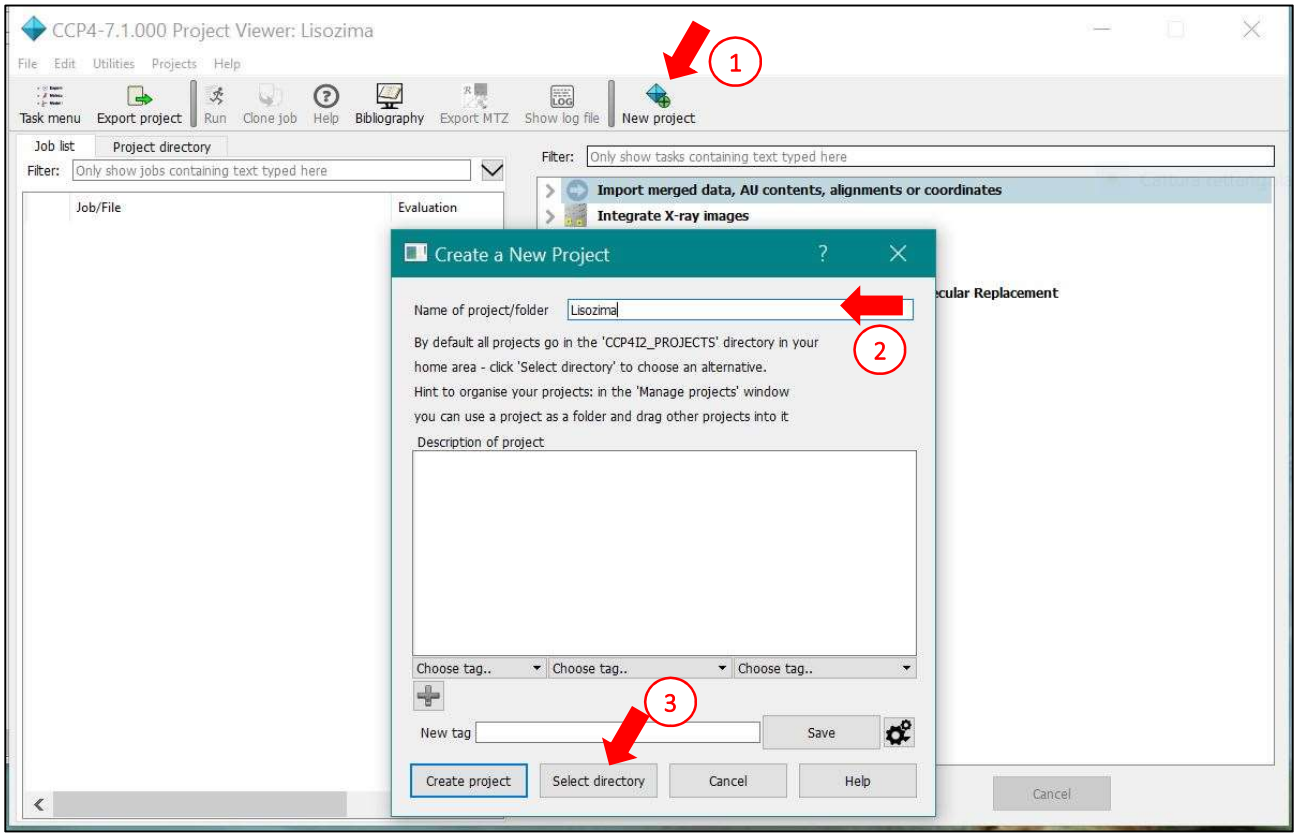
# DATA REDUCTION FROM SINGLE CRYSTAL DATA COLLECTION: INDEXING, INTEGRATION AND SCALING

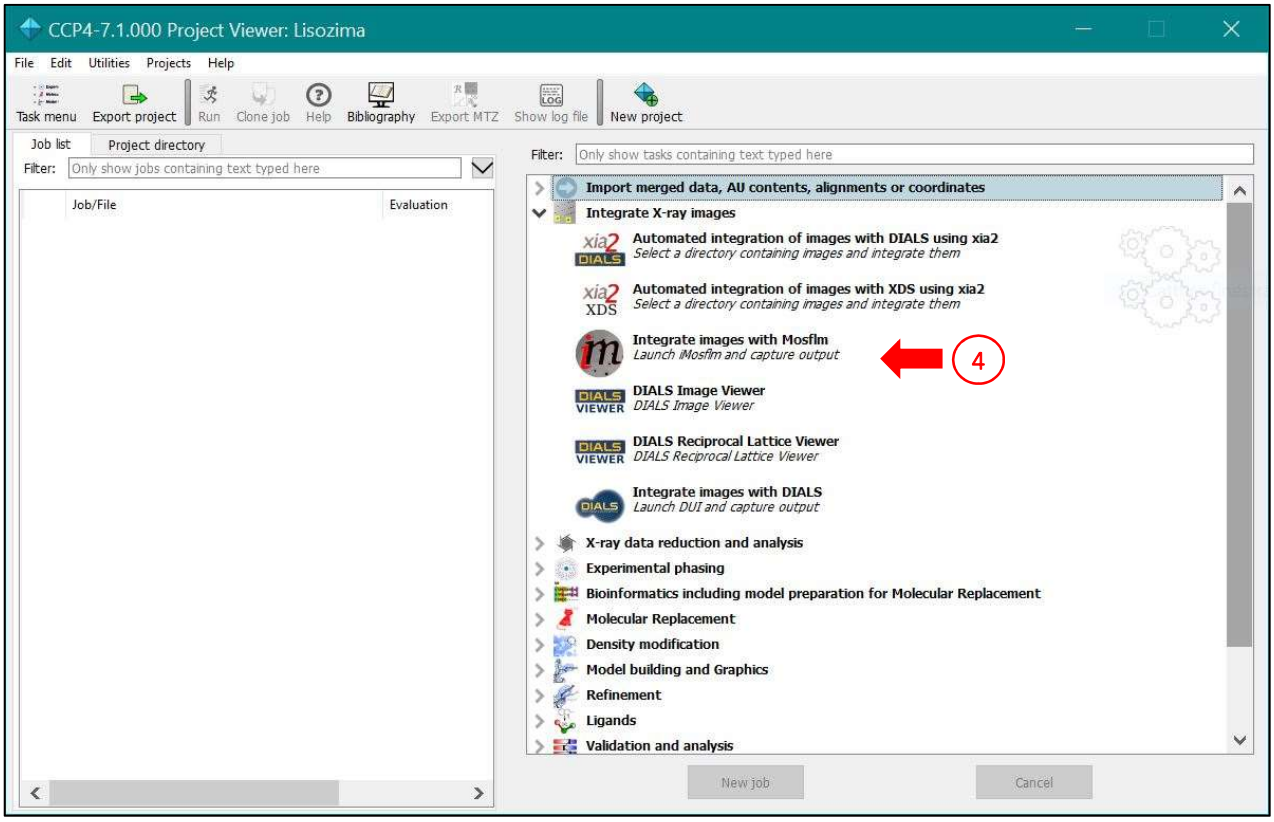
A data collection using synchrotron radiation yielded diffraction images corresponding to rotations of the crystal along the φ, axis, perpendicular to the incident radiation. Data will be indexed and integrated using the Mosflm software [1], under the CCP4 suite interface [2]. After integration, systematic absences will be analyzed to identify the space group, manually or using the Pointless software [3]. Scaling in the correct space group will be performed using the software Aimless [4] of the CCP4 suite.

