Data Collection & Data Analysis

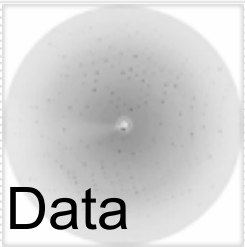
Basic Processing with Mosflm

Gwyndaf Evans
Diamond Light Source, UK.

Contents

- some thoughts on data collection
- simple processing with *Mosflm*
- data collection strategy with *Mosflm*

MOSFLM and SCALA in CCP4



Data

(----------MOSFLM----------)

Index → Strategy → Integrate

(---SCALA---)

Reindex → Sort → Scale/Merge

Detwin → Convert I to F

(-TRUNCATE-)

Optimization of Data Collection

- Pre-process at least one image (preferably two at 90º to each other) to obtain:
  - Cell parameters, crystal orientation and putative Laue group
- Estimate of mosaicity
- Effective resolution limit
  - Crystal to detector distance
  - Exposure time
- Strategy for data collection
- **Remember!**
  - This is the last experimental stage - if you collect bad data now you are stuck with it.
  - No data processing program can rescue the irredeemable!

What is needed prior to running Mosflm?

- X-ray images
- Experimental details (e.g. detector type, direct beam position, wavelength, etc)
- The program itself and a computer to run it on!

```
[localhost:~/test/muldlx1] harry% ipmosflm

 ********** Version 6.2.3 for Image plate and CCD data 10th July 2003 **********
 A.G.W. Leslie, MRC Laboratory Of Molecular Biology, HILLS ROAD, CAMBRIDGE CB2 2QH, UK
 E-mail andrew@mrc-lmb.cam.ac.uk
 New auto-indexing using DPS due to Ingo Steller Robert Bolotovsky and Michael Rossmann
 (1998) J. Appl. Cryst. 30, 1036-1040
 Original auto-indexing using REFIX due to Wolfgang Kabsch (Kabsch, W. (1993),
J. Appl. Cryst. 24, 795-800.)
 X-windows interface using xdl_view due to John Campbell (Daresbury Laboratory, UK.)
 (Campbell, J. W. (1995) J. Appl. Cryst. 28, 236-242.


 MOSFLM => image muldlx1_301.mar2000
 MOSFLM => go

  (Q) QOPEN: file opened on unit 1     Status: READONLY
<B><FONT COLOR="#FF0000"><!--SUMMARY_BEGIN-->
 Logical Name: muldlx1_301.mar2000     Filename: muldlx1_301.mar2000
<!--SUMMARY_END--></FONT></B>

 Crystal to detector distance of 250.00 mm taken from image header

 Wavelength of 1.54180A taken from image header
```

Crystal to detector distance of 250.00 mm taken from image header

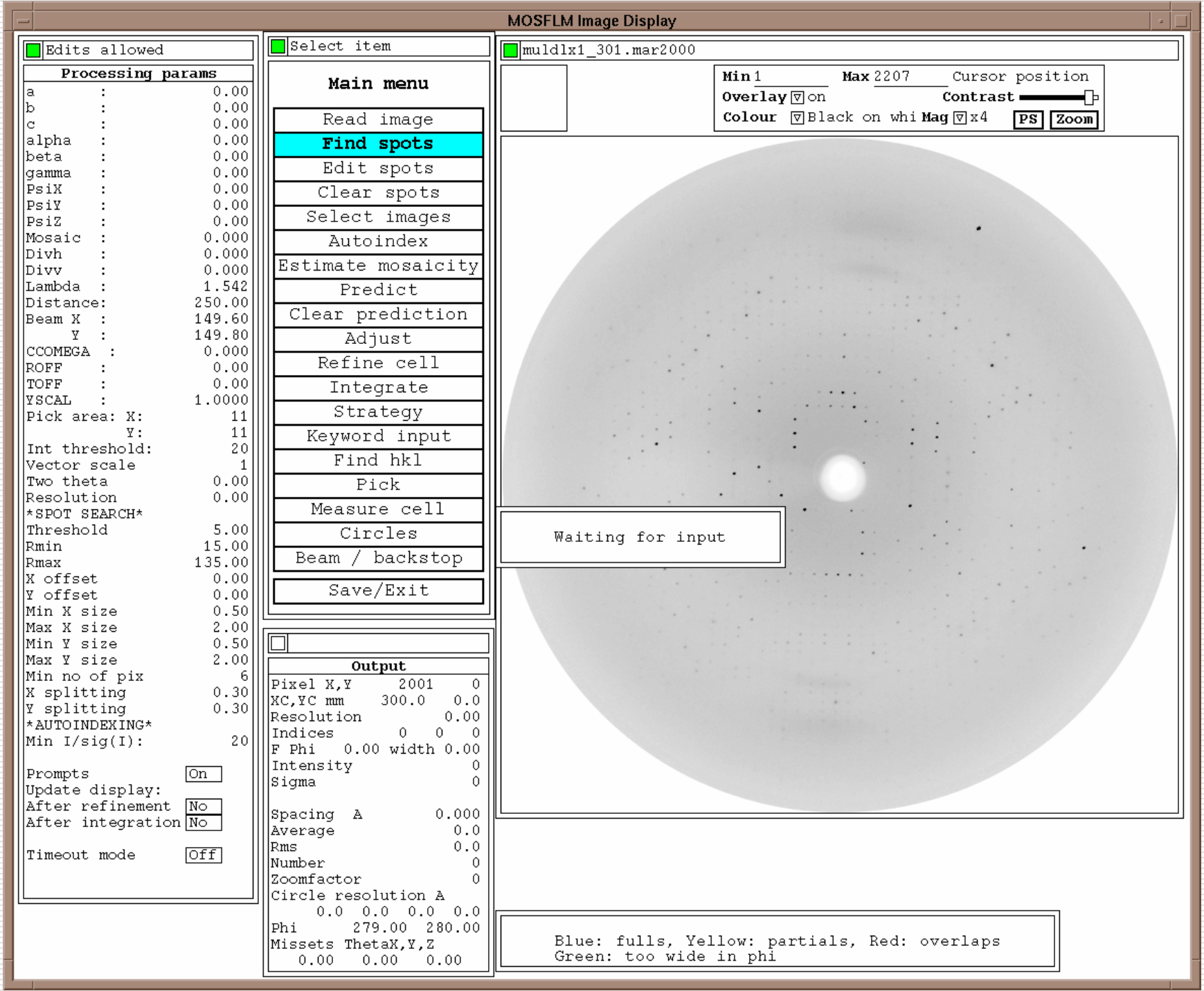
Wavelength of 1.54180A taken from image header

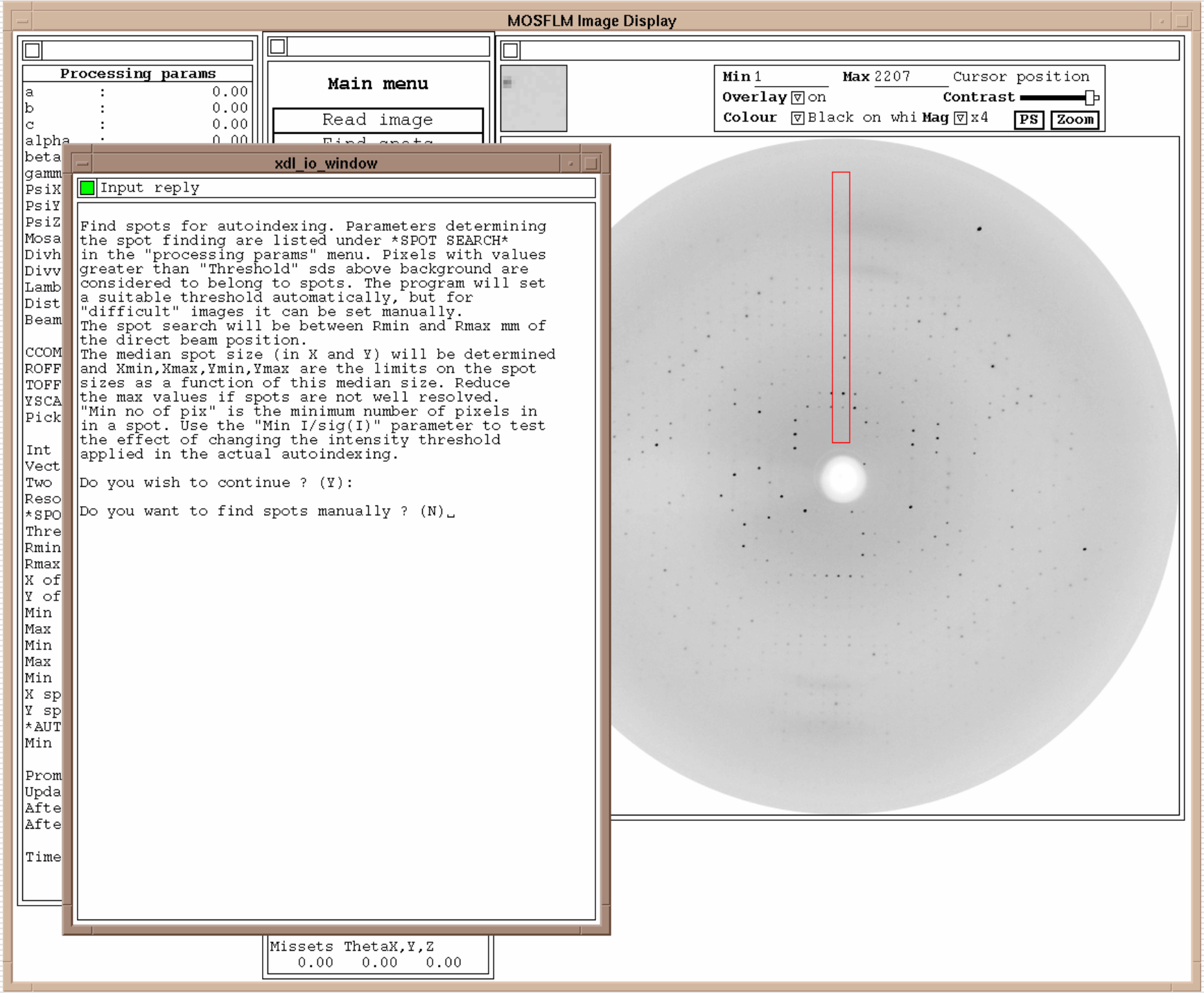
Pixel size of 0.1500 mm taken from image header.

Start and end phi values for image 1 from image header are 279.00 and 280.00 degrees.
image FILENAME: muldlx1_301.mar2000

(Q) QOPEN: file opened on unit 1 Status: READONLY
<B><FONT COLOR="#FF0000"><!--SUMMARY_BEGIN-->
Logical Name: muldlx1_301.mar2000 Filename: muldlx1_301.mar2000
<!--SUMMARY_END--></FONT></B>

The red circle denotes the region behind the backstop shadow
(Use BACKSTOP keyword to set this.)


MOSFLM Image Display
Edits allowed
Processing params
a : 0.00
b : 0.00
c : 0.00
alpha : 0.00
beta : 0.00
gamma : 0.00
PsiX : 0.00
PsiY : 0.00
PsiZ : 0.00
Mosaic : 0.000
Divh : 0.000
Divv : 0.000
Lambda : 1.542
Distance: 250.00
Beam X : 149.60
Y : 149.80
CCOMEGA : 0.000
ROFF : 0.00
TOFF : 0.00
YSCAL : 1.0000
Pick area: X: 11
Y: 11
Int threshold: 20
Vector scale 1
Two theta 0.00
Resolution 0.00
SPOT SEARCH
Threshold 5.00
Rmin 15.00
Rmax 135.00
X offset 0.00
Y offset 0.00
Min X size 0.50
Max X size 2.00
Min Y size 0.50
Max Y size 2.00
Min no of pix 6
X splitting 0.30
Y splitting 0.30
AUTOINDEXING
Min I/sig(I): 20
Prompts On
Update display:
After refinement No
After integration No
Timeout mode Off
Select item
Main menu
Read image
Find spots
Edit spots
Clear spots
Select images
Autoindex
Estimate mosaicity
Predict
Clear prediction
Adjust
Refine cell
Integrate
Strategy
Keyword input
Find hkl
Pick
Measure cell
Circles
Beam / backstop
Save/Exit
Output
Pixel X,Y 2001 0
XC,YC mm 300.0 0.0
Resolution 0.00
Indices 0 0 0
F Phi 0.00 width 0.00
Intensity 0
Sigma 0
Spacing A 0.000
Average 0.0
Rms 0.0
Number 0
Zoomfactor 0
Circle resolution A
0.0 0.0 0.0 0.0
Phi 279.00 280.00
Missets ThetaX,Y,Z
0.00 0.00 0.00
muldlx1_301.mar2000
Min 1
Max 2207
Cursor position
Overlay on
Contrast
Colour Black on whi
Mag x4
PS
Zoom
Waiting for input
Blue: fulls, Yellow: partials, Red: overlaps
Green: too wide in phi


MOSFLM Image Display
Processing params
a : 0.00
b : 0.00
c : 0.00
alpha :
beta
gamm
PsiX
PsiY
PsiZ
Mosa
Divh
Divv
Lamb
Dist
Beam
CCOM
ROFF
TOFF
YSCA
Pick
Int
Vect
Two
Reso
*SPO
Thre
Rmin
Rmax
X of
Y of
Min
Max
Min
Max
Min
X sp
Y sp
*AUT
Min
Prom
Upda
Afte
Afte
Time
Main menu
Read image
Min 1
Max 2207
Cursor position
Overlay on
Contrast
Colour Black on whi
Mag x4
PS
Zoom
xdl_io_window
Input reply
Find spots for autoindexing. Parameters determining
the spot finding are listed under *SPOT SEARCH*
in the "processing params" menu. Pixels with values
greater than "Threshold" sds above background are
considered to belong to spots. The program will set
a suitable threshold automatically, but for
"difficult" images it can be set manually.
The spot search will be between Rmin and Rmax mm of
the direct beam position.
The median spot size (in X and Y) will be determined
and Xmin,Xmax,Ymin,Ymax are the limits on the spot
sizes as a function of this median size. Reduce
the max values if spots are not well resolved.
"Min no of pix" is the minimum number of pixels in
in a spot. Use the "Min I/sig(I)" parameter to test
the effect of changing the intensity threshold
applied in the actual autoindexing.
Do you wish to continue ? (Y):
Do you want to find spots manually ? (N)
Missets ThetaX,Y,Z
0.00 0.00 0.00

MOSFLM Image Display

Edits allowed

Processing params

a :	126.19
b :	126.19
c :	74.31
alpha :	90.00
beta :	90.00
gamma :	120.00
PsiX :	0.00
PsiY :	0.00
PsiZ :	0.00
Mosaic :	0.510
Divh :	0.000
Divv :	0.000
Lambda :	1.542
Distance:	250.00
Beam X :	149.69
Y :	149.71
CCOMEGA :	0.000
ROFF :	0.00
TOFF :	0.00
YSCAL :	1.0000
Pick area: X:	11
Y:	11
Int threshold:	20
Vector scale	1
Two theta	0.00
Resolution	2.88
SPOT SEARCH	
Threshold	4.92
Rmin	15.00
Rmax	135.00
X offset	0.00
Y offset	0.00
Min X size	0.50
Max X size	2.00
Min Y size	0.50
Max Y size	2.00
Min no of pix	6
X splitting	0.30
Y splitting	0.30
AUTOINDEXING	
Min I/sig(I):	20

Prompts [On]
Update display:
After refinement [No]
After integration [No]

Timeout mode [Off]

Select item

Main menu

- Read image
- Find spots
- Edit spots
- Clear spots
- Select images
- Autoindex
- Estimate mosaicity
- Predict
- Clear prediction
- Adjust
- Refine cell
- Integrate
- Strategy
- Keyword input
- Find hkl
- Pick
- Measure cell
- Circles
- Beam / backstop

