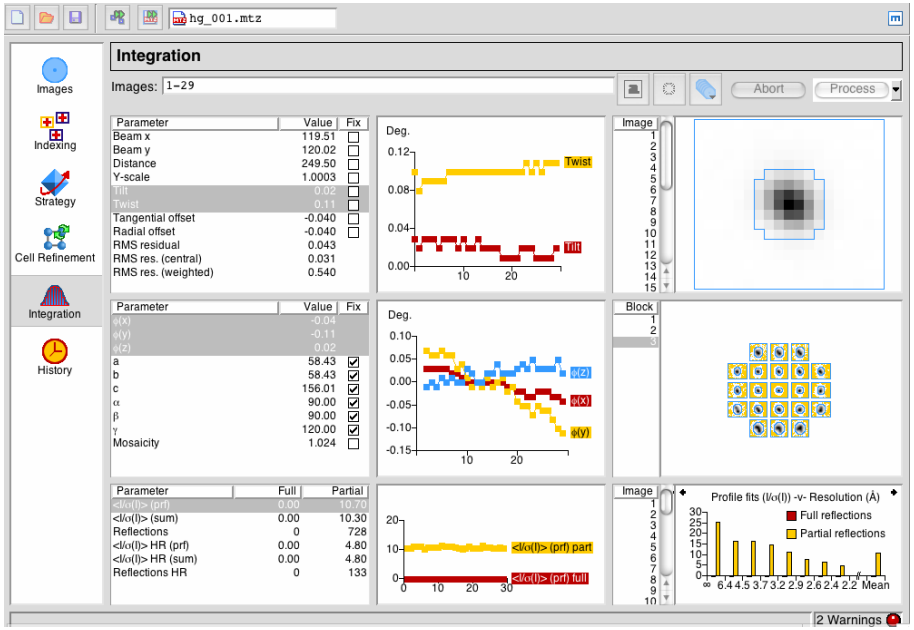
DLS-CCP4 workshop, December 2016

Data Processing with iMosflm

Andrew GW Leslie, MRC LMB, Cambridge, UK



Software is distributed with CCP4 or can be downloaded from (Google “imosflm”): www.mrc-lmb.cam.ac.uk/harry/imosflm

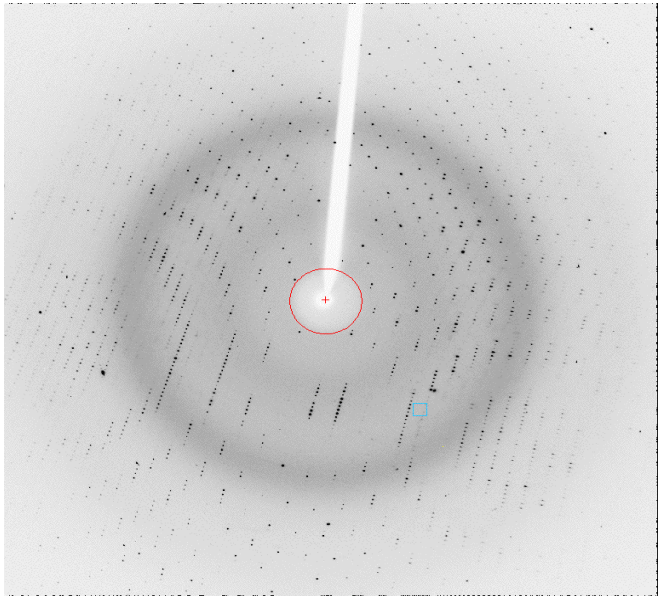
Documentation is a detailed tutorial for iMosflm available on the web site.

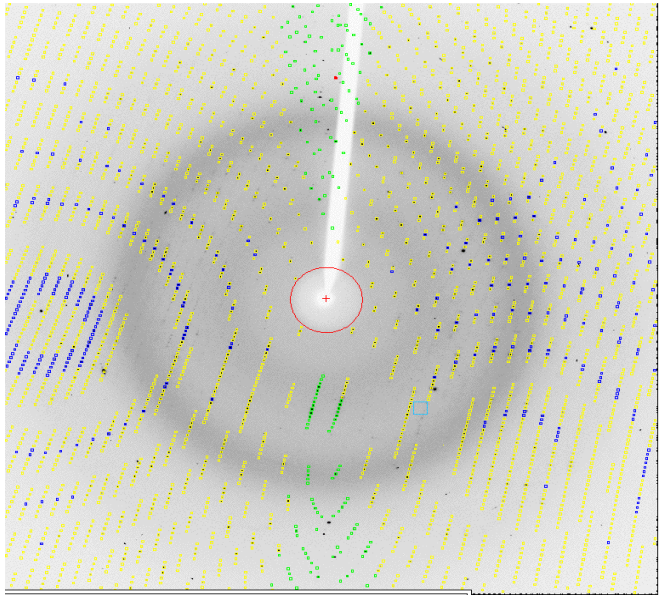
Integration of Diffraction Images

Starting Point: A series of diffraction images, each recorded on a 2D area detector while rotating the crystal through a small angle (typically 0.1-1.0 degrees per image) about a fixed axis (the Rotation/Oscillation Method).

Outcome: A dataset consisting of the indices (h,k,l) of all reflections recorded on the images with an estimate of their intensities and the standard uncertainties of the intensities: h, k, l, I(hkl), σ(I)

This requires the prediction of which image(s) each reflection will occur on and also the precise position of the reflection in the images. (Typically each reflection will be spread out over several images, and therefore only partially recorded on any single image).





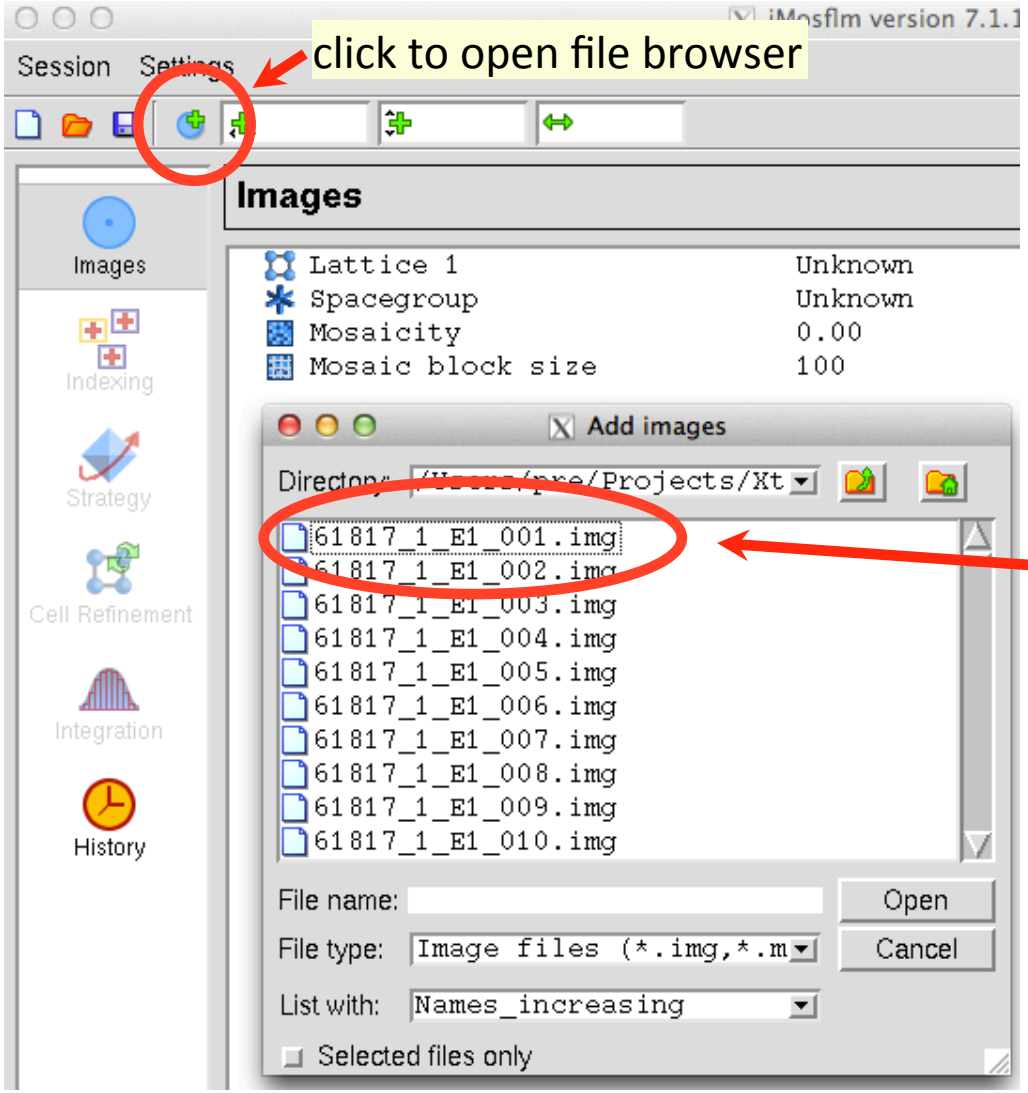
In practice, defects in the crystals (or detectors) make this operation far from trivial. eg weak diffraction, crystal splitting, multiple lattices, radiation damage, anisotropic diffraction, diffuse scattering, ice rings/spots, high mosaicity, unresolved spots, overloaded spots, zingers/cosmic rays, “hot” pixels.

Integration procedure in iMosflm

1. **Read Images**
   read all images with a common template
2. **Find spots and Index**
   find *lattice* which fits spots
3. **Check prediction**
   is the indexing correct?
4. **Estimate mosaicity**
   improve estimate later
5. **Refine cell**
   use two wedges at 90°, or more in low symmetry
6. **Mask backstop shadow**
   not (yet) done automatically by program
7. **Integrate one (or few) image**
   to check resolution etc
8. **Integrate all images**
   optionally run in background for speed

Strategy option, for use before data collection

Select and load images



double-click on any image to import the whole series

This also opens the image display window for the first image

MOSFLM tries to extract useful information from the header

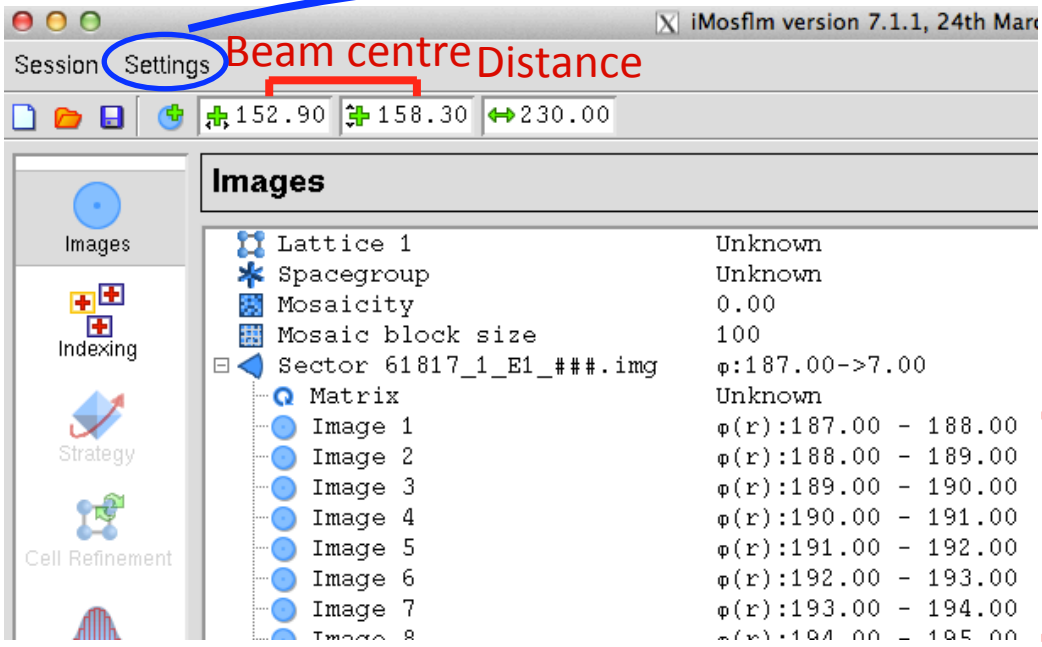
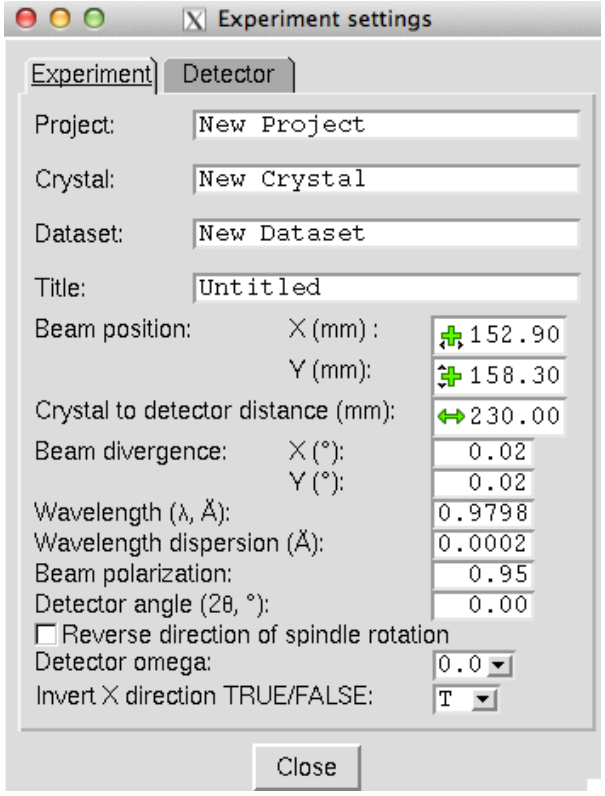


