

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 slides



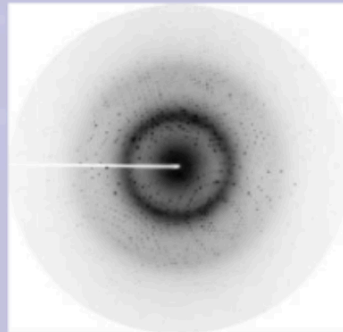
$$\rho(xyz) = \frac{1}{V} \sum_{hkl} F_{hkl} \exp(i\alpha_{hkl}) \exp[2\pi i(hx + ky + lz)]$$

Crystal



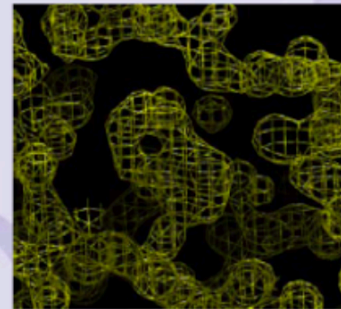
Diffraction Experiment

Diffraction pattern



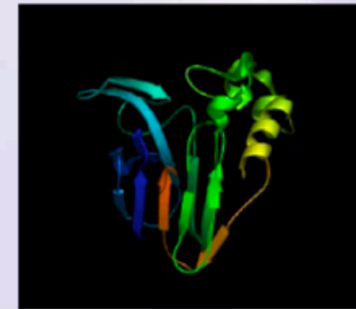
Phasing

Electron density



Model Building
Refinement

Structure



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

Data collection is important!

This is the last experimental step

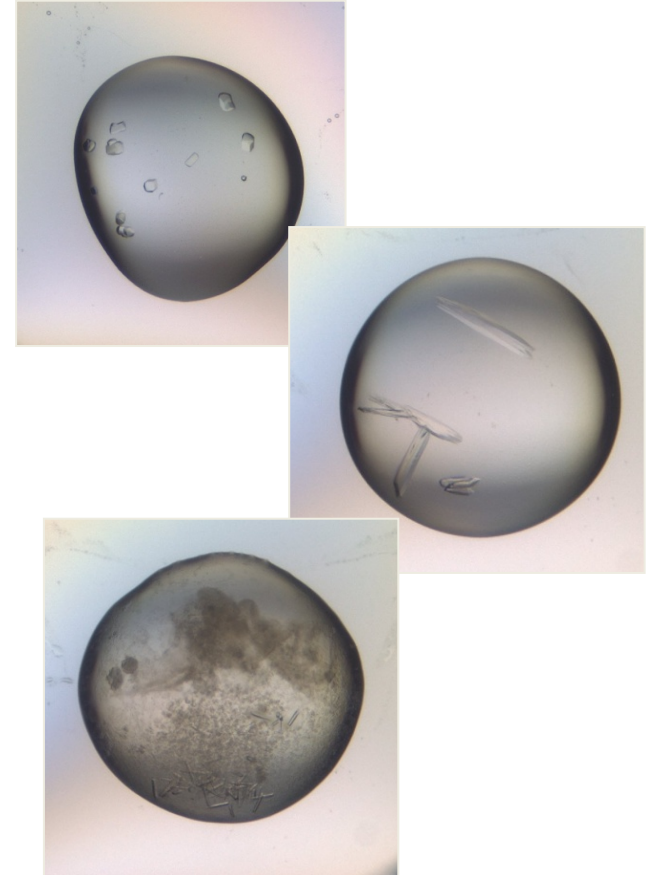
Time taken now to collect the best possible data can prevent pain and possible failure later on.

"One good friend is better than a thousand poor ones." Indian Proverb

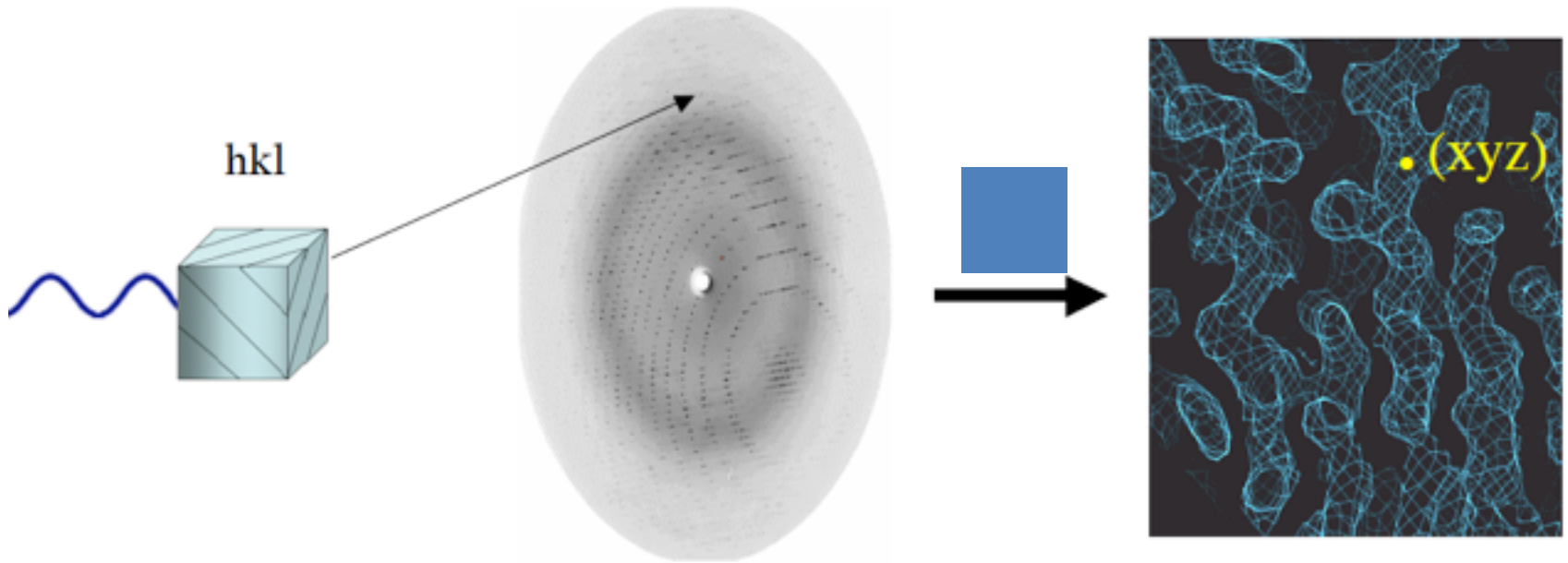
If the original data are poor then not much can be done about it later.



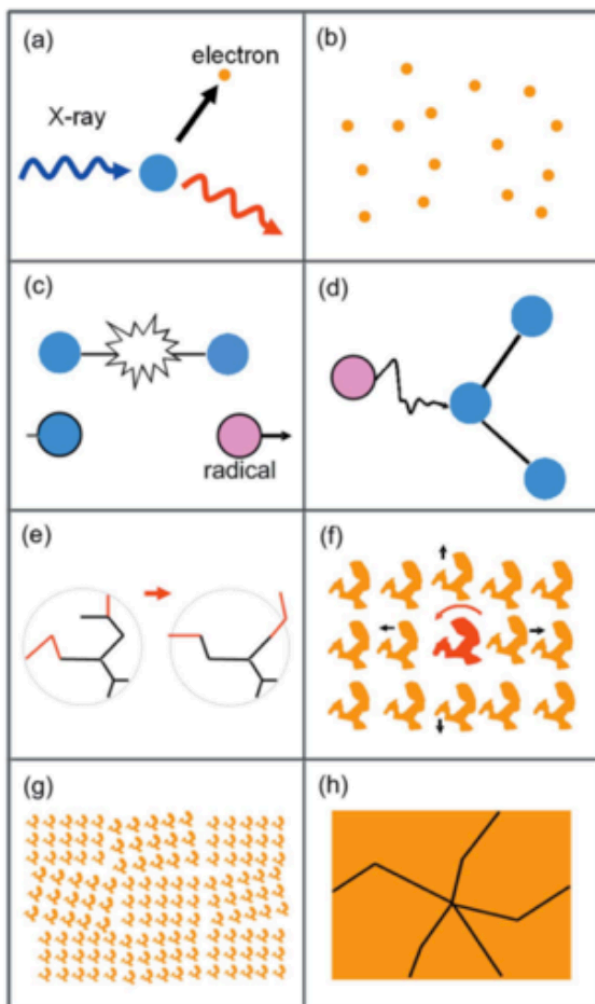
With fast sample changers, goniometers and detectors a lot of data can be collected very fast.
Quantity $\neq$ Quality



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



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)

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

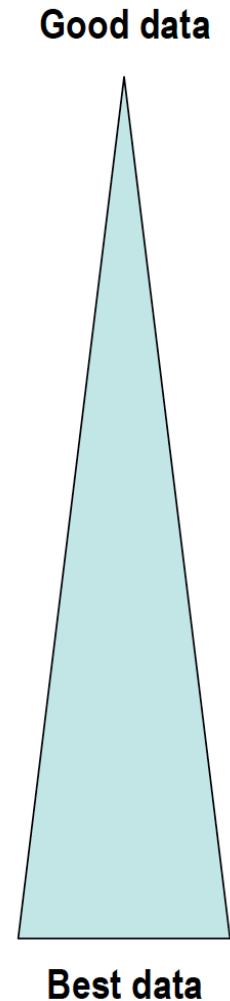
$$k_{dose} = 2000\lambda^{-2}$$

What's the goal?

Aim	High priorities	Lower priority	Max Dose per dataset (e.g. per 120, 180 or 360 degrees) For a ~2Å resolution dataset
Native data collection	High resolution (e.g < 1.8 Å) 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 - 5 MGy (heavy atom will increase dose for same exposure)
Sulphur SAD	As above + extra high redundancy (40-200fold) if not on I23 beamline		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 Å), 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

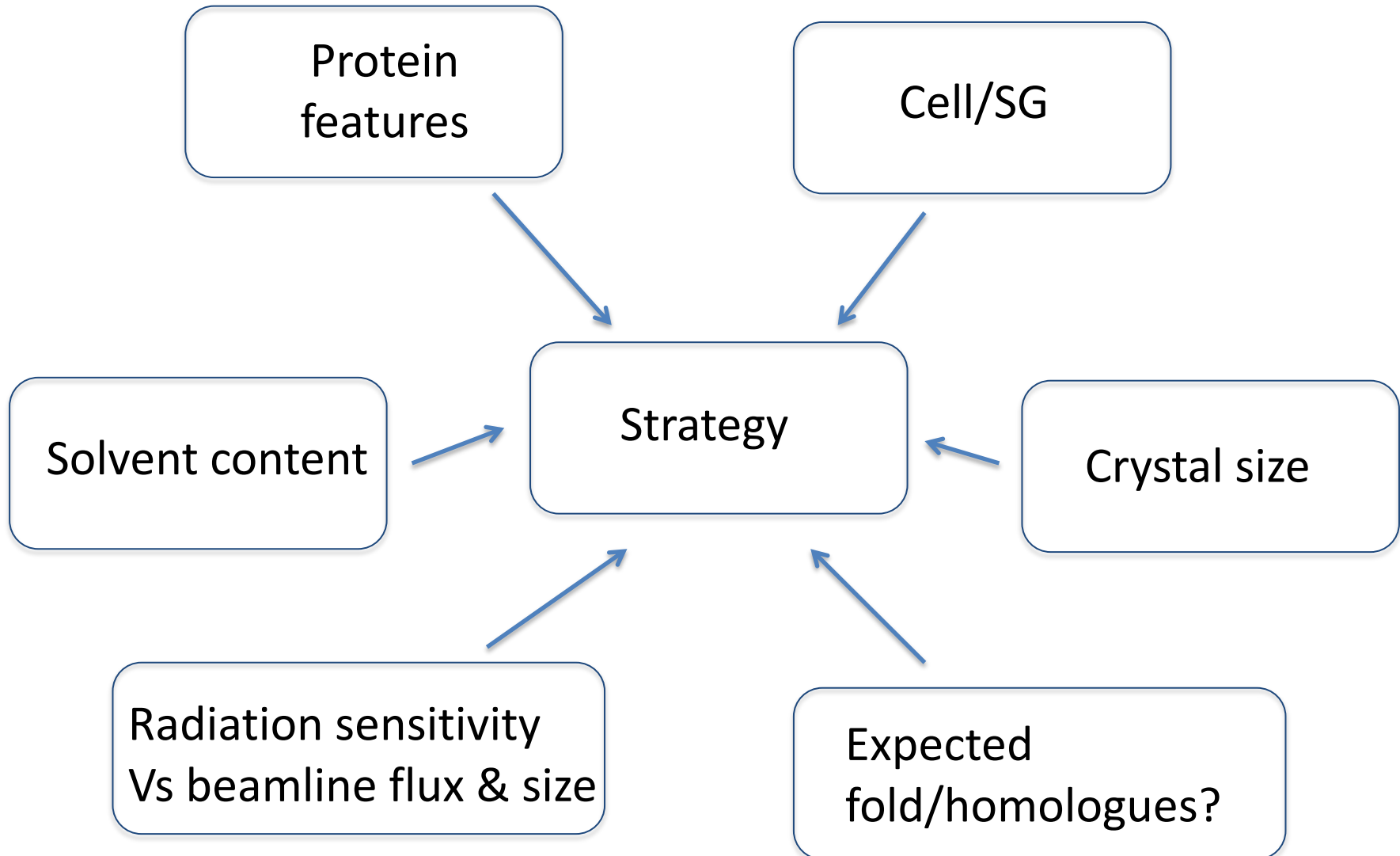


Data collection & evaluation



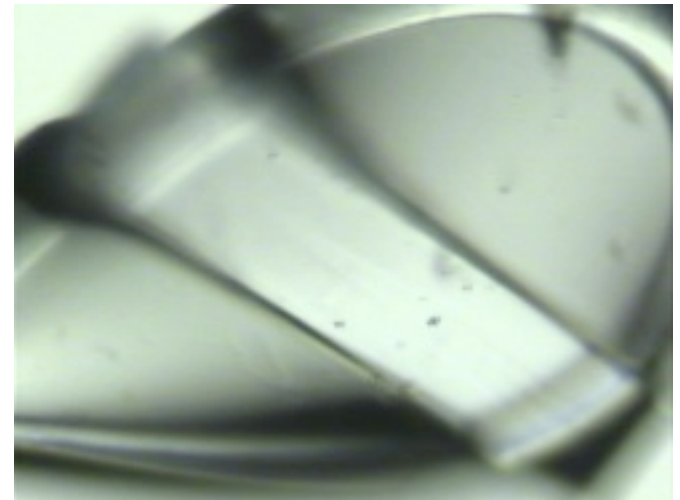
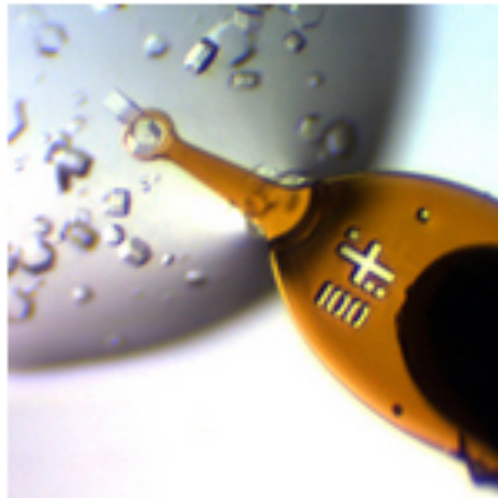
Post beamtime

