Institute of High Energy Physics, Chinese Academy of Sciences

PXRD-12

20-24 May 2013 Institute of Physics
Chinese Academy of Sciences, Beijing, China

Hard X-ray Free Electron Lasers for Structural Biology: Structure Determination of Nano-crystals and Single Particles

Yuhui Dong

BSRF, IHEP, CAS

2012/02/24

This document was presented at PPXRD - Pharmaceutical Powder X-ray Diffraction Symposium

Sponsored by The International Centre for Diffraction Data

PPXRD Website – www.icdd.com/ppxrd

ICDD Website - www.icdd.com

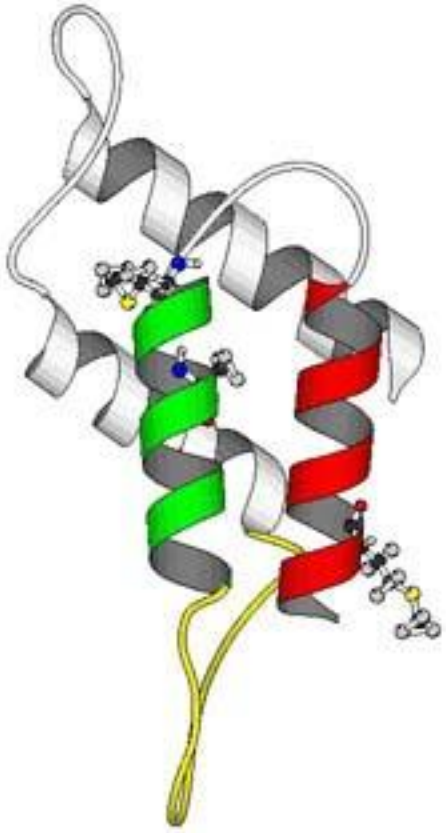
outline

- A brief introduction of structural biology
- Coherent X-ray: X-ray free electron laser
- Solving structures by XFEL

✓ Nano-crystals

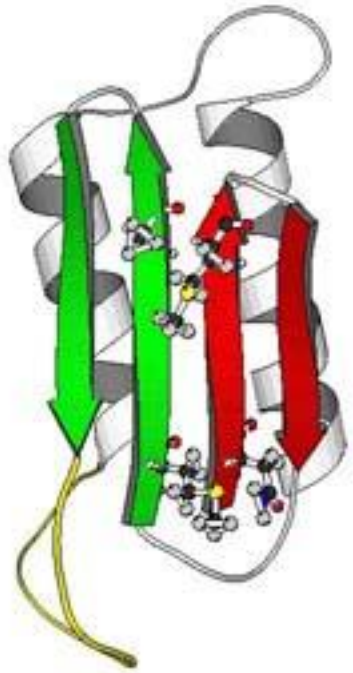
✓ Single particle

An example for the importance of structures of proteins—Prion



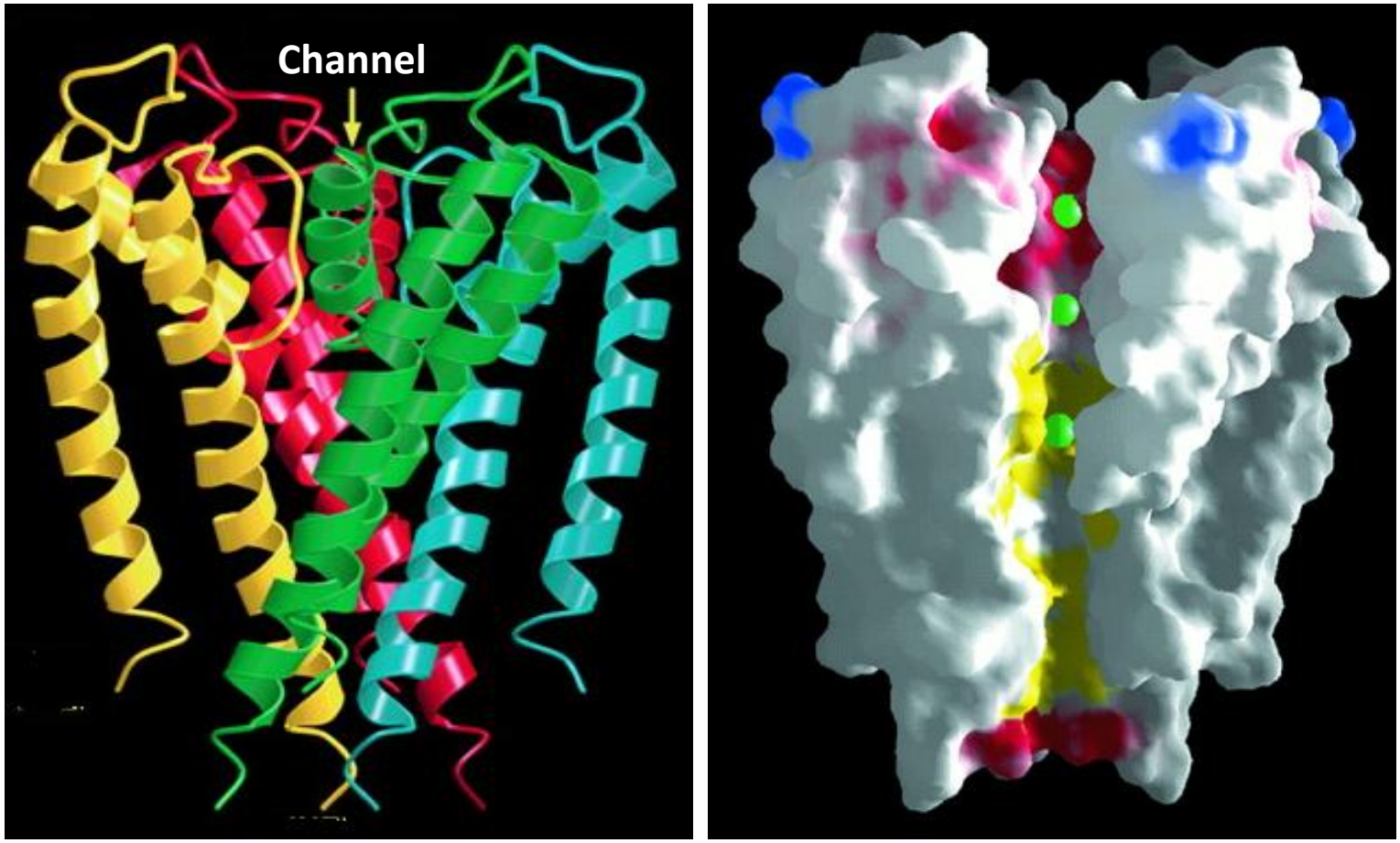
Normal prion

Same sequences
Same stechiometries
Difference in structures



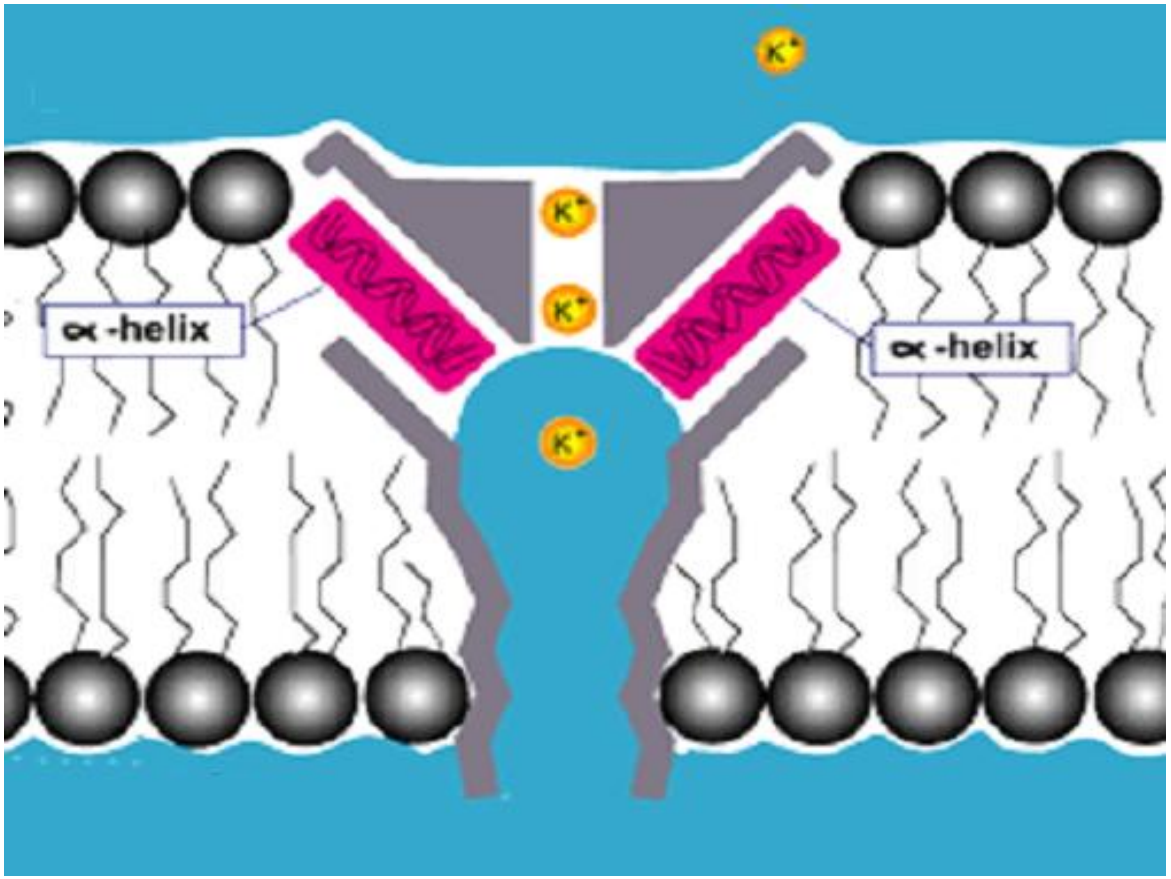
Prion caused mad-cow-disease

2003 Nobel prize for chemistry: Potassium ion channel



Why K^+ ion can pass through the membrane while the smaller Na^+ ions can't?

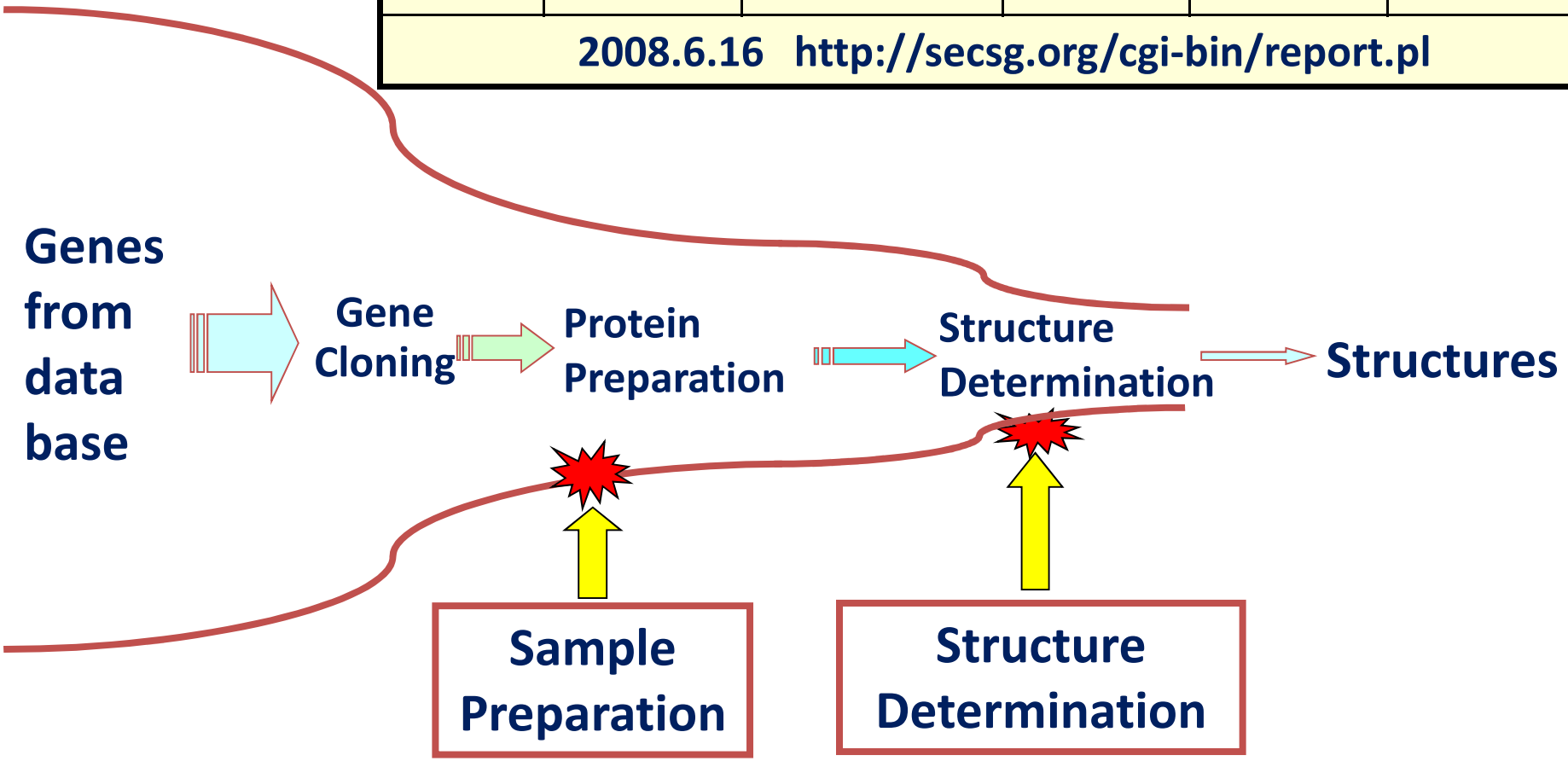
The mechanism revealed by structures



1) Hydrated ions;
2) The channels mimic the environments of hydrated K^+ ions, but not Na^+ ions;
3) Therefore Cs^+ ions, with the similar sizes and hydration, can also pass.

The way from genes to structures

Genes	Cloning	Expression	Soluble	Purified	Structures
15272	14866	6408	1474	843	80
2008.6.16 http://secsg.org/cgi-bin/report.pl					

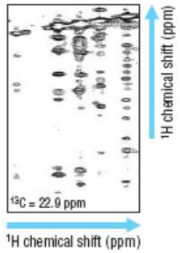


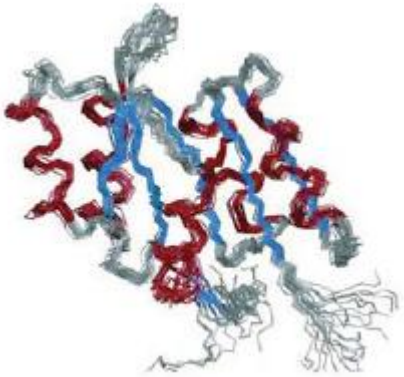
The methods for obtaining protein structures

Methods	Structures
X-ray	79916
NMR	9928
EM	545
Hybrid	52
Other	170
Total	90611

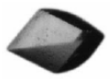
http://www.pdb.org/pdb/statistics/holdings.do, 2013/05/14

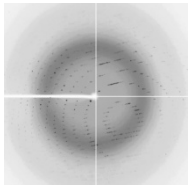
NMR

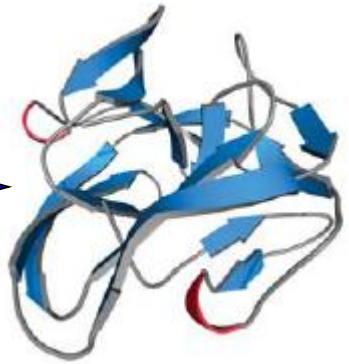

^{13}C = 22.9 ppm
^{1}H chemical shift (ppm)
^{1}H chemical shift (ppm)



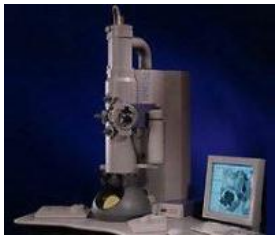
X-ray

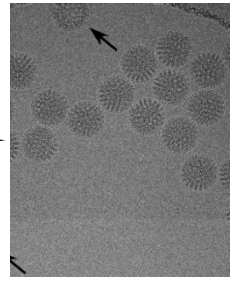


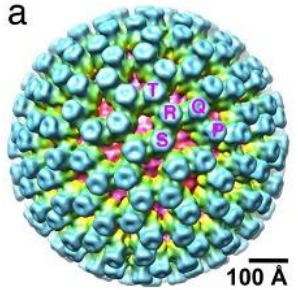




EM

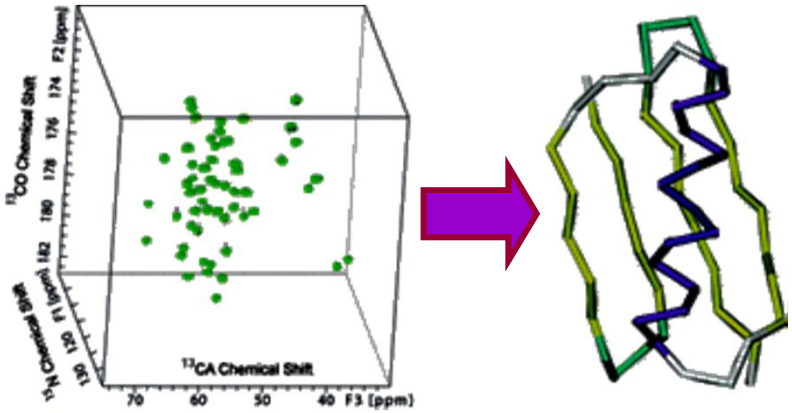





a
T
R
Q
S
P
100 Å

NMR can solve the structures in solution and suitable for studying the interactions

- Series of 2D, 3D NMR spectra
- The positions of the NMR peaks give the information of stereochemistry
- The structures, also the interaction can be reconstructed



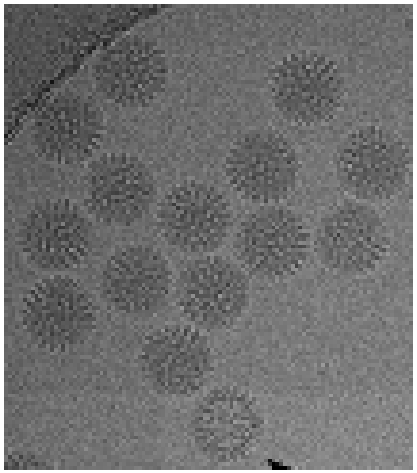
Limitation:

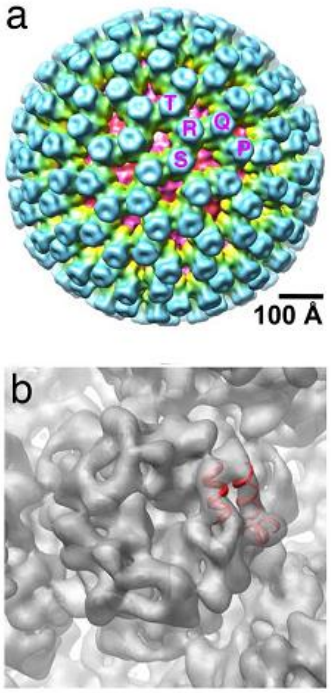
- Sensitivity: High concentration
- Resolution: Low molecular weight (<39kD)
- Slow: High stability
- Difficult to study weak interactions and transient states
- Difficult to solve the structures of membrane proteins

Pictures from Prof. Maili Liu, WIPM, CAS

Cryo-EM can solve the structures of huge complexes

- Many projections of identical particles
- The orientation of individual particle
- The 3D structures can be reconstructed



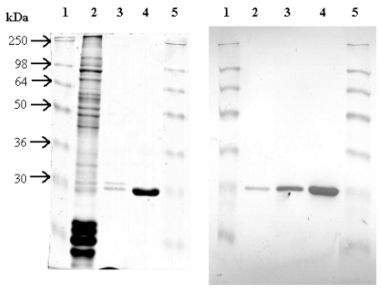


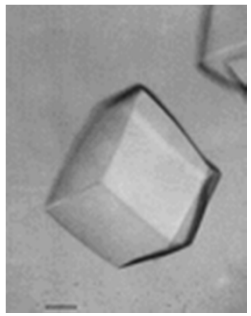
Limitation:

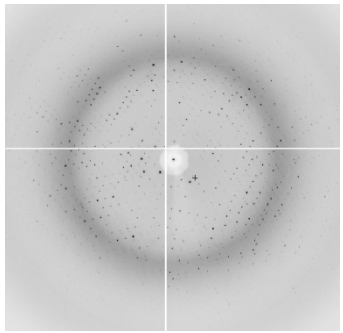
- Molecular weight (better >1MD)
- Resolution: typically ~1nm, near atomic resolution for high symmetry particles
- Low signal-to-noise ratio
- Huge amount of particles

Pictures from Zhang et al. Cell 141, 472, 2010

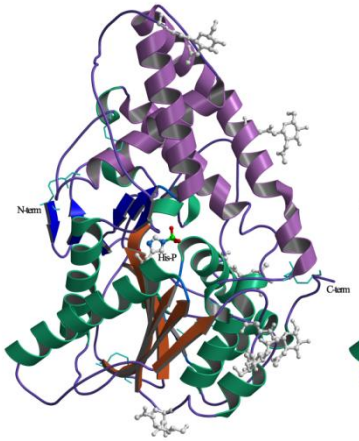
Protein crystallography is suitable for all kinds of macromolecules

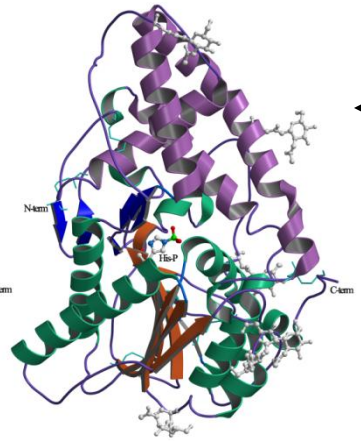


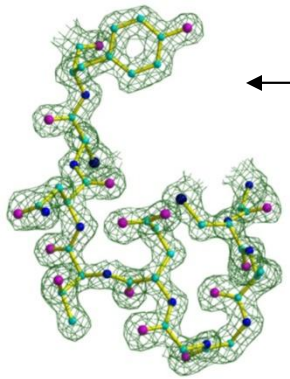




Expression/purification → crystallization → X-ray diffraction







$$F(h,k,l) = |F(h,k,l)| e^{i\alpha(h,k,l)}$$

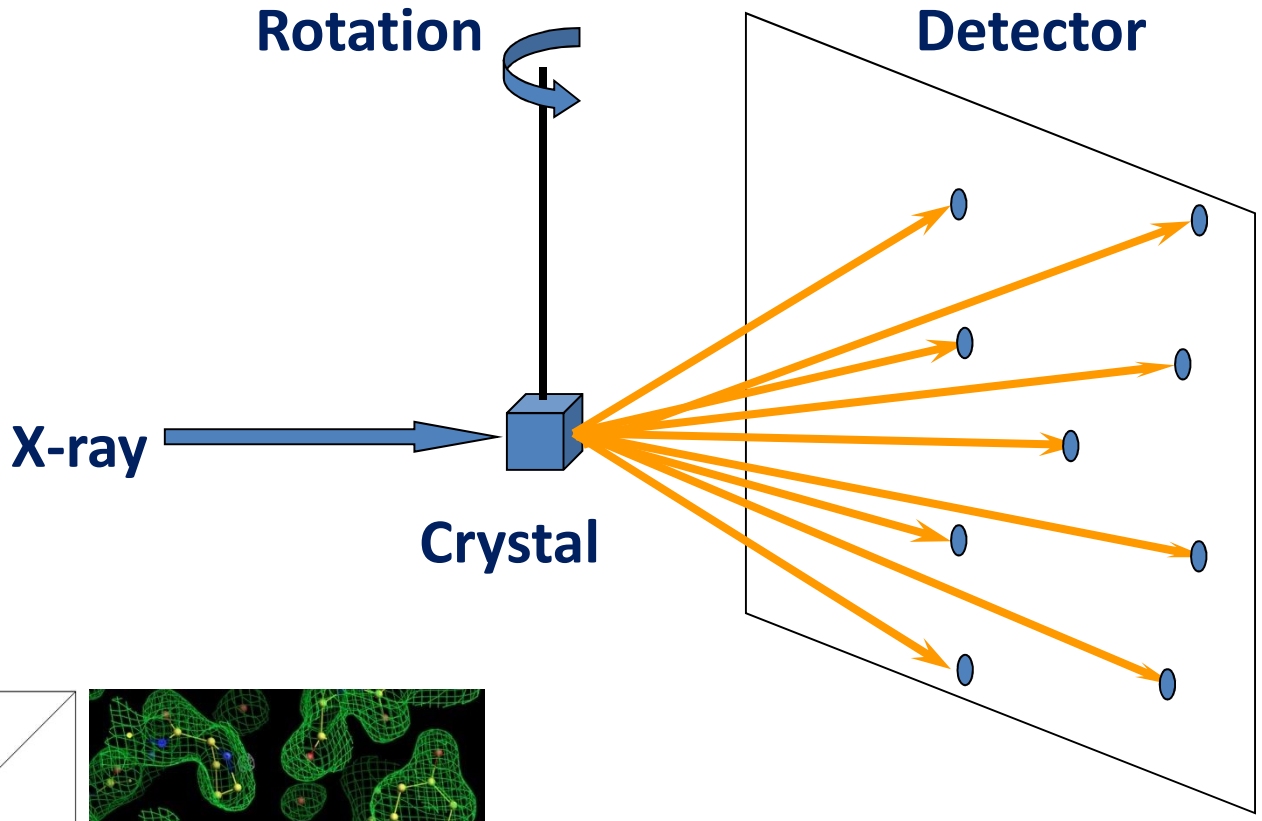
$$\alpha(h,k,l) = ?$$

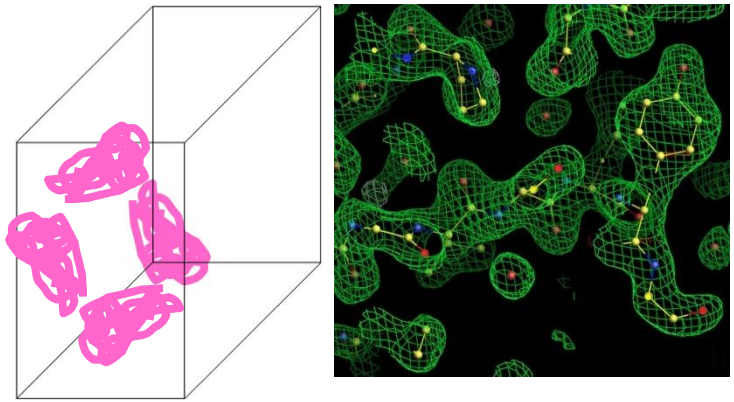
Refined structure ← Model building ← phasing

High-quality crystals are necessary, sometimes special crystals are needed.

