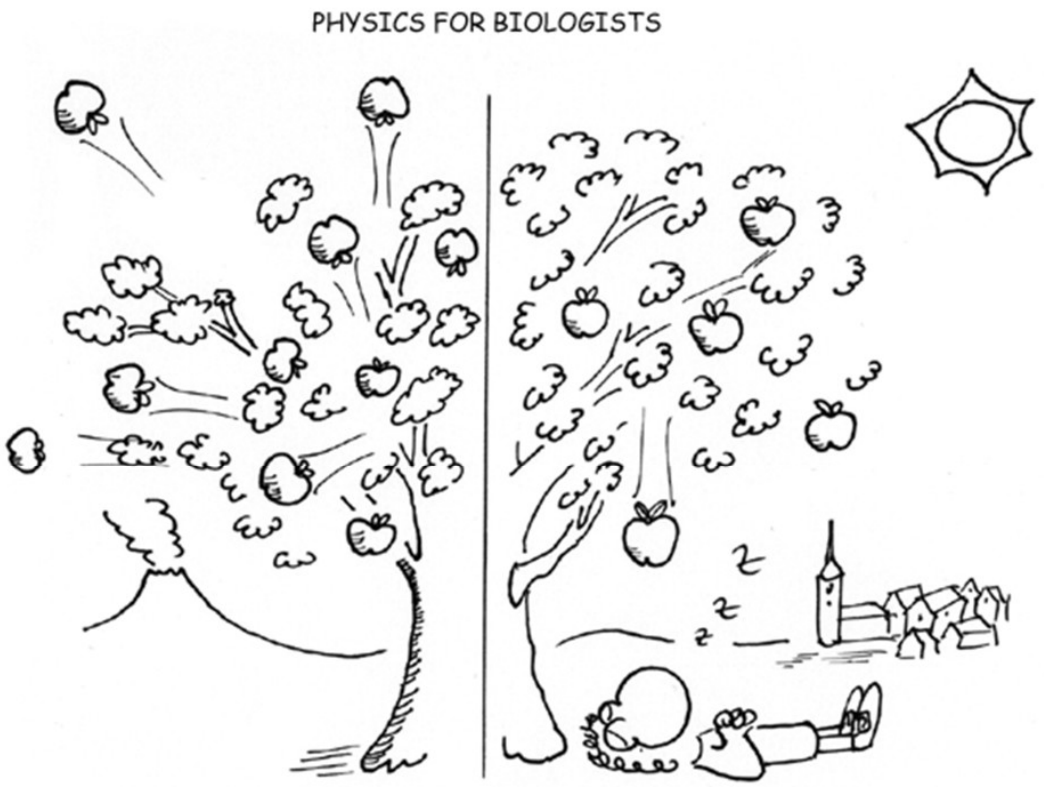
Structural Biology Techniques

Practical Structural Biology from Crystals to Cryo-EM

Edward H. Snell

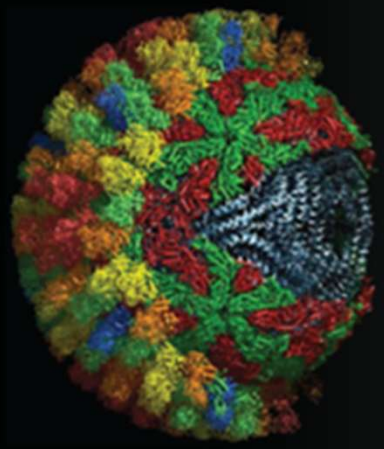
University at Buffalo Hauptman Woodward Research Institute, Buffalo, NY 14203, USA.

esnell@buffalo.edu



A long time ago the apple trees used to shoot the apples in all directions. Only those that did it downward got reproduced. Then, after millions years of natural selection and evolution, gravity was finally discovered.

www.thescientificcartoonist.com

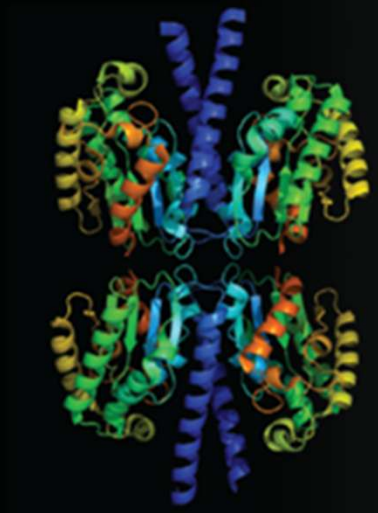


X-Ray Crystallography

sample must be crystallized in a solid frozen structure

any size macromolecule

atomic resolution but crystallization may take years and damage protein structure

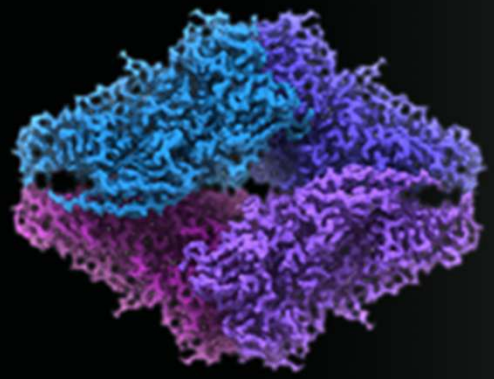


Nuclear Magnetic Resonance

sample must be dissolved and radioactively labelled

small molecules

closer to real protein structure but larger proteins can not be resolved



Cryo-Electron Microscopy

sample is frozen in its native state

any size macromolecule

near-atomic resolution, fast sample preparation

The goal of structural biology

- Structural biology is like trying to understand how a complex machine works, not by reading the manual, but by carefully examining every cog, lever, and wire, and watching how they move and interact.
- In other words, the goal of structural biology is to determine the shapes of biological molecules, e.g. proteins, nucleic acids, complexes, and understand how their form enables their function, and how they interact with other molecules.

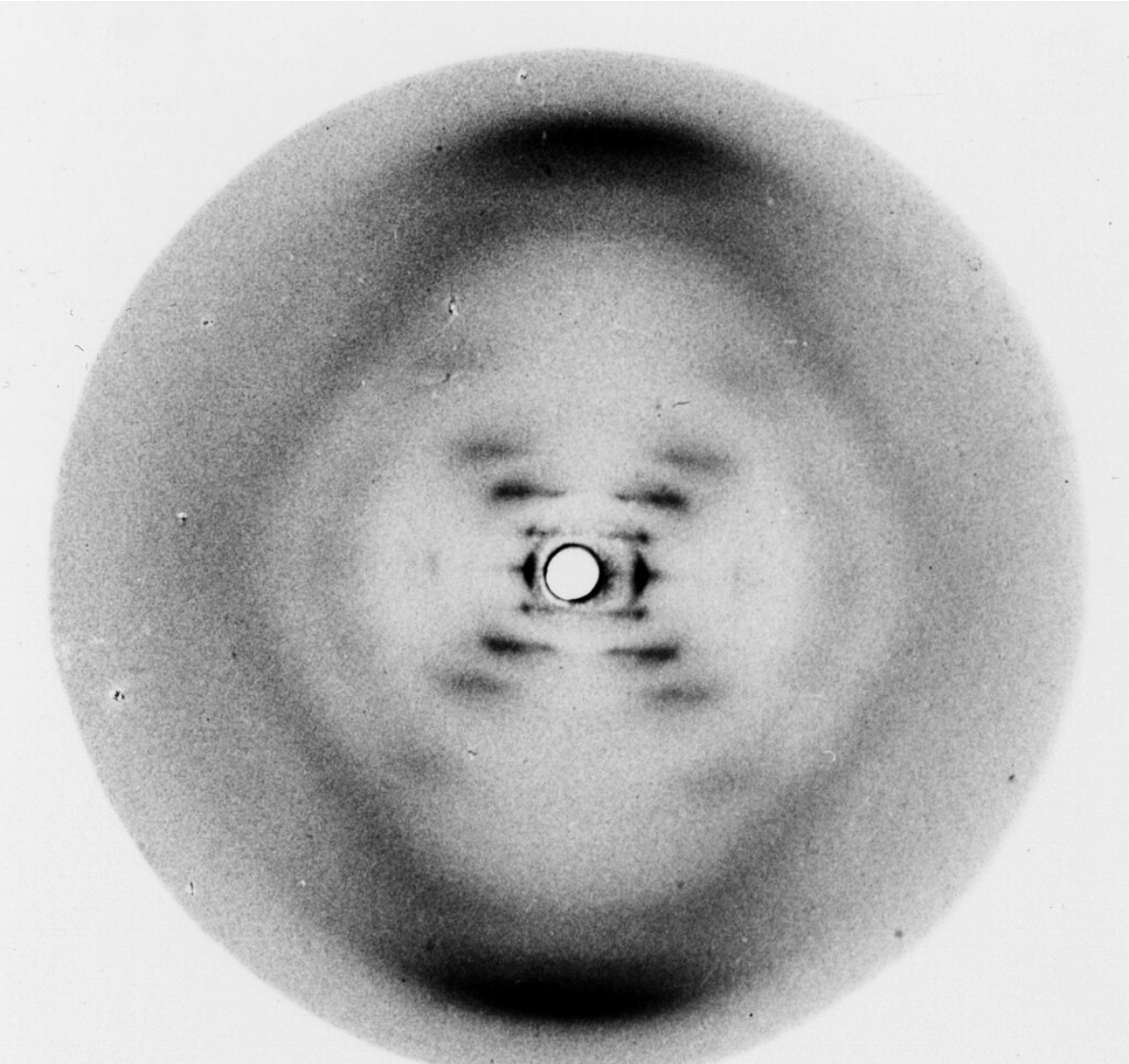
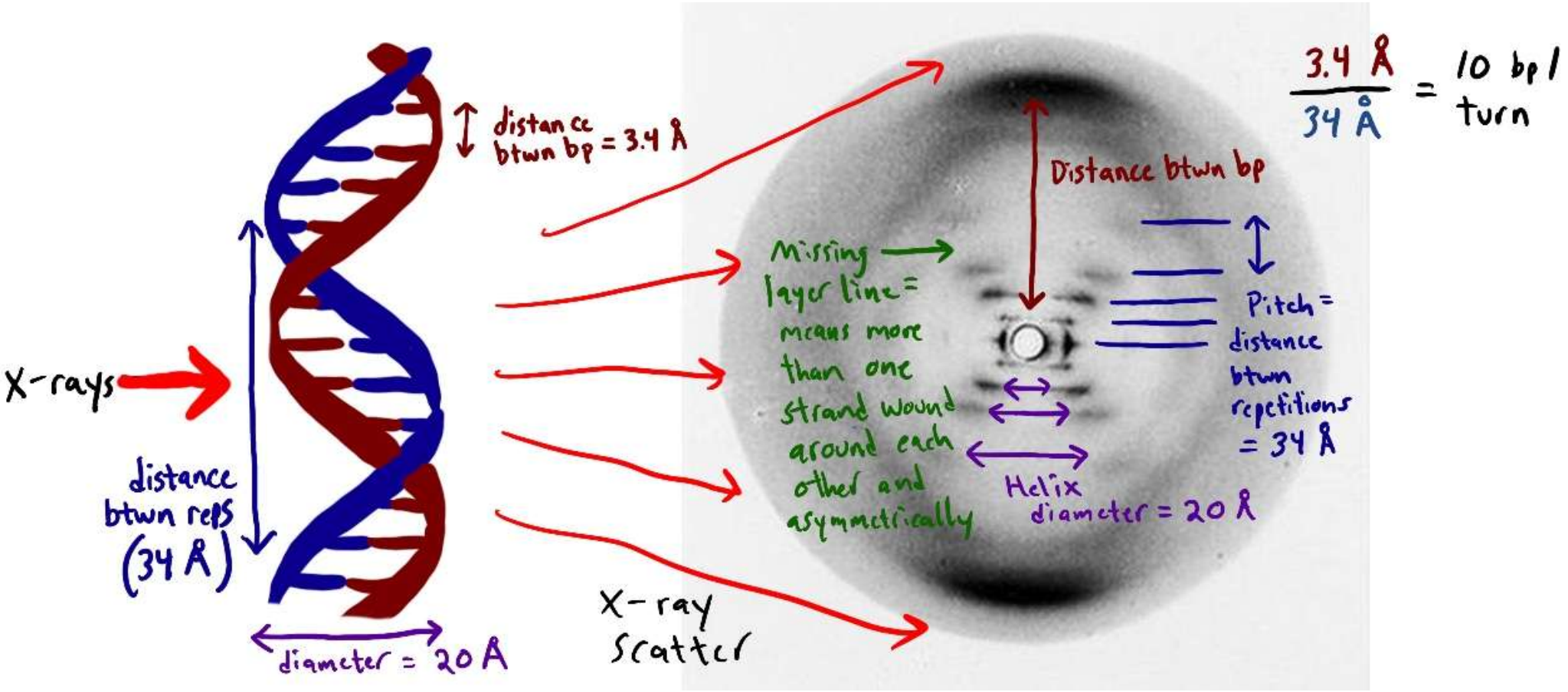


Photo 51 from Rosalind Franklin (1990-1958)

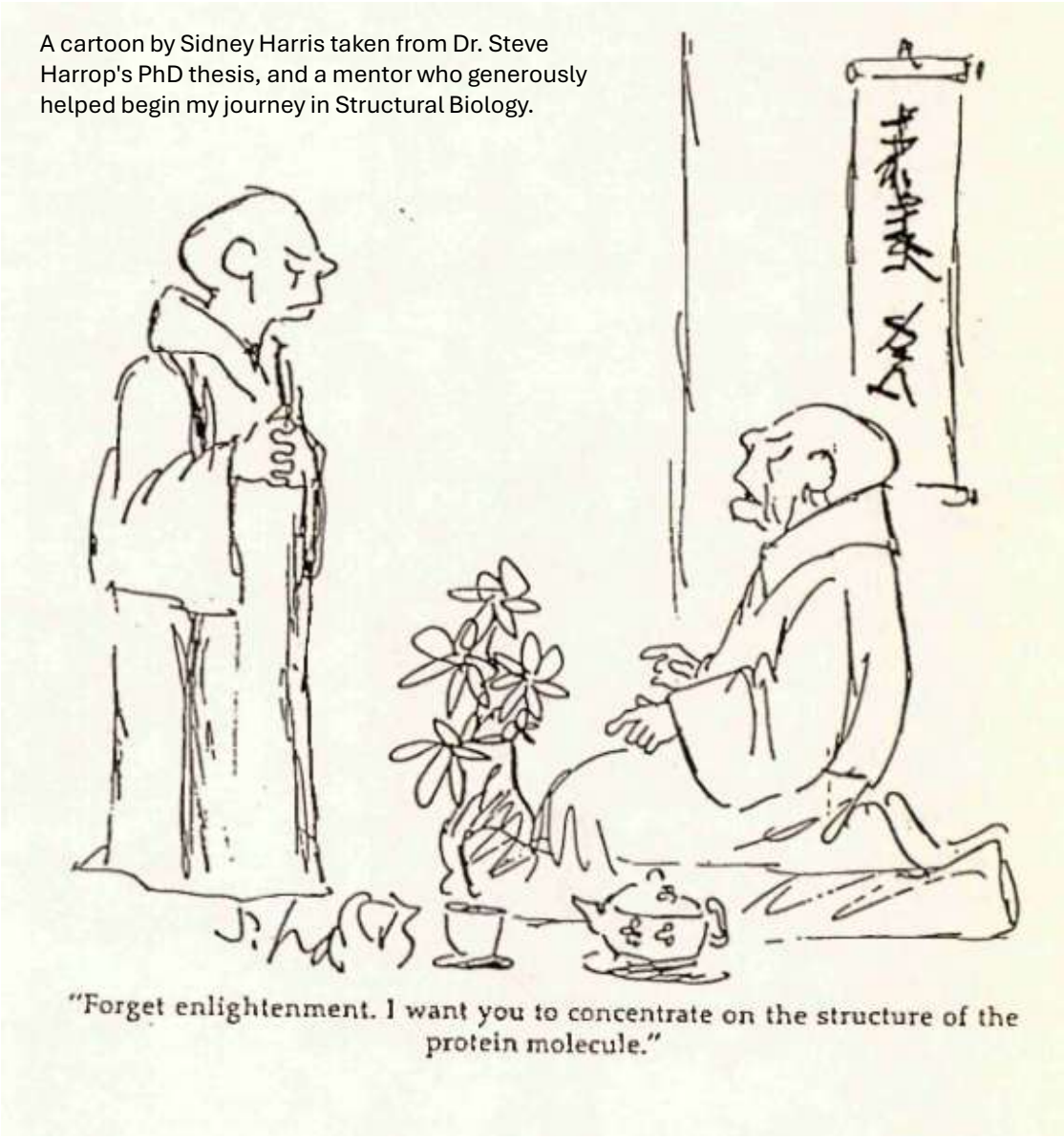


Crick and Watson, Proc. R. Soc. Lond. A 223, 80–96 1954, *without Franklin's data, "the formulation of our structure would have been most unlikely, if not impossible*"

The four main techniques

- X-ray Crystallography
- Small Angle X-ray Scattering (SAXS)
- Nuclear Magnetic Resonance (NMR)
- Cryo-electron microscopy (Cryo-EM)

A cartoon by Sidney Harris taken from Dr. Steve Harrop's PhD thesis, and a mentor who generously helped begin my journey in Structural Biology.



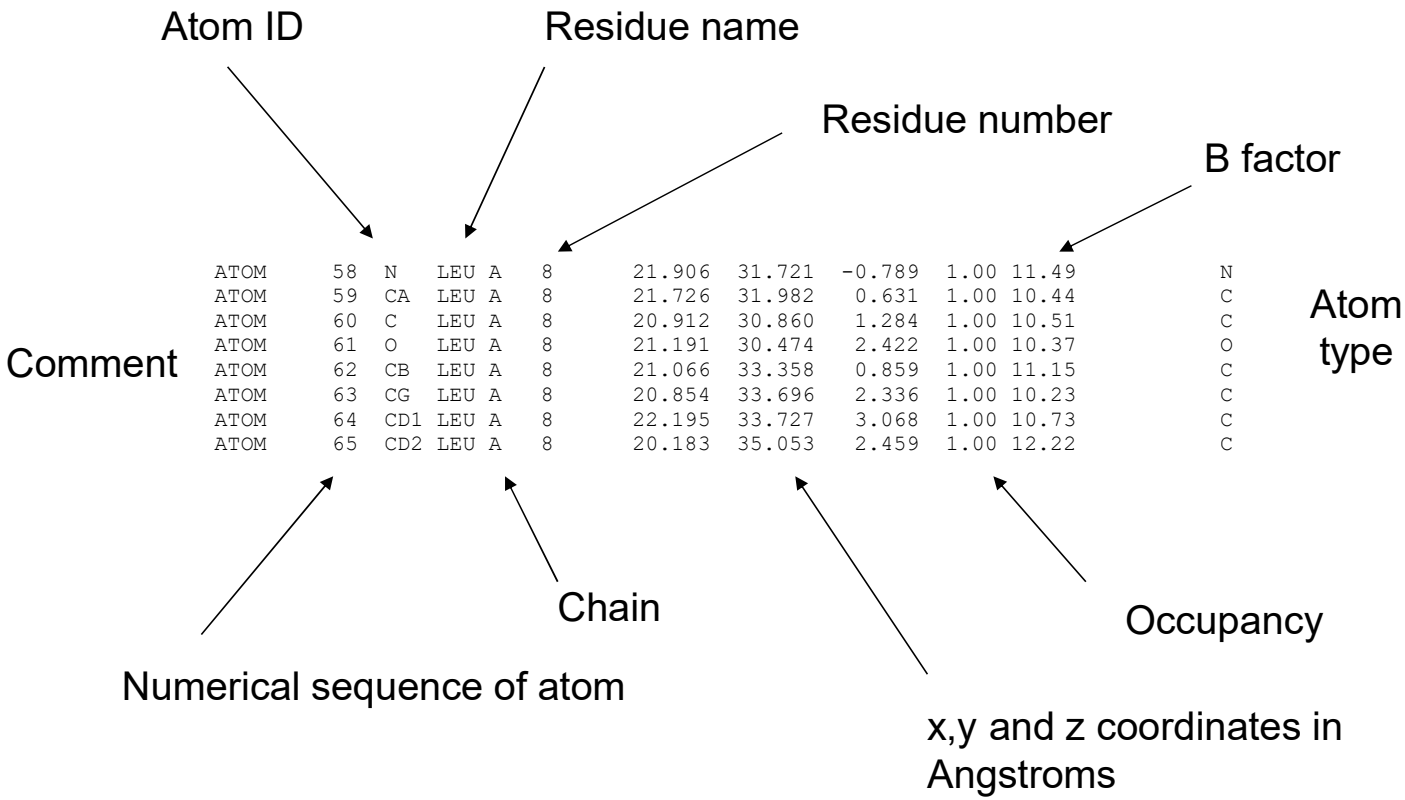
How do we store structures (all atomic techniques)?

