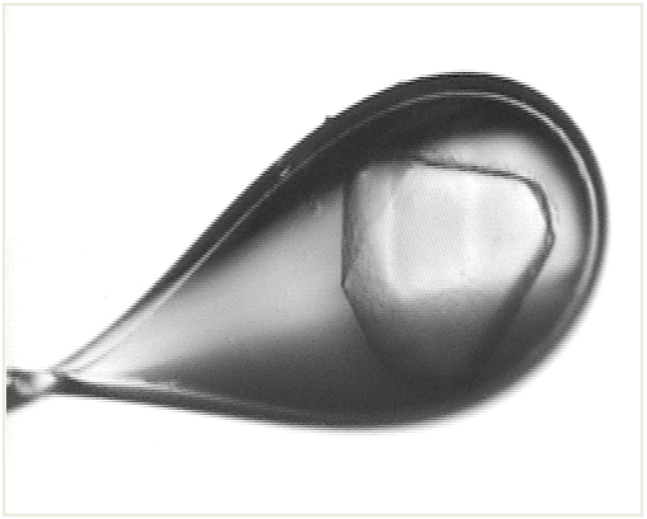
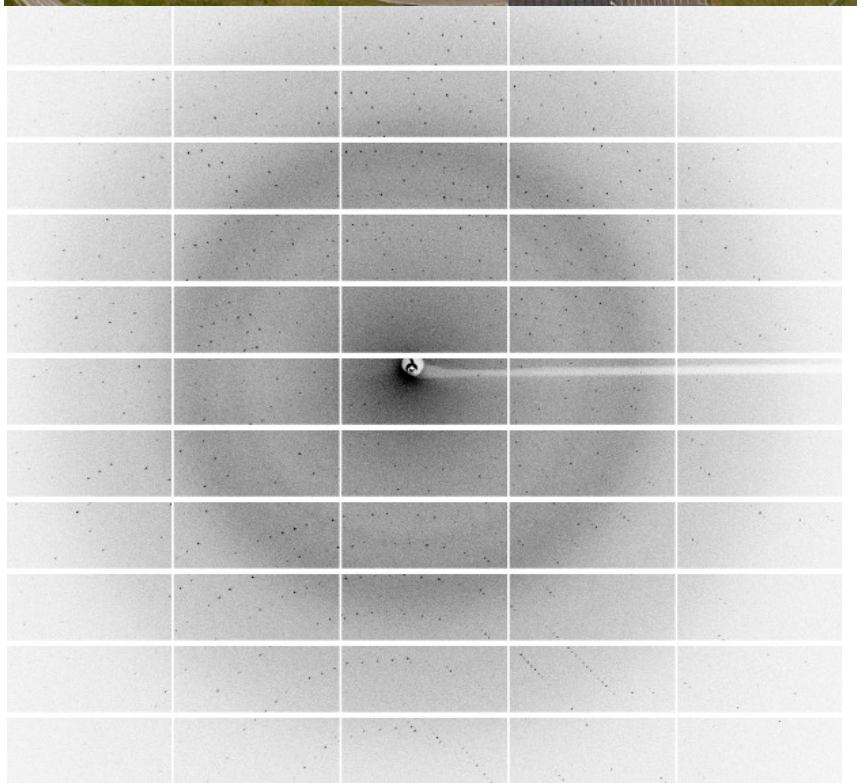
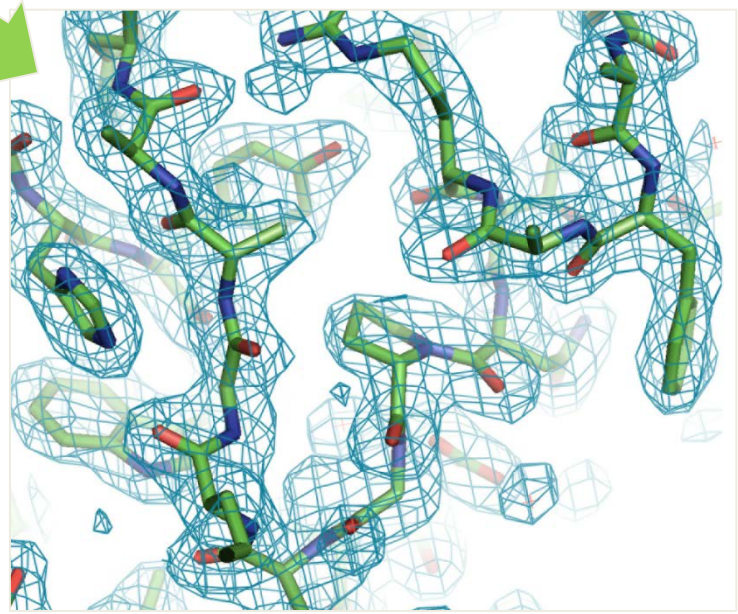
Practical MX data collection

Ralf Flaig
ralf.flaig@diamond.ac.uk

Thanks to
Neil Paterson
Gwyndaf Evans
for slides







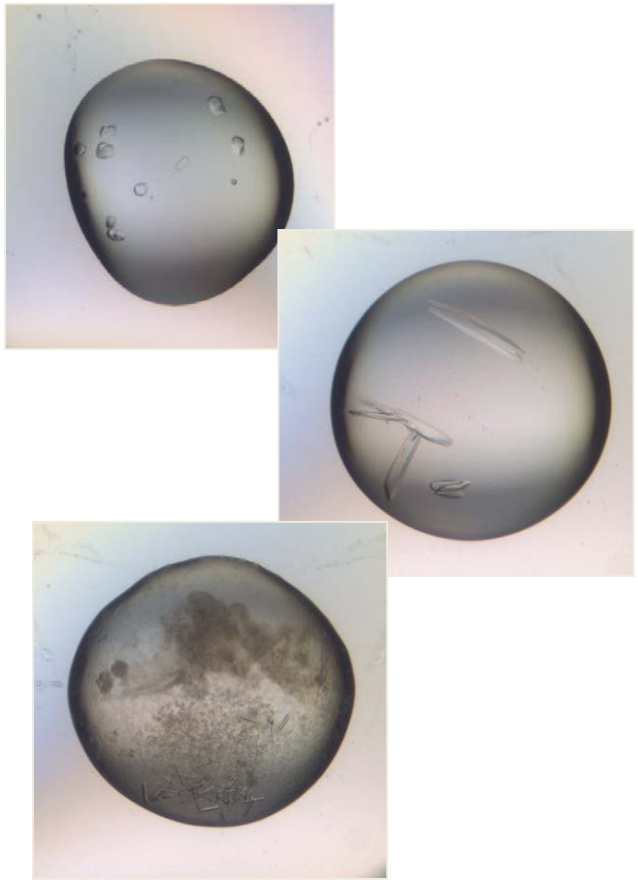
Data collection is important!

This is the last experimental step

Time taken now to collect the best possible data can prevent pain and possible failure later on.

If the original data are poor then not much can be done about it later.

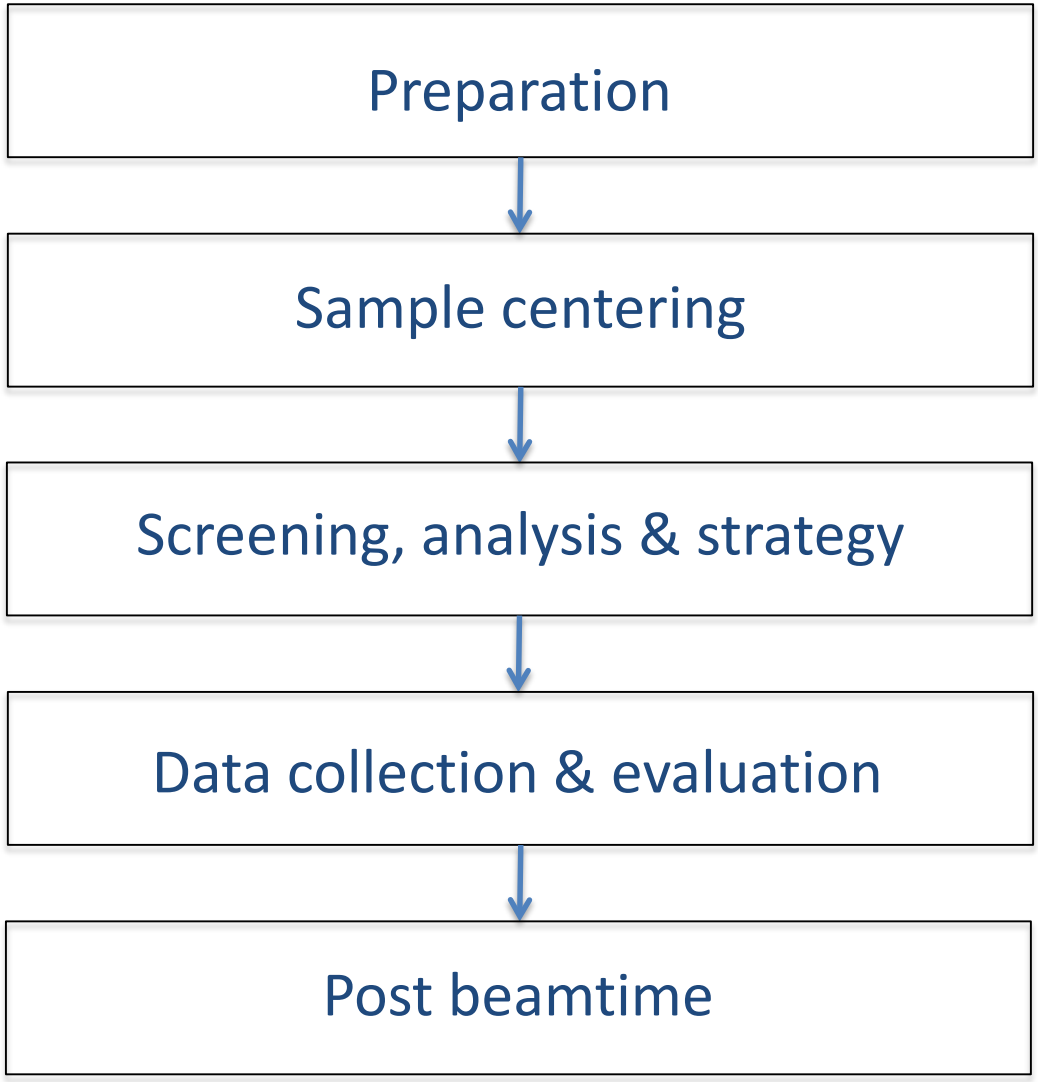
Ask your local contact for advice if you are unsure, ideally drop an email through before your trip.



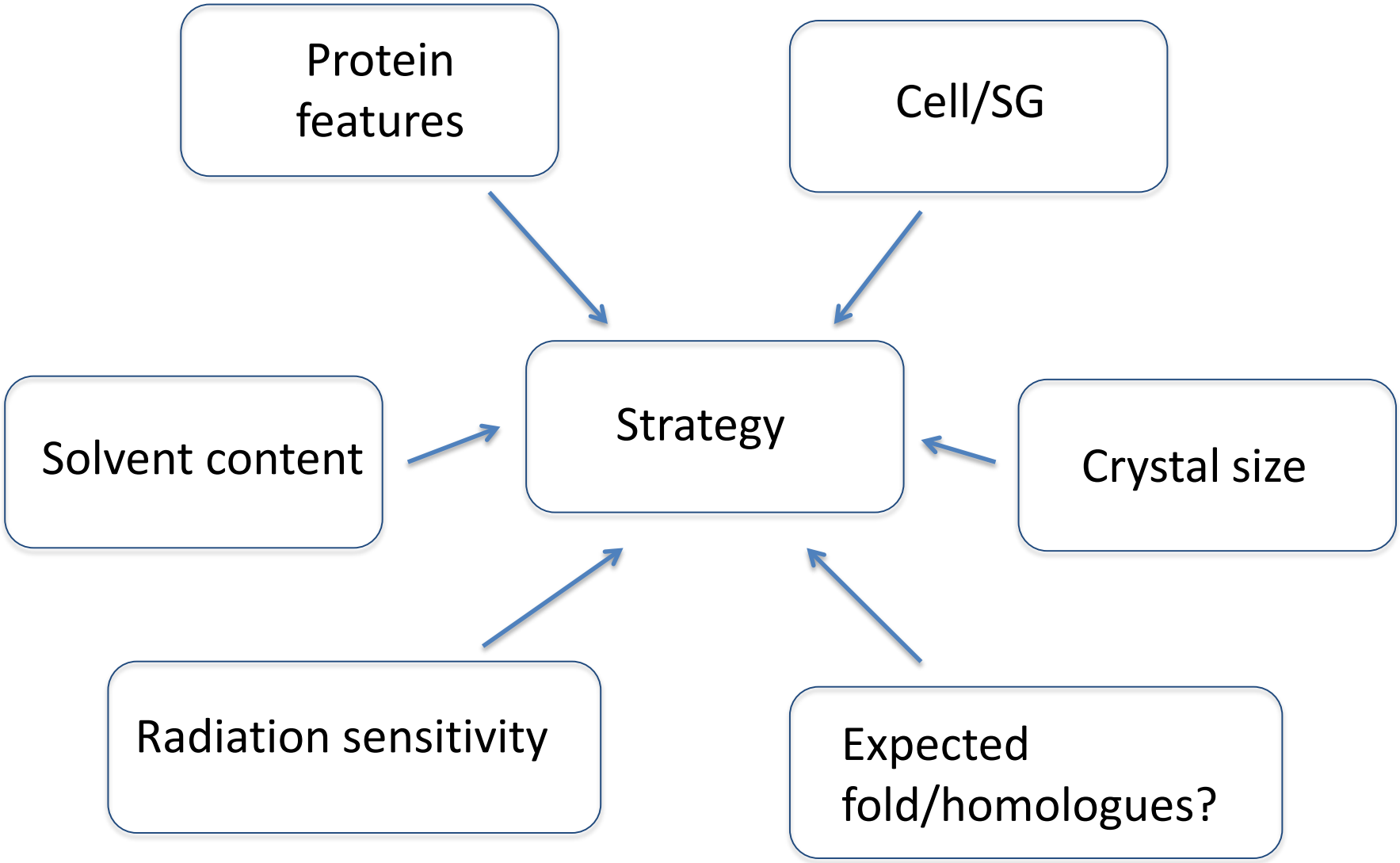
What's the goal?

