X-ray diffraction set-up & data processing

Maurizio Polentarutti

Elettra, XRD1 beamline

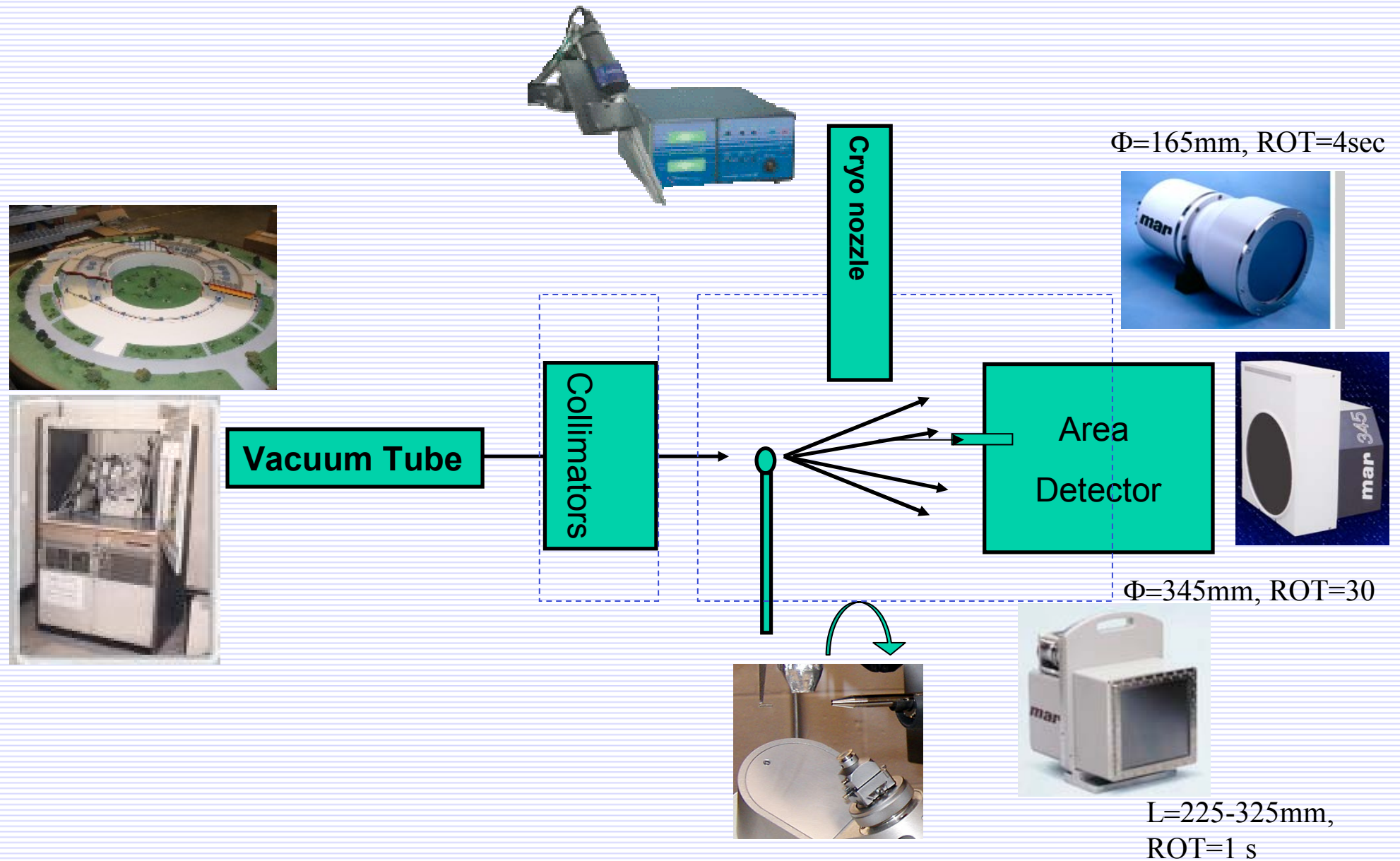# Part 1: X-ray (*protein crystallography*) diffraction set-up

- Common to all the PX beamlines and *home* labs

- Main components (in order of appearance):

  - *Slits (beam shapers)*
  - *Sample (protein single crystal)*
  - *Sample cooler system*
  - *Sample manipulator system (horizontal spindle axis)*
  - *Fluorescence detection system*
  - *(Primary)-Beam stopper*
  - *Detector*
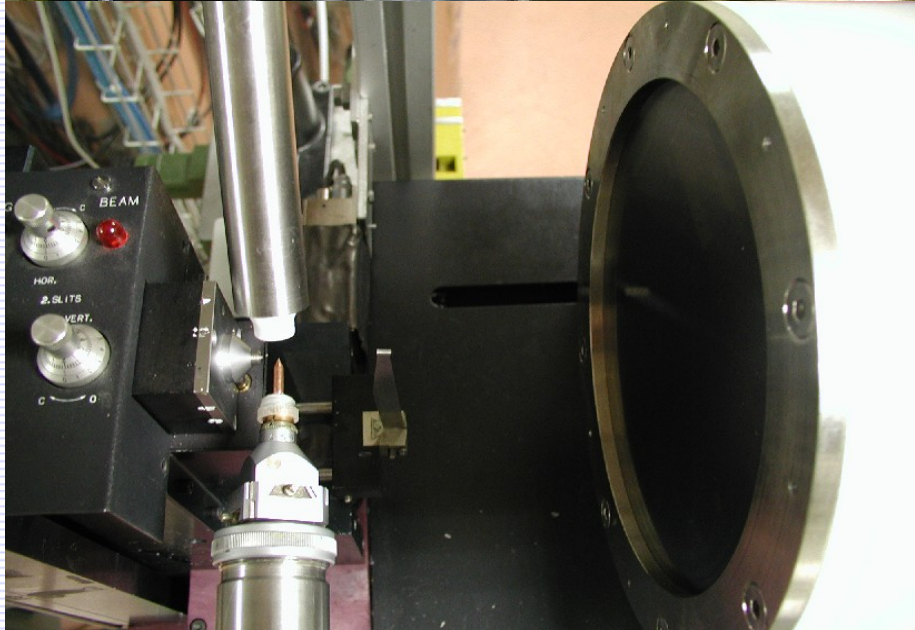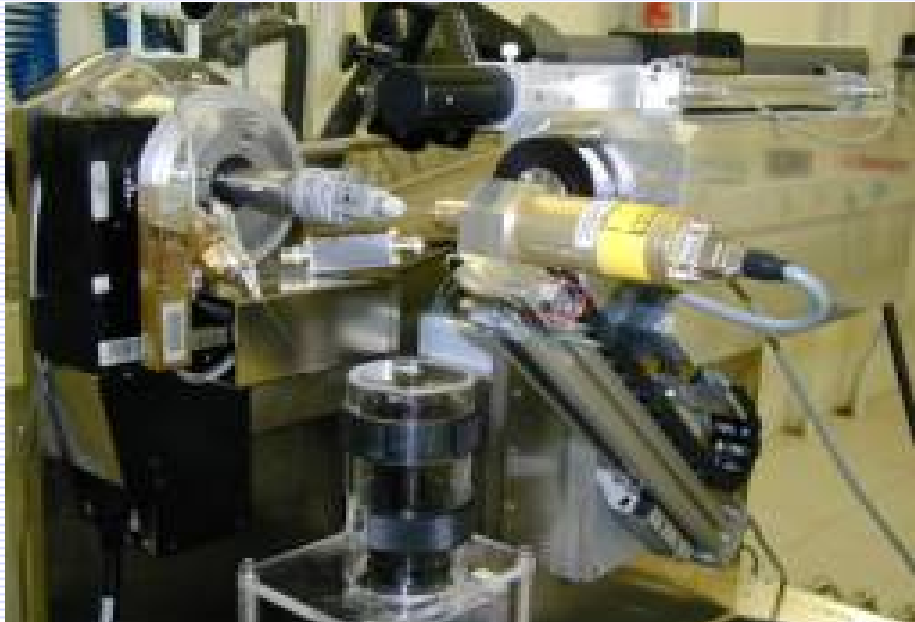
- Experimental key parameters

  Sample-to-detector distance, wavelength, detector surface, beam stopper position, sample external and unit-cell dimensions, sample orientation, detector angular position.

