version 18/01/17

iMosflm BAG training - 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.

For this tutorial, we suggest you start from ccp4i2 (1c below)

(1a) Start iMosflm from the command-line

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

When started from the command line, iMosflm will create its output files in the starting directory

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:

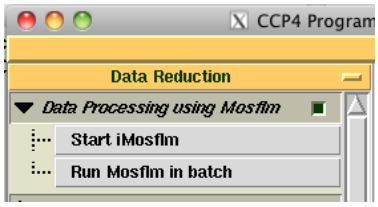
```
mkdir MyProject
cd MyProject
imosflm
```

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"

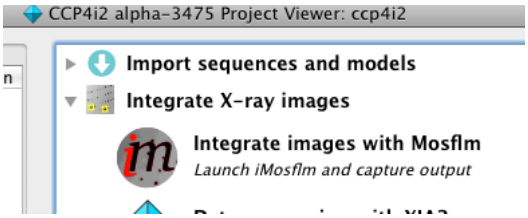


Output files will be created in the project directory

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

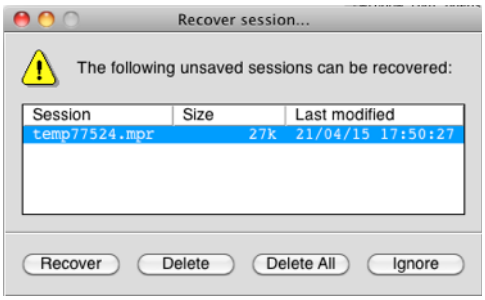
select "Start a new crystallography project"
in the new window that opens, select your home directory and then click on the folder icon with the star on it (near the top right of the window) and create a new folder for your work. Create a new project for your work. A new window opens -

- 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".



Output files will be created in the project job directory CCP4_JOBS/job_N/

In each case, 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.

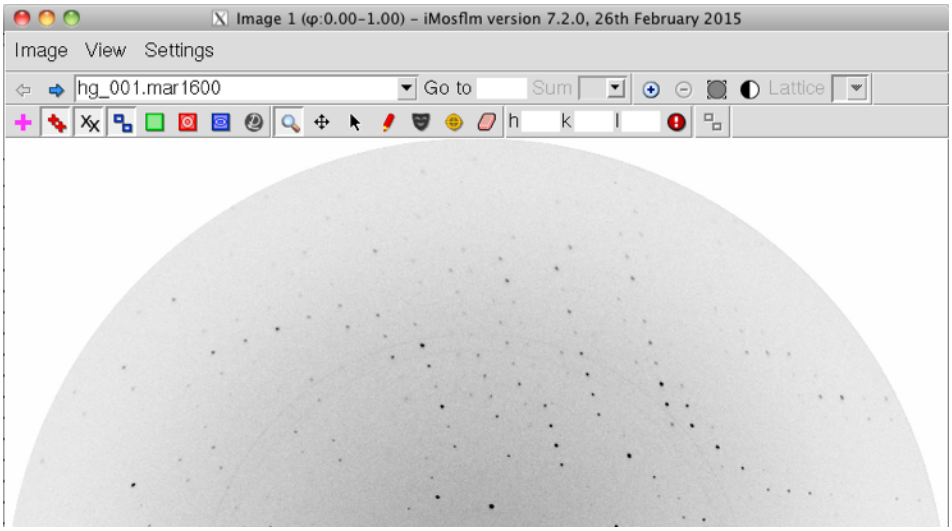
Double-click on the image "hg_001.mar1600"; this will load all the images with the same “template”.

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 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.

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 << 20) and the "best bad" solution (with a penalty >> 20).

iMosflm will refine each of the solutions with a penalty <= 50; good fits will have a σ(xy) value very close to that for the triclinic solution.

iMosflm chooses the solution with a good penalty and small σ(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/σ(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.

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.


Integration

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 symmtery 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.

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/σ(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° 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) More advanced use of iMosflm

These exercises works through some examples of datasets which display relatively common pathologies. See separate file for details (mosflm-examples.pdf)
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