

Supplementary Tutorial

A total of nine examples illustrating different aspects of data processing referred to in the text are given here. Images for these examples can be downloaded from www.mrc-lmb.cam.ac.uk/harry/imosflm/examples. The program behaviour described here, and the numerical values for cell dimensions, mosaicity, indexing rmsd values etc. were obtained by running the Linux version of iMosflm version 7.2.1. Different numerical values and possibly different behavior may be seen for other versions of the program or from running the program on other platforms (Windows, OSX).

Example 1. Reverse phi direction and incorrect beam coordinates

The first four 1.0° oscillation images (numbers 1-4) and image 90 are provided. Displaying the direct beam position (magenta cross) on image 1 shows that it is displaced by over 20 mm from the centre of the image. The selection tool can be used to “drag and drop” the direct beam to the centre of the circular region of the backstop shadow (coordinates ~80.8 mm in X and Y). Indexing with images 1 and 90 (selected by default) gives a P1 solution with an rmsd ~1.1 mm. Estimating the mosaicity gives a value > 2° and the resulting prediction is very poor (Fig. S1a). Indexing using only image 1 results in the correct tetragonal solution ($a=78.5$ Å, $c=37.4$ Å) with an rmsd of 0.12mm. Estimating the mosaicity now gives a value of 0.7° and the prediction matches the image very well (Fig S1b). However, the prediction becomes progressively worse as images 2, 3, and 4 are selected in the image display. This is a clear indication that the goniometer is rotating in the opposite direction to that conventionally adopted. Selecting the “Reverse direction of spindle rotation” checkbox in Experiment settings and indexing with images 1 and 90 gives the correct tetragonal solution with an rmsd of 0.13 mm. The prediction now matches well for all images.

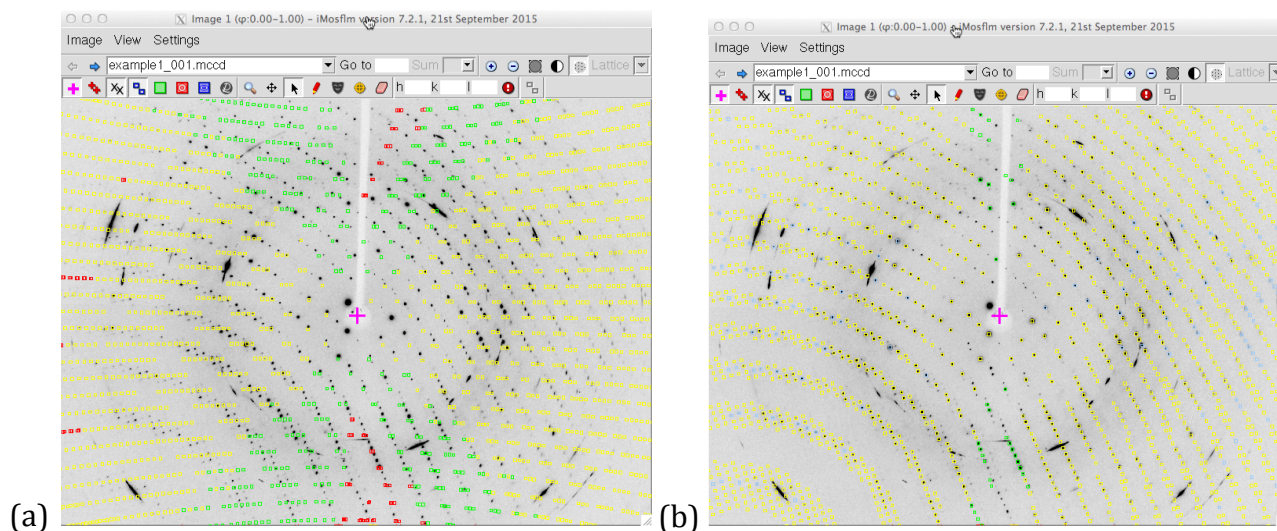


Figure S1. Incorrect indexing solution due to a “reverse-phi” goniometer. (a) Indexing using two images results in a prediction (yellow boxes) that is in very poor agreement with the diffraction spots. (b) Indexing using only the first image results in an excellent prediction. Reflections shown as green boxes are considered too having too wide a reflection range (>5°) and will not be integrated. Red boxes represent spatially overlapped reflections. The magenta cross shows the direct beam position. The intense arcs are due to diffraction from ice crystals.

Example 2. Elongated spot shape

Two images separated in ϕ by 90° are used for indexing. An orthorhombic solution ($a=48.5$ Å, $b=52.9$ Å, $c=64.0$ Å) is suggested, but the rmsd (0.99 mm) is unusually large. However, examination of either image shows that the spots are very elongated with the stronger spots being up to 20 pixels (3.4 mm) in width (Fig. S2). In spite of this, there are very few cases where two spots have been found within a single actual spot. Mosaicity estimation gives a value of 0.5° and the resulting predicted reflections match the shape of the lunes of spots on the image, suggesting that this indexing solution is correct and the high rmsd is due to the error in locating the true spot positions due to their elongated shape. A direct beam search also suggests that the original direct beam coordinates are correct, confirming the indexing solution. Many reflections at low resolution are not predicted with the estimated mosaicity. While this may be partly due to small errors in the crystal orientation derived from the indexing, it does suggest that the mosaic block size (100 microns) is too large. Decreasing the mosaic block size to 1 micron gives a greatly improved prediction at low resolution.

