MACROMOLECULAR CRYSTALLOGRAPHY

lecture notes:
David C. Richardson

aka
Arendall & Richardson

BCH 291
Physical Biochemistry
Duke University

http://kinemage.biochem.duke.edu/teaching/bch291

Macromolecular Crystallography

Overview

Notes are divided into 10 Chapters reflecting the usual size of this section of the Duke University Physical Biochemistry course. (BCH291, SBB291)

Kinemages
Problem Sets

Web Site: http://kinemage.biochem.duke.edu/teaching/bch291

Other Web sites:
http://molprobity.biochem.duke.edu/ protein & RNA analysis, evaluation,...
http://www.rcsb.org/ Protein Data Bank get coordinates here
http://eds.bmc.uu.se/eds/ Electron Density Server get maps here
http://xray.bmc.uu.se/gerard/embo2001/modval/index.html model validation tutorial
http://www.ysbl.york.ac.uk/~cowtan/ Insightful crystallography exercises

Suggested Textbooks for reference or different presentation:

Alexander McPherson "Introduction to Macromolecular Crystallography", 2nd ed., somewhat different organization than these notes, more information about crystallization and the initial stages of structure determination. This is the nearest thing to a textbook I've found to accompany my notes and lectures.

Stout and Jensen, 2nd ed. "X-ray Structure Determination" A very good general crystallography text.

Gale Rhodes 3rd ed. "Crystallography Made Crystal Clear" Informal, less detail but more context.

David Blow "Outline of Crystallography for Biologists" More detail than Rhodes, less math than Drenth, still not quite enough detail in deriving equations and developing the constructions needed by a practicing crystallographer. However more complete description of the scope of information available from macromolecular crystallography.

Jan Drenth "Principles of Protein X-ray Crystallography" More "mathematical", vector and matrix notation...

McPherson for review when you start to actually do crystallography,
Rhodes for overall concepts,
Blow for appreciating results,
Stout & Jensen or Drenth to understand what you are actually doing in crystallography..

2009:

This year we have 12 lecture-periods scheduled: (Overall exam at semester end includes the material of this section). Besides allowance for slipage, this gives room for extra emphasis on relating quality of results with quality of data and processing.

Lecture notes continue to evolve: this year continuing the process of reducing font size and rearranging page layout so that written, class-time note-taking could be better done directly on these pages.

Each page is marked with a topic and a chapter-page-number. The sections designations are numbered to match the "Chapter" numbers. Version control date added.

Note are distributed using three-hole punched paper so they can be accummulated, inserted, etc.

Specific items listed under each "Chapter" might change or be rearranged as we go through them these next 4 weeks.

HOMEWORK is due at the next lecture time.
Late Homework accepted only when the homework problem is discussed with me.
Improved homework grade possible by discussing actual homework problem with me..

This course will not teach you to use current machinery and computer programs to solve crystal structures; it will, however, develop the concepts and equations you would need to build your own diffractometers and write your own computer programs to solve crystal structures. The approach is geometrical rather than algebraic, but the equations derived are complete (but not necessarily in the form for an efficient computer program).

For those who will be producers of crystallographically determined structures, the complete do-it-yourself approach is not, in general, advised, but in the rapidly developing world of crystallographic machinery and programs, it is valuable to understand the fundamental strengths and limitations of the experiments.

Everyone these days is a consumer of crystallographically determined structures. Evaluating the reliability of structural details is critical to using them. Knowing what goes into producing structural models gives an appreciation of the strengths and limitations of those models.

Apology:

These are the notes from which I lecture.

These notes have accreted over the years, in fits and starts as Jane and I worked over how to present the concepts of crystallography.

They started as just annotated drawings, and were handed out so that I didn't have to make my chalk-board drawings quite as accurate, nor did students have to spend all the lecture time trying to reproduce the board drawings. So these notes are indeed quite sketchy and lack much of the words of my actual lectures which I rearrange and make up as I talk each time. (Which is why these notes were not posted on the kinemage/teaching/ web site!)

When these notes were started and for many years the macromolecules that were crystallized were proteins. Thus much of the terminology (like that in many crystallography textbooks) is Protein centric, e.g. F_P for native structure factor and F_{PH} for native-with-Heavy-atom derivative.

Bryan Arendall has imported and improved the original hand drawings into electronic form, as well as added commentary and some reorganization to the notes. If ever these notes become a textbook, then it would be "Arendall and Richardson". Until that time, I own all the mistakes. -Dave Richardson

Chapter 3, sec 1: Light Scattering

Light: different representations
The Electromagnetic Spectrum wavelength range of x-rays
Light as a wave: Interaction of light and matter
Combination of Light waves (visualize, add point-by-point)
Vector notation for waves (sum phase vectors)
Phase shift during interaction
(damped driven simple harmonic motion: ball on string)
Light scattered by an individual oscillator

Combining waves: "interference" constructive (destructive)

adding-waves.2.kin shown in KiNG

Problem Sets "Adding Waves" point-by-point by hand for once

Chapter 3, sec 2: Bragg Diffraction

Bragg's Law and The Diffraction Vector

How crystals affect the scattering of x-rays
Sharpness of diffraction maxima
Diffraction from a 2D grid of atoms (cones)
Diffraction from a 3D array (lines = rays)
Scattering and Diffraction
Diffraction as a wave-front phenomenom
The appearance of the Diffraction Pattern
Diffraction as a probe for spacings within a crystal

Problem Sets "Bragg's Law limits"
"3 planes --> 3 spots"

Chapter 4, sec 1, Diffraction Vector

The Diffraction Vector

Chapter 4, sec 2, Reciprocal Lattice

The Reciprocal Lattice and diffraction

Ewald Sphere: Crystal and Reciprocal Lattice

Chapter 4, sec 3, Crystal diffraction

Physical properties of crystal diffraction

Crystal of molecules: side-by-side alpha helices
Bragg Planes in 3D
Reciprocal Lattice vs Bragg Planes
Geometrical factors that affect measured intensities
Structure factor of an individual atom
Size of the spot
Mosaic Blocks - perfectly-imperfect crystals

Kinemage HelixBearHair.kin
Problem Sets "Reciprocal Lattice & Diffraction from a Protein Crystal {Ewald}"

Chapter 9: Electron Density Maps, Resolution, and Refinement

Kinemages 1HJ8_1.0A.kin, 1C9P_2.8A.kin
Trypsin structures at 2 different resolutions
Maps 1HJ8, 1C9P 2Fo-Fc from EDS
Problem Sets "Resolution and model-to-map fitting"
"KiNG on MolProbity site: resolution in trypsin models and maps"

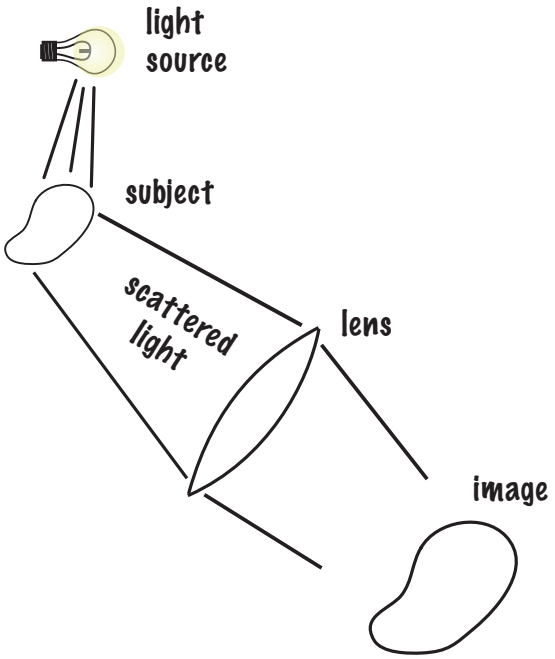
Chapter 10: Quality Evaluation and Validation

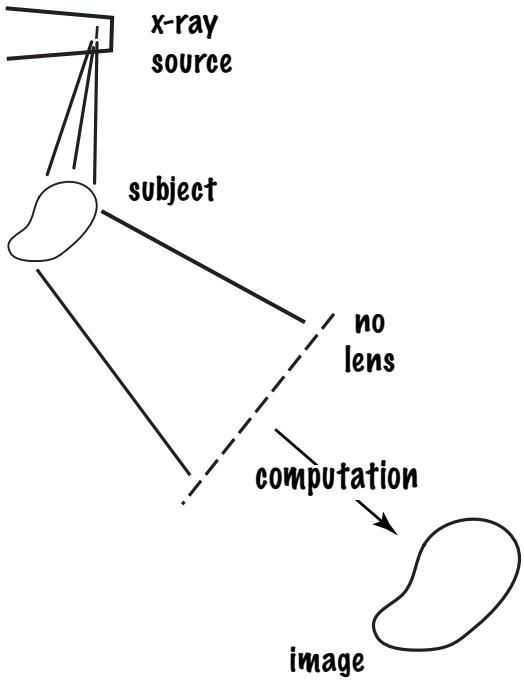
Kinemages HowDotsWork.kin
1JIRon1S83_Arg66_supr.kin, 1JIR.pdb in MolProbity
Problem Sets "Model quality and validation: MolProbity and KiNG"

The Technique of X-ray Crystallography

The goal in macromolecular crystallography is to obtain a three-dimensional picture of a molecule. The general method, and the source of some of the main difficulties involved in applying it, can be described by an analogy with an ordinary light microscope.

The light microscope uses a lens to recombine the light scattered by the subject into an image. But it is necessary to use light of a wavelength not much greater than the size of the features we want to see in the images. for molecular structures this means using x-rays.

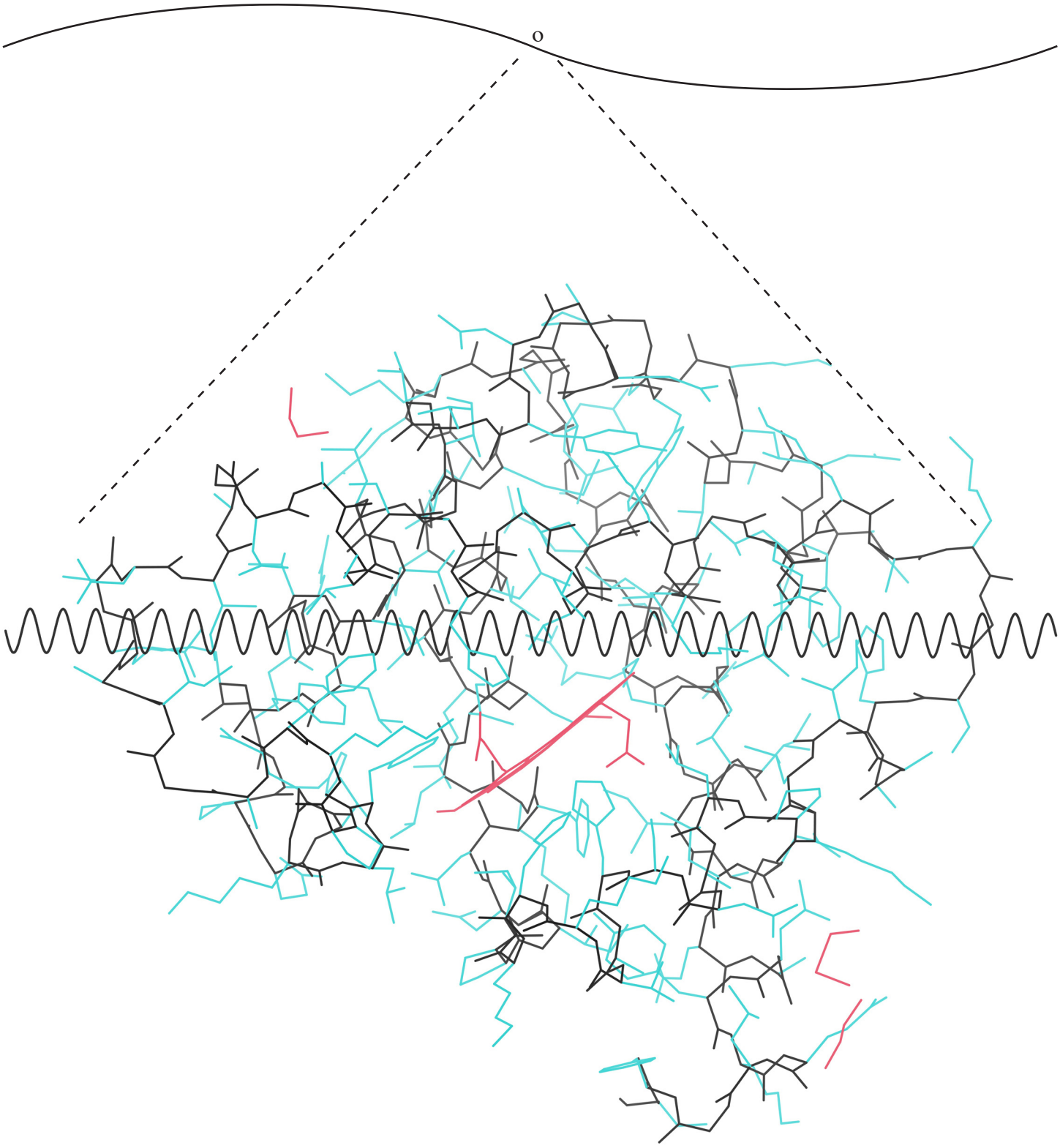




The characteristic x-ray emission of copper, which is often used for protein crystallography, has a wavelength of 1.54 Å. There is no known way of making a lens that will directly refract x-rays to make an image, so we are forced to measure the scattered x-rays at the place where the lens should have been, and use a computational method of recombining them into an image. This would not be too difficult a procedure if it were not for two very serious technical problems that arise: the **size problem** and the **phase problem**.

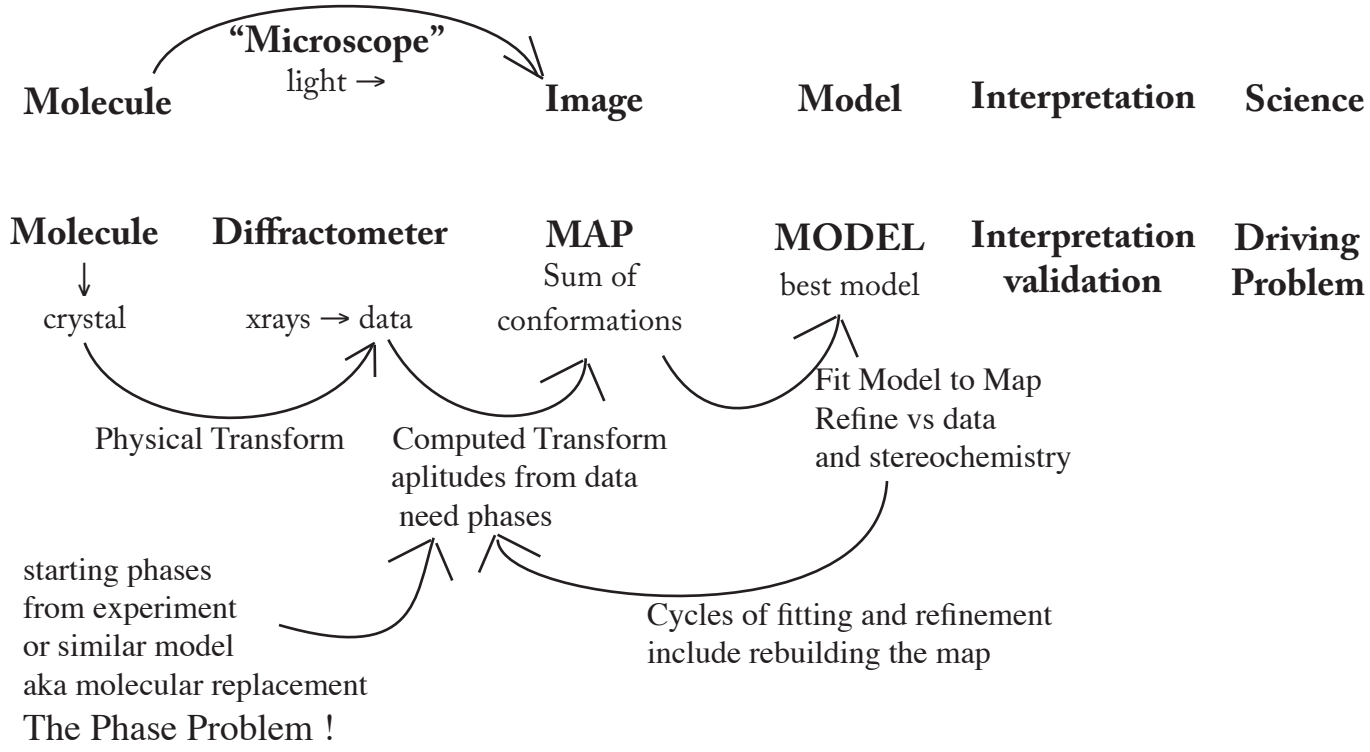
Match of wavelength to features we want to see:
x-ray: 1 Å, atom: 3Å, protein molecule: 50Å, (bonded atoms: 1.5Å apart)
Green light: 5000 Å, scale to 5meters = 5000 mm , scaled protein: 50mm = 5cm, (bonded atoms: 1.5 mm)

→ Draw 1 wavelength across both boards (of 147Nanaline),(Note change of phase along wave)
look at two 5 cm patchs: (or two neighboring 3mm patches), myoglobin would be about 5cm in diameter.
even if they would scatter this light, could we tell we had two of them?



32 cycles of 1.5Å across the diameter of myoglobin: → 2NRL.kin 2nrl.omap in KiNG

Macromolecular Crystallography, **The Experiment**



General Issues

What is the information content of the experiment?
distance & direction

How does one get information out of the experiment and into the model?
measure, calculate, make image, interpret...

What can we know about the reliability of the model, both in general and in detail?
validation and feedback

Specific Toe-catchers

All atoms contribute to all Data, all Data contribute to all parts of the Map.

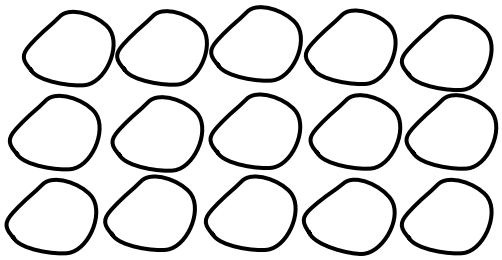
Map shows SUM of all conformations.

There are TWO kinds of waves:
real x-rays
computed waves to construct the image

The Size Problem ---> CRYSTALS

A single molecule (even a large protein molecule) is so small that it does not by itself scatter a sufficient amount of x-rays in a reasonable time to accumulate enough information to form a detailed picture. Also, x-rays of a wavelength short enough to resolve molecular details are energetic enough that the molecule would be destroyed long before its image could be produced.

The solution to this problem is to use a very great number of identical molecules packed together in a very orderly three dimensional array: a crystal.

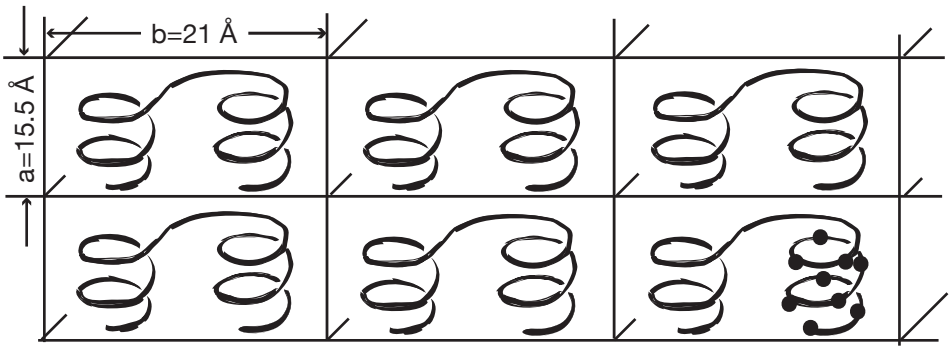


This gives **scattering power.**

But one will need to understand how x-rays interact with crystals. This is the general subject of x-ray diffraction.

One will also need to grow suitable crystals (of dimensions on the order of a few tenths of a millimeter,) which is a very tricky art indeed.

CRYSTAL Packing

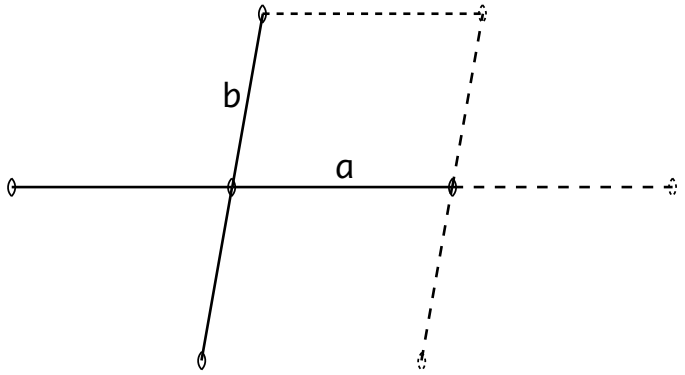


Unit Cell & possible symmetry in a crystal lattice

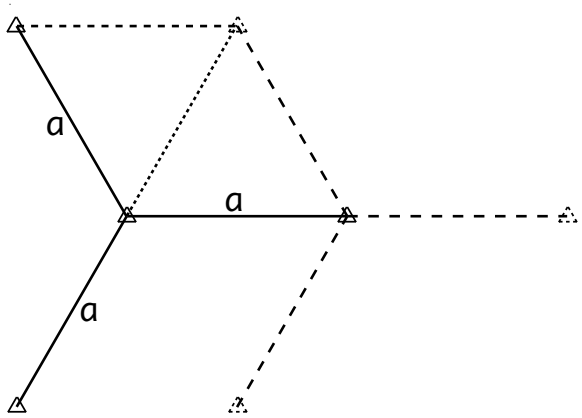
A unit cell fills all space by translations along its edges.
Such a unit-cell translation relates a point in space to an identical point. That is, the space surrounding one point looks exactly the same from any other point related by unit-cell translations.

n-fold axes and unit cell translations.

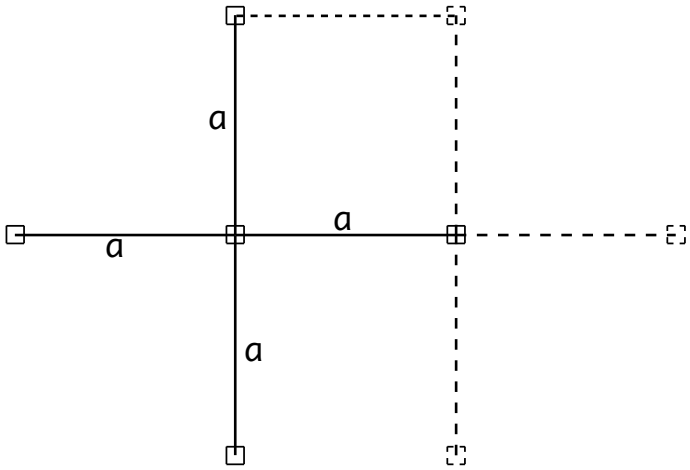
2-fold axes:



3-fold axes (also 6-fold):

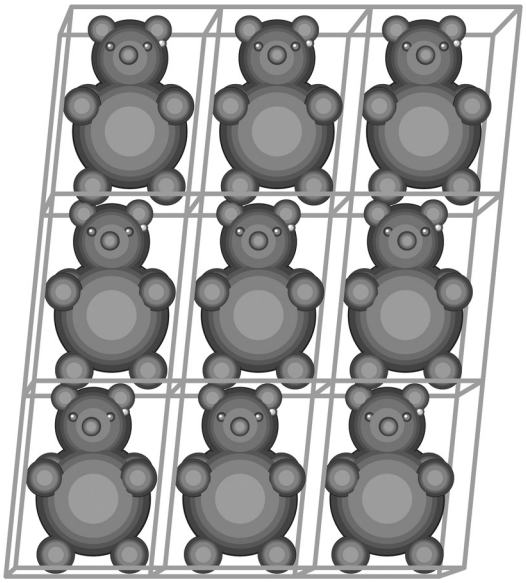


4-fold axes:



Objects in a crystal

1: Only unit cell translations, and there are no restrictions on the angles relating the three directions (axes) that describe the crystal.
All three angles free leads to the term "**triclinic**" for this kind of crystal.

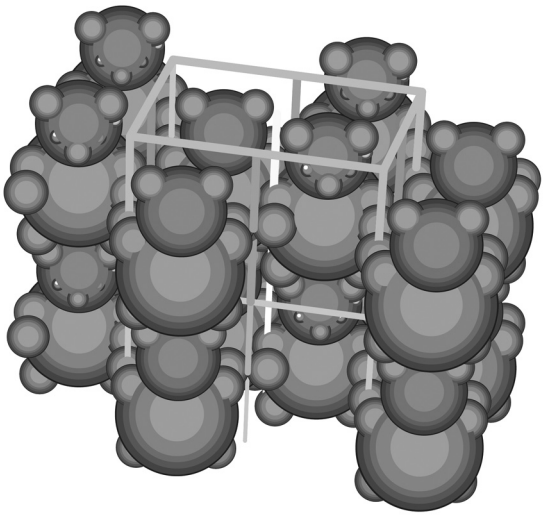


Here, the whole "unit cell" (that box that fills all space by translations along three axes) is the "asymmetric unit" (that volume whose contents are all the unique parts that exactly repeat through 3D space to make this crystal).

Note that this allows further "Non-Crystallographic Symmetry" within the asymmetric unit, (e.g. 2 arms, 2legs, etc.) but these relationships are not constrained to be exact.

However, for a single bear per cell, each unique feature, e.g. the earring, is exactly related to all other identical features by unit-cell translations.

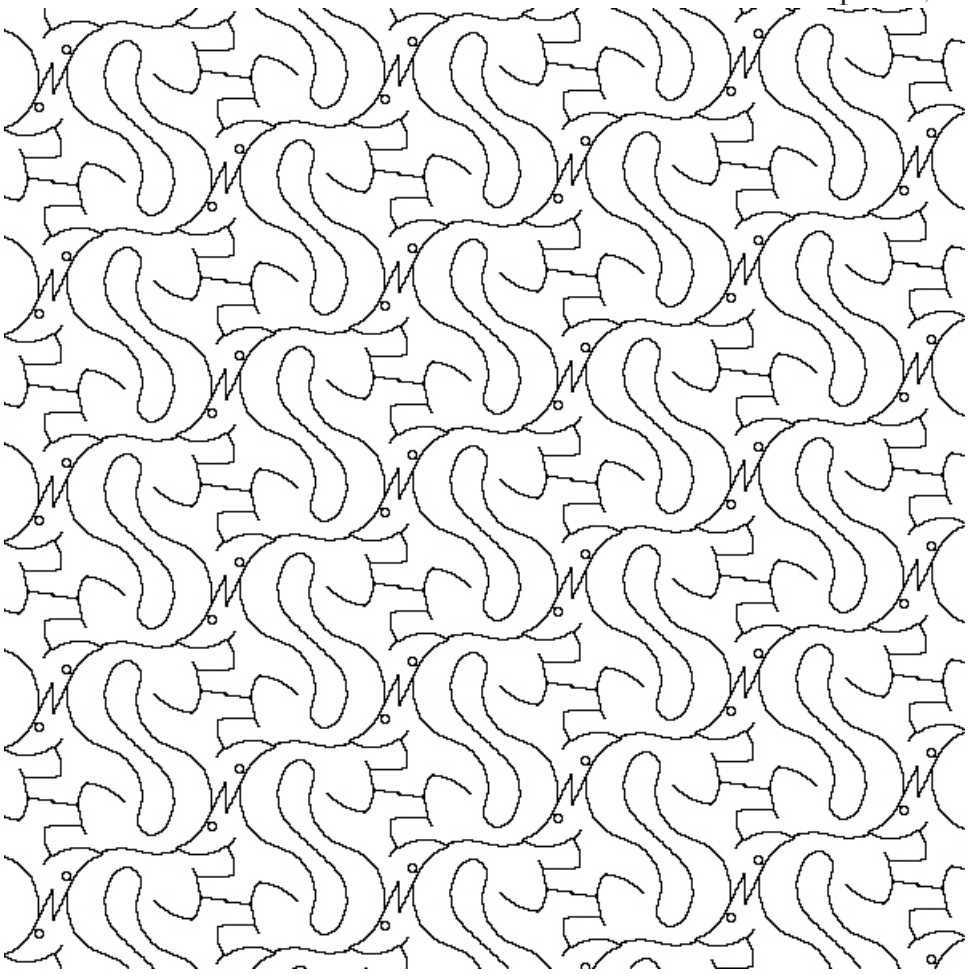
2-fold axes constrain 2 of the angles relating crystal axes to be 90°, but the third is not constrained.
So this type of crystal is termed "**monoclinic**" for its one free angle.



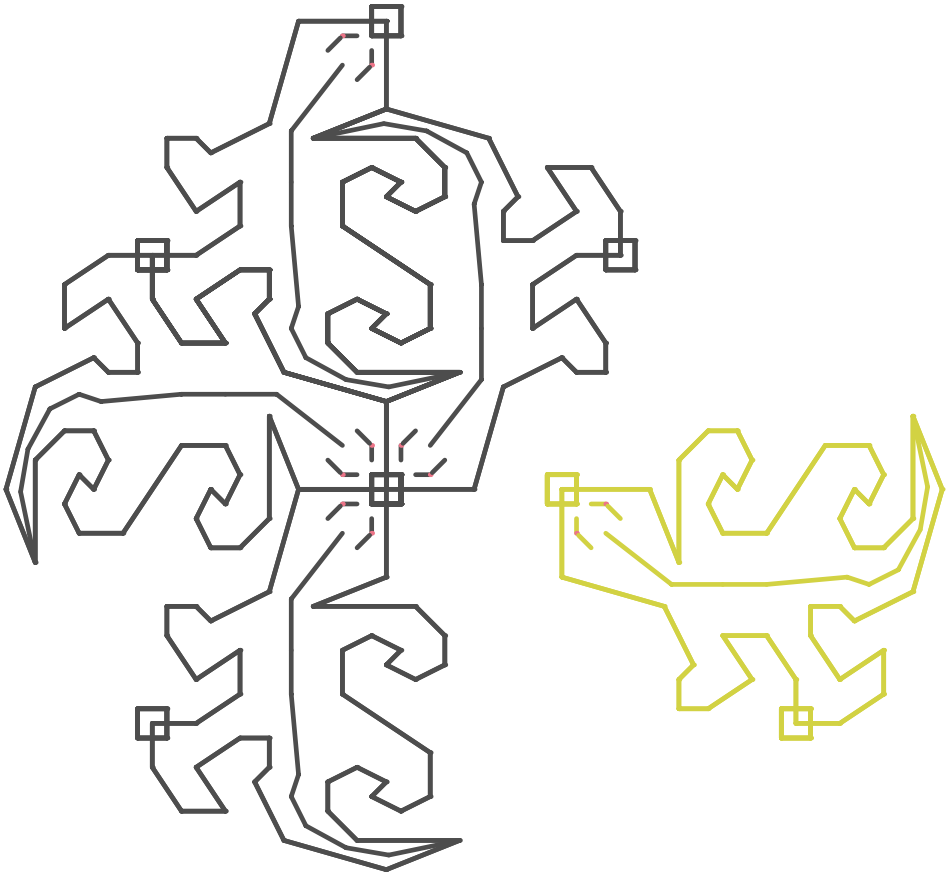
<u>2-fold axes and unit cell translations.</u>
The 2-fold axes run through the whole crystal and relate not only the contents of the "asymmetric unit" with its pair within the unit cell, but also the asymmetric units throughout all 3-D space.

Note that this allows further "Non-Crystallographic Symmetry" within the asymmetric unit, (e.g. 2 arms, 2legs, etc.) but these relationships are not constrained to be exact.

Additional 2-fold axes perpendicular to the first will contrain all angles to be 90°.
Hence "**orthorhombic**". A 4-fold axis constrains 2 of the axes to be equal, thus "**tetragonal**"
A 3-fold at least constrains the three axes to be equal, with "rhombohedral" or at least "trigonal" shape.
A 6-fold leads to "**hexagonal**". Combinations of axes can lead to a "**cubic**" overall shape.



Squirrels



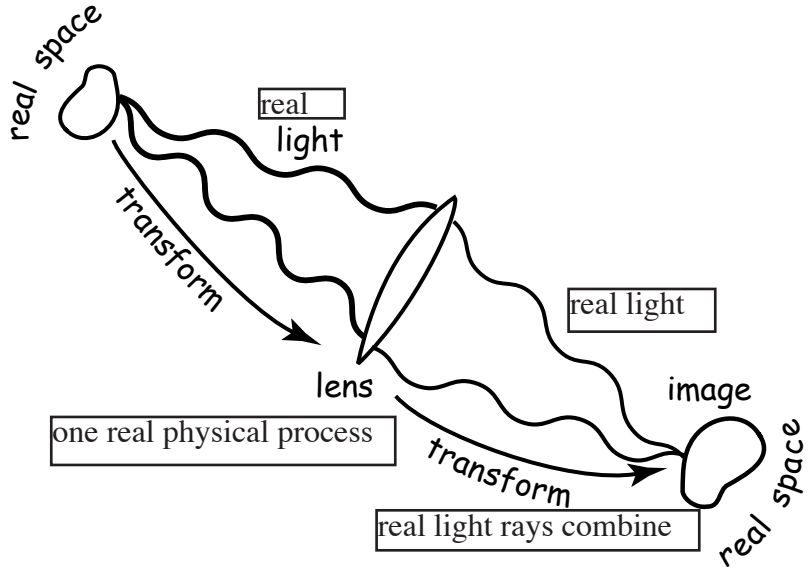
Lizards

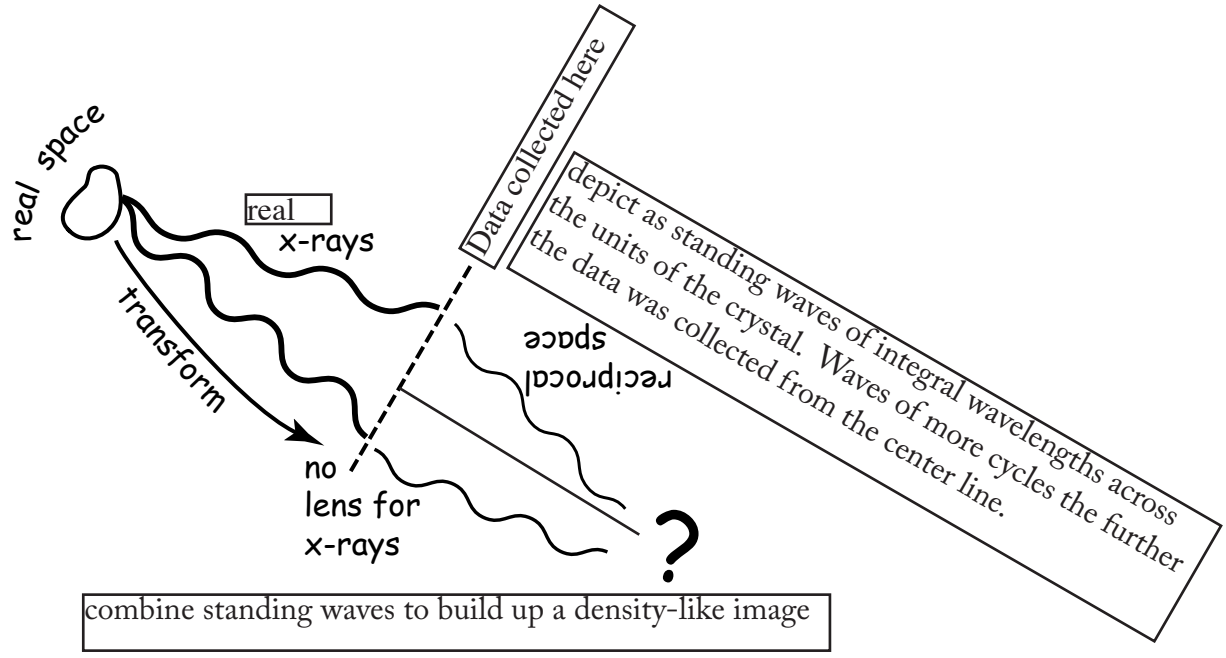
MAGE make kinemage
Practice Docking
tetramers extra

The Phase Problem --> Two Kinds of Waves

The Crystallography experiment: X-rays illuminate the crystal, diffract into discrete rays which leave the crystal. The rays are captured (instead of entering a lens), and the image is produced by computing the recombination of those scattered rays (instead of recombining naturally when a lens focuses the rays).

This recombination of the scattered rays is computationally simulated using equations.





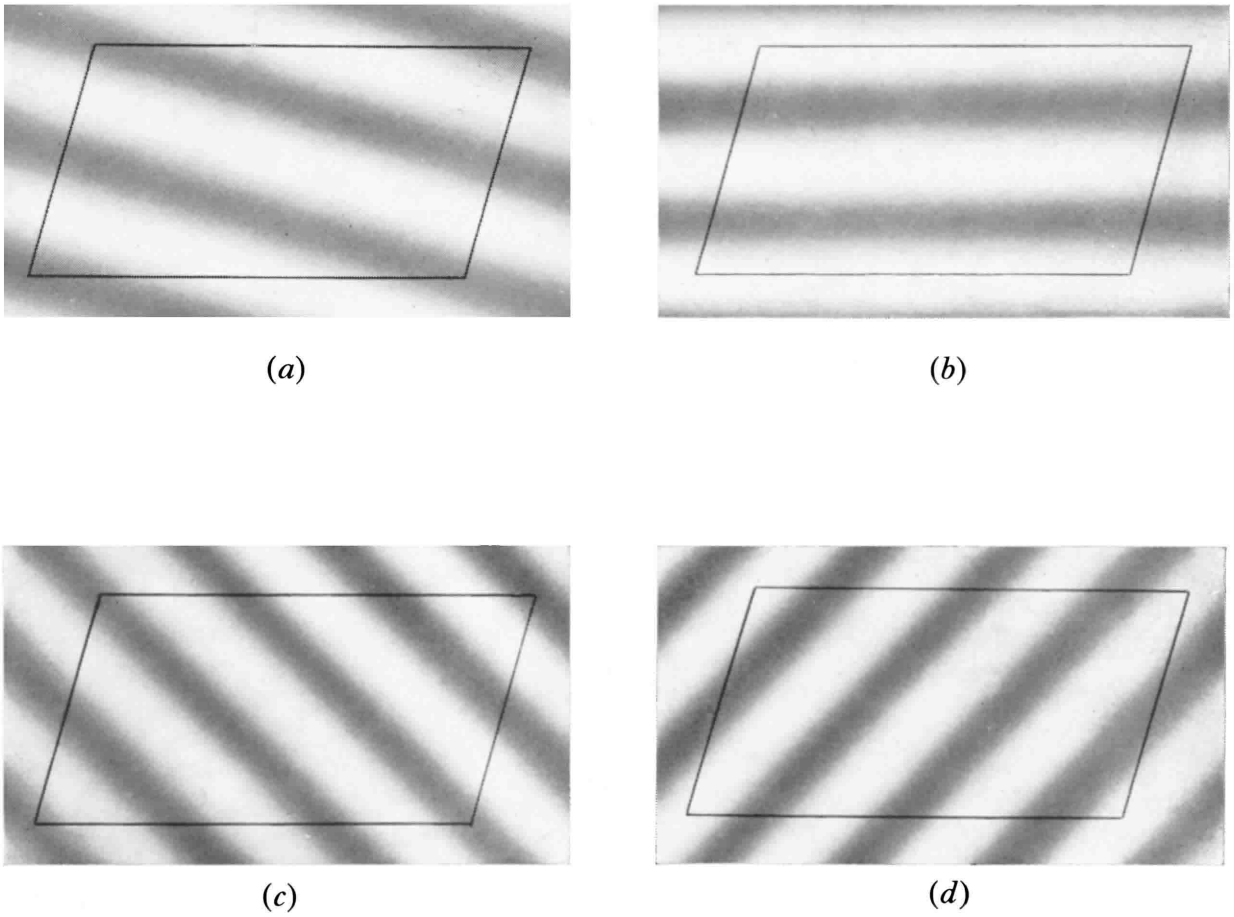
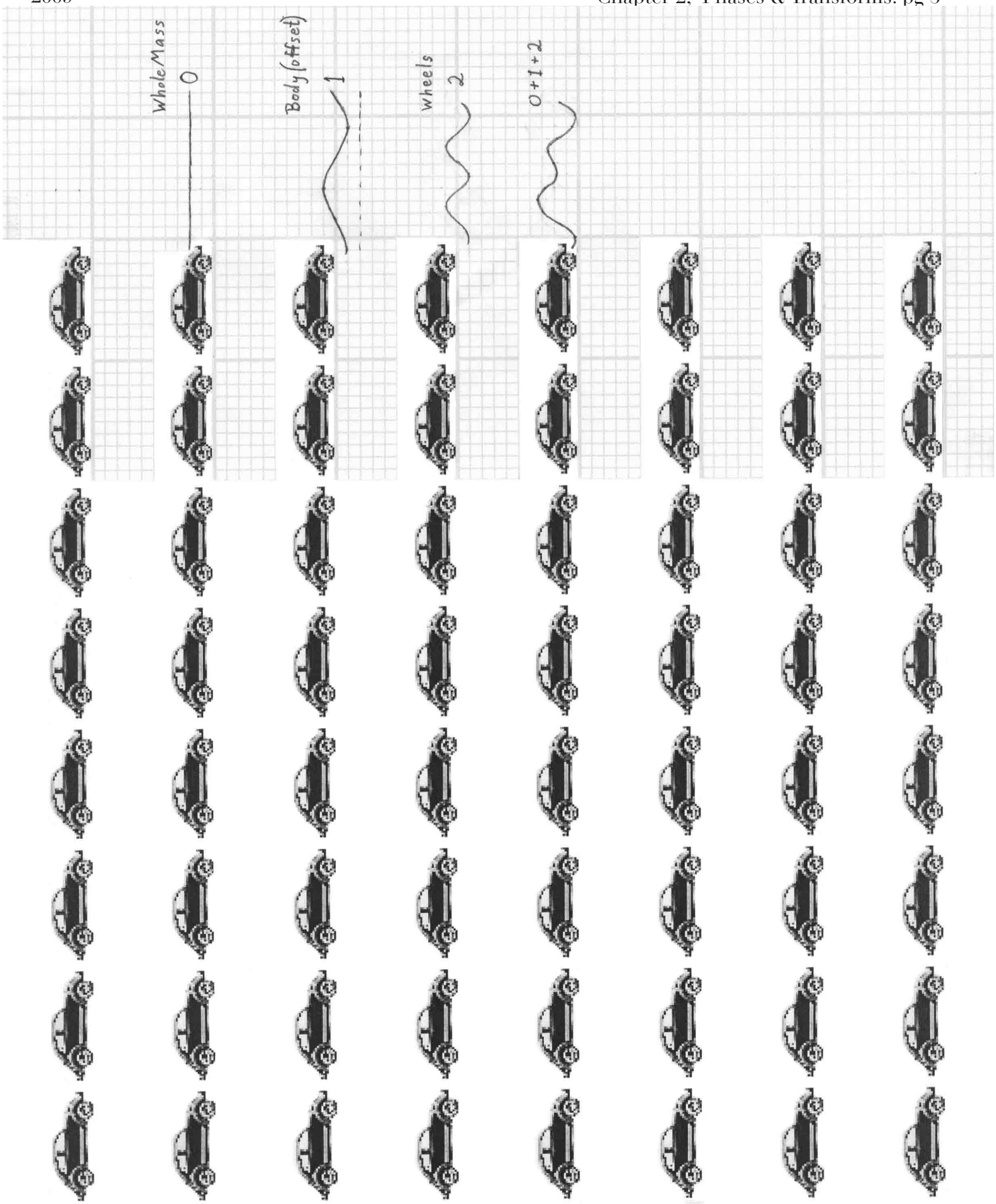


FIG. 138.—Sinusoidal alternations of light and shade. The bands in the figure represent the contributions to the image due to the following spectra:

(*a*) F(102), phase negative (*b*) F(002), phase positive
(*c*) F(302), phase negative (*d*) F(30$\bar{1}$), phase positive

(*Zeit. f. Krist.*, **70,** 483, 1929)

The concept of building up the image (of the molecule(s) in the unit cell) is to first uniformly fill the box with average electron density, then one-by-one add in standing waves where density is either added or subtracted according to the position and height of each wave. The VW Beetle figure shows starting this process using an example suggested in "Introduction to Macromolecular Crystallography" by Alexander McPherson, who also once had a '63 beetle. (His was yellow, mine was red.)

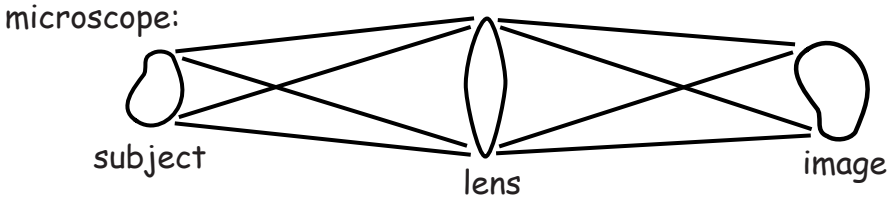


→ → see fftoys on web site

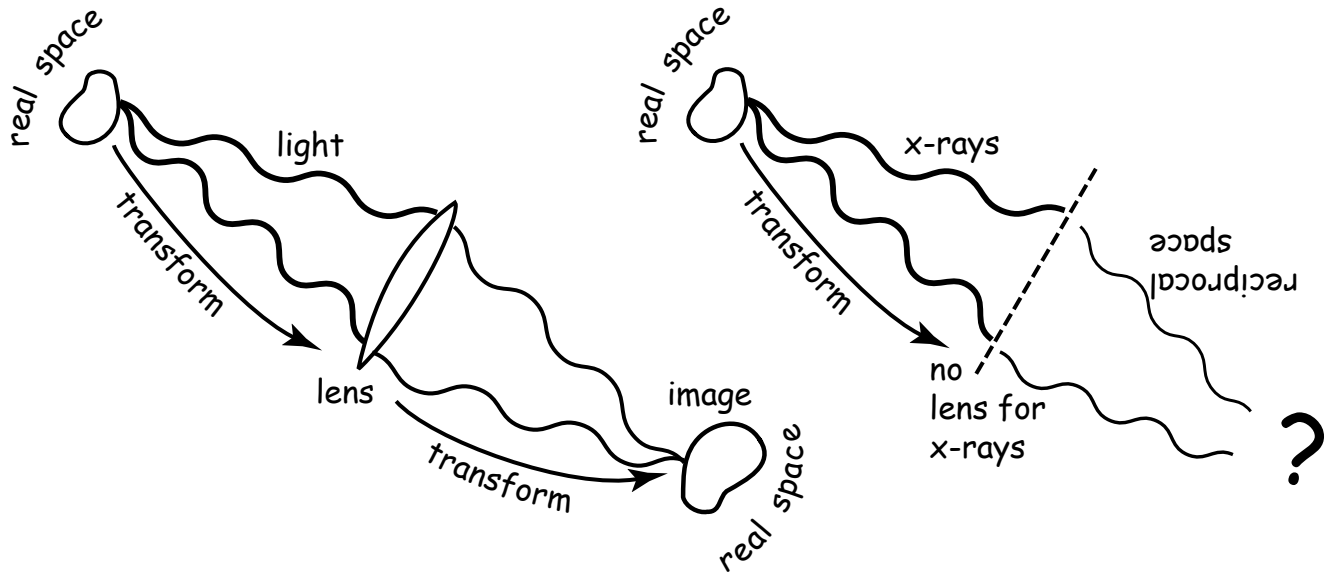
The Phase Problem restated

DATA: what can be measured and what is missing...

The only devices available to measure radiation respond only to the energy and not to the relative phases of the x-rays.



In a microscope every point on the lens receives light from every point on the subject, and the lens combines the light so that every point on the image receives light from every point on the lens. The physical combination of all these light rays depends on each ray's magnitude and phase..



In the x-ray case, we can measure the intensity of each scattered ray at the place where the lens would have been. Intensity = energy = (amplitude)^2. (The intensity of an x-ray is just its number of photons per unit time.)

$$I = F^2$$

in standard crystallographic notation. But we also must know the relative phases of those rays in order to compute what the image should look like, and unfortunately film, scintillation tubes, area detectors, CCD's, etc. do not record anything about the relative phases of the various rays. (That is, there is no way to tell when the crest of a wave passes, and for wavelengths of about 1Å going past at the speed of light in a machine made of real materials at room temperature that would be a real trick indeed!) The only way to get relative phases is to use interference effects with a reference wave, and that is exactly what one does!

