Macromolecular crystallography

Pavel Plevka

- Importance of crystallography
- Development of crystallography
- Waves, radiation, and diffraction
- Phase problem
- Macromolecular structures

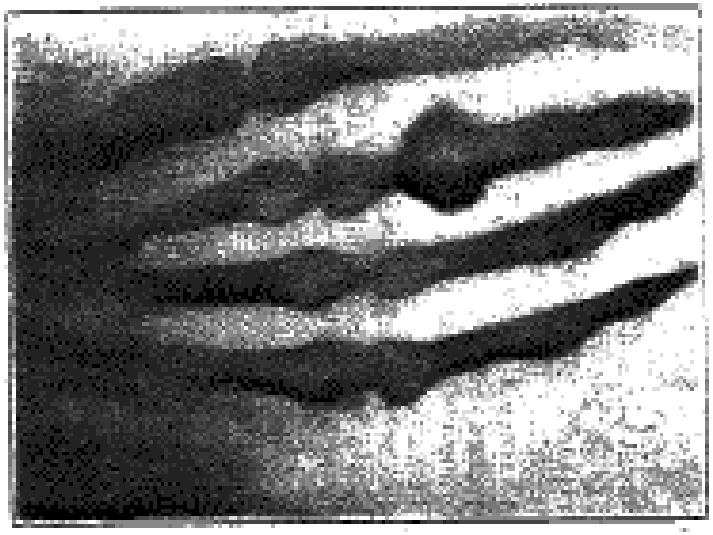
X-ray crystallography

- First method to determine structure of molecules with atomic resolution
- As of September 17, 2013 there were more than 70,000 structures determined by protein crystallography in Protein Data Bank
- Macromolecular structures are crucial for our understanding of life at the molecular level
- 28 Nobel prizes

WILHELM CONRAD RÖNTGEN (1845-1923)

- **1901 Nobel Laureate in Physics**

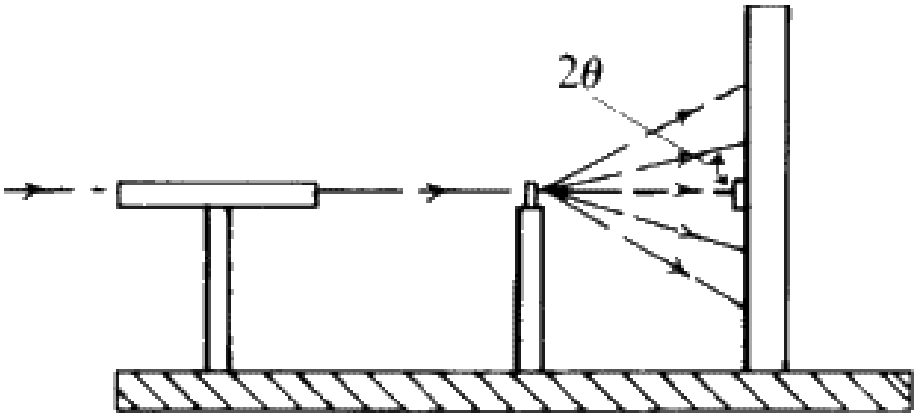
  in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him.

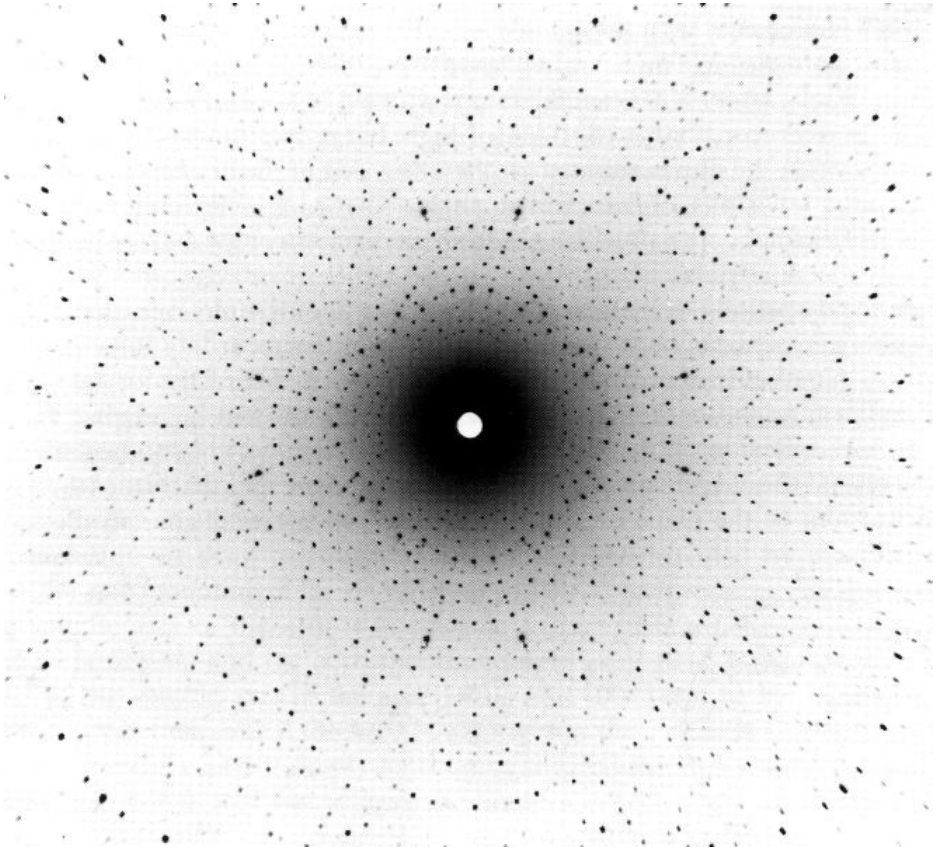


MAX VON LAUE (1879-1960)

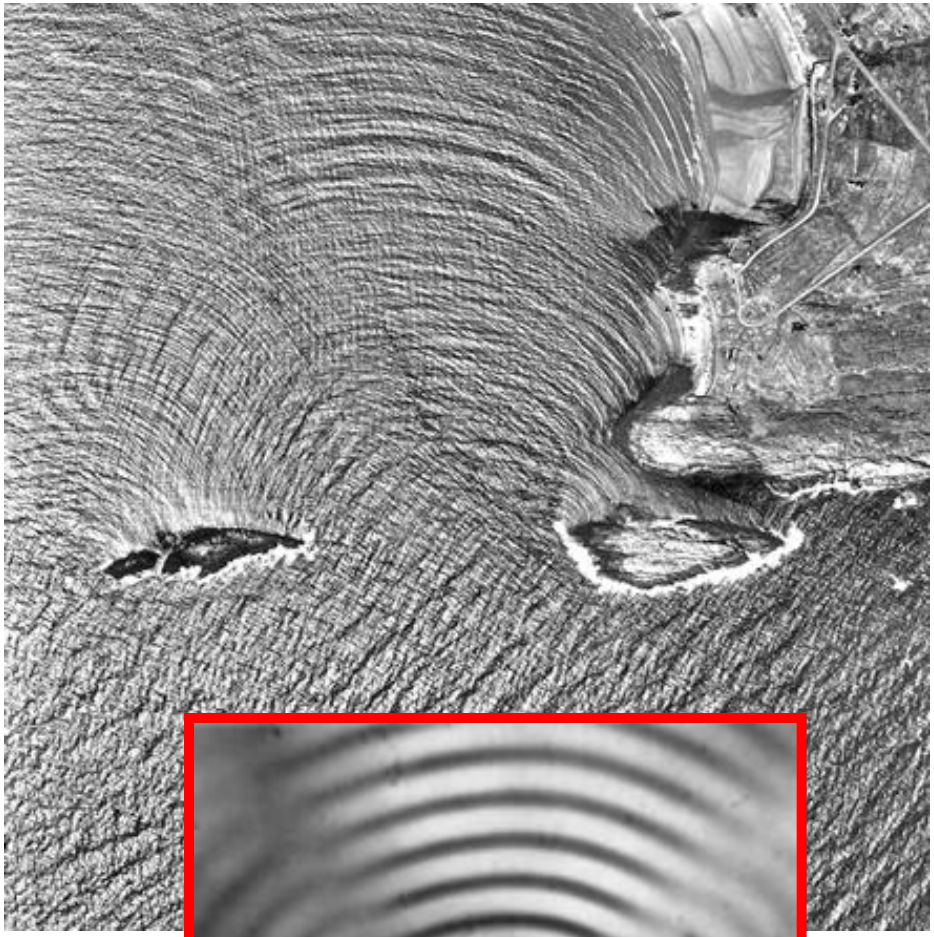
- **1914 Nobel Laureate in Physics**

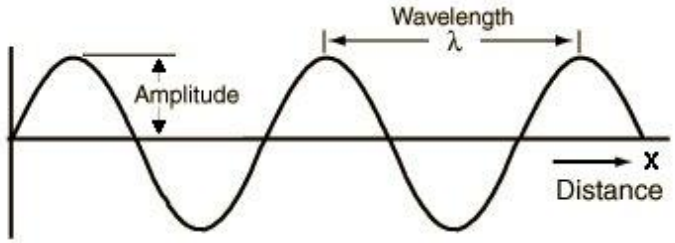
  for his discovery of the diffraction of X-rays by crystals



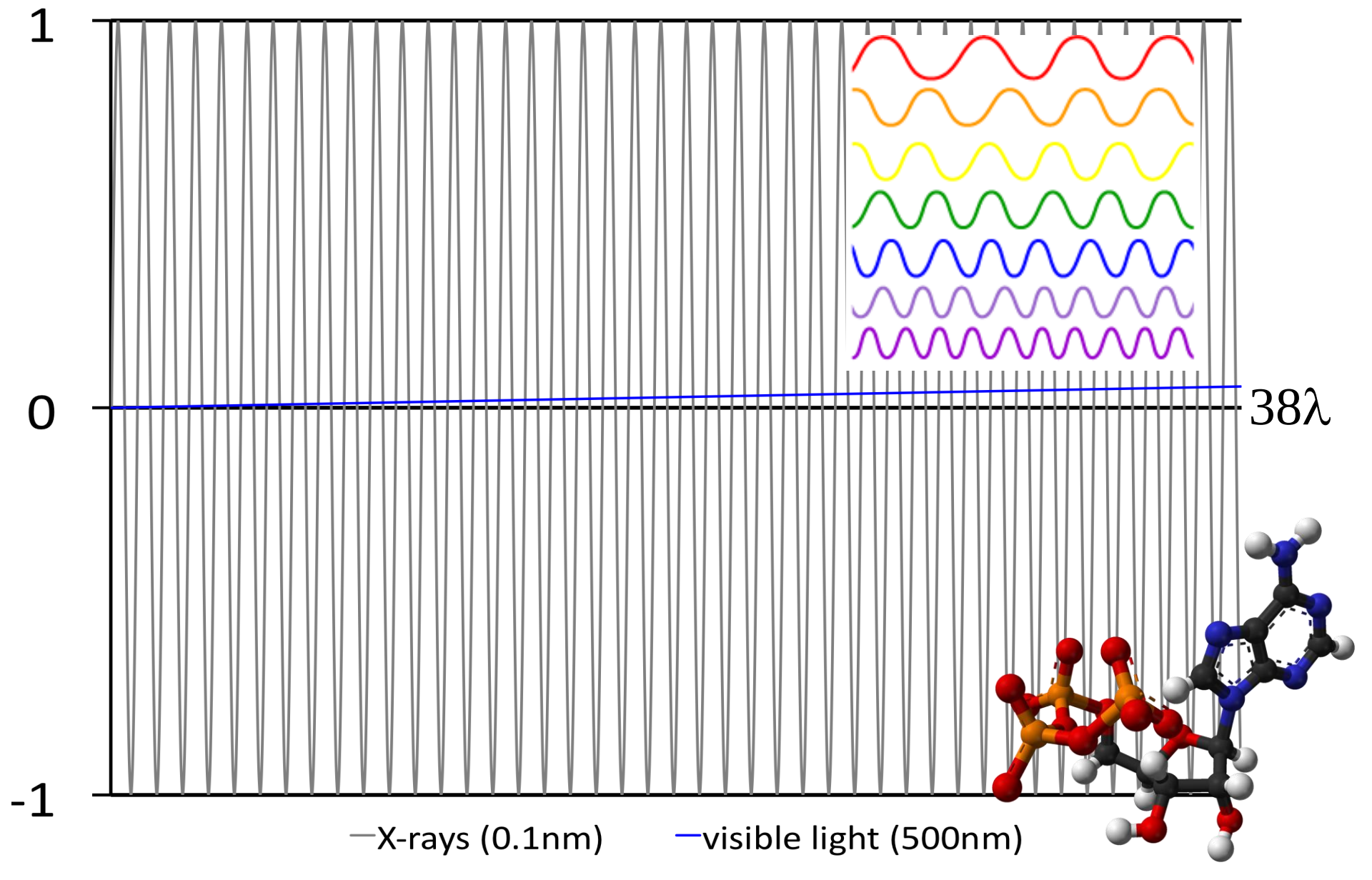


Wavelength and diffraction





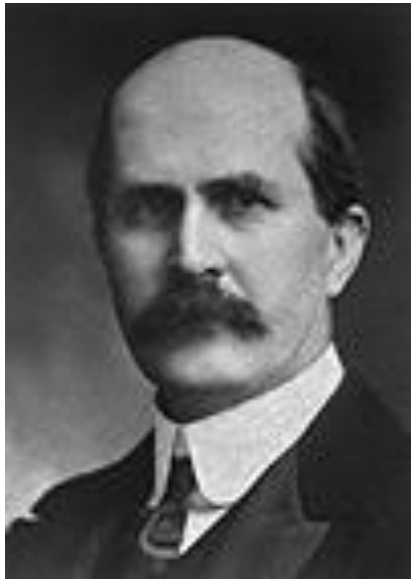
Wavelength comparison of X-rays and visible light



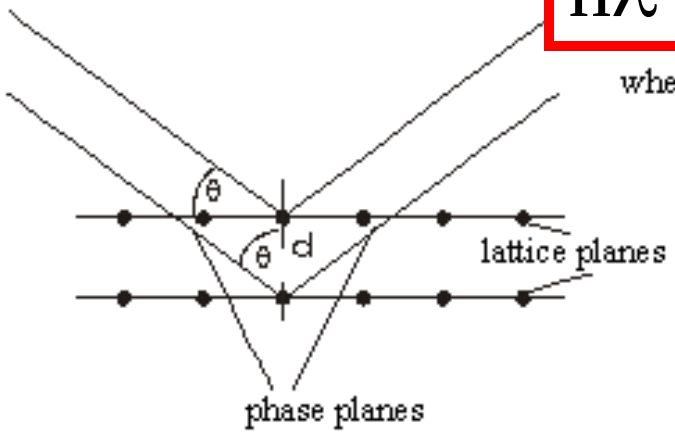
SIR WILLIAM HENRY BRAGG (1862-1942)
SIR WILLIAM LAWRENCE BRAGG (1890-1971)

- **1915 Nobel Laureates in Physics**

  for their services in the analysis of crystal structure by means of X-rays.

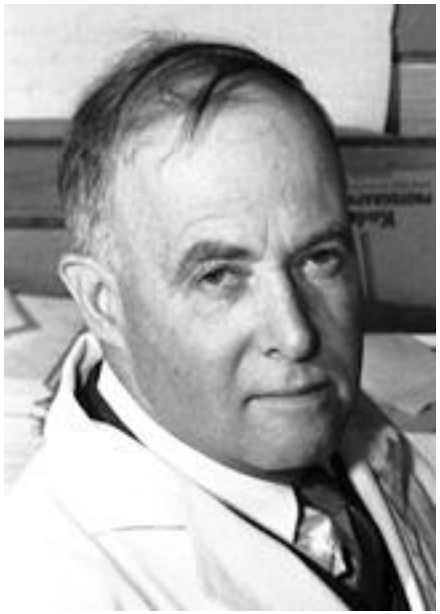


$$n\lambda = 2d \sin\theta$$



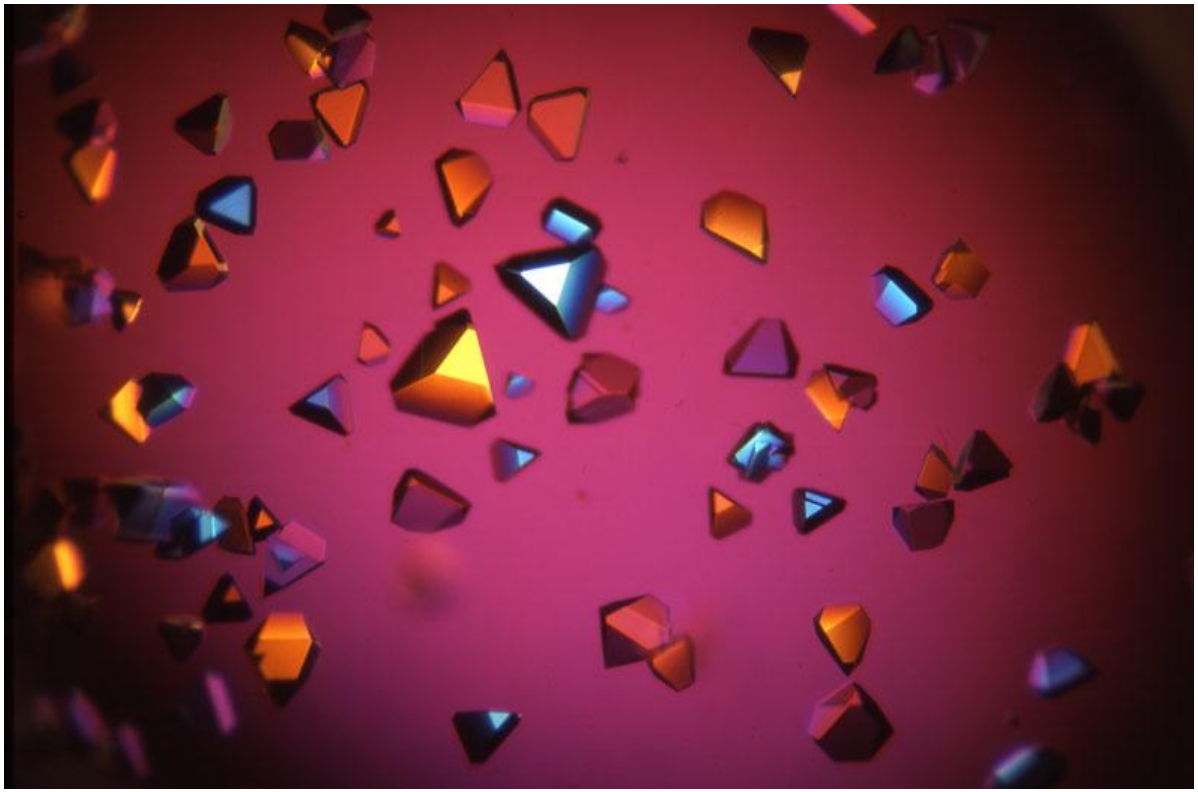
where
d: lattice spacing
Θ: incident angle
λ: wavelength
m: integral multiple of wavelength

James Batcheller Sumner (1879-1960)



- **1946 Nobel Laureate in Chemistry**

  for his discovery that enzymes can be crystallized

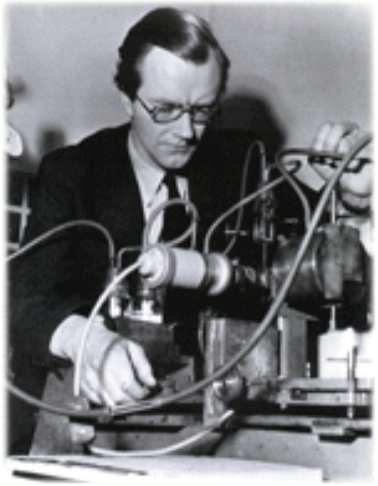


FRANCIS HARRY COMPTON CRICK (1916~2004)
JAMES DEWEY WATSON (1928~)
MAURICE HUGH FREDERICK WILKINS (1916~2004)

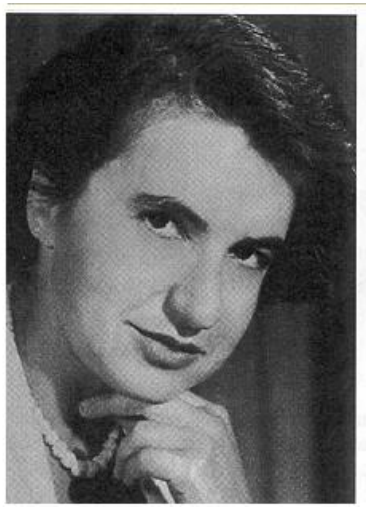
- **1962 Nobel Laureates in Physiology and Medicine**

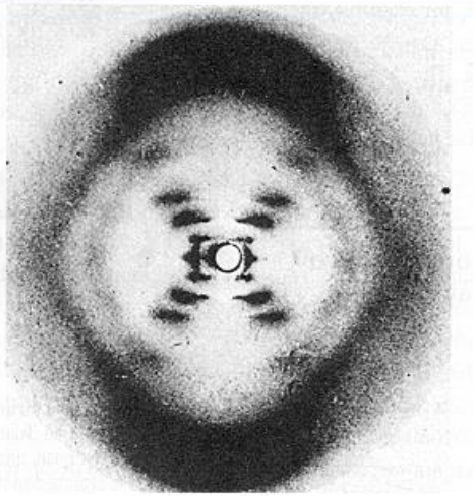
for their discoveries concerning the molecular structure of nuclear acids and its significance for information transfer in living material.

