Advanced Workshop on Structural Biology:

Using Synchrotron Radiation to Visualise Biological Molecules

15-19 December 2014

Concerning the nature of things: the data collection itinerary

michele.cianci@embl-hamburg.de

The Abdus Salam International Centre for Theoretical Physics
50th Anniversary 1964-2014

EMBL 40 YEARS

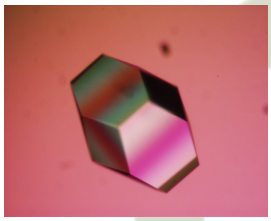
This set of talks

- Gives an overview of the data collection:
  - Experimental set up;
  - sample mounting;
  - Philosophy of data collection;
  - Radiation damage;
  - Pilatus data collection;
- Slides unashamedly adapted from:
  - Zbigniew Dauter (NCI, Brookhaven, NY);
  - Kay Diederich (Konstanz University, DE);
  - James Holton (LBNL, Berkeley, CA);
  - Elspeth Garman (Oxford University, UK);
  - Thomas Schneider (EMBL, Hamburg, DE);
  - Paul Tucker (EMBL, Hamburg, DE);



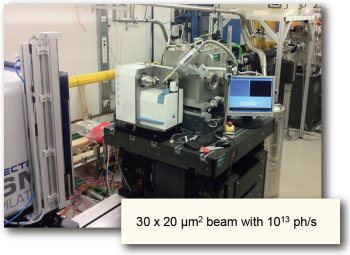



Essentials:



P13 - large cells, low energies

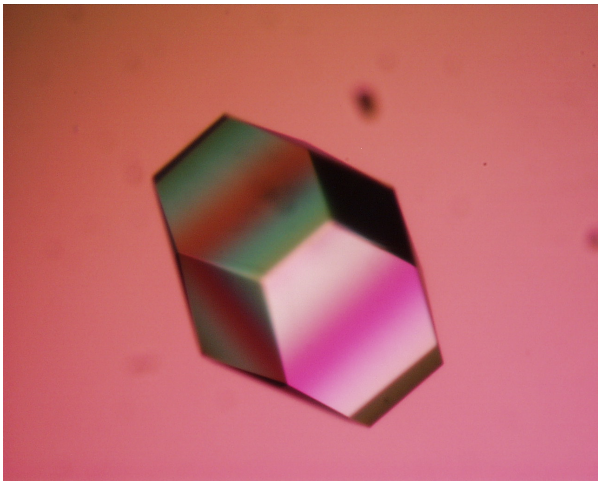
Endstation



You are essential because of:

- Your knowledge:
  - What's protein?
  - How has it been cloned/expressed/purified?
- Your questions:
  - What am I expecting to learn?
  - New structure? New bound ligand?
- Your decisions:
  - How to mount my sample?
  - What's the best instrument for my experiment?
  - How many data I need?

No crystal? no party…yet



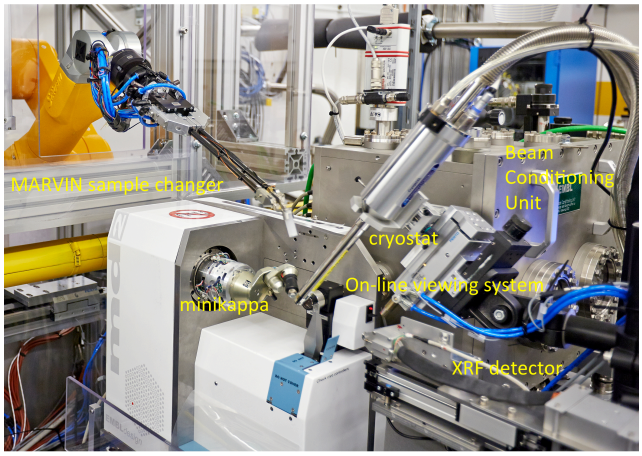
Know your instrument – I: basic parameters

application	parameter	MX1
Highly brilliant and highly stable beam,	**Main purpose**	Petra III, 3(1/2)rd generation light source, low emittance 1 nmrad
Wide range tunability, broad spectrum of experimental phasing methods *in crystallo* spectroscopy	**Energy range**	5(4)–16 keV
Matching the beam to the size of the crystal	**Focus H / V**	29 x 23 μm (10-5 μm with collimation) (100 μm defocussed)
Large unit cells, low mosaicity crystals	**Divergence H x V**	0.2 mrad x 0.15 mrad
Fast data collection, small beam	**Intensity, ph/s**	Up to $1*10^{13}$ ph/s

- Pilatus 6M has also been optimized for low energy data collection.
- Focused beam data collection will reduce data collection time to few minutes per data set. Highly redundant/Multi crystal data collections feasible.

