

iMosflm introductory & intermediate lessons

## READ THIS FIRST

Read the text carefully - most questions that you are likely to have are addressed in this document.

If you have not used iMosflm before, or only processed straightforward datasets, do section (1) below, then go onto section (2).

If you have used iMosflm extensively and think you know what you are doing, try section (3). If you need help, read the corresponding text in section (2) carefully before asking a demonstrator.

You will be told the location of the images for the tutorials at the start of the session.

Extra information that explains things is in boxes like this

advice on running the program and interpreting output is in boxes like this

## (1) Introduction to iMosflm

Even if you have run iMosflm before you should start here; if you are an experienced iMosflm user, this will take about 5 or 10 minutes. If you have not run iMosflm before, it will take considerably longer because you will be learning a lot of new things.

You will learn how to do the following:

1. start iMosflm from the command-line, or from within ccp4i or ccp4i2
2. load a set of images into iMosflm
3. use circle fitting to mask the beamstop shadow
4. use circle fitting to locate the direct beam position
5. index a set of straightforward images & obtain an initial estimate for the mosaicity
6. refine the detector and crystal parameters (including the unit cell dimensions)
7. improve your initial estimate of the space group
8. integrate a dataset
9. use Pointless and Aimless to merge and scale a dataset
10. find and examine an unusual image
11. use the Strategy task to design a more efficient data collection strategy

-----

### (0) Before you start

Provided that a recent version of CCP4 has been installed (and all updates have

been applied) you can run iMosflm from within CCP4 and it will make sure a record is kept of everything you do in a convenient way.



To start ccp4i (the original GUI, ccp4i) or ccp4i2 (the new GUI, ccp4i2, which may not be available on your system yet), type “ccp4i” or “ccp4i2” on the command-line in a terminal window, or double-click the appropriate icon on the Desktop.



#### (1a) Start iMosflm from the command-line

iMosflm can be run from outside CCP4 – this can be the quickest and easiest way to proceed.

You need to run iMosflm from a directory on /localdisk, not from your home directory – integration will not work if run anywhere in /home, because of special features of the Oulu set-up.

DO NOT run iMosflm in the directory that the images are in – create a new working directory that you have write access to (here we are calling it “tutorial”). You may find it helpful to work in a separate directory for each of the examples. On the command-line:

```
mkdir /localdisk/tutorial
cd /localdisk/tutorial
imosflm
```

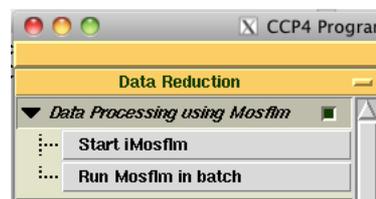
Provided that iMosflm has been set up correctly, the interface should now appear.



#### (1b) ccp4i Start iMosflm from within ccp4i

Select (at the top-right) “Directories & Project Dir” and create a new project by pressing the “Add Project” button near the bottom of the window. Choose a new project name and a directory to work in, then select your new project using the drop-down menu on the line starting “Project for this session...”. Press “Apply & Exit”. When you have done that, on the drop-down menu at the left,

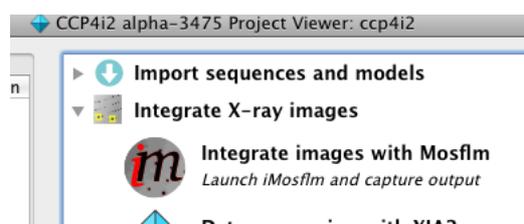
- select the "Data Reduction" module
- select "Data processing using Mosflm"
- select "Start iMosflm"



#### (1c) ccp4i2 Start iMosflm from within ccp4i2

select "Start a new crystallography project"

in the new window that opens, select your home directory and then click on the folder icon with the star on it (near the top right of the window) and create a new folder for your work. Create a new project for your work. A new window opens

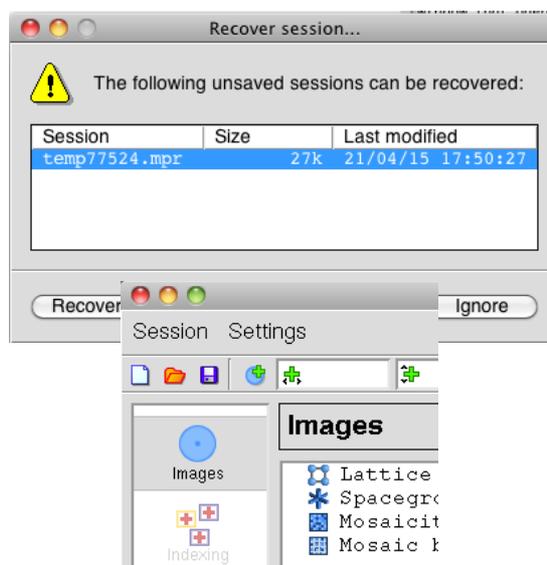


- select "Integrate X-ray images"
- select "Integrate images with Mosflm"
- enter a Job title and select "Run". Do not try to "use data from job".

The iMosflm interface should appear.

If a small window titled "Recover session..." appears, choose "Delete", "Delete All" or "Ignore". This option is only of use if a previous iMosflm job has exited unexpectedly.

Unless you choose to recover a session from a previous job, iMosflm *always* starts a new session - there is no need to choose this option again unless you want to process a second (or subsequent) dataset without restarting iMosflm.