Aim	High priorities	Lower priority
Native data collection	High resolution, complete data	High redundancy
MAD, SAD	Accurate, complete and highly redundant Very little radiation damage Choice of wavelength	High resolution data
-> Sulphur SAD	As above + extra high redundancy (40-200fold) if not on I23	
Molecular replacement	Completeness at low resolution (< 3Å)	
Ligand/mutation	High resolution, complete data. High throughput	High redundancy

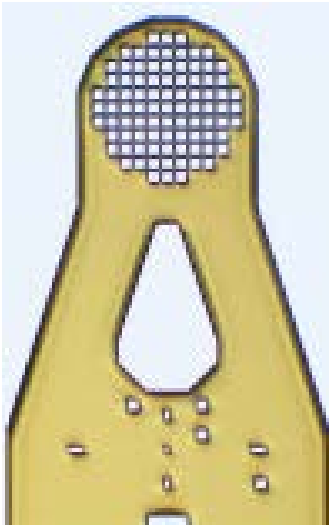
Trying to collect everything at once will usually result in failure

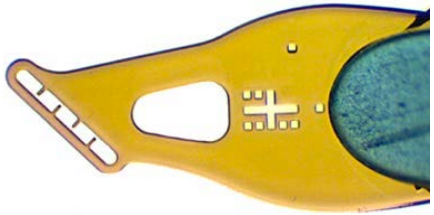

Preparation
Sample centering
Screening, analysis & strategy
Data collection & evaluation
Post beamtime

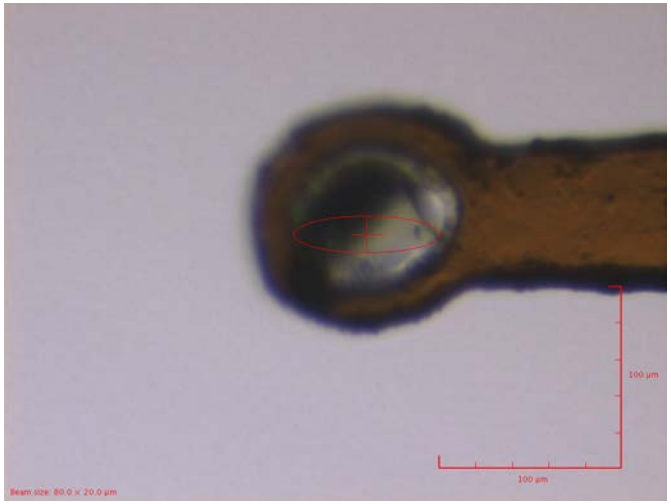
Preparation - Prior knowledge

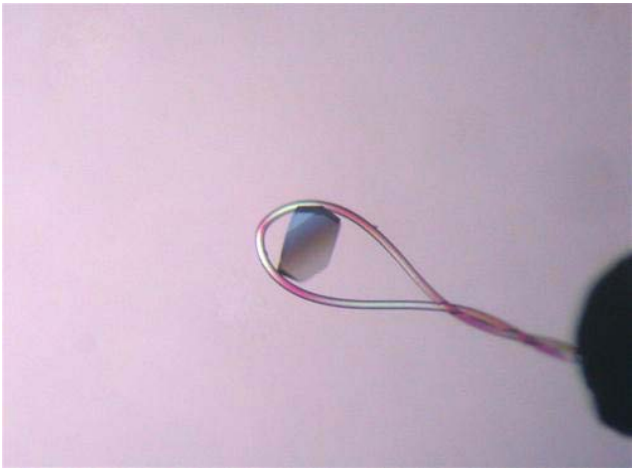


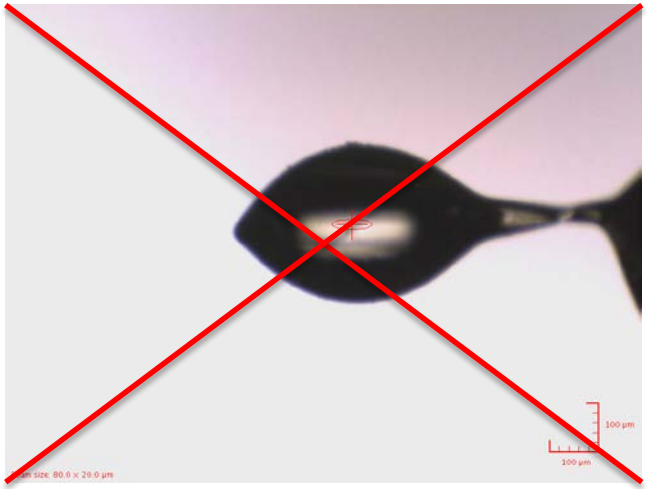
Preparation - Sample mounting











Preparation - Choice of beamline

Energy range – does it cover what you need?

Multi-axis goniometer? P1, experimental phasing

Biological containment?

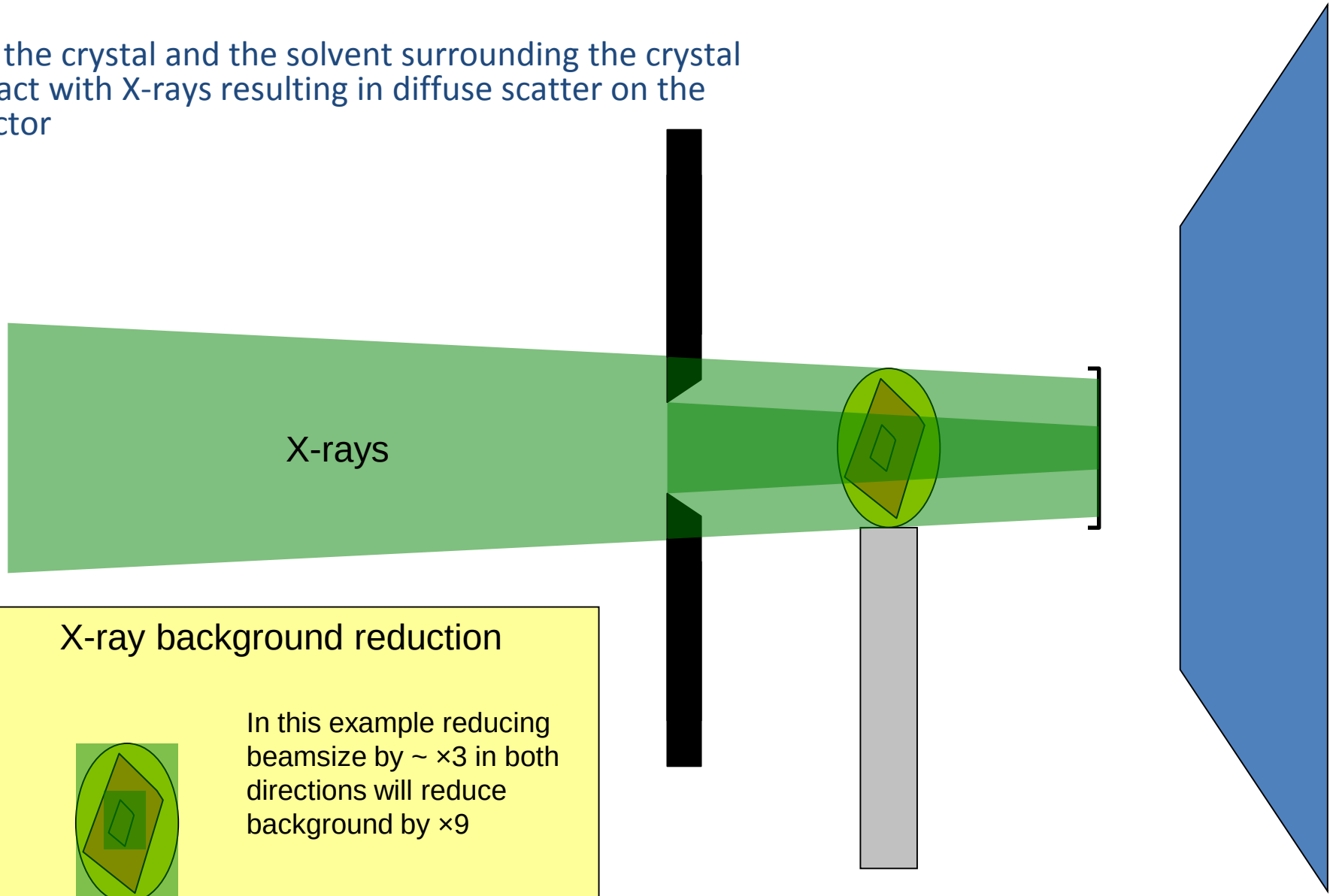
Cryo or room temperature collection?

Can the beam size match your crystals?

Flux – important for radiation damage

Beam and sample size

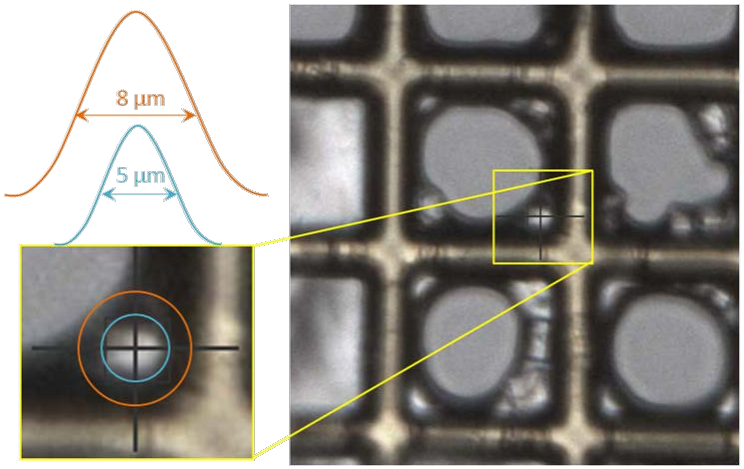
Both the crystal and the solvent surrounding the crystal interact with X-rays resulting in diffuse scatter on the detector



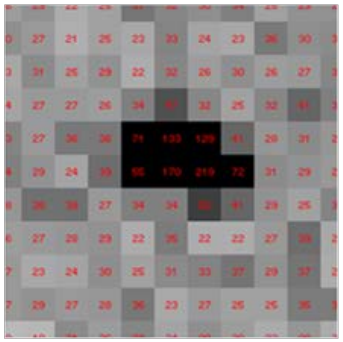
X-ray background reduction

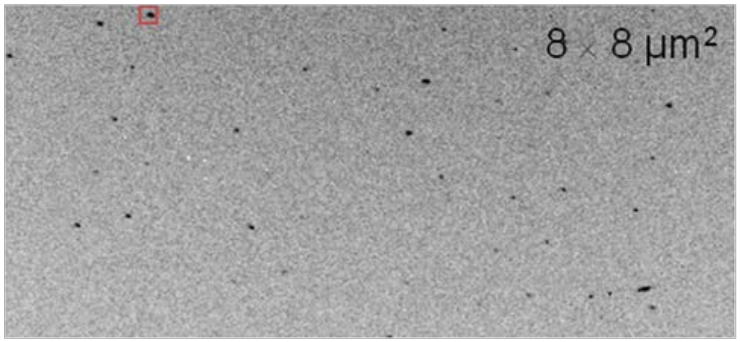
In this example reducing beamsize by ~ ×3 in both directions will reduce background by ×9

Match beam to sample

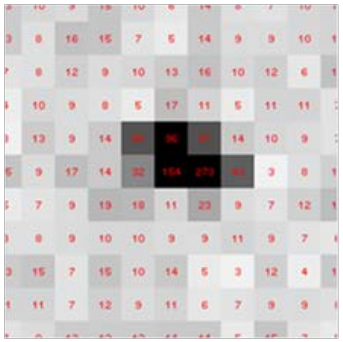


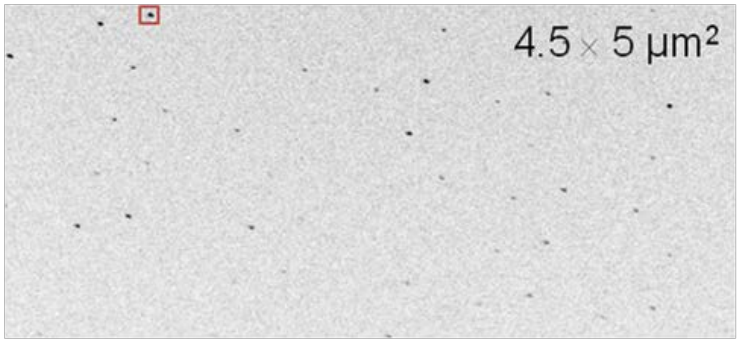
AcMNPV polyhedra crystals approx 5×5×5 μm^3 in size were mounted on a micromesh grid. Data were collected with two beamsizes: 8×8 μm^2 and 4.5×5 μm^2