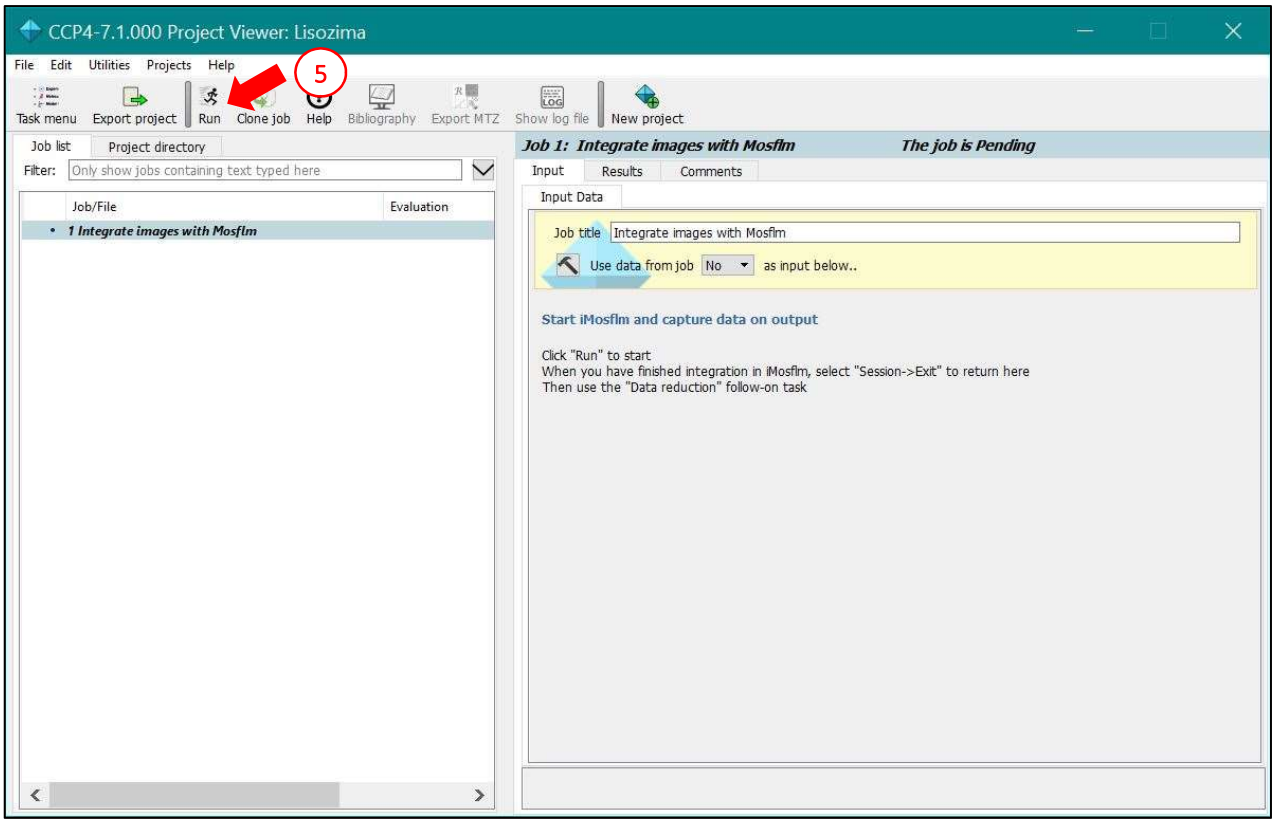
After opening the software CCP4i2 (new graphical interface), first a new project should be created (**1**). The name of the project will be selected (**2**) and the user should choose the working directory where input and output files will be saved by the software (**3**).



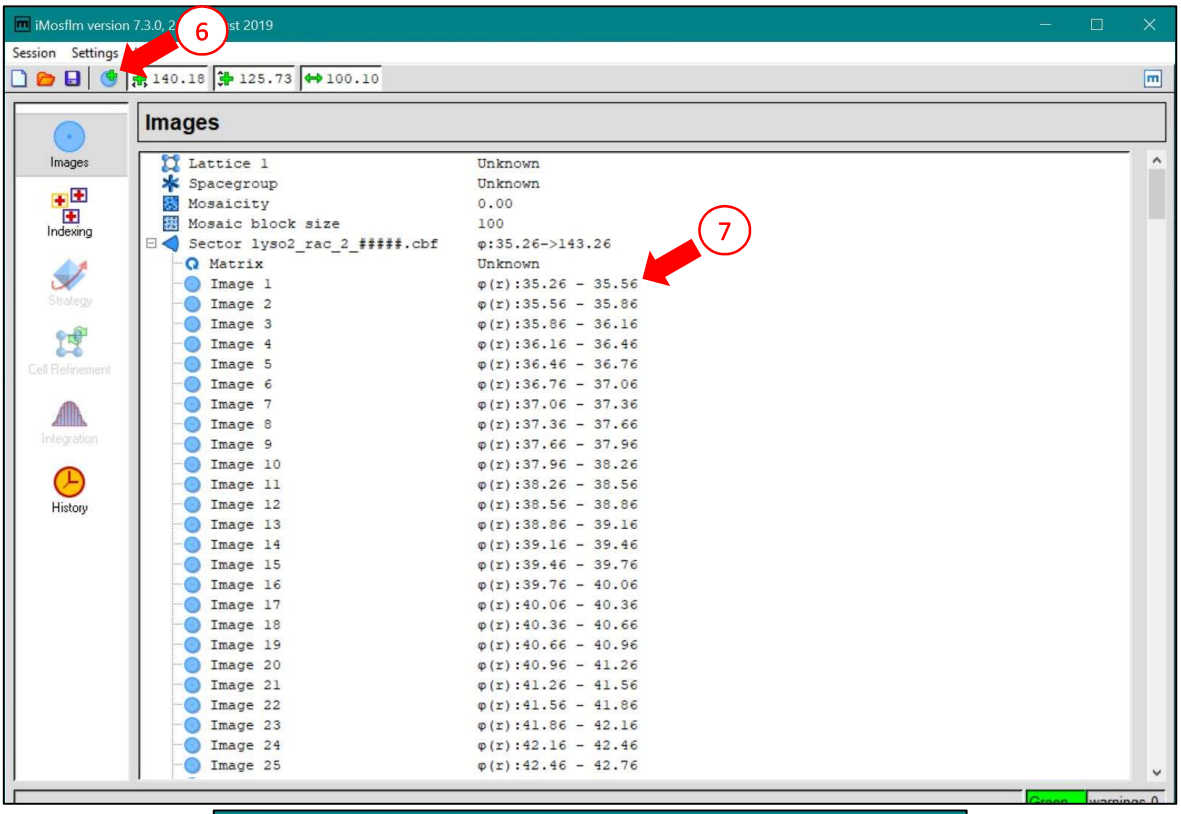
## Indexing and integration of diffraction spots.

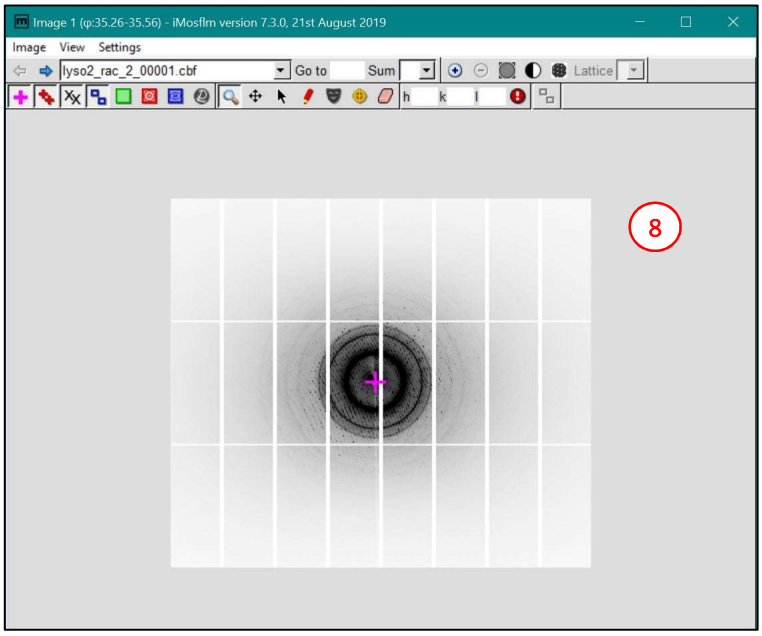
In the window of the previously created project, the Mosflm software is selected from the list on the right (under “Integrate X-ray images”) (**4**). If the task menu is not available in the main project window, the user can recall it at any time by pressing the “Task menu” button on the upper bar, left side. Once the software has been selected by double-click, to lunch it click “Run” (**5**) on the upper bar.



