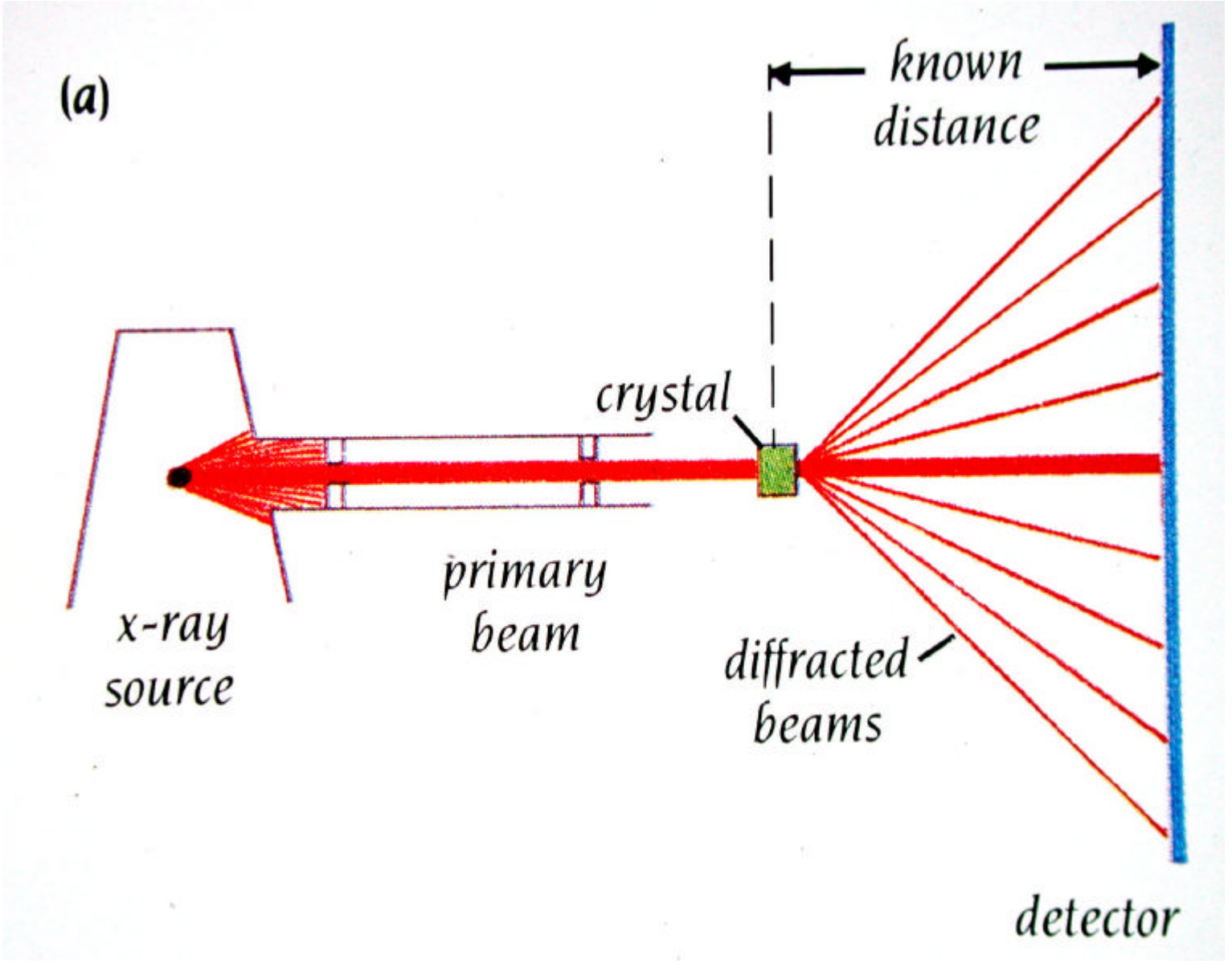
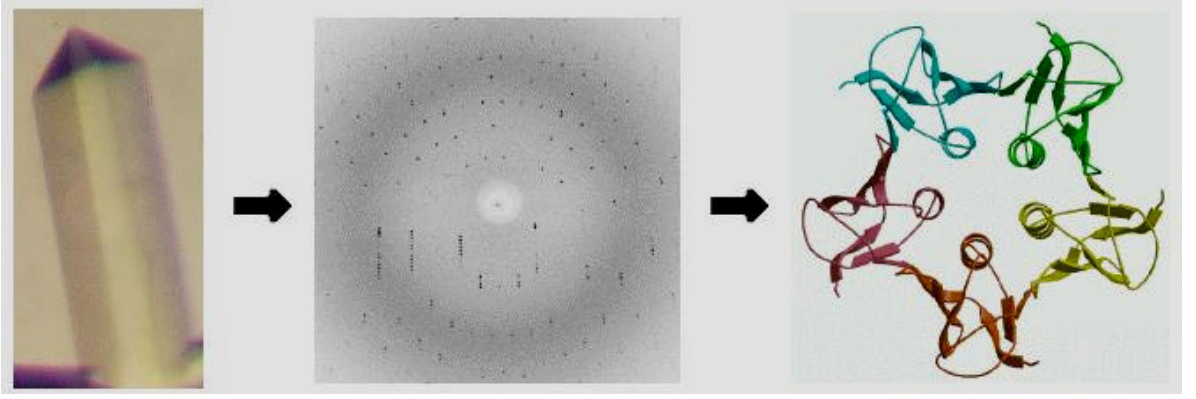
X-ray crystallography



X-ray crystallography (www-structmed.cimr.cam.ac.uk/course.html)



X-ray sources:

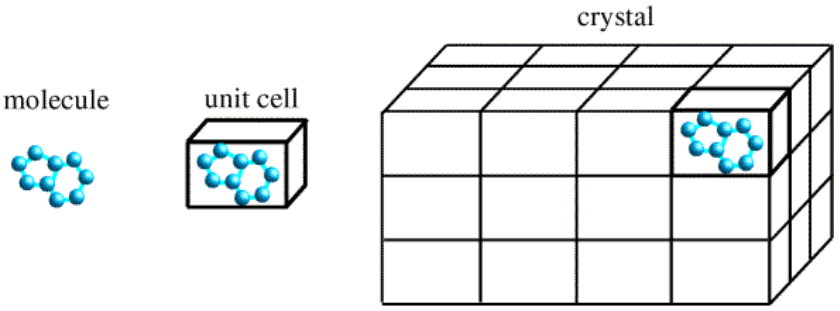
1. Conventional X-ray generator (Cu K_α-line ~ 1.5418 Å)
2. Synchrotron source (Variable wavelength)

Goal: Determine the electron density distribution of molecules.

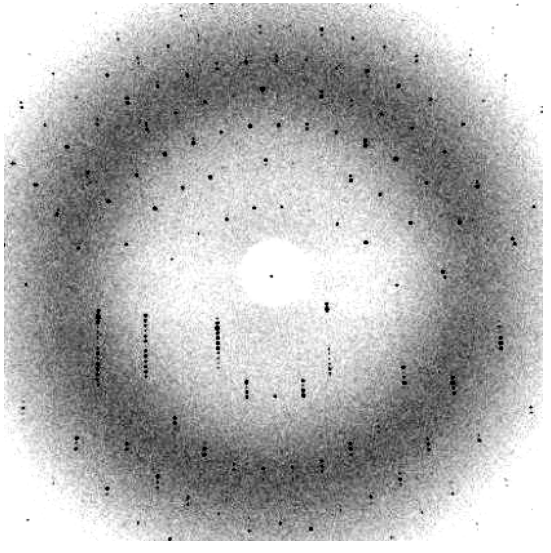
Why crystals?

X-ray scattering from a single molecule would be unimaginably weak and could never be detected above the noise level, which would include scattering from air and water. A crystal arranges huge numbers of molecules in the same orientation, so that scattered waves can add up in phase and raise the signal to a measurable level. In a sense, a crystal acts as an amplifier. Of course, if the waves add up in phase in some directions, they have to cancel out in a lot of other directions. That is why the diffraction pattern from a crystal is an array of spots.

Diffraction:



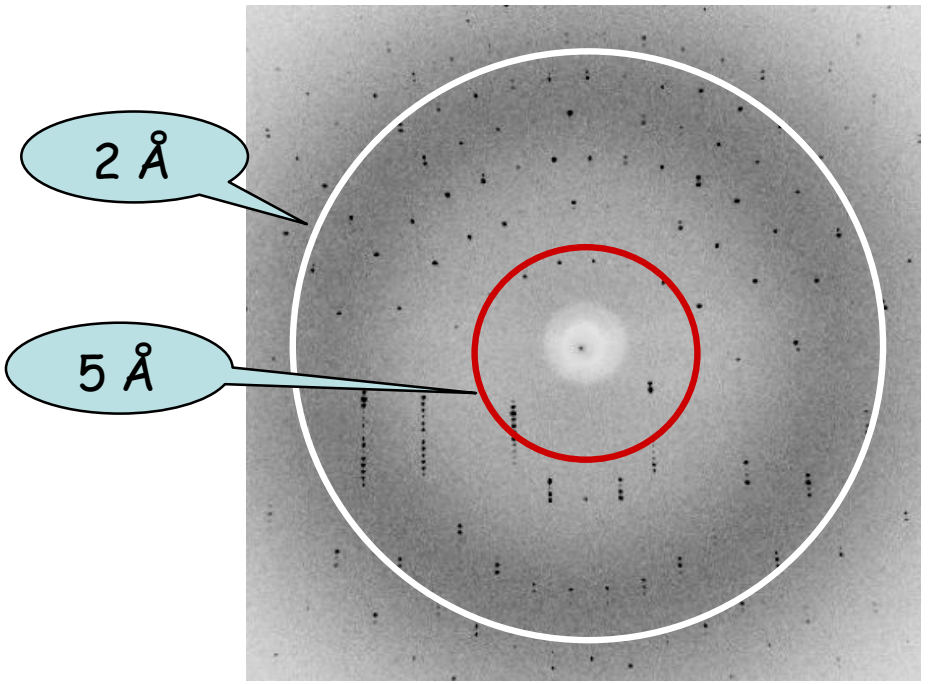
- The observed pattern is the result of light interference due to diffraction from the lattice.
- Diffraction pattern depends on the crystal symmetry and the intensity of each spot is modulated by the protein structure.