I/ σ(I) = 1.5 at 2.9Å





Average background reduced by 3

I/ σ(I) = 1.5 at 2.5Å

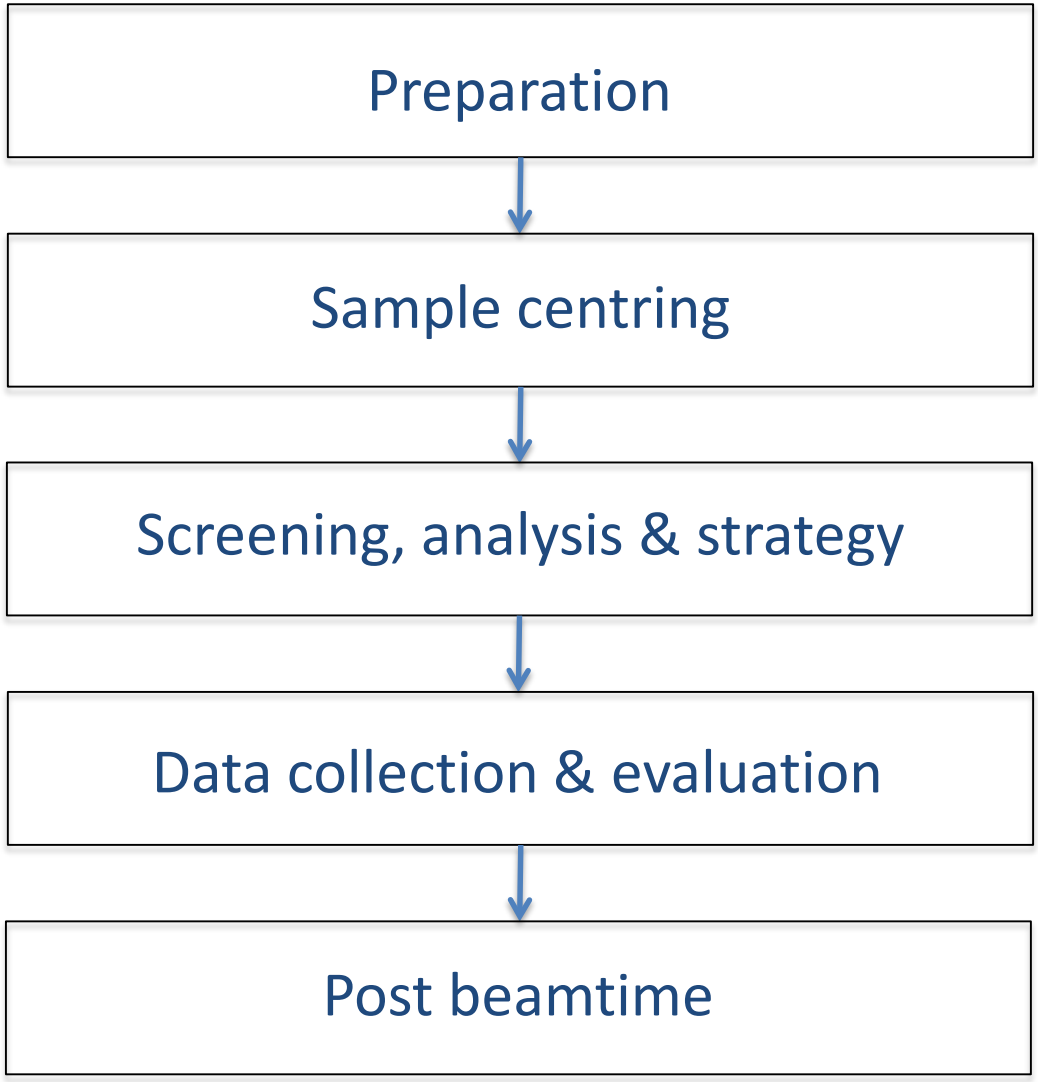
Beam size and flux density

Be mindful of how the beamline generates a small beam size. Focusing (I04 +I24) can increase the flux density massively compared to apertures (I03)

Radiation damage is related to the flux density so small beam sizes on I04/I24 can kill a crystal very quickly.

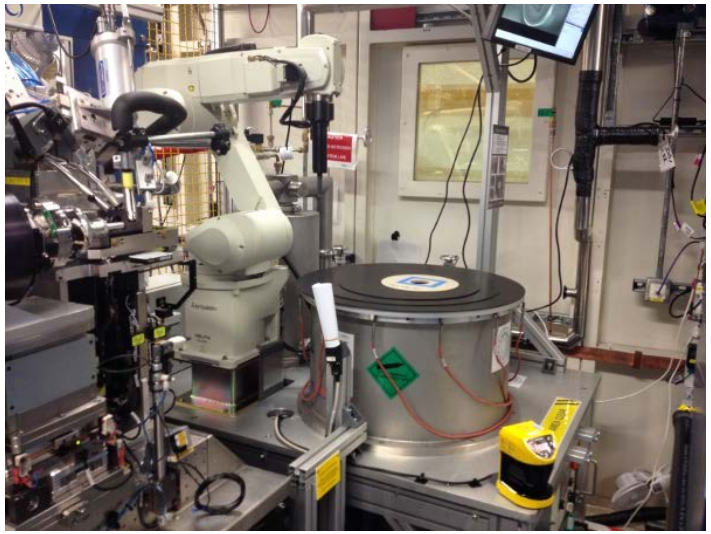
E.g. 50x50 μm^2 beam will deliver around 45x less dose than 9x6 μm^2 beam on I24

So match beam to crystal, but attenuate when required

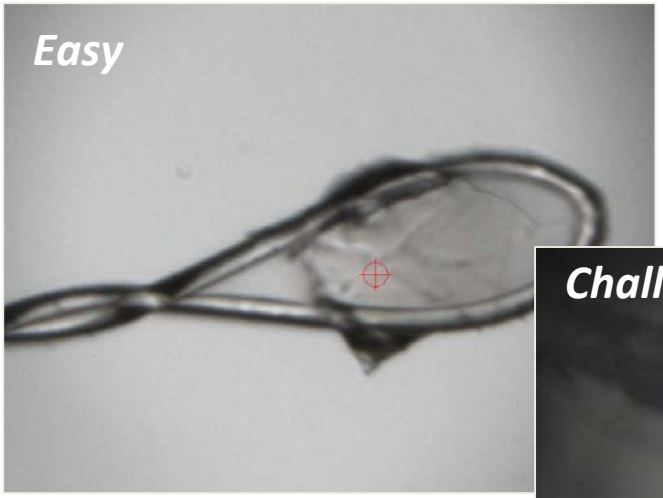

Preparation
Sample centring
Screening, analysis & strategy
Data collection & evaluation
Post beamtime

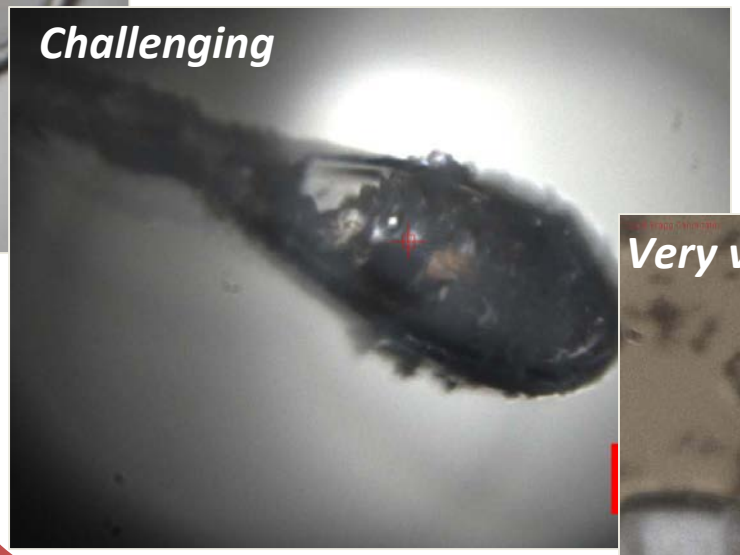
Mount and centre crystal

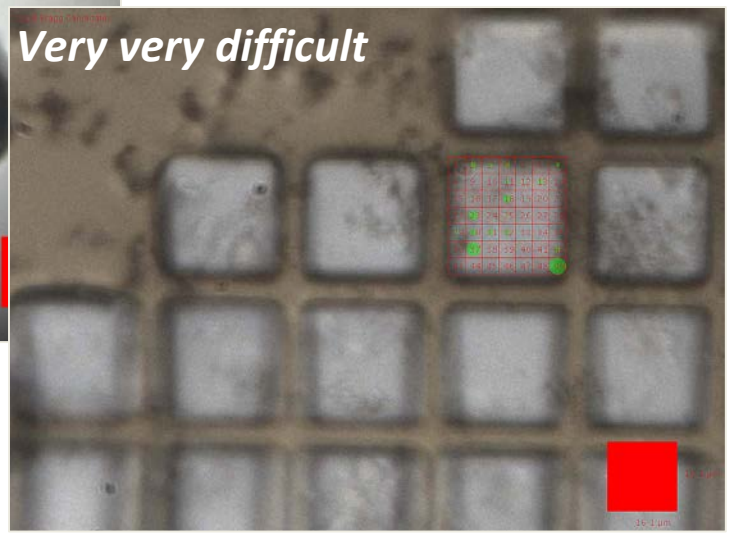
Mounting now automated at most beamlines.



Centring can be:



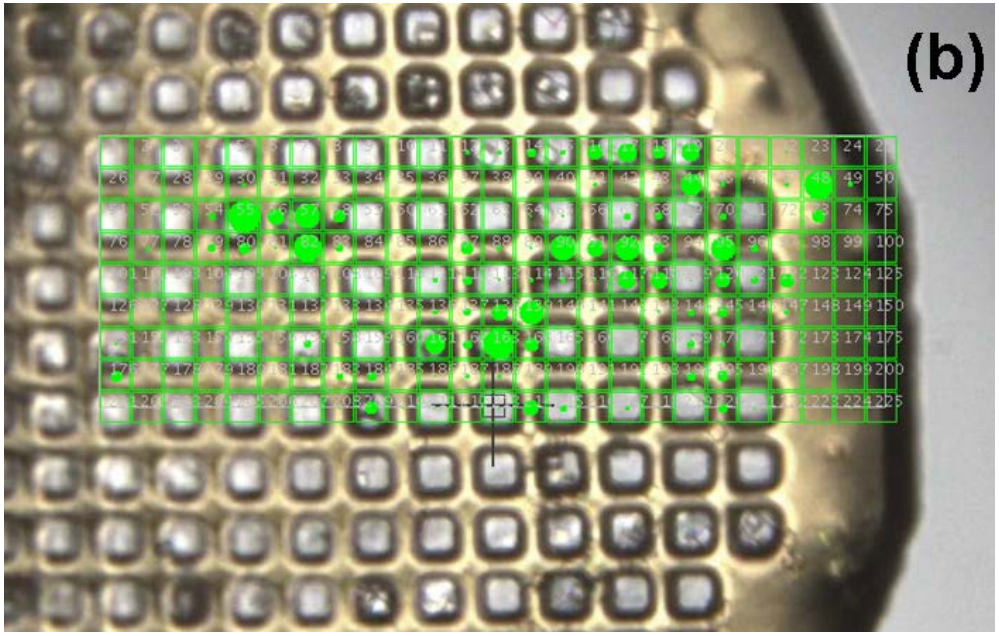




Make the most of tools developed to make these cases easier

Centring in difficult cases

Making this easier: grid scanning



Allows large areas to be quickly rastered over and small or hard to see crystals found.

Also eliminates optical effects: crystal may not be where it seems..

