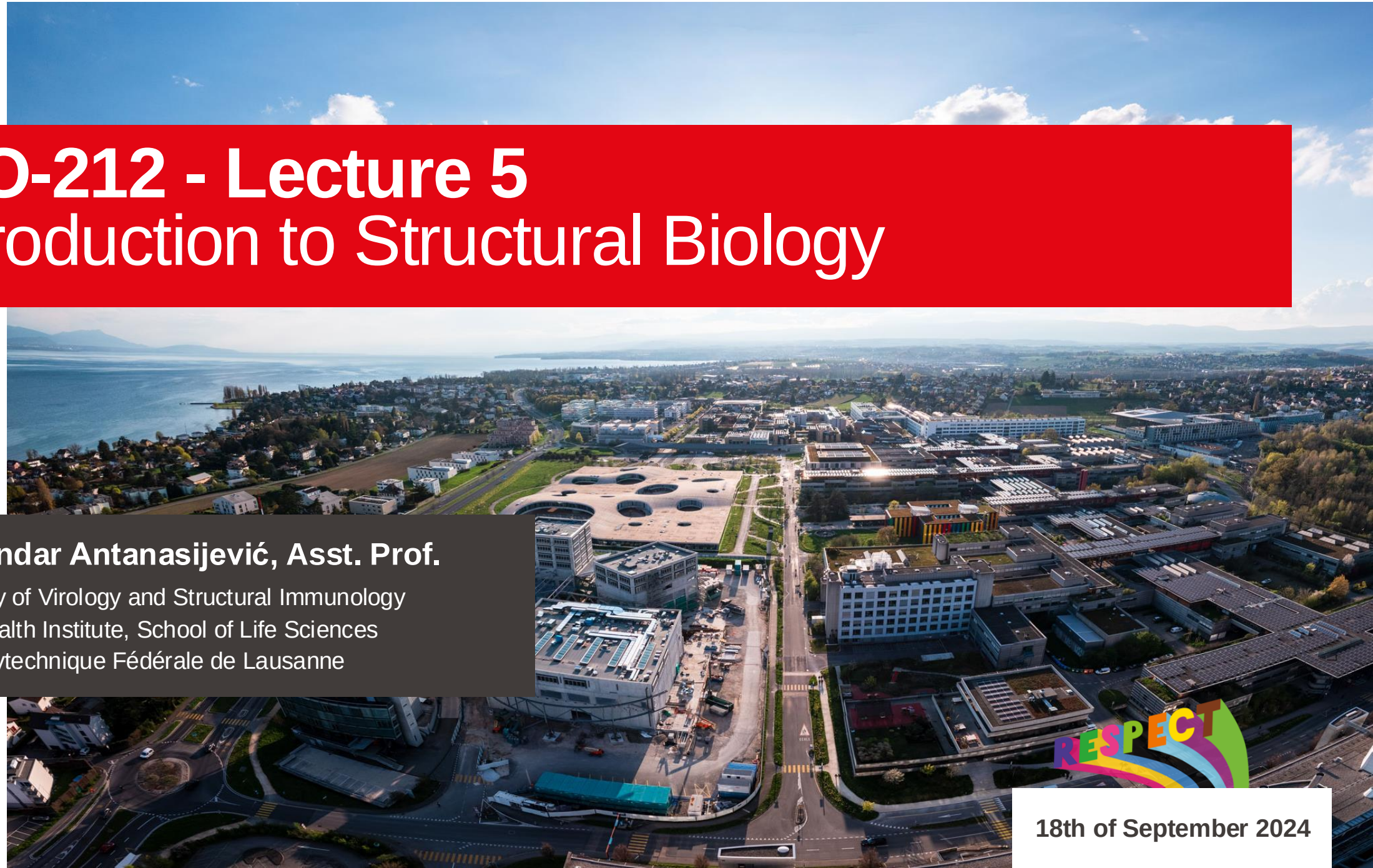
BIO-212 - Lecture 5
Introduction to Structural Biology



Aleksandar Antanasijević, Asst. Prof.

Laboratory of Virology and Structural Immunology
Global Health Institute, School of Life Sciences
École Polytechnique Fédérale de Lausanne

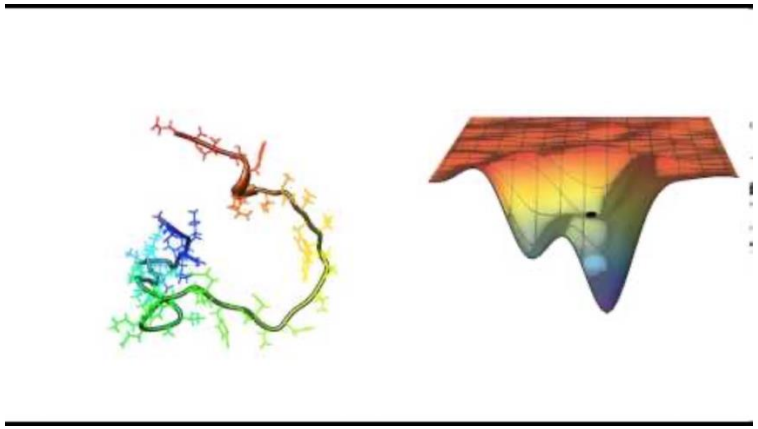
École polytechnique fédérale de Lausanne

18th of September 2024

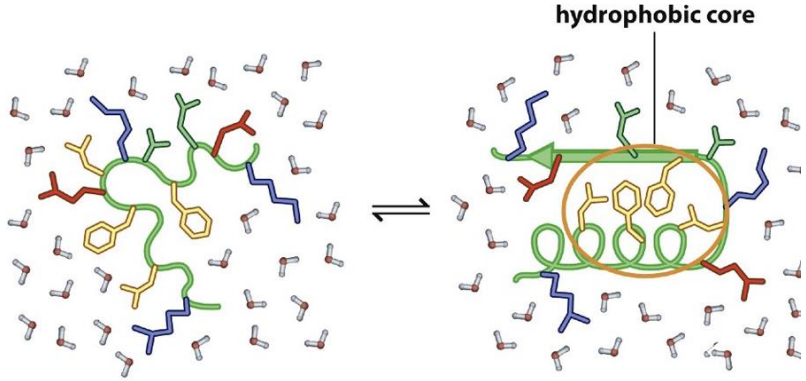
Lecture 4 – Quick Summary

• Thermodynamics of protein folding

- Proteins folding is about reaching the state of minimum free energy

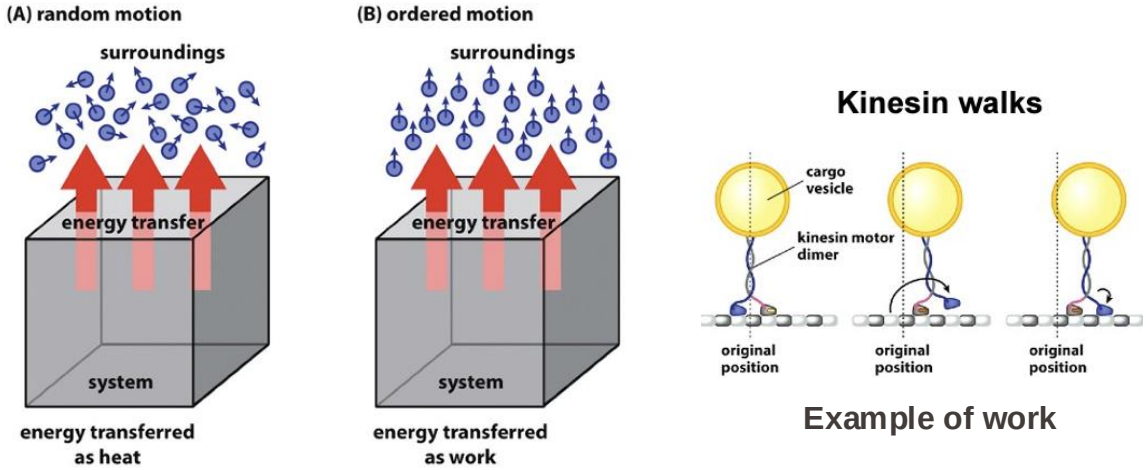


- Hydrophobic core formation as an essential driver



• Work and heat in the context of biological systems

- Energy released by chemical reactions is converted into heat and work



- Free energy, internal energy, enthalpy and enthropy

$$G = H - TS$$

- Heat capacity of macromolecules changes depending on the conformational (energy) state

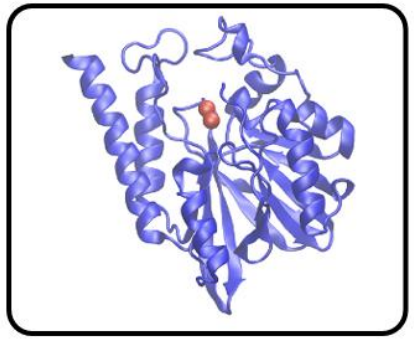
Why do we care about biomolecule structures?

- Basic paradigm in biochemistry

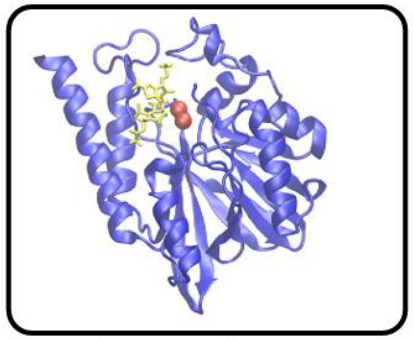
Sequence

```
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
TPDDHLVARWCATPVAEVAEKSARF
PPRPEELLLPREETLPEMYSLSFTA
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
```

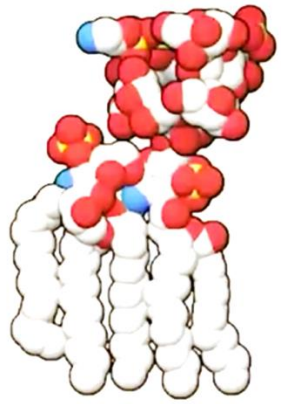
Structure



Function

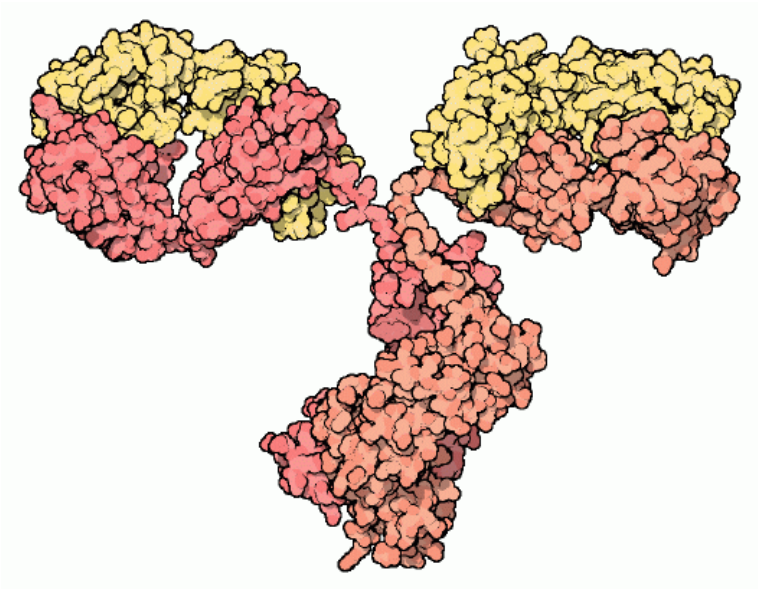


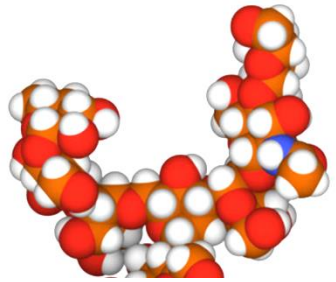
- Examples of different biomolecules and their functions



Glycerophospholipids
(building membranes)

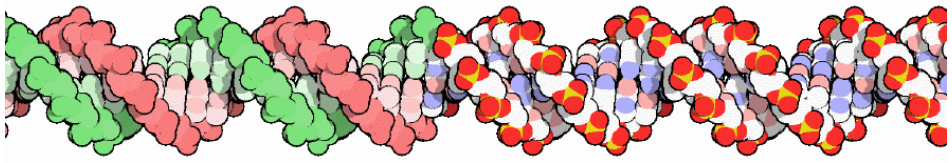
Antibody (immune defense)





Carbohydrates
(energy storage)

DNA (carrier of genetic information)

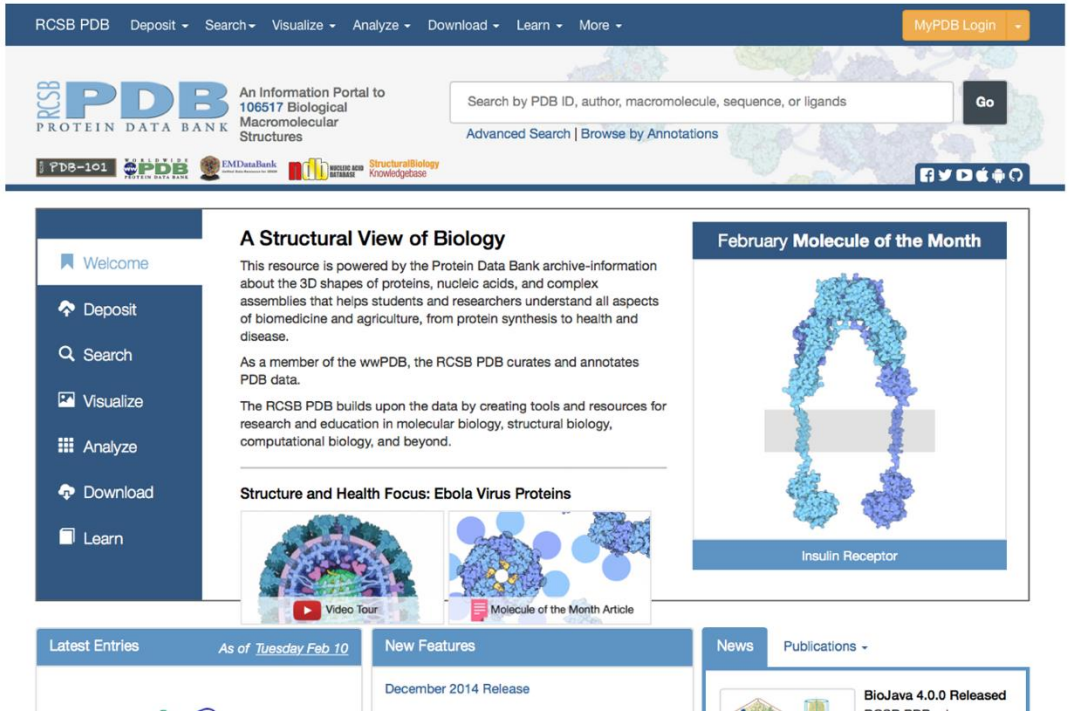


What is a structure?

- Generally speaking, a structure is a file containing information on (1) atoms comprising the biomolecule, (2) their coordinates in 3D space and (3) connections via covalent bonds
- The files are typically in a *.pdb or *.cif formats and can be opened with programs such as Pymol

Structures are deposited to Protein Data Bank (PDB)

http://www.rcsb.org



Structures are text files with atom information

```
ATOM      3  C    ASN C  33      94.970 129.157 158.704  1.00120.39           C
ATOM      4  O    ASN C  33      95.150 128.023 159.150  1.00120.34           O
ATOM      5  CB   ASN C  33      93.857 130.343 160.569  1.00122.96           C
ATOM      6  CG   ASN C  33      93.999 131.366 161.635  1.00123.84           C
ATOM      7  OD1  ASN C  33      95.115 131.727 162.029  1.00124.33           O
ATOM      8  ND2  ASN C  33      92.891 131.856 162.121  1.00127.60           N
ATOM      9  N    LEU C  34      94.730 129.414 157.432  1.00120.36           N
ATOM     10  CA   LEU C  34      94.679 128.344 156.455  1.00116.36           C
ATOM     11  C    LEU C  34      95.989 128.297 155.723  1.00111.18           C
ATOM     12  O    LEU C  34      96.690 129.303 155.594  1.00112.82           O
ATOM     13  CB   LEU C  34      93.509 128.521 155.492  1.00118.74           C
ATOM     14  CG   LEU C  34      92.120 128.631 156.175  1.00125.86           C
ATOM     15  CD1  LEU C  34      91.033 128.827 155.119  1.00135.13           C
ATOM     16  CD2  LEU C  34      91.840 127.407 157.005  1.00123.16           C
ATOM     17  N    TRP C  35      96.320 127.106 155.280  1.00108.83           N
ATOM     18  CA   TRP C  35      97.545 126.765 154.595  1.00107.97           C
ATOM     19  C    TRP C  35      97.293 126.124 153.267  1.00108.62           C
ATOM     20  O    TRP C  35      96.242 125.525 153.041  1.00112.10           O
ATOM     21  CB   TRP C  35      98.348 125.842 155.490  1.00115.33           C
ATOM     22  CG   TRP C  35      98.782 126.548 156.676  1.00115.40           C
```

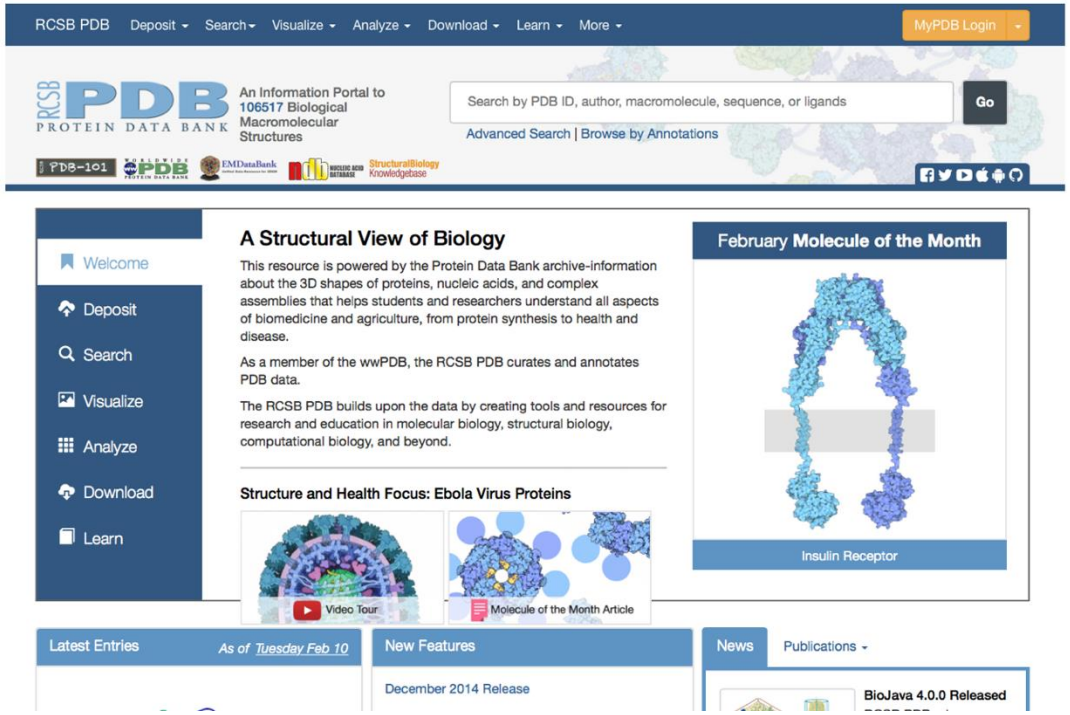
Atom number
Amino acid
Sequence position
Atom type
Atom denotation
Polypeptide chain
XYZ coordinates

What is a structure?

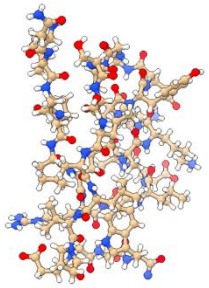
- Generally speaking, a structure is a file containing information on (1) atoms comprising the biomolecule, (2) their coordinates in 3D space and (3) connections via covalent bonds
- The files are typically in a *.pdb or *.cif formats and can be opened with programs such as PyMol

Structures are deposited to Protein Data Bank (PDB)

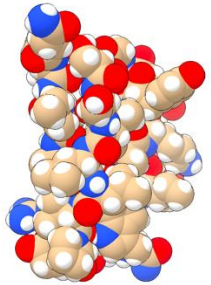
http://www.rcsb.org



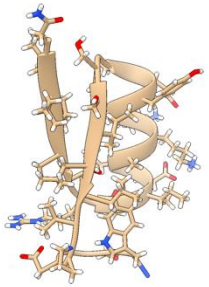
Different representations of a structure



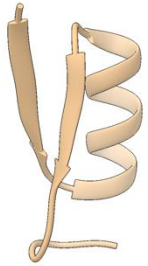
All atom (sticks)



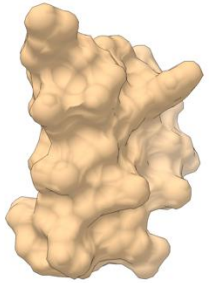
All atom (balls)



Ribbon + Side chains



Ribbon

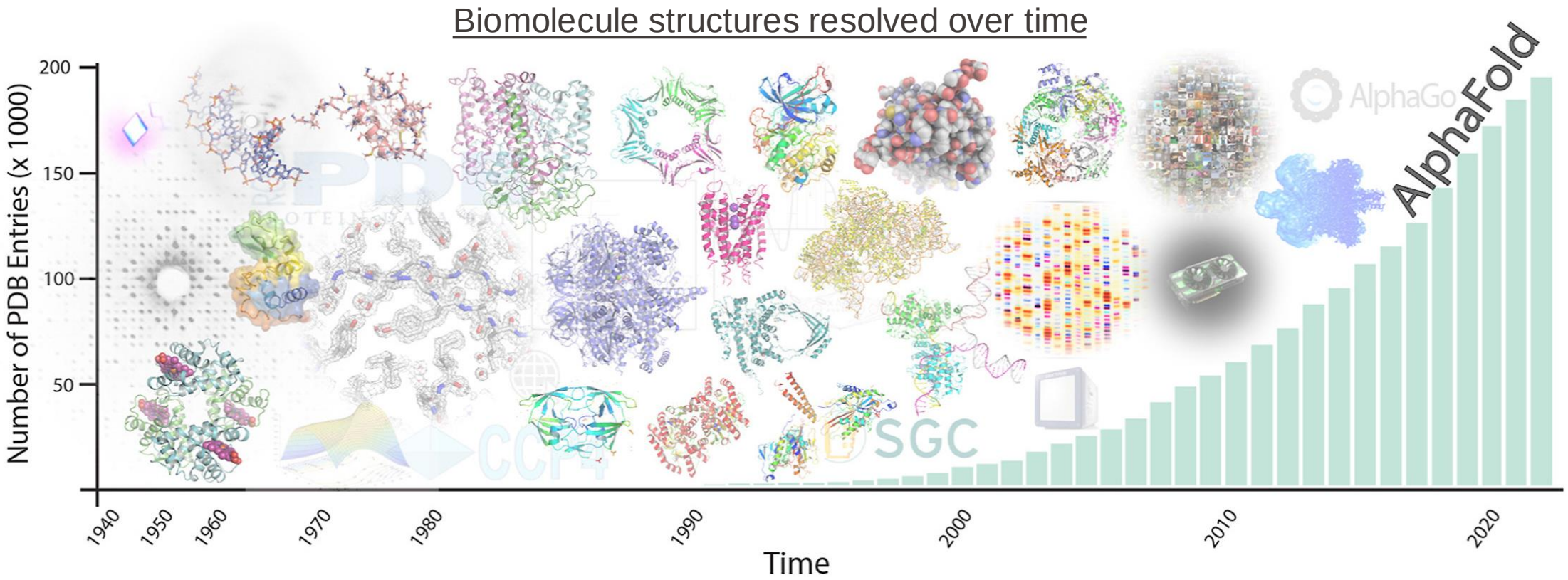


Surface

All atom views can be busy so it is often advantageous to show secondary structure or whole surface view.

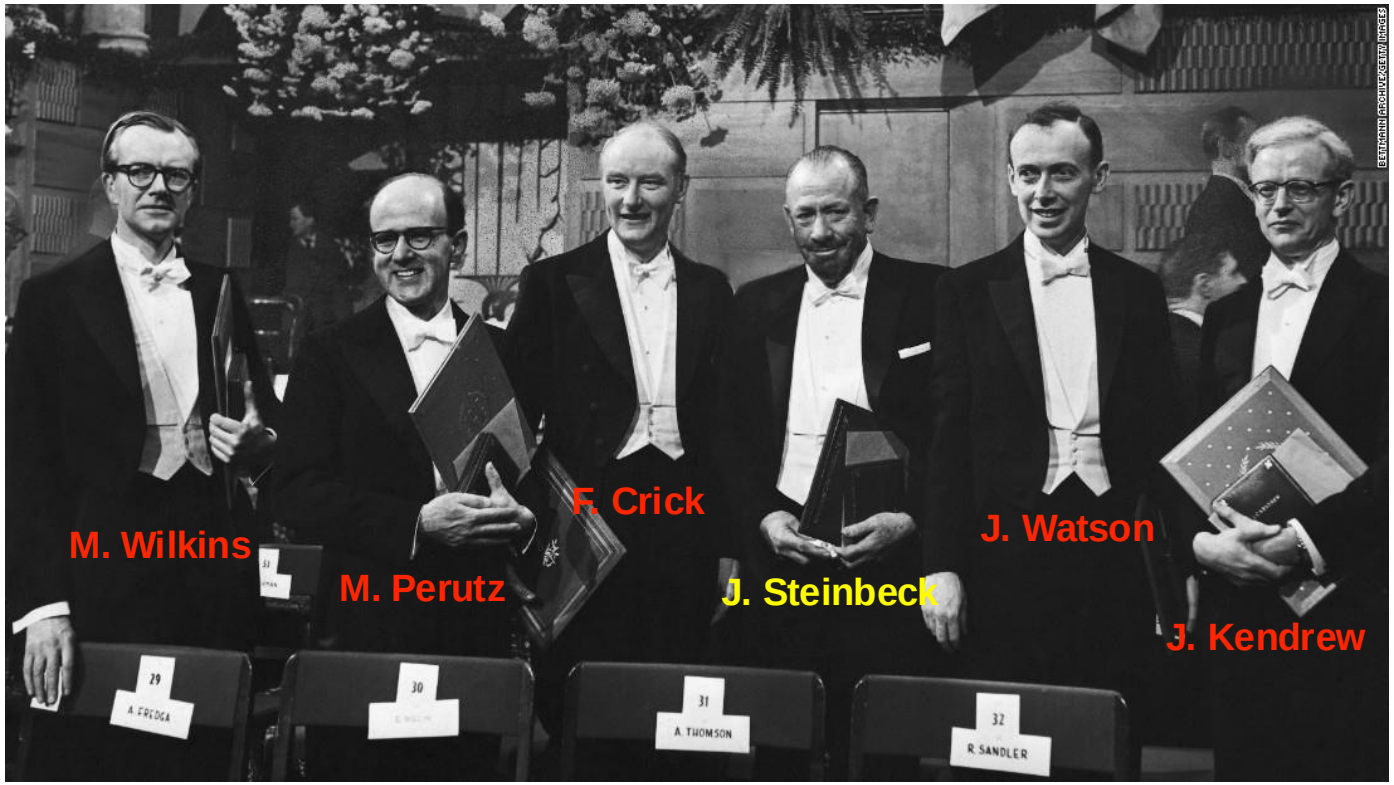
Structural biology - A subfield of Biochemistry

- Structural biology is the study of molecular structure and dynamics of biological macromolecules, particularly proteins and nucleic acids, and how alterations in their structure affect their function.
- Total number of available experimentally determined structures to date is ~225k
- Computational tools (e.g., AlphaFold) can predict macromolecular structures *in silico* allowing to expand the available structure space



History of experimental structure determination

- Due to the challenging nature of the experiments and the importance for understanding biological systems, major structural biology efforts have been rewarded with Nobel Prizes

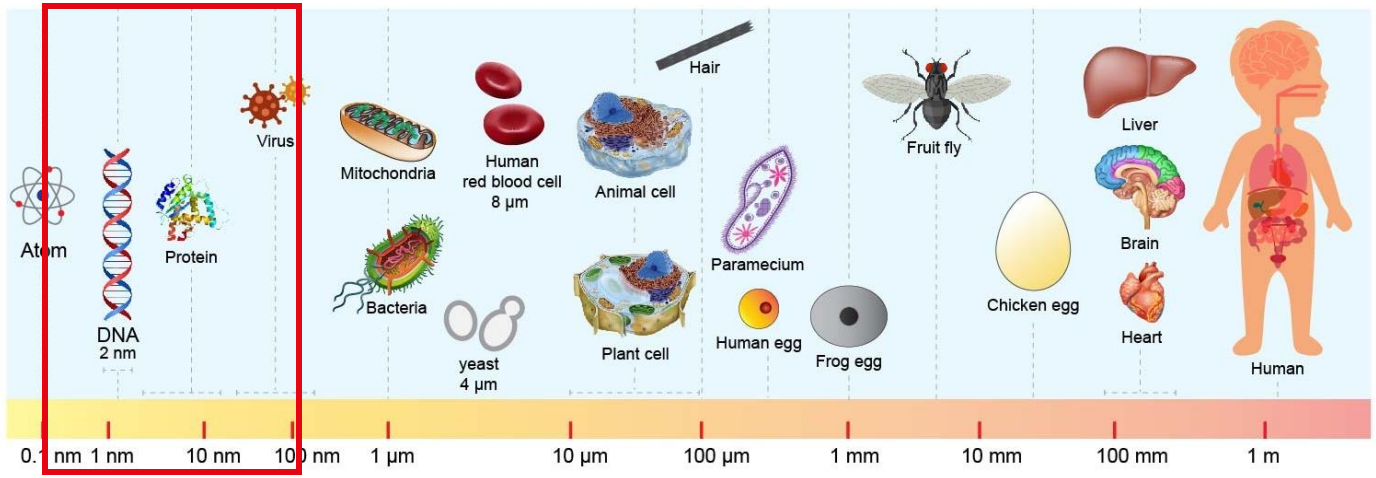


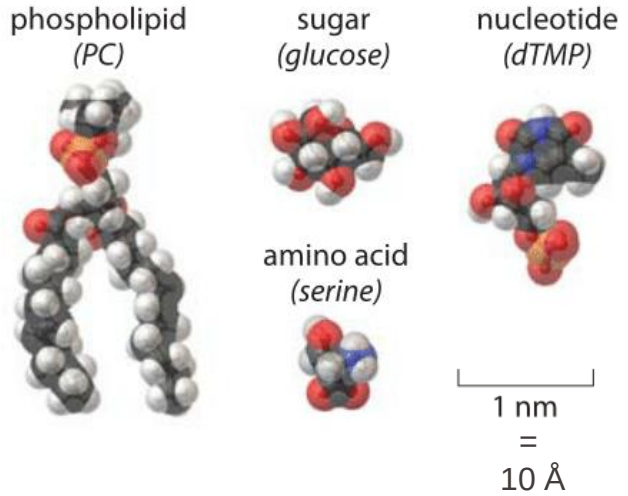
1962 – Nobel Year of Structural Biology

2020: Chemistry	Charpentier and Doudna	CRISPR-Cas9x
2017: Chemistry	Dubochet, Henderson, Frank	development of cryoEM
2013: Chemistry	Karplus, Levitt and Warshel	Molecular modeling
2012: Chemistry	Brian Kobilka	G-protein coupled receptors
2009: Chemistry	Venki Ramakrishnan, Thomas Steitz, Ada Yonath	Ribosome
2006: Chemistry	Roger D. Kornberg	RNA polymerase
2003: Chemistry	Roderick MacKinnon	potassium channel
2002: Chemistry	Kurt Wüthrich	development of protein NMR
1997: Chemistry	John E. Walker	FoF1-ATPase
1991: Chemistry	Richard R. Ernst	development of protein NMR
1988: Chemistry	Johann Deisenhofer, Robert Huber, Hartmut Michel	photosynthetic reaction centre
1982: Chemistry	Aaron Klug	development of cryo-EM
1972: Chemistry	Christian B. Anfinsen	ribonuclease
1964: Chemistry	Dorothy Hodgkin	penicillin
1962: Medicine	Francis Crick, James Watson, Maurice Wilkins	DNA double helix
1962: Chemistry	Max Perutz, John Kendrew	myoglobin and haemoglobin
1946: Chemistry	James Sumner	first crystals of an enzyme (urease)