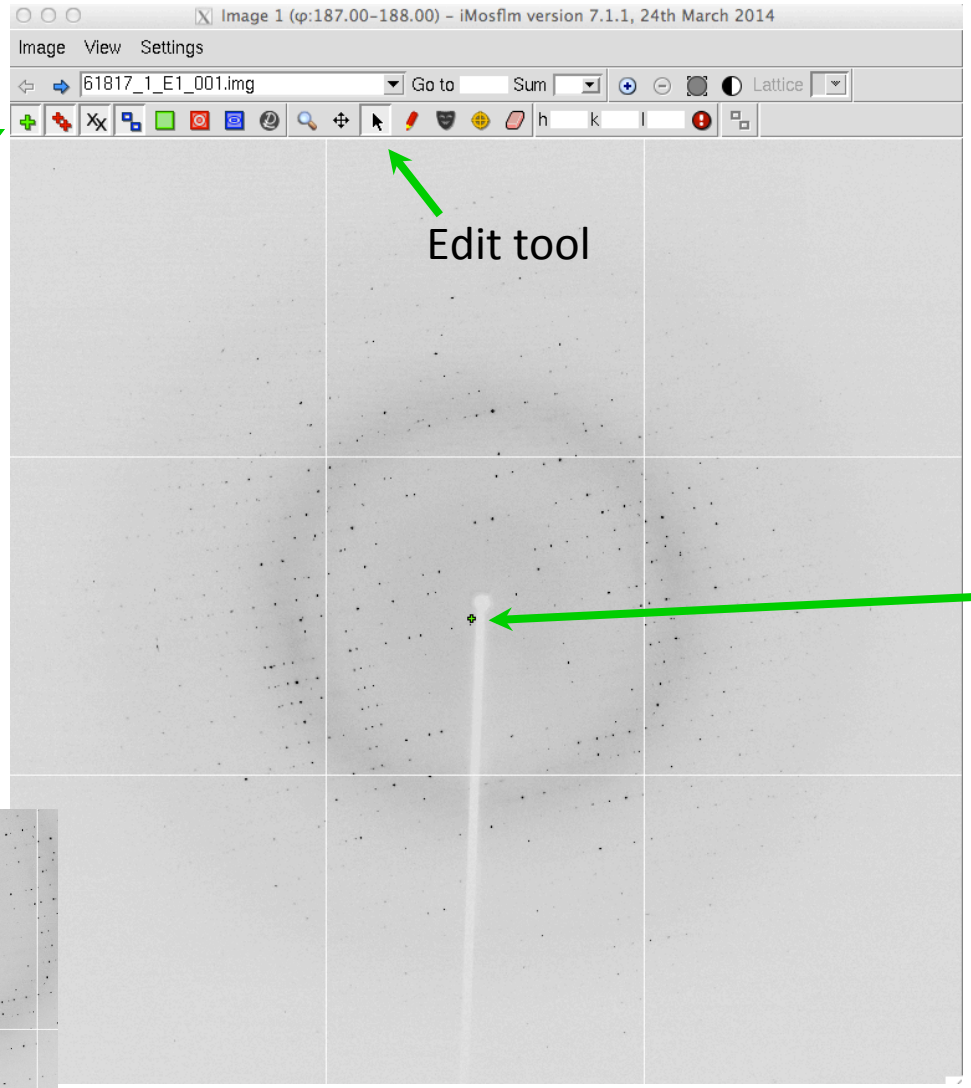
Image phi ranges



Parameters may be reset in the main window, or in the Settings → Experiment settings window

A few beamlines have a "reversed" spindle: this can be indicated here

Image display window

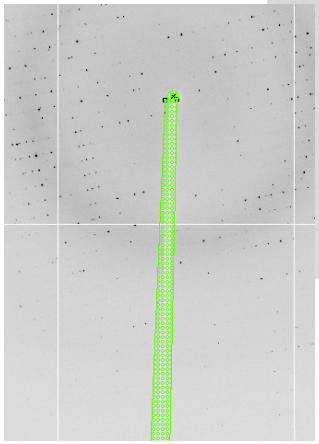


Display main beam position

Edit tool

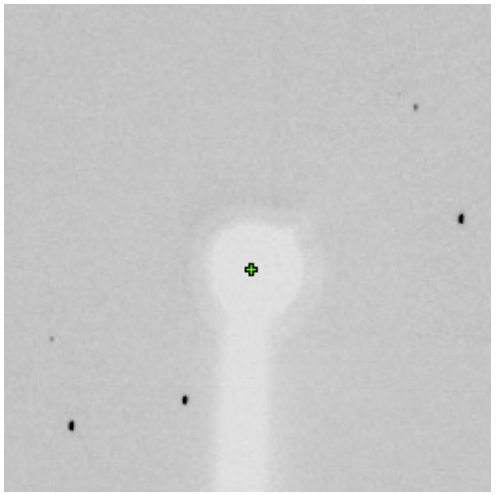
Controls:
Zoom & pan.
Choose overlays eg spots, predictions, resolution limits.
Move main beam position.
Edit backstop mask.
etc

Main beam in wrong position



Beam-stop shadow masked out

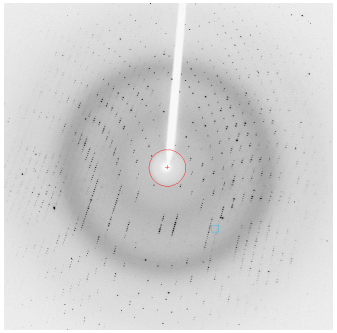
(see Tutorial for details)



Moved to roughly right position

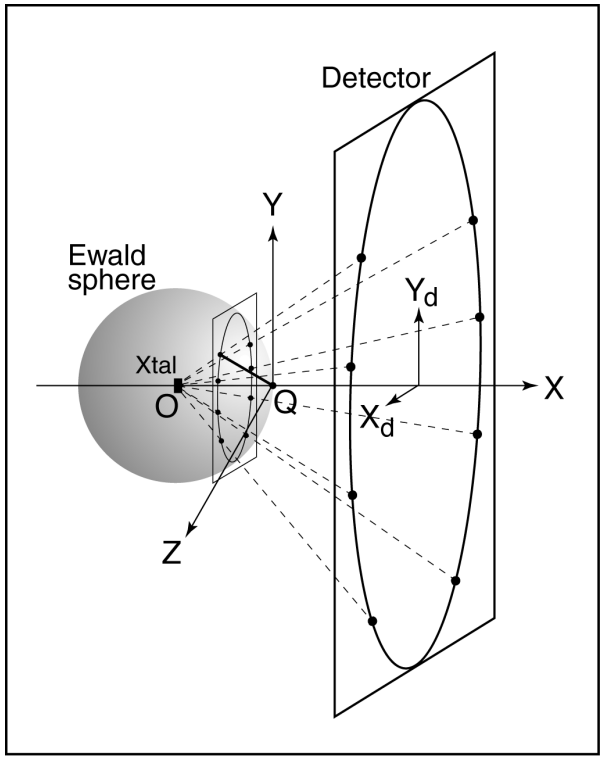
Indexing - Mapping spot positions back to reciprocal space vectors

The spot positions in a diffraction image are a distorted projection of the reciprocal lattice. Using the Ewald sphere construction, the observed reflections (X_d, Y_d, Φ) can be mapped back into reciprocal space giving a set of scattering vectors $\mathbf{s}_i$.



$$\mathbf{s} = \begin{bmatrix} D/r - 1 \\ X_d/r \\ Y_d/r \end{bmatrix}$$

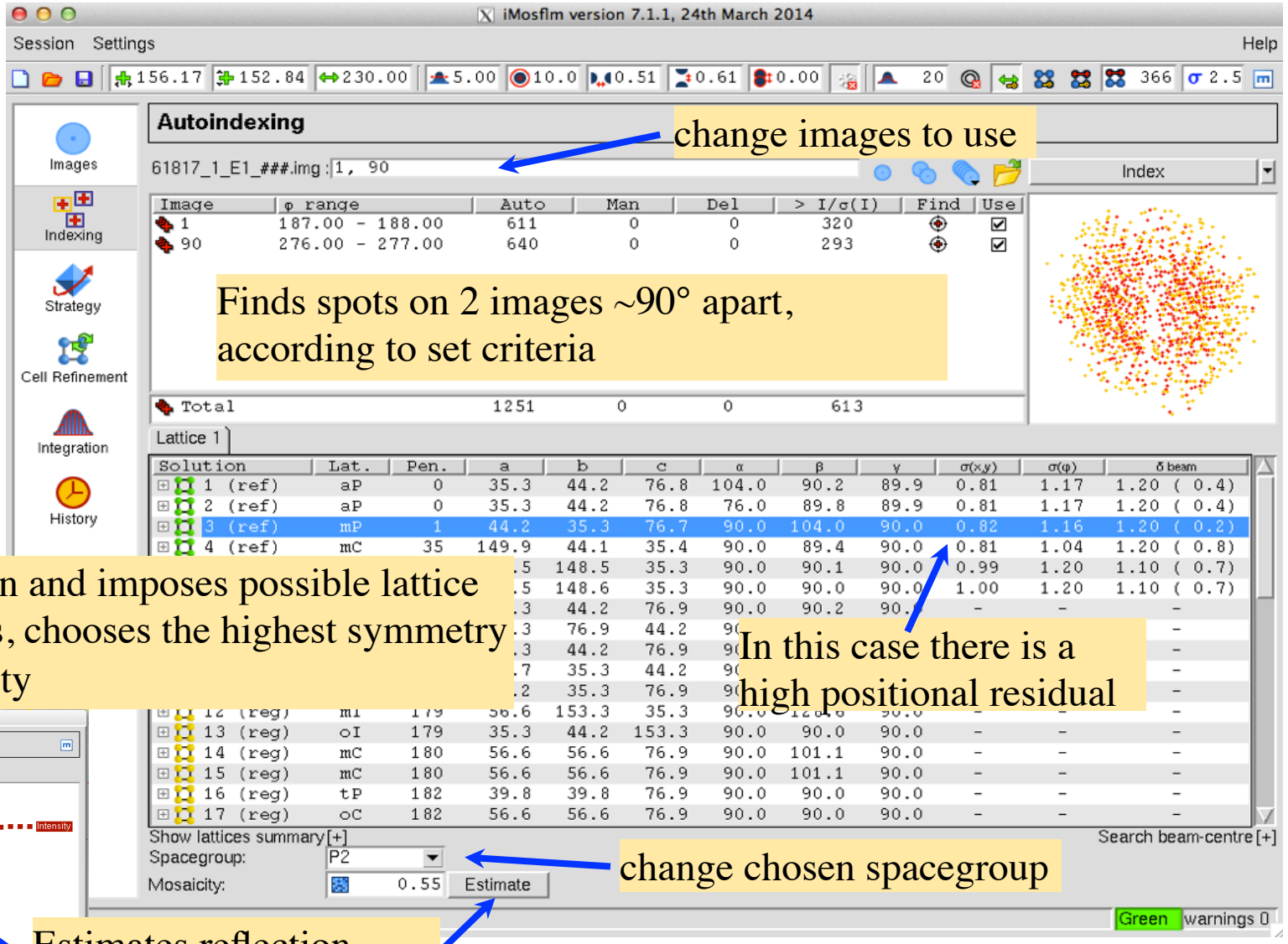
$$r = \sqrt{(X_d^2 + Y_d^2 + D^2)}$$



D is crystal to detector distance. Φ is taken as the midpoint of the oscillation range.