Save/Exit

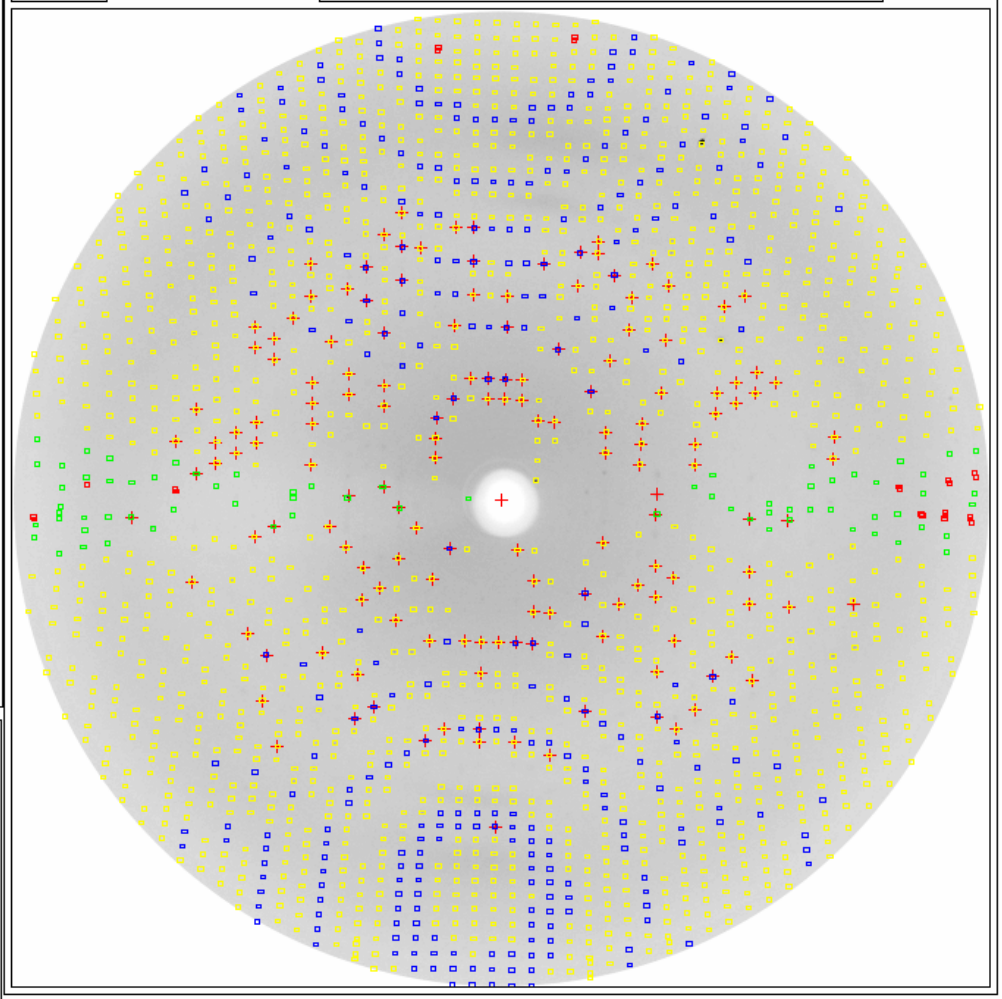
Output

Pixel X,Y	2001	0
XC,YC mm	300.0	0.0
Resolution		0.00
Indices	0 0	0
F Phi	0.00 width	0.00
Intensity		0
Sigma		0
Spacing A		0.000
Average		0.0
Rms		0.0
Number		0
Zoomfactor		0

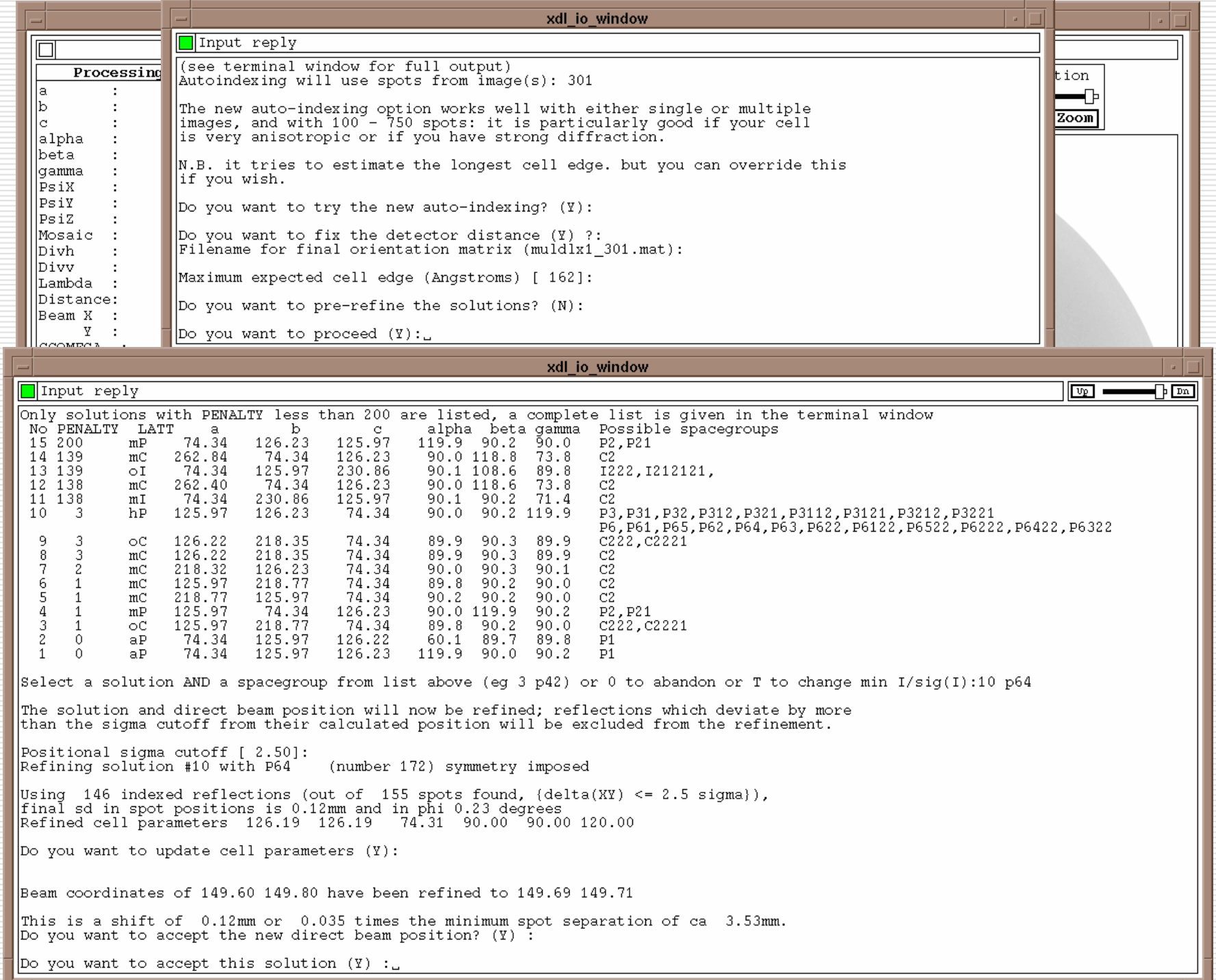
Circle resolution A
0.0 0.0 0.0 0.0
Phi 279.00 280.00
Missets ThetaX,Y,Z
0.00 0.00 0.00

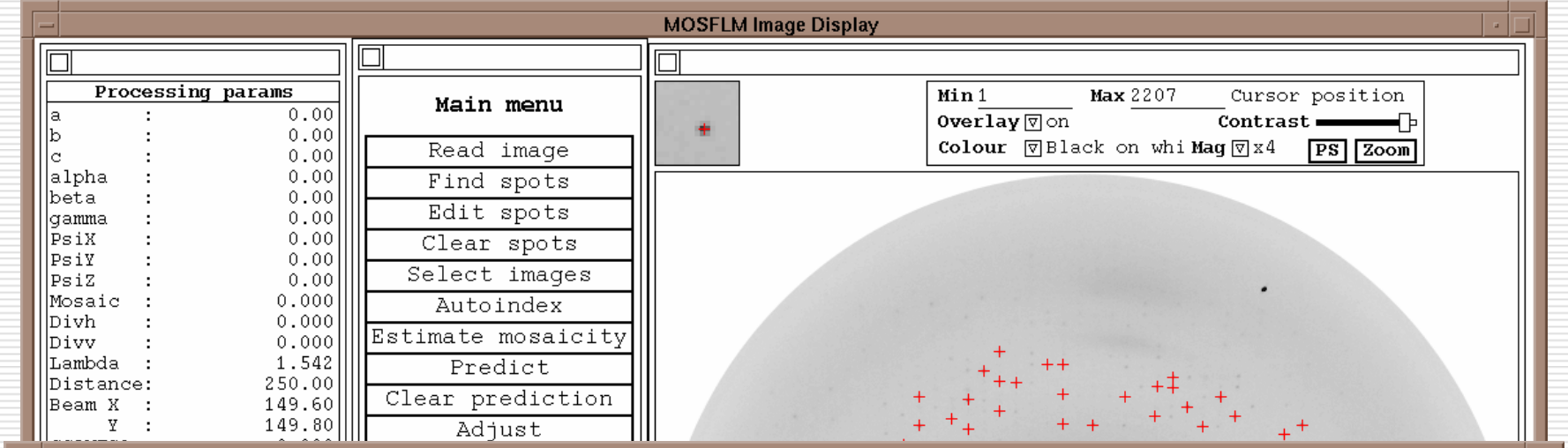
muldlx1_301.mar2000

Min 1 Max 2207 Cursor position
Overlay on Contrast
Colour Black on whi Mag x4 PS Zoom



Blue: fulls, Yellow: partials, Red: overlaps
Green: too wide in phi


xdl_io_window
Input reply
(see terminal window for full output)
Autoindexing will use spots from image(s): 301
The new auto-indexing option works well with either single or multiple
images, and with 100 - 750 spots: it is particularly good if your cell
is very anisotropic or if you have strong diffraction.
N.B. it tries to estimate the longest cell edge. but you can override this
if you wish.
Do you want to try the new auto-indexing? (Y):
Do you want to fix the detector distance (Y) ?:
Filename for final orientation matrix (muldlx1_301.mat):
Maximum expected cell edge (Angstroms) [162]:
Do you want to pre-refine the solutions? (N):
Do you want to proceed (Y):
Processing
a :
b :
c :
alpha :
beta :
gamma :
PsiX :
PsiY :
PsiZ :
Mosaic :
Divh :
Divv :
Lambda :
Distance:
Beam X :
Y :
Zoom
xdl_io_window
Input reply
Up
Dn
Only solutions with PENALTY less than 200 are listed, a complete list is given in the terminal window
No PENALTY LATT a b c alpha beta gamma Possible spacegroups
15 200 mP 74.34 126.23 125.97 119.9 90.2 90.0 P2,P21
14 139 mC 262.84 74.34 126.23 90.0 118.8 73.8 C2
13 139 oI 74.34 125.97 230.86 90.1 108.6 89.8 I222,I212121,
12 138 mC 262.40 74.34 126.23 90.0 118.6 73.8 C2
11 138 mI 74.34 230.86 125.97 90.1 90.2 71.4 C2
10 3 hP 125.97 126.23 74.34 90.0 90.2 119.9 P3,P31,P32,P312,P321,P3112,P3121,P3212,P3221
P6,P61,P65,P62,P64,P63,P622,P6122,P6522,P6222,P6422,P6322
9 3 oC 126.22 218.35 74.34 89.9 90.3 89.9 C222,C2221
8 3 mC 126.22 218.35 74.34 89.9 90.3 89.9 C2
7 2 mC 218.32 126.23 74.34 90.0 90.3 90.1 C2
6 1 mC 125.97 218.77 74.34 89.8 90.2 90.0 C2
5 1 mC 218.77 125.97 74.34 90.2 90.2 90.0 C2
4 1 mP 125.97 74.34 126.23 90.0 119.9 90.2 P2,P21
3 1 oC 125.97 218.77 74.34 89.8 90.2 90.0 C222,C2221
2 0 aP 74.34 125.97 126.22 60.1 89.7 89.8 P1
1 0 aP 74.34 125.97 126.23 119.9 90.0 90.2 P1
Select a solution AND a spacegroup from list above (eg 3 p42) or 0 to abandon or T to change min I/sig(I):10 p64
The solution and direct beam position will now be refined; reflections which deviate by more
than the sigma cutoff from their calculated position will be excluded from the refinement.
Positional sigma cutoff [2.50]:
Refining solution #10 with P64 (number 172) symmetry imposed
Using 146 indexed reflections (out of 155 spots found, {delta(XY) <= 2.5 sigma}),
final sd in spot positions is 0.12mm and in phi 0.23 degrees
Refined cell parameters 126.19 126.19 74.31 90.00 90.00 120.00
Do you want to update cell parameters (Y):
Beam coordinates of 149.60 149.80 have been refined to 149.69 149.71
This is a shift of 0.12mm or 0.035 times the minimum spot separation of ca 3.53mm.
Do you want to accept the new direct beam position? (Y) :
Do you want to accept this solution (Y) :



xdl_io_window

Input reply

Up Dn

Only solutions with PENALTY less than 200 are listed, a complete list is given in the terminal window

No	PENALTY	LATT	a	b	c	alpha	beta	gamma	Possible spacegroups
15	200	mP	74.34	126.23	125.97	119.9	90.2	90.0	P2,P21
14	139	mC	262.84	74.34	126.23	90.0	118.8	73.8	C2
13	139	oI	74.34	125.97	230.86	90.1	108.6	89.8	I222,I212121,
12	138	mC	262.40	74.34	126.23	90.0	118.6	73.8	C2
11	138	mI	74.34	230.86	125.97	90.1	90.2	71.4	C2
10	3	hP	125.97	126.23	74.34	90.0	90.2	119.9	P3,P31,P32,P312,P321,P3112,P3121,P3212,P3221 P6,P61,P65,P62,P64,P63,P622,P6122,P6522,P6222,P6422,P6322
9	3	oC	126.22	218.35	74.34	89.9	90.3	89.9	C222,C2221
8	3	mC	126.22	218.35	74.34	89.9	90.3	89.9	C2
7	2	mC	218.32	126.23	74.34	90.0	90.3	90.1	C2
6	1	mC	125.97	218.77	74.34	89.8	90.2	90.0	C2
5	1	mC	218.77	125.97	74.34	90.2	90.2	90.0	C2
4	1	mP	125.97	74.34	126.23	90.0	119.9	90.2	P2,P21
3	1	oC	125.97	218.77	74.34	89.8	90.2	90.0	C222,C2221
2	0	aP	74.34	125.97	126.22	60.1	89.7	89.8	P1
1	0	aP	74.34	125.97	126.23	119.9	90.0	90.2	P1

Select a solution AND a spacegroup from list above (eg 3 p42) or 0 to abandon or T to change min I/sig(I):10 p64

The solution and direct beam position will now be refined; reflections which deviate by more
than the sigma cutoff from their calculated position will be excluded from the refinement.

Positional sigma cutoff [2.50]:
Refining solution #10 with P64 (number 172) symmetry imposed

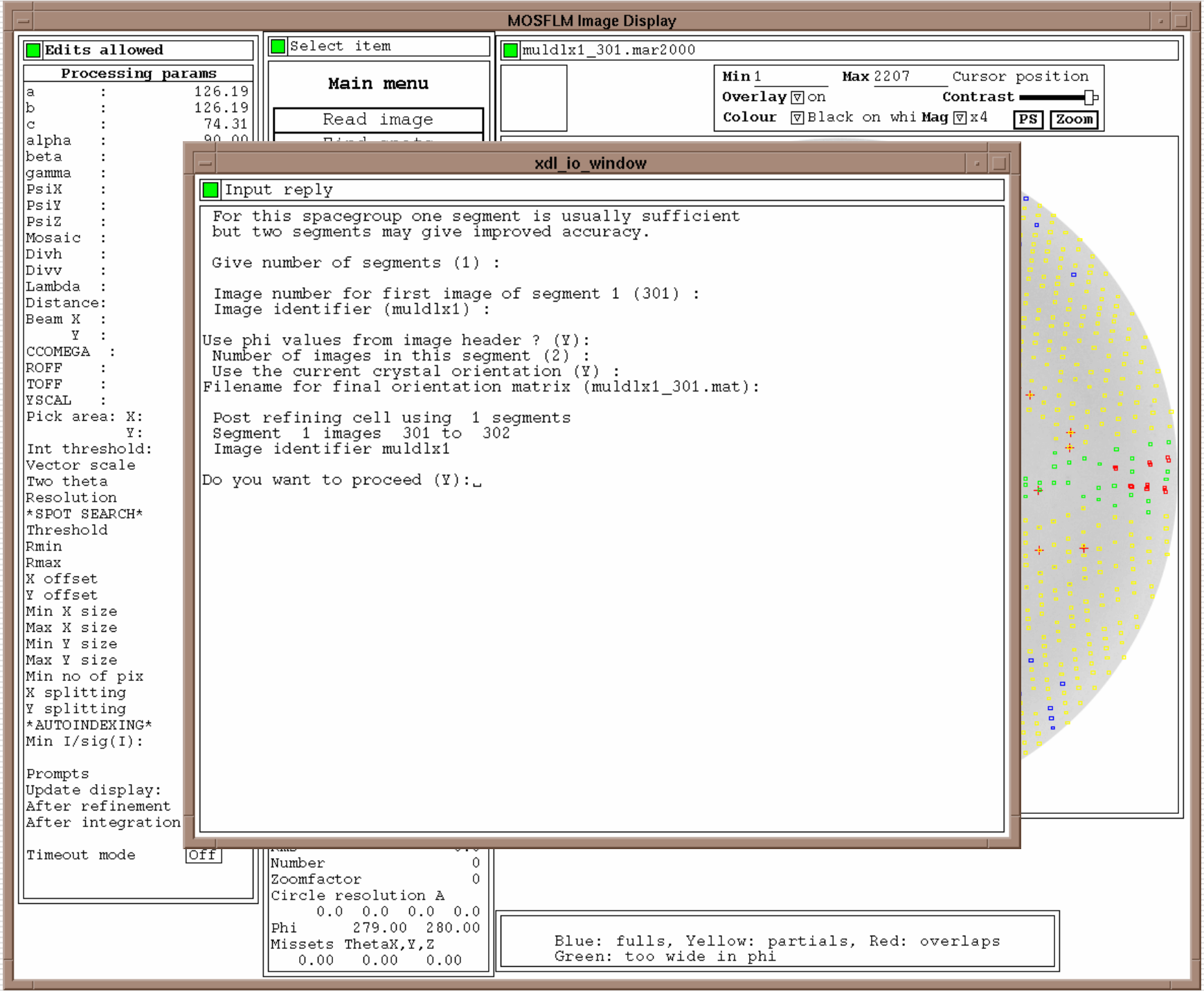
Using 146 indexed reflections (out of 155 spots found, {delta(XY) <= 2.5 sigma}),
final sd in spot positions is 0.12mm and in phi 0.23 degrees
Refined cell parameters 126.19 126.19 74.31 90.00 90.00 120.00

