Integrating macromolecular X-ray diffraction data with the graphical user interface iMOSFLM

Harold R Powell, **T Geoff G Battye**, **Luke Kontogiannis**, **Owen Johnson**, and **Andrew GW Leslie**

MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK

Abstract

X-ray crystallography is the overwhelmingly dominant source of structural information for biological macromolecules, providing fundamental insights into biological function. Collection of X-ray diffraction data underlies the technique, and robust and user-friendly software to process the diffraction images makes the technique accessible to a wider range of scientists. iMosflm/ MOSFLM (www.mrc-lmb.cam.ac.uk/harry/imosflm) is a software package designed to achieve this goal. The graphical user interface (GUI) version of MOSFLM (called iMosflm) is designed to guide inexperienced users through the steps of data integration, while retaining powerful features for more experienced users. Images from almost all commercially available X-ray detectors can be handled. Although the program only utilizes two-dimensional profile fitting, it can readily integrate data collected in "fine phi-slicing" mode (where the rotation angle per image is less than the crystal mosaic spread by a factor of at least 2) that is commonly employed with modern very fast readout detectors. The graphical user interface provides real-time feedback on the success of the indexing step and the progress of data processing. This feedback includes the ability to monitor detector and crystal parameter refinement and to display the average spot shape in different regions of the detector. Data scaling and merging tasks can be initiated directly from the interface. Using this protocol, a dataset of 360 images with ~2000 reflections per image can be processed in approximately four minutes.

Keywords

Introduction

X-ray crystallography remains the predominant technique for the determination of high-resolution structures of biological macromolecules. Advances in molecular biology, crystallization protocols, data collection techniques and structure solution algorithms have led to a dramatic increase in the number of structures being determined, as demonstrated by the recent achievement of 100,000 structures deposited in the Protein Data Bank[1]. Optimal processing of the diffraction data is a crucial step in the structure determination process, as all subsequent computational steps depend on this.

Development of the protocol

The MOSFLM program is designed to process diffraction data collected with 2D area detectors using the rotation method[2]. It was originally written to process data collected on X-ray film, and early versions were able to run on computers with as little as 32kB of memory, with the whole operation being broken down into a series of individual steps[3]. Subsequently the program was extensively restructured and extended to handle image plate, charge-coupled device (CCD), complementary metal–oxide–semiconductor (CMOS) and pixel array detector data[4]. More recently a greatly improved graphical user interface, iMosflm[5], was developed with the particular aim of guiding less experienced users through the data processing procedures, while retaining a considerable level of control for more experienced users. iMosflm sets default values for the large number of parameters that influence processing, many of which are assigned values on the basis of information in the image file headers (wavelength, detector distance, direct beam coordinates, start and end phi values) and an initial assessment of the images to be processed. However, these defaults can be overridden via a number of menus in the user interface, allowing improved processing of particularly challenging cases.

MOSFLM has been widely used by the macromolecular crystallographic community for over 30 years. For example, more than 14,000 depositions in the Protein Data Bank have acknowledged use of MOSFLM for data processing. Specific examples include membrane proteins such as a photosynthetic reaction centre[6], cytochrome c oxidase[7], G protein coupled receptors[8,9], a bacterial ATP synthase[10] and a voltage-gated sodium channel[11], histone complexes[12], molecular chaperones[13], RNA polymerase II[14], T7 RNA polymerase–T7 promoter complex[15] and RNA structures such as the hammerhead ribozyme[16]. More recently, MOSFLM has also been used to process diffraction data collected by serial femtosecond crystallography using an X-ray free electron laser[17] and for electron diffraction data[18,19].

Diffraction images from all widely-used commercial detectors can be handled by the program (See also Box 1). With relatively slow read-out detectors, such as image plates and some CCD detectors, data are typically collected with an oscillation or rotation angle per image that is greater than the mosaic spread of the crystals (coarse-sliced mode), while with very fast read-out detectors (pixel, CMOS and some CCD detectors), the oscillation angle is usually less than the mosaic spread by a factor of at least 2 (fine ϕ-slicing mode). Both types of images can be dealt with successfully. Although MOSFLM does not employ three-dimensional profile fitting when dealing with fine-sliced data, this approach is used in other

programs (DIALS20, XDS21, d*TREK22) and in some circumstances can provide an improvement in data quality.

Overview of the procedure

While it is possible to use the program in a non-interactive mode (with keywords as input), this requires a reasonably detailed knowledge of the program and therefore only the graphical user interface iMosflm5 will be described here. The graphical user interface is made up of two components; a control panel (Fig. 1), where individual tasks can be selected and the results are being displayed, and an image display panel (Fig. 2). The control panel has six major tasks: Reading images, Indexing, Strategy calculation, Cell refinement, Integration and History. A different panel will be displayed according to the task selected. Within these panels, parameters that influence that specific task can be adjusted and the obtained results will be displayed. The control panel also has a menu bar, with "Session" and "Settings" drop-down menus. The Session menu allows a processing session to be saved, or previously saved sessions to be opened. The Settings menu provides control over parameters of the actual data collection experiment (Experiment settings) or the processing of the images (Processing options).

The image display panel (Fig. 2) allows a detailed examination of the diffraction images (with the usual zoom, contrast and panning controls) and enables overlaying of the spots selected for indexing and the predicted reflections once the indexing has been carried out. Within this panel, inactive or shadowed regions of the detector can be masked out, and resolution limits for spot finding and for integration can be adjusted separately. There is also a circle-fitting tool to determine the direct beam position from ice rings (if present).

MOSFLM writes out a list of integrated reflections, with indices, intensity estimates and their standard uncertainties, spot coordinates and other information in an MTZ file for input to further data reduction programs (POINTLESS26 and AIMLESS27) in the CCP4 suite28.

Details of the processing are recorded in a date-stamped log file, and a date-stamped summary file contains graphical output (similar to that displayed in the iMosflm GUI) that can be viewed with the CCP4 program LOGGRAPH.

Comparison with other methods

One of the principal advantages of the MOSFLM program when compared to other data processing software such as XDS21 and HKL200023 is the design of the graphical user interface (DIALS20 does not yet have a GUI, although this is currently under development). This makes the software very straightforward to use, especially for inexperienced users, while retaining the functionality required for more difficult cases. The strategy option, that calculates the initial rotation angle and the rotation range required to collect a complete dataset, is also very powerful. It is especially useful when multiple crystals are required to complete a dataset, and has the advantage that it takes account of the gaps between the tiles of some detectors (hybrid pixel detectors and some CCDs) when calculating expected data completeness. MOSFLM is also currently the only software that will allow integration of images where multiple lattices are present, taking proper account of the overlap between different lattices. However, MOSFLM has the disadvantage that it does not employ three-

dimensional profile fitting22 which is present in XDS, DIALS and d*TREK. This procedure can lead to an improvement in data quality when data are collected using high frame rate detectors with no readout noise, such as hybrid pixel detectors, that allow data collection in "shutterless" mode with a small oscillation angle per image (fine ϕ-sliced data). This mode of data collection is now common on many synchrotron beamlines. If attempting to measure a very weak anomalous scattering signal for phasing, then using three dimensional profile fitting (rather than two dimensional as in MOSFLM) can make the difference between success and failure in structure solution. The error modeling is more sophisticated in HKL2000 than in MOSFLM (and possibly other software) and in some cases this can also lead to meaningful improvements when attempting to extract a weak anomalous signal for phasing. XDS has a reputation of being very robust when dealing with poor quality images (poor spot shape, very weak diffraction, presence of more than one lattice) although this is difficult to quantify. Finally, MOSFLM and DIALS are both open source, unlike HKL2000 and XDS. MOSFLM, XDS and DIALS are free to academic users.

Materials

Equipment Setup

System requirements

- MOSFLM runs on most widely available desktop computing systems; this includes all Linux platforms of which we are aware, Mac OSX (Intel-based machines) and MS-Windows (all versions from Windows 95 to Windows 8). It is also known to work on several obsolete systems such as SGI Irix and Alpha Tru64 UNIX. At present there is no implementation for mobile computing devices such as those running Android or Apple iOS.

- iMosflm runs on all platforms for which Tcl/Tk version 8.4 is available (a bug in the image display library of Tk 8.4.13 means that this version is not recommended for use with iMosflm).

MOSFLM/iMosflm software and its availability

- iMosflm and MOSFLM are packaged together and distributed according to common installation methods. For all platforms, a .*zip* file can be downloaded from the iMosflm website (www.mrc-lmb.cam.ac.uk/harry/imosflm) and uncompressed in any suitable location.

- For OSX, an *App* containing the whole package can be downloaded and placed in the */Applications* folder.

- For Windows, an installer unpacks the program files into any chosen folder.

- iMosflm and MOSFLM are both freely available, and when obtained from the MOSFLM website are free of charge to all users. They depend on CCP422 libraries, which are free of charge and without onerous licensing restrictions for both academic and commercial users.

- MOSFLM is supplied as pre-compiled executable binary files for machines running Linux (32- and 64-bit), Apple Macintosh OSX (64-bit Intel-based machines) and Windows (32-bit). The executables for Linux and OSX (Intel) are compiled with the Intel *ifort* and *icc* compilers. The executable for Windows is compiled using a cross-compiler version of *gfortran* and *gcc* run on either an Intel-based Macintosh running OSX or a Linux computer.

- The source code of MOSFLM is also freely available from the MOSFLM website, with a *build* script written to simplify compilation. Since MOSFLM is written in standard FORTRAN 77 and ANSI C, it can be built on any computer on which compilers for these languages exist. The compiled object files are linked against the basic FORTRAN and C libraries in the CCP4 distribution to produce an executable. For platforms that support X11 graphics a version of MOSFLM can be built which contains the obsolete ipmosflm GUI, which also needs the CCP4 xdl_view libraries to link against.