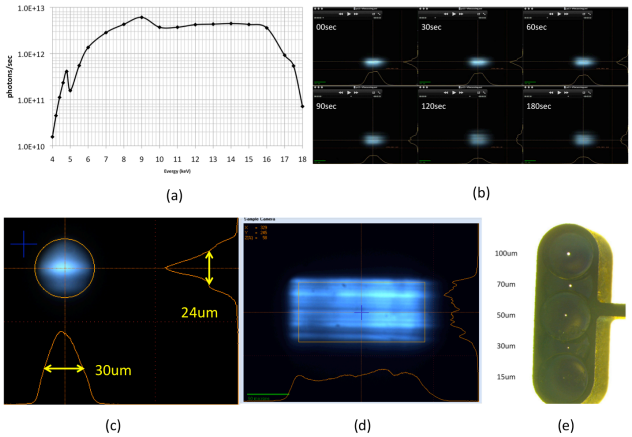
(data taken shown for P13@EMBL-Hamburg.de)

Know your instrument – II: geometry of the camera



(data taken shown for P13@EMBL-Hamburg.de)

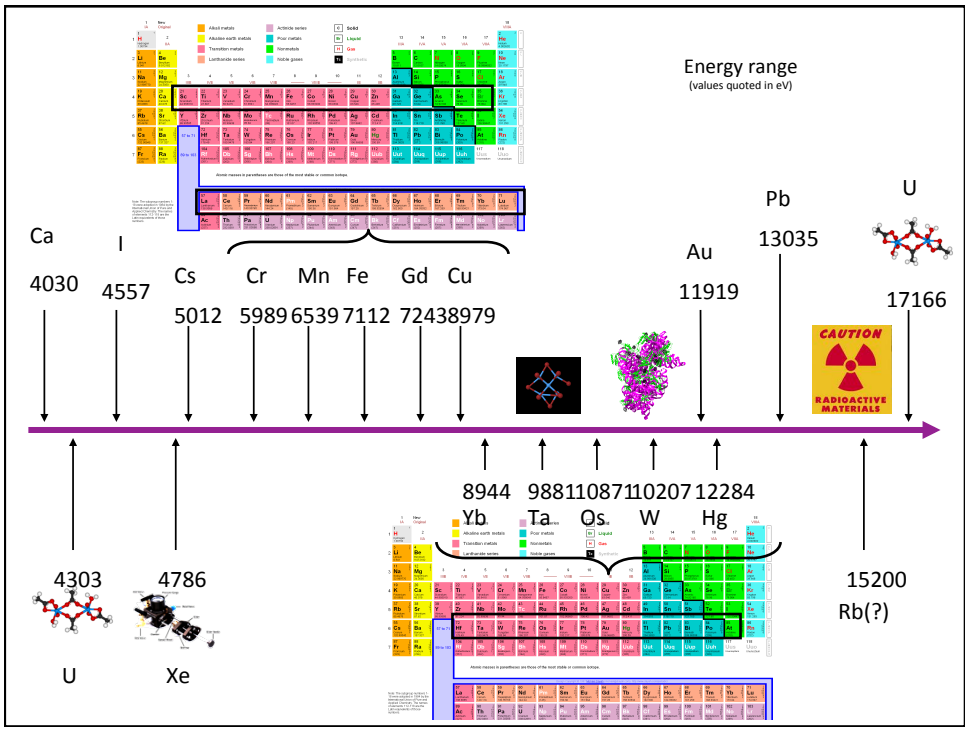
Know your instrument – III: photon flux and beam size



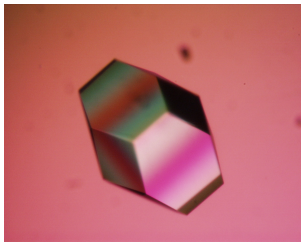
(a) (b)

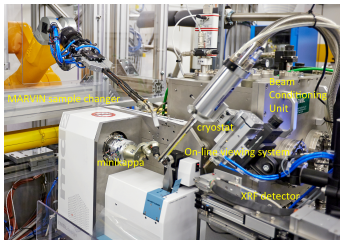
(c) (d) (e)

(data taken shown for P13@EMBL-Hamburg.de)



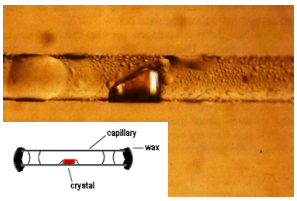
So match the instrument to your crystal!





Use Your knowledge and know Your questions!

Sample mounting options



Cryoconditions: what's that?

LEAD ARTICLE

J. Appl. Cryst. (1997). **30**, 211–237

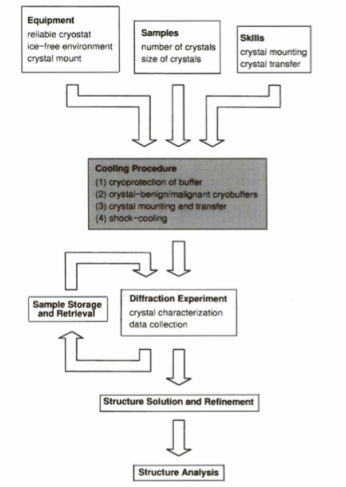
Macromolecular Cryocrystallography

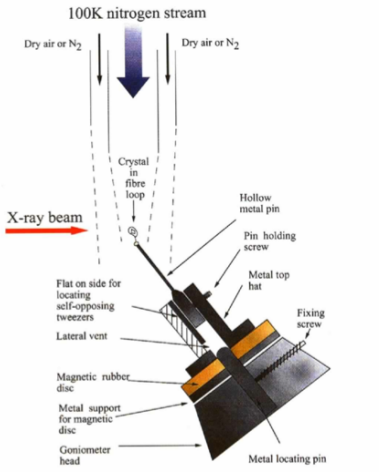
ELSPETH F. GARMAN[a] AND THOMAS R. SCHNEIDER[bc]

