Structure factors again

- Remember 1D, structure factor for order h
  - $F_h = |F_h|\exp[i\alpha_h] = \int_0^1 \rho(x)\exp[2\pi ihx]dx$
  - Where x is fractional position along "unit cell" distance (repeating distance, origin arbitrary)
- As if scattering was coming from a single "center of gravity" scattering point
- Presence of "h" in equation means that structure factors of different orders have different phases
- Note that $\exp[2\pi ihx]dx$ looks (and behaves) like a frequency, but it's not (dx has to do with unit cell, and the sum gives the phase

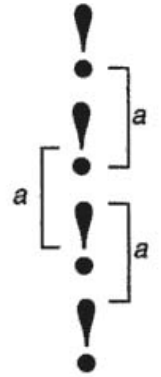
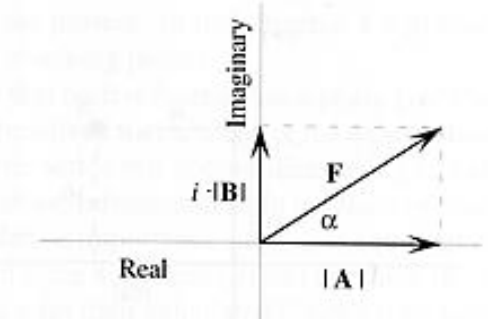


Fig. 4.4 Regular repeating array of scattering density. The density at any point is the same as the density at a distance *a* from it.

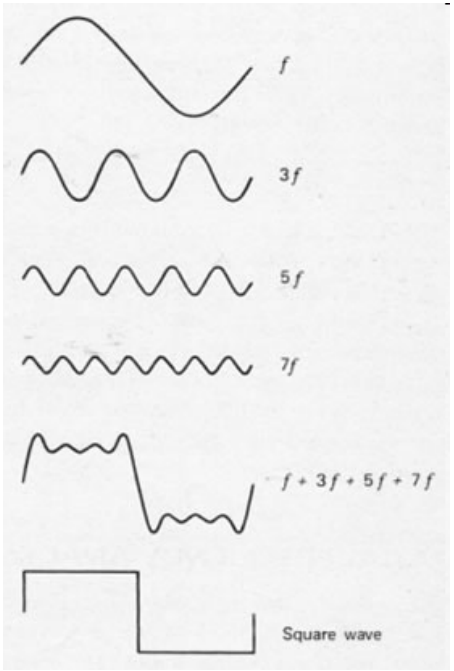


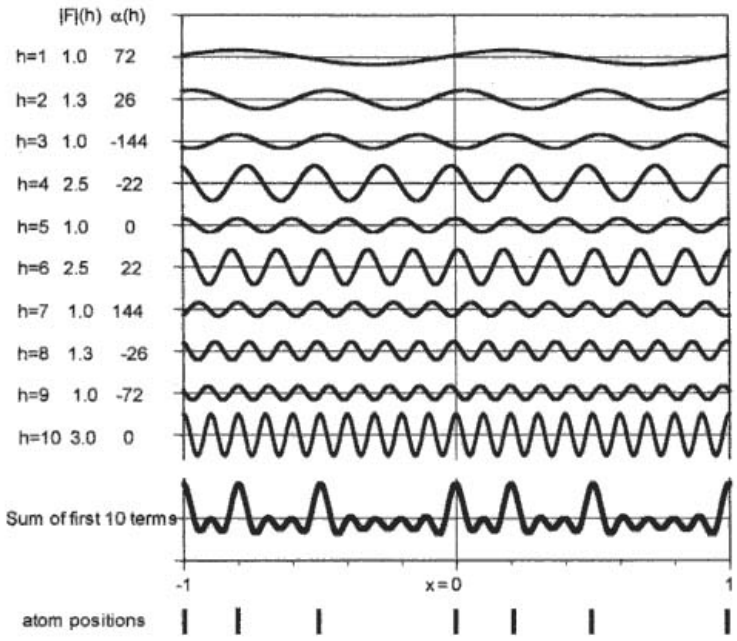
Back and Forth

- Fourier sez
  - For any function f(x), there is a “transform” of it which is
  - $F(h) = \int f(x)\exp(2\pi i(hx))dx$
  - Where h is reciprocal of x (1/x)
  - Structure factors look like that
- And it works backward
  - $f(x) = \int F(h)\exp(-2\pi i(hx))dh$
  - Or, if h comes only in discrete points
  - $f(x) = \Sigma F(h)\exp(-2\pi i(hx))$

Structure factors, cont'd

- Structure factors are a "Fourier transform" - a sum of components
- Fourier transforms are reversible
  - From summing distribution of $\rho(x)$, get hth order of diffraction
  - From summing hth orders of diffraction, get back $\rho(x) = \Sigma F_h \exp[-2\pi ihx]$





Two dimensional scattering

- In Frauenhofer diffraction (1D), we considered scattering from points, along the line
- In 2D diffraction, scattering would occur from lines.
- Numbering of the lines by where they cut the edges of a unit cell
- Atom density in various lines can differ

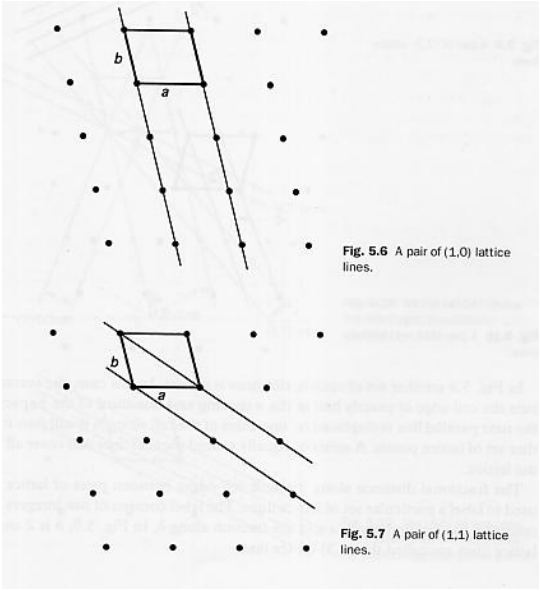


Fig. 5.6 A pair of (1,0) lattice lines.

Fig. 5.7 A pair of (1,1) lattice lines.

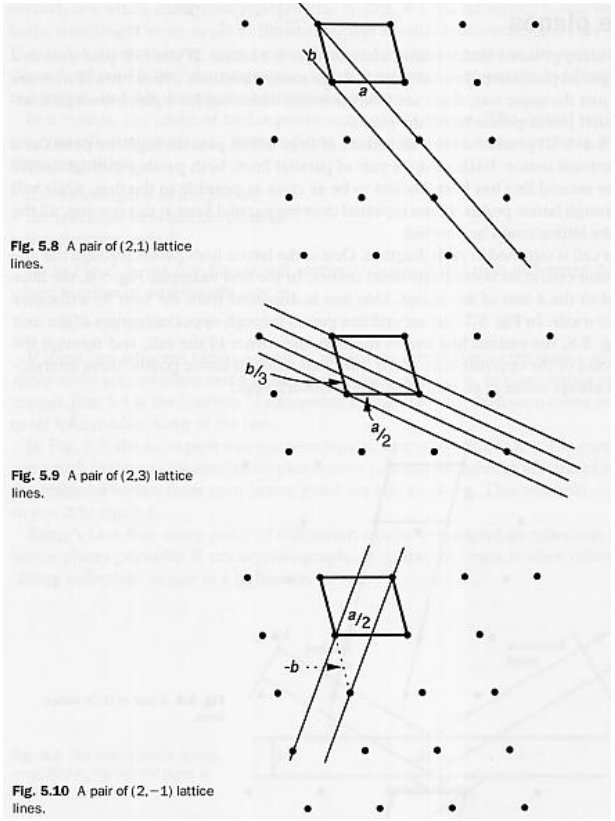


Fig. 5.8 A pair of (2,1) lattice lines.

Fig. 5.9 A pair of (2,3) lattice lines.

Fig. 5.10 A pair of (2,−1) lattice lines.

