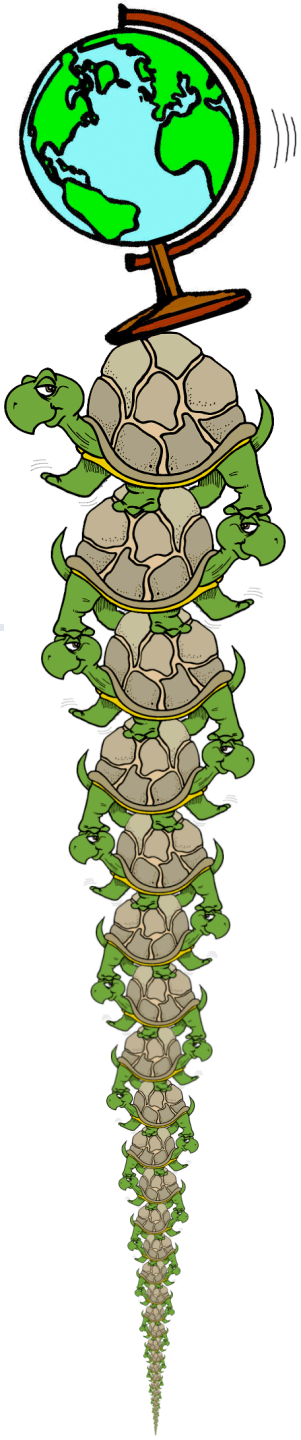
Experimental Phasing

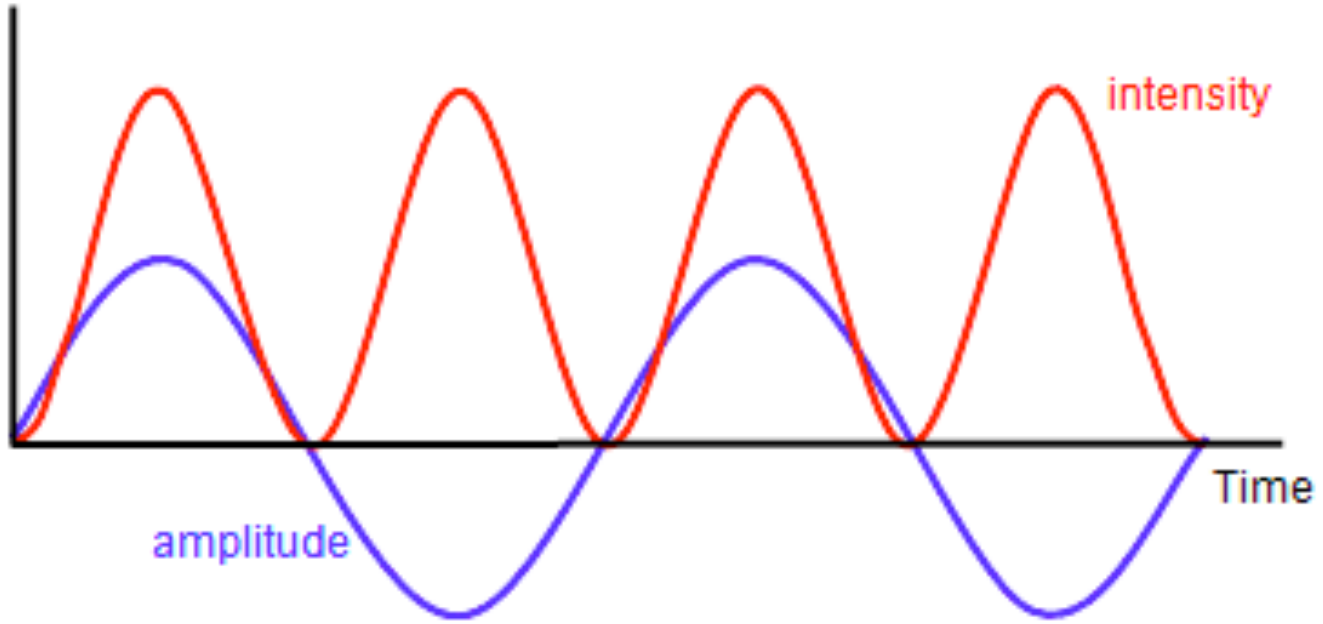
Airlie McCoy

UNIVERSITY OF CAMBRIDGE



Interference

- X-ray detectors only detect X-ray intensities

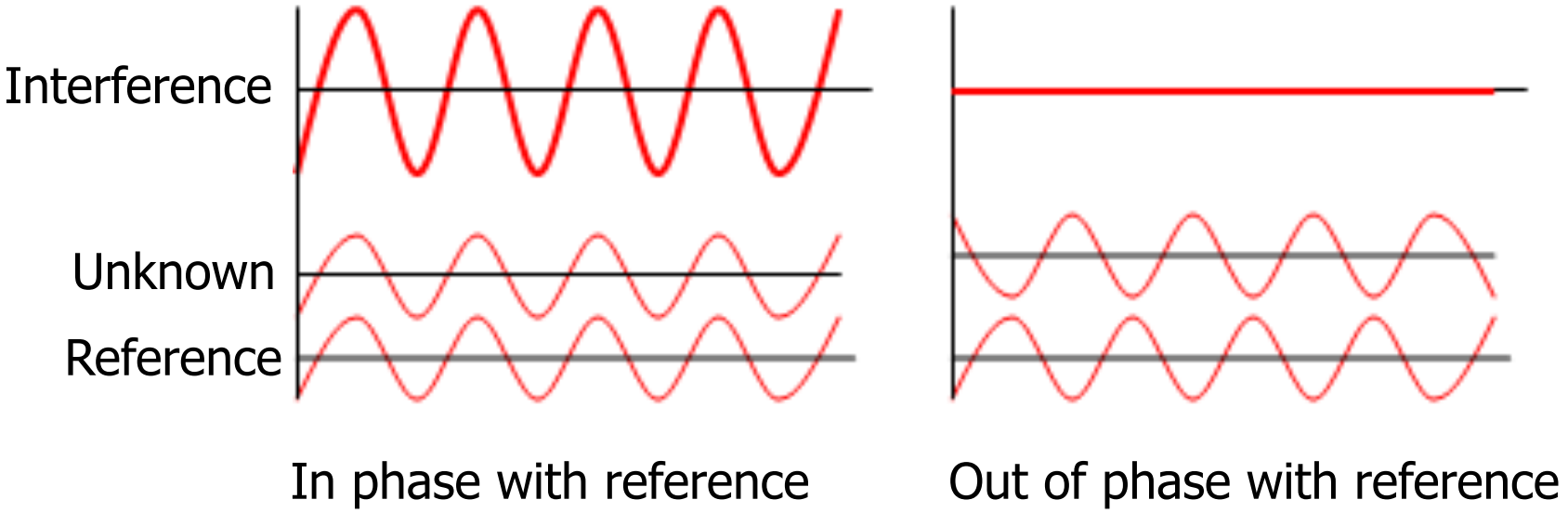


$$E \propto A^2$$

