iMosflm Tutorial

Table of Contents

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, AIMLESS, CTRUNCATE) 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. For details of installation visit http://www.mrc-lmb.cam.ac.uk/harry/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 iMosflm.

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. This contains considerably more detail about how ipmosflm processes diffraction images than this tutorial does. 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 ipmosflm) 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"

See also: Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. & Leslie, A.G.W. 2011. iMosflm: a new graphical interface for diffraction image processing with MOSFLM. Acta Cryst. **D67**, 271-281.

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 amino acids) 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.

Note: You can equally well choose to process your own diffraction images if you wish

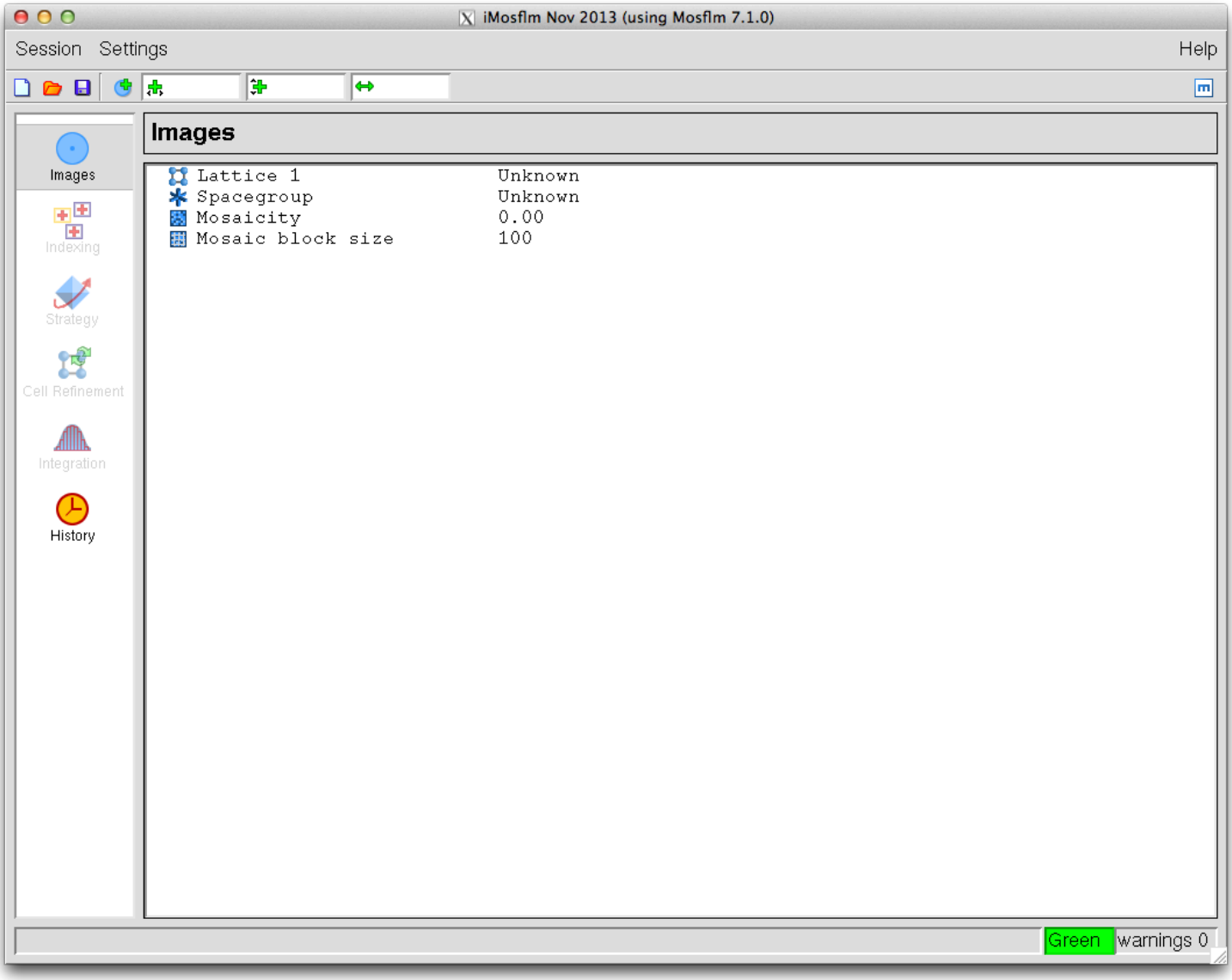
1.5 Mouse and key functions

For simplicity, the mouse and key functions are all described here, although they will also be referred to at various points in the tutorial.

When positioned over the image, right mouse button will return the display to the full size image if it has been zoomed. Scrolling middle mouse will zoom/shrink the image.
Holding down the "Ctrl" key (or "Alt" or "Cmd" (for Macs)) will display the resolution and the current mouse position on the image in mm and pixels. If positioned over a found spot, the spot coordinates and I/σ(I) will be given. If positioned over a predicted spot position the *hkl* indices will be shown.
Holding down the "Shift" key while the "Zoom" icon in the image is selected will show the image within the small dotted square zoomed within the solid square.

2. Overview of iMosflm

Start the program by typing "imosflm". After a brief pause, the following window will appear:



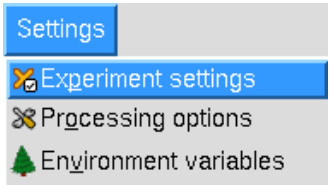
The basic operations listed down the left hand side (Images, Indexing, Strategy, Cell Refinement, Integration, History) can be selected by clicking on the appropriate icon, but those that are not appropriate will be greyed-out and cannot be selected.

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 read/save a site file (see §14.1) or add images):

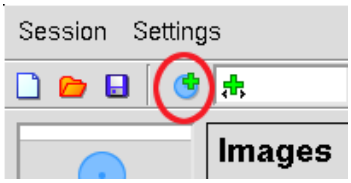
Clicking on "Settings" will allow you to see (and modify) Experiment settings, Processing options and Environment variables. These will be described later.



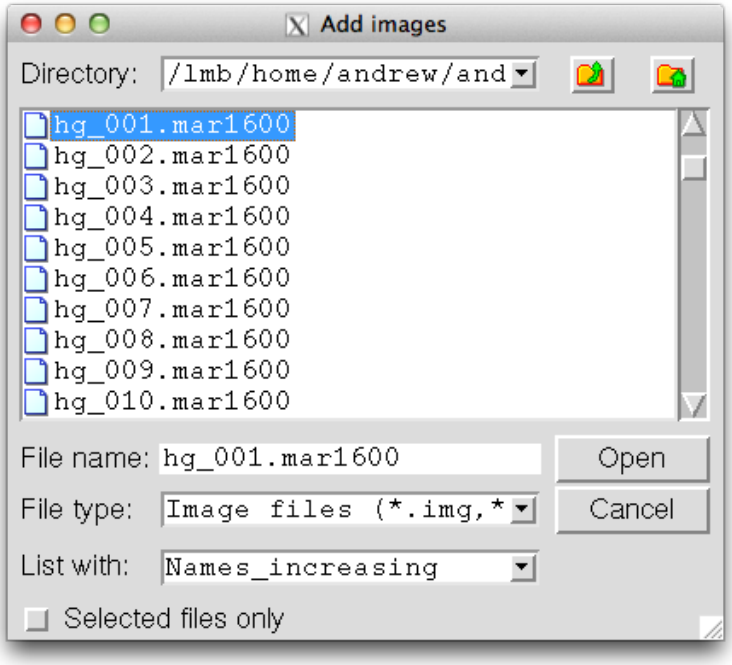
The three small icons below "Session" allow you to start a new session, open a saved session or save the current session. Moving the mouse over these icons will result in display of a tooltip describing the action taken if the icon is clicked.

3. Adding images to a session

To add images to a session, use the "Add images..." icon:

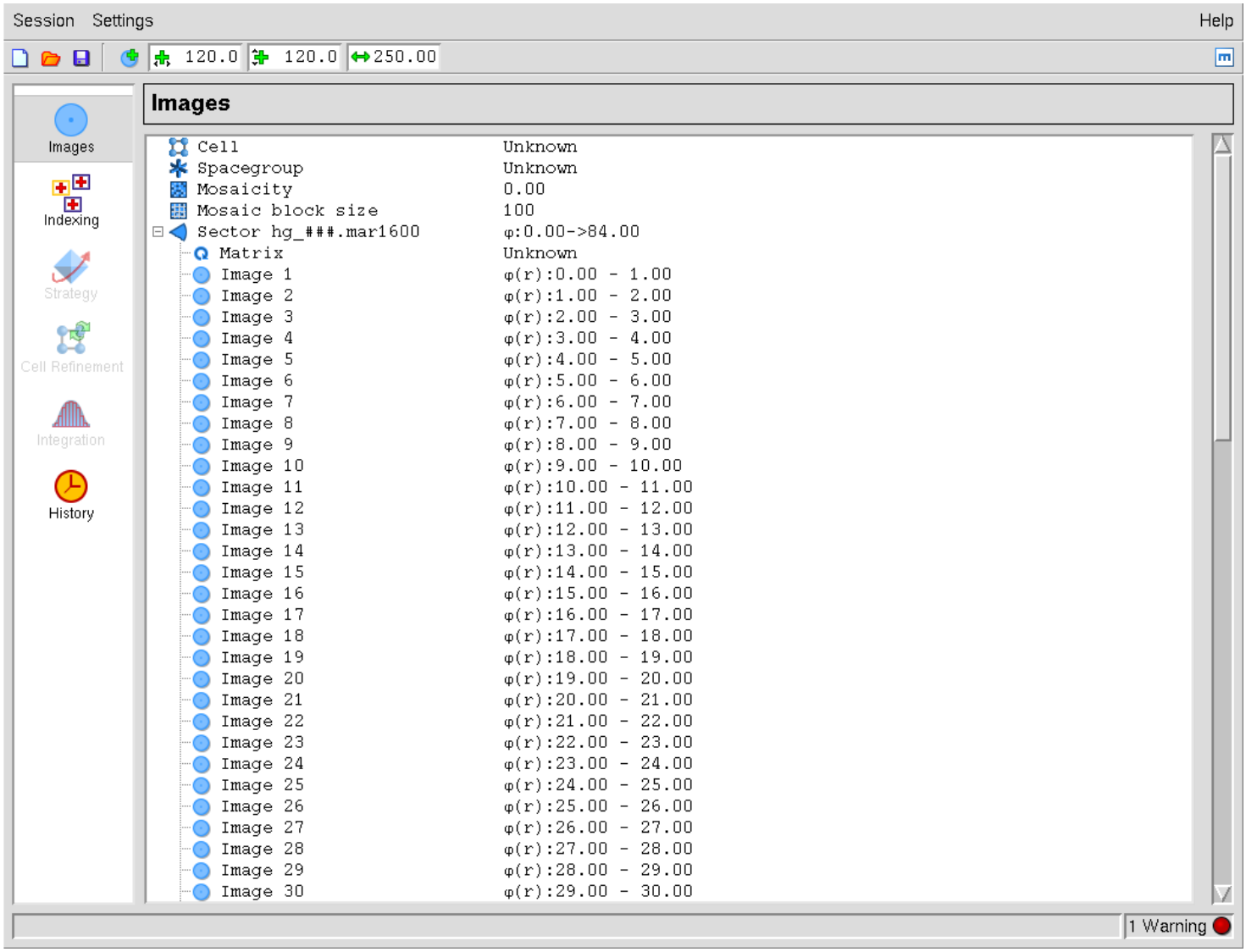


Select the correct directory from the pop-up Add Images window (the default is the directory in which iMosflm was launched). All files with an appropriate extension (which can be selected) will be displayed. Double-clicking on any file will result in all images with the same template being added to the session. The template is the whole part of the filename prefix, except the number field which specifies the image number. An alternative is to single-click on one image filename and then click on Open.



To open one or several images only, check the 'Selected images only' box and then use click followed by Shift+click to select a range of images or Control+click to select individual image files as required.

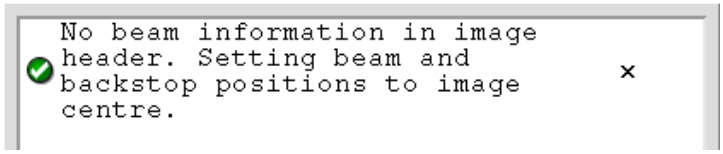
Loaded images will be displayed in the Images window with the start & end phi values displayed (as read from the image header).



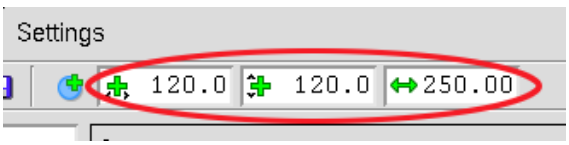
All images with the same template belong to the same "Sector" of data.

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 on the “1 Warning” text to get a brief description of the warning. In this case it is because the direct beam coordinates stored in the 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 green tick on the right hand side to dismiss this warning; click anywhere outside the warning box to collapse the box; double-click the warning text itself and more details, hints and notes may be available from MOSFLM.



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

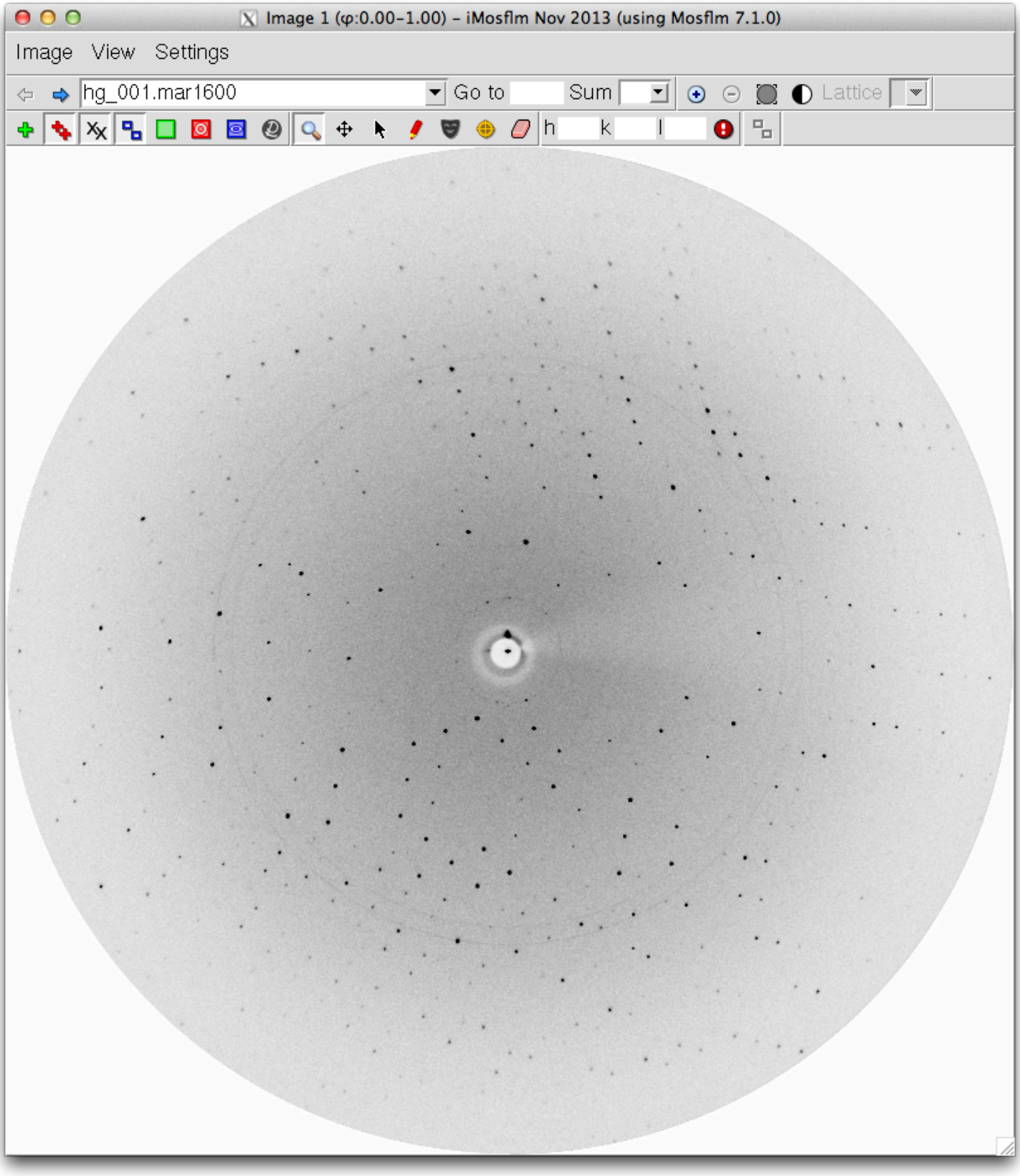
1. The phi values of images can be edited in the Image window. First click on the Image line to highlight it then click with the mouse over the phi values to make them editable allowing new values to be given. These new values are propagated for all following images in the same sector. Starting values for the three missetting angles (PHIX, PHIY, PHIZ) may also be entered following the image phi values.