- iMosflm is written in Tcl/Tk version 8.4 with the object-oriented extensions [incr Tcl] and [incr Tk]. It uses several other (standard) extensions to the core languages for image display (tkImg), parsing XML responses from MOSFLM (tDom), displaying lists of items (treectrl) and for producing customizable widgets (Iwidgets).

- iMosflm/MOSFLM is also available in the standard CCP4 distribution, and can be launched from both the older the CCP4 graphical user interface, CCP4i, either from the "Program List" menu or in the "Data Reduction and Analysis Task", and the recently released interface CCP4i2 from the "Integrate X-ray Images" task.

- The communication between MOSFLM and iMosflm occurs via a TCP/IP socket. MOSFLM reads commands from iMosflm, and the results of its calculations are returned through the same socket as XML documents. iMosflm parses these documents and updates the appropriate displays (whether text or graphical) accordingly.

- MOSFLM can also be run independently of iMosflm from the command line; this is intended for expert users, for whom comprehensive instructions and lists of commands are published on the MOSFLM website.

MOSFLM/iMosflm documentation—A detailed tutorial that describes all features of the software is available from the iMosflm website (www.mrc-lmb.cam.ac.uk/harry/imosflm). The tutorial can either be followed with images downloaded from the website or with the users' own datasets. The tutorial is also accessible from the Help item on the menu bar of the iMosflm control window.

Procedure

Reading the images

1| Select the "Add images" icon from the Images pane (Fig. 1) and read in the images to be processed by double clicking the filename of any image in the series in the file browser. The image file that was double clicked is now displayed in the image display panel, other images can be selected using the "Go to" entry box.

A subset of images can be selected using shift+click or ctrl+click providing the "Selected files only" checkbox is ticked. This series of images is known as a "Sector". For inverse beam experiments, the two sets of images should have a different filename template (e.g. a different "run" number) and hence treated as separate sectors, or be numbered so that the image numbers do not overlap. The phi values of each image are displayed in the Images pane.

TROUBLESHOOTING

! CRITICAL STEP The direction of rotation of the spindle is reversed compared to the conventional direction on some synchrotron beamlines. For some beamlines, this is detected automatically by MOSFLM (from the serial number of the detector) and dealt with accordingly. If this is not the case, select the checkbox "Reverse direction of spindle rotation" via the "Settings -> Experiment settings" menu. (Supplementary Tutorial Example 1)

2| Examine the displayed image (using the zoom, pan and contrast tools if necessary) to see if the diffraction quality is sufficiently good to warrant processing (ie diffraction is visible to a resolution that is sufficiently high to be useful) and whether the diffraction spots are well defined.

3| Check that the direct beam position is credible, for example, that it is within the central backstop shadow (the direct beam position can be displayed by the "Show beam centre" icon in the Image display panel).

TROUBLESHOOTING

4| Check for the presence of multiple lattices, badly split spots, ice rings, bad shadows or inactive regions of the detector. If necessary, mask out bad regions of the detector with the masking tool (Fig. 2). Ideally, also examine the image that corresponds to a 90° spindle rotation from the first image, as this will be used for indexing (Steps 5-8). For images recorded on Dectris Pilatus or Eiger detectors with a very small oscillation angle and short exposure times, it can be helpful to sum several images (use the Sum entry box in the image display).

Indexing the images

5| The Indexing icon becomes active (no longer grayed-out) once the images have been read. Select the Indexing icon to open the Indexing pane in the control panel.

The program will identify diffraction spots on the first image and on a second image that is as close as possible to a 90° spindle rotation from the first image. The found spots will be shown in the image display as crosses, and a representation of the found spots for both images is shown in the Indexing pane. The program sets a threshold for the intensity of spots to be used in indexing (corresponding to I/σ (I) of 20 for strong images, 10 for weaker images or 5 for very weak images). The initial threshold value of 20 is reduced if necessary to obtain at least 100 spots for indexing. Red crosses in the image display indicate spots with an intensity above this threshold, spots below this threshold are displayed in yellow. After finding spots, the program immediately indexes the two images using the selected spots, no user action is required (this behaviour, and the automatic spot finding on entering the Indexing pane, are controlled by check boxes in the Indexing and Spot finding tabs, respectively, of the Processing options).

Assuming the indexing is successful, a list of possible solutions corresponding to different lattice types is presented, with the suggested solution highlighted in blue. The selection of the suggested solution is based on the Penalty value (Pen.) and the rms difference (rmsd) between predicted and observed spot positions (denoted $\sigma(x,y)$ in the interface).

Provided a solution is found, the mosaicity of the crystal will be estimated (if not automatically indexing after finding spots, the mosaicity can be estimated by clicking the Estimate button). The mosaicity is defined in terms of a spherical model of a reciprocal lattice point[29] and is numerically a factor of 2-3 times larger than the value derived when processing data with XDS.

CRITICAL STEP All subsequent steps depend critically on obtaining the correct indexing solution. The success of the indexing step is best judged by visual inspection of the predicted reflections in the Image display. The predicted reflections should be a good match to the Bragg spots on the image; in particular the pattern of the lunes should be correct (Supplementary Tutorial Example 1). This can be judged most easily by toggling the display of the predicted reflections on and off, with the found spots display turned off. Providing the mosaicity estimation has worked correctly, all spots in the image should be predicted (except possibly at low resolution). In addition, the penalty (Pen.) will normally be below 20 for the correct solution. The rmsd values will usually be in the range 0.1–0.2 mm (the larger values are common for data collected using a laboratory source as a result of the larger spots sizes). However, if the spots are elongated or badly split, the rmsd can be 0.5-1.0 mm or larger for the correct solution (Supplementary Tutorial Example 2). In these cases only visual inspection of the predicted reflections will indicate if the solution is correct. Note that if the initial direct beam coordinates are only slightly in error, the indexing will probably find the correct solution, but an improved estimate of the cell parameters (resulting in a lower value of rmsd) can be found by simply repeating the indexing step (use the Index button), as this will use the refined values for the direct beam coordinates. If indexing is not successful, see Box 2.

CRITICAL STEP The assignment of symmetry at this stage is based solely on the shape of the unit cell. The true symmetry can only be determined when integrated intensities are

available (see step 25). For example, it is not uncommon for a monoclinic space group to have a β angle close to 90°, in which case the indexing will suggest an orthorhombic lattice.

TROUBLESHOOTING

6| If the predicted reflections appear to match the observed spots, but do not account for all spots on the image even after adjusting the mosaic spread and mosaic block size (see step 23), it is possible that there is more than one lattice present in the image. Sometimes this is evident from an inspection of the image itself, by the presence of two (or more) sets of intersecting lunes. The presence of multiple lattices usually also results in larger than expected values of σ(x,y) (e.g., 0.3–1.0 mm). Invoke multiple lattice indexing[31] by clicking the multiple lattice icon (next to the Index button) and repeat the indexing. Solutions for the different lattices will be presented in different tabs of the Indexing pane. (Supplementary Tutorial Example 7)

CRITICAL STEP Even if multiple lattices are known to be present, always do a conventional indexing first (step 3) as this will provide more accurate direct beam coordinates, which can be critical for multi-lattice indexing.

CRITICAL STEP Critical examination of the results of multiple lattice indexing is necessary. All lattices are normally expected to have the same symmetry. The predicted reflections for each lattice should be checked individually (the lattice can be selected in the Images pane), and when reflections from all lattices are displayed by clicking the multi-lattice icon this should account for all Bragg spots on the image.

7| Estimate the mosaic spread. If not already carried out automatically, use the "Estimate" button to get an initial estimate, which will normally lie between 0.1° and 1.5°. This value will be refined during cell refinement and integration.

CRITICAL STEP Cell refinement and Integration will not work correctly if the mosaic spread is left at zero. If either of these tasks is selected with a value of zero for the mosaic spread, then the mosaic spread estimation will be carried out automatically.

8| If the true space group is known, select it from the list of options in the drop-down menu at the bottom of the Indexing pane. The selection of a particular space group from this list will not have any effect on either cell refinement or integration, but will in some cases affect the strategy calculation.

Calculating a data collection strategy

CRITICAL If processing initial reference images (typically two) prior to a data collection, calculate a data collection strategy using the Strategy task (steps 9-10). If the data have already been collected, go to step 12. Note that the strategy calculation takes no account of the presence of multiple lattices. If multiple lattices are present, the calculations will be incorrect as they do not take account of overlapping spots from different lattices.

9| Select the Strategy task and click the Auto-complete button, which will open a pop-up window (Fig. 3). If collecting a complete dataset from this crystal, check the tick box if

anomalous data are required and then select Ok. The circle in the lower left pane shows the phi start and end values that should be used to collect a complete dataset. The histogram in the lower right shows the resulting data completeness as a function of resolution (other statistics are also available for display).

10| In order to investigate the effect of changing the phi start and end values proceed to option A. To use two or three segments rather than one, proceed to option B. To determine a strategy for the case where multiple crystals are required to collect a complete dataset, see option C.

(A) Changing phi start and end values

i. Selecting the yellow cross-hatched area in the lower left pane will highlight the sector in bold.

ii. Drag the small filled squares denoting the start and end phi values to new positions. The new sector will become red and the completeness will be recalculated.

(B) Using multiple segments

i. In the pop-up window, select the total rotation angle to be used and the number of segments (maximum of 3). The start and end phi values for each segment will now be displayed in the lower left pane. Using this approach it is sometimes possible to collect an almost complete dataset using substantially less than the expected total rotation; for example, a 60° total rotation in two segments will typically give a 95% complete dataset for orthorhombic symmetry.