James Watson and Francis Crick



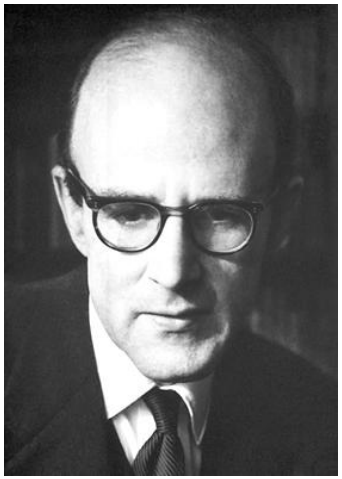
Maurice Wilkins

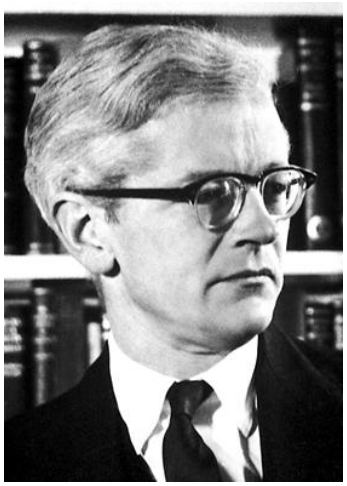




Rosalind Franklin

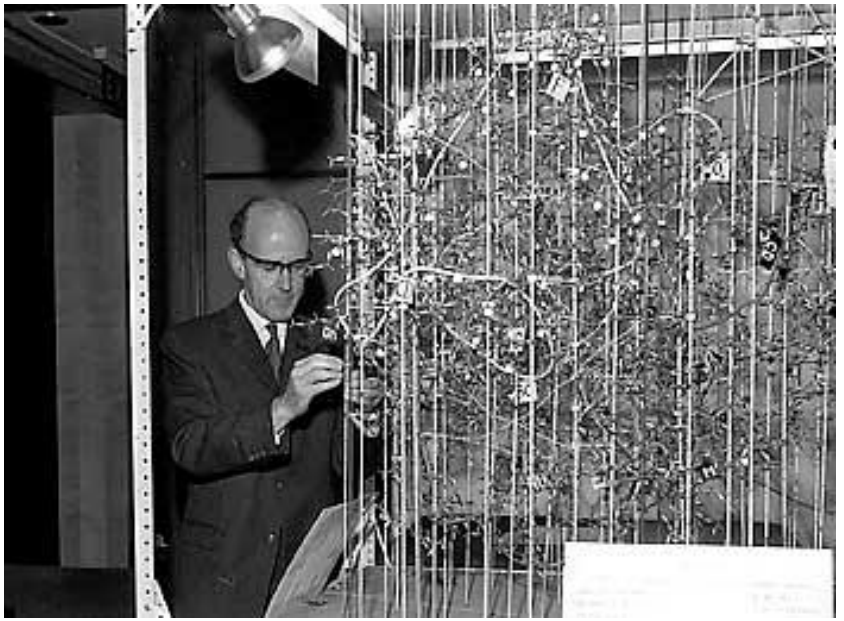
Max Ferdinand Perutz (1914 – 2002)
John Cowdery Kendrew (1917 – 1997)



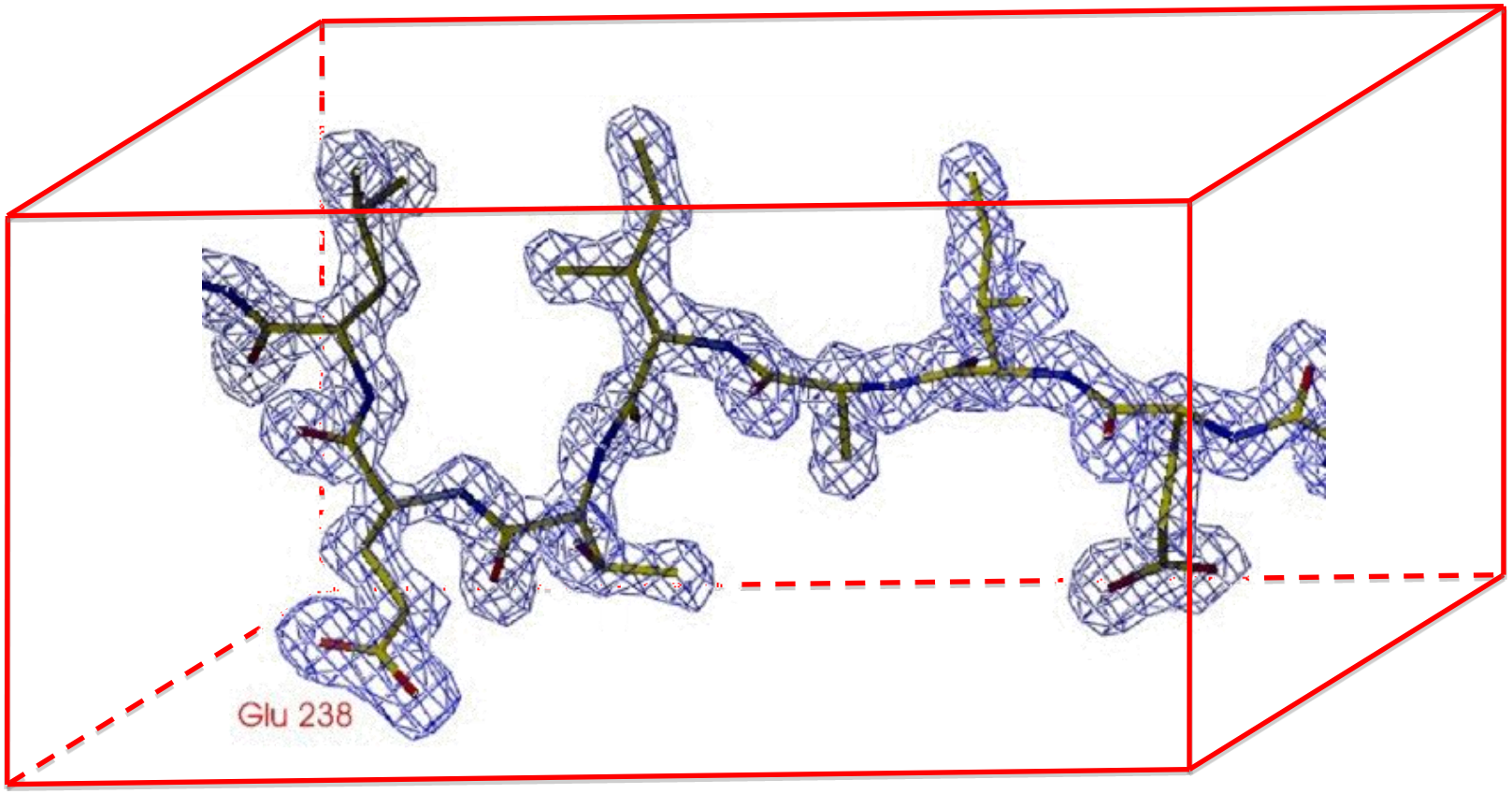


- **1962 Nobel Laureate in Physics**

  for their studies of the structures of globular proteins

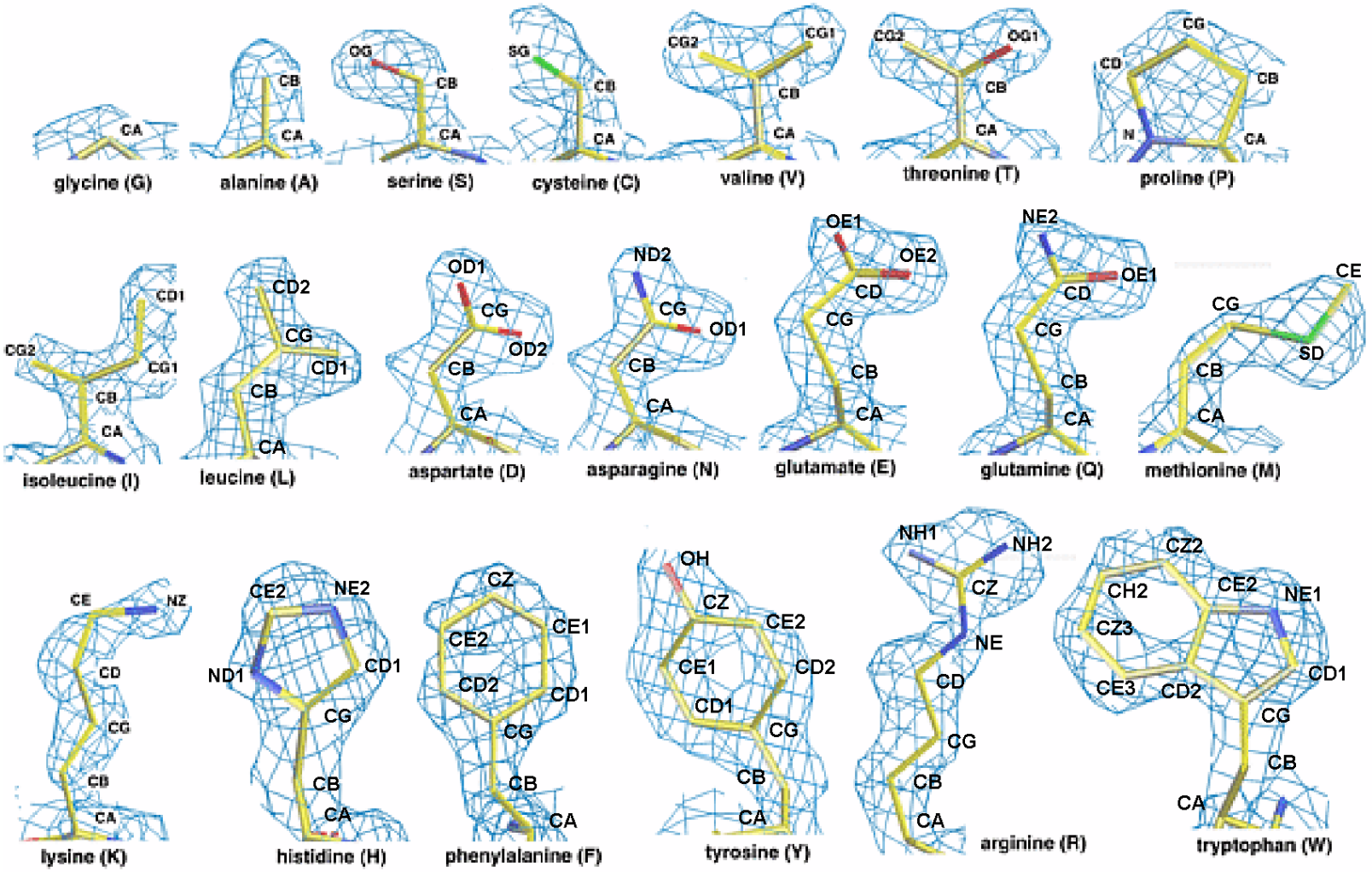


Information from X-ray diffraction experiment



$$\rho(x\,y\,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F(h\,k\,l)\right|\exp\left[-2\pi i(hx+ky+lz)+i\alpha(h\,k\,l)\right]$$

Representative electron density for amino acid side chains

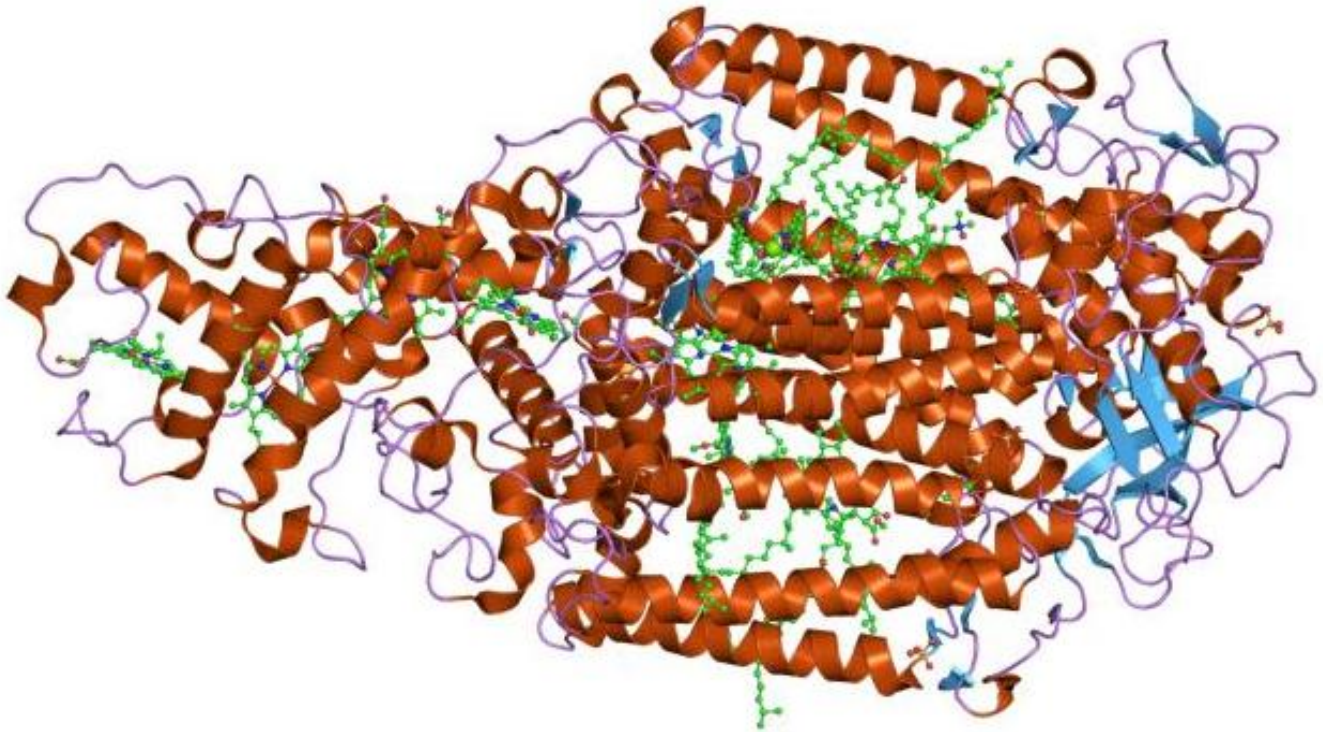


Electron density maps calculated at 1.5 Angstrom resolution.