To summarize the phase problem:

We need to know the phases of the scattered x-rays in order to perform the computation step to get the image, but there is no way of measuring the phases directly. The solution to this problem is the main work of crystallography.

Two kinds of waves, one kind of equation:

(Amplitude factor) • (Phase factor) the general equation of a wave

two varieties:

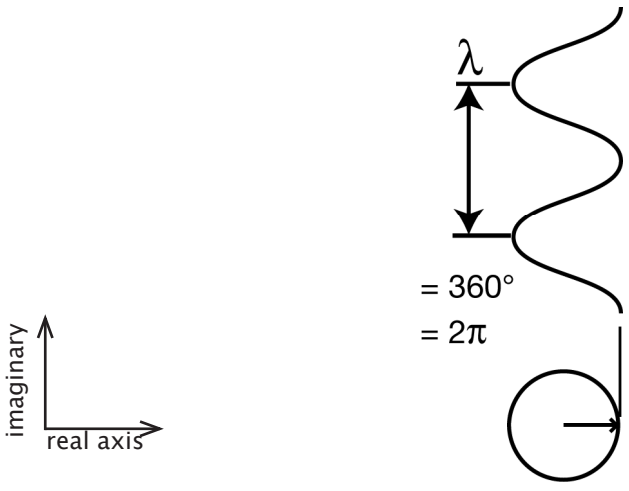
1) $\left|F_{hk\ell}\right| \bullet e^{i\phi_{hk\ell}}$ the expression for a real x-ray, with a real, experimentally fixed, wavelength.

A resultant diffracted x-ray wave from the crystal is the sum of x-ray waves scattered from each and every atom in that crystal in a particular direction. Each resultant wave, indexed as *h,k,l* , travels out of the crystal in that particular direction, so we will need to learn how to combine parallel x-ray waves to form a resultant wave.

2) (amplitude factor)• $e^{-i\,2\pi(\,hx + ky + \ell z\,)}$ some other kind of wave, with wavelengths that turn out to be integral fractions of the dimensions of the unit cell.

The electron density in a model of the crystal is the sum of these second kind of waves. Not only are the wavelengths of these density waves different from each other, each wave is going in its own particular direction. The wavelengths are integral fractions of unit cell dimensions (i.e. 1,2,3,... complete cycles within the bounds of the unit cell), thus they are standing waves. So we will need to learn how to combine standing waves in a box (the unit cell) to build up a density-like image.

Representing waves, the phase clock (with radius = amplitude):



ϕ factor ($e^{i\phi}$): exponential form convenient to talk about; $e^{i\phi} = \cos(\phi) + i\,\sin(\phi)$
cos() & sin() form (real and imaginary components) sometimes more convenient for computation.

Transforms:

Transform of a set of planes is a row of points.

Transform of a lattice is a lattice.

Transform of a single object is a continuous distribution.

Crystallography:

A convolution is a thing repeated by the rule of another thing.

e.g. a crystal of molecules is the molecule repeated at all points of the crystal lattice.

Transform of a convolution of two functions is the product of their individual transforms.

Transform of a lattice populated by objects is the transform of the object sampled at the points of the transformed lattice.

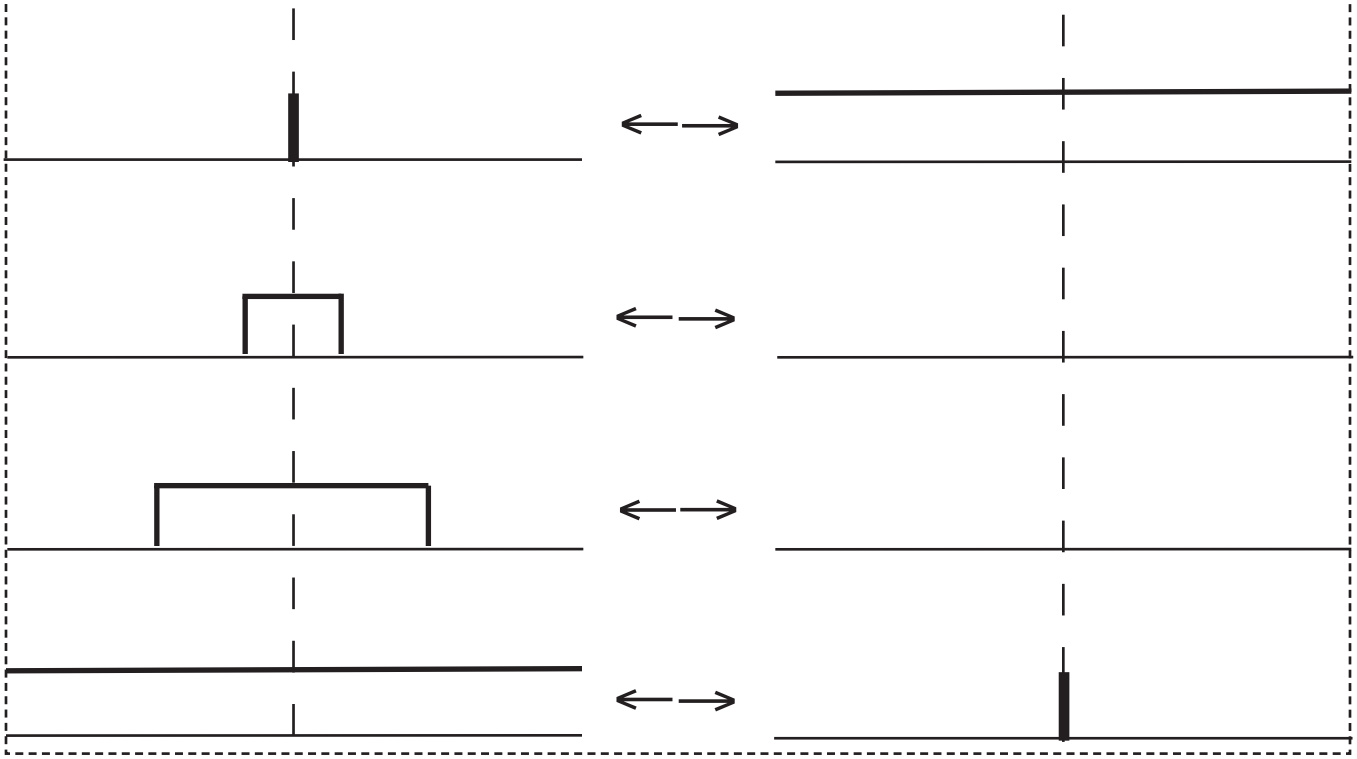
Getting it on the blackboard or on "film"

Transform of a projection of a 3D object is a planar section through the 3D transform of the object. (If stay in plane for transform: third dimension position of features of the object don't matter.)

Transforms for Dummies:

Transform of Short and Fat is Tall and Thin.

All these are reversible.→



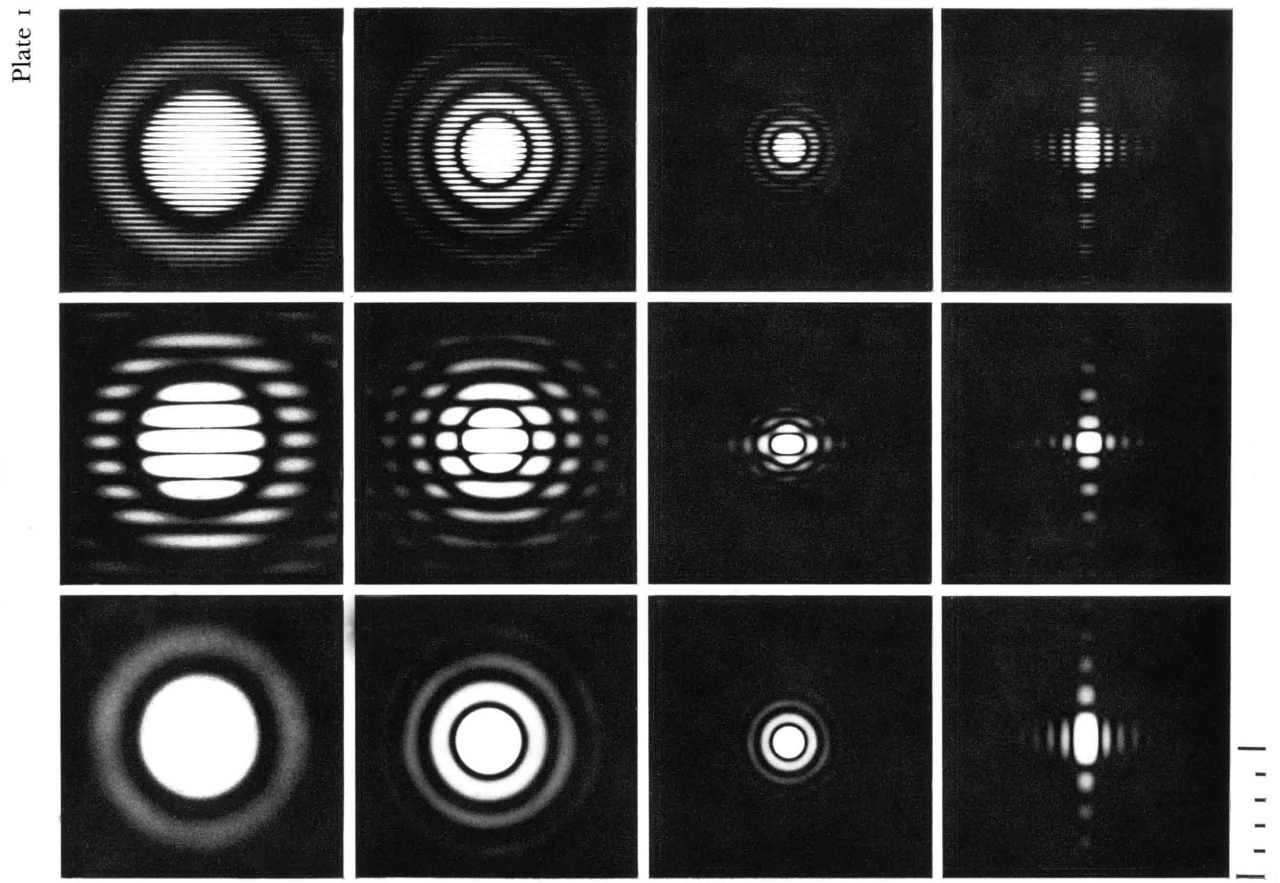


Plate 1

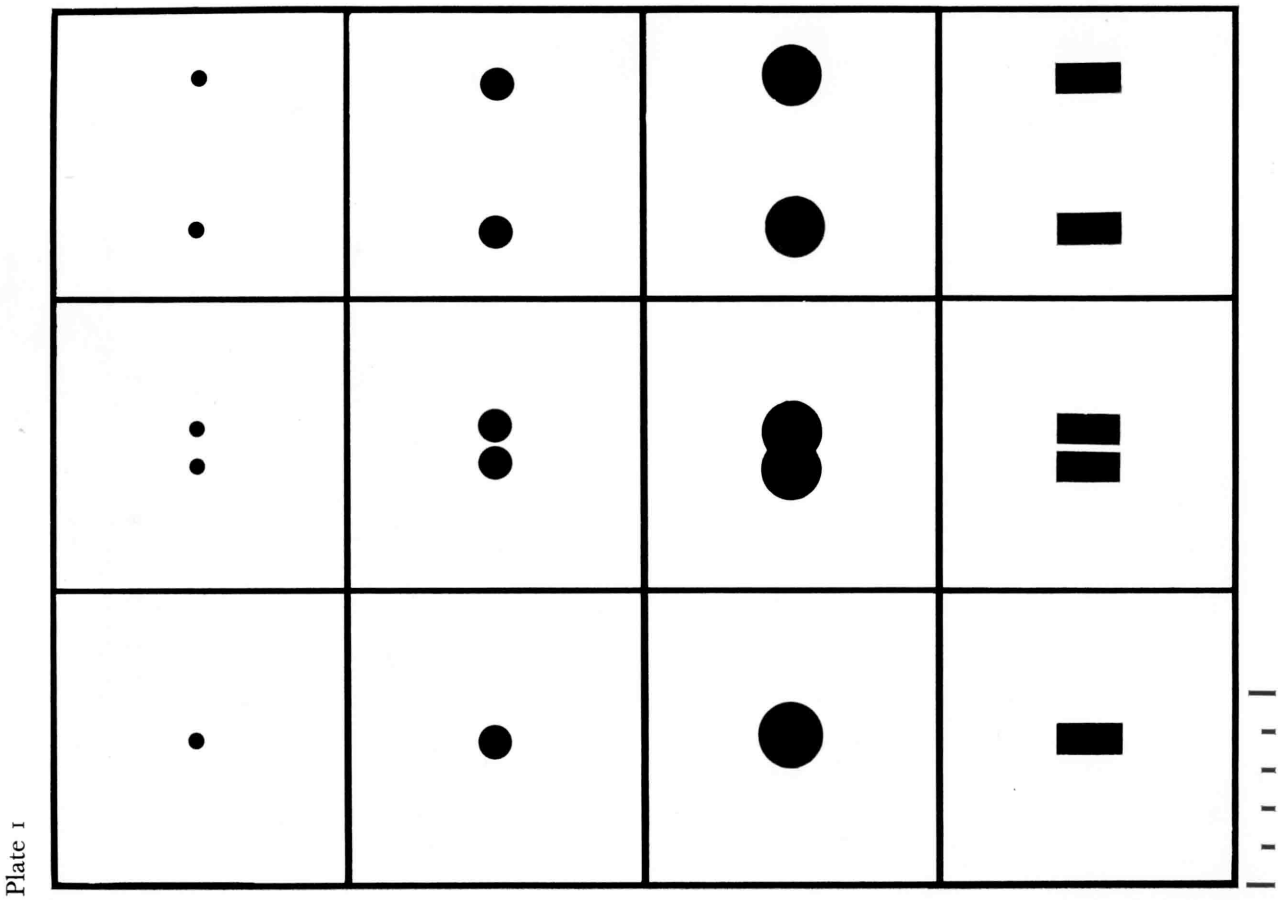
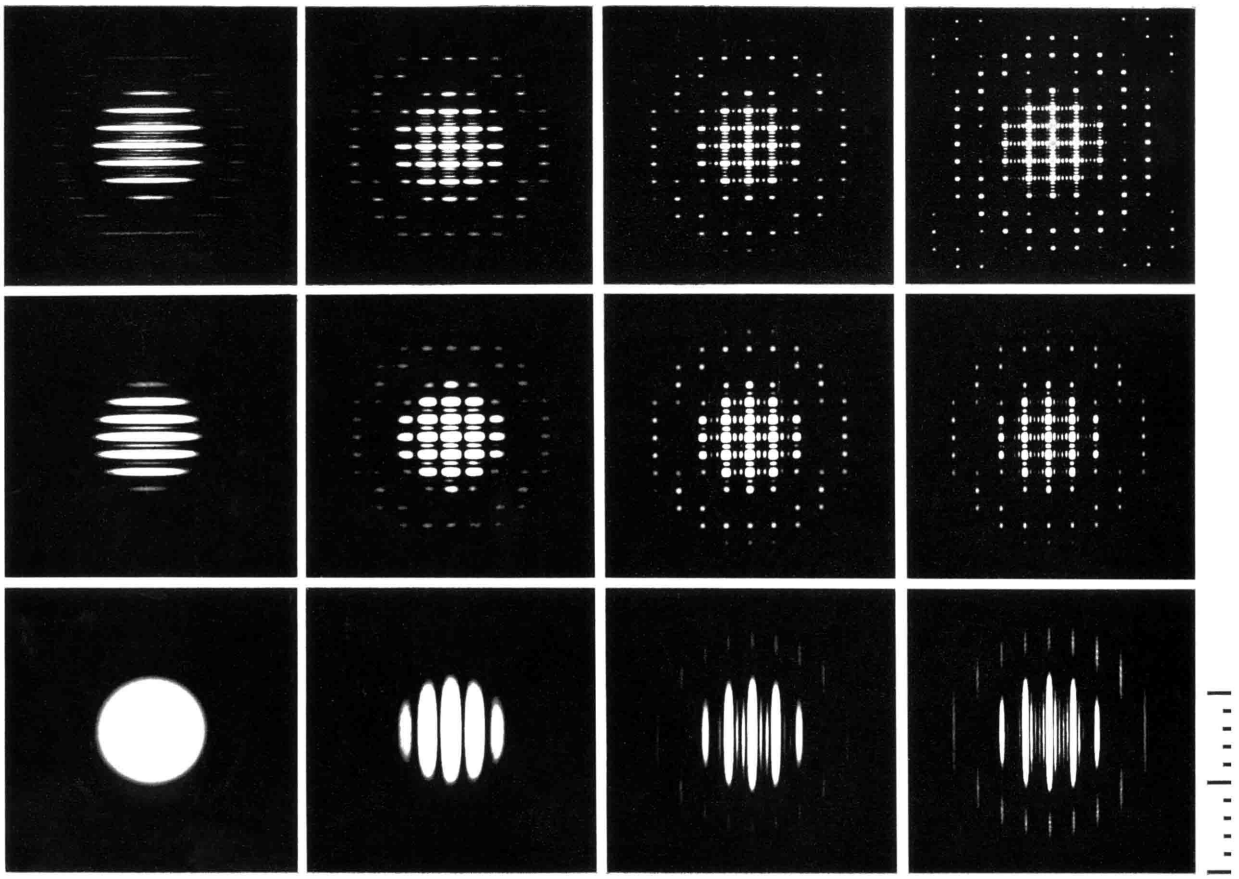


Plate 1

Plate 10



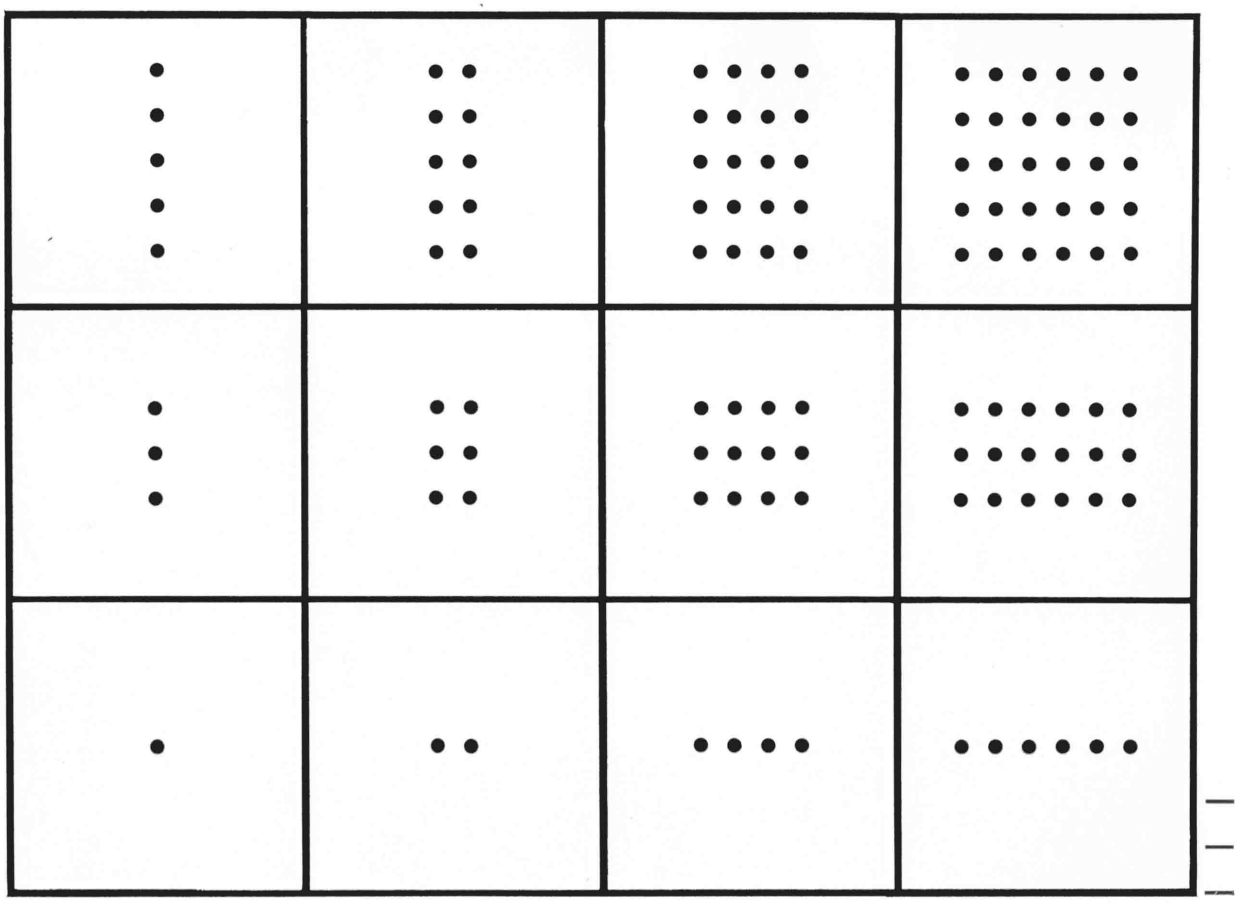


Plate 10

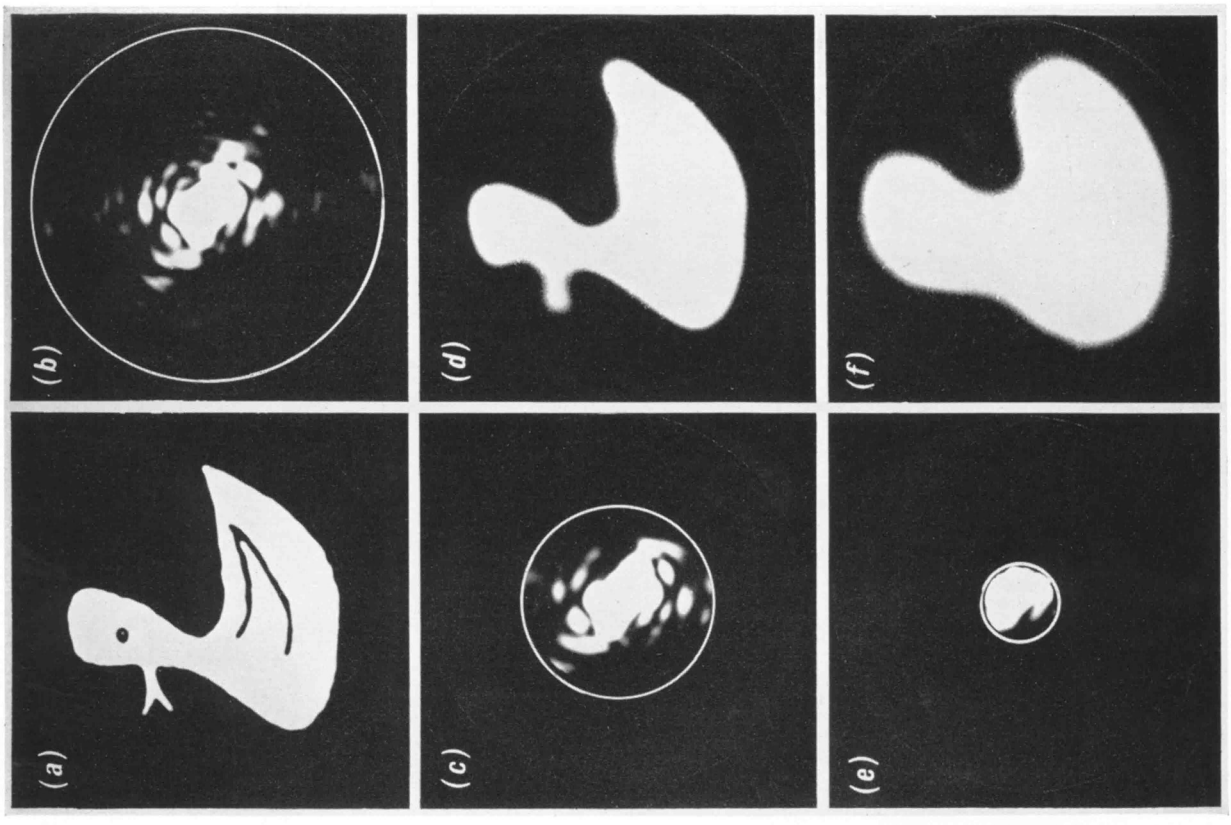


PLATE 44. (*a*) is an irregular object and (*b*) is its optical transform. (*d*) and (*f*) are the recombined images of the portions of the transform shown at (*c*) and (*e*) respectively. *See* **9.7**

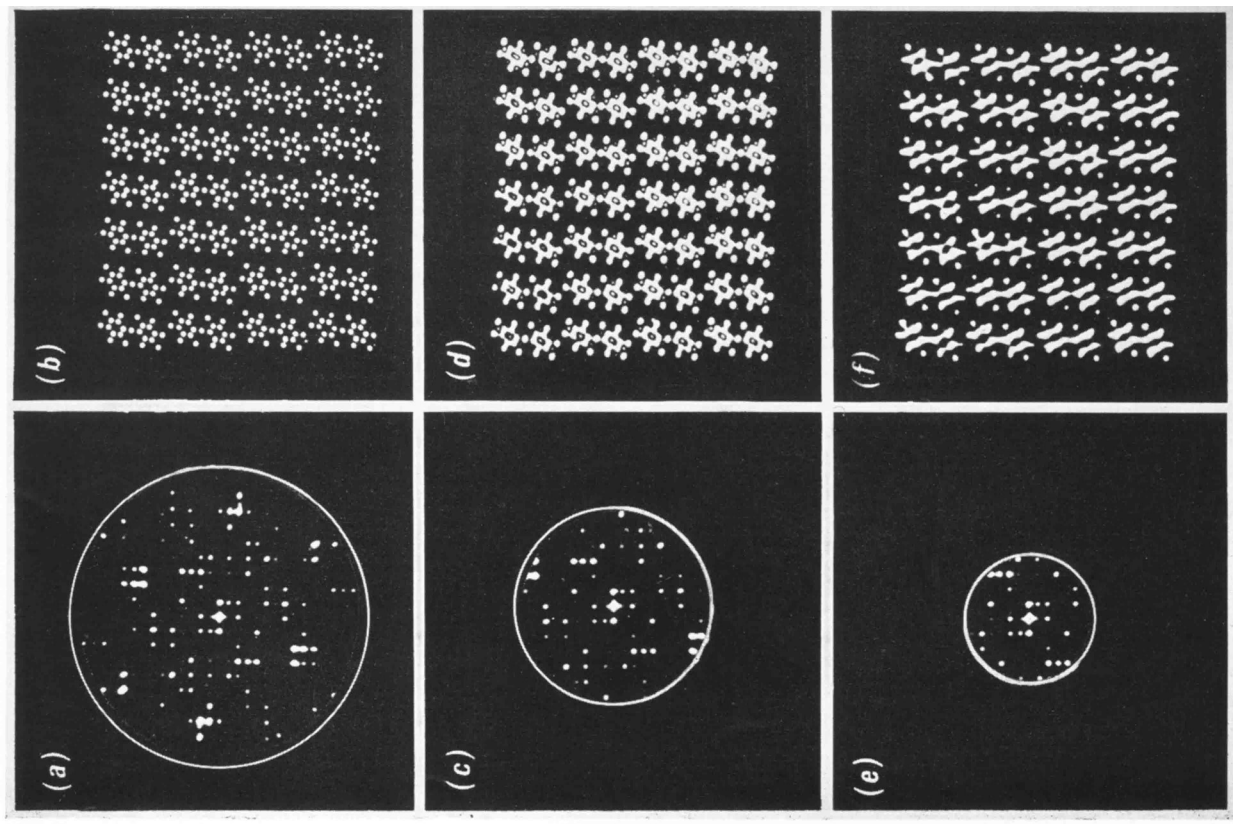


PLATE 45. Optical illustration of the effects of series termination in Fourier synthesis; (*b*), (*d*) and (*f*) are the recombined images of the portions of the transforms shown at (*a*), (*c*) and (*e*) respectively. *See* **9.8**

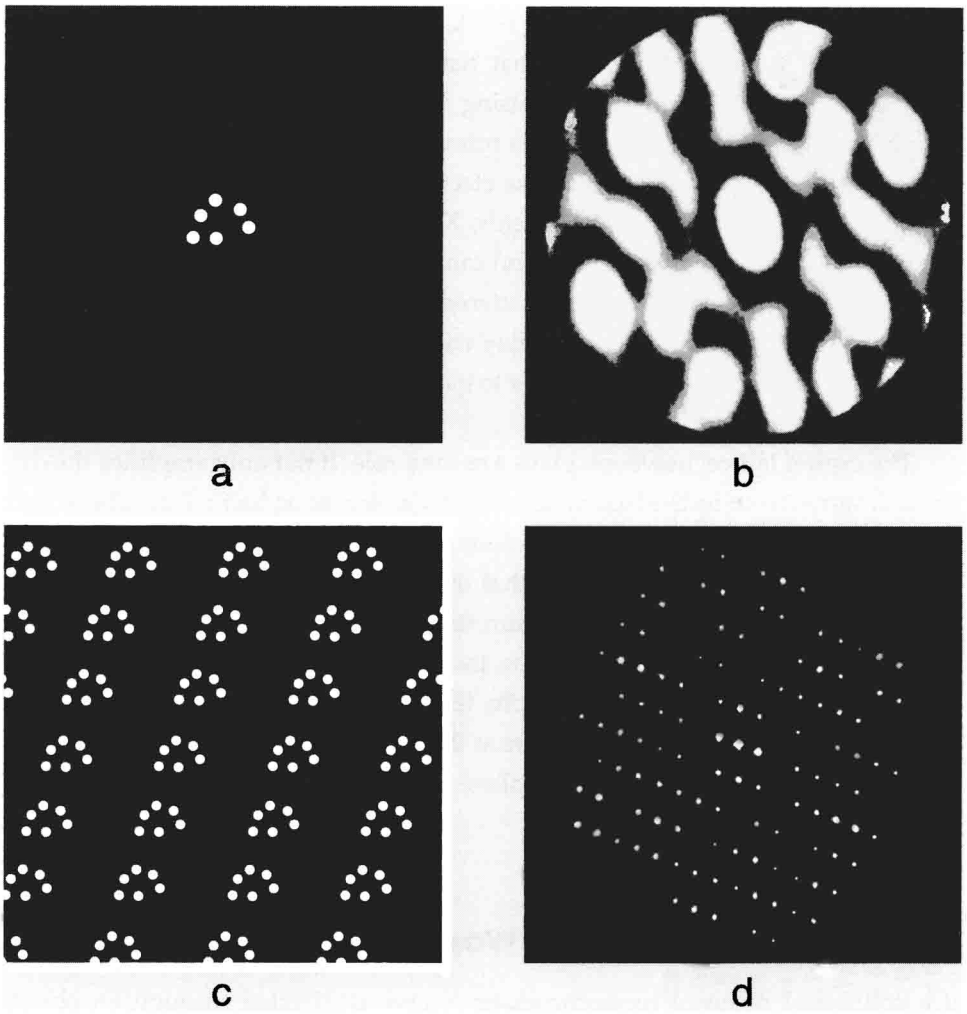
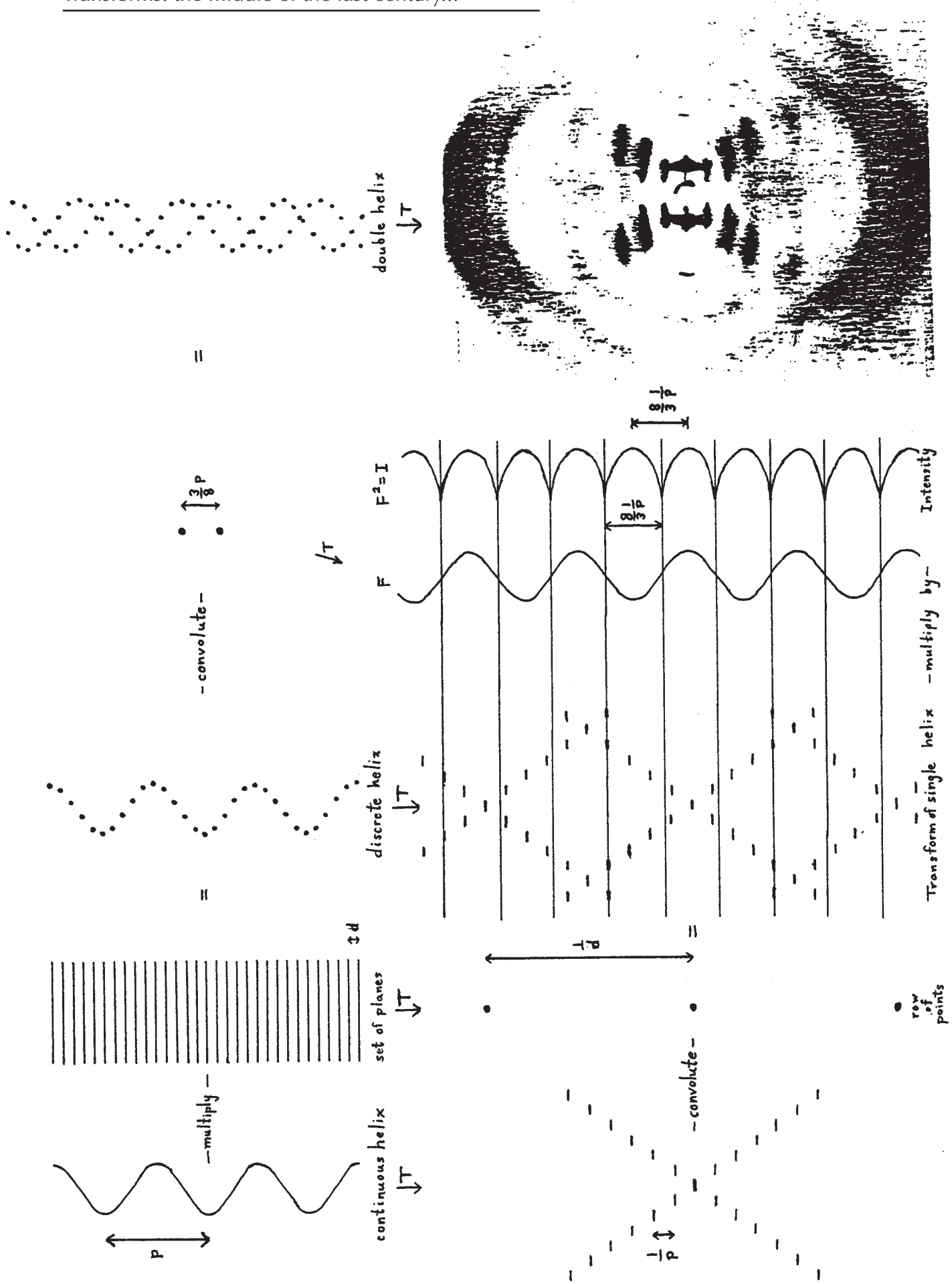


FIGURE 1.8 In (*a*) is an arbitrary set of points that might represent the atoms in a molecule, and in (*b*) is the optical diffraction pattern of that set of points. It is a continuum of light and dark over the whole surface of the screen. The mask (object) in the optical diffraction experiment in (*c*) is a periodic arrangement of the fundamental set of points in (*a*) in two dimensions, that is, the repetition of the object according to the instruction of a lattice. The diffraction pattern of (*c*) is shown in (*d*). We would find, if we superimposed the point array in (*d*) upon the continuous transform in (*b*), that the intensity at each point in (*d*) corresponded to the value of the continuous transform beneath. That is, the diffraction pattern in (*d*) samples the continuous transform in (*b*) at specific points determined by the periodic lattice of (*c*) (Taylor and Lipson, 1964).

this version from: Alexander McPherson "Introduction to Macromolecular Crystallography"

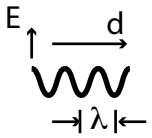

double helix
T
=
3/8 P
-convolute-
T
F²=I
F
Intensity
-multiply by-
Transform of single helix
discrete helix
T
=
1/d
row of points
-convolute-
1/P
-multiply-
continuous helix
T
P
set of planes
T
1 d

Light

There is a particular model for light and matter which allows a consistent explanation of a very large part of the subjects which can be included under the general term "Spectroscopy".

3 different representations

Wave of wavelength, λ, each wavetrain (broad and long enough to perform as we wish)



Particle specific quantum of energy $\propto 1/\lambda$, a photon = 1 quantum, a chunk of a wavetrain (broad and long enough to perform as we wish)

Ray direction of wave propogation, or direction of a stream of photons

Light has both electric and magnetic properties, e.g. electromagnetic wave. One can almost always ignore the magnetic component and still adequately explain experimental phenomena.

Light interacts with matter: driving wave interacts with an oscillator.
Oscillator must involve an oscillating dipole, this could be an induced dipole in a polarizable object. So we are getting a measure of polarizability of a medium = molecules.

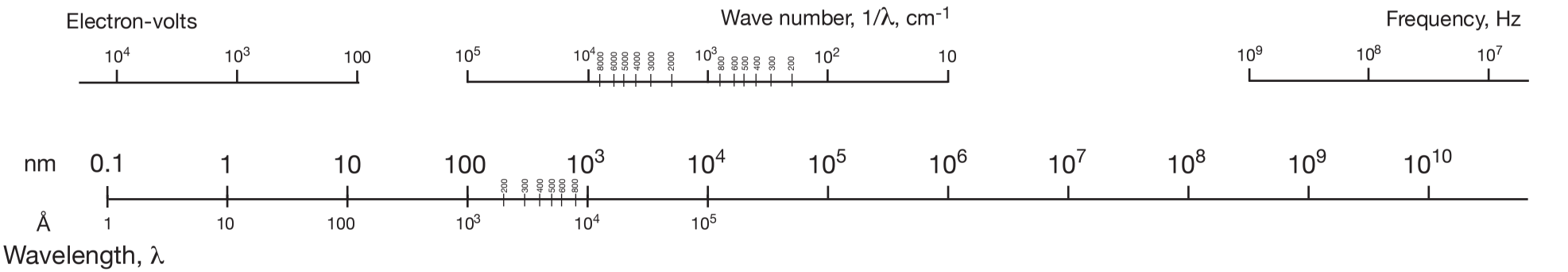
Energy can be transfered from (or to) this oscillator, for this introduction we will often ignore most processes except re-radiation of light. (i.e. we can creep up to an absorption band, but avoid actual absorption). Absorption is just the loss of energy from the oscillator before it re-radiates. Absorption thus can be thought of as energy lost in the process of getting the oscillator started and keeping it going.

Oscillator (oscillating dipole) can radiate energy (light)
So we must consider certain points:

1) Must investigate properties of original interaction
2) Must know character of re-emitted light: is it different from original wave? If so, how?
3) Must investigate how light waves interact with each other since we have opened the possibility of various light waves: e.g. original wave and emitted waves from various oscillators.

Avoid quantum mechanics: except note that any system, including our oscillators, can exist in only certain energy levels; so energy is handled only in discrete chunks which will be important occasionally to consider. But, the picture of a driving wave and a simple oscillator does, in fact, explain much of the observed phenomena!

The Electromagnetic Spectrum



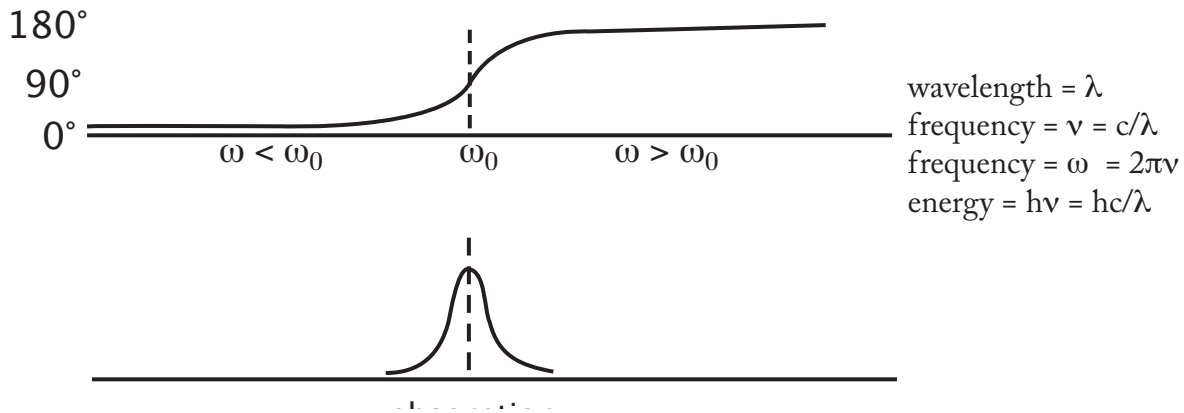
Light As A Wave

Interaction of light and matter

Analogy: Damped Driven Simple Harmonic Motion

→ → Example: bound "electron" and driving "wave"

Phase lag of an oscillator, scattered wave always in phase with oscillator:

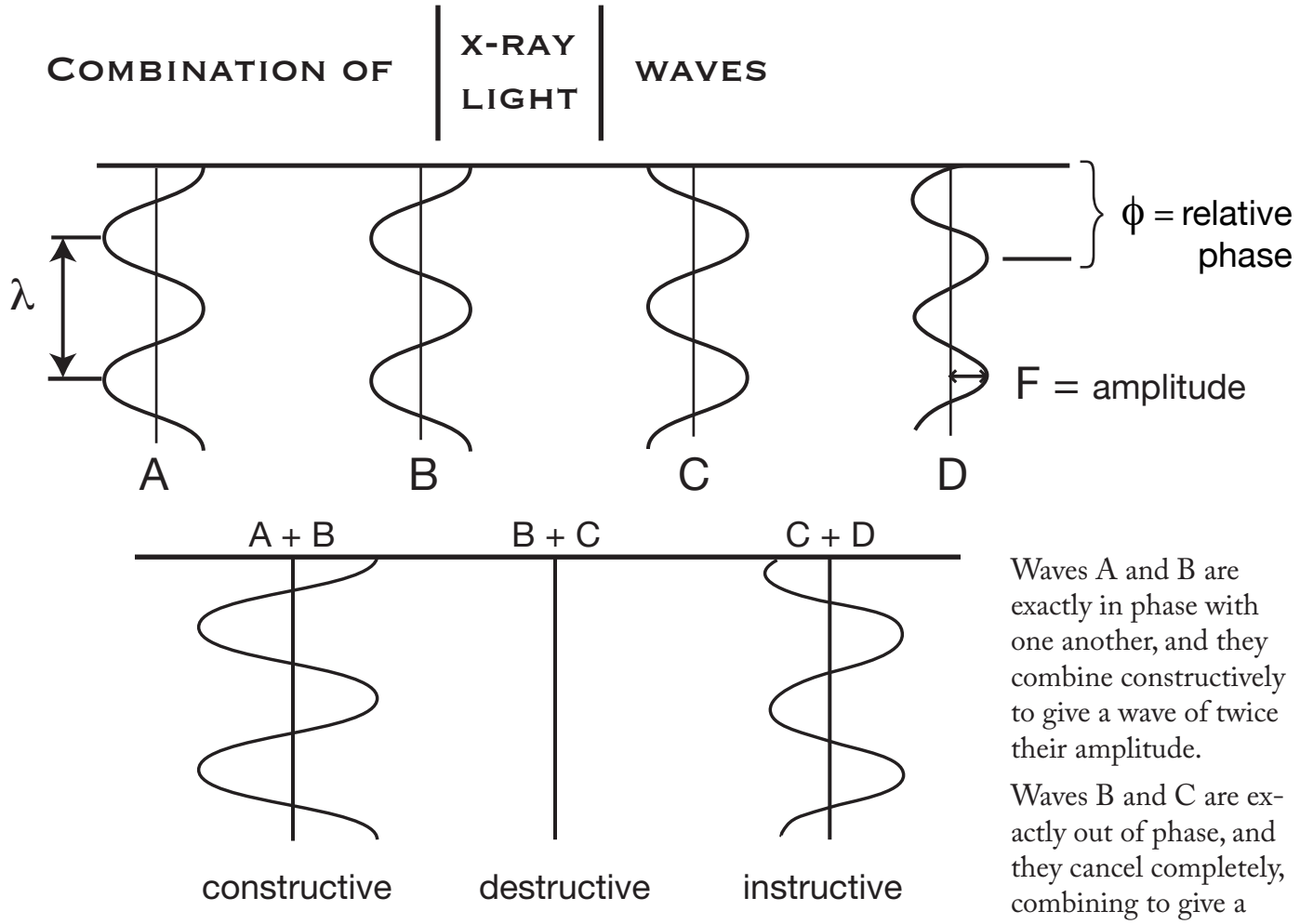


Quantum mechanics has its own way of getting broad absorption bands.

Identify the "oscillators" and we have the field of Absorption Spectroscopy.

Points to note:

1. Oscillator acts as a radiation source: radiated wave is in phase with the oscillator.
2. The energy that can be pumped into the system, and thus the amount reradiated increases as $\omega \rightarrow \omega_0$
3. For visible light $\omega < \omega_0$, so oscillator and emitted wave in phase (0°) with the driving wave.
   For x-rays $\omega > \omega_0$ for most electrons in atoms, so oscillator and resulting scattered wave are 180° out of phase with the driving wave. Since one is usually concerned with the scattered wave, this is often defined so that the scattered wave is at 0° and the driving wave at 180°.
4. Usual to treat phase of actual scattered light as having a usual component at 0° and an anomalous component with 90° phase difference (lag). For x-rays with scattered wave redefined to be 0° this is 90°phase advance, see later for a way to draw this).



Waves A and B are exactly in phase with one another, and they combine constructively to give a wave of twice their amplitude.

Waves B and C are exactly out of phase, and they cancel completely, combining to give a wave of zero amplitude.

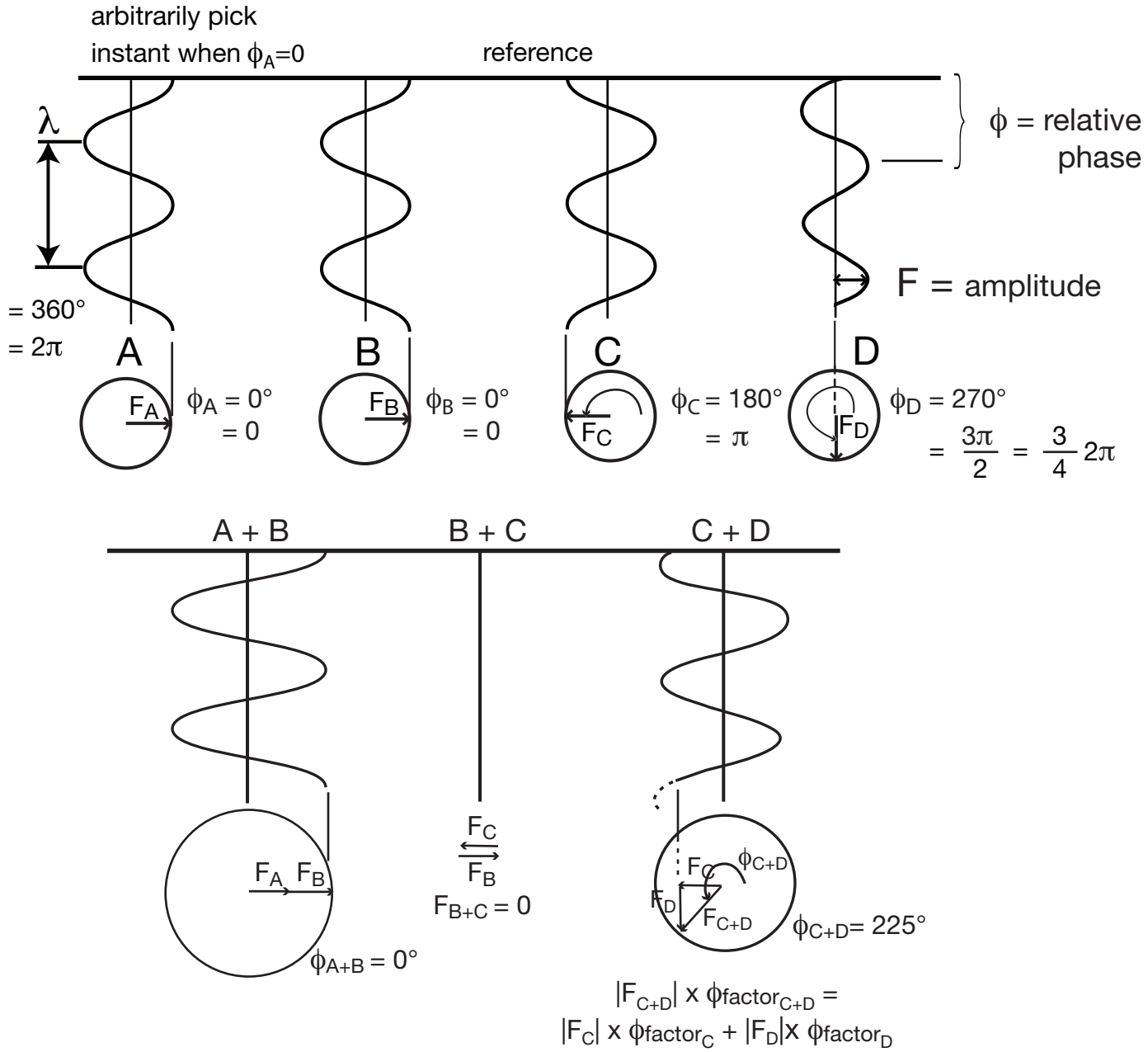
Waves C and D are somewhat out of phase, and they combine to give a wave with an intermediate phase and less than twice their original amplitude.

