Data Analysis and Sample delivery

John C H Spence*

ASU Physics/LBNL

Printed bibliography provided.

See Rep Prog Phys, Spence, Weierstall, Chapman. 2012 for review

Projector color

The three most common questions regarding BioXFEL

1. Why use an XFEL instead of a synchrotron ?
   (*XFEL proposals for work which could be done on a synchrotron are rejected* !)

* Avoid damage (not electronic, assuming MR, 2/3 info comes from phases, 2Ang)
* RT, native environment. Use short pulses instead of freezing to avoid damage.
* Time resolution (ps), study dynamics because not frozen.
* Try new ideas. (2 color SAD, pump-probe, FSS, SP, SFX, Shape transform phasing)
* Xtals large enough for MX not available
* Rapid diffusion into nanoxtals, irreversible processes

2. How to make nanoxtals ? See bibliogaphy at end

3. How does data analysis differ for SFX ?

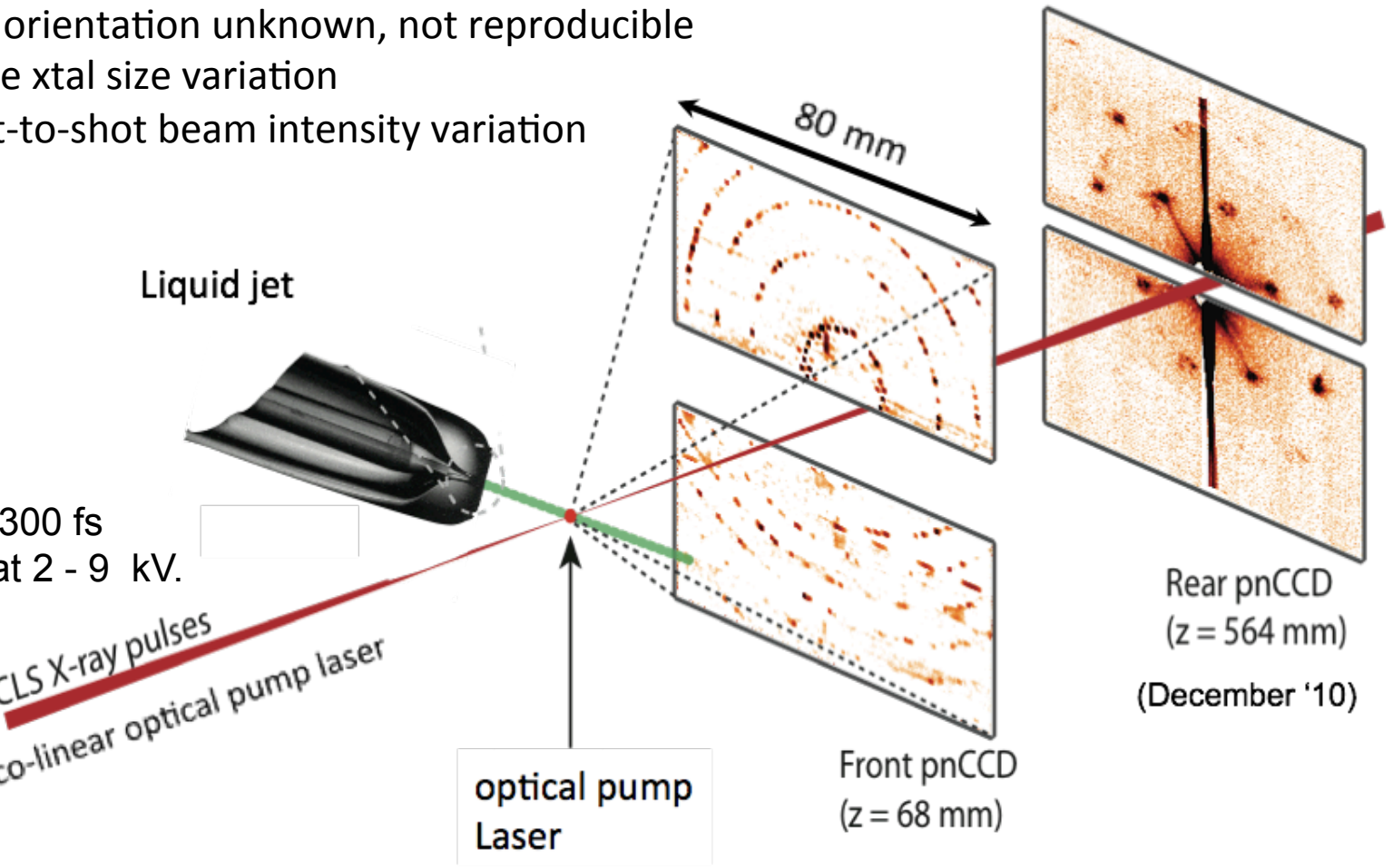
Experimental geometry

XFELS are terrible sources for MX.

Major sources of error

1. Xtal orientation unknown, not reproducible
2. Large xtal size variation
3. Shot-to-shot beam intensity variation

F

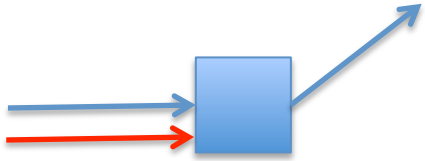


Barty et al

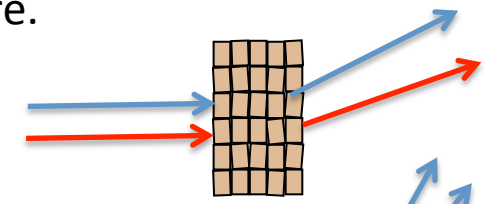
Debye period is ~100fs

The four cases of SFX diffraction.

