Protein Crystallography

Lecture 1
Introduction
Symmetry and diffraction principles
protein crystallization
Structure factor and phase problems
Experimental phasing methods

Lecture 2
Molecular replacement
Direct method
Solvent modification
Refinement and evaluation
Cross validation

Protein crystallography

Lecture 3

Applications (case study)
Use of Coot (practice in Ling lab)

Lecture 4 Short presentations by students
crystallization (seeding, new methods)
Introduction of heavy atoms
-Halides, Sel-Met, I-phe, Xe chamber etc.
MR method
MAD, SAD, MIR etc
Rfree/v alidation
Applications (structure case study)

Protein Crystallography

Crystallography

-technique reveals the three dimensional structures of molecules by X-ray diffraction.

Why X-ray

-λ = 1-2 Å, comparable to the C-C bond length of ~1.5 Å

-Interaction of X-ray with matter is through electrons which are around nulei, so electron distribution tells us where atomic nuclei (*i.e.* atoms) are.

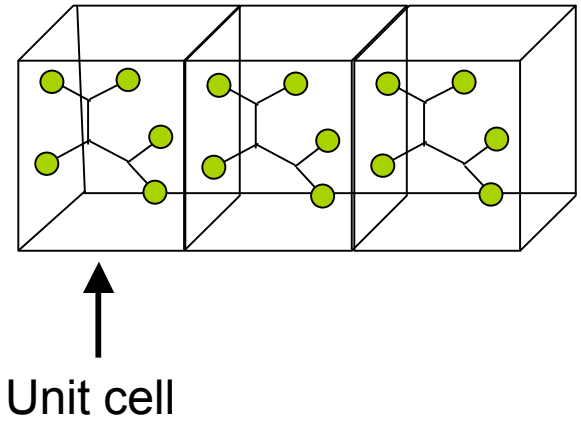
Why Crystal

- signal from one molecule is too weak
- diffraction of hundreds of thousands molecules in a crystals can be detected

Crystal and Lattice

Crystal

3-D solid composed of an arrangement of atoms, molecules and ions that are regularly repeated throughout the volume of the solid.

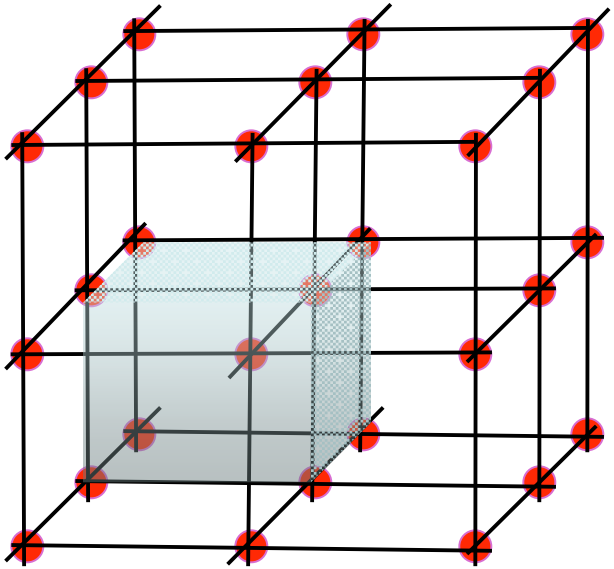


Crystal lattice: 3-D array of unit cells

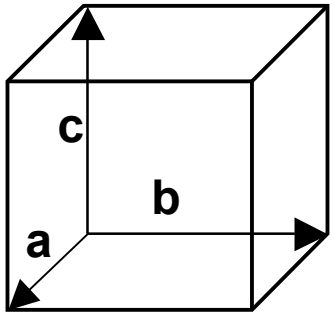
periodical translational repeats of basic pattern - "unit cell"

Crystal Lattice and Unit Cell

Three dimensional lattice

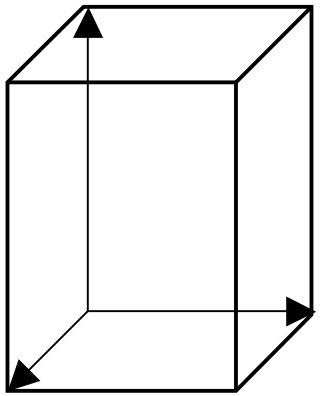


Unit cell



cubic

a=b=c

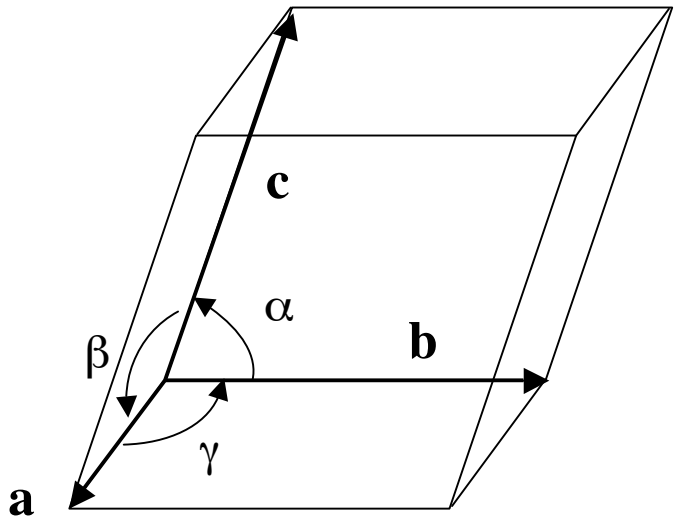


orthorhombic

a≠b≠c

Unit Cell and Symmetry

Cell defined by dimensions and angles



a, b, c, α, β, γ

7 crystal systems:

Triclinic	**$a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90^\circ$**
Monoclinic	**$a \neq b \neq c$, $\beta \neq 90^\circ$**
Orthorhombic	**$a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$**
Tetragonal	**$a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$**
Trigonal	**$a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$**
Hexagonal	**$a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$**
Cubic	**$a = b = c$, $\alpha = \beta = \gamma = 90^\circ$**

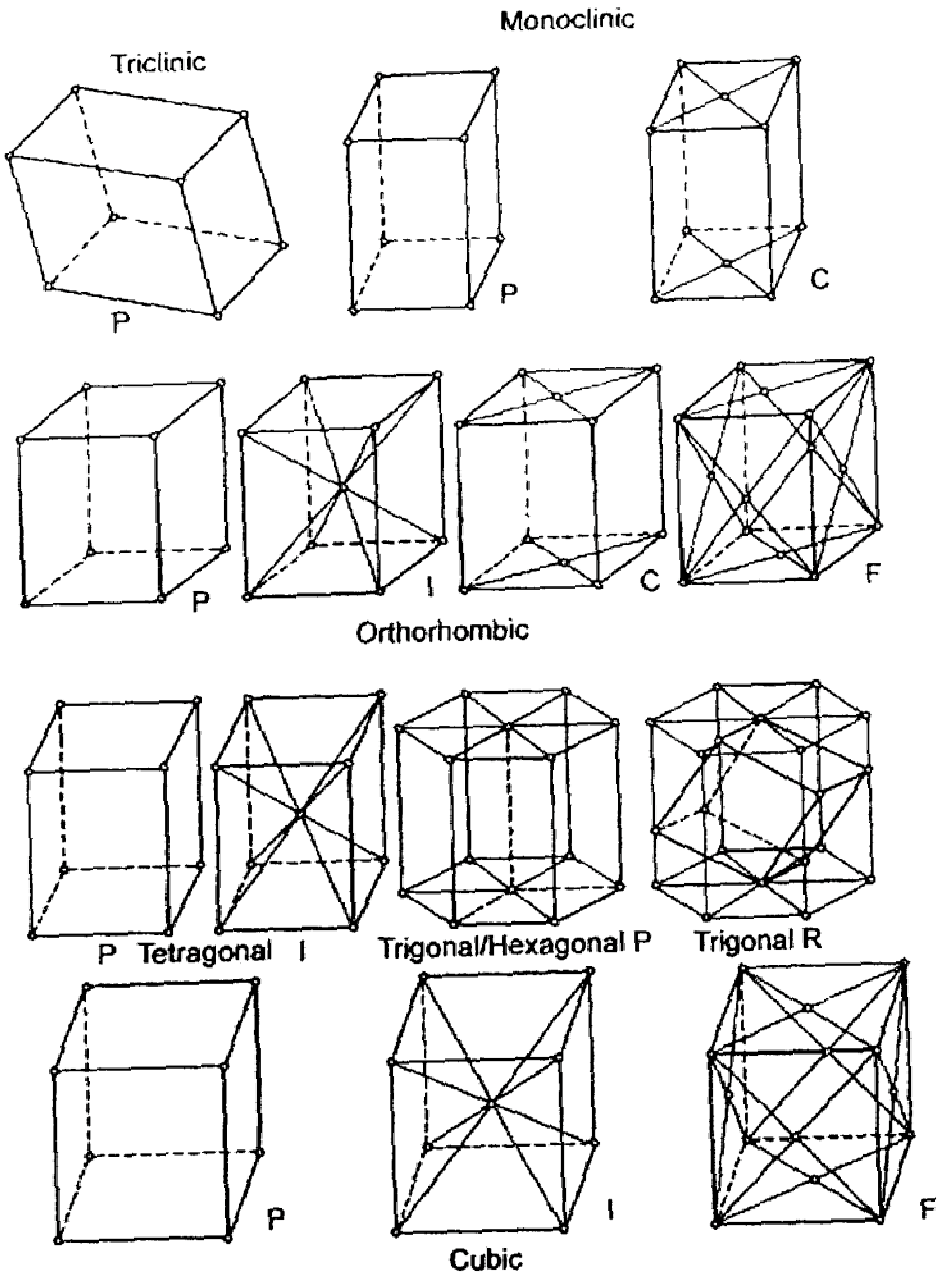
Crystallography 101
http://ruppweb.dyndns.org/Xray/101index.html