Scattering \'skad-er-ən\ :
Electrons in the path of an x-ray wave are set into forced vibrations by the periodically changing electric field of the x-ray wave passing by. These oscillating electrons are themselves sources of x-ray waves. This forced oscillation is of the same frequency as the incident x-ray wave, and the emitted x-ray waves are thus of this same frequency. By this interaction the electrons are said to scatter the original x-ray wave.

Diffraction \dȧ -'frak-kshən\:
Cooperative combination of scattered waves. This can occur if the scattering points are arranged in space in some regularly repeating manner. In certain directions with respect to the incident wave and the array of the scatterers, the scattered waves will combine constructively. This can be used to increase the amount of scattered radiation to a point where meaningful measurements can be made. This application is used in crystallography; for, in our case, the study of globular proteins. Conversely, the existence of a diffraction pattern and its characteristics can be used to detect and deduce something about a repeating structure. This aspect is exploited in the case of fibers, fibrous structures, and other extended polymers.

Vector notation for waves

The length of the vector represents the amplitude of the wave, and the direction of the vector represents the phase of the wave.



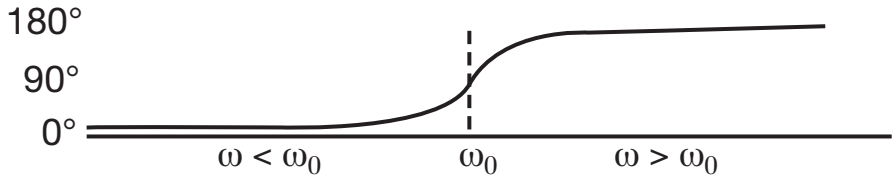
The construction of adding the two vectors head to tail gives the correct amplitude and phase for the combined wave, as above. These are just vectors in the complex plane and can be written as:

$$|F_{C+D}|e^{i\phi_{C+D}} = \mathbf{F}_{C+D} = |F_C|e^{i\phi_C} + |F_D|e^{i\phi_D}$$

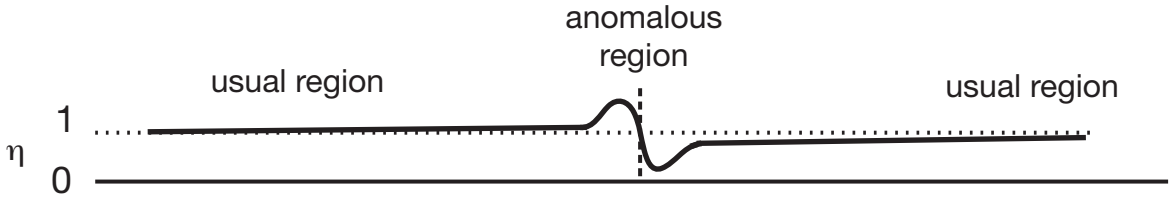
ϕ factor ($e^{i\phi}$): exponential form convenient to talk about; $e^{i\phi} = \cos(\phi) + i\sin(\phi)$
cos() & sin() form (real and imaginary components) sometimes more convenient for computation.

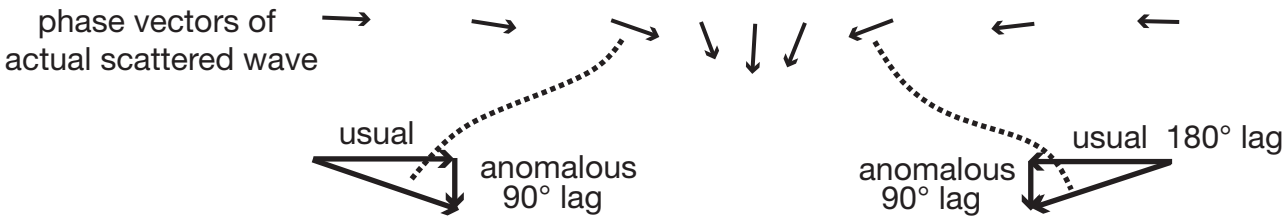
Phase shift

Phase lag of an oscillator, scattered wave always in phase with oscillator:



Refractive index is a measure of interaction:





Consider wave as sum of 2 components

Redefine usual scatter to be at 0°
i.e. rotate phase diagram 180°:

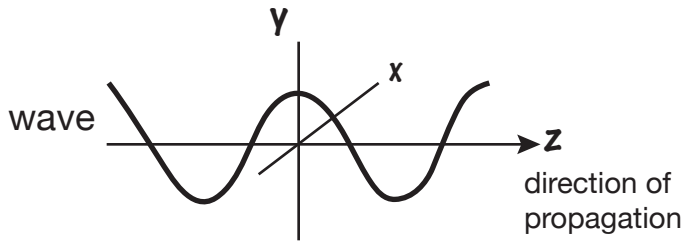
anomalous part now 90° phase advance with respect to usual

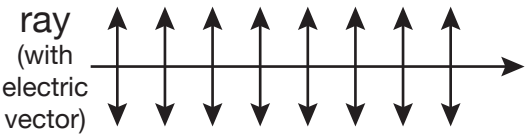
η is a measure of the interaction of light with matter: interaction affects relative phase of the oscillator with respect to the driving wave and a probability of feeding energy into the oscillator. The more energy in the oscillator, the more can be lost to other processes like absorption.

In any real situation there may be many types of oscillators in the medium interacting with the light but perhaps only one type near enough to resonance to have a significant anomalous scattering component. Also, in common situations almost all of the effective oscillators will have a resonance frequency either greater than the light frequency, as is the case for visible light; or much smaller, as is the case for x-rays. Any odd oscillator which happens to be far on the other side of resonance will make its unique and minor contribution 180° out of phase with the bulk of the scattered radition and thus very slightly diminish the scattered intensity.

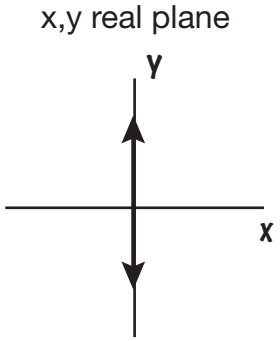
Representation of Light Waves

Plane polarized light: the simplest case

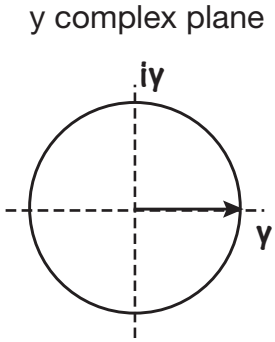




i.e. in plane of paper

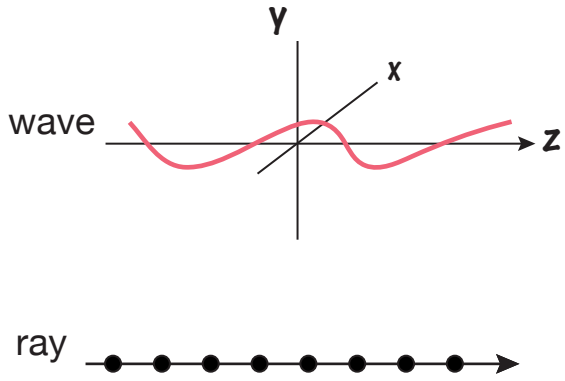


electric field vector
in y direction

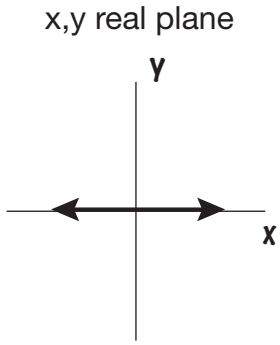


phase vector
$\phi = 0°$

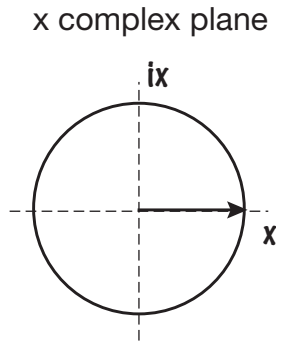
convention:
z-axis coming toward viewer;
right-handed coordinates



i.e. ⊥ to plane of paper



electric field vector
in x direction

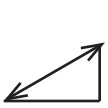


phase vector
$\phi = 0°$

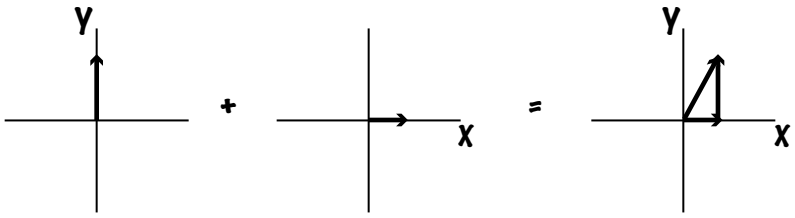
Unpolarized Light

Electric field vectors  in all x,y directions;

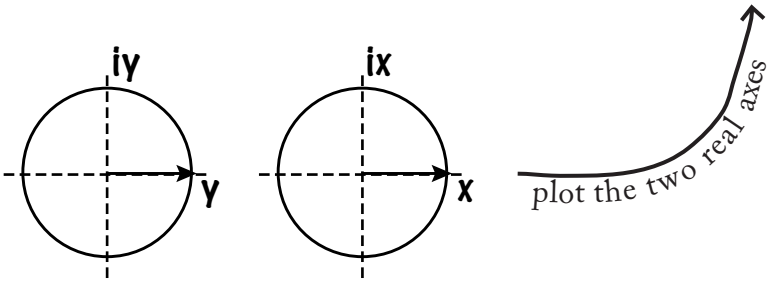
each component, e.g.  or 

can be treated as composed of x and y components in phase with each other:  

The components add directly,

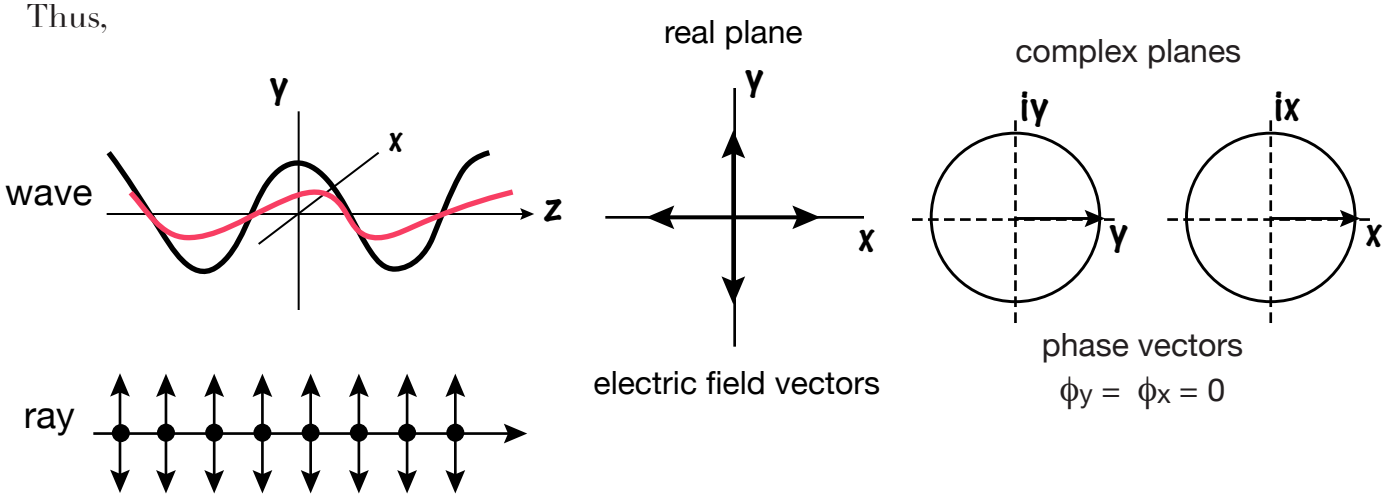


only if ϕ angle same;
i.e. can thus define
$\phi_y = \phi_x = 0$:



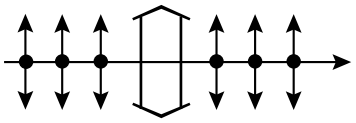
Thus can treat unpolarized light as the resultant of two plane polarized rays in phase. Any arbitrary photon in an unpolarized light ray can be broken into two components, and a great number of photons will then average out to give equal intensities in the two directions of polarized light chosen.

Thus,



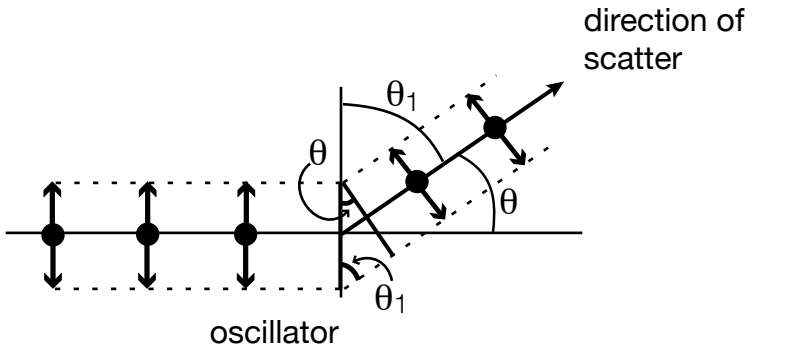
UNPOLARIZED LIGHT -- INDIVIDUAL OSCILLATOR

Straight through:



Perhaps a phase shift with respect to the original wave but otherwise no effect.
As seen from an angle, i.e. those photons scattered in a particular direction.
For example: direction of scattering in the plane of the paper:

$F_{\bullet scattered} = F_{\bullet original}$
$F_{\updownarrow scattered} = F_{\updownarrow original} \sin \theta_1$
$= F_{\updownarrow original} \cos \theta$



$F_{\updownarrow scattered}$ is a measure of the projection of the dipole motion ⊥ to the direction of scatter.

$$I_{original} = F^2_{original} \; ; \quad I_{scattered} = (F^2_{original}) \times (\sin^2 \theta_1)$$

The state of polarization and total intensity, $I = F^2$, will vary as a function of θ. In this simple case at $\theta = 90^\circ$, the ray is completely polarized.

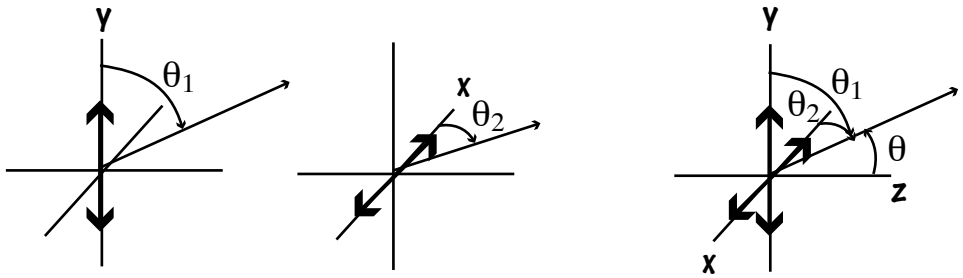
Later we will find that we can describe diffraction as a "reflection" from a plane (plane of oscillators) and see that this simple 2D diagram works as a projection to explain the general case for crystals.

UNPOLARIZED LIGHT -- INDIVIDUAL OSCILLATOR (DIGRESSION FROM CRYSTALLOGRAPHY)

General Case: single particles tumbling in solution

For scattering from discrete objects tumbling in solution we would need to consider the general case of scattering in all directions.

Example: general direction of scatter



y component x component

Scattering of I_o unpolarized = scattering from $\frac{1}{2}I_{o(x)} + \frac{1}{2}I_{o(y)}$ components, each identical except one has Intensity $\propto \frac{1}{2}I_o \sin^2\theta_1$; the other has Intensity $\propto \frac{1}{2}I_o \sin^2\theta_2$

$$I_{total\ scattered} = I_x + I_y \propto \frac{1}{2} I_o (\sin^2\theta_1 + \sin^2\theta_2)$$
$$I_s = I_x + I_y \propto \frac{1}{2} I_o (\quad 1 \quad + \cos^2\theta)$$

This describes Intensity in any one direction = solid angle

Field strength = Amplitude per unit area (on a sphere at distance r) decreases as 1/r

Intensity per unit area decreases as $1/r^2$

This is important since measuring scatter from isolated objects the 1/r term is needed to match observed scatter at the measuring device. (Note that for scattering from a crystal the allowed direction of scatter is limited and the measuring device can capture all the light scattered in a particular direction, so there is not a fall-off of intensity just due to radial spread.)

Now we have some of the geometrical terms in the equation for light scattering from particles whose diameter is $<< \lambda$; i.e., where we can treat the particle as a single oscillator.

In practice that diameter is : $d < \lambda/20$

Small Angle X-ray Scattering (SAXS). Solution scattering gets especially interesting when the wavelength of the light is of the same length as features within the tumbling molecule, i.e. just the same relations that make crystal scattering really useful for detailed structure determination. In addition to the Intensity equations above, must be added intensity scattered from features within the molecules averaged by tumbling. Rather than building a model into the image obtained from crystallography, the best that can be done is to propose shapes for the molecule and see if the data is consistent with what would be observed from such shapes tumbling in solution. Of course, if one already knows a great deal about the possible structures that a particular molecule might have, like opening/closing between subunits, or donuts around DNA, then the models can be quite detailed. Later, we will derive the general scattering equation - though the crystallographer does not have to describe the tumble averaging.

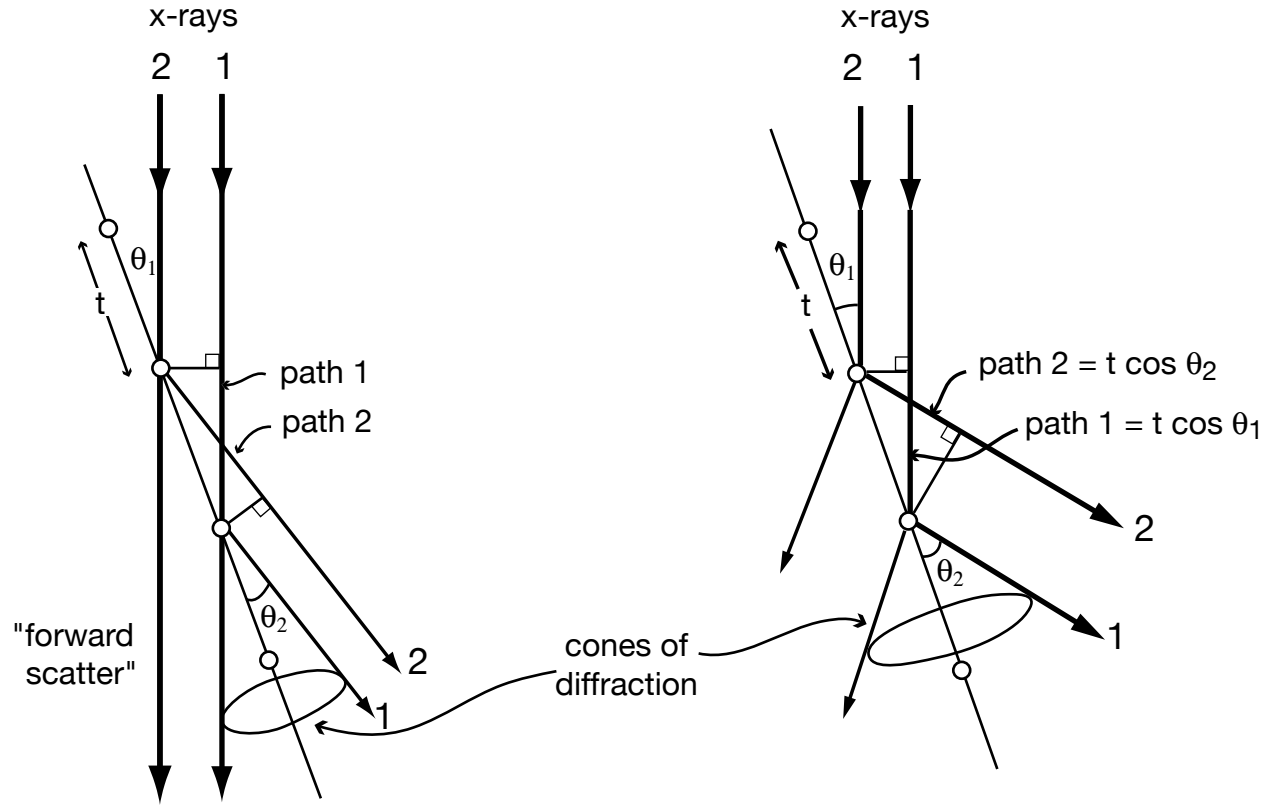
Bragg's Law --
How Crystals Affect the Scattering of X-Rays

The following is true and useful for all light (and electrons (EM) and neutrons). But, the development we will take will direct us toward 3-dimensional crystal diffraction of x-rays with only passing comments about fiber diffraction.

Now we consider the result of taking our scattering unit (an atom, or a molecule) and repeating it regularly in space. We must use our methods of combining waves to appreciate how the scattered light from multiple points combines to form a resultant wave. (Refer to the earlier section: "Vector Notation for Waves".) A wave front can be considered to be made up of component individual rays. (Later, there is a diagram that shows how diffraction occurs when a wave front encounters an array of scatterers.)

Diffraction from a row of molecules

The incoming ray is in phase with itself, and in general the diffracted ray will be in phase with itself if all of the component rays' travel paths differ in length only by an integral number of wavelengths.



If $\theta_1 = \theta_2$, then path difference is zero.

Zero order diffraction (zero path length difference) is independent of d.

If $d\cos(\theta_1) - d\cos(\theta_2) = n\lambda$, then scattered rays are in phase.

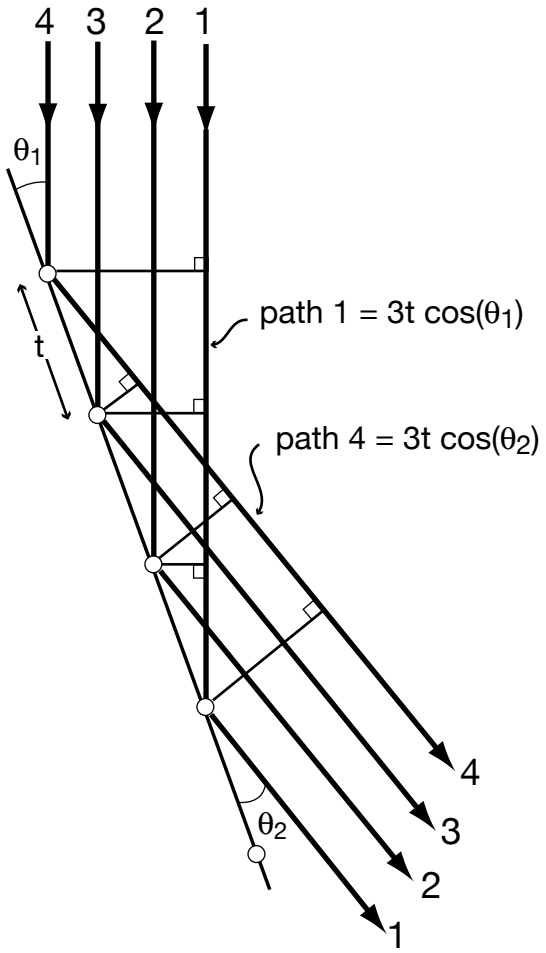
Higher order diffraction (finite path length difference) is dependent on d.

Note that these angle relationships hold all around the cone; thus diffraction from a row of atoms gives cones of scattered x-rays.

Sharpness of Diffraction Maxima

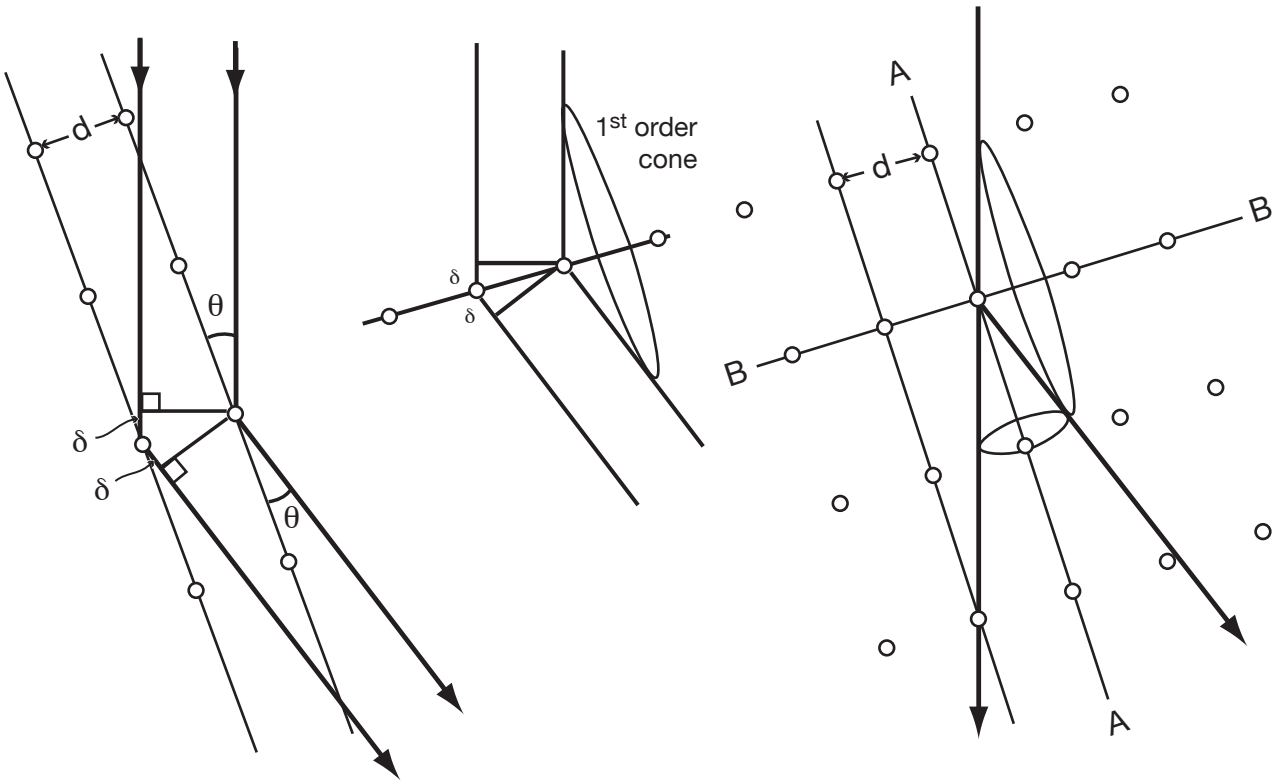
So far we have only considered conditions for diffraction maxima (where scattered rays are exactly in phase.) For a row of 2 or 3 atoms, if the path lengths are just slightly wrong then the intensity of the diffracted beam is just slightly less than maximum, so the distribution of scattered x-rays is rather smooth.

But consider the case of zero order diffraction from a very long row of atoms:
If θ_1, is just slightly smaller than θ_2, then path 2 is shorter than path 1 by $\Delta = t\cos(\theta_1) - t\cos(\theta_2)$.
Likewise, path 3 is shorter than path 1 by $2\Delta = 2t\cos(\theta_1) - 2t\cos(\theta_2)$, and
path 4 by $3\Delta = 3t\cos(\theta_1) - 3t\cos(\theta_2)$, and so on.
If ray 2 was only 1˚ out of phase with ray 1, then ray 181 will be 180˚ out of phase with ray 1; furthermore, ray 182 will be 180˚ out of phase with ray 2, ..., and they will all cancel each other out.



For a real crystal with 10^4 to 10^6 atoms in a row, cancellation is complete for even slightly incorrect angles, giving extremely sharp diffraction maxima (spots.) In fact, the spots on an x-ray photograph are somewhat smeared out because of the size of the x-ray source, the size of the crystal, the mosaic character of the crystal, and the spectral dispersion of the x-rays; but in spite of all that, the spots are still quite small and sharp.

Diffraction from a Two-Dimensional Grid of Atoms



Path difference between ray 1 and ray 2 is 2δ. If 2δ = nλ, then rays will be in phase.

$$\delta = d \sin(\theta), \text{ so}$$

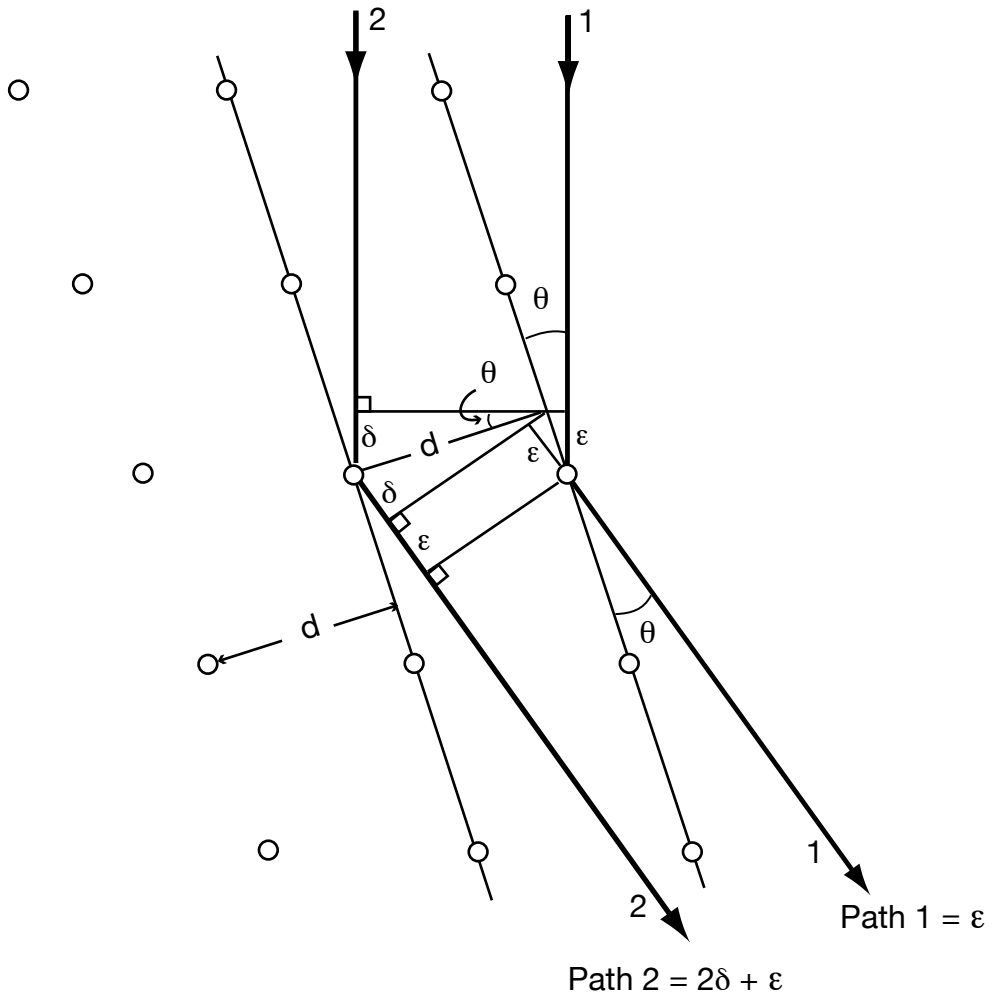
Bragg's Law

$$\lambda = 2(d/n) \sin(\theta)$$

Where d is the distance between repeating units measured perpendicular to the A-A rows (planes in 3D), and d/n is the distance between Bragg Planes which cut that d into n equal intervals.

This can also be thought of as the intersection of a zero-order diffraction cone from row A-A with a higher-order diffraction cone from row B-B. (This generalizes to the third dimension.) Intersections of cones give straight lines, so that diffraction from a grid is in discrete rays

Consider the case where the grid of atoms is non-orthogonal:



Distances ε are all equal, and the path difference between rays 1 and 2 is still 2δ = 2d sin θ

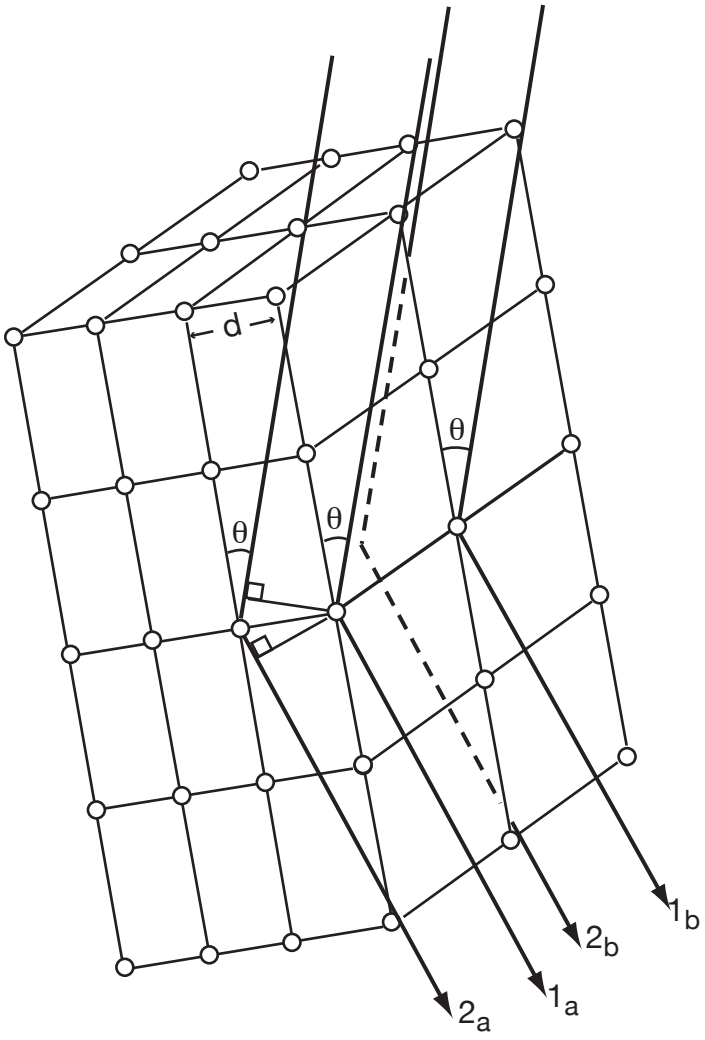
In the general case, therefore, **the d in Bragg's Law is the perpendicular <u>distance between planes</u>, not the distance between atoms.**

Thus for diffraction in a particular direction (defined by the normal to the "Bragg planes"), it is the <u>projected distance between</u> scattering atoms along that direction, not the raw distance between atoms, that determines the relative phases.

For instance, the diffraction pattern from the so-called "alpha" form of some proteins implied regular spacing between residues as well as a helical arrangement. However, polypeptide models that arranged the atoms directly in line along the helix axis failed to explain the details of the diffraction pattern. Linus Pauling proposed a compact model that did not have the atoms in line along the axial direction and he pointed out that this would still diffract! With this in mind, Max Perutz showed that his hemoglobin crystals contained α-helices -- which in turn comfirmed Pauling's model. (But they did not then have enough information to know whether it was right-handed or left-handed!)

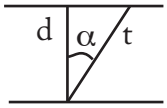
Diffraction from 3-D array

If the x-rays are in the plane of the paper in the left-hand diagram on page Bragg-3, then the condition can be generalized to three dimensions by visualizing a set of identical diagrams parallel to the paper above and below. Rays scattered from all the atoms in any one diagram have been shown to be in phase, and since the path lengths are identical for any two corresponding atoms on different diagrams, all the rays from the entire three-dimensional array are in phase when Bragg's Law is satisfied.



Restricted ray from single Bragg plane: direction limited and $\theta_i = \theta_r$, i.e mirror reflection; then deeper layers restrict θ.

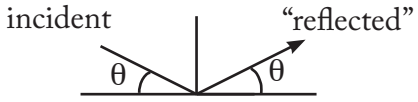
Bragg planes are $\perp$ to "d"; if t of cell at angle α then:



$d = t \cos \alpha$

This may look rather special but when diffraction conditions are satisfied one can always describe these conditions in terms of "Bragg planes".

3-D lattice constrains diffraction to be in single beams, so can always draw diffraction on a 2-D diagram:



A diffracted "ray" is produced in the direction of the arrows, as a beam whose cross-sectional area is approximately the size of the crystal as seen from that direction.

The lattice limits when diffraction can occur, however, the relative intensity is determined by the relative intensity scattered from the molecule in that particular direction.

DEFINITIONS

Scattering:

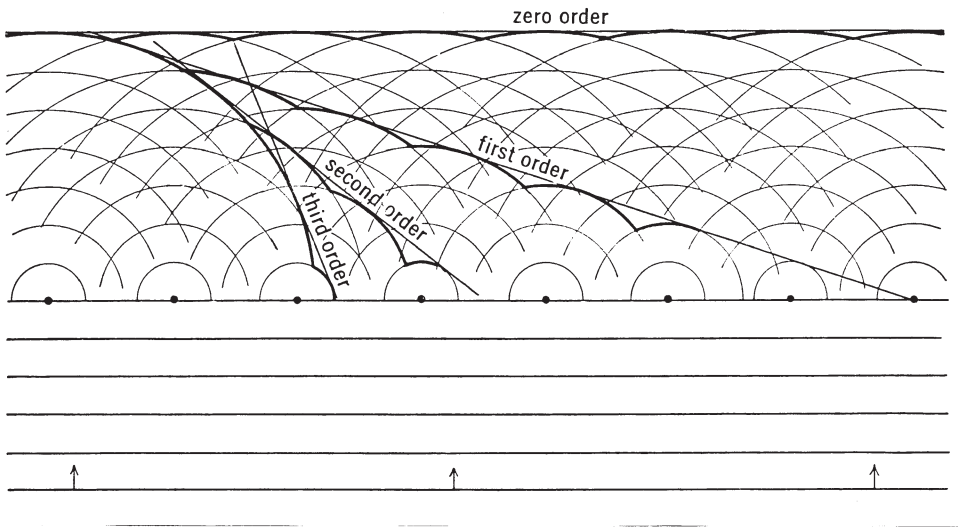
Electrons in the path of an x-ray wave are set into forced vibrations by the periodically changing electric field of the x-ray wave passing by. These oscillating electrons are themselves sources of x-ray waves. This forced oscillation is of the same frequency as the incident x-ray wave, and the emitted x-ray waves are thus of this same frequency. By this interaction the electrons are said to scatter the original x-ray wave.

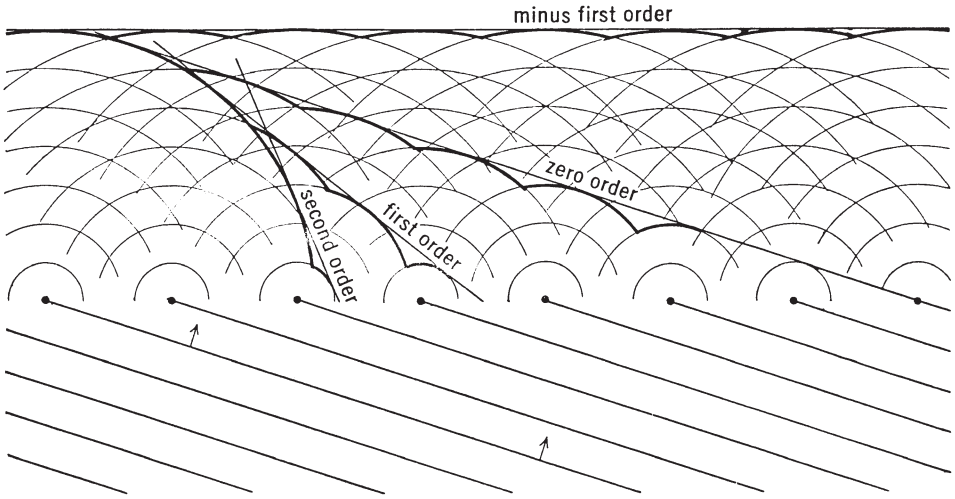
Diffraction:

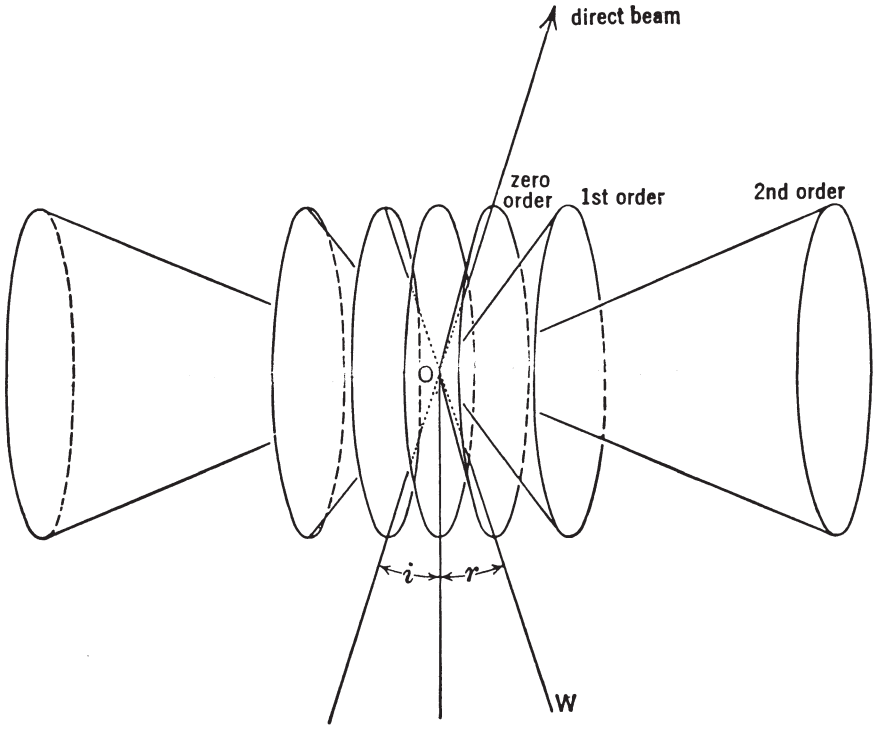
Diffraction is cooperative combination of scattered waves. This can occur if the scattering points are arranged in space in some regularly repeating manner. In certain directions with respect to the incident wave and the array of scatterers, the scattered waves will combine constructively

This can be used to increase the amount of scattered radiation to a point where meaningful measurements can be made. This application is used in crystallography; for, in our case, the study of "globular" proteins and nucleic-acids.

Conversely, the existence of a diffraction pattern and its characteristics can be used to detect and deduce something about a repeating structure. This aspect is exploited in the case of fibers, fibrous structures, and other extended polymers..


zero order
first order
second order
third order

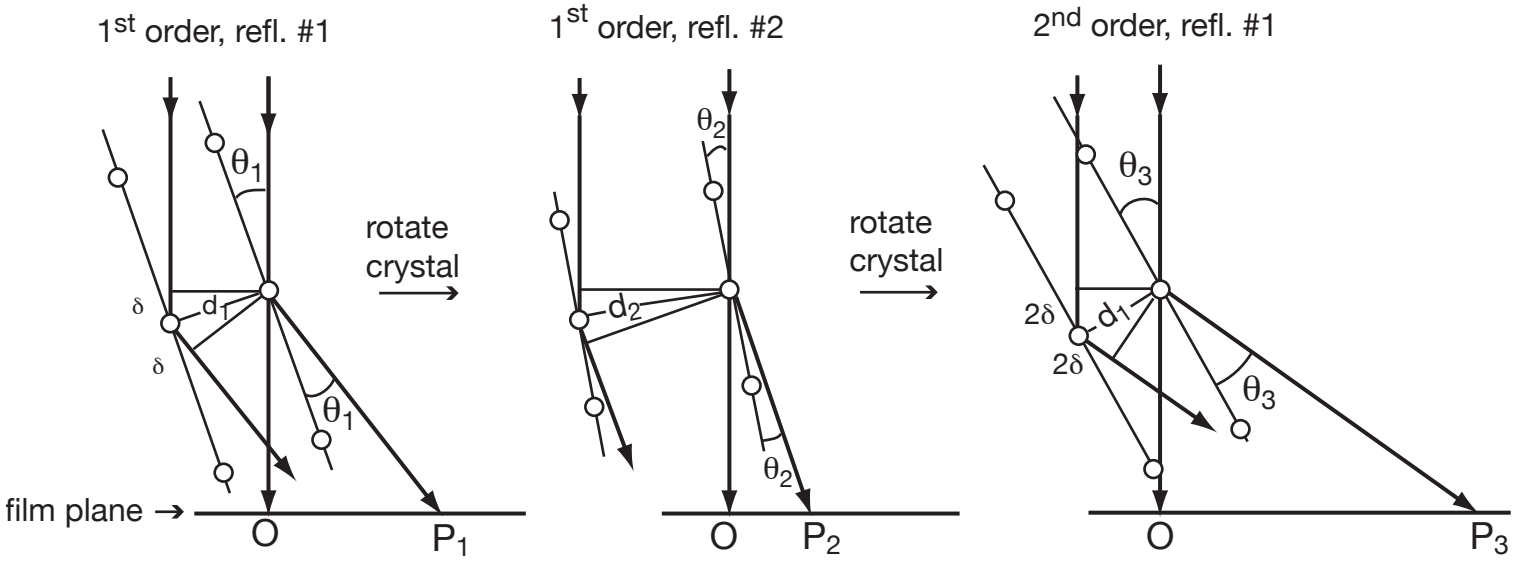

minus first order
zero order
first order
second order


direct beam
zero order
1st order
2nd order
O
i
r
W

The Appearance of the Diffraction Pattern

The 2D crystal shown here has repeats t_1 and t_2 at right angles to each other. As we saw earlier, this perpendicular relationship is not necessary, but sure makes it easier to think about. In any case, d_1 and d_2 are the perpendicular distances between the rows. t_1 and t_2 define the repeating unit of the crystal, a.k.a. the unit cell. In these examples where the Bragg Planes run along the unit cell edges, the Bragg Planes of 1st order reflections span the unit cell, while the Bragg Planes of 2nd order reflections cut the unit cell in half (see next page).

1st order: d_1 = full t_1 of repeat $d_2 = t_2$ 2nd order: $d_3 = (t_1)/2 = (d_1)/2$



$\lambda = 2(d_1)\sin(\theta_1)$ $\lambda = 2(d_2)\sin(\theta_2)$ $\lambda = 2(d_3)\sin(\theta_3) = 2(d_1/2)\sin(\theta_3)$

If a piece of photographic film is put in place to intercept the x-rays, then it gets a mark at "0" from the direct beam and a spot at "P" from the diffracted (reflected) ray.