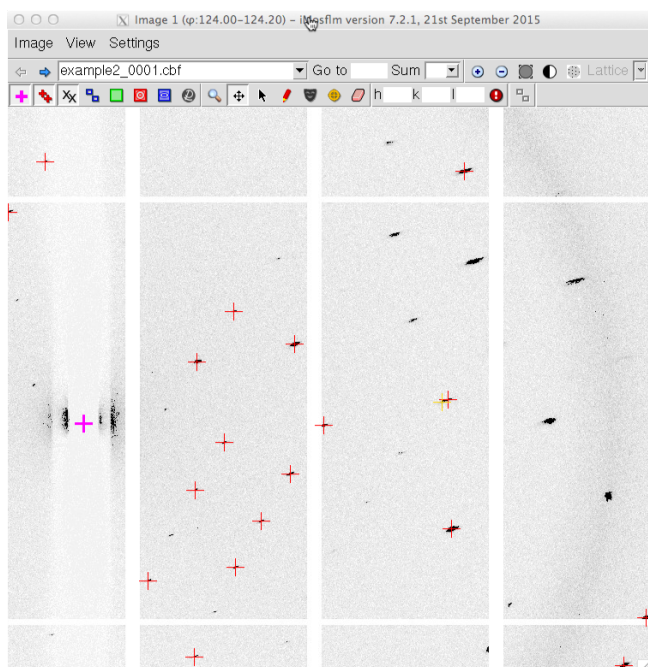


Figure S2. High positional residuals in indexing due to elongated spot shape. Reflections used in indexing are overlaid with a red cross, while those that are below the intensity threshold are overlaid with a yellow cross. Some diffraction spots are more than 20 pixels (3.4 mm) in length, although the very longest spots have been rejected by the spot finding routine. There is one case where both a red cross and a yellow cross have been overlaid on the same diffraction spot, but only coordinates of the red cross will be used in indexing.

Example 3. Incorrect but plausible direct beam coordinates

Three contiguous 0.5° images are available. These images have a numerical suffix (0001, 0002, 0003) that is produced by some beamline software. To be able to see these files in the file browser when adding the images, select “Numbered files” or “All files” as the “File type”. The initial direct beam coordinates are $X=149.1$ mm, $Y=149.9$ mm. Indexing using images 1 and 3 results in a P1 solution with a large rmsd (>0.5 mm), which is much larger than expected for these images, which have small well-defined spots. After estimating the mosaicity ($\sim 0.5^\circ$) the predicted reflections do show the correct pattern of lunes (Fig. S3a), but many spots are poorly predicted (Fig. S3b). Performing a direct beam search gives a set of solutions with refined beam coordinates of $X=150.1$ mm, $Y=150.4$ mm and an rmsd of 0.05 mm and another set of solutions with refined coordinates 148.2, 150.2 mm and an rmsd of 0.07 mm (the beam coordinates in the image header are $X=149.1$ mm, $Y=149.9$ mm). Selecting (double click) one of the solutions with the lower rmsd of 0.05 mm gives an orthorhombic solution ($a=66$ Å, $b=117$ Å, $c=190$ Å) with an rmsd of 0.05 mm. Estimating the mosaicity results in a value of $\sim 0.5^\circ$. Integrating the 3 images and running QuickSymm confirms that the symmetry is

orthorhombic with a high degree of confidence (probability ~ 0.8) in spite of the fact that this is based on a total rotation angle of only 1.5° (this is atypical, and a result of the particular section of reciprocal space recorded on these images).

If one of the solutions with refined coordinates $X=148.2$ mm, $Y=150.2$ mm and an rmsd of 0.07 mm is selected, this gives an orthorhombic solution with very similar cell dimensions and an rmsd of 0.08 mm. Integrating the three images using this solution appears to work well, but running QuickSymm gives a monoclinic solution (probability ~ 0.6) because the images have been miss-indexed.