14 Bravis Lattices

4 lattice types
P, I, C, F

7 crystal systems

P- Primitive
I- body centered
C- C centered
F- Face centered



Symmetry and Unit Cell

- **Unit cell: A minimum repeating unit (parallelepiped) in a crystal lattice for translational repeats in 3-D space.**
  - **a translational symmetry**
  - **cell parameters: a, b, c, α, β, γ, in right-handed axis systems**

- **Symmetry and Space group**
  - **relationships of identical objects in 3-D space (operations, elements)**

  e.g. rotations: 2, 3, 4, 6 fold axes

  mirror: m

  screw axis: 2_1, 3_1, 3_2, 4_1, 4_2, 4_3, 6_1, 6_2, *etc*
  - **a set of symmetry operations existing in a certain crystal**

    ~230 space groups in total

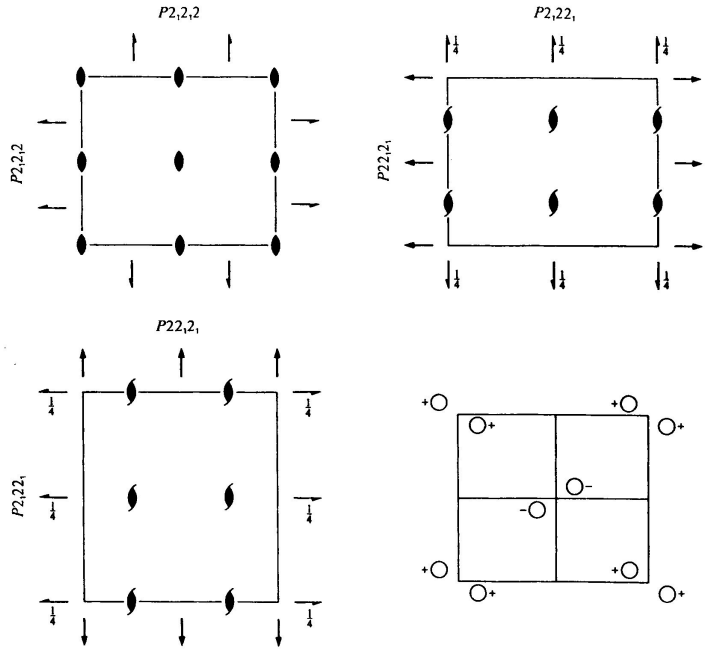
    - 65 space groups for protein crystal

Space group and International table

- **Symmetry contents**
  -International tables for crystallography (Vol. I)
- **Asymmetric unit (AU)**
  -multiplicity (Z)
  e.g. P21212, four AUs,
  so $V_{au} = 1/Z \times V_{unit\ cell}$

Space group P21212

$P2_12_12$	D_2^3	222	Orthorhombic
No. 18	$P2_12_12$		Patterson symmetry $Pmmm$



Origin at intersection of 2 with perpendicular plane containing 2_1 axes

Asymmetric unit $0 \le x \le \frac{1}{2}$; $0 \le y \le \frac{1}{2}$; $0 \le z \le 1$

Symmetry operations

(1) 1 (2) 2 $0,0,z$ (3) $2(0,\frac{1}{2},0)$ $\frac{1}{4},y,0$ (4) $2(\frac{1}{2},0,0)$ $x,\frac{1}{4},0$

Space group P21212

CONTINUED No. 18 $P2_12_12$

Generators selected (1); $t(1,0,0)$; $t(0,1,0)$; $t(0,0,1)$; (2); (3)

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates				Reflection conditions
							General:
4	c	1	(1) x,y,z	(2) $\bar{x},\bar{y},z$	(3) $\bar{x}+\frac{1}{2},y+\frac{1}{2},\bar{z}$	(4) $x+\frac{1}{2},\bar{y}+\frac{1}{2},\bar{z}$	$h00$: $h=2n$ $0k0$: $k=2n$
							Special: as above, plus
2	b	. . 2	$0,\frac{1}{2},z$	$\frac{1}{2},0,\bar{z}$			$hk0$: $h+k=2n$
2	a	. . 2	$0,0,z$	$\frac{1}{2},\frac{1}{2},\bar{z}$			$hk0$: $h+k=2n$

Symmetry of special projections

Along [001] $p2gg$
$\mathbf{a}'=\mathbf{a}$ $\mathbf{b}'=\mathbf{b}$
Origin at $0,0,z$

Along [100] $p2mg$
$\mathbf{a}'=\mathbf{b}$ $\mathbf{b}'=\mathbf{c}$
Origin at $x,\frac{1}{4},0$

Along [010] $p2gm$
$\mathbf{a}'=\mathbf{c}$ $\mathbf{b}'=\mathbf{a}$
Origin at $\frac{1}{4},y,0$

Maximal non-isomorphic subgroups

I [2] $P12_11$ ($P2_1$, 4) 1; 3
[2] $P2_111$ ($P2_1$, 4) 1; 4
[2] $P112$ ($P2$, 3) 1; 2

IIa none

IIb [2] $P2_12_12_1$ ($\mathbf{c}'=2\mathbf{c}$) (19)

Maximal isomorphic subgroups of lowest index

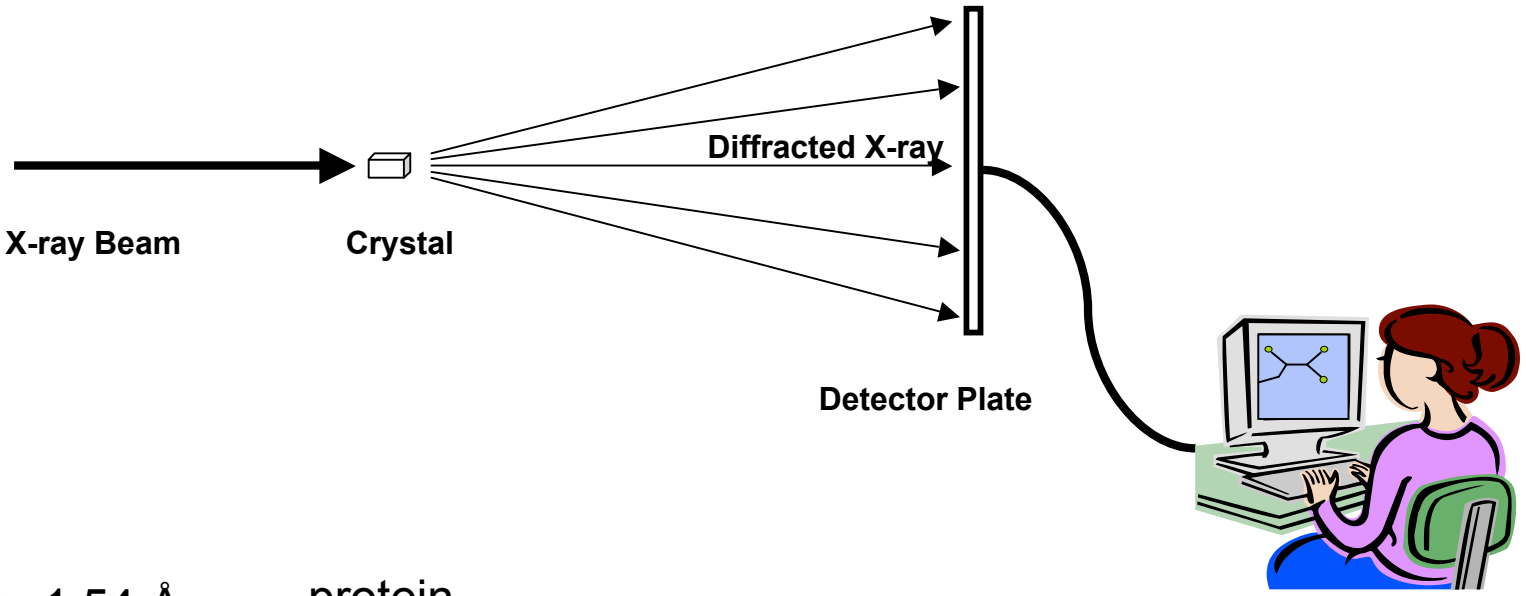
IIc [2] $P2_12_12$ ($\mathbf{c}'=2\mathbf{c}$) (18); [3] $P2_12_12$ ($\mathbf{a}'=3\mathbf{a}$ or $\mathbf{b}'=3\mathbf{b}$) (18)

Minimal non-isomorphic supergroups

I [2] $Pbam$ (55); [2] $Pccn$ (56); [2] $Pbcm$ (57); [2] $Pnnm$ (58); [2] $Pmmn$ (59); [2] $Pbcn$ (60); [2] $P42_12$ (90); [2] $P4_22_12$ (94); [2] $P\bar{4}2_1m$ (113); [2] $P\bar{4}2_1c$ (114)

II [2] $A2_122$ ($C222_1$, 20); [2] $B22_12$ ($C222_1$, 20); [2] $C222$ (21); [2] $I222$ (23); [2] $P22_12$ ($\mathbf{a}'=\frac{1}{2}\mathbf{a}$) ($P222_1$, 17); [2] $P2_122$ ($\mathbf{b}'=\frac{1}{2}\mathbf{b}$) ($P222_1$, 17)