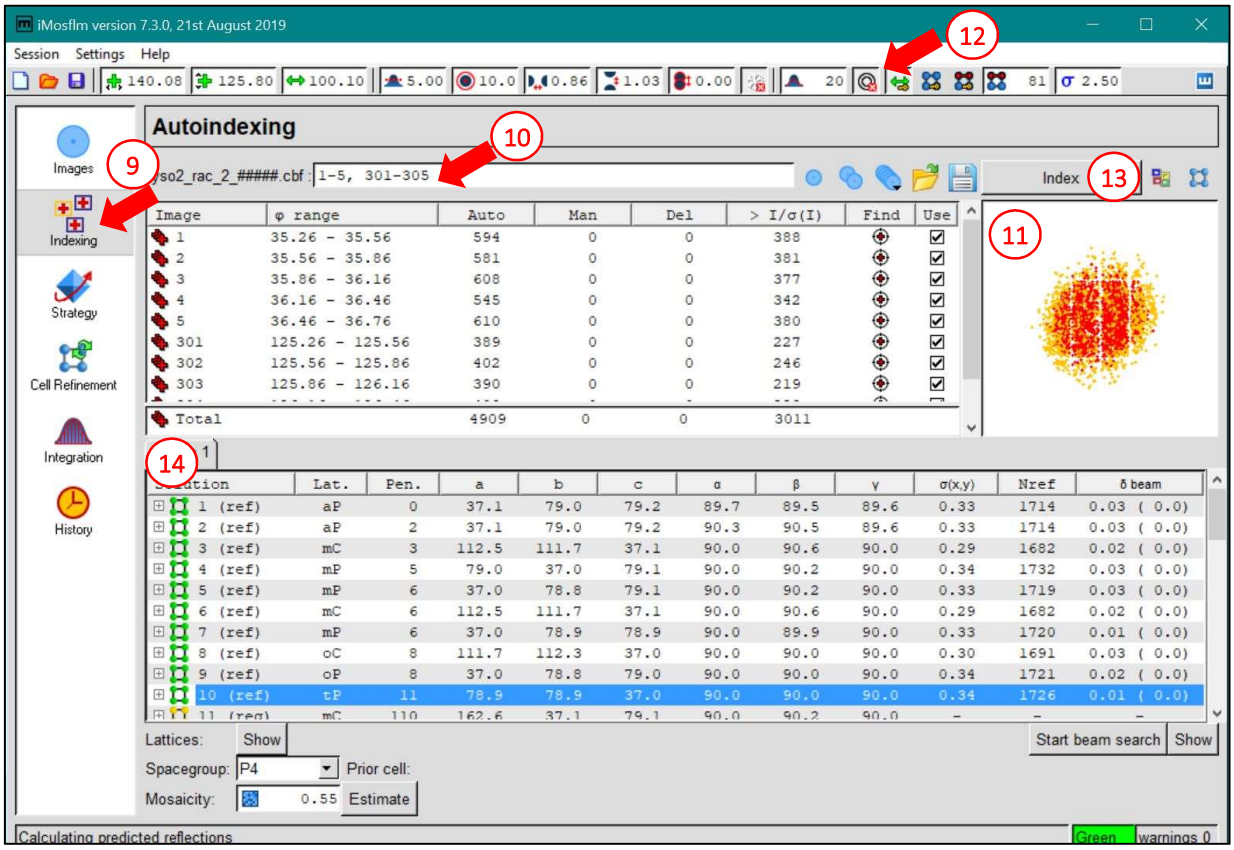



A new window will open, the main window of the Mosflm software. Images to be processed can be selected using the “Add images” icon (**6**). If the images are stored in a single directory, selecting a single image will be sufficient and the software will automatically upload all the images of the data collection. In the list that will appear in the main window, each image is followed by the Δφ interval (**7**). At the same time, a second window will automatically open allowing to manually analyze the diffraction images (**8**). In the image window, the user can see the selected spots using in the following processing (small red crosses) and the predicted position of the spots obtained during indexing, squares of different colors: blue for the full spots, yellow for the partials, red for spots discarded due to overlapping, and green for spots discarded due to too wide Δφ interval. In this window, the user can also manually correct the position of the main beam (large purple cross) and set the resolution limits for data analysis (blue circles).

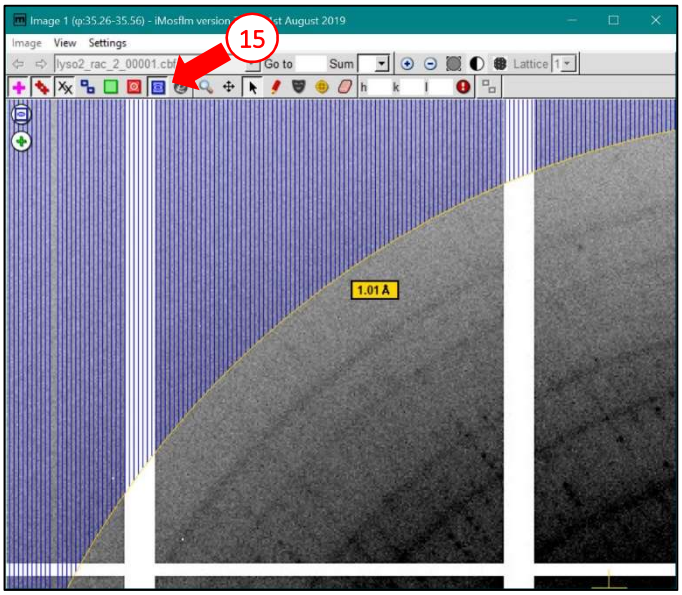


In the following tab of the software, selected from the left column (**9**), some images can be selected (**10**) for the initial determination of the unit cell. Usually, 10 images are sufficient, ideally selected at different diffraction angles. The software automatically selects the diffraction spots from the images (**11**). When present, rings due to the presence of crystalline ice can be excluded from analysis, by clicking the specific button from the main window (**12**). Presence of ice rings is indicative of a non-ideal cryoprotection of the crystal before freezing. The "index" button allows the start of the indexing process (**13**).

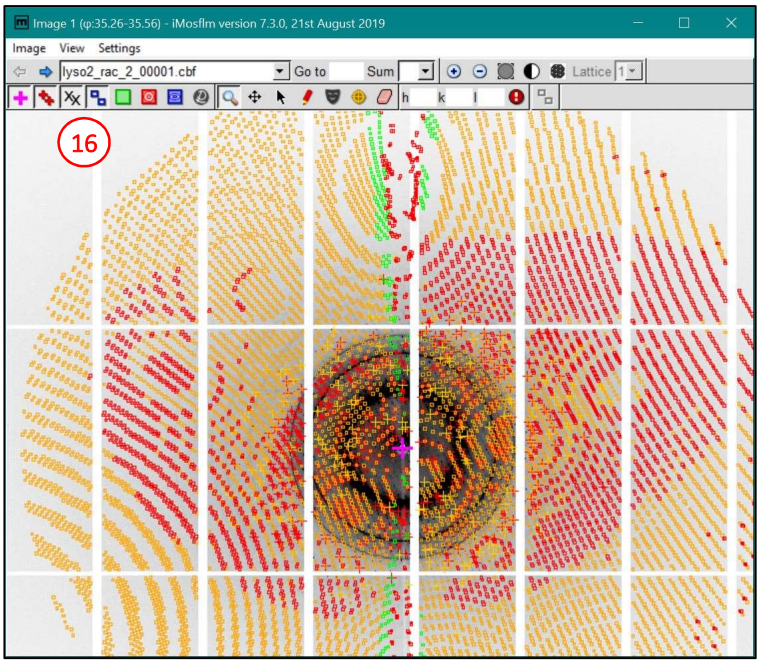



At the end of the indexing, the software shows a list of possible unit cells (**14**), with the respective Bravais lattice (a = triclinic, m = monoclinic, o = orthorhombic, t = tetragonal, h = hexagonal, c = cubic; P = primitive, C = base-centered, F = face-centered, I = body centered, R = rhombohedral centering), cell parameters, and penalty. The penalty value indicates how many spots do not match the proposed unit cell. When selecting the unit cell, the user should prefer higher symmetry cells (lower in the list) among those marked by the software with a green color, indicating a satisfactory indexing. Structure solution can be performed also in a lower symmetry option, but a higher number of parameters would reduce the refinement quality. If a wrong higher symmetry is selected at this stage, the scaling will yield a high $R_{merge}$ value. In this second case, the indexing should be done again decreasing the symmetry of the unit cell.