(C) A strategy for multiple crystal data collection

i. Enter the number of degrees of data that will be collected from the first crystal (e.g., 20°) in the "Rotation" entry box of the pop-up window. Now select the yellow cross-hatched area (this will result in small filled squares appearing at the start and end phi values). Select either the starting or ending phi value so that the sector turns red.

ii. For the second crystal, index using the reference images as normal. On entering the Strategy panel, the sector from the first crystal will already be shown. In the pop-up window, enter the number of degrees of data to be collected from this crystal. The phi start value for the second crystal will be calculated that gives the highest overall completeness for data from both crystals.

iii. Repeat selection of the sector for the second crystal as in 10C(i) so that it turns red.

iv. For the third and subsequent crystals, repeat steps (10C(i) – 10C(iii)). The phi start values will again be calculated to give the highest completeness for all the wedges of data.

v. Select the "Save" option in the tool bar to save the current multiple crystal strategy as an ASCII file. If, for any reason, the iMosflm session is closed before

data has been collected from all the available crystals, start a new iMosflm session and index the next crystal. Then in the Strategy panel, read in the details of all the sectors collected from previous crystals using the "Add" option in the toolbar, selecting the saved strategy file. Calculate the strategy for the new crystal, making sure to select the Matrix for the new crystal in the pop-up window.

vi. The end phi values for each wedge can be changed as in option A if it is found that radiation damage resulted in a smaller wedge of useful data than expected. Subsequent strategy calculations will take this into account.

CRITICAL STEP The strategy calculation depends on the Laue group of the chosen indexing solution. From the indexing it is not possible to distinguish between some Laue groups (e.g., 3, 32; 6, 622 or 4, 422). By default, the lowest space group symmetry consistent with the chosen lattice is selected at the indexing stage. If the true space group is known, it should be selected. If not, it is safest to assume the lowest Laue group symmetry.

11| Calculate the maximum allowable oscillation angle per image to avoid spatial overlaps by selecting the "check for overlaps" button. With the default options, this plots the maximum angle as a function of phi angle over the phi range selected by the strategy calculation. Alternatively, the percentage of overlapping reflections for a range of different oscillation angles can be plotted.

CRITICAL STEP For large unit cells, the number of overlaps will critically depend on the maximum resolution, the mosaic spread and the spot size. The diffraction image should be examined to check that all these parameters have sensible values. New values can be entered via the toolbar.

CRITICAL STEP The program only reports the *maximum* oscillation angle that can be used to avoid spatial overlaps. For small unit cells (less than 100 Å) this can be several degrees. In practice it is advisable to collect data with an oscillation angle substantially less than one degree.

Refining the unit cell

12| If the diffraction data extend to a resolution higher than 3.5 Å, proceed to option A for steps on post-refinement of the unit cell parameters. If the resolution is below 3.5 Å, proceed to option B for cell refinement using the Indexing task.

(A) Images in two or more wedges of data are integrated in order to provide the intensities required for post-refinement[29] of the unit cell parameters, crystal orientation and crystal mosaicity.

i. Select the Cell Refinement task (Fig. 4). Images belonging to two wedges of data (or more for lower symmetries) are automatically selected but these can be edited manually.

ii. Select the "Process" button. The images will be integrated and the cell refinement carried out. If any of the cell parameters change by more than 2.5x

their estimated standard deviation, the images will be re-integrated and the refinement repeated. A maximum of 5 cycles of integration and cell refinement is allowed.

iii. Check the plots of the refined detector and crystal parameters to ensure that the refinement is stable. In particular the tilt and twist should not vary by more than 0.1–0.2° and the beam *x,y*-coordinates should not change by more than 0.1 mm. Within each wedge of data, Y-scale should not vary by more than 0.001 and the distance should not change by more than 1 mm. Between wedges, the Y-scale and distance can vary more than this, as these values compensate for cell parameter changes due to radiation damage. Plots of individual parameters can be displayed by highlighting the appropriate parameter. The plots can be expanded to full screen by a combination of shift+left mouse button.

iv. Inspect the rms error in spot positions (RMS residual), detector distance and Y-scale parameter which are plotted for each image as a function of cycle number in the lower right window. The RMS residual should decrease in later cycles. The distance and Y-scale values may change between cycles, but ideally the values for the different wedges used in refinement should be more similar to each other in later cycles.

v. When the refinement has converged (or after 5 cycles) the initial and refined cell parameters are displayed, together with error estimates. The errors should be less than 1 part in 1000.

TROUBLESHOOTING

TROUBLESHOOTING

(B) Cell refinement using the Indexing task

i. Select 4-7 images for the indexing rather than the default of two. Ideally these images should be well separated in phi, covering a total range of 90° (e.g., phi values of 0, 15, 30, 45, 60, 75 and 90).

ii. Repeat the indexing. This will generally provide more accurate cell dimensions than using two images, especially for lower symmetry space groups.

Integrating the dataset

13| Select the Integration task. By default, all images in the sector will be selected for integration. However, we advise restricting the initial integration to the first 5-10 degrees of data, so that any possible issues can be detected quickly.

14| The MTZ output filename will default to a name based on the image filename and the number of the first image processed. If desired, change this by typing a new name into the entry box.

15| By default, the image display is not updated when images are being processed, as thissubstantially slows down the processing. In difficult cases it can be helpful to display

each image with the predicted reflections superposed. To do this, select the icon with the tool-tip "Show predictions on images during processing" from the toolbar.

16| If the images show strong rings or spots due to diffraction from crystalline ice, select the "snowflake" icon from the toolbar to exclude all reflections that lie in resolution shells corresponding to crystalline ice.

CAUTION Selecting this option will result in a substantial drop in completeness of the measured data. In practice it is generally better to define specific resolution shells to be excluded using the Settings -> Processing Options -> Processing tab.

17| When processing data collected on a laboratory source, where the data collection time can be several hours (rather than a few minutes when using a synchrotron beamline), it is possible to start integration before all the images have been collected. Select the "clock" icon, to allow the final image number for the integration to be numerically greater than the last image read into the session in step 1.

18| If more than one sector of data has been read in, select the sector to be processed from the drop-down menu.

19| If processing images with multiple lattices, select the lattice to be integrated from the drop-down menu.

20| Select the Process button. This will integrate the selected images and generate plots (Fig. 5) of the refined detector and crystal parameters, intensity statistics and the number of reflections that are overloaded, spatially overlapped or are classified as "bad spots". Gray-scale representations of the average spot profile for reflections in the central region of the detector and the standard profiles used for integration are shown in the upper right and central panels. Mean $I/\sigma(I)$ values are displayed as a function of resolution in the lower right panel.

CRITICAL STEP The plots of refined detector and crystal parameters should be examined to check for stable refinement, as detailed in step 12A. In particular it is important that the mosaic spread refinement is stable and does not refine to a value that is clearly too small.

TROUBLESHOOTING

21| Check for program warnings. The number of warnings is shown in the lower right of the display. A "traffic signal" warning colour is also shown. Green indicates no issues with the processing, orange indicates that there are minor issues and red indicates serious issues that should be addressed. Select the "warnings" box to display a list of the individual warning messages; double click on individual warnings to show more details and suggestions of ways to address them.

22| Adjust the mosaic block size. If some reflections at low resolution are not predicted following the initial integration (when the crystal orientation has been refined) then the mosaic block size should be reduced from its default value of 100 microns. Different values

should be entered and the reflection predictions examined. A value as low as 0.5 microns may be necessary in some cases. (Supplementary Tutorial Examples 2, 8, 9)

TROUBLESHOOTING

23|If the processing is satisfactory, proceed to integrate all images in the dataset (selecting the "**a**" icon next to the image number entry box will automatically select all images in this sector).

CRITICAL STEP Examine the plots for Overloads, Bad Spots and Spatial overlaps. If there are many overloaded reflections, the option to include profile fitted estimates of the intensity of overloaded reflections should be selected when running "QuickScale" (step 25).

TROUBLESHOOTING

Determining the Laue Symmetry

24| Select "QuickSymm" from the toolbar of the Integration panel. This will run POINTLESS26 which will attempt to determine the Laue symmetry and, if possible, the space group symmetry. The output is displayed in a new qtRView window, with both the log file and a graphical representation of the important results. The Laue symmetry can often be determined from 5–10 degrees of data. If the symmetry is not known in advance it is worth running QuickSymm using the small wedge of data integrated in step 13 in order to confirm that the correct symmetry was chosen at the Indexing step.

CRITICAL STEP If the Laue symmetry indicated by POINTLESS26 is lower than that of the indexing solution (and the Laue group confidence is >0.5) it is essential to return to the Indexing pane and select a solution that matches the Laue symmetry indicated by POINTLESS26. Examine the values of the correlation coefficient for each symmetry operator carefully, as non-crystallographic symmetry or merohedral twinning can both result in POINTLESS26 incorrectly selecting a higher Laue group symmetry than the true symmetry. For example, if an orthorhombic symmetry is being tested, the values of the correlation coefficient for all three 2-fold axes should be similar (providing there are a reasonable (e.g.>50) number of reflections). If POINTLESS is selecting the wrong symmetry, the correct space group should be selected at the indexing stage, and the option to use the iMosflm symmetry in QuickScale should be selected from the "Sort Scale and Merge" tab of the Processing options. (Supplementary Tutorial Example 9).

