AN INTRODUCTION TO MACROMOLECULAR CRYSTALLOGRAPHY

Seán McSweeney

October 7, 2015

Structural Biology
NSLS-II
BNL.
smcsweeney@bnl.gov

MACROMOLECULES

In macromolecular crystallography we investigate the form and function of *Biological* molecules and their complexes in order to understand the function of biological systems.

We bring together, Physics, Chemistry, Engineering and robotics in order to allow the study of the structure of biological molecules.

Its all in the genes

`tctgaagctaattaaatgtgctcttaqtctcctgaagcaaagtct`→
SEAN*MCS*SPEAKS

DNA codons are converted in to *amino-acids*, in one dimension these are polypeptide chains. Then life gets more complicated. [1]

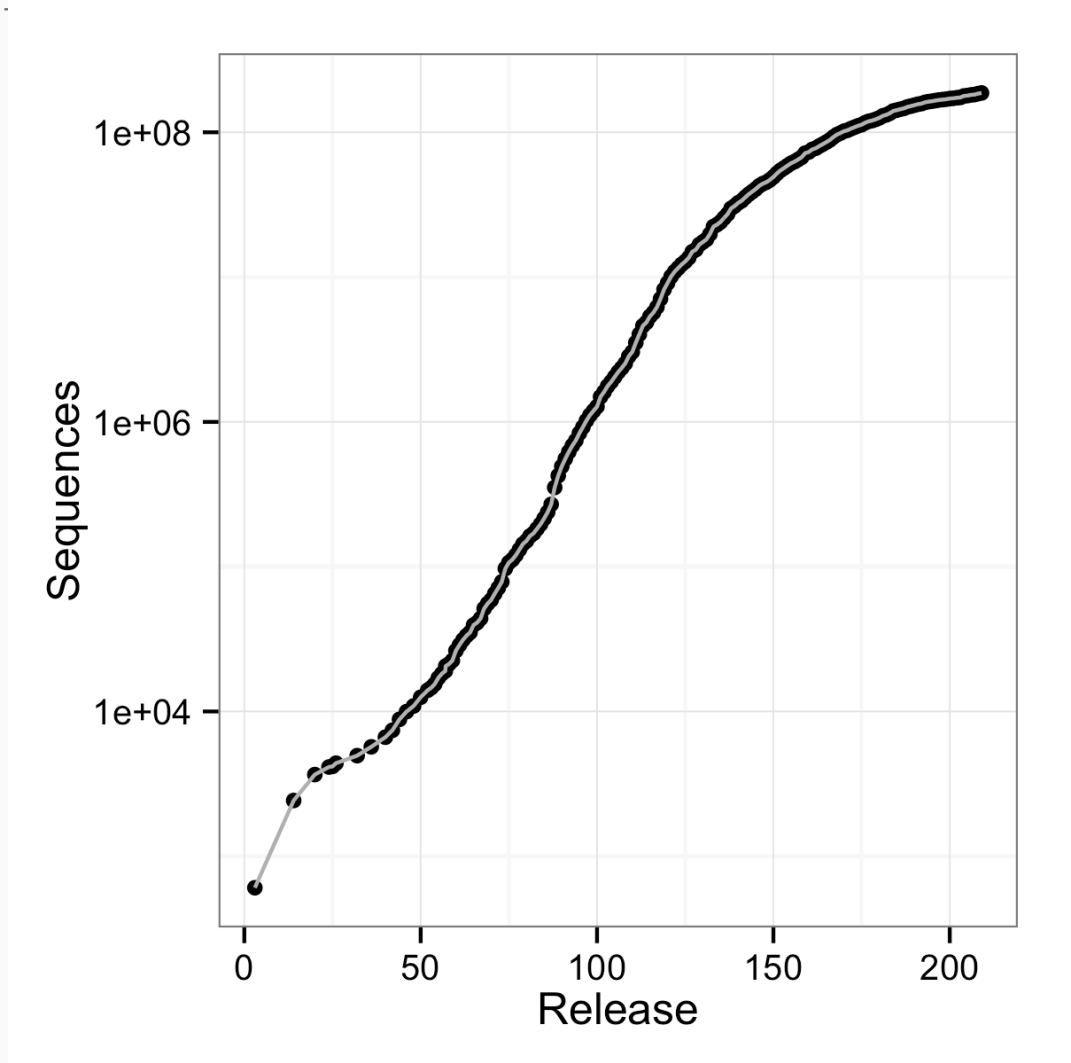
[1] The conversion steps and the error checking and the regulation...are done with macromolecular machines.

WHY ARE MACROMOLECULES IMPORTANT?

We are all ensembles of macromolecules.

Every living thing, in sickness and in health, are collections of functioning and mis-functioning macromolecules. Revealing the structure and higher order organization of these complexes will help understand the organism.

GROWTH OF GENE DATA

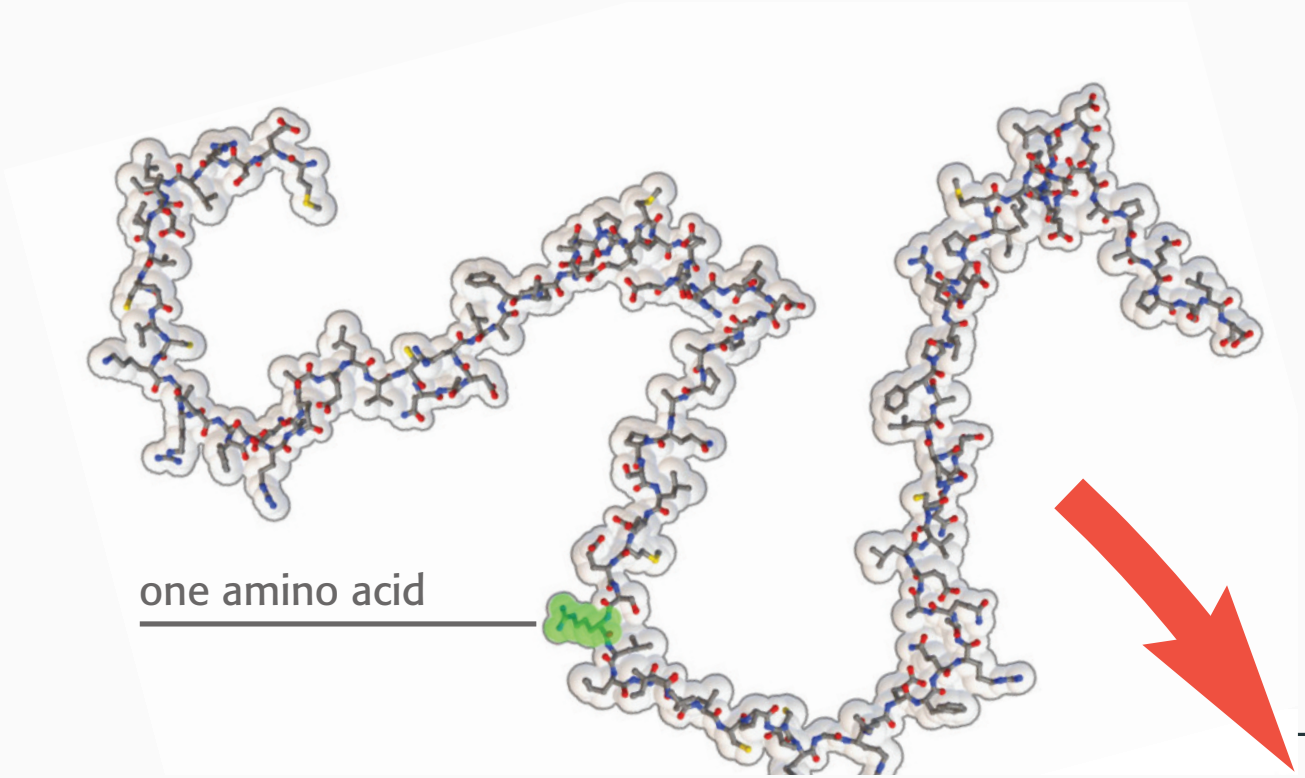


- The amount of publicly available gene data has doubled about every 18 months since 1982.
- 1982 : 606 sequences
- 2015: 187,066,846 sequences

[2]

[2] Data from `http://www.ncbi.nlm.nih.gov/genbank/statistics`

THE STRUCTURE OF PROTEINS.

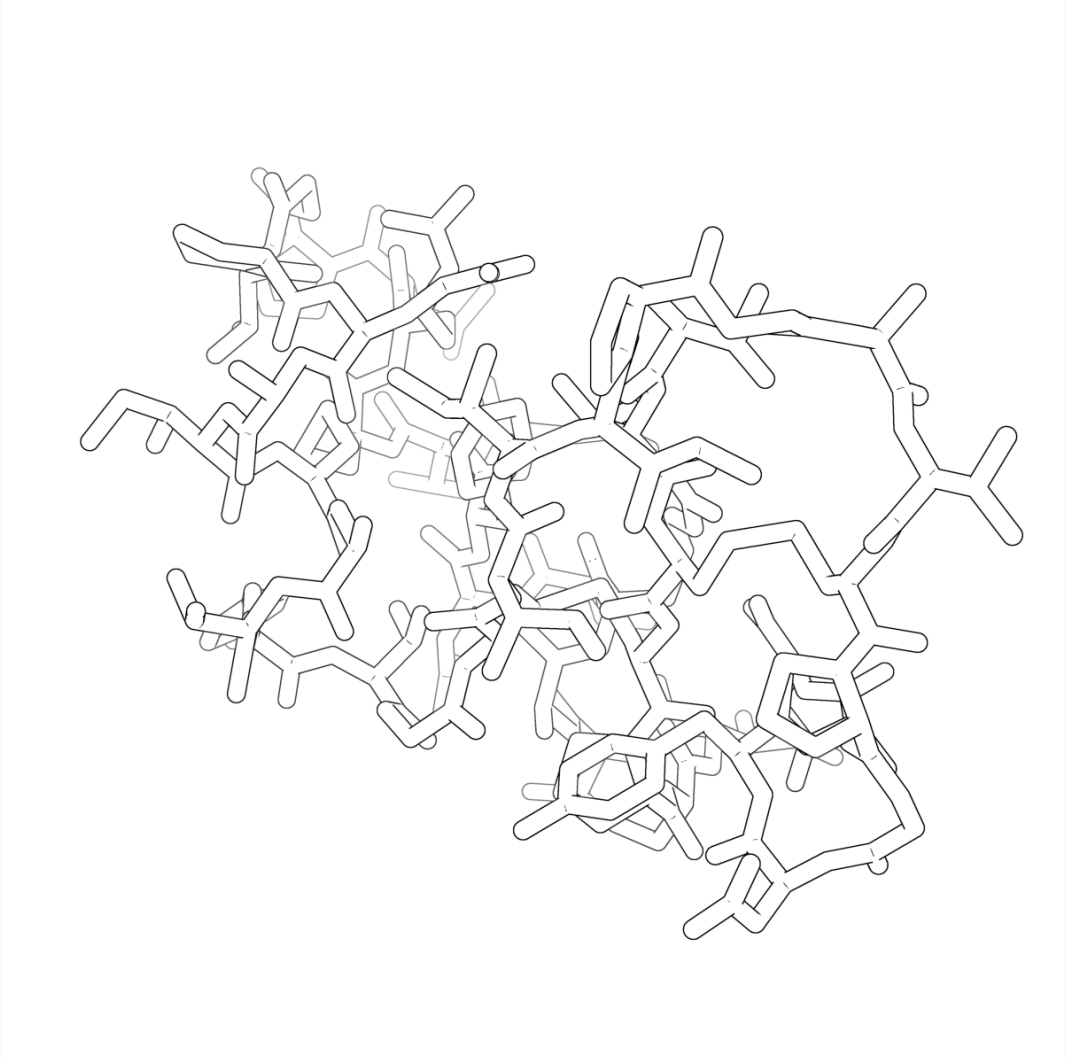


The amino acids don't just stay in a linear(ish) alignment but *fold* into more complex shapes.

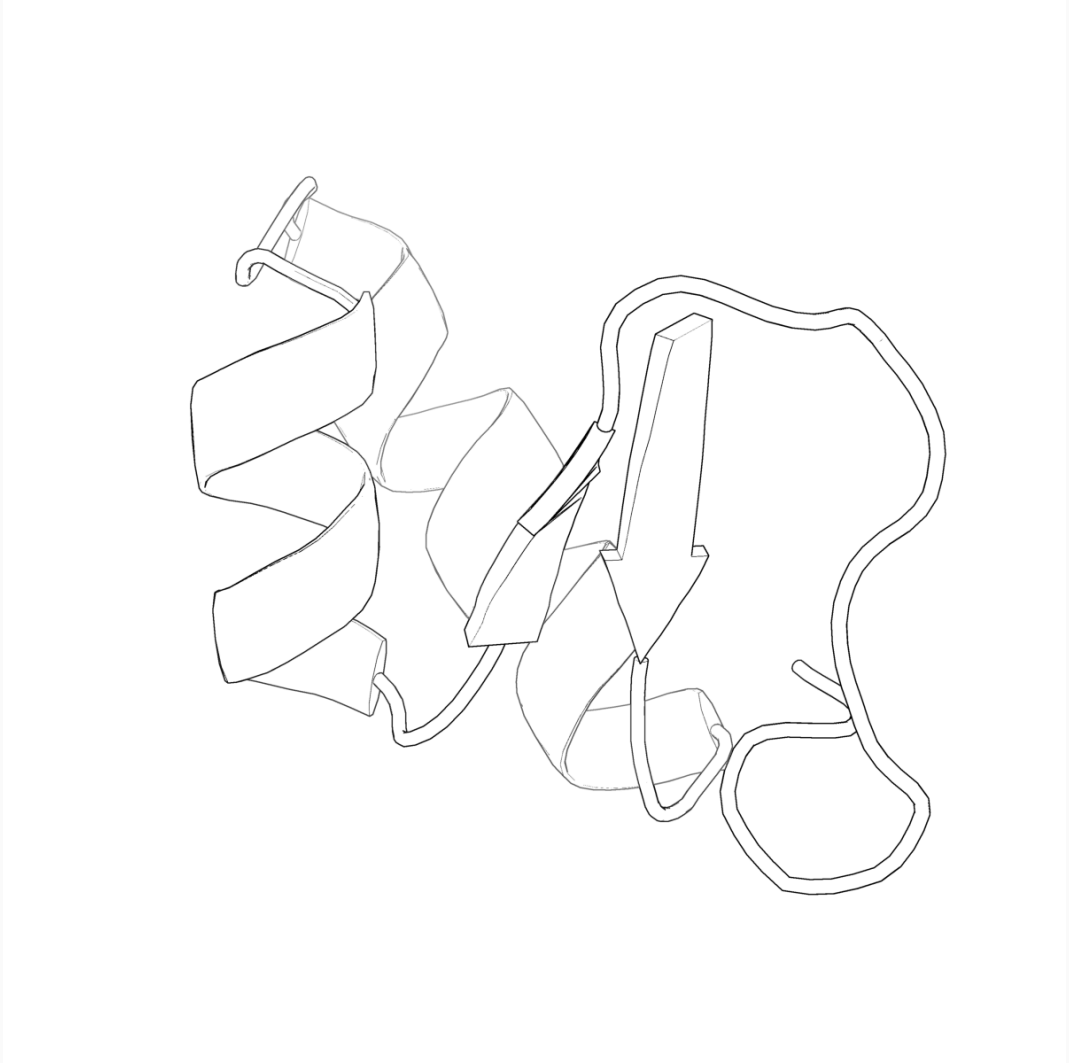
Why do we make the effort?

The structure and organisation of a biological macromolecule is intimately related to the biological function. Crystallography happens to be one of the best methods for revealing the structure. This activity has lead to several Noble prizes, and many millions of dollars or revenue amongst other things.

A SIMPLE PROTEIN (SIMPLIFIED)

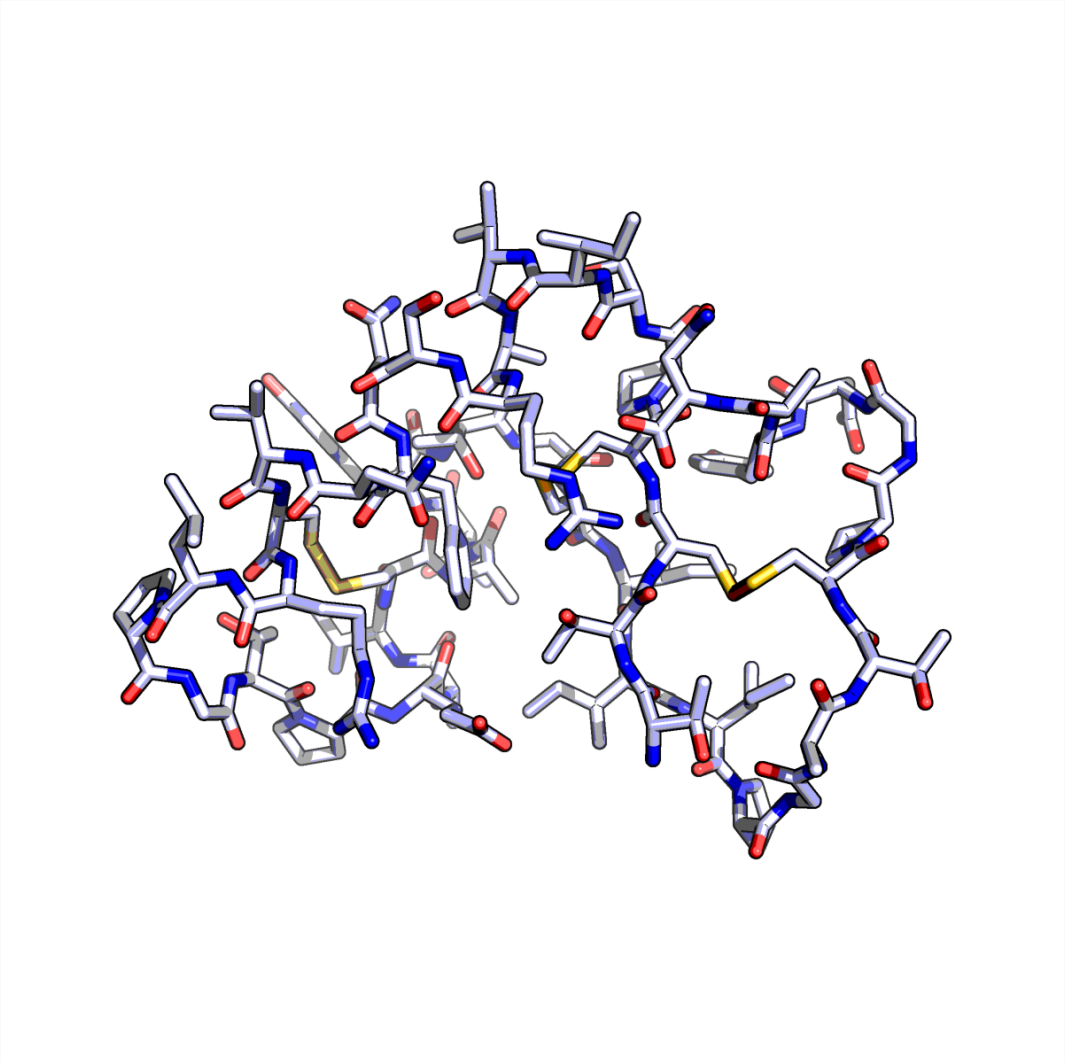


SIMPLIFIED FURTHER...



SOME ATOMS MAY BE IMPORTANT

Sulfur atoms displayed according to their atomic radius



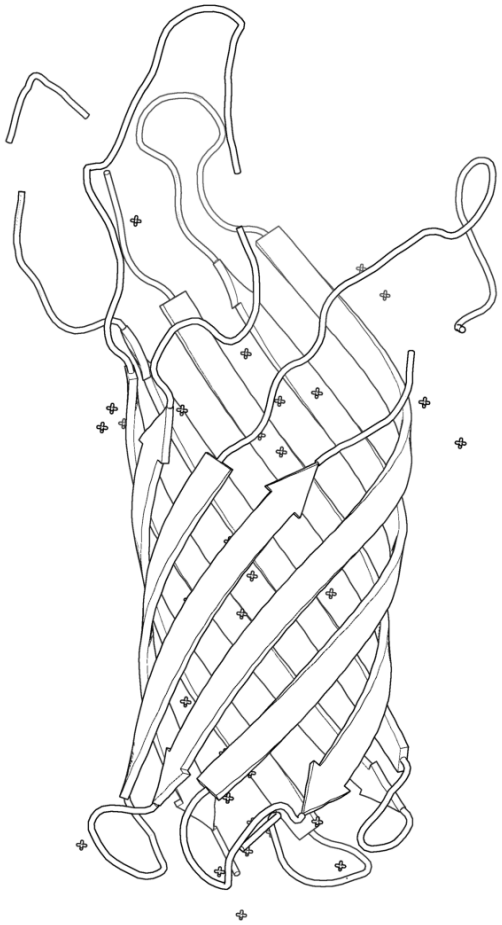
Crambin with the atoms coloured, C=white, N=Blue, O=red, S=yellow.

Protein structure contains many types of local structure.

We describe a number of typical structural types as α-helix and β-sheet, these are created acording the hydrogen bonding properties of the amino-acids. In general the molecules we observe are combinations of these two basic forms with loops and strands connecting them.

What may happen in folding a biomolecule?

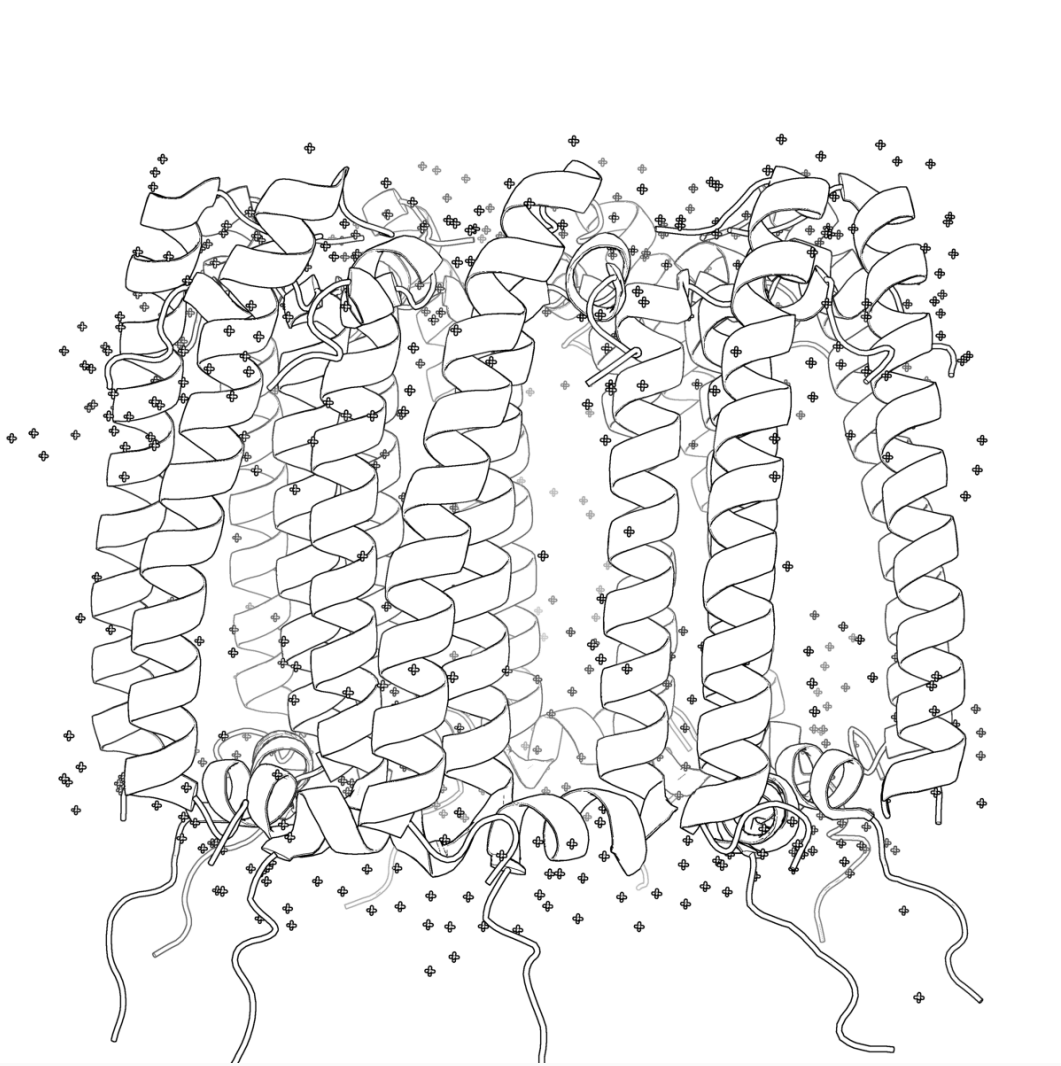
MOLECULE 1BXW



- some proteins are almost entirely β-sheet.
- strands protrude linking the sheets, in the model these strands are incomplete.