Preparation - Prior knowledge

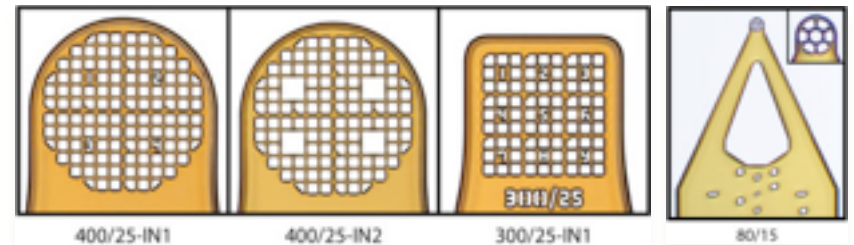
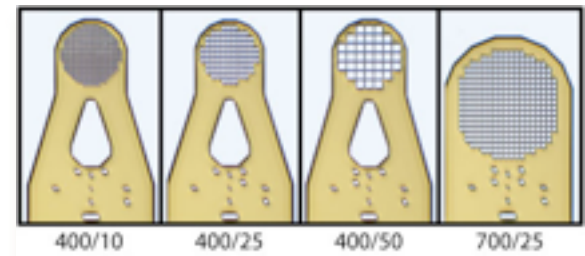
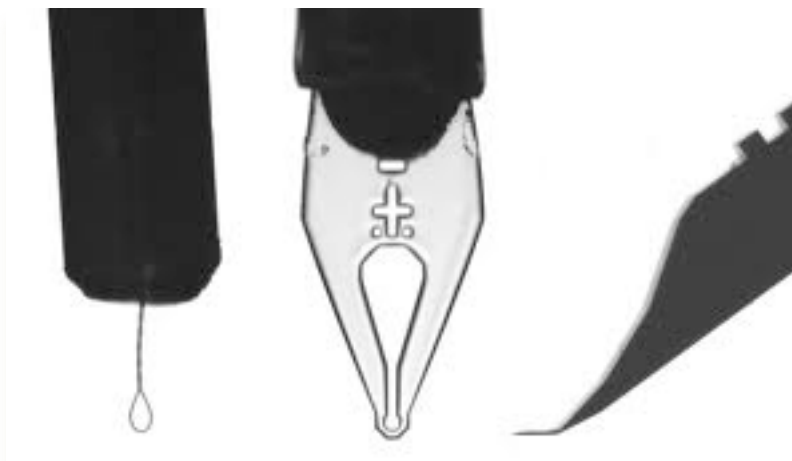
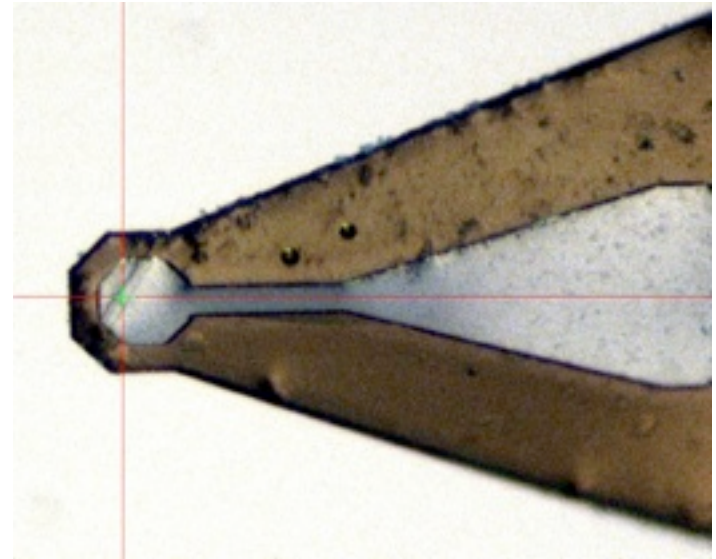
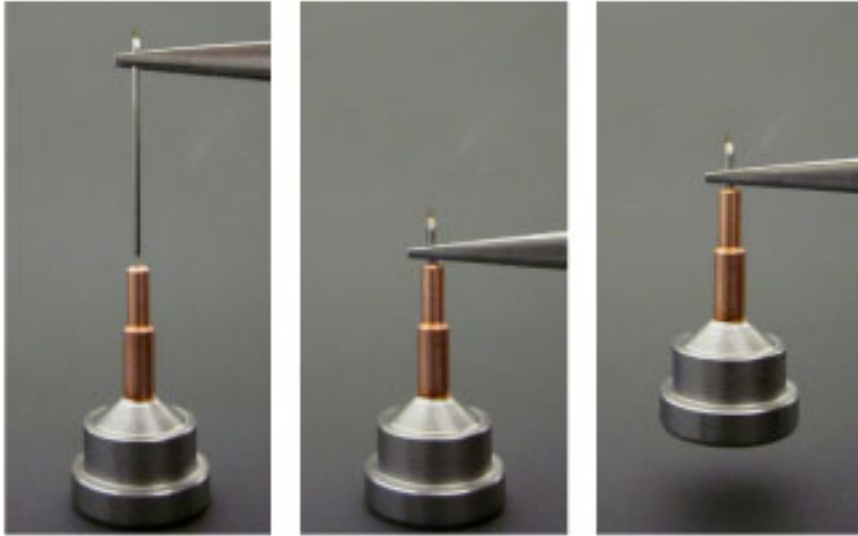


Data collection - crystal mounting

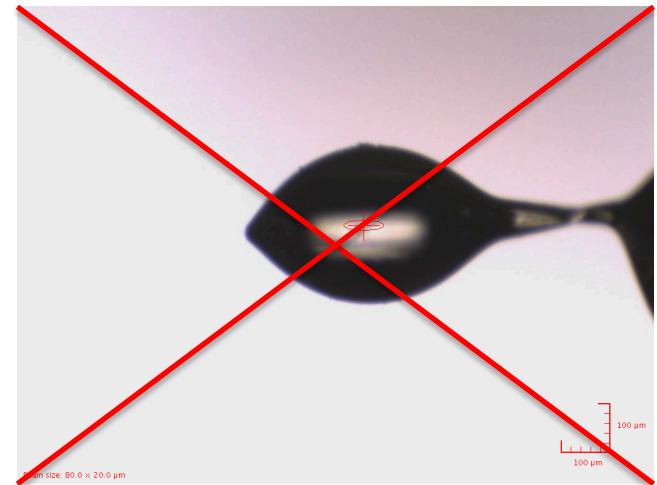
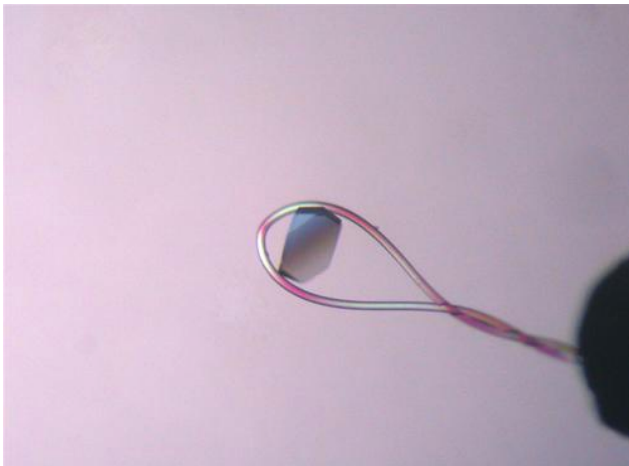
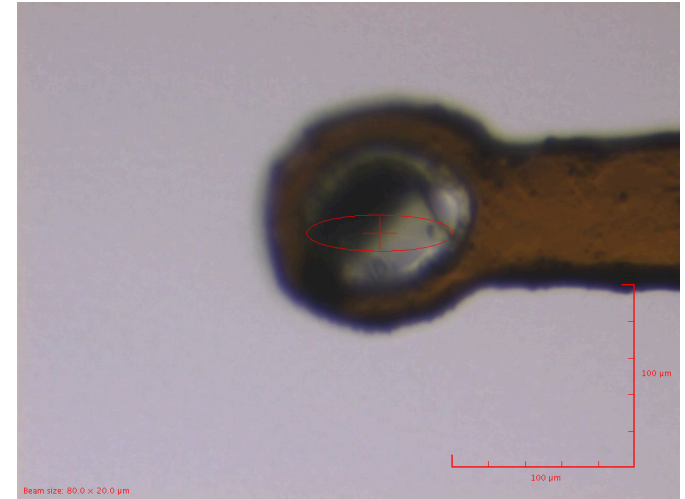
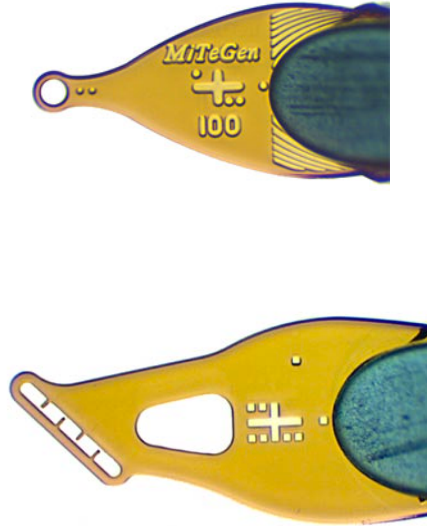
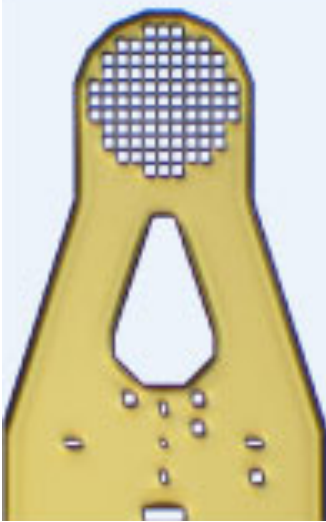
- if crystallisation condition does not contain cryo protectant, xtals need to be transferred to appropriate solution (e.g. add 20% glycerol, ethylene glycol, etc to crystallisation solution) before freezing; see A. Evdokimov's protocol for example
https://mcl1.ncifcrf.gov/nihxray/Tips-and-Tricks_CryotectionOfDelicateCrystals-ArtemEvdokimov.pdf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

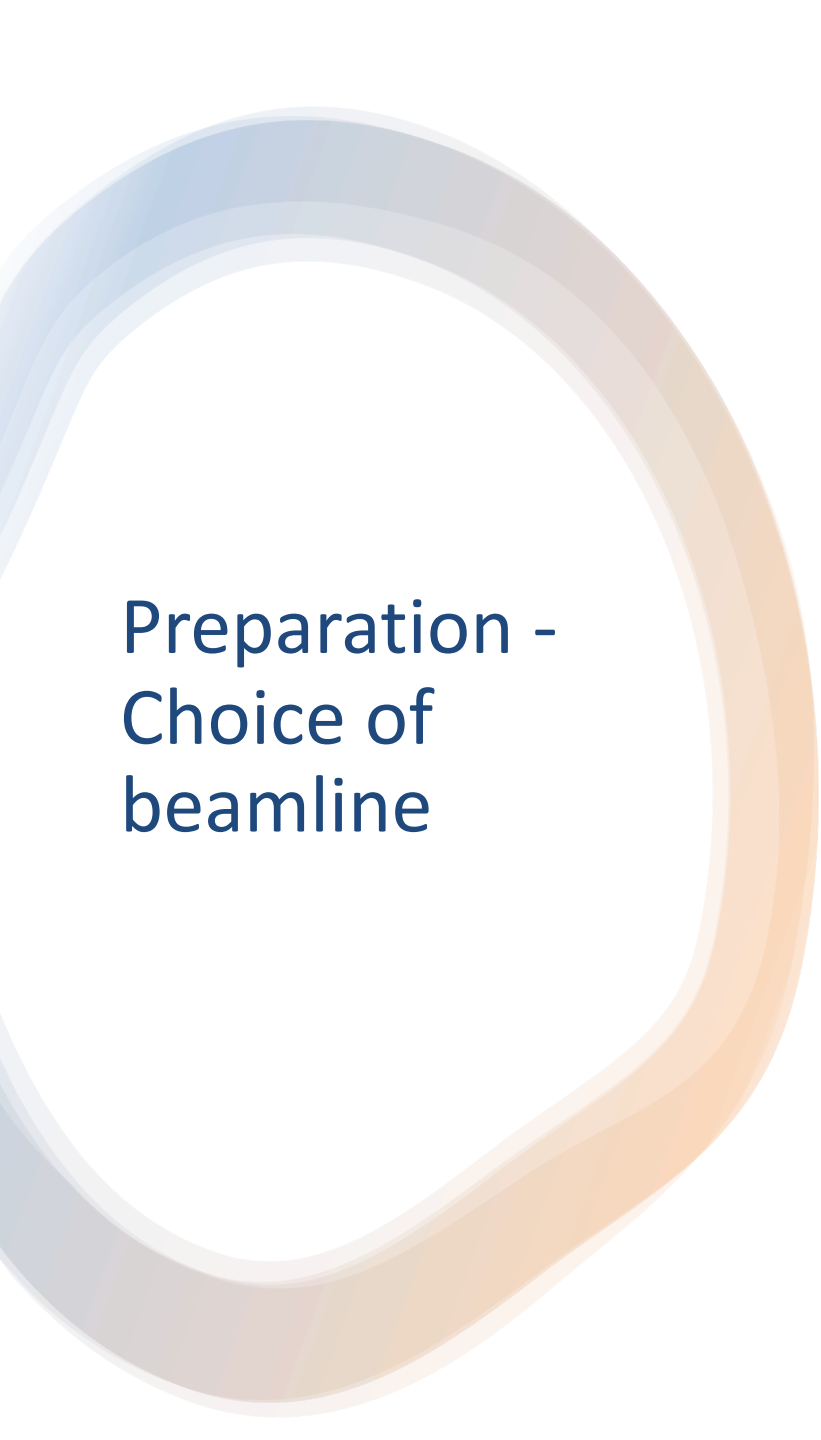


Data collection - crystal mounting



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

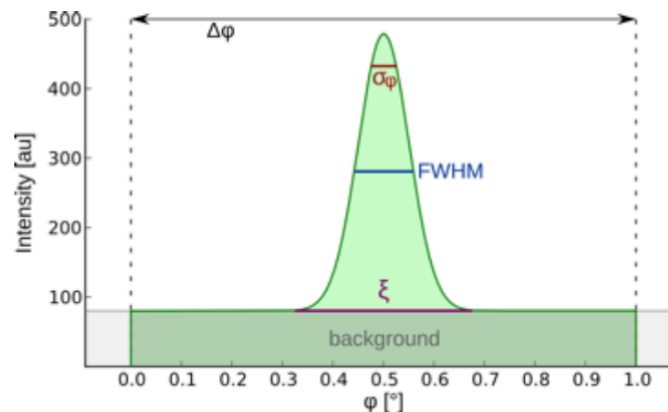
Random errors, counting statistics

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

$$I = N_p - N_b \quad \text{Signal is the difference}$$

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

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



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

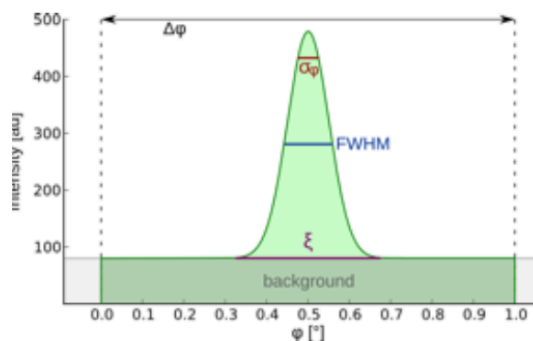
Random errors, counting statistics

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

$$I = N_p - N_b \quad \text{Signal is the difference}$$

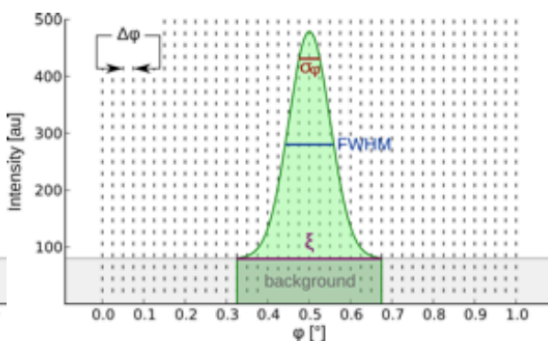
$$\sigma_I = (\sigma_p^2 + \sigma_b^2)^{1/2}$$