- Small differences….

Part 1: X-ray (*protein crystallography*) diffraction set-up

Cryo nozzle

Φ=165mm, ROT=4sec

Collimators

Vacuum Tube

Area Detector

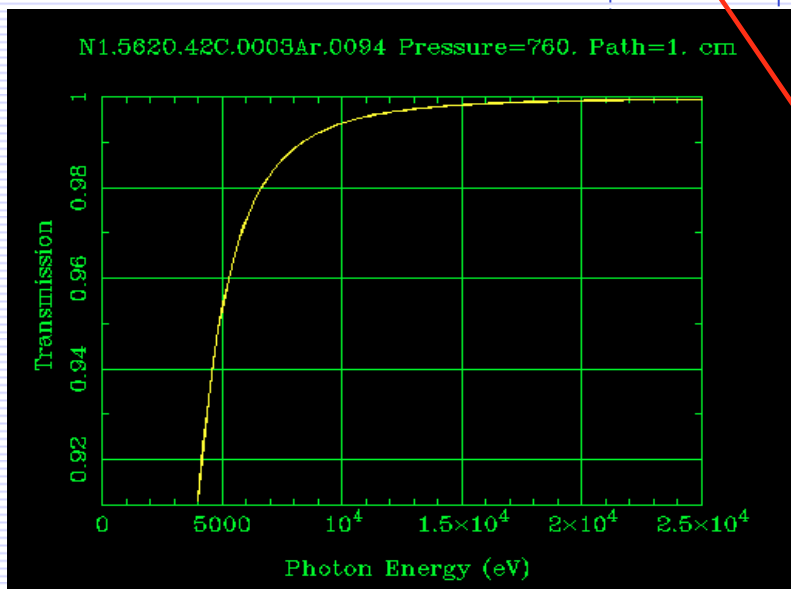Φ=345mm, ROT=30

L=225-325mm, ROT=1 s

Part 1: X-ray (*protein crystallography*) diffraction set-up

Last shutter to expose the sample to the beam: how long?

Beam size should be comparable to sample size

Which should be the stopper position?

Cryo nozzle

Collimators

Vacuum Tube

Detector

Detector

N1.5620.42C.0003Ar.0094 Pressure=760. Path=1. cm

Transmission

0.92   0.94   0.96   0.98   1

0   5000 10^4 1.5×10^4 2×10^4 2.5×10^4

Photon Energy (eV)

How many spots there will be on the images?

Absorption and scattering!

How much should it rotate?

Part 1: X-ray (*protein crystallography*) diffraction set-up

Why do we want to know the diffracted-beams intensities?

1. The electron density equation:

$$\rho(x, y, z) = \frac{1}{V_c} \sum_h \sum_k \sum_l \left| F(h, k, l) \right| \cdot e^{-2\pi i(hx + ky + lz) + i\alpha(h, k, l)}$$

2. The integrated intensity for the reflection (h k l):

$$I(int, \ h \ k \ l) = \frac{\lambda^3}{\omega \cdot V_{cell}^2} \times \left(\frac{e^2}{mc^2} \right)^2 \times V_{cr} \times I_0 \times L \times P \times A \times \left| F(hkl) \right|^2$$