Pictures from Prof. Quan Hao, HK Univ.

The experiments of PX





$$\rho(\vec{r}) = \sum_h \sum_k \sum_l \mathrm{F}(\vec{s}) \exp(-2\pi i \vec{r} \bullet \vec{s})$$

The methods for phasing

$$\rho(\vec{r}) = \sum_h \sum_k \sum_l \left|\mathrm{F}(\vec{s})\right| \exp[\mathrm{i}\varphi(\vec{s})] \exp(-2\pi \mathrm{i}\vec{r} \bullet \vec{s})$$

- **Molecular replacement (MR)**
- **Isomorphous replacement (IR)**
- **Anomalous dispersion (AD)**

- **For *de nuovo* structure determination: additional sample preparation or special experiments are necessary.**

Structures without crystals?

- NMR can only solve the structures of small proteins;
- Cryo-EM can only solve the structures of huge protein complexes;
- Protein crystallography is the most precise and widely-used method;
- Obtaining crystals is the bottleneck, especially good-quality crystals.

- Can we obtain structures by small crystals, or even without crystals?

Coherence of X-ray

- The spatial (transverse) coherent of light:

$$4\pi \cdot \sigma \cdot \sigma' = \lambda$$

where

σ: the size of light source;

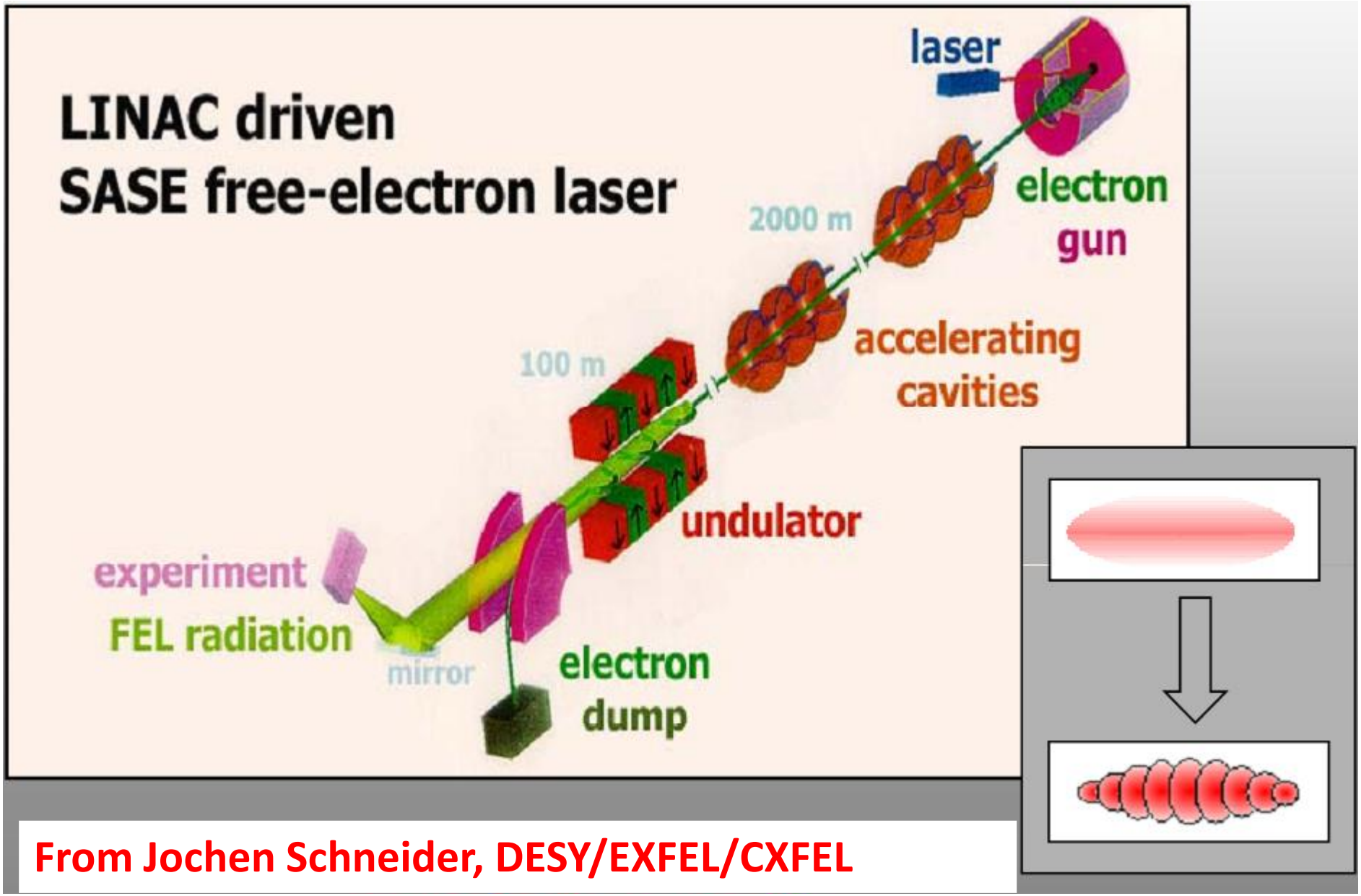
σ': the divergence of light source;

l: the wavelength.

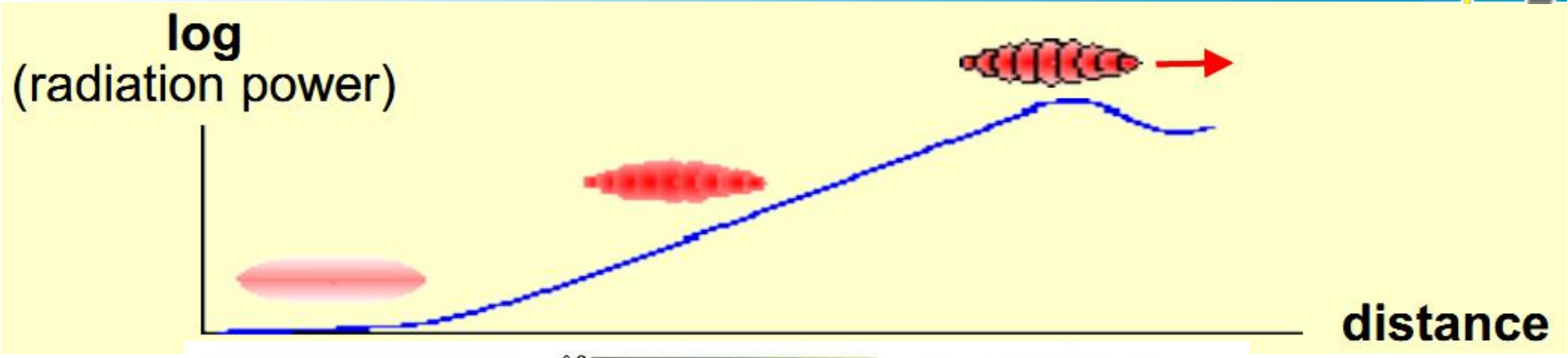
If the source is small enough (point source), or the divergence is small enough (parallel beam), then we obtain coherent light.

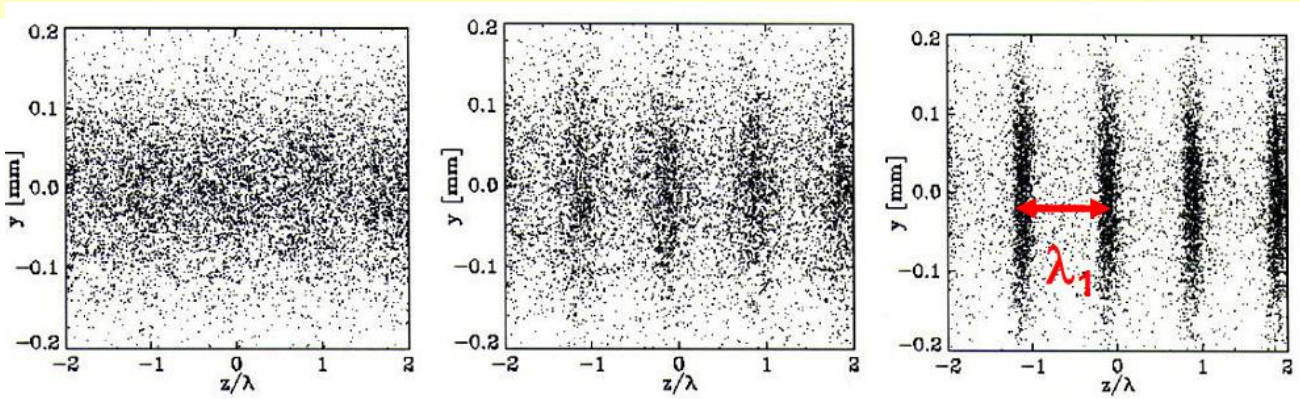
Free Electron Laser

- High energy Linac (~1000m long, the energy of electrons ~10GeV, $\gamma \sim 10^6$)
- Long undulator（~200m）
- The interaction between X-ray and electrons modulates the beam inside undulator: same phases, saturation.



From Jochen Schneider, DESY/EXFEL/CXFEL





e-gun: 1mmrad

Linac: 1nmrad
(γ=10^6)

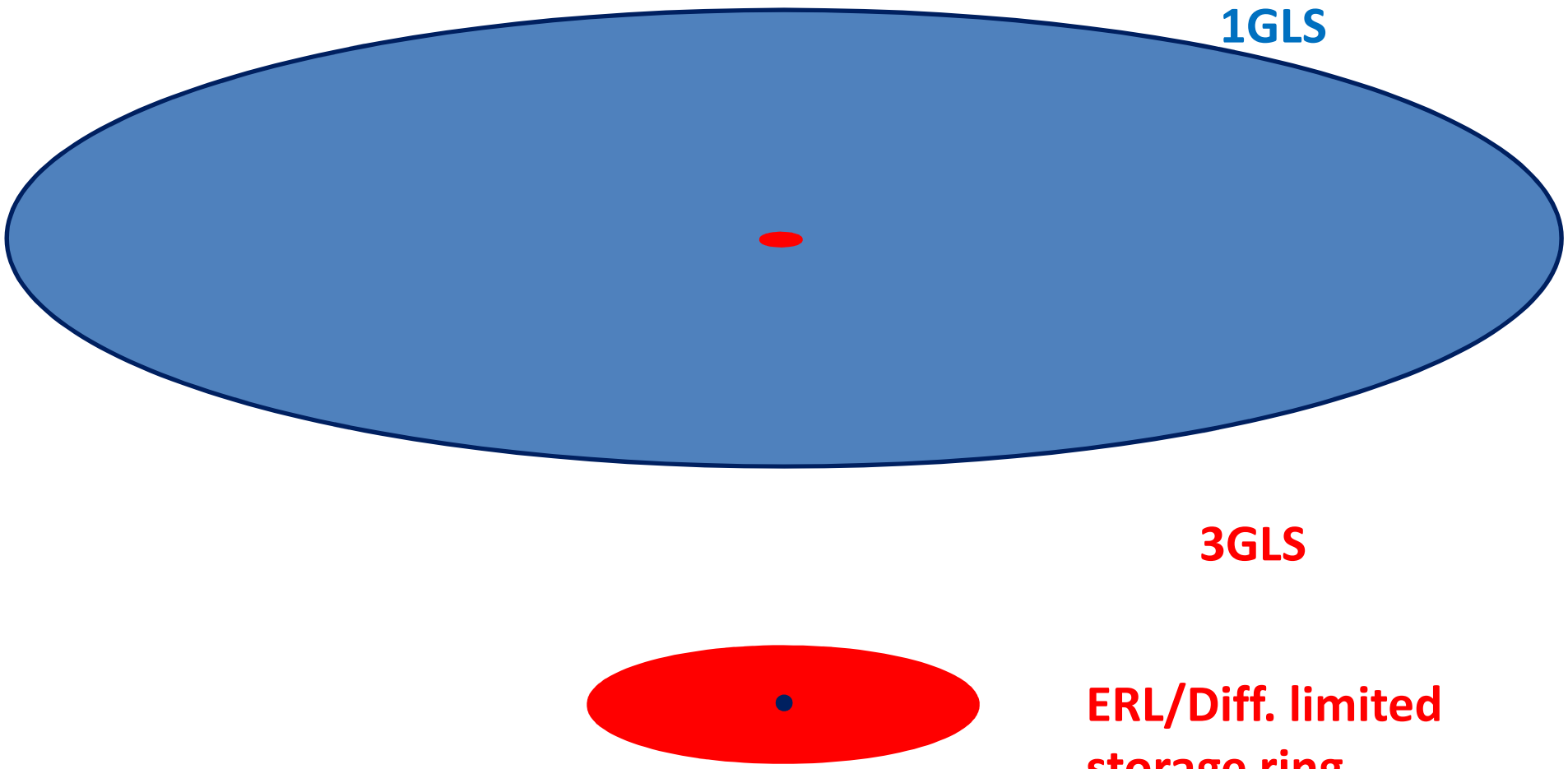
Undulator: slicing,
~10pmrad

$4\pi\cdot\sigma\cdot\sigma' = \lambda$

$4\pi\cdot 10\text{pm} = 1.256\text{Å}$

Pictures from Zhirong Huang, LCLS

Storage ring



The improvement of storage rings

- **1GLS**

  emittance ~100nmrad, coherent wavelength: 1256nm

  Coherent length at sample @1Å X-ray: L_t~500nm.

- **3GLS (SSRF)**

  emittance: 4nmrad, coherent wavelength: 50nm

  L_t~5mm

- **3GLS (NSLS-II)**

  1nmrad, 12.56nm

  L_t~20mm

- **ERL**

  0.1nmrad, 1.256nm

- **Diffraction limit ring (PEP-X & SPring-8 upgrading plan)**

  10pmrad, 1.256Å

Longitude (temporal) coherence

- **Longitude coherence is related to energy resolution:**

$$L_c = \lambda^2/\Delta\lambda$$

- **1Å X-ray, $\Delta\lambda/\lambda=10^{-4}$, $L_c=10^4$Å; $\Delta\lambda/\lambda=10^{-7}$, $L_c=10^7$Å.**

- **Usually the longitude coherent length is enough for structure determination.**

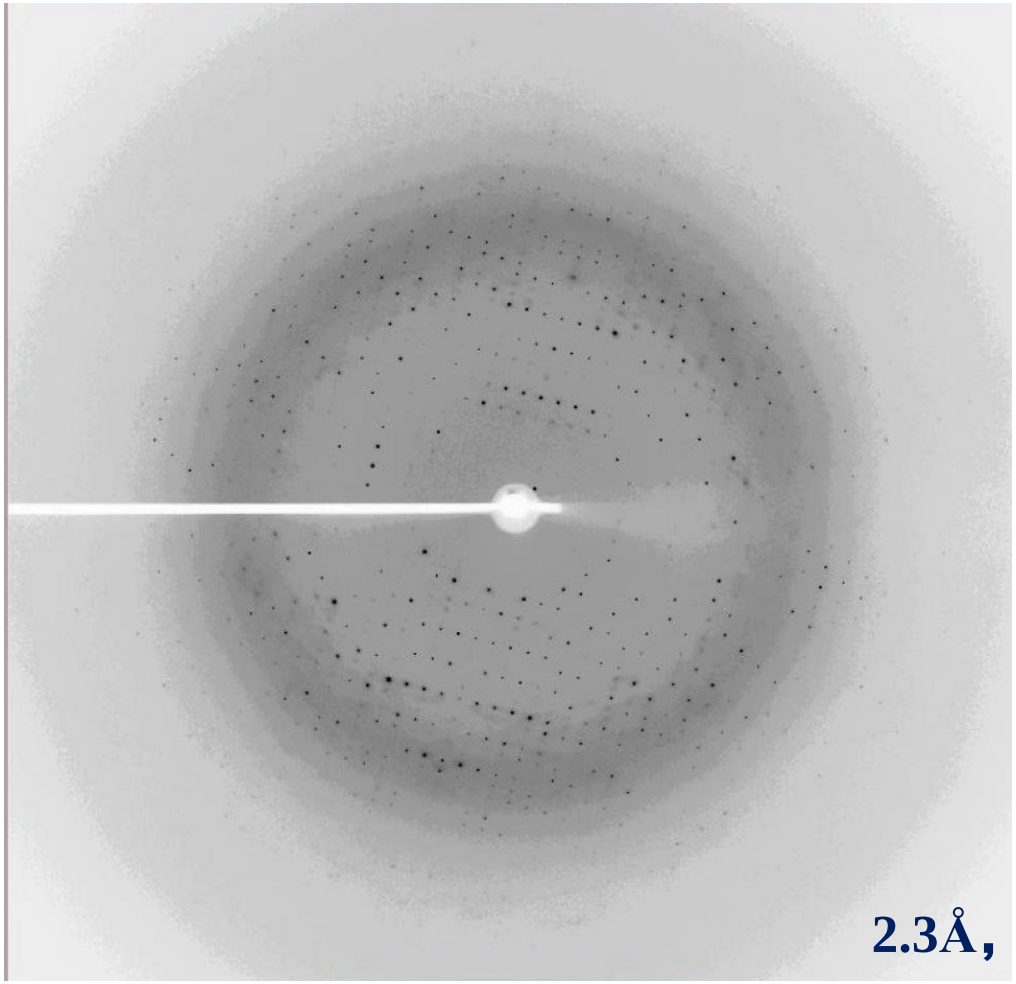
The key point: emittance

- Low emittance source = coherent X-ray source
- The only available source now: X-FEL
- Maybe diffraction limit storage ring or ERL?

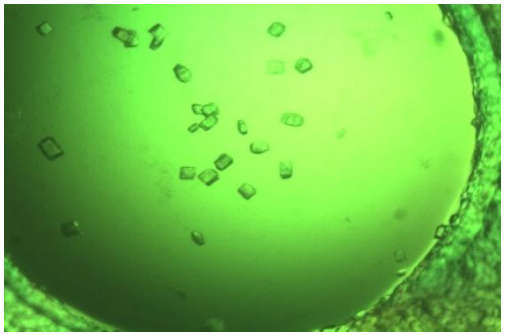
- Good coherence: smaller focus spot, better signals from smaller crystals, even detectable signals from non-crystals.

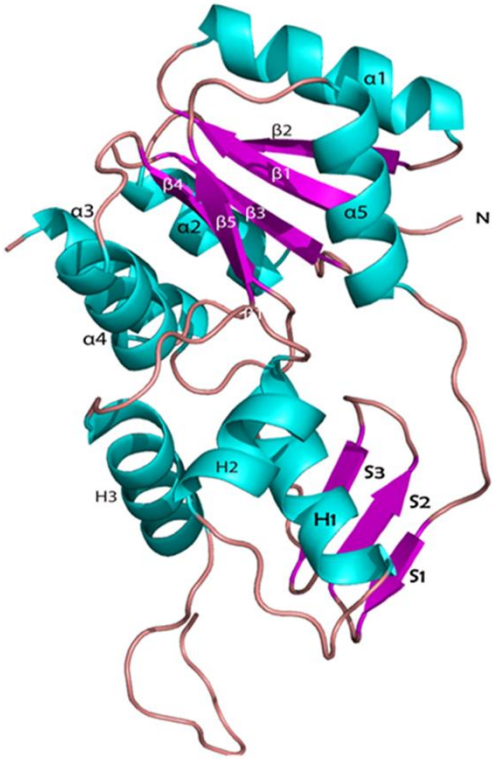
A case of small crystal: Dr-rrA-TD7(1-215)

SeMet-DrRRA-TD7@SSRF-BL17U, ~15μm

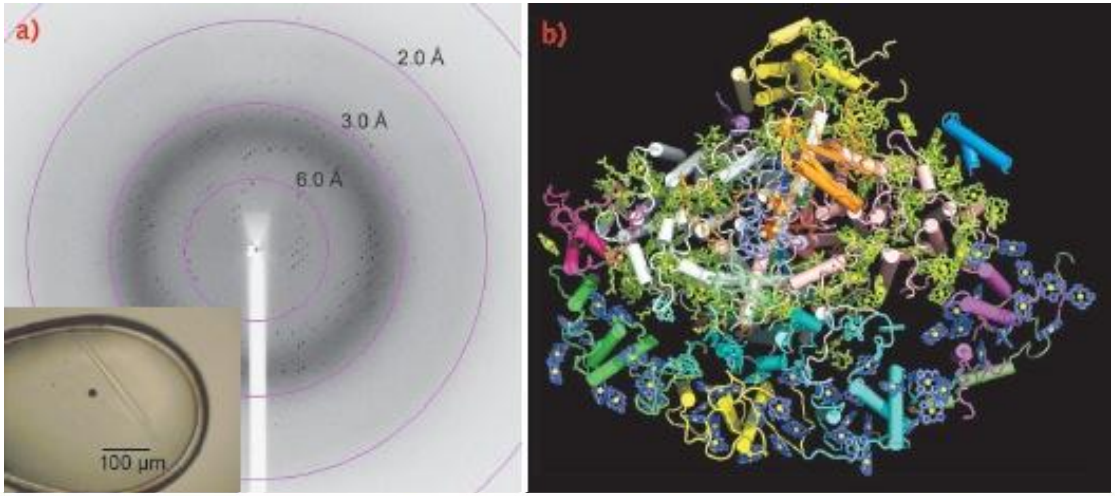


2.3Å, SAD

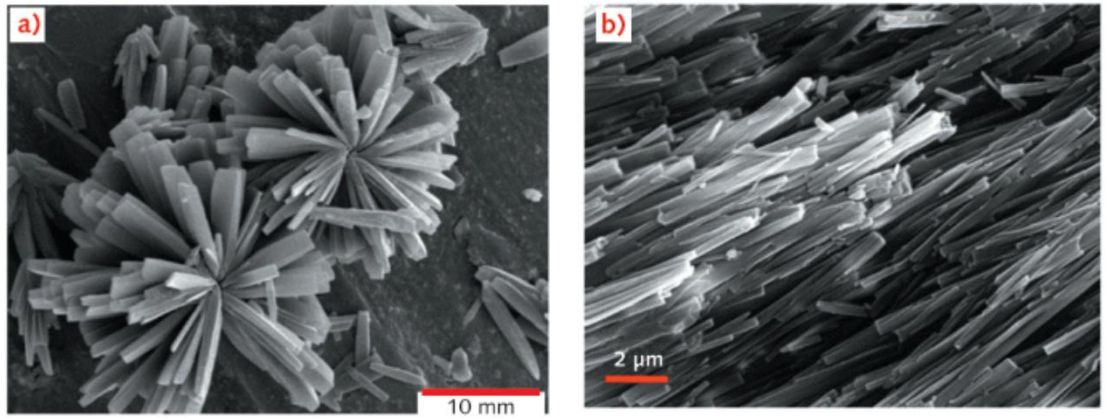




Protein structure determination for crystals smaller than 1μm



The structures of 10μm-size crystals can be easily solved（ESRF，J. Mol. Biol., 367, 310-318. 2007）
New SR facilities: 5mm



It is not difficult to obtain 1μm-size crystals, but not easy to grow big crystals.