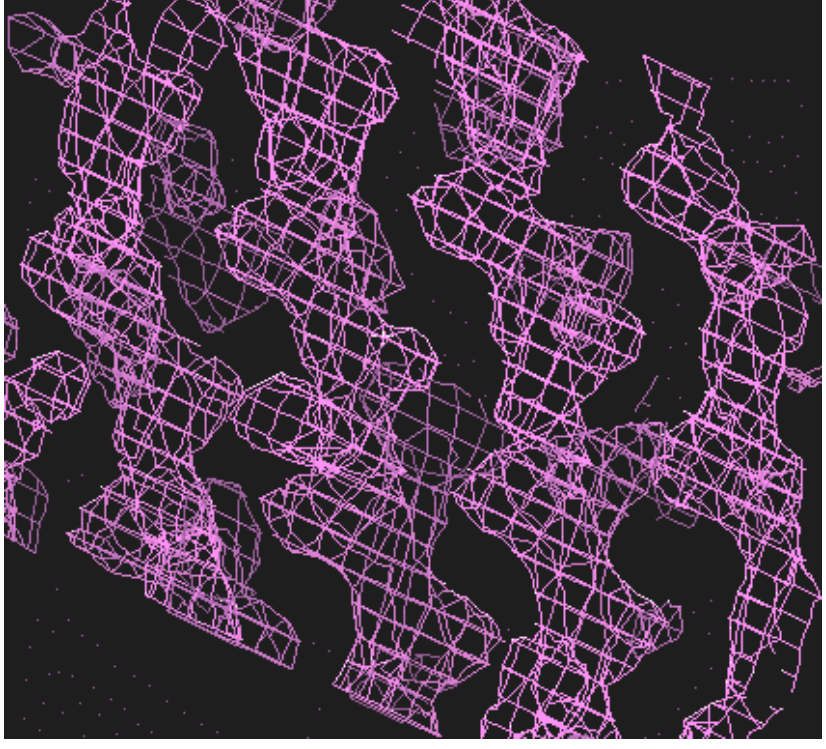
- The diffraction pattern is related to the object diffracting the waves through a mathematical operation called the Fourier transform.
- If you think of the electron density as a mathematical function, then the diffraction pattern is the Fourier transform of that function.
- Spots near the center are due to larger spacing and those on the edge are due to smaller space (Reciprocal space).
- We can only measure the intensity, not the phase of the spots
  → **The phase problem.**
- Solving the phase is a major challenge in X-ray crystallography.
  → Multiple Isomorphic Replacement (MIR) with heavy atom labeling.
  → multiple-wavelength anomalous dispersion (MAD) with methionine replaced by selenomethionine in synchrotron source.
  → Molecular replacement (Guess the phase).

Resolution:

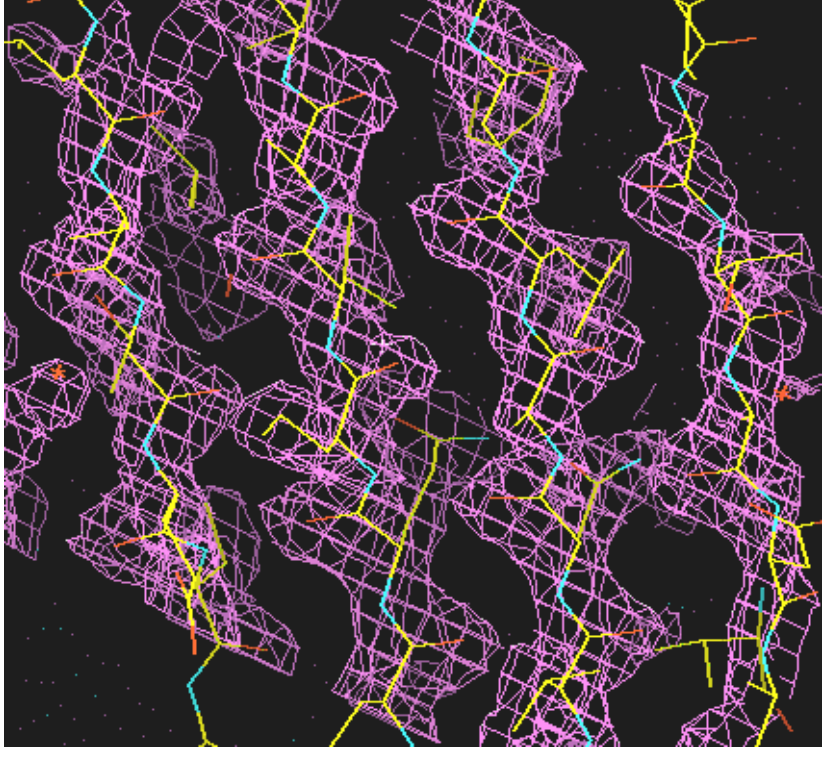
Due to various factors, such as crystal imperfection, molecular motion or wavelength limitation one can only observe diffraction pattern to certain up to certain angle, corresponding to certain distance, thus limit the resolution one can get. The smallest distance one can observe is called the resolution.



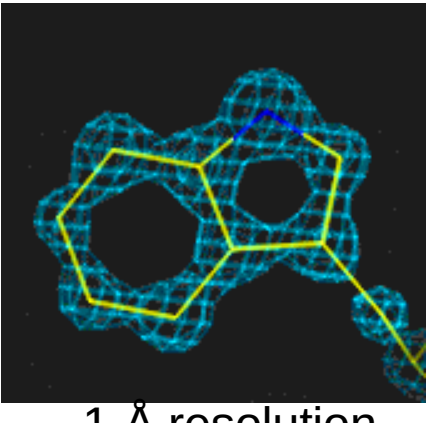
2 Å resolution means diffraction data up to 2 Å are used for calculating the structure.



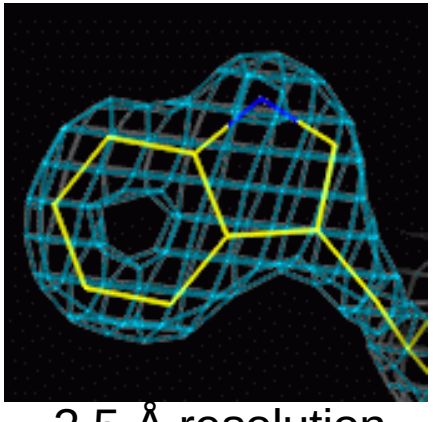
Electron density map



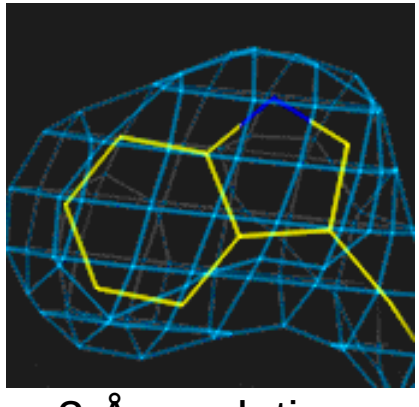
Fitting of electron density map



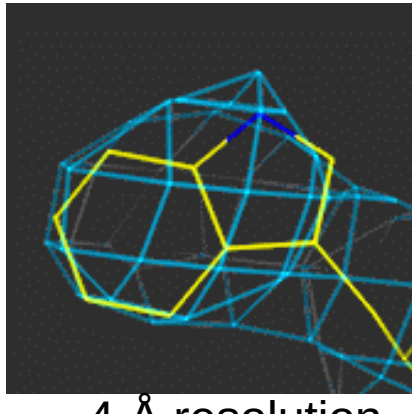
1 Å resolution



2.5 Å resolution



3 Å resolution



4 Å resolution

Fitting and refinement

Because the density map doesn't resolve individual atoms, fitting models to density is a bit of an art. It requires the use of computer graphics programs such as O or XtalView. Early in a structure determination, the phase information is usually poor as well, so the electron density maps are not ideal. As a result, the initial model that one builds will have a lot of errors.

An atomic model can never be perfect, but it can be improved a great deal by a process called refinement, in which the atomic model is adjusted to improve the agreement with the measured diffraction data. The success of an atomic model is often judged through the standard crystallographic R-factor, which is simply the average fractional error in the calculated amplitude compared to the observed amplitude. Though it depends on a number of factors, as a rule of thumb a good structure will have an R-factor in the range of 15% to 25%.

The R-factor: $R = [\Sigma_{h,k,l}|F_{obs}(hkl) - F_{cal}(hkl)|]/[\Sigma_{h,k,l}F_{obs}(hkl)]$

The smaller the R the better the structure. R = 0.2 is good and 0.59 is random.

Refinement of the model:

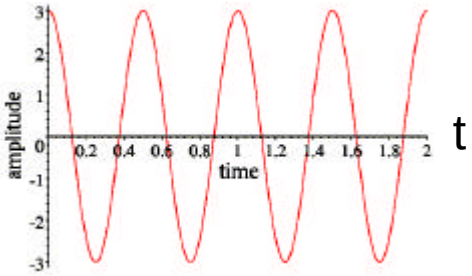
The R-factor: $R = [\Sigma_{h,k,l}|F_{obs}(hkl) - F_{cal}(hkl)|]/[\Sigma_{h,k,l}F_{obs}(hkl)]$

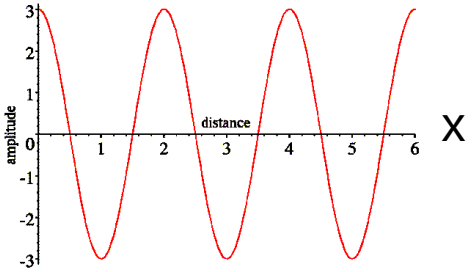
The smaller the R the better the structure. R = 0.2 is good and 0.59 is random.