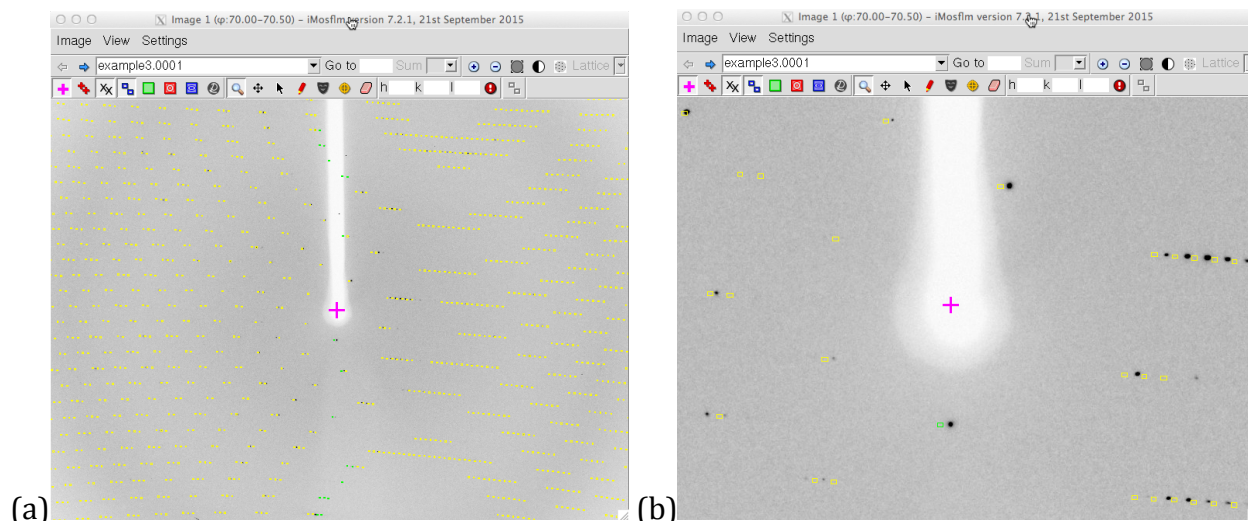


Figure S3. Correct pattern of lunes for an incorrect indexing solution. (a) Small errors in the direct beam coordinates can result in a reflection prediction that matches the pattern of the lunes well. (b) Closer examination of the prediction in the central region of the detector shows that individual spots are not correctly predicted. The initial direct beam coordinates were plausible insofar as the direct beam position (magenta cross) is close to the centre of the backstop shadow, but in fact they were in error by 1.0 mm in X and 0.5 mm in Y.

Example 4. Finding very weak spots

The initial images (example4_1_001.mar2300, example4_1_001.mar2300) are one degree oscillation images from a Mar Research image plate detector, collected using a laboratory source. The default values for the spot finding parameters are different for laboratory sources (recognized by the wavelength) and synchrotron sources, and sufficient spots are found on these images to allow indexing. However, additional spots can be found by reducing the “Threshold $I/\sigma(I)$ ” value to 2.5 and the “Spot rms variation” value to 0.5. Discrimination against finding noise is provided by the “Minimum pixels per spot” parameter that defaults to 30. In difficult cases, adjusting these three parameters can result in more correct spots being found, but it is important to guard against finding noise, so the found spots should be checked on the image display. Additional information about the spot finding on a particular image can be obtained by highlighting the image in the indexing pane, pressing the right mouse button, and then selecting the resulting “details” button. This will give the total number of spots found and the number rejected by various criteria. This can be helpful in deciding what spot finding parameters to change.

Varying these parameters and, in addition, decreasing the $I/\sigma(I)$ threshold for including spots in indexing, allows indexing using the 3 images in the second test set (images example4_2_001.osc etc) which were collected using a Rigaku image plate detector. Note that a

new run of iMosflm is required to handle the new detector type. The indexing based on a total of ~40 spots from these three images gives an orthorhombic C-face centred solution ($a=44\text{ \AA}$, $b=86\text{ \AA}$, $c=165\text{ \AA}$) with an rmsd of only 0.08 mm. These images can be integrated, but there is insufficient data to determine the point group with QuickSymm.

Example 5. Poor spot shape and highly mosaic crystal

Two one degree oscillation images are provided, with starting phi values of 0° and 92° . iMosflm will select images 1 and 2 for indexing, but the indexing will fail. Examination of image 2 reveals badly split spots, in this case due to severe radiation damage, and several cases where the spot finding has identified the two parts of a split spot as being two separate spots (Fig. S4). By setting the minimum spot separation (spot finding tab of Processing options) in X and Y to 1.5 mm and re-running the spot finding on image 2 (by clicking the red cross under the “Find” heading for image 2 in the Indexing pane) these “doubled” spots are eliminated. Repeating the indexing gives the correct C-face centred monoclinic solution ($a=139\text{ \AA}$, $b=67\text{ \AA}$, $c=104\text{ \AA}$, $\beta=97^\circ$) with an rmsd of 0.3 mm. An alternative way to index the images, without changing the spot finding parameters, is to increase the $I/\sigma(I)$ threshold for spots used in indexing from 20 (the default for these images) to 60 or above. There are two reasons why this is helpful. Firstly, it eliminates some (but not all) of the doubled spots found on image 2. Secondly, the crystal is very mosaic ($>2^\circ$), and this is apparent from the absence of clear lunes in the diffraction pattern for image 1. In these cases, setting a high threshold will generally make the lune appearance clearer in the selected spots (indicated by red crosses overlaid on the diffraction image). This in turn will reduce the errors when spot positions are mapped back to reciprocal lattice vectors, improving the likelihood of successful indexing. This effect is particularly clear if indexing is attempted using only image 1. Setting the $I/\sigma(I)$ threshold to 100 results in correct indexing, while the indexing fails for other values close to 100 (e.g. 80, 90, 110, 120). Here there is a trade off between setting a high threshold (to improve the apparent lune appearance) and keeping a significant number of spots with a good distribution in reciprocal space. This example also makes it clear that using two images well separated in phi rather than a single image results in improved robustness in indexing. Images 1 and 2 together will give the correct indexing for threshold values between 50 and 140.