The low emittance (<1nmrad) SR source would provide the beam smaller than 1μm. However, the radiation damage becomes the main problem.

Pictures from ESRF upgrading CDR

What structure biologists want?

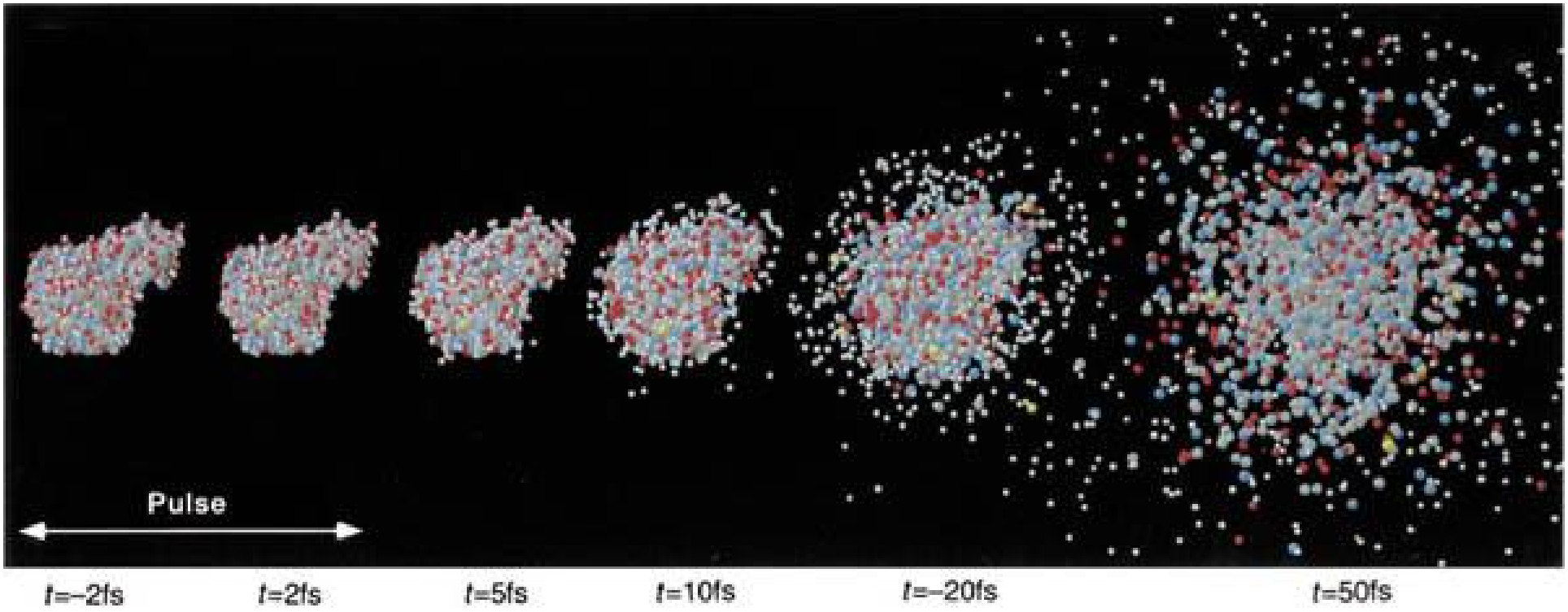
- Structures without crystals
- Crystallization is always a problem, especially high-quality crystals

- Nano-crystallography
- If your crystals are not good enough, usually due to the qualities of your proteins…
- But it will be very helpful if the structures can be solved via small crystals.
- Coherent X-ray may help.

Does it work?

- Can we obtain the diffraction/scattering signals from nano-crystals/single particle?
- How to solve the phases?
- In the case of non-crystal, how to reconstruct the 3D structure?

The protein molecules will be destroyed by XFEL



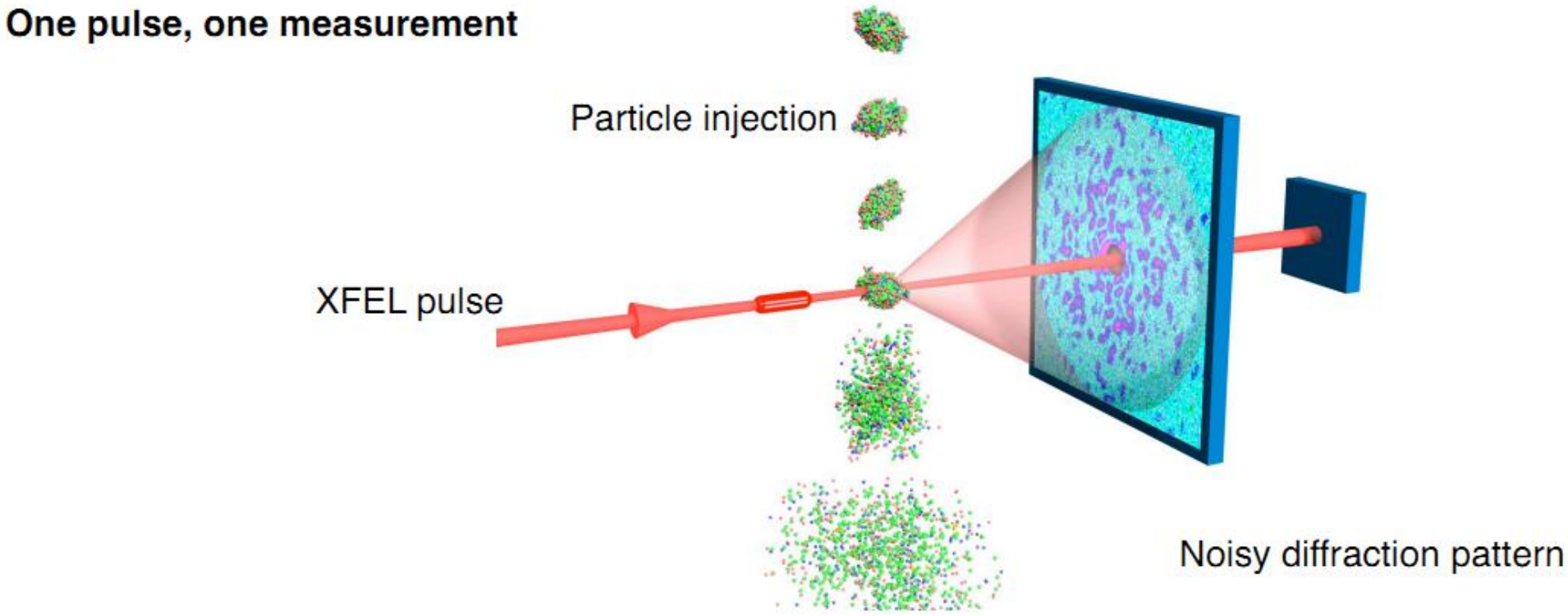
The pulse shorter than 2fs: reliable structures;

5fs: some errors;

>10fs: Distorted structures

R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, Nature 406 (2000)

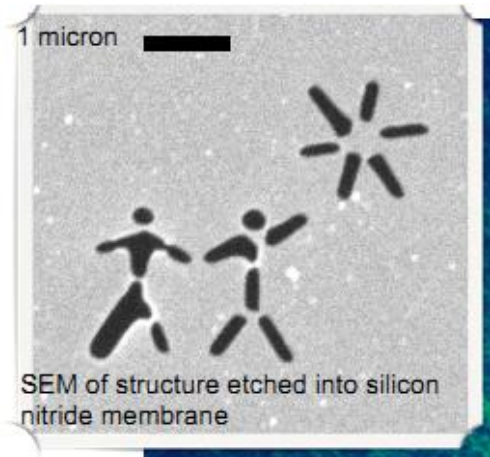
Diffraction before destruction

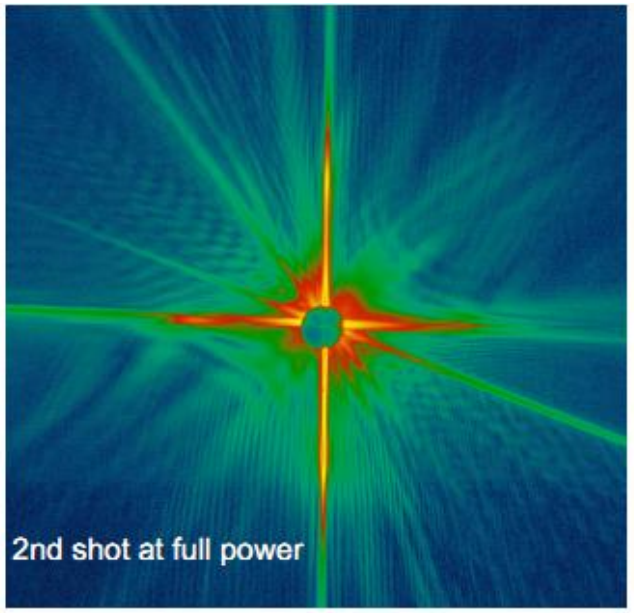


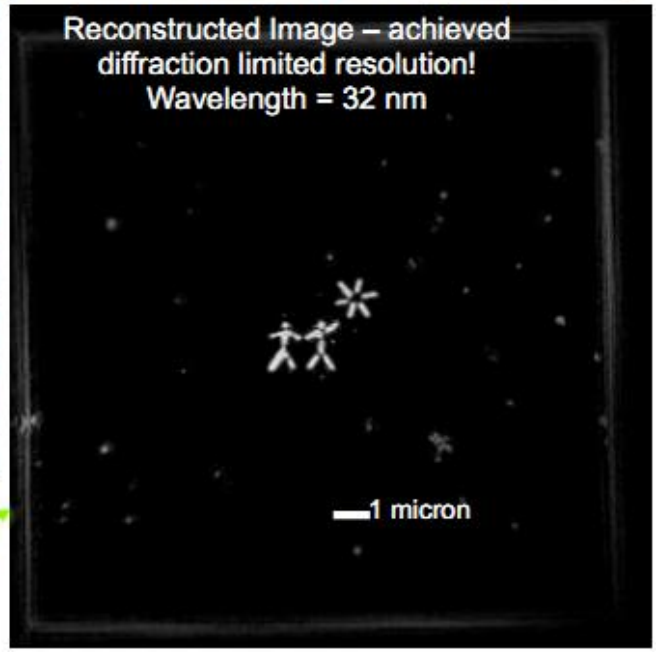
If the XFEL pulses are short enough, reliable structural information can be recorded before the samples are destroyed.

From Henry Chapman

Demonstrated with soft X-rays at FLASH



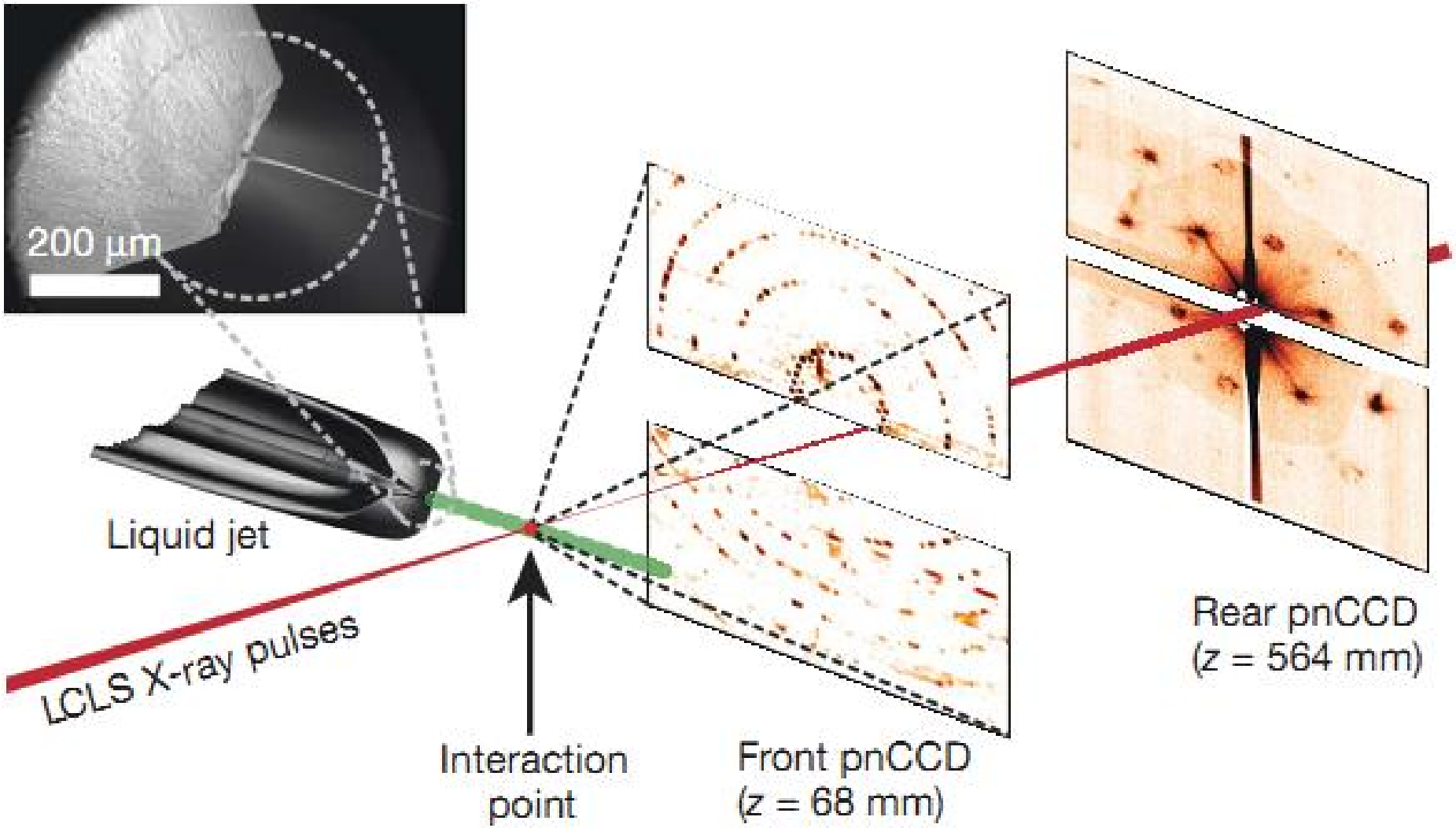




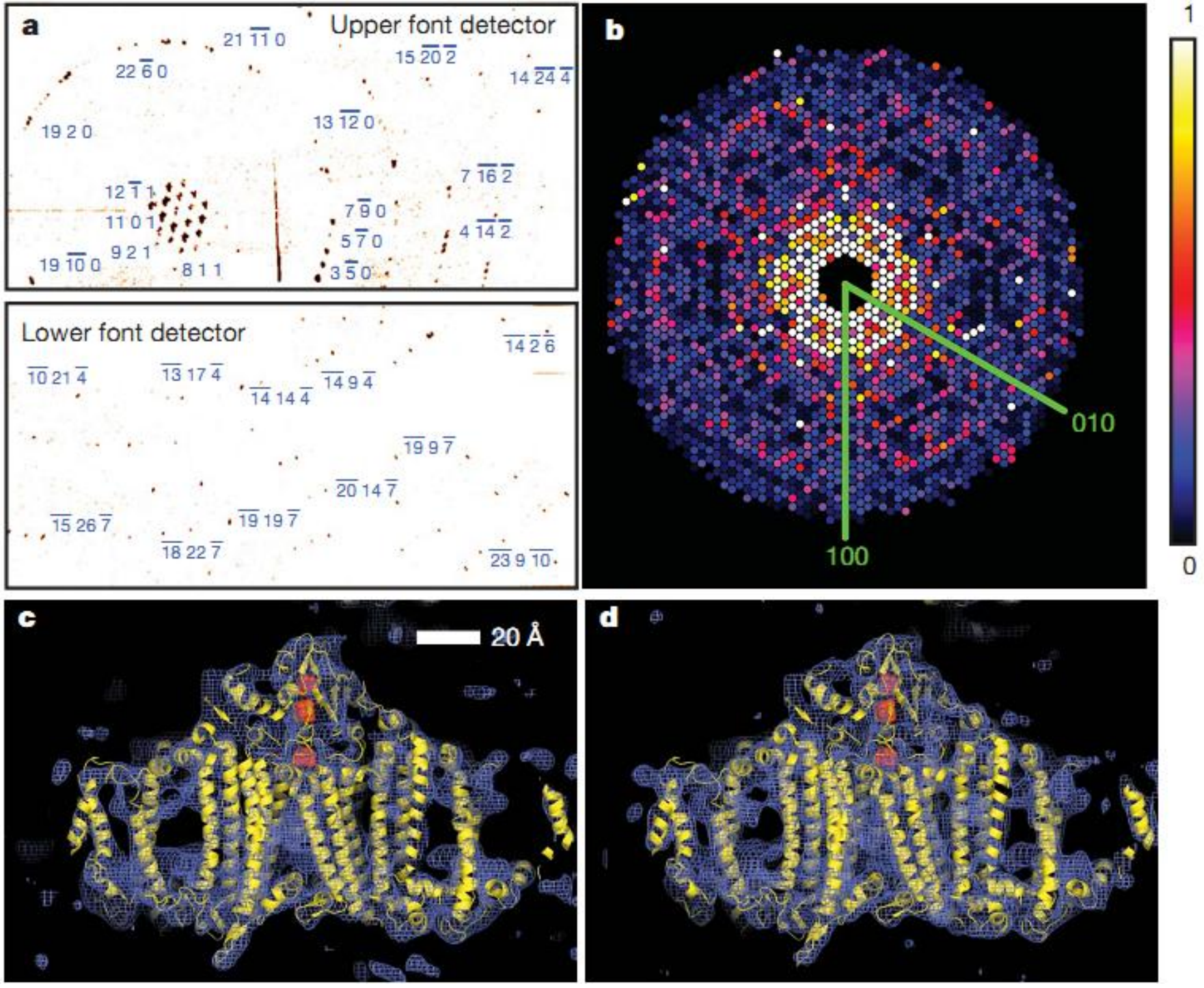
Edge of membrane support also reconstructed

Chapman et al, Nature Physics 2 839 (2006)

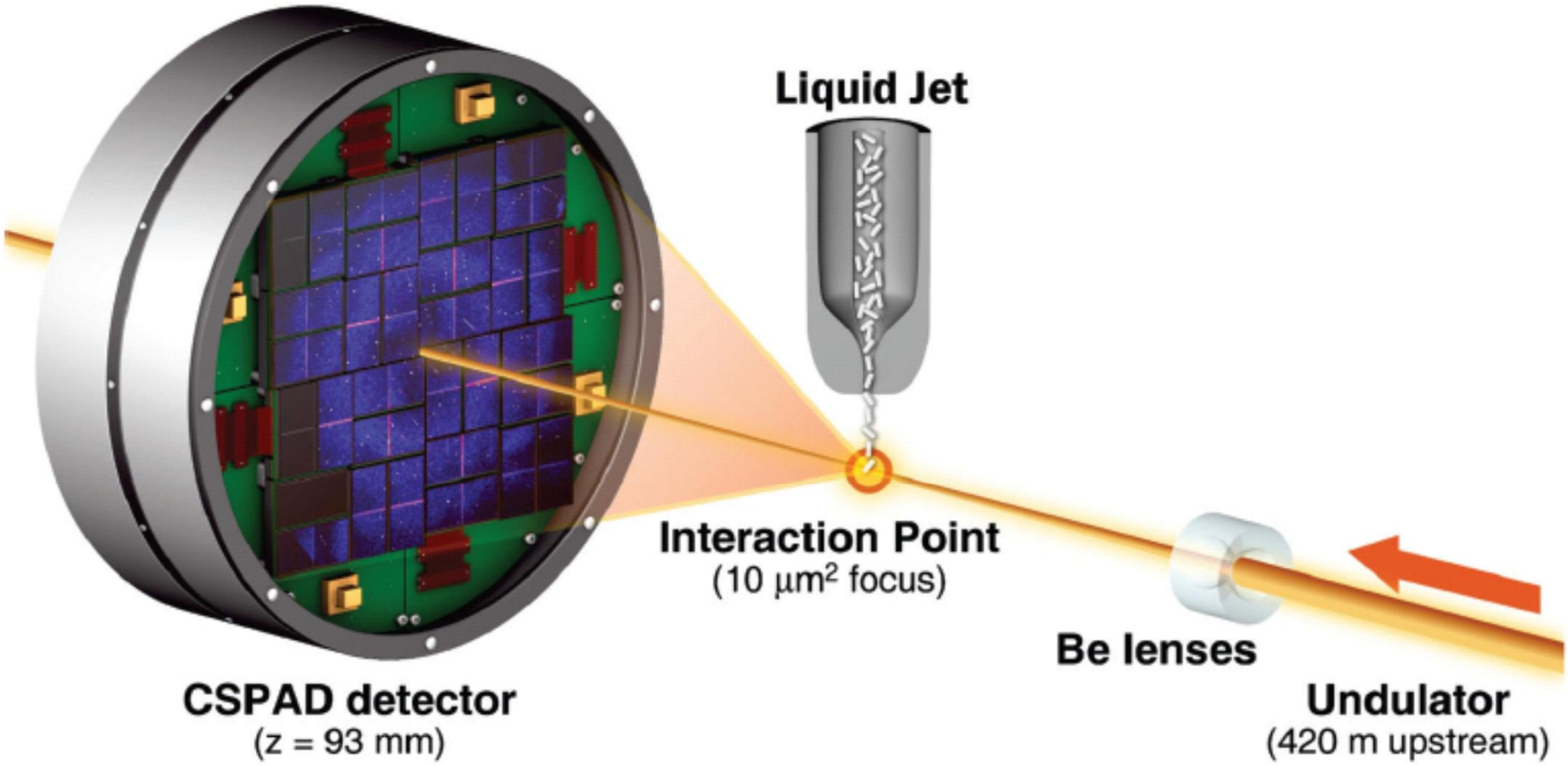
Verified with nano-crystals in LCLS



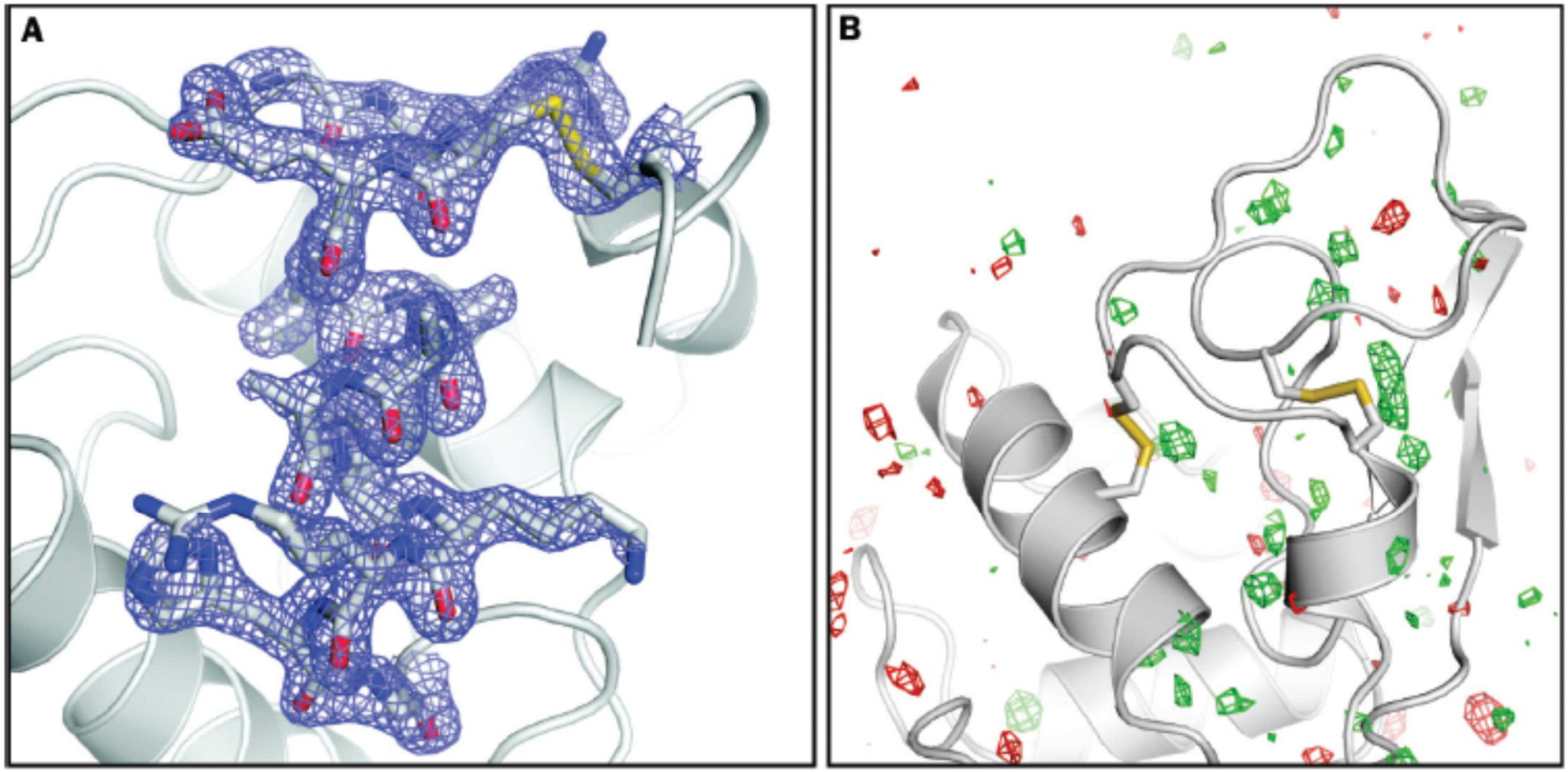
Chapman et al. Nature 470 73–77 (2011)