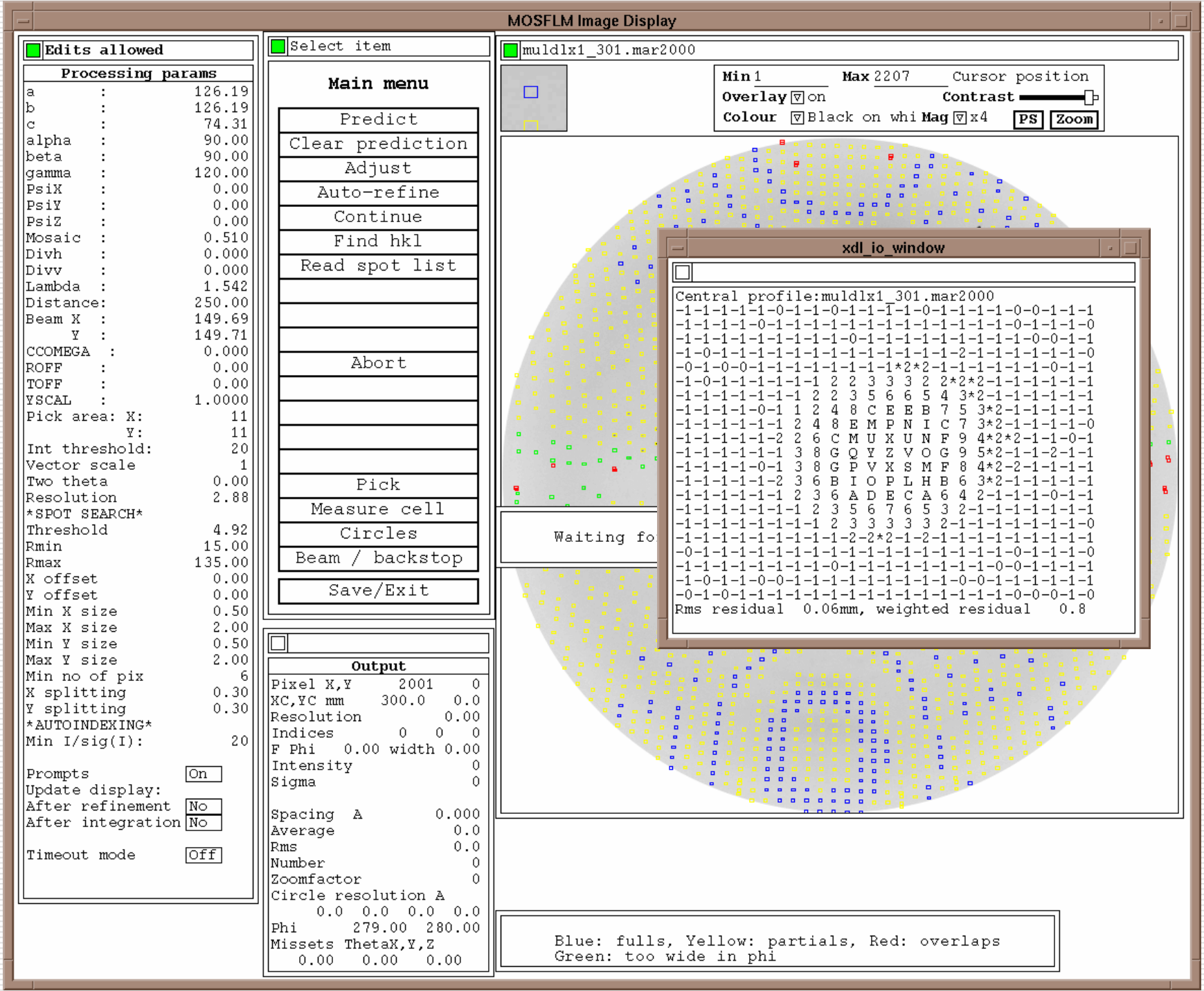
Do you want to update cell parameters (Y):

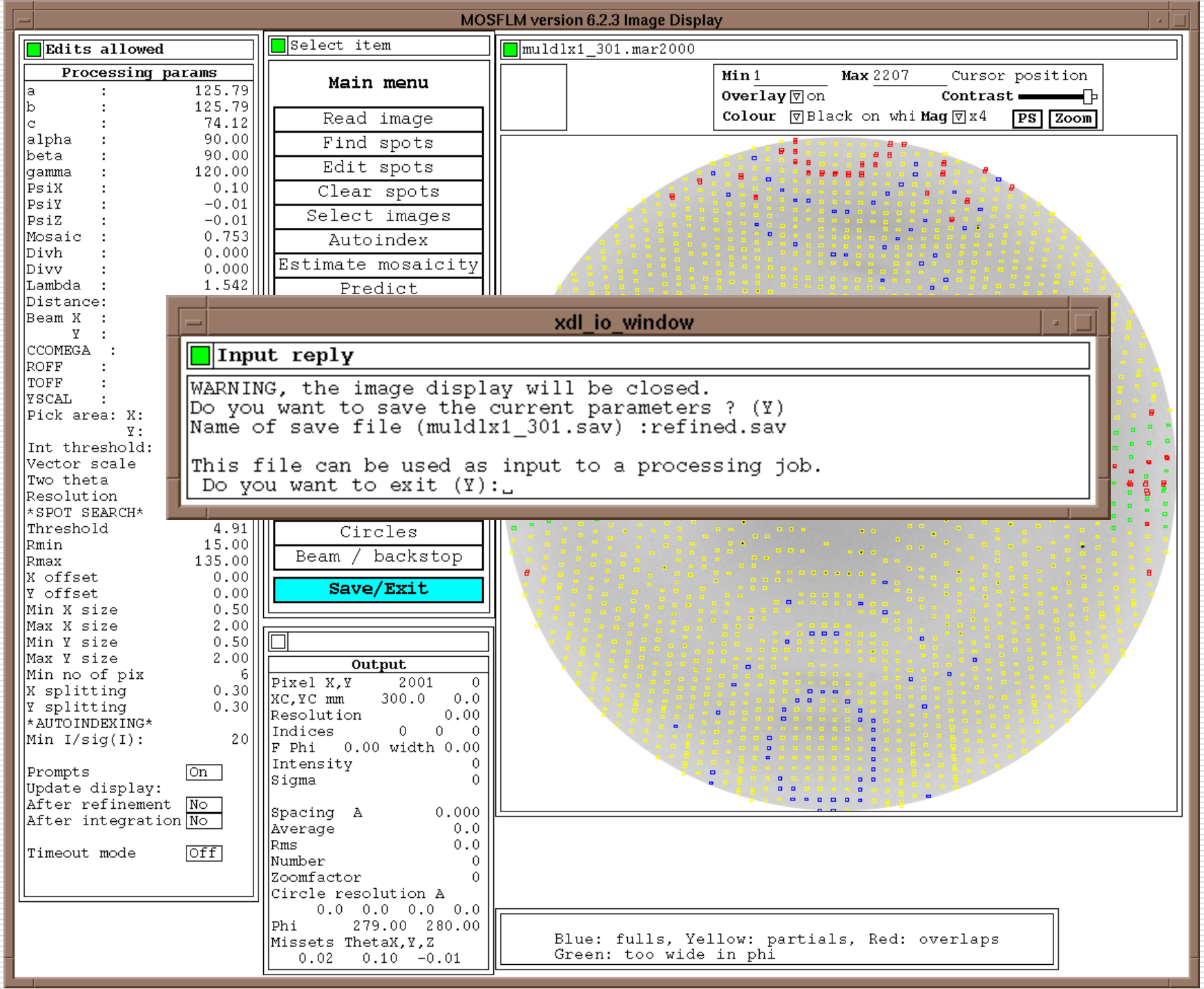
Beam coordinates of 149.60 149.80 have been refined to 149.69 149.71

This is a shift of 0.12mm or 0.035 times the minimum spot separation of ca 3.53mm.
Do you want to accept the new direct beam position? (Y) :

Do you want to accept this solution (Y) :


MOSFLM Image Display
Edits allowed
Processing params
a : 126.19
b : 126.19
c : 74.31
alpha :
beta :
gamma :
PsiX :
PsiY :
PsiZ :
Mosaic :
Divh :
Divv :
Lambda :
Distance:
Beam X :
Y :
CCOMEGA :
ROFF :
TOFF :
YSCAL :
Pick area: X:
Y:
Int threshold:
Vector scale
Two theta
Resolution
SPOT SEARCH
Threshold
Rmin
Rmax
X offset
Y offset
Min X size
Max X size
Min Y size
Max Y size
Min no of pix
X splitting
Y splitting
AUTOINDEXING
Min I/sig(I):
Prompts
Update display:
After refinement
After integration
Timeout mode
Off
Select item
Main menu
Read image
muldlx1_301.mar2000
Min 1
Max 2207
Cursor position
Overlay on
Contrast
Colour Black on whi
Mag x4
PS
Zoom
xdl_io_window
Input reply
For this spacegroup one segment is usually sufficient
but two segments may give improved accuracy.
Give number of segments (1) :
Image number for first image of segment 1 (301) :
Image identifier (muldlx1) :
Use phi values from image header ? (Y):
Number of images in this segment (2) :
Use the current crystal orientation (Y) :
Filename for final orientation matrix (muldlx1_301.mat):
Post refining cell using 1 segments
Segment 1 images 301 to 302
Image identifier muldlx1
Do you want to proceed (Y):
Number 0
Zoomfactor 0
Circle resolution A
0.0 0.0 0.0 0.0
Phi 279.00 280.00
Missets ThetaX,Y,Z
0.00 0.00 0.00
Blue: fulls, Yellow: partials, Red: overlaps
Green: too wide in phi


MOSFLM Image Display
Edits allowed
Processing params
a : 126.19
b : 126.19
c : 74.31
alpha : 90.00
beta : 90.00
gamma : 120.00
PsiX : 0.00
PsiY : 0.00
PsiZ : 0.00
Mosaic : 0.510
Divh : 0.000
Divv : 0.000
Lambda : 1.542
Distance: 250.00
Beam X : 149.69
Y : 149.71
CCOMEGA : 0.000
ROFF : 0.00
TOFF : 0.00
YSCAL : 1.0000
Pick area: X: 11
Y: 11
Int threshold: 20
Vector scale 1
Two theta 0.00
Resolution 2.88
SPOT SEARCH
Threshold 4.92
Rmin 15.00
Rmax 135.00
X offset 0.00
Y offset 0.00
Min X size 0.50
Max X size 2.00
Min Y size 0.50
Max Y size 2.00
Min no of pix 6
X splitting 0.30
Y splitting 0.30
AUTOINDEXING
Min I/sig(I): 20
Prompts On
Update display:
After refinement No
After integration No
Timeout mode Off
Select item
Main menu
Predict
Clear prediction
Adjust
Auto-refine
Continue
Find hkl
Read spot list
Abort
Pick
Measure cell
Circles
Beam / backstop
Save/Exit
Output
Pixel X,Y 2001 0
XC,YC mm 300.0 0.0
Resolution 0.00
Indices 0 0 0
F Phi 0.00 width 0.00
Intensity 0
Sigma 0
Spacing A 0.000
Average 0.0
Rms 0.0
Number 0
Zoomfactor 0
Circle resolution A
0.0 0.0 0.0 0.0
Phi 279.00 280.00
Missets ThetaX,Y,Z
0.00 0.00 0.00
muldlx1_301.mar2000
Min 1
Max 2207
Cursor position
Overlay on
Contrast
Colour Black on whi
Mag x4
PS
Zoom
xdl_io_window
Central profile:muldlx1_301.mar2000
Rms residual 0.06mm, weighted residual 0.8
Waiting fo
Blue: fulls, Yellow: partials, Red: overlaps
Green: too wide in phi


MOSFLM version 6.2.3 Image Display
Edits allowed
Processing params
a : 125.79
b : 125.79
c : 74.12
alpha : 90.00
beta : 90.00
gamma : 120.00
PsiX : 0.10
PsiY : -0.01
PsiZ : -0.01
Mosaic : 0.753
Divh : 0.000
Divv : 0.000
Lambda : 1.542
Distance:
Beam X :
Y :
CCOMEGA :
ROFF :
TOFF :
YSCAL :
Pick area: X:
Y:
Int threshold:
Vector scale
Two theta
Resolution
SPOT SEARCH
Threshold 4.91
Rmin 15.00
Rmax 135.00
X offset 0.00
Y offset 0.00
Min X size 0.50
Max X size 2.00
Min Y size 0.50
Max Y size 2.00
Min no of pix 6
X splitting 0.30
Y splitting 0.30
AUTOINDEXING
Min I/sig(I): 20
Prompts On
Update display:
After refinement No
After integration No
Timeout mode Off
Select item
Main menu
Read image
Find spots
Edit spots
Clear spots
Select images
Autoindex
Estimate mosaicity
Predict
Circles
Beam / backstop
Save/Exit
Output
Pixel X,Y 2001 0
XC,YC mm 300.0 0.0
Resolution 0.00
Indices 0 0 0
F Phi 0.00 width 0.00
Intensity 0
Sigma 0
Spacing A 0.000
Average 0.0
Rms 0.0
Number 0
Zoomfactor 0
Circle resolution A
0.0 0.0 0.0 0.0
Phi 279.00 280.00
Missets ThetaX,Y,Z
0.02 0.10 -0.01
muldlx1_301.mar2000
Min 1
Max 2207
Cursor position
Overlay on
Contrast
Colour Black on whi
Mag x4
PS
Zoom
xdl_io_window
Input reply
WARNING, the image display will be closed.
Do you want to save the current parameters ? (Y)
Name of save file (muldlx1_301.sav) :refined.sav
This file can be used as input to a processing job.
Do you want to exit (Y):
Blue: fulls, Yellow: partials, Red: overlaps
Green: too wide in phi

Running the STRATEGY option

- In the GUI, click on the STRATEGY button
- On the command-line, type STRATEGY.
- In either case, determining a suitable strategy for data collection once you have a cell, orientation and crystal symmetry is straightforward.
- Then run TESTGEN to check for overlaps

xdl_io_window

Input reply

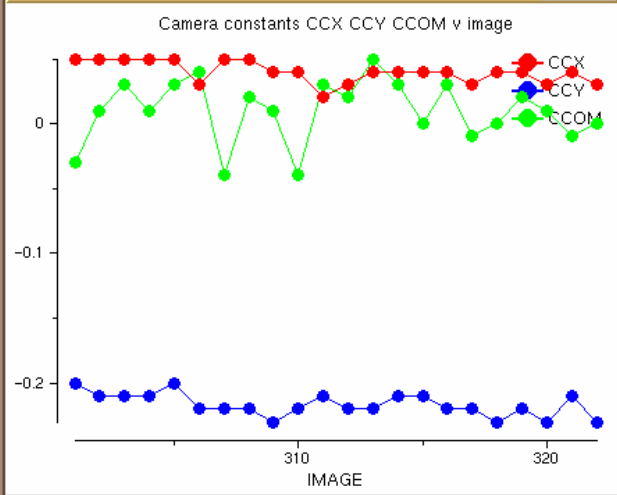
```
To test the full rotation for this Laue group just type STRATEGY at the prompt.
To test a smaller rotation (eg 60 degrees) in 2 different segments type:
STRATEGY SEGMENTS 2 ROTATE 60
A suitable speedup factor will be set by default but can be overridden
eg STRATEGY ROTATE 60 SEGMENTS 2 SPEEDUP 10
To try specific phi ranges type : STRATEGY START -20 END 60

Do you want to proceed (Y):
MOSFLM =>strategy
Running strategy in default mode. Type GO to continue or ABORT to stop.
MOSFLM =>go

SPEEDUP factor set to    15.0
Angles between a axis and X,Y,Z axes  54.552  35.492  88.463
Angles between b axis and X,Y,Z axes 174.303  84.515  91.537
Angles between c axis and X,Y,Z axes  91.615  90.736   1.775
The c axis is closest to the rotation axis (angle  1.77)
Rotation angle to get the a axis in XZ plane  -54.54
Rotation angle to get the b axis in XZ plane    5.49
Unique axis is: c
Rotation angle to get the unique axis into the YZ plane:  245.5
***** WARNING *****
***** WARNING *****
***** WARNING *****
***** WARNING *****
With the crystal in the current orientation it is not possible to collect
100% of the data, some will be lost in the cusp.
To avoid a cusp, the c-axis must be at an angle between 74.5 and 15.5 degrees
away from the rotation axis (Z)
You may want to adjust the orientation with the goniometer arcs.

Generating reflection list

Generated reflections for phi range -145.0 to  -25.0
Sorting the generated list.
Generating the list of unique reflections.
Merging  the generated and unique reflection list.

Checking completeness of data
Optimum rotation gives  97.7% of unique data.
This corresponds to the following rotation range(s):
From  255.0 to 315.0 degrees
Type "STATS" at prompt for full statistics.

**********************************************************************
It may be possible to get a higher completeness using two segments,
or a slightly greater rotation angle.
**********************************************************************

Use the TESTGEN option to determine appropriate oscillation angles to avoid
spatial overlaps.
STRATEGY =>
```

Checking the output (1)

- There are two useful log files;
  - SUMMARY - this is of most use when viewed with the CCP4 graph viewer LOGGRAPH, as it contains graphs of parameters which have varied through the data processing.
  - mosflm.lp - this can be very large, and contains a complete record of the experiment.