X-ray crystallography

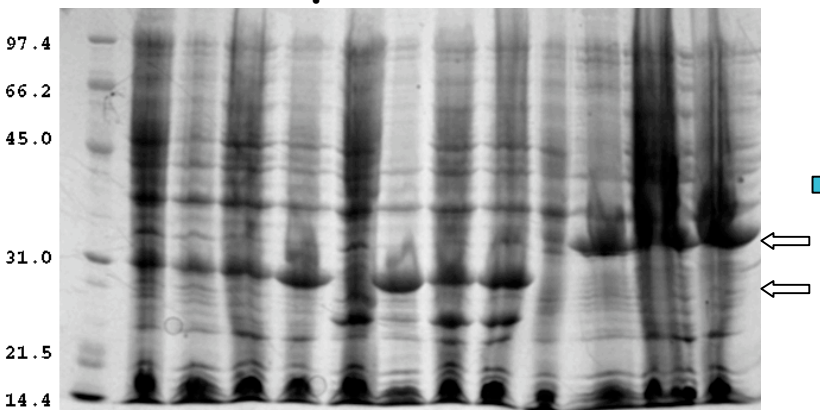


λ=1.54 Å protein

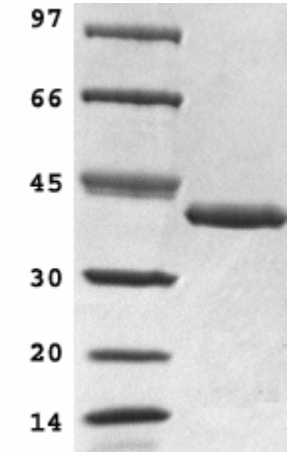
Image of molecules in crystal

General Steps

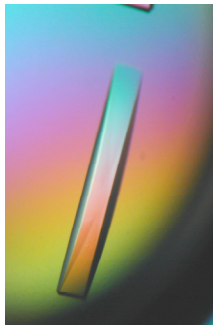
Expression

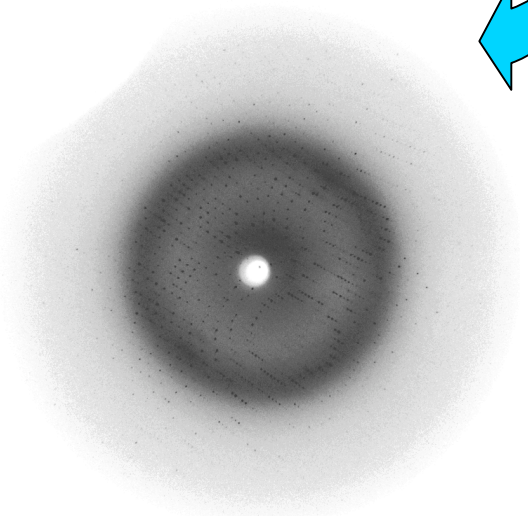


Purification



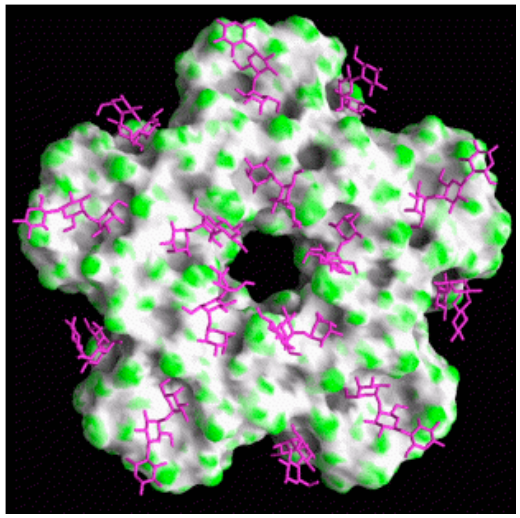
Crystallization





X-ray diffraction

Structural analysis

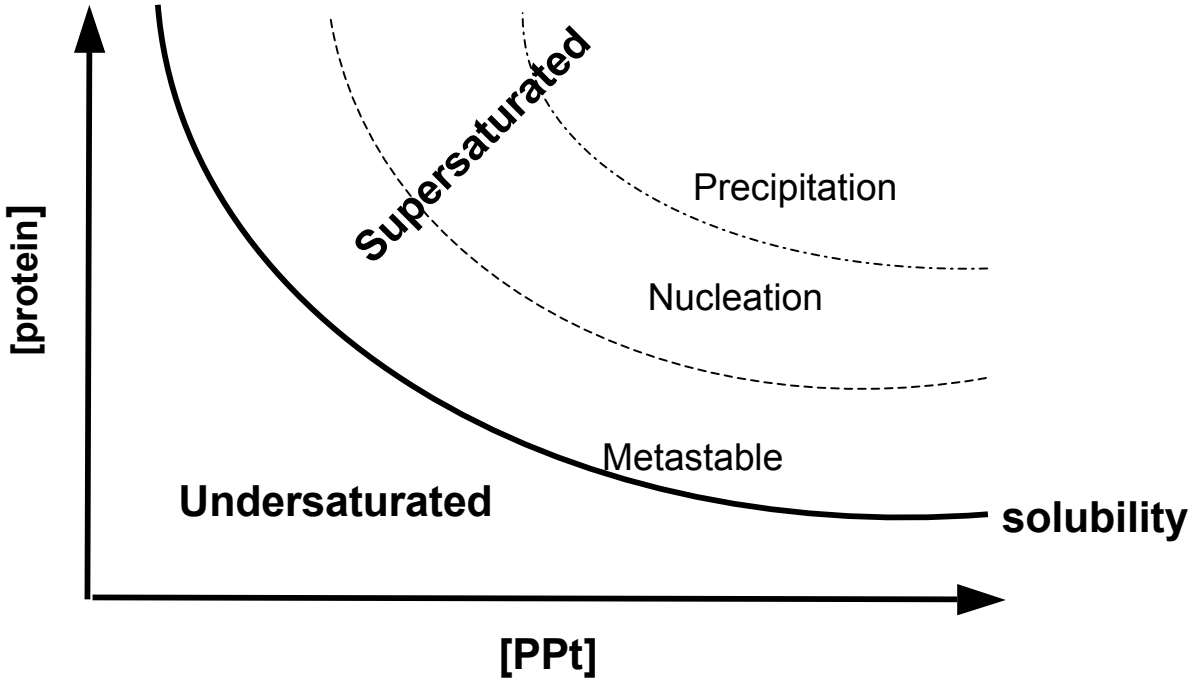


Major Steps in Crystallography

- **Grow crystals**
- **Measure the intensities of the diffraction pattern**
- **Compute an electron density map**
- **Interpret the electron density map and construct model**
- **Refine molecular model to fit experimental observations**

Protein Crystallization

- **Principles of protein solubility**



Crystallization:

-To create supersaturation in protein solution by Increasing protein concentration beyond their saturation points gently

Crystallization

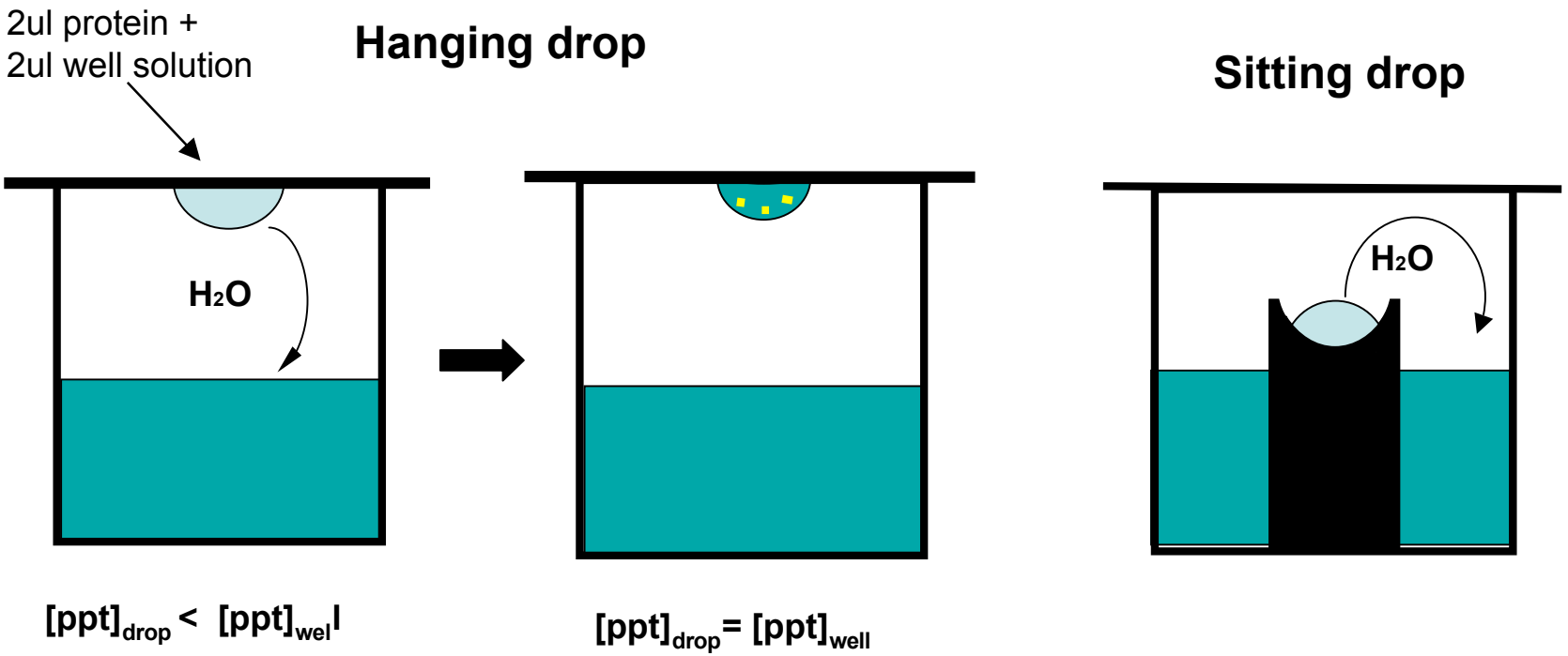
Crystallization

-precipitate protein from solution in organized fashion.