a
Upper font detector
21 $\bar{1}$ 0
22 $\bar{6}$ 0
15 $\bar{2}$0 $\bar{2}$
14 $\bar{2}$4 $\bar{4}$
19 2 0
13 $\bar{1}$2 0
7 $\bar{1}$6 $\bar{2}$
12 $\bar{1}$ 1
7 $\bar{9}$ 0
11 0 1
4 $\bar{1}$4 $\bar{2}$
5 $\bar{7}$ 0
9 2 1
19 $\bar{1}$0 0
8 1 1
3 $\bar{5}$ 0
Lower font detector
$\bar{1}$4 $\bar{2}$ $\bar{6}$
$\bar{1}$0 21 $\bar{4}$
$\bar{1}$3 17 $\bar{4}$
$\bar{1}$4 14 $\bar{4}$
$\bar{1}$4 9 $\bar{4}$
$\bar{1}$9 9 $\bar{7}$
$\bar{2}$0 14 $\bar{7}$
15 26 $\bar{7}$
$\bar{1}$9 19 $\bar{7}$
$\bar{1}$8 22 $\bar{7}$
$\bar{2}$3 9 $\bar{1}$0
b
010
100
1
0
c
20 Å
d

LCLS-lysozyme structure

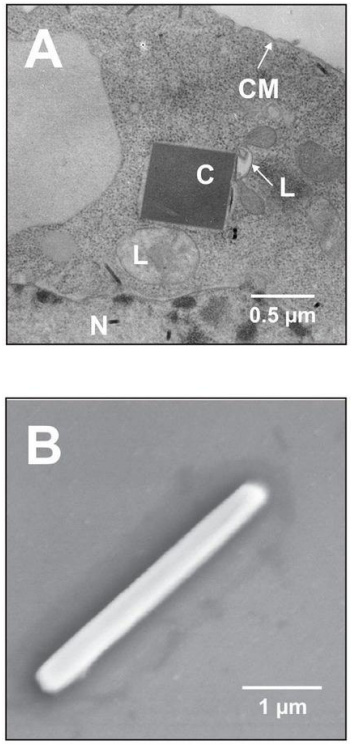


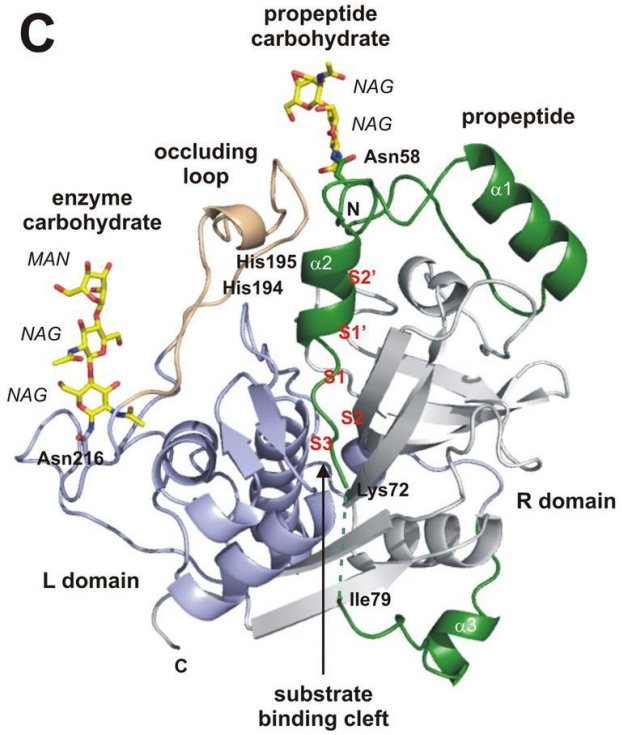
Structure solved via MR

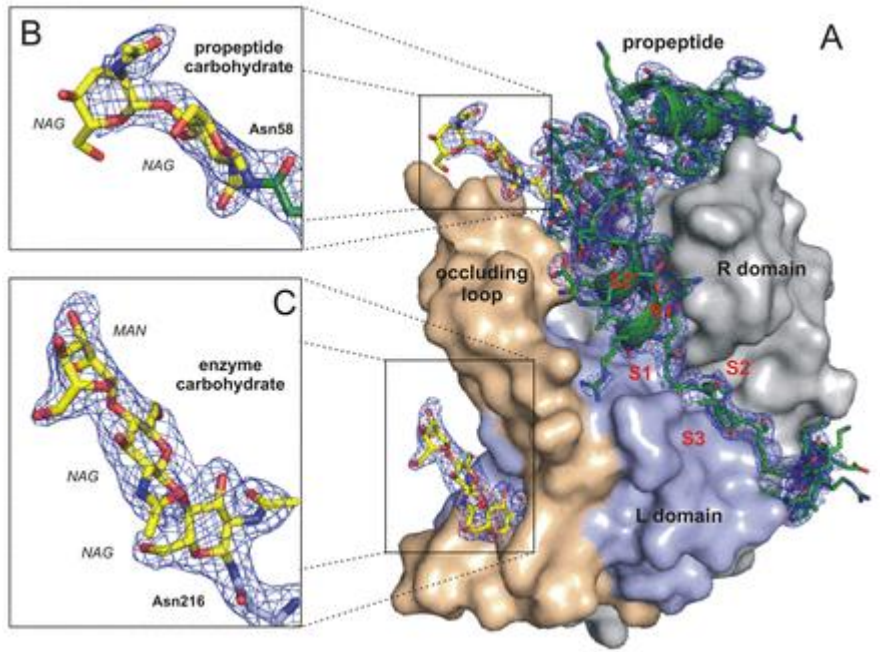


Boutet et al. Science 2012

Natively Inhibited *Trypanosoma brucei* Cathepsin B







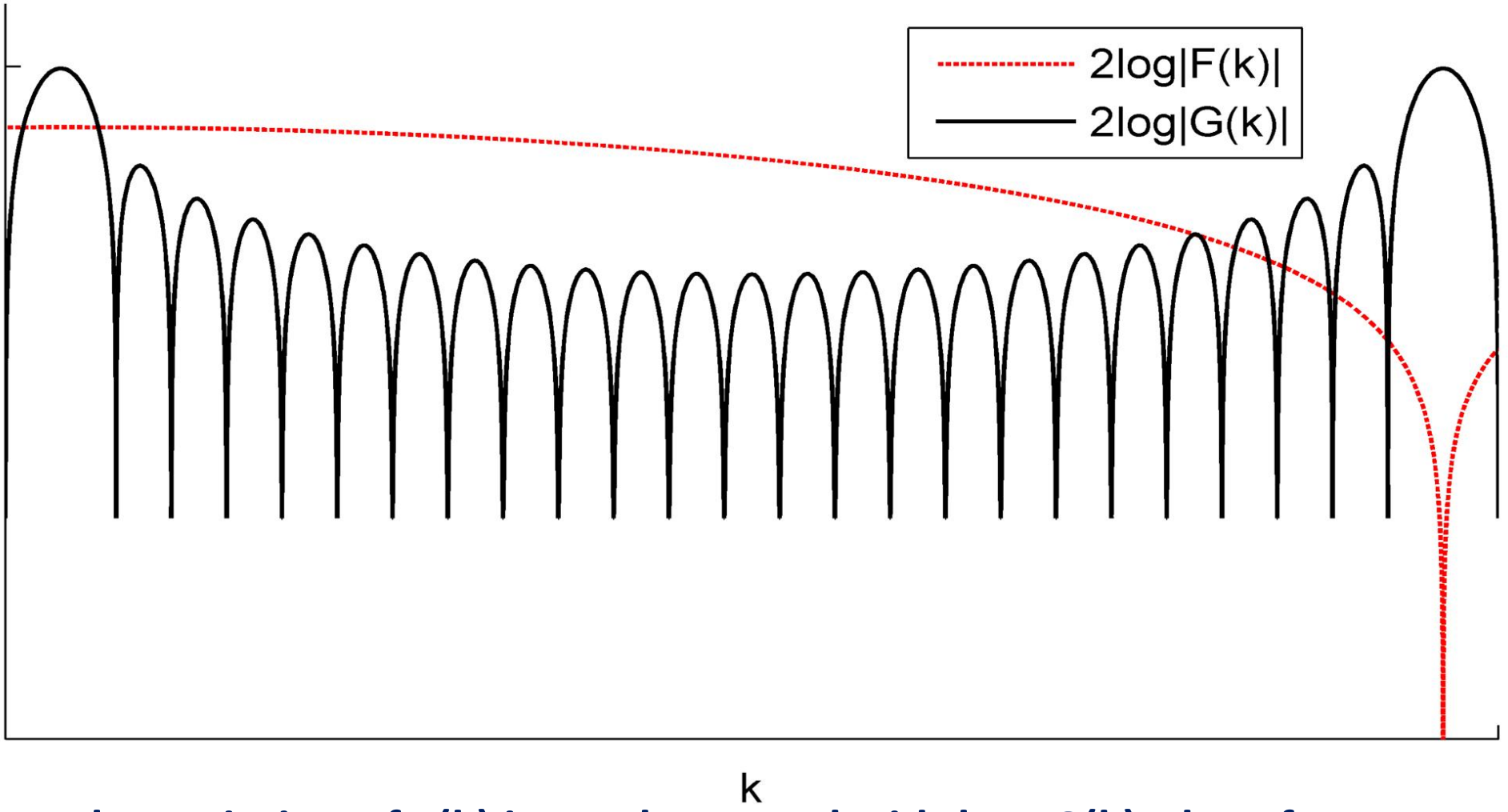
By combining two recent innovations, *in vivo* crystallization and serial femtosecond crystallography, we obtain the room-temperature 2.1 Å resolution structure of the fully glycosylated precursor complex of TbCatB.

Redecke et al. Science 2012

These experiments shows

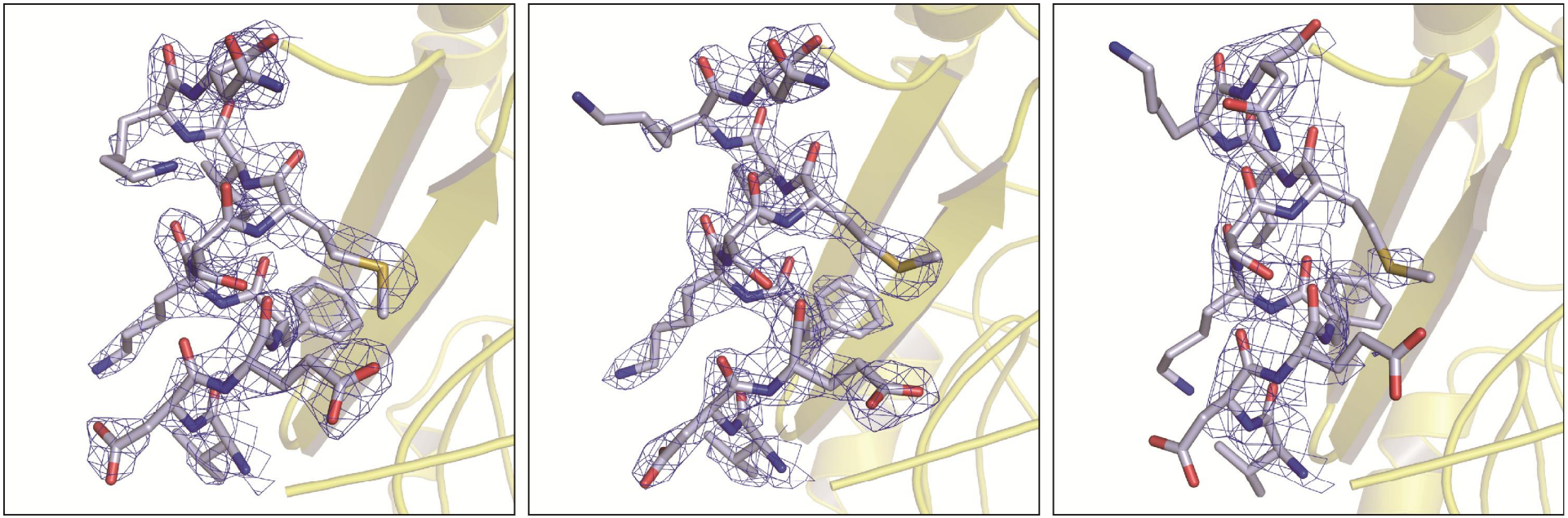
- "Diffraction before destruction" works to atomic resolution, the diffraction patterns can be obtained.
- The structures can be solved by molecular replacement, but too many patterns used (20,000-40,000 patterns).
- However, the phases of diffractions can not be solved via anomalous signals, due to the uncertainty of central symmetry related orientations.
- What structure biologists want: Solving *de nuovo* structures.

Geometric factor G(k) & structure factor F(k)



The variation of F(k) is much more placid than G(k), therefore G(k) can be calibrated to obtain the F(k) of a single cell.

More precise F(k), fewer patterns



(a) (b) (c)

Calibrating of G factors calculated at 5.0 Å resolution

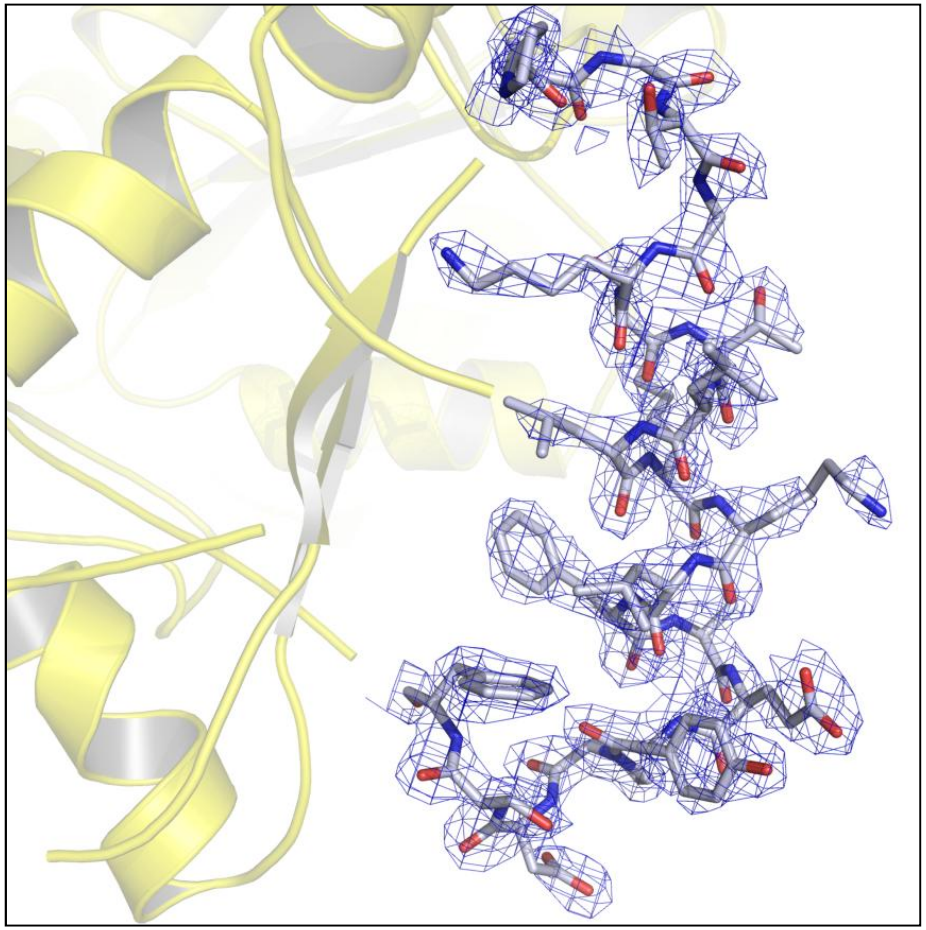
Calibrating of G factors calculated at 2.5 Å resolution

No calibration

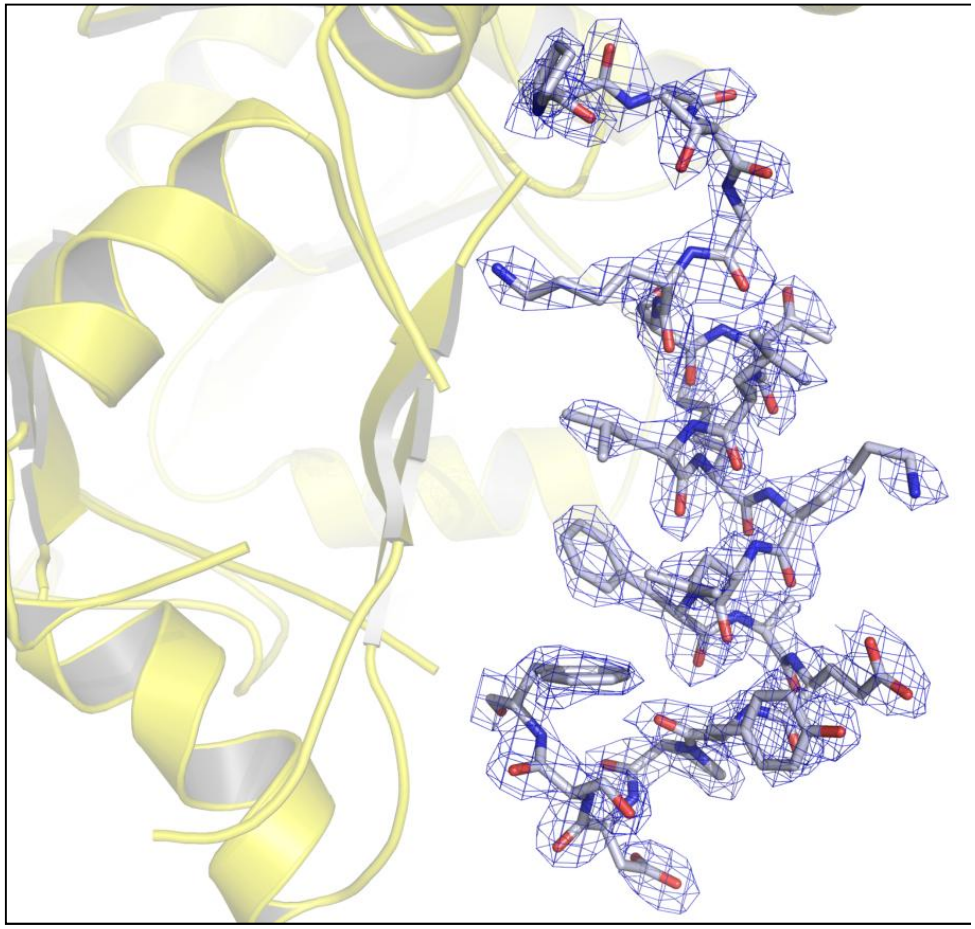
Only 2300 patterns used.

Qu et al. submitted.

The resolution ranges used to calculate G factors



G at 2.5 Å
R/Rsplit: 0.3440/0.1871
C: 98.35%



G at 5.0 Å
R/Rsplit: 0.1119/0.1225
C: 97.87%

The anomalous signals used in PX for *de nuovo* structure determination

To solve the structures of protein crystals, the electron densities can be calculated by:

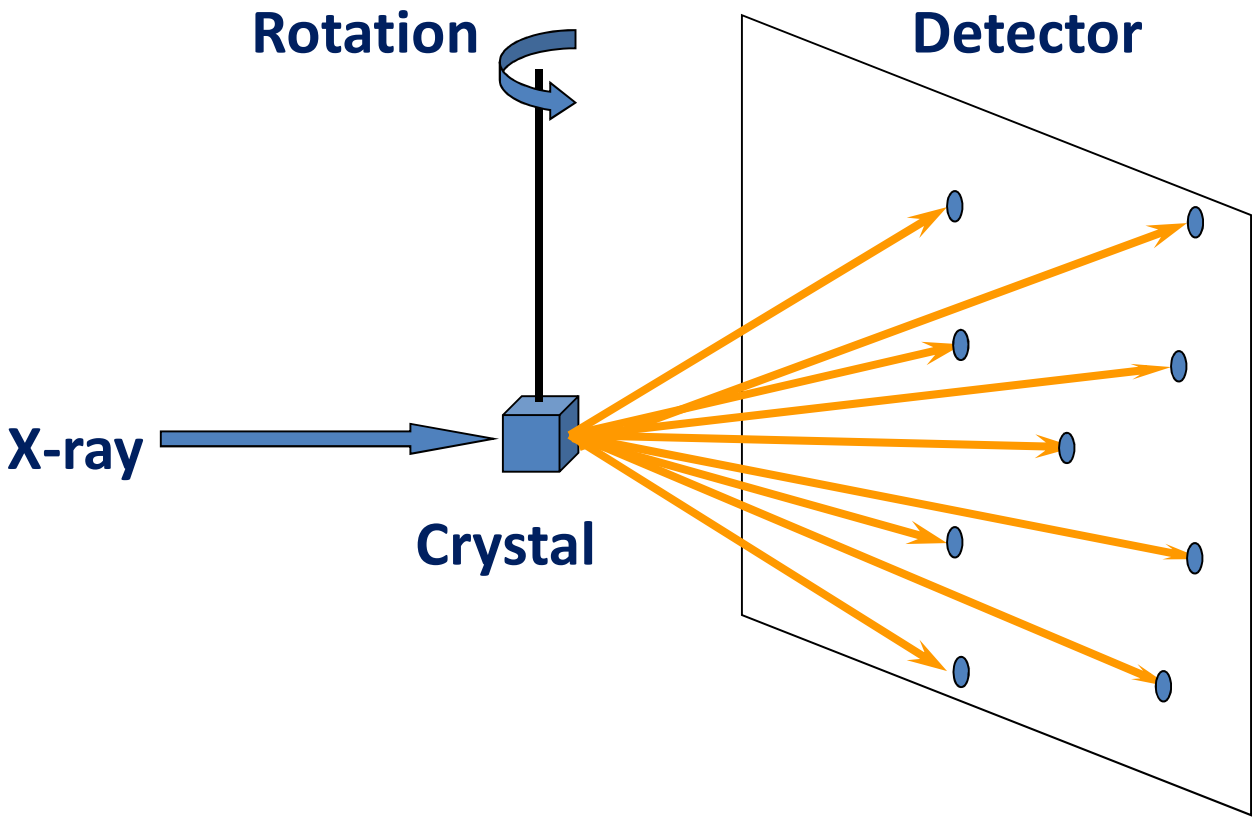
$$\rho(x_j, y_j, z_j) = \sum_m F(h_m, k_m, l_m) \exp[2\pi i(x_j h_m + y_j k_m + z_j l_m) + \varphi(h_m, k_m, l_m)]$$

Since ρ must be real, therefore, in the case of no anomalous scattering, F(h,k,l)=F(-h,-k,-l), φ(h,k,l)=φ(-h,-k,-l)+π

With anomalous scattering, F(h,k,l)≠F(-h,-k,-l). The difference can be used for solving the phases.