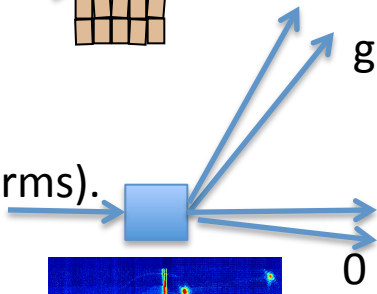
1. Big perfect crystal, polychromatic beam. Sample monochromates.
   Plane-wave to plane-wave. Picks out one wavelength)
   *J.Holton & Frankel Acta D66, 393 (2010)



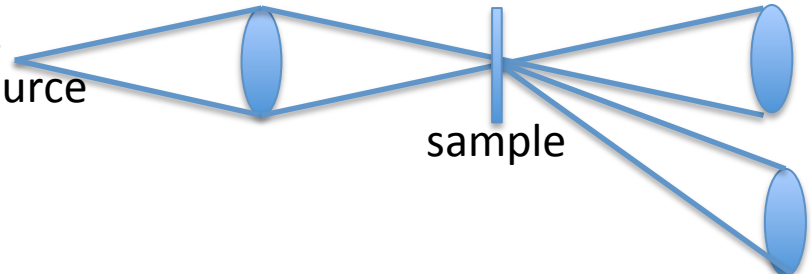
2. **Big mosaic crystals*** (>1 micron) . Every wavelength goes somewhere.
   *J.Holton & Frankel Acta D66, 393 (2010);
   *J.Hattne….Sauter Nature Meth. (2014)
   *T.White Phil Trans Roy Soc. B 369 (2014)
   *E. Snell Meth Enzymol (2003). *What is mosaicity* ?



3. Nanocrystal (< micron). Monochromatic beam give big spots (shape transforms).
   *Kirian et al Optics Express 18, 5713 (2010); Acta A 67, 131 (2011).
   *Spence et al Optics Express 19, 2866 (2011); Kirian et al (2014) in press
   (for phasing from shape transforms).

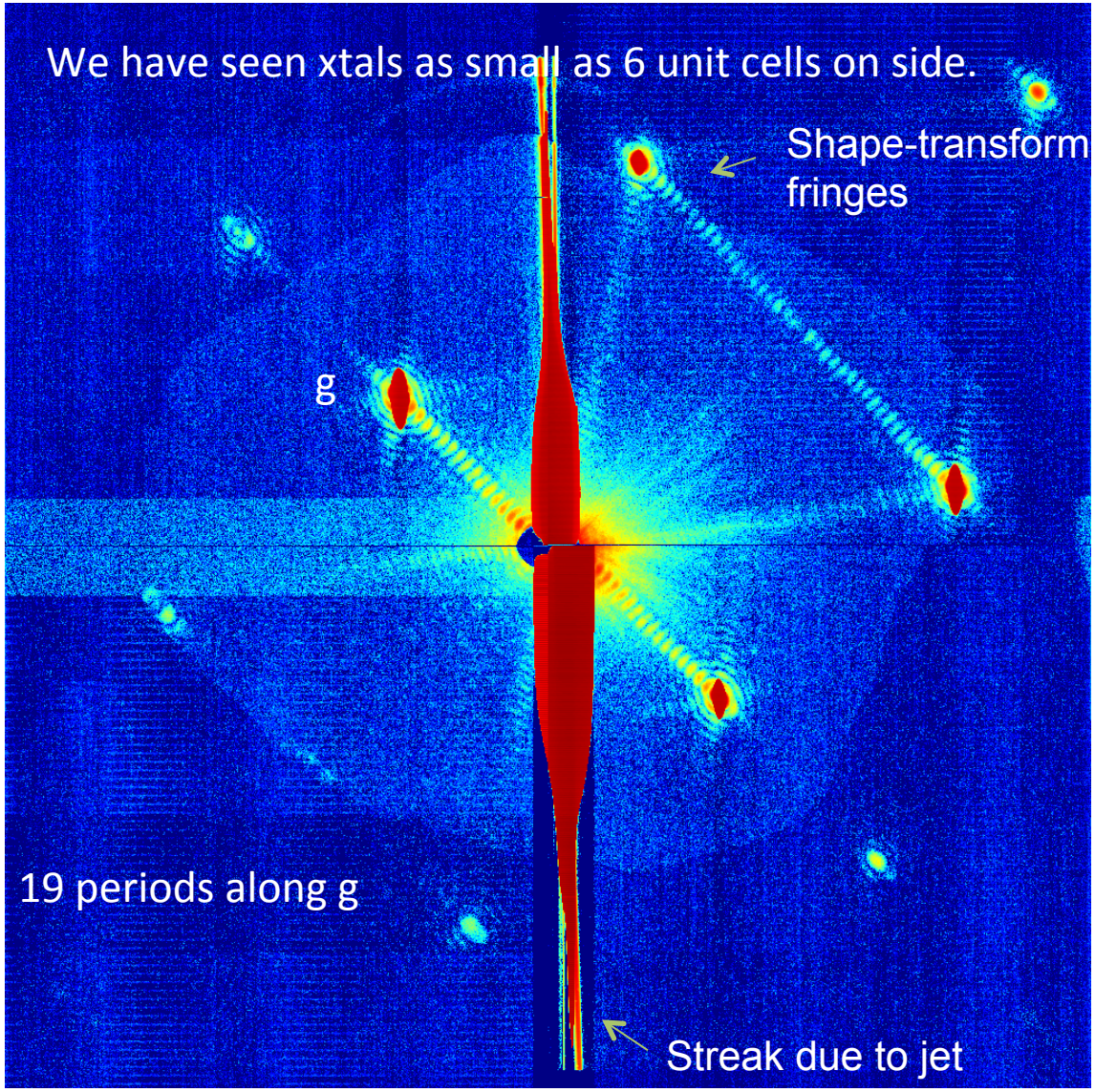


4. Beam smaller (eg 0.1 micron) than mosaic block. Same as 1, but big beam divergence (CCBD).
   *J.Spence et al Phil Trans *B* **369**, 20130325 (2014).
   Nanodiffraction. Small coherent focussed beam.



*The case treated in this course.

The smallest crystals show "shape transforms" .