Part 1: X-ray (*protein crystallography*) diffraction set-up
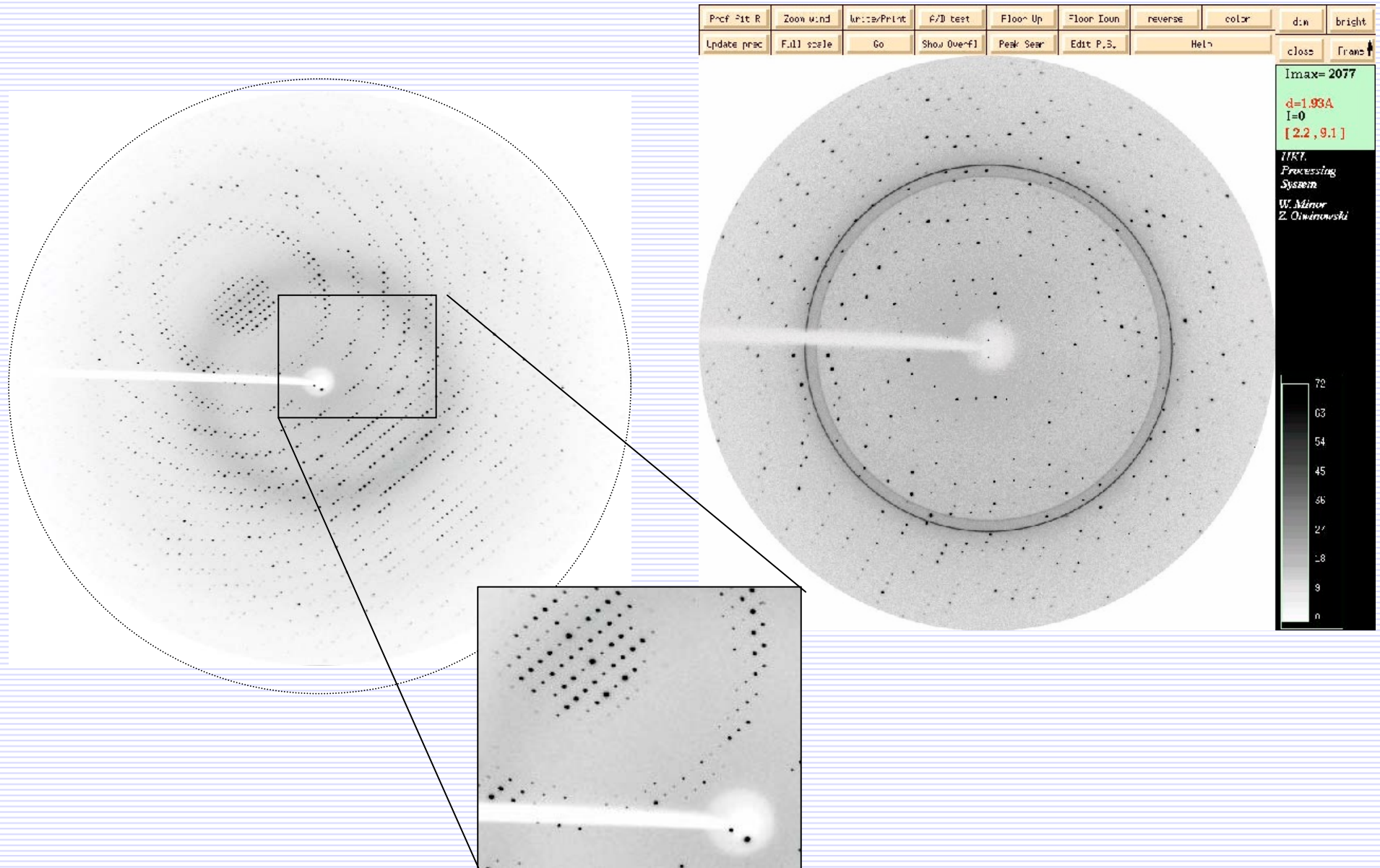
Part 1: X-ray (*protein crystallography*) diffraction set-up
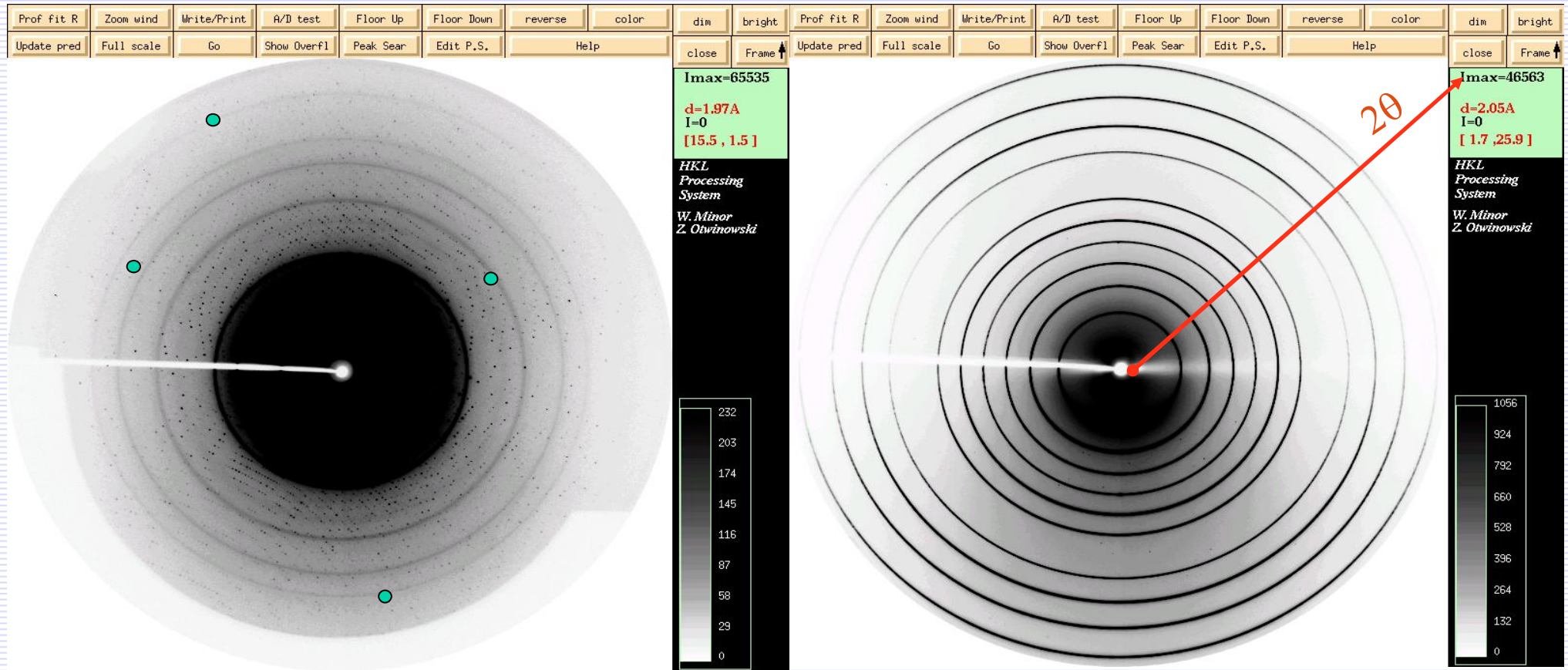
Protein (and bad cryosolution)
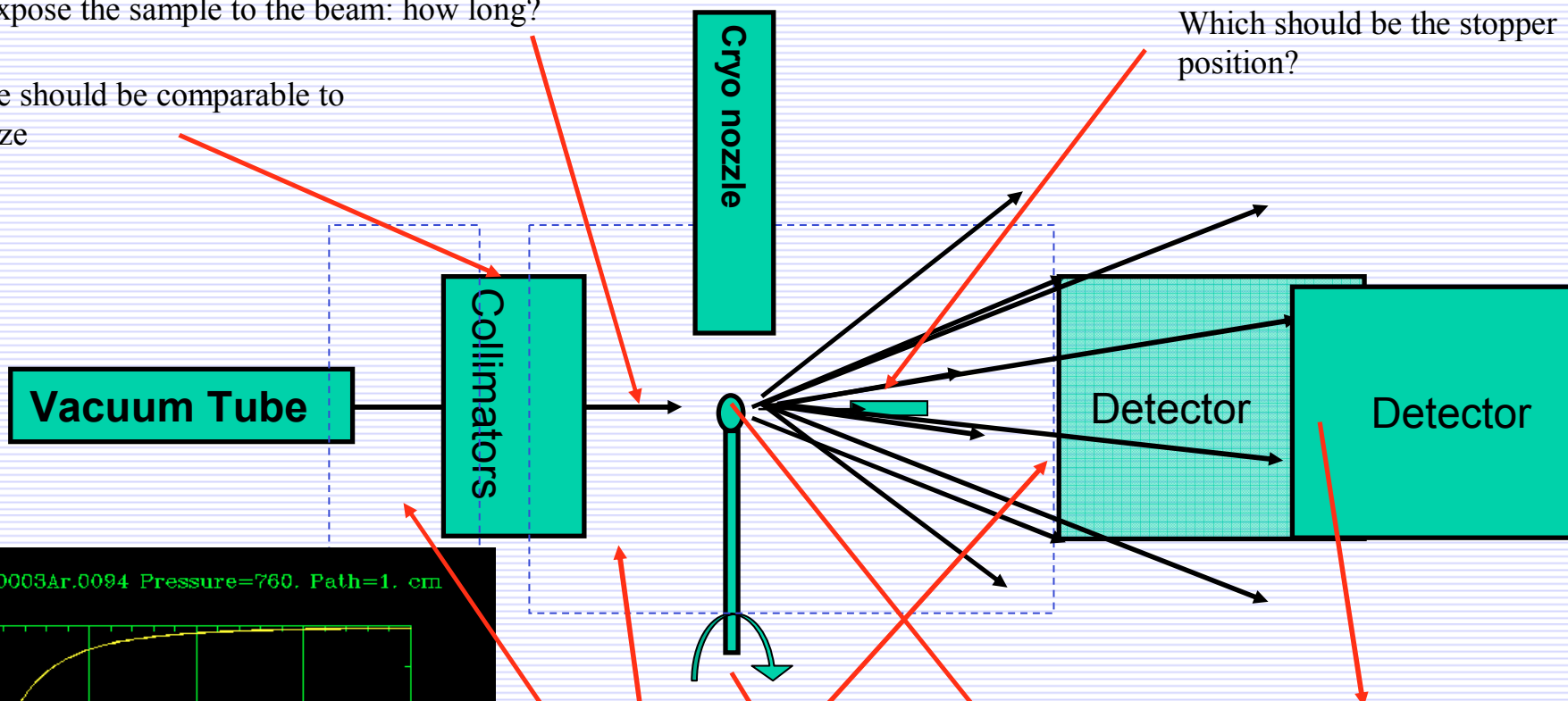
LaB₆ – policristalline powder in capillary (not well-centered)

How the integration works:

If the members of a set of reciprocal-lattice planes perpendicular to a chosen direction t are well separated, then the projections of the reciprocal-lattice vectors onto **t** will have an easily recognizable periodic distribution.

About 7300 separate roughly equally spaced directions. For each direction **t**, the distribution of the corresponding F(k) coeficients is surveyed to locate the largest local maximum. The and ' values associated with the 30 largest maxima are selected for refinement by a local search procedure to obtain an accuracy of 10-4 rad (0.006   ).