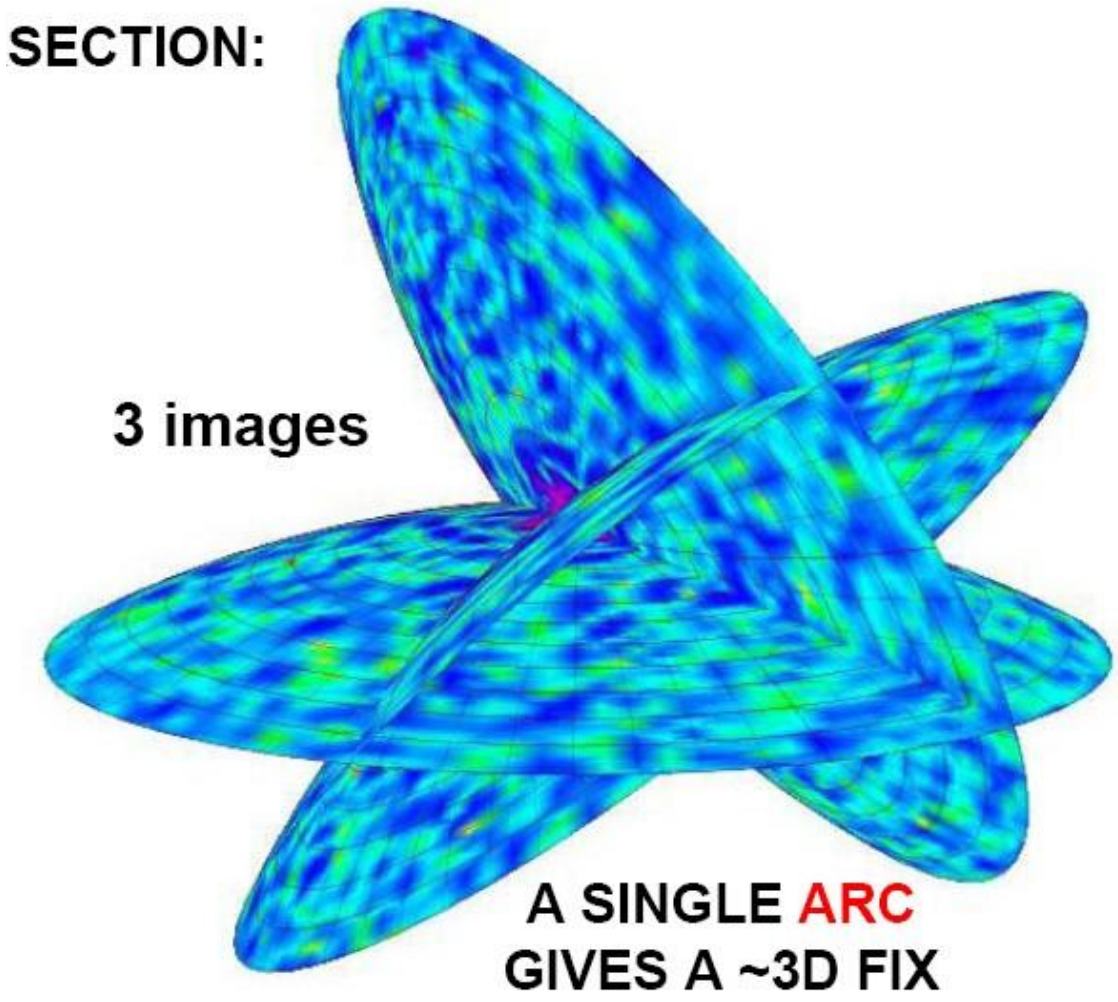
The diffraction spots (h, k, l) and (-h, -k, -l) can not be distinguished by crystallography only.

Traditional PX experiments



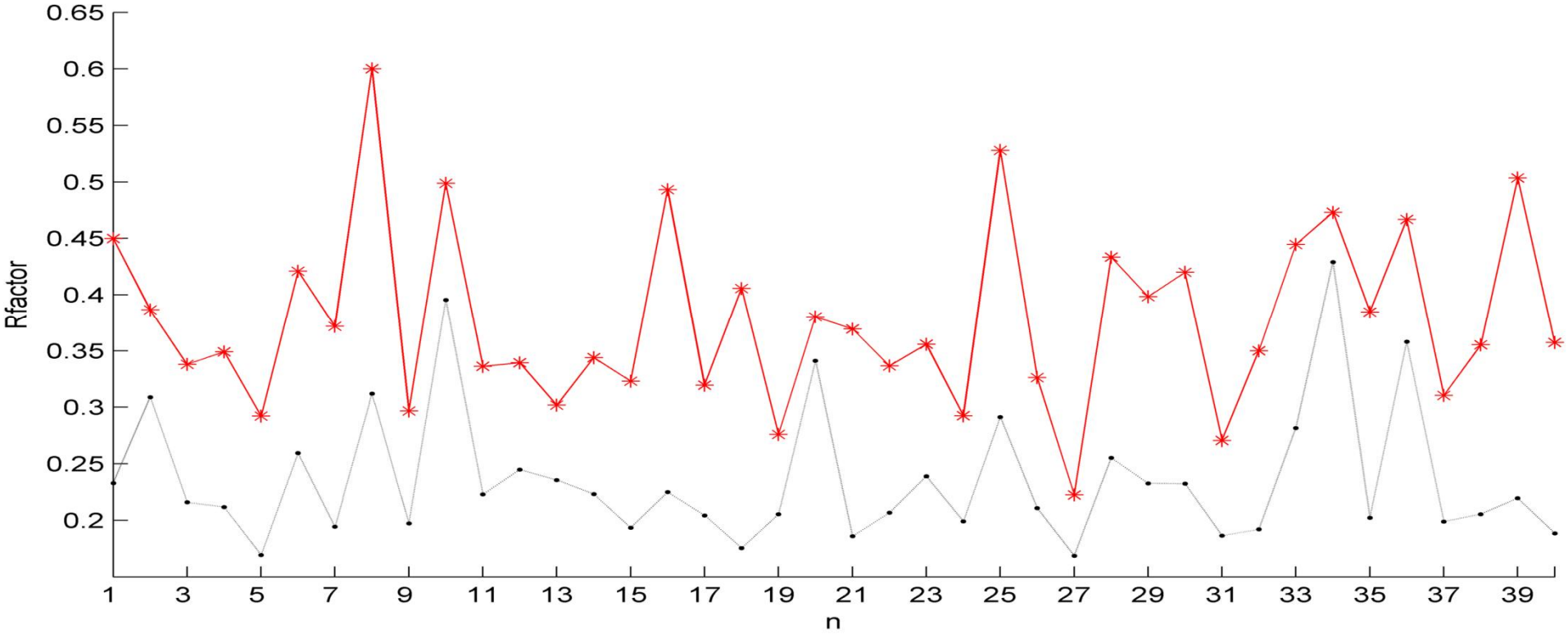
- Only 1 crystal is used, one can define the “implicit” coordinate system of crystal cell.

Solution: Common-line method



High resolution: the curvature of the Edwald sphere.

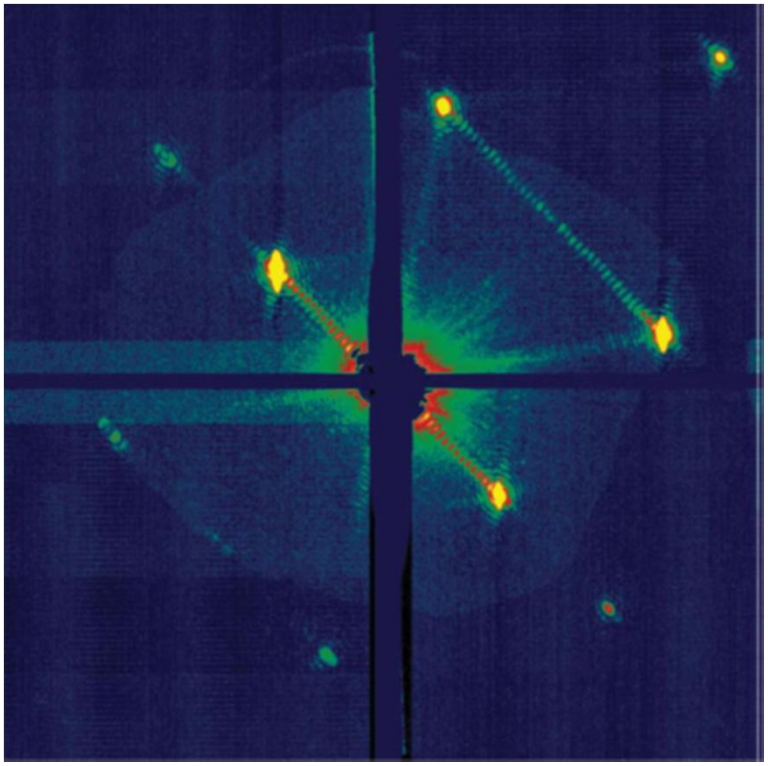
"implicit" coordinate system of crystal cell



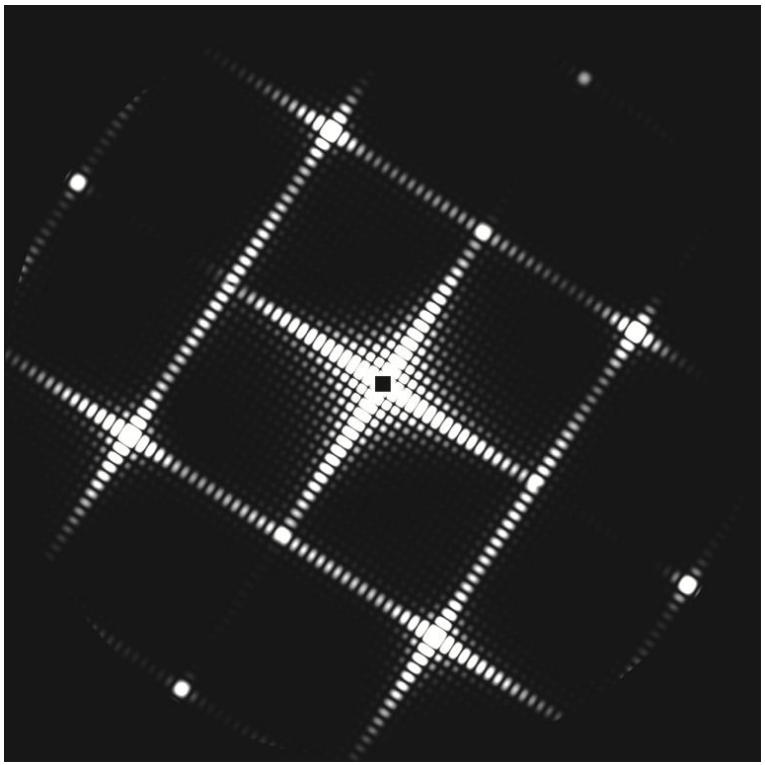
From the structure factors of a single cell obtained from different nano-crystals, by comparing the differences on common lines, the "+" and "-" coordinate systems can be distinguished.

Zhou et al. *Chinese Phys. C*.

"non-Bragg" peaks



$P\,6_3$



$P\,4_32_12$

No. collected diffraction images	1,471,615	1,997,712
No. of hits/indexed images	66,442/12,247	40,115/10,575

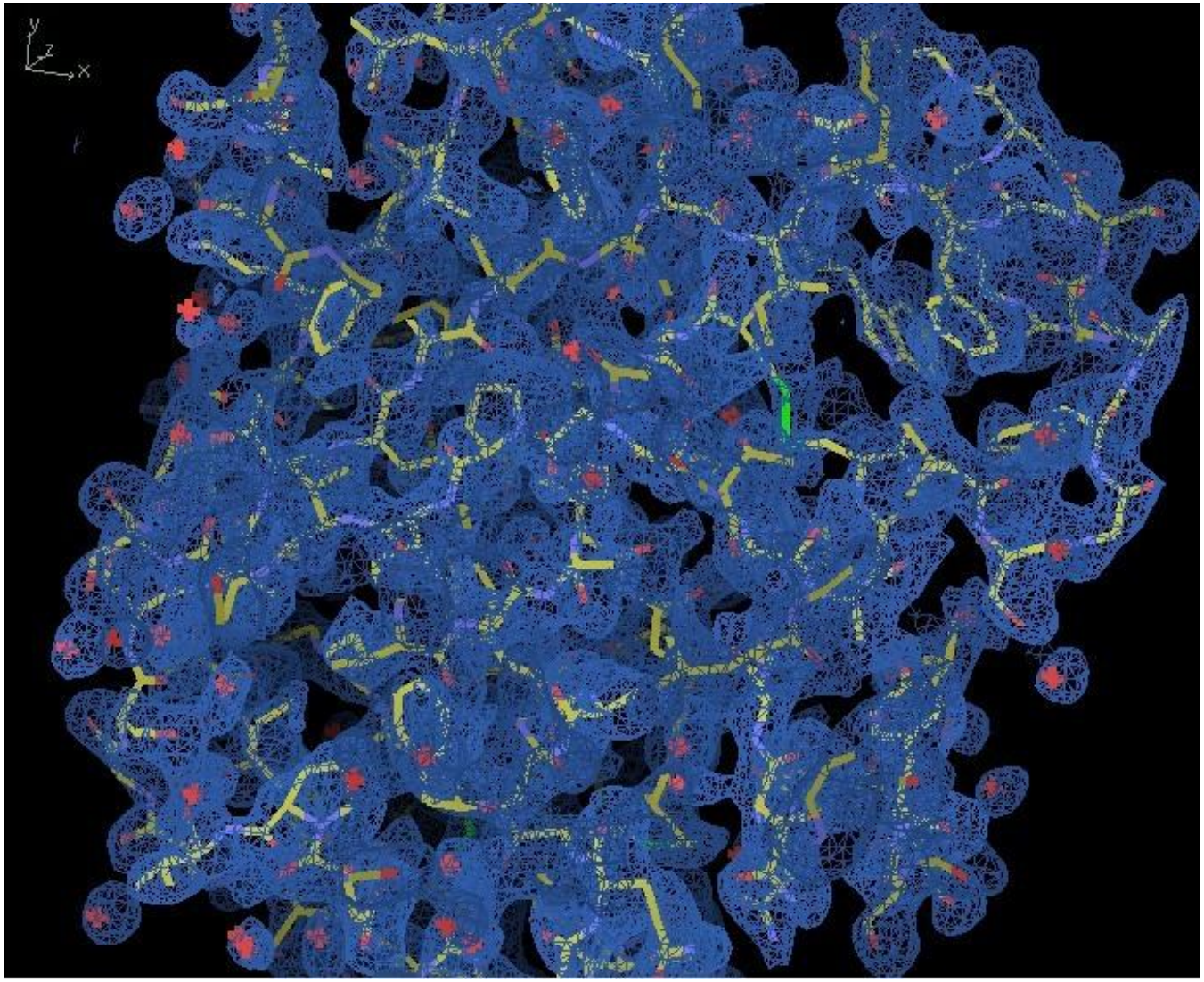
Chapman, H. N. *et al*. *Nature* 470, 73–77 (2011).
Boutet, S. *et al*. *Science,* 337, 362 (2012).

Oversampling in crystallography

- If the sizes of the crystals are small enough (e.g. the numbers of cells are small), one can observe the signals (non-Bargg peaks) between two Bragg peaks.

- The intensities of non-Bragg peaks provide additional information for phasing: oversampling becomes possible.

The result of simulation



Zhou et al. in preparation.

Nano-crystals for structure determination

- Much easier to obtain nano-crystals.

- Developing methods for:

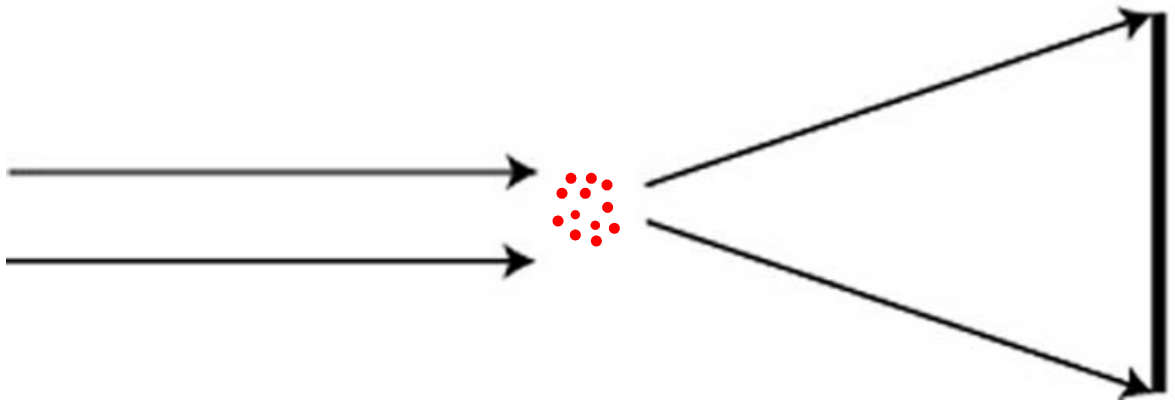
✓ Fewer patterns (2000 vs 20000)

✓ Anomalous dispersion for *de nuovo* structures

✓ Oversampling for *ab initio* phasing

Final solution: Coherent scattering

Coherent light source Molecules Detector



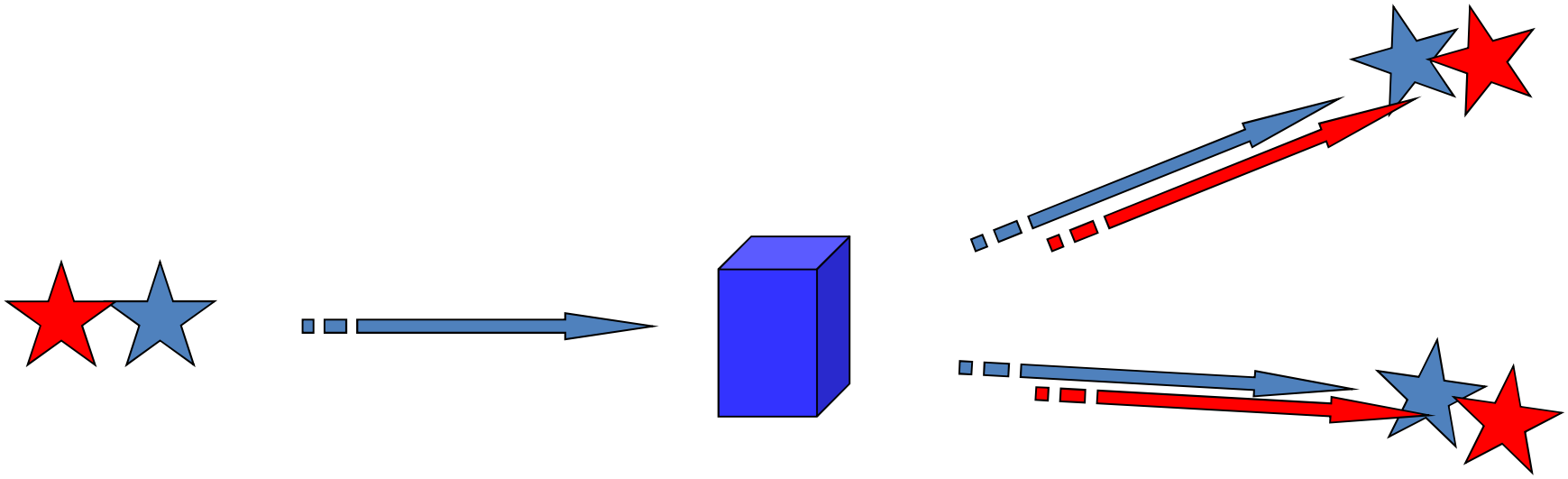
Crystals: Alignment of molecules

Alignment of photons: Coherent light (laser)

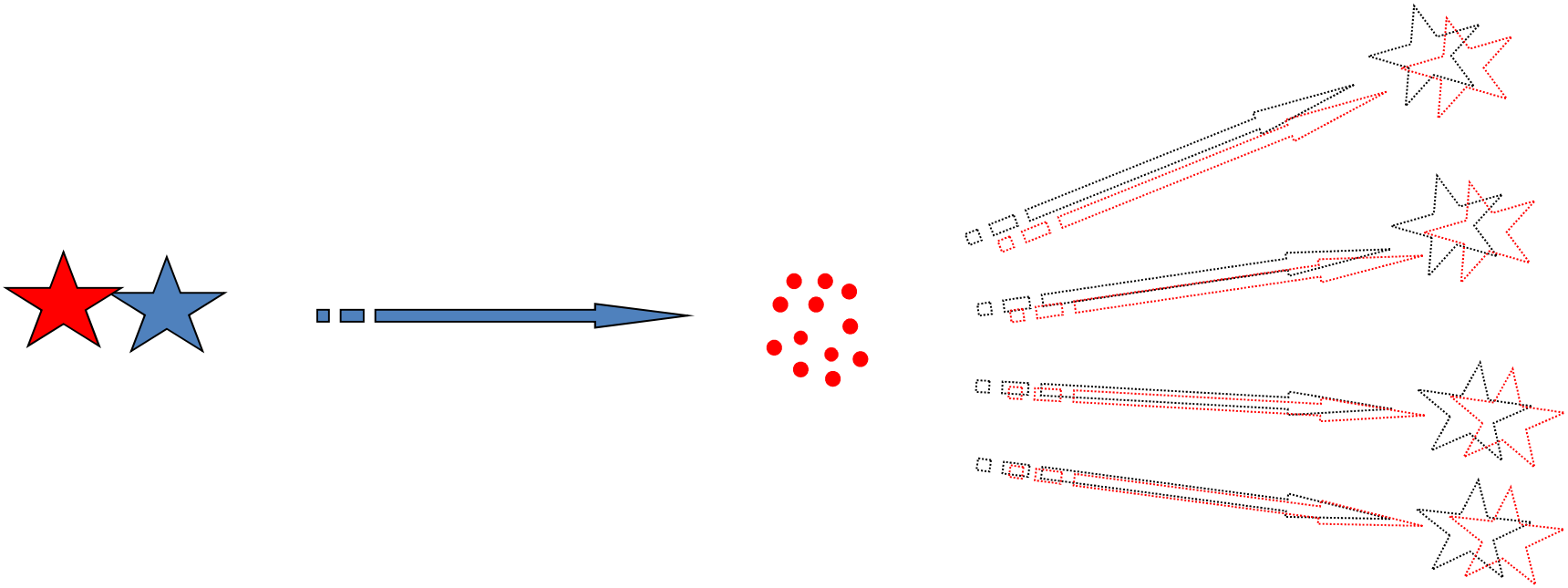
Coherent, high-intensity X-ray also can provide enough scattering signals from one molecule

$$\rho(\vec{r}) = \int \left|F(\vec{q})\right| \exp[i\phi(\vec{q})] \exp[2\pi i(\vec{r} \bullet \vec{q})] d^3\vec{q}$$