Red vertical streak is diffraction from column of liquid in jet.

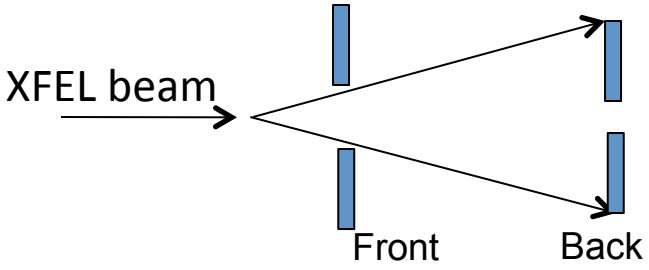
Subsidiary maxima give number of unit cells.

Each nanoxtal is smaller than one mosaic block

Beam size is about 1 micron (or less)

How to merge millions of patterns ?

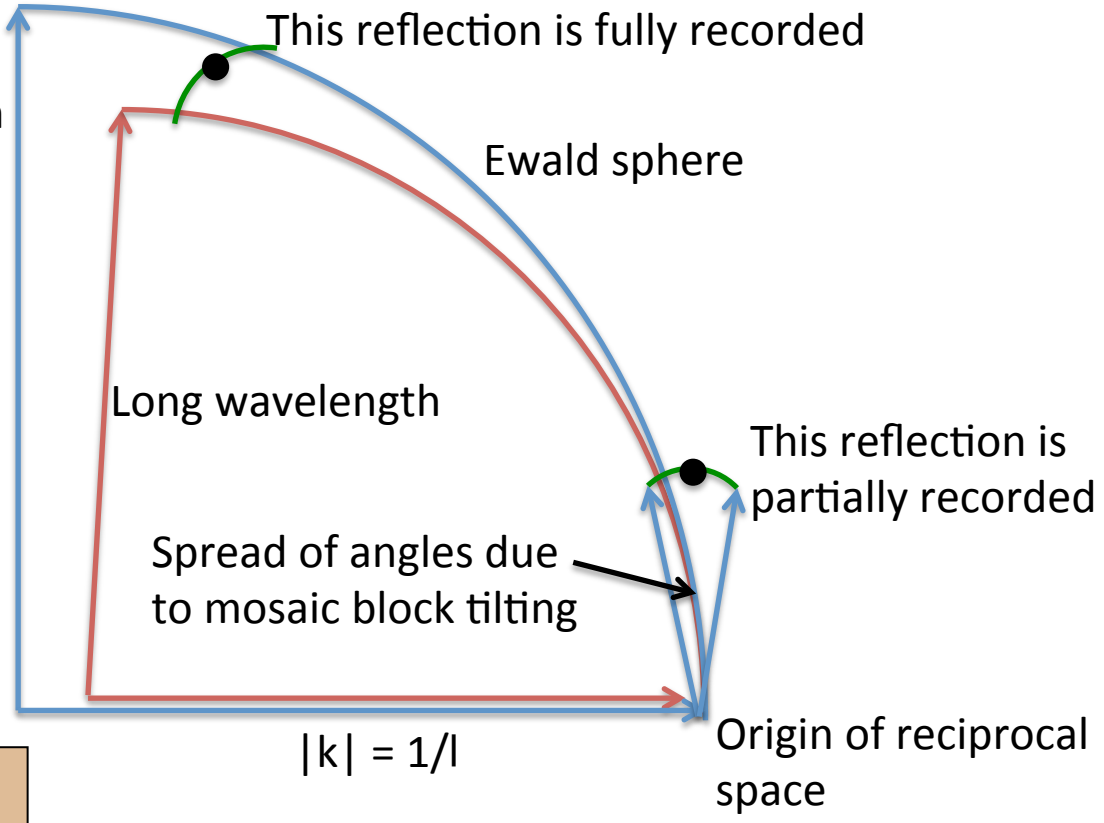
Every nanoxtal is in a different (random) orientation.



Photosystem I nanocrystals at 2 kV (6.9 Ang wavelength).
Single Shot (10^{12} photons incident).

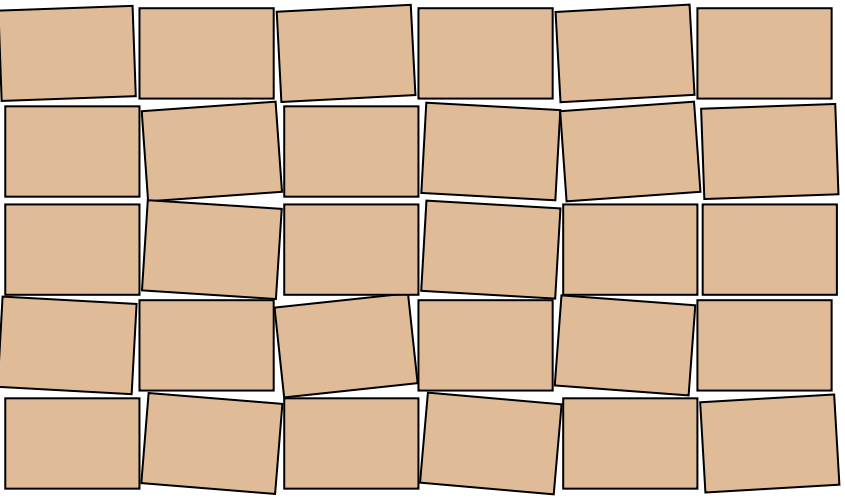
Chapman, Fromme….Spence 2011

The wavelength spread in XFEL beam spans wider spots at higher angle.



