Data processing and scaling with Mosflm and Aimless

James Parkhurst

NJUST / CCP4 crystallography school

September 2015

Acknowledgements

MRC Laboratory of Molecular Biology

CCP4


Andrew Leslie
(MOSFLM)


Harry Powell
(MOSFLM)


Phil Evans
(AIMLESS)


Geoff Battye
(iMOSFLM)

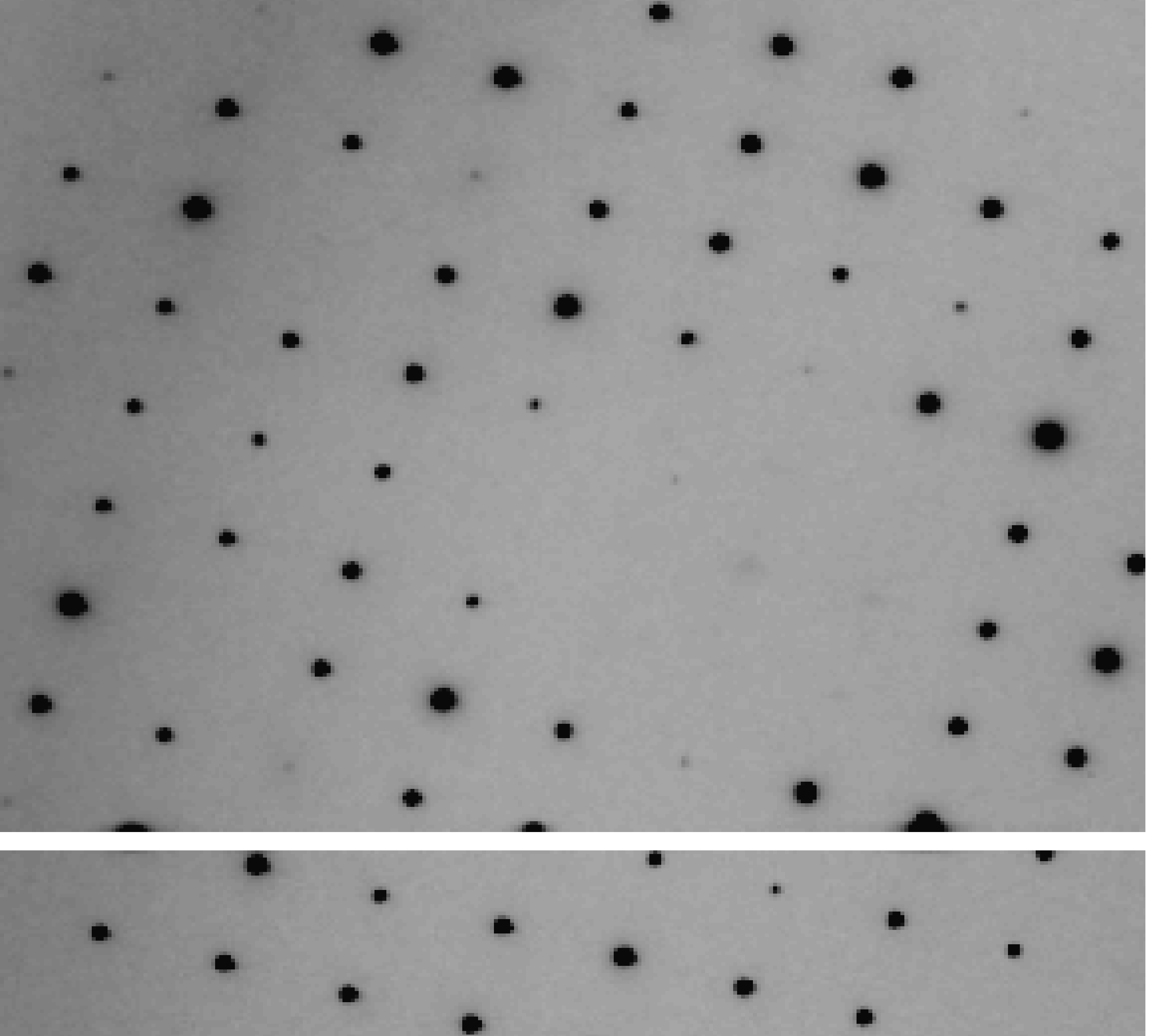

Luke Kontogiannis
(iMOSFLM)


Owen Johnson
(iMOSFLM)

Overview

- How to process and scale your data with Mosflm and Aimless
- Focus on interactive data processing in iMosflm
- Careful scaling using the CCP4i graphical interface

What are we doing and why are we doing it?



Compute the intensity (and structure factor) of each bragg spot in a set of diffraction images

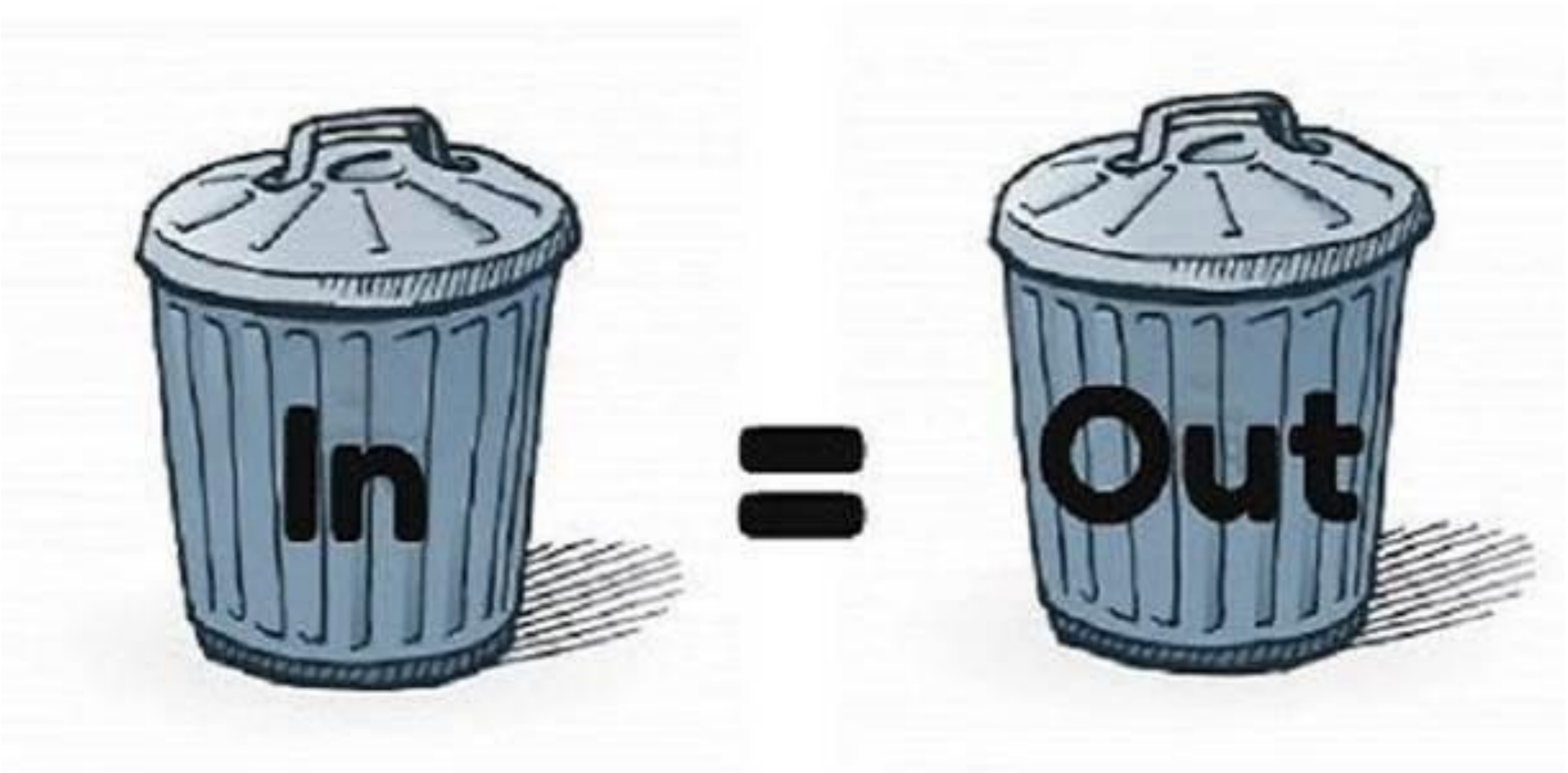
$$|F_{hkl}| = \left(\frac{KI_{hkl}}{Lp}\right)^{\frac{1}{2}}$$

K = constant for given crystal
L = Lorentz factor
p = polarization factor

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl}e^{-2\pi i(hx+ky+lz)}$$

Electron density at every point in the cell depends on the intensity of every reflection. We need to measure out intensities as well as possible!

Warning: garbage in, garbage out

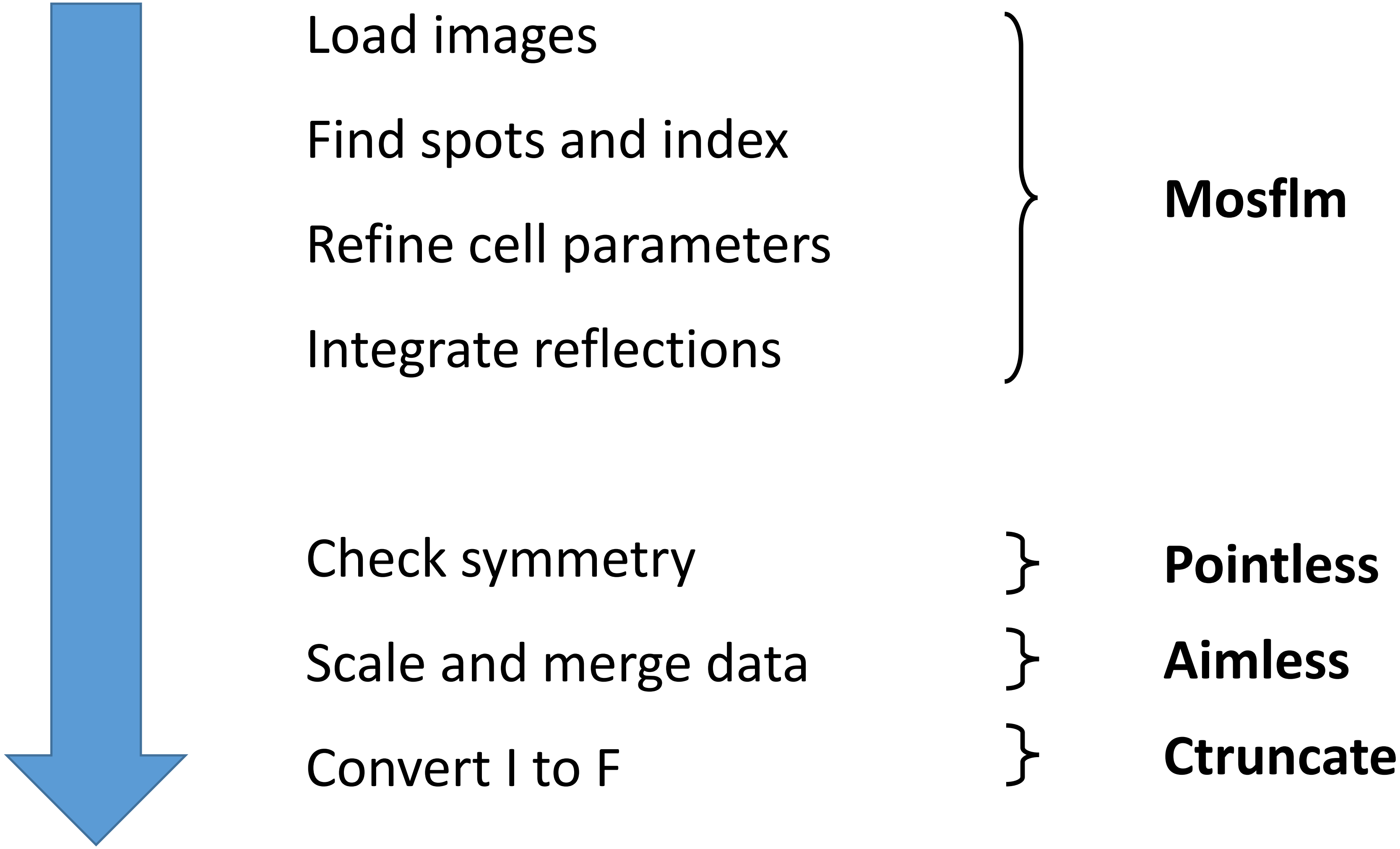


Data collection is the last experimental stage; if you collect bad data you are stuck with it! Data processing programs won’t be able to rescue you!

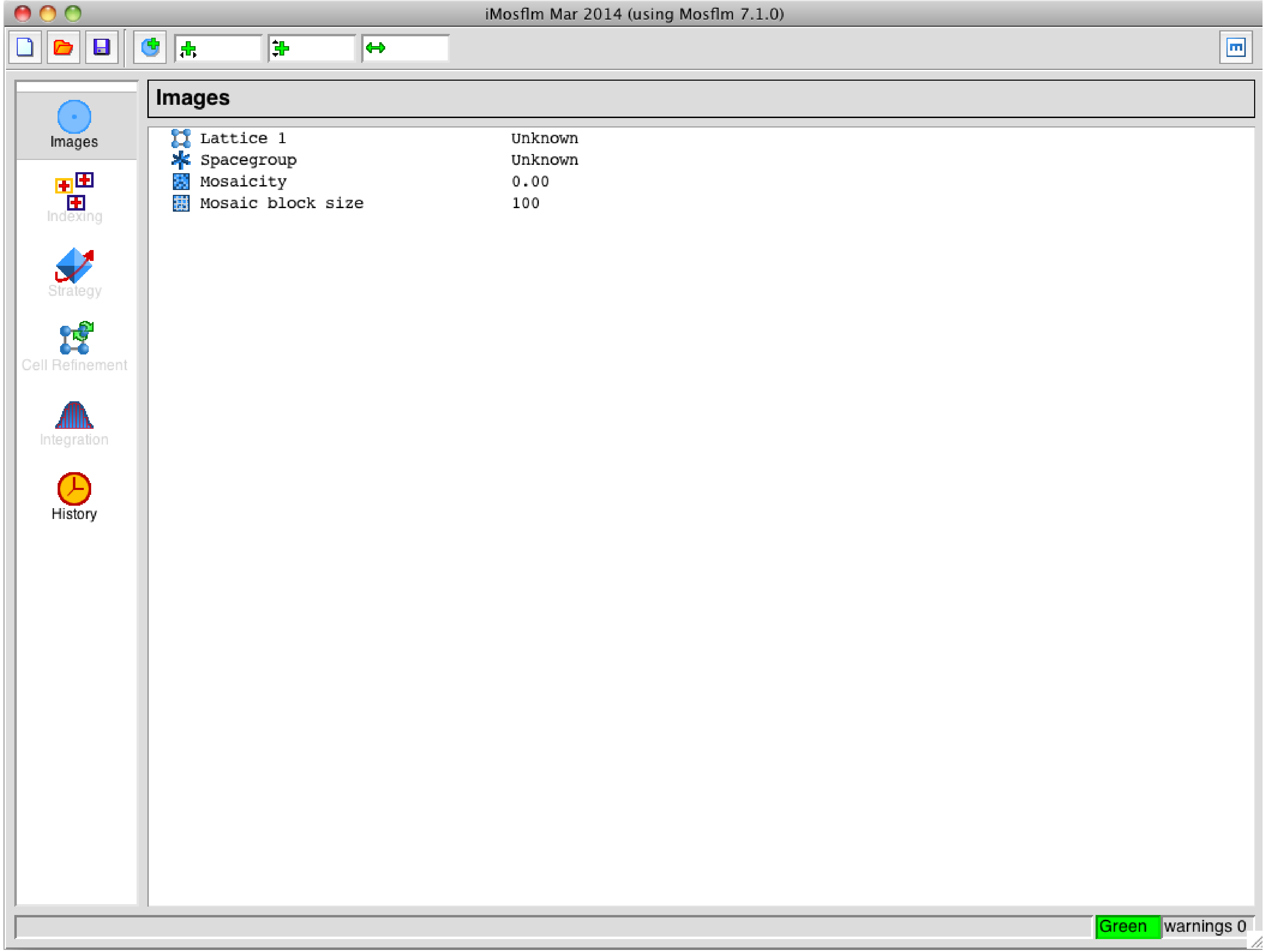
iMosflm

Data processing and scaling with Mosflm and Aimless

Programs



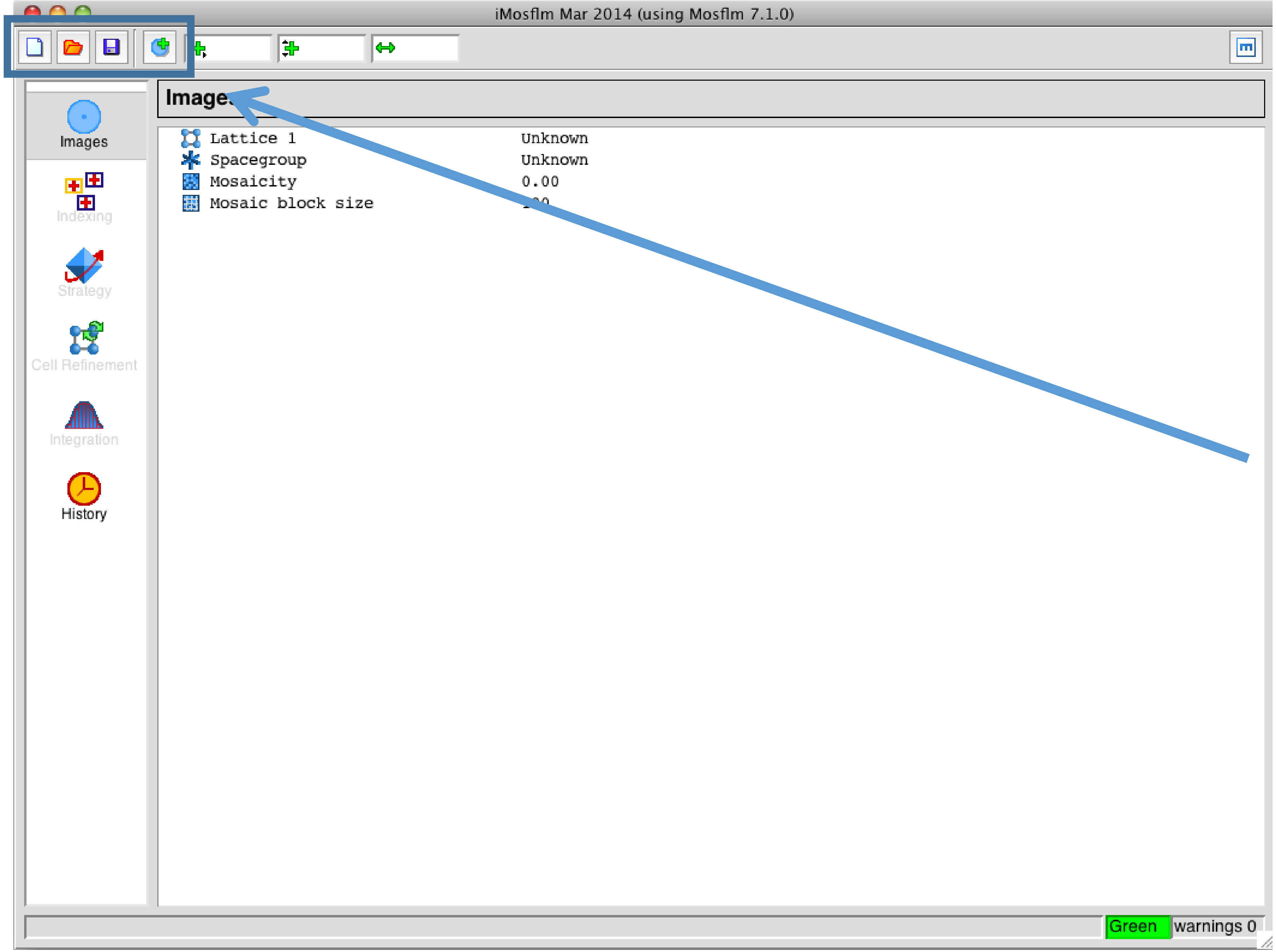
Load images	**Mosflm**
Find spots and index	
Refine cell parameters	
Integrate reflections	
Check symmetry	**Pointless**
Scale and merge data	**Aimless**
Convert I to F	**Ctruncate**



Images

Lattice 1	Unknown
Spacegroup	Unknown
Mosaicity	0.00
Mosaic block size	100

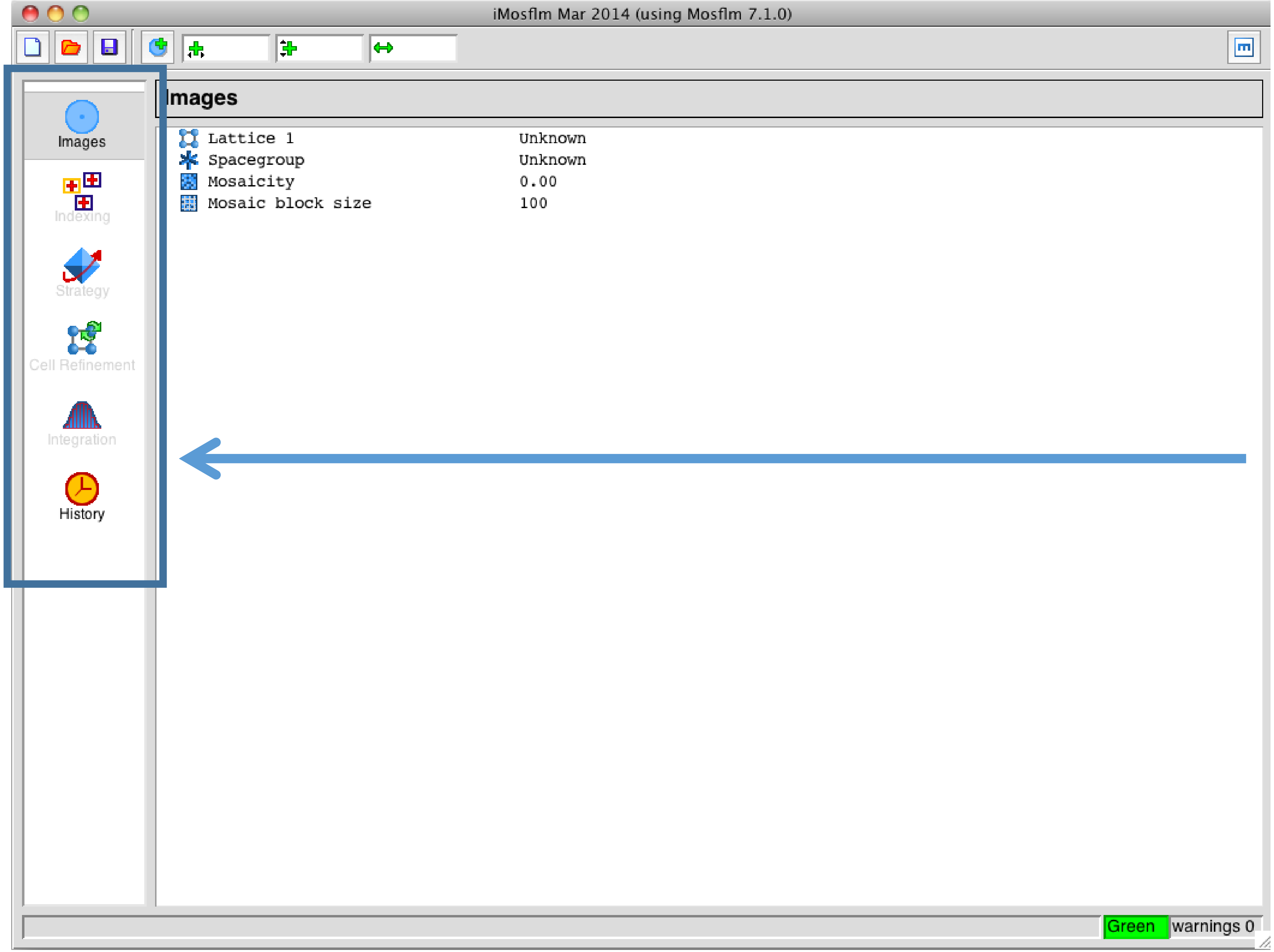
The iMosflm GUI



Lattice 1 Unknown
Spacegroup Unknown
Mosaicity 0.00
Mosaic block size 1.00

Action buttons:

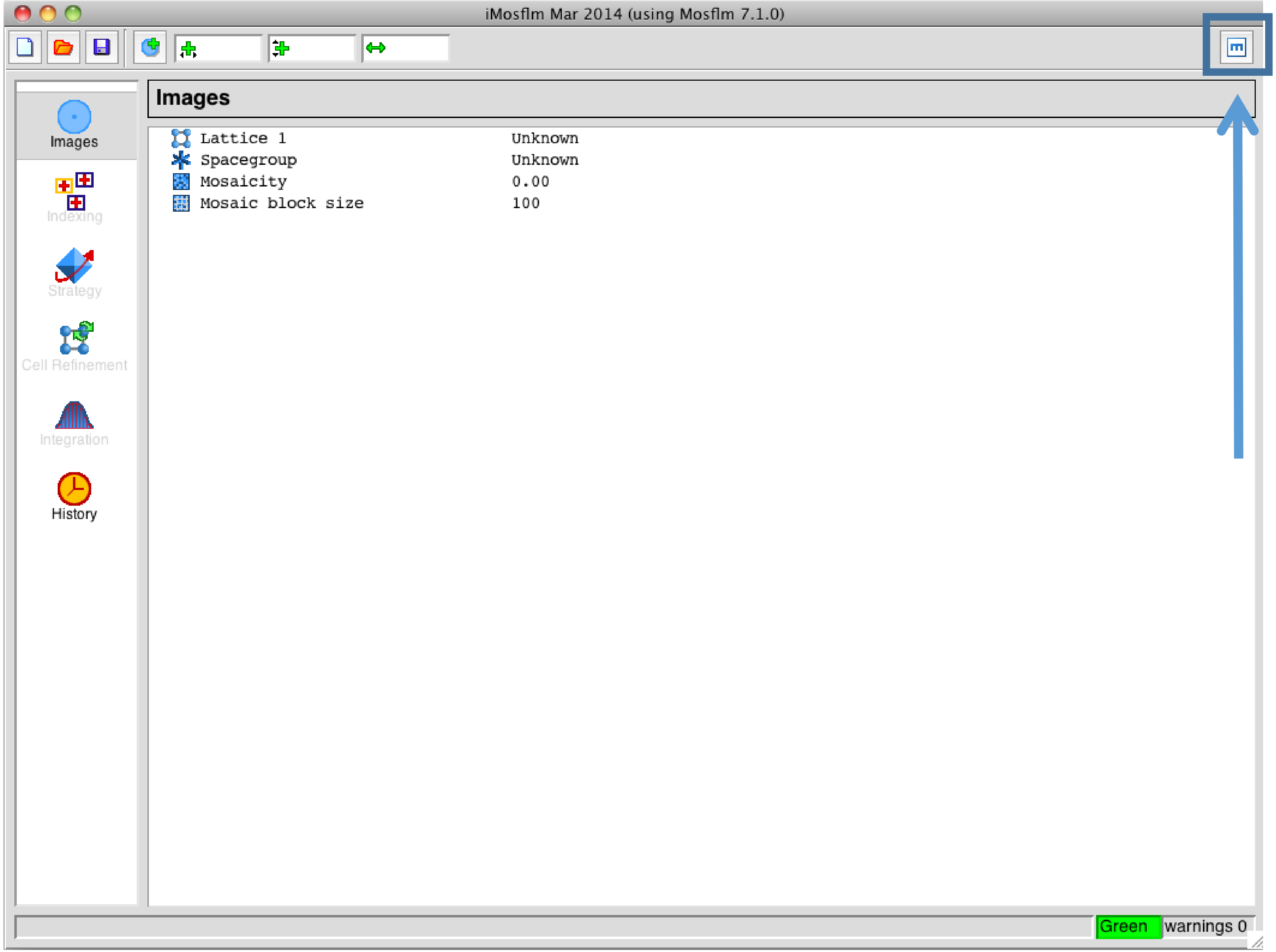
1. Create session
2. Open session
3. Save session
4. Add images