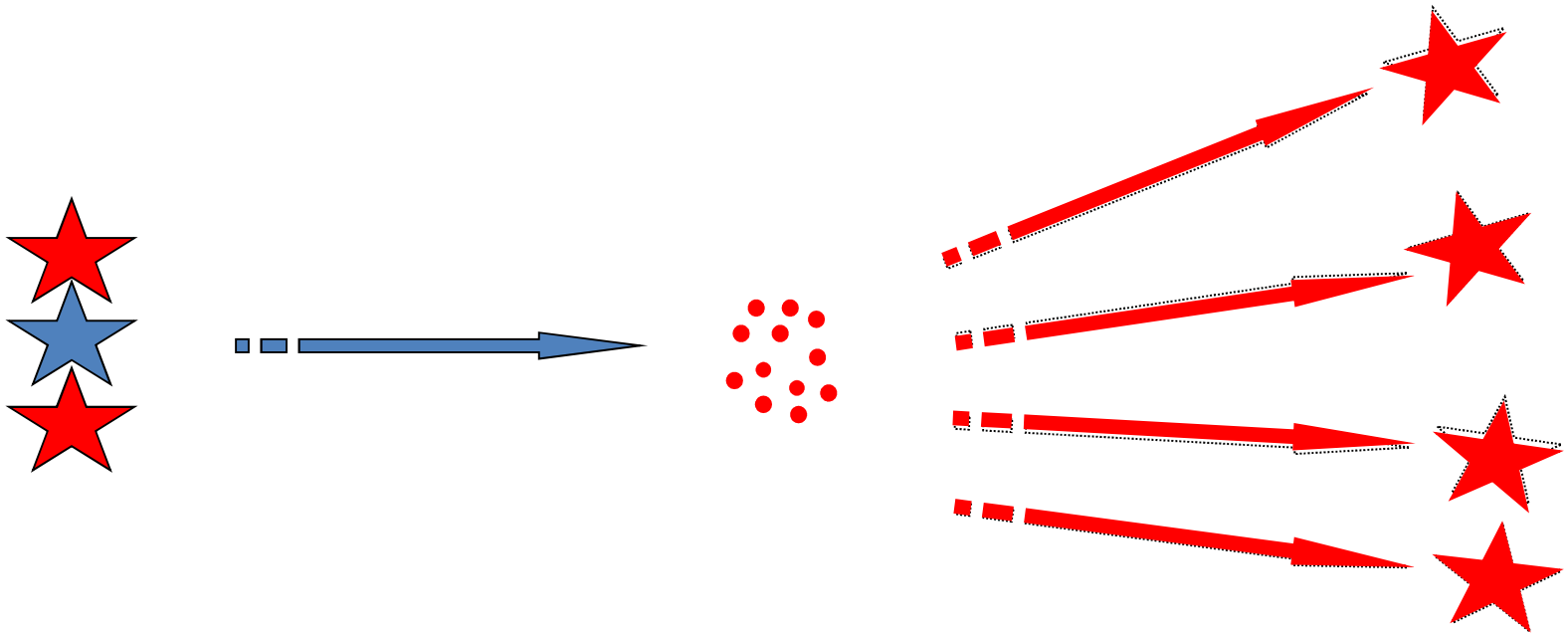
Pictures from Prof. Jianwei Maio, UCLA



The diffraction of a crystal: The periodic arrangement of molecules limits the directions of diffracted spots. Although the time of a photon arriving (phase) is random, the intensities of diffractions are strong enough to be detected.



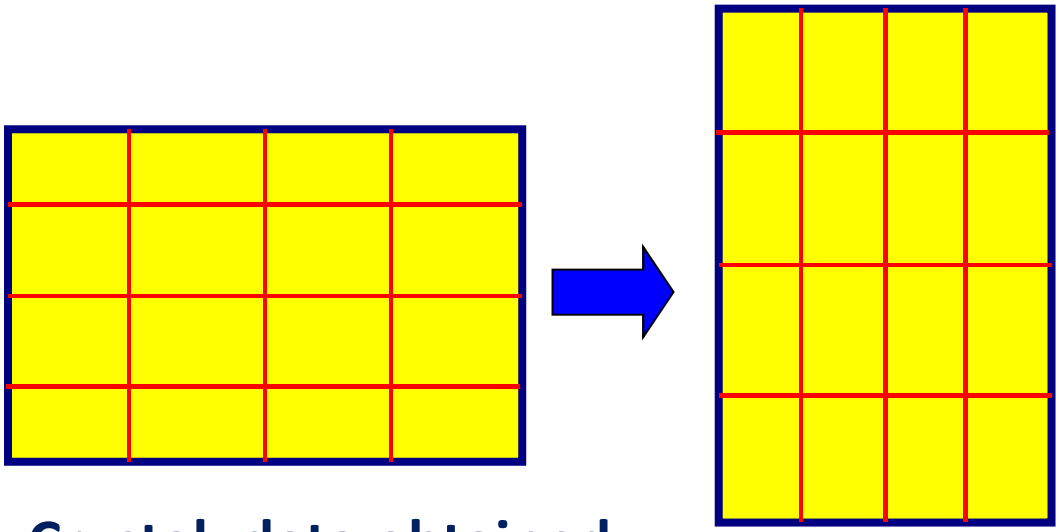
Incoherent photon scattering by a molecule: No periodicity, scattering in any directions. Without coherence, N photons produce N times of signals, still too weak to be detected.



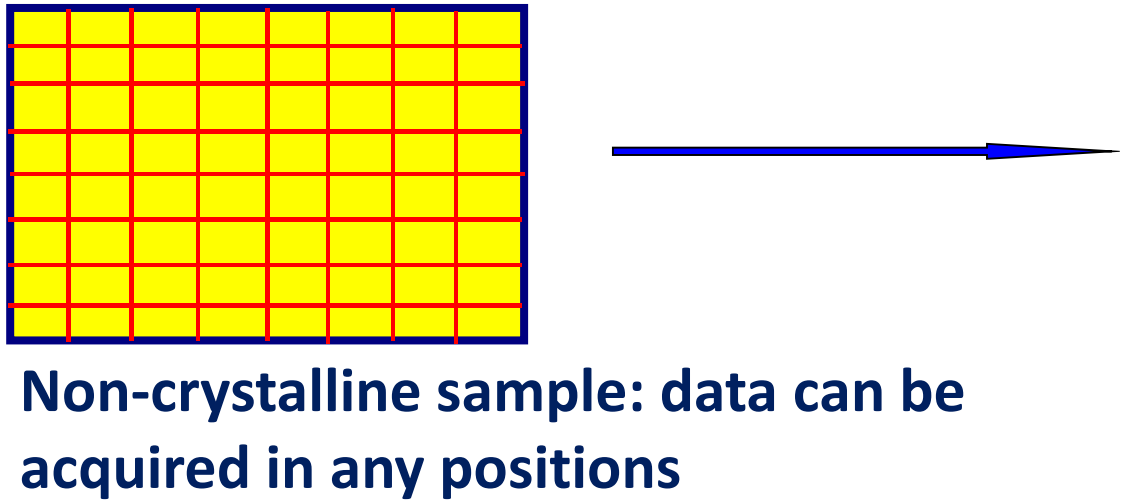
Unless the photons are coherent, N photons provide N^2 times of intensities.

Single molecule scattering

- Should be the final solution for structure determination
- Non-crystalline samples provide the possibility of oversampling

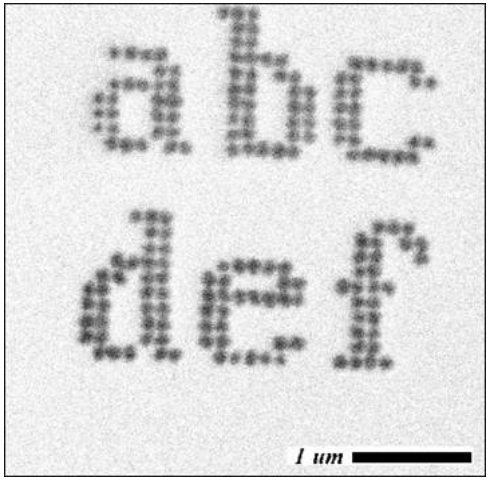


Crystal: data obtained only in some positions satisfied the Laue's law.

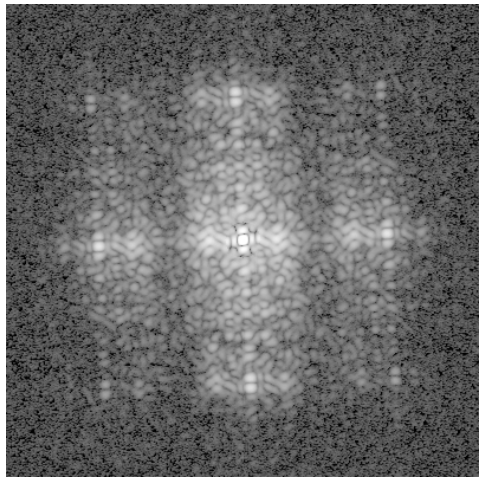


Non-crystalline sample: data can be acquired in any positions

Oversampling – Random phase + DM



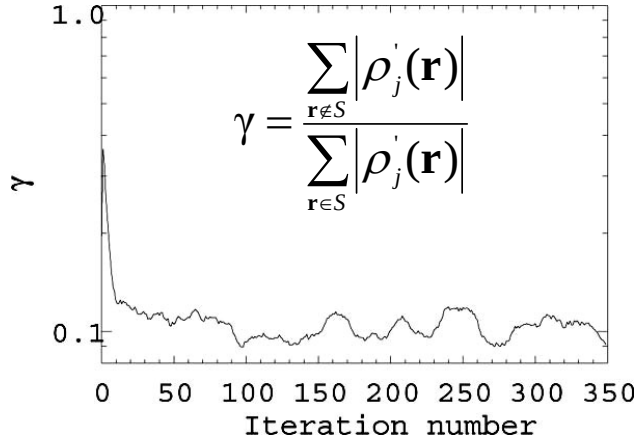
(a) A SEM image



(b) An oversampled diffraction pattern (in a logarithmic scale) from (a).

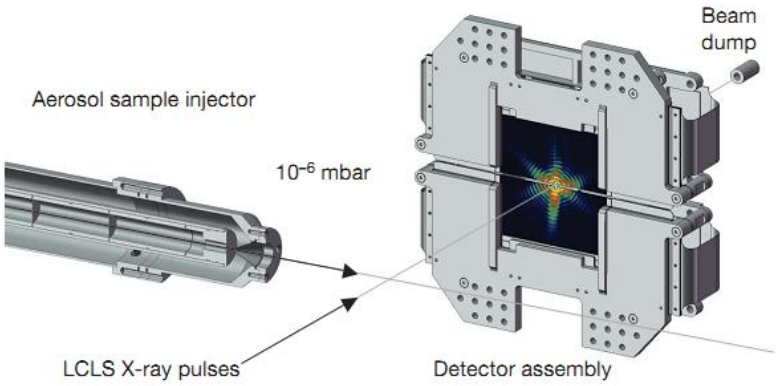
(c) An image reconstructed from (b).

Miao *et al.*, *Nature* 400, 342 (1999).



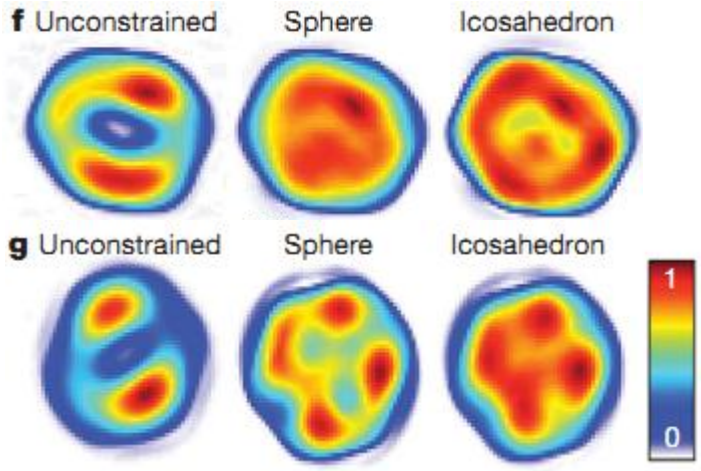
(d) The convergence of the reconstruction.

Single mimivirus particles

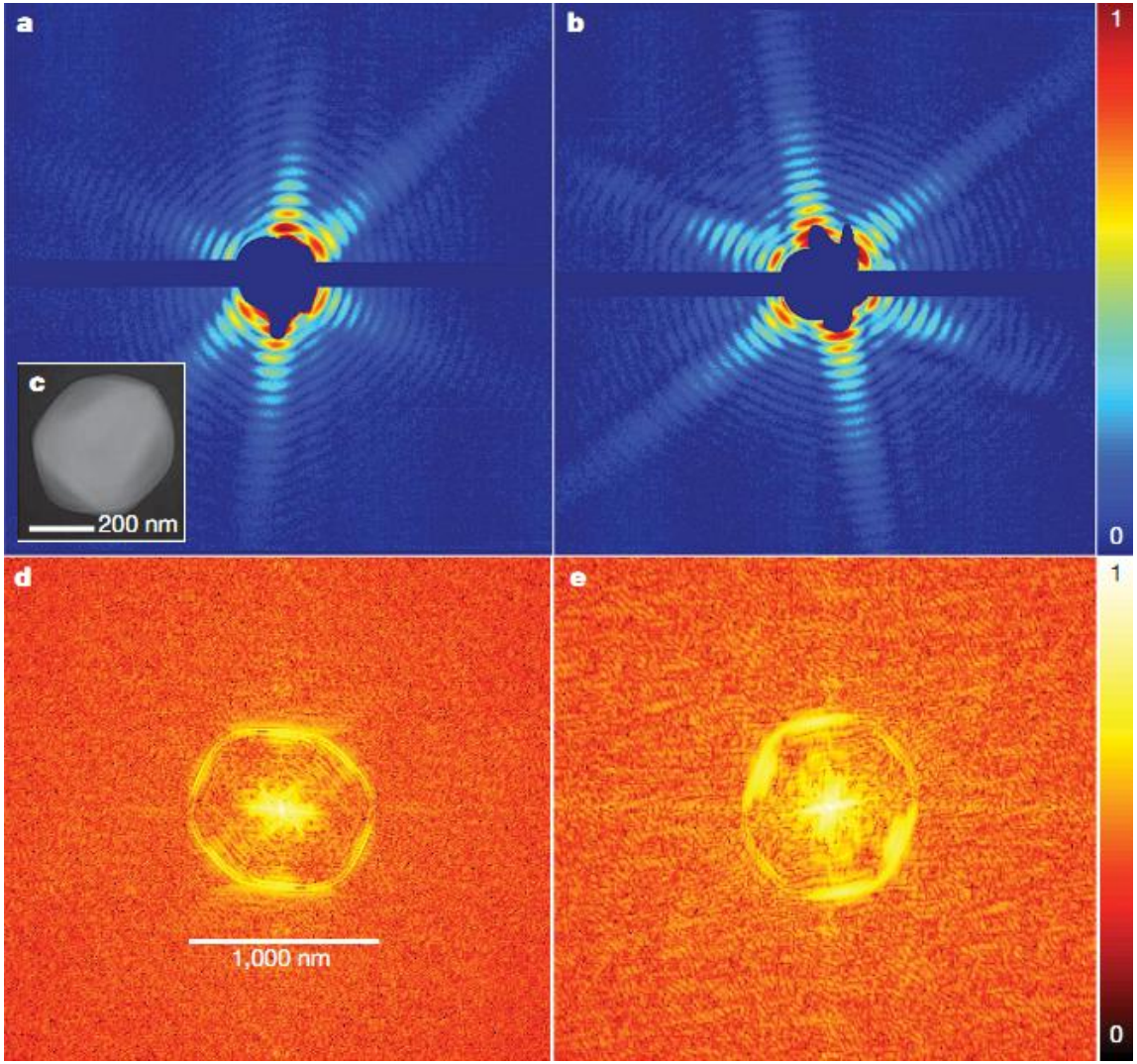


"Diffraction before Destruction" also works in the case of single particle.

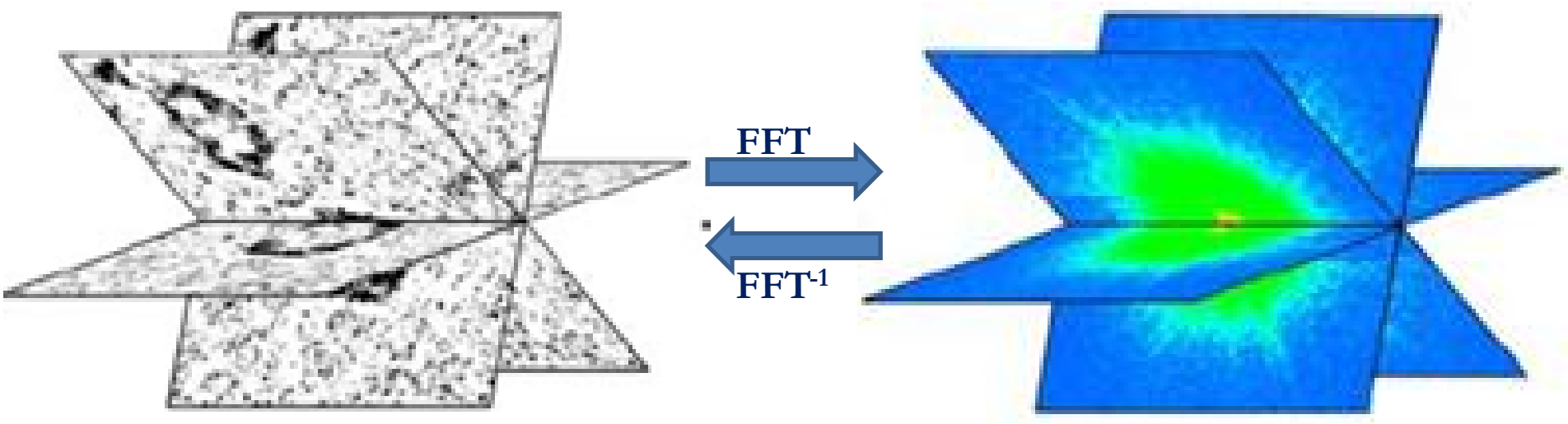
The experiment need to be improved, especially the "lost region" due to the beamstop.



M. Marvin Seibert *et al.* Nature 470, 78 (2011)



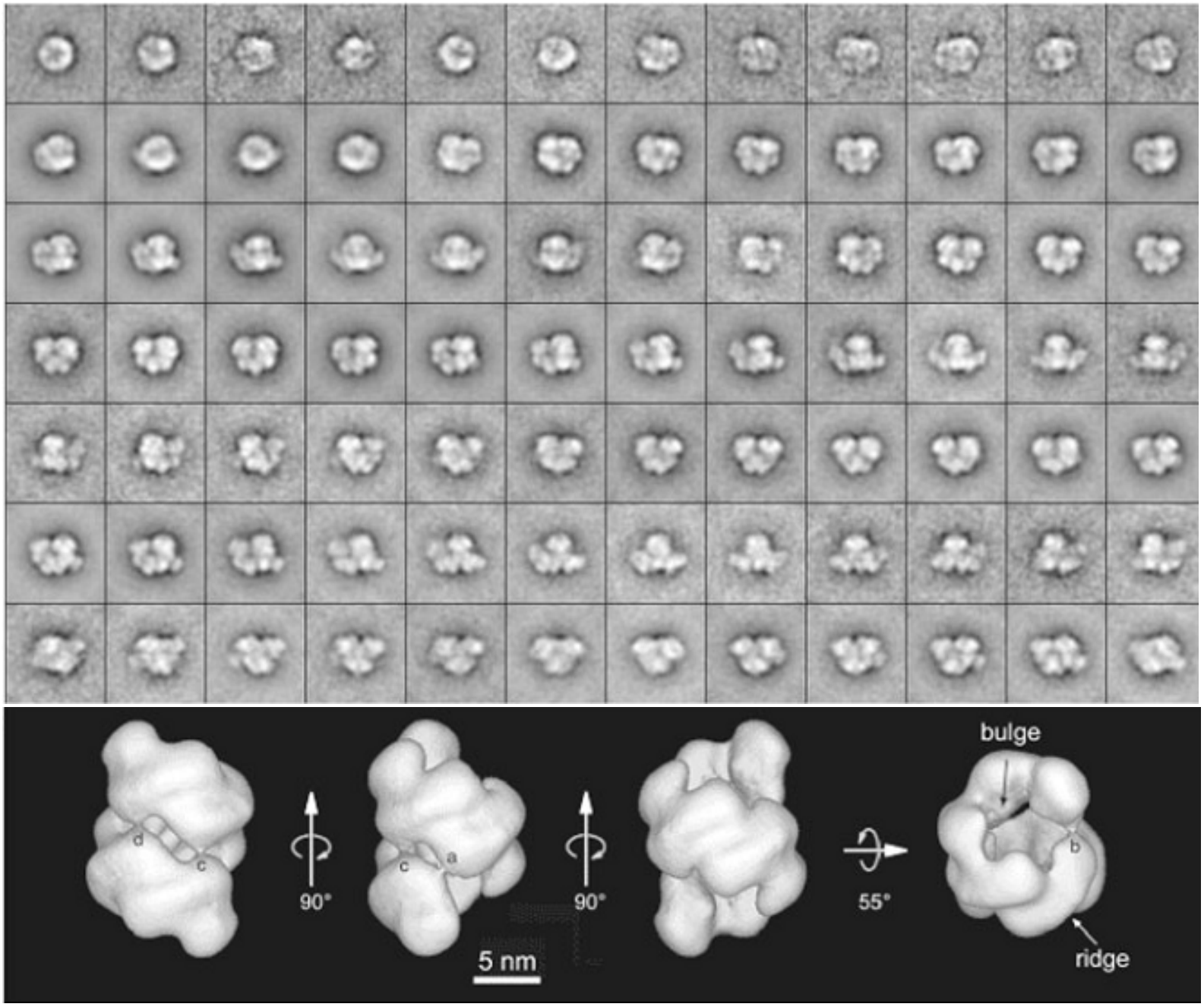
Reconstruction the 3D structure



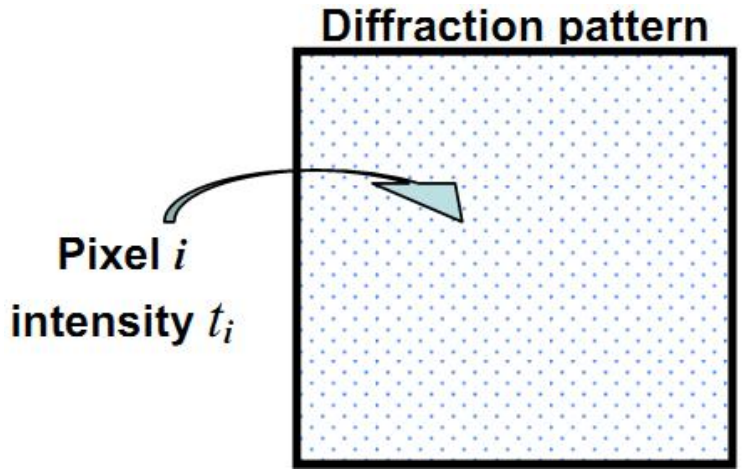
- From a series of 2D projections, the 3D structure of a particle can be reconstructed.
- The orientation of individual projection need to be known before.
- The technique is similar to the "Three-Dimensional Reconstruction" of cryo-EM.

