Macromolecular crystallography

Lecture 1

Pavel Plevka

- Lecture 1 – Introduction to X-ray crystallography, basic diffraction

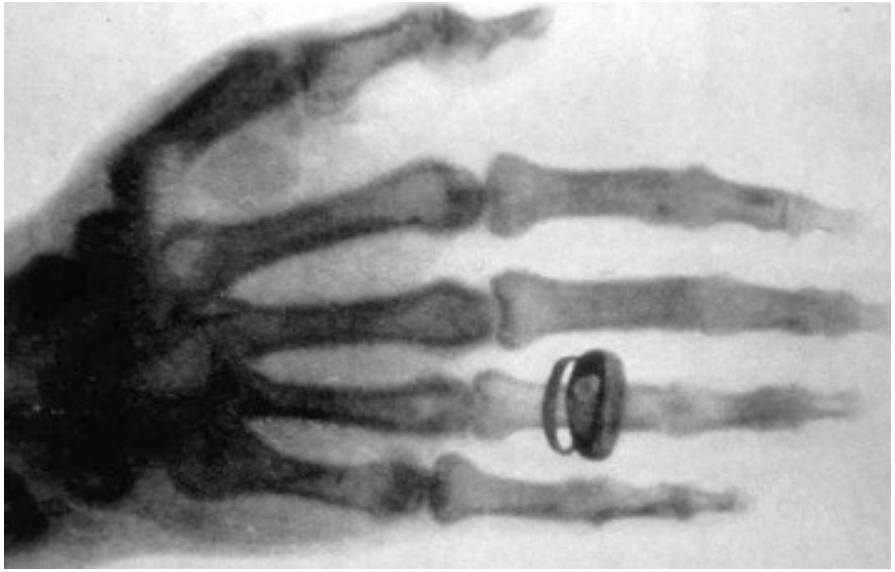
- Lecture 2 – Solution of phase problem, model building, and structure validation

- Development of crystallography
- Waves, radiation, and diffraction

WILHELM CONRAD RÖNTGEN (1845-1923)

- **1901 Nobel Laureate in Physics**

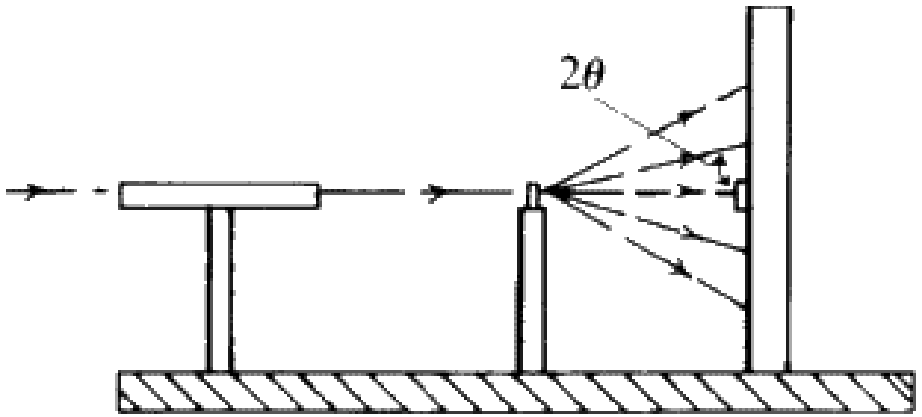
  discovery of the remarkable rays subsequently named after him.



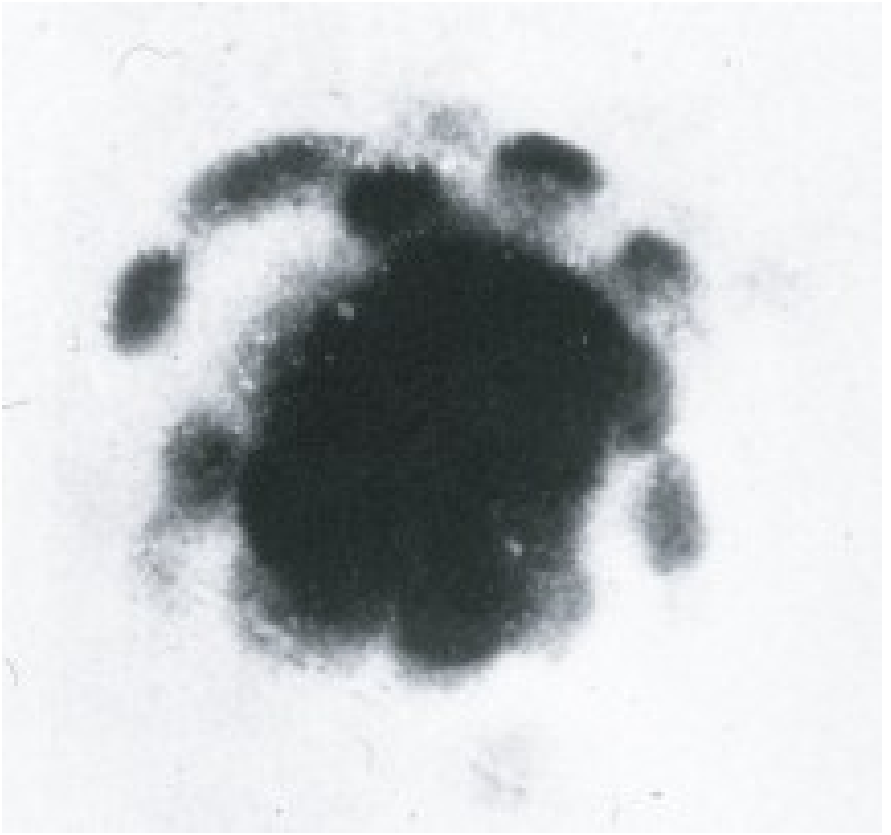
MAX VON LAUE (1879-1960)

- **1914 Nobel Laureate in Physics**

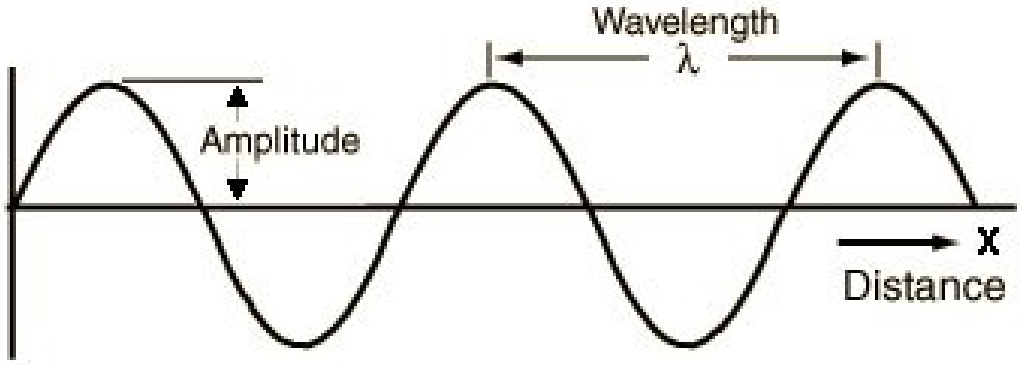
  for his discovery of the diffraction of X-rays by crystals

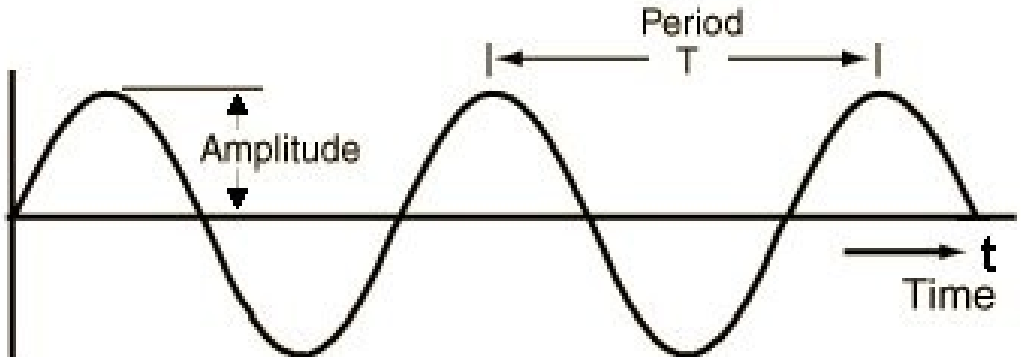


Friedrich and Knipping

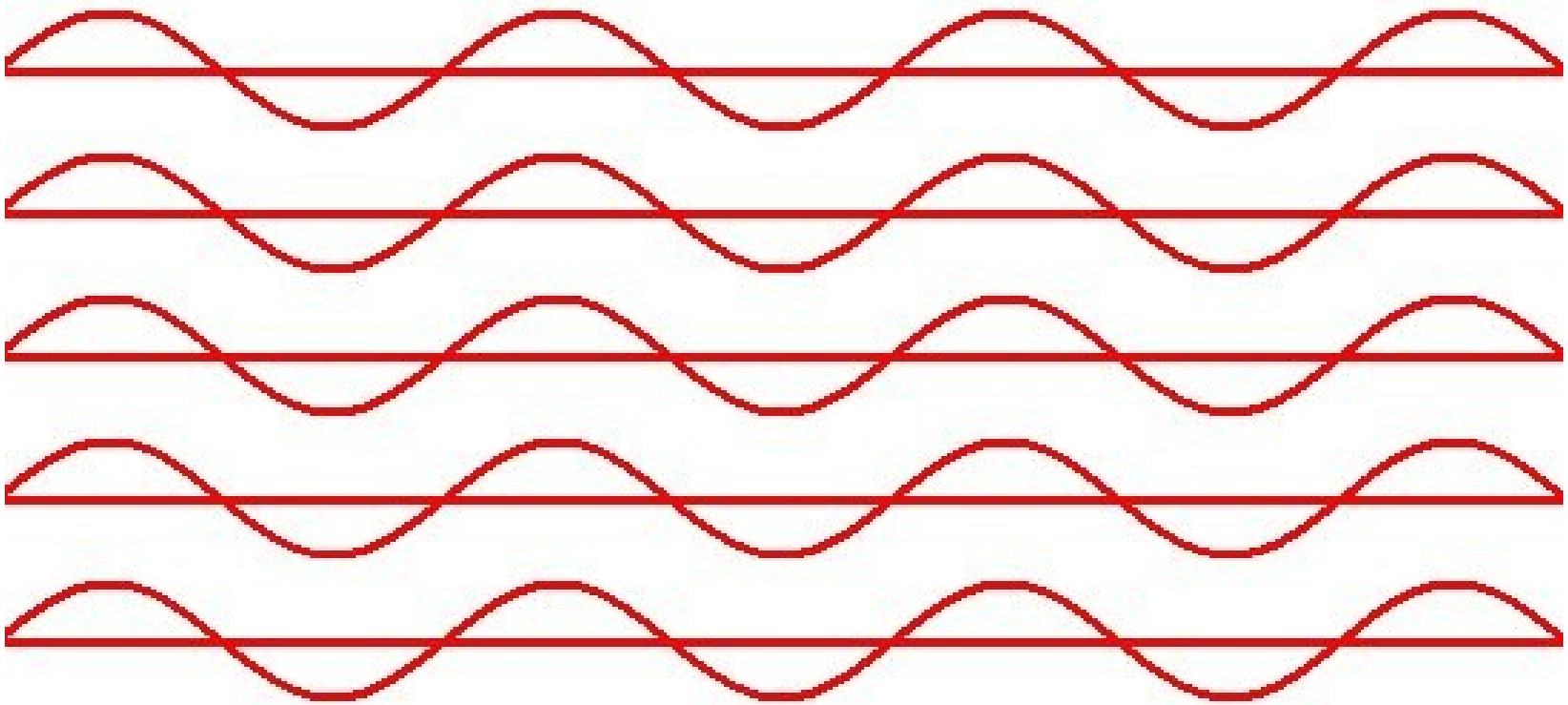


Waves

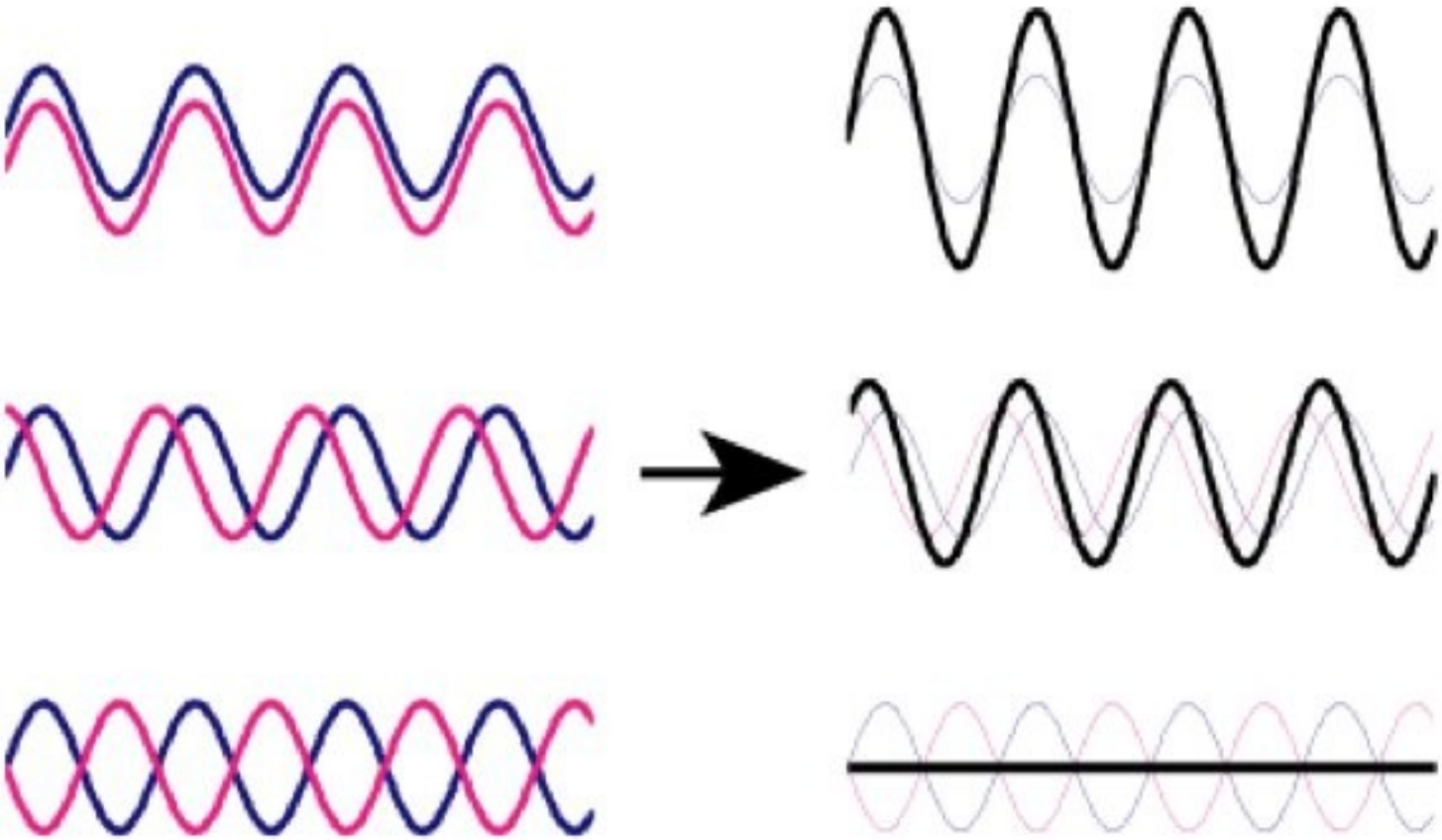

Wavelength
λ
Amplitude
x
Distance


Period
T
Amplitude
t
Time

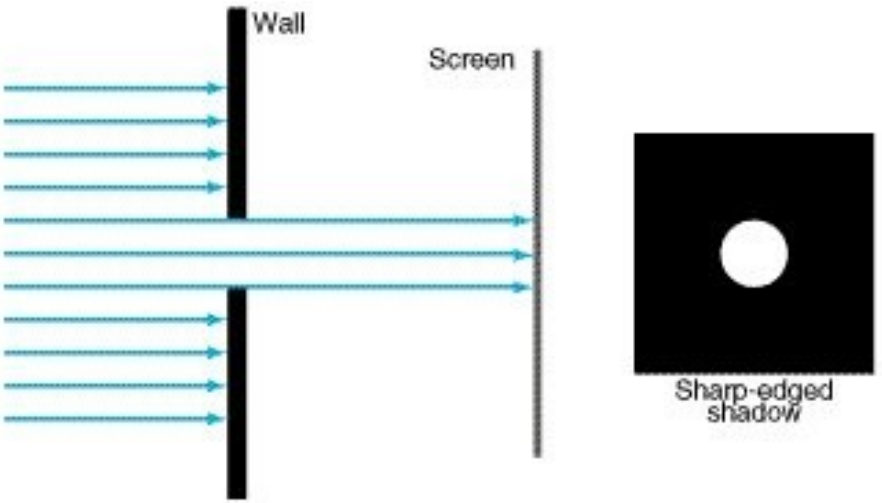
Coherent beam

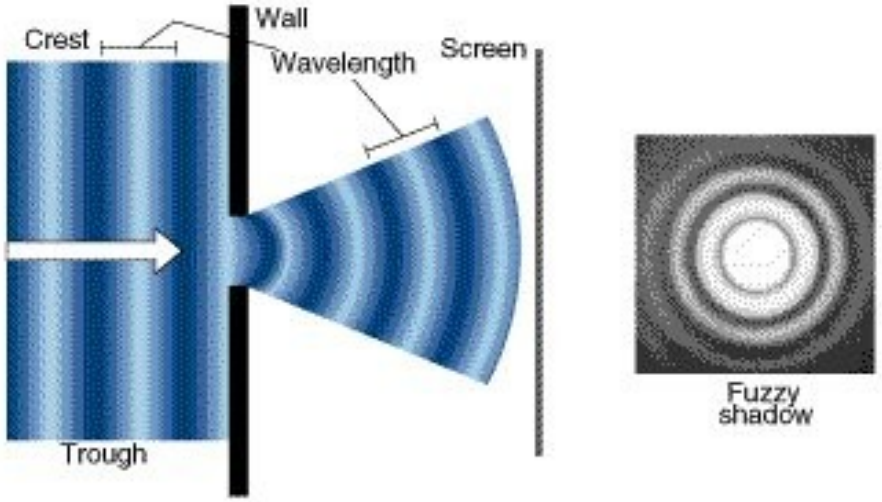


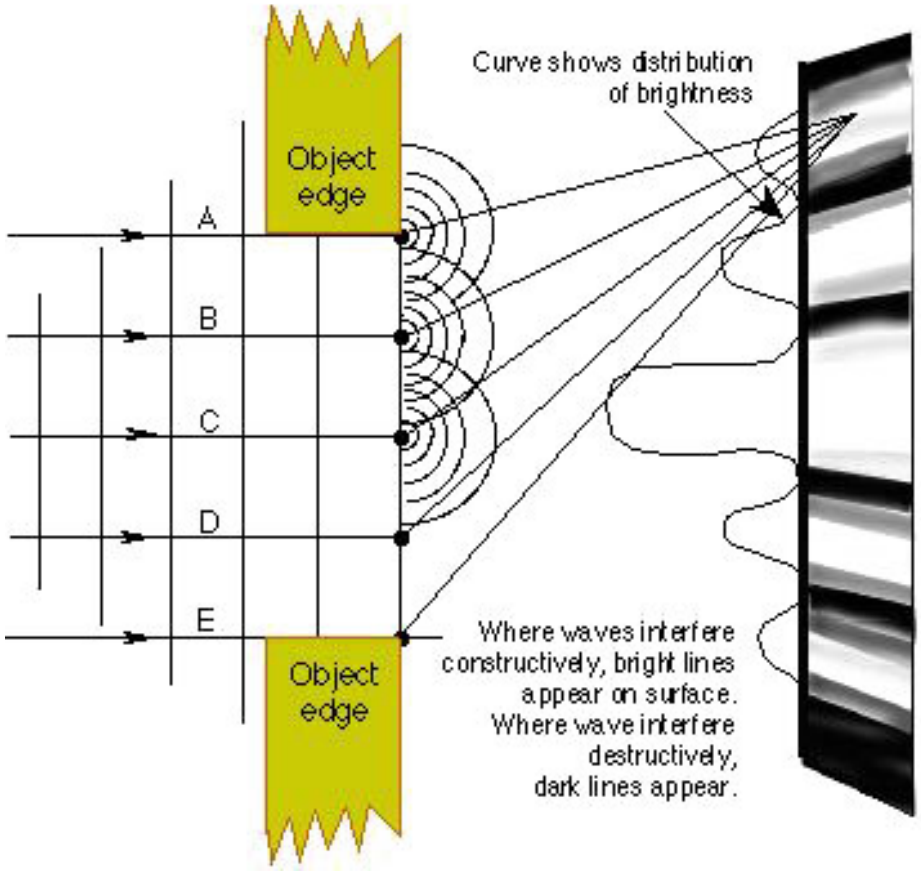
Addition of waves



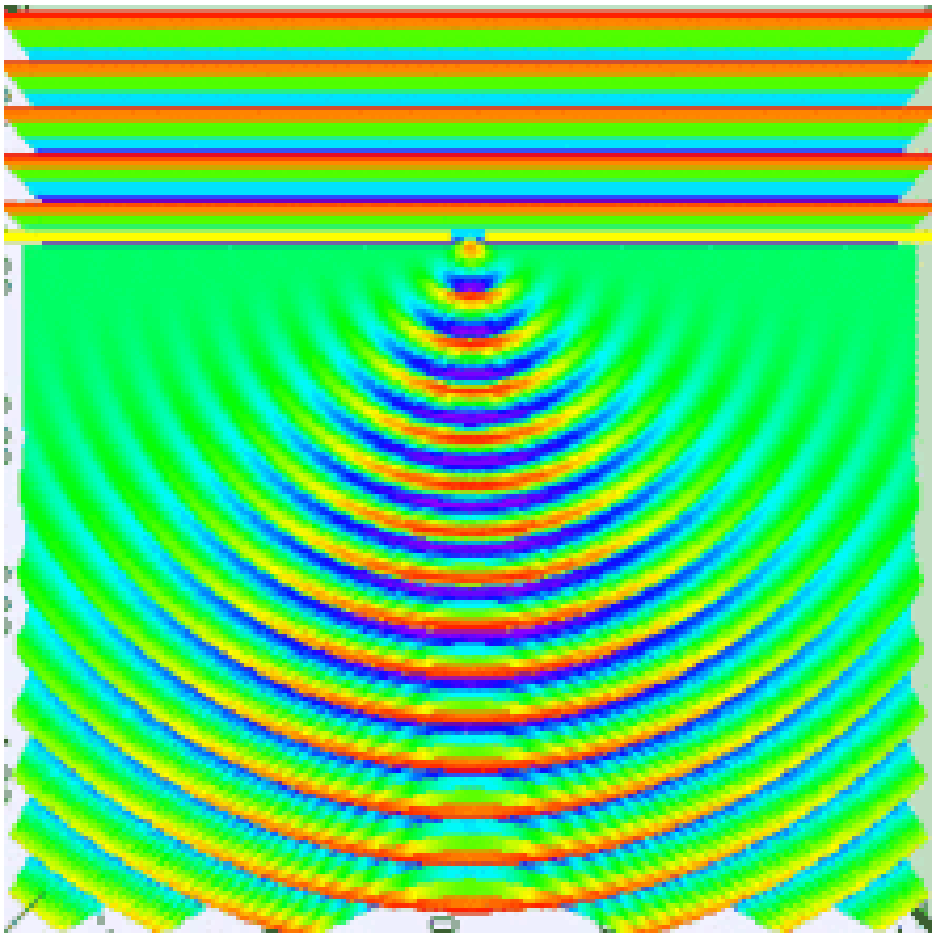
Particles & waves



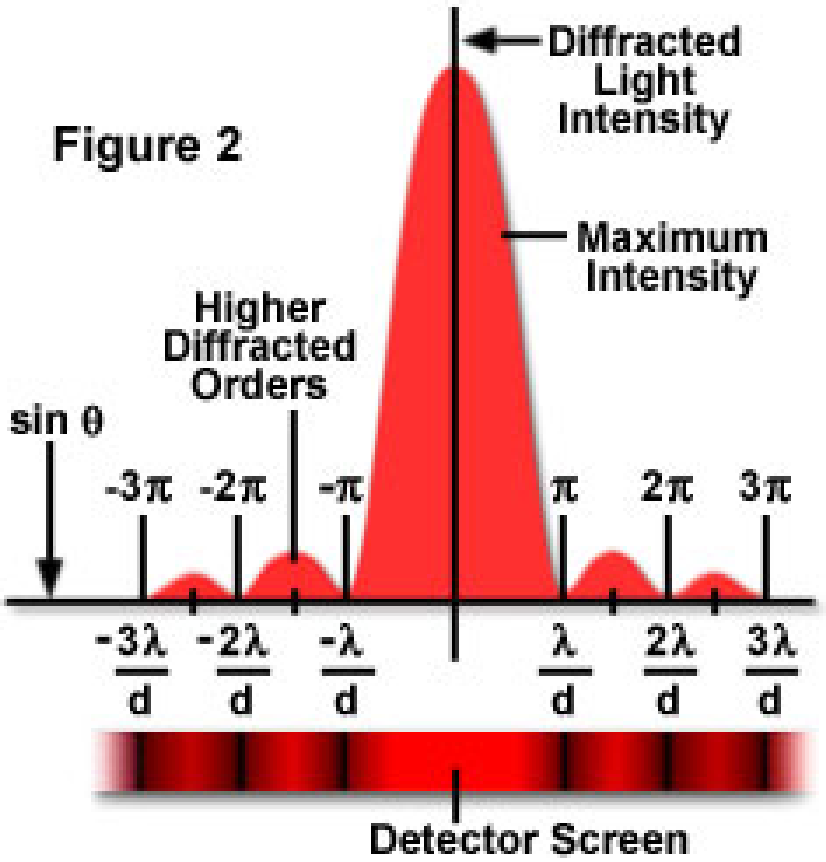




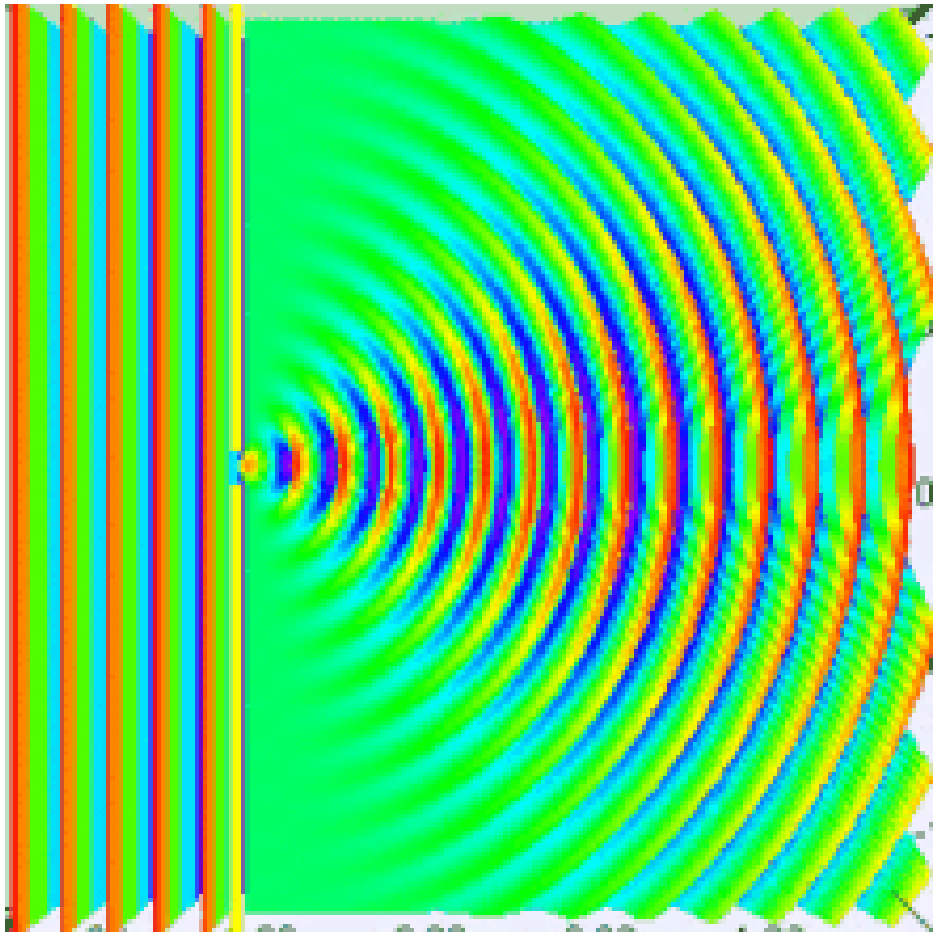
Diffraction of light

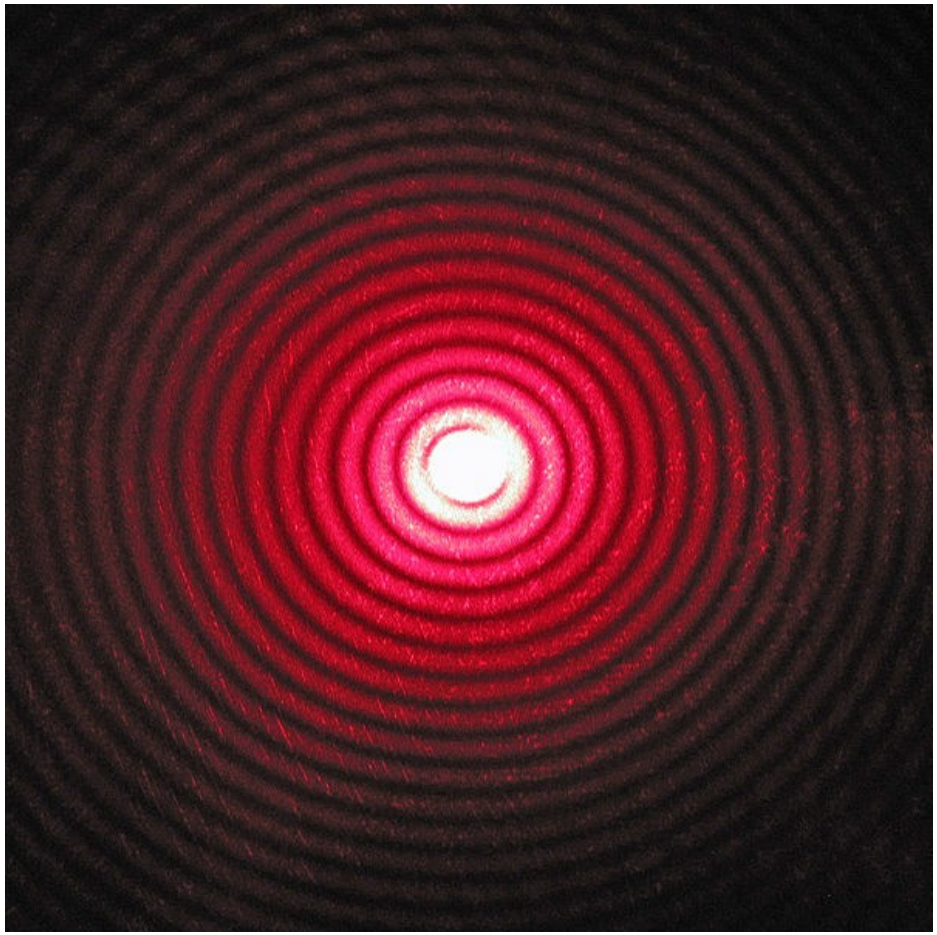


Intensity Distribution of Diffracted Light



Diffraction of light

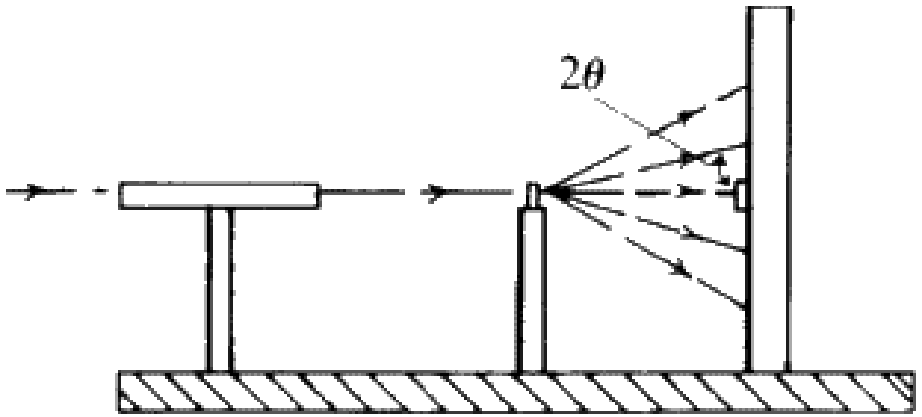


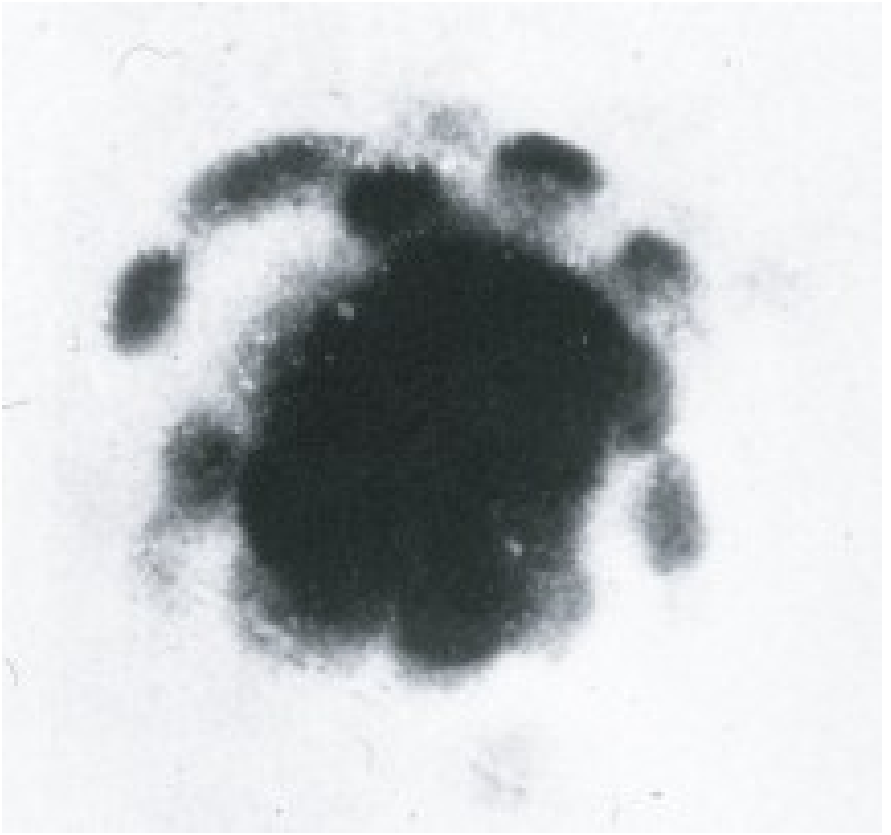


MAX VON LAUE (1879-1960)

