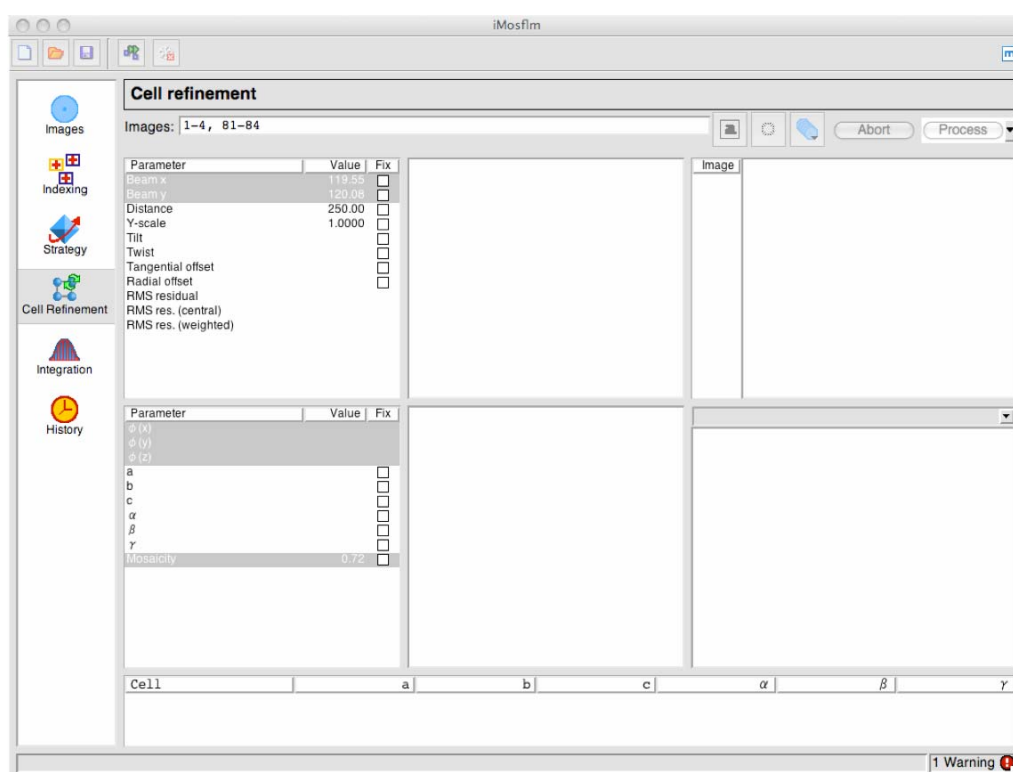
8.1 Selecting the images

There are several ways to select the images to be used in cell refinement. Whichever method is used, the image numbers will be displayed in the Images list box.
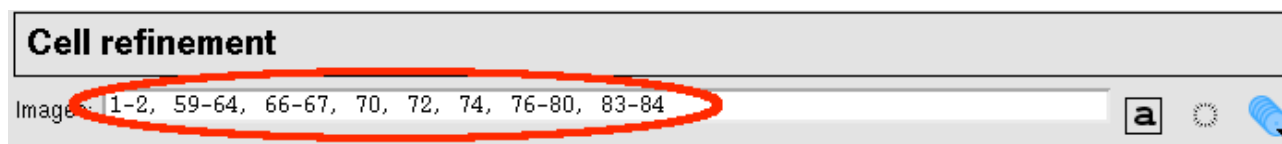

8.1.1 Manual selection

A list of images can simply be typed into this box. An image series can be specified as n-m where n and m are the first and last images. Different series must be separated by white space or comma.

For orthorhombic or lower symmetries, two segments should be given, approximately 90 apart in φ. For trigonal and higher symmetries, one segment can be sufficient, but for all but cubic symmetry, if the c axis happens to be approximately parallel to the X-ray beam for the chosen segment then it will not be well defined. Thus it is safer to use two segments in all cases, and for triclinic data three or four segments are best (eg φ = 0,45,90,135).

The number of images that should be included in each segment depends on the mosaic spread and oscillation angle. A reasonable number is 2*(mosaic spread/oscillation angle +1). MOSFLM will automatically select two appropriate segments (for monoclinic/triclinic systems you may wish to add additional segments)
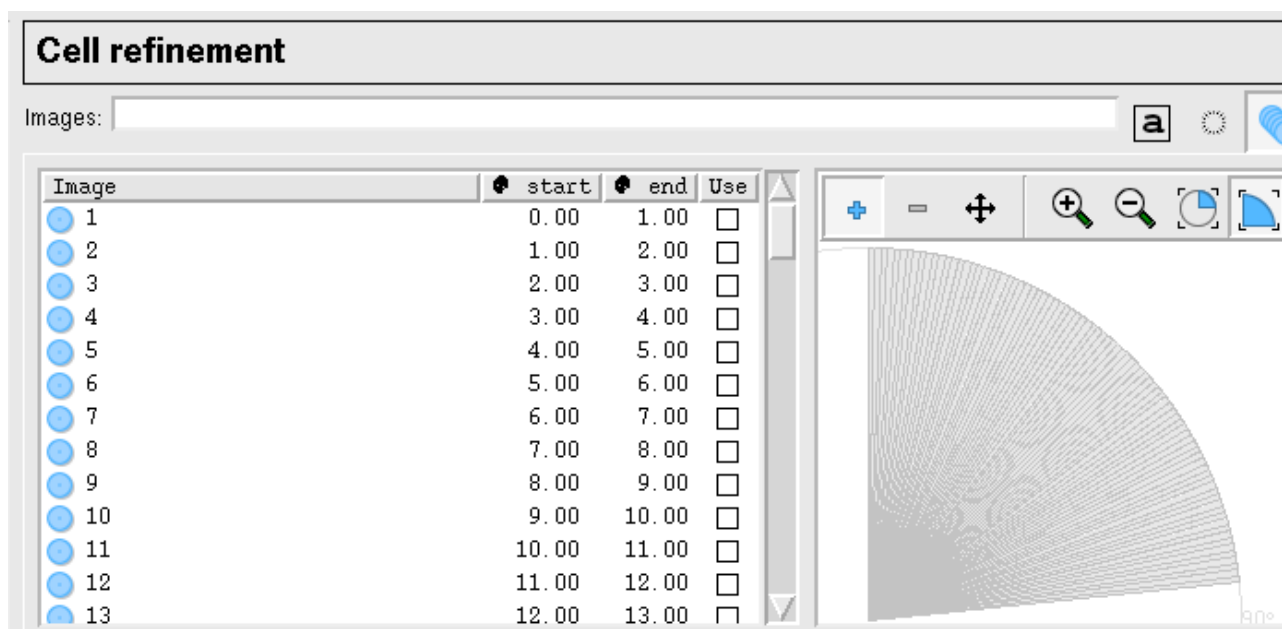
8.1.2 Automatic selection
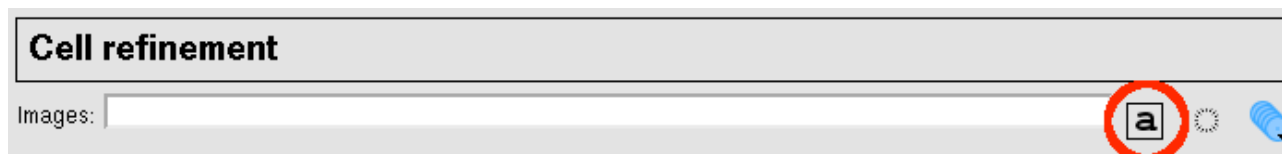
Selecting the "Automatically select images" icon will select a suitable set of images.