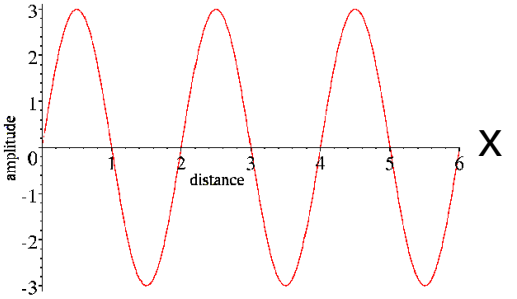
Teaching of X-ray crystallography web site:

(www-structmed.cimr.cam.ac.uk/course.html)

Basic theory: Wave equation: $A\cos[2\pi(\nu t - x/\lambda)]$

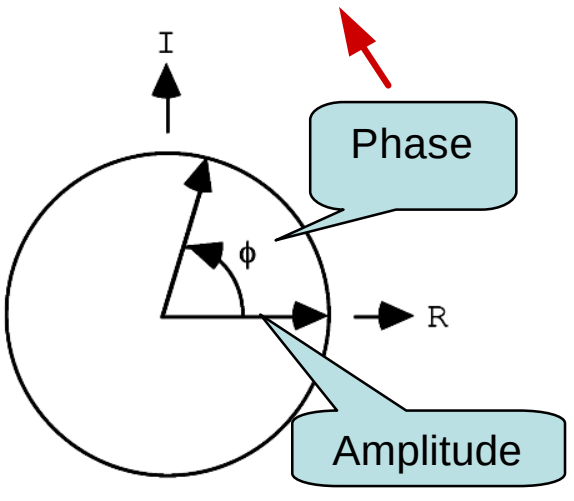






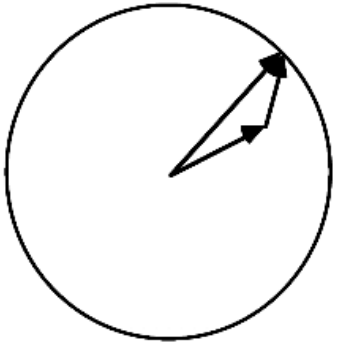
Vector representation:

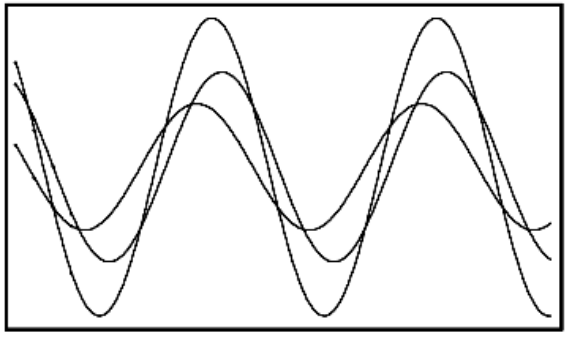
$A\cos(a+f_1)$



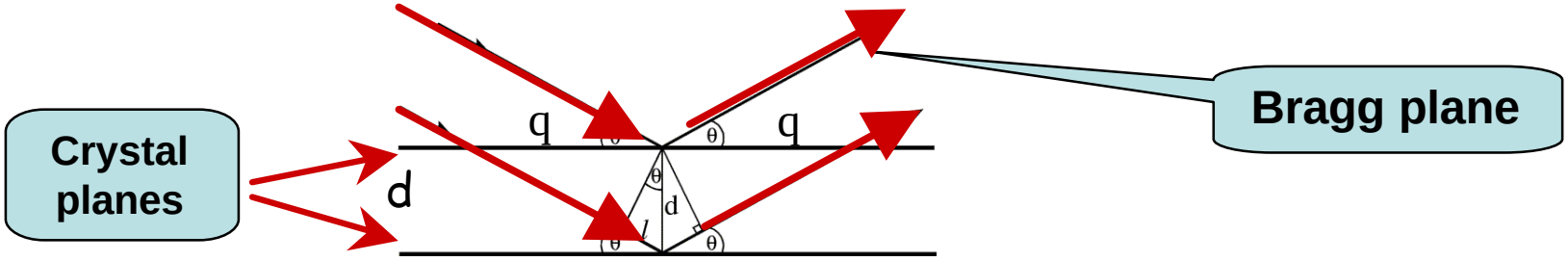
Adding wave together:

$A\cos(a+f_1) + B\cos(a+f_2)$





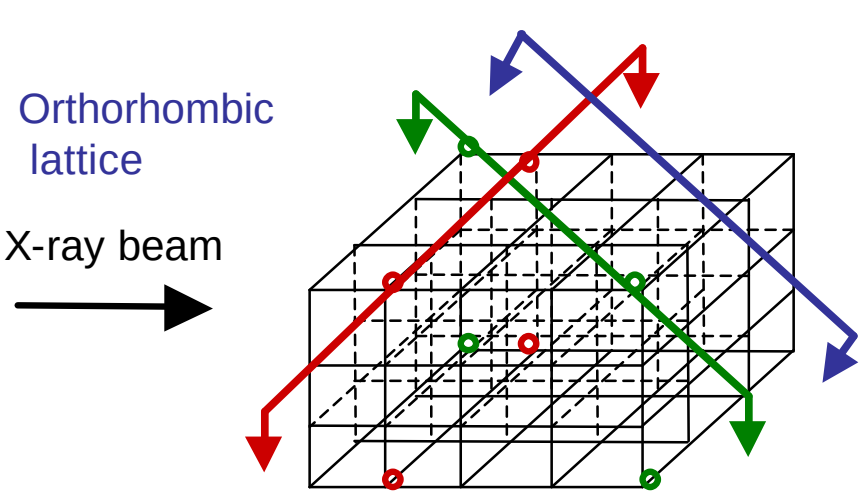
Interference due to diffraction:

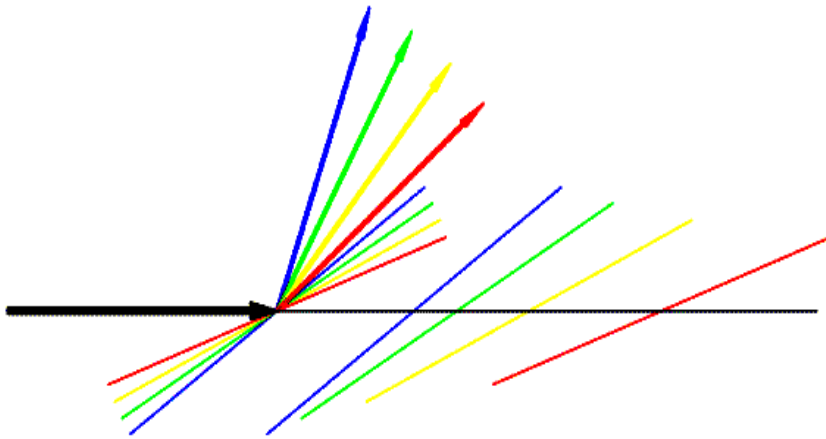


Bragg's law: $\lambda = 2\,d\sin\theta$ or $\sin\theta/\lambda = 1/2d$

(Constructive interference)

Sinq μ 1 / d ➔ Larger d smaller angle θ ➔ Reciprocal space





One can either rotate the crystal or the x-ray beam to find all planes

In real crystal:

1. Lattice symmetry determines the diffraction pattern.
2. Protein structure in a unit cell modulates the intensities of the spots.
3. We want to determine the structure of the protein, thus need to measure the spot intensities and phases.

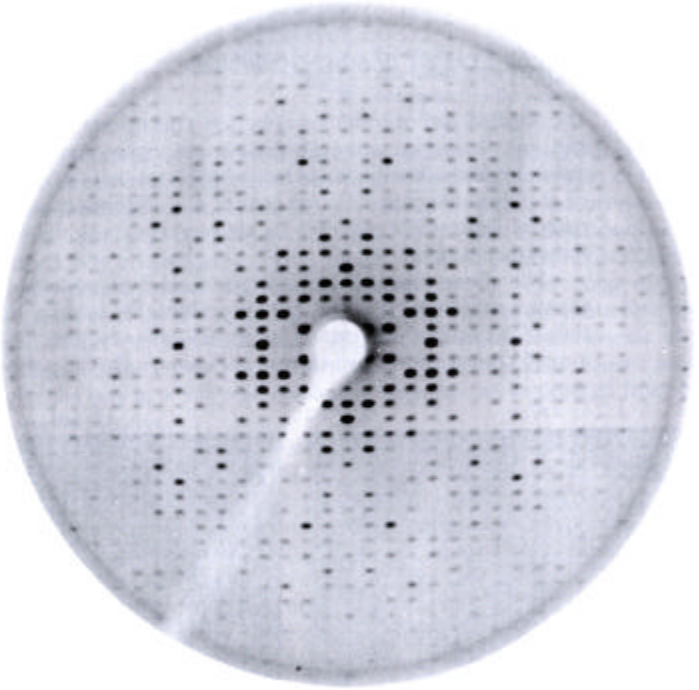
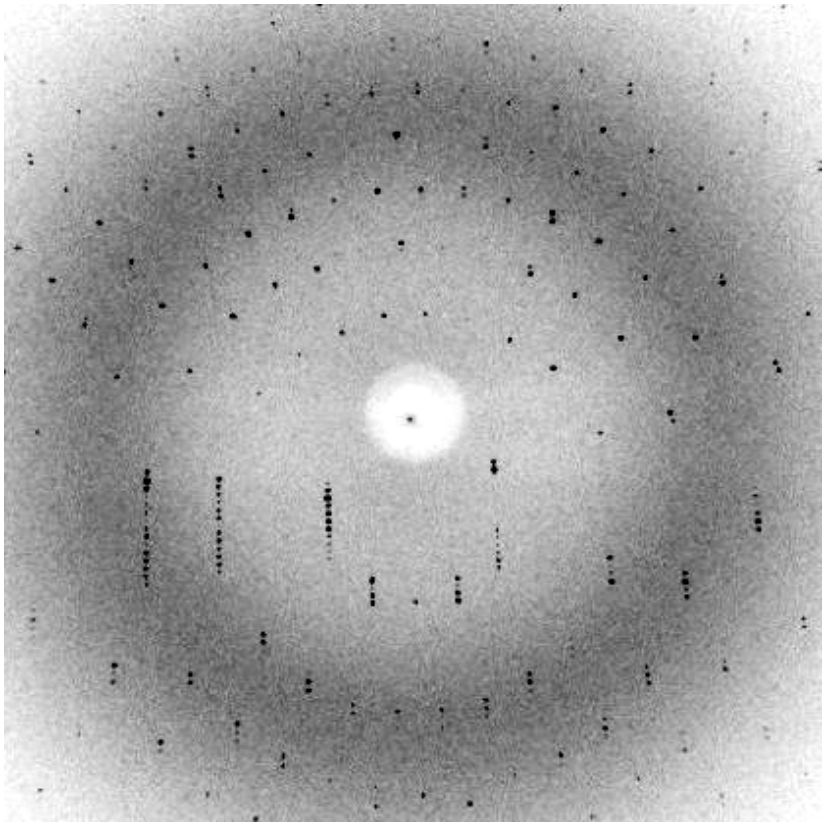
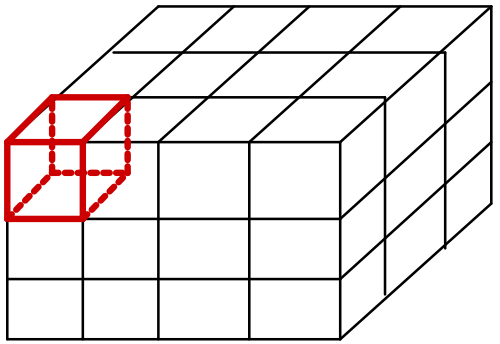


Figure 11.10 X-ray diffraction pattern from a crystal of horse heart oxidized cytochrome *c*. Courtesy of R. E. Dickerson.