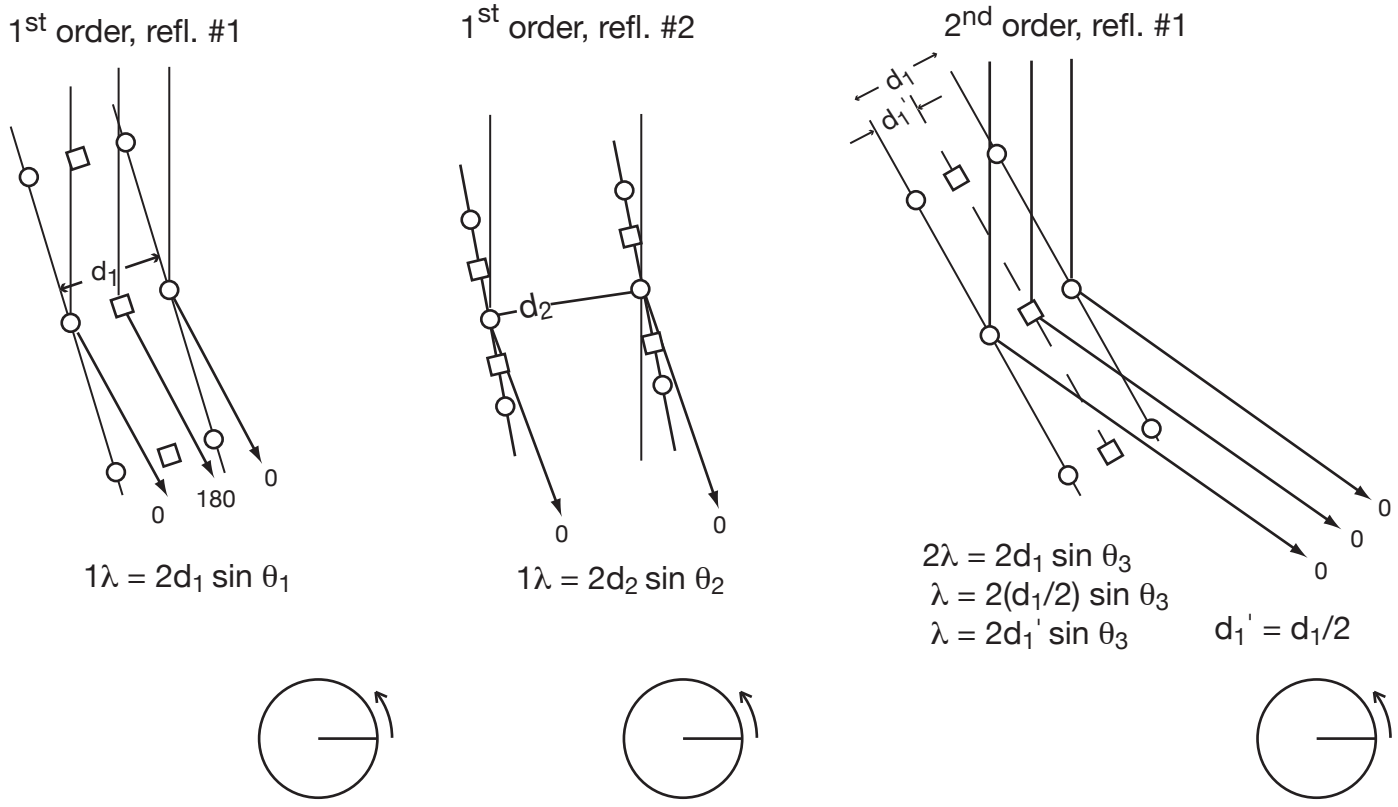
If $d_2 > d_1$ then $\theta_2 < \theta_1$ for diffraction to occur.

Notice that the distance O-P is approximately proportional to θ and inversely proportional to d. Spots far out from the center of the film come from planes that are quite close together in the crystal.

This reciprocal relationship between the spacings d in the crystal and the distances O-P in the diffraction pattern is the origin of the term "reciprocal space" used to describe the diffraction pattern.

Diffraction as a probe for spacings of d within a crystal

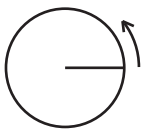
Diagrams are those of the previous page but with a square atom halfway between the round atoms that are spaced at distance d_1



1st order reflection #1 is weak because for that solution of the Bragg equation - with d_1 being the distance between planes of atoms with 1 λ scattering path difference - the square atoms scatter 180° out of phase from the round ones.

1st order reflection #2 is strong because for that solution of the Bragg equation - with d_2 being the distance between planes of atoms with 1 λ scattering path difference - the square atoms scatter in phase with the round ones, since they are both on the same planes.

2nd order reflection #1 is strong because for that solution of the Bragg equation - with d_1' being the distance between planes of atoms with 1 λ scattering path difference = the square atoms scatter in phase from the round ones. (The unique index of this reflection implies d_1')



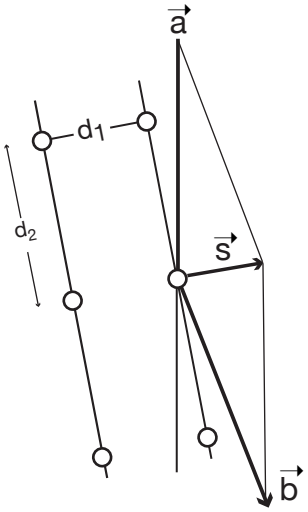
As one travels along the normal direction from plane to plane, there is one phase clock revolution for each λ scattering path difference.

The Diffraction Vector $\lambda = 2d \sin\theta$

Refer to Molecular Scatter chapter for a general description of scattering.
Diagrams: same examples as drawn earlier for Bragg Diffraction, but now showing the diffraction vector which is perpendicular to Bragg Planes and of a length proportional to the reciprocal distance between Bragg Planes.

1st order reflection
Bragg Plane d = d_1

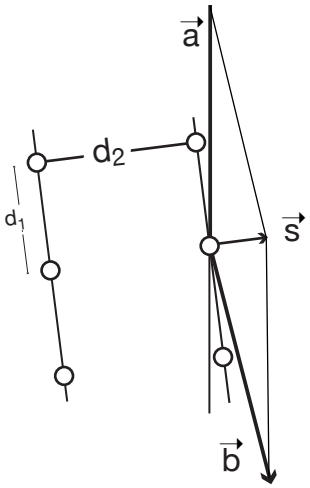
1st order, refl. #1



$1\lambda = 2d_1 \sin\theta_1$

1st order reflection
Bragg Plane d = d_1

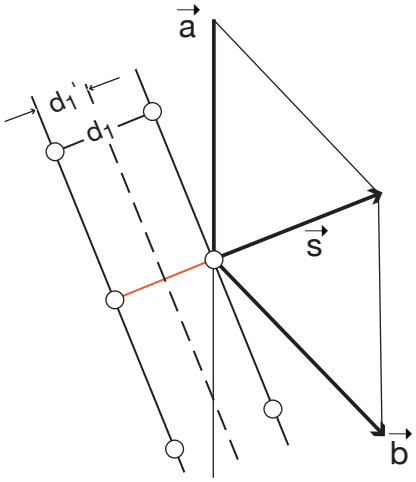
1st order, refl. #2



$1\lambda = 2d_2 \sin\theta_2$

2nd order reflection
Bragg Plane d = $d_1' = d_1/2$

2nd order, refl. #1



$2\lambda = 2d_1 \sin\theta_3$ $\quad d_1' = d_1/2$
$\lambda = 2(d_1/2) \sin\theta_3$
$\lambda = 2d_1' \sin\theta_3$

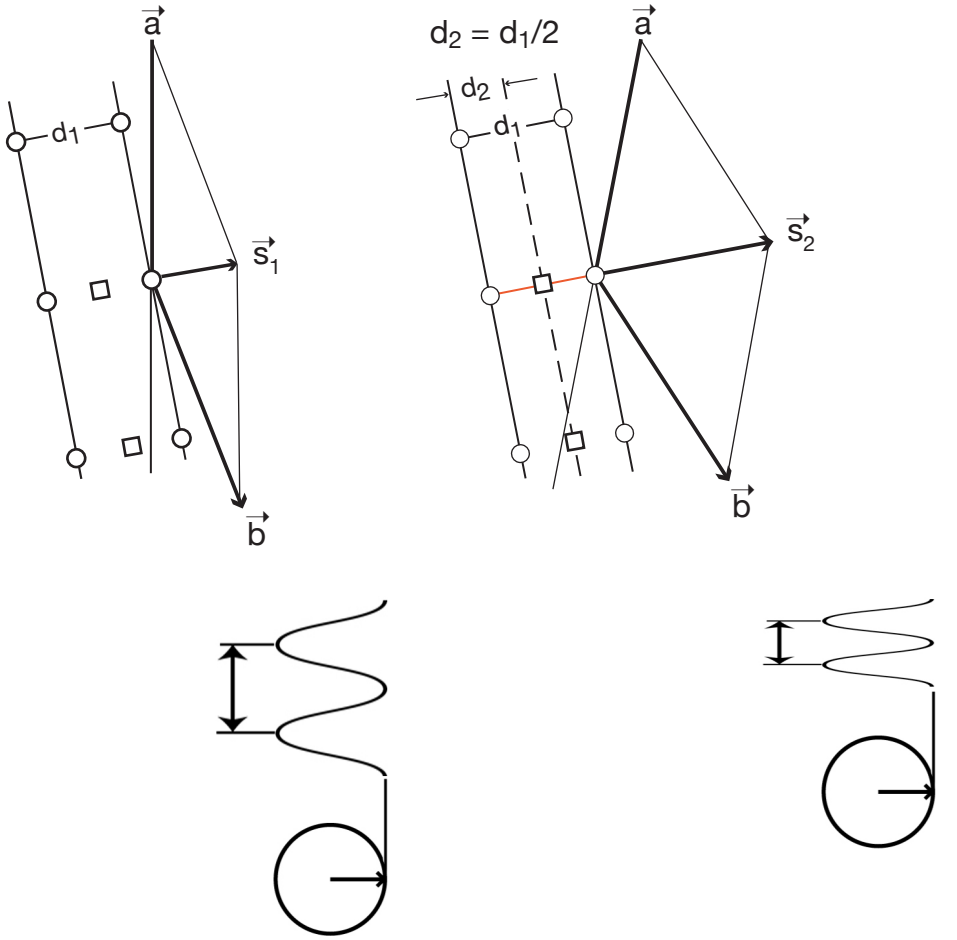
Parallelogram construction: $\vec{s}$ is normal to the Bragg planes (and thus is fixed with respect to the crystal). The end of $\vec{s}$, then, is a point which by its position (direction and distance out) describes a set of Bragg planes. Note that the length of $\vec{s}$ is inversely proportional to the Bragg plane spacing d. The $\vec{s}$ of the third diagram is twice the length of the $\vec{s}$ of the first diagram, and for the third case any point on a plane halfway between the original ones will scatter in phase with points on the original planes.

Every set of Bragg Planes has its own unique diffraction vector. Just as the Bragg Planes divide up the real crystal in a regular manner, all the diffraction vectors describe the crystal. The crystal is a lattice and the ends of the diffraction vectors describe a lattice, the reciprocal lattice.

Draw Diffraction Vectors to show reciprocal relationship

Parallelogram construction: $\vec{s}$ is normal to the Bragg planes (and thus is fixed with respect to the crystal). The end of $\vec{s}$, then, is a point which by its position (direction and distance out) describes a set of Bragg planes. Note that the length of $\vec{s}$ is inversely proportional to the Bragg plane spacing d.

The $\vec{s}$ for half-size spacing is twice the length of the $\vec{s}$ of the single spacing, and for the second case any point on a plane halfway between the original ones will scatter in phase with points on the original planes.



Diffraction vector $\vec{s}_1$ describes a set of Bragg planes spaced by d_1, $\vec{s}_2$ describes a set of Bragg planes spaced by d_2. The diffraction vectors are a way describing both the sets of Bragg planes and the diffracted rays from those planes.

In the case where there is a square atom halfway between round atoms as shown, what is the relative intensities of the diffracted x-rays described by $\vec{s}_1$ and $\vec{s}_2$?

Scale of the diffraction experiment: 1cm = 100,000,000Å

unit cell size (range 50 to 200+) 100Å ,

crystal size (range 0.05 to 1 mm) 0.1 mm = 1,000,000Å

Spot size on detector just a little larger than the crystal cross-section. 1,000,000Å

crystal to detector distance (range 5 to 50 cm) 10 cm = 1,000,000,000 Å

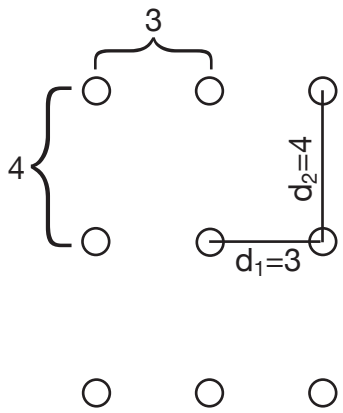
0.1 mm crystal about 10,000 unit cells across. (Note: a crystal is a mosaic of crystal domains)

crystal to dectector in unit-cells: 10,000,000 uc

Unit cell on chalk-board: 1 uc = 1 ft, crystal size 10,000 ft = 2 miles (size of Duke University)

detector distance 10,000,000 ft = 2,000 miles. (somewhat beyond the chalk tray, perhaps in Arizona)

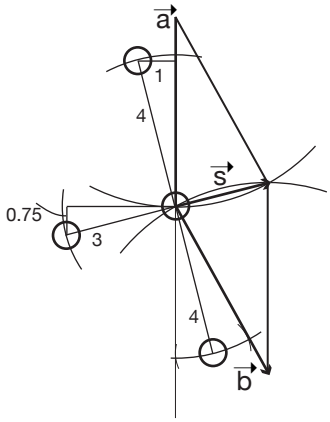
Construction of the Diffraction Vector



Setup:

1. Determine the unit cell dimensions, wavelength of x-rays; scale if needed. Here we have lattice spacings of 3 & 4, and the wavelength is 1.5 - all are in Å. The grid being used to construct the drawings at left has divisions in mm and the scale being used is 5mm = 1Å.
2. Pick one of the lattice points as an origin.
3. Pick an arbitrary, but suitable length/scale for the incident vector $\vec{a}$. The construction here uses $|\vec{a}|$ = 5Å.

Construction:



1. First place the lattice point chosen as origin and draw a of $\vec{a}$ length 5, as a vertical line.

2. Determine the location of the other lattice points given that the distances to neighbors is 4Å and 3Å. This restricts the neighbors to the perimeters of circles of radii 4Å and 3Å. Construct this constraint by drawing two circles (or sectors if you don't want to draw the full circle) centered at the origin.

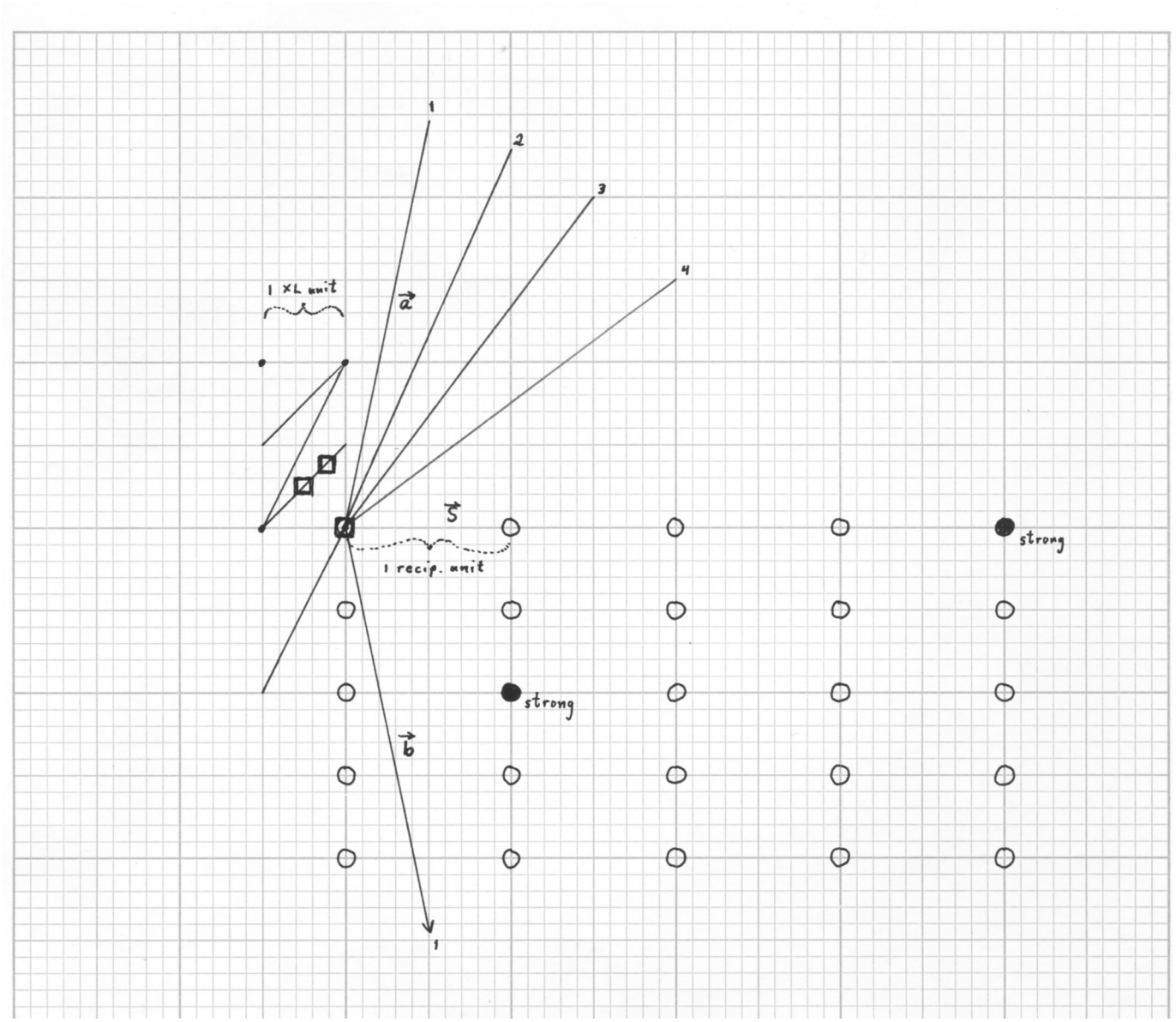
3. Another constraint is Bragg's Law; $\lambda = 2d\sin\theta = 2(t/n)\sin\theta$
   For n=1: $\lambda = 2d\sin\theta$ (d is the full repeat distance t = 3)
   $\lambda/2 = 1.5/2 = 0.75 = d\sin\theta$
   for d=3, $\sin\theta = 0.75/3$ and for d=4, $\sin\theta = 1/4$
   .75 and 1 are just the short sides of triangles of hypotenuse lengths 3 and 4, respectively, the sector radii drawn in 2.

4. Draw three neighbor lattice points by locating the intersection of a chord of length 1 from a (upper and lower two points), and of a chord of length 0.75 from the $\perp$ to a (through the origin point).

5. Construct b. b is at angle 2θ from the straight through of the incident ray. The vector b is the same length as a and is tangent to a circle of radius 1 centered on the lower lattice point. The length of b is the same as a.

6. Construct s. s is located by drawing two circles of radius a, one circle is centered at the base of a, the other at the head of b. The head of s is located at one intersection of these two circles, the base of s is at the other intersection i.e., the origin. So s extends from the origin and is perpendicular to the line of lattice points

The Reciprocal Lattice



$\vec{s}$ is normal to the Bragg planes (and thus is fixed with respect to the crystal). The end of $\vec{s}$, then, is a point which by its position (direction and distance out) describes a set of Bragg planes. Note that the length of $\vec{s}$ is inversely proportional to the Bragg plane spacing d.

Every set of Bragg Planes has its own unique diffraction vector. Just as the Bragg Planes divide up the real crystal in a regular manner, all the diffraction vectors describe the crystal. The crystal is a lattice and the ends of the diffraction vectors describe a lattice, the reciprocal lattice.

Ewald Sphere: Crystal and Reciprocal Lattice

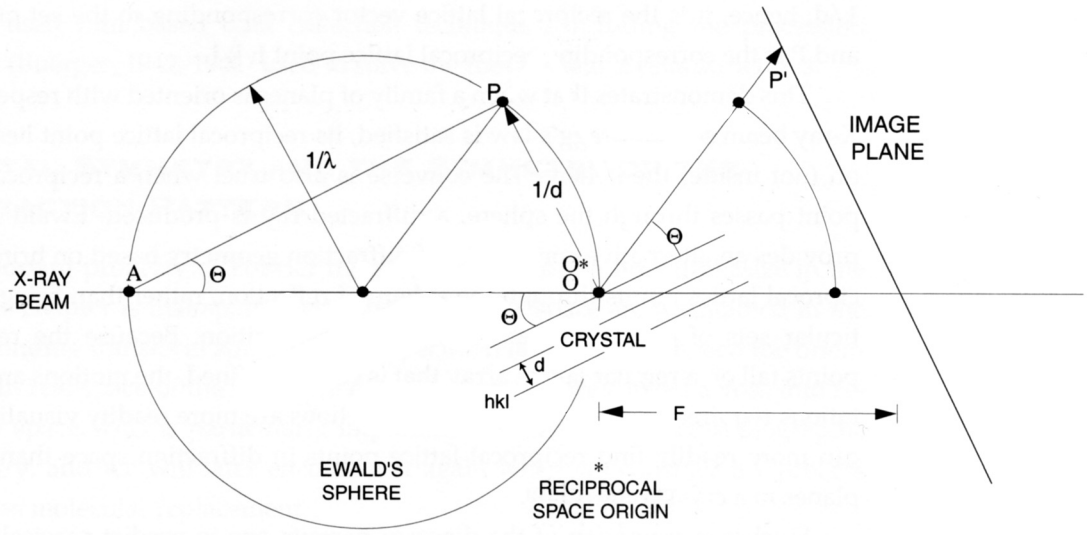
The crystal lattice and the reciprocal lattice are duals. That is, each one describes the other and they are logically linked together. It is convenient to make the origin of the crystal lattice and the origin of the reciprocal lattice to be the same point. Then, during data collection, as the crystal rotates, the reciprocal lattice rotates.

The Ewald sphere helps describes diffracting conditions when the crystal is oriented to the incoming x-ray beam such that particular sets of Bragg Planes can "reflect" the x-rays.

The center of the Ewald sphere is placed along the incoming x-ray beam with the center of the reciprocal lattice on the x-ray beam at the circumference of the Ewald sphere. It is convenient to make the radius of the Ewald sphere be $1/\lambda$. When the crystal is rotated, the reciprocal lattice rotates, and when a reciprocal lattice point is on the surface of the Ewald sphere, the associated Bragg Planes are reflecting. This is just the conditions of our parallelogram construction!

The actual diffracted x-ray goes out from the crystal at an angle θ to the Bragg Planes and at angle 2θ to the direction of the original x-ray beam.

There are several advantages to placing the crystal at the center of reciprocal space on the circumference of the Ewald sphere besides emphasizing that the crystal lattice and the reciprocal lattice are duals that rotate together and are both descriptions of the real crystal.
1) We developed the reciprocal lattice using the parallelogram construction which builds the reciprocal lattice around the crystal as a center.
2) Simultaneous diffraction from a range of wavelengths can be shown without any rescaling.
3) The kinemage showing diffraction conditions with the crystal and the reciprocal lattice locked together around a common center is much easier to make than the case where two rotation centers must be correlated.



HOWEVER: Some textbooks show a different diagram.

e.g. McPherson places the crystal at the center of the Ewald sphere, and the center of the reciprocal lattice on the circumference (where it must be). This does makes drawing the direction of the diffracted ray easier since the x-rays follow a radius of the Ewald sphere from the crystal.

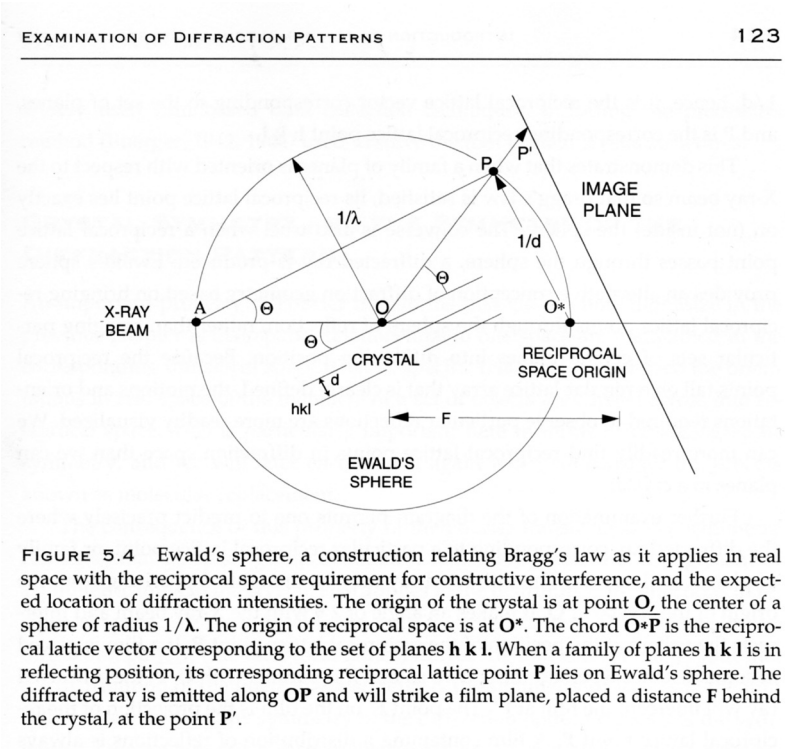
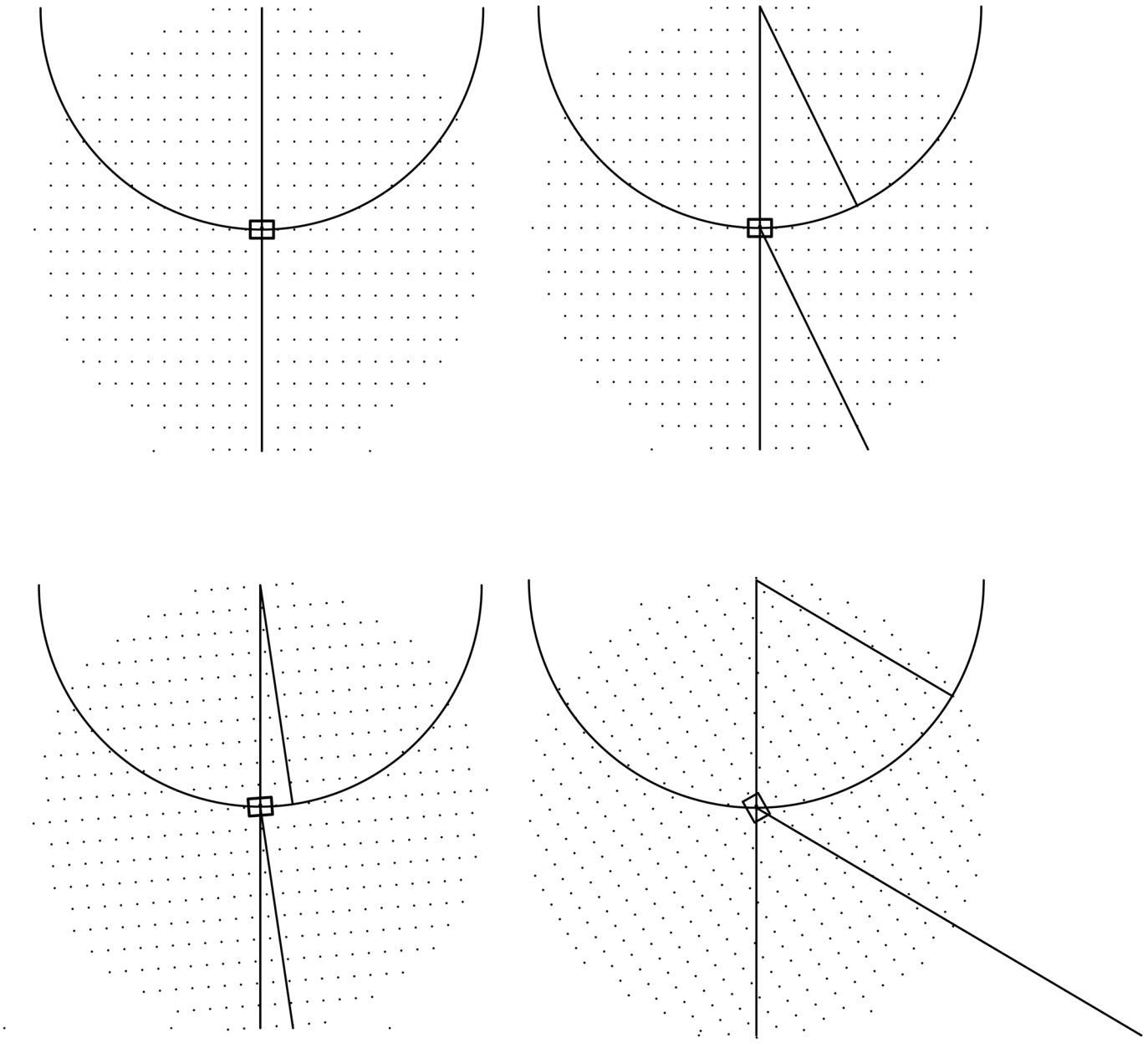


FIGURE 5.4 Ewald's sphere, a construction relating Bragg's law as it applies in real space with the reciprocal space requirement for constructive interference, and the expected location of diffraction intensities. The origin of the crystal is at point **O**, the center of a sphere of radius $1/\lambda$. The origin of reciprocal space is at **O***. The chord $\overline{\mathbf{O^*P}}$ is the reciprocal lattice vector corresponding to the set of planes **h k l.** When a family of planes **h k l** is in reflecting position, its corresponding reciprocal lattice point **P** lies on Ewald's sphere. The diffracted ray is emitted along **OP** and will strike a film plane, placed a distance **F** behind the crystal, at the point **P′**.

ref: Alexander McPherson "Introduction to Macromolecular Crystallography"

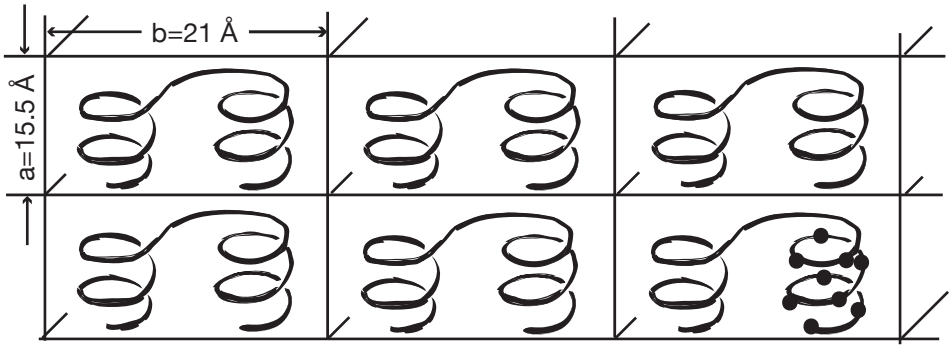
Ewald Sphere: Set up for HelixBearHair.kin

Diffraction from a crystal of molecules described by how the Ewald-sphere (the generalization of all possible paralellogram constructions) intersets the reciprocal lattice (the diffraction-vector description of all Bragg Planes within the crystal.)



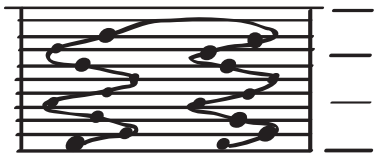
Crystal of Molecules

Consider a crystal of molecules, each with two α-helical regions.

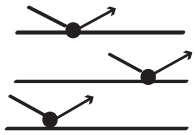


α-helix:
side-to-side packing ≈ 10 Å
3.6 residues / turn
1.5 Å rise / residue
5.4 Å rise / turn

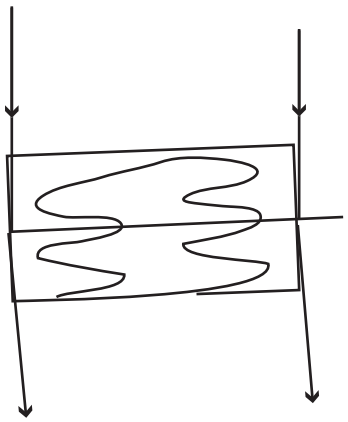
View ⊥ to ab planes



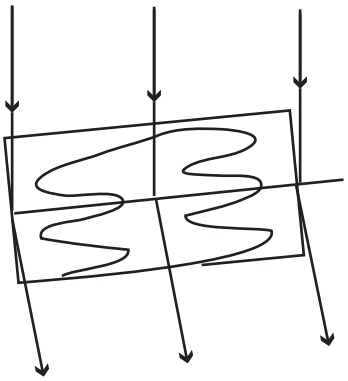
10, 0, 0 reflection



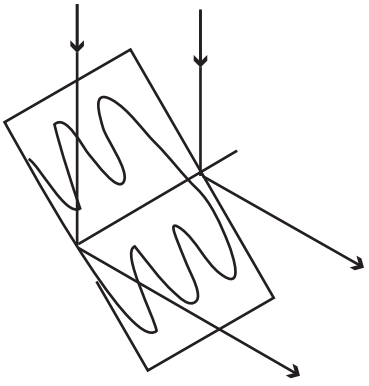
Lines mark equi-phase planes. Scattering from, e.g. alpha carbons, nearly in phase over all the unit cell (of course, exactly in phase with corresponding one in next unit cell).



0, 1, 0 weak



0, 2, 0 strong



10, 0, 0 strong

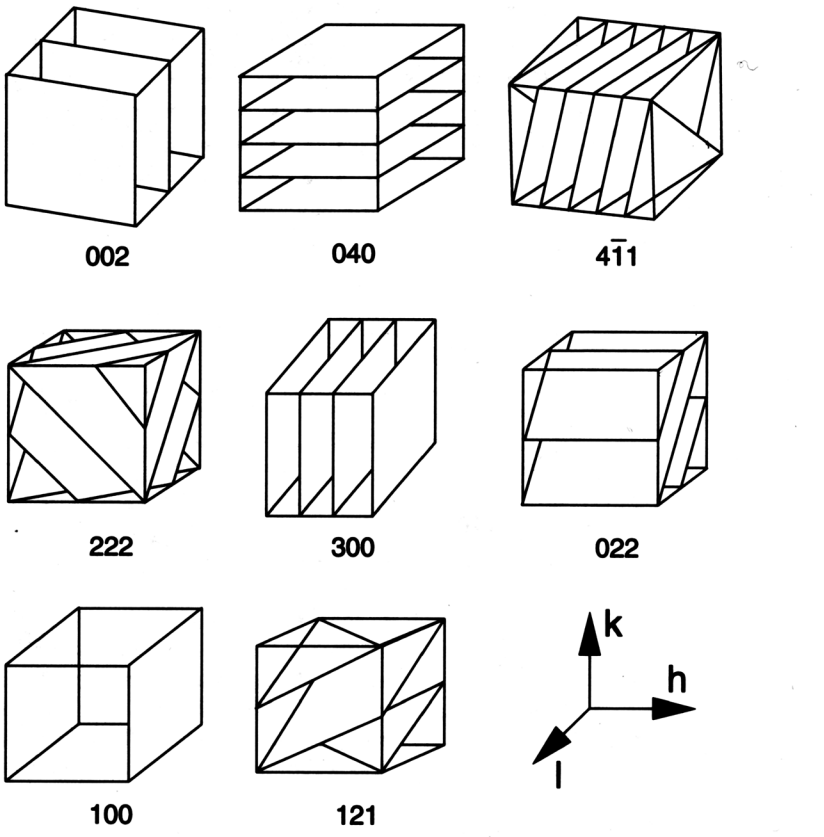


FIGURE 2.17 Families of planes making rational intercepts with the unit cell edges are identified by a set of three integers, **h k l,** known as Miller indices. The values of these indices are equal to the number of times the family intercepts each unit cell edge. Each family of planes has associated with it a vector called the reciprocal lattice vector, which is normal to the planes and has a length inversely proportional to the interplanar spacings. The various families of planes may be considered to be the spectral or harmonic components of the electron density in the unit cell, each having a wavelength equal to the interplanar spacing, or a frequency equal to its reciprocal. By mathematical recombination of many sets of planes, each having a characteristic frequency, the electron density of the cell can be reconstructed. It is these families of planes that give rise, by constructive interference, to the maxima that comprise the X-ray diffraction pattern.

Concepts: Reciprocal Lattice vs Bragg Planes

We go on to further development of diffraction from a lattice by using the parallelogramconstruction to show the relationship of diffraction vectors to specific distances and directions in the real crystal. Two concepts emerge from this approach:

One concept is that the reciprocal lattice is a description of the Bragg planes within the real space lattice of the crystal. The reciprocal lattice points are "named" (indexed) by the same "names" (indices) that describe the Bragg planes. The reciprocal lattice points are in a direction from the origin that is normal to their corresponding Bragg planes, and at a distance inversely proportional to the "d" spacing (the perpendicular displacement) of those Bragg Planes. These indices are integers, and describe how the Bragg planes divide the Unit Cells of the crystal. Each reciprocal lattice point corresponds to a possible diffracted ray from the crystal -- it is the parallelogram construction that illustrates that correspondence -- and the **Ewald sphere** construction (describes when diffraction will occur from a set of Bragg planes) is a general summary of all the possible parallelogram constructions.

Another concept is that a value can be assigned to each reciprocal lattice point that is the relative intensity of the corresponding diffracted ray. Of course, each diffracted ray also has a relative phase which can be assigned to its reciprocal lattice point, but we can not measure that phase directly. The intensity of a diffracted ray is determined by the relative offsets of the scattering points, the atoms, from the diffracting Bragg plane. We will explore that by filling in a few atoms in a crystal and seeing how their perpendicular position with respect to the Bragg planes affects the phase of their scattering in the direction of the diffracted ray, and consequently, affects the intensity of that diffracted ray. This phase as a function of the distance between Bragg planes can be described as "clock" turns around the circle that describes the magnitude and phase of the wave from an atom.

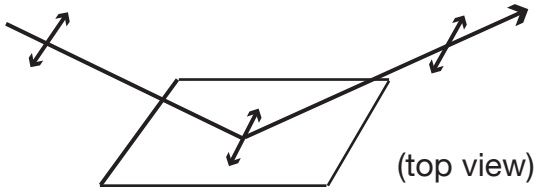
Bragg Plane definition:
Bragg Planes for crystals of few atoms (and as W. L. Bragg seems to have first described them) are conviently thought of as the planes of atoms making up the crystal. However, as you have seen already, sometimes only some atoms line up on the Bragg planes even for relatively simple molecules, and for proteins few if any atoms are actually exactly on a given Bragg plane.

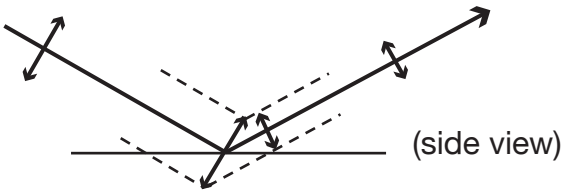
The more robust way of describing Bragg Planes is in terms of the unit cell, however you happen to define it. Then the Bragg Planes cut the unit cell edges into integer fractions, and the strength of diffraction from a Bragg Plane set has to do with the spacing between atoms in the direction of that Bragg Plane normal where atoms separated by that "d" scatter in phase with each other.

The **Unit Cell** is the small parallelepiped built upon the three translations selected as unit translations. The unit cell repeats by those 3 translations to fill all space.

Geometrical Factors that Affect Measured Intensities

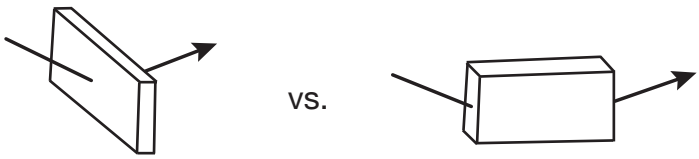
1) Polarization of the reflected x-rays:
The component of the incoming oscillation that is parallel to the plane of reflection is undiminished. But the perpendicular component is only "seen" by the reflected ray in projection (since the wave cannot oscillate parallel to its direction of travel;) thus some energy is lost, and the amount of intensity lost is a function of θ.





2) Lorentz factor:
To collect the full scattering pattern the crystal must be moved. If, because of this motion, some reflections get more time than others their relative measured intensities will be increased. This is known as the Lorentz factor and is a function of almost every variable involved in the motion, including θ.

3) Absorption:
Matter (electrons) absorb x-rays as well as scattering them. If the crystal is not a sphere some paths through it are longer than others and will absorb relatively more of the x-ray beam. This absorption correction is sometimes measured empirically and sometimes calculated geometrically from the known shape of the crystal and its position during each reflection.



None of these geometrical effects contains any information we are interested in: they are simply factors we must correct for before using the intensities to find out about the structure of our protein molecule.

Size of the SPOT

4 main factors contribute the the actual size of the observed spot that the diffracted ray makes on the detection device (e.g. film).
1) Size of the crystal.
2) Mosaic spread of the crystal.
3) Size of the x-ray source (presume each point of the effective source radiates in all directions).
4) Wavelength dispersion of the source.

This presumes that the crystal is rotated such that all possible points in the crystal have opportunity to diffract.

In addition, two other factors that are part of the experimental setup will have an effect on spot size.
5) How parallel the rays are coming from the source.
6) How thick the detector is and the angle the x-rays hit it. For instance, a multi-wire area-sensitive detector has to have a finite depth in which to catch the x-ray photons -- if the x-ray beam comes in at an angle, then there is a latteral uncertainty as to when the photon ionizes the detector gas.

Size of Mosaic Blocks

The trick is for a crystal to be perfectly-imperfect for the wave-length/atom-types such that the mosaic blocks are large enough to make sharp diffraction rays, yet small enough that the chance of a scattered ray to diffract again within the mosaic block is relatively small.

Effective interaction probability, i.e. efficiency of an X-ray interacting with electrons in a crystal according to James p 53 is 10^{-4} for a "strong" reflection (perhaps from NaCl, which I suppose would interact more strongly with xrays than CNO molecules. Apparently a multiplier to the atomic scattering factor.

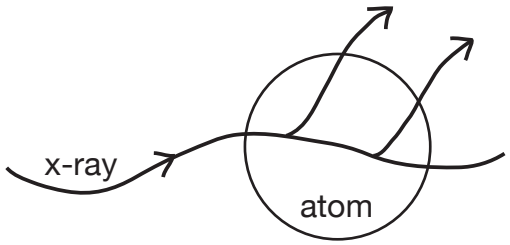
REALLY NEED A DRAWING OF MOSAIC CRYSTAL...

google mosaic block protein crystal

need IUCR code to download pdf's Albrecht Messerschmidt X-ray Crystallography of Biomolecules page 77: mosaic block of size ca. 0.1 micrometer with average tilt angle 0.1 - 0.5 degrees for protein crystals. 0.1u = 0.0001 mm= 0.00001 cm = 1000Å

How an Individual Atom Affects the Structure Factor

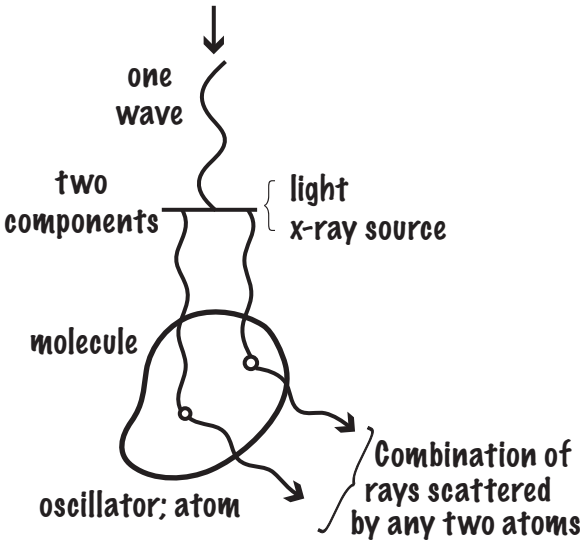
1) f, the atomic scattering factor:
Atoms (i.e., the distribution of electrons) have a real, finite size, and so it is possible for various parts of an atom to scatter out of phase with one another. The same sort of path-Iength arguments apply as when we considered two separate atoms scattering out of phase. But for a given atom, the electron distribution is approximately spherically symmetric, and the contribution of this factor can be calculated as a function of θ or looked up in tables.

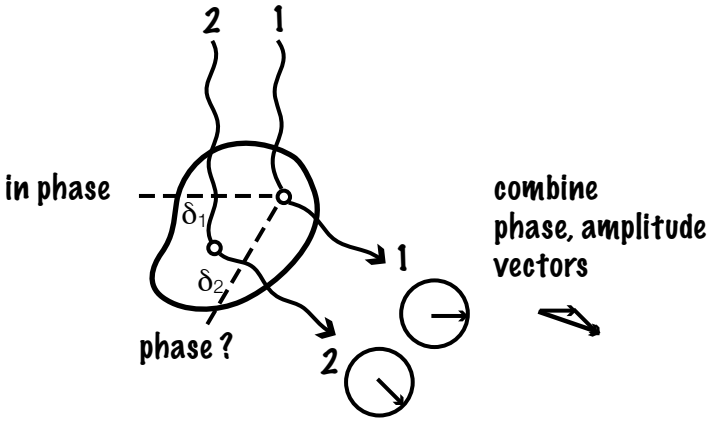


2) the B factor :
All uncertainties in the position of an atom is put into the B factor. Thermal motion is only a minor part of the B-factor for atoms of a macromolecule. Atoms seem to vibrate, so their effective scattering is spread over a larger volume. (It may not be a spherically symmetric volume, such as when vibration is along the direction of a bond.) This effect is also a function of θ. The most common formula (for the spherically symmetric case) is:

$$f_{n,\theta}\, e^{-B(\sin^2\theta)/\lambda^2}$$

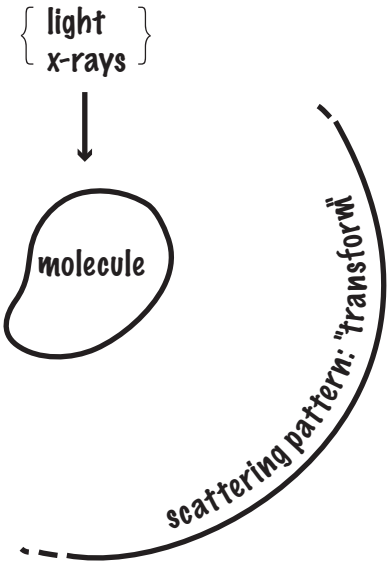
Scattering of | Light / X-Rays | by a Molecule





The combination of rays scattered by any two (or more) atoms depends on the amplitude and phase of each scattered ray. The amplitude depends on the [oscillator strength | number of electrons] in that atom (since it is electrons that scatter x-rays). The relative phase of the rays depends on the relative path lengths they have travelled. Wave 2 travels $(\delta_1 + \delta_2)$ further than wave 1, so the phase change in that length is $(\delta_1 + \delta_2)/\lambda$.

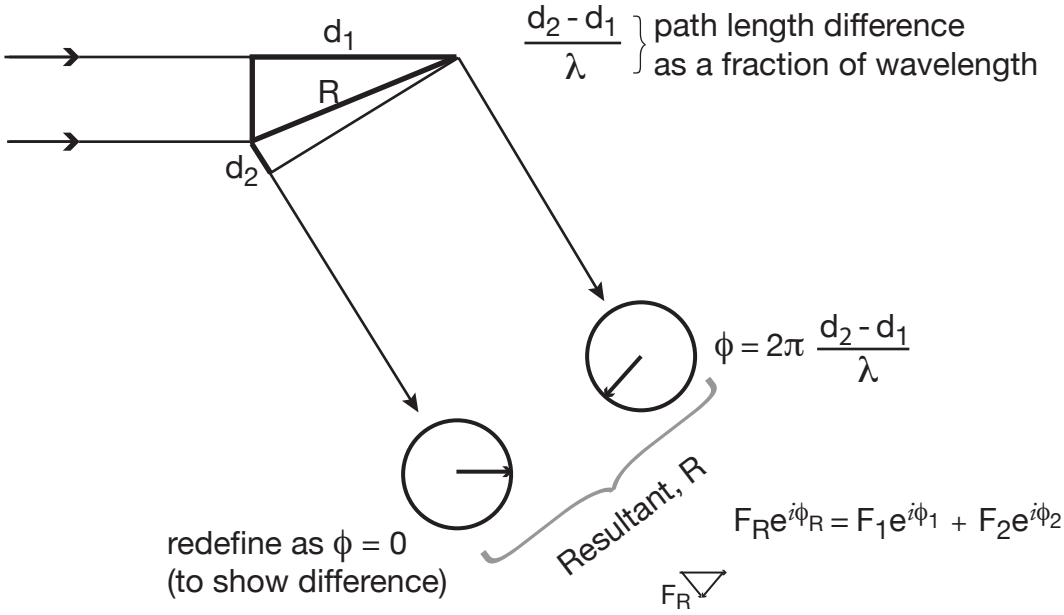
The scattering pattern is distributed in all directions, and varies fairly smoothly.