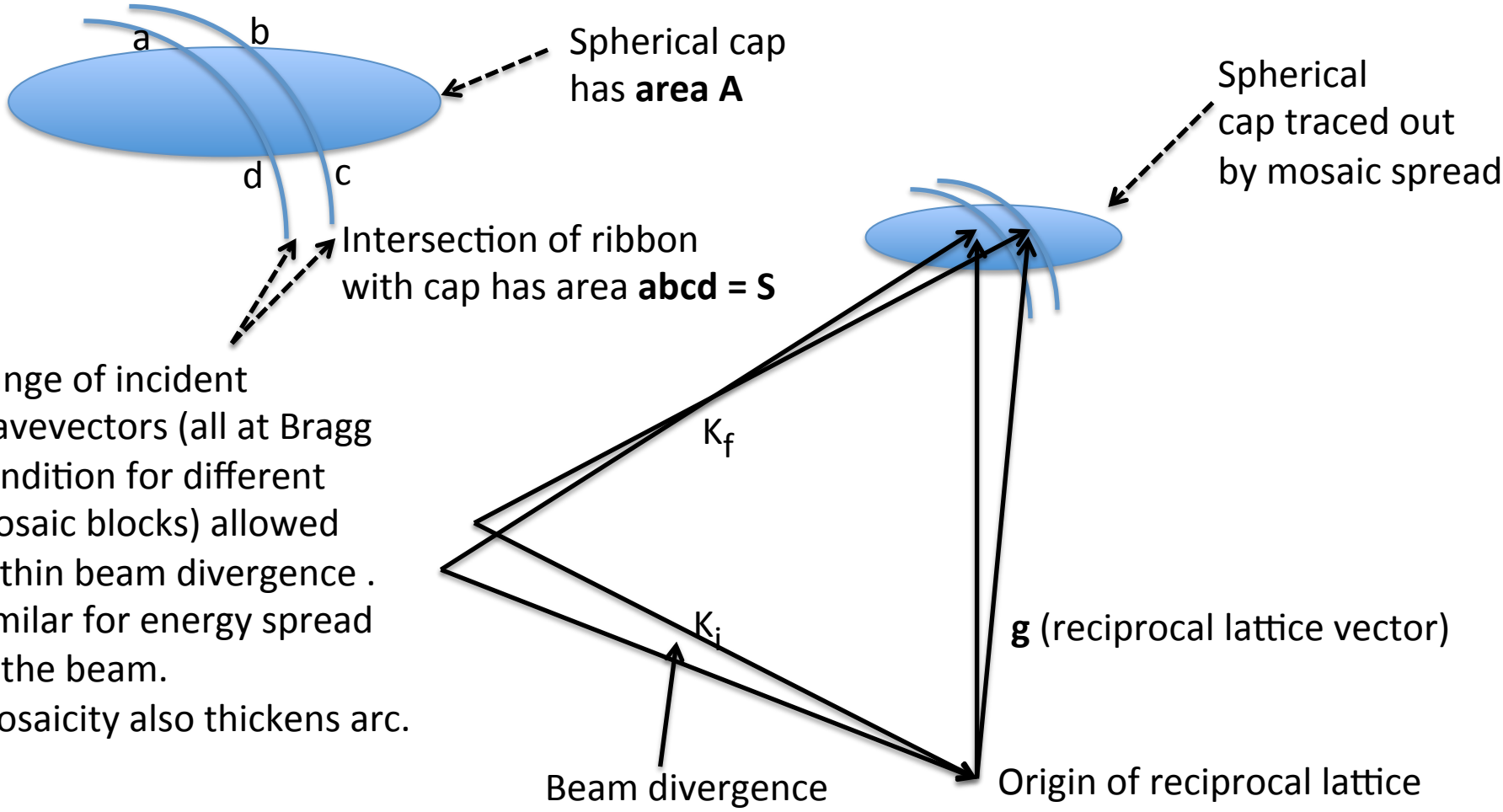
Terry Sabine – theory of extinction

Each block acts as a monochromator for a different wavelength in the beam



Conclude: different fractions of different reflections are uncovered by the range of wavelengths.
(some reflection arcs are just clipped, others fully spanned)
Hence "partial reflections" (also beam divergence).

Partiality and Mosaic Spread.