Diffraction from a protein crystal

Orthorhombic lattice



Protein crystal

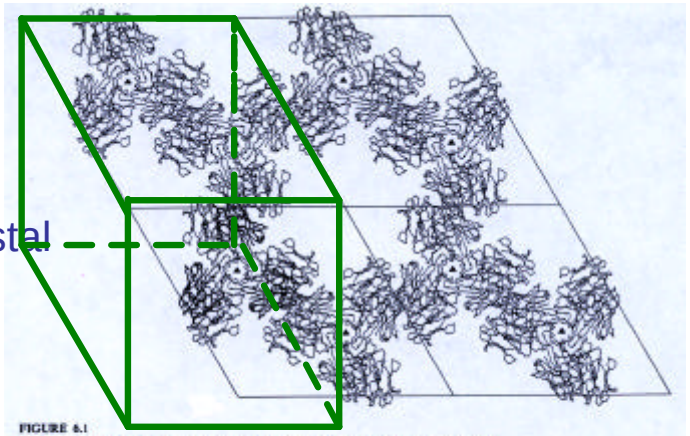
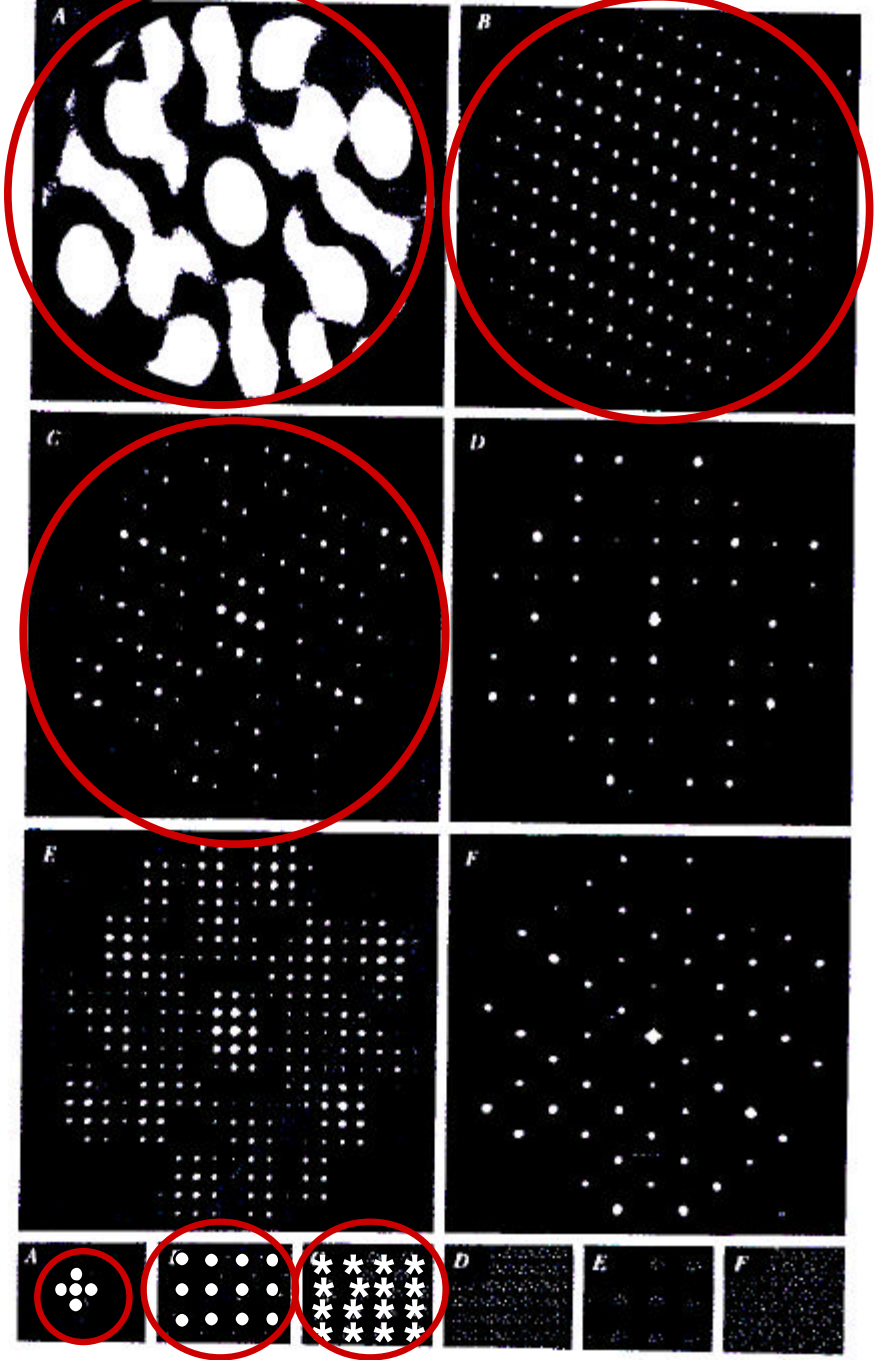


FIGURE 6.1
Crystal structure of the immunoglobulin Fab McPC603 in which 70% of the volume is solvent. Four unit cells are shown, with the Fab molecules indicated by the backbones of the heavy and light polypeptide chains that comprise each molecule. The heavy chains have been drawn bolder in the three molecules that are clustered about a crystallographic three-fold axis (triangle) in the lower left-hand cell. The clusters of three Fab molecules are maintained principally by a set of hydrogen bonds and van der Waals contacts between the molecules. The space group is $P6_122$; twofold screw axes are located midway between each adjacent pair of threefold axes, and a 6_2 axis is located at each corner of the unit cell.
(From Y. Satow et al., *J. Mol. Biol.* 190:593–604, 1986.)

Determine the structure of protein, not the lattice !



Von Laue conditions for constructive inerference:

$l\lambda = c \cdot \cos\gamma$ Where $\cos\gamma$ is the direction cosine.

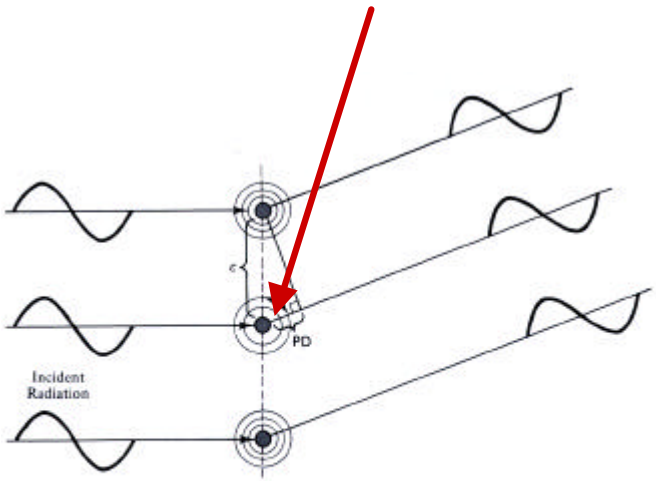
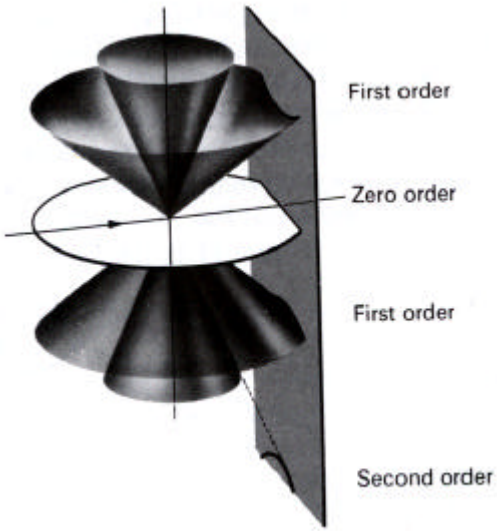


Figure 11.2 Demonstration of the von Laue condition for radiation incident perpendicular to a row of scatterers with regular spacing, c. In scattering directions where the path difference equals a multiple of λ, all scatterers will reinforce. At any other angles, there will be interference, which will become more complete the longer the row of scatterers.



Diff. pattern:

Hyperbolic layer lines a the intersection of the plate and the cones

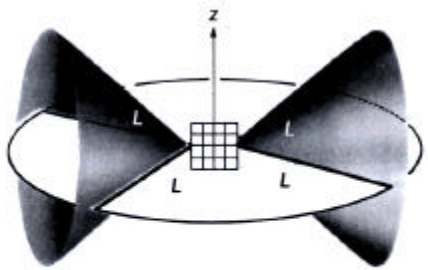
Figure 11.3 The cones along which constructive interference can occur for a row of scatterers. These are shown intersecting a photographic plate. On the plate, layer lines corresponding to $l = 0, 1, 2$, and so forth will appear.

For three dimensional lattice of dimensions a, b, c:

$$l\lambda = c \cdot \cos\gamma$$
$$h\lambda = a \cdot \cos\alpha$$
$$k\lambda = b \cdot \cos\beta$$

$\cos\alpha$, $\cos\beta$ and $\cos\gamma$ are the direction cosines.

Diff. pattern: Spots at the intersections of the plate and the three cones



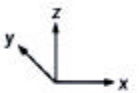


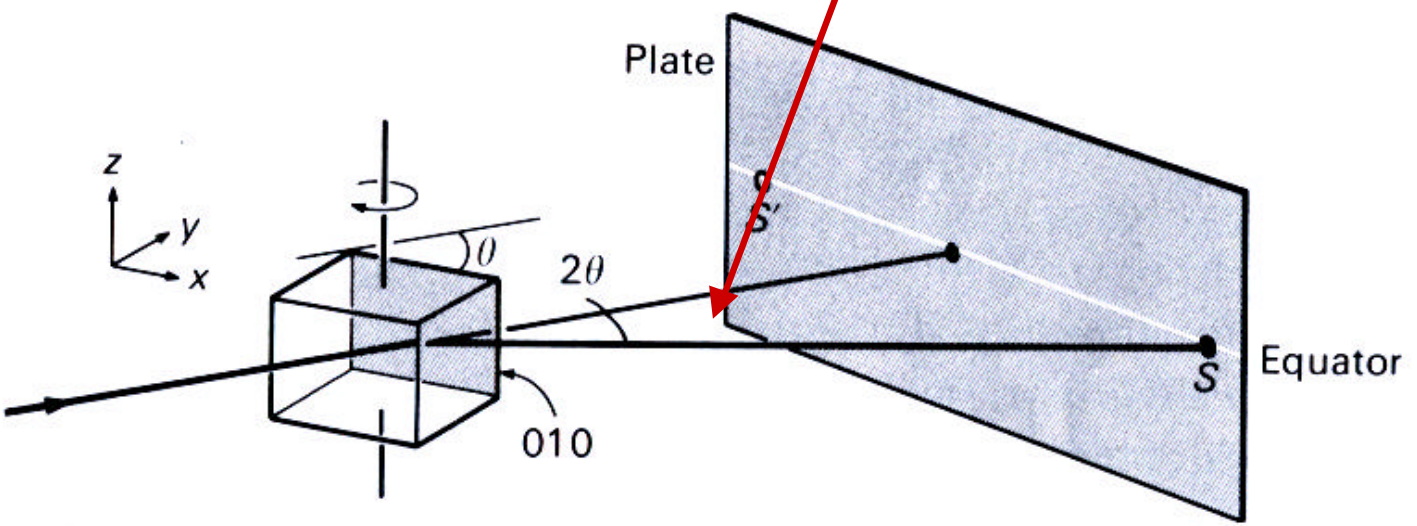
Figure 11.4 Diffraction by a rectangular lattice in the xz plane. For clarity, onl one set of cones of diffraction produced by the x spacings is shown, and only th circle (zero-order reflection) from the z axis is shown. Diffraction will be possibl along the lines (L) where these surfaces intersect. These could be called the 1, reflections.