Directions are chosen from these vectors to give a linarly independent set of three basis vectors of a primitive real-space unit cell. These are then converted to the basis vectors of the reciprocal cell. The resultant unit cell is then reduced and analyzed in terms of the 44 lattice types (Burzlaff et al., 1992).
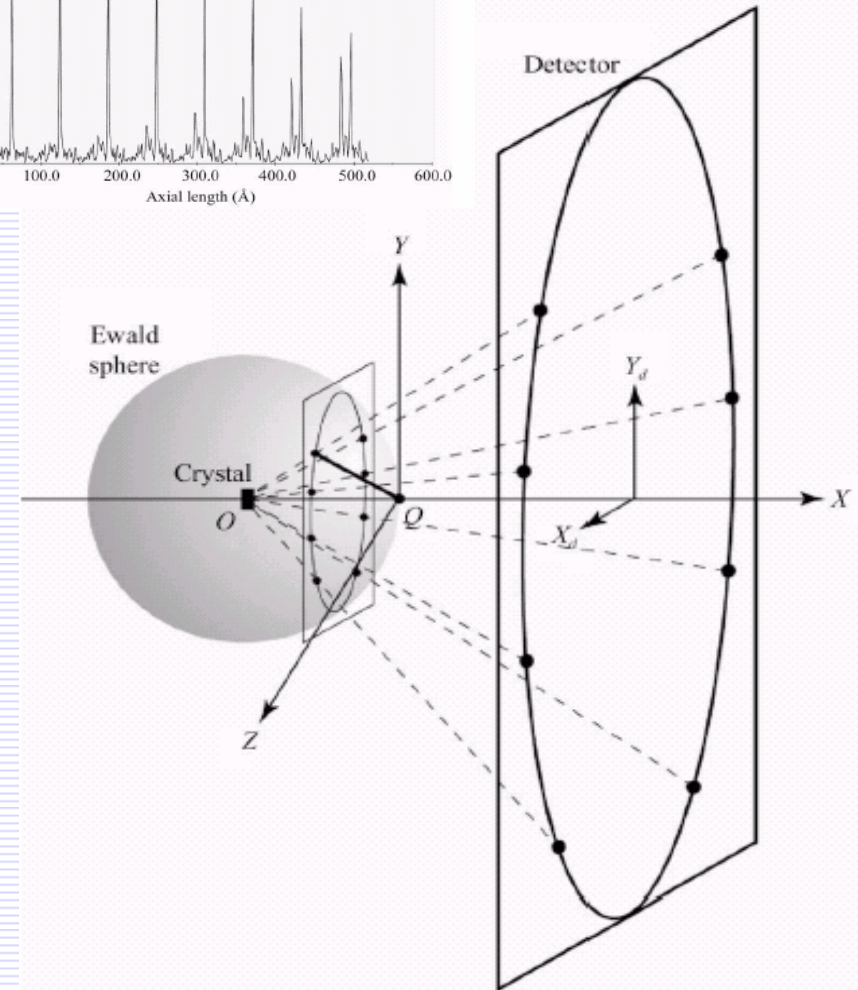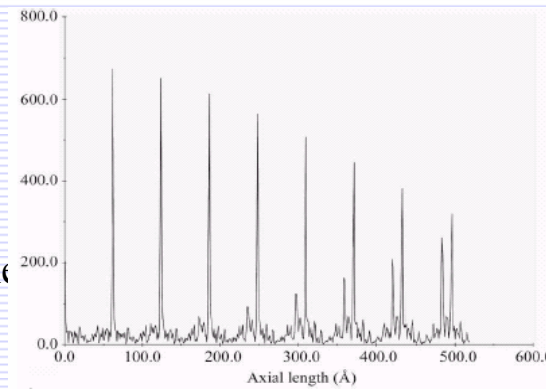
Optimization of Data Collection

- 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!

Part 2: Data Processing (*integration and scaling*)

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!

Part 2: Data Processing (*integration and scaling*)

```
[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  250.00mm taken from image header


 Wavelength of 1.54180A taken from image header
```