Johann Deisenhofer (1943)
Robert Huber (1937)
Hartmut Michel (1948)

- **1988 Nobel Laureates in Chemistry**

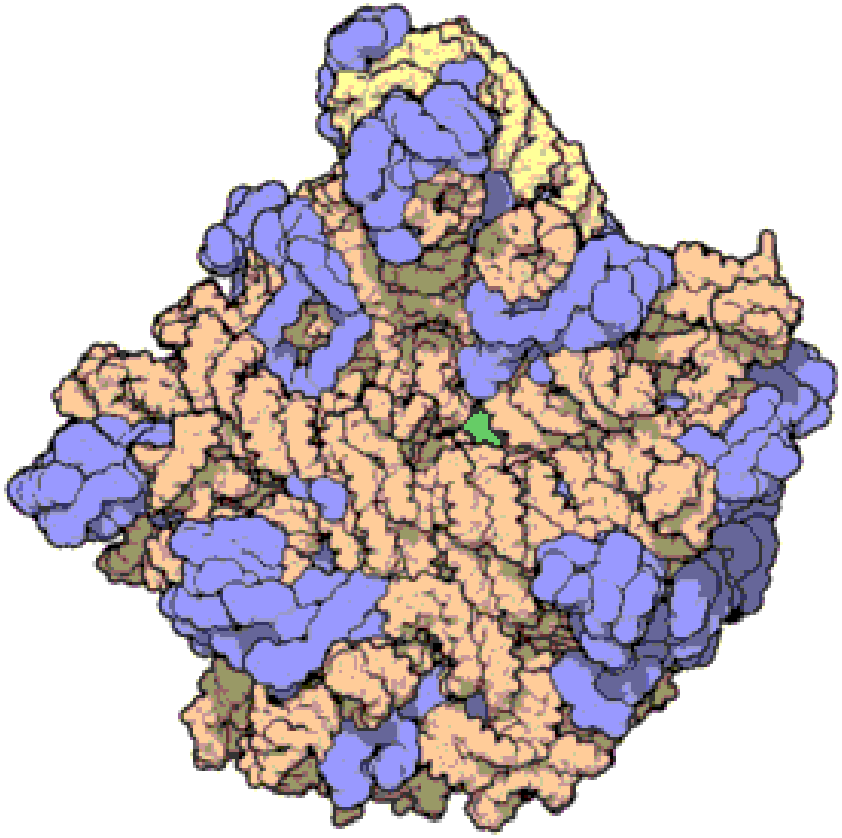
for the determination of the three-dimensional structure of a photosynthetic reaction centre

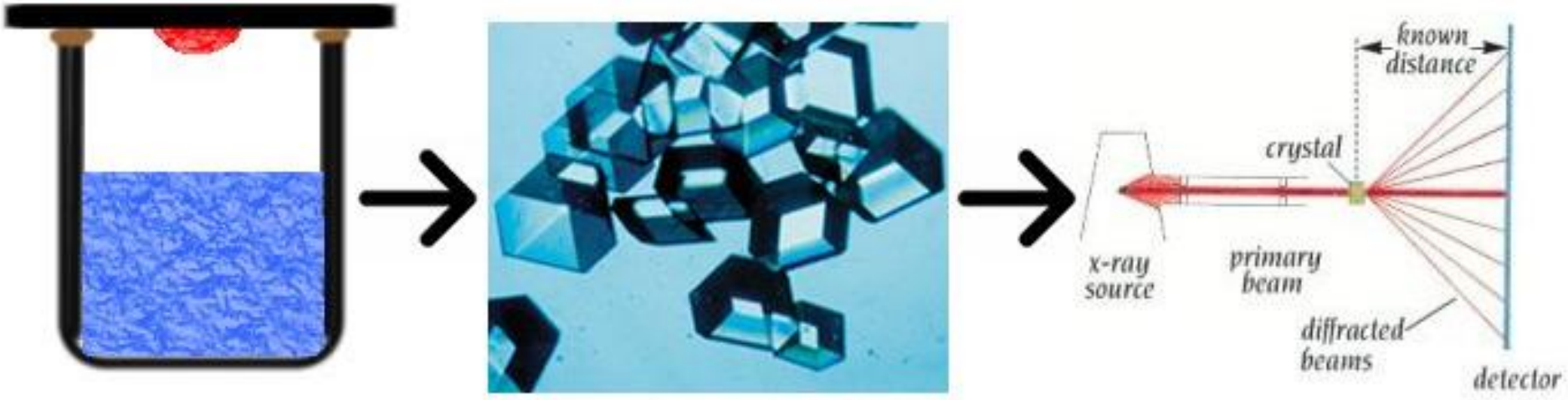


Venkatraman Ramakrishnan (1952)
Thomas A. Steitz (1940)
Ada E. Yonath (1939)

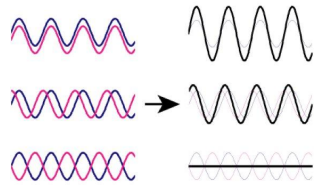
- **2009 Nobel Laureates in Chemistry**

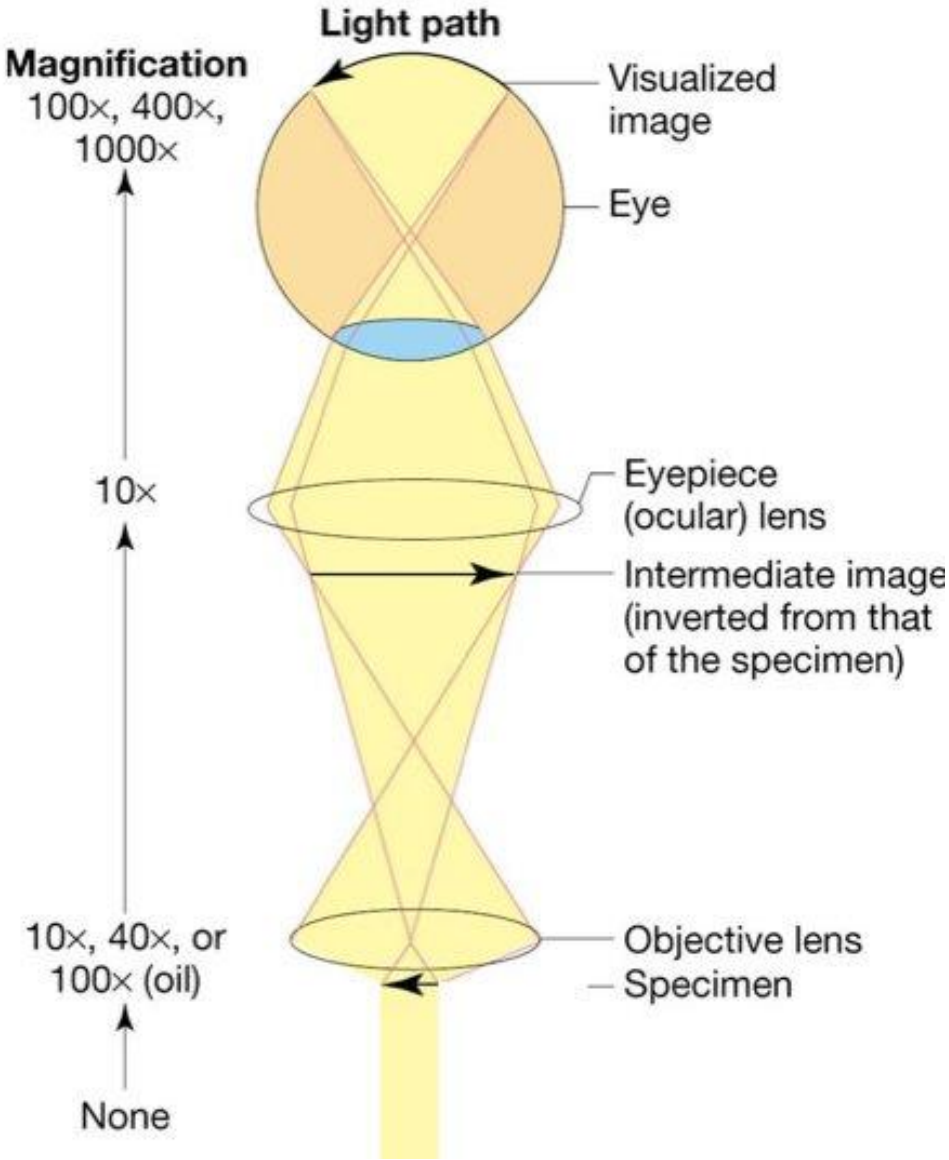
or studies of the structure and function of the ribosome