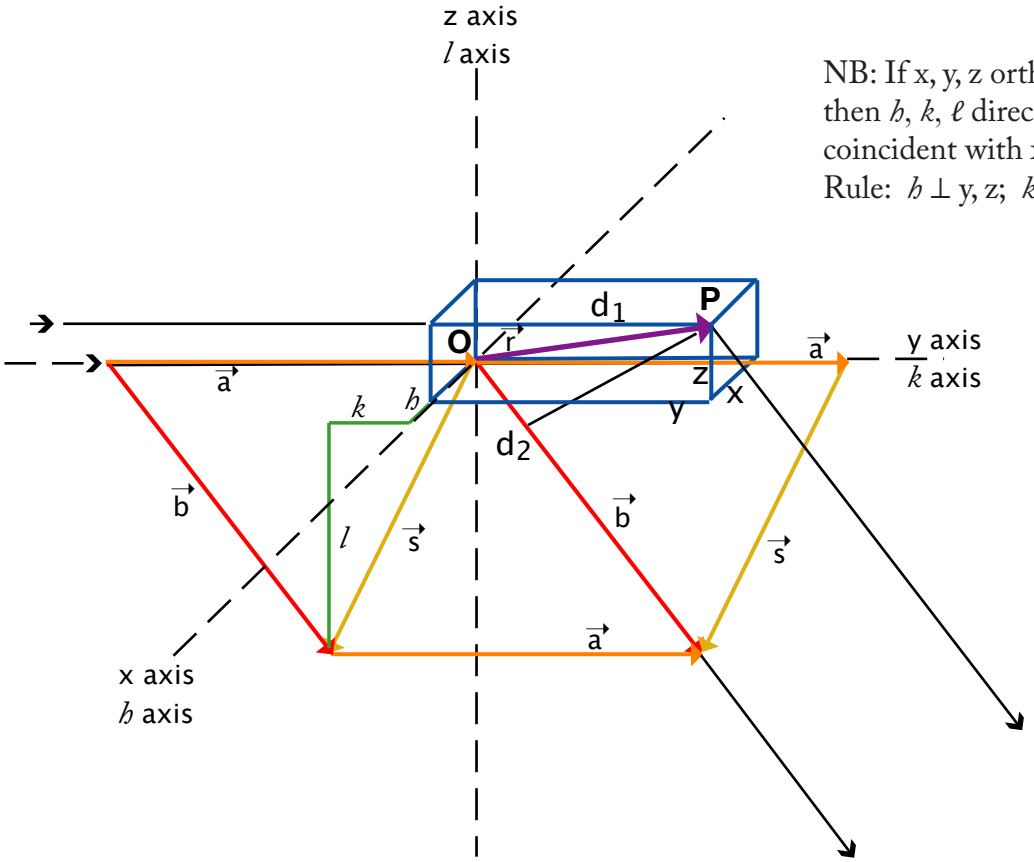
Intensity in any particular direction is a function of the [strength of the oscillators | number of electrons in the atoms] and of their relative positions. So if the molecules were randomly positioned, e.g. by free rotation, one could only find out some general property concerning distance such as how big the diameter (radius of gyration) is of the region where the oscillators are distributed. e.g. low angle x-ray scattering, light scattering experiments. This is of mathematical interest only, unless we can relate the pattern seen (or imagined) here to the pattern of structure in the molecule.

For x-rays the scattered ray has 180° phase shift from the incident ray - but this is true for all scattered rays so it makes no difference to how they combine.

Scattering from 2 oscillators in a plane



Scattering from Multiple Oscillators



- Point **P** is an oscillator
- **O** is an arbitrary origin, which may or may not be an oscillator
- $\vec{r}$ is the vector from **O** to **P**, with components x, y, z
- $\vec{a}$ and $\vec{b}$ are standard vectors in the directions of the incident and scattered rays: $\vec{s} = \vec{b} - \vec{a}$
- $\vec{a}, \vec{b}$ scaled from unit vectors by $1/\lambda$; so all these related dimensions will be in $Å^{-1}$ and express distance as fractions of wavelength.

Path length difference $\delta = d_2 - d_1$ where,

d_1 is projection of $\vec{r}$ on $\vec{a}$ direction : $d_1 = \vec{r} \bullet \vec{a}$

d_2 is projection of $\vec{r}$ on $\vec{b}$ direction : $d_2 = \vec{r} \bullet \vec{b}$

So, $\delta = d_2 - d_1 = \vec{r} \bullet \vec{b} - \vec{r} \bullet \vec{a} = \vec{r} \bullet (\vec{b} - \vec{a}) = \vec{r} \bullet \vec{s}$

d_1 and d_2 are unitless as is δ. i.e. $\delta = \delta'/\lambda$ where δ' is the real measurement in Å and λ is wavelength in Å so δ is unitless fractional length in terms of wavelength.

$$\delta = \underbrace{\vec{r}}_{x,\,y,\,z} \bullet \underbrace{\vec{s}}_{h,\,k,\,\ell}$$

components: x, y, z h, k, ℓ $\xrightarrow{\text{vector multiplication}}$ $(hx + ky + \ell z)$

δ is dimensionless and x, y, z are in Å; h, k, ℓ are therefore in $Å^{-1}$ (reciprocal Å): they give us a useful way of describing a scattered ray.

Relative phase $\phi = 2\pi (hx + ky + \ell z)$ as in $|F|e^{i\phi} = |F|\, e^{i2\pi(hx + ky + \ell z)}$, expression for a light ray scattered from P.

In this general case h, k, ℓ are not constrained to be integers.

More wrestling with units, dimensions:

We need $\delta = \vec{r} \cdot \vec{s}$ as unitless fractions of wavelength,
so components x,y,z h,k,ℓ must multiply out to cancel units:

if x, y, z in Å, then h, k, ℓ in reciprocal Å, and
if x, y, z in fractions of something, then h, k, ℓ in related reciprocal fractions.

Note that the equations that we use imply repeating functions:

$$F_{hk\ell}\, e^{i\,\phi_{hk\ell}} = \sum_x \sum_y \sum_z \rho_{xyz}\, e^{i2\pi(\,hx + ky + \ell z\,)}$$

equation of a wave

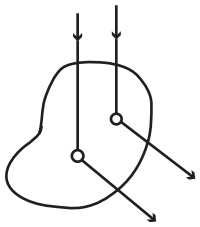
$$\rho_{xyz} = \sum_h \sum_k \sum_\ell F_{hk\ell}\, e^{i\phi_{hk\ell}}\, e^{-i2\pi(\,hx + ky + \ell z\,)}$$

So the idea of a box around the molecule of interest in which we map ρ_{xyz} is not only a convenient enclosure, it can be thought of as one box of many identical ones, i.e., a repeating function, a crystal. Each such box is called an unit cell.

If we express x, y, z as fractional coordinates in terms of the box dimensions, then h, k, ℓ must be in terms of reciprocal fractions relating to the box dimensions.

If the data is in terms of unitless h, k, ℓ we can get back to the molecule in terms of unitless x, y, z. We would need to know the wavelength to predict where the scattered rays would appear in our experiment, and to get the size of the box in real dimensions for describing the distances between atoms in real dimensions.

Scattering from a Molecule of Multiple Atoms Using Vector Notation



Intensity in any given direction (if molecule is in a crystal then only the directions given by the Bragg Law are allowed) is a function of the number of electrons in the atoms (oscillators) and the relative positions of the atoms.

Let h, k, and ℓ index the direction of the diffracted (scattered) rays (reflections).

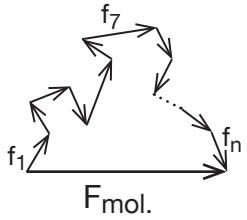
By definition, Intensity = (Amplitude)^2 or $I = F^2$ and $|F_{hk\ell}| = \sqrt{I_{hk\ell}}$

$$|F_{hk\ell}|\, e^{i\phi_{hk\ell}} = \sum_{n=1}^{N} f_n\, e^{i\phi_n}$$

where f_n is the scattering power of the n^{th} atom (oscillator), and
ϕ_n is the phase of the ray scattered by the n^{th} atom
(ϕ_n is a function of the x, y, z coordinates of that atom)

$$|F_{hk\ell}|\, e^{i\phi_{hk\ell}} = \sum_{n=1}^{N} f_n\, e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

That is, $\vec{F}_{molecule} = \sum \vec{f}_{atoms}$



If we knew or guessed all the atom positions we could calculate |F| and ϕ and so could check a trial structure against the real one by comparing $|F_{calculated}|$ vs $|F_{observed}|$ for each diffracted ray (reflection). But protein molecules have too many atoms to guess positions for!

NB: actually f_n is dependent on θ, the scattering angle because of polarization effect, size of atom, etc.

Equations and the Fourier Transform relationship

The Fourier transform is a general concept applicable to all kinds of light scattering, but it really comes into its own with x-ray crystallography where there is sufficient information to apply the details of the equations.

Since it is electrons that are scattering x-rays, the scattering factor can be thought of in terms of the electron density, ρ, throughout the molecule. Rewriting the formula from the previous page in terms of ρ:

$$\left|F_{hk\ell}\right| e^{i\phi_{hk\ell}} = \underset{\left(\begin{smallmatrix}\text{of repeat-}\\ \text{ing unit}\end{smallmatrix}\right)}{\text{Volume}} \sum_x \sum_y \sum_z \rho_{xyz}\, e^{i\,2\pi(\,hx + ky + \ell z\,)}$$

where the sum is over the entire unit cell.

The $\vec{F}$'s are the Fourier transform of the electron density, and it conveniently happens to be true that the inverse relation also holds:

$$\rho_{xyz} = (\text{Vol})^{-1} \sum_h \sum_k \sum_\ell \left|F_{hk\ell}\right| e^{i\phi_{hk\ell}}\, e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

where in theory, but not in practice, the *h, k,* and *ℓ* sums are from -∞ to +∞.
So the electron density (which is our desired final result) is the Fourier transform of the F's. Everything on the right-hand side of the equation is known except, still, for $\phi_{hk\ell}$!

We assume that the molecule stays still, oriented in a definite position. For stationary molecules in a crystal, one must consider the consequences of a regular array. (e.g. crystal diffraction)

If however, the molecules tumble in solution, then one must average over all positions.
e.g. Small Angle X-ray Scattering, SAXS : refer to :

Chapter 3, sec 1, Light Scattering: pg 10: Scattering, point particle

The Equations...

$$\lambda = 2d_{h,k,l}\sin(\theta_{h,k,l})$$

$$\vec{F} = |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \cdot f_{n,\theta} \cdot e^{-B_n(\sin\theta/\lambda)^2} \cdot e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

Amplitude • phase — Amplitude • phase

$$|F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \left(\begin{matrix}\text{Volume}\\ \text{of repeat-}\\ \text{ing unit}\end{matrix}\right) \sum_{x}\sum_{y}\sum_{z} \rho_{xyz} \cdot e^{i\,2\pi(\,hx + ky + \ell z\,)}$$

Amplitude • phase — Amplitude • phase

$$\rho_{xyz} = (Vol)^{-1} \sum_{h}^{-\infty\to\infty}\sum_{k}\sum_{\ell} m_{hk\ell} \cdot |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} \cdot e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

"Amplitude" — Amplitude • phase

Symbols used:

λ	:	wavelength of radiation
d	:	interplanar spacing, the effective distance associated with a particular diffracted ray
$\theta_{hk\ell}$	:	angle of incident beam to the *h, k, ℓ* Bragg plane
h, k, ℓ	:	integer index numbers of a particular Bragg Plane, a diffracted ray, a "reflection", index of a point of the reciprocal lattice "reciprocal space"
$\lvert F_{hk\ell}\rvert$	:	amplitude of the $hk\ell^{th}$ diffracted ray
$\phi_{hk\ell}$	:	the phase of the $hk\ell^{th}$ diffracted ray
B	:	B-factor (historically the Temperature Factor, but dominated by other uncertainties)
$f_{n,\theta_{hk\ell}}$	:	individual atomic scattering factor of the n^{th} atom as a function of $\theta_{hk\ell}$
x_n, y_n, z_n	:	coordinates of the n^{th} atom
$m_{hk\ell}$	:	figure of merit for phase of the $hk\ell^{th}$ diffracted ray
N	:	total number of atoms in the repeating unit of the crystal
O_n	:	occupancy of the n^{th} atom.
Vol	:	volume of the repeating unit
ρ_{xyz}	:	electron density at coordinates x, y, z in the crystal "real space"

Patterson Function added to the equations

$$\lambda = 2d_{h,k,l}\sin(\theta_{h,k,l})$$

$$\vec{F} = \left|F_{hk\ell}\right| \bullet e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \bullet f_{n,\theta} \bullet e^{-B_n(\sin\theta/\lambda)^2} \bullet e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

Amplitude • phase

$$\left|F_{hk\ell}\right| \bullet e^{i\phi_{hk\ell}} = \frac{\text{Volume}}{\left(\begin{smallmatrix}\text{of repeat-}\\ \text{ing unit}\end{smallmatrix}\right)} \sum_{x}\sum_{y}\sum_{z} \rho_{xyz} \bullet e^{i\,2\pi(\,hx + ky + \ell z\,)}$$

$$\rho_{xyz} = (\text{Vol})^{-1} \sum_{h}^{-\infty\,\to\,\infty}\sum_{k}\sum_{\ell} m_{hk\ell} \bullet \left|F_{hk\ell}\right| \bullet e^{i\phi_{hk\ell}} \bullet e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

Amplitude • phase

Patterson map

$$P_{xyz} = (\text{Vol})^{-2} \sum_{h}^{-\infty\,\to\,\infty}\sum_{k}\sum_{\ell} \left|F_{hk\ell}\right|^2 \bullet e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

Amplitude • phase

Symbols used:

λ	:	wavelength of radiation
d	:	interplanar spacing, the effective distance associated with a particular diffracted ray
$\theta_{hk\ell}$	:	angle of incident beam to the *h, k, ℓ* Bragg plane
h, k, ℓ	:	integer index numbers of a particular Bragg Plane, a diffracted ray, a "reflection", index of a point of the reciprocal lattice "reciprocal space"
$\left\|F_{hk\ell}\right\|$	:	amplitude of the $hk\ell^{th}$ diffracted ray
$\phi_{hk\ell}$	:	the phase of the $hk\ell^{th}$ diffracted ray
B	:	B-factor (historically the Temperature Factor, but dominated by other uncertainties)
$f_{n,\theta_{hk\ell}}$	:	individual atomic scattering factor of the n^{th} atom as a function of $\theta_{hk\ell}$
x_n, y_n, z_n	:	coordinates of the n^{th} atom
$m_{hk\ell}$	:	figure of merit for phase of the $hk\ell^{th}$ diffracted ray
N	:	total number of atoms in the repeating unit of the crystal
O_n	:	occupancy of the n^{th} atom.
Vol	:	volume of the repeating unit
ρ_{xyz}	:	electron density at coordinates x, y, z in the crystal "real space"
P_{xyz}	:	Patterson function value at coordinates x, y, z value is the product of electron densities separated by that vector distance "Patterson space" has same shape and dimensions as real space usually indexed u, v, w to distinguish distances-between from atom positions at x,y,z .

Patterson Maps

- a possible method for locating atom positions

We can experimentally measure the intensity (I, or $|F|^2$) of each spot in the diffraction pattern; but we do not know the phase of each $\vec{F}$, so we cannot calculate the electron density map:

$$\rho_{xyz} = \sum_h \sum_k \sum_\ell \vec{F}_{hk\ell}\, e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

However, in order to make the best use of what information we **do** have we can calculate a special Fourier transform of the intensities:

$$P_{xyz} = \sum_h \sum_k \sum_\ell \left|F_{hk\ell}\right|^2 e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

A plot of this function is called a Patterson map, and it has a peak for each vector between a pair of atoms in the crystal The size of each peak is the product of the number of electrons in each atom of that pair. (The Patterson function is usually expressed in terms of u,v,w, instead of x,y,z because it shows separations between atoms rather than real positions in space.) A sample simple structure and its Patterson map are shown on the next page.

In very simple cases one can reason backward from the Patterson map and discover the atom positions. Also, if there are a few heavy atoms that dominate the Patterson map then their positions can be determined and used to help solve the rest of the structure. With dominant heavy atom(s):

$$|F_T| \sim |F_H| \text{ and } |\phi_T| \sim |\phi_H|$$

Patterson map using $|F_T|^2$ (measured) Patterson map with $|F_H|^2$

To use this trick, one locates the heavy atoms, calculates $|\phi_H| \sim |\phi_T|$, calculates an approximation to the electron density map using those phases. That map will hopefully show at least some of the lighter atoms, which are used to get a better approximation to the phase for an improved electron density map which may show a few more atoms, etc., etc.

This method is used very frequently in small-molecule crystallography, but it is hopeless for proteins, where no atom could be big enough to dominate all those thousands of C,N, and O.

Patterson Maps

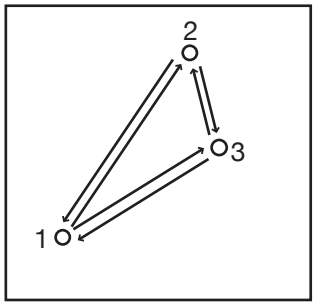


Fig. 1

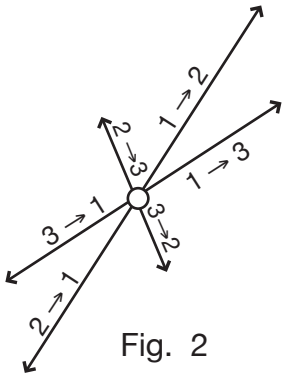


Fig. 2

Figure 1 shows a single unit cell (repeating unit) of a simple 3-atom crystal structure, with all the vectors between pairs of atoms drawn in.

Figure 2 shows these vectors as starting from a common origin. The Patterson map has a peak at the end of each vector, plus an origin peak.

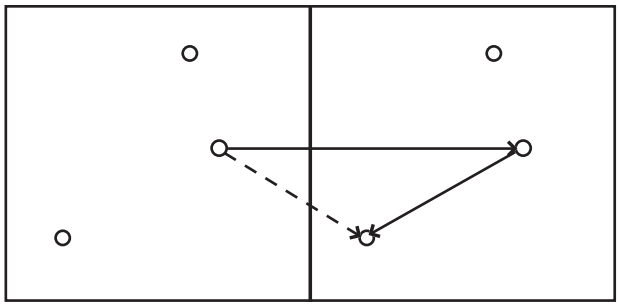


Fig. 3

Any vector between atoms in different unit cells, as in the dotted arrow of **Figure 3**, is just the unit cell translation plus one of the vectors we already plotted.

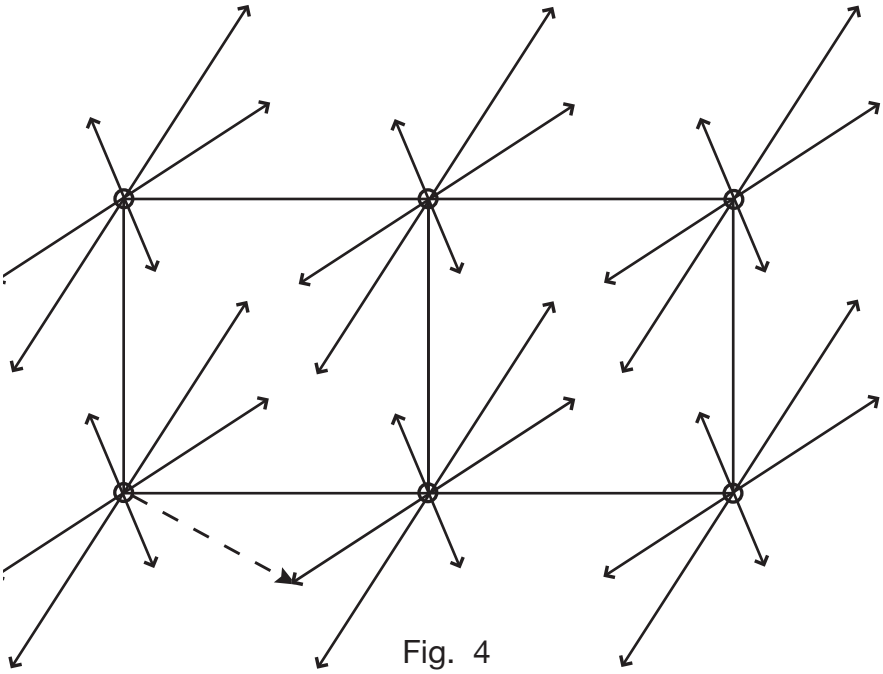


Fig. 4

Thus if we repeat the same pattern of peaks at each corner of the unit cell, as in **Figure 4**, we include all vectors between atoms.

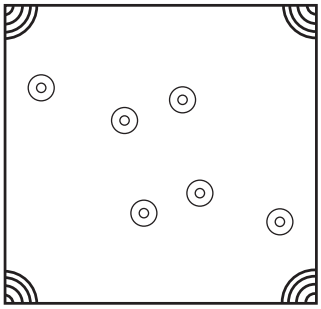


Fig. 5

Figure 5 shows what this same Patterson map would look like when plotted in the customary way as a contour map.

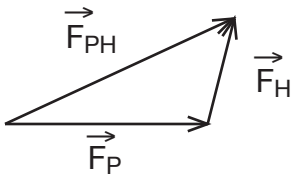
Phasing - Isomorphous Derivative Method

Add just one or a few heavy atoms to the crystal without disturbing the rest of the structure. This can be done either by:

Substitution -- Ba^{++} for Ca^{++} , I for CH_3

or Addition -- $PtCl_4$ in space between molecules

or changing wavelength so a particular "heavy" atom scatters more or less -- a small effect but useful in special circumstances. (Anomalous scattering, MAD, etc. considered in another section.)



$\vec{F}_P$ is from native protein crystals.

$\vec{F}_{PH}$ is from derivative crystals.

$\vec{F}_H$ is from an imaginary crystal containing only the heavy atoms.

(The native protein and the derivatives are usually separate crystals.*)

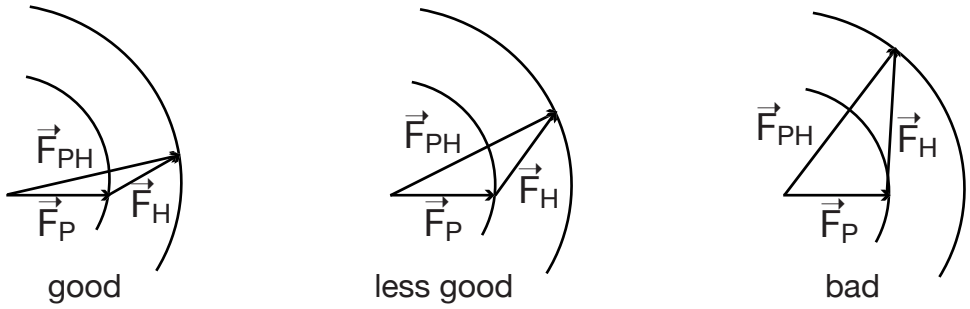
*Same crystal: changing wavelength, or adding heavy atoms to a crystal in a flow cell.

$$\vec{F}_P + \vec{F}_H = \vec{F}_{PH}$$

$$|F_{PH}| - |F_P| \approx |F_H|$$

$|F_P|$ and $|F_{PH}|$ are measured quantities so an approximate $|F_H|$ can always be calculated.

We do a Patterson map using $\approx |F_H|^2$, find the heavy atoms, and calculate ϕ_H. This works even though this approximation is only really good if $\vec{F}_{PH}$ and $\vec{F}_P$ are of nearly the same phase.

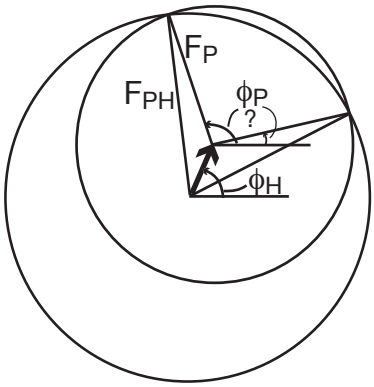


Fortunately,

1) Sometimes the phases are constrained to be either 0° or 180°.
2) There is a trick we can use: anomalous scattering by the heavy atoms reveals the difference of ϕ_H from ϕ_P .
3) Simple subtraction actually works fairly well because
   a) Error can only reduce $|F_H|$, so that term merely contributes less to the Patterson than it should have.
   b) We are using hundreds or thousands of measurements to determine a few heavy atom positions, so the Patterson peaks build-up in the right places.

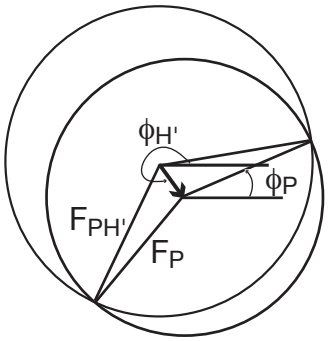
Phasing -- Determination of ϕ_P

$|F_P|$ can be measured; we must find ϕ_P (for each *h, k, ℓ* reflection).
$|F_{PH}|$ can be measured; $|F_H|$ and $|\phi_H|$ can be calculated for each *h, k, ℓ*.



Lay out circles of radii $|F_P|$ and $|F_{PH}|$ at the ends of the calculated vector F_H (arrow in diagrams). The intersections of the circles give possible values for ϕ_P; but there is a twofold ambiguity!

So, get a second heavy-atom derivative: measure $|F_{PH'}|$, calculate $IF_{H'}|$ and $\phi_{H'}$.
Layout the circles for this second case also: *



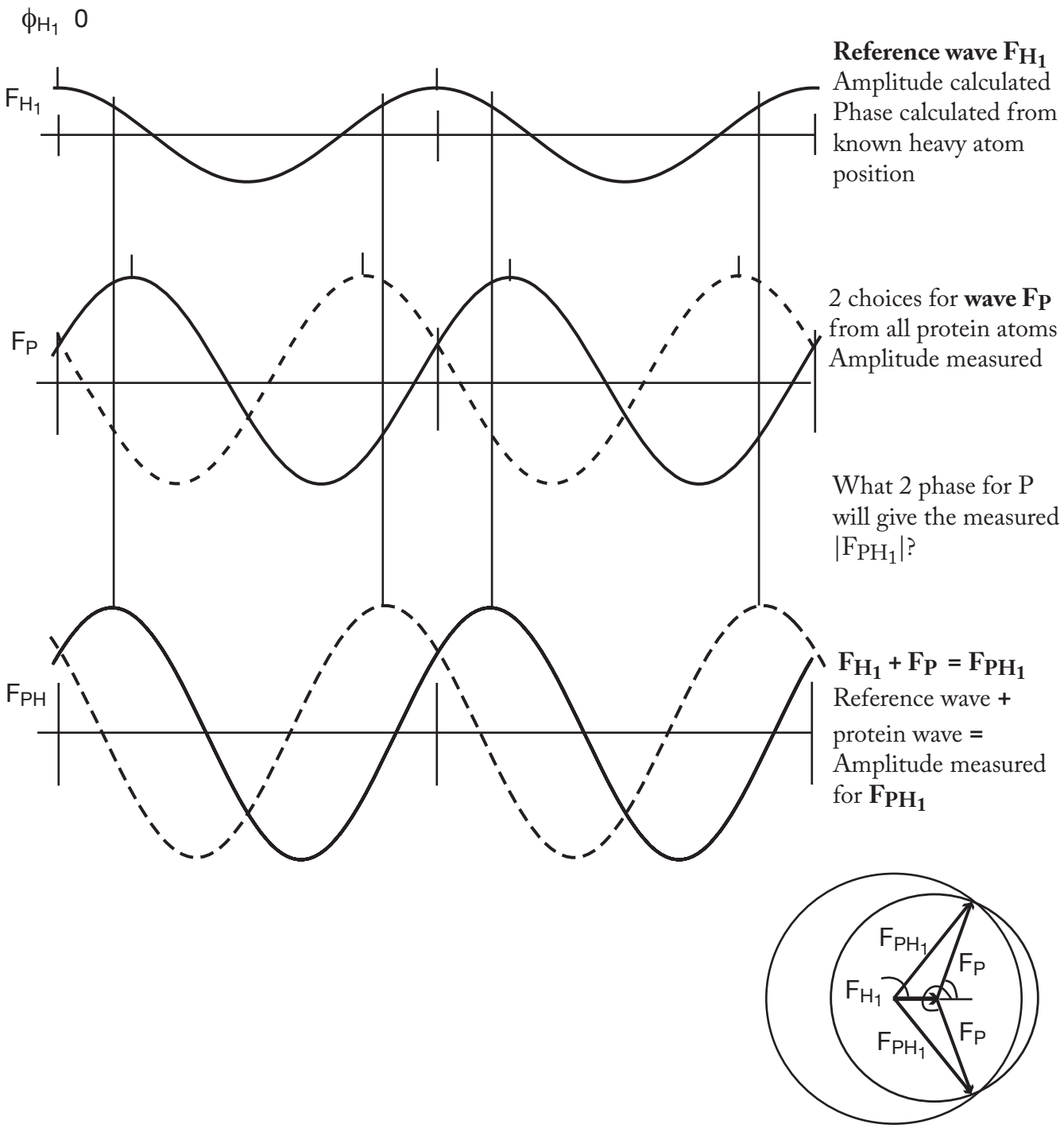
This case will give a different twofold ambiguity, and the alternative that agrees with one of the intersections from the first case should give the correct ϕ_P.

Do this for all reflections.

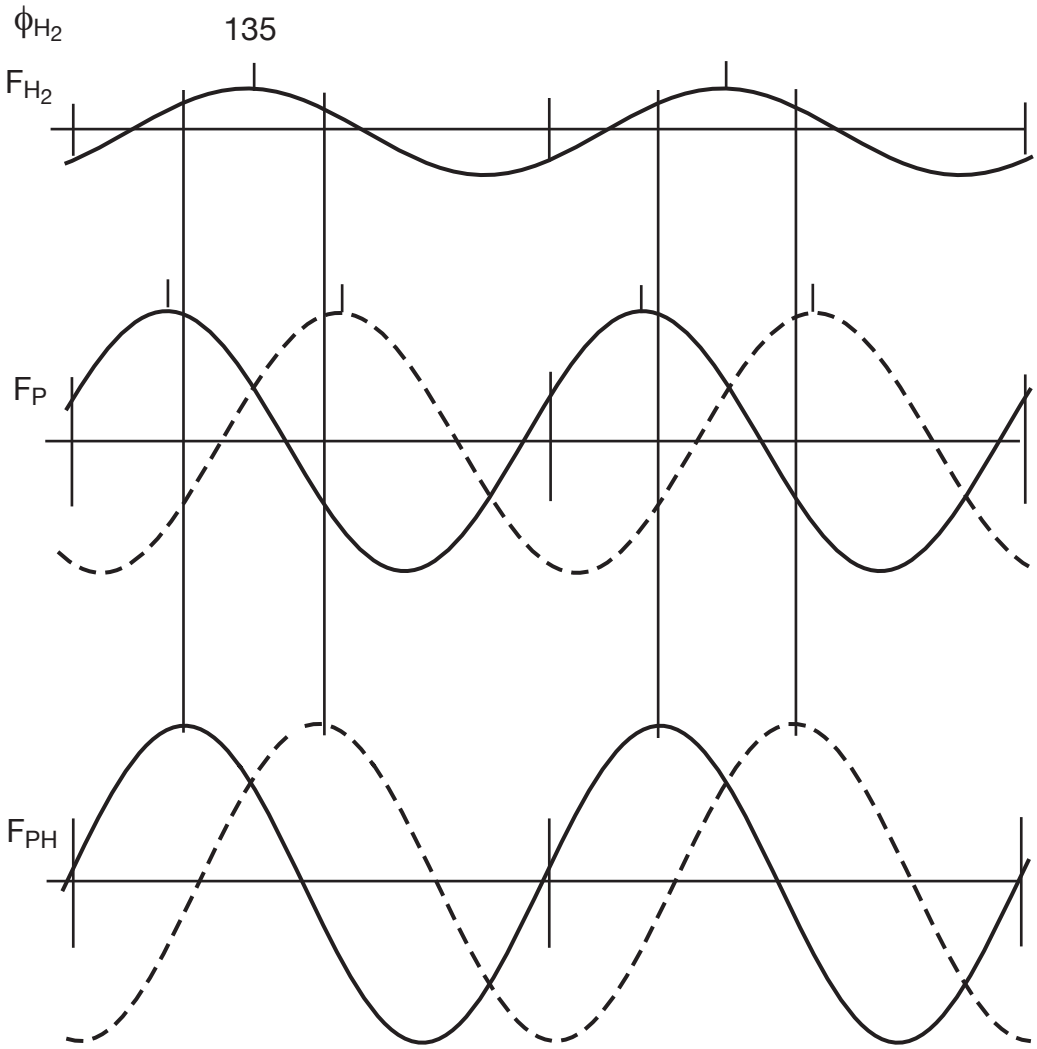
Using $|F_P(h, k, \ell)|$ and $\phi_P(h, k, \ell)$, plug into the Fourier transform formula to calculate an electron density map.

* For each derivative, calculate a $|F_{PH}|$-$|F_P|$ Patterson map and solve for those heavy atoms. Thus for each derivative $|F_H|$ and $|\phi_H|$ can be calculated for each *h, k, ℓ*.
The heavy atom models are independent, and before they can be combined, they must be related to the same origin. There are various tricks that take advantage of symmetry and patterns in the diffraction from the native protein molecule to do this. However, the best way to relate the derivatives is to make a crystal with BOTH derivatives at once. The Patterson map from this this combined derivative will give the relative postions of the heavy atoms, even if the crystal is of poorer quality.

Heavy**A**atom**I**somorphous**R**eplacement **#1**



HAIR #2

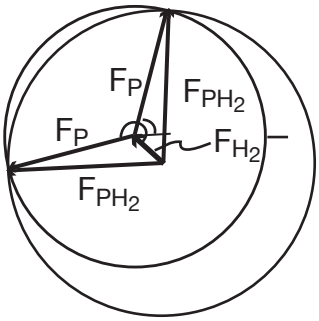


Reference wave F_{H_2}
Amplitude calculated
Phase calculated from known heavy atom position

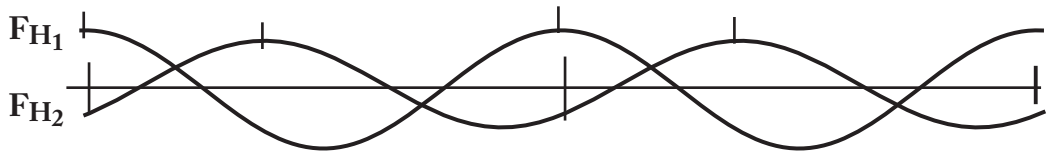
2 choices for **wave F_P** from all protein atoms
Amplitude measured

What 2 phases for P will give the measured $|F_{PH_2}|$?

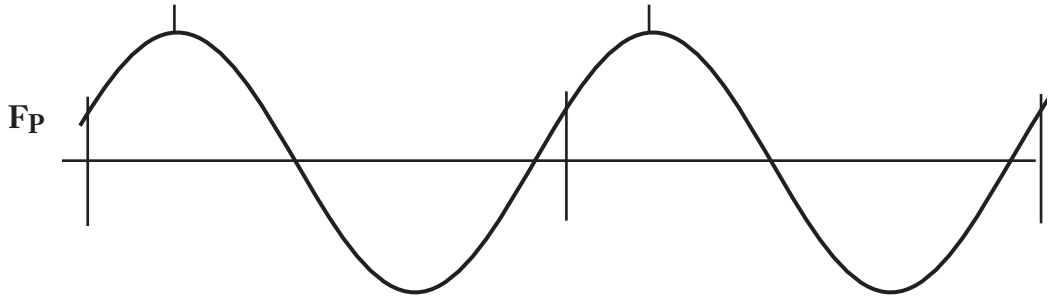
$F_{H_2} + F_P = F_{PH_2}$
Reference wave **+** protein wave **=** Amplitude measured for **F_{PH_2}**



MIR Multiple Isomorphous Replacement

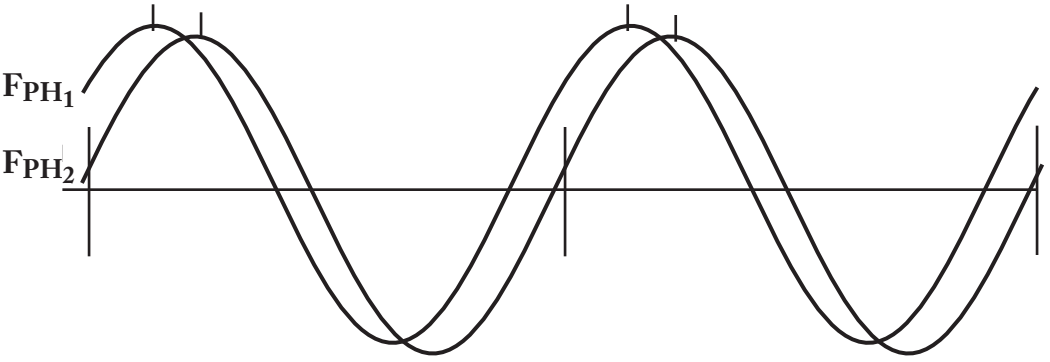


Reference waves 1 & 2
Amplitudes calculated
Phases calculated from known heavy atom positions

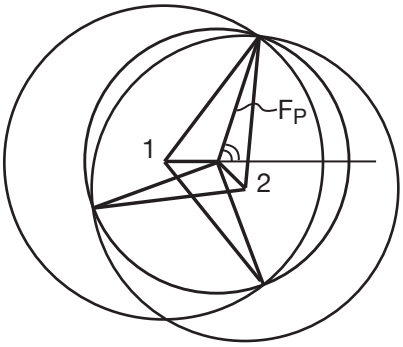


wave F_P
from all protein atoms
Amplitude measured

What phase for P will give the measured $|F_{PH}|$'s?



$\mathbf{F_{H_1} + F_P = F_{PH_1}}$
$\mathbf{F_{H_2} + F_P = F_{PH_2}}$
Amplitudes measured for $\mathbf{F_{PH_1}}$ and $\mathbf{F_{PH_2}}$



Phase Probability Distributions

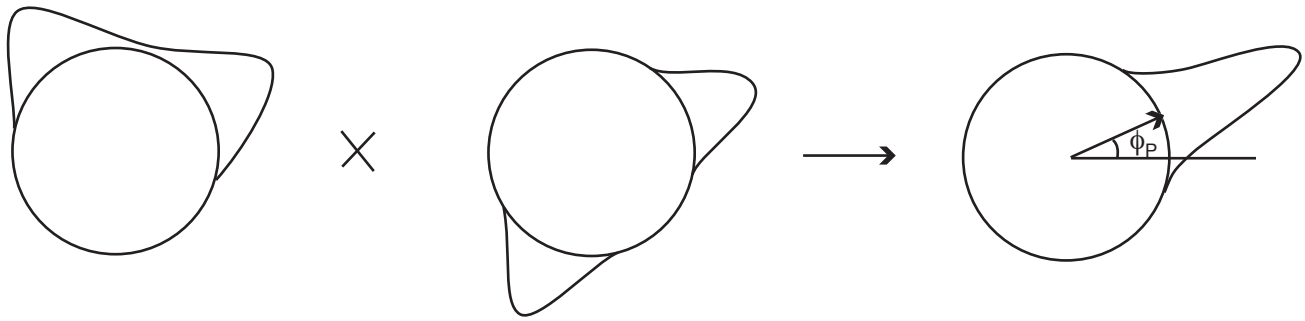
Because the data are by no means perfect, one does not try to find a unique ϕ_P where for both derivatives the circles intersect. Instead, a probability of intersection is defined that is related to the distance between these circles (as measured along a F_{PH} radius.)

$$P = e^{\left(\frac{-dist}{\varepsilon}\right)^2}$$

where ε is an appropriate estimate of the error in the data.

Each probability distribution will be bimodal. They can be plotted on circular graphs, with radial distance out beyond a reference circle representing the probability of a given phase angle for ϕ_P. The probability distributions then can be multiplied together to give a new probability distribution that hopefully will have only one large peak at the true ϕ_P, as shown below for the same two derivatives discussed in Phasing -- Isomorphous 2.

Normally, more than two derivatives are used in solving a protein structure. These probability distributions provide a convenient method of combining several sets of imperfect data.



prob. dist. for derivative H

prob. dist. for derivative H'

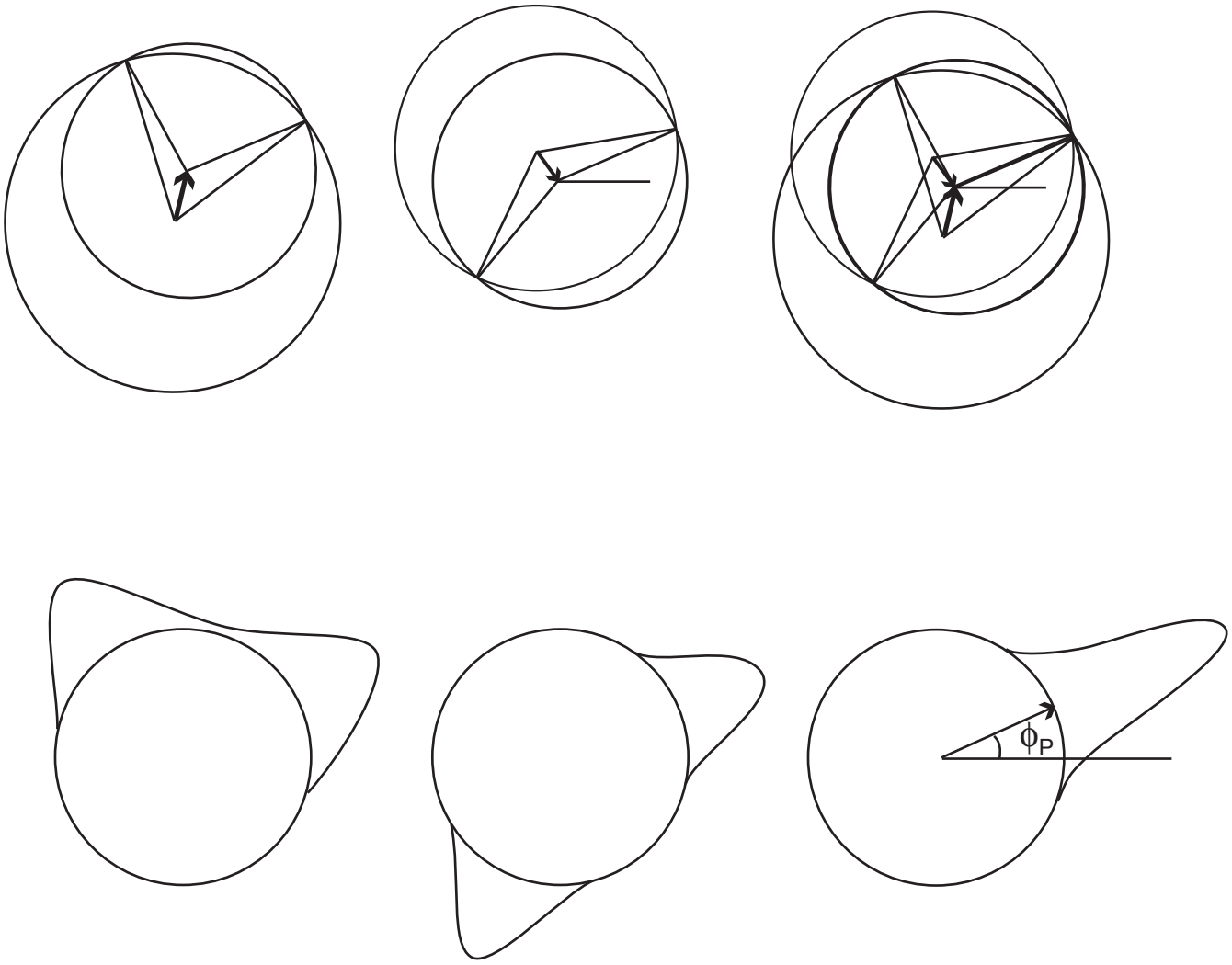
combined prob. dist.

When multiplying probability distributions: point by point multiply the probability at that point of one distribution by the probability of the other. It is important to put a "floor" under the low values: Otherwise, good probability in one distribution could be multiplied by an erroneous "zero" value in another and thus wipe out the correct value.

Triangles and Phase Probability Distributions

Probability distributions correlated with triangles. The F_P circle is centered on the arrow head. The probability distributions are drawn around the circumference of the F_P circle.

GOOD CASE

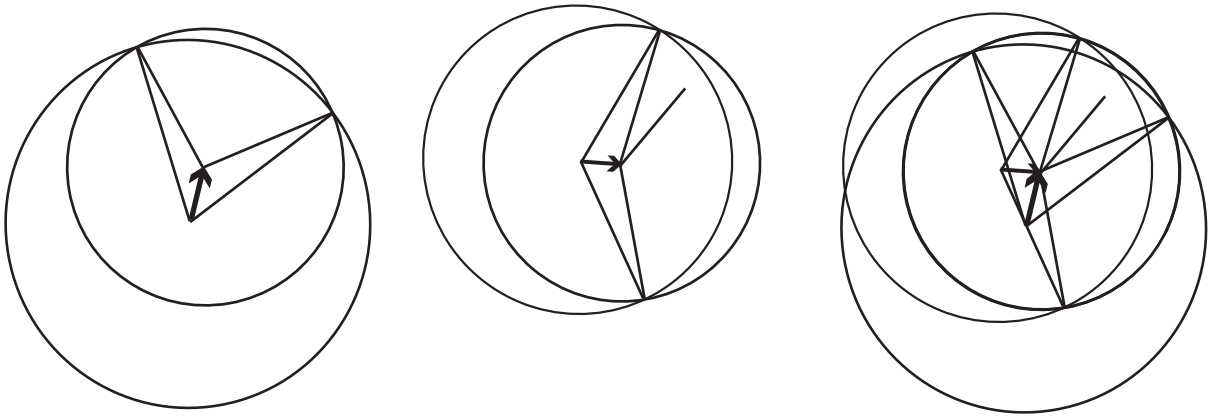


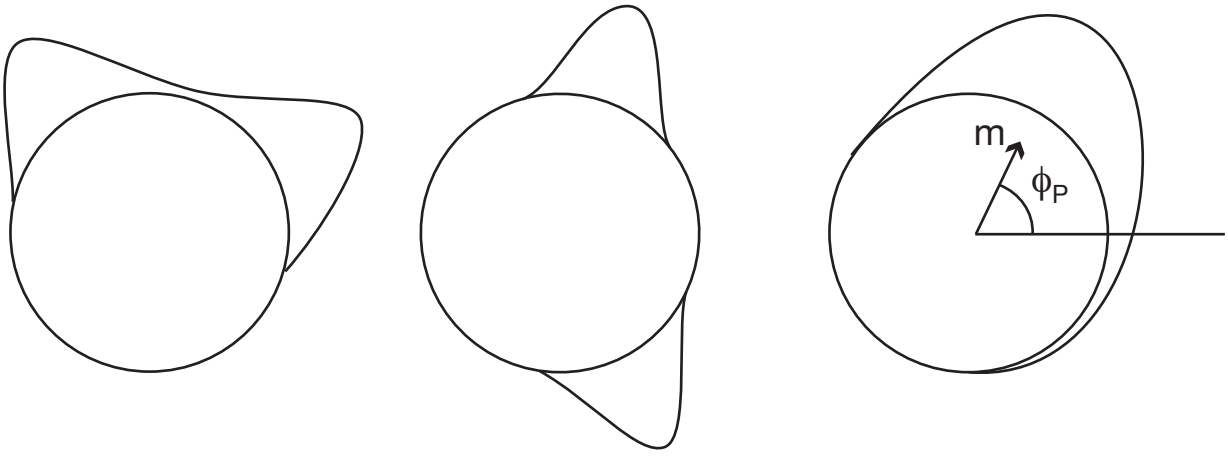
This is a good determination of the Phase! If the distribution as shown really measures the density on the circumference of the F_P circle, then the center of mass of the resultant probability distribution is very close to the tip of the F_P arrowhead. This reflection would have a figure of merit very close to 1.0.

Triangles and Phase Probability Distributions

Probability distributions correlated with triangles. The F_P circle is centered on the arrow head. The probability distributions are drawn around the circumference of the F_P circle.

NOT SO GOOD CASE



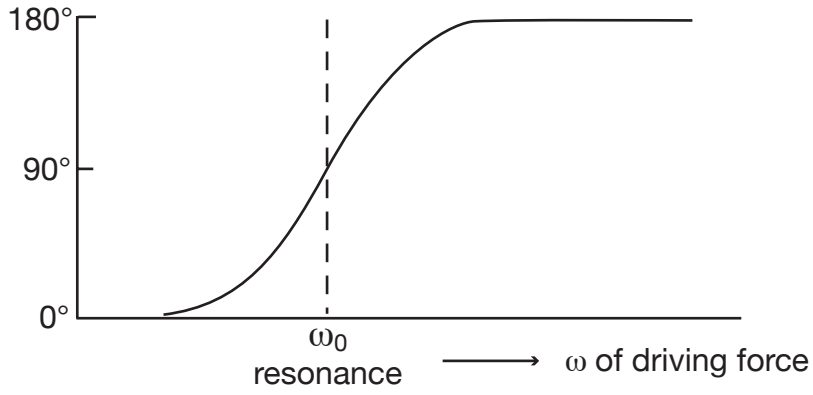


This reflections has a less sure determination of its Phase!
If the distribution as shown really measures the density on the circumference of the F_P circle, then the fractional radius of the center of mass of the resultant probability distribution is the figure of merit. This reflection would have a figure of merit about 0.7.