Relative sizes of biological macromolecules

- The relevant size range is: ~0.1nm - 100nm
- Molecular weight range: 1Da - 100MDa





Some reminders:

- Molecular weight of biomolecules is usually expressed in Dalton (1 Da = 1 g/mol)
  - Average amino acid has a MW of ~110 Da
  - Average nucleotide pair has a MW of ~650 Da
- Length and resolutions are often expressed in Angstrom (Å) which is 10^{-10} m
  - Average atom diameters and bond lengths in biomolecules are ~1-2 Å
  - Basic building blocks (amino acids, nucleic acids) will be ~5-15 Å

Abbé's law on diffraction limited optical systems

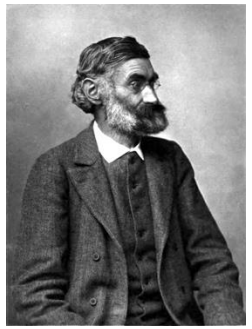
- The wavelength of electromagnetic radiation needs to be at a similar or lower order of magnitude as the features that are visualized

Resolution limit of a microscope

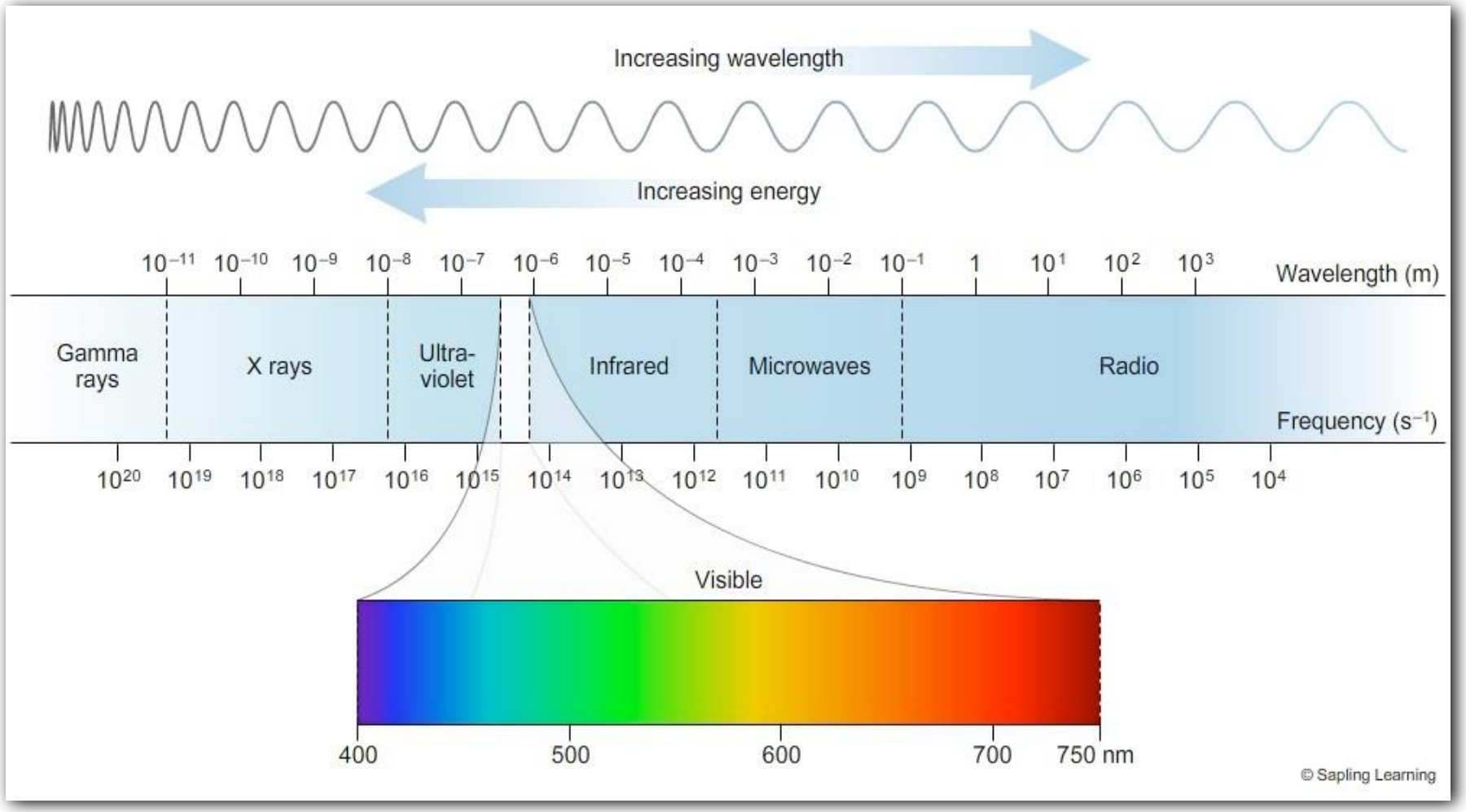
Light wavelength

$$d = \frac{\lambda}{2n\sin\theta} = \frac{\lambda}{2\mathrm{NA}}$$

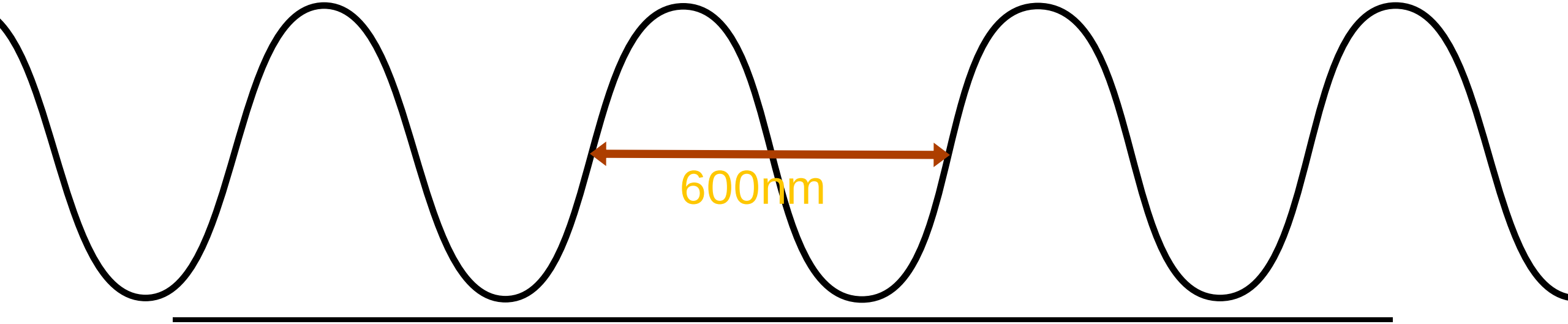
Numerical Aperture



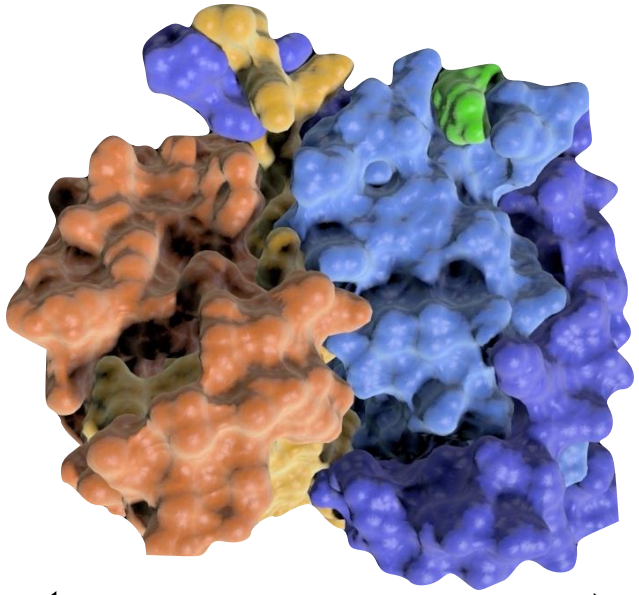
Ernst Abbe
(1840-1905)



Visible light is too “coarse” to study biomolecular structures



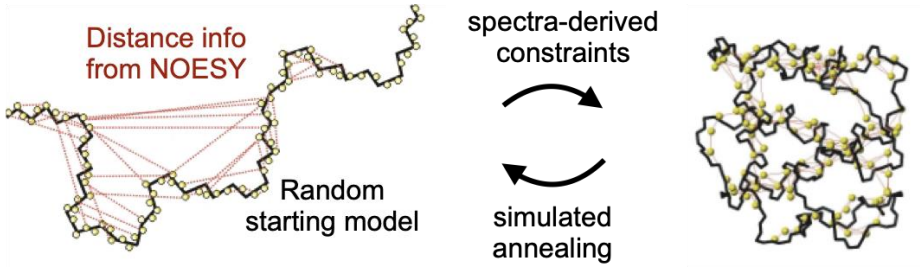
Wavelength of visible light is 400-750nm



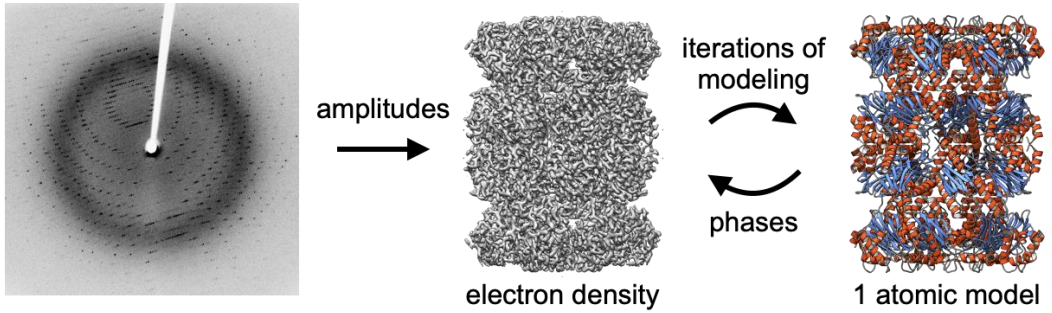
The main structural biology methods

- 3 main methods for experimental structure determination:

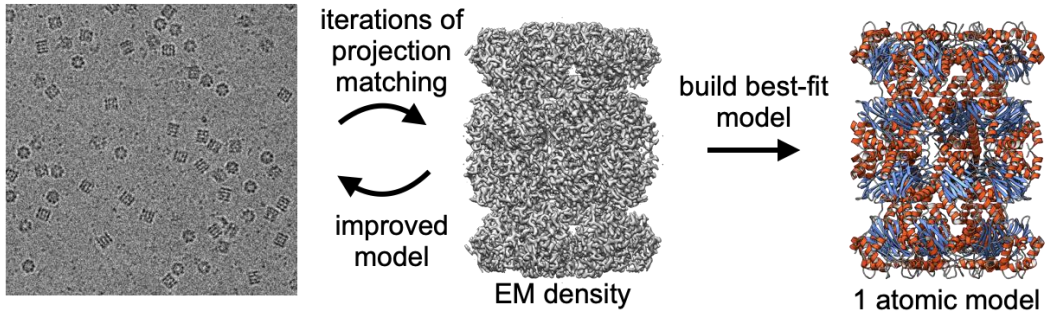
Nuclear Magnetic Resonance



X-ray Crystallography



Cryo-electron Microscopy

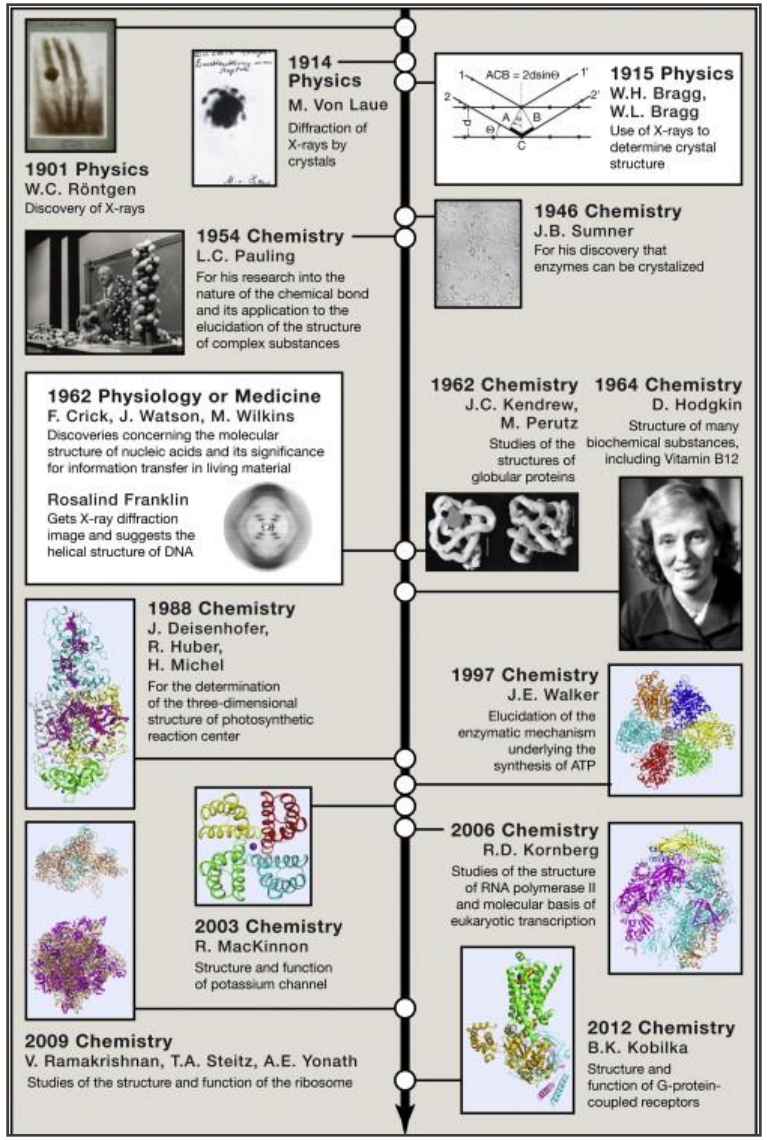


Molecular Type	X-ray	EM	NMR
Protein (only)	166,790	15,369	12,516
Protein/Oligosaccharide	9,624	2,600	34
Protein/NA	8,710	4,654	286
Nucleic acid (only)	2,867	137	1,507
Other	170	10	33
Oligosaccharide (only)	11	0	6
Total	188,172	22,770	14,382

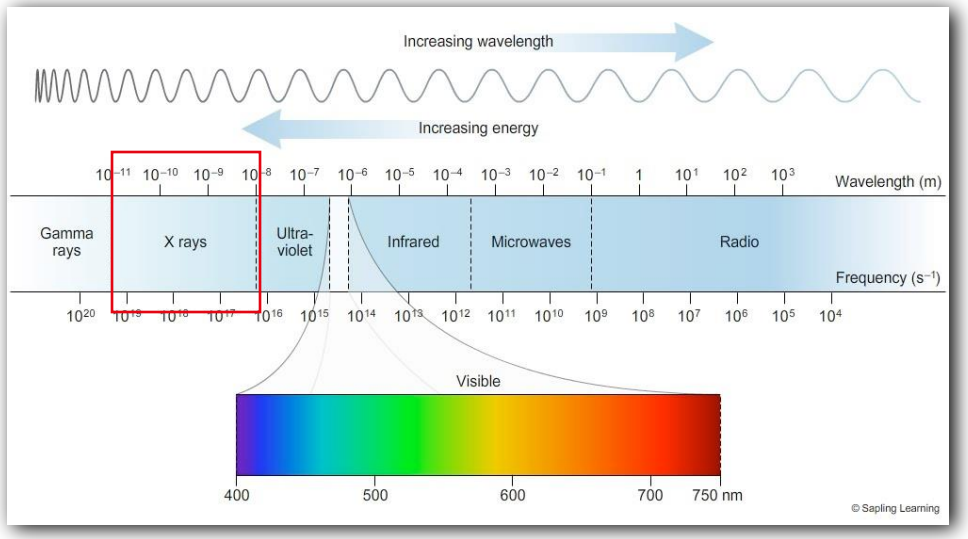
- Most structures came from X-ray crystallography since it is historically the oldest method

- The outputs of computational (AI) prediction of biomolecular structures are usually not referred to as “structures” but rather as “models”
- This is because they have not been experimentally verified

X-ray crystallography



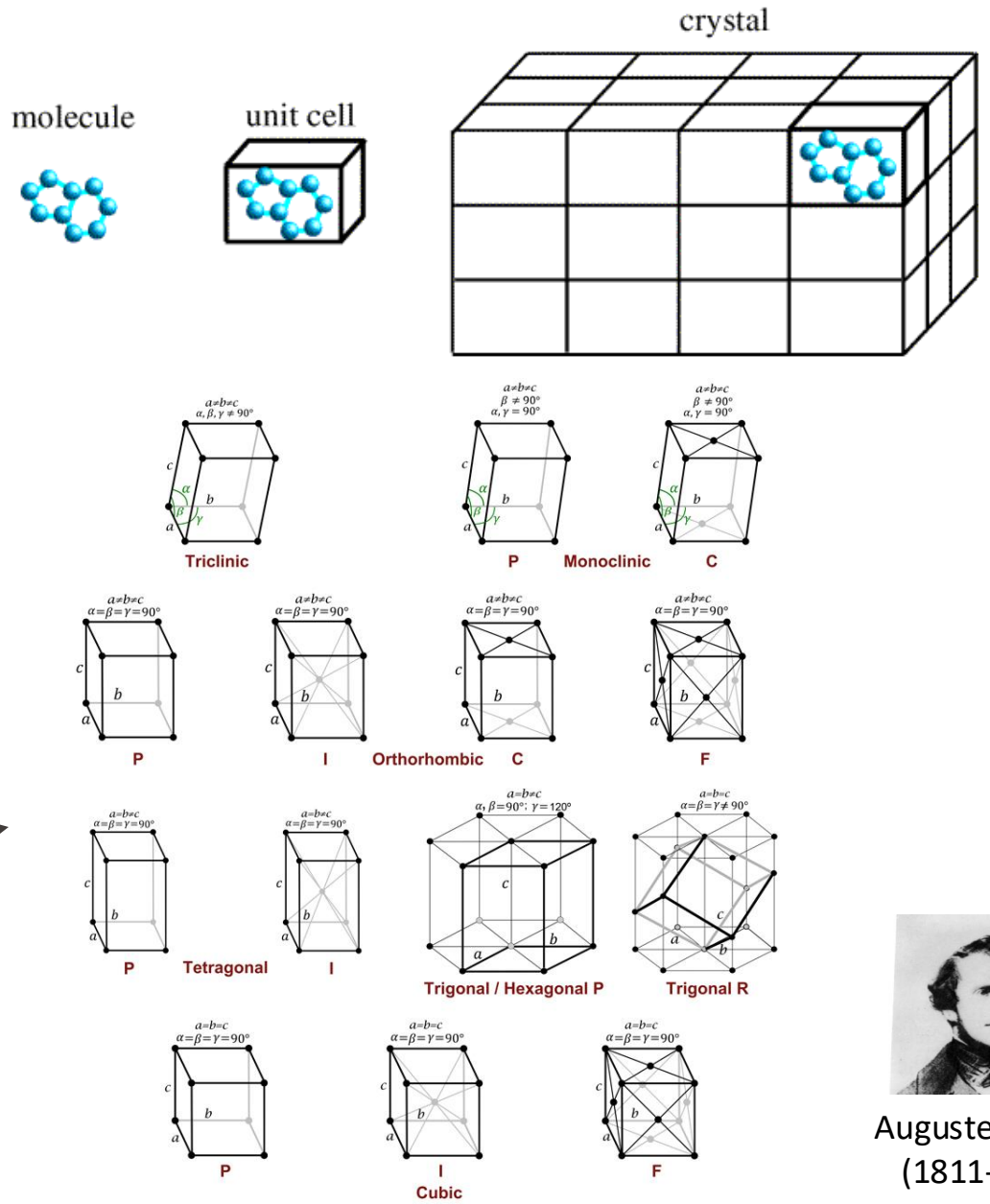
- X-ray diffraction applied to study the structural properties of crystalline materials in atomic detail



- First applied to determine the structures of inorganic molecules (e.g. salts) or organic molecules
- Relatively easy for small molecules but the proteins are much more challenging. Compare:
  - C_6H_5OH (Phenol)
  - $C_{2717}H_{4155}N_{735}O_{810}S_{19}$ (Protein with ~500 amino acids)

What are crystals?

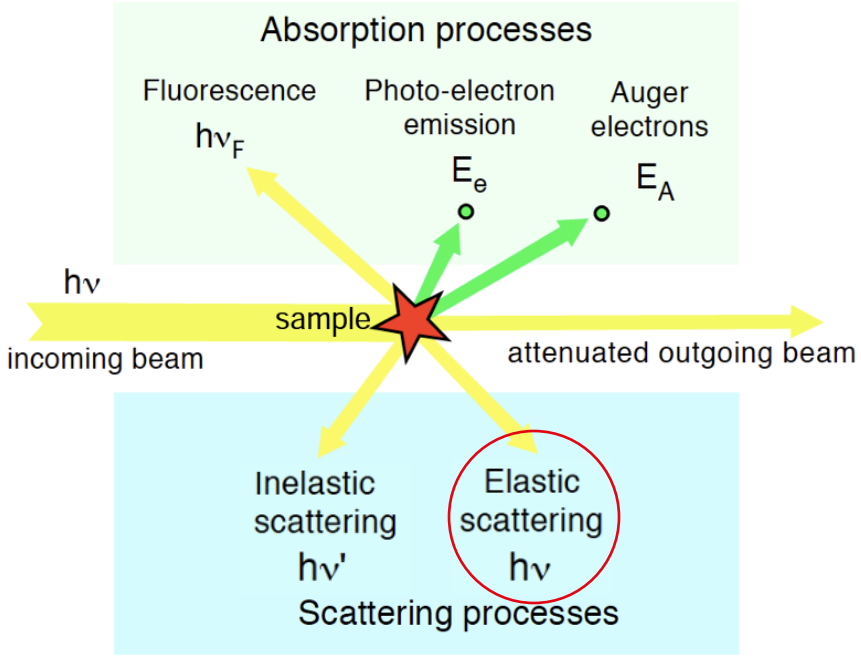
- A form of solid state where molecules are packed together in an ordered lattice held together by non-covalent bonds
- **Unit cell** is the smallest repeating structural unit of dimensions (a,b,c) and angles (α, β, γ)
- **Bravais lattice** defines how atoms can assemble into unit cells compatible with highly regular crystal pattern
- **14 different Bravais lattice systems**
- **230 space groups** = lattice + rotational and translational symmetries

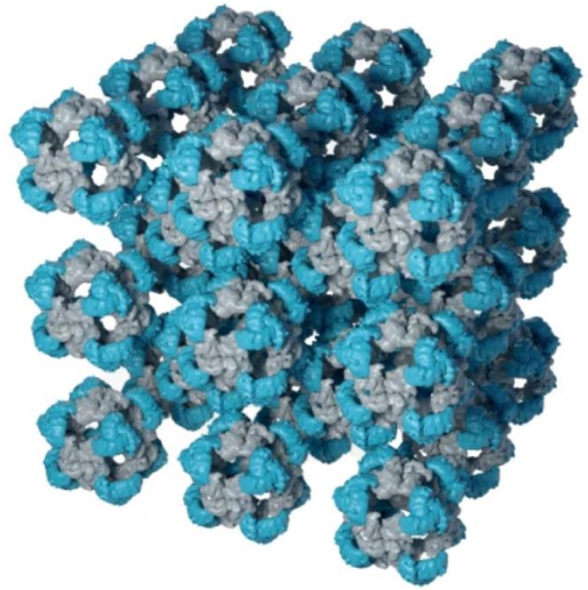


Auguste Bravais
(1811-1863)

Why do we need crystals?

- X-rays can interact with matter in different ways including **elastic scattering**
- X-rays are **scattered** by the electrons in atoms comprising the sample molecule



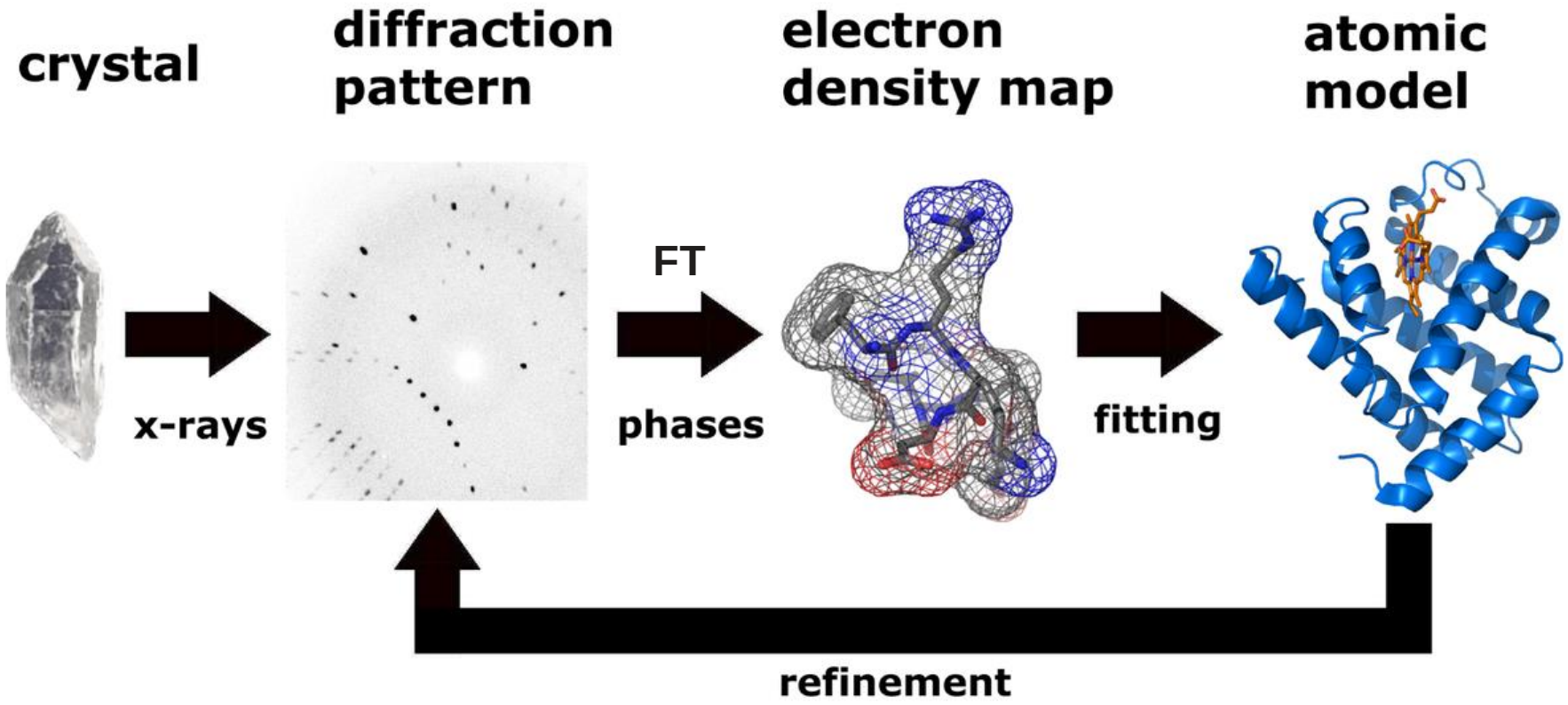


Protein crystal

- X-ray scattering from a single atom/molecule is too low to generate contrast while imaging.
- Protein crystals (0.2 mm cube) contain ~10^{15} molecules. They scatter in phase and **amplify the signal**.