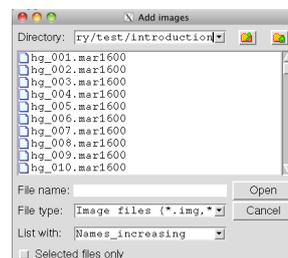


You will notice that there is a group of icons on the left-hand side (Images, Indexing, Strategy, Cell Refinement, Integration, History); these are the six basic "tasks" that iMosflm can perform. Some of the icons are "greyed-out" and do not respond to the mouse; by design, iMosflm will only allow you to perform tasks that are appropriate for the current stage of data processing.

## (2) Load a set of images

Click on the small blue circle with a green cross  near the upper left of the iMosflm display to add images to the current session.

In the file browser that appears, locate the directory that the images are in. It may be faster to type the name of the directory that the images are in in the entry field (see the "READ THIS FIRST" section) rather than using the file browser buttons to navigate.



By default, images with "normal" file extensions are listed - see the "File type" entry field. Some images have unusual file extensions, and you can change the extension that iMosflm looks for using the drop-down menu next to this

entry field.

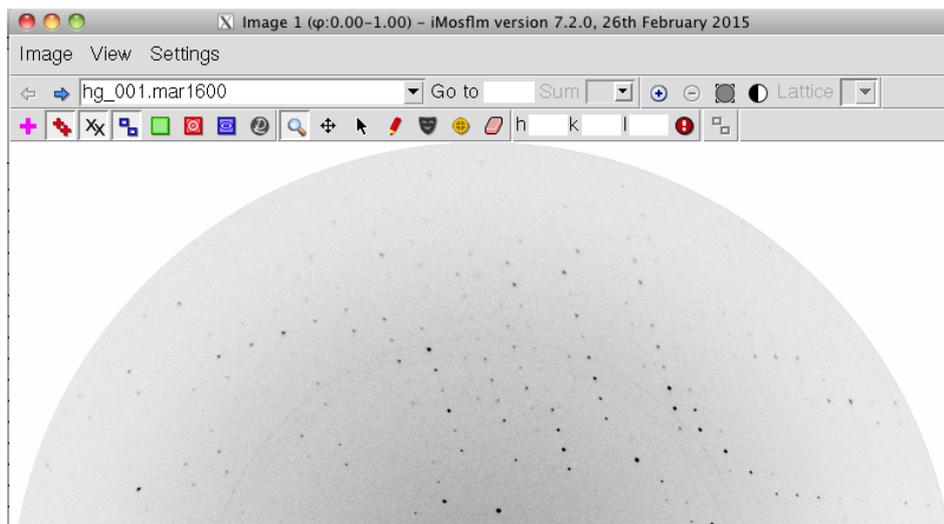
Double-click on the image "hg\_001.mar1600"; this will load all the images with the same "template".

There is no need to select all the images in the dataset - iMosflm assumes that (normally) you will want to load a complete dataset, so it finds all images that fit the "image template" if you pick any file in the dataset. This dataset contains 84 images.

If you highlight an image and press "Open" it will have exactly the same effect as double-clicking an image.

The image that you double-clicked will be displayed in a new window.

Selecting *any* of the images in the dataset will find all the images in the dataset, but the displayed image will be the one you chose.



All the buttons near the top of the display have "tool-tips" - if you let the mouse cursor rest over any of them for a few seconds, the function of the button is displayed.

Note carefully - the red and blue square buttons look as if they have similar actions, but their effects are very different. The red button only affects the spot finding for autoindexing, and the blue button only affects the resolution limits used in refinement and integration.

-----

### (3) Use circle fitting to mask the beamstop shadow

Zoom in on the central part of the image by pressing the  twice (on the top row of buttons, next to the "Sum" entry box).

There is a faint “powder ring” with a few intense spots close to the beamstop. You can safely ignore this ring.

The current direct beam position is displayed with a magenta cross (in this case it is set to the centre of the image because the image headers for this dataset do not have this information – most images do!)

Press the yellow cross & circle button . Three new buttons will appear on the top-right of the image display area.

Using the mouse, click to select four or five points around the edge of the beamstop shadow (there is a very clear shadow and a larger, more diffuse shadow – choose the larger one).

Each picked point is displayed as a small yellow cross; if you click on one of these, it will disappear and its coordinates will not be used in this calculation.

Select the icon with the green diagonal cross to fit the beamstop mask (also called the backstop mask).



This will display the new beamstop mask. iMosflm will ignore all points in this region of the detector in this session (i.e. it will not search for spots or try to process any here) unless you change the mask.

If you press the “show masked areas” button , the mask can be hidden or made visible and you can see the yellow circle that has been calculated for the fitting procedure.

**VERY IMPORTANT** – after fitting the beamstop, select the icon with the red diagonal cross to clear the points used in this calculation.



Both the points used and the fitted circle will disappear from the display.

If you do not clear the points, they will be stored and might be used in calculating the direct beam position from a circle.

-----

#### (4) Use circle fitting to locate the direct beam position

Zoom out by either

- clicking on  twice or
- clicking on  once (which zooms right out in one action)

You should see two faint ice rings on the image. Either of these can be used to give a good estimate of the direct beam position.

Using the mouse, click to select four or five points around one of the ice

rings.

Select the icon with the green orthogonal cross to fit the beam from the circle. This updates the beam position used by iMosflm, and you should notice that the magenta cross moves a little from its original position.



VERY IMPORTANT – after fitting the direct beam position, select the icon with the red diagonal cross to clear the points used in this calculation.



Both the points used and the fitted circle will disappear from the display.

If you do not clear the points, they will be stored and might be used in calculating a circular beamstop from a circle.