Uncertainty in Φ leads to errors in **s**, the size of the errors depends on both the oscillation angle and the mosaic spread.

Indexing



change images to use

Finds spots on 2 images ~90° apart, according to set criteria

Finds a solution and imposes possible lattice cell constraints, chooses the highest symmetry with low penalty

In this case there is a high positional residual

change chosen spacegroup

Estimates reflection width (mosaicity)

Auto-indexing procedures generate a list of possible cells

Solutions:

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	58.5	58.6	62.1	90.1	118.0	120.0	0.20	0.36	0.45 (0.1)
2 (ref)	aP	0	58.5	58.6	62.1	61.9	62.0	60.0	0.20	0.35	0.45 (0.1)
3 (ref)	mC	1	101.5	58.5	62.1	90.0	123.0	90.0	0.21	0.39	0.45 (0.2)
4 (ref)	mC	2	101.5	58.5	62.1	90.0	123.0	90.0	0.21	0.39	0.45 (0.2)
5 (ref)	mC	2	101.4	58.6	62.2	90.0	123.0	90.0	0.19	0.42	0.44 (0.2)
6 (ref)	mC	2	101.5	58.5	62.1	90.0	122.9	90.0	0.20	0.36	0.46 (0.2)
7 (ref)	hR	4	58.6	58.6	156.4	90.0	90.0	120.0	0.21	0.36	0.46 (0.3)
8 (reg)	mC	59	85.3	85.5	58.5	90.0	133.2	90.0	-	-	-
9 (reg)	mC	59	103.5	62.1	58.5	90.0	124.3	90.0	-	-	-
10 (reg)	mC	60	101.5	58.5	62.1	90.0	123.0	90.0	-	-	-
11 (reg)	oI	60	58.5	62.1	85.5	90.0	90.0	90.0	-	-	-
12 (reg)	oI	60	58.6	62.2	85.3	90.0	90.0	90.0	-	-	-
13 (reg)	tI	61	60.4	60.4	85.3	90.0	90.0	90.0	-	-	-
14 (reg)	mC	61	101.5	58.5	62.1	90.0	123.0	90.0	-	-	-
15 (reg)	hR	61	60.4	60.4	144.9	90.0	90.0	120.0	-	-	-

Spacegroup: h3

Mosaicity: 0.00 Estimate

Lattice: a Triclinic; m monoclinic; o orthorhombic; t tetragonal; h hexagonal; c cubic
P primitive; C C-face centred; I body-centred; F face-centred; R rhombohedral

A “penalty” is associated with each solution, which reflects how well the determined cell obeys the constraints for that lattice type. The solution with the highest symmetry from the group of solutions with low penalties (highlighted in blue) is usually chosen as the correct solution (but beware of pseudosymmetry)

Note that all these solutions are based on the same reduced cell

An example of pseudo-symmetry

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	81.3	81.9	82.3	90.0	90.1	90.0	0.17	0.43	0.01 (0.0)
2 (ref)	mP	0	81.3	81.9	82.3	90.0	90.1	90.0	0.16	0.43	0.00 (0.0)
3 (ref)	aP	0	81.3	81.9	82.3	90.0	89.9	90.0	0.17	0.43	0.01 (0.0)
4 (ref)	mP	1	81.3	82.3	81.8	90.0	90.0	90.0	0.18	0.44	0.00 (0.0)
5 (ref)	mP	1	81.8	81.3	82.3	90.0	90.0	90.0	0.18	0.44	0.00 (0.0)
6 (ref)	oP	1	81.3	81.8	82.3	90.0	90.0	90.0	0.18	0.44	0.00 (0.0)
7 (ref)	mC	3	116.2	115.8	81.4	90.0	90.2	90.0	0.18	0.48	0.00 (0.0)
8 (ref)	tP	3	82.1	82.1	81.3	90.0	90.0	90.0	0.22	0.45	0.00 (0.0)
9 (ref)	oC	3	116.0	116.1	81.3	90.0	90.0	90.0	0.21	0.46	0.00 (0.0)
10 (ref)	mC	3	116.2	115.8	81.4	90.0	90.2	90.0	0.18	0.48	0.00 (0.0)
11 (ref)	mC	4	115.5	115.3	82.2	90.0	90.1	90.0	0.24	0.43	0.01 (0.0)
12 (ref)	oC	4	115.5	115.3	82.3	90.0	90.0	90.0	0.24	0.42	0.01 (0.0)
13 (ref)	tP	4	81.6	81.6	82.3	90.0	90.0	90.0	0.24	0.42	0.01 (0.0)
14 (ref)	mC	4	115.5	115.3	82.2	90.0	90.1	90.0	0.24	0.43	0.01 (0.0)
15 (ref)	hR	6	115.5	115.5	142.4	90.0	90.0	120.0	0.28	0.49	0.01 (0.0)
16 (ref)	hR	6	115.7	115.7	141.7	90.0	90.0	120.0	0.35	0.44	0.01 (0.0)
17 (ref)	cP	7	81.8	81.8	81.8	90.0	90.0	90.0	0.34	0.44	0.01 (0.0)
18 (reg)	mC	248	81.3	183.3	81.9	90.0	90.0	90.0	-	-	-
19 (reg)	oC	248	81.3	183.3	81.9	90.0	90.0	90.0	-	-	-

Spacegroup: P23

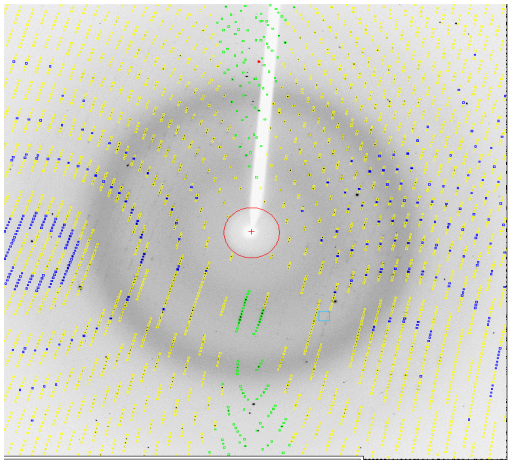
Mosaicity: 0.92 Estimate

Based on the penalties, the lattice would be assigned as cubic, but the rms error in spot positions σ(x,y) is 0.34 mm for the cubic solution, but only 0.18 mm for the (correct) orthorhombic solution.

Judging the success of the auto-indexing

- *Does the predicted pattern match the image* ?

Unless mosaic spread has been estimated, not all spots will be predicted. If two non-sequential images were used the prediction may be very slightly off. However, the pattern of the lunes should be correct.



Beware of under-predicting (if there is a systematic variation in the intensity of adjacent spots, eg every 2nd spot is weak).

Beware of over-predicting if a cell edge has been doubled. This happens most often when the direct beam coordinates are wrong.

Judging the success of the auto-indexing (contd.)

- *The rms error in spot positions $\sigma(x,y)$.*

  The magnitude of the error depends on the spot size and shape, on the the number of images included and how these images were collected.

  Typical values are 0.1-0.2 mm for image plate data collected on a lab source (where spots tend to be larger) and 0.05-0.15 mm for data collected on CCD/Pixel detectors at synchrotrons.

  However, if two images are used which were not collected sequentially (eg the first and last image of a data collection) the error is often higher (eg by a factor of 2) because the crystal orientation has changed during data collection or the rotation axis is not orthogonal to the X-ray beam.

  If the crystal is imperfect and gives split spots, the error can be up to 1.0 mm for a correct indexing.

Autoindexing….Criteria for success

- Having a sufficient number of spots (preferably a few hundred although 30 may be enough). A relatively small number of "false" spots (eg zingers, ice spots), can cause problems. iMosflm normally rejects all spots at resolutions corresponding to ice rings.
- Correct wavelength, direct beam position, detector distance.
- Reasonable mosaic spread (no overlap of adjacent lunes).
- Resolved spots.

Multiple lattices can now be indexed in many cases.

Absence of a clear separation between solutions with low penalties and solutions with high penalties can indicate errors in direct beam position, wavelength, distance etc (or a triclinic solution).

Results from a single image can be misleading for low symmetries.

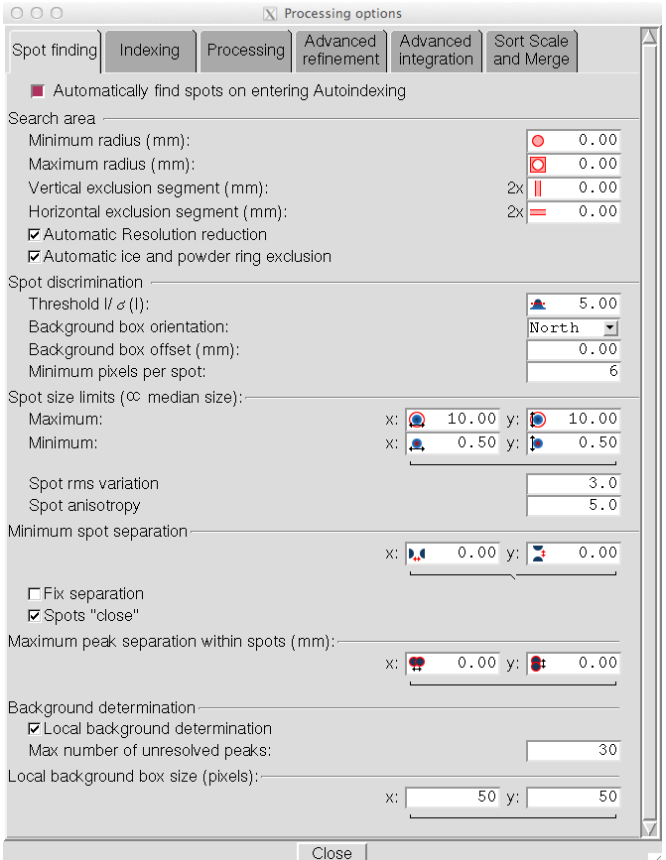
If Indexing Fails:

- check spot finding
- check main beam position
- try main beam search (button)
- try different images – try single images or more than two
- change maximum cell edge
- change spot selection criterion (I/sig(I) threshold)
- if indexing works on one image, but not on two, consider reversed phi

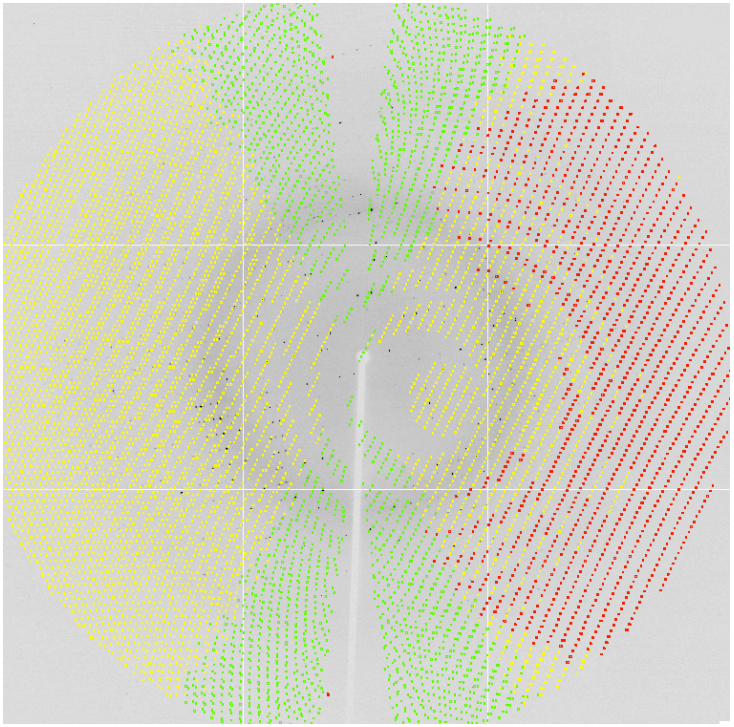
For weak images, to locate more spots reduce the threshold (default 5) to 2.5-3.0 and the spot rms variation (default 3.0) to 1.0.