Structures are deposited in the protein data bank or PDB (which also includes other biological macromolecules)

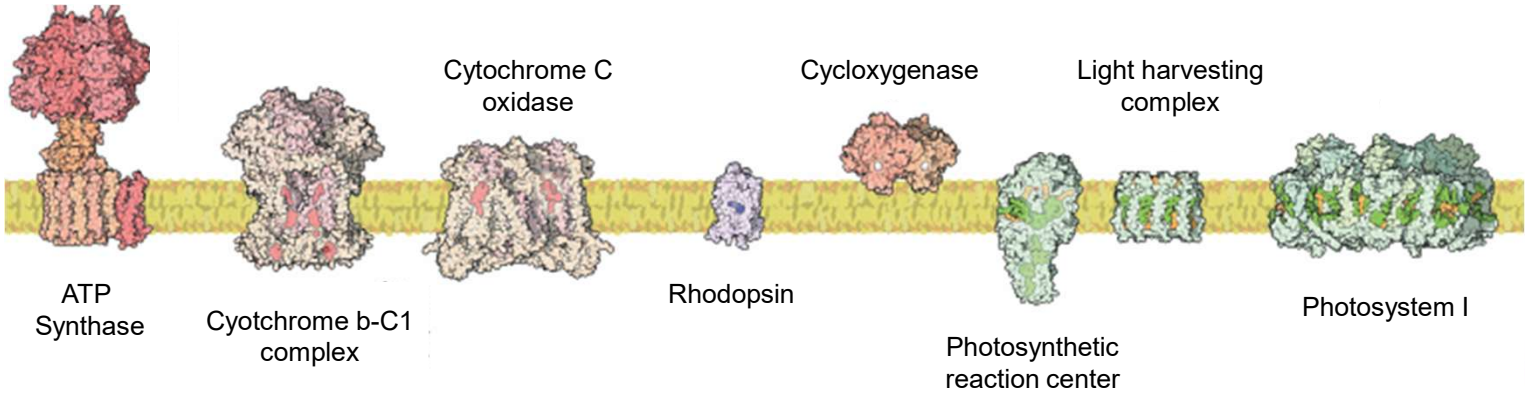
http://www.rcsb.org

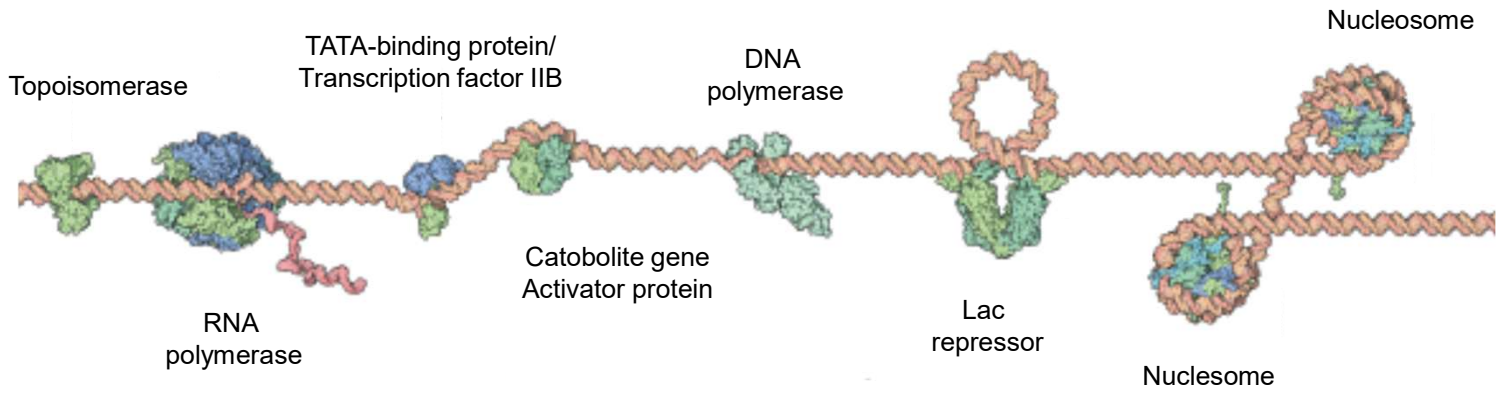
Good introductory material is available at https://pdb101.rcsb.org/

What information is stored?



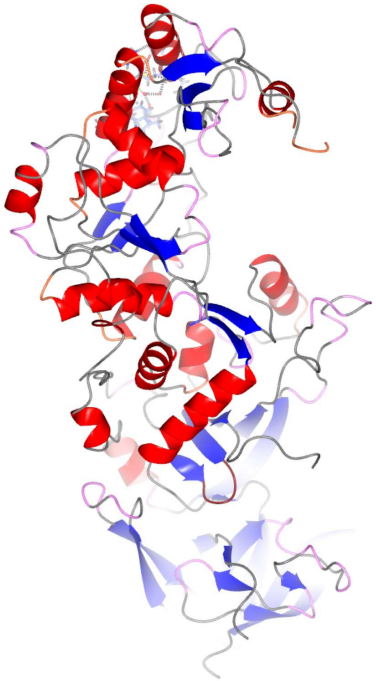
And the experimental data used to produce the model



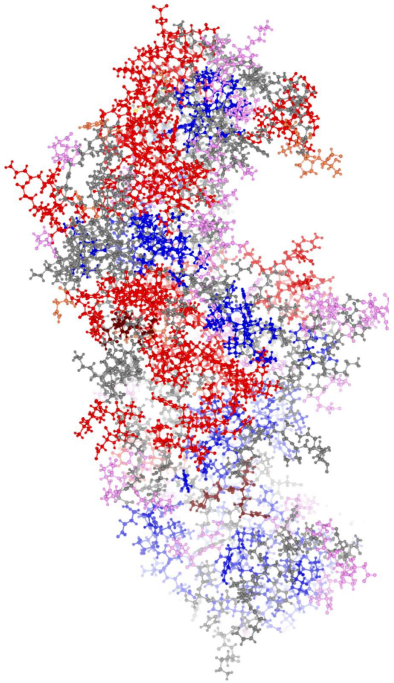


Form (or structure) gives a **clue** to the function.

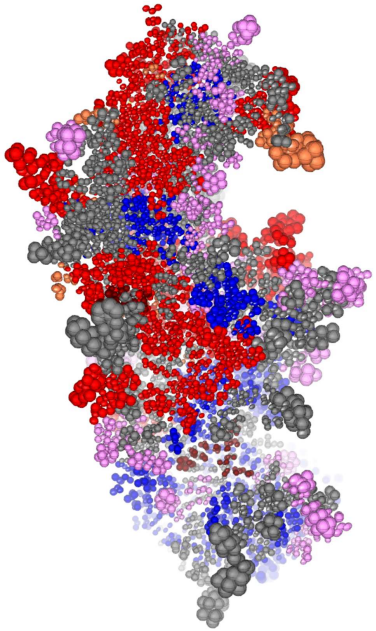
Adapted from Molecular Machinery: A tour of the Protein Data Bank, http://www/rcsb/org



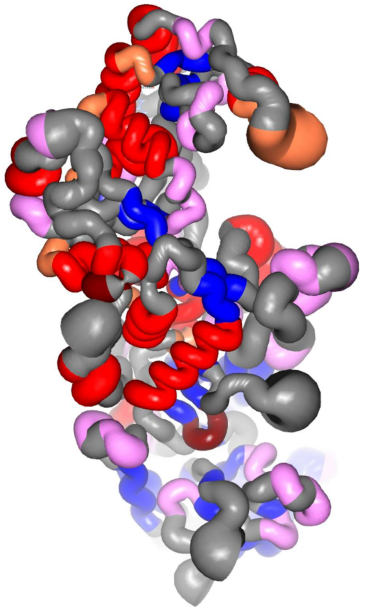
Ribbon



Ball and stick



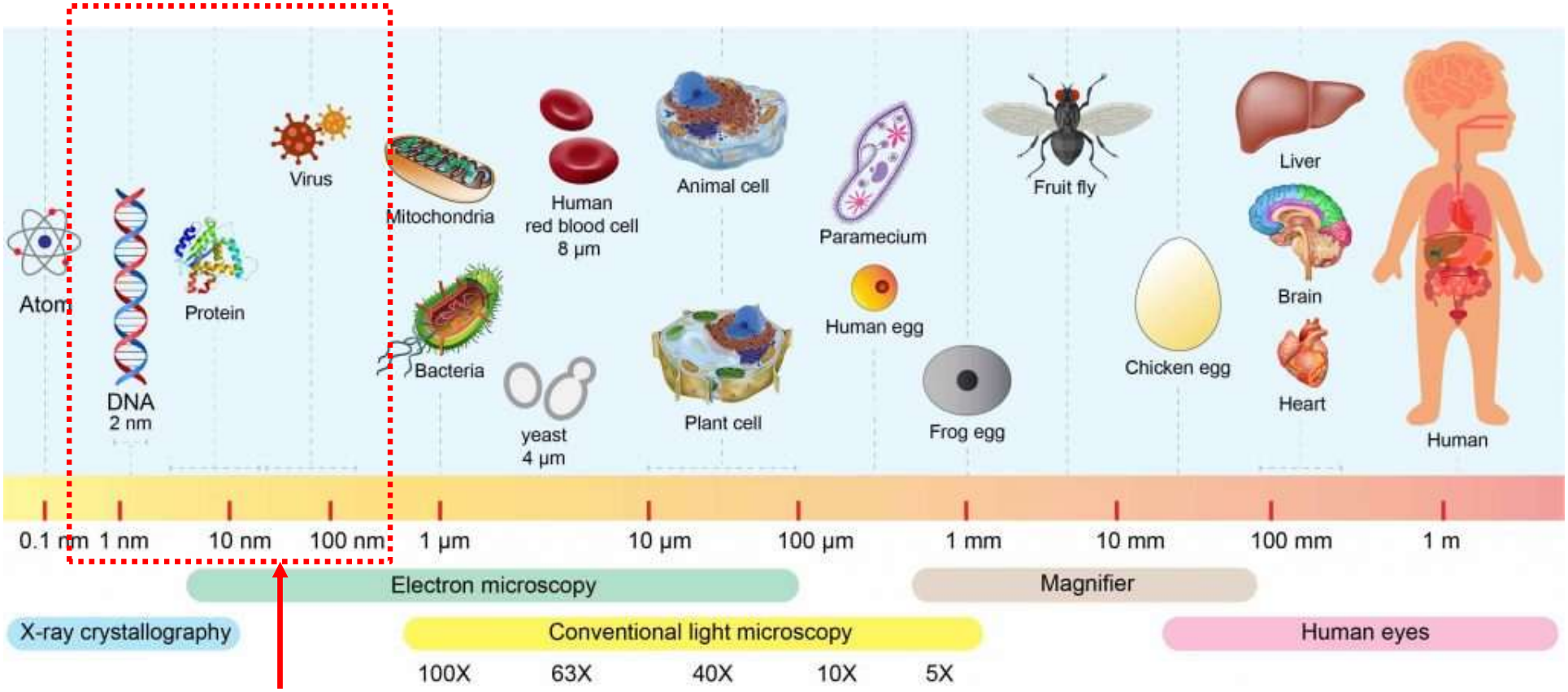
Atom as sphere



Thermal motion

Common concepts in structural studies (from a physics viewpoint)

- Scale
  - We are interested in structural information several orders of magnitude smaller than visible light.
- Waves
  - We use X-rays, electrons, neutrons, and even radio waves, and the interaction between waves to extract information.
- Interference
  - The interaction can be constructive or destructive, and we can infer structural information from that (Bragg's law in crystallography).
- Fourier transform
  - We use Fourier transforms to transform the information from waves into structural detail
- Signal amplification
  - Observation of many data points allows the signal to be seen in the noise

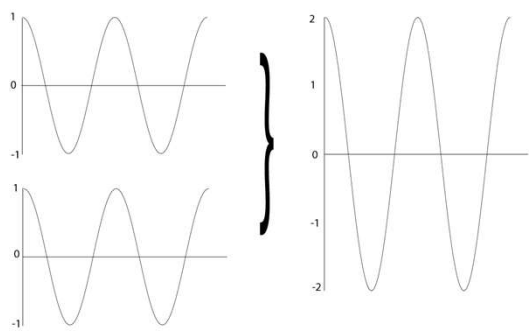


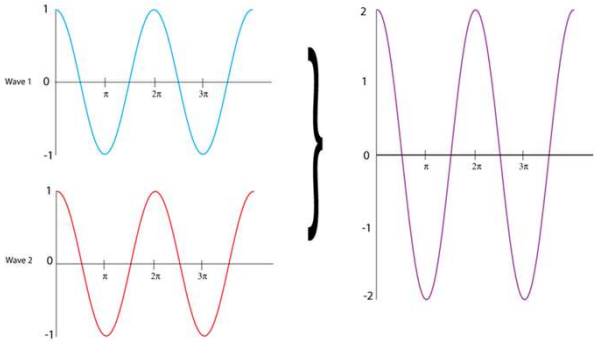
The size of objects that can be observed is limited by the wavelength. Resolution ~λ/2

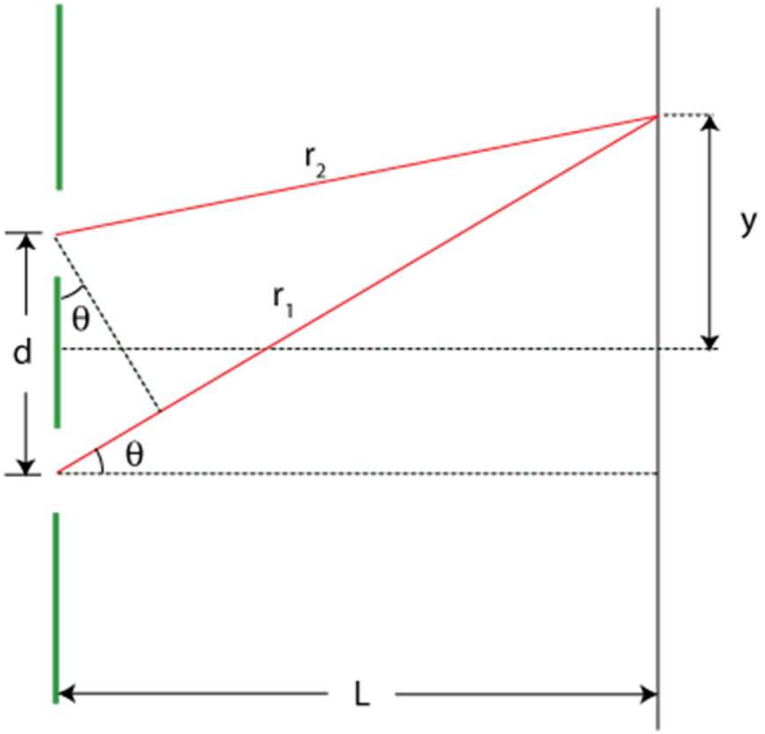
Visible light 400-700 nm, X-rays 0.07-0.15 nm

www.rsscience.com

Waves and interference







For constructive interference the path difference between r_1 and r_2 has to be a whole wavelength λ

$r_1 - r_2 = \lambda = d\sin\theta$

remembering that sine is the ratio of the opposite over hypotenuse.

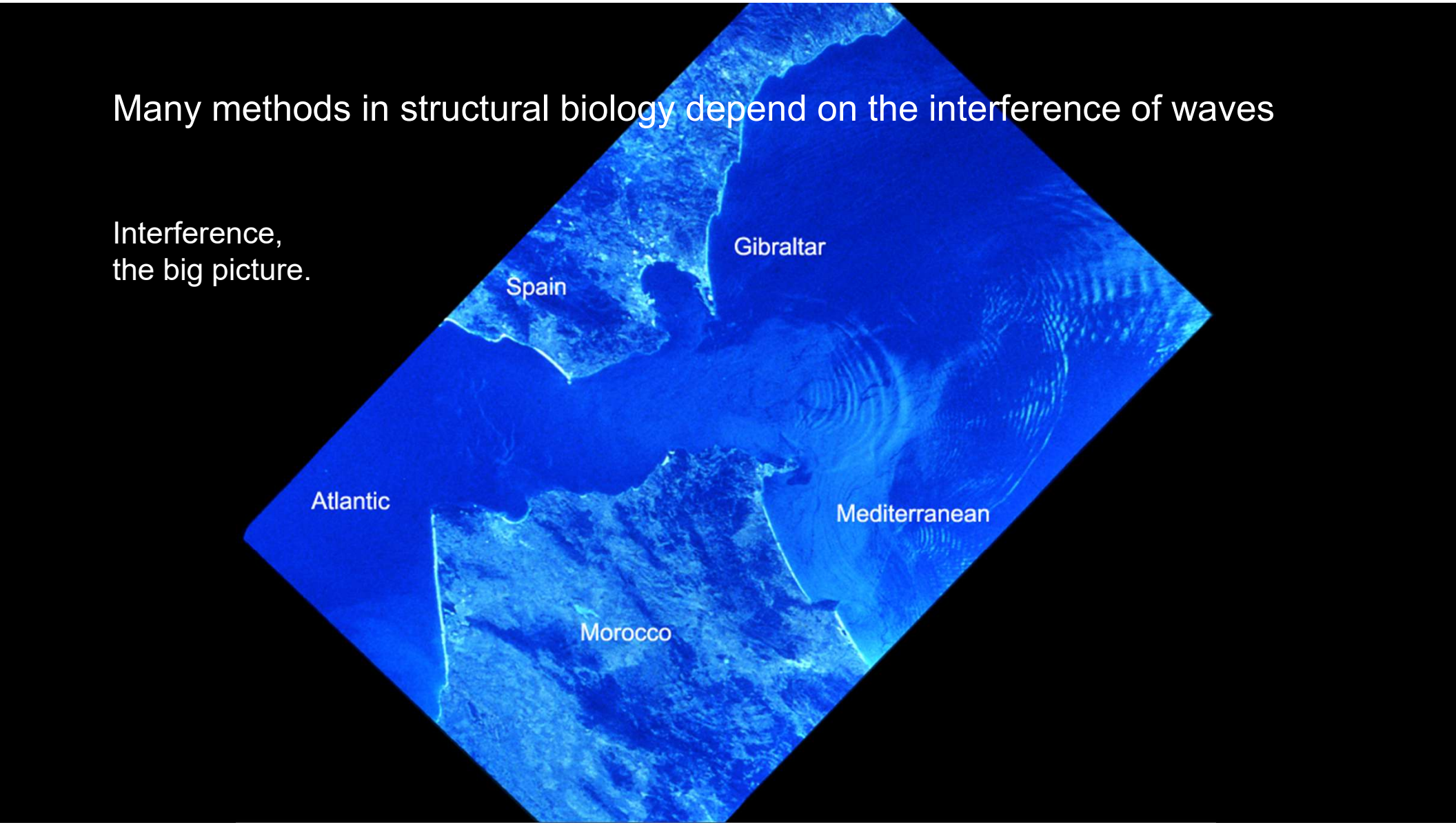
For interference at successive λ path differences the equation becomes

$m\lambda = d\sin\theta$

Where m is +/- 1, +/-2 etc.

Many methods in structural biology depend on the interference of waves

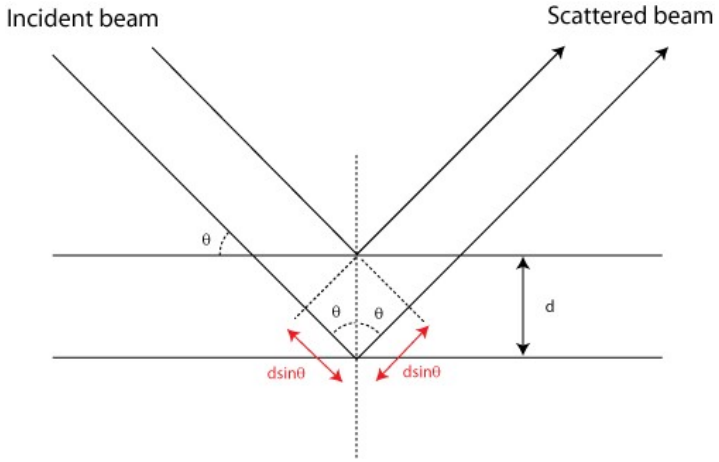
Interference,
the big picture.