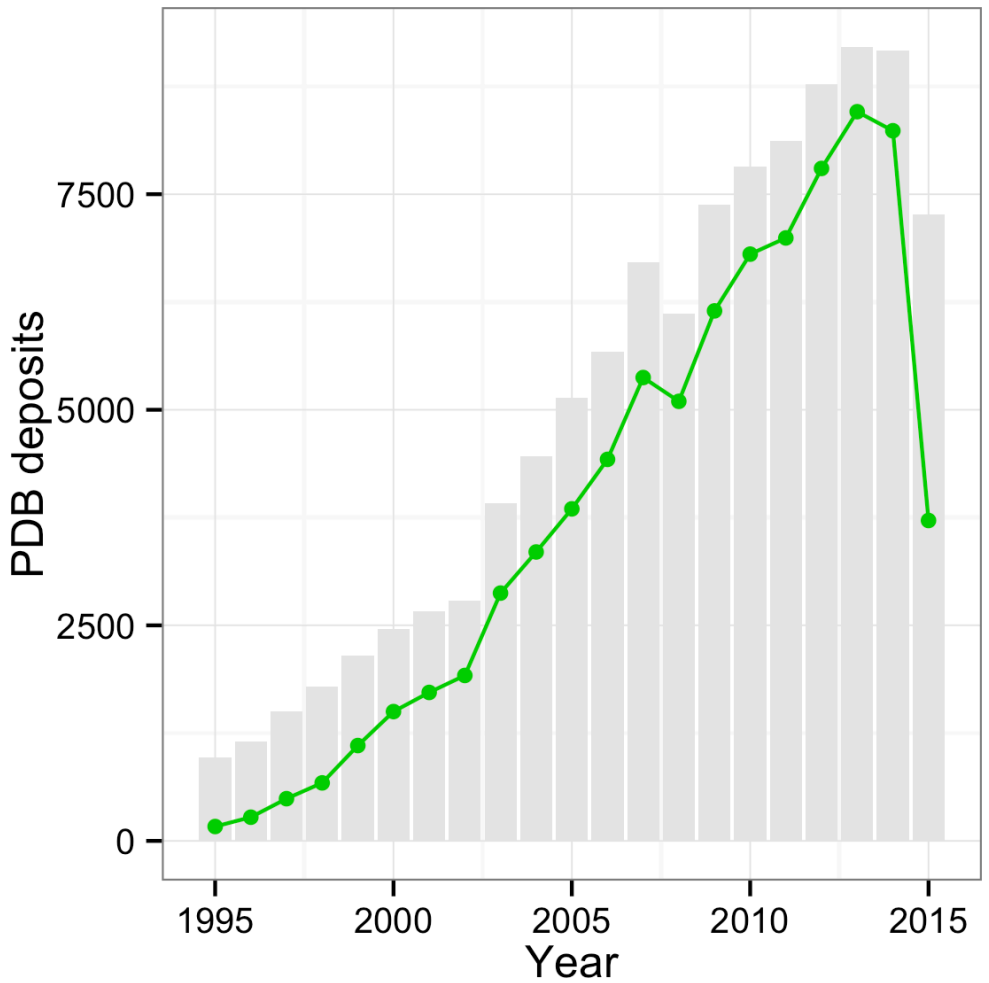
[3] http://www.rcsb.org

MOLECULE 2FKW



- other proteins are essentially all α-helical.
- again the linking strands are incomplete in the model.

CRYSTAL STRUCTURES.



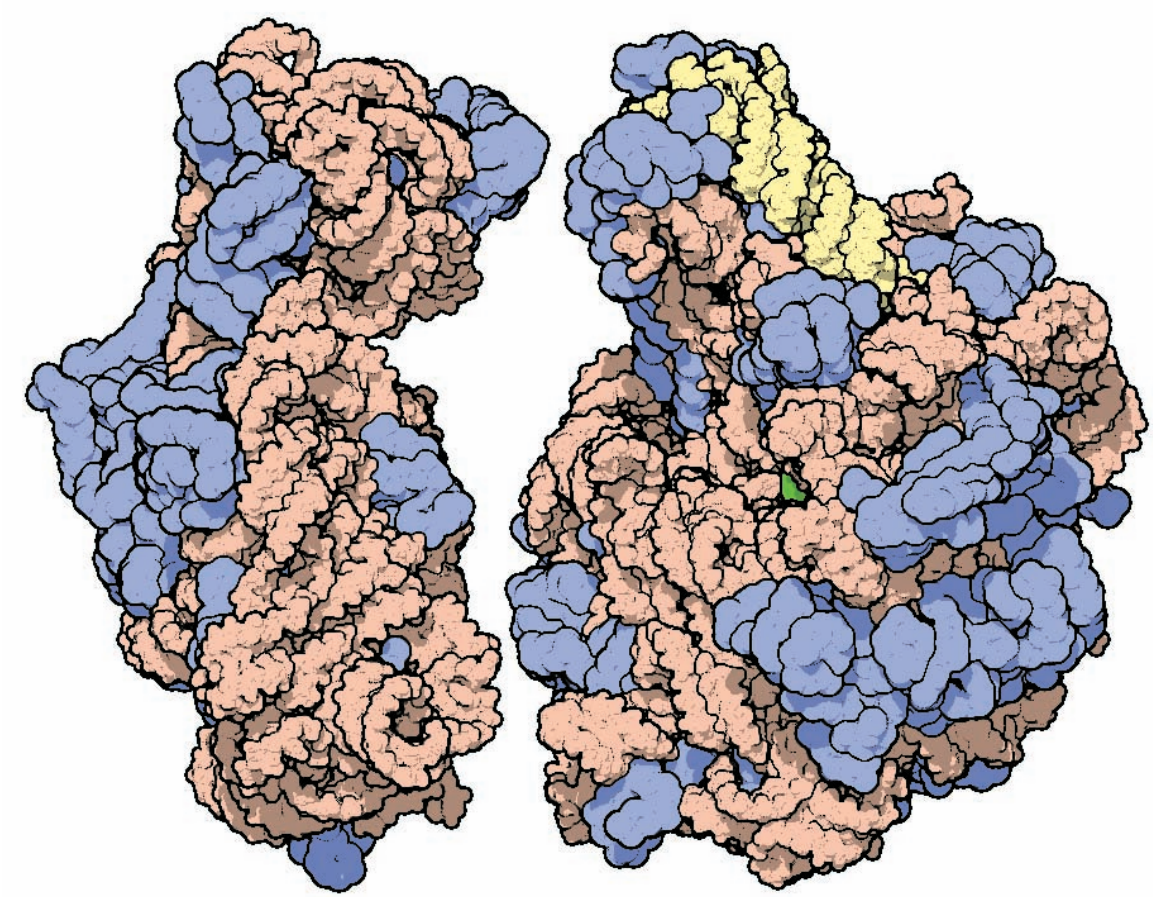
- The principle method for determining macromolecular structures is crystallography.
- The major source of these structures is through synchrotron radiation.
- *To understand the experiments we need to understand a little crystallography.*

[4]

[4] http://biosync.sbkb.org

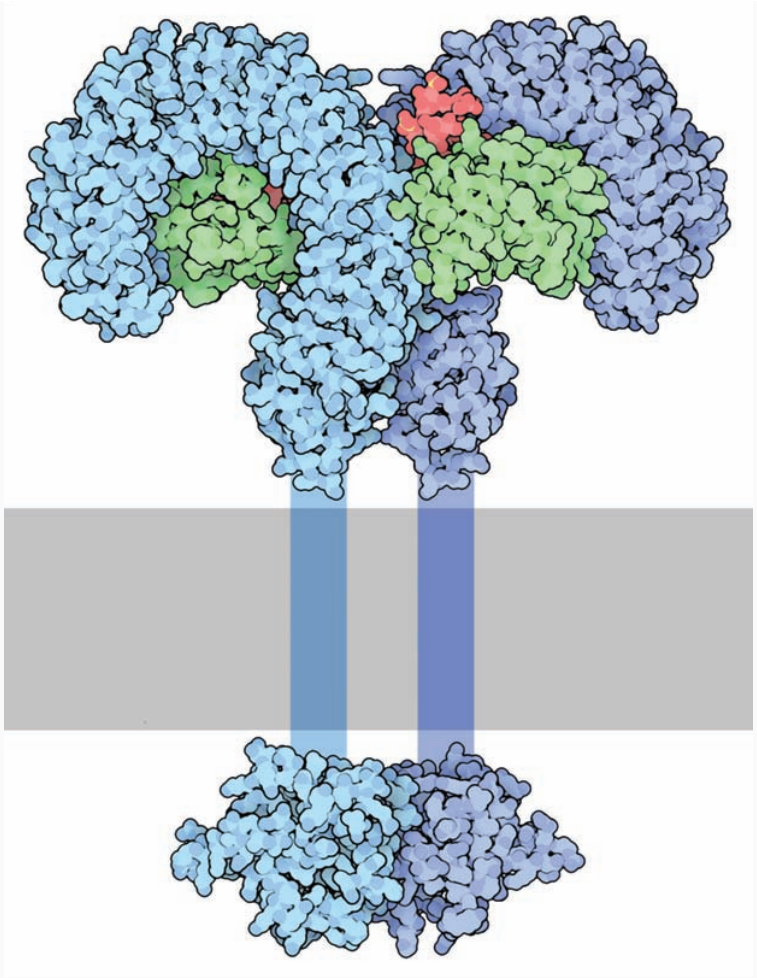
RIBOSOME

the factory of protein synthesis

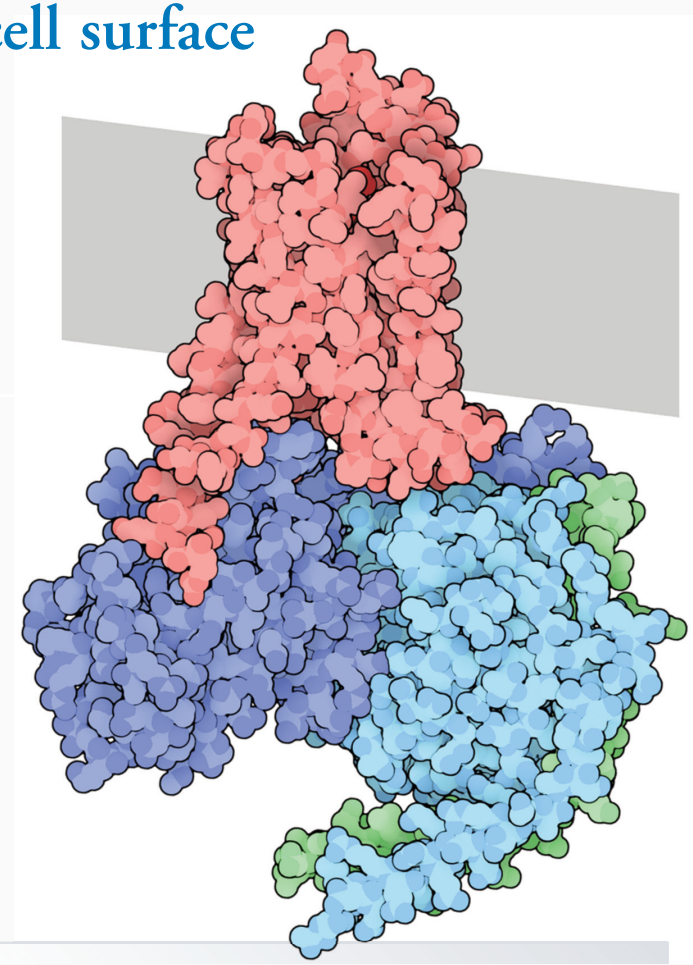


TOLL-LIKE RECEPTORS

frontline for the immune system



GPCRs
G-Protein-Coupled Receptors
Signaling across the cell surface



There are thousands of G-protein-coupled receptors on our cell surfaces, each waiting for its own particular messenger molecule. When the messenger is sensed, the G-protein associated with the GPCR initiates a chain reaction that amplifies the signal and yields an immediate cellular response. Our sense of sight relies on GPCRs that are sensitive to light, and our sense of smell is controlled by a thousand different forms of GPCR, each recognizing a different odorant molecule. Others are used in the nervous system to transmit nerve signals. Many widely-used drugs, such as Prozac, Claritin, and Zoloft, act by binding to proteins involved in GPCR signaling.

Related Resources:

Molecule of the Month: Adrenergic Receptors
www.rcsb.org/pdb/101/motm.do?momID=100

Molecule of the Month: G Proteins
www.pdb.org/pdb/101/motm.do?momID=58

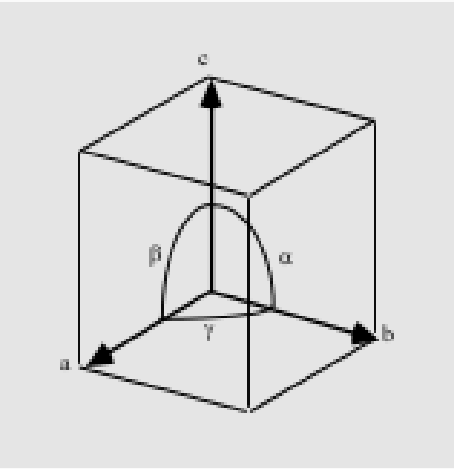
Author Profile: Brian K. Kobilka, **bit.ly/OoCTLX**

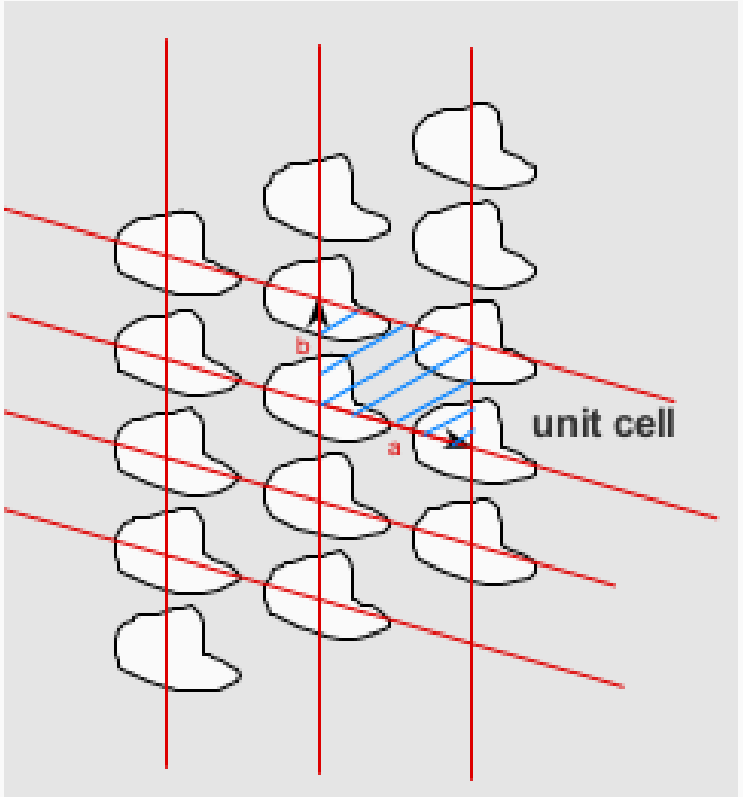
THE PROPERTIES OF A CRYSTAL.

What is a crystal?

THE CRYSTAL LATTICE.

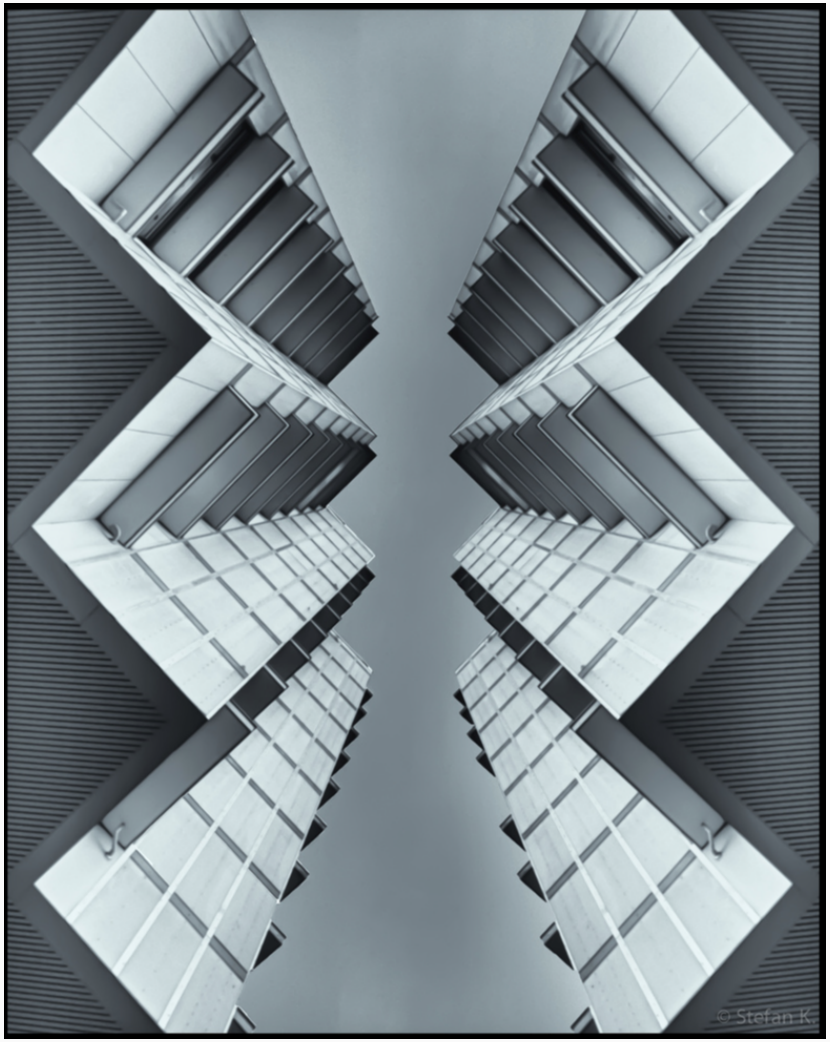
A periodic arrangement of atoms or molecules in three dimensions.



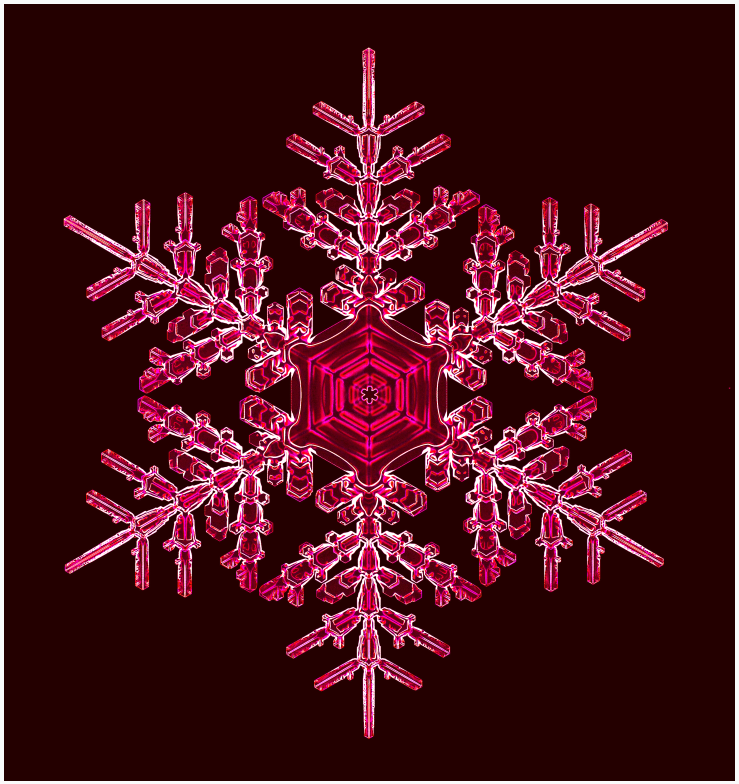


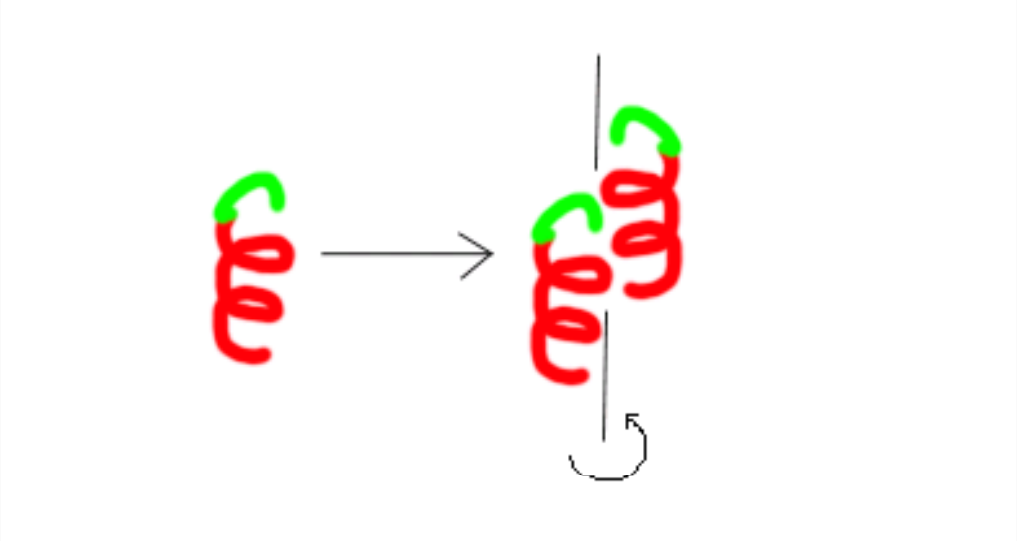
What is symmetry?

SYMMETRY?

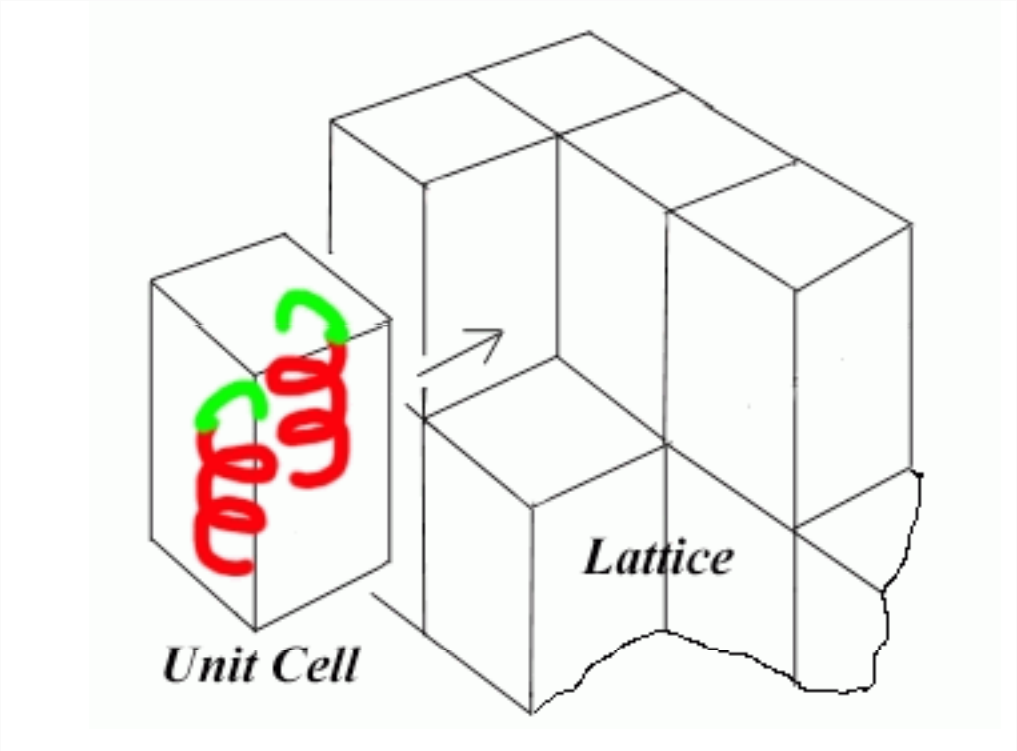


MORE OR BETTER SYMMETRY?



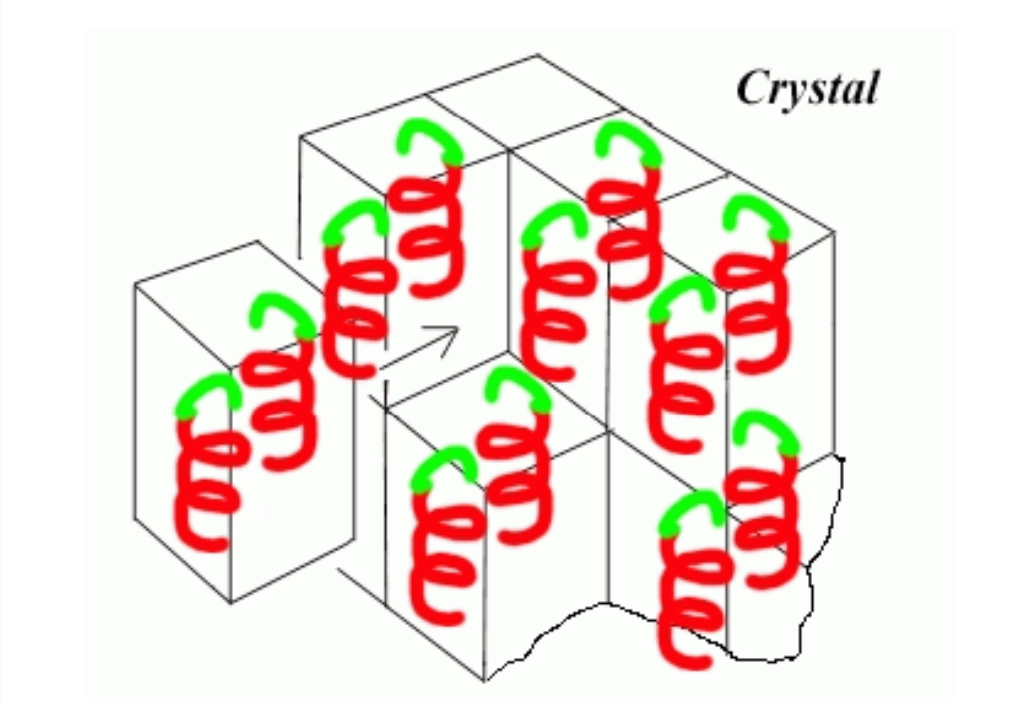


the molecules are arranged with (*two-fold*) symmetry.



the repeating unit fills the *unit–cell*.

MACROMOLECULAR CRYSTALS.