because they respond to the energy of the wave

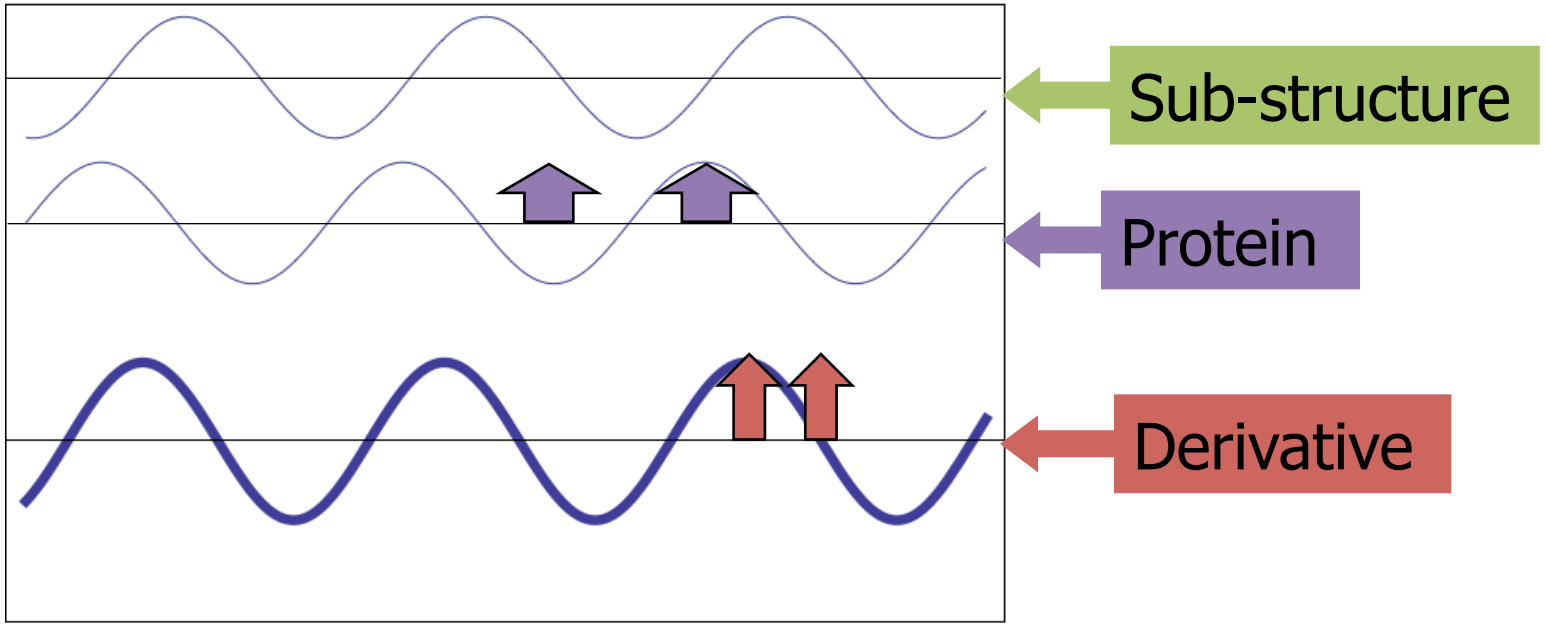
Reference Wave

- To phase a wave of known amplitude by interference, look at the amplitude of the wave after interference with a reference wave of known phase and amplitude