8.1.3 Graphical selection
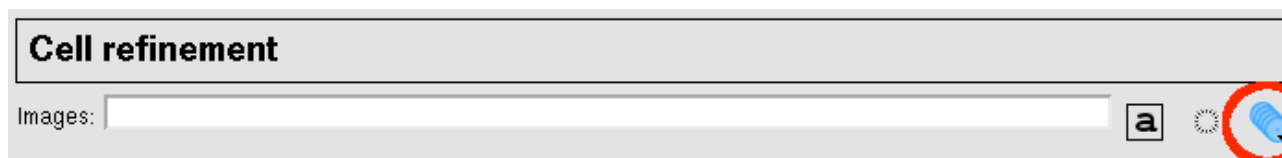
The graphical selection tool is the most powerful way of selecting images. Selecting this icon will give a new window:

Images can be selected by clicking in the "Use" checkbox in the list of images, or by using the mouse to drag a selection of images from the sector displayed. The other icons allow the sector to be zoomed in or out and the viewed area to be moved to make the selection easier (**these icons do not work in iMosflm 0.4.5 and generate tcl errors so do not use them unless you have iMosflm 0.5.1 or later**).

The final two icons fit the segment display to fit the window as either a full circle or a quadrant.

8.2 Integrating the images and refining the cell

*Use one of the methods above to select images to use in cell refinement and click on the Process button. Before starting cell refinement, check that the prediction for the first image to be used is OK. **The majority of the spot positions must be correctly predicted, or else the refinement will not work.***

The selected images will be integrated, and following integration the cell parameters are refined. Selectable refined detector parameters will be plotted, as will the changes in crystal orientation and (if selected) mosaic spread. If the shift in cell parameters is greater than 2.5 standard deviations, another cycle of integration and cell refinement is carried out. Typically, two or three cycles are required for convergence. Note that the current program activity is shown at the bottom of the window.
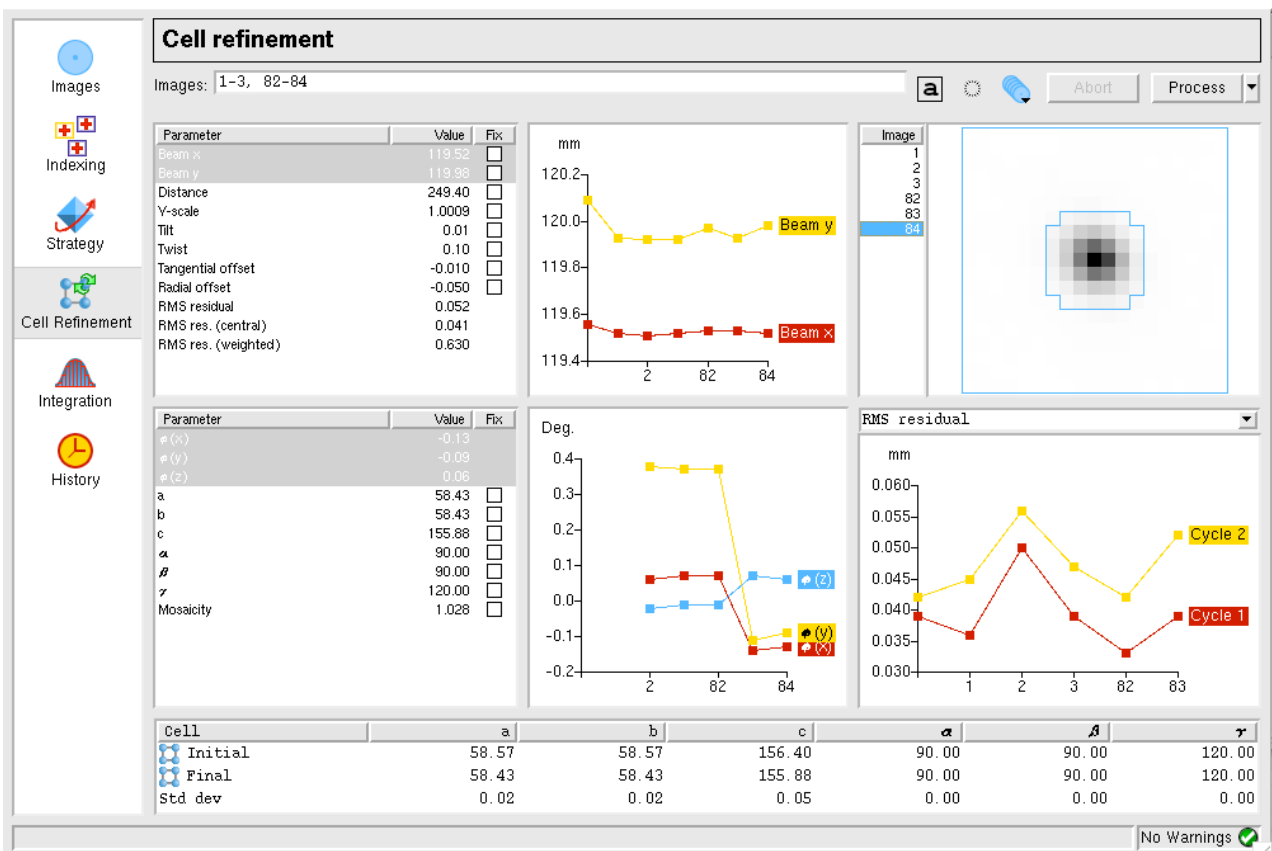
The different windows are explained below. Moving the mouse over any parameter will give a full description of that parameter.