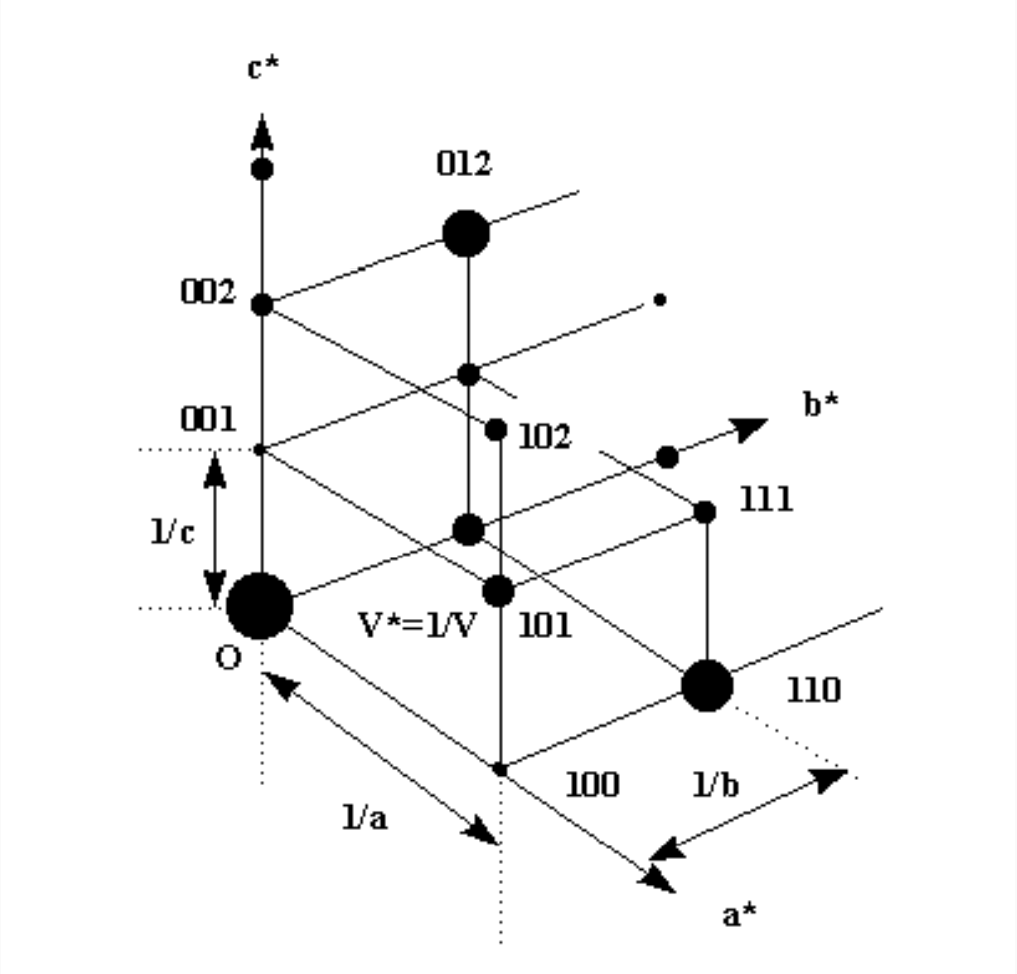


creating many units cells enable a crystal to be formed.

`http://proteinformatics.charite.de/ngl/html/ngl.html`

(or search “ngl viewer”)

THE RECIPROCAL LATTICE.



The observed intensity I of the diffraction spots can be thought of as corresponding to the 'size' of the reciprocal lattice point ($I_{(hkl)}$ is proportional to $|F_{(hkl)}|^2$). Clearly, either depends on the contents of the unit cell

What is the magnitude of the $F_{(000)}$ reflection? Have a guess...

The scattering from an atom in a regular structure may be calculated.

$$f_{hkl} = f_j \exp\left[2\pi i \left(hx_j + ky_j + lz_j\right)\right]$$

Where the f_j represents the scattering from an individual atom. The (h, k, l) represent the *Bragg* lattice, and the fractional coordinates take care of the phase.

To consider a crystal we will need to consider all the atoms in the molecule.

$$F_{hkl} = \sum_{\text{atoms}} f_j \exp\left[2\pi i \left(hx_j + ky_j + lz_j\right)\right]$$

$$F_{hkl} = \int_V \rho(x, y, z) \exp\left[+2\pi i \left(hx_j + ky_j + lz_j\right)\right] dV$$

The structure factor F_{hkl}, is related to the *electron–density* within a volume V via a Fourier transform.

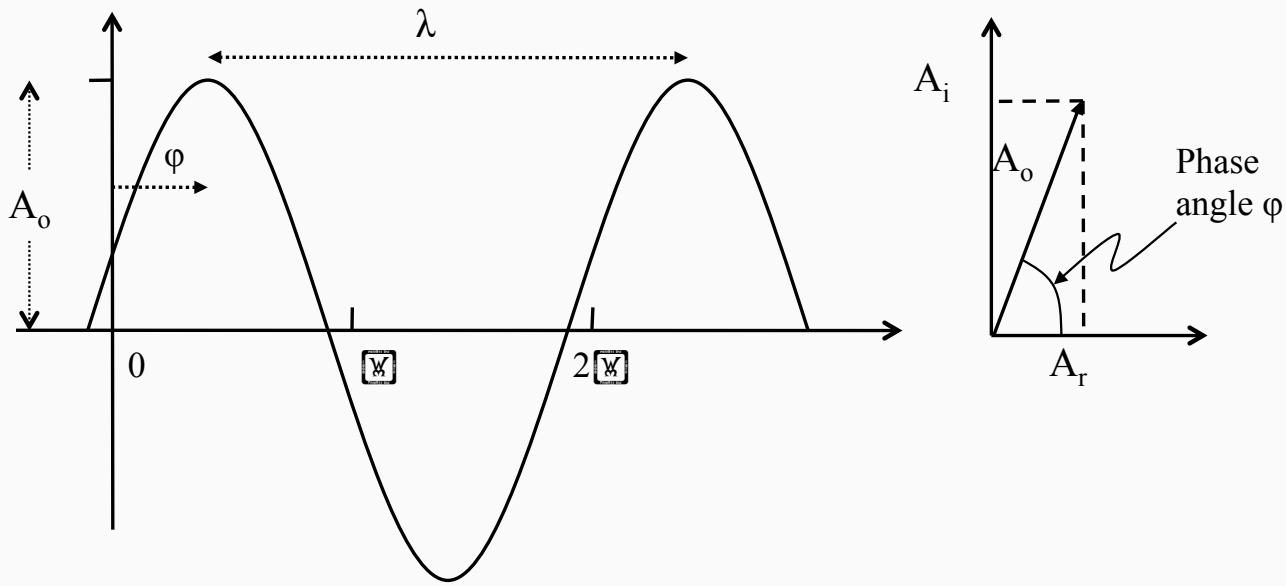
$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k} \sum_{l} F_{hkl} \exp\left[-2\pi i \left(hx_j + ky_j + lz_j\right)\right]$$

DATA COLLECTION.

THE GOAL OF THE DIFFRACTION EXPERIMENT

The goal of the experiment is to determine $\rho_{(x,y,z)}$, the electron density for all x, y, z in the unit cell. With this achieved we can *interpret* the electron density in terms of the chemical and structural *prior knowledge* of the molecules.

- What we measure in the experiment is $I_{(hkl)} \propto |F_{(hkl)}|^2$
- What we still need is $\phi_{(hkl)}$ - this is the phase problem of crystallography.



why a problem?

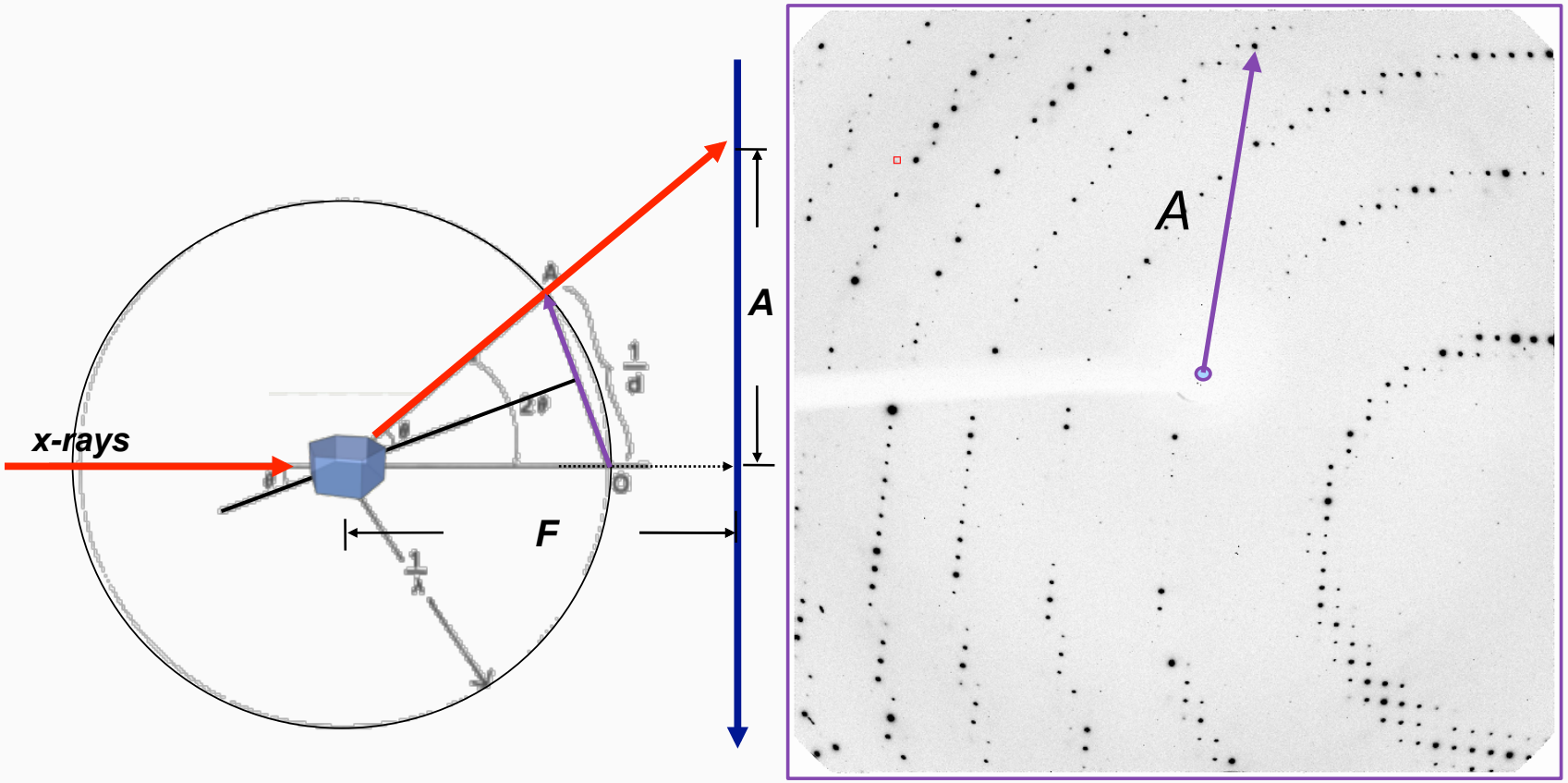
The detectors we use measure intensity only, not phase. The phase is related to the (fractional) distance of an atom for one of the crystallographic planes. Without the phases the Fourier transform is *problematic*.

- put your crystal in the beam.
- make sure it stays in the beam at all orientations.
- where the *Bragg* condition is satisfied diffraction orders may be seen

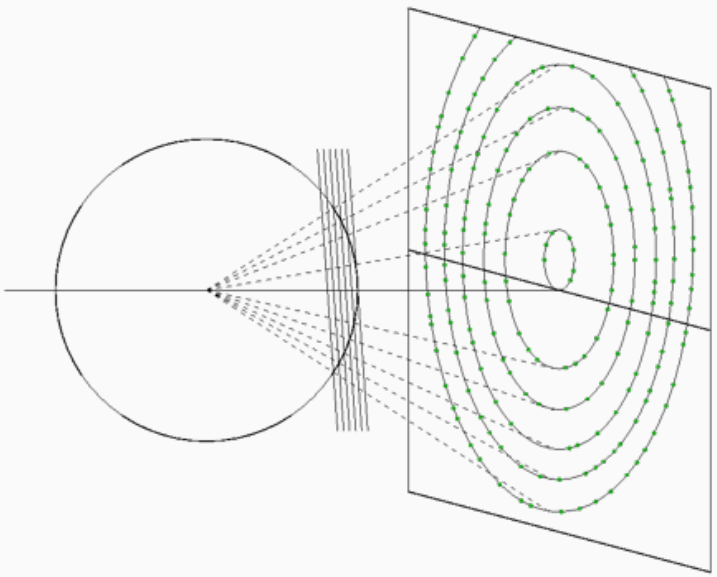
Bragg's Law

$$2d\sin\theta = n\lambda$$

THE DIFFRACTION EXPERIMENT

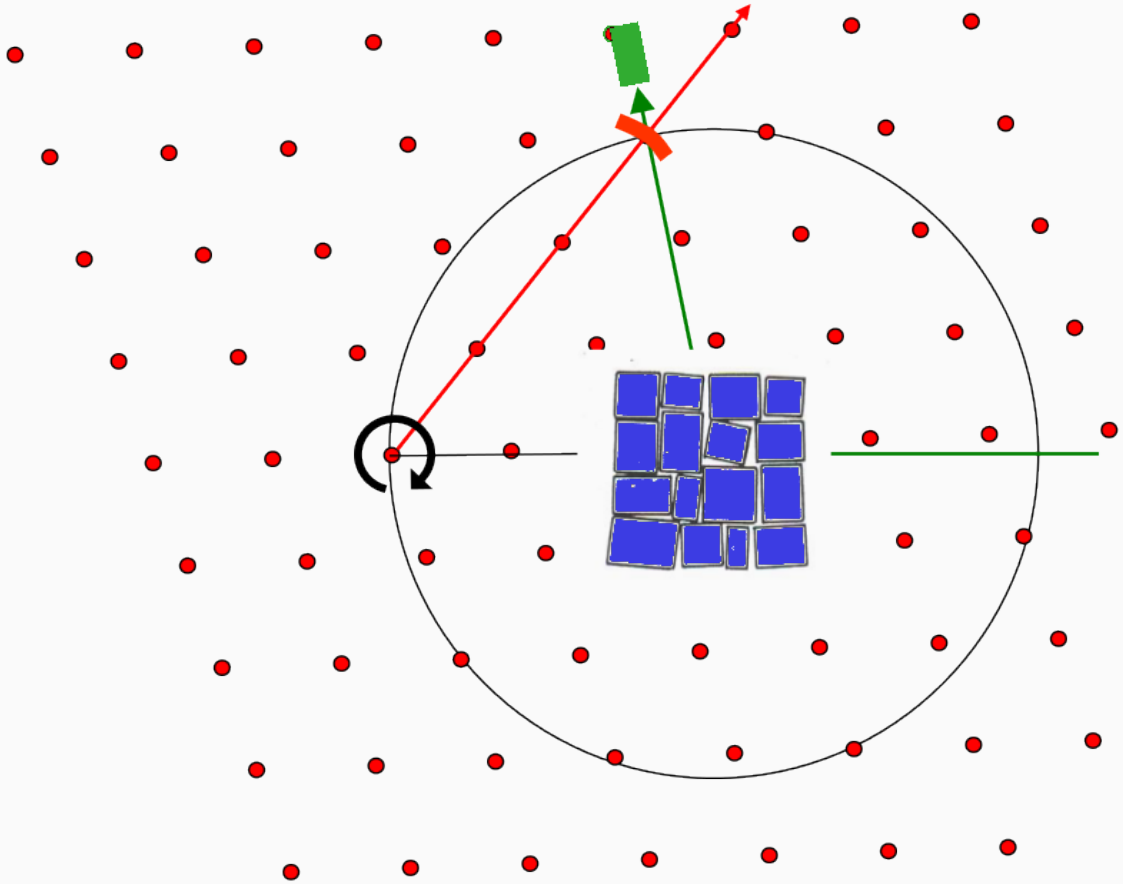


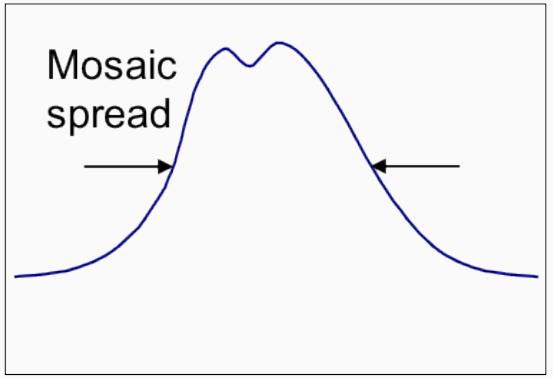
Bragg's Law is obeyed and diffraction occurs when a vector of length $1/d_{(hkl)}$, that is perpendicular to the lattice planes. (hkl)



The planes of spots in reciprocal space appear as circles of spots on an area-sensitive x-ray detector. If the crystal is rotated more (different) spots obey Bragg's Law and the circles become lunes.

A SMALL REALITY CHECK

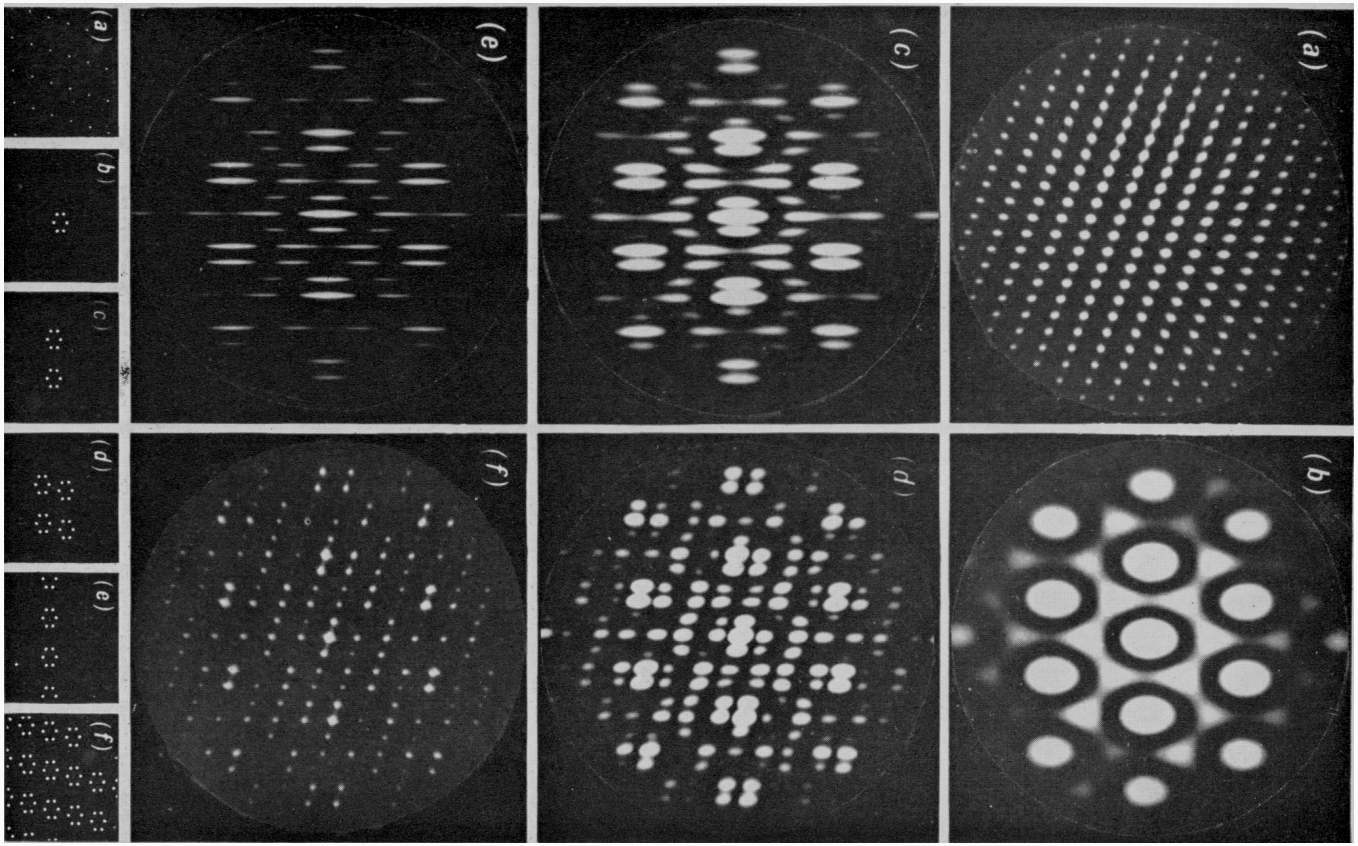




Omega

1. Detector stationary at 2θ angle.
2. Crystal is rotated about θ by +/- ω.
3. FWHM is the mosaic spread.

RELATIONSHIP MOLECULE TO CELL TO DIFFRACTION.

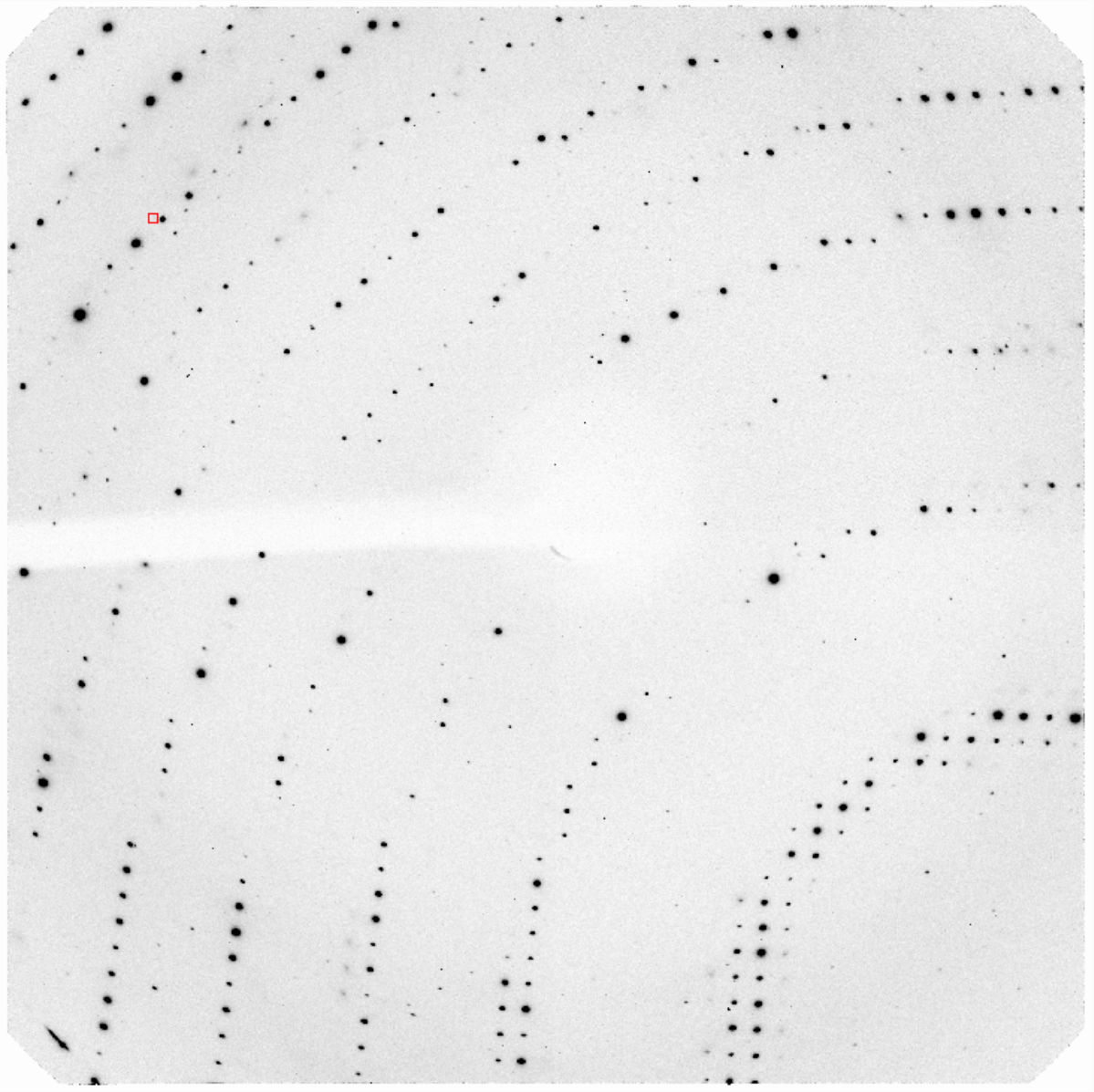


Notice (1) The symmetry, and (2) how the continuous diffraction pattern of one molecule (b) is “sampled” by the lattice of diffraction points.