Part 2: Data Processing (*integration and scaling*)

```
Crystal to detector distance of  250.00mm taken from image header

Wavelength of 1.54180A taken from image header

Pixel size of 0.1500mm 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	:	0.00
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

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

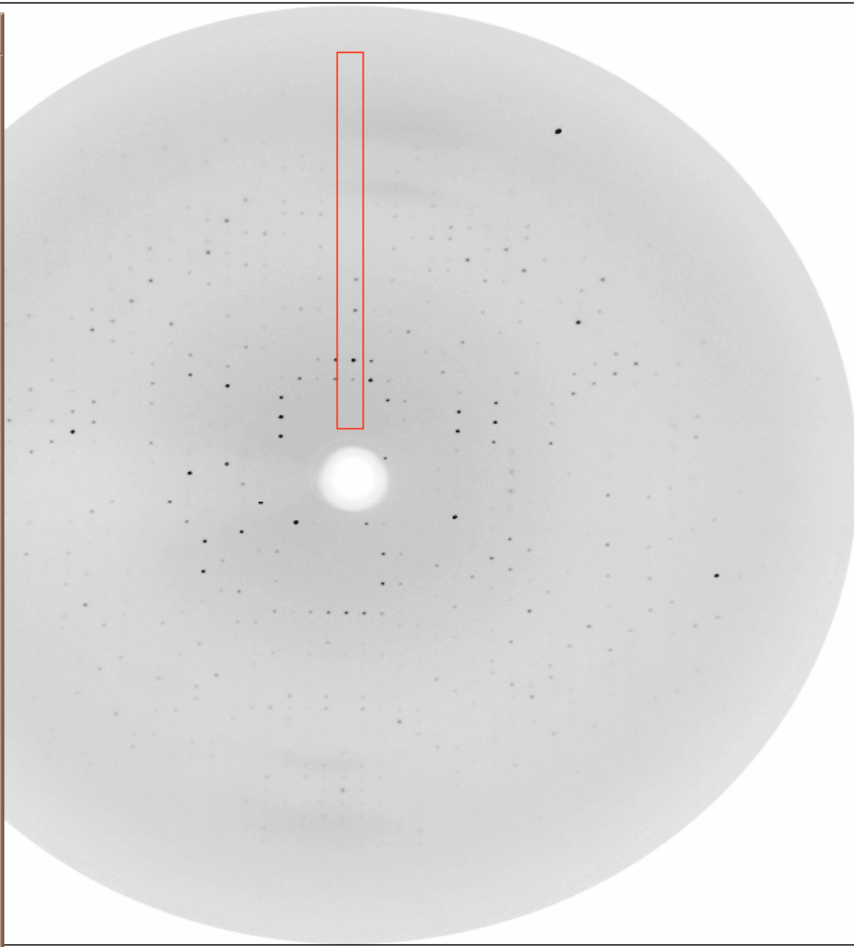|		
a	:	0.00
b	:	0.00
c	:	0.00
alpha		0.00
beta		
gamm		
PsiX		
PsiY		
PsiZ		
Mosa		
Divh		
Divv		
Lamb		
Dist		
Beam		
CCOM		
ROFF		
TOFF		
YSCA		
Pick		
Int		
Vect		
Two		
Reso		
*SPO		
Thre		
Rmin		
Rmax		
X of		
Y of		
Min		
Max		
Min		
Max		
Min		
X sp		
Y sp		
*AUT		
Min		
Prom		
Upda		
Afte		
Afte		
Time		

Main menu

Read image

Find spots

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)␣

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

Missets ThetaX,Y,Z
    0.00   0.00   0.00

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	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

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
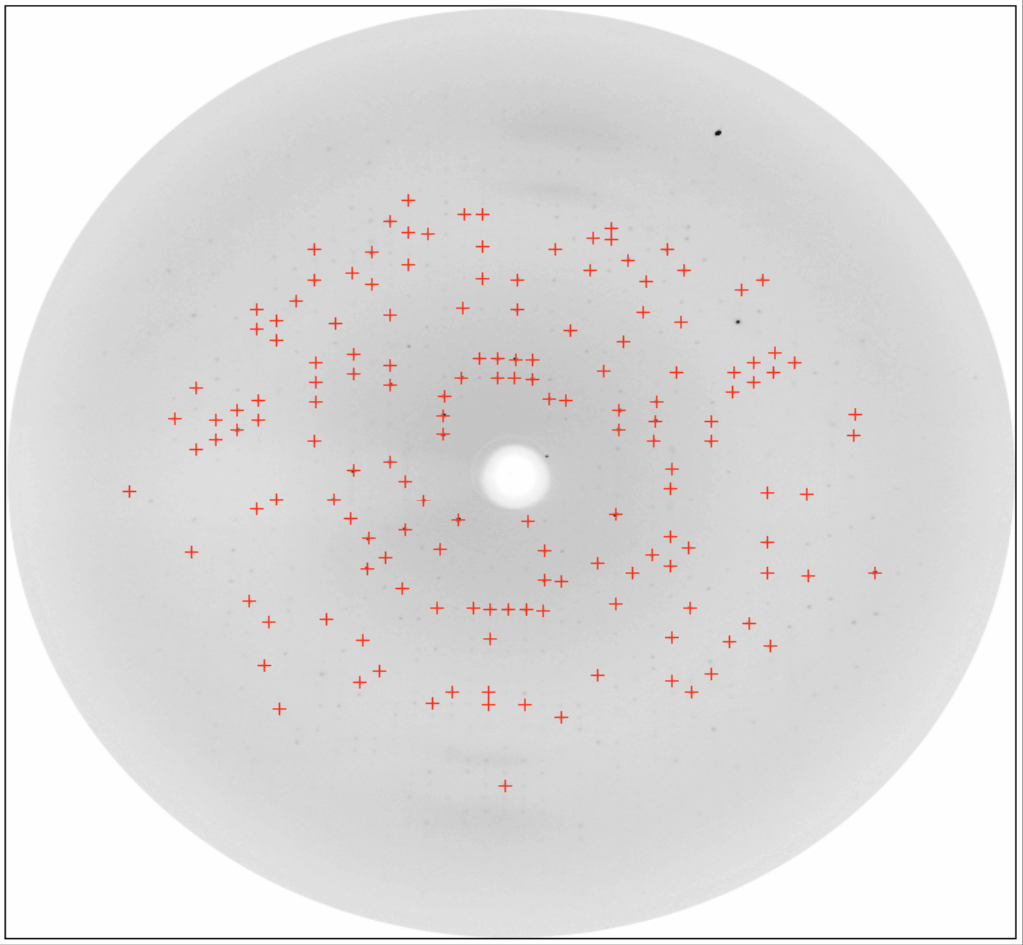- Circles
- Beam / backstop
- Save/Exit

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

Min 1   Max 2207   Cursor position

Overlay ▽ on   Contrast

Colour ▽ Black on whi Mag ▽ x4   PS   Zoom

MOSFLM Image Display

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	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

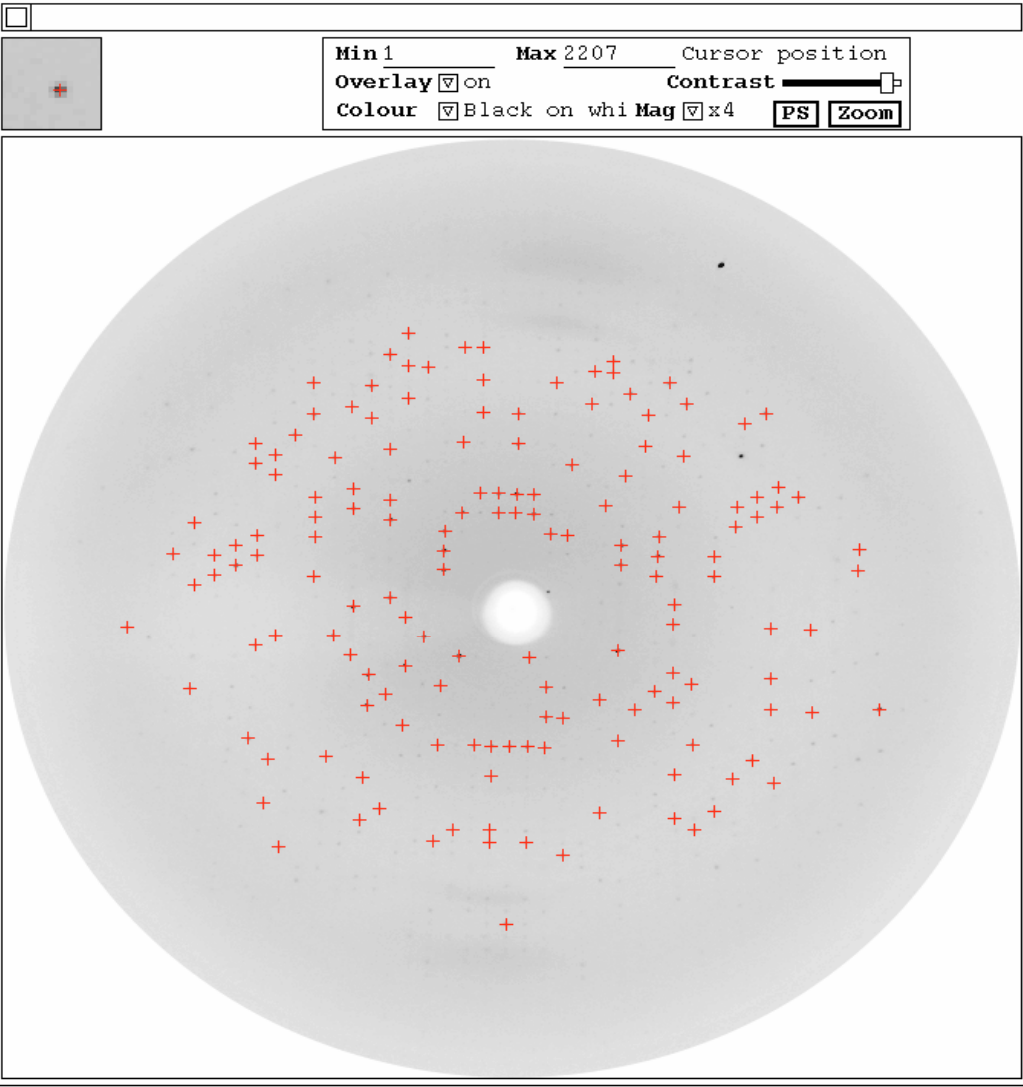|	
Prompts	On
Update display:	
After refinement	No
After integration	No
Timeout mode	Off

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	

Min 1   Max 2207   Cursor position

Overlay ▽ on   Contrast ▬▬▬

Colour ▽ Black on whi Mag ▽ x4   PS Zoom

MOSFLM Image Display

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

Edits allowed

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

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	:	1.54
Distance:		250.0
Beam X	:	149.6
Y	:	149.7
CCOMEGA	:	0.00
ROFF	:	0.0
TOFF	:	0.0
YSCAL	:	1.000
Pick area: X:		1
Y:		1
Int threshold:		2
Vector scale		
Two theta		0.0
Resolution		2.8

SPOT SEARCH

Threshold	4.9
Rmin	15.0
Rmax	135.0
X offset	0.0
Y offset	0.0
Min X size	0.5
Max X size	2.0
Min Y size	0.5
Max Y size	2.0
Min no of pix	
X splitting	0.3
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

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
-1-1-1-1-1-0-1-1-0-1-1-1-1-0-1-1-1-1-0-0-1-1-1
-1-1-1-1-0-1-1-1-1-1-1-1-1-1-1-1-1-1-1-0-1-1-0
-1-1-1-1-0-1-1-1-1-0-1-1-1-1-1-1-1-1-1-0-0-1-1
-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-2-1-1-1-1-1-1-0
-0-1-0-0-1-1-1-1-1-1-1-1*2*2-1-1-1-1-1-0-1-1
-1-0-1-1-1-1-1 2 2 3 3 3 2 2*2*2-1-1-1-1-1-1
-1-1-1-1-1-1-1 2 2 3 5 6 6 5 4 3*2-1-1-1-1-1-1
-1-1-1-1-0-1 1 2 4 8 C E E B 7 5 3*2-1-1-1-1-1
-1-1-1-1-1-1 2 4 8 E M P N I C 7 3*2-1-1-1-1-0
-1-1-1-1-1-2 2 6 C M U X U N F 9 4*2*2-1-1-0-1
-1-1-1-1-1-1 3 8 G Q Y Z V O G 9 5*2-1-1-2-1-1
-1-1-1-1-0-1 3 8 G P V X S M F 8 4*2-2-1-1-1-1
-1-1-1-1-2 3 6 B I O P L H B 6 3*2-1-1-1-1-1
-1-1-1-1-1-1 2 3 6 A D E C A 6 4 2-1-1-1-0-1-1
-1-1-1-1-1-1 2 3 5 6 7 6 5 3 2-1-1-1-1-1-1-1
-1-1-1-1-1-1-1 2 3 3 3 3 2-1-1-1-1-1-1-1-1-0
-1-1-1-1-1-1-1-1-2-2*2-1-2-1-1-1-1-1-1-1-1
-0-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-0-1-1-1-0
-1-1-1-1-1-1-1-1-0-1-1-1-1-1-1-1-1-0-0-1-1-1-1
-1-0-1-1-0-0-1-1-1-1-1-1-1-1-1-0-0-1-1-1-1-1
-0-1-0-1-0-1-1-1-1-1-1-1-1-1-1-1-1-1-0-0-0-1-0
Rms residual  0.06mm, weighted residual   0.8
```
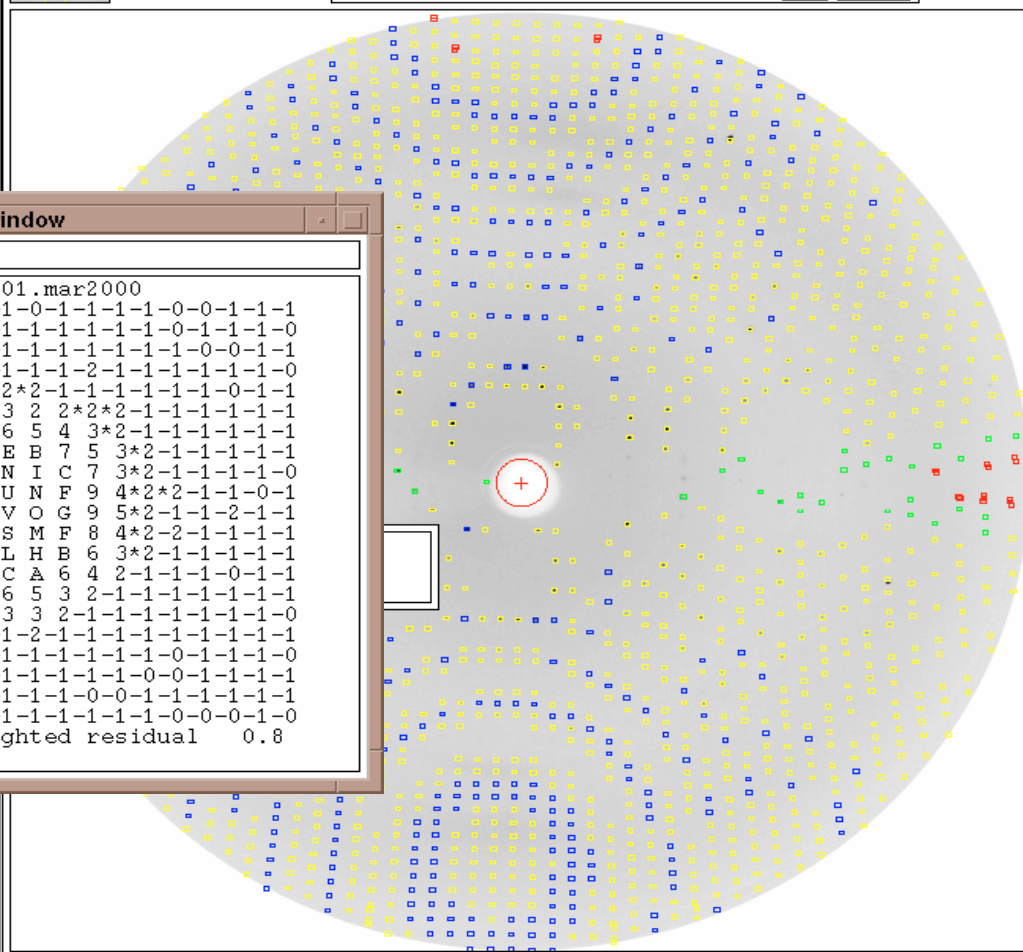
Blue: fulls, Yellow: partials, Red: overlaps
Green: too wide in phi