[a]*Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 3QU, England,* [b]*European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany, and* [c]*Max-Planck-Institute for Molecular Physiology, Rheinlanddamm 201, 44139 Dortmund, Germany. E-mail: elspeth@biop.ox.ac.uk*

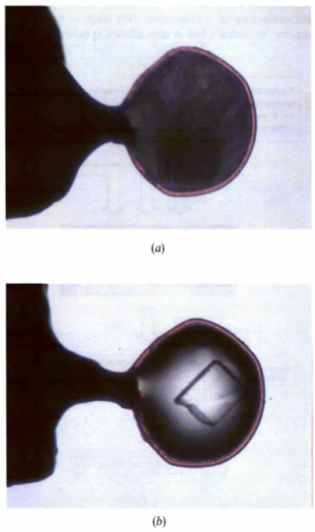
(*Received 3 August 1996; accepted 10 February 1997*)

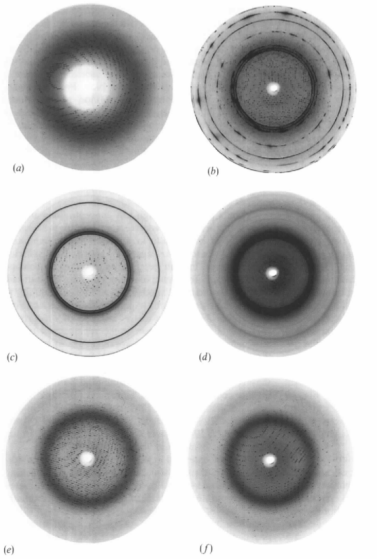
Cryoconditions: continued…



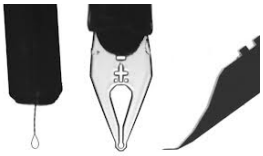


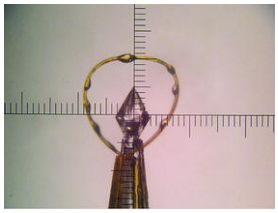
Cryoconditions: results…





Choosing the crystal support





Two basic principles:

- Choose a loop of a similar size to your sample;
- Minimize the amount of buffer around your sample;

Basic results:

- Easier identification of the sample in the loop;
- Easier and better alignment to the beam;
- Reduction of image background;
- Better measurement of intensities;
- General improvement of data quality;

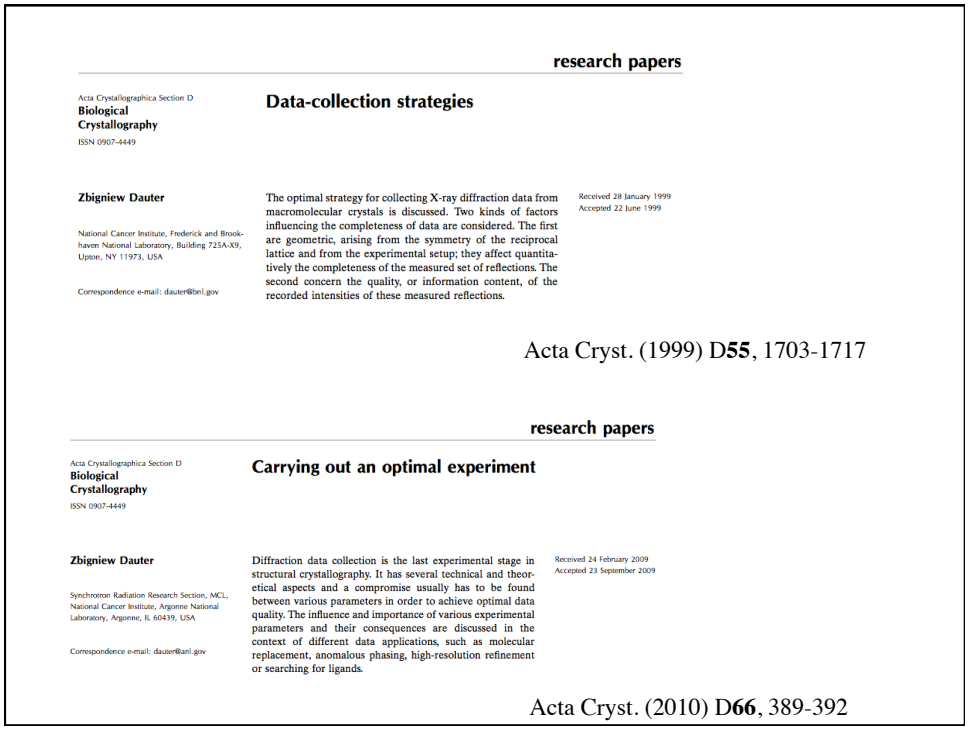
Kitago et al., *Acta Cryst.* (2005). D**61**, 1013-1021

The sample is mounted: now what?

A particular protocol and the relative importance of various quality criteria depend on the intended application of the data and on the type of the experiment performed.

Mostly your experiments will conducted away from your lab.

Think in advance and be prepared.



research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Data-collection strategies

Zbigniew Dauter

National Cancer Institute, Frederick and Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA

Correspondence e-mail: dauter@bnl.gov

The optimal strategy for collecting X-ray diffraction data from macromolecular crystals is discussed. Two kinds of factors influencing the completeness of data are considered. The first are geometric, arising from the symmetry of the reciprocal lattice and from the experimental setup; they affect quantitatively the completeness of the measured set of reflections. The second concern the quality, or information content, of the recorded intensities of these measured reflections.