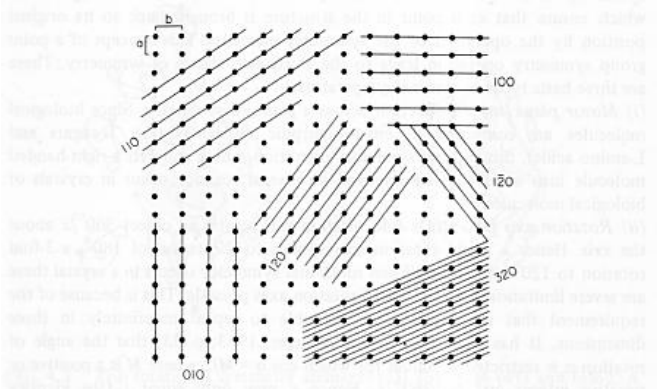


FIG. 4.2 A two dimensional lattice, illustrating the indices of certain sets of planes. (After Bunn, 1961.)

Extension to 3D

- Reflections now from planes
  - Planes defined by extension from 2D case
  - Unit cells differ
    - Depends on arrangement of materials in 3D lattice
    - = "Space Group"
    - Limited number of ways to pack irregular objects
  - Low order planes correspond to faces of macroscopic crystal
- Bragg's law
  - Suppose planes reflect X-rays like a mirror
  - In phase again implies $n\lambda = 2d\sin\theta$
  - Each plane will generate a reflection corresponding to order (h,k,l) of plane

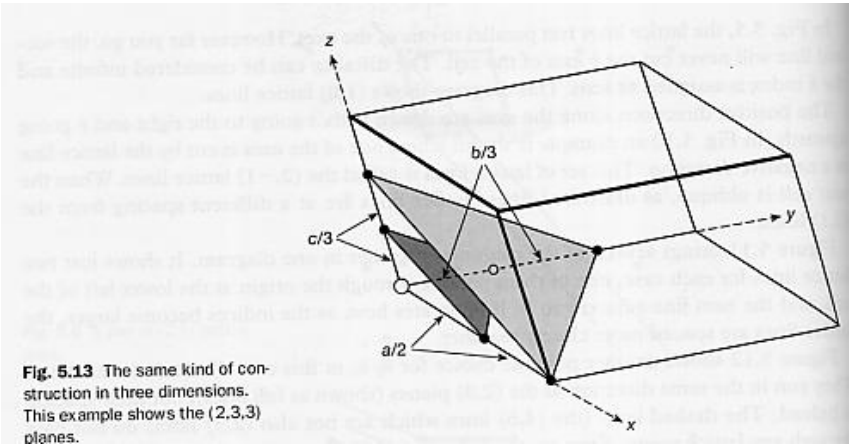
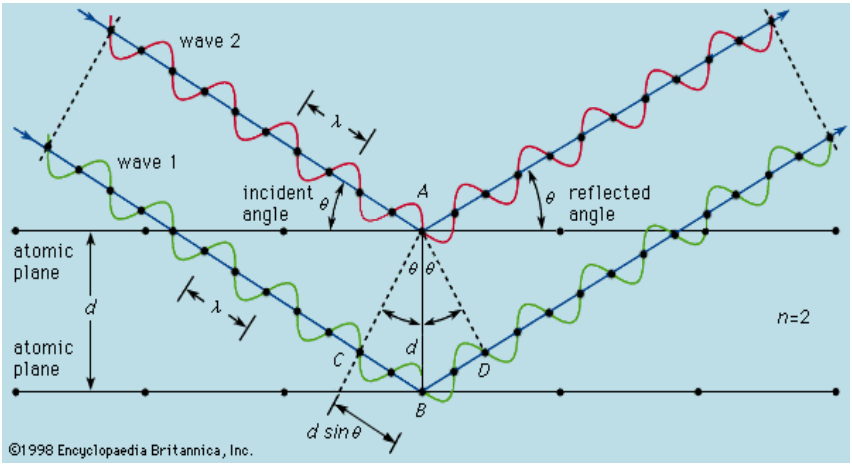


Fig. 5.13 The same kind of construction in three dimensions. This example shows the (2,3,3) planes.



Structure factors in 3D

- Remember 1D
  - $F_h = |F_h|\exp[i\alpha_h] = \int_0^1 \rho(x)\exp[2\pi ihx]dx$
- 3D similar
  - $F_{h,k,l} = |F_{h,k,l}|\exp[i\alpha_{hkl}]$
  - $= V\int_0^1\int_0^1\int_0^1 \rho(x,y,z)\exp[2\pi i(hx+ky+lz)]dx$
    - V (olume of unit cell) adjusts from electrons/volume in $\rho(x,y,z)$ to electrons per unit cell
    - Still calculatable from $\rho(x,y,z)$
- Reverse transformation
  - $\rho(x,y,z) = 1/V \sum\sum\sum F_{hkl} \exp[-2\pi i(hx+ky+lz)]$
  - (I.e, sum over all combinations of h, k, and l)
  - Still need phases (they're in F_{hkl})

Data collected - "reflections"

- Detected in a plane
  - Need a volume of reflections
  - Different reflections depending on orientation of unit cell (crystal) with respect to Xray beam
  - Actually collect many images like this one (10^3 to 10^6 total reflections"
- Measure intensity at each reflection

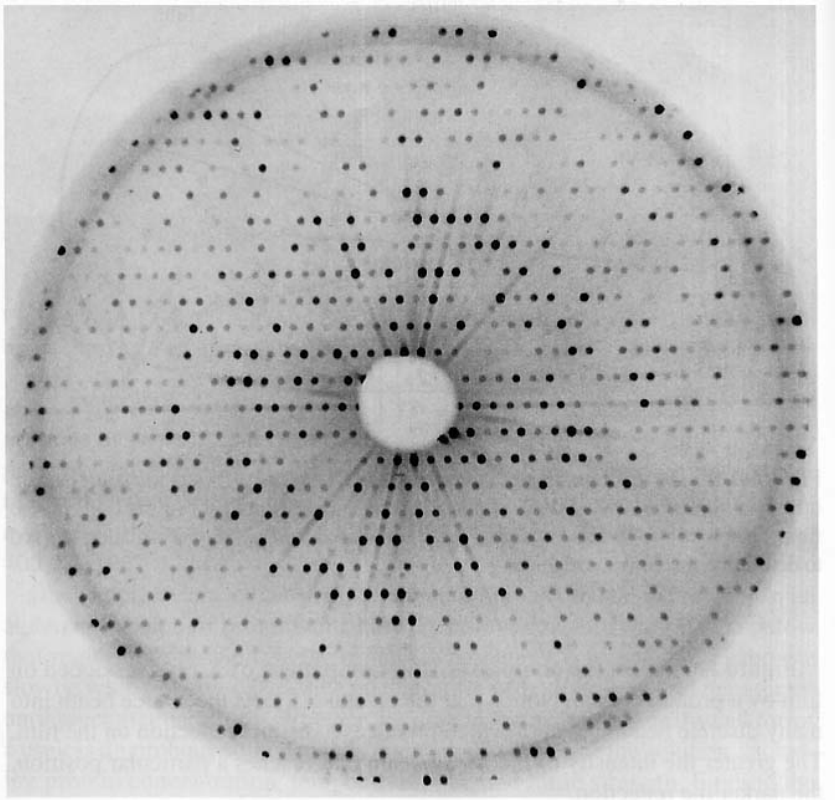
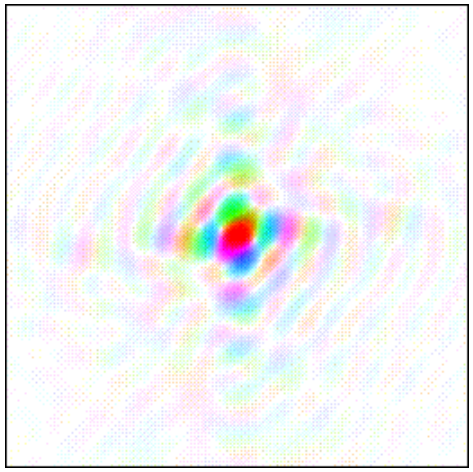
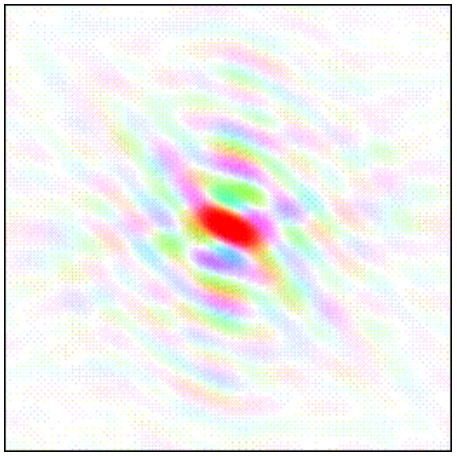
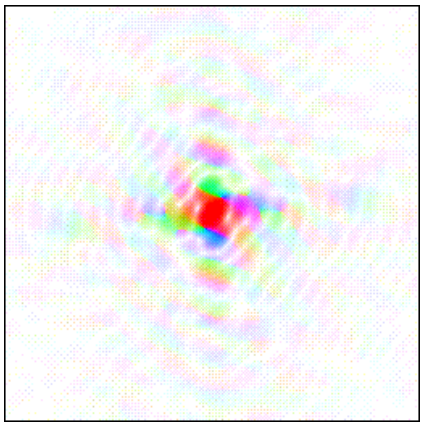