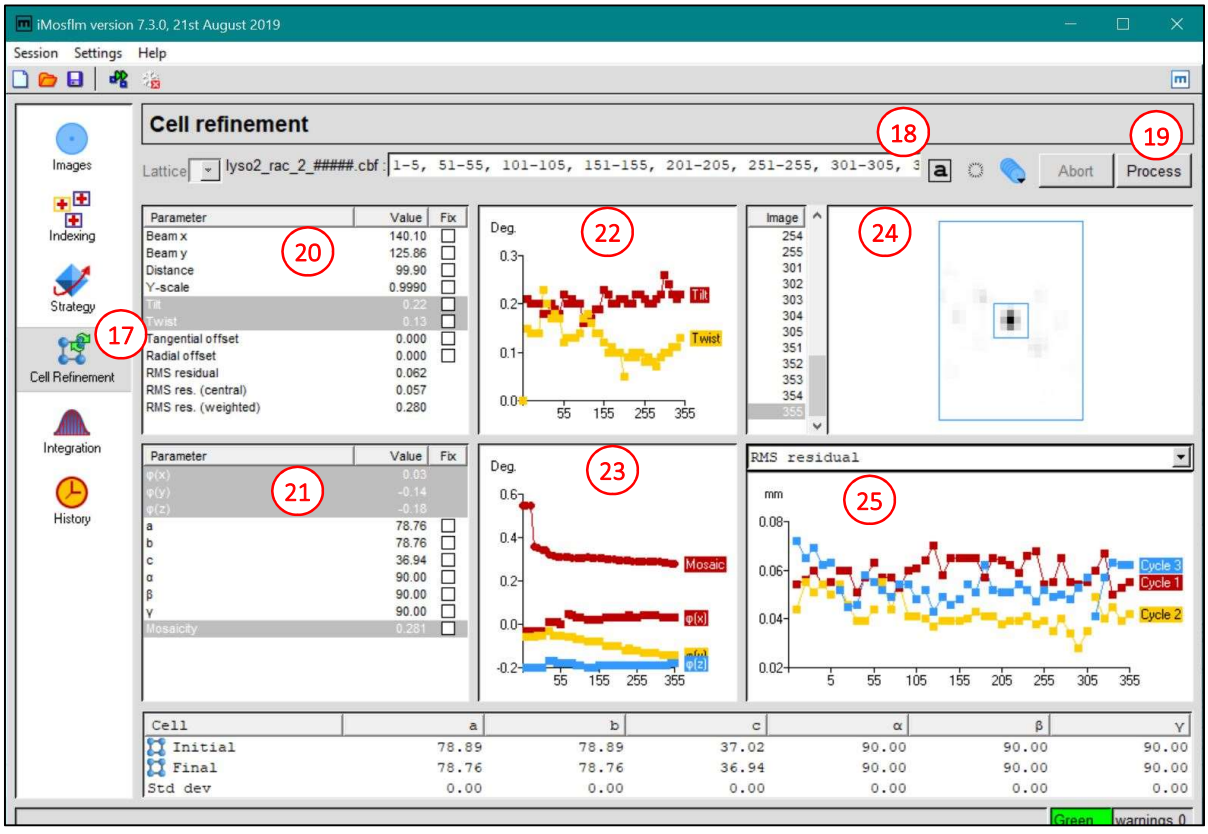



In the diffraction image window, the user can limit the resolution using the button (**15**) and moving the external circle to modify the highest resolution value (to move the circle, the arrow button on the same bar should be selected).

Once the images have been indexed and the unit cell parameters have been selected from the list, the image window shows the predicted positions of the diffraction spots (**16**), with the mentioned color code.

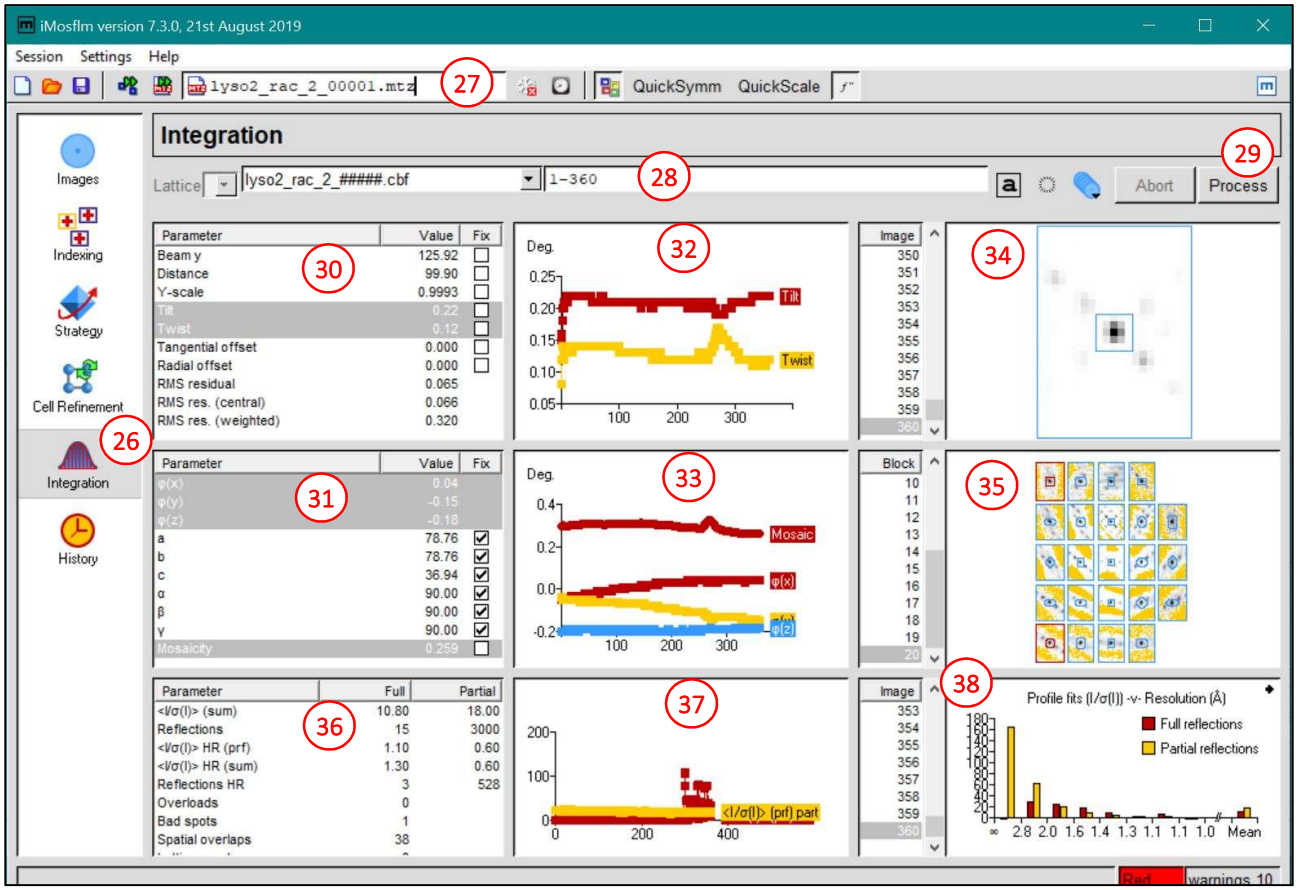



Once the indexing is complete, the tab "Cell refinement" (**17**) can be selected from the main window. This window allows to refine the values of the unit cell parameters using a larger number of diffraction images (**18**). After the process is started with the suitable button (**19**), the software starts to calculate optimized values of parameters in order to reduce the Root Mean Square residuals ("RMS res.") between measured and calculated spot positions. The **20** and **21** panels allow to observe the variation of each parameter while the optimization is carried on. Graphs **22** and **23** show a graphical representation of the change of some parameters along with image number: "Tilt" e "Twist" in panel **22** are detector positioning parameters, while panel **23** shows mosaicity and orientation of the crystal. During the refinement, the spot position determined from indexing is optimized to cover the maximum intensity registered on the diffraction image (**24**). Finally, graph **25** shows the behavior of the RMS residues for the selected images.