2. To delete a sector click on the sector name to select it (it turns blue) then use the right mouse button on the sector to bring up a "delete" button. Move the mouse over the delete button (it turns blue) and click to delete the sector and its images. Individual images can be deleted from the Images pane in the same way.

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. Alternatively, it is possible to "drag and drop" the matrix from the first sector onto the second sector (drop onto the sector itself, not the Matrix for the second sector).

4. Image Display

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



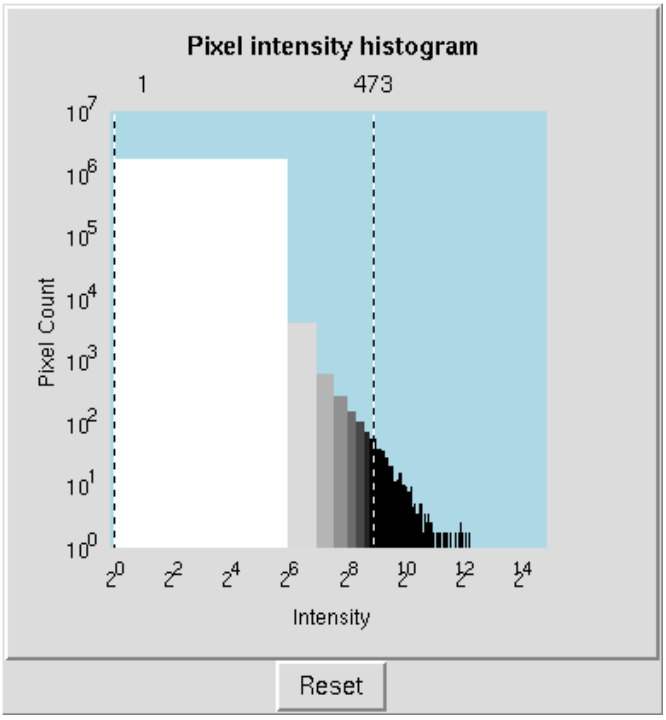
The "Image" drop-down menu allows display of the previous or next image in the series.

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 "Settings" drop-down menu allows the colour of the predictions to be changed (see Indexing §5.2 for the default colours),

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. Alternatively the image number can be typed into the "Go to" entry box. The "Sum" entry box will allow the

display of the sum of several images, but this is not yet fully implemented.The image being displayed can also be changed by double-clicking on an image name in the "Images" pane of the main window. This will also re-active the image window if it has been closed.



"+" 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. The "Lattice" entry window can be used when processing images with multiple lattices present, see multiple lattices §5.3)

4.1 Display Icons

The eight icons on the left, control the display of the direct beam position, spots found for indexing, bad spots, predicted spots, masked areas, spot-finding search area, resolution limits and display of the active mask for Rigaku detectors respectively.

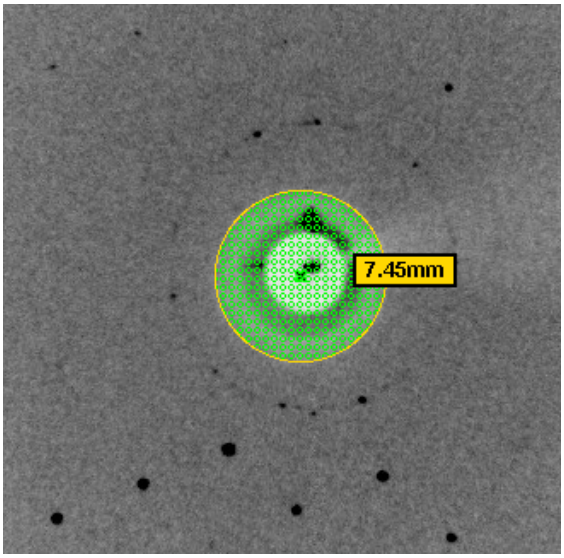
These are followed by icons for Zoom, Pan and Selection tools, and tools for adding spots manually (for indexing), editing masks, circle fitting and erasing spots or masks.

Next on this toolbar are the entry boxes for *h, k & l* and a button to search for this *hkl* among the predicted spots displayed on the image. The button resembles a warning sign if the given *hkl* cannot be found, but a green tick if it is found. The final icon relates to processing images with multiple lattices, (see multiple lattices §5.3) and is not described here.

4.1.1 Masked areas - circular backstop shadow

Select the masked area icon. A green circle will be displayed 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 zoomed area.



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

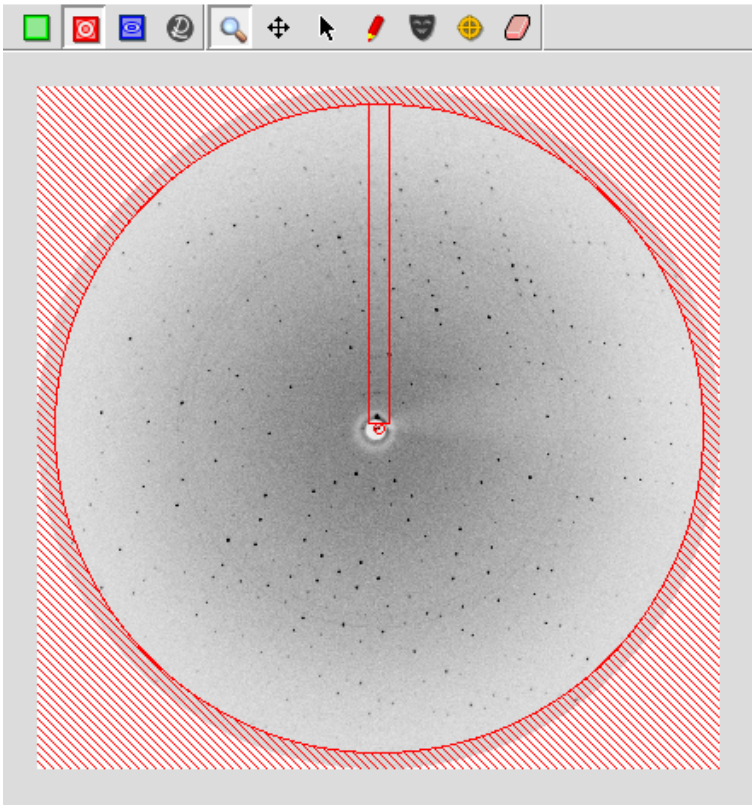
Choose the Masking tool. Any existing masked areas will automatically be displayed. *Use LMB to define the four corners of the region to be masked*. When the fourth position is given, the masked region will be shaded. This region will be excluded from spot finding and integration. This provides a powerful 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*.

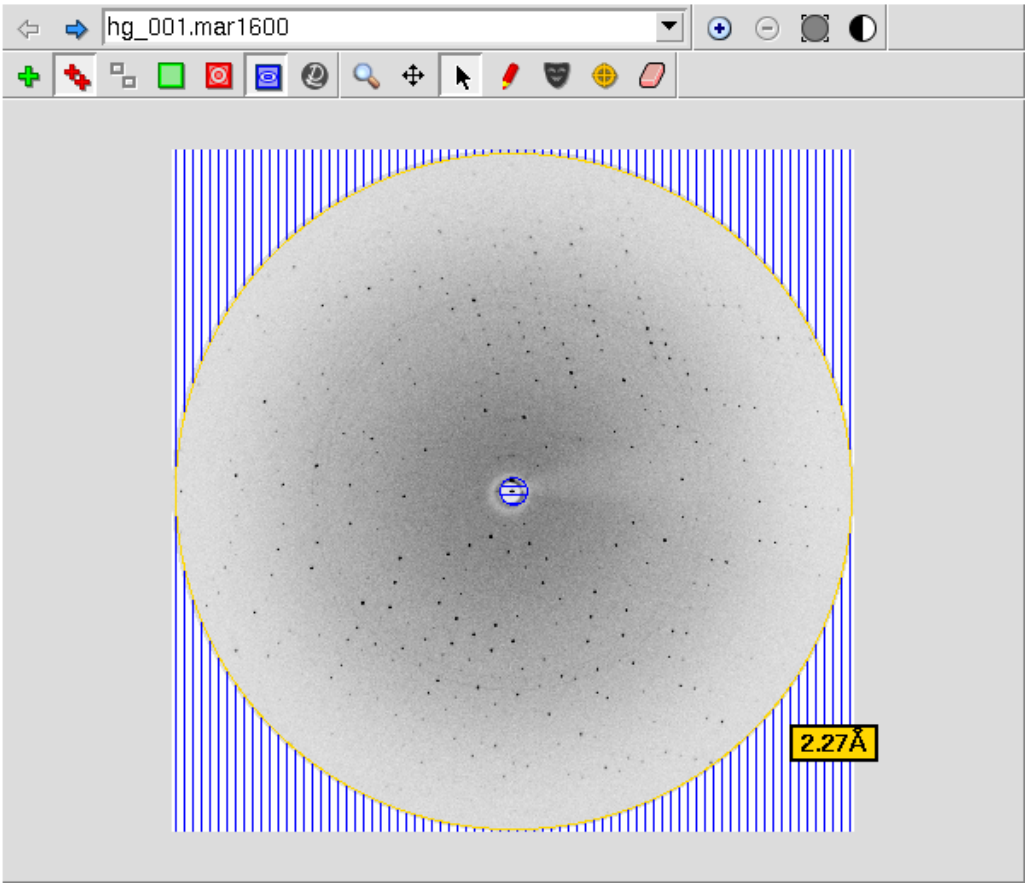


Advanced Usage

The red rectangle displays the area used to determine an initial estimate of the background of the image. It is important that this does not overlap significant shadows on the image. It can be shifted laterally or changed in orientation (in 90° steps) by dragging with the LMB.

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.

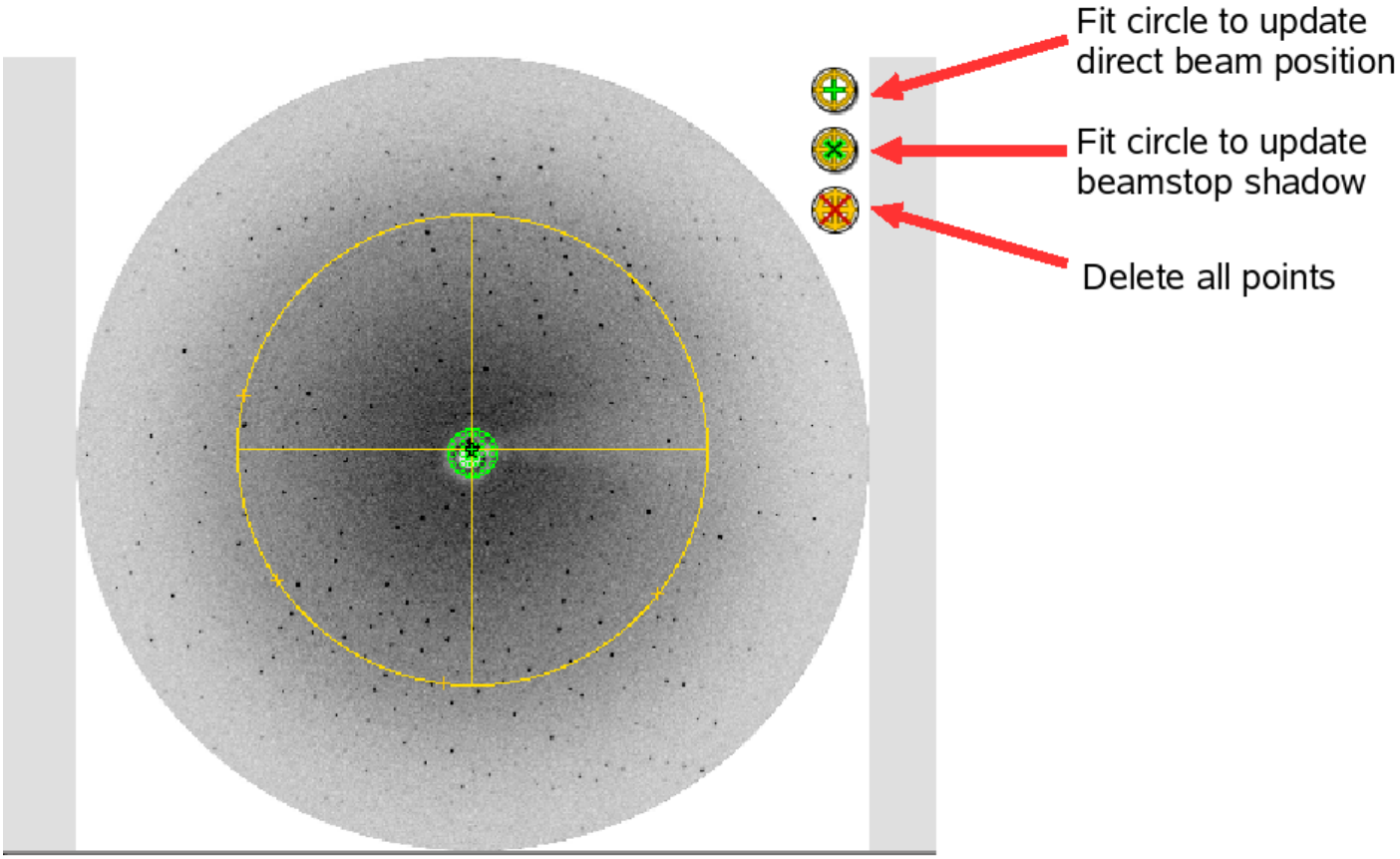
Select the Pan tool and pan the displayed area by holding down LMB and moving the mouse. This is very rapid on a local machine, but may be slow if run on a remote machine over a network.

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.



5. Spot finding, indexing and mosaicity estimation

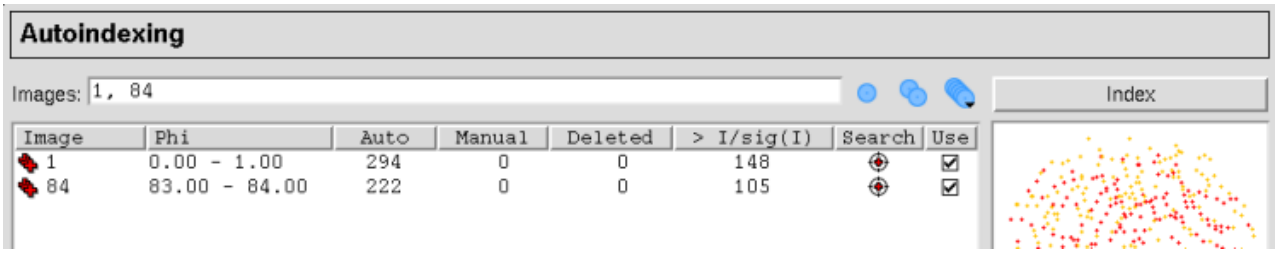
When images have been added, the "Indexing" operation becomes accessible (it is no longer greyed-out).

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 degrees apart in phi (or as close to 90 as possible) will be selected and a spot search carried out on both images (This is controlled by the "Automatically find spots on entering Autoindexing" button in the Spot finding tab of Processing options).

Found spots will be displayed as crosses in the Image Display window (red for those above the intensity threshold, yellow for those below). 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:

1. Simply type in the numbers of the images (e.g. 1, 84 above).
2. Use the "Pick first image" icon (single blue circle).
3. Use the "Pick two images ~90 apart" icon (two blue circles) . This is the default behaviour.
4. Use the "Select images ..." icon (multiple circles). If selected, all images in the sector are displayed in a drop-down list. Click on a image to select it, then double-click on the search icon (resembling a target) 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 "Settings" drop-down window and selecting "Processing options". The resulting new window contains six tabs relating to Spot finding, Indexing, Processing, Advanced refinement, Advanced integration and Sort, Scale and Merge.