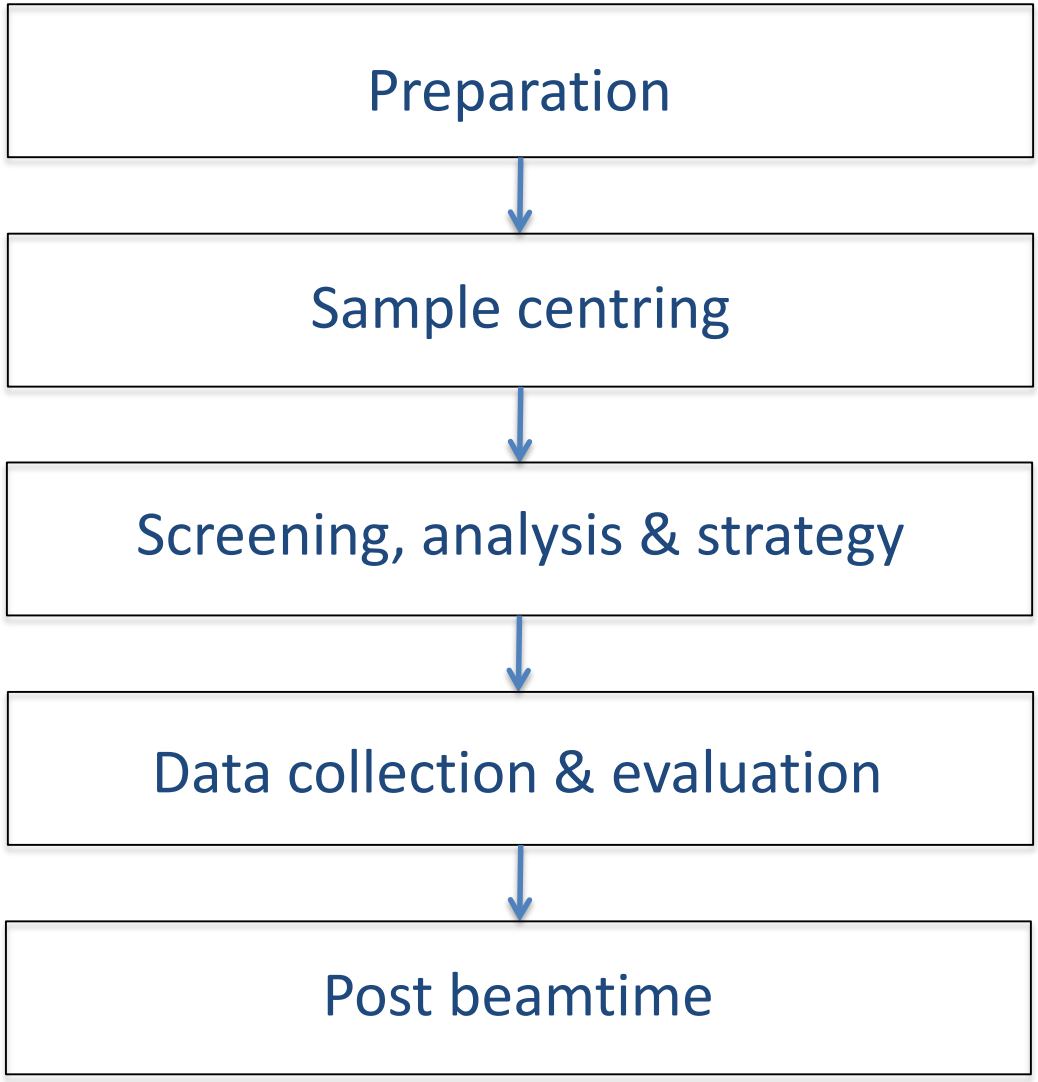
X-ray centring option in GDA

Try to crudely centre first and collect a test shot. Is it worth spending a lot of time on this sample?

If yes ⇨ Spend more time

If no ⇨ Move on

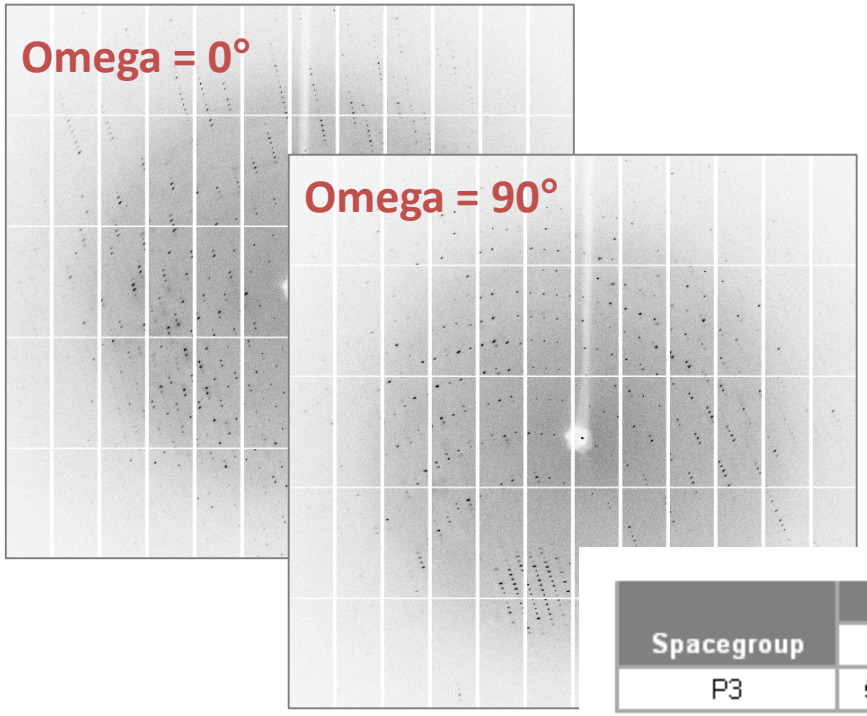
If maybe ⇨ Take advantage of the robot: dismount and see what your other samples are like


Preparation
Sample centring
Screening, analysis & strategy
Data collection & evaluation
Post beamtime

Screening and analysis

Determine crystal diffraction properties, orientation matrix and appropriate settings for data collection

Usually by collecting 3 or 4 images, each separated by 45 degrees.



Important to check the images

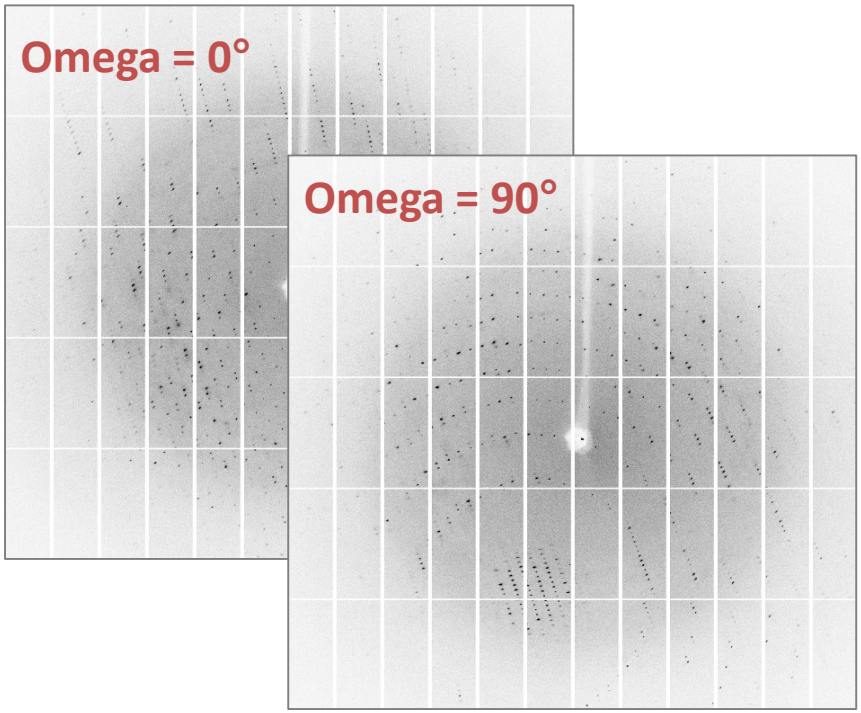
Spot quality – circular, well spaced
Similar at 0° and 90°?
Resolution appropriate?
Autoindexing successful?

Spacegroup	Refined unit cell parameters (Å/degrees)						Est Mosaicity
	a	b	c	α	β	γ	
P3	55.697	55.697	119.332	90.000	90.000	120.000	1.00

Description	Phi Start	Phi Oscillation	No Of Images	Phi End	Completeness	Resolution
Native	202.0°	1.9°	64	322.0°	1.00	2.20Å
Anomalous	202.0°	1.9°	64	322.0°	0.93	2.20Å

Strategy

Programs such as BEST go further than just providing a start angle and oscillation range..



plus

- Estimate of crystal lifetime
- Aim of experiment (*eg* native or anomalous data collection)
- Desired multiplicity

gives

Comprehensive strategy including omega start and end, but also suggested oscillation angle, exposure time and filter transmission.

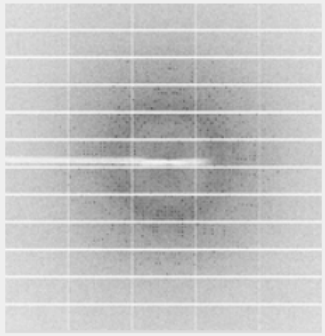
BEST: Bourenkov and Popov *Acta Cryst*. (2010). **D66**, 409-419

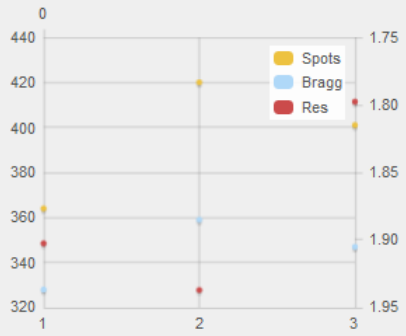
IspyB/Synchweb will display indexing results and suggested strategies for several scenarios. Results depend on the input quality! Repeat if necessary.

0 0 29-11-2018 15:51:45 - Marco/screening/Thau/Sel-Thaum_4/Sel-Thaum_4_4_####.cbf

Sample: Sel-Thaum_4	Flux: 1.21e+11
Ω Start: 0.0°	Ω Osc: 0.15°
Ω Overlap: -44.85°	No. Images: 3
φ: 345.82999° χ: 39.36°	Resolution: 1.50Å
Wavelength: 0.9780Å	Exposure: 0.040s
Transmission: 100.00%	Beamsize: 32x20µm
Type: SAD	

Comment: (-139,-765,316) (-590,-422,410) (-652,261,509) (-572,229,428) (-207,229,356) (110,-123,71) (-484,178,229) (-485,166,242) (-248,-165,196) Aperture: LargeAperture: LargeAperture: LargeAperture: LargeAperture: LargeAperture: LargeAperture: LargeAperture: Lar





Strategies — Mosflm: ✔ EDNA: ✔ Xia2: ?

mosflm

Space Group	A	B	C	α	β	γ
P4	57.55	57.55	150.04	90.00	90.00	90.00

Lookup Cell

Strategy	Description	Ω Start	Ω Osc	Res (Å)	Ranking Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
anomalous Wedge1		86	0.20	1.48	0.00	0.0	0	0.000	450
native Wedge1		86	0.20	1.48	0.00	0.0	0	0.000	450

EDNA MXv1

Space Group	A	B	C	α	β	γ
P4	57.48	57.48	149.67	90.00	90.00	90.00

Lookup Cell

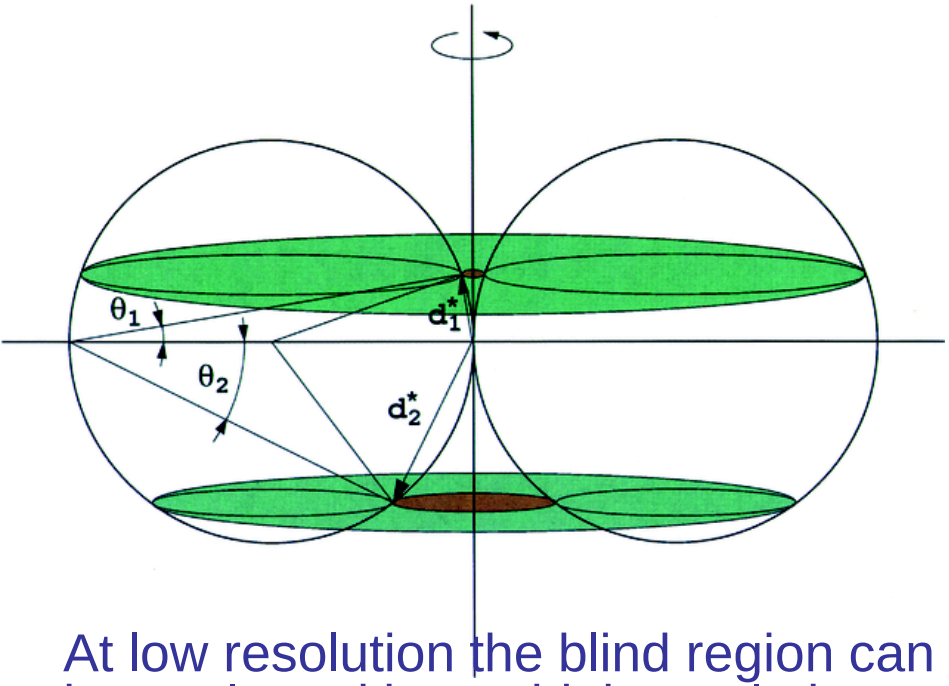
Strategy	Description	Ω Start	Ω Osc	Res (Å)	Ranking Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
Strategy1 Wedge1	Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=201.0 s	144	0.05	1.63	1.63	40.4	40.4	0.040	2400
Strategy2 Wedge1	Standard Anomalous Dataset Multiplicity=3 I/sig=2 Maxlifespan=201.0 s	0	0.10	1.97	1.97	42.9	42.9	0.040	3600
Strategy3 Wedge1	High multiplicity Multiplicity=16 I/sig=2 Maxlifespan=201.0 s	0	0.05	1.65	1.65	10.7	10.7	0.040	7200
Strategy4 Wedge1	Gentle Multiplicity=2 I/sig=2 Maxlifespan=201.0 s	144	0.05	1.63	1.63	40.4	40.4	0.040	2400