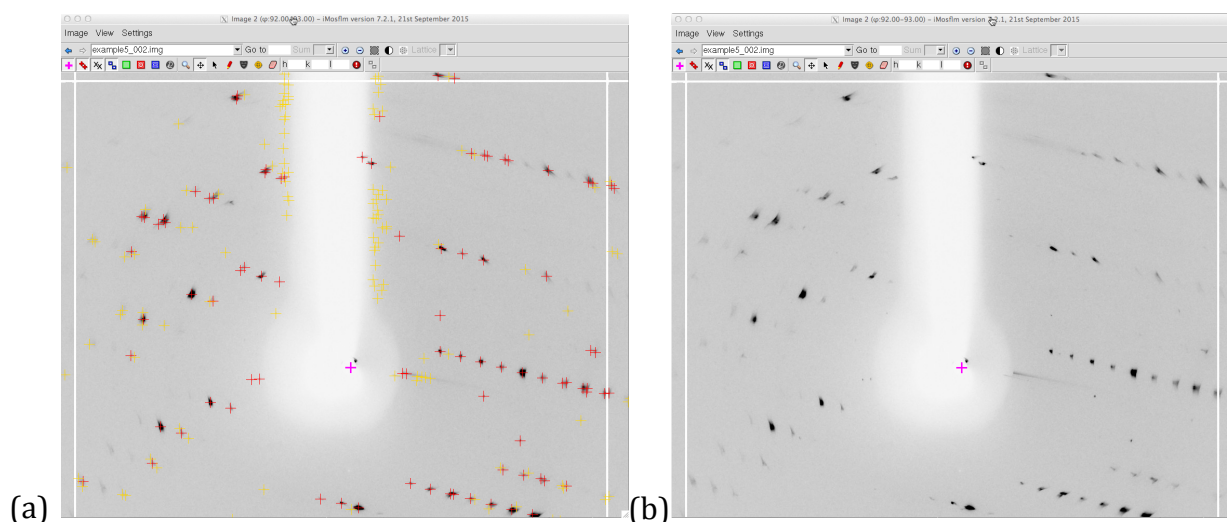


Figure S4. Split spots can result in errors in spot finding. (a) Radiation damage has resulted in badly split diffraction spots. (b) The spot finding results in two (or more) found spot positions (red crosses) for a single (split) diffraction spot. This can be avoided by adjusting the spot finding parameters (see Example 5).

Example 6. Incorrect beam coordinates, poor prediction due to change in crystal orientation during data collection

The first 10 images and image number 360 (corresponding to a 90° phi rotation from the first image) are provided. The oscillation angle is 0.25°. Indexing using images 1, 360 suggests a P1 solution with an rmsd of 0.48 mm, much higher than expected. A direct beam search gives a solution with an rmsd of only 0.10 mm, for beam coordinates X=103.2 mm, Y=105.5 mm (starting values were X=104.2mm, Y=104.8 mm). Selecting this solution gives a tetragonal cell, $a=170 \text{ \AA}$, $c=321 \text{ \AA}$, with an rmsd of 0.10 mm, but although the appearance of the lunes of the predicted reflections seems correct, the predicted reflections do not match the spot positions well. Mosaicity estimation then gives a mosaicity of 2.4°, which, from the widths of the lunes, is clearly far too large (Fig. S5). This is an example of incorrect mosaicity estimation due to a poor indexing solution. Indexing using image 1 alone gives an rmsd of 0.07 mm, the prediction is much closer to the observed spots and the mosaicity is estimated to be 1.0°. The poor prediction when using both images is because the crystal orientation changed during data collection (for reasons that are not clear). Integration of the full dataset shows that two of the missetting angles change by 0.7° over images 1- 360. The long c-axis (321 Å) results in a spot separation of only ~1 mm. Because the direct beam coordinates changed by 1.0 mm in X and 0.7 mm in Y, there is a significant risk of miss-indexing. Integrating images 1-10 and running QuickSymm confirms the point group as P422 with a high degree of confidence, verifying that the direct beam position is correct.