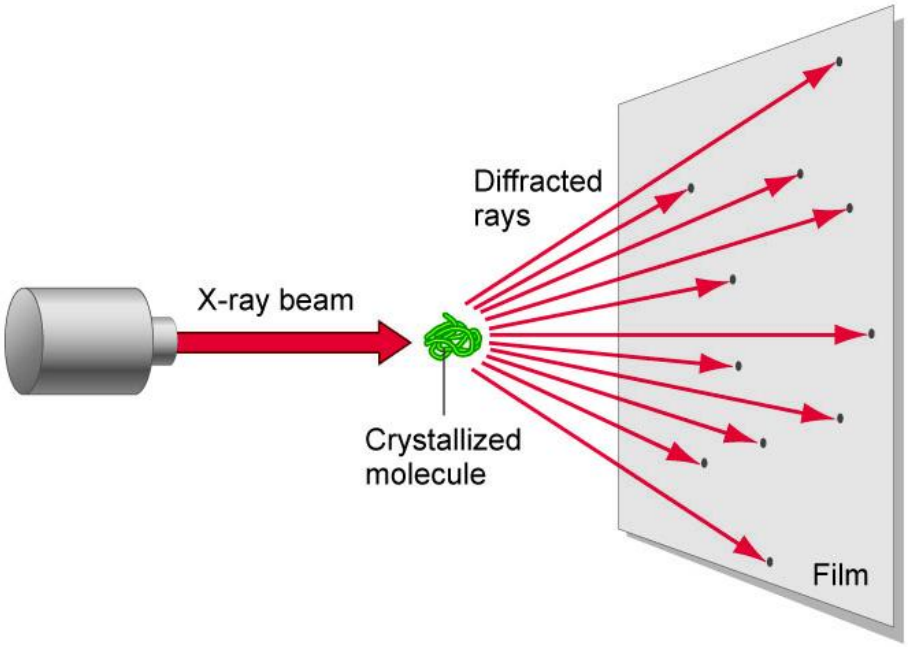


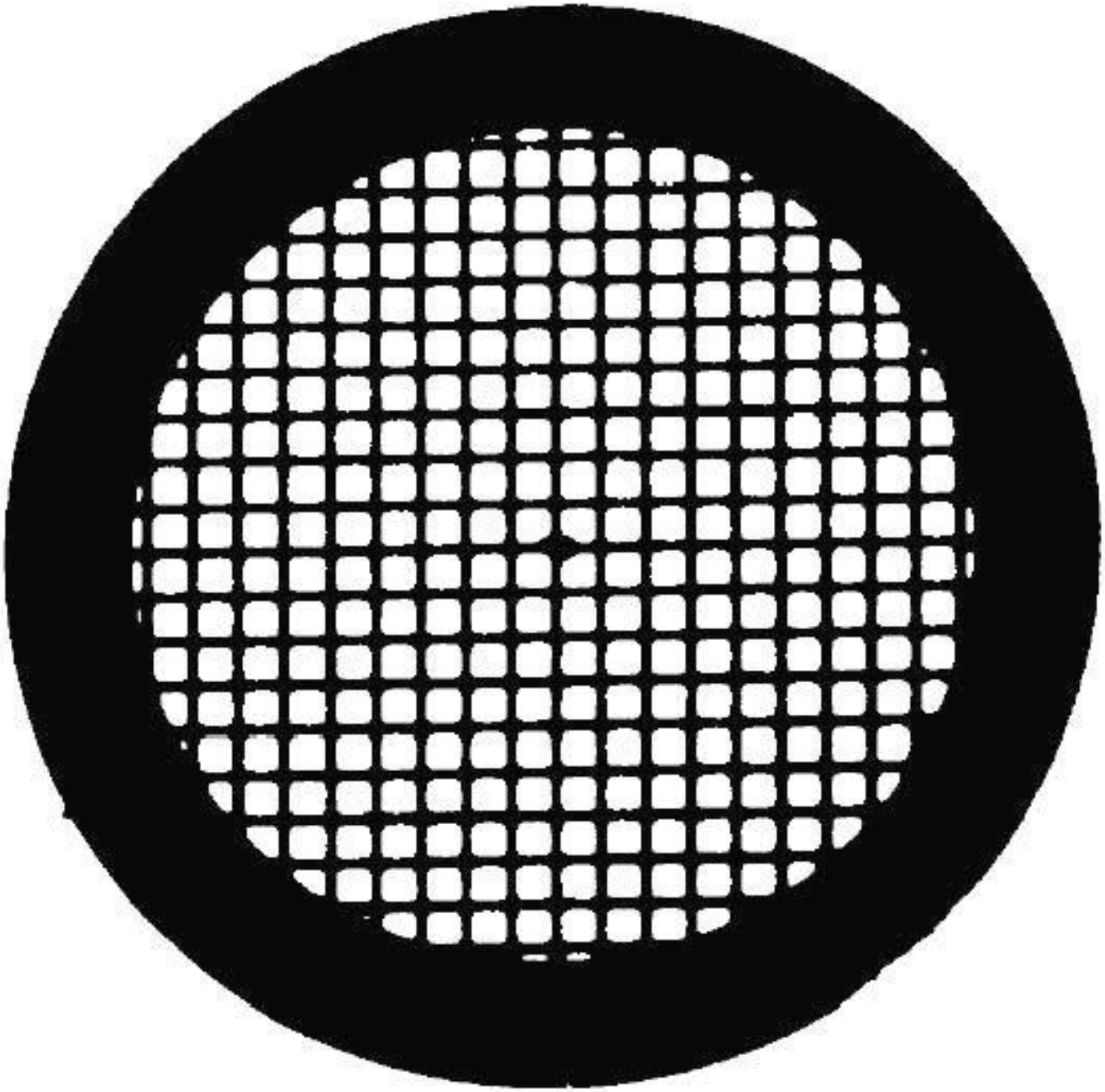

known
distance
crystal
x-ray
source
primary
beam
diffracted
beams
detector

Comparison of microscope and diffraction

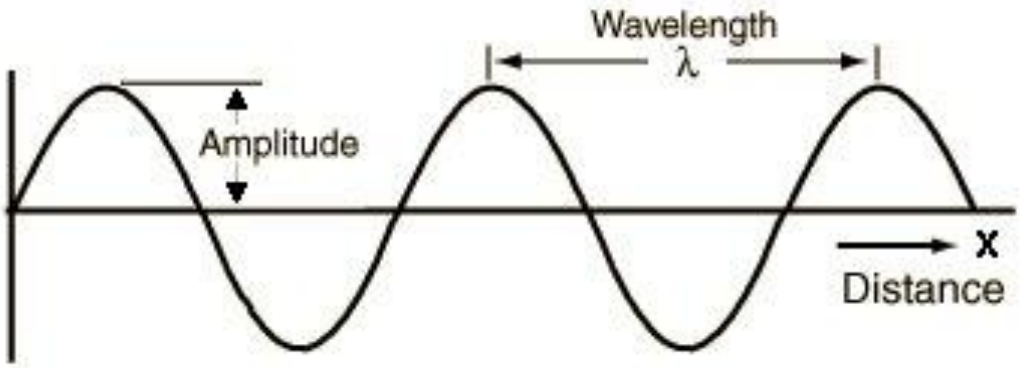


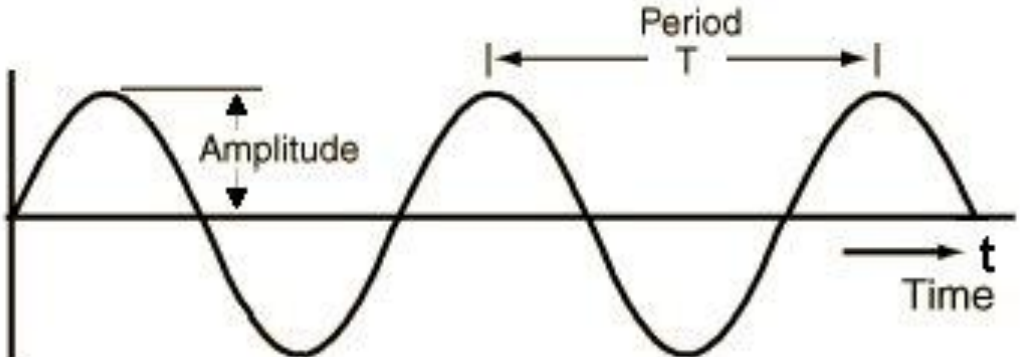




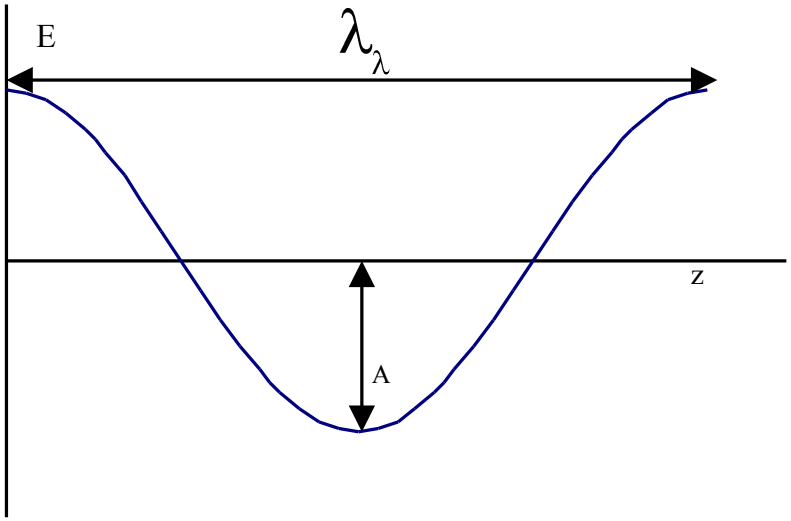


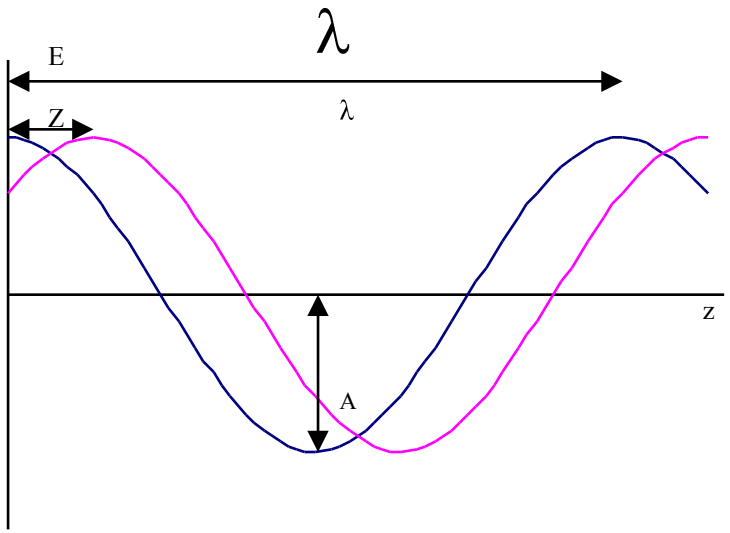
Waves and Radiation





Description of lectromagnetic waves





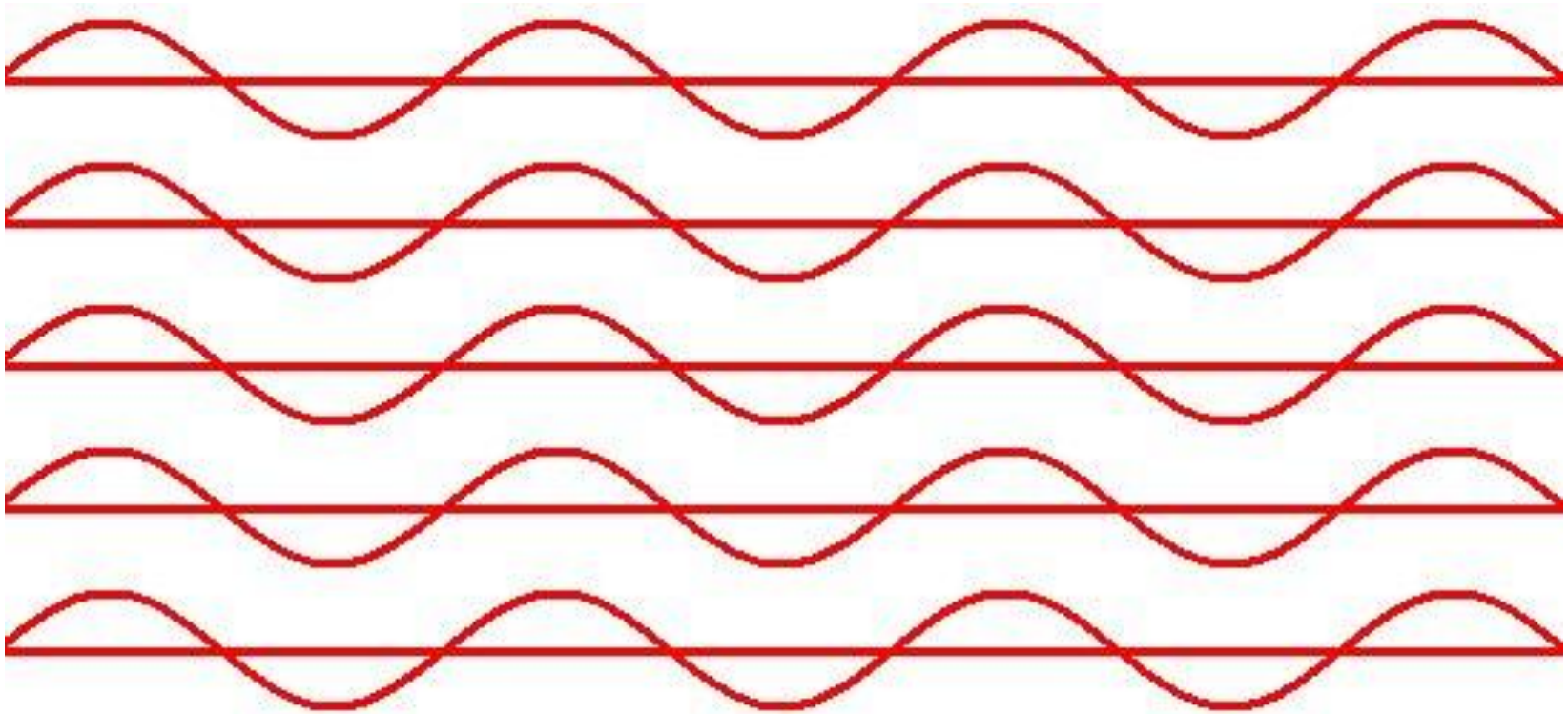
$E = A \cos 2\pi z/\lambda$

$E = A \cos (\alpha+2\pi z/\lambda)$

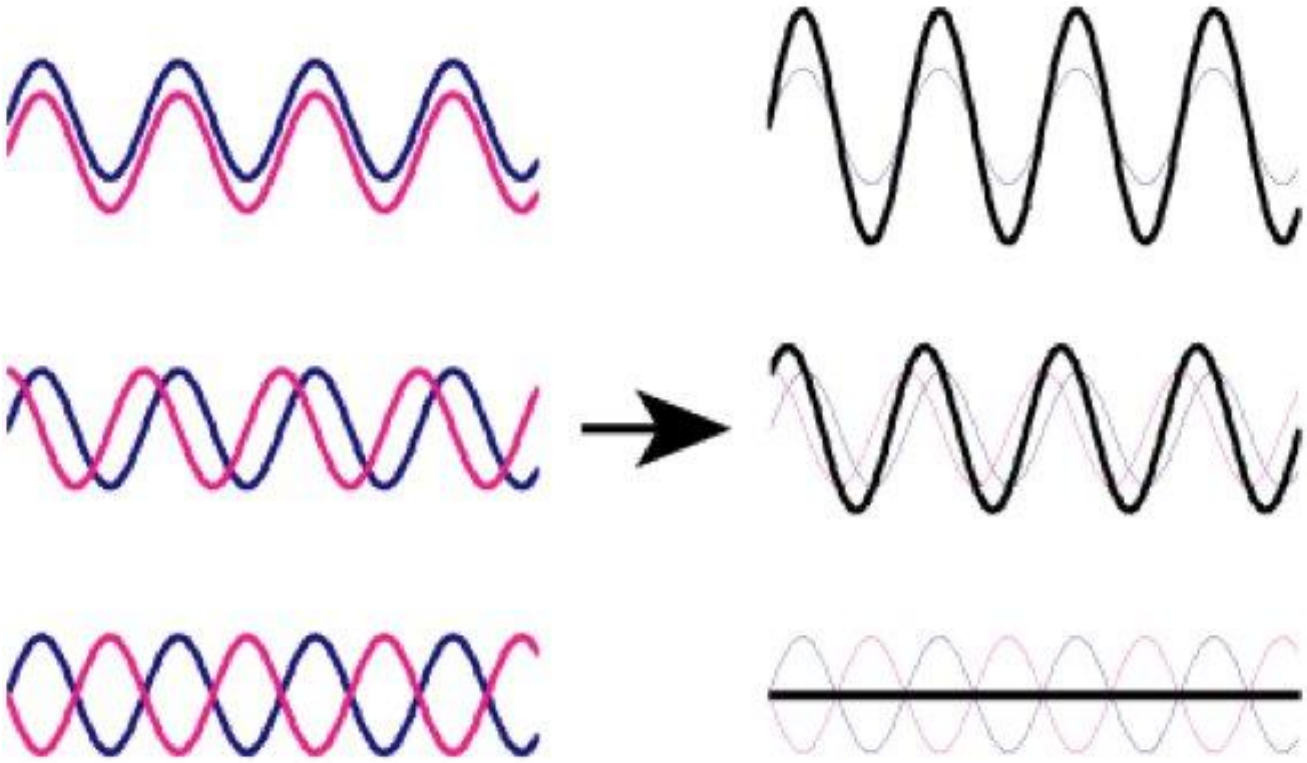
- E- electric field strength
- A- amplitude
- λ - wavelenght

- z - position along beam path
- α - phase

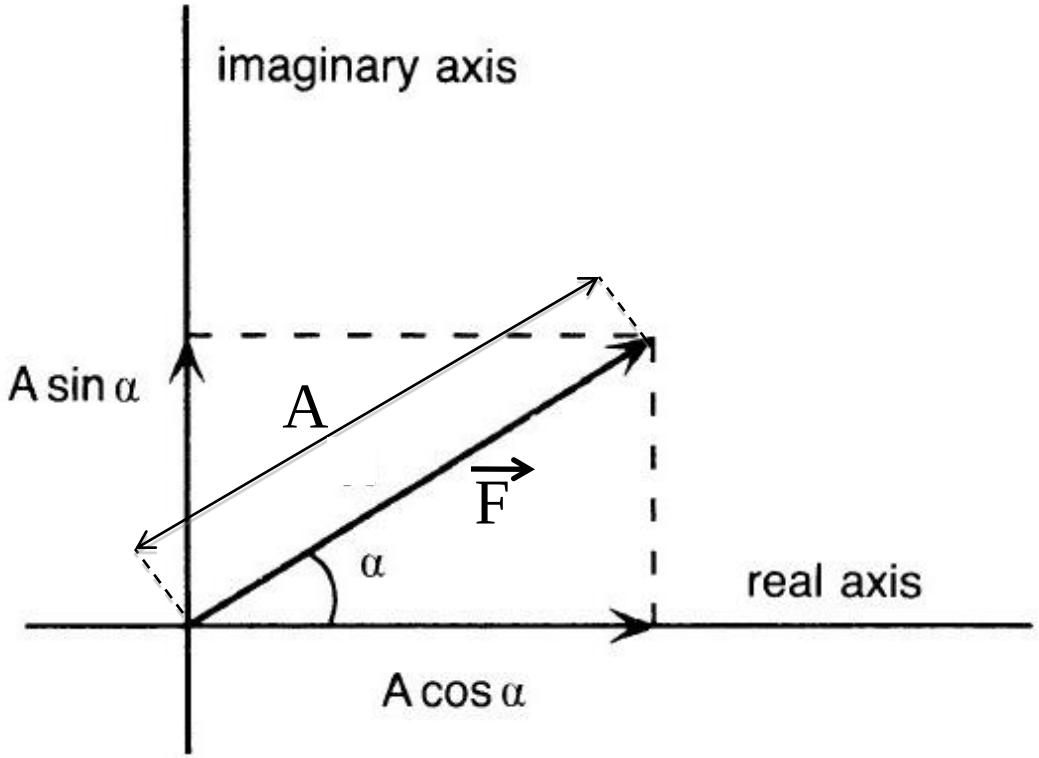
Coherent beam



Addition of waves



Wave as a vector

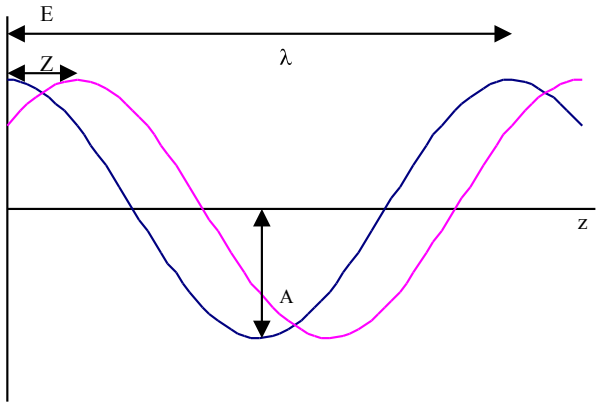


A- wave amplitude

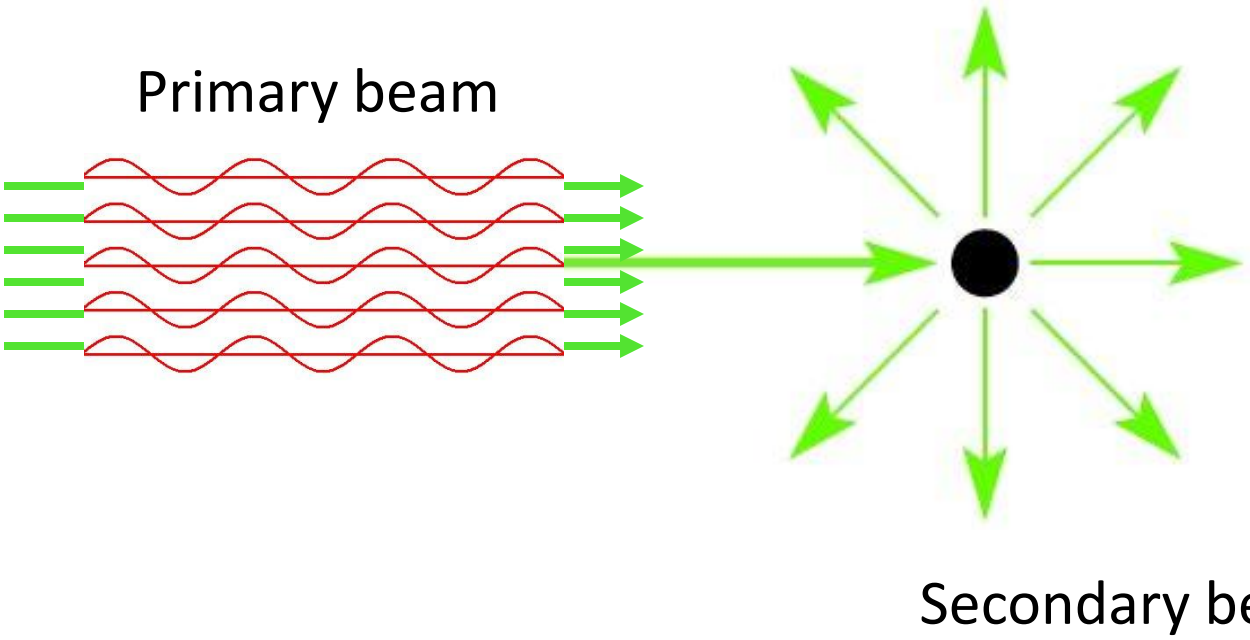
α- wave phase

$\vec{F}=A\cos\alpha+iA\sin\alpha$

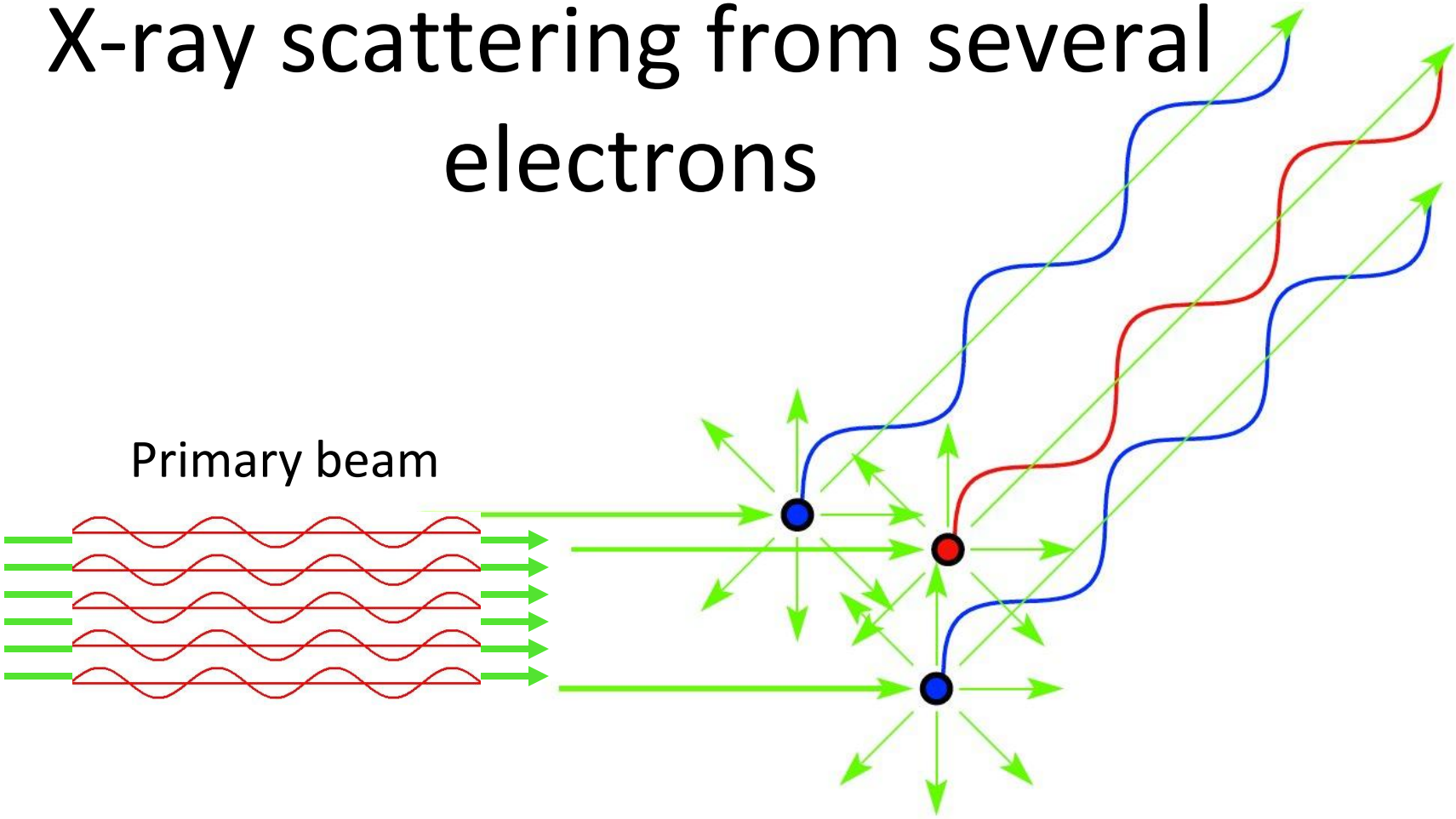
$\vec{F}=\exp(i\alpha)$



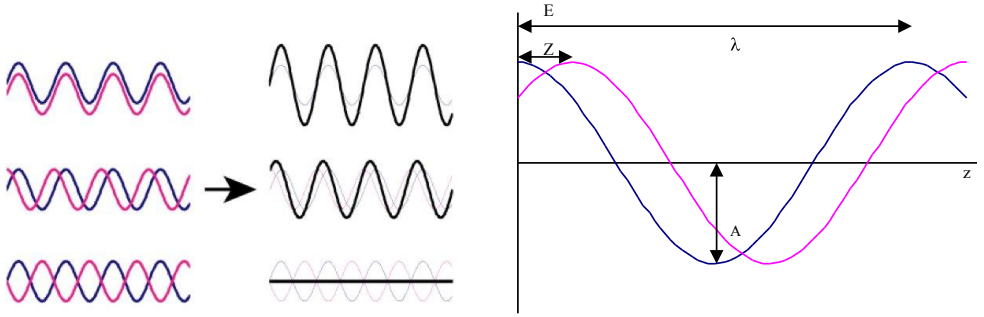
X-rays scatter from electrons in all directions



X-ray scattering from several electrons

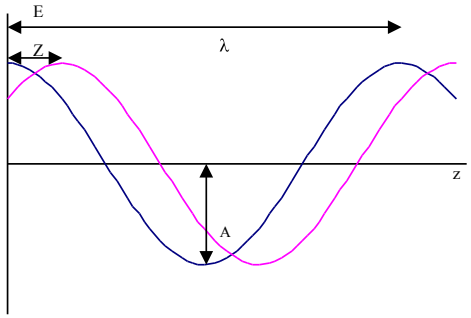


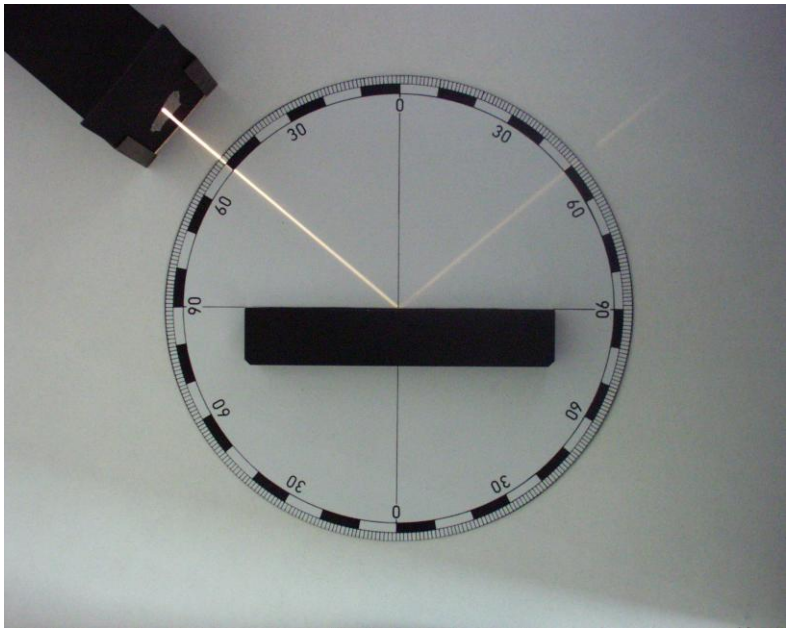
When do electrons scatter “in phase” – waves add constructively?

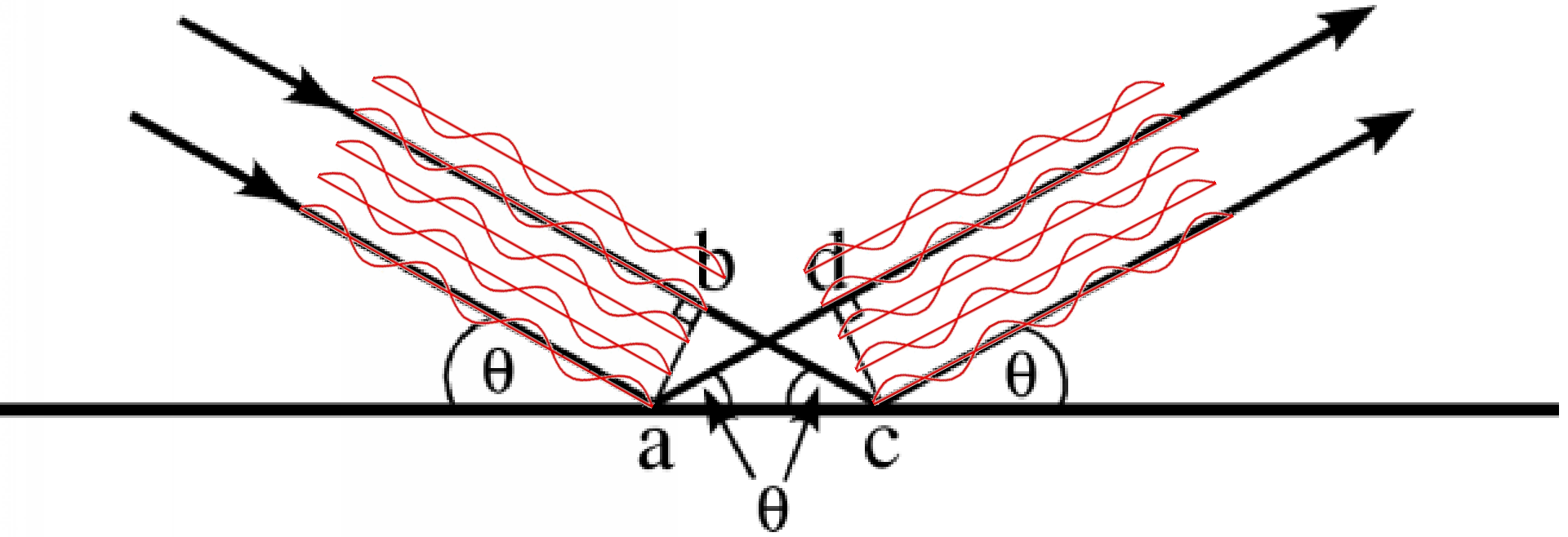


- Scattering from a single molecule is weak
- If molecules are all oriented in the same way, the scattering from individual molecules will add in certain directions
  - Which directions?