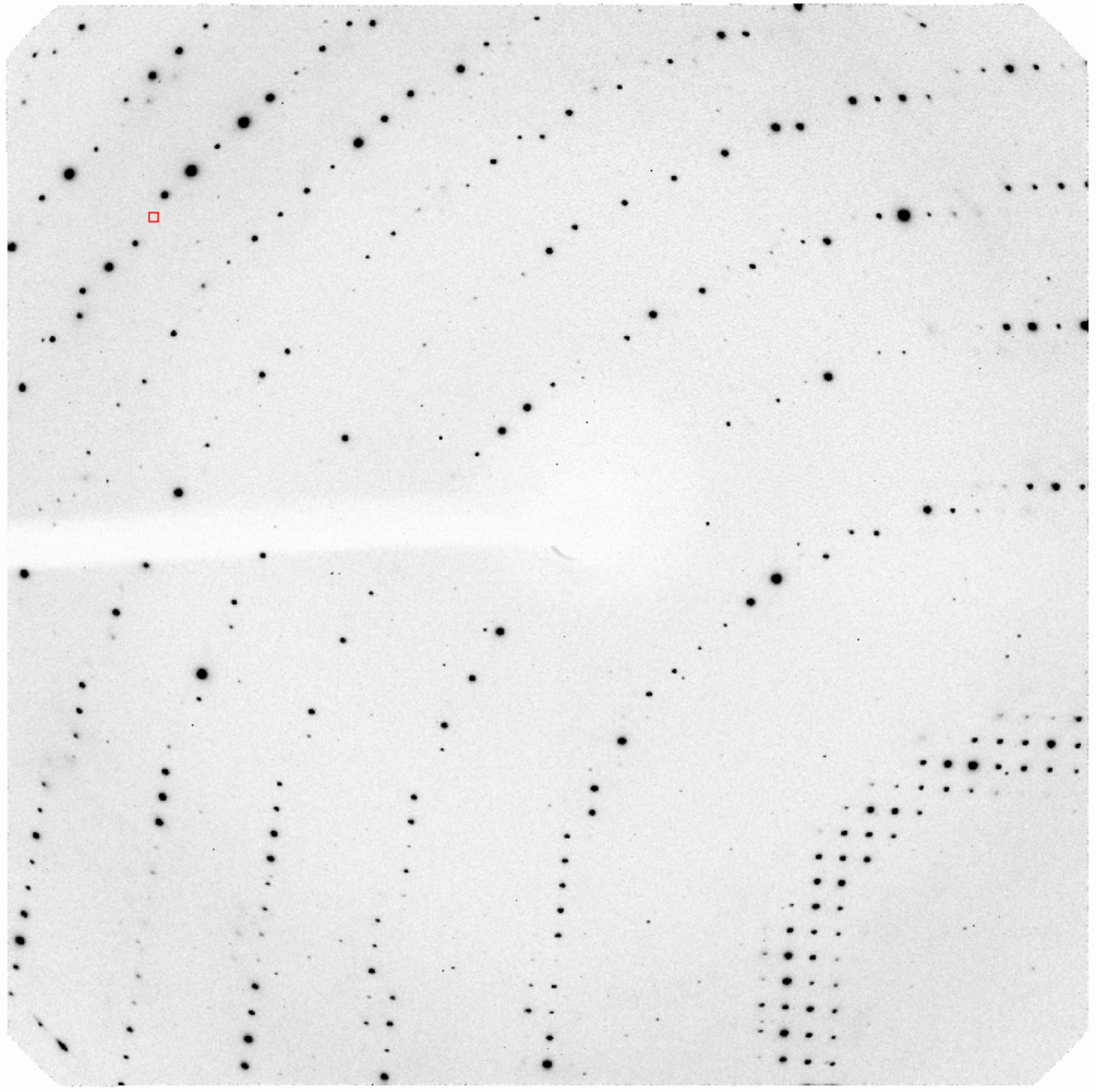
Taylor and Lipson, *Optical Transforms*, 1964

What will the diffraction pattern during a data collection look like? Explain your thinking.

