Tutorial

Data processing with iMosflm

The CCP4 suite of programs

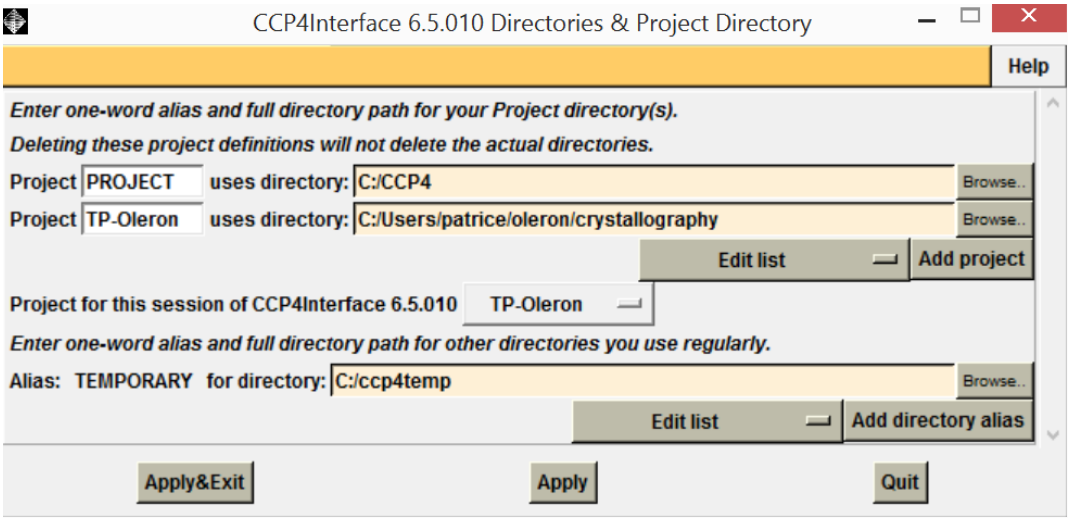
CCP4 provides software for every stage of the crystallographic structure solution process, from processing X-ray diffraction images through to producing images and movies for presentations and publications (documentation available at http://ccp4wiki.org/).

Launch the CCP4 suite: *ccp4i*

The first time you run the interface, the window which enables you to set up directory aliases and define a project directory comes up automatically. Alternatively it can be accessed via the "*Directories & ProjectDir*" button near the top right of the main window.

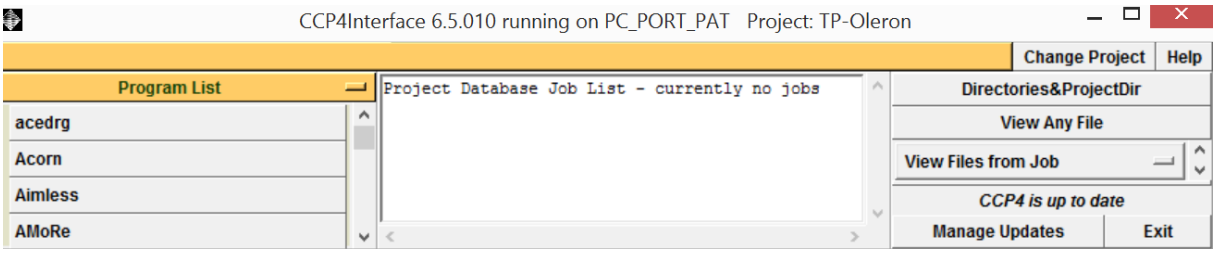
In the new window, select "*Add project*"

Enter the project name (TP-Oleron) and the full directory path in the *"uses directory:*" field (e.g. C:/Users/username/oleron/crystallography).

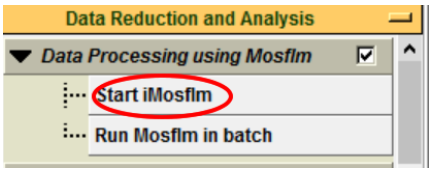


Select *'TP-Oleron*" from "*Project for this session of CCP4 interface*"

Select "*Apply and Exit*". Dismiss the warning message that comes up. The CCP4 main window appears. This is composed of windows, menus and buttons. "TP-Oleron" is written on the top bar.



Click on the gold "*Program List*" drop down menu. Select "*Data reduction and anaysis*"and open the top bullet field "*Data processing using Moslm*". Then click on *"start iMosflm"*



iMOSFLM

This tutorial is based on the full iMosflm tutorial available at "http://www.mrc-lmb.cam.ac.uk/harry/imosflm" and written by Harry Powell, Andrew Leslie and Owen Johnson.

1. Introduction

1.1 Background

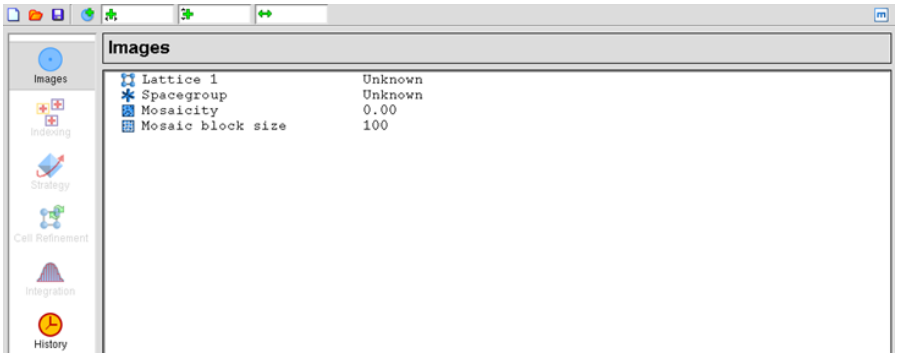
MOSFLM can process diffraction images from a wide range of detectors and produces, as output, an MTZ file of reflection indices with their intensities and standard deviations (and other parameters). This MTZ file is passed onto other programs of the CCP4 program suite (POINTLESS, SORTMTZ, AIMLESS, CTRUNCATE) for further data reduction.