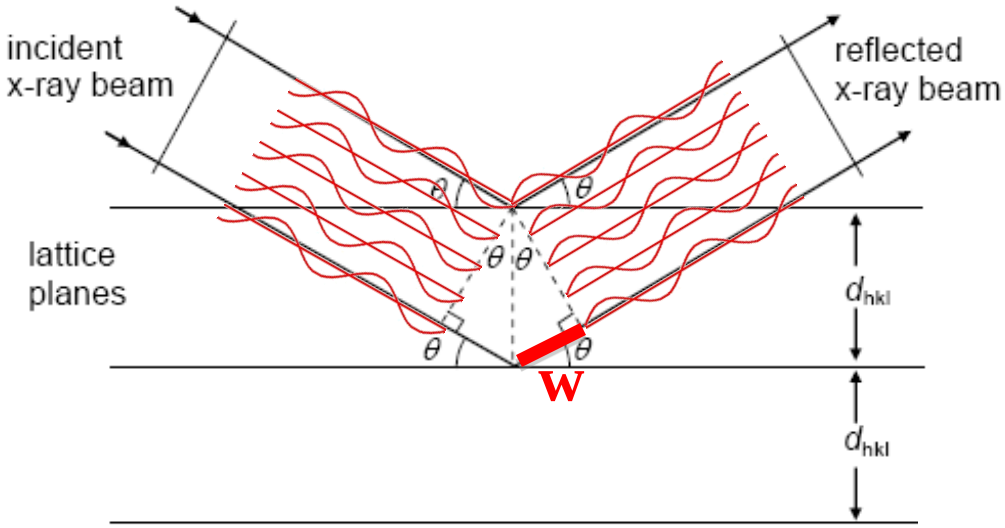
There is no path and PHASE DIFFERENCE when rays reflect from a plane





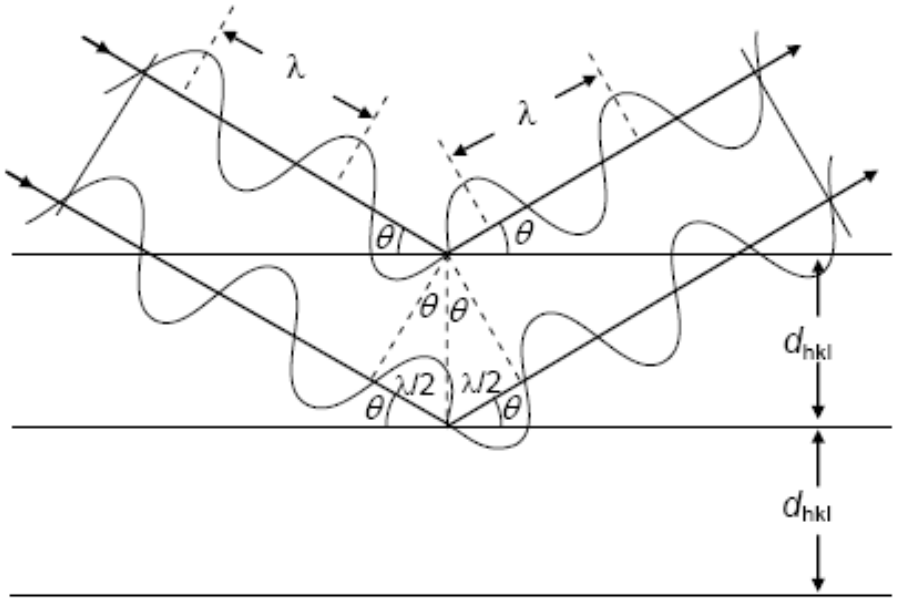


There is NO PHASE DIFFERENCE if the path differences are equal to prime number multiplies of wavelength (λ)



Bragg's law:

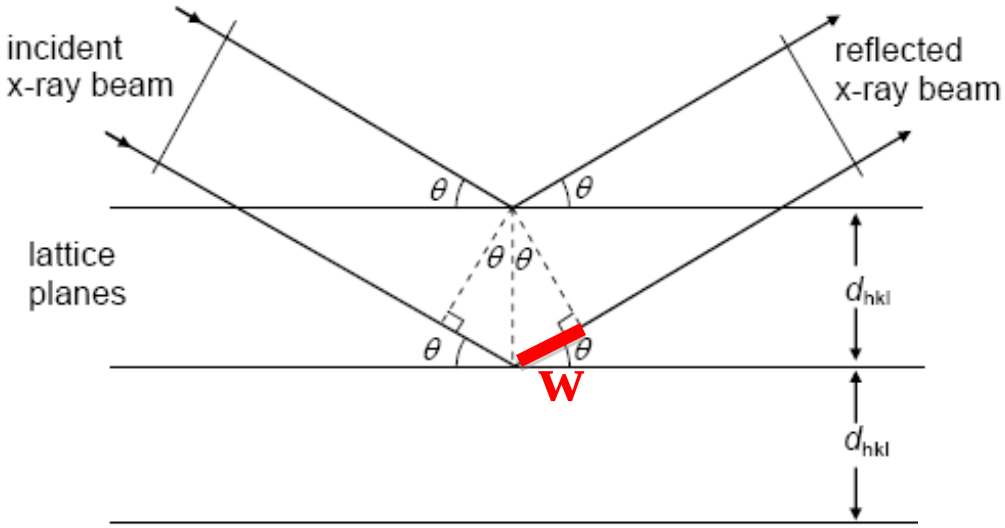
$$n\lambda = 2d \sin\theta$$



$\sin\theta = \mathbf{w}/d$

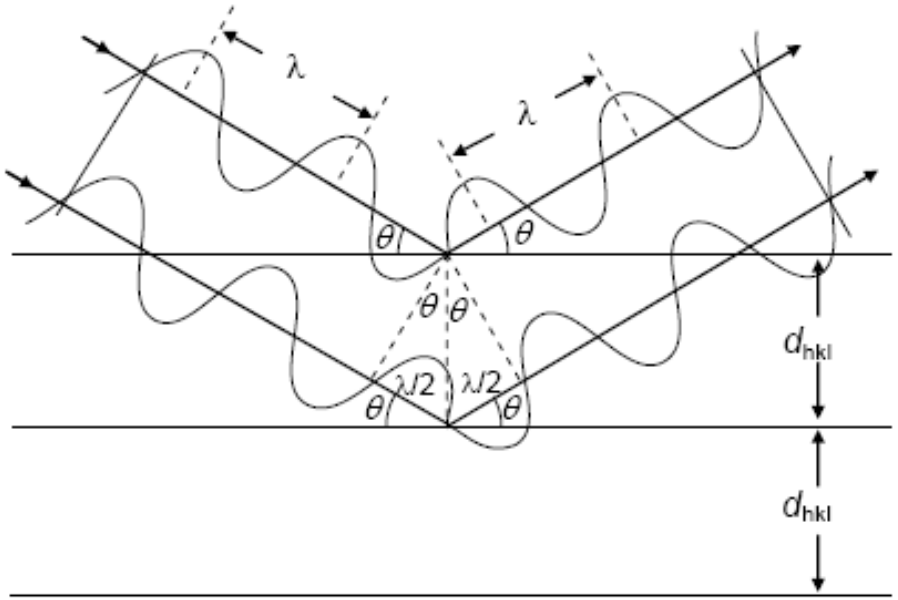
$2\mathbf{w} = n\lambda$

There is NO PHASE DIFFERENCE if the path differences are equal to prime number multiplies of wavelength (λ)



Bragg's law:

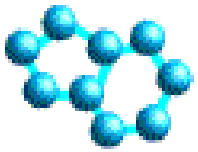
$$n\lambda = 2d \sin\theta$$



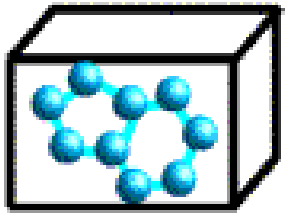
$\sin\theta = \mathbf{w}/d$

$2\mathbf{w} = n\lambda$

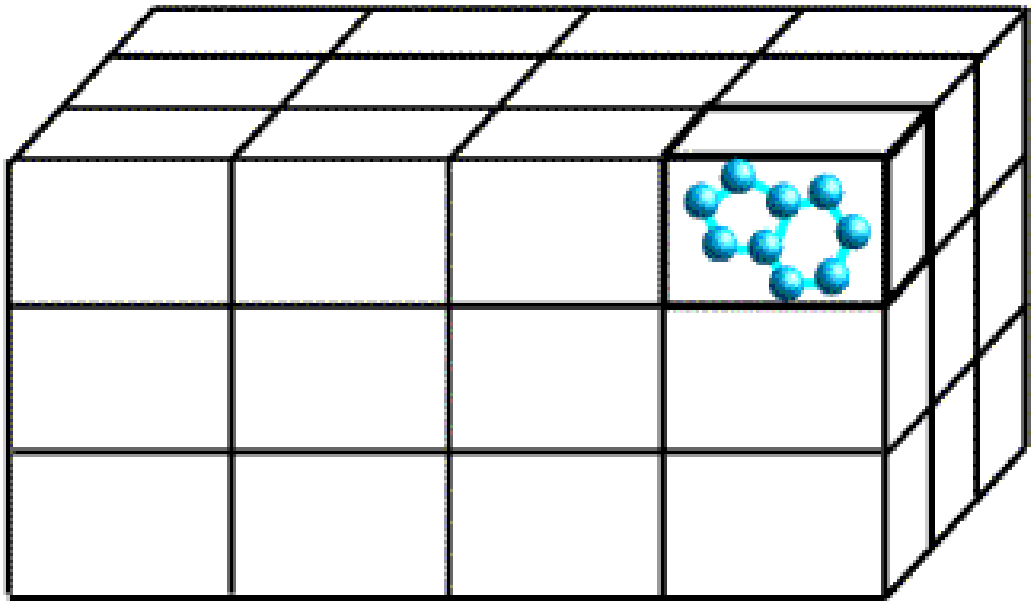
molecule

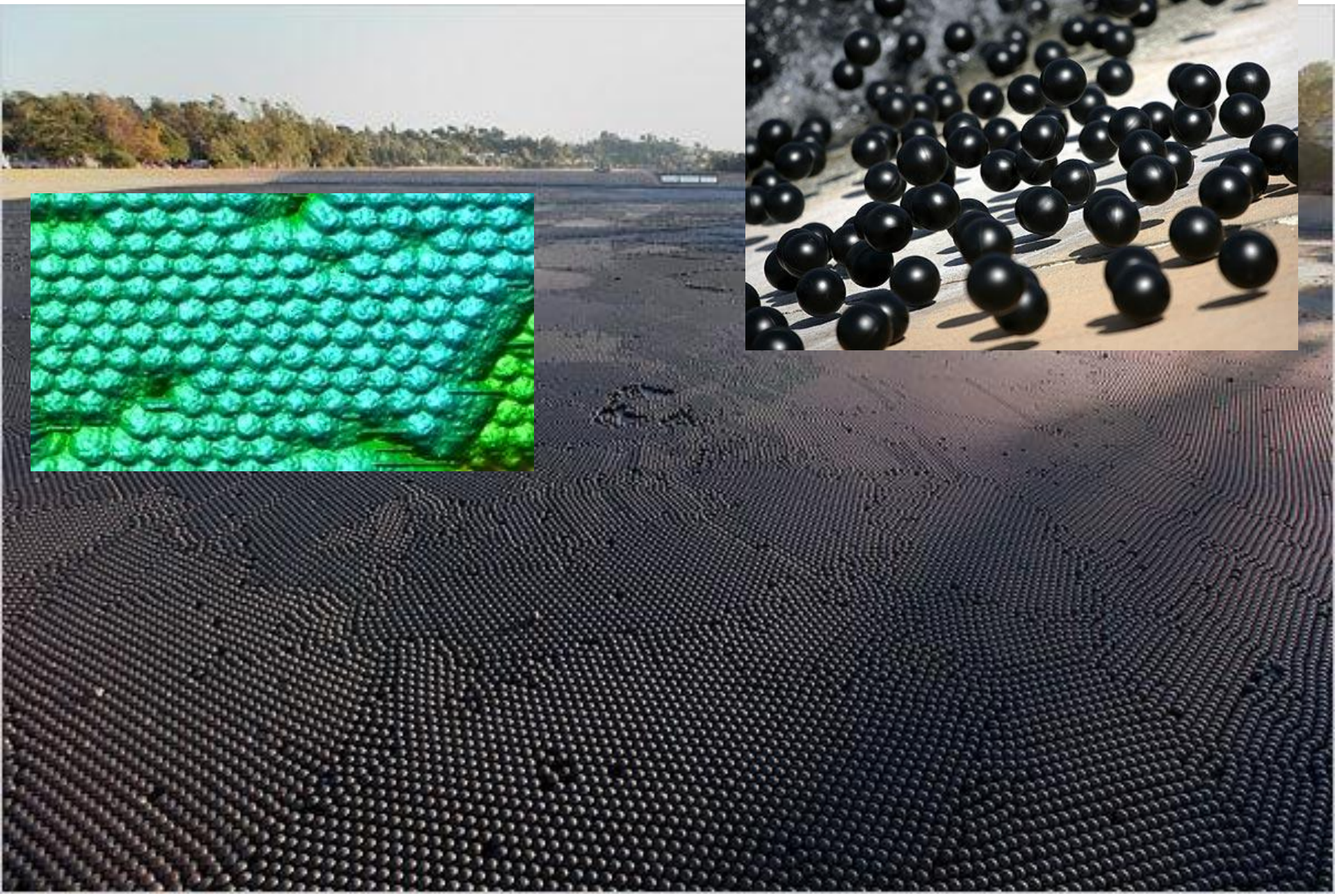


unit cell

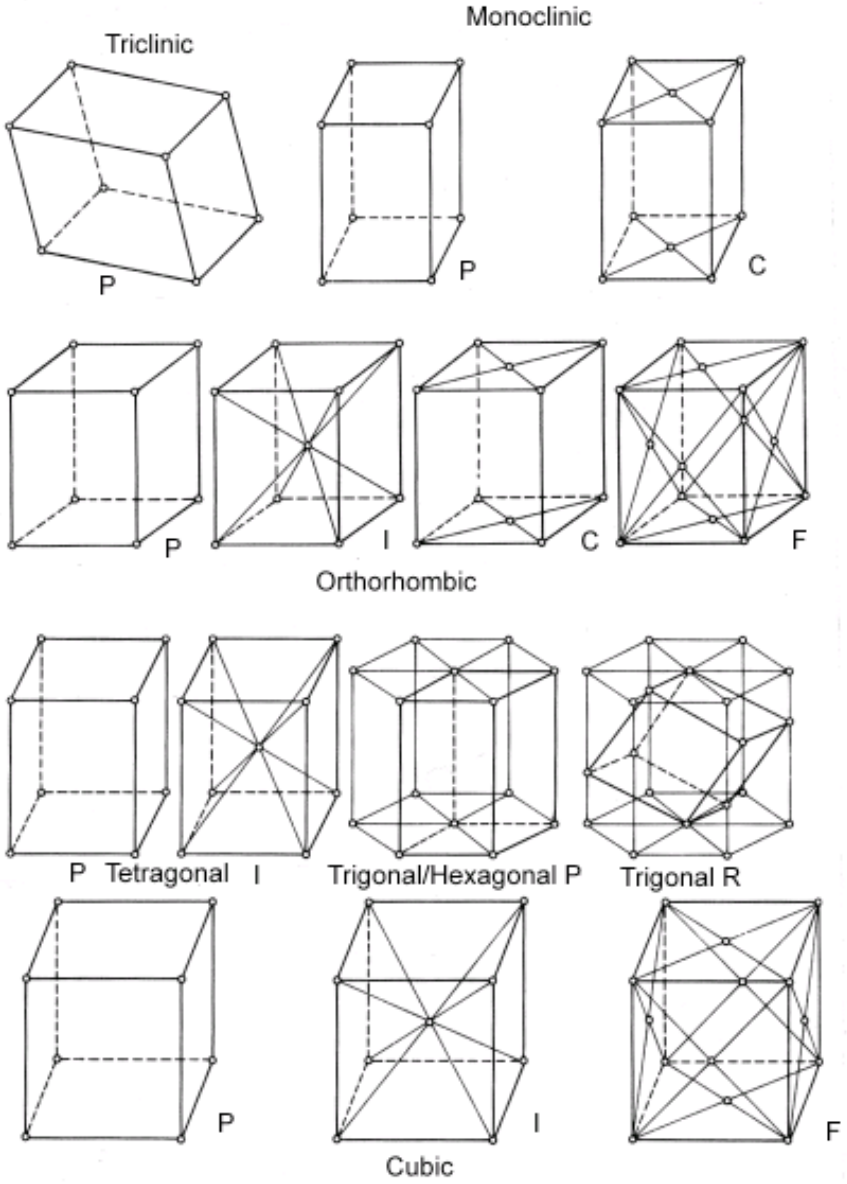


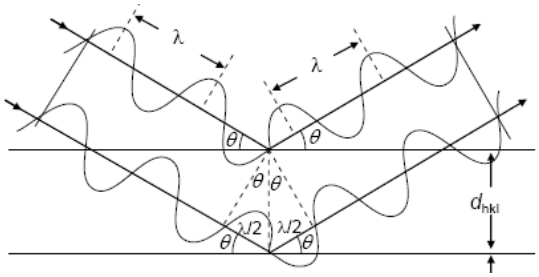
crystal

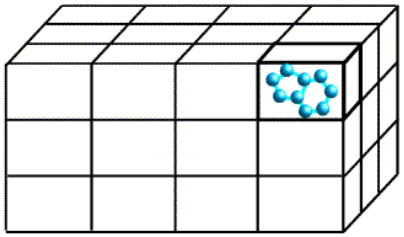


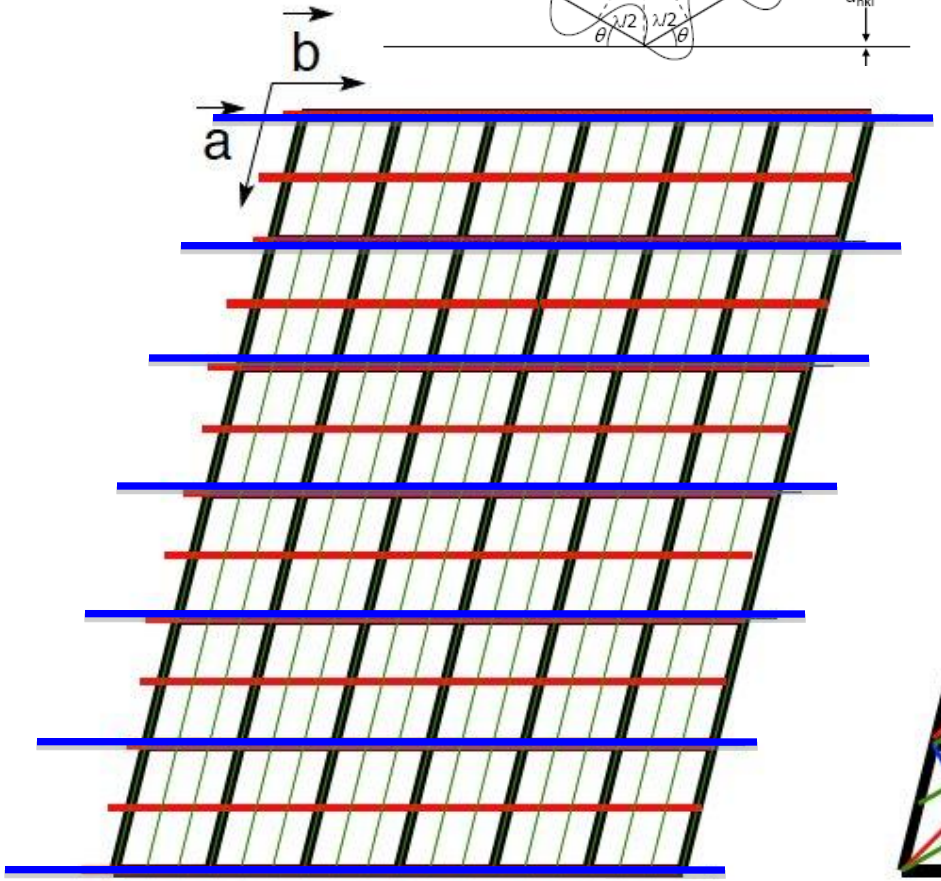


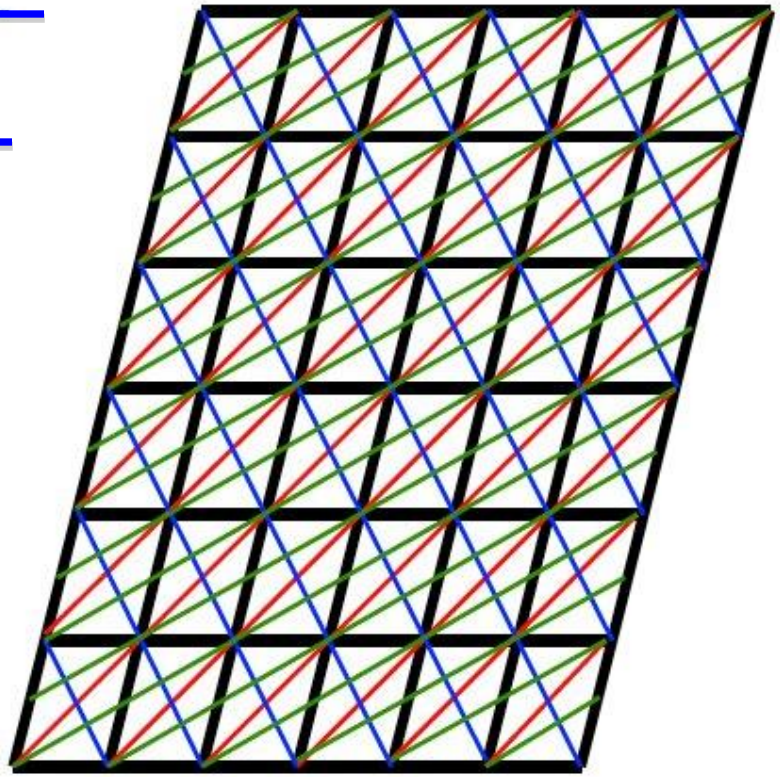
14 Bravais Lattices

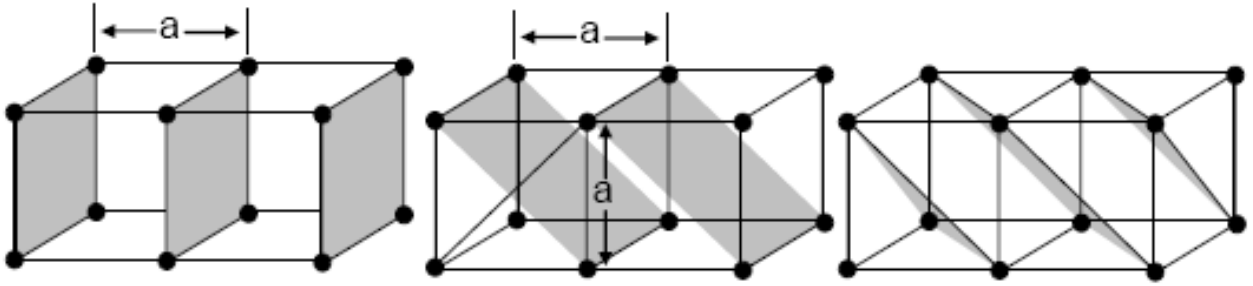






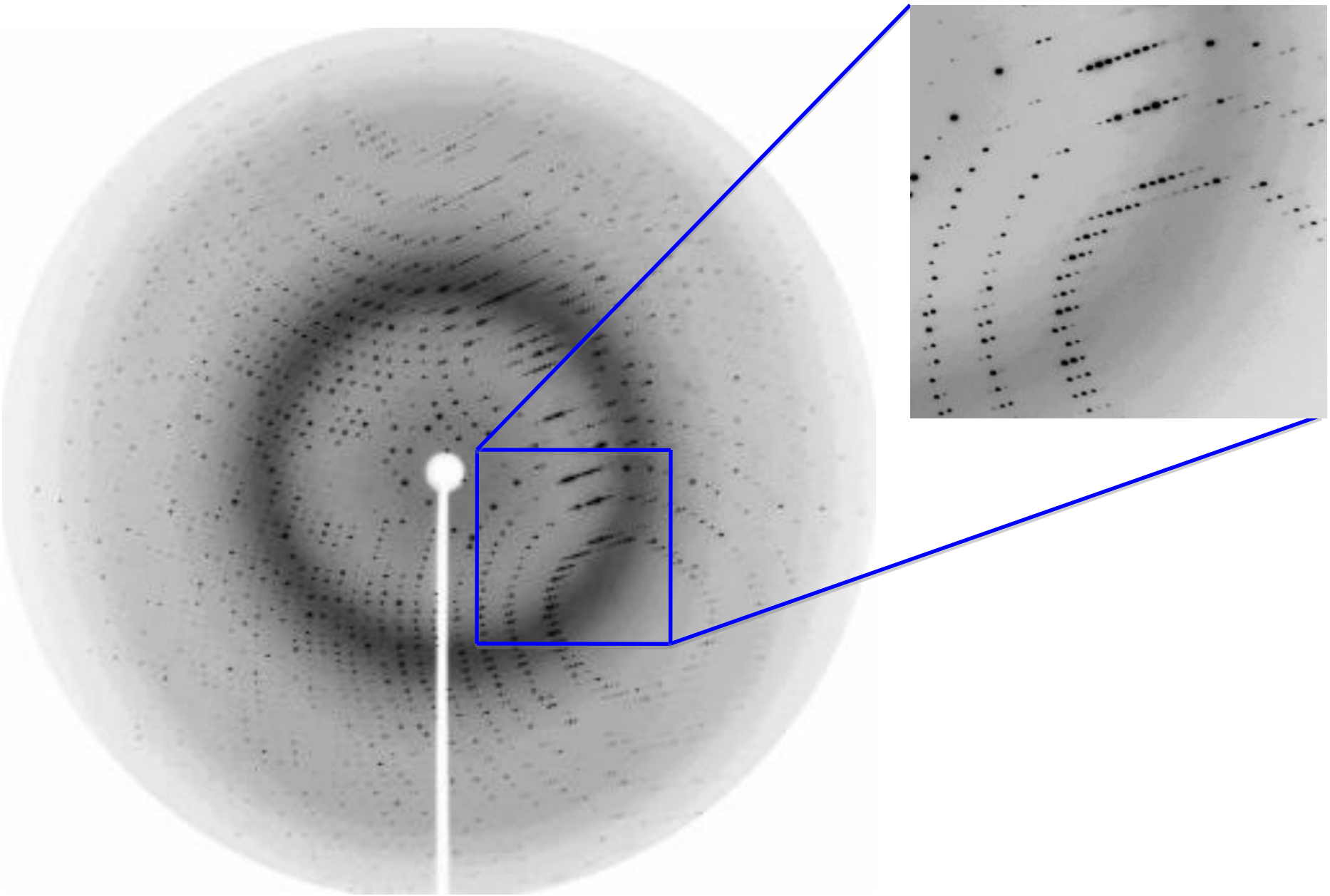




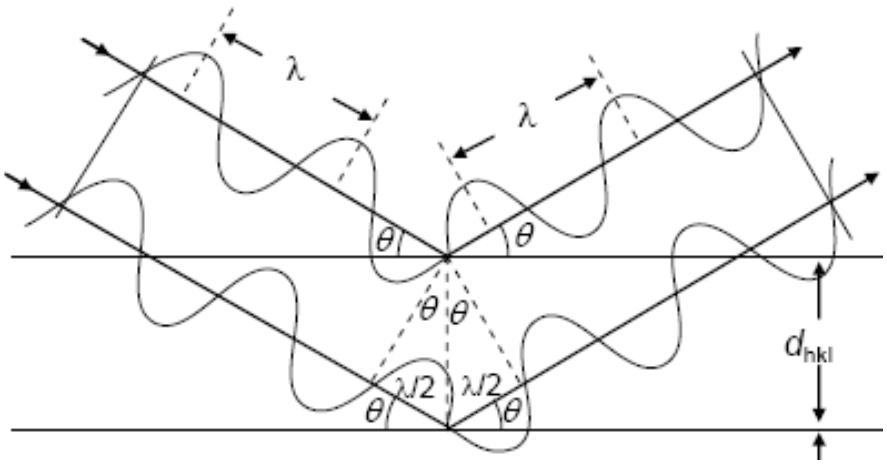


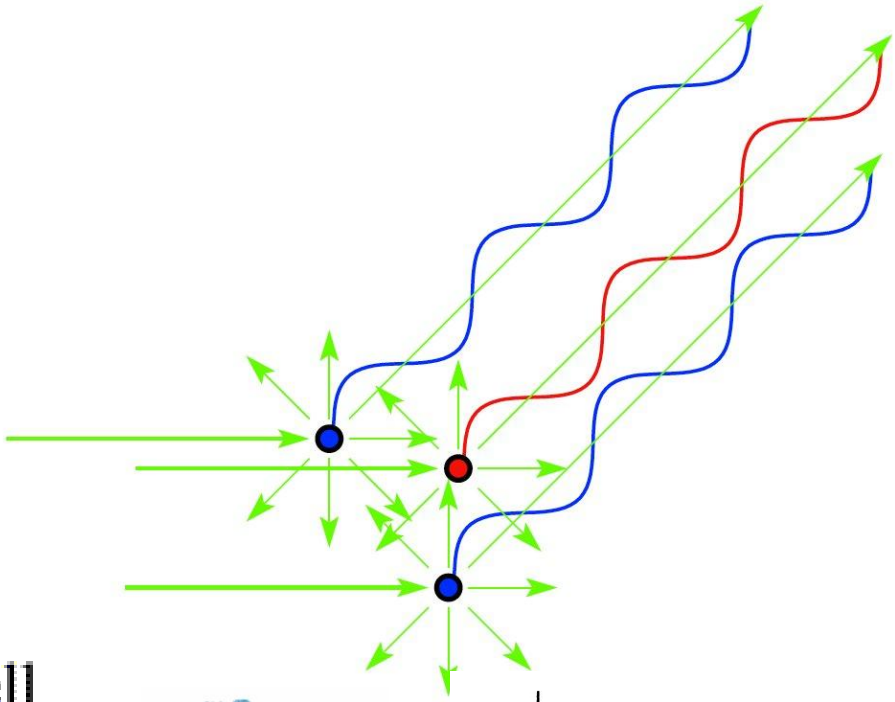
(h, k, l)

Diffraction pattern from a protein crystal



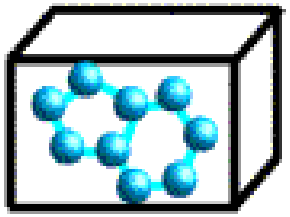
Diffraction pattern from a protein crystal

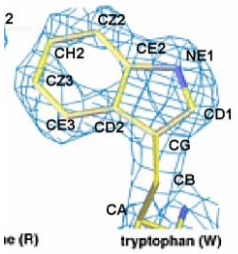


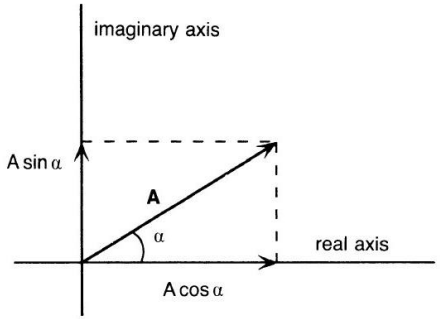


$$n\lambda = 2d \sin\theta$$

unit cell







$$F(h\,k\,l) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x\,y\,z) \exp[2\pi i(hx + ky + lz)]\,dx\,dy\,dz$$

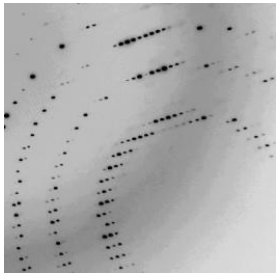
$$F(h\,k\,l) = |F(h\,k\,l)|e^{i\alpha(h\,k\,l)}$$

Electron density equation

$$\rho(x\,y\,z) = \frac{1}{V}\sum_h\sum_k\sum_l \left|F(h\,k\,l)\right| \exp\left[-2\pi i(hx + ky + lz) + i\alpha(h\,k\,l)\right]$$

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

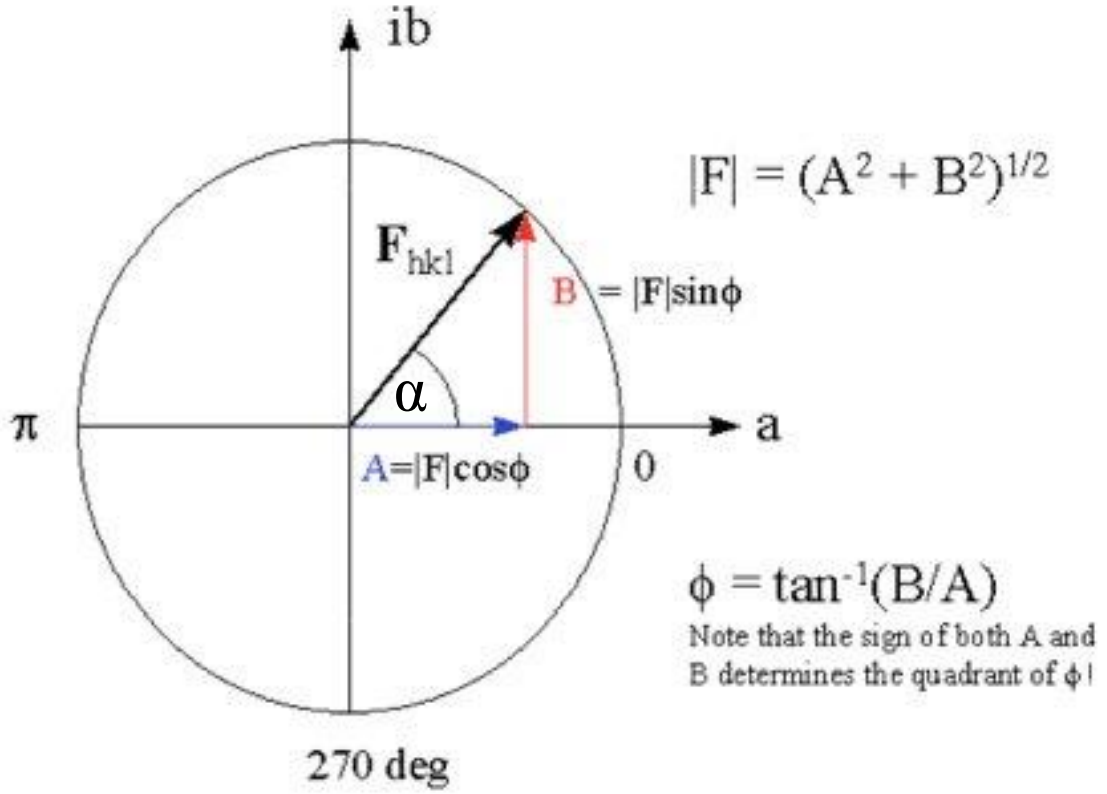
- only the intensities of reflections can be measured
- phase information is lost



- we must obtain phase information in some other way

Phase problem

- F_{hkl} is complex and can be represented with an Argand diagram.
- F_{hkl}=A+iB
- We measured $|F_{hkl}|$ in the experiment but we still need α_{hkl}.



Solving the phase problem

Molecular replacement

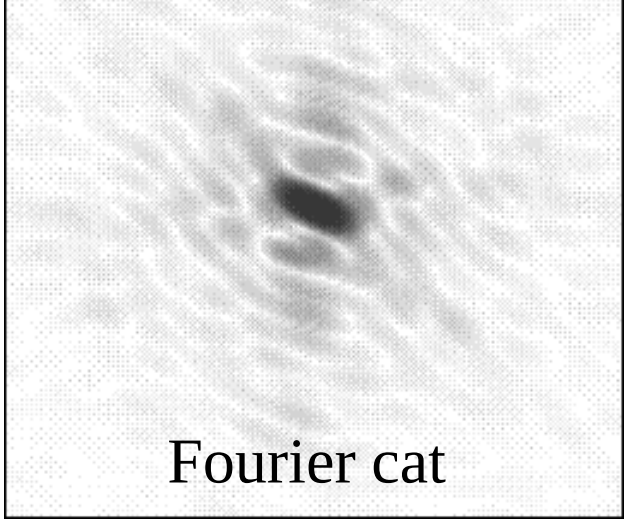
- source of initial phases is structure of similar molecule (model)
- the model is repositioned (replaced) to obtain the best agreement with the x-ray data
- phases are calculated from the model (using the structure factor equation)
- calculated phases are combined with the experimental data

Solving the phase problem 2

Multiple/Single Isomorphous Replacement (MIR/SIR)

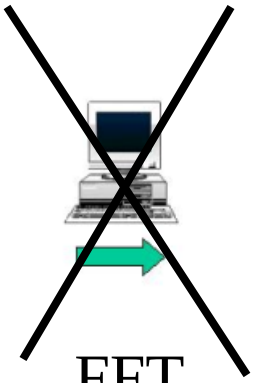
- source of phases – intensity differences between data from native and derivative (heavy atom containing) crystals
- Positions of heavy atoms identified from isomorphous difference Patterson maps

Observed amplitudes



Fourier cat

Phases unknown!