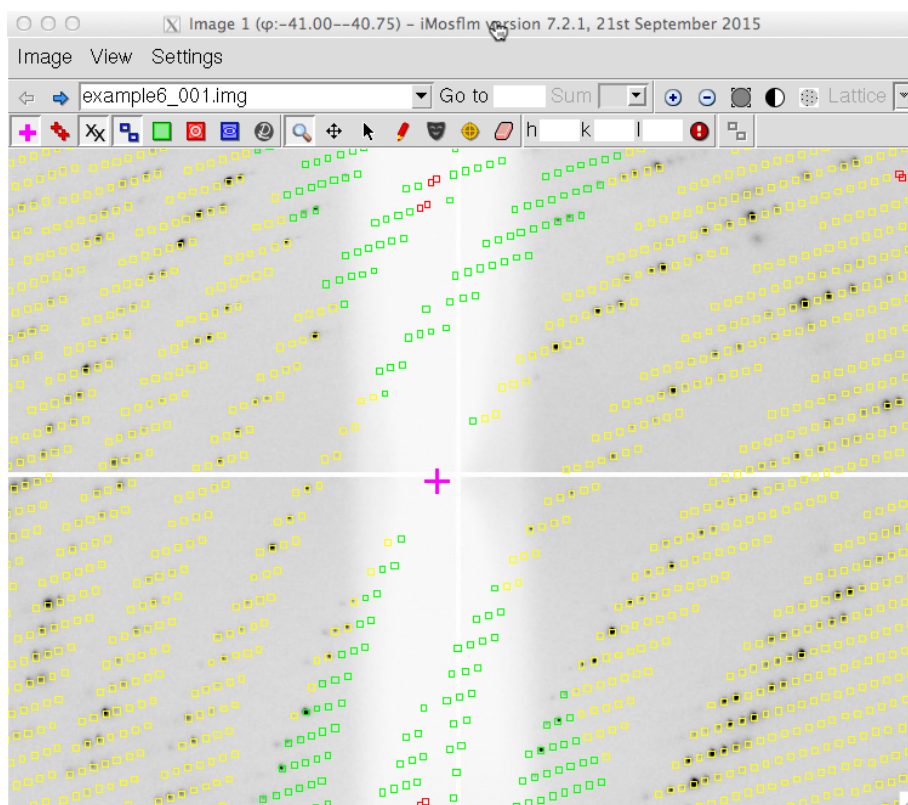


Figure S5. Overestimate of crystal mosaicity. A genuine change in crystal orientation between the two images used for indexing resulted in a poor indexing solution, which in turn resulted in an overestimated value for the mosaic spread (2.4°). This is apparent in the widths of the lunes for the predicted reflections that are far greater than the actual lune widths.

Example 7. A multilattice dataset

The dataset consists of 120 images (1° oscillation angle) that have two lattices present. The images are rather weak, so it is not easy to detect the presence of two lattices by visual inspection. Because the multilattice integration feature is still under development, it is important that the exact sequence of processing described below is followed closely.

Indexing using images 1 and 90 (the default) gives a C-face centred monoclinic cell with a high rmsd (0.67 mm). This could be due to incorrect direct beam coordinates, but in this case doing a direct beam search does not result in a significantly lower rmsd. Selecting the multilattice icon in the indexing pane results in two lattices being found, but lattice one (C-face centred monoclinic) still has a rather high rmsd (0.36 mm) while lattice two (orthorhombic) has a much lower rmsd (0.10 mm). This is an example of the sensitivity of the multilattice indexing to the maximum cell dimension, which in this case is 672 Å. If this is reset to 300 Å and the indexing repeated, both lattices are orthorhombic ($a=73.6$ Å, $b=124.3$ Å, $c=128.5$ Å) with rmsds of 0.20 mm and 0.10 mm. Very similar solutions are found if the maximum cell dimension is set between 200 Å and 400 Å. The angle between the two lattices (denoted "Split") is 3.4°. Estimating the mosaicity gives a value of 0.5°. The predicted reflections for the two lattices can be displayed separately using the lattice selector in the image display window, and this confirms that there are two lattices present. Reflections where the two lattices overlap are shown in magenta (Fig. S6). Selecting the multilattice icon in the image display window shows the predictions for lattice one in blue and for lattice two in yellow. Together, these predictions account for all the spots visible in the image, again confirming the presence of two (but not more) lattices. It is important to confirm that the multilattice indexing is correct before attempting the next stage in the processing.

The maximum resolution is set to 2.5 Å after examination of the images. The Cell Refinement can then be carried out on lattice one, which only gives small changes in the cell parameters. This is followed by Integration of lattice one. The images for integration are automatically assigned and it will be noticed that images 10, 20 and 45 are missing (simply because there were data transfer errors when storing the images). This does not affect the integration. During integration, selecting the "Lattice overlaps" parameter in the lower left panel will give a plot of the number of reflections flagged as being spatially overlapped between the two lattices. The next step is to return to the Cell Refinement step and select lattice 2 from the lattice selector, and refine the cell for lattice 2. This second lattice can then be integrated. Note that when working with multiple lattices the default name for the output MTZ file is appended with the lattice number, so this does not need to be updated when integrating lattice two. When integration of lattice two is complete, selecting QuickScale will run three programs, FECKLESS, POINTLESS and AIMLESS. FECKLESS will merge the data from the two lattices checking for any inconsistencies, POINTLESS will determine the most probable Point Group and Space Group and AIMLESS will scale and merge the data. The statistics in AIMLESS are based only on "singletons", that is, reflections that are not overlapped. Overlapped reflections are integrated as a single spot and this summed intensity is written to the MTZ file, but at present no downstream programs can make use of these intensities and they are not present in the merged MTZ output file written by AIMLESS.

This example works well because there is a reasonably large angle between the two lattices (if the angle is two degrees or less, almost all the reflections will be spatially overlapped) and because both lattices are present on all images and there is not a large difference in the intensity (i.e. diffracting volume) of the two lattices.