-----

#### (5) Index the images & estimate mosaicity

Go back to the main window; you will notice that the "Indexing" icon in the column at the left is no longer greyed-out.



Indexing

Click on the Indexing icon to enter the Auto-indexing task.

By default, iMosflm will now automatically find a suitable selection of spots for indexing on two images as close to 90° apart, run the indexing, choose a solution, and estimate the mosaicity.

Provided that indexing has given a clear solution, that is all you need to do here - but you should ask "what is a clear solution"?

iMosflm attempts to index the images you gave it, and if the algorithm runs successfully, it will list 44 solutions in the results window, sorted on increasing penalty. All these solutions are based on the triclinic basis solution (no. 1 in the list), but have extra Bravais lattice symmetry enforced; the "penalty" indicates how well each higher symmetry solution fits the triclinic basis - there should be a large jump between the "worst good" solution (with a penalty  $\ll 20$ ) and the "best bad" solution (with a penalty  $\gg 20$ ).

iMosflm will refine each of the solutions with a penalty  $\leq 50$ ; good fits will have a  $\sigma(xy)$  value very close to that for the triclinic solution.

iMosflm chooses the solution with a good penalty and small  $\sigma(xy)$  with the highest Bravais lattice symmetry, assigns the lowest symmetry space group with that Bravais lattice symmetry, then runs the mosaicity estimation to give an initial idea of what the mosaicity is.

Go back to the image viewer window, and confirm that the prediction boxes are displayed over the diffraction spots. Provided that most of the spots are

covered by predictions, and most of the predictions cover spots, you can be confident that indexing has been successful.

If you zoom in on the image, you can use the “panning tool”  to move the display around to see different parts of it in close-up.

You will see that red and yellow crosses are displayed on the image, along with yellow and green boxes (you might also see a few blue boxes and red boxes). If you don't see any crosses, make sure that you are looking at one of the images that has been used for autoindexing.

The crosses mark spots that have been found on the image for autoindexing –

\* Red crosses are stronger than a threshold ( $I/\sigma(I) > 20$  in this example) and have been used in the autoindexing.

\* Yellow crosses have intensities below the threshold.

You can change the threshold by changing the value in the entry box just above the “Read spots file” and “Save spots file” buttons, but in this exercise you should not do this.



The boxes show where iMosflm is predicting where spots should be on the image using the results of autoindexing.

\* Yellow boxes show where partially recorded reflections (“partials”) are – these spots will be spread over more than one image

\* Red boxes show overlapped reflections – where more than one reflection is predicted to be on this image

\* Blue boxes (rare these days) show fully recorded reflections (“fulls”), which are only present on a single image

\* Green boxes show partials that are spread over too many images to be measured reliably, so iMosflm will not try to measure them.

The mosaicity will affect how many images iMosflm expects a reflection to be spread over.

-----

## (6) Refine the detector and crystal parameters

Choose the "Cell Refinement" task.



Cell Refinement

iMosflm will automatically choose an appropriate set of images for a good refinement, based on the symmetry chosen in the Indexing task and the mosaicity.

All you should need to do is to press the "Process" button. iMosflm will finish the refinement when it has converged, so there is usually no need to repeat by pressing the "Process" button again. The prediction boxes should match the spots better after refinement.

Check the predictions on other images to make sure they match the spot positions well.

Do not try to choose your own selection of images to refine the cell.

There are seven smaller panels in the task window. Each of these can be expanded to fit the main window by using the mouse to shift-click, control-click or by using the mouse scroll wheel.

On the left there are two sets of parameters; the top pane lists those that are refined using the spot positions on the images, and the bottom pane lists those that are refined with "post-refinement", which is based on the relative intensities of reflections that are partial across several images. The two sets of parameters are refined independently, which means that, for example, the crystal to detector distance can be refined at the same time as the unit cell dimensions - this cannot be done reliably in most other programs.

In the middle panels, the values of the refined parameters are plotted against image number. At the end of refinement, these plots should vary reasonably smoothly and not have any large jumps.

The top right-hand pane shows the average spot profile in the central region of the detector for each image used. Most of the intensity should be inside the inner blue region. You can look at the profiles for different images by selecting them in the image list on the left of this pane.

The bottom right-hand pane shows the RMS residual for each image - it should be about one-quarter to one-third of the pixel size for a good dataset; these images have pixels 0.15mm on a side, so values of ~0.04 - 0.05mm would indicate a good processing run.

The pane at the bottom shows the cell parameters before and after refinement, with standard deviations for each.

-----

### (7) Improve your initial hypothesis for the space group

As noted above, the Indexing task chooses the lowest symmetry space group with the highest symmetry Bravais lattice; so screw axes or extra Laue symmetry are not considered. We may be able to get a better idea of the true space group if we analyse some images more closely before proceeding.



Choose the "Integration" task.

iMosflm will automatically choose all the images in the dataset - change this to process the first few images (say, about 5 or 10). Do not make any other changes.

Press "Process"; iMosflm will integrate your chosen images (the plots produced will be discussed in the next section).

Press the "QuickSymm" button at the top of the main window. This will run the program Pointless and launch a results browser; check the suggested best Space Group. If Pointless indicates that the Laue group is different from the one you chose, it will print a warning.

Pointless will report the best fitting Laue group to the data, but iMosflm reports Bravais lattices. The Laue group symmetry is what can be determined from the diffraction data (*i.e.* in reciprocal space), but the Bravais lattice (when combined with the Point group and any translational symmetry) determines the Space group (in real space).

(i) if the Space Group is different to the one given by iMosflm, but the Bravais lattice is the same, you can just go back to the Indexing task and change the Space Group in the drop-down entry box at the bottom-left of the main window. There is no need to run the Cell refinement again.