One can show that for a rhombic crystal with spacings a, b, and c the following condition defines the constructive interference spot (Bragg law):

$$\lambda (h^2/a^2 + k^2/b^2 + l^2/c^2) = 2 \cdot \sin\theta$$

Where θ is the angle between incident beam and diffracted beam

h, k, l are the Miller indices.



Miller indices which define the lattice planes

001 plane (h = 0, k = 0, l = 1) of an orthorhombic unit cell (^ to Z-axis)

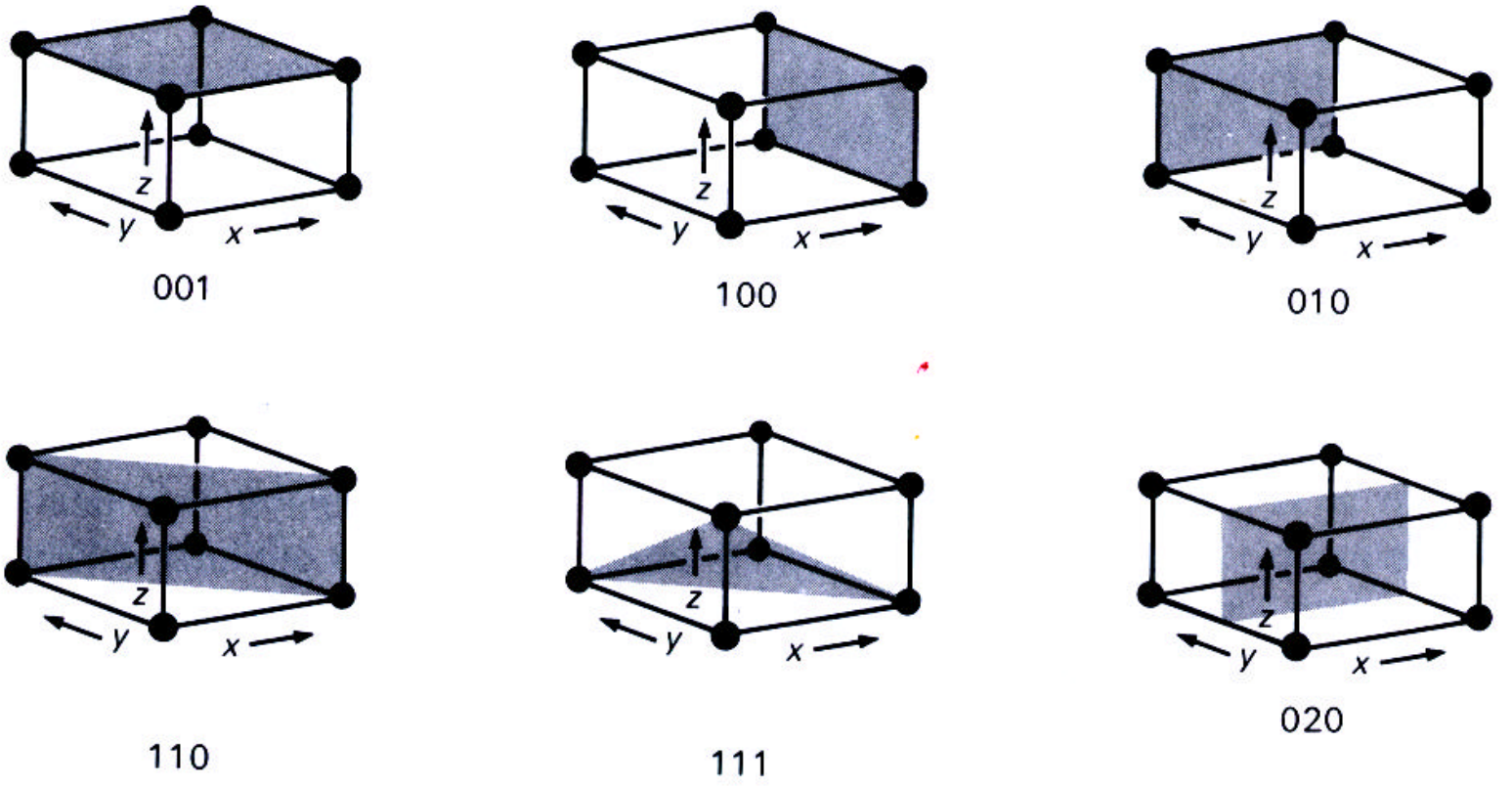


Figure 11.5 Miller indices of some of the possible planes in an orthorhombic unit cell (a rectangular parallelepiped) with sides a, b, c. The numbering rule is as follows: If the plane cuts an axis at a/h, b/k, c/l, the indices are h, k, l.

One needs to rotate either the crystal or the X-ray beam to satisfy the Bragg conditions.

Only certain orientations which satisfy the Bragg conditions give spots.

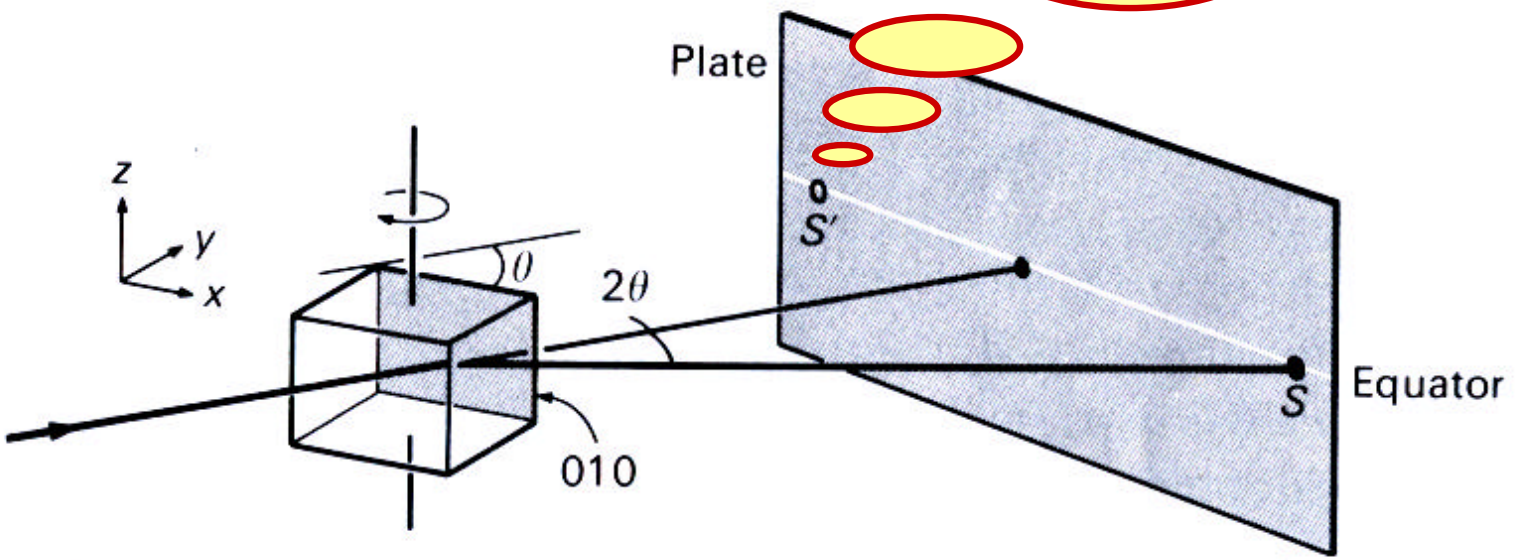


Figure 11.6 The rotating crystal method. The crystal has been oriented with the z axis perpendicular to the beam and is rotated about this axis. When the 010 planes make the Bragg angle θ with respect to the beam, a diffraction spot (S) is produced on the equator of the image at an angle 2θ from the incident beam. Another spot will be observed at S', when the angle is $-\theta$. Higher-order reflections will also be seen. Higher-order spots will be farther out on the pattern.

Consider a protein crystal with a unit cell containing a diatomic molecule

Consider a set of planes in z-direction, i.e. ⊥ to paper

Waves diffracted from plane of the black dots will differ from that diffracted from the open circle atoms by

$$\Delta\phi = 2\pi hx/a$$

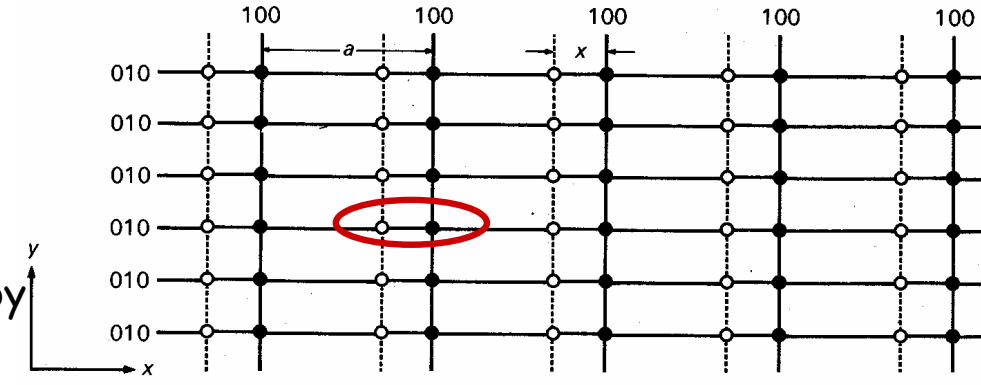


Figure 11.11 The *xy* plane of a hypothetical lattice of diatomic molecules. Note that the same lattice (displaced) may be used either for the white atoms or the black atoms. The molecules are assumed to lie in the plane of the paper.

