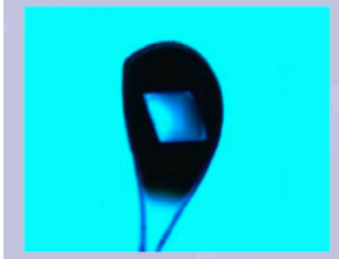
Practical MX data collection

David Aragao
David.aragao@diamond.ac.uk

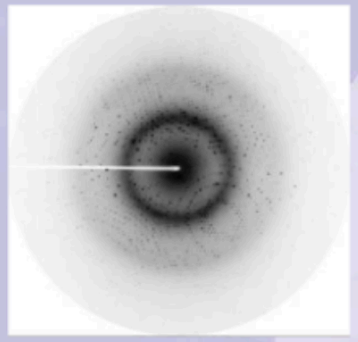
Thanks to
Neil Paterson
Gwyndaf Evans
Ralf Flaig
Daniela Stock
Murray Stuart
Meitian Wang
Tom Caradoc-Davies
for different figures and slides.

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}\left|F_{hkl}\right|\exp(i\alpha_{hkl})\exp\left[2\pi i(hx+ky+lz)\right]$$

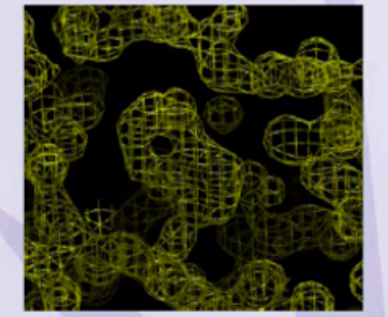
Crystal



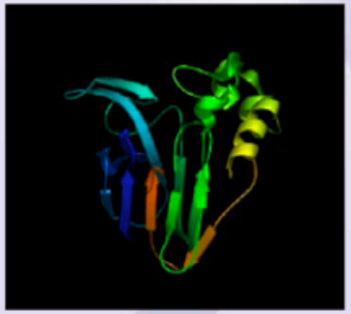
Diffraction pattern



Electron density



Structure



Diffraction Experiment

Phasing

Model Building
Refinement

$$I(hkl) = I_0 \cdot \frac{\lambda^3}{\omega V_{cell}^2} \cdot V_{cr} \cdot L \cdot P \cdot T_r \cdot r_e^2 \cdot |F_{hkl}|^2$$

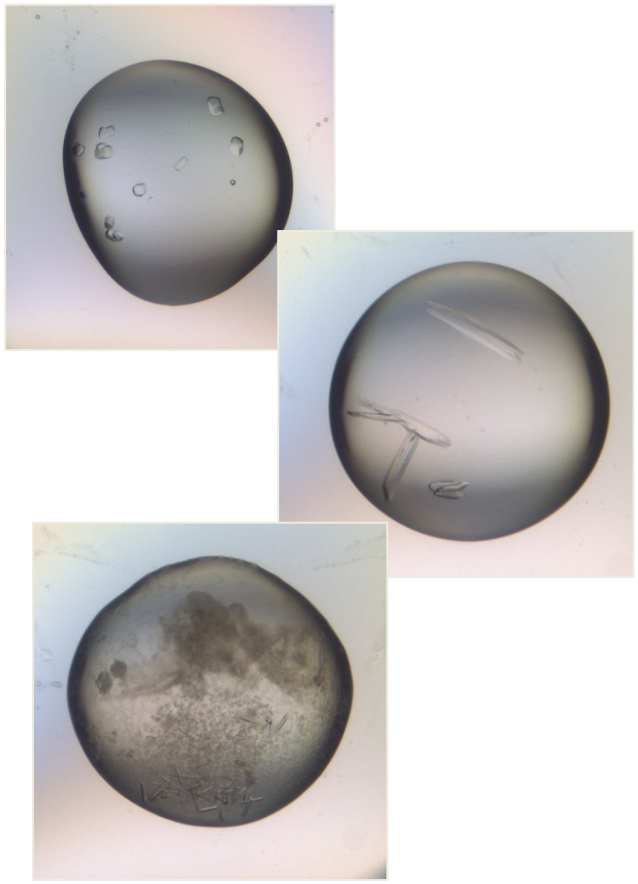
Data collection is important!

This is the last experimental step

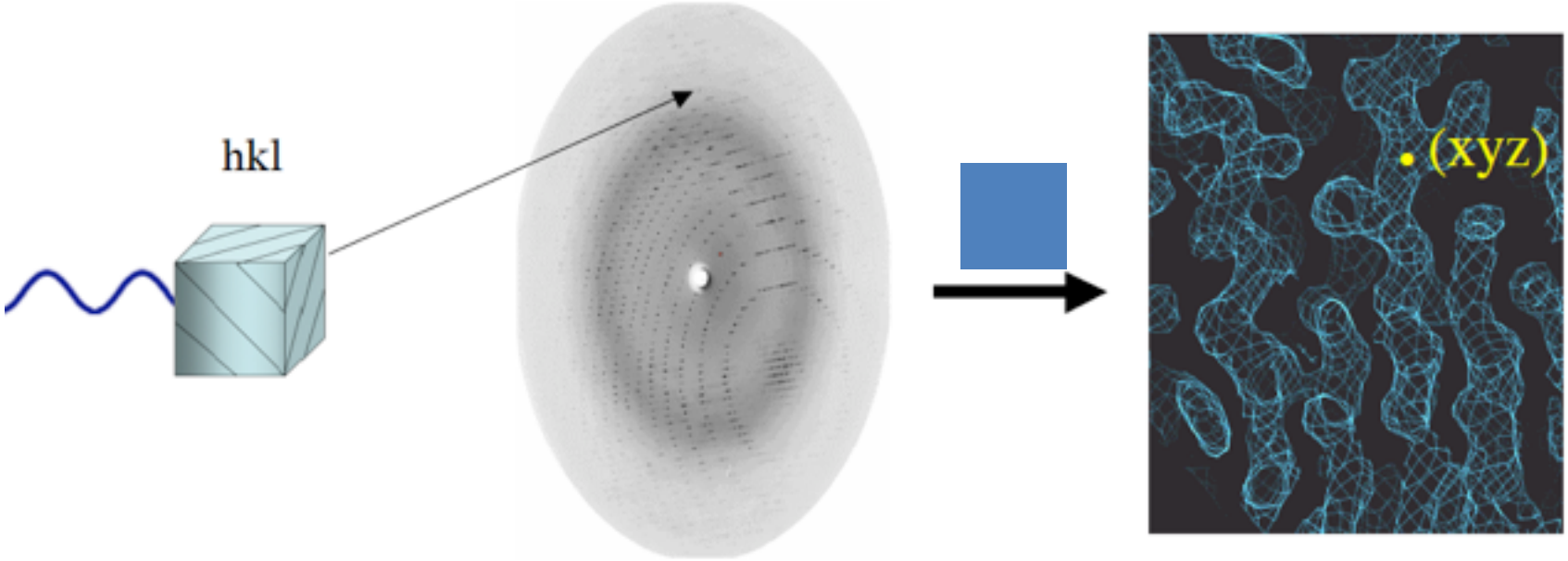
Time taken now to collect the best possible data can prevent pain and possible failure later on.

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. During the day don't be afraid to reach out. We are here to help.



Data collection: the goal



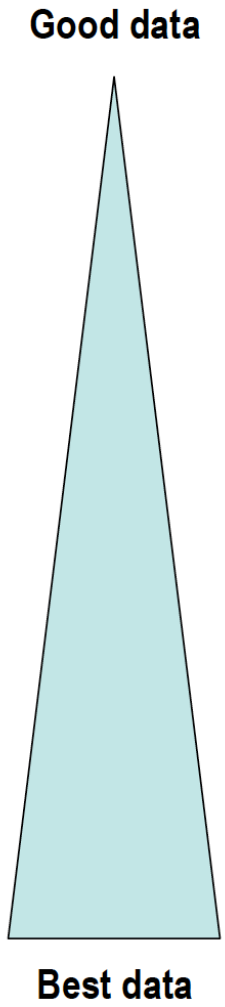
- record complete set of X-ray diffraction intensities from a crystal
- the better the data (high resolution, high completeness, low noise), the easier the following steps and the better refinement works
- like all previous steps this might need some optimisation and a lot of patience

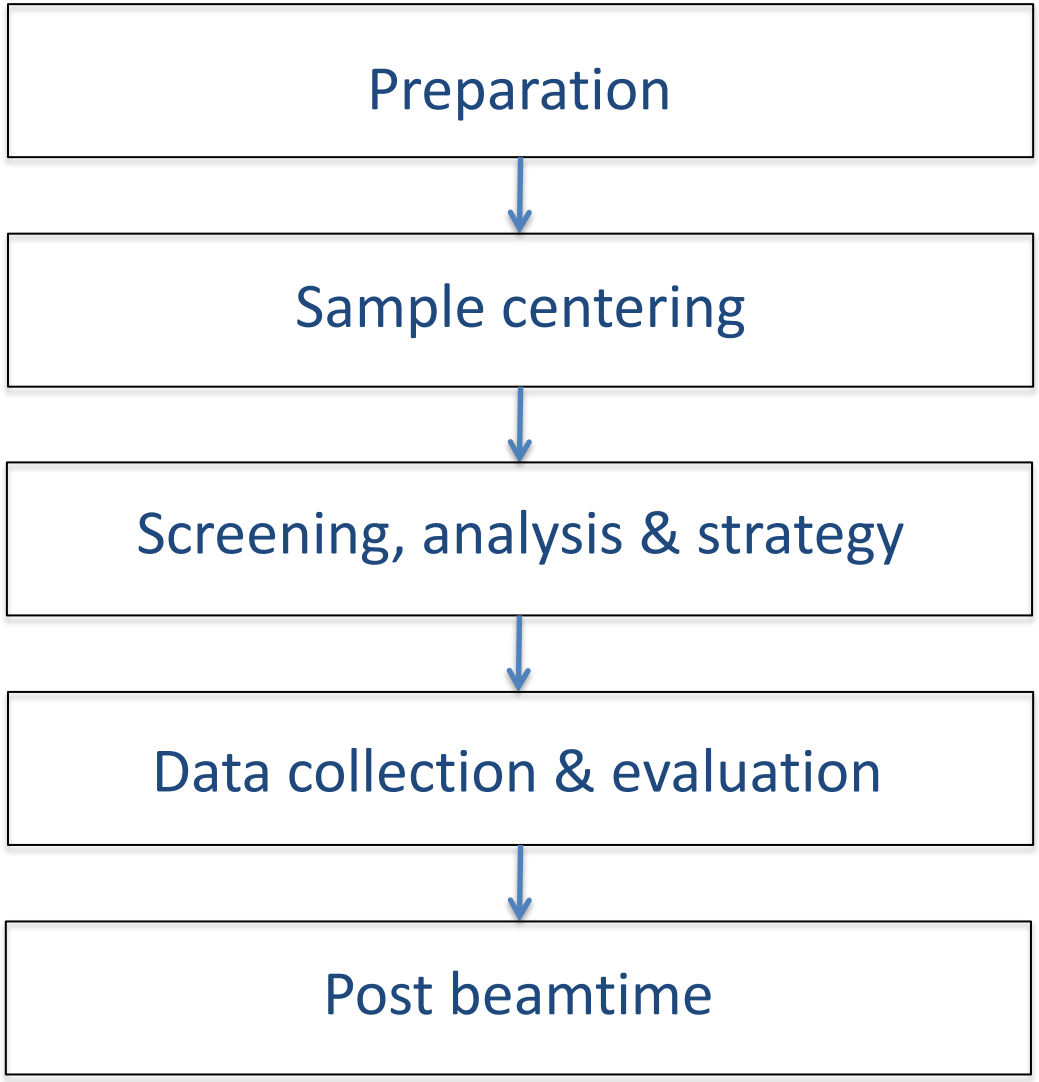
What's the goal?

Aim	High priorities	Lower priority	Max Dose per dataset (e.g. per 120, 180 or 360 degrees)
Native data collection	High resolution (e.g < 1.8 A) Complete data	High redundancy	20 MGy
Multi-wavelength anomalous dispersion (MAD) Single Anomalous dispersion (SAD)	Accurate, complete and highly redundant Very little radiation damage Choice of wavelength	High resolution data	1 MGy
Sulphur SAD	As above + extra high redundancy (40-200fold) if not on I23		0.3-0.5 MGy
Molecular replacement	Completeness at low resolution (< 3Å) Good quality low resolution	High resolution data High redundancy	20 - 30 MGy
Ligand/mutation	Medium resolution (e.g. 2 A), complete data. High throughput	High redundancy	15-20 MGy

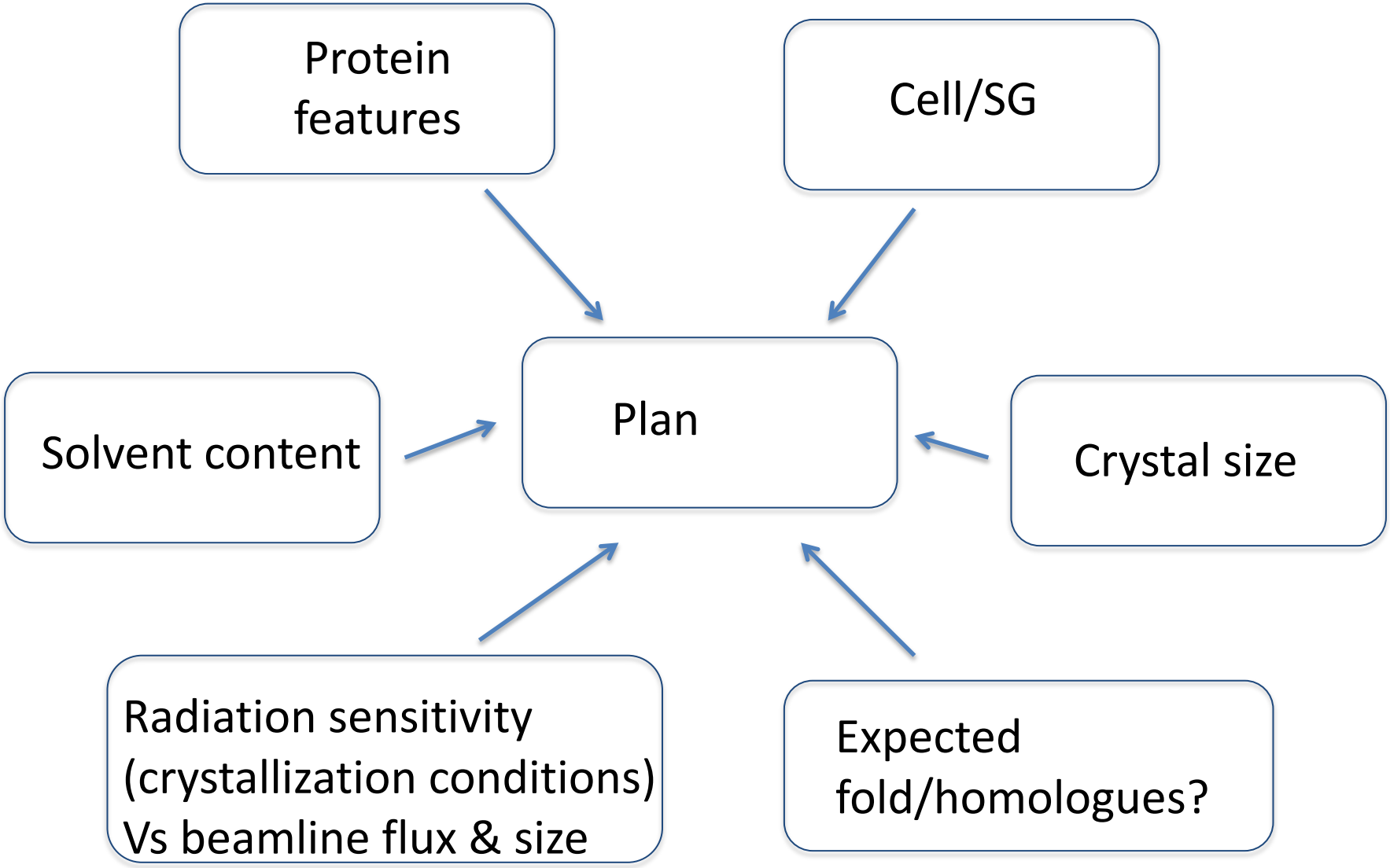
Trying to collect everything at once will usually result in failure