1.2 Aim of the tutorial

Your task is to process 164 images, "prefix_1_00001.img" to "prefix_1_00164.img", collected from a lysozyme crystal on a ADSC CCD detector at the ESRF beamline FIP-BM30A,Grenoble. The 164 images shoud be stored in the "BM30A-2014-11-19/img" folder under your "TP-Oleron" folder (*e.g.* C:/Users/username/oleron/crystallography/BM30A-2014-11-19/img".

2. Overview of iMosflm

Launch the program "*iMosflm*" as previouly described. The following window will appear.

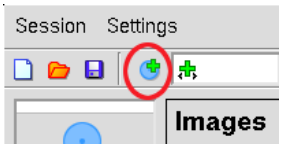


The basic operations listed down the left hand side (*Images, Indexing, Strategy, Cell Refinement, Integration, History*) can be selected by clicking on the appropriate icon.

We will start with the default "Images" window.

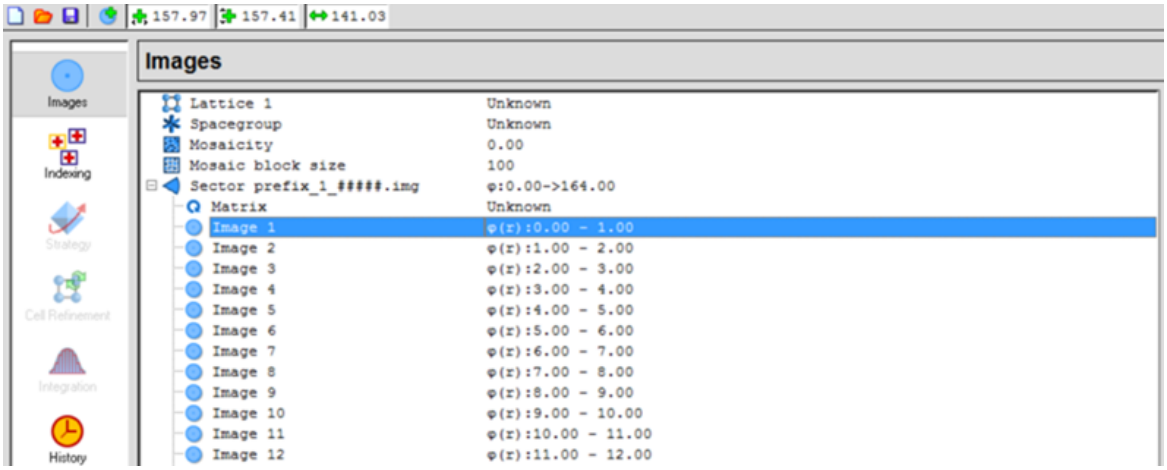
3. Adding images to a session

To add images to a session, use the "*Add images*" icon (circled in red below):



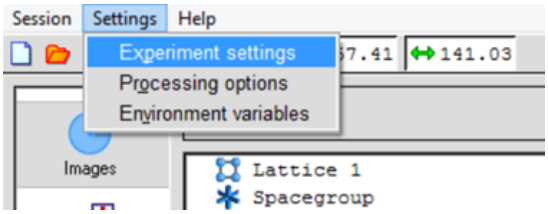
Select the correct directory from the pop-up window. All files with an appropriate extension will be displayed. Double-click on the first image "*prefix_1_00001.img*".

The 164 loaded images will be displayed in the Images window with the start & end phi values displayed (as read from the image header), the beam centre and the crystal-detector distance values.

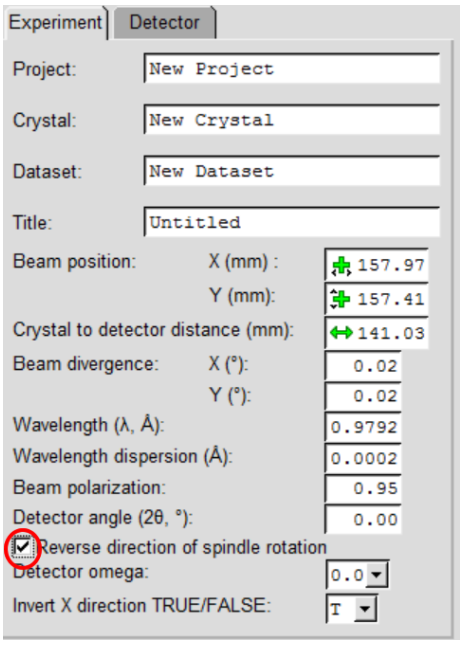


3.1 The FIP-BM30A experimental settings

Because of the particular geometry of the goniometer at FIP-BM30A, you have to open the "*Settings*" drop-down menu and select "*Experiment settings*" as shown below



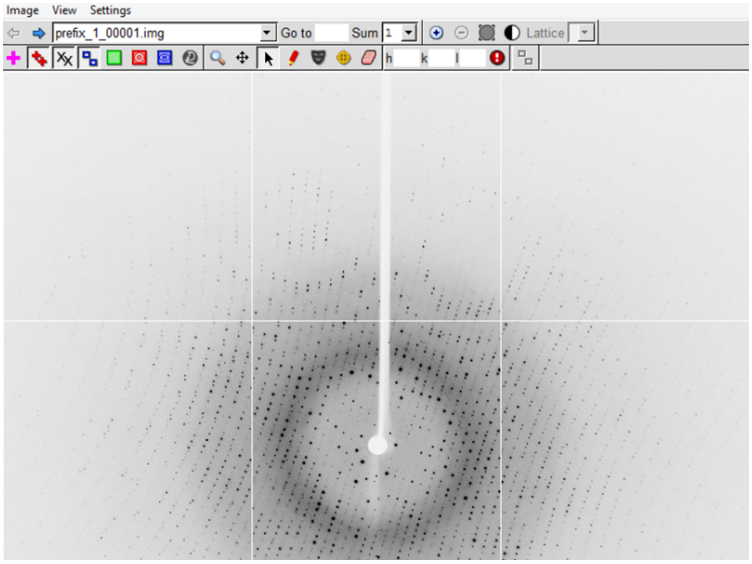
The "*Experiment settings*" window will pop-up and you have to check the "*Reverse direction of spindle rotation*" checkbox (as shown in red below).