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



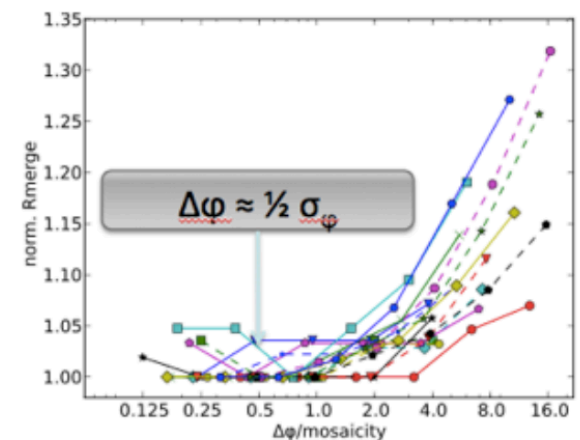
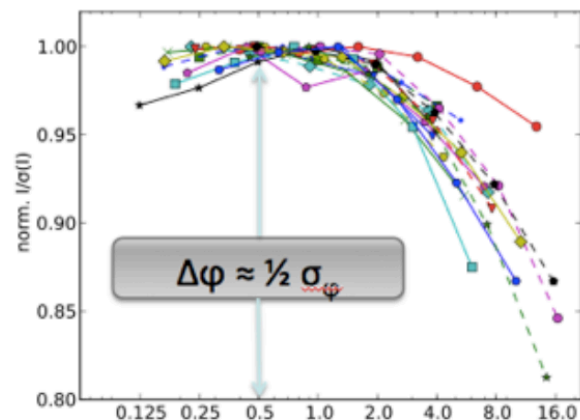
Wide ϕ -slicing

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



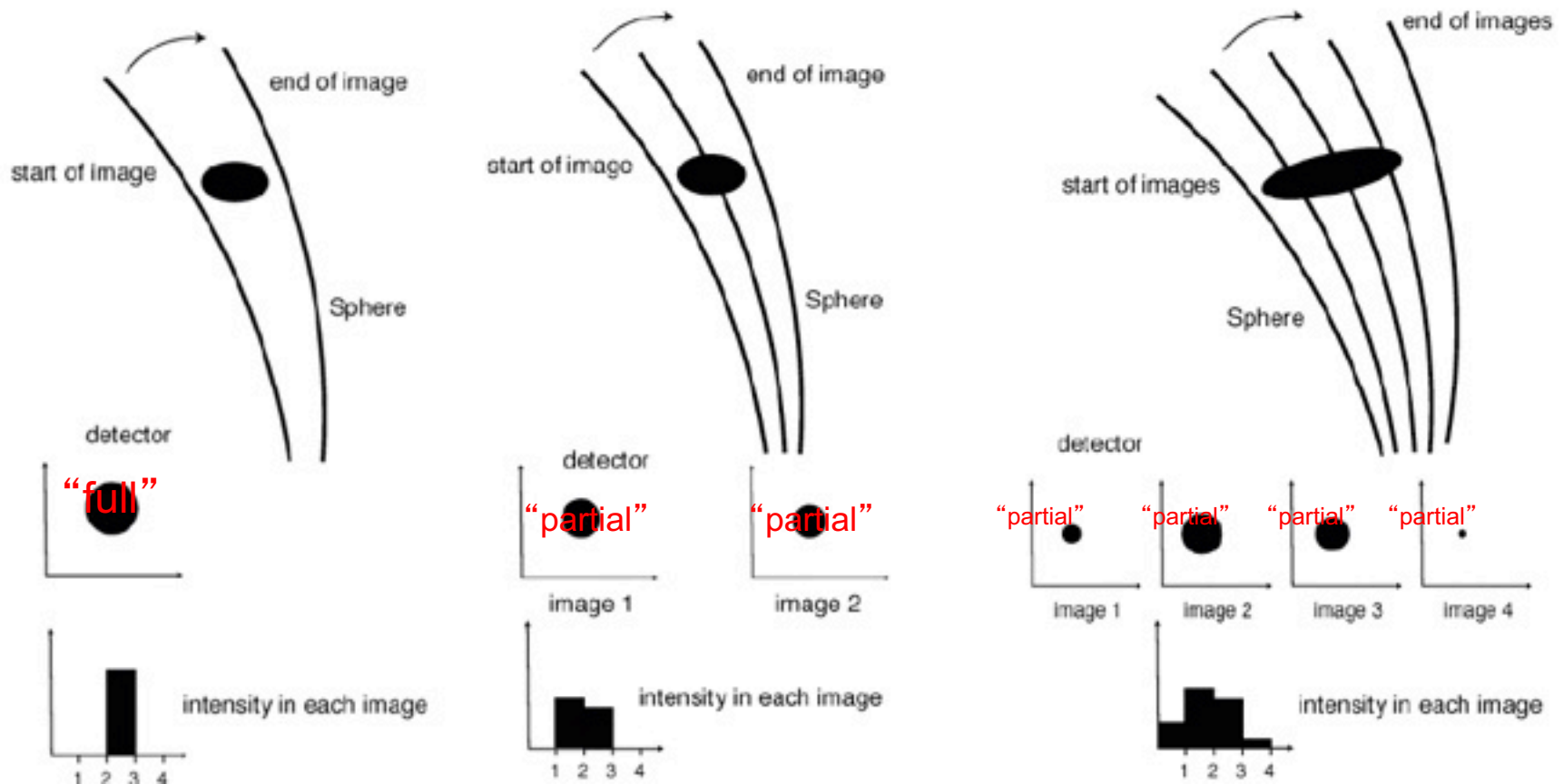
Fine ϕ -slicing

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



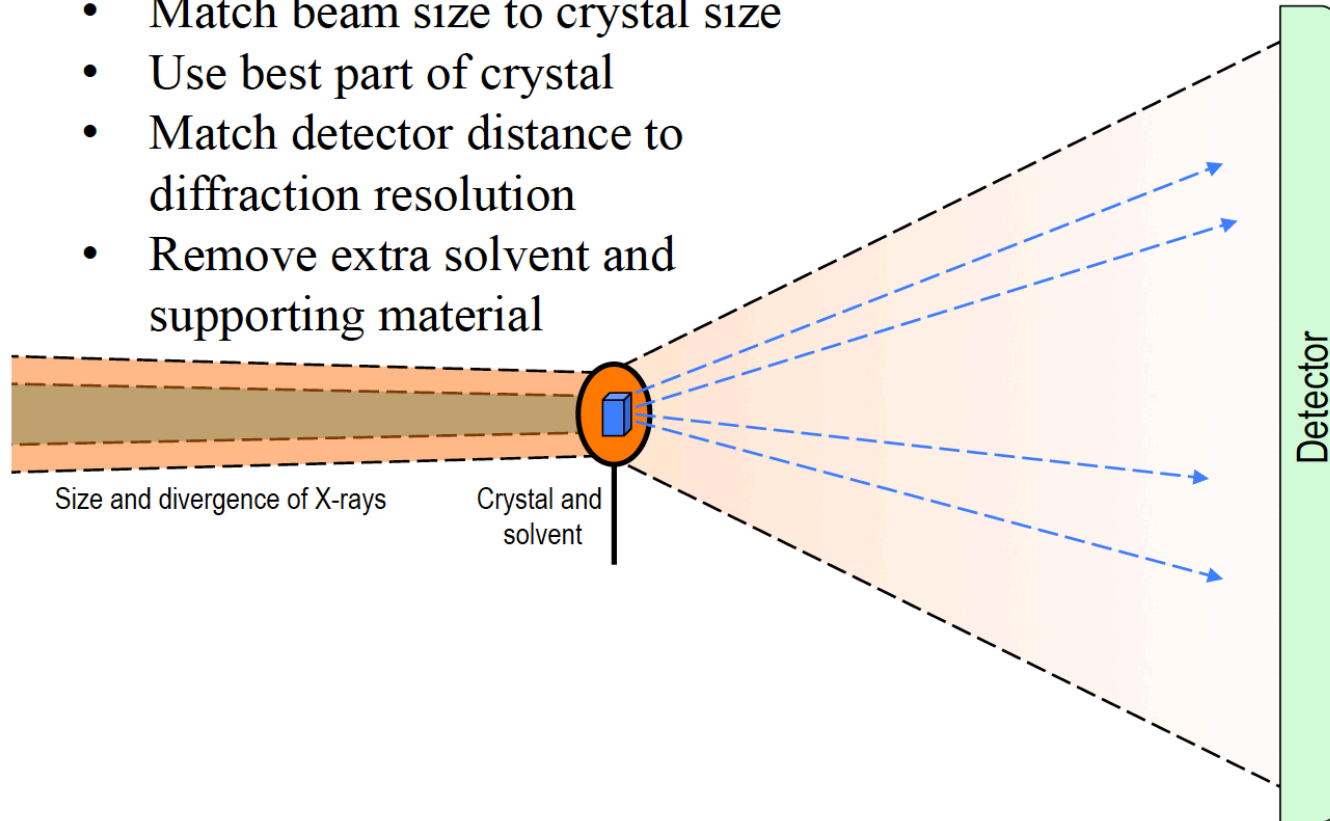
Fine-phi slicing data collection is enabled by the pixel array detector (PILATUS, EIGER), which has single-photon sensitivity and no readout noise

Data collection: the rotation/oscillation method

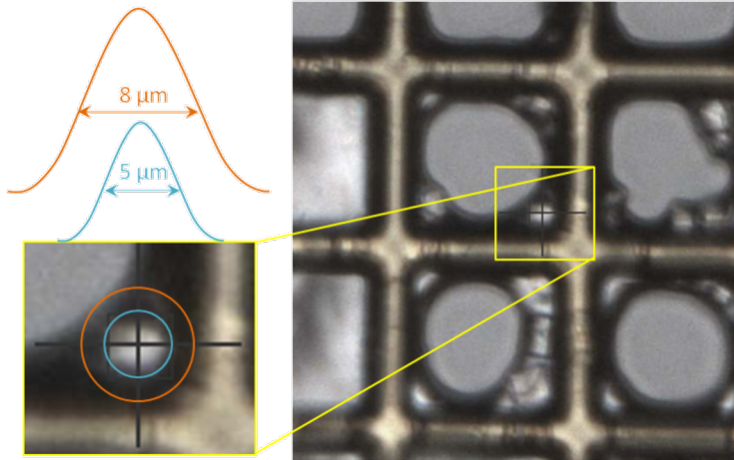


- spot size is equally affected by mosaicity, beam size and beam divergence
- programs that work with 3-d profile fitting (e.g. XDS, DIALS, ...) are better for dealing with "partials" spread over many images

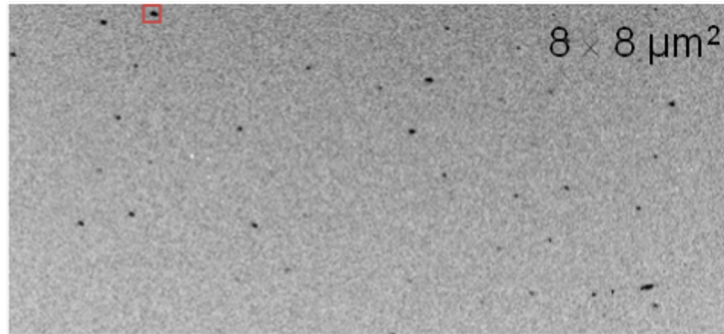
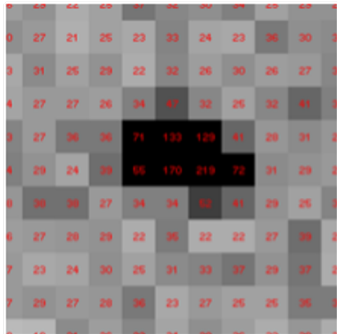
- Match beam size to crystal size
- Use best part of crystal
- Match detector distance to diffraction resolution
- Remove extra solvent and supporting material



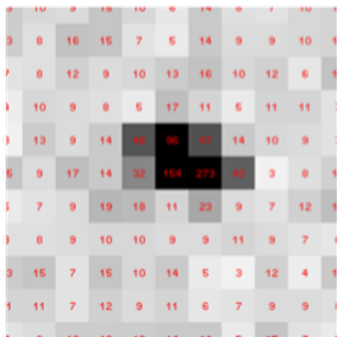
Match beam to sample



AcMNPV polyhedra crystals approx $5\times 5\times 5\text{ }\mu\text{m}^3$ in size were mounted on a micromesh grid. Data were collected with two beamsizes: $8\times 8\text{ }\mu\text{m}^2$ and $4.5\times 5\text{ }\mu\text{m}^2$



$$I/\sigma(I) = 1.5 \text{ at } 2.9\text{\AA}$$



Average background reduced by 3

$$I/\sigma(I) = 1.5 \text{ at } 2.5\text{\AA}$$

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. $50 \times 50 \mu\text{m}^2$ beam will deliver around 45x less dose than $9 \times 6 \mu\text{m}^2$ beam on I24

So match beam to crystal, but attenuate when required

Preparation



Sample centring



Screening, analysis & strategy



Data collection & evaluation



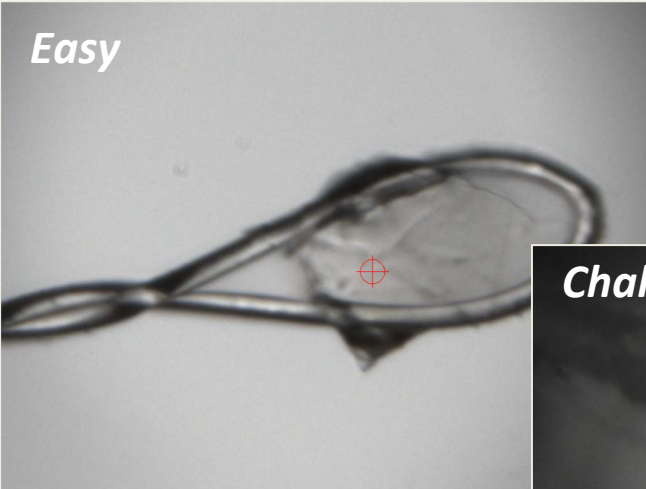
Post beamtime

Mount and centre crystal

Mounting now automated at most beamlines.