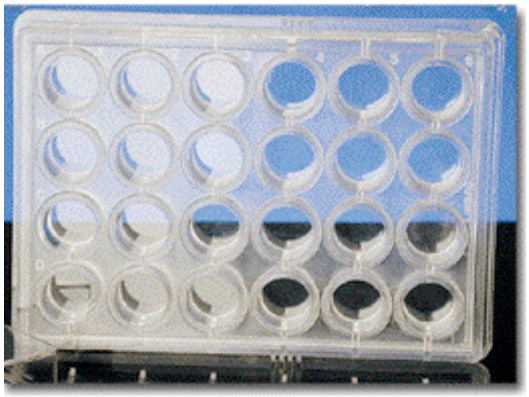
-by increasing protein concentration > S_o (solubility)

Methods

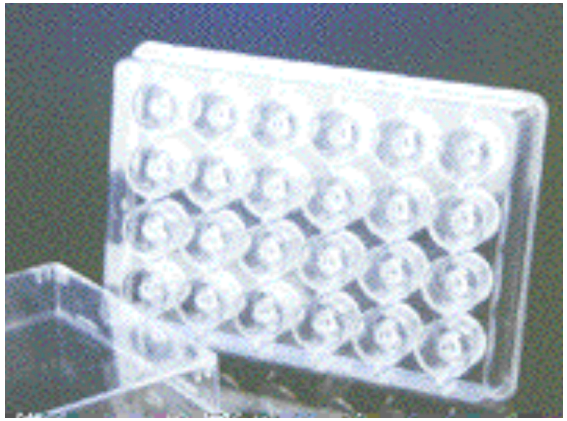
-vapor diffusion



Crystallization Plates



A Linbo plate for hanging drop



24-well plate for sitting drop

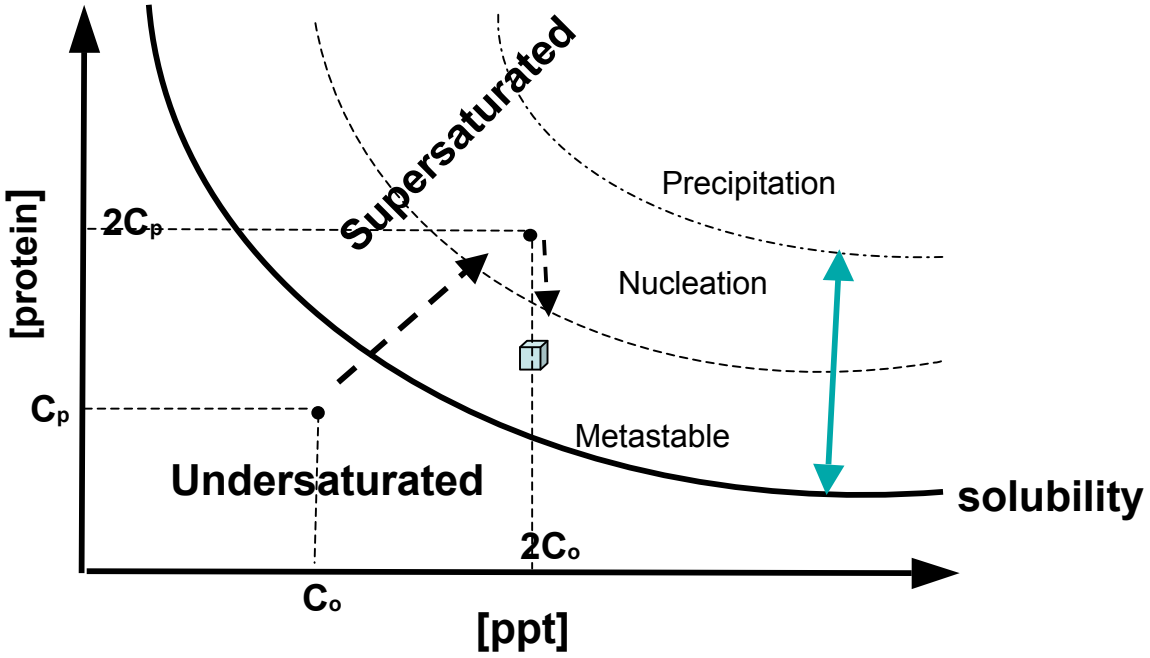
Hampton site: http://hamptonresearch.com/support/

Protein Crystallization

- **Screening/optimization**

  crystal growth depends on:

  type of precipitants, [ppt], pH, T, salts, organic solvents, etc.



Purpose: to find the conditions that nucleation and mtastable regions exist in the phase diagram and hit the right concentrations.

Seeding

Problems:

-Nucleation hard to occur or to be controlled
-Supersaturation for nucleation is too high to slow crystal growth that leads to well-ordered , large-sized crystals (small crystal clusters)
-lack of reproducibility

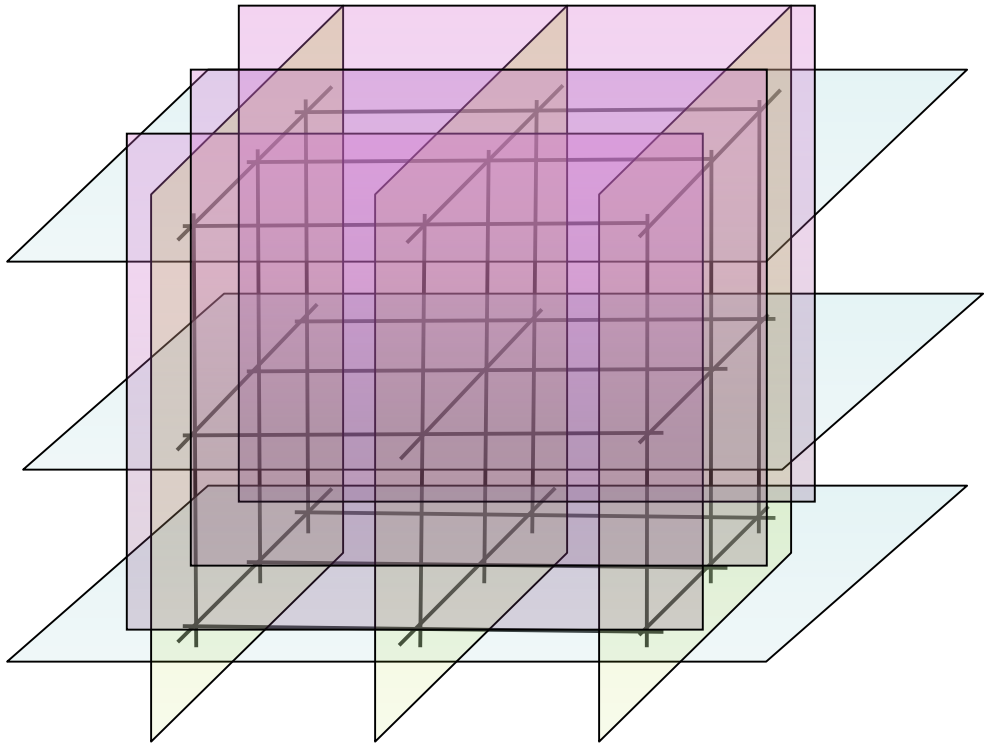
Solutions:

-Skip the nucleation stage
-Separate the nucleation and crystal growth stages
-increase size of crystals
-reduce showers of crystals
-improve crystal quality

Major Steps in Crystallography

- **Grow crystals**
- **Measure the intensities of the diffraction pattern**
- **Compute an electron density map**
- **Interpret the electron density map and construct model**
- **Refine molecular model to fit experimental observations**

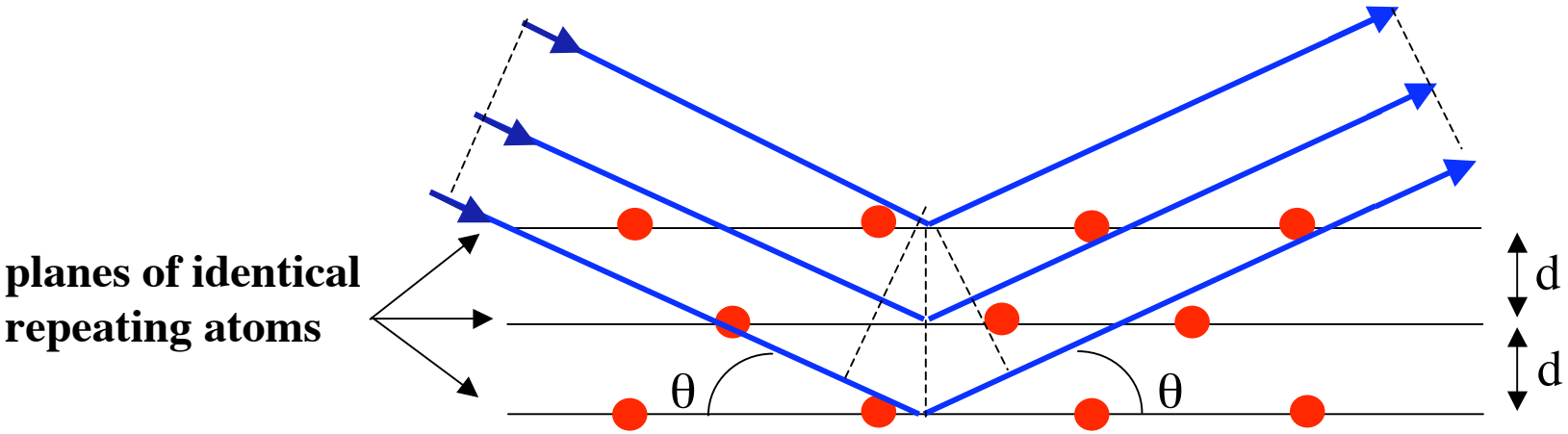
Lattice and Planes



Assembly of a set of planes

Bragg's Law

X-rays are diffracted by lattice planes of atoms when the path difference equal to integral (n) folds of the wavelength.