For Pilatus images, changing the minimum number of pixels can sometimes help.

Sometime specifying a minimum spot separation can help, eg if spots are badly split.



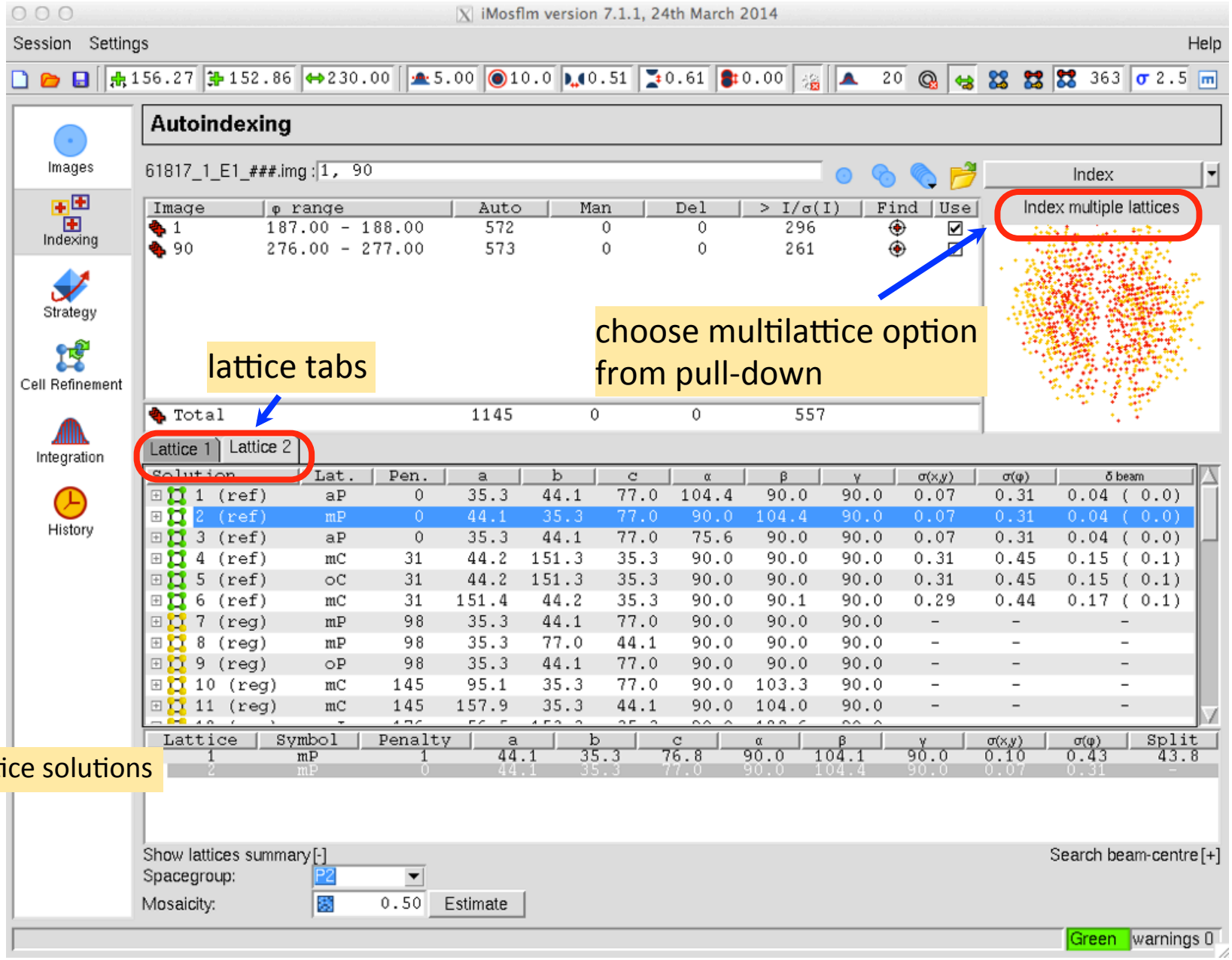
Appearance of the predicted reflections



- Yellow reflections are partials (orange for Pilatus images)
- Blue reflections (if present) are fully recorded.
- Red reflections are overlapped (closer than minimum separation)
- Green reflections are too wide (in phi) to measure.

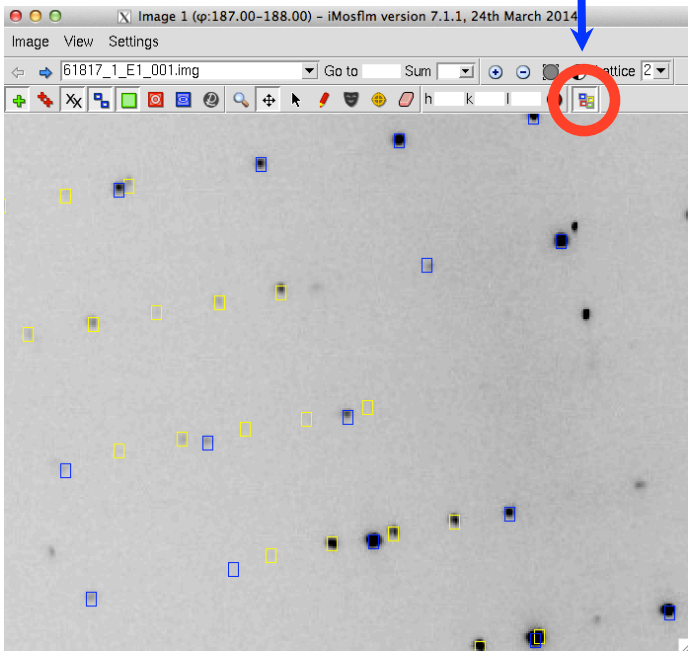
Neither red or green reflections are measured. In some cases the number of overlapped reflections can be reduced by manually specifying the minimum separation or reducing (and fixing) the mosaic spread. *If there are many "green" reflections, increase the maximum reflection width (default 5°)*

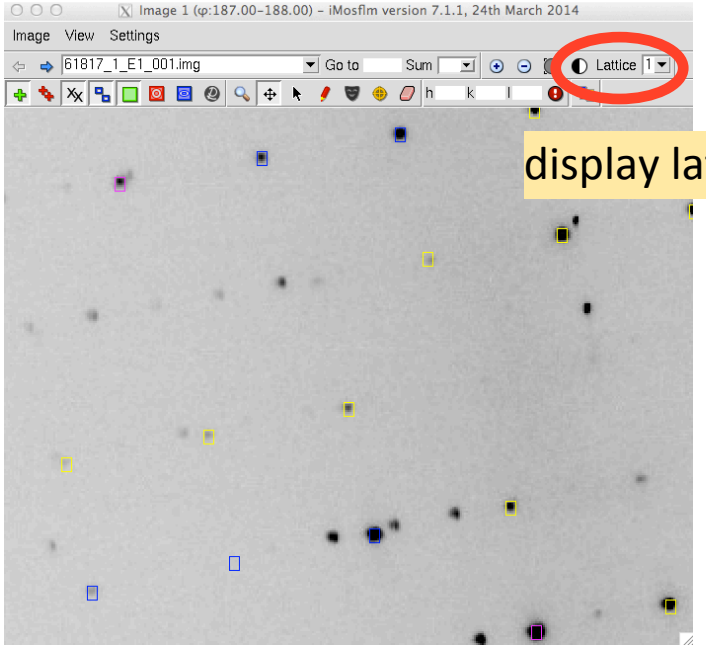
Indexing multiple lattices



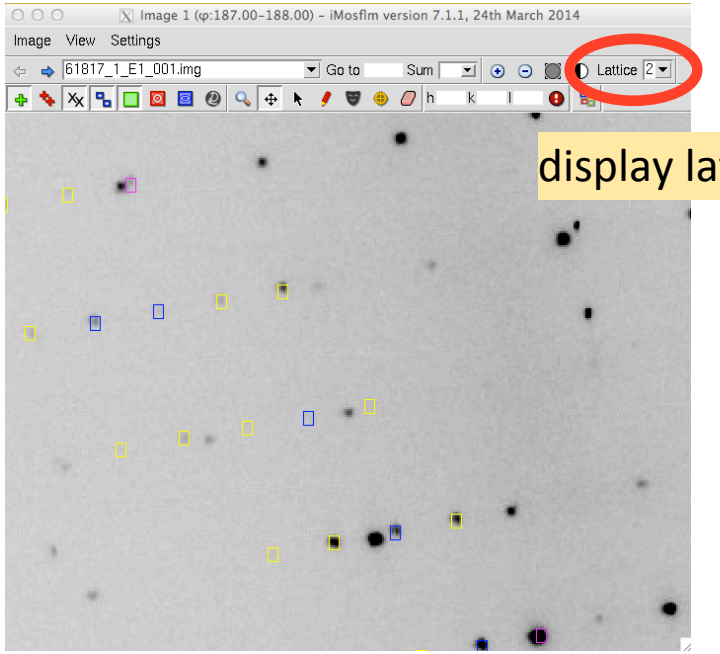
Display of multiple lattice solutions

display all lattices, in different colours





display lattice 1

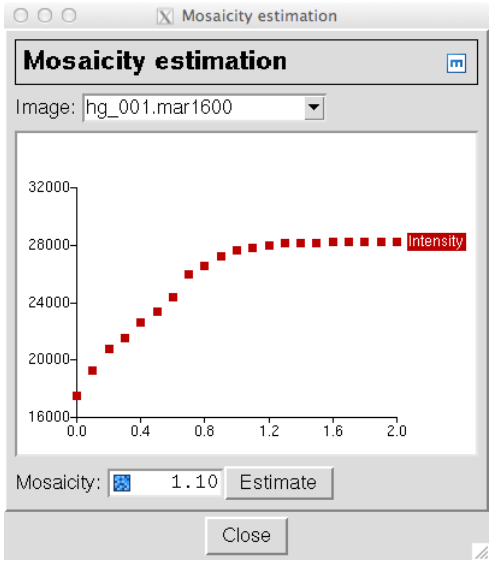


display lattice 2

Processing multiple lattices: warning

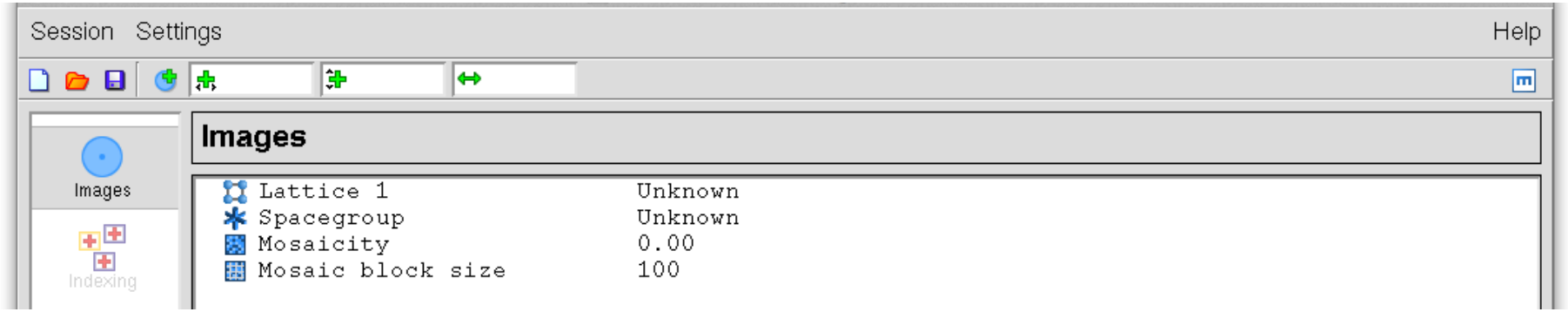
- Very much under development and test
- It may be better to ignore a weak 2nd lattice