In the following "Integration" tab (**26**), the software allows integration of the diffraction data. In this window the user can select the output file name (**27**), containing the integrated diffraction intensity corresponding to each reflection identified from Miller indexes *hkl*, and the number of images to be integrated (**28**). If no radiation damage is present during data collection, usually all collected images are integrated. When

integration process starts (**29**), the software automatically performs a new optimization of the image parameters using a smaller subset of subsequent images, showing the same parameter variations as in the previous cell refinement (panels **30** and **31**), relative graphs for subsequent images (panel **32** and **33**), and the integration profiles of diffraction spots with the respective background (panels **34** and **35**). Panel **36** offers some indexes to monitor the behavior of the integration, among which the value of $\langle I/\sigma_I \rangle$. The same value for the single images is also reported in a graph (**37**) for both the full (red line) and partial (yellow line) reflections. A significant reduction of the $\langle I/\sigma_I \rangle$ value may indicate a signal deterioration, often due to the presence of radiation damage. Finally, graph **38** allows to evaluate the value of $\langle I/\sigma_I \rangle$ in the resolution shells, for both full and partial reflections.

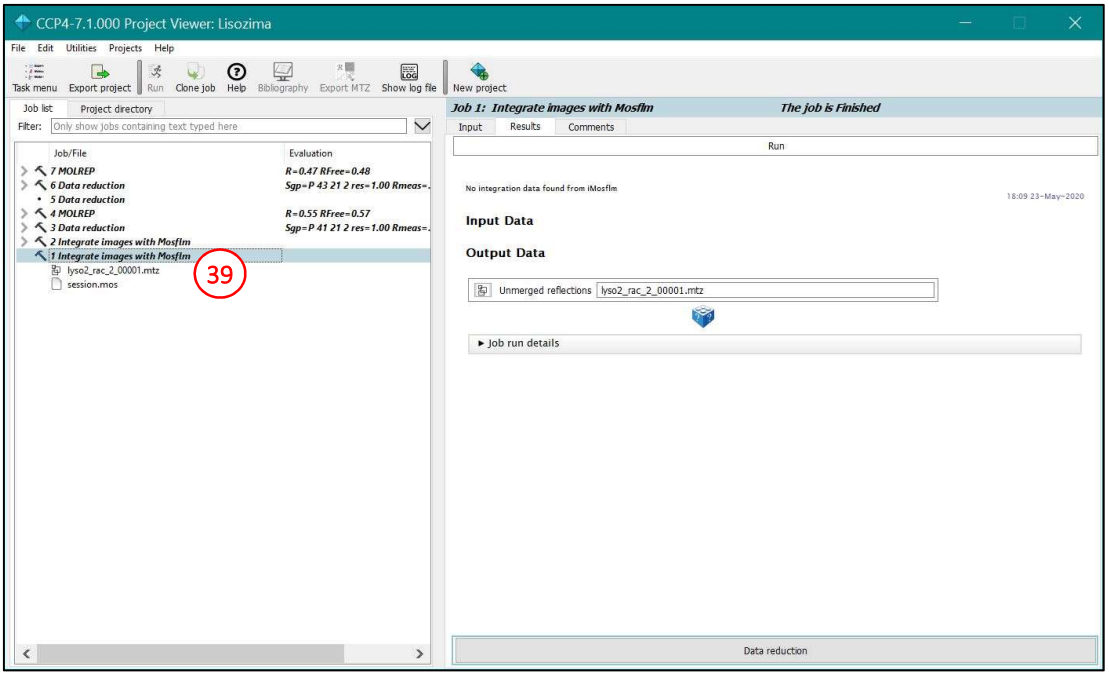



At the end of the integration process, the software provides an *mtz* output file containing intensity values and standard deviations for each of the integrated diffraction spots. In addition, the software provides a second file, with extension *lp*, containing the details of the processing and the results for each image. The *lp* file is very useful when the integration process did not yield the expected results and the user should analyze it carefully before re-running the protocol.

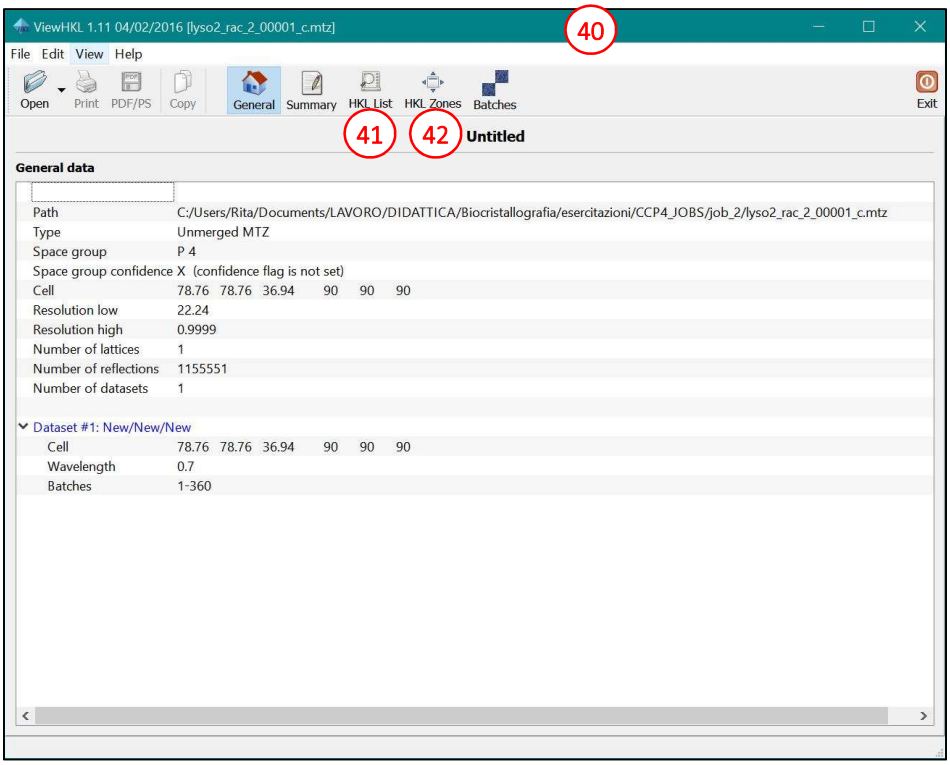
After integration, the Mosflm software can be closed. Data are saved in a subfolder of the main folder of the project (selected at the beginning).