Figure 2.6 Diffraction pattern from a crystal of the MoFe (molybdenum–iron) protein of the enzyme nitrogenase from *Clostridium pasteurianum*. Notice that the re-

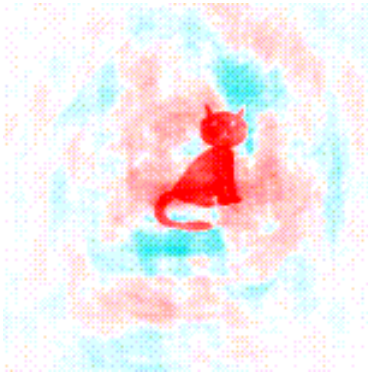
Phases and intensities







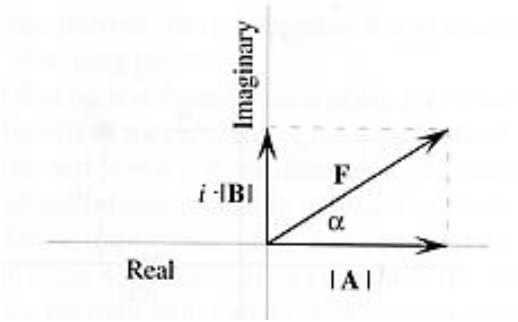
Duck intensities,
but cat phases

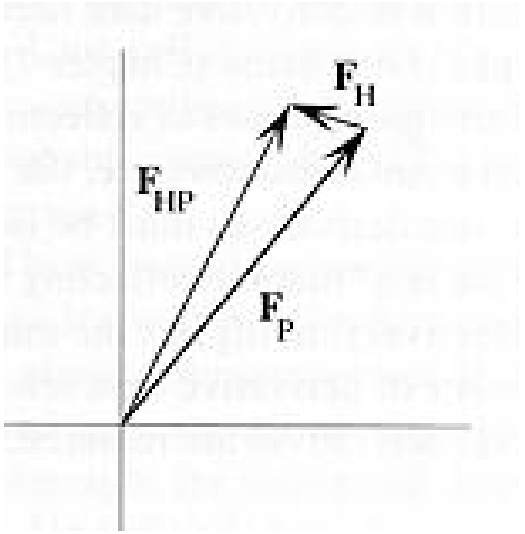


"In X-ray diffraction experiments, we collect only the diffraction magnitudes, and not the phases. Unfortunately the phases contain the bulk of the structural information. That is why crystallography is difficult." - Kevin Cowhan

Multiple isomorphous replacement

- Consider structure factors as complex vectors
  - In complex plane
  - Length from amplitude (proportional to $(I_{hkl})^{1/2}$
  - Angle α is phase
- Isomorphous heavy atom derivative of protein
  - Same reflections (same unit cell, so same planes)
  - Different structure factors
    - $\mathbf{F}_{HP} = \mathbf{F}_P + \mathbf{F}_H$





For example…

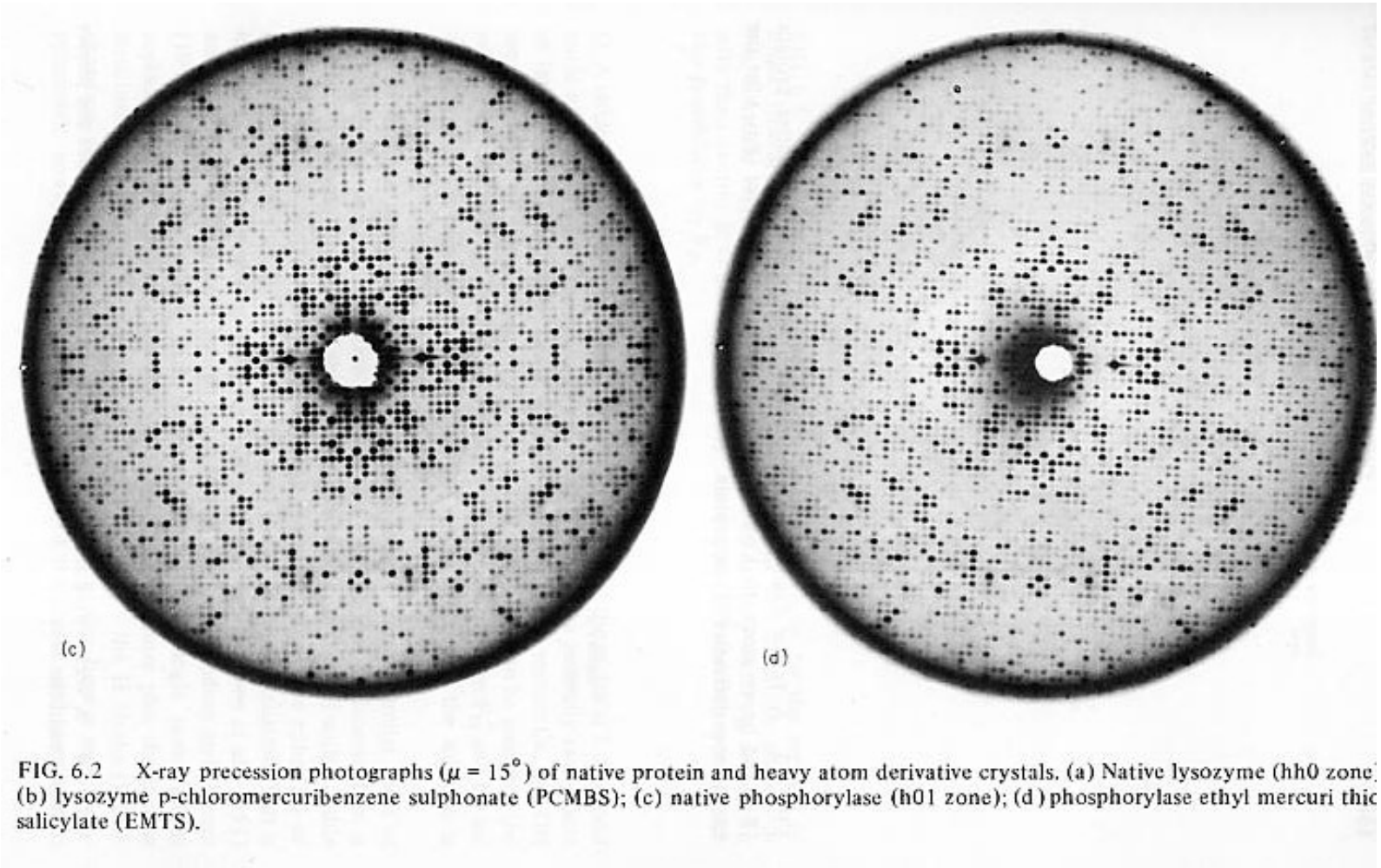


FIG. 6.2 X-ray precession photographs ($\mu = 15^{\circ}$) of native protein and heavy atom derivative crystals. (a) Native lysozyme (hh0 zone) (b) lysozyme p-chloromercuribenzene sulphonate (PCMBS); (c) native phosphorylase (h01 zone); (d) phosphorylase ethyl mercuri thio salicylate (EMTS).