Interference and Phase

- The phase and amplitude of the reference wave are known from calculating them from a substructure of atoms

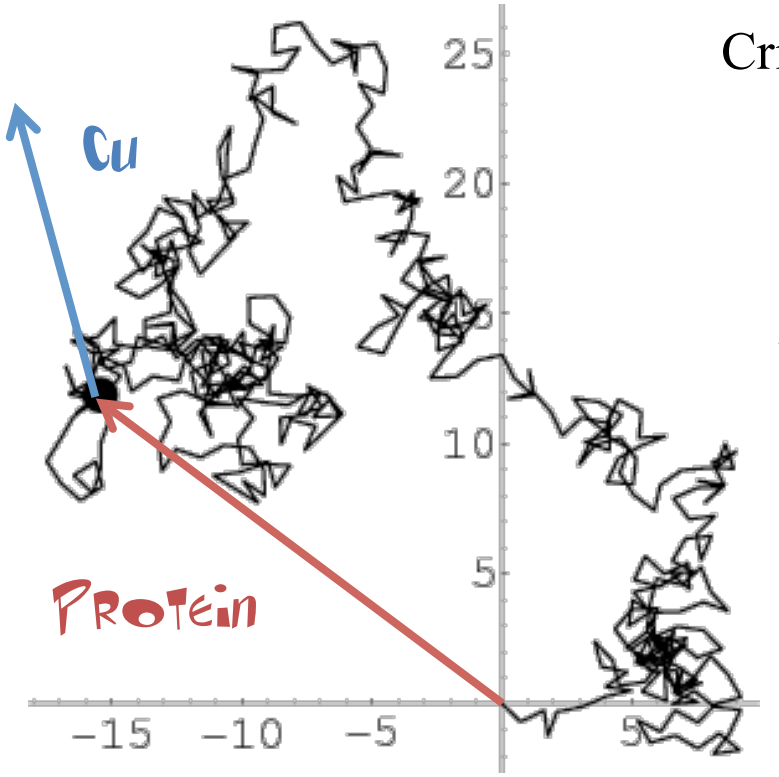


Substructure determination

- How do you determine the substructure?
- If you had the intensities of the substructure atoms floating in space you could solve the substructure
  - Patterson methods
  - Direct methods
  - Dual space methods
- PROBLEM
  - You don’t have these intensities!
- SOLUTION
  - Use approximations (see e.g. Blundell & Johnson 1976)

Heavy Atom Scattering is Significant

- How can e.g. a single Cu atom in a 100 kDa protein make any difference to the intensities?
- Structure factors add up as a "random walk"