Centring can be:

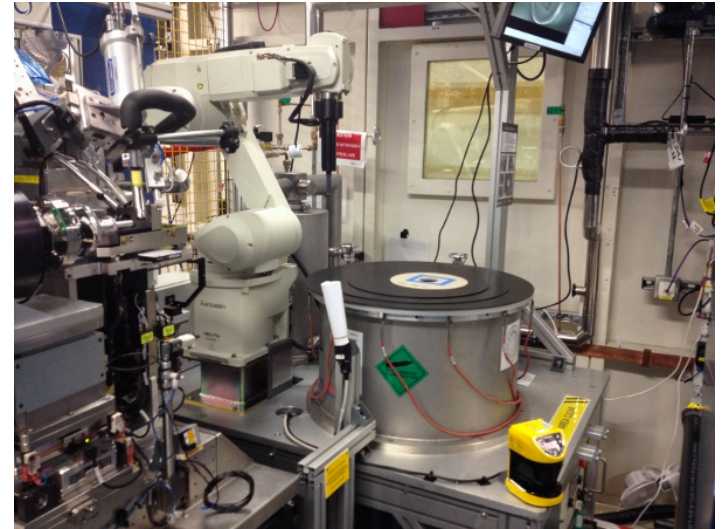
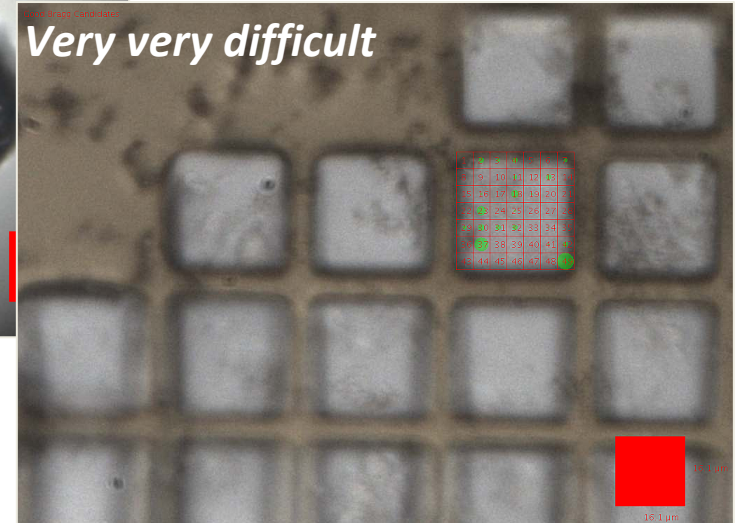
Easy



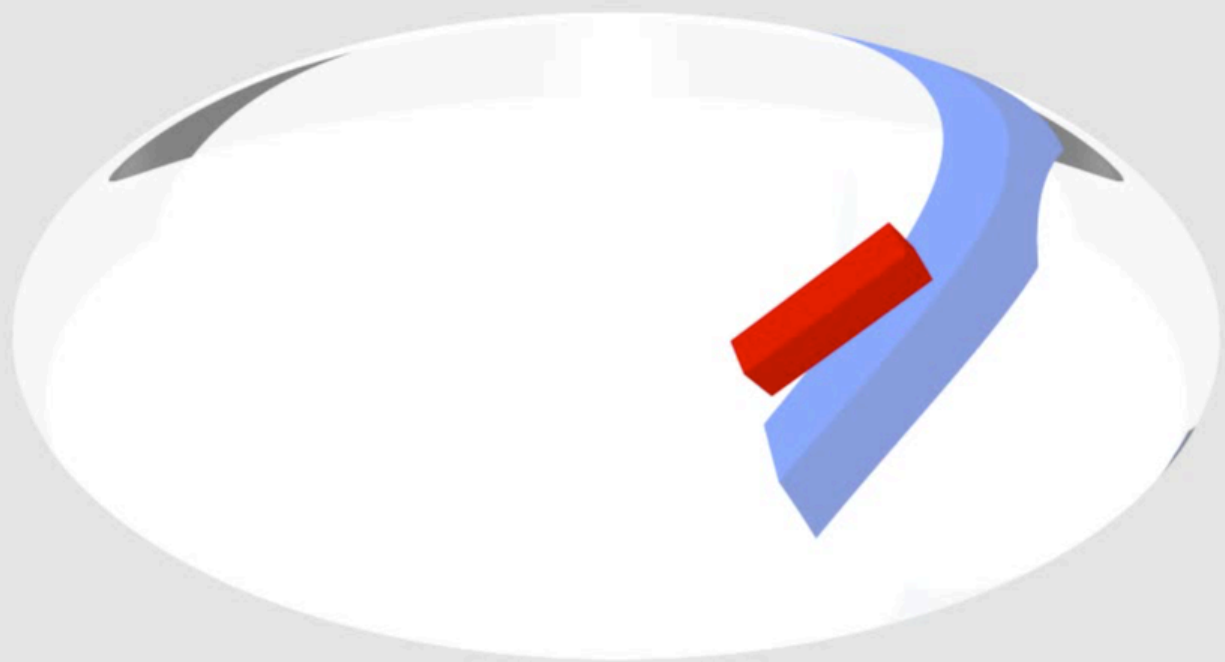
Challenging



Very very difficult

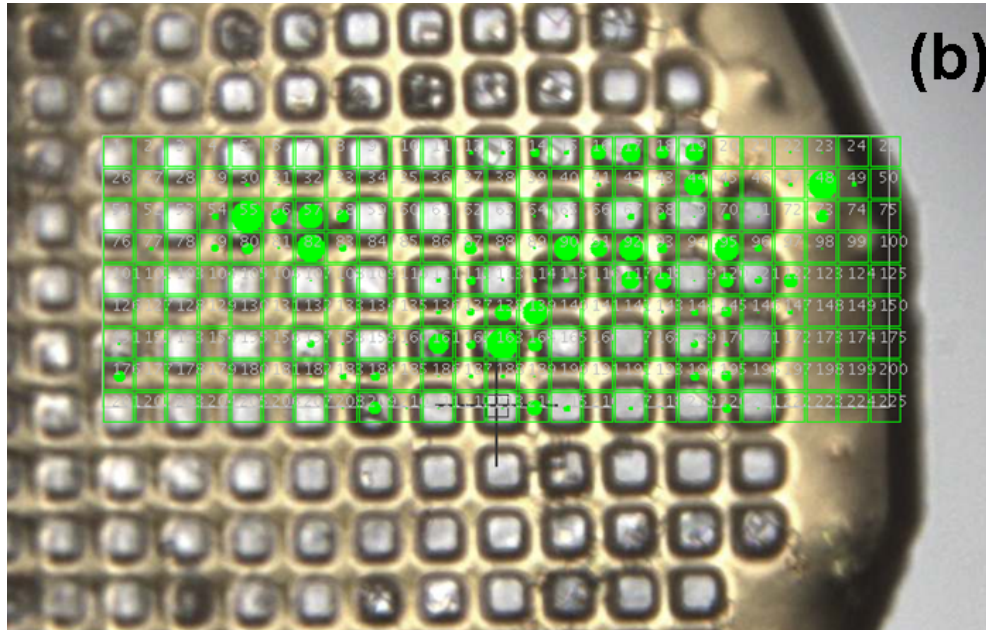


Make the most of tools
developed to make these
cases easier



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

28-11-2019 21:05:04 - 20191128/Sethaumatina/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

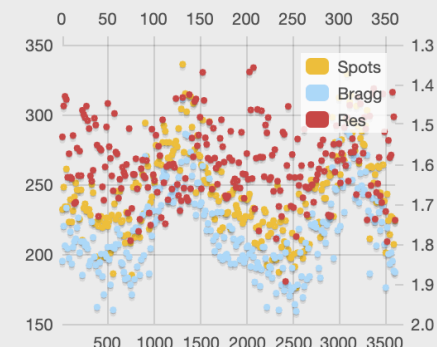
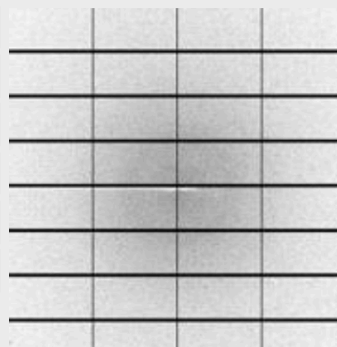
Transmission: 100.00%

Beamsize: 32x20μm

Type: SAD

Group: [4 Data Collections](#)

Comment: (34,-126,78) Aperture: Large



28-11-2019 20:58:37 - 20191128/Sethaumatina/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

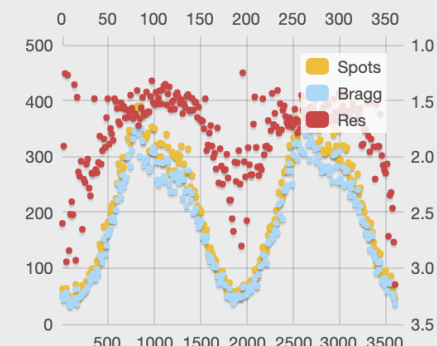
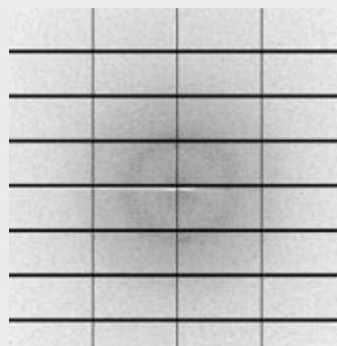
Transmission: 100.00%

Beamsize: 32x20μm

Type: SAD

Group: [4 Data Collections](#)

Comment: (-1111,-315,146) Aperture: Large



29-11-2019 09:14:05 - 20191129/Sethaumatina/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

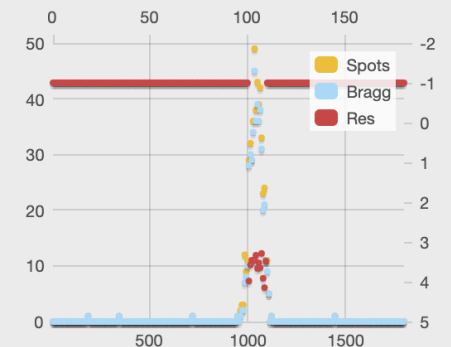
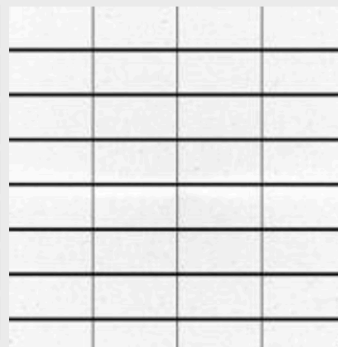
Exposure: 0.010s