8.2.1 The detector parameters window

The values of the refined detector parameters are displayed for each image as it is integrated. There is the option to fix any of the parameters during integration by checking the "Fix" box on the right. The positional residuals (overall, central and weighted) are also listed.

Parameter	Value	Fix
Beam X	119.52	☐
Beam y	119.98	☐
Distance	249.40	☐
Y-scale	1.0009	☐
Tilt	0.01	☐
Twist	0.10	☐
Tangential offset	-0.010	☐
Radial offset	-0.050	☐
RMS residual	0.052	
RMS res. (central)	0.041	
RMS res. (weighted)	0.630	

Any of these parameters can be selected for plotting on the graph that appears to the right of this box by simply clicking on that parameter (which is then highlighted in blue, as shown for Beam X and Beam Y).

Check the stability of the refined parameters by displaying the appropriate graphs. To get sensible scaling of the graphs, only parameters with similar numerical values should be plotted together.

8.2.2 The crystal parameters window

The values of the refined crystal missetting angles $\phi(x)$, $\phi(y)$, $\phi(z)$, the unit cell parameters and mosaic spread for the current image is displayed in the table. Selectable parameters will be plotted as for the detector parameters. Specific unit cell parameters can also be fixed.

| Parameter	Value	Fix
$\phi(x)$	-0.13	
$\phi(y)$	-0.09	
$\phi(z)$	0.06	
a	58.43	☐
b	58.43	☐
c	155.88	☐
α	90.00	☐
β	90.00	☐
γ	120.00	☐
Mosaicity	1.028	☐

If more than one segment of data is being used for cell refinement, it is not unusual to see a change in orientation of the crystal between the two segments.

Observe how the crystal orientation is different for images in the two segments. See how the mosaic spread varies during refinement.

8.2.3 The central spot profile

The average spot profile for the central region of the detector is plotted for each image. This confirms that the spot prediction is good. If the profile is not well defined and central in the box, or if the blue border for the peak region of the spot is much larger than the apparent spots size, it suggests a problem with the integration and the initial prediction should be checked. In some cases it may be necessary to increase the "Profile Tolerance" parameters to ensure that the blue box fits the spot. To do this, select the Advanced tab of Processing Options. *Check the spot profile for the images used in the cell refinement.*

Image
1
2
3
82
83
84
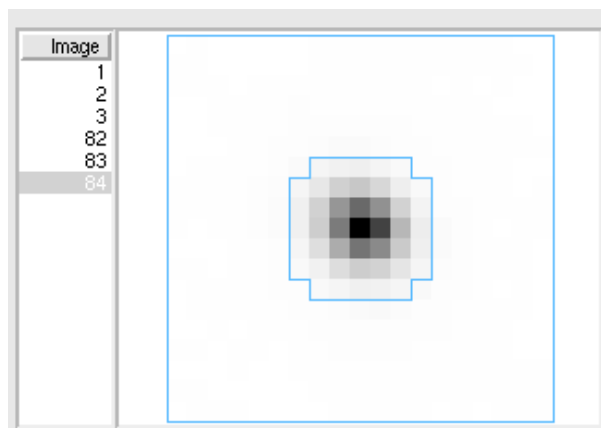
8.2.4 The summary window

This window presents a summary of how selected parameters have changed in different cycles of the cell refinement. The behaviour of these parameters gives a good indication of whether the cell refinement has been successful.

If the refinement has been successful, the RMS residual should be lower for the final cycle than in previous cycles. However, errors in the cell can be compensated quite well by changing the crystal to detector distance, so the differences are not always dramatic. In this case, the RMS error is actually slightly higher for all images in cycle 2, but not by a significant amount.
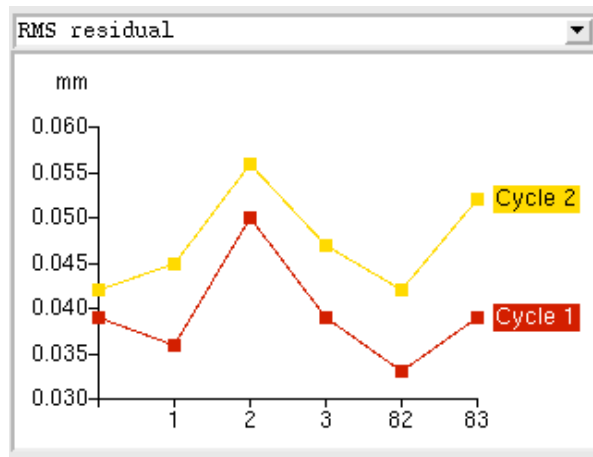
The other parameters that can be plotted are the detector distance and the Pixel ratio (Yscale).

The detector distance should be the same for all images in the refinement. If the original indexing was based on autoindexing a single image (say the first), then the detector distance might change for images from the second segment to compensate for an error in the original cell. If the variation in distance for different images increases with the cycle number, this indicates that the refinement is not stable.

The Pixel ratio (YSCALE) parameter may also show an initial variation for images in different segments, because refining YSCALE (which should be 1.0000 for all except RaxisII and Raxis HTC detectors) can compensate for errors in cell dimensions in the same way that the distance can. The final values (last cycle) should be very close to identical for all images and should also be very close to unity (to within 0.0004).

Cell refinement can be unstable if the data do not extend to at least 3.5Å. If it appears unstable, use the cell derived from autoindexing a series of images well spaced in φ instead.

Examine the plots to see if the parameters are stable and correct (YSCALE).

8.2.5 The final results

Cell	a	b	c	α	
Initial	58.57	58.57	156.40	90.00	
Final	58.43	58.43	155.88	90.00	
Std dev	0.02	0.02	0.05	0.00	

When refinement has converged (or after 5 cycles) the initial and final cell parameters and the estimated errors (std dev) are listed.