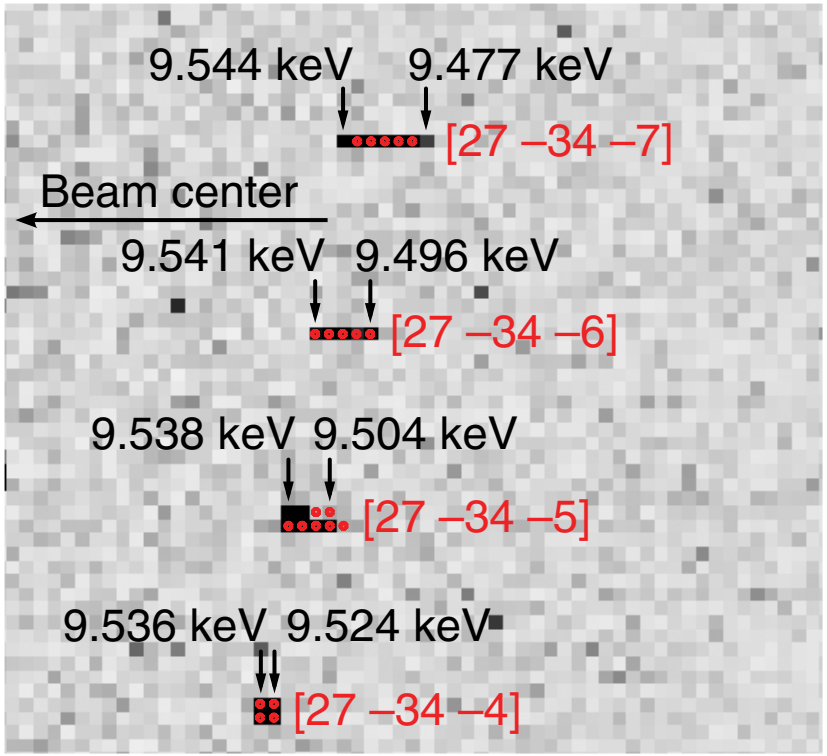


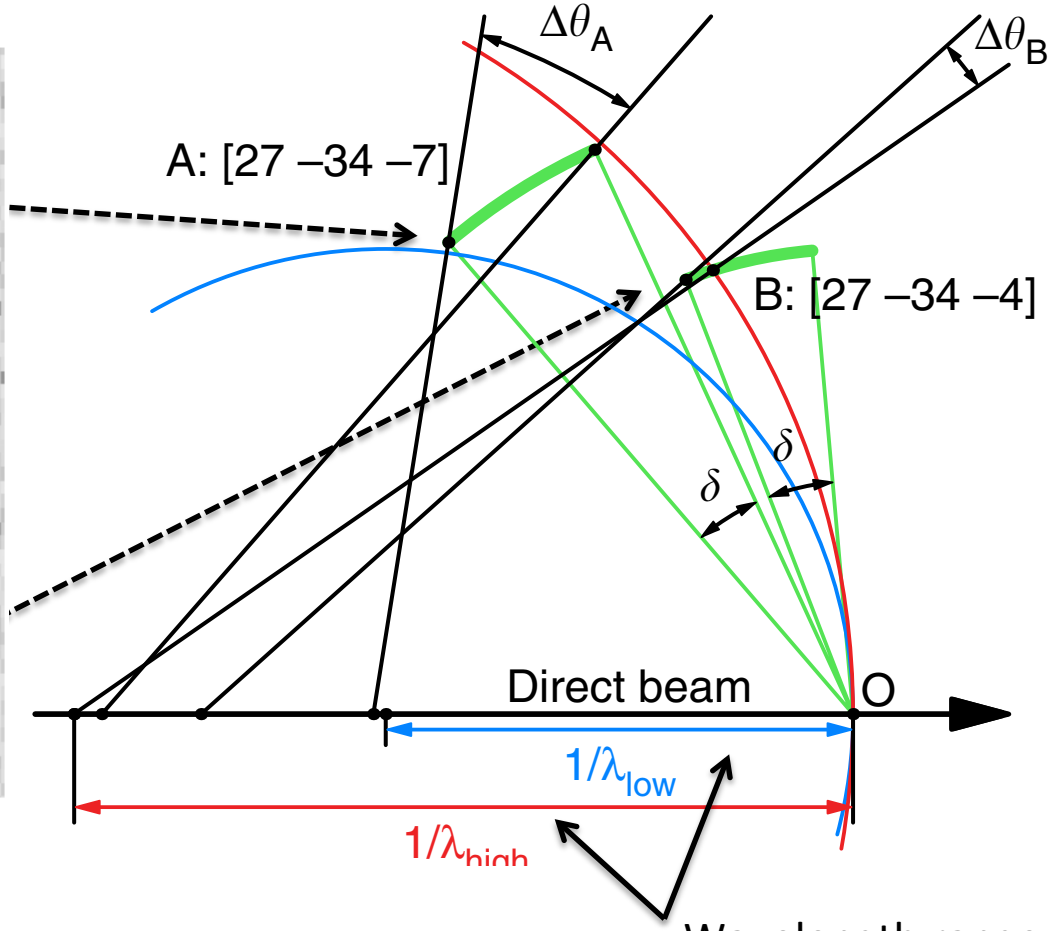
eg Rossman and Van Beek. Acta D55 1631 (1999) for review of post-refinement, partials.

Partiality is ratio of areas P = abcd/A = S/A . (volumes in a full theory).

The wavelength spread in XFEL beam spans wider spots at higher angle.

so different fractions of different reflections are uncovered by the range of wavelengths.
(some reflection blobs are just clipped, others spanned fully)





Each tilted mosaic block acts as a monochromator for a different component wavelength in the beam. Energy spread in the beam also thickens the concentric arcs due to mosaicity.

From Hattne/Sauter et al Nature Methods 2014

Cryo-em images of mosaic blocks in Lysozyme.