Structure-solving using X-ray crystallography

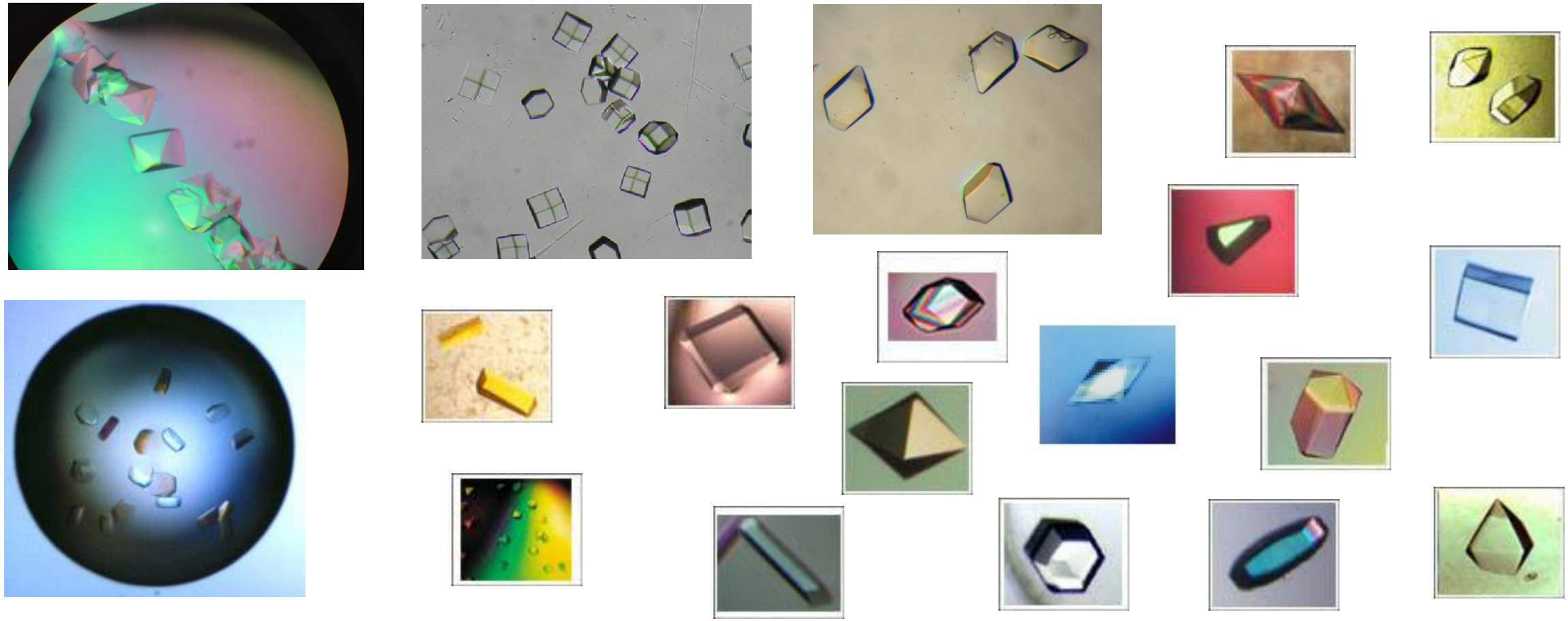
- Protein crystals are placed in the X-ray beam which produces a diffraction pattern (diffractogram)



- Fourier transform of the diffractogram produces real-space map that can be used for model building

Protein crystals - Difficult to make but very beautiful

- Proteins do not naturally crystallize, and most of the work in the process is spent on their production

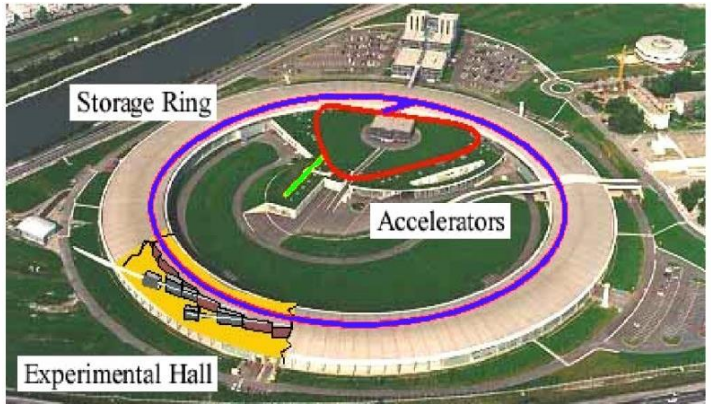


- Proteins are crystallized by **precipitation with high amounts of salts, organic solvents or polymers**
- Typical sizes needed for structural analysis are **0.1 - 0.3 mm**

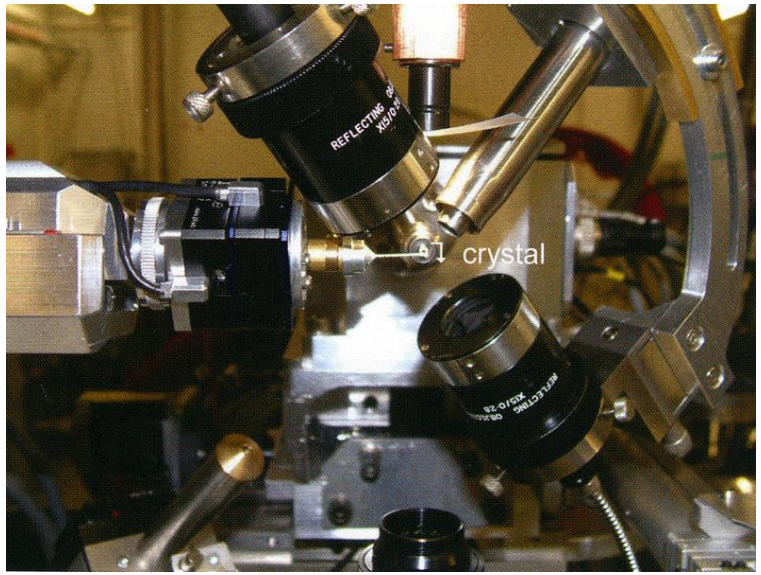
Sources of X-rays: Synchrotrons

- Synchrotrons are cyclic particle accelerators in which the magnetic field bends the particles (e.g., accelerated electrons) into a closed path which results in production of X-rays.
- The radiation is projected at a tangent to the electron storage ring creating a very intense, coherent beam at a single frequency

SLS in Villigen



ESRF Grenoble

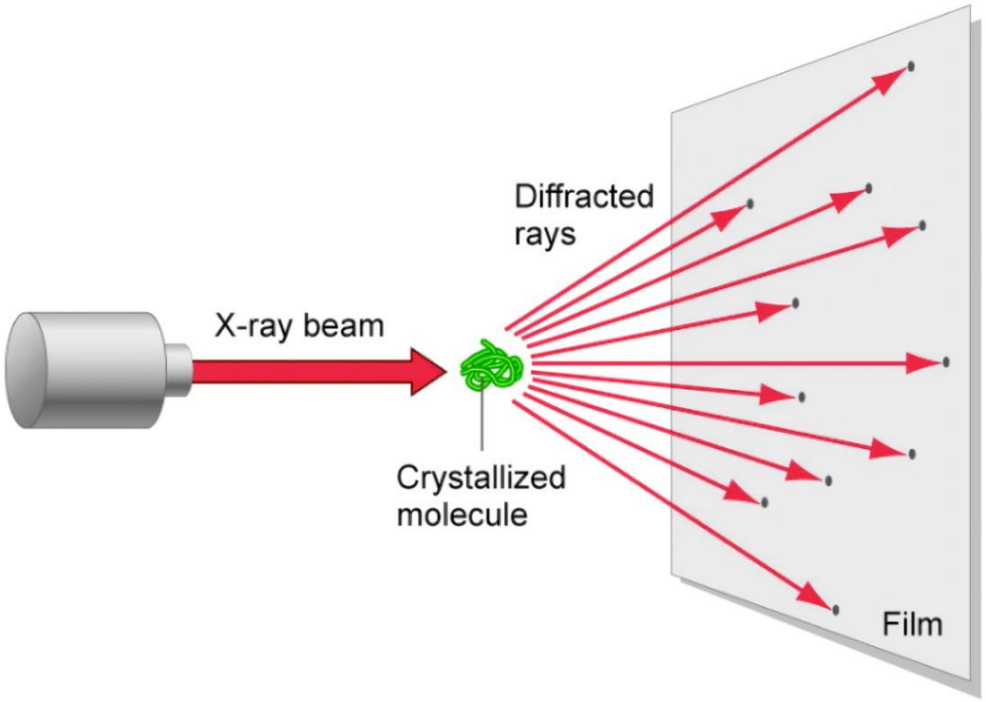


Crystal is mounted onto a goniometer-controlled stage and shot with X-rays while it rotates

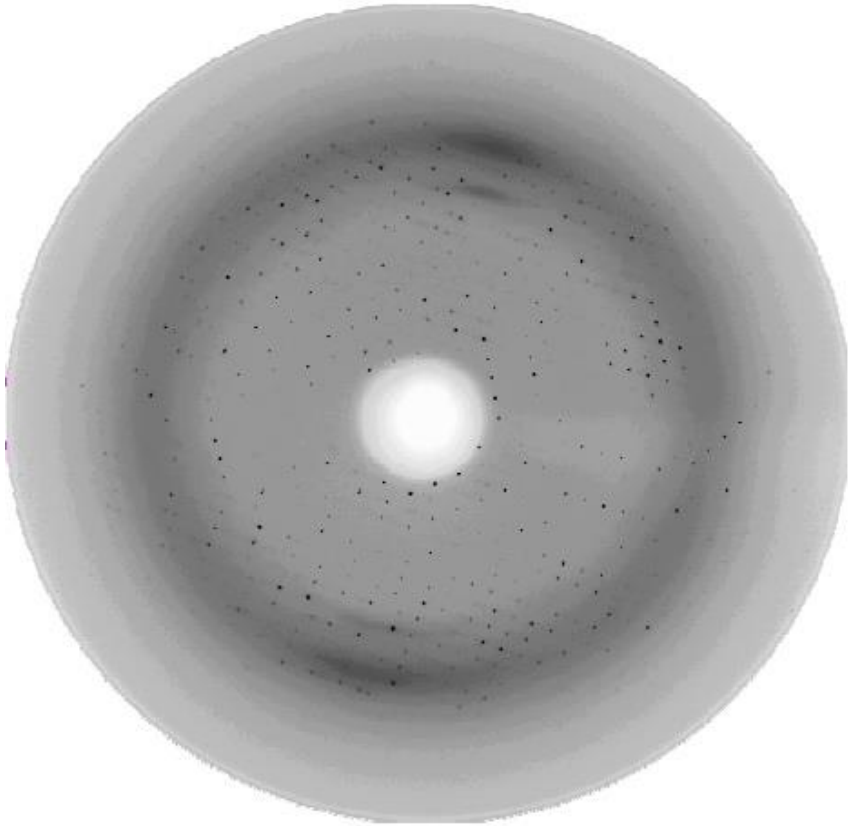
- Note: There are also smaller X-ray anode-based sources housed in research labs that are very much in use today

Diffraction pattern

- Diffraction pattern is the raw output of X-ray scattering experiments from crystalline samples

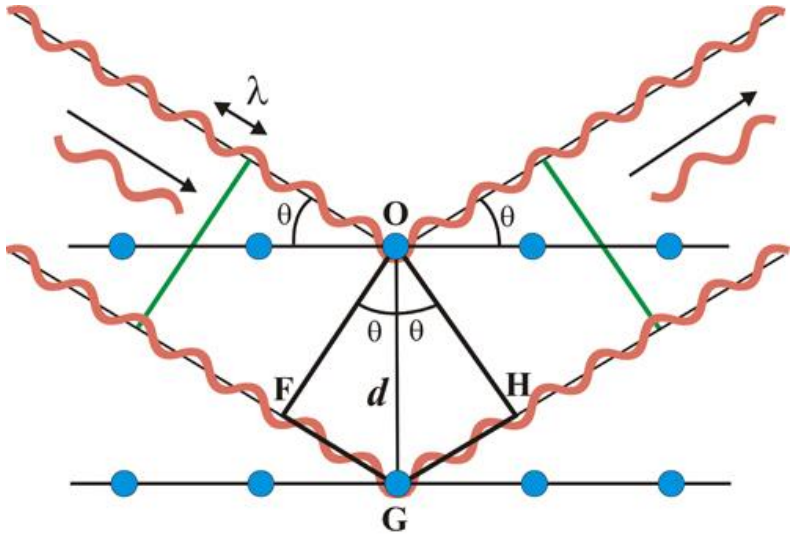


Where are the spots coming from?



Diffraction pattern origin

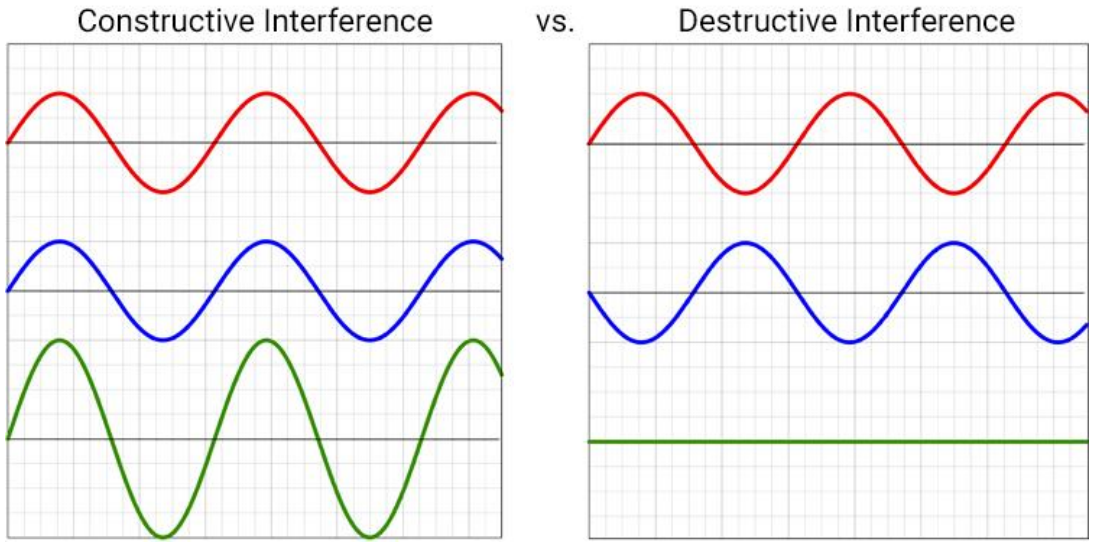
- Diffraction pattern is generated as a result of (constructive and destructive) **interference of X-rays elastically scattered from electrons** in the crystallized biomolecule
- Bragg's law connects the scattering angles and intensities of spots to the information on spacing in the crystal, i.e. on the position of atoms in a protein molecule



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

- λ is the wavelength of incident X-rays
- d is the spacing between the planes in the atomic lattice
- θ is the angle between the incident ray and the scattering planes

- Depending on the phase offset between the two X-rays the waves will interfere differently

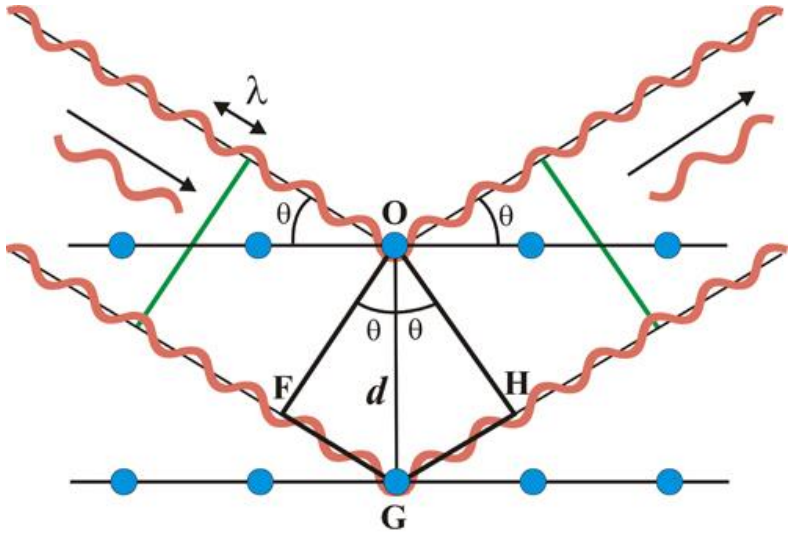


Intensity ~ (Amplitude)^2

Resolution can be estimated from diffraction patterns

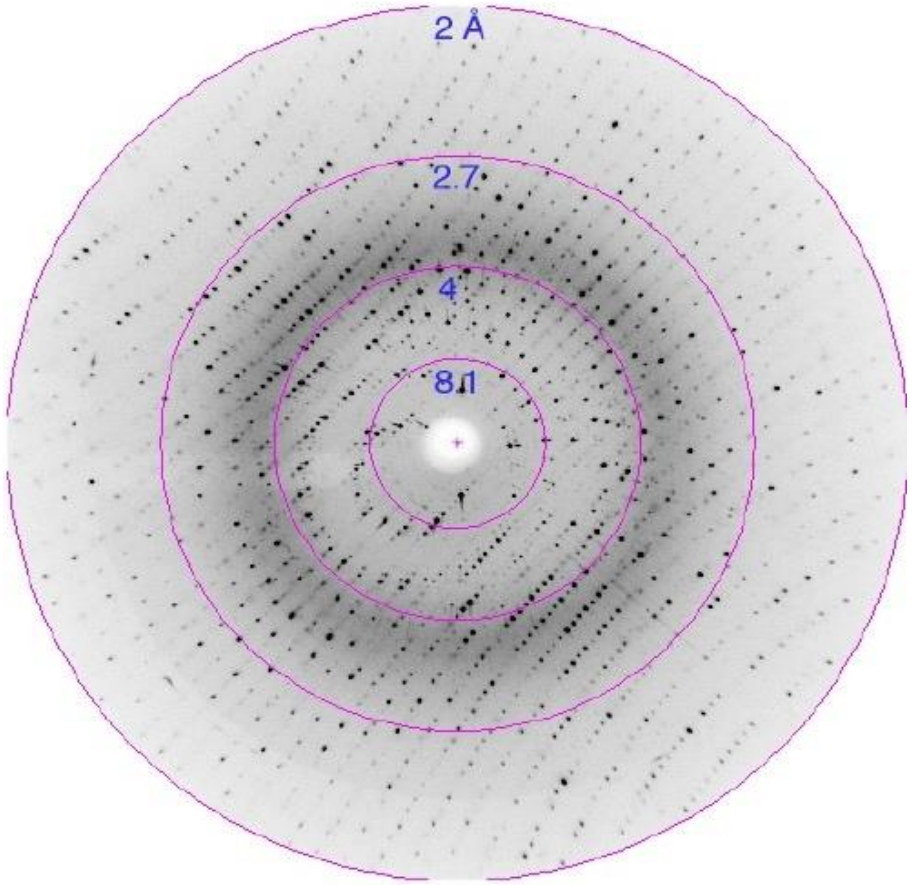
- The further out spots (wider **θ** angles) are visible the finer the distance between diffracting atoms
- This **minimum measurable distance translates into the resolution** of reconstructed 3D map

Bragg's law



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

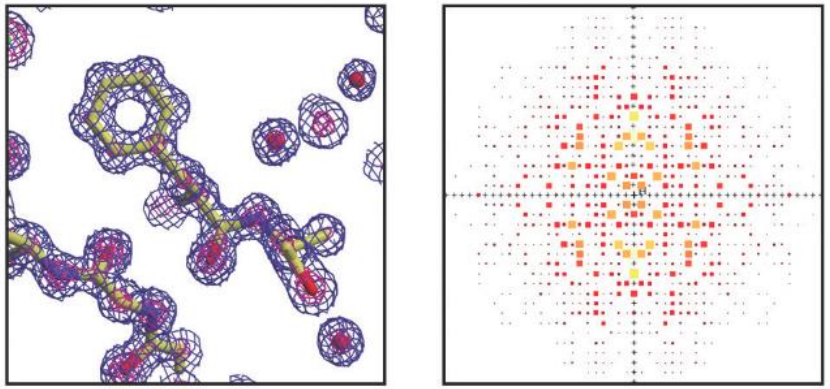
$d \sim 1/\sin\theta$



Reconstructing electron density map

- The diffraction pattern is related to the scattering object by a Fourier transform:

$$\rho(x,y,z) = \frac{1}{V}\sum_{-h}^{h}\sum_{-k}^{k}\sum_{-l}^{l} F_{hkl} \cdot \exp[-2\pi i(hx+ky+lz-\alpha_{hkl})]$$



The crystallographic phase problem

F_{hkl} - Structure Factor at coordinates h,k,l
h,k,l - Miller indices (dimensions in reciprocal space)
ρ_{xyz} - Electron density at position xyz
x,y,z - Coordinates in real space
α_{hkl} - Phase angle of X-rays hitting position h,k,l

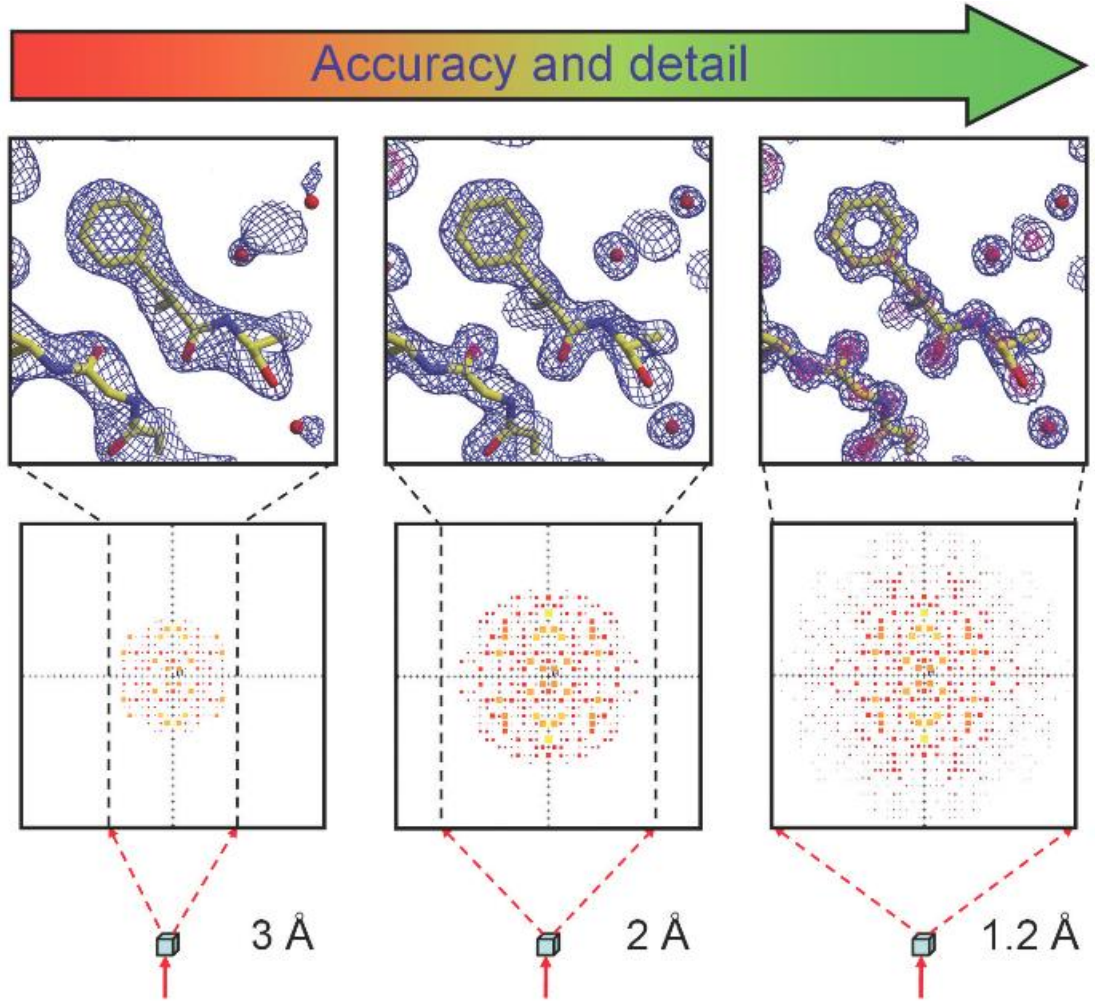
- Detectors measure only intensity (amplitudes) of X-rays but not the phase of the resulting X-rays

- Phase problem can be solved using heavy metals during crystallization or a known starting model of the protein during reconstruction.

- Structure factors F(h,k,l) are calculated from the diffraction pattern and “inverted” to reveal the 3D structure of a scattering object in real space (x,y,z)

Interpreting the map and reconstructing the model

- Resulting electron density maps are used for building atomic models of the biomolecules
- The higher the resolution (lower value in Å) the greater the detail in EM maps



In case of proteins and nucleic acids:

10 - 20 Å	Molecule shape
5 - 10 Å	Secondary structure elements
4 - 5 Å	Main-chain trace (+ bulky side-chains)
1.5 - 4 Å	Side-chain densities
<1.5 Å	Atoms

The interpretation of data is confined by the final map resolution (i.e., you cannot build atoms in a 20Å map)

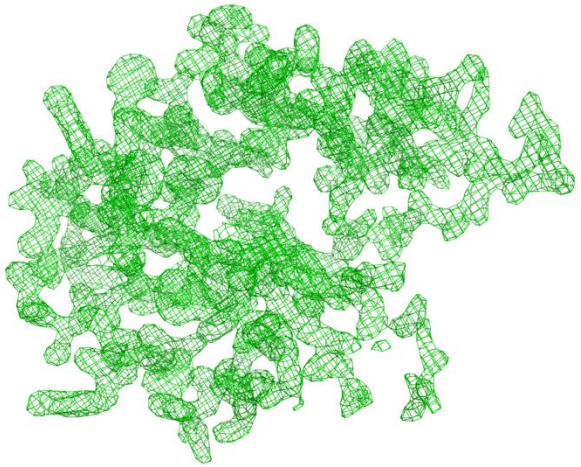
Reconstructing the atomic model

- The structure is built using:
  - Electron density map restraints
  - Known sequence of the model (e.g., amino-acid sequence)
  - Geometric restraints (e.g., Cis-Trans peptide bonds, allowed Ramachandran angles)

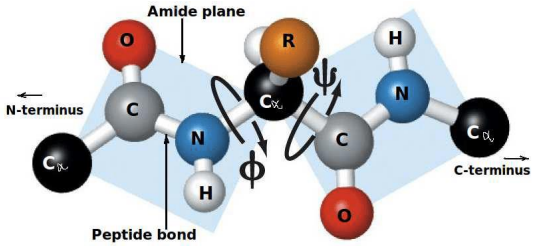
Sequence:

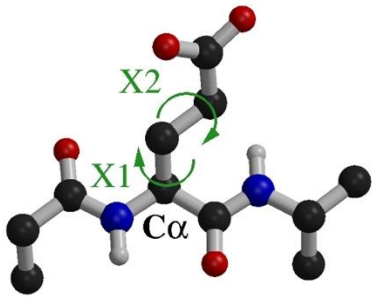
PVNSLEKHSWYHGPVSRNAAEYLLS
SGINGSFLVRESESSPGQRSISLRYE
GRVYHYRINTASDGKLYVSSESRFNT
LAELVHHHSTVADGLITTLHYPAPKRN

Electron density map

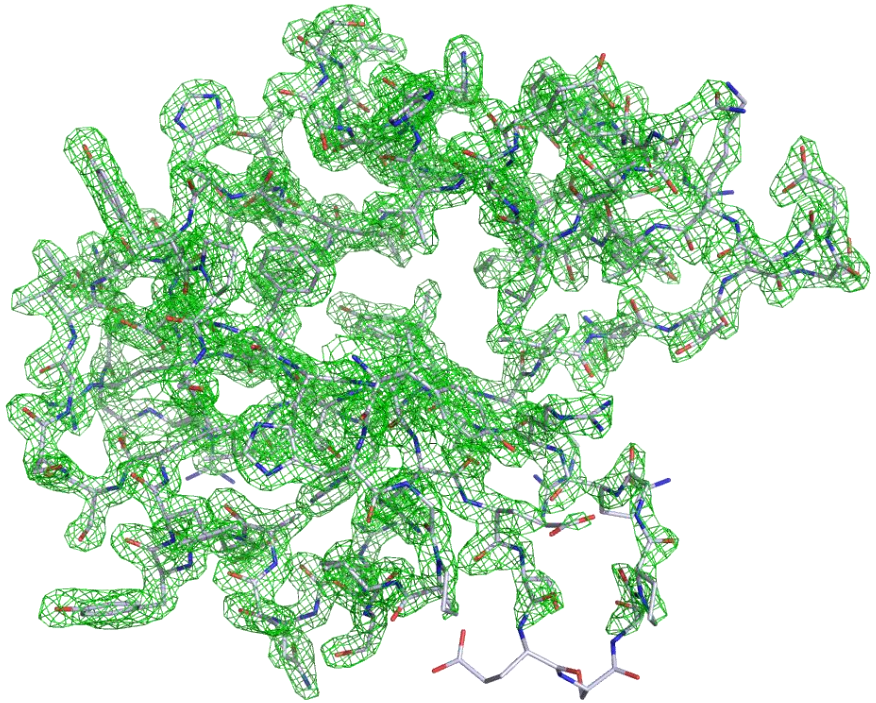


Geometric restraints:



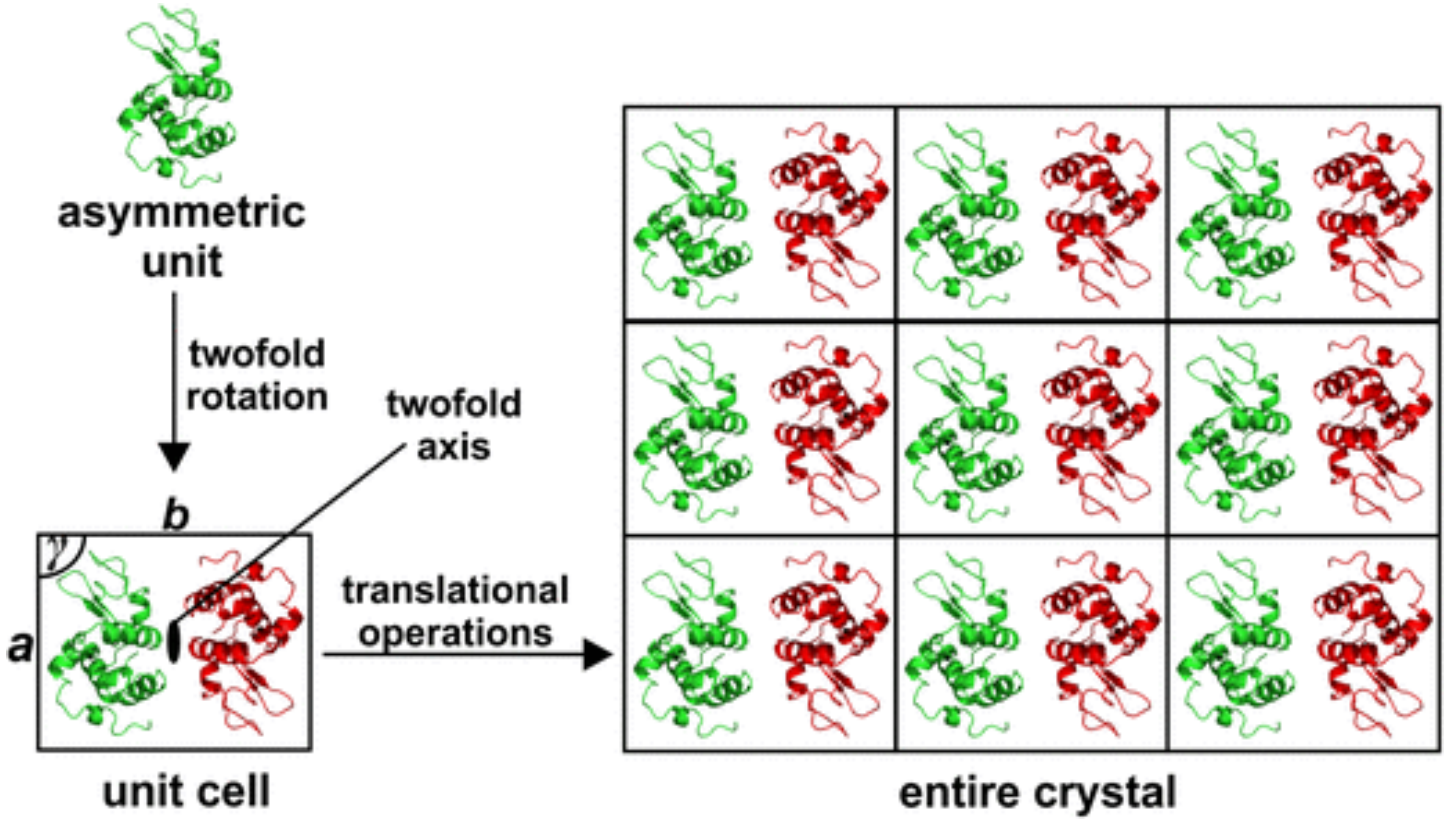


Atomic model



Reconstructing the atomic model

- The structure needs to be built only for the smallest **asymmetric unit** of the crystal
- The full crystal structure (if needed) can be generated from the asymmetric unit using the crystallographic symmetry

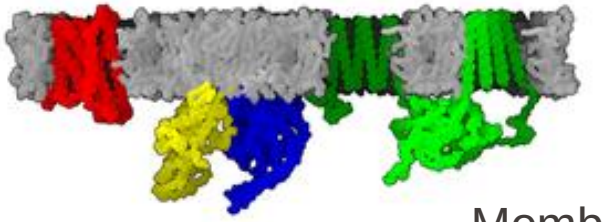


X-ray crystallography - Pros and Cons

- **Positive sides**
- Versatile and highly applicable
- Historically, it is the most fruitful method for structure determination
- No limitations with respect to molecular weights of biomolecules that are studied
- No limitations in terms of achievable resolutions

Molecular Type	X-ray	EM	NMR
Protein (only)	166,790	15,369	12,516
Protein/Oligosaccharide	9,624	2,600	34
Protein/NA	8,710	4,654	286
Nucleic acid (only)	2,867	137	1,507
Other	170	10	33
Oligosaccharide (only)	11	0	6
Total	188,172	22,770	14,382

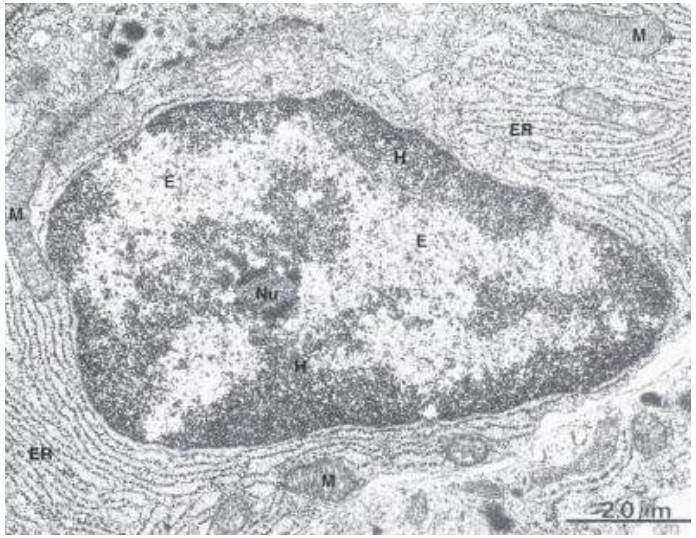
- **Negative sides**
- Needs highly purified and homogenous biomolecule material
- Biomolecule crystallization is a very challenging and unpredictable process
- Cannot be used to study heterogeneous samples (e.g., flexible complexes, pleomorphic viruses)
- High material requirements (many milligrams of purified biomolecule)
- It is a static picture of a molecule and cannot capture dynamic information
- Difficult to solve membrane proteins due to heterogeneous membrane environment