Received 28 January 1999
Accepted 22 June 1999

Acta Cryst. (1999) D**55**, 1703-1717

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Carrying out an optimal experiment

Zbigniew Dauter

Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA

Correspondence e-mail: dauter@anl.gov

Diffraction data collection is the last experimental stage in structural crystallography. It has several technical and theoretical aspects and a compromise usually has to be found between various parameters in order to achieve optimal data quality. The influence and importance of various experimental parameters and their consequences are discussed in the context of different data applications, such as molecular replacement, anomalous phasing, high-resolution refinement or searching for ligands.

Received 24 February 2009
Accepted 23 September 2009

Acta Cryst. (2010) D**66**, 389-392

Data collection process

- Easy to screw-up in many ways
- Involves lots of technical problems
- But it is science, not technicality
- Pays off to “engage your brain”
- Last truly experimental step later mostly computing (and writing-up) which may be repeated many times
- good quality data make all subsequent steps much easier

Type of experiments

- Molecular replacement
- Anomalous phasing
- High resolution refinement
- Ligand complexes for drug development
- Exhaustive search for diffracting crystals

Dauter et al., Acta Cryst. (2010) D**66**, 389-392

Table 1
Relative importance of various aspects of data collection in different applications.

The priorities of different aspects of data are graded from very high (++++) to not very important (+).

	Molecular replacement	Anomalous phasing	High-resolution refinement	Ligand search
Accuracy	+	++++	++	++
Low-resolution completeness	+++	+++	++	++
Resolution	+	+	+++	++
Overall completeness	++	++	++	++
Automation	++	+	++	+++

Molecular replacement (MR)

- Phasing using an homology model;
- MR is based on comparison of Patterson Maps;
- Strong reflections are very important;
- Calculations done at low resolution (4 – 50 Å) ;

- Requirements:
  - Complete low resolution
  - Accuracy and high resolution, not a priority

Anomalous phasing

- Anomalous signal is small;
- Patterson and direct methods heavy atoms searches done at medium resolution (≈ 3 Å);

- Requirements:
  - Complete low resolution;
  - High accuracy;
  - No radiation damage;
  - high resolution, not a priority

High resolution refinement

- Full capability of crystal diffraction;
- Multiple passes may be needed (geometry, resolution and overload);

- Requirements:
  - Complete high resolution (~70%);
  - Some radiation damage may be tolerated;

Ligand complexes for drug development

- You need to see if the ligand is bound;

- Requirements:
  - Complete data set (~70%);
  - Decent resolution;
  - Data accuracy not so high;

Search for diffracting crystals

- Crystals of large protein complexes are problematic…
- Requirements:
  - patience

A Phasing/refining data set?

- Maybe possible sometimes…
- Requirements:
  - Compromises…
  - Accurate anomalous signal
  - Be happy with less resolution
  - Reduce risk of radiation damage

You can't have your cake and eat it, too..

Ideal data

- Goal of the experiment is: collect - (qualitatively) *complete* and (quantitatively) *accurate* data
- After data processing, the intensities of (ideally) *all* unique reflections of the asymmetric unit of reciprocal space should be known *accurately*

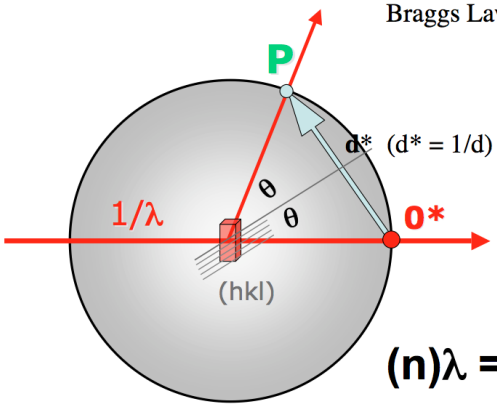
Completeness

Complete data set means that all the reflections I(h,k,l) within the asymmetric unit are measured.

- The Ewald sphere (radiation or crystal orientation at the goniostat);
- Reciprocal lattice (crystal and crystal symmetry);

The Ewald sphere construction

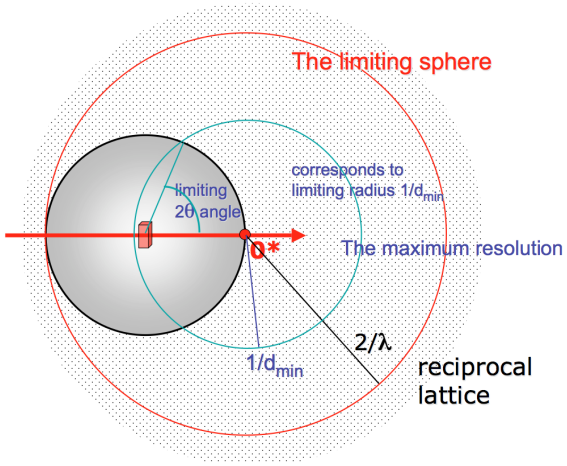
- a graphical representation of Braggs Law