Question. Check the used wavelength then close the "Experimental settings" window.

4. Image Display

When images are added to a session, the first image of the sector is displayed in a separate window.

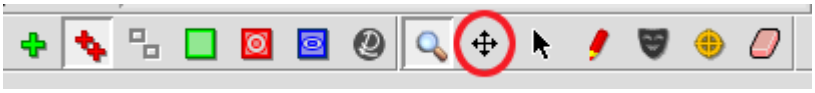


4.1 Resolution limits

Select the Show resolution limits icon (circled in red above). The low and high resolution limits will be displayed. The resolution limits can be changed by dragging the perimeter of the circle with the "left-mouse-button" (**LMB**) (make sure that the "*Selection Tool*" icon has been chosen in the bar).

Question: Check the diffraction limit of the crystal.

4.2 Zooming and Panning



Select a region of the image to be zoomed with the Zoom tool  **.** Then select the Pan tool (circled in red above) and pan the displayed area by holding down the LMB and moving the mouse.

Question: Zoom in on a diffraction spot and check its profile. Check that the detector is divided in pixels, each pixels counting a number of photons.

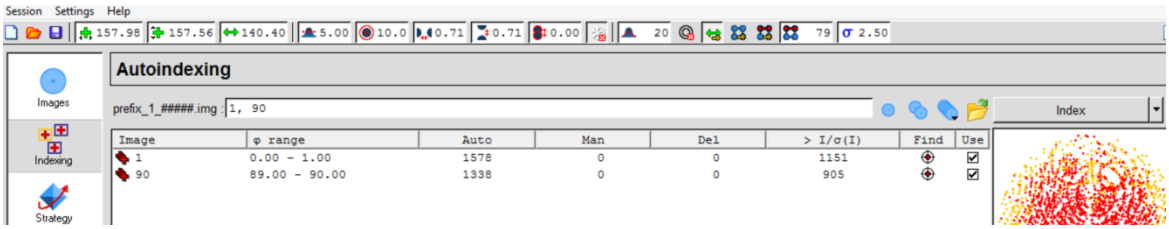
5. Spot finding, indexing and mosaicity estimation

When images have been added, the "*Indexing*" operation becomes accessible (it is no longer greyed-out).

Click on *"Indexing".* This will bring up the major Indexing window in place of the Images window.

5.1 Spot Finding

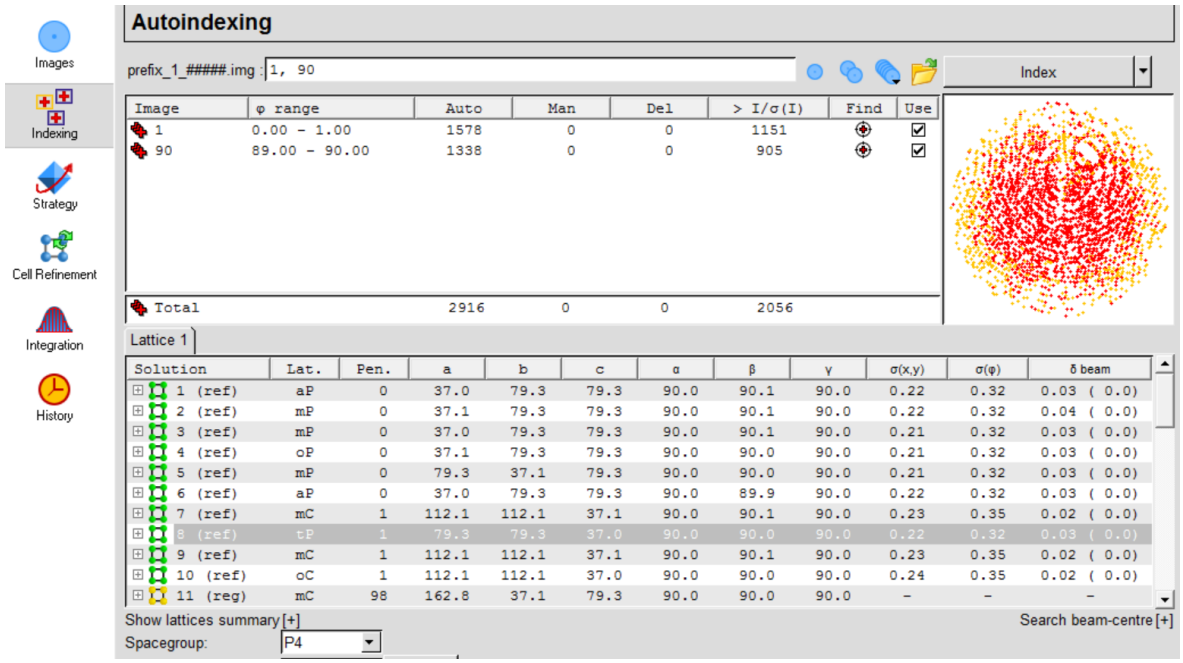
By default, two images 90 degrees apart in phi (or as close to 90 as possible) will be selected and a spot search carried out on both images.



Found spots will be displayed as crosses in the "*Image*" Display window (red for those above the intensity threshold, yellow for those below). The intensity threshold normally defaults to 20.

5.2 Indexing

Providing there are no errors during spot finding, indexing will be carried out automatically



.

The rms deviation (rmsd) or error in predicted spots positions (σ(x,y) in mm) and the rms error in (σ(φ) in degrees) are given for each solution.Usually the penalty will be less than 20 for the correct solution, although it could be higher if there is an error in the direct beam coordinates (or distance/wavelength).