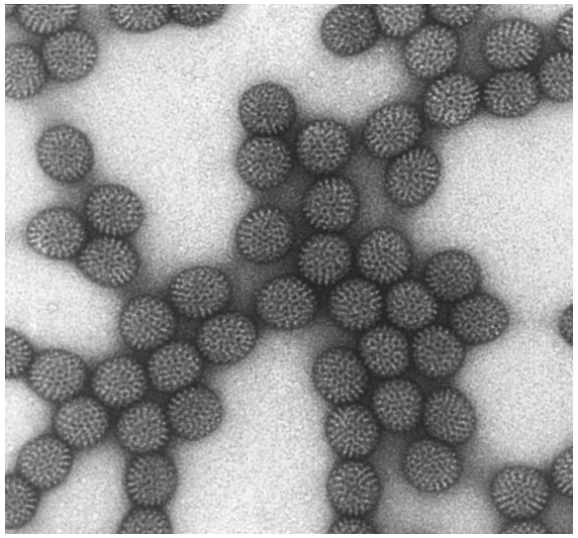
Membrane proteins

Electron Microscopy

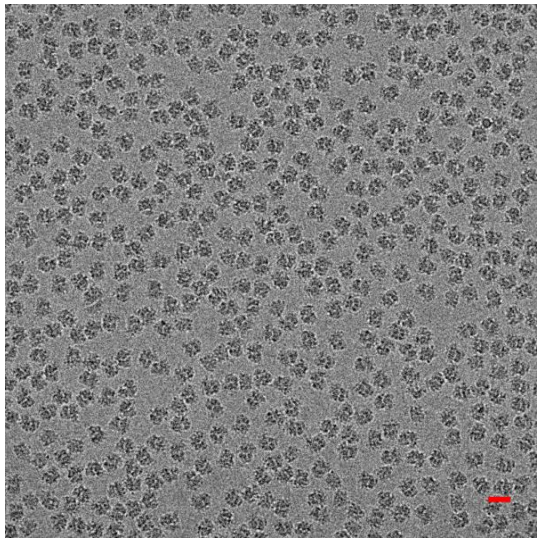
Nucleus of a cell



Rotavirus



Ribosomes

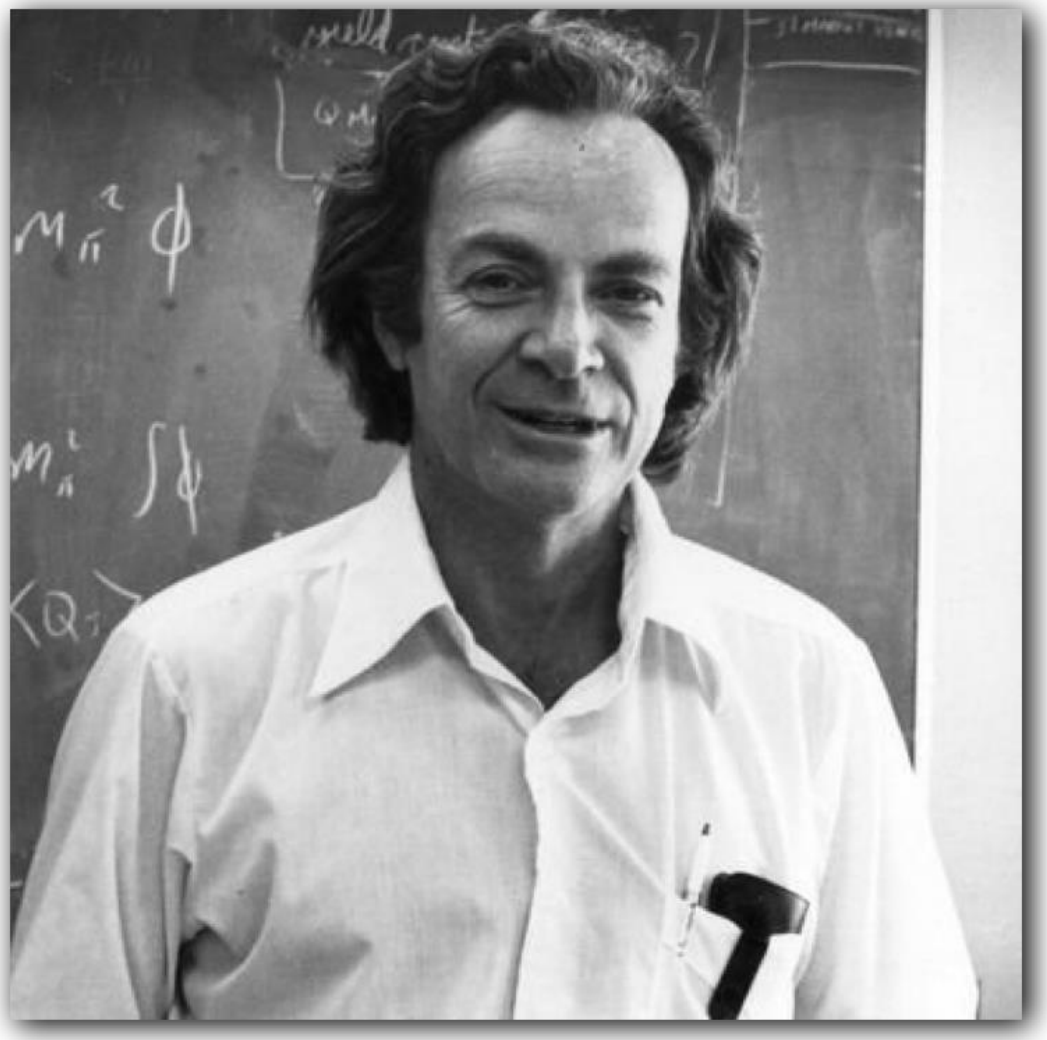


… same technique, but at different magnification, as protein are much smaller than cells!

How to visualize biomolecules?

"It is very easy to answer many of these fundamental biological questions;
you just look at the thing!... Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier.
I exaggerate, of course, but the biologists would surely be very thankful
to you ….."

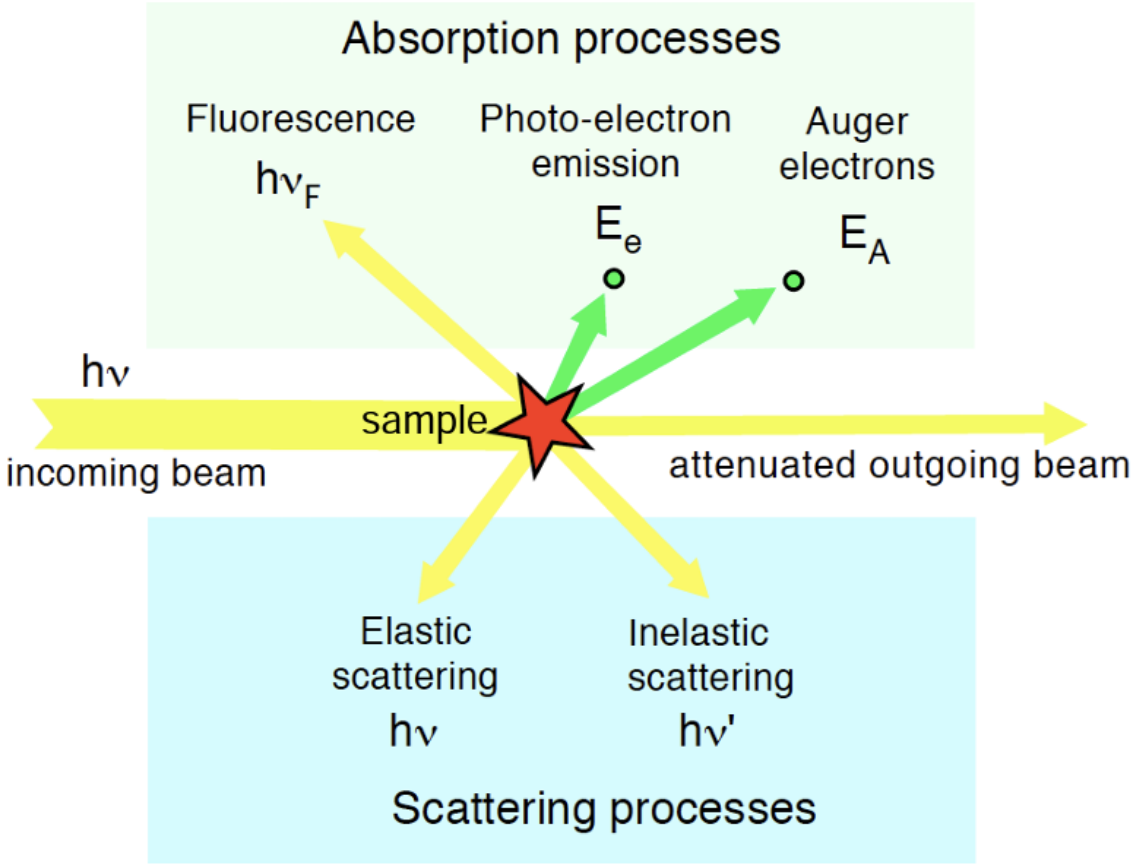
Richard P. Feynman



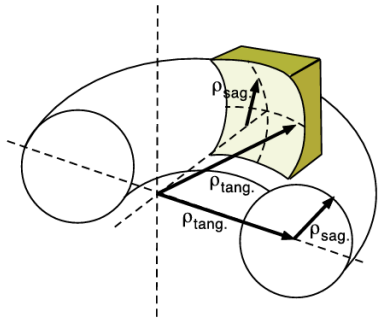
"There's plenty of room at the bottom"
lecture given at the APS in 1959

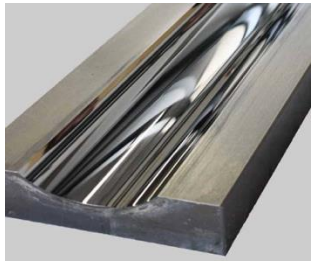
X-rays for microscopy applications

- X-rays can interact with matter in different ways but are relatively poorly scattered

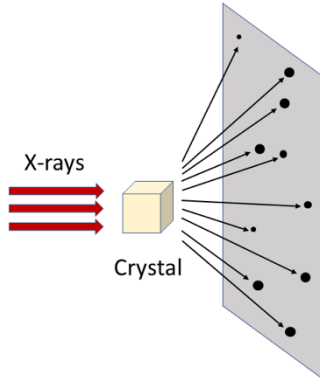


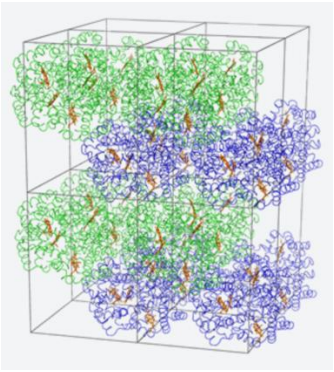
- Difficult to produce good lenses/mirrors





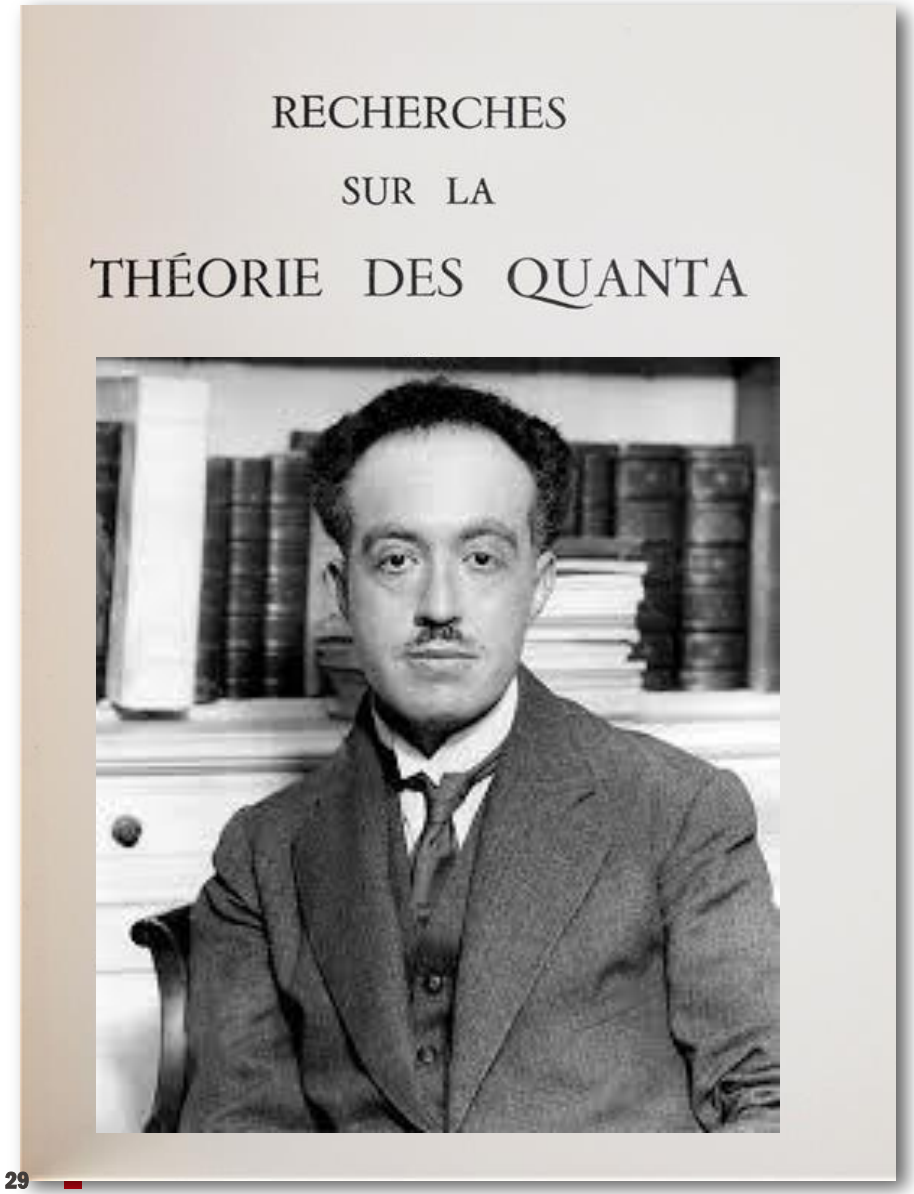
- Poor image contrast in a microscope





- Protein crystals help amplify the scattering signal

Accelerated electrons as an alternative for X-rays



1924 thesis of Louis de Broglie

Combined the equations of special relativity & quantum theory to suggest that electrons could be thought of as waves as well as particles.

et la fréquence ν des ondes s'exprime par :

$$\nu = \frac{1}{T} = \frac{\nu_0}{\sqrt{1-\beta^2}} = \frac{m_0 c^2}{h\sqrt{1-\beta^2}}$$

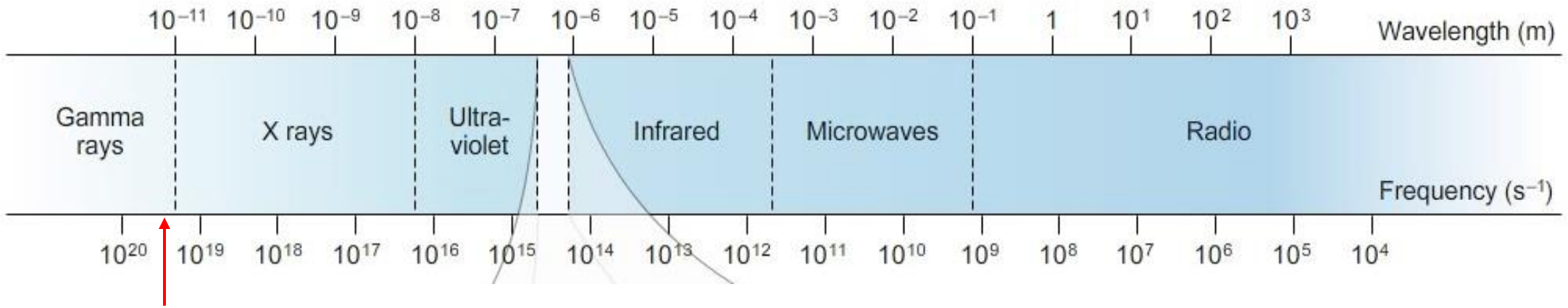
$$\lambda = \frac{h}{\gamma m_0 v} = \frac{h}{m_0 v}\sqrt{1-\frac{v^2}{c^2}}$$

$$f = \frac{\gamma m_0 c^2}{h} = \frac{m_0 c^2}{h} \Big/ \sqrt{1-\frac{v^2}{c^2}}$$

$$\lambda = \frac{h}{p}$$

Accelerated electrons as an alternative for X-rays

- Electrons with energy of **~100keV** have a wavelength of **3.88pm**



- Typical voltage in a microscope is 100-300kV

- 100X shorter **λ** than typical X-rays in a synchrotron

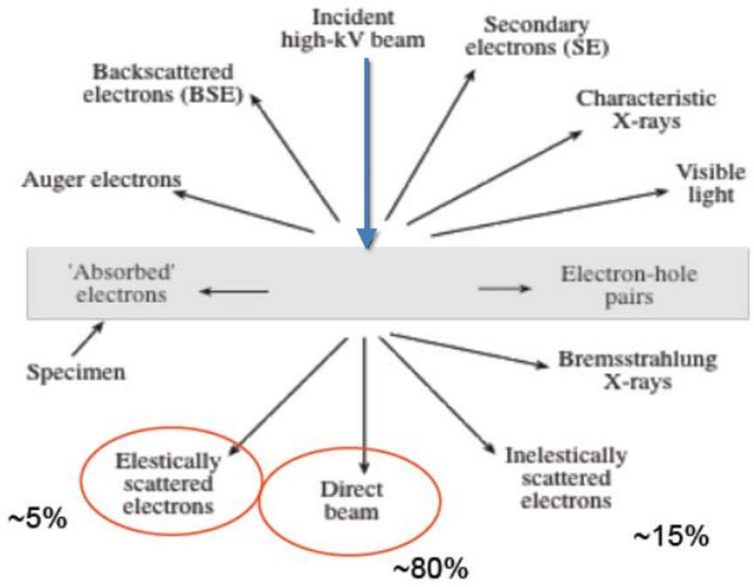
$$\lambda = \frac{6.626\times10^{-34}}{\sqrt{2V\times1.6\times10^{-19}\times9.11\times10^{-31}}}$$

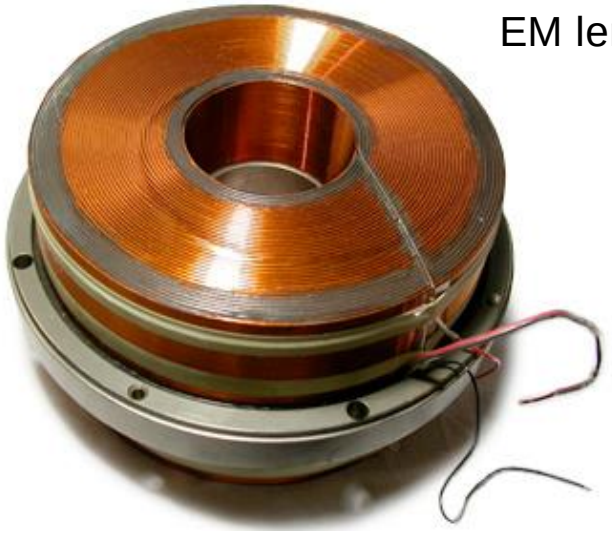
$$\Rightarrow \lambda = \frac{12.27\times10^{-10}}{\sqrt{V}}\text{meter}$$

$$\text{(or)}\ \lambda = \frac{12.27}{\sqrt{V}}\text{Å}$$

What are the advantages of electrons?

- The main benefit comes from the negative charge
- Compared to X-rays, elastic scattering of electrons is ~10^3-10^6 times stronger
- Electrons are easily accelerated using potential difference
- Electron beam can be readily collimated with electromagnetic lenses
- Readily detected by films and cameras

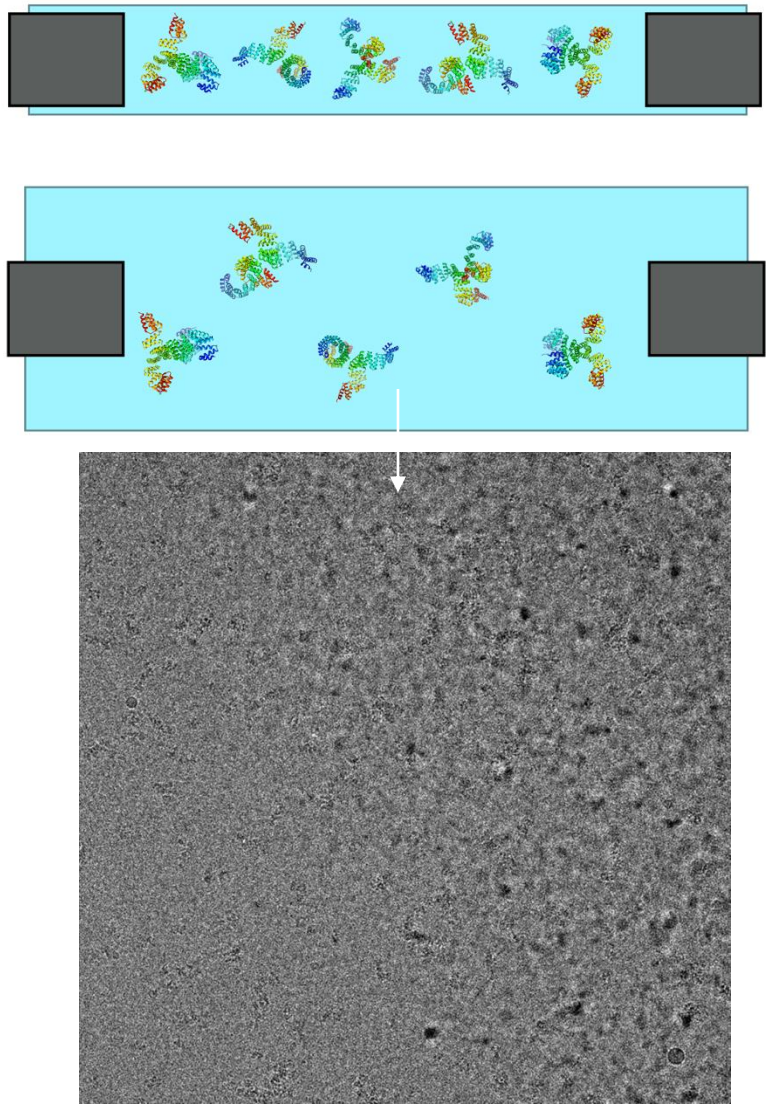




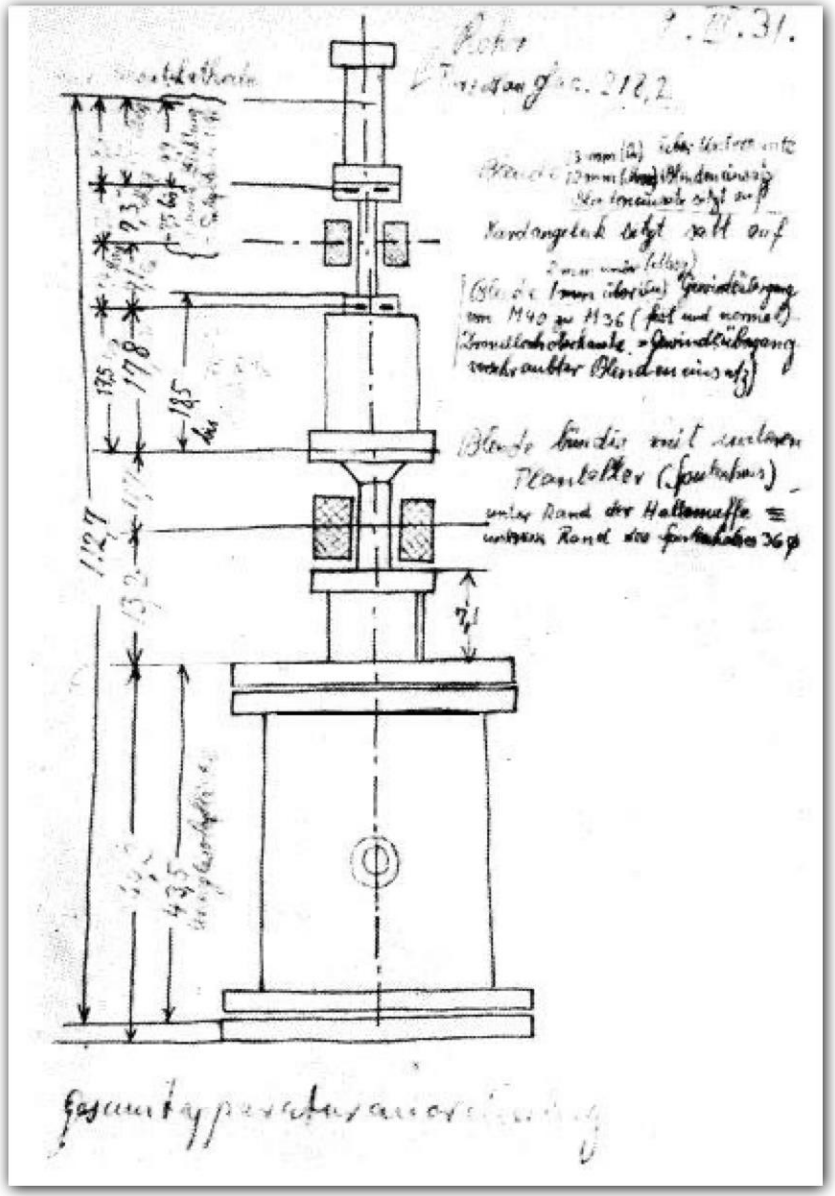
EM lens

What are the disadvantages of electrons?

- Need high vacuum within the microscope to reduce interactions with air molecules
- Samples need to be applied as a very thin layer (~10-100nm) to achieve sufficient contrast against surrounding ice
- Radiation damage in the sample caused by inelastic scattering of electrons (i.e., collisions with biological material leading to energy transfer)
- Insufficient contrast when imaging small proteins (lower molecular weight limit for EM is ~50kDa)



The first Transmission Electron Microscope



The first Transmission Electron Microscope

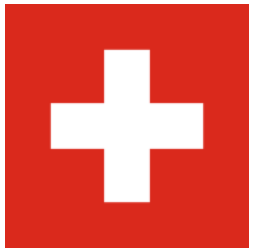
The Nobel Prize in Physics 1986

Ernst Ruska
Prize share: 1/2

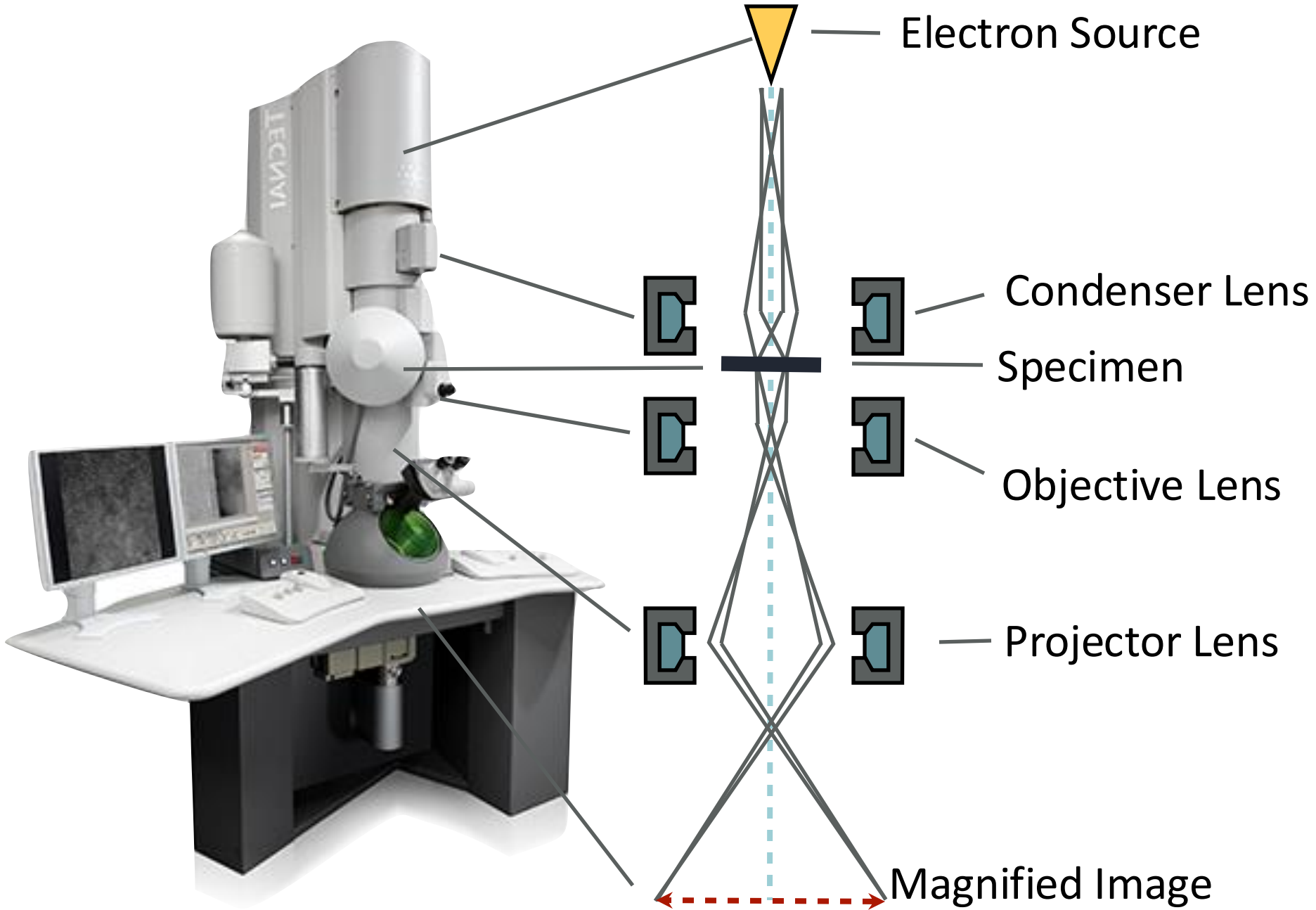
Gerd Binnig
Prize share: 1/4

Heinrich Rohrer
Prize share: 1/4

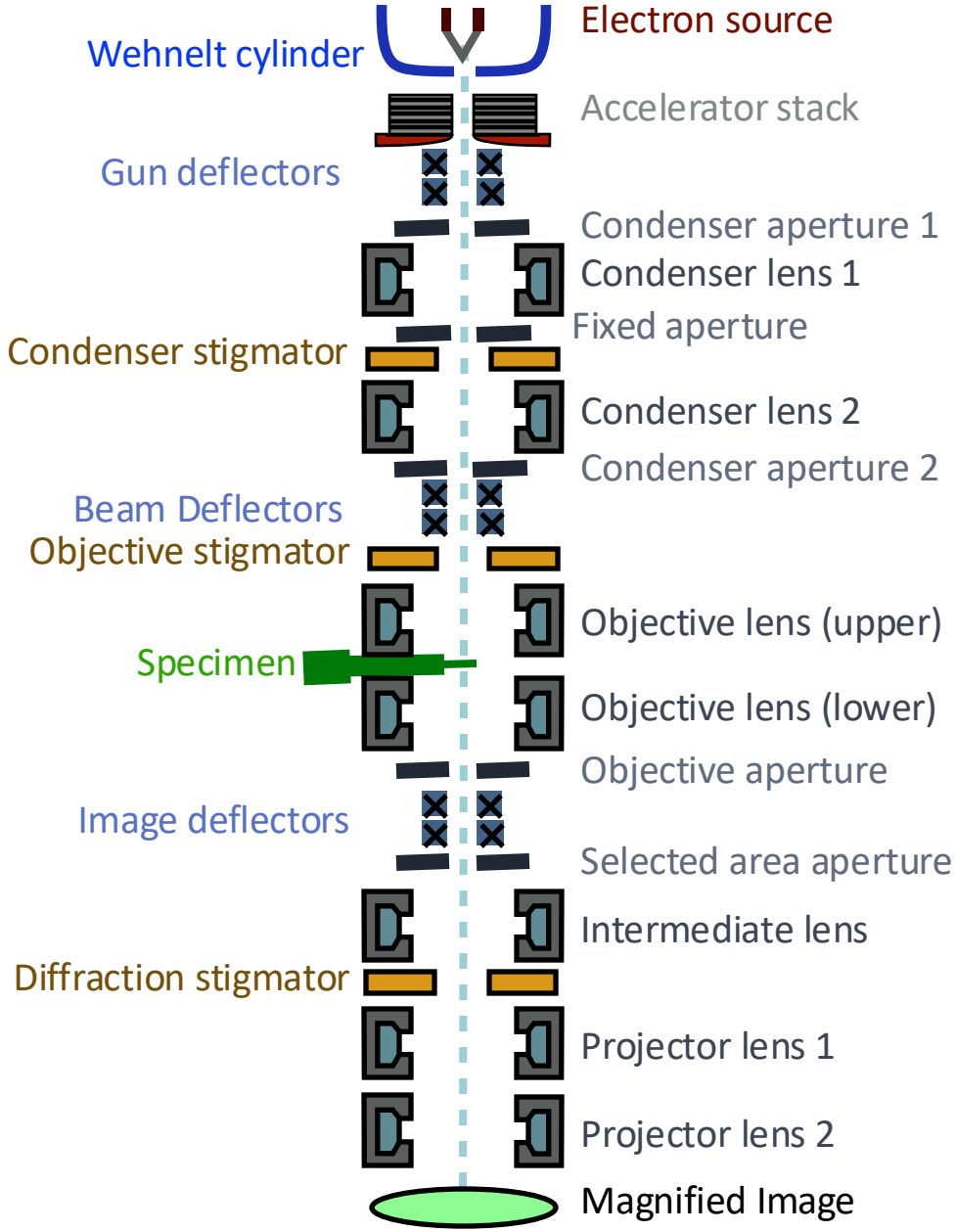
The Nobel Prize in Physics 1986 was divided, one half awarded to Ernst Ruska *"for his fundamental work in electron optics, and for the design of the first electron microscope"*, the other half jointly to Gerd Binnig and Heinrich Rohrer *"for their design of the scanning tunneling microscope"*.