Transmission: 100.00%

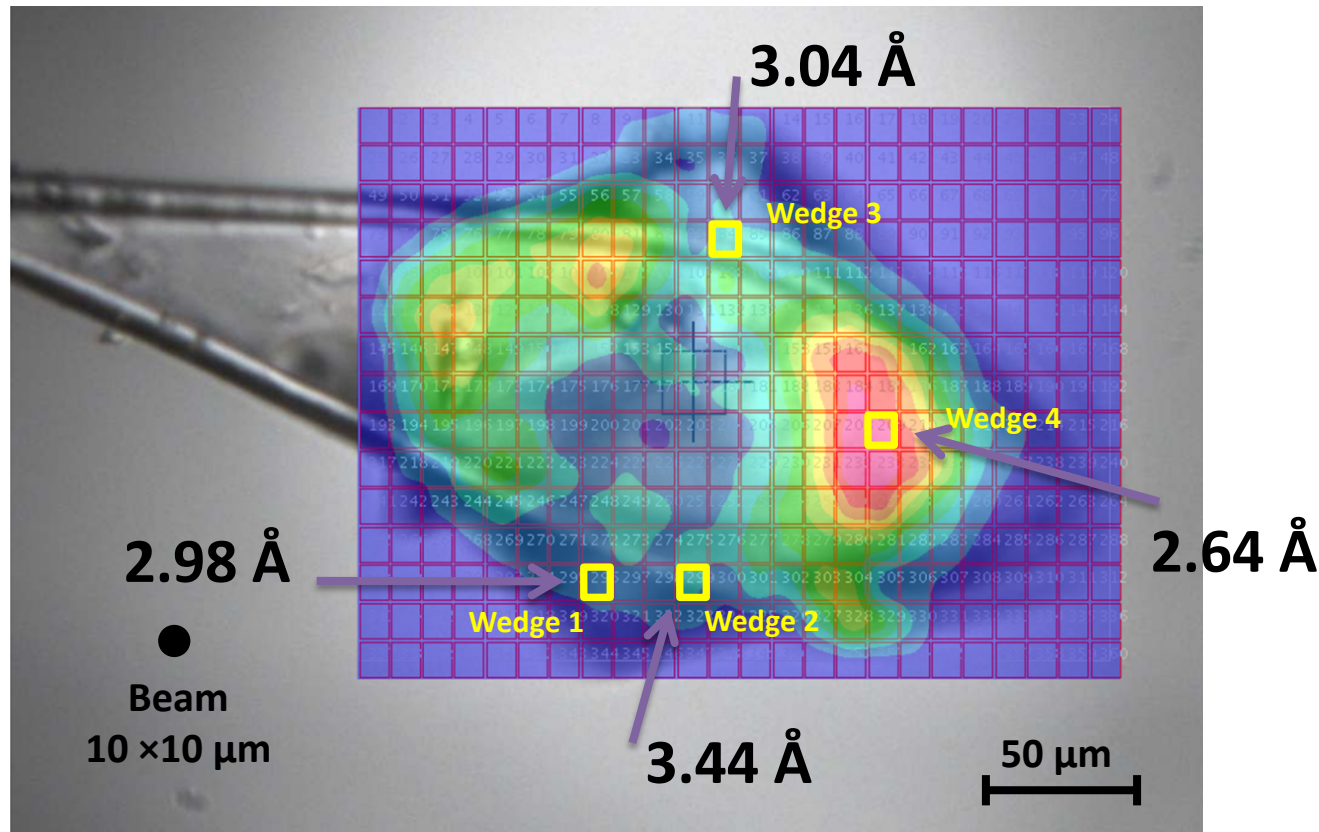
Beamsize: 32x20μm

Type: SAD

Comment: (0,0,0) Aperture: Large



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



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

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



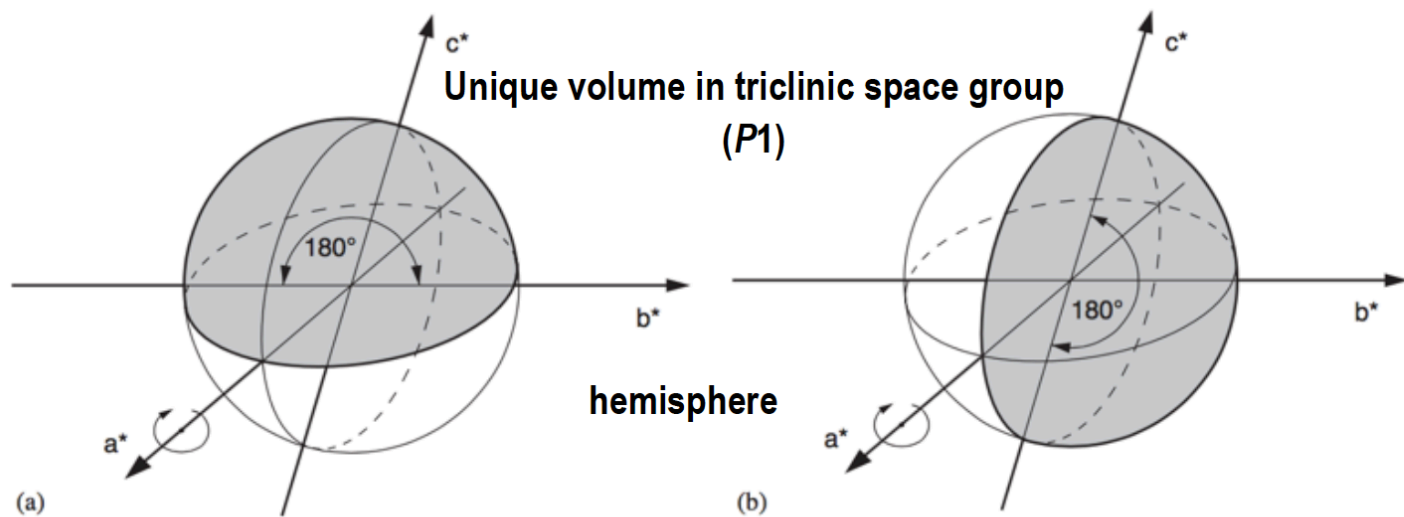
Data collection & evaluation



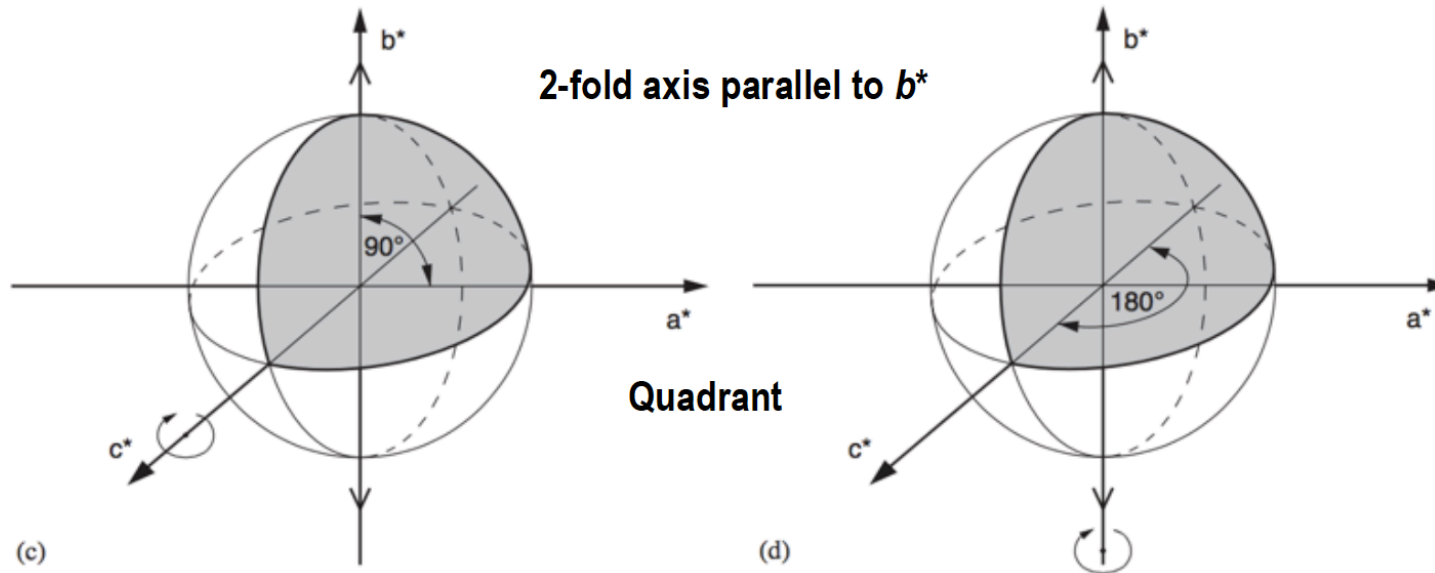
Post beamtime

Completeness - crystal symmetry

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



Completeness - crystal symmetry



- 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 $> 180^\circ$

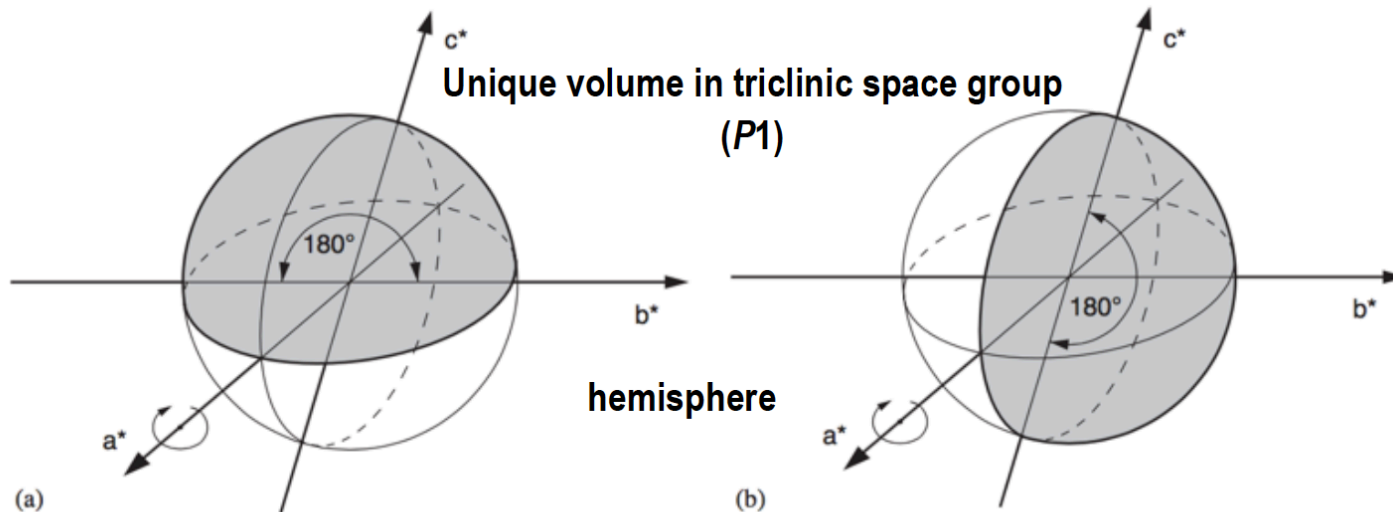
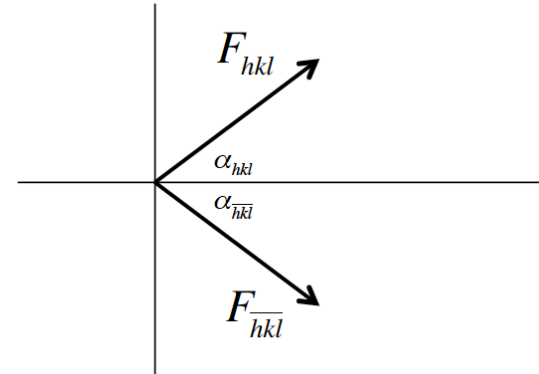
Symmetry in the diffraction pattern

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

$$F_{\bar{h}\bar{k}\bar{l}} = V \int_{\text{cell}} \rho(xyz) \exp[2\pi i(-hx - ky - lz)] dx dy dz$$

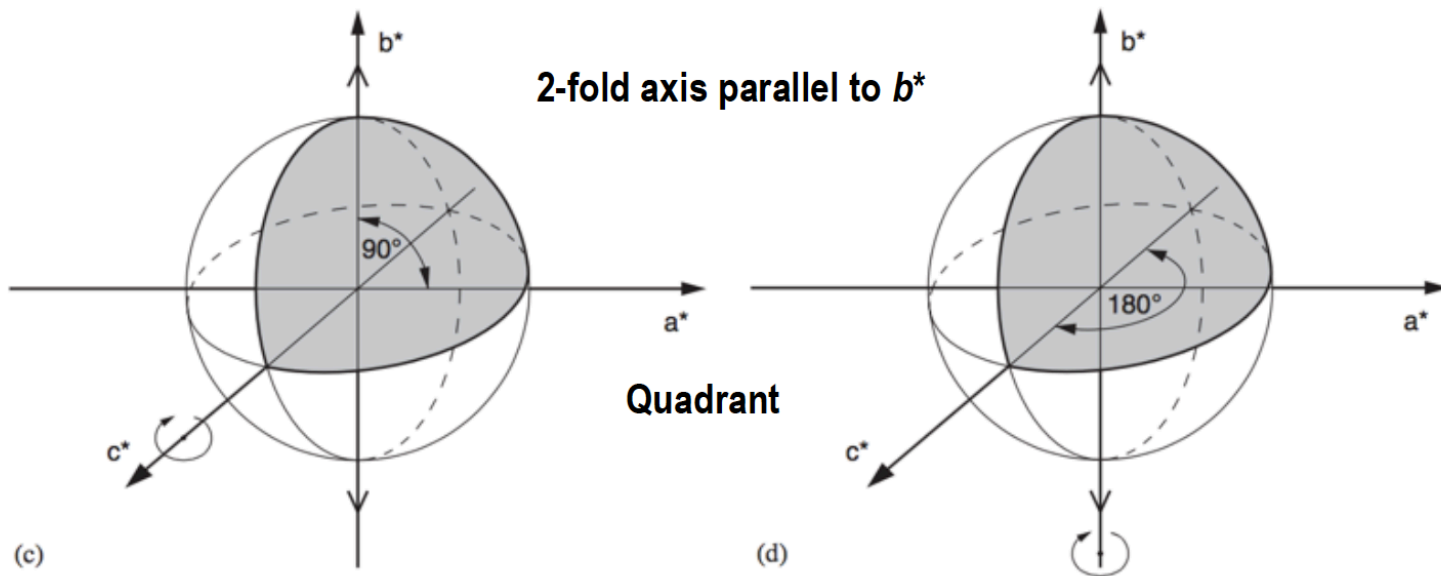
$$I_{hkl} \propto (F_{hkl})^2$$

$I_{hkl} = I_{\bar{h}\bar{k}\bar{l}}$ Friedel's law; However, $I_{hkl} \neq I_{\bar{h}\bar{k}\bar{l}}$ when anomalous scattering is present



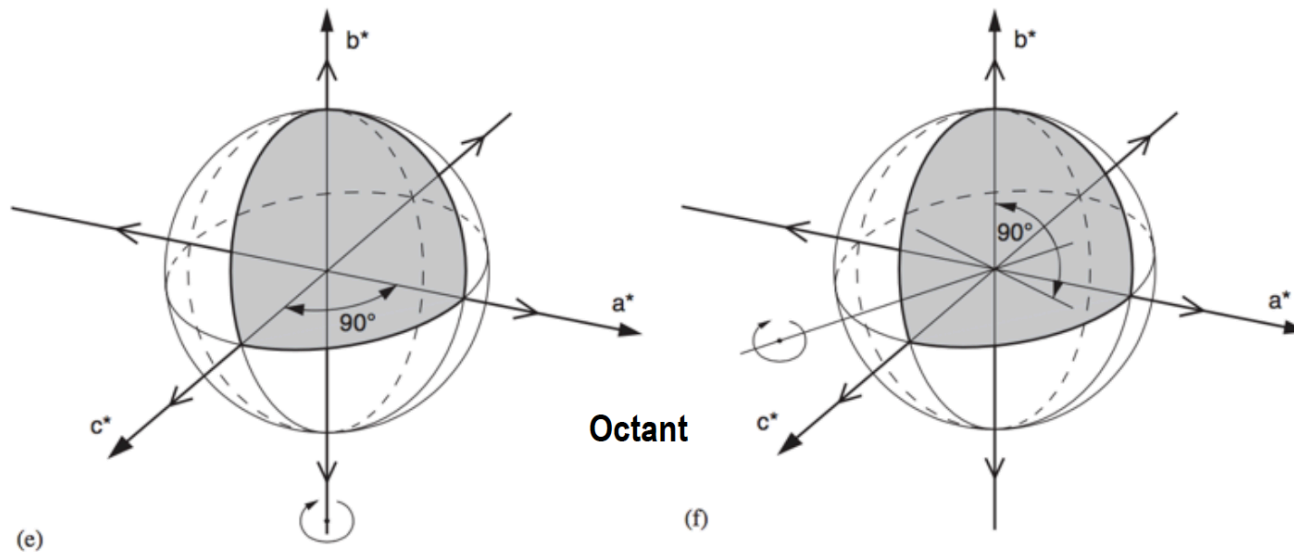
Unique volume in monoclinic space group ($P2$, $P2_1$, $C2$)

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



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

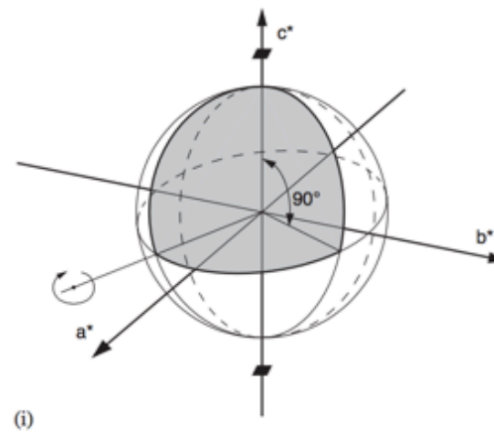
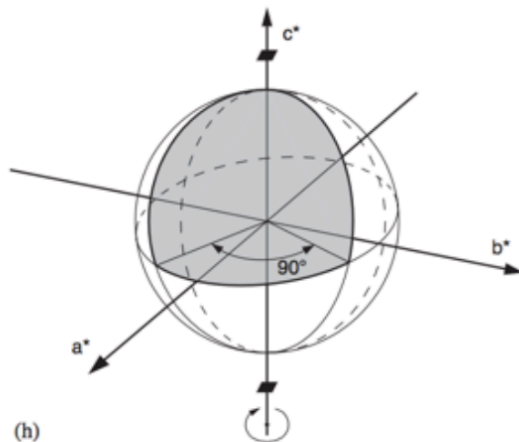
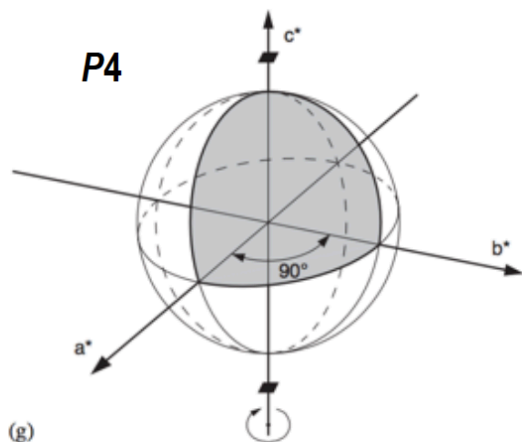
$$I_{hkl} = I_{\bar{h}kl} = I_{h\bar{k}l} = I_{hk\bar{l}} = I_{\bar{h}\bar{k}\bar{l}} = I_{h\bar{k}\bar{l}} = I_{\bar{h}k\bar{l}} = I_{hkl}$$