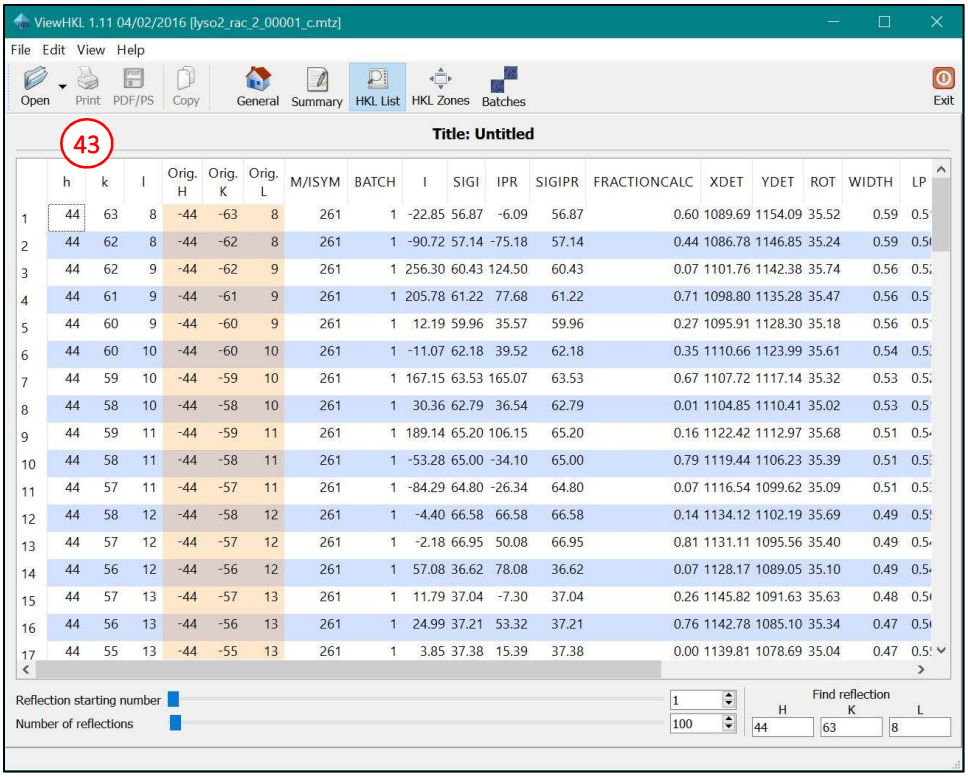
## Analysis of systematic absences.

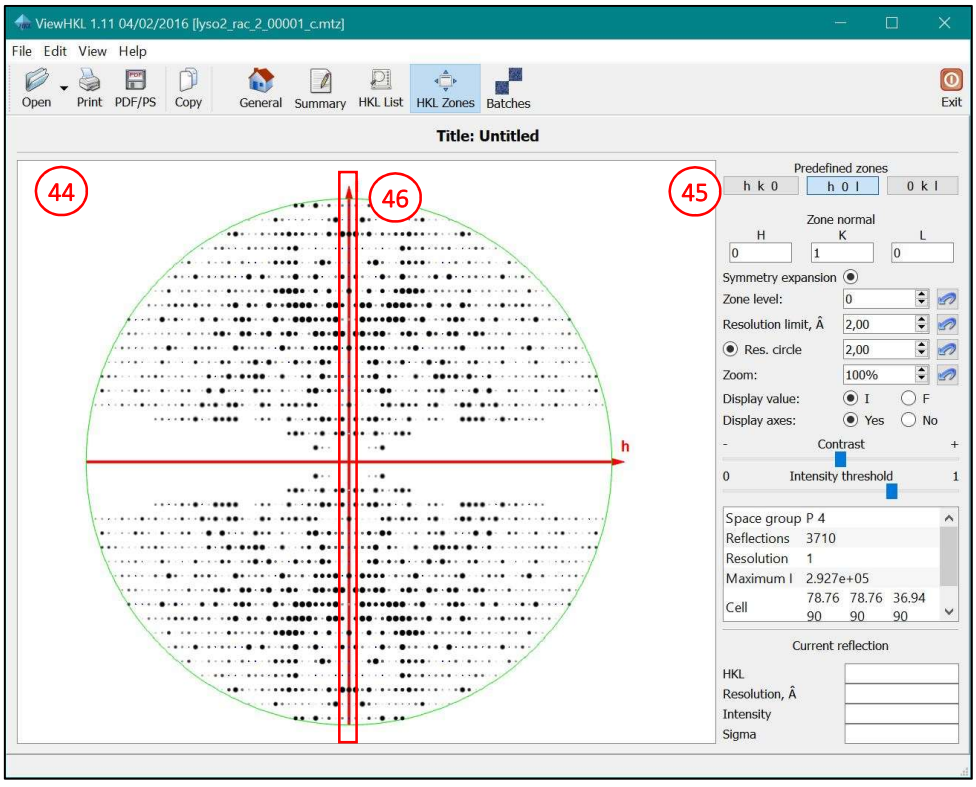
Before starting the scaling protocol, it is advisable to analyze the file containing integrated reflections (with *mtz* extension). This file can be opened from the main CCP4i2 interface by clicking on the specific job in the left panel of the main window to open the list of output and input files of the job (**39**). Double-click on the *mtz* file opens it.




The file is opened by another software of the CCP4 suite, ViewHKL (**40**), that allows visualization and analysis of the contents of the *mtz* file. Two additional buttons in the window of the ViewHKL software allow to visualize the list of reflections (**41** and **43**) and the sections of the reciprocal lattice showing the intensities of the nodes (**42** and **44**). In the latter window, a single section can be selected and analyzed with the buttons on the right (**45**).

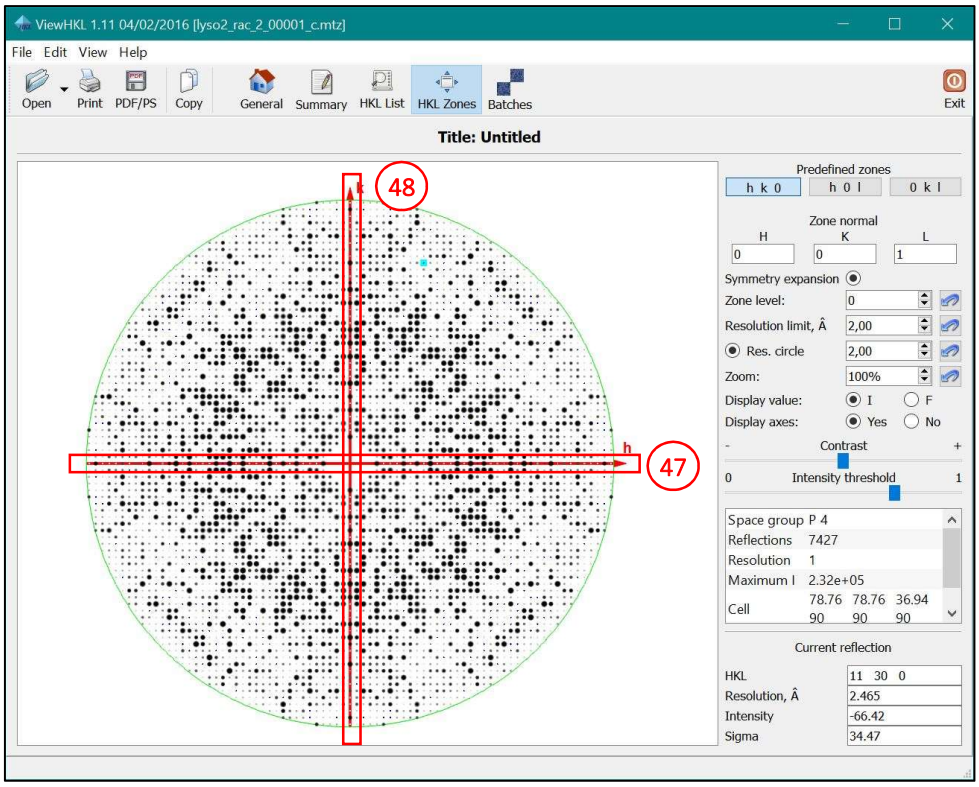





In this exemple, reflections with $h$ = 0 and $k$ = 0 are present only when $l$ is a multiple of 4 (**46**). This kind of systematic absence is indicative of the presence in the unit cell of an element of roto-translational symmetry $4_1$ or $4_3$ along the unique axis of the tetragonal system ($c$ axis).