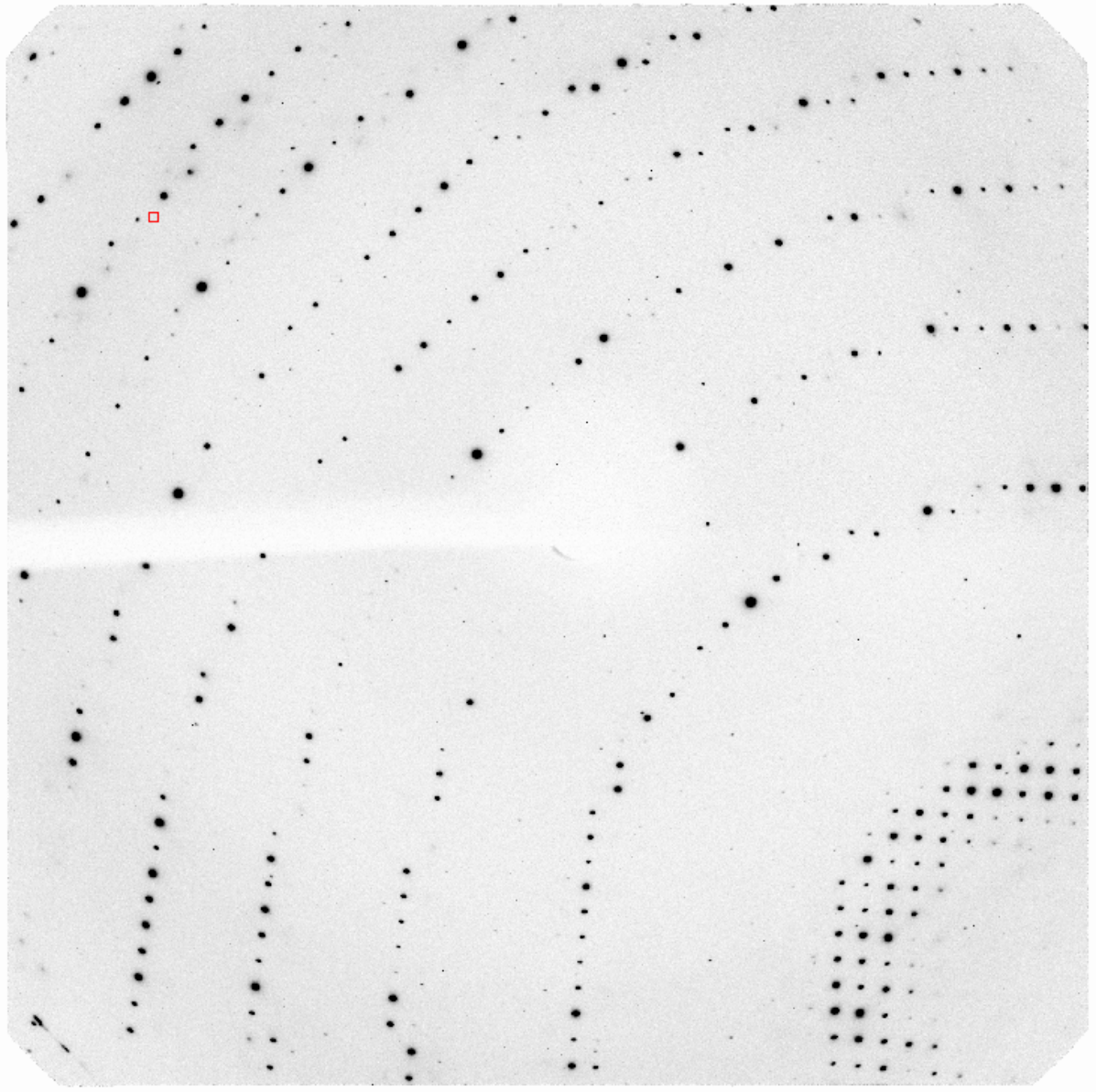
DIFFRACTION PATTERN



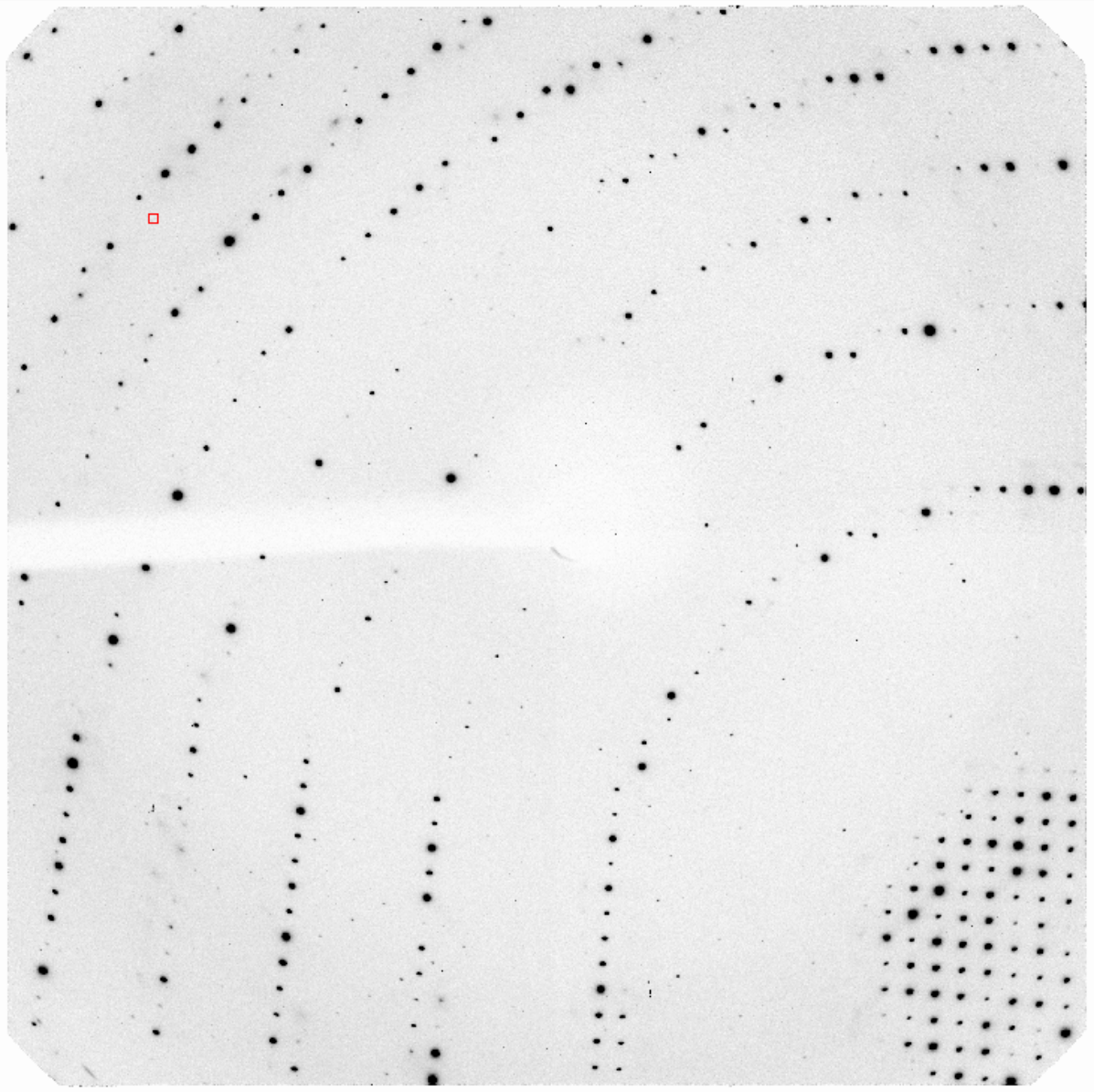
DIFFRACTION PATTERN



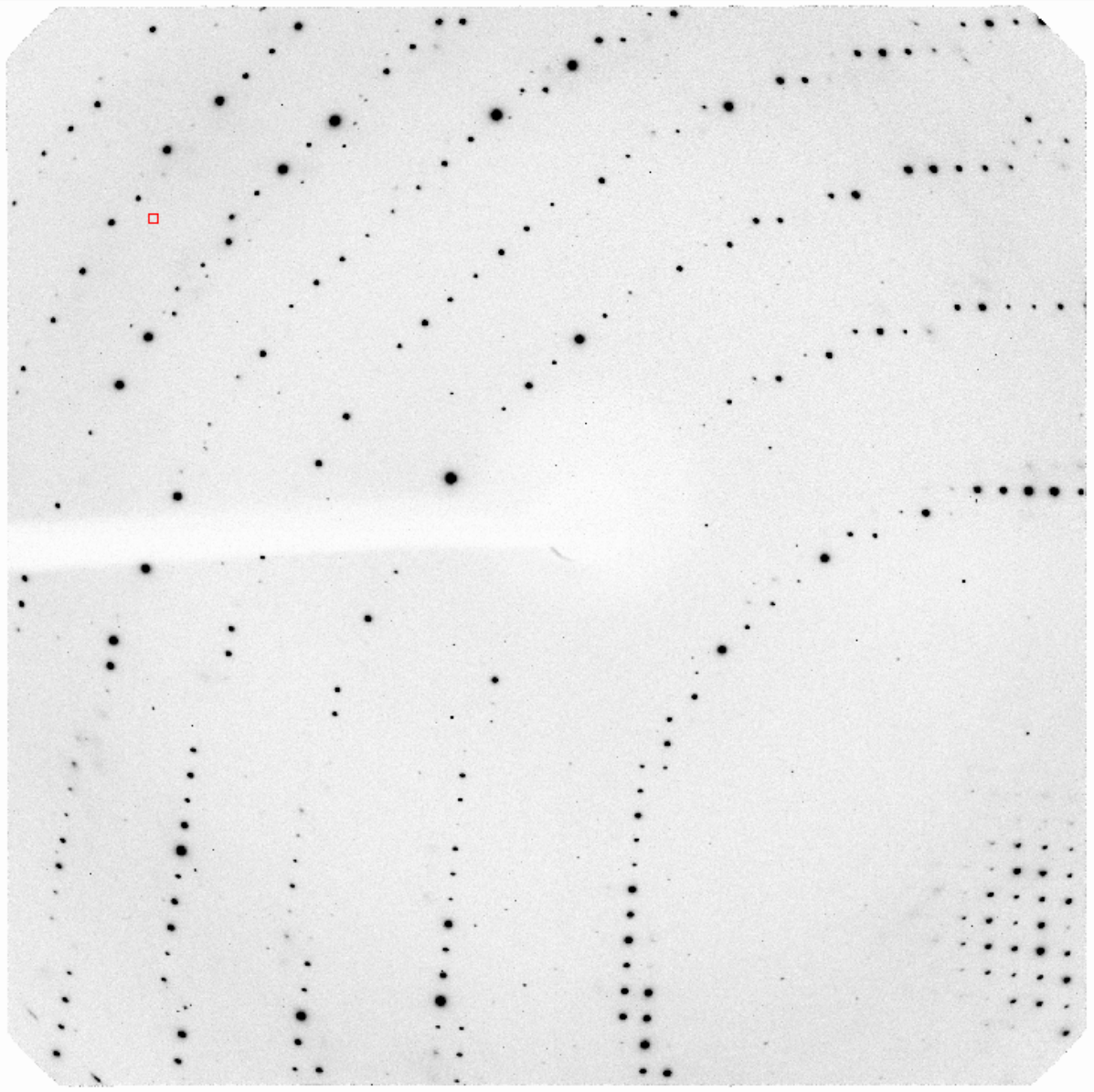
DIFFRACTION PATTERN



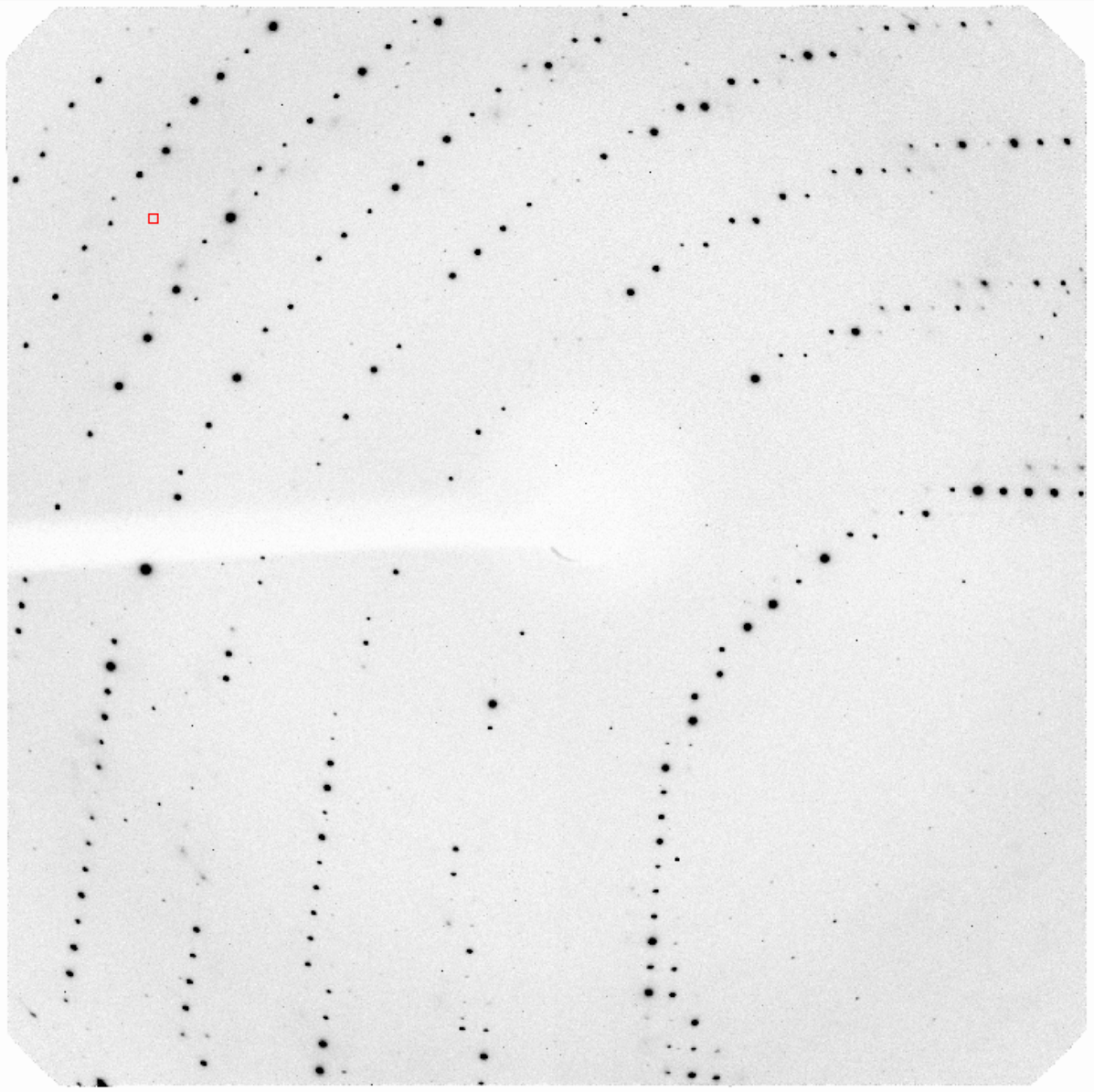
DIFFRACTION PATTERN



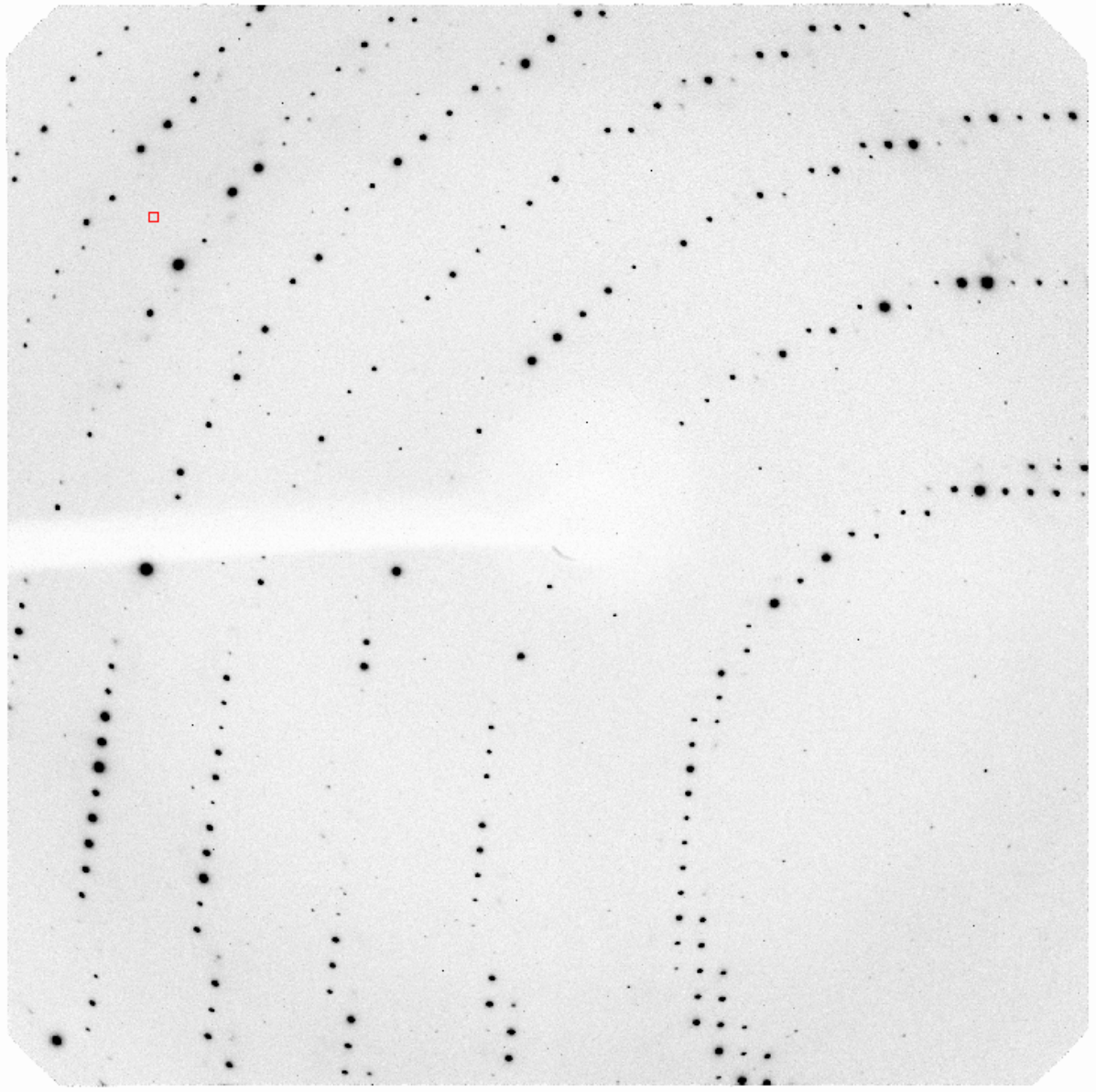
DIFFRACTION PATTERN



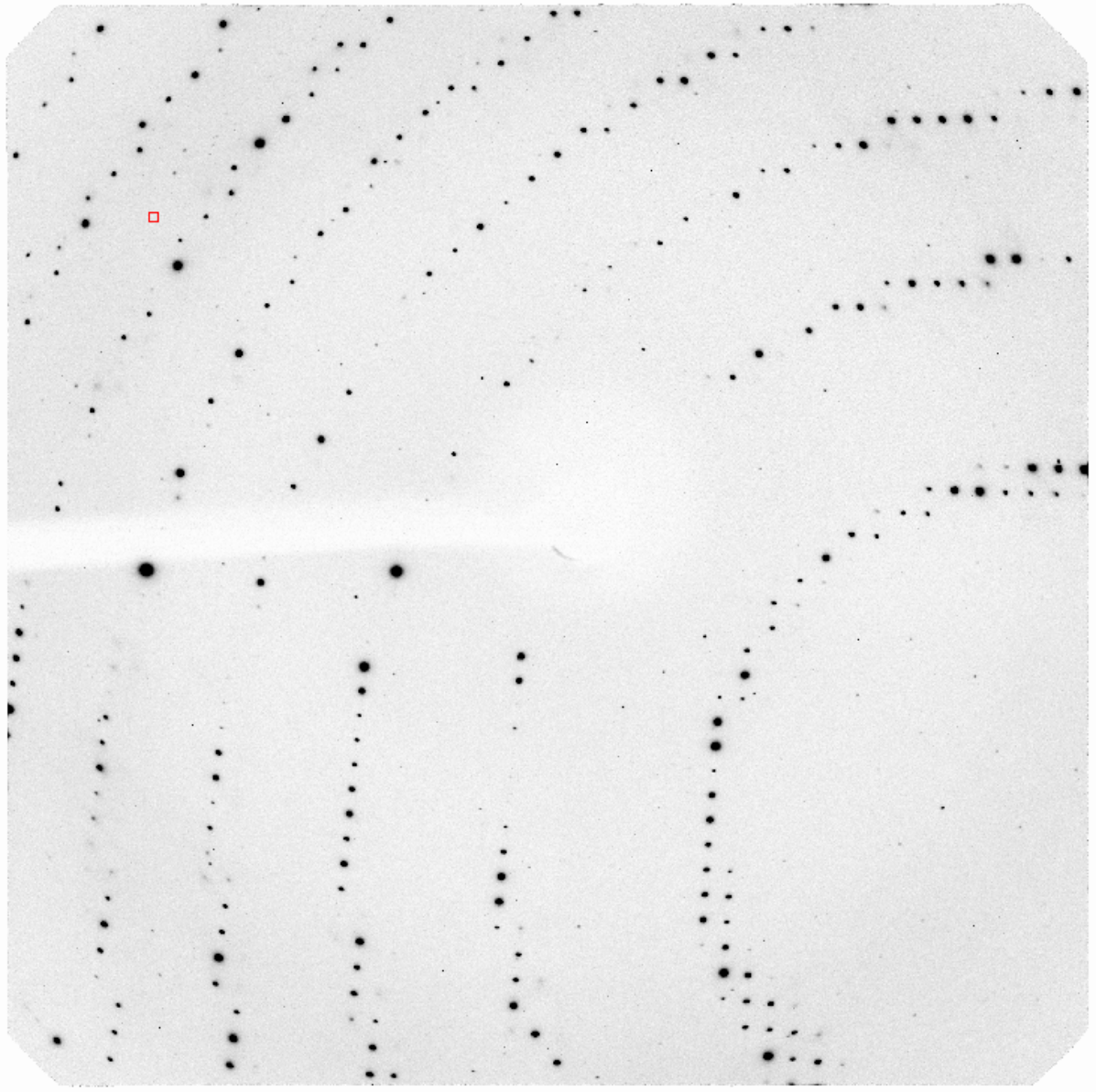
DIFFRACTION PATTERN



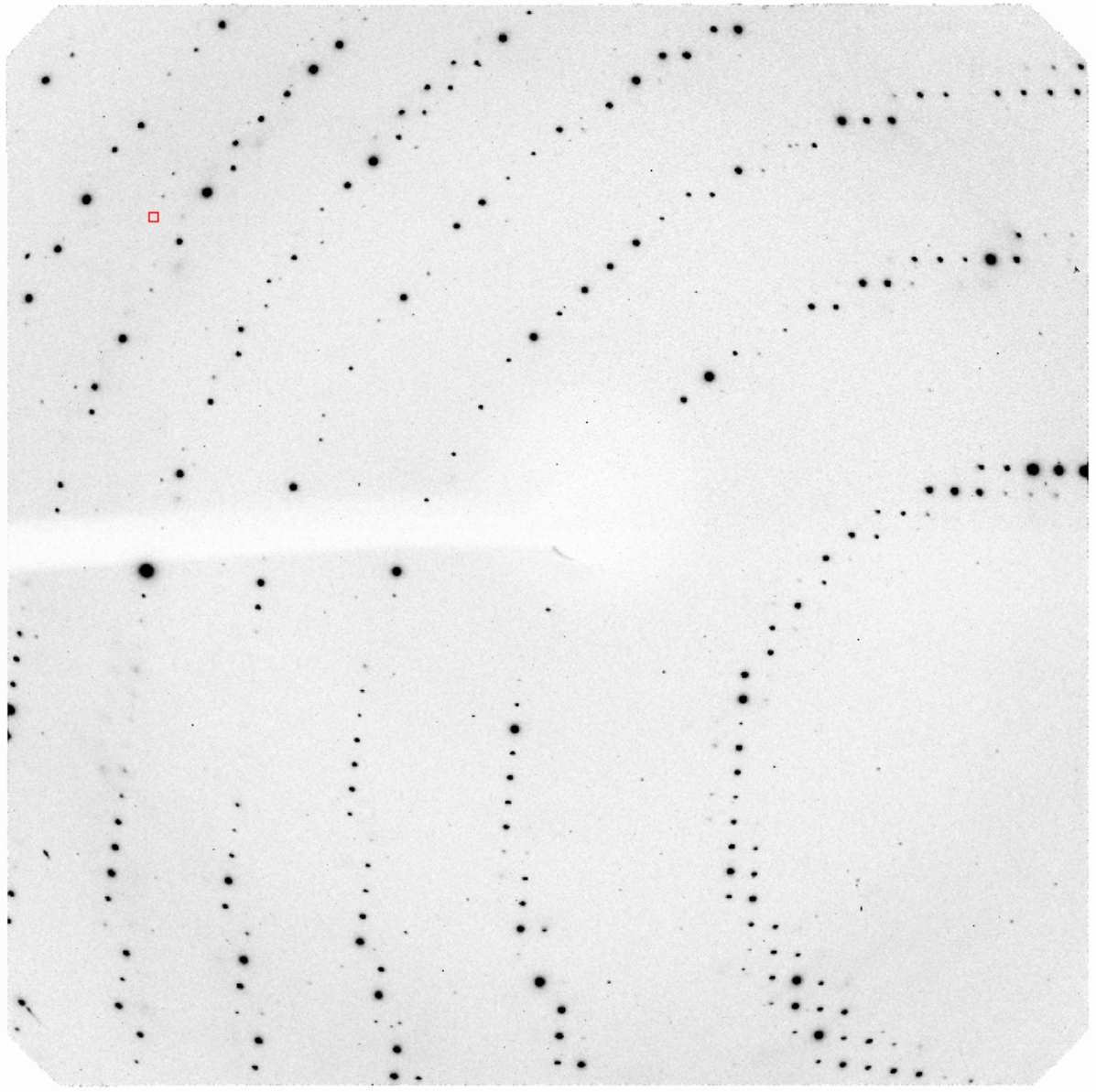
DIFFRACTION PATTERN



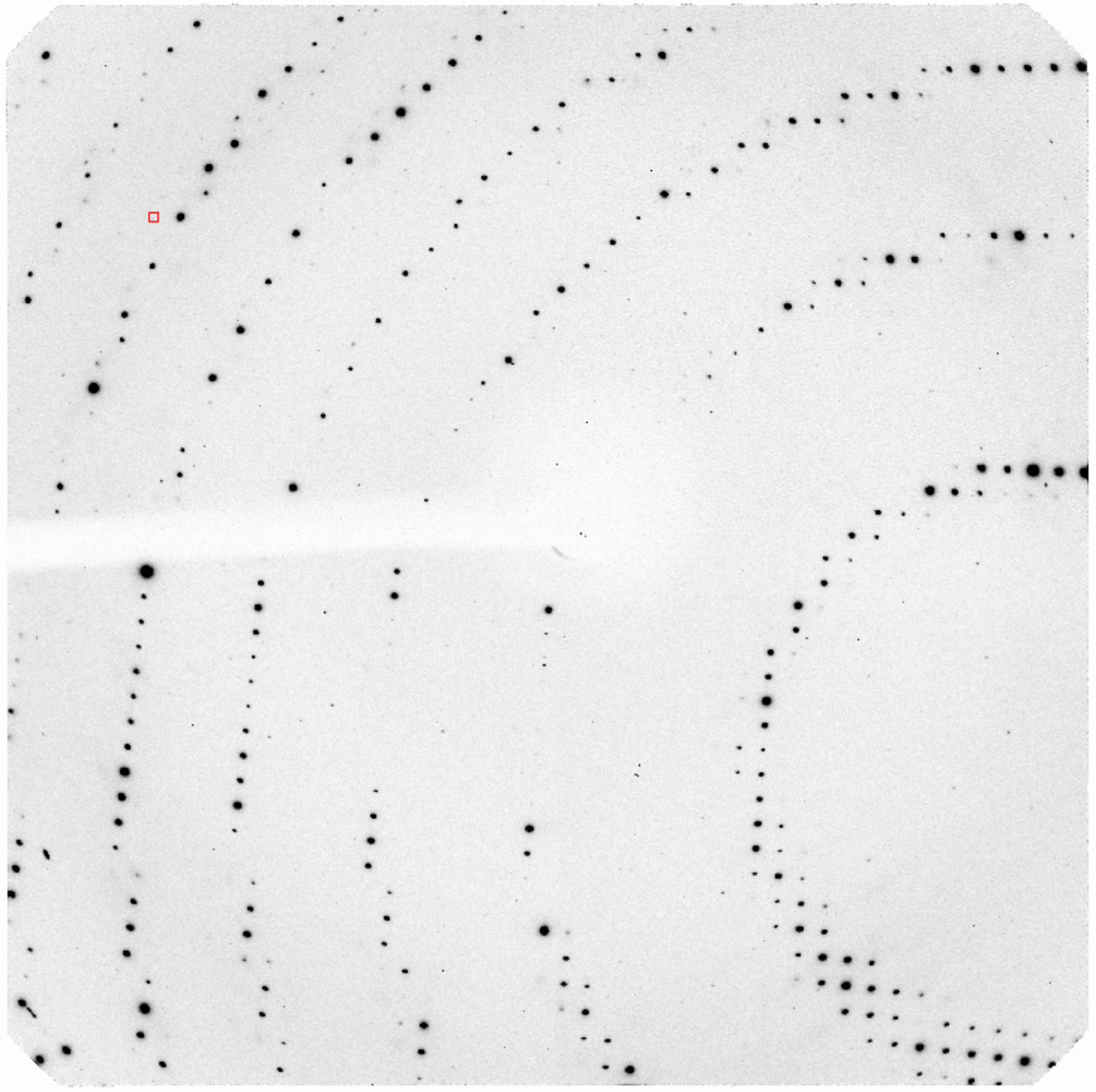
DIFFRACTION PATTERN



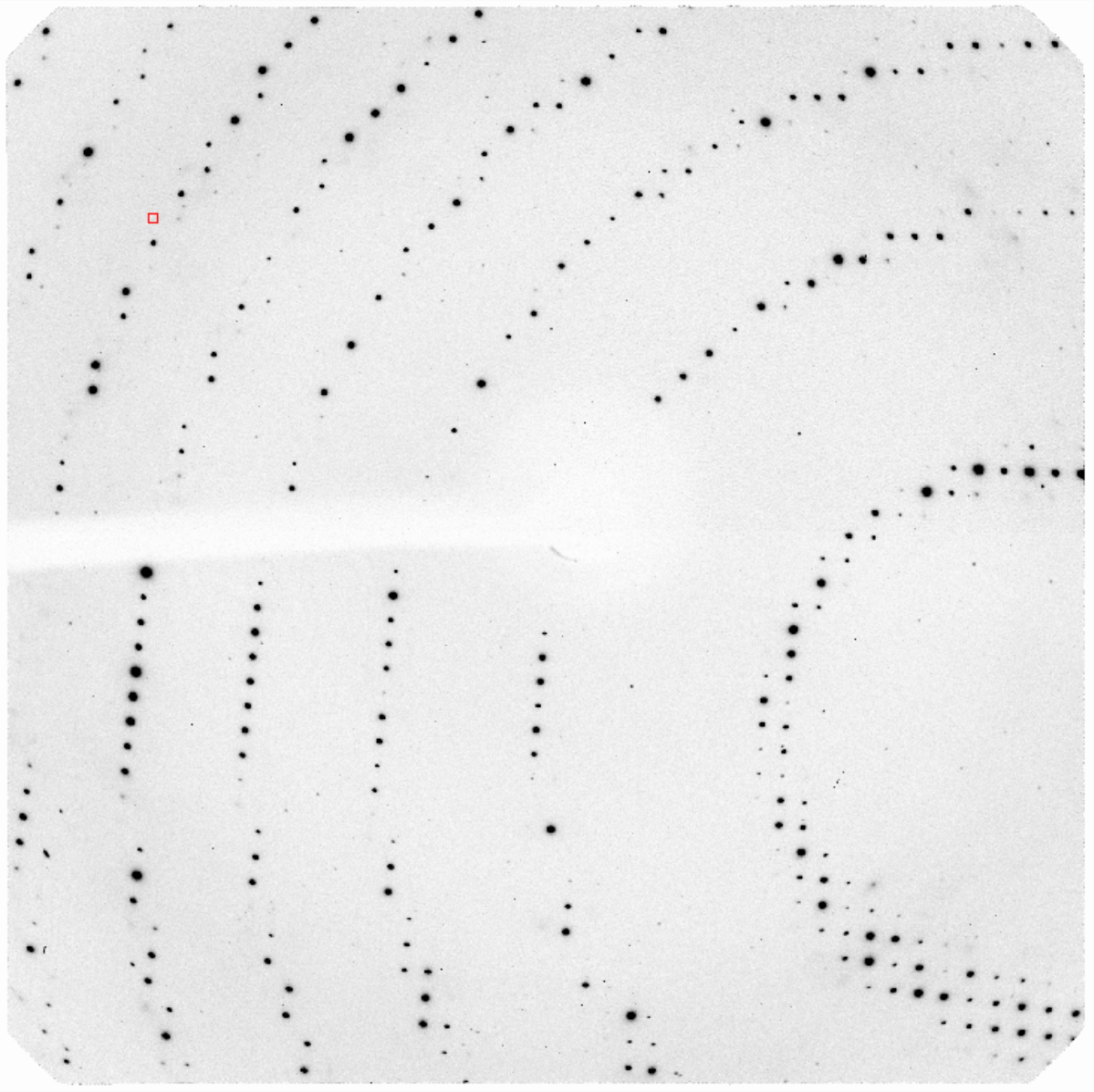
DIFFRACTION PATTERN



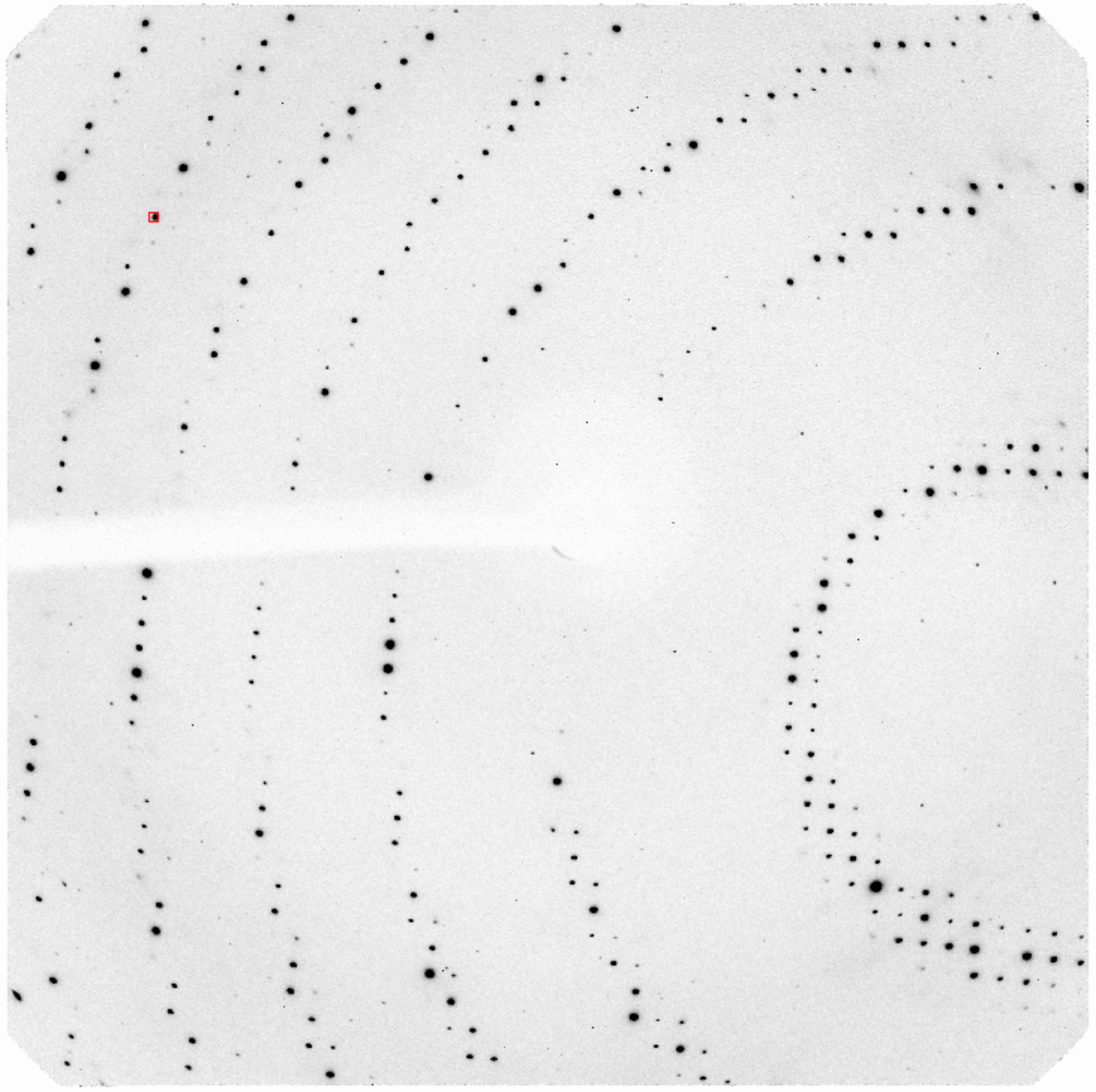
DIFFRACTION PATTERN



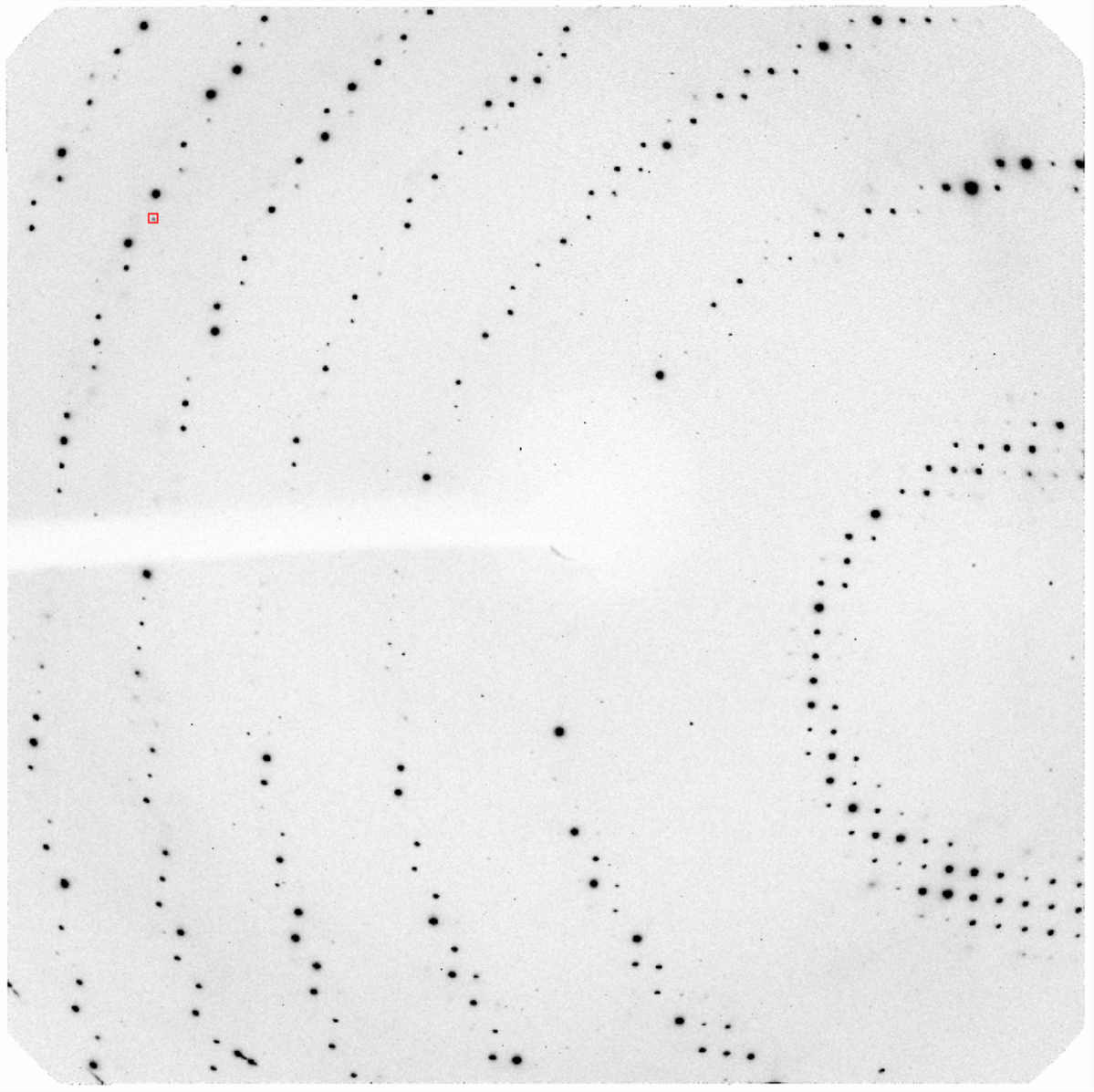
DIFFRACTION PATTERN