- Molecular substitution
- Molecular replacement
- Isomorphous replacement (Hg, Pt)
- MAD/SAD phasing
- Native-SAD phasing (Sulfur-SAD)




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

Preparation - Prior knowledge



Data collection - crystal mounting

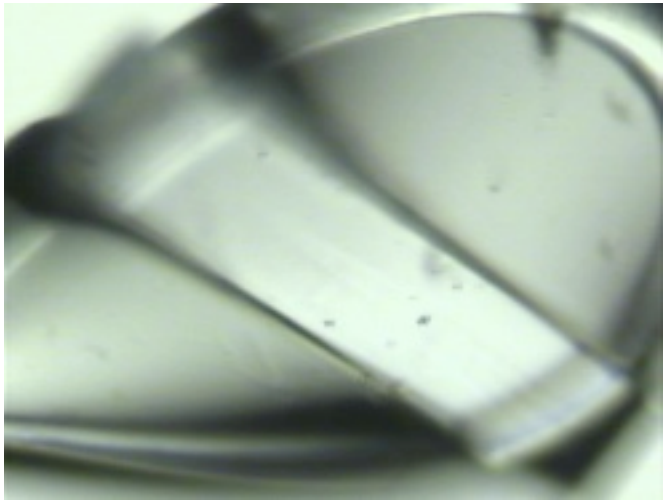
• if crystallisation condition does not contain cryo protectant, xtals need to be transferred to appropriate solution (e.g. add 20% glycerol, ethylene glycol, etc to crystallisation solution) before freezing; see A. Evdokimov's protocol for example https://mcl1.ncifcrf.gov/nihxray/Tips-and-Tricks_CryotectionOfDelicateCrystals-ArtemEvdokimov.pdf

• test new cryoprotectant in X-ray beam / cryo-stream before mounting crystal

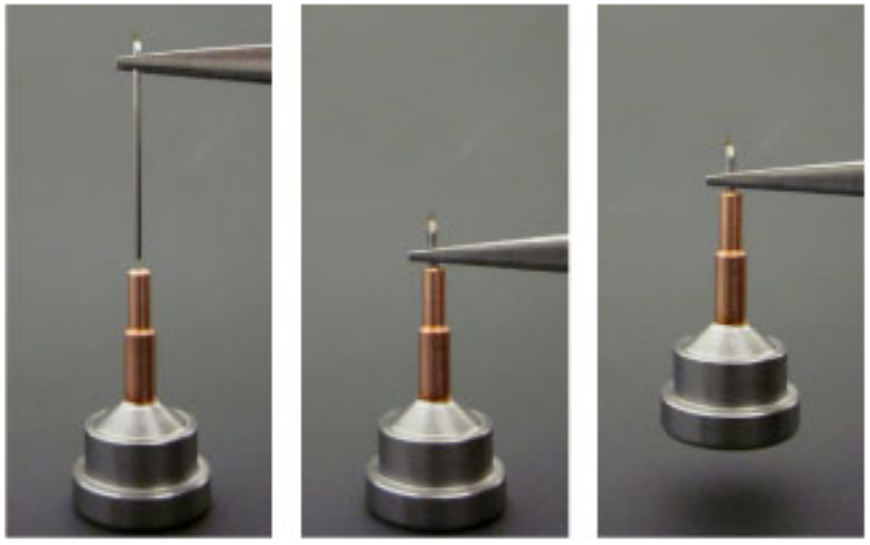
• consider mounting + shooting crystals at room temperature (particularly on a home source) to know how well they diffract without potentially damaging cryoprotectants (MiTeGen MicroRT sleeves)

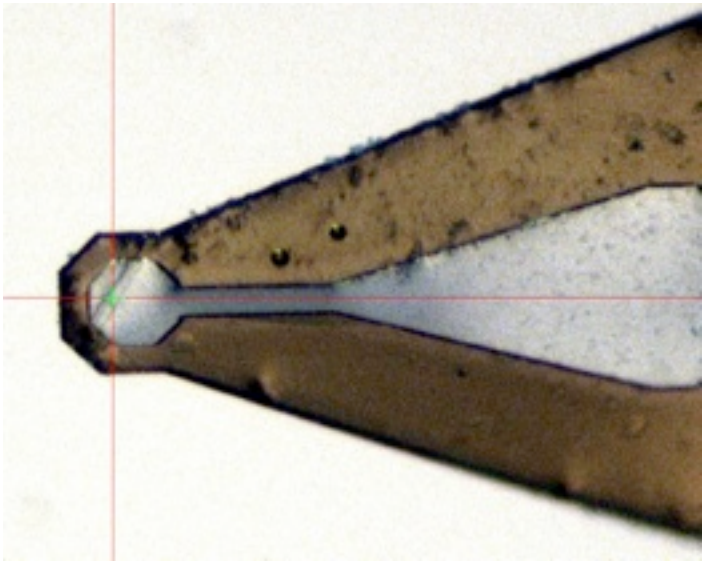


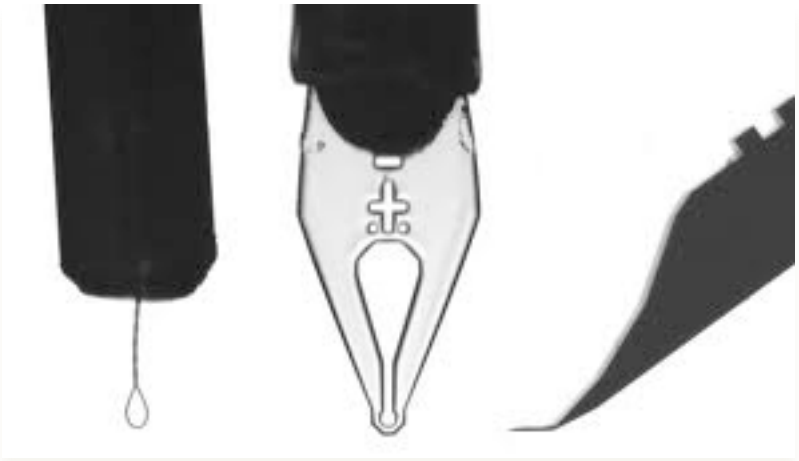


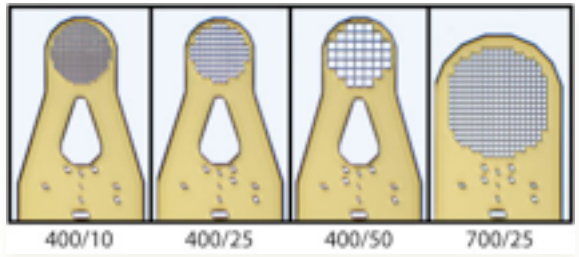


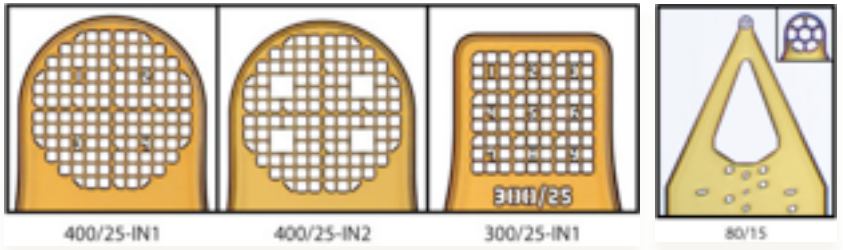
Data collection - crystal mounting





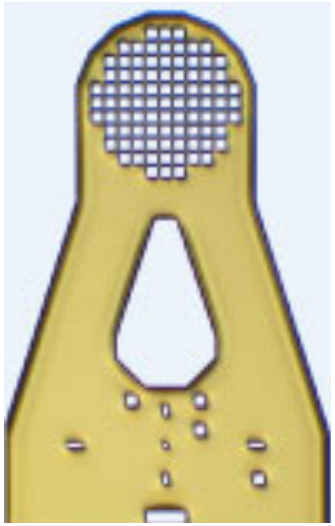




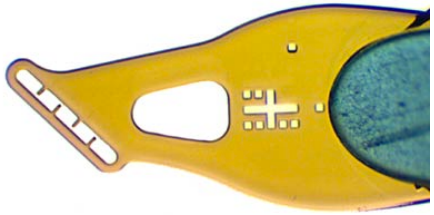


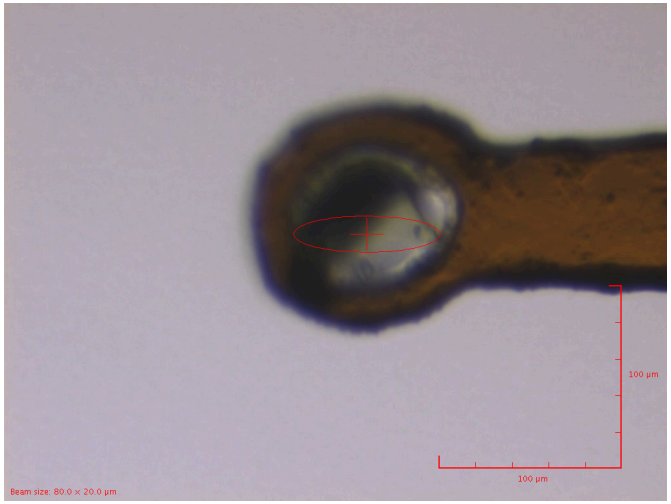
http://www.mitegen.com/

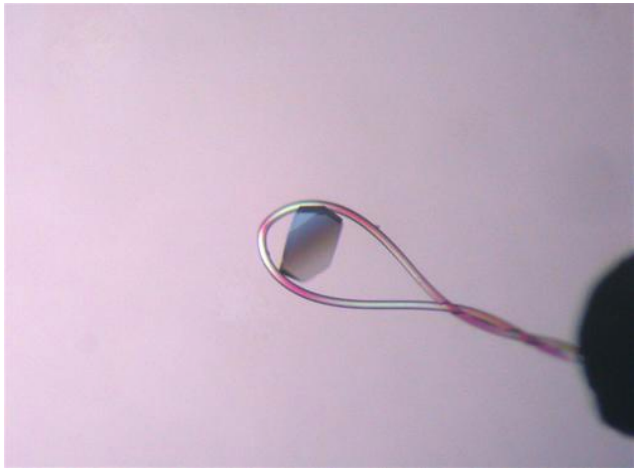
Preparation - Sample mounting

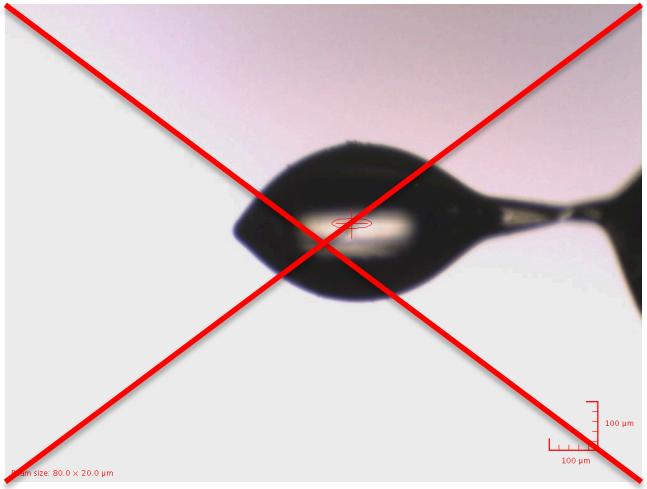












Preparation - Choice of beamline

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 (seconds exposure for max dose)
– important for radiation damage

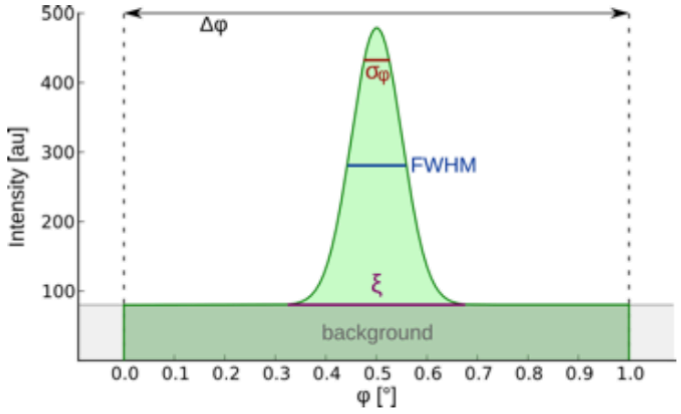
Random errors, counting statistics

$$\sigma_{count} = N^{1/2}$$

$$I = N_p - N_b \qquad \textbf{Signal is the difference}$$

$$\sigma_I = (\sigma_p{}^2 + \sigma_b{}^2)^{1/2}$$

$$\sigma_I = (N_p + N_b)^{1/2} \qquad \textbf{Uncertainty is the sum}$$



Relative errors (%) for hypothetical integrated intensity and background values.

Example: Intensity = 100, Background = 500, therefore Peak = 600 and $R_{err} = \sigma(I)/I = (600 + 500)^{1/2}/100 = 33.2\%$.

Integrated intensity (photons)	Background (photons)		
	0	500	1000
100	10.0	33.2	45.8
500	4.5	7.7	10.0
1000	3.3	4.5	5.6
10000	1.0	1.0	1.1

Pflugrath, *Acta Cryst*. **D55**, 1718 (1999)

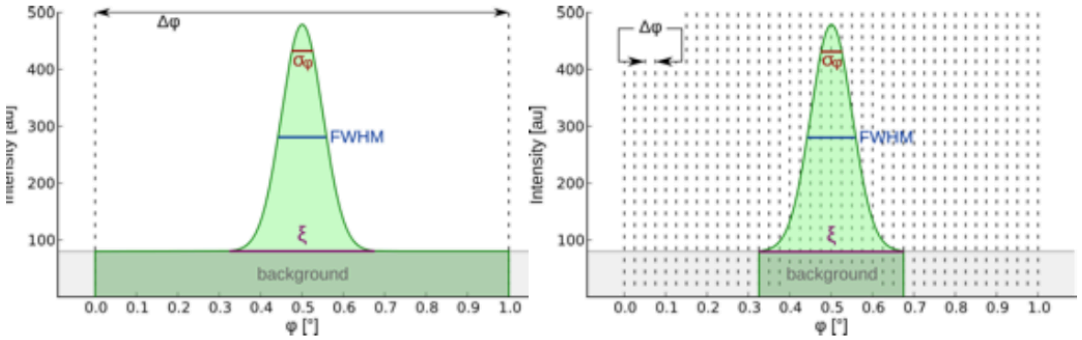
Random errors, counting statistics

$$\sigma_{count} = N^{1/2}$$

$$I = N_p - N_b \qquad \textbf{Signal is the difference}$$

$$\sigma_I = (\sigma_p{}^2 + \sigma_b{}^2)^{1/2}$$

$$\sigma_I = (N_p + N_b)^{1/2} \qquad \textbf{Uncertainty is the sum}$$

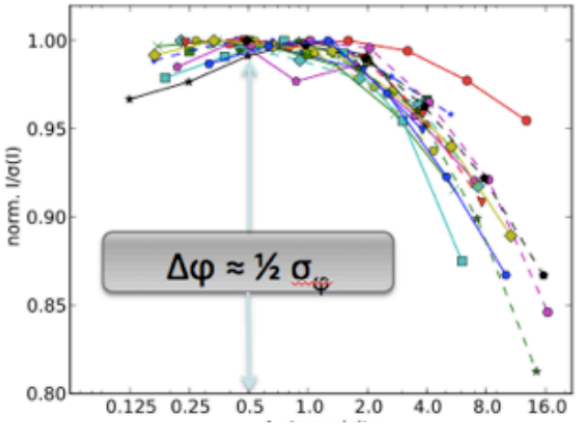


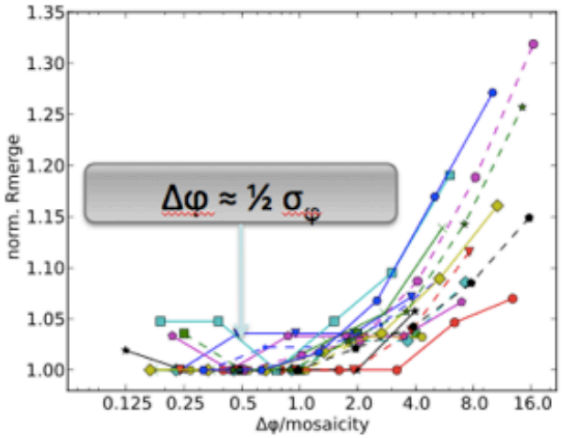
Wide φ-slicing

- Large $\Delta\varphi$ ($\Delta\varphi > \xi$)
- Large overlap of reflections and background along φ
- Few images