TROUBLESHOOTING

Merging and scaling the dataset

25| Select "QuickScale" from the toolbar of the Integration pane. This will run POINTLESS26 to determine the Laue symmetry and, if possible, the space group symmetry, and will then scale and merge the data in AIMLESS27 using that symmetry (unless this is over-ridden as described in step 24). The output of both steps is displayed in a new qtRView window. The TRUNCATE28 program will then generate an MTZ file containing structure factor amplitudes (as well as intensities) and provide intensity statistics that can be used to

detect twinning. An R-free flag will also be added to the MTZ file, giving a file that is ready to be used in downstream programs for phasing and/or structure solution.

CRITICAL STEP A limited number of parameters that affect the scaling and merging can be set from the "Sort Scale and Merge" tab of the Processing options. These include specifying multiple input MTZ files (corresponding to different sectors of data or different crystals), changing resolution limits, excluding batches and controlling the standard deviation correction terms.

Integrating multiple lattices

26| If multiple lattices have been identified at the indexing stage, the integration needs to be carried out separately for each lattice. Once the first lattice has been integrated, integrate the second and subsequent lattices by going back to step 19. The name of the MTZ file is automatically updated for each lattice. When all lattices have been integrated, merge the data using the program FECKLESS prior to running POINTLESS26 and AIMLESS27 by selecting the QuickSymm or QuickScale options. (Supplementary Tutorial Example 7).

CRITICAL STEP If multiple lattices have been found, but QuickSymm or QuickScale is to be run only on one of the lattices, it is necessary to click on (and therefore de-select) the multiple lattice icon in the toolbar (immediately next to QuickSymm), otherwise an error message will be given.

Integrating images in a background job

27| When processing a large number of images, or processing the same images with different parameter values (such as mosaicity, mosaic block size or spot separation) it can be helpful to set up a command script rather than running all the jobs using iMosflm. Select the "Batch" option from the drop-down menu next to the "Process" button. A pop-up window will display all the keywords necessary to run MOSFLM as a background job; these can be cut and pasted into a suitable command script. When integration is carried out in this way the graphical iMosflm output is lost. However, a summary file is produced (with the extension *.sum* and many of the plots can be reproduced by running the CCP4 program LOGGRAPH on this file.

Integrating images in parallel

28| For large datasets the integration can be carried out in parallel over a number of processors. In order to do this, select the "Parallel" option from the drop-down menu next to the "Process" button. The images to be processed are divided into a number of batches, depending on the number of cores available.

CRITICAL STEP It is important that the mosaic spread has already been refined before submitting a parallel processing job.

TIMING In straightforward cases, after reading in all the images, an indexing solution (steps 5-8) can be obtained in a few seconds. However in challenging cases, which could involve changing some parameters associated with spot finding and indexing, carrying out a direct beam search and/or trying different images and different resolution limits, this step

can take several minutes. The Strategy calculation (steps 9-11) only takes a few seconds. Cell refinement (step 12) will typically take between 30 seconds and two minutes, depending on the quality of the indexing solution and the resolution of the data. Integration (steps 13-23) is the most time-consuming task and will depend on the number of images to be processed and the typical number of reflections on each image. For example, a dataset with ~700 reflections per image will take ~0.3 seconds per image, while a larger dataset with ~2000 reflections per image takes ~0.6 seconds per image. For the latter case, a complete dataset of 1200 images would take twelve minutes to integrate. Using the parallel processing option (step 28) for large datasets can provide a speedup by a factor of 4–8. Determining the Laue group or space group (step 24) will normally take only a few seconds, although it will take up to a minute or longer if run on an entire dataset. Scaling and merging the data (step 25) is normally complete in one to two minutes.

TROUBLESHOOTING

Most common error conditions result in a warning being displayed in a pop-up window, with a description of the error and suggestions on how to correct for it. However, not all errors are trapped in this way, so if the program stops with no obvious warning message, the MOSFLM log file should be examined. This is achieved by selecting the History task and then the Log tab. An error message will be in the log file, typically 50–100 lines from the end.

Some errors will result in the activity icon (top right of the iMosflm window) spinning continuously even though no processing is being carried out. If this occurs, the spinning icon should be clicked and in some cases it will then be possible to continue processing (but do not simply repeat the action that led to the error). In other cases, it will not be possible to execute further tasks via the interface (no action occurs when selecting a new task) and the iMosflm program should be aborted (Ctrl-C in the terminal window where the job was launched) and then restarted. On restarting, a "Recover session" pop-up window will be displayed. Selecting "Recover" in this window will restore the session to the point at which the last task (that caused the error) was started. This saves repeating steps such as reading in images, indexing, etc. Processing sessions can also be saved to a file (Session menu). Reading in a saved session will restore all the graphical output of that session.

Troubleshooting advice on specific errors can be found in Table 1.

Anticipated Results

Provided that a satisfactory indexing solution has been achieved, the quality of the output data will be determined by the quality of the diffraction images themselves. Overall quality indicators can therefore vary widely between different datasets and there are no "correct" values. One possible exception to this is the value of Rmerge at low resolution. Rmerge is a measure of the agreement between the intensities of reflections related by crystallographic symmetry, and for low resolution reflections (say 20-5Å) should normally be in the 2–4% range. Higher Rmerge values indicate a problem either with the processing (e.g. wrong indexing solution, incorrect masking of the direct beam shadow, incorrect choice of mosaic

block size) or with the data itself (e.g. radiation damage, the merging of data from non-isomorphous crystals). Data from a well diffracting crystal (e.g. thaumatin) collected at a synchrotron beamline can give overall Rmerge values of 5.2% at 1.57 Å resolution (2.0% at low resolution) while important biological information was obtained for a bacterial ATP synthase where the overall Rmerge was 18.0% (3.7% at low resolution)10. Rmeas is a statistically more meaningful parameter than Rmerge, as it takes into account the multiplicity of the data, but again a wide range in values is observed in the literature. Importantly, the final resolution limit for the dataset should be based on half dataset correlation coefficients CC(1/2) or mean (I/σ(I)) values for the merged data, and not on the Rmerge or Rmeas values, as these can exceed 100% at resolutions where there is still meaningful information in the merged data32. The statistics provided at the data scaling and merging step (AIMLESS27) are far better indicators of data quality than any statistics provided during the integration, as the latter take no account of the improvement provided by averaging multiple measurements.

It is worth noting that MOSFLM is very robust and can sometimes succeed in integrating datasets even when the indexing solution is incorrect. In order to avoid this, it is very important to carefully inspect the results of the indexing step (through the comparison of predicted and observed spots) and the subsequent integration (by monitoring the stability of refined parameters, appearance of standard profiles).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Berman, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]

2. Arndt, UW., Wonacott, AJ. The rotation method in crystallography. North-Holland Publishing company; Amsterdam, The Netherlands: 1977.

3. Nyborg, J., Wonacott, AJ. Computer Programs. The rotation method in crystallography. North-Holland Publishing company; Amsterdam, The Netherlands: 1977. p. 139-152.

4. Leslie, AGW., Powell, HR. Processing diffraction data with MOSFLM. Evolving Methods for Macromolecular Crystallography. Read, RJ., Sussman, JL., editors. Springer Press; the Netherlands: 2007. p. 41-51.

5. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. iMosflm: a new graphical interface for diffraction image processing with MOSFLM. Acta Crystallogr D. 2011; 67:271–281. [PubMed: 21460445]

6. Katona G, et al. Conformational regulation of charge recombination reactions in a photosynthetic bacterial reaction center. Nat Struct Mol Biol. 2005; 12:630–631. [PubMed: 15937492]

7. Aoyama H, et al. A peroxide bridge between Fe and Cu ions in the O2 reduction site of fully oxidized cytochrome c oxidase could suppress the proton pump. Proc Natl Acad Sci USA. 2009; 106:2165–2169. [PubMed: 19164527]

8. Warne T, et al. Structure of a β₁-adrenergic G-protein-coupled receptor. Nature. 2008; 454:486–491. [PubMed: 18594507]

9. Lebon G, et al. Agonist-bound adenosine A$_{2A}$ receptor structures reveal common features of GPCR activation. Nature. 2011; 474:521–525. [PubMed: 21593763]

10. Morales-Rios E, Montgomery MG, Leslie AGW, Walker JE. Structure of ATP synthase from *Paracoccus denitrificans* determined by X-ray crystallography at 4.0 Å resolution. Proc Natl Acad Sci USA. 2015; 112:13231–13236. [PubMed: 26460036]

11. McCusker EC, et al. Structure of a bacterial voltage-gated sodium channel pore reveals mechanisms of opening and closing. Nat Commun. 2012; 3:1102. [PubMed: 23033078]

12. Natsume R, et al. Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. Nature. 2007; 446:338–341. [PubMed: 17293877]

13. Prodromou C, et al. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. Cell. 1997; 90:65–75. [PubMed: 9230303]

14. Cheung ACM, Sainsbury S, Cramer P. Structural basis of initial RNA polymerase II transcription. EMBO J. 2011; 30:4755–4763. [PubMed: 22056778]

15. Cheetham GM, Jeruzalmi D, Steitz TA. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature. 1999; 399:80–83. [PubMed: 10331394]

16. Martick M, Scott WG. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell. 2006; 126:309–320. [PubMed: 16859740]

17. Hirata K, et al. Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. Nat Methods. 2014; 11:734–736. [PubMed: 24813624]

18. Nederlof I, van Genderen E, Li Y-W, Abrahams JP. A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals. Acta Crystallogr D. 2013; 69:1223–1230. [PubMed: 23793148]

19. Nannenga BL, Shi D, Leslie AGW, Gonen T. High-resolution structure determination by continuous-rotation data collection in MicroED. Nat Methods. 2014; 11:927–930. [PubMed: 25086503]

20. Waterman DG, et al. The DIALS framework for integration software. CCP4 Newslett Protein Crystallogr. 2013; 49:16–19.

21. Kabsch W. XDS. Acta Crystallogr D. 2010; 66:125–132. [PubMed: 20124692]

22. Pflugrath JW. The finer things in X-ray diffraction data collection. Acta Crystallogr D. 1999; 55:1718–1725. [PubMed: 10531521]

23. Otwinowski, Z., Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology Vol. 276: Macromolecular Crystallography, part A. Carter, CW., Jr, Sweet, RM., editors. Academic Press; New York, USA: 1997. p. 307-326.