A cleaner example

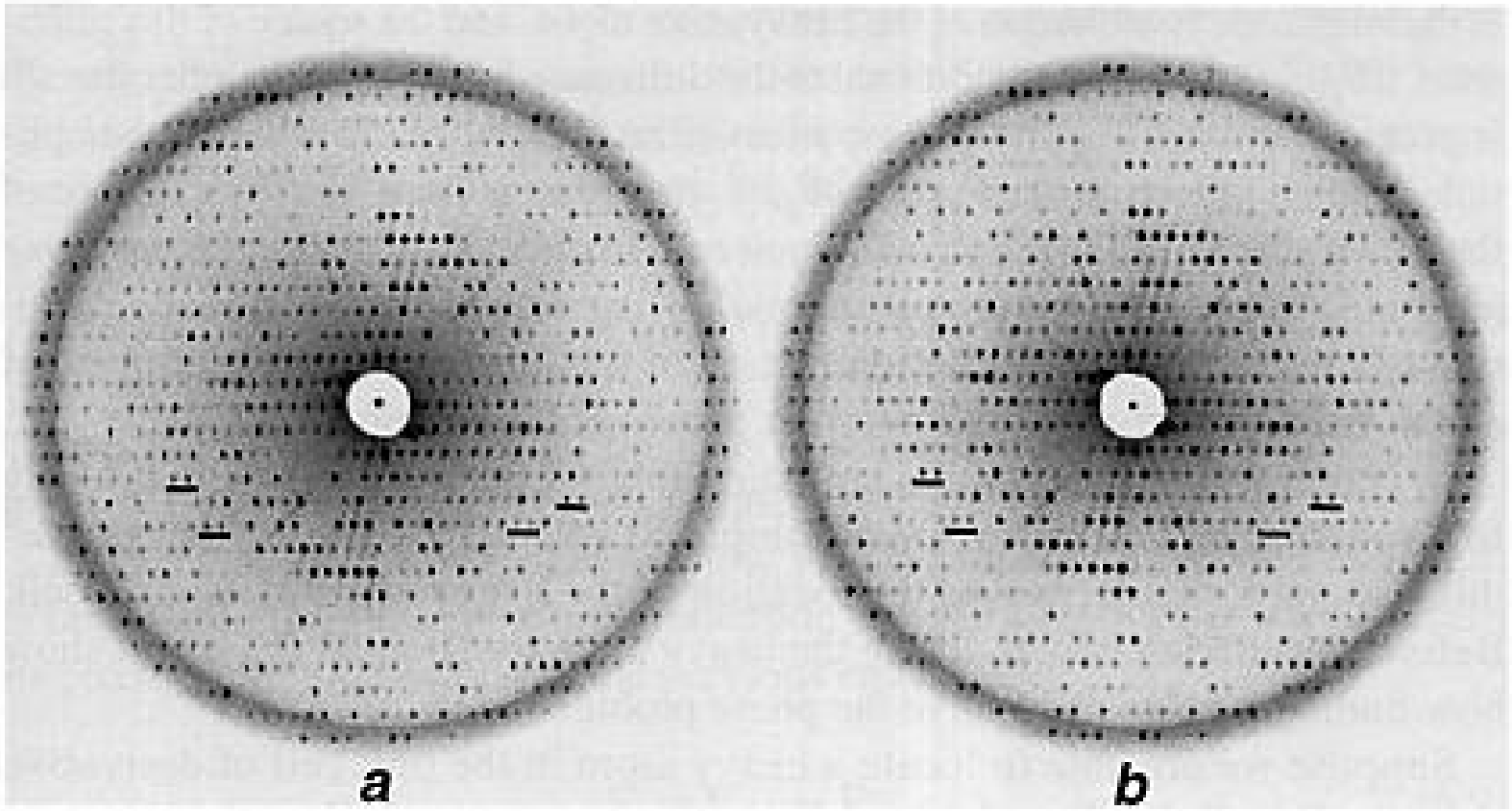
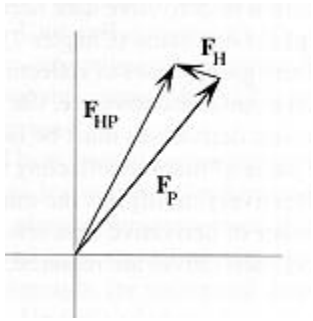


Figure 6.5 Precession photographs of the *hk*0 plane in native (*a*) and heavy-atom (*b*) crystals of the MoFe protein from nitrogenase. Corresponding underlined pairs in the native and heavy-atom patterns show reversed relative intensities. Photos courtesy of Professor Jeffrey Bolin.

What differences tell you

- Remember, for each reflection, structure factor can be represented as
  - $\mathbf{F}_{HP} = \mathbf{F}_P + \mathbf{F}_H$
  - Or $\mathbf{F}_P = \mathbf{F}_{HP} - \mathbf{F}_H$
- Don't know structure factor phases for $\mathbf{F}_{HP}$ or $\mathbf{F}_P$, but know intensities
- Suppose you knew $\mathbf{F}_H$
- You could reduce the possibilities to two



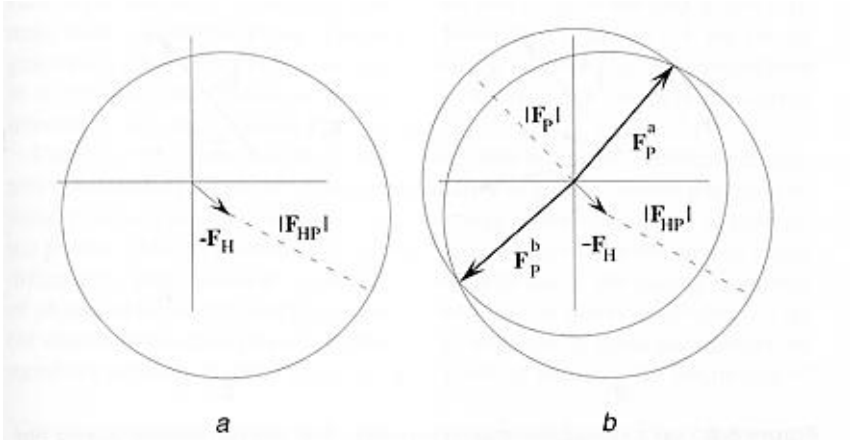


Figure 6.7 Vector solution of Equation (6.9). (*a*) All points on the circle equal the vector sum $|\mathbf{F}_{HP}| - \mathbf{F}_H$. (*b*) Vectors from the origin to intersections of the two circles are solutions to Eq. (6.9).

Choosing…

- Make a different derivative (at a different place in crystal)
- Repeat exercise for each reflection
  - Some better than others
  - Rarely as clean as image (more derivatives helpful)

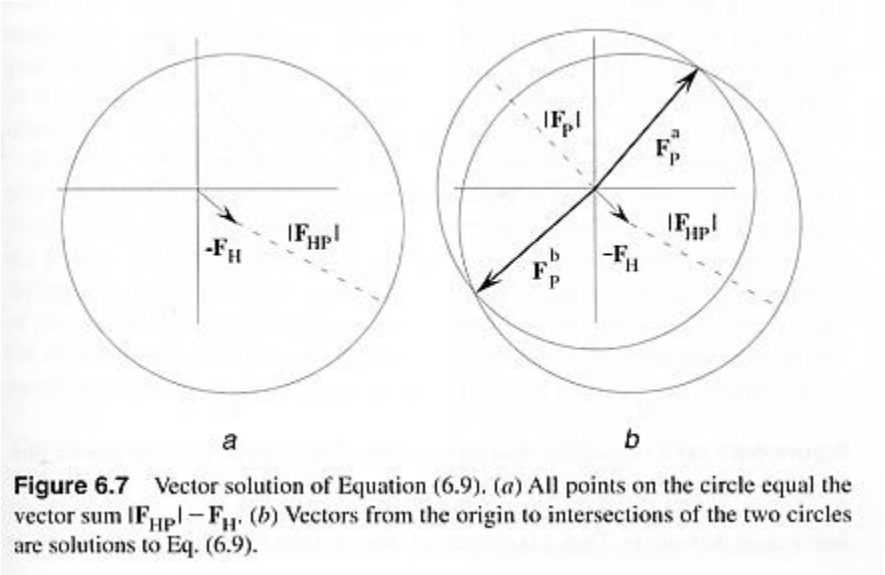


Figure 6.7 Vector solution of Equation (6.9). (*a*) All points on the circle equal the vector sum $|F_{HP}| - F_H$. (*b*) Vectors from the origin to intersections of the two circles are solutions to Eq. (6.9).