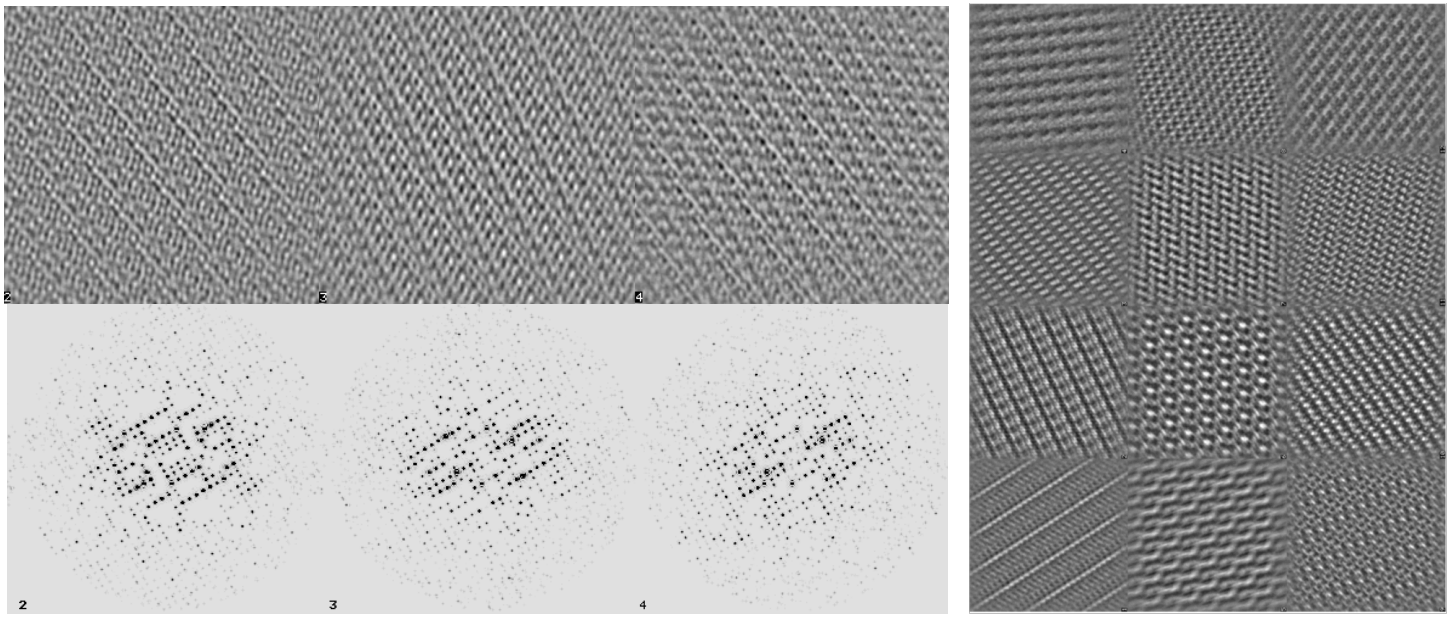


Figure 2 Left panel: *Three projection class averages from a single image of a lysozyme crystal with three mosaic domains with subtly different orientations (top). The images are 230x230 Å^2. The bottom panel shows the square roots of the amplitudes of their Fourier transforms after subtracting the radially averaged background.* ***Right panel:*** *A selection of some high-resolution electron images of 3D crystals of the protein lysozyme (1/4 magnification of the left hand panel).*

Mosaic blocks imaged by TEM in Lysozyme. Unit cell dimensions give scale.

From…

Nederlof et al (2013) *Acta Crystallogr.* **D69**, 852-859

See Subramanian et al Ultramic (2014) for multiple scattering limits on thickness for electron microscope diffraction from protein nanoxtals.

Scaling and partiality for SFX.

See White, Phil Trans B369, 20130330 (2014)
Hattne, Nature Methods 11, 545 (2014). Kabsch

1. Bragg spot intensity depends on xtal size (s), incident beam intensity (I), structure factor and partiality (p). (Ignore Lorentz factor ?, polarization…).

2. Simplest analysis. Sum intensity of same reflection from every "nanoxtal" to get average. (Monte Carlo). Averages over variations in xtal size, orientation and beam.

3. Better method.

i) Use sum of all Bragg beam intensities per shot to monitor xtal size, giving a xtal size and incident beam scale factor (s x I).

ii) Use detector metrology to get angular deviation of a given spot from exact reference Bragg condition (can't). Assume width of mosiac spread (spherical cap). Hence calculate geometric partiality p for this reflection in each shot. (OR: average one Bragg spot intensity over all shots , then go back and compare each value with the average to get "partialities" p due to misorientation. Iterate ?). EMC ?

iii) Form sum of shots for given reflection weighted by s, I and p in i) and ii).

4. Even better method.

Model mosaicity spot-broadening function. *Each spot in one shot samples it at a different place* ! Use neighboring spots (Hattne). Scale using Wilson plot ?

5. If Shape Transforms : Dividing out. (Too slow ?. Kirian Acta A 2011) EM?

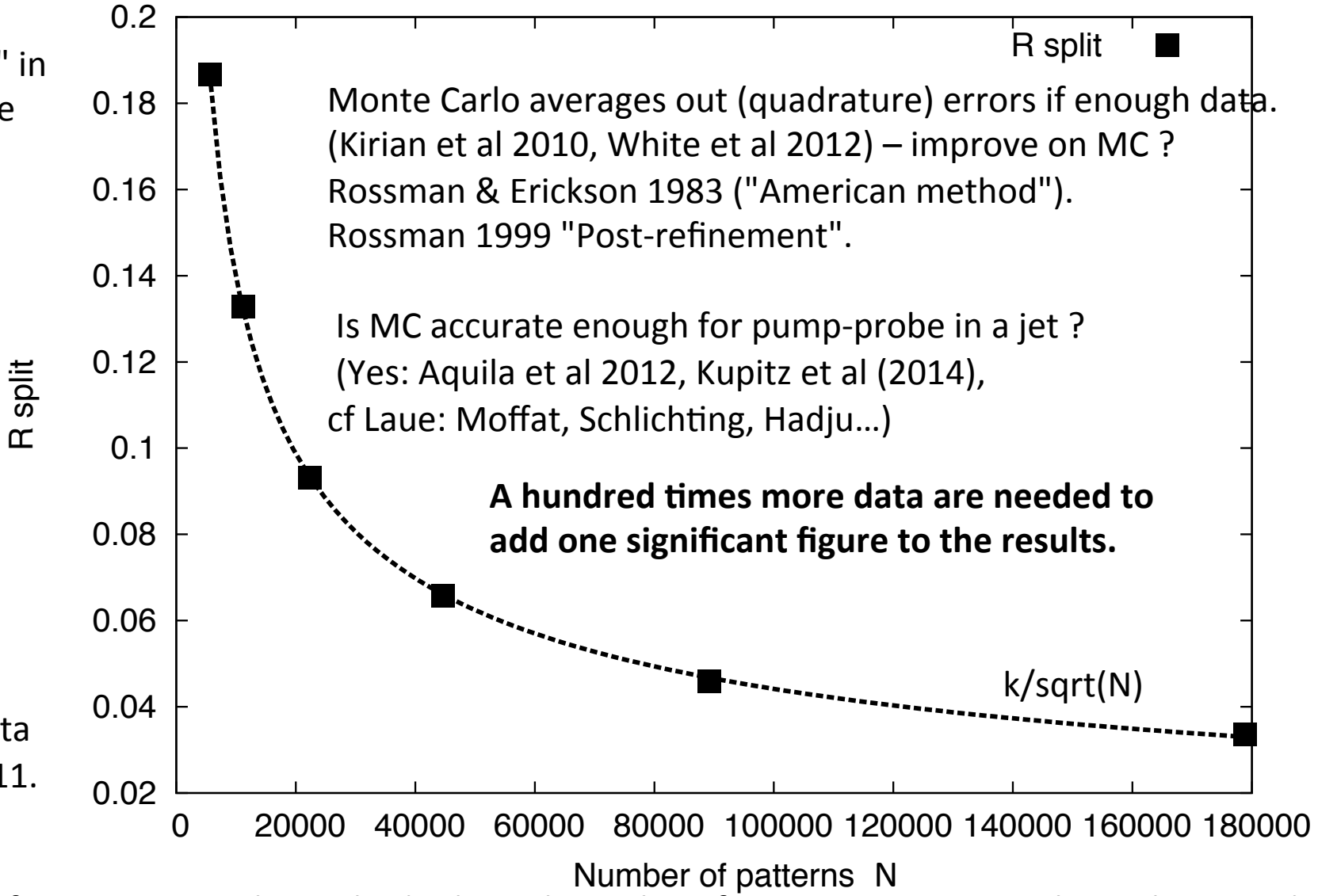
Issues: Beam divergence, energy spread in beam, pixel size…Plot I vs number of spots. (H. Ginn)
Beam smaller than one block ? Shape transforms ?. Rossman's "American method"