Loggraph summary.1

File Appearance Edit Utilities Help



317.6,0.083

Tables in File

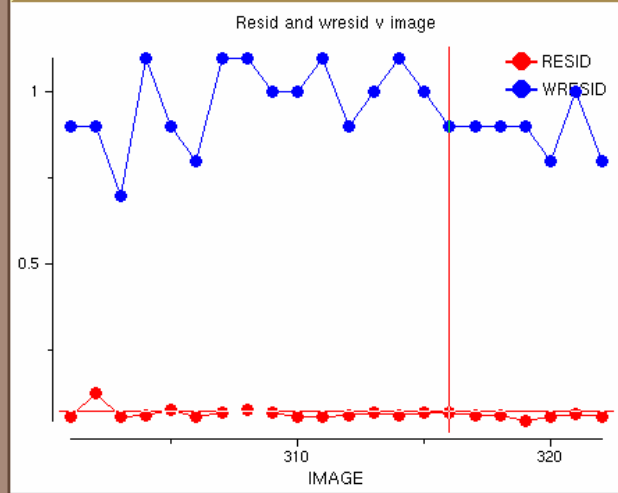
- Refined detector parameters
- Post refinement

Graphs in Selected Table

- Camera constants CCX CCY CCOM v image
- Distance v image
- Yscale v image
- Tilt and twist v image
- ROFF and TOFF v image

Loggraph summary.1

File Appearance Edit Utilities Help



315.9,0.076 RESID 316.0 , 0.07

Tables in File

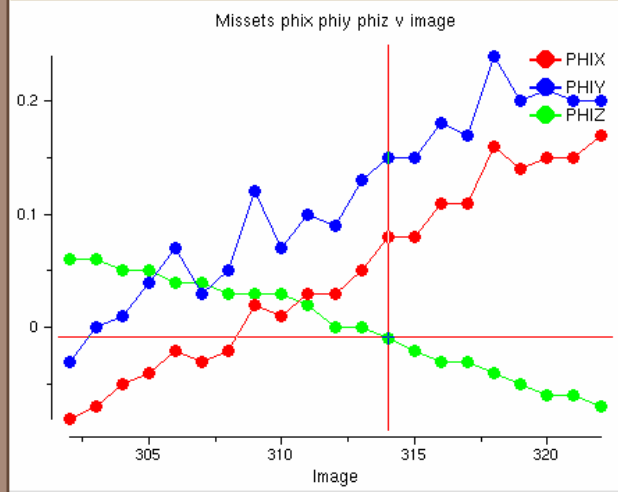
- Refined detector parameters
- Post refinement

Graphs in Selected Table

- Tilt and twist v image
- ROFF and TOFF v image
- Resid and wresid v image
- I/sig(I) overall and outer v image
- Rsym v image

Loggraph summary.1

File Appearance Edit Utilities Help



313.9,-0.00 PHIZ 314.0 , -0.01

Tables in File

- Refined detector parameters
- Post refinement

Graphs in Selected Table

- Missets phix phiy phiz v image
- Cell parameters A,B,C v image
- Cell angles alpha beta gamma v image
- Mosaic spread v image
- Beam divergences v image

Checking the output (2)

- If everything has gone right so far;
  - check the MTZ file - is it as you'd expect?

More on Processing with Mosflm

- what do you do about the warnings?
- what if are real problems?
- the new ccp4i Mosflm task

```
HEADER INFORMATION FROM OUTPUT MTZ FILE
 Logical Name: muldlx1_301.mtz  Filename: muldlx1_301.mtz

 <snip>

 * Number of Columns = 18

 * Number of Reflections = 43904

 * Missing value set to NaN in input mtz file

 * Number of Batches = 22

 * Column Labels :

 H K L M/ISYM BATCH I SIGI IPR SIGIPR FRACTIONCALC XDET YDET...
 * Column Types :

 H H H Y B J Q J Q R R R R R I I R
```

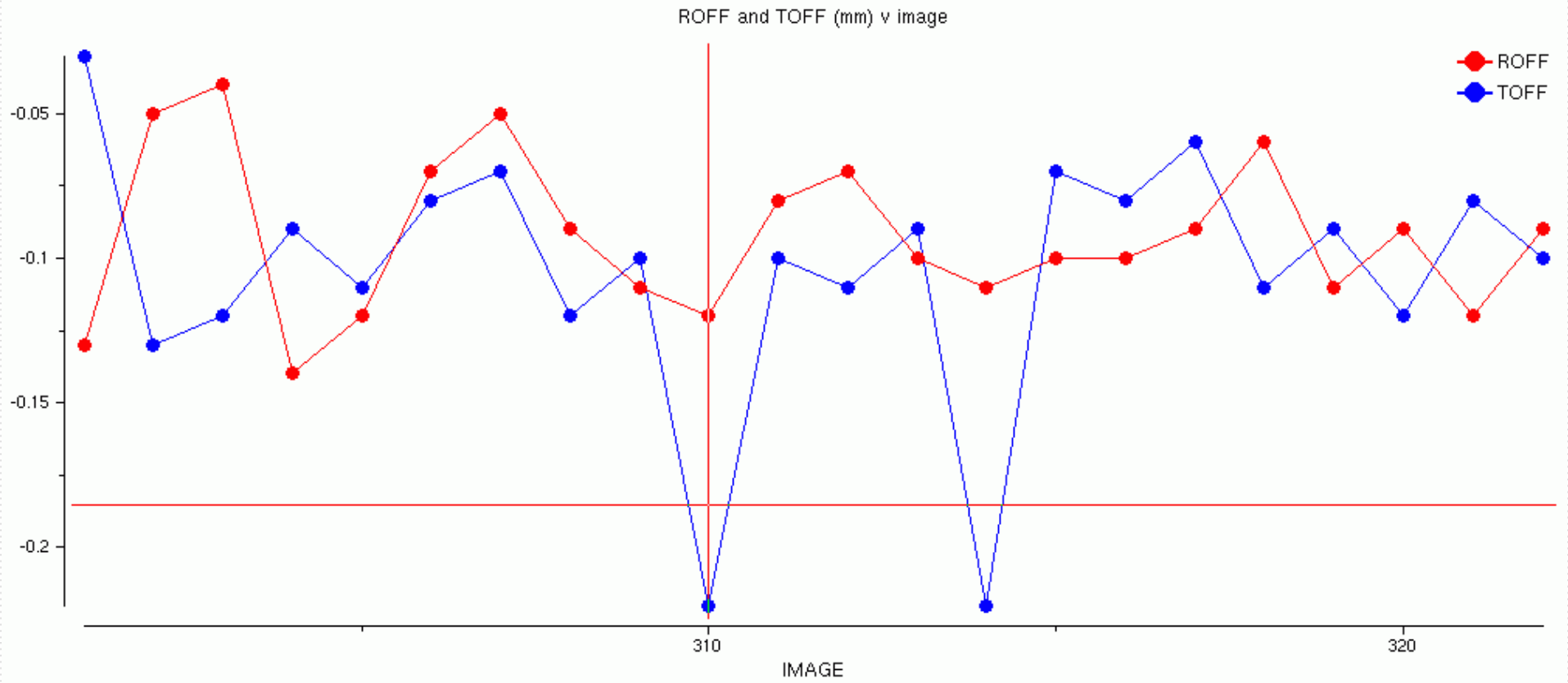
```
*** For information only. ***

 PARTIALS INCLUDED IN POSITIONAL REFINEMENT AND PROFILES
 ======================================================
 Because there were rather few fully recorded reflections...
<snip>
 *** Warning messages ***

 TANGENTIAL OFFSET UNSTABLE
 ==========================
 The tangential offset parameter (TOFF) is varying more...
<snip>
 SPOT OVERLAP
 ============
 Adjacent spots overlap. This will produce systematic errors...
<snip>
 EXCESSIVE NUMBER OF BADSPOTS
 ============================
 At least some images have rather a lot of badspots...
<snip>
 TOO MANY BACKGROUND PIXELS OVERLAPPED BY NEIGHBOURING SPOTS
 ===========================================================
 For some of the standard profiles, more than half the backgr...
```

*** Warning messages ***

TANGENTIAL OFFSET UNSTABLE
==========
The tangential offset parameter (TOFF) is varying more than it should. (Maximum variation is 0.15 mm)
If there are large changes in both TOFF and ROFF or CCOMEGA, this suggests that the refinement is unstable.
In this case, it is best to fix the TOFF parameter:
REFINEMENT FIX TOFF
If known the correct value can be input: eg
DISTORTION TOFF 0.17 If not known, the mean refined value can be used. In such cases ROFF should also be FIXED.

Loggraph SUMMARY

File Appearance Edit Utilities Help



309.9,-0.18

Tables in File

- Refined detector parameters
- Post refinement

Graphs in Selected Table

- Camera constants CCX CCY (mm) CCOM (deg) v image
- Distance (mm) v image
- Yscale v image
- Tilt and twist (hundredths of deg) v image
- ROFF and TOFF (mm) v image

SPOT OVERLAP

==========

<snip>

The minimum allowed spot separation (SEPARATION keyword) was 1.6 1.6mm. The actual spot size determined by the mask optimisation is 1.7 by 1.4 mm in the centre of the image and the largest spot size is 2.8 by 2.8 mm.

The separation given should be at least as large as the spot size in the centre of the image (keyword SEPARATION).

Check standard profiles carefully to ensure that the optimisation of the raster parameters has worked correctly.

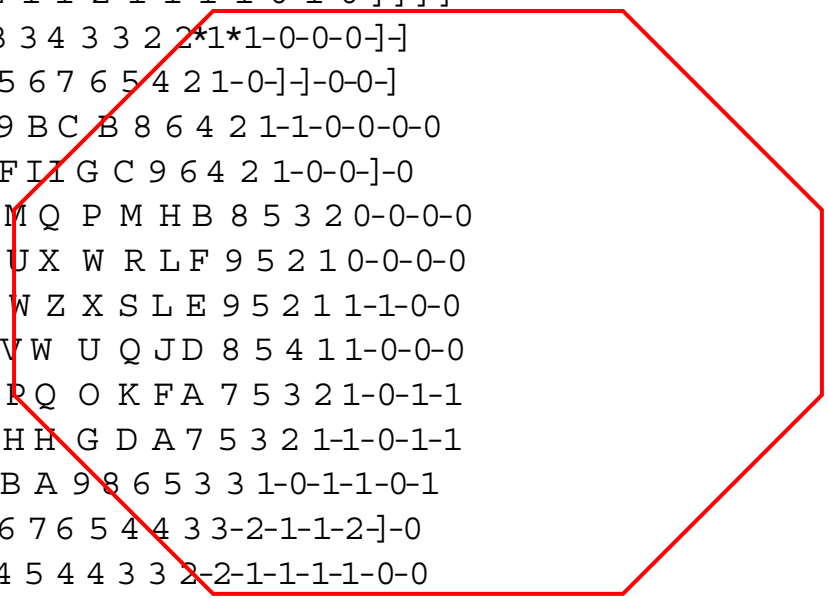
The effective size of the spots can be controlled by PROFILE TOLERANCE keywords. If the peak regions look too large (ie they include too much of the tails of the spot), try increasing TOLERANCE (current value 0.010) by eg 0.005 and see if profiles look better. (Increasing TOLERANCE will decrease spot size).

As a last resort the profile optimisation can be turned off using keywords PROFILE NOOPT.

In cases of serious overlap, (ie if the pattern is very dense), then the SEPARATION CLOSE option should be used (eg SEPARATION 1.0 1.0 CLOSE) and it may also help to suppress profile optimisation in these cases (PROFILE NOOPT) keyword.

See help library for details.

```
 **** WARNING **** Peaks of neighbouring reflections overlap, data
 quality will be impaired
<snip>
Profile for box  1
 X limits   0 to  92  mm, Y limits   0 to  92 mm
 Number of reflections in profile  299  RMSBG   9.5 Profile factor 0.50
           -0-]-0-0*0*]*0*0*0*0*]*0*]*0*0*0*0*]*]*]*]*]*]*]-]
           -0-0-0-0*0*0*0*0*0*0*0*0*1*0*]*0*0*]*]*]*]*]*]*]-]
           -0-]-]-0*1*1*1*1*1*0*0*0*1*1*0*0*0*0*0*]*]*]*]-]-]
           -0-0-0-0-0*0*0*0*1*2*2*1*1*2*1*0*0*1*0*]*]*]-]-]-]
           -1-0-0-0-0-0*1*1*1*2*2*1*1*2*1*1*1*1*0*1*0-]-]-]-]
           -0-]-0-1-1-1-1 2 2  1 2 3 3 4  3 3 2 2*1*1-0-0-0-]-]
           -0-0-0-1-2-1 1 2 3  3 4 5 6 7  6 5 4  2 1-0-]-]-0-0-]
           -1-1-1-1-1 1 2 3 4  5 7 9 B C  B 8 6 4  2 1-1-0-0-0-0
           -2-1-1-1 1 2 2 4 5  8 B F I I  G C 9 6 4  2 1-0-0-]-0
           -1-1-1 1 1 2 3 4 7  B H M Q  P M H B 8  5 3 2 0-0-0-0
           -1-2-1 2 2 2 2 5 9  G N U X  W R L F 9  5 2 1 0-0-0-0
           -0-1-1 2 2 3 3 6 C  J Q W Z  X S L E 9  5 2 1 1-1-0-0
           -1-2-2 3 2 2 4 6 C  J P V W  U Q J D 8  5 4 1 1-0-0-0
           -1-2-2 2 2 2 4 6 A  G L P Q  O K F A 7  5 3 2 1-0-1-1
           -1-1-1-3 2 3 3 5 8  C F H H  G D A 7 5  3 2 1-1-0-1-1
           -1-1-1-2-2 2 3 4 6  8 A B A  9 8 6 5 3  3 1-0-1-1-0-1
           -0-0-0-1-1-2 2 3 4  5 6 6 7  6 5 4 4 3  3-2-1-1-2-]-0
           -]-1-1-1-2*2*2 3 3  4 4 4 5  4 4 3 3 2  2-1-1-1-1-0-0
           -]-]-1-2*2*2*2*2*3*4*3*3*3*3*3*2*2*1*2-2-1-1-1-1-0
           -]-]-0*1*2*2*1*2*2*2*3*2*2*2*2*2*3*3*1*2-1-]-0-0-1
           -]-]*]*1*2*2*2*2*1*2*2*3*2*2*1*2*2*2*1*1*0-1-0-0-1
           -]*]*]*2*2*2*1*1*2*2*2*2*2*1*1*1*1*2*0*0-1-0-0-0
           -]*]*]*]*1*1*2*1*1*1*2*2*2*1*1*2*2*1*0*0-1-0-]-0
```