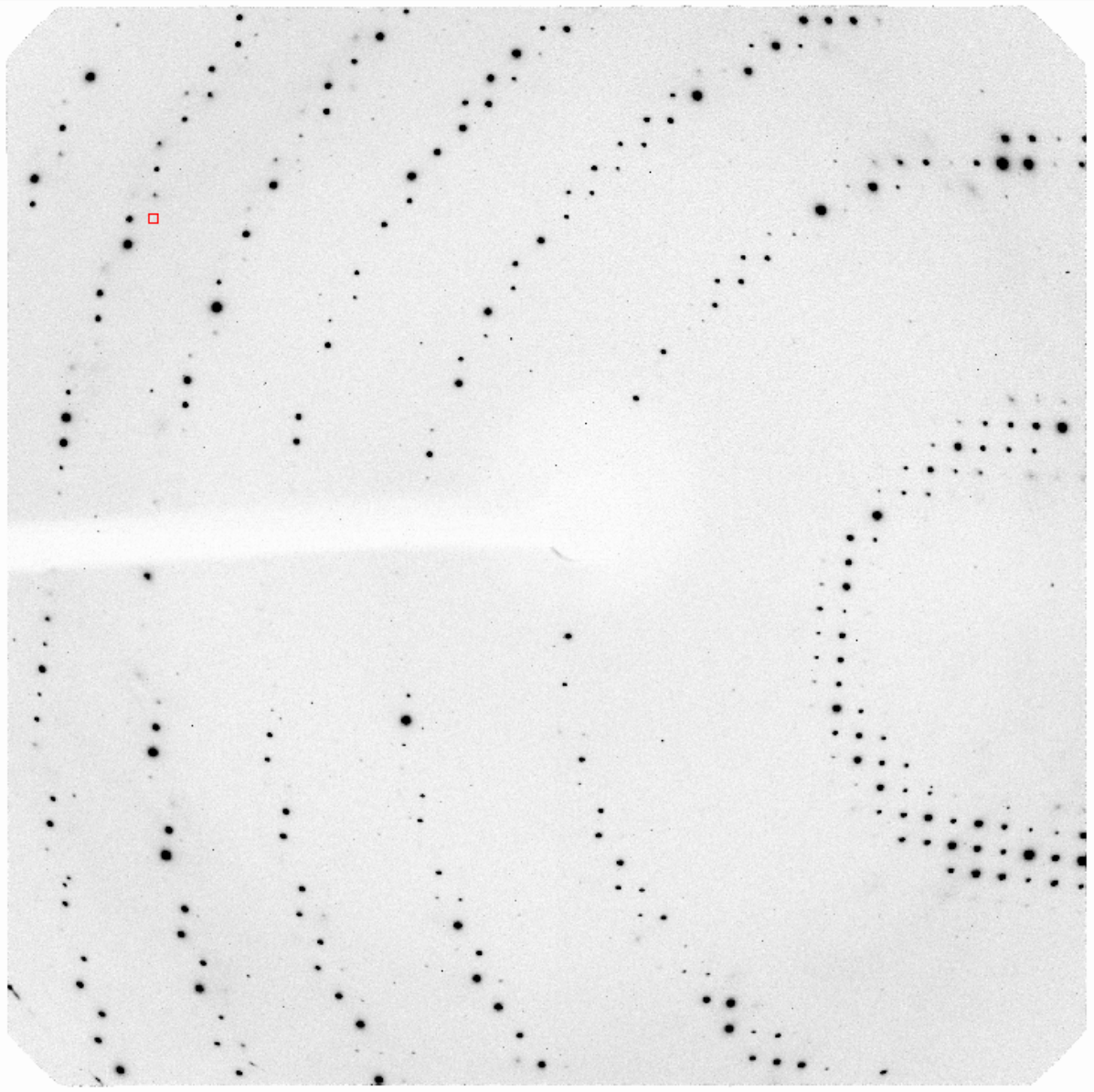
DIFFRACTION PATTERN



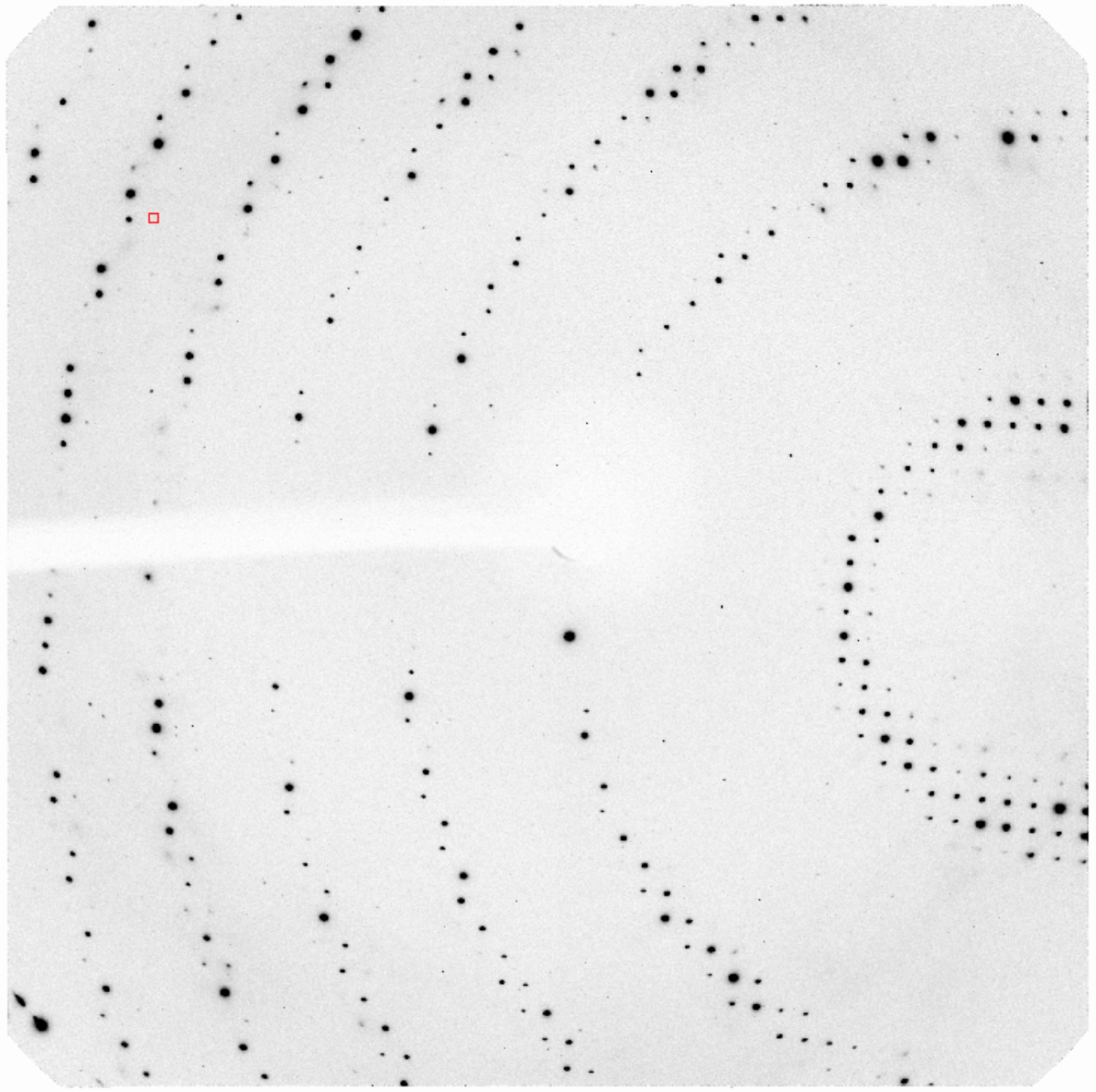
DIFFRACTION PATTERN



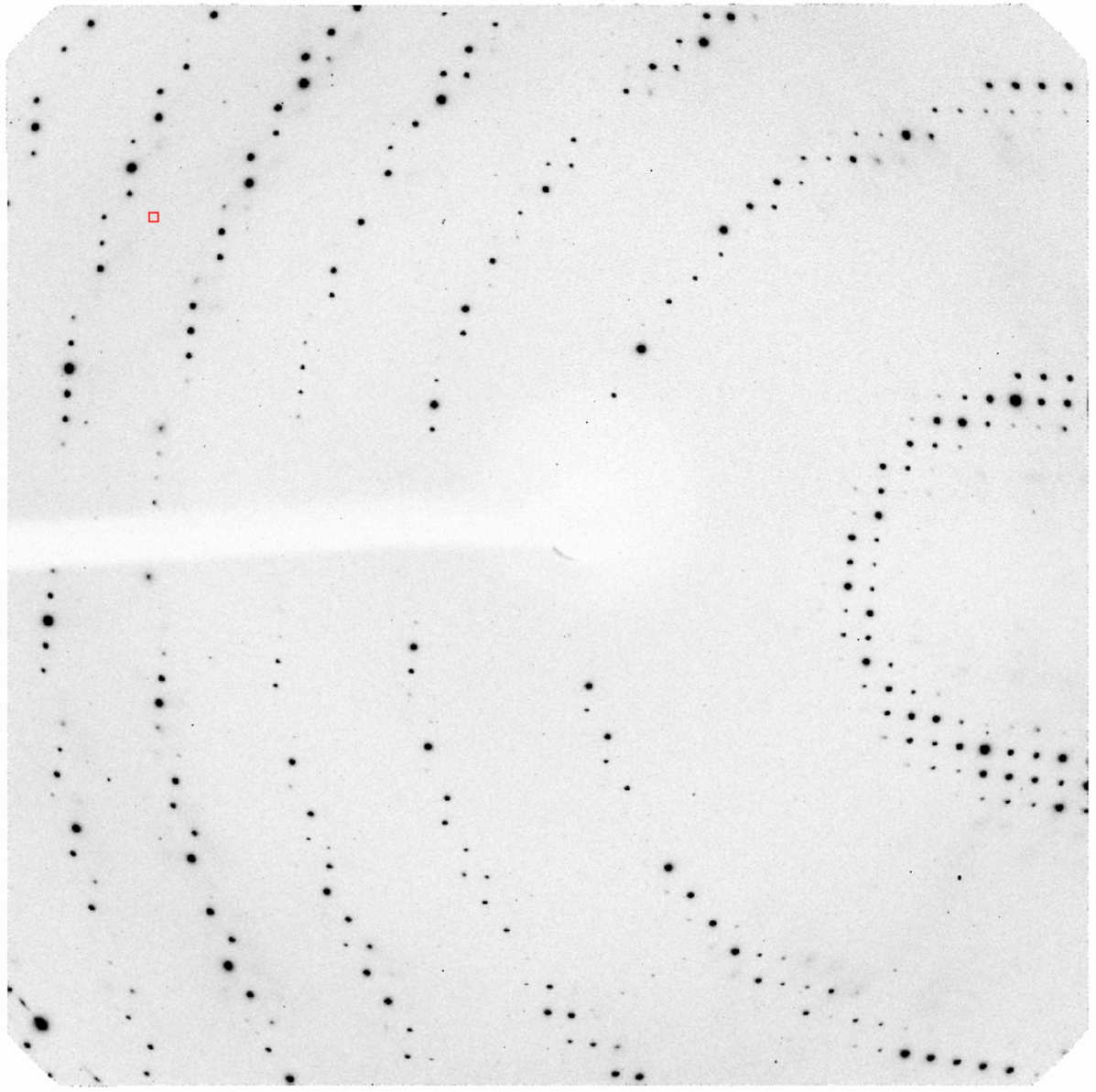
DIFFRACTION PATTERN



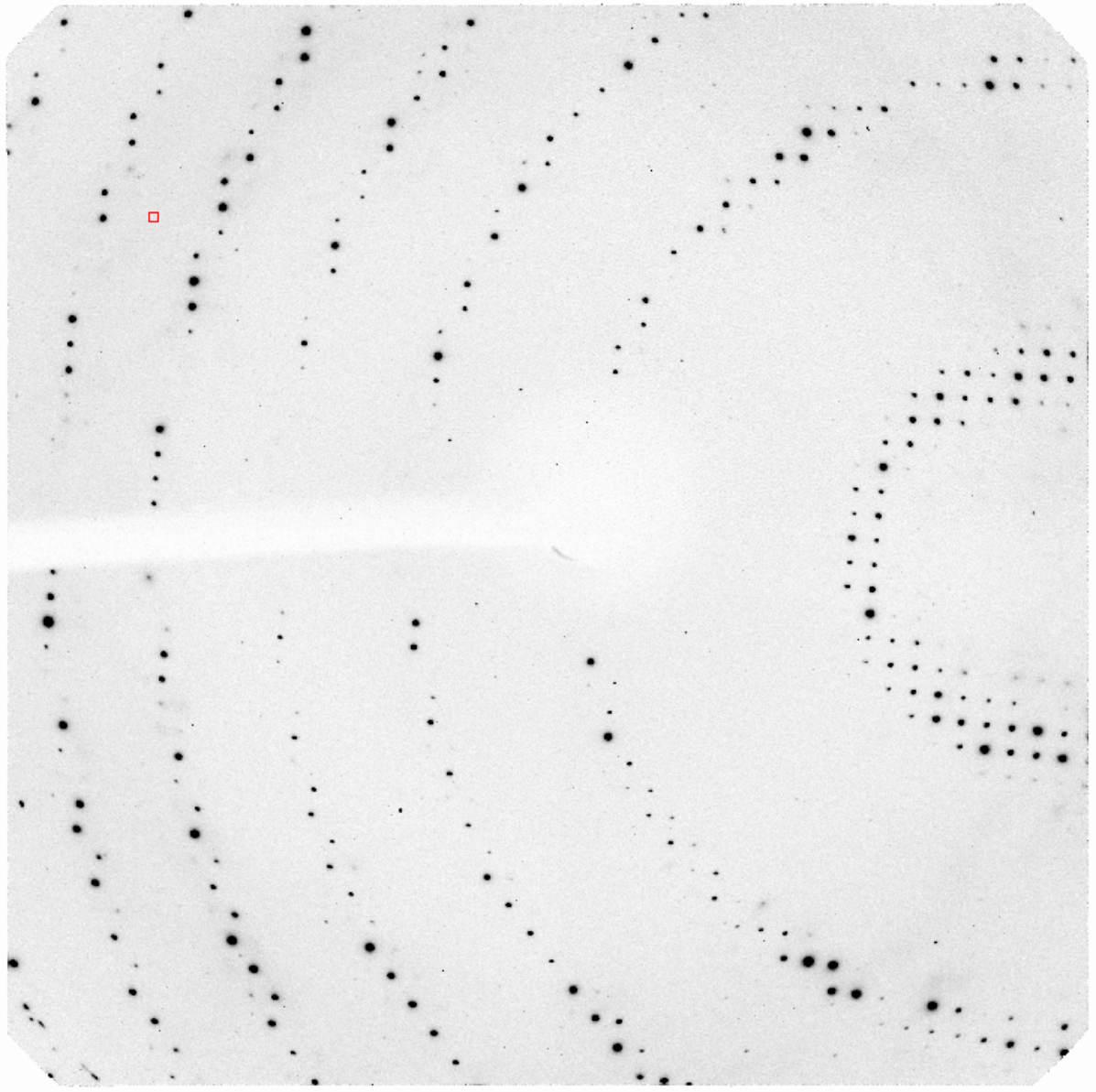
DIFFRACTION PATTERN



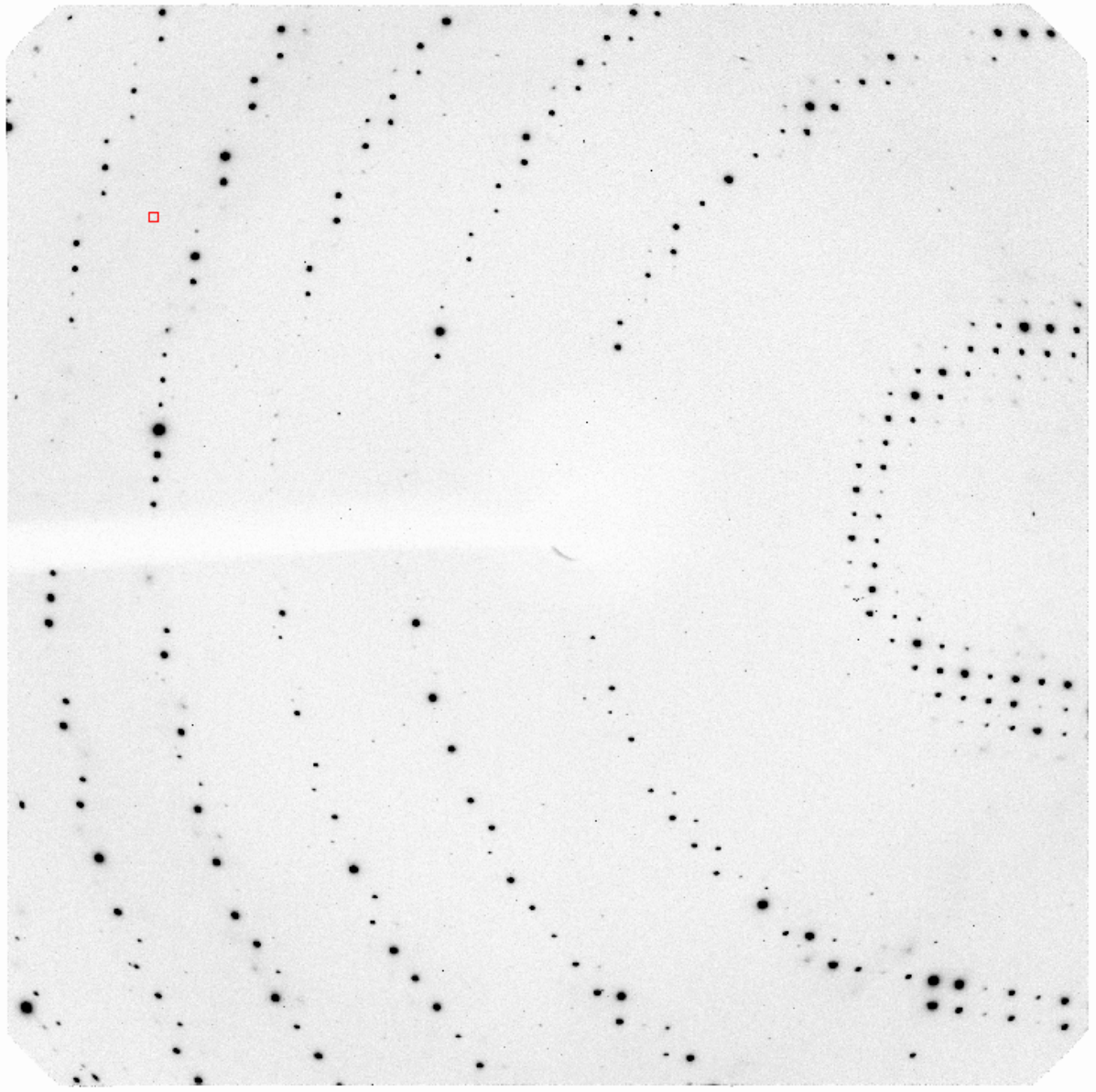
DIFFRACTION PATTERN



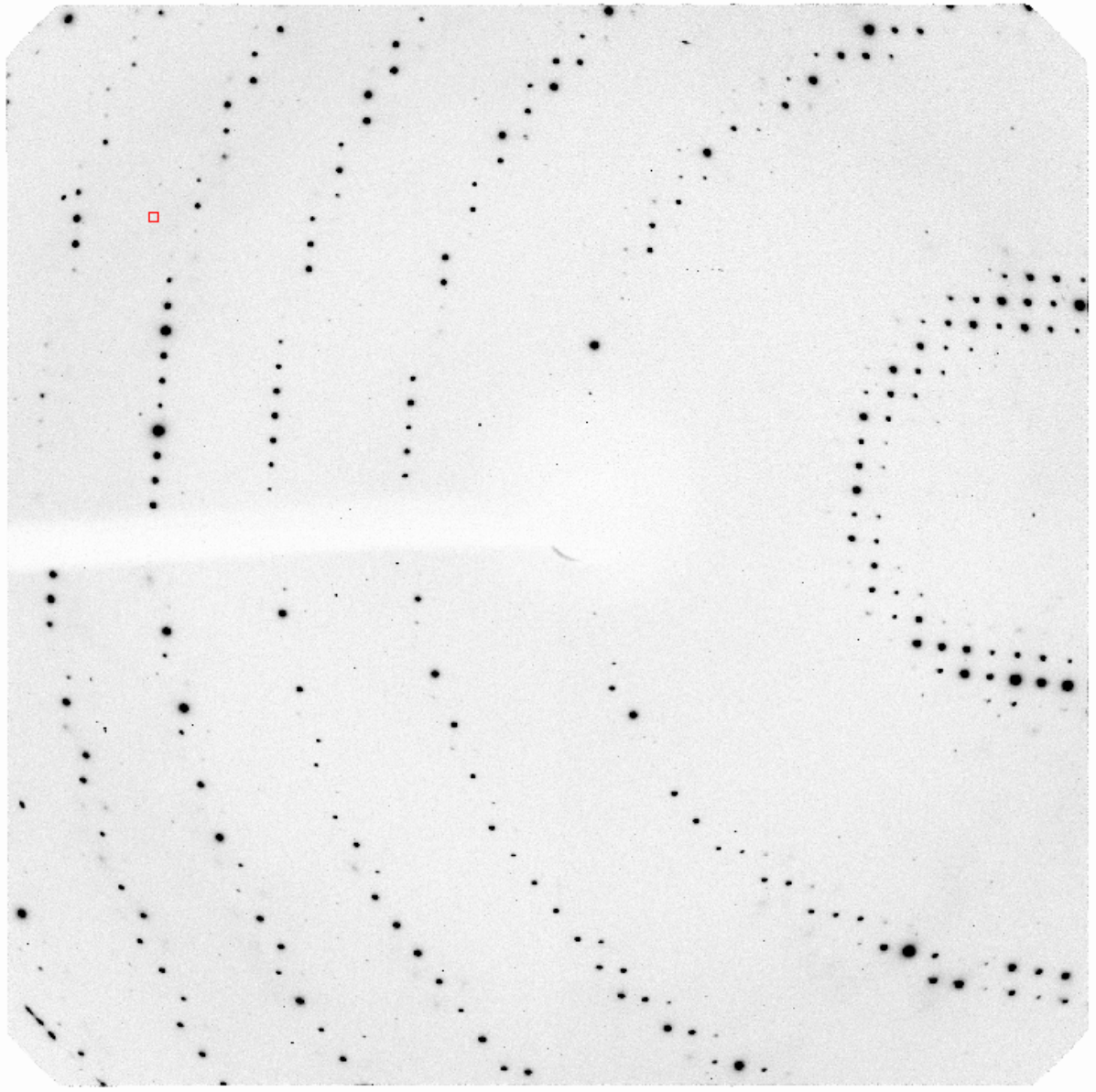
DIFFRACTION PATTERN



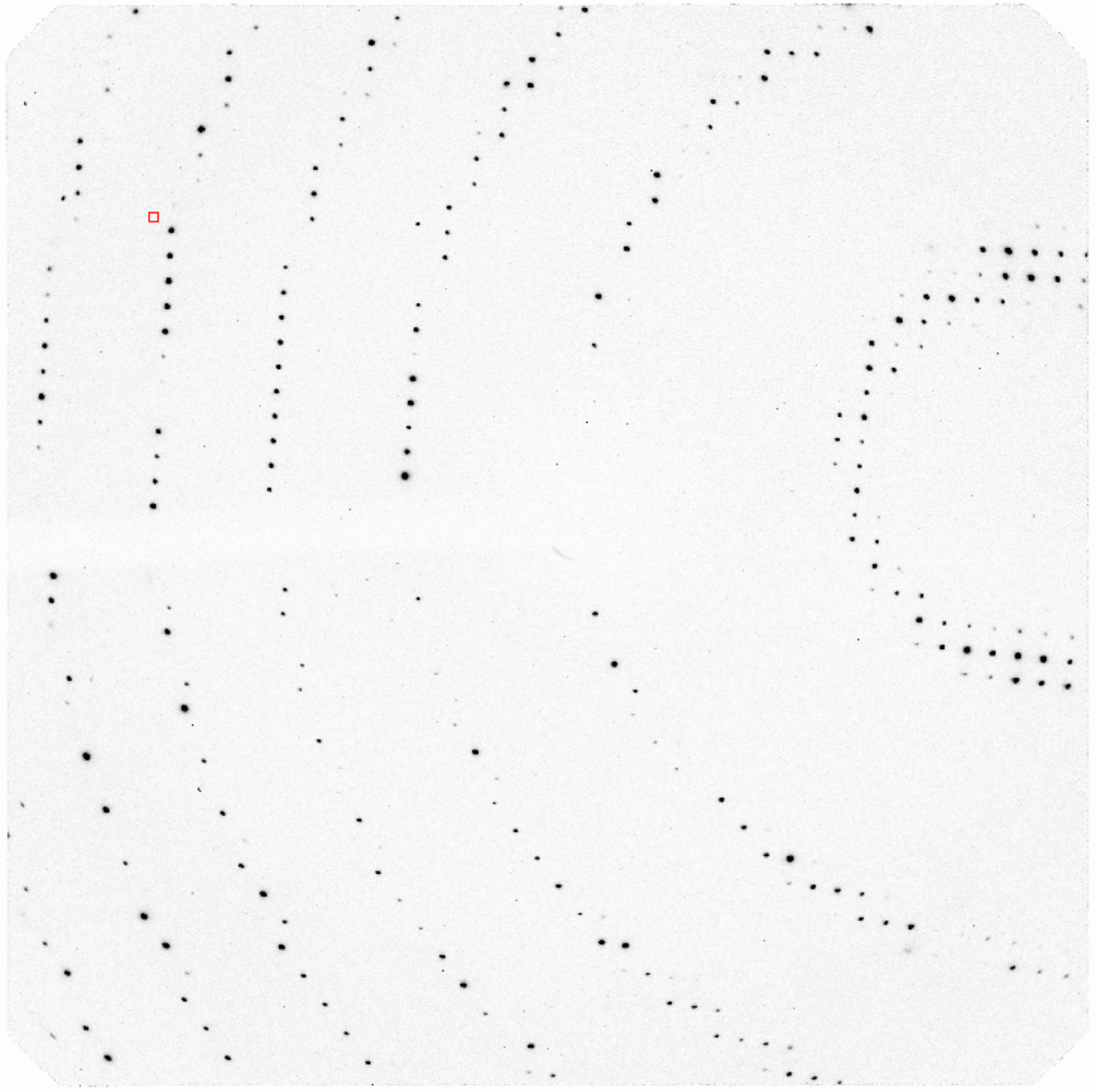
DIFFRACTION PATTERN



DIFFRACTION PATTERN



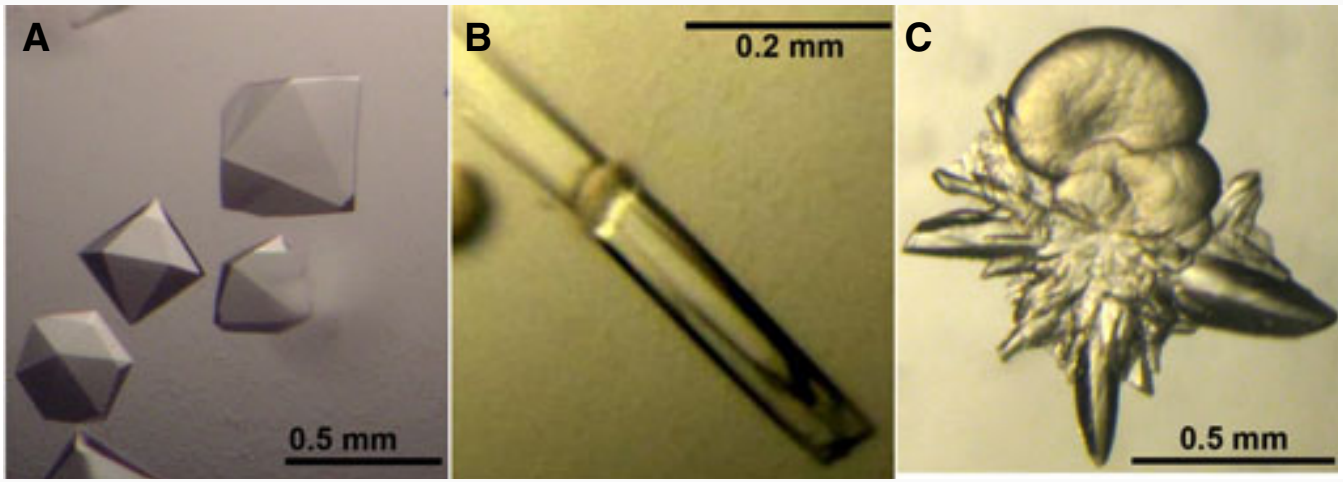
DIFFRACTION PATTERN



What happened during the experiment?

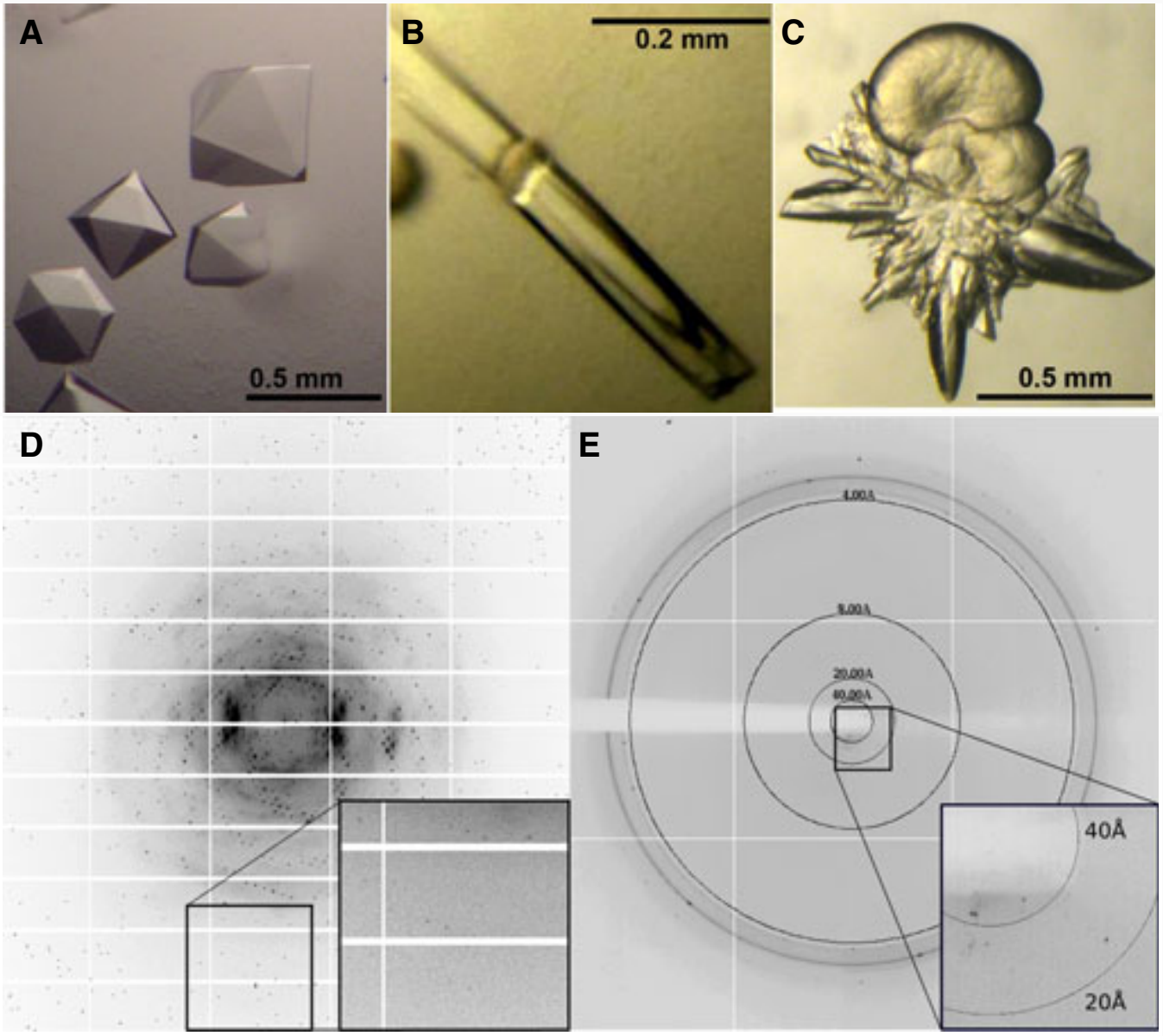
WHAT CAN GO RIGHT OR WRONG?

DIFFRACTION FROM CRYSTALS.



Which of these crystals give the best diffraction?