Fine φ-slicing

- Small $\Delta\varphi$ ($\Delta\varphi << \xi$)
- Minimal overlap of reflections and background along φ
- Many images

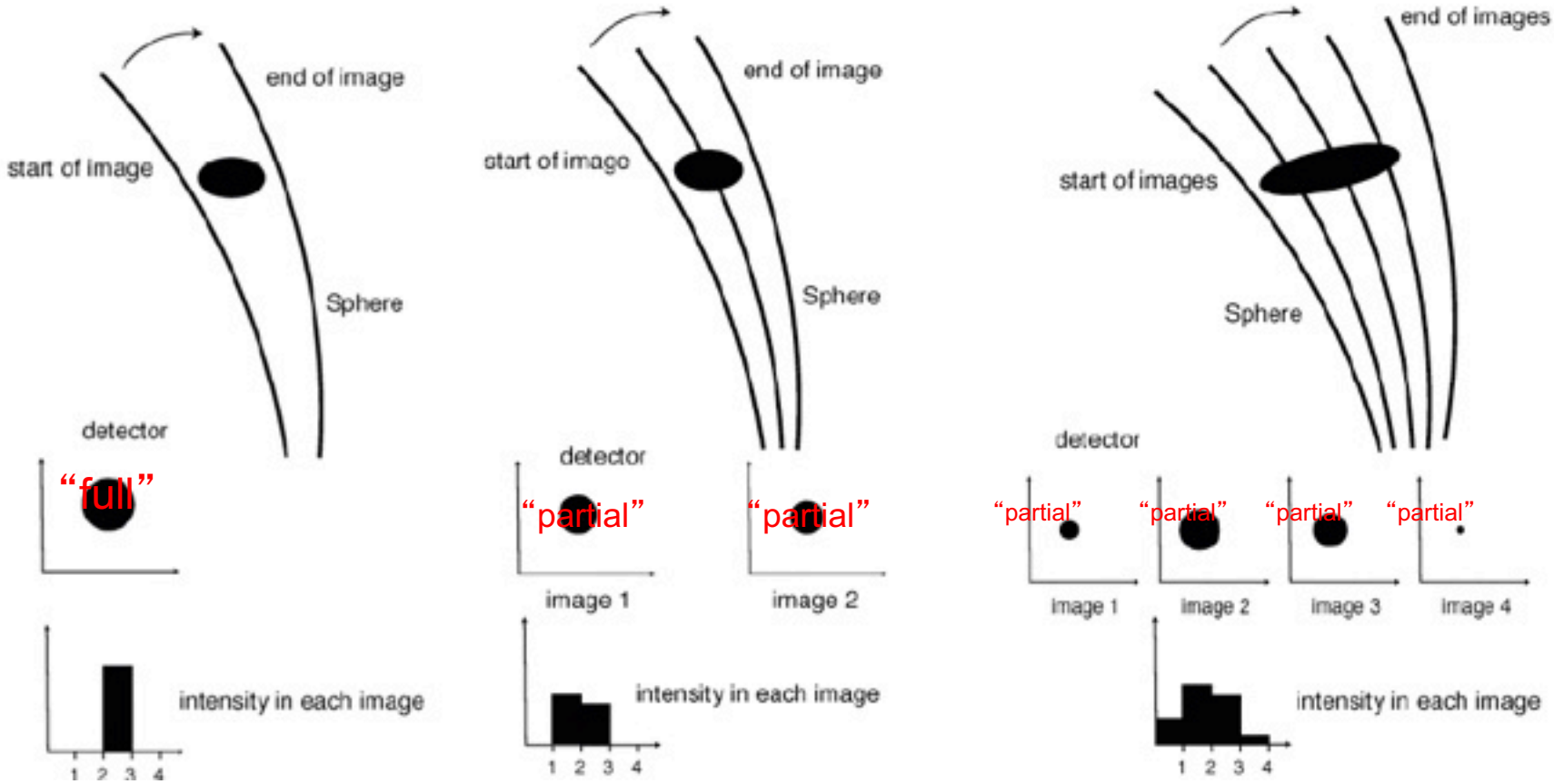




Fine-phi slicing data collection is enabled by the pixel array detector (PILATUS, EIGER), which has single-photon sensitivity and no readlout noise

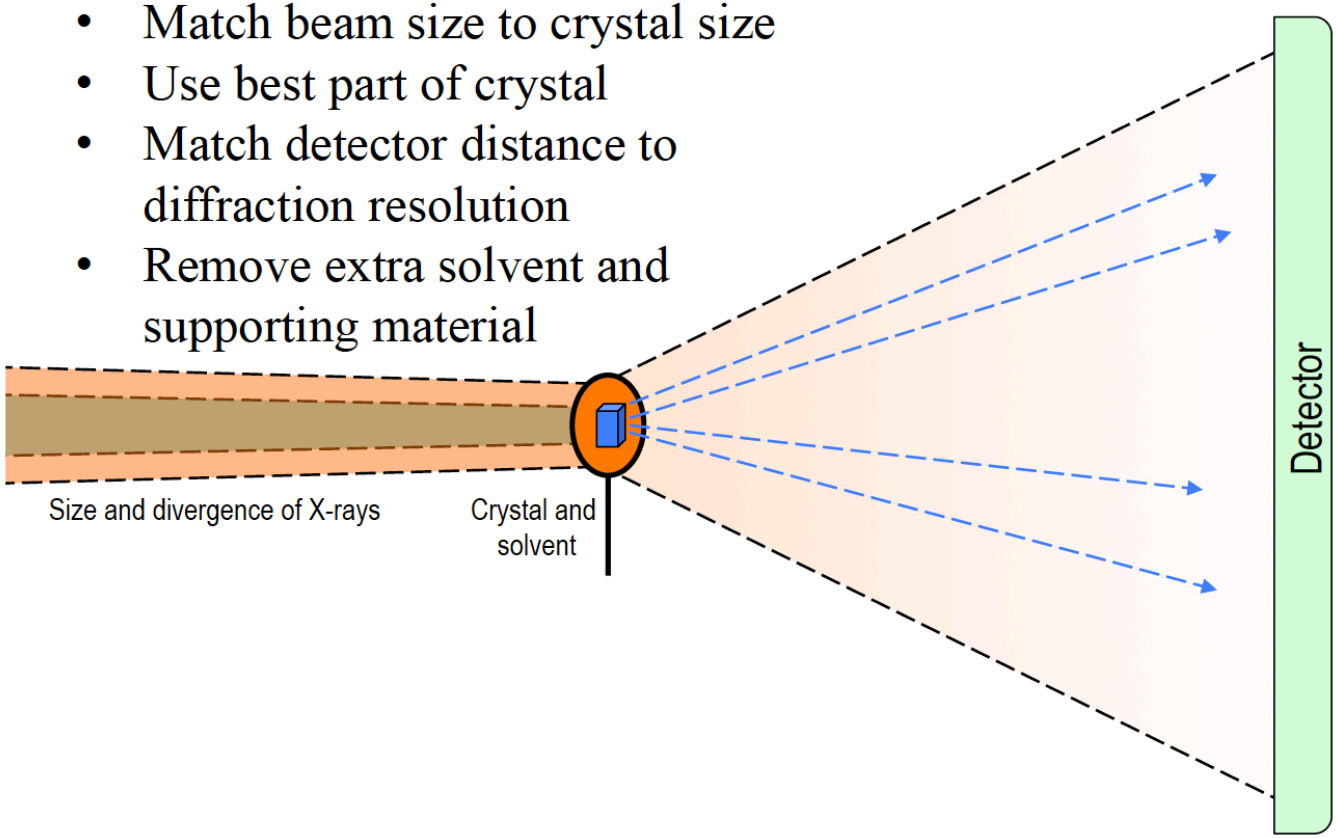
Mueller, *et al*, *Acta Cryst*. **D68**, 42 (2012)

Data collection: the rotation/oscillation method

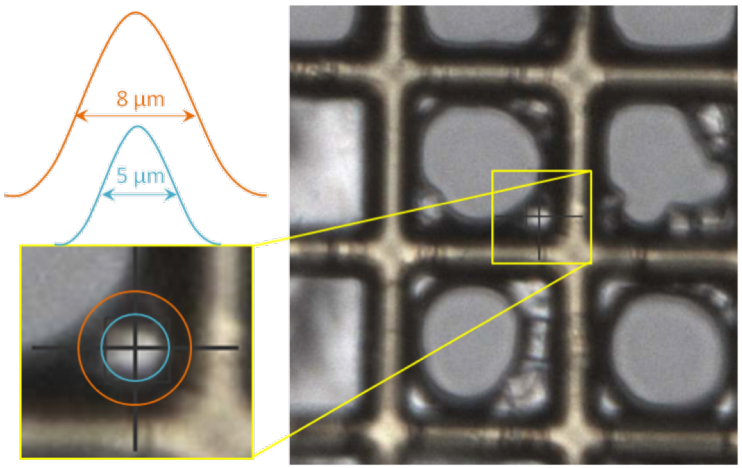


- spot size is equally affected by mosaicity, beam size and beam divergence
- programs that work with 3-d profile fitting (e.g. XDS) are better for dealing with "partials" spread over many images

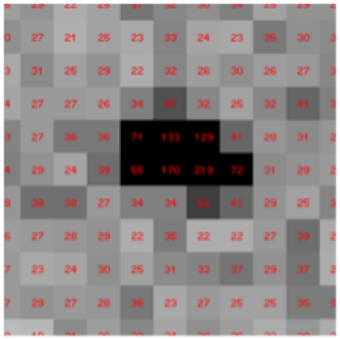
- Match beam size to crystal size
- Use best part of crystal
- Match detector distance to diffraction resolution
- Remove extra solvent and supporting material

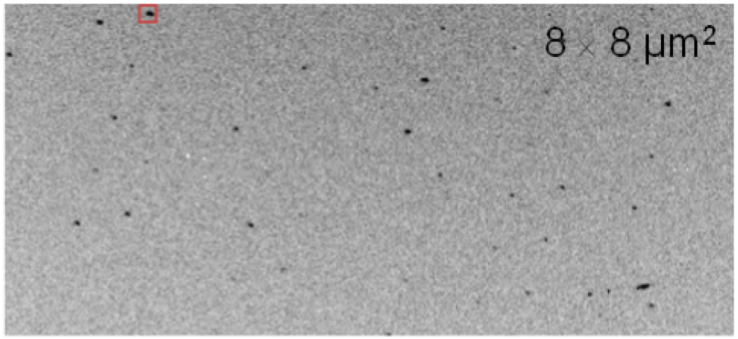


Match beam to sample

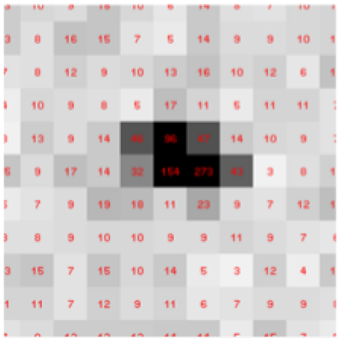


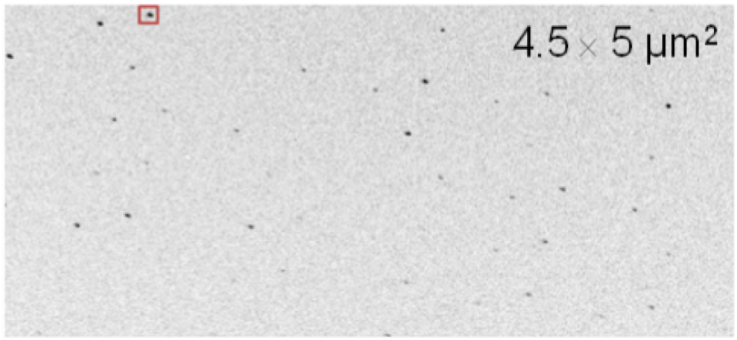
AcMNPV polyhedra crystals approx 5×5×5 μm^3 in size were mounted on a micromesh grid. Data were collected with two beamsizes: 8×8 μm^2 and 4.5×5 μm^2





I/ σ(I) = 1.5 at 2.9Å





Average background reduced by 3

I/ σ(I) = 1.5 at 2.5Å

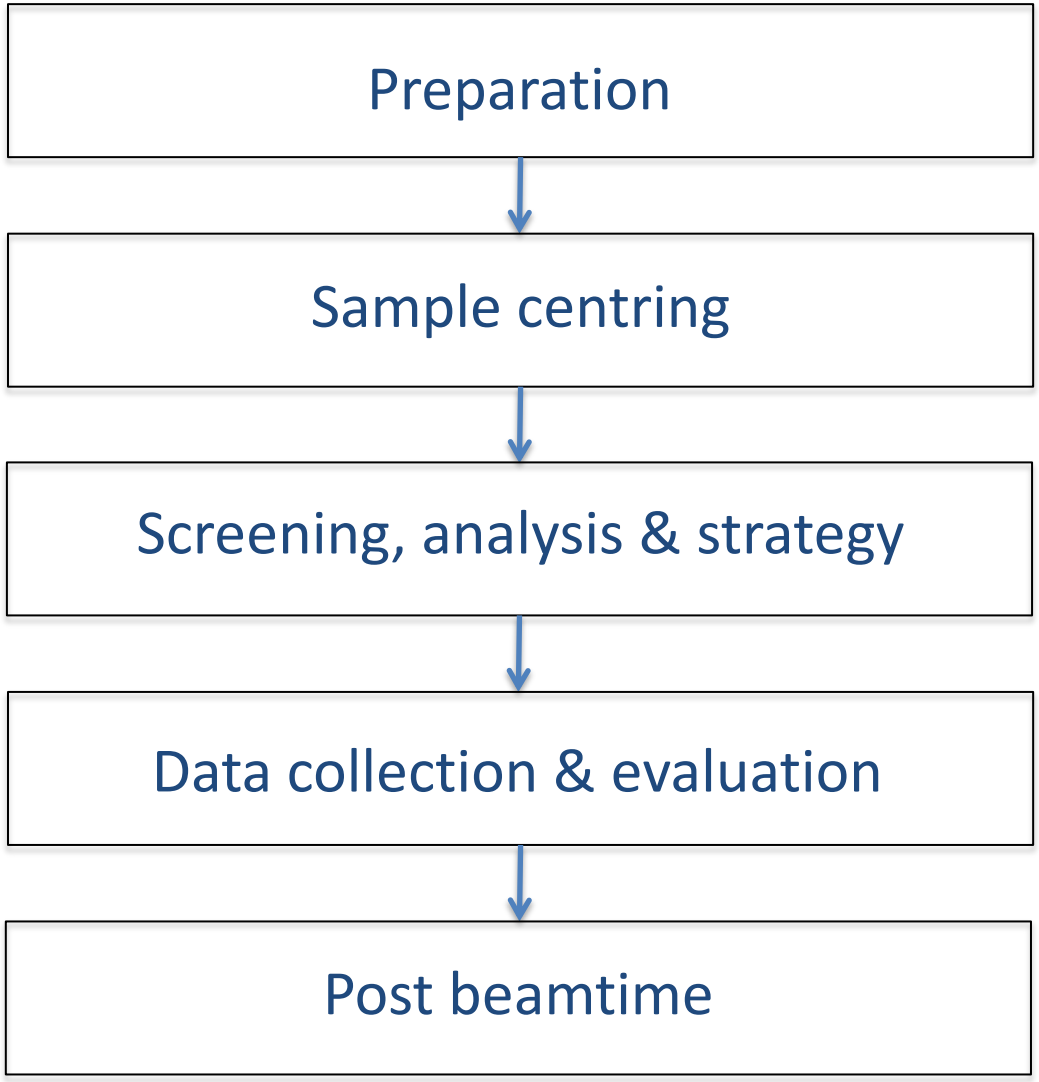
Beam size and flux density

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.