[Note: There is now a more elaborate method called Maximum Likelihood, which explicitly takes into account estimates of possible uncertainty in both the amplitude measurement and in knowledge of the phase of all components.]

Anomalous Scattering

Usually the x-ray wave which drives the electrons into forced vibrations is of a much higher frequency than the resonant frequency of the electrons. In this case, the forced damped simple harmonic oscillations of the electrons will be 180° out of phase with the driving wave. This behavior is a completely general characteristic of forced simple harmonic motion. It depends on the interaction of the driving force and the natural restoring force of the oscillator, and it can easily be demonstrated with a pendulum or spring. The relationship of phase lag to frequency is shown on the diagram:



Sometimes the binding of some inner electrons of an atom is of such a strength that the resonant frequency is shifted from that usual for free electrons toward the frequency of the x-ray wave. These electrons will oscillate nearer to 90° out of phase (and also more violently, since they are nearer resonance). For a given atom, then, most of the scattered ray has a 180° phase lag and there will be a few electrons-worth of scattering with a 90° phase lag. Relative to the normal diffracted ray, therefore, the anomalous part of the scattering has a 90° phase advance.

The amount of anomalous scattering for a given atomic species is known and can be looked up in tables (International Tables for Crystallography, Vol. III). For a given x-ray wavelength an atom's scattering will be described as:

$$f_{total} = f_0 + f' + if''$$

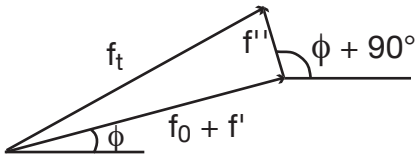
usual scattering at that 2θ

correction to usual scattering

number electrons-worth of anomalous scattering

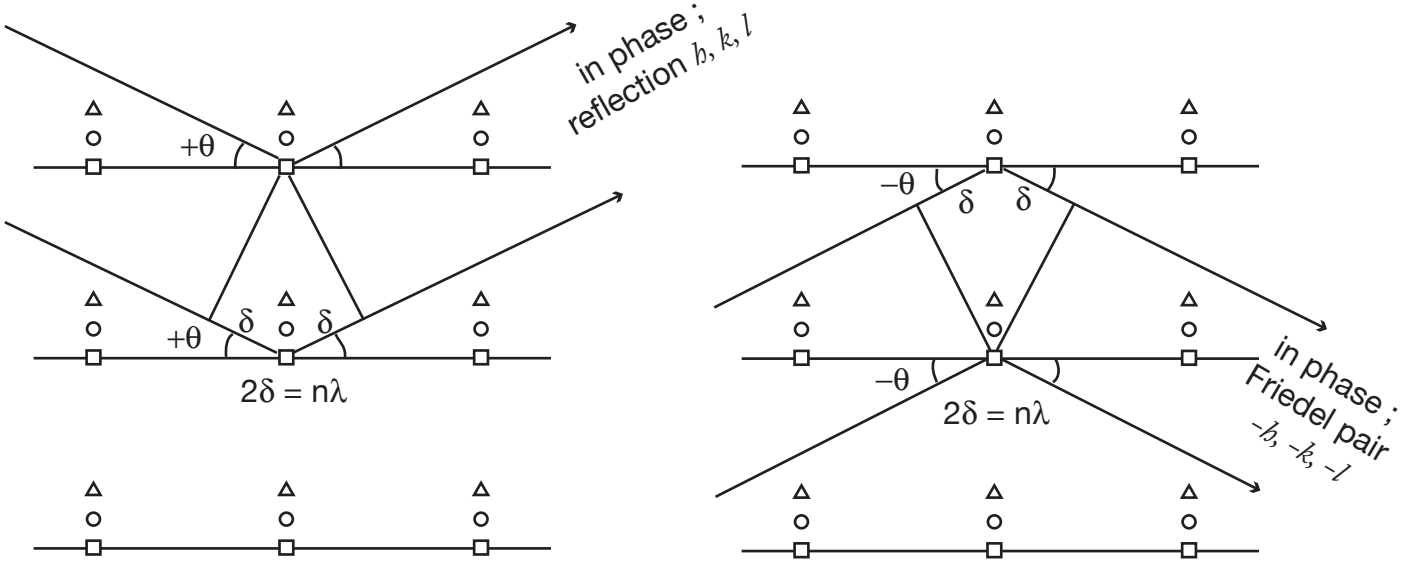
means 90° out of phase

This is what a vector diagram looks like with anomalous scattering, for a reflection of phase ϕ:

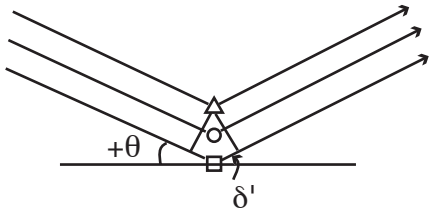


This turns out to have consequences for the diffraction pattern, if we consider a crystal in which only a few of the atoms (usually, only the heavy ones) exhibit anomalous scattering.

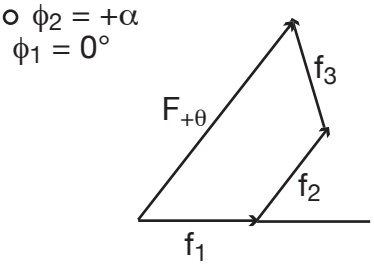
Friedel pairs are similar reflections that are taken on opposite sides of a crystal. Normally their intensities are exactly equal (as we will show on the next page); it is this equivalence that introduces a center of symmetry into all diffraction patterns, whether the crystals have a center of symmetry or not.

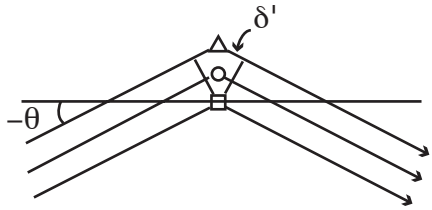


Now let us consider the phase relationships within each triatomic molecule of the crystal lattice on the last page. The geometry is the same for reflections from opposite sides of the crystal, except that for the +θ case the rays from the triangle and circle atoms are ahead of the reference ray (square atoms) and for the -θ case they are behind.

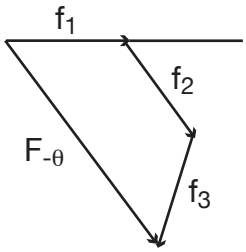


△ $\phi_3 = +\beta$
○ $\phi_2 = +\alpha$
□ $\phi_1 = 0°$



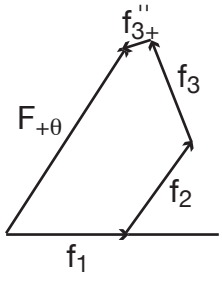


△ $\phi_3 = -\beta$
○ $\phi_2 = -\alpha$
□ $\phi_1 = 0°$

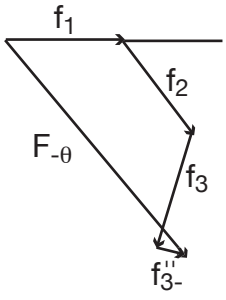


$|F(+\theta)| = |F(-\theta)|$ because the vector diagrams would superimpose if the -θ one were reflected up onto the +θ one; so the Friedel pair reflections have the same intensity. But suppose one of the atoms (say, the triangles) has anomalous scattering: then an f" term with a 90° phase advance is added onto its f vector. Now when we superimpose the two diagrams as at the right the f" parts do <u>not</u> fall on top of each other and $|F(+\theta)| < |F(-\theta)|$.

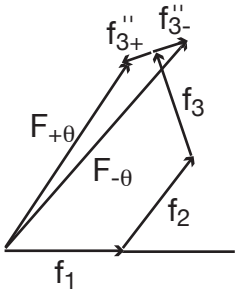
These intensity differences between $|F_{hk\ell}|^+$ and $|F_{hk\ell}|^-$ are small, but measurable.



△ $\phi_{3+} = +\beta$
△ $\phi''_{3+} = +\beta + 90°$

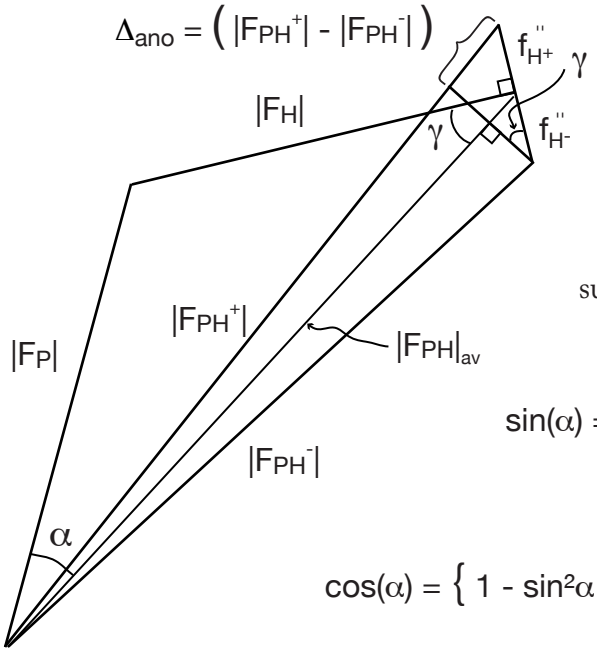


△ $\phi_{3-} = -\beta$
△ $\phi''_{3-} = -\beta + 90°$



Uses of Anomalous Scattering

1) To help get an accurate value (see phasing by isomorphous replacement, page 1) for the heavy atom contribution $|F_H|$, given measured values of $|F_P|$ (native protein crystal) and the Friedel pairs $|F_{PH}^+|$ and $|F_{PH}^-|$ from the heavy atom derivative protein crystal.



in the small (approximately) right triangle:

$$\sin(\gamma) = \Delta_{ano}/2f''_H$$

applying the law of sines to the large triangle:

$$\sin(\alpha) = \frac{|F_H|}{|F_P|}\sin(\gamma)$$

substituting, and letting

$$\kappa = |F_H| / f''_H$$

$$\sin(\alpha) = \frac{|F_H|\,\Delta_{ano}}{|F_P|\,2f''_H} = \frac{\kappa\,\Delta_{ano}}{2\,|F_P|}$$

$$\cos(\alpha) = \left\{ 1 - \sin^2\alpha \right\}^{1/2} = \left\{ 1 - \frac{\kappa^2\,\Delta_{ano}^2}{4\,|F_P|^2} \right\}^{1/2}$$

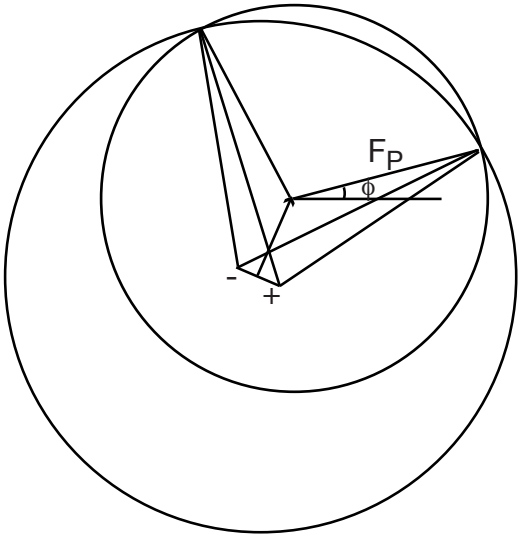
Since we know two sides ($|F_P|$ and $|F_{PH}|_{av}$) and the cosine of the included angle in the large triangle, we can use the **law of cosines** to solve for the third side, which is what we wanted to find.

$$|F_H|^2 = |F_P|^2 + |F_{PH}|^2_{av} - 2|F_P||F_{PH}|_{av}\left\{ 1 - \frac{\kappa^2\,\Delta_{ano}^2}{4\,|F_P|^2} \right\}^{1/2}$$

$\kappa = |F_H| / f''_H = \frac{\text{normal}}{\text{anomalous}}$ scattering of heavy atom can be looked up. All other quantities on the right hand side are measured experimentally. This gives us an accurate way of finding $|F_H|$, rather than approximating it by $|F_{PH}| - |F_P|$ as we did in the section on phasing by isomorphous replacement.

Uses of Anomalous Scattering, cont'd

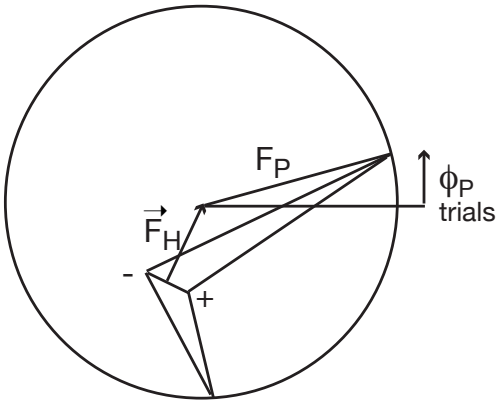
2) To help find the protein phases:



As in the diagram at the left, anomalous scattering resolves the twofold ambiguity in each protein phase determination, because if $|F^{+}_{PH}| < |F^{-}_{PH}|$ the triangle on the right must be the correct one. This means that theoretically one heavy-atom derivative is sufficient to determine protein phases and make an electron density map (this SIRAS is now done quite routinely with the help of a few additional tricks); or, if you have several isomorphous derivatives then anomalous scattering measurements will improve the accuracy of your ϕ_P values.

The best way to handle the anomalous scattering to get ϕ_P is illustrated below. Instead of just using the direction of the difference between $|F^{+}_{PH}|$ and $|F^{-}_{PH}|$ it makes use of the magnitude of that difference. That way you get an accurate phase if the measurements are exact, and you go wrong less spectacularly if they are inexact.

After laying out the known $\vec{F}_H$ vector with its $^{+}\theta$ and $^{-}\theta$ anomalous terms, you make a series of trials at each ϕ_P direction. At each trial there will be a trial difference between $|F^{+}_{PH}|$ and $|F^{-}_{PH}|$ which can be compared with the difference you measured experimentally.



If Δ = trial difference - experimental difference, then

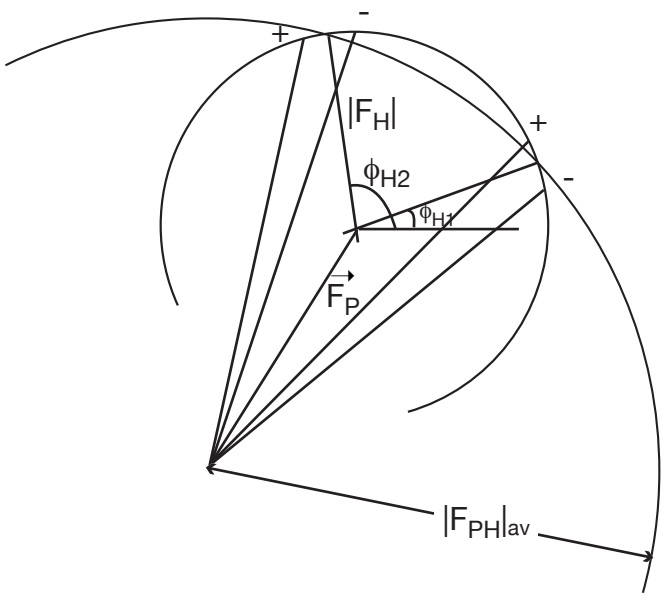
$$P = e^{-(\Delta^2 / \varepsilon^2)}$$

(where ε is an estimate of the error in the anomalous difference measurements) will give a phase probability distribution that can be combined with the ones derived in Phasing-IR-6.

Uses of Anomalous Scattering, cont'd

3) **"Cross-Fourier"** maps to locate additional heavy atoms derivatives

If, from one or more previously analyzed derivatives, one has a fairly good value for ϕ_P, then if $|F^+_{PH}|$ and $|F^-_{PH}|$ are measured on a new derivative an initial value for ϕ_H can be obtained immediately.



Laying out Fp as a known vector and swinging circles from either end of it of lengths $|F_H|$ (from law of cosines formula) and $|F_{PH}|_{av}$, the two intersections will give two possible values for ϕ_H. If $|F^+_{PH}| > |F^-_{PH}|$ then ϕ_{H1} is correct, and if $|F^+_{PH}| < |F^-_{PH}|$ then ϕ_{H2} is correct.

Now that we know both $|F_H|$ and ϕ_H for every *h,k,ℓ* we can calculate an electron density map showing the heavy atom positions in the new derivative. This technique is especially valuable if the new heavy atom occupies a number of different sites on each protein molecule, in which case the Patterson map could be too complex to be unravelled.

4) Determines handedness (absolute configuration, enantiomorph)

Right-handed and left-handed molecules are indistinguishable in their effects on the ordinary intensities, but they would give the opposite anomalous scattering terms. So, using anomalous scattering one can see in the electron density map that the molecule indeed has *ℓ*-amino acids and right-handed α-helices, rather than having to use their occurrence as a criterion for whether one had guessed the correct enantiomorph.

Anomalous Scattering: SIRAS

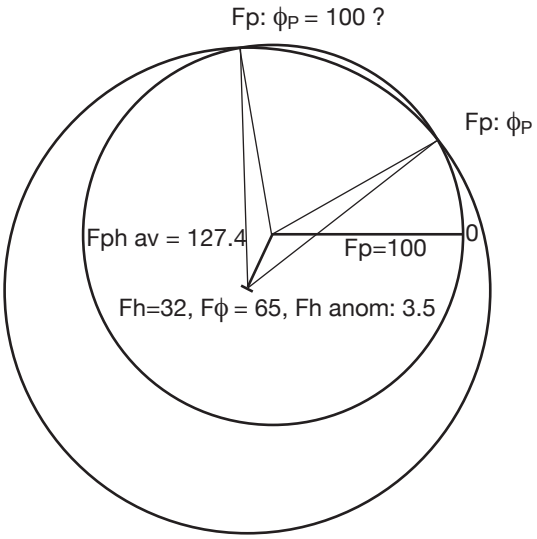
SIRAS, Single Isomorphous Replacement with Anomalous Scattering, is illustrated with PhaseSIRAS.kin in either Mage or KiNG. The following few pages are frames from this kinemage taken to show key steps in understanding this method.

PhaseSIRAS.kin

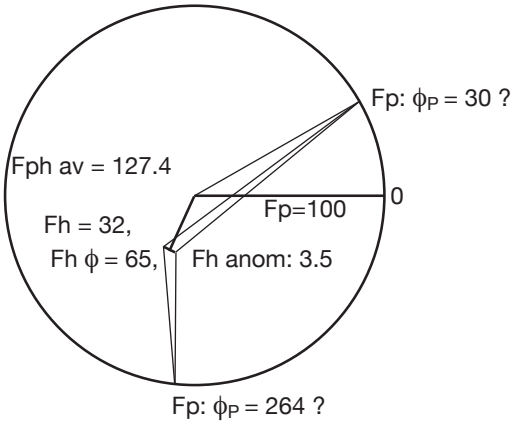
```
@text
SIRAS.kin
Single Isomorphous Replacement with Anomalous Scattering
|Fp| = 100
IFph-| = 128.965  e.g. “L3” measurement of MAD experiment
IFph+| = 125.849
del anom = 3.116 with - > +
IFph|av = 127.407
```

Take “L3” as typical for SIRAS since often one is working from a fixed wavelength x-ray source which is not optimal for the particular anomalous scatterer in the experiment’s crystal.

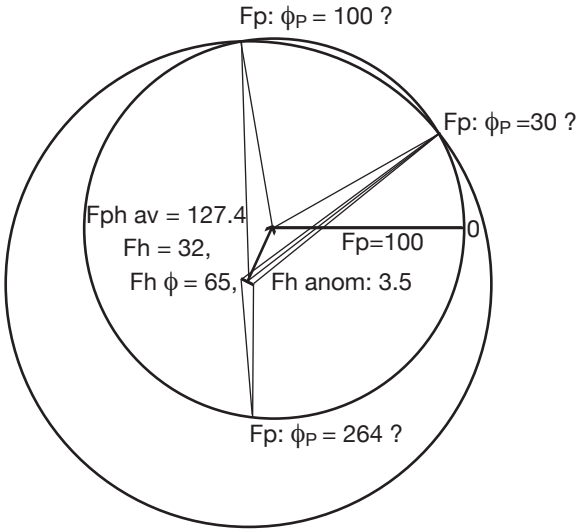
PhaseSIRAS.kin Isomorphous part



PhaseSIRAS.kin Anomalous Scattering part



PhaseSIRAS.kin Isomorphous part combined with Anomalous Scattering part



MAD: Multiple wavelength Anomalous Dispersion

There are two approaches to abstracting information from a MAD experiment. One is to use all information in a set of simultaneous equations, mathematically satisfying perhaps, but using a different intuition then our geometrical approach. Another is to treat the experiment as a combination of isomorphous changes in the scattering power of the "heavy" atoms along with associated anomalous scattering differences. This uses the same conceptual machinery already developed for the SIRAS method, and, indeed, one of the commonly used computer programs for MAD phasing was originally developed for SIRAS.

Theoretically, all that is really needed is data at two wavelengths. Recall that in SIRAS a difference in isomorphous scattering along with one pair of anomalous scattering sets is sufficient. However, not only does redundancy of taking data at more than 2 wavelengths help in this experiment where one is interested in small differences between large numbers, three wavelengths can be selected to optimize the phasing information (and often data at 4 wavelengths is taken for further assurance).

Three wavelengths can be chosen, in terms of the SIRAS analysis, such that one pair has maximal "isomorphous" differences, and an intermediate wavelength gives maximal anomalous differences. At the absorption edge of an element, inner electrons come into resonance with the incident x-ray energy, these electrons scatter more strongly and with a different phase shift (i.e. more like 90 phase lag) to the incident wave. There also are slightly fewer electrons scattering with the usual 180 phase lag, so the usual scattering falls off while the anomalous scattering increases as the absorption edge is crossed. The effect is specific to the element but is subject to slight shifts because of interactions with surrounding atoms, so good practice is to run a scattering scan across the wavelength region and determine exactly the scattering profiles for the crystal under study.

Scattering from an element can be expressed as f(total) = fo + f' + f" where fo is the expected scattering for that element at the angle of the measurement, f' is the change in the usual 180 deg. phase lagged scatter, and f' is the amount of scatter with a 90 deg. phase lag. (2 components at 90 deg. of course, can represent all scattering.)

f' hits a minimum just at the absorption edge where f" is rapidly changing, and f" has a sharp maximum at the high energy side of the edge. Let L1 be the point of minimum usual scattering, L2 the point of maximal anomalous scattering, and L3 be remote from the edge where the usual scattering is no longer depressed. Then most of the phasing power can be extracted from the experiment just by taking the difference between L1 and L3 as the "isomorphous" part (averaging the Bijvoet pairs taken at those wavelengths) and the Bijvoet pair at L2 for the anomalous signal. Of course, with 3 sets of Bijvoet pairs there are other ways to combine the data sets to enhance the accuracy, but most of the power is gained from those optimal combinations.

Modern "area" detectors and high intensity x-rays from a synchrotron source on frozen crystals, allow for efficient collection of the Bijvoet pairs and redundancy to provide accurate intensity measurements.

TABLE I
VALUES OF f' AND f'' FOR A THREE-WAVELENGTH EXPERIMENT AT THE SELENIUM K-EDGE

Data set	Wavelength (Å)	f' (electrons)	f'' (electrons)
L1	0.9802	−9.52	3.15
L2	0.9795	−7.35	5.92
L3	0.9300	−2.19	3.46

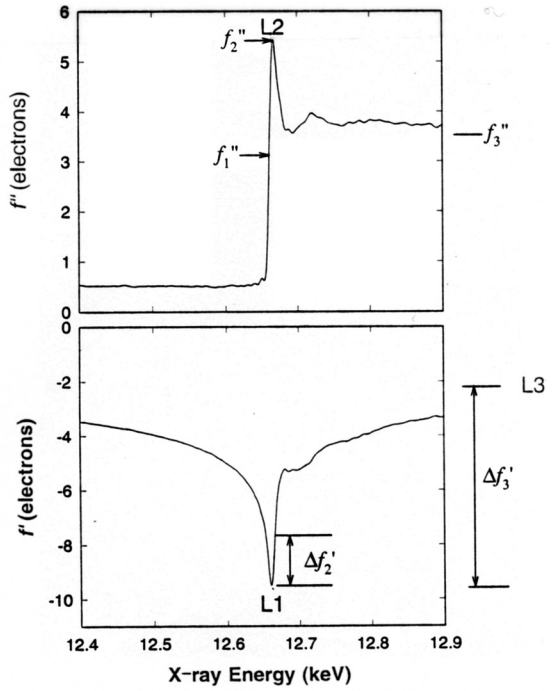


FIG. 1. The variation of the real (f') and imaginary (f'') components of the anomalous scattering around the K-edge of selenium, from a crystal containing selenomethionine. The values were calculated from measured values of the absorption spectrum.[10] Data were taken on beam line X12C at the NSLS at Brookhaven National Laboratory. The three wavelengths chosen for diffraction measurements in a MAD experiment are designated as follows: L1 is the minimum of f', called variously the "inflection point," "edge," or "rising edge"; L2 is the maximum of f'', called the "white line"; and L3 is a remote point that is at 0.930 Å, which lies well to the right of the interval shown in the graph.

While the distribution of intensity in the white radiation depends primarily on the accelerating voltage and only to a small extent on the nature of the target material, X-ray spectra show in addition a number of sharp spikes of high intensity whose positions change from one material to another (Fig. 1.2). These peaks are the ***characteristic lines*** for the element of which the target is made.[3] When the electrons bombarding the target reach certain critical energies (***threshold potentials***) they are capable of knocking electrons out of their atomic orbitals. In particular, at energies of about 10,000 eV (for elements with atomic number ~30) they can remove electrons from the innermost (K) shell. The vacancy in the K shell is then filled by the descent of an electron from the next higher shell (L) or the one above that (M). The decrease in potential energy in going from the higher level to the lower appears as radiation, and as the energies of the shells are

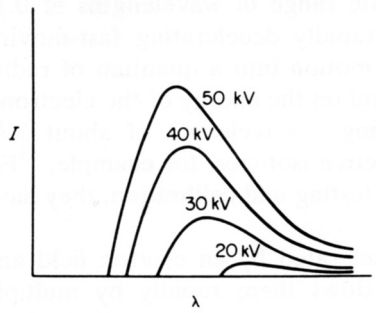


Figure 1.1. Continuous X-ray spectra as a function of accelerating voltage.

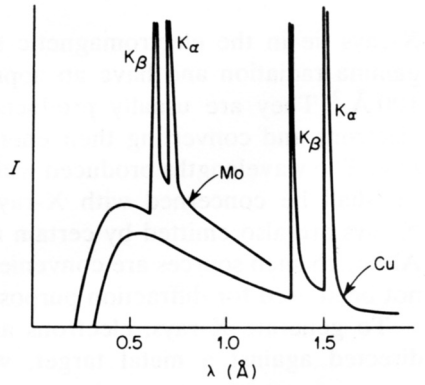


Figure 1.2. X-ray spectra with characteristic peaks: Mo K_α, 50 kV; Cu K_α, 35 kV.

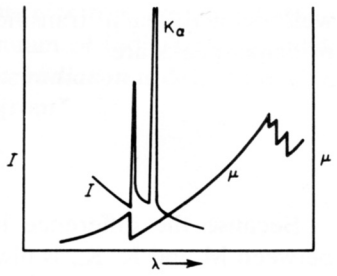


Figure 1.3. X-ray spectrum showing characteristic peaks (I curve) and absorption coefficient (μ curve) for the same element as a function of wavelength λ.

(Figures from Stout and Jensen, 2nd Edition)

PhaseMADasSIRAS.kin

@text
MAD.kin
"L1" measurement near adsorption close to minimum of f'
|Fph-| = 122.396 |Fph|av = 120.902 take average as Fph, i.e. isomorphous derivative data
|Fph+| = 119.408
del anom = 2.99 with - > +

"L2" measurement near adsorption close to maximum of f"
|Fph-| = 125.658 |Fph|av = 122.893
|Fph+| = 120.127
del anom = 5.531 with - > +

"L3" measurement far from adsorption edge
|Fph-| = 128.965 |Fph|av = 127.407 take average as "Native"== Fp
|Fph+| = 125.849
del anom = 3.116 with - > +

MAD treated as SIRAS

L1 chosen for maximal Δf'

L2 chosen for maximal Δf"

L3 chosen for minimal Δf'

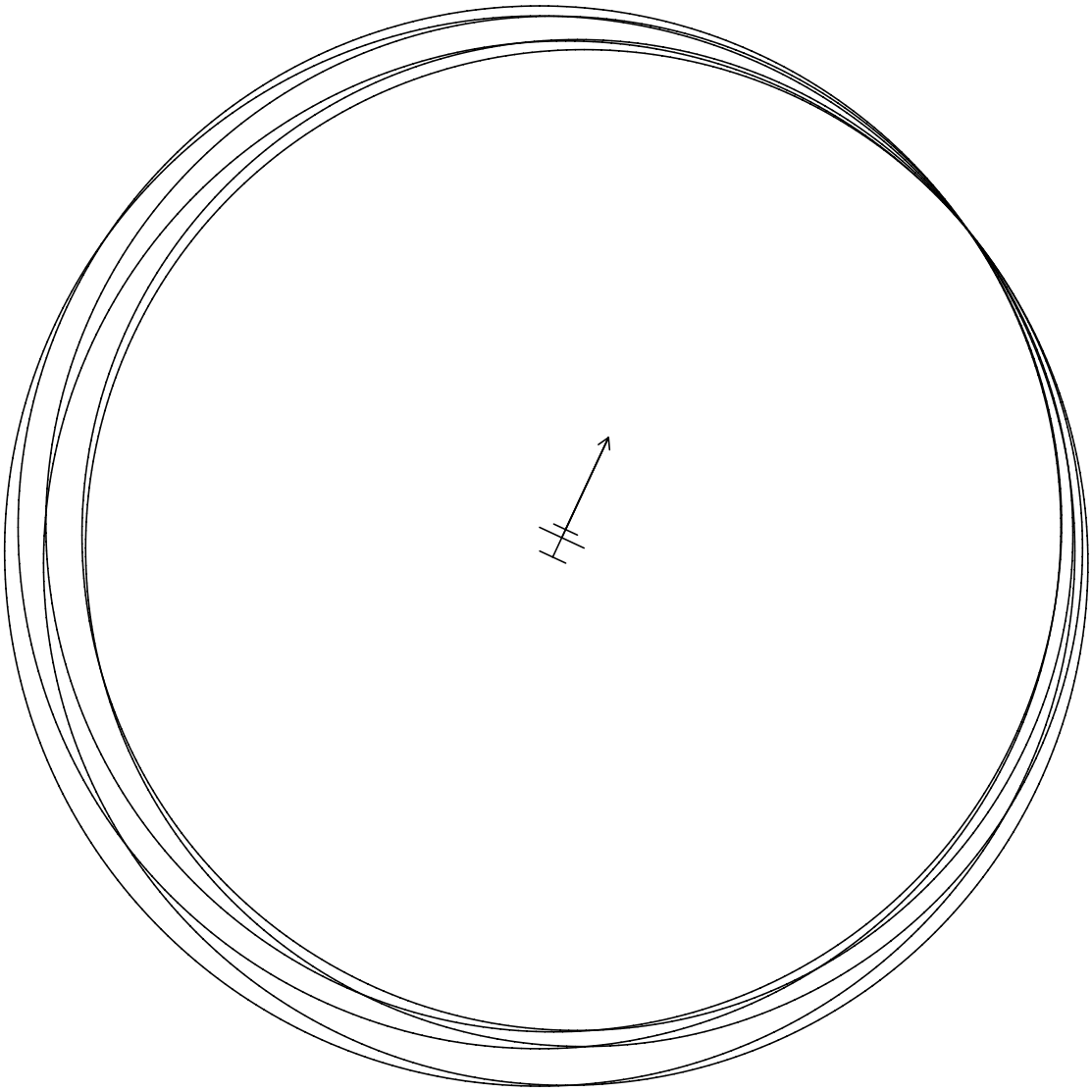
Thus L1av - L3av is the maximum "isomorphous" difference: analogous to "derivative - native"
Where L1av = average of $L1^+$ and $L1^-$, and L3av = average of $L3^+$ and $L3^-$.

Note that this exactly isomorphous "derivative" has less scattering power than the "native", so in general, L1av - L3av = "$F_{PH} - F_P$" will be negative.
Also $L2^+ - L2^-$ is the maximum anomalous difference.

One then computes both an isomorphous and an anomalous Patterson map.
The isomorphous uses $F_H = F_{PH} - F_P =$ L1av - L3av, and
the anomalous uses $L2^+ - L2^-$. Each Patterson map could theoretically be solved by itself, but in practice looking at both makes finding the correct peaks much easier. Then one has coordinates of the "heavy" atoms (in this case, the anomalously scattering atoms). With these coordinates, for any diffracted ray (spot, reflection, reciprocal spot, etc.) one can compute the F_H vector. Since the only difference between the isomorphous pairs is $\Delta\Delta$f' that is what the f_n of the "heavy" atom is.

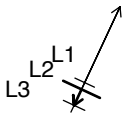
PhaseMADasSIRAS.kin

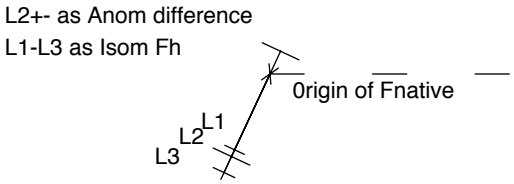
MAD: + & - at 3 different wavelengths for 6 data sets. For each hkl reflection, the six circles should intersect at one point when drawn from appropriately positioned vectors representing the waves from the anomalously scattering atoms. .

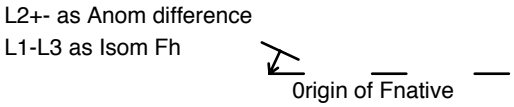


PhaseMADasSIRAS.kin

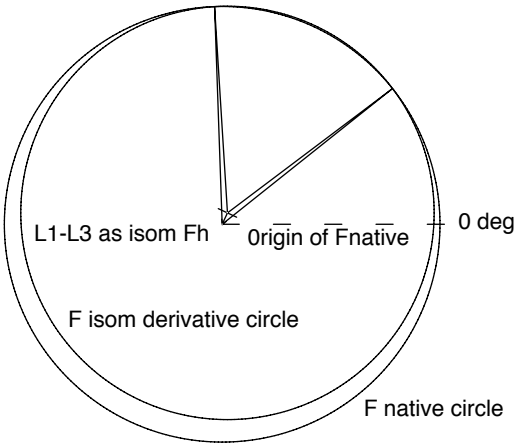
MAD as SIRAS: diagrams to show how to set up the effective "heavy" atom wave vector for a particular reflection once the isomorphous and anomalous Patterson maps have been solved to yield the position of that heavy atom in the unit cell.



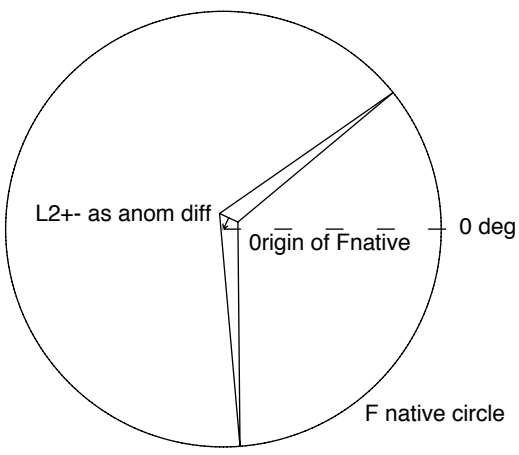




MAD as SIRAS: Draw Fnative circle from the head of the isomorphous Fheavy-atom phase vector, draw F derivative circle from the base of the isomorphous Fheavy-atom phase vector. The intersection points are the 2 possible phases of the "native" phase vector.

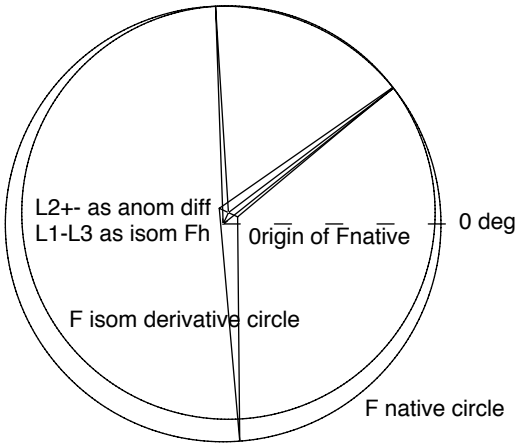


MAD as SIRAS: At regular intervals around the Fnative circle, measure the distance from those points to both the + and - ends of the anomalous vectors. The closer the difference between those lengths is to the actual measured anomalous difference, the more probable that is the phase of the "Fnative"..

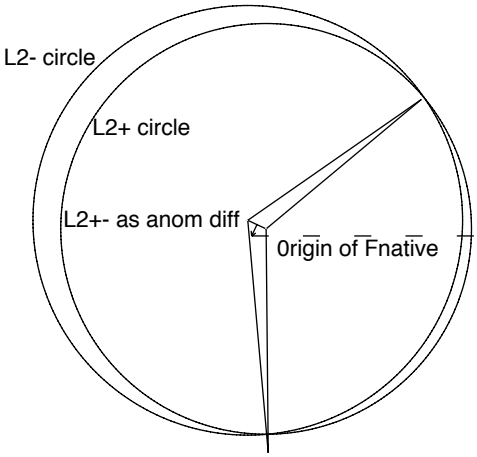


PhaseMADasSIRAS.kin

MAD as SIRAS: Multiplying the two phase probability distributions will give a combined phase probability distribution. A maximum shows where the phase triangles coincide, or at least a best phase and a figure of merit as a measure of the spread of the distribution..



MAD as SIRAS: Extra diagram showing that the phases from the anomalous data are in the direction of where the two separate anomalous circles ovelap.



When we know the ϕ_P for each reflection, we can use the Fourier transform formula to get a picture of the electron density in the molecule:

Electron density is the Fourier transform of all diffracted waves.

$$\rho_{xyz} = (Vol)^{-1} \sum_{h}^{-\infty \rightarrow \infty} \sum_{k} \sum_{\ell} m_{hk\ell} \cdot \left| F_{hk\ell} \right| \cdot e^{i\phi_{hk\ell}} \cdot e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

Amplitude (electron density has no phase)

(Amplitude-factors) • (Phase-factors) here the x,y,z is the point in the unit cell

The x,y,z coordinates actually used in the calculaton are fractions of the unit cell edges a,b,c. This is in accord with the *h,k,l* actually representing spacings defined in terms of Bragg planes that cut the edges of the unit cell into integral fractions. Just as *h,k,l* defines directions $h \perp y,z$; $k \perp z,x$; $l \perp x,y$ perpendicular to the planes of the unit cell parallelepiped, the x,y,z directions are aligned with the edges of the unit cell. Thus the natural coordinate system of the model is (often) non-orthogonal and non-normalized. However, most graphics programs (and most people) work with ortho-normal Cartesian coordinates in standard units (usually Ångstroms). Since a,b,c can be expressed in Å at the known angles of the unit cell, Cartesian model atom coordinates are calculated from the size and shape of the unit cell.

experimental phases, model phases

When some sort of starting values of the phases are available, a model can be built into the electron density. This electron density is dependent on both the amplitudes and the phases of the *h,k,l* data points. The appearance of the image turns out to be most dependent on the phases, i.e. on the part not known directly from experiment, and thus quite susceptible to errors and misconceptions. Initially errors in deriving the starting phases, and later as the model itself is used to calculate phases, from errors and misconceptions about molecule.

When we know (some) of the coordinates of a model, we can use a Fourier transform of these to get calculated phases for each reflection in order to make a (hopefully) better electron-density image

A diffracted wave is the sum of contributions from all atoms.

$$\vec{F} = \left| F_{hk\ell} \right| \cdot e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \cdot f_{n,\theta_{hk\ell}} \cdot e^{-B_n (\sin\theta_{hk\ell}/\lambda)^2} \cdot e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

Also, the calculated amplitudes can be used to evaluate the model!

Residuals, R-values, assess agreement between datasets. (here model vs experimental)

$$R_{cryst} = \frac{\sum \left| F_{obs} - F_{calc} \right|}{\sum \left| F_{obs} \right|}$$

But since R_{cryst} can be forced to appear good by warping the model -- the model must also be evaluated by other criteria. Also, a small subset of the data can be withheld from the refinement process. These data points can be used to calculate an R_{free} which should get better as Rcryst gets better as long as the model changes are really an improvement toward matching what the molecule really is.

R_{free} is a very valuable control against over-fitting.

R_{free} is usually just a few percentage points greater than R_{cryst}

Refinement

$$\vec{F} = |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \cdot f_{n,\theta_{hk\ell}} \cdot e^{-B_n (\sin\theta_{hk\ell}/\lambda)^2} \cdot e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

Model improvement involves (re)building a model into the electron density (real space "refiinement"), and shifting parameters to improve the fit of the calculated "structure factors" (data) to the observed data (reciprocal space refinement -- what is commonly called "Refinement Cycles" since the relationship of parameters to data is non-linear and is matched through a series of successive approximations).

The main target is to match the calculated |F| with that observed.

The parameters are x_n y_n z_n B_n (and O_n when needed). The molecule has a certain number of atoms, which sets the number of parameters, but the number of data points increases by the volume of reciprocal space that is measured. So at poorer resolutions there is a problem of numbers of parameters vs number of data points.

Geometrical target functions can be defined based on "previous knowledge" like bond lengths, bond angles, etc.

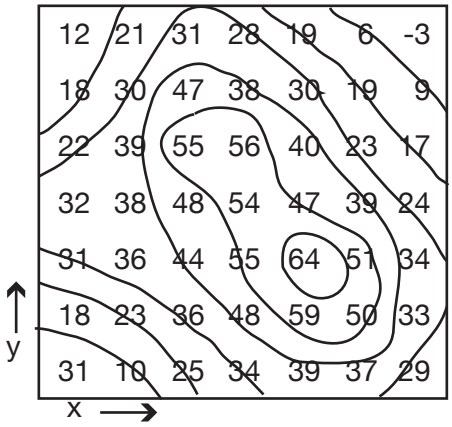
Even at the best (highest) (smallest-value) resolution, there can be regions where the electron density is weak and geometrical target functions are needed.

Refinement balances fit to data and fit to stereo-chemistry.

Electron Density Maps

Now that we know the ϕ_P for each reflection, we can use the Fourier transform formula to get a picture of the electron density in the molecule:

$$\rho_{xyz} = (\text{Vol})^{-1} \sum_{h}\sum_{k}\sum_{\ell} |F_{P\,hk\ell}|\, e^{i\phi_{P\,hk\ell}}\, e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

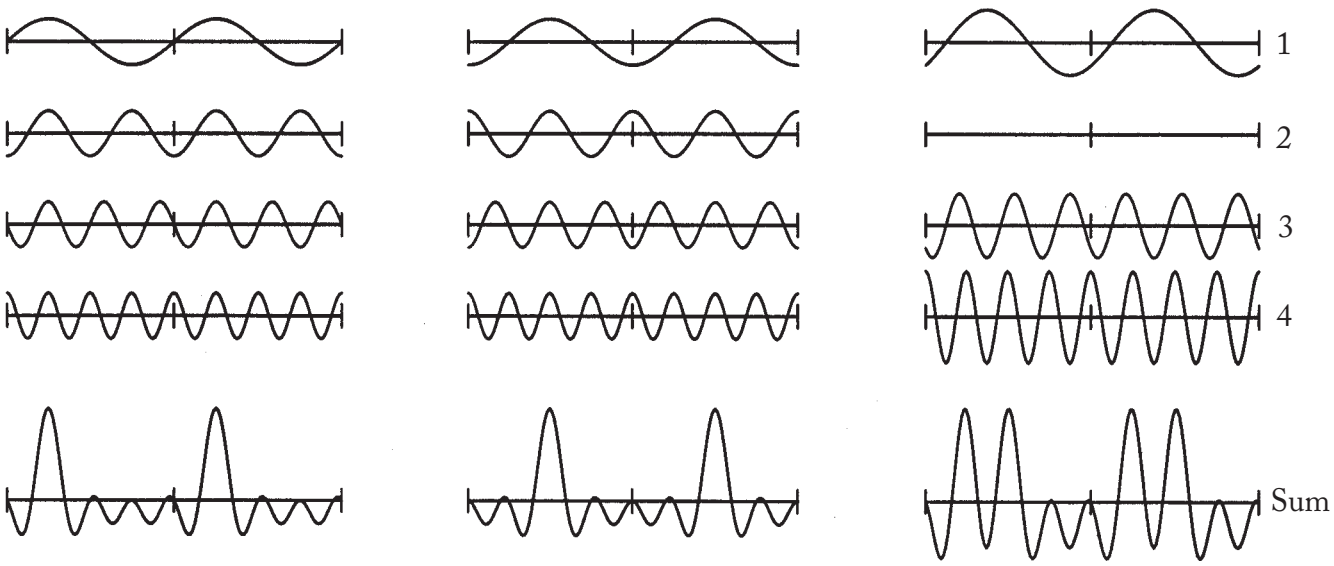


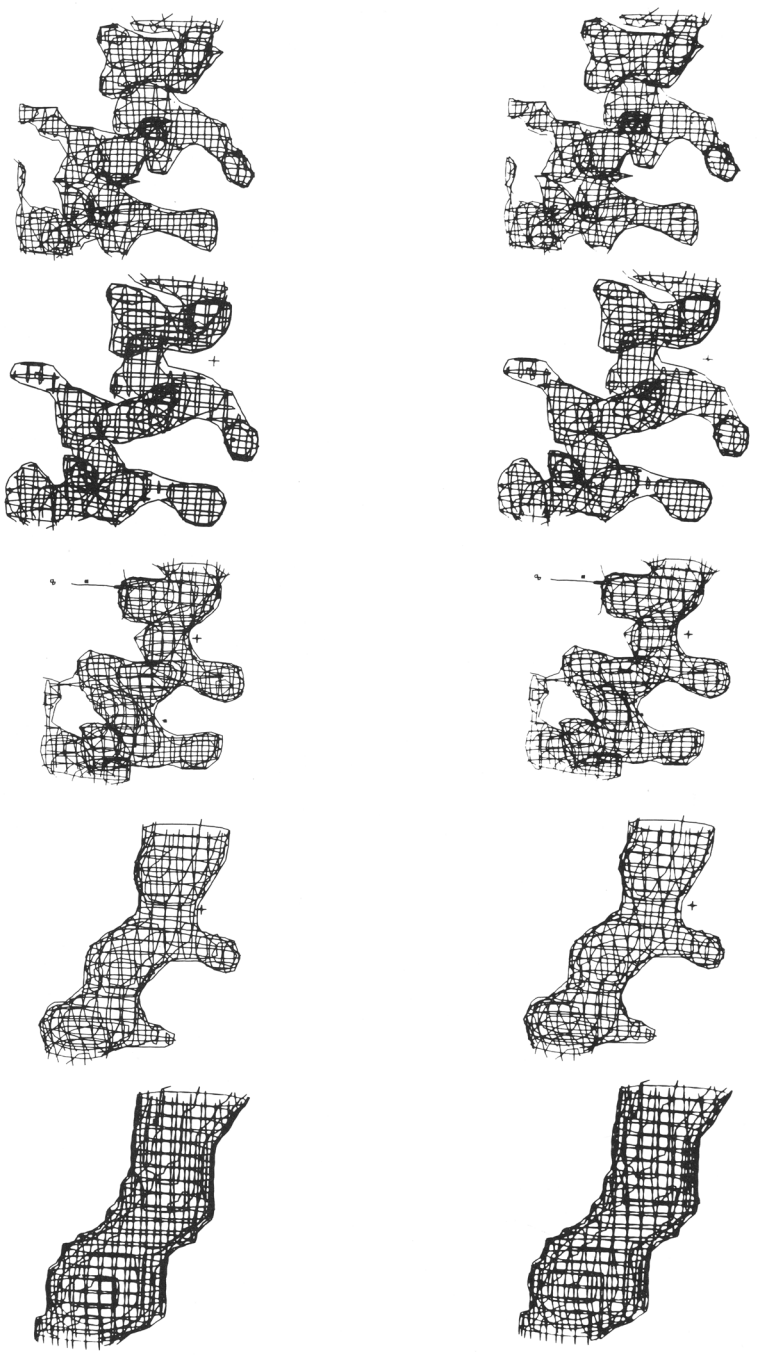
For each chosen grid point x,y,z in the repeating unit of the crystal, the above expression must be summed over all the measured reflections *h, k, ℓ*. The top illustration shows part of one layer of such a map, where each number is the value of the electron density (on an arbitrary scale) at that grid point. Contours have been drawn at 10, 20, 30, etc.