The Monte Carlo method for TR-SFX reduces error as 1/sqrt(N)

Evidence that S/N improves as 1/sqrt(N) for N shots from Cathepsin data. (K.Nass Feb 2013)

XFELS cause big errors in SFX (Shot-to-shot 15%, xtal orientation, size)

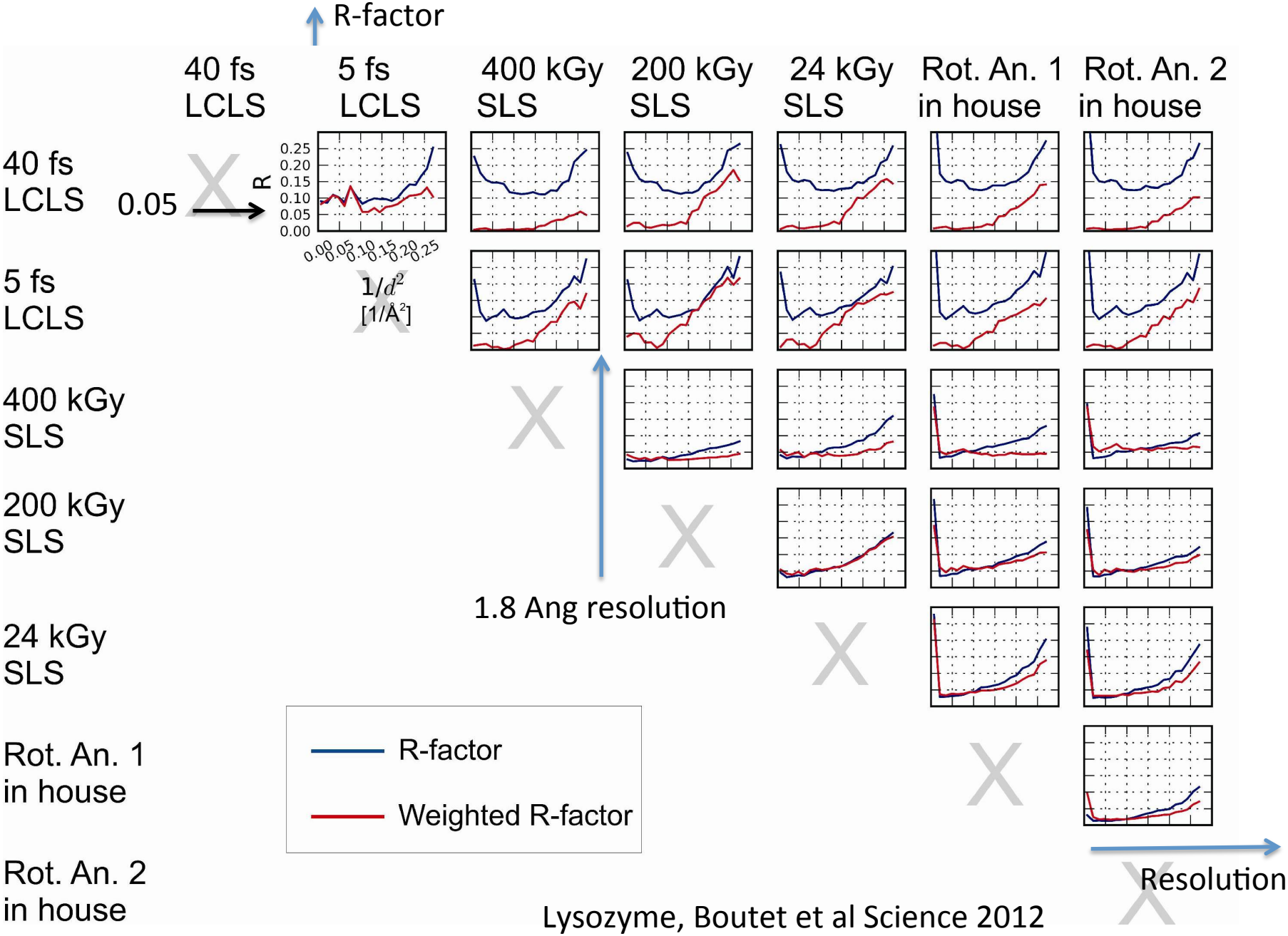
~ "Error" in structure factors



See also Kirian Acta Cryst 2011.

Validity of our Monte Carlo method is best shown by R factor comparison with synchrotron data

R vs Resolution for various doses, LCLS vs Synchroton, Rotating Anode



Lysozyme, Boutet et al Science 2012

The time for SFX data collection and analysis has been reduced greatly between 2011 and 2014.

Protein	2011 Cathepsin	2014 Phycocyanin
Sample injection	GVDN	LCP
Protein size	37 kDa protease	209 kDa hexameric antenna complex
Year of experiment	2011	2014
Data collection took	**5 days**	**2.5 hours**
Crystal hits	293,195	18,794
Indexed hits	**178,875**	**6,629**
Time data took to analyze	**1 year**	**3 months**
Resolution	2.1 Å	1.95 Å

The improvement is due to many incremental advances and high quality nanocrystals.