If time is available, make a note of the std dev values and repeat the cell refinement using image 1-5 and 80-84. Are the standard deviations smaller when using more images ?

9. Integration

The accurate cell parameters are now used in the integration. Note that although the images are integrated during the cell refinement, the intensities are not saved and no MTZ file is generated.

It is good practice to start by integrating a block of about ten images, to check that the parameters do not need further adjustment. MOSFLM will generate warning messages after integration if there are any difficulties, and it may be possible to improve the situation by changing some of the default parameters

(see 11).

Before integration, ensure that the backstop shadow has been defined 4.1.1 and 4.1.2)

9.1 Image selection

Image selection is performed in exactly the same way as in the cell refinement. In this case, the "Automatic" selection will simply include all images in the current sector.

Select images 1-10

9.2 Setting the MTZ filename and controlling updating of the image display

Two icons appear when the integration option is chosen.

If the "Show predictions ..." icon is clicked, the display window will be updated as each image is processed. This will slow down the processing, but allows the accuracy of the prediction to be checked for each image.
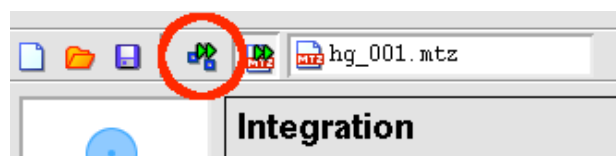
Click on this icon so that the display will be updated.

The filename of the MTZ file containing the results of the integration is generated automatically, but can be edited manually.
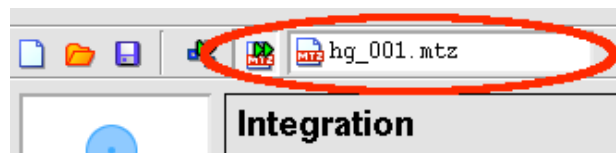
9.3 Integrating the images

Click on the Process button to start integration.

Integration of the images occurs in two passes. In the first pass, for each block of images (typically 5-10, selected automatically by MOSFLM), the detector and crystal parameters are refined for each image in turn and the "measurement boxes" for all predicted spots are written to a file for use in the second pass. In the second pass, the standard profiles are formed from reflections present on all of the images in this block, and each image is then integrated and the results written to the MTZ file. The display is not updated during the second pass, and will remain on the last image of the block (assuming image updating has been selected).

Note that the unit cell dimensions are normally fixed during integration, and only the crystal orientation and mosaic spread are refined. Crystal parameter refinement only starts when a sufficient number of images have been integrated to provide data for the post-refinement, typically after 2-3 images but this can be more depending on the mosaic spread and oscillation angle.

9.4 Parameter display windows

As described in 8.2, the refined detector and crystal parameters will be displayed in tables and selected parameters will be plotted in graphs. The average spot profile for each image will also be displayed.

The size of these graphs (and the profile) can be expanded to fill the whole window by holding down "shift" and clicking LMB anywhere in the graph window. A second "shift+click" will revert the graph to the original size.

In addition to these windows, there are windows that tabulate and plot (as a function of image number) the mean I/σ(I) for profile fitted and summation integration intensities. The overall values and the values for the highest resolution bin are given.

A display of the standard profiles for different regions of the detector is also provided. Poor profiles are "averaged", by including reflections from inner regions of the detector. Averaged profiles are indicated by a red line around the spot (rather than blue). The original "unaveraged" profile can be viewed a clicking on the profile (providing there were sufficient spots in that region to allow formation of a profile).
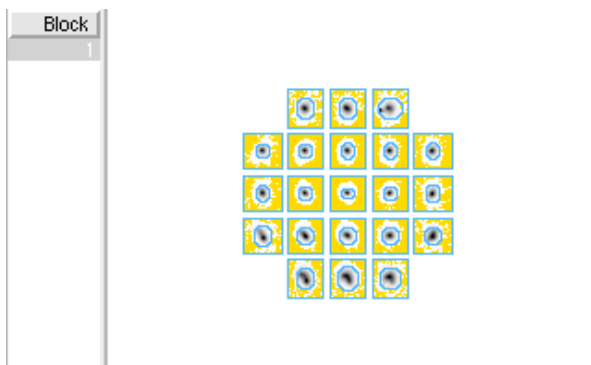
The profiles should be checked to see that they are well defined and centred within the box. (Poorly defined, diffuse or n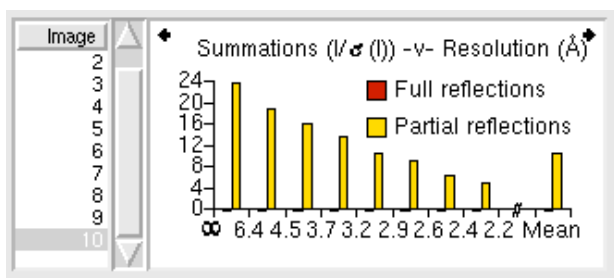on-centred profiles may suggest that the prediction is not very good, in which case this should be checked on the image).

It may be useful to expand the standard profiles to fill the window using "Shift + mouse click" in the profiles window).

Finally, the lower left window plots I/σ(I) as a function of resolution for any selected image.

Different parameters can be plotted by selecting the right or left pointing arrows.

9.5 Integrating the whole dataset

Assuming that everything looks OK, the entire dataset can now be integrated. However, it is strongly recommended that Project/Crystal/Dataset names are assigned first. These are written to the output MTZ file and used by downstream programs, in particular SCALA will use this information when deciding which images belong to the same dataset.

Select "Experiment settings" from the "View" drop down menu (2.1) to define these names and optionally give a title for the MTZ file.

Select images 1-84 and click on "Process".