XOalign

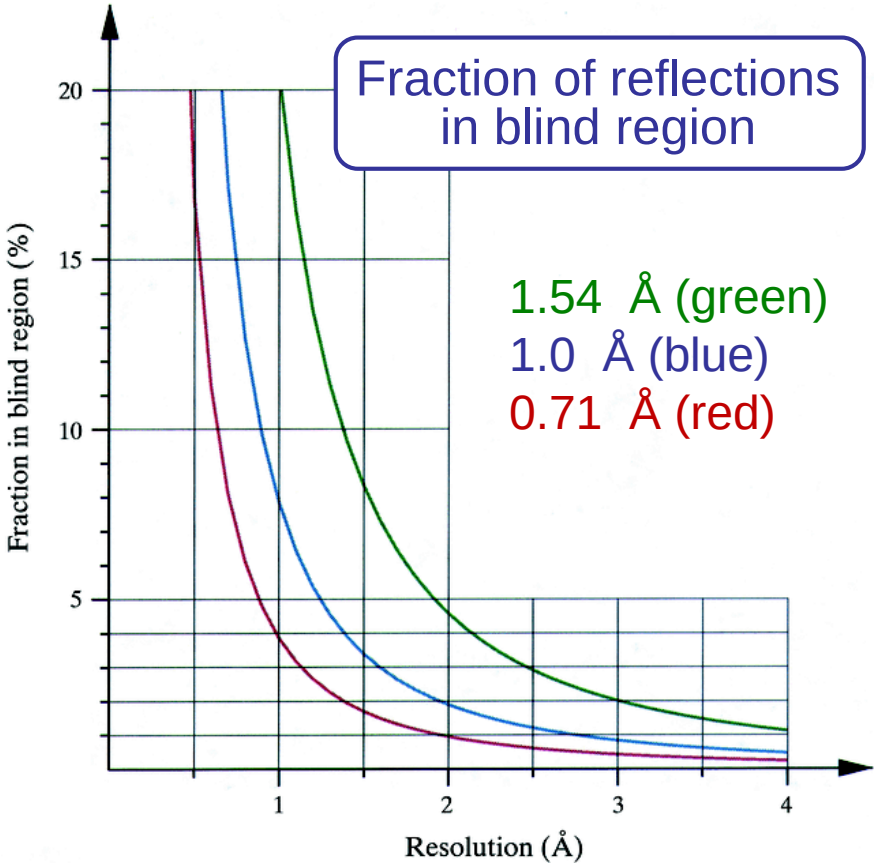
Axes	Kappa	Chi	Phi
[(b*, a*), (b*, c*)]		38.6	345.87

The blind region

When rotating around a single axis some reflections close to the rotation axis will never pass through the Ewald sphere.



The blind region is larger at long wavelengths as $1/\lambda$, and hence the radius of the Ewald sphere, is smaller.



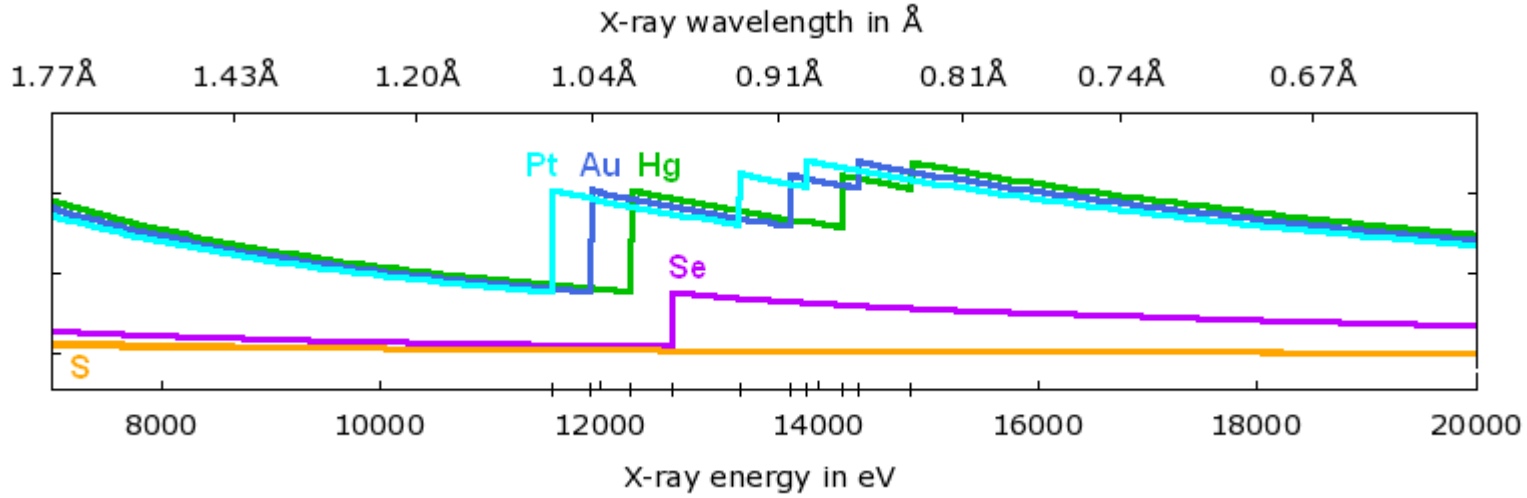
At low resolution the blind region can be neglected but at high resolution and in low symmetry space groups it may be necessary to offset the rotation axis.

Collecting anomalous data

More data is required because Friedel's law is no longer true

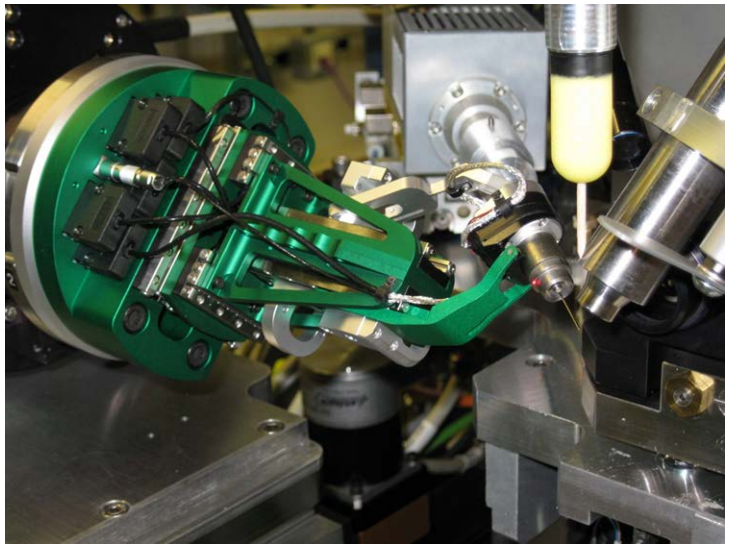
i.e. I(*h*,*k*,*l*) ≠ I(*-h*,*-k*,*-l*)

This occurs because electron resonance occurs in elements at certain wavelengths. The resonant wavelength depends on the element



Point group	Native data	Anomalous data
1	180 (any)	$180 + 2\theta_{max}$ (any)
2	180 (*b*); 90 (*ac*)	180 (*b*); $180 + 2\theta_{max}$ (*ac*)
222	90 (*ab* or *ac* or *bc*)	90 (*ab* or *ac* or *bc*)
4	90 (*c* or *ab*)	90 (*c*); $90 + \theta_{max}$ (*ab*)
422	45 (*c*); 90 (*ab*)	45 (*c*); 90 (*ab*)
3	60 (*c*); 90 (*ab*)	$60 + 2\theta_{max}$ (*c*); $90 + \theta_{max}$ (*ab*)
32	30 (*c*); 90 (*ab*)	$30 + \theta_{max}$ (*c*); 90 (*ab*)
6	60 (*c*); 90 (*ab*)	60 (*c*); $90 + \theta_{max}$ (*ab*)
622	30 (*c*); 90 (*ab*)	30 (*c*); 90 (*ab*)
23	~60	~70
432	~35	~45

Multi-axis goniometer



- crystal alignment/orientation
- get equivalent images from different crystals by reorienting them
- smart data collection strategies, better completeness of data
- reduced radiation damage by using smaller overall oscillation range
- help in point group determination
- comparing crystals in the same orientation
- optimised MAD data collection; Bijvoet pairs on the same frame
- improved high multiplicity SAD data collection protocols
- reducing/avoiding spot overlap, better spot separation (long unit cell problem)
- multiple crystal data collection: collect only what has not been collected before

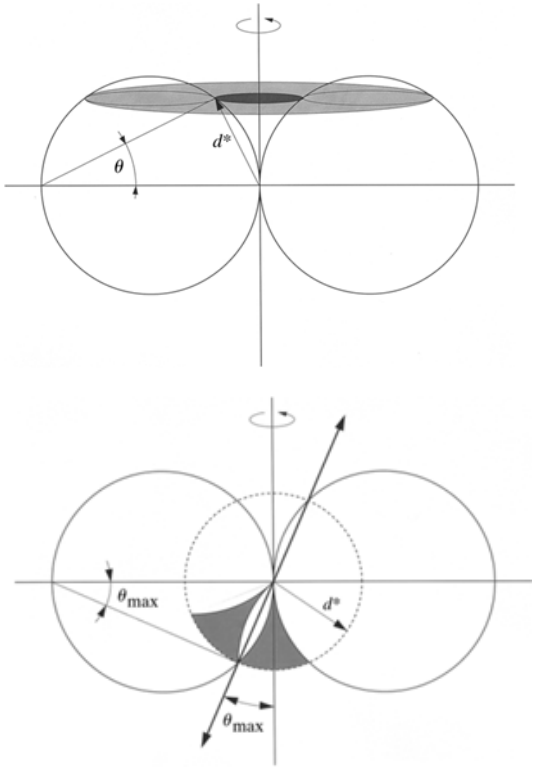
Complete data sets

OAV View | Data Collection Table Input

Visit Folder: /dls/i04/data/2017/cm16781-5 | Default Folder: gw/${date}/${proteinacronym}/${samplename} | Default Prefix: ${samplename} | Run All | Run Selected

Sample ID	Code	Folder	Prefix	Holder	Position	Omega Start (°)	Omega Oscillatio (°)	Omega Delta (°)	Chi (°)	Phi (°)	X (µm)	Y (µm)	Z (µm)	Number of Images	Time per Image (s)	Maximum Resolutio (Å)	Distance (mm)	Wavelength (Å)	Energy (eV)
thaum_8	CA00AD1	${date}/${proteinac	${samplename}	35	8	0.00	0.100	0.00	0.000	0.000	-9999.0	-9999.0	-9999.0	3600	0.040	1.5000	266.8	0.97950	12657.9
thaum_8	CA00AD1	${date}/${proteinac	${samplename}	35	8	0.00	0.100	0.00	45.000	0.000	-9999.0	-9999.0	-9999.0	3600	0.040	1.5000	266.8	0.97950	12657.9

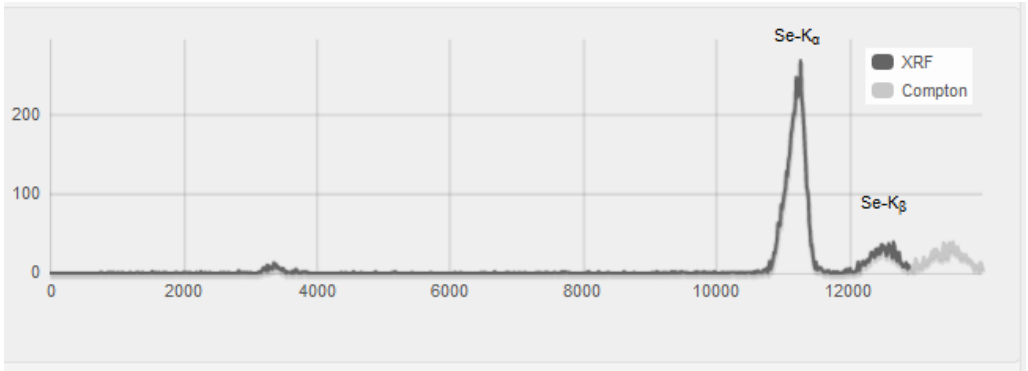
- particularly problematic for triclinic (*P* 1) samples
  - To collect missing reflections re-orient crystal by at least $2\theta_{max}$
- more pronounced with long wavelength radiation (increased curvature of Ewald sphere)
- high resolution reflections more affected than low resolution reflections
- Beware of nearly aligned crystals
  - skewing the axis by at least θ_{max} from the spindle axis ensures there will be no loss of completeness as a result of the blind region

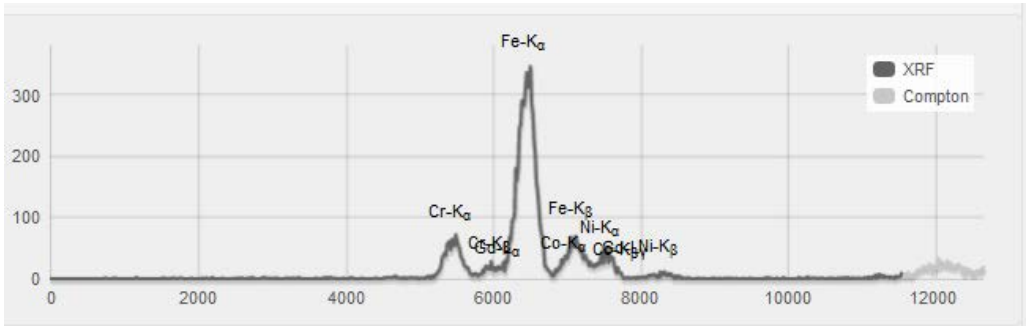


More screening?

Fluorescence spectrum/scans give you information about the elements in your sample.

Spectrum will suggest what elements are present – remember to set energy to a high value above absorption edge (e.g. 15-18keV) first.