William Lawrence Bragg (1890-1971)

Sir William Lawrence Bragg, was an Australian-born British physicist and X-ray crystallographer, discoverer (1912) of Bragg's law of X-ray diffraction, which is basic for the determination of crystal structure. He was joint recipient (with his father, William Henry Bragg) of the Nobel Prize in Physics in 1915.

Bragg is the youngest ever Nobel laureate in physics, having received the award at the age of 25 years. Bragg was the director of the Cavendish Laboratory, Cambridge, when the discovery of the structure of DNA was reported by James D. Watson and Francis Crick in February 1953.



Path difference for constructive interference is when the wavelength is a multiple of 2dsinθ or Braggs law:

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

Braggs law

Jean Baptiste Fourier (1763-1830)

In 1807 came up with an idea ….

Any periodic function could be rewritten as a weighted sum of sines and cosines of different frequencies.

This was not regarded as possible by other mathematicians of the time and it was not until 1878 that the idea was published in English.

The name of the idea is the Fourier series

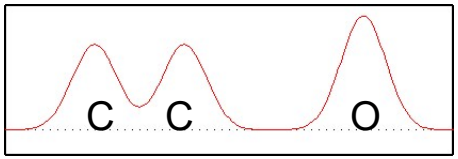
To change a signal to a Fourier series we perform a Fourier transform. To change the Fourier transform to a signal we perform an inverse Fourier transform.

Fourier's theorem is not only one of the most beautiful results of modern analysis, but it may be said to furnish an indispensable instrument in the treatment of nearly every recondite question in modern physics.

Lord Kelvin (1824-1907)

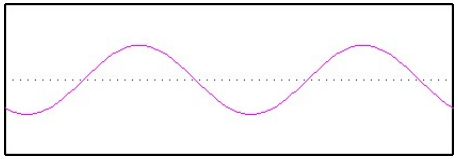
Other Lord Kelvin quotes

- Heavier-than-air flying machines are impossible.
- Radio has no future.
- In science there is only physics; all the rest is stamp collecting.



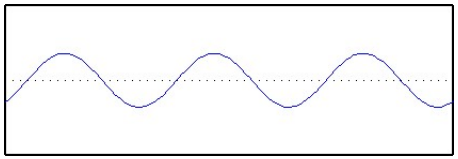
=

A unit cell of a crystal with 2 carbons and an oxygen



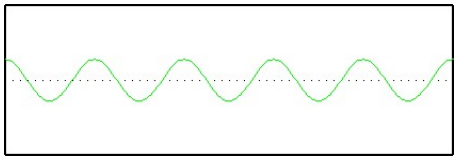
A cosine wave with frequency of 2, one peak represents the oxygen and the other the two carbons

+



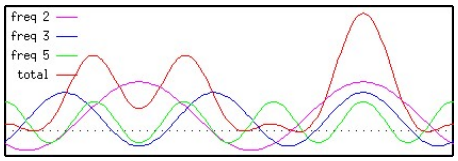
Add a cosine wave with frequency of 3, three repeats across the crystal. Note the phase is different, it starts in a different place.

+

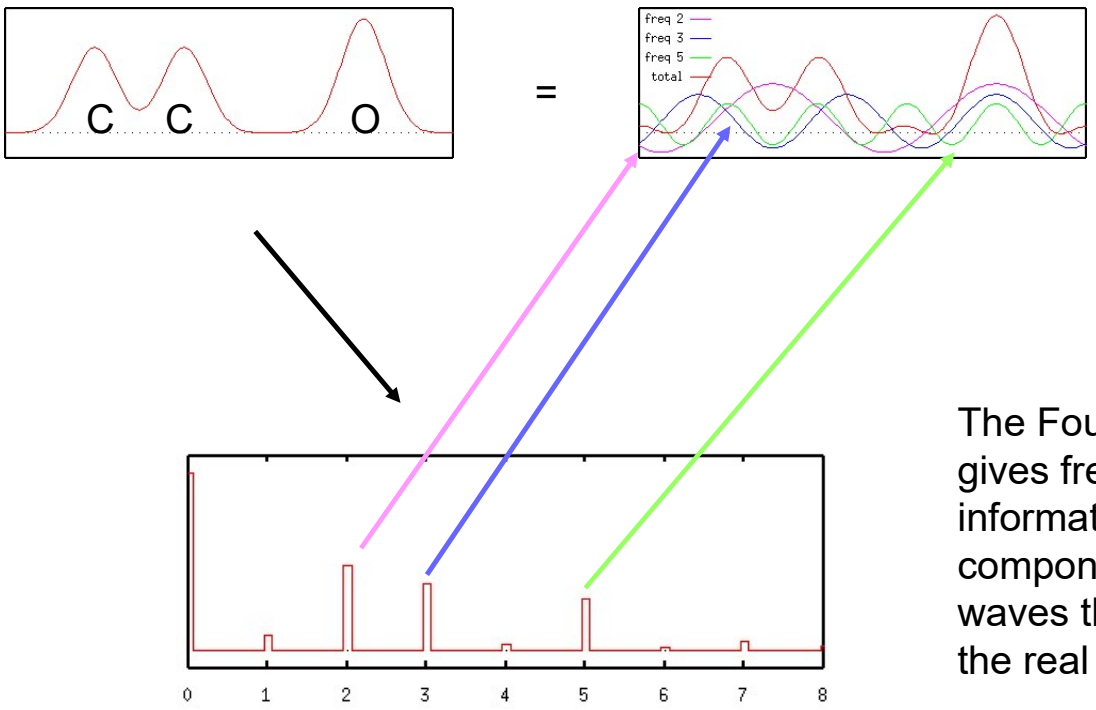


Add a third cosine wave with frequency of 5, with the peaks lined up on the carbons

=



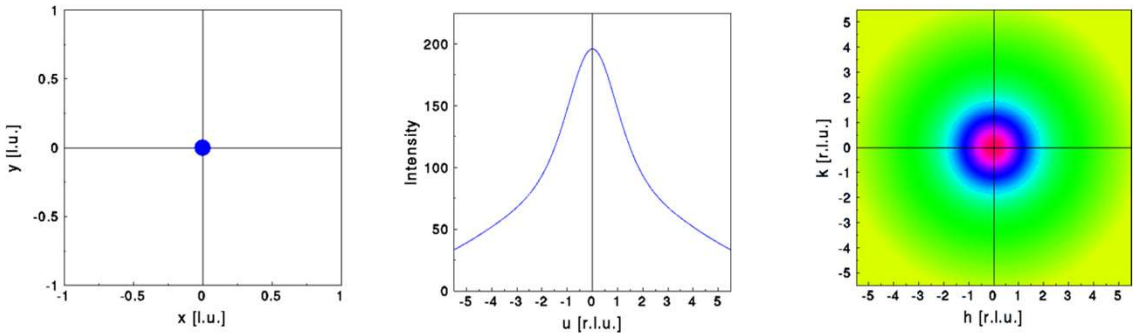
Add all the waves and the result is the original unit cell of the crystal



The Fourier transform gives frequency information on the components of the waves that describe the real space.

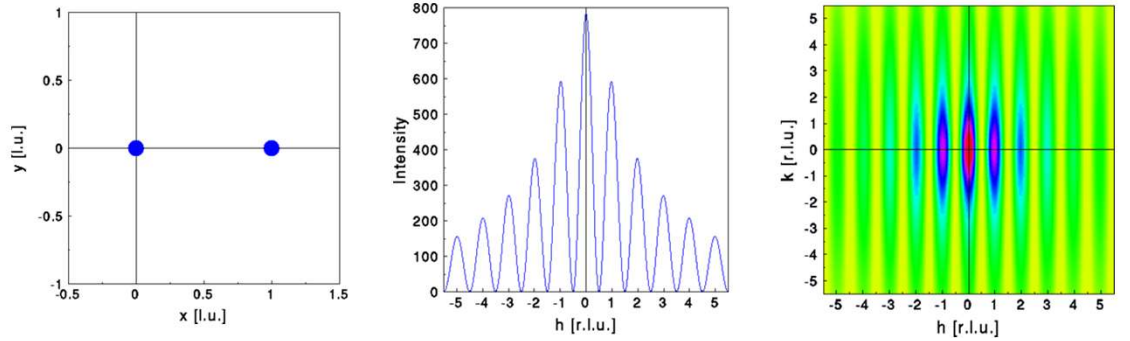
The Fourier transform of the unit cell

Check out http://www.ysbl.york.ac.uk/~cowtan/fourier/ftheory.html



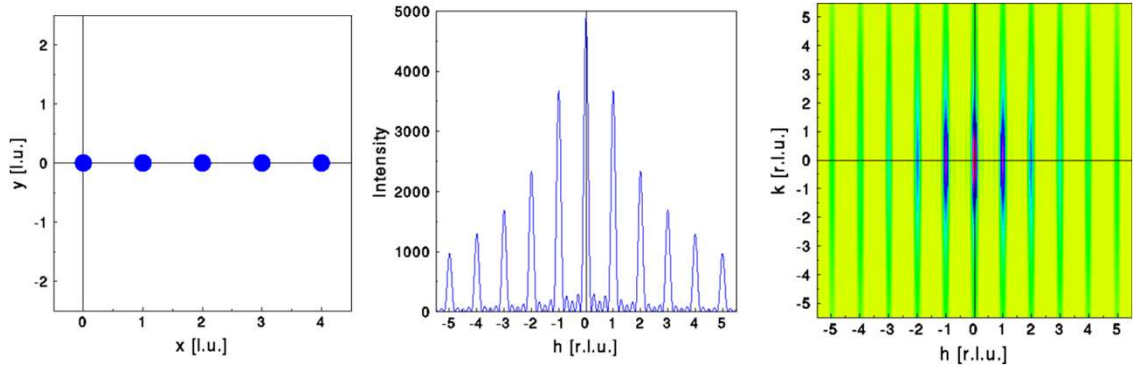
One atom

Max I = 200
(200x1^2)



Two atoms

Max I = 800
(200x2^2)

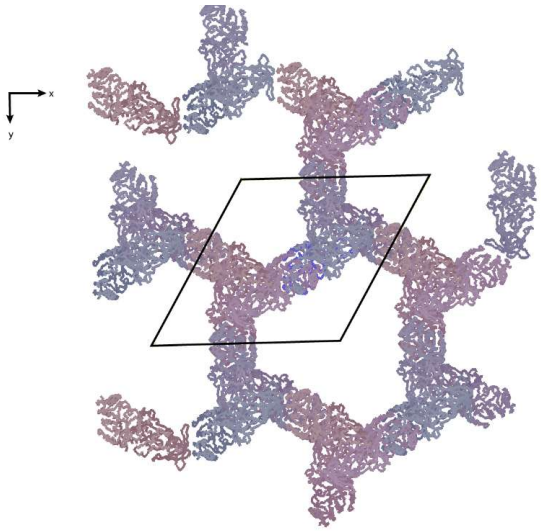


Five atoms

Max I = 5000
(200x5^2)

Amplification of the signal

How many atoms are there in a single molecule in the crystal?



Atomic composition

Carbon	4166
Hydrogen	6525
Nitrogen	1129
Oxygen	1248
Sulfur	23
Total number of atoms	13091
Formula = $C_{4166}H_{6525}N_{1129}O_{1248}S_{23}$	

On the order of 500,000,000 molecules in the crystal.

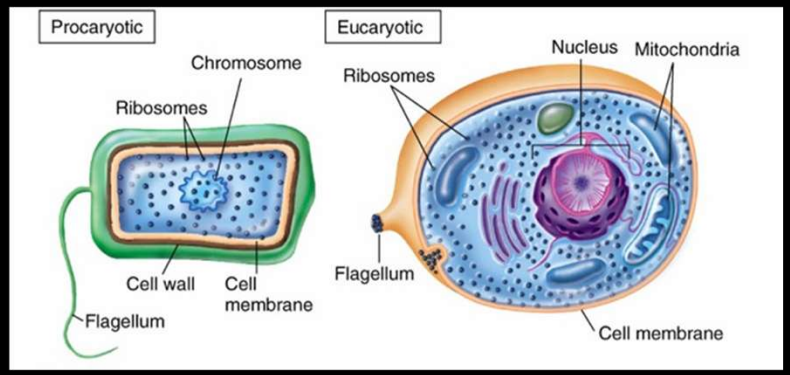


Image from *Microbiology: A Systems Approach*.

Case study - Glutamine tRNA Synthetase

Prokaryotes

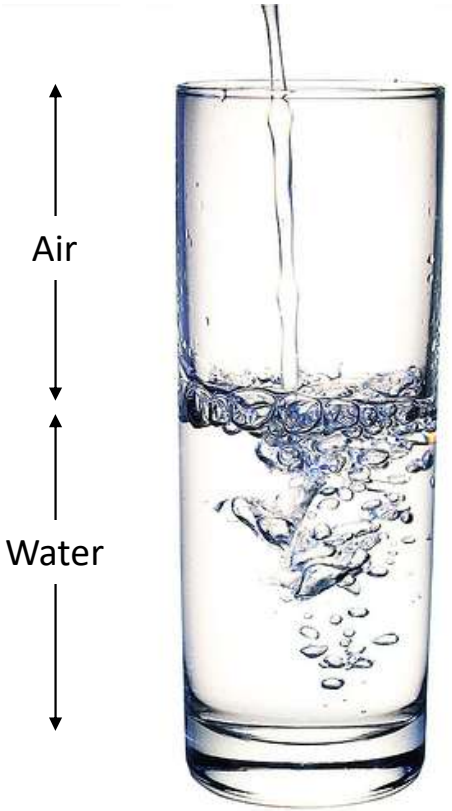
Catalytic Region | Anti-codon binding

40% Sequence Identity

Eukaryotes

N-term Domain tRNA Binding | Middle Domain Catalytic Region | C-term Domain Anti-codon binding

Pessimists, Optimists, and Crystallographers



Consider a glass of water

Pessimist
(the glass is half empty)

Optimist
(the glass is half full)

Crystallographer
(the glass is completely full)

Fantasy

A typical crystallographer and crystal …

Wile E. Coyote (Genius)

Overconfidentii Vulgaris

(Cristali Coltivatore Optimista)

And the crystal of interest …

Road Runner
(Beep beep)

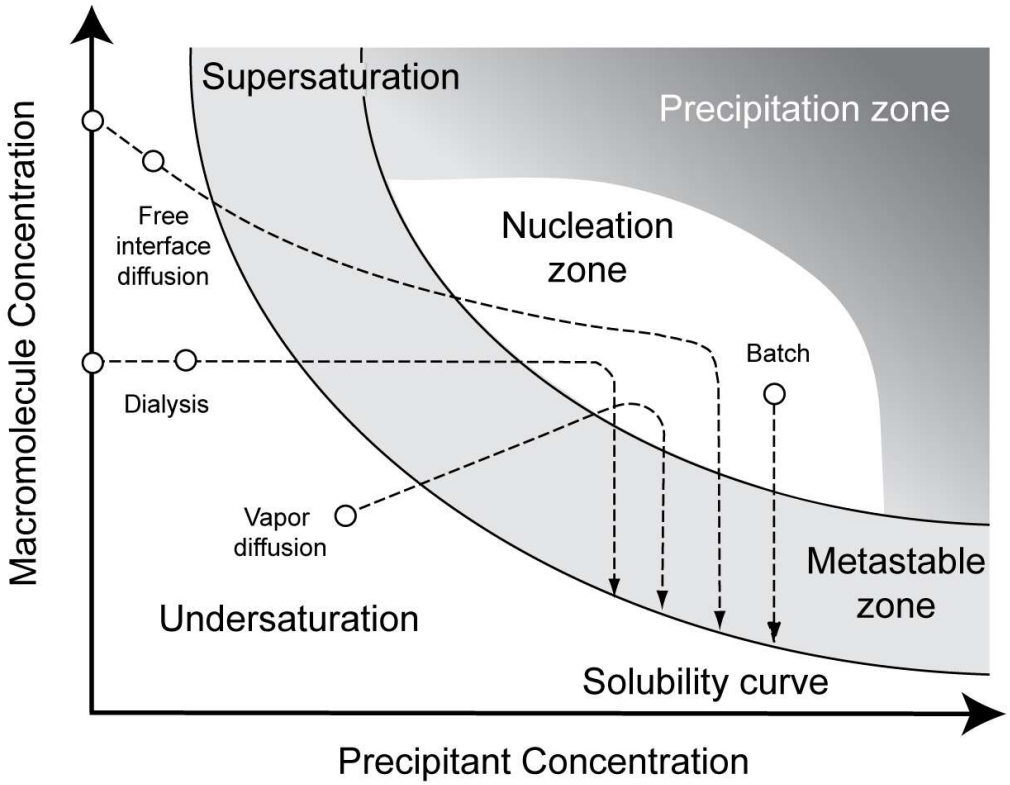
Disappearialis Quickius

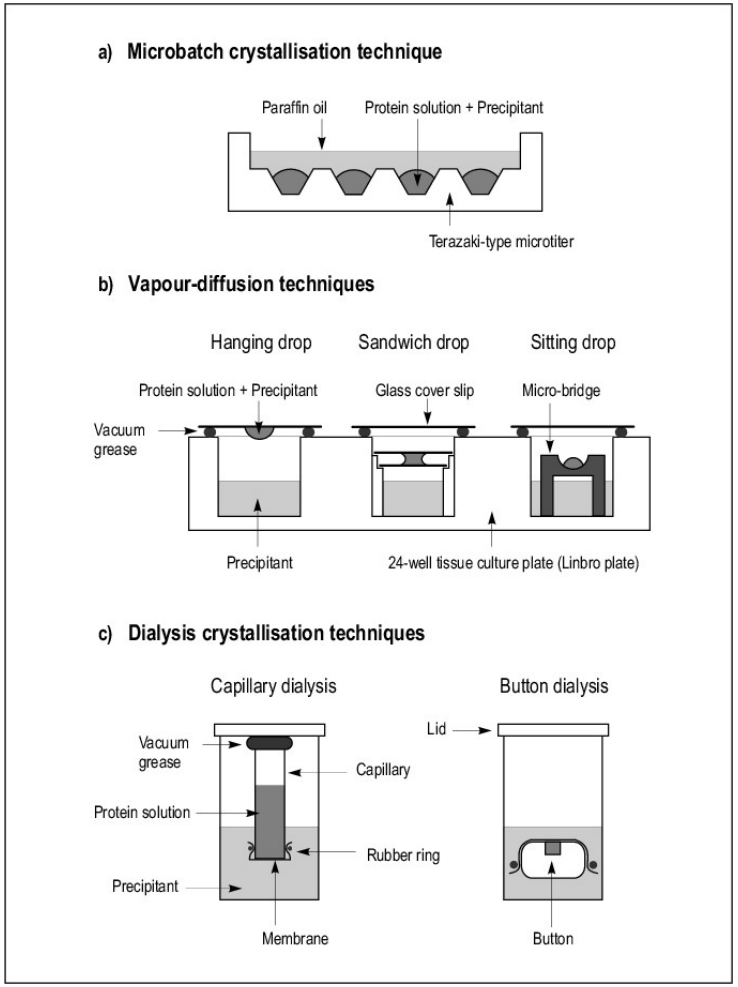
(Cristallio Perfetto)

For crystallography, there are many potential steps where failure can occur, with only 20-25% of soluble purified proteins yielding crystals.

However, the structural information that results is of the highest resolution and arguably the gold standard for structural information.