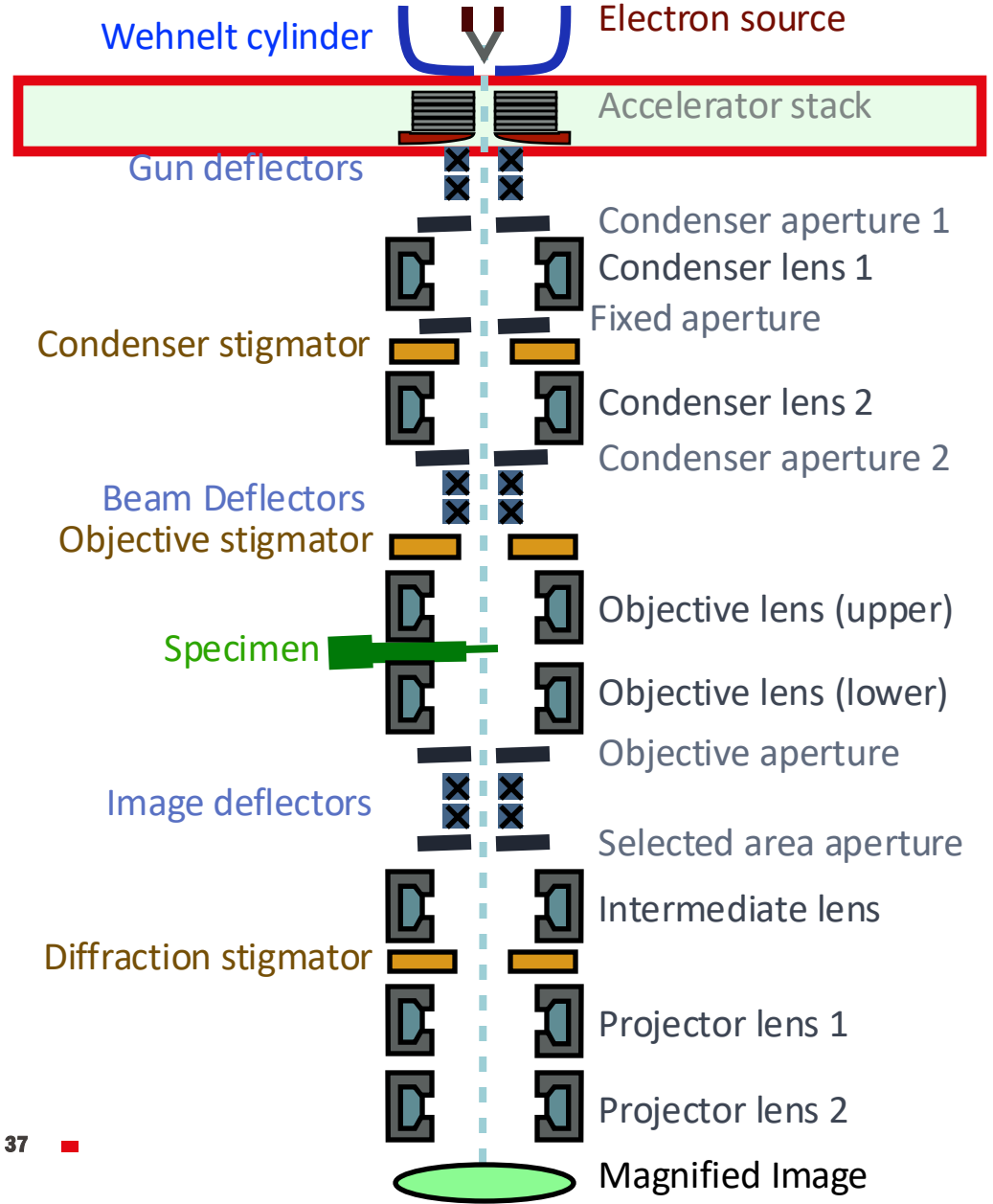
Microscope principle and components



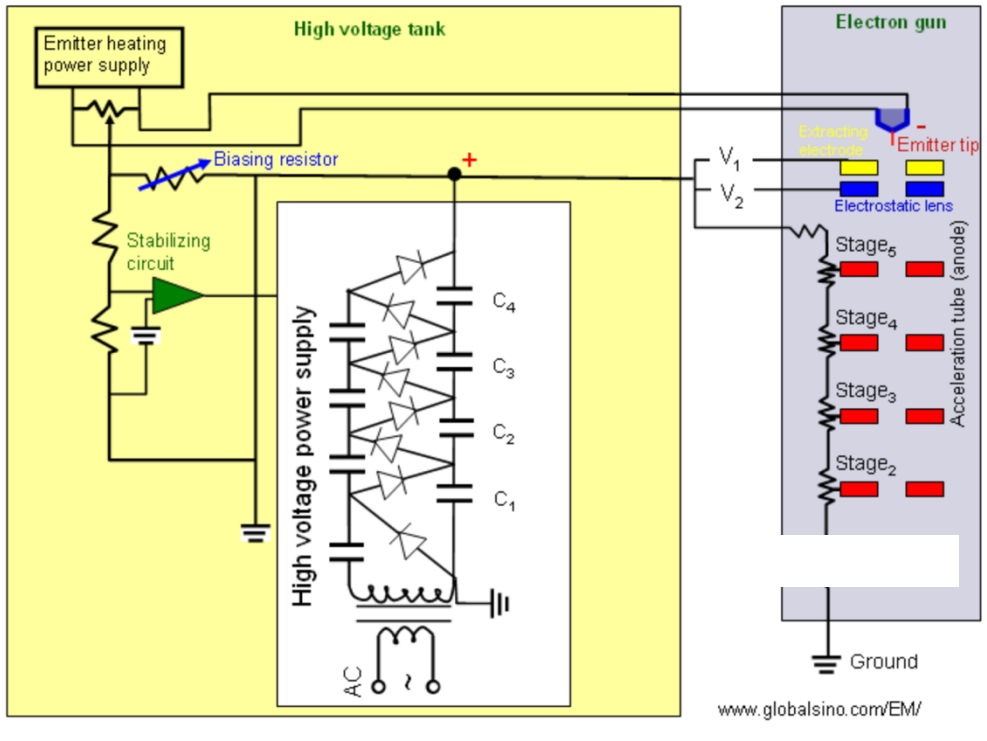
Microscope components



Accelerating electrons

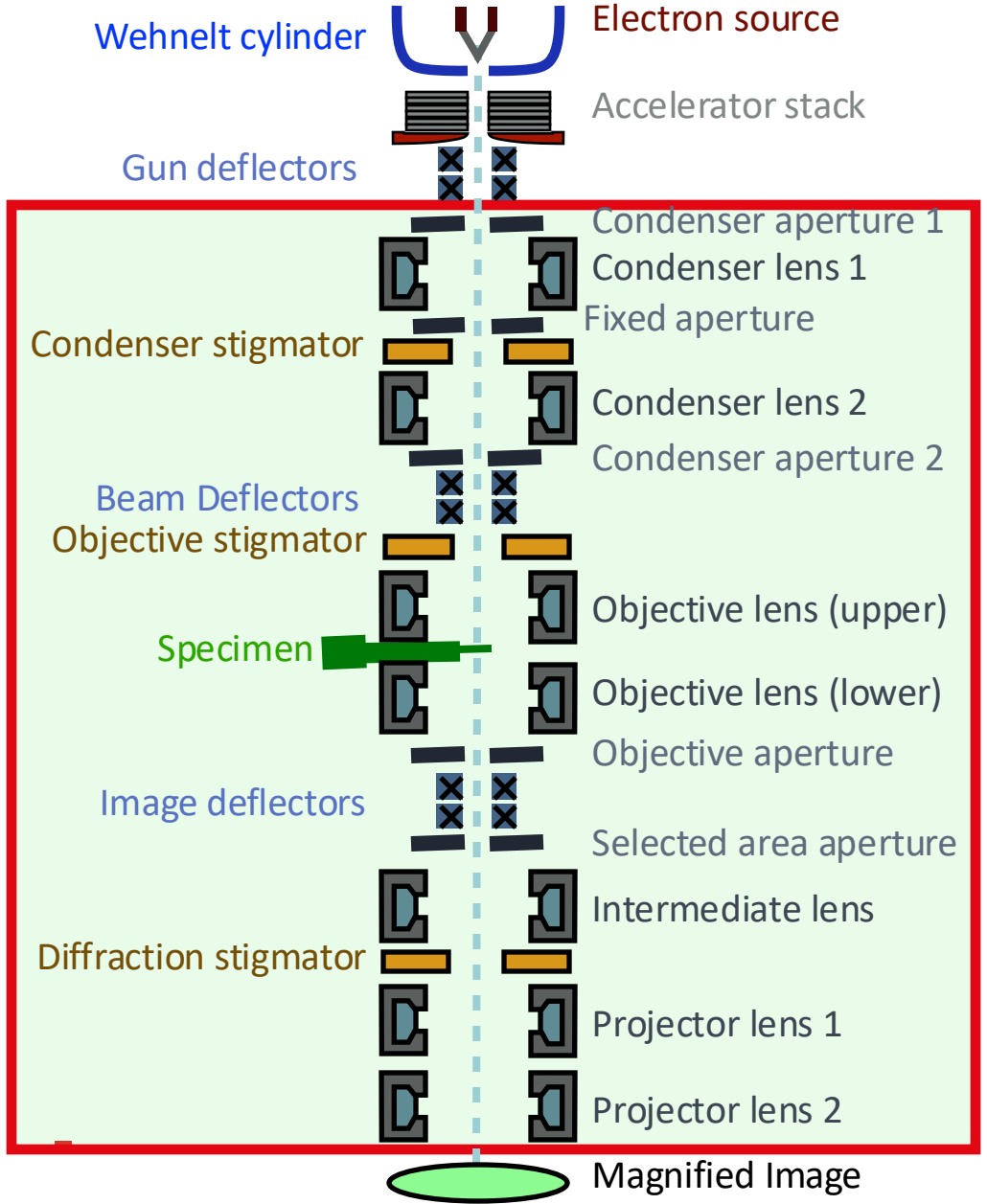


- The accelerator stack is based on **Cockcroft–Walton** voltage generator (Nobel Prize 1951)
- Originally designed by **Heinrich Greinacher**, a Swiss physicist



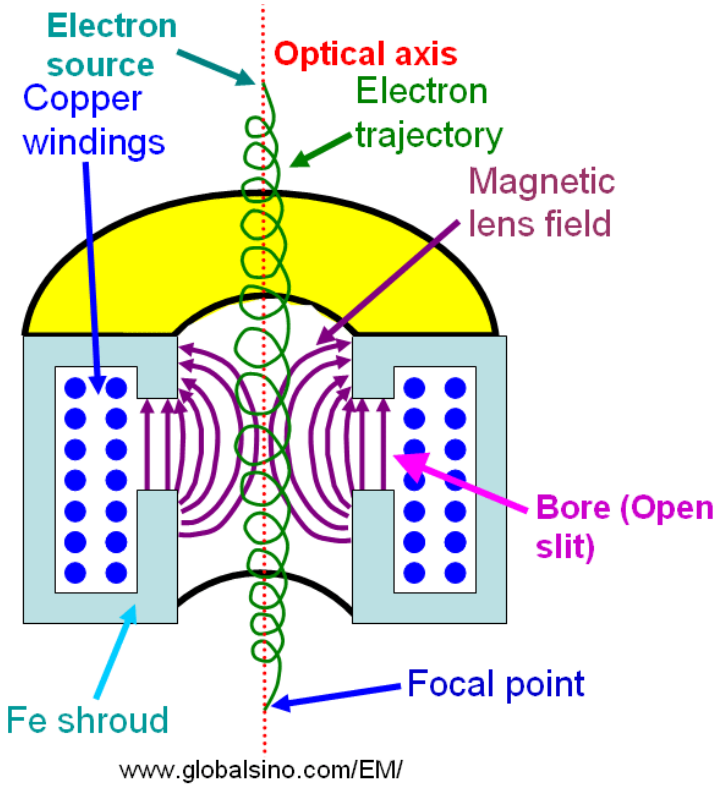
Accelerating voltage is typically **100-300kv**

Electromagnetic lenses



- Electromagnetic lenses change strength as excitation current changes.

- The magnetic field of the lens gives electrons a spiral path.

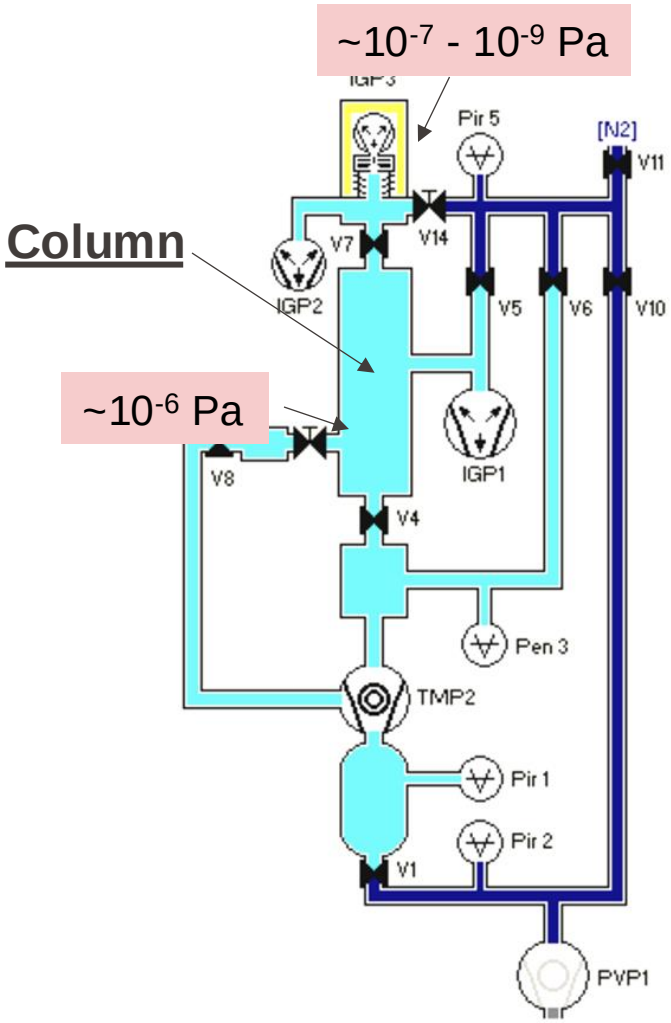


- The further the electrons are from the optical axis, the stronger the focusing effect.

Vacuum system

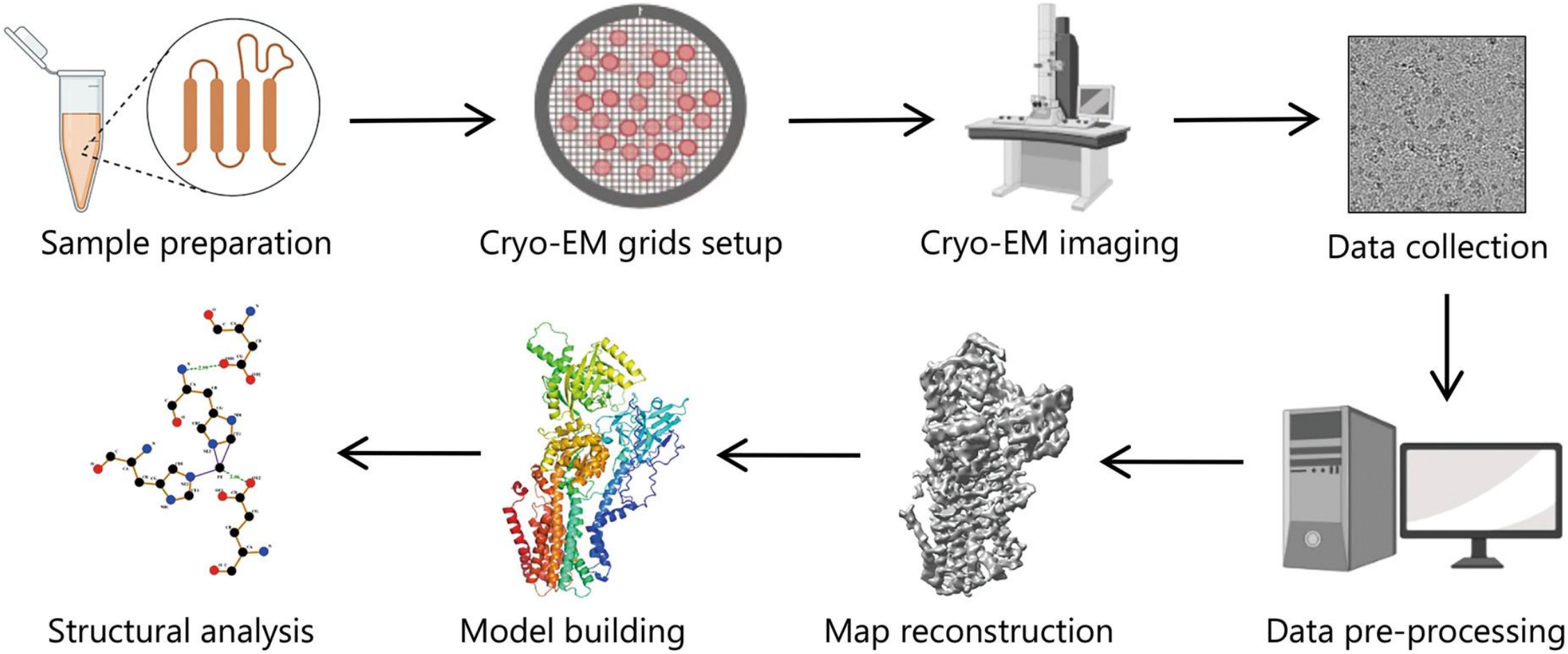
- Electrons interact with air, thereby the column needs to be kept at high vacuum

- Different types of vacuum pumps:
  - Rotary pump
  - Oil diffusion pump
  - Turbomolecular pumps
  - Ion-getter pumps



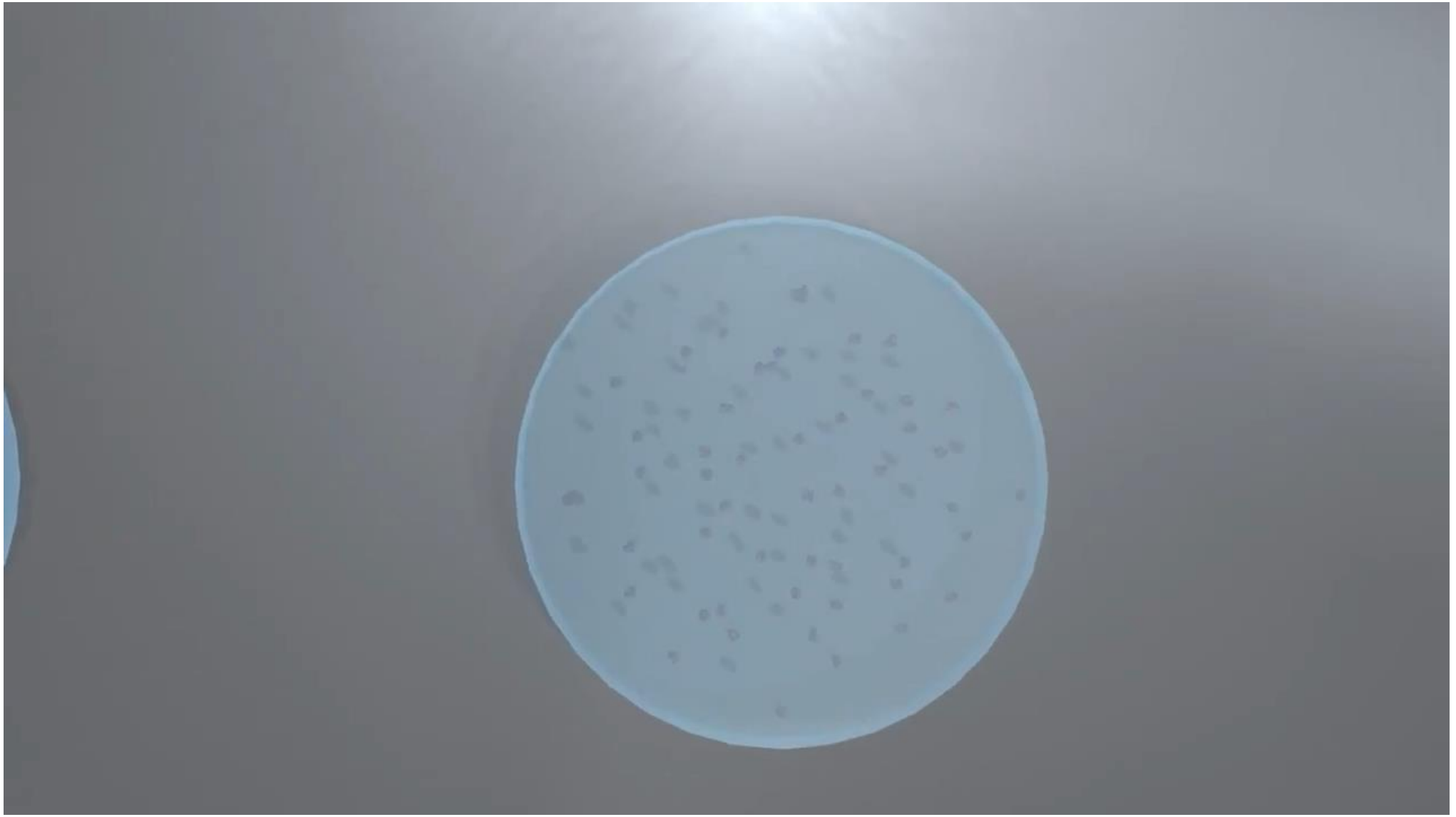
Fun fact: Even at high vacuum (~10^{-10} Pa) 30k molecules per cm^3 will still be present.

Typical cryo-EM workflow



- Unlike X-ray crystallography, cryo-EM maps are recovered in real-space

What does the grid look like?

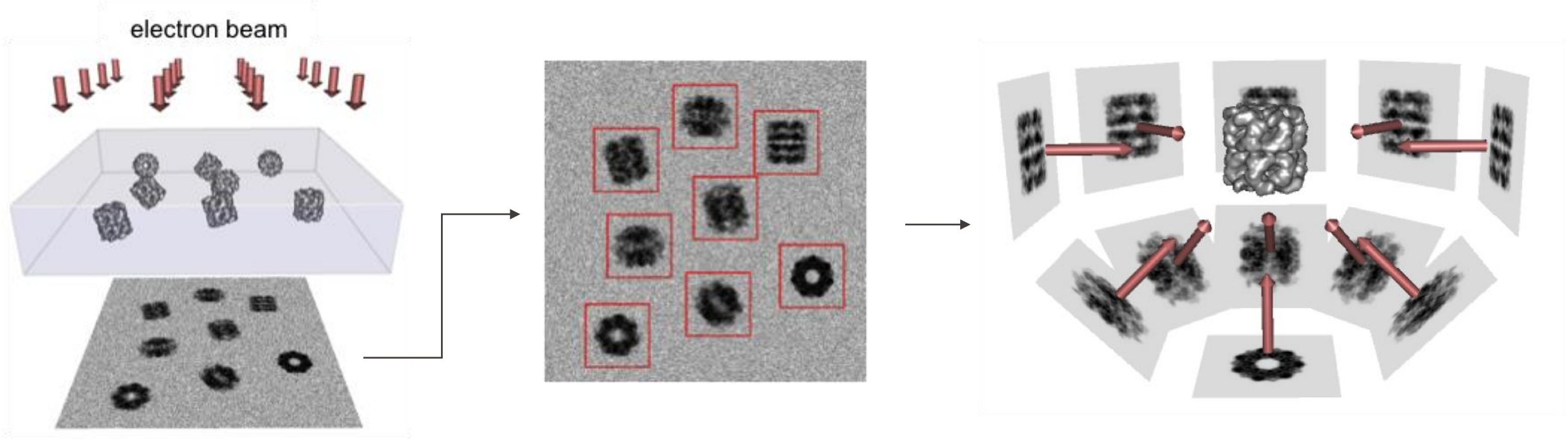


Data analysis and reconstruction of 3D models

Imaging

2D projection images

3D reconstruction



- Samples are vitrified in a thin layer of ice which provides structural support to the layer and reduces radiation damage
- 2D projection images are extracted from each micrograph and used to recover the 3D model of the object(s)

Adapted from: https://people.csail.mit.edu/gdp/cryoem.html

“Resolution Revolution” and “Democratization” of EM

Improved resolution of EM data

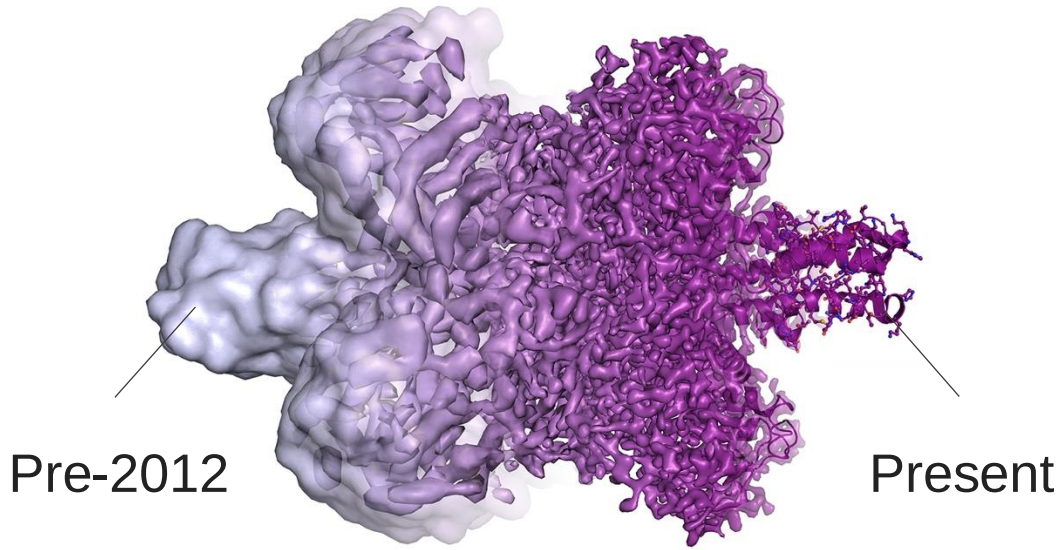
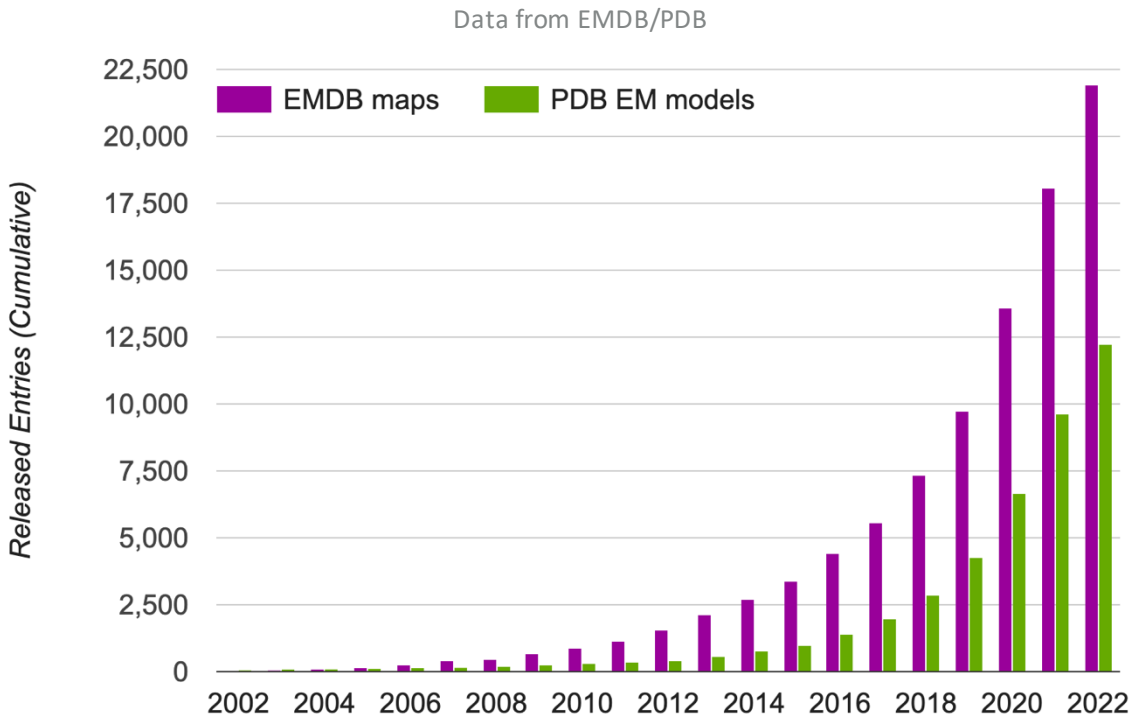


Illustration by Martin Högbom; The Royal Swedish Academy of Science

EM map/model depositions



- High-resolution maps became readily attainable starting ~2012 as a result of technological breakthroughs in the field, such as the development of direct electron detectors, new electron sources, 300kV microscopes and computational tools