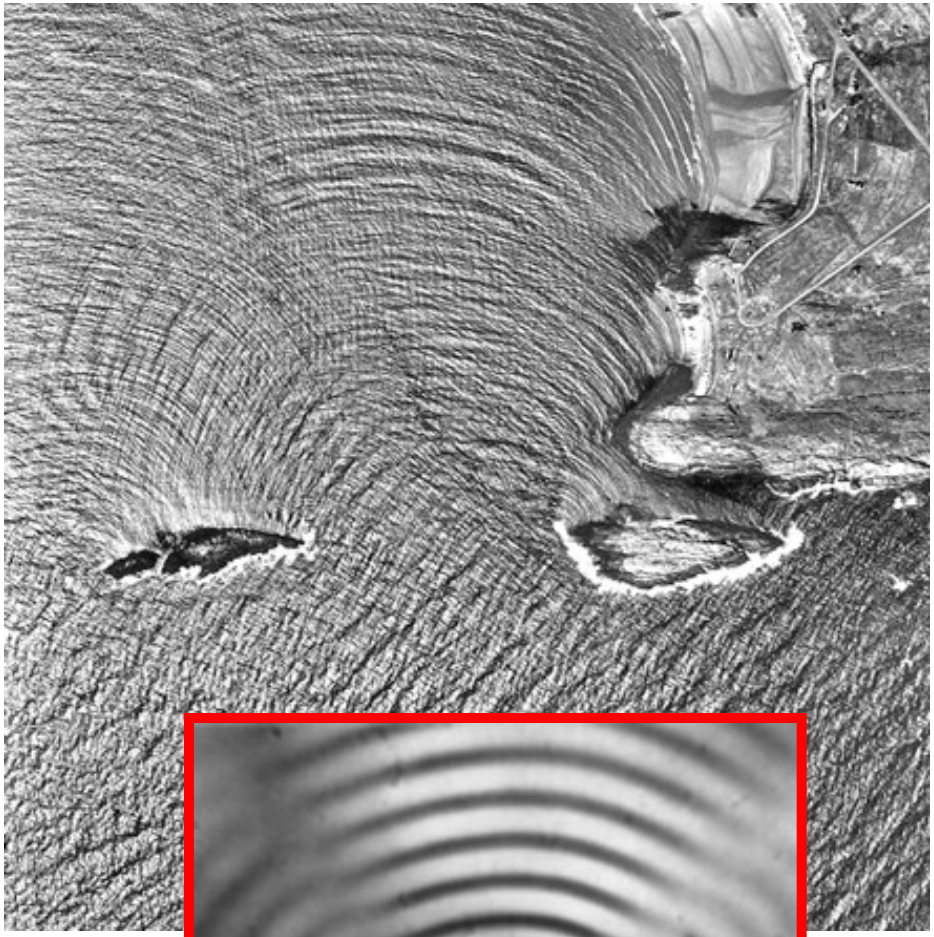
- **1914 Nobel Laureate in Physics**

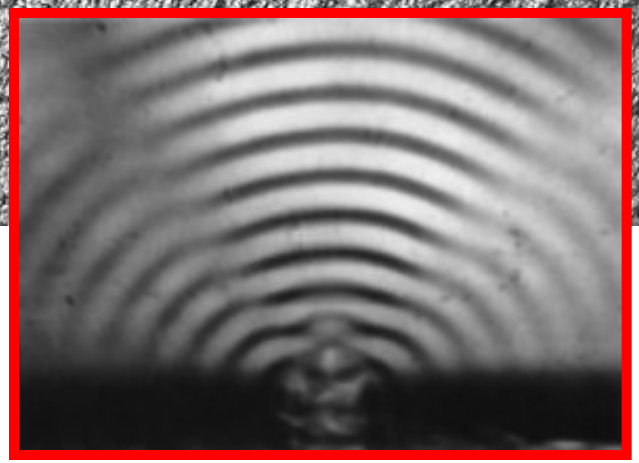
  for his discovery of the diffraction of X-rays by crystals

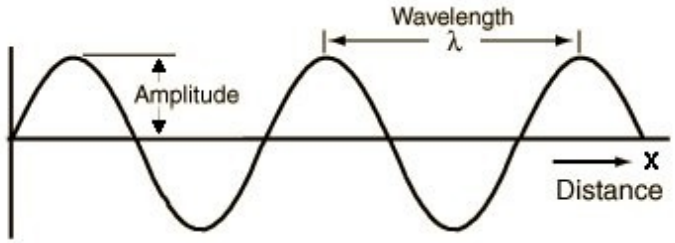




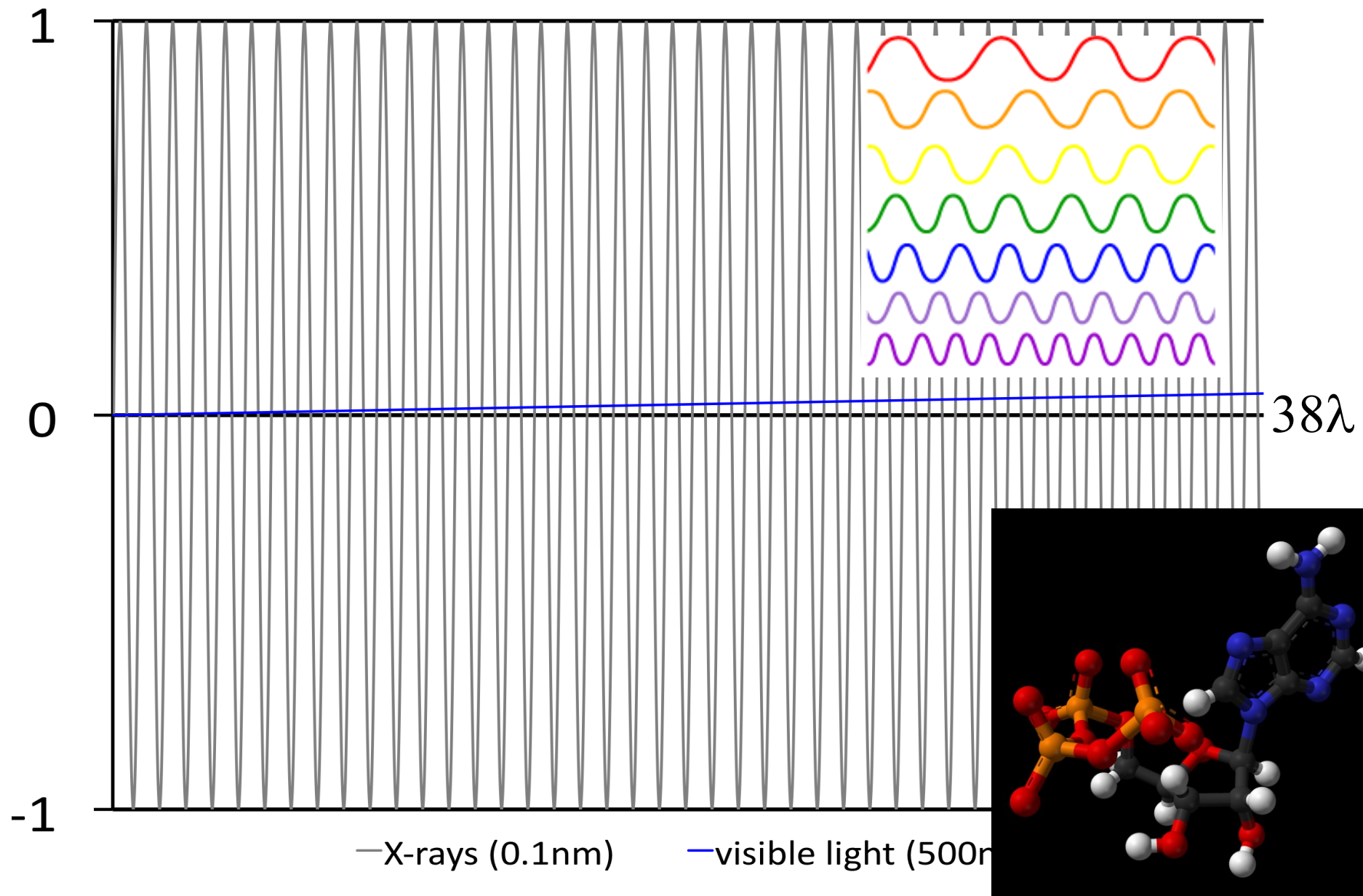
Wavelength and diffraction

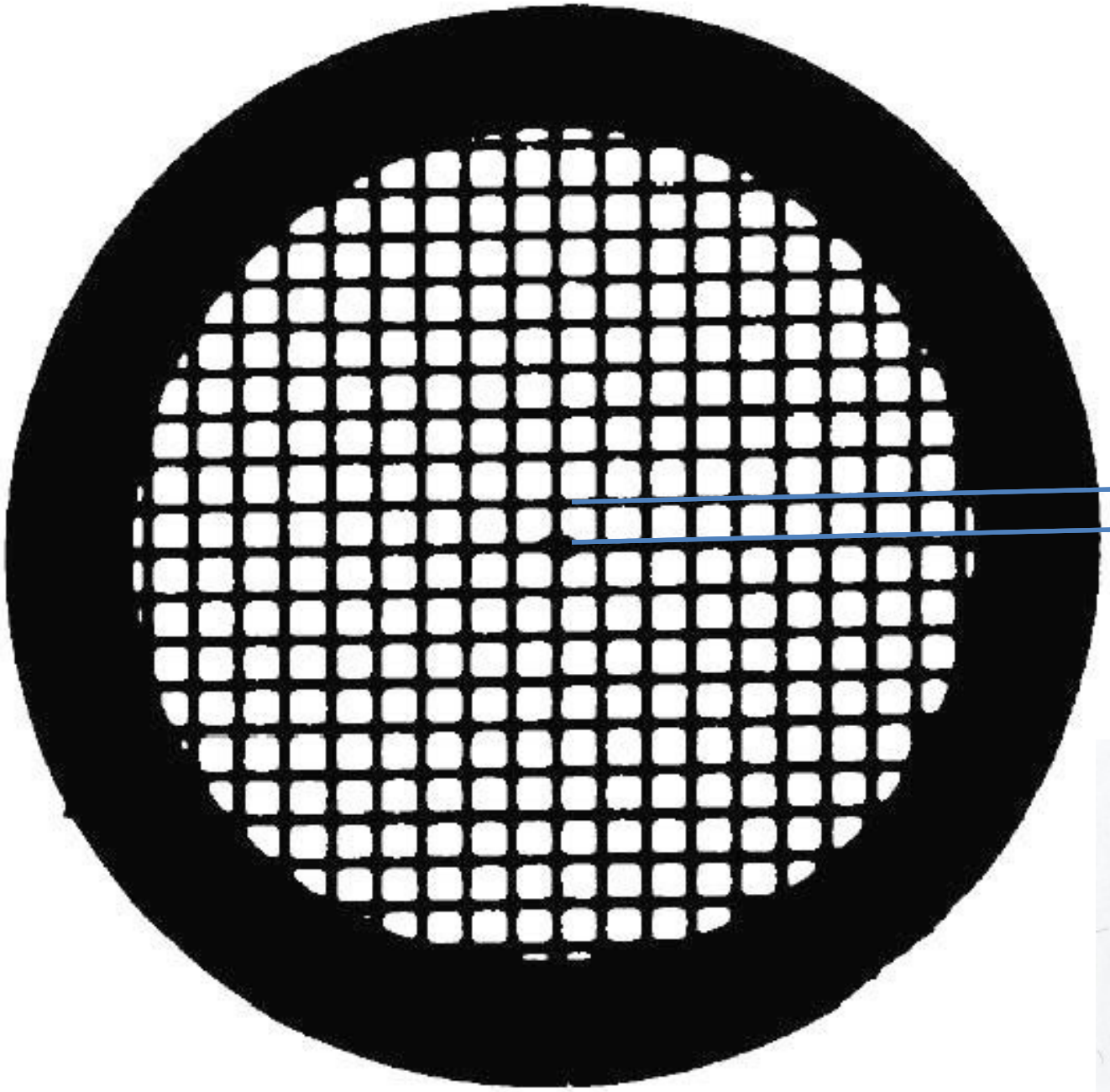


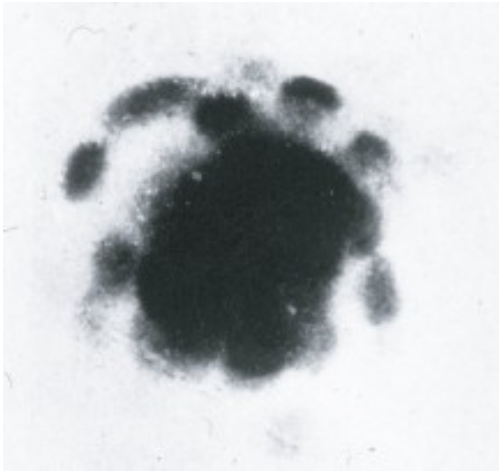




Wavelength comparison of X-rays and visible light



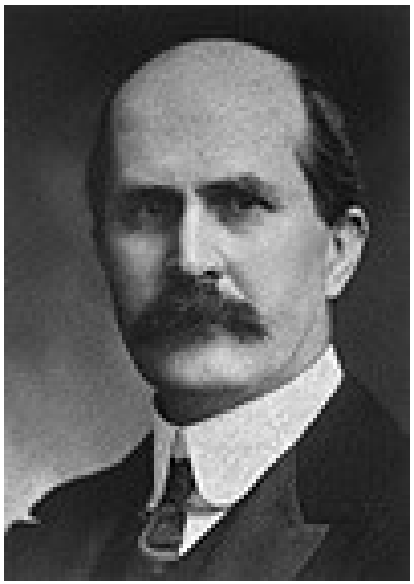

70µm

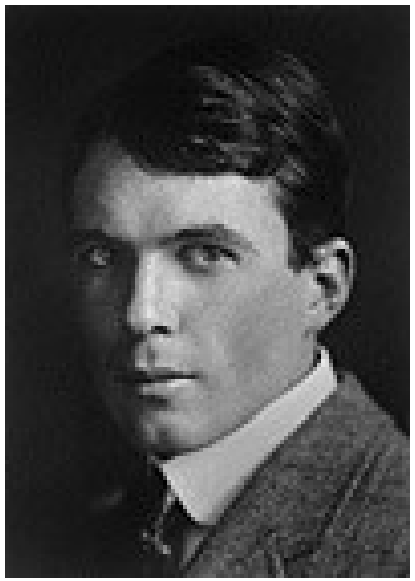


SIR WILLIAM HENRY BRAGG (1862-1942)
SIR WILLIAM LAWRENCE BRAGG (1890-1971)

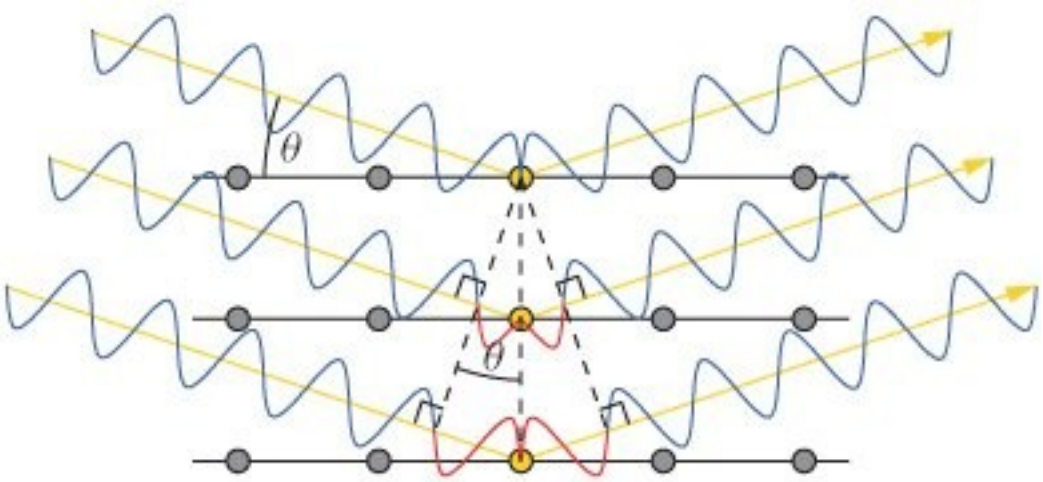
- **1915 Nobel Laureates in Physics**

  for the analysis of crystal structure by means of X-rays

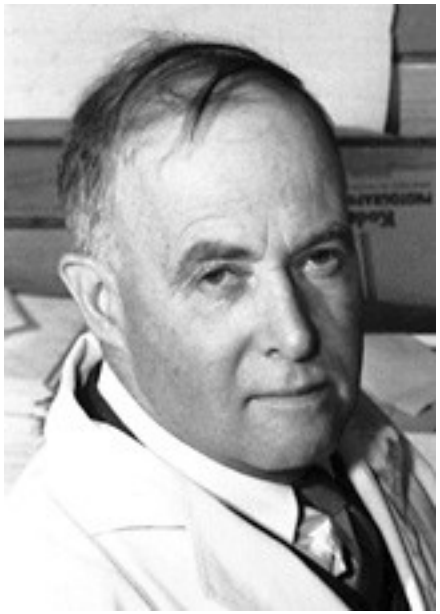




$$n\lambda = 2d \sin\theta$$

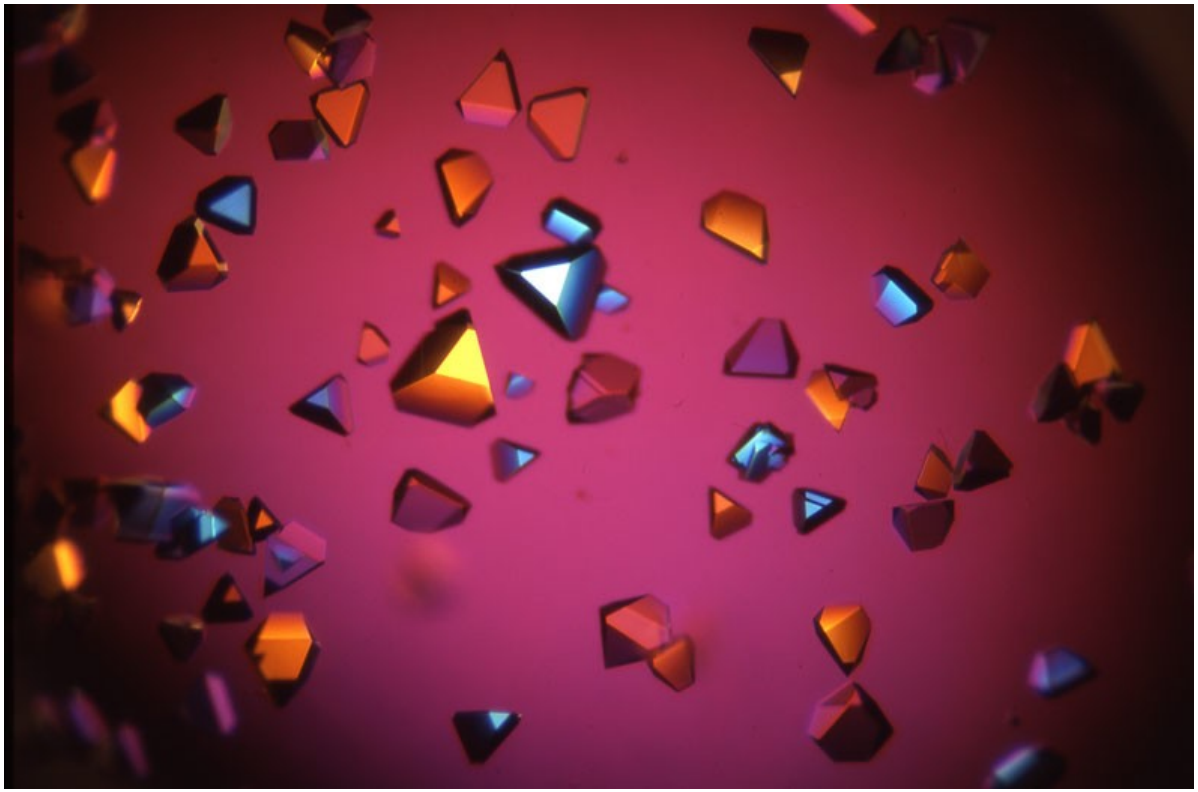


James Batcheller Sumner (1879-1960)



- **1946 Nobel Laureate in Chemistry**

  for his discovery that enzymes can be crystallized

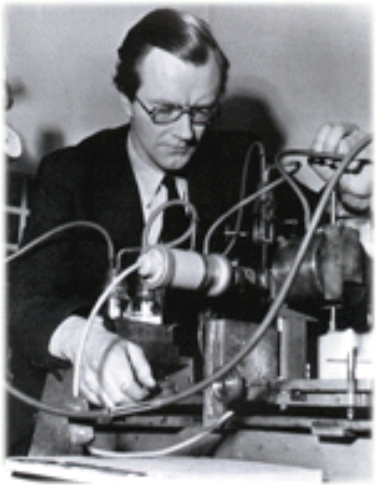


FRANCIS HARRY COMPTON CRICK (1916~2004)
JAMES DEWEY WATSON (1928~)
MAURICE HUGH FREDERICK WILKINS (1916~2004)

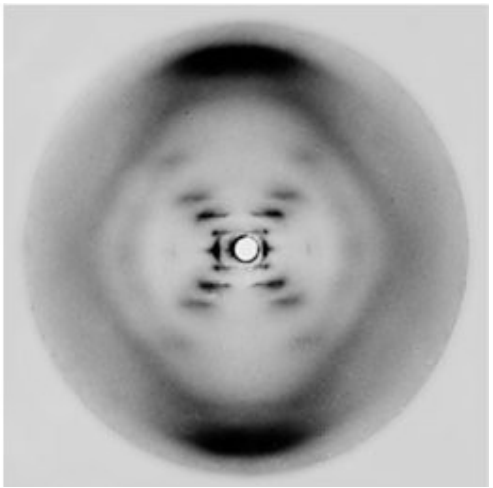
- **1962 Nobel Laureates in Physiology and Medicine**

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

James Watson
and Francis Crick

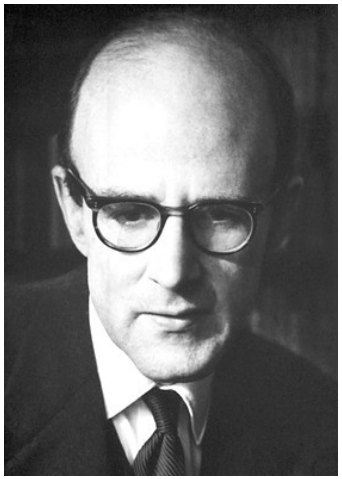


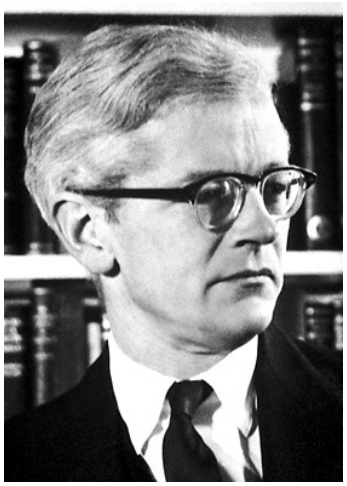
Maurice Wilkins



Rosalind Franklin

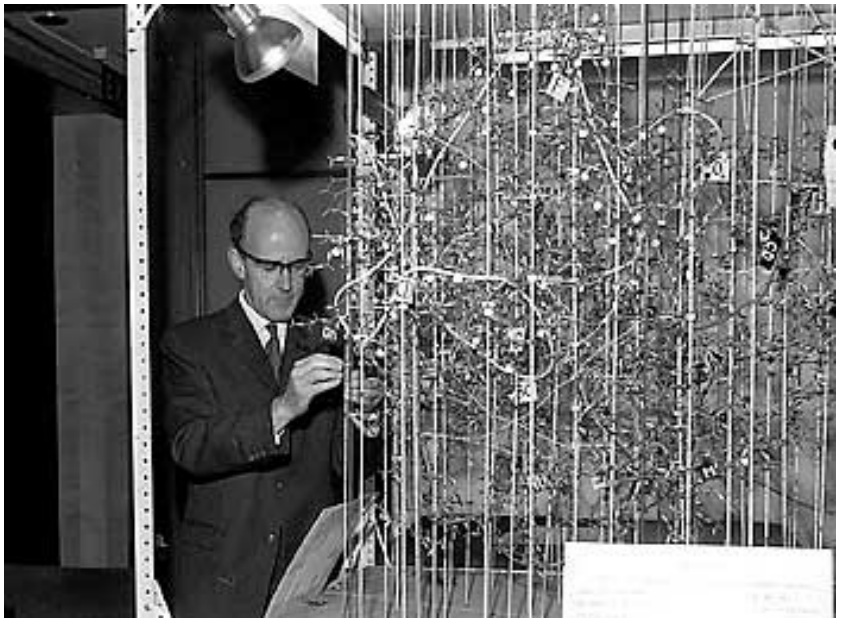
Max Ferdinand Perutz (1914 – 2002)
John Cowdery Kendrew (1917 – 1997)



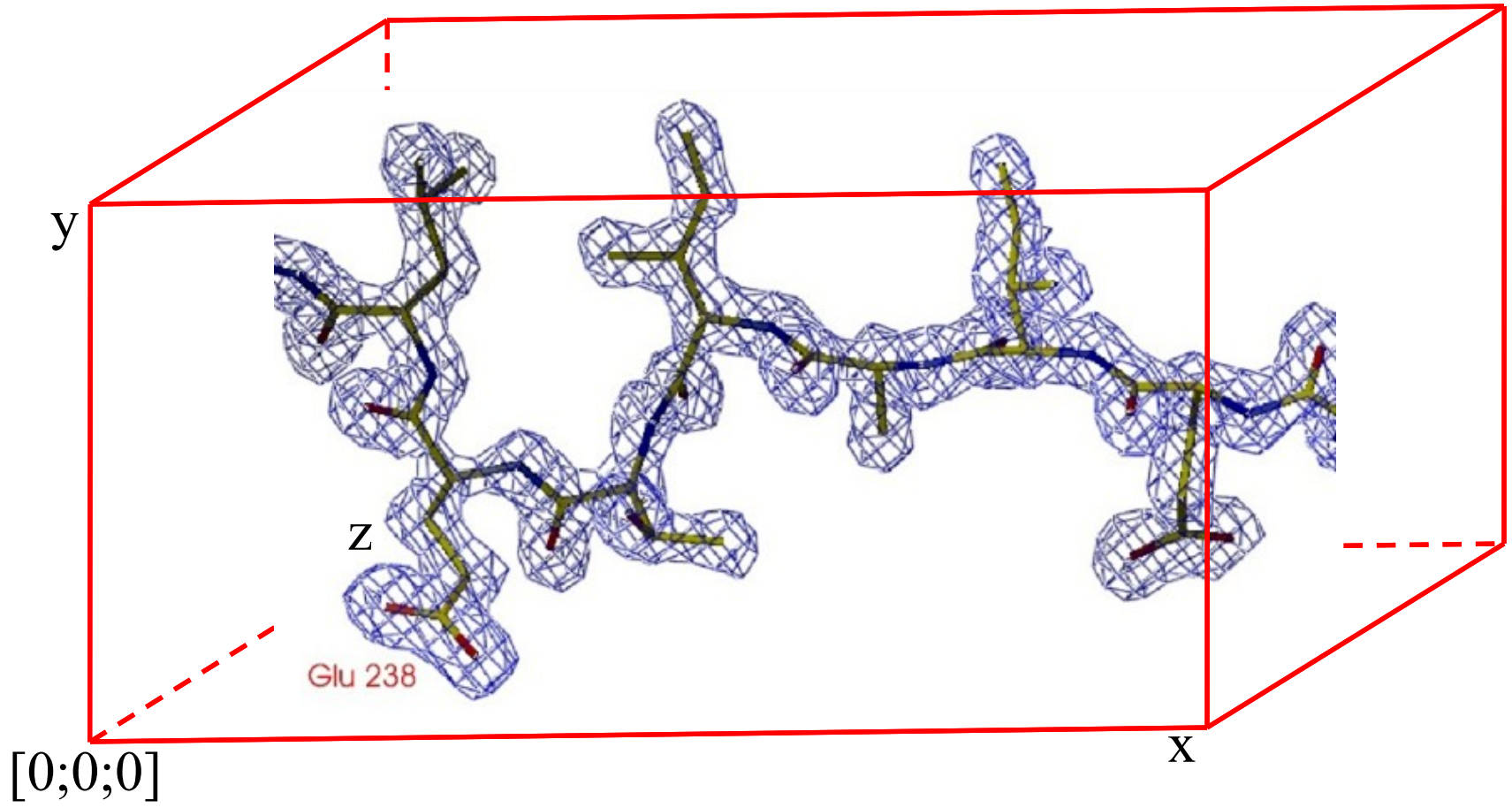


- **1962 Nobel Laureates in Physics**

  for their studies of the structures of globular proteins

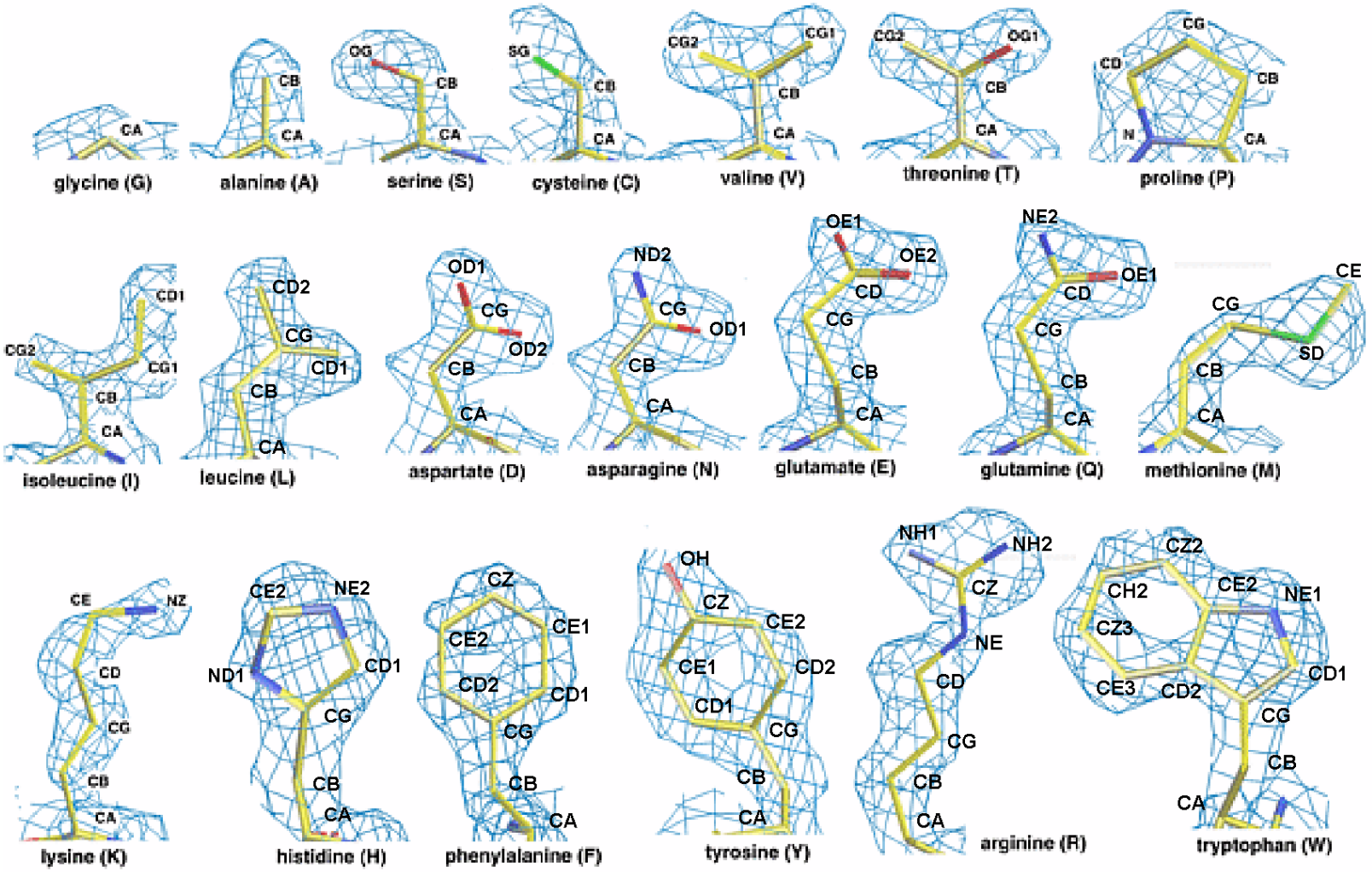


Information from X-ray diffraction experiment

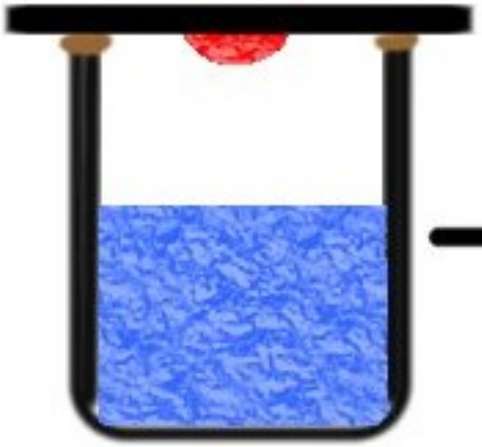


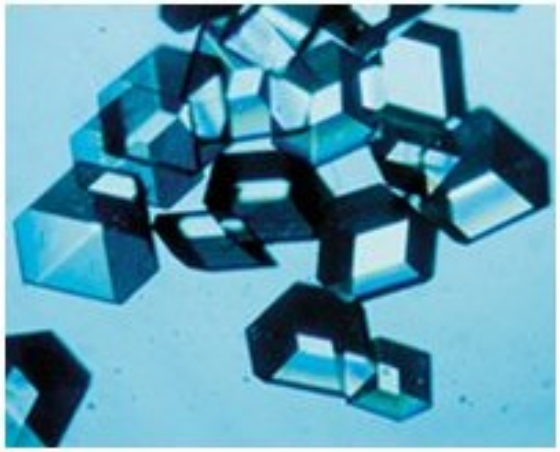
$$\rho(x\,y\,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F(h\,k\,l)\right|\exp\left[-2\pi i(hx+ky+lz)+i\alpha(h\,k\,l)\right]$$

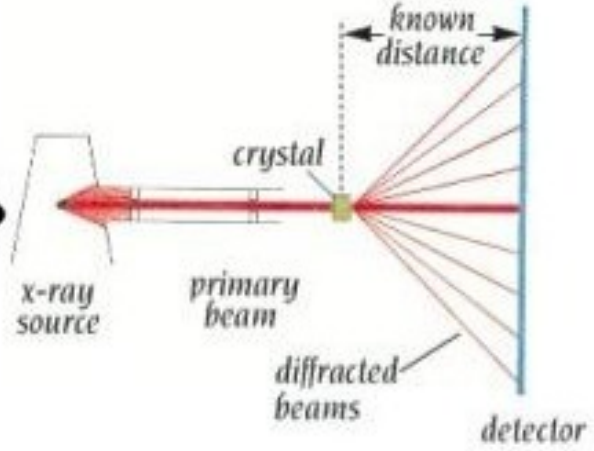
Representative electron density for amino acid side chains

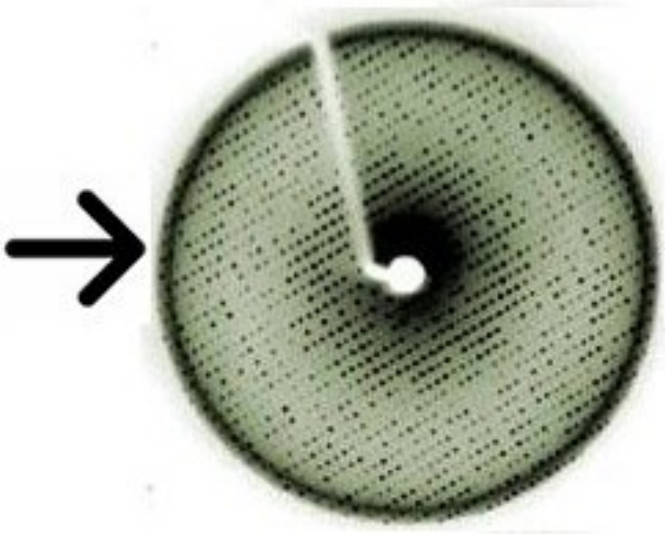


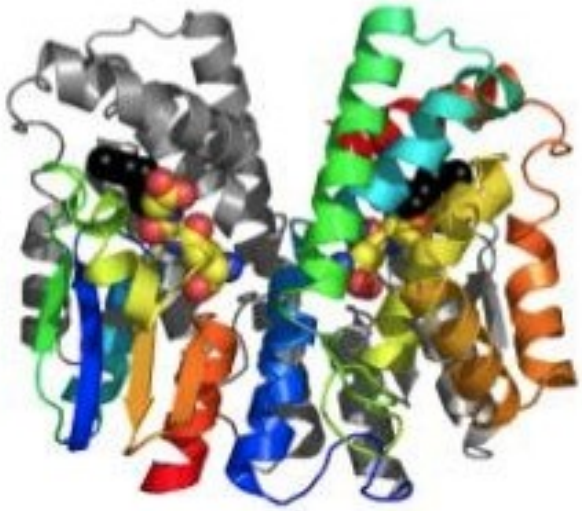
Electron density maps calculated at 1.5 Angstrom resolution.