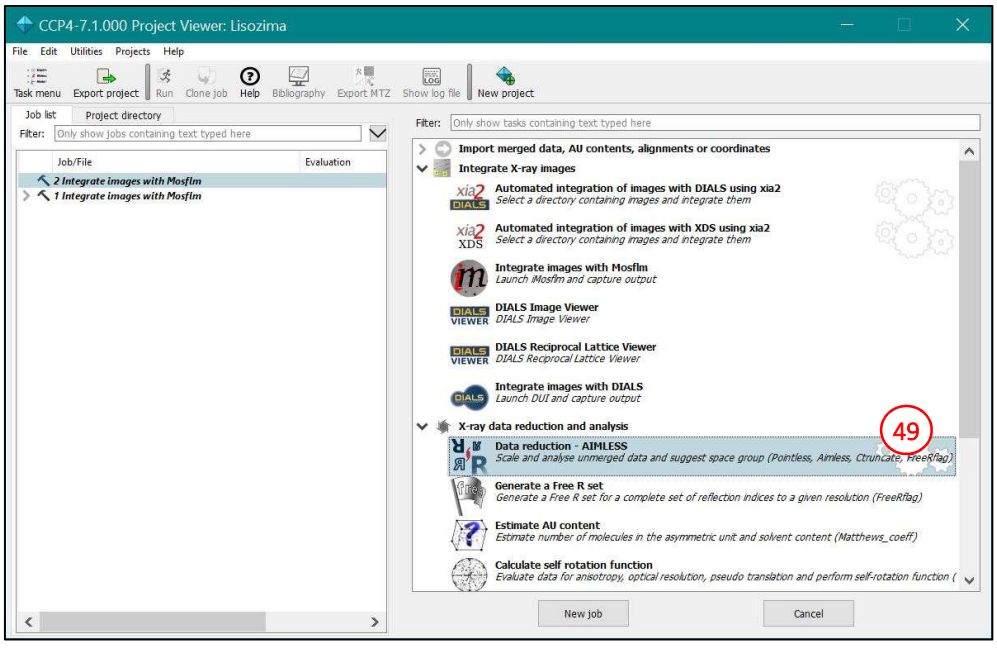
A similar analysis can be performed for systematic absences on the other 2 reciprocal axes, $h$ (with $k$ = 0 and $l$ = 0, **47**) and $k$ (with $h$ = 0 and $l$ = 0, **48**), showing the presence of axes of roto-translational symmetry of order 2, namely $2_1$, along axes $a$ and $b$ in the real space.

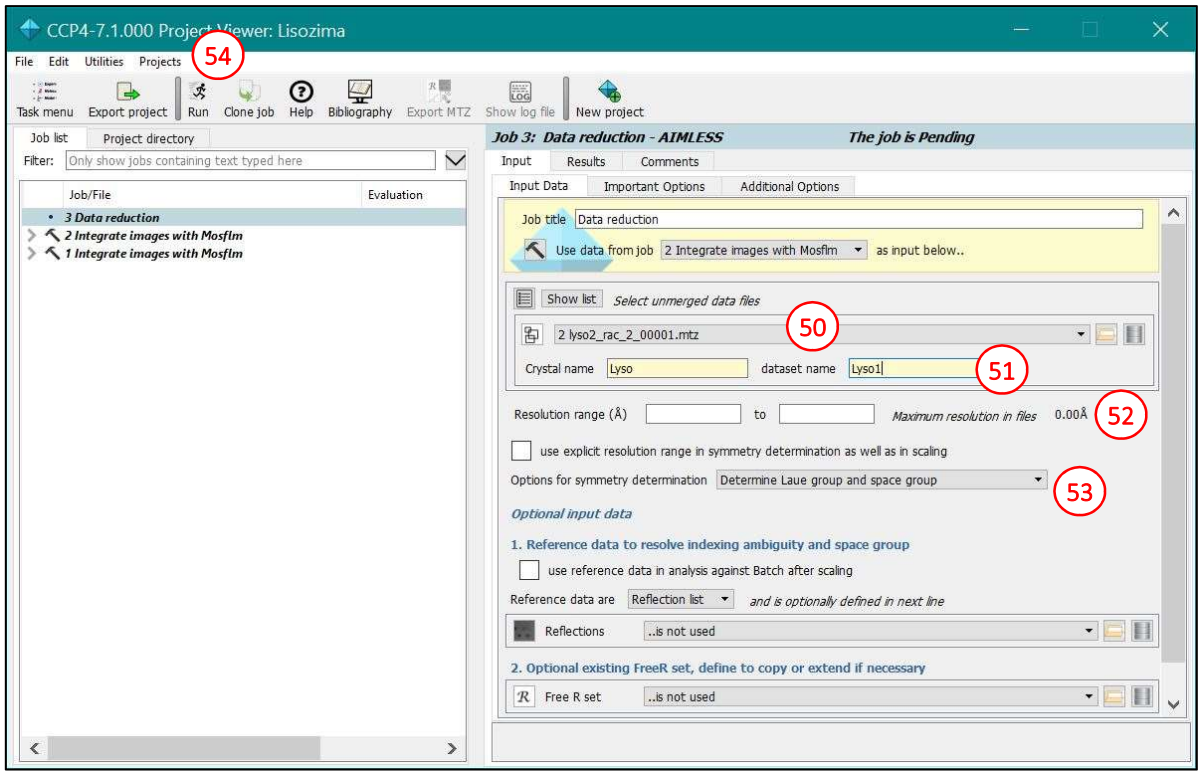



This analysis can be performed automatically using the software Pointless [3] of the CCP4 suite. The user should be aware, however, of the possible limitations of the selection, such as the intrinsic ambiguity between enantiomorphic space groups.

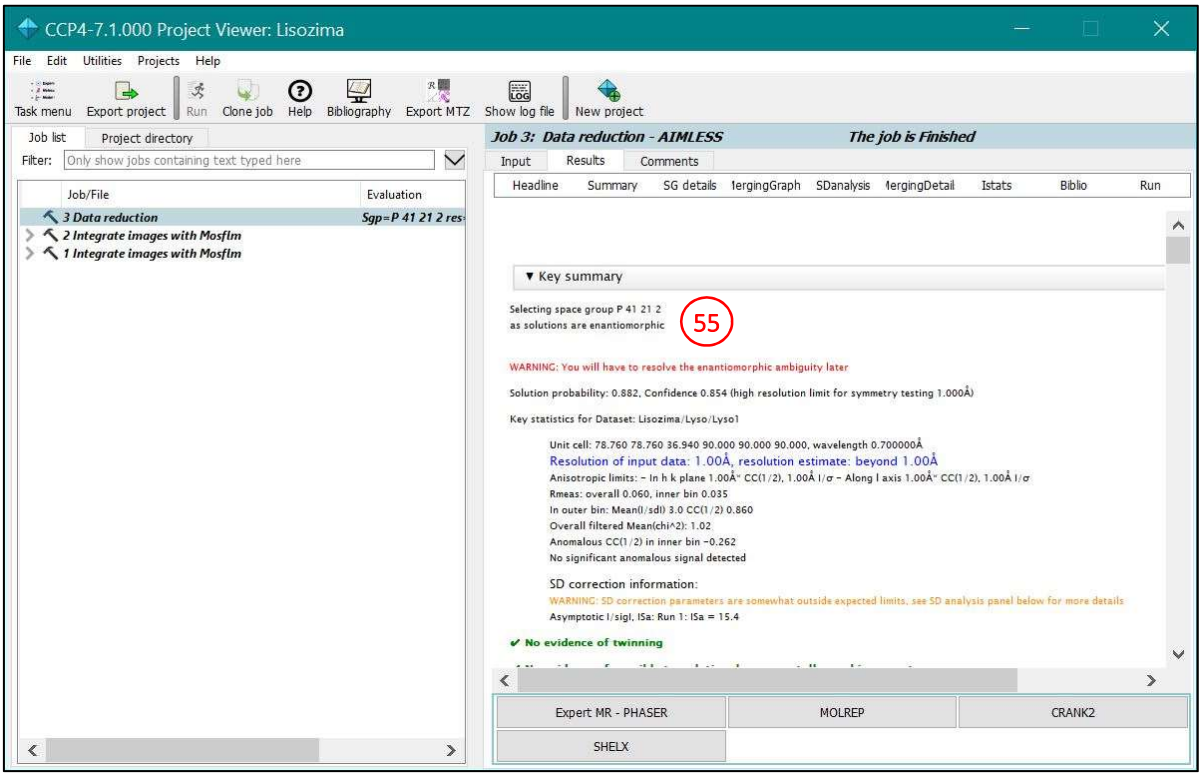
## Scaling and merging of diffraction data.