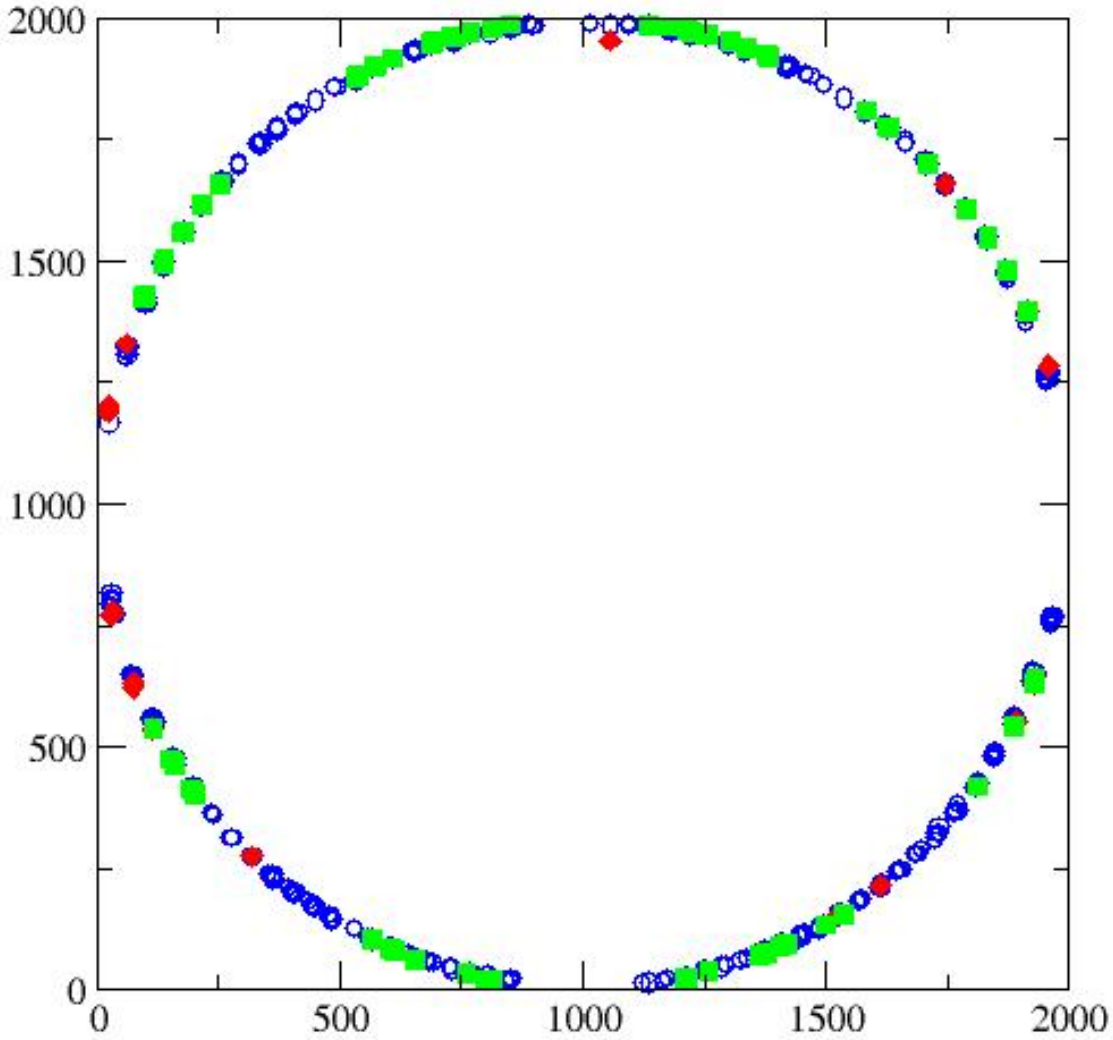
EXCESSIVE NUMBER OF BADSPOTS

At least some images have rather a lot of badspots (Maximum number 42). They are rejected on the basis of:

1) Poor profile fit (PKRATIO >3, controlled by REJECTION PKRATIO). *6*
2) Too large a BGRATIO (too much background variation, controlled by REJECTION BGRATIO).
3) Too large a background gradient (controlled by REJECTION GRADMAX) *578*
4) Intensity negative and more than 5 sigma. *38*

Look at the list of badspots to see what category they fall under.

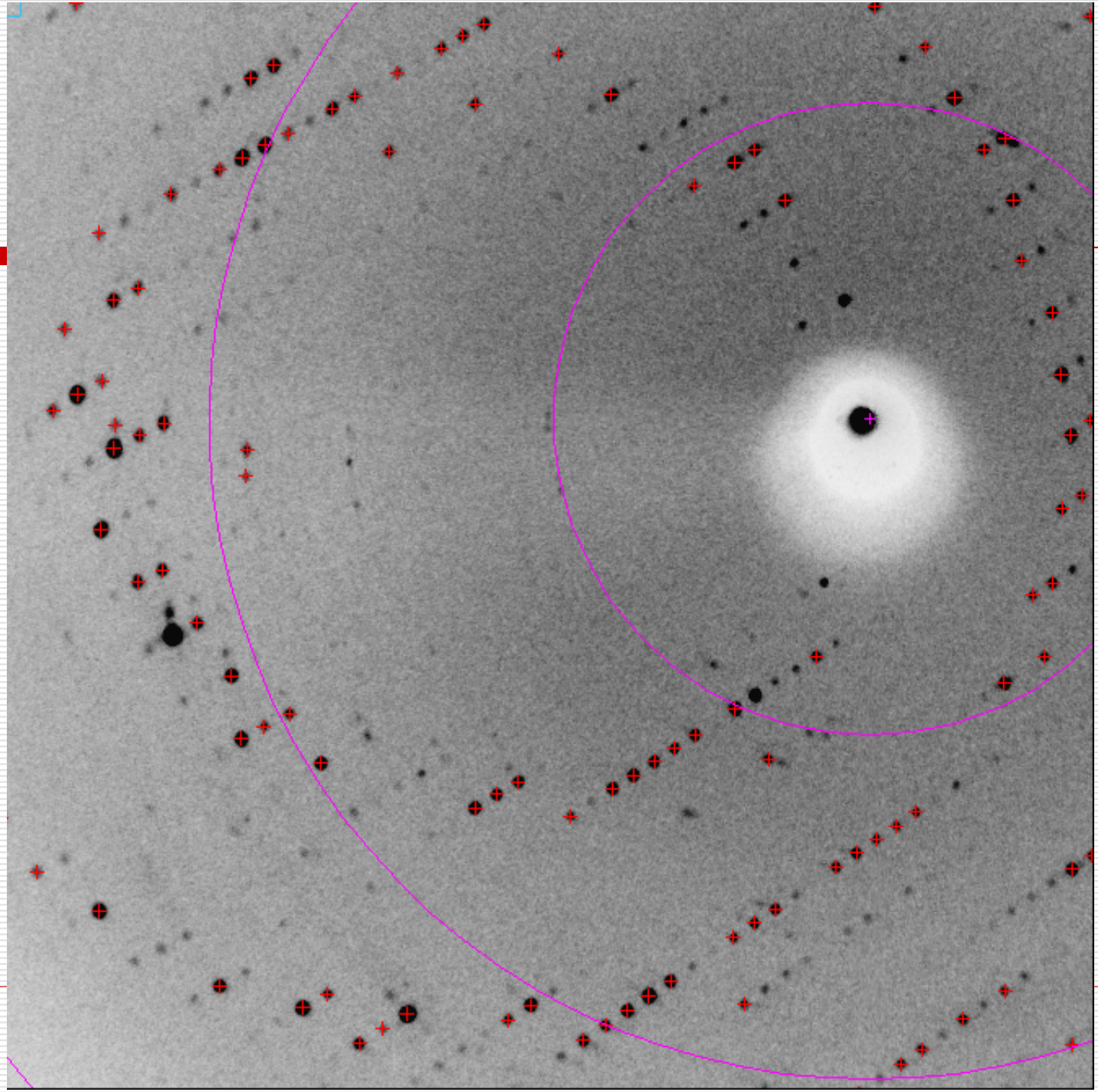
Poor profile fit is often the result of changes in ROFF, TOFF or CCOMEGA between successive images when using the ADDPART option.
Very intense images can have unusually large gradients, GRADMAX may have to be changed from the default
A pixel dump of the BADSPOTS can be obtained using REJECTION PLOT if the reason for their rejection is not clear


2000
1500
1000
500
0
0
500
1000
1500
2000

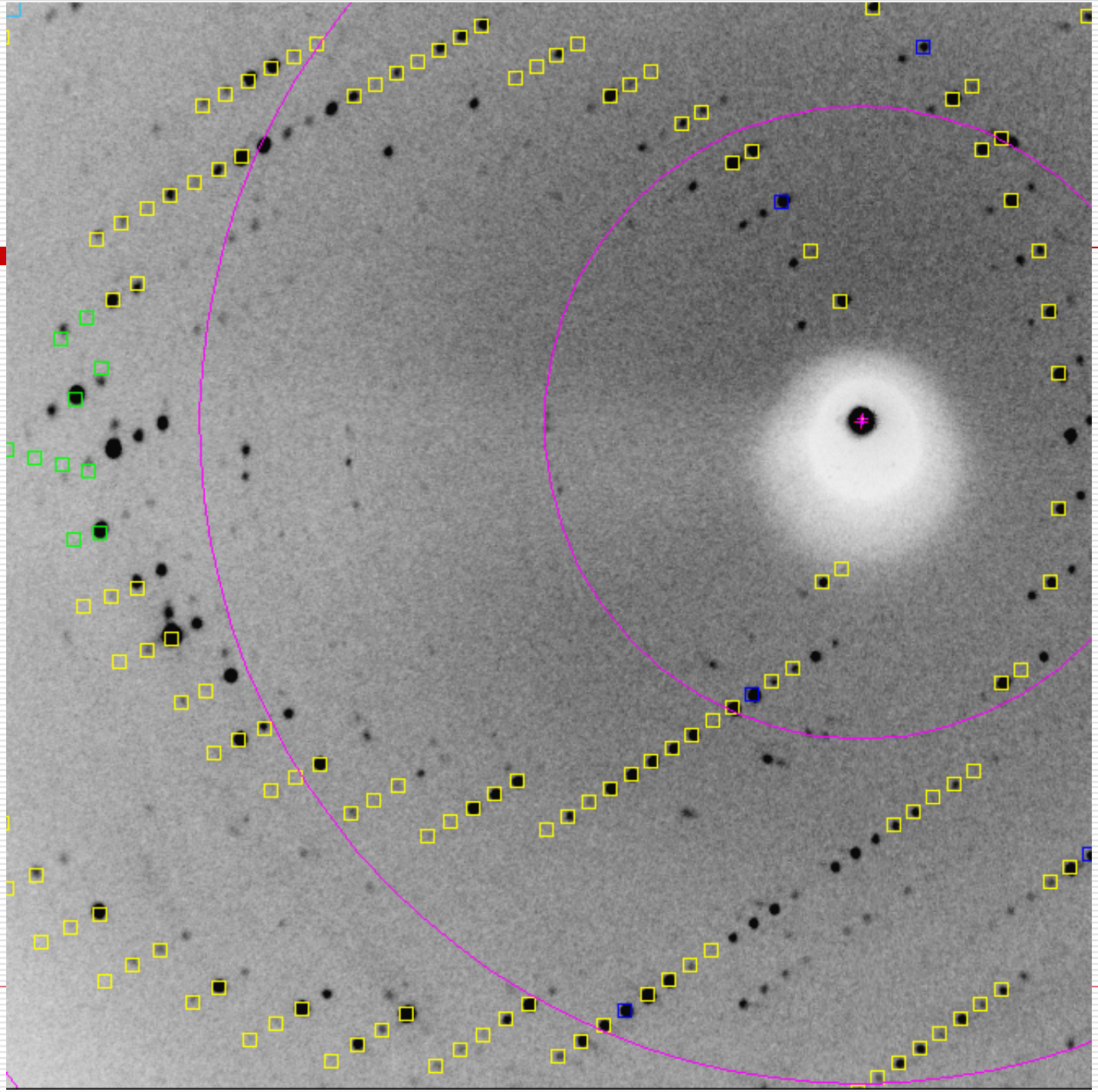
TOO MANY BACKGROUND PIXELS OVERLAPPED BY NEIGHBOURING SPOTS
==========
For some of the standard profiles, more than half the
background pixels are flagged as being overlapped by
neighbouring spots (in the worst case, 52.2 % are overlapped).
You should use the SEPARATION CLOSE keywords, eg
SEPARATION 1.0 1.0 CLOSE

More on Processing with Mosflm

- what do you do about the warnings?
- what if there are real problems?
- the new ccp4i Mosflm task



```
 12 103     mC    43.79   219.23    43.93    90.0  90.2  79.6   C2
 11 103     oC    43.79   219.23    43.93    90.0  90.2 100.4   C222,C2221
 10   7     tP    43.79    43.93   107.84    90.0  91.1  90.2   P4,P41,P42,P43,P422,P4212
  9   6     mC    61.93    62.13   107.84    89.2  90.8  89.8   C2
  8   6     mC    62.13    61.93   107.84    90.8  90.8  90.2   C2
  7   6     oP    43.79    43.93   107.84    90.0  91.1  90.2   P222,P2221,P21212,P212121
  6   6     mP    43.79   107.84    43.93    90.0  90.2  91.1   P2,P21
  5   6     oC    61.93    62.13   107.84    89.2  90.8  89.8   C222,C2221
  4   1     mP    43.79    43.93   107.84    90.0  91.1  90.2   P2,P21
  3   0     mP    43.79    43.93   107.84    90.0  91.1  90.2   P2,P21
  2   0     aP    43.79    43.93   107.84    90.0  88.9  89.8   P1
  1   0     aP    43.79    43.93   107.84    90.0  91.1  90.2   P1

Select a solution AND a spacegroup from list above (eg 3 p42) or 0 to abandon or T to cha

The solution and direct beam position will now be refined; reflections which deviate by m
than the sigma cutoff from their calculated position will be excluded from the refinement

Positional sigma cutoff [ 2.50]:
Refining solution #10 with P4     (number  75) symmetry imposed

Using  322 indexed reflections (out of  366 spots found, {delta(XY) <= 2.5 sigma}),
final sd in spot positions is 0.38mm and in phi 0.78 degrees
Refined cell parameters   44.03   44.03  107.87  90.00  90.00  90.00

Do you want to update cell parameters (Y):


Beam coordinates of 150.30 150.00 have been refined to 150.82 149.63

***** WARNING ***** WARNING ***** WARNING ***** WARNING ***** WARNING ***** WARNING

This is a shift of  0.64mm or  0.298 times the minimum spot separation of ca  2.14mm.
Do you want to accept the new direct beam position? (answer Y or N!) :y

Do you want to accept this solution (Y) :␣
```



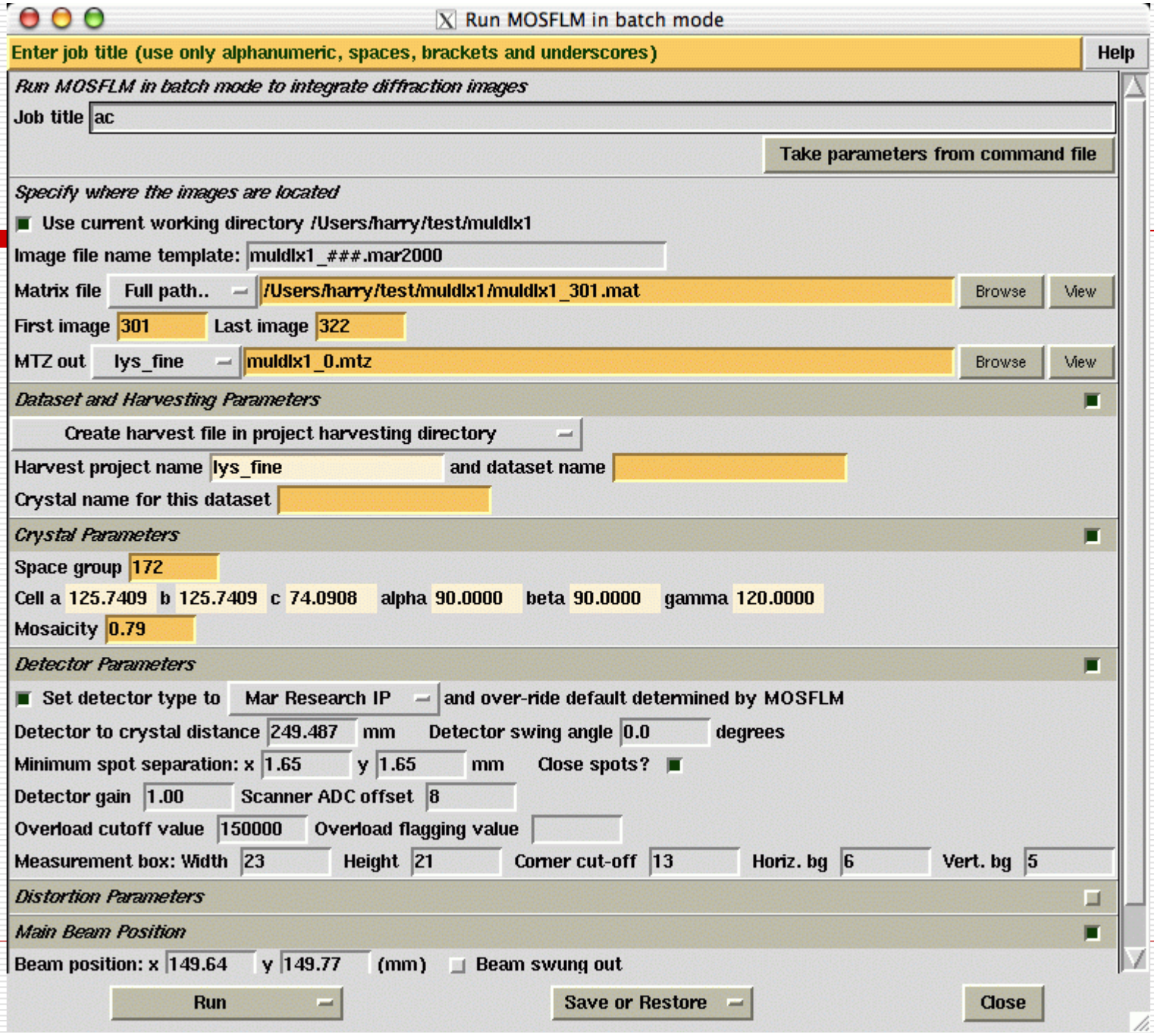
```
Cell refinement is complete
Starting cell   44.033   44.033  107.873   90.000   90.000   90.000
Refined cell    43.953   43.953  107.685   90.000   90.000   90.000

Rms positional error (mm) as a function of cycle for each image.
     Image   1     2     3    88    89    90
Cycle 1    0.160 0.168 0.181 0.738 0.663 2.608


YSCALE as a function of cycle for each image:
     Image   1     2     3    88    89    90
Cycle 1    1.000 1.002 1.000 0.966 0.964 0.970


Detector distance as a function of cycle for each image:
     Image   1     2     3    88    89    90
Cycle 1    149.2 149.1 150.1 151.3 151.0 151.6


Refined mosaic spread (excluding safety factor):   0.12



Missets for first image (   1) -0.03  0.02 -0.02
Missets for last image  (  90)  0.06 -0.27 -0.03

The current missets are for the last image to be processed.
If you want to integrate the data starting at the first image, you should
reset the misseting angles.

Reset missets to those of the first image ? (Y)
```