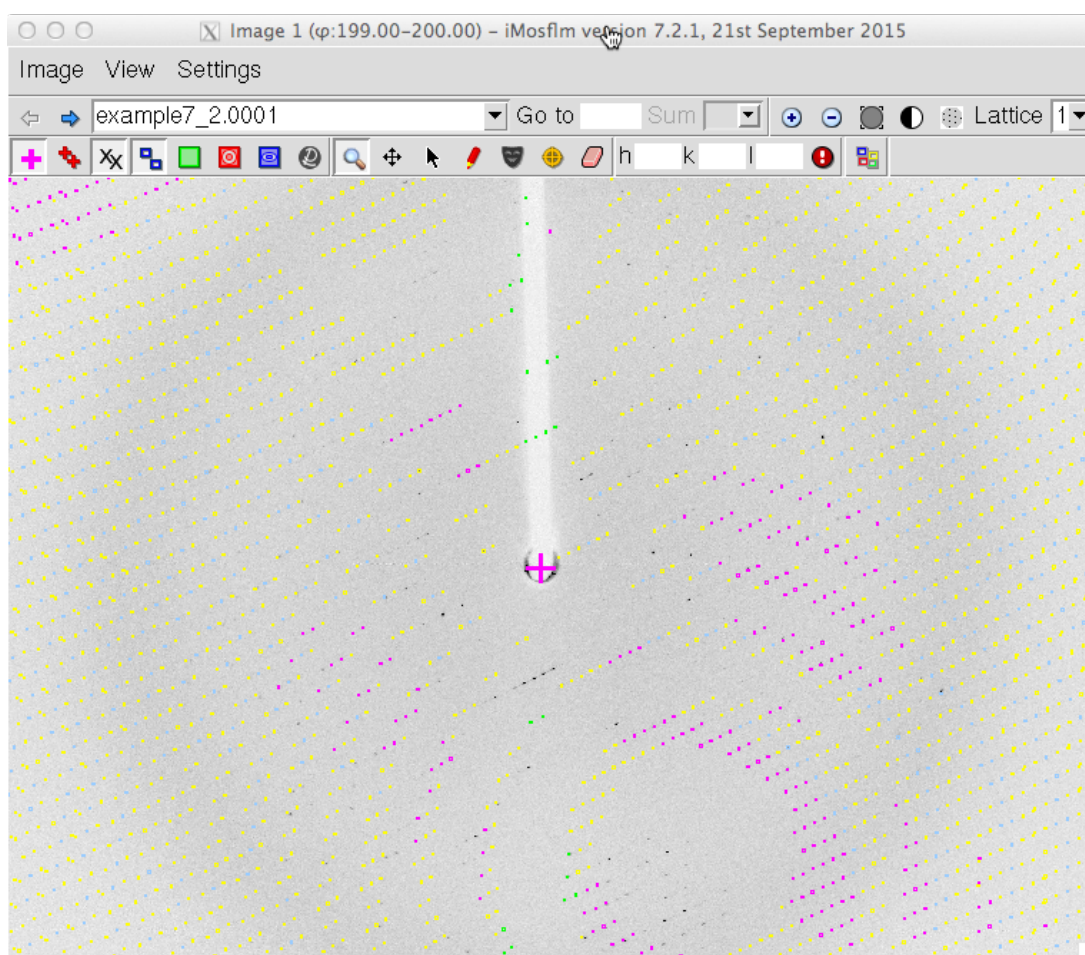


Figure S6. Multiple lattice indexing. Predicted reflections are shown for lattice 1 of the two lattices present on the image. Reflections that are overlapped by reflections of the second lattice, which are primarily at low resolution, are shown in magenta. Diffraction spots that are not predicted are from the second lattice.

Example 8. A challenging dataset with one very large cell dimension

Six hundred images are provided corresponding to a total rotation of 90° . Indexing using image 1 and 600 (the default) gives an orthorhombic solution with an rmsd of 0.39 mm, larger than expected for the very small spot size. The input beam coordinates (X=206.1 mm, Y=214.5 mm) refine to X=205.3 mm, Y=214.6 mm. The indexing solution does predict the observed spots, but also predicts rows of reflections between the observed spots, indicating that the unit cell is too large.

A direct beam search gives two possible solutions, the first at X=205.3 mm, Y=213.8 mm and the second at X=205.3 mm, Y=215.1 mm, both with an rmsd of 0.20 mm and the same cell parameters. The difference in the Y coordinate corresponds to the separation of the spots in the Y direction (vertical in the image display). At this stage it is not possible to distinguish which solution is correct. Selecting the first solution (Y=213.8 mm) gives a hexagonal cell $a=105.2 \text{ \AA}$ $c=624.6 \text{ \AA}$ and rmsd 0.18 mm. To test if this solution is correct, the first 60 images can be integrated after setting the resolution limit to $6 \text{ \AA}$ (to speed up the integration). Because the diffraction only extends to $\sim 4 \text{ \AA}$, the Cell Refinement will not improve the accuracy of the cell dimensions, so this step can be omitted for this dataset. During integration, the tilt and twist are unstable for the first 6-7 images but then stabilize. Running QuickSymm on the resulting data gives a correlation coefficient of only 0.17 for the Identity operator (Friedel pairs), showing that the pattern is miss-indexed. Selecting the alternative solution (Y=215.1