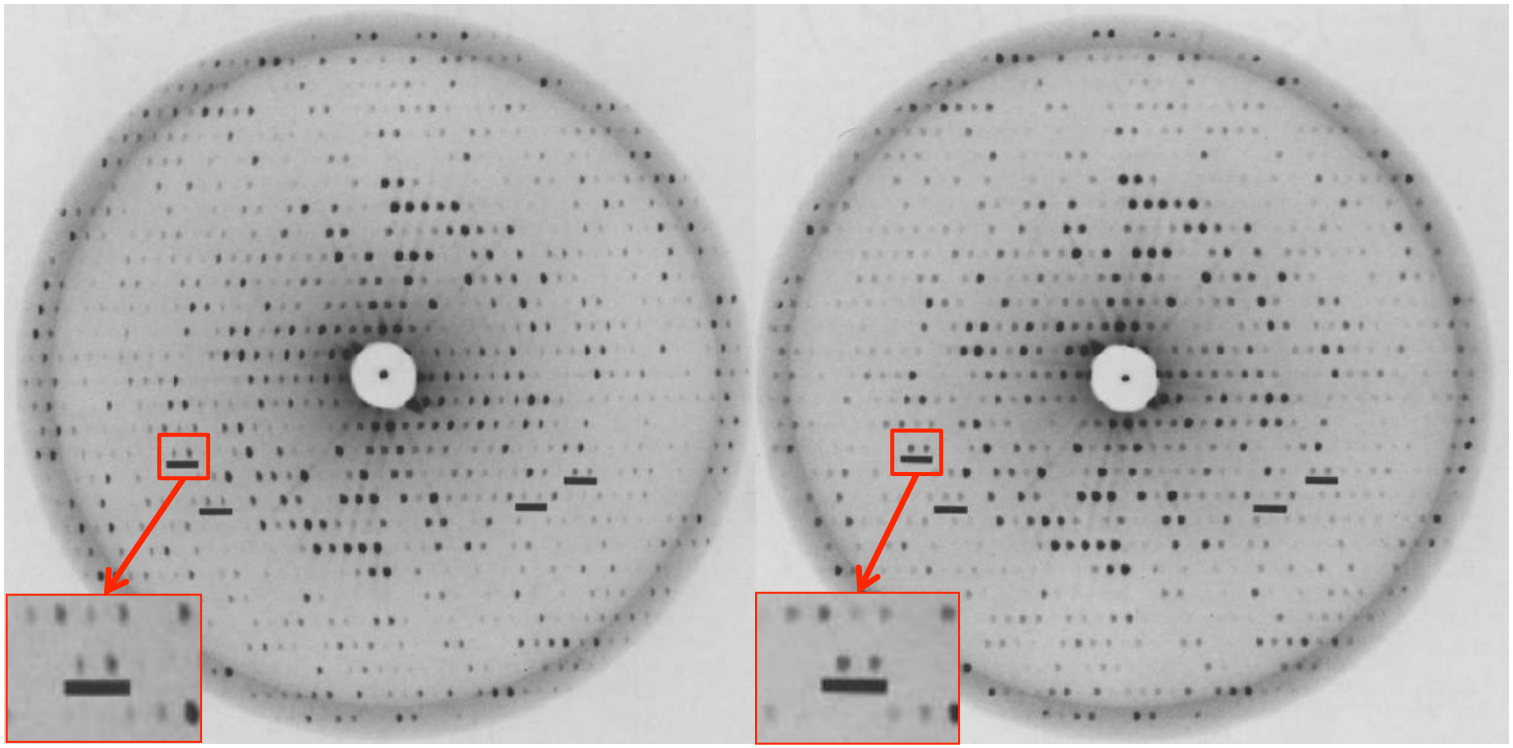


Crick and Magdoff (1956)

$$\frac{\langle \Delta F \rangle}{\langle |F| \rangle} = \frac{Z_H}{Z_{eff}} \sqrt{\frac{2N_H}{N_p}}$$

N_h, N_p are number of heavy, protein atoms
Z_h, Z_{eff} are atomic numbers of heavy atom and average of protein atoms (~6.7)