The 90° rotation must be between two axial orientations

Unique volume in tetragonal space groups

P4



P422

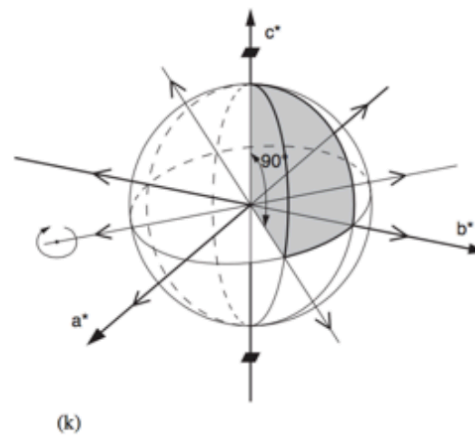
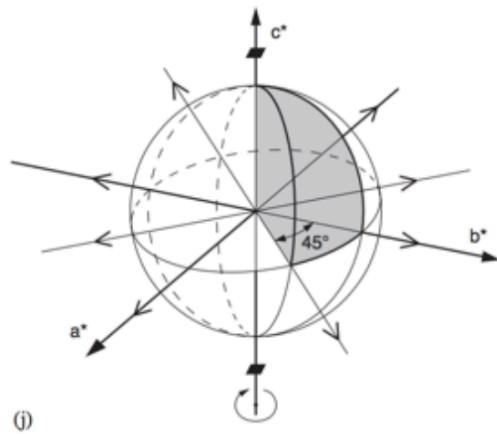
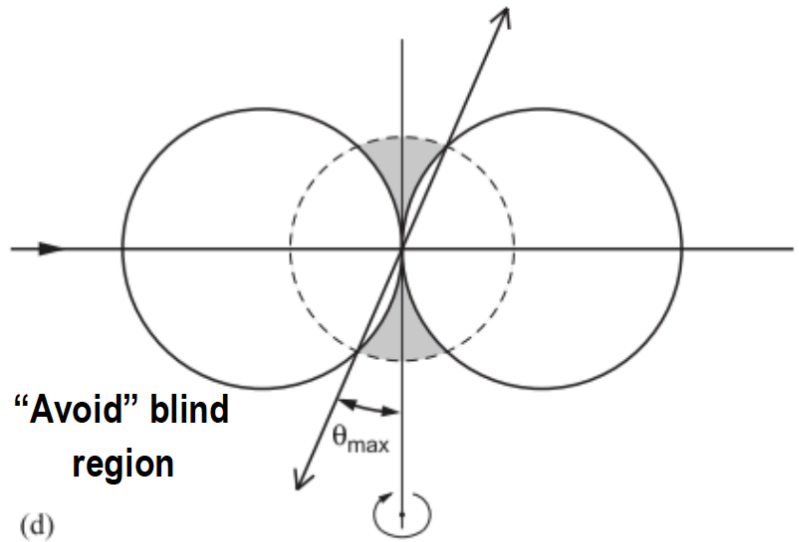
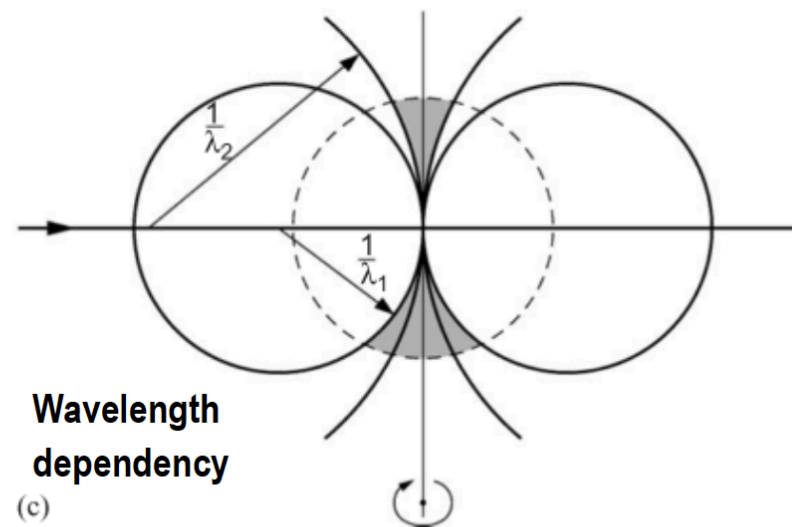
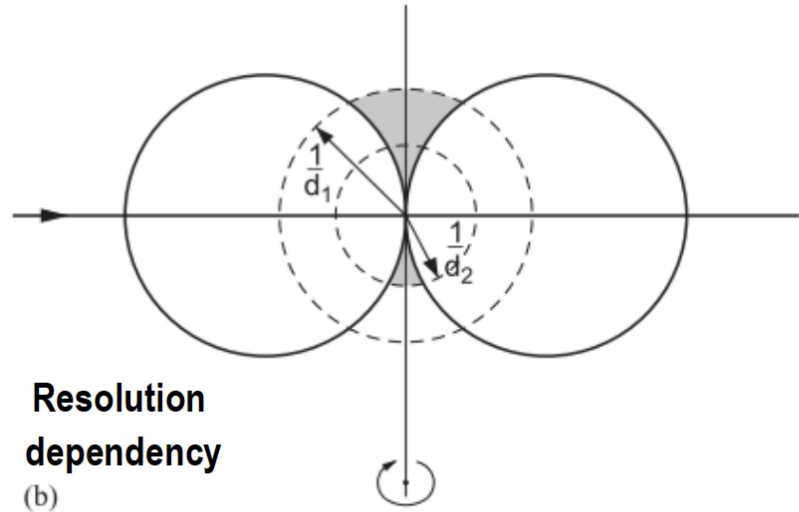
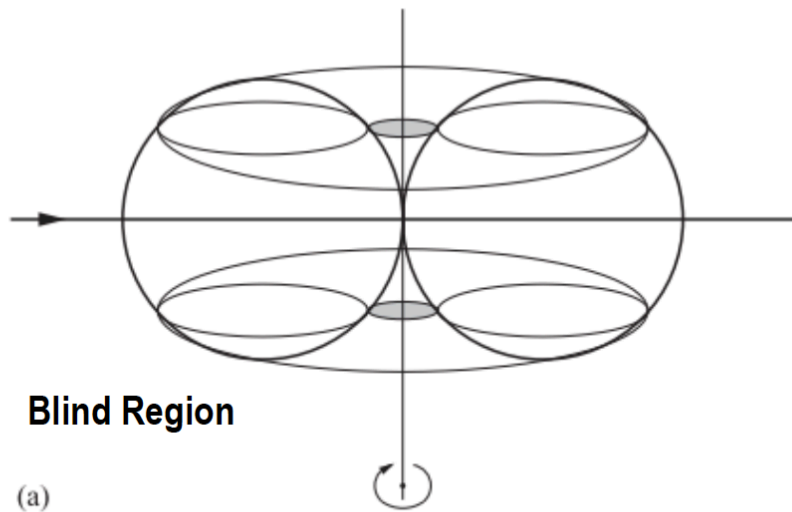


Table 1

Rotation range ($^{\circ}$) required to collect a complete data set in different crystal classes.

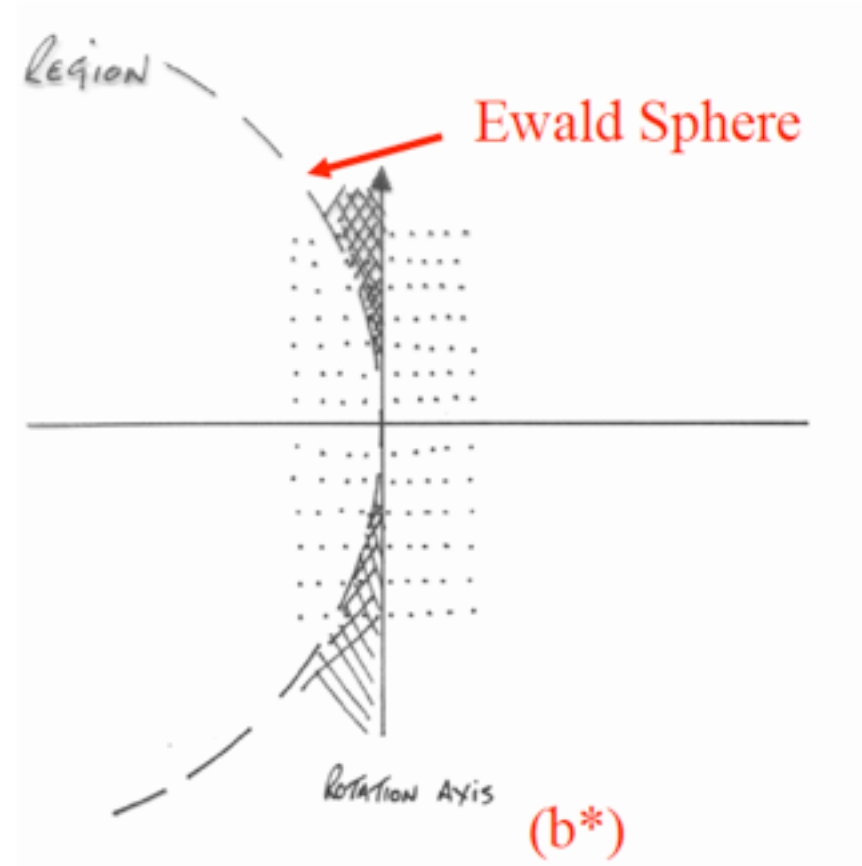
The direction of the spindle axis is given in parentheses; *ac* means any vector in the *ac* plane.

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



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.



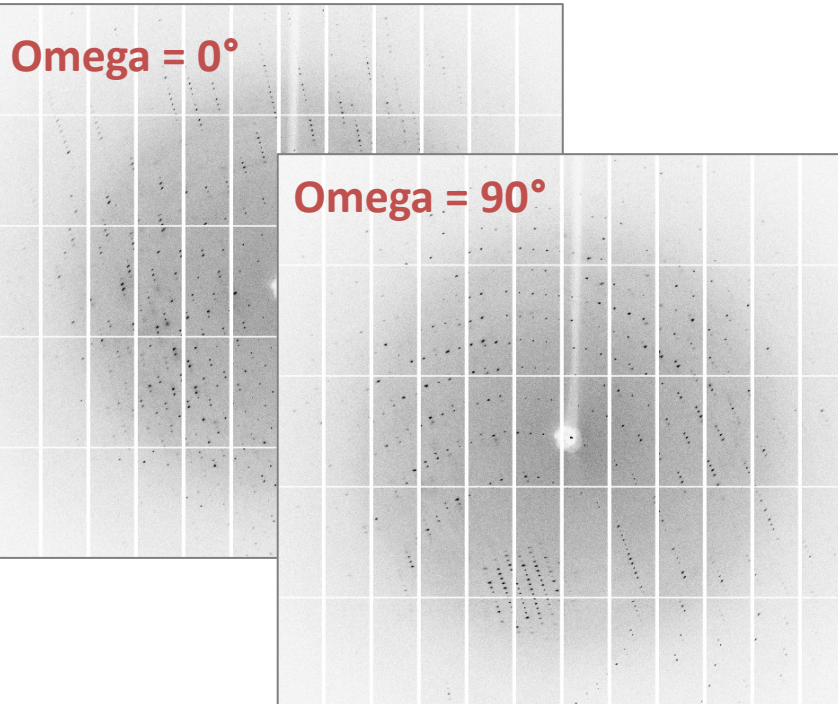
Screening and analysis

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?



Screening ?

- strategy / crystal assessment prior to dataset?

Depends if you are eager for speed, data quality or unattended data collection.

- 3 images at 0, 45 and 90 degrees are slow to take (3 arming of the detector) and slow to get results from but are useful particularly for new projects – implemented in GDA
 - Important if you want to use Chi/Phi with the smargon for example
- A small wedge of data (e.g. 20 degrees) that does not need to be “thrown away” but later merged with the dataset is also a good way – Gives more accurate resolution

- Stepped transmission? (*)

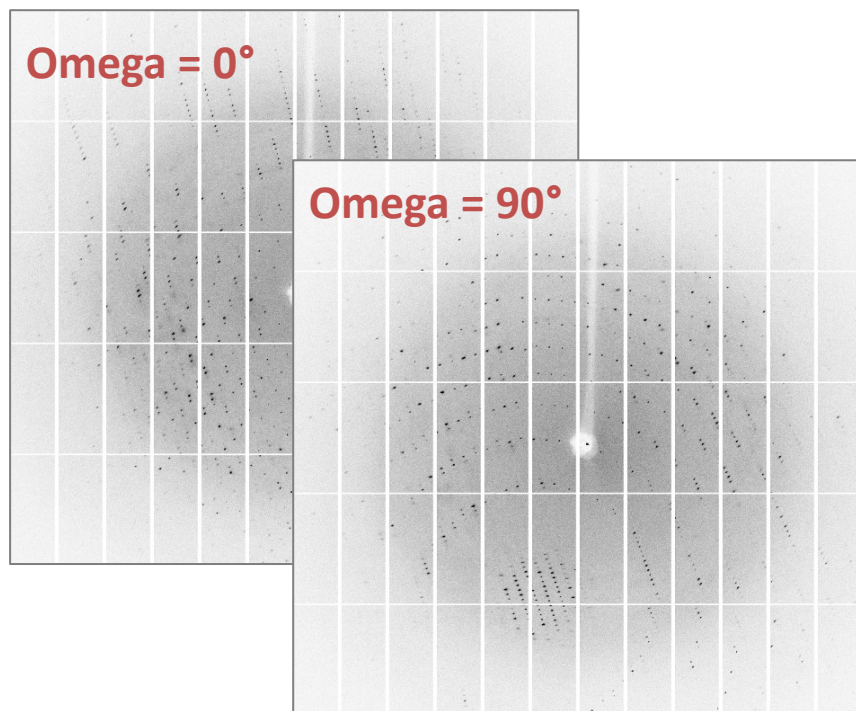
Is a good idea and prone to automation.