θ = incident light ∠ = diffracted light ∠

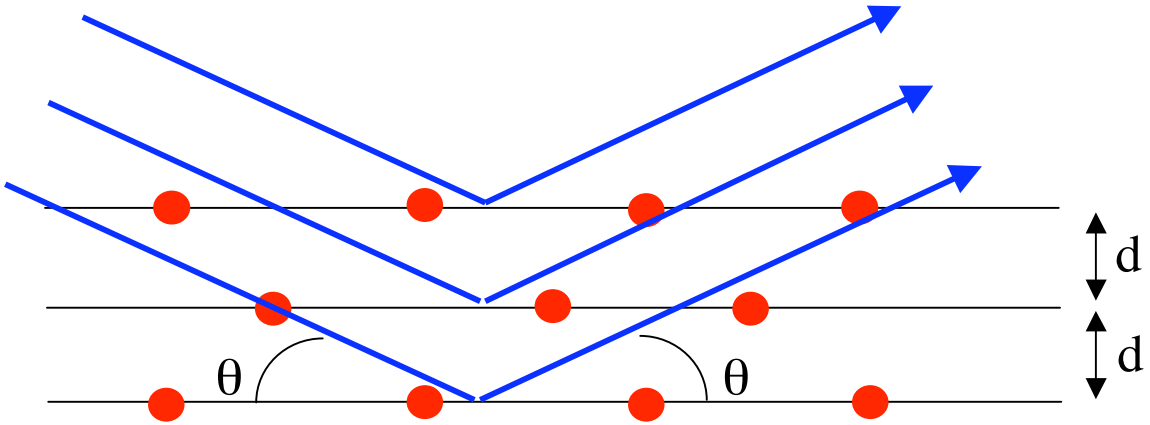
The diffraction is analog to reflection, so it also called X-ray reflection.

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

Optical path difference

Wavelength of X-ray

Bragg's Law and Data collection

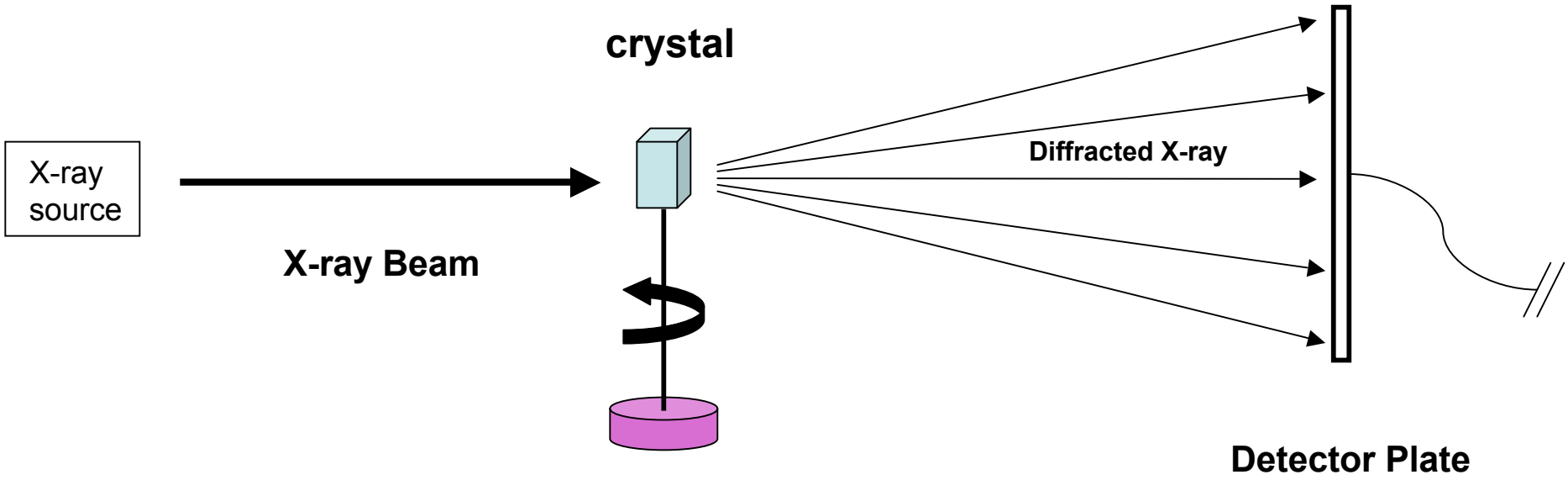


$2d\sin\theta = n\lambda$

- **n** = integer
- **λ** = constant
- **d** = distance between planes or resolution
- **θ** = incident/diffraction angle

- Since **n** is integral, the diffraction is discrete
- Diffraction occurs only when X-ray with certain incident angle **θ** that meets the Bragg's Law.

Diffraction Data Collection

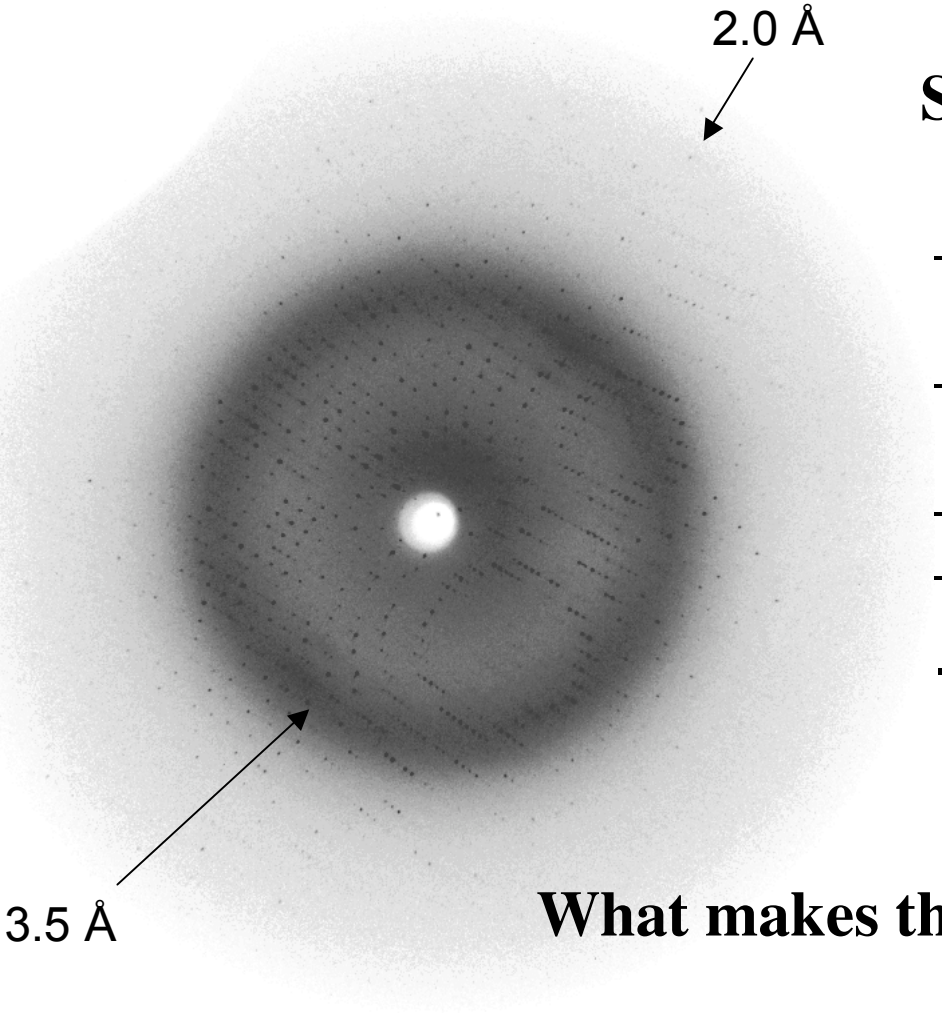


As the crystal rotates along the axis perpendicular to X-ray beam, lattice planes in different orientations will be in diffraction angles (θ) to X-ray; *i.e.*, when sets of planes meet the condition of Bragg's Law, diffraction occurs.