<u>Question</u>. Check that the solution selected by the program (tP bravais lattice) is the best possible with the highest symmetry in the reciprocal space. P4 is the space group selected by default by this bravais lattice: what are the other possible choices (look at the scrolling menu)?

*Check the **image display***. The predicted spot positions for the highlighted solution (P4) will be shown on with the following default colour codes:

Blue: Fully recorded reflection

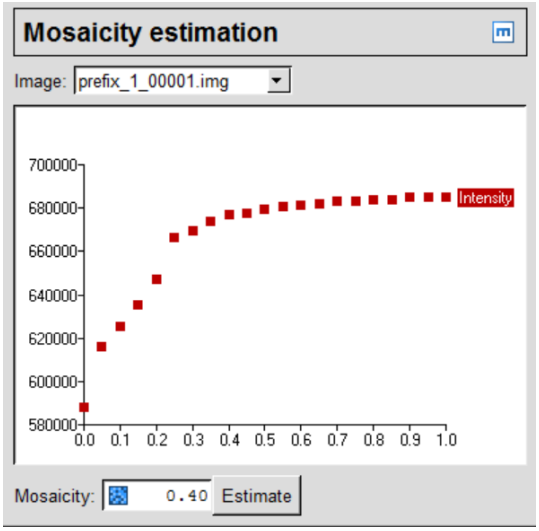
Yellow: Partially recorded reflection

Red: Spatially overlapped reflection... these will NOT be integrated

Green: Reflection width too large (more than 5 degrees)... not integrated.

<u>Question</u>. Select other solutions with a higher penalty in the "Indexing display" and see in the "Image display" how well the predicted patterns match the diffraction image. Go back to the "tP" solution.

5.3 Mosaicity estimation



The mosaicity (mosaic spread) will be estimated automatically for the preferred solution. However, if another solution is chosen, the mosaicity should be estimated again.

Click on the *"Estimate"* button to estimate the mosaicity.

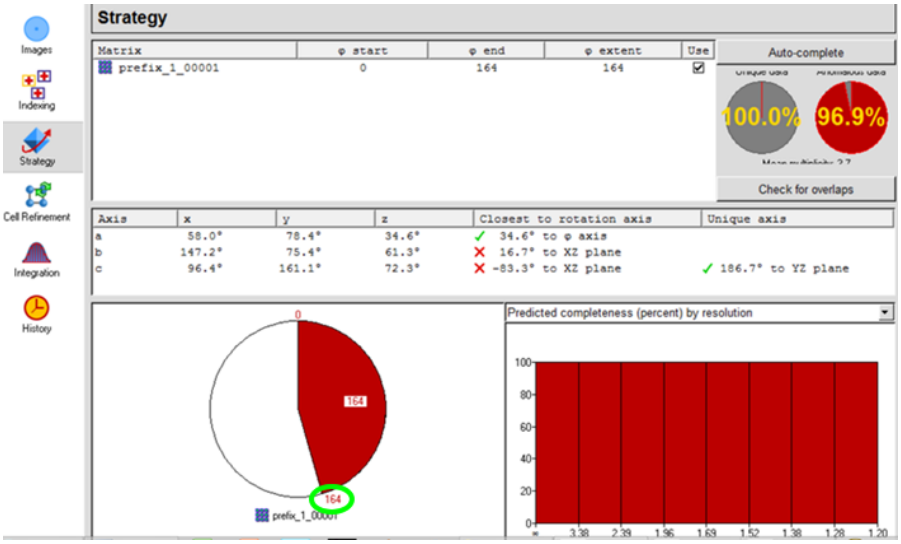
A window will appear which plots the total predicted intensity as a function of mosaic spread, from which an estimate is determined.

Question . Change mosaicity value to 2.0 degrees and check the Image display. Same with a mosaicity value of 0.1degrees. Then re-enter the estimated mosaicity value.

6. Data collection strategy

Once the crystal orientation and (probable) Laue group have been determined, it is possible to calculate a data collection strategy and the Strategy icon is no longer greyed-out.

Select the *"Strategy"* icon. This will open the Strategy window.



The orientation of the crystal, expressed as the angles between the a,b,c unit cell axes and the X,Y,Z coordinate frame, is given. X is along the X-ray beam, Z is the rotation axis (see **Annex**). A warning will be given if the unique axis is so close to the rotation axis that there will be missing cusp data.

The completeness of any given segment of data can be determined manually using the circle at the bottom left of the window: *use the LMB to click on the black box at "164", (circled in green) and drag this around the circle.*

When the LMB is released, the statistics for that sector will be displayed. The statistics are presented in two ways. Near the top right of the window, the overall completeness is shown as red circles for unique data and anomalous pairs. The average multiplicity is given beneath these circles.

Question. Check that the predicted completeness is already good in the range 0 to 60 degrees. Check the "mean multiplicity by segment" by clicking on the drop and down menu (on the middle-right).

7. Cell refinement

It is important to determine the cell parameters accurately before integrating the images. This procedure requires the integration of a series of images in ideally two or more separate segments at widely different φ values. The distribution of the intensity of partially recorded reflections over the images on which they occur is used to refine the unit cell, crystal orientation and mosaic spread.