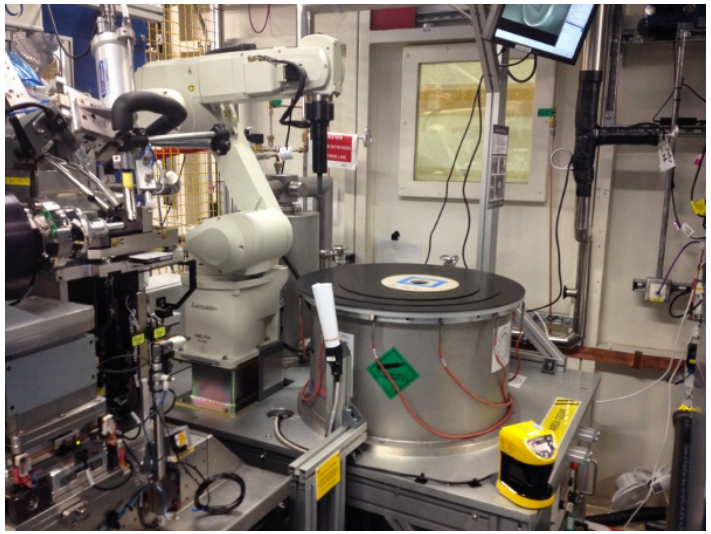
E.g. 50x50 μm^2 beam will deliver around 45x less dose than 9x6 μm^2 beam on I24

So match beam to crystal, but attenuate when required

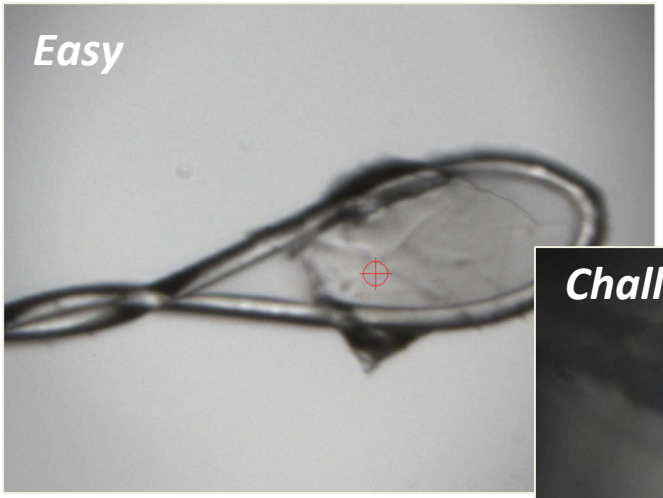

Preparation
Sample centring
Screening, analysis & strategy
Data collection & evaluation
Post beamtime

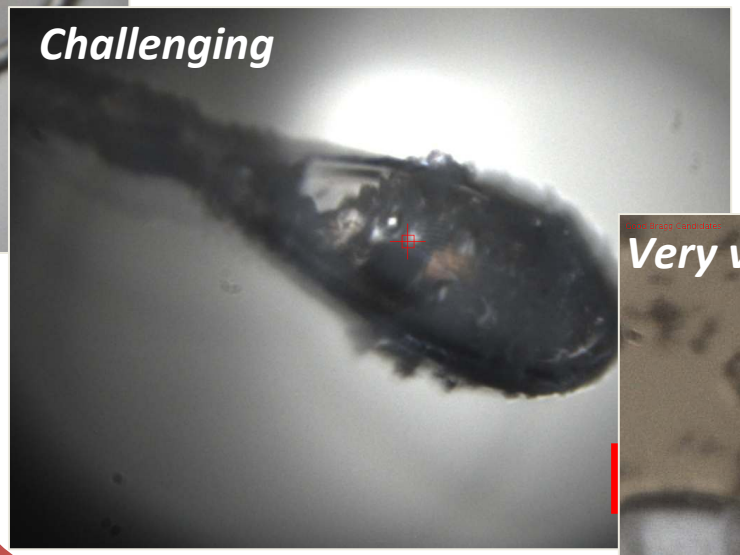
Mount and centre crystal

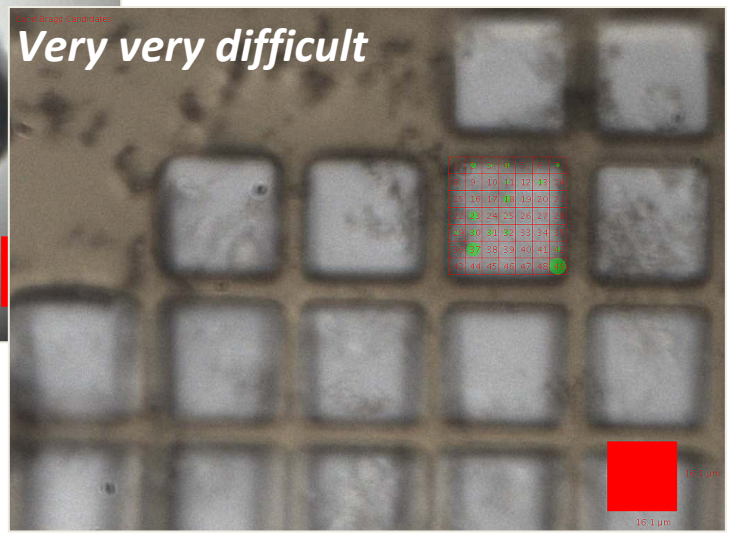
Mounting now automated at most beamlines.



Centring can be:



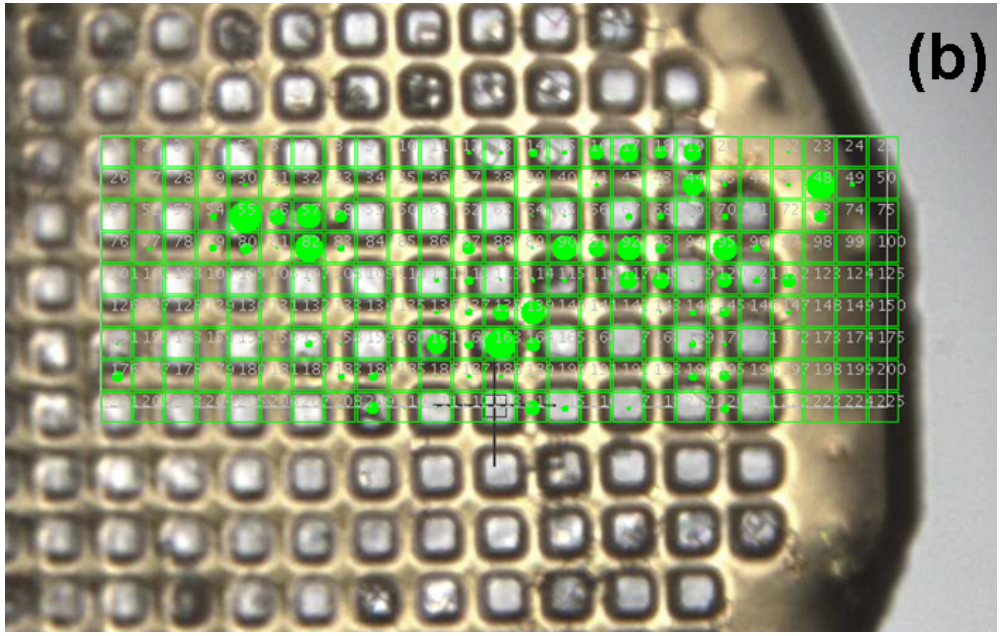




Make the most of tools developed to make these cases easier

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

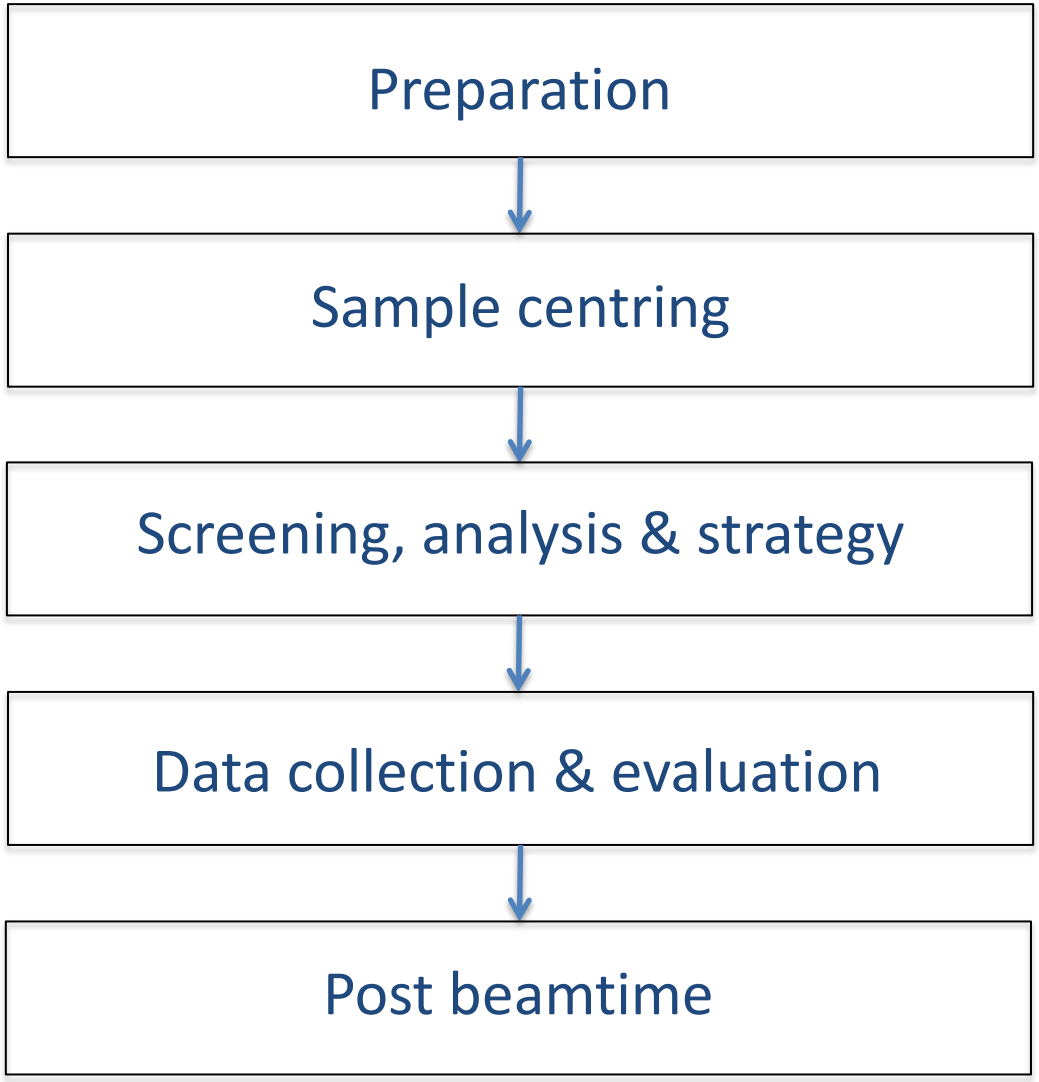
If no ⇒ Move on

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

Data collection - think!

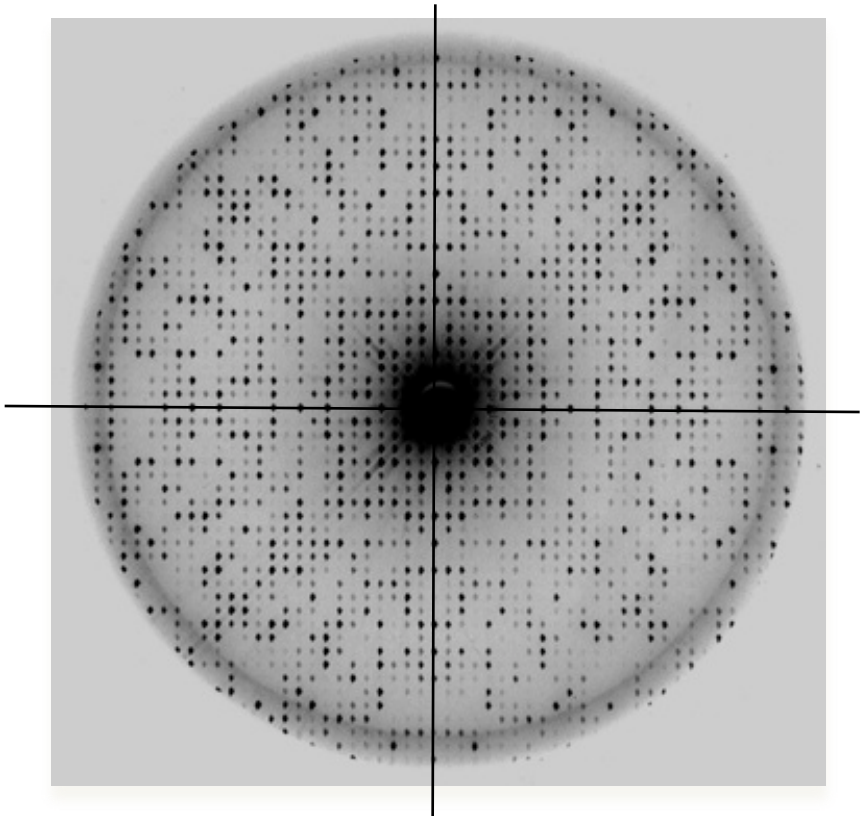


- before starting data collection make sure xtal is properly centered and does not “walk” out of beam
- consider exposure time, attenuation, wedge angle, detector distance, wavelength ... before it is too late!


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

Completeness - crystal symmetry

- reciprocal space symmetry = crystal symmetry plus
- Friedel's law : I (h k l) = I (-h -k -l)

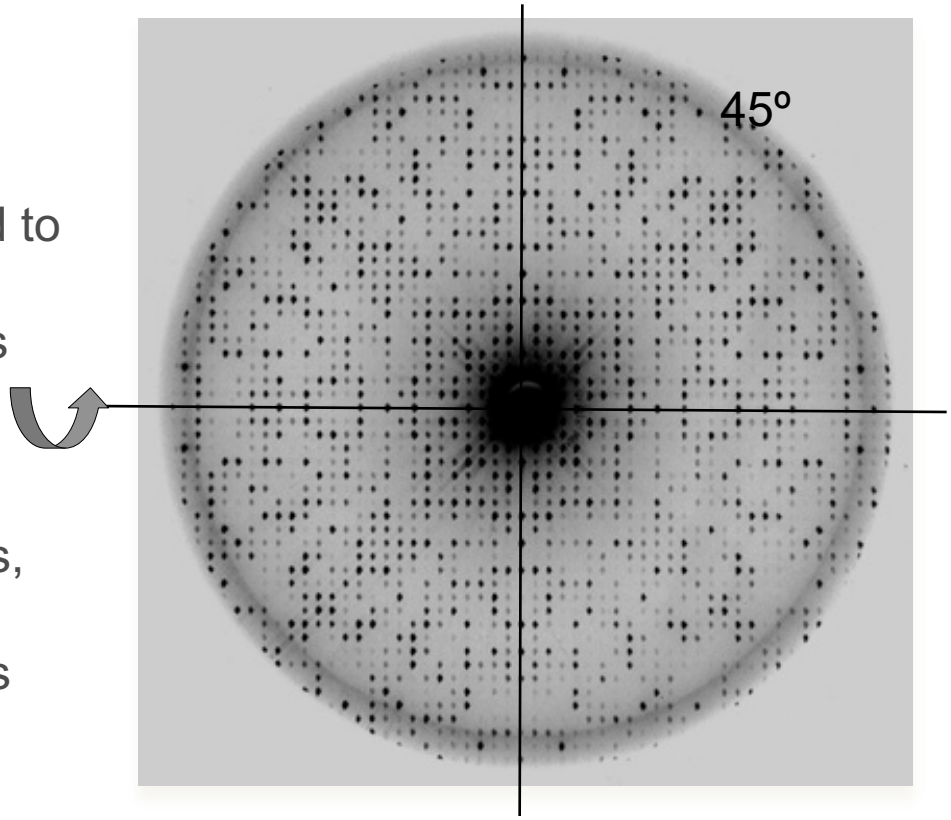


Completeness - crystal symmetry

- reciprocal space symmetry = crystal symmetry plus
- Friedel's law : I (h k l) = I (-h -k -l)
- thus even for P1 only 180º are needed to obtain complete data set of unique reflections - unless there is anomalous scattering

90º

- for N-fold symmetry along rotation axis, need 180/N degrees of data to obtain complete data set (assuming Friedel's law holds)
- you have to start collecting at the right angle, unless collecting > 180º

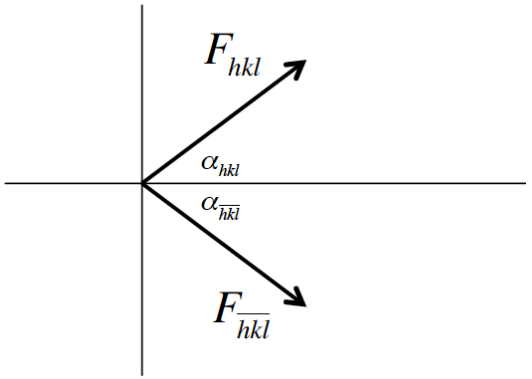


shot along 4-fold
(perpendicular to rotation axis)