During the rotation of a crystal, many sets of planes will be in diffraction angles, so we collect lots reflections

X-ray diffraction

Diffraction Pattern



Spots are reflections:

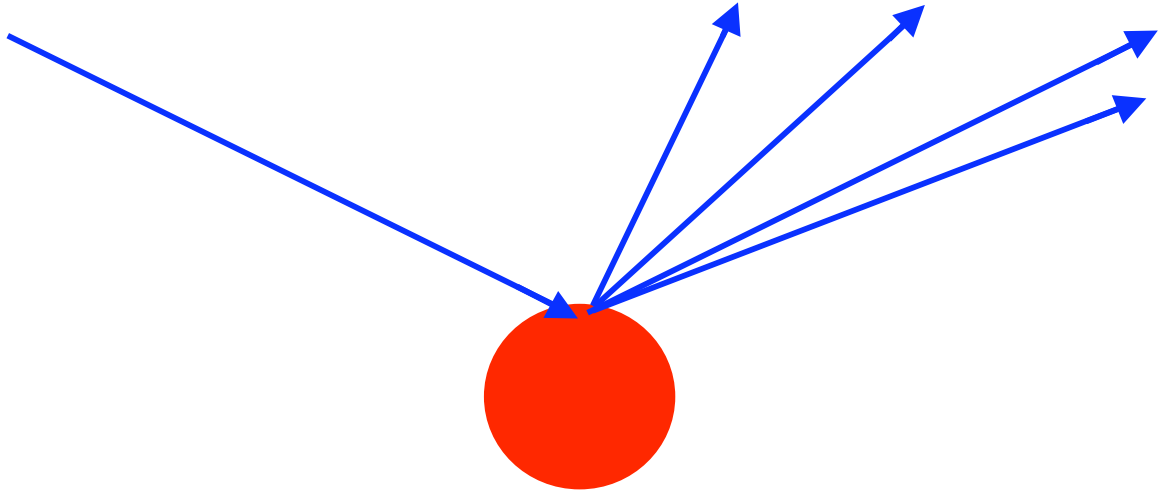
-the positions determined by the cell parameters
- the intensities (I) determined by the protein structure
- Intensity with hkl index (I_{hkl})
- data is a set of I_{hkl} (s)

- resolution at the edge, θ_{max}

What makes the intensity different in spots?

X-ray scattered by atom

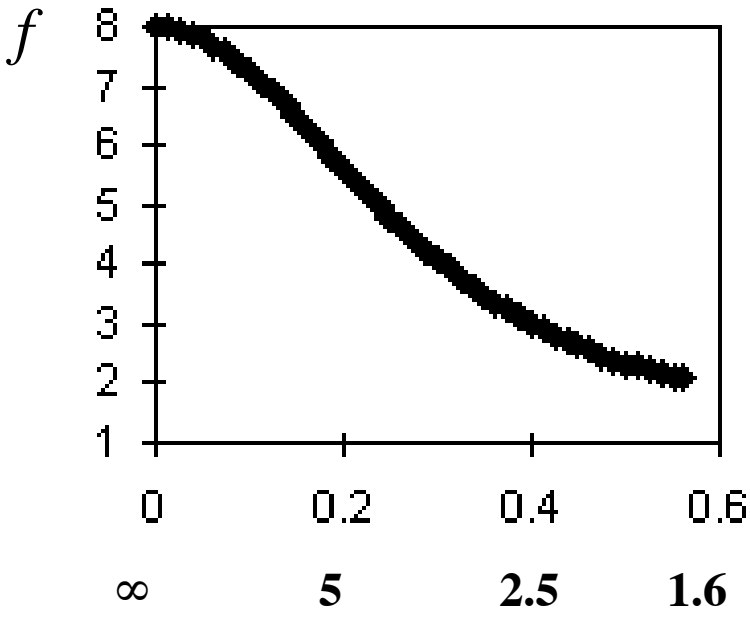
Scattering is in 3D



Diffraction is actually from atoms in planes

Intensities have 3D information

Atomic Scattering Factor



$$f^{o}(\sin\theta/\lambda) = \sum_{i=1}^{4} a_i \cdot e^{-b_i(\sin\theta/\lambda)^2} + c$$

Oxygen atomic scattering factor
Z = 8

Diffraction from different atoms - scattering factor f_j

based on atom, numbers of electrons ⟶ At zero degree equal to Z

Larger atoms - more electrons diffract strongly

Small atoms - fewer electrons diffract weakly

Structure Factor

Atomic structure factor

$$f_{(hkl)} = f_{(j)}\, exp\, [2\pi * i(hx_{(j)} + ky_{(j)} + lz_{(j)})]$$

hkl = index from diffraction pattern
$f_{(j)}$ = atomic scattering factor for atom j
x, y, z - coordinates of an atom

For a group of atoms (i.e. molecules in unit cells)

$$F_{(hkl)} = \sum_{j=1}^{atoms} f_{(j)}\, exp\, [i2\pi(hx_{(j)} + ky_{(j)} + lz_{(j)})]$$

$F_{(hkl)}$ = structure factor

Structure factor represents x-ray diffraction of molecules in a unit cell and is determined by all the atoms in the unit cell and their positions

Structure Factor and Intensity

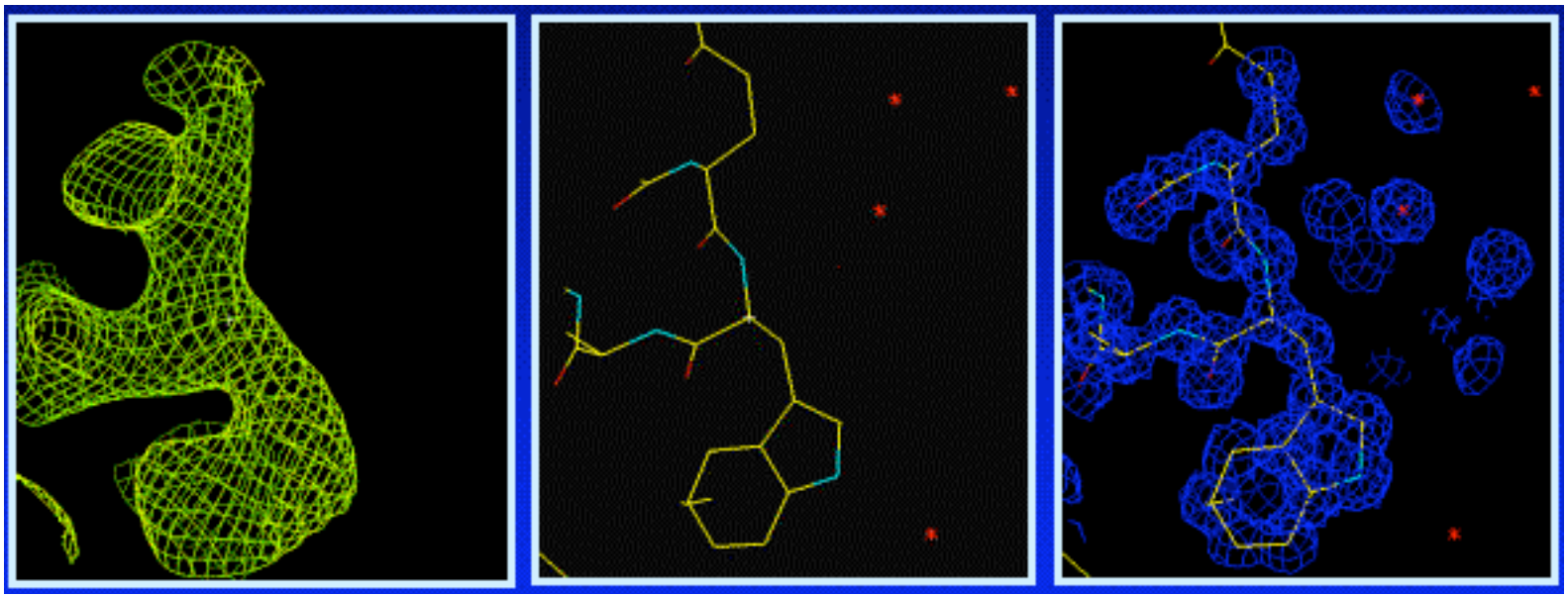
$$I_{hkl} = k\,|F_{hkl}|^2$$

$$|F| = k\,I^{1/2}$$

The amplitude of F_{hkl} is proportional to the square root of the intensity, so I_{hkl} determined by molecular structure in the crystal.

The amplitudes ($|F_{hkl}|$) can be obtained from the measured intensities (I_{hkl}) - X-ray diffraction data.

Electron Density Map



Initial map model refined model and map

Crystallography 101
http://ruppweb.dyndns.org/Xray/101indexhtml

Electron Density Calculation and Phase Problem

$$\rho(x,y,z) = 1/V \sum_{h}\sum_{k}\sum_{l} F_{hkl}\, exp\,(-2\pi i(hX+kY+lZ))$$

X, Y, Z: any point in the unit cell

$$F_{hkl} = |F_{hkl}|\, exp(i\,\alpha_{hkl})$$

From diffraction

phase of F_{hkl} : $\alpha_{hkl} = 2\pi(hx_j+ky_j+lz_j)$

x_j, y_j, z_j: coordinates for atom j, unknown

Ways of obtaining phases (solving structure)

Guess:

- **Molecular Replacement (homology> 30%)**

Heavy atom methods (experimental phasing)

- **MIR: multiple isomorphous replacement**
- **MAD: multiple wavelength anomalous diffraction**
- **SIRAS: single isomorphous replacement with anomalous scattering**
- **SAD: single wavelength anomalous diffraction**

Direct method

- **mathematical/statistical method**

Introduction of heavy atoms into protein crystals

-Heavy metal soaking (Hg, Au, Pt, Cs… etc)
-traditional methods
-Se-Met incorporation
-use autotrophic strains or deplete Met in media
-Xe into crystals by Xe gas chamber
-Halide ions soaking
-I-Phe incorporation (modified genetic code)

Multiple data sets for experimental phasing

Native protein crystal: $|F_P|$

Plus:

MIR: heavy atom derivative crystal 1: $|F_{PH1}|$

" 2: $|F_{PH2}|$

....

" n: $|F_{PHn}|$

MAD: heavy atom derivative crystal at λ_1: $|F_{PH1}|$

" λ_2: $|F_{PH2}|$

....