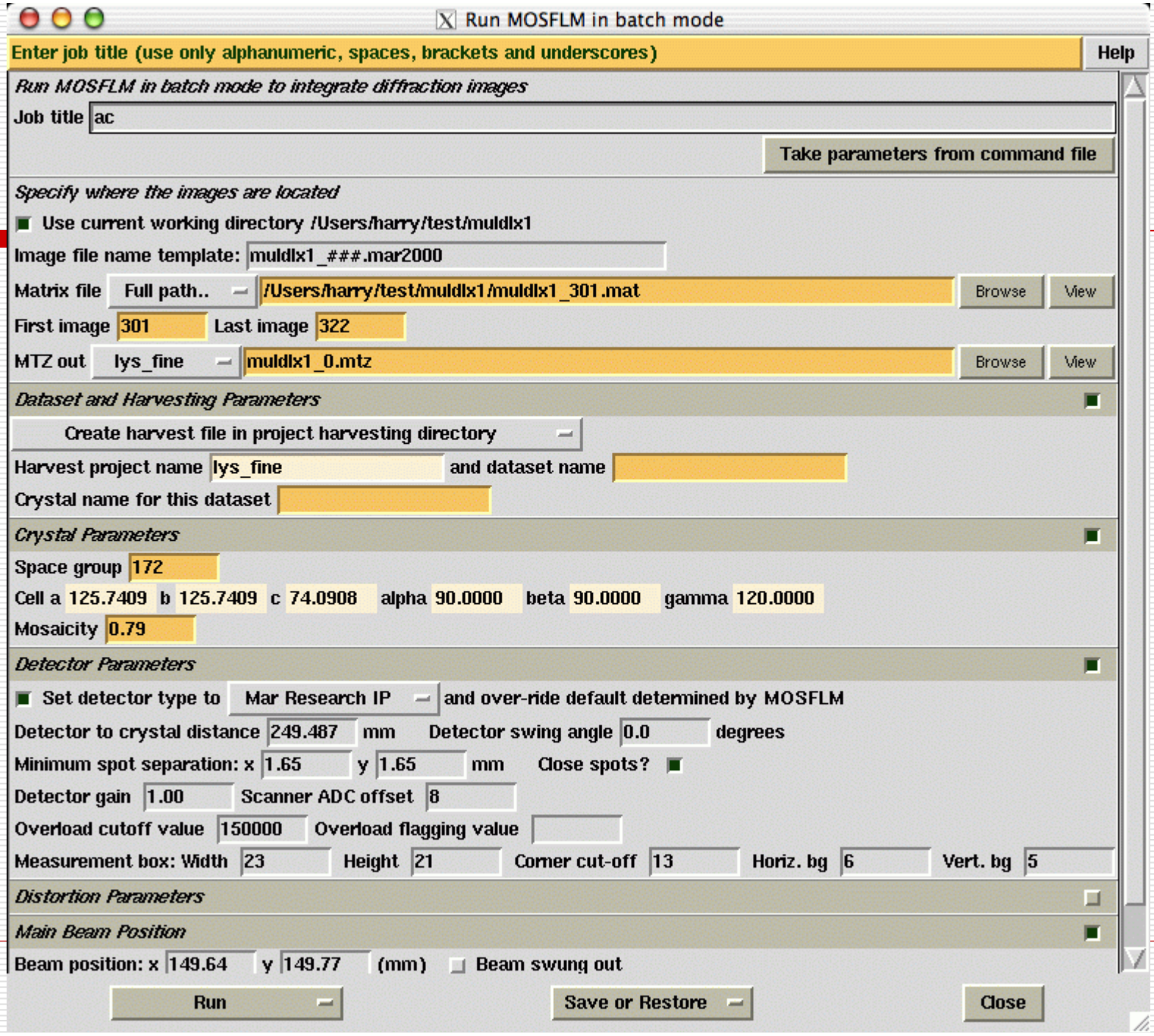
```
 12 103     mC    43.79   219.23    43.93    90.0   90.2   79.6   C2
 11 103     oC    43.79   219.23    43.93    90.0   90.2  100.4   C222,C2221
 10   7     tP    43.79    43.93   107.84    90.0   91.1   90.2   P4,P41,P42,P43,P422,P4212
  9   6     mC    61.93    62.13   107.84    89.2   90.8   89.8   C2
  8   6     mC    62.13    61.93   107.84    90.8   90.8   90.2   C2
  7   6     oP    43.79    43.93   107.84    90.0   91.1   90.2   P222,P2221,P21212,P212121
  6   6     mP    43.79   107.84    43.93    90.0   90.2   91.1   P2,P21
  5   6     oC    61.93    62.13   107.84    89.2   90.8   89.8   C222,C2221
  4   1     mP    43.79    43.93   107.84    90.0   91.1   90.2   P2,P21
  3   0     mP    43.79    43.93   107.84    90.0   91.1   90.2   P2,P21
  2   0     aP    43.79    43.93   107.84    90.0   88.9   89.8   P1
  1   0     aP    43.79    43.93   107.84    90.0   91.1   90.2   P1

Select a solution AND a spacegroup from list above (eg 3 p42) or 0 to abandon or T to cha

The solution and direct beam position will now be refined; reflections which deviate by m
than the sigma cutoff from their calculated position will be excluded from the refinement

Positional sigma cutoff [ 2.50]:
Refining solution # 7 with P222   (number  16) symmetry imposed

Using  322 indexed reflections (out of  366 spots found, {delta(XY) <= 2.5 sigma}),
final sd in spot positions is 0.22mm and in phi 0.64 degrees
Refined cell parameters   45.77   43.87  107.79  90.00  90.00  90.00

Do you want to update cell parameters (Y):


Beam coordinates of 150.30 150.00 have been refined to 150.61 149.60

This is a shift of  0.51mm or  0.100 times the minimum spot separation of ca  5.05mm.
Do you want to accept the new direct beam position? (Y) :

Do you want to accept this solution (Y) :
```

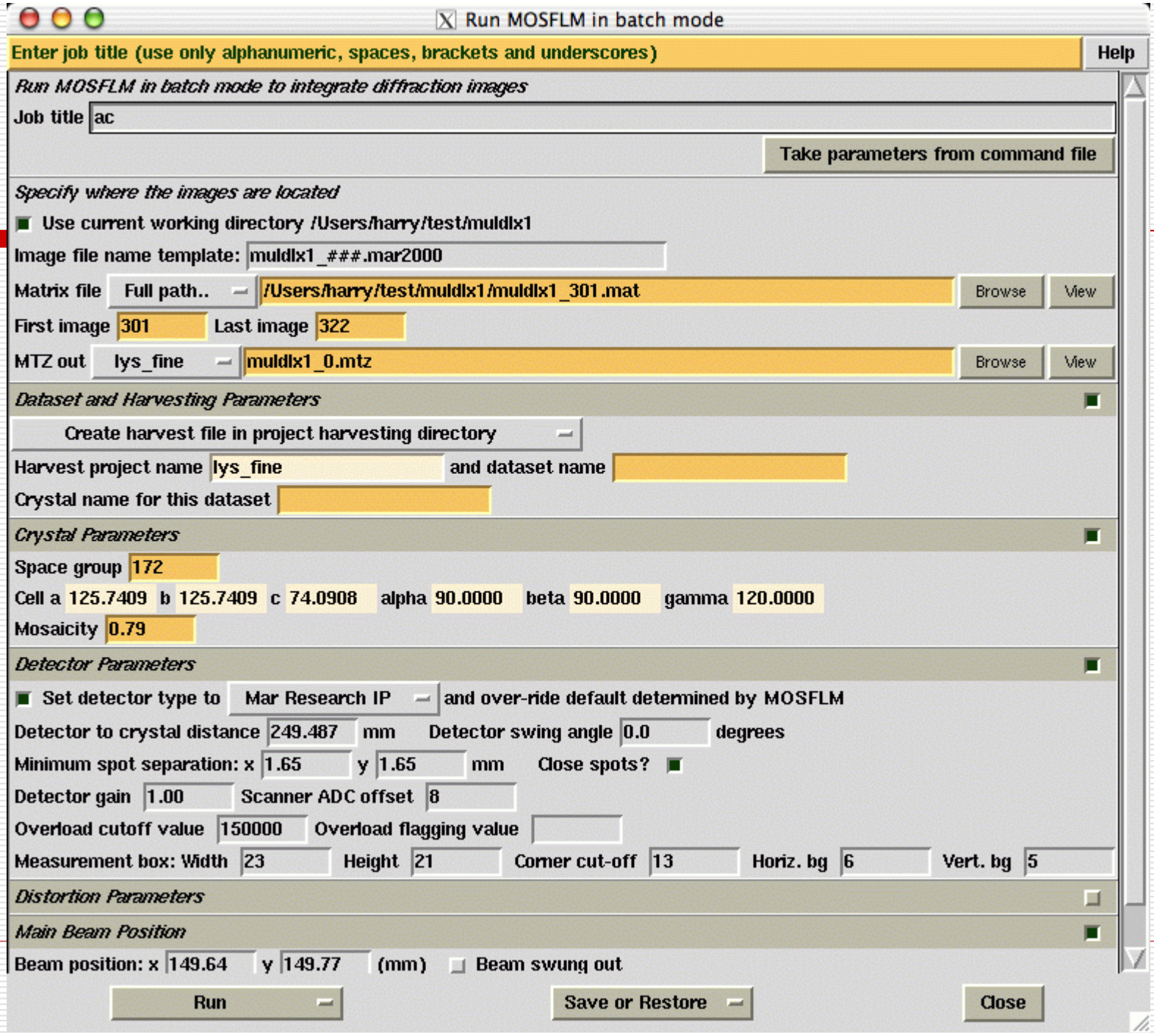
```
Cell refinement is complete
Starting cell   43.864   45.758  107.796   90.000   90.000   90.000
Refined cell    43.962   45.897  108.009   90.000   90.000   90.000

Rms positional error (mm) as a function of cycle for each image.
     Image   1     2     3    88    89    90
Cycle 1    0.066 0.062 0.061 0.256 0.305 0.190
Cycle 2    0.112 0.148 0.110 0.231 0.274 0.220


YSCALE as a function of cycle for each image:
     Image   1     2     3    88    89    90
Cycle 1    1.000 1.000 1.000 1.000 0.996 0.997
Cycle 2    1.000 0.999 1.000 0.997 0.999 0.999


Detector distance as a function of cycle for each image:
     Image   1     2     3    88    89    90
Cycle 1    150.0 150.0 150.1 149.8 150.4 150.1
Cycle 2    150.5 150.6 150.5 150.9 150.6 150.6


Refined mosaic spread (excluding safety factor):   1.60



Missets for first image (   1)  0.06 -0.01  0.15
Missets for last image  (  90)  0.17 -0.33  0.21

The current missets are for the last image to be processed.
If you want to integrate the data starting at the first image, you should
reset the misseting angles.

Reset missets to those of the first image ? (Y)␣
```

More on Processing with Mosflm

- what do you do about the warnings?
- what if there are real problems?
- the new ccp4i Mosflm task


Run MOSFLM in batch mode
Enter job title (use only alphanumeric, spaces, brackets and underscores)
Help
Run MOSFLM in batch mode to integrate diffraction images
Job title ac
Take parameters from command file
Specify where the images are located
Use current working directory /Users/harry/test/muldlx1
Image file name template: muldlx1_###.mar2000
Matrix file Full path.. /Users/harry/test/muldlx1/muldlx1_301.mat Browse View
First image 301 Last image 322
MTZ out lys_fine muldlx1_0.mtz Browse View
Dataset and Harvesting Parameters
Create harvest file in project harvesting directory
Harvest project name lys_fine and dataset name
Crystal name for this dataset
Crystal Parameters
Space group 172
Cell a 125.7409 b 125.7409 c 74.0908 alpha 90.0000 beta 90.0000 gamma 120.0000
Mosaicity 0.79
Detector Parameters
Set detector type to Mar Research IP and over-ride default determined by MOSFLM
Detector to crystal distance 249.487 mm Detector swing angle 0.0 degrees
Minimum spot separation: x 1.65 y 1.65 mm Close spots?
Detector gain 1.00 Scanner ADC offset 8
Overload cutoff value 150000 Overload flagging value
Measurement box: Width 23 Height 21 Corner cut-off 13 Horiz. bg 6 Vert. bg 5
Distortion Parameters
Main Beam Position
Beam position: x 149.64 y 149.77 (mm) Beam swung out
Run
Save or Restore
Close


Run MOSFLM in batch mode
Enter job title (use only alphanumeric, spaces, brackets and underscores)
Help
Run MOSFLM in batch mode to integrate diffraction images
Job title ac
Take parameters from command file
Specify where the images are located
Use current working directory /Users/harry/test/muldlx1
Image file name template: muldlx1_###.mar2000
Matrix file Full path.. /Users/harry/test/muldlx1/muldlx1_301.mat Browse View
First image 301 Last image 322
MTZ out lys_fine muldlx1_0.mtz Browse View
Dataset and Harvesting Parameters
Create harvest file in project harvesting directory
Harvest project name lys_fine and dataset name
Crystal name for this dataset
Crystal Parameters
Space group 172
Cell a 125.7409 b 125.7409 c 74.0908 alpha 90.0000 beta 90.0000 gamma 120.0000
Mosaicity 0.79
Detector Parameters
Set detector type to Mar Research IP and over-ride default determined by MOSFLM
Detector to crystal distance 249.487 mm Detector swing angle 0.0 degrees
Minimum spot separation: x 1.65 y 1.65 mm Close spots?
Detector gain 1.00 Scanner ADC offset 8
Overload cutoff value 150000 Overload flagging value
Measurement box: Width 23 Height 21 Corner cut-off 13 Horiz. bg 6 Vert. bg 5
Distortion Parameters
Main Beam Position
Beam position: x 149.64 y 149.77 (mm) Beam swung out
Run
Save or Restore
Close


Run MOSFLM in batch mode
Enter job title (use only alphanumeric, spaces, brackets and underscores)
Help
Run MOSFLM in batch mode to integrate diffraction images
Job title ac
Take parameters from command file
Specify where the images are located
Use current working directory /Users/harry/test/muldlx1
Image file name template: muldlx1_###.mar2000
Matrix file Full path.. /Users/harry/test/muldlx1/muldlx1_301.mat Browse View
First image 301 Last image 322
MTZ out lys_fine muldlx1_0.mtz Browse View
Dataset and Harvesting Parameters
Create harvest file in project harvesting directory
Harvest project name lys_fine and dataset name
Crystal name for this dataset
Crystal Parameters
Space group 172
Cell a 125.7409 b 125.7409 c 74.0908 alpha 90.0000 beta 90.0000 gamma 120.0000
Mosaicity 0.79
Detector Parameters
Set detector type to Mar Research IP and over-ride default determined by MOSFLM
Detector to crystal distance 249.487 mm Detector swing angle 0.0 degrees
Minimum spot separation: x 1.65 y 1.65 mm Close spots?
Detector gain 1.00 Scanner ADC offset 8
Overload cutoff value 150000 Overload flagging value
Measurement box: Width 23 Height 21 Corner cut-off 13 Horiz. bg 6 Vert. bg 5
Distortion Parameters
Main Beam Position
Beam position: x 149.64 y 149.77 (mm) Beam swung out
Run
Save or Restore
Close

Data Collection & Data Analysis

Scaling and merging with Scala

Gwyndaf Evans
Diamond Light Source, UK.