Lattice 1	Unknown
Spacegroup	Unknown
Mosaicity	0.00
Mosaic block size	100

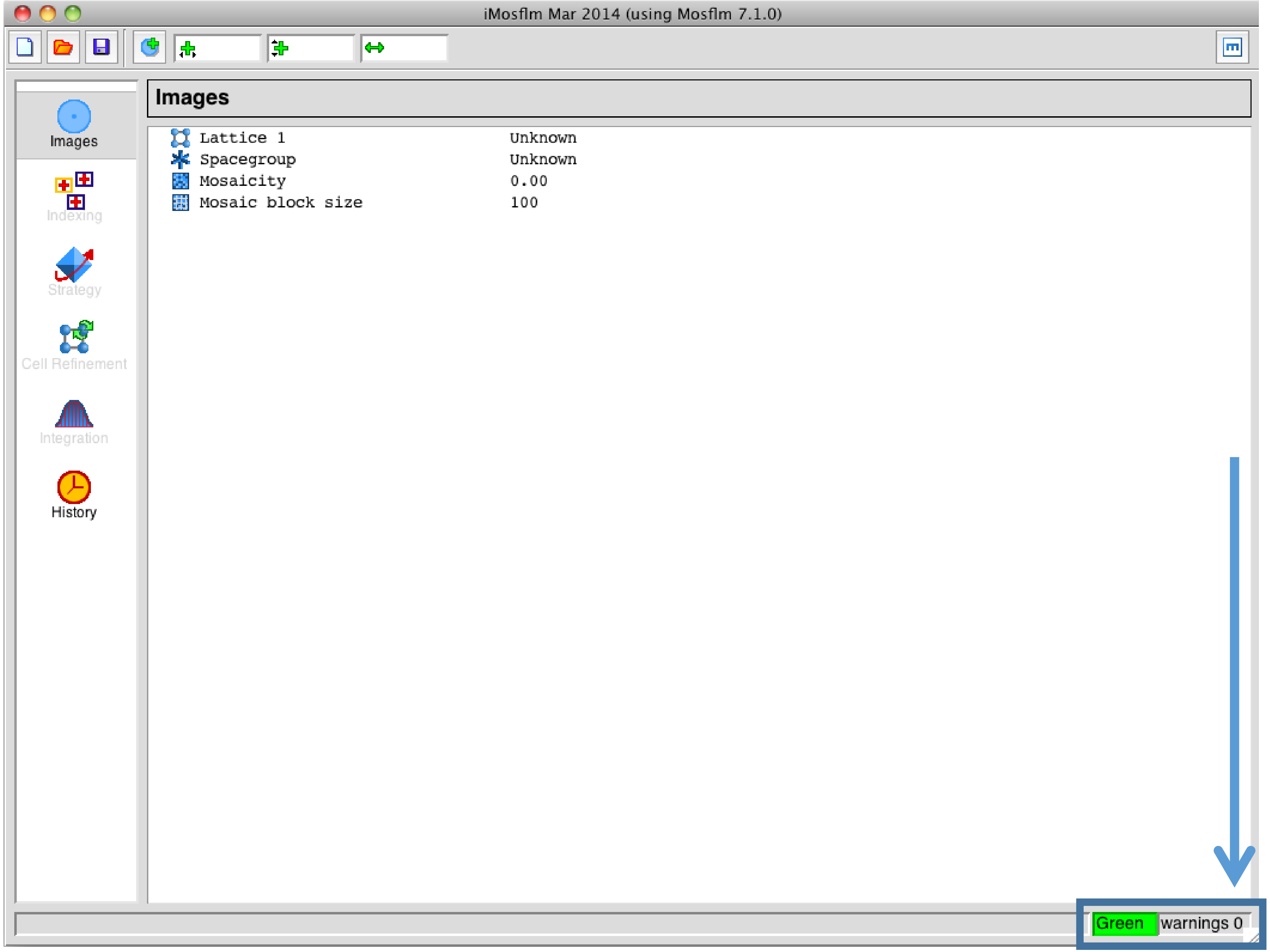
Task palette:

Only available tasks are live, *i.e.* when no images have been loaded, you can only load images (not index, refine, *etc.*)



Lattice 1	Unknown
Spacegroup	Unknown
Mosaicity	0.00
Mosaic block size	100

Activity indicator: "M" spins while a task is active; can be used as to kill mosflm process

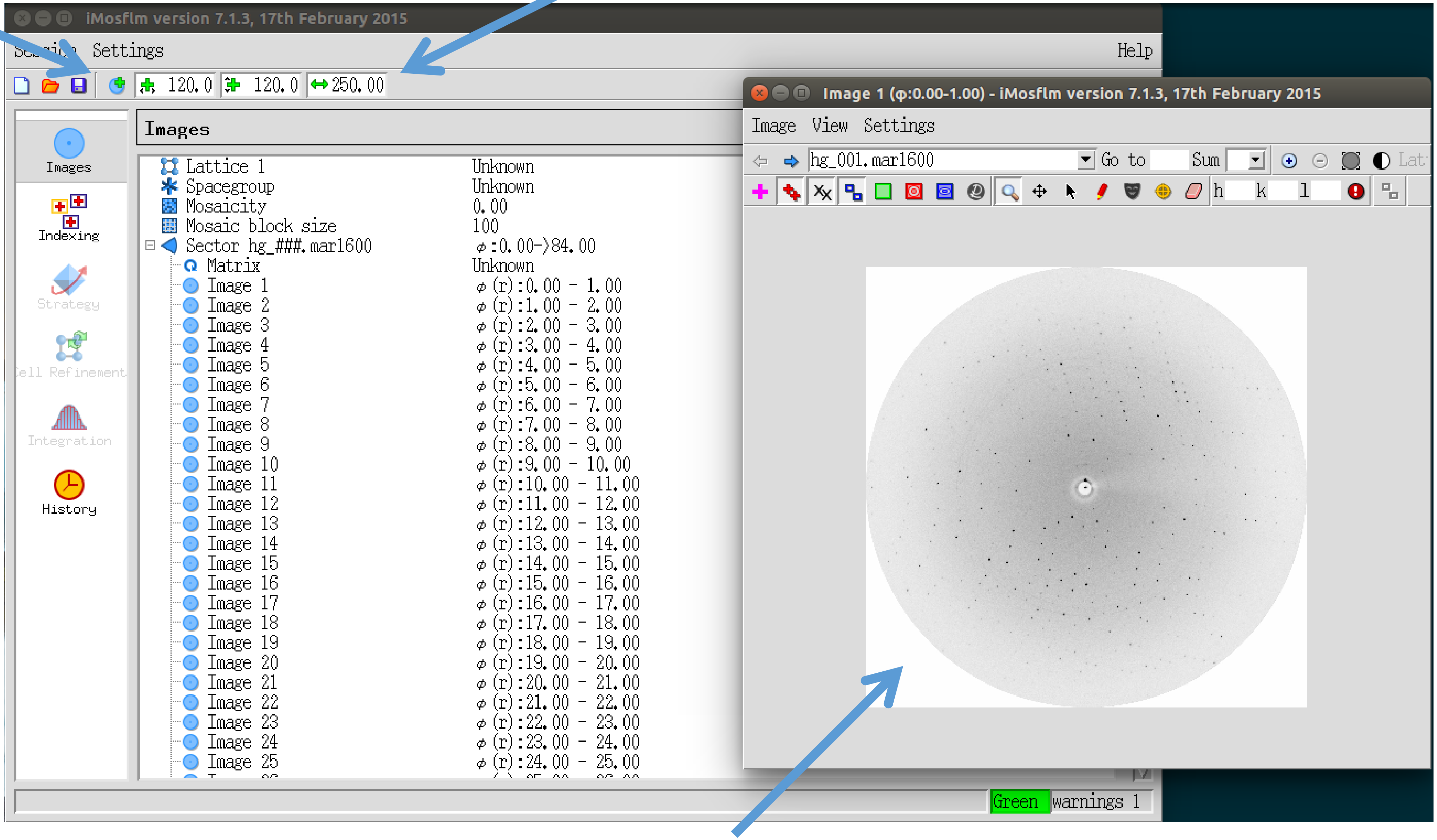


Lattice 1 Unknown
Spacegroup Unknown
Mosaicity 0.00
Mosaic block size 100

Errors and Warnings: Clicking will reveal popup with more information

Click here to load images

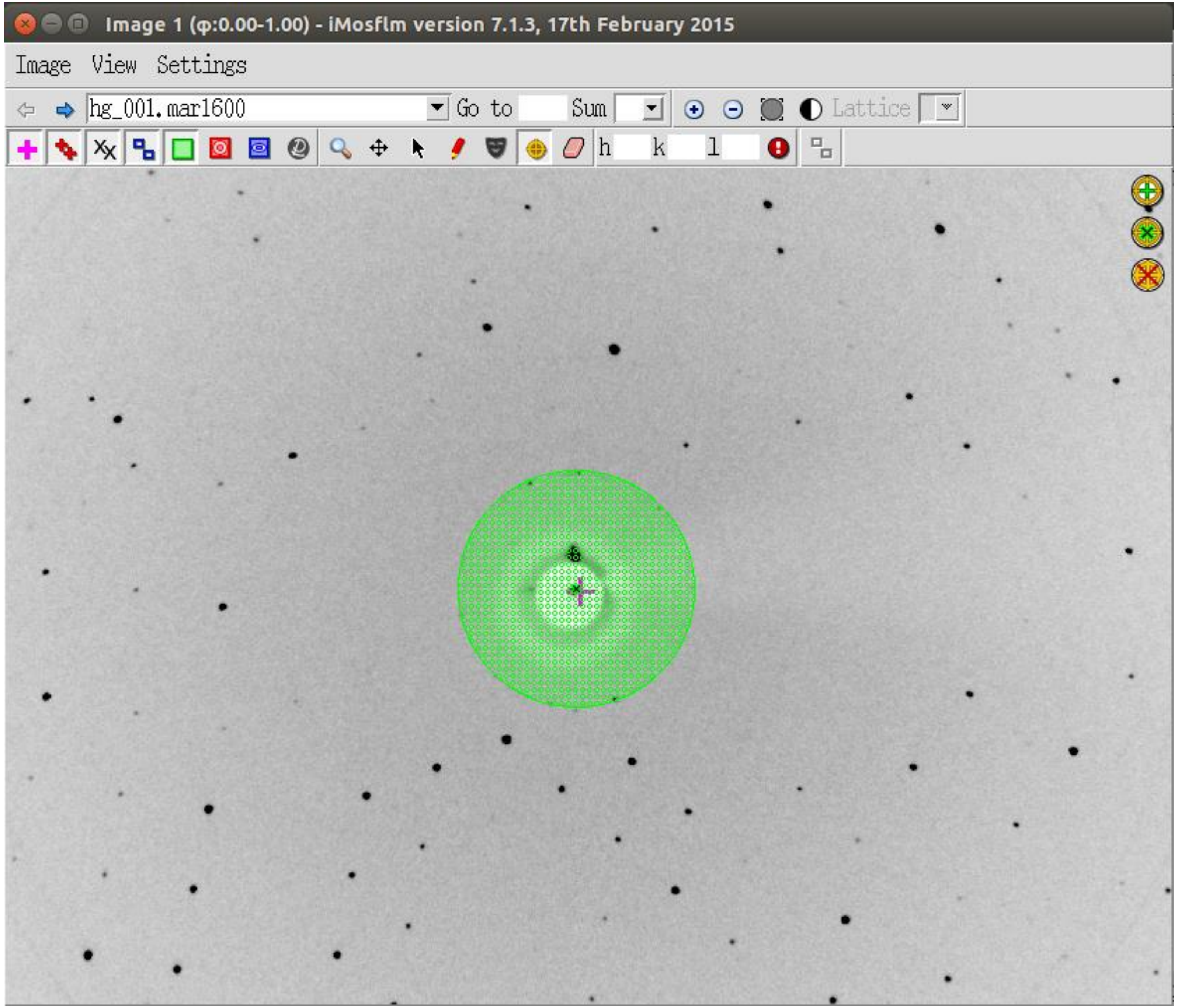
Adjust beam centre and detector distance manually



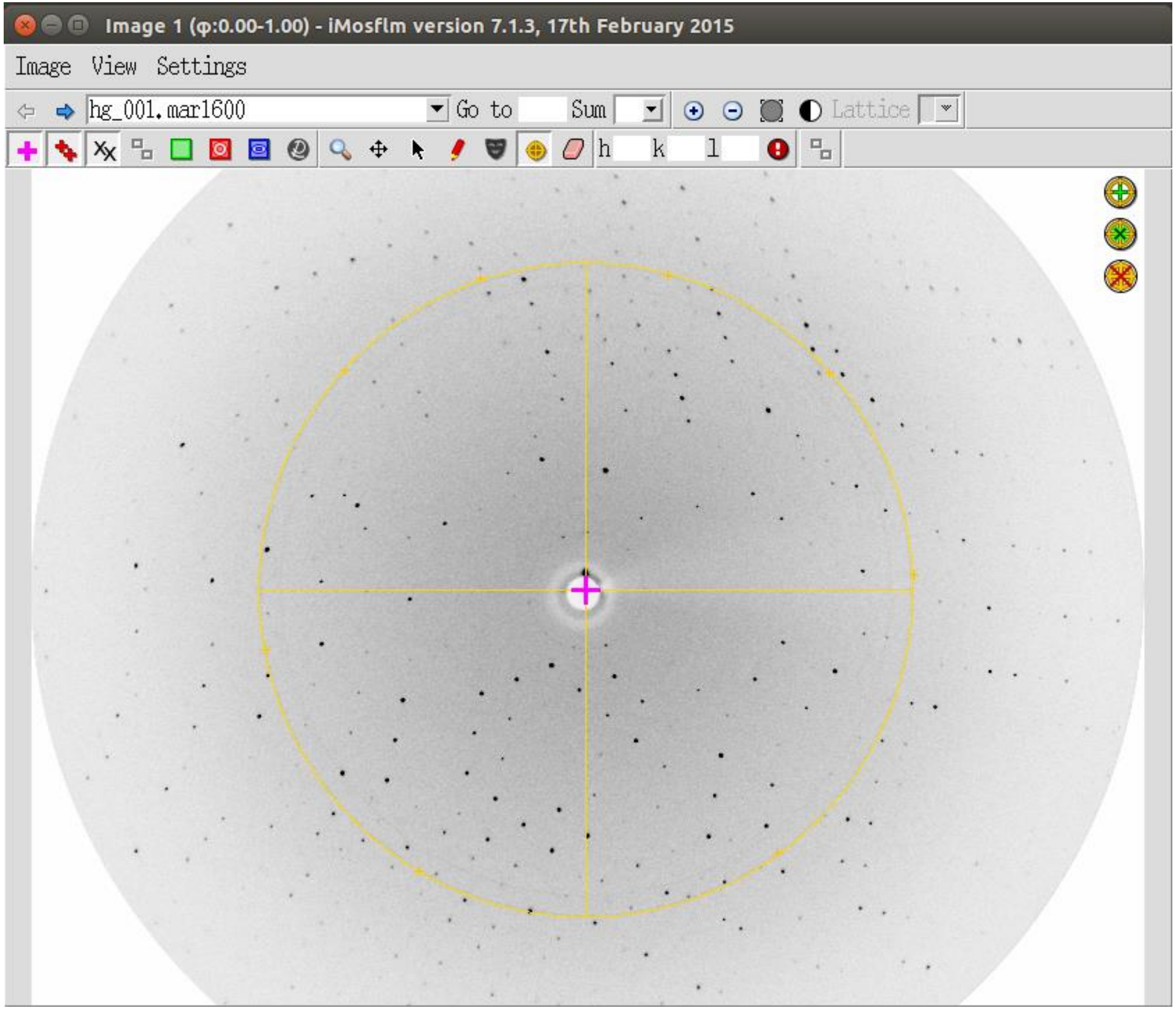
Images

Lattice 1	Unknown
Spacegroup	Unknown
Mosaicity	0.00
Mosaic block size	100
Sector hg_###.mar1600	φ:0.00->84.00
Matrix	Unknown
Image 1	φ(r):0.00 - 1.00
Image 2	φ(r):1.00 - 2.00
Image 3	φ(r):2.00 - 3.00
Image 4	φ(r):3.00 - 4.00
Image 5	φ(r):4.00 - 5.00
Image 6	φ(r):5.00 - 6.00
Image 7	φ(r):6.00 - 7.00
Image 8	φ(r):7.00 - 8.00
Image 9	φ(r):8.00 - 9.00
Image 10	φ(r):9.00 - 10.00
Image 11	φ(r):10.00 - 11.00
Image 12	φ(r):11.00 - 12.00
Image 13	φ(r):12.00 - 13.00
Image 14	φ(r):13.00 - 14.00
Image 15	φ(r):14.00 - 15.00
Image 16	φ(r):15.00 - 16.00
Image 17	φ(r):16.00 - 17.00
Image 18	φ(r):17.00 - 18.00
Image 19	φ(r):18.00 - 19.00
Image 20	φ(r):19.00 - 20.00
Image 21	φ(r):20.00 - 21.00
Image 22	φ(r):21.00 - 22.00
Image 23	φ(r):22.00 - 23.00
Image 24	φ(r):23.00 - 24.00
Image 25	φ(r):24.00 - 25.00

Opens a window for you to inspect images



Mask beam stop shadow
Uses circle fitting to fit the beamstop shadow. This region of the detector will not be used during spot finding or integration.



Locate beam centre

Uses circle fitting to locate direct beam position. It is important to check since you can’t always rely on the information in image headers to be correct!

Indexing

Data processing and scaling with Mosflm and Aimless

Purpose

Things we do know:

- detector position *etc*.
- where spots appear to be
- goniometer settings for each image

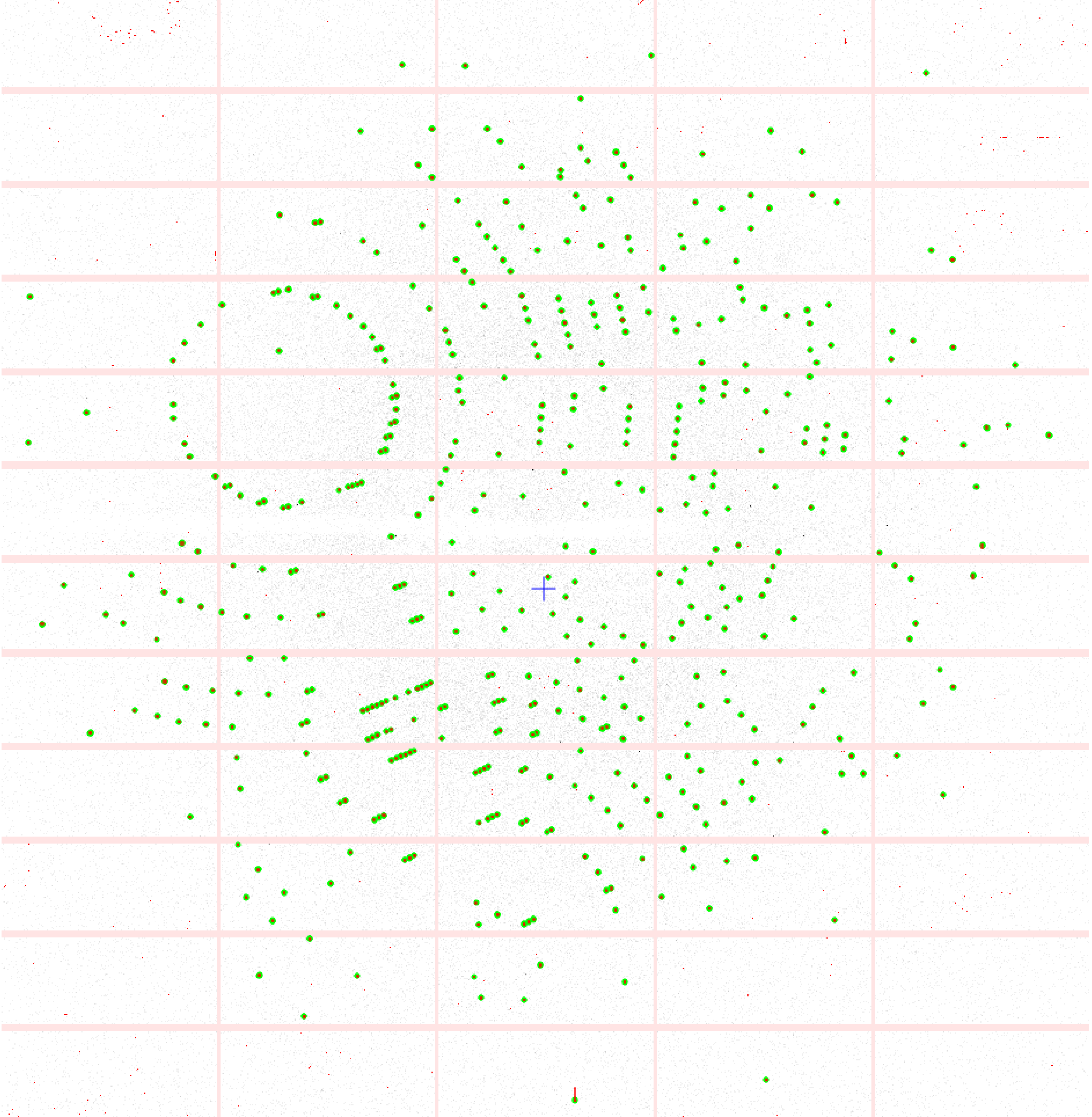
Things we don’t know:

- unit cell
- miller index hkl
- sample orientation
- crystal lattice type

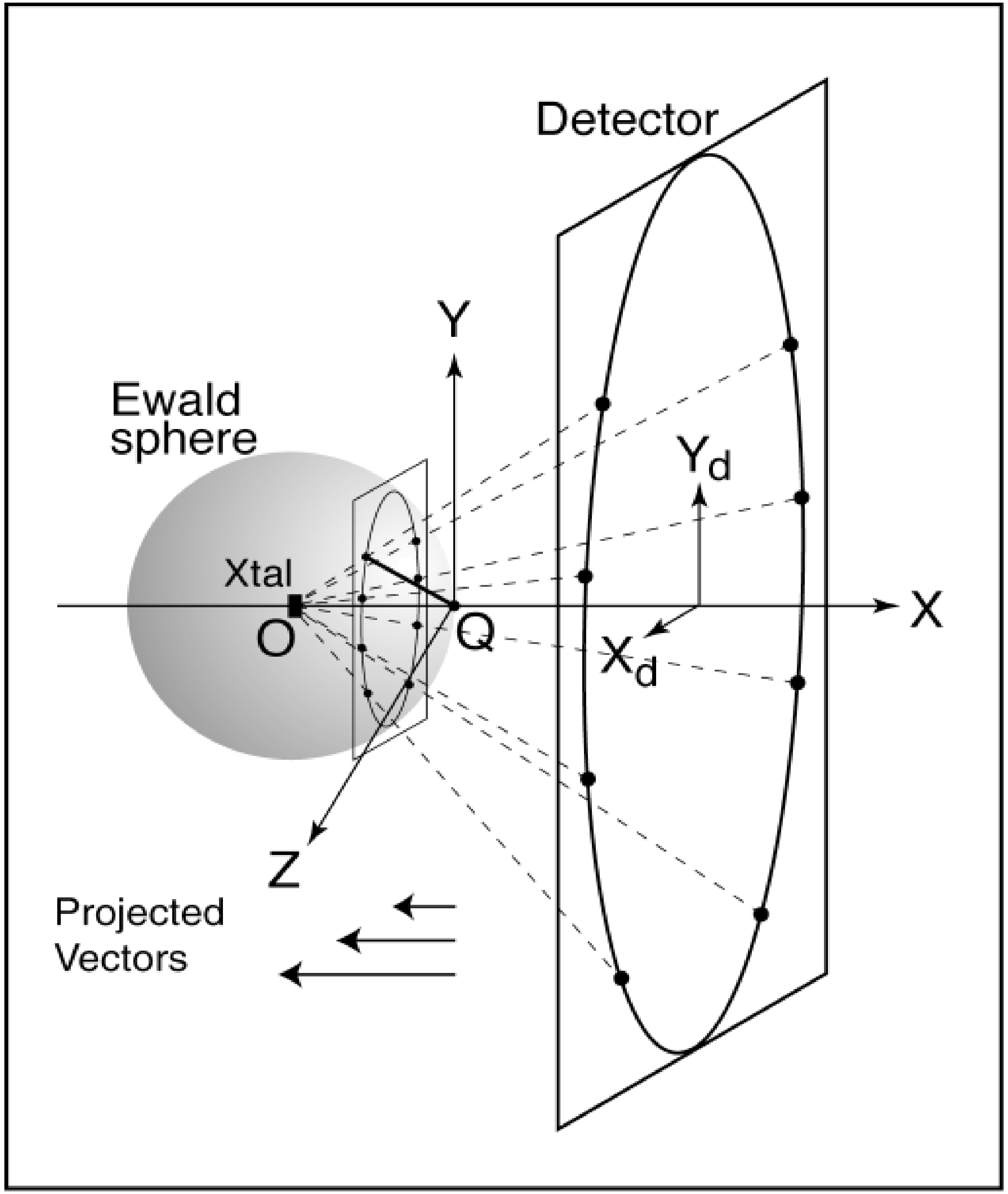
This is Indexing

Indexing in Mosflm

1. Find spots on the image
2. Convert 2D image co-ordinates to 3D scattering vectors
3. Index *via* Fast Fourier Transform
4. Reduce the cell
5. List 44 characteristic lattices for cell, with penalty value for each
6. Pick a likely solution
7. Estimate mosaic spread of the crystal