" λ_4: $|F_{PH4}|$

MAD data sets collected at synchrotron centers

Heavy atoms positions determined by different Patternson methods

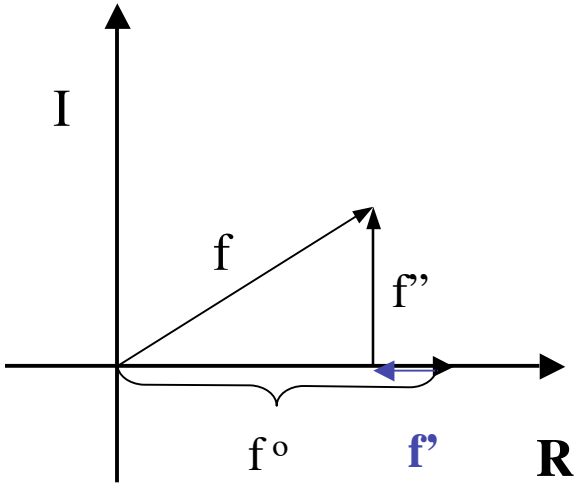
$$|F_H| = |F_{PH}| - |F_P|$$

Different Patterson function:

$$P(u,v,w) = 1/v \sum_{h} \sum_{k} \sum_{l} |F_H|^2 \, exp\,[-2\pi i(hu + kv + lw)]$$

The locations of the Patterson peaks represent vectors between heavy atoms, thus, the position of the heavy atoms can be deduced from the Patterson peaks. F_H derived from heavy atom positions.

Anomalous scattering

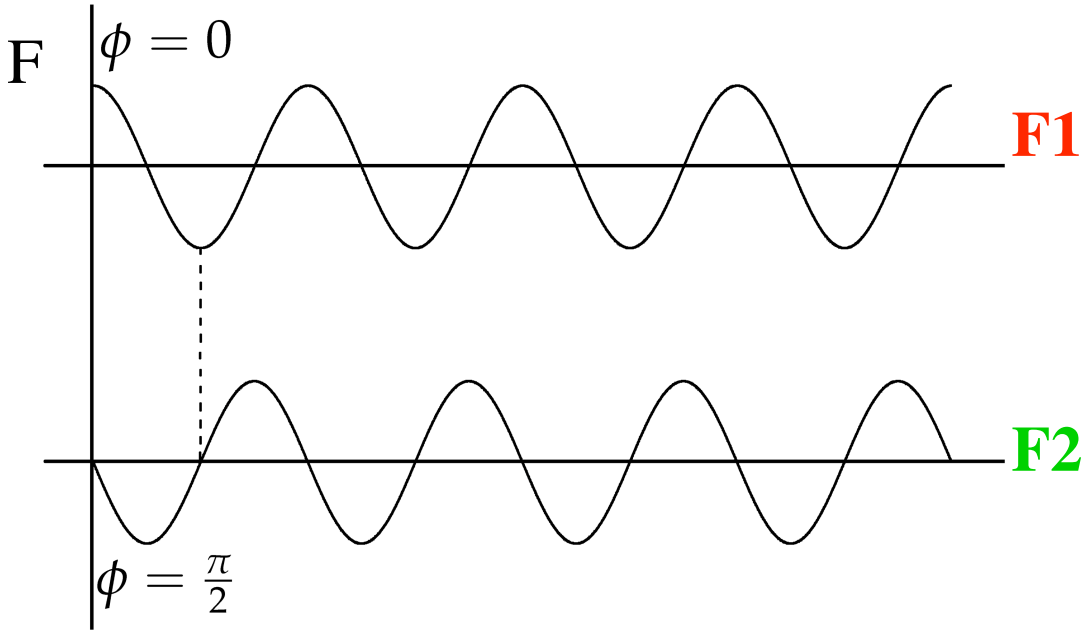


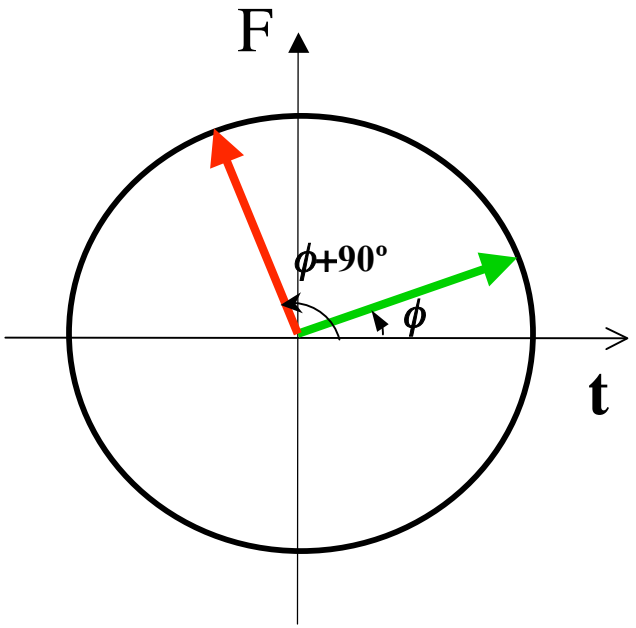
$$f = f^o + f' + if''$$

f’ and f” are wave length dependent and detectable only near absorption edges

$f = f^o$ when there is no anomalous scattering

Periodic Function as a Wave

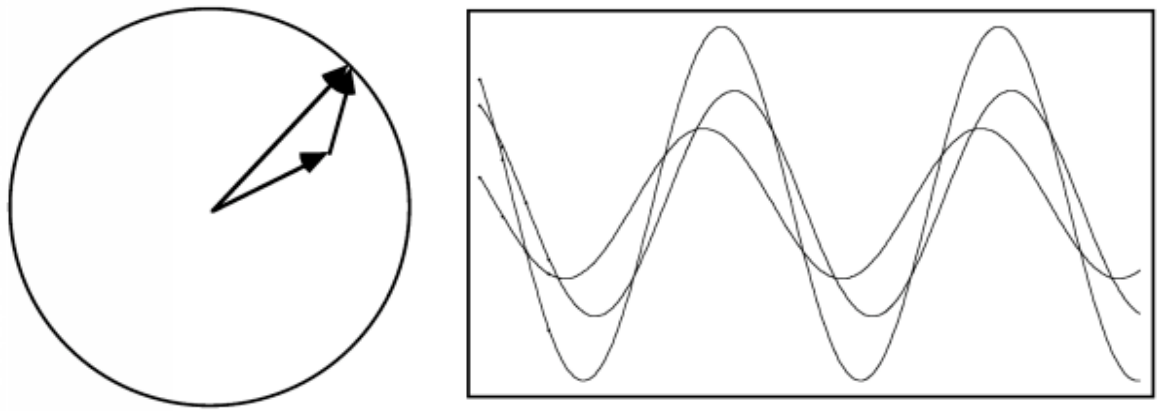




$F = A \cos(\omega t - \phi)$

$\omega t = 2\pi \upsilon t$

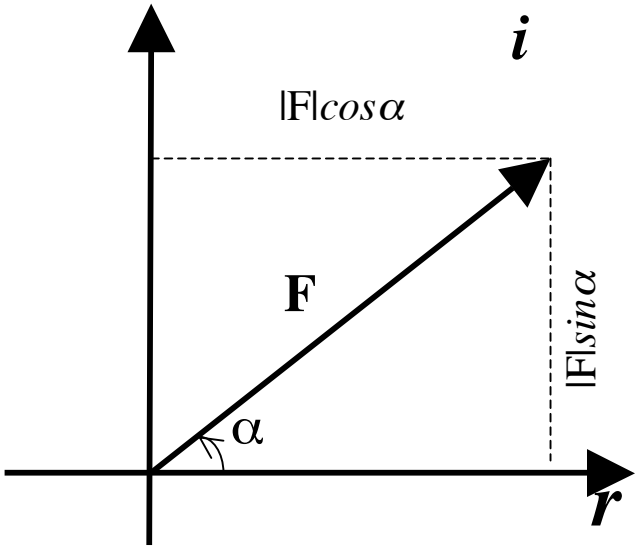
Adding waves as vectors



Cambridge X-ray course

http://www-structmed.cimr.cam.ac.uk/course.html

Argand diagram of the vector of a wave

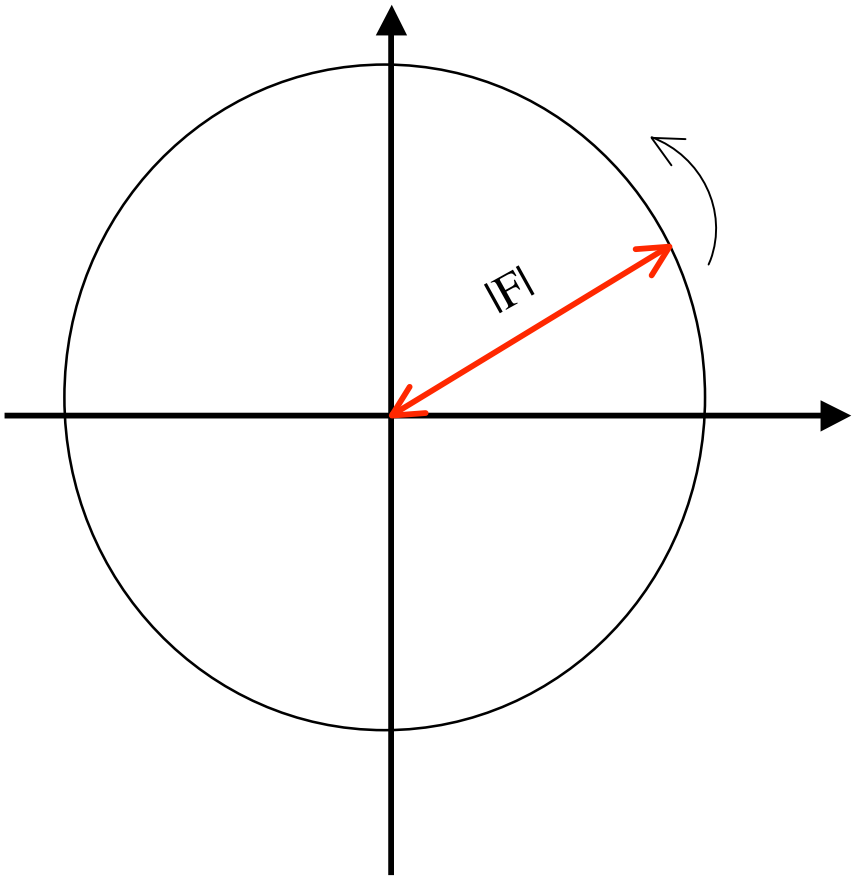


$$\mathbf{F} = |F|\cos\alpha + \boldsymbol{i}|F|\sin\alpha$$

Euler Formula: $\boldsymbol{e}^{\pm i\alpha} = \cos\alpha \pm \boldsymbol{i}\sin\alpha$