Contents

- sources of systematic errors
- parameterization of scaling
- estimates of data quality
- estimation of standard errors
- outlier rejection

Introduction

- scaling and merging is the most important diagnostic step in terms of data quality.
- it is important that it be performed as soon as possible after data collection and preferably during.
- in many cases it is a straightforward procedure but can become complicated. An understanding of the underlying principles is important.

Steps in scaling

- Choose scaling model
  - Should reflect the experiment
    - X-ray source intensity variation
    - Changes in diffracting volume
    - Sample or air absorption
    - Radiation damage
- Analyze results
  - Should the sample be discarded?
  - What is the useful resolution?
  - Are there outliers or bad images?
  - What is the spacegroup?

Factors affecting scale

- Incident beam
  - intensity
  - size
  - primary beam absorption
- Sample
  - diffracting volume
  - diffracted beam absorption
- Detector
  - calibration
  - time stability
  - bad pixels
- Miscellaneous
  - beam stop and cryo-stream shadows

Incident beam related factors

- Synchrotron
  - smooth decay of beam intensity
  - any discontinuities (e.g. beam injection) should be noted and included in scaling model
  - illuminated volume
  - shutter synchronization/goniometer rotation speed

Crystal related factors

- Sample absorption
  - diffracted beam absorption (shape dependent)
  - important for weak anomalous signal
- Radiation damage
  - can be significant on high brilliance sources
  - difficult to correct for
  - modeled as change in relative B-factor
  - extrapolation to zero dose

Detector related factors

- calibration errors
  - spatial distortion
  - non-uniformity of response
  - time stability
  - bad pixels

Miscellaneous factors

- unavoidable
  - zingers
- avoidable
  - beam stop shadow
  - cryo-stream shadow
  - should be dealt with at integration stage

Determination of scale factors

What information do we have?

Scales are determined by comparison of symmetry-related reflections, i.e. by adjusting scale factors to get the best internal consistency of intensities. Note that we do not know the true intensities and an internally-consistent dataset is not necessarily correct. Systematic errors will remain

$$\text{Minimize } \Delta\Phi = \Sigma_{hl}\, w_{hl}\, (I_{hl} - 1/k_{hl}\langle I_h\rangle)^2$$

I_{hl} l'th intensity observation of reflection **h**

k_{hl} scale factor for I_{hl} $\langle I_h\rangle$ current estimate of I_h

$g_{hl} = 1/k_{hl}$ is a function of the parameters of the scaling model

g_{hl} = g(□ rotation/image number) . g(time) . g(**s**)

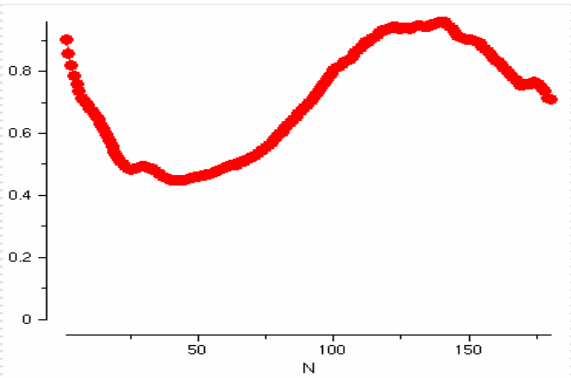
...*other factors*

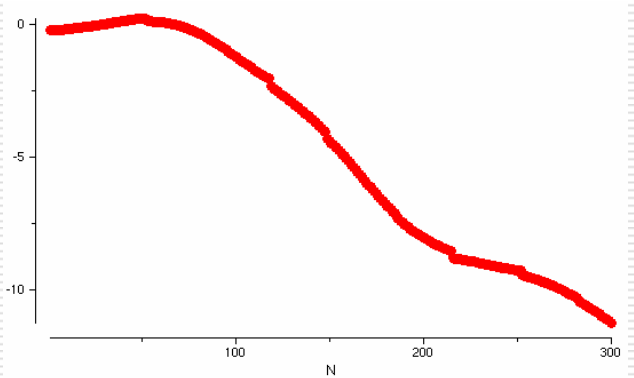
Primary beam $\mathbf{s_0}$ *B-factor Absorption*

Parametrization

g_{hl} = g(rotation/image number) . g(*time*) . g(**s**) ...*other factors*

Primary beam $\mathbf{s_0}$ *B-factor* *Absorption* *eg "tails"*





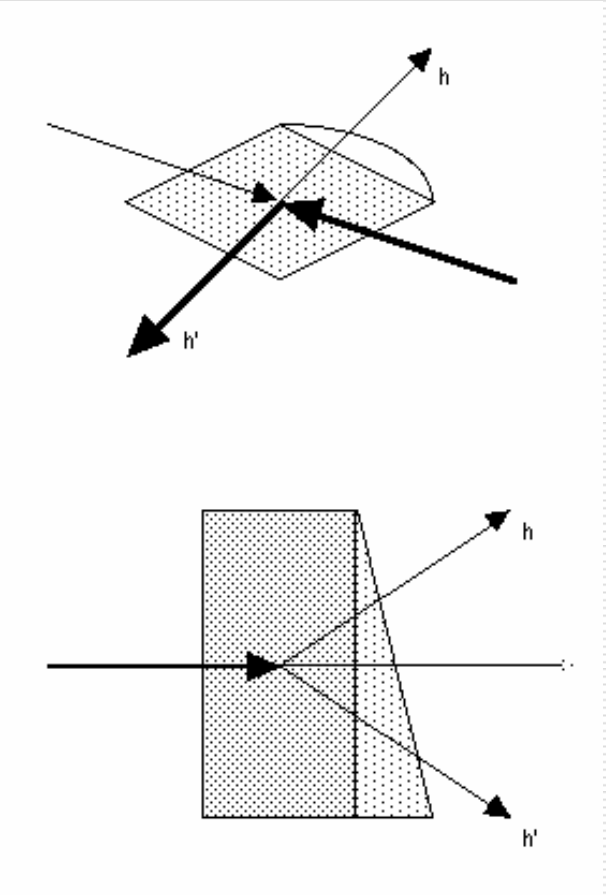
scale is smooth function of spindle rotation *or* discontinuous function of image (batch) number (usually less appropriate)

$g(time) = \exp[+2B(time) \sin^2\theta/\lambda^2]$

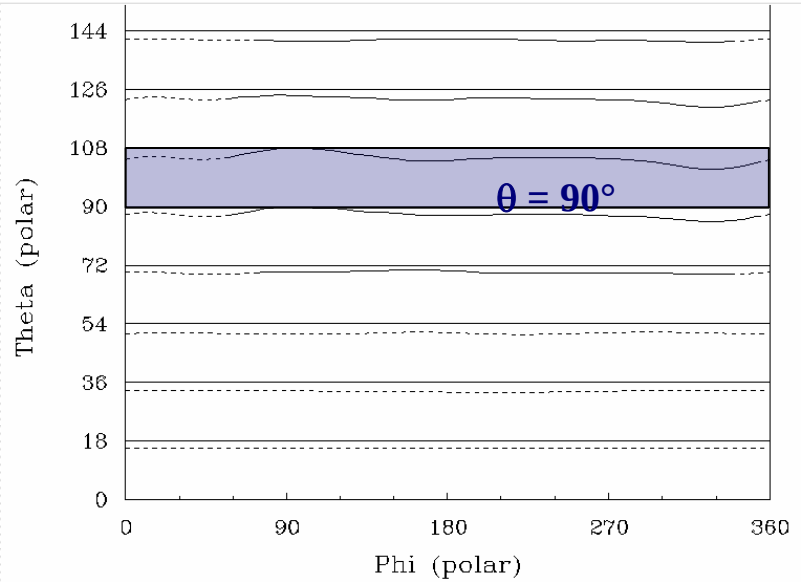
essentially a time-dependent radiation damage correction

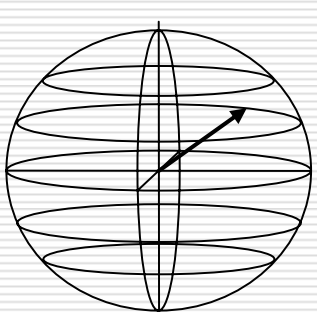
Absorption correction

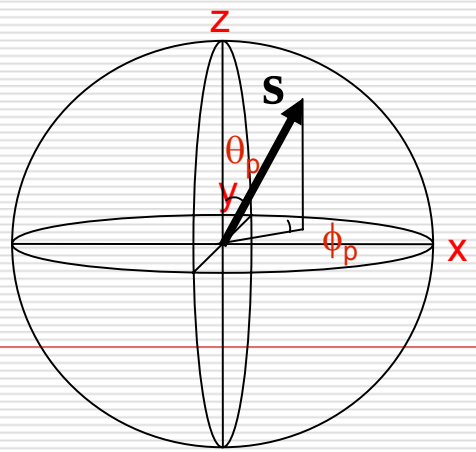
- serious for longer wavelengths and larger irregular samples
- we typically do not have enough data to find true absorption corrections
- good corrections require data recorded with sample in multiple orientations
- some strategies can minimize effect of absorption on quality of observed anomalous differences
- inverse beam
- simultaneous measurement of I+ & I-



Spherical harmonics





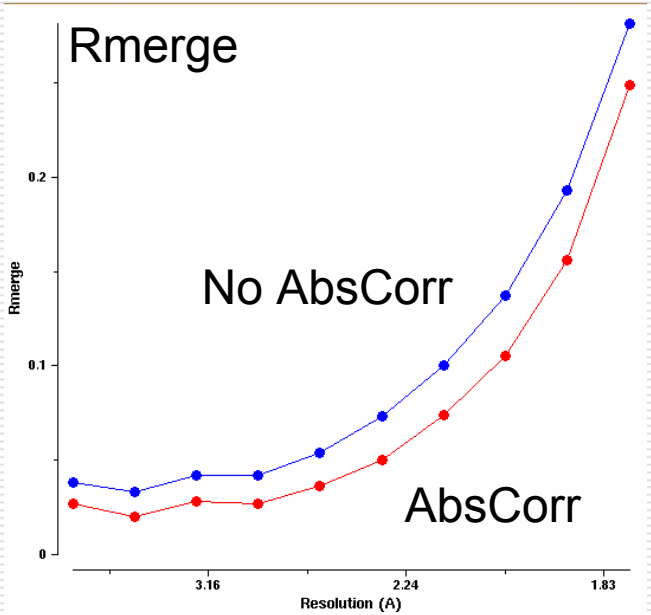


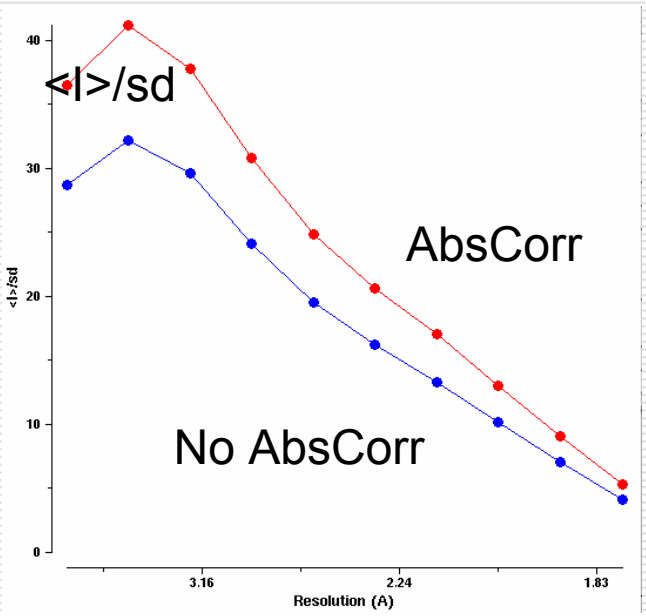
$$A(\mathbf{s}) = \Sigma_{lm}\, c_{lm} Y_{lm}(\theta_p, \phi_p)$$

- linear coefficients clm determined as parameters.
- Note the surface is not centrosymmetric (see e.g. equator $\theta = 90°$)
- i.e. different corrections are applied to I+ & I-
- Graph along lines of latitude
- Dashed lines where there is no data

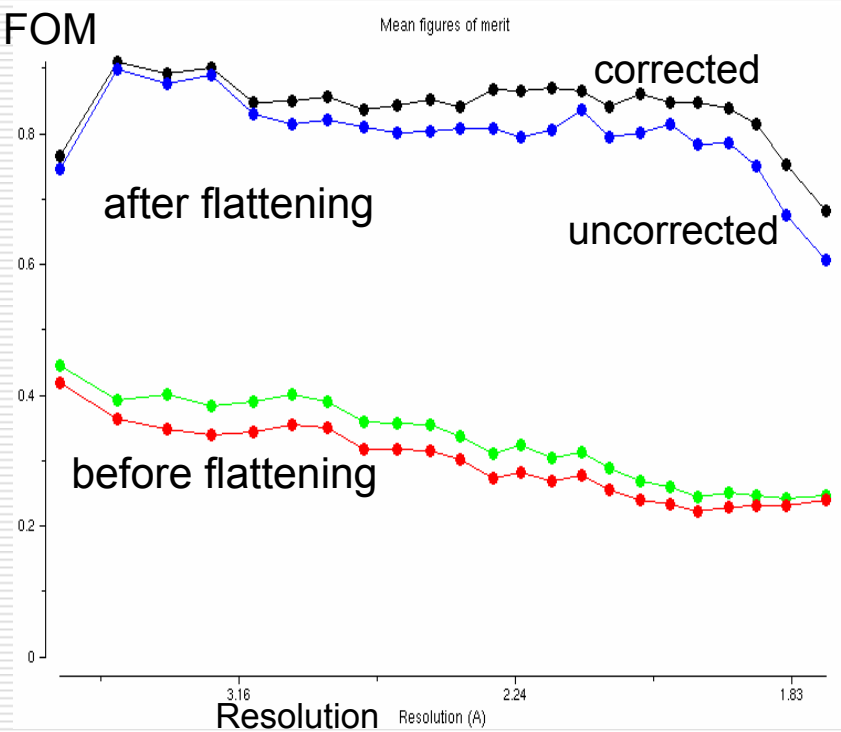
Absorption: sample dataset

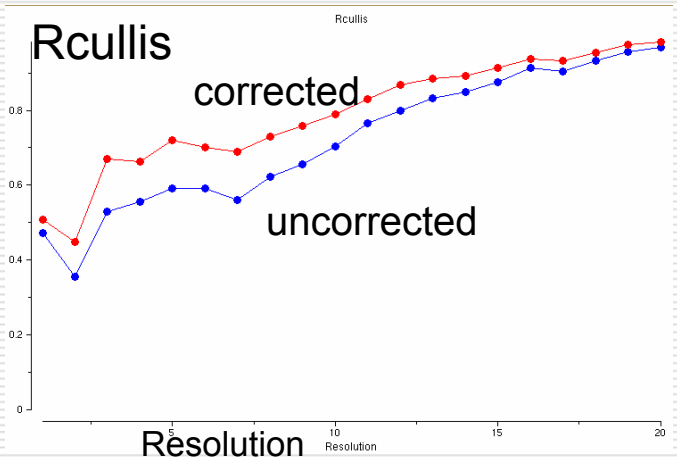
- Rotating anode (RU200, Osmic mirrors, Mar345)
- 100 images, 1°, 5 min/°, resolution 1.8Å
- Scaled with and without secondary beam correction (sub-keyword: absorption)
- Secondary beam correction (absorption) improves the data

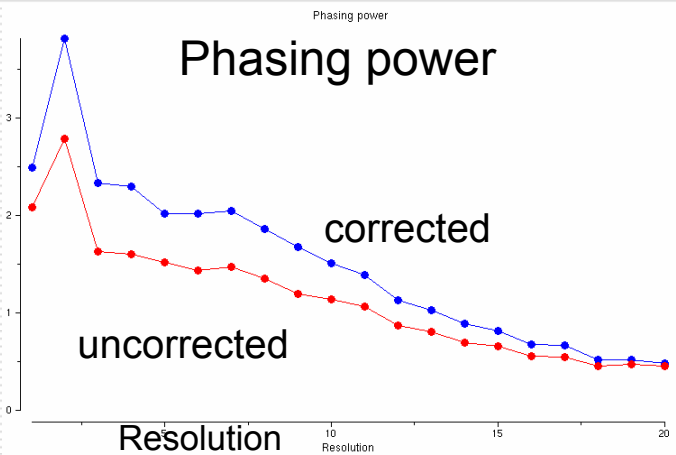




Absorption correction and phasing







- Phasing (SHARP) is better with absorption correction even after solvent flattening

SCALA 'scales' options

- Typical scaling options are

```
scales rotation spacing 5 absorption 6  bfactor on brotation spacing 20

tie surface 0.005          # restrain secondary beam surface
                           # maybe relax (eg 0.05) for high absorption
link surface all           # same surface for all runs
tie bfactor 0.3            # stabilise B-factors
```

- When to use other options (more complicated cases):-
  - data collected from more than one crystal
    - use different absorption surfaces
  - very low resolution - don’t use B-factor

How well are the scales determined?