Mosaicity Estimation



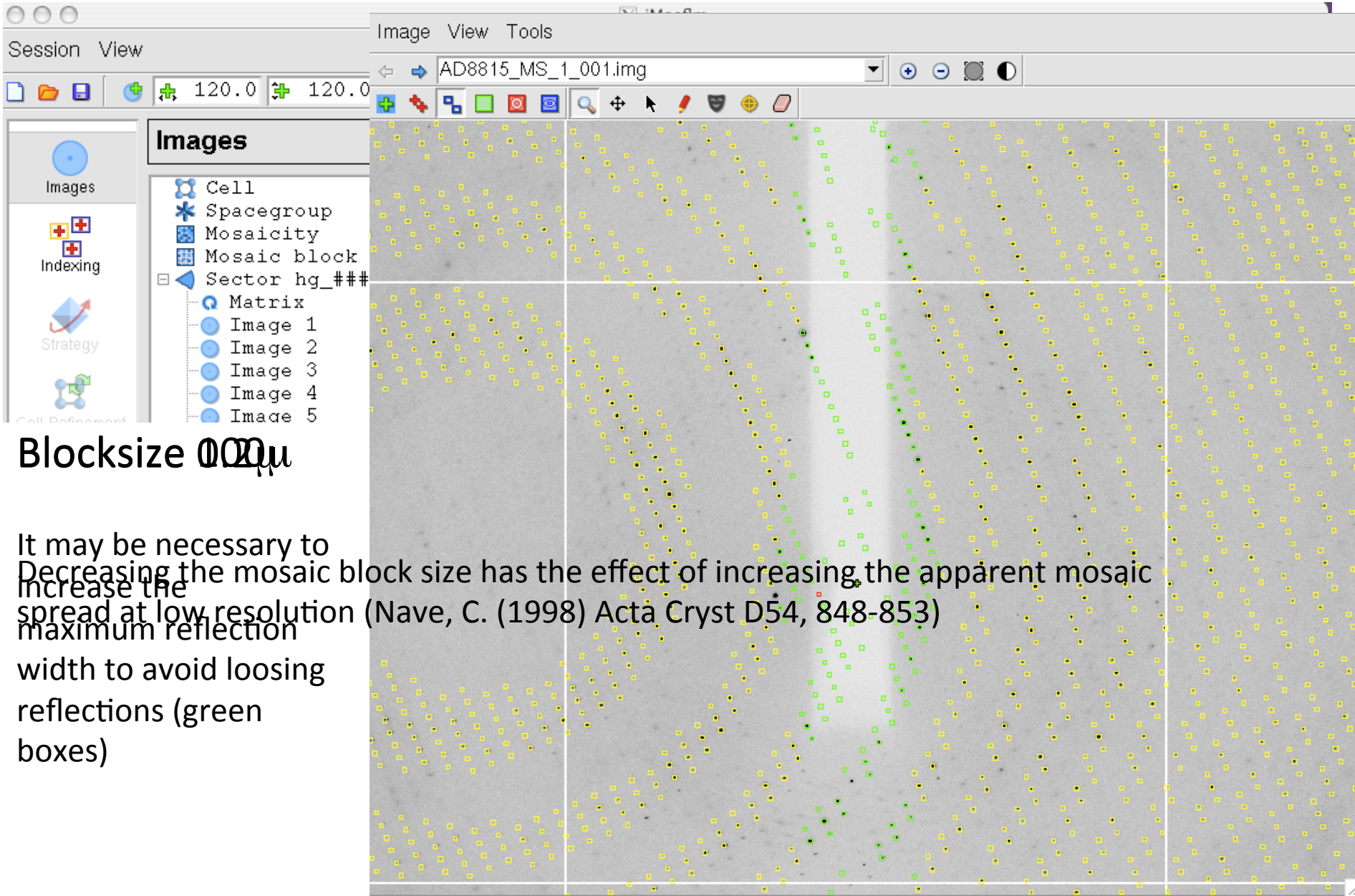
Predict pattern with increasing values for the mosaic spread (eg 0.0, 0.05, 0.1, 0.15 degrees). In each case, measure the total intensity of all predicted reflections. The mosaicity can be estimated from the plot of total intensity vs mosaic spread (after Rossmann, 1979).

NOTE: *If the indexing is incorrect, the mosaicity estimation will give the wrong answer !*



Mosaic block size: Defaults to 100 microns. Decreasing this results in more reflections being predicted at low resolution (with little or no effect at high resolution). Can be as low as 1 micron in some cases. *A negative partial bias (> 3%) and high Rmerge (> 0.04) at low resolution indicate that the mosaic block size should be reduced.*

Mosaicity – the effect of the mosaic block size



Blocksize 0.2°

Blocksize 0.02°

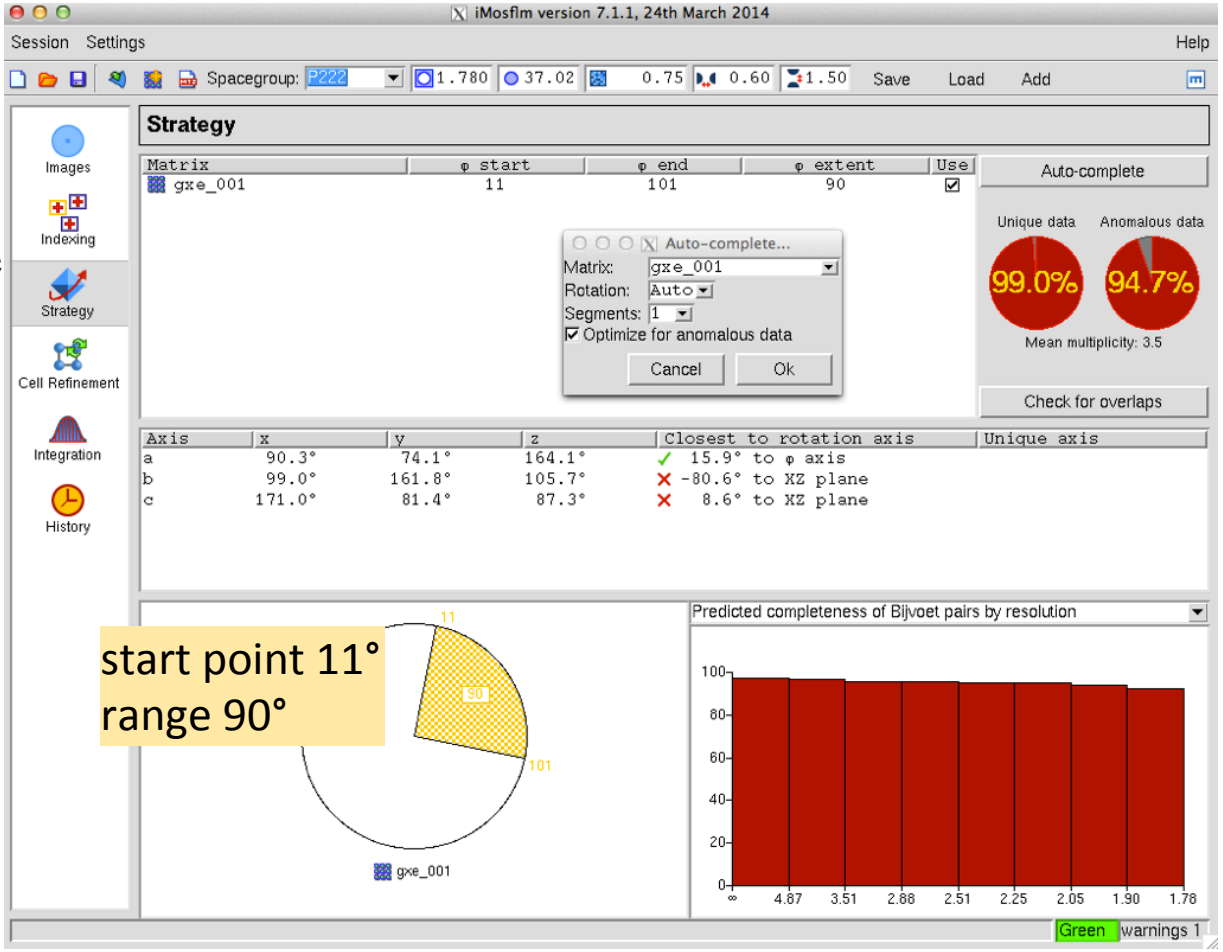
It may be necessary to increase the maximum reflection width to avoid loosing reflections (green boxes)

Decreasing the mosaic block size has the effect of increasing the apparent mosaic spread at low resolution (Nave, C. (1998) Acta Cryst D54, 848-853)

Strategy

Useful before data collection, after collecting (2 or 3) reference images and indexing

- gives starting phi angle and total rotation range
- can explore maximum rotation range/image (NB never 1°!) to avoid or minimise overlaps
- can be used to find range to collect on eg 2nd crystal to complete partial data from first
- multi-segment strategies



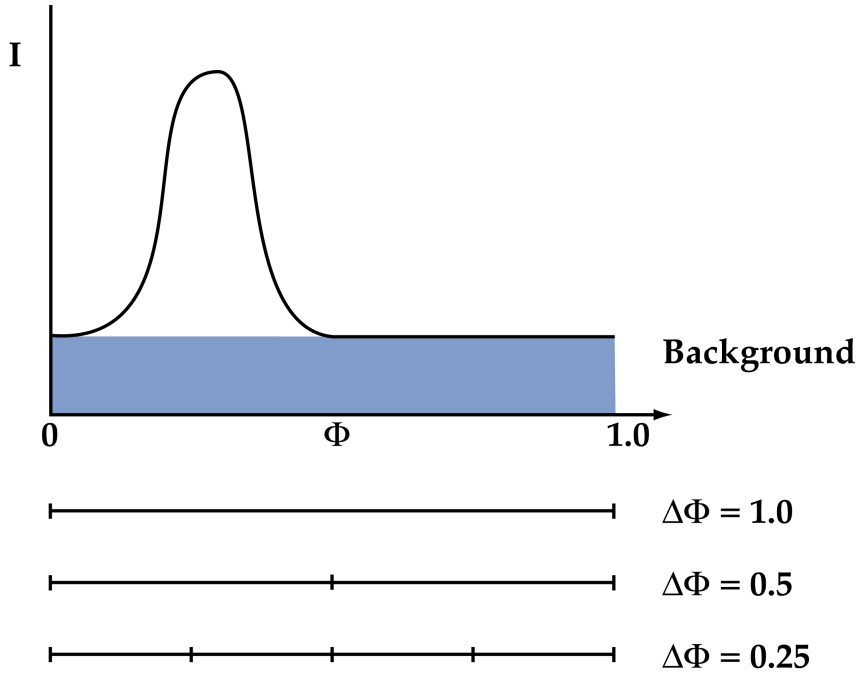
Sensitive to parameters from indexing:

- point group
- mosaicity
- maximum resolution
- minimum spot separation

see tutorial for more details

Advantages of fine phi slicing

Using an oscillation angle smaller than the reflection width (in phi) will improve signal to noise for weak reflections by minimising the background:



$$I_{spot} = I_{tot} - I_{bck}$$

$$var(I_{spot}) = var(I_{tot}) + var(I_{bck})$$

$$= I_{tot} + I_{bck}$$

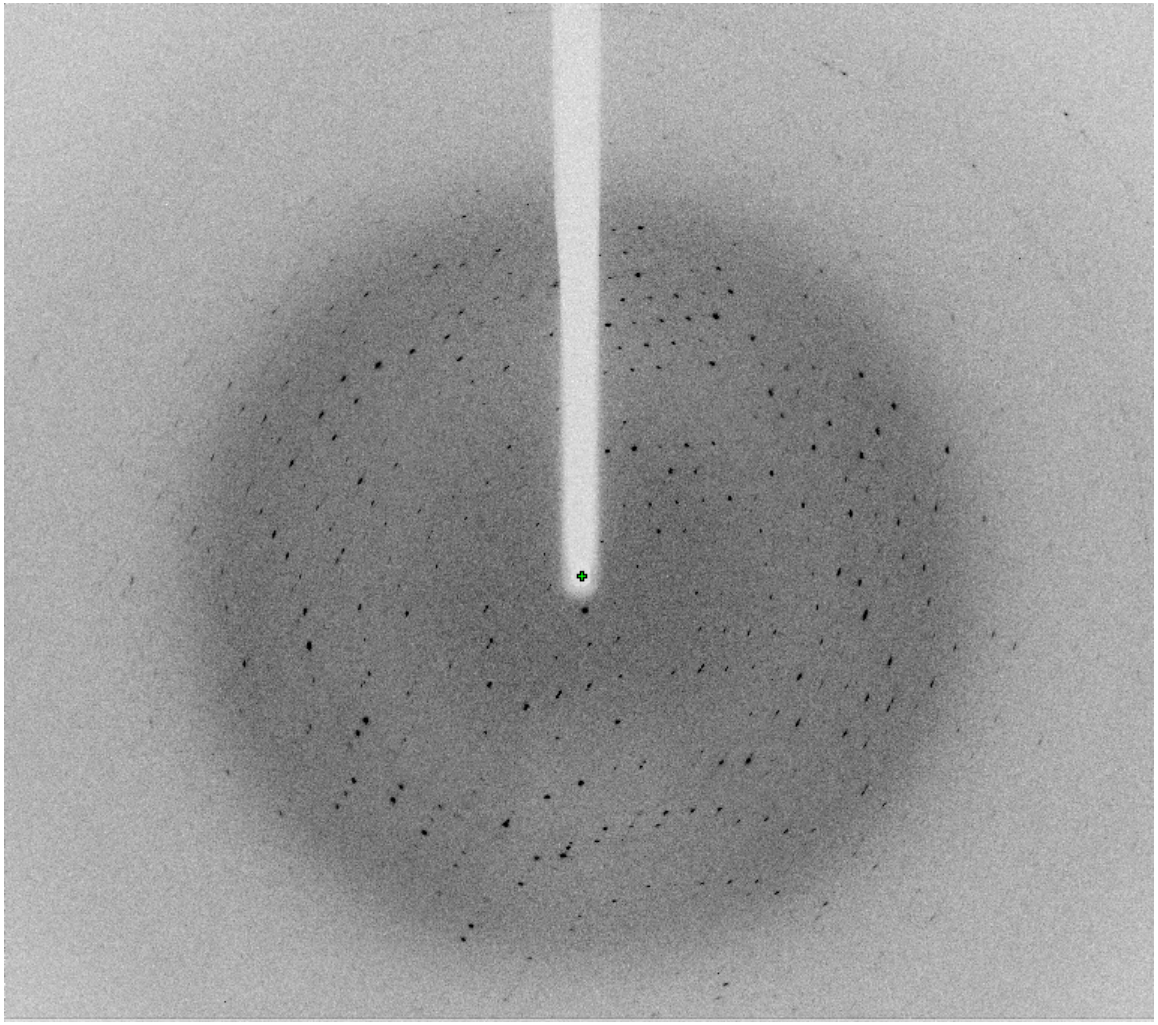
$$I/\sigma(I) = (I_{tot} - I_{bck})/(I_{tot} + I_{bck})^{1/2}$$

For weak reflections, the background will dominate $I/\sigma(I)$.

Fine phi slicing *should* give better data but:

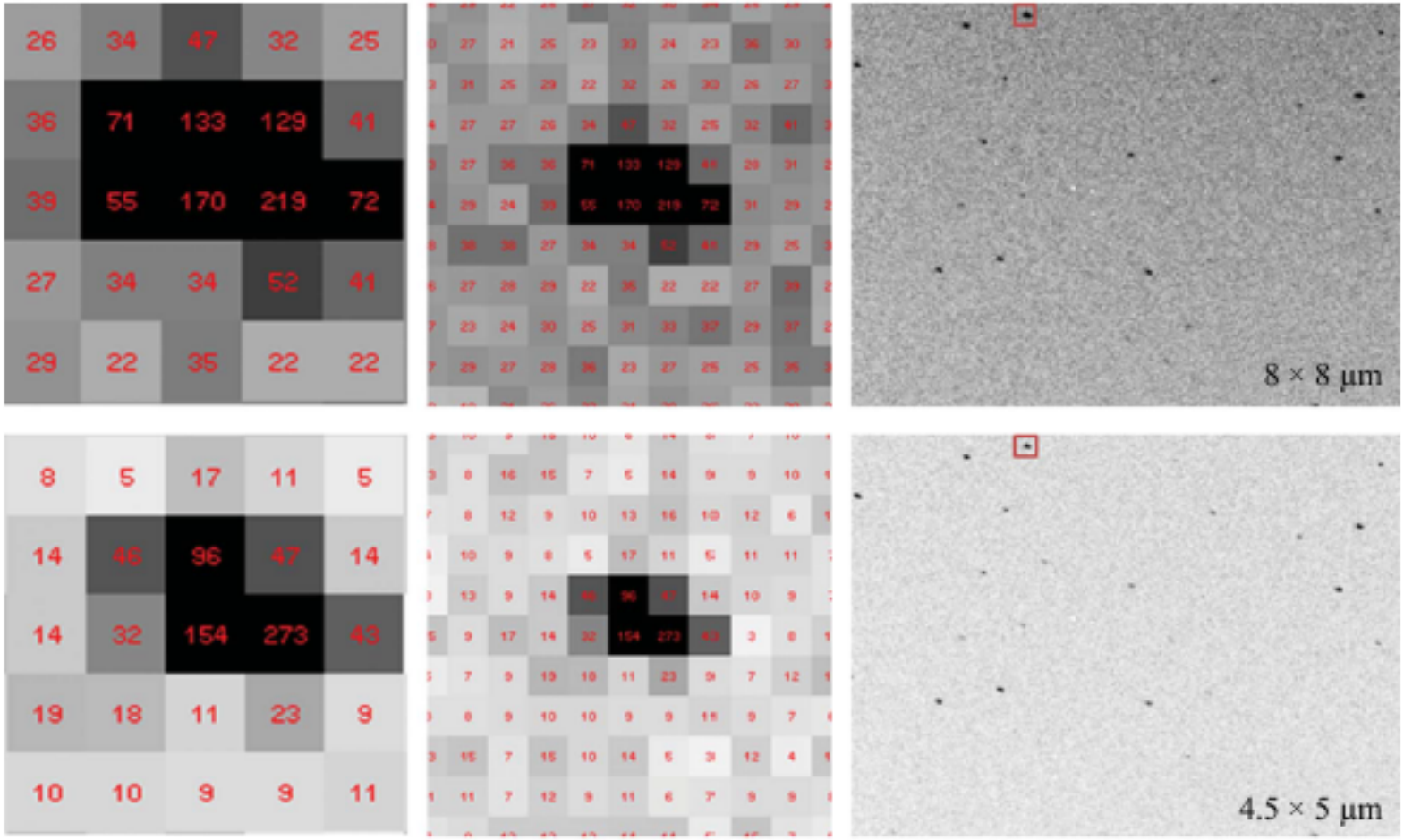
1) Assumes no errors in shutter synchronisation (demanding for very short (<0.5 sec) exposures, assuming a shutter is used).
2) There will be detector readout noise in each image (except Pilatus).

(see J. Pflugrath, Acta Cryst **D55**, 1718-1725, 1999).



Weak, high resolution, diffraction is “lost” in the background

The importance of matching beam size to crystal size



The upper figures represent the diffraction obtained using a 8x8 micron beam with a 5x5x5 micron crystal. In the lower figures, a 5x5 micron beam was used, greatly reducing the background. From Evans et al., Acta **D67**, 261-270, 2011.

Parameter Refinement

Generally, once an orientation matrix and cell parameters have been derived from the autoindexing procedures described, these parameters are refined further using different algorithms.

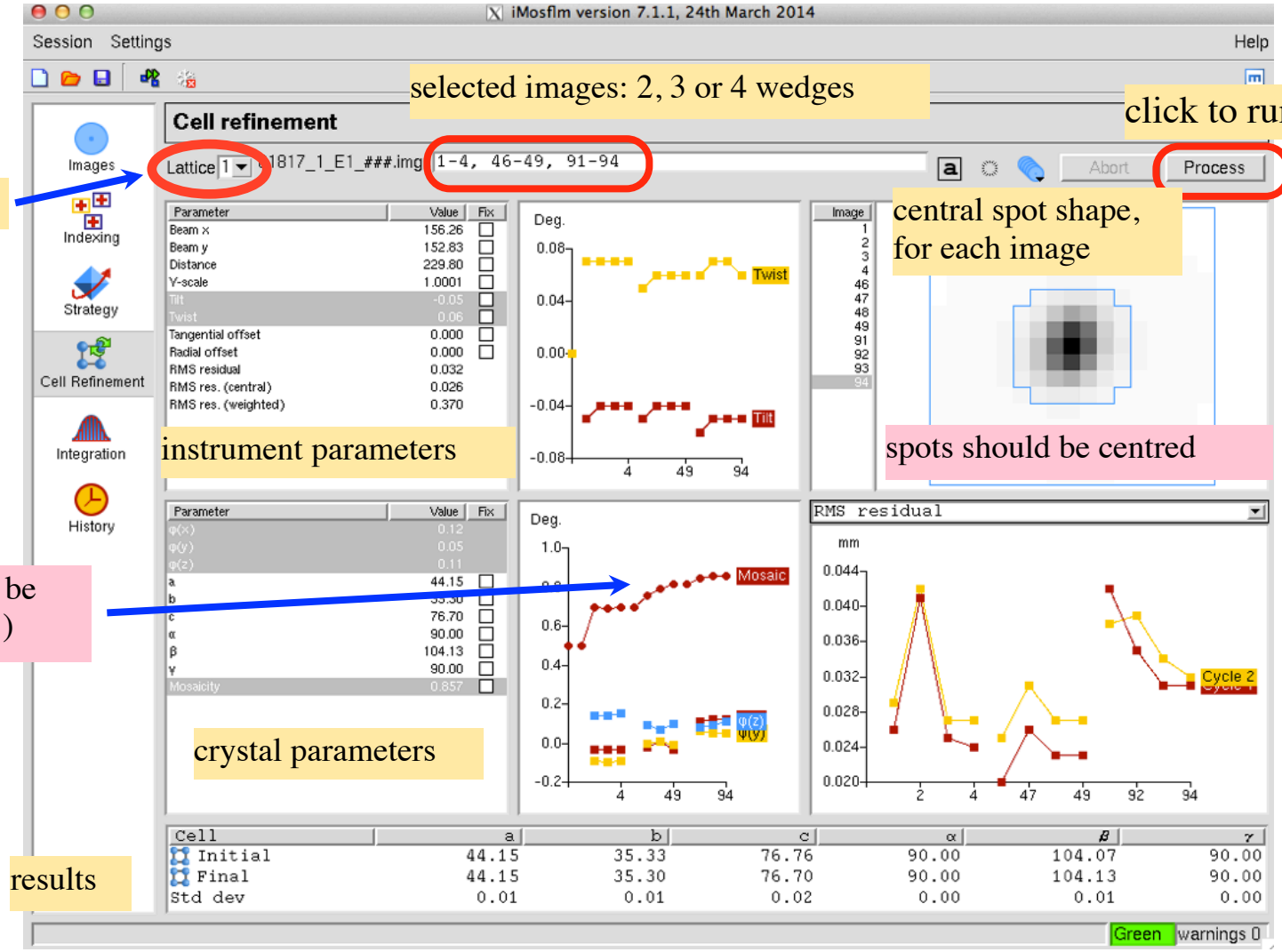
Parameters to be refined:

1) Crystal parameters:
   Cell dimensions, orientation, mosaic spread.