Phycocyanin: SFX results of crystals delivered in the liquid jet and embedded in LCP were in good agreement, but showed significant difference from results on similar samples recorded from cryo-cooled samples at a synchrotron.

Phycocyanin in LCP:
Only 6 µL of crystal suspension of globular protein transferred into LCP

Ginn, Garman et al . **used 6000 patterns collected in 30 mins to get 1.74 Ang.** map of Polyderin CPV17 (Aug 2014). Adjust l for max number of spots, optimize orient matrix

This greatly advances our goal of increased capacity for SFX work and XFEL availability for biology

A CXIDB Data Base for XFEL data has been established at http://cxidb.org/index.html

Including these SFX data sets…
CXIDB #15 (Lysozyme Boutet)
CXIDB #21 (Serotonin/ergotamine)
CXIDB #22 (Gd-lys SAD from Heidelberg)
CXIDB #23 (thermolysin from Sauter et al)

Indexing ambiguity (symmetry of lattice higher than structure):
See Brehm and Diederichs Acta D70, p101 (2014)

Oct 2013: There are 15 data sets, both nanoxtal and single-particle..

F. Maia Nature Methods 9, 854 (2012)

Bibliography

A bibliography of some papers on structural biology using X-ray lasers (XFELS). Aug. 2014.

1. Review articles, early papers.
i) "Femtosecond X-ray protein nanocrystallography". Chapman et al. (2011) Nature **470**, p. 77.
ii) "Mimivirus particles imaged on-the-fly with an X-ray laser". M. Seibert et al. Nature **470** p.78 2011.
iiii) "X-ray lasers for structural and dynamic biology". J.C.H.Spence, U. Weierstall and H.N.Chapman. Rep. Prog. Phys. **75**, 2601 (2012). A review.
iv) Twenty papers on this topic in the volume begining : "The birth of a new field". J.C.H.Spence and H.N.Chapman. *Phil. Trans. R. Soc.* ***B*** **369**, p. 20130309 (2014)

2. Data analysis for serial femtosecond crystallography (SFX) and single particles (SP)
i) "Femtosecond protein nanocrystallography - data analysis methods". R. Kirian et al Optics Express 18, p.5713 (2010)., also Acta Cryst A 67, p. 131-140 (2011).
ii) "CrystFEL: Software for serial crystallography. T. White et al. J. Appl Cryst. 45, 335 (2012)
iii) "Accurate structures from XFEL data " J. Hattne et al . Nature Methods, 11, 545 (2014)
iv) "Processing of X-ray snapshots from crystals in random orientations". W. Kabsch. Acta Cryst. D70, 2204 (2014). Also Barty et al Acta 47, 1118 (2014) on primary data processing, hit finding.
v) "Hawk: the image reconstruction package for coherent X-ray diffractive imaging" F. Miai et al. J. Appl Cryst. 43, p.1535 (2010). See also Yoon et al Optics Express 19, p.16542 (2011).
vi) Resolving the indexing ambiguity in SFX. Brehm and Diederichs Acta D70, p101 (2014)
vii) T. White, Phil Trans B369, 20130330 (2014)

3. Sample delivery
"Injector for scattering measurements on fully solvated species" U. Weierstall, J.C.H.Spence and R.B.Doak. Rev. Sci Instr. **83**, 035108 (2012). Electrospinning: R. Sierra et al Acta **D68** (2012)
"Liquid Mixing Jet for XFEL Study of Kinetics". D. Wang et al. J. Synch. Rad. (2014).
"Lipidic cubic phase injector for SRX". U.Weierstall, Nature Communications **5**, p. 3309 (2014)

4. Time-resolved XFEL crystallography (TR-SFX) and solution scattering (TR-FSS)
"Time-resolved nanocrystallography using XFEL". Aquila et al. Optics Express **20**, 2706 (2012)
"Serial Time-resolved crystallography of Photosystem II using a femtosecond X-ray laser**".** C. Kupitz et al. Nature. July (2014). doi:10.1038/nature13453.
"A protein quake in photosynthesis". D. Arnlund et al. Nature Methods August. 2014.
"Three-dimensional single-particle imaging using angular correlations from X-ray laser data". H. Liu et al. Acta Cryst A69. P.365-373 2013. *and many other single-particle papers referenced therein.*
"Single-particle structure determination by correlations of snapshot X-ray diffraction patterns". D. Starodub et al. Nature Comms. 3, 1276 (2012). See also Loh et al PRL 104, 225501 (2010).
"New light on disordered ensembles: Ab-initio structure determination of one particle from scattering fluctuations of many copies". D. Saldin et al Phys Rev Letts. **106,** 115501. (2011).

4. Structural biology.
SFX of G Protein-coupled receptors in lipidic cubic phase. W. Liu et al. Science **342** 1521 - 1524. (2013).
"Protein Structure Determination by SFX". S. Boutet et al . Science (2012). Vol **337** 362
"Cathepsin B structure solved to 2.1 Å using an X-ray laser L. Redecke et al. Science **339**, 227 (2013).

5. Phasing.
"De novo protein crystal structure determination from XFEL data" Barends et al Nature. **244**,505 (2014)
"Phasing SFX from size-varying nanocrystals". J. Spence et al. Optics Express. **19**, 2866-2873 (2011)
"MAD at high X-ray intensity". S. Son et al Phys Rev Letts **107**, 218102 (2011).

6. Radiation damage with XFEL.
"Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements". A Barty et al Nature Photonics, **6,** 35 (2012), References therein and papers by C. Caleman and S. Hau-Riege et al

7. Making nanocrystals.
"Microcrystallization techniques for SFX". Kupitz et al *Phil. Trans. R. Soc. B* 2014 369, 20130316
B. G. Abdallah et al, ACS Nano (2013) 7, 10534-10543
B. G. Abdallah et al, ACS Nano (2013) 7, 9129-9137