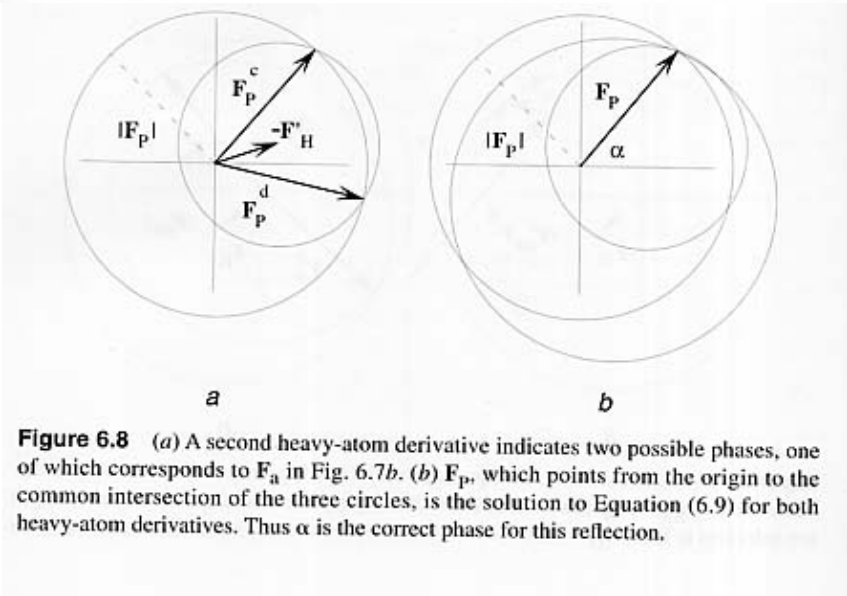


Figure 6.8 (*a*) A second heavy-atom derivative indicates two possible phases, one of which corresponds to F_a in Fig. 6.7*b*. (*b*) F_P, which points from the origin to the common intersection of the three circles, is the solution to Equation (6.9) for both heavy-atom derivatives. Thus α is the correct phase for this reflection.

How do you know $\mathbf{F}_H$?

- Can calculate if you know where heavy atoms are.
- Relatively easy, because only a limited number in each unit cell
  - If only one, easiest of all
  - If more than one, there's a way to get location from just intensities.
  - Patterson function

Patterson functions

- Convolution of structure factor and its inverse
  - Makes phases disappear
  - Result calculable from intensities of reflections alone
  - Corresponds to collection of vectors between atoms
  - Can figure out from that info alone what arrangement was, if number of atoms was small
- Calculate Patterson function of difference between native protein and derivative = Patterson function of heavy atoms only.

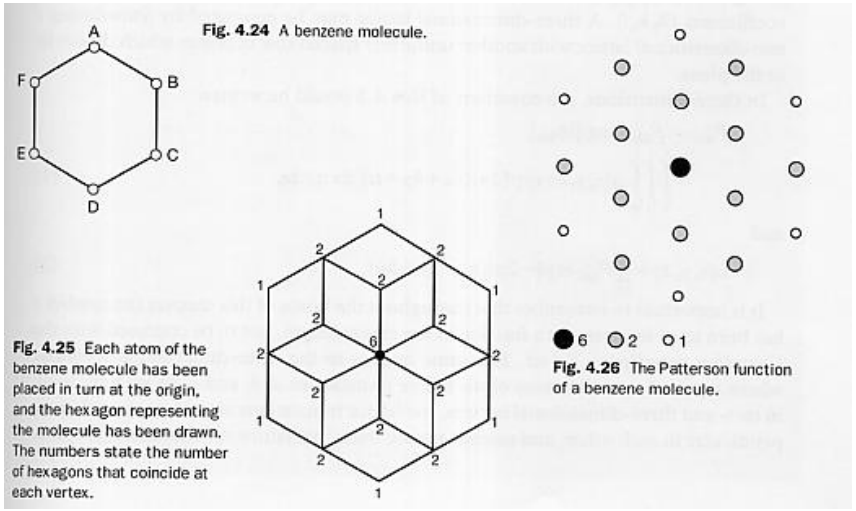


Fig. 4.24 A benzene molecule.

Fig. 4.25 Each atom of the benzene molecule has been placed in turn at the origin, and the hexagon representing the molecule has been drawn. The numbers state the number of hexagons that coincide at each vertex.

Fig. 4.26 The Patterson function of a benzene molecule.

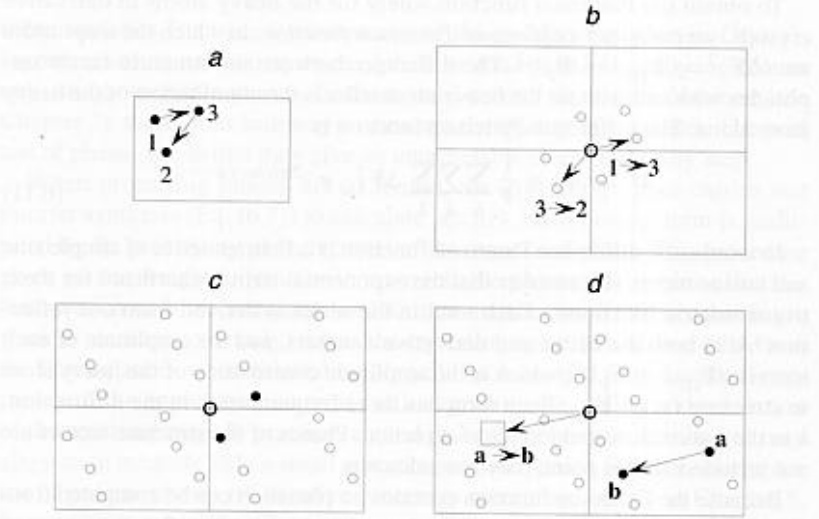


Figure 6.10 Construction and interpretation of a Patterson map. (*a*) Structure of unit cell containing three atoms. Two of the six interatomic vectors are shown. (*b*) Patterson map is constructed by moving all interatomic vectors to the origin. Patterson "atoms" (peaks in the contour map) occur at the head of each vector. (*c*) Complete Patterson map, containing all peaks from (*b*) in all unit cells. Peak at origin results from self-vectors. Image of original structure is present (origin and two darkened peaks) amid other peaks. (*d*) Trial solution of map (*c*). If origin and Patterson atoms **a** and **b** were the image of the real unit cell, the interatomic vector **a** → **b** would produce a peak in the small box. Absence of the peak disproves this trial solution.

Next steps

- Patterson function results (or other method such as anomalous scattering)
  - Give location of heavy atoms in unit cell
  - Calculate structure factors for heavy atoms (F_h)
  - Then calculate phases for each reflection
  - Then combine observed intensities I_{hkl} with estimated phases to calculate electron density $\rho(x,y,z)$
  - (Contour lines to represent density distribution)
- Resulting crude map must be refined.

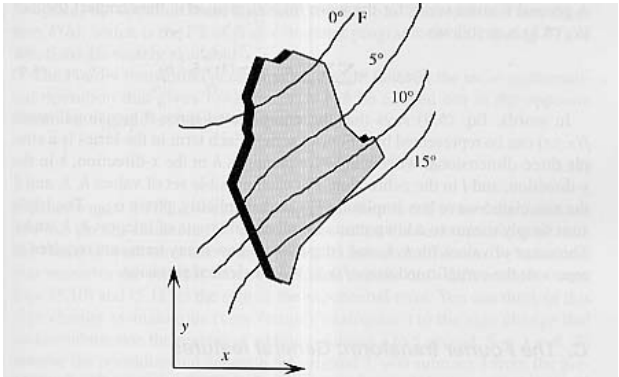
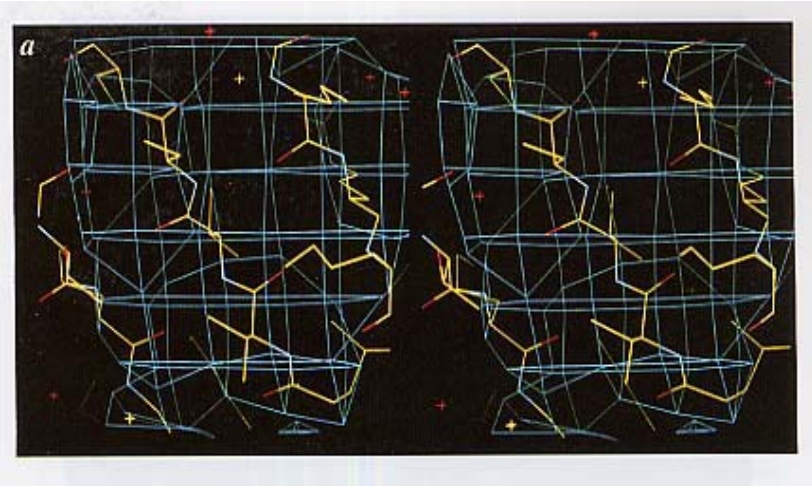


Figure 5.1 Seasonable February morning in Maine. Lines of constant temperature (isotherms) allow plotting a two-dimensional function without using the third dimension. This is a contour map of $t(x,y)$, giving the temperature t at all locations (x,y). Along each contour line lie all points having the same temperature. A planar contour map of a function of two variables takes the form of contour lines on the plane. In contrast, a contour map of a function of three variables takes the form of contour surfaces in three dimensions (see Plate 2).