9.5.1 Integrating as a background job

hg_001.mtz   QuickSymm   QuickScale

Integration

Images: 1-84   a   Abort   Process ▼

Batch

Parameter	Value	Fix
Beam x		

Image

Indexing
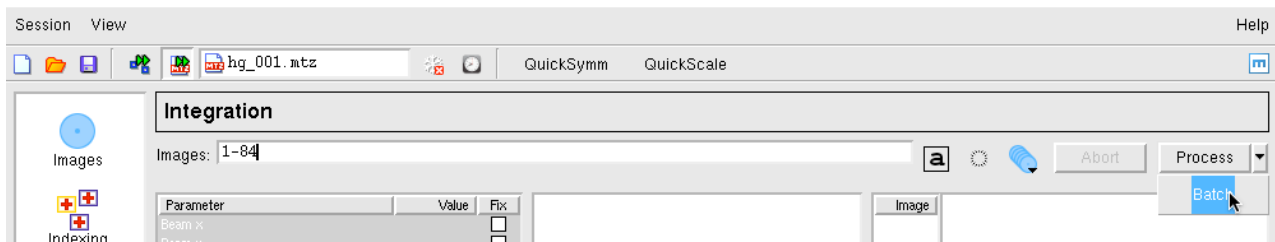
By selecting "Batch" in the drop down menu by the Process button, the integration can be run as a background job rather than interactively, and the GUI can be used to look at a different dataset.

Selecting "Batch" from the drop down menu will put up a new window showing the script that will be used to process the images. This script can be edited if additional keywords are required or if some parameters are to be changed. It is possible to configure job

Batch job submission

Script:

```
directory /andrewgrp0/andrewgrp1/andrew/test/examples/camillo
template hg_###.mar1600
pname New Project
dname New Dataset
xname New Crystal
title Untitled
beam 119.52 119.99
distance 249.60
wavelength 0.999700
twotheta 0.00
divergence 0.02 0.02
gain 1.00
pixel 0.15000
bias 0
dispersion 0.0002
polarisation synchrotron 0.95
matrix 0.00687576 -0.01132296 -0.00230916 0.01846737 0.01487094 0.00073012 0.00040468 -0.0
0625535 0.00591558
cell 58.44 58.44 156.02 90.00 90.00 120.00
symmetry h3
mosaicity 1.024
hklout /andrewgrp0/andrewgrp1/andrew/test/newgui/tutorial/hg_001.mtz
genfile /ss0/andrew/.mosflm/genfile
resolution exclude none
```

Send to: ls1.lmb.internal ▼   Configure...

Cancel   Ok

submission
for remote
machines.

The logfile
for batch
submission
jobs is
written to
the
".mosflm"
directory
(created by
imosflm if it
does not
exist) in
your home
directory.
The MTZ
and
SUMMARY
files are
written to
the directory
in which
imosflm was
launched.

9.6 Checking the integration

The detector and crystal parameter plots should be examined carefully to check for any instability in the refinement. If there are large and random variations in some parameters (eg the detector twist and tilt) then it may be better to fix them and repeat the integration. Discontinuities due to blank images should also show up, and the offending images removed from the scaling run.

If adjacent spots are incompletely resolved on the detector, it may be possible to improve the processing by increasing the PROFILE TOLERANCE parameters by 1-2%. These parameters can be set in the Advanced tab of the Processing Options menu (use View).

9.7 Advanced features for integration

The "View"/"Processing options" drop down menu allows additional control over the integration. Parameters such as the minimum spot separation, resolution limits, block size, MTZ filename and BATCH ADD can be changed in the Processing tab. In the Advanced tab, the measurement box parameters, profile tolerance and profile averaging parameters can be set.

Additional MOSFLM parameters will be added based on user requests.

10. Running Pointless to check the symmetry

Once the images have been integrated (even a subset of the whole dataset) the program POINTLESS can be run determine the true Laue symmetry and try to determine the spacegroup.
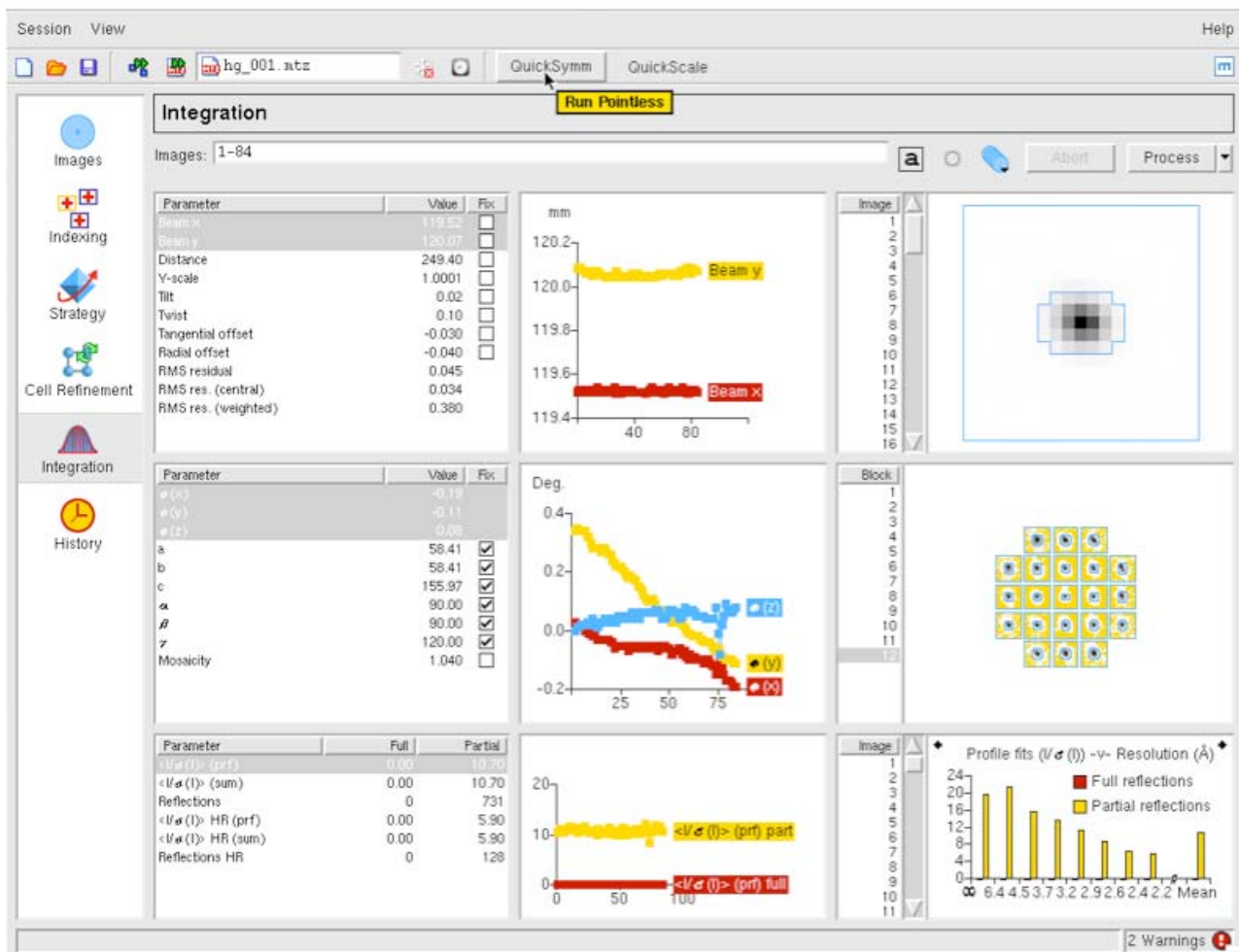
Select the QuickSymm button. A summary of the Pointless results as shown below will appear in a browser window. (Note that the environment variable CCP4_BROWSER must be set to a web browser that is installed on your machine for this to work).