“Resolution Revolution” and “Democratization” of EM

Improved resolution of EM data

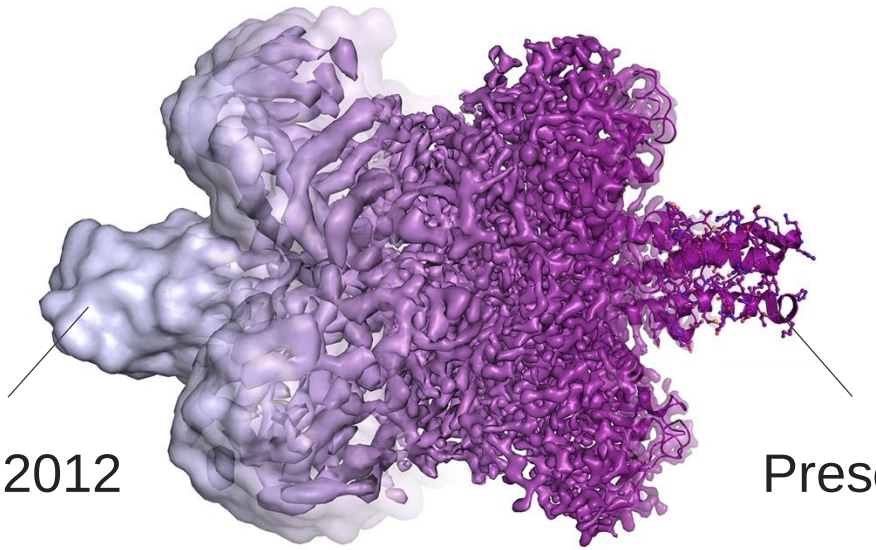
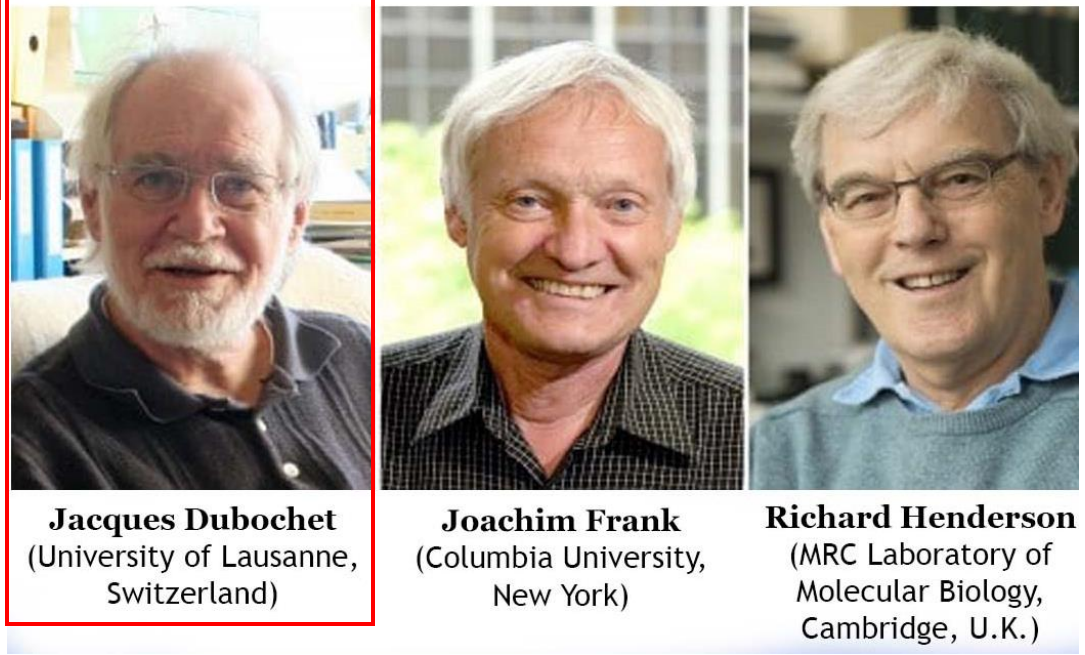


Illustration by Martin Högbom; The Royal Swedish Academy of Science

Nobel Prize in Chemistry, 2017



“for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution”

- High-resolution maps became readily attainable starting ~2012 as a result of technological breakthroughs in the field, such as the development of direct electron detectors, new electron sources, 300kV microscopes and computational tools

Bicycle parking spot at Biophore

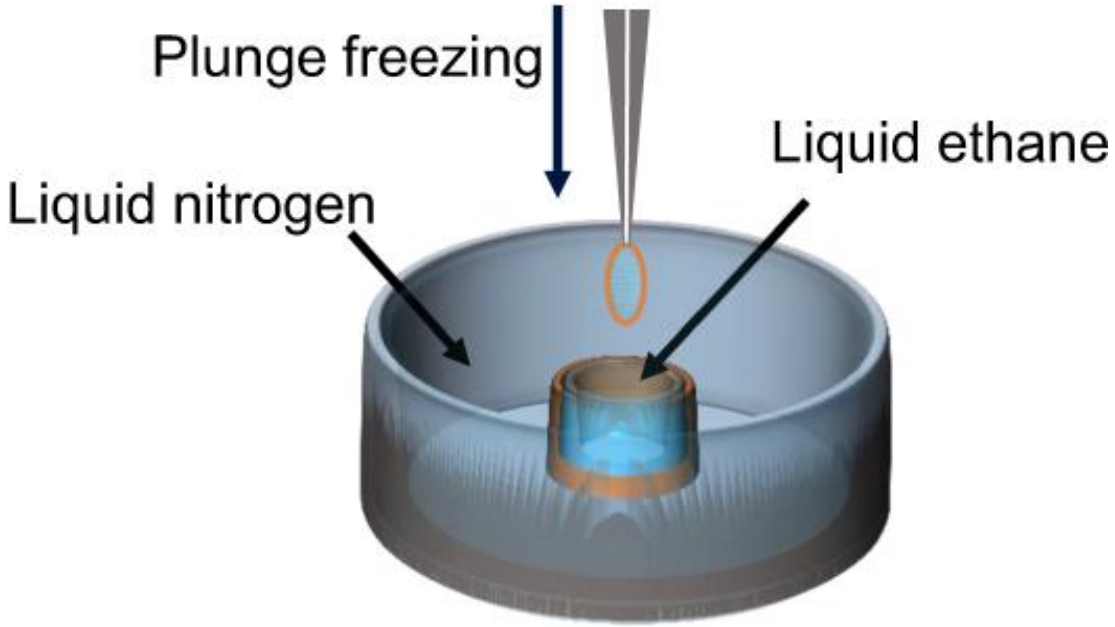
©DCI – DUBOCHET CENTER FOR IMAGING

Preserving Biological Specimens for EM Imaging

- Biological samples are about 80% water
  - Placing biological samples into the vacuum of electron microscope will cause immediate dehydration and destruction of sample
  - Need to preserve ultrastructure to withstand vacuum
- Biological samples are susceptible to radiation damage
  - Conversion of electron beam flux to radiation dose:
    $1 Coulomb/m^2$ @100keV = 40Mrad ($4x10^7$ rad)
  - To view samples in EM, typically use $30C/m^2/sec$ (1200Mrad/sec)

Equivalent to standing 30m from a 10Mt H-bomb explosion

Plunge-freezing method (the “cryo” part of cryo-EM)

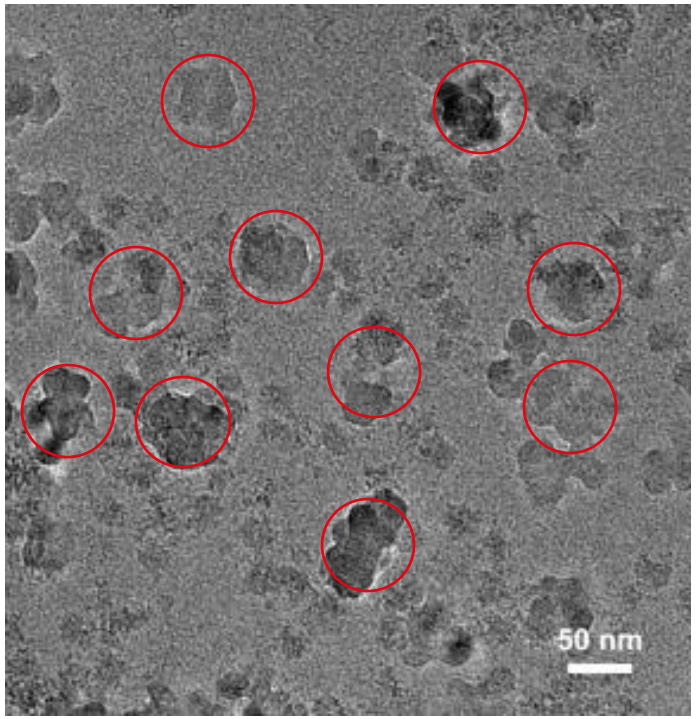


Ethane has a much higher heat capacity than liquid nitrogen.

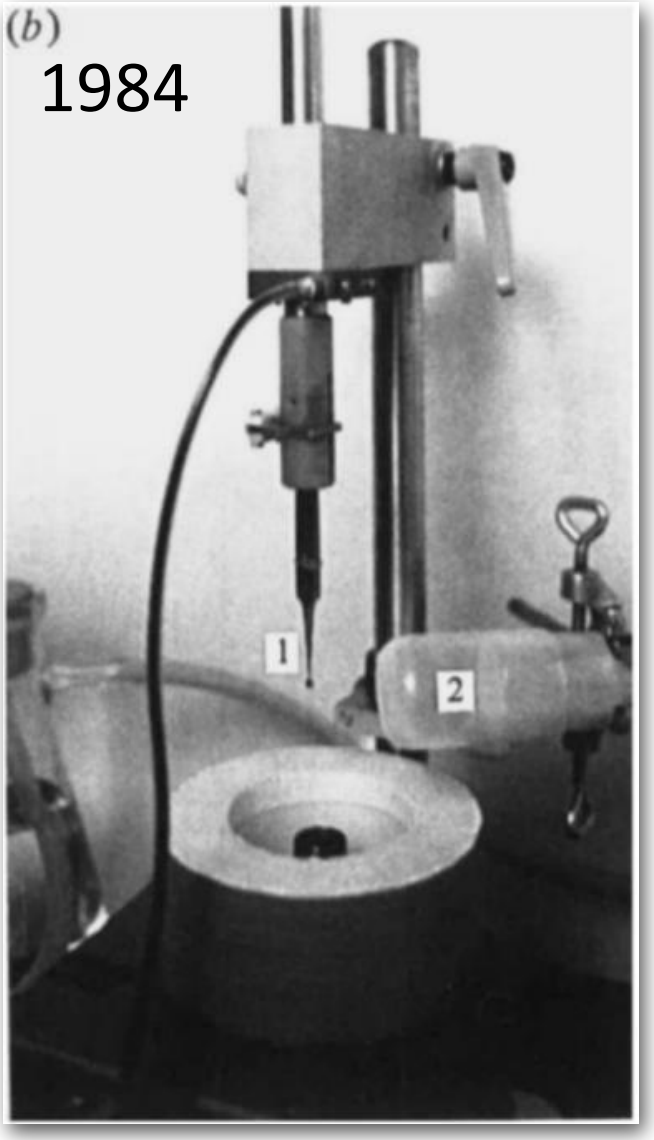
Ethane is liquid at temperatures just slightly above those of liquid nitrogen (its melting point is -188 °C)

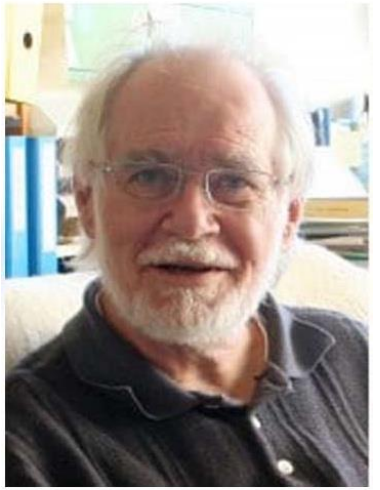
Therefore, liquid ethane is cold enough to vitrify water quickly, while not boiling off in the process.

Rapid vitrification is critical to avoid crystalline ice formation!



Rapid vitrification is critical – early design

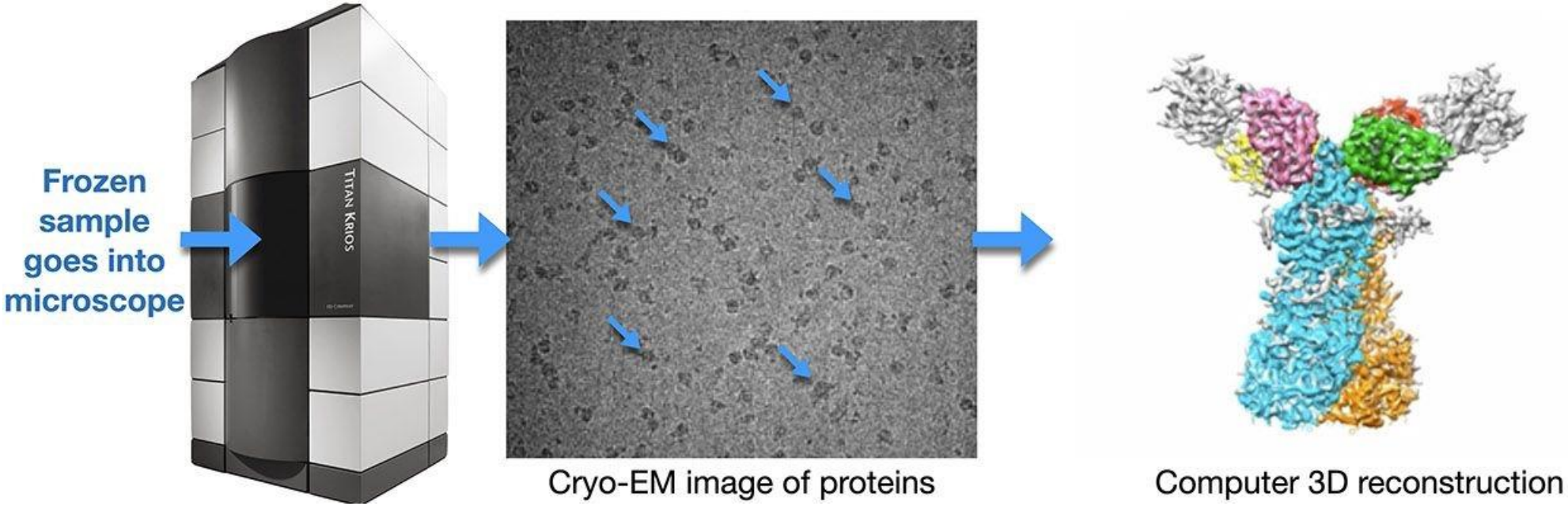




Jacques Dubochet
(University of Lausanne, Switzerland)

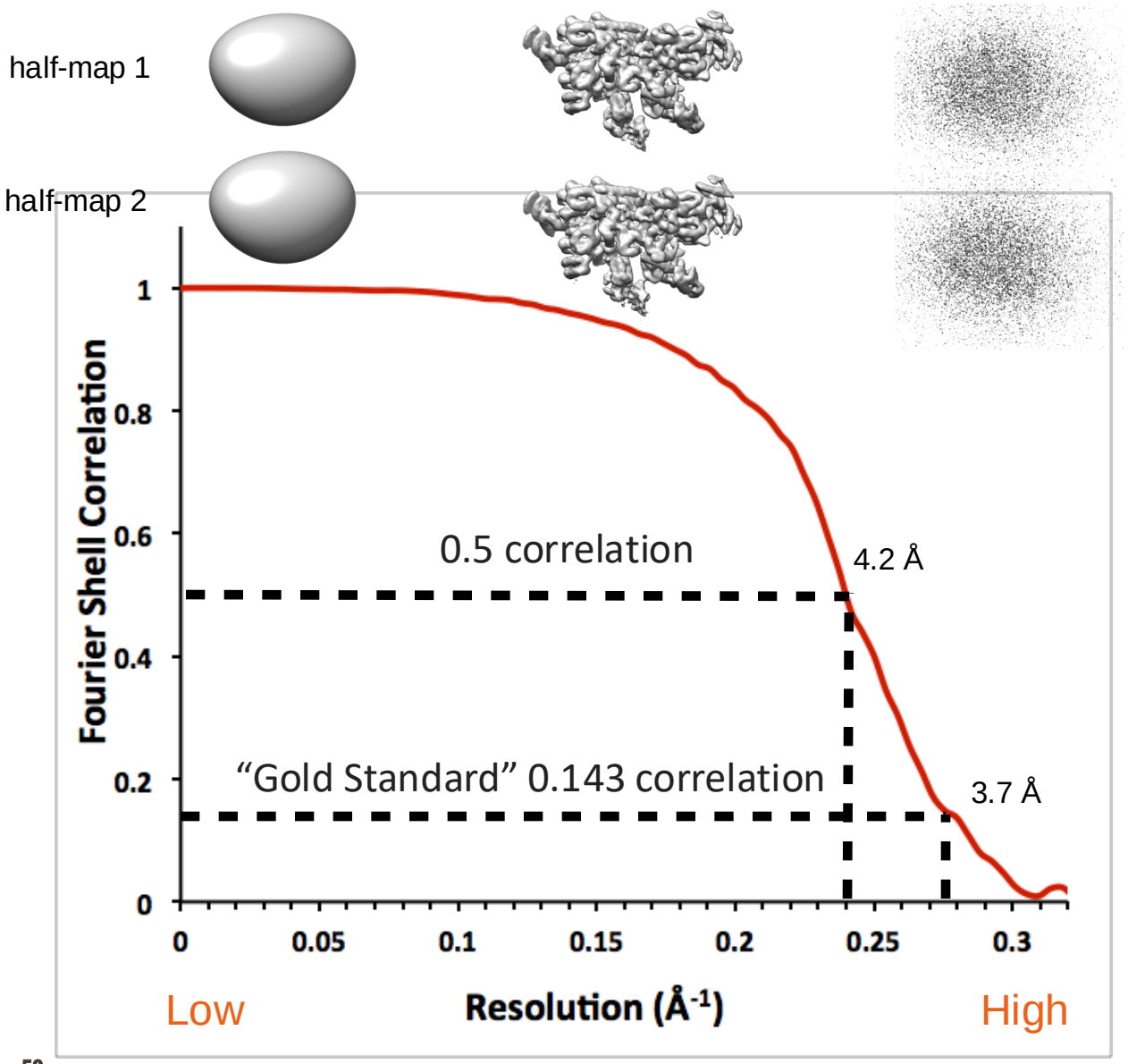
Many particles needed to recover maps from 2D projections

- As many as **~10^3 - 10^6 particles go into final map reconstructions**. Typically, the more -> the better



But how is map resolution estimated?

Estimating the resolution of a cryo-EM map



- **Fourier Shell Correlation (FSC) plot**

- Correlation between resolution shells in 2 independently refined half-sets of data (particles)

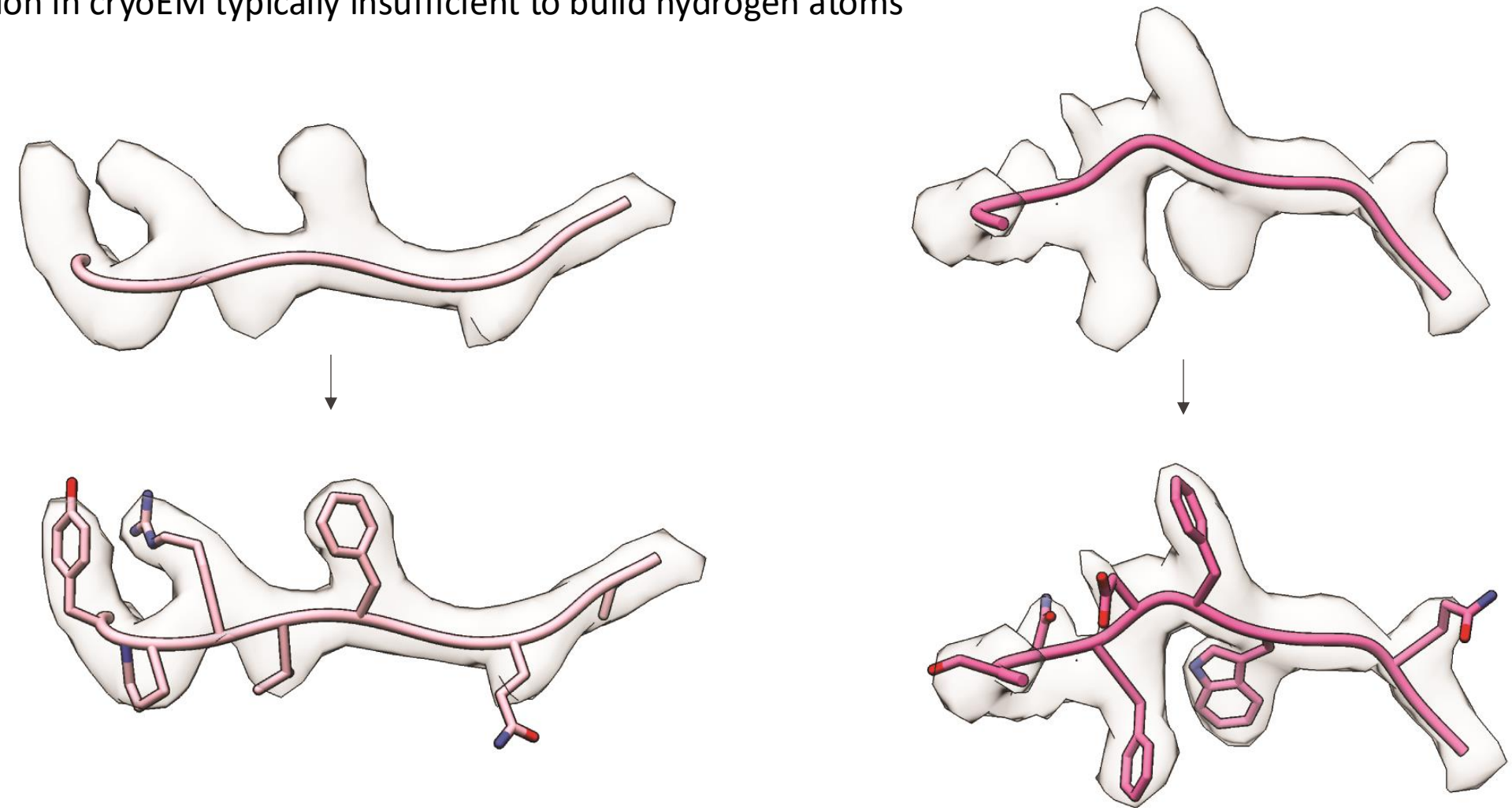
$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}}$$

F1 – Structure factors for volume 1
F2* – Complex conjugate of the structure factors for volume 2
$\mathbf{r_i}$ – Voxel element at the radius r

- "Gold Standard" is the most commonly used criterion to define global resolution

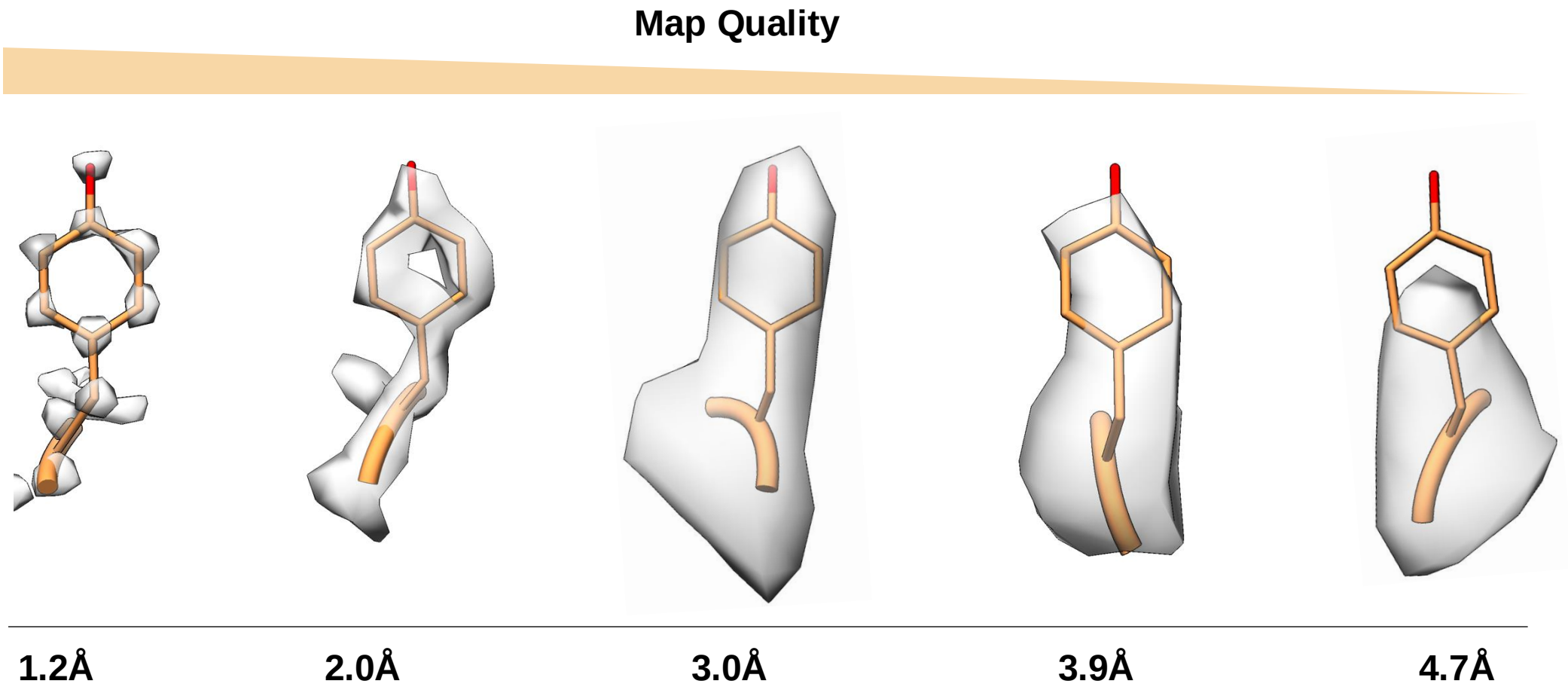
cryoEM map is used to build atomic models

- cryoEM map is used as a set of 3D restraints to approximate an atomic model that best recapitulates the reconstructed map.
- Amino-acid, nucleotide or monosaccharide sequence (depending on the biomolecule) needs to be known beforehand.
- Resolution in cryoEM typically insufficient to build hydrogen atoms



cryoEM maps at different resolutions

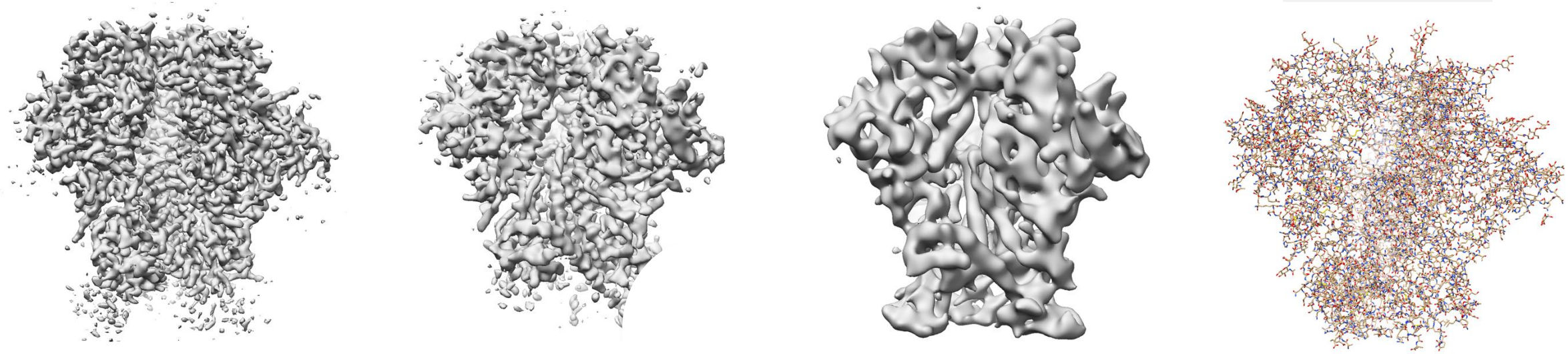
- Side-chain of residue Y28 of apoferritin at different EM map resolutions
- Structural data alone is insufficient to unambiguously build atomic models in maps >5Å



PDB IDs: **7A6A, 6WX6, 5YI5, 6HPI, 4V1W**

Let’s compare some protein maps

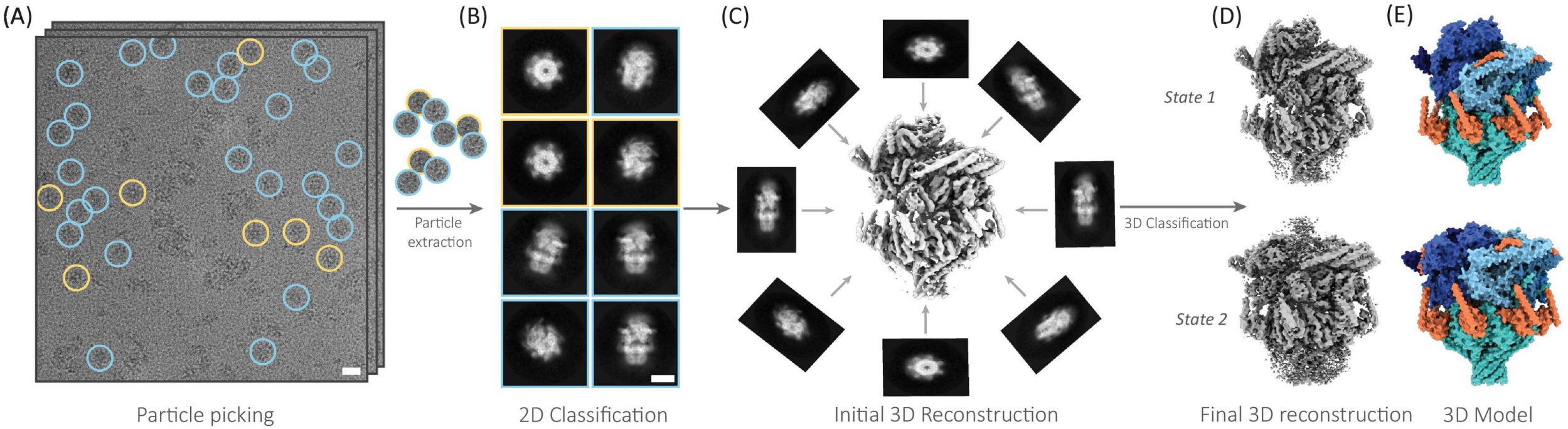
- What map has the highest resolution?
- Can you guess what the values may be?