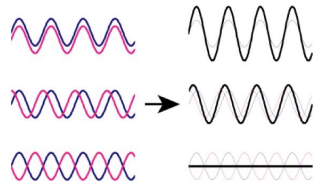


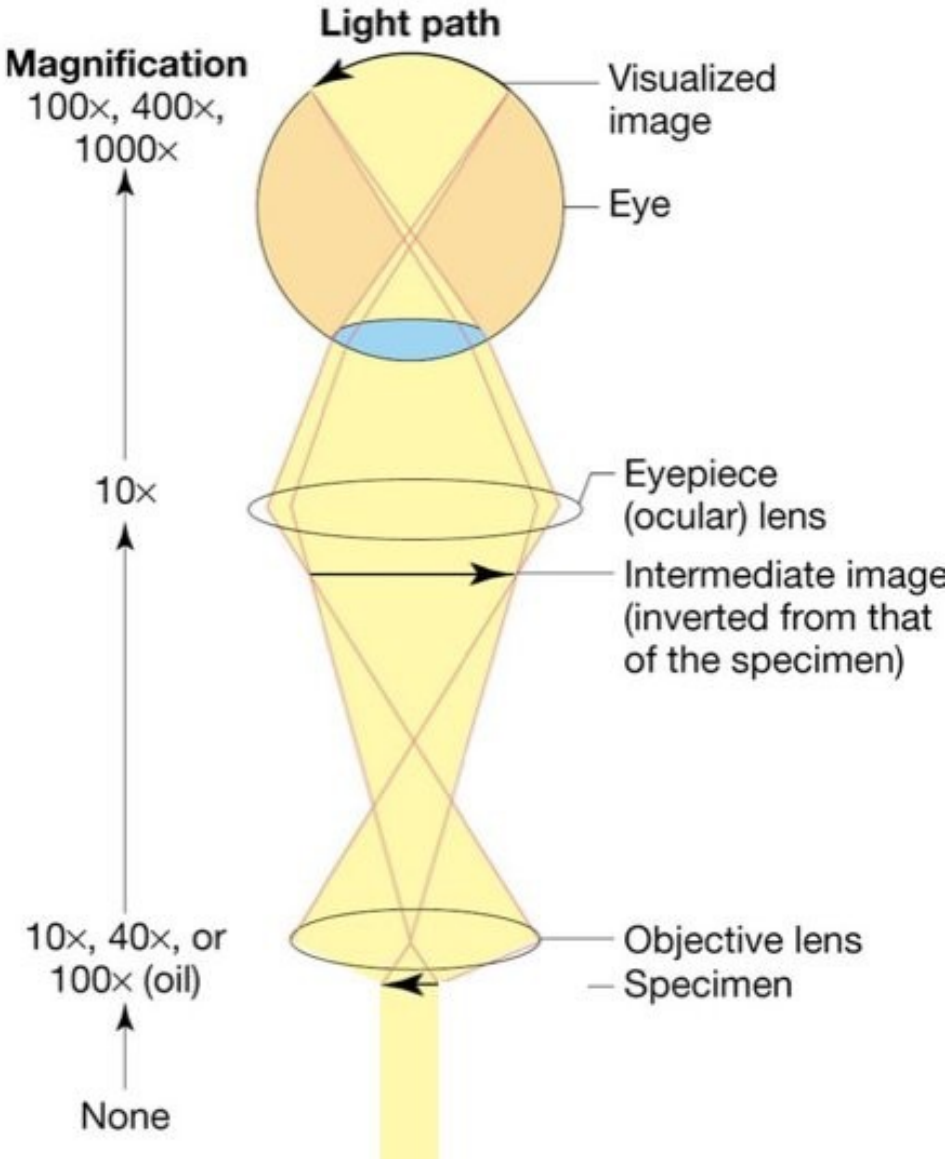

known
distance
crystal
x-ray
source
primary
beam
diffracted
beams
detector

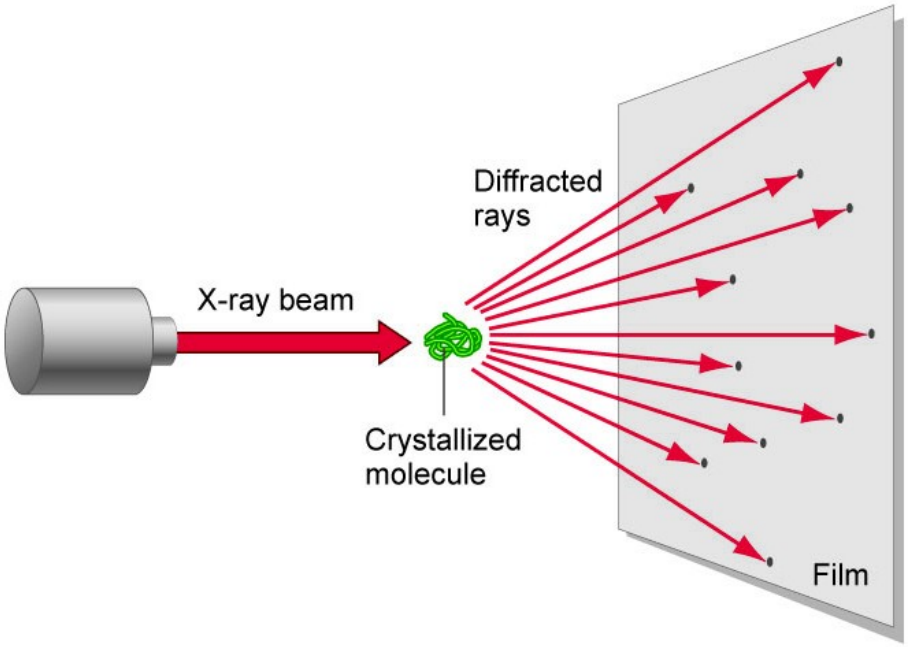




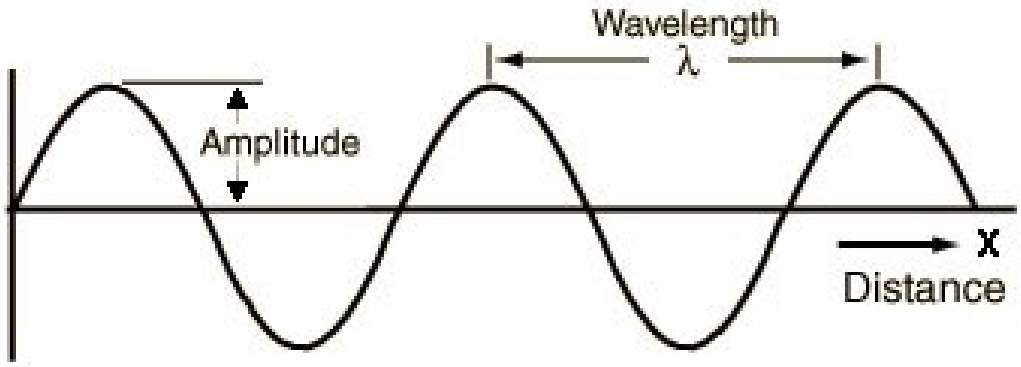
Comparison of microscope and diffraction

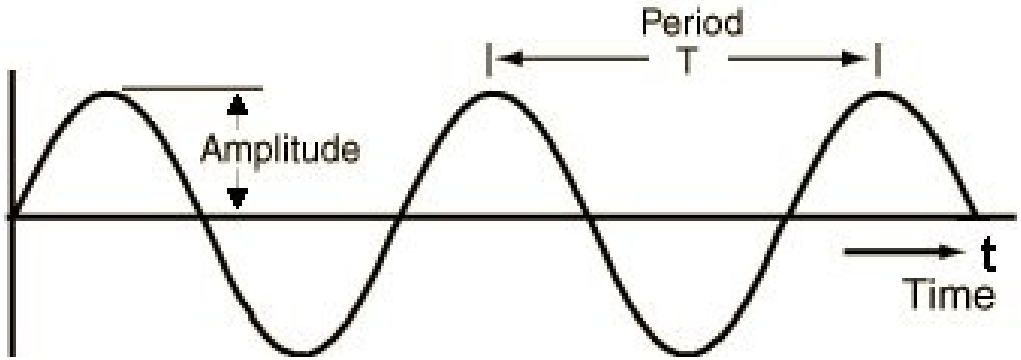




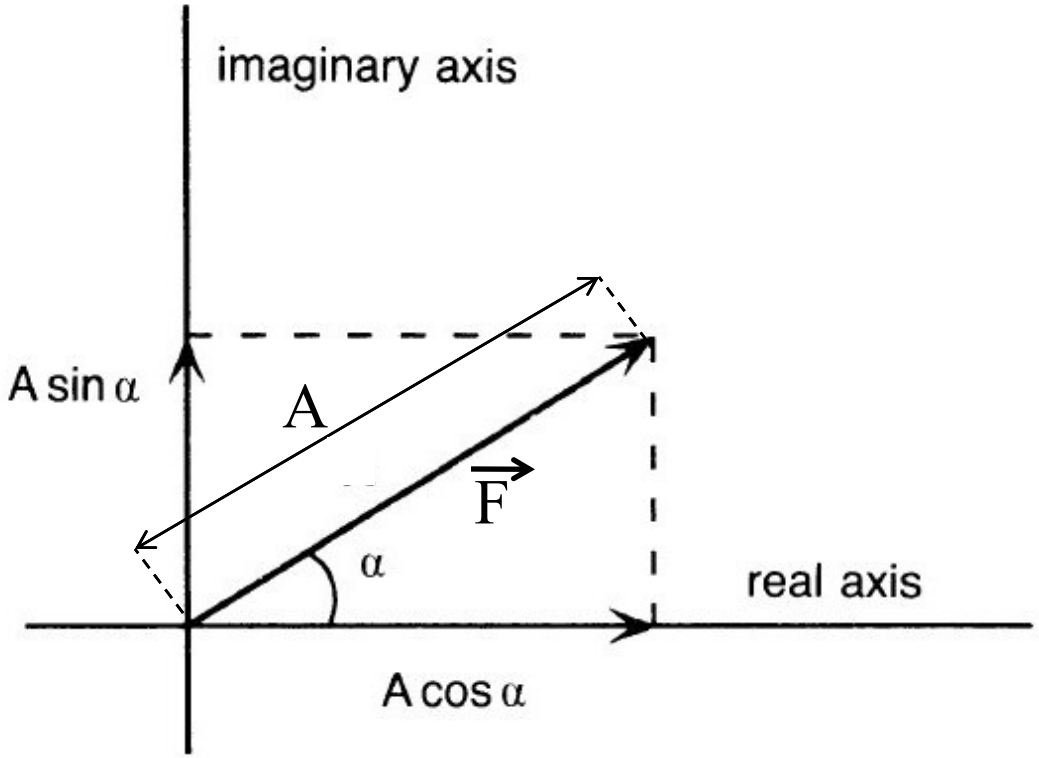


Waves and Radiation





Wave as a vector

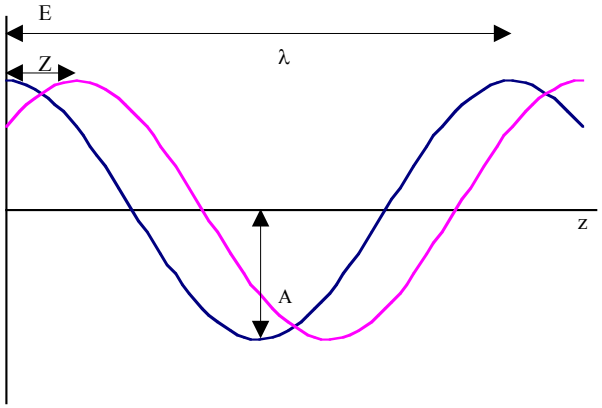


A- wave amplitude

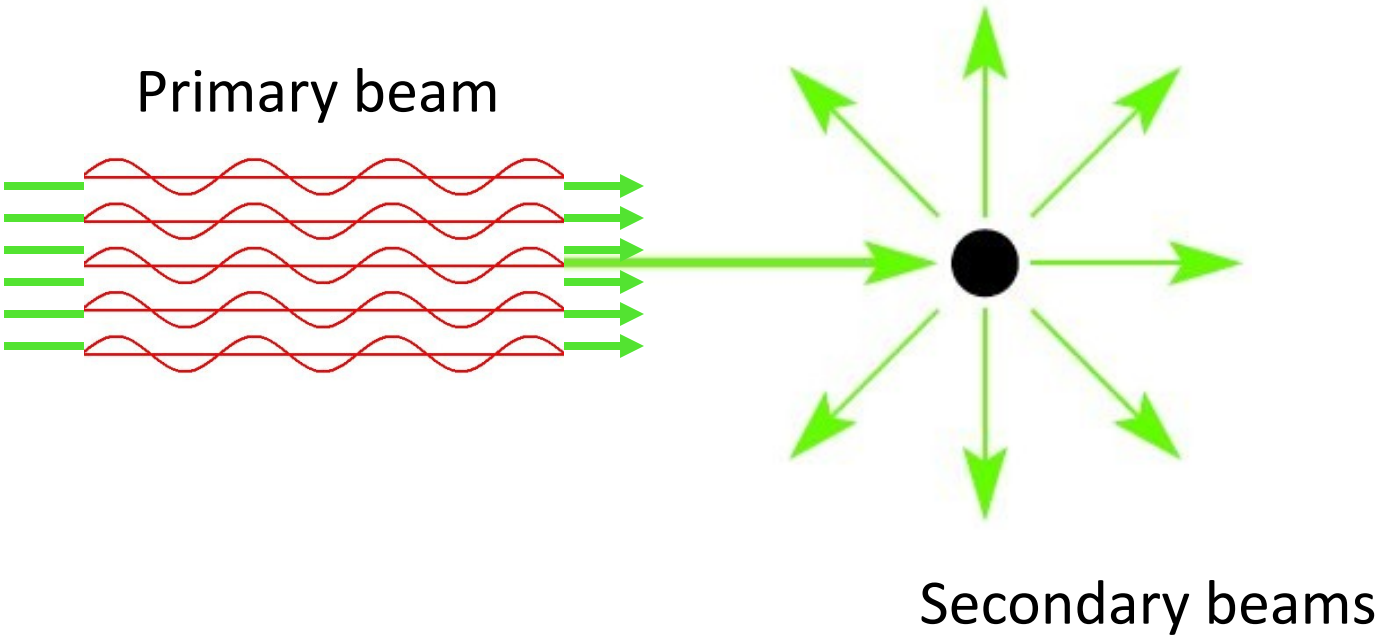
α- wave phase

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

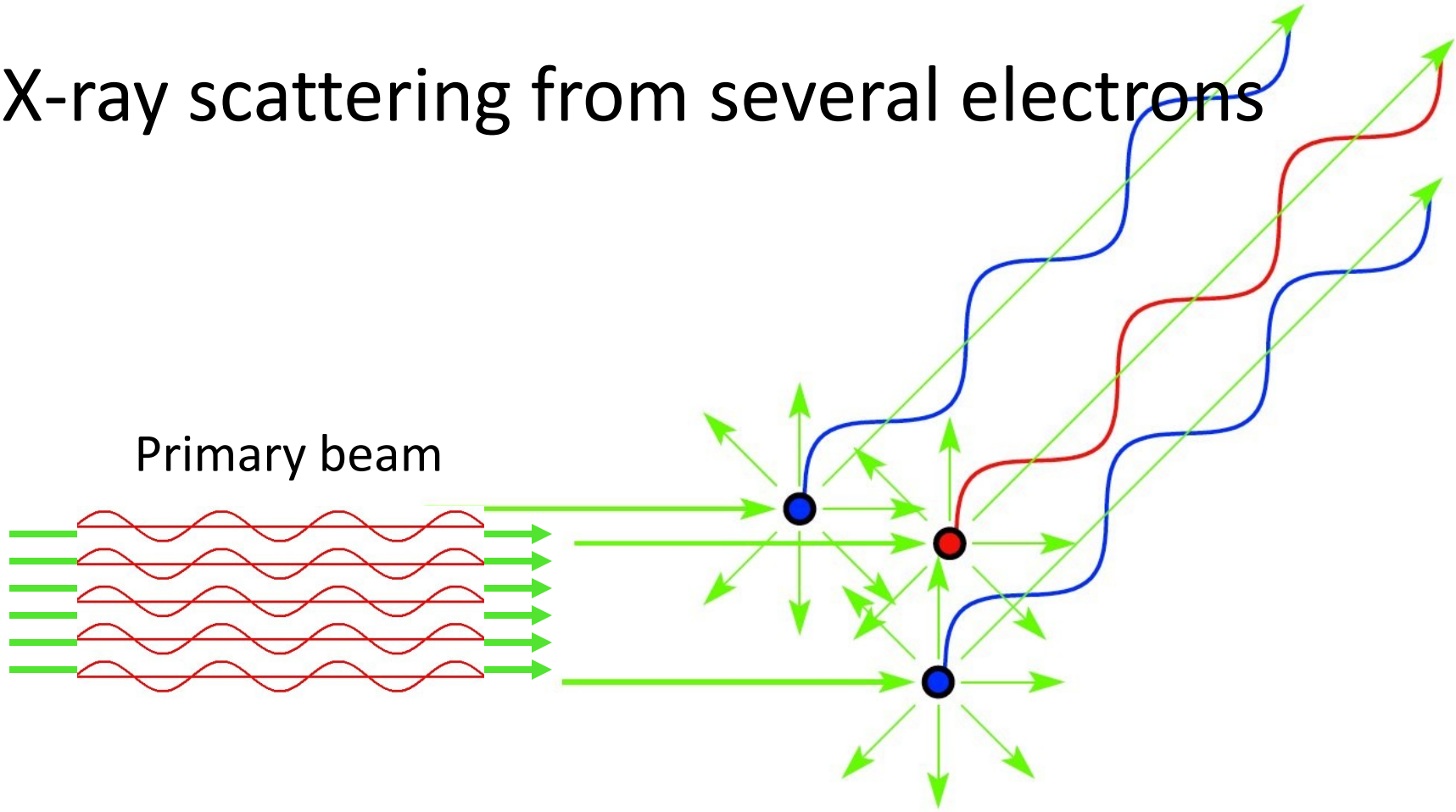
$\vec{F}=A\exp(i\alpha)$



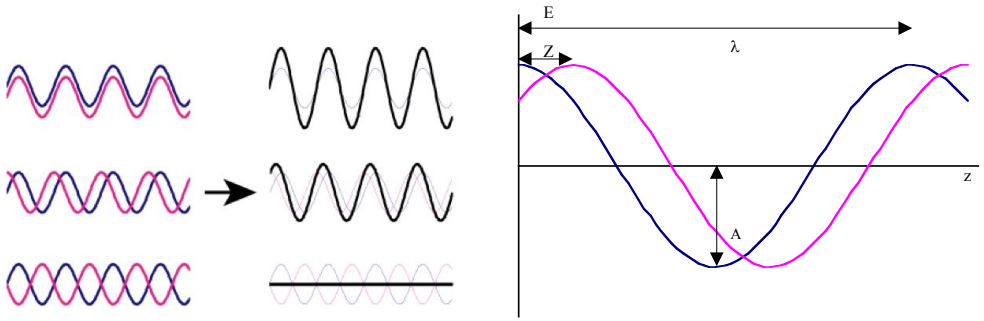
X-rays scatter from electrons in all directions



X-ray scattering from several electrons



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

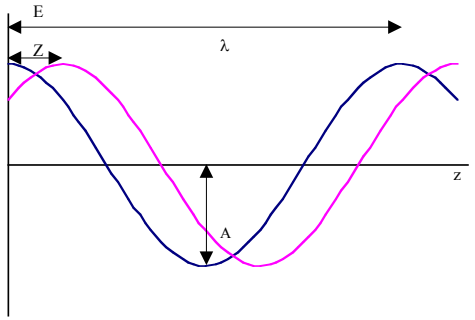


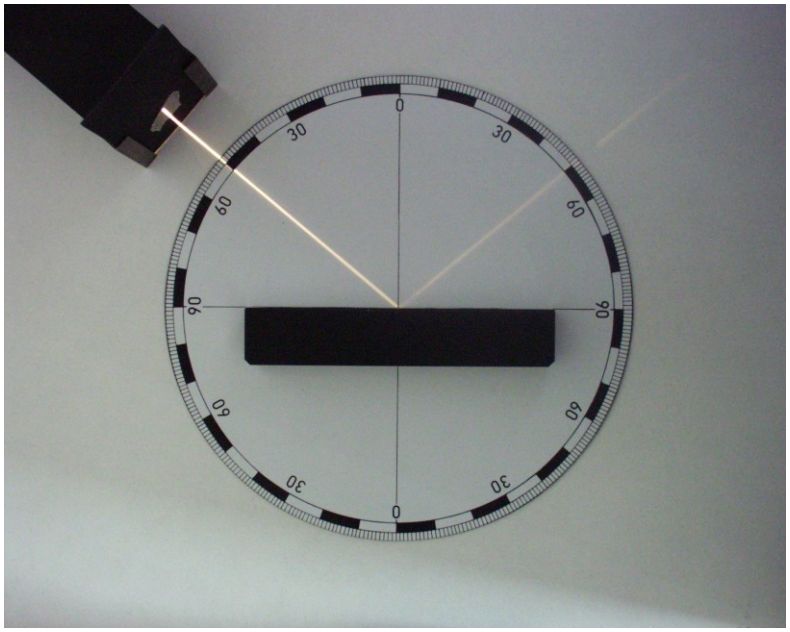
- Scattering from a single molecule is not detectable

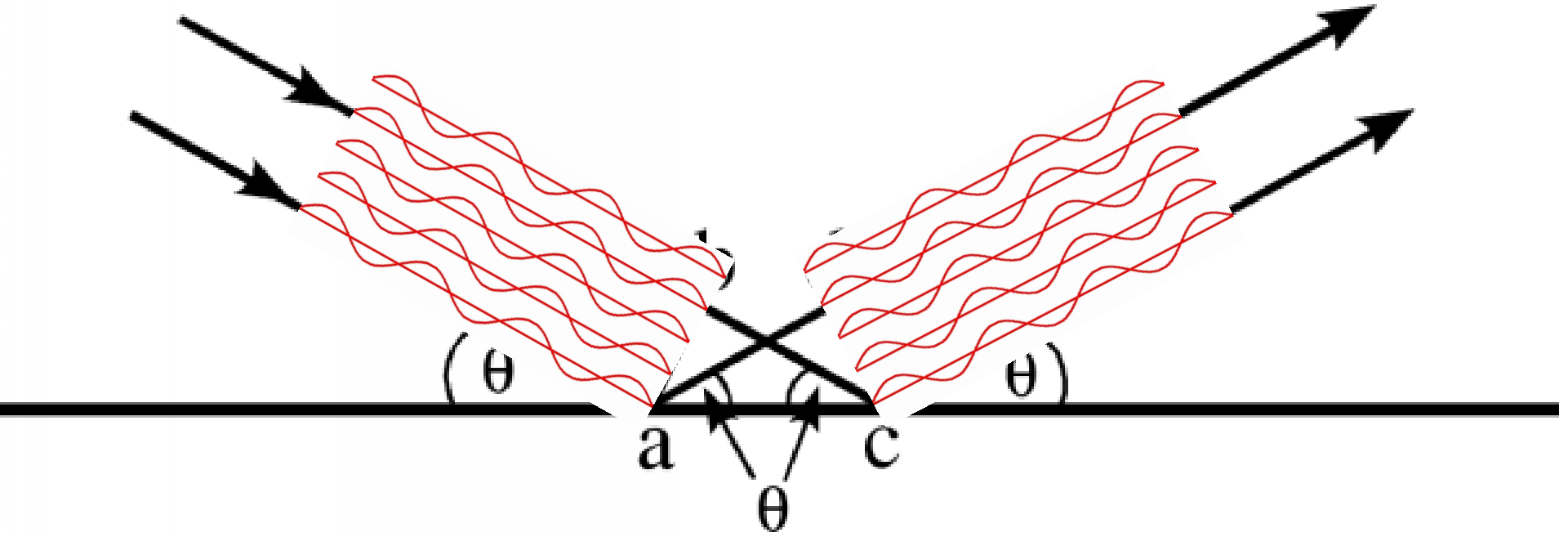
- If molecules are all oriented in the same way, the scattering from individual molecules will add in certain directions

  –Which directions?

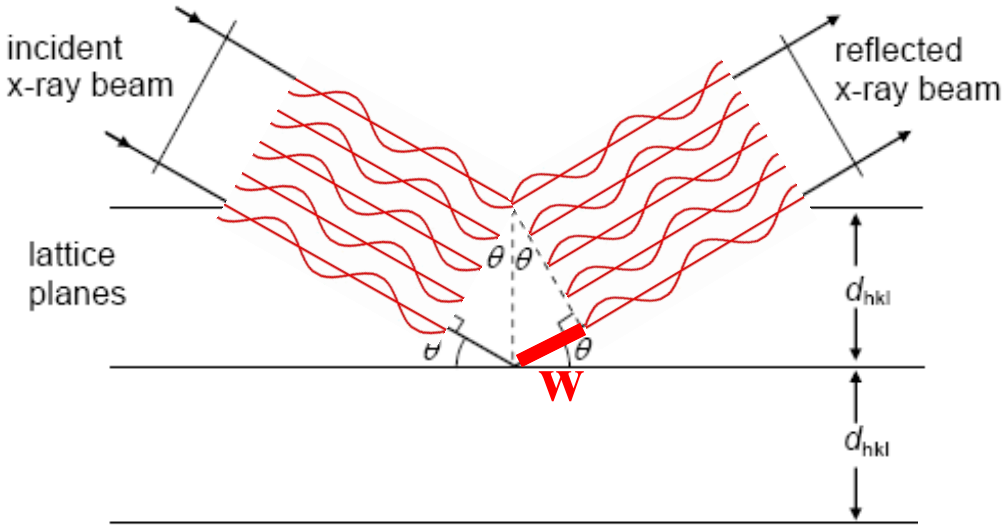
There is no path and PHASE DIFFERENCE when rays reflect from a plane





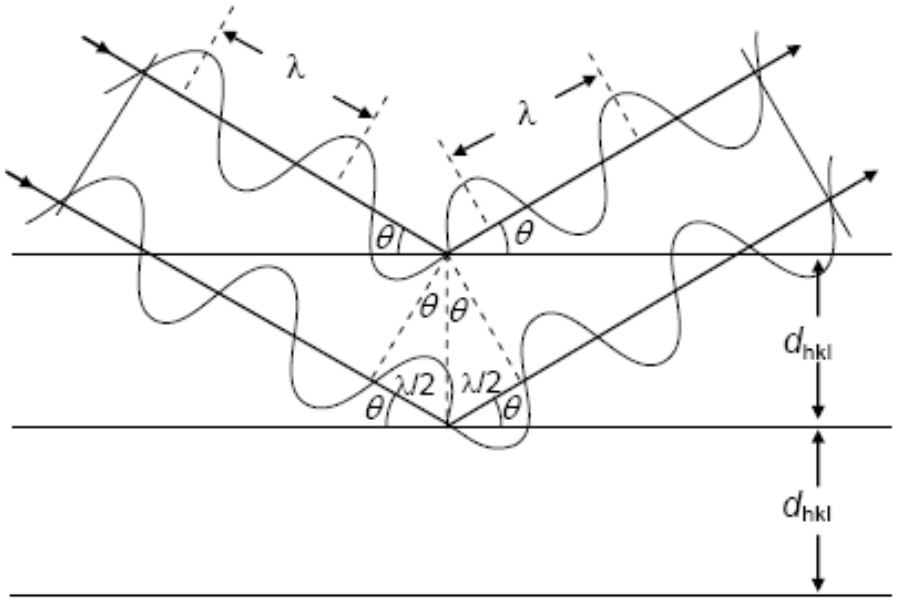


There is NO PHASE DIFFERENCE if the path differences are equal to whole number multiplies of wavelength (λ)



Bragg's law:

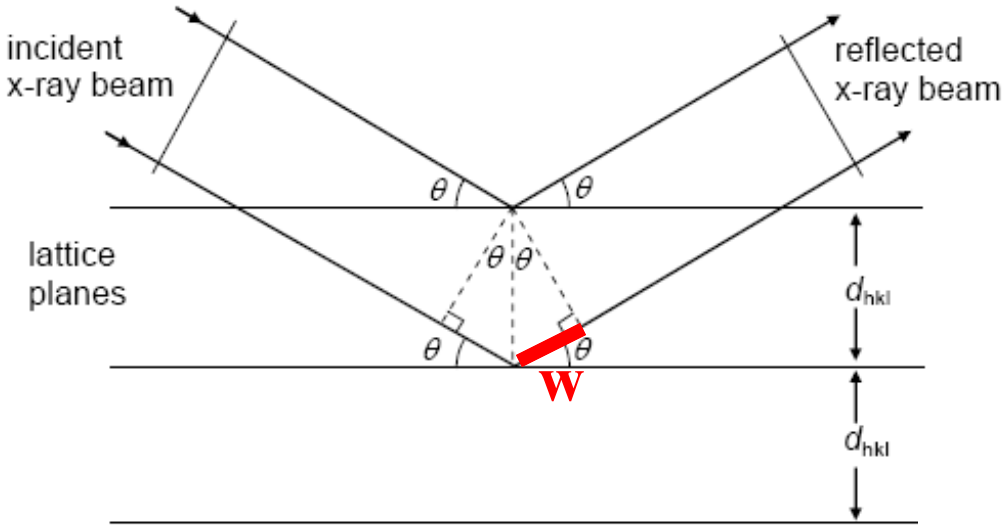
$$n\lambda = 2d \sin\theta$$



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

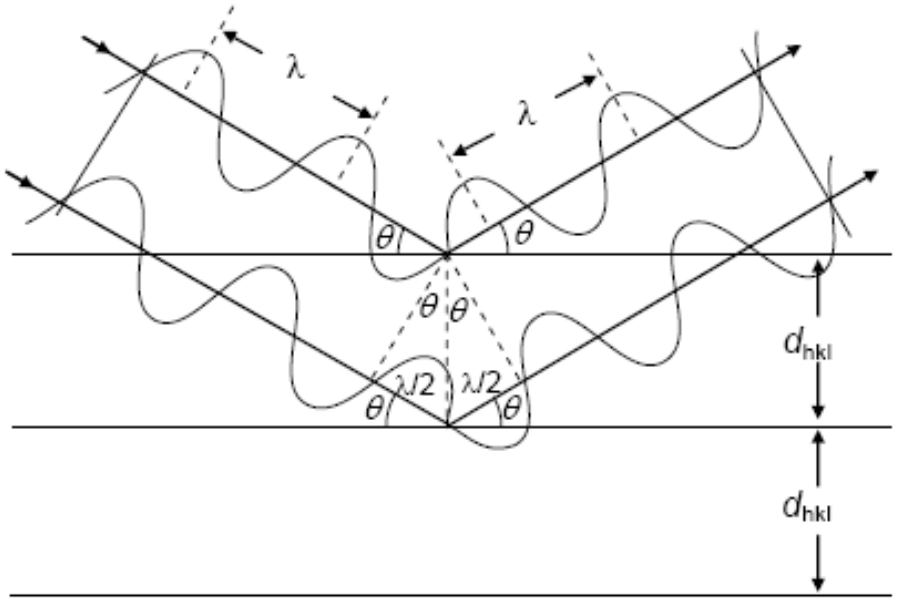
$2\mathbf{w} = n\lambda$

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



Bragg's law:

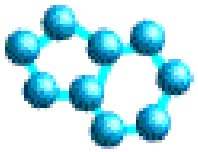
$$n\lambda = 2d \sin\theta$$



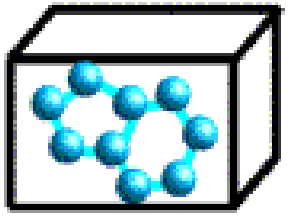
$\sin\theta = \mathbf{w}/d$

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

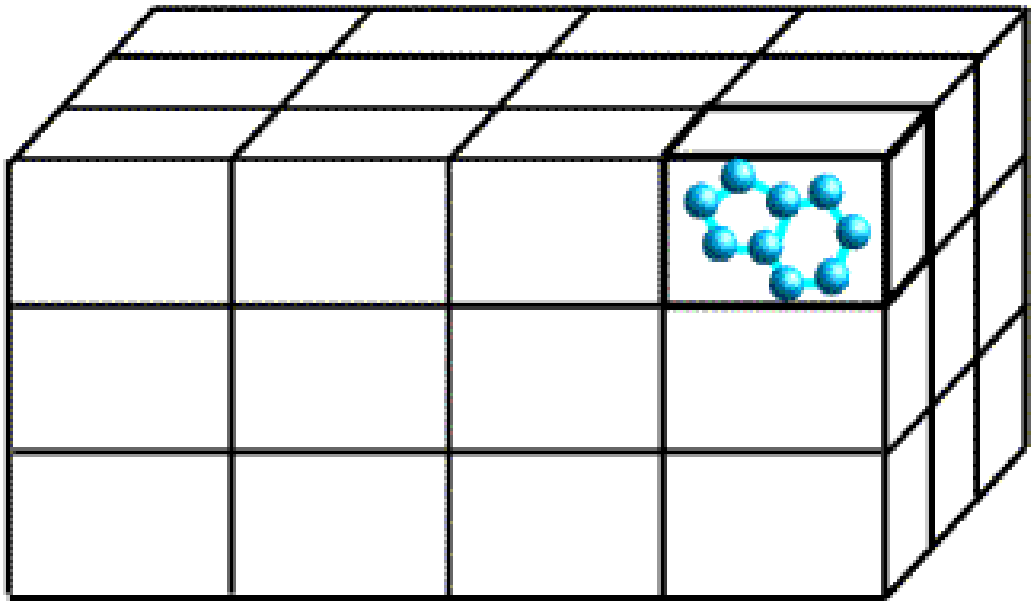
molecule

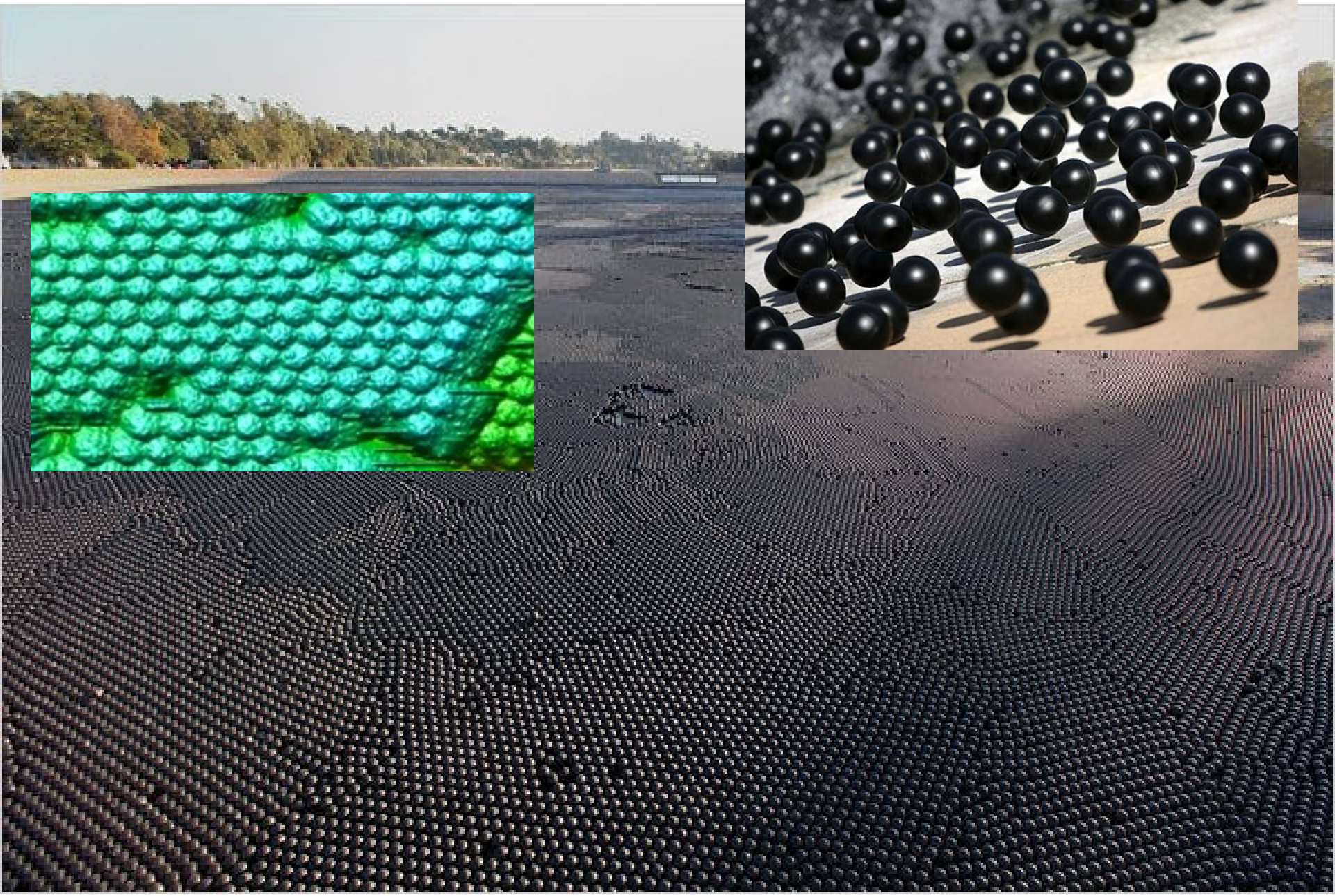


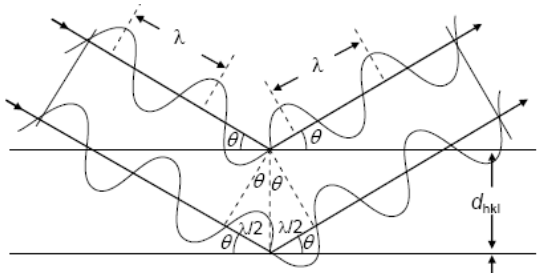
unit cell

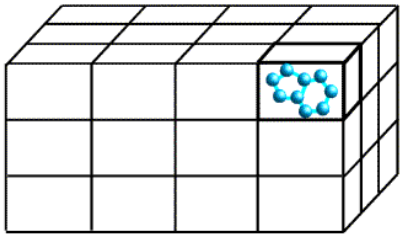


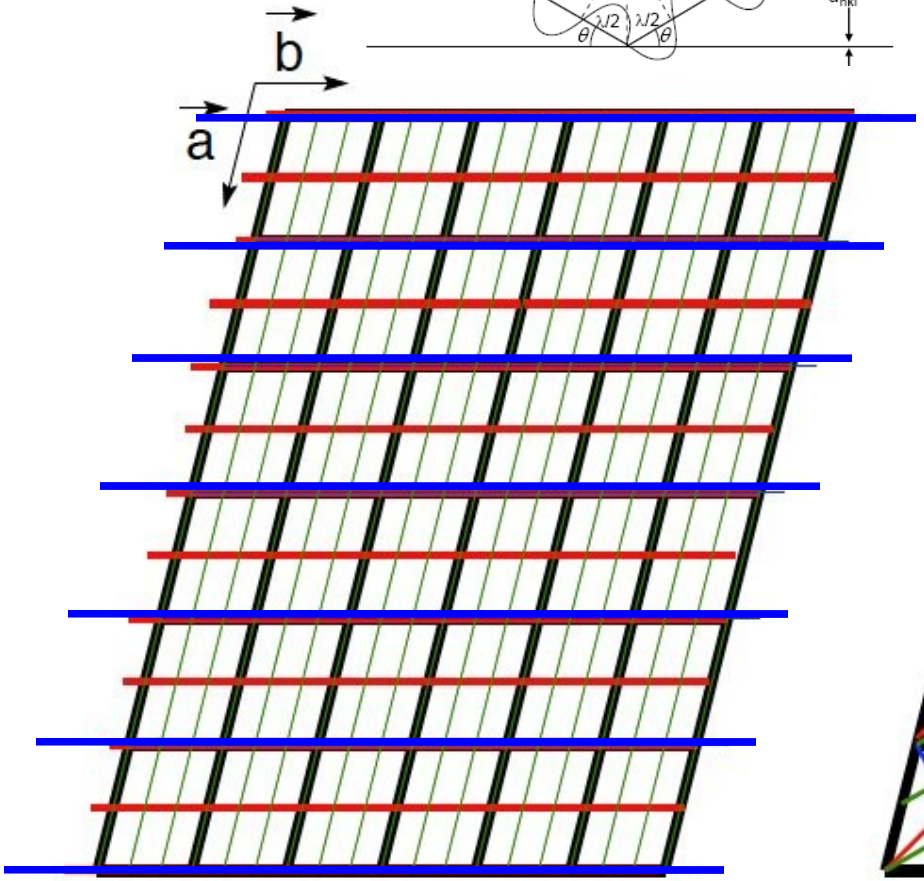
crystal

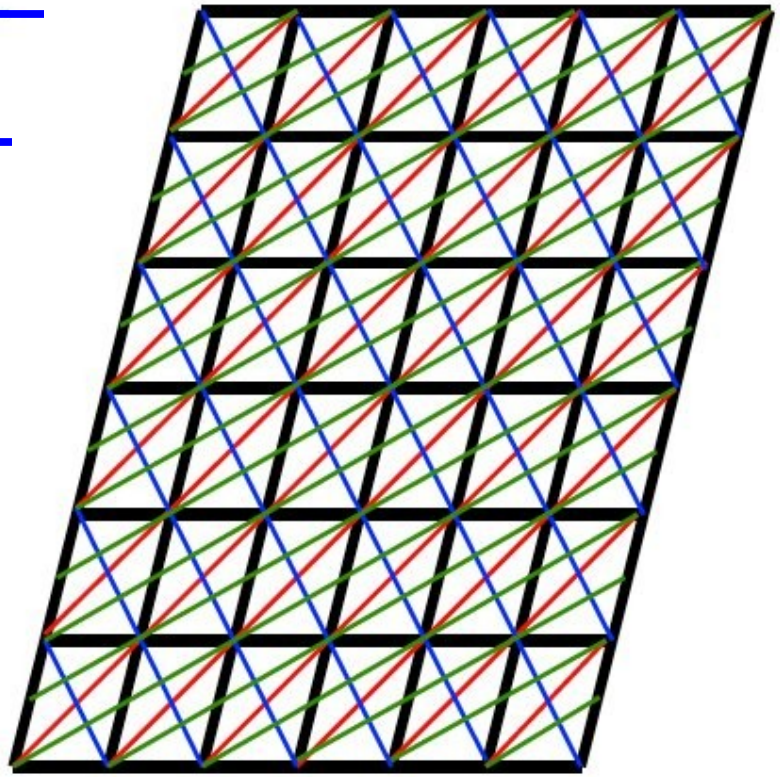


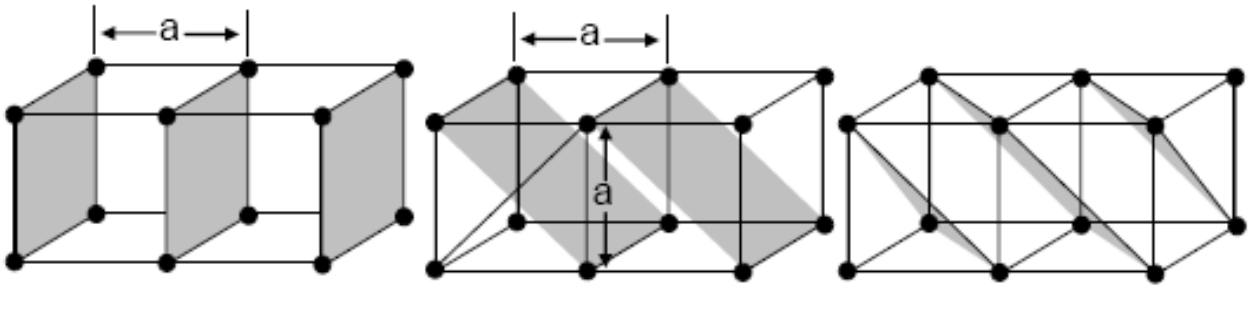






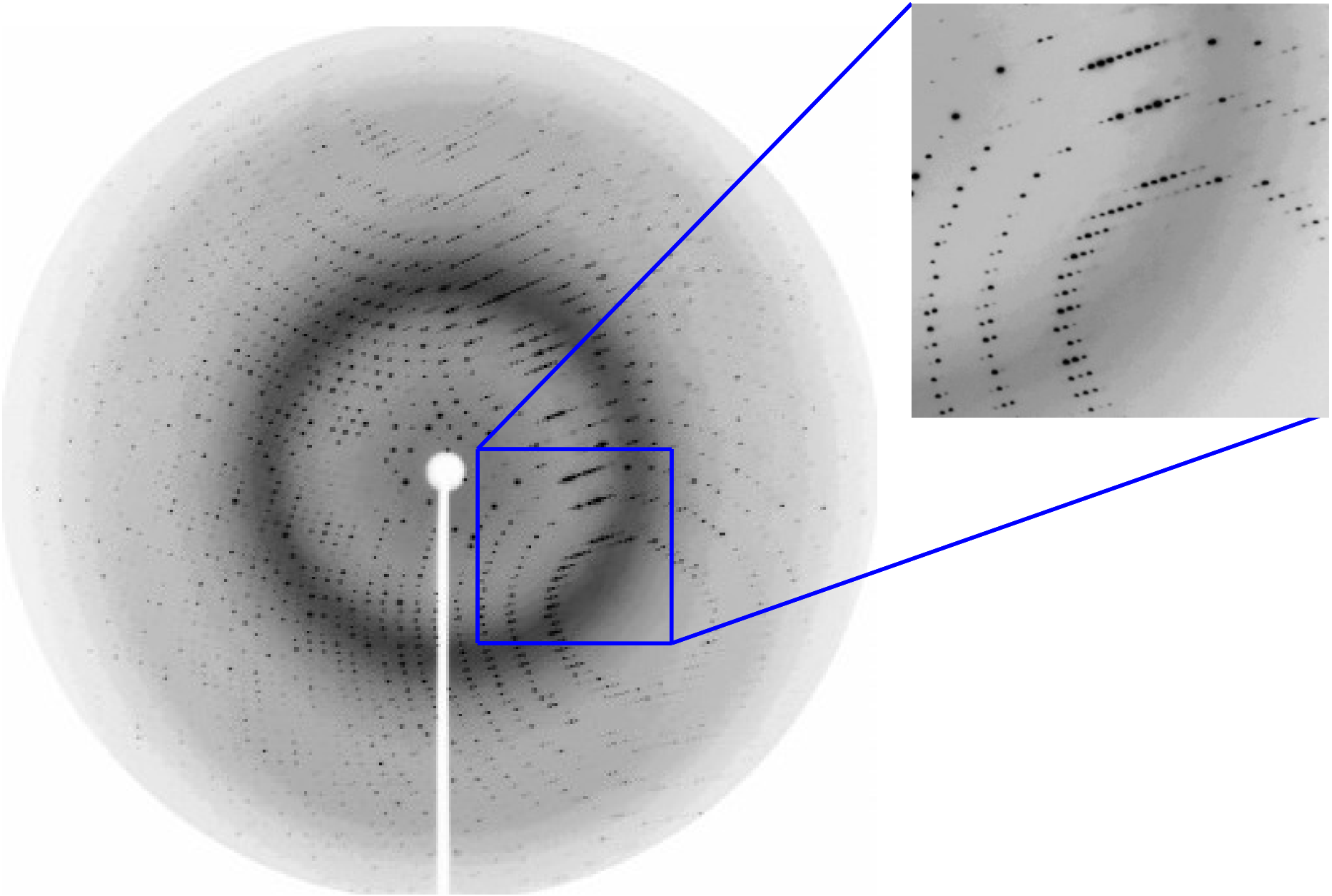


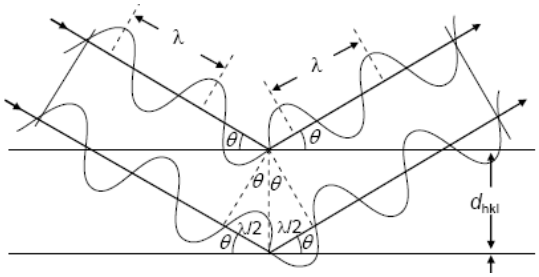


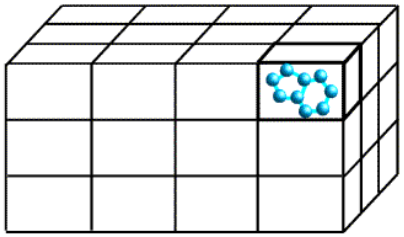


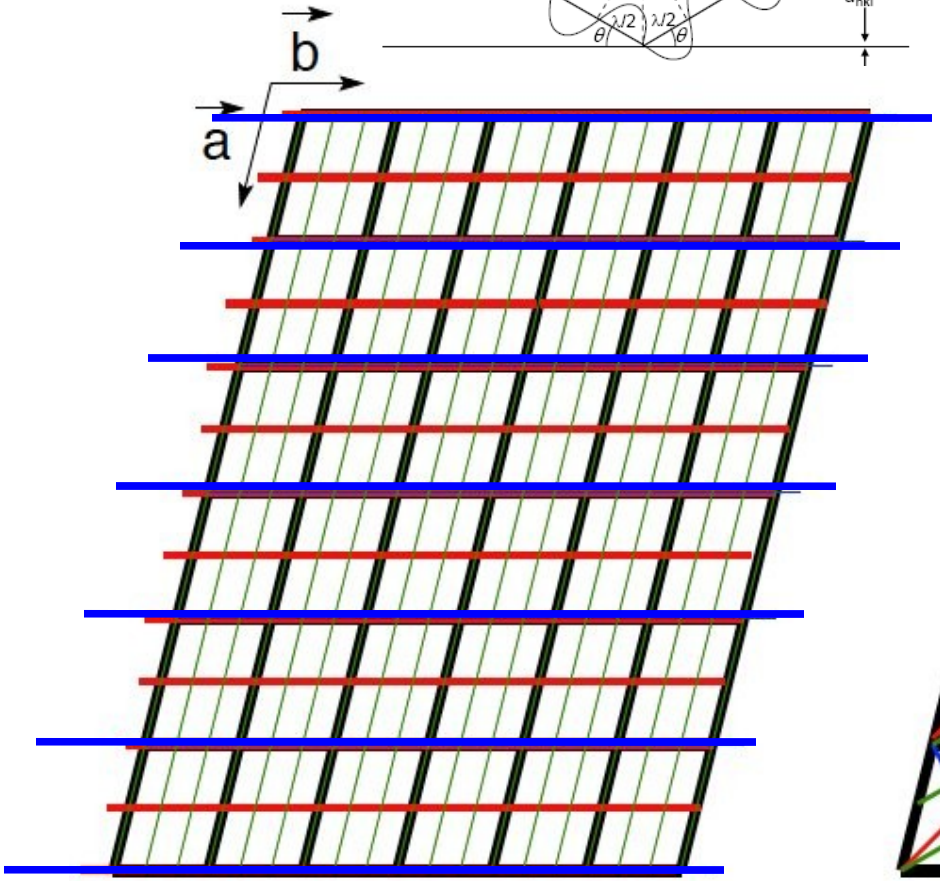
(h, k, l)

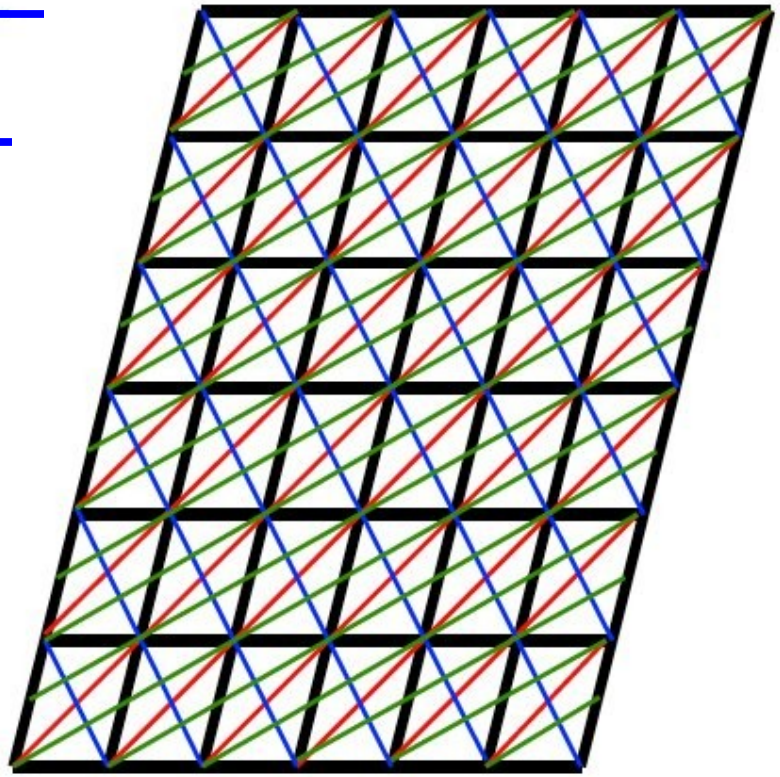
Diffraction pattern from a protein crystal

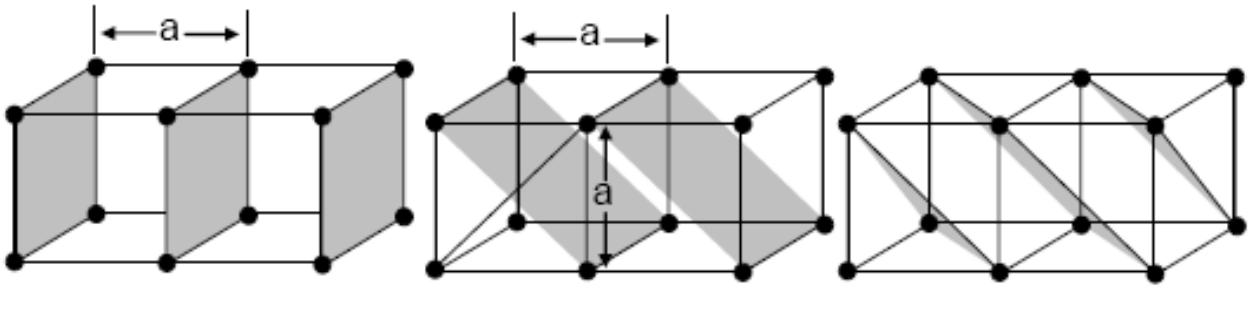




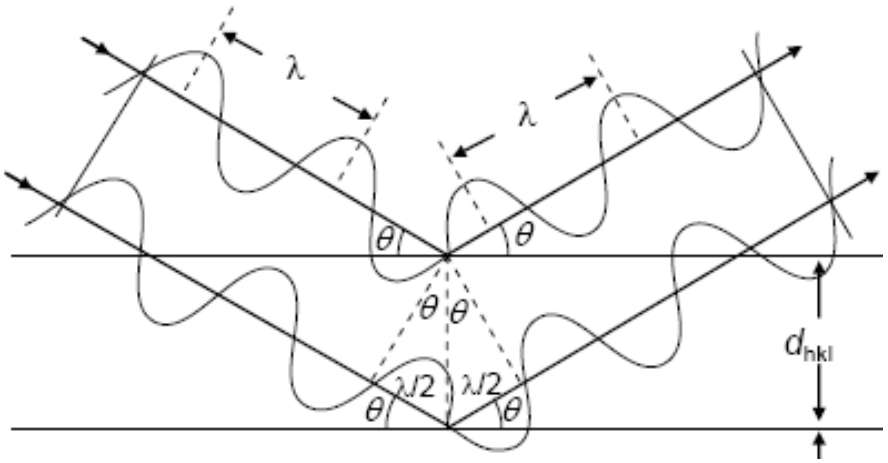


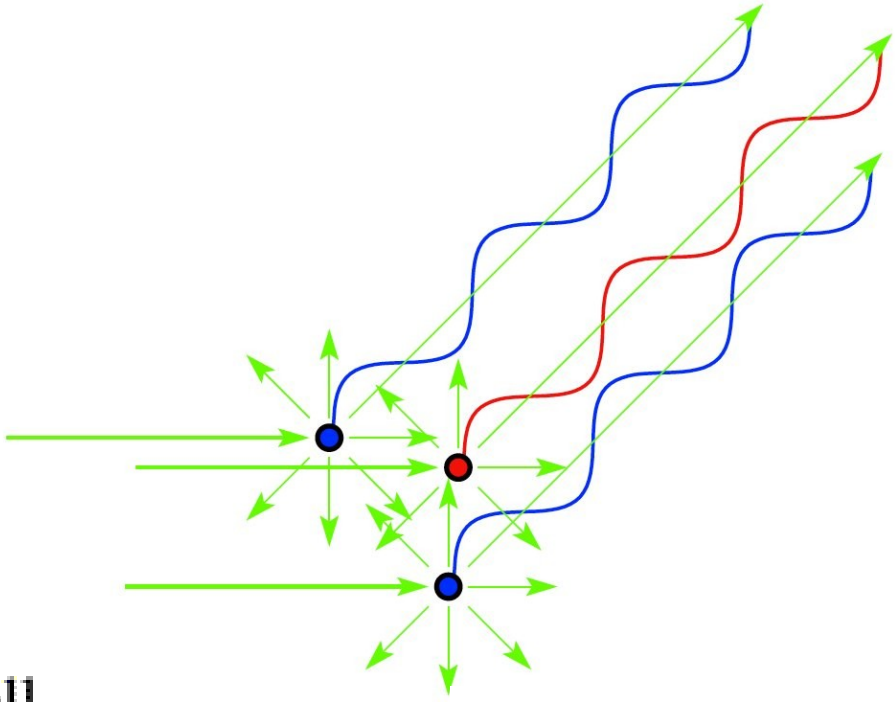






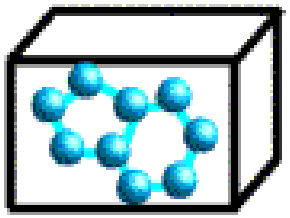
(h, k, l)

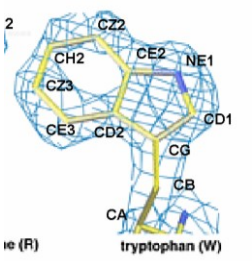


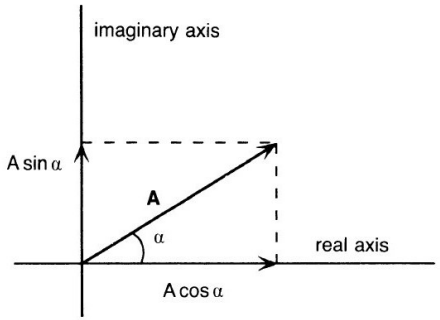


$$n\lambda = 2d \sin\theta$$

unit cell







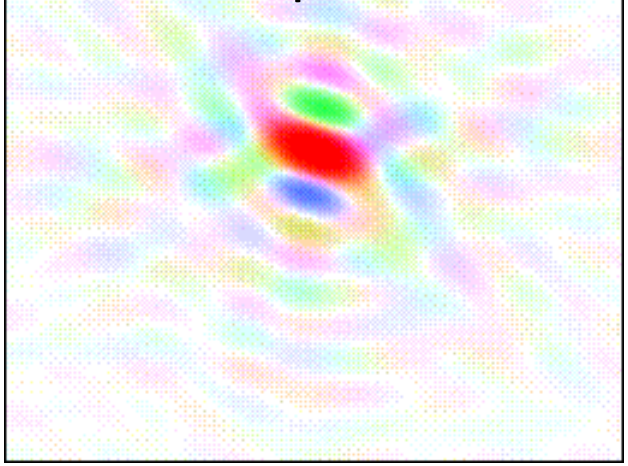
$$F(h\,k\,l) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x\,y\,z) \exp[2\pi i(hx + ky + lz)]\,dx\,dy\,dz$$

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

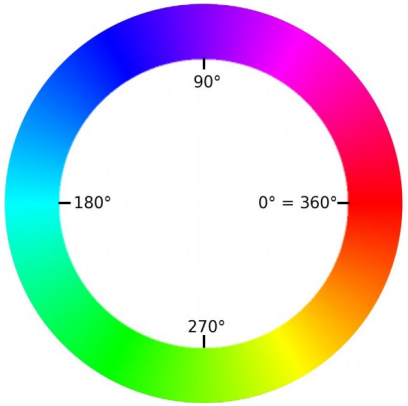
Real space cat

Fourier transform

Fourier amplitudes and phases

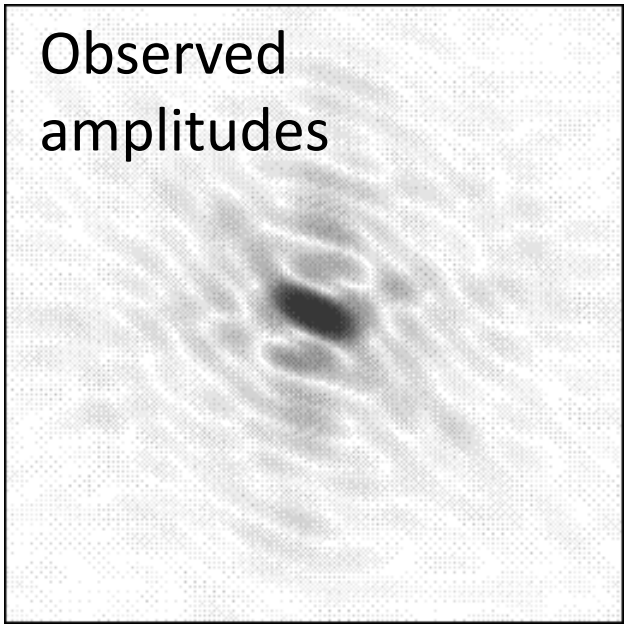


Circular rainbow scale of phases



Linear intensity scale of amplitude size

Observed amplitudes

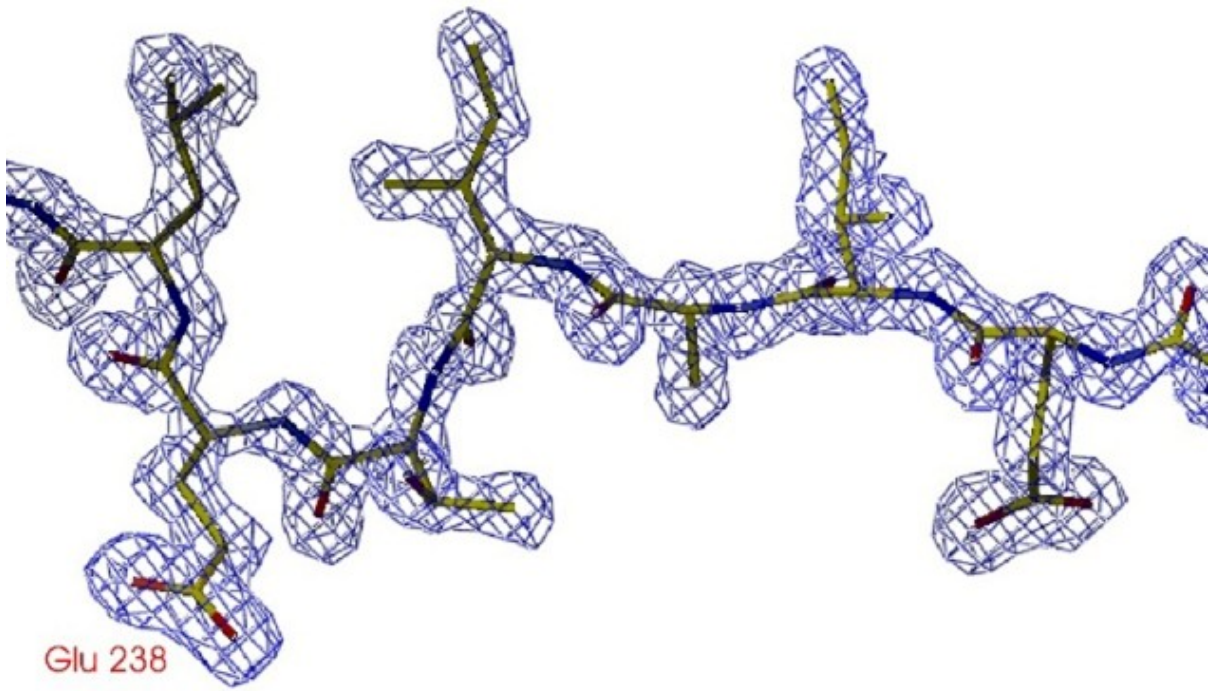


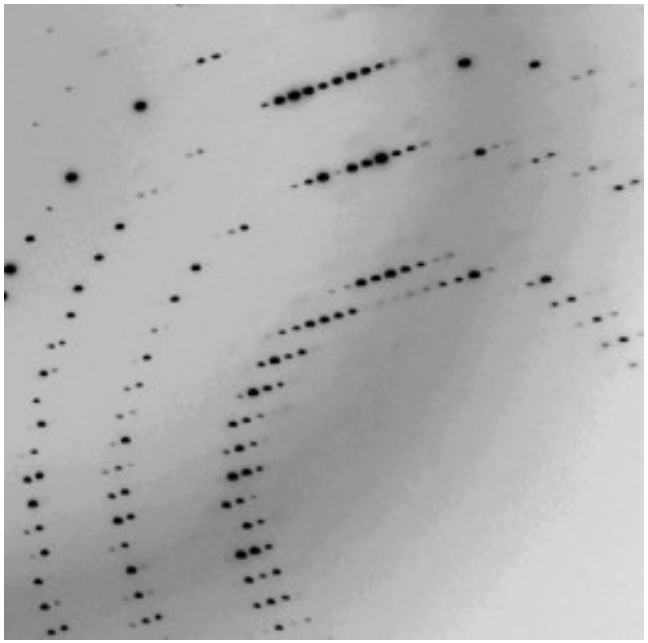
$$F(h\,k\,l) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x\,y\,z) \exp[2\pi i(hx + ky + lz)]\,dx\,dy\,dz$$