100 kDa protein with Copper (Z=28)
5.6%



Native

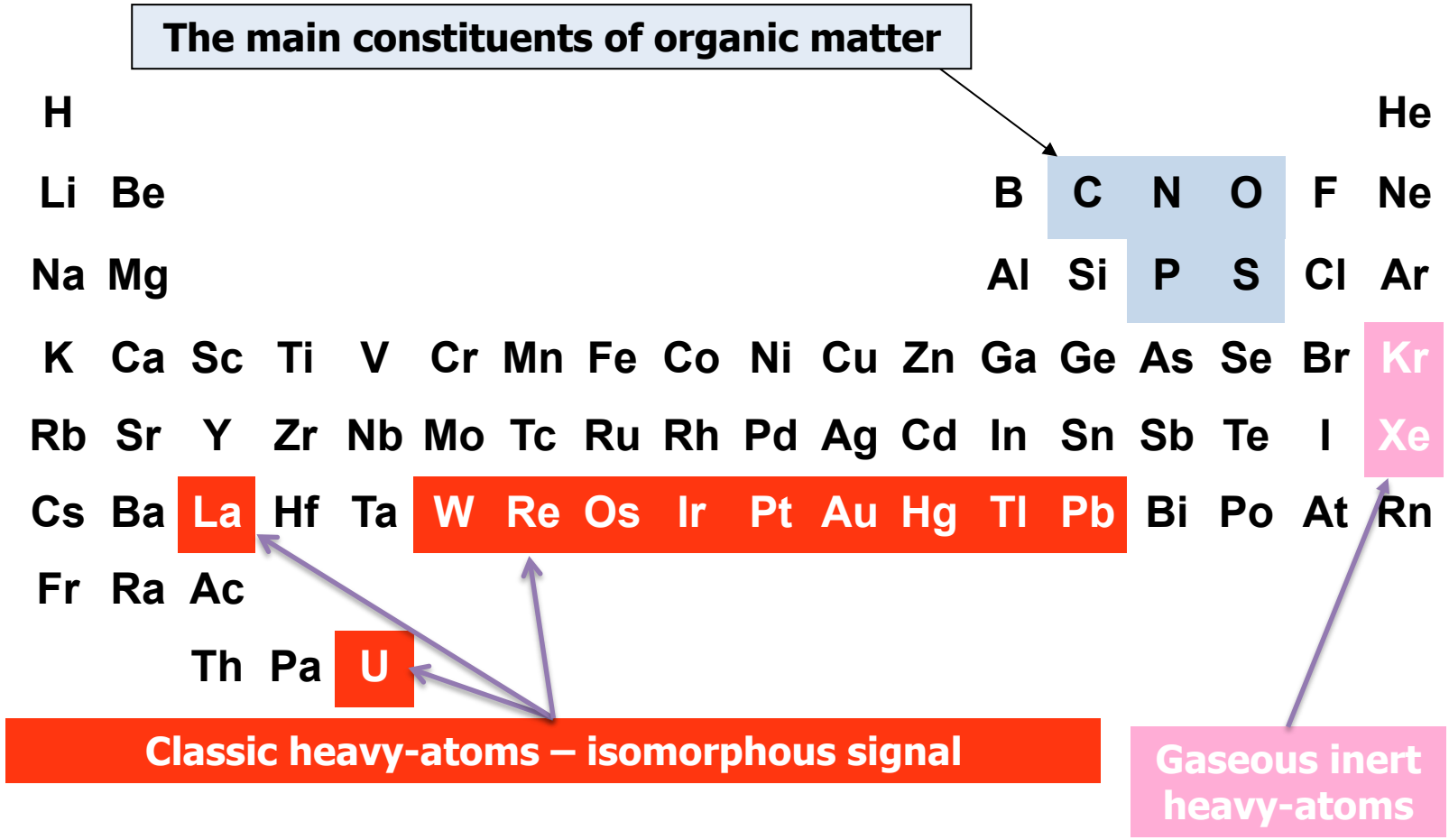
Heavy atom derivative

Intensity reversals (underlined)
indicate heavy atom has bound

Isomorphous Replacement

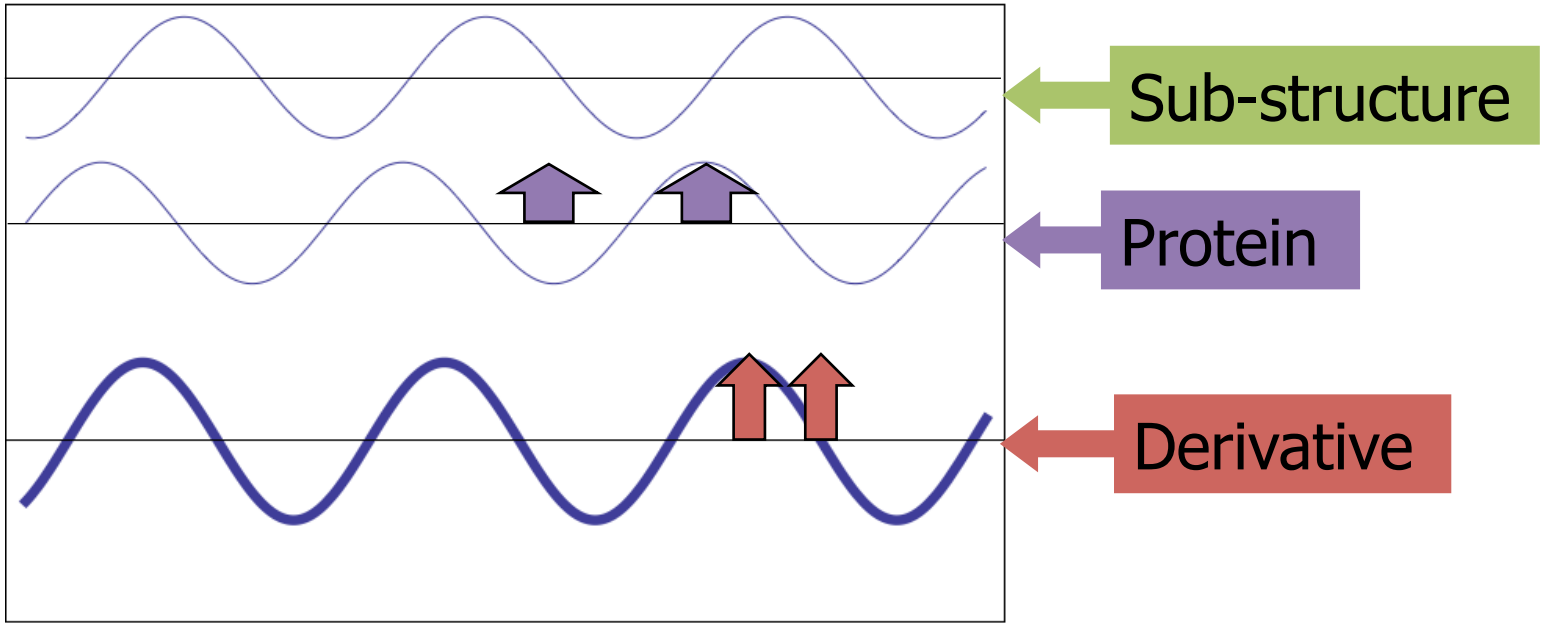
- Native and derivative must be isomorphous
  - Same unit cell and space group
  - Same position and orientation of protein in unit cell
- Can require searching many different compounds to find one or two isomorphous ones
  - Merge datasets from different crystals
  - e.g. BLEND to find isomorphous (partial) datasets