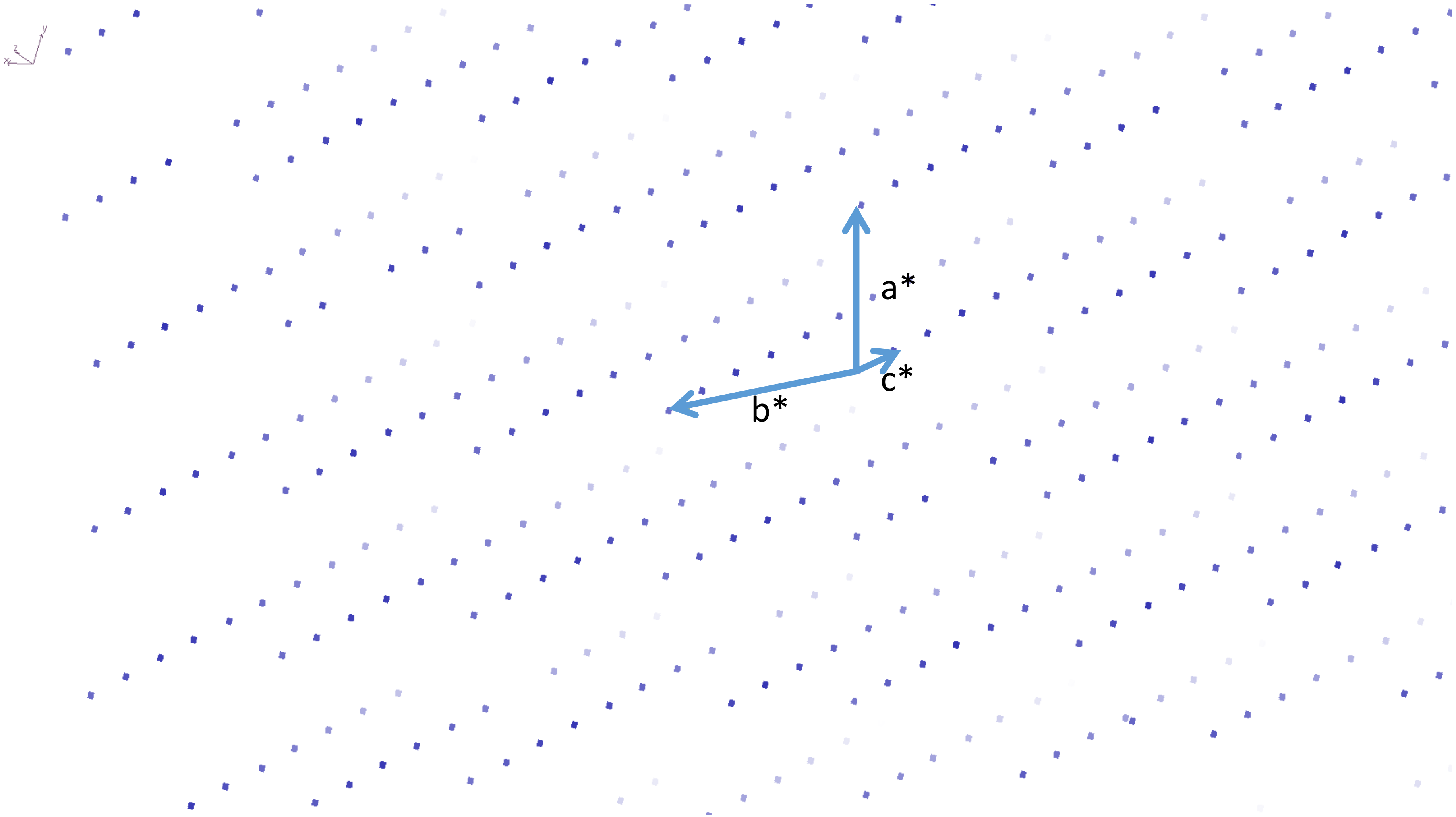


Find spots on the diffraction images

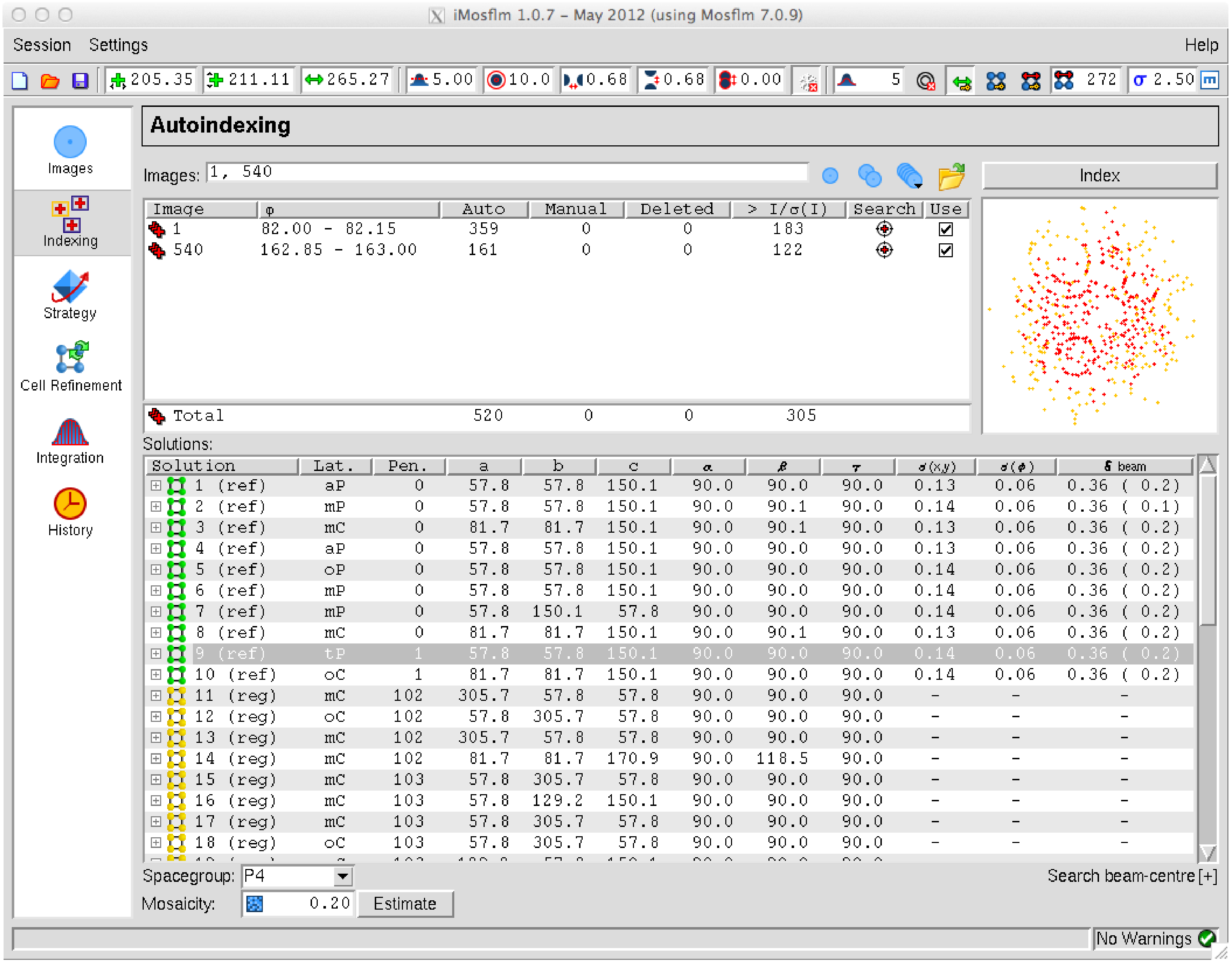


Using the Ewald sphere construction, the observed spots ($X_{d,}$ Y_d, ϕ) can be mapped back into reciprocal space giving a set of scattering vectors $\mathbf{s}_i$

This results in a set of points in reciprocal space



Auto-indexing then attempts to use these points to find the cell and orientation of the crystal



Autoindexing

Images: 1, 540

Image	φ	Auto	Manual	Deleted	> I/σ(I)	Search	Use
1	82.00 – 82.15	359	0	0	183	⊕	☑
540	162.85 – 163.00	161	0	0	122	⊕	☑
Total		520	0	0	305		

Solutions:

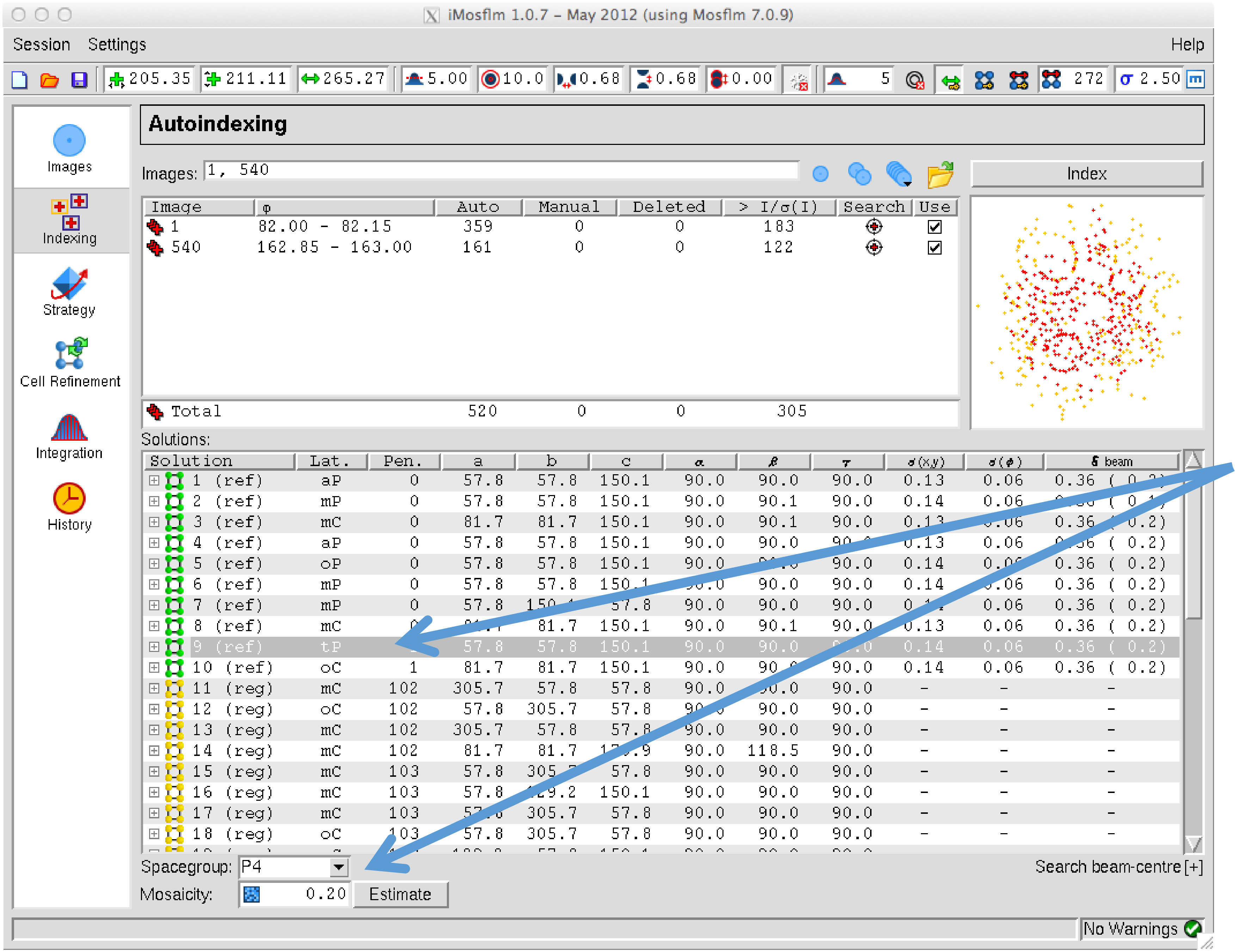
Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
2 (ref)	mP	0	57.8	57.8	150.1	90.0	90.1	90.0	0.14	0.06	0.36 (0.1)
3 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
4 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
5 (ref)	oP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
6 (ref)	mP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
7 (ref)	mP	0	57.8	150.1	57.8	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
8 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
9 (ref)	tP	1	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
10 (ref)	oC	1	81.7	81.7	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
11 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
12 (reg)	oC	102	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
13 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
14 (reg)	mC	102	81.7	81.7	170.9	90.0	118.5	90.0	-	-	-
15 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
16 (reg)	mC	103	57.8	129.2	150.1	90.0	90.0	90.0	-	-	-
17 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
18 (reg)	oC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-

Spacegroup: P4

Mosaicity: 0.20

Indexing tab:

iMosflm finds spots and automatically indexes when you click the indexing tab



Images: 1, 540

Image	φ	Auto	Manual	Deleted	> I/σ(I)	Search	Use
1	82.00 - 82.15	359	0	0	183	⊕	☑
540	162.85 - 163.00	161	0	0	122	⊕	☑
Total		520	0	0	305		

Solutions:

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
2 (ref)	mP	0	57.8	57.8	150.1	90.0	90.1	90.0	0.14	0.06	0.36 (0.1)
3 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
4 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
5 (ref)	oP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
6 (ref)	mP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
7 (ref)	mP	0	57.8	150.1	57.8	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
8 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
9 (ref)	tP	1	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
10 (ref)	oC	1	81.7	81.7	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
11 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
12 (reg)	oC	102	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
13 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
14 (reg)	mC	102	81.7	81.7	[illegible]	90.0	118.5	90.0	-	-	-
15 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
16 (reg)	mC	103	57.8	129.2	150.1	90.0	90.0	90.0	-	-	-
17 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
18 (reg)	oC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-

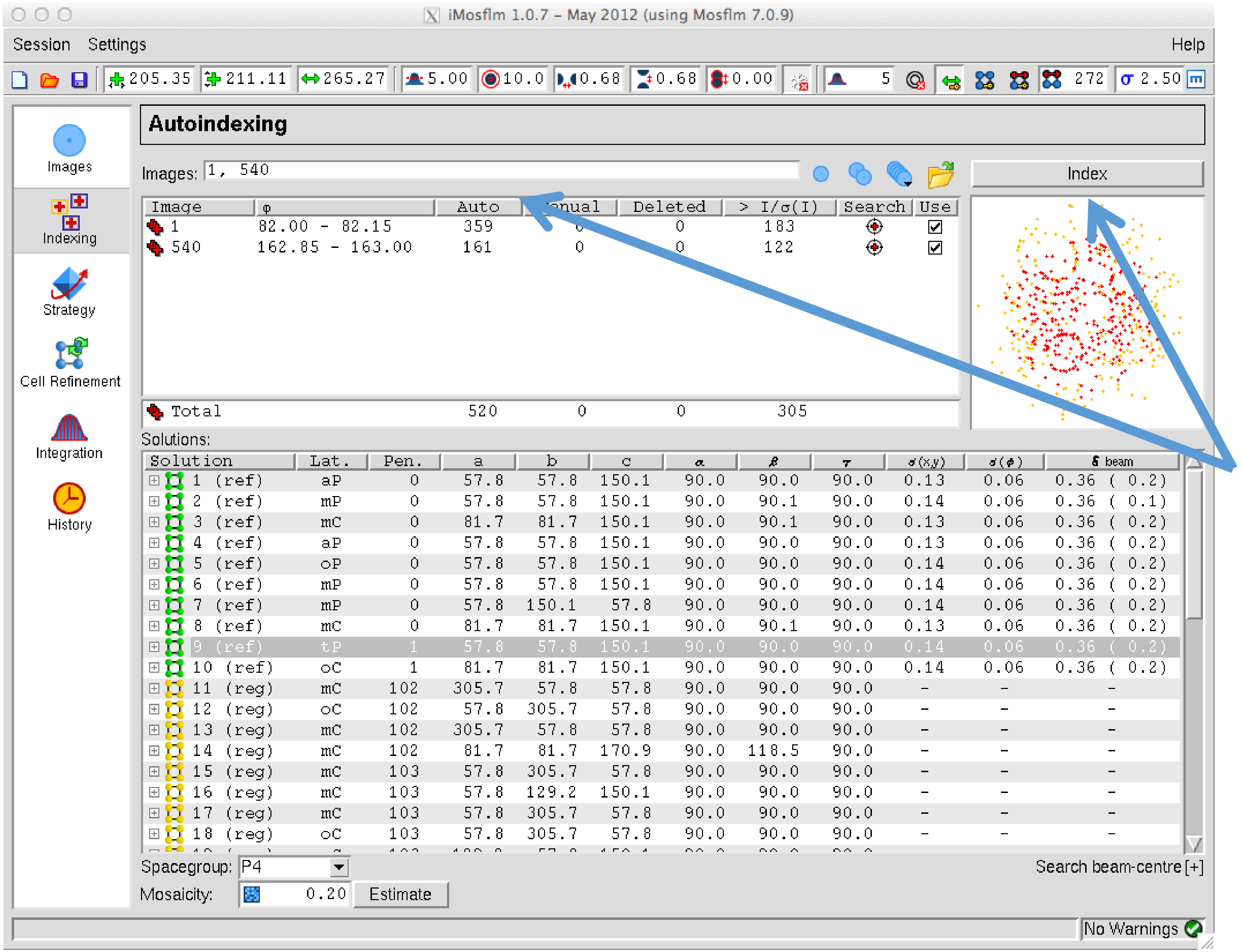
Spacegroup: P4

Search beam-centre [+]

Mosaicity: 0.20 Estimate

Solutions:

The 44 possible solutions are listed with penalty values for each. The most likely solution is highlighted. Also gives the best guess for the lattice.



Images: 1, 540

Image	φ	Auto	nual	Deleted	> I/σ(I)	Search	Use
1	82.00 - 82.15	359	0	0	183	⊕	☑
540	162.85 - 163.00	161	0	0	122	⊕	☑
Total		520	0	0	305		

Solutions:

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
2 (ref)	mP	0	57.8	57.8	150.1	90.0	90.1	90.0	0.14	0.06	0.36 (0.1)
3 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
4 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
5 (ref)	oP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
6 (ref)	mP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
7 (ref)	mP	0	57.8	150.1	57.8	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
8 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
9 (ref)	tP	1	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
10 (ref)	oC	1	81.7	81.7	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
11 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
12 (reg)	oC	102	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
13 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
14 (reg)	mC	102	81.7	81.7	170.9	90.0	118.5	90.0	-	-	-
15 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
16 (reg)	mC	103	57.8	129.2	150.1	90.0	90.0	90.0	-	-	-
17 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
18 (reg)	oC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-

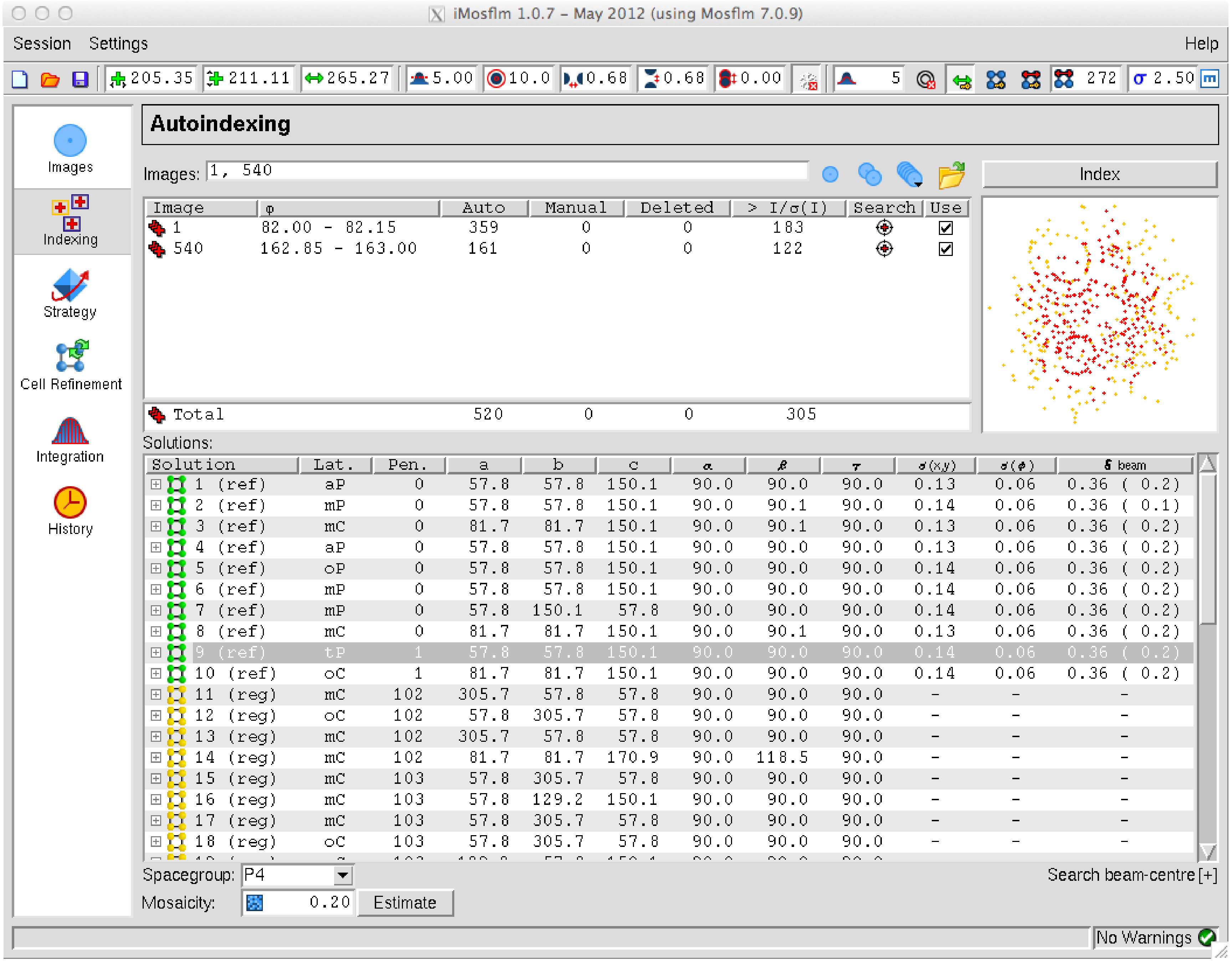
Spacegroup: P4

Search beam-centre [+]

Mosaicity: 0.20 Estimate

Reindexing:

You can choose which images to use and reindex if necessary



Autoindexing

Images: 1, 540

Image	φ	Auto	Manual	Deleted	> I/σ(I)	Search	Use
1	82.00 – 82.15	359	0	0	183	⌖	☑
540	162.85 – 163.00	161	0	0	122	⌖	☑
Total		520	0	0	305		

Solutions:

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
2 (ref)	mP	0	57.8	57.8	150.1	90.0	90.1	90.0	0.14	0.06	0.36 (0.1)
3 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
4 (ref)	aP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.13	0.06	0.36 (0.2)
5 (ref)	oP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
6 (ref)	mP	0	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
7 (ref)	mP	0	57.8	150.1	57.8	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
8 (ref)	mC	0	81.7	81.7	150.1	90.0	90.1	90.0	0.13	0.06	0.36 (0.2)
9 (ref)	tP	1	57.8	57.8	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
10 (ref)	oC	1	81.7	81.7	150.1	90.0	90.0	90.0	0.14	0.06	0.36 (0.2)
11 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
12 (reg)	oC	102	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
13 (reg)	mC	102	305.7	57.8	57.8	90.0	90.0	90.0	-	-	-
14 (reg)	mC	102	81.7	81.7	170.9	90.0	118.5	90.0	-	-	-
15 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
16 (reg)	mC	103	57.8	129.2	150.1	90.0	90.0	90.0	-	-	-
17 (reg)	mC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-
18 (reg)	oC	103	57.8	305.7	57.8	90.0	90.0	90.0	-	-	-

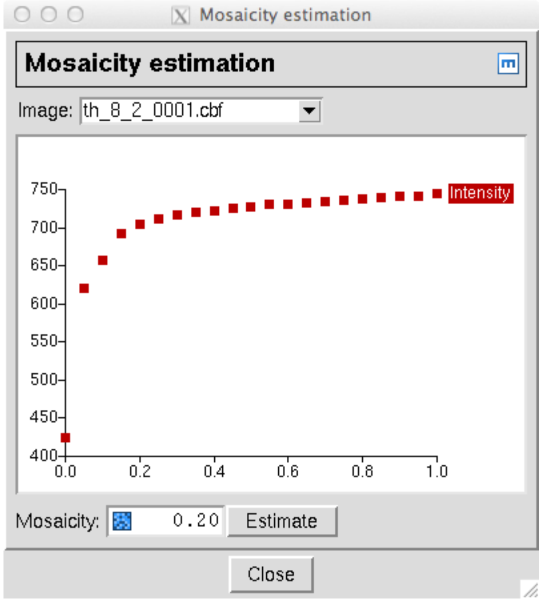
Spacegroup: P4

Search beam-centre [+]

Mosaicity: 0.20 Estimate

Mosaicity:

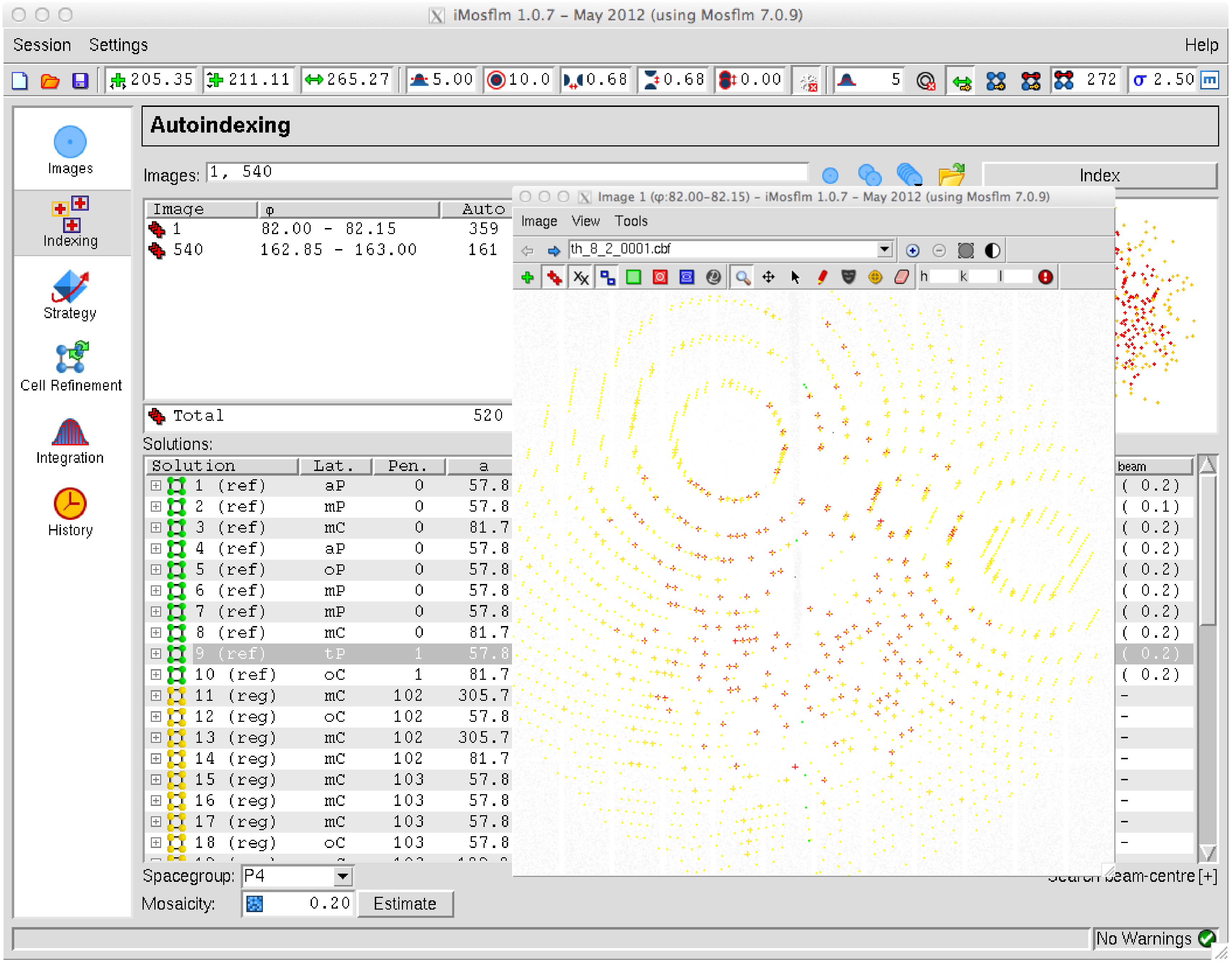
During indexing, mosflm will also estimate the mosaicity



Mosaicity estimation

Image: th_8_2_0001.cbf

Mosaicity: 0.20



Autoindexing

Images: 1, 540

th_8_2_0001.cbf

Image	φ	Auto
1	82.00 – 82.15	359
540	162.85 – 163.00	161
Total		520

Solutions:

Solution	Lat.	Pen.	a	beam
1 (ref)	aP	0	57.8	(0.2)
2 (ref)	mP	0	57.8	(0.1)
3 (ref)	mC	0	81.7	(0.2)
4 (ref)	aP	0	57.8	(0.2)
5 (ref)	oP	0	57.8	(0.2)
6 (ref)	mP	0	57.8	(0.2)
7 (ref)	mP	0	57.8	(0.2)
8 (ref)	mC	0	81.7	(0.2)
9 (ref)	tP	1	57.8	(0.2)
10 (ref)	oC	1	81.7	(0.2)
11 (reg)	mC	102	305.7	-
12 (reg)	oC	102	57.8	-
13 (reg)	mC	102	305.7	-
14 (reg)	mC	102	81.7	-
15 (reg)	mC	103	57.8	-
16 (reg)	mC	103	57.8	-
17 (reg)	mC	103	57.8	-
18 (reg)	oC	103	57.8	-

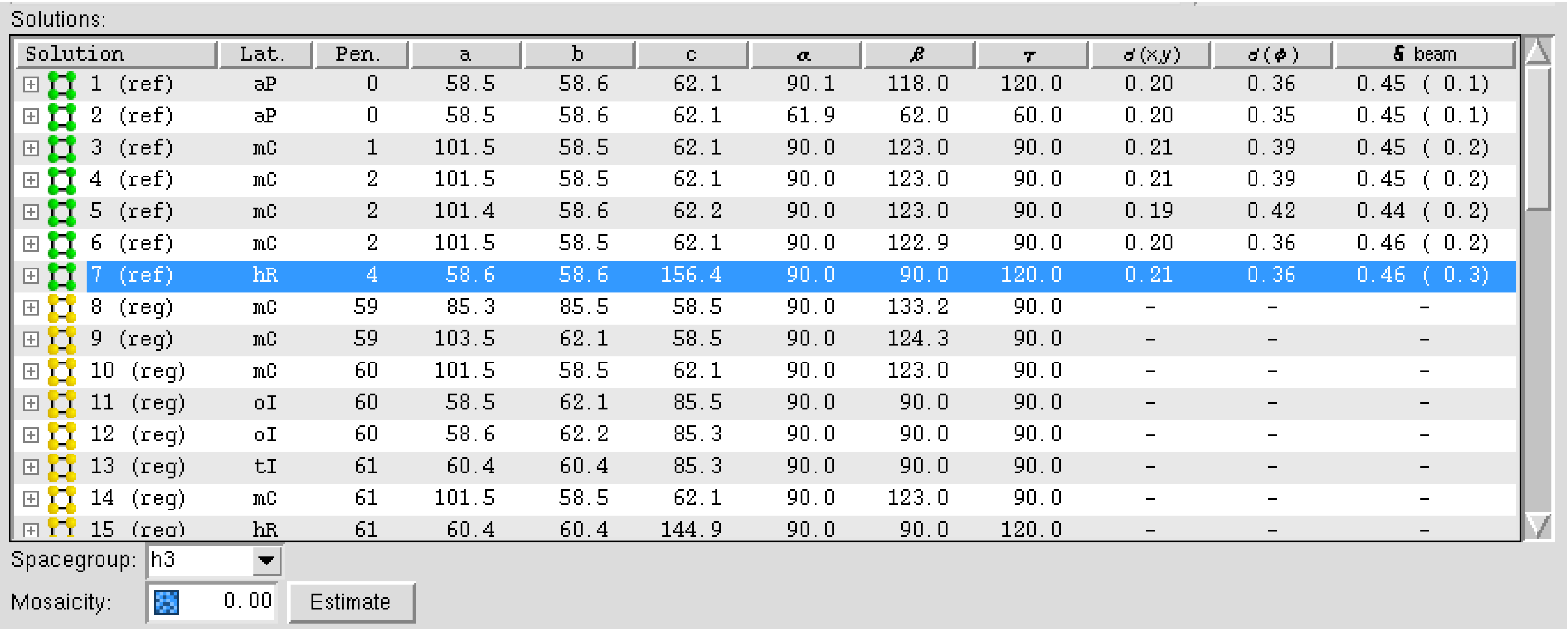
Spacegroup: P4

Mosaicity: 0.20

Viewing predictions:

The image viewer will be updated showing the predictions.