$$(n)\lambda = 2\, d_{hkl} \sin\theta$$

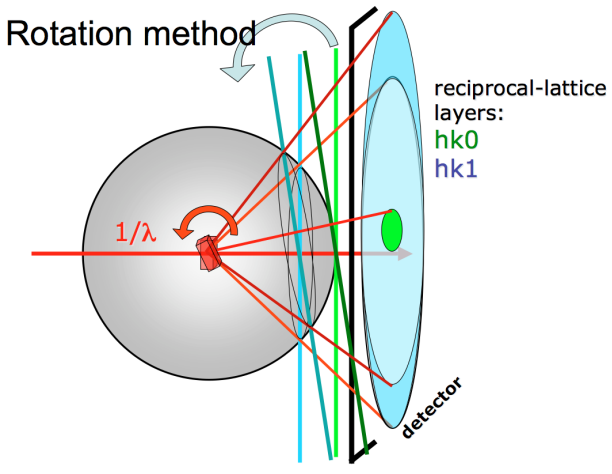
The Ewald sphere represents radiation
Reciprocal lattice represents crystal

The Ewald sphere and reciprocal lattice

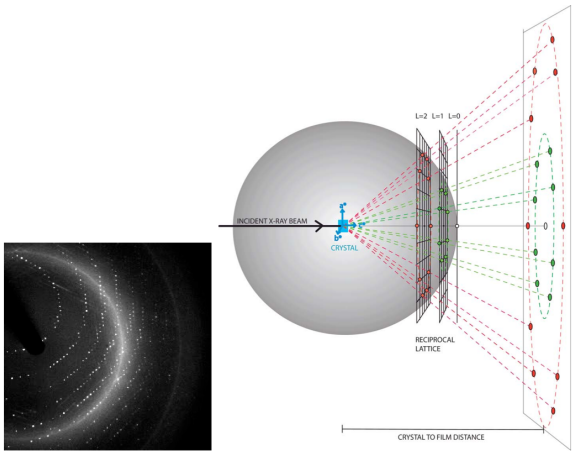


The Ewald sphere represents radiation
Reciprocal lattice represents crystal

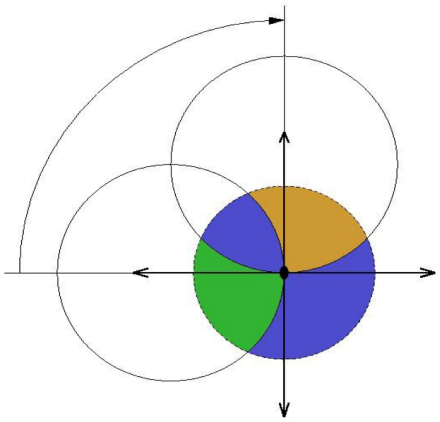
Rotation method



From crystals to diffraction

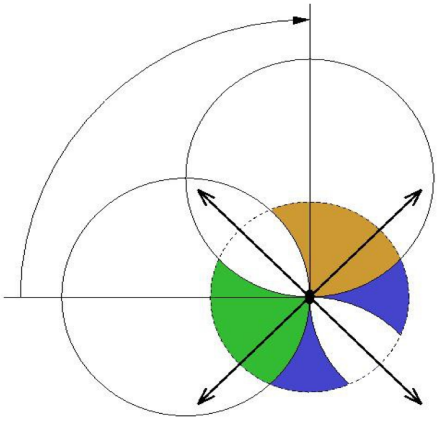


Asymmetric unit in 222 – 90° axis rotation



Zbigniew Dauter, Acta Cryst. (1999) D**55**, 1703-1717

Asymmetric unit in 222 – 90° axis rotation



Zbigniew Dauter, Acta Cryst. (1999) D**55**, 1703-1717

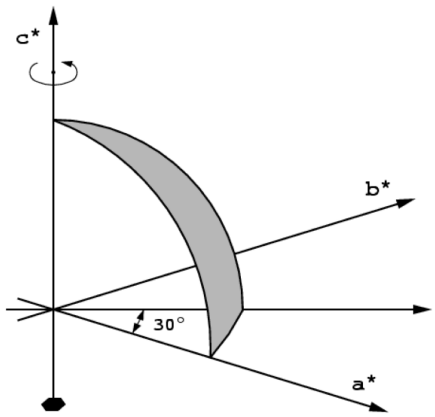
Table 1
Rotation range (°) required to collect a complete data set in different crystal classes.

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

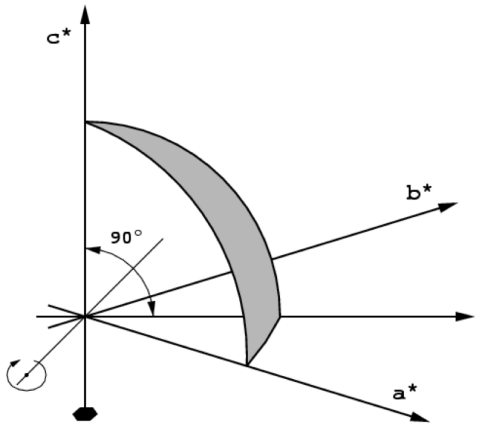
Zbigniew Dauter, Acta Cryst. (1999) D**55**, 1703-1717