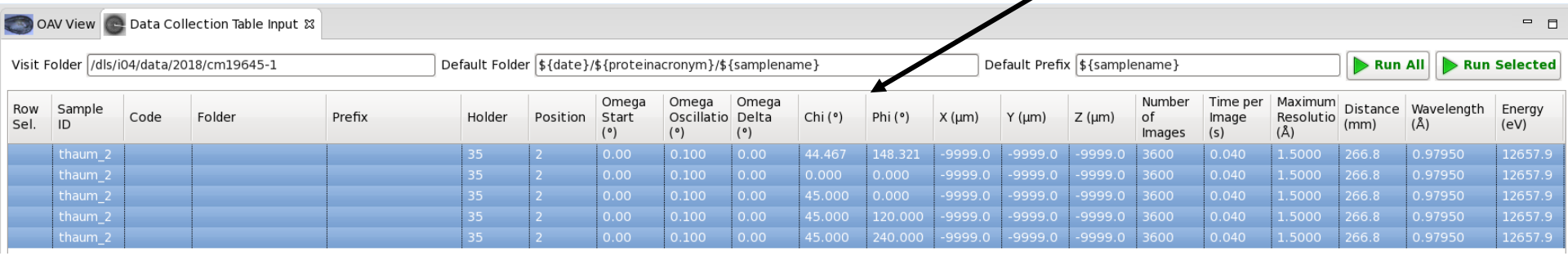


If an anomalous scattering atom is present then you can also scan a particular edge to obtain peak/inflection energies and values.

Multi-sweep SAD data

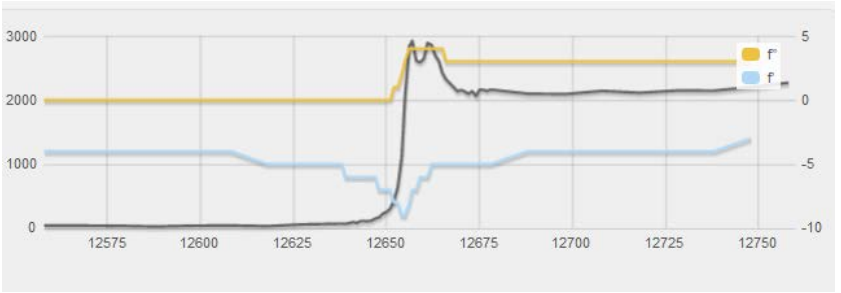
- target: high multiplicity
- collect at peak wavelength
- fluorescence scan
- single axis goniometer
  - rotate 360 deg and repeat
  - inverse beam
- multi-axis goniometer
  - use of different crystal orientations
  - aligned and random

Omega Start (°)	Omega Oscillatio (°)	Omega Delta (°)	Chi (°)	Phi (°)
0.00	0.100	0.00	44.467	148.321
0.00	0.100	0.00	0.000	0.000
0.00	0.100	0.00	45.000	0.000
0.00	0.100	0.00	45.000	120.000
0.00	0.100	0.00	45.000	240.000



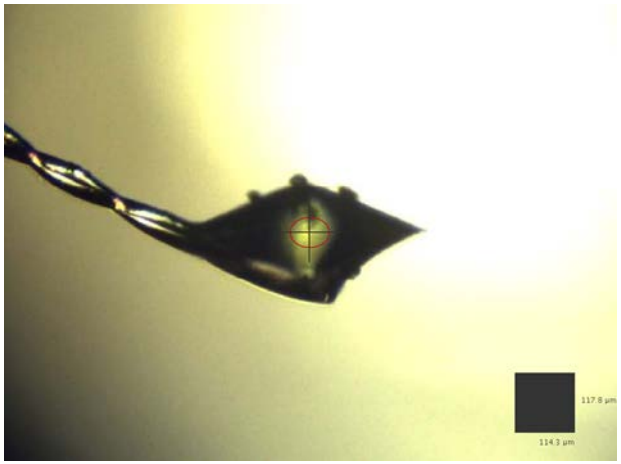
MAD data collection

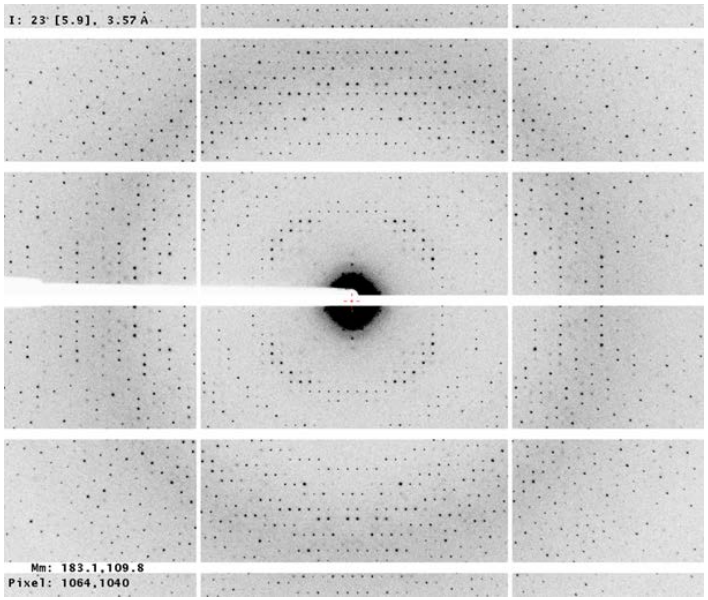
- 3 wavelengths: peak, inflection, remote
- fluorescence scan
- data as complete as possible
- single axis goniometer
  - wedged MAD (takes time)
- multi-axis goniometer
  - use of different crystal orientations
  - aligned crystal
  - Bijvoet pairs on the same frame
  - offset crystal to get better completeness



29-11-2018 15:28:00 - Se Edge Scan

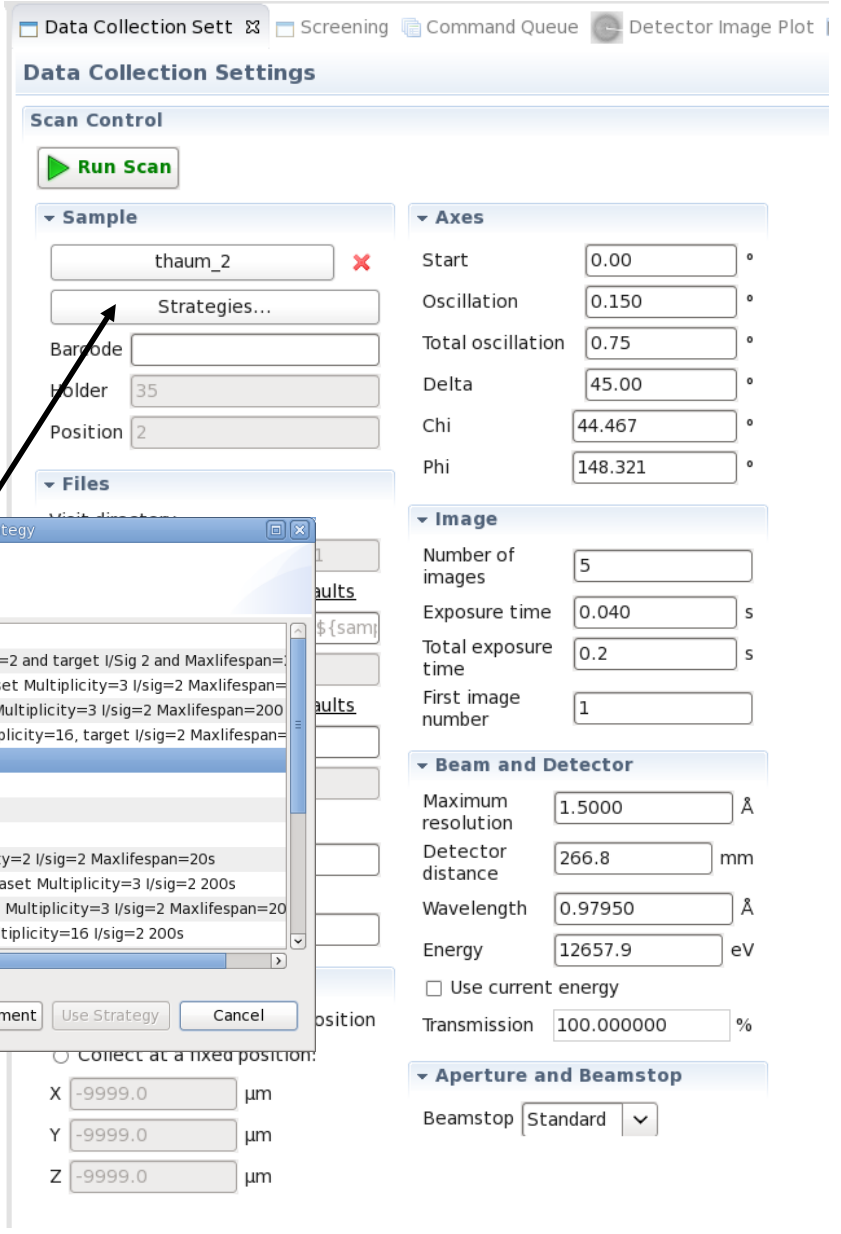
Sample: Sel-Thaum_4	Scan File: my_crystal.fluo
E(Peak): 12656.9eV (0.9796Å)	f': 4.92 / f'': -7.66e
E(Inf): 12654.9eV (0.9797Å)	f': 3.53 / f'': -9.66e
Exposure: 1s	Transmission: 1.00%
Beamsize: 31.72598x20µm	





New crystal orientation

1) take screening images
2) check indexing and alignment calculation in Synchweb
3) use alignment via GDA strategies button
4) take screening images at new crystal orientation
5) look at strategies in Synchweb
6) data collection