- Does not require any calculations on dose and knowing a priori beam flux, size or crystal composition as it collects several datasets from very conservative doses to higher doses.
- Does require analysis of which dataset is best, sometimes merging of datasets.
- Takes a little longer

(*) “How best to use photons” (2019) Acta Crystallogr D Struct Biol. 1;75:242-261. doi: 10.1107/S2059798319003528

Strategy

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

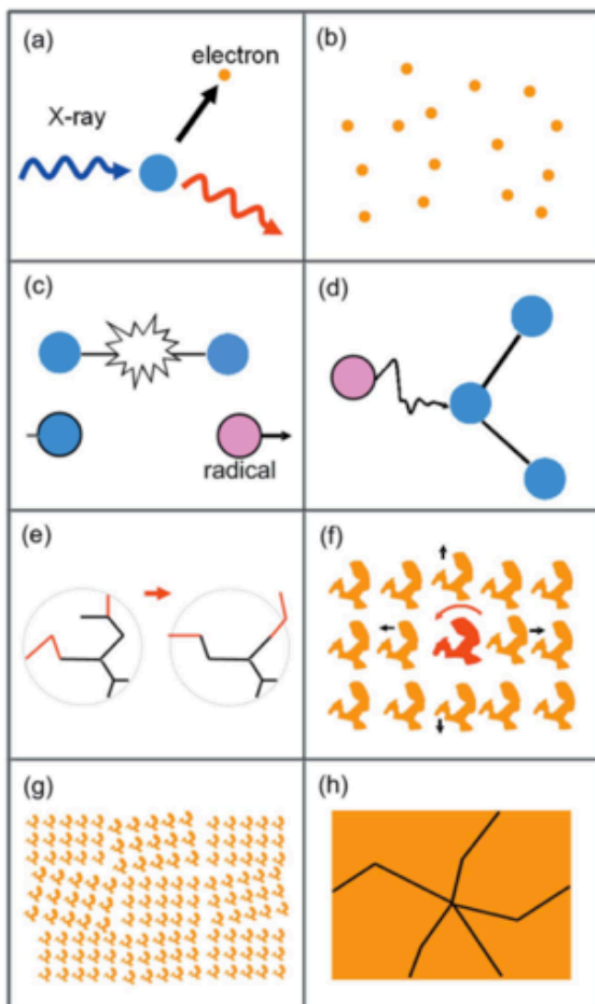


plus

- 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



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)

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

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

Radiation damage estimation at 100 K

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

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

dose-rate =

$$4 \times 10^{11} \text{ p/s} / (2000 \times 100 \mu\text{m} \times 100 \mu\text{m}) = 0.02 \text{ MGy/s}$$

$$4 \times 10^{11} \text{ p/s} / (2000 \times 10 \mu\text{m} \times 10 \mu\text{m}) = 2 \text{ MGy/s}$$

$$4 \times 10^{11} \text{ p/s} / (2000 \times 5 \mu\text{m} \times 5 \mu\text{m}) = 8 \text{ MGy/s}$$

$$4 \times 10^{11} \text{ p/s} / (2000 \times 1 \mu\text{m} \times 5 \mu\text{m}) = 40 \text{ MGy/s}$$

$$4 \times 10^{11} \text{ p/s} / (2000 \times 1 \mu\text{m} \times 1 \mu\text{m}) = 200 \text{ MGy/s}$$

- Another > flux beamline
- Phasing experiment

Single crystal
crystallography

10 seconds dataset
360 Hz (3600 images at 0.1
oscillation, 100 transmission)

Micro-crystallography

7.2 seconds dataset
500 Hz (3600 images at 0.1
oscillation, 35% transmission)

Serial crystallography

7.2 seconds dataset
500 Hz (3600 images at 0.1
oscillation, 7% transmission)

Serial crystallography

7.2 seconds dataset
500 Hz (3600 images at 0.1
oscillation, 1.4% trans)

Serial femtosecond
crystallography (xFEL)

2020

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

Online calculator: <https://bl831.als.lbl.gov/xtallife.html>. (allows resolution as a factor)

expected crystal lifetime calculator

<https://bl831.als.lbl.gov/xtallife.html>

Beamline settings	Flux (ph/s)	Beam size (um)	Exposure (s)	Dose (MGy)
A	4.5×10^{11}	100 x 100	1316	20
B	3.1×10^{11}	30x 20	80	20
C	3.0×10^{12}	80x15	19	20
D	3.0×10^{12}	8x6	1	20
E	3.0×10^{12}	20x20	6	20

Assumptions:

Beamline energy at 0.97 A

Crystal of the size of the beam

100% transmission

“standard” sample = no heavy metals (e.g. no Se, no As, average amount of S)

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 analysis

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

Use these results to update your data collection strategy.

```
Fast_DP installed in: /dls_sw/apps/fast_dp/2395
Starting image: /dls/i03/data/2016/cml4451-5/20161209/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/cml4451-5/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  1.91
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
```

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/cml4451-4/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/cml4451-4/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

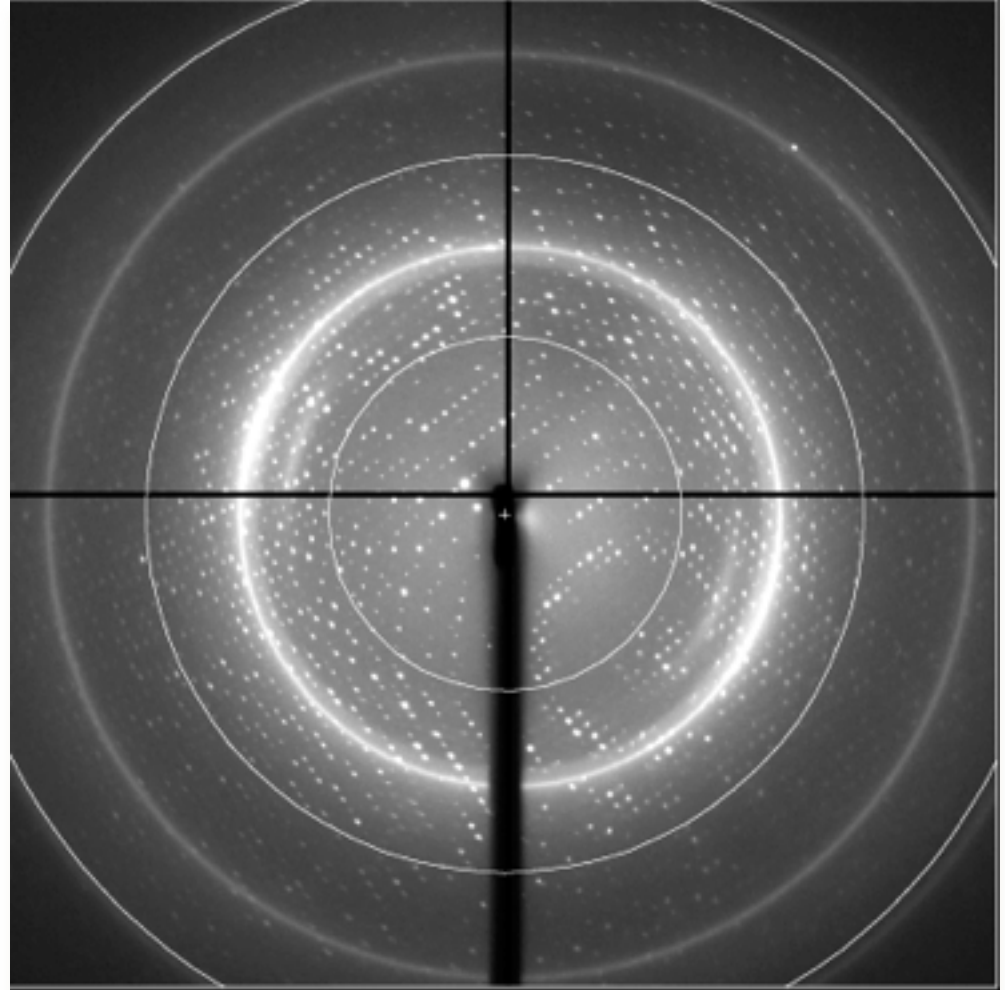
Data collection - the “native” data set

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)

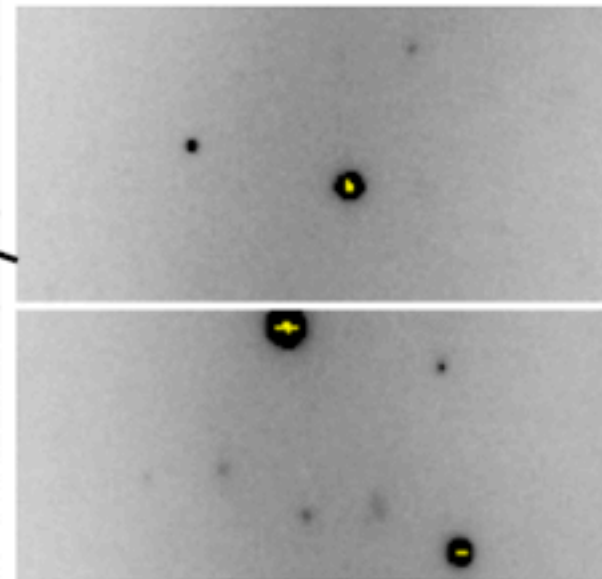
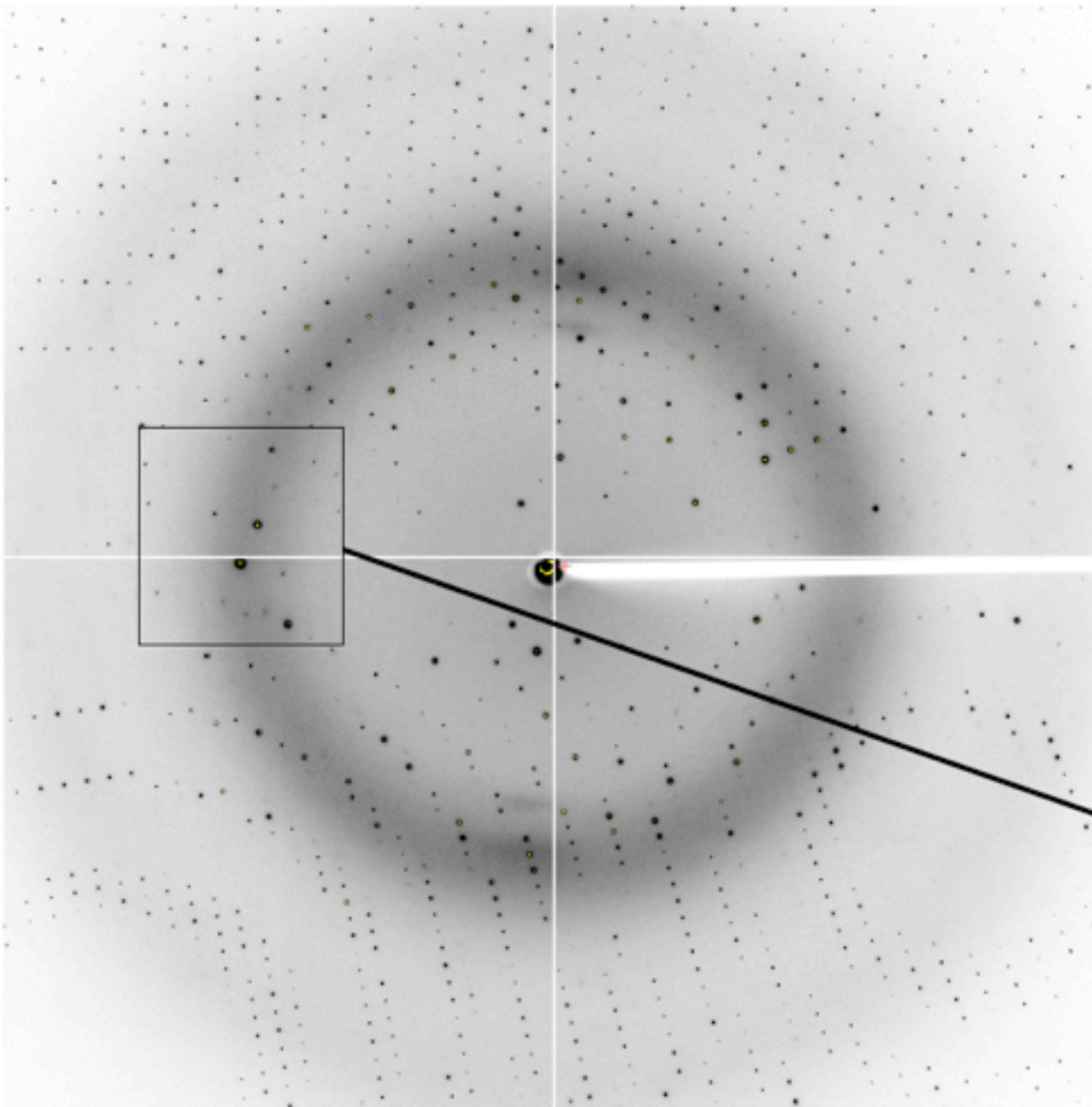
Data collection - molecular replacement

- 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



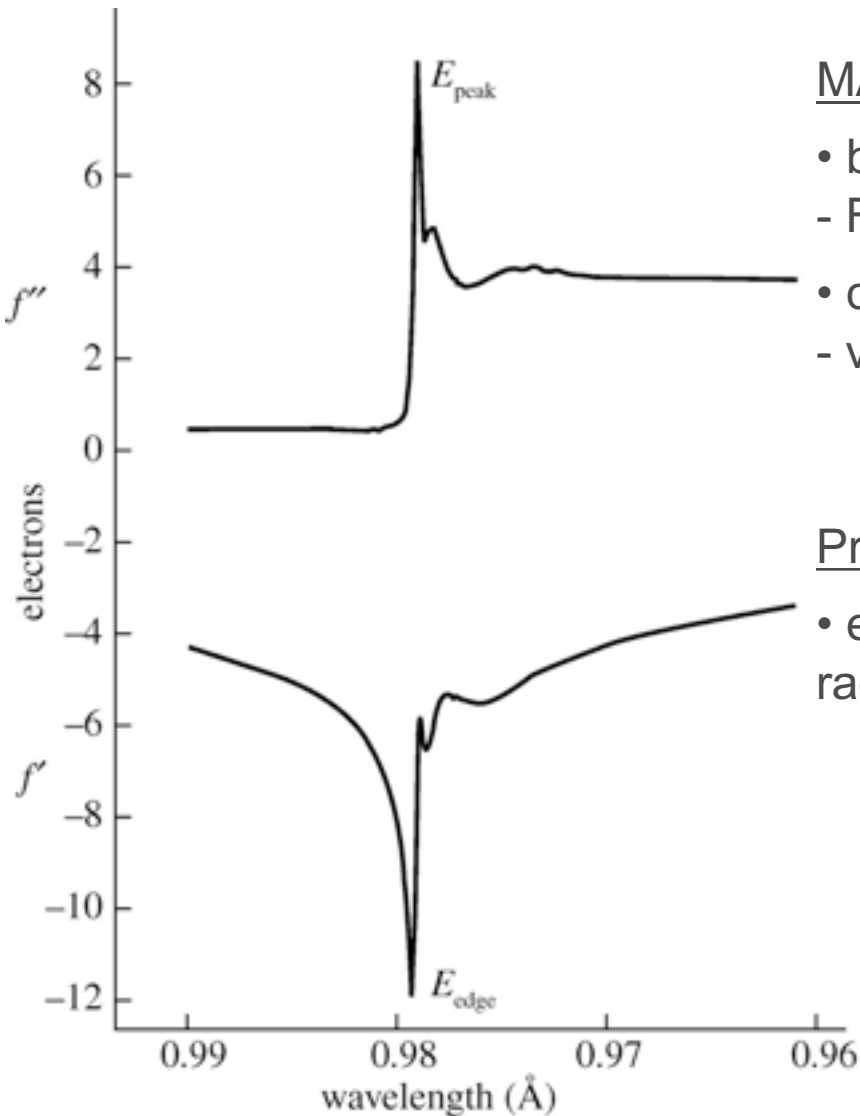
Overloads

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



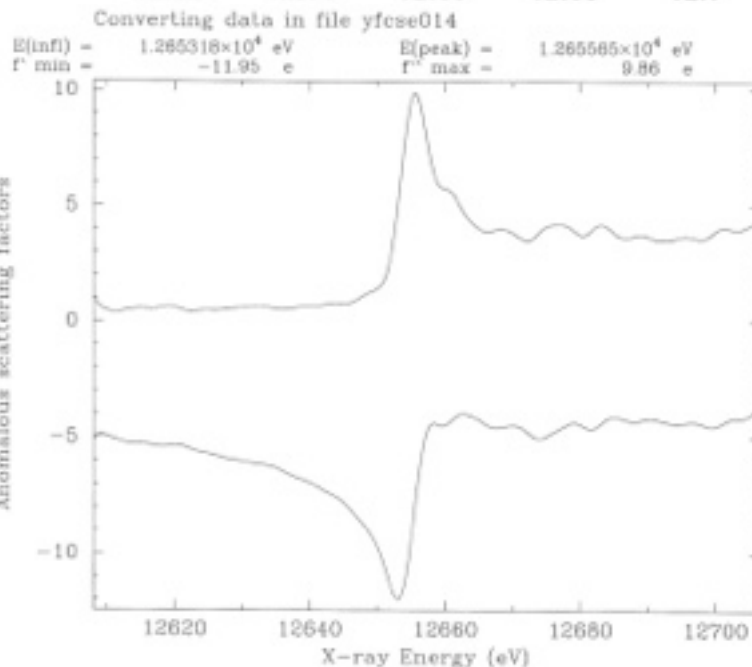
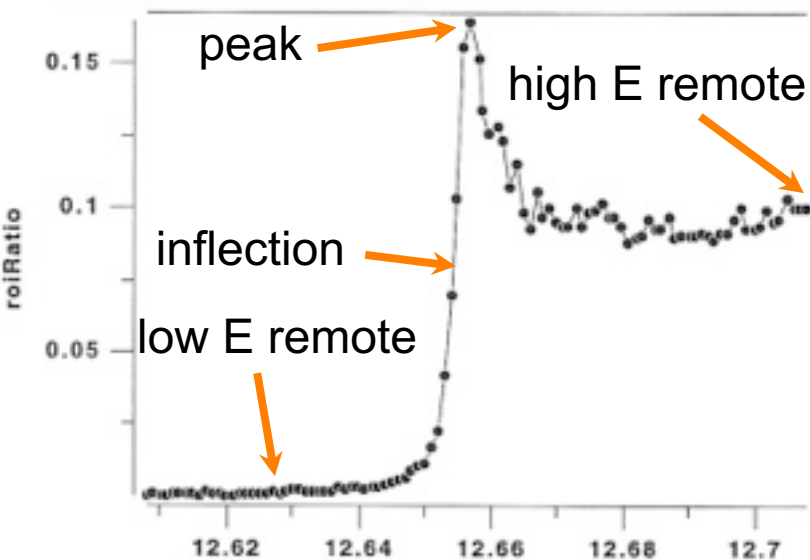
MAD/SAD:

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

Practical considerations:

- especially for SAD: the more **redundancy** (without radiation damage), the better!!!