Asymmetric unit in 622 – c axis rotation



Zbigniew Dauter, Acta Cryst. (1999) D**55**, 1703-1717

Asymmetric unit in 622 – a/b axis rotation



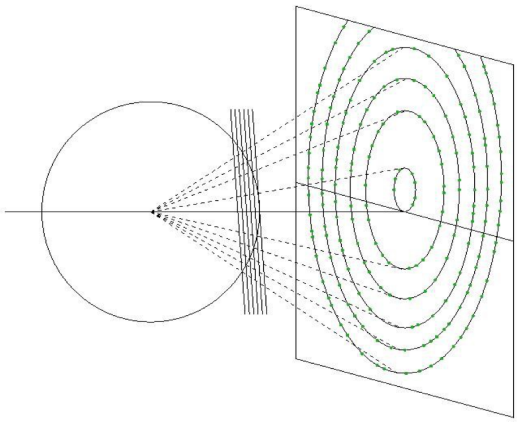
Zbigniew Dauter, Acta Cryst. (1999) D**55**, 1703-1717

„Oscillation“ range

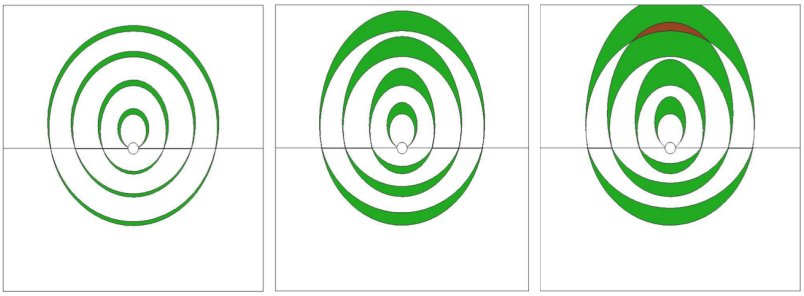
It influences:

- „partials“ versus „fullies“
- Overloads
- Overlap
- Background

Still image



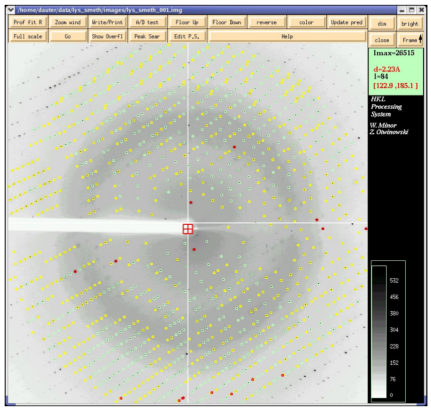
Increasing rotation

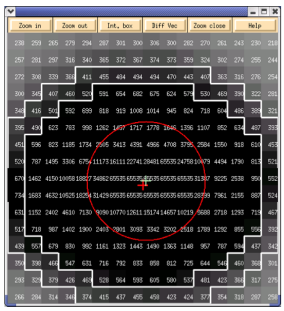


$$\Delta\varphi_{max} = \frac{180 \cdot d}{\pi \cdot a} - \eta$$

d – resolution
a – cell parameter || beam
η - mosaicity

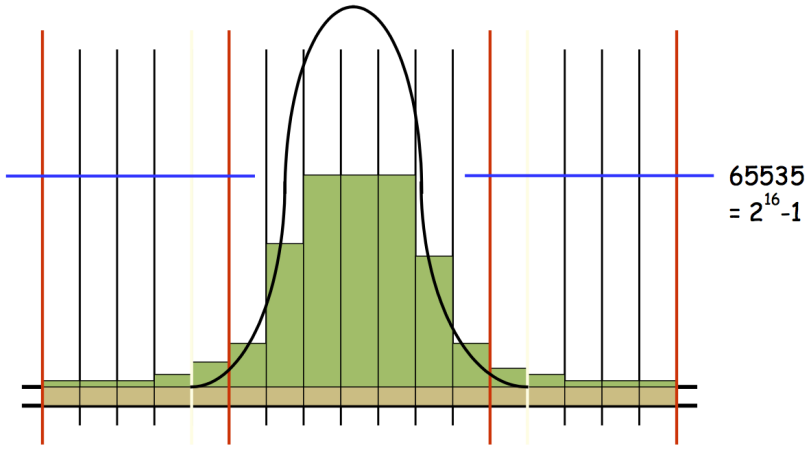
Overloaded profiles





Best, strongest reflections – very important for Fourier maps, Pattersons, direct methods, phasing

Overload extrapolated



standard profile fitted on shoulders and extrapolated above overload value

Multiplicity

- More measurements of equivalent reflections
- lead to more accurate average and σ estimation
- Also scaling and merging is more effective
- But beware of radiation damage

Radiation damage (to our crystals)

PNAS

Predicting the X-ray lifetime of protein crystals

Oliver B. Zeldin[a], Sandor Brockhauser[b], John Bremridge[a], James M. Holton[c], and Elspeth F. Garman[a,1]

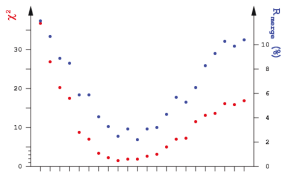
[a]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom; [b]European Molecular Biology Laboratory, Grenoble Outstation, and Unit for Virus Host-Cell Interactions, University of Grenoble Alpes–European Molecular Biology Laboratory–Centre National de la Recherche Scientifique, 38042 Grenoble, France; and [c]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, and Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Edited* by Douglas C. Rees, Howard Hughes Medical Institute, Caltech, Pasadena, CA, and approved November 8, 2013 (received for review August 21, 2013)

Radiation damage is a major cause of failure in macromolecular crystallography experiments. Although it is always best to evenly illuminate the entire volume of a homogeneously diffracting crystal, limitations of the available equipment and imperfections in the sample often require a more sophisticated targeting strategy, involving microbeams smaller than the crystal, and translations of damage remains difficult to predict. Most experienced investigators know that subjecting a protein crystal to a lower dose will give them less radiation damage, but it will also give them less diffraction, and striking the appropriate balance is the key to success. This paper presents a method for optimizing this ratio, allowing the best data to be gained from a given diffracting

Zeldin et al., (2013) PNAS, **110**, 20551–20556

Radiation damage



Typical syndrome of radiation damage – first and last data do not agree with average.