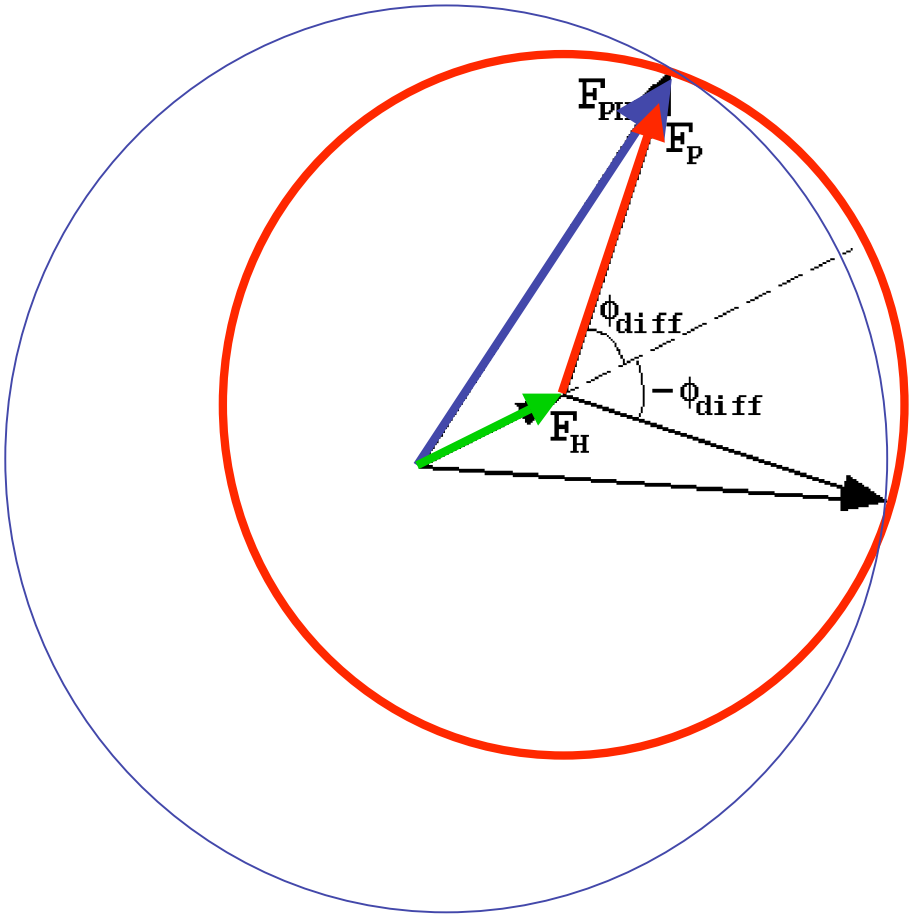
$$\mathbf{F} = |F|\boldsymbol{e}^{i\alpha}$$

Probability of phases of unknown structure



$$|F| = k\, I^{1/2}$$

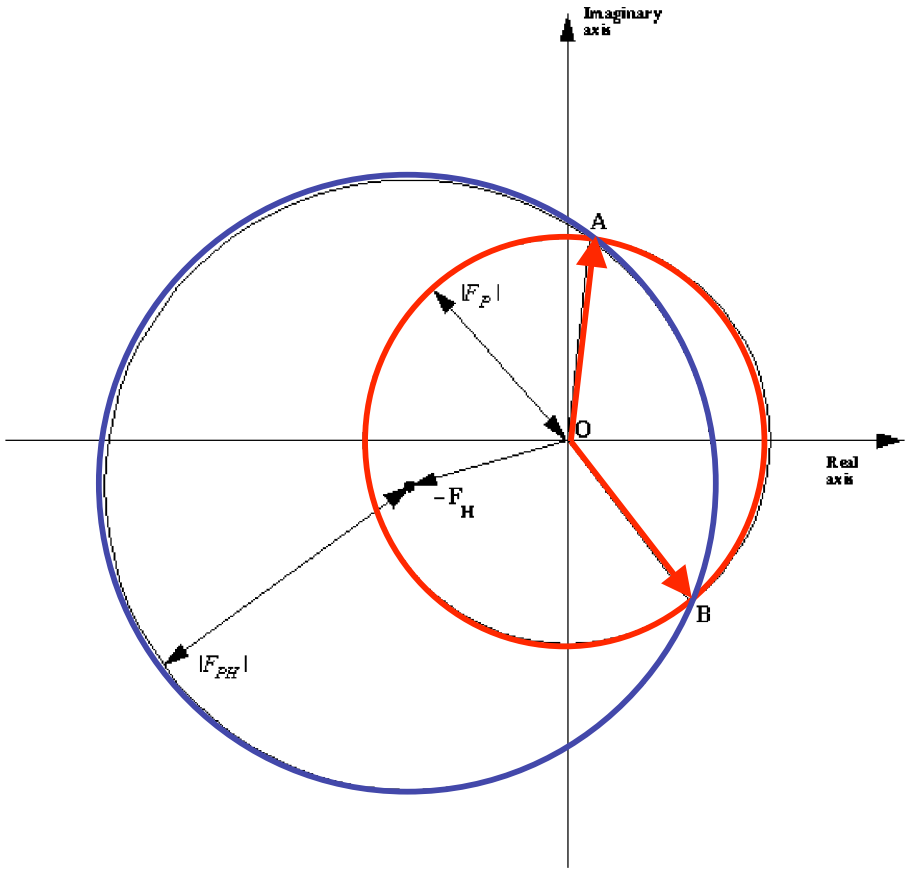
Phases of a heavy atom derivative



F_{PH}: heavy atom derivative
F_P: protein
F_H: heavy atoms

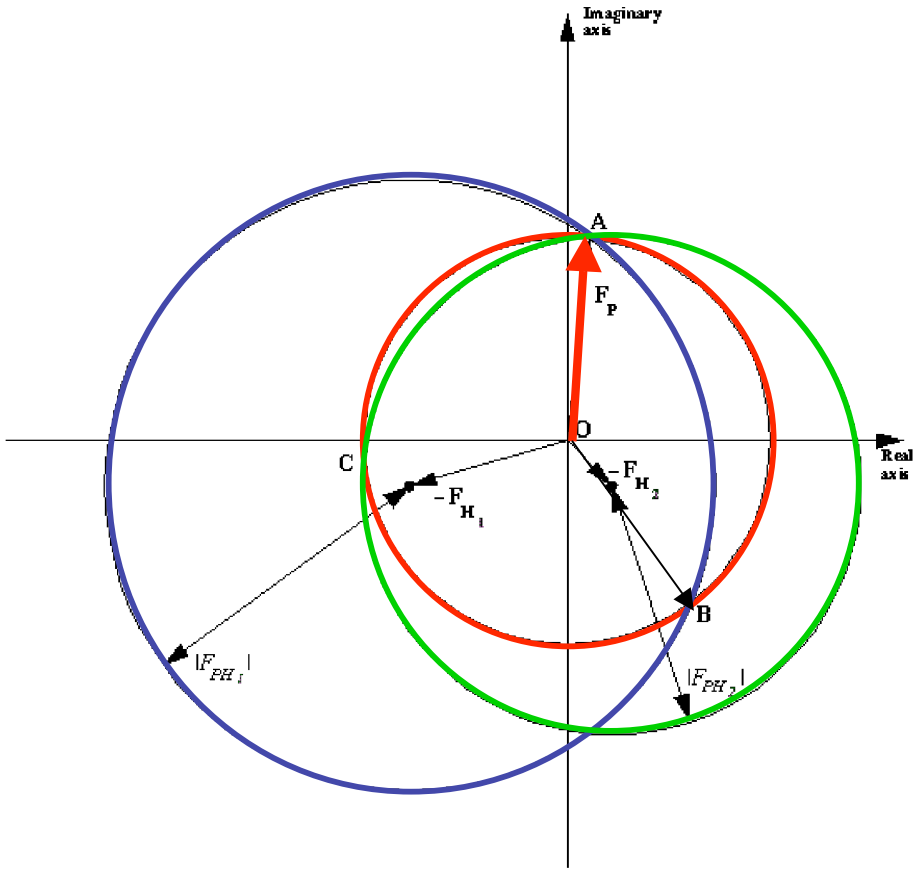
$$F_{PH} = F_P + F_H$$

Phasing diagram for SIR



CCP4: (basic math for crystallographers)
http://www.ccp4.ac.uk/dist/html/pxmaths/index.html

Phasing diagram for MIR



CCP4: (basic math for crystallographers)
http://www.ccp4.ac.uk/dist/html/pxmaths/index.html

Phasing diagram for SIRAS