(ii) if the Bravais lattice is different, go back to the Indexing task, choose the appropriate Bravais lattice (by clicking on the solution in the list) and also choose the Space Group indicated by Pointless. You must now run the Cell refinement again.

-----

## (8) Integrate a dataset

Choose the "Integration" task.



Integration

Press the "select all"  button (to the right of the image selection entry box) to choose all the images in the dataset. Do not make any changes to its selection.

iMosflm will choose to refine the detector parameters (by positional refinement) and the crystal mis-setting angles and mosaicity (by post-refinement). The unit cell parameters are fixed by default in the integration task and should not normally be refined here.

Press "Process"; select "Yes" when iMosflm asks if you want to overwrite the MTZ file (which was just written in step 7). iMosflm will integrate your dataset.

There are nine smaller panels in the task window; the top three are the same as those in the Refinement task window, as are the two left-hand panels in the middle row.

Of the four new panels, the most useful are:

(i) middle row, right-hand pane containing the profiles for each region of the detector, calculated for each block of images. iMosflm divides up the dataset into evenly sized blocks - in this case, 12 blocks each containing 7 images. You can look at the profiles for different blocks by selecting them in the list on the left of this pane.

(ii) bottom row, right-hand pane displays (by default) a plot of average intensity against resolution. For this example, the crystal diffracted well beyond the limits of the detector, so the  $I/\sigma(I)$  in the high resolution bin is around 4 or 5 for all images.

For a well-processed dataset, all the graphs should be reasonably smooth without any large jumps in values. You should notice that there are jumps in the plots for post-refined parameters around image 75. You will find out why in a later section.

-----

#### (9) Use QuickScale to merge and scale a dataset

Stay in the Integration task.

Press the "QuickScale" button. This runs Pointless and Aimless; a results browser will open, which contains overall processing statistics in the classic Crystallographic "Table 1", as well as numerous graphs.

It is a matter of debate (see various discussions on the ccp4 bulletin board over many years!) which values are important in Table 1; however, in general the statistics should be better for the inner shell than for the overall statistics, which should in turn be better than for the outer shell. Overall, this is a particularly well-diffracting crystal which diffracts strongly to the edges of the detector, so its statistics are rather good. It is a mercury derivative of a protein, and a strong anomalous signal is apparent (as can be seen from the "DelAnom correlation between half-sets" (which would be close to 0.0 for no anomalous signal; correlations over about 0.25 - 0.3 indicate an anomalous signal) and "Mid-Slope of Anom Normal Probability" (which would be ~1.0 for no anomalous signal; values >~ 1.1 indicate some signal, >~ 1.2 would normally be considered quite large).

Because this dataset was collected for the mercury signal, no attempt was made to collect the very high resolution data.

Scroll down to the graph window for Aimless, and choose "Analysis against all

Batches for all runs". By default, the graphs for "Rmerge v batch for all runs" is displayed (if it is not, choose it). Which batch (i.e. image) does not fit with the rest?

-----

(10) Find and examine an unusual image

Go to the image viewer and select the image you just identified - do you notice anything odd about it? Use the back and forward arrows on the image display to view images close in sequence, and confirm that Aimless has identified the outlier.

Having identified the outlier, it is a useful exercise to find out how removing it from the data analysis affects the overall data quality; this could be measured by the merging statistics given by Aimless, or in the electron density map quality, or the ease of solution with SHELXC/D/E, or a variety of other ways - there is no single right answer.

Exactly *how* the image is removed from processing is up to the user; the simplest (and fastest) option is to just remove the image from QuickScale (see Settings->Processing Options->Sort, Scale and Merge), rather than repeating the integration and excluding the image in that step, but if you are interested it is worthwhile trying the different options to determine which (if any) is best.

-----

(11) Use the Strategy task

iMosflm has a function to calculate an efficient data collection strategy based on images that have already been collected. This is of use when the crystals suffer severe radiation damage and a dataset requires data from several crystals.

Go to the "Strategy" task. You will notice that there is a "pie" in the pane at the bottom of the main window, with a red section marking the images that have been collected already. The object of this exercise is to find out what would have been the most efficient way to collect data from this crystal.

In the top pane, uncheck the "use" box for the matrix; this will let the Strategy task use the "Matrix" (*i.e.* the orientation matrix, which is what was actually determined in the autoindexing step) but ignore the images that have been collected for this matrix.

Press "Autocomplete", and use the defaults in the pop-up window. Press "Ok"; the displays update to show what rotation range would give the highest completeness for the dataset based on the crystal's unit cell parameters, symmetry, and current orientation.

The rotation range for this calculation is the minimum required for a 100% complete dataset for a crystal with the current symmetry - note that the suggested data collections do not start at  $0^\circ$  in phi. If the Space group

were H3, 120° total rotation would be necessary, but for the correct H32 space group, only 60° is needed.

Try Autocomplete with both H3 and H32 - make sure you uncheck "Use" (in the "matrix" pane before each run.

Run Autocomplete again for both Space groups, but in the pop-up window check the box for "anomalous data". Do the rotation ranges match those found previously?

If the normal strategy option is run, iMosflm will choose the rotation range that will give the maximum completeness for unique reflections in the minimum total rotation for the current space group. No attempt is made to optimise the completeness of Friedel mates (which correspond to Bijvoet pairs in the presence of an anomalous signal) or other symmetry-related reflections.