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:		249.71
Beam X	:	149.65
Y		149.69
CCOMEGA	:	0.000
ROFF	:	-0.08
TOFF	:	0.07
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	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

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]
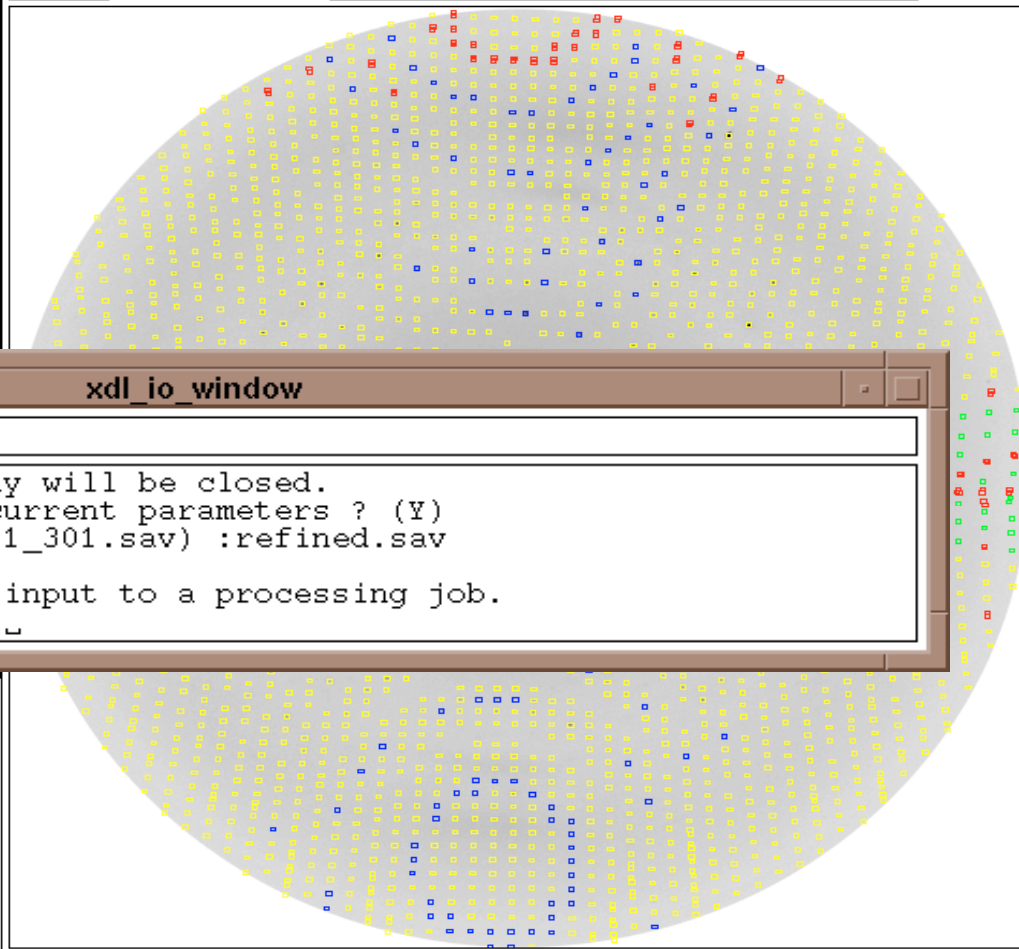
xdl_io_window

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

Scaling and merging with "*Scala*"

- 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.