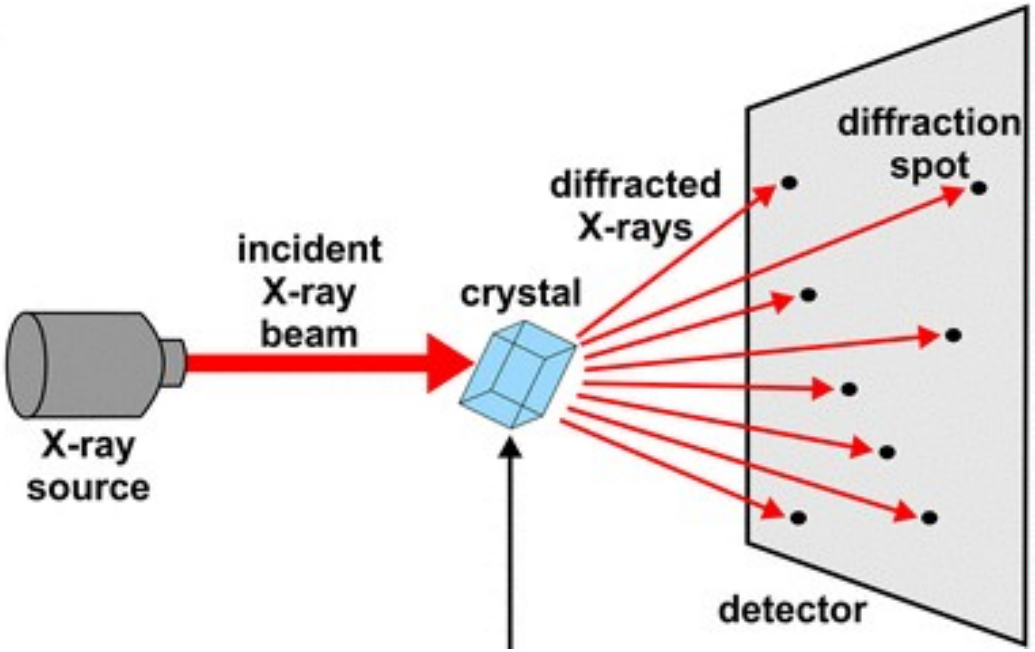
Simplified phase diagram for crystallization



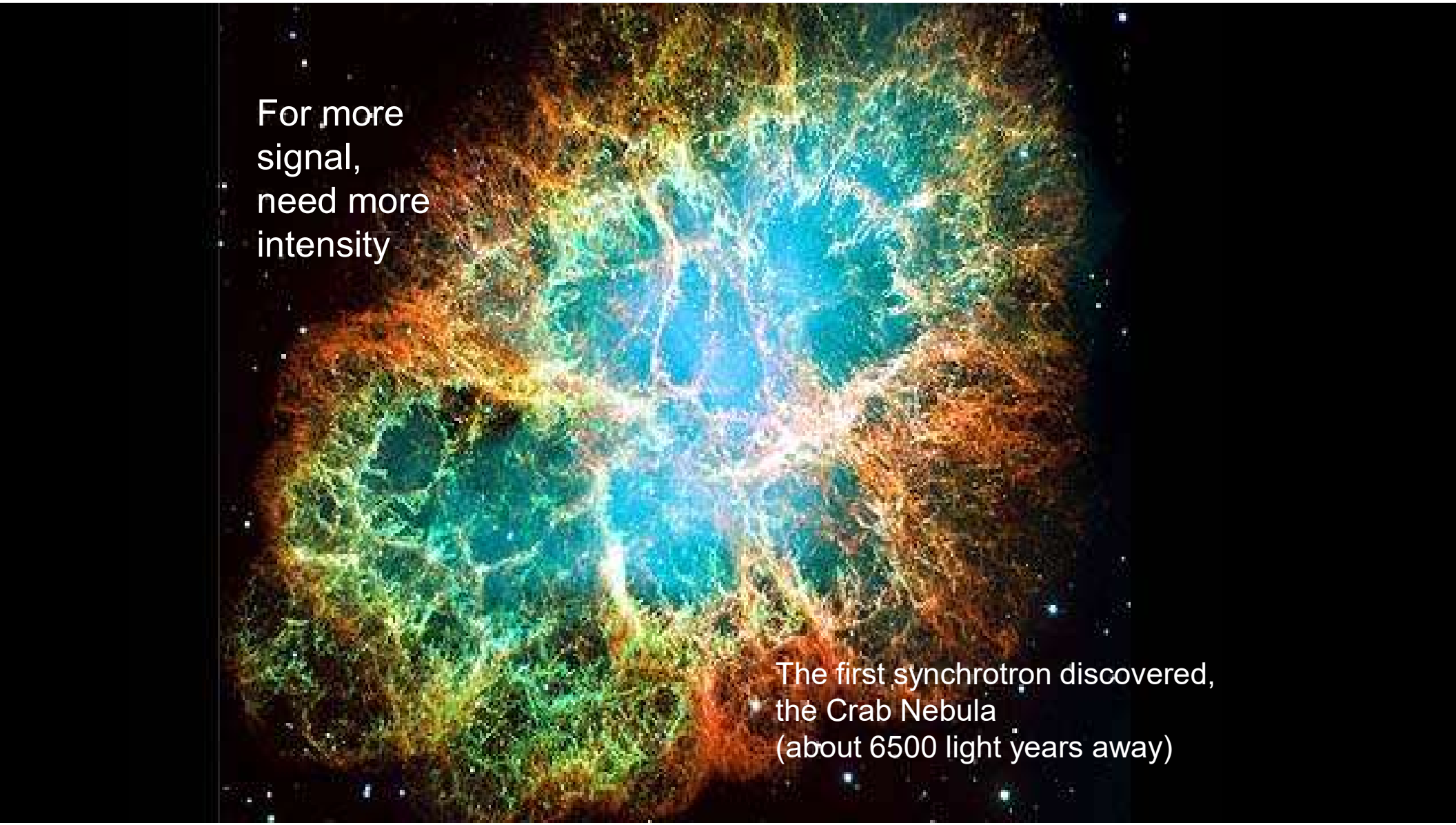


X-ray Crystallography – the most common structural technique

- Sample Requirements:
  - Pure, stable protein, mg quantities, must crystallize
- Where/How:
  - Crystal screens, home/source setups, synchrotrons
- Data/Output:
  - High-resolution atomic models, electron density
- Pros/Cons:
  - Best resolution, but crystals can be a bottleneck
- Use Cases:
  - Enzymes, drug binding, well-folded domains



Bijelic, A., Rompel, A. Polyoxometalates: more than a phasing tool in protein crystallography. ChemTexts 4, 10 (2018)


For more
signal,
need more
intensity
The first synchrotron discovered,
the Crab Nebula
(about 6500 light years away)

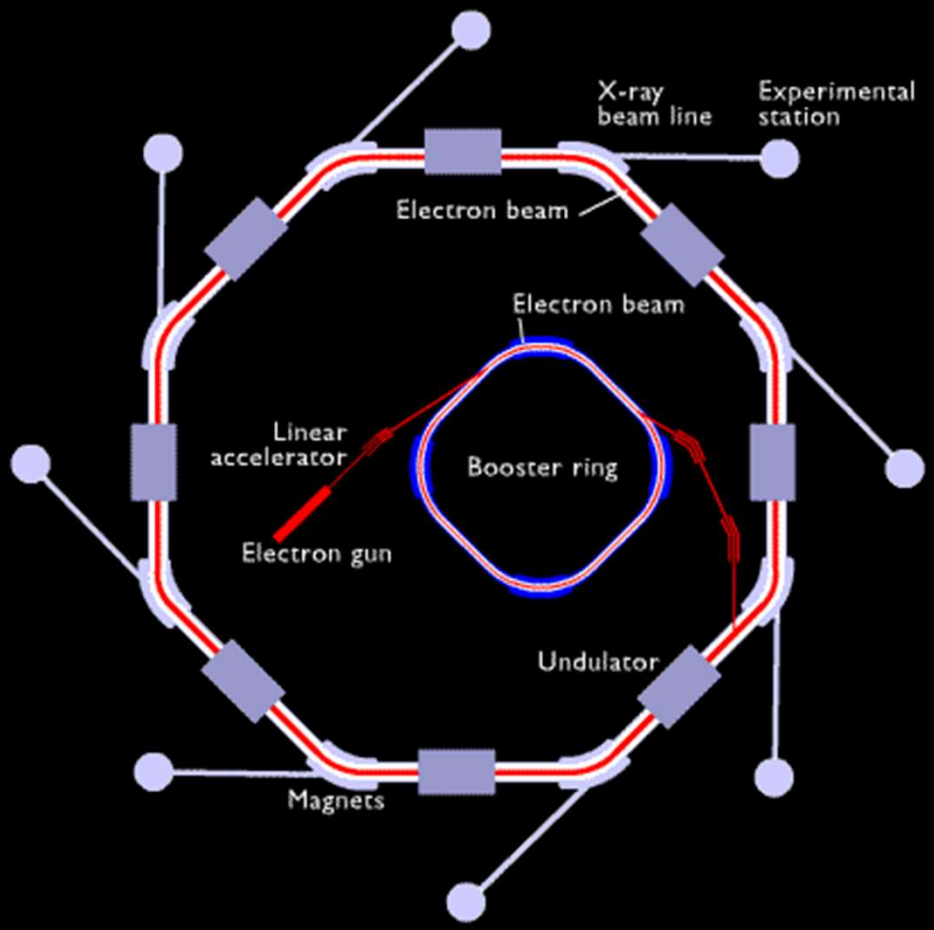
A synchrotron accelerates and stores particles (electrons or protons) moving at speeds close to that of light.

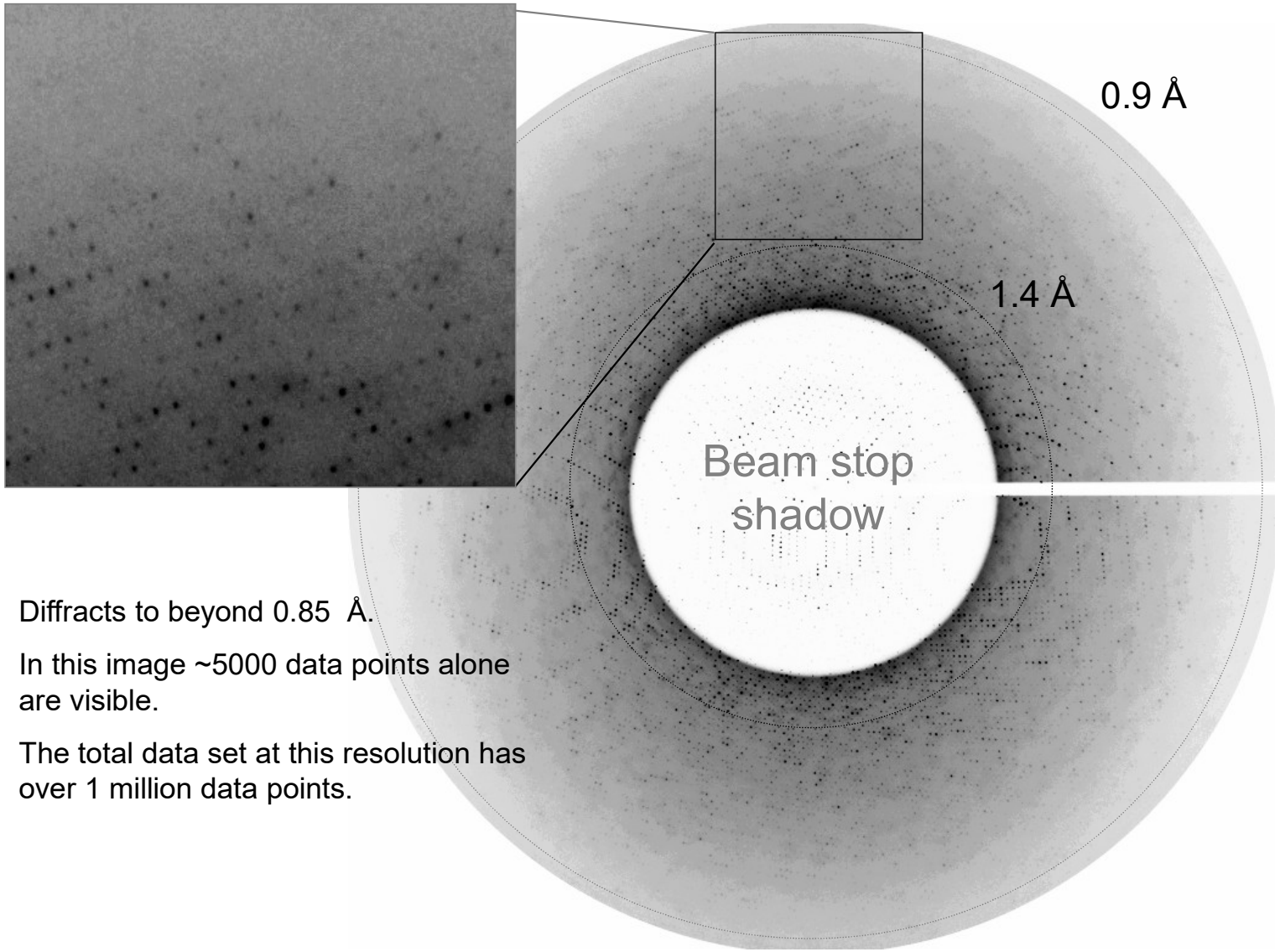
As the particles loose energy they give of electromagnetic radiation.

The particles are steered by magnetic fields.

Electromagnetic radiation (photons) is not affected by these fields and is emitted at the tangent to the change in direction.

Insertion devices (undulators and wigglers) ‘amplify’ this radiation

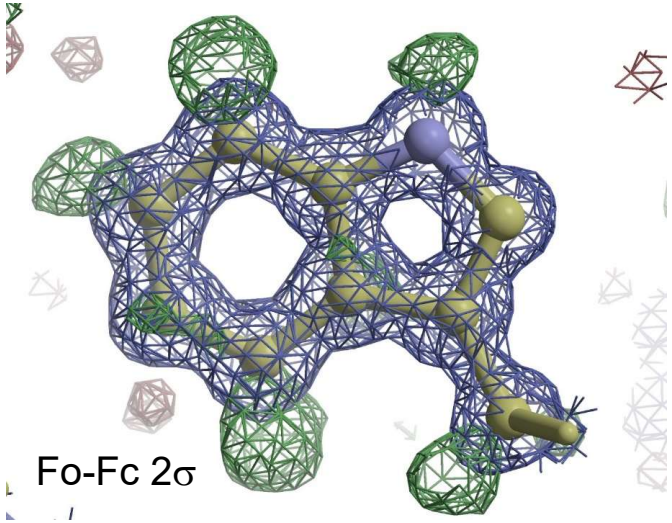


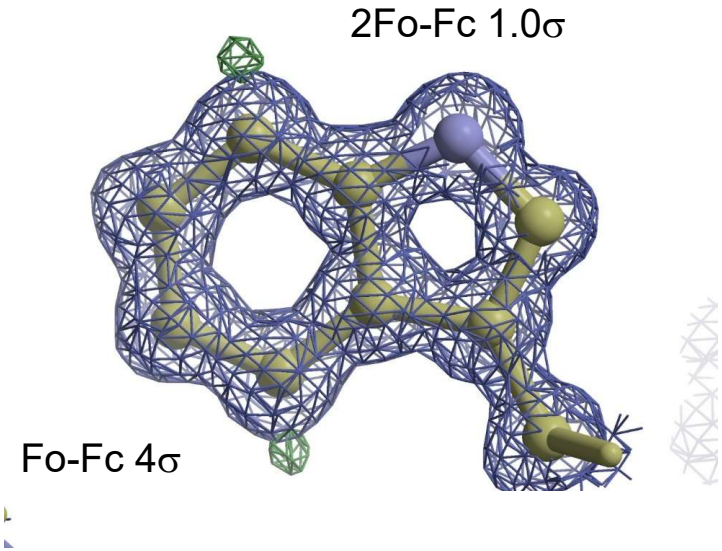


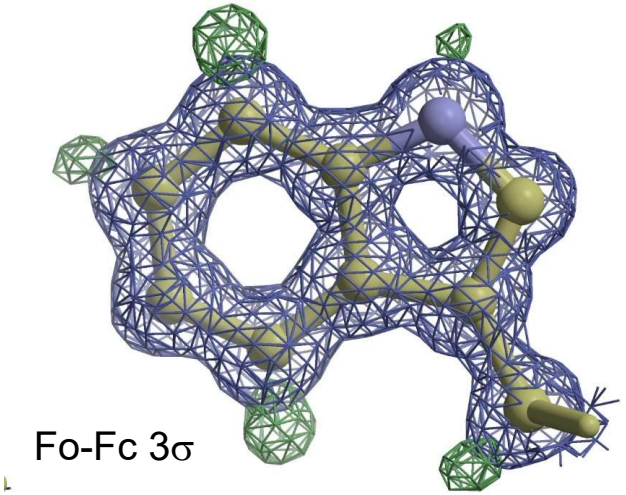
Diffracts to beyond 0.85 Å.

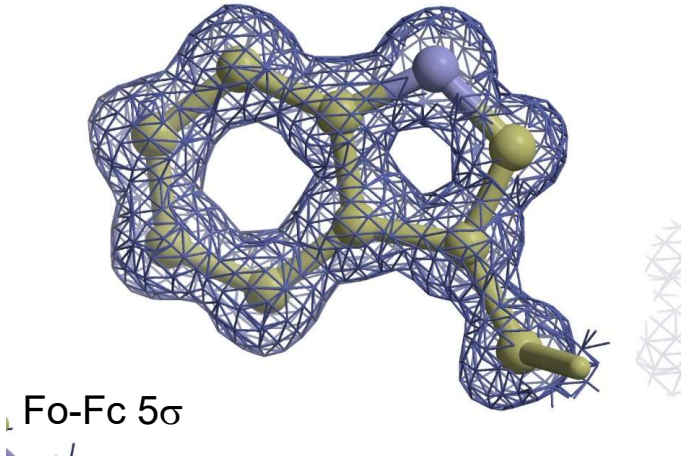
In this image ~5000 data points alone are visible.

The total data set at this resolution has over 1 million data points.