For a three dimensional crystal this will be:

$$\Delta\phi = 2\pi(hx/a + ky/b + lz/c)$$

Atomic factor

Considering a jumble of waves given by: $\mathbf{E} = e^{i2\pi\nu t}\Sigma f_k \exp(i\Delta\phi_x)$

Where the amplitude: $F = \Sigma f_k \exp(i\Delta\phi_x)$

For the diatomic molecule: $F_{100} = f_B\exp(i\Delta\phi_B) + f_W\exp(i\Delta\phi_W)$ is called the Structural factor for the 100 reflection.

The intensity is given by: $\mathbf{I}(hkl) = F(hkl)F^*(hkl)$

For the $I(100) = f_B^2 + f_B f_W[\exp(2\pi lx/a) + \exp(-2\pi lx/a) + f_W^2$

In general: $F(hkl) = \int_{\text{Unit cell}} \int \int \rho(x,y,z)\exp(2\pi i(hx/a + ky/b + lz/c))\, dV$ (Fourier Transform)

Where $\rho(x,y,z)$ is the electron density in the volume dV in the unit cell, i.e.:

$$\rho(x,y,z) = 1/V \sum_{h,k,l=-\infty}^{h,k,l=\infty} \sum \sum F(hkl)\exp[-2\pi i(hx/a + ky/b + lz/c)]$$

➔If we know the structure factors F(hkl)'s we would be able to calculate $\rho(x,y,z$

➔But we know only the intensity, i.e. F(hkl)F*(hkl) not the phase (phase problem)

Determine phase with Multiple Isomorphic Replacement:

(MIR with heavy atom derivatives)

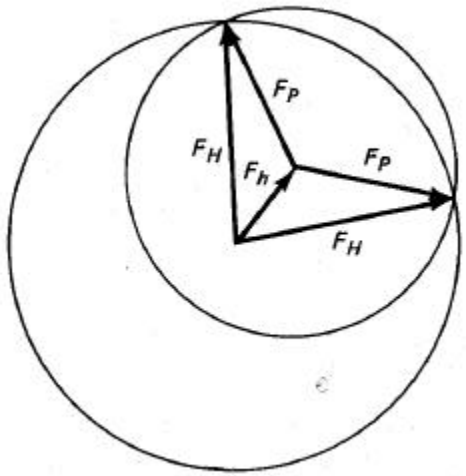
The phase of a protein with heavy atom, $F_H(hkl)$ is:

$$\mathbf{F_H}(hkl) = \mathbf{F_P}(hkl) + \mathbf{F_h}(hkl)$$ (Vector sum)

Atomic diffraction power:
H: 1; C: 36; N: 49; O: 64
Pt.: 6084; Au: 6241; Hg: 6400

The terms on the left are due to native protein alone and heavy atom , respectively. If we can determine $F_{h+}(hkl)$ for several heavy atoms (Both amplitude and phases) and the amplitude of $F_H(hkl)$ and $F_P(hkl)$ then we can determine the phase graphically.

For a given F_h there are two solutions for F_H and F_P, as shown on the left figure. By obtaining several heavy atoms the intercepts can be determined quite precisely.



Determining the positions of heavy atoms by "Patterson map":

Patterson function: $P(hkl) = 1/V \sum_{h,k,l=-\infty}^{h,k,l=\infty} \sum \sum F^2(hkl)\exp[-2\pi i(hx/a + ky/b + lz/c)]$

1. Contains only the intensities F^2 and can be calculated readily)
2. P(hkl) is a measure of the densities of distances between groups.

For a unit cell containing three scatterers. If we measure all vectors between these Centers, move the tails of these vectors to the origin, and graph the positions of their heads, we will have the resulting Patterson map.

The map for a protein is too complex but the "difference Patterson Map" between native protein and protein with heavy atoms is rather simple and can be used to locate the heavy atoms in a unit cell.

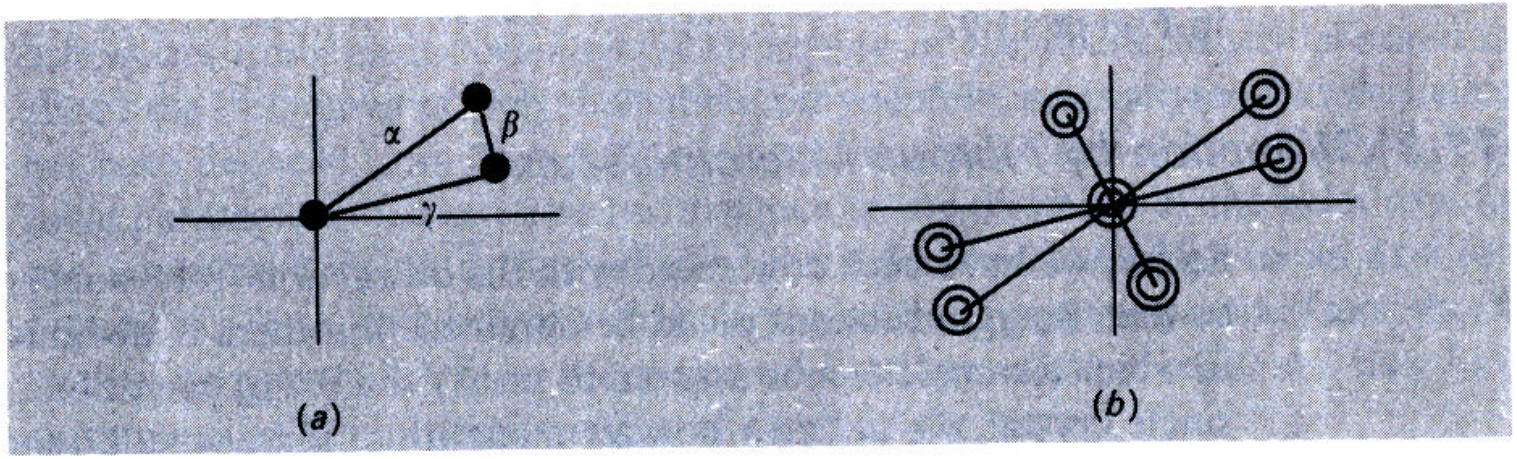


Figure 11.14 A planar three-atom "molecule" (*a*) and its corresponding Patterson function (*b*).

In real crystal:

1. Lattice symmetry determines the diffraction pattern.
2. Protein structure in a unit cell modulates the intensities of the spots.
3. We want to determine the structure of the protein, thus need to measure the spot intensities and phases.

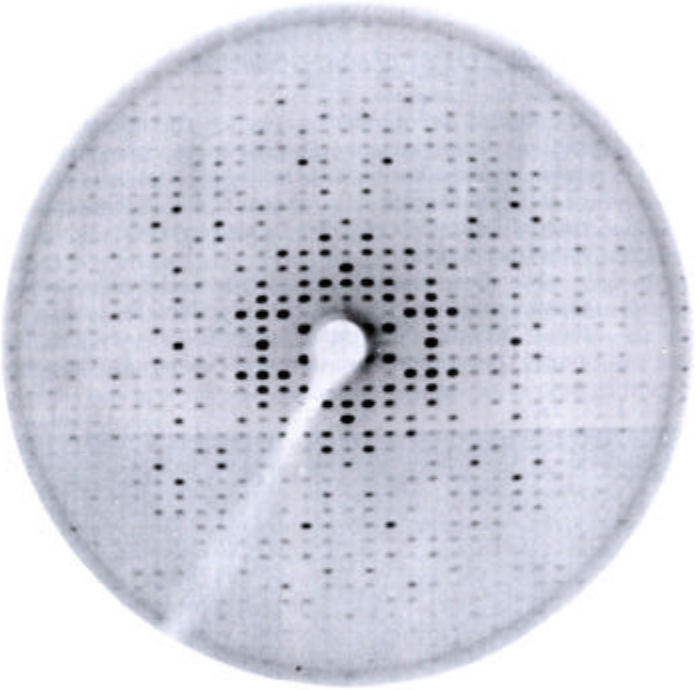
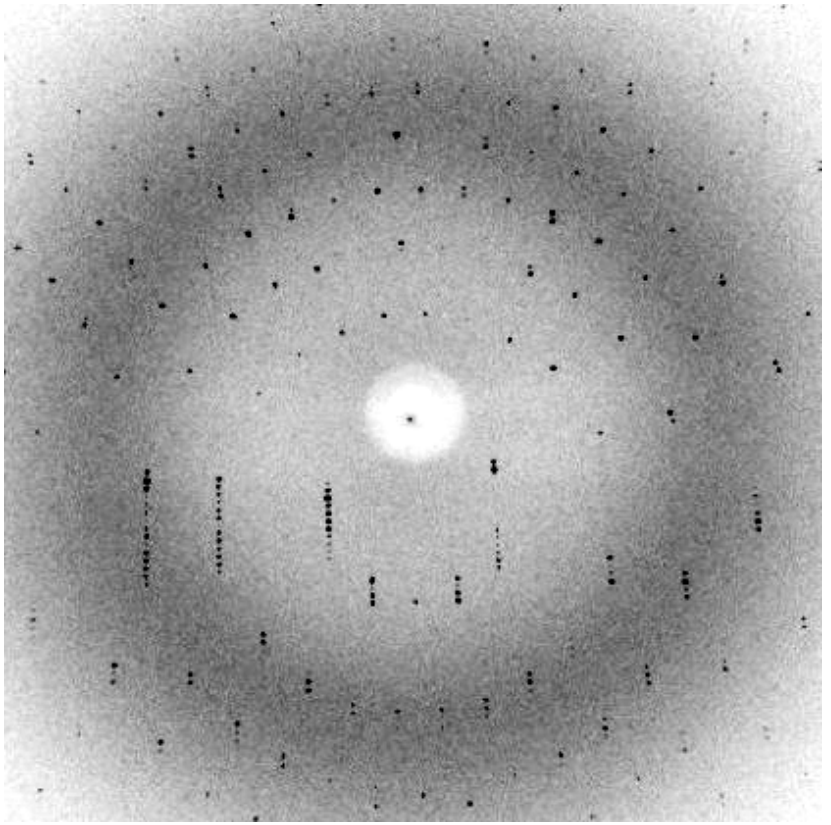


Figure 11.10 X-ray diffraction pattern from a crystal of horse heart oxidized cytochrome *c*. Courtesy of R. E. Dickerson.



Refinement of the model:

The R-factor: $R = [\Sigma_{h,k,l}|F_{obs}(hkl) - F_{cal}(hkl)|]/[\Sigma_{h,k,l}F_{obs}(hkl)]$

The smaller the R the better the structure. R = 0.2 is good and 0.59 is random.

The B-factor (temperature factor or Debye-Waller factor):

The extent of disorder of each atom to the diffraction pattern can be taken into account by weighting the factor:

$$Exp(-B_i \sin^2\theta/\lambda^2)$$

Teaching of X-ray crystallography web site:

(www-structmed.cimr.cam.ac.uk/course.html)

Powder diffraction:
See lattice spacing but not the structure of a protein in a unit cell.