Electron density equation + PHASE PROBLEM

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

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





Summary:

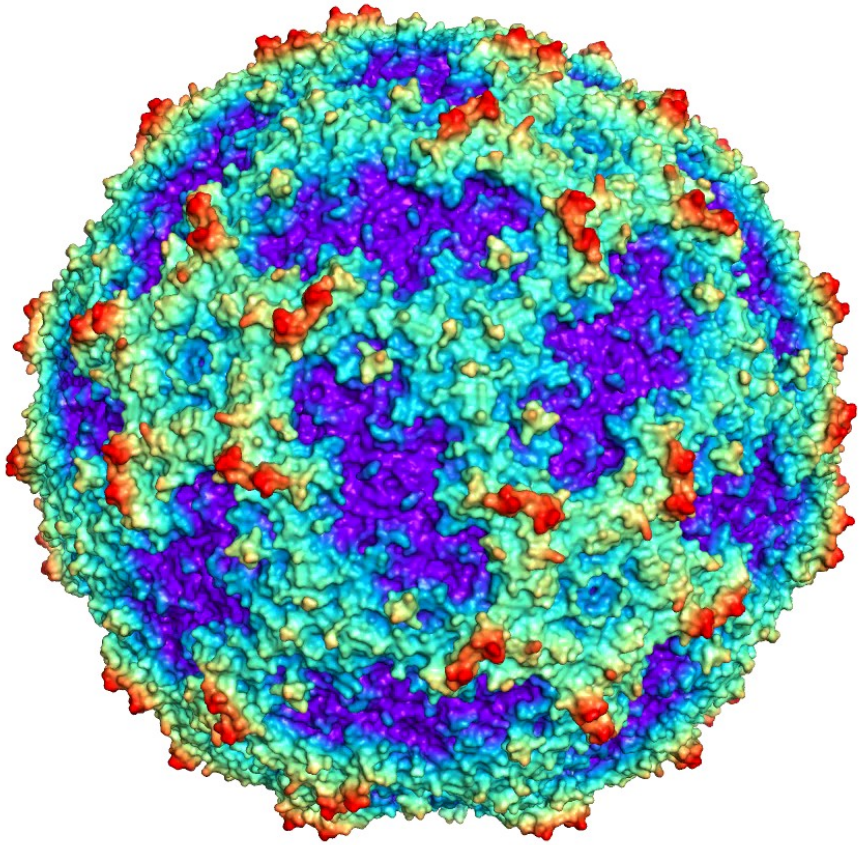
1. X-rays have suitable wavelength for study of molecular structures

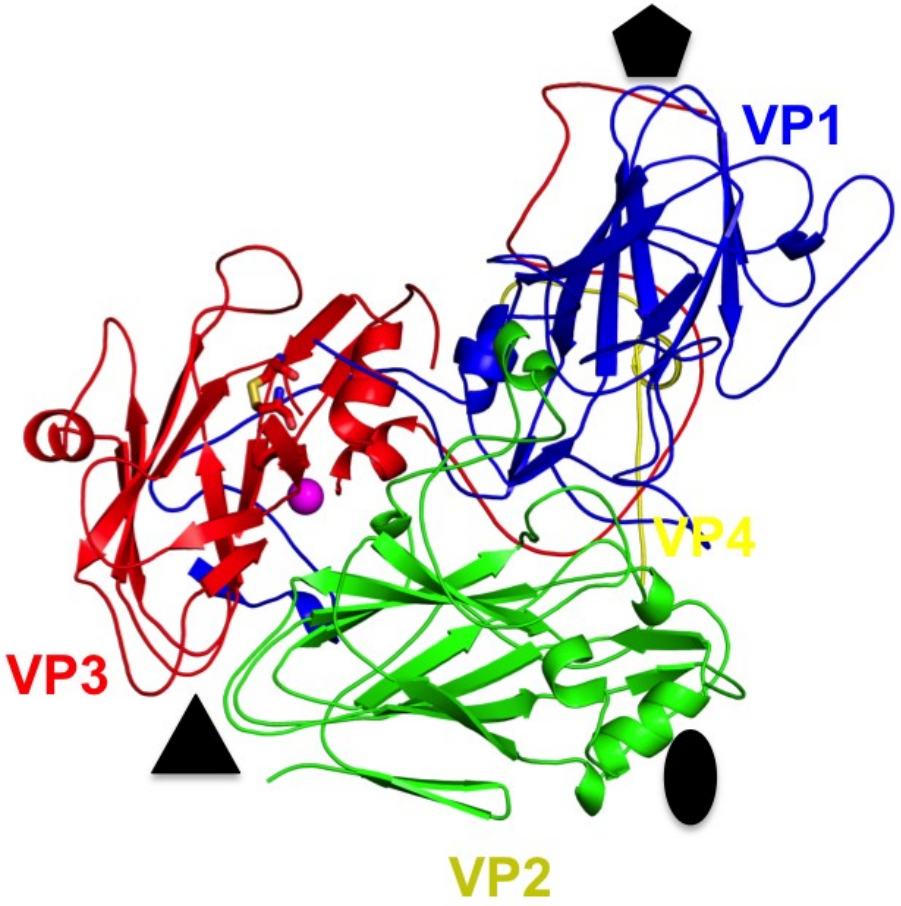
2. Crystals allow measurement of useful diffraction data because they diffract strongly in certain directions

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

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

Human cardiovirus Saffold virus 3 (2.5Å resolution)





1. Rentgenové paprsky se používají ke studiu makromolekulárních struktur protože:

A.) Mají vlnovou délku podobnou meziatomovým vzdálenostem.

B.) Jako jediné elektromagnetické záření interagují s biologickým materiálem.

C.) Byly objeveny v době intenzivního zájmu o strukturu makromolekul a z historických důvodů se používají dodnes.

2. To, že makromolekuly tvoří krystaly znamená že:

A.) Mají enzymatickou aktivitu

B.) Jsou součástí kostry buňky (cytoskeletu)

C.) Mají stabilní strukturu.

3. Mapa elektronové hustoty, která je výsledkem rentgenové analýzy krystalů:

A.) Ukazuje tvar molekul, které tvoří krystal

B.) Má vždy bílou barvu

C.) Ukazuje tvar molekuly po denaturaci

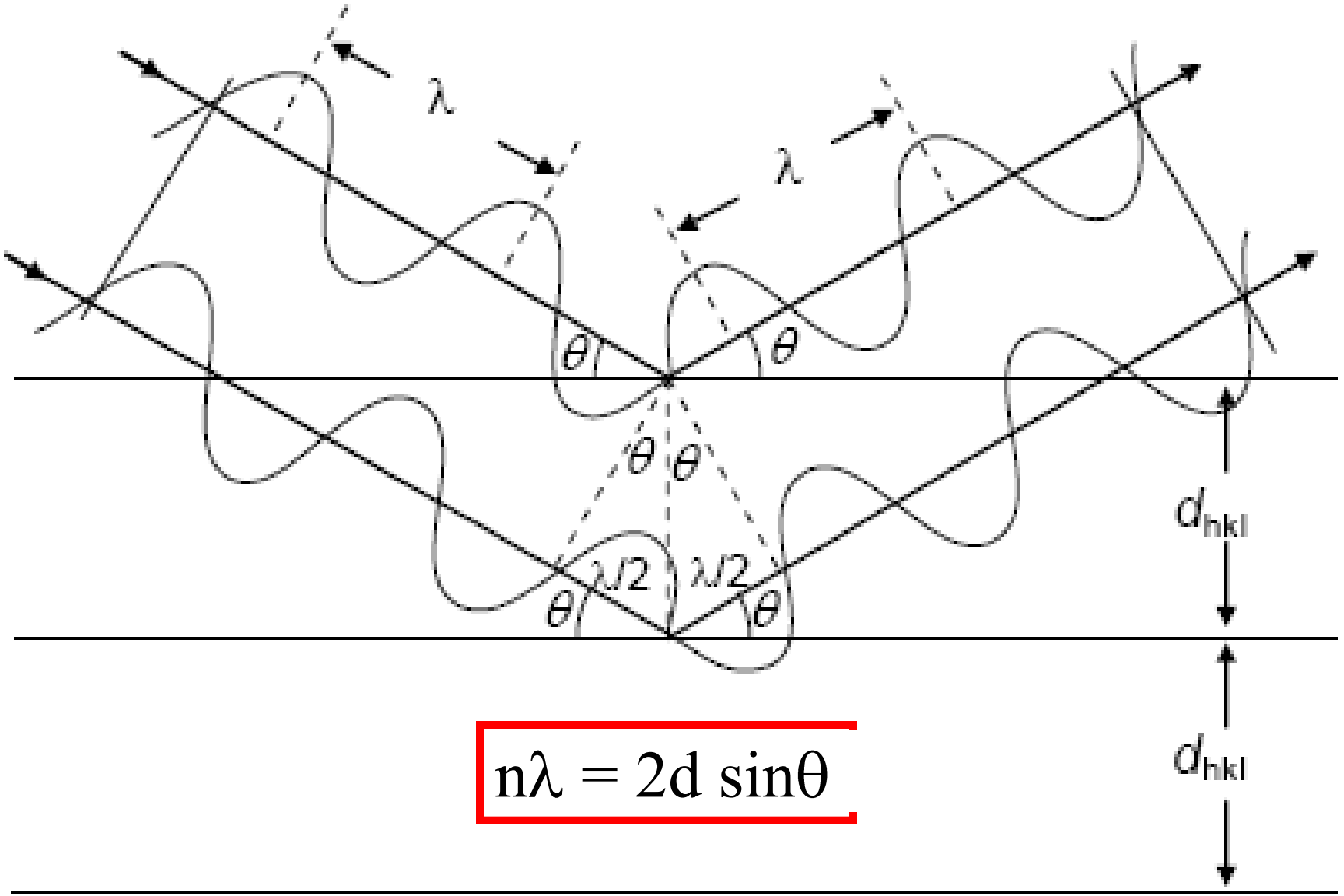
Macromolecular crystallography

Lecture 2

Pavel Plevka

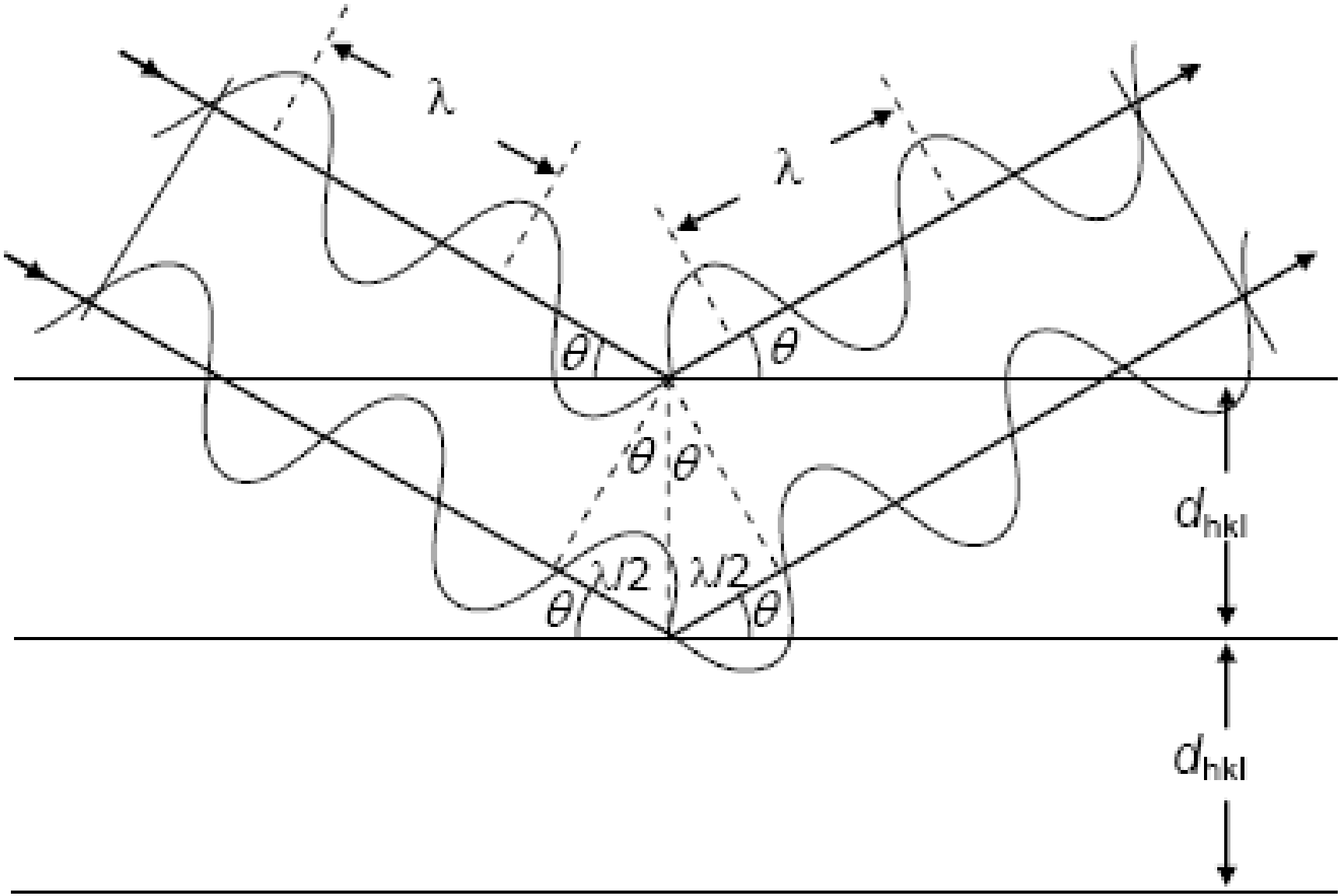
- Phase problem and its solution
- Building macromolecular structures based on X-ray diffraction data
- Validation of macromolecular structures

Bragg's law

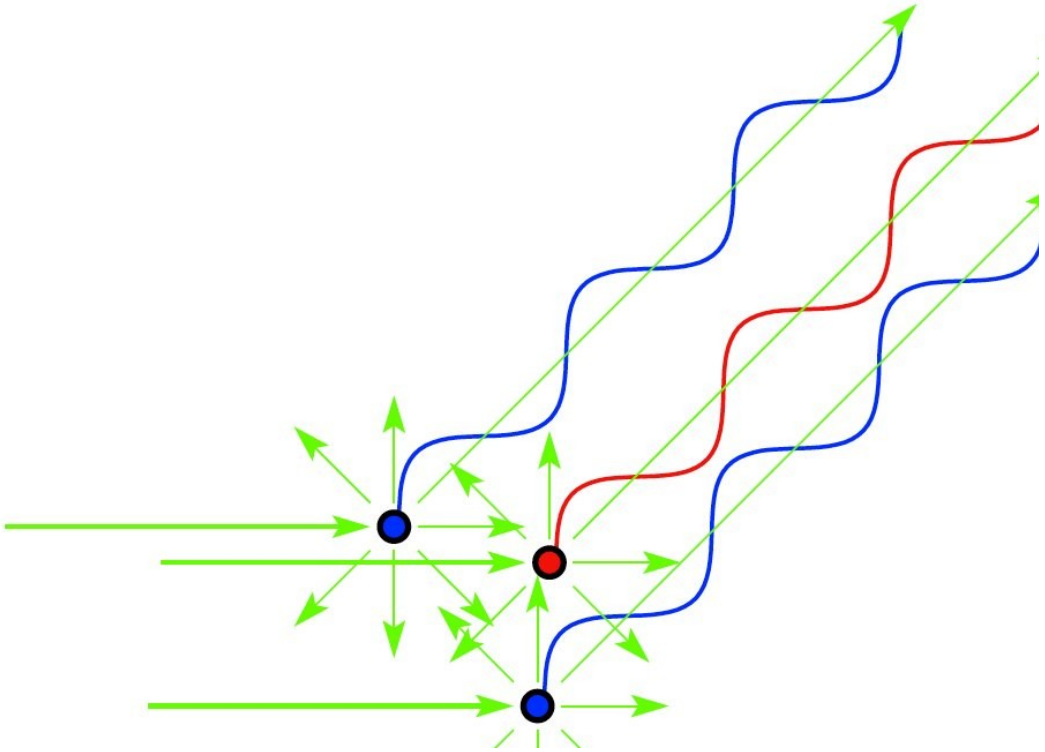


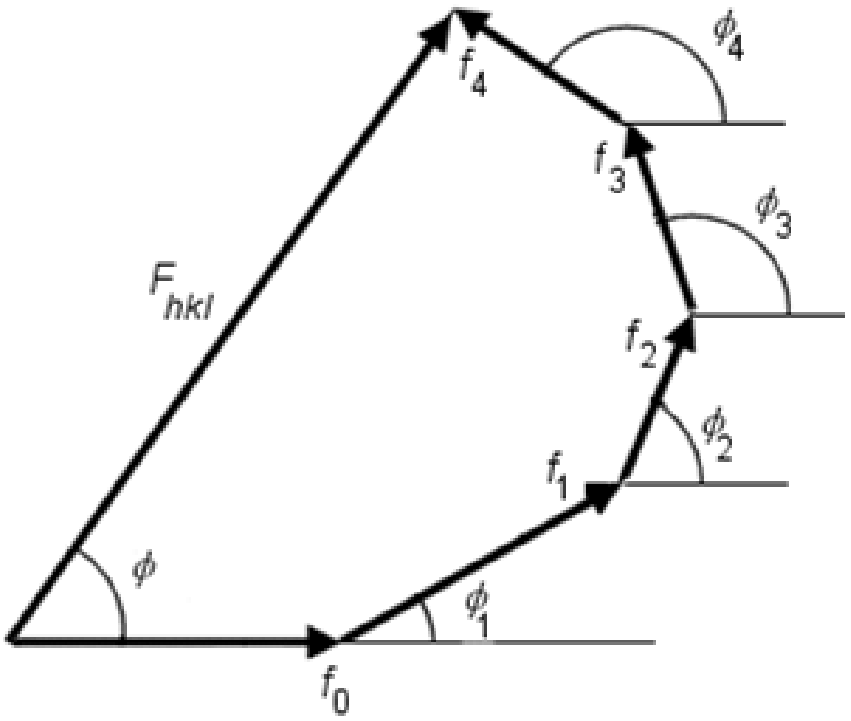
$$n\lambda = 2d \sin\theta$$

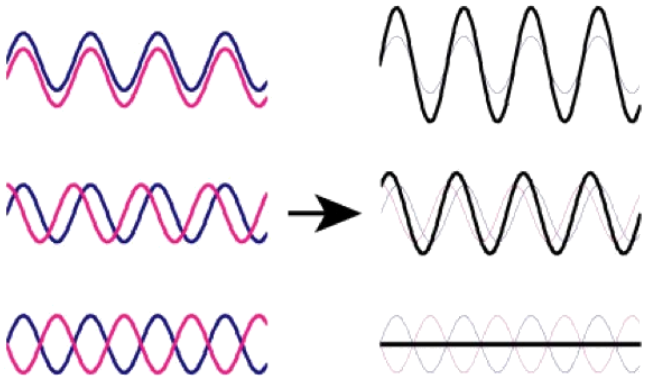
Bragg's law



Adition of waves





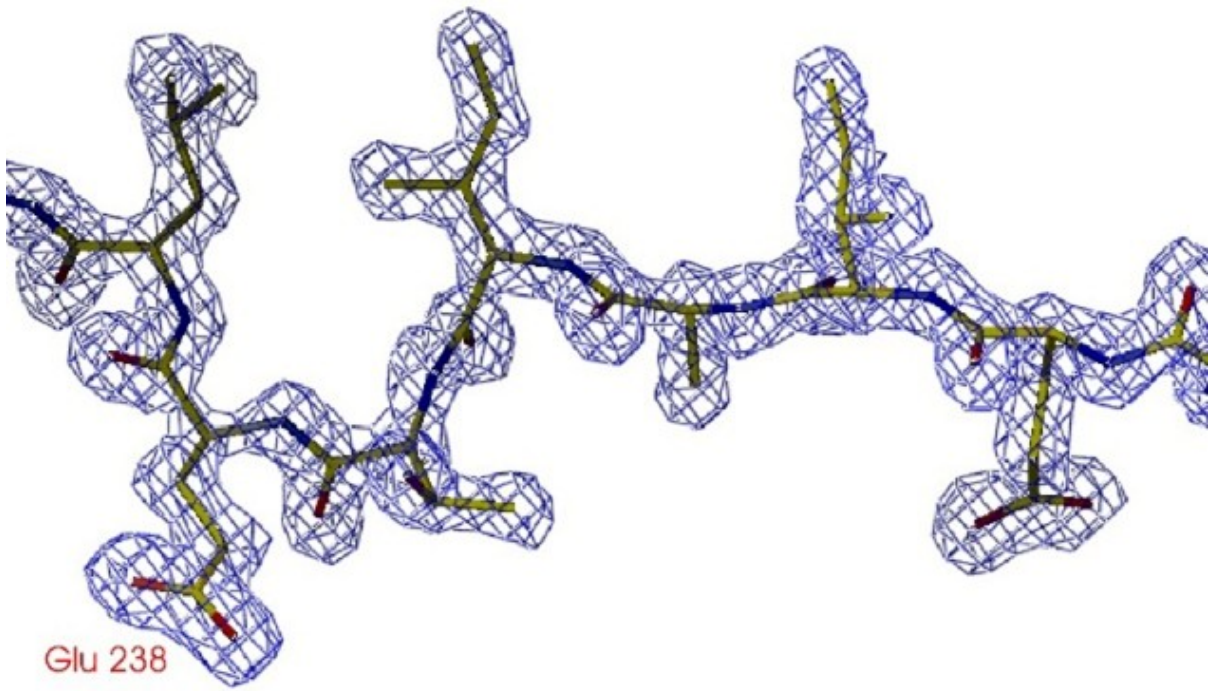


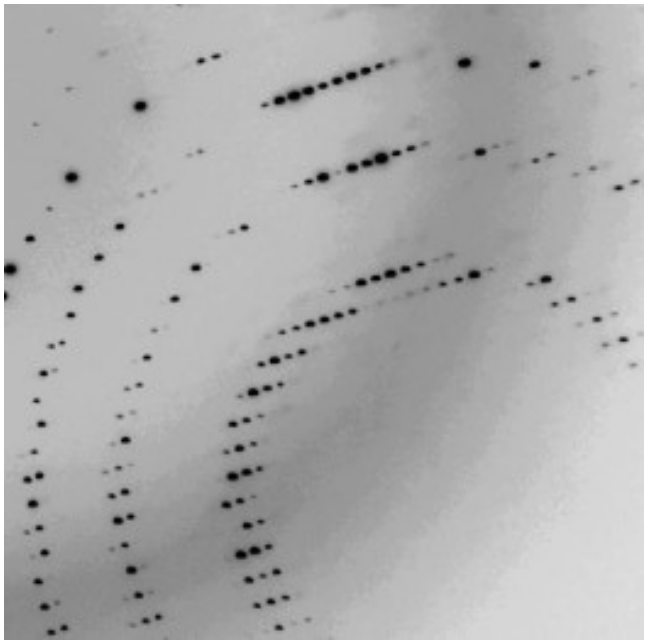
$F=A\cos\alpha+iA\sin\alpha$

Electron density equation & PHASE PROBLEM

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

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

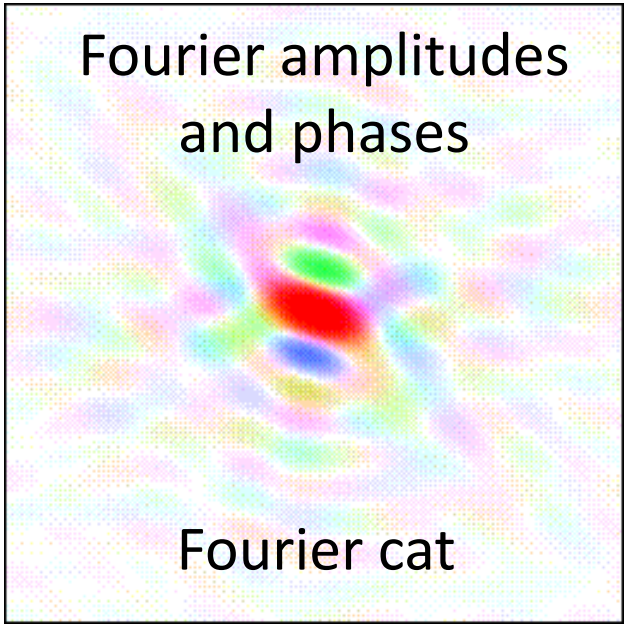




Real space cat

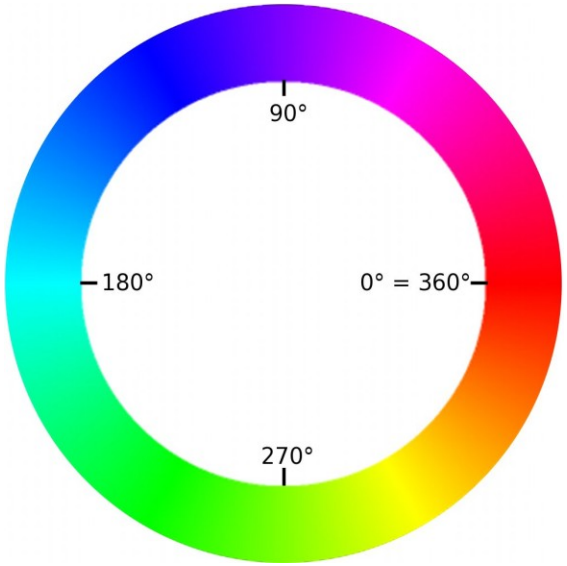
Fourier transform

Fourier amplitudes and phases



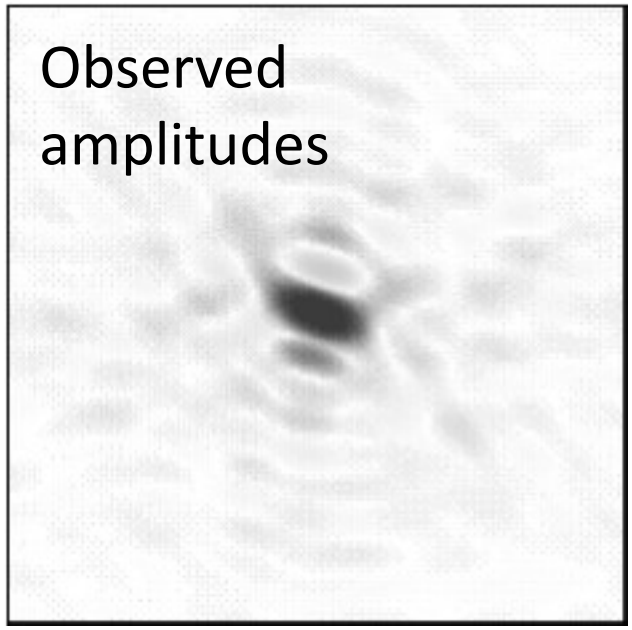
Fourier cat

Circular rainbow scale of phases



Linear intensity scale of amplitude size

Observed amplitudes



Solving the phase problem by:

Molecular replacement

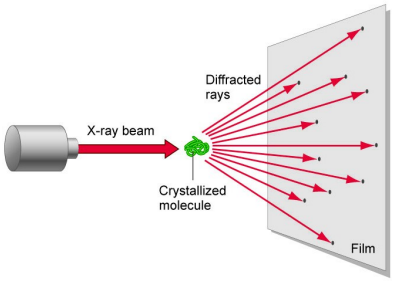
1. source of initial phases is a model
2. the model is oriented and positioned to obtain the best agreement with the x-ray data
3. phases are calculated from the model
4. The calculated phases are combined with the experimental data

Molecular **R**eplacement was invented by **M**ichael **R**ossmann

Unknown structure, unknown orientation

Cat

Diffraction experiment



Observed amplitudes Phases unknown!

Fourier cat

Known structure

Manx cat

Fourier transform

Calculated amplitudes and phases

FT of Manx cat

Wrong orientation!

Known structure

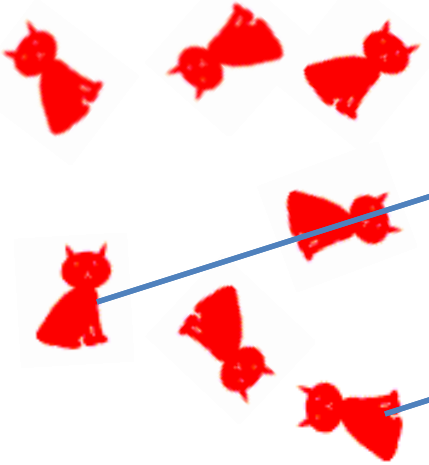
Manx cat

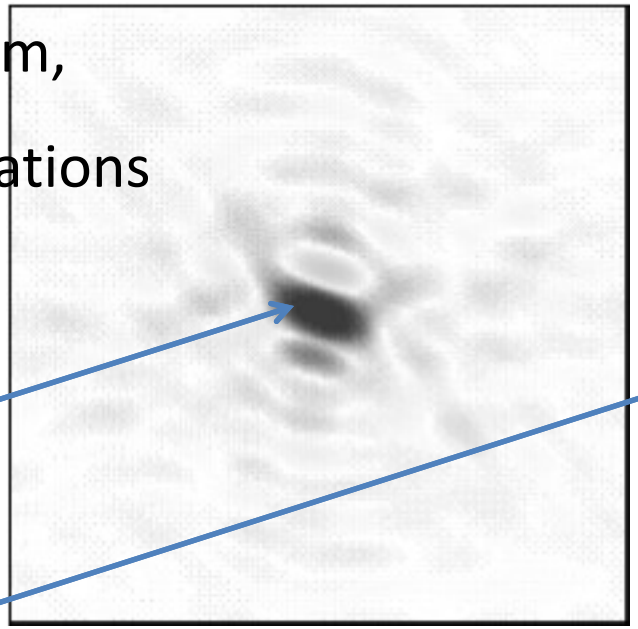
Observed amplitudes Phases unknown!

Fourier cat

Fourier transform,

try different orientations



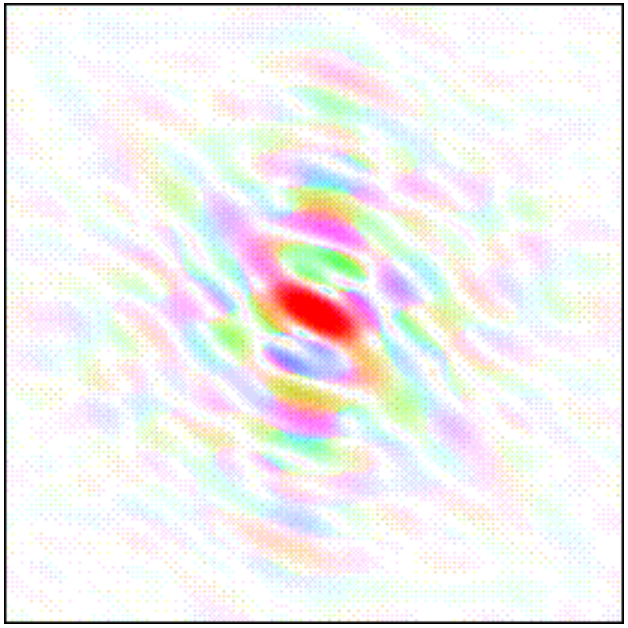


Calculated amplitudes and phases

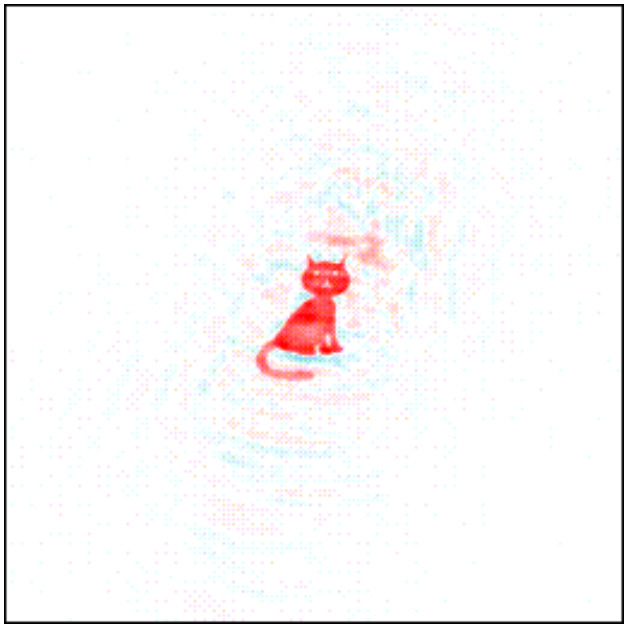
FT of Manx cat

Wrong orientation!

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



Inverted Fourier transform



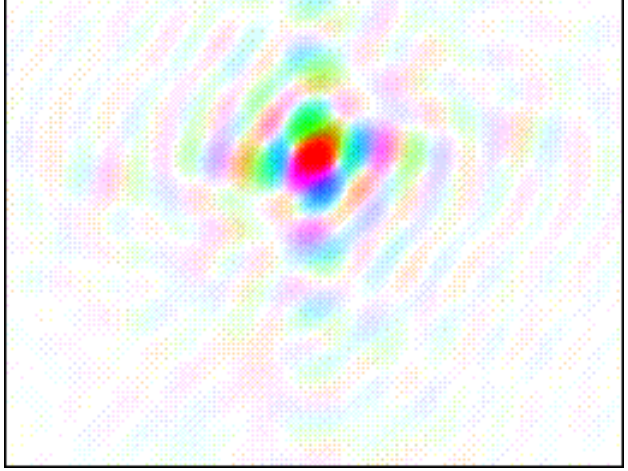
Even the tail becomes visible!

Model Bias

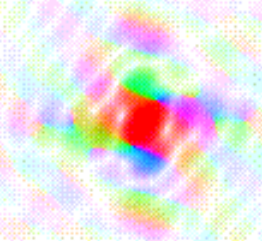
Duck

Fourier transform

Fourier transform of duck

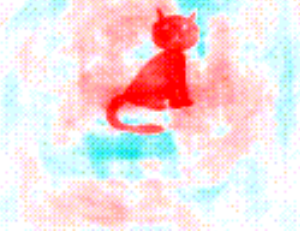


Duck amplitudes + cat phases



Inverted Fourier transform

Looks like a cat!!