mosflm

Space Group	A	B	C	α	β
P4	57.68	57.68	150.07	90.00	90.0

Strategy	Description
anomalous Wedge1	
native Wedge1	

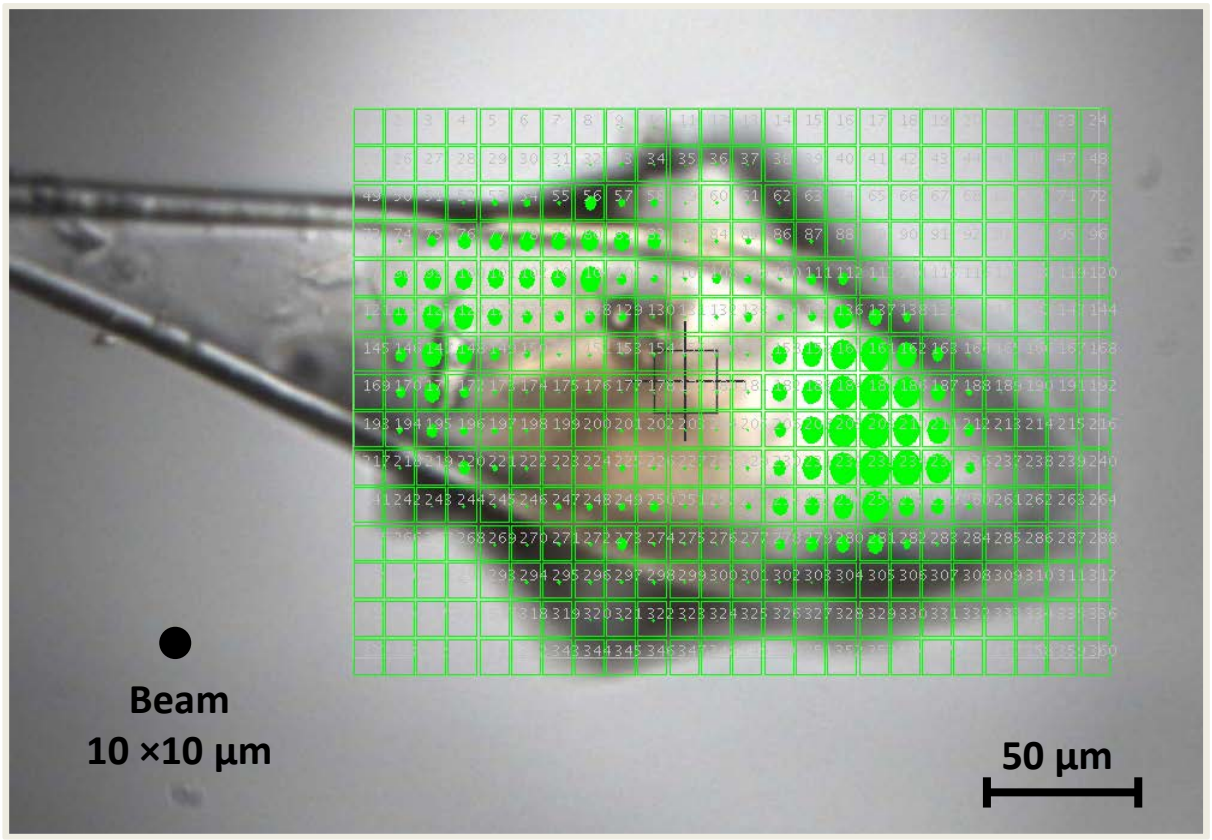
XOalign

Axes	Kappa	Chi	Phi
[(c*, b*), (c*, a*)]		44.467	148.321

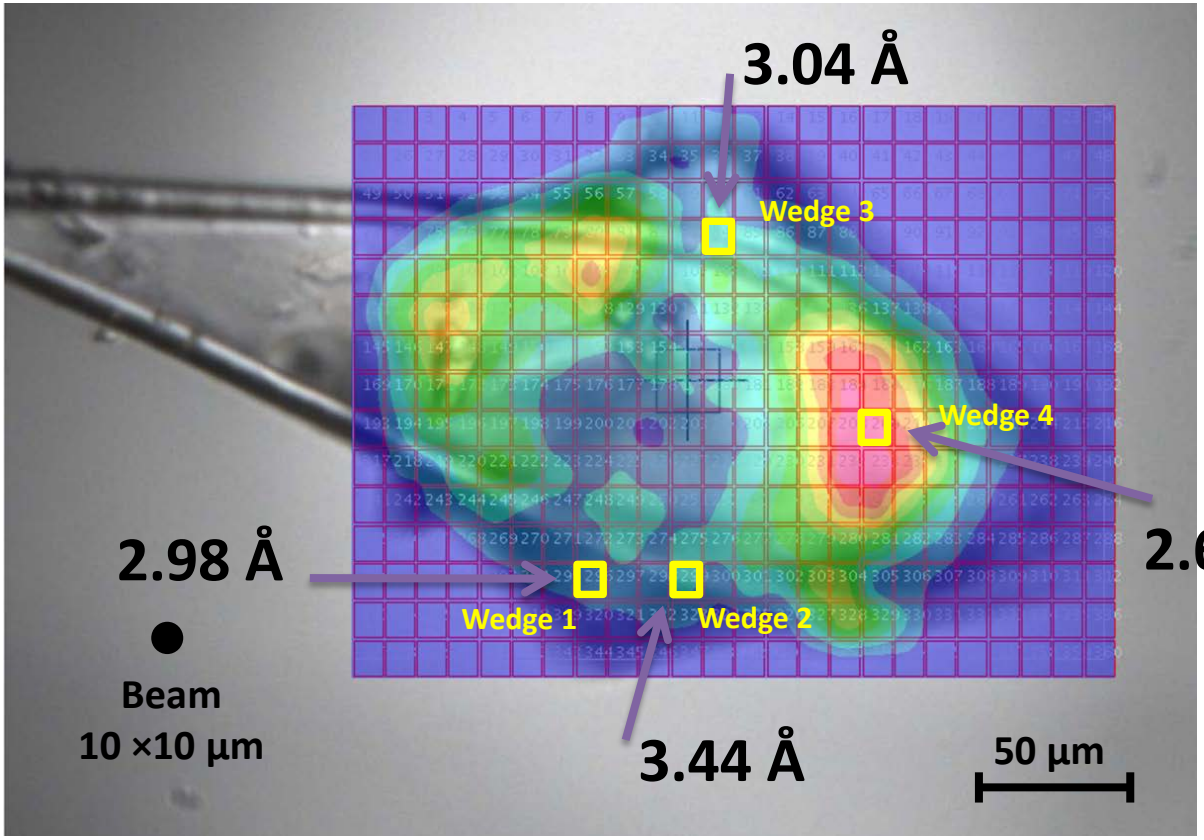
Other practicalities: does all of my crystal diffract

i) well

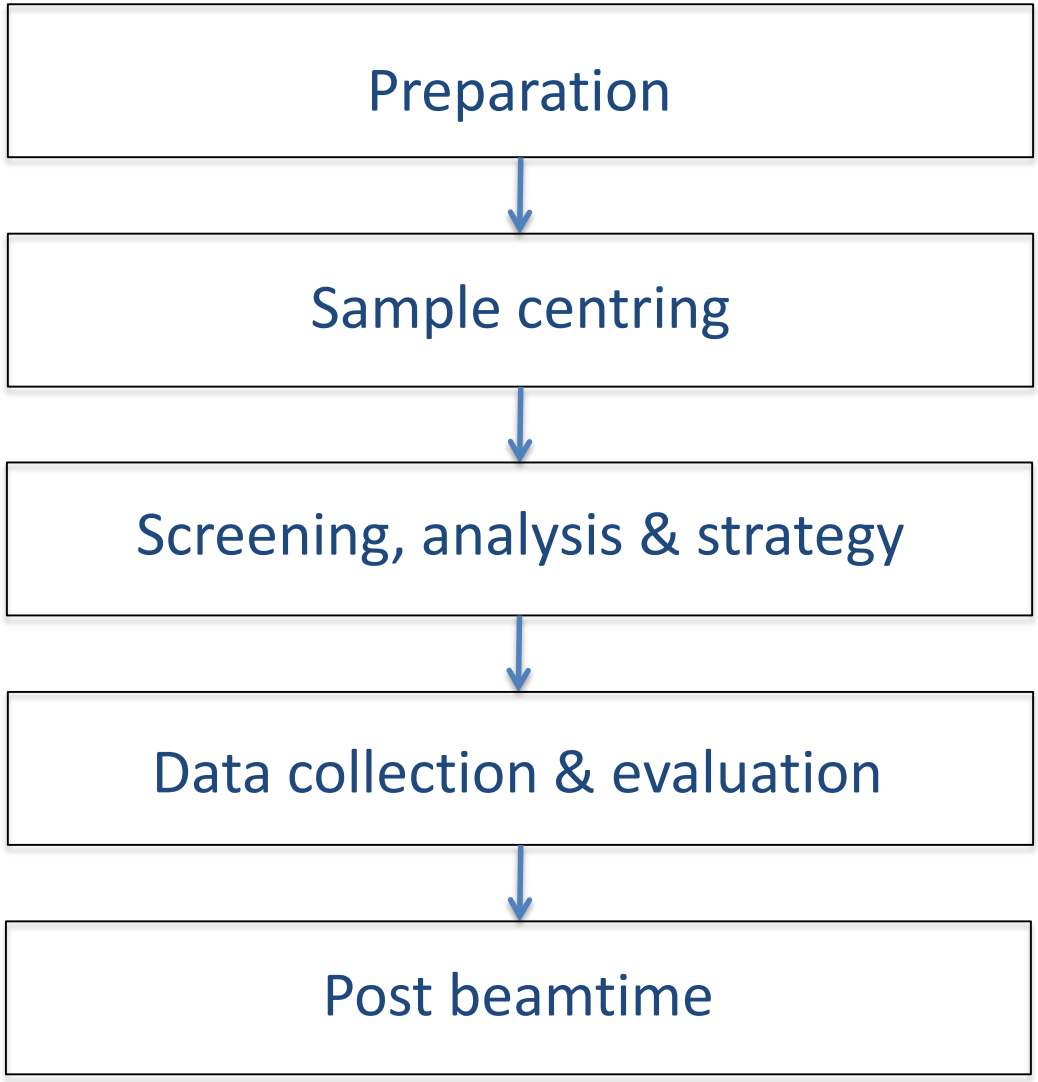
ii) badly?



Even the same crystal can show a lot of variation..



Sometimes worthwhile, sometimes not. Learn when to move on.


Preparation
Sample centring
Screening, analysis & strategy
Data collection & evaluation
Post beamtime

Decision time

Remember the aim of your experiment – will this sample give you what you need, or get you closer to your goal? If not, move on.

Experimental phasing is the most critical at this point, a trivial case can be made borderline by bad choices. Most common error is overexposing the crystal, next is not collecting enough data.

Remember that low energy = stronger diffraction and faster radiation damage. Transmission of <5% not uncommon. If the edge is very different to the screening energy then consider repeating the screening shots.

Data collection

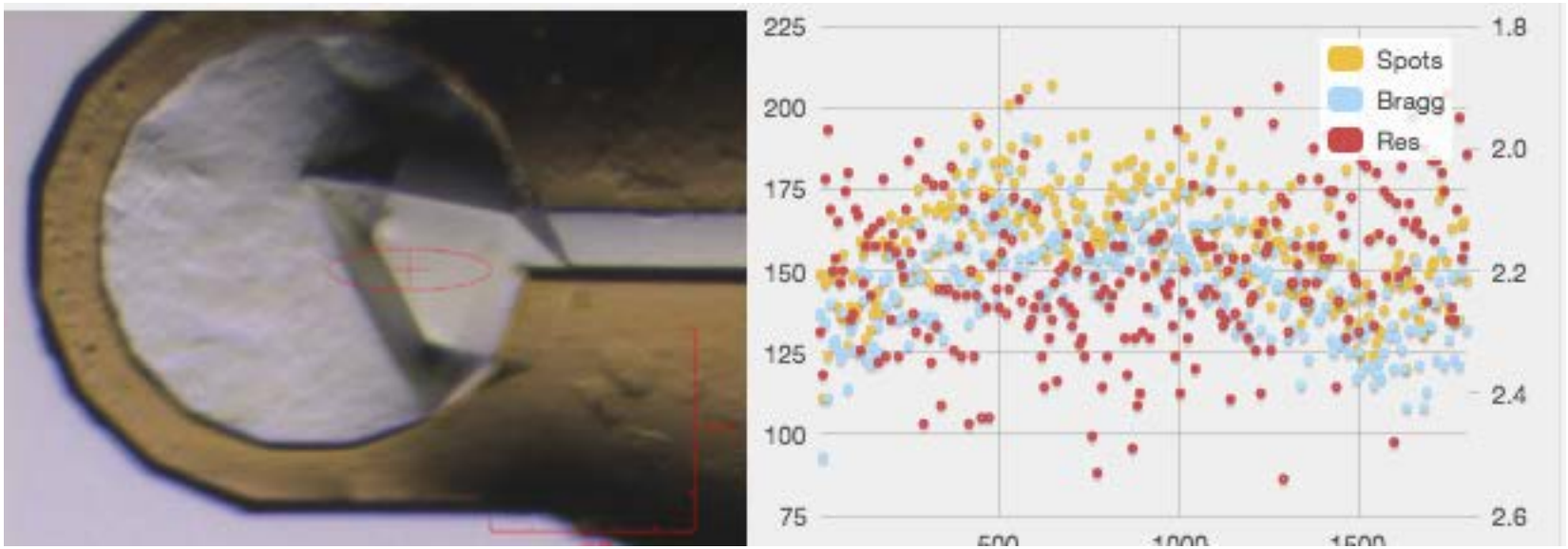
Strategies are a good place to start. Choose the appropriate one for your experiment aim.

Space Group	A	B	C	α	β	γ
P4	67.23	67.23	101.43	90.00	90.00	90.00

Lookup Cell

Strategy	Description	Ω Start	Ω Osc	Res (Å)	Ranking Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
Strategy1 Wedge1	Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=200 s	129	0.05	1.16	1.02	18.5	18.5	0.010	2200
Strategy2 Wedge1	Standard Anomalous Dataset Multiplicity=3 I/sig=2 Maxlifespan=200 s	105	0.05	1.16	1.04	24.9	24.9	0.010	3000
Strategy3 Wedge1	strategy with target multiplicity=16, target I/sig=2 Maxlifespan=200 s	0	0.05	1.16	1.02	5.8	5.8	0.010	7200
Strategy4 Wedge1	Gentle: Target Multiplicity=2 and target I/Sig 2 and Maxlifespan=20 s	129	0.05	1.16	1.10	18.5	18.5	0.010	2200
Strategy5 Wedge1	UnderDEV Anomalous Dataset, RadDamage of standard protein	105	0.05	1.16	1.04	24.9	24.9	0.010	3000

Watch the frames as they come in, look for signs of radiation damage. Per image analysis is performed automatically.



Data analysis

Process as you go, thankfully automatic these days! Fast_dp results usually keep up with data collection.

```
Fast_DP installed in: /dls_sw/apps/fast_dp/2395
Starting image: /dls/i03/data/2016/cm14451-5/20161209/protk/protk_4_0001.cbf
Running on: cs04r-sc-com09-32
Number of jobs: 18
Number of cores: 0
Processing images: 1 -> 1800
Phi range: 0.00 -> 180.00
Template: protk_4_####.cbf
Wavelength: 0.97943
Working in: /dls/i03/data/2016/cm14451-5/processed/20161209/protk/protk_4_/fast_dp
All autoindexing results:
Lattice      a      b      c  alpha   beta  gamma
     tP  67.15  67.15 101.10  90.00  90.00  90.00
     oC  95.00  95.00 101.10  90.00  90.00  90.00
     oP  67.10  67.20 101.10  90.00  90.00  90.00
     mC  95.00  95.00 101.10  90.00  90.10  90.00
     mP  67.10  67.20 101.10  90.00  90.10  90.00
     aP  67.10  67.20 101.10  90.00  90.10  90.00
Mosaic spread: 0.09 < 0.12 < 0.14
Happy with sg# 89
 67.15  67.15 101.10  90.00  90.00  90.00
------------------------------------------------------------------------------
      Low resolution  28.83  28.83   1.85
     High resolution   1.81   8.08   1.81
              Rmerge  0.081  0.036  0.152
             I/sigma  16.80  42.60   1.60
        Completeness   87.1   97.9   20.5
        Multiplicity    8.4   10.3    1.1
             CC 1/2    99.8  100.0   66.7
  Anom. Completeness   75.0  100.0    1.4
  Anom. Multiplicity    4.2    6.8    1.0
   Anom. Correlation   10.8   -5.9    0.0
               Nrefl 160500   3165    343
            Nunique  19180    307    316
          Mid-slope  0.794
                dF/F  0.070
         dI/sig(dI)  0.825
------------------------------------------------------------------------------
Merging point group: P 4 2 2
Unit cell:  67.26  67.26 101.28  90.00  90.00  90.00
Processing took 00h 00m 51s (51 s) [160500 reflections]
RPS: 3106.9
```

Analysis:

Use these results to update your data collection strategy.

