

iMosflm examples for BAG training

READ THIS FIRST

If you have not used iMosflm before, do section (1) below. Do not try either section (2) or section (3).

If you have used iMosflm, but only on straightforward datasets, do section (2). Do not try (1) or (3).

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

(1) Introduction to iMosflm

You should work through the tutorial available at

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

using 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

The images are located on the MRC Linux cluster at -

`/lmb/home/harry/test/bag-examples`

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

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° apart, but indexing using only a single image gives good predictions for that image only.

Open up “Settings → Experiment settings → 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.

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 $\sim 3.5\text{\AA}$. 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?

(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?

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 $\sim 0.67\text{mm}$. The detector is a Mar CCD with square pixels of size 0.0732mm - so for a good, high resolution dataset, you would hope to see a $\sigma(xy)$ of $\sim 0.03 - 0.04\text{mm}$ for the correct solution. Since the triclinic basis solution (the first one in the list) has a $\sigma(xy)$ of $\sim 0.52\text{mm}$, 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 $\sim 2.2\text{mm}$ - one has a $\sigma(xy)$ of 0.05mm (8 solutions refining to 150.05, ~ 150.37), and the other has a $\sigma(xy)$ of 0.07mm (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 $\sim 430\text{\AA}$ (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 ~ 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 $140\text{\AA}$, 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 - $> 2.0\text{mm}$, so you should be very suspicious.

Q4: (i) using only image 1, you should get an orthorhombic cell with dimensions 29.1, 32.9, $136.4\text{\AA}$ 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, $137.1\text{\AA}$, $\sigma(xy) = 0.36\text{mm}$ 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.