HIV glycoprotein at different resolutions

cryoEM analysis allows to reconstruct multiple maps

- 2D and 3D image classification algorithms allow to separate biomolecules that are compositionally or conformationally different

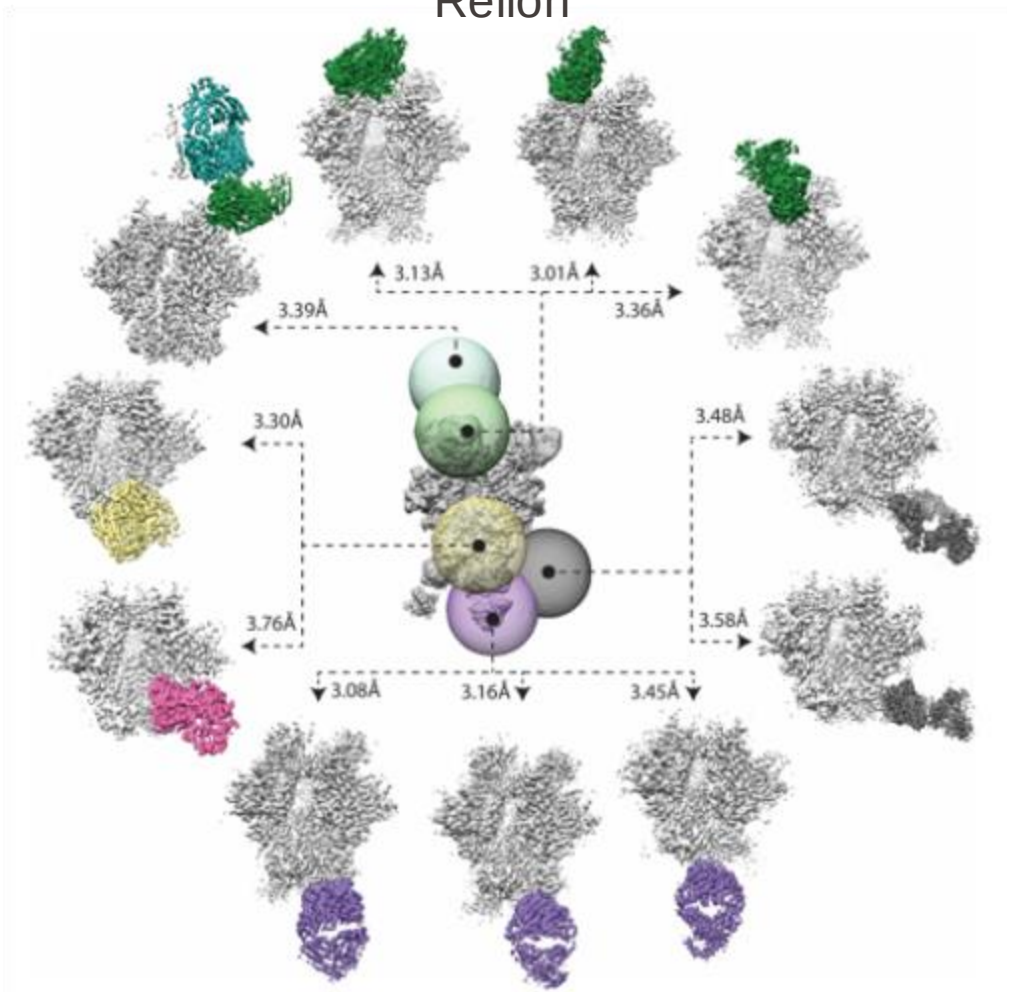


- In the above example, the authors identified 2 different states of the same protein from a single dataset

Image classification algorithms in cryoEM

Studying non-uniform assembly

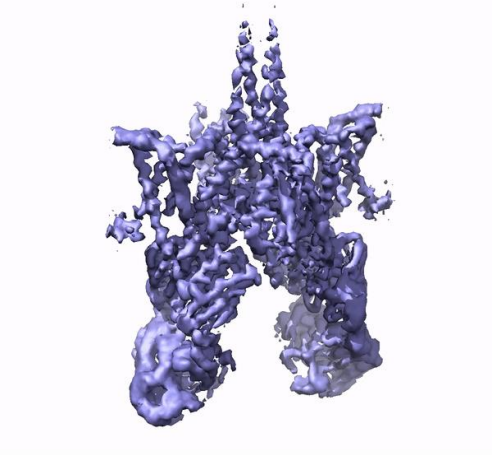
Relion



Antanasijevic et al., Nature Comm 2021

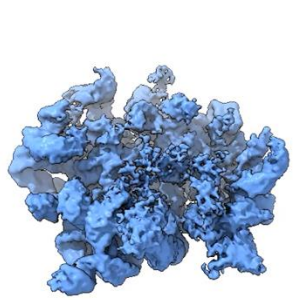
Studying conformational (dynamic) states

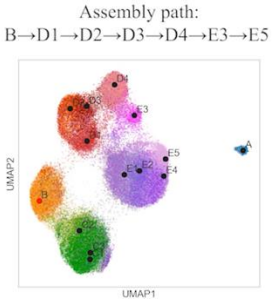
Variability Analysis (CryoSPARC)



Punjani & Fleet JSB 2021

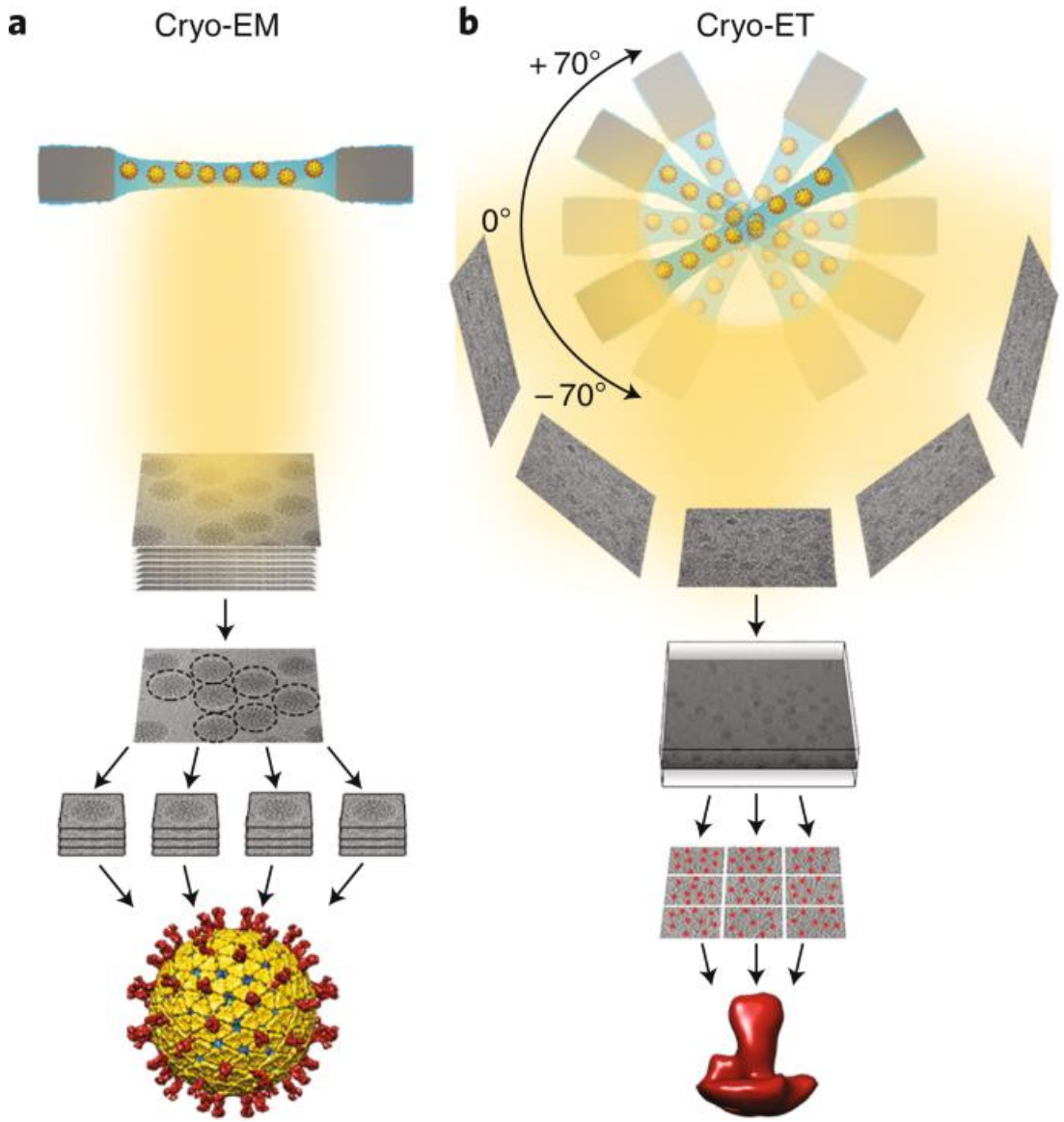
CryoDRGN





Zhong et al. Nat Methods 2021

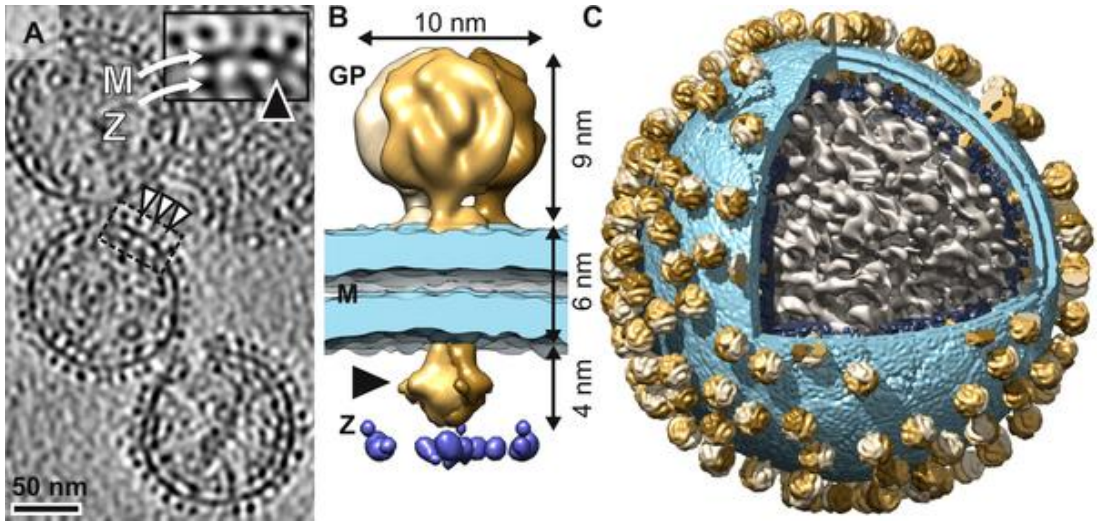
Cryo-electron tomography (cryo-ET)



- CryoET allows to reconstruct ~nm resolution three-dimensional views of complex assemblies such as protein complexes, pleomorphic viruses, bacterial pathogens, cells etc.
- Very flexible and versatile
- Tilt-series of images (typically every 1-2˚ in the range of -60˚ to +60˚) is used for tomographic reconstruction
- Subtomogram averaging can be applied to extract the signal corresponding to different components in the complex

LASV GP reconstructed by subtomogram averaging

Whole LASV particle



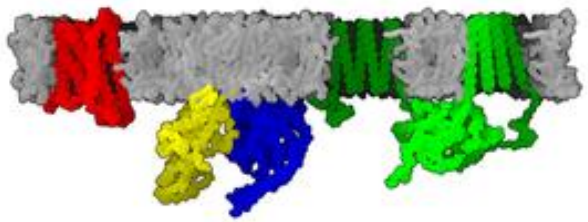
Luque et al., **Nature Chem Bio**, 2020

Li et al., **PLOS Path**, 2016

cryo-electron-microscopy - Pros and Cons

- **Positive sides**
- Very versatile and quick
- No requirement for protein labeling
- No requirement for crystallization
- No requirement for homogenous samples
- Real space imaging – no phase problem
- Can be used to study protein dynamics
- Can be expanded to larger assemblies (e.g., viruses and cells) and very heterogeneous samples (e.g., membrane proteins)

- **Negative sides**
- Grid preparation procedure requires screening
- Limited to samples with MW >40kDa
- Preferred orientation problems (i.e., cannot reconstruct a 3D map if there is limited distribution of orientations in 2D images)
- Computationally heavy (TBs of data + requirement for GPU processing)

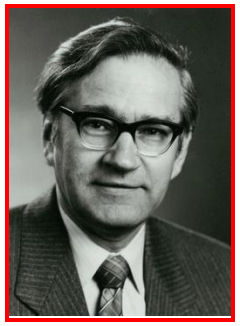


Membrane proteins

Nuclear Magnetic Resonance (NMR)

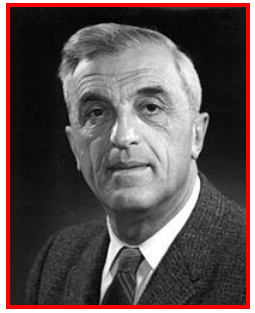
Nuclear Magnetic Resonance Spectroscopy

- Indirect determination of relative atomic coordinates using nuclear magnetic spins and radio-frequent radiation (~MHz-GHz range)

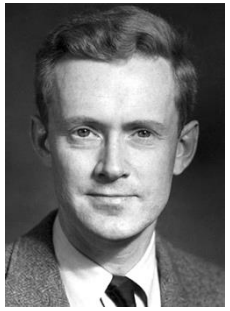


Richard Ernst
(1933-2021)

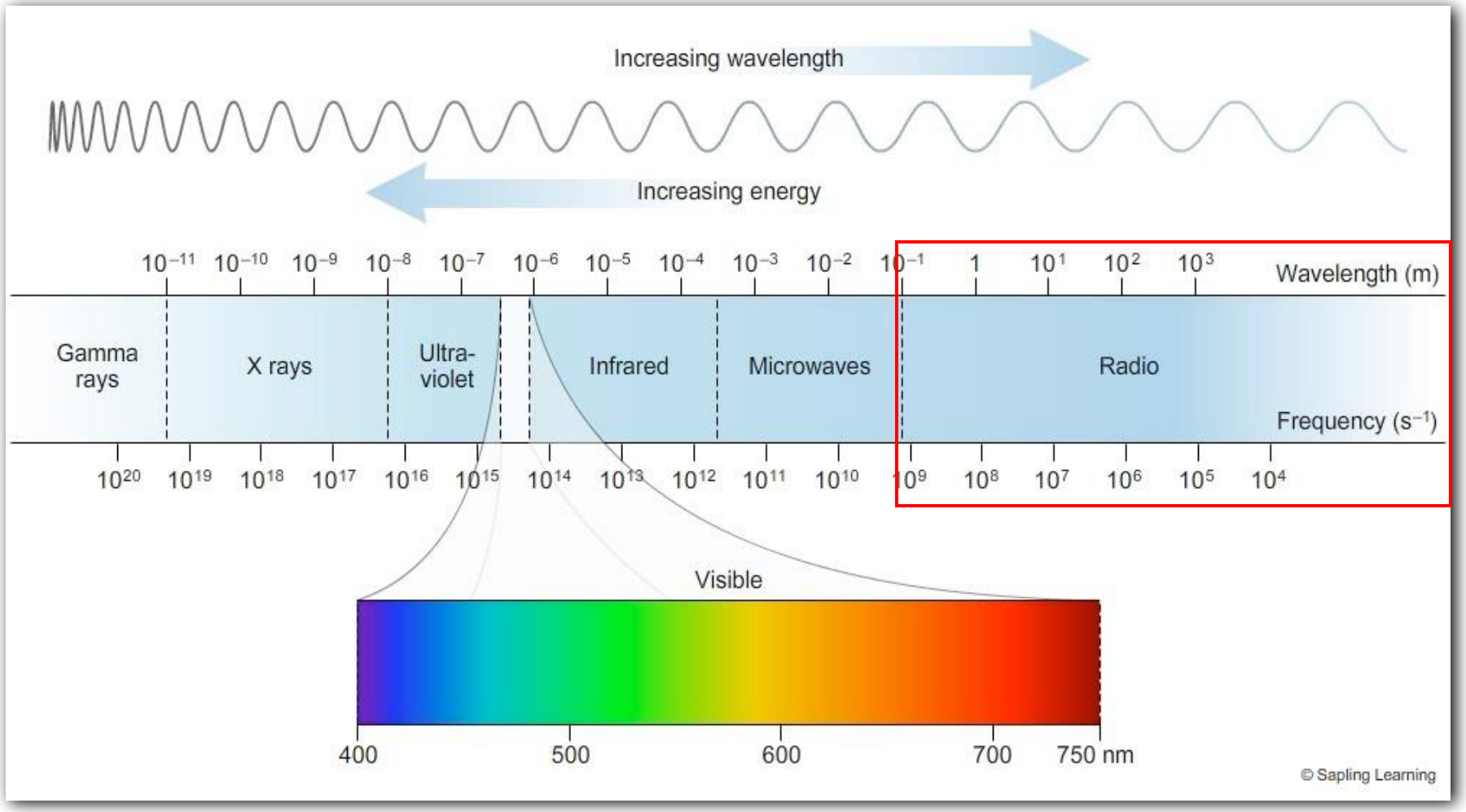
Isidor Isaac Rabi
(1898-1988)



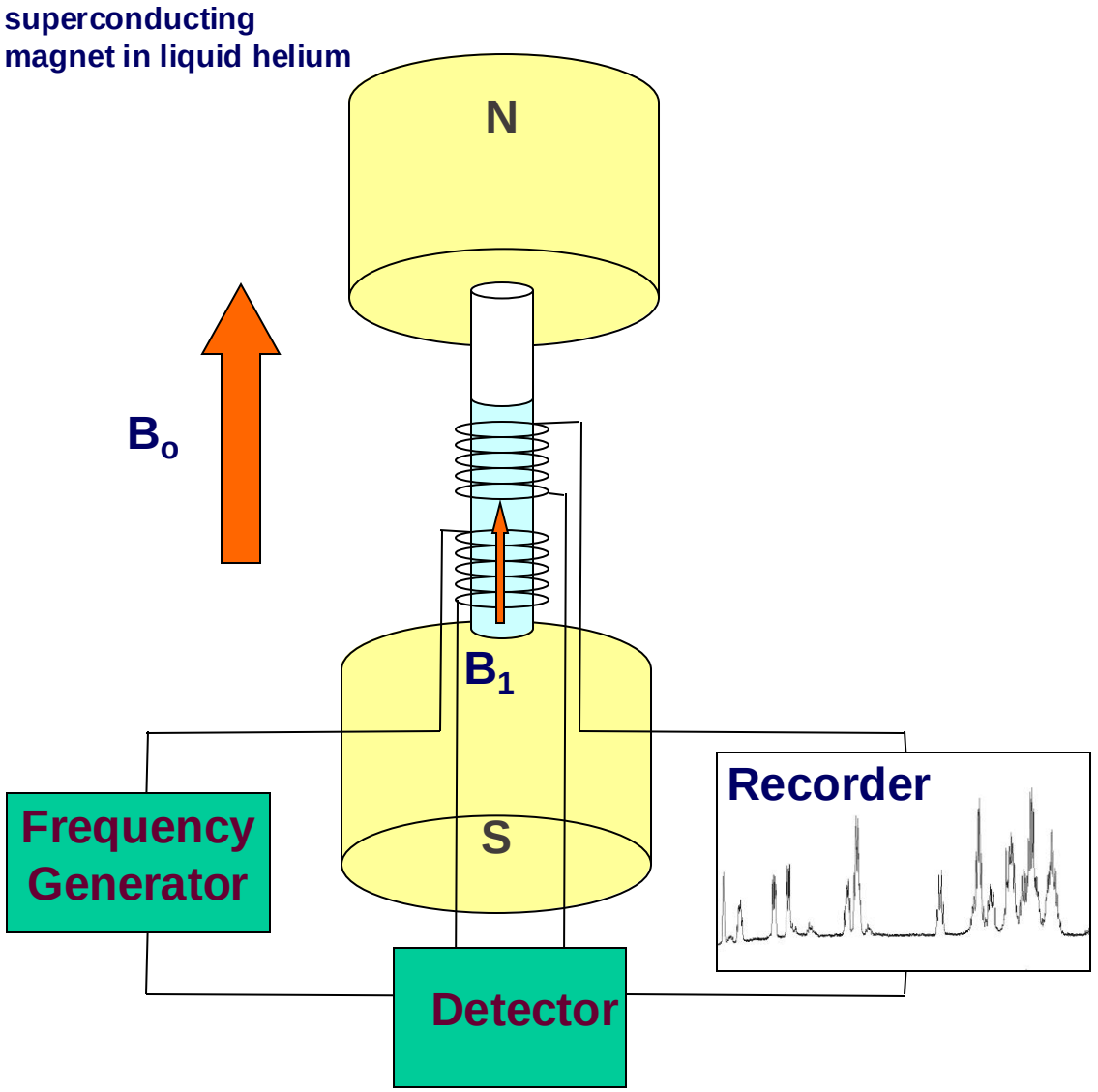
Felix Bloch
(1905-1983)



Edward Purcell
(1912-1997)



NMR Spectrometer

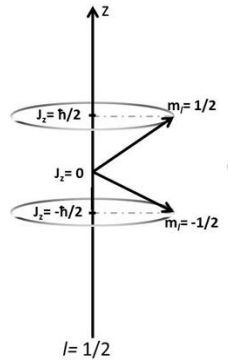


B_0 = 21.1 Tesla magnetic field > ^{1}H resonates at 900 MHz.
(magnetic field of the Earth is around 50 µT)

NMR sample tube
(400-500 µl protein needed)

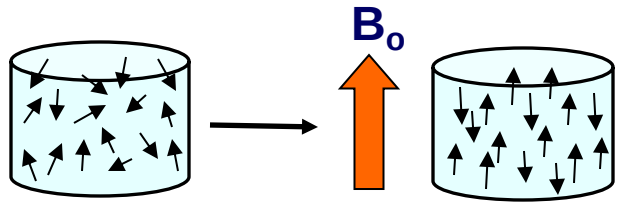
Nuclei in the magnetic field

S =1/2 implies $m_s = \pm 1/2$
S = nuclear spin
m_s= quantum spin number



• NMR is the phenomenon occurring when “NMR active” nuclei at equilibrium in a static magnetic field are exposed to an external oscillating magnetic field.

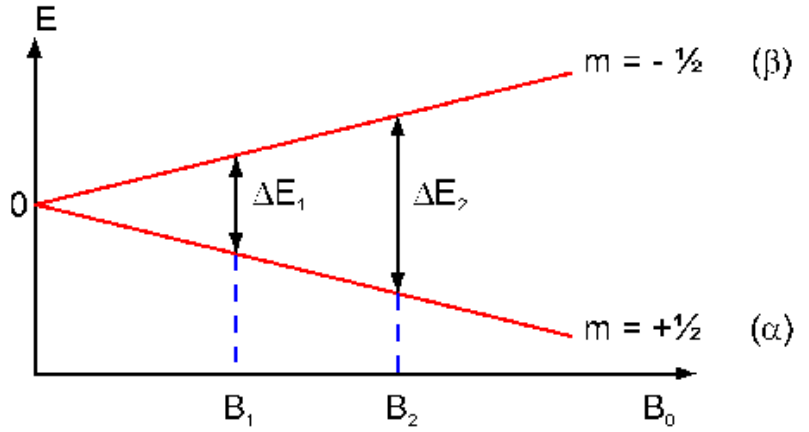
• NMR active nuclei are those that have a magnetic spin ideally of 1/2 (e.g.,: **^{1}H, ^{13}C, ^{15}N, ^{31}P**)



• In a magnetic field, the nuclei orient either parallel or antiparallel to the direction of the field creating energy levels

$\Delta E = \gamma h B_o / 2\pi$ γ : Gyromagnetic constant

• The difference in energy levels is directly proportional to the strength of the magnetic field

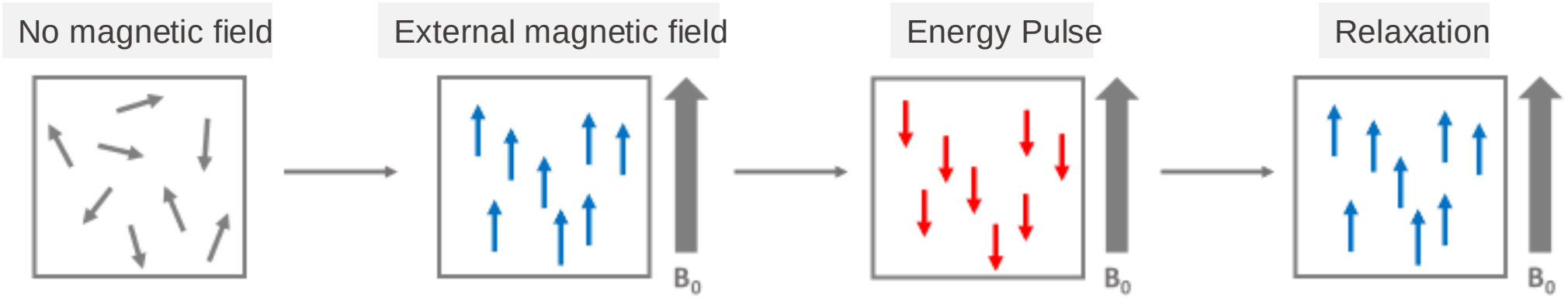


• The difference in energy levels matches the energy of radio-frequent radiation

Note: To readily detect C and N atoms in NMR experiments isotope labelling with ^{13}C and ^{15}N is necessary

Nuclear Magnetic Resonance Spectroscopy

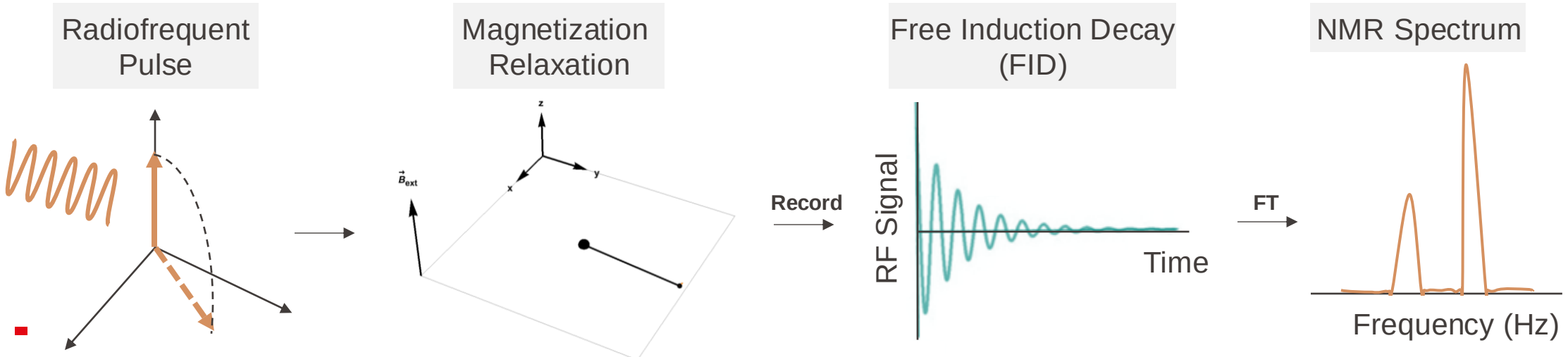
- Radio-frequent pulse is used to deflect the nuclei out of their equilibrium distribution



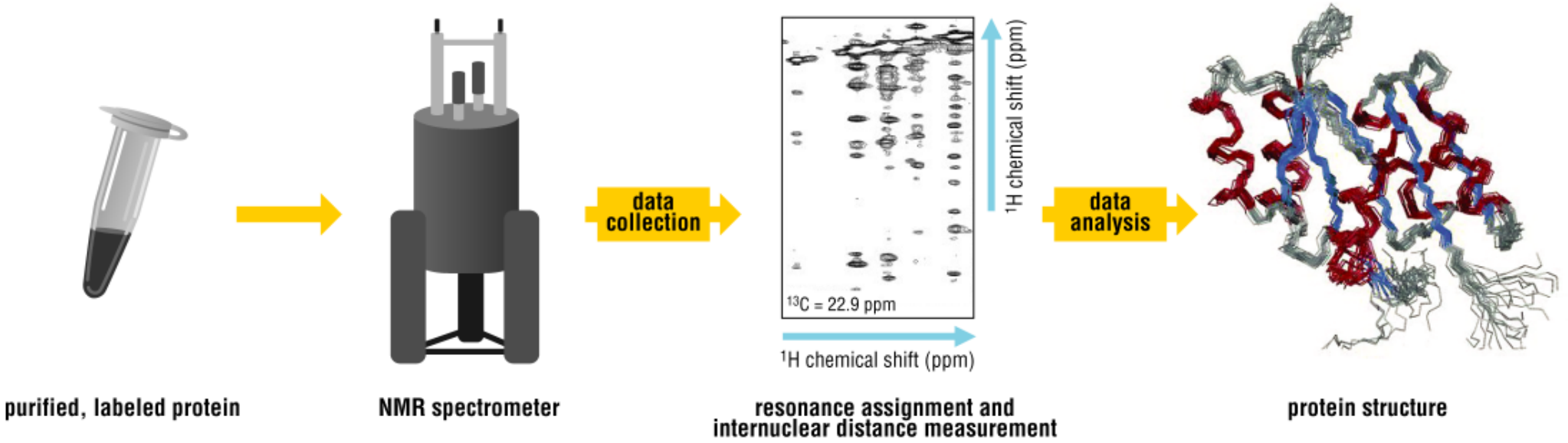
Larmor Frequency

$$\nu = \gamma B_o / 2\pi$$

- To return to equilibrium perturbed nuclei emit RF radiation which is recorded as FID and then converted to frequencies by Fourier transform (FT)



From a protein in solution to a structure using NMR



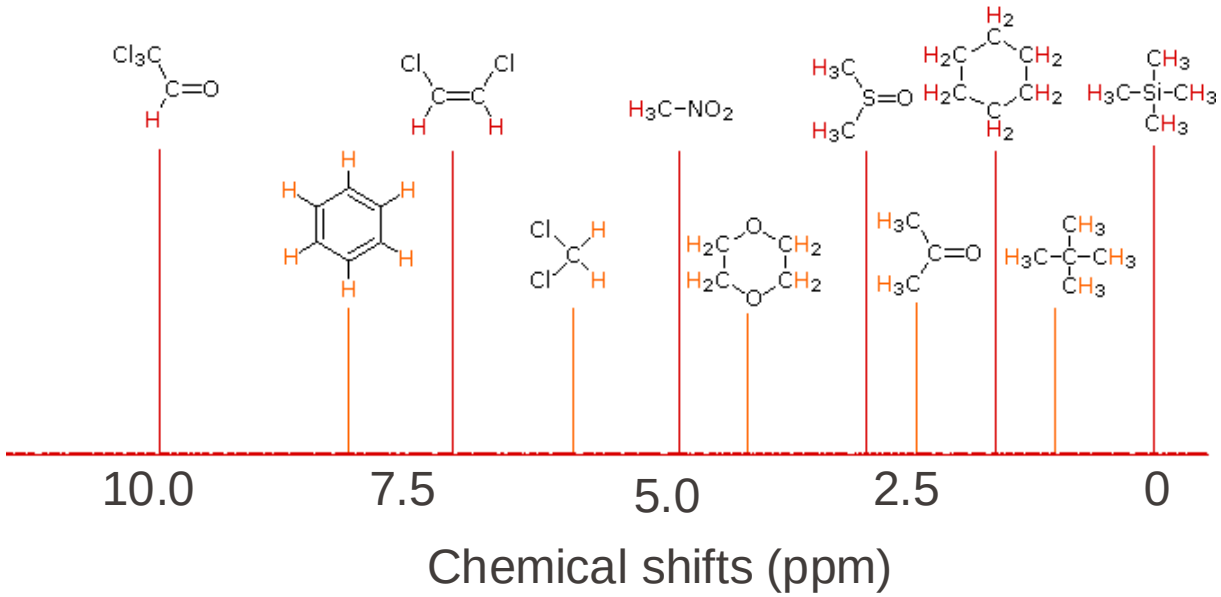
- Completely different to crystallography and cryoEM, as proteins are studied **in solution**
- Several different NMR experiments give you information on spatial restraints of particular nuclei (e.g. proton of NH group of Ala-25 is close to C alpha of Tyr-88)
- The collection of several hundreds of such restraints is used to calculate **several models of the protein** that adheres to the experimental restraints

Chemical shifts (δ)

- All “NMR active” nuclei resonate at a characteristic frequency depending on their chemical environment (i.e., the degree of electron shielding)

- Chemical shifts are proportional to frequency δ reported in ppm with respect to a standard resonance (ppm=“parts per million”)
  - ^{1}H range ~ 10 ppm or 5000 Hz of 500,000,000 Hz or 9000 Hz of 900,000,000 Hz
  - ^{13}C range ~ 200 ppm or 25000 Hz of 125,000,000 Hz
  - ^{15}N range ~ 70 ppm or 3500 Hz of 50,000,000 Hz

* The detected frequencies (in Hz) are referenced against TMS (tetramethylsilane, Me_4Si), which is assigned the chemical shift of zero.



^{1}H-NMR spectrum of a protein is very complex

- 60 kDa protein has >4000 H-atoms which makes the ^{1}H spectrum “busy”

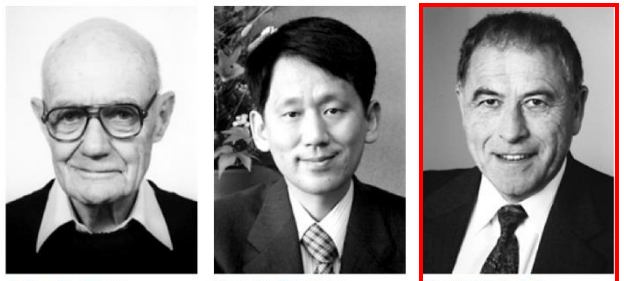
- One can distinguish folded from unfolded (denatured) proteins, but otherwise not too useful to get the structure of a protein since we cannot connect resonance to the atomic nuclei.