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

Part 2: Data Processing (*integration and scaling*)

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

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

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

k_{hl} scale factor for I_{hl} $<I_h>$ current estimate of I_h

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

$g_{hl} = g(\quad \text{rotation/image number}) \cdot g(\text{time}) \cdot \quad g(\mathbf{s})$

...other factors
        Primary beam $\mathbf{s}_0$ *B-factor   Absorption*

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
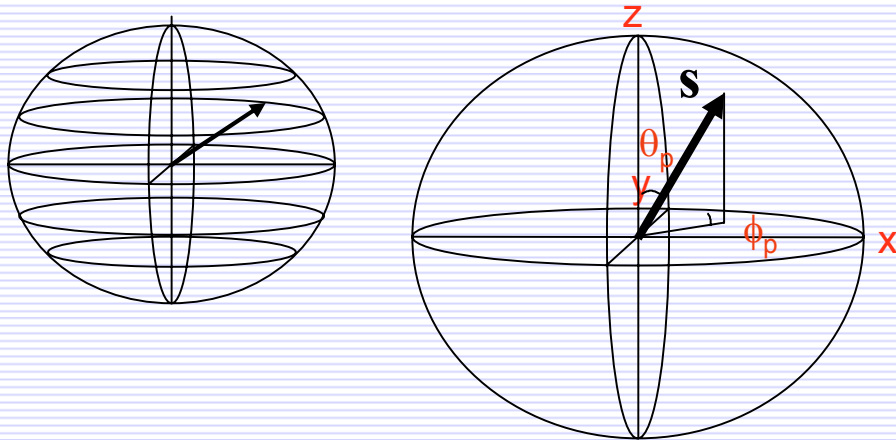- inverse beam
- simultaneous measurement of I+ & I-
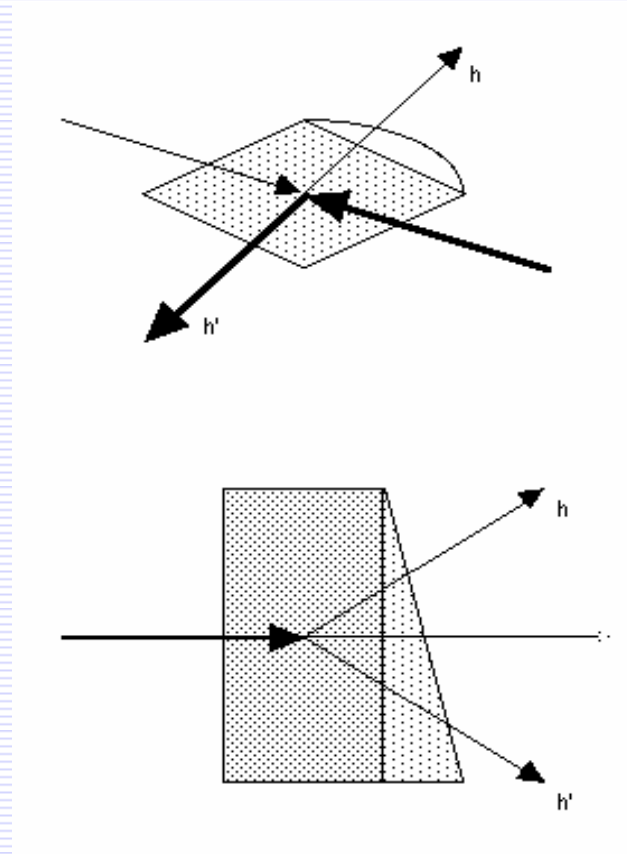
Spherical harmonics
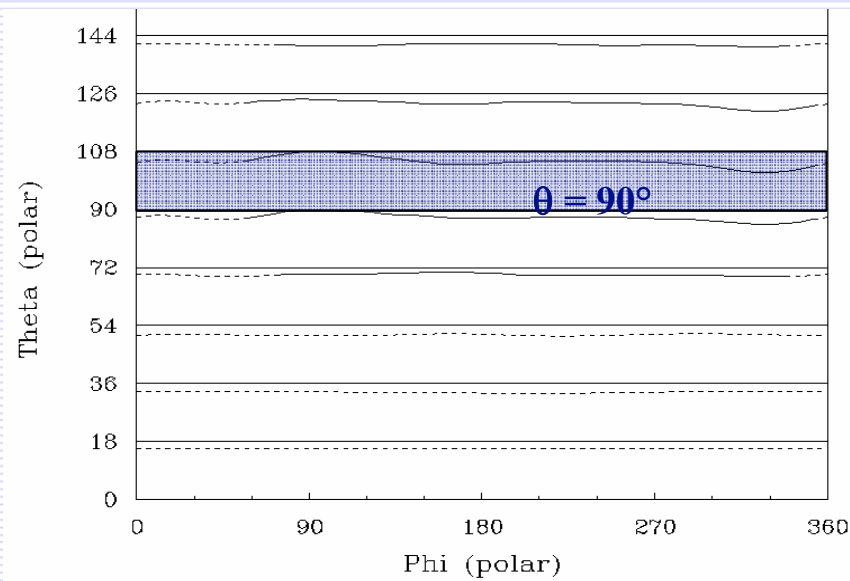
$$A(\mathbf{s}) = \Sigma_{lm}\, c_{lm} Y_{lm}(\theta_p, \phi_p)$$

- 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

Part 2: Data Processing (*integration and scaling*)

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

Results of scaling: R-factors

- **Data quality indicators**:

    Rmerge (Rsym) = $\Sigma \mid Ihl - <Ih> \mid / \Sigma \mid <Ih> \mid$

   *This is the traditional measure of agreement, but it increases with higher multiplicity even though the merged data is better*

- **Diffraction intensities values**

- **Corrected $\sigma(I)$** is compared with the intensities: the most useful statistic is $< <I>/\sigma(<I>) >$ (labelled Mn(I)/sd in tables)

Part 2: Data Processing (*integration and scaling*)
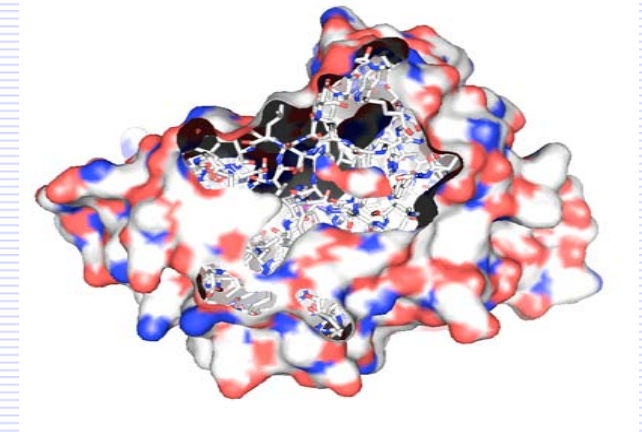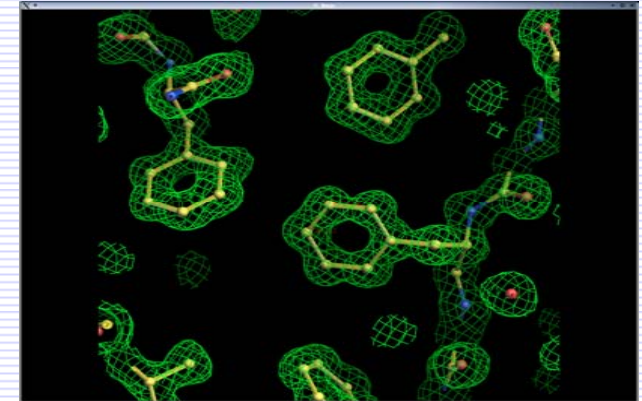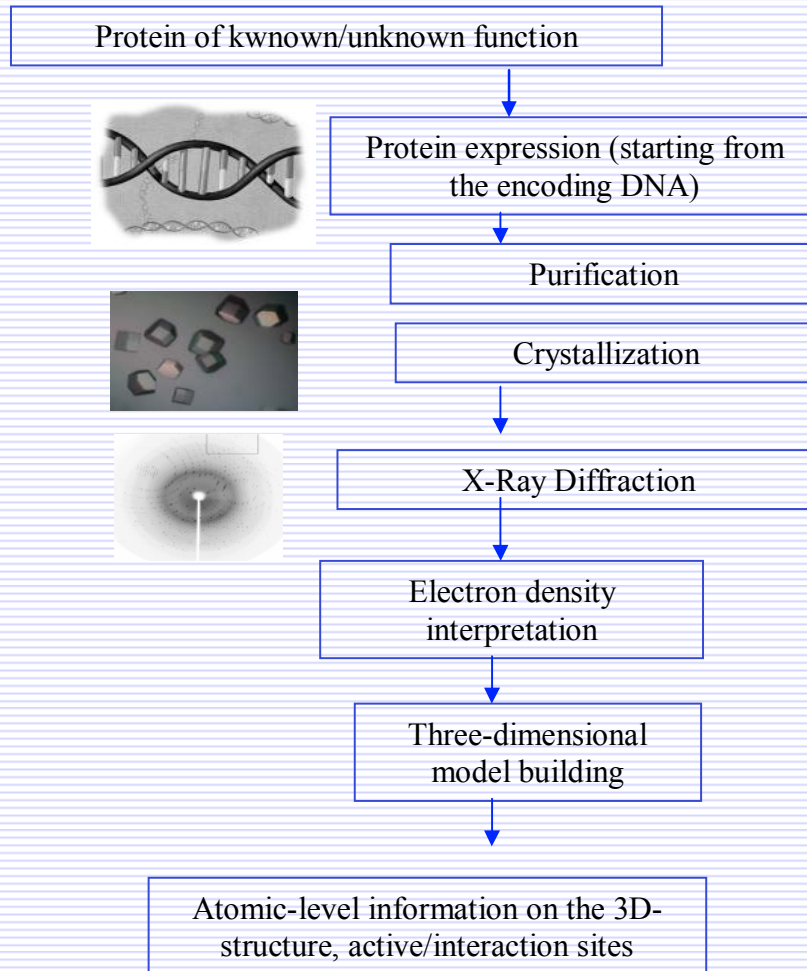
Results of scaling: Common sense checks