Picture from Jianwei Miao, UCLA

An example of cryo-EM

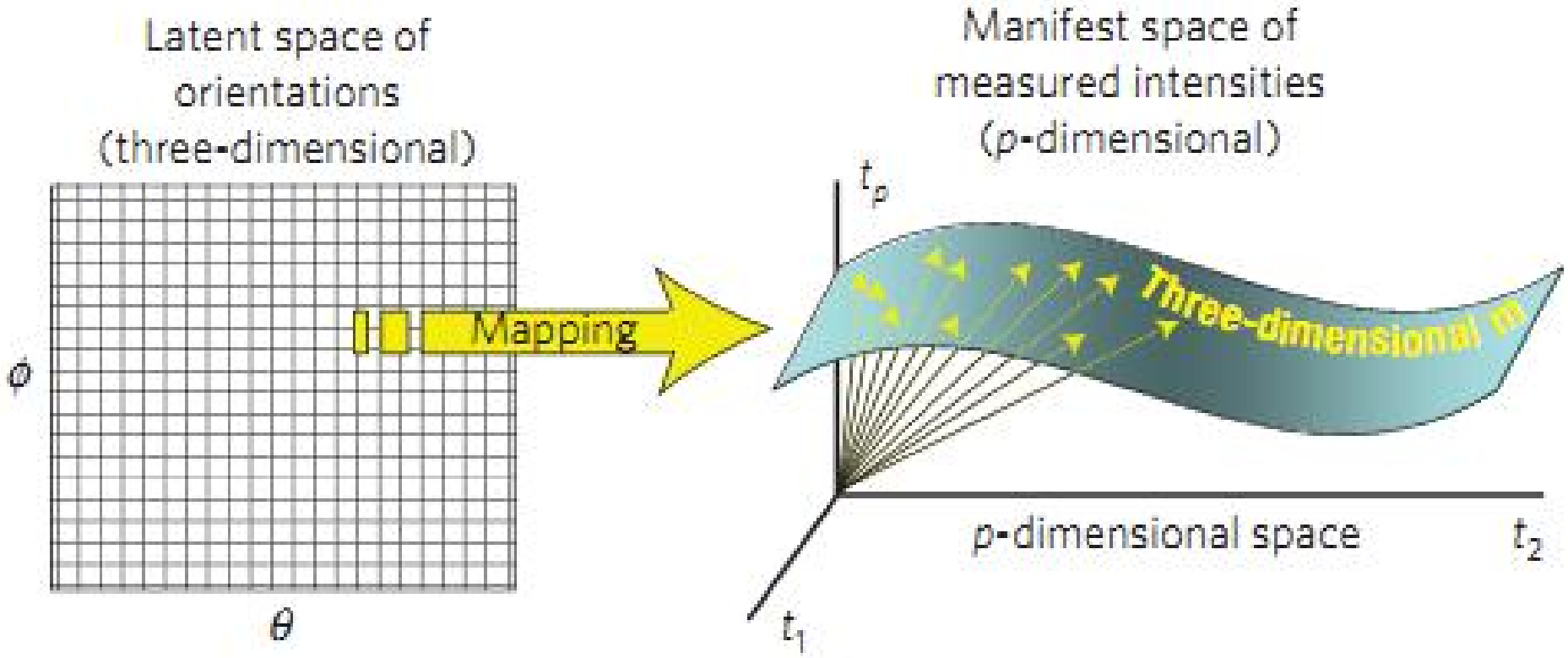


Classification method

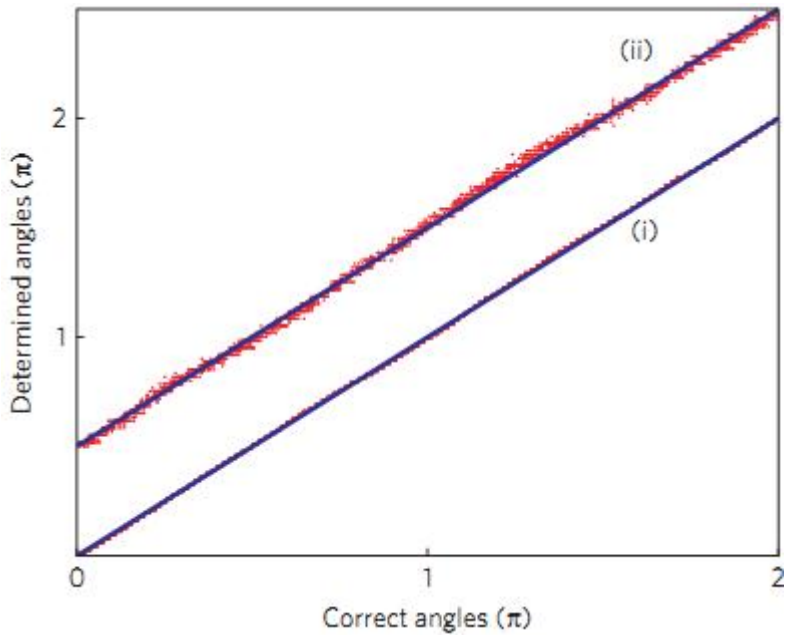


Diffraction pattern represented by vector:

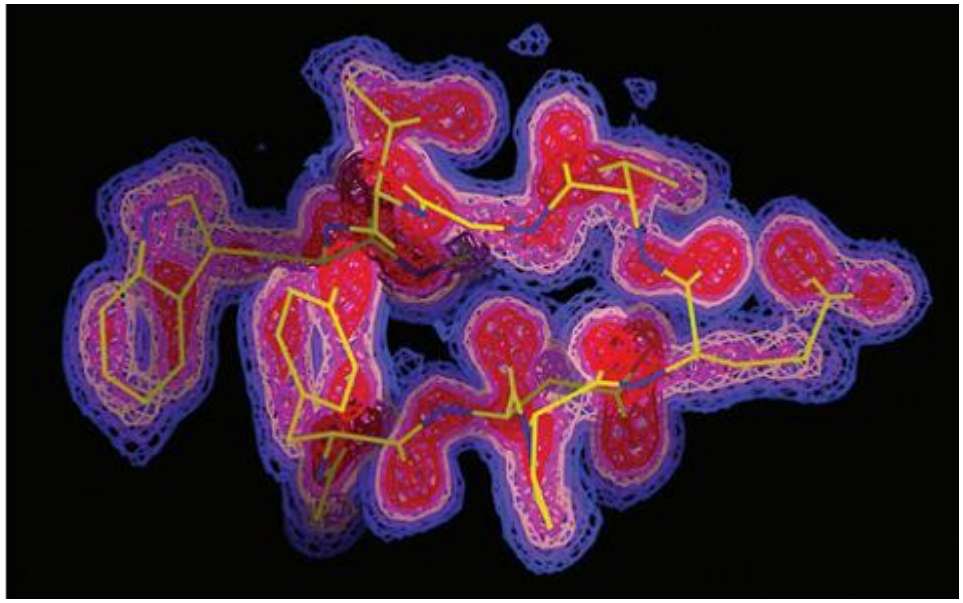
$$\mathbf{t} = \left(t_1, \ldots, t_p \right)$$



Simulation for the classification method



The orientation recovery

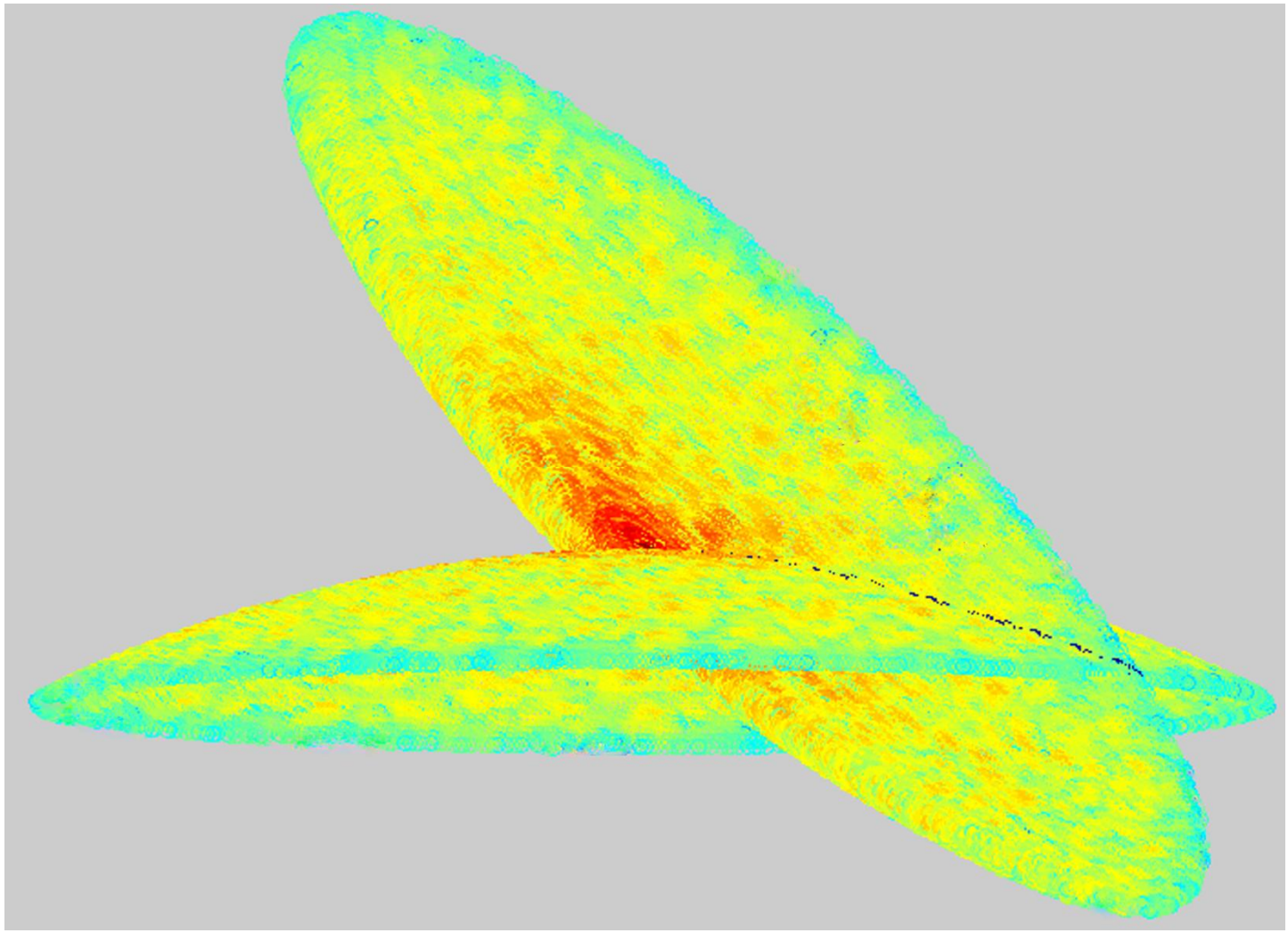


The structure recovery

- The number of the particles are huge: 72,000
- The r.m.s.d of the orientation is still large: 3.8 degree
- Higher resolution, more particles: e.g. 1.8Å, 10^6 diffraction patterns.

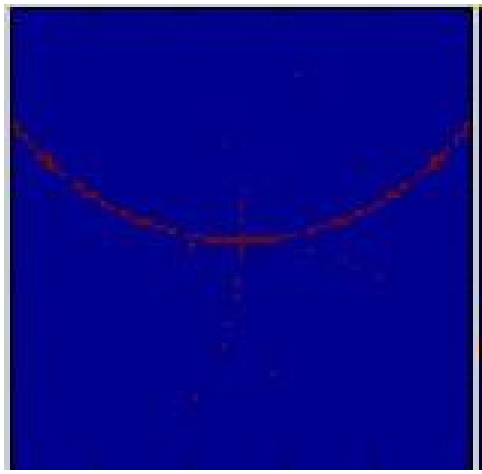
Russell Fung et al. Nature Physics, 5, 64, 2009

Common-line method

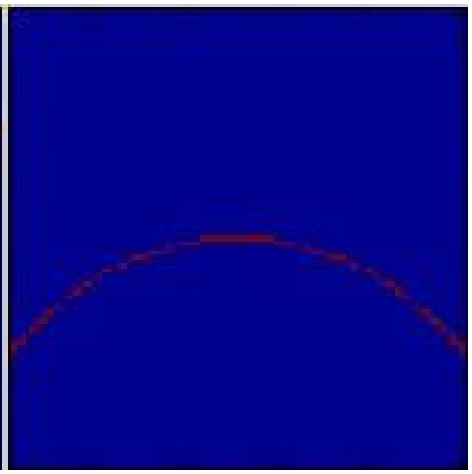


High resolution: the curvature of the Edwald sphere.

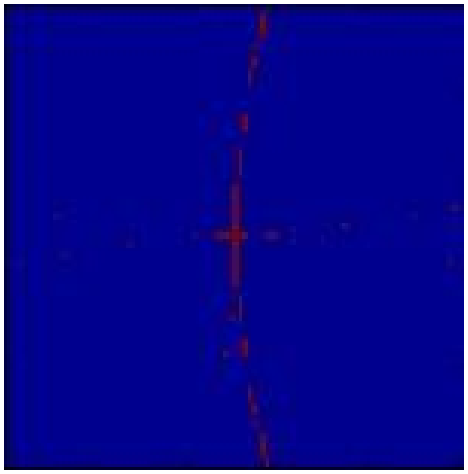
The projections of common lines on detector



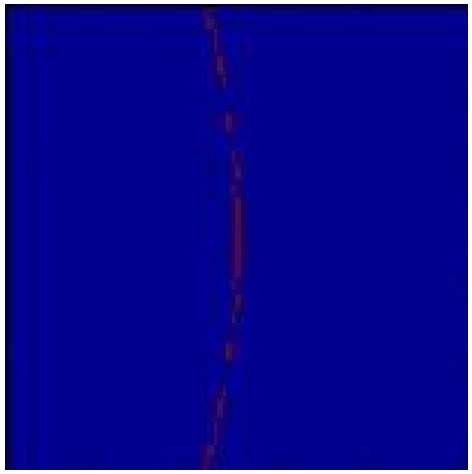
(0, 0, 0)



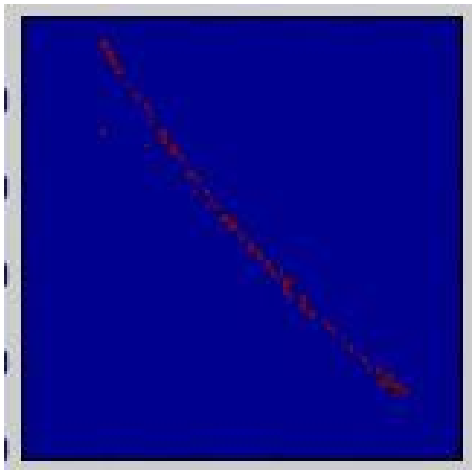
(120, 0, 0)



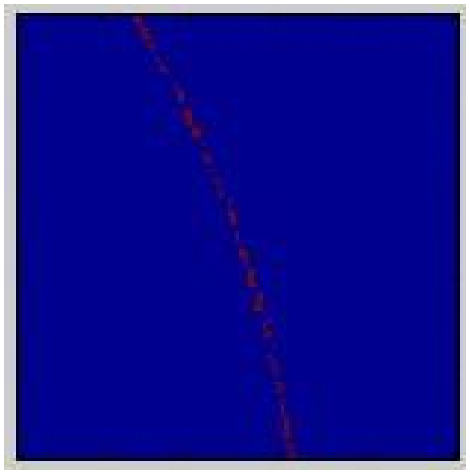
(0, 0, 0)



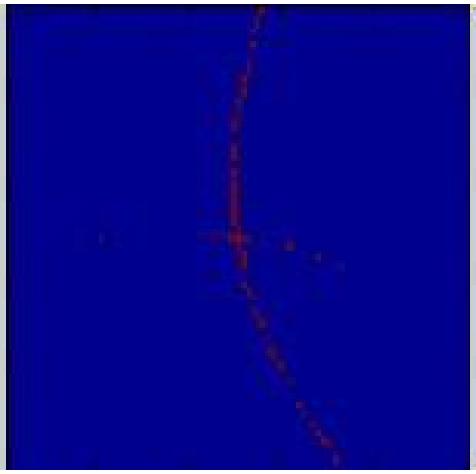
(0, 60, 0)



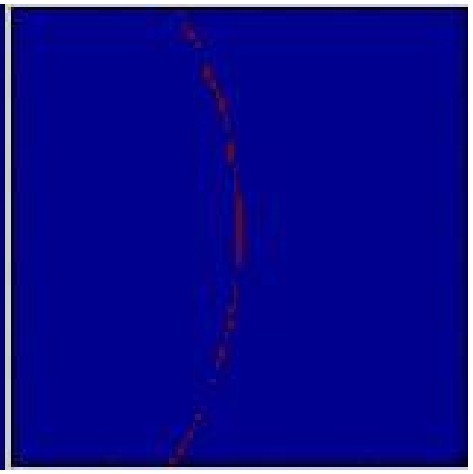
(0, 0, 0)



(30, 30, 30)



(0, 0, 0)



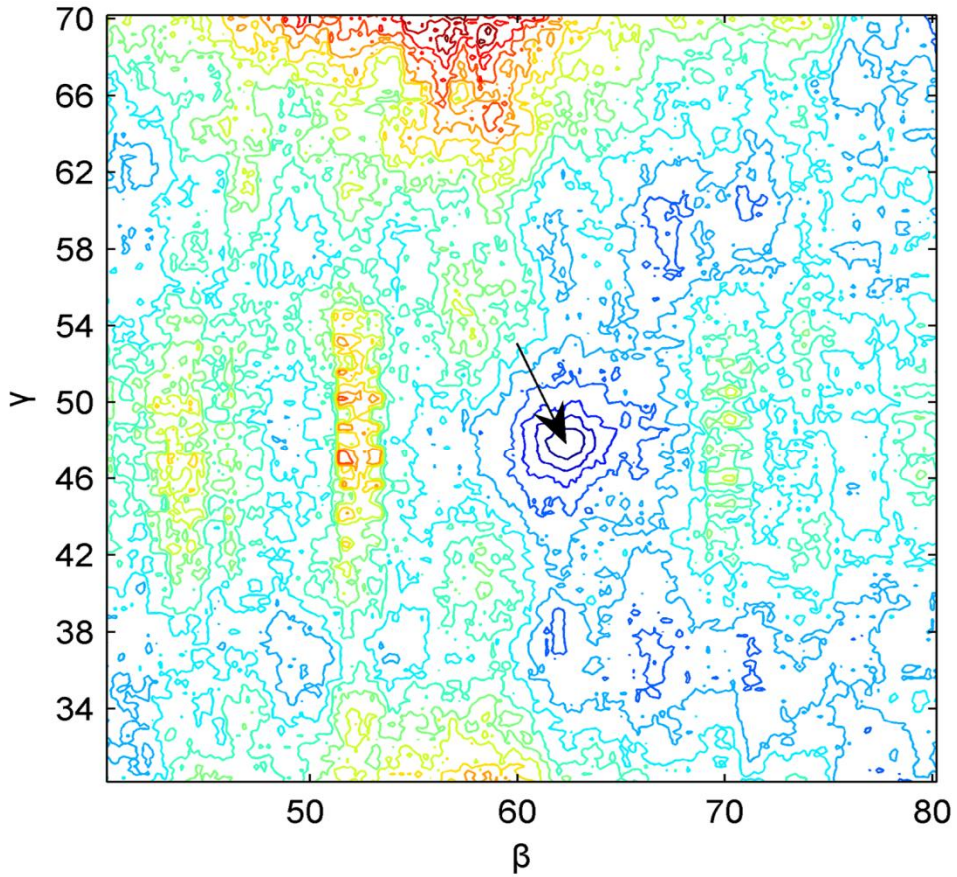
(120, 80, 120)

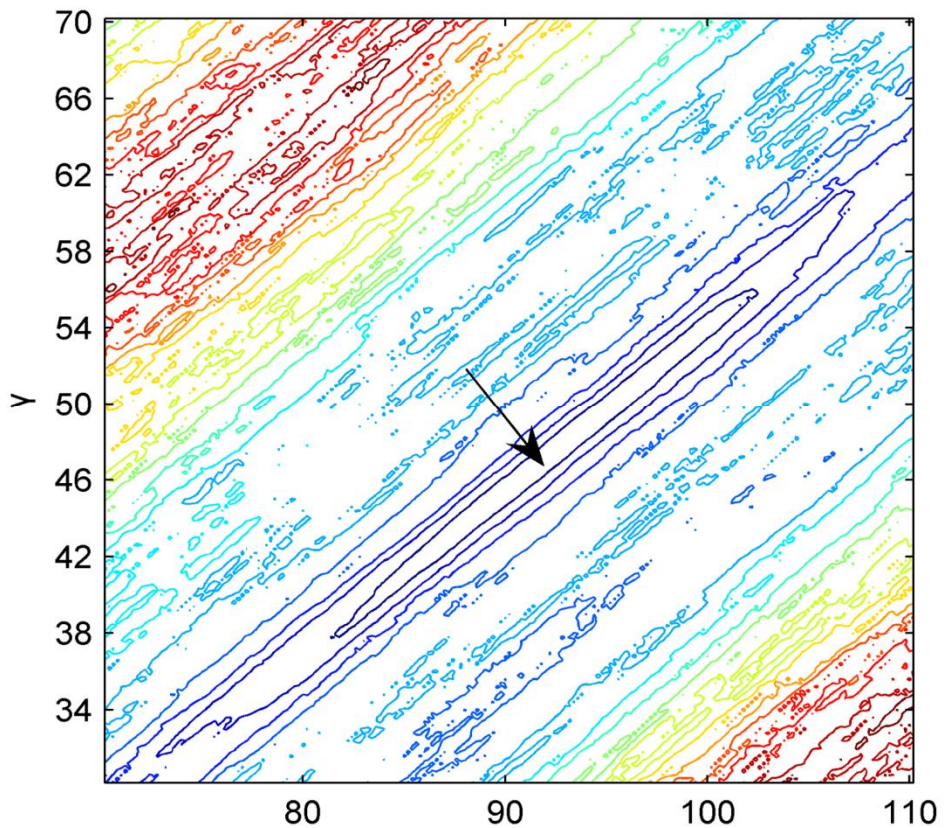
Single-common-line method

- Select one pattern as reference, calculating the difference between two patterns at the common line.
- The intensities can be scaled via the intensities at common lines.