The middle illustration shows 7 superimposed layers of a part of the 2.5 Å resolution map of staphylococcal nuclease .

In the bottom illustration this same piece of map is shown interpreted as two residues of backbone in extended chain conformation, with arginine and phenylalanine side groups.

Example of summing waves of 1 , 2 , 3 , 4 wavelengths across a unit cell, of different amplitudes and different relative phases to get different number and positions of reconstructed “atoms”.

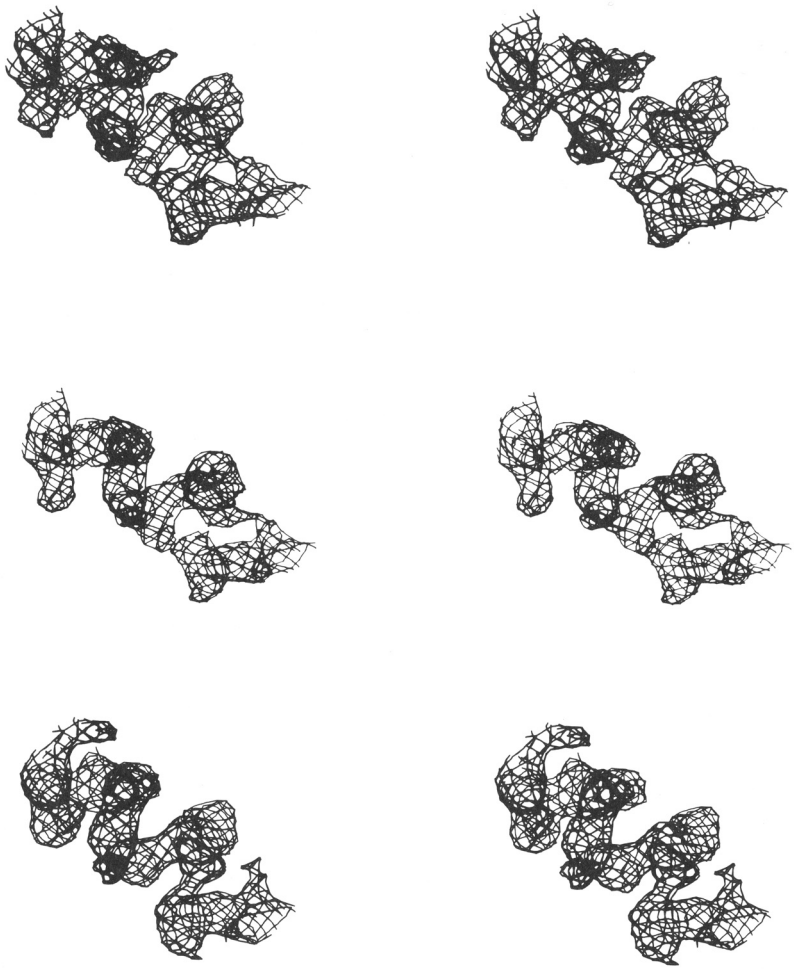




Stereo views of the electron density for an α-helix in staphylococcal nuclease at (from top to bottom) 2, 3, 4, 5, and 6 Å resolution. All maps were made using F_{obs}, the original MIR phases, and the same grid spacing. Viewpoint is the same, and contour levels were adjusted to be approximately equivalent.

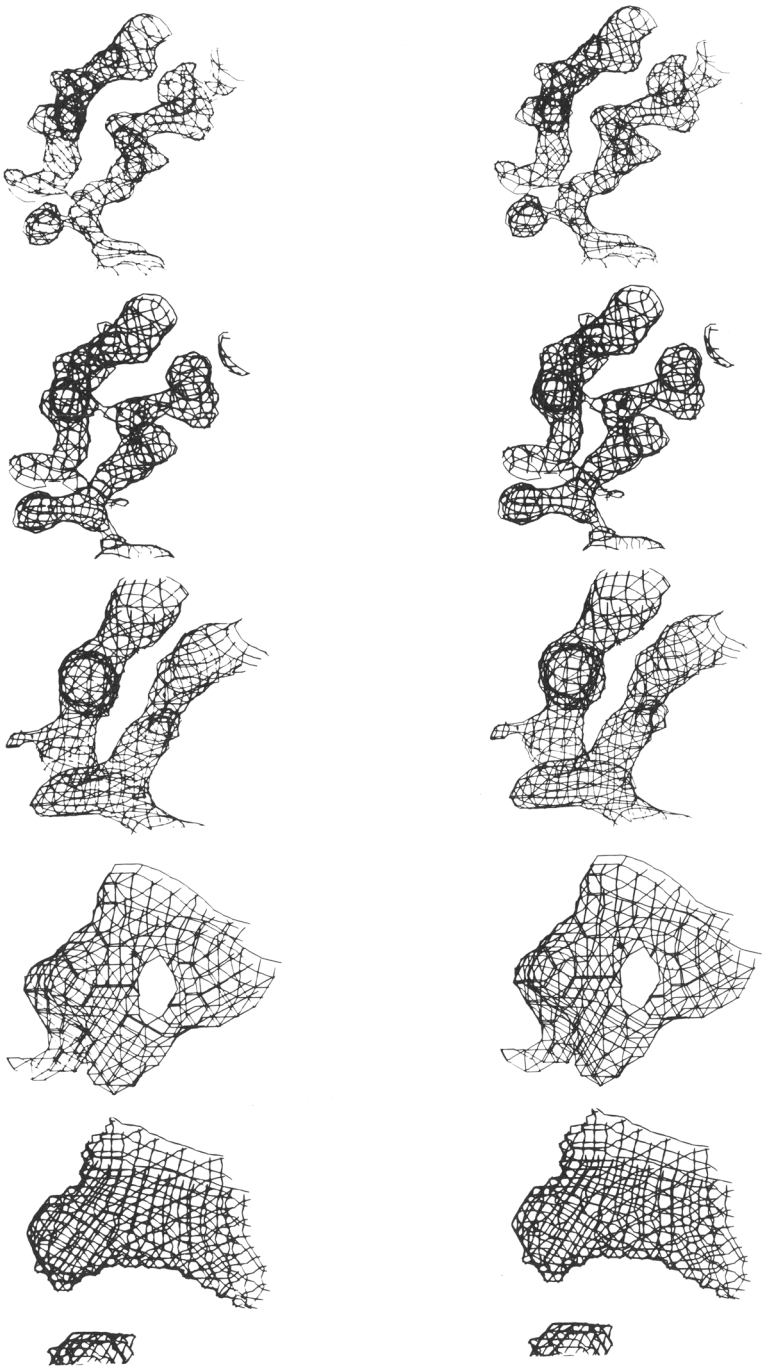
All carbonyl oxygens are clear at 2 Å, but almost all of them are absent at 3 Å, although side chains can still be judged. At 4 Å, density has begun to coalesce along the helix axis, and there is a false connection between side chains at the lower left.

This and the next two figures are from JS and DC Richardson (1985) "Interpretation of Electron Density Maps" in Methods in Enzymology; HW Wyckoff, CHW Hirs, SN Timasheff, eds.; **115**: 189-206.



Side views of the same helix as on previous page, at 2, 3, and 3.5 Å resolution.

At intermediate resolution the density connects through a hydrogen bond (lower right) more strongly than through the nearby helical main chain, although the connectivity is correct at both higher and lower resolutions.



Stereo views of the electron density for two strands of antiparallel β-sheet in staphylococcal nuclease at 2, 3, 4, 5, and 6 Å resolution.

In this case the strands separate correctly at 4 Å but that would not always be true. At 5 and 6 Å the density is sheetlike, but with holes in variable locations. At 6 Å the right-hand side extends further out because it is no longer separated from a third strand.

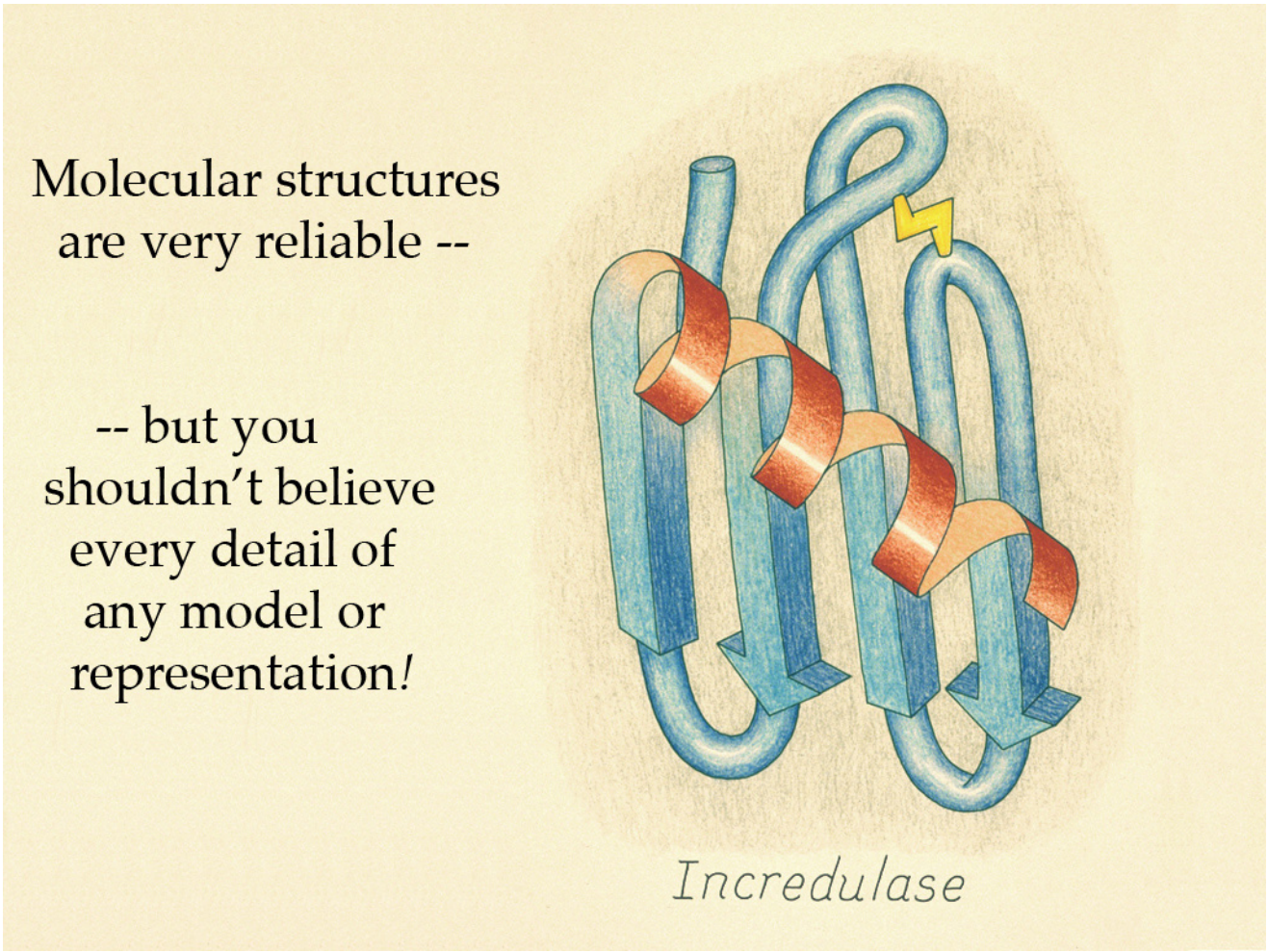
Quality Evaluation and Validation

Model evaluation and validation considers the same sort of factors used in building and refining the model, but uses different tools and often stricter criteria. Properties based on previous knowledge are used in model building and evaluation.

with respect to data
with respect to stereochemistry and physics
All Atom Contact Analysis and the MolProbity website

Kinemages HowDotsWork.kin
1JIRon1S83_Arg66_supr.kin, 1JIR.pdb in MolProbity

Quality indicators:
- Inherently Global:
  - Resolution
  - R_{cryst}
  - R_{free}
- Local in detail and global as averages
  - Bond lengths and angles, stereochemistry
  - B-factors
  - Match to electron density
  - Match with previous knowledge
    - Rotamers, Ramachandran, C-beta deviation, Suiteness
    - All-atom contact analysis



Diffraction Table

B-Factors and Resolution

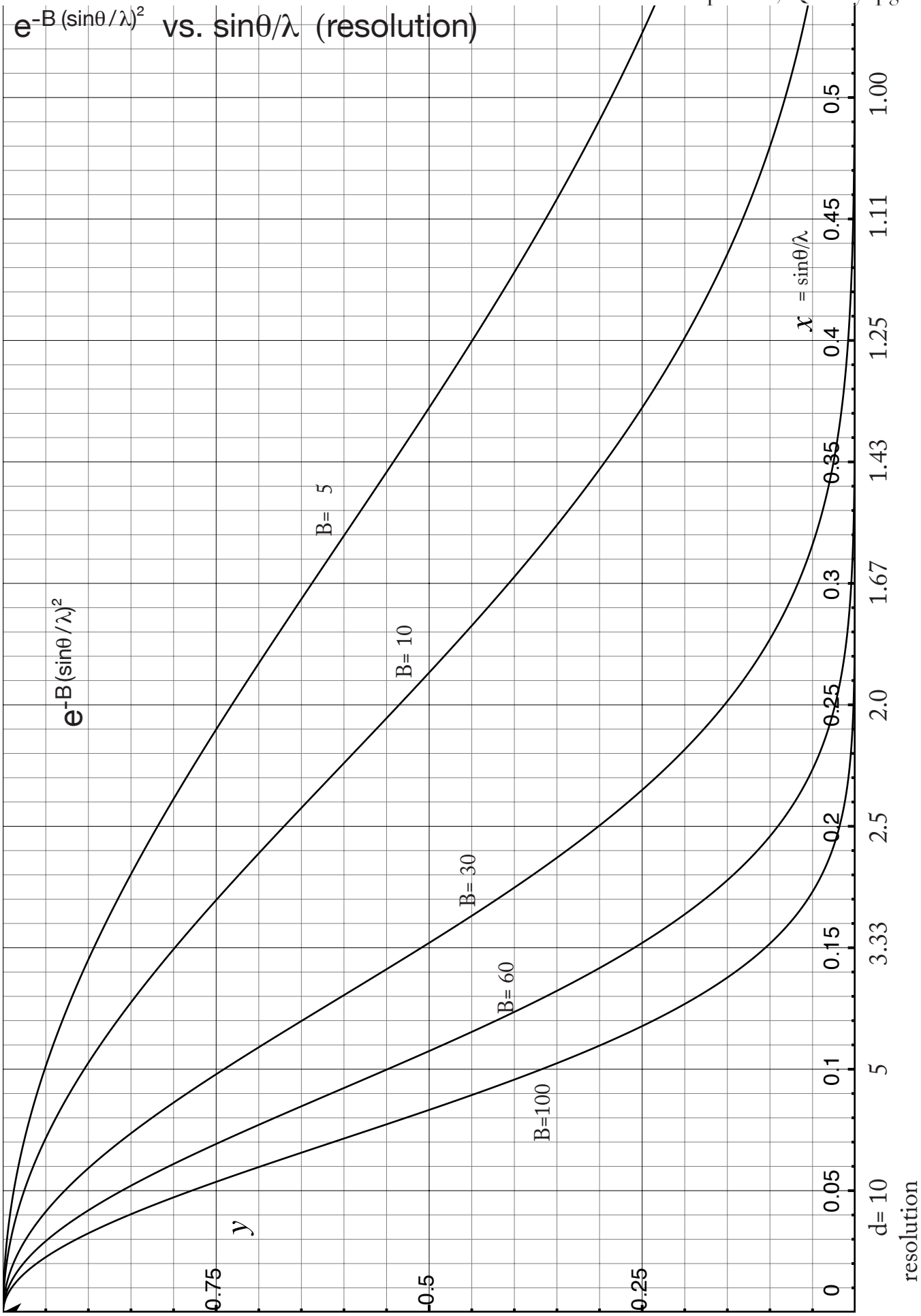
	(sin θ)/λ	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60	0.65
geometry	2θ (°)	0.00	8.84	17.74	26.74	35.92	45.34	55.10	65.32	76.15	87.87	100.87	135.36	180.00
	θ (°)	0.00	4.42	8.87	13.37	17.96	22.67	27.55	32.66	38.08	43.93	50.43	67.68	90.00
	d		10.00	5.00	3.33	2.50	2.00	1.67	1.43	1.25	1.11	1.00	0.83	0.77

B-Factors (disorder factors) expressed as $e^{-B(\sin\theta/\lambda)^2}$

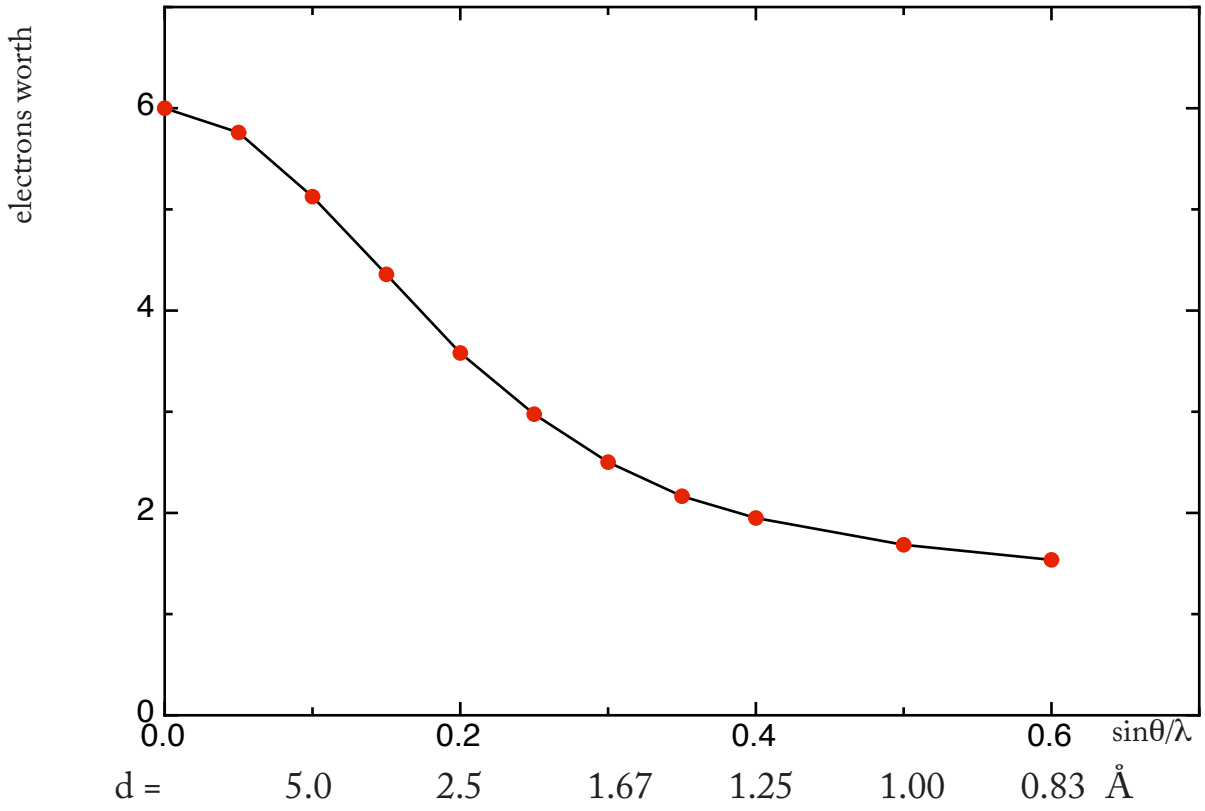
	(sin θ)/λ	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60	0.65
	$(\sin\theta/\lambda)^2$	0.0000	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500	0.3600	0.4225
B value	1		0.9975	0.9901	0.9778	0.9608	0.9394	0.9139	0.8847	0.8521	0.8167	0.7788	0.6977	0.6554
	10		0.9753	0.9048	0.7985	0.6703	0.5353	0.4066	0.2938	0.2019	0.1320	0.0821	0.0273	0.0146
	20		0.9512	0.8187	0.6376	0.4493	0.2865	0.1653	0.0863	0.0408	0.0174	0.0067	0.0007	0.0002
	30		0.9277	0.7408	0.5092	0.3012	0.1534	0.0672	0.0254	0.0082	0.0023	0.0006		
	40		0.9048	0.6703	0.4066	0.2019	0.0821	0.0273	0.0074	0.0017	0.0003			
	50		0.8825	0.6065	0.3247	0.1353	0.0439	0.0111	0.0022	0.0003				
	100		0.7788	0.3679	0.1054	0.0183	0.0019							

scattering factors for select atoms as a function of θ (λ = 1.542 Å); with some anomalous terms

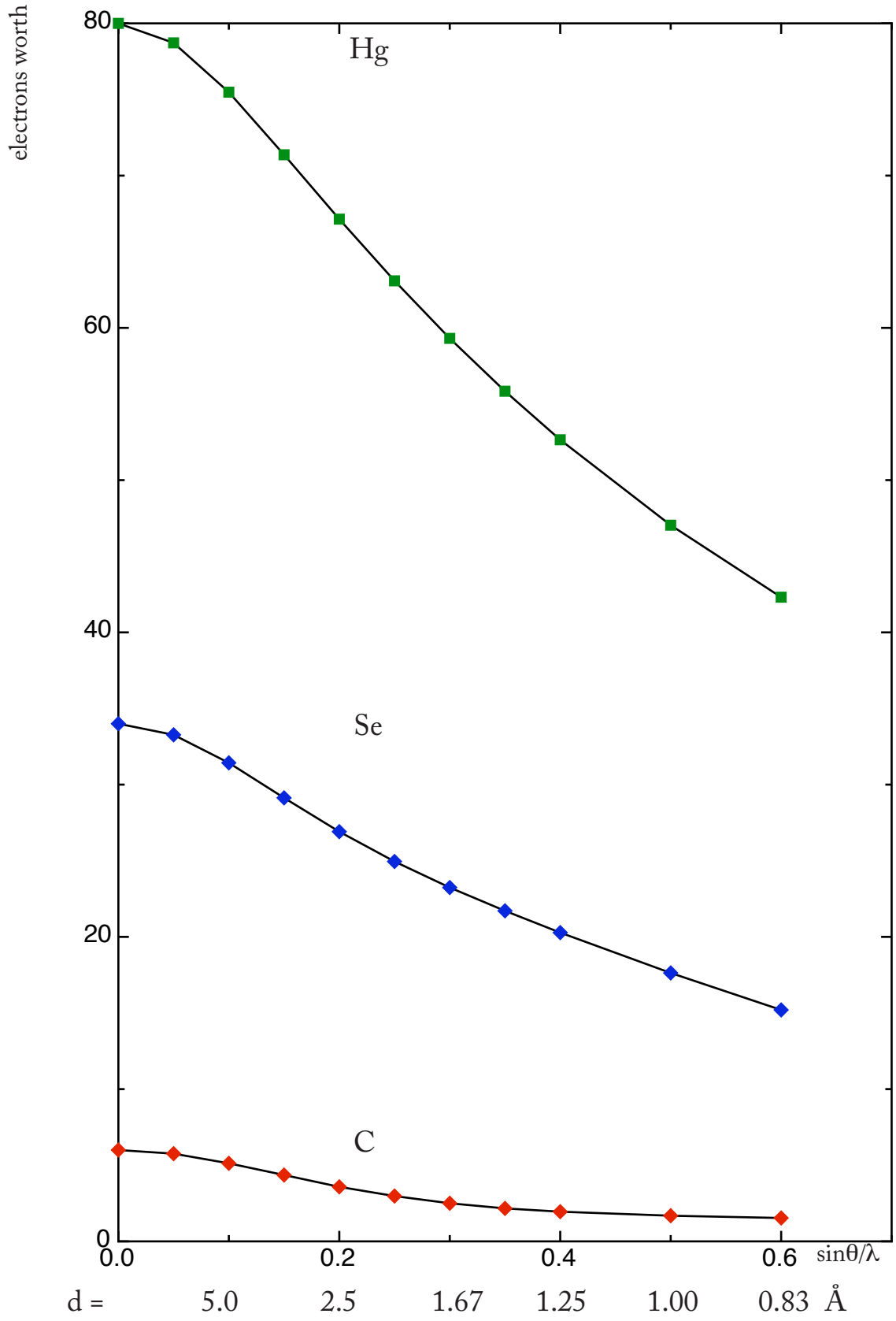
	(sin θ)/λ	0.00			0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40			0.45	0.50	0.60			0.65
			anomalous										anomalous					anomalous		
		f_n	f'	f''									f'	f''				f'	f''	
Atom Type	C	6.000	0.0	0.0	5.760	5.126	4.358	3.581	2.976	2.502	2.165	1.950	0.0	0.0		1.685	1.536	0.0	0.0	
	N	7.000	0.0	0.0	6.781	6.203	5.420	4.600	3.856	3.241	2.760	2.397	0.0	0.0		1.944	1.698	0.0	0.0	
	O	8.000	0.0	0.1	7.796	7.250	6.482	5.634	4.814	4.094	3.492	3.010	0.0	0.1		2.338	1.944	0.0	0.1	
	S	16.00	0.3	0.6	15.54	14.33	12.75	11.21	9.930	8.990	8.320	7.830	0.3	0.6		7.050	6.310	0.3	0.6	
	Se	34.00	-1.0	1.3	33.27	31.43	29.13	26.91	24.95	23.24	21.71	20.28	-1.0	1.2		17.63	15.20	-1.1	-1.1	
	Hg	80.00	-5	9	78.72	75.48	71.37	67.14	63.09	59.31	55.84	52.65	-5	8		47.04	42.31	-5	8	


e-B (sinθ / λ)2 vs. sinθ/λ (resolution)
e-B(sinθ / λ)2
B= 5
B= 10
B= 30
B= 60
B=100
y
0.75
0.5
0.25
x = sinθ/λ
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
d= 10
5
3.33
2.5
2.0
1.67
1.43
1.25
1.11
1.00
resolution

Carbon scattering factor vs. resolution



Hg, Se, C scattering factor vs. resolution



Bragg's Law

$$n\lambda = 2d_{\text{unit cell}} \sin(\theta)$$

$$\lambda = 2d_{h,k,l} \sin(\theta_{h,k,l})$$

smallest $d_{h,k,l}$ is resolution

$$\frac{1}{d_{h,k,l}} = \frac{2 \sin(\theta_{h,k,l})}{\lambda} \quad \textbf{so} \quad \frac{\sin(\theta_{h,k,l})}{\lambda}$$ **is a measure of resolution**

Scalar value of amplitude equals the square root of the intensity (energy) of the wave.

$$|F_{hk\ell}| = (I_{hk\ell})^{1/2} \qquad I_{hk\ell} = |F_{hk\ell}|^2$$

"Intensity" has been corrected for polariztion and other consequences of data collection method.

A diffracted ray is the sum of contributions from all atoms.

$$\vec{F} = |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \cdot f_{n,\theta} \cdot e^{-B_n (\sin\theta / \lambda)^2} \cdot e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

A diffracted ray is the sum of contributions from all electron density.

$$|F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \frac{\text{Volume}}{\left(\begin{smallmatrix}\text{of repeat-}\\ \text{ing unit}\end{smallmatrix}\right)} \sum_{x}\sum_{y}\sum_{z} \rho_{xyz} \cdot e^{i\,2\pi(\,hx + ky + \ell z\,)}$$

Electron density is the Fourier transform of all diffracted rays.

$$\rho_{xyz} = (\text{Vol})^{-1} \sum_{h}^{-\infty\rightarrow\infty}\sum_{k}^{\infty}\sum_{\ell}^{\infty} m_{hk\ell} \cdot |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} \cdot e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

Patterson map is the Fourier transform of the intensities.

$$P_{xyz} = (\text{Vol})^{-2} \sum_{h}^{-\infty\rightarrow\infty}\sum_{k}^{\infty}\sum_{\ell}^{\infty} |F_{hk\ell}|^2 \cdot e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

product of all electron densities separated by the x,y,z vector distance
same shape and dimension unit cell

Residuals, R-values, assess agreement between datasets. (here model vs experimental)

$$R_{cryst} = \frac{\sum \big|\, |F_{obs}| - |F_{calc}| \,\big|}{\sum |F_{obs}|}$$

R_{free} calculated from otherwise unused 5%

λ	:	wavelength of radiation
d	:	interplanar spacing, the effective distance associated with a particular diffracted ray
$\theta_{hk\ell}$	:	angle of incident beam to the *h, k, ℓ* Bragg plane
h, k, ℓ	:	integer index numbers of a particular Bragg Plane, a diffracted ray, a "reflection", index of a point of the reciprocal lattice "reciprocal space"
$\|F_{hk\ell}\|$	:	amplitude of the $hk\ell^{th}$ diffracted ray
$\phi_{hk\ell}$	:	the phase of the $hk\ell^{th}$ diffracted ray
B	:	B-factor (historically the Temperature Factor, but dominated by other uncertainties)
$f_{n,\theta_{hk\ell}}$	:	individual atomic scattering factor of the n^{th} atom as a function of $\theta_{hk\ell}$
x_n, y_n, z_n	:	coordinates of the n^{th} atom
$m_{hk\ell}$	:	figure of merit for phase of the $hk\ell^{th}$ diffracted ray
N	:	total number of atoms in the repeating unit of the crystal
O_n	:	occupancy of the n^{th} atom.
Vol	:	volume of the repeating unit, the unit cell
ρ_{xyz}	:	electron density at coordinates x, y, z in the crystal "real space"
P_{xyz}	:	Patterson function value at coordinates x,y,z : vector distances-between atom positions

The equations: (Good-Parts version: what you really need to remember)

Bragg's law (as used to track each individual scattered ray):

$$\lambda = 2d_{h,k,l} \sin(\theta_{h,k,l})$$

Remember this equation!

Note the reciprocal relationship between $d_{h,k,l}$ and $\theta_{h,k,l}$

$$\frac{1}{d_{h,k,l}} = \frac{2 \sin(\theta_{h,k,l})}{\lambda}$$

d: distance, but direction only implied!

$d_{h,k,l}$ is a particular indexed distance in the crystal, and is the "resolution" of the *h,k,l* diffracted ray. The smaller the d, the better the resolution, thus the best resolution data is from diffracted rays with the largest angles. The index carries a lot of the "direction" information.

These "d"s are distances between "Bragg Planes". Bragg Planes are simply ways of dividing the unit-cells of the crystal into equally spaced intervals (the "d" spacing). So Bragg Planes are oriented families of parallel planes with a particular separation. The *h,k,l* indices, along with the unit cell dimensions and shape, define both the orientation and the d-spacing of a particular set of Bragg Planes.
(The unit cell is the basic repeating unit of the crystal: the repeat is done entirely by translations, and the unit cells "tile" (completely fill) the volume of the crystal.)

The "θ" is the angle the ray makes to the Bragg Plane, and in the defining construction the angle of the incident ray is equal to the angle of the diffracted ray, just like a classical reflection, so the diffracted rays are often called "reflections".

CAVEAT:
When discussing crystallography, many people remember Bragg's law this way:

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

more specifically: $n\lambda = 2d_{\text{unit cell}} \sin(\theta)$

where λ is the wavelength of the x-rays, d is a dimension of the unit cell of the crystal, and θ is the angle of the scattered ray from "Bragg Planes", and the integer n is the number of wavelengths of path difference between a ray scattered from one side of the unit cell and a ray scattered from the opposite side of the unit cell.

The "n" formula brings to mind vibrating string harmonics. Indeed, we will see how the electron density can be thought of as a sum of standing waves in the box of the unit cell.

This n, the order of the diffraction, is useful for fiber diffraction and is reasonably convenient for small molecule crystals with very small unit cells, BUT can be very confusing when thinking about diffraction from macromolecular crystals. (For convenience, the examples drawn in text books and lecture notes show small unit cells, so drawings with the "n" form seem to make sense. However, in real life, or with general equations, dealing with large unit cells, everything is done in terms of uniquely indexed Bragg planes, i.e. uniquely indexed diffraction events.).

The equations: (Good-Parts version: what you really need to remember, cont'd.)

A light wave ray diffracted from a molecule in a crystal:

A wave is defined by an Amplitude $|F_{hk\ell}|$ and a Phase: $e^{i\phi_{hk\ell}}$

Each diffracted "wave" is a sum of contributions from ALL atoms:

$\sum_{n}^{N}$ (Amplitude-factors) • (Phase-factor) ...the sum is over all atoms.

Amplitude-factors are a property of the atom, including uncertainty about its position.

(Uncertainties are put into the B-factor. Many people think they understand the B-factor because they remember hearing about the Temperature Factor from Physics, for macromolecules the temperature part of the B-factor is only a minor part, most of the effect comes from other kinds of disorder.)

Phase-factor is just dependent on the position of the atom.

And now the equation we will derive (and that crystallographers should be familiar with):

Remember this equation for your prelim!

Amplitude Phase = (Amplitude-factors) • (Phase-factor)

$$\vec{F} = |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \cdot f_{n,\theta_{hk\ell}} \cdot e^{-B_n (\sin\theta_{hk\ell}/\lambda)^2} \cdot e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

h, k, ℓ	:	integer index numbers of a particular Bragg Plane, a diffracted ray, a "reflection".
$\theta_{hk\ell}$	:	angle of incident beam to the h, k, ℓ Bragg plane
$\|F_{hk\ell}\|$	:	amplitude of the $hk\ell^{th}$ diffracted ray
$\phi_{hk\ell}$	:	the phase of the $hk\ell^{th}$ diffracted ray
B	:	B-factor (historically the Temperature Factor, but dominated by other uncertainties)
$f_{n,\theta_{hk\ell}}$	:	individual atomic scattering factor of the n^{th} atom as a function of $\theta_{hk\ell}$
x_n, y_n, z_n	:	coordinates of the n^{th} atom
N	:	total number of atoms in the repeating unit of the crystal
O_n	:	occupancy of the n^{th} atom. (e.g. for a half-occupied ligand, sidechain alternate rotamer, etc.)

The equations: (Good-Parts version: segue into a trick...)

The wave equation depends on a model of the structure (i.e. atomic coordinates) ---
but that is the result of the process, the whole problem is that we don't know any of the coordinates in the beginning! We really want to run that equation backwards and calculate the model from the diffracted waves. The trick is to cast the equation into the form of a Fourier Transform equation which can be reformulated backwards.

So, take this equation:

A diffracted wave is the sum of contributions from all atoms.

$$\vec{F} = |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \sum_{n}^{N} O_n \cdot f_{n,\theta_{hk\ell}} \cdot e^{-B_n (\sin\theta_{hk\ell}/\lambda)^2} \cdot e^{i\,2\pi(\,hx_n + ky_n + \ell z_n\,)}$$

And recast it in terms of continuous density instead of individual atoms:

A diffracted wave is the sum of contributions from all electron density.

$$|F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} = \left(\begin{matrix}\text{Volume}\\ \text{of}\\ \text{unit cell}\end{matrix}\right) \sum_x \sum_y \sum_z \rho_{xyz} \cdot e^{i\,2\pi(\,hx + ky + \ell z\,)}$$

And now we can work it backwards: the **Trick**

Electron density is the Fourier transform of all diffracted waves.

$$\rho_{xyz} = (\text{Vol})^{-1} \sum_{h}^{-\infty \rightarrow \infty} \sum_{k} \sum_{\ell} m_{hk\ell} \cdot |F_{hk\ell}| \cdot e^{i\phi_{hk\ell}} \cdot e^{-i\,2\pi(\,hx + ky + \ell z\,)}$$

Amplitude = (Amplitude-factors) • (Phase-factors)
(electron density has no phase) here the x,y,z is the point in the unit cell where we are calculating the density, if we do this for the whole unit cell, then we'll see everything that is possible to see.

So, can we do it? What is known about the factors on the right side of the equation?
$m_{hk\ell}$ is the figure of merit for phases, we'll get that when we figure out the phases.

$|F_{hk\ell}|$ is the amplitude of the $hk\ell^{th}$ diffracted wave, $|F_{hk\ell}| = (I_{hk\ell})^{1/2}$ and $I_{hk\ell}$ is the Intensity of the diffracted wave, which we can measure!

$e^{-i\,2\pi(\,hx + ky + \ell z\,)}$ is easy: we know h,k,l for each wave and the x,y,z of the point being calculated.
$e^{i\phi_{hk\ell}}$ is the hooker, we do NOT know $\phi_{hk\ell}$ the phase for the $hk\ell^{th}$ diffracted wave.

So now we have to do some work -- find ways to recover the phase of each diffracted wave that was lost when we measured just the intensity.

Two kinds of waves: these last big equations are of a form:
(Amplitude factor) • (Phase factor) Which is just the general equation of a wave.
However, there are two varieties:

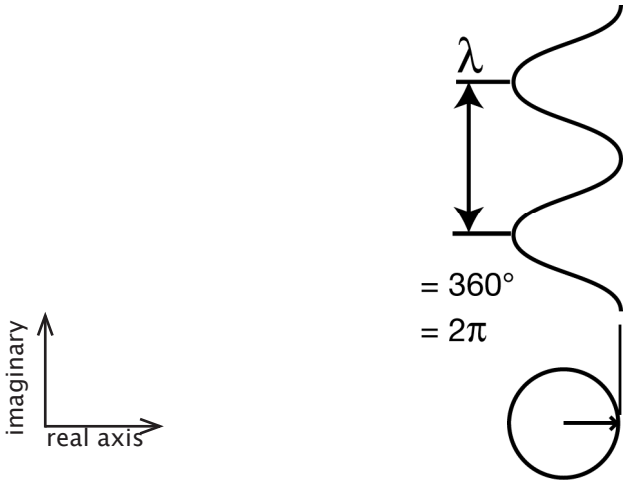
1) $\left|F_{hk\ell}\right| \cdot e^{i\phi_{hk\ell}}$ the expression for a real x-ray, with a real, experimentally fixed, wavelength.

A resultant diffracted x-ray wave from the crystal is the sum of x-ray waves scattered from each and every atom in that crystal in a particular direction. Each resultant wave, indexed as *h,k,l* , travels out of the crystal in that particular direction, so we will need to learn how to combine parallel x-ray waves to form a resultant wave.

2) (amplitude factor)• $e^{-i\,2\pi(\,hx + ky + \ell z\,)}$ some other kind of wave, with wavelengths that turn out to be integral fractions of the dimensions of the unit cell.

The electron density in a model of the crystal is the sum of these second kind of waves. Not only are the wavelengths of these density waves different from each other, each wave is going in its own particular direction. The wavelengths are integral fractions of unit cell dimensions (i.e. 1,2,3,... complete cycles within the bounds of the unit cell), thus they are standing waves. So we will need to learn how to combine standing waves in a box (the unit cell) to build up a density-like image.

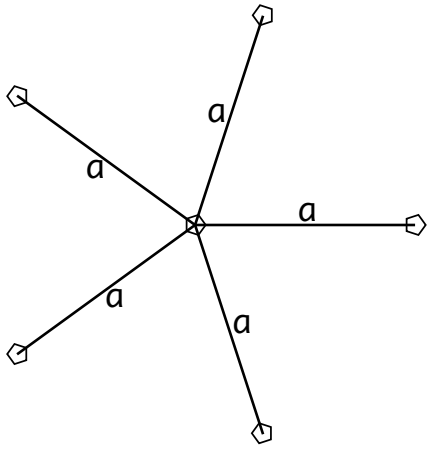
Representing waves, the phase clock (with radius = amplitude):



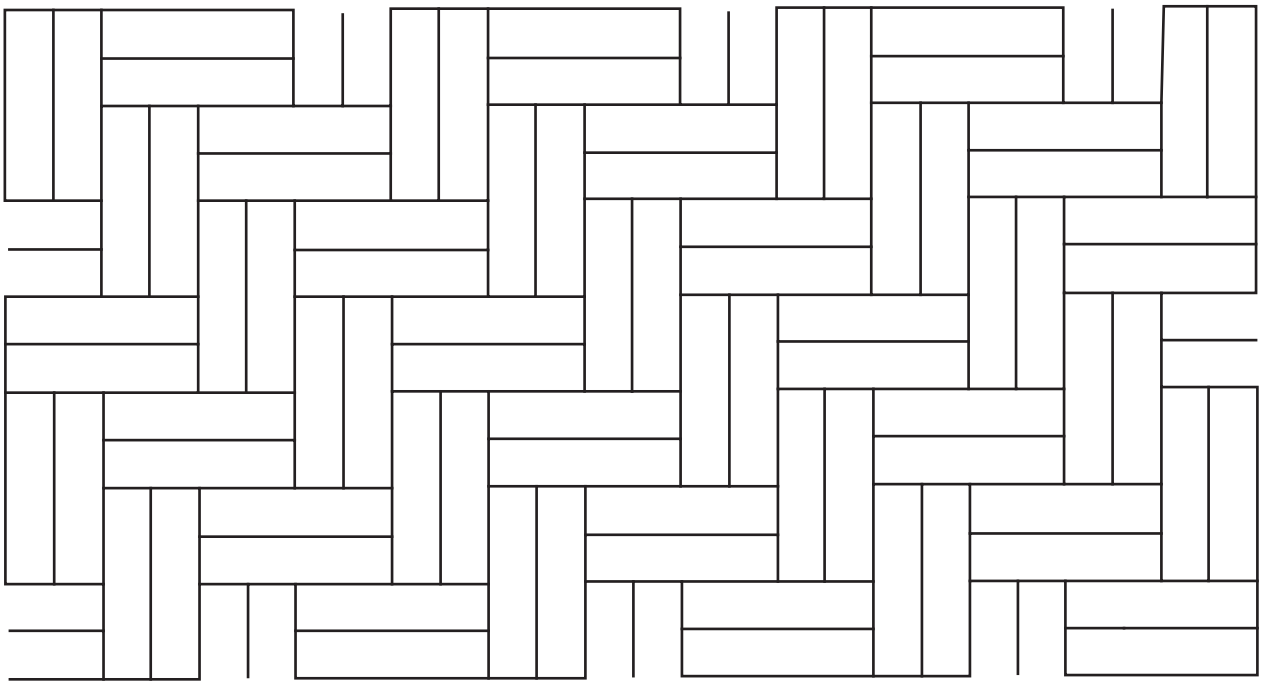
ϕ factor ($e^{i\phi}$): exponential form convenient to talk about; $e^{i\phi} = \cos(\phi) + i\,\sin(\phi)$
cos() & sin() form (real and imaginary components) sometimes more convenient for computation.

Show on this diagram why a 5-fold axis is incompatible with a crystal of translationally related Unit cells.

Explain briefly why.



On this diagram of a two dimensional crystal draw a sensible choice for the unit cell. Explain why your choice is a unit cell.



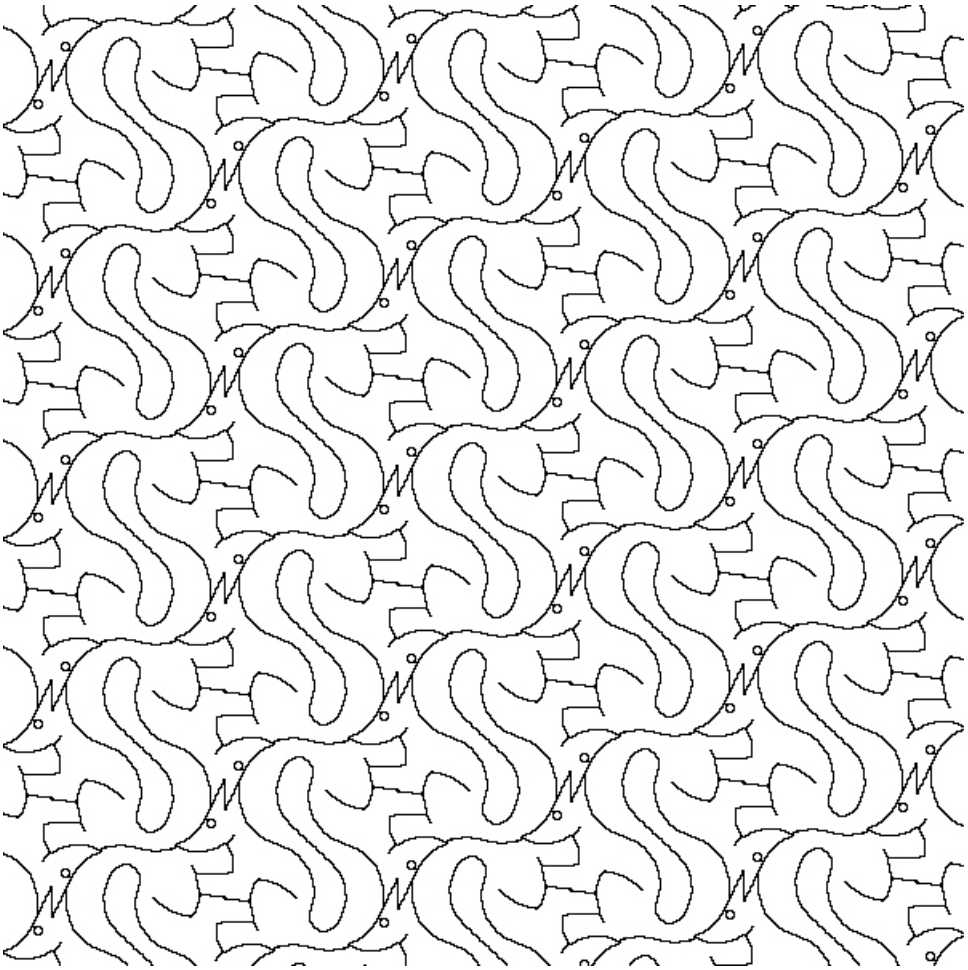
 Name:_____

This 2-D drawing shows a crystal of squirrels:

a) MARK at least one of each of the 4 distinctly different squirrel 2-fold axes: Use different symbols for each type axes (i.e. that touch different parts of the squirrels that are related by those 2-fold axes.).

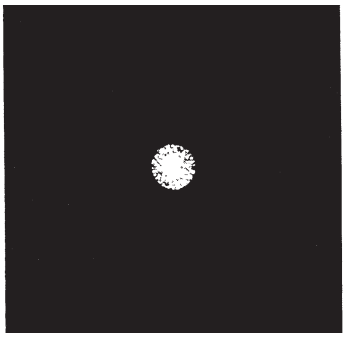
b) DRAW a rhombus-shaped unit cell with the MINIMUM amount of squirrel that can be translated to completely cover the space.

c) DRAW another unit cell, a rectangle, that holds TWICE as much but has right angles at the corners.



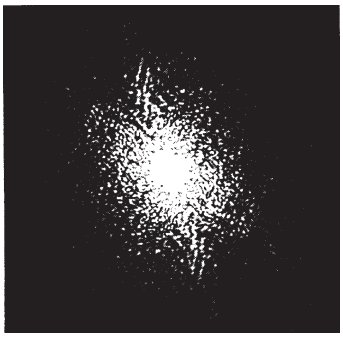
Match Transforms

Match the diffraction patterns in the top row by number to the images on the next row.