Symmetry in the diffraction pattern

$$F_{hkl} = V \int_{cell} \rho(xyz) \exp[2\pi i(hx + ky + lz)]\, dx\, dy\, dz$$

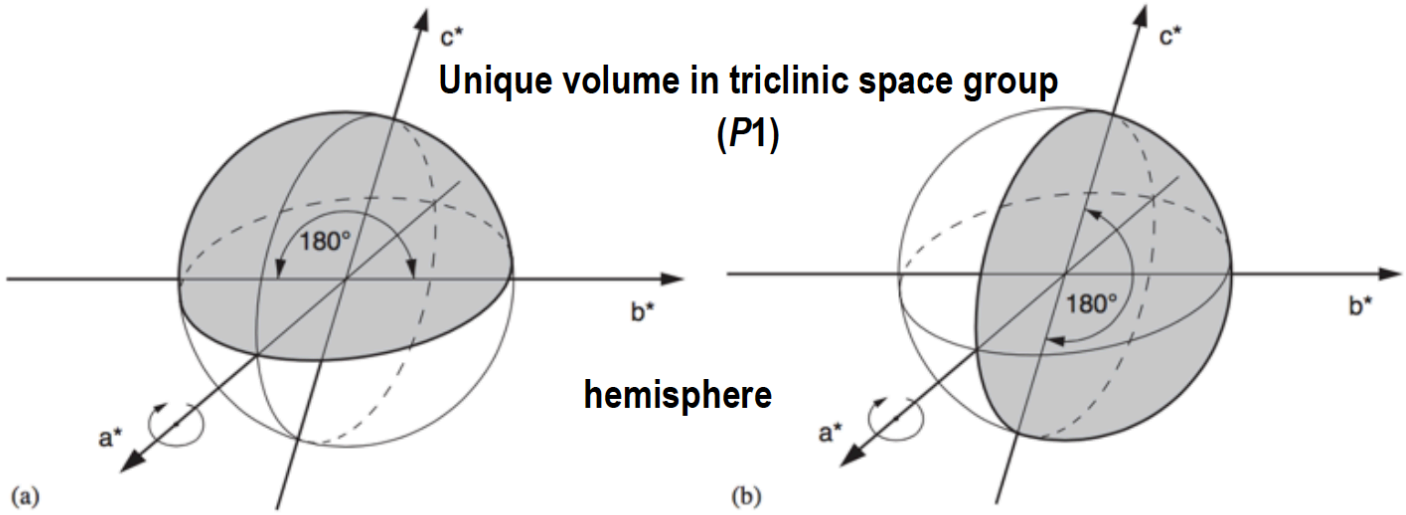
$$F_{\overline{hkl}} = V \int_{cell} \rho(xyz) \exp[2\pi i(-hx - ky - lz)]\, dx\, dy\, dz$$



$$I_{hkl} \propto (F_{hkl})^2$$

$I_{hkl} = I_{\overline{hkl}}$ ***Friedel*'s law;** **However, $I_{hkl} \neq I_{\overline{hkl}}$ when anomalous scattering is present**

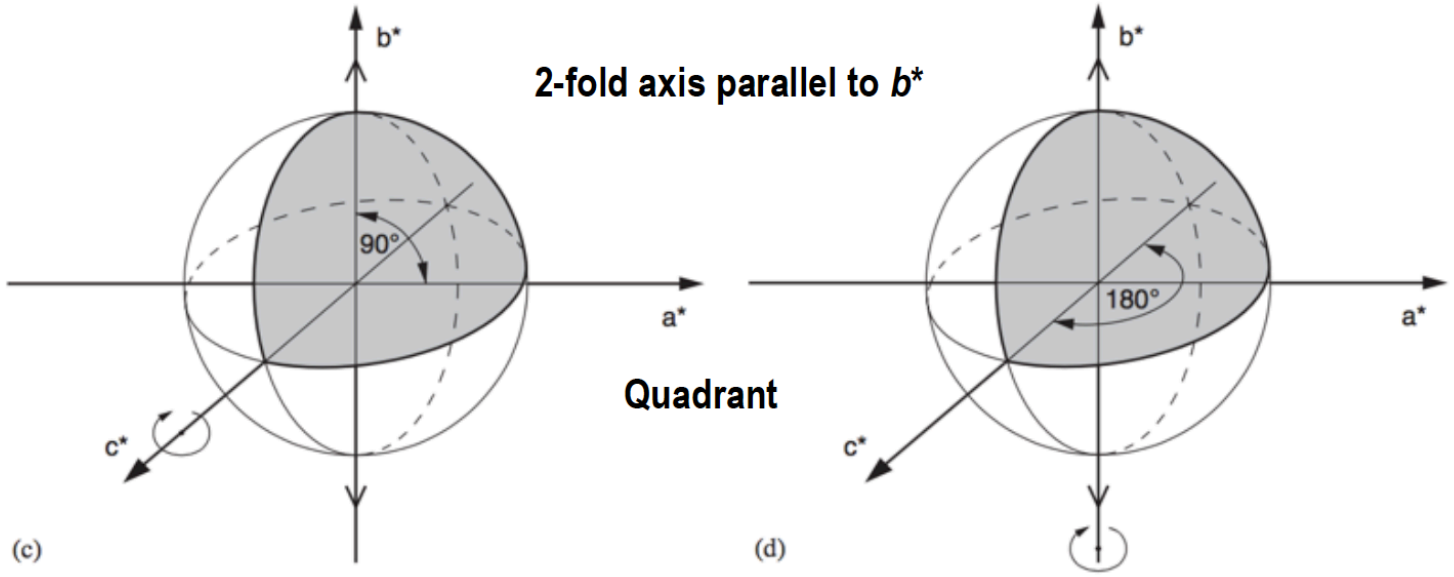
Unique volume in triclinic space group (*P*1)



hemisphere

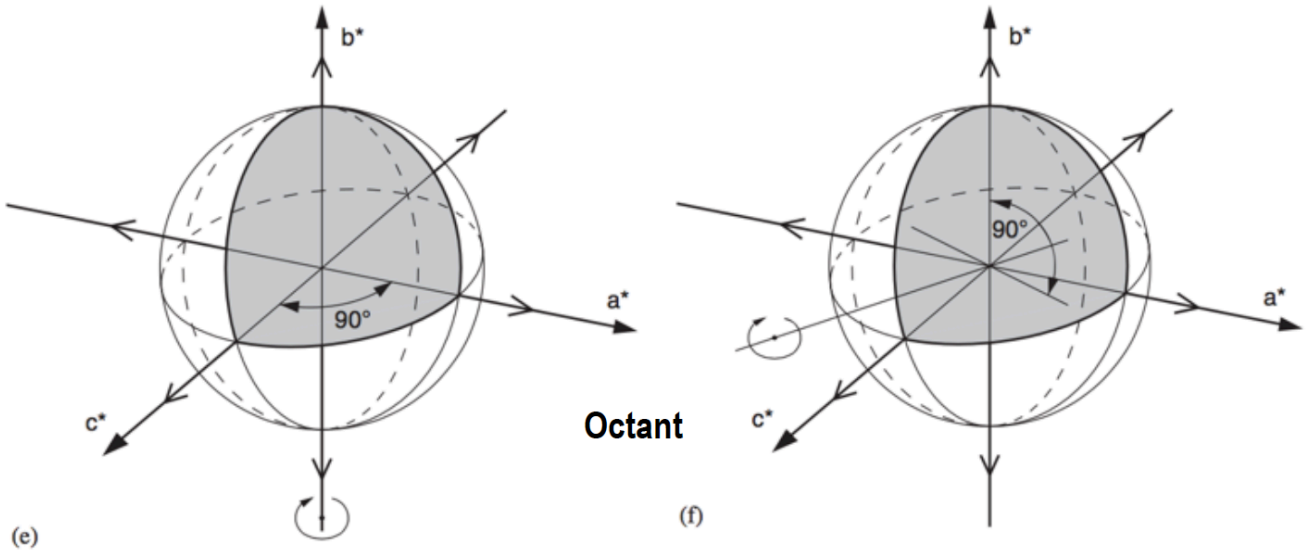
Unique volume in monoclinic space group (*P*2, *P*2$_1$, *C*2)

$$I_{hkl} = I_{\overline{hkl}} = I_{\bar{h}k\bar{l}} = I_{h\bar{k}l}$$



Unique volume in orthorhombic space group ($P2_12_12_1$, $C222_1$, $F222$, $I222$ …)

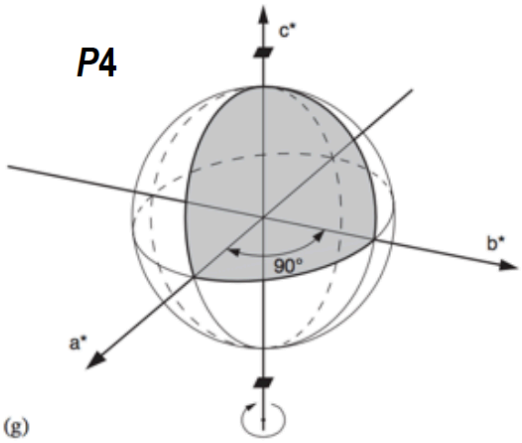
$$I_{hkl} = I_{\bar{h}kl} = I_{h\bar{k}l} = I_{hk\bar{l}} = I_{\bar{h}\bar{k}l} = I_{h\bar{k}\bar{l}} = I_{\bar{h}k\bar{l}} = I_{\bar{h}\bar{k}\bar{l}}$$



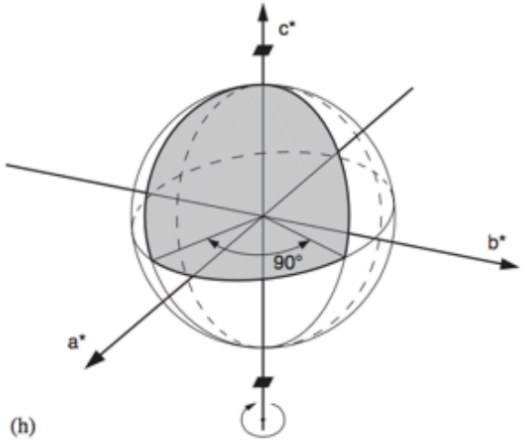
The 90° rotation must be between two axial orientations

Unique volume in tetragonal space groups

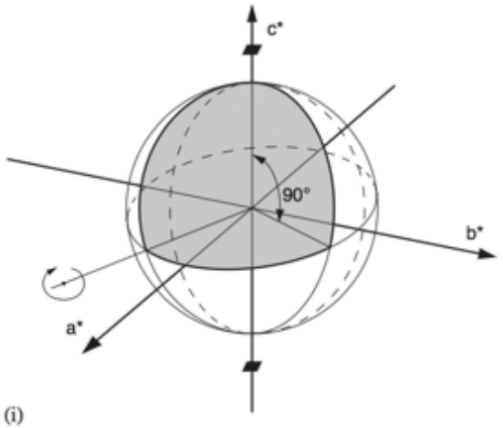
*P*4



(g)

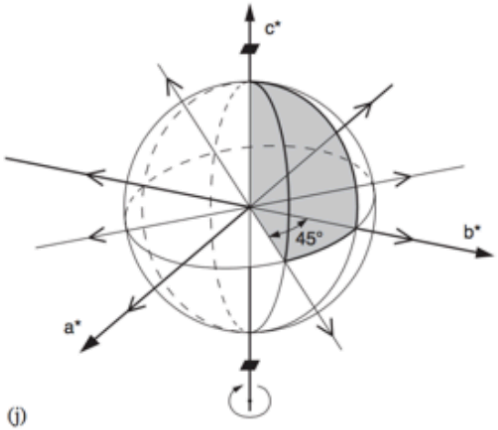


(h)

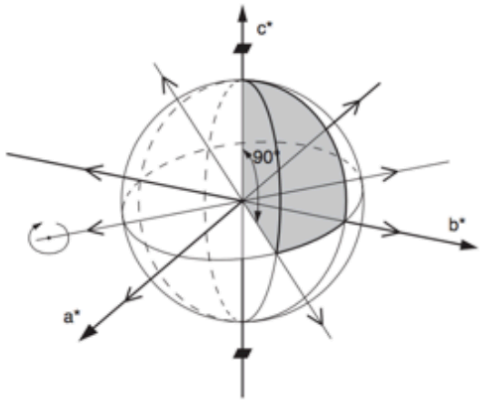


(i)

*P*422



(j)



(k)

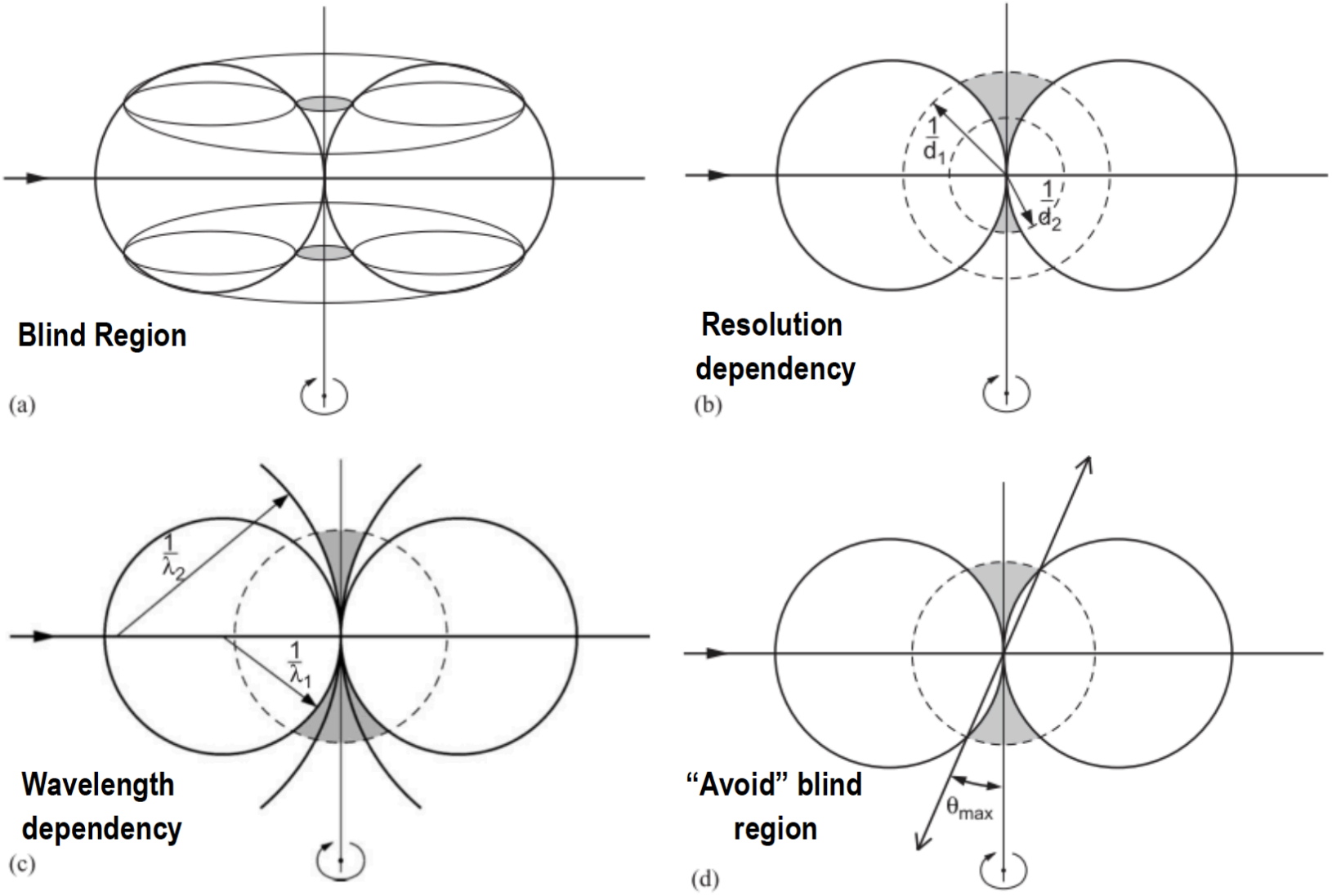
Table 1

Rotation range (°) required to collect a complete data set in different crystal classes.

The direction of the spindle axis is given in parentheses; *ac* means any vector in the *ac* plane.