24. Hattne J, et al. MicroED data collection and processing. Acta Crystallogr A. 2015; 71:353–360.

25. Crowther RA, Henderson R, Smith JM. MRC Image Processing Programs. J Struct Biol. 1996; 116:9–16. [PubMed: 8742717]

26. Evans PR. Scaling and assessment of data quality. Acta Crystallogr D. 2006; 62:72–82. [PubMed: 16369096]

27. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D. 2013; 69:1204–1214. [PubMed: 23793146]

28. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D. 2011; 67:235–242. [PubMed: 21460441]

29. Leslie AGW. The integration of macromolecular diffraction data. Acta Crystallogr D. 2006; 62:48–57. [PubMed: 16369093]

30. Steller I, Bolotovsky R, Rossmann MG. An algorithm for automatic indexing of oscillation images using Fourier analysis. J Appl Cryst. 1997; 30:1036–1040.

31. Powell HR, Johnson O, Leslie AGW. Autoindexing diffraction images with *iMosflm*. Acta Crystallogr D. 2013; 69:1195–1203. [PubMed: 23793145]

32. Karplus PA, Diederichs K. Assessing and maximising data quality in macromolecular crystallography. Curr Opin Struct Biol. 2015; 34:60–68. [PubMed: 26209821]

Box 1

Detector requirements for datasets processed with MOSFLM

The MOSFLM package can process diffraction images from the following detectors:

- Image Plates: Mar Research/MarXperts, Rigaku, Mac Science.

- CCD detectors: ADSC, Rayonix, Rigaku, Bruker, Oxford Instruments.

- Hybrid pixel detectors: Dectris, ADSC.

- Specialist detectors developed by Ed Westbrook (CMOS and CCD) and the ESRF (CCD).

Other considerations to take into account:

- Images from CCD detectors must have already been corrected for spatial distortion and non-uniformity of response.

- Images from the custom-made CSPAD detector at the LCLS free electron laser cannot be processed, but data recorded on commercial CCD detectors at both the LCLS and the Japanese free electron laser SACLA can be processed routinely.

- Electron diffraction images cannot be processed directly, but need to be converted to a format recognized by the program. Software is available[24] to convert images written by the TVIPS TemCam-F416 CMOS detector to Super Marty View (SMV) format (this is the format adopted for ADSC CCD detectors) and a conversion program is available on the MOSFLM website to convert MRC format images[25] to SMV format. For these images it is still necessary to define the orientation of the detector relative to the MOSFLM coordinate frame, which is defined in Appendix III of the "iMosflm User Guide" available on the MOSFLM website.

Box 2

Ways to achieve successful indexing in difficult cases

There are a variety of actions to undertake if the initial indexing attempt fails. Use the Index button to repeat the indexing after changing any of the parameters listed below. If a solution is found then continue to step 6.

i. Check the direct beam position, the detector distance and the X-ray wavelength. If the direct beam position is uncertain, use the "Start beam search" option. The parameters for this search can be changed in the Indexing tab of Processing options. (Supplementary Tutorial Example 3)

ii. Include more images for indexing at phi values quite different to those already being used. For example, if the phi values of the initial images are 0° and 90°, try including images at 30° and 60°. If possible, include images that show a clear separation of the lunes and well-defined spot shapes. Avoid including any images that show signs of serious radiation damage (if necessary, exclude the second image of the two chosen initially).

iii. If there are fewer than 100 reflections being used for the indexing, increase this either by reducing the I/σ (I) threshold for spots to be included (but do not use less than 3) or by including additional images in the spot finding stage by entering the image numbers in the entry box.

iv. If many spots are not found because they are very weak and/or rather diffuse, use the Spot finding tab of Processing options to reduce the threshold for finding spots ("Threshold I/σ (I) ") from the default value of 5.0 (do not go below 2.0) and decrease the "Spot rms variation" parameter from its default (3.0 for synchrotron data, 1.0 for laboratory sources) and repeat the spot finding by clicking the "Find" icon for each image (Supplementary Tutorial Example 4).

v. Inspect the images to make sure that all spots being used in the indexing (red crosses) are genuine Bragg spots and that split spots are not being treated as two separate spots (see ix below). (Supplementary Tutorial Example 5)

vi. If spots are not found because they are extremely small (2–3 pixels), reduce the "Minimum pixels per spot" parameter in the Spot finding tab. If the spots are extremely large, increasing this parameter can help discriminate real Bragg spots from noise in the image.

vii. In very unusual cases where following steps iii, iv and vi does not result in the detection of spots that are clearly visible by eye, manual spot selection (Fig. 2) can be used.

viii. Attempt indexing using only one image. If the direction of spindle rotation is wrong, indexing will work with one image but not with two or more. If indexing with the first (or second) image alone works, check the prediction for the subsequent images. If these do not match, then the spindle direction is

probably wrong (see step 1 and Supplementary Tutorial Example 1). If only one image shows poor spot shape (split, streaky etc.) omitting this image might result in successful indexing. Also, if the crystal has changed orientation during data collection by more than a few tenths of a degree, indexing using only the first image may give a much better prediction than when using both images and this may give improved integration (Supplementary Tutorial Example 6).

ix.   Try using different values for the "I/σ (I) cutoff". For example, if the program selected a value of 20, try values of 5, 10, 15, 30 and 50.

x.   Try changing the "Max cell edge" parameter. This is displayed after an indexing attempt. For example, if the displayed value is 300 Å, try values of 200, 250, 400 and 500 Å.

xi.   If the Bragg spots are badly split and the two (or more) components of a single Bragg spot are being treated as separate spots, then defining the "Minimum spot separation" in the "Spot finding" tab to be slightly larger than the true spot size can help. The true spot size (in mm) can be estimated by examining a (strong) reflection, measuring its size in pixels, and using the known pixel size (given in the Detector tab of Experiment settings) to convert this to mm. Alternatively, simply entering a value of 1.5 mm will work in the majority of cases. (Supplementary Tutorial Example 5).

xii.   It is not uncommon to find "false" spots close to regions with a strong background gradient, for example due to the backstop shadow, or adjacent to tile boundaries in a tiled detector. Normally these spots will be excluded from indexing by the I/σ (I) threshold. However, if it is necessary to reduce the I/σ (I) threshold to a low value in order to include a sufficient number of reflections, then this region of the detector should be masked out with the masking tool (Fig. 2) to avoid including these false spots in the indexing.

xiii.   In rare cases it can help to change parameters associated with the Fourier based indexing[30,31] itself (rather than spot finding). The "maximum deviation from integral hkl" parameter can be reduced from the default value of 0.3 to values of 0.25 or 0.2, and the "number of vectors to find for indexing" can be increased from 30 to 50. Both parameters are set in the Indexing tab of Processing options.