- **How can we identify the corresponding atomic nuclei in a protein?**

Enter these guys:

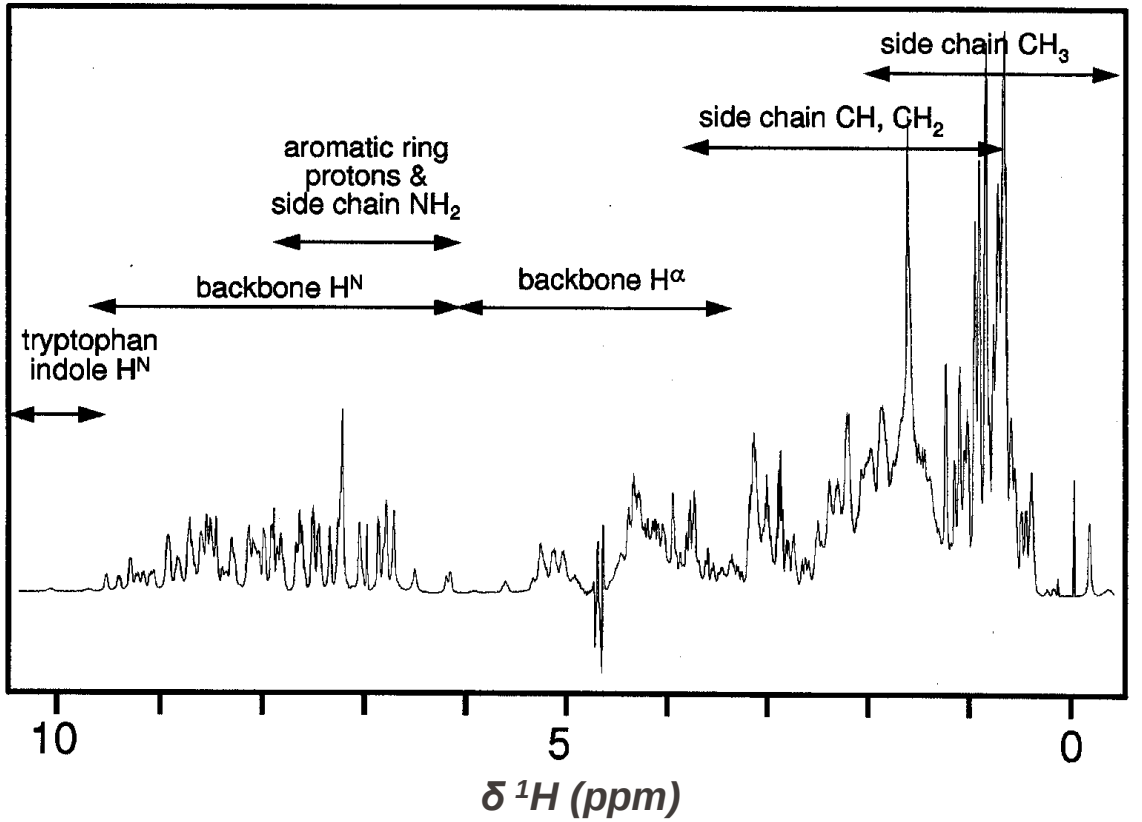
The Nobel Prize in Chemistry 2002
John B. Fenn, Koichi Tanaka, Kurt Wüthrich

The Nobel Prize in Chemistry 2002
Nobel Prize Award Ceremony
John B. Fenn
Koichi Tanaka
Kurt Wüthrich



John B. Fenn | Koichi Tanaka | Kurt Wüthrich

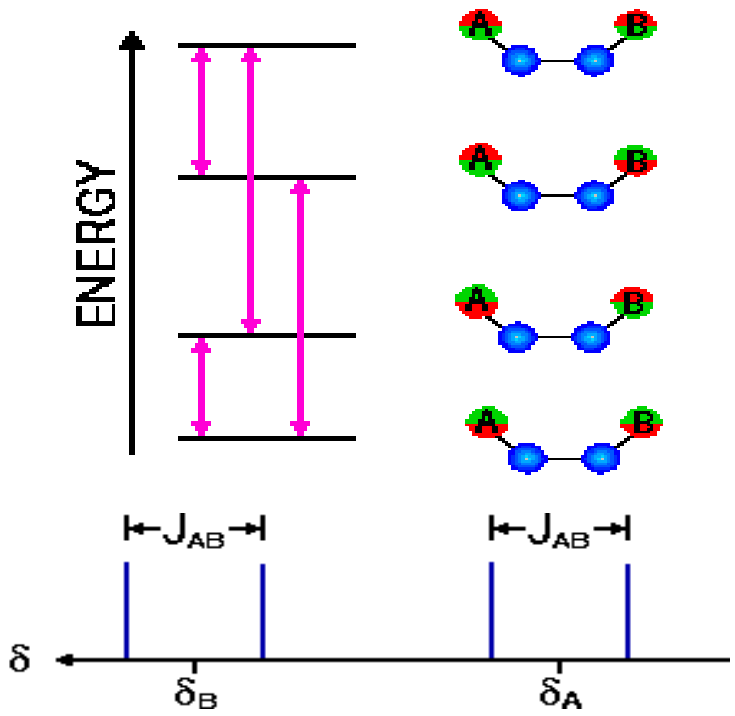
The Nobel Prize in Chemistry 2002 was awarded *"for the development of methods for identification and structure analyses of biological macromolecules"* with one half jointly to John B. Fenn and Koichi Tanaka *"for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules"* and the other half to Kurt Wüthrich *"for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution"*.



Magnetic coupling in NMR

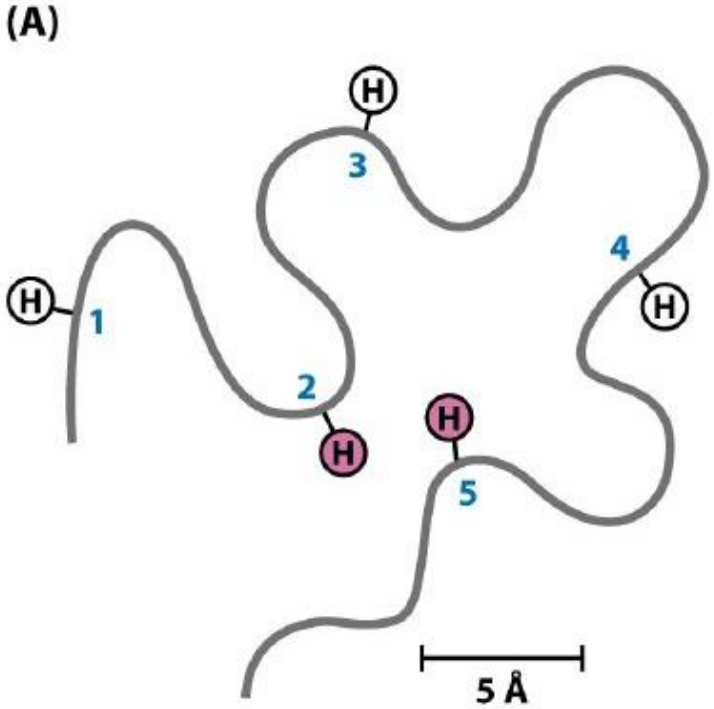
- NMR-active nuclei can “sense” each other if close in 3D space

- **Scalar (J) coupling**



Coupling of nuclei connected by 1-3 chemical bonds

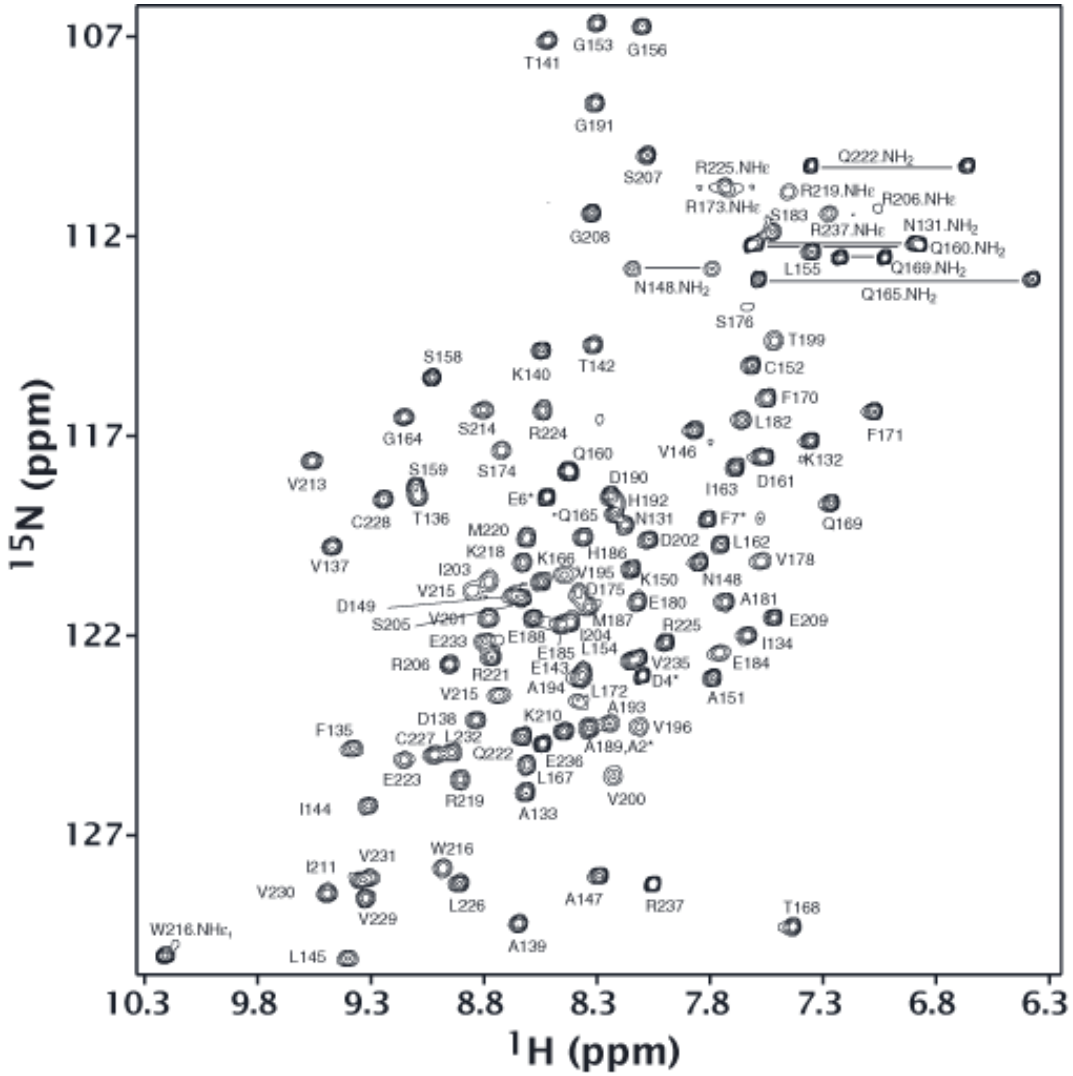
- **Dipolar (NOE) coupling**



Coupling of nuclei proximal in 3D space

- The coupling can be detected by the methods developed by Wutrich and others, allowing to determine which atoms are in close proximity via bonds or through space

2D NMR experiments: $^{15}N/^{1}H$-HSQC



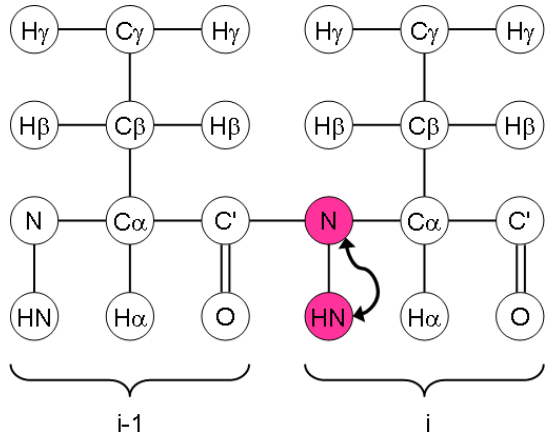
- **Correlates amide nitrogen with its proton**
- One peak for every amino acid (except Pro) i.e. **'fingerprint' of protein**
- Additional peaks for aa with NH group in sidechain, i.e. N, Q, K

Commonly used in protein NMR:

- Rapidly acquired (ca. 10 min)
- Cheap uniform ^{15}N labeling

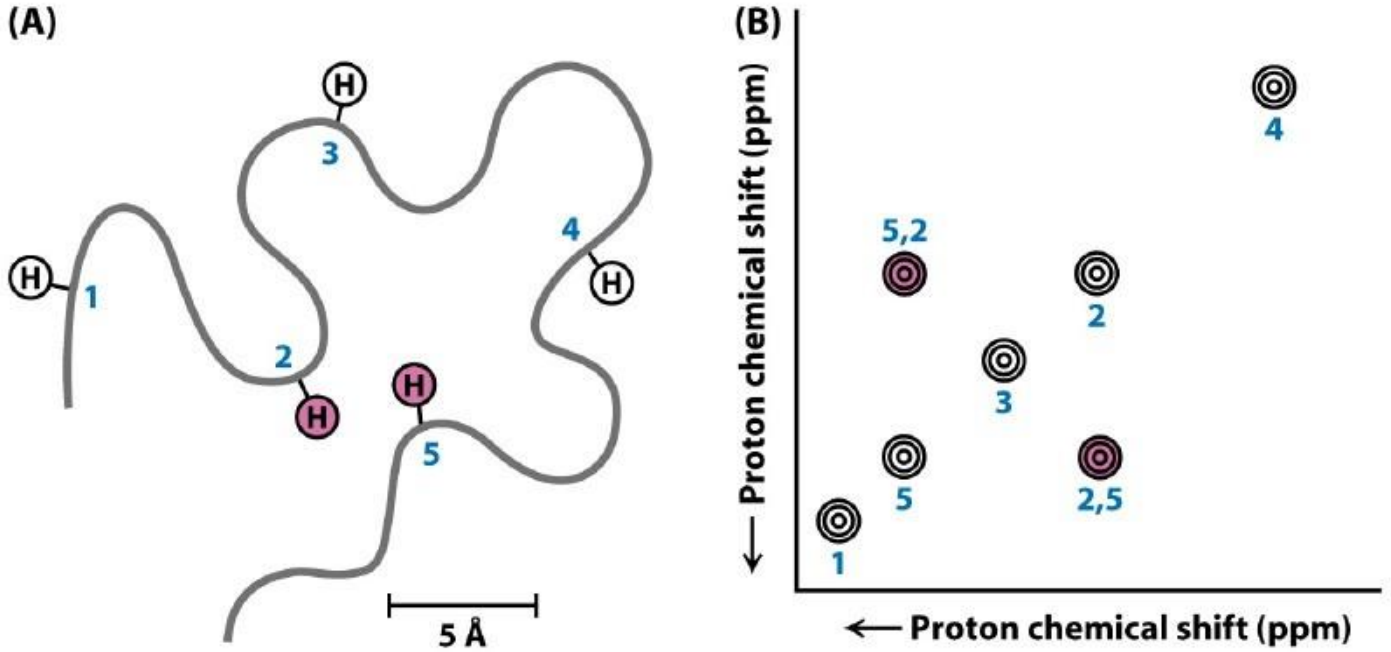
What it can assess:

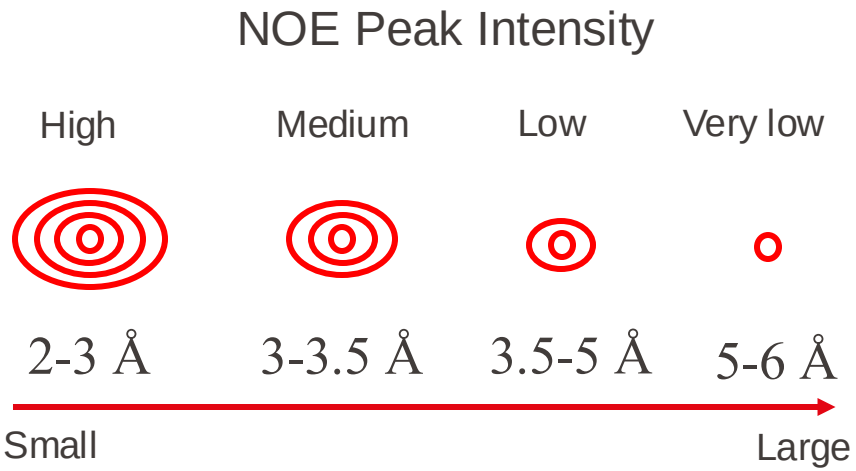
- Interactions (ligand, protein)
- Conformational changes
- Folding



* Heteronuclear Single Quantum Correlation

Nuclear Overhauser Effect Spectroscopy (NOESY)



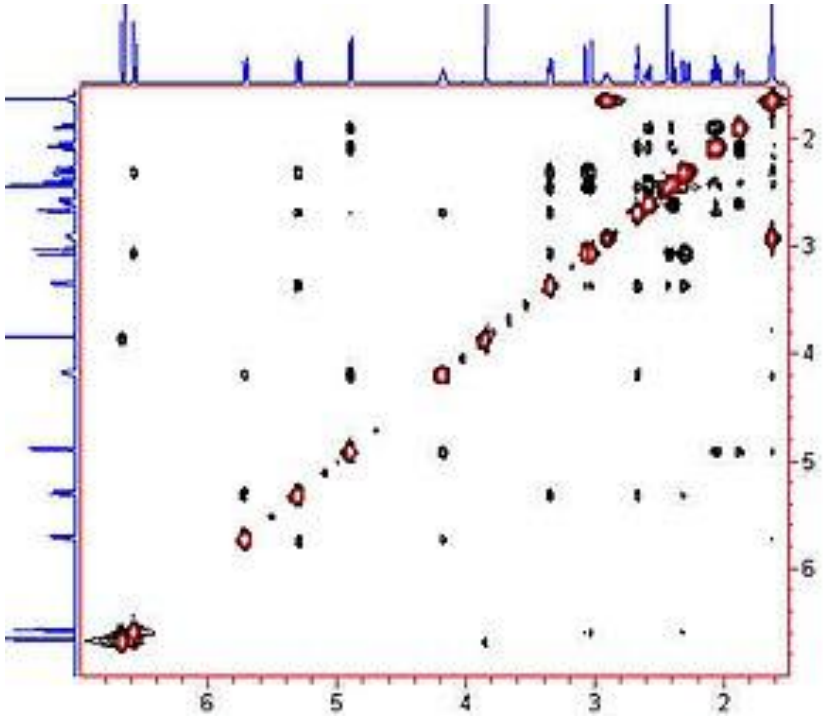


- Also protons that are not coupled directly via a few bonds, but are close in space influence each others' chemical shift (Nuclear Overhauser Effect)
- **The NOE is strongly distance dependent (~1/r^6), therefore detectable only for distances < 5-6 Å**
- Provides information on **spatial proximity of protons** in a protein

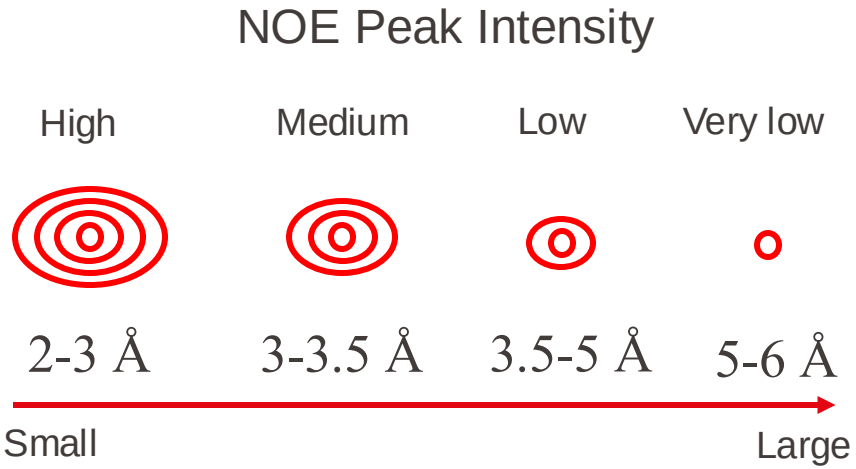

Kurt Wüthrich

Nuclear Overhauser Effect Spectroscopy (NOESY)

Real example



All photons off-diagonal originate from ^{1}H-^{1}H pairs close in 3D space

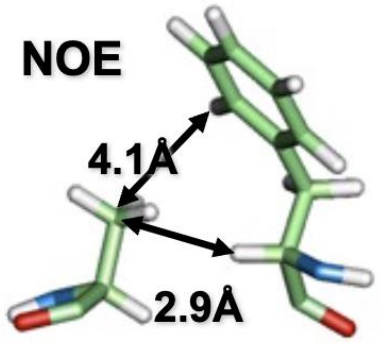


- Also protons that are not coupled directly via a few bonds, but are close in space influence each others' chemical shift (Nuclear Overhauser Effect)
- **The NOE is strongly distance dependent (~$1/r^6$), therefore detectable only for distances < 5-6 Å**
- Provides information on **spatial proximity of protons** in a protein

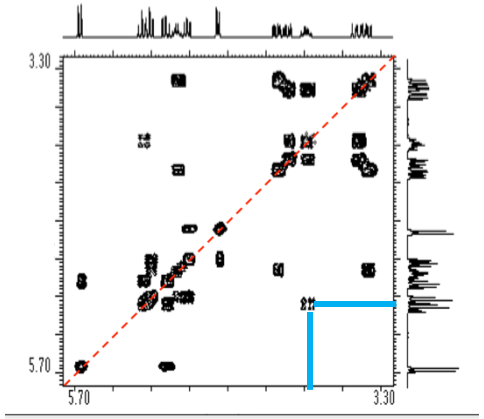

Kurt Wüthrich

How to get a structure from NMR experiments?

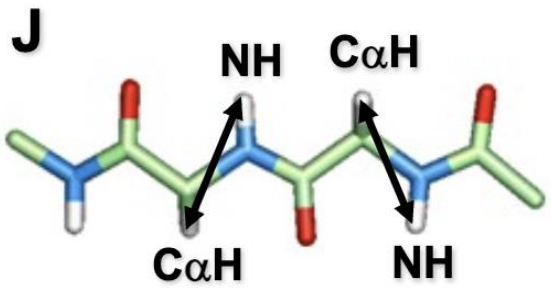
- The essence of NMR structure solving is to acquire as many connections between atoms (either through covalent bonds or through space) which then serve as restraints for model building



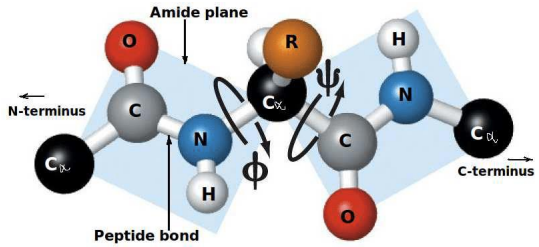
Atom proximity in 3D space



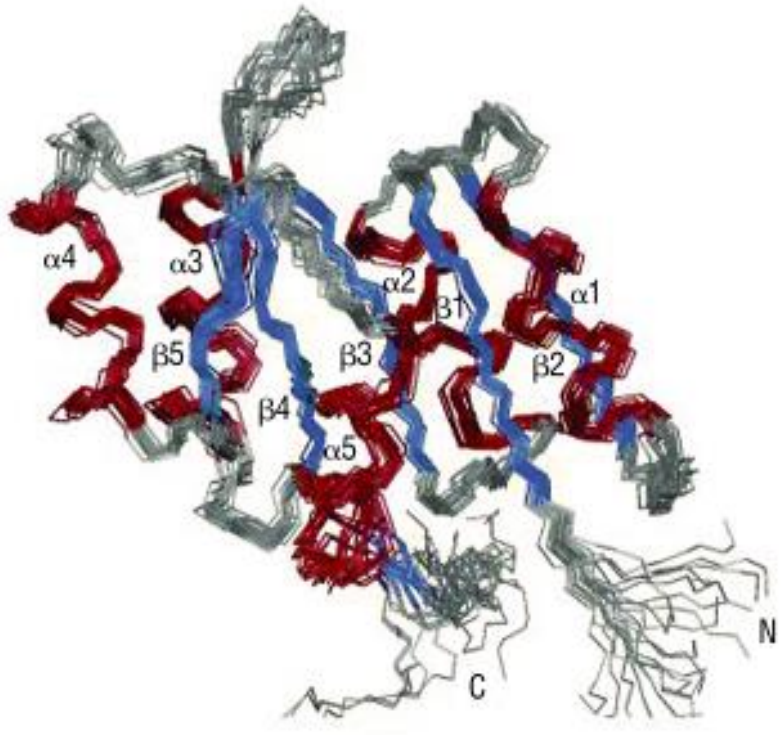
Chemical shifts (ppm)



Atoms connected by bonds



Geometric restraints:



Ensemble of structures that satisfy all the imposed restraints.

NMR Structure is an ensemble

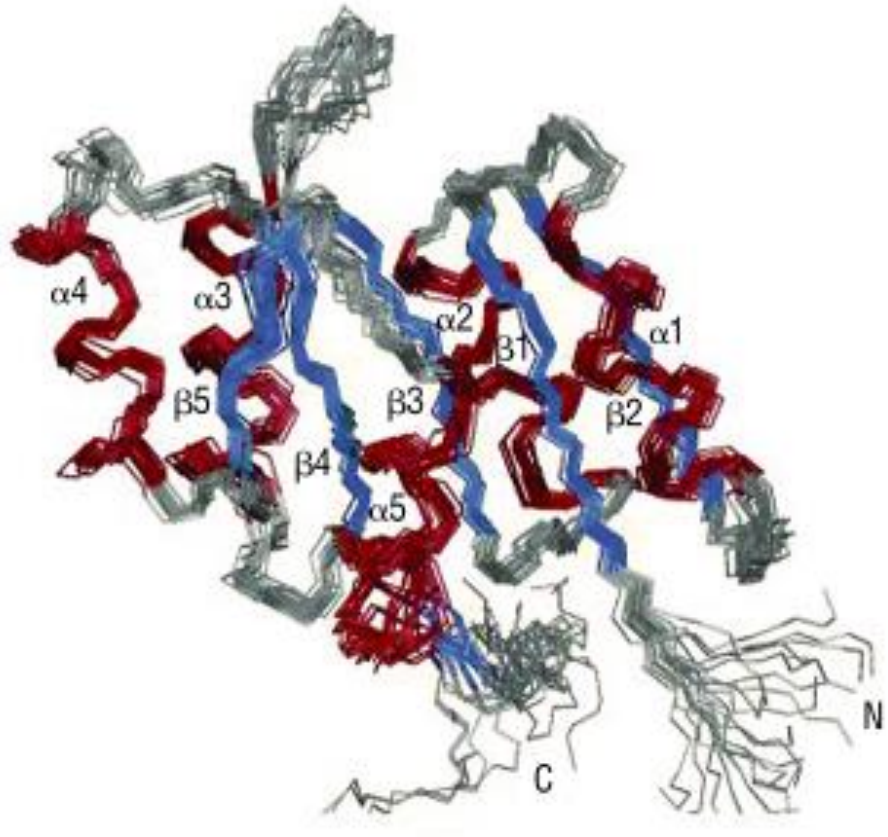
• An NMR structure is in reality a **model** based on the available restraints (e.g., recovered by restrained simulated annealing)

• In contrast to crystal structure, ensemble of 10-20 lowest energy structures

• Gives appreciation of rigidity and dynamics of structure

• Does not have a given resolution

• It is usually modelled with reported H atoms (in comparison to X-ray crystallography and cryoEM that usually do not have enough resolution to solve H atoms)



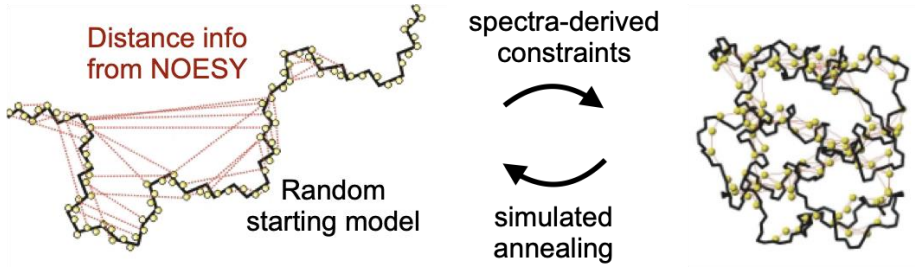
Nuclear Magnetic Resonance - Pros and Cons

- **Positive sides**
- Proteins studied in solution
- Applicable to many more questions than just the structure determination
- Computationally light
- Can be used to measure/approximate protein dynamics
- Hydrogens are “visible” by this method

- **Negative sides**
- Limited to studies of biomolecules with MW < 50kDa
- Complex procedure to assign resonances to atoms
- Requires isotopic labeling (^{13}C, ^{15}N) which is expensive and not very practical

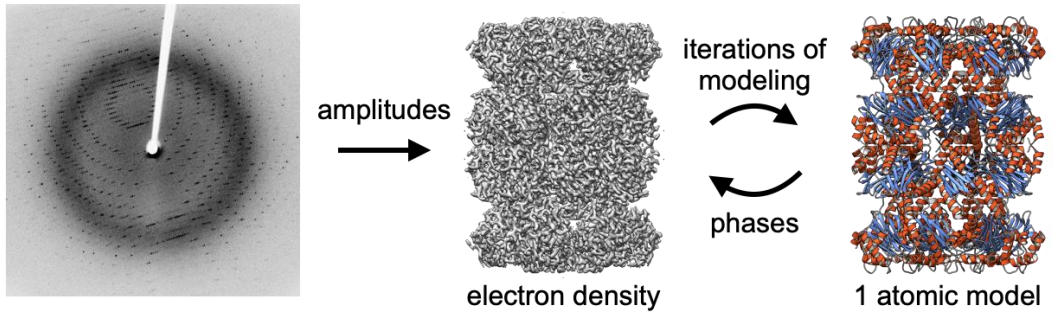
Methods for determining biomolecular structures

NMR



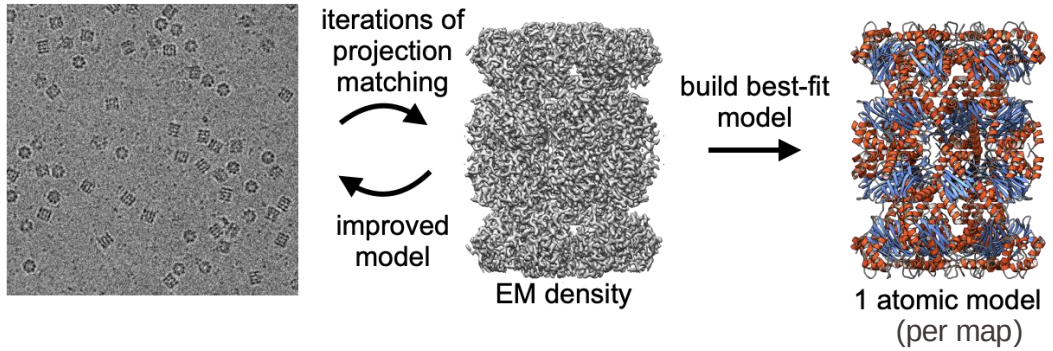
- **Versatile tool for studying protein structure and dynamics**
- Computationally light
- Full structural analysis limited to smaller proteins (<50kDa)
- Requires isotopic labeling
- Results in model ensemble

X-ray



- Gold-standard method for solving protein structures
- **Not limited in size or achievable resolution**
- Computationally light
- Requires highly homogenous, crystallizable sample
- Requires screening of crystallization conditions
- Phase problem
- Results in a single model

CryoEM



- Versatile tool for studying protein assembly, structure, dynamics
- Limited to proteins >40kDa
- No requirement for protein labeling
- **Does not require homogenous samples**
- Grid preparation procedure requires screening
- Real space imaging – no phase problem
- Can be used to study protein dynamics
- Can be expanded to larger assemblies (e.g., viruses and cells)
- Results in 1 or more models per dataset
- Computationally heavy (TBs of data + requirement for GPU processing)

Additional learning resources

Extra material:

- Grant Jensen course on EM
https://cryo-em-course.caltech.edu/

- Lectures and courses on X-ray crystallography:
https://www.youtube.com/watch?v=mDbmfyOGLIM

- Lectures and Courses on NMR:
https://www.youtube.com/watch?v=_rjYI8XxEOs
https://www.khanacademy.org/science/