pointandscale.log

Please consider citing the following papers:

- Pointless
  - P.R.Evans, 'Scaling and assessment of data quality' Acta Cryst. D62, 72-82 (2005).

Pointless *Version 1.2.18* Run at 17:52:42 on 23/ 9/2008

Result:

Best Solution space group H 3 2

Reindex operator:	[h,k,l]
Laue group probability:	0.999
Systematic absence probability:	1.000
Total probability:	0.999
Space group confidence:	0.999
Laue group confidence	0.999

Unit cell: 58.41 58.41 156 90 90 120

[Show logfile summary] [Show full logfile]   [Documentation]

Generated for you by baubles 0.0.7 on Tue Sep 23 17:52:44 2008

Caution: The results can be ambiguous if run with a partial dataset.

11. Running Scala to scale and merge the data

Select the QuickScale button. This will first run POINTLESS, then change the spacegroup for the data to that selected by POINTLESS and then run SCALA to scale and merge the data in that spacegroup. SCALA is run with default parameters. To change any of these, SCALA will need to be run from the ccp4i GUI (see section 16). The SCALA summary will be displayed in the web browser, as shown below, but graphical output (see below) and the full logfile can also be examined.

Scala summary file:

Scala Version 3.3.6 Run at 17:54:02 on 23/ 9/2008 Finished with: ** Normal termination **

Result:

Summary data for Project: New Crystal: New Dataset: New

	Overall	InnerShell	OuterShell
Low resolution limit	48.12	48.12	2.37
High resolution limit	2.25	7.12	2.25
Rmerge	0.084	0.073	0.148
Rmerge in top intensity bin	0.052	-	-
Rmeas (within I+/I-)	0.094	0.082	0.166
Rmeas (all I+ & I-)	0.094	0.082	0.166
Rpim (within I+/I-)	0.041	0.037	0.074
Rpim (all I+ & I-)	0.041	0.037	0.074
Fractional partial bias	-0.003	-0.010	-0.011
Total number of observations	23891	775	2776
Total number unique	4933	172	596
Mean((I)/sd(I))	12.2	24.4	6.7
Completeness	97.1	94.8	82.5
Multiplicity	4.8	4.5	4.7

Average unit cell: 58.41 58.41 155.97 90.00 90.00 120.00

Part of the Scala graphical output:

The following tables were found in the logfile:

>>> Scales v rotation range, New

Analysis against all Batches for all runs , New

Analysis against resolution , New

Analysis against intensity, New

Completeness, multiplicity, Rmeas v. resolution, New

Correlations within dataset, New

Axial reflections, axis h, New

Axial reflections, axis l, New

Run 1, standard deviation v. Intensity, New

Rmerge v Batch for all runs
Imean & RMS Scatter
Imean/RMS scatter
Number of rejections v Batch

Analysis against all Batches for all runs , New

[Show logfile summary] [Show full logfile]   [Documentation]

12. History and mosflm logfile

Selecting History allows the history of the session to be examined, or, by selecting the Log tab, the full mosflm.lp file can be viewed.

The History has a tree structure, and will show when various parameters have been changed. It is possible to "Undo" some of the actions shown in the History, although this feature has not been extensively tested.

13. Warning Messages

As already mentioned in sections 3 and 9, warning messages are generated by MOSFLM during processing and the abbreviated messages can be obtained by clicking on the "Warnings" icon (see 3). The full warnings will eventually be available via the GUI, but at present it is necessary to look at the mosflm.lp file (or the Log file via the History) to get full details and suggestions.

> **NOTE This is the end of the part of the tutorial which deals directly with iMosflm (but see section 19 which has some suggestions for further work if you have time). The remainder of this document deals with integrating and scaling data with MOSFLM and Scala via the CCP4i user interface, examining the traditional MOSFLM SUMMARY file with the CCP4 utility loggraph.**
>
> **There are also two appendices, one which contains a shell script for running MOSFLM as a traditional batch job, and one which gives an example of using the TESTGEN option in MOSFLM in order to work out the optimal oscillation angle for the images in the data collection.**

14. Using the CCP4i GUI to integrate images

By selecting to submit a background processing job (see 9.5.1) it is possible to use the CCP4i GUI to integrate the images. As described in 9.5.1, a window with the MOSFLM commands will appear. These commands should be pasted into a file (eg hg.sav). It is also necessary to save the orientation matrix to a file. To do this, select the Images window and double click on the matrix. A window will pop up which allows the matrix to be saved to a file.

```
> ccp4i
```

Select "Directories & ProjectDir" (Top left)

In the new window, select "Add project"

Type in the Project and the full directory path (uses directory)

Select this project from "Project for this session..."

Select "Apply and Exit" ... Dismiss the warning message that comes up.

Select "Data reduction" from the pull down menu of modules (orange bar at top left; if you haven't used CCP4i before, it probably says "Coordinate utilities", but the state is saved from the previous session) so it may be different.

Select "Integrate Images" (the first option).

A window will be displayed which allows you to set up a batch MOSFLM job.

You need to provide the information stored in the "save" file.

Give a job title. On the next line ("Load parameters from command file") select "browse" and locate the ".sav" file that you saved from the MOSFLM session. Once you have selected the file, click on "Load parameters from command file" and the information will be read from the save file.

You should then supply a "Crystal" and a "Dataset" names.

Then in the box headed "Images to Integrate", click on "Add processing run", and in the resulting block click on "Add block of images". Supply the number of the first and last image (1 and 84). The job can then be submitted.