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 (usually 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 absorption of X-rays by air at short wavelength / high energy
 - radiation damage
 - also check max. resolution and spot overlaps after changing wavelength (Bragg's Law: $n\lambda = 2d\sin\theta$)!
- Please calculate Dose (e.g. Raddose3D)

Should I use the same recipe for any energy?

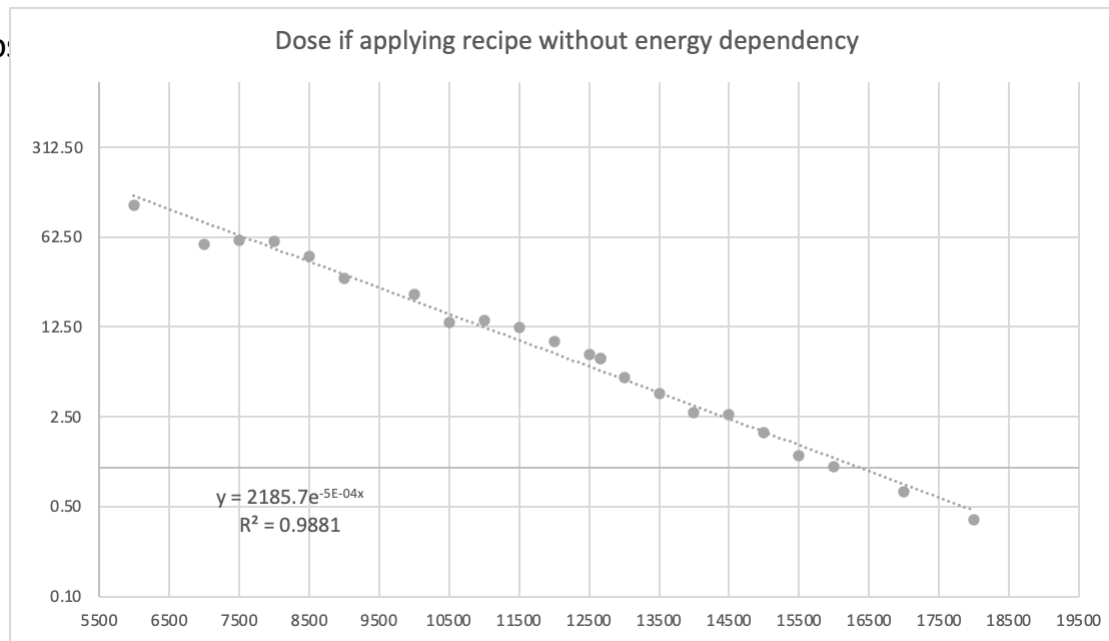
- 36 total exposure (two datasets of 18 seconds)
- 0.005 s/frame
- 7200 frames (2x 3600)
- Beam ~30x20 microns
- Energy: varying
- Same sample (for calculation assuming lysozyme but could be standard protein without any metals or special scatterer).

I04 Dose (*) vs Energy profile

(fixed exposure)

Set point 20 μm vertical

Log Dose
(Mgy)



110 MGy

Energy
(eV)

110X

0.4 MGy

energy	Wavelegh	beamsize.F.n_lenses		Dose if applying recipe without energy dependency
6000	2.070313	[28.6, 17.4]	9	109.76
7000.1	1.774528	[34.0, 21.9]	12	54.74
7500	1.65625	[29.7, 18.4]	14	59.44
8000.1	1.552715	[28.6, 17.4]	16	57.66
8500.1	1.46138	[29.5, 18.2]	18	44.46
9000	1.380208	[31.8, 20.1]	20	29.90
10000.1	1.242175	[28.6, 17.4]	25	22.23
10500	1.183036	[34.0, 21.9]	27	13.57
11000	1.129261	[30.8, 19.2]	30	14.04
11500.1	1.080154	[29.1, 17.8]	33	12.33
12000.1	1.035148	[28.6, 17.4]	36	9.71
12500	0.99375	[29.0, 17.7]	39	7.57
12658	0.981346	[28.9, 17.7]	40	7.14
13000	0.955529	[30.1, 18.7]	42	5.07
13500.1	0.920132	[31.8, 20.1]	45	3.79
14000	0.887277	[34.0, 21.9]	48	2.70
14500	0.856681	[31.4, 19.7]	52	2.58
15000	0.828125	[29.7, 18.4]	56	1.90
15500.1	0.801406	[33.3, 21.3]	59	1.25
16000.1	0.776362	[32.7, 20.8]	63	1.02
16999.9	0.730703	[31.1, 21.2]	71	0.65
18000.1	0.6901	[31.8, 20.1]	80	0.40

(Outliers removed)

(*) as calculated with raddose using current i04 recipe (2x 0.005s*3600frames = 2x 18 seconds = 36 seconds exposure) and a few assumptions (E.g. xtal size = beam size, standard protein composition etc)

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?).

Practical rules

These seem to work well on a majority of cases

- Default total angle: 360 ° per dataset ^(*),^(**)
- Default oscillation: 0.1° for all datasets ^(**)
- Use total exposure for dataset. For example for a beamline with flux and beamsize of 104 ^(***)
 - MR → 20 MGy (greedier on Max resolution), 10 MGy or 5MGy (greedier on time per dataset)
 - 72s at 100% transmission for 360° with 0.1° giving a total of 3600 images (50 Hz)
 - 36s at 100% transmission for 360° with 0.1° giving a total of 3600 images (100 Hz)
 - 18s at 100% transmission for 360° with 0.1° giving a total of 3600 images (360 Hz)
 - SAD, MAD → 5 MGy ^(***) in reasonable time
 - 29s at 62% transmission for 1,440° with 0.1° giving a total of 14,400 images (~500 Hz)
 - Low anomalous signal SAD, MAD (e.g. S, P etc) → Spread the dose through more images
 - 60s at 5% transmission for 2,400° with 0.1° giving a total of 24,000 images (400 Hz)

Issues:

- Needs to index and as you drop flux this can be harder but trust that indexing happens in 3D so if you need to rationalise then merge images to 1° using ADXV, Albula, Dials and do not look at fine sliced images.

(*) even for high symmetry space groups but if possible get best starting angle from strategy. Also consider collecting datasets with different Chi/Phi if goniometer allows

(**) Discussion yesterday on slack channel [#01-Monday-30th-november](#)

(***) Don't think in terms of exposure per frame and think in total dose you want to load the crystal with. That directly translates to a number of seconds full beam for each particular beamline / crystal composition, size and maximum resolution you need to obtain . Estimate this number using the crude online calculator (<https://bl831.als.lbl.gov/xtallife.html>) or even better the online Raddose 3D interface (<http://www.raddo.se/rd3d/prepare.php>)

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

(*) Heads up that Diamond does not run EP pipelines if you don't define a heavy atom on your pucks/pins & won't run MR pipelines if you don't give a PDB (code or coordinates)

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, $n \times 360^\circ$ 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

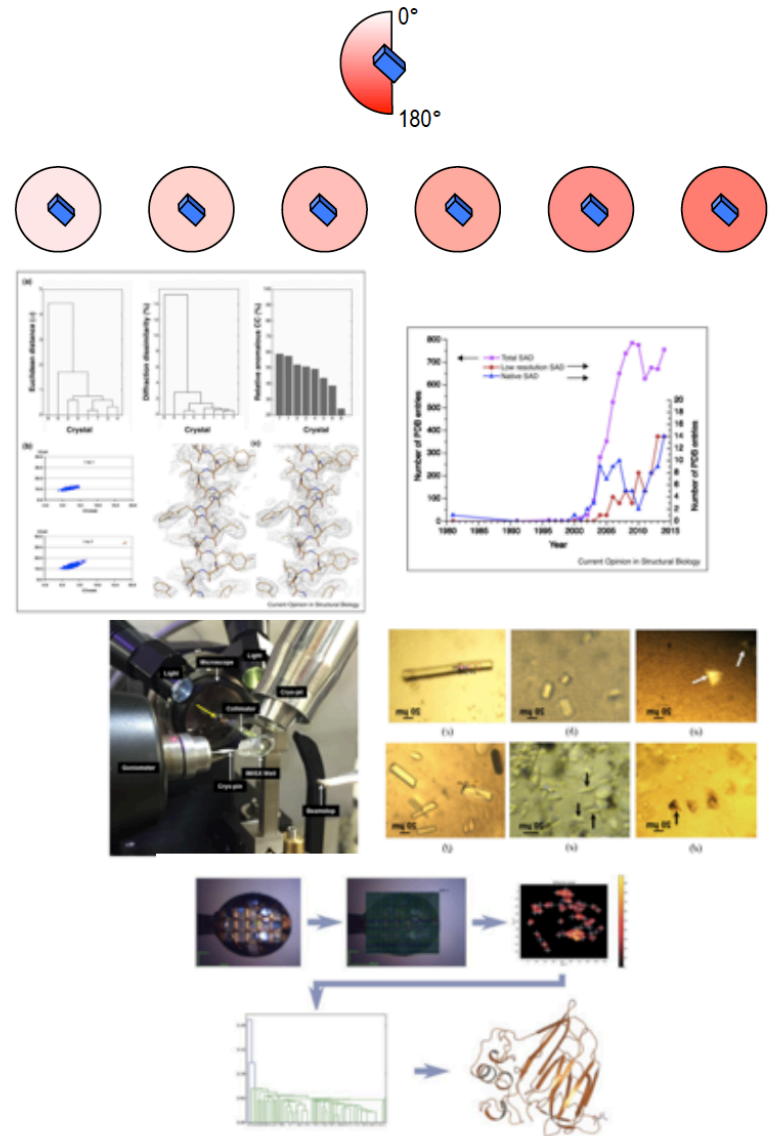
- Distribute x-ray dose to multiple data sets and multiple crystals
- Liu & Hendrickson *Curr. Opin. Struc. Bio.* **34**, 99 (2015)
- Weinert *et al. Nature Methods* **12**, 131, (2015)

Micro-crystallography

- 10-20° data per crystal $\times n$ 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)



References on dose choice

10 MGy/A resolution:

Howells MR, Beetz T, Chapman HN, Cui C, Holton JM, Jacobsen CJ, Kirz J, Lima E, Marchesini S, Miao H, Sayre D, Shapiro DA, Spence JHC & Starodub D (2009)*J. Electron Spectrosc. Relat. Phenom.* **170**, 4-12.

1% non-isomorphism per MGy:

Banumathi S, Zwart PH, Ramagopal UA, Dauter M & Dauter Z (2004)*Acta Cryst. D* **60**, 1085-1093.

200 kGy for Room Temperature:

Warkentin M, Badeau R, Hopkins JB, Mulichak AM, Keefe LJ & Thorne RE (2012). *Acta Cryst. D* **68**, 124-133.

Barker AI, Southworth-Davies RJ, Paithankar KS, Carmichael I & Garman EF (2009)*J. Synch. Rad.* **16**, 205-216.

Blake CCF & Phillips DC (1962)., pp. 183-191. Vienna: IAEA.

5 MGy for Se-Met:

Holton JM (2007)*J. Synch. Rad.* **14**, 51-72.

4 MGy for Hg-Cys:

Ramagopal UA, Dauter Z, Thirumuruhan R, Fedorov E & Almo SC (2005).", *Acta Cryst. D* **61**, 1289-1298.

2 MGy for Cys-Cys:

Murray JW & Garman EF (2002)*J. Synch. Rad.* **9**, 347-354.

500 kGy for Br-RNA:

Olieric V, Ennifar E, Meents A, Fleurant M, Besnard C, Pattison P, Schiltz M, Schulze-Briesse C & Dumas P (2007, *Acta Cryst. D* **63**, 759-768.

500 kGy for Photosystem II:

Yano J, Kern J, Irrgang K-D, Latimer MJ, Bergmann U, Glatzel P, Pushkar Y, Biesiadka J, Loll B, Sauer K, Messinger J, Zouni A & Yachandra VK (2005)*PNAS USA* **102**, 12047-12052.

60 kGy for putidaredoxin:

Corbett MC, Latimer MJ, Poulos TL, Sevrioukova IF, Hodgson KO & Hedman B (2007, *Acta Cryst. D* **63**, 951-960.

60 kGy for bacteriorhodopsin:

Borshchevskiy V, Round E, Erofeev I, Weik M, Ishchenko A, Gushchin I, Mishin A, Willbold D, Buldt G & Gordeliy V (2014)*Acta Cryst. D* **70**, 2675-2685.

20 kGy for Fe reduction in myoglobin:

Radiat Phys Chem Oxf Engl 1993 **76**, 714-721.

rough rotisserie factor:

Holton JM (2009). "A beginner's guide to radiation damage", *J. Synch. Rad.* **16**, 133-142.

more accurate rotisserie factor calculations:

Zeldin OB, Brockhauser S, Bremridge J, Holton JM & Garman EF (2013)*PNAS USA* **110**, 20551-20556.

GENERAL
ASSEMBLY
& CONGRESS
- IUCr 2023

MELBOURNE, AUSTRALIA

22-29 AUGUST 2023



**GENERAL
ASSEMBLY
& CONGRESS
- IUCr 2023**

MELBOURNE, AUSTRALIA

22-29 AUGUST 2023



**GENERAL
ASSEMBLY
& CONGRESS
- IUCr 2023**

MELBOURNE, AUSTRALIA

22-29 AUGUST 2023