Make sure to inspect the images to see if predictions match observations!



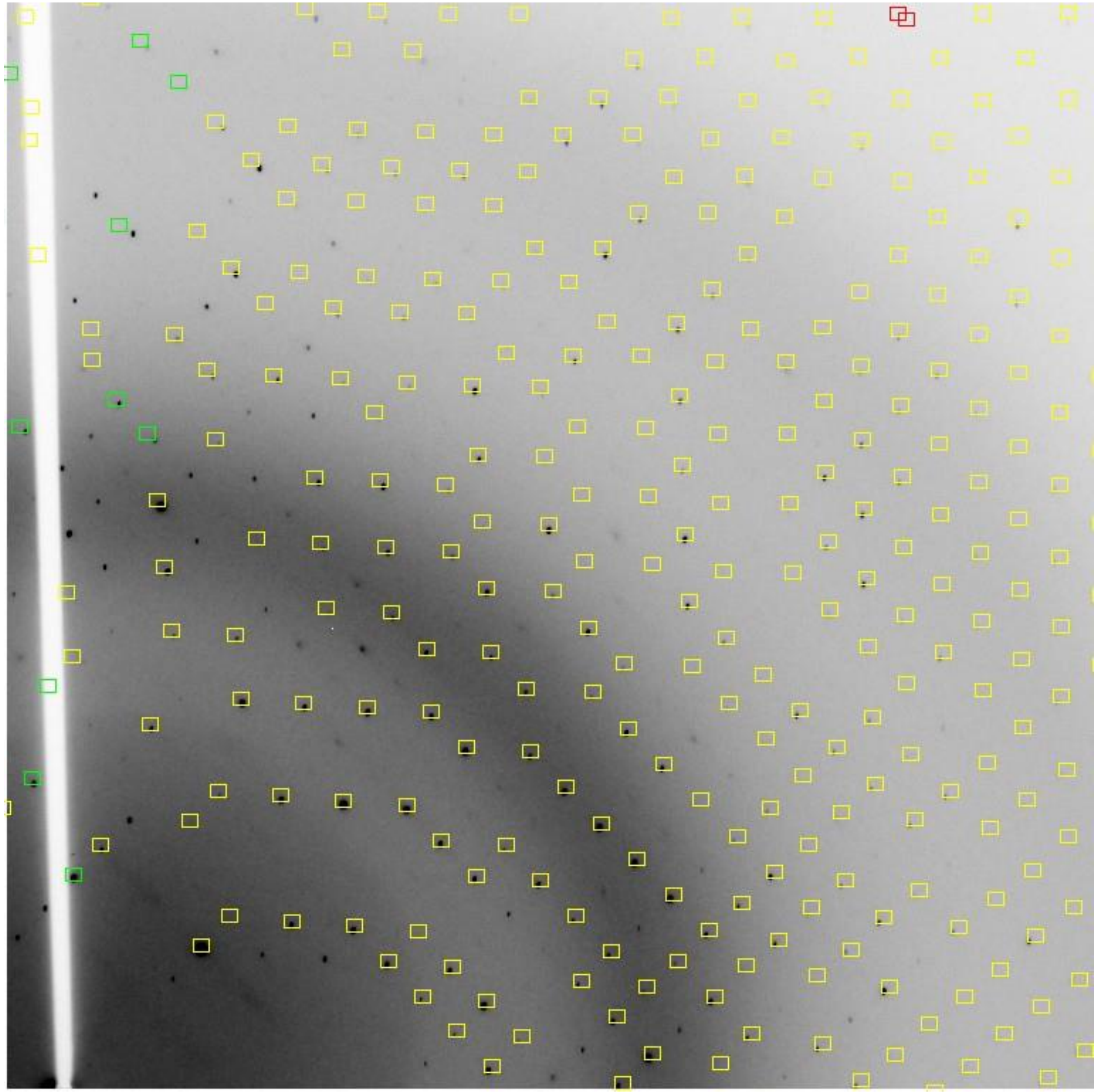
Solutions:

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

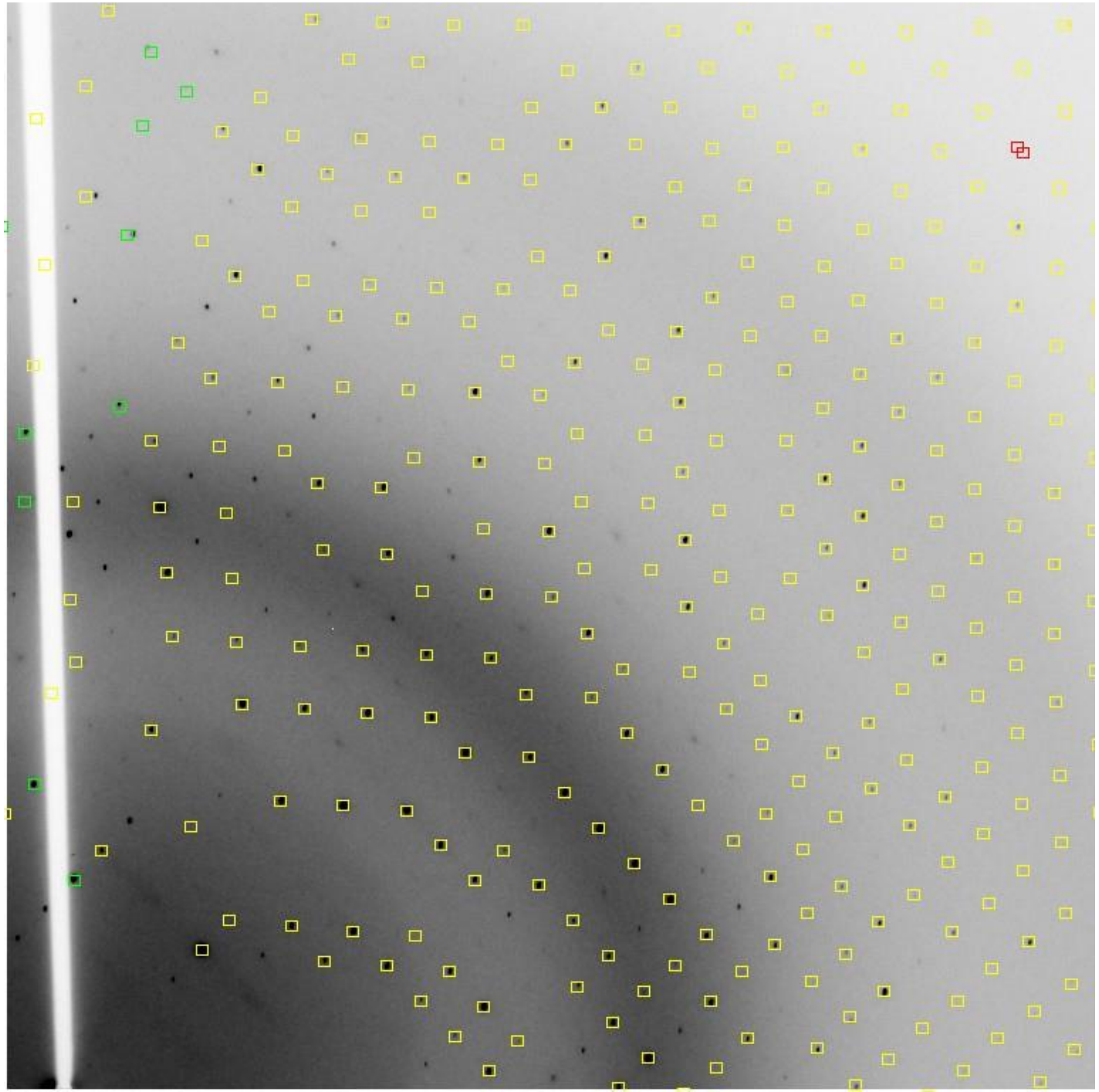
Mosaicity: 0.00

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 in cases of pseudosymmetry the rms error in spot positions (σ(x,y)) is also important. Reasonable solutions in labelled in green.



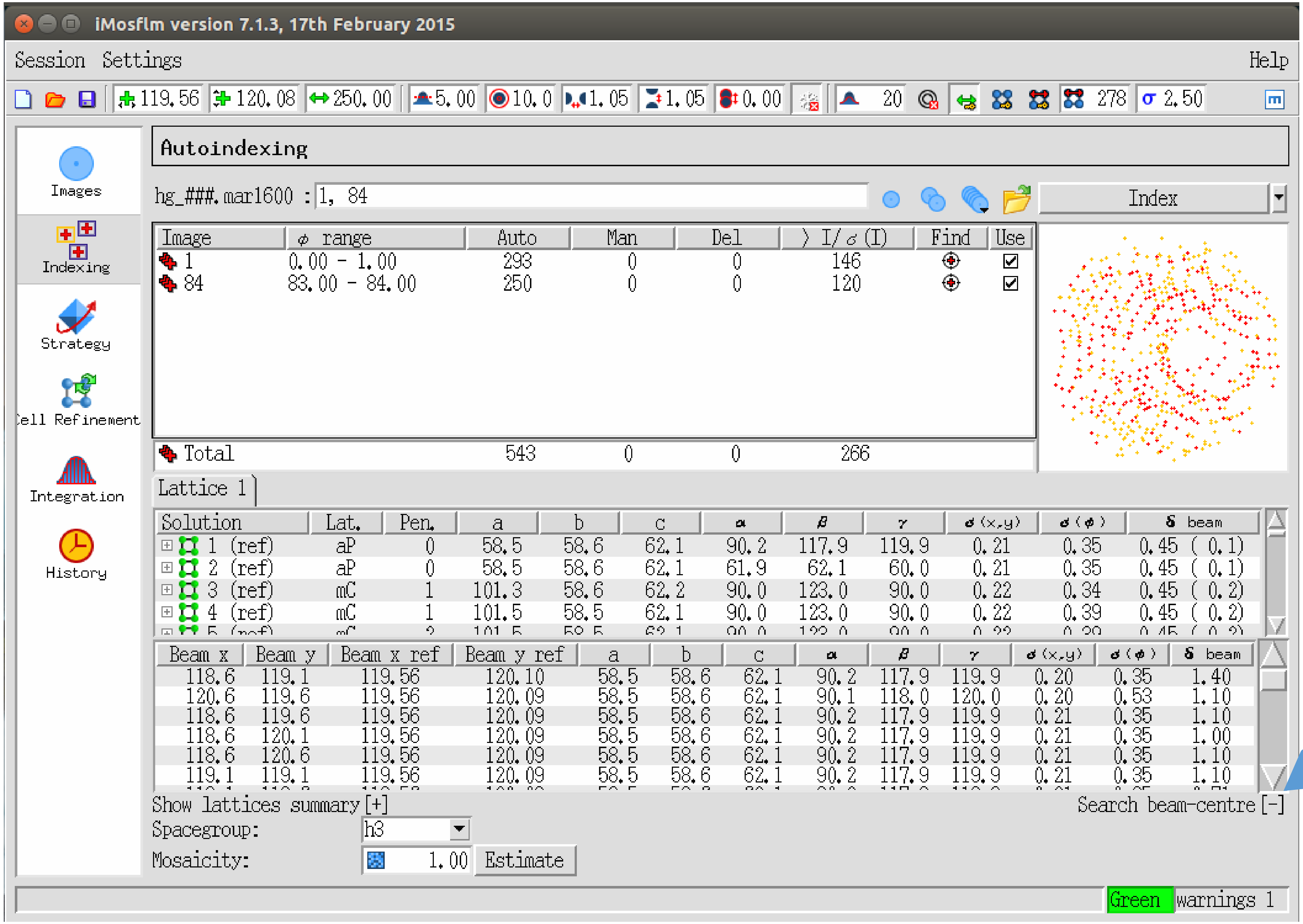
Check the predictions match the observations:

In this case the predictions are not nicely aligned with the spots



Check the predictions match the observations:

In this case the predictions are better aligned with the spots



Autoindexing

hg_###.mar1600 : 1, 84

Image	φ range	Auto	Man	Del	⟩ I/σ (I)	Find	Use
1	0.00 - 1.00	293	0	0	146	⌖	☑
84	83.00 - 84.00	250	0	0	120	⌖	☑
Total		543	0	0	266		

Lattice 1

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	90.2	117.9	119.9	0.21	0.35	0.45 (0.1)
⊞ 2 (ref)	aP	0	58.5	58.6	62.1	61.9	62.1	60.0	0.21	0.35	0.45 (0.1)
⊞ 3 (ref)	mC	1	101.3	58.6	62.2	90.0	123.0	90.0	0.22	0.34	0.45 (0.2)
⊞ 4 (ref)	mC	1	101.5	58.5	62.1	90.0	123.0	90.0	0.22	0.39	0.45 (0.2)

Beam x	Beam y	Beam x ref	Beam y ref	a	b	c	α	β	γ	σ (x,y)	σ (φ)	δ beam
118.6	119.1	119.56	120.10	58.5	58.6	62.1	90.2	117.9	119.9	0.20	0.35	1.40
120.6	119.6	119.56	120.09	58.5	58.6	62.1	90.1	118.0	120.0	0.20	0.53	1.10
118.6	119.6	119.56	120.09	58.5	58.6	62.1	90.2	117.9	119.9	0.21	0.35	1.10
118.6	120.1	119.56	120.09	58.5	58.6	62.1	90.2	117.9	119.9	0.21	0.35	1.00
118.6	120.6	119.56	120.09	58.5	58.6	62.1	90.2	117.9	119.9	0.21	0.35	1.10
119.1	119.1	119.56	120.09	58.5	58.6	62.1	90.2	117.9	119.9	0.21	0.35	1.10

Show lattices summary [+]

Search beam-centre [-]

Spacegroup: h3

Mosaicity: 1.00 Estimate

Green warnings 1

After indexing you can refine the direct beam position to get better predictions

Cell refinement

Data processing and scaling with Mosflm and Aimless

Purpose

To obtain a good estimate of cell parameters which can then be fixed during integration

Crystal	Detector	Beam
Cell dimensions	Position	Orientation
Orientation	Orientation	Divergence
Mosaic spread	Distortion	

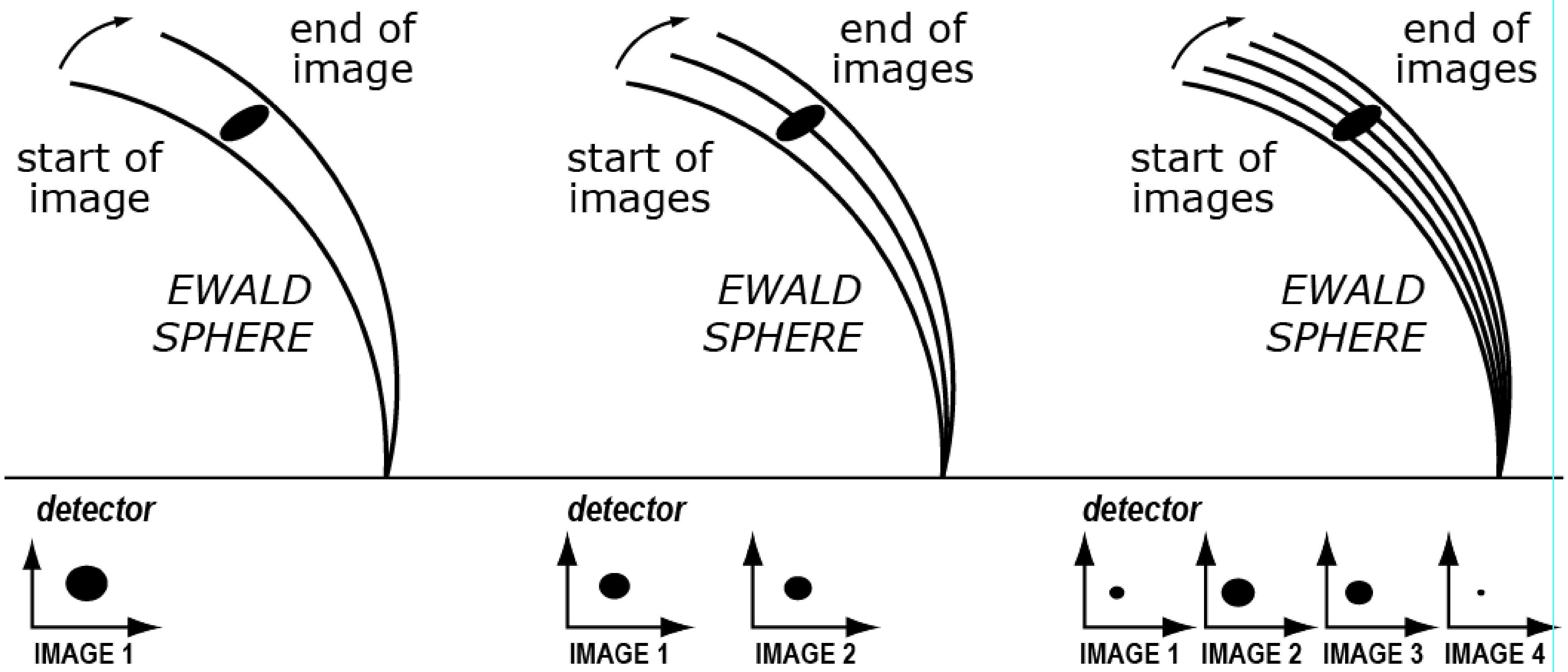
Refined parameters

Cell refinement in Mosflm

- Uses the relative intensities of the parts of partial reflections that are spread across multiple images
- Minimises the differences between the observed and calculated spots on the Ewald sphere.
- Provides very accurate cell parameters

- Requires:
  - A reasonable knowledge of intensities: mosflm selects a few frames to be integrated
  - A model for how we expect the intensities of the parts to vary between images – the “rocking curve”

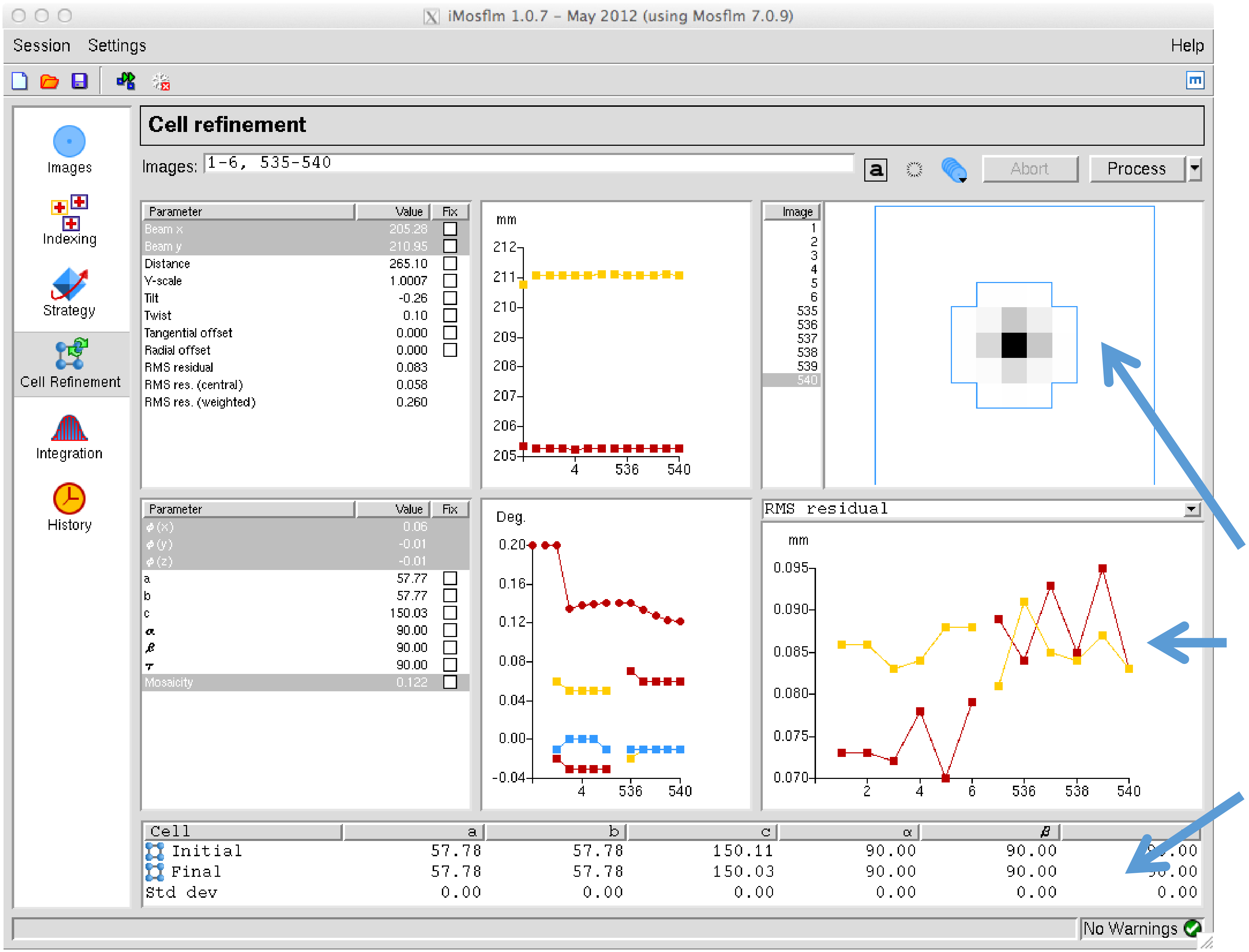
Fully and partially recorded reflections



A fully-recorded spot is entirely recorded on one image

Partials are recorded on two or more images

“Fine-sliced” data has spots sampled in 3-dimensions



Cell refinement

Images: 1-6, 535-540

Parameter	Value	Fix
Beam x	205.28	☐
Beam y	210.95	☐
Distance	265.10	☐
Y-scale	1.0007	☐
Tilt	-0.26	☐
Twist	0.10	☐
Tangential offset	0.000	☐
Radial offset	0.000	☐
RMS residual	0.083	
RMS res. (central)	0.058	
RMS res. (weighted)	0.260	

Image
1
2
3
4
5
6
535
536
537
538
539
540

Parameter	Value	Fix
ϕ(x)	0.06	
ϕ(y)	-0.01	
ϕ(z)	-0.01	
a	57.77	☐
b	57.77	☐
c	150.03	☐
α	90.00	☐
β	90.00	☐
γ	90.00	☐
Mosaicity	0.122	☐

Cell	a	b	c	α	β	
Initial	57.78	57.78	150.11	90.00	90.00	90.00
Final	57.78	57.78	150.03	90.00	90.00	90.00
Std dev	0.00	0.00	0.00	0.00	0.00	0.00

Cell refinement tab:

Shows average profile

RMS residuals should be small

Shows initial and final unit cell parameters

Integration/Refinement

Data processing and scaling with Mosflm and Aimless

Purpose

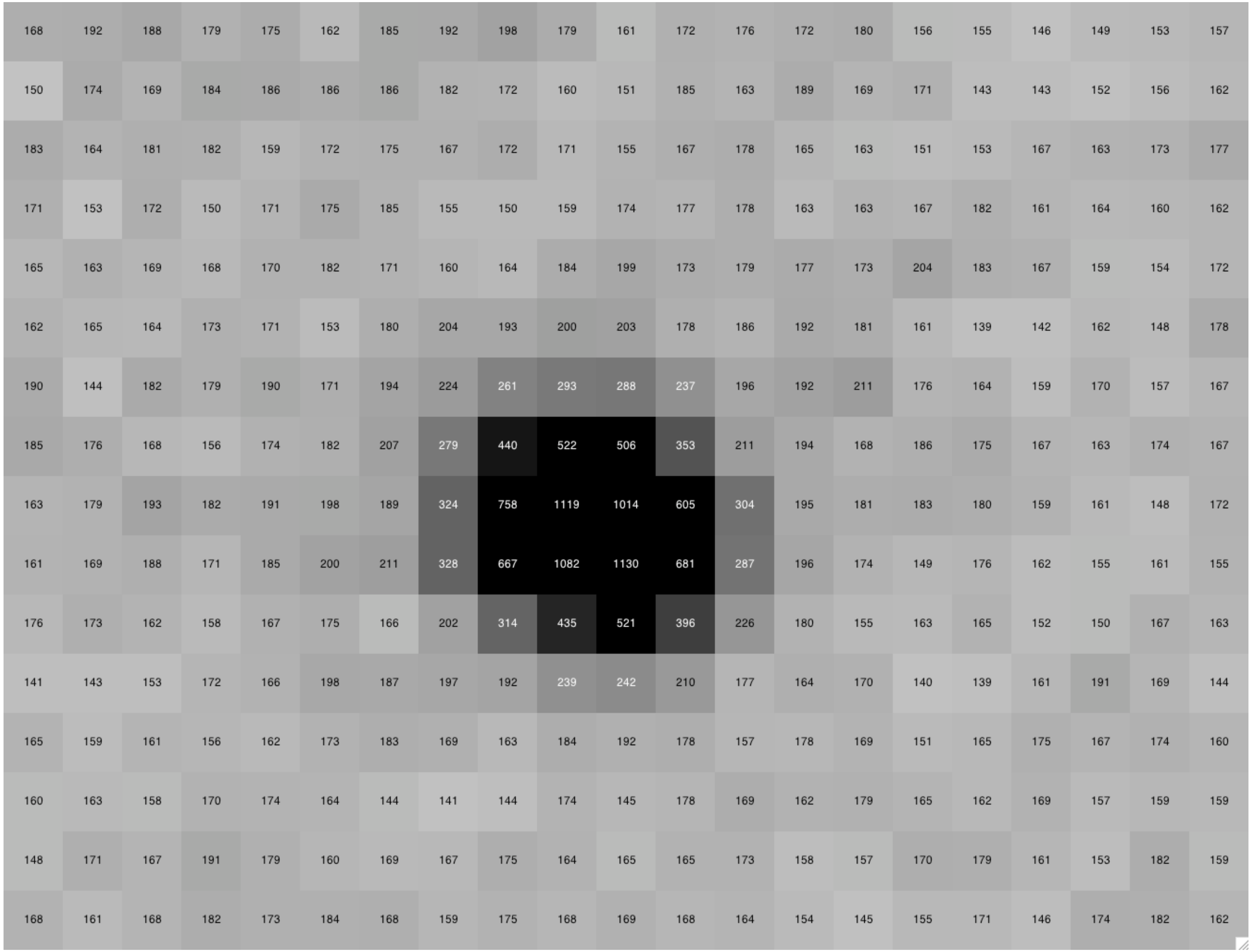
What we know:

- Where the spots should be (i.e. orientation, cell, geometry)
- What Miller indices belong to the spots
- How many images the spots are on (mosaic spread)

What we don’t know:

- Spot intensities

This is Integration



168	192	188	179	175	162	185	192	198	179	161	172	176	172	180	156	155	146	149	153	157
150	174	169	184	186	186	186	182	172	160	151	185	163	189	169	171	143	143	152	156	162
183	164	181	182	159	172	175	167	172	171	155	167	178	165	163	151	153	167	163	173	177
171	153	172	150	171	175	185	155	150	159	174	177	178	163	163	167	182	161	164	160	162
165	163	169	168	170	182	171	160	164	184	199	173	179	177	173	204	183	167	159	154	172
162	165	164	173	171	153	180	204	193	200	203	178	186	192	181	161	139	142	162	148	178
190	144	182	179	190	171	194	224	261	293	288	237	196	192	211	176	164	159	170	157	167
185	176	168	156	174	182	207	279	440	522	506	353	211	194	168	186	175	167	163	174	167
163	179	193	182	191	198	189	324	758	1119	1014	605	304	195	181	183	180	159	161	148	172
161	169	188	171	185	200	211	328	667	1082	1130	681	287	196	174	149	176	162	155	161	155
176	173	162	158	167	175	166	202	314	435	521	396	226	180	155	163	165	152	150	167	163
141	143	153	172	166	198	187	197	192	239	242	210	177	164	170	140	139	161	191	169	144
165	159	161	156	162	173	183	169	163	184	192	178	157	178	169	151	165	175	167	174	160
160	163	158	170	174	164	144	141	144	174	145	178	169	162	179	165	162	169	157	159	159
148	171	167	191	179	160	169	167	175	164	165	165	173	158	157	170	179	161	153	182	159
168	161	168	182	173	184	168	159	175	168	169	168	164	154	145	155	171	146	174	182	162

This is a diffraction spot. All we want to do is measure the number of counts in the peak minus the background counts. Simple!