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

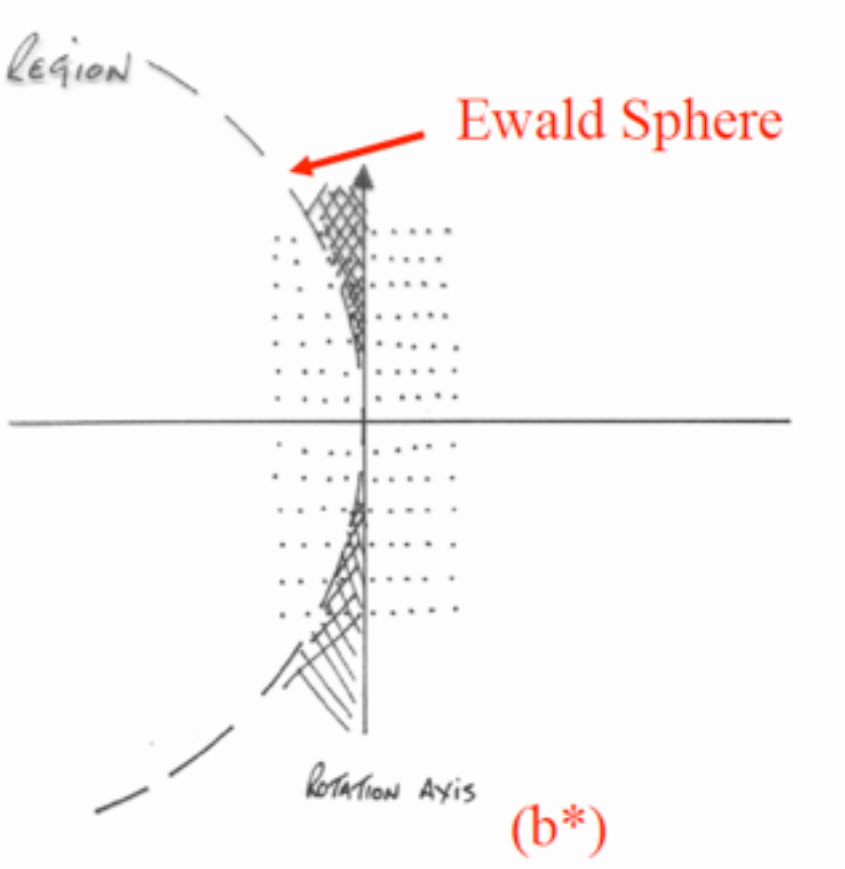
Dauter, *Acta Cryst.* *D***55,** 1703 (1999)



Dauter, *Prog. Biophys Mol. Bio.* **89,** 153 (2005)

Completeness - blind region

- blind region is due to curvature of Ewald sphere - more data is lost at higher resolution
- *this is why the best orientation is with the highest symmetry axis almost aligned with the rotation axis of data collection - tumbling.



Data collection - Summary

SYMMETRY

Better to assume a lower symmetry when planning data collection if there is some uncertainty.

For orthorhombic, assume P222

If there is a possibility of P2 rather than P222, assume P2

Similarly, assume P3 rather than P6

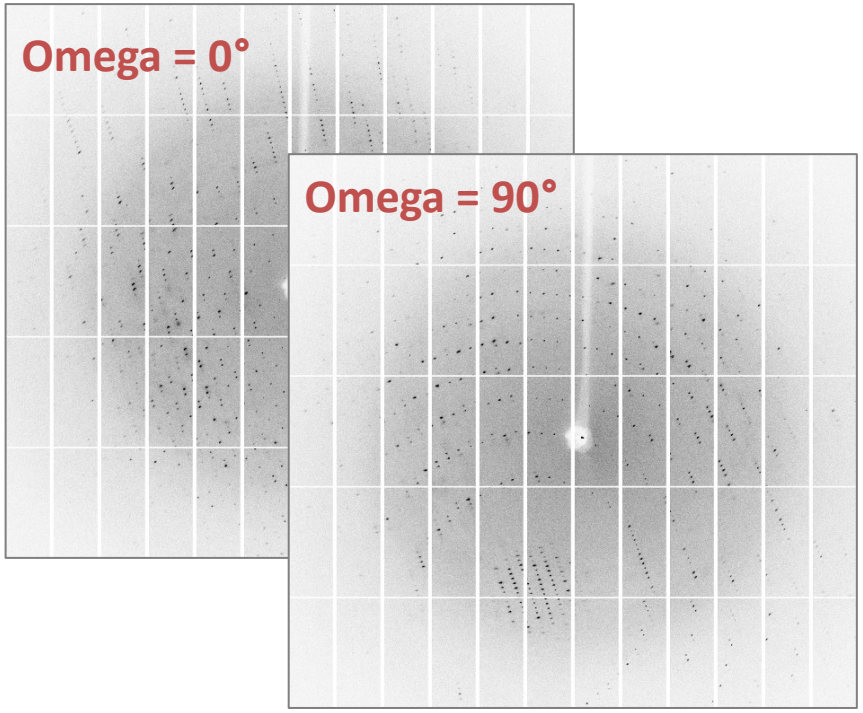
Indexing tells you about crystal lattice not about intensities - only after integrating and merging and checking Rmerge you will know the symmetry

Use ccp4 program pointless to get a first idea

Screening and analysis

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

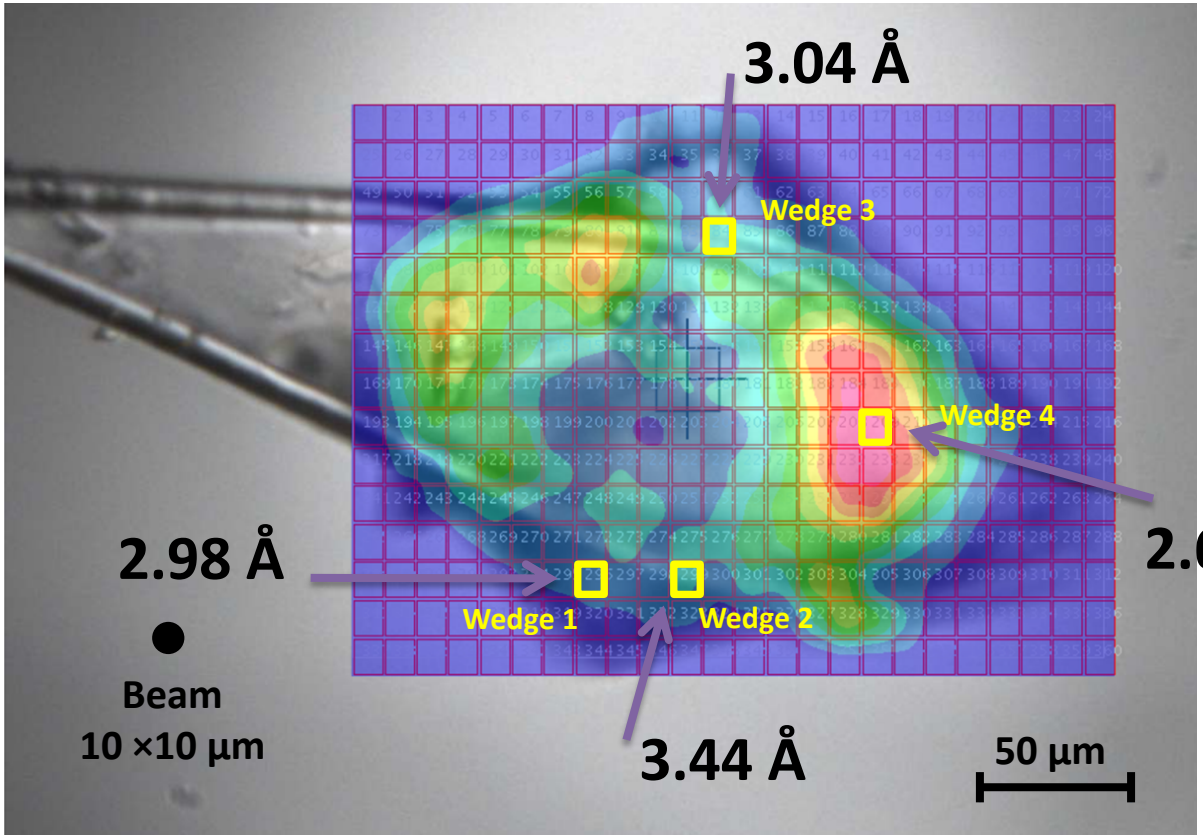
Usually by collecting 3 or 4 images, each separated by 45 degrees but can also efficiently be done by collecting an attenuated small wedge of 20 degrees for example.



Important to check the images

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

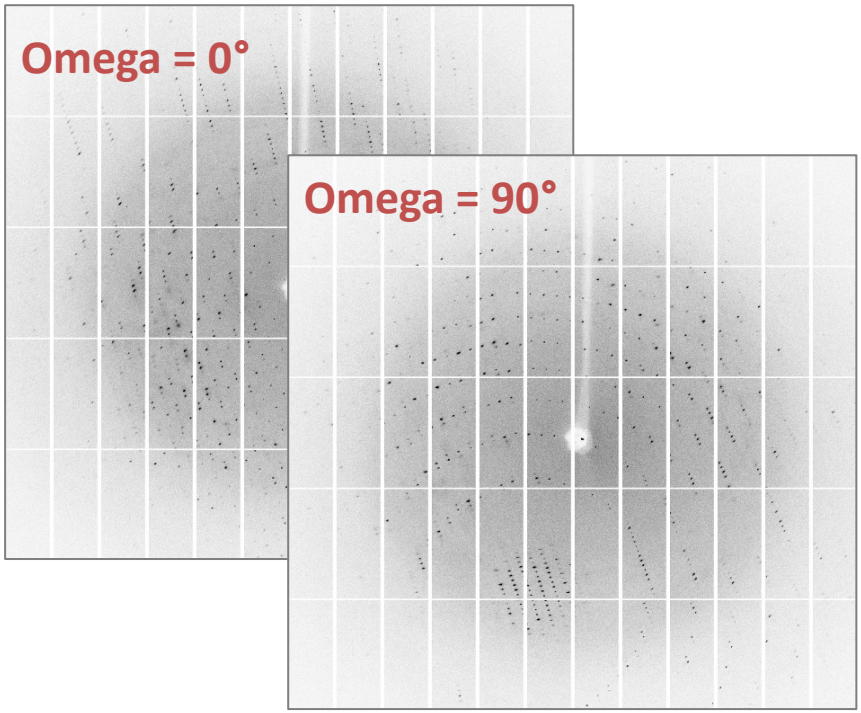
Even the same crystal can show a lot of variation..



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

Strategy

Programs such as BEST & Raddose 3D go further than just providing a start angle and oscillation range..



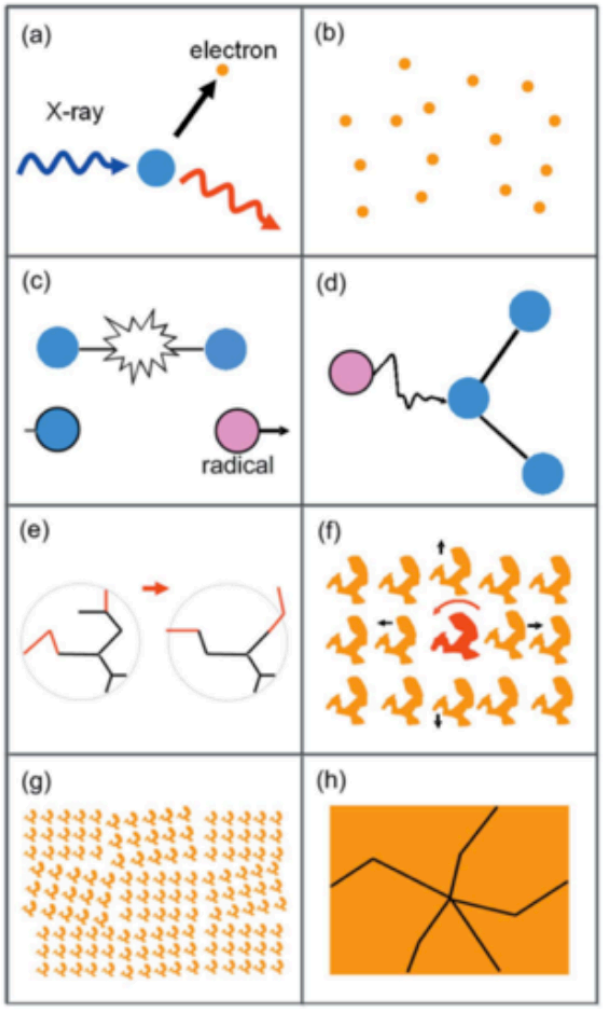
plus

- 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 and possible a step transmission strategy

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



M. Warkentin *et al. J. Synchrotron Rad.* **20,** 7 (2013)

Room temperature 298 K

(0.1 – 0.5 MGy)

- Owen, *et al. Acta Cryst.* **D68**, 810 (2012)
- Warkentin, *et al. J. Synchrotron Rad*, **20**, 7 (2013)

Cryo-temperature 100 K

Native data collection (20 MGy)

- Henderson, *Proc. R. Soc. B.* **241**, 6 (1990)
- Owen, *et al. Proc. Natl. Acad. Sci. USA*, **103**, 4912 (2006)

Experimental phasing (< 5 MGy)

- Holton, J. M. *J. Synchrotron Rad.* **14**, 51 (2007)
- Olieric, *et al. Acta Cryst.* **D63**, 759 (2007)

Rule of thumb

- Resolution dependency of 10 MGy / Å, Howells *et al. J. El. Spect. & Rel. Phen.* **170**, 4 (2009)
- Does estimation, Holton, *J. Synchrotron Rad.* **16**, 133 (2009)

$$Dose = (t_{expo} \times flux) / (k_{dose} \times I_{H\text{-}beam} \times I_{B\text{-}beam})$$

$$k_{dose} = 2000\lambda^{-2}$$

Radiation damage estimation at 100 K

$Dose = (t_{expo} \times flux) / (k_{dose} \times I_{H\text{-}beam} \times I_{B\text{-}beam})$; $k_{dose} = 2000\lambda^{-2}$ (Gy, sec, photon, μm)

e.g. 12.4 keV (1.0Å), 4×10^{11} photon/sec,
dose-rate =

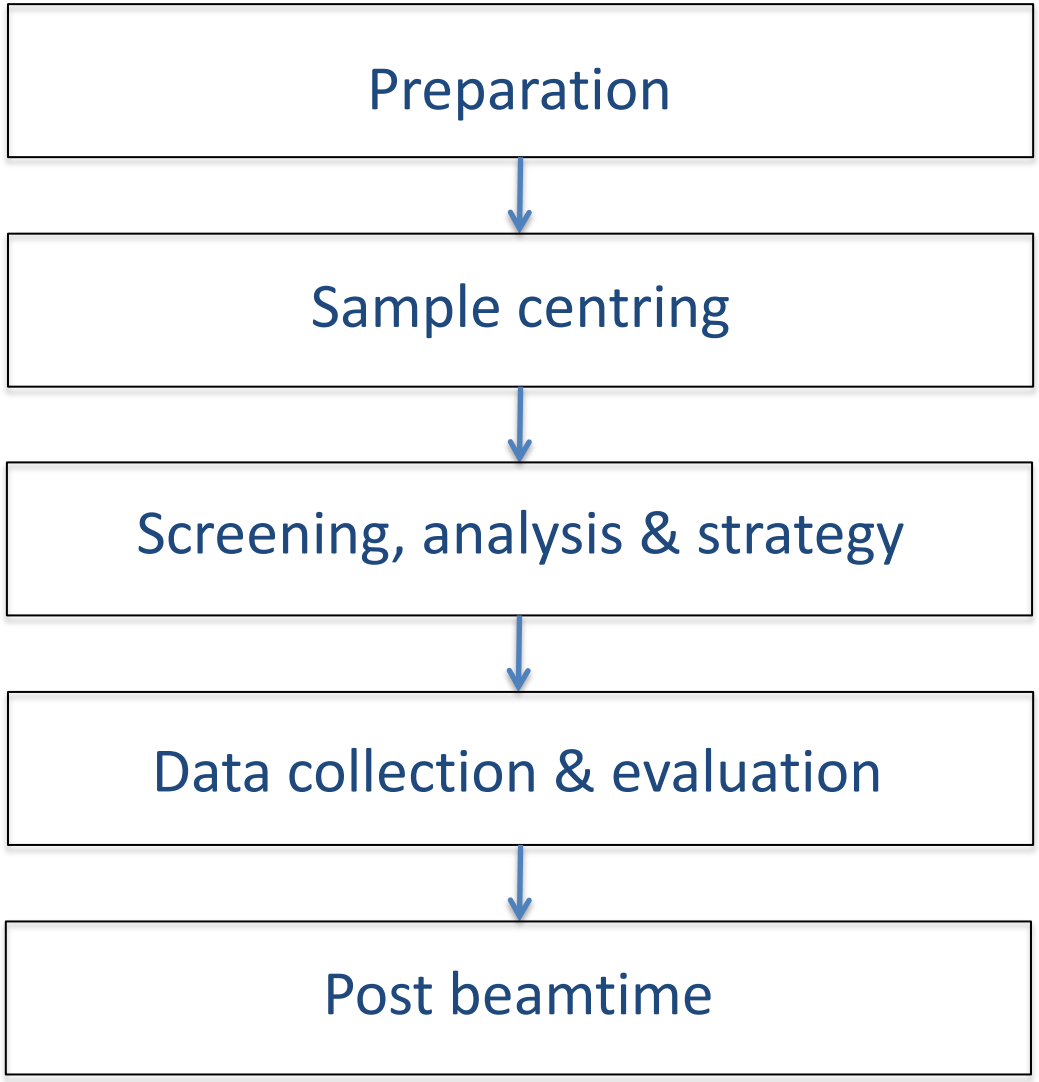
4×10^{11} p/s / (2000 × **100 μm × 100 μm**) = 0.02 MGy/s	→ one xtal	Single crystal crystallography
4×10^{11}p/s / (2000 × **10 μm × 10 μm**) = 2 MGy/s	→ a few xtals	Micro-crystallography
4×10^{11}p/s / (2000 × **5 μm × 5 μm**) = 8 MGy/s	→ tens xtals	Serial crystallography
4×10^{11}p/s / (2000 × **1 μm × 5 μm**) = 40 MGy/s	→ hundreds xtals	Serial crystallography
4×10^{11}p/s / (2000 × **1 μm × 1 μm**) = 200 MGy/s	→ thousands xtals	Serial femtosecond crystallography (xFEL)