mm) gives the same unit cell as the initial solution and an rmsd of 0.19 mm. At this point it is necessary to manually update the Y beam coordinate to 215.1 mm in the indexing pane, as this will have reverted to the original value. Repeating the integration of the first 60 images with the updated Y beam coordinate and running QuickSymm gives a correlation coefficient of 0.98 for the identity operator, and in addition a correlation coefficient of 0.98 for reflections related by a 2-fold axis in the [1 1 0] direction, confirming that the indexing is now correct. The ambiguity in the direct beam coordinates could be avoided by noticing that there are distinct ice rings on image 600. Using the circle fitting option in the image display window gives direct beam coordinates X=205.6 mm, Y=215.4 mm, which refine to the correct beam coordinates when indexing images 1 and 600.

Close examination of image 1 shows that adjacent reflections along the c-axis are not completely resolved. The minimum spot separation allowed before reflections are flagged as spatial overlaps (and not integrated) is determined from the size of the average spot profile for spots in the central region of the detector. This in turn depends on the profile tolerance parameters, which default to 0.02 for low-resolution reflections and 0.03 for higher resolution. Increasing the profile tolerance parameters will result in a decrease in size of the peak region of the measurement box, and a corresponding decrease in the minimum spot separation. If the closest possible spot separation determined from the diffraction geometry (unit cell dimensions, wavelength, detector distance) is less than a factor of 1.25 times the minimum spot separation determined from the average spot profile, the profile tolerance values are automatically increased. This will result in a smaller minimum spot separation and reduce the likelihood of reflections being flagged as spatial overlaps. In this case, the profile tolerance values are increased automatically to 0.04/0.05 as can be seen in the Advanced Integration tab of the Processing options. (Note that the check box that allows this option to be turned off is not fully implemented in this release of the program). The profile tolerance will only be increased to a maximum of 0.05/0.06 by this procedure. In very challenging cases larger values can be set manually.

Prior to integrating the whole dataset, the backstop shadow needs to be masked and the mosaic block size adjusted. Looking at the predicted low-resolution reflections, for example on image 1, it is clear that some spots are not being predicted. In order to predict these spots, the mosaic block size needs to be set to ~4 microns. This value is chosen so that ideally all low-resolution spots are predicted, but there are not too many reflections predicted for which there are no obvious spots. The mosaic block size should only be adjusted after an initial trial integration of the first few degrees of data has been carried out, so that the crystal orientation has been refined, as the orientation determined from the indexing step will not be as accurate as that determined by post-refinement during integration.

The resolution limit can now be set to 4 Å (judged as the diffraction limit based on inspection of the images) and integration of all 600 images can be carried out. Running QuickScale, the results from AIMLESS estimate the resolution limit as 4 Å (Mean $(I/\sigma(I))$ is 2.1 in the highest resolution bin), ie the resolution limit of the integration. In these circumstances the resolution limit would be increased to (say) 3.7 Å and the integration repeated. In this case, caution is required in setting the high-resolution limit because of the very poor spot shape at high resolution as seen in the standard profiles (Fig. S7), especially for the later images.