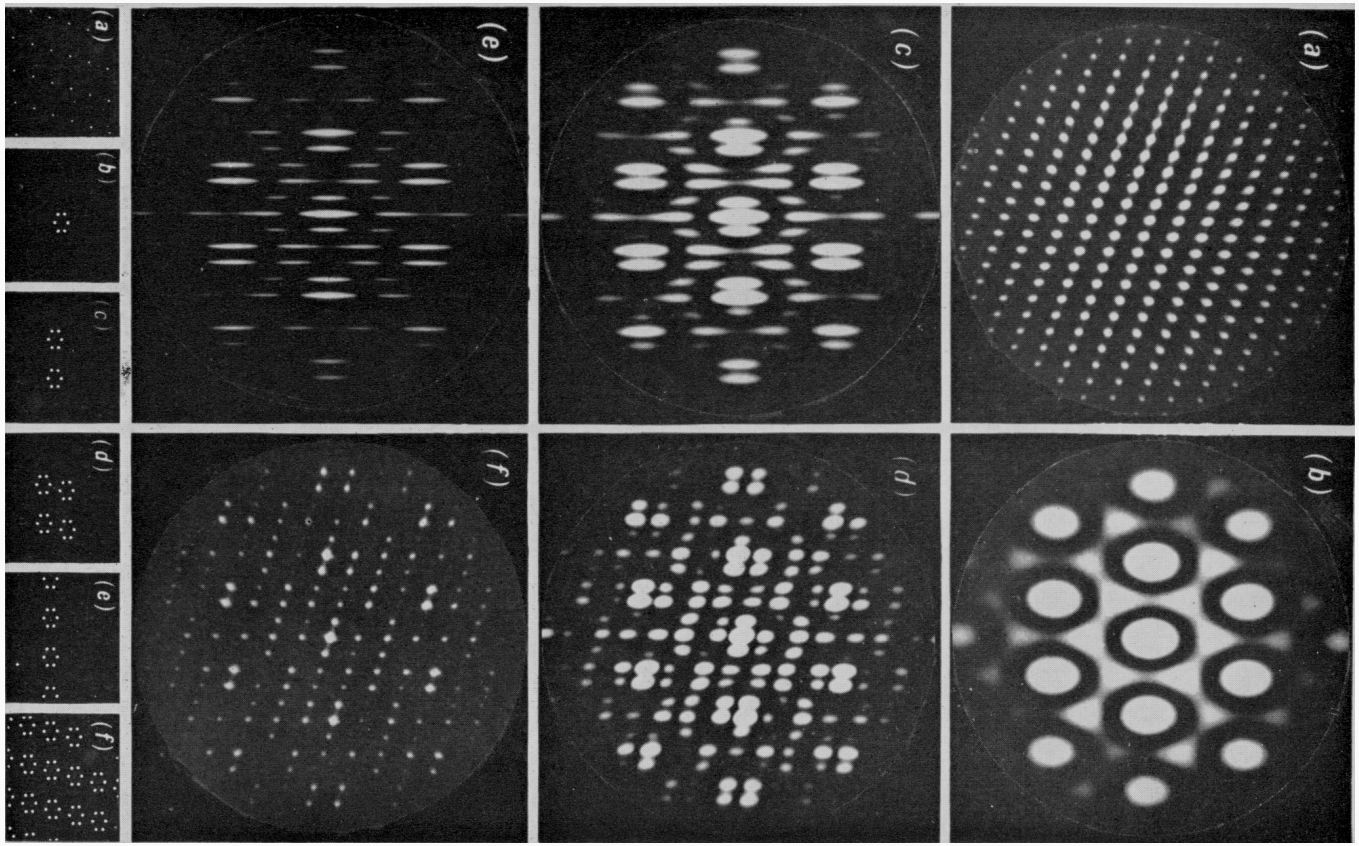
DIFFRACTION FROM CRYSTALS.



An orthorhombic crystal form of a protein

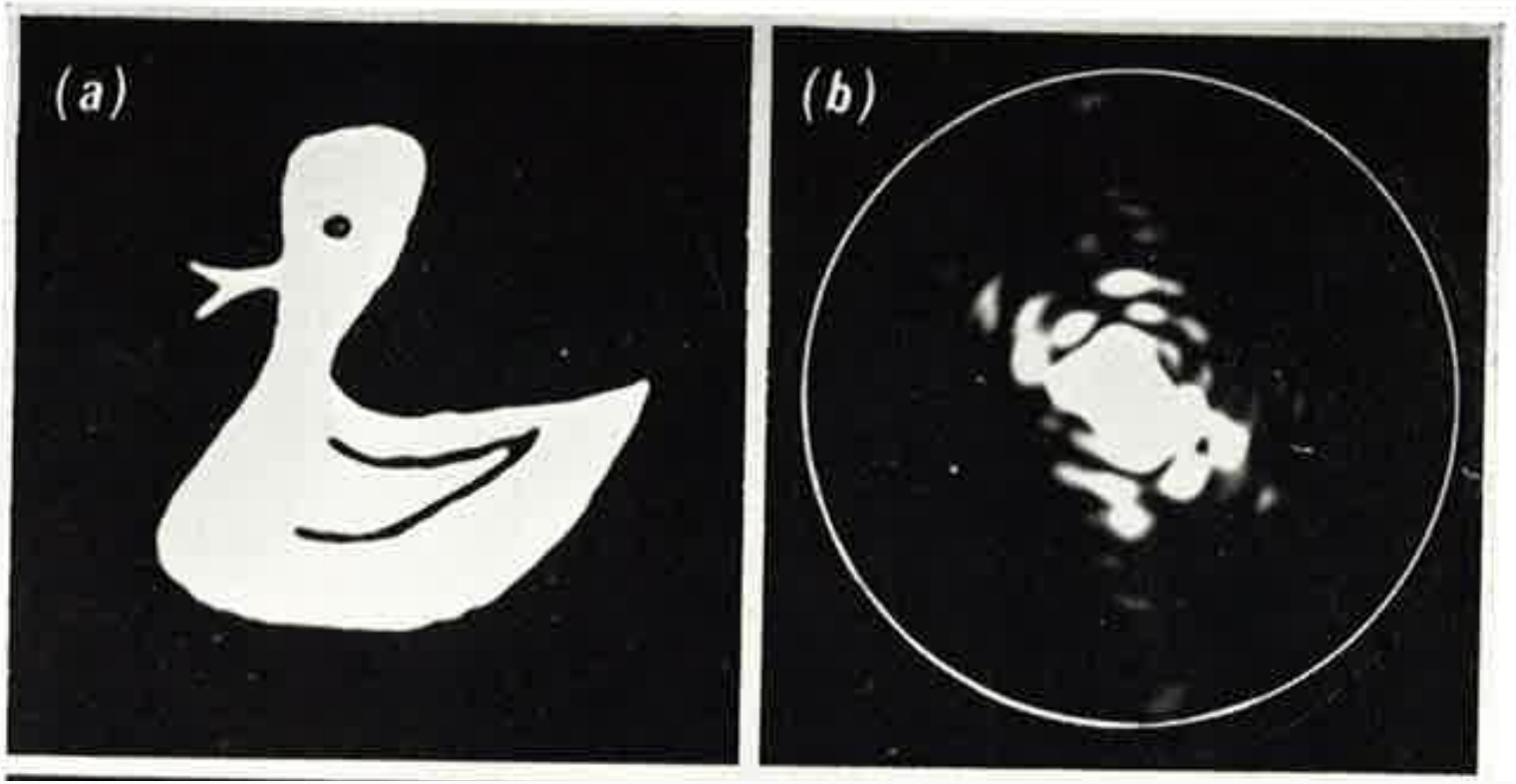
- Has a unit cell of 45 Å, 55 Å, 50 Å.
- How many unit cells will be illuminated by a 10μm x-ray beam?
- suppose the unit cell dimensions are 250Å, 255Å, 350Å and the beam 1μm ?

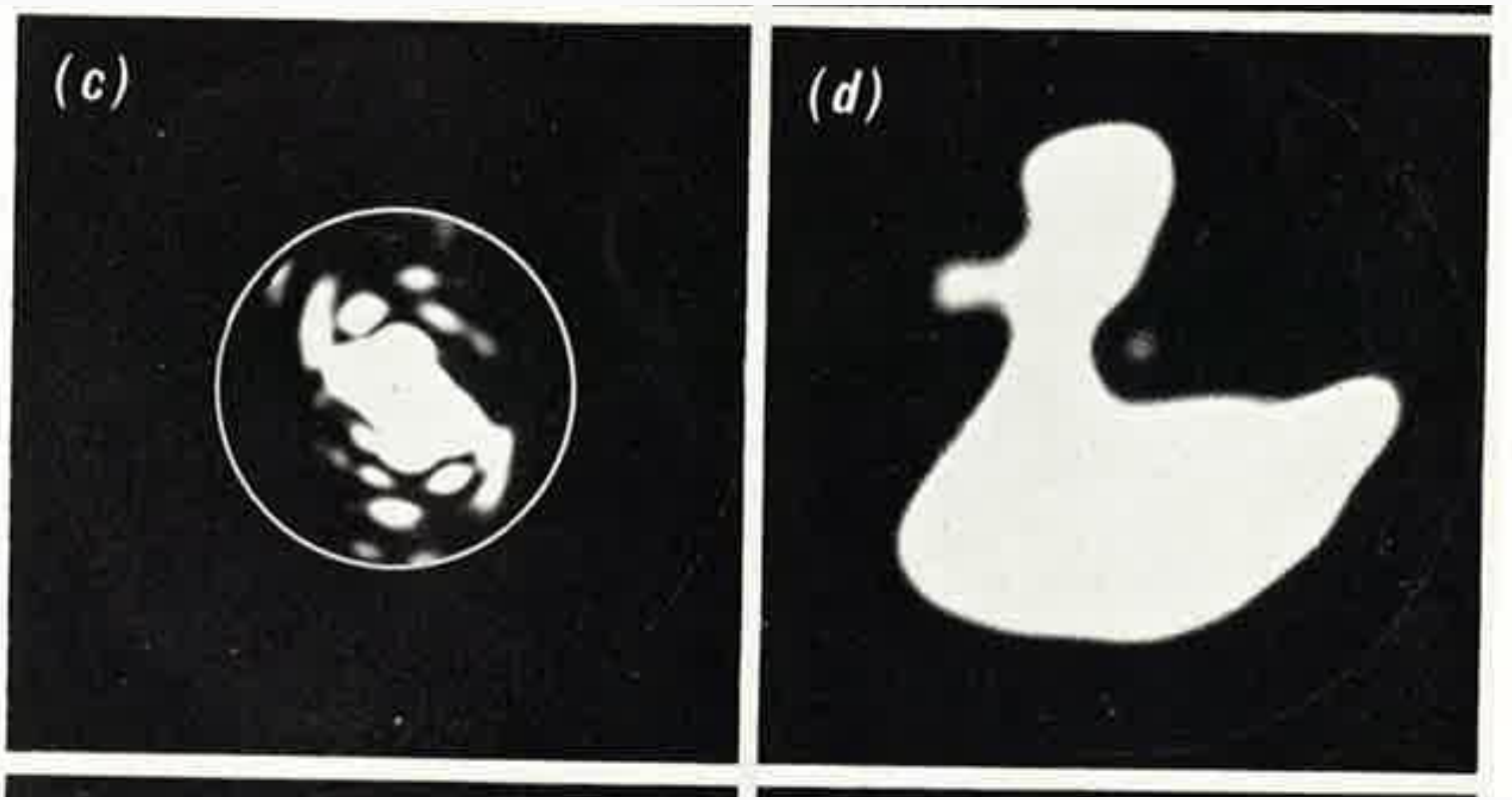
RELATIONSHIP MOLECULE TO CELL TO DIFFRACTION.



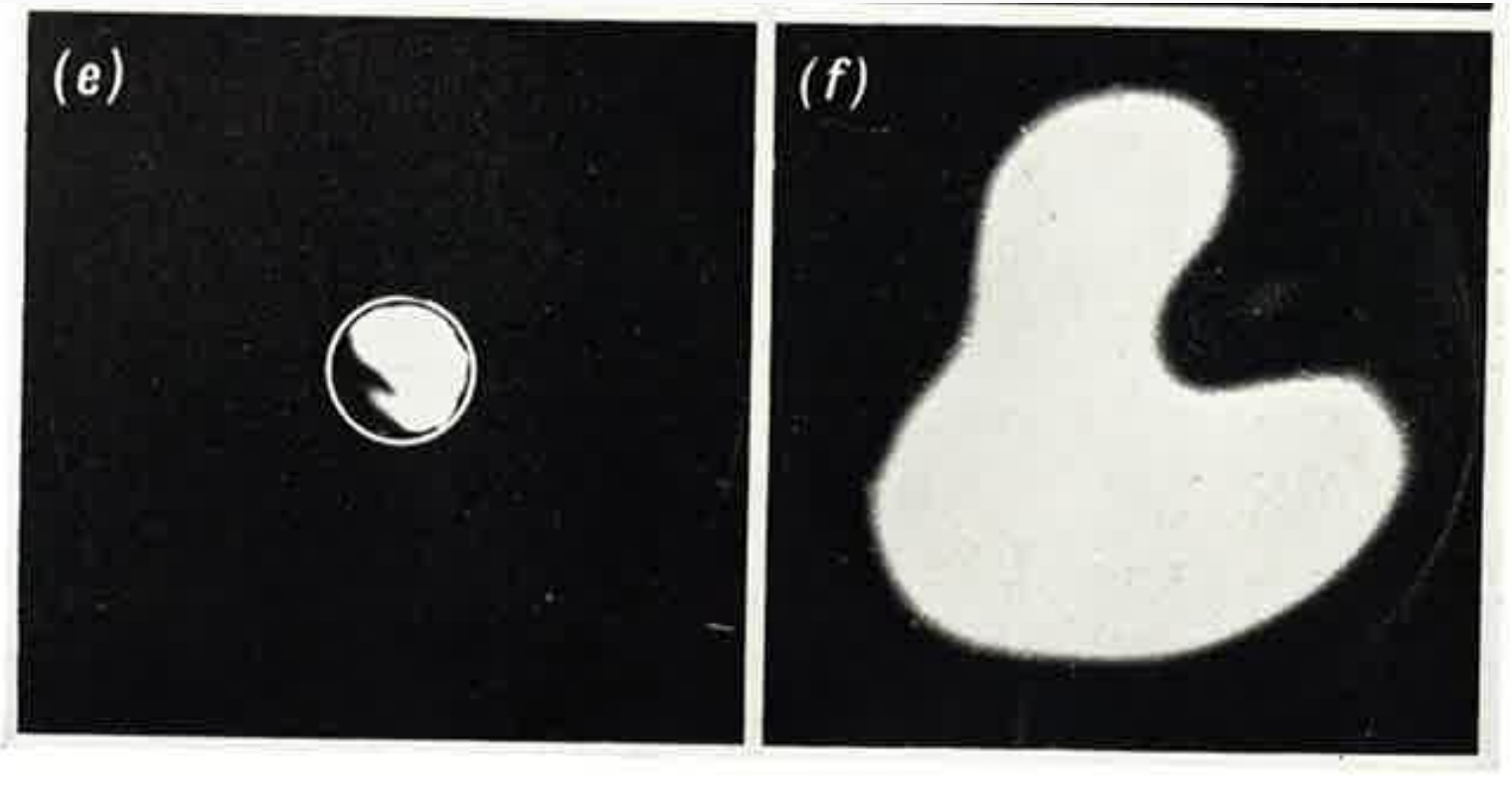
Notice (1) The symmetry, and (2) how the continuous diffraction pattern of one molecule (b) is “sampled” by the lattice of diffraction points.

DATA COMPLETENESS



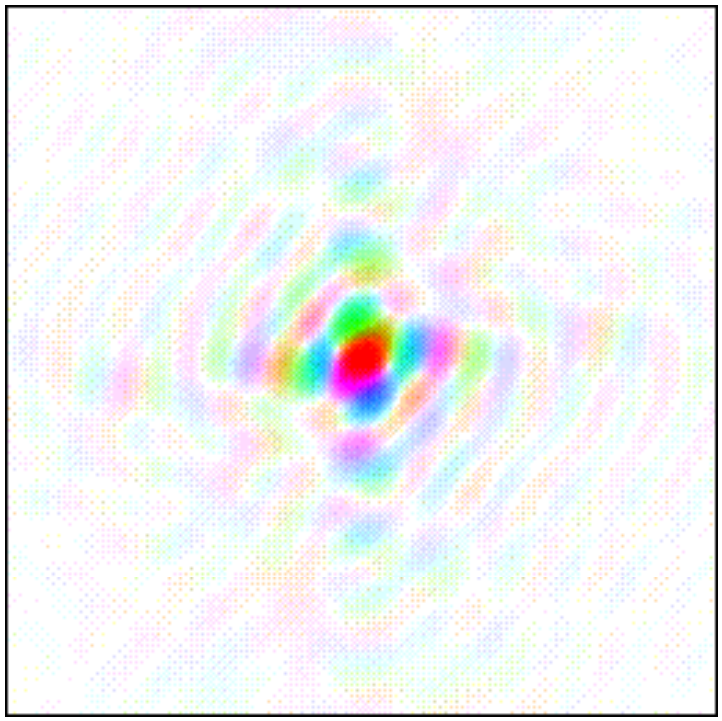

(c)
(d)

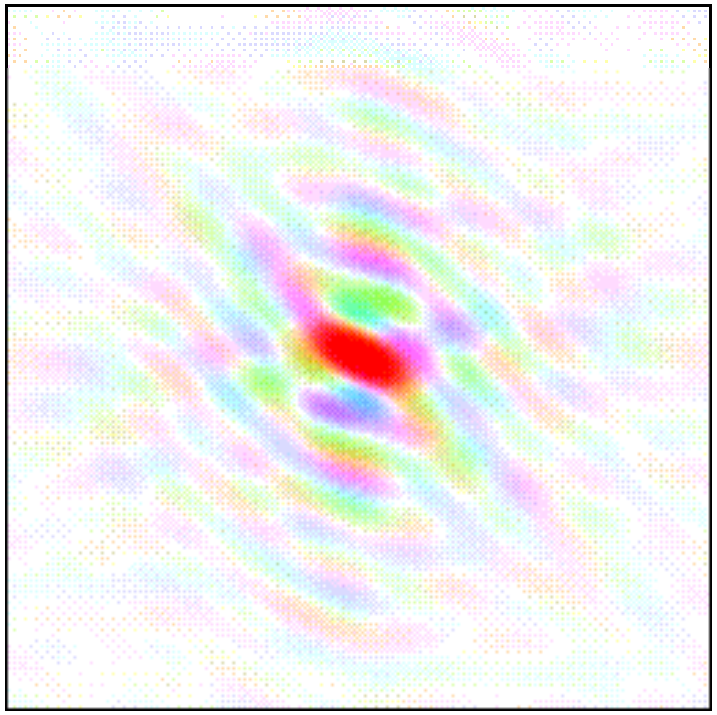
DATA COMPLETENESS



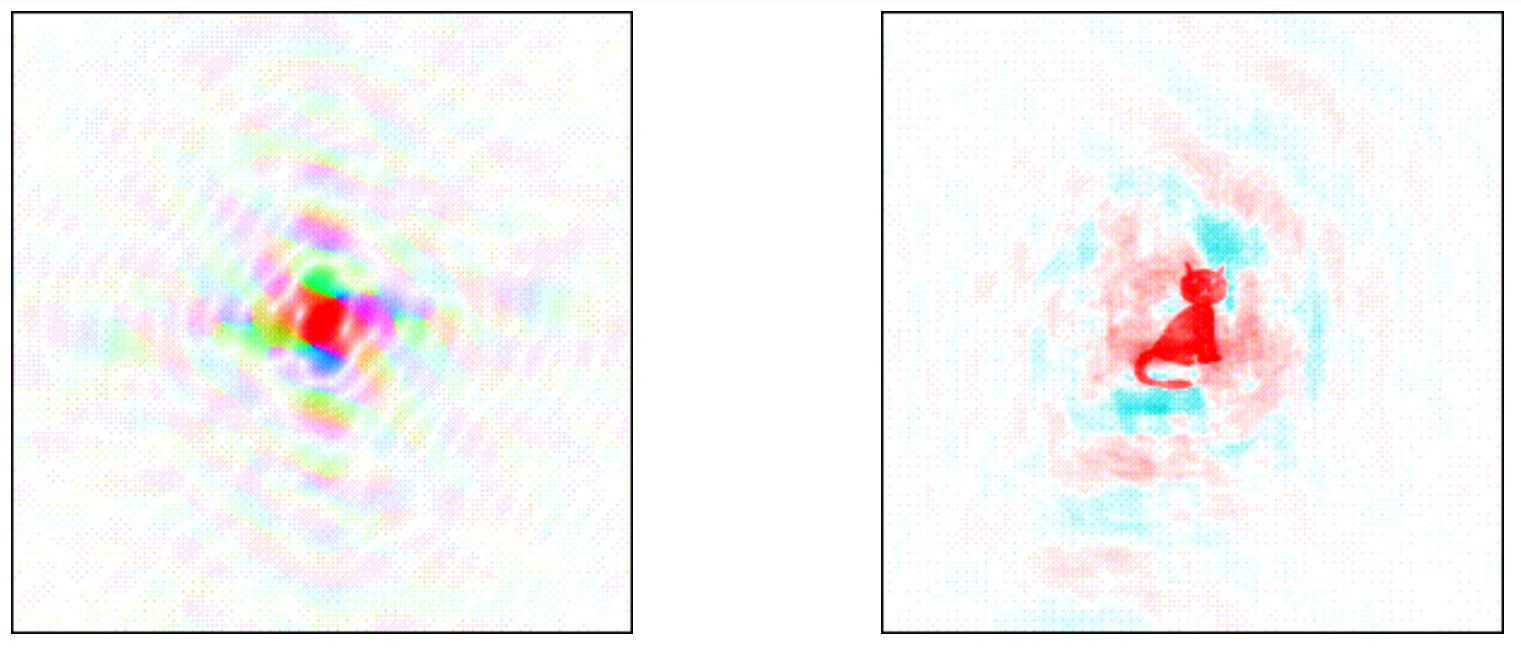
Which is more important *intensity* or *phase*? Why?

DUCKS AND CATS





PHASE MAGIC, MIXING CATS AND DUCKS.



Note

The image which contributed the phases is still visible, whereas the image which contributed the magnitudes has gone

[6]

[6] http://www.ysbl.york.ac.uk/ cowtan/fourier/magic.html

The crystallographic experiment.

In a X-ray diffraction experiment, we collect only the diffraction magnitudes, and not the phases. Unfortunately the phases contain the bulk of the structural information. So in addition to collection all the intensities possible we need to estimate the phase as accurately as possible.

RESOLVING THE PHASE PROBLEM.

- If there exists a structure that is *similar enough*
- Position it within the unit cell to get best agreement with the x-ray data.
- Calculate phases from the model and combine with the experimental intensities.

What could go wrong?

Two principle methods:

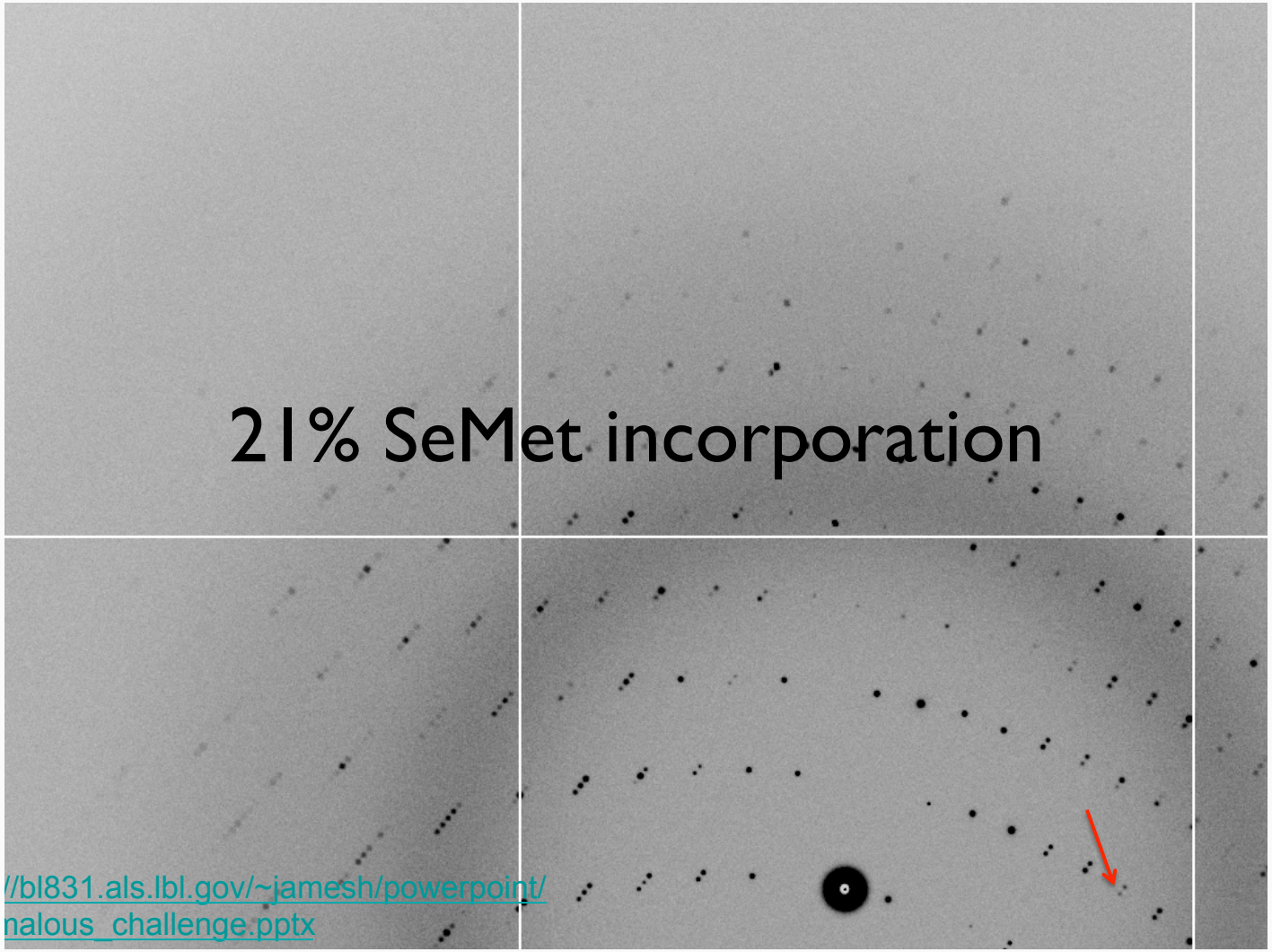
- Multiple (or single) isomporphous replacement.
- Anomalous dispersion methods.