Select the "Cell Refinement" icon

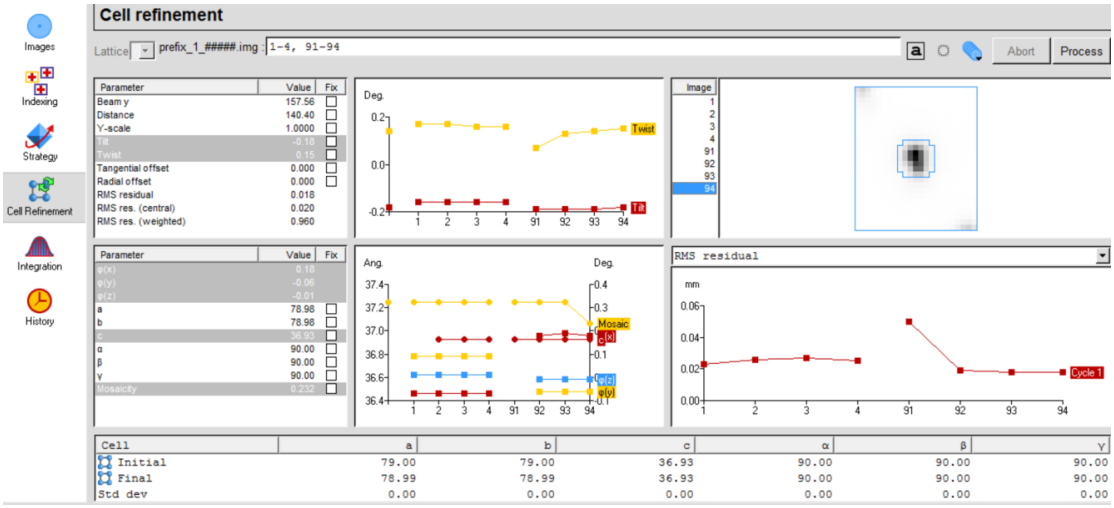
Selecting the "Automatically select images" icon [a] will select a suitable set of images. The chosen sector can be visualized with the graphical selection tool .

8.2 Integrating the images and refining the cell

Before starting cell refinement, check that the prediction for the first image to be used is **OK**

Click on the Process button.

The selected images will be integrated, and following integration the cell parameters are refined. Selectable refined detector parameters will be plotted, as will the changes in crystal orientation and (if selected) mosaic spread.



Questions: Examine the plots to see if the parameters are stable and correct. Check the final cell parameters and estimated errors (std dev).

Note that although the images are integrated during the cell refinement, the intensities are not saved and no reflection file is generated at that stage.

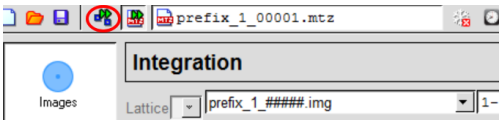
9. Integration

The accurate cell parameters are now used in the integration.

Select the "Integration" icon. This will open the Integration window.

9.1 Image selection

Image selection is performed in exactly the same way as in the cell refinement. In this case, the "*Automatic*" selection will simply include all images in the current sector.



If the "*Show predictions*" icon is clicked (circled in red above), the display window will be updated as each image is processed. This will slow down the processing, but allows the accuracy of the prediction to be checked for each image.

The filename of the MTZ file containing the results of the integration is generated automatically (**prefix_1_00001.mtz**).

9.3 Integrating the images

Integration of the images occurs in two passes. In the first pass, for each block of images (a number corresponding to a few degrees of data, selected automatically by MOSFLM), the detector and crystal parameters are refined for each image in turn and the "measurement boxes" for all predicted spots are written to a file for use in the second pass.

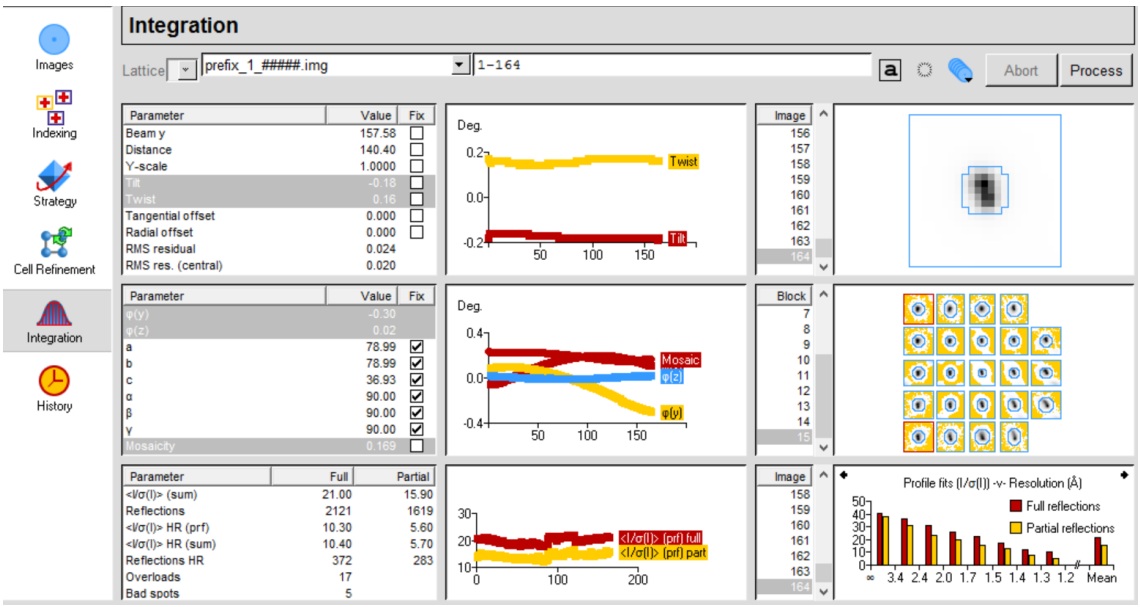
In the second pass, the standard profiles are formed from reflections present on all of the images in this block, and each image is then integrated and the results written to the MTZ file.

Note that the unit cell dimensions are normally fixed during integration, and only the crystal orientation and mosaic spread are refined.

Click on the "Process" button
The refined detector and crystal parameters are displayed in tables and selected parameters are plotted in graphs. The average spot profile for each image is also displayed.

A display of the standard profiles for different regions of the detector is also provided. Poor profiles are "averaged", by including reflections from inner regions of the detector. Averaged profiles are indicated by a red line around the spot.

Finally, the lower left window plots I/σ(I) as a function of resolution for any selected image.