Good low resolution statistics

Detector too far away, good data at edge

No anomalous signal of note

Data analysis

```
Fast_DP installed in: /dls_sw/apps/fast_dp/2395
Starting image: /dls/i03/data/2016/cm14451-4/tmp/2016-10-07/fake094409/INS1_1_0001.cbf
Running on: cs04r-sc-com09-21
Number of jobs: 18
Number of cores: 0
Processing images: 1 -> 4800
Phi range: 0.00 -> 720.00
Template: INS1_1_####.cbf
Wavelength: 1.20001
Working in: /dls/i03/data/2016/cm14451-4/processed/tmp/2016-10-07/fake094409/INS1_1_/fast_dp
All autoindexing results:
Lattice      a      b      c  alpha   beta  gamma
     cI  78.00  78.00  78.00  90.00  90.00  90.00
     hR 110.30 110.30  67.50  90.00  90.00 120.00
     tI  78.00  78.00  78.00  90.00  90.00  90.00
     oI  78.00  78.00  78.00  90.00  90.00  90.00
     oF  78.00 110.30 110.40  90.00  90.00  90.00
     mC 110.30  78.00  78.00  90.00 135.00  90.00
     aP  67.50  67.60  67.60 109.50 109.50 109.50
Mosaic spread: 0.06 < 0.07 < 0.08
Happy with sg# 197
 78.00  78.00  78.00  90.00  90.00  90.00
-------------------------------------------------------------------------------
      Low resolution  27.62  27.62   1.63
     High resolution   1.59   7.12   1.59
              Rmerge  0.067  0.045  0.783
             I/sigma  52.90 148.20   7.90
        Completeness  100.0   98.4  100.0
        Multiplicity   78.4   71.0   76.5
               CC 1/2  100.0  100.0   97.7
  Anom. Completeness  100.0   98.9   99.9
  Anom. Multiplicity   40.4   41.8   38.8
   Anom. Correlation   52.7   52.6    2.6
               Nrefl 847313   9869  61017
             Nunique  10812    139    798
           Mid-slope  1.262
                dF/F  0.021
         dI/sig(dI)  1.237
-------------------------------------------------------------------------------
Merging point group: I 2 3
Unit cell:  78.11  78.11  78.11  90.00  90.00  90.00
Processing took 00h 02m 23s (143 s) [847313 reflections]
RPS: 5892.0
```

Strong anomalous signal, fast_ep successful – note the anomalous multiplicity for sulfur SAD

Trickier cases

Low solvent content, low resolution, low anomalous signal cases may require more than 1 dataset or even crystal to solve.

Low solvent content usually requires MAD phasing, 2 or 3 energies - some excellent talks covering this during the course!

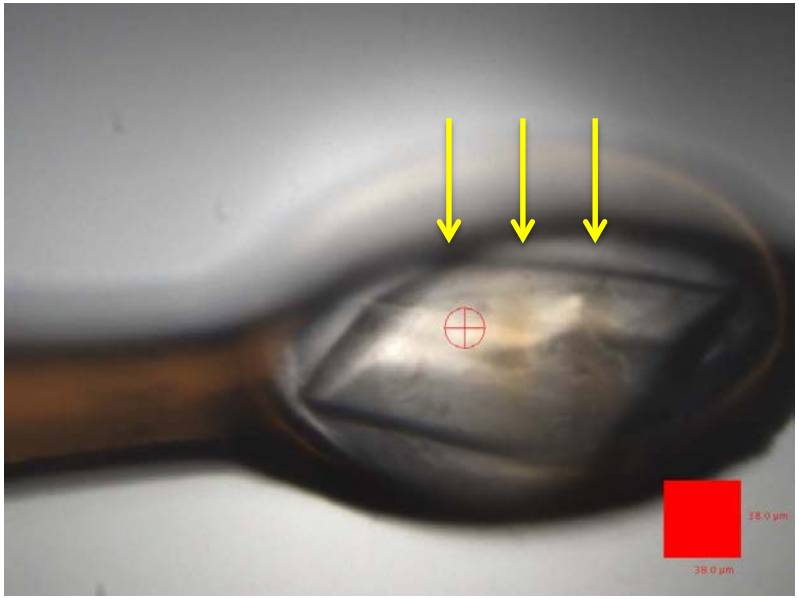
From a data collection point of view you have options here:

- Collect all on same part of crystal, interleaved. Chief concern is radiation damage.
- Collect each energy on a separate part of the same crystal, small beam and large crystal.
- Collect data from several crystals and combine (isomorphous?).

LptDE

Outer-membrane pore, beta barrel. Low resolution (3 Å) and low symmetry (I2). 20 Se per monomer and 2-3 in the ASU. 3 crystals available. 90min beamtime left.

Mainly beta strand so good experimental map needed for model building.

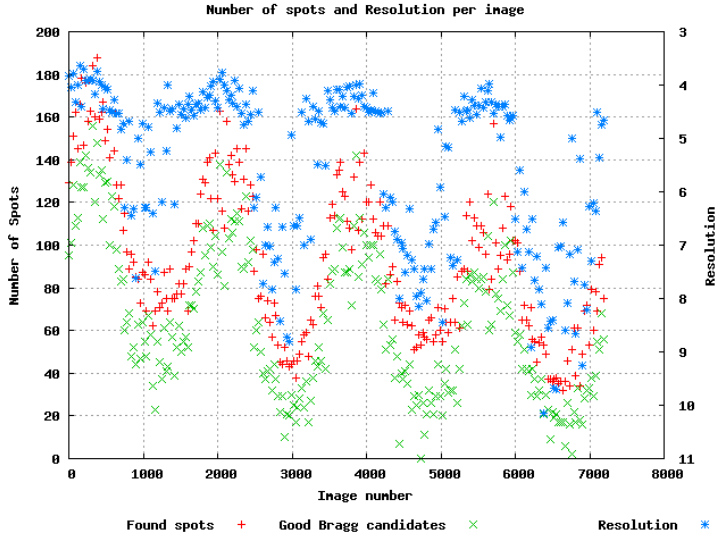


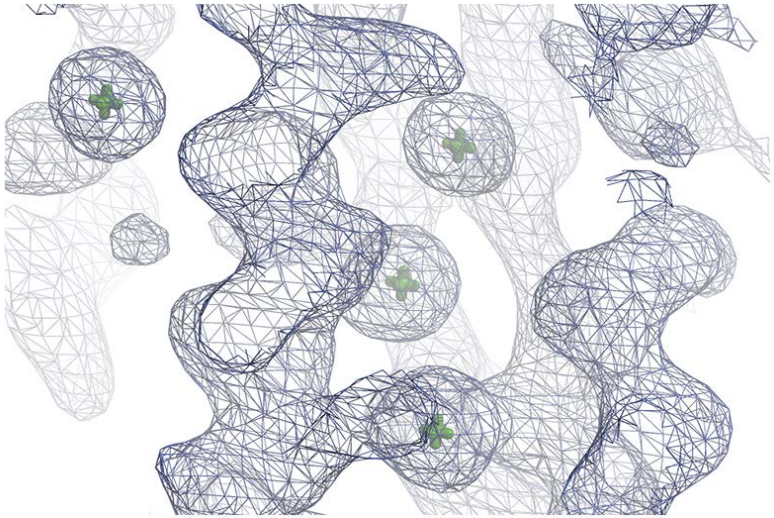
Strategy: 3-wavelength MAD using 20x20 µm beam on I24. Collect 720 deg at 1 energy, then translate, change energy and repeat. Do the same on all 3 crystals.

LptDE

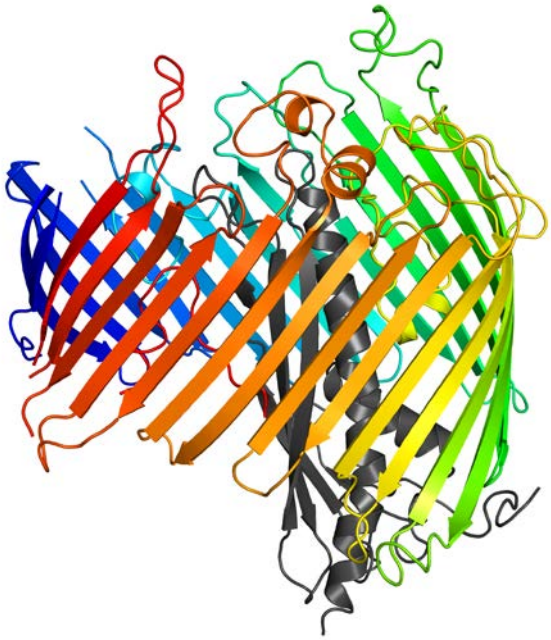
Fast_dp integration files combined in AIMLESS, last 1200 frames trimmed from each for radiation damage.

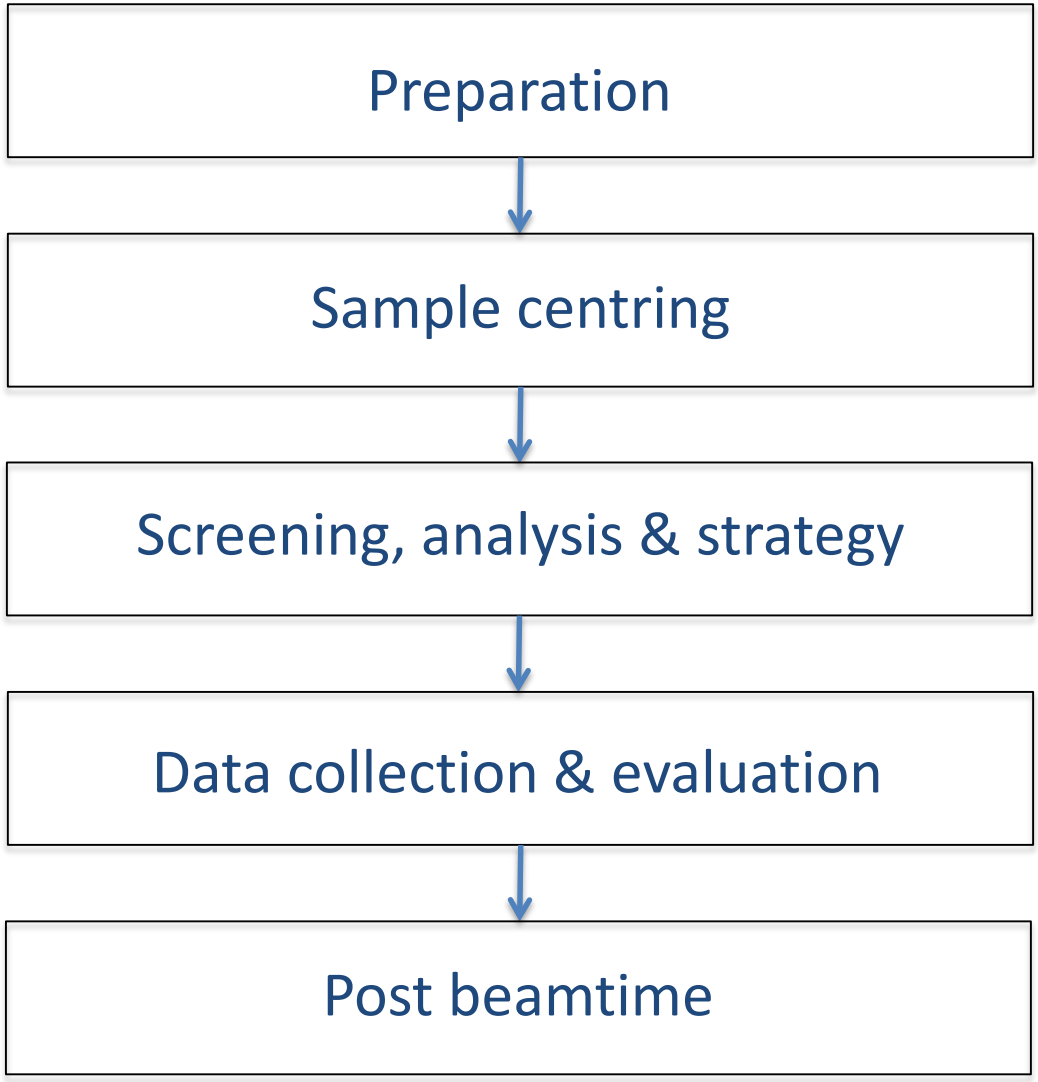
Phasing with SHARP and SHELXD, 40 Se found. 715 residues built automatically (from ~1500).





Data collection	Peak
Wavelength (Å)	0.97835
Resolution (Å)[b]	43.92-3.00 (3.09-3.00)*
Space group	I2
Cell dimensions (Å/°)	173.4, 76.1, 213.6 / 111.5
Unique reflections	52322 (4472)
Mn(I/σ(I))	15.4 (1.8)
Anomalous completeness (%)	99.9 (100.0)
Anomalous redundancy	15.9 (16.3)
R_{meas}	0.21 (2.96)
CC(1/2)	0.999 (0.798)




Preparation
Sample centring
Screening, analysis & strategy
Data collection & evaluation
Post beamtime

Post beamtime

Ideally you'd want to confirm on the beamline that your structure is solved and gives you what you need. Doesn't always happen.

Check structure solution pipelines, FastEP, BigEP, Dimple, MrBump

Post beamtime is your chance to take your time and do things carefully. Reprocessing, exclusion of radiation damage, data combination etc. Don't assume that autoprocessing failure means your data is useless.

How to do this is what you'll learn through the rest of this course