FFT

Unknown structure

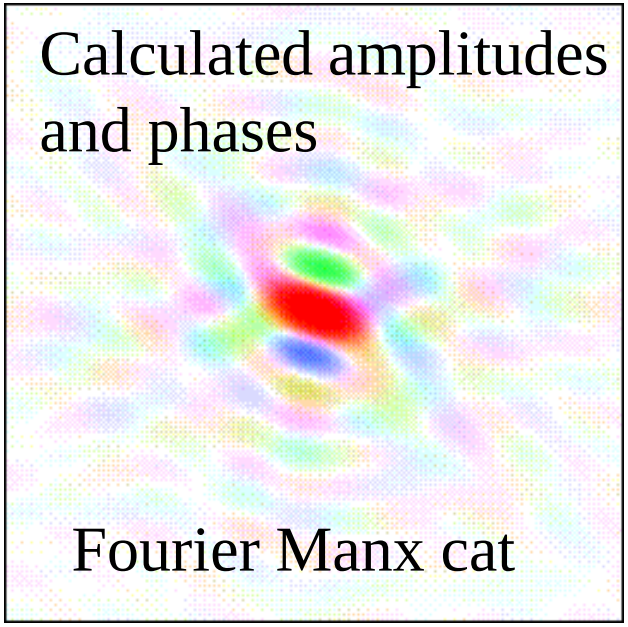
Cat

Known structure

Manx cat

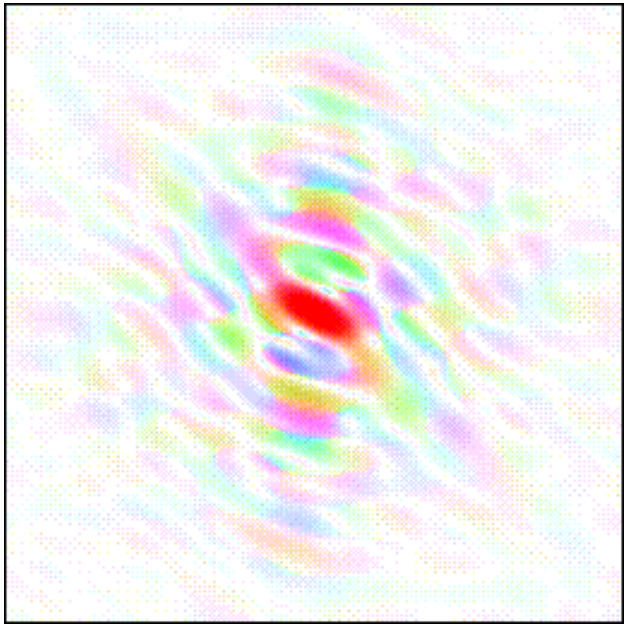
FFT

Calculated amplitudes and phases

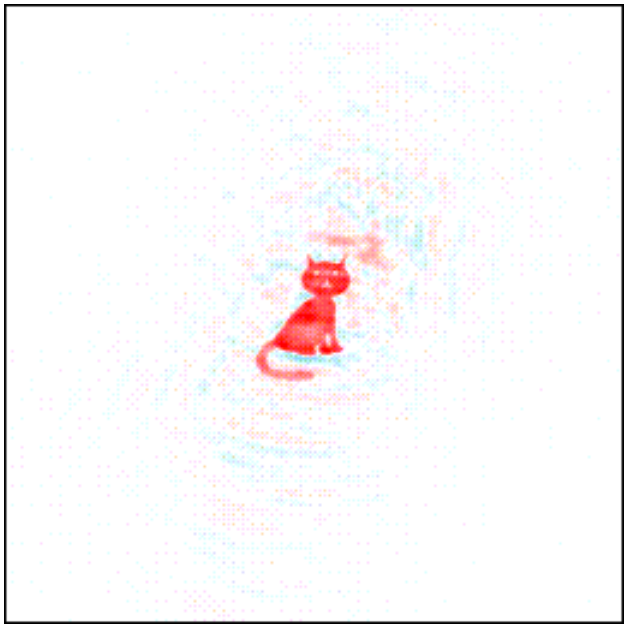


Fourier Manx cat

Observed amplitudes (tailed cat),
calculated phases (Manx cat)



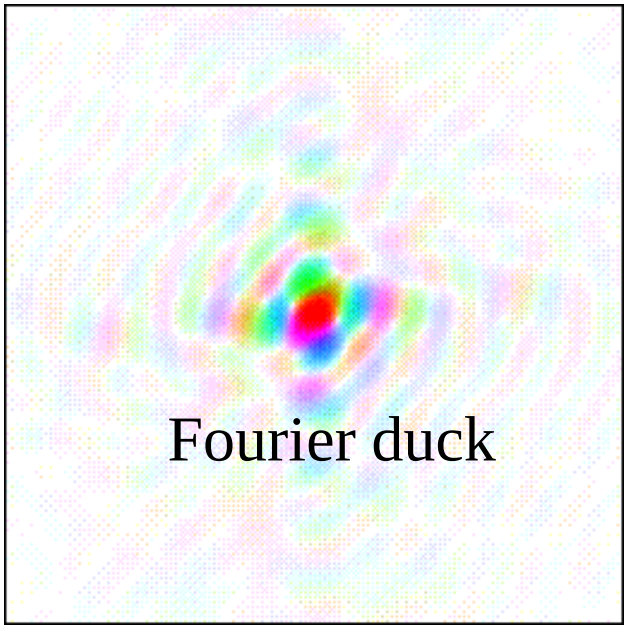
FFT



The tail becomes visible!

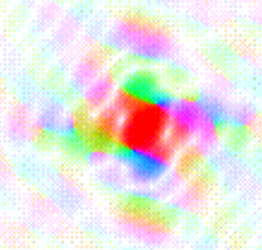
Model Bias

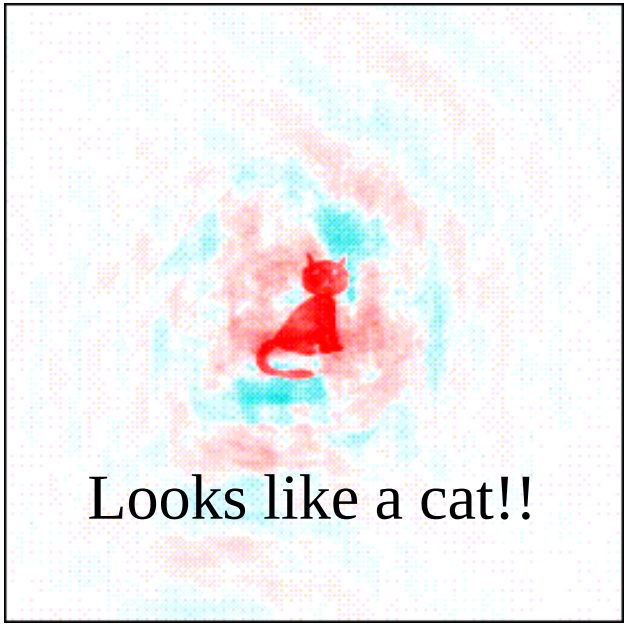
Duck



Fourier duck

Duck amplitudes + cat phases

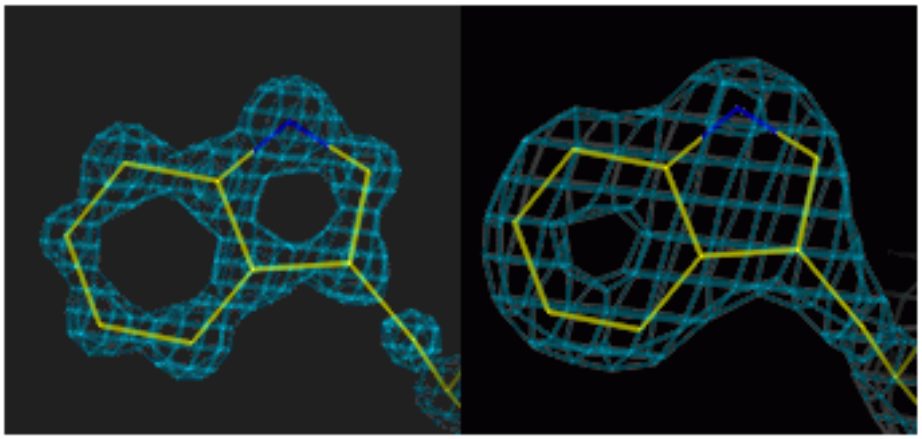




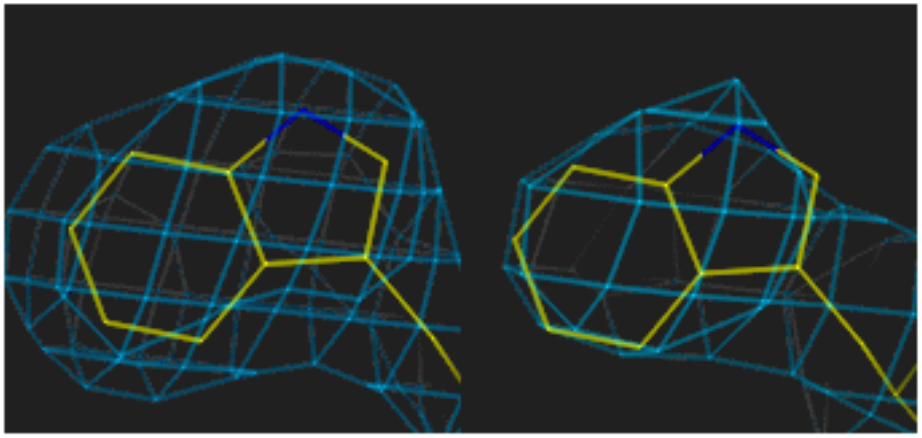
Looks like a cat!!

Model building

- Fitting of protein sequence in the electron density
- Easy in molecular replacement
- More difficult if no initial model is available
- Unambiquous if resolution is high enough (better than 3.0 Å)
- Can be automated, if resolution is close to 2Å or better

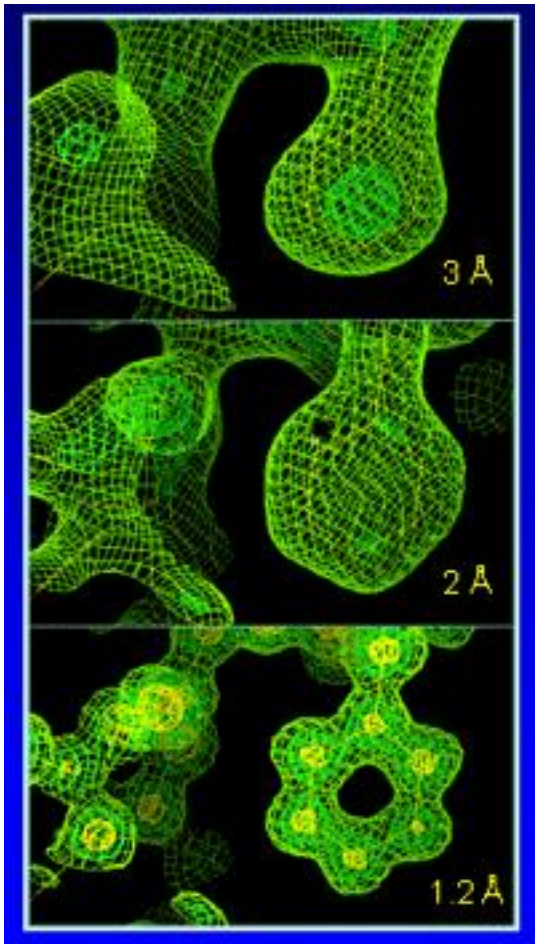


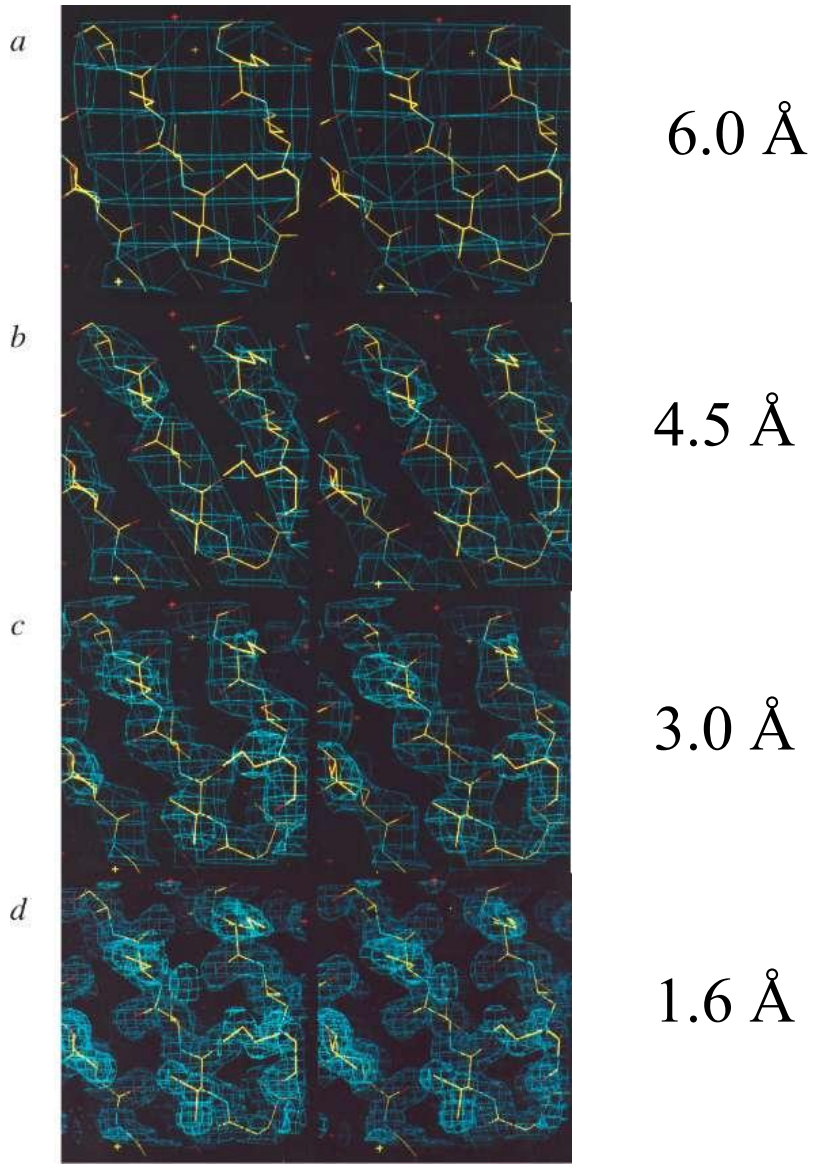
1.0Å 2.5Å



3.0Å 4.0Å

What does resolution mean in practice?





Refinement

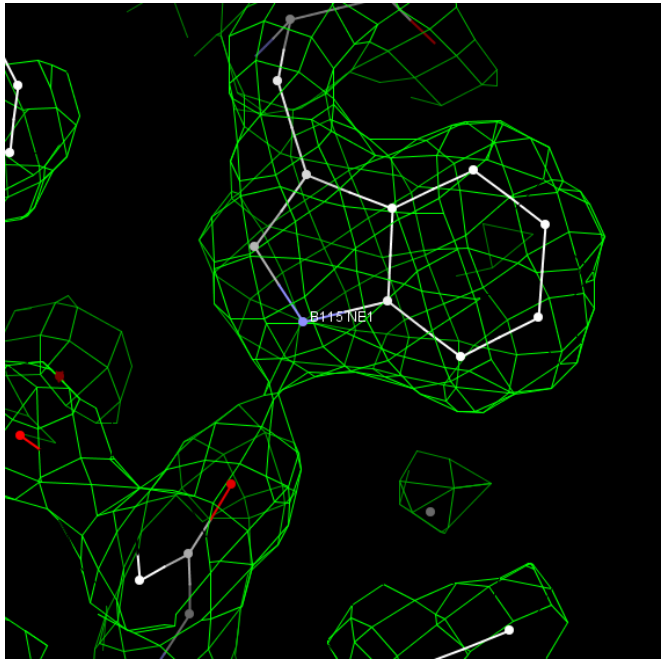
- Automated improvement of the model, so it explains the observed data better
- The phases get improved as well, so the electron density maps get better

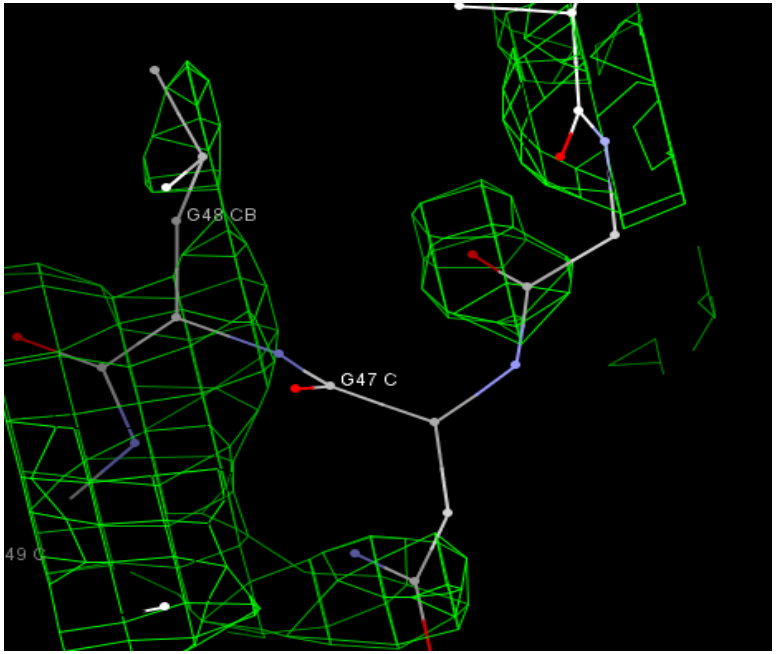
Validation

- Assesment of the final(?) model quality
- How the geometry of amino acids look like? (Ramachandran plot)
- Are non-covalently atoms far enough from each other? (no atom bumps)
- Are residues “happy” in their environment? (hydrophobic in core, polar on surface)
- Are the hydrogen donors/acceptors satisfied?