Does estimation, Holton, *J. Synchrotron Rad.* **16**, 133 (2009)

expected crystal lifetime calculator
https://bl831.als.lbl.gov/xtallife.html

Beamline settings	Flux (ph/s)	Beam size (um)	Exposure (s)	Dose (MGy)
A	4.5x10^11	100 x 100	1316	20
B	3.1x10^11	30x 20	80	20
C	3.0x10^12	80x15	19	20
D	3.0x10^12	8x6	1	20
E	3.0x10^12	20x20	6	20

Assumptions:
Beamline energy at 0.97 A
Crystal of the size of the beam
100% transmission


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

Decision time

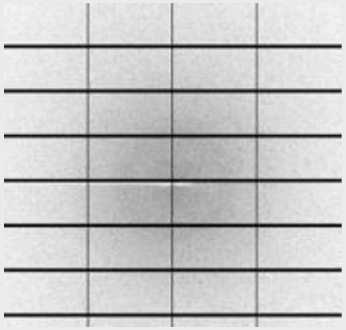
Remember the aim of your experiment – will this sample give you what you need, or get you closer to your goal? If not, move on.

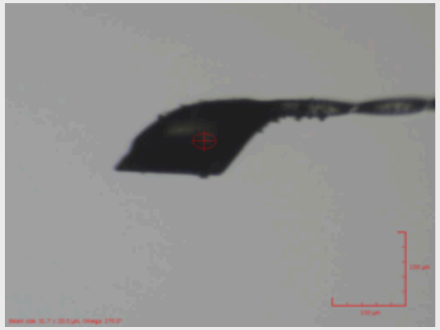
Experimental phasing is the most critical at this point, a trivial case can be made borderline by bad choices. Most common error is overexposing the crystal, next is not collecting enough data.

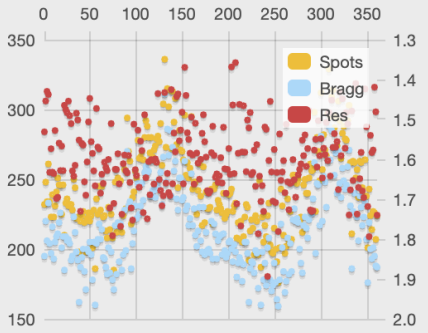
Remember that low energy = stronger diffraction and faster radiation damage. Transmission of <5% not uncommon. If the edge is very different to the screening energy then consider repeating the screening shots.

0 0 28-11-2019 21:05:04 - 20191128/Sethaumatin/Sethau_14/stepped/Sethau_14_1_master.h5

Sample: Sethau_14	Flux: 3.04e+11
Ω Start: 0.0°	Ω Osc: 0.10°
Ω Overlap: 0°	No. Images: 3600
Resolution: 1.33Å	Wavelength: 0.9763Å
Exposure: 0.002s	Transmission: 100.00%
Beamsize: 32x20µm	Type: SAD
Group: 4 Data Collections	

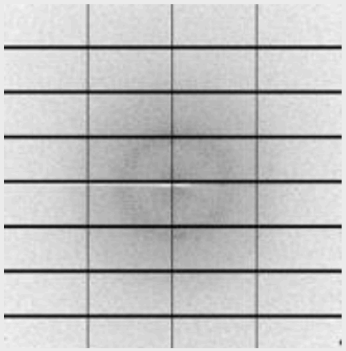




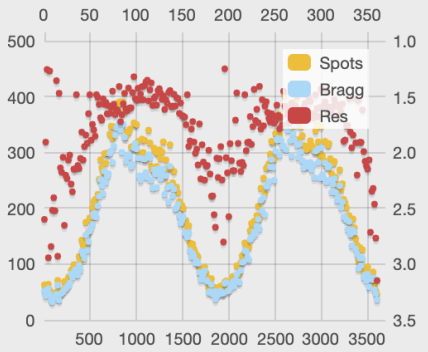


0 0 28-11-2019 20:58:37 - 20191128/Sethaumatin/Sethau_13/stepped/Sethau_13_1_master.h5

Sample: Sethau_13	Flux: 3.06e+11
Ω Start: 0.0°	Ω Osc: 0.10°
Ω Overlap: 0°	No. Images: 3600
Resolution: 1.33Å	Wavelength: 0.9763Å
Exposure: 0.002s	Transmission: 100.00%
Beamsize: 32x20µm	Type: SAD
Group: 4 Data Collections	
Comment: (-1111,-315,146) Aperture: Large	

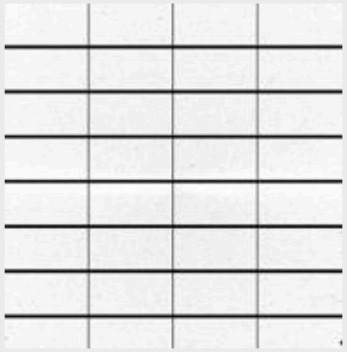




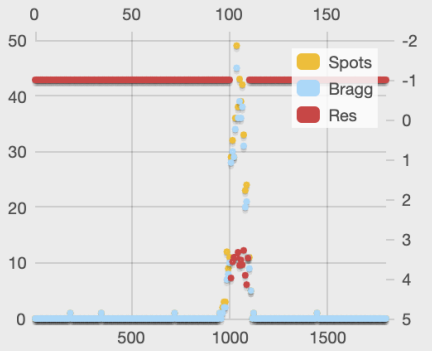


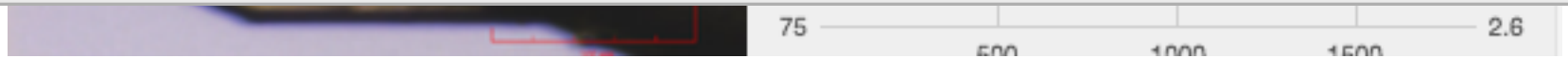
0 0 29-11-2019 09:14:05 - 20191129/Sethaumatin/Sethau_6/Sethau_6_700_1_master.h5

Sample: Sethau_6	Flux: 3.14e+11
Ω Start: 0.0°	Ω Osc: 0.10°
Ω Overlap: 0°	No. Images: 1800
Resolution: 4.32Å	Wavelength: 0.9795Å
Exposure: 0.010s	Transmission: 100.00%
Beamsize: 32x20µm	Type: SAD
Comment: (0,0,0) Aperture: Large	









Data analysis

Process as you go! Fast_dp results will be the first to be available

```
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
```

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

What's the goal?

Aim	High priorities	Lower priority	Max Dose per dataset (e.g. per 120, 180 or 360 degrees)
Native data collection	High resolution (e.g < 1.8 A) Complete data	High redundancy	20 MGy
Multi-wavelength anomalous dispersion (MAD) Single Anomalous dispersion (SAD)	Accurate, complete and highly redundant Very little radiation damage Choice of wavelength	High resolution data	1 MGy
Sulphur SAD	As above + extra high redundancy (40-200fold) if not on I23		0.3-0.5 MGy
Molecular replacement	Completeness at low resolution (< 3Å) Good quality low resolution	High resolution data High redundancy	20 - 30 MGy
Ligand/mutation	Medium resolution (e.g. 2 A), complete data. High throughput	High redundancy	15-20 MGy

Trying to collect everything at once will usually result in failure

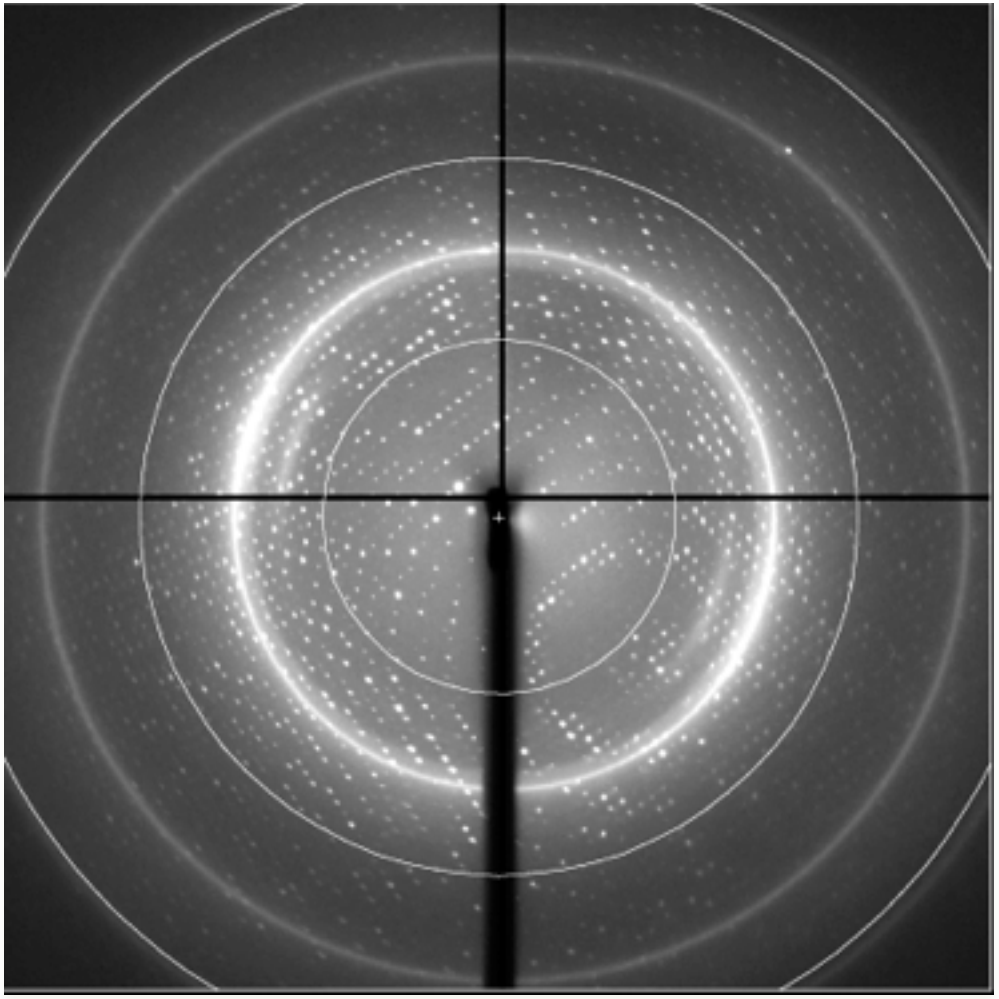
Data collection - the “native” data set

if you have solved your structure and need a “native” data set of the highest possible resolution:

- need complete (> 90%) but not highly redundant data set
- optimise resolution - “fry” as much as possible
- may have to collect multiple “wedges” - STRATEGY to calculate start and end angles to minimise number of images
- or collect datasets from multiple xtals and merge the most isomorphous ones until complete (extreme: XFEL - 1 crystal per image)

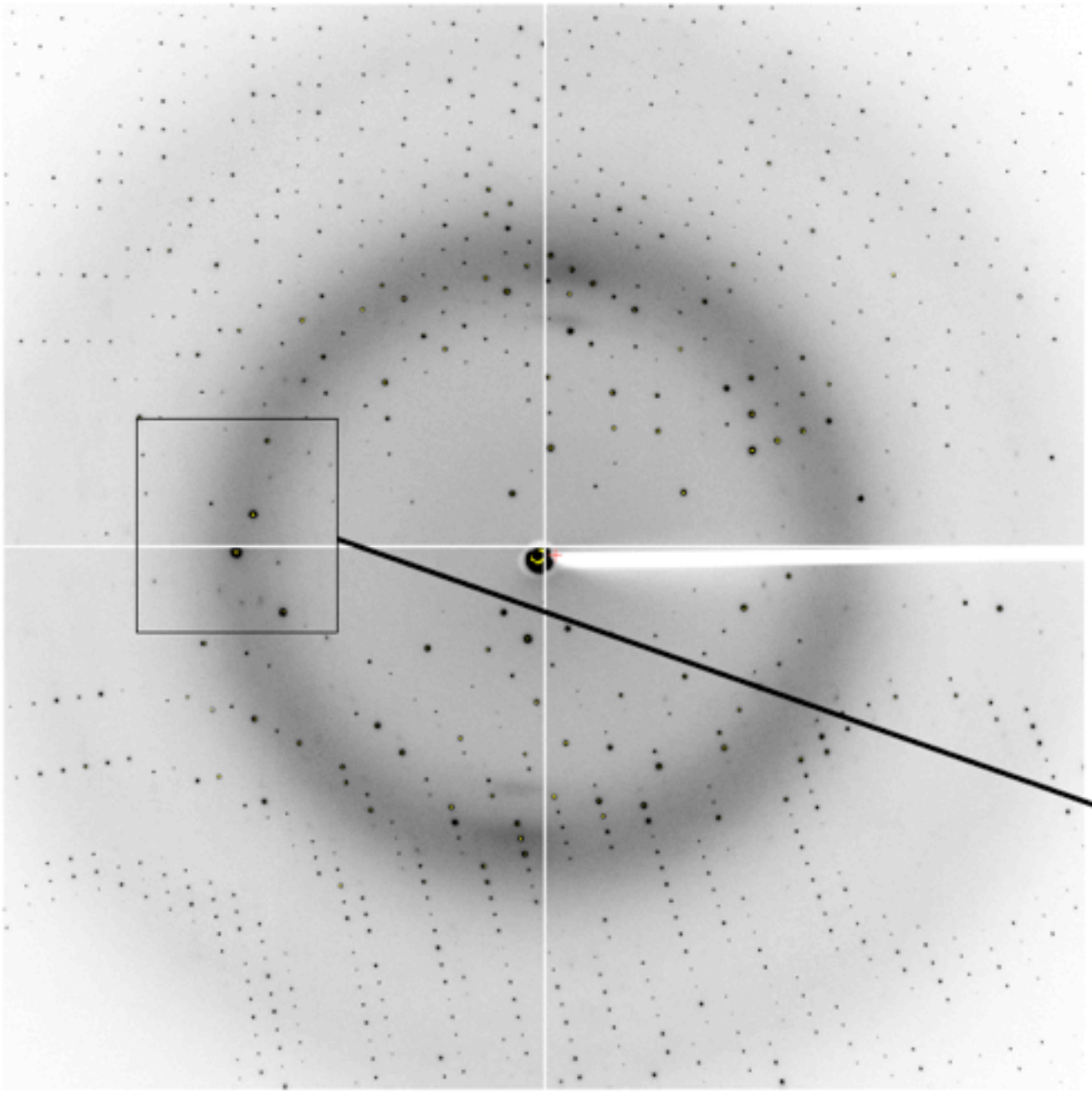
Data collection - molecular replacement

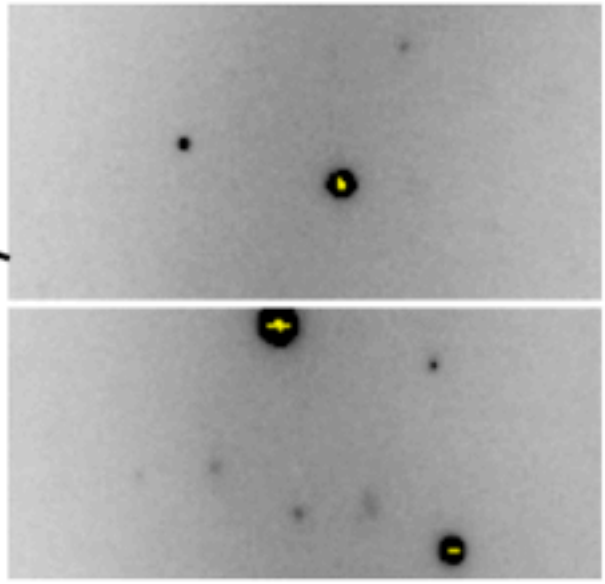
- needs good + complete low resolution data!
- take some time to adjust beam stop - move as far away from xtal as possible without exposing detector to primary beam
- be careful with overloads/count rate correction + spot overlaps
- overall redundancy of data not so important, but as always should be > 90% complete