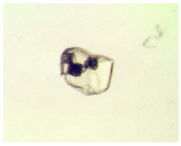
The 'fingerprint' that X-rays can leave on structures:

- atomic B factors increase;
- unit-cell volumes increase;
- protein molecules undergo slight rotations and translations
- disulphide bonds break
- decarboxylation of acidic residues occurs Unit cell variations

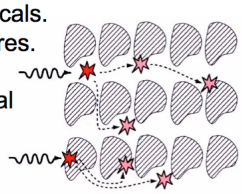
Ravelli McSweeney (2000), Structure, **8**, 315-328.

Primary and secondary damage

Primary damage
- Caused by photoelectrons (photoelectric effect).
- Still occurs at cryogenic temperatures.
- Probably not temperature dependent below 100K.
- No dose rate effect.

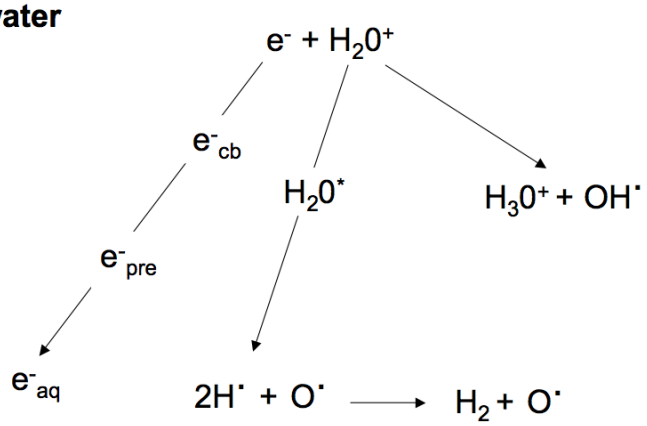


Secondary damage
- Caused by the diffusion of reactive radicals.
- Does not occur at cryogenic temperatures.
- Temperature dependent.
- Higher dose rates result in longer crystal lifetimes.



Processes involved

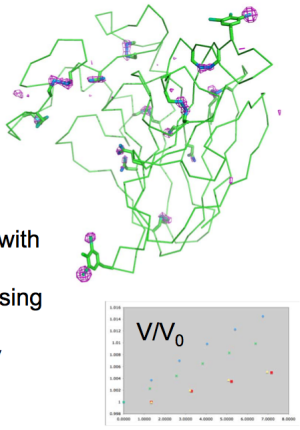
In water



Effects of radiation damage at 100K

Specific
- Breakage of the weakest bonds (e.g. S-S, decarboxylation)
- Loss of anomalous signal
- Not all bonds of the same chemical type show the same susceptibility
- Metals will be reduced!!

Non-specific
- Increase in unit cell dimensions with increasing dose
- Increase of Wilson B with increasing dose
- Non-isomorphism => loss of any dispersive signal => SAD in preference to MAD



Is there a cure?

At 100K - not really:
- The global damage depends quantitatively on absorbed dose.
- Radicals are not mobile - electron scavengers might work, but they have to be soaked into the crystal prior to flash cooling. Where some effect has been reported (ascorbic acid, DNTB, nicotinic acid) it is crystal dependent. Specific damage might be altered.
- Avoid metal containing scavengers - this will just increase the dose.
- Zero-dose extrapolations will only work if the damage is not excessive and the multiplicity is high.

Is there a cure? continued

At 100K avoidance is better:

- Know the photon flux and use RADDOSE to calculate the dose.
- Your experiment should not involve a dose of more than 20 MGy (The Henderson Limit). ~40 MGy results in diffracted intensity dropping by one half.
- If you want to measure an anomalous signal try to keep the dose below 5 MGy.

RADDOSE

computer programs

Journal of
Applied Crystallography
ISSN 0021-8898

Received 22 January 2013
Accepted 26 April 2013

RADDOSE-3D: time- and space-resolved modelling of dose in macromolecular crystallography

Oliver B. Zeldin, Markus Gerstel and Elspeth F. Garman*

Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. Correspondence e-mail: elspeth.garman@bioch.ox.ac.uk

RADDOSE-3D allows the macroscopic modelling of an X-ray diffraction experiment for the purpose of better predicting radiation-damage progression. The distribution of dose within the crystal volume is calculated for a number of iterations in small angular steps across one or more data collection wedges, providing a time-resolved picture of the dose state of the crystal. The code is highly modular so that future contributions from the community can be easily integrated into it, in particular to incorporate online methods for determining the shape of macromolecular crystals and better protocols for imaging real experimental X-ray beam profiles.