Depositing

- Depositing structure and diffraction data in PDB is required for the paper to be accepted in most journals





Summary:

1. Our goal is to obtain three dimensional electron density distribution, because it shows the shape of a molecule

2. X-rays have suitable wavelength for study of molecular structures

3. Crystals allow measurement of diffraction data because they diffract strongly in certain directions

4. Diffraction experiments provide only amplitudes of structure factors => Phase problem

5. Solution of the phase problem:

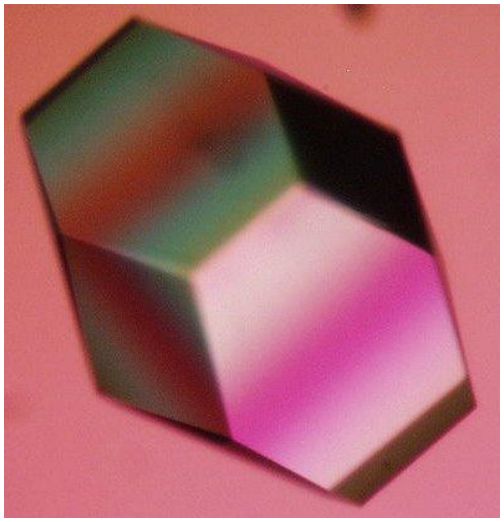
    Molecular replacement

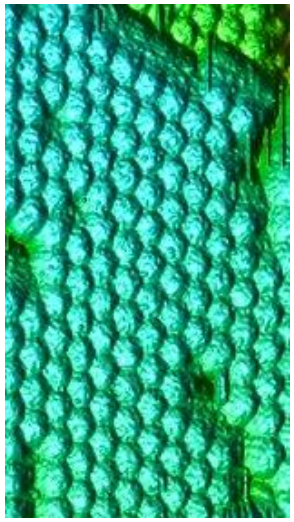
    Isomorphous replacement

6. Model building, refinement, deposition

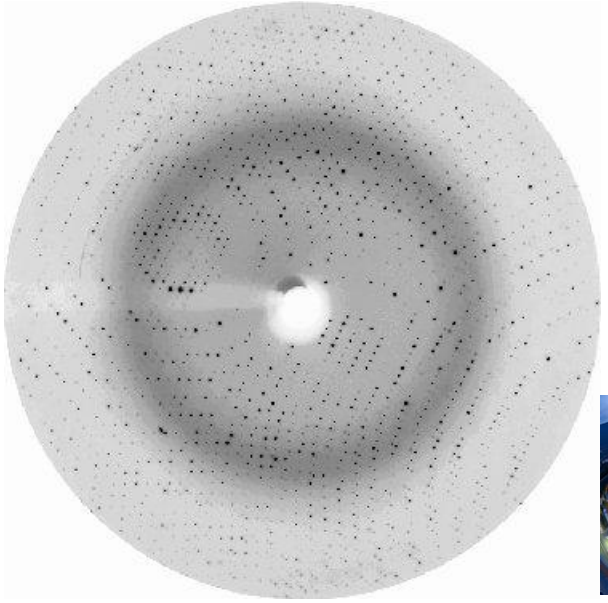
1. Virus purification

2. Crystallization

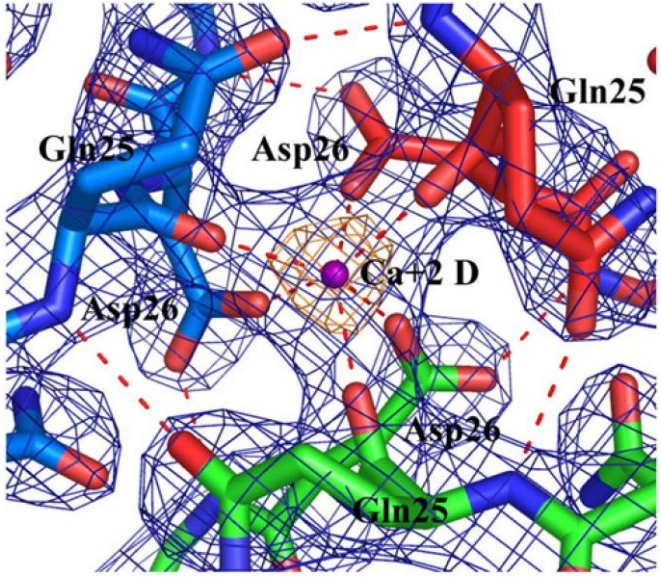




3. Diffraction data

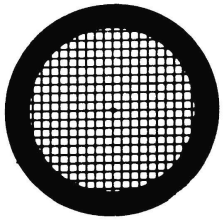


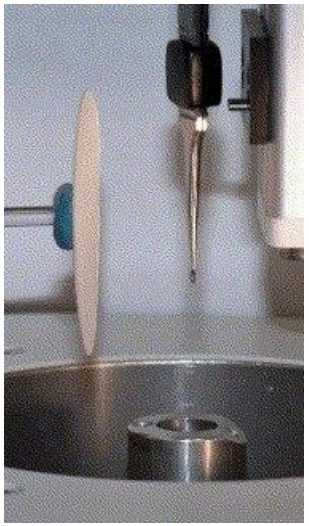
4. Solve structure



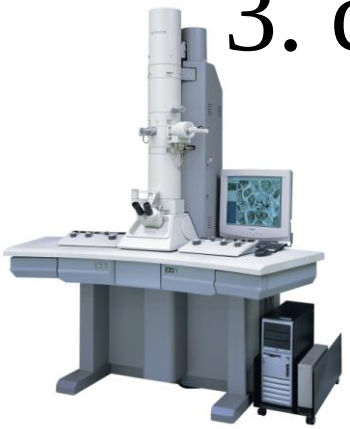
1. Virus purification

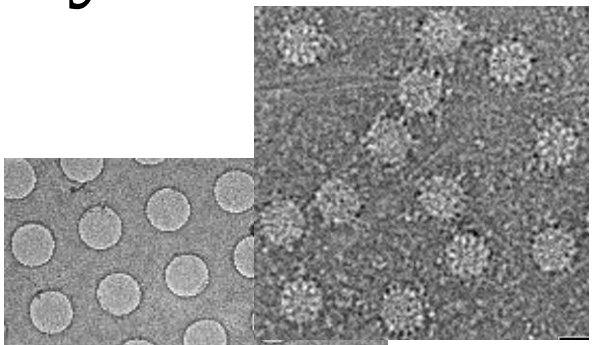
2. Grid preparation

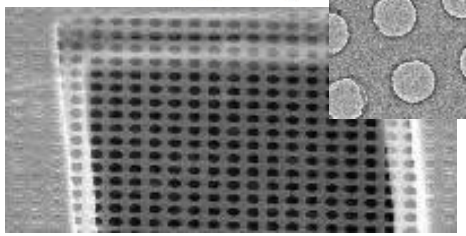




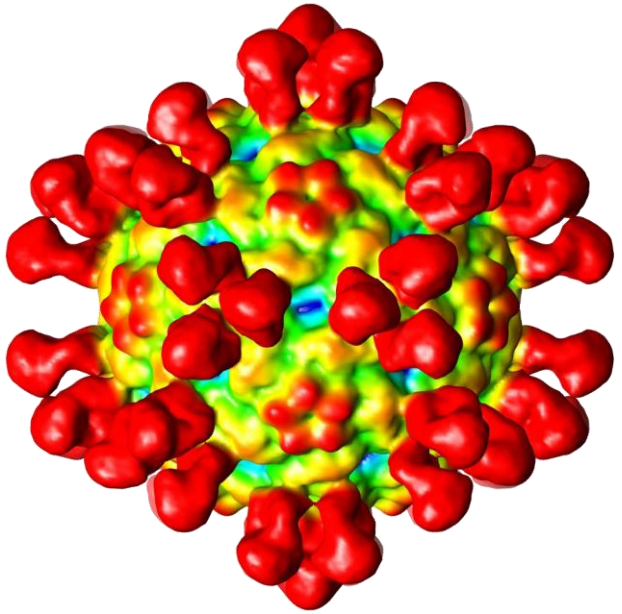
3. cryo-EM







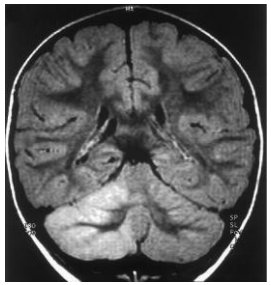
4. Reconstruction

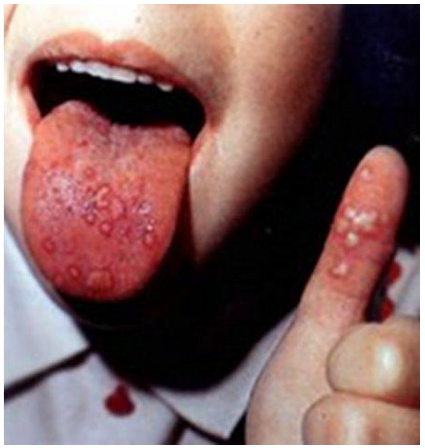


Structural studies of human picornaviruses

Rhinoviruses
- 40% of common cold cases
- economic losses $16bn/year in USA

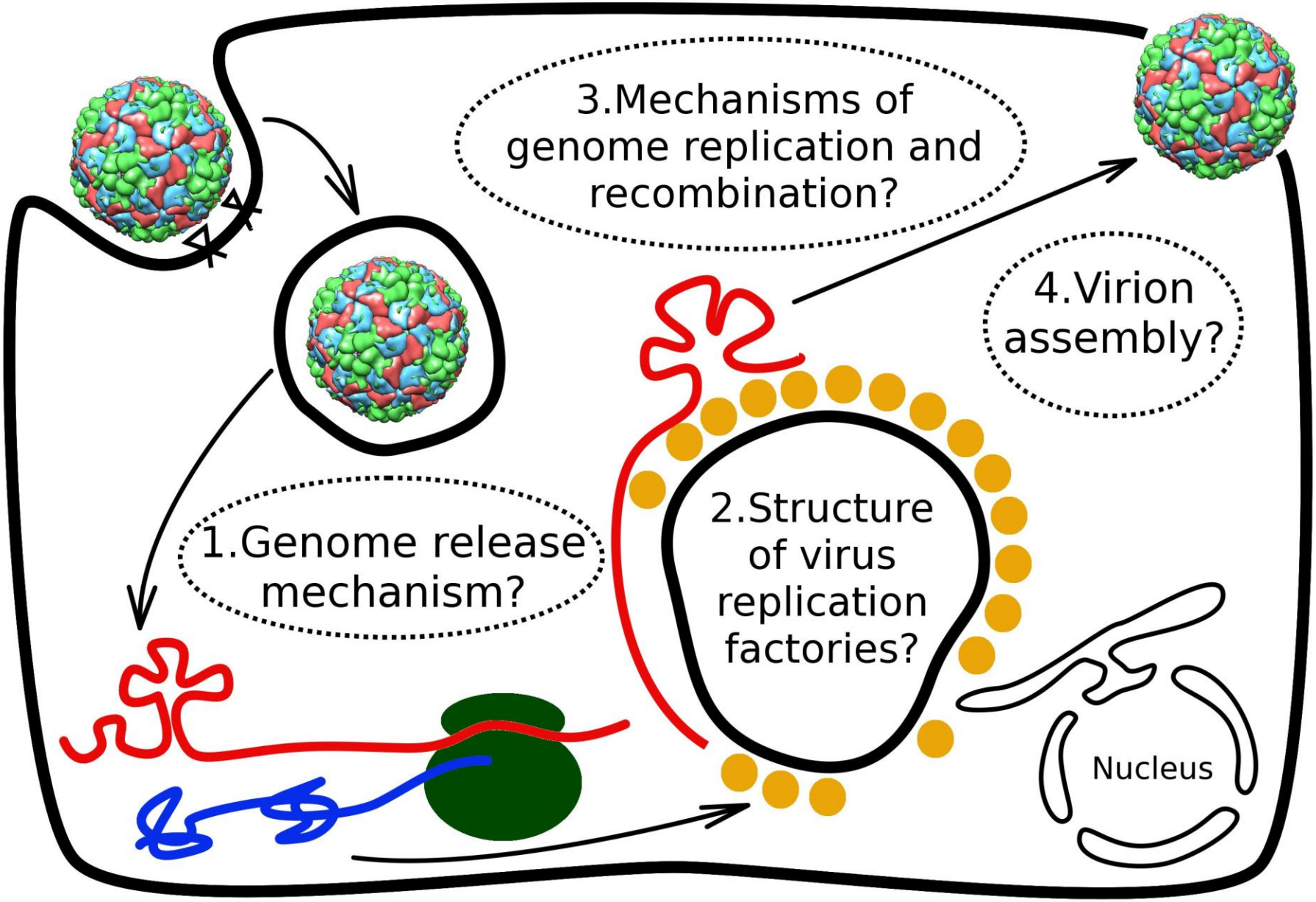
Enteroviruses (EV71)
- hand-foot-and-mouth-disease
- encephalitis



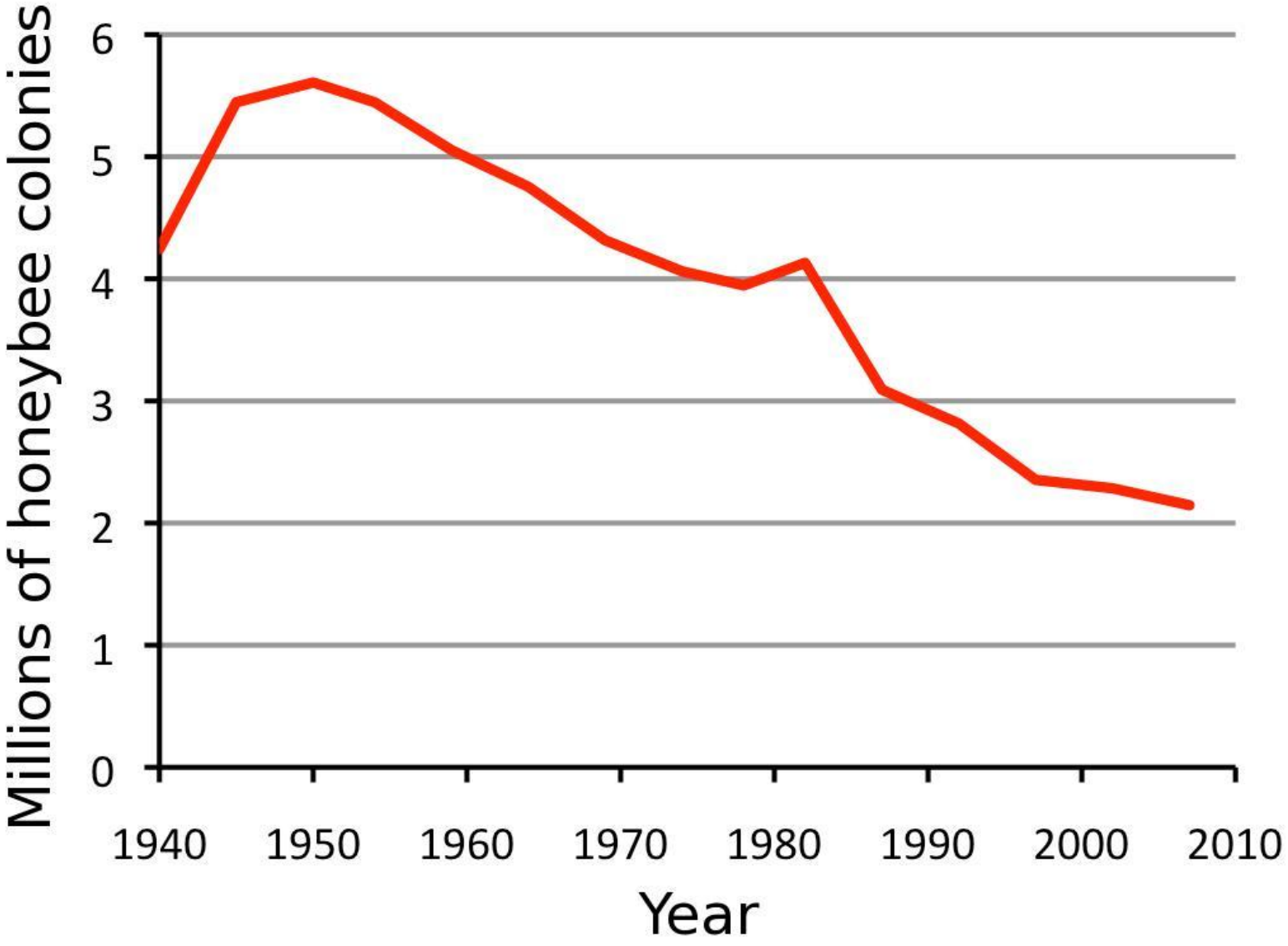


McMinn et al. Clin Infect Dis 2001. Image by Heng Soy, KI Media 2012.

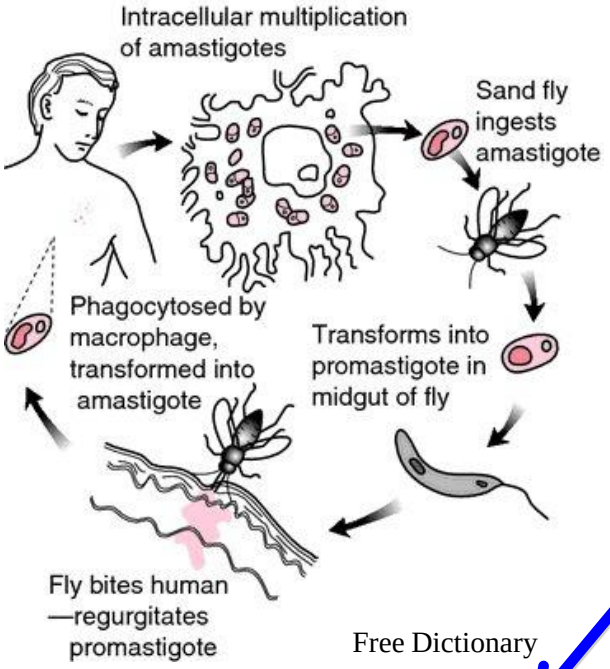
Picornavirus replication cycle



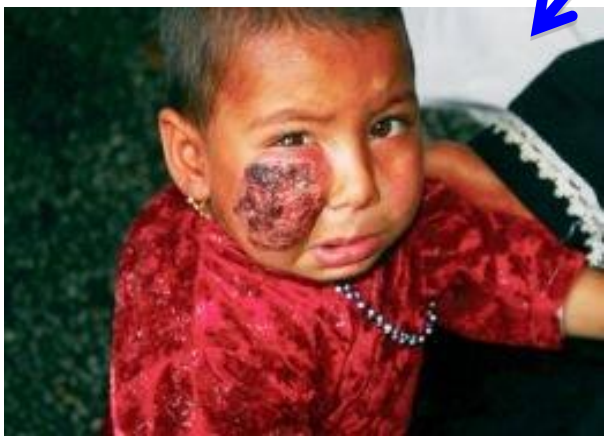
Honeybee viruses



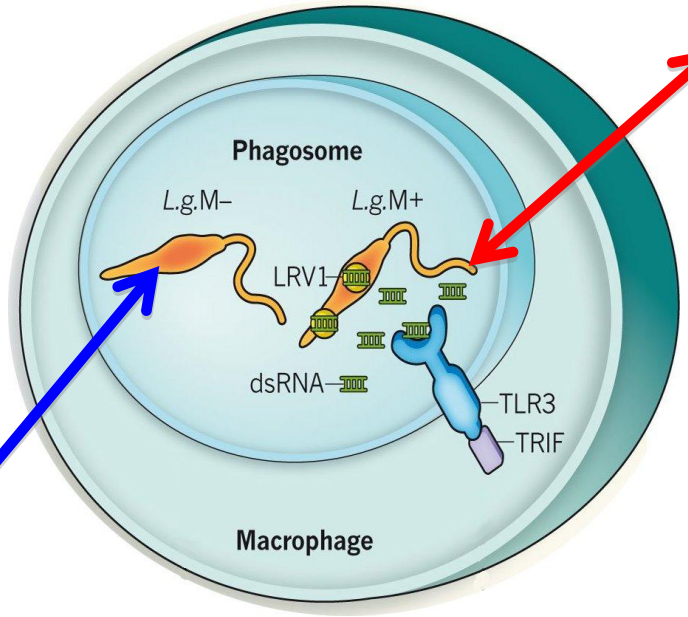
Leishmania RNA virus 1



Free Dictionary



cutaneous leishmaniasis



Ives et al. Science, 2011
Olivier. Nature, 2011

metastasis to nasopharyngeal tissues