Refinement and prediction

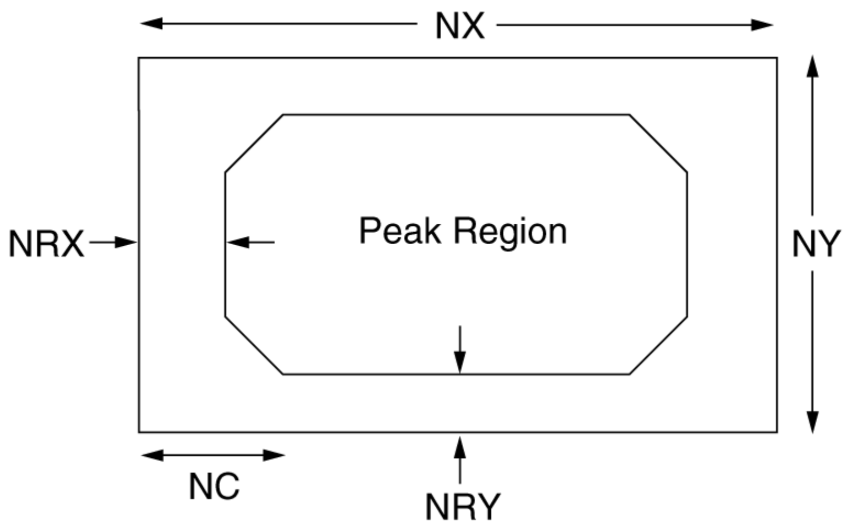
- Need to predict reflection positions accurately to avoid systematic errors in integration
- The detector parameters, crystal orientation and mosaic spread will be refined for every image during integration
- Cell parameters are kept fixed during integration

- Refined parameters such as the detector distance and YSCALE would not be expected to change during an experiment
- However, these parameters can compensate for errors in the cell parameters

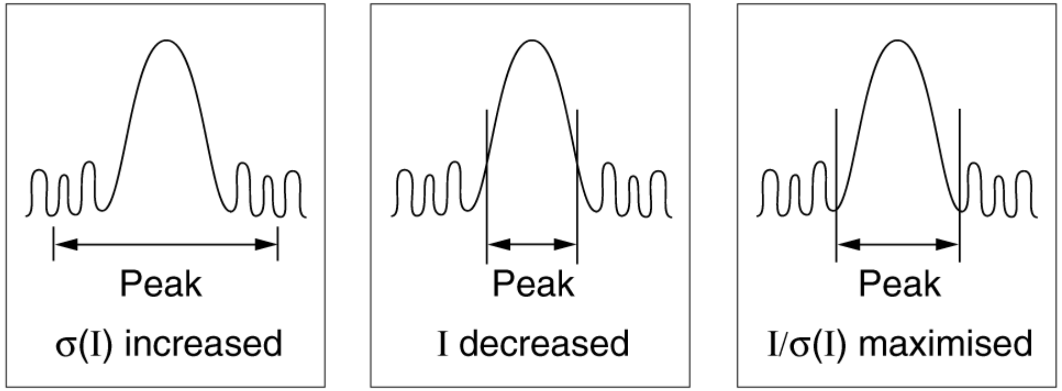
Background subtraction

- Create a peak/background mask by maximizing I/sigma
- Estimate background in peak region by fitting a plane to pixels in the background region and interpolating
- Errors in mask definition can give systematic errors in intensities

OPTIMISATION OF PEAK BACKGROUND MASK



Parameters NRX, NRY, NC are optimised by maximising $I/\sigma(I)$ (Lehmann & Larsen)



Integration in Mosflm

Two methods:

- Summation integration: *signal = total counts – background*
- Profile fitting: form standard profiles and fit to each reflection to estimate the intensity minimize $\sum \omega_i (x_i - K p_i)^2$

Profile fitting

Spots in the same region of the detector have similar profiles

Assume that the shape or profile 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.

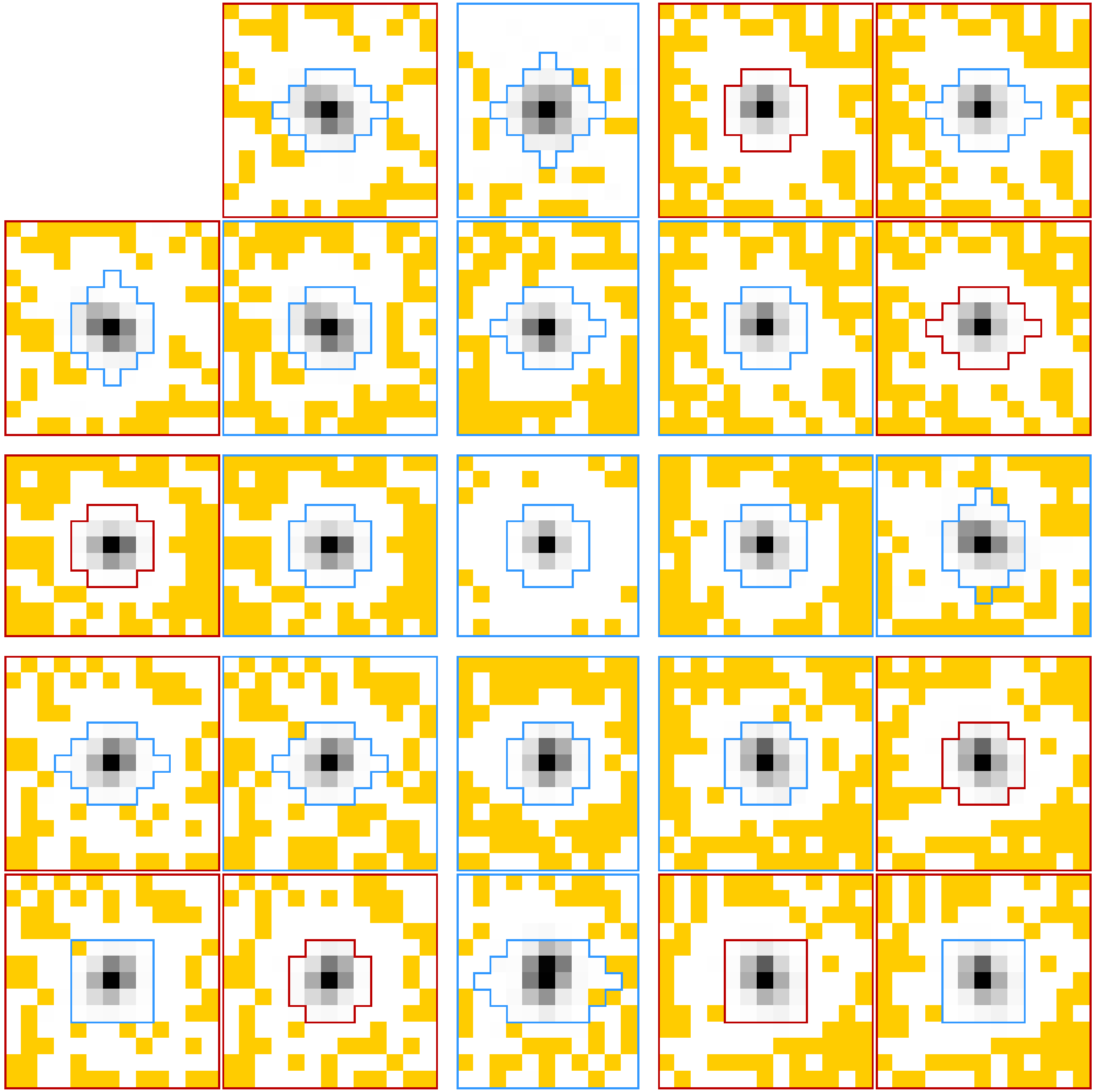
Profile fitting is more reliable for weak reflections than summation integration. Effectively “down-weight” peripheral peak pixels relative to the central pixels.

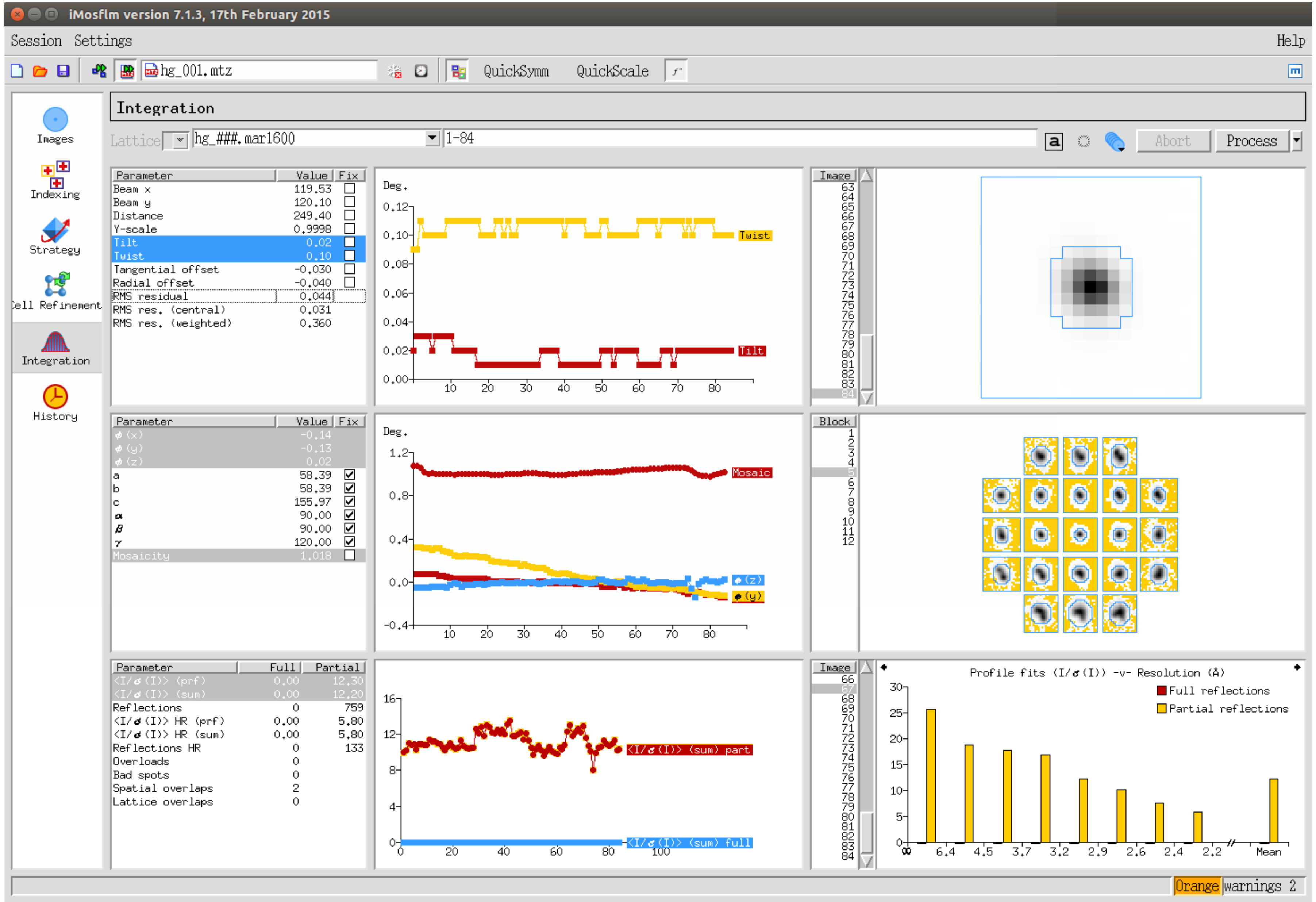
Standard profiles:

Determined empirically by computing the weighted mean of many spots.

The spot shape varies according to position on the detector.

This is done by forming a grid of standard profiles





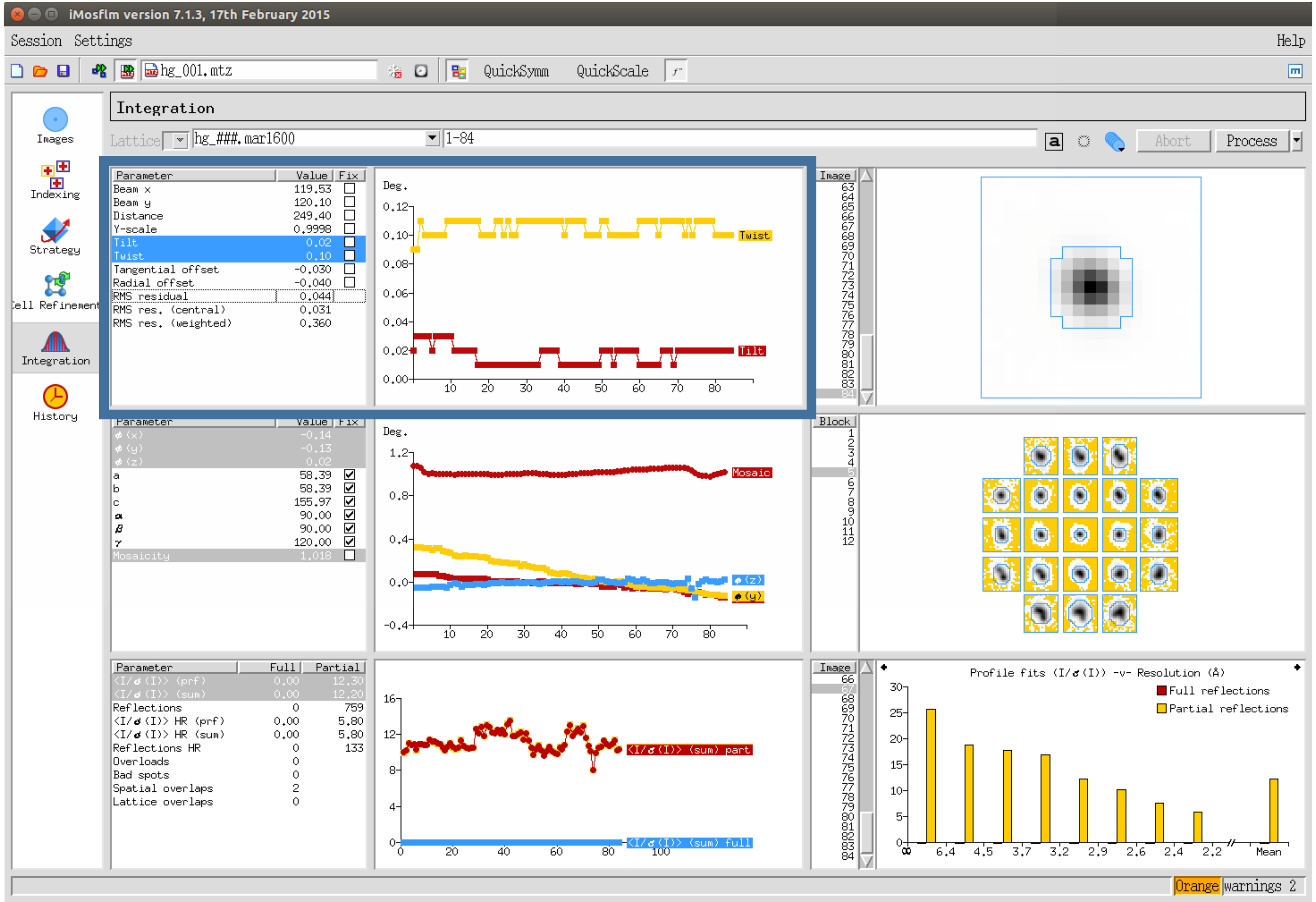
Parameter	Value	Fix
Beam x	119.53	☐
Beam y	120.10	☐
Distance	249.40	☐
Y-scale	0.9998	☐
Tilt	0.02	☐
Twist	0.10	☐
Tangential offset	-0.030	☐
Radial offset	-0.040	☐
RMS residual	0.044	
RMS res. (central)	0.031	
RMS res. (weighted)	0.360	

Parameter	Value	Fix
φ (x)	-0.14	
φ (y)	-0.13	
φ (z)	0.02	
a	58.39	☑
b	58.39	☑
c	155.97	☑
α	90.00	☑
β	90.00	☑
γ	120.00	☑
Mosaicity	1.018	☐

Parameter	Full	Partial
<I/σ(I)> (prf)	0.00	12.30
<I/σ(I)> (sum)	0.00	12.20
Reflections	0	759
<I/σ(I)> HR (prf)	0.00	5.80
<I/σ(I)> HR (sum)	0.00	5.80
Reflections HR	0	133
Overloads	0	
Bad spots	0	
Spatial overlaps	2	
Lattice overlaps	0	

Integration tab:

To perform the integration, click on the integration tab and then click “Process”



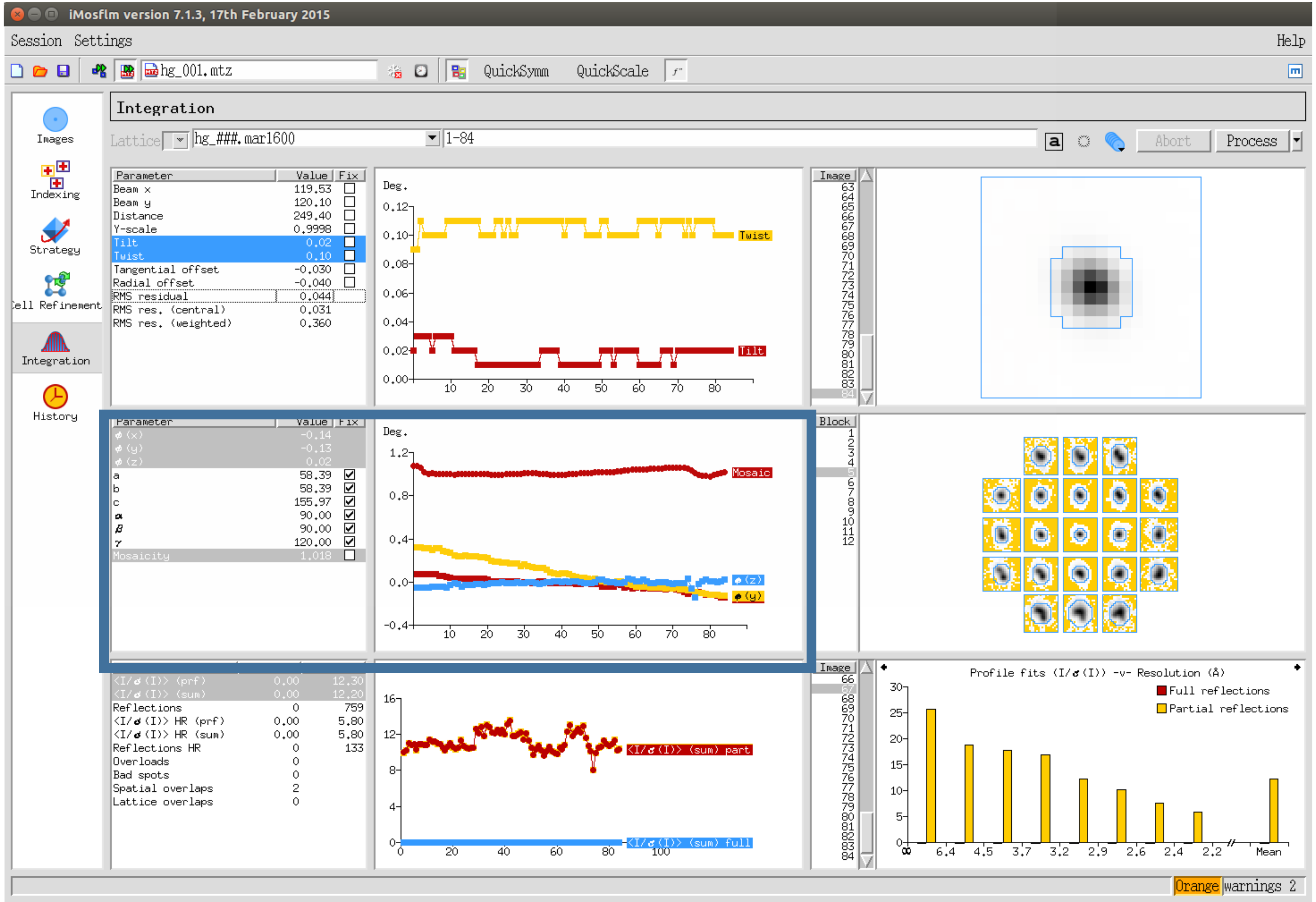
Parameter	Value	Fix
Beam x	119.53	☐
Beam y	120.10	☐
Distance	249.40	☐
Y-scale	0.9998	☐
Tilt	0.02	☐
Twist	0.10	☐
Tangential offset	-0.030	☐
Radial offset	-0.040	☐
RMS residual	0.044	
RMS res. (central)	0.031	
RMS res. (weighted)	0.360	

Parameter	Value	Fix
φ(x)	-0.14	
φ(y)	-0.13	
φ(z)	0.02	
a	58.39	☑
b	58.39	☑
c	155.97	☑
α	90.00	☑
β	90.00	☑
γ	120.00	☑
Mosaicity	1.018	☐

Parameter	Full	Partial
<I/σ(I)> (prf)	0.00	12.30
<I/σ(I)> (sum)	0.00	12.20
Reflections	0	759
<I/σ(I)> HR (prf)	0.00	5.80
<I/σ(I)> HR (sum)	0.00	5.80
Reflections HR	0	133
Overloads	0	
Bad spots	0	
Spatial overlaps	2	
Lattice overlaps	0	

Refined detector parameters:

Check tilt and twist are stable; might need to fix. Yscale should also be ~1.0



hg_001.mtz

Integration

hg_###.mar1600

1-84

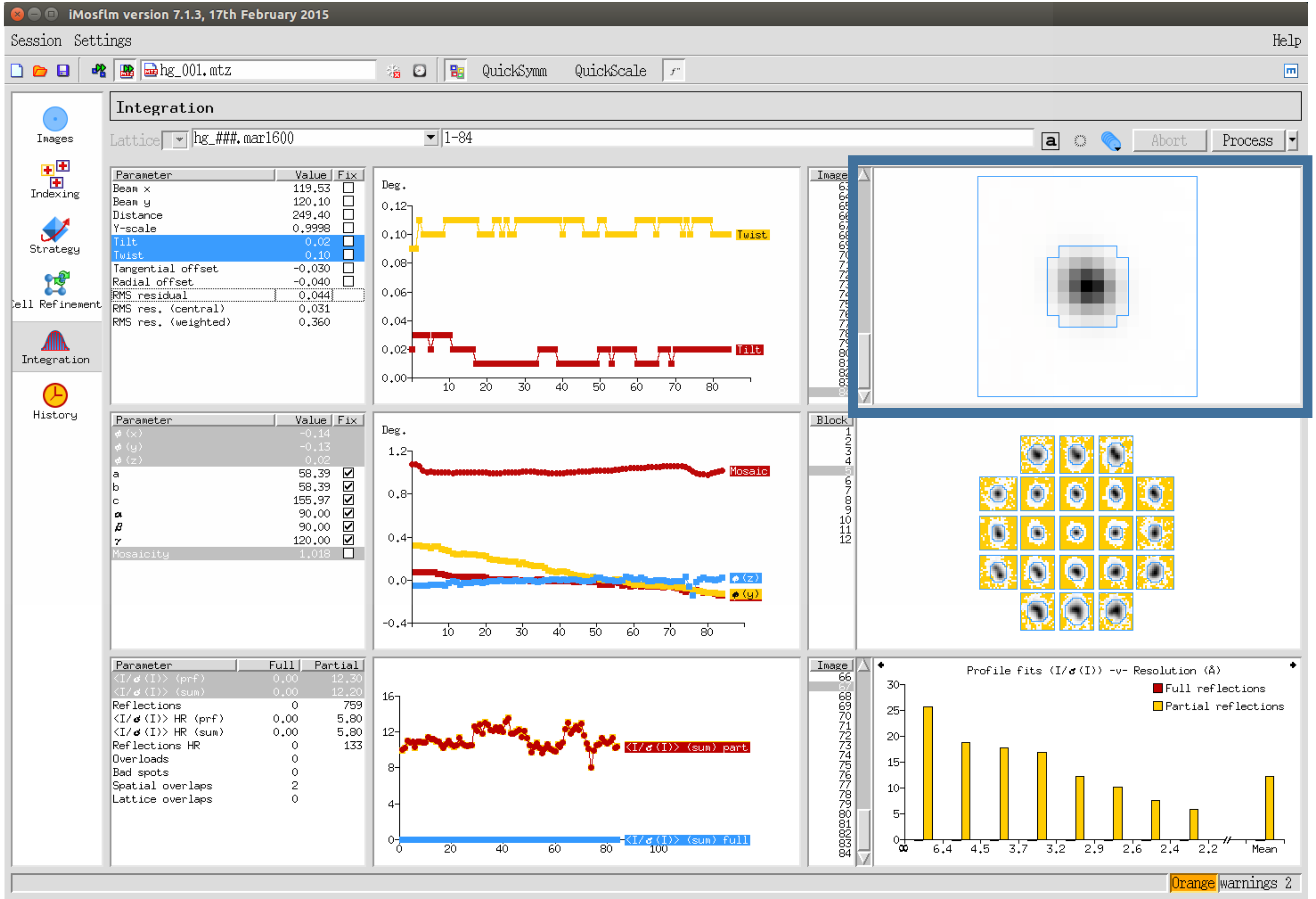
Parameter	Value	Fix
Beam x	119.53	☐
Beam y	120.10	☐
Distance	249.40	☐
Y-scale	0.9998	☐
Tilt	0.02	☐
Twist	0.10	☐
Tangential offset	-0.030	☐
Radial offset	-0.040	☐
RMS residual	0.044	
RMS res. (central)	0.031	
RMS res. (weighted)	0.360	

Parameter	Value	Fix
ɸ (x)	-0.14	
ɸ (y)	-0.13	
ɸ (z)	0.02	
a	58.39	☑
b	58.39	☑
c	155.97	☑
α	90.00	☑
β	90.00	☑
γ	120.00	☑
Mosaicity	1.018	☐

<I/σ(I)> (prf)	0.00	12.30
<I/σ(I)> (sum)	0.00	12.20
Reflections	0	759
<I/σ(I)> HR (prf)	0.00	5.80
<I/σ(I)> HR (sum)	0.00	5.80
Reflections HR	0	133
Overloads	0	
Bad spots	0	
Spatial overlaps	2	
Lattice overlaps	0	

Refined crystal parameters:

Ensure that cell parameters are fixed

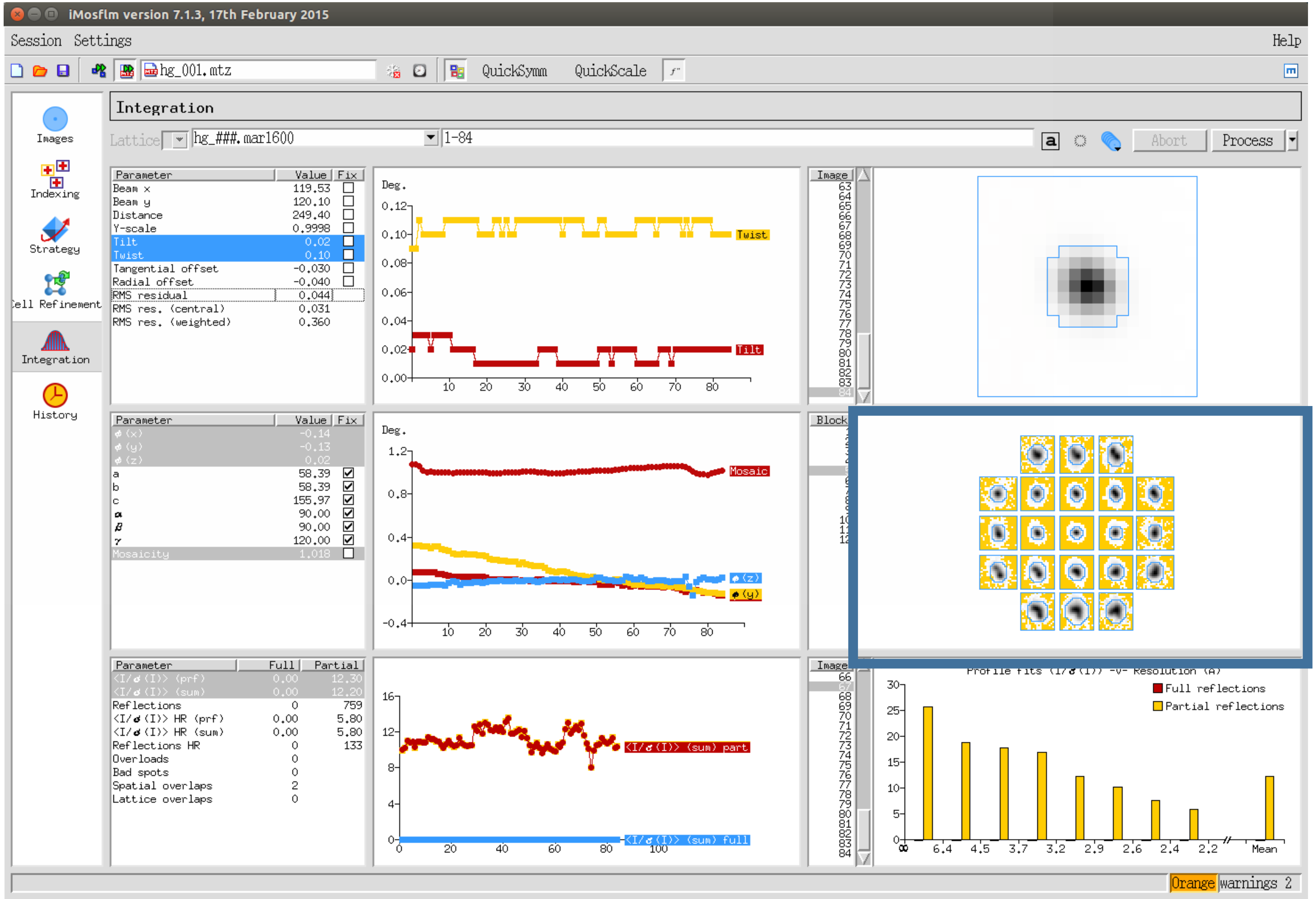


Parameter	Value	Fix
Beam x	119.53	☐
Beam y	120.10	☐
Distance	249.40	☐
Y-scale	0.9998	☐
Tilt	0.02	☐
Twist	0.10	☐
Tangential offset	-0.030	☐
Radial offset	-0.040	☐
RMS residual	0.044	
RMS res. (central)	0.031	
RMS res. (weighted)	0.360	

Parameter	Value	Fix
φ(x)	-0.14	
φ(y)	-0.13	
φ(z)	0.02	
a	58.39	☑
b	58.39	☑
c	155.97	☑
α	90.00	☑
β	90.00	☑
γ	120.00	☑
Mosaicity	1.018	☐

Parameter	Full	Partial
<I/σ(I)> (prf)	0.00	12.30
<I/σ(I)> (sum)	0.00	12.20
Reflections	0	759
<I/σ(I)> HR (prf)	0.00	5.80
<I/σ(I)> HR (sum)	0.00	5.80
Reflections HR	0	133
Overloads	0	
Bad spots	0	
Spatial overlaps	2	
Lattice overlaps	0	

Average reflection profile

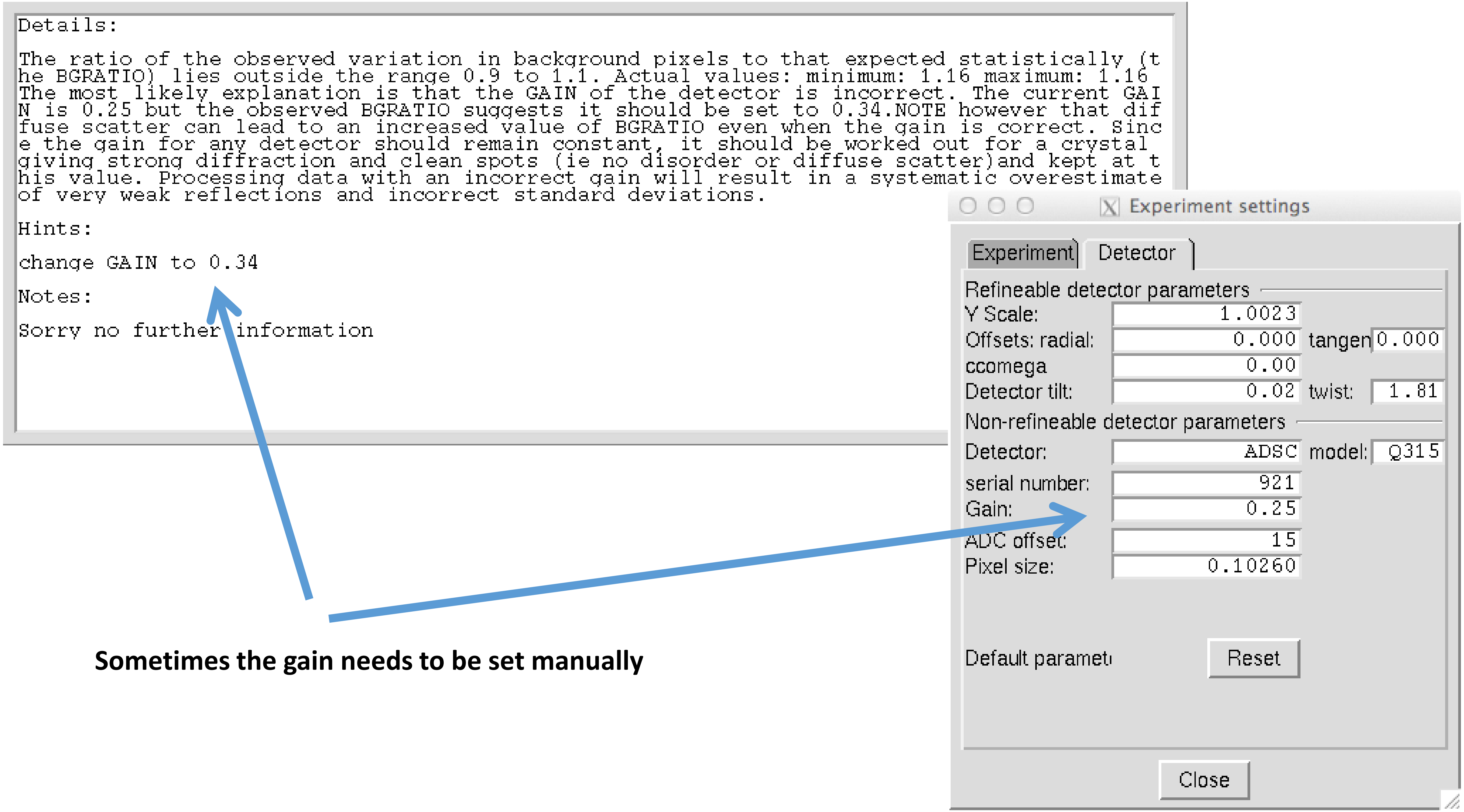


Parameter	Value	Fix
Beam x	119.53	☐
Beam y	120.10	☐
Distance	249.40	☐
Y-scale	0.9998	☐
Tilt	0.02	☐
Twist	0.10	☐
Tangential offset	-0.030	☐
Radial offset	-0.040	☐
RMS residual	0.044	
RMS res. (central)	0.031	
RMS res. (weighted)	0.360	

Parameter	Value	Fix
φ (x)	-0.14	
φ (y)	-0.13	
φ (z)	0.02	
a	58.39	☑
b	58.39	☑
c	155.97	☑
α	90.00	☑
β	90.00	☑
γ	120.00	☑
Mosaicity	1.018	☐

Parameter	Full	Partial
<I/σ(I)> (prf)	0.00	12.30
<I/σ(I)> (sum)	0.00	12.20
Reflections	0	759
<I/σ(I)> HR (prf)	0.00	5.80
<I/σ(I)> HR (sum)	0.00	5.80
Reflections HR	0	133
Overloads	0	
Bad spots	0	
Spatial overlaps	2	
Lattice overlaps	0	

Standard profiles



```
Details:

The ratio of the observed variation in background pixels to that expected statistically (t
he BGRATIO) lies outside the range 0.9 to 1.1. Actual values: minimum: 1.16 maximum: 1.16
The most likely explanation is that the GAIN of the detector is incorrect. The current GAI
N is 0.25 but the observed BGRATIO suggests it should be set to 0.34.NOTE however that dif
fuse scatter can lead to an increased value of BGRATIO even when the gain is correct. Sinc
e the gain for any detector should remain constant, it should be worked out for a crystal
giving strong diffraction and clean spots (ie no disorder or diffuse scatter)and kept at t
his value. Processing data with an incorrect gain will result in a systematic overestimate
of very weak reflections and incorrect standard deviations.

Hints:

change GAIN to 0.34

Notes:

Sorry no further information
```

Experiment settings

Refineable detector parameters

Y Scale: 1.0023

Offsets: radial: 0.000 tangen 0.000

ccomega 0.00

Detector tilt: 0.02 twist: 1.81

Non-refineable detector parameters

Detector: ADSC model: Q315

serial number: 921

Gain: 0.25

ADC offset: 15

Pixel size: 0.10260

Sometimes the gain needs to be set manually

Quick scaling

Data processing and scaling with Mosflm and Aimless

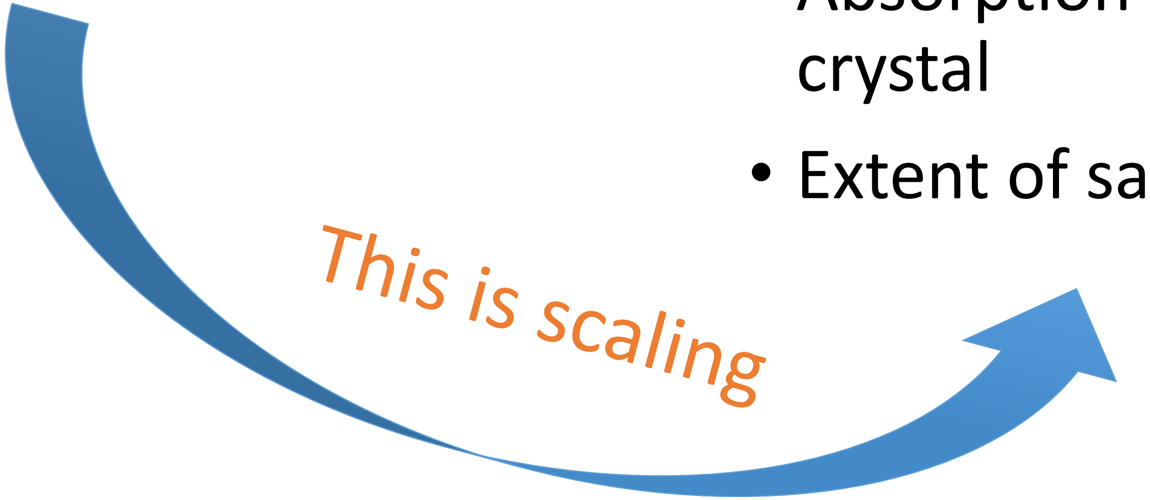
Purpose

Things we know:

- I, sig(I), corrected for geometric effects
- Lots of observations
- Symmetry

Things we don’t know:

- $|F|^2$
- Beam intensity
- Illuminated volume
- Absorption path through crystal
- Extent of sample decay



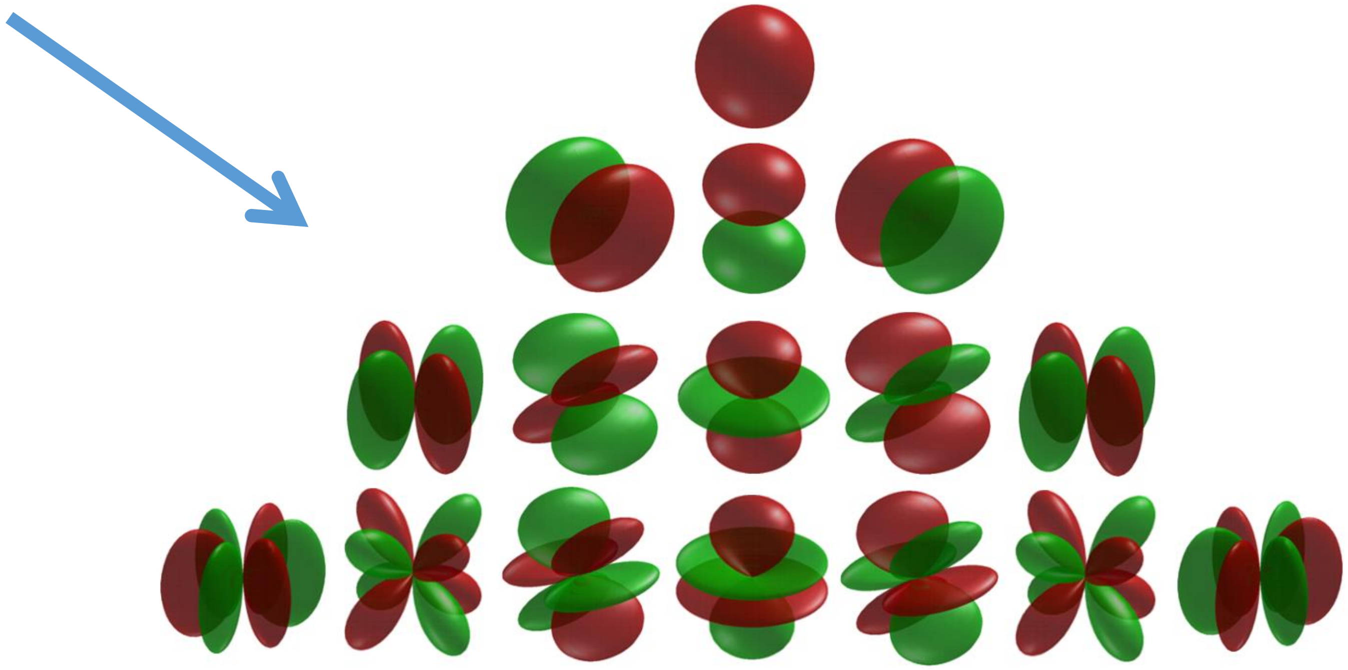
This is scaling

Scaling

- Corrections for some of the things we don’t know can be determined experimentally
- In most cases however empirical corrections are determined
- Have a model for: overall scale (beam intensity + illuminated volume) sample decay and absorption
- Refine model against data, to minimise differences between symmetry related intensities

Scaling models

- Time or frame # dependent – overall scale
- Time and resolution dependent – decay
- Direction dependent – absorption – for example as spherical harmonics
- All depends on multiplicity



Objective of scaling

- To model all of the unknown contributions to the measured intensity
- To recover $I=k|F|^2$ for each observation
- Achieved by minimizing the differences between observations – internally consistent not necessarily correct!
- Final result of scaling is average $I=k|F|^2$ for each unique Miller index
- May want to keep I+ and I- separate

Merging statistics

- Level of agreement between unmerged I is usually guide to data quality
- For strong measurements this is dominated by experimental effects e.g. radiation damage
- For weak reflections dominated by signal-to-noise
- Best merging statistics give a guide to how useful the averaged measurements are

Programs

Pointless

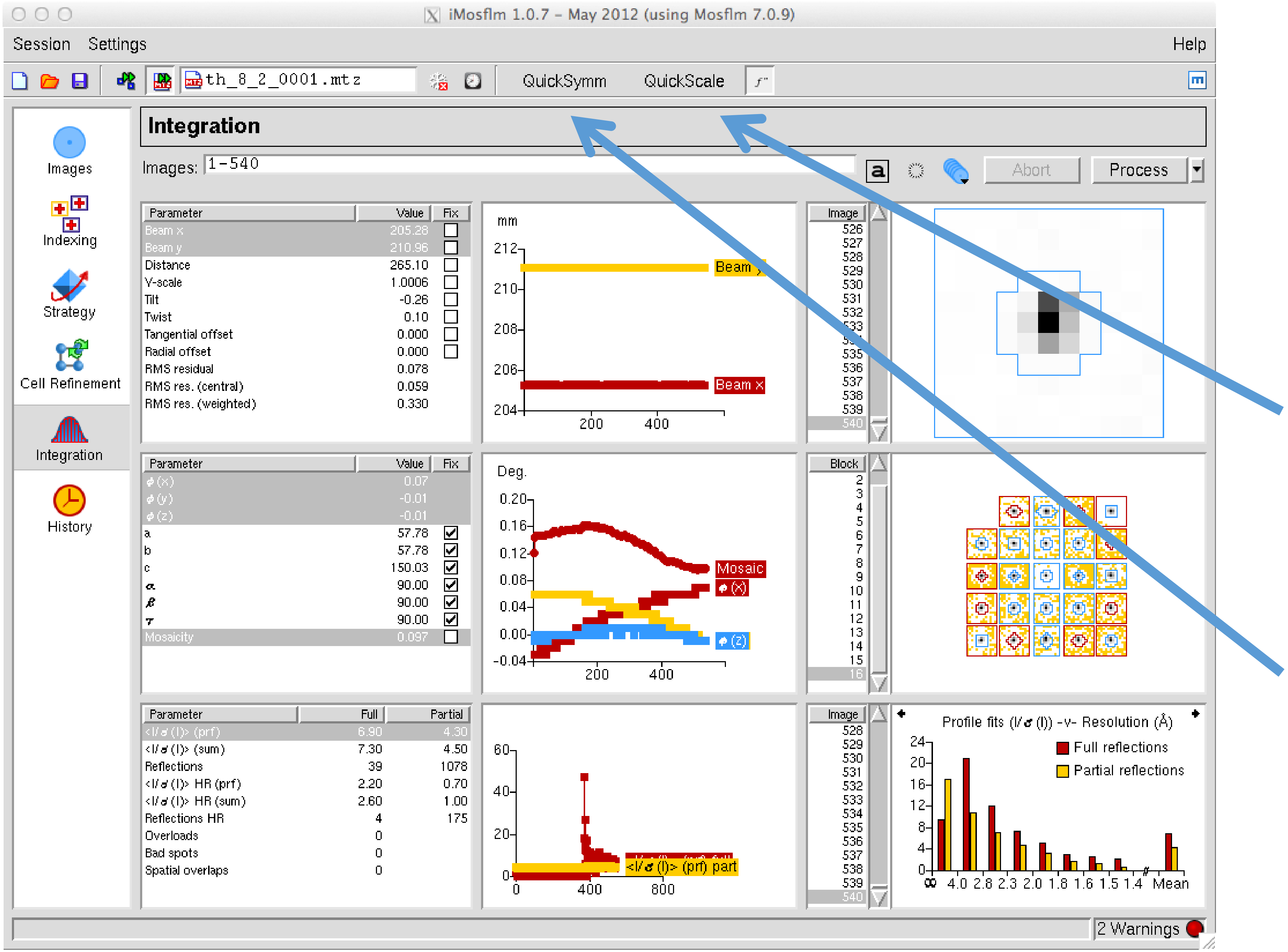
- Determines likely point group
- Corrects space group *if* sufficient information
- Sorts reflections
- Detects screw axes & glide planes
- Re-indexes multiple datasets to a common setting

Aimless

- Merges partial reflections together
- Puts data onto a common scale
- Merges each set of symmetry equivalent reflections into a single observation

CTruncate

- Analyses scaled data according to an expected physical model
- Gives statistics on intensity distribution - *e.g.*
    - Wilson statistics
    - twinning analysis
- Outputs |F| values



Integration

Images: 1-540

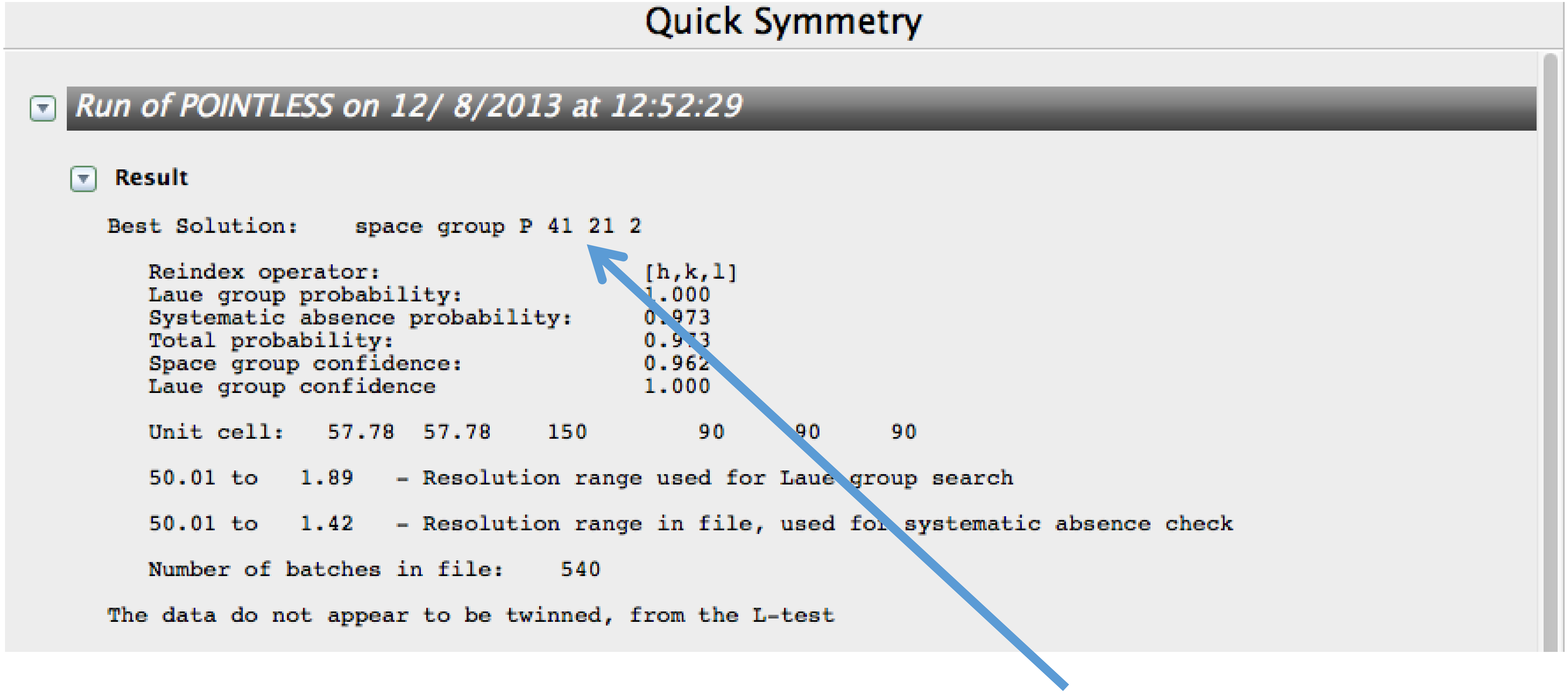
Parameter	Value	Fix
Beam x	205.28	☐
Beam y	210.96	☐
Distance	265.10	☐
Y-scale	1.0006	☐
Tilt	-0.26	☐
Twist	0.10	☐
Tangential offset	0.000	☐
Radial offset	0.000	☐
RMS residual	0.078	
RMS res. (central)	0.059	
RMS res. (weighted)	0.330	

Parameter	Value	Fix
φ (x)	0.07	
φ (y)	-0.01	
φ (z)	-0.01	
a	57.78	☑
b	57.78	☑
c	150.03	☑
α	90.00	☑
β	90.00	☑
γ	90.00	☑
Mosaicity	0.097	☐

Parameter	Full	Partial
<I/σ(I)> (prf)	6.90	4.30
<I/σ(I)> (sum)	7.30	4.50
Reflections	39	1078
<I/σ(I)> HR (prf)	2.20	0.70
<I/σ(I)> HR (sum)	2.60	1.00
Reflections HR	4	175
Overloads	0	
Bad spots	0	
Spatial overlaps	0	

Run Pointless, Aimless and cTruncate to scale the data

Run Pointless to check pointgroup



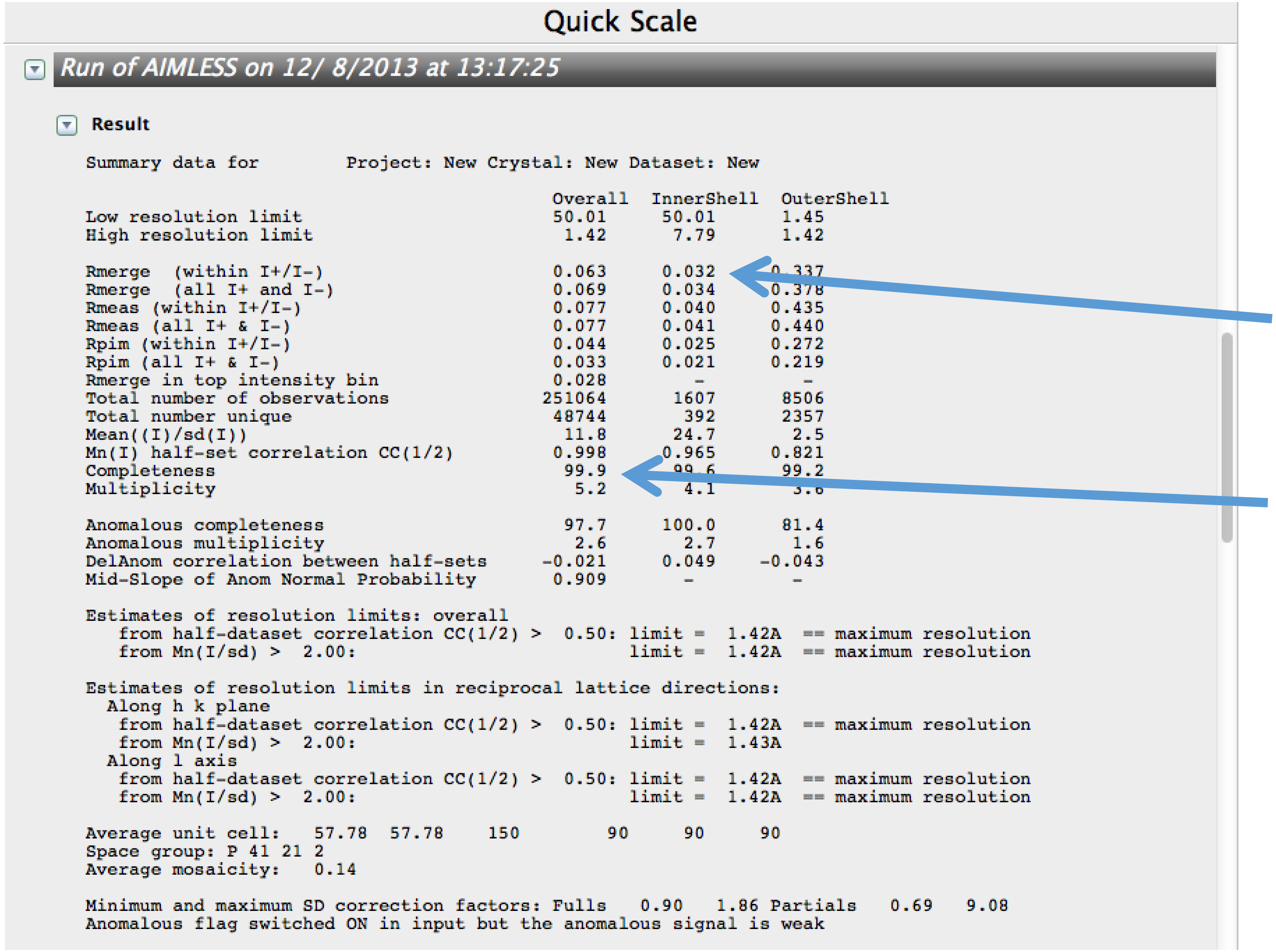
Quick Symmetry

Run of POINTLESS on 12/ 8/2013 at 12:52:29

Result

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

   Reindex operator:                [h,k,l]
   Laue group probability:          1.000
   Systematic absence probability:  0.973
   Total probability:               0.973
   Space group confidence:          0.962
   Laue group confidence            1.000