Solving the phase problem by:

Multiple/Single Isomorphous Replacement (MIR/SIR)

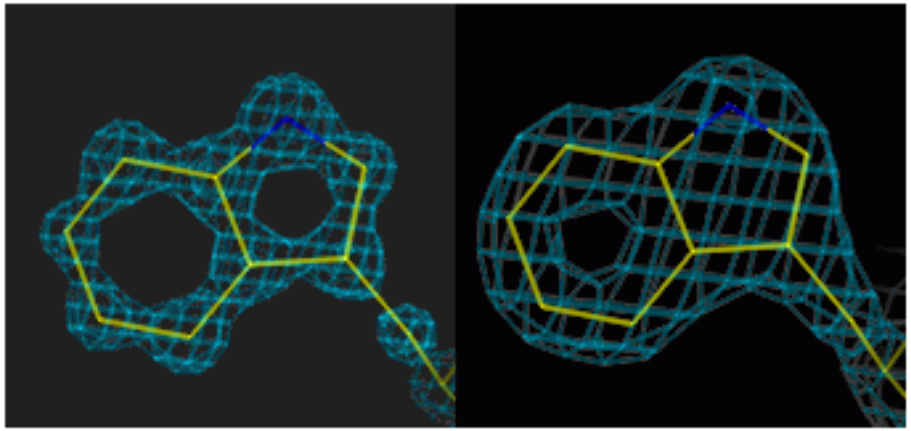
- source of phases – intensity differences between data from native and derivative (heavy atom containing) crystals
- Positions of heavy atoms identified from isomorphous difference Patterson maps

Solving the phase problem 3

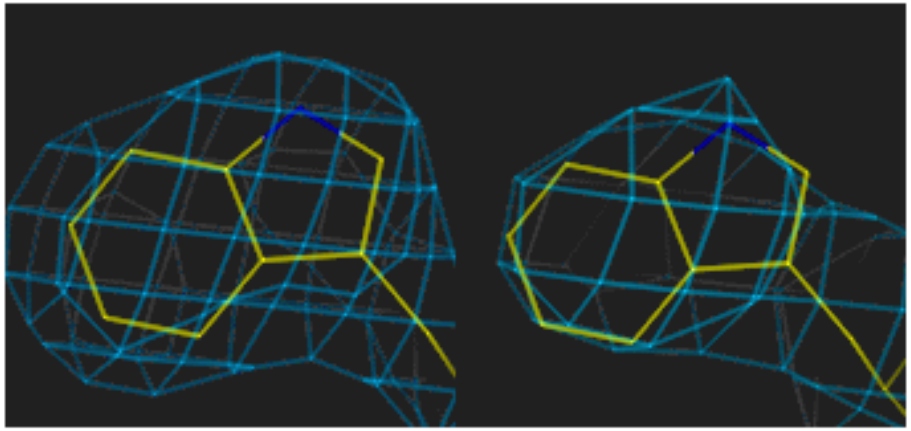
Multiple/Single-wavelength anomalous diffraction (MAD/SAD)

- source of phases – intensity differences between structure factors due to the presence of atom that specifically interacts with X-rays of a given wavelength
- Positions of heavy atoms identified from anomalous difference Patterson maps

Model building & resolution

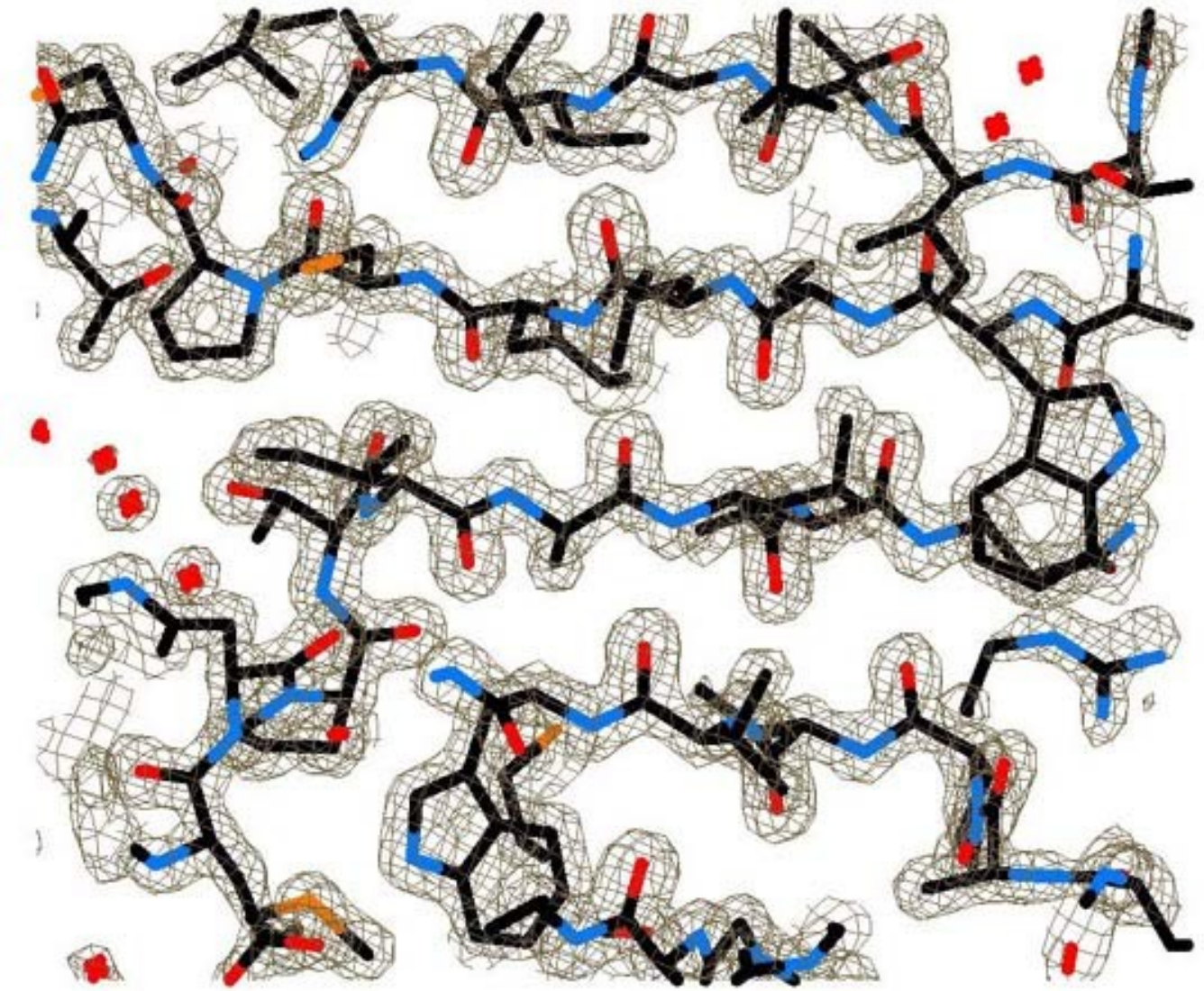


1.0Å 2.5Å

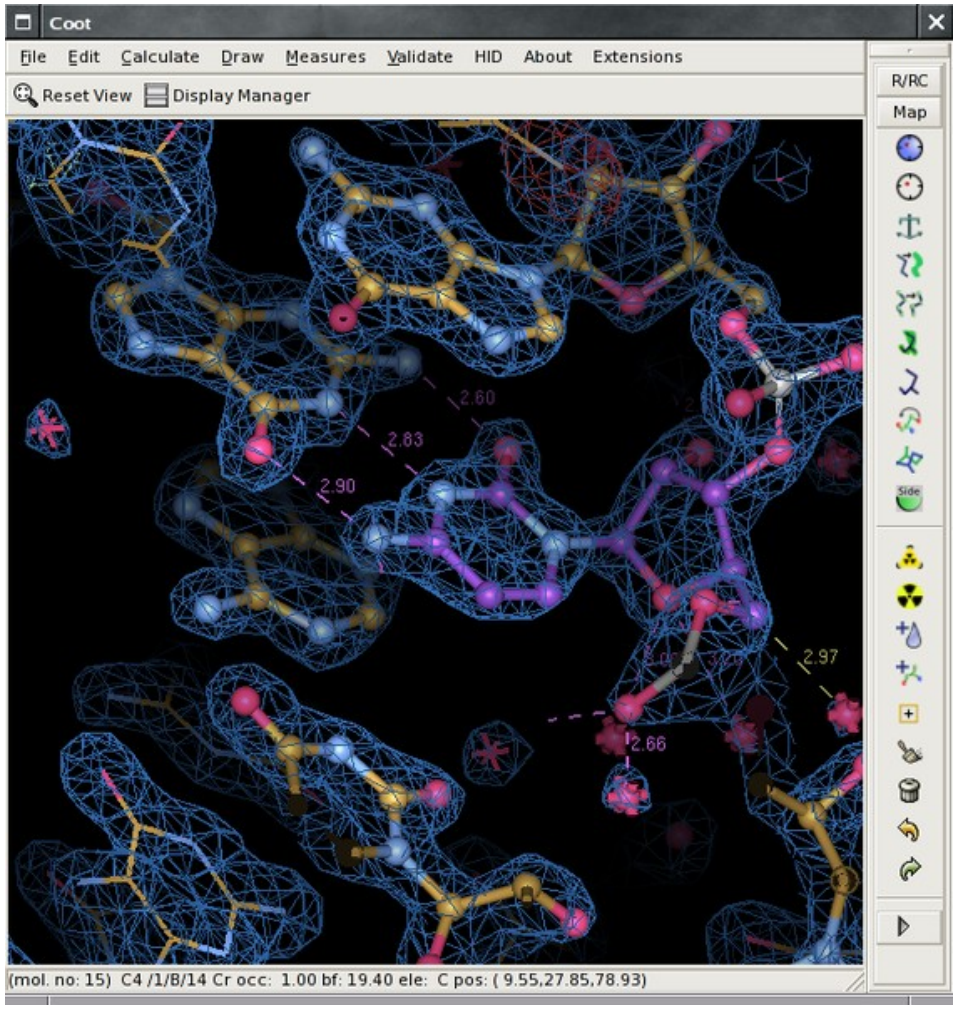


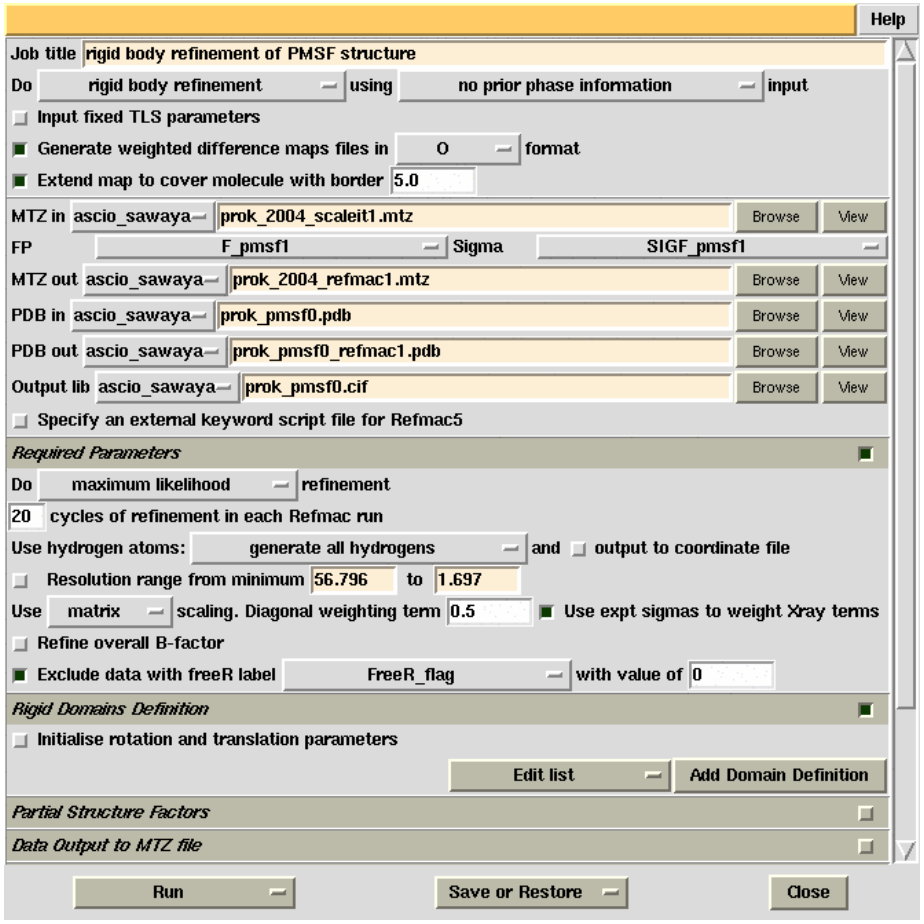
3.0Å 4.0Å

Model building & refinement



Model building & refinement



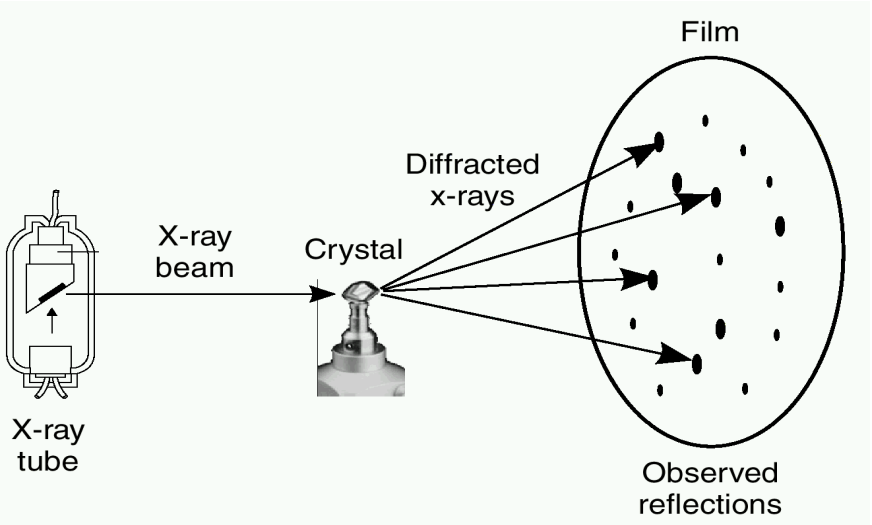


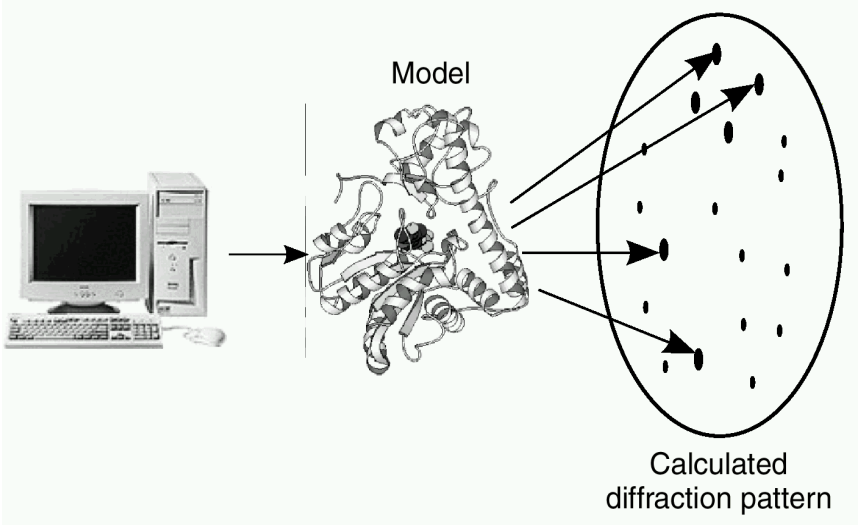
Validation

Assesment of model quality:

- Is the model in agreement with experimwntal data?
- How the geometry of amino acids look like?
- Are atoms far / close enough from each other?
- Are residues “happy” in their environment?
- Are the hydrogen donors/acceptors satisfied?

R-factor, R_{free} factor





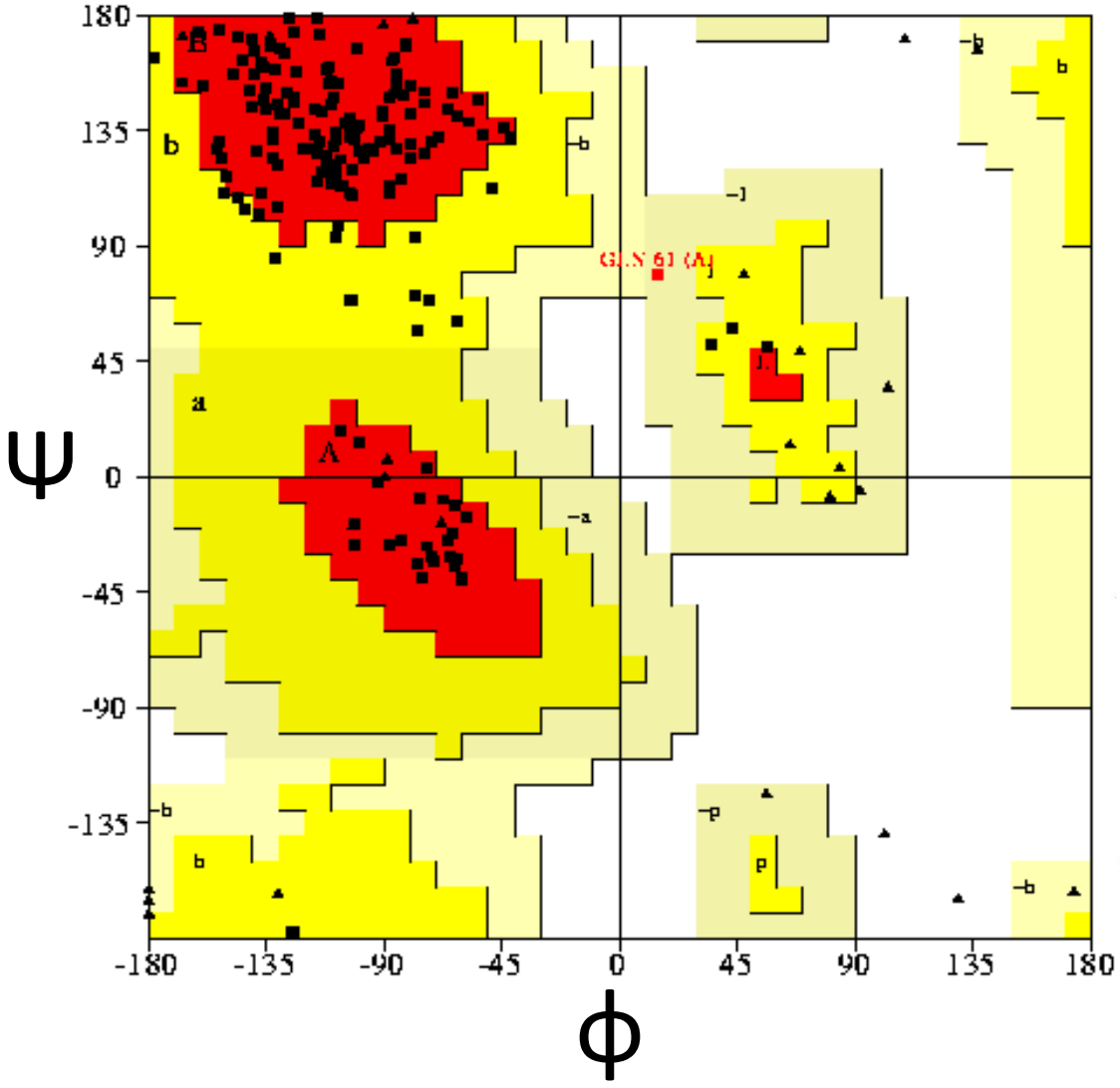
R-factor

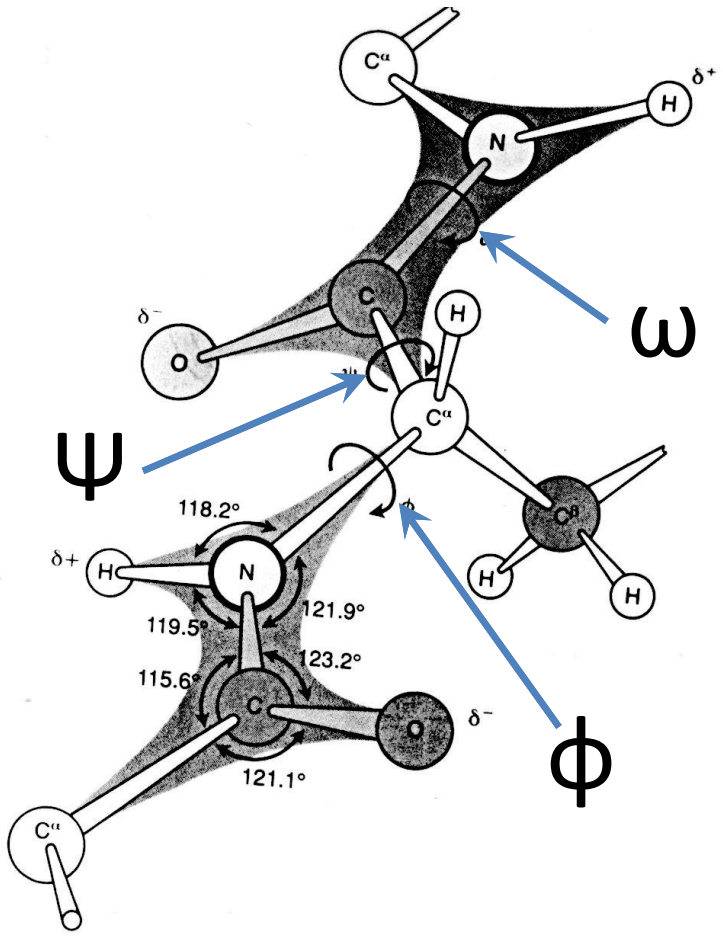
$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

R_{free} factor

$$R_{free} = \frac{\sum_{hkl \subset T} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl \subset T} |F_{obs}|}$$

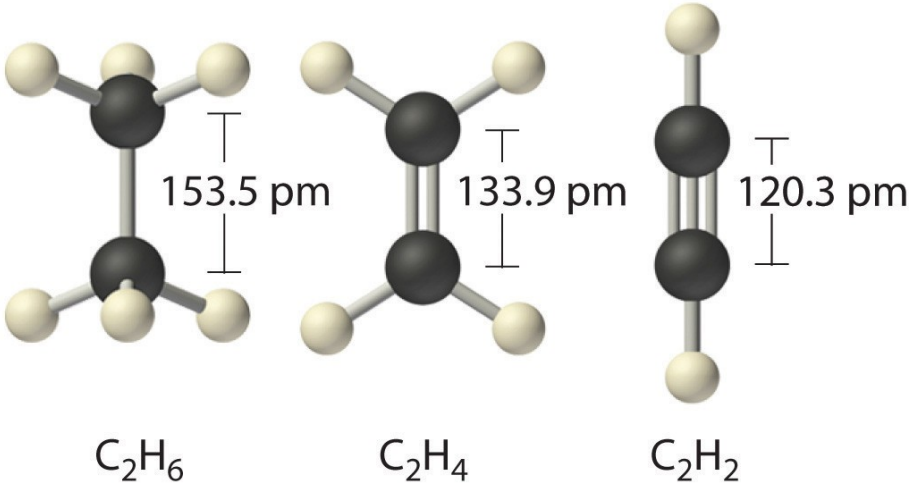
Ramachandran plot



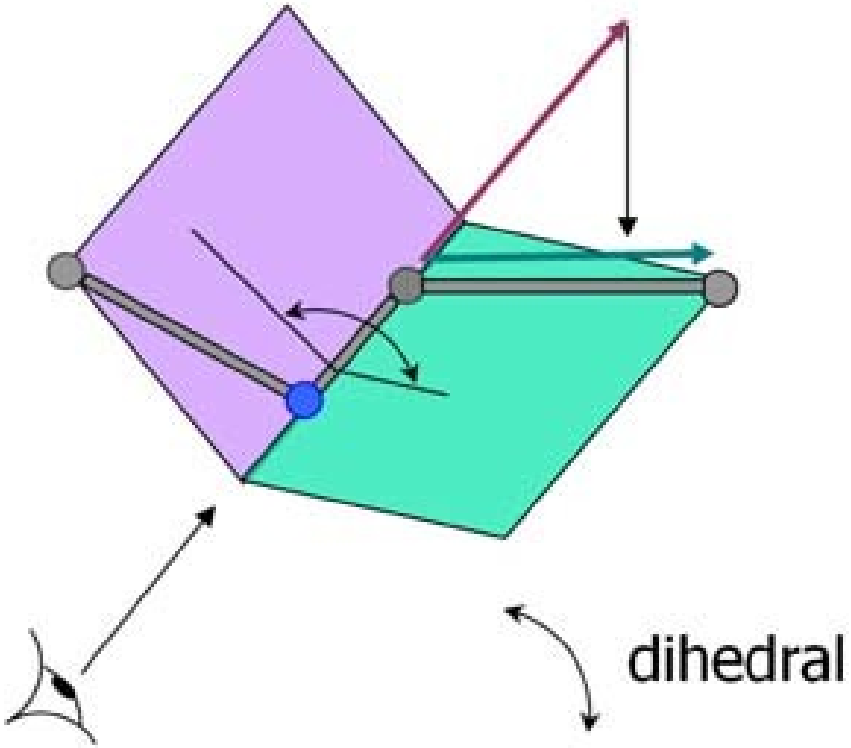


Geometry and stereochemistry

Bond lengths

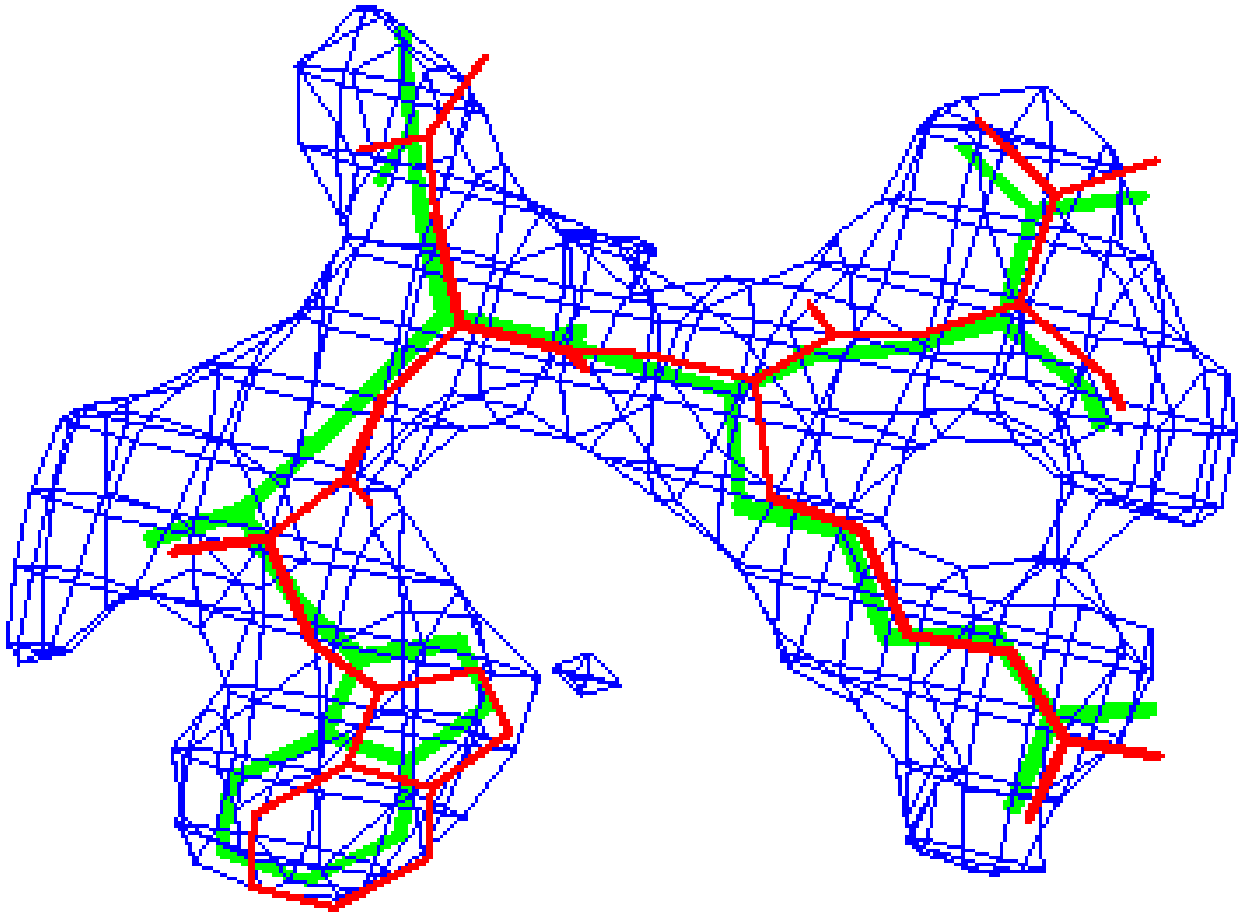


Dihedral angles



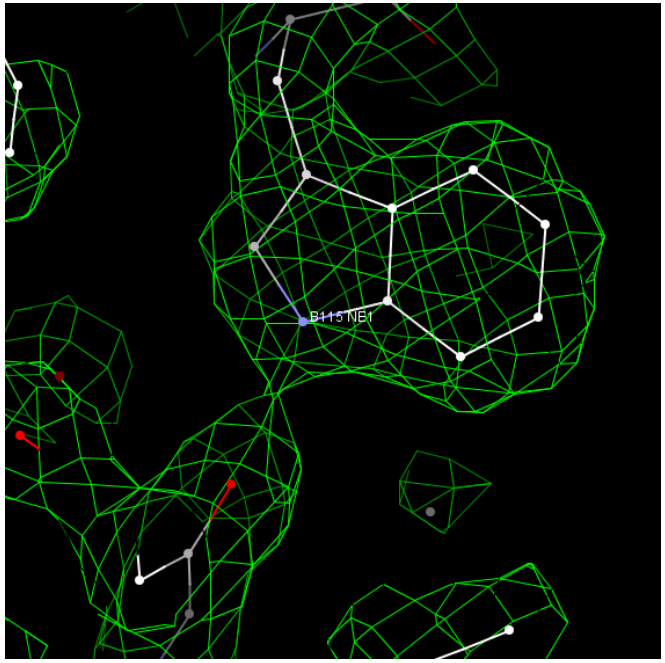
$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}}.$$

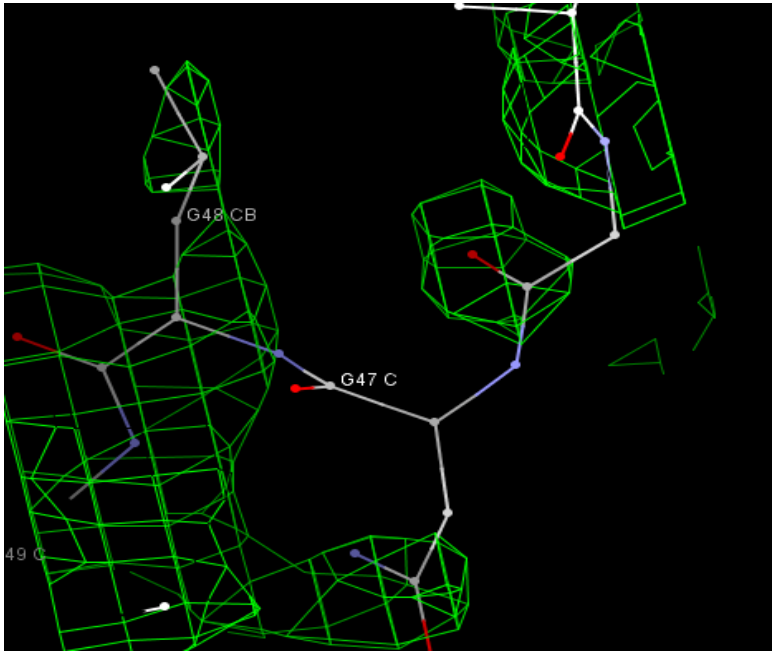
Real-space fit



Data deposition

- Protein Data Bank (PDB)
- Some structures are wrong!