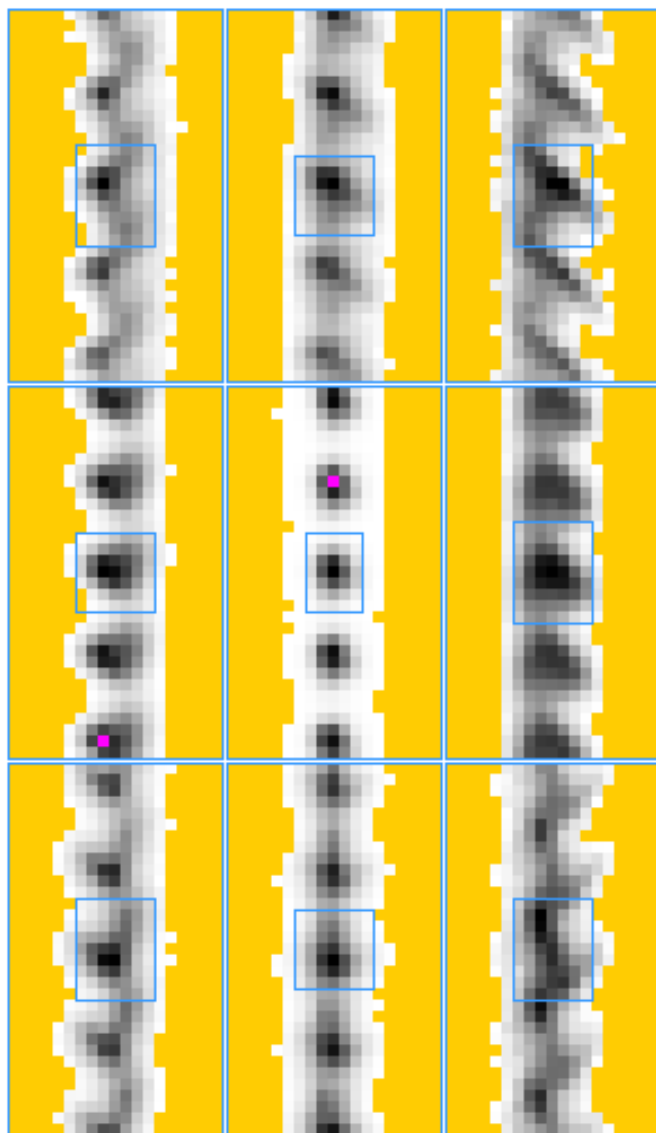


Figure S7. Poor spot shape as shown in the standard profiles. Defects in the crystal result in very split and streaky spots, especially at higher resolution. While these are visible on close inspection of the images, they are very clear in the appearance of the standard profiles. This will have an adverse effect on the quality of the high resolution data.

Example 9. An example of pseudosymmetry

Thirty degrees of data recorded with an oscillation angle of 0.2° are provided. Indexing (using images 1, 360) gives a monoclinic solution ($a=108 \text{ \AA}$, $b=270 \text{ \AA}$, $c=467 \text{ \AA}$, $\beta=90.4^\circ$) with an rmsd of 0.14 mm. While the prediction matches the pattern of the lunes well and the rmsd is quite small, close examination of image 1 shows that far more reflections are being predicted than are present, so the suggested unit cell is larger than it should be. This is usually the result of an incorrect direct beam position. A direct beam search gives a set of solutions with an rmsd of 0.12 mm. Selecting one of these gives a monoclinic unit cell ($a=108 \text{ \AA}$, $b=270 \text{ \AA}$, $c=156 \text{ \AA}$, $\beta=90.4^\circ$) with an rmsd of 0.13 mm, which is the correct solution (the original indexing solution corresponded to a tripling of the true dimension of the c axis). However, the orthorhombic solution ($a=108 \text{ \AA}$, $b=156 \text{ \AA}$, $c=269 \text{ \AA}$) has an rmsd of 0.27 mm. This is

significantly larger than the rmsd of the monoclinic solution, strongly suggesting that it is not correct, but for the purpose of illustration, this solution should be selected. Following cell refinement, all 360 images can be integrated without difficulty, but there is a clear indication that the indexing solution is incorrect because the standard profiles are not well centred in the measurement box (Fig. S8). There is also a large change in the detector “Twist” parameter from -0.5° to $+0.2^\circ$ over the first 25 images. However, QuickSymm suggests that the Laue symmetry is orthorhombic with a probability of ~ 0.7 . This is because the crystal is indeed pseudo-orthorhombic. The details of the QuickSymm logfile show that the correlation coefficient is 0.91 for the Friedel operation, 0.90 for one 2-fold axis and 0.79 for the other two 2-fold axes. In an orthorhombic space group, if the correlation coefficient is significantly higher for one 2-fold axis than the other two, this suggests the possibility that the symmetry is actually monoclinic. In this case, running AIMLESS (QuickScale option) results in an Rmerge at low resolution of 10%, far higher than would be expected and strongly suggesting that the true symmetry is monoclinic. If the images are integrated using the monoclinic solution, the standard profiles are correctly centred and the twist parameter only changes by 0.02° over the entire dataset. However, QuickSymm still selects orthorhombic solution. By forcing QuickSymm to accept the symmetry selected in iMosflm (Use the checkbox in the Sort, Scale and Merge tab in Processing options) the Rmerge at low resolution drops to 5.8%. Changing the mosaic block size results in a further reduction. Inspection of the prediction for image 1 shows that there are a significant number of reflections at low resolution that are not predicted. Reducing the mosaic block size from 100 microns to 2 microns gives a much better prediction. This also results in the crystal mosaicity refining to a larger value. The resulting Rmerge at low resolution is less than 2%.

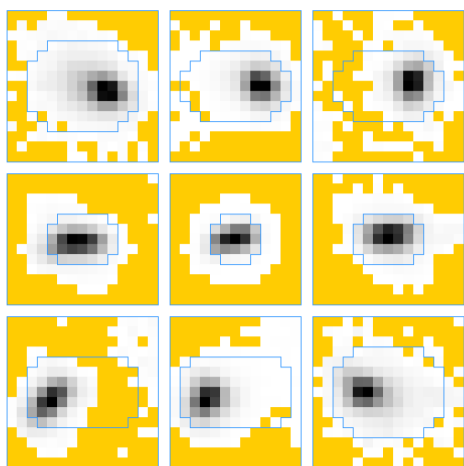


Figure S8. Off-centred standard profiles resulting from an incorrect indexing solution. An orthorhombic indexing solution was selected for a monoclinic, pseudo-orthorhombic, unit cell ($\beta=90.4^\circ$). Integration of the data results in standard profiles that are significantly off-centre in the measurement boxes, especially in the four corners of the detector. See Example 9 for details.