If you choose the option to optimise for anomalous data, iMosflm will calculate the appropriate range that includes as many Bijvoet pairs as possible. These reflections are used for calculating anomalous differences in SAD, MAD, SIRAS and MIRAS experiments).

The two ranges may well overlap but it is likely that they are different.

If you have data from more than one crystal, it is possible to read the matrices for the other crystals and calculate a strategy for the current crystal. See Example 6 in Section 2, on Intermediate use of iMosflm.

-----

This tutorial is based on our comprehensive tutorial available at

<http://www.mrc-lmb.cam.ac.uk/harry/imosflm/ver711/documentation/tutorial.html>

which uses the 84 images contained in the file at

<http://www.mrc-lmb.cam.ac.uk/harry/imosflm/images/hypF.tar>

---

## (2) Intermediate use of iMosflm

This exercise works through some examples of datasets which display relatively common pathologies.

In EACH CASE, load the images into iMosflm and try to index them using default settings before doing anything else.

Do not work in the image directories, and use separate work directories for each example.

Note that “predict correctly” means

- (1) most of the spots are covered by prediction boxes
- (2) most of the prediction boxes are over spots
- (3) it doesn't matter too much if a few spots are not predicted (particularly those near the rotation axis)
- (4) it doesn't matter too much if a few prediction boxes appear where there are no spots

-----

### Example 1

There are 3 images in this example - directory “example1”, image filenames Example1\_1.0001 - 0003.

This example illustrates

1. the importance of having the correct beam centre in processing
2. the use of the “beam search” function in iMosflm
3. how to add images which do not have “normal” filenames (*i.e.* without an extension like “.img” or “.cbf”) to the current session

You will need to use the iMosflm “Add images” option “Numbered files” or “All files”, as these image filenames do not have an extension.

Q1: Is there a clearly correct solution?

Q2: What is the  $\sigma(xy)$  for iMosflm’s chosen solution?

Q3: Do the predictions for iMosflm’s chosen solution match up with the spot positions on the images?

Q4: Do any of the solutions predict correctly?

Correct indexing depends critically on having the right

- \* wavelength,
- \* crystal to detector distance
- \* direct beam position

These three pieces of information are usually in the image header, but occasionally one or more of them is wrong. Because the direct beam position is the one that is most often incorrect, iMosflm has a “Beam Search” function, which tries 25 beam positions close to the current setting to determine which is “best”.

Try this function and answer the following questions.

Q5: How many of the trials refine to the same position?

Q6: What is the  $\sigma(xy)$  for the top solution? Is it the same for all of the trials that refine to the same solution?

For a good dataset collected at a synchrotron, the  $\sigma(xy)$  for the right

solution will be around half the pixel size for any given detector (which can be found in “Settings → Experiment settings → Detector”. For a good dataset collected on a home source, the  $\sigma(xy)$  will be about 1 or 2 pixels.

Choose a solution by double-clicking it with the left mouse button.

Index the images again using the “best” beam position.

Q7: Does the list of solutions have a clearer cut-off between the “worst good” solution and the “best bad” solution?

Q8: Do the predictions match the spots better?

-----

## Example 2

There are 59 images in this example - directory “example2”, image filenames Example2\_001.img - Example2\_059.img

This example shows

- (1) that some image files have incorrect information in their headers
- (2) some datasets cannot easily be indexed using the default settings
- (3) some images are more useful than others when indexing
- (4) the graphs in iMosflm can be used to identify problems in processing

Q1: Does the indexing give a clear solution?

Q2: Does using the “Beam Search” help?

Note that the beam centre displayed by iMosflm is not in the same place as an intense spot close to the middle of the beamstop. Many beamstops are artificially thinned so that they don't completely block the direct beam, in order to show where the beam position is, even if (as in this case) the beam position is recorded incorrectly in the image header.

Correct the beam position so that it coincides with the intense spot in the middle of the beamstop.

Q3: Does this help with the indexing?

Try indexing using

- (i) only image 1
- (ii) only image 59

Q4: Does either give a solution that allows you to accurately predict the spot positions on nearby images? Remember to look at image 59 when you index using it.

At least one of the “single image” indexing attempts should work. In the “Cell Refinement” task, refine this solution until the cell edges do not

change and the predictions match the spots for both image 1 and image 59 (you may need to refine several times).

Integrate the dataset.

Q5: Do you notice anything unusual about the graphs?

Q6: What is significant about any bad images (look at them in the image viewer)?

-----

### Example 3

There are 2 images in this example - directory "example3", image filenames Example3\_001.mccd & Example3\_090.mccd. The sample is tetragonal lysozyme, space group  $P4_32_12$ ,  $a = b \approx 79\text{\AA}$ ,  $c \approx 38\text{\AA}$

This example shows

- (1) images of questionable quality can be processed
- (2) features other than the desired diffraction can be useful in indexing
- (3) knowledge of the individual beamline can be very important
- (4) prior knowledge of the unit cell dimensions can be important

Q1: What do you notice immediately about the beam centre cross in iMosflm?

Address the issue you noticed in Q1. Try indexing now.

Q2: Do the predictions match either image 1 or image 90?

Try indexing using

- (i) only image 1
- (ii) only image 90

Q3: Does either give a solution that allows you to accurately predict the spot positions on either image?

This is an example of some images from a beamline where the crystal is rotating in the opposite direction to the way we expect - there are several beamlines at synchrotrons (e.g. Shanghai, Melbourne) around the world where this is the case. The classic way to detect the problem is to observe that indexing fails (or gives rise to predictions that don't match the spots) if we use two images  $90^\circ$  apart, but indexing using only a single image gives good predictions for that image only.