Fo-Fc 2σ


2Fo-Fc 1.0σ
Fo-Fc 4σ


Fo-Fc 3σ


Fo-Fc 5σ

The Phase Problem

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F(h,k,l)\,e^{-i\,2\pi\,(hx+ky+lz)}$$

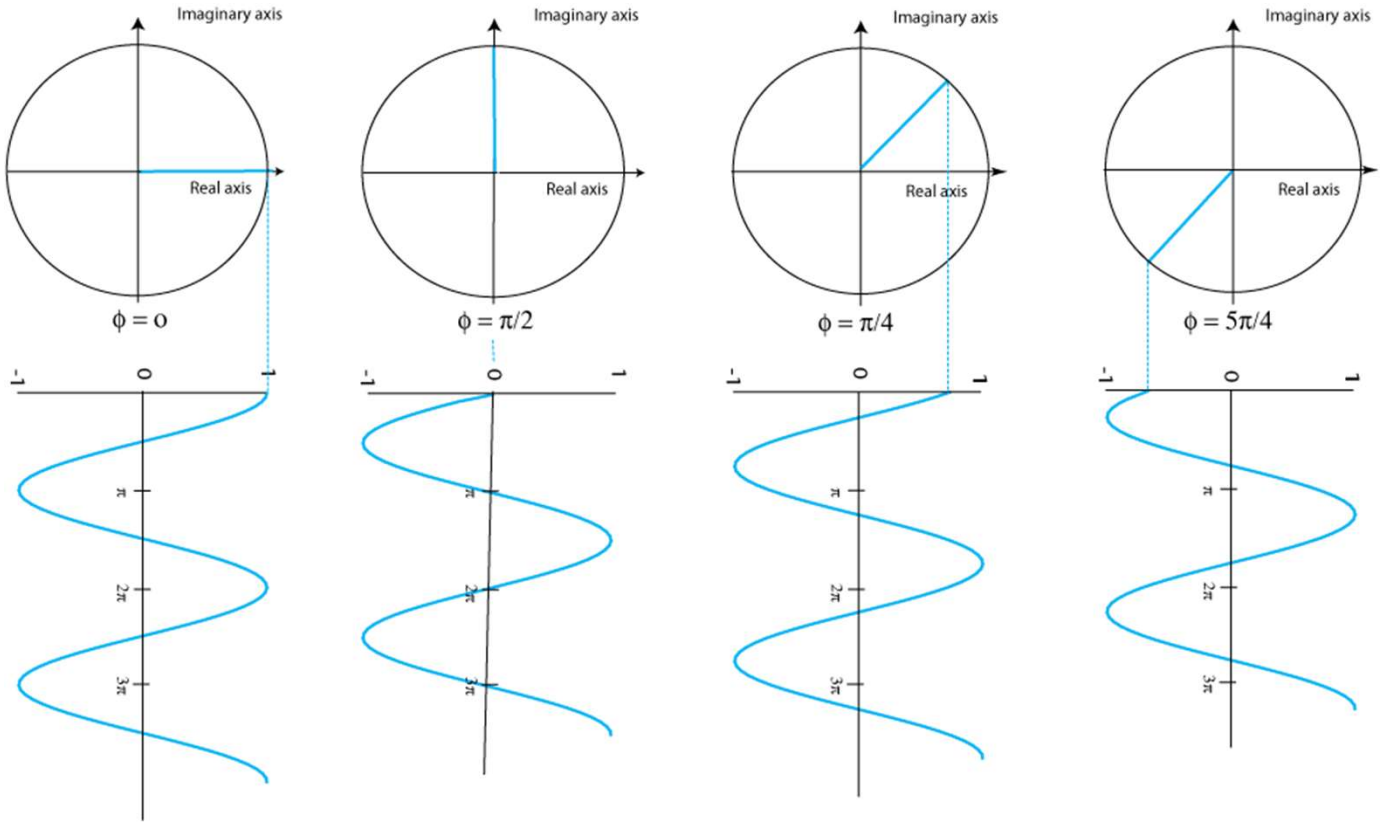
Electron density

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

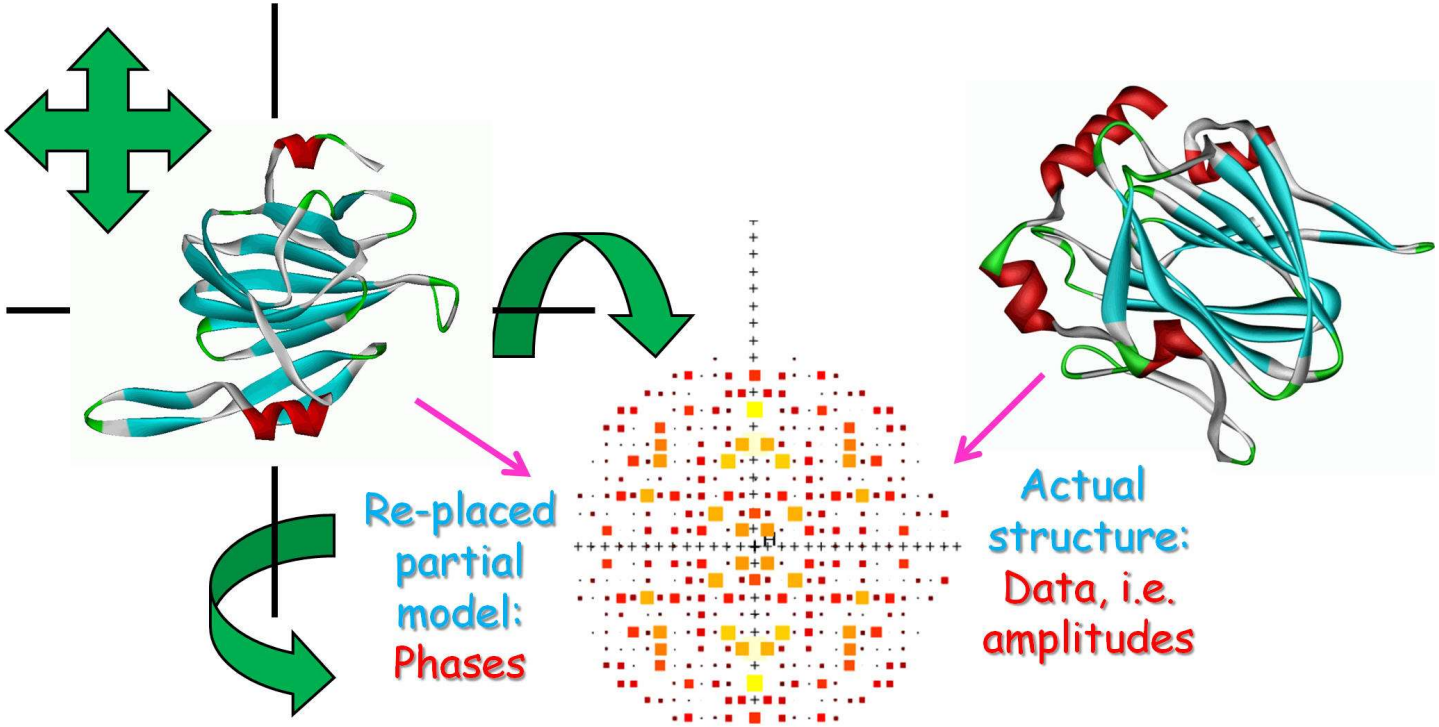
Amplitude of the wave
(measurable)

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

Phase of the wave
(unmeasurable crystalographically)

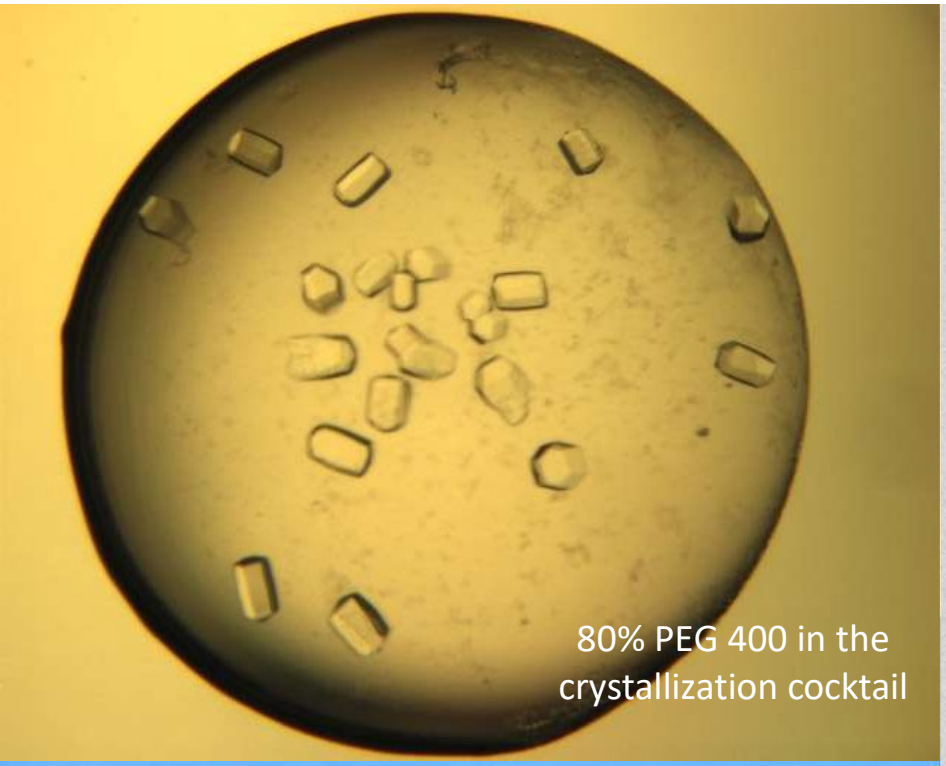

Imaginary axis
Real axis
ϕ = o
-1
0
1
π
2π
3π
Imaginary axis
Real axis
ϕ = π/2
-1
0
1
π
2π
3π
Imaginary axis
Real axis
ϕ = π/4
-1
0
1
π
2π
3π
Imaginary axis
Real axis
ϕ = 5π/4
-1
0
1
π
2π
3π

Molecular Re-Placement of a structurally similar search model

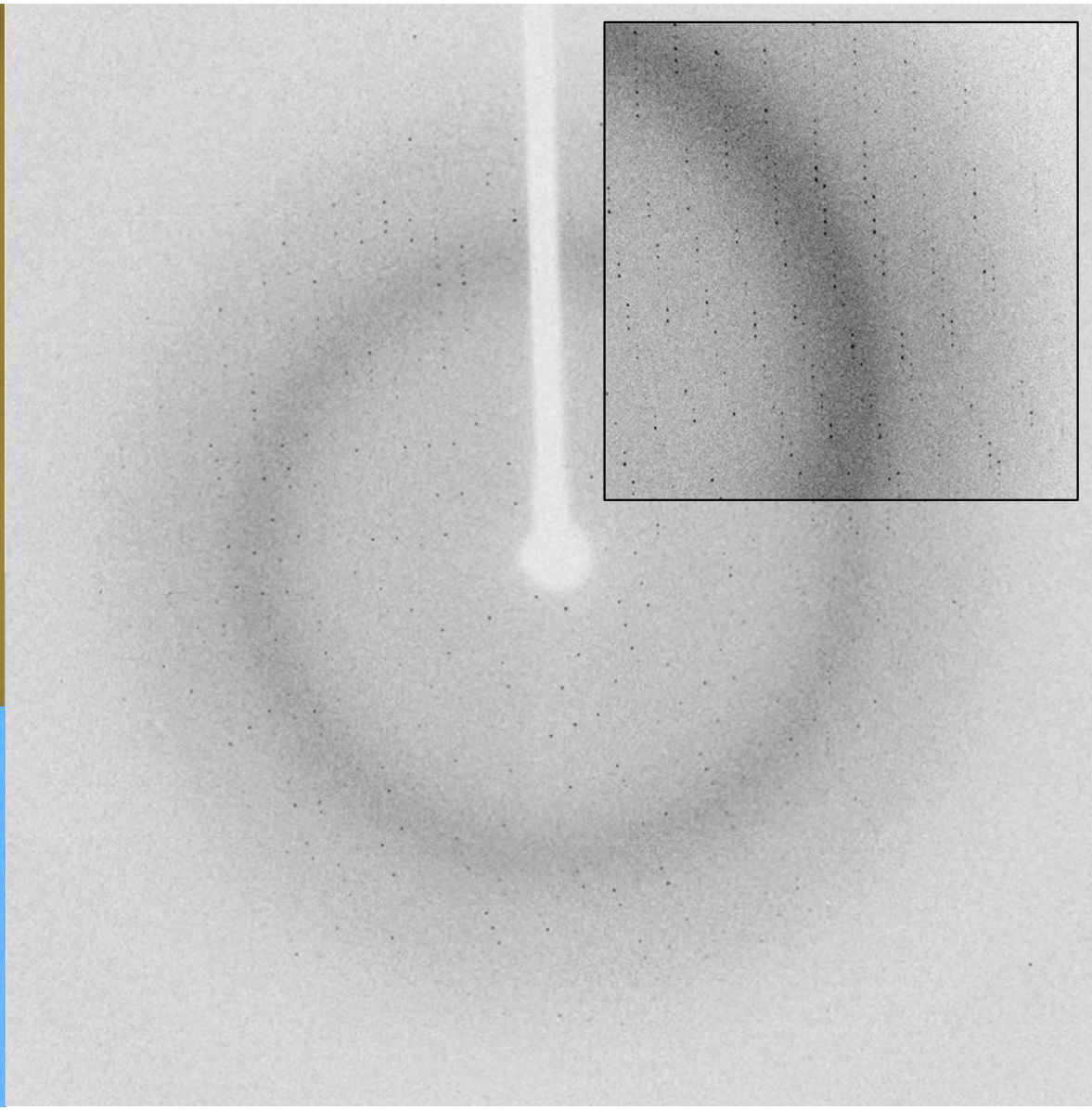


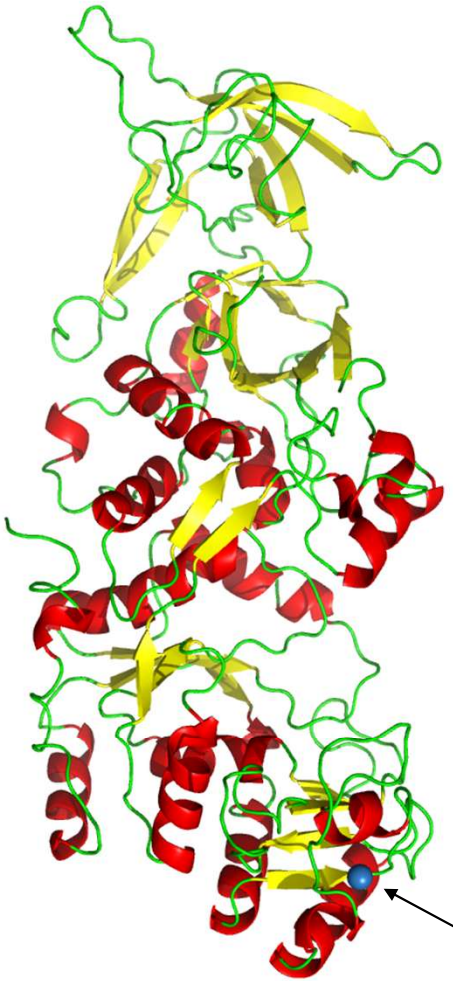
Either in one shot – 6-dimensional stochastic/evolutionary search

or by successive 3-d rotation and 3-d translation search using Patterson methods


80% PEG 400 in the
crystallization cocktail

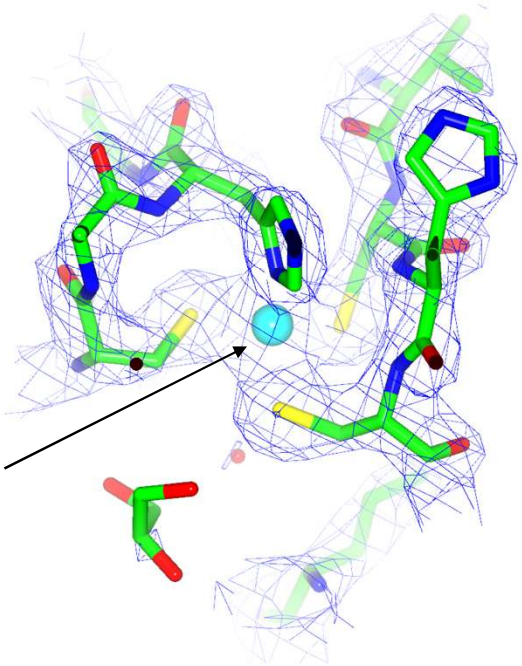

200 micron beam

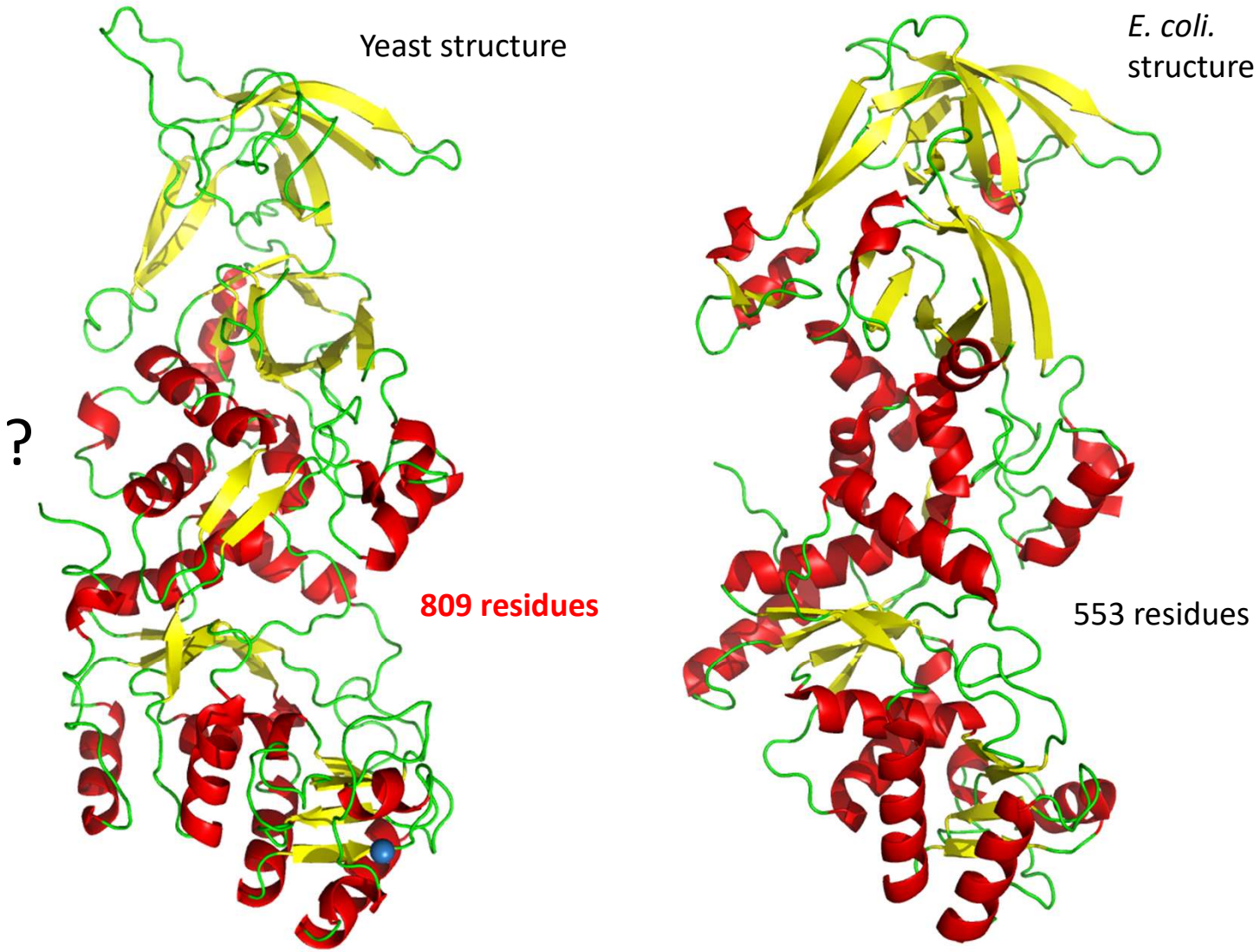


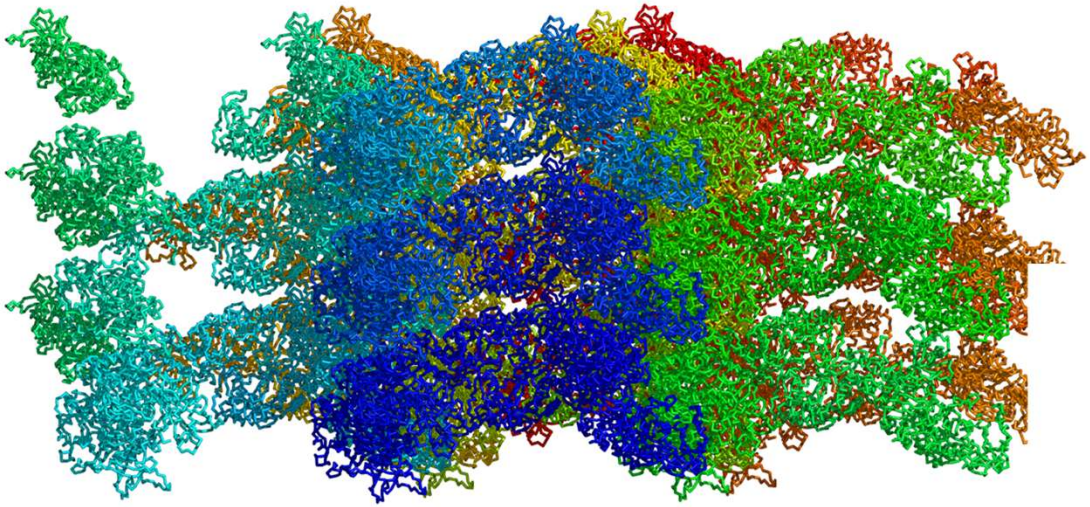


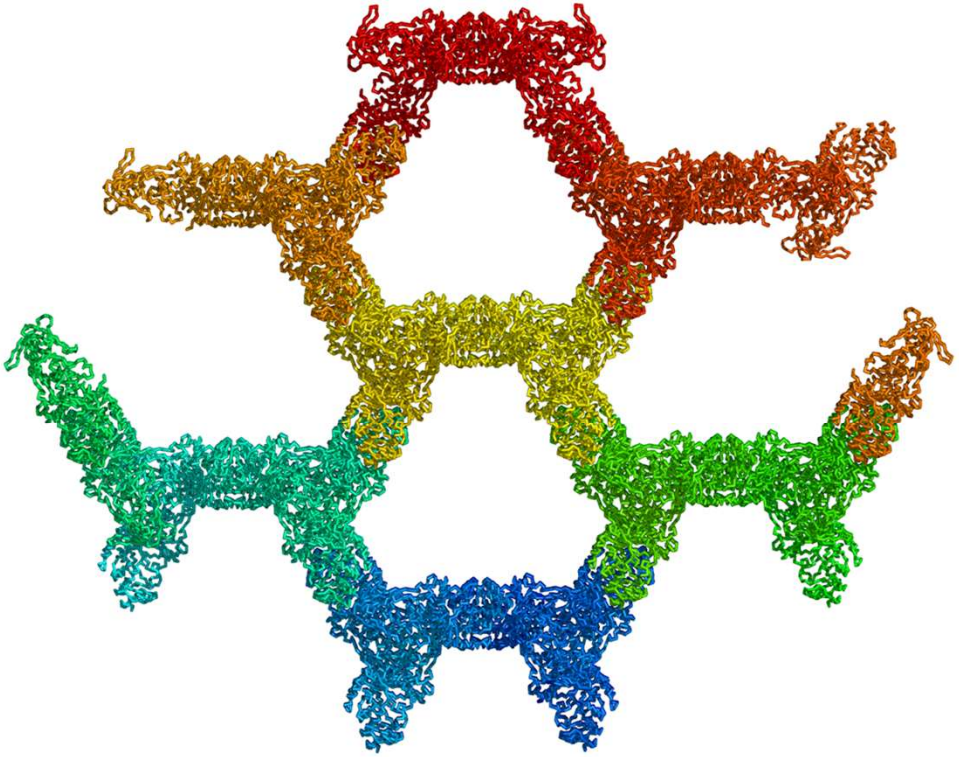
Structure solved (with help of the zinc signal) and refined with Phenix.