- **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
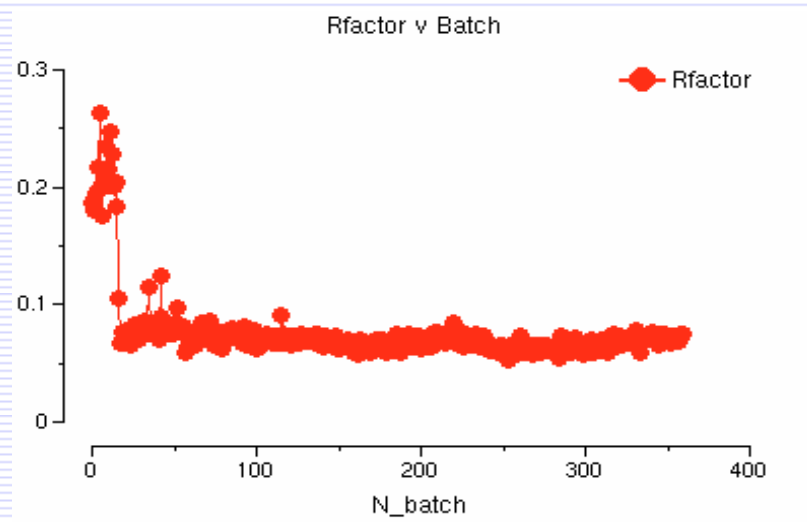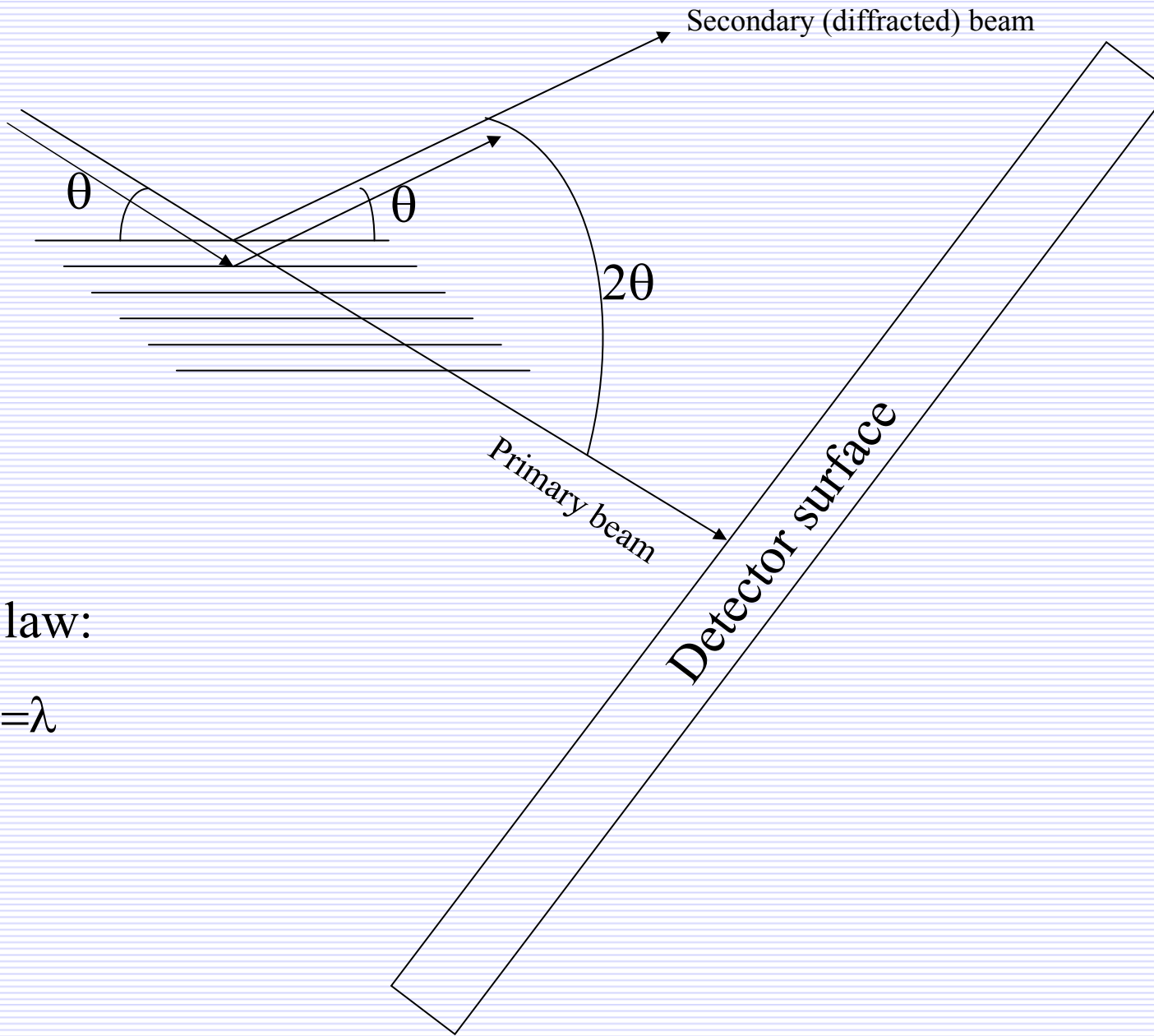

- Here the beginning of the dataset is wrong due to problems in integration (e.g. poor orientation matrix in MOSFLM at start of job.)

Part 2: Data Processing (*integration and scaling*)

Now we have the |F(hkl): what's missing?

$$\rho(x, y, z) = \frac{1}{V_c} \sum_h \sum_k \sum_l |F(h, k, l)| \cdot e^{-2\pi i(hx + ky + lz) + i\alpha(h, k, l)}$$

Protein of kwnown/unknown function

Protein expression (starting from the encoding DNA)

Purification

Crystallization

X-Ray Diffraction

Electron density interpretation

Three-dimensional model building

Atomic-level information on the 3D-structure, active/interaction sites

Secondary (diffracted) beam

θ

θ

2θ

Primary beam

Detector surface

Bragg's law:

$2d\sin(\theta)=\lambda$

Spots: Full/ partials