Isomorphous replacement

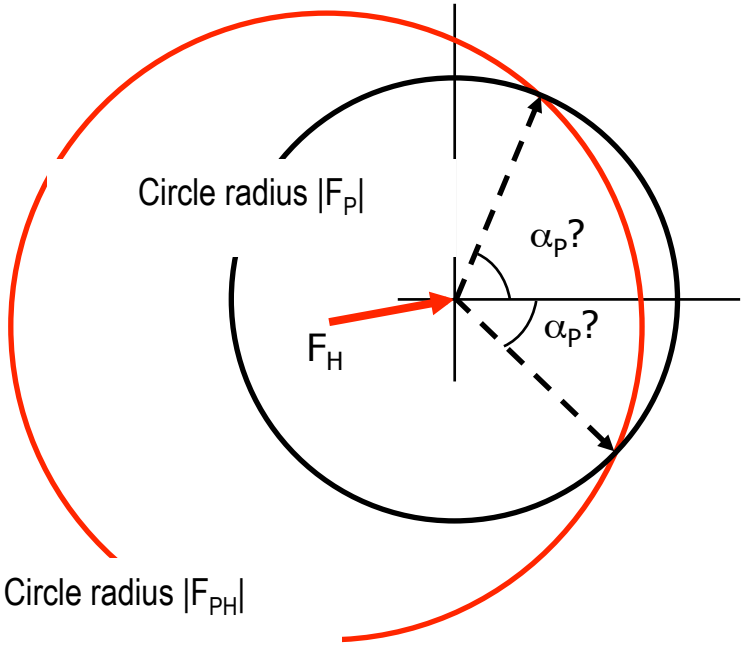


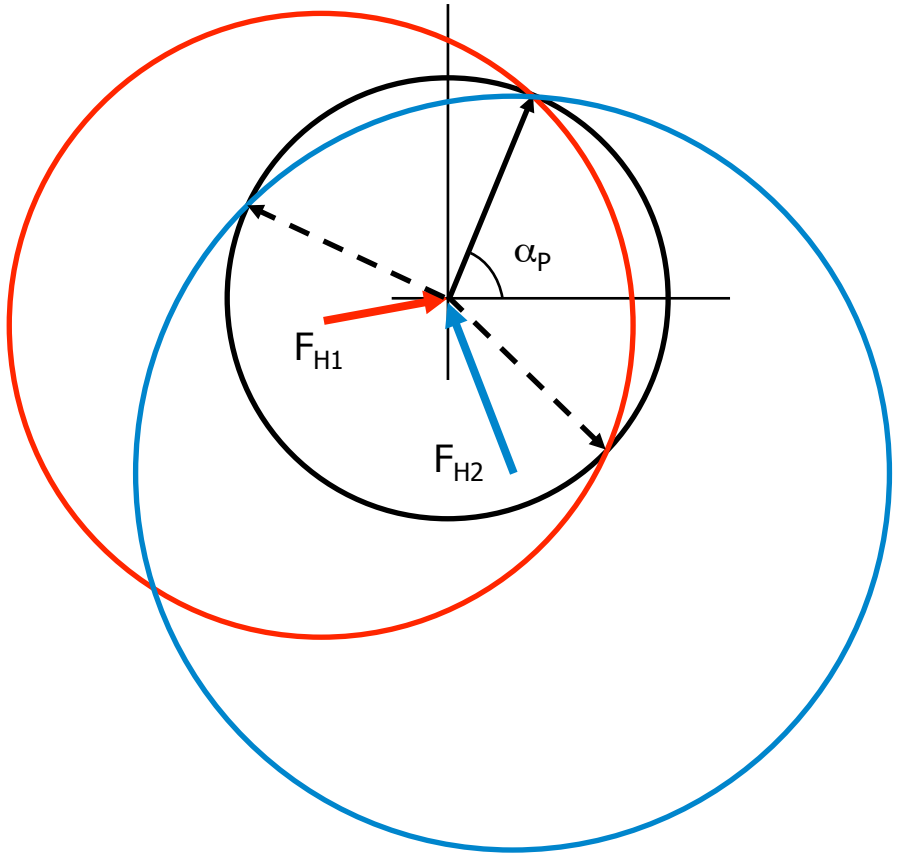
Harker Diagrams

- Two solutions for phase of unknown wave
- Also two solutions for phase of interference wave