At this stage, the **Medium** and **Advanced** groups can process the gadolinium dataset on another computer (ignore the distance warning when using cell refinement in iMosflm).

10. Running Pointless to check the symmetry

Once the images have been integrated, the program POINTLESS can be run to determine the true **Laue symmetry** and try to determine the **space group**.

Select the QuickSymm button in the top bar of the Integration window.

A summary of the Pointless results as shown below will appear in a "qtRView" window.

Results | Log File

Quick Symmetry

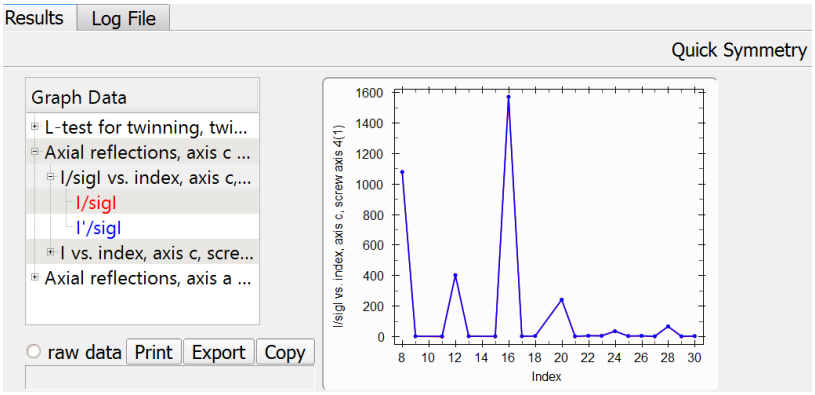
Run of POINTLESS on 23/ 5/2015 at 14:34:24

Result

```
Best Solution:    space group P 41 21 2

   Reindex operator:                  [h,k,l]
   Laue group probability:            1.000
   Systematic absence probability:    0.962
   Total probability:                 0.961
   Space group confidence:            0.953
   Laue group confidence              0.999

WARNING:
You will have to resolve the enantiomorphic ambiguity later

   Unit cell:   78.99  78.99  36.93     90.00  90.00  90.00

   39.49 to   1.20   - Resolution range used for Laue group search

   39.49 to   1.20   - Resolution range in file, used for systematic absence check

   Number of batches in file:    164
```

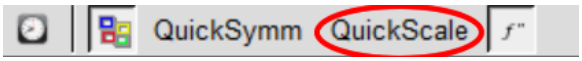
<u>Question</u>: Find in the Log of QuickScale the identified enantiomorphous pair of space groups

<u>Question</u>. Check the identified space group $P4_12_12$ by using the graph data tool of the qtRView window and observing the systematic absence check along the 00l and h00 axes.



11. Running Aimless to scale and merge the data

Select the QuickScale button.



This will first run POINTLESS then change the space group for the data to that selected by POINTLESS, run AIMLESS to **scale** and **merge** the data in that space group and then run CTRUNCATE which calculates a number of statistics from the intensity data.

AIMLESS can be run from the QuickScale button either treating anomalous data (the default) or ignoring anomalous data (uncheck the f'' button next to the QuickScale button).

The AIMLESS results will be displayed in the qtRView, as shown below, along with graphical output and the full log file can also be examined.

Quick Scale

Result

Summary data for Project: New Crystal: New Dataset: New

	Overall	InnerShell	OuterShell
Low resolution limit	39.49	39.49	1.26
High resolution limit	1.20	3.79	1.20
Rmerge	0.043	0.025	0.146
Rmerge in top intensity bin	0.030	-	-
Rmeas (within I+/I-)	0.047	0.028	0.159
Rmeas (all I+ & I-)	0.047	0.028	0.158
Rpim (within I+/I-)	0.019	0.011	0.063
Rpim (all I+ & I-)	0.014	0.009	0.045
Fractional partial bias	-0.016	-0.023	0.014
Total number of observations	456668	14062	65995
Total number unique	37271	1286	5352
Mean((I)/sd(I))	32.7	52.4	15.7
Mn(I) half-set correlation CC(1/2)	1.000	0.999	0.993
Completeness	99.9	97.1	100.0
Multiplicity	12.3	10.9	12.3
Anomalous completeness	99.9	96.4	100.0
Anomalous multiplicity	6.3	6.5	6.3
DelAnom correlation between half-sets	-0.066	0.099	-0.054
Mid-Slope of Anom Normal Probability	0.982	-	-

Question. Check completeness, multiplicity, Rmerge,vs resolution in the Graph data

$$R_{merge} = \frac{\sum_h \sum_i |I_i - \langle I \rangle|}{\sum_h \sum_i I_i}$$

At the bottom of the qtRView window, you can see that an output file named "**ctruncate_prefix_1_00001.mtz**" was created. This file contains your merged amplitudes from the lower to the higher resolution limit.

Rename the output file to **"lyso-1-1-A-IF-P41212.mtz"** *with the export button available under the "reflection data" bullet field.*

12. Changing the space group to $P4_32_12$ and creating a new reflection file

Now, we will create a file equivalent to "**lyso-1-1-A-IF-P41212.mtz"** in the enantiomorphous $P4_32_12$