Open up "Settings  $\rightarrow$  Experiment settings  $\rightarrow$  Experiment" and check the box "Reverse direction of spindle rotation".

Index again (using both images 1 and 90) and check the predictions again.

Q4: What do you notice about the unit cell dimensions now compared to the values

found for each image individually without setting “Reverse direction of spindle rotation”?

-----

#### Example 4

There is no example 4 in the “Intermediate” exercises.

If you are using iMosflm 7.1.2 or later, the example is trivial, but if you are using an earlier version it's *very* hard. If you are interested, the images are in directory “example4”, image filenames Example4\_001.img - Example4\_020.img & Example4\_440.img - Example4\_460.img.

-----

#### Example 5

There are 22 images in this example - directory “example5”, image filenames Example5\_001.img - Example5\_020.img, Example5\_050.img & Example5\_100.img.

This example shows the importance of

- (1) using sufficient images to index
- (2) using Pointless to check the symmetry
- (3) the effect of changing the mosaic block size

Try setting the beam centre to the intense spot behind the beamstop, then indexing.

Q1: How well do the predictions match the spots?

Use the beam search, then index again with the “best” solution.

Q2: How well do the predictions match the spots now?

Q3: Are all the low resolution spots predicted?

Q4: What do you estimate the maximum resolution of the data in the region perpendicular to the rotation axis to be?

Set the resolution for processing to about 0.2Å higher than you estimated in Q4.

Do not try to refine the cell if the resolution is worse than ~3.5Å. If you have chosen the orthorhombic solution, you will probably find that the cell refinement will crash; if you have chosen the tetragonal solution, the refinement will probably be okay.

Integrate the first 20 images and run the QuickScale option - make a note of the low resolution limit, Rmeas, Rpim, I/sig(I) and total number of observations.

Q5: What symmetry does Pointless indicate?

Q6: What is the maximum resolution indicated by Aimless?

If the symmetry suggested by Pointless does not agree with that found from indexing, go back to the indexing task and choose a solution that agrees with Pointless. Then integrate & scale images 1 - 20 again.

Go to the “Images” task and change the “mosaic block size” to 5 microns.

Q7: What happens to the predictions?

Integrate the first 20 images again, run the “QuickScale” option and compare the values of low resolution limit, Rmeas, Rpim, I/sig(I) and total number of observations with those obtained with the default mosaic block size of 100 microns.

Q8: What do you notice about these values?

Repeat the previous step with mosaic block sizes of 1 and 0.5 microns.

Q9: Which processing gives the best results?

-----

### Example 6

There are 59 images in this example, collected on four different isomorphous crystals; there are 3 partial datasets (each containing 19 images) and two initial reference images from a fourth crystal.

This example shows

1. the use of the multi-crystal data collection strategy
2. that you can save and retrieve a strategy for use later

This is useful if your crystals die before you can collect a full dataset from any one crystal.

Load the images with prefix “xtal1” (i.e. the first 19 images only) and index them (use the button “pick 2 images ~90 degrees apart”, which will pick spots only on images from the last crystal added), and estimate the mosaicity. Go to the “Strategy” task.

Q1: What is the unique completeness for this crystal? What is the anomalous completeness?

Go back to the “Images” task, and load the images for the second crystal (xtal2) into the same session; index them (be careful to choose images from only one crystal when indexing!), estimate the mosaicity and go to the “Strategy” task.

Q2: What are the unique and anomalous completenesses for both datasets together?

Note that

1. there are now two small circles in a pane on the left showing the rotation ranges collected for each crystal.
2. the completeness is calculated for both segments together.

Load the images for the third crystal (xtal3); index them, estimate the mosaicity and go to the "Strategy" task.

Q3: What are the unique and anomalous completenesses for the three datasets now?

Load the two images from the fourth crystal, and index (and estimate mosaicity). Go back to the "Strategy" task. Select "Autocomplete", and in the pop-up window change the rotation from "Auto" to  $10^\circ$ . Click "Okay".

Changing the rotation from "Auto" to  $10^\circ$  forces iMosflm to calculate which  $10^\circ$  rotation will give the largest increase in completeness.

Q4: What are the optimum start and end phi values for this crystal?

Q5: Are they the same as the images that have already been collected?

Q6: What is the total completeness for these four crystals using the optimum start and end phi values?

If we want to check the completeness for other phi ranges for any of the crystals, we can do so as follows.

- (1) Click on the small circle corresponding to the crystal that we want to check.
- (2) Click on the yellow hatched (it will have changed to red if you have gone to a different task and returned to the "Strategy" task) wedge in the large circle, so that black squares appear at the start and end of the sector.
- (3) Drag either of these squares around the circle to give a different phi range. On releasing the square, the sector will turn red and the completeness will be recalculated.

Q7: What is the total completeness for these four crystals using the start and end phi values used for the reference images for the fourth crystal?

If a large number of crystals are being used, it is best to save the details from all the crystals that have been indexed in a file; this can be read by a later iMosflm session to allow the multi-crystal strategy to be calculated without having to re-enter all the images that have been collected. A new file should be saved after each crystal is added.

Save the current multi-crystal settings by selecting the "Save" icon from the right-hand side of the menu bar (\*not\* the "Session → Save" option!) and save the current strategy calculation to a file (you should specify a suffix ".str" explicitly).

Exit iMosflm, start a new session, read in the images for xtal4 and index.

Go to the "Strategy" task, select "Add" from the menu bar and specify the name of the saved strategy file.

Note that the values displayed for overall completeness etc. are the same as before.

Imagine that, after integrating, we have decided that crystal 2 "died" after 7 degrees.