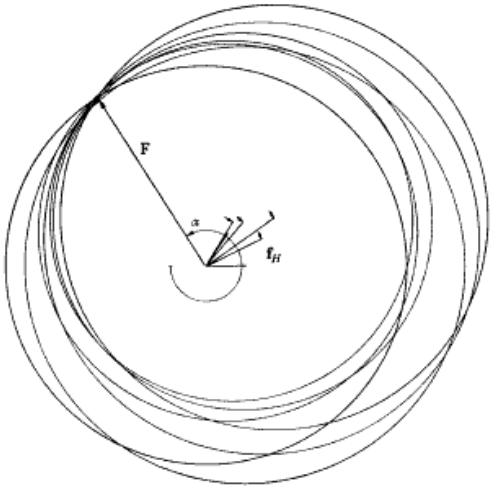
Harker Diagrams

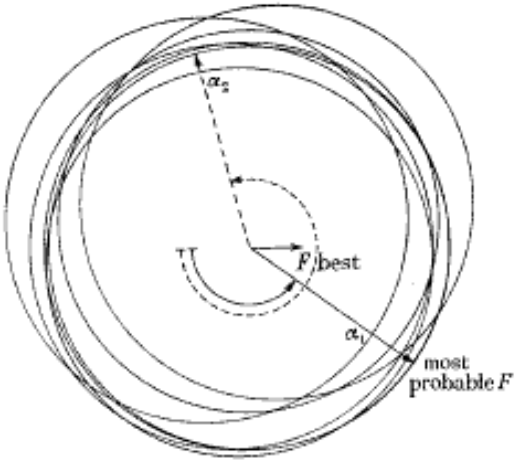


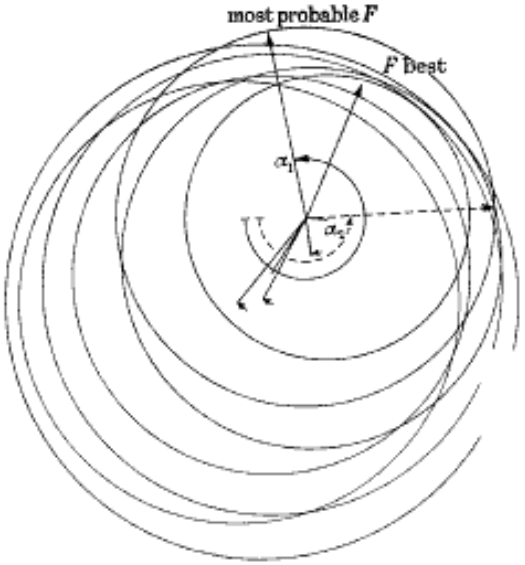


Reality…

- Some real Harker diagrams from the phasing of haemoglobin with 6 derivatives



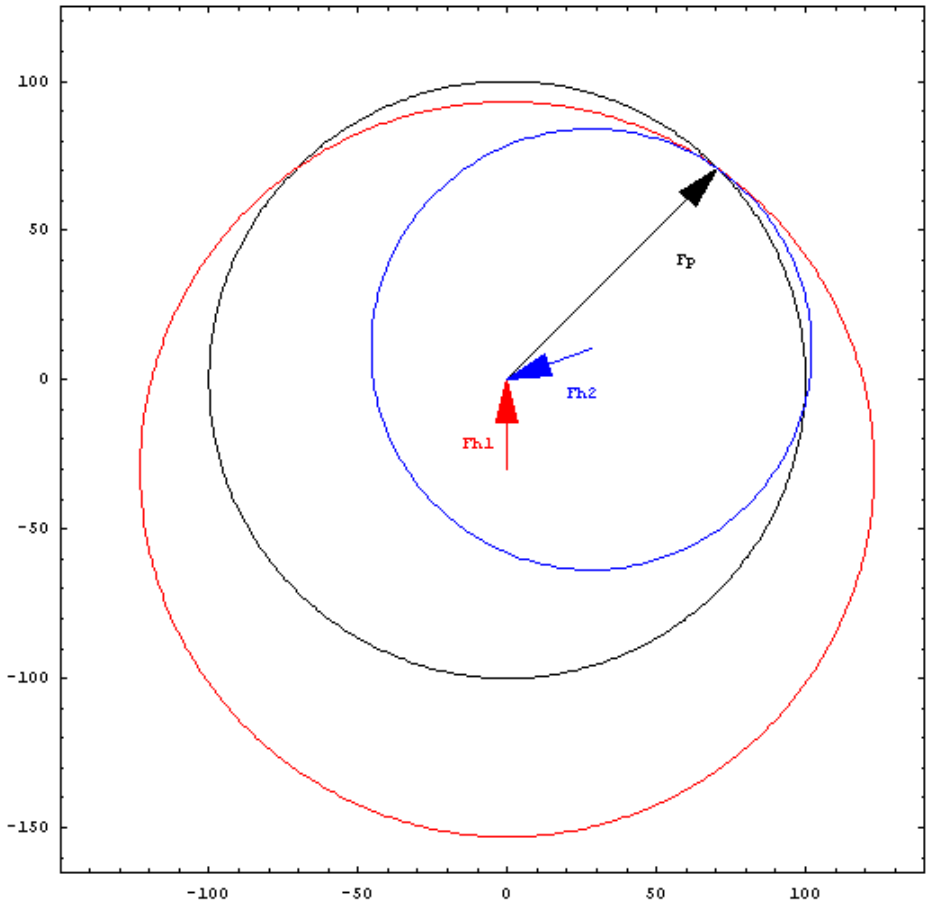




- Phase circles rarely cross exactly
- Need a **probabilistic** approach to determining the phase