Go back to the ***CCP4i interface***

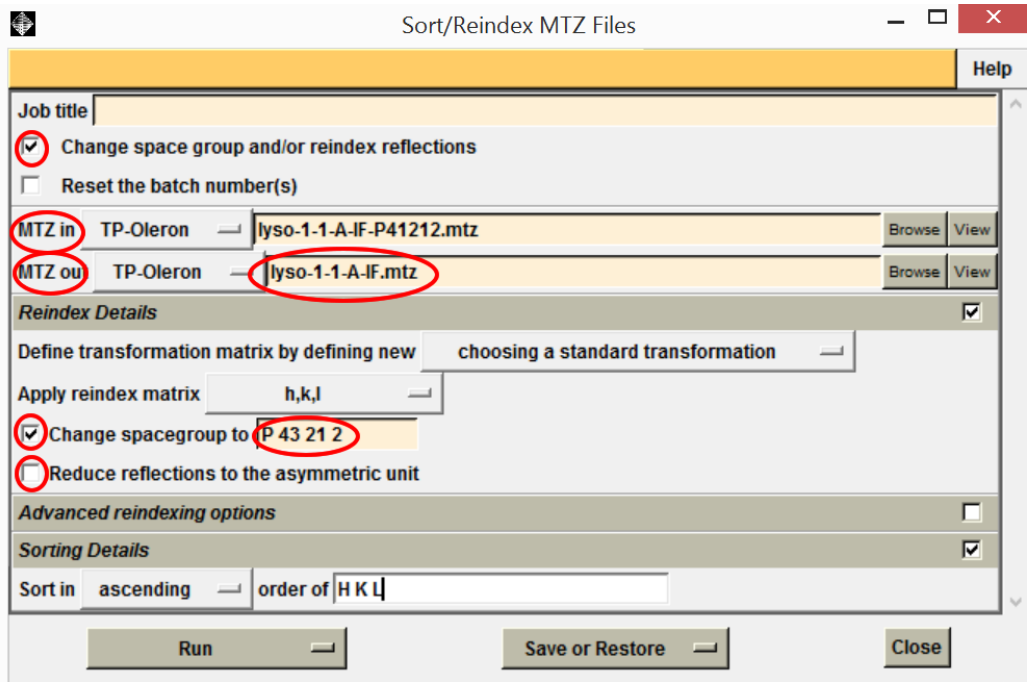
From the list of "*Data Reduction and Analysis*" options at the left hand side of the main window, select "*Utilities*" and under that "*Sort/Modify/Combine MTZ Files*"

Click on the Box "*Change space group and/or reindex reflections*"

Use the "browse" option to select the input MTZ file (First orange bar), the output filename will be generated automatically. Change it to " **lyso-1-1-A-IF.mtz**"

Check the box next to "*Change space group to*" and enter *P 43 21 2* in the box.

Uncheck the box "*Reduce reflections to the asymmetric unit*"



Select "*Run now*" from the "*Run*" option button. This will create the output file.

Conclusion. The two files "**lyso-1-1-A-IF-P41212.mtz**" and "**lyso-1-1-A-IF.mtz**" contain your observed amplitudes (Fobs) indexed in the enantiomorphic space groups $P4_12_12$ and $P4_32_12$ respectively. They will be used for data phasing.

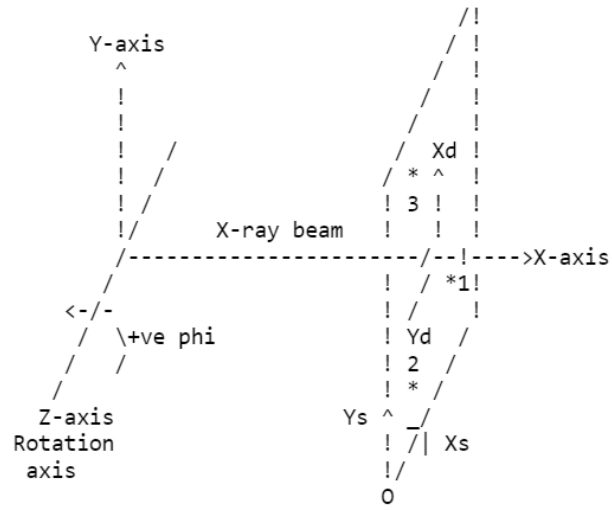
At this stage:

- The **Beginner** groups should proceed to the molecular replacement section.

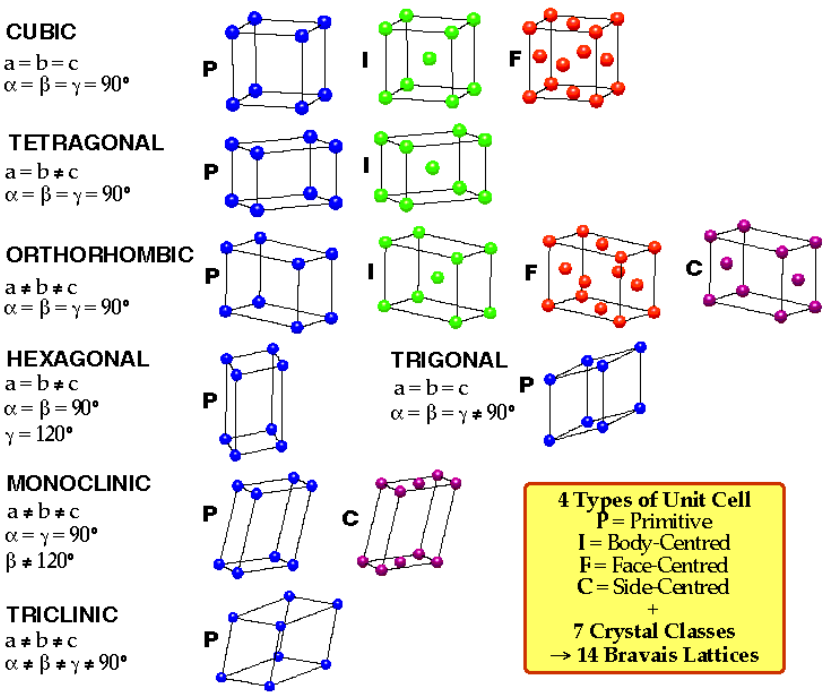
- The **Medium** and **Advanced** groups should have checked that the gadolinium dataset has a strong anomalous signal.