15. The MOSFLM SUMMARY file

MOSFLM produces a summary file listing refined parameters for each image, and when using the new GUI this file will be called always be called SUMMARY (and will be overwritten by any subsequent job unless renamed). This contains the same information as the plots produced by the new GUI, but in addition it lists the number of "badspots" and overloads, which are not yet available in the new GUI.

This can be inspected graphically with the CCP4 program "loggraph":

```
> loggraph SUMMARY
```

This is very useful to identify "rogue" images (with unusually high positional residual, or low $I/\sigma(I)$). If you find any, read that image into MOSFLM and see what is wrong with it.

If you have done more than one integration run during your MOSFLM session, you will find multiple entries of the tables. Look at the last set for integration of the 84 images.

Click on the "Refined detector parameter" tables and check on the stability of parameters like the TILT and TWIST of the detector (units are hundredths of a degree) and, for Mar Research (or DIP) image plate data, the distortion parameters ROFF and TOFF (units are mm). If they are varying a lot (more than 20 for TWIST/TILT) or 0.15 for ROFF/TOFF then it is probably a good idea to fix these parameters at the average value (or the known values for this detector, if they are available) (see section 8.2.1).

Click on the "Post refinement" table and check the missets...it does not matter if they change slowly and by an amount (per image) that is less than 0.1*mosaic spread. If they are changing more than this then there could be a problem with processing the data (there is not much you can do about this).

Check how the mosaic spread is changing. If it is unstable, or if you think it is not refining to sensible values (as judged by looking at the prediction) then you can fix the input value in the same way as the detector parameters (see section 8.2.2).

BADSPOTS AND OVERLOADS

The program will set a rejection flag for those reflections that fail certain tests, for example too much variation in the background level, a poor profile fit, an intensity that is very negative (more than 5 sds), a very high gradient for the background plane. These reflections are called "Bad spots" and they are listed individually for each image in the mosflm.lp file, together with the reason for flagging them. The number of bad spots is also written to the summary file. Generally there should be very few (5-10) bad spots on each image. If there are more, it could be because the backstop shadow is not correctly allowed for (consider using the NULLPIX keyword).

There should also be very few overloaded reflections on each image. If there are a lot, then a separate low resolution data collection pass should be made with a much shorter exposure time.

IMPORTANT: Both "Bad spots" and "Overloads" are written to the output MTZ file, but will be rejected by default by SCALA. The intensity of overloaded reflections is estimated by profile fitting that part of the spot that is not saturated. To include these reflections in the final merged data, use the ACCEPT keyword in SCALA.

16. Scaling data with CCP4i

```
> ccp4i
```

Select "Directories & ProjectDir" (Top left) In the new window, select "Add project"

Type in the Project and then the full directory path (uses directory) or use the "Browse.." button to select it.

Select this project from "Project for this session..."

Select "Apply and Exit" ... Dismiss the warning message that comes up.

Select "Data reduction" from the pull down menu of modules (orange bar at top left)

Select "Scale and Merge intensities"

In the window that appears:

1. Give a job title (top line)

2. Click on the box next to "Separate anomalous pairs for merging/output" (the crystals have been soaked in a mercury compound).

3. In the next orange bar (line starts "MTZ in") select Browse and select the MTZ file output by MOSFLM (probably called hg_001.mtz) Select "OK"

4. In the next orange box ("Estimated number of residues...") Enter 91

   Under "Scaling Protocol":

5. From the pull down menu on the line starting "Scale", select "On rotation axis with secondary beam correction"

6. On the same line, from the pull down menu select "Isotropic B" (default is no B-factor correction).

7. At bottom of window, from pull down menu named "Run" select "Run and view command file"

   This will put up a new window showing the command file for running the "Sort" step, select "Continue". After a brief pause it will show the command file for running SCALA. (This can be edited before submitting the job). Select "Continue"

   In the main window, it will show that the SCALA job is running.

   When SCALA has finished, the command file for running the TRUNCATE step will appear...select "Continue".

   When TRUNCATE has run, the main window will say "FINISHED".

   At that point, select the pull down menu "View files from job" and select "View log graphs". This will give you the same graphs are running loggraph on the SCALA logfile. To look at the logfile itself (eg the overall merging statistics) select "View logfile" and it will appear in a new window.

   When SCALA has finished, the command file for running the TRUNCATE step will appear...select "Continue".

16.1 Looking at the SCALA output

The simplest way to look at the output from SCALA is to use loggraph on the log file (or via the CCP4i GUI). This can display many different graphs.

Use "Scales vs rotation range" to check for smooth variation in the scale factor. Check the variation in B factor with image, and if there is no real variation it is best to turn off the B-factor refinement.

Use the "Analysis against batch" to detect "bad" images (high R-factor, large number of rejected spots).

Check the "Analysis against resolution" Fractional bias to see if there is any indication of "Partial bias" (A negative partial bias will result if the mosaic spread is underestimated, or if there is a lot of diffuse scatter. The TAILS correction can be used in SCALA to correct for diffuse scatter.

The Fractional bias should be less than 1-2%, although it will often exceed this for weak data (eg in the high resolution bins).

The "Axial reflections" graphs are useful for detecting systematic absences, which can be used to identify the true space group.

It can also be useful to look at the SCALA logfile itself, in particular the table giving statistics as a function of resolution. The effective resolution limit of the data can be determined by looking at the Mn(I)/sd column (this is the mean $I/\sigma(I)$ AFTER merging symmetry mates, and is the best indicator of data quality). This table also shows shell and cumulative R-factors.

The Mn(I)/sd values depend on having realistic values of the standard deviations (errors) in the intensities. As the values that come from MOSFLM are always underestimates of the true error, these values are scaled up in SCALA using a two-term correction:

sdcorrected = SdFac * Sqrt[sd(I)**2 + SdB*LP*I + (SdAdd*I)**2]

Here "SdFac" is an overall scale factor, and "SdB" and "SdAdd" are intensity dependent factors.

SCALA will automatically work out a suitable value for all three parameters, in order to make the mean value of (observed scatter)/(sdestimate) equal to 1.0, where "observed scatter" is the differences between an individual estimate of intensity and the mean of all other estimates (from symmetry mates).

This table comes under the heading:

ANALYSIS OF STANDARD DEVIATIONS

in the SCALA logfile.