Summary

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

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

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

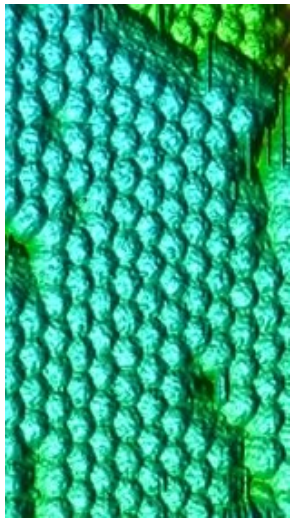
4. Diffraction experiments provide only amplitudes of structure factors => Phase problem

5. Solution of the phase problem:
   - Molecular replacement
   - Isomorphous replacement
   - Anomalous diffraction

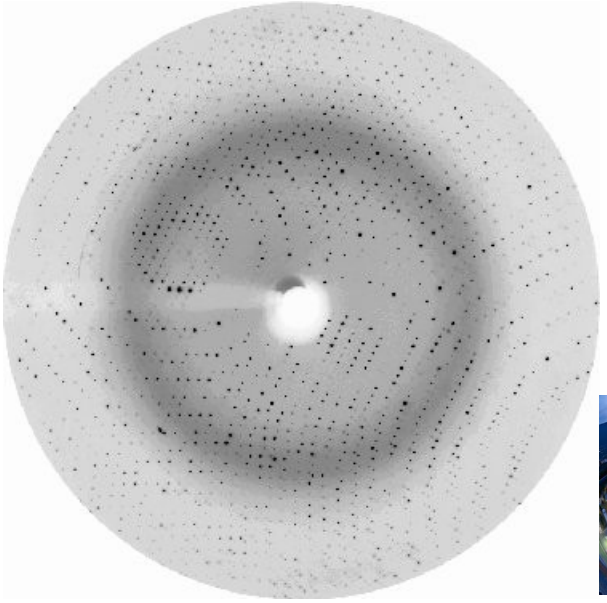
6. Model building, refinement, validation, deposition

1. Virus purification

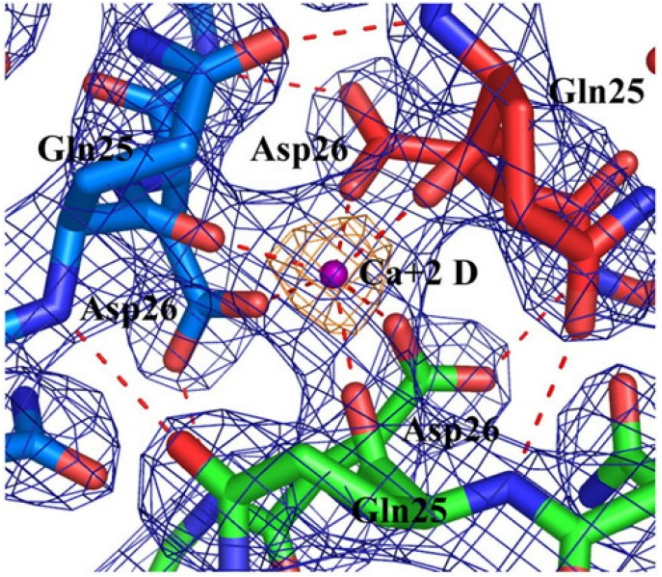
2. Crystallization



3. Diffraction data



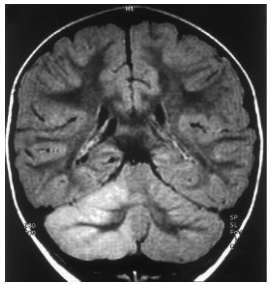
4. Solve structure

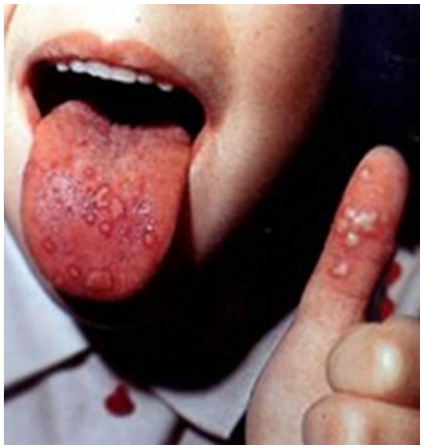


Structural studies of human picornaviruses

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

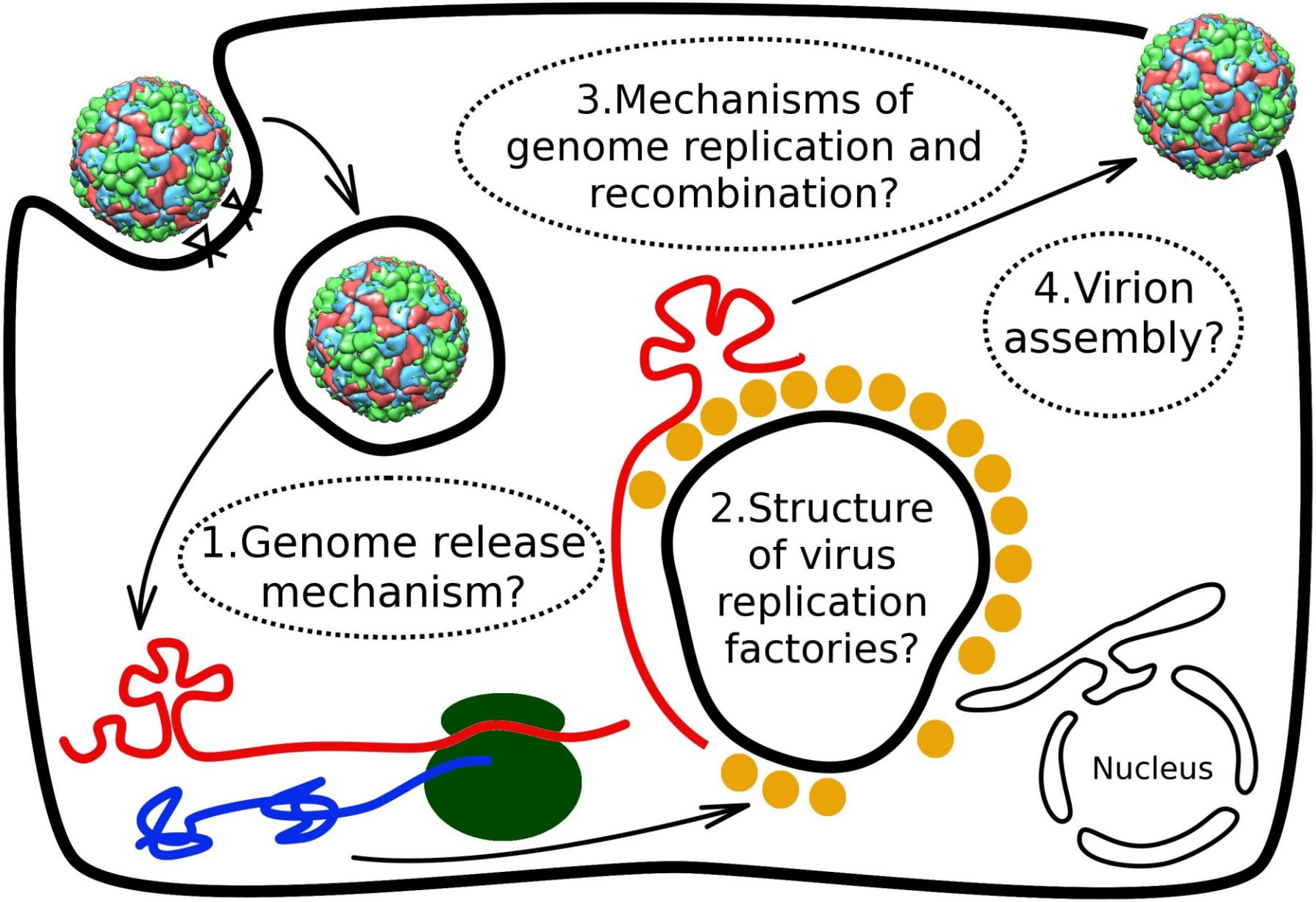
Enteroviruses (EV71)
- hand-foot-and-mouth-disease
- encephalitis



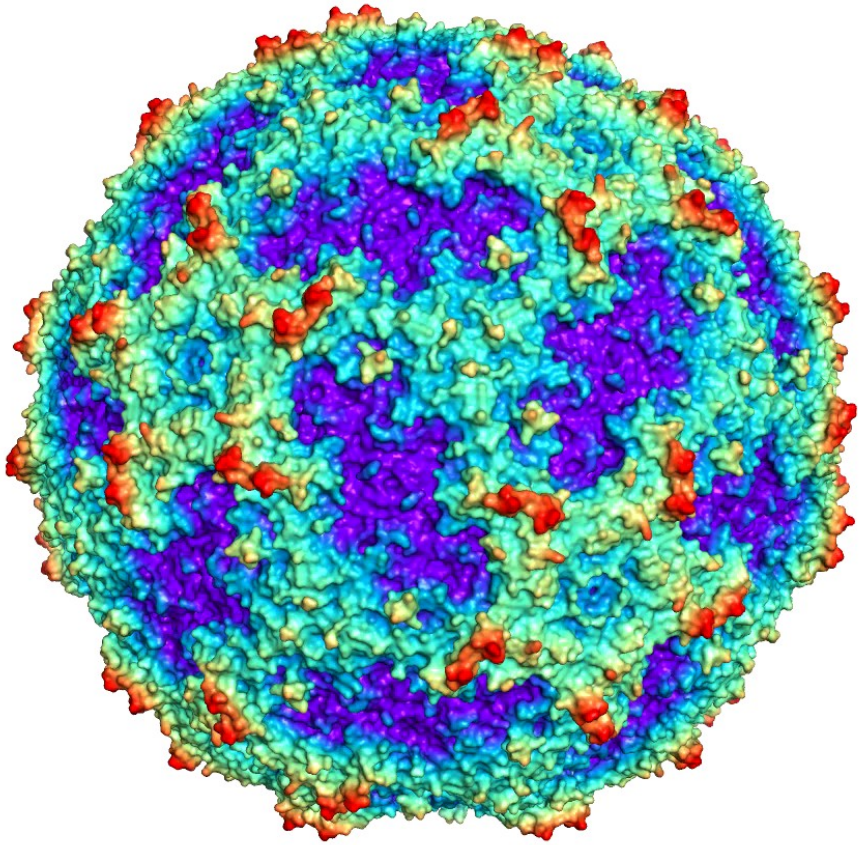


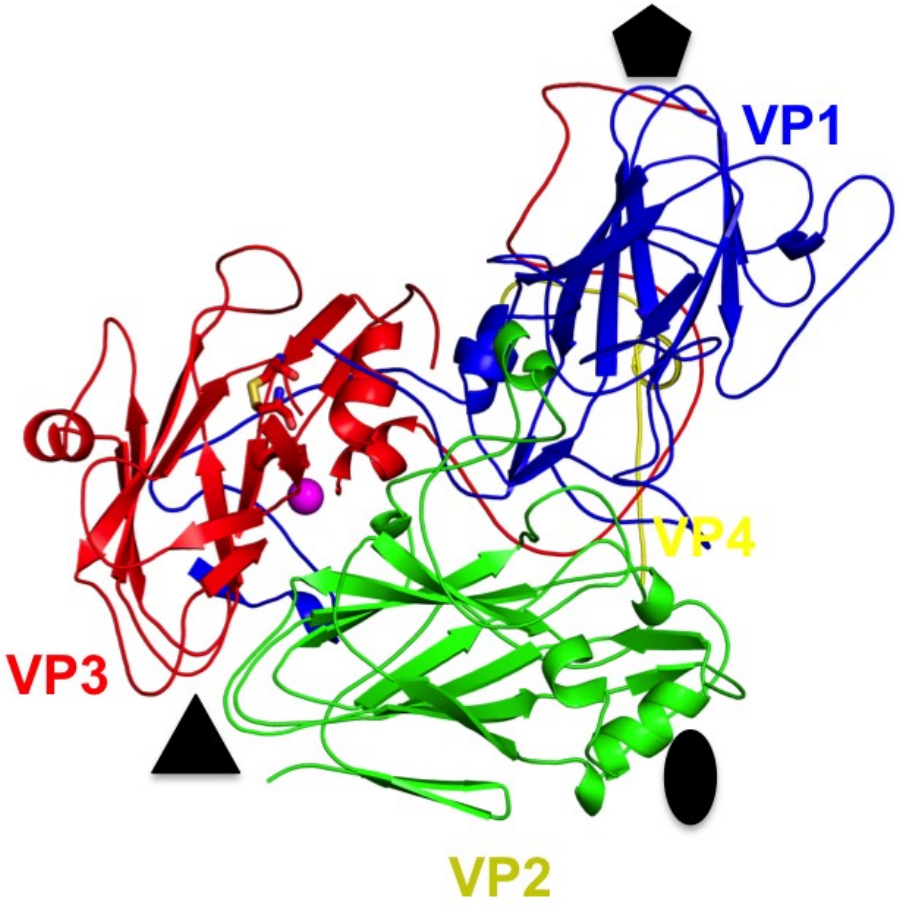
McMinn et al. Clin Infect Dis 2001. Image by Heng Soy, KI Media 2012.

Picornavirus replication cycle

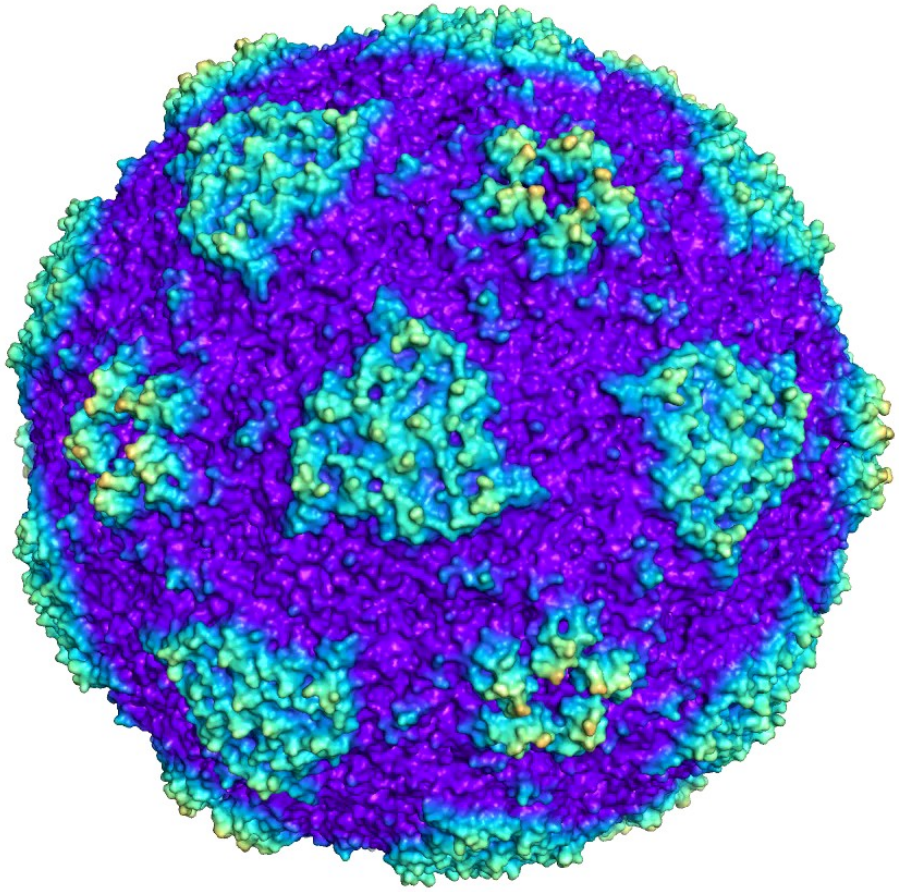


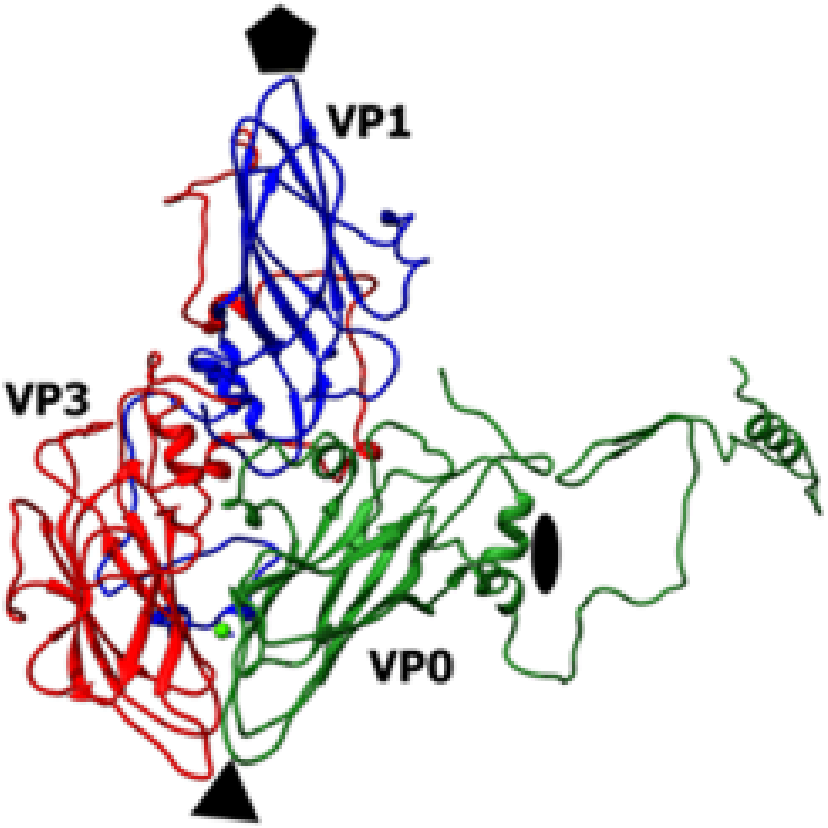
Human cardiovirus Saffold virus 3 (2.5Å resolution)





Human Parechovirus 1 @ 3.1Å





1.) Jakou část strukturního faktoru můžeme změřit v difrakčním experimentu:

a) amplitudu (ve formě intensity)

b) fázi

2.) Nejčastější metoda pro získání fází je:

a) molekulární nahrazení (molecular replacement)

b) isomorfní nahrazení

c) anomální diffrakce

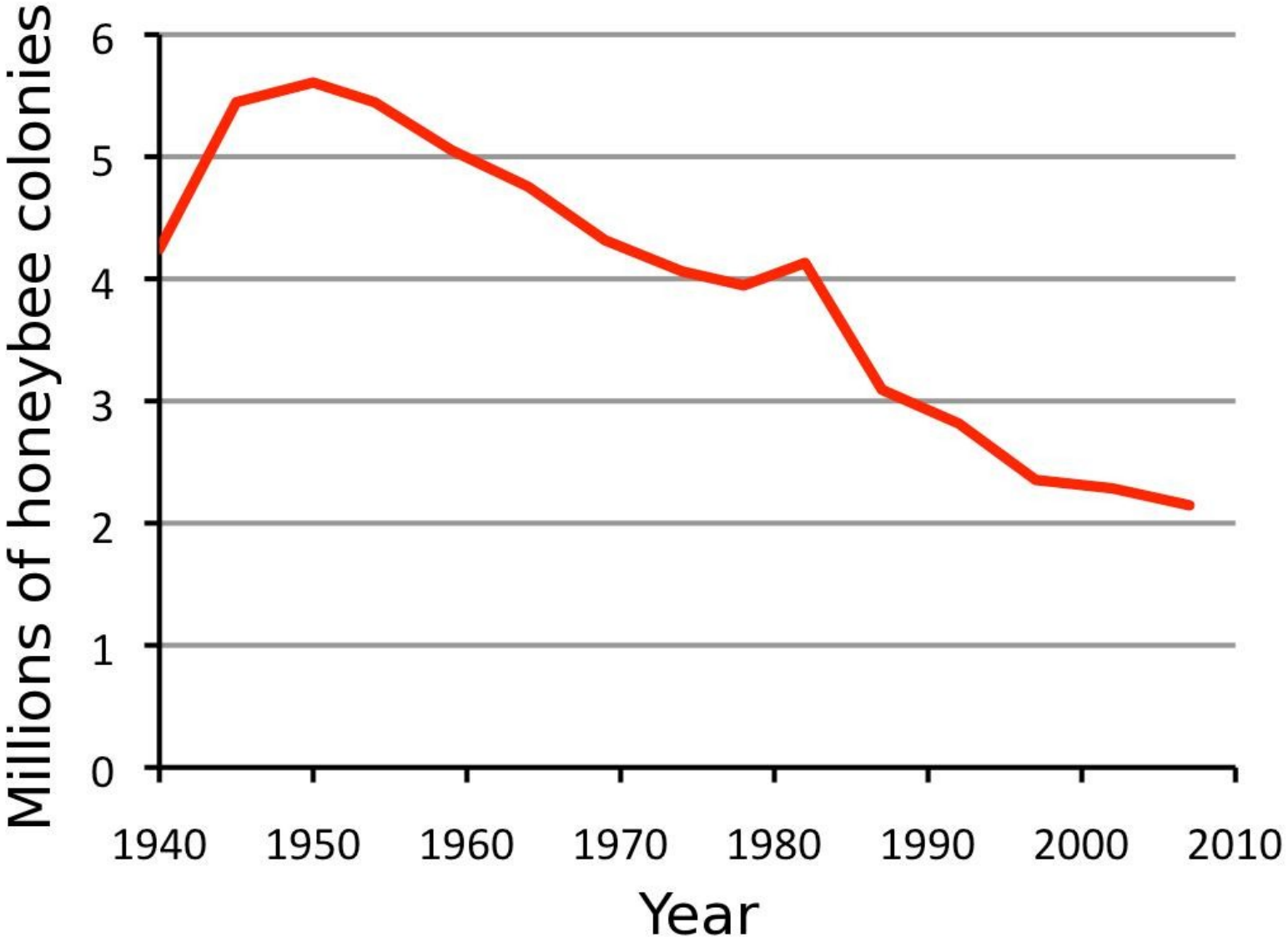
3.) Ramachandran plot ukazuje:

a.) distribuci úhlů v hlavním řetězci proteinu

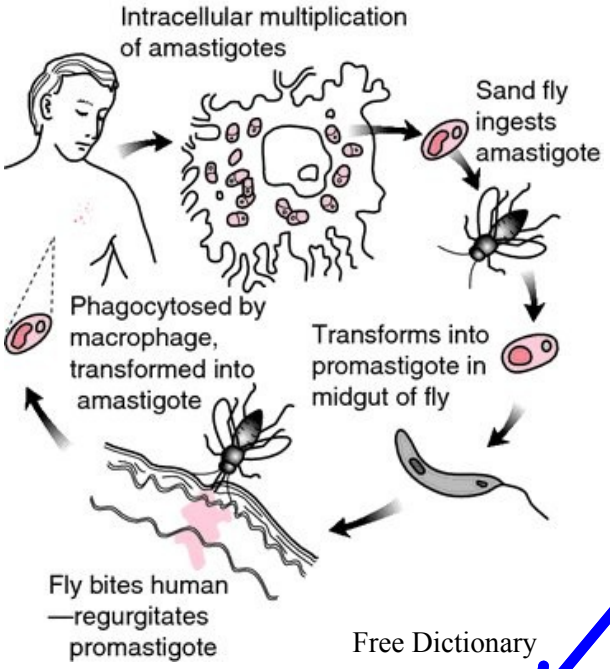
b.) vzdálenosti mezi atomy

c.) konformace postranních řetězců aminokyselin

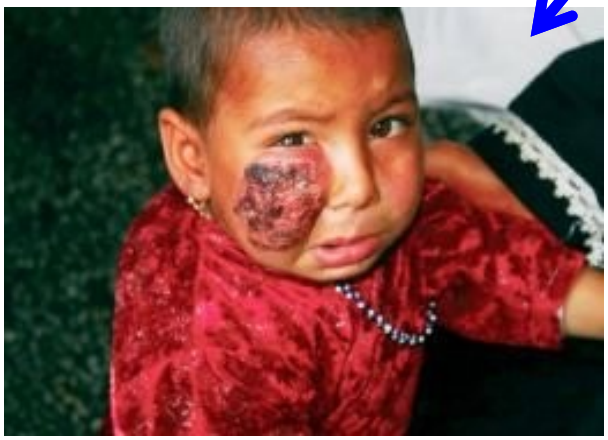
Honeybee viruses



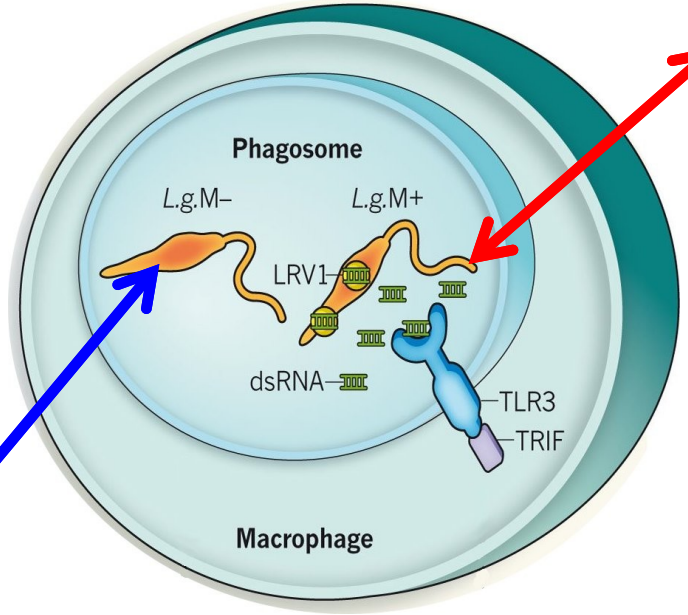
Leishmania RNA virus 1



Free Dictionary

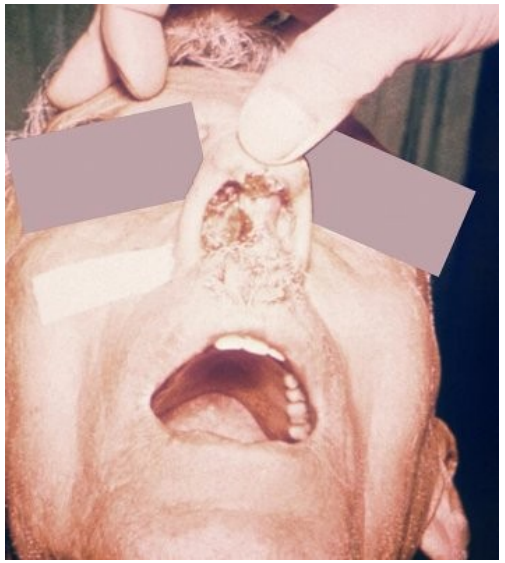


cutaneous leishmaniasis



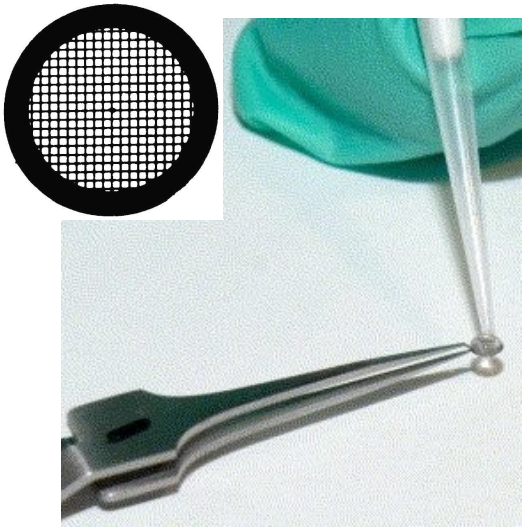
Ives et al. Science, 2011
Olivier. Nature, 2011

metastasis to nasopharyngeal tissues

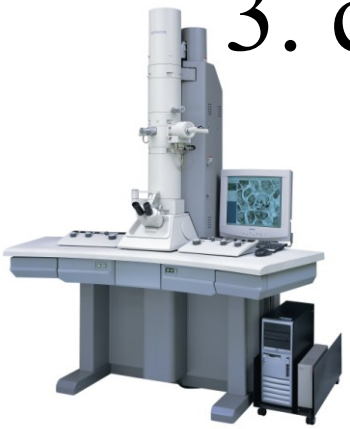


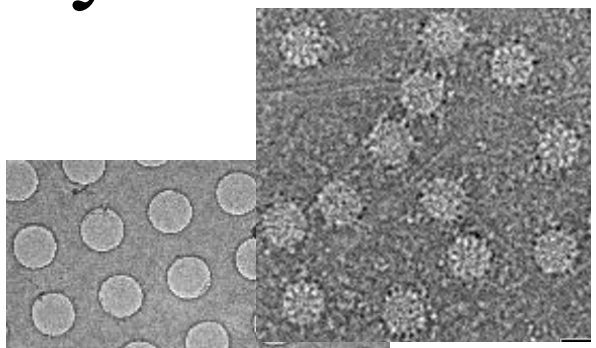
1. Virus purification

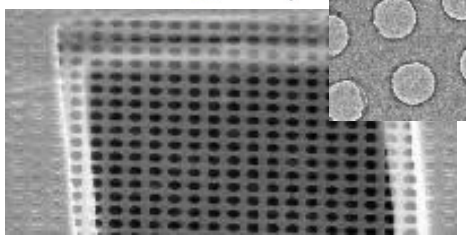
2. Grid preparation



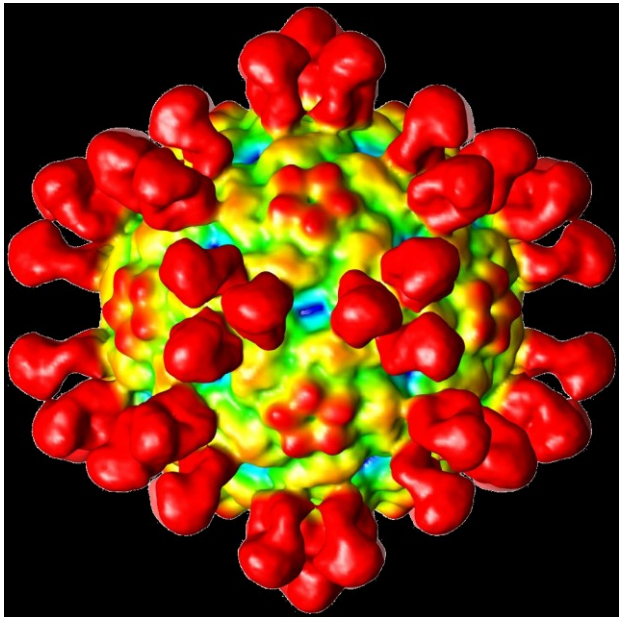
3. cryo-EM







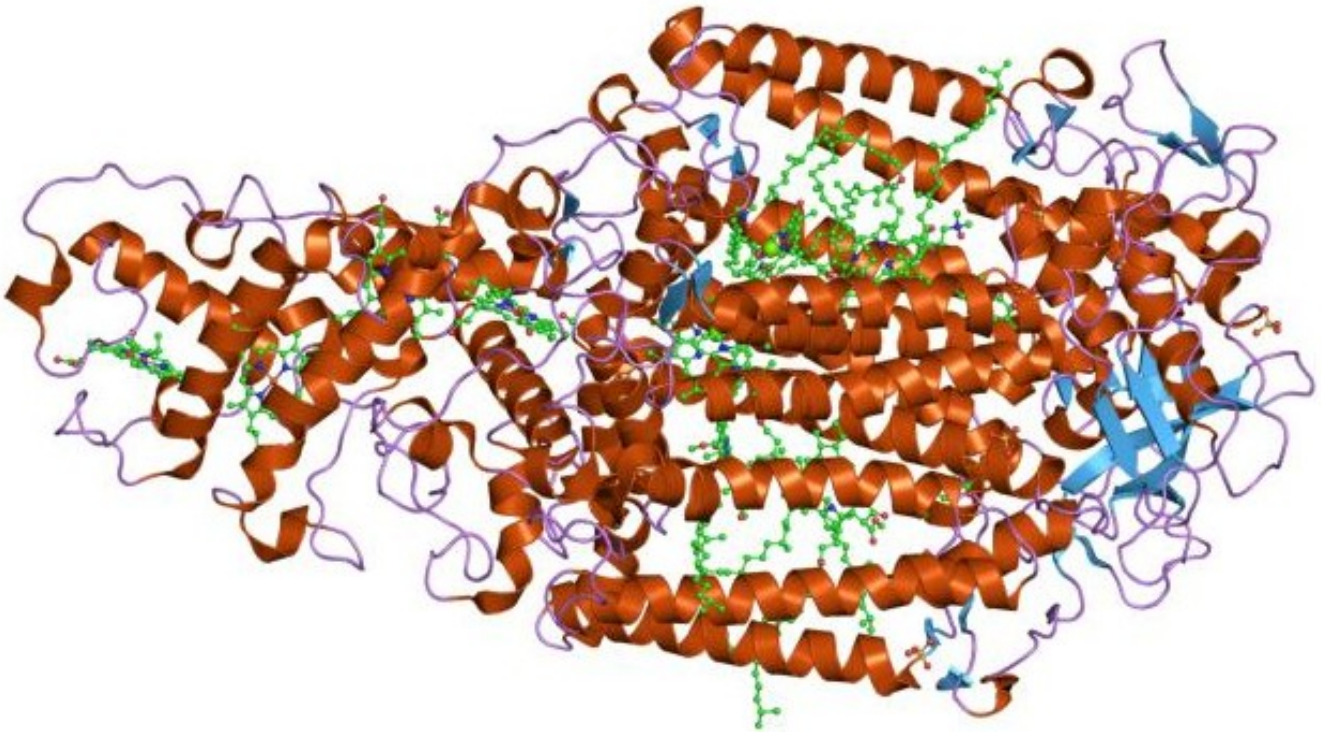
4. Reconstruction



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

- **1988 Nobel Laureates in Chemistry**

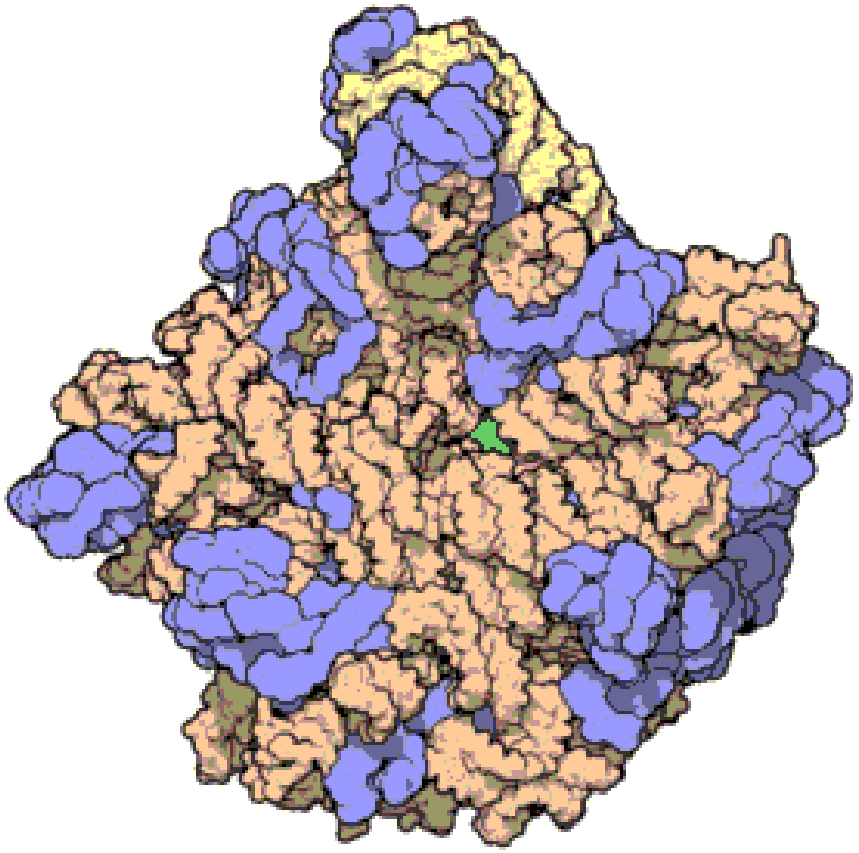
for the determination of the structure of a photosynthetic reaction centre