Select the small circle for crystal 2, and change the phi end value so that the total range for this crystal is  $7^\circ$  rather than  $10^\circ$ .

Q8: What are the overall and anomalous completenesses now?

If we decide that any of the crystals have completely died, or are not isomorphous, we can remove them from the calculations - the easiest way to do this is to select the crystal's small circle and reduce the rotation range to  $0^\circ$  (simply deleting the crystal does not cause the calculation to be updated).

### (3) Advanced exercise

Only attempt this if you really are an expert - the solutions are not obvious!

-----

#### Example 1

Index and integrate the three images.

Q1: What is the Laue Group?

-----

#### Example 2

Integrate the dataset.

Q1: Are there any outliers, and if so, what is the cause?

-----

#### Example 3

Index these images.

Q1: What is iMosflm's indexing choice (symmetry, cell dimensions, penalty,  $\sigma(xy)$ )?

-----

#### Example 4

There is no example 4 if you are using iMosflm 7.1.2 or later, because it's really easy.

If you are using iMosflm 7.1.1 or earlier -

Q1: What are the unit cell dimensions?

Q2: What is the symmetry?

-----

### Example 5

Integrate the first 20 images of this dataset with mosaic block sizes of 100, 5, 1 and 0.5 microns (a total of 4 integration runs); run QuickSymm in each case.

Q1: What is the Laue group for this dataset?

Q2: Which processing run gives the best statistics?

-----

### Example 6

Q1: What are the unique and anomalous completenesses calculated if you use the images from

1. crystal 1?
2. crystals 1 & 2?
3. crystals 1, 2 and 3?

Q2: What phi range for crystal 4 would give the best overall completeness (from all four crystals)?

Q3: What are the unique and anomalous completenesses now?

Q4: Does the best range for crystal 4 correspond with its start and end phi values?

Q5: If crystal 3 died after 7°, what would the unique and anomalous completenesses be?

-----

## Answers to exercises (section (2))

### Example 1

Q1: No. For a clear solution there should be a large jump in penalty between the “worst good solution” and the “best bad solution”. In this case, it's hard to be sure which solutions are the “worst good” and the “best bad”.

Q2: iMosflm's chosen solution has a  $\sigma(xy)$  of  $\sim 0.67$ mm. The detector is a Mar CCD with square pixels of size  $0.0732$ mm - so for a good, high resolution dataset, you would hope to see a  $\sigma(xy)$  of  $\sim 0.03 - 0.04$ mm for the correct solution. Since the triclinic basis solution (the first one in the list) has a  $\sigma(xy)$  of  $\sim 0.52$ mm, you have an immediate indication that the indexing has not succeeded properly.

It should be remembered that all 44 solutions are actually based on solution 1 (but with added symmetry imposed); if the predictions for solution 1 do not match the spots, indexing has failed even if it gives you a solution.

Q3: The predictions don't match where the spots are for the chosen solution.

Q4: Even if you look at the basis triclinic solution, you should see that the predictions don't match the diffraction very well, although they are closer than for iMosflm's chosen solution.

Q5, Q6: There are actually two groups of refined beam positions which differ by  $\sim 2.2$ mm - one has a  $\sigma(xy)$  of  $0.05$ mm (8 solutions refining to 150.05,  $\sim 150.37$ ), and the other has a  $\sigma(xy)$  of  $0.07$ mm (9 solutions, refining to 148.17, 150.22).

Q7: Plainly the answer is “yes” - there is a distinct jump between 2 (solution 6, oP) and 87 (solution 7, oC).

Q8: Again, the answer should be “yes” - if this is not the case, ask a demonstrator to help!

-----

### Example 2

Q1: For most people, the indexing will fail completely, and you will get a pop-up message suggesting that you try again with

1. a larger or smaller longest cell edge
  - note that Mosflm has chosen a value based on the images here - a bug means that it has not been printed in the entry box, but the value is  $\sim 430$ Å (you can see this if you go to the “History” task, click on the “Log” tab and scroll the “light green on dark green” window to the start of the autoindexing output). It's obviously much too big for these images, but not the real cause of the problem in this case.
2. using more or fewer reflections
  - since  $\sim 300$  have been used in this example automatically, this probably isn't a problem
3. using more and/or different images

- this may be a problem, but without examining the images you can't tell at this stage
4. checking your direct beam position carefully
- this is the most common problem and should be addressed first

Q2: None of the beam search solutions really help here. You might find that one or more may seem to help, but none are actually correct. There is a slight indication that the longest cell edge might be around 140Å, since this comes up in a number of solutions, but since none are correct, it's hard to place much faith in this.

Q3: The images should index now; there is a reasonably large jump in penalty from ~15 to ~40, and there are prediction (i.e. measurement) boxes close to some of the spots, but close is not good enough - you want to measure the spots, not the background close to them! The  $\sigma(xy)$  is huge for any of these solutions - >2.0mm, so you should be very suspicious.

Q4: (i) using only image 1, you should get an orthorhombic cell with dimensions 29.1, 32.9, 136.4Å and  $\sigma(xy)$  of 0.30mm - this predicts pretty well for the early images, but not for the later images. If you estimate the mosaicity, it gives an odd looking graph and a value of around 0.2°.

(ii) using only image 59, you will get a pop-up saying

Refinement of indexing solution failed, sorry. unstable refinement.

However, the cell is similar to that found using only image 1 (29.1, 32.9, 137.1Å,  $\sigma(xy) = 0.36$ mm for the oP solution). This solution predicts for the later images but not for the earlier ones. The mosaicity estimates at around 0.55°.