Can you identify a bad image in this dataset ? Can you work out why it is bad ?

(Clue...the image headers contain quite a lot of information that is not used by MOSFLM, including the date and time that the image was collected).

You can exclude a given batch from the scaling using the EXCLUDE keyword:

RUN ALL EXCLUDE imagenumber

In this case, runnumber will be "1".

Rerun the scaling, SENDING THE OUTPUT TO A NEW LOGFILE, excluding this image, and see the effect on the heights in the anomalous Patterson (in the logfile). DO NOT repeat the integration. In the command file, uncomment the line (near the top) "goto scala" so it skips the integration part.

17. Changing the symmetry

From the list of options at the left hand side of the main window, select "Sort/Reindex MTZ files" In the new window give a new title, eg reindex as H32

Click on the Box "Change space group or reindex reflection"

Use the "browse" option to select the input MTZ file (First orange bar), the output filename will be

generated automatically (but you can change it).

Under "Reindex details"

In the last option "Reduce to asymmetric unit for space group ..." (which is the one selected by default) enter H32 in the box.

Select "Run" from "Run" pull down menu.

The go back to the scaling window, and select the new MTZ file (reindexed) as the input for scaling (use "Browse"). CHANGE THE JOB TITLE

Then select "RUN"

The scaling will now be performed in space group H32. Look at the merging statistics in the logfile, in particular the multiplicity weighted R-factor, called Rmeas in the logfile.

18. Looking at the TRUNCATE output

The TRUNCATE program converts intensities to amplitudes, and also compiles some useful statistics. The so called "cumulative intensity statistics" tabulated by Truncate (and plotted by loggraph) provides the only point at which you will be able to detect merohedral twinning (when the two lattices of the twin components exactly overlap, and every measured intensity is actually the sum of two intensities).

For a good dataset, the observed distribution should be within 1% of the theoretical distribution. Is this the case for this data ?

19. If you have time ....

19.1 Different mosaic spreads

Try repeating the processing with the mosaic spread set to a value 25% smaller than the refined one and fix it by checking the "Fix" box in the Integration window.

Can you tell from the merging statistics if this data is better ?

19.2 Checking up on outliers

SCALA writes a file called ROGUES which lists reflections which show very poor agreement between symmetry mates, or which are implausibly large.

Look in this file (it is ASCII), and try to work out why this has happened. The ROGUES file gives you the image (Batch) on which the reflections have been recorded (for partials, this is the image nearest the centre of the reflection, so you may need to look on the preceding and following image as well). Select a reflection which shows very poor agreement with its symmetry mates (A value greater than 10 in the DelI/sd column in the ROGUES file). Read in the offending image using "Read image" in the Main menu. Predict the reflections, then use "Find hkl" to locate the offending reflection in the image (you must give the measured indices, the first set of values in the ROGUES file, when doing this). See if there is anything odd about the spot. Remember to check adjacent images for partials.

19.3 How accurate does the unit cell have to be ?

Try changing the cell parameters by (say) 1.0% and integrate the images again. What is the effect on the merging statistics ?

Appendix I

A command script for running MOSFLM, using the "save" file from the old GUI or from background processing of the new GUI. Note that keyworded input for MOSFLM and other CCP4 programs is case-insensitive apart from file and directory names.

```csh
#!/bin/csh -fv
#
# Maxinf Workshop on Data Processing, Cambridge, December 2002
#
# A generic command file for processing data. For completeness,
# I have left in the steps that calculate an anomalous Patterson map.
# Andrew Leslie
# Variables to generate unique filenames. Change the "job" variable
# if processing multiple times, eg with different mosaic spread values.
# "scr" defines a scratch disk
#
set ident = "hg"
set job = "e"
set scr = "/scr0/andrew"
#
#
Define Title, first and last image numbers, project and dataset names
# Number of amino acids
#
set title = "Camillo Rosano Hg deriv HypF N-terminal domain"
set img1 = 1
set img2 = 84
set project = HypF
set dataset = Hg
set nres = 91

# Beam parameters
set diverg = "0.02 0.02"
set disp = 0.0002
set polar = 0.95

start:

mosflm:
ipmosflm spotod $scr/{$ident}_{$job}_{$img1}to{$img2}.spotod \
        COORDS $scr/{$ident}_{$job}_{$img1}to{$img2}.coords \ summary
        {$ident}_{$job}_{$img1}to{$img2}.sum \
        << eof-mosflm_ip
TITLE $title
!!!!
!!!! Insert the "save" file here
!!!!
! Project and dataset names ... strongly recommended
PNAME $project
DNAME $dataset
! MTZ and "generate" filenames (genfile is a scratch file)
HKLOUT ${ident}_${job}_${img1}to${img2}.mtz
GENFILE ${scr}/${ident}_${job}_${img1}to${img2}.gen
! Beam parameters
DIVERGENCE $diverg
DISPERSION $disp
POLARISATION SYNCH $polar
! Do not refine the cell during integration
POSTREF FIX ALL PROCESS $img1 TO $img2
PLOT
RUN
eof-mosflm
#
# Delete temporary files coords, genfile
#
/bin/rm ${scr}/${ident}_${job}_${img1}to${img2}.coords
/bin/rm ${scr}/${ident}_${job}_${img1}to${img2}.gen
#
# sort: sortmtz hklout ${ident}_${job}_sort_${img1}to${img2} << end-sort
H K L M/ISYM BATCH
```

```
${ident}_${job}_${img1}to${img2}.mtz
end-sort
# scala:
#
# Smooth
scaling, Bfactors ON, absorption correction
#
scala hklin ${ident}_${job}_sort_${img1}to${img2} \
     hklout ${ident}_${job}_scala_${img1}to${img2} \
     rogues ${ident}_${job}_sort_${img1}to${img2}.rogues \
     << end-scala
title $title
scales rotation spacing 5 secondary 6 bfactor on tails
sdcorrection 1.5 0 0.03
anomalous on
end-scala

truncate:
truncate hklin ${ident}_${job}_scala_${img1}to${img2} \
        hklout ${ident}_${job}_trunc_${img1}to${img2} \
        << end-trunc
title $title
nresidues $nres
labout F=FHg$job SIGF=SIGFHg$job
anomalous yes
end-trunc
```

Harry Powell
Andrew Leslie
Last modified: Wed Oct 02 11:48:30 BST 2008