Overloads

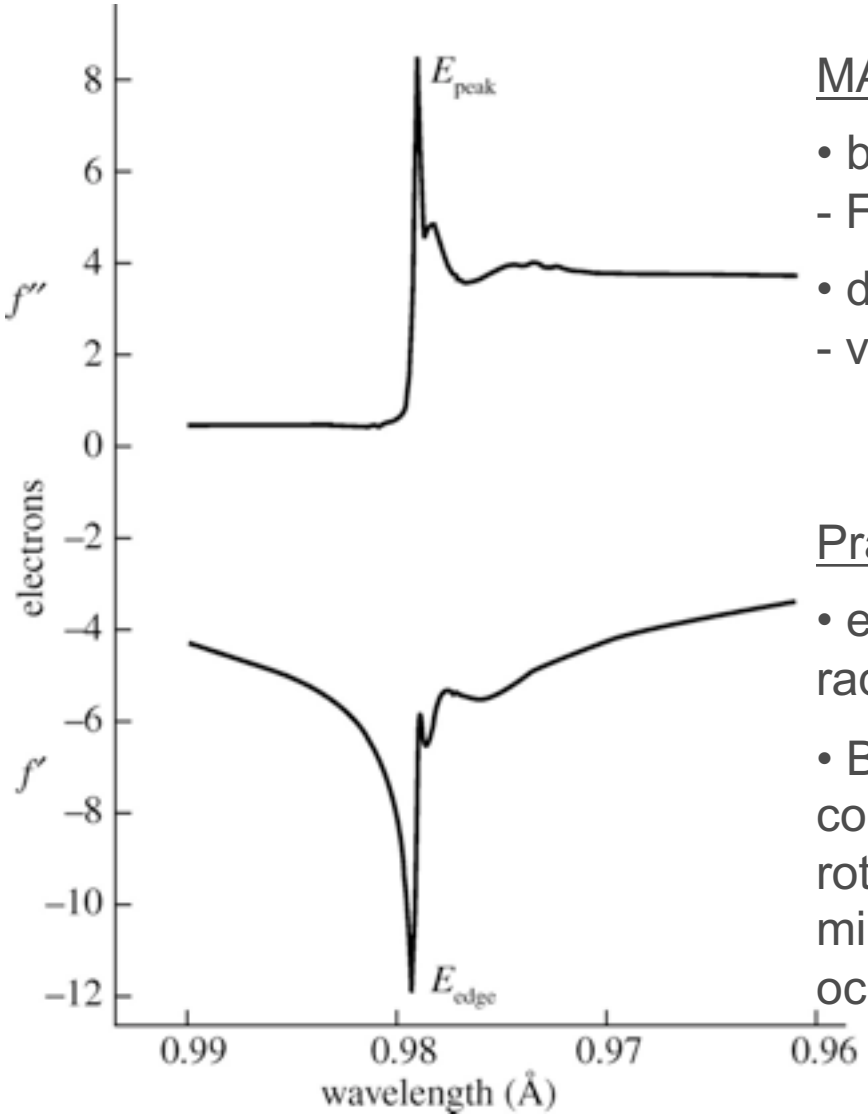
Many reflections in central portion of image are overloaded, need to ***adjust exposure time or attenuate beam.*** If we don't then data set will be incomplete at low resolution





The Pilatus 6M and Eiger 16M have a large dynamic range. When the dynamic range is exceeded, the counter starts from zero again. The only indication for this is that holes may be observed in a few peaks. The issue is also that some highly count rate corrected peaks won't show any visible pathology.

Data collection - anomalous diffraction



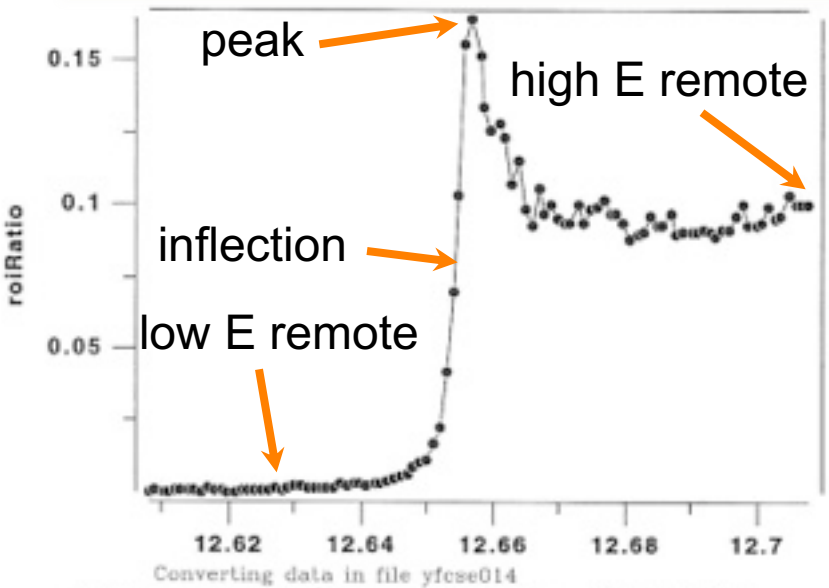
MAD/SAD:

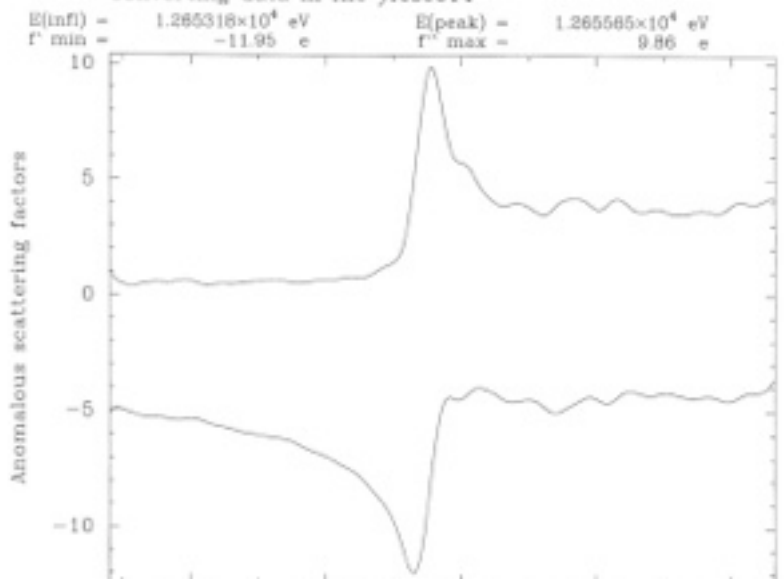
- breaking Friedel's symmetry: I(hkl)≠I(-h-k-l)
  - Friedel pairs become Bijvoet pairs
- detecting anomalous signal is tricky
  - very weak ~ 2 - 4 % of total signal!

Practical considerations:

- especially for SAD: the more redundancy (without radiation damage), the better!!!
- Bijvoet pairs are on images 180º apart
  consider collecting in "inverse beam" mode, i.e. rotate +/- 180º every image or in small wedges, so minimal radiation damage, changes in beam etc occurs between Bijvoet pairs

Data collection - anomalous diffraction





MAD: 3-4 data sets of one crystal at different wavelengths around absorption edge of heavy atom

SAD: 1 data set of 1 crystal at absorption peak of heavy atom or slightly above the peal (e.g Se)

- MX beamlines: automated scans of metal edges
- otherwise see e.g. http://skuld.bmsc.washington.edu/scatter/AS_index.html

Consider

- effects of changing wavelengths:

-- less/short absorption of X-rays by air at short/longer wavelength / high energy

-- radiation damage

-- also check max. resolution and spot overlaps after changing wavelength (Bragg's Law: nl=2dsinth)!

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

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?).

Data collection - Summary

SYMMETRY

Better to assume a lower symmetry when planning data collection if there is some uncertainty.

For orthorhombic, assume P222

If there is a possibility of P2 rather than P222, assume P2

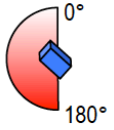
Similarly, assume P3 rather than P6

Indexing tells you about crystal lattice not about intensities - only after integrating and merging and checking Rmerge you will know the symmetry

Use ccp4 program pointless to get a first idea

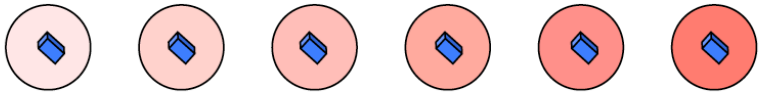
Traditional high-dose coarse-phi slicing strategy

- 0.5-1° per image,180°, 4-fold redundancy, and 10 MGy
- CCD detector



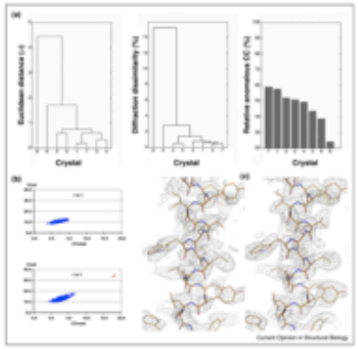
low dose, fine-phi slicing, high redundancy strategy

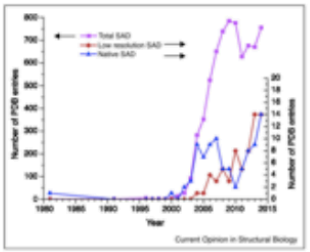
- ½ mosaicity slicing, *n* × 360° Mueller, *et al*, *Acta Cryst*. **D68**, 42 (2012)
- Distribute x-ray dose to multiple data sets. Liu *et al. Acta Cryst.* **A67**, 544, (2011)
- PILATUS detector



Multi-orientation and multi-crystal strategy

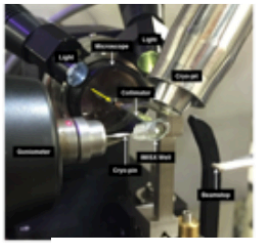
- Distribute x-ray dose to multiple data sets and multiple crystals
- Liu & Hendrickson *Curr. Opin. Struc. Bio.* **34**, 99 (2015)
- Weinert *et al. Nature Methods* **12**, 131, (2015)

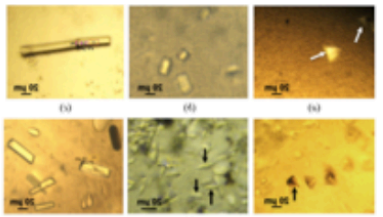




Micro-crystallography

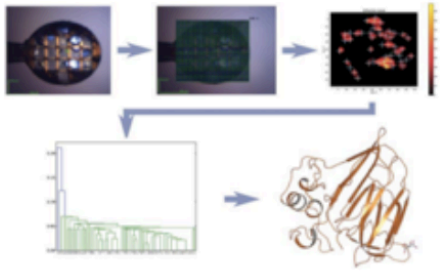
- 10-20° data per crystal × *n* crystals
- Smith *et al. Curr. Opin. Struc. Bio.* **22**, 602 (2012)

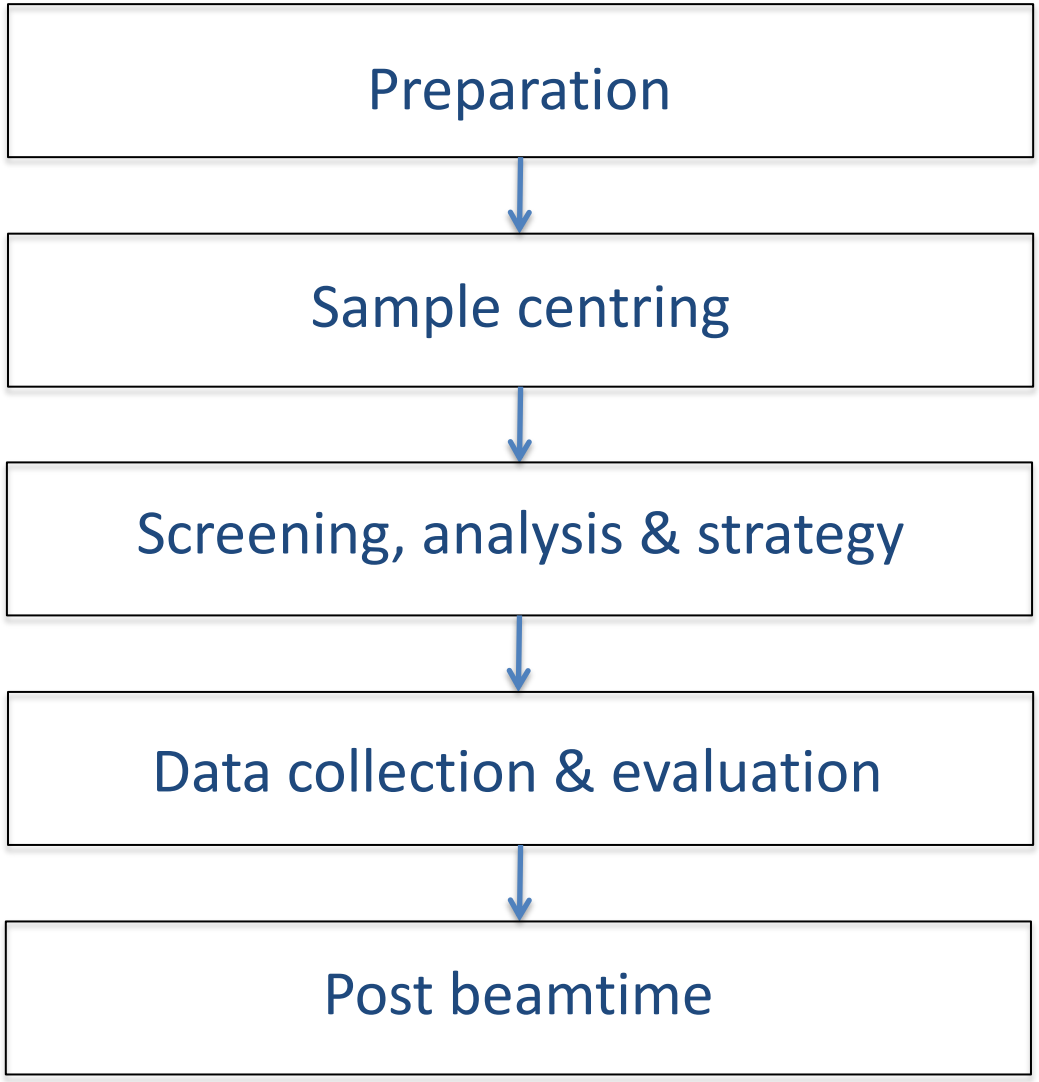




Serial crystallography

- Injector method. Botha *et al. Acta Cryst.* **D71**, 397 (2015), Nogly et al IUCrJ, **2**, 168 (2015)
- *In-situ* method. Huang *et al. Acta Cryst.* **D71**, 1238 (2015), **D72**, 93 (2016)
- Scanning method. Gati *et al. IUCrJ*, **1**, 87 (2014), Zander *et al.*, *Acta Cryst.* **D71**, 2328 (2015)




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