   Unit cell:   57.78  57.78   150       90      90      90

   50.01 to   1.89   - Resolution range used for Laue group search

   50.01 to   1.42   - Resolution range in file, used for systematic absence check

   Number of batches in file:    540

The data do not appear to be twinned, from the L-test
```

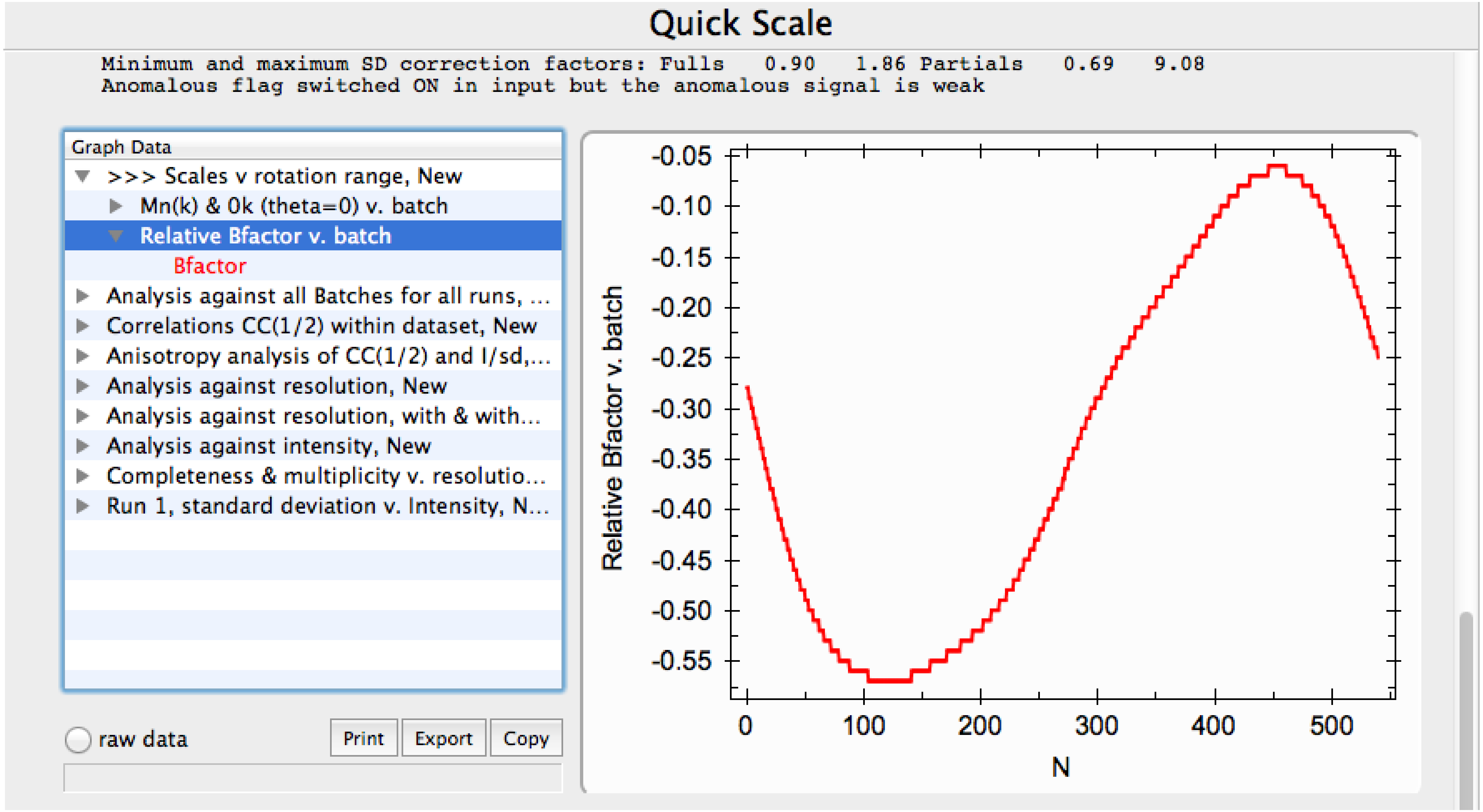
Check pointgroup is consistent with lattice used for integration



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

                                           Overall  InnerShell  OuterShell
Low resolution limit                       50.01     50.01       1.45
High resolution limit                       1.42      7.79       1.42

Rmerge  (within I+/I-)                     0.063     0.032     0.337
Rmerge  (all I+ and I-)                    0.069     0.034     0.378
Rmeas (within I+/I-)                       0.077     0.040     0.435
Rmeas (all I+ & I-)                        0.077     0.041     0.440
Rpim (within I+/I-)                        0.044     0.025     0.272
Rpim (all I+ & I-)                         0.033     0.021     0.219
Rmerge in top intensity bin                0.028        -         -
Total number of observations              251064      1607      8506
Total number unique                        48744       392      2357
Mean((I)/sd(I))                             11.8      24.7       2.5
Mn(I) half-set correlation CC(1/2)         0.998     0.965     0.821
Completeness                                99.9      99.6      99.2
Multiplicity                                 5.2       4.1       3.6

Anomalous completeness                      97.7     100.0      81.4
Anomalous multiplicity                       2.6       2.7       1.6
DelAnom correlation between half-sets     -0.021     0.049    -0.043
Mid-Slope of Anom Normal Probability       0.909        -         -

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.50: limit =   1.42A  == maximum resolution
   from Mn(I/sd) >  2.00:                         limit =   1.42A  == maximum resolution

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.50: limit =   1.42A  == maximum resolution
   from Mn(I/sd) >  2.00:                         limit =   1.43A
  Along l axis
   from half-dataset correlation CC(1/2) >  0.50: limit =   1.42A  == maximum resolution
   from Mn(I/sd) >  2.00:                         limit =   1.42A  == maximum resolution

Average unit cell:   57.78   57.78    150        90      90      90
Space group: P 41 21 2
Average mosaicity:   0.14

Minimum and maximum SD correction factors: Fulls   0.90   1.86 Partials   0.69   9.08
Anomalous flag switched ON in input but the anomalous signal is weak
```

Ideally this should be a small number

… and this should be close to 100%



Quick Scale

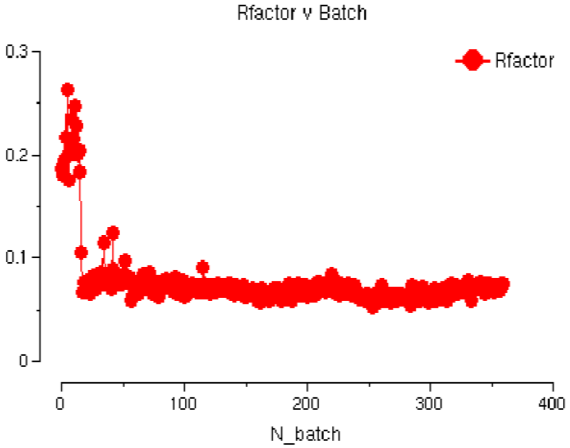
```
Minimum and maximum SD correction factors: Fulls   0.90   1.86 Partials   0.69   9.08
Anomalous flag switched ON in input but the anomalous signal is weak
```

Judging data quality

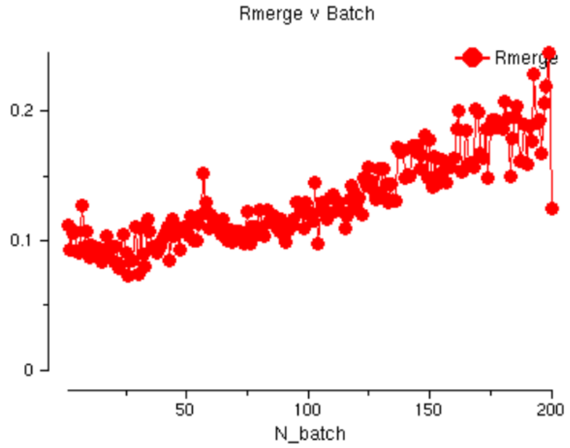
- Are there bad batches?
- Was the radiation damage such that you should exclude the later parts?
- What is the real resolution? Should you cut the high-resolution data?
- Is there any apparent anomalous signal?
- Is the outlier detection working well?
- Are the data twinned?
- What is the overall quality of the dataset?

Rmerge: finding bad batches

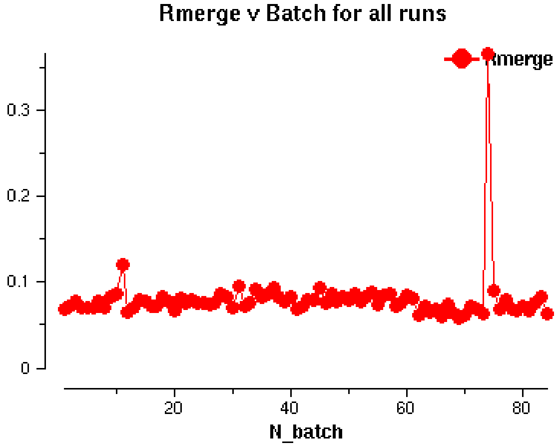
Horribly wrong at beginning



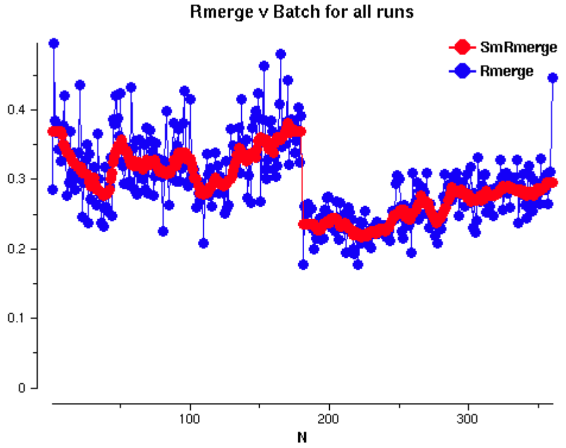
Steady decline in quality



One bad batch



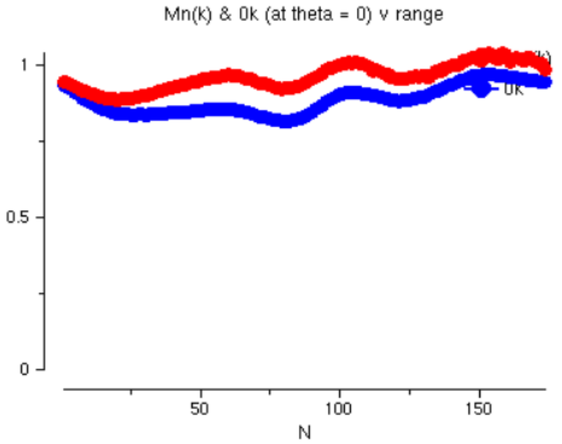
Batches for 2 crystals



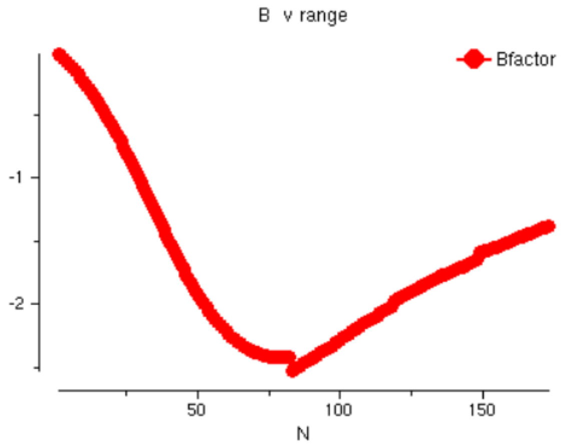
Would like to have relatively stable rmerge across all batches

Scales and B-factors: radiation damage

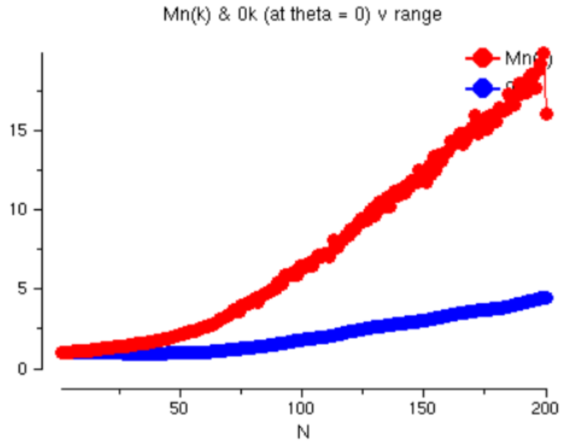
Good: scales uniform



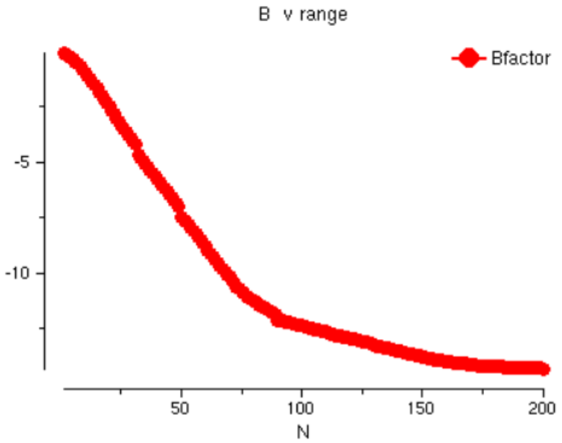
Good: small B-factors



Bad: scales increase sharply



Bad: B-factors large and negative



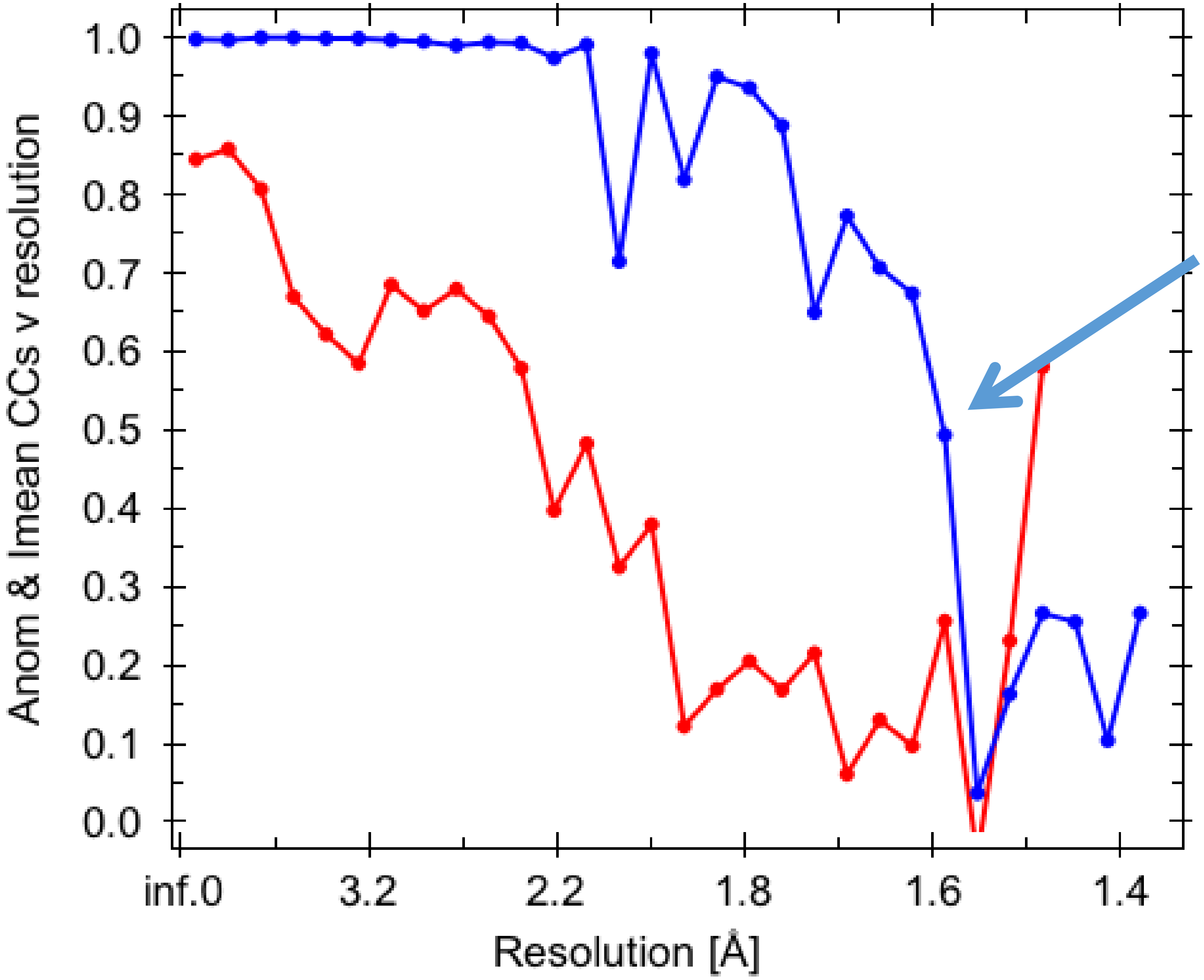
Ideally have constant scaling factor of 1; except if crystals have an irregular shape.
Drop in B factor below -10 indicates radiation damage

Number of rejections per image

```
N  Run.Rot MidPhi   Batch Bfactor  Mn(k)   0k      Number NumReject
1  1.1     -49.50   1     -0.694   1.0651  0.9940  1703   0
2  1.2     -48.50   2     -0.688   1.0622  0.9905  2193   0
3  1.3     -47.50   3     -0.677   1.0564  0.9851  2219   0
4  1.4     -46.50   4     -0.668   1.0453  0.9774  2202   0
5  1.5     -45.50   5     -0.656   1.0339  0.9671  2198   0
6  1.6     -44.50   6     -0.641   1.0180  0.9542  2217   1
7  1.7     -43.50   7     -0.629   1.0017  0.9395  2208   0
8  1.8     -42.50   8     -0.614   0.9811  0.9185  2217   0
```

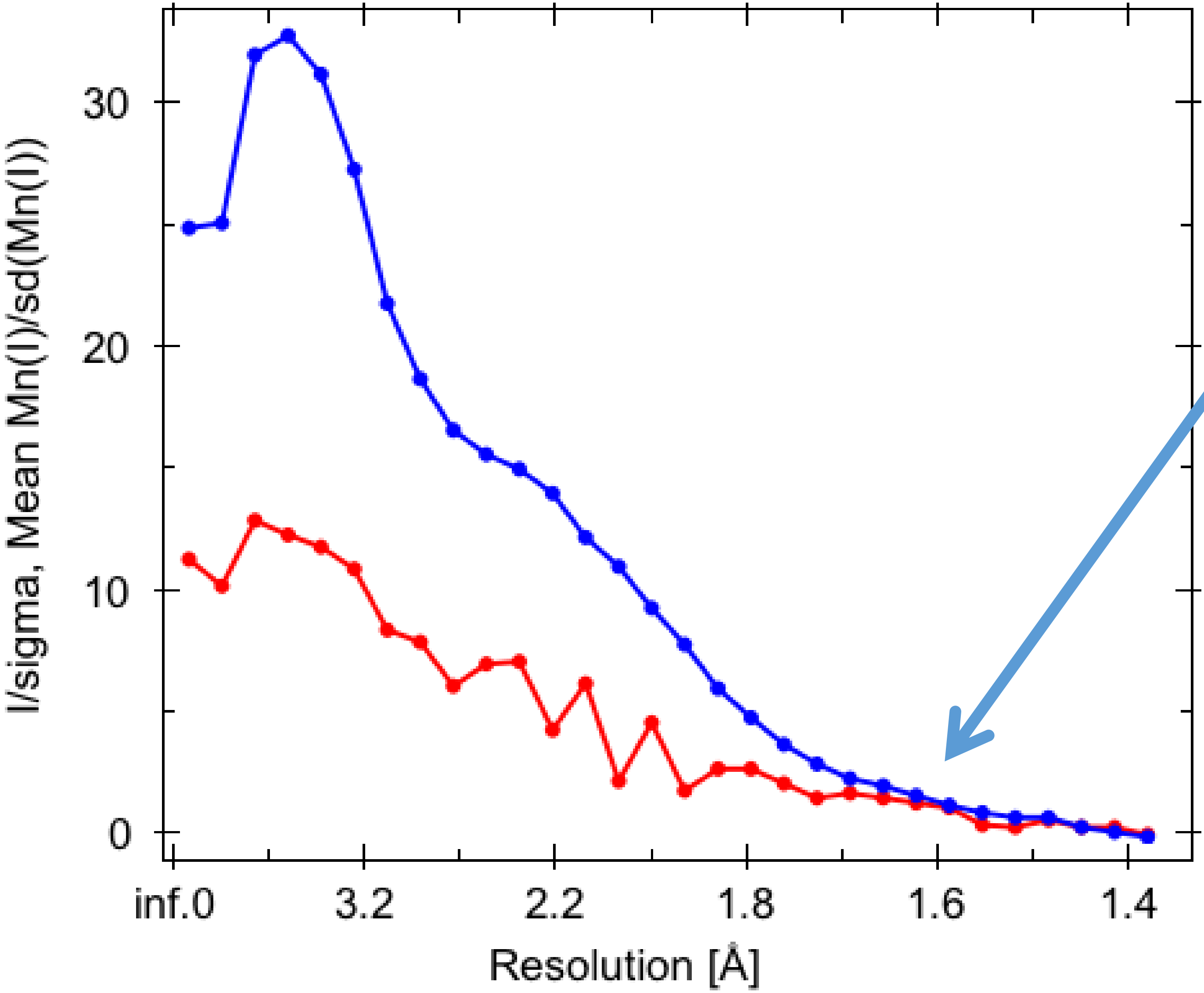
Want low number of rejected reflections per image; a maximum of around 5

Resolution: CC 1/2



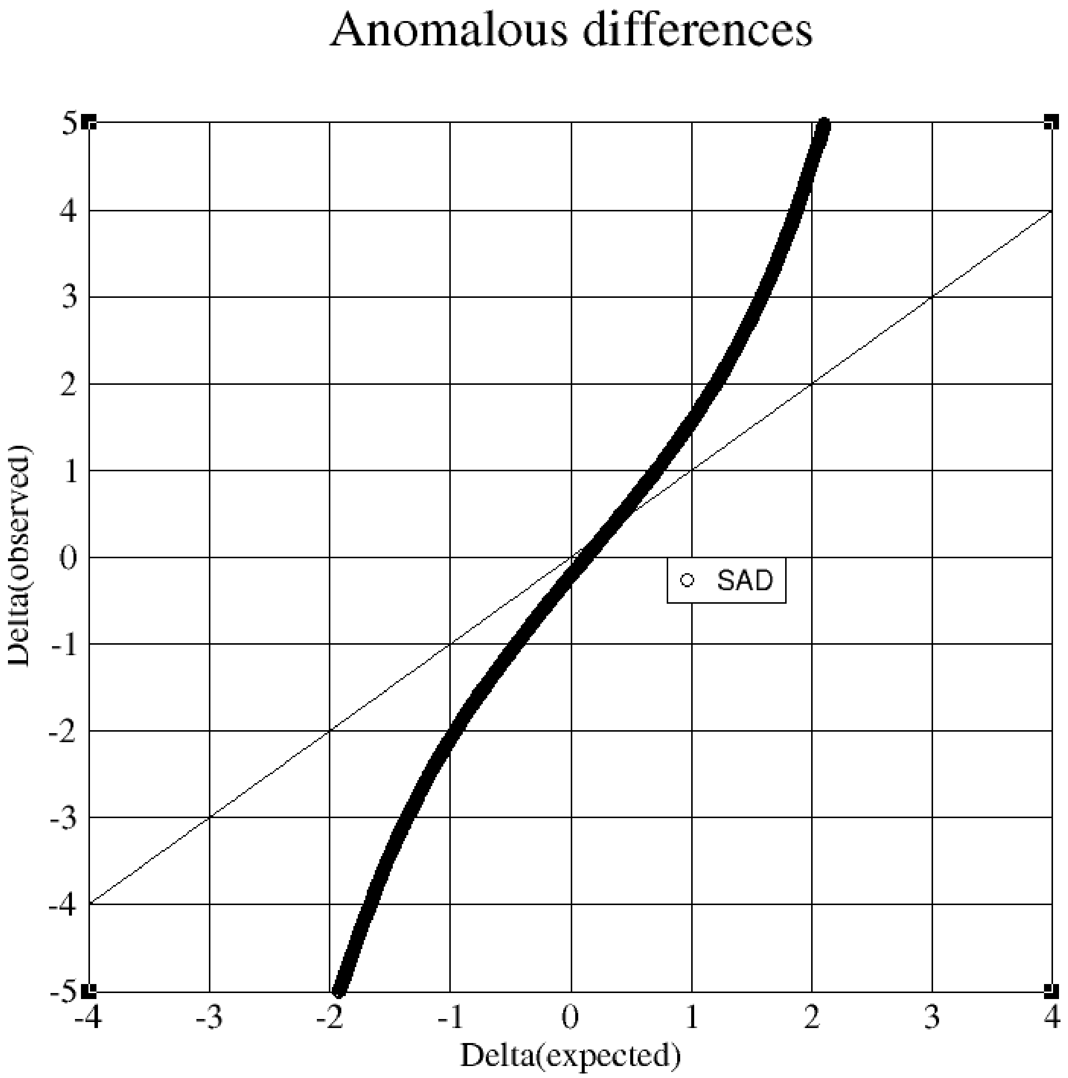
CC ½ around 0.5

Resolution: I/sig(I)



I/sig(I) around 1 (unmerged)