2) Detector parameters:
   Detector position, orientation and (if appropriate) distortion parameters.

3) Beam parameters:
   Orientation, beam divergence.

Cell refinement



It is best to refine the cell from 2 or more widely-spaced wedges, then fix it during integration. Refinement is not reliable at low resolution (< ~ 3.5Å), particularly in low symmetry: just use the cell from indexing with several images (eg 5 or 6)

Integration of the Images

Step 1: Predict the position in the digitised image of each Bragg reflection.

Step 2: Estimate its intensity (need to subtract the X-ray background) and an error estimate of the intensity.

1) Predicting reflection positions

Accuracy in prediction is crucial. Ideally, cell parameters should be known to better than 0.1%. Errors in prediction will introduce systematic errors in profile fitting.

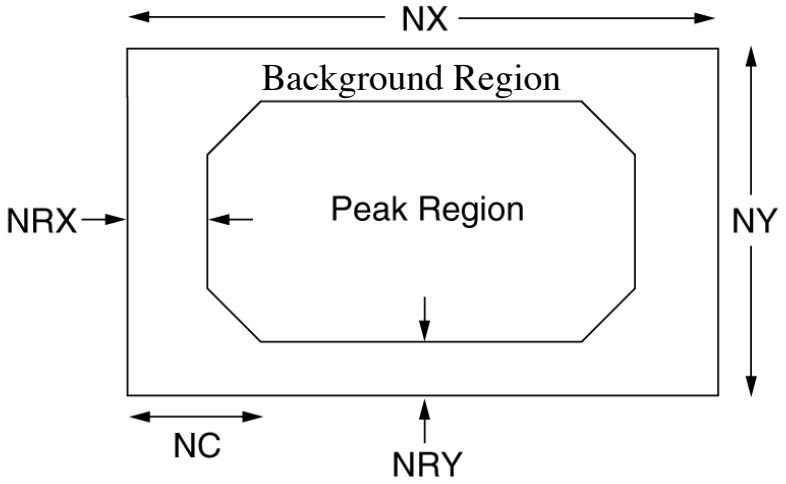
Typically the detector parameters, crystal orientation and mosaic spread will be refined for every image during the integration. The cell parameters are not normally refined.

2) Determining the X-ray background

The background can only be measured in a region around the spot either in two dimensions (X, Y, the detector co-ordinates) for coarse phi-slices or in 3 dimensions (X, Y and phi) for fine phi slices.

Requires definition of a peak/background mask. A plane is fitted to pixels in the background region (rejecting any outliers), and the equation of this background plane is used to calculate the background for pixels in the peak region.

Errors in the mask definition will give systematic errors in intensities.



Summation integration and Profile Fitting

Summation integration:

Sum the pixel values of all pixels in the peak area of the mask, and then subtract the sum of the background values calculated from the background plane for the same pixels.

Profile fitting:

Assume that the shape or profile (in 2 or 3 dimensions) of the spots is known. Then determine the scale factor which, when applied to the known spot profile, gives the best fit to be observed spot profile. This scale factor is then proportional to the profile fitted intensity for the reflection. Minimise:

$$R = \Sigma\, w_i\, (X_i - KP_i)^2$$

X_i is the background subtracted intensity at pixel i

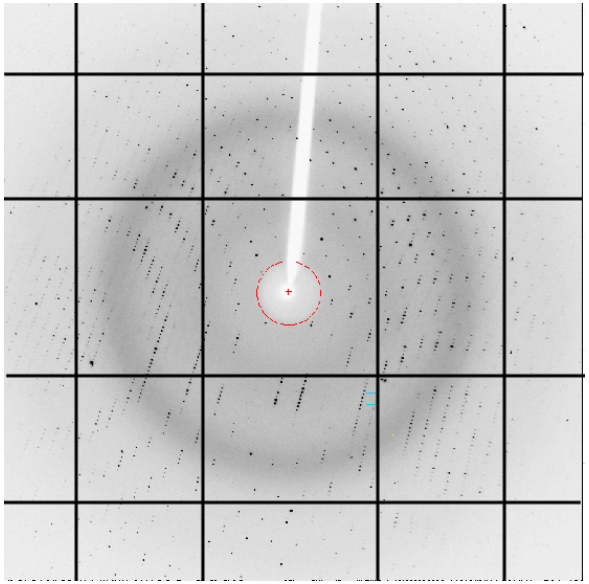
P_i is the value of the standard profile at the corresponding pixel

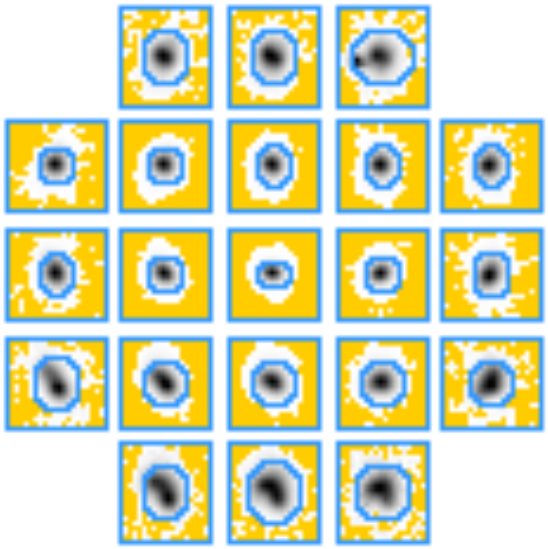
w_i is a weight, derived from the expected variance of X_i

K is the scale factor to be determined

Determining the "Standard" Profile

The profiles are determined empirically (as the average of many spots). The spot shape varies according to position on the detector, and this must be allowed for (different programs do this in different ways).



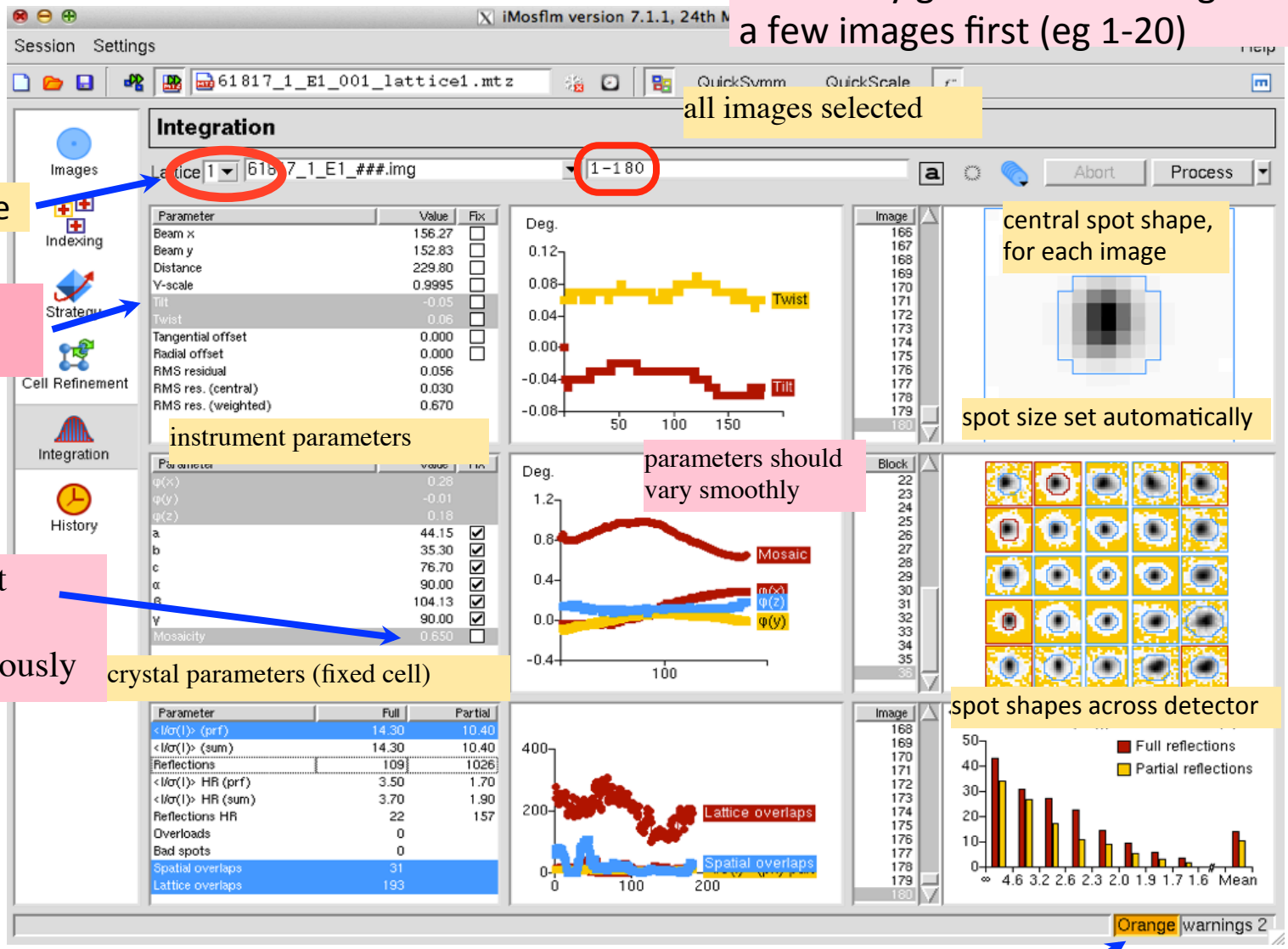


Need to take precautions to avoid introducing systematic errors due to broadening profiles during averaging.

For each reflection integrated, a new profile is calculated as a weighted mean of the “standard” profiles for the adjacent regions.

Integration

It is a very good idea to integrate a few images first (eg 1-20)



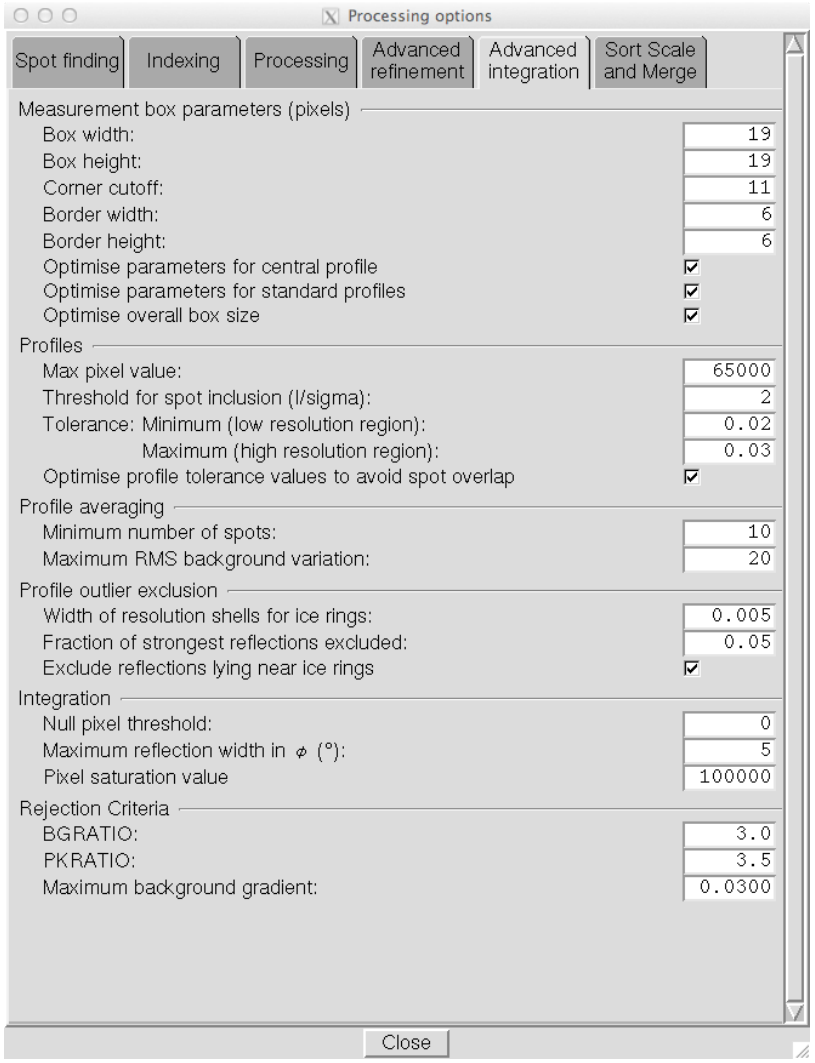
For multiple lattices, integrate each lattice in a separate run, writing a different MTZ file

Challenging Cases

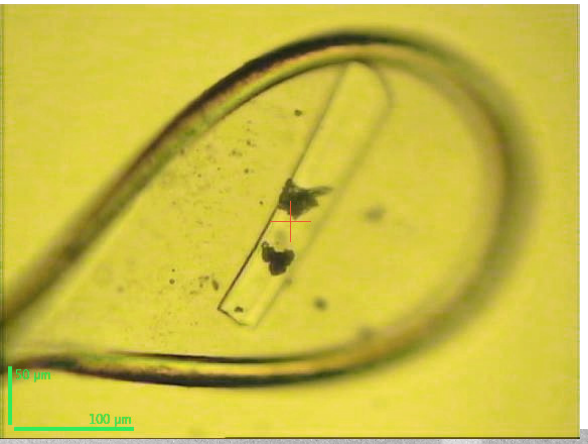
Many parameters can be changed to deal with "difficult" cases. The most common ones are listed below. In addition, for very small cells, the minimum number of reflections for refinement (20) may need to be reduced (Advanced refinement).