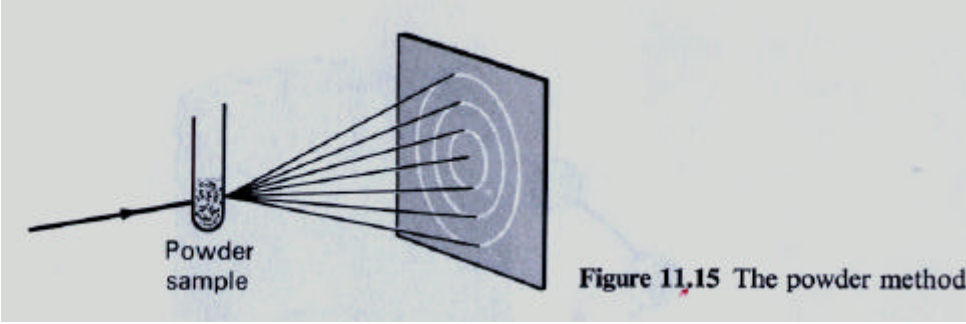


Figure 11.15 The powder method

Fiber diffraction:

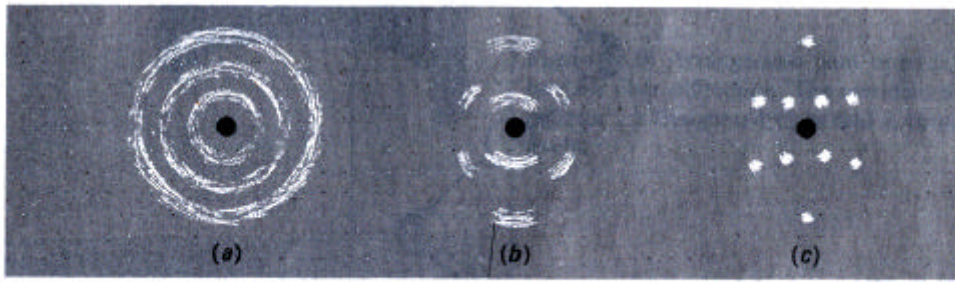


Figure 11.16 Diffraction from a fiber in which molecules are (*a*) unoriented, (*b*) partly oriented, and (*c*) highly oriented.

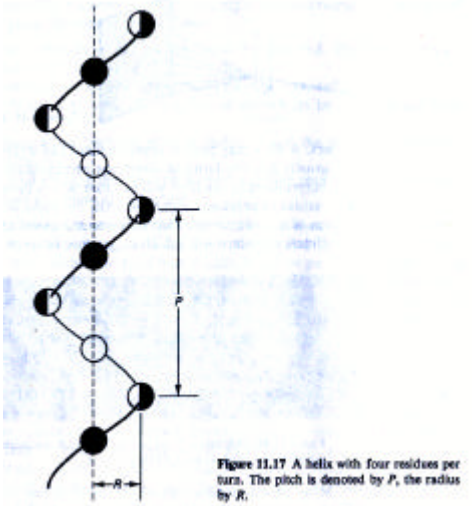
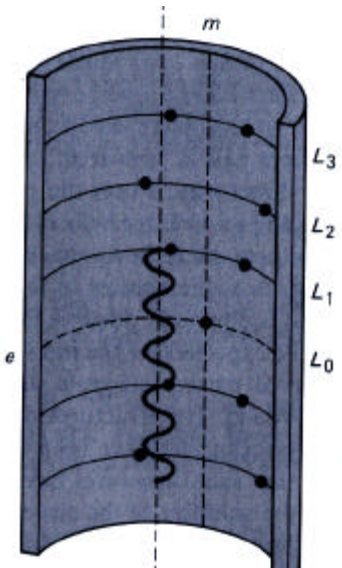


Figure 11.17 A helix with four residues per turn. The pitch is denoted by *P*, the radius by *R*.



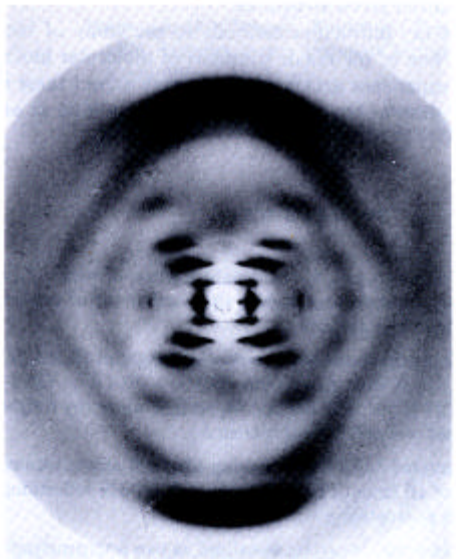
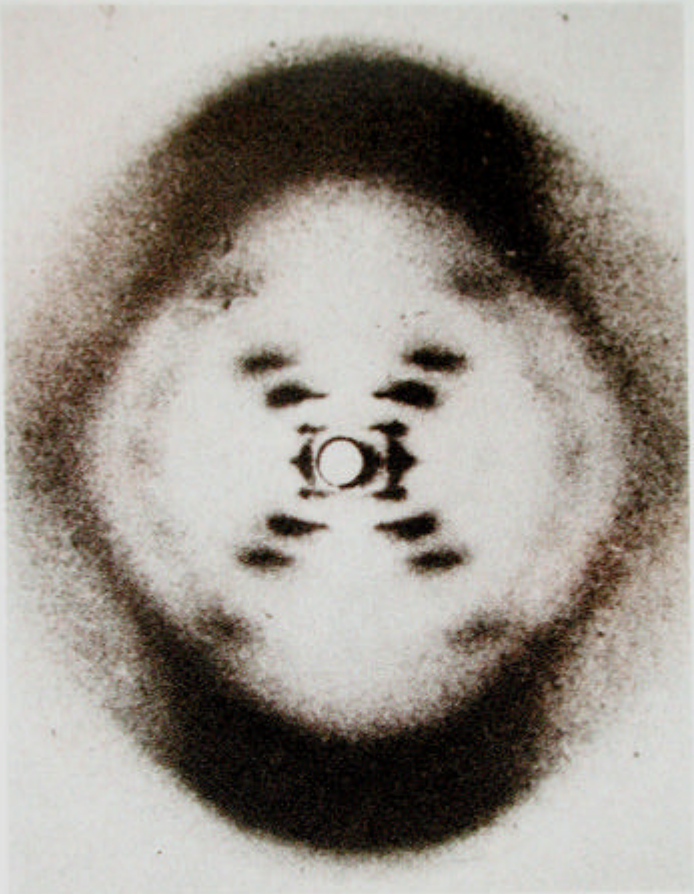
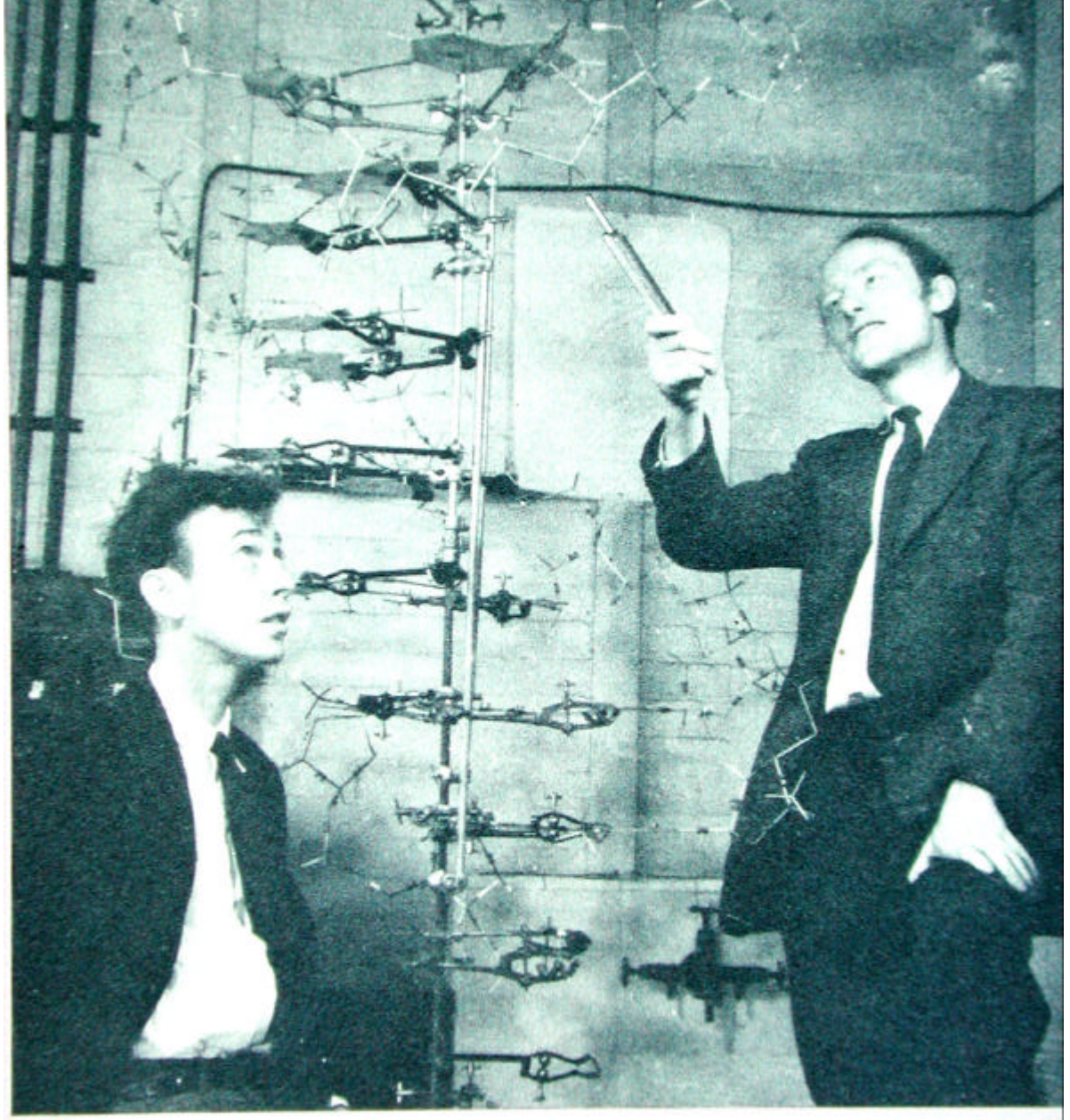


Figure 11.18 Arran
ture for fiber diffr

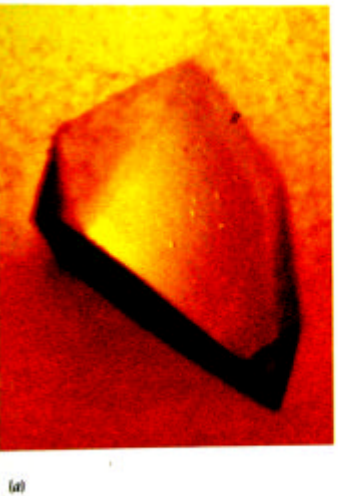
NA, which has been
xtracted from cells of the
acterium *E. coli*, on the end
f a glass rod.