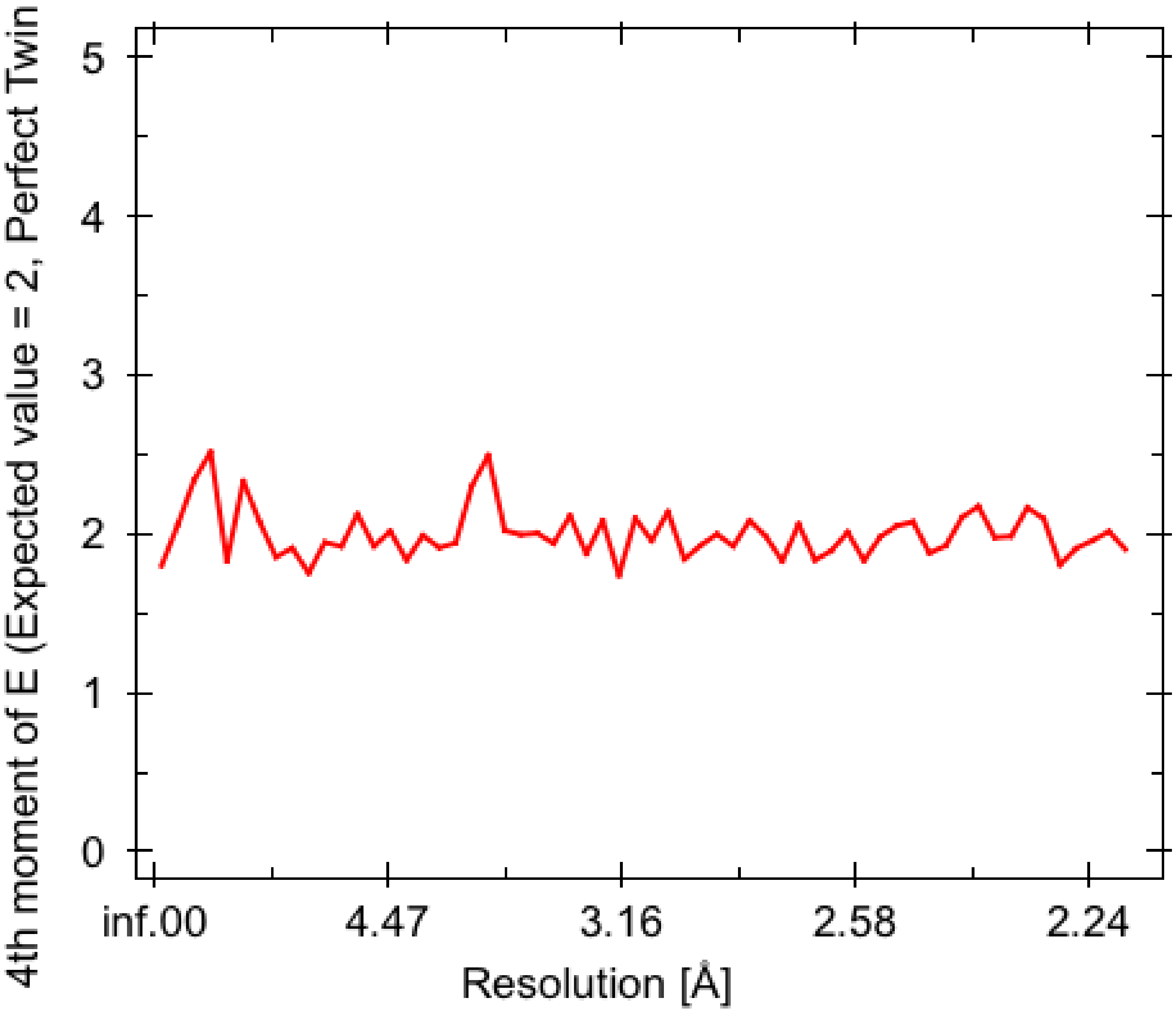
Anomalous differences



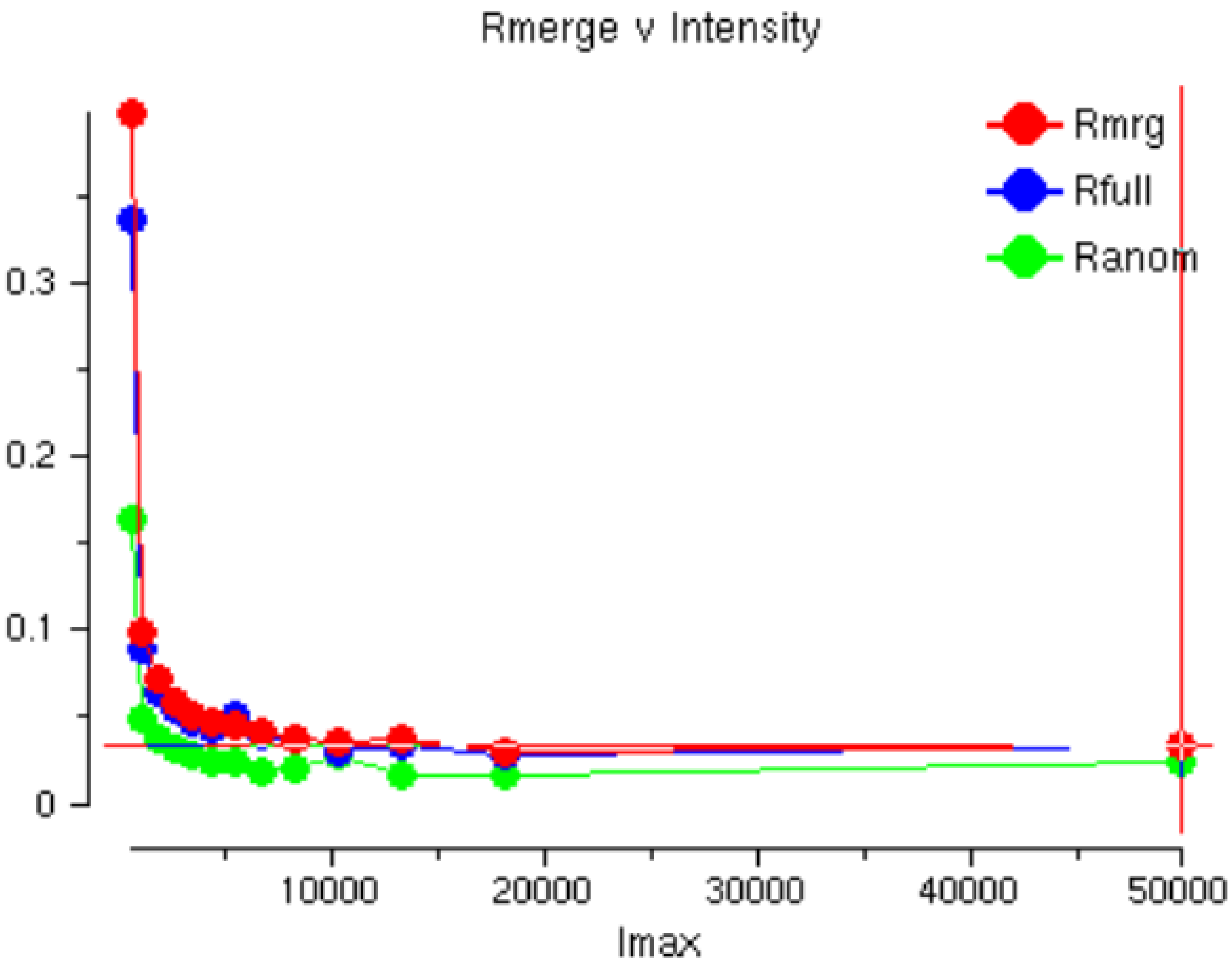
A slope > 1 in the centre indicates that the measured anomalous differences are greater than would be expected from the standard deviations.

Detecting twinning

- Depends on moments of intensity distributions
- Acentric E^4 is useful: if 2 probably not twinned, if 1.5 probably twinned
- Measures the spread of the merged intensity distribution
- Look at ctruncate output
- More twinning tests are performed, check ctruncate log

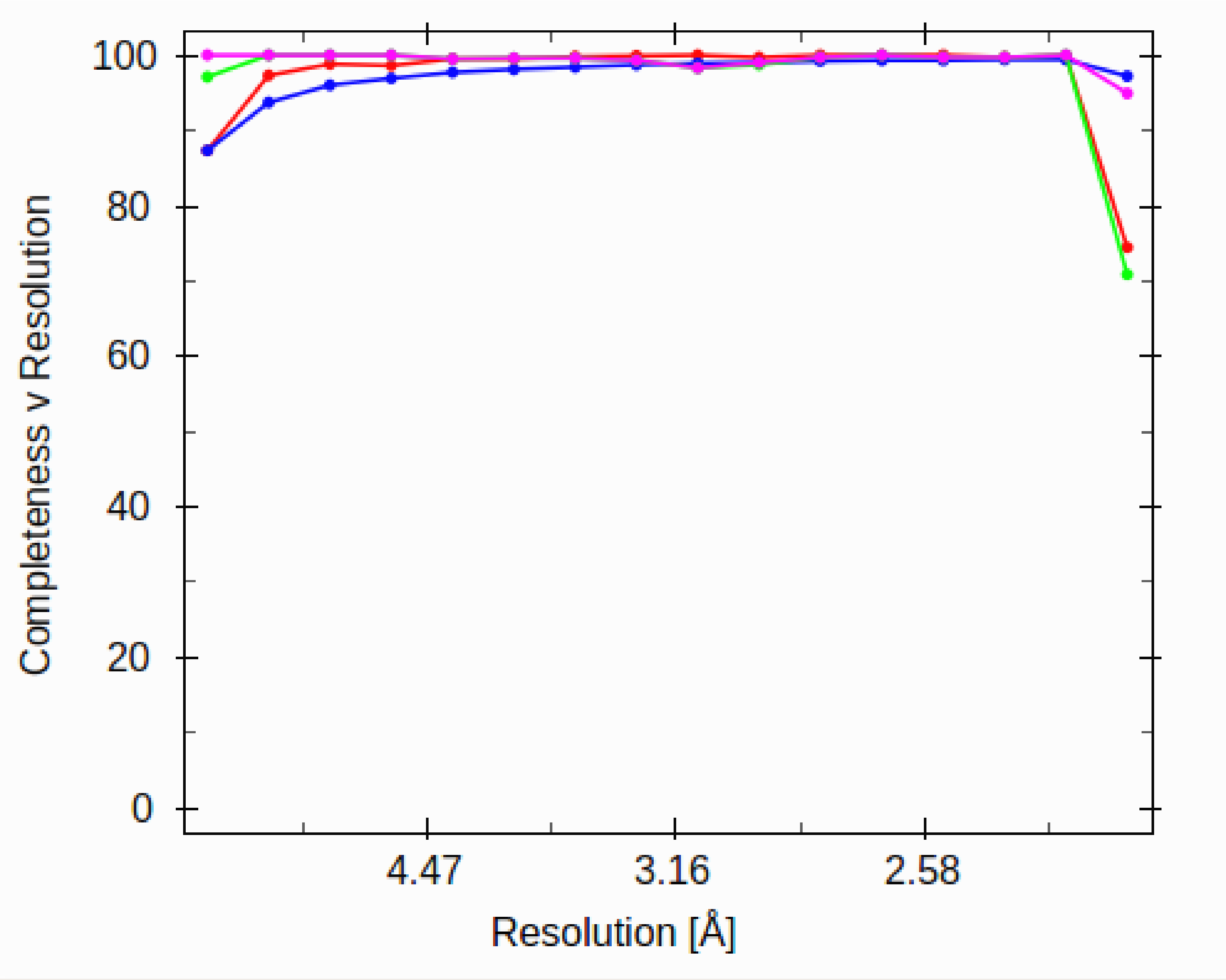

4th moment of E (Expected value = 2, Perfect Twin
5
4
3
2
1
0
inf.00
4.47
3.16
2.58
2.24
Resolution [Å]

Data Quality: Rmerge vs intensity



Rmerge is always large for small intensities. For large intensities it should be in the range 0.01 to 0.04 for good data. Larger values suggest that there are systematic errors.

Data Quality: completeness



Completeness of data should be as close to 100% as possible. Watch out for data with < 95% completeness.

Some loss of completeness can be tolerated in the outermost resolution bins.

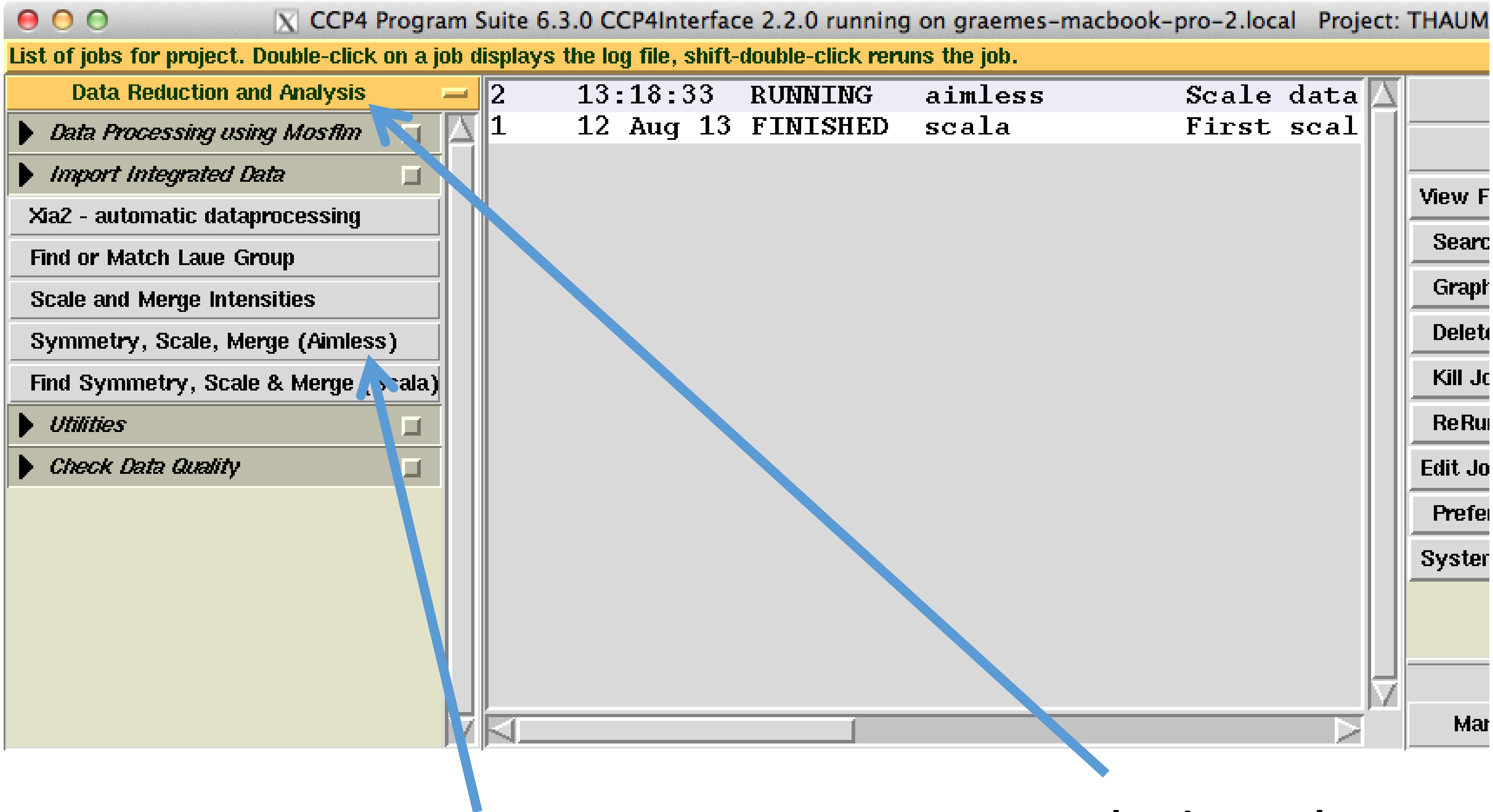
If you integrate to the corners of the detector, you may have low completeness at high resolution.

More scaling

Data processing and scaling with Mosflm and Aimless

After iMosflm

- Quick scaling useful to see if there is nothing pathological with your data set however…
- Move on to using CCP4i
- Probably easiest to create a new “project” and copy in the MTZ file from iMosflm
- N.B. take care with consistent indexing scaling multiple data sets



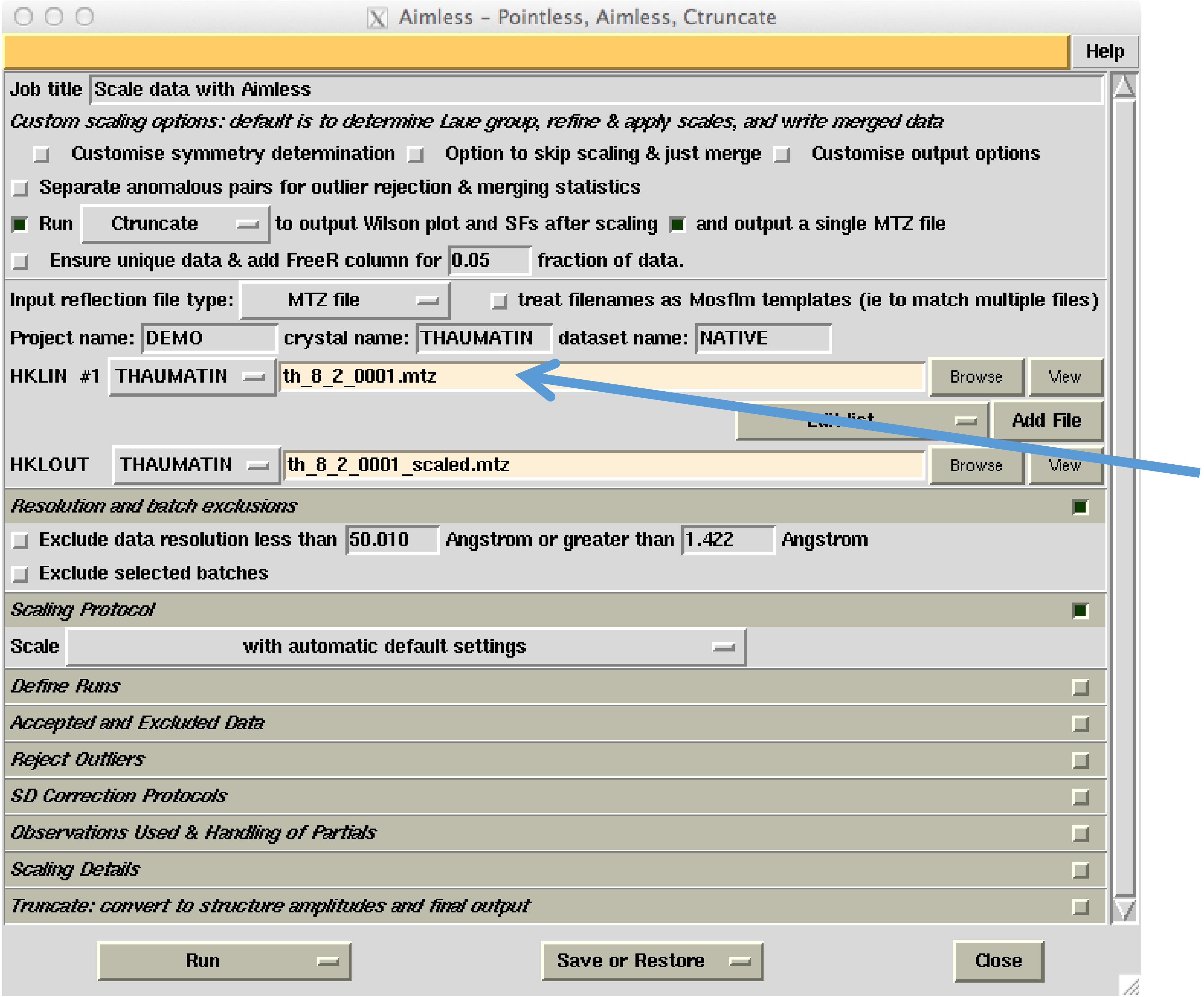
CCP4 Program Suite 6.3.0 CCP4Interface 2.2.0 running on graemes-macbook-pro-2.local Project: THAUM

List of jobs for project. Double-click on a job displays the log file, shift-double-click reruns the job.

```
2     13:18:33  RUNNING   aimless       Scale data
1     12 Aug 13 FINISHED  scala         First scal
```

Scale and merge data with Aimless

Data reduction and analysis module


Aimless - Pointless, Aimless, Ctruncate

Job title: Scale data with Aimless

Custom scaling options: default is to determine Laue group, refine & apply scales, and write merged data

☐ Customise symmetry determination ☐ Option to skip scaling & just merge ☐ Customise output options

☐ Separate anomalous pairs for outlier rejection & merging statistics

☑ Run Ctruncate to output Wilson plot and SFs after scaling ☑ and output a single MTZ file

☐ Ensure unique data & add FreeR column for 0.05 fraction of data.

Input reflection file type: MTZ file ☐ treat filenames as Mosflm templates (ie to match multiple files)

Project name: DEMO crystal name: THAUMATIN dataset name: NATIVE

HKLIN #1 THAUMATIN th_8_2_0001.mtz

HKLOUT THAUMATIN th_8_2_0001_scaled.mtz

Resolution and batch exclusions ☑

☐ Exclude data resolution less than 50.010 Angstrom or greater than 1.422 Angstrom

☐ Exclude selected batches

Scaling Protocol ☑

Scale with automatic default settings

Define Runs ☐

Accepted and Excluded Data ☐

Reject Outliers ☐

SD Correction Protocols ☐

Observations Used & Handling of Partials ☐

Scaling Details ☐

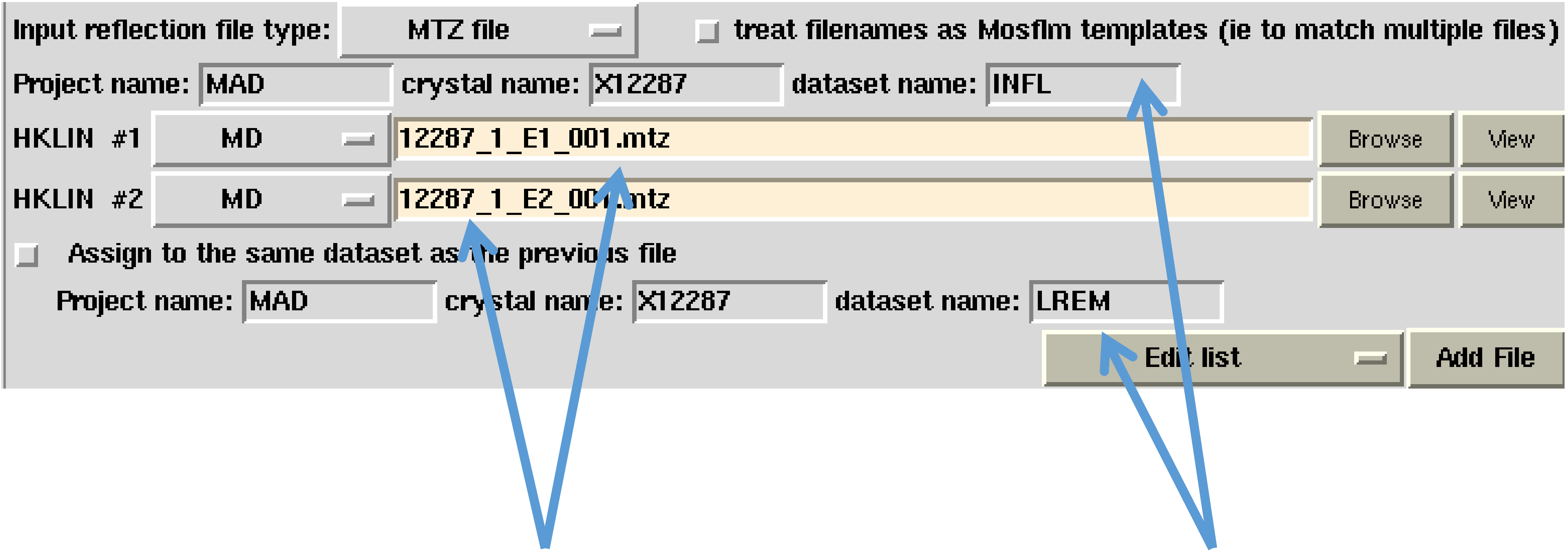
Truncate: convert to structure amplitudes and final output ☐

Mosflm MTZ file name

Selecting scaling models

- If you have or expect an anomalous signal, tell the scaling program
- Try different scaling models – if you don’t need a correction, don’t apply it
- If you have multiple data sets (MAD data set or multi-pass) scale them together now
- For multiple wavelengths assign different DATASET names

MAD



Input reflection file type: MTZ file

☐ treat filenames as Mosflm templates (ie to match multiple files)

Project name: MAD crystal name: X12287 dataset name: INFL

HKLIN #1 MD 12287_1_E1_001.mtz

HKLIN #2 MD 12287_1_E2_001.mtz

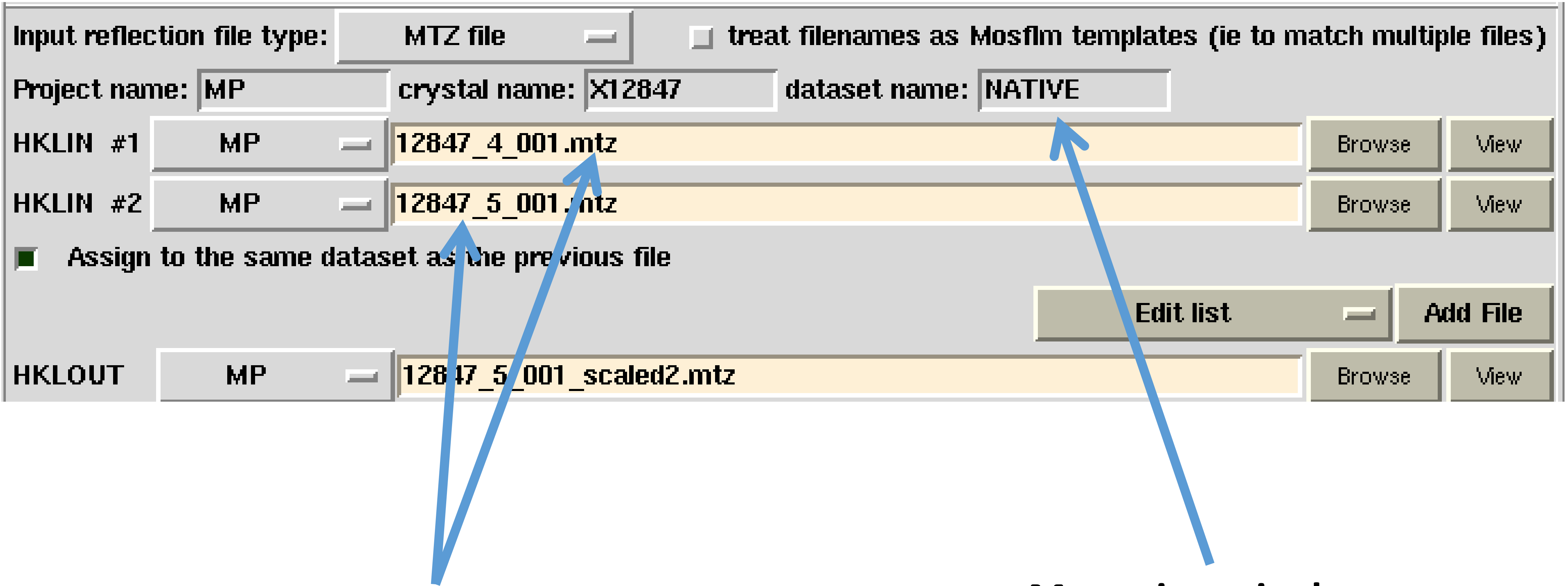
☐ Assign to the same dataset as the previous file

Project name: MAD crystal name: X12287 dataset name: LREM

Data from two iMosflm runs

Dataset names

Multi-pass



Input reflection file type: MTZ file

☐ treat filenames as Mosflm templates (ie to match multiple files)

Project name: MP crystal name: X12847 dataset name: NATIVE

HKLIN #1 MP 12847_4_001.mtz

HKLIN #2 MP 12847_5_001.mtz

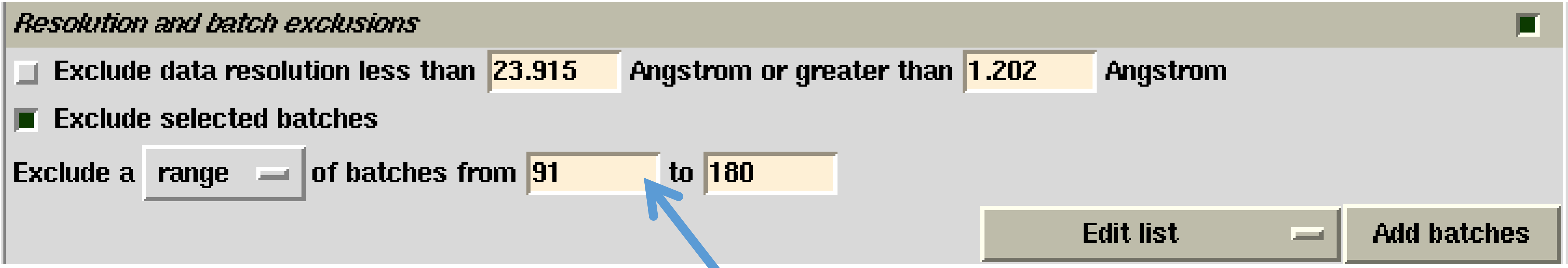
☑ Assign to the same dataset as the previous file

HKLOUT MP 12847_5_001_scaled2.mtz

Data from two iMosflm runs

Merge into single “native” data set

Radiation damage



Resolution and batch exclusions ☑

☐ Exclude data resolution less than 23.915 Angstrom or greater than 1.202 Angstrom

☑ Exclude selected batches

Exclude a range of batches from 91 to 180

This is the last half of the data

Thank you for listening!

http://www.ccp4.ac.uk