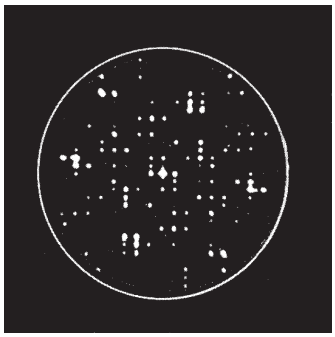


(1)

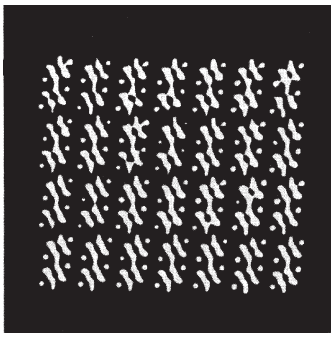
(2)



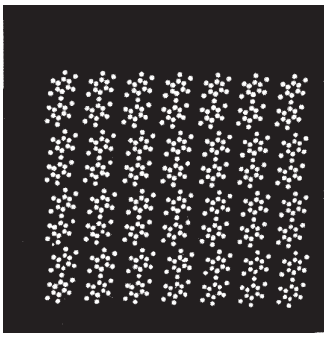
(3)



(4)



(a)



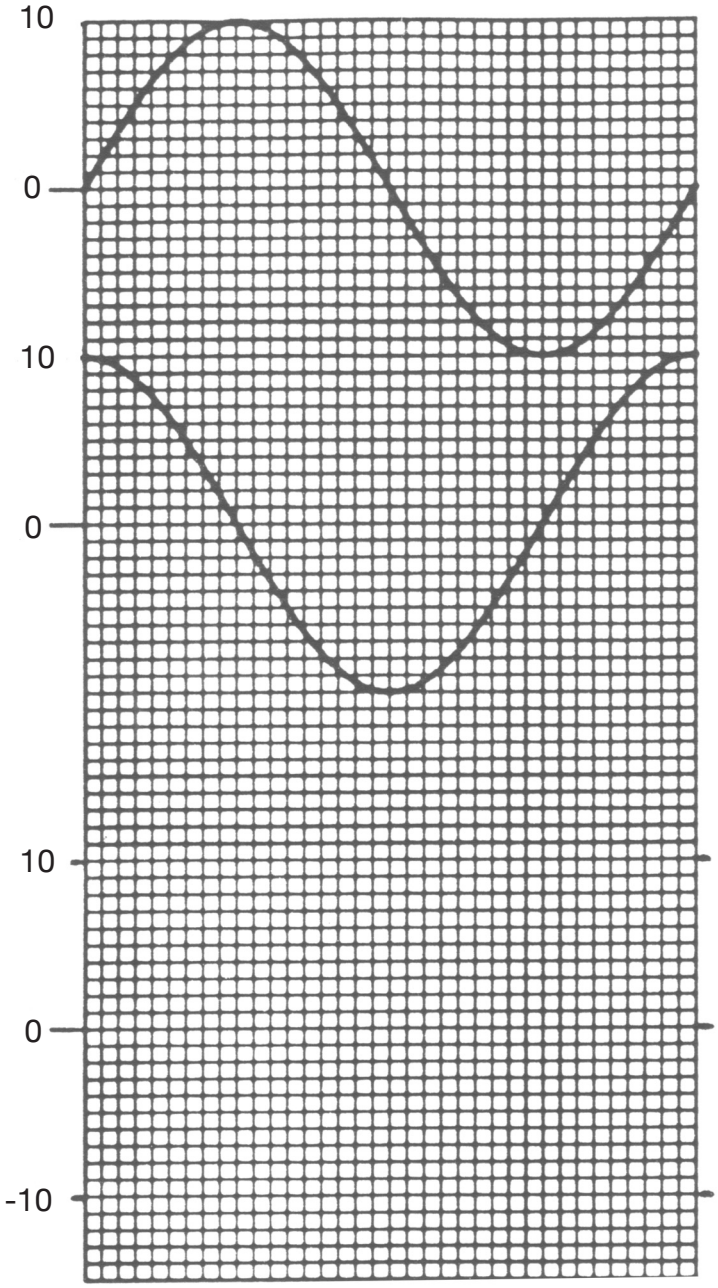
(b)

(c)

(d)

Discuss a common theme to this pairing. What principle is being illustrated by the pair of crystal images and the pair of Mickey images ?

Adding Waves:
Sum the two curves "point-by-point" using the empty graph space below the two.



3-2. Bragg's Law limits

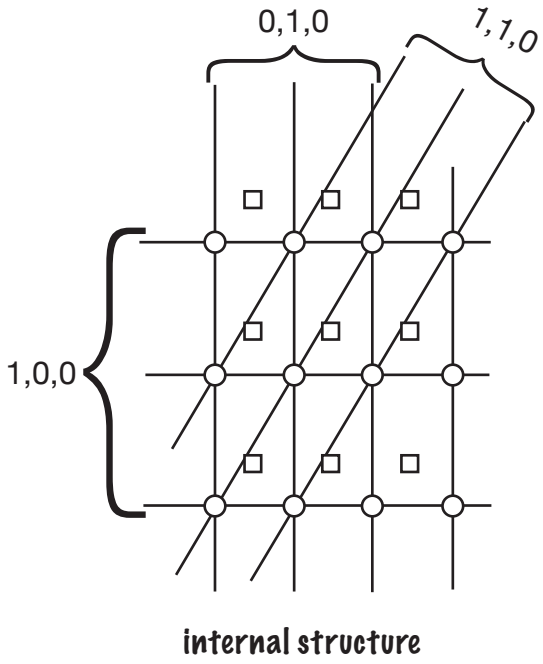
For X-rays of wavelength A = 1.5 Å, what is the smallest Bragg plane d spacing which could produce a diffracted ray? In what direction with respect to the incident beam is this diffracted ray going?

Make a drawing to show this, including direction of incoming and "reflected" rays.

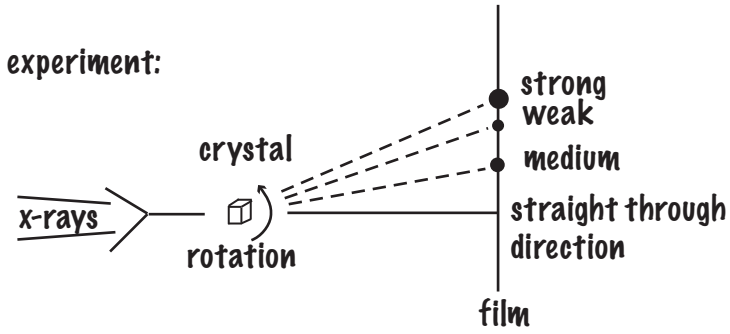
What is the longest wavelength of x-rays which will diffract from a crystal with largest Bragg plane d spacing = 100 Å ?

3-3Planes-->3Spots

As one rotates the crystal shown below, three of the spots on the film closest to the straight-through direction are produced by diffraction from the three indicated sets of Bragg planes. Identify them (that is, match the Bragg planes with the spots on the film). Explain your assignments including reasons for the relative strengths of the spots and for the relative positions of the spots on the film.



internal structure

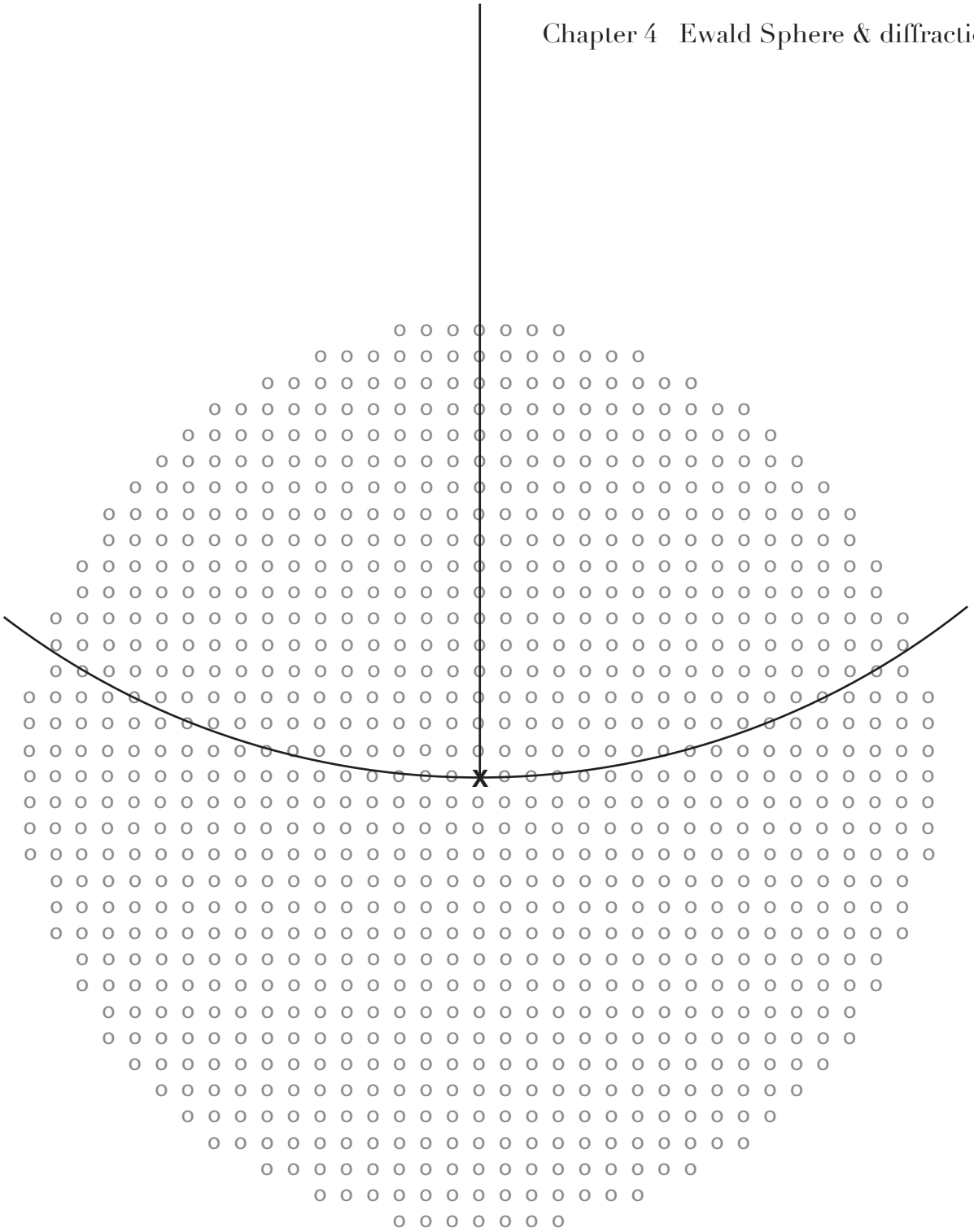


Reciprocal Lattice & Diffraction from a protein crystal

On the next page. a crystal of orthogonal unit cell, 50x50x50 Angstroms, is at the 'X' in the center of its reciprocal space. The spots representing the ends of diffraction vectors are approximately a realistic size, accounting for wavelength spread, source size, crystal size, and the mosaic character of a real crystal. A beam of x-rays is shown hitting the crystal, and a circle is drawn that shows the location of solutions to the parallelogram constructions (Ewald sphere) for this orientation of x-ray beam and crystal.

a) Show on the film plane where x-rays would hit, that come from **all** of the Bragg planes that are effectively in diffracting condition as shown by this diagram. (Compass and straight edge needed for these constructions.)

b) What is the highest resolution, calculated in Å (i.e. spacing of the closest spaced Bragg planes) represented on this slice of reciprocal space, whether or not that plane is actually diffracting at this time? (Mark the reciprocal space spot to identify the one you choose.)



film plane

Helix-spacing Problem

For the crystal shown below:

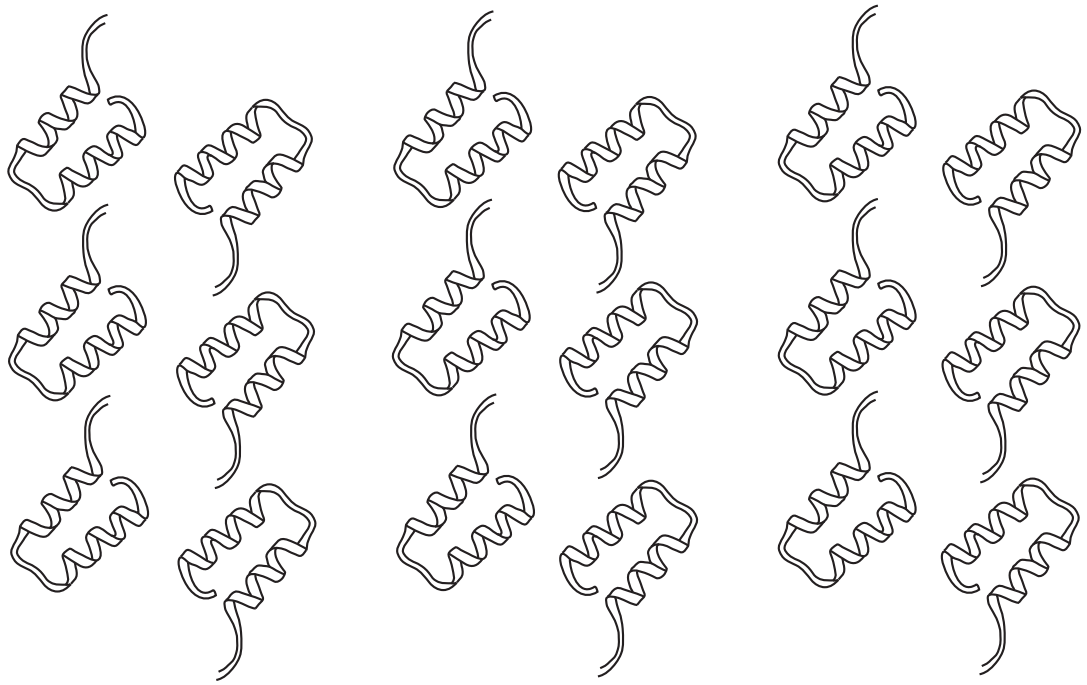
a) Draw a unit cell. Actually tiling the area of molecules shown below with your choice of unit cell will help think about how to draw the corresponding reciprocal lattice.

b) Draw the reciprocal lattice out to the 5th spot in each direction, with approximately the correct relative spacing. The origin of reciprocal space does not have to be in the crystal, in any case, draw the reciprocal lattice down below the real-space crystal lattice. However, it is useful to line up the lattices so that directions are correlated.

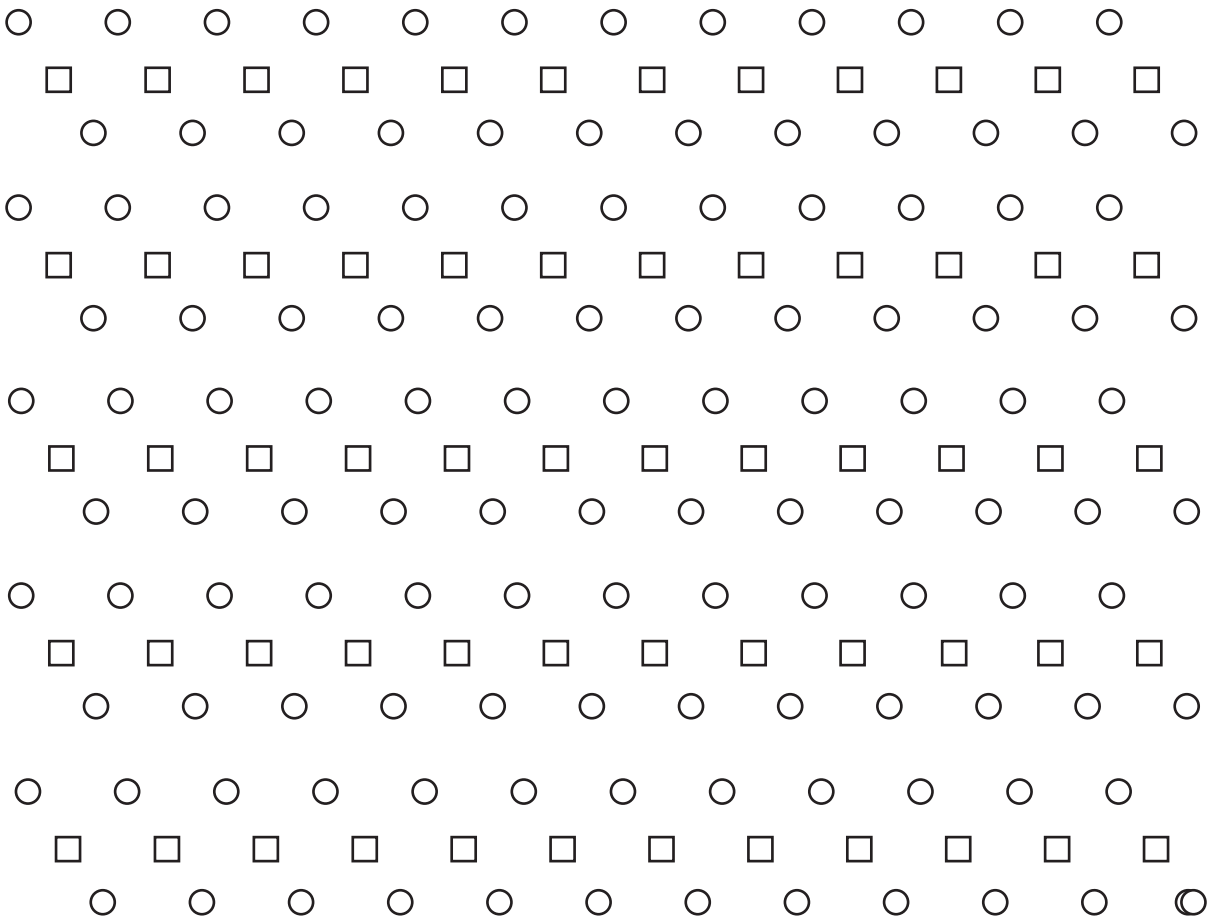
c) Circle a spot that is especially strong because of the side-to-side helix packing in this crystal.

(Draw in the corresponding Bragg Planes in the real space area.)

USE A RULER TO DO YOUR DRAWING:
SPACING, RELATIVE PROPORTIONS AND ORIENTATION IS IMPORTANT.



1. The pattern below represents a two-dimensional crystal of a linear three atom molecule (like O=C=O). Label clearly, using separate parts of the pattern for each:
   a) A reasonable unit cell
   b) A set of Bragg planes with the largest possible d spacing.
   c) A set of Bragg planes that would give a very strong diffracted ray.
   d) A set of Bragg planes that would give a relatively weak diffracted ray.



Below is a very similar drawing to that of the general case of molecular scattering on pg 3 of chapter 5. However, here the point P is part of a molecule embedded in a unit cell of a crystal, and a special direction of view down the x axis is used to simplify the diagram.

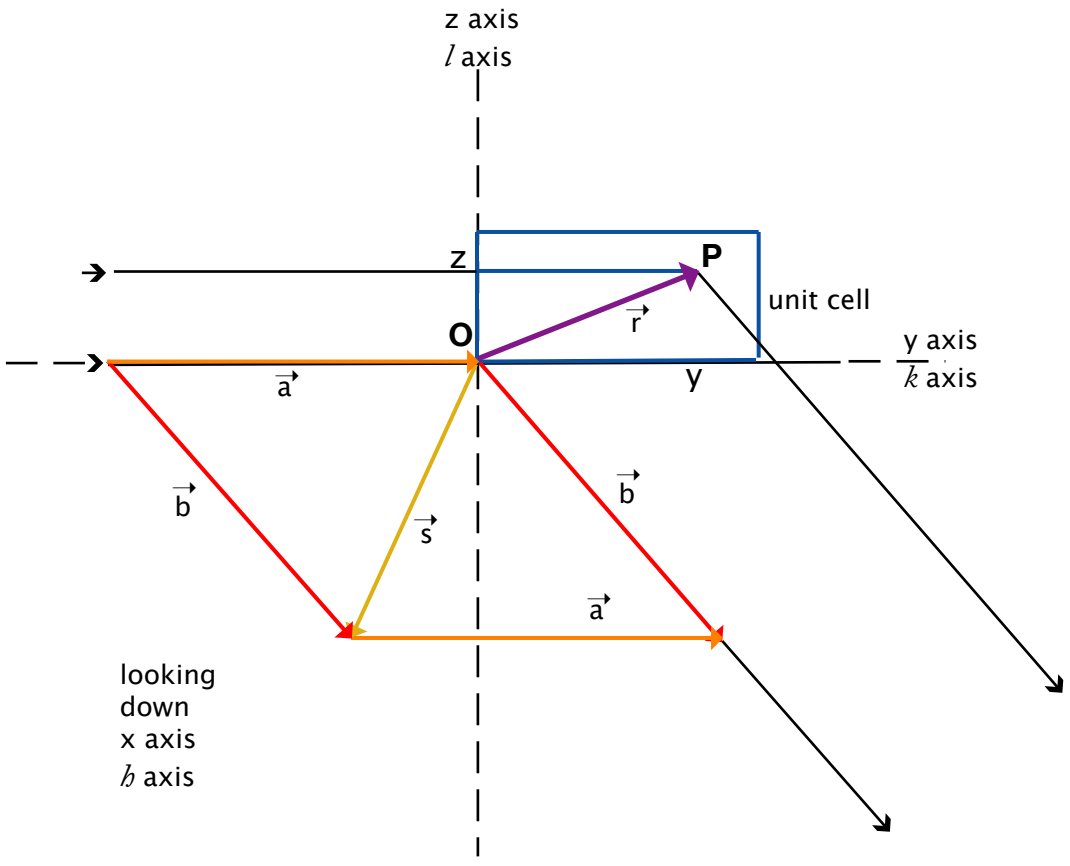
A crystal can only diffract in special directions, so only then can the point P be contributing to a diffracted ray.

a) Show a set of Bragg planes that determining this diffraction event. Indicate θ in and θ out.

Draw carefully: the construction should be easy and show obvious answers.

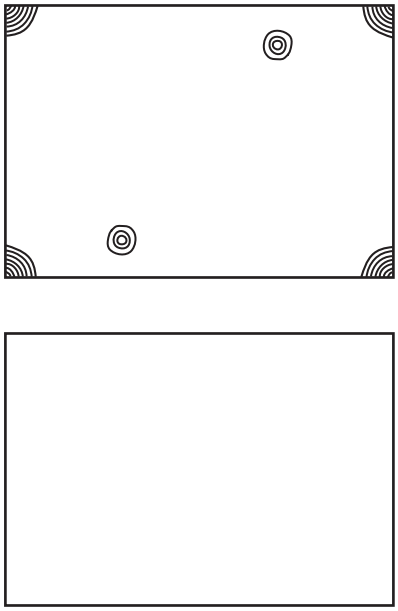
b) What are the indices (h, k, ℓ) of this Bragg plane?

c) For this reflection, approximately what is the phase of the ray from P with respect to a ray of phase = 0 from the Origin (O)?



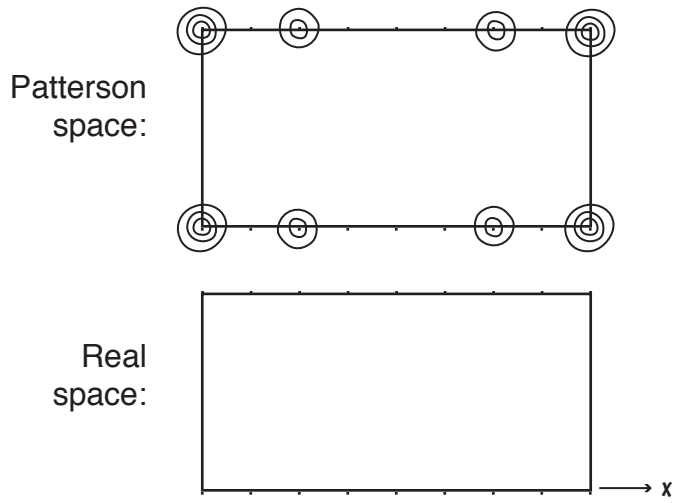
Patterson Maps

1. Since your protein has a single free SH group you try soaking the crystals in solutions of mercury compounds. Data from one of them gives you the simple Patterson map shown below. On the unit cell shown empty, draw in a possible set of heavy-atom positions that would satisfy the Patterson map.



Patterson Maps, cont'd

2. Solve the Patterson map shown below; put one atom at the origin in real space.



If one of the atoms is at the origin, explain why for the 1,0,0 diffracted ray the second atom would give a phase of 90°.

If atom one and atom two are the same, draw a phase-vector diagram describing the combined scattering from the two atoms for this 1,0,0 reflection.

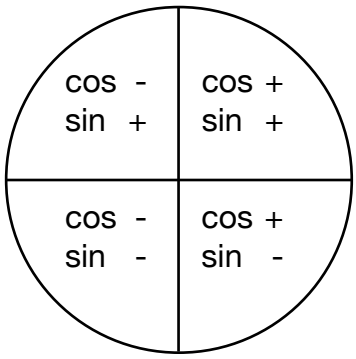
Phase Triangles

Consider a one dimensional crystal (or, if you like, call it a double projection onto one axis of a 3-D crystal).

The unit cell can be described as going from x=0 to x=1.
The diffracted spots can be indexed simply in h.
So the equation describing the diffraction from atoms is:

$$\vec{F}_h = \sum_{n=1}^{N} f_n e^{i\,2\pi(hx)} = \sum_{n=1}^{N} f_n \left[\cos(2\pi\, hx_n) + i \sin(2\pi\, hx_n) \right]$$

Possibilly useful relationships to help position the heavy atom vector on the graph paper before drawing the $|F_P|$ and $|F_{PH}|$ vectors,



		cos	sin
	0, 2π	1	0
	π/2	0	1
	π	-1	0
	3π/2	0	-1
120°	2π/3	-0.5	+0.87

The data from crystals of the Native and 2 derivatives for 3 reflections are given in the table.

Note that data was also taken from a crystal with both heavy atom derivatives. The Patterson Map from that showed the relaive positions of the heavy atoms in the unit cell.

reflection: h =	1	2	3
$\lvert F_P \rvert_{obs}$	200	200	400
$\lvert F_{PH_1} \rvert_{obs}$	150	150	450
$\lvert F_{H_1} \rvert_{calc}$	50	50	50
$\lvert F_{PH_2} \rvert_{obs}$	230	230	450
$\lvert F_{H_2} \rvert_{calc}$	50	50	50

$|F_P|$ is measured; we must find ϕ_P (for each reflection).
$|F_{PH_1}|$ & $|F_{PH_2}|$ are measured;
$|F_{H_1}|$, $|F_{H_2}|$ and heavy atom phases can be calculated for each reflection:

The Patterson map from derivative 1 had no other feature than the origin peak, nor did the map for derivative 2. The double derivative was used to get relative positions. When the position of the heavy atom of derivative 1 was taken as the origin, x=0, then the heavy atom of derivative 2 was at x=1/3.

Question: what are the phases of $\vec{F}(1)$, $\vec{F}(2)$, and $\vec{F}(3)$?

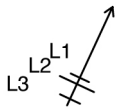
Method behind the MADness

MAD (Multiple wavelength Anomalous Dispersion) experiments often collect data at 3 different wavelengths, although collecting anomalous pairs of data at 2 wavelengths is sufficient to determine the phase.

Besides just accumulating more data (which always helps if the crystals can stand the additional radiation dose), what is the main reason for collecting MAD data at 3 different wavelengths?

(i.e., What camparisons are made among these data sets, and what particular properties are maximized by particular comparisons?)

remember:



and think about what is special at L1, L2, and L3

Reciprocal Lattice & Diffraction from a protein crystal

A number of very exciting new developments in macromolecular crystallography exploit the remarkable features of the synchrotron as an x-ray source. The high flux of radiation giving quicker data collection, allowing the use of smaller crystals, and often permitting vastly more data to be collected from one crystal is one such feature. Another is the tunable wavelength so that data can be collected that, for instance, exactly matches the anomalous scattering resonance of a particular type of atom in the molecule. A third point which we will illustrate in this question, is the combination of the high flux of photons over a broad range of wavelengths which allows a lot of different data points to be collected simultaneously very quickly. If one can trigger an enzymatic reaction in an enzyme-substrate complex, then this high speed crystallography will be able to collect data on transient intermediates.

On the next page, a crystal of orthogonal unit cell, 50x50x50 Angstroms, is at the 'X' in the center of its reciprocal space. The spots representing the ends of diffraction vectors are approximately a realistic size, accounting for effective source size, crystal size, and the mosaic character of a real crystal. A beam of x-rays is shown hitting the crystal, and a circle is drawn for a particular wavelength that shows the location of solutions to the parallelogram constructions (Ewald sphere) for this orientation of x-ray beam and crystal. The scale here is the "natural" one of a typed page, that is, 12 characters per inch across and 6 lines per inch down. Thus the n=6 spot is 1 inch from the origin and there are 6/50 = 0.12 $Å^{-1}$ per inch in this representation of reciprocal space.

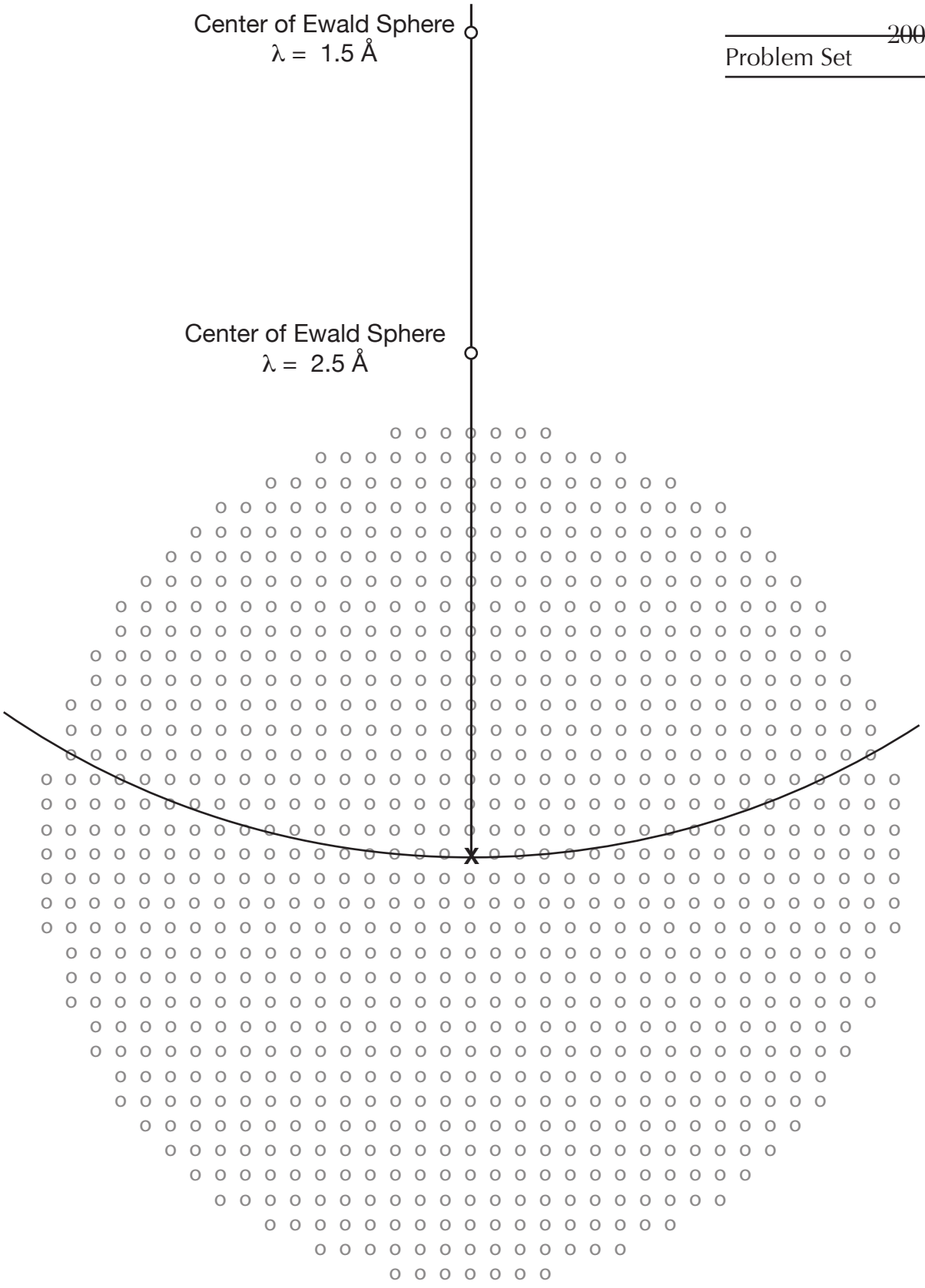
To illustrate this efficiency of data collection, consider a range of wavelengths;
to make this tractable as a pencil and paper construction, make this a rather narrow range: from 1.5 Å to 2.5 Å. The circle is drawn at one extreme for a wavelength of 1.5 Å.

(1/1.5 = 0.67 $Å^{-1}$ which at 0.12 $Å^{-1}$ per inch. makes the radius of the Ewald sphere about 5.5 inches.

The other extreme of our range is 2.5 Å (that is, 1/2.5 = 0.4 Å or 3.3 inches radius).

a) Counting partial hits as well as spots completely diffracting, how many simultaneous diffracted rays are being produced as shown in the plane of the paper? (Note that even with a continuous wavelength distribution, discrete rays are made.) Show how you counted them, presumably by coloring in all spots that are diffracting.

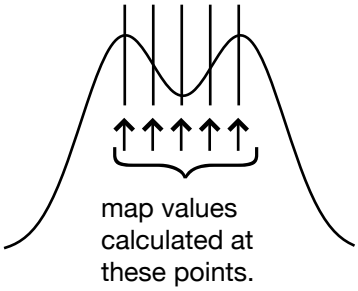
b) Show the parallelogram construction for two reciprocal lattice points, one diffracting wavelengths near 1.5 Å, the other diffracting wavelengths near 2.5 Å, and indicate on the film plane where the diffracted beams would hit. (Use compass and straight-edge and show all 4 sides of each parallelogram.)



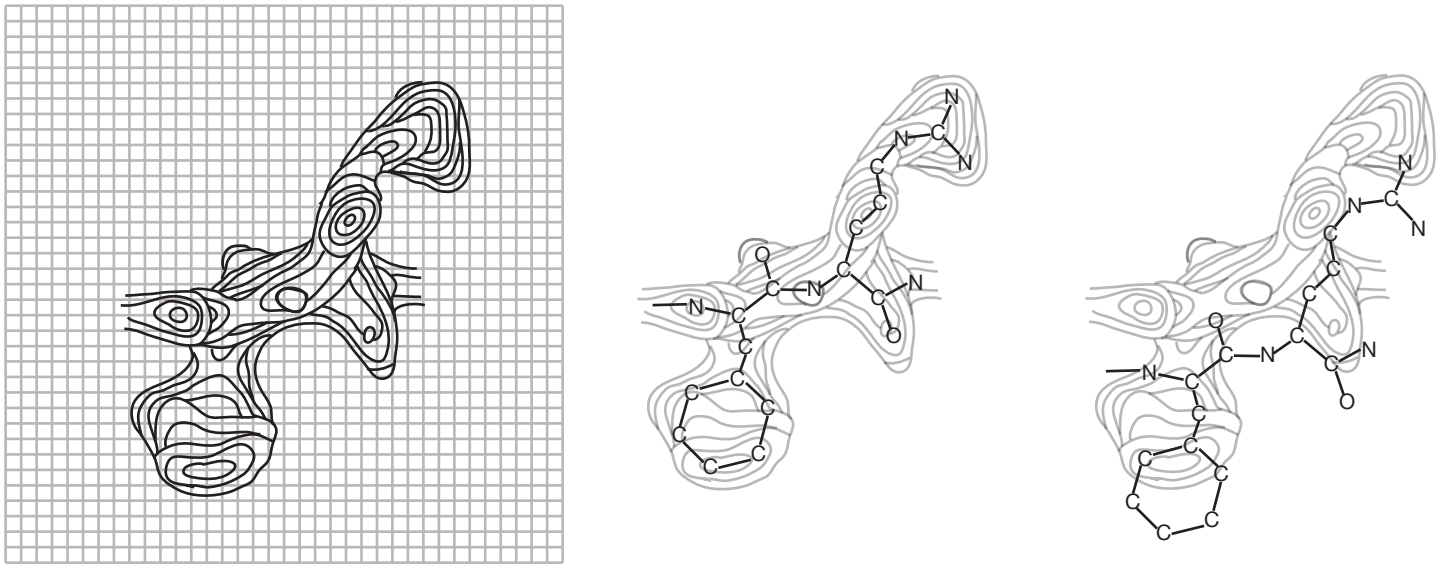
film plane

Resolution & Model-to-Map Fitting

The minimum Bragg plane spacing (that is, the furtherest out in reciprocal space data is collected) is a good measure of the effective resolution. Resolution is just the minimum distance between peaks where one can tell that there really are two peaks. One needs to calculate the map on a grid with closer intervals between points than the resolution distance in order to see a dip between two peaks separated by the resolution distance.



The contours shown in the frames below are from a 2.5 Å resolution map. The intersection points of the grid in the 1st frame show where the map values were calculated. The contours were drawn interpolating and smoothing between points. (On this three dimensional map the space inside contours blocks the view of contours behind them, but is transparent to the imbedded model.) The model is shown in a best fit position in frame 2, and displaced by 2.5 Å in frame 3.



Just how accurately do you think the actual coordinates of the relative position of each atom in the best fitted model really are - that is, are they accurate to 2.5 Å, or to 0.25 Å, or to 0.025 Å?

What other information besides the contours of electron density is being used to increase the accuracy of the relative atom positions in the model?

Exercise on crystallographic resolution, in trypsin electron density maps

Files needed for this exercise: 1HJ8_1.0A.kin, 1C9P_2.8A.kin, downloaded from the BCH 291 web page
***1HJ8 TRYPSIN AT 1.0A RESOLUTION ***

Take your browser to http://kinemage.biochem.duke.edu and click MolProbity in the navigation bar. [If you don't have Java, follow the directions to get it.]

Set the "browse" file type to "kinemage", browse to find the 1HJ8_1.0A.kin, and upload it. Continue. Set the "fetch" file type to "2Fo-Fc (EDS)" and the pdbID to 1HJ8, and fetch it (takes ~15 sec). Back on the MolProbity main page, expand the kinemage entry in the file list and ask to view your kin in KiNG.

Choose "beta" on the Views menu. You should see several vertical beta strands, with Val 52 at the center (click on an atom, to see its identity on the info line at the bottom of the graphics window). Drag right from "Structural biology" on the Tools menu, and release on "electron density map". Choose your 2Fo-Fc map, and OK its format. Be patient for it to load - it's very big. Move the contour window off to the side, and turn on both 1.2 (gray) and 3.0 sigma (purple) contour levels. You should be able to see clear density for all the non-H atoms. Note that the Val 52 sidechain is in a staggered orientation relative to its backbone.

Turn off the gray contours and move or click the slider for the purple contours up to 8.0 sigma, where the difference in x-ray scattering power between different atom types becomes evident. Click on some of the atoms with the largest peaks: what element type are they? _____ Most but not all of the C atoms have disappeared at this contour level. What element type here has intermediate size peaks (and thus intermediate scattering power)? _____

Not all atoms of a given type show up equally strongly, because they are not all equally well ordered. Click on several Calpha atoms (where the sidechain joins the backbone) that do show small purple density peaks; what are their B-factors (given on the info line, along with their identity)?

_____ , _____ , _____ .

Pick several Calphas without peaks; what are their B-factors? _____ , _____, _____ .

Choose the "helix" view. Almost no peaks are visible at 8 sigma, but if you shift to 5 sigma most backbone atoms show, and at 4 sigma most sidechain atoms. This helix is at the C-terminal chain end and has somewhat higher B's. Right-click on the double ring of the Trp sidechain to center there, and turn on the gray 1.2 sigma contour. Is there a hole thru both 5-membered and 6-membered rings? _____ Zoom out (right-drag up), center near the chain end, and notice that the end is disordered enough for even the gray contour to disappear.

Look at the views for the Ser-His-Asp catalytic triad, benzamidine inhibitor, and Arg 66, which are all extremely clear and well-ordered. In the SS 42-59 view, note that the S atoms have even bigger peaks than oxygens. Radiation damage has oxidized and opened the disulfide bond in some fraction of the molecules. Click on the S atom of the open form; what is its occupancy? _____

The 2 conformations of that S atom are quite distinct, not just a smear.

Choose the "Gln 192" view. The backbone CO has two widely separated alternate conformations, one H-bonded to an SO4 (pink). The Gln Calpha and Cbeta densities are smeared between two close alternates, and beyond that the density essentially disappears (lower contour level doesn't help very much). What is the occupancy and B-factor for one of the Cg atoms? occ:_____ , B:_____ ;

for one of the terminal N or O atoms? occ:_____ , B:_____

When an atom is not visible at all in the electron density, some crystallographers omit it from the model, some set its occupancy to zero, and some let its B refine very high. Note that the information content is much worse for this sidechain with high B (or low occupancy in each of probably many conformations) at atomic resolution than for the well-ordered parts of a much lower-resolution structure like the one below.

Choose the "neighbor e.d." view, which is part of a neighboring molecule in the crystal with no model shown in this kinemage. What is the amino-acid type of the residue at center? _____ Are the 5 atoms in its ring planar or puckered? _____

Keep in mind that initial maps seldom look this good; the phases and the density quality both improve during refinement.

Close the KiNG window.

Exercise on crystallographic resolution, in trypsin electron density maps (cont'd)

*** 1C9P TRYPSIN AT 2.8A RESOLUTION ***

Repeat the above procedure to upload the 1C9P kin file and fetch its 2Fo-Fc map from the EDS. Open the kin file from the kin list on the MolProbity main page, for viewing in KiNG, and go to the "beta" view. Click on some backbone atoms; what is the best (lowest) B-factor you find for one? _____ How does that compare with the B-values you found for similar atoms in the high-resolution structure? _____ Open the 2Fo-Fc map, and drag slowly back&forth to judge the density shape of the Val 52 sidechain in 3D. Is it concave left, concave right, or symmetrical? _____
Center on the Ile sidechain to the right of the Val, and drag left to view it from the side. Does the density show which branch is the longer one with the extra Cdelta atom? _____

Choose the "helix" view. Drag back&forth gently to see the spiral shape of the backbone density and the small bumps for the backbone O atoms. Center on the Trp sidechain. Is there a hole in either ring? _____ Would you know from the density that this was a Trp? _____

Choose the "His end" view to see the ring cross section. Could it be plausibly turned 90 degrees to sit crosswise in the density? _____ ; turned by 15 degrees? _____ Note the nearby Asp; the His orientation is fine-tuned by its hydrogen bonds, in the absence of higher resolution.

Look at the other views. Note that the disulfide S atoms are not resolved into separate peaks, altho their positions are clear. Click on one, then the other, S: what is their distance? _____A.
For Gln 192, note that it has quite reasonable density in this structure, presumably because it interacts with the BPTI inhibitor molecule.

Close the KiNG window and log out of MolProbity.

Model quality, validation exercise.

You will need a web link to MolProbity (with Java), and the file 1JIRon1S83_Arg66_supr.kin downloaded from the kinemage.biochem.duke.edu BCH291 web site.

Part 1: MolProbity

Go to the MolProbity web service (at http://kinemage.biochem.duke.edu, click MolProbity on the navigation bar) and fetch PDB file 1JIR (not case sensitive). Check that you got a trypsin at 2.0Å resolution. What is the R value? _____%; the Rfree? _____% That is very good for 2Å, presumably because of information from previous structures at higher resolution. Continue to the main page, ask to add hydrogens, and run with the default settings.
The resulting chart shows no His flips but 10 amide flips; the largest score differences are for Asn _____ and Gln _____ .

Pick "View in KiNG" for 1jirH-flipnq.kin, and animate between the two orientations for some of the views marked * for flips. Gln 30 has no clashes in the unfavored (pink) position, but in the clearly better flipped version (green) it makes _____ H-bonds.
Asn 48 makes a pseudo-turn H-bond to the backbone _____ atom of residue _____, but in the incorrect original position the NH2 has really dire clashes (not evident, of course, if the crystallographer had not added those H atoms).
Gln 64 is similar, but the clashes or H-bond are to the sidechain of _____ _____.
Close the KiNG window, and "regenerate H", accepting the flips; continue.

On the main page, chose "Analyze all-atom contacts and geometry", and run with the defaults. While waiting, you can preview the Ramachandran kin or pdf, seeing that this structure has excellent phi,psi values with no outliers. The summary statistics are also good, almost all evaluated as green; the clashscore of 7.94 is at the _____ percentile for this resolution. But good average scores do not protect against local errors. Click on "Multi-criterion chart" for per-residue scores. Click on "Rotamer" to sort by increasing rotamer quality. The worst rotamer is for Arg 66 (0% of the high-quality data is this bad, giving a a score of 0%); note that it also has a serious clash, with an overlap of _____Å. Sort on "clashes", to see that no other sidechain has both a bad rotamer and a bad clash.

Close the chart window, and view the multi-criterion kinemage in KiNG. On a backdrop of the Calpha trace and the non-water "het" groups (in pink, or gray balls for metals), this kinemage shows bad sidechain rotamers in gold and serious clashes as clumps of hotpink spikes. Find the gold sidechain for Arg 66; how many clash clumps does it have? _____ [Before flipping Gln 64, there would have been more.] Center on the Arg, zoom in, and turn on sidechains. The planar Arg guanidinium is stacked between the sidechains of residues _____ _____ and _____ _____. We will study Arg 66 further in the next part.

Close the KiNG window and continue to the main page. In the file list, click on the triangles to expand the outline, to see all the viewable or downloadable file you have accumulated. This time you will look at a further modified version of the multi-kin, so logout of MolProbity now: "logout" on left side panel, then click "Destroy all my files and log out" to clear your workspace on the server.

Model quality, validation exercise (cont'd).

Part 2: local comparison of 2Å and 1.25Å structures.

One of the few problems with the 1JIR bovine trypsin structure at 2Å resolution is Arg 66, with serious clashes and a very bad rotamer although it fits quite acceptably in the electron density. Rotamer and all-atom contact criteria were used to refit Arg 66, with the Asn 64 flip corrected and the Arg guanidinium group flipped over in its density to make two good H-bonds. To test the validity of that correction, we will coompare with a more recent porcine trypsin structure at 1.25Å resolution.

Open the 1JIRon1S83_Arg66_supr.kin kinemage in either Mage or KiNG (or upload the file in Molprobity: browse then set file type to kin and hit upload.)

Note the green Ser-His-Asp sidechains of the trypsin active site. Go to the "Arg66" view, which shows the original 2Å 1JIR model (gold) in its 2.0Å density, and its all-atom contacts, with several bad clashes (red spikes) [with the starting button selection with "*1JIRa"].

Animate to "*refit Arg 66", the model refit by adding additional steric and dihedral constraints (orange bonds). It is an excellent rotamer; are all the clashes gone? _____

Now animate to "*1S83Ha", the actual 1S83 model refined at 1.25Å (cyan). Are the 1S83 atoms cleanly centered in their atomic-resolution density peaks? _____

All 5 guanidium NH's make H-bonds, 2 to Gln 64 Oe1, one to a water, and the other two to _____.

Turn off the "Arg dots" and the "1jir map" buttons;

turn on the "*refit Arg 66" button as well as the "*1S83Ha" button. Stay on or re-choose View Arg 66.

Click on pairs of equivalent atoms in these two Arg sidechain models to find their separation (reported on the info line at the bottom of the graphics window); what is the largest difference (C,N atoms)? _____Å

Turn off the "*refit Arg 66" and turn on the original "*1JIRa". Is the original model clearly wrong? _____

What is the distance between its cd atom and the 1S83 cd? _____Å.

What is the largest distance between two equivalent (same name C,N) atoms? _____Å for the ____ atom.

Protein structures always need to use extra information in the form of bond lengths and bond angles (known from high-resolution small-molecule crystal structures and from quantum calculations); at medium to low resolution we have seen that accuracy can be improved by also adding in knowledge about dihedral-angle preferences and all-atom sterics.

Go to the "Gln 64" view with only the original 1JIR model on; is the N or the O near Arg 66? _____

For 1JIR, Gln 64 was flagged by MolProbity as needing an amide flip for steric and H-bonding reasons.

Switch to the 1S83 model; is the N or the O near Arg 66? _____.

At this resolution, does one branch have clearly higher electron density? _____; which? _____.

The 2Å model seemed well centered in the 2Å map, but that model was displaced slightly from the position of the well-fit 1.25Å model. Note that the phases for the maps come from the model and this model bias tends to make the map fit whatever is the model!

Arg 66 and Gln 64 were incorrectly fit and refined into the wrong local-minimum conformation. Look at the 3 views (Trp 141, Phe 82, Asn 34) with the 1S83 map and both 1JIR and 1S83 models on, to compare the basic accuracy of these correctly-fit sidechains at 2Å resolution. Which of the 3 sidechains matches the atomic-resolution map and model almost perfectly? _____ Which one deviates the most? _____, by what maximum atom separation? _____Å for the _____ atom.

So, at 2Å resolution, in a model that fits well into its electron density with good stereo-chemistry, would you judge that a typical atom is known to an accuracy of about 2Å, 1Å, 0.2Å, or 0.1Å? _____Å However, you have seen that a few atoms may be displaced by very large amounts: 2-3Å if a group is flipped over (and even 5-10Å occasionally, if the local density is very poor).