Overall R and R_{free} are 15.9 and 21.1% respectively.



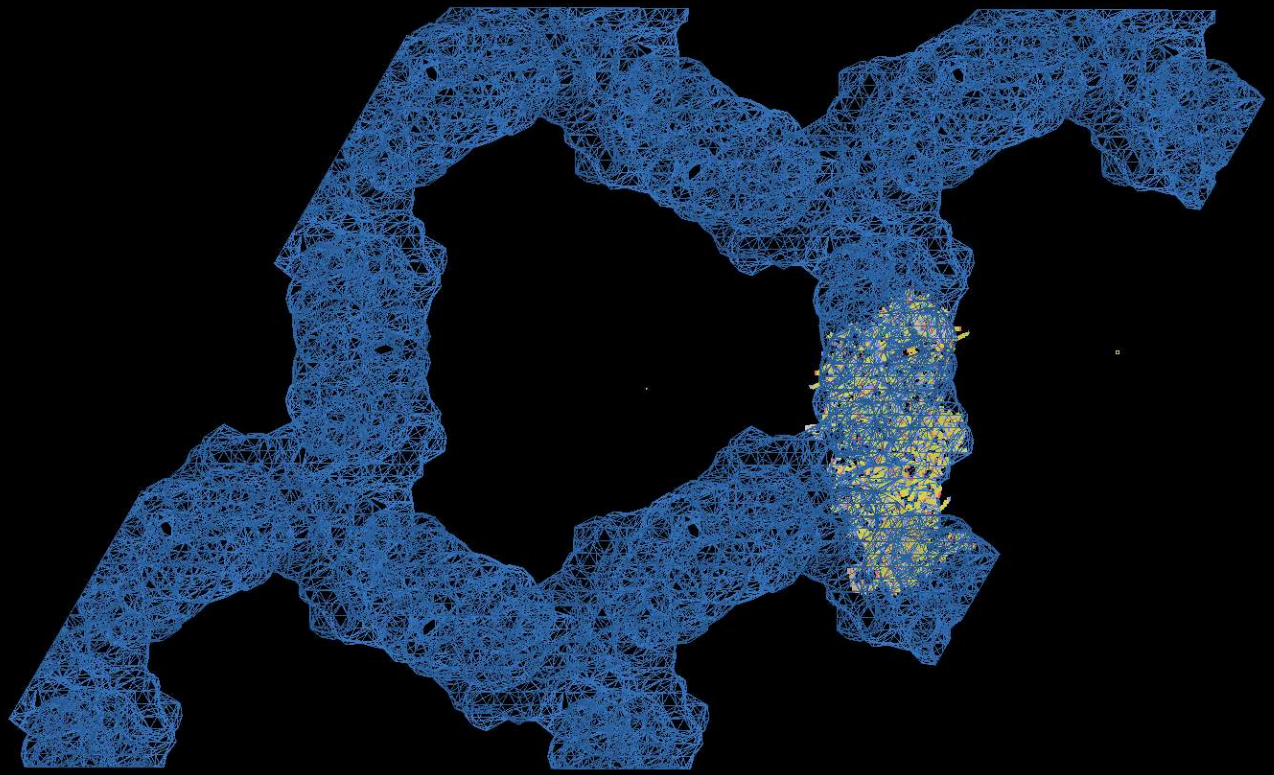

Yeast structure
?
809 residues
E. coli.
structure
553 residues





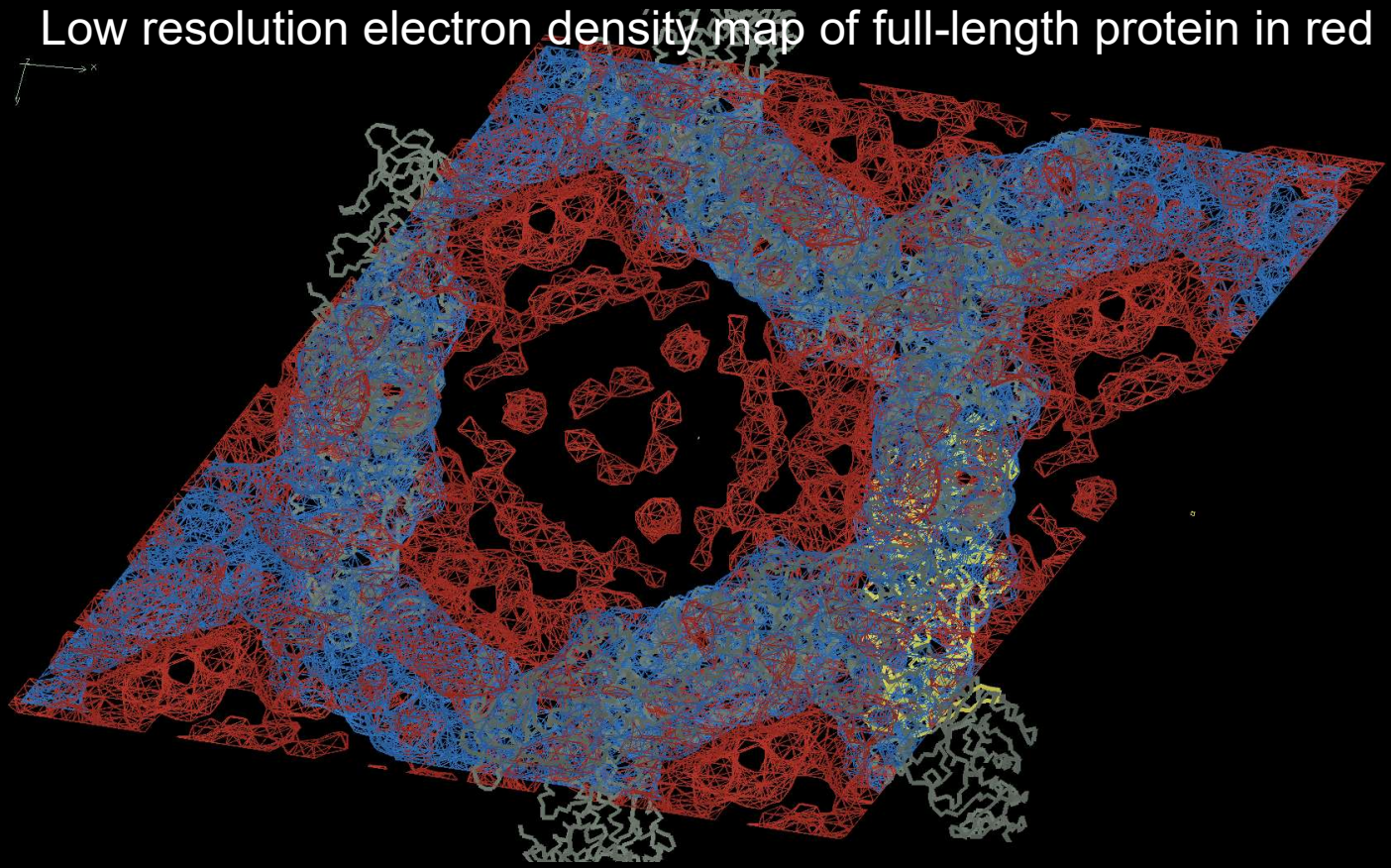
Packing diagram

Protein with N-terminal arm cleaved



Crystallized, data truncated to 20A (data to 78A still plenty of reflections due to geometry and wavelength used purposely used for data collection)

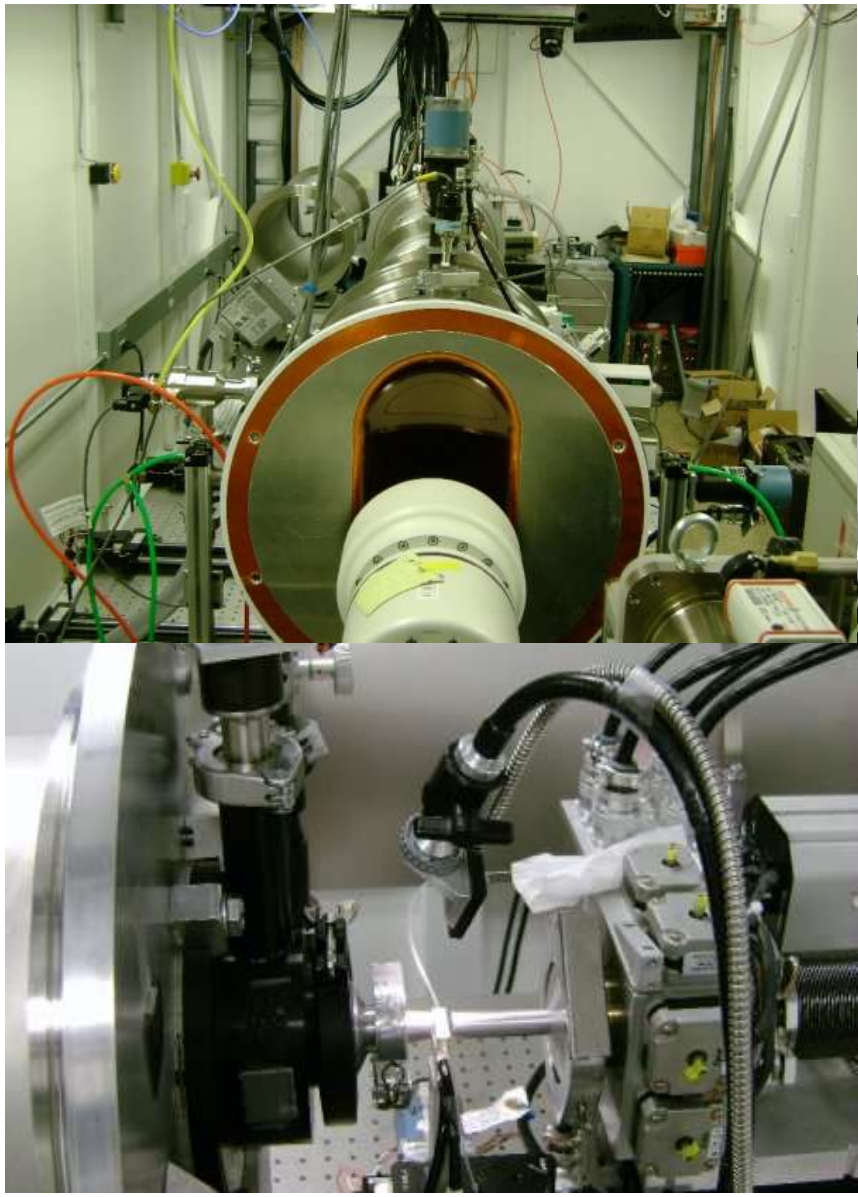
Low resolution electron density map of full-length protein in red



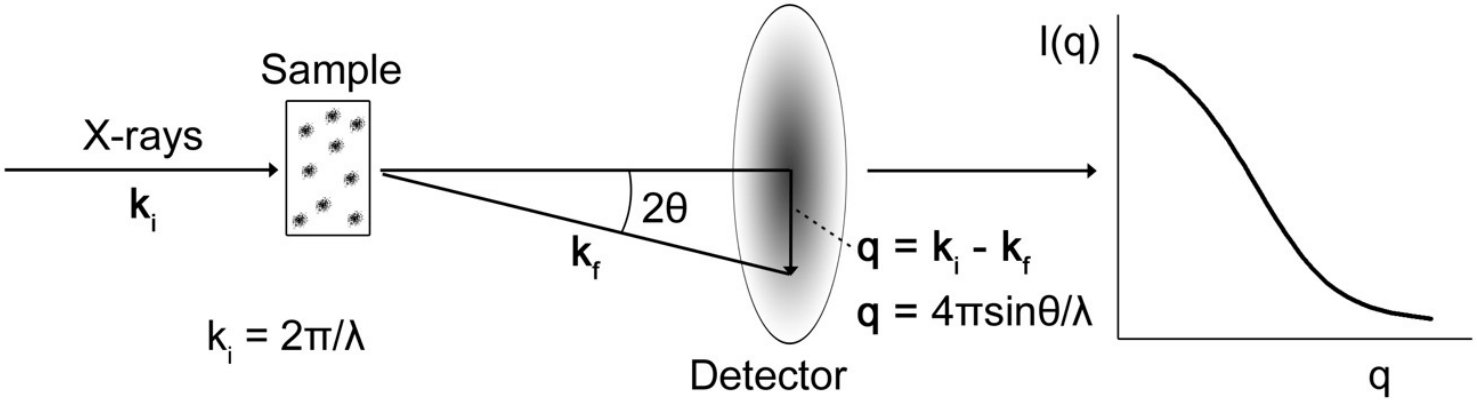
Data truncated to 20A (data to 78A still plenty of reflections due to geometry and wavelength used purposely used for data collection)

Small-Angle X-ray Scattering

- Sample Requirements:
  - Monodisperse, soluble protein, 10s of µL. A series of concentrations and a buffer blank.
- Where/How:
  - Synchrotrons or lab setups (e.g., SSRL, APS, CHESS, ALS, ESRF).
- Data/Output:
  - Shape envelopes, flexibility info, oligomerization state
- Pros/Cons:
  - Works in solution, low resolution
- Use Cases:
  - Disordered regions, dynamics, complexes



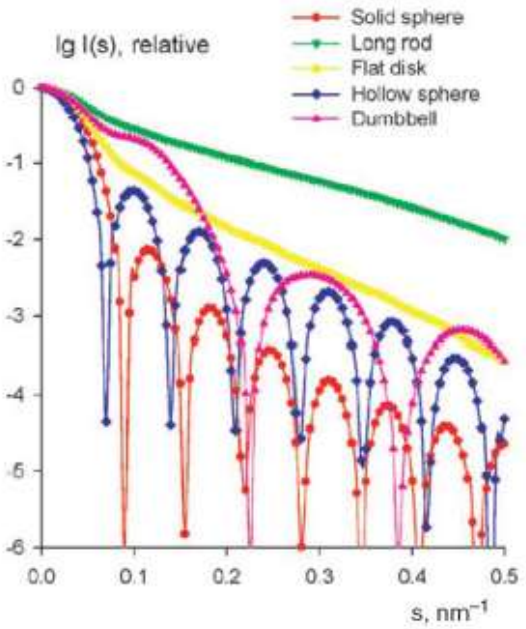
SAXS images everything behind the beamstop



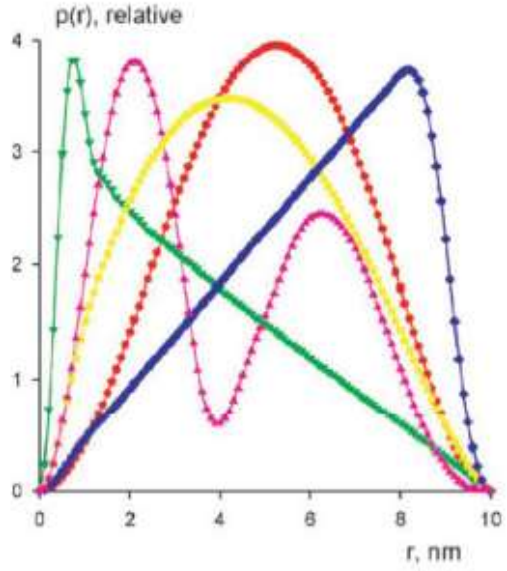
- Particles in solution tumble – spherically averaged intensity is recorded.
- Radial integration results in one dimensional SAXS profile.
- Larger particles scatter at smaller angles.
- Analysis of the 1D profile yields information about size and shape.