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

- **2009 Nobel Laureates in Chemistry**

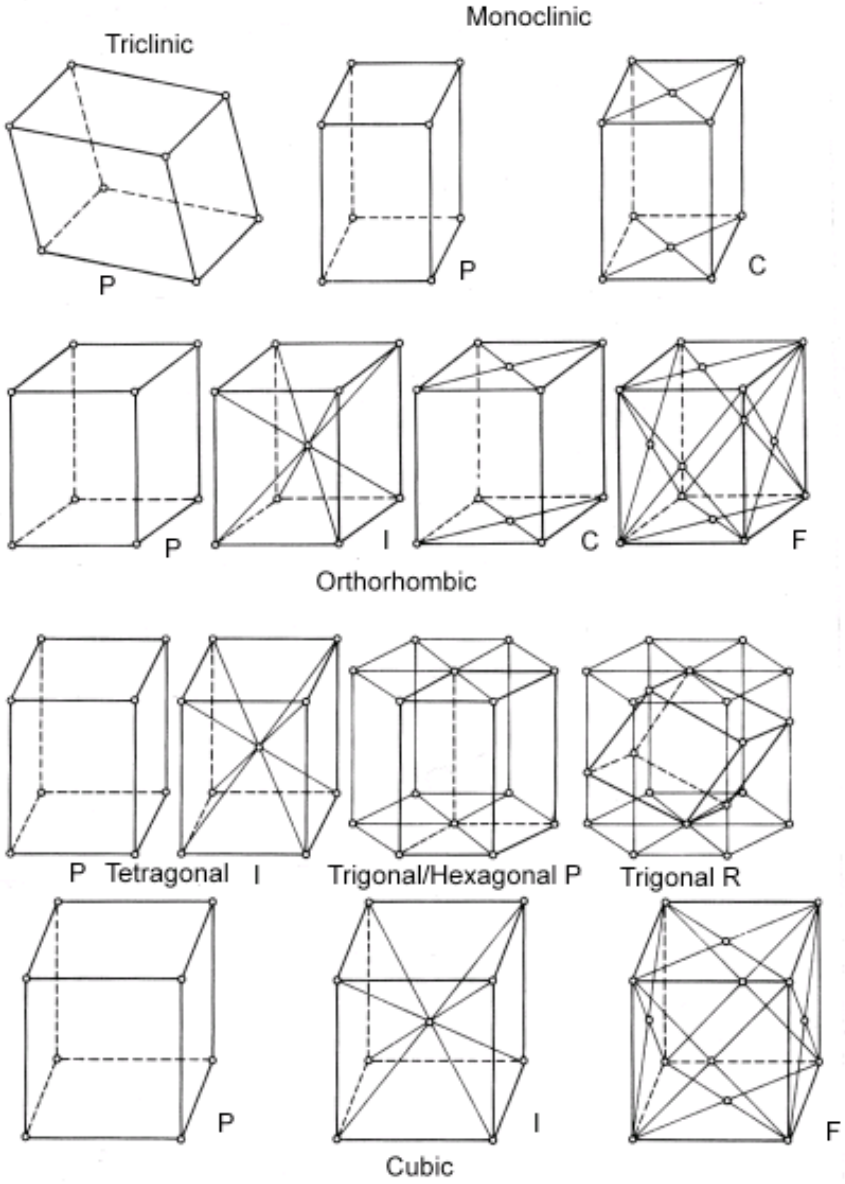
for studies of the structure and function of the ribosome



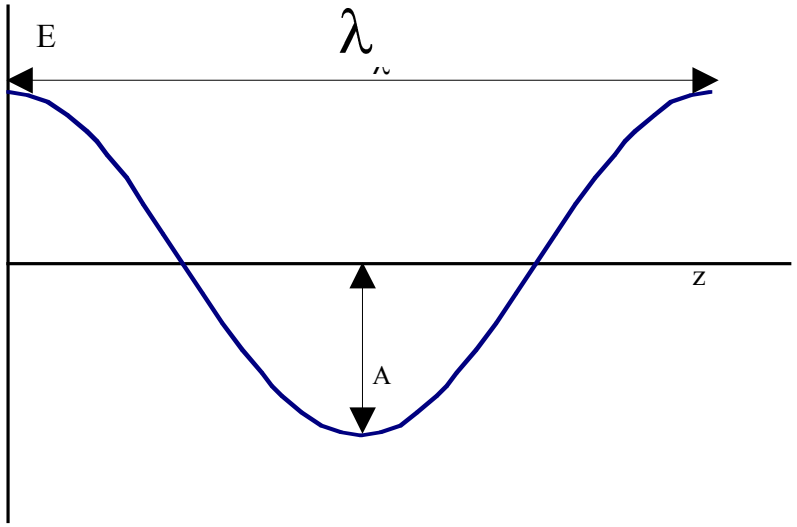
X-ray crystallography

- 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

14 Bravais Lattices

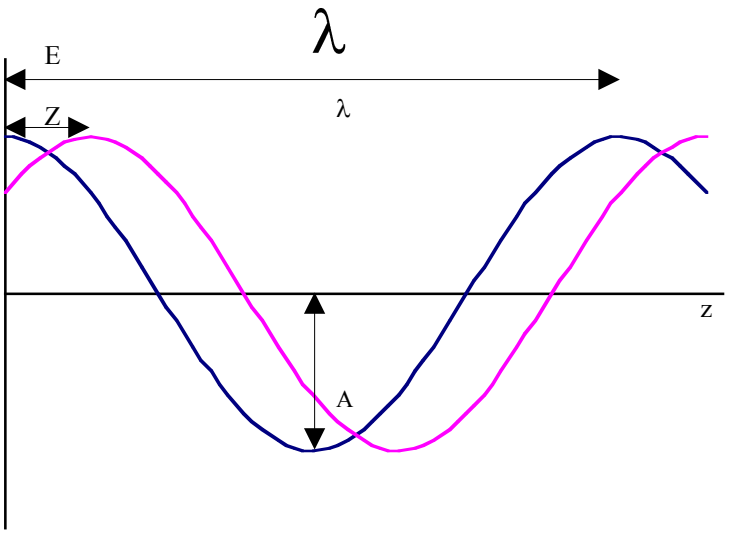


Description of lectromagnetic waves



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

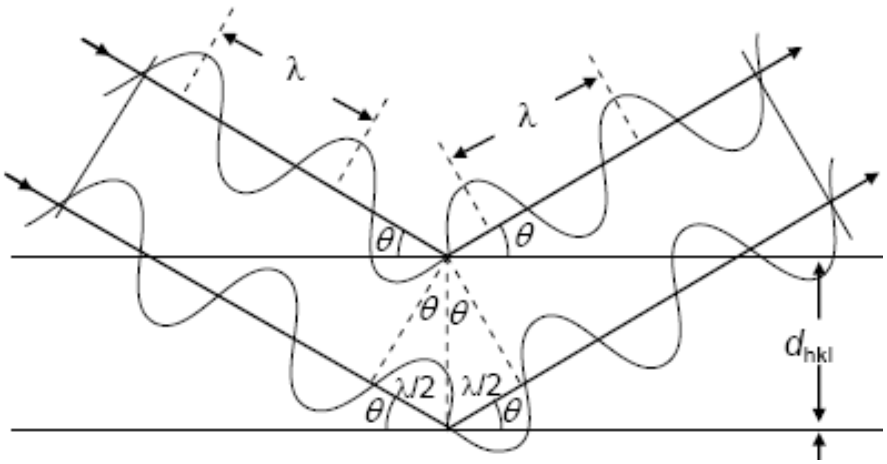
- E- electric field strength
- A- amplitude
- λ - wavelenght

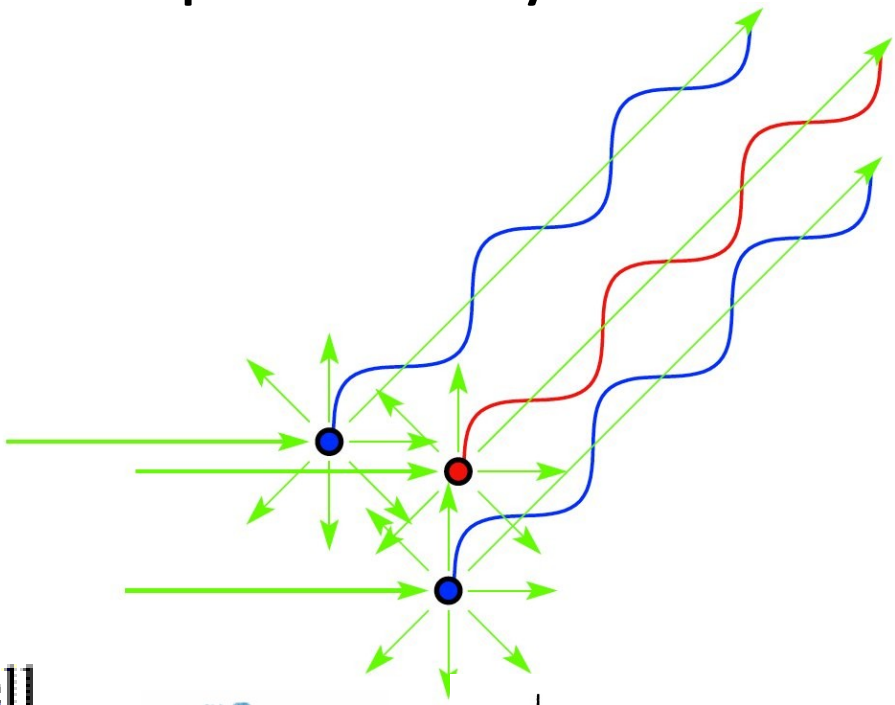


$$E = A \cos (\alpha + 2\pi z/\lambda)$$

- z - position along beam path
- α - phase

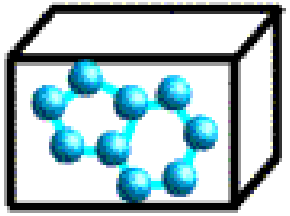
Diffraction pattern from a protein crystal

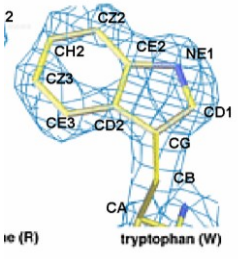


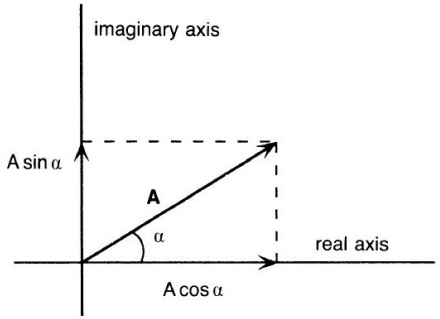


$$n\lambda = 2d \sin\theta$$

unit cell



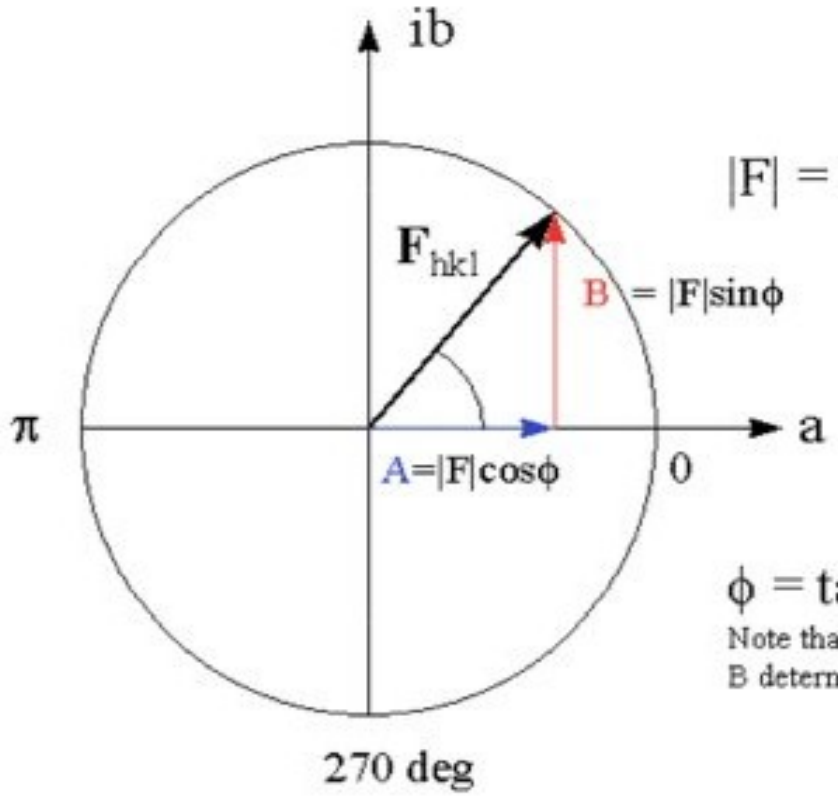




$$F(h\,k\,l) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x\,y\,z) \exp[2\pi i(hx + ky + lz)]\, dx\, dy\, dz$$

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

Phase problem



$|F| = (A^2 + B^2)^{1/2}$

α

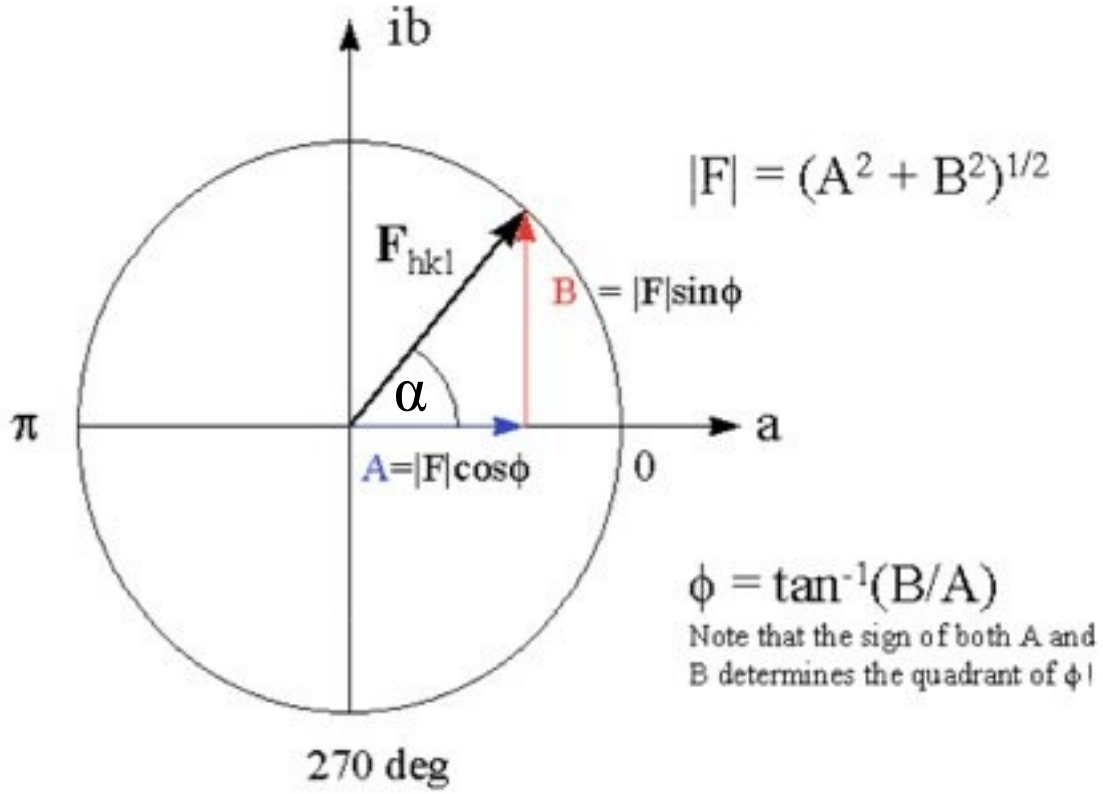
$\phi = \tan^{-1}(B/A)$

Note that the sign of both A and B determines the quadrant of ϕ!

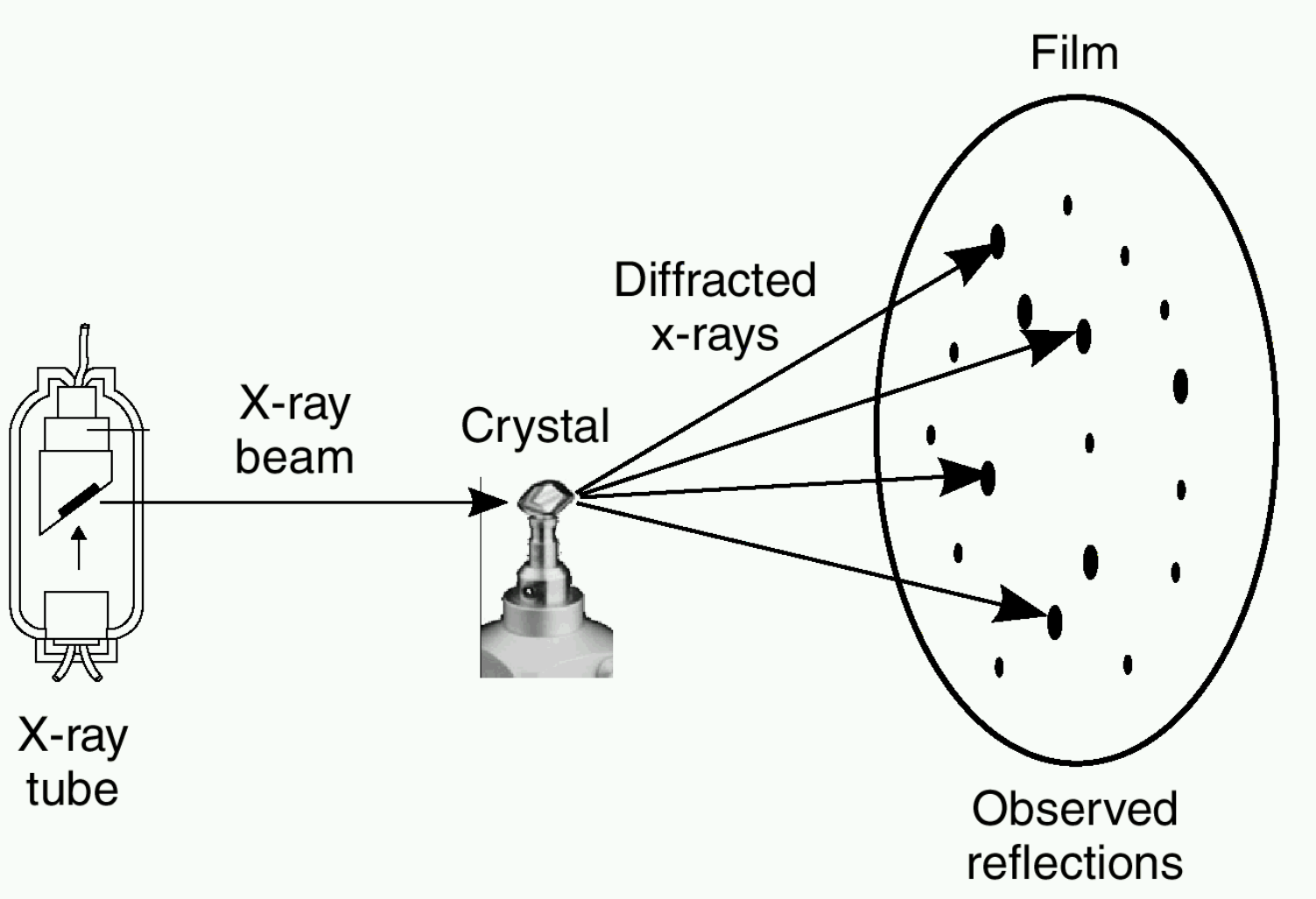
α

Phase problem

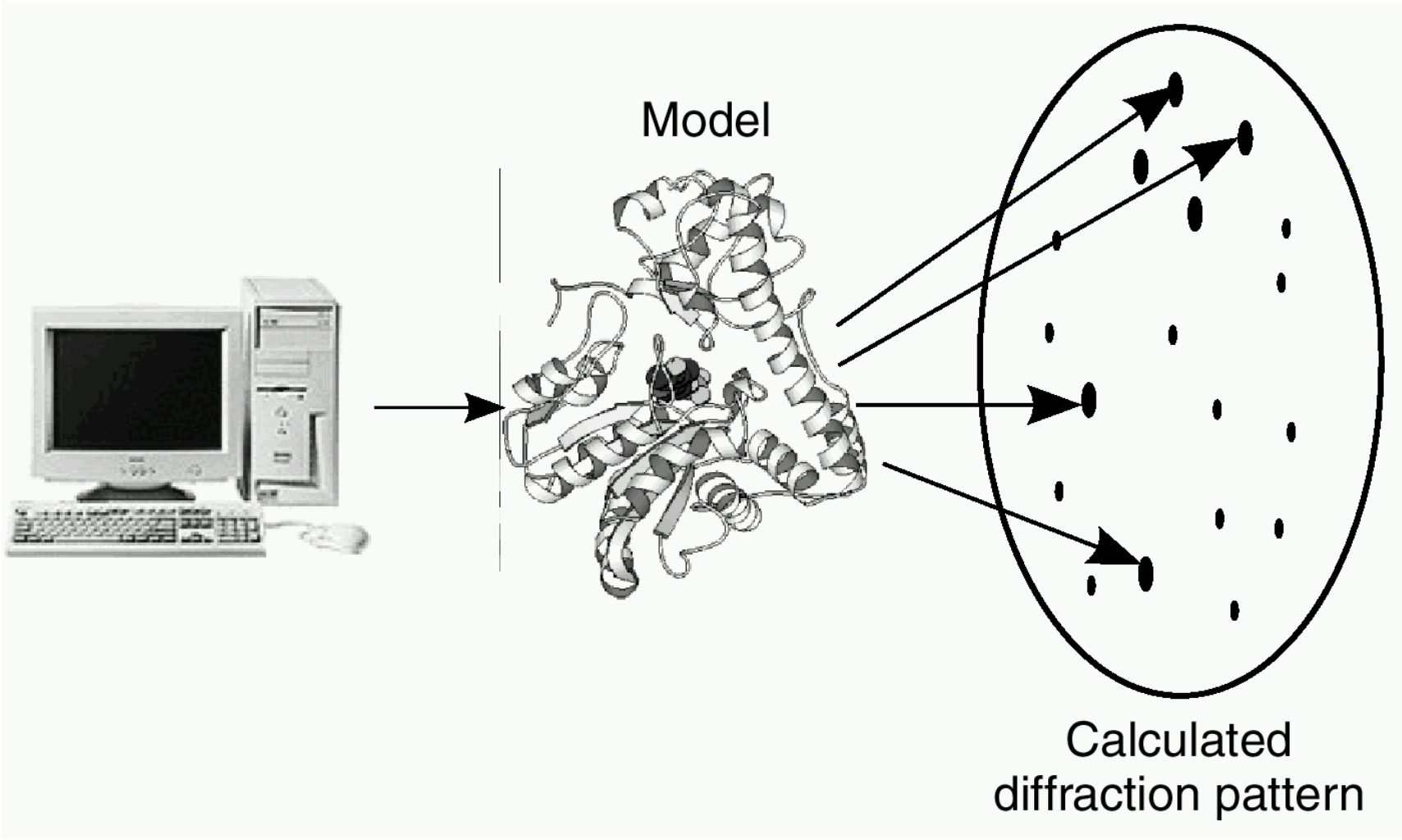
- F_{hkl} is complex and can be represented with an Argand diagram.
- F_{hkl}=A+iB
- We measured $|F_{hkl}|$ in the experiment but we still need α_{hkl}.



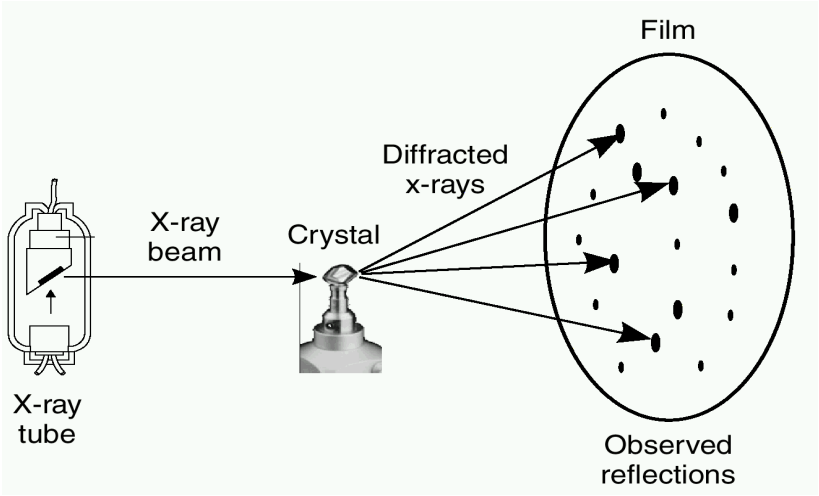
Building macromolecular structures

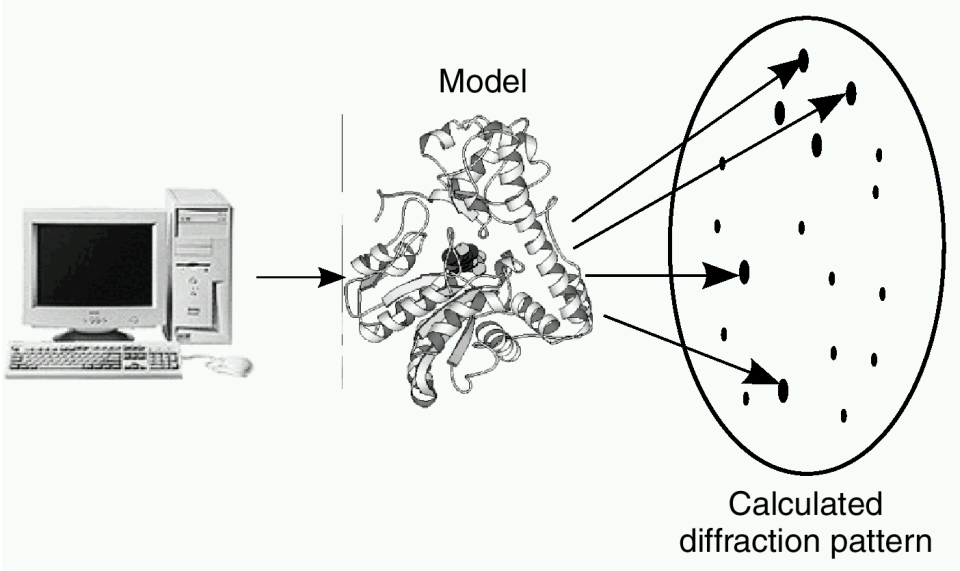


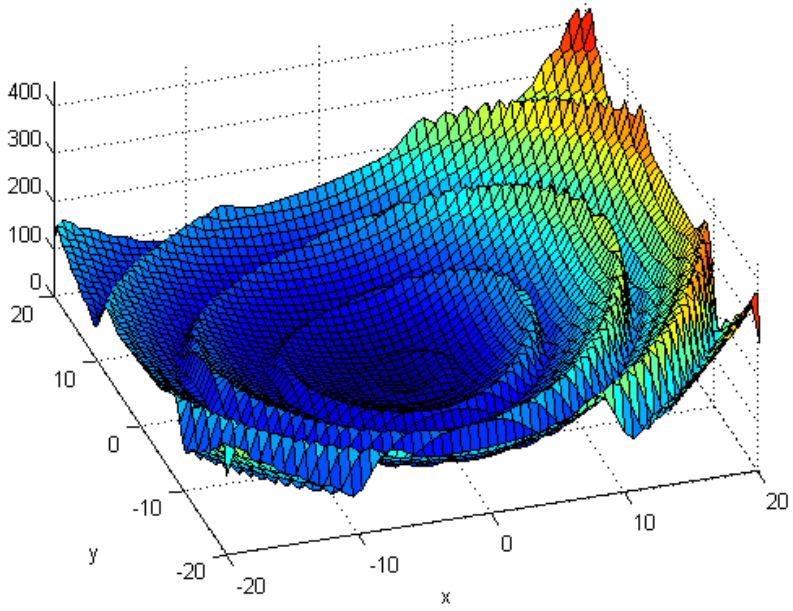
Building macromolecular structures



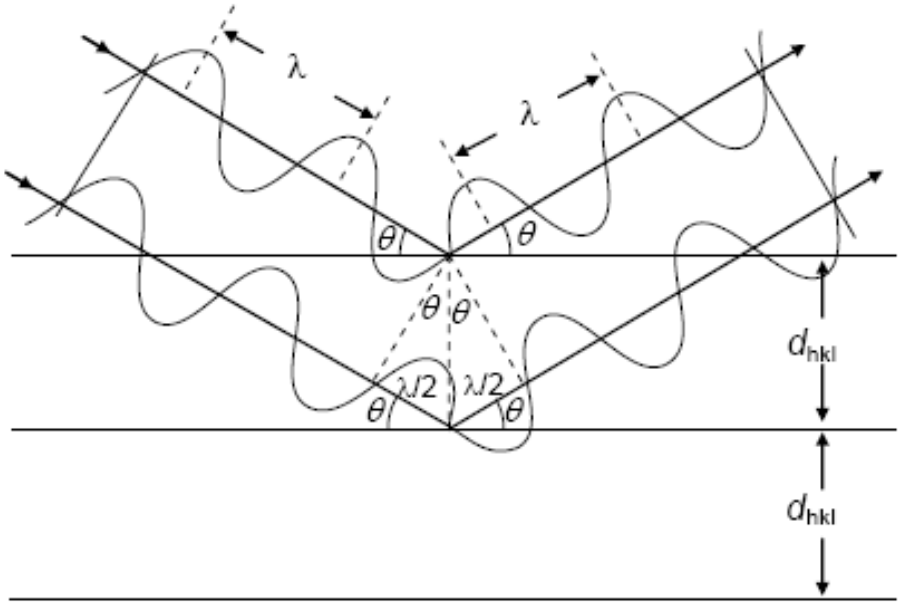
Building macromolecular structures

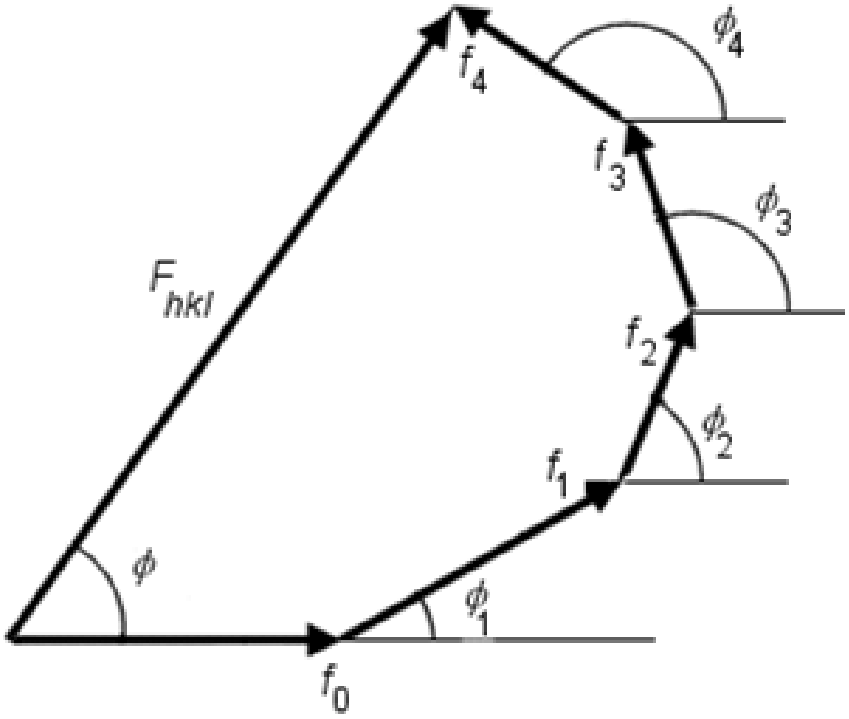






Phase problem



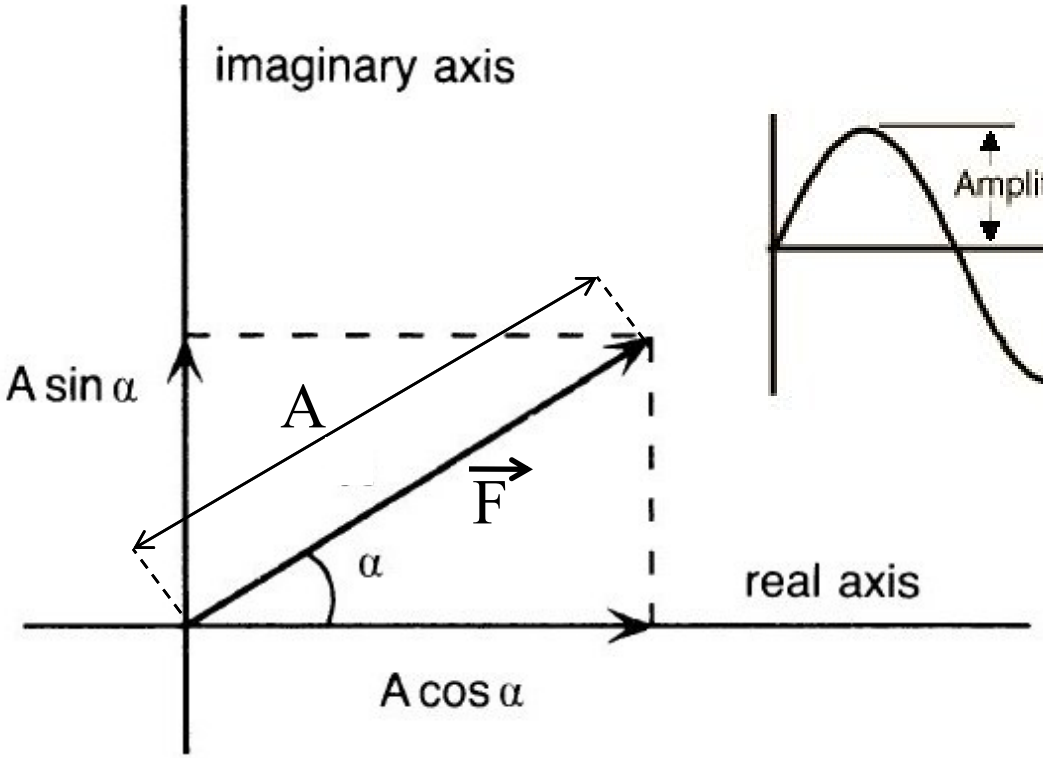


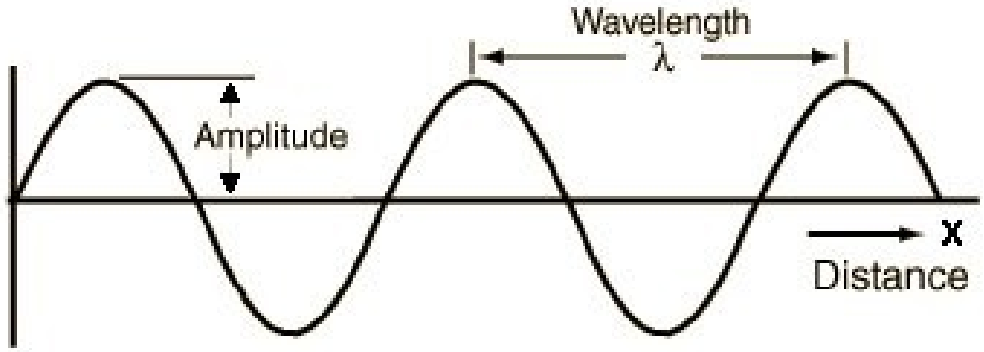
Electron density equation

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

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

Wave as a vector





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

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

A - wave amplitude

α - wave phase

Patterson function, Patterson space