The next step in image analysis is the scaling of diffraction data, the Aimless software can be selected from the main CCP4 interface (**49**). When the software is selected, the input file can be specified (**50**), together with the name of the dataset (**51**). The user can select further scaling options, such as resolution limits of the data to be scaled (**52**) and the space group (**53**), when previously known. Once parameters are selected, the software can be started with the "Run" button (**54**).

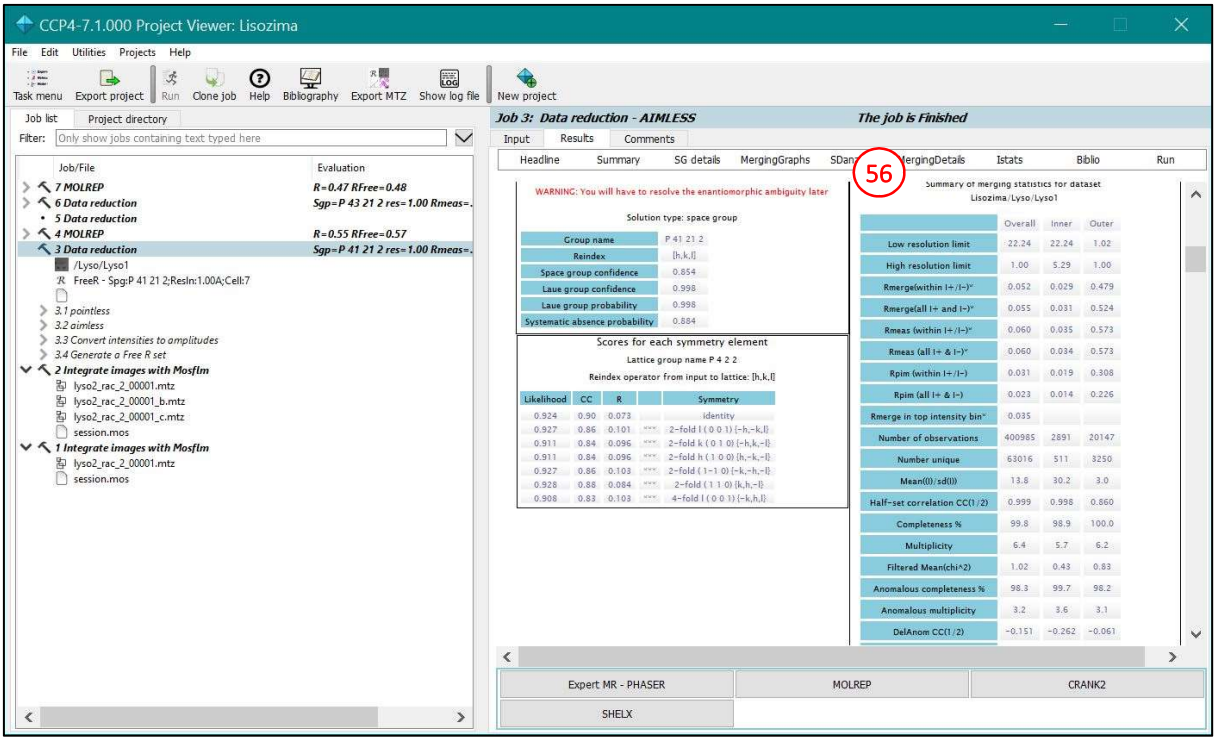


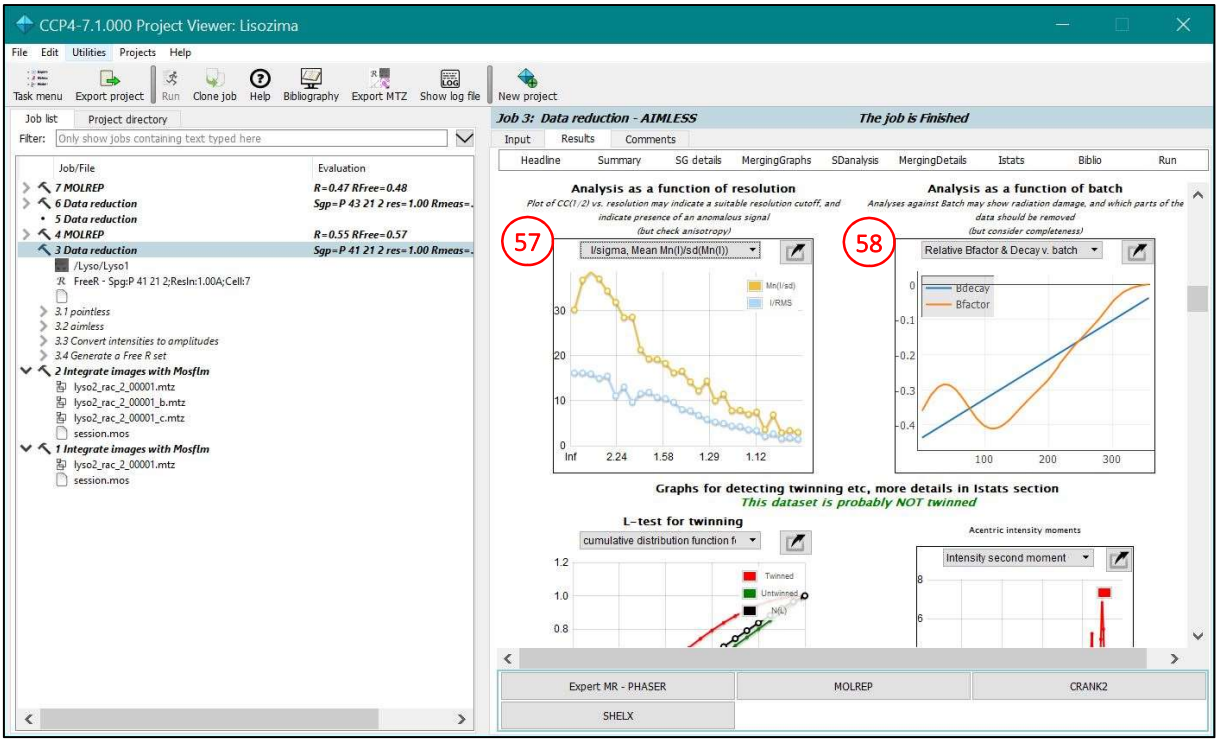



Once the scaling process is complete, the software provides the output reflection file, together with the space group selected by the Pointless software (automatically run by the Aimless protocol) (**55**) and the scaling statistics. For the lysozyme crystal analyzed in this example, the selected space group (*P* $4_1$ $2_1$ 2) has an enantiomorphic group (*P* $4_3$ $2_1$ 2). The choice of the correct space group is not possible at the scaling level, because the two possible groups have the same symmetry relations in the reciprocal space. When enantiomorphic space groups are present, it is advisable to perform in parallel two data reductions using the alternative groups and test the correctness of each solution in the following phasing step.



The software provides statistical values that allow to evaluate the quality of the scaling processing. These values are summarized in a table (**56**). In addition, analysis of the graphs of statistical parameters against resolution shells (**57**) is crucial to determine the resolution limit of the data collection, to be used in the following steps. Statistical analysis against the progress of data collection (**58**) may indicate radiation damage accumulating in the crystal.



## References


[1] T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell and A.G.W. Leslie, *"IMosflm: a new graphical interface for diffraction-image processing with MOSFLM"*. Acta Cryst. 2011; D67, 271-281.

[2] M. D. Winn et al., *"Overview of the CCP4 suite and current developments"*. Acta Cryst. 2011; D67, 235-242

[3] P.R. Evans, *"An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics"*. Acta Cryst. 2011; D67(Pt 4), 282-92.

[4] P.R. Evans, and G.N. Murshudov, *"How Good Are My Data and What Is the Resolution?"*. Acta Cryst. 2013; D69(Pt 7), 1204-14.