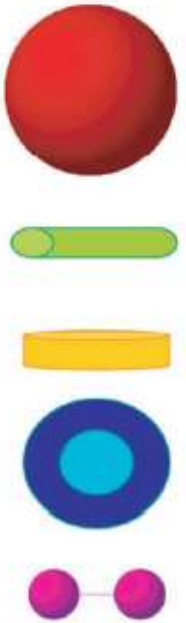
Shapes & scattering patterns

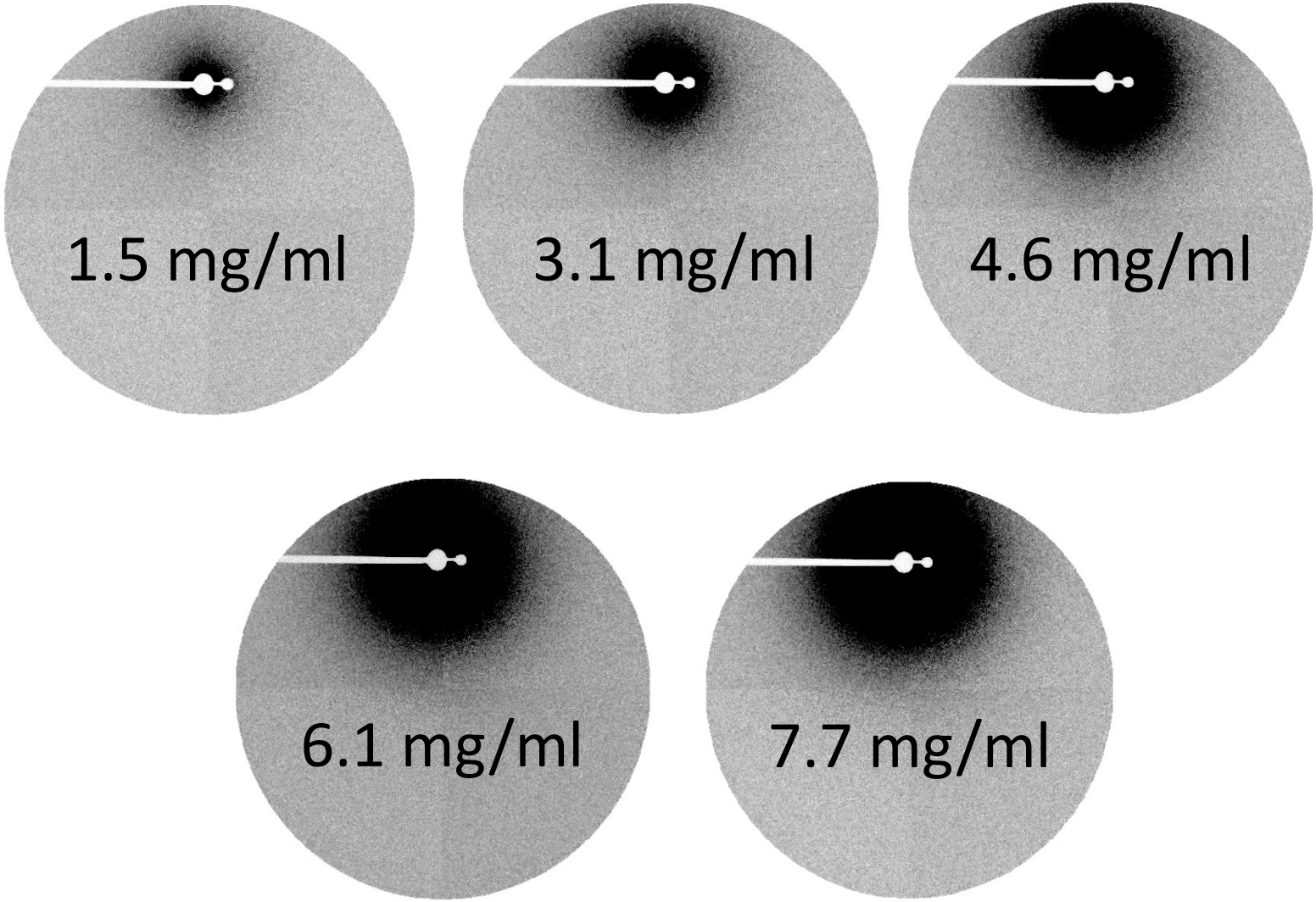
SAXS patterns



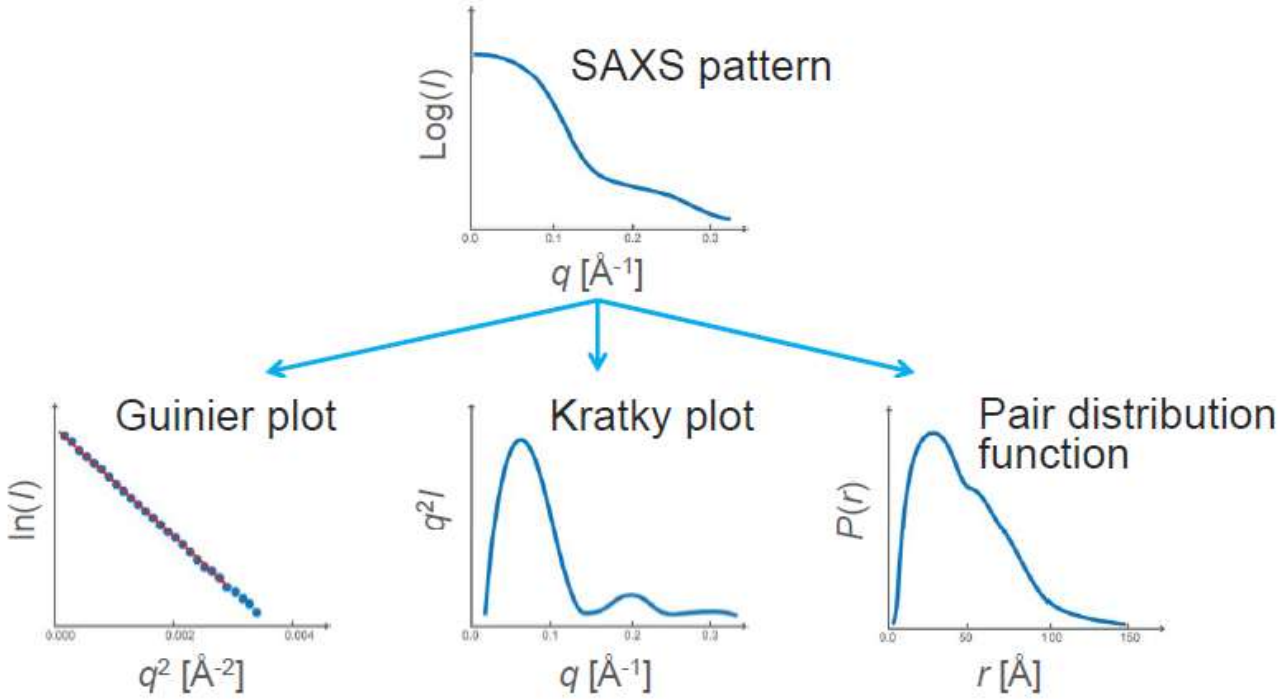
Pair distribution functions

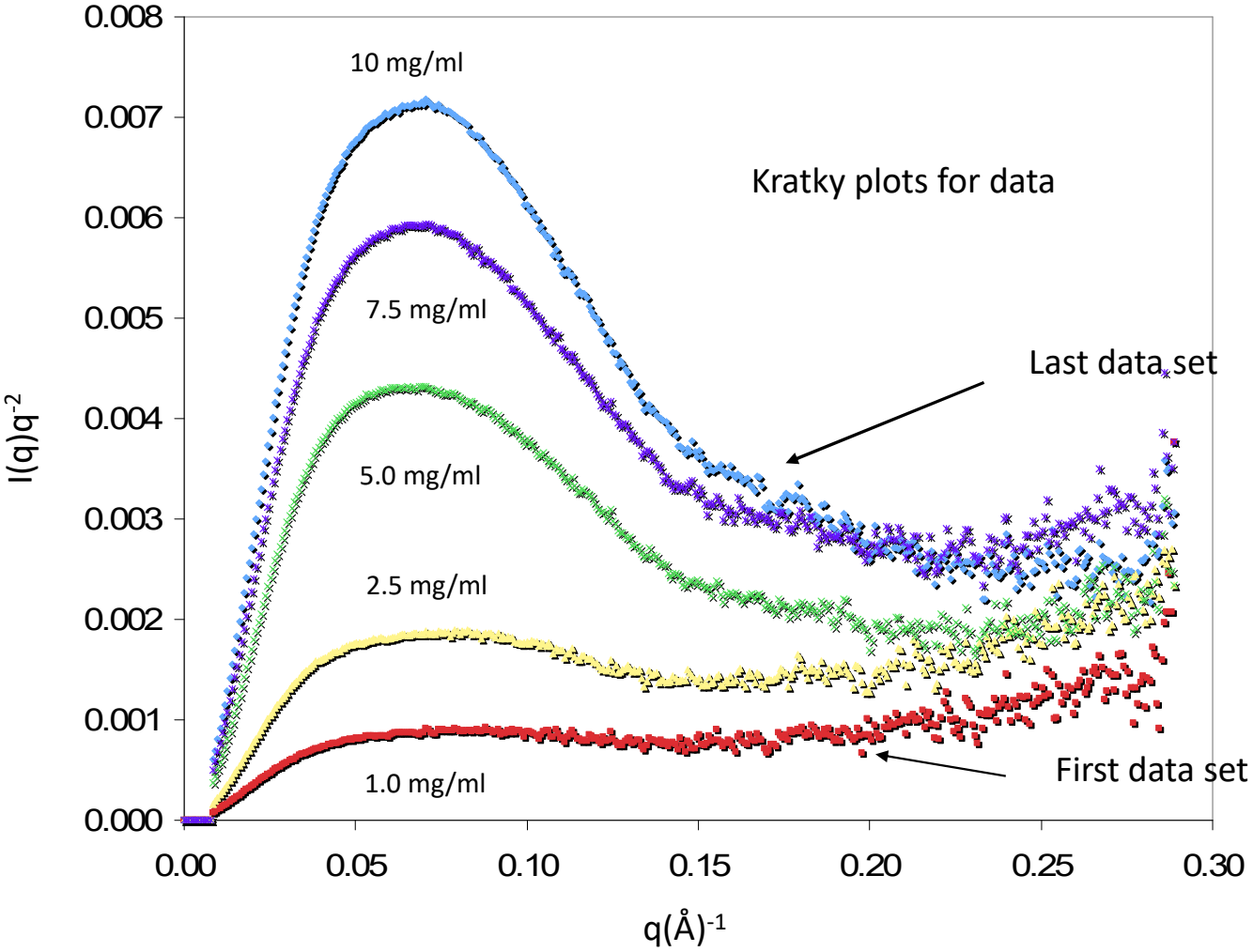




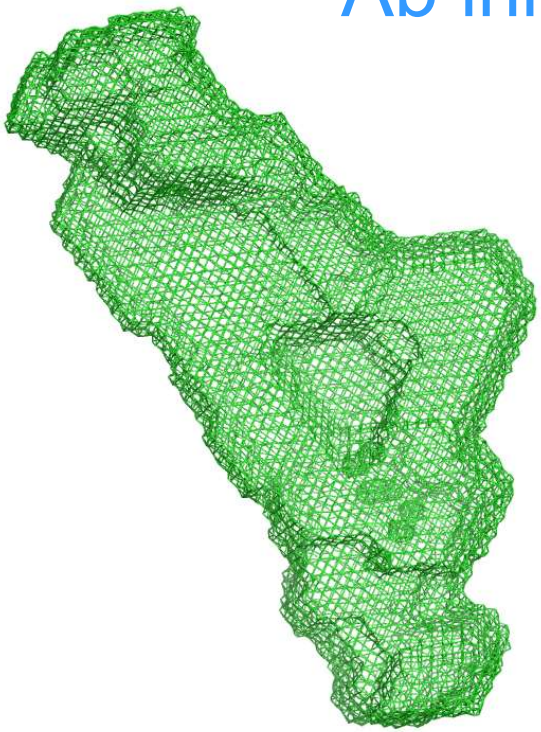

1.5 mg/ml
3.1 mg/ml
4.6 mg/ml
6.1 mg/ml
7.7 mg/ml

Profile conversion

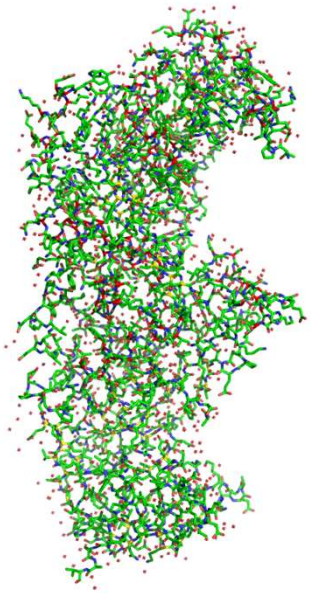



0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
I(q)q-2
10 mg/ml
7.5 mg/ml
5.0 mg/ml
2.5 mg/ml
1.0 mg/ml
Kratky plots for data
Last data set
First data set
0.00
0.05
0.10
0.15
0.20
0.25
0.30
q(Å)-1

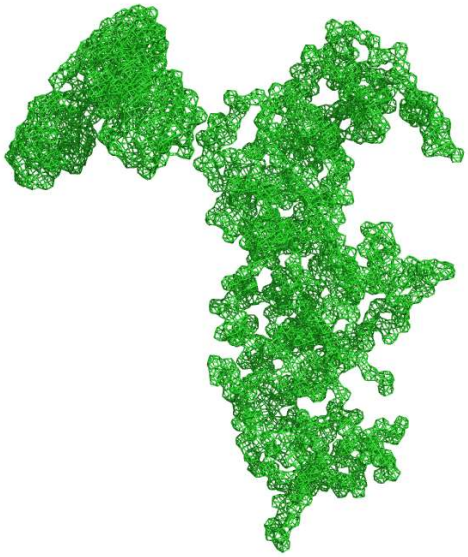
Ab initio reconstruction



Completely *ab initio*
(samples dynamic range)

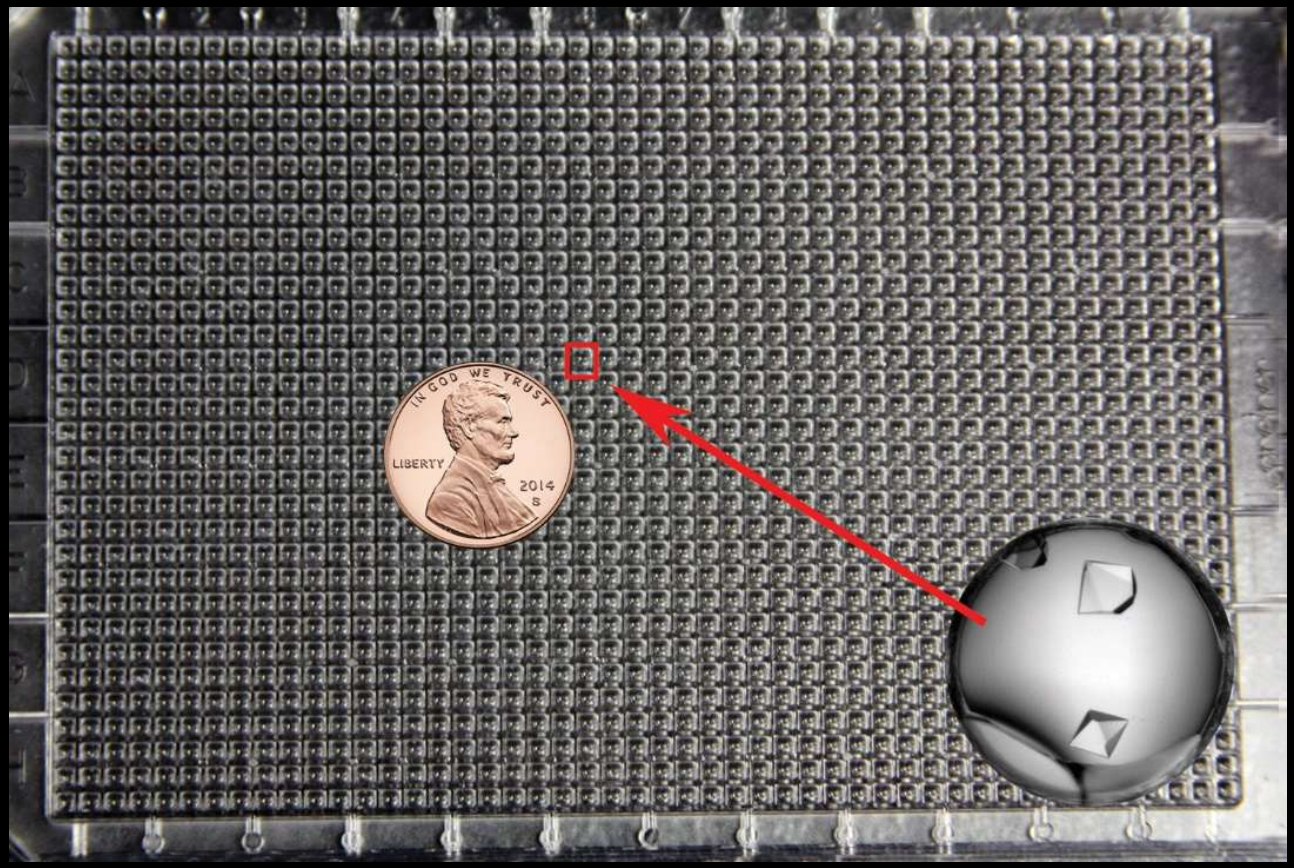


The crystal
structure



SAXS with the
crystal structure

Minimize sample volume

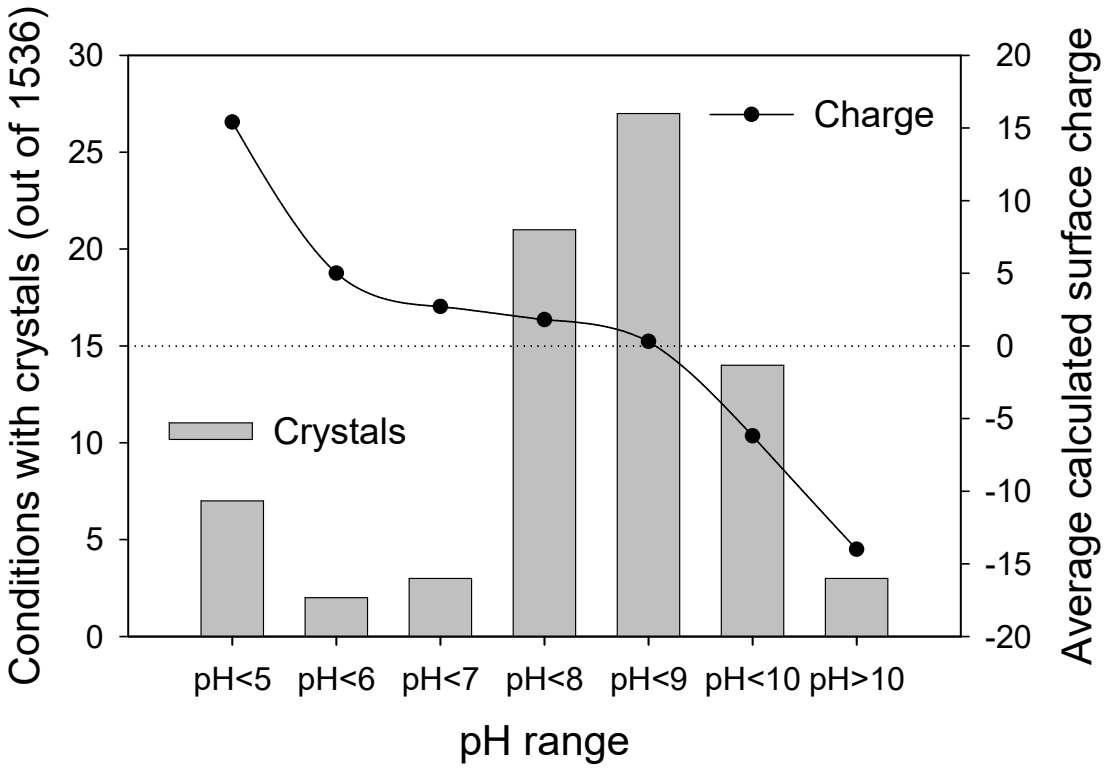


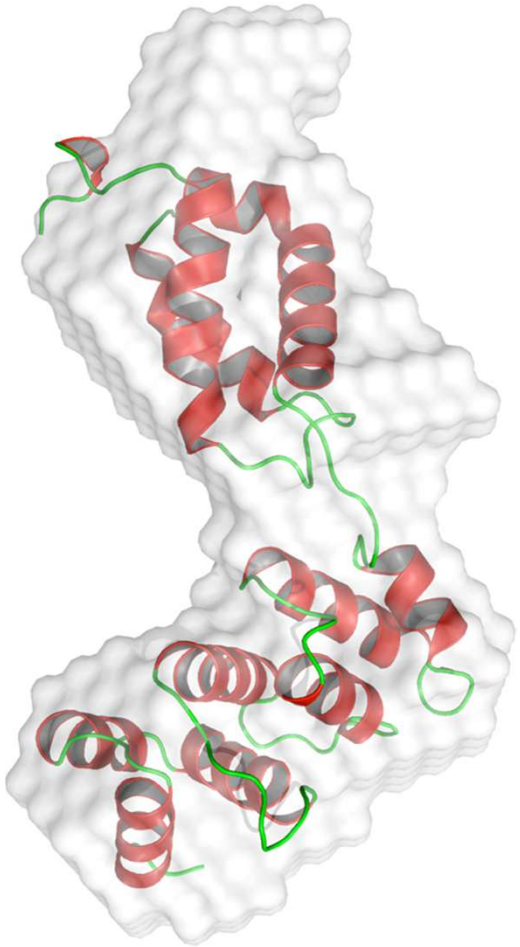
NATIONAL CRYSTALLIZATION CENTER

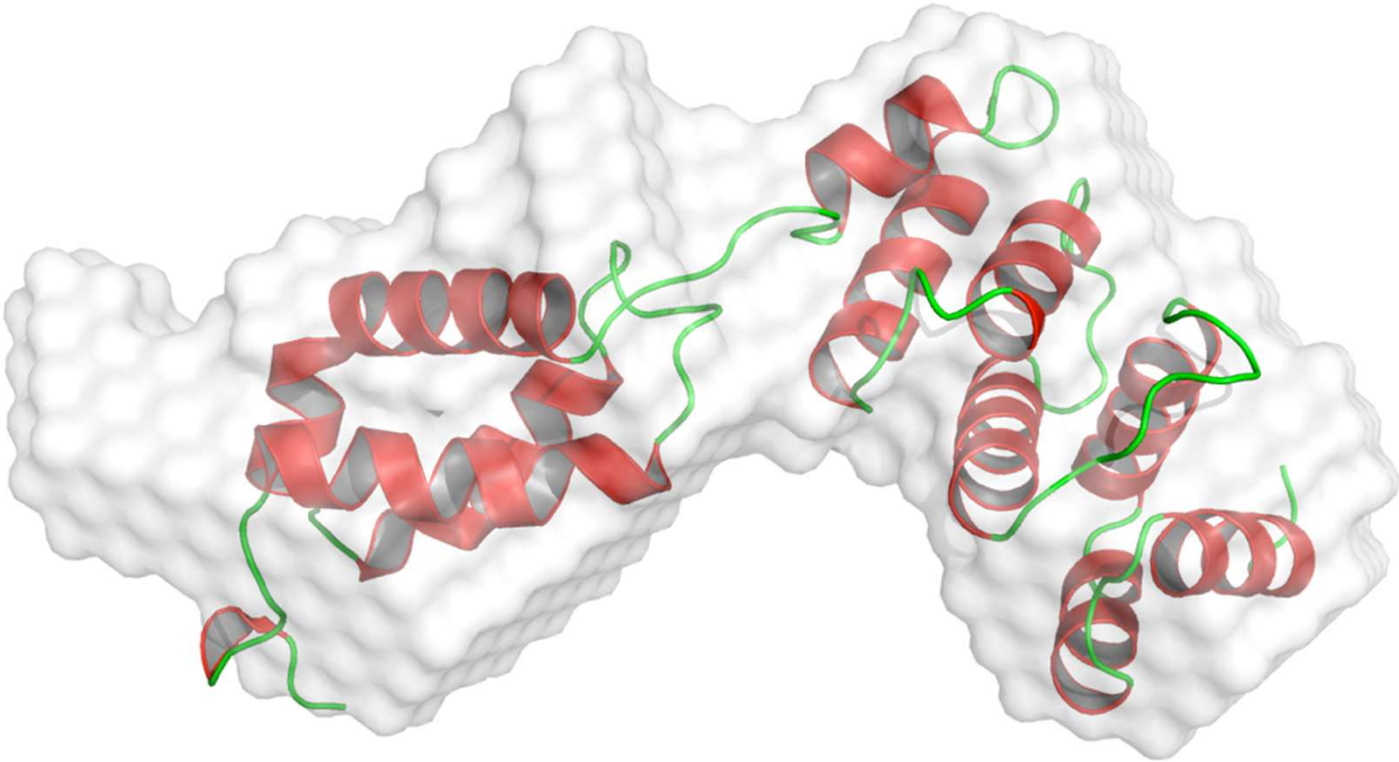
https://hwi.buffalo.edu/high-throughput-crystallization-center