ANNEX

ESRF beamline FIP-BM30A



Diffractometer geometry in iMosflm

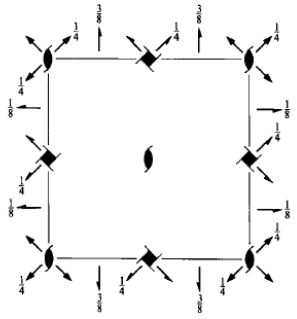


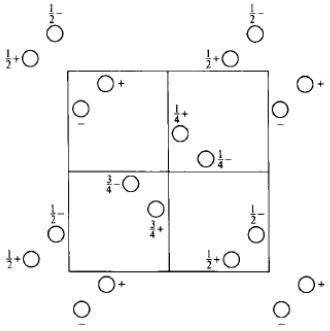
Crystal System	Minimum Symmetry*	Constraints on unit cell
Triclinic	None	None
Monoclinic	One 2-fold (along b)	$\alpha = \gamma = 90$
Orthorhombic	Three 2-folds (along a,b,c)	$\alpha = \beta = \gamma = 90$
Trigonal	3-fold (along c)	$a = b$; $\alpha = \beta = 90$; $\gamma = 120$
Tetragonal	4-fold (along c)	$a = b$; $\alpha = \beta = \gamma = 90$
Hexagonal	6-fold (along c)	$a = b$; $\alpha = \beta = 90$; $\gamma = 120$
Cubic	Four 3-fold axes (along body diagonal)	$a = b = c$; $\alpha = \beta = \gamma = 90$

System	Laue class	Space Groups
Triclinic	1	P1
Monoclinic	2	P2, $P2_1$, C2
Orthorhombic	222	P222, $P222_1$, $P2_12_12$, $P2_12_12_1$, $C222_1$, C222, F222, I222, $I2_12_12_1$
Quadratic	4 422	P4, $P4_1$, $P4_2$, $P4_3$, I4, $I4_1$, P422, $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, I422, $I4_122$
Trigonal	3 32	P3, $P3_1$, $P3_2$, R3, P312, P321, $P3_112$, $P3_121$, $P3_212$, $P3_221$, R32
Hexagonal	6 622	P6, $P6_1$, $P6_5$, $P6_2$, $P6_4$, $P6_3$, P622, $P6_122$, $P6_522$, P622, $P6_422$, $P6_322$
Cubic	23 432	P23, F23, I23, $P2_13$, $I2_13$, P432, $P4_232$, F432, $F4_132$, I432, $P4_332$, $P4_132$, $I4_132$

$P4_12_12$ D_4^4 422 Tetragonal

No. 92 $P4_12_12$ Patterson symmetry $P4/mmm$





Origin on $2\,[1\,1\,0]$ at $2_1\,1\,(1,2)$

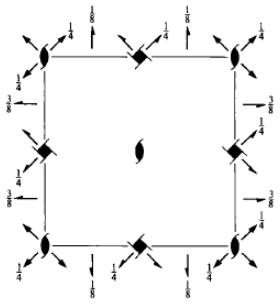
Asymmetric unit $0 \le x \le 1;\quad 0 \le y \le 1;\quad 0 \le z \le \frac{1}{8}$

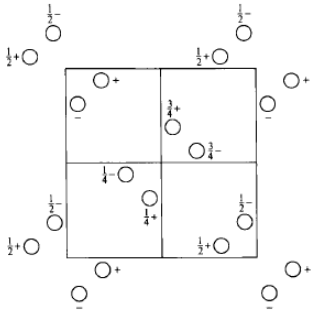
Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates				Reflection conditions
							General:
8	b	1	(1) x,y,z	(2) $\bar{x},\bar{y},z+\frac{1}{2}$	(3) $\bar{y}+\frac{1}{2},x+\frac{1}{2},z+\frac{1}{4}$	(4) $y+\frac{1}{2},\bar{x}+\frac{1}{2},z+\frac{3}{4}$	$00l:\ l=4n$ $h00:\ h=2n$
			(5) $\bar{x}+\frac{1}{2},y+\frac{1}{2},\bar{z}+\frac{1}{4}$	(6) $x+\frac{1}{2},\bar{y}+\frac{1}{2},\bar{z}+\frac{3}{4}$	(7) $y,x,\bar{z}$	(8) $\bar{y},\bar{x},\bar{z}+\frac{1}{2}$	
							Special: as above, plus
4	a	. . 2	$x,x,0$	$\bar{x},\bar{x},\frac{1}{2}$	$\bar{x}+\frac{1}{2},x+\frac{1}{2},\frac{1}{4}$	$x+\frac{1}{2},\bar{x}+\frac{1}{2},\frac{3}{4}$	$0kl:\ l=2n+1$ or $2k+l=4n$

$P4_32_12$ D_4^8 422 Tetragonal

No. 96 $P4_32_12$ Patterson symmetry $P4/mmm$





Origin on $2\,[1\,1\,0]$ at $2_1\,1\,(1,2)$

Asymmetric unit $0 \le x \le 1;\quad 0 \le y \le 1;\quad 0 \le z \le \frac{1}{8}$

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates				Reflection conditions
							General:
8	b	1	(1) x,y,z	(2) $\bar{x},\bar{y},z+\frac{1}{2}$	(3) $\bar{y}+\frac{1}{2},x+\frac{1}{2},z+\frac{3}{4}$	(4) $y+\frac{1}{2},\bar{x}+\frac{1}{2},z+\frac{1}{4}$	$00l:\ l=4n$ $h00:\ h=2n$
			(5) $\bar{x}+\frac{1}{2},y+\frac{1}{2},\bar{z}+\frac{3}{4}$	(6) $x+\frac{1}{2},\bar{y}+\frac{1}{2},\bar{z}+\frac{1}{4}$	(7) $y,x,\bar{z}$	(8) $\bar{y},\bar{x},\bar{z}+\frac{1}{2}$	
							Special: as above, plus
4	a	. . 2	$x,x,0$	$\bar{x},\bar{x},\frac{1}{2}$	$\bar{x}+\frac{1}{2},x+\frac{1}{2},\frac{3}{4}$	$x+\frac{1}{2},\bar{x}+\frac{1}{2},\frac{1}{4}$	$0kl:\ l=2n+1$ or $2k+l=4n$