- Note that determination of scaling parameters depends on symmetry-related observations having **different** scales. If all observations of a reflection have the same value of the scale component, then there is no information about that component and it remain as a systematic error in the merged data (this may well be the case for absorption for instance)
- Thus to get intensities with the lowest absolute error, the symmetry-related observations should be measured in as different way as possible (eg rotation about multiple axes). This will increase Rmerge, but improve the estimate of <I>.
- Conversely, to measure the most accurate differences for phasing (anomalous or dispersive), observations should be measured in as similar way as possible

Scaling datasets together

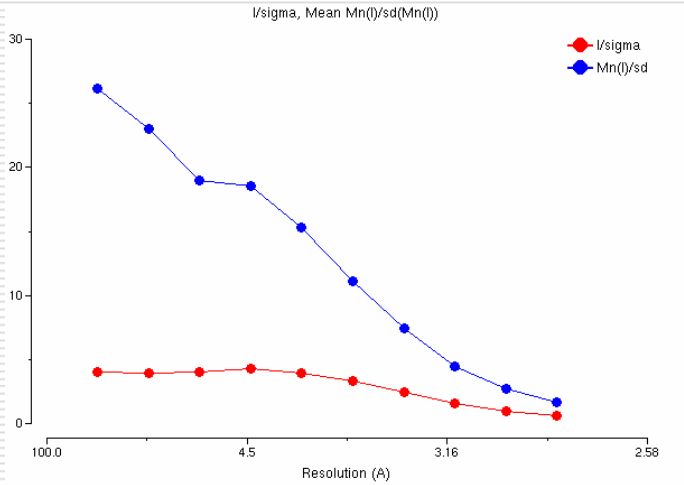
- For multiple-wavelength datasets, it is best to scale all wavelengths together simultaneously. This is then a ***local*** scaling to minimize the difference between datasets, reducing the systematic error in the anomalous and dispersive differences which are used for phasing
- Other advantages of simultaneous scaling:-
- rejection of outliers with much higher reliability because of higher multiplicity
- correlations between ΔF_{anom} and ΔF_{disp} indicate the reliability of the phasing signal
- approximate determination of relative f" and relative f' values
- *In Scala, this is now automatic if multiple datasets are present in the input file*

Results of scaling: R-factors

- (a) Rmerge (Rsym) = Σ | Ihl - <Ih> | / Σ | <Ih> |
- This is the traditional measure of agreement, but it increases with higher multiplicity even though the merged data is better
- (b) Rmeas = Rr.i.m.= Σ (n/n-1) | Ihl - <Ih> | / Σ | <Ih> |
- The multiplicity-weight R-factor allows for the improvement in data with higher multiplicity. This is particularly useful when comparing different possible point-groups
- Diederichs & Karplus, *NSB.*, **4**, 269-275 (1997)

Results of scaling: Intensities and standard deviations

- Scala compares the estimated standard deviation ♦(I) to the observed scatter, and tries to correct ♦(I) by a multiplication factor. This is done using a *normal probability plot*. A correction as a function of intensity is also done, but this is not yet automatic
- $\sigma(I)' = Sdfac * Sqrt [\sigma^2(I) + (Sdadd * I)^2]$
- The corrected $\sigma(I)$ is compared with the intensities: the most useful statistic is $< <I>/ \sigma(<I>) >$ (labelled Mn(I)/sd in table)

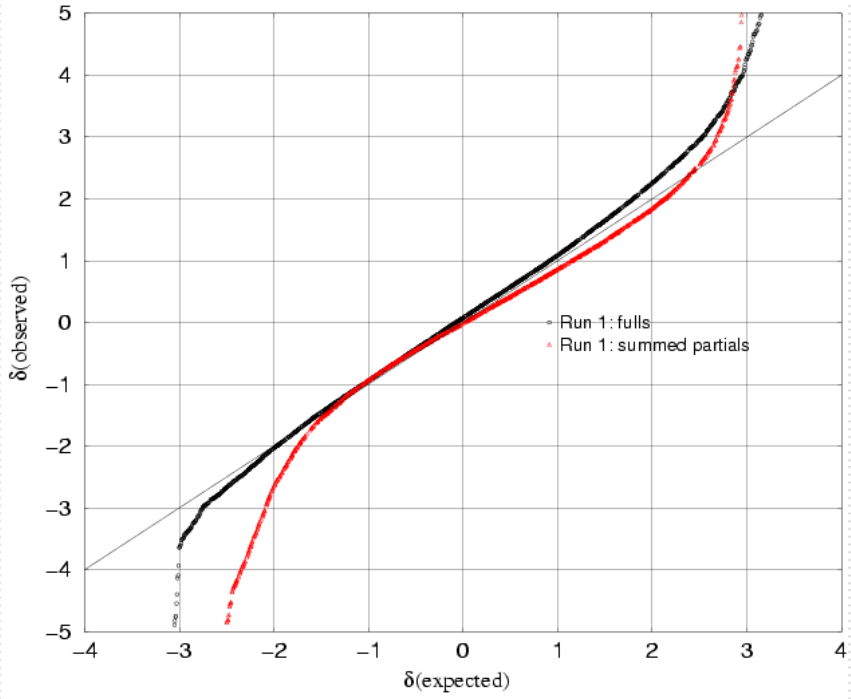


$< <I>/ \sigma(<I>) >$ .gt. ~ 2

Estimation of errors

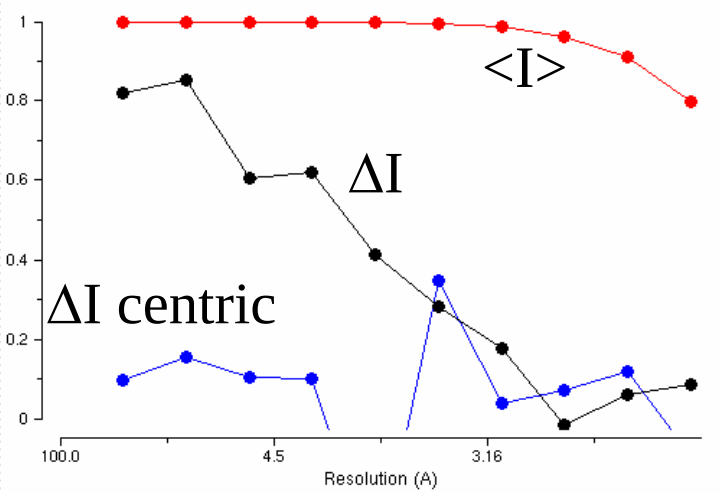
- keyword for estimation of standard errors
  - **sdcorrection sdfac 1.5 sdadd 0.015**

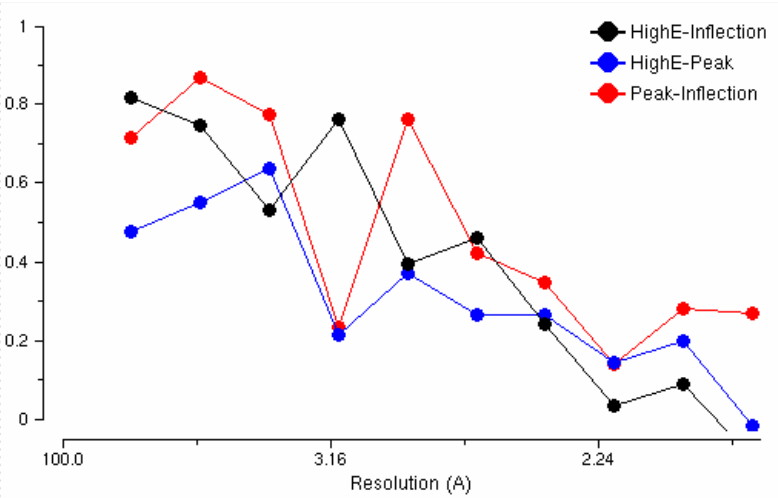
Normal probability plots



Results of scaling: Correlation coefficients

- CC's between different estimates of anomalous or dispersive differences (DI) indicate reliability of the phasing signal



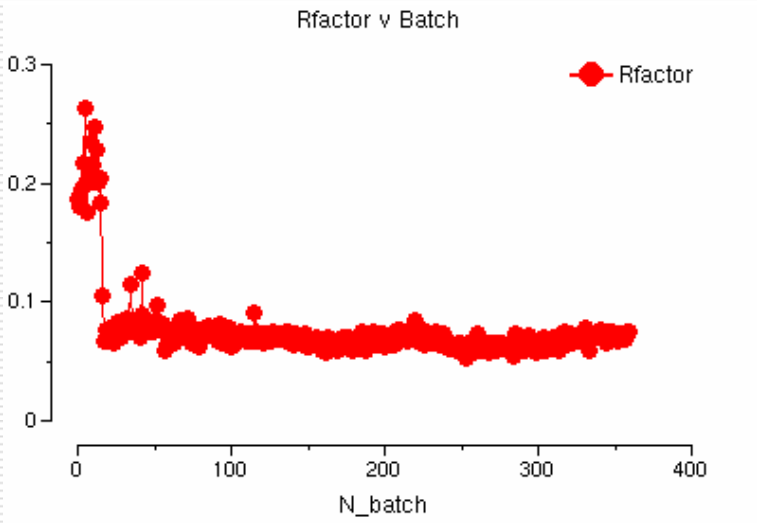


Correlations between half-sets (data randomly halved)

Correlations between (I^+ - I^-) for MAD data

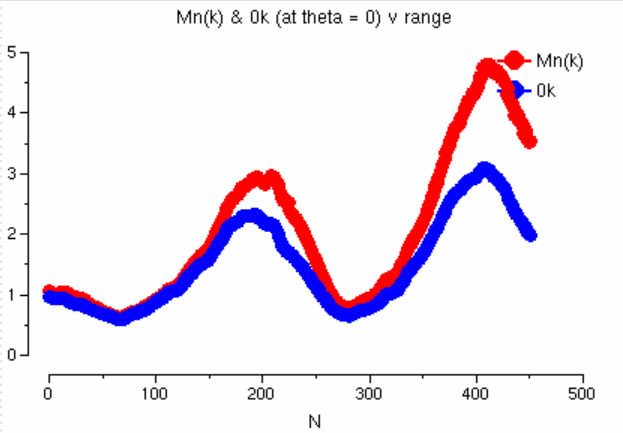
Results of scaling: Common sense checks

- **Are some parts of the data bad?**
- Analysis of Rmerge against batch number gives a very clear indication of problems local to some regions of the data. Perhaps something has gone wrong with the integration step, or there are some bad images

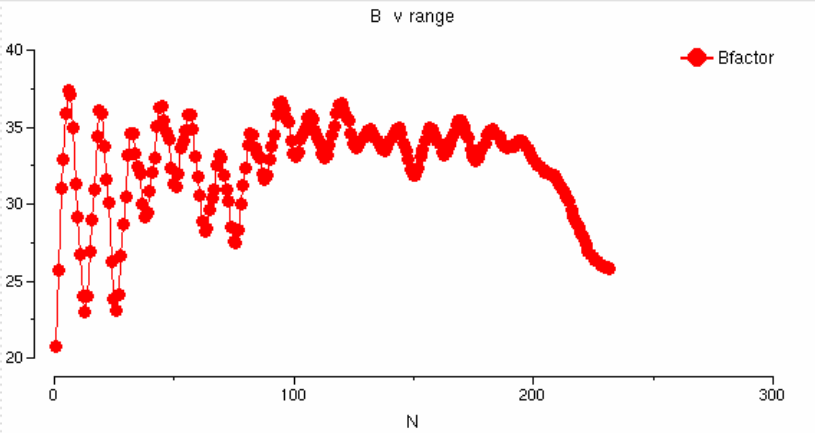


- Here the beginning of the dataset is wrong due to problems in integration (e.g. poor orientation matrix in MOSFLM at start of job.)

Results of scaling: Physical sense checks

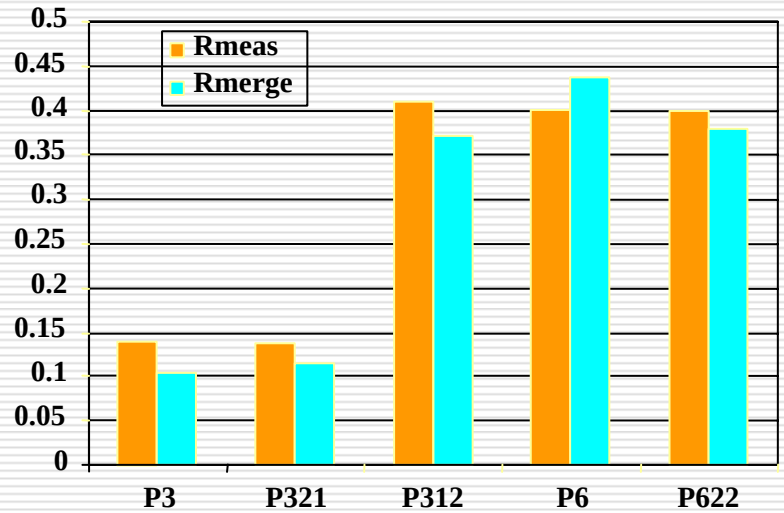


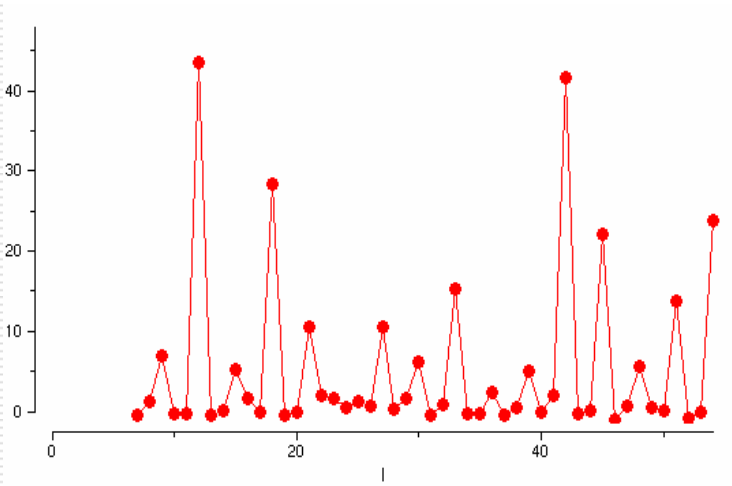
- These scale factors follow a reasonable absorption curve



- These B-factors are not sensible (fluctuations may be damped by "tie bfactor")

Spacegroup determination





- Lattice type: hexagonal (from indexing)
- Merge in different Laue groups
- Point-group P321 is highest symmetry with good merging (Rmeas)

- Systematic absences
  - l = 3n
- Spacegroup = P3121
  - could be P3221

Outliers

- Some (a small proportion) of observations may be rejected as being improbable
- Unreasonably large
  - default limit in Scala E >10
  - removes **some** ice|salt spots & zingers
- Unreasonable disagreement with other observations
  - easier with high multiplicity
  - not clear what to do between *e.g.* I+ & I- where there is real difference

Outliers: possible causes

- outside reliable area of detector (*e.g.* behind shadow)
  - specify backstop shadow, calibrate detector
- ice spots
  - do not get ice on your crystal!
- zingers
- bad prediction (spot not there)
  - improve prediction
- spot overlap
  - lower mosaicity, smaller slice, move detector back
  - deconvolute overlaps
- multiple lattices
  - find single crystal

Outliers: detection

- Detection of outliers is easiest if the multiplicity is high
- Removal of spots behind the backstop shadow does not work well at present: usually it rejects all the good ones, so **tell Mosflm where the backstop shadow is**
- Inspect the ROGUES file to see what is being rejected (at least occasionally)

```
The ROGUES file contains all rejected reflections (flag "*", "@" for I+- rejects, "#" for
Emax rejects)
    TotFrc = total fraction, fulls (f) or partials (p)
    Flag I+ or I- for Bijvoet classes
    DelI/sd = (Ihl - Mn(I)others)/sqrt[sd(Ihl)**2 + sd(Mn(I))**2]
   h   k   l     h   k   l  Batch      I  sigI    E  TotFrc Flag Scale   LP   DelI/sd d(A)
Xdet   Ydet    Phi
   (measured)     (unique)

  -2  -2   0     2   2   0   1220  24941  2756  1.03  0.95p  I-  2.434  0.031   -1.1 30.40
1263.7 1103.2  210.8
  -4   2   0     2   2   0   1146   9400  2101  0.63  0.99p *I+  3.017  0.032   -6.7 30.40
1266.4 1123.3  151.3
   4  -2   0     2   2   0   1148  27521  2972  1.08  1.09p  I-  2.882  0.032    0.0 30.40
1058.8 1130.0  153.2
   2  -4   0     2   2   0   1075  29967  2865  1.13  0.92p  I+  2.706  0.032    1.1 30.40
1060.9 1106.6   94.4
```

Acknowledgments

mosflm Andrew Leslie and Harry Powel

scala Phil Evans