- 375 µl of protein sample
- ~10 mg/ml concentration
- 1,536 conditions set up
- Visual, UV, and SONICC imaging
- $400 per sample

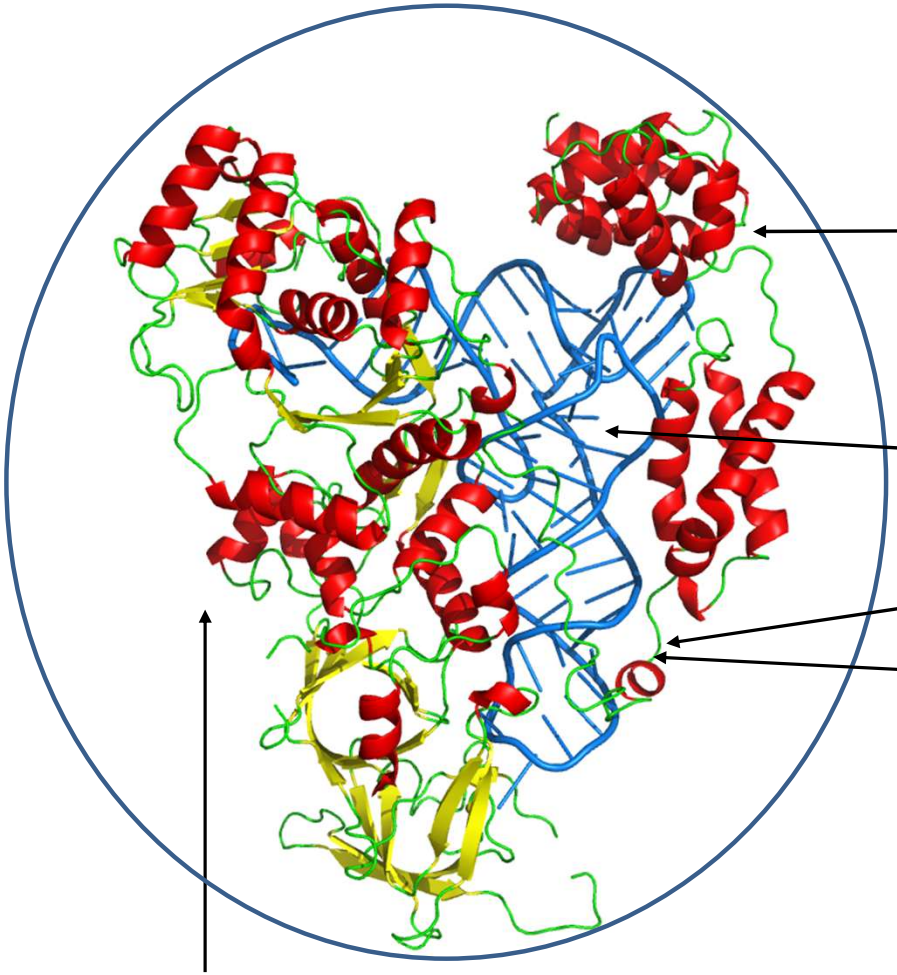
Crystallization trials of the N-terminal domain







Eukaryotic Gln tRNA synthetase



SAXS data indicating a larger but well folded system in solution

The truncated terminal was crystallized

It was extracted directly from the screening plate and X-rayed to give the structure.

tRNA was docked in

SAXS aided by sequence analysis identified a flexible region

Homology modeling indicated a flexible region

Crystallized the C-terminal in the standard screen, conditions chosen that were already known to be good cryo-conditions.

A combination of crystallography, SAXS, homology modeling and computational modeling was used to give the complete structure and tested by biochemical analysis.

Structural studies alone are not enough to fully understand mechanism

Nuclear Magnetic Resonance (NMR)

Magnetic Field Alignment

- Certain atomic nuclei (like hydrogen-1, carbon-13, nitrogen-15) have a property called spin, which makes them behave like tiny magnets. When placed in a strong magnetic field, these spins align either with or against the field.

Radio Frequency (RF) Pulse

- A short RF pulse is applied to the sample, perturbing the spin alignment. After the pulse is turned off, the nuclei gradually return to their original alignment emitting energy.

Signal Detection

- The emitted energy depends on the chemical environment of the nucleus, i.e., the atoms nearby and their bonds.

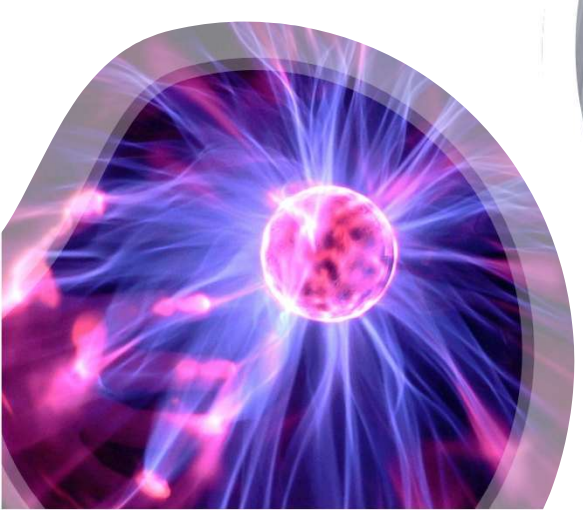
Data Interpretation

By analyzing the chemical shifts, coupling patterns, and relaxation times, we can:

- Identify which atoms are present.
- Determine how atoms are connected (bonded and through space).
- Infer distances between atoms (via NOE – Nuclear Overhauser Effect).
- Determine the 3D structure of the protein.
- Study flexibility and motion within the protein.

Nuclear Magnetic Resonance

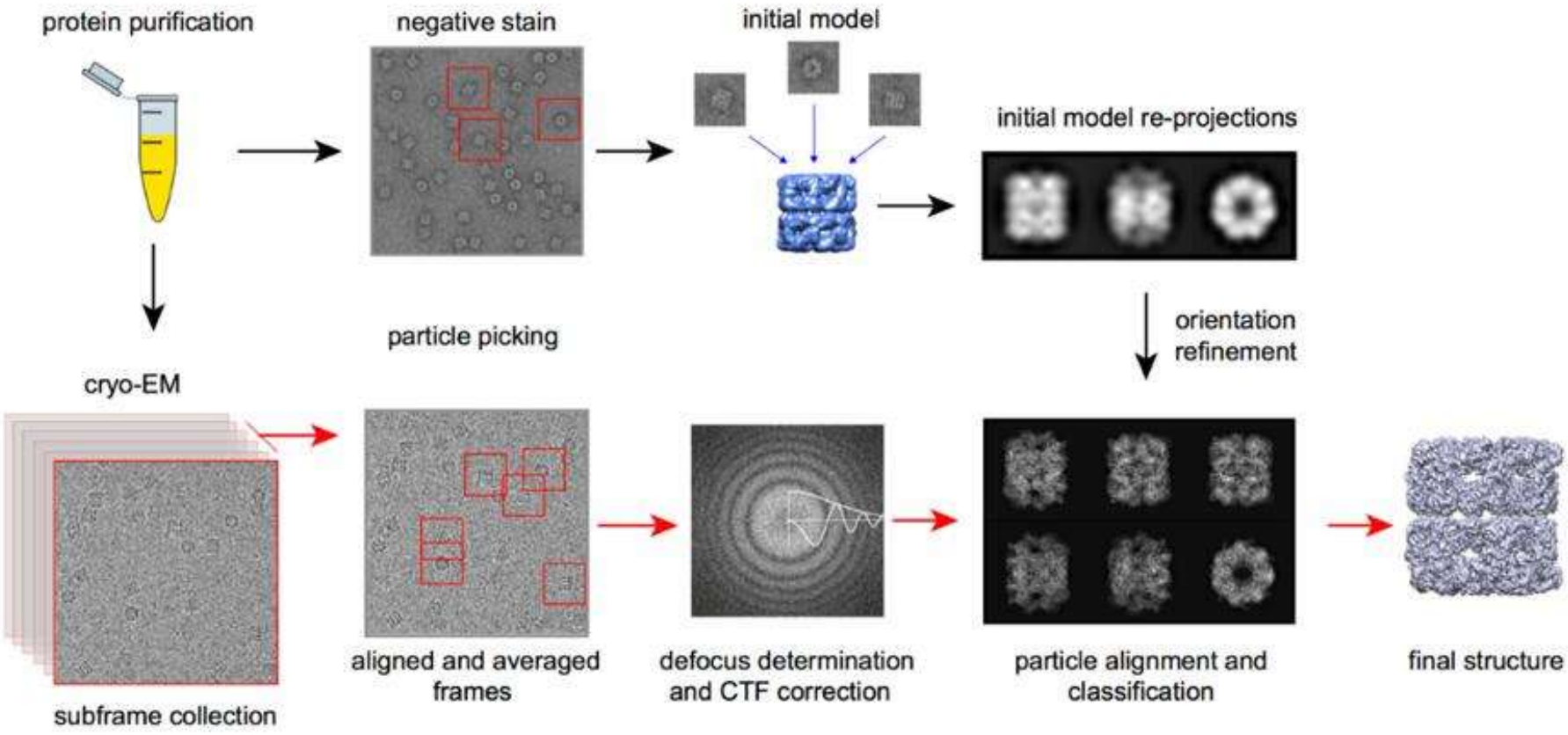
- Sample Requirements:
  - Isotopically labeled proteins, soluble, ~0.5 mM
- Where/How:
  - 600+ MHz instruments, isotopic labeling labs
- Data/Output:
  - Dynamics, local structure, binding interfaces
- Pros/Cons:
  - Excellent for small flexible proteins, size limits
- Use Cases:
  - Disordered proteins, kinetics, allostery



Cryo-Electron Microscopy

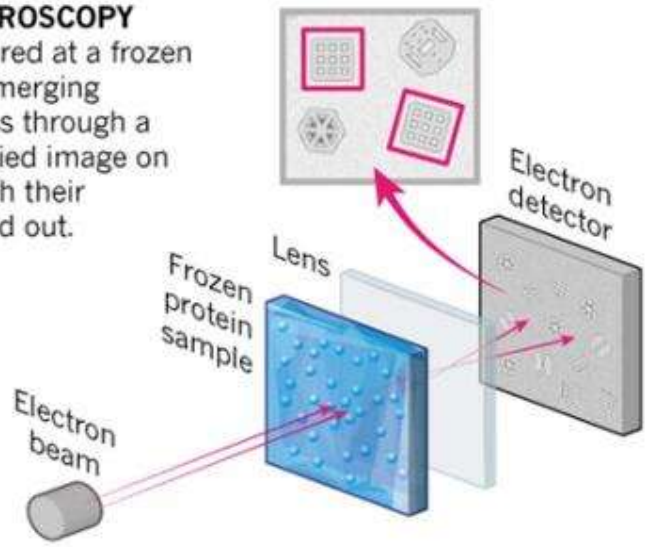
- Sample Requirements:
  - Frozen-hydrated samples, ~0.1 mg/mL, ideally >100 kDa
- Where/How:
  - Access centers (e.g., NYSBC, Stanford, EMDB sites)
- Data/Output:
  - 3D density maps; atomic resolution for larger complexes
- Pros/Cons:
  - No crystallization, great for large complexes, expensive setup
- Use Cases:
  - Membrane proteins, ribosomes, virus particles

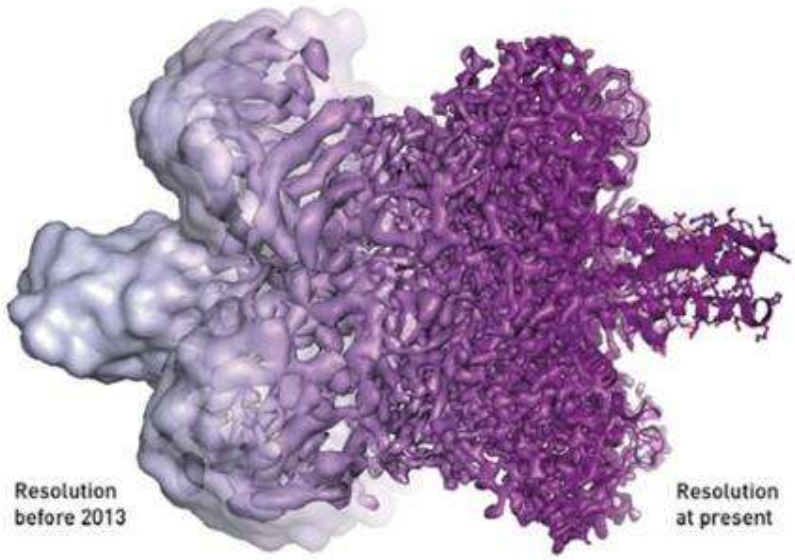
Cryo-EM



From https://www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

CRYO-ELECTRON MICROSCOPY
A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.





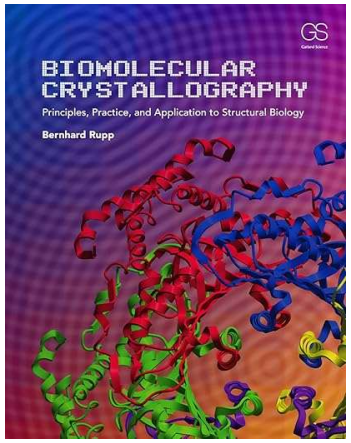
From https://www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

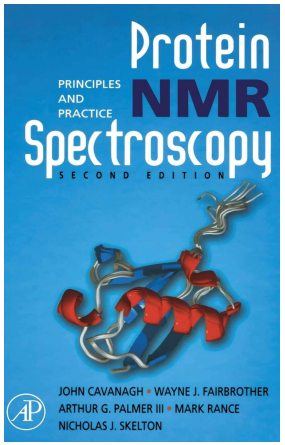
Technique	Advantages	Disadvantages	Samples	Resolution
X-ray Crystallography	• Well developed • High resolution • Broad molecular weight range • Easy for model building • Many instruments available	• Difficult for crystallization • Difficult for diffraction • Solid structure preferred • Static crystalline state structure • Destructive	• Crystallizable samples • Soluble proteins, membrane proteins, ribosomes, DNA/RNA and protein complexes	High
Nuclear Magnetic Resonance	• High resolution • 3D structure in solution • Good for dynamic study • Non-destructive	• Need for high sample purity • Difficult for sample preparation • Difficult for computational simulation • Smaller MW samples	• MWs below 40–50 kDa • Water soluble samples	High
Cryo-Electron Microscopy	• Easy sample preparation • Small sample size	• Applicable to samples of high molecular weights only • Highly dependent on EM techniques • Costly EM equipment • Significant data processing requirements • Destructive	• >100 kDa, but that limit is falling • Virions, membrane proteins, large proteins, ribosomes, complex compounds	Relatively Low (<3.5 Å) but rapidly improving
Small Angle X-ray Scattering	• Easy sample preparation • Many instruments available	• Very low resolution • Interpretation of data can be somewhat qualitative • Sensitive to aggregation • Destructive	• Solution samples • Few mg/ml • Buffer blank • Several concentrations	Low

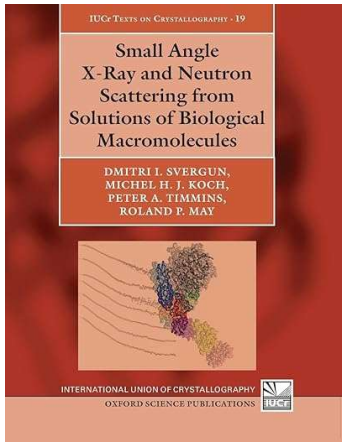
How to choose?

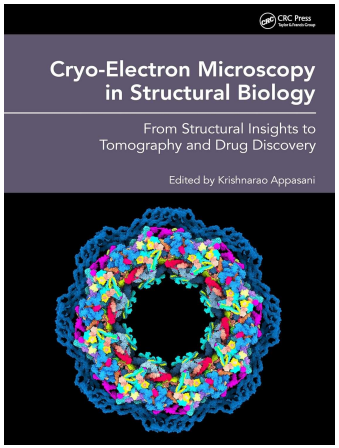
- Sample type and stability
- Size
- Flexibility
- Resolution needs
- Time/resource availability

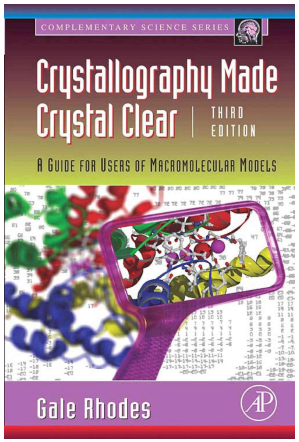
Recommended books

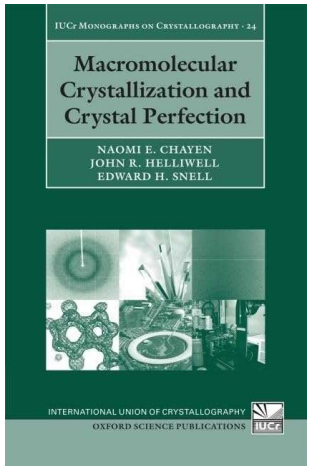


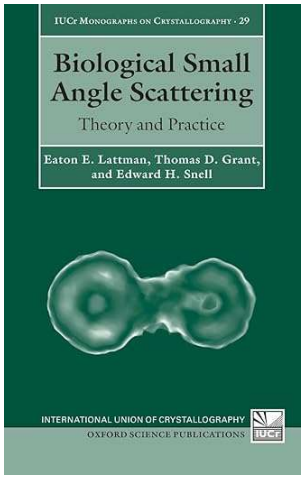








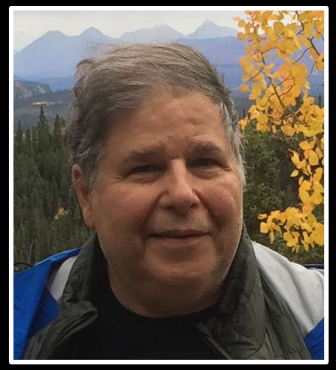




Thanks to an enormous number of people







BioXFEL

University at Buffalo
Hauptman-Woodward Research Institute

SSRL

Funding from

NSF

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

DEFENSE THREAT REDUCTION AGENCY · UNITED STATES OF AMERICA

Questions?

esnell@buffalo.edu