Editorial Summary

The MOSFLM software is widely used for X-ray diffraction data integration. The graphical user interphase version iMosflm now also makes this powerful software program accessible even to inexperienced users.
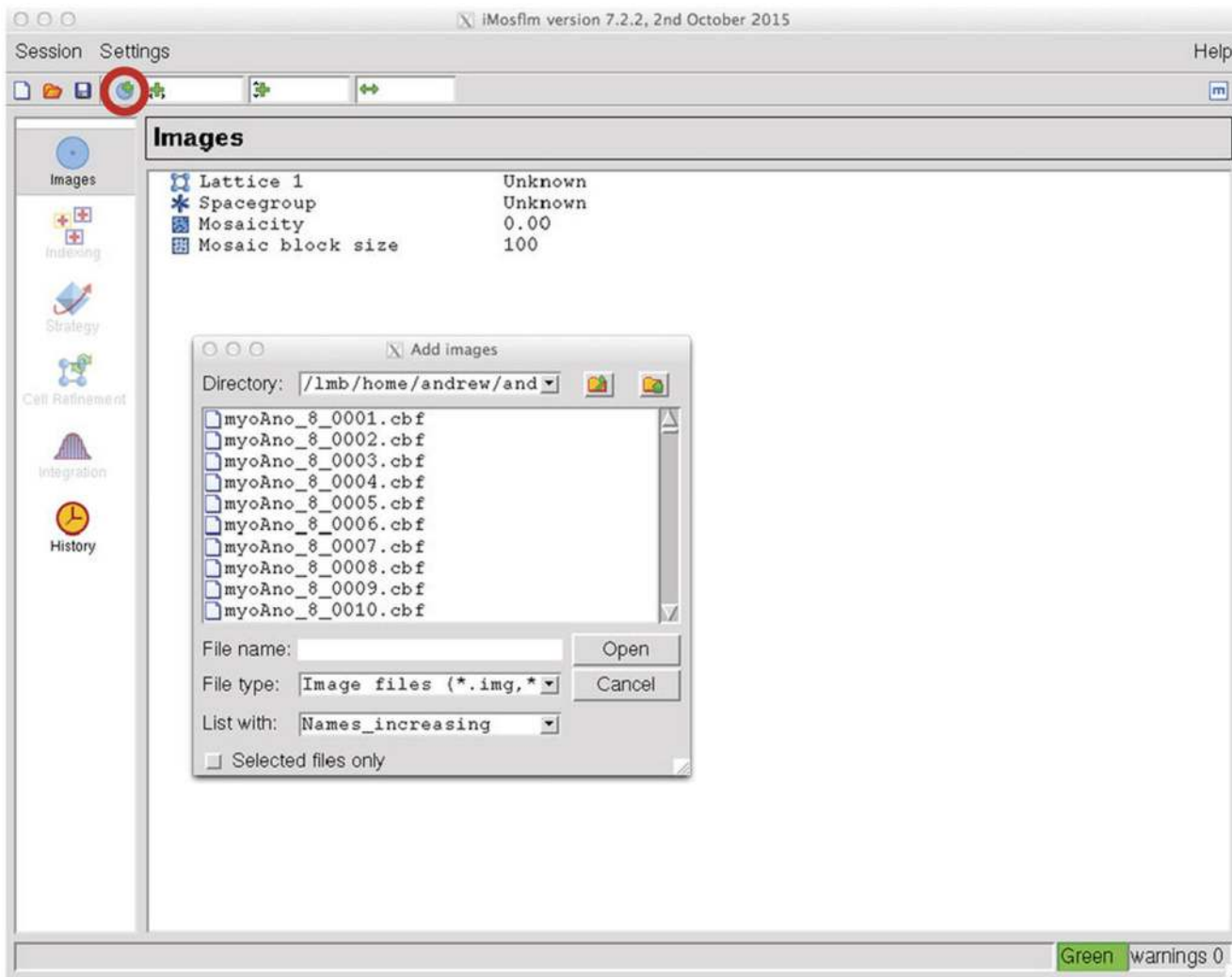
Figure 1.
The iMosflm Images window. Images are selected from the "Add images" browser that is activated by the "Add images" icon (circled in red). At this point all other operations (Indexing, Strategy etc.) are inactive (grayed out).
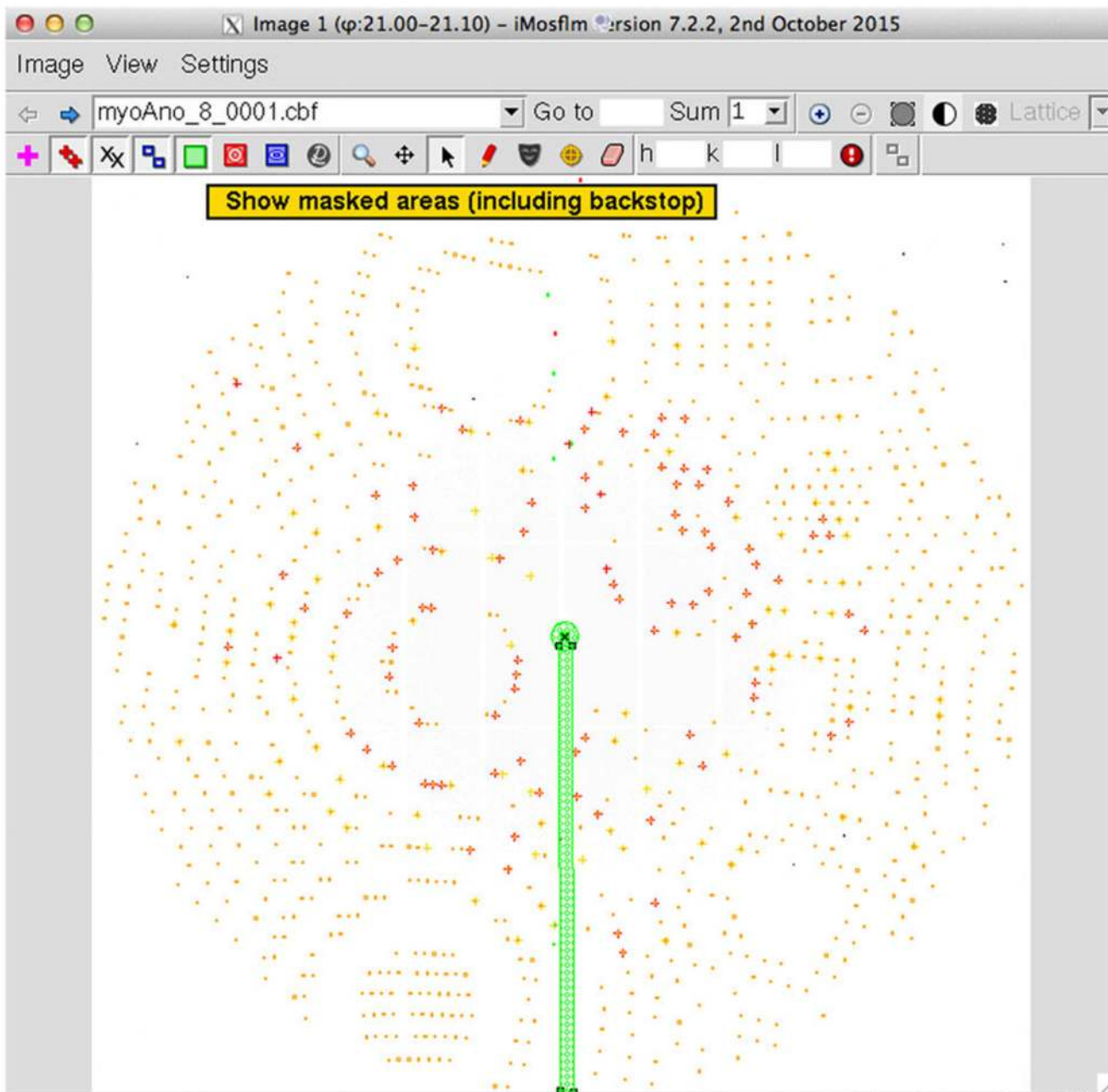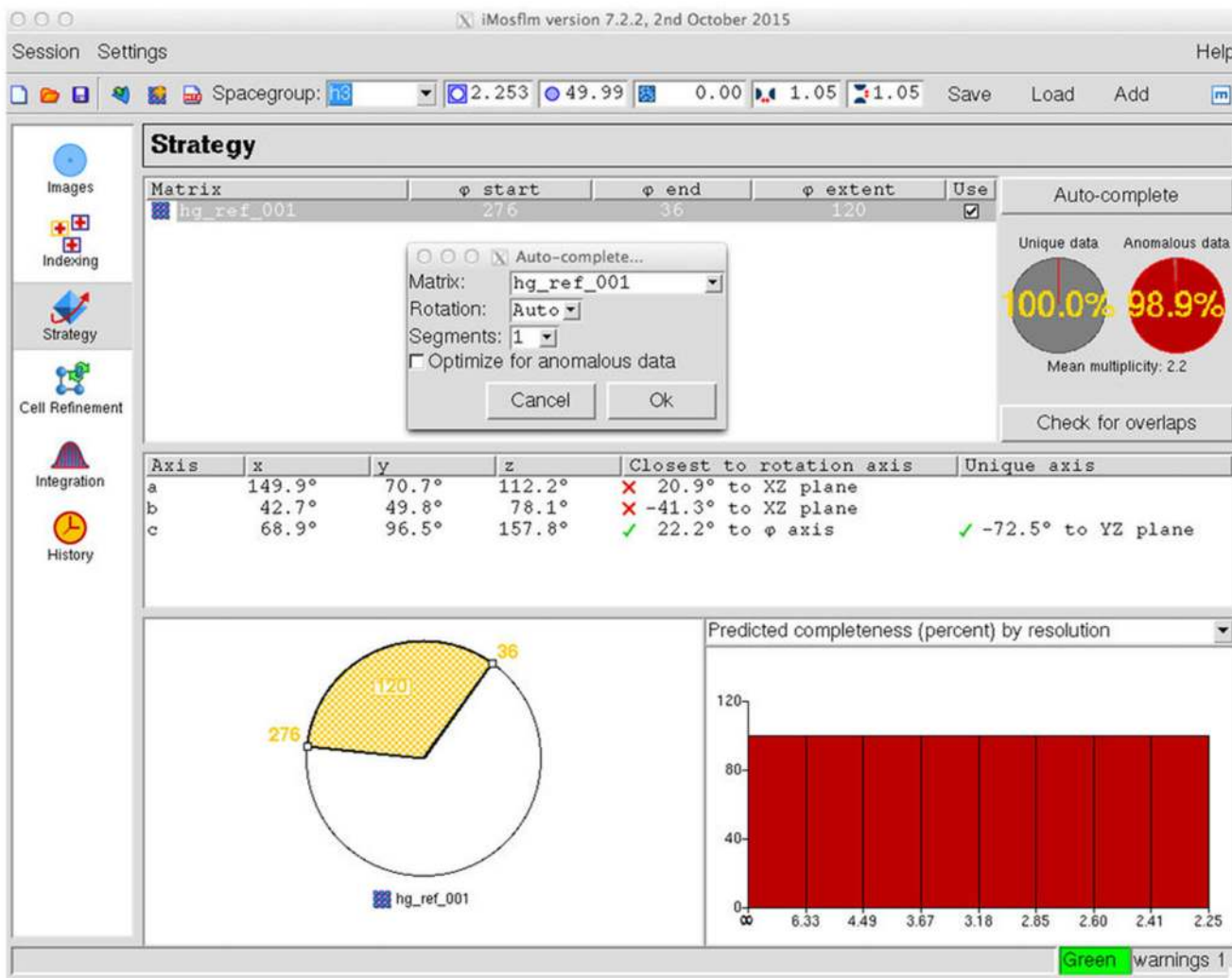
Figure 2.
The iMosflm image display window. The fifteen icons (from left to right) control display of the direct beam position (magenta cross); found spots for autoindexing; "bad" spots from the integration; predicted reflections (yellow – partially recorded, green – too wide in phi); masked areas (eg backstop); spot finding area; resolution limits; Rigaku active mask; zoom control; pan control; selection pointer; manual spot addition tool; masking tool; circle fitting tool; mask eraser. These are followed by entry boxes to allow finding a particular reflection and icons for "reflection successfully found" and multi-lattice display. Each icon has an explanatory "tool tip" that is activated when the mouse is placed over the