Basic refinements

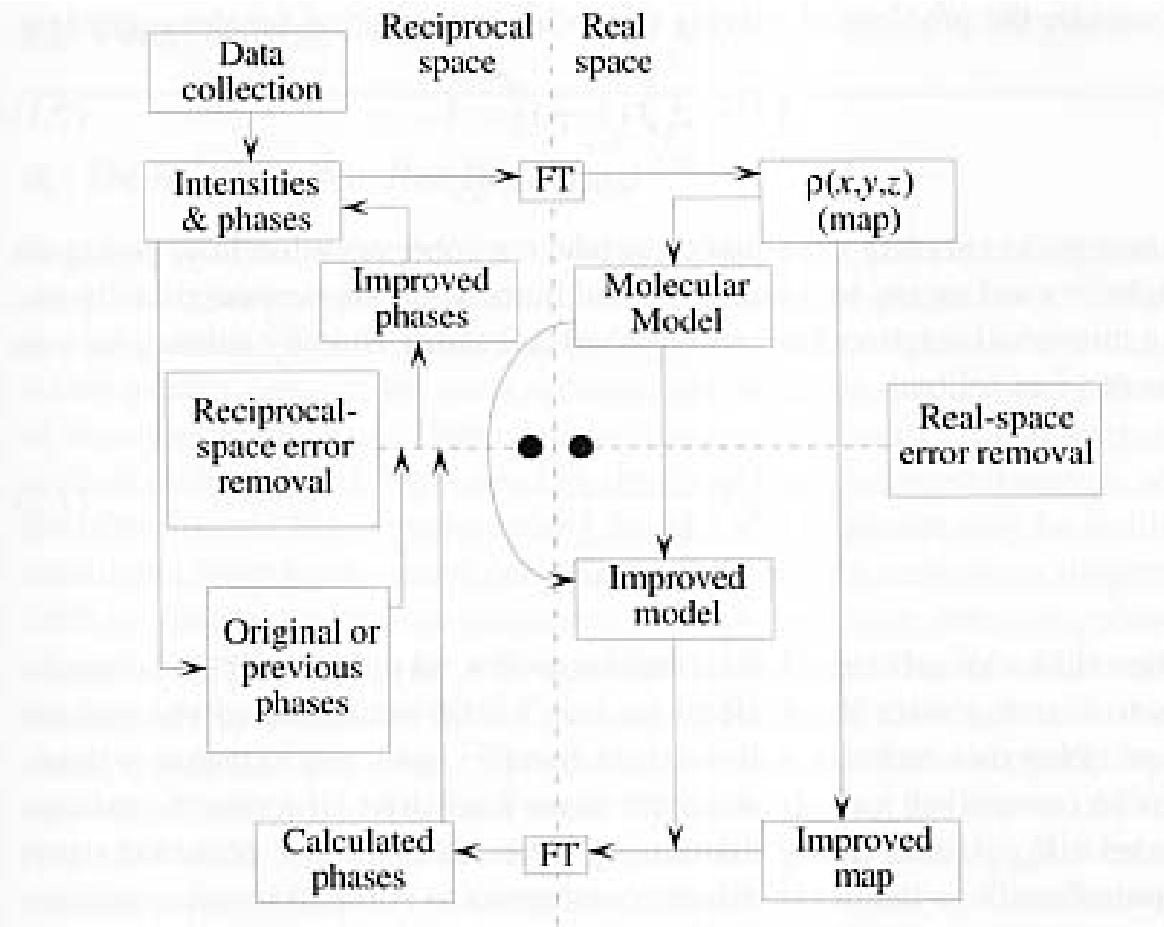


Figure 7.1 Block diagram of crystallographic structure determination.

Solvent flattening

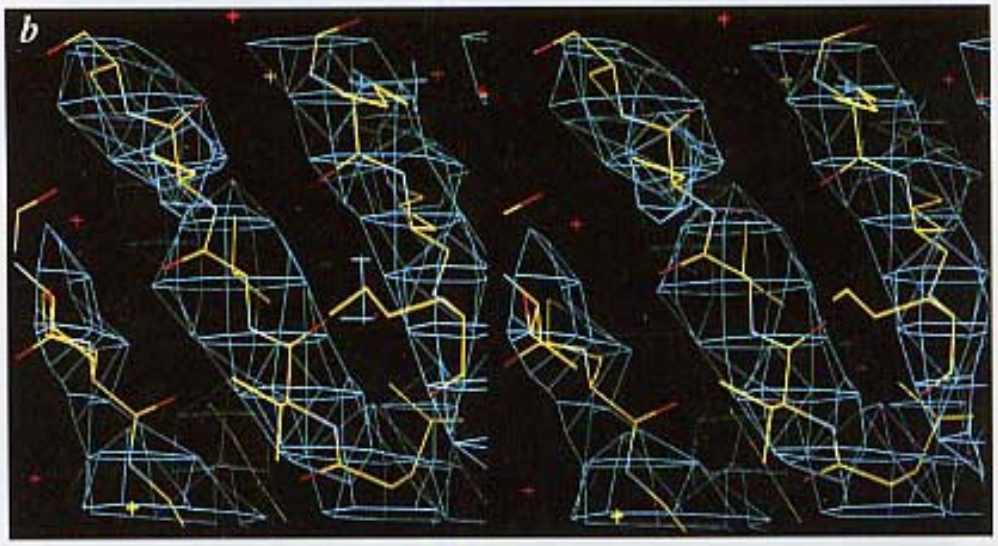
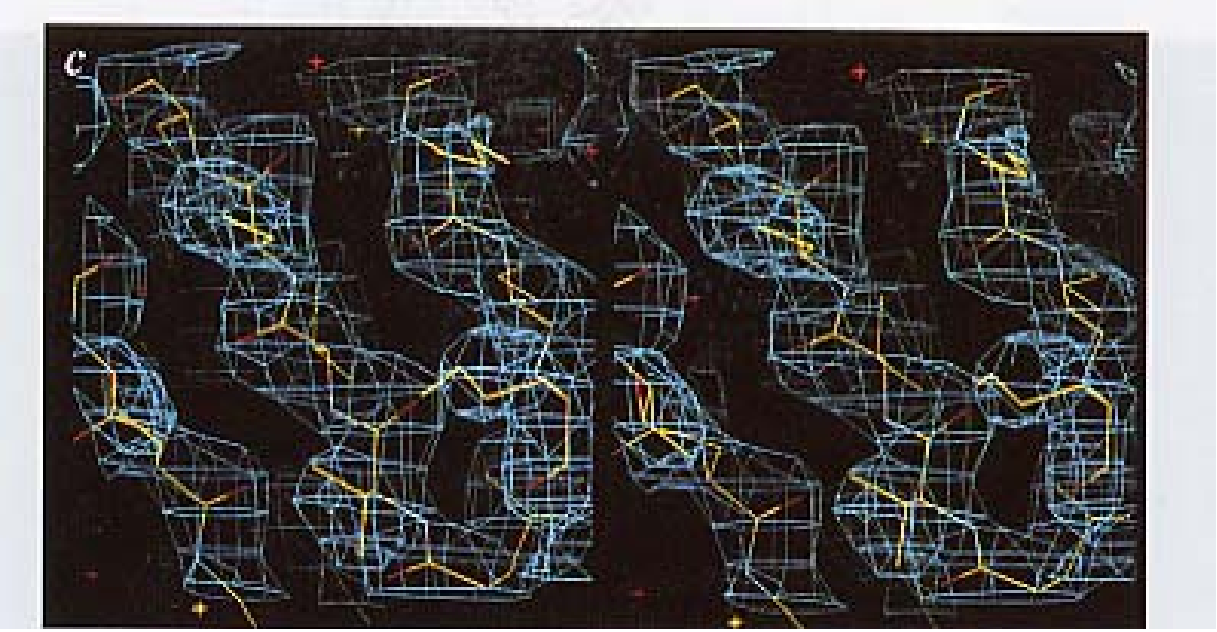