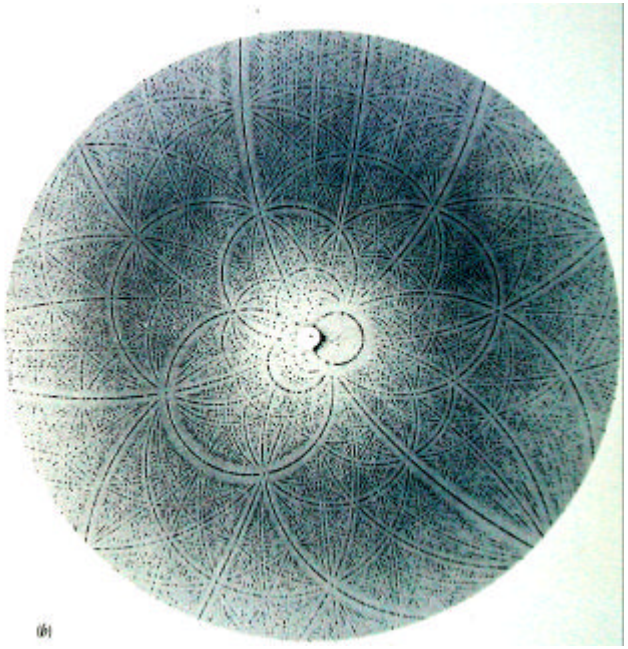


The X-ray diffraction photograph of DNA taken by Rosalind Franklin that was published with the paper by Watson and Crick in *Nature*, April 25, 1953.

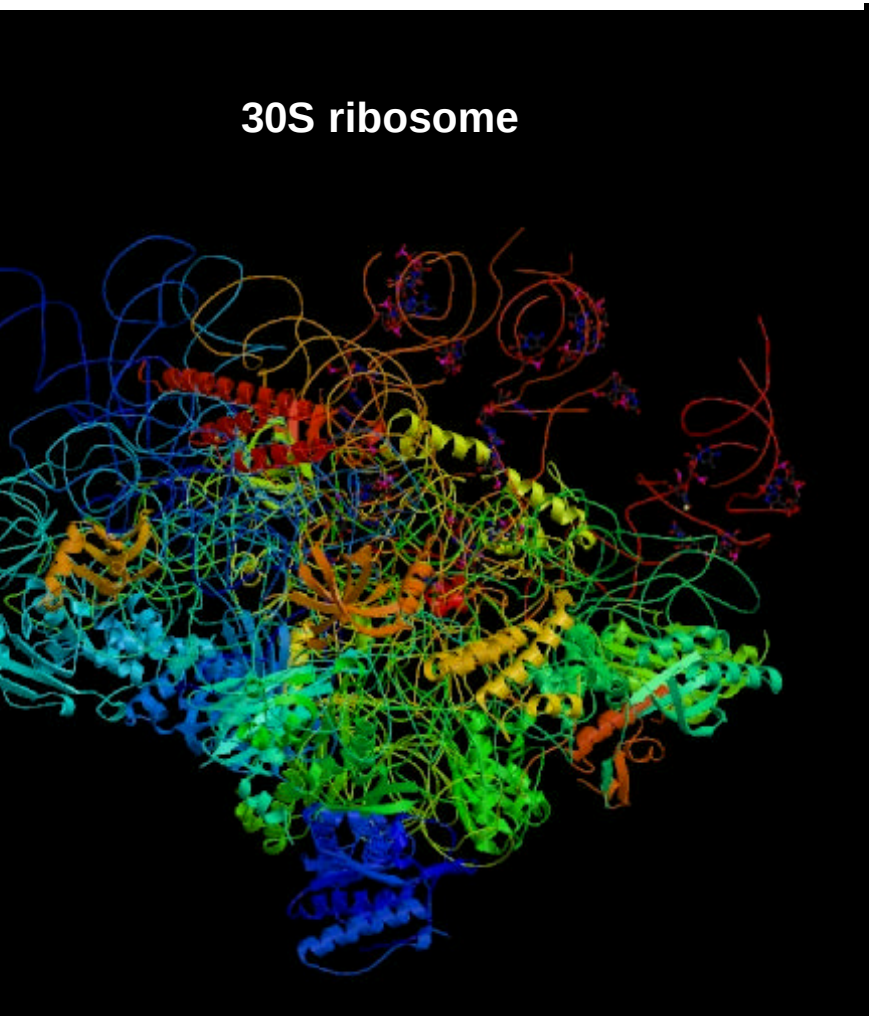


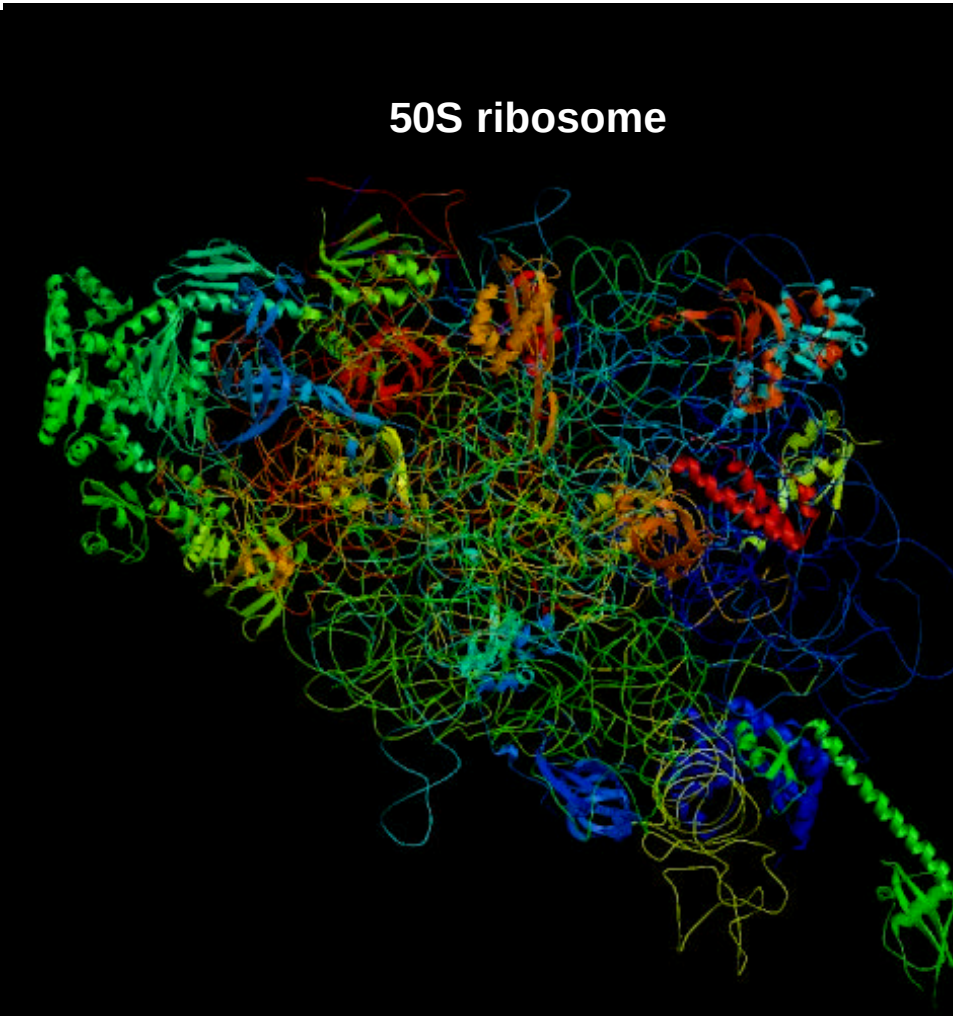
James Watson and Francis Crick with the first full model





John Kendrew with the first protein model (Myoglobin)



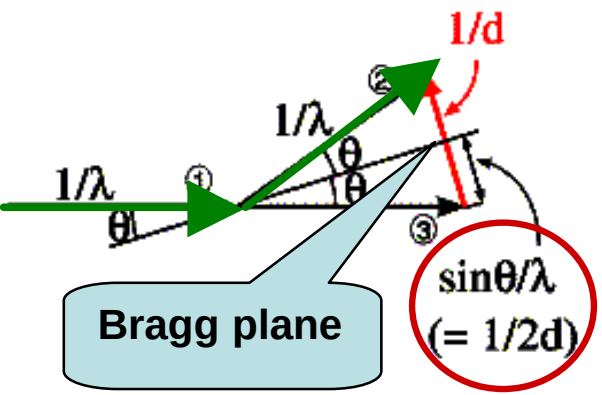


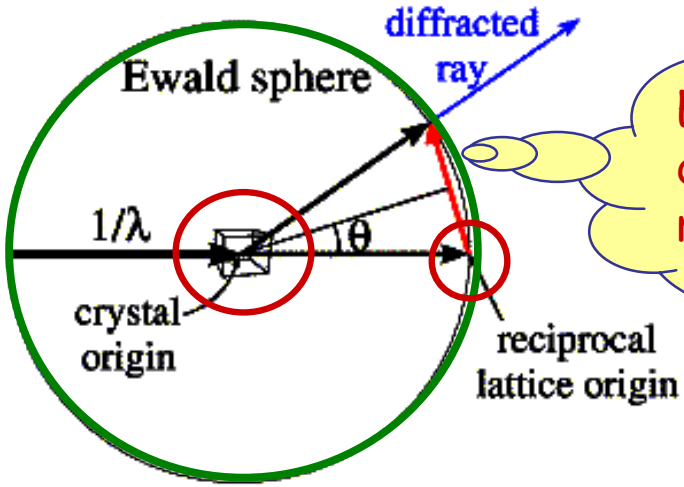
Protein data bank (PDB) (Research Consortium of Structural Biology, RCSB)
http://www.rcsb.org/pdb/

The Ewald sphere:

The sphere defined with radius = **1/l** is called the Ewald sphere.

1. All diffraction spots lies on the surface of the sphere.
2. Has two origins, crystal origin and reciprocal lattice origin.
3. Define the relationship between real space and reciprocal space.





Ewald sphere defines the reciprocal space