icon, as shown for the display of the masked areas (stippled in green). Other features include image selection, summing of images, zooming and shrinking, contrast control, reversed video display and selection of the lattice when processing images with multiple lattices. The size of the display can be adjusted from the View menu and the colour of the prediction boxes can be controlled from the Settings menu.
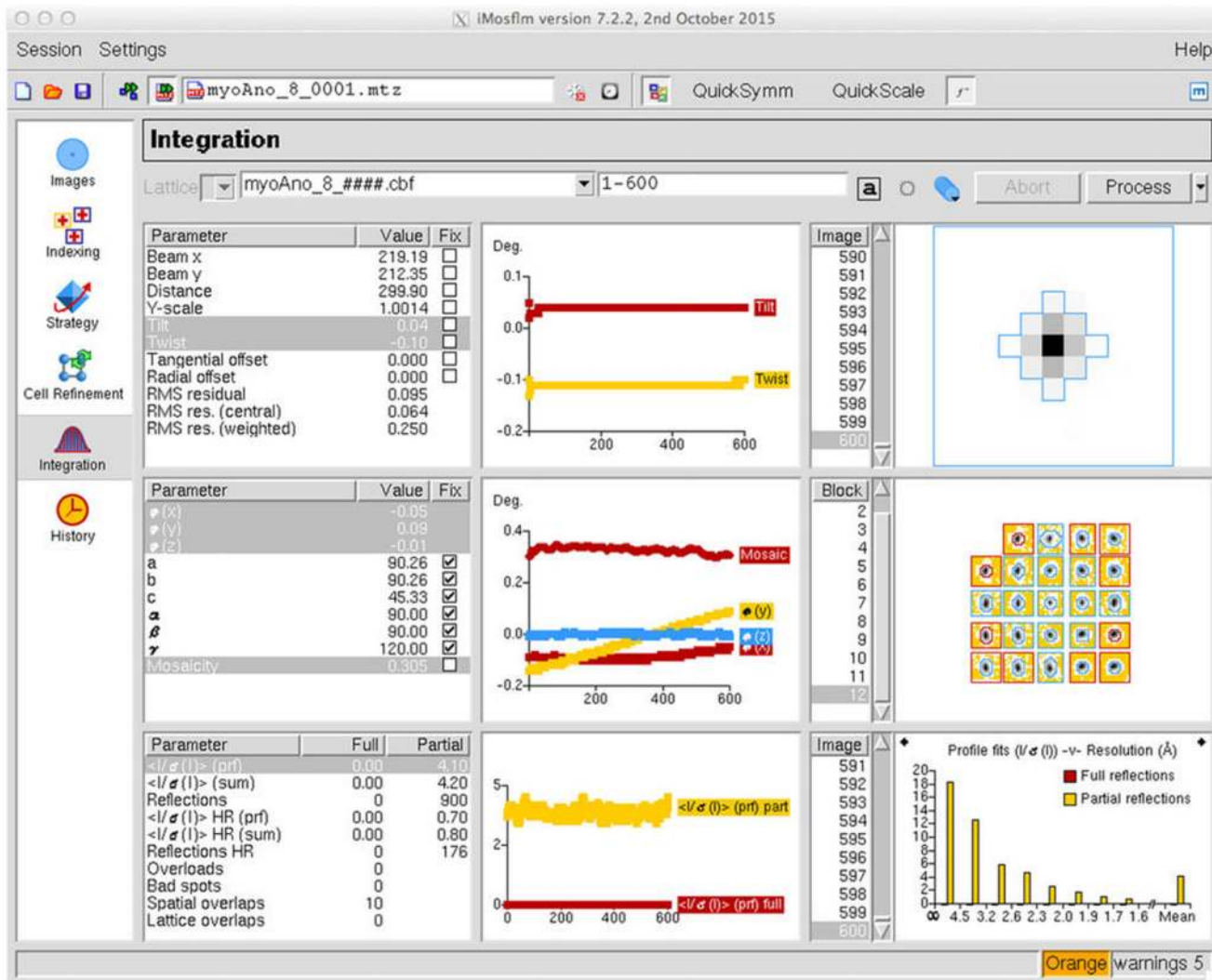
Figure 3.
The iMosflm Strategy window. On completion of the Auto-complete pop-up window, the program will calculate the starting rotation angle and rotation range required to collect a complete dataset for the indexing solution chosen at the Indexing stage (in this case, spacegroup h3). This is shown in the hatched yellow sector in the circle in the lower left of the pane. A histogram of the completeness as a function of resolution is shown in the lower right. The orientation of the crystal at a phi value of zero is shown in the centre of the pane in terms of the angles between the a, b and c axes of the unit cell and the laboratory X, Y, Z coordinate frame, where X is along the X-ray beam, Z is along the rotation axis and Y forms a right-handed set.
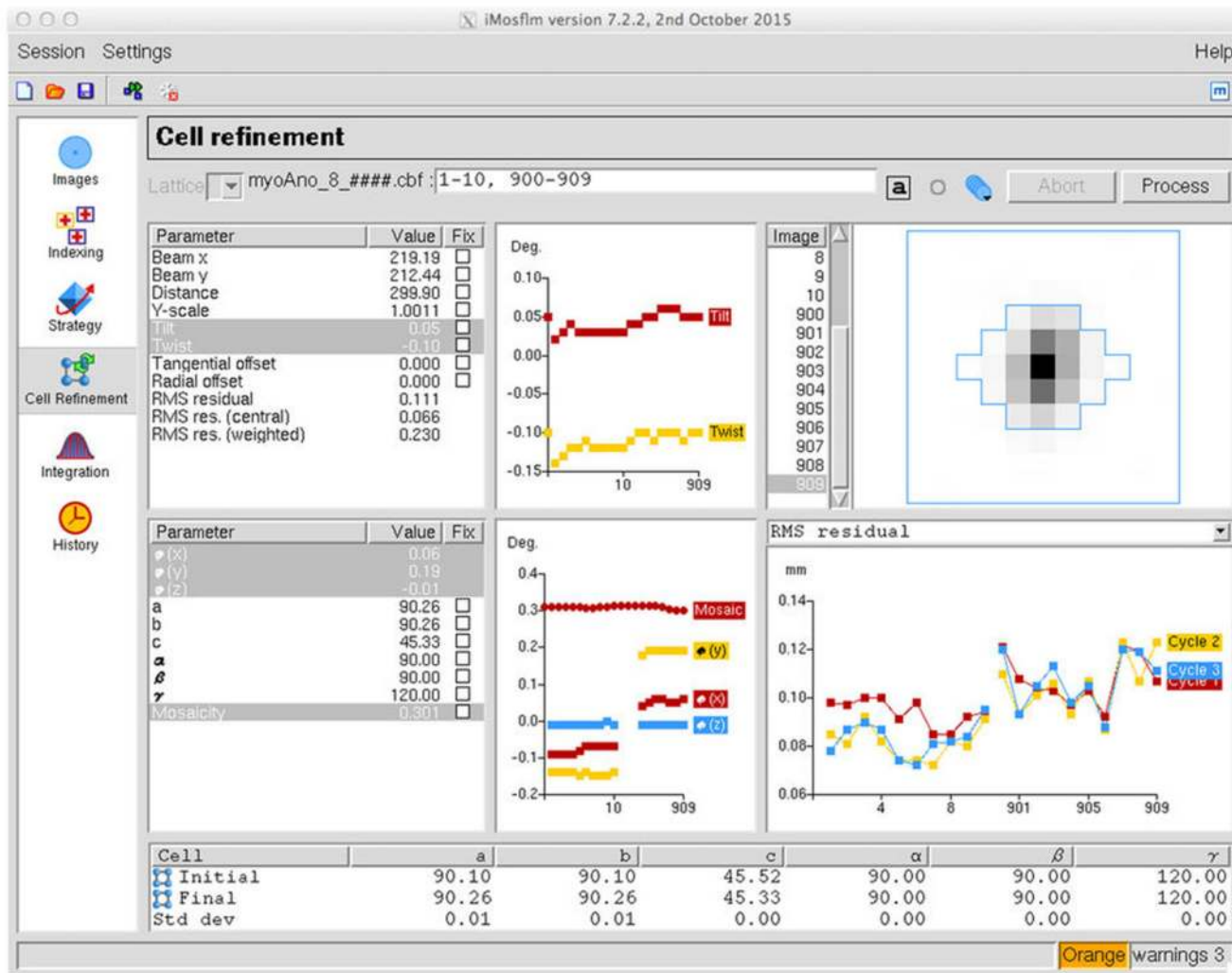
Figure 4.
The iMosflm Cell Refinement window. In this example, images 1-10 and 900-909 were integrated and the resulting data used to refine the unit cell parameters. Selected detector parameters (tilt and twist) and crystal parameters (mosaicity and missetting angles) are plotted as a function of image number. Parameters or statistics to be plotted are selected by clicking on the lists in the left hand panes (these are shown grayed out in the figure). The average spot profile for reflections in the central region of the detector are shown in the upper right panel. The initial and refined cell parameters, and an estimate of the uncertainties of the refined values, are given in the lower part of the panel.