Plate 12 Electron-density maps at increasing resolution. Maps were calculated using final phases, and Fourier series were truncated at the resolution limits indicated: (*a*) 6.0 Å; (*b*) 4.5 Å; (*c*) 3.0 Å; (*d*) 1.6 A. (For discussion, see Chapter 7.) (*Continues*)

- Refining the crude map
  - Eliminate electron densities on physical grounds (negative or isolated densities)
  - Separate densities into positive (protein) and zero (solvent)
  - Reset ratio of positive to low density to match amount of protein in unit cell/crystal
  - Recalculate structure factors, yielding intensities and phases for all reflections
  - Combine phases with observed intensities to recalculate new, better model
  - Aiming for phases at progressively higher resolution (increased scatter angle)

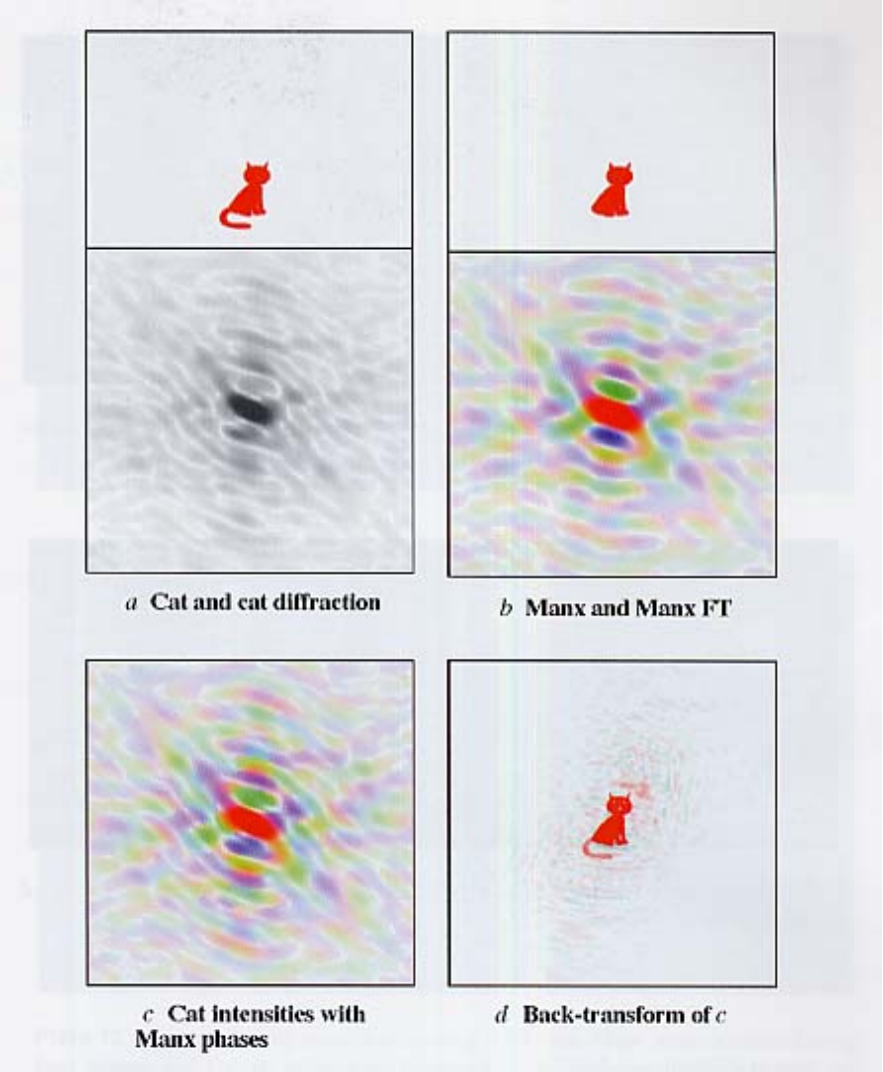
Molecular models

- Introduce atoms into the model from known arrangements in α helices and β-sheets
  - Polyalanine (because it contains a β carbon attached to an α carbon in the backbone)
  - Reasonable fits in protein interior, where little room for movement
  - Calculate new phases from models and combine with observed intensities to get more refined models
- Identify sequence
  - Large bulky hydrophobics as bulges in electron densithy



Atomic positions from relatives

- Known relatives from sequence homology
- Assume common structure
- Take phases from homolog
- Can apply early (to get initial phases)
  - Easy if structure from isomorphous crystal (often so if new protein is previously determined one with some change imposed)
  - Harder if nonisomorphous
    - Must place model in unit cell
    - Done with gigantic Patterson functions
- Intensities from real protein can then show the way to differences



How are you doing?

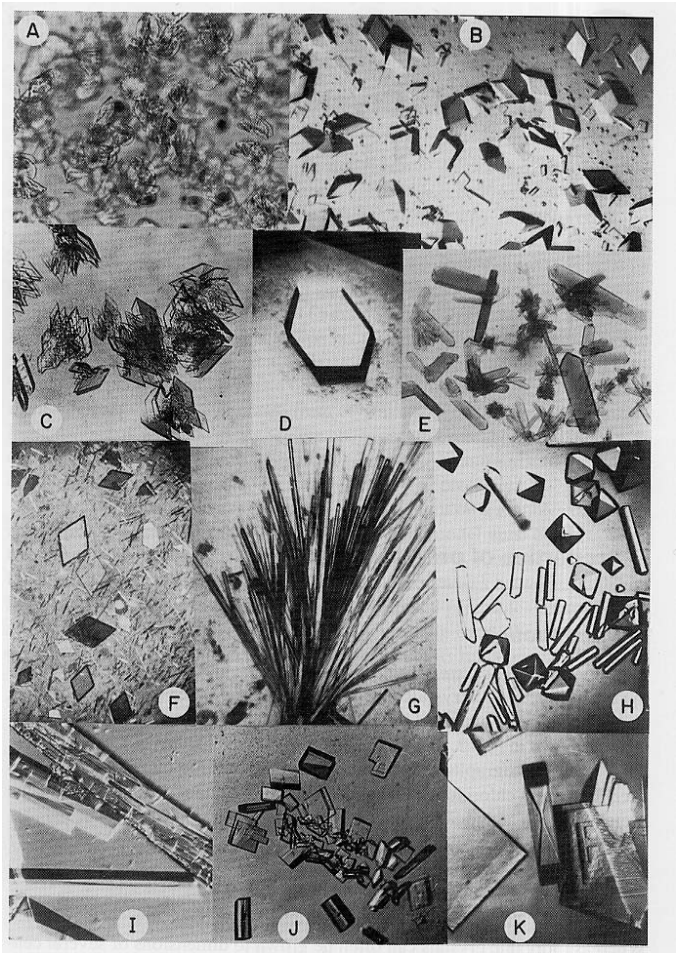
- At every stage, you get computed intensities ($|F_{calc}|$), which can be compared with observed intensities ($|F_{obs}|$)
  - Minimize these differences (least squares calculations)
  - To minimize differences, can allow for atoms to ocscillate within a larger space ("temperature factors" "B factors" "B_j")
  - Can allow certain atoms to be present in a particular position only part of the time ("occupancy")
- Residuals
  - Calculate difference between observed and actual intensities
  - $R = \sum (|F_{obs}| - |F_{calc}|) / \sum |F_{obs}|$
  - Even better, set aside some reflections that aren't used in the model building, and calculate R for just those reflections (R_{free})
  - Good values < 0.2

…cont'd

- Model should match sequence
- Model should become chemically reasonable
  - Bond lengths and bond angles should be normal
- Model should be conformationally reasonable
  - No hydrogen clashes
  - Planar peptide bonds
  - Acceptable Ramachandran angles
  - Reasonable side chain confomations

Crystals

- Protein crystals …
- …are soft (high water content)
- …fragile (few molecule/molecule interactions)
- …small
- …imperfect assemblies of microcrystals



Crystal growing

- Precipitation by salt, precipitating agents
- …from “mother liquor”
- Hanging drop methods
- Seeds of small, perfect crystals
- Add materials (ligands, heavy metals) in mother liquor to already formed crystals

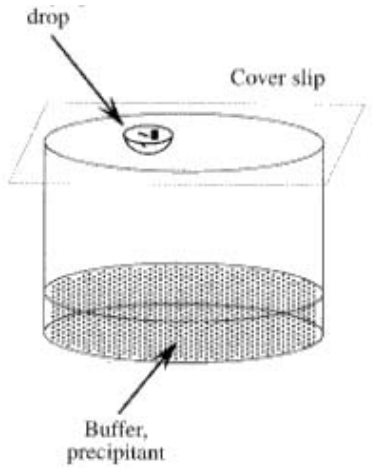


Figure 3.2 Growing crystals by the hanging-drop method. The droplet hanging under the cover slip contains buffer, precipitant, protein, and, if all goes well, growing protein crystals.