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:

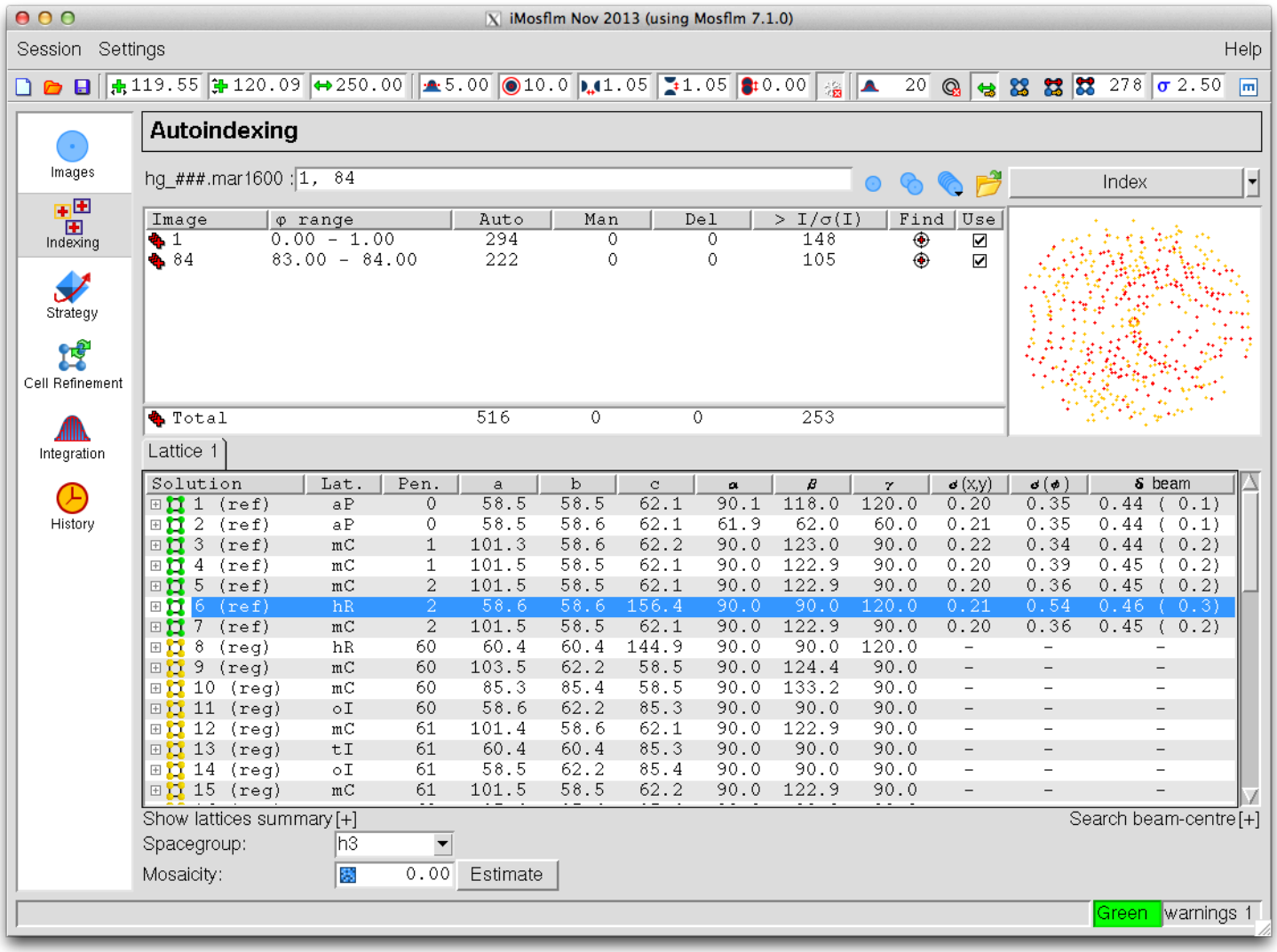
1. Minimum spot separation. This should be the size (in mm) of an average spot (not a very strong spot). Change to values estimated by manual inspection of spots if there are difficulties due to badly split spots. This separation parameter is very important when spots are very close, but usually the program will determine a suitable value.
2. Minimum pixels per spot. Default value 6, but this will be reduced automatically if spots are very small (eg in situ screening and microfocus beam).
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 parameters are automatically changed, but some improvement may still be obtained by the following:

1. Reduce the spot finding threshold (default 3.5), e.g. to 2.
2. Increase the minimum number of pixels per spot (default 30) to 20 or 40.
3. Set the minimum spot separation to a sensible value, e.g. 1.5mm

5.2 Indexing

Providing there are no errors during spot finding, indexing will be carried out automatically after spot finding (This is controlled by the "Automatically index after spot finding" button in the Indexing tab of Processing options). If the image selection or indexing parameters are changed, the "Index" button must be used to carry out the indexing.



Autoindexing will be carried out by MOSFLM using spots (above the threshold) from the selected images. The threshold is set by MOSFLM but can be changed using the entry box 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 50, 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 deviation (rmsd) or error in predicted spots positions (σ(x,y) in mm) and the rms error in (σ(φ) 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 spot positions for the highlighted solution will be shown on the image display with the following default 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.

These colours may be adjusted via the "Settings - Prediction colour" item in the Image display window.

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

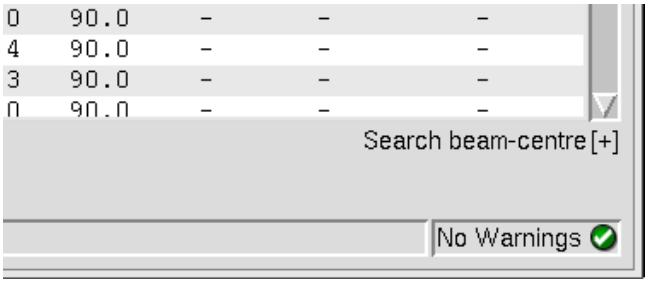
Select other solutions with a higher penalty and see how well the predicted patterns match the diffraction image.

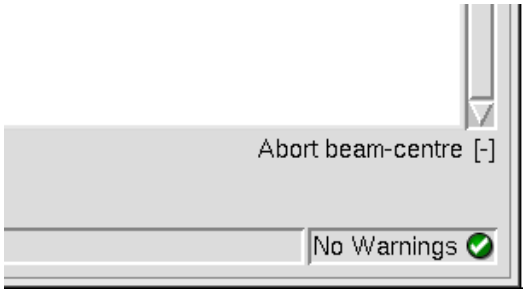
The rmsd and visual inspection of the predicted pattern are the best ways of checking if a solution is correct. If the agreement is not good, then the autoindexing has probably failed.

5.2.1 If the indexing fails - Direct beam search

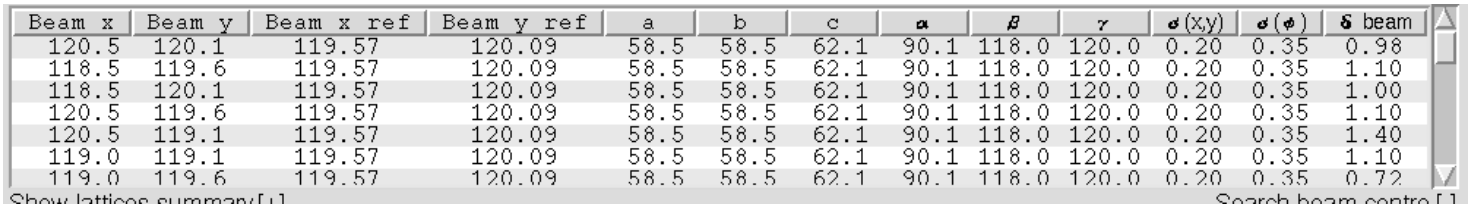
The indexing is very sensitive to errors in the direct beam coordinates. For a correct indexing solution, these should be correct to better than half the minimum spot separation. Check, for example, that the current direct beam position is behind the backstop. If there are any ice rings, these can be used to determine the direct beam position (see Circle fitting §4.1.6).

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 (see Settings - Processing options §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 text "Search beam-centre" in the Indexing pane. While in progress, it can be stopped by clicking on the same place, the text of which is now "Abort beam-centre".





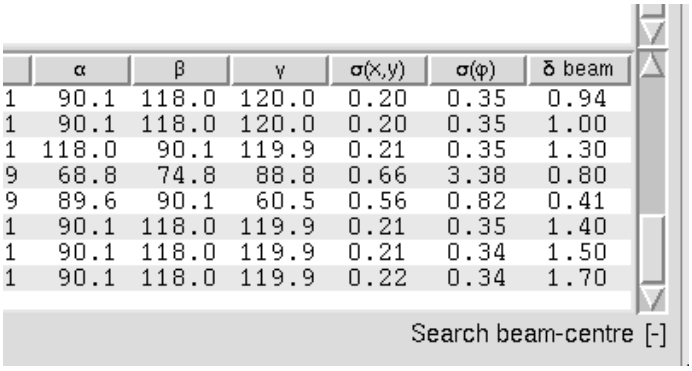
The indexing will be carried out for each set of starting coordinates (Beam x and Beam y in the table which appears) and, if a solution is found, the refined beam coordinates (Beam x ref, Beam y ref), the unit cell parameters of the triclinic solution and the rmsd error in spot coordinates and in phi will be listed. The correct solution will generally be the one with the smallest rms error in spot positions (σ(x,y)). To make it simpler to identify the correct solution, once the grid search has finished the solutions will be sorted on σ(x,y), so the correct solution will be at the top of the list. When sorting the solutions, those which have smaller cell parameters will be listed first, as it is possible to have a smallσ(x,y) value for an incorrect solution with a much larger unit cell.



Beam x	Beam y	Beam x ref	Beam y ref	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
120.5	120.1	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.98
118.5	119.6	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	1.10
118.5	120.1	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	1.00
120.5	119.6	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	1.10
120.5	119.1	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	1.40
119.0	119.1	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	1.10
119.0	119.6	119.57	120.09	58.5	58.5	62.1	90.1	118.0	120.0	0.20	0.35	0.72

Show lattices summary[+] Search beam-centre [-]

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 "Search beam-centre" and recovered by clicking on [+]



	α	β	γ	σ(x,y)	σ(φ)	δ beam
1	90.1	118.0	120.0	0.20	0.35	0.94
1	90.1	118.0	120.0	0.20	0.35	1.00
1	118.0	90.1	119.9	0.21	0.35	1.30
9	68.8	74.8	88.8	0.66	3.38	0.80
9	89.6	90.1	60.5	0.56	0.82	0.41
1	90.1	118.0	119.9	0.21	0.35	1.40
1	90.1	118.0	119.9	0.21	0.34	1.50
1	90.1	118.0	119.9	0.22	0.34	1.70

Search beam-centre [-]

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 entry fields and buttons that appear in the Indexing toolbar.

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 Difficult images §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. Select images that show the clearest lunes and/or have the best spot shape.
3. In case the crystal orientation has changed, or if you are unsure of the direction of positive rotation in phi, try indexing using only one image.
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, but use with caution as often this needs to be ~twice the largest cell for successful indexing.
5. If the detector distance is uncertain, and the images are high resolution (e.g. 2Å), allow the detector distance to refine during cell refinement. Use with caution.

5.3 Indexing multiple lattices simultaneously

This is a new feature and is still undergoing development. It is most likely to succeed when the different lattices are clearly separated, at least at higher resolution. If the images shows badly split spots with several components (each arising from a sub-crystal with a slightly different orientation) then indexing may well still be impossible.

After spot finding, examine the image to ensure that spots that are close to each other but belong to different lattices have been identified as separate spots, and not treated as a single split spot (in which case the red cross will lie midway between the two spots). Adjusting the minimum spot separation (Settings -> Processing options -> Spot finding) can help in resolving close spots. It may also help to increase the inner radius of the spot search region (see Spot search area §4.1.3) so that low resolution reflections where the spots are not really separated are excluded from the search.

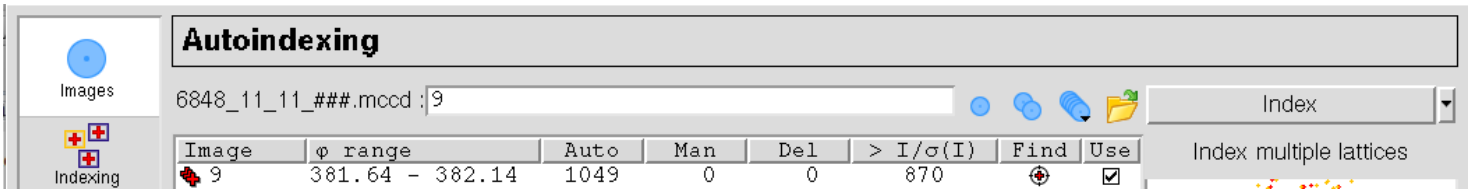
The following procedure usually provides the best chance of success:

1. Select appropriate images for use in indexing. The separation of the lattices can be clearer on some images than others. Ideally choose several images widely separated in phi.

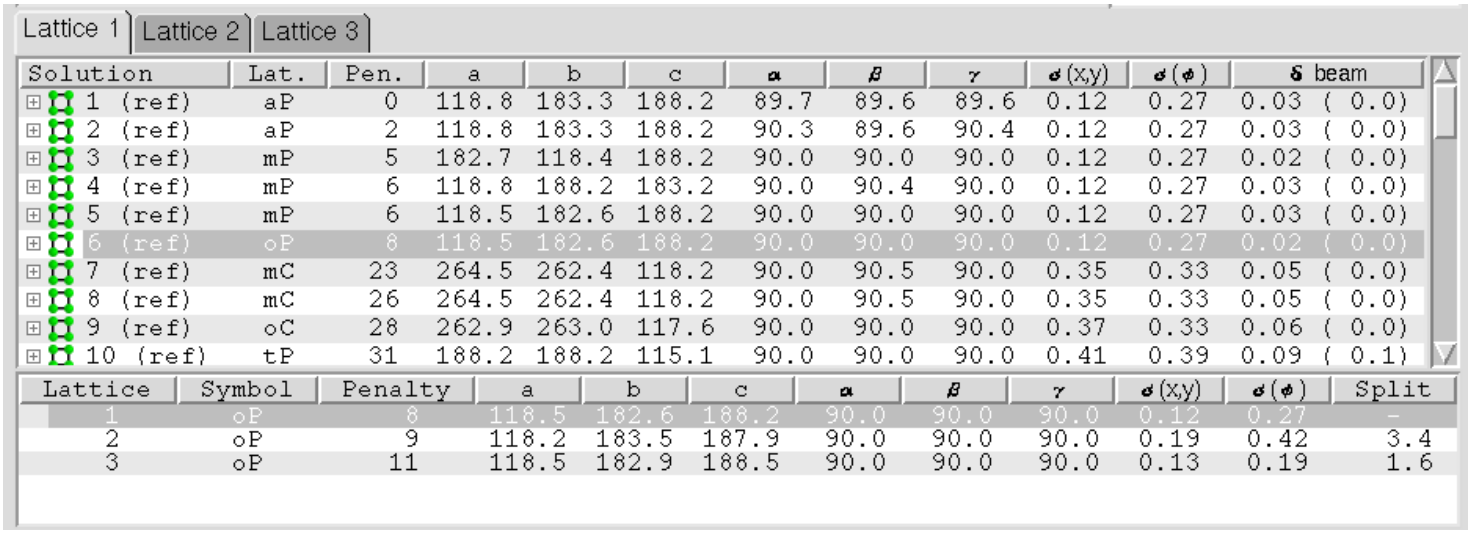
2. Do a normal (single lattice) indexing first. This will usually work, but will give a high positional rmsd because of the presence of more than one lattice. This step is important because it will result in more accurate direct beam coordinates, which are crucial for

multi-lattice indexing.

3. Select “Index multiple lattices” from the drop-down option next to the “Index” button.



The indexing results for different lattices will be presented in different tabs, while the selected solution for each lattice will be shown in the lower "Summary" window. Currently a maximum of 5 lattices can be handled. The figure below shows a case where 3 different lattices were found.

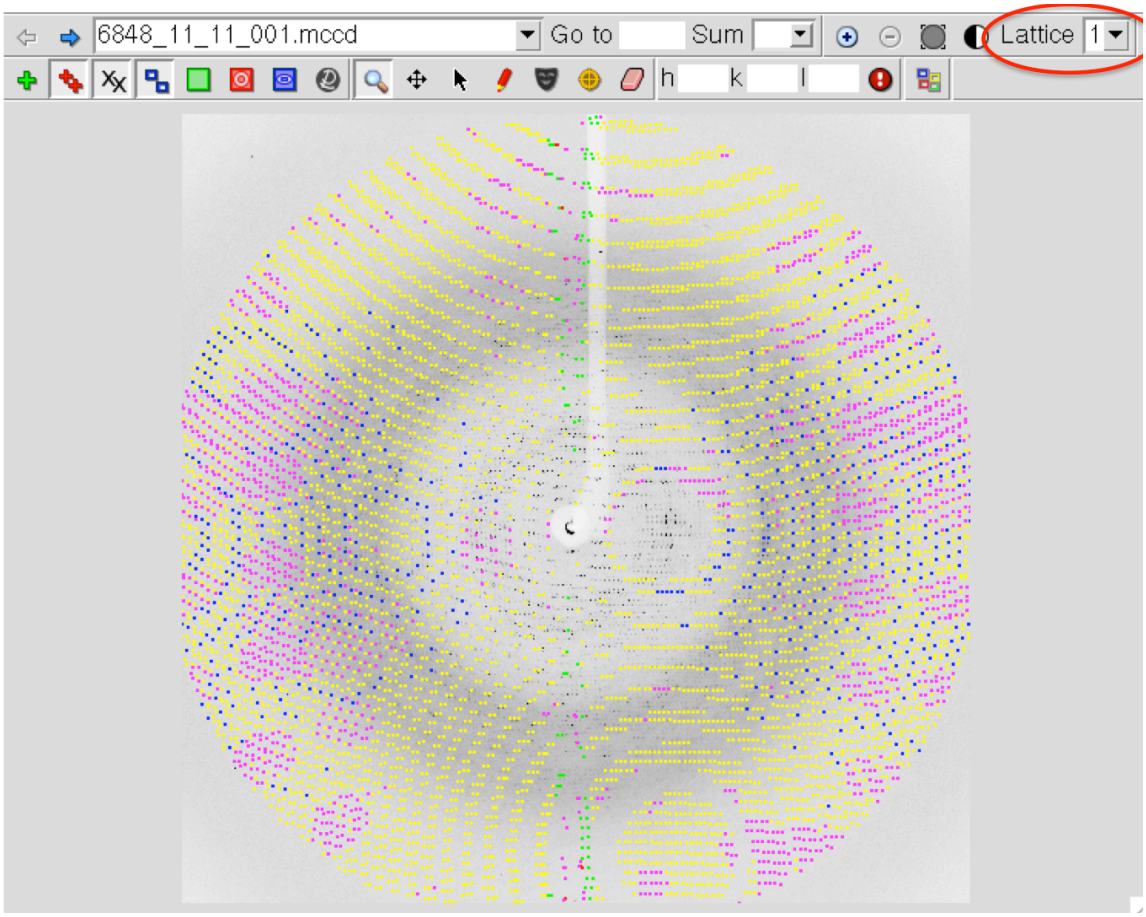


4. Judge the success of indexing by examining the lattice type and the cell parameters obtained for the different lattices (the cells should be very similar) and checking that the predictions match the observed spots. The rmsd values for correct lattices should also be significantly smaller than the value obtained in the initial (single lattice) indexing. It may be necessary to manually select a different autoindexing solution for some lattices in order to get consistent results for the different lattices.