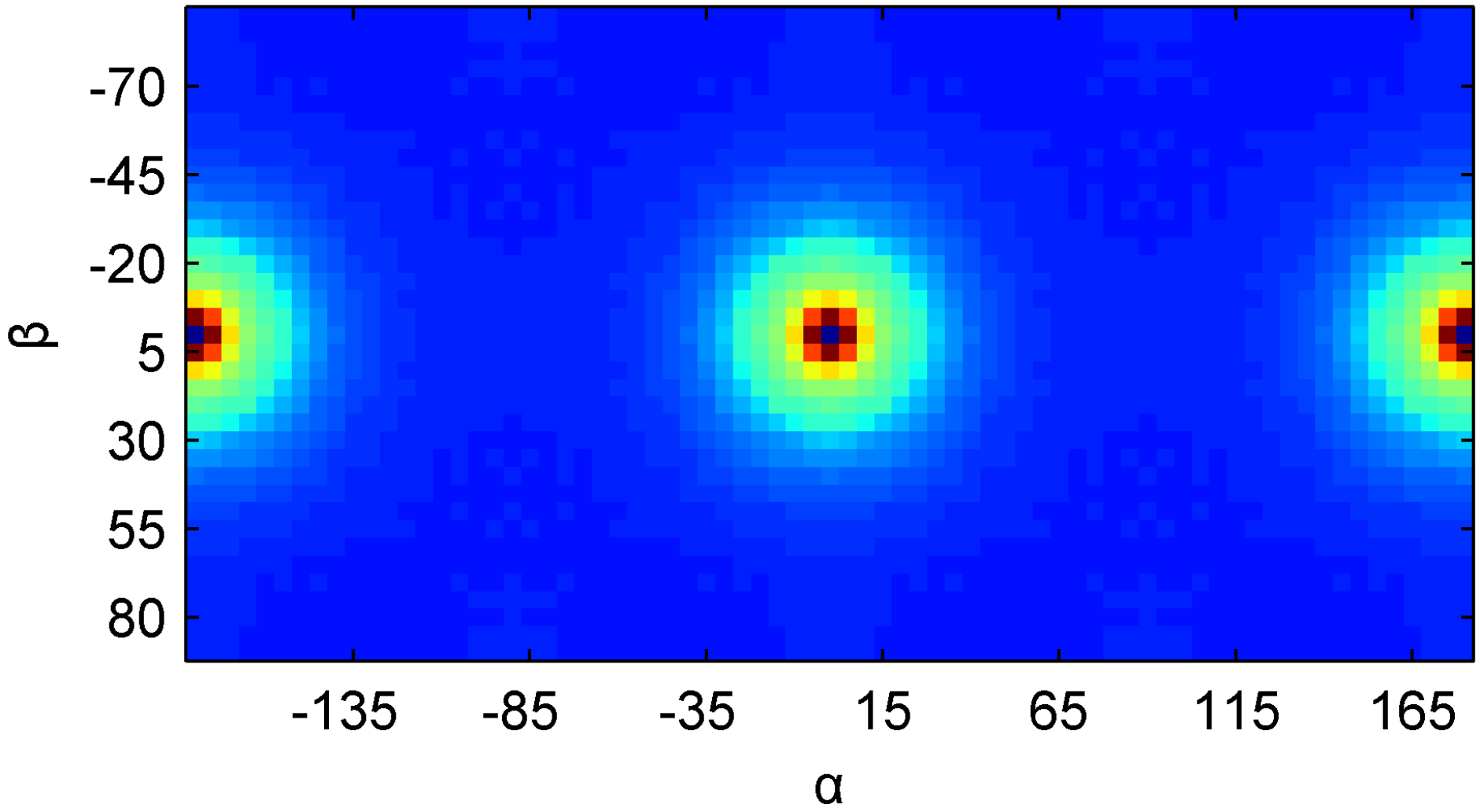
$$Rfactor(\alpha,\beta,\gamma,n)=\frac{\sum_{hkl}\frac{\left\|F_{hkl}(\alpha,\beta,\gamma,n)\right|-\left|F_{hkl}(0,0,0,1)\right\|}{|F_{hkl}(0,0,0,1)|}}{\sum_{hkl}1}$$

The result of single-common-line method





The number of data points used in single-common-line method

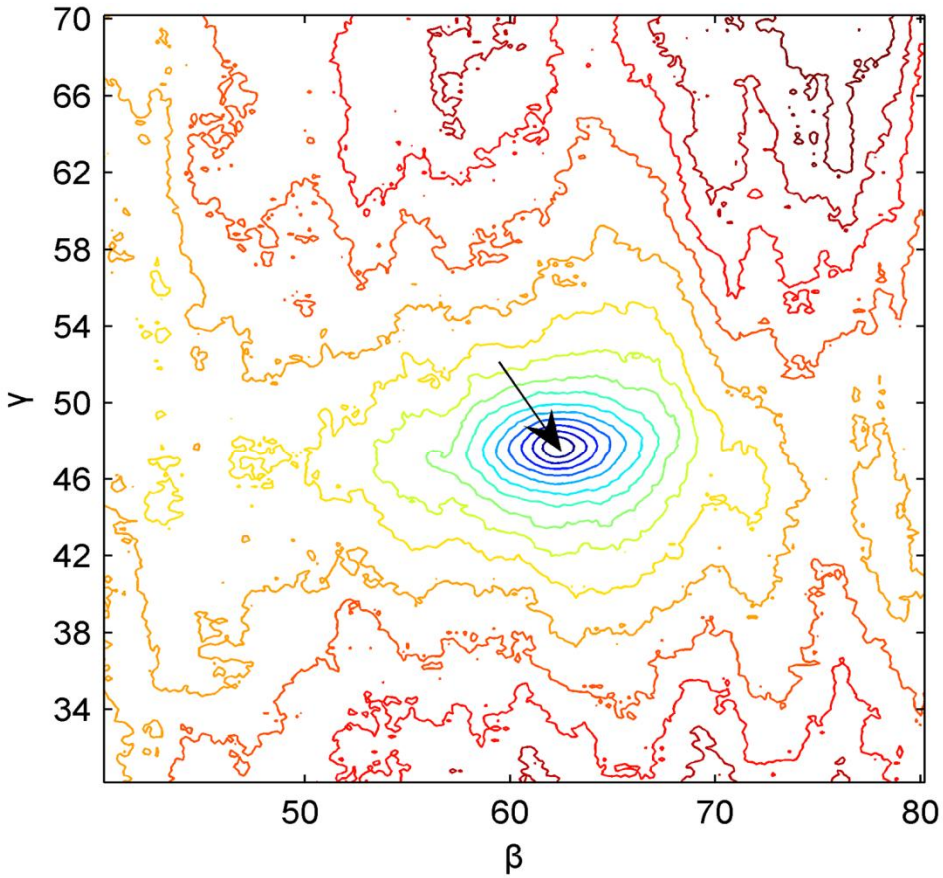


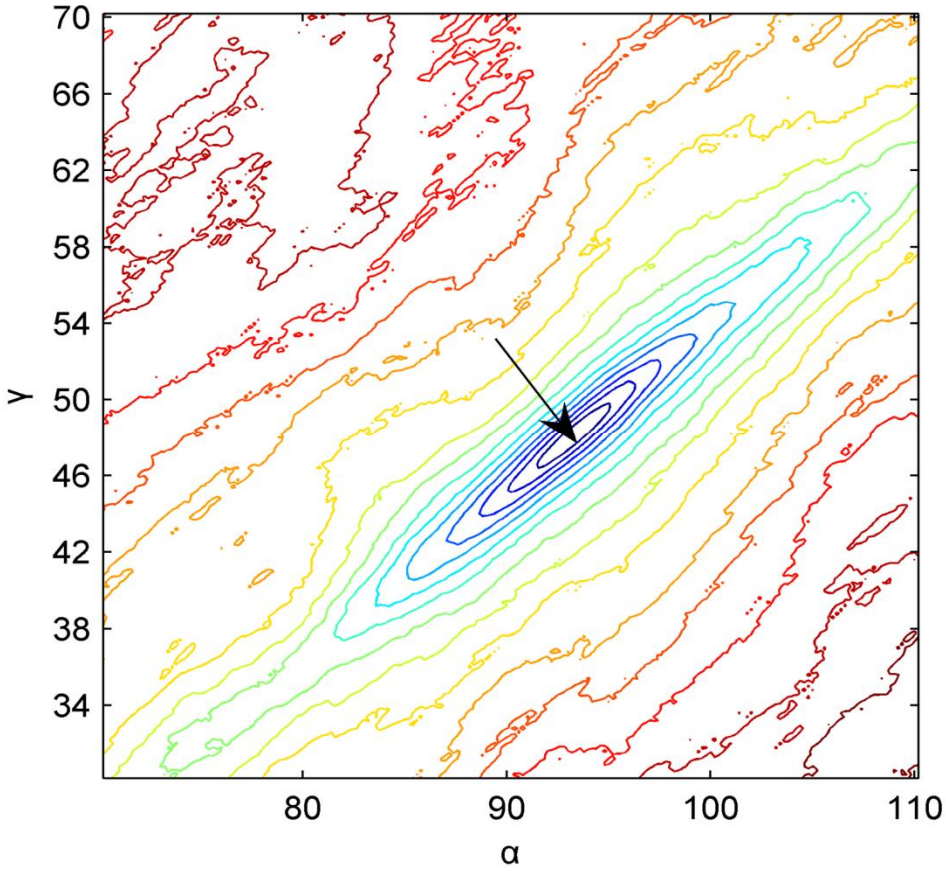
Multi-common-line

- **Use all of the other patterns as references, refining the orientation**
- **Iteration**

$$multi-Rfactor(\alpha,\beta,\gamma,n)=\frac{\sum_{hkl}\frac{\left|\left|F_{hkl}(\alpha,\beta,\gamma,n)\right|-\left\langle\left|F_{hkl}\right|\right\rangle\right|}{\left\langle\left|F_{hkl}\right|\right\rangle}}{\sum_{hkl}1}$$

Result of multi-common-line method

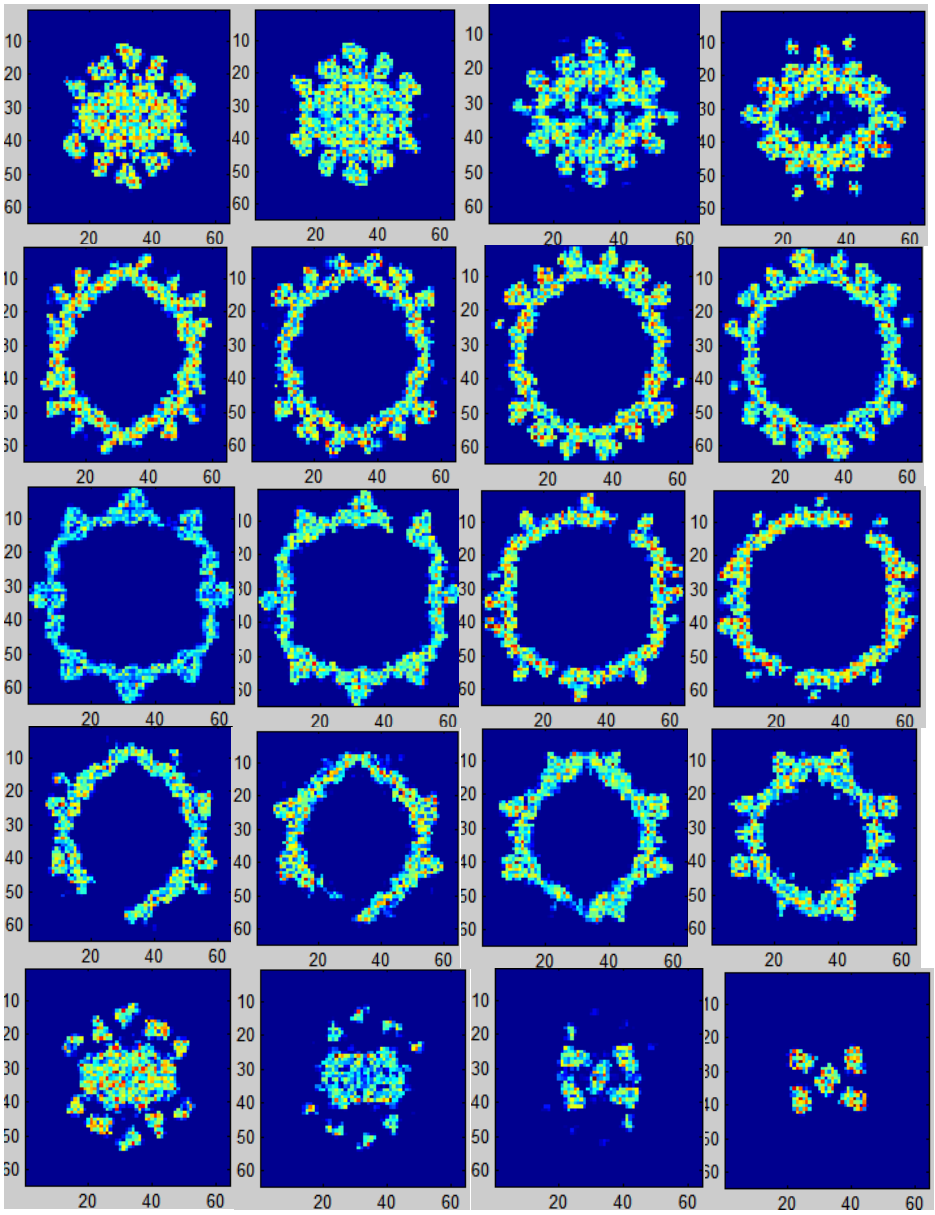


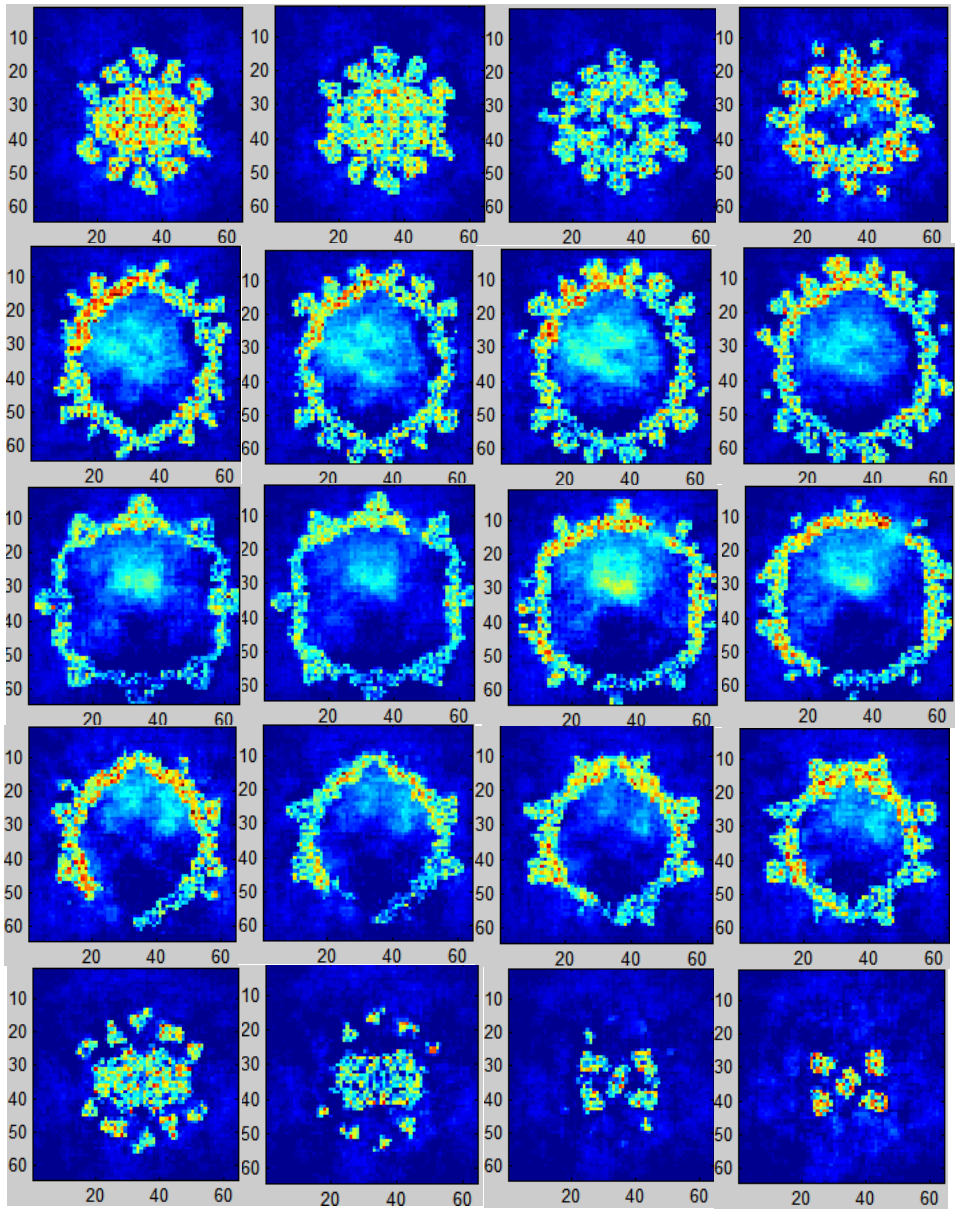


Determining the orientations

- ✓A method similar to the molecular replacement method in PX
- ✓Only ~10^4 particles are needed
- ✓The r.m.s.d. of orientations are 0.34°, 0.14° and 0.19° respectively
- ✓10^{13} photons per pulse, near atomic resolution for 5MDa virus (diameter 345Å)

patterns	α (actual/recovered)	β (actual/recovered)	γ (actual/recovered)	Nr. of grids used
2nd	18.6 / 18.8	21.2 / 21.2	67.6 / 67.8	1668
3rd	18.4 / 18.4	-34.8 / -34.8	-32.2 / -32.4	1110
4th	-20.8 / -20.8	-28.2 / -28.2	122.8 / 122.8	1238
5th	-31.7 / -31.4	-27.7 / -27.6	97.1 / 97.0	1056
6th	8.8 / 9.2	12.4 / 12.4	-98.3 / -98.6	2456
7th	-116.6 / -117.2	-34.4 / -34.2	144.4 / 144.6	662
8th	-117.5 / -117.6	-46.2 / -46.2	-14.3 / -14.2	669





Actual structure
2tbv, 5MDa

Recovery structure

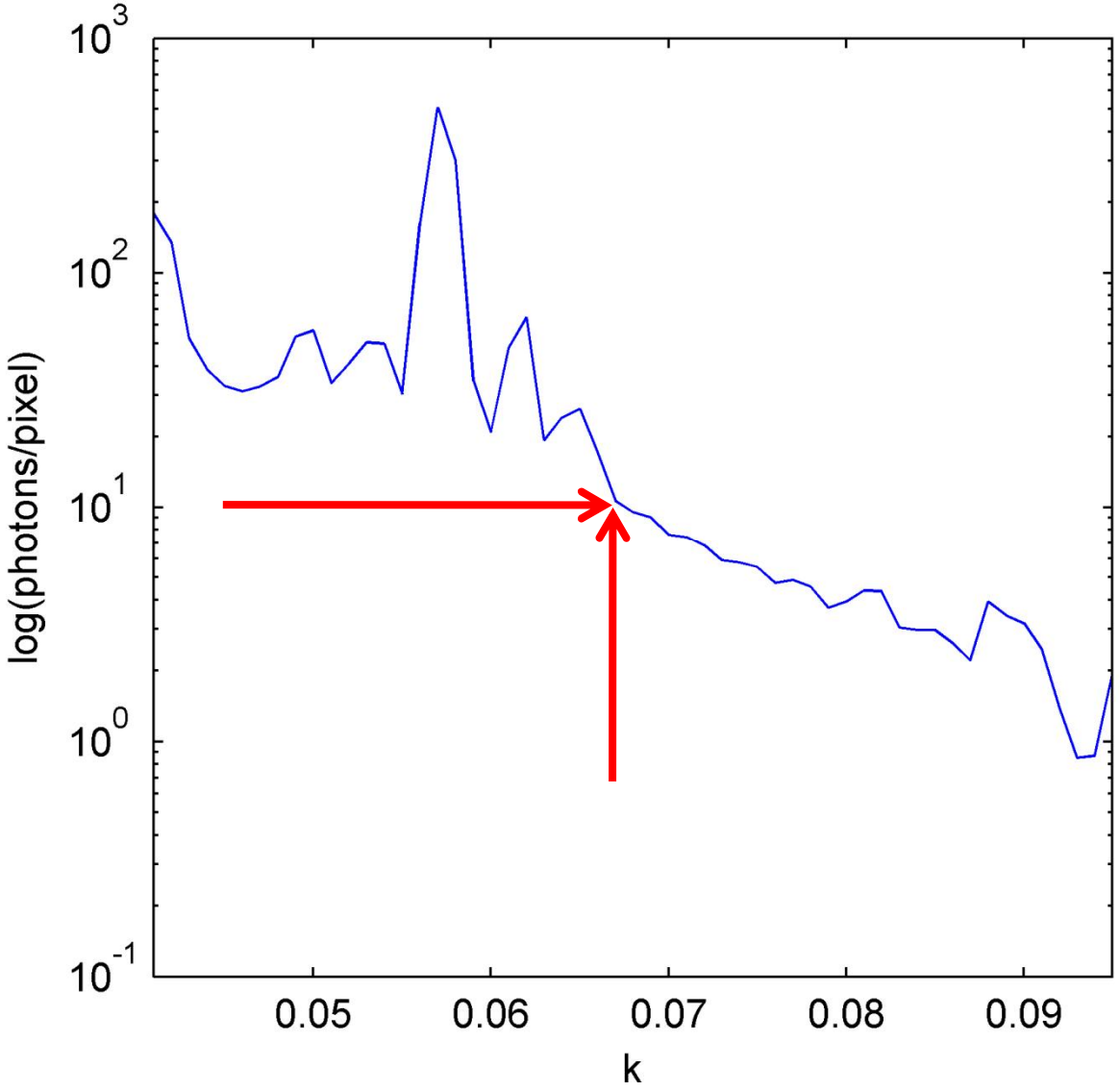
Zhou et al. in preparation.

Still long way to go …

- The duration of the pulse need to be shorter: at least 10fs, better 2fs.

- Not enough coherent photons: now 10^{12}phs/pulse (2mJ@8.3keV) can be archived at LCLS. But 10^{13}phs/pulse (20mJ@8.3keV) is required for molecule of 5MDa.
- Strong enough for destruction but not for signals!

Simulation for single particle scattering



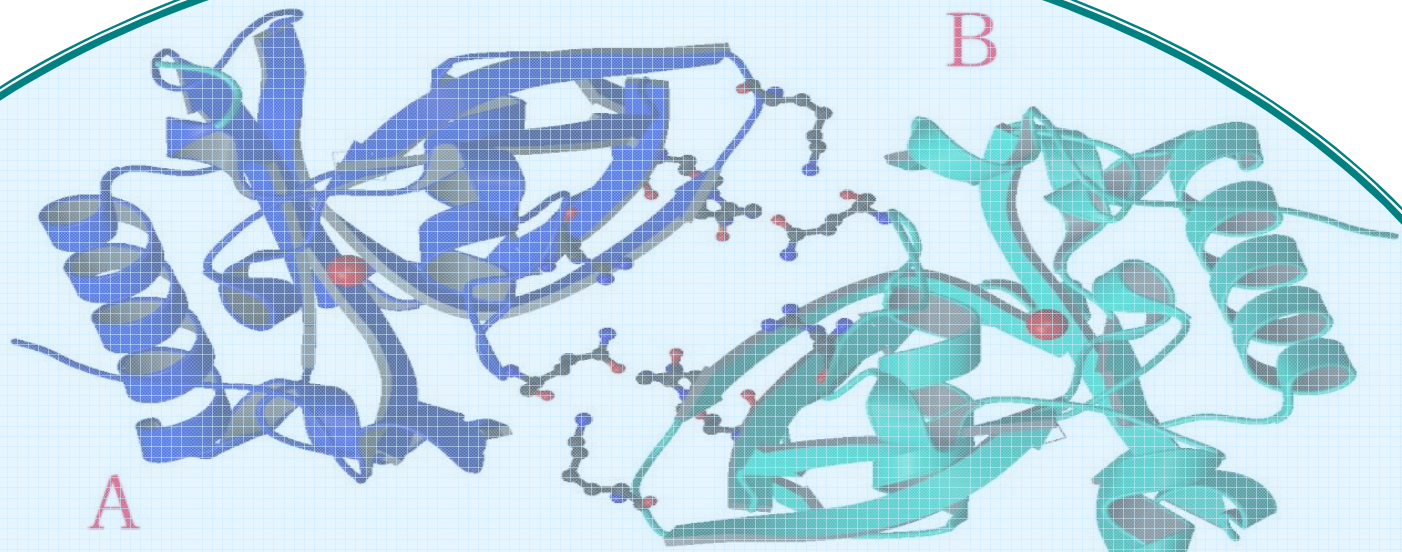
Virus particle, 5MDa, 34.5nm

λ=5.0 Å, 5×10^{13} photons per pulse focused to 100 nm

10 counts at 14.7 Å resolution

Summary

- ✓ Coherent X-ray source: low emittance accelerator-based source
- ✓ X-FEL, or diffraction limit storage ring/ERL
- ✓ Can we obtain the scattering signal?
- ✓ Yes, "Diffraction before Destruction" for nano-crystals or single particles.
- ✓ How to solve the phase?
- ✓ MR, AD, oversampling are all possible
- ✓ How to reconstruct the 3D structures?
- ➢ Similar to cryo-EM, need orientations before.
- ➢ Classification/Common-line
- ➢ Accurate structure factors!
- ➢ Coherent X-ray open a new area for structure biology: small crystals, structure without crystal



Thank you for your attention.