Holding crystals up to the light

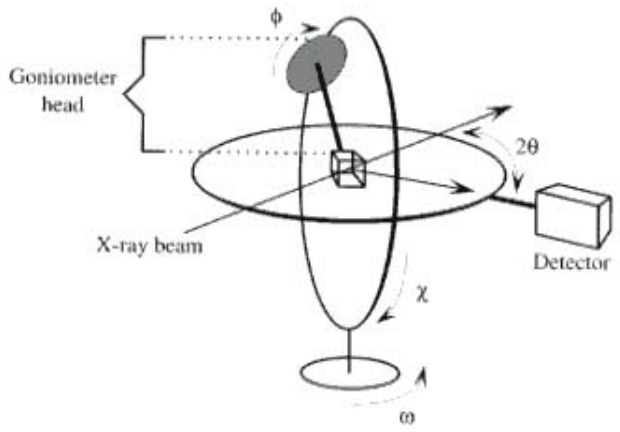


Figure 4.21 System of circles in diffractometry. The crystal in the center is mounted on a goniometer head.

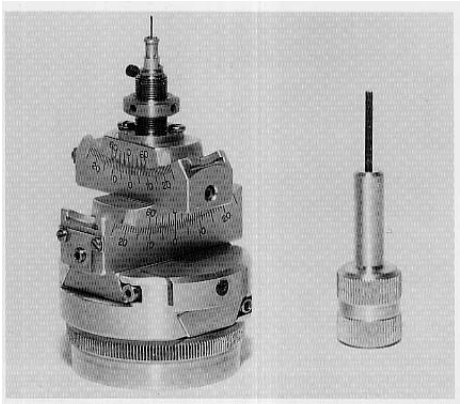


Figure 4.20 Goniometer head, with capillary tube holder at top. The tool (right) is an Allen wrench for adjusting arcs and sledges. Photo courtesy of Charles Supper Company.

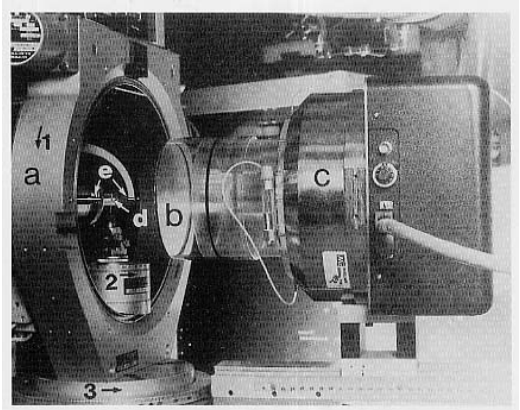


Figure 4.22 Diffractometer and area detector. Photo courtesy of Professor Leonard J. Banaszak.

Light sources

- In lab sources
  - Improved
  - Rarely used now
- Synchrotron

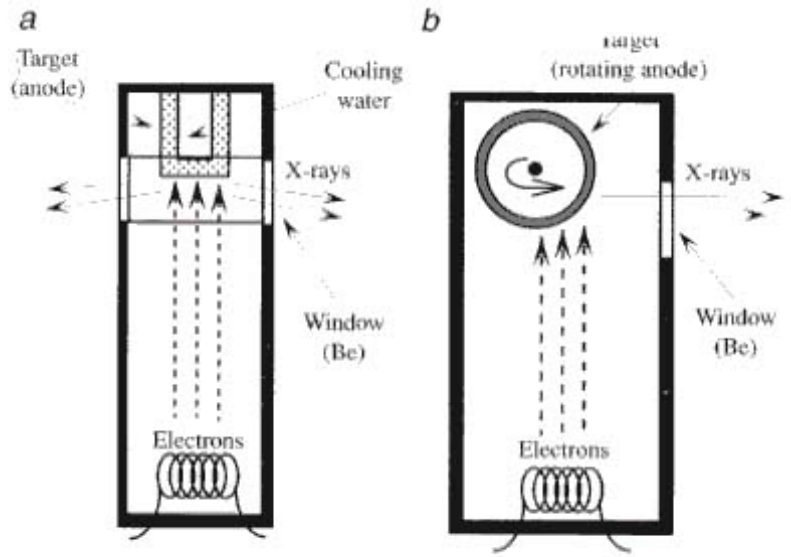


Figure 4.17 (*a*) X-ray tube. (*b*) Rotating anode tube.

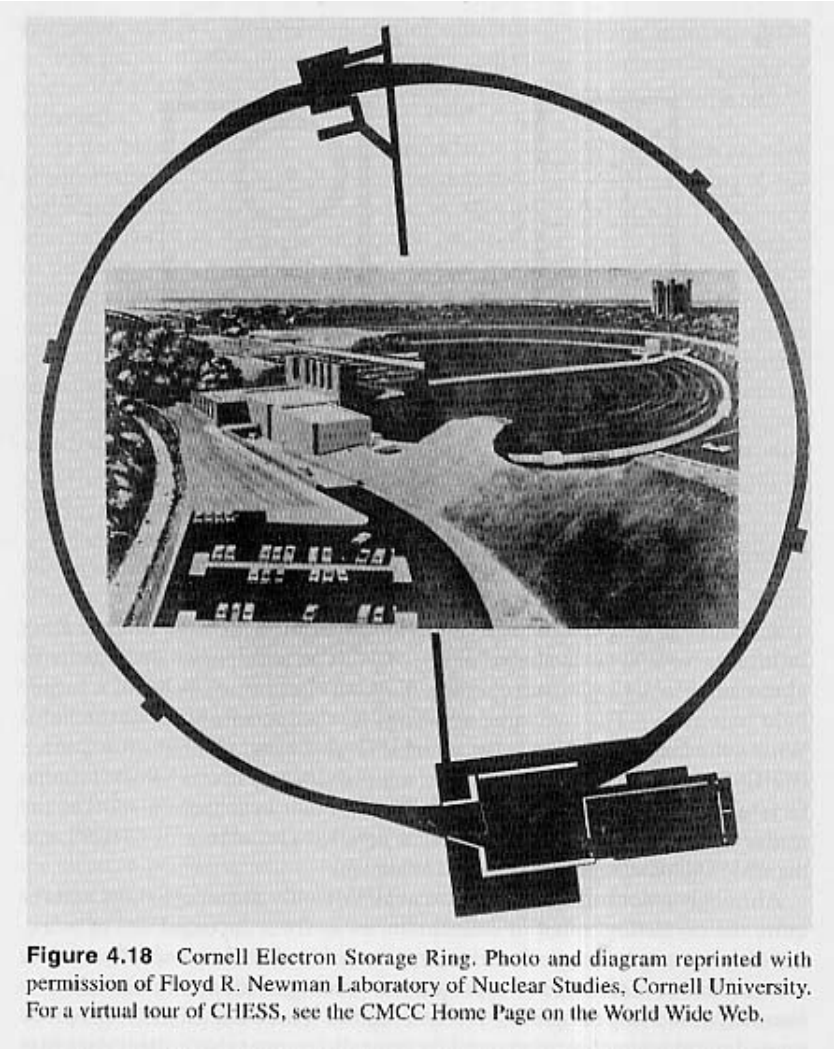


Figure 4.18 Cornell Electron Storage Ring. Photo and diagram reprinted with permission of Floyd R. Newman Laboratory of Nuclear Studies, Cornell University. For a virtual tour of CHESS, see the CMCC Home Page on the World Wide Web.