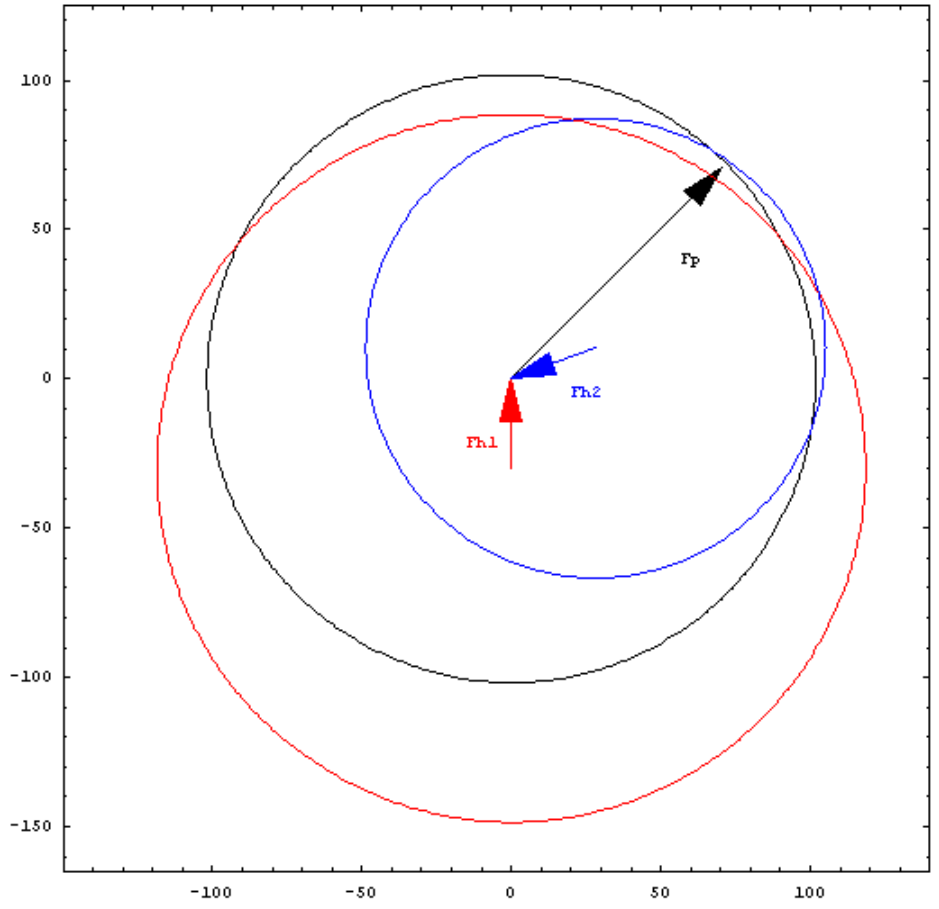
Harker construction

- Phasing of one refection using two derivatives with no errors
- Phase determined with very high probability



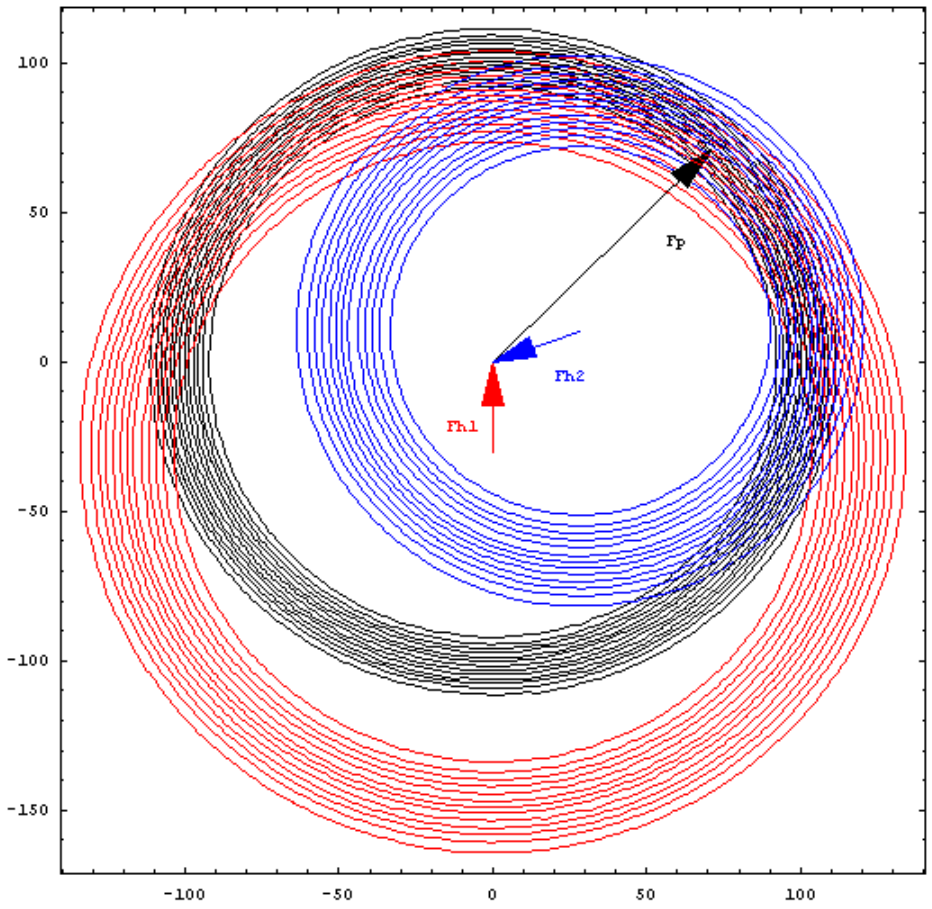
Harker construction

- There are many sources of error in the experiments
  - Mainly model errors
  - Also data errors
- The errors are large
- We are looking for the best phase
- We therefore need a probability function