Figure 5.
The iMosflm pane for the Integration task. Detector parameters are listed in the upper left pane and plotted as a function of image number in the upper central pane. Crystal parameters are listed in the central left pane and plotted in the adjacent pane to the right. Other statistics are listed in the lower left pane and plotted in the adjacent pane to the right. The upper right pane shows the average spot profile for reflections in the central region of the detector for each image; the pane below shows the standard reflection profiles in different regions of the detector for each block of images and the lowest pane shows intensity statistics as a function of resolution. Yellow pixels in the spot profiles are those with negative values following subtraction of the background.

Table 1

Troubleshooting Table

Step	Problem	Possible reason	Solution
1	The Phi values are clearly incorrect (e.g., all images have the same start and end phi values or all angles are zero)	The phi values are incorrect in the image header.	Enter start and end phi values for the first image. This is done by highlighting that image in the list, click over the phi values and then enter the correct start and end phi. The phi values for all other images will be calculated, assuming all images have the same oscillation angle. A sector with gaps in phi will be handled correctly. However, all images that are contiguous in serial number are assumed to have the same oscillation angle as the first image in the series, regardless of the information in the image headers.
1	A warning is given that the detector distance and/or wavelength values are zero	The values have incorrectly been read from the image header.	The "Experiment settings" window will be displayed so that the correct parameters can be provided
3	Direct beam position is clearly not correct (not within backstop shadow) (Supplementary Tutorial Example 1)	The direct beam position (read from the image header) is substantially in error.	Use the selection tool to drag and drop the direct beam marker to the most likely position in the centre of the central backstop shadow. If there are any powder rings present, use the "Circle-fitting" tool to determine the direct beam position.
5	The mosaicity estimation does not work correctly and results in a value that is too large (Supplementary Tutorial Example 1)	The indexing solution found is not correct.	Follow the checklist steps in Box 2. If a solution is found, then continue to step 5.
	The predicted reflections are close to the observed spot positions but do not match very well, particularly at high resolution (Supplementary Tutorial Example 8)	The direct beam position may be slightly in error. Even small errors can cause miss-indexing if one or more unit cell dimensions are large (>150Å) or if the spot shape is poor (split or fuzzy).	Integrate a few degrees of data to see if refinement of the crystal orientation and detector parameters results in an improved fit. First, the resolution should be reduced so that within the new resolution limit the spots are reasonably strong and the predictions at least partly overlap the observed spots. Integrate 5–10° of data following steps 13–23 and check the symmetry as described in step 24 (do not refine the cell parameters at this stage). If the reported symmetry is the same as that selected from the Indexing step proceed to step 12, otherwise select a different indexing solution and repeat the process. If the Friedel pairs have a poor correlation coefficient reported by POINTLESS26 (step 24) then the direct beam position is almost certainly incorrect and a direct beam search should be carried out, possibly searching over a larger area. When a satisfactory integration has been achieved, proceed to step 12
6	Multiple lattices are not detected even though they are visible in the images	Success in indexing multiple lattices can depend critically on some indexing parameters.	Vary the "Max cell edge" parameter in the Indexing pane (try smaller and larger values, varying by ~25%). Vary the "maximum deviation from integral hkl" parameter in the Indexing tab of the Processing options. This parameter defaults to 0.2 for multiple lattice indexing; try values from 0.15 to 0.3)
12A	Errors in cell parameters are greater than 0.1%	Too few images included in refinement or initial cell. parameters are inaccurate.	Repeat the cell refinement including additional images
12	Cell refinement does not complete due to inaccurate prediction of reflections in the second (or subsequent) wedge of data	Crystal orientation has changed more than a few tenths of a degree during data collection.	Examine the reflection prediction for the first image in each wedge of data. If the prediction is a poor match to the spots on the image, skip the cell refinement step and integrate the images as far as the first image in the final wedge used for cell refinement. This will provide initial values for the crystal orientation and subsequent cell refinement should be successful. Alternatively, select the pattern matching orientation refinement from the Advanced refinement tab of Processing options. This will carry out an orientation refinement for the first image of each wedge prior to integrating the images in that wedge.

Step	Problem	Possible reason	Solution
			This can correct for initial errors of up to one degree in the crystal orientation.
12	Crystal mosaicity refines to a value close to zero on the final cycle of cell refinement	Inaccurate initial cell parameters.	Estimate the mosaicity by finding a value that gives a good match between the number of predicted and observed reflections on the first image, and then fixed at that value by checking the "Fix" tick box. Once the cell refinement has been completed, it should be possible to refine the mosaicity.
12	Detector tilt and twist parameters vary substantially (more than 0.1-0.2°) from one image to the next	All spots in the outer regions of the detector are very weak, so the tilt and twist parameters are poorly defined.	Set the tilt and twist parameters to their average value in the Detector tab of Experiment settings and fix them (check the "Fix" tick box).
12	Tilt and twist parameters refine (stably) to values greater than 1°	Incorrect indexing solution, for example, the selected solution is orthorhombic but the true solution is monoclinic with a β angle close to 90°.	Select a solution with lower symmetry from the Indexing panel and the refinement repeated.
12	The RMS residual increases substantially (by more than 25% of initial value) in later cycles of refinement	Refinement is not working correctly. This may be due to selecting the wrong indexing solution or because the effective resolution of the data is less than 3.5Å.	If the indexing solution appears correct, refine the cell using the Indexing task (step 12(B)) rather than post-refinement.
12	Cell refinement fails to complete and there is a large number of reflections (>20%) that are spatially overlapping	Too few reflections are available to give stable integration of the images, as a result of the number of overlapped reflections.	Reduce the minimum spot separation (Processing tab of Processing options) by a small amount (as small as possible, not more than 25%) to reduce the number of overlapped reflections..
12 or 20	Detector parameters (tilt, twist, detector distance, YSCALE) refine to physically unreasonable values, resulting in a warning in a pop-up window	Incorrect indexing solution. Serious radiation damage (can result in large change in apparent detector distance).	Unless the indexing solution is known to be correct, select the option in the pop-up window to reset the parameters to their initial values. Failure to do so may result in the failure of subsequent steps (indexing, integration). Note that choosing the Reset option will set the direct beam parameters to the values read from the image header and the missetting angles for all images will be set to zero.
20	The average spot profile or some of the standard profiles are either not well defined (fuzzy/split) or not located centrally within the peak region (outlined in blue). See Supplementary Tutorial Example 9.	Fuzzy or split profiles can be the result of poor spot shape due to defects in the crystal. Profiles not located centrally suggests errors in cell parameters or crystal orientation, or that the refinement of detector or crystal parameters is unstable.	Examine the images to see if individual spots are also fuzzy/split. If they are not, this suggests errors in the unit cell parameters, crystal orientation or detector parameters. For this situation, or when profiles are not centrally located, try alternative indexing solutions.
20	Standard profiles are outlined in red rather than in blue	Profiles outlined in red are those that were initially poorly determined, and spots closer to the centre of the detector have been included to improve the definition of the profile.	If many of the standard profiles have been averaged in this way, then this may indicate that the outer resolution limit is too high. The image should be inspected to see if the high-resolution limit is appropriate. This would normally be set to a value just beyond where any spots are visible in the image (after optimising the zoom level and contrast). For fine-sliced data it may be necessary to sum images when estimating the resolution limit (see step 2).
20	Spots are very close together and the peak regions of the standard profiles include some of the adjacent spots. See Supplementary Tutorial Example 8.	Adjacent spots are not completely resolved on the detector, resulting from a very long cell dimension or very elongated spot shape.	Increase the profile tolerance values via the "Advanced Integration" tab of Processing options. Default values of 0.02 (at low resolution) and 0.03 (high resolution) can be increased up to ~0.05–0.06 in challenging cases. In very challenging cases it may not be possible to avoid this situation.
22	A substantial number of reflections (>5%) are flagged as "too wide in phi" (green prediction boxes)	Either a large crystal mosaicity or a small mosaic block size.	Increase the maximum reflection width should from the default value of 5° to 10° or if necessary 15° in the Advanced integration tab of Processing options
23	Processing is not well behaved, for example, unstable or excessively large	Incorrect indexing solution.	Select a different indexing solution or change the parameters used for indexing (e.g., images used, direct beam position, number of reflections included).

Step	Problem	Possible reason	Solution
	tilt and twist detector parameters, mosaic spread refining to zero or unstable crystal orientation refinement		
23	More than 10–20 Bad Spots per image	Many possible reasons.	Examine the warning messages (step 21), double click on the warning to get more information and advice.
23	Large number of spatially overlapped reflections leading to a substantial loss in data completeness when the data is scaled and merged	Mosaic spread is large, or has incorrectly refined to a value that is too large. Alternatively, spots are not completely resolved on the detector (usually due to a large unit cell parameter).	Check that the refined mosaicity is correct by examining the predicted reflections, if necessary reduce the mosaicity and fix it. Alternatively, reduce the minimum spot separation via the "Processing" tab of Processing options. This needs to be done with caution, as reducing the minimum separation to a value less than the actual spot size will result in systematic errors in the integrated intensities; in severe cases, the program will fail to integrate the data. In some cases, increasing the Profiles Tolerance values in the Advanced integration tab of Processing options will result in reducing the minimum spot separation determined by the program. Values above 0.06 should be avoided.
24	Correlation coefficient for the identity operator in POINTLESS is low (<0.3) (Supplementary Tutorial Examples 3, 8)	Images are misindexed. This is most likely to happen if one or more of the unit cell parameters are very large and the direct beam coordinates are inaccurate.	Return to the indexing step and carry out a direct beam search as described in step 4(A)i. The solution with the smallest value of $\sigma(x,y)$ should be correct, but the difference may not be very large between correct and incorrect indexing solutions. If necessary, integration should be repeated with different solutions, using the correlation coefficient for the identity operator to identify the correct solution.