Q5: If you refine the cell using the solution from image 59 (which is most likely to index - sometimes image 1 gives a solution, sometimes not - it depends on exactly what you have done beforehand!), in the first cycle you may see that the initial images do not give a clear spot profile in the "central profile" plot (the top right hand window in the "Cell Refinement" task).

Q5: You should notice a big jump in the "positional refinement" graphs at image 50. There is no corresponding jump in the "postrefinement" graphs, because these are calculated using good data from several adjacent images, whereas the positional refinement uses strong reflections from individual images.

Q6: Image 50 has strong ice rings - the specific reason in this case is unknown, but it may have been caused by a blob of ice falling off the tip of the Cryostream.

The  $I/\sigma(I)$  bar chart for image 50 is quite distinct compared to those for the other images.

-----

### Example 3

Q1: The beam centre cross is in completely the wrong position.

You should notice that there are strong intense arcs on the image - these are partial ice rings, and are normally a very bad thing. However, in this case it is possible to use them to get a reasonably accurate position for the beam centre. You have to be a little careful to make sure you choose arcs that are part of the same ice ring - there are three strong low resolution arcs which are unambiguously from the same ring, but these do not give an accurate beam centre. Towards the edge of the image, there are two obvious arc and a third that is less obvious, and these do give a good beam centre.

Q2: You might find that indexing appears to work, but the predictions will not match the spots for either image 1 or image 90.

Q3: You should find that (using the correct beam centre of 81.04 80.71mm), you can index with image 1 and also with image 90; in each case, the predictions match the image they were indexed from, and iMosflm's chosen solution is tP with the right cell dimensions for tetragonal hen egg white lysozyme (HEWL). However, image 90 will not have the right predictions for the solution from image 1, and vice-versa. Also, the values of  $\sigma(xy)$  will be reasonable

Q4: You should find that the result from indexing with both images together and "reverse phi" should agree well with both the solutions from using each image individually (and no "reverse phi") and with the expected cell dimensions for tetragonal HEWL.

-----

#### Example 4

This example is almost impossible to index with versions of iMosflm before 7.2.0 unless you are a real expert or Mosflm developer! The integration is also difficult, and presents real problems. If you have access to newer versions (including 7.2.0), however, recent improvements in both Mosflm and iMosflm make the whole process trivial.

-----

#### Example 5

To emphasise the point that the most common problem in indexing is that the beam centre in the image header is incorrect, this example also has the wrong position in the headers.

Q1: The predictions don't match the spots well.

Q2: The beam search gives a good beam centre (114.88, 115.94mm) which allows indexing to proceed, giving a tetragonal solution with cell dimensions  $a=b=155.8$ ,  $c=250.9\text{\AA}$ , penalty 6,  $\sigma(xy)$  0.10mm. Most spots are predicted.

If you have the wrong beam centre, indexing will appear to work and give solutions with good penalties and  $\sigma(xy)$ , but you will get a very high estimate of the mosaic spread, which will refine to an unrealistic value in cell

refinement.

Q3: No matter what mosaic spread you choose, the low resolution reflections are not all predicted. If you use “the eye of faith” you can convince yourself that there is a second set of weak lunes perpendicular to the main pattern.

Q4: If you are being optimistic, the spots extend to  $\sim 4.5\text{\AA}$  perpendicular to the rotation axis. They extend to higher resolution close to the rotation axis but these reflections are in the cusp region so cannot be measured well.

Q5: Pointless should indicate primitive orthorhombic. Since there is insufficient information to determine the presence of screw axes, Pointless will only suggest the point group P222 and not a space group.

Q6: Aimless should indicate a maximum resolution of around 3.4 - 3.6 $\text{\AA}$ .

Q7: More low resolution reflections will be predicted with a 5 micron mosaic block size, and the number of high resolution reflections will not change significantly.

Q8: The overall statistics for the integration with 5 micron block size are better than those with 100 microns - but the low resolution limit is higher. This latter issue is because more low resolution reflections are overlapped and therefore are not integrated.

Q9: I found that using a 1 micron block size gave the best overall statistics. This parameter is not currently refined by Mosflm so the optimum figure can only be obtained by trial and error.

-----

### Example 6

Q1: I get 19.7% for unique completeness and 0.7% for anomalous completeness.

Q2: 35.4% and 3.3%.

Q3: 41.4% and 5.8%.

Q4: The best range to optimise completeness is  $199 \rightarrow 209^\circ$ .

Q5: No - the two images collected already are at  $-230^\circ (= 130^\circ)$  and  $-221^\circ (= 139^\circ)$ .

Q6: 55.9% and 9.4%.

Q7: Collecting the range between the reference images for crystal 4 will give 55.2% and 8.6%.

This should remind you that collecting data using the “American method” (i.e. shooting first and asking questions afterwards) is not as efficient as thinking about the problem, and using available tools. The strategy tool was developed

for use with fragile samples such as virus crystals, which typically only survive for a few images.

Q8: 53.0% and 8.2% - i.e. these 3° of data contain nearly 3% of the reflections available in this crystal orientation.

It is possible to recalculate the strategy for different ranges and different crystals in other ways, e.g. if we want to repeat the calculation for xtal4, we can delete the current calculations then use Auto-complete again;

- click on "xtal4\_001" in the top pane to highlight that entry,
- use right mouse to delete it
- Select Auto-complete and enter the desired rotation, then OK and a strategy will be calculated.

Note: When selecting Auto-complete, be sure that the matrix file (top line of pop-up window) is correct (this can be chosen from the drop down menu).