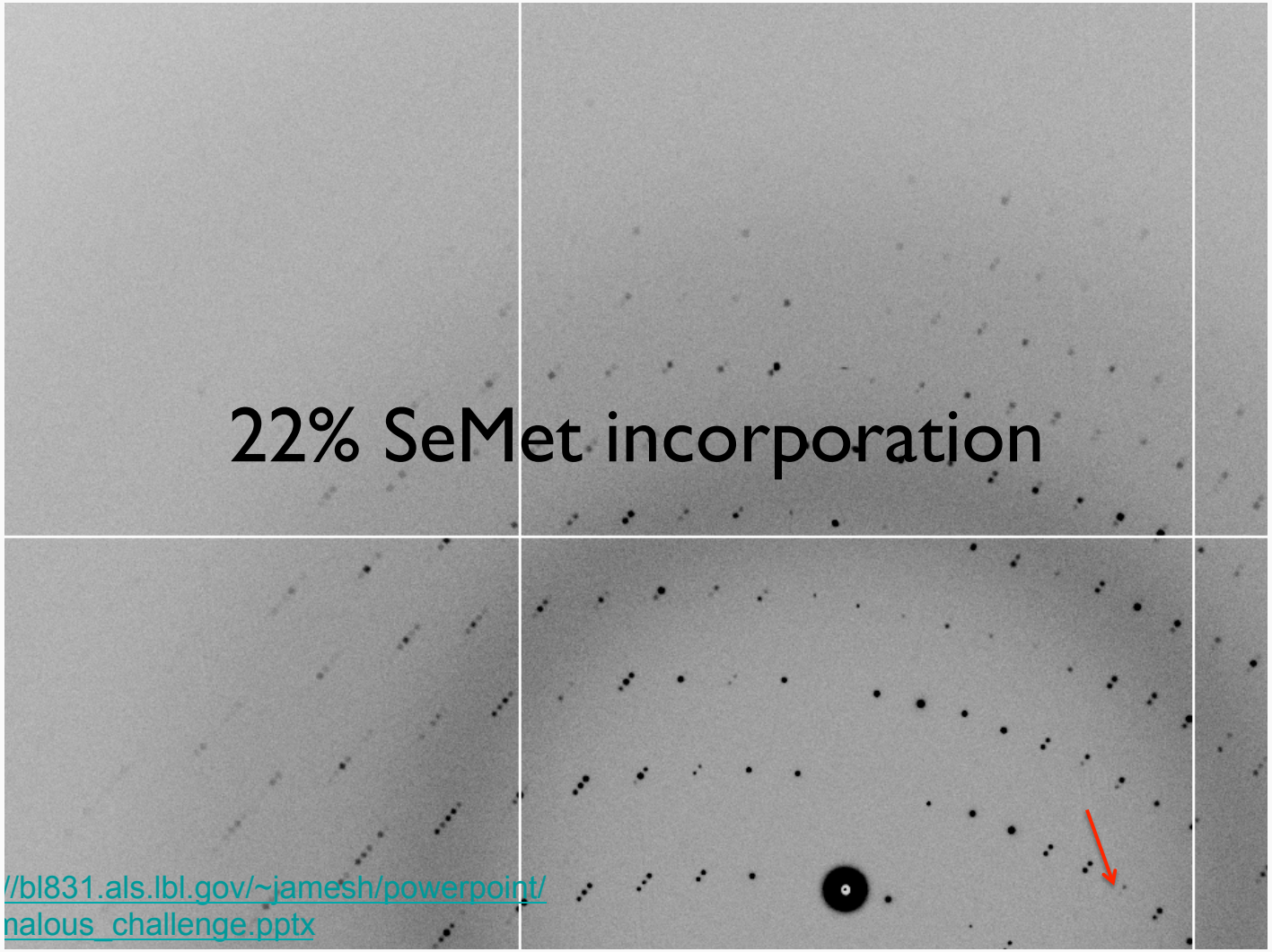
The basic idea

- Add another atom that has enough electrons to significantly change the scattering.
- If the addition doesn't change the cell (*too much*) or the molecule's orientation (*too much*)
- Compare the data with and without the "*heavy atom*"
- The position of this heavy atom allows an estimate of its phase to be made.
- Allows estimation of the protein phases.

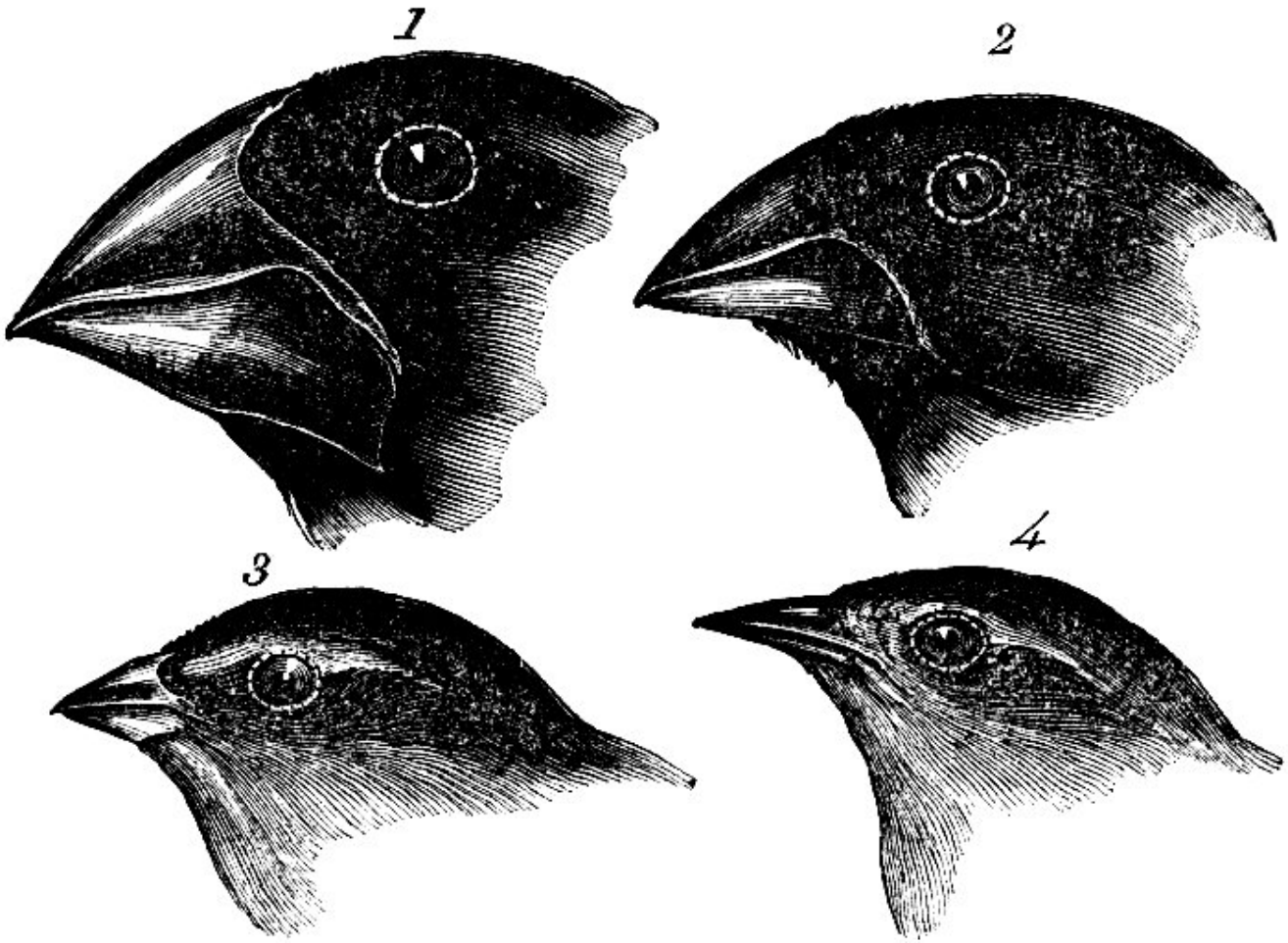
The atomistic scattering factor is wavelength dependent.

- scattering factor of an atom depends on energy - absorption edges...
- suitable changes can be measured - in general the changes are small.
- symmetry related reflections will be different: *Bijvoet differences*
- changes of wavelength will induce “significant” change in scattering for some atoms: *Dispersive diferences*
- The chief problem is that the signal is at about the level of the noise in the data.


21% SeMet incorporation
//bl831.als.lbl.gov/~jamesh/powerpoint/
nalous_challenge.pptx

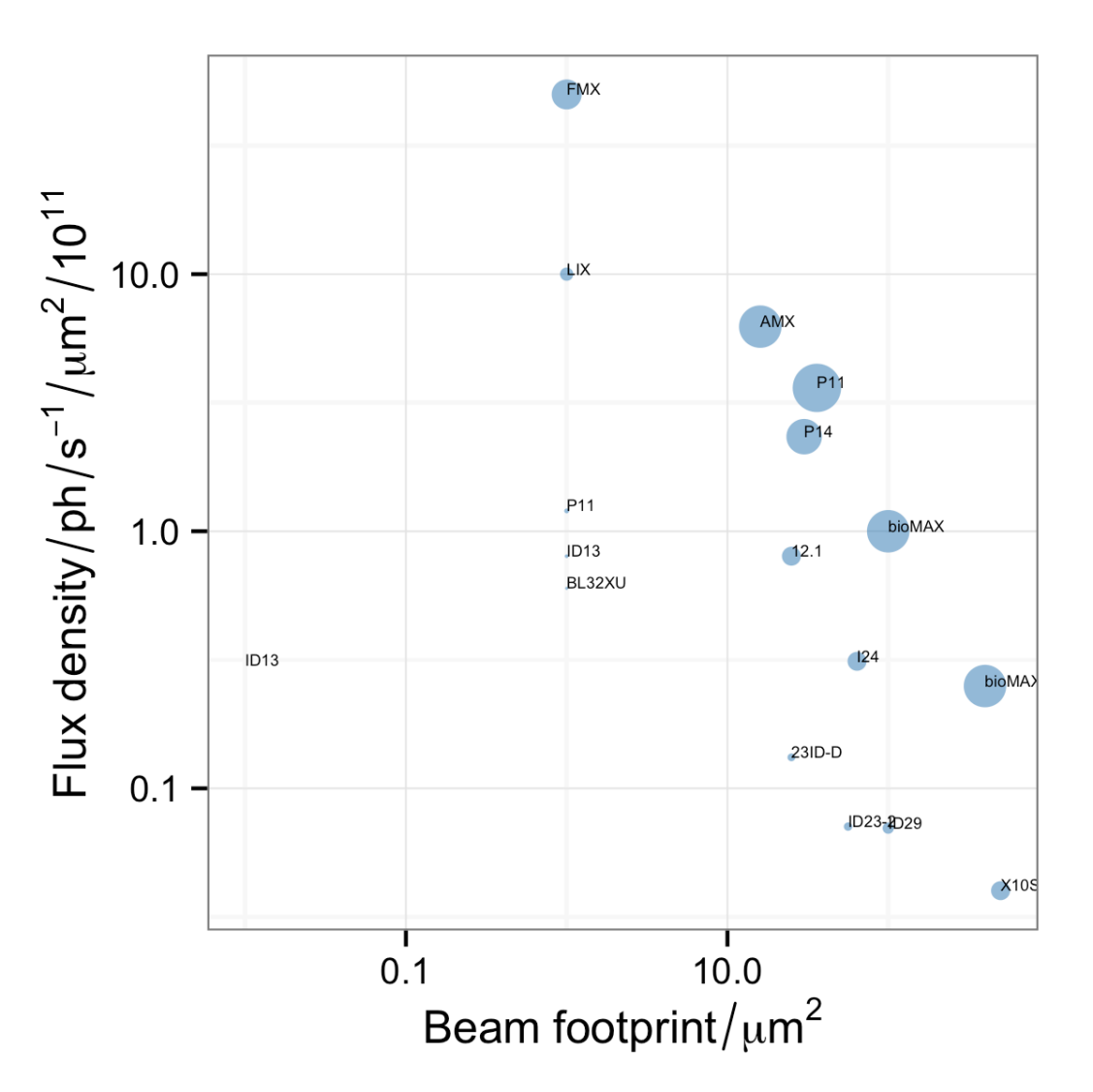


//bl831.als.lbl.gov/~jamesh/powerpoint/
nalous_challenge.pptx



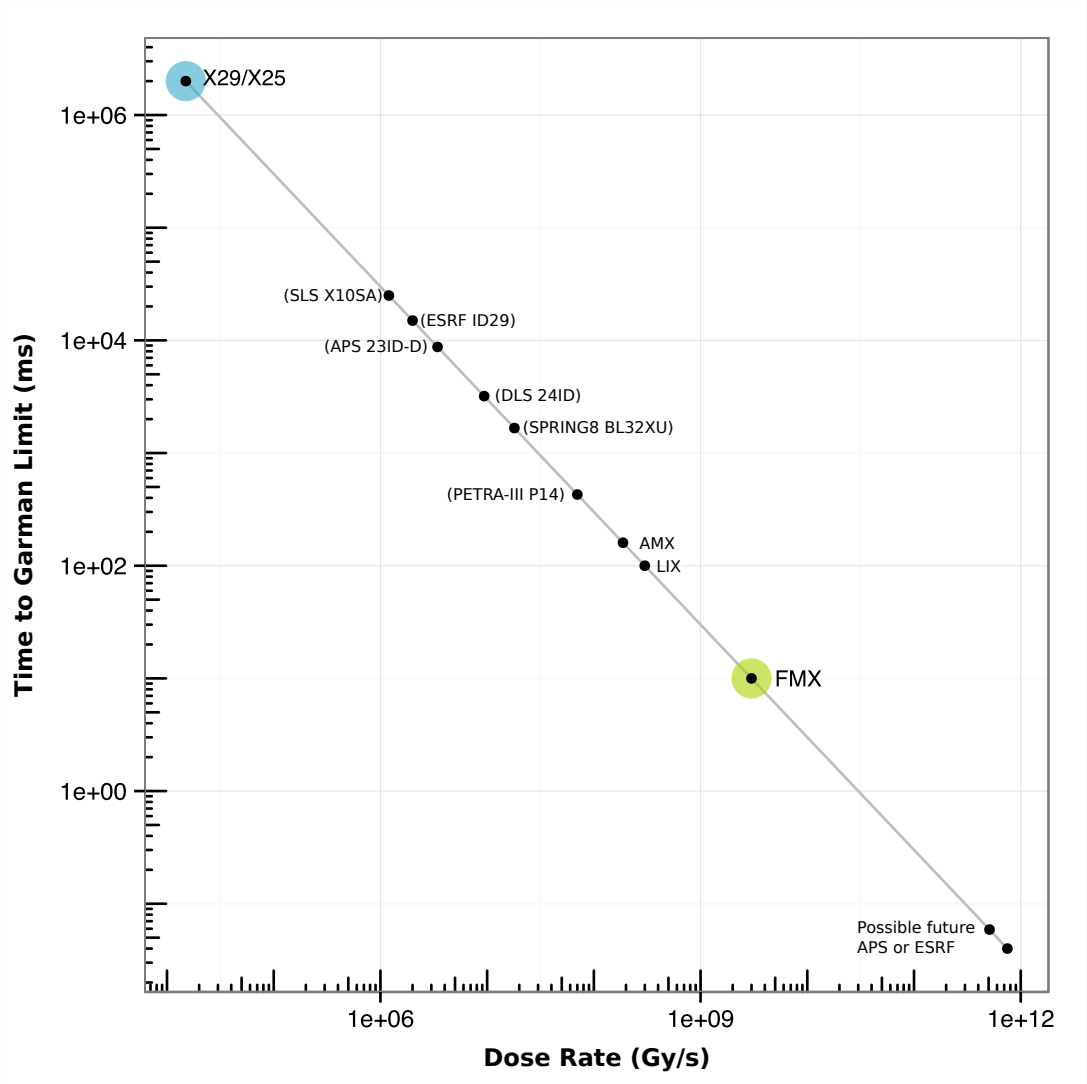
1. Geospiza magnirostris.
2. Geospiza fortis.
3. Geospiza parvula.
4. Certhidea olivacea.

Making use of the potential at NSLS-II.

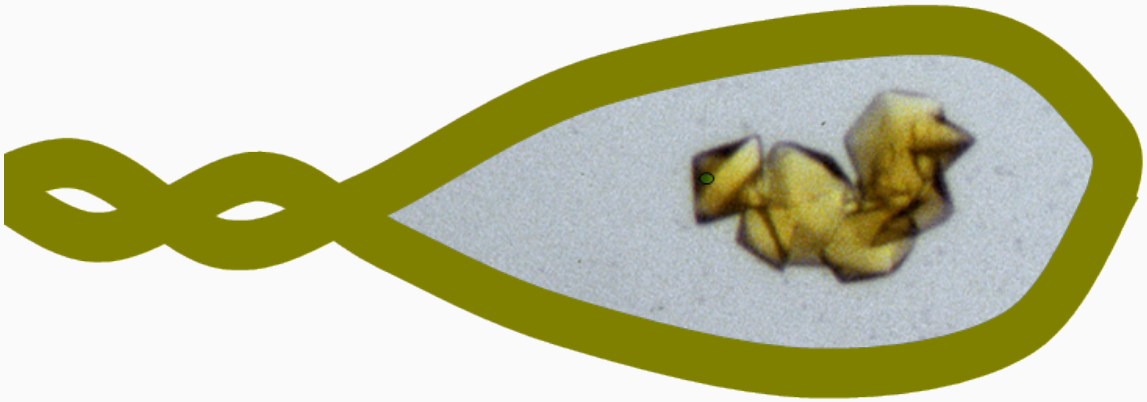


WHERE ONE OF THE NEW CHALLENGES LIES

Making use of the potential at NSLS-II.



THIS WILL ALSO BE ONE.



Making small beams is now “easy”. But **You** need to know how to use it!

Theoretical Dose Limit.

- Dose is measured in Gy = J/Kg
- Henderson: $2x10^7$Gy
- Garman: $4.3x10^7$Gy

ref

Henderson(1999) Proc. R. Soc. Lond. **B241**, 6-8
Owen, Rudiño-Piñera and Garman PNAS(2006), **103**, 4912–17

INDICATIONS OF RADIATION DAMAGE.

- Mosaicity (generally increases)
- Discrepancy between (partial) data sets increases.
- Noise in data sets increases.
- Specific structural changes or disorder will be observed in maps.
- The unit cell changes
- $I/\sigma(I)$
- resolution limit of the data.

ATOMIC MOVEMENTS.

- Damage is specific.
- disulphide bonds break.
- Decarboxylation of acidic amino acids.
- Active sites are disrupted.
- Tyrosine hydroxyl group lost.

This is not primary damage alone! Secondary events including diffusion are implicated.

Examples include:

- Incomplete data from the crystal or part of.
- Specific structural changes.
- Wrong Biological interpretation.
- Structure determination fails. Probably due to structural changes **during the experiment** Leading to non–isomorphism within the data.

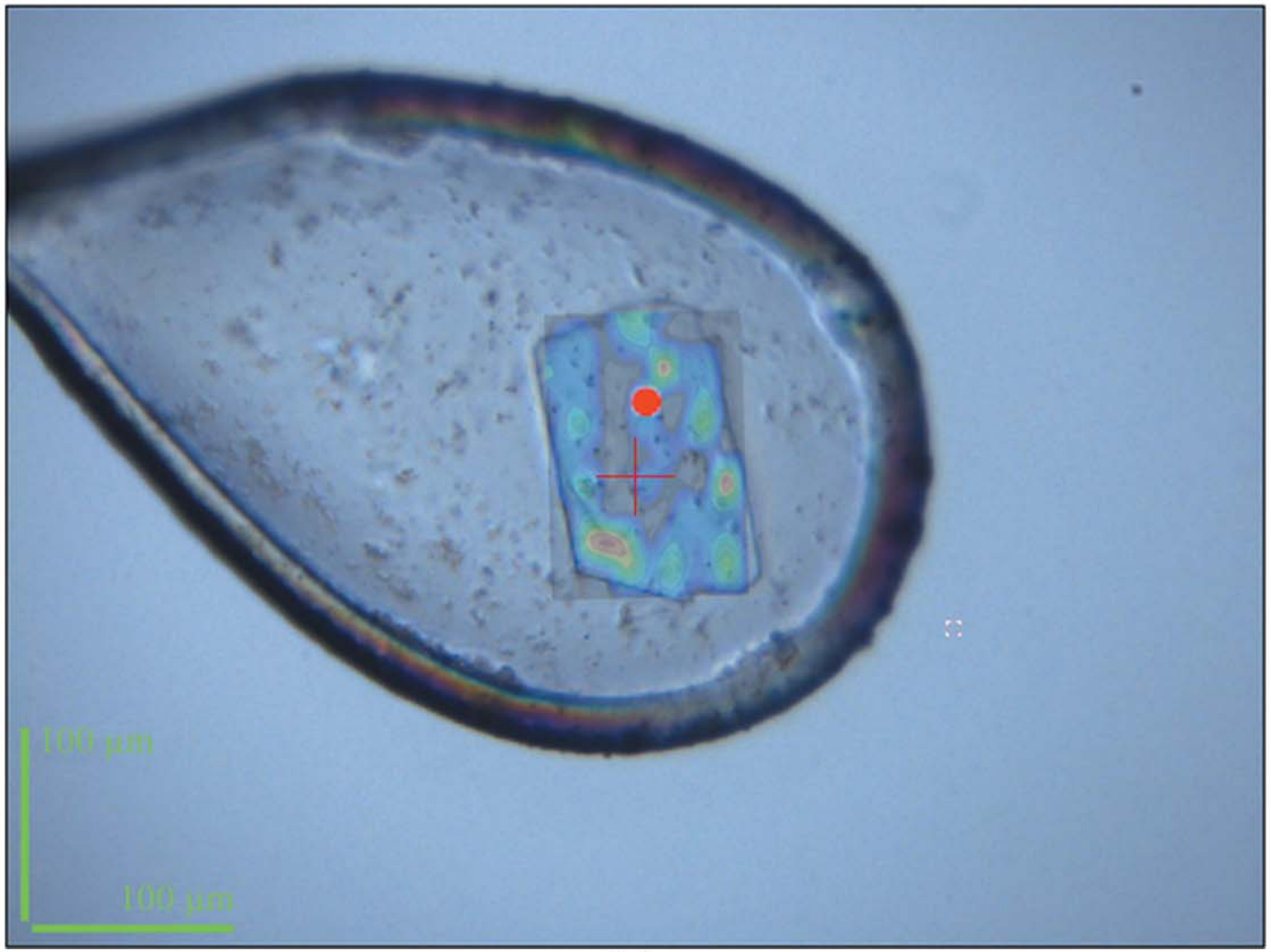
My opinion is

- We are far from exploiting the potential of modern sources.
- Beamlines need to improve the precision with which we can measure diffraction data.
- Detectors need better correction and uniformity.
- Better understanding of isomorphism (esp. in partial datasets)
- We need to perform “better” experiments.
- Analysis software will need to improve.
- Automation will make some of these problems tractable.

SUMMARY

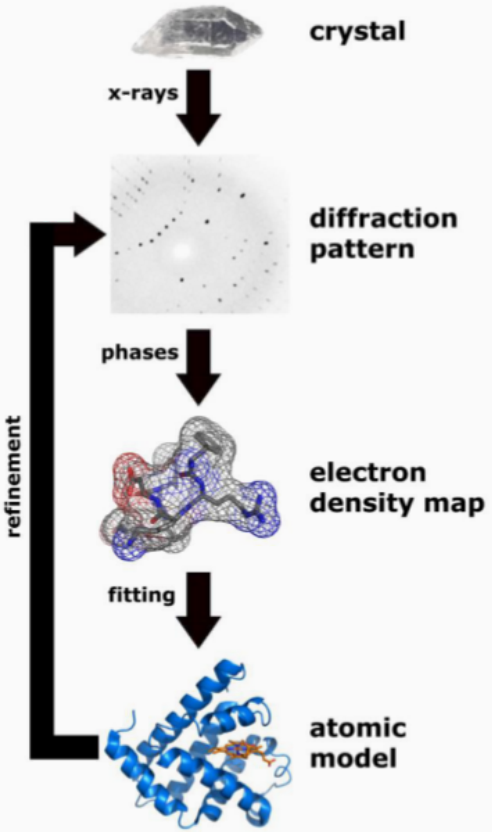
$$\rho(x,y,z) = \frac{1}{V}\sum_{h=-\infty}^{\infty}\sum_{k}\sum_{l} F_{hkl}\exp\left[-2\pi i\left(hx_j + ky_j + lz_j\right)\right]$$

- Solve the electron density equation which reveals the contents of the crystal.
- The diffraction pattern provides data on the positions and intensities of the reflections.
- Intensities $I_{(hkl)}$ are proportional to the square of the structure factor magnitudes $\left|F_{(hkl)}\right|^2$
- $F_{(hkl)}$ is the vector sum of the scattering factors of all the atoms in the unit cell.



(c)

SUMMARY



Measured intensity

$$I_{hkl} \propto |F_{hkl}|^2$$

Electron (X-ray) or nuclear (neutron) density at point x,y,z in the unit cell

$$\rho(xyz) = \frac{1}{V}\sum_{hkl} F_{hkl}\, e^{-2\pi i\,(hx+ky+lz)}$$

Phase angle

$$F_{hkl} = |F_{hkl}|e^{-i\phi} = |F_{hkl}|\cos\phi + i|F_{hkl}|\sin\phi = A + iB$$

$$\phi = \tan^{-1}\frac{B}{A}$$

$$F_{hkl} = \int_{cell} \rho_{xyz}\, e^{2\pi i(\boldsymbol{s}\cdot\boldsymbol{r})} d\boldsymbol{v} = \sum_{j} b_j\, e^{2\pi i\,(hx_j+ky_j+lz_j)}$$

Sum over j atoms in the unit cell

Neutron scattering length or X-ray form factor for j^{th} atom

DAVID S. GOODSELL, 2002