Profile tolerances can be increased to reduce the size of the peak region of the measurement box if spot separation is an issue

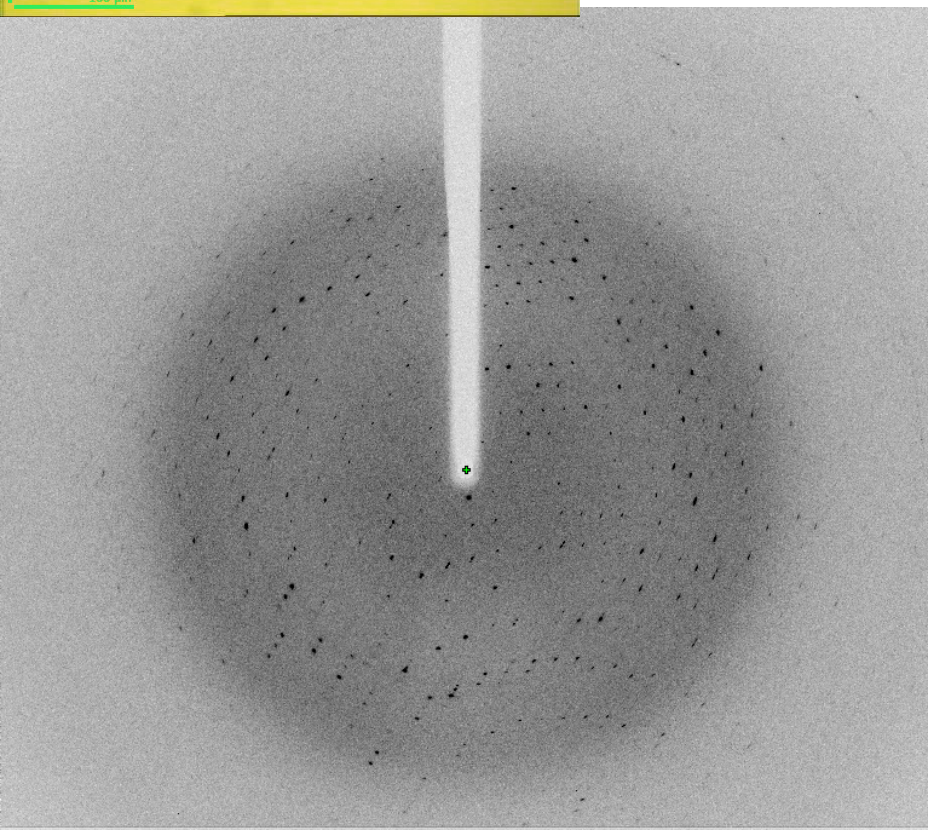
The maximum reflection width should be increased for high mosaicity or low mosaic block size (to reduce the number of "green" prediction boxes)

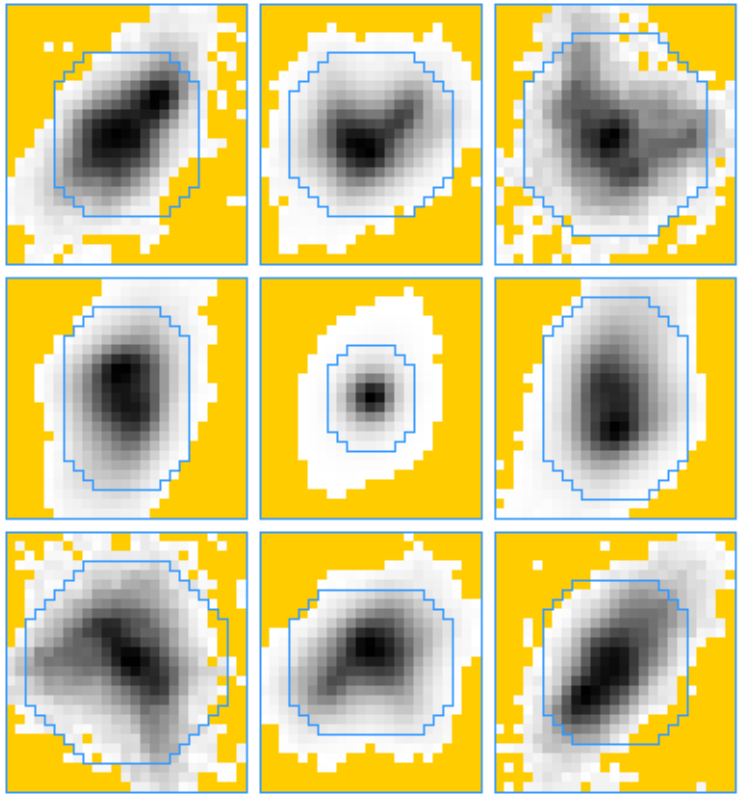


Weak diffraction and poor crystal order results in very variable spot shapes



GPCR crystal, size 240x40x10µ





The advantages of profile fitting

Profile fitting provides a more reliable estimate of the spot intensity for weak reflections than summation integration. It can be shown that the effect of profile fitting is to effectively "down-weight" the peripheral peak pixels (where the signal to noise is lowest) relative to the central pixels (where the signal to noise is highest).

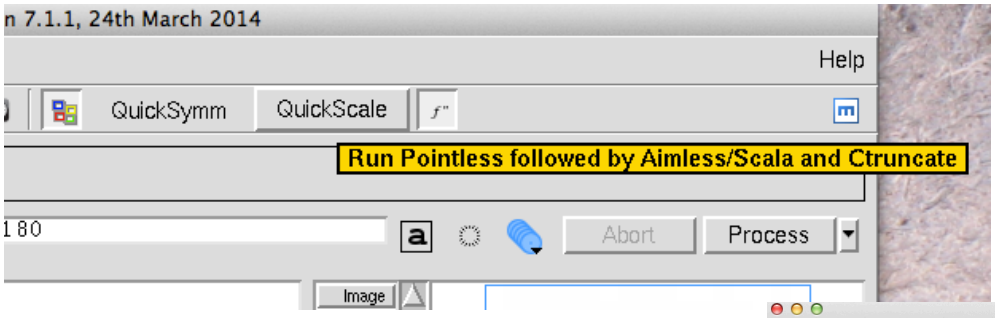
(See Leslie, Acta Cryst **D55**, 1696-1702,1999)

Standard Deviation Estimates

Standard deviations can be obtained based on Poisson statistics (the uncertainty in a total of N photons is √N). This requires a knowledge of the detector "gain", the number that converts pixel counts into the equivalent number of X-ray photons.

These will generally under-estimate the true errors, and should be modified accordingly at the merging step so that they reflect the actual differences between multiple (symmetry related) measurements. It is important to get realistic estimates of the errors in the intensities, because many of the downstream structure solution programs use maximum likelihood algorithms that depend on good error estimates.

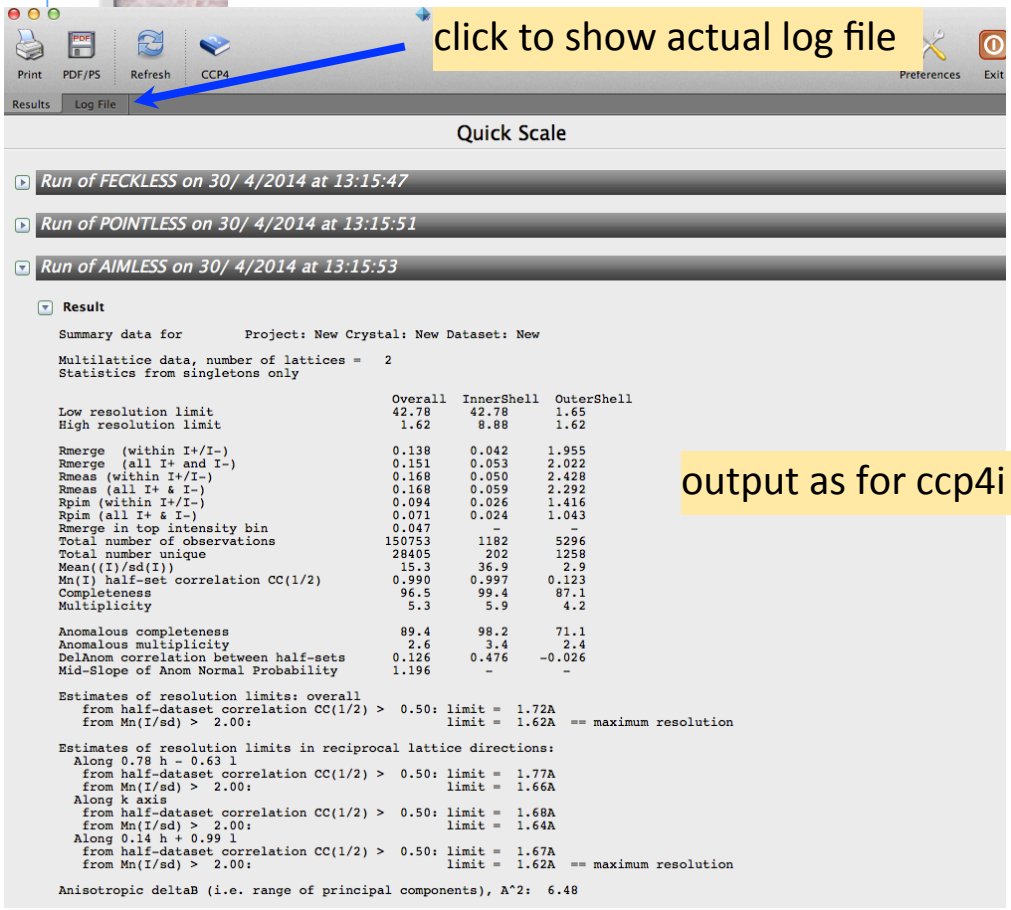
QuickSymm and QuickScale



QuickSymm (POINTLESS) can often be run successfully on just a few degrees of data

QuickScale runs:
[• If there are multiple lattices, FECKLESS to combine lattice files]
• POINTLESS to determine the point group and maybe the space group
• AIMLESS to scale and merge the data
• CTRUNCATE to convert I to F and detect twinning
• FREERFLAG to generate a Free-R set

Options may be set from Settings (more options are available from ccp4i)



output as for ccp4i

Acknowledgements

Harry Powell

Geoff Battye

Luke Kontogiannis

Owen Johnson

Phil Evans and all users for feedback

CCP4, MRC and BBSRC for support