The "Split" column in the lower pane indicates the angle between the reference lattice (with value "-") and the other lattices. This is only calculated when the lattices have the same lattice type. In this case, the angles between lattice 1 and lattices 2 & 3 are 3.4 and 1.6 degrees.

If a lattice is clearly incorrect, it can be deleted by highlighting that lattice (in the lower pane that summarises the results) and using the right mouse button to select “delete”.

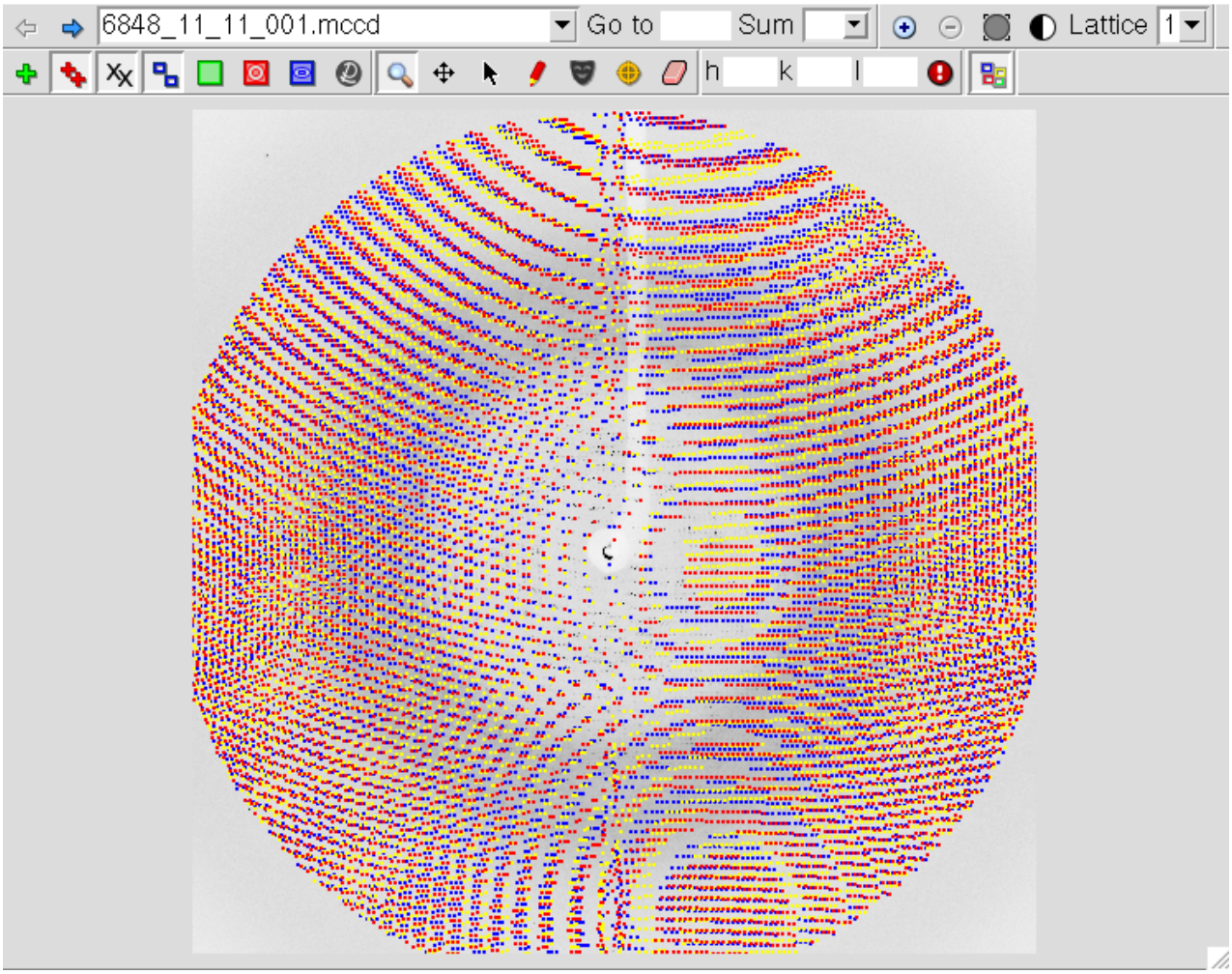
The predictions for different lattices can be displayed either by highlighting the lattice in the Summary window or by using the lattice selector in the image display pane.



In the image display, spots in different lattices that overlap with each other are shown in magenta. Only the summed intensity will be measured for these reflections.

The predictions for all lattices can be shown by selecting the “merge lattices” icon in the image display.

This can be useful to check if all the lattices have been successfully indexed, because all spots in the image should then be predicted. In the display, predictions from lattice 1,2,3,4,5 are shown in blue, yellow, red, green and magenta with no distinction made between fully and partially recorded reflections or spatial overlaps.



5.3.1 Practical Tips

Beware of solutions where the lattices only vary in orientation by a small angle (less than 1 degree) as these are often different solutions for the same lattice. The difference in orientation between lattices is given by the "split" column in the Summary window, but this is only calculated if the lattices have the same lattice type. Differences are all relative to one lattice, which will display a split angle as "-". To calculate the angles relative to a different lattice, open the tab for that lattice, select a different solution to that currently selected and then select the original solution – this will trigger recalculation of the split angles relative to that lattice.

Other parameters that can influence the indexing are the "Maximum deviation from integral hkl" (default 0.2 when indexing multiple lattices, otherwise 0.3) and "Number of vectors to find for indexing" (default 30). These can be changed via the Indexing tab of the "Processing Options" that can be selected from the "Settings" menu item. Values between 0.15 and 0.25 might be helpful for the first parameter and increasing the number of vectors from 30 to 50 can also help. At present there is no defined protocol for changing the default parameter values for particular cases, it has to be done by trial and error.

Experience so far has shown that selecting appropriate images and adjusting the maximum cell length used in indexing have the biggest effect. The optimum value for the maximum cell length will normally lie between one and three times the largest cell dimension, but it is worth varying this in steps of between 25 and 50 Å, as the indexing can be very sensitive to this parameter.

For a more complete discussion of multiple lattice indexing see: Powell, H.R., Johnson, O. & Leslie, A.G.W. 2013. Autoindexing diffraction images with iMosflm. Acta Cryst. D69, 1195-1203.

5.4 Space group selection

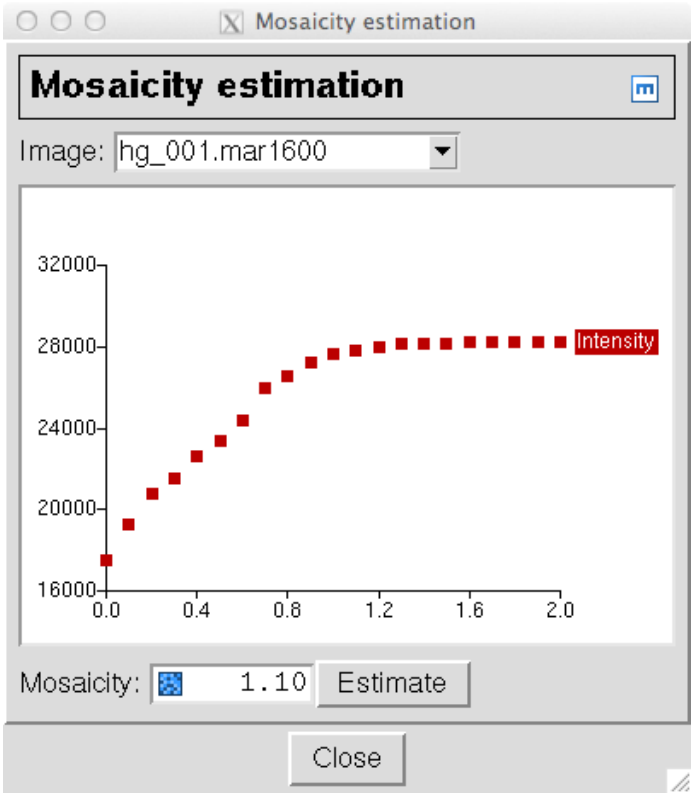
Note that the indexing is based solely on information about the unit cell parameters. It will therefore be very difficult (or impossible) to determine the correct Laue group in the presence of pseudosymmetry. For example, a monoclinic space group with $\beta \sim 90$ will appear to be orthorhombic; an 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 (i.e. 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 for the strategy calculation 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.

5.5 Mosaicity estimation



The mosaicity (mosaic spread) 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 predictions displayed on the image will be updated. Entering different values of mosaic spread in the box (followed by return) will allow a visual estimate of the effect of changing the mosaic spread.

The mosaic spread can also be entered in the Images pane by clicking once on the Mosaicity line and once again on the value displayed.

5.5.1 The mosaic block size

In some cases, the apparent mosaic spread is much larger 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) from the default of 100 microns will result in predicting more spots at low resolution but have virtually no effect at high resolution.

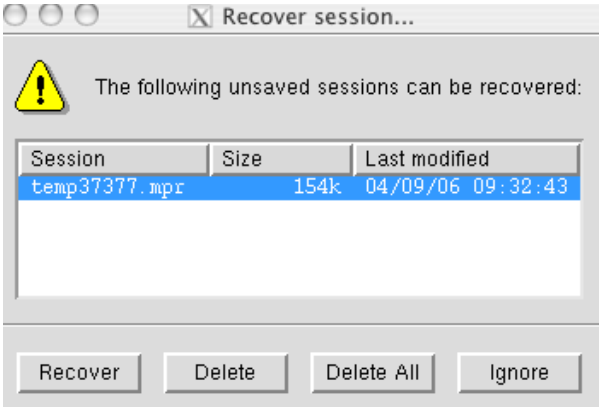
6. Saving and recovering a session

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.

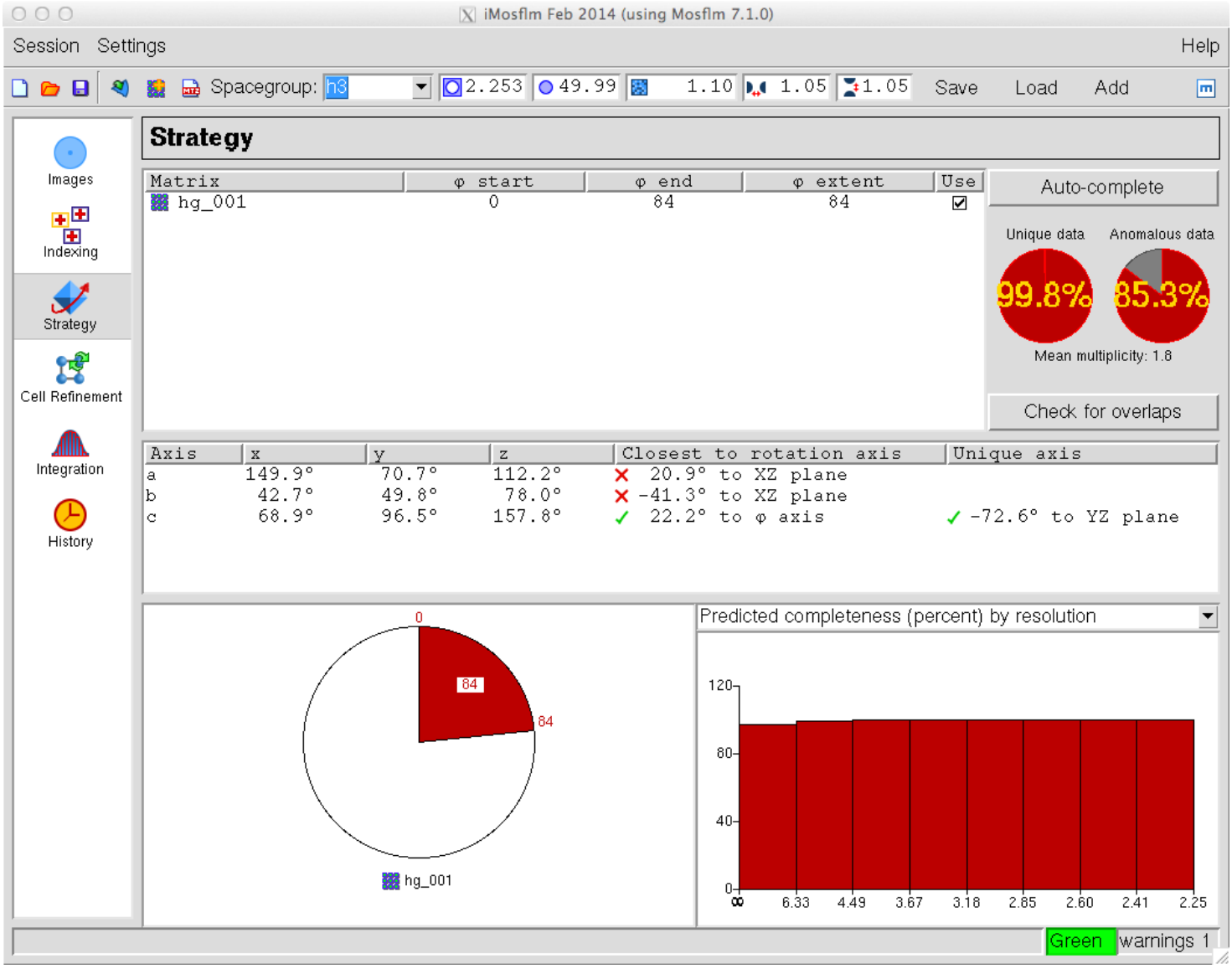
Selecting the Recover button will restore the session as far as possible. This file is written to the directory ".mosflm" in the user's home directory.



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.



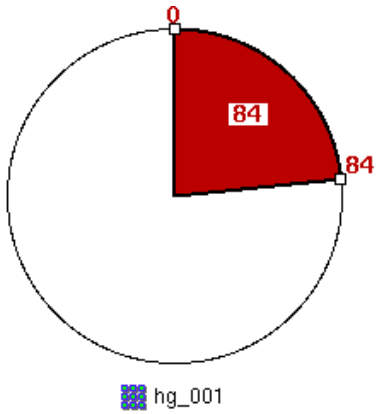
The statistics initially presented in the window are based on the assumption that all images between the smallest phi value and the largest phi value in the current sector(s) have been collected (the phi ranges are listed). If processing a dataset that has already been collected (as in this tutorial)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

More usually, the Strategy option will be used after two initial images, 90° apart in ϕ, have been collected. In such cases a complete, shaded segment will not be shown in the lower part of the screen, rather an empty circle will be shown with the matrix label beneath it. In cases where two reference images have been collected and indexed, click the 'Auto-complete' button and, in the pop-up menu that appears, select a rotation angle or "sweep" of data to collect - or leave the "Rotation:" setting as "Auto" for a calculation of maximum completeness for this crystal orientation. Click on Ok and the completeness results will be presented.

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

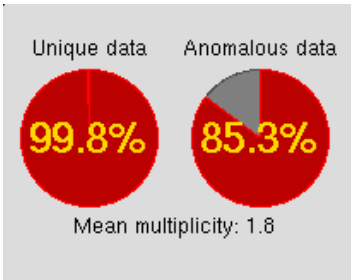
The completeness of any given segment of data can be determined manually using the circle at the bottom left of the window. *Click on the red segment shown in this circle*.



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.

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.

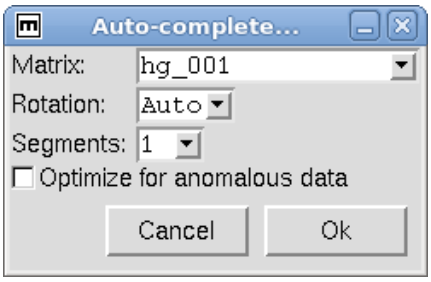


More detailed statistics (such as completeness as a function of resolution) are given in a series of histograms that can be selected by a drop-down menu.

7.2 Calculating a strategy automatically

Uncheck *the 'Use' checkbox at the right side of the hg_001 line listed in the sector list near the top of the Strategy pane. This means that the 84 images used thus far will not be included in the completeness calculation; only the matrix derived from indexing them will be used. Select the Auto-complete button to calculate a strategy automatically.*

A strategy calculation data pop-up 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 (the Auto setting). However, it is sometimes possible to achieve a high completeness with a significantly smaller total rotation (e.g. 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 5° and 90° can be selected from the drop-down list. The number of segments to be used can be set between 1 and 3.