Zeldin et al. J. Appl. Cryst. (2013). 46, 1225–1230

raddose_6kev.inp

```
~/RADDOSE/raddose <<EOF          < Unix command line
ENERGY 6                         < X-ray beam energy used
CELL 78 78 78 90.0 90.0 90.0     < unit cell parameters (I2_3 insulin)
NRES 51                          < number a.a. resides/molecules
NMON 1                           < number of molecule
PATM S 6                         < type/number heavy atoms
BEAM  0.035 0.025                < X-ray beam size
CRYST 0.05 0.1 0.1               < crystal size
PHOSEC 4.8E+10                   < X-ray photon flux @given energy
EXPO 1                           < exposure time
IMAGE 1                          < number of images
END
EOF
```

raddose_6kev.inp / results

Total absorbed dose (Gy)	.110E+06
Absorbed dose per image (Gy)	.110E+06

DOSE LIMITS:

** Time in sec to reach Henderson limit calculated from electron diffraction (20 MGy)	181
** Time in sec to reach experimental dose limit (30 MGy)	271

So if you want to collect a full revolution ie. 360°, the max allowed exposure time would be 0.5 sec/degree.

Exposure time: "dose slicing"

Again compromises…

- short exposure time (or strong attenuation) reduces radiation damage and avoids overloads,
- long exposure time (or little attenuation) improves the signal-to-noise
- small oscillation better samples the reflection profiles and reduces background,
- large oscillation saves readout time and minimizes the damage from shutter flicker
- short crystal-detector distance ensures that even the highest resolution is recorded,
- long crystal-detector distance avoids reflection overlap and increases signal-to-noise

Overall picture

- Aim of the data collection
- Completeness of the data collection
- Accuracy of the data collection
- Radiation damage

Parameter selection

So each data collection will be result of several parameters:

- Wavelength choosen;
- Total rotation range;
- Photon flux (or total radiation dose);
- Image width;
- Exposure time;
- Detector distance;
- Etc.

How to find „best“ compromise?

- some parameters are ill-defined because they involve non-proven concepts
- some parameters are qualitative only: what is „high enough completeness“, „too much radiation damage“?
- choice based on past experience of „experts“
- build up your own experience by trial and error
- based on strategy programs: BEST

Popov Bourenkov (2003) Acta Cryst D**59**, 1145-1153
Popov Bourenkov (2006) Acta Cryst D**62**, 58-64

Features of Pilatus detector

- pixel-array detector (PAD): each pixel is a detector with electronics
- counts (instead of accumulates) each photon as it hits the detector
- Point spread function: one photon affects only one pixel (if the photon hits the detector at right angle)
- noise-free readout, no intrinsic background
- can count up to 20 bits (>1.000.000)
- fast readout (ms)

This changed the rules

- lack of intrinsic and read-out noise improves signal-to-noise ratio
- very low counts (0,1,2,...) are possible: low exposure allows to avoid overloads
- ideal for fine slicing: less background
- enables shutterless (i.e. continuous) data collection: no shutter jitter
- for the same signal-to-noise, one can expose less: this means less radiation damage, higher multiplicity
- multiple passes not required

... and give us freedom!

Examples:

- to adapt the oscillation range to the mosaicity as shown by Müller et al (2012) „Optimal fine ϕ-slicing for single-photon-counting pixel detectors“, Acta D68, 42
- to slice the tolerable dose into many low-dose frames such that we obtain more meaningful partially complete datasets from microcrystals or at RT

... and simplifies planning:

	Conventional way	*Pilatus way*
Statistics	R_{merge}	$CC_{1/2}$
Exposure	Reflections visible; tolerate some overloads	Low-dose, high multiplicity
	Expose such that reflections can be seen visually	Expose weakly and rather increase multiplicity
Oscillation range	0.25-1°	0.05-0.2° CCD:
Rotation range	strategy, xplan	180° / native 360°/ anomalous
High resolution	Multiple passes	Single pass

FAQ questions

- How much completeness is enough?

For high-quality data obtained with synchrotron radiation, completeness > 93% and observable data > 70% should be achievable for the highest resolution shell.

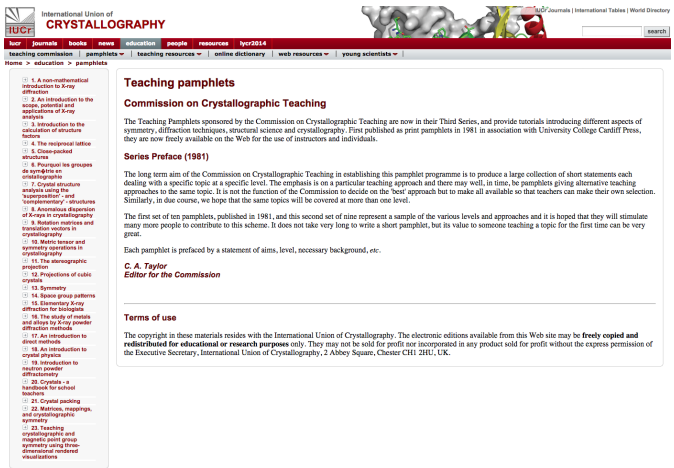
(Notes for authors 2012. *Acta Crystallographica Sec. D68,* 194-199).

- How much radiation damage can be tolerated?
- How good do the data have to be, to be able to solve a structure?

Suggestions

1. always use a test crystal
2. process data on site and adjust parameters for the next crystal, based on the results. But it's important to look at the right indicators!

Thank you for your attention



Please visit: http://www.iucr.org/education/pamphlets