If the check box is checked, data completeness will be calculated to optimize the proportion of anomalous data collected.

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

Choose the default values (Rotation: Auto, Segments: 1) 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.

****** SKIP OVER SECTION 7.3 IF DOING THE NORMAL TUTORIAL ******

7.3 Calculating a strategy using multiple crystals

The Strategy pane may also be used when, due to radiation damage, many crystals are required to collect a complete dataset .

IMPORTANT: This procedure will only work correctly for space groups where there is no indexing ambiguity. For those space groups where there is an indexing ambiguity(....) it is necessary to integrate images to determine whether different crystals have been indexed in the same way. This is not achievable within the Strategy option at present.

Suppose the first crystal has been indexed from the reference images.

- Enter the Strategy pane, an empty circle is displayed in the bottom part of the screen, and click on the 'Auto-complete' button near the top-right of the pane. In the Auto-complete... window that appears choose a suitable 'Rotation:' angle for the amount of data that can be collected from one crystal, say 20 degrees for this example. Leave 'Segments:' set to 1 and do not check 'Optimize for anomalous data'.

- Click the Ok button and the completeness of the data will be displayed. The matrix name, start and end phi values and the phi extent will be displayed in the top area of the pane with a checkbox ticked under the 'Use' column. This means this segment will be used in further strategy calculations.

- Accept and save this result by clicking on the yellow, hatched segment displayed in the large circle near the bottom of the pane. Click once, and the segment will be outlined in black (you can adjust the start or end phi value calculated in one degree increments by click+dragging the respective black square), click again, this time on the black square denoting the start or end phi (or release if you have adjusted by click+dragging) and the segment will turn red. *The segment will not be saved unless it has turned red !*

- This result will be saved if you leave this pane (e.g. to load more images) and return later in this session. The strategy file can also be saved to a permanent file using the 'Save' button on the tool-bar. *This is highly recommended when working with many crystals*.

Return to the Strategy pane when you have loaded and indexed the reference image(s) from the second crystal and repeat the procedure above. You will see the calculated segment from the first matrix displayed at the bottom-left of the screen. The optimum phi range for the second crystal will be displayed. Save this in the same way as for the first crystal (clicking twice). Repeat this procedure for any subsequent crystals, and in each case the optimum phi range will be calculated for the latest crystal taking into account the data from previous crystals.

It is a good idea to save the strategy to a file as each new crystal is added, in case the program crashes and this information is lost.

In practice, it is probably safest to have a separate iMosflm session that is reserved for the strategy calculations, and not used to process any data. The strategy for each new crystal can be calculated as described above. If there is a problem and the session is corrupted, or if iMosflm is closed down for any reason, it is straightforward to restore the status of the session providing the strategy has been saved. The following steps should be taken:

1. Read in the reference images for the latest crystal and index them.

2. Select the Strategy pane.

3. Read in the optimised strategy for all previous crystals using the “Add” button and giving the appropriate filename.

4. Making sure that the circle for the latest crystal is highlighted in the lower left window, click on ‘Auto-complete’ and follow the steps described above.

Advanced Usage

The start/end phi values for each crystal/matrix can be updated. This might be required if, for example, only 10 degrees of useful data were collected from a particular crystal instead of the anticipated 20 degrees. To change the phi values, click either on the matrix name in the top pane or on the circle showing the segment of data in the lower left pane (these are highlighted in a blue box as the mouse moves over them). The corresponding segment will now be displayed in the larger circle. Click on the red segment so that a black square appears at the start/end phi values. Then simply drag the appropriate phi to the desired value. The completeness will automatically be updated (but the strategy calculation for the latest crystal will not be repeated, to do this the current segment must be deleted by highlighting the matrix name and using right click and then repeating the Auto-complete step).

If matrix files, saved from iMosflm, are available from crystals for which data have been collected in previous sessions, these can be added into the calculation using the 'Add matrix' button on the tool-bar.

If the saved strategy result has become complicated or is incorrect and you wish to restart the calculations afresh, there is a tool-bar button, 'Reload sectors from session'. This will remove any calculated sectors while leaving any known matrices listed at the bottom-

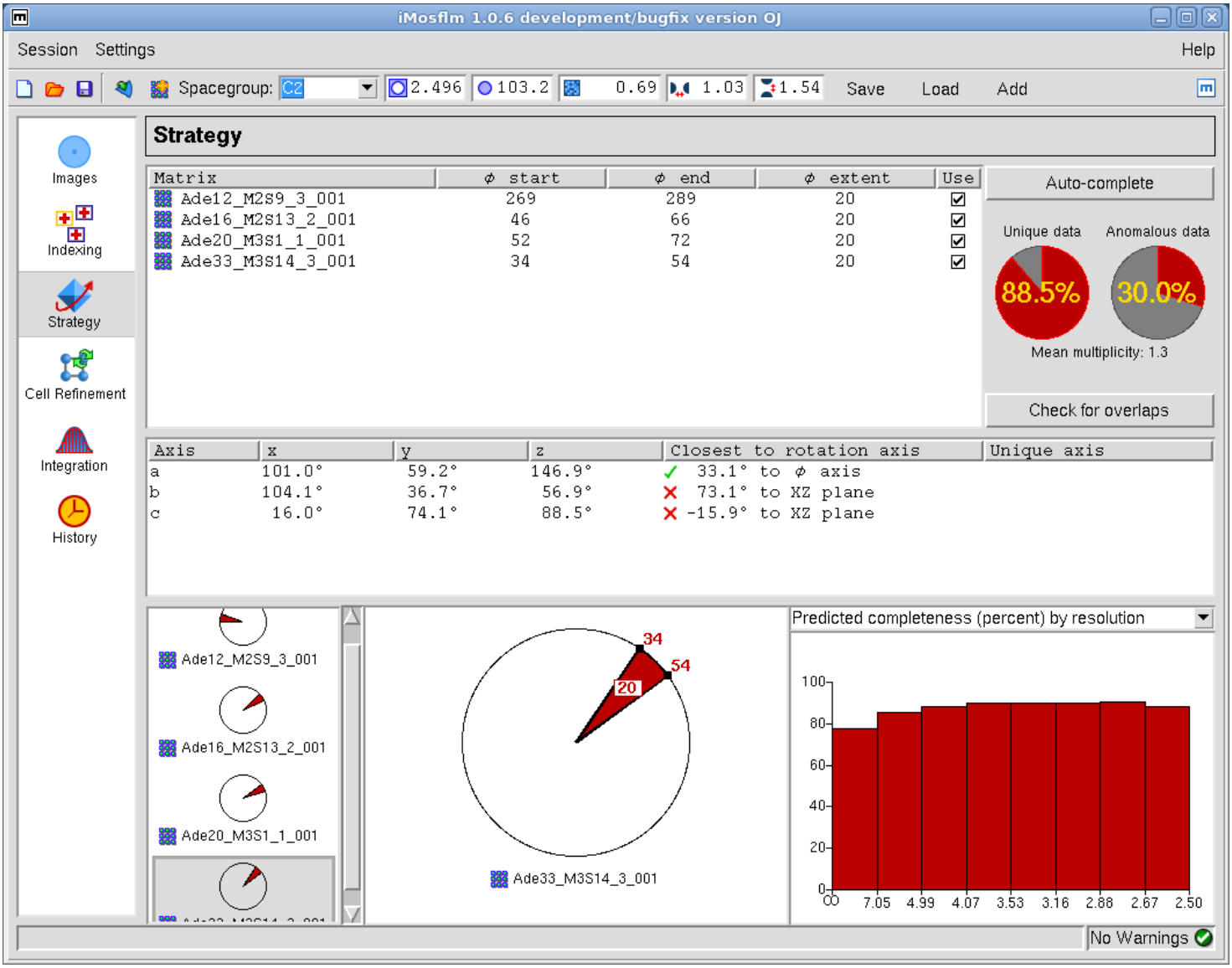
left of the Strategy pane.

If you have saved a strategy file from a previous session, this can be used to re-initialize the Strategy pane using the 'Load' button on the tool-bar.

A 'Spacegroup:' pull-down menu has been added to the tool-bar to permit different choices of space group to be used in the calculations (e.g. h3 and h32). This menu is populated from the list offered on the Indexing pane. However, the space group symbol text is editable. Any space group symbol entered will be validated by Mosflm together with the current cell parameters.

As a calculator for a multi-segment strategy, the Strategy pane is reasonably flexible; you can Save or Delete highlighted matrix lines by right-clicking. As you highlight different matrix lines 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 re-displayed in the centre of the pane.

The following screen shot shows the cumulative results in the Strategy pane after this procedure has been followed for four crystals (ignore the iMosflm version number).

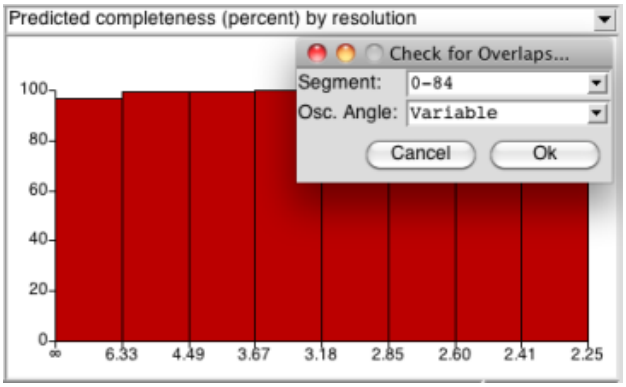


7.3.1 Calculating a strategy when some data have already been collected and processed

If data have already been processed (to give an MTZ file) but the images are not currently available (eg they were collected on a previous synchrotron trip, possibly at a different synchrotron), then the MTZ file containing this data can be read and used to calculate a strategy for data collection from new crystals.
The MTZ icon in the tool bar is used to read the MTZ file, **but this option is not yet implemented.**

7.4 Calculating a suitable oscillation angle

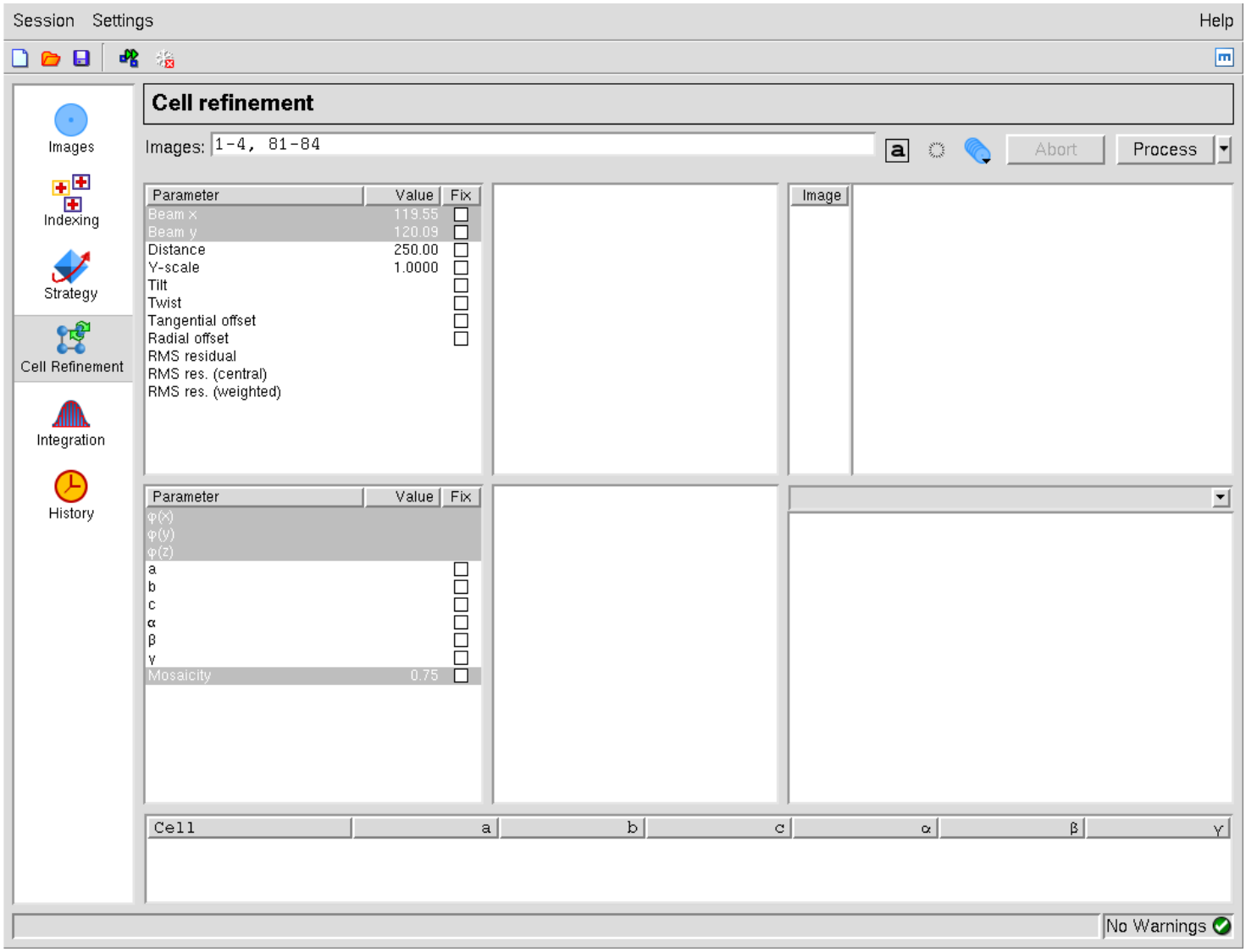
To calculate the maximum possible oscillation angle while avoiding spatial overlaps, click on the "Check for overlaps" button. A pop-up window will appear, from which the option to calculate the maximum oscillation angle (as a function of ϕ) or check how many overlaps will occur for different (user selected) 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

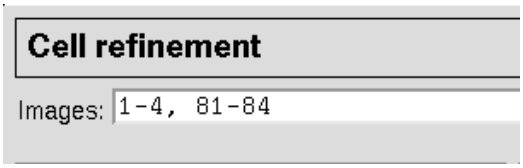
It is important to determine the cell parameters accurately before integrating the images. Although the unit cell is refined as part of the autoindexing, providing the diffraction extends beyond ~3.5Å resolution, it is possible to obtain more accurate cell parameters using a procedure known as post-refinement. 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 entry field. iMosflm will automatically select appropriate images, but these can be changed.



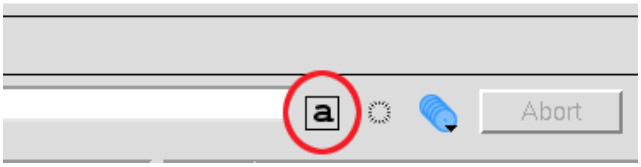
8.1.1 Manual selection

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 (e.g. ϕ = 0,45,90,135).

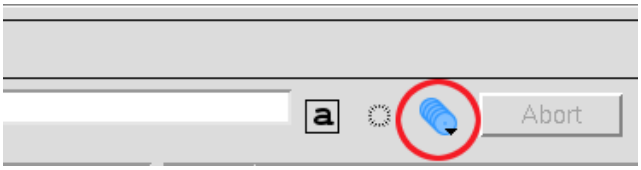
The number of images that should be included in each segment depends on the mosaic spread and oscillation angle. A reasonable number is 2*(mosaic spread/oscillation angle +1). MOSFLM will automatically select two or more segments as appropriate. Note that there is a limit on the total number of images that can be used.

8.1.2 Automatic selection

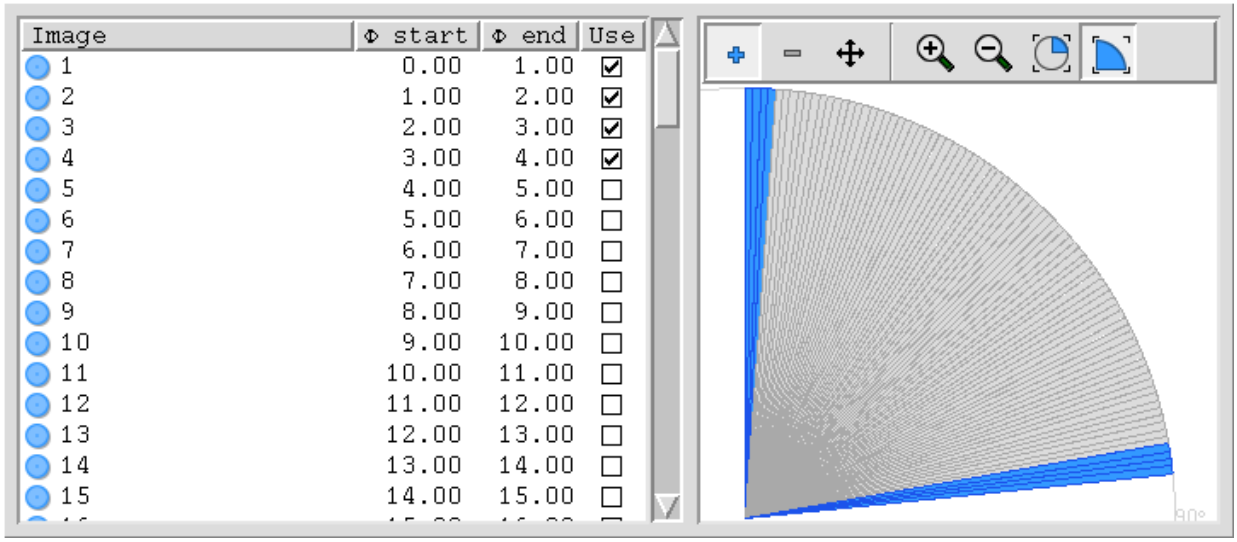


Selecting the "Automatically select images" icon will select a suitable set of images.

8.1.3 Graphical selection (No longer recommended)



Selecting this icon will give a new window:

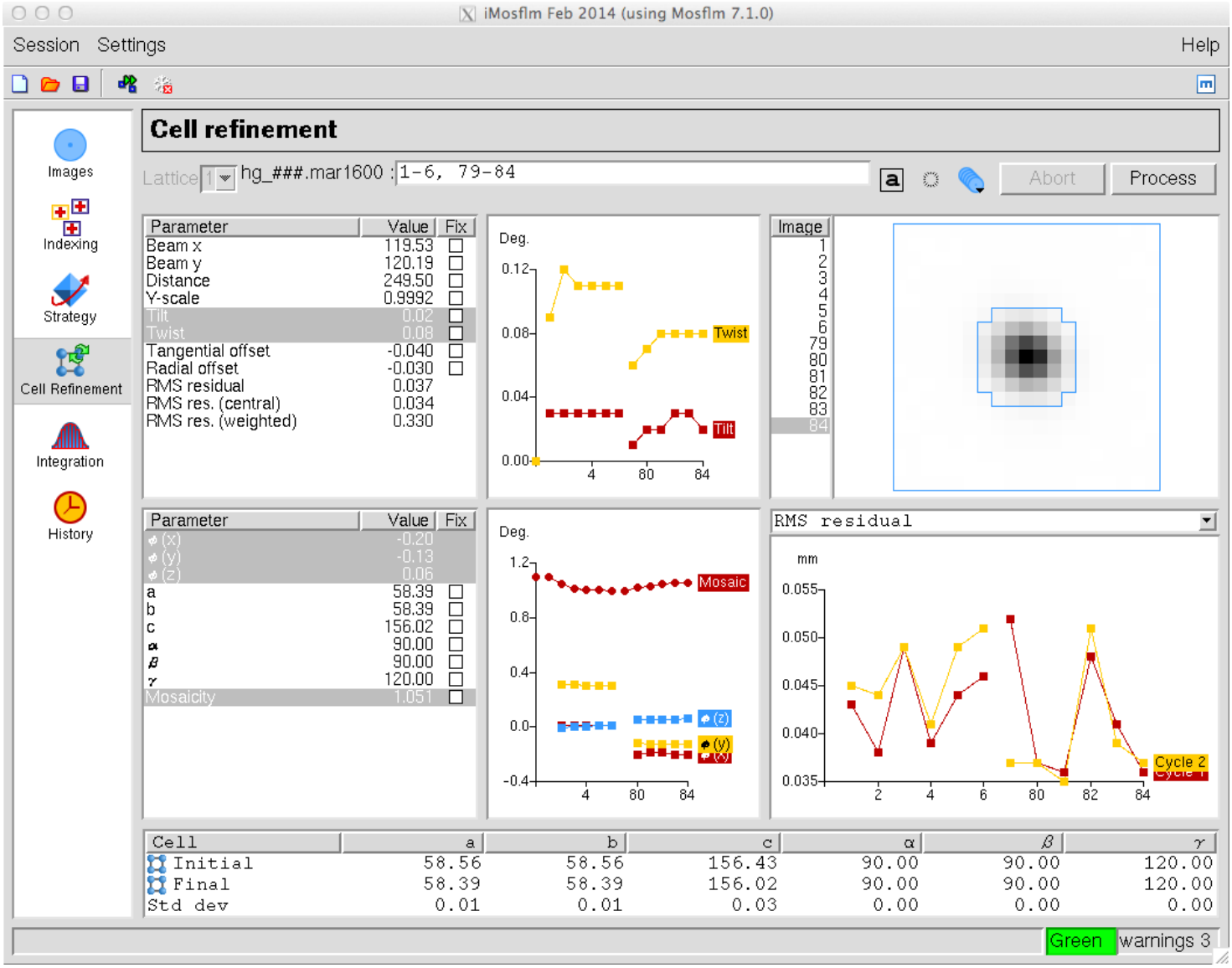


Images can be selected by clicking in the "Use" check-box in the list of images, or by using the mouse to drag a selection of images from the sector displayed. The other icons allow the sector to be zoomed in or out and the viewed area to be moved to make the selection easier. The final two icons on the tool-bar switch the segment display to fit the display space as either a full circle or a quadrant. This Image 'palette' can be closed by clicking outside its area but will close after 60 seconds anyway.

8.2 Integrating the images and refining the cell

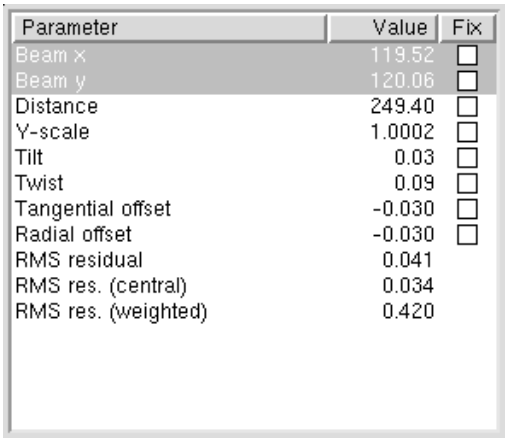
Use one of the methods above to select images to use in cell refinement and click on the Process button. Before starting cell refinement, check that the prediction for the first image ***in each segment*** *to be used is OK.* ***The majority of the spot positions must be correctly predicted for the first image in each segment, 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 in the bottom left-hand corner of the window in the status message area.



The different windows are explained below. Moving the mouse over any parameter or plotted graph point will give a full description of it.

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.

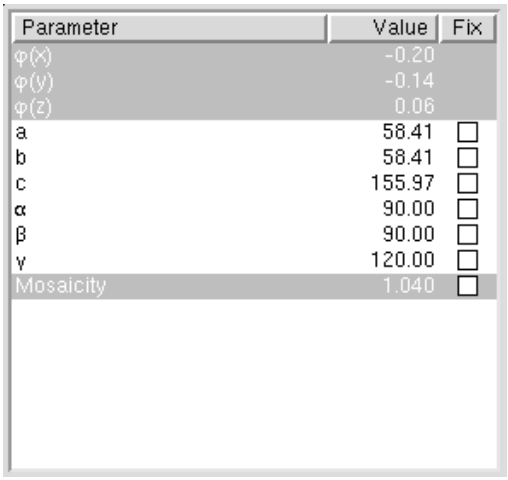
Any of these parameters can be selected for plotting on the graph that appears to the right of this box by simply clicking on that parameter (which is then highlighted in blue, as shown for Beam X and Beam Y).

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.

For weak images the tilt and twist may not be well defined, and might refine to large values (greater than one degree). In such cases these parameters should be fixed. It is not generally recommended to fix any other detector parameters.

8.2.2 The crystal parameters window

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

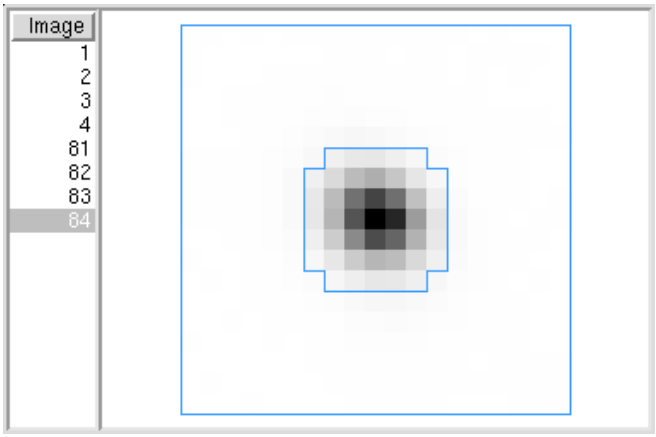


If more than one segment of data is being used for cell refinement, it is not unusual to see a change in orientation of the crystal between the two segments.

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 integration tab of Processing options. *Check the central spot profiles for the images used in the cell refinement.*

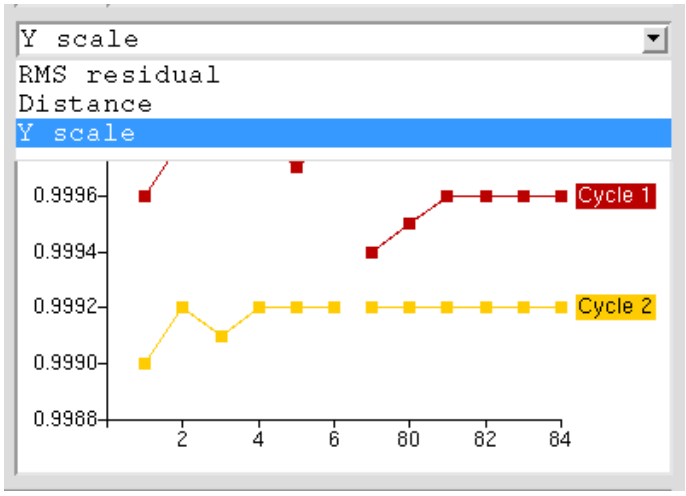


8.2.4 The summary window

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.

The other parameters that can be plotted from the pull-down list 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 auto-indexing 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 Raxis II 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 auto-indexing 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

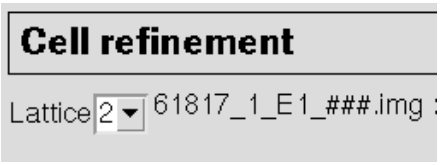
Cell	a	b	c	α	β	γ
Initial	58.56	58.56	156.43	90.00	90.00	120.00
Final	58.41	58.41	155.97	90.00	90.00	120.00
Std dev	0.01	0.01	0.05	0.00	0.00	0.00

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 ?

8.3 Refining cell parameters for multiple lattices

Providing the resolution is high enough (generally better than ~3.5-3.2Å) and not too many of the spots in the different lattices are overlapped, it is possible to refine the cell parameters for each lattice individually. The lattice to be refined is selected from the Lattice selector:



This selector will be grayed out when processing a single lattice.

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 at that stage.

It is good practice to start by integrating a block of about 10-20 images (a few degrees of data), 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 Warning messages §13).

Before integration, ensure that the backstop shadow has been defined (see Masked areas backstop §4.1.1 and Masked areas general §4.1.2)

9.1 Image selection

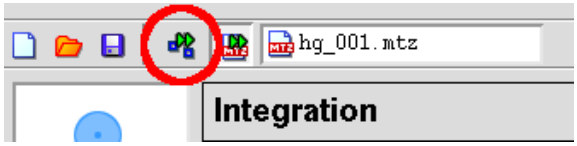
Image selection is performed in exactly the same way as in the cell refinement. In this case, the "Automatic" selection will simply include all images in the current sector.

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 (a number corresponding to a few degrees of data, 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 ~0.5-1 degree (more for high mosaicity) .

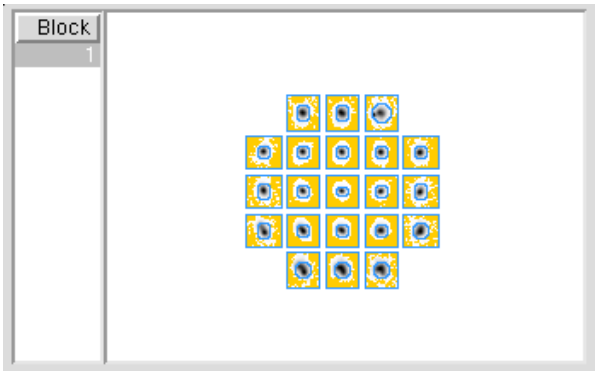
9.4 Parameter display windows

As described in Section 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.

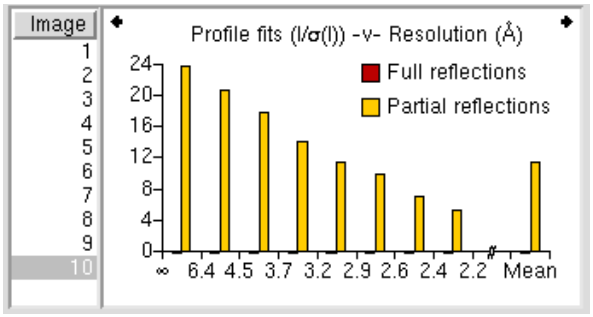
A display of the standard profiles for different regions of the detector is also provided. Poor profiles are "averaged", by including reflections from inner regions of the detector. Averaged profiles are indicated by a red line around the spot (rather than blue). The original "unaveraged" profile can 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 non-centred profiles may suggest that the prediction is not very good, in which case this should be checked on the corresponding image(s).

It may be useful to expand the standard profiles to fill the window using "Shift + 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 small right- or left-pointing arrows on the title line above the plot.

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 AIMLESS will use this information when deciding which images belong to the same dataset.

Select "Experiment settings" from the "Settings" drop-down menu (see Drop down menus §2.1) to enter these names and optionally give a title for the MTZ file.

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

9.5.1 Parallel Integration Using multiple cores

An option is available on the pull-down menu (next to the "Process" button in the Integration task) which splits up the integration job over available cores on the local machine. All aspects are currently determined automatically at present; user-configurable options will be available in a later version. The algorithm used is as follows:

1. iMosflm determines the number of cores available on the local machine.
2. If the number of cores (N) is greater than 2, iMosflm splits the integration run into N-1 batches (up to a maximum of 16) containing approximately the same number of images. If $N \leq 2$, iMosflm splits the run into N batches.
3. If the total number of images in each batch is less than 32, the number of batches is decreased until this condition is satisfied.
4. The first batch is submitted as a background task, consisting of a refinement stage and an integration stage, with its input and output files (including the MTZ file) in a new directory, its name being determined from the number of segments of data being integrated, and the lattice number (if the sample has been determined to contain multiple lattices). If the directories exist already, all the expected directories will be put into a new directory with a time-stamped name. Once the refinement stage is complete, the refined parameters are written to a file.
5. The launch of subsequent batches are delayed until the corresponding preceding batch has written its refined parameters. This file is read, and the new batch is launched, again with a refinement stage and integration stage.

At present, the job cannot be aborted once it has started.

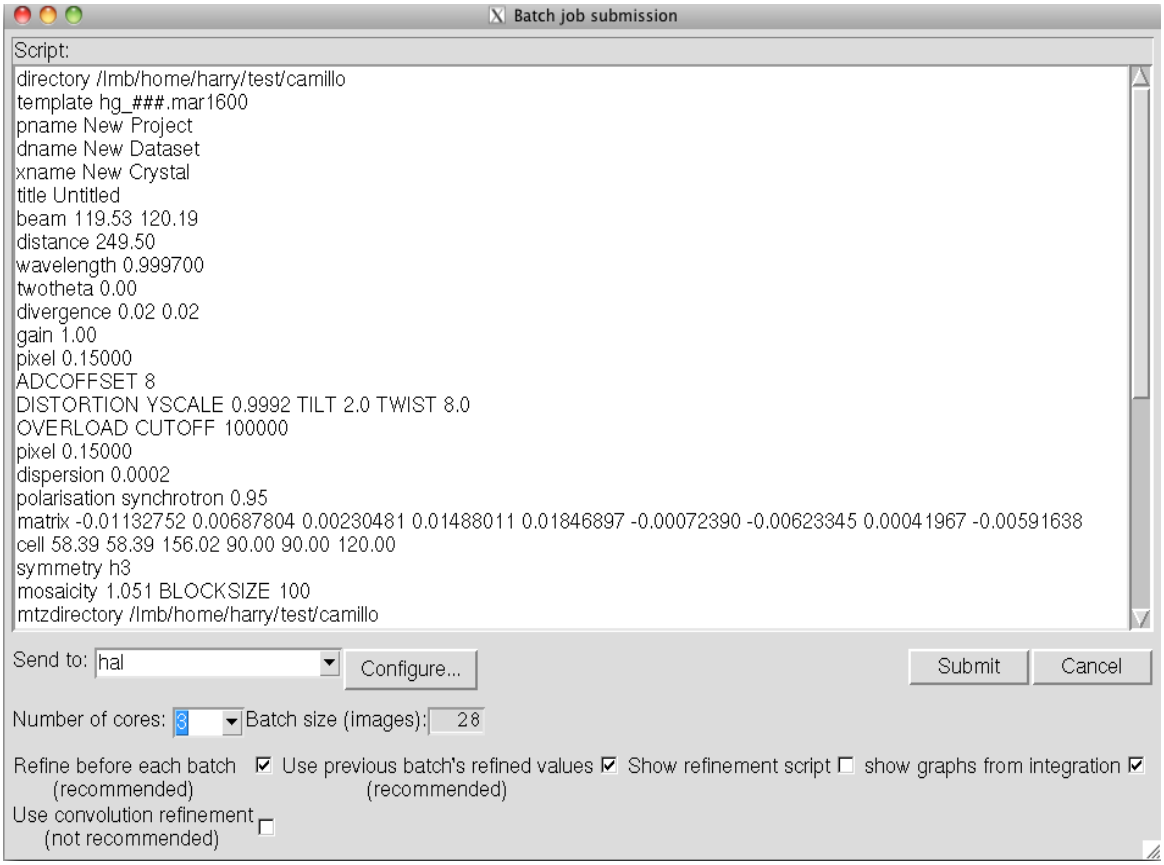
9.5.2 Integrating as a background job

By selecting "Batch" in the drop-down menu adjacent to 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 display a new window showing the script that will be used to process the images, along with a number of processing options. The script can be edited if additional keywords are required or if some parameters are to be changed, and can also be copied and pasted into a file for off-line processing.

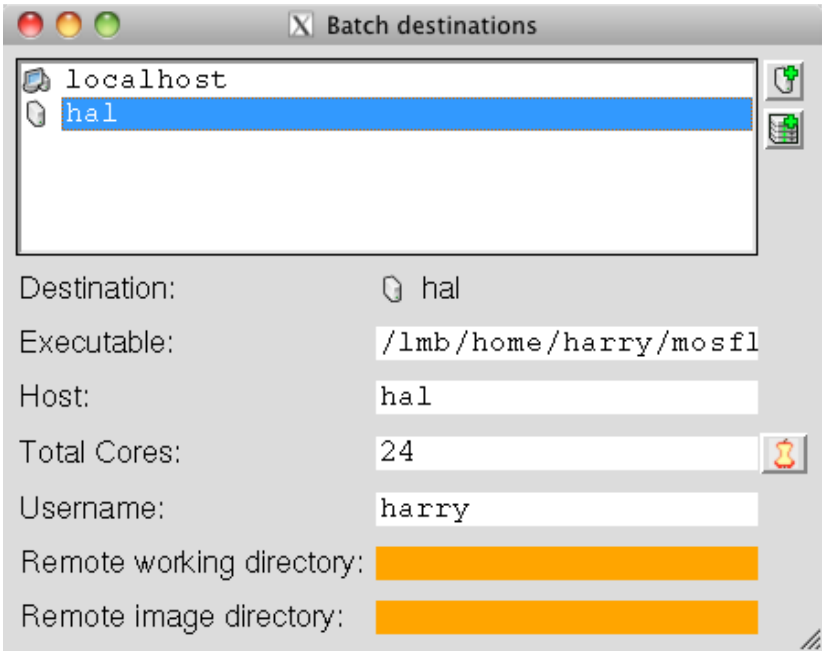
By default, the job is divided into batches across multiple cores as determined for the parallel integration described in §9.5.1, but the number of cores can be chosen by the user from a drop-down menu; in this case, the size of each batch is determined by the interface.

Plotting the results of the integration can be turned off; this can make processing much faster.



It is possible to configure job submission for remote machines (which can be added by depressing the "add host" button). Ensure you have enabled passwordless logins for all the remote machines you enter and that you specify the full path to the Mosflm executable in the "Configure..." window. The remote working directory (in which the log and MTZ files are written) and the remote image directory must both be specified with their full path names. Any required parameter fields that are not completed by the user are highlighted in orange; if the user tries to submit a job without completing these fields, a new window containing only those fields (again highlighted in orange) pops up with a request to do so.

The number of cores available on the selected host can be determined by depressing the "apple core" button on the "Batch destinations" configuration window.



The log file for batch submission jobs, 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 (e.g. 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 should be 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 integration tab of the Processing options menu (use Settings).

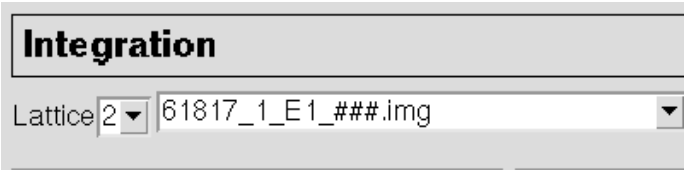
9.7 Advanced features for integration

The "Settings" - "Processing options" 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 integration tab, the measurement box parameters, profile tolerance and profile averaging parameters can be set.

Additional MOSFLM parameters will be added based on user requests.

9.8 Integrating multiple lattices

When multiple lattices have been successfully indexed, each lattice can be integrated individually by selecting the appropriate lattice number:

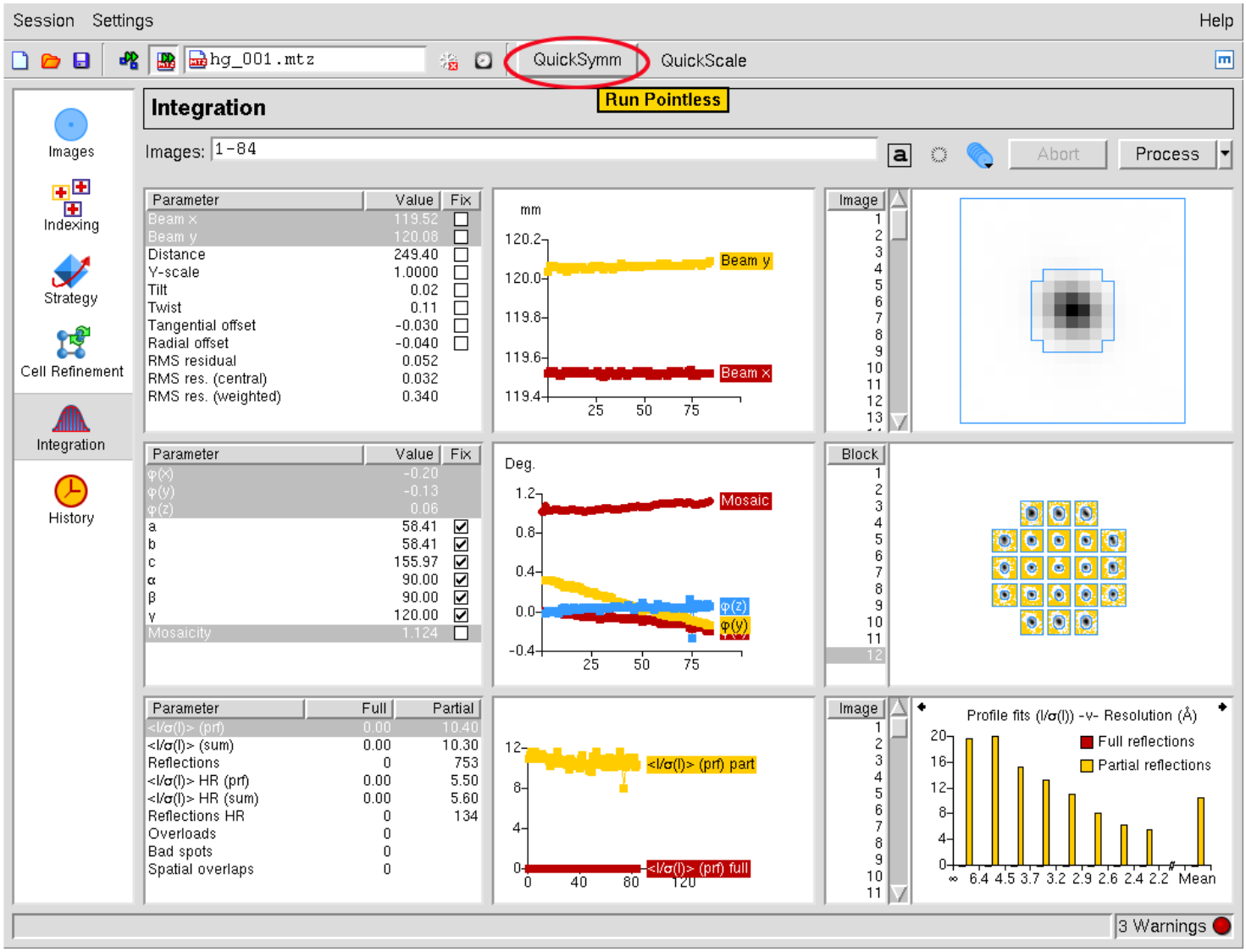


Reflections in the chosen lattice that are overlapped by a reflection from another lattice will be integrated, but the measurement box will be expanded so that the total intensity of the two (or more) reflections is measured. These reflections will be flagged for special treatment. Reflections that are not overlapped are referred to as "singletons". They will be dealt with in the normal way by POINTLESS and AIMLESS. It is often the case that by measuring both lattices (in separate integration runs), and inputting both MTZ files to POINTLESS, the overall completeness is high using ONLY the singletons. To make use of the overlapped reflections, both MTZ files must be input to the program FECKLESS, which will derive the best estimate of the summed intensity for overlapping reflections. The resulting MTZ file can then be input to POINTLESS and AIMLESS. At this time, no downstream programs can make use of the summed intensities of overlapping reflections, but in the future it should be possible to use this data in REFMAC.
At the present time, the best course is to use only singletons from all the lattices; FECKLESS is not required in this case.

The MTZ filename is automatically updated to reflect which lattice has been chosen, to avoid overwriting the MTZ file of the first lattice when integrating the second one.

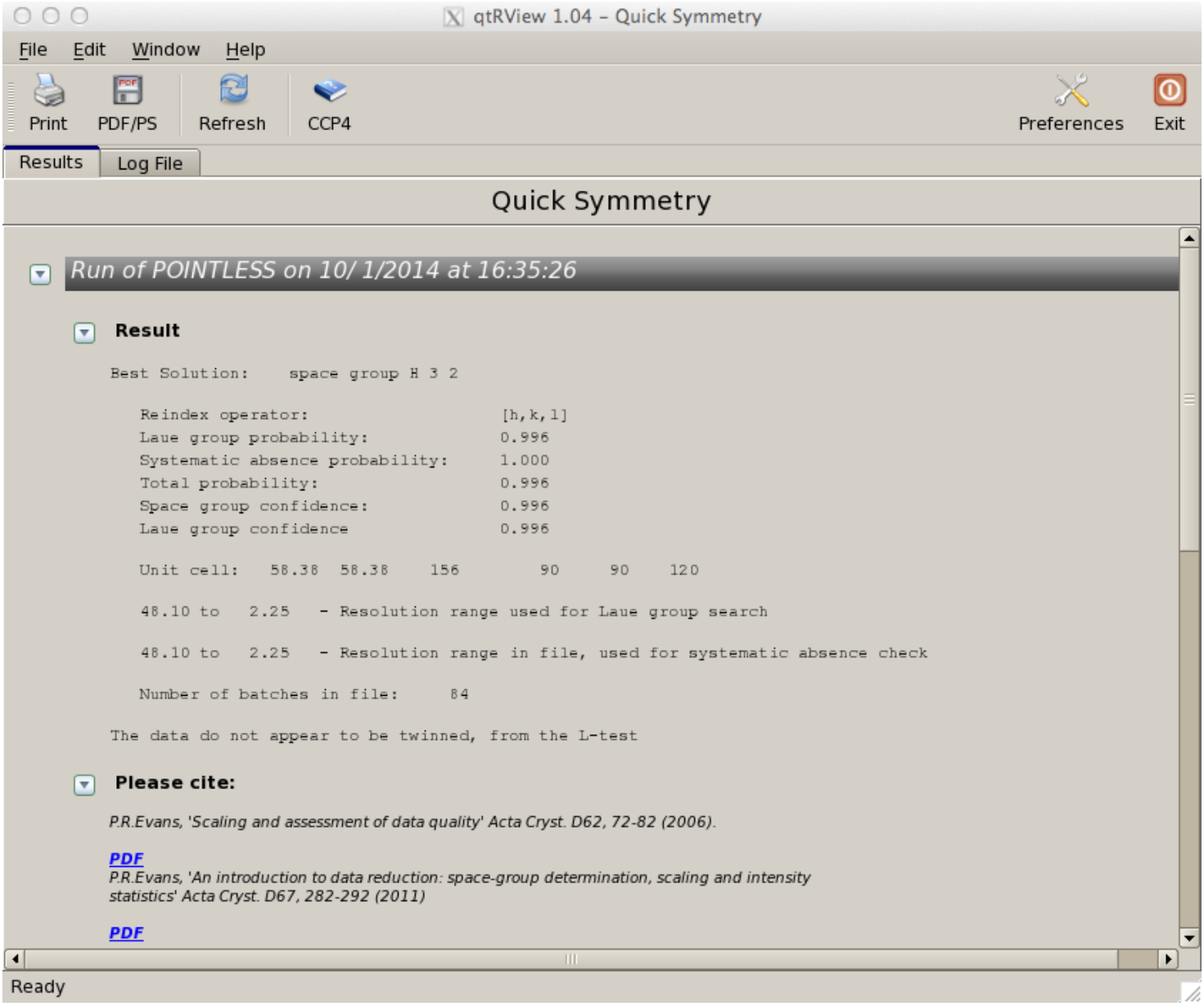
10. Running Pointless to check the symmetry

Once the images have been integrated (even a few degrees of data in many cases) the program POINTLESS can be run to determine the true Laue symmetry and try to determine the space group.



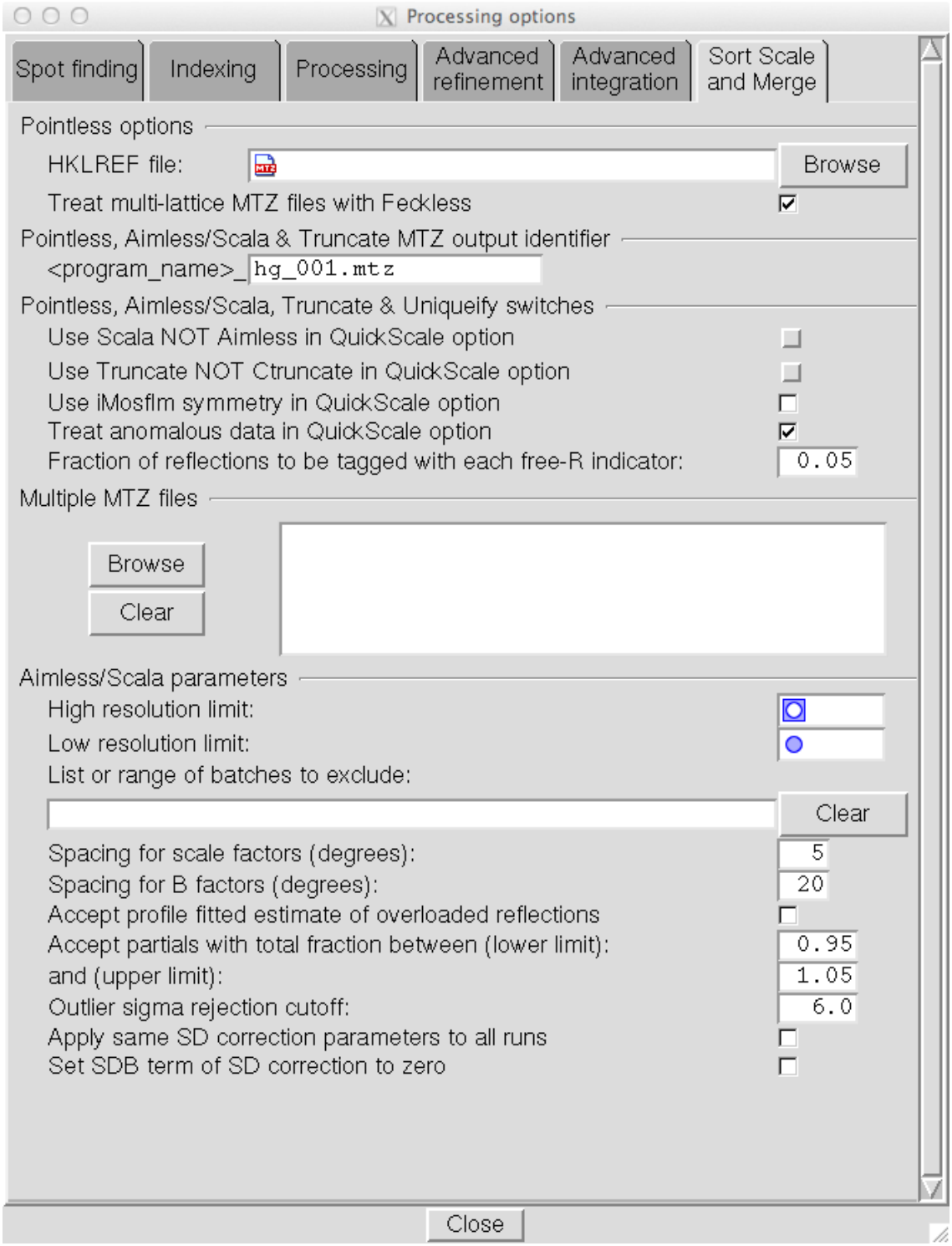
Select the QuickSymm button.

A summary of the Pointless results as shown below will appear in a "qtRView" window.



Various graphs (L test for twinning, ViewHKL to view the reciprocal lattice) can also be viewed in this window.

Various options for POINTLESS, such as a reference MTZ file, forcing the symmetry selected in iMosflm and specifying multiple input MTZ files can be selected from the Settings -> Processing options -> Sort Scale and Merge tab (see below).



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

11. Running Aimless to scale and merge the data

Select the QuickScale button.

This will first run POINTLESS then change the space group for the data to that selected by POINTLESS (unless the iMosflm symmetry is set to override POINTLESS), run AIMLESS to scale and merge the data in that space group and then run CTRUNCATE which calculates a number of statistics from the intensity data. The CCP4 UNIQUEIFY script will then be run which generates all the unique reflections for that spacegroup to the resolution of the measured data and will flag reflections (5% by default, but this can be changed) for use as the Rfree set.

AIMLESS can be run from the QuickScale button either treating anomalous data (the default) or ignoring anomalous data. If only

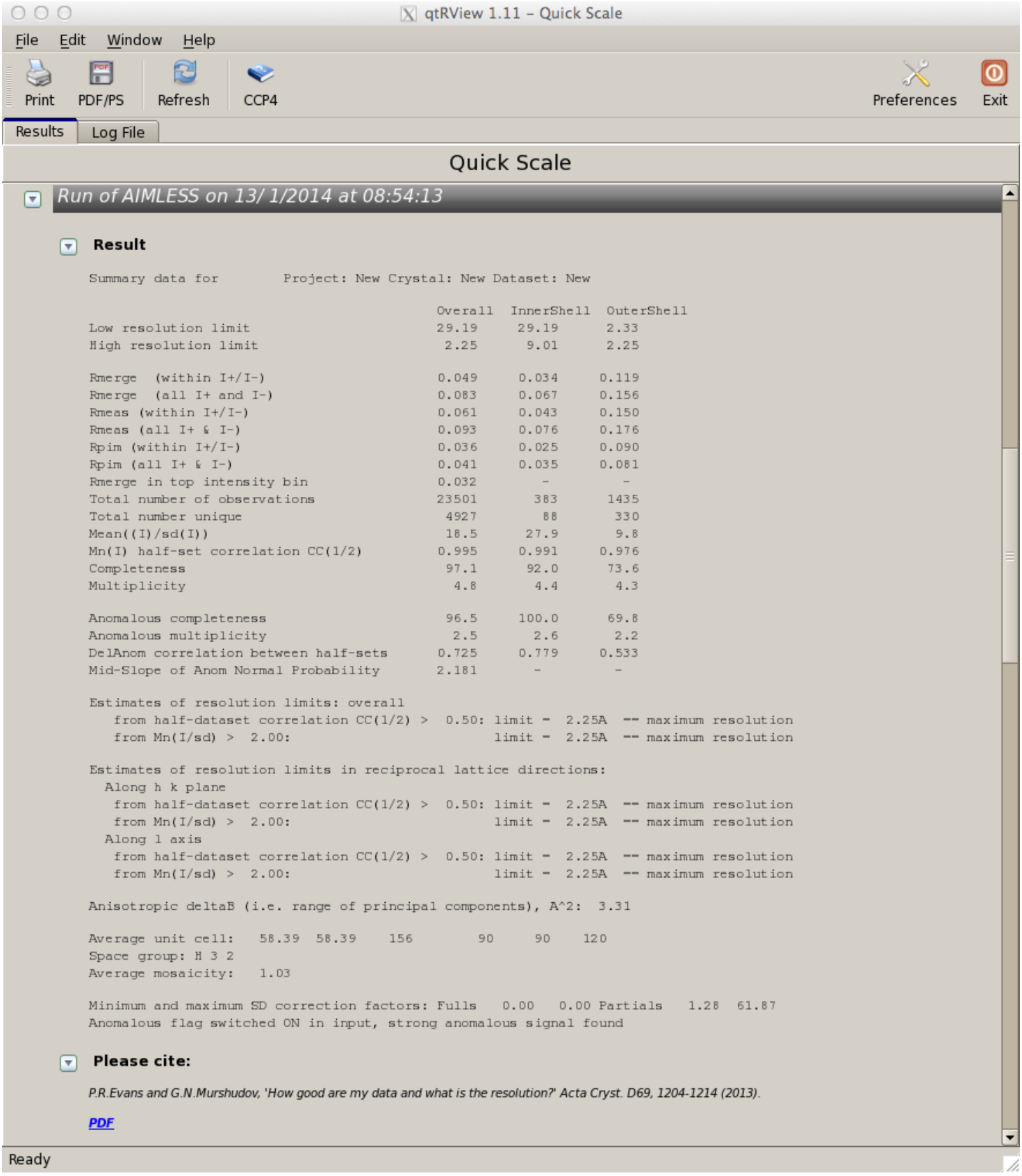
processing a few degrees of data it is usually better to ignore anomalous, or no merging statistics will be given in AIMLESS.

SCALA can be selected as an alternative to AIMLESS for scaling **(not recommended).**

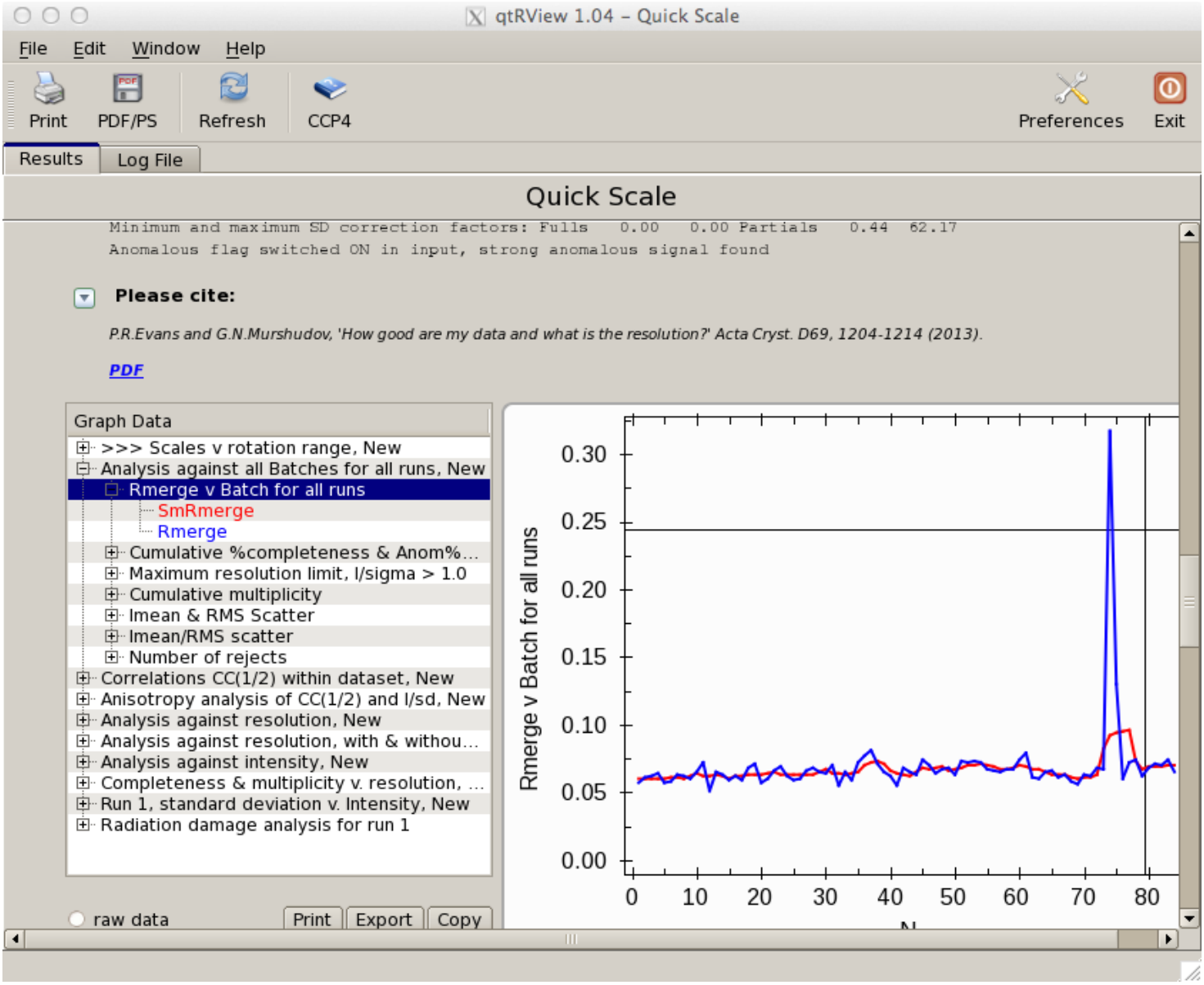
There is limited control over running AIMLESS (see figure above), including re-setting resolution limits, rejecting batches (images), changing the treatment for accepting partials, changing the rejection criterion and changing the procedure for updating the standard deviation estimates.

The AIMLESS results will be displayed in the qtRView, as shown below, along with graphical output (see below) and the full log file can also be examined.

Aimless summary file:



An example of the Aimless graphical output:



The AIMLESS logfile can also be examined by selecting the "Log File" tab.

12. History and mosflm log file

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 mosflm.lp file is also available in a time-stamped file, mosflm_*yyyymmdd_hhmmss*.lp, (e.g. mosflm_20100518_131019.lp) and can be read with a text editor.

The History has a tree structure, and will show when various parameters have been changed. This has not been fully implemented and is not generally useful.

The logfile may provide useful clues when processing has failed.

13. Warning Messages

As already mentioned in Adding images §3 and Integration §9, warning messages are generated by MOSFLM during processing. A "traffic light" colour coding has now been assigned based on the warnings and is shown next to the Warnings box:

A Red status implies a serious problem that will affect data quality and should be addressed.

An Orange status implies minor issues that should be checked to see if things can be improved.

A Green status means no significant problems, messages are for information only.

Abbreviated messages can be obtained by clicking on the "Warnings" icon. The full warning Details together with Hints & Notes text (if any are available) can be obtained by double-clicking the relevant warning. Details are displayed in a pop-up text box which will disappear with the next click outside the box. The same full details and suggestions are, of course, available in the mosflm.lp files where even more information may be available.

NOTE This is the end of the part of the tutorial which deals directly with iMosflm but see If you have time §18 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 Aimless via the CCP4i user interface, examining the traditional mosflm SUMMARY file with the CCP4 utility loggraph.

The Appendix contains a shell script for running MOSFLM as a traditional batch job.

14. Useful command line options

iMosflm has a number of command line options, and these can be listed by typing:
imosflm --help
Current options are:
--debug Creates a large output file for debugging purposes
Don't use this unless asked to by a developer
--expert Permits access to advanced detector settings.
This should not normally be required
--help Displays this message, then exits
--imagedir <directory>
Starts the image browser in this directory; if <directory>
is given as an environment variable, it will be expanded
to its full value
--allimages <filename>
Starts mosflm loading all files that match the template
generated from <filename>
--singleimage <filename>
Starts mosflm loading only the image from <filename>
--startdir <directory>
Starts mosflm in directory <startdir> rather than the current
directory. This will only work if the directory exists and
you have write permission to it, otherwise iMosflm will exit.
All the normal mosflm output files will be in this directory
--init <filename>
Starts Mosflm and reads from the given saved session file
--site <filename>
--version Displays the program version and required Mosflm version,
then exits

14.1 The site file

"imosflm --site <sitefile>" will read keywords from the file <sitefile> that will overwrite incorrect information in image headers. This saves having to manually correct values in every iMosflm session. The site file can also be read or written as an option from the Session menu (see Drop down menus §2.1). Allowed keywords are:
WAVELENGTH <x>
DISPERSION <x>
POLARISATION [PINHOLE | MIRRORS | MONO | SYNCHROTRON <x>]
DIVERGENCE <xh xv>

BEAM <beamx beamy>
DISTANCE <x>
DISTORTION TILT <x>
DISTORTION TWIST <x>
GAIN <x>
DETECTOR REVERSEPHI
DETECTOR OMEGA <x>

15. The mosflm SUMMARY file

MOSFLM produces a summary file listing refined parameters for each image, and when using iMosflm this file will be date- & time-stamped and have a name such as "mosflm_20100518_131019.sum". This contains the same information as the plots produced by the iMOSFLM GUI which now also include the number of "bad spots" and overloads.

This can be inspected graphically from the CCP4i GUI using the program loggraph. Select the pull down menu "View Files from Job" and under Output files.. select Additional log graphs.. where the *_SUMMARY.loggraph file will be listed. Select this file to display its tables in loggraph.

This is very useful to identify "rogue" images (with unusually high positional residual, or low I/σ(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 Detector parameters §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 Crystal parameters §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 standard deviations), 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 AIMLESS. 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 AIMLESS.

16. Scaling data with CCP4i

```
% ccp4i
```

Select "Directories & ProjectDir" (top right)

In the new window, select "Add project"

Enter the project name in the 'Project' field and the full directory path in the 'uses directory:' field. Default values may have been

entered already here.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 "Symmetry, Scale, Merge (Aimless)"

In the window that appears do the following.

Give a job title (top line)

Click on the box next to "Separate anomalous pairs for outlier rejection & merging statistics" (the crystals have been soaked in a mercury compound).

In the next orange bar (line starts "HKLIN #1") select Browse and select the MTZ file output by MOSFLM (probably called hg_001.mtz) Select "OK"

If you did not supply the Project name, Crystal name or Dataset name to Mosflm, fill in the boxes for these now.

Under "Scaling Protocol":

From the pull-down menu on the line starting "Scale", select "with automatic default settings"

At the bottom of window, from the pull-down menu named "Run" select "Run&View Com File"

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

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

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

When CTRUNCATE 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 as running loggraph on the AIMLESS log file; they are also the same as those that appeared in IMOSFLM's "QuickScale" task. To look at the log file itself (e.g. the overall merging statistics) select "View Log File" and it will appear in a new window.

16.1 Looking at the AIMLESS output

The simplest way to look at the output from AIMLESS is to use loggraph on the AIMLESS 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 all Batches for all runs, 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 Fractional bias should be less than 1-2%, although it will often exceed this for weak data (e.g. in the high resolution bins).

The "Axial reflections" graphs in AIMLESS were useful for detecting systematic absences, which can be used to identify the true space group. However, the detection of screw axes is thought to be sufficiently robust in POINTLESS that these graphs are no longer necessary.

It can also be useful to look at the AIMLESS log file itself, in particular the table giving statistics as a function of resolution. The effective resolution limit of the data can be estimated by looking at the Mn(I)/sd column (this is the mean I/σ(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 AIMLESS 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.

AIMLESS 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 difference between an individual estimate of intensity and the mean of all other estimates (from symmetry mates).

Probably the best estimate of the true high resolution limit of the data can be obtained from the CC(1/2) values; these are the correlations between equivalent observations that have been divided randomly into two half-datasets. The resolution limit is found when the CC drops below ~0.5 - 0.3.

This table comes under the heading:

ANALYSIS OF STANDARD DEVIATIONS

in the AIMLESS log file and 'Run *n*, standard deviation v. Intensity, test' (where *n* is 1 or more) in the CCP4i GUI 'View Files from Job' - 'View Log Graphs' display.

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, run number 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 log file). DO NOT repeat the integration. In the command file, uncomment the line (near the top) "goto aimless" so it skips the integration part.

In the CCP4i GUI click to highlight the AIMLESS job which is listed as FINISHED and select “ReRun Job..” from the control buttons down the right-hand side of the main interface window. Under “Excluded Data” in the window that appears, check the box next to 'Exclude selected batches' and pull-down on the menu button to exclude a 'range'. Give the start & end range to exclude (e.g. 74 to 74). Rerun the job as before.

17. Changing the symmetry

From the list of options at the left hand side of the main window, select “Utilities” and under that "Sort/Modify/Combine MTZ Files"

In the new window give a new title, e.g. reindex as H32

Click on the Box "Change space group and/or reindex reflections"

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 check the box next to “Change space group to” and enter H32 in the box. This is above the final item, "Reduce reflections to the asymmetric unit ", which is checked by default.

Select "Run" from the "Run" option button.

Then 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 log file, in particular the multiplicity weighted R-factor, called Rmeas in the log file.

18. Looking at the CTRUNCATE output

The CTRUNCATE program converts intensities to amplitudes, and also compiles some useful statistics. The so-called "cumulative intensity statistics" tabulated by CTRUNCATE (and plotted by loggraph) provide 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 in iMosflm.

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

19.2 Checking up on outliers

AIMLESS 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).

Select the offending image by double-clicking it in the Images pane of the iMosflm main window. Predict the reflections, then enter the *hkl* values into the boxes and hit Return (or click the red warning button following the entry box for *l*) 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.

```
#!/bin/csh -fv
#
# Maxinf Workshop on Data Processing, Cambridge, December 2002
#
# A generic command file for processing data.
# Andrew Leslie
# Variables to generate unique filenames. Change the "job" variable
# if processing multiple times, e.g. 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
#
aimless:
#
# Smooth scaling, Bfactors ON, absorption correction
#
aimless hklin ${ident}_${job}_sort_${img1}to${img2} \
      hklout ${ident}_${job}_aimless_${img1}to${img2} \
      rogues ${ident}_${job}_sort_${img1}to${img2}.rogues \
      << end-aimless
title $title
scales rotation spacing 5 secondary 6 bfactor on tails
sdcorrection 1.5 0 0.03
anomalous on
end-aimless

truncate:
ctruncate hklin ${ident}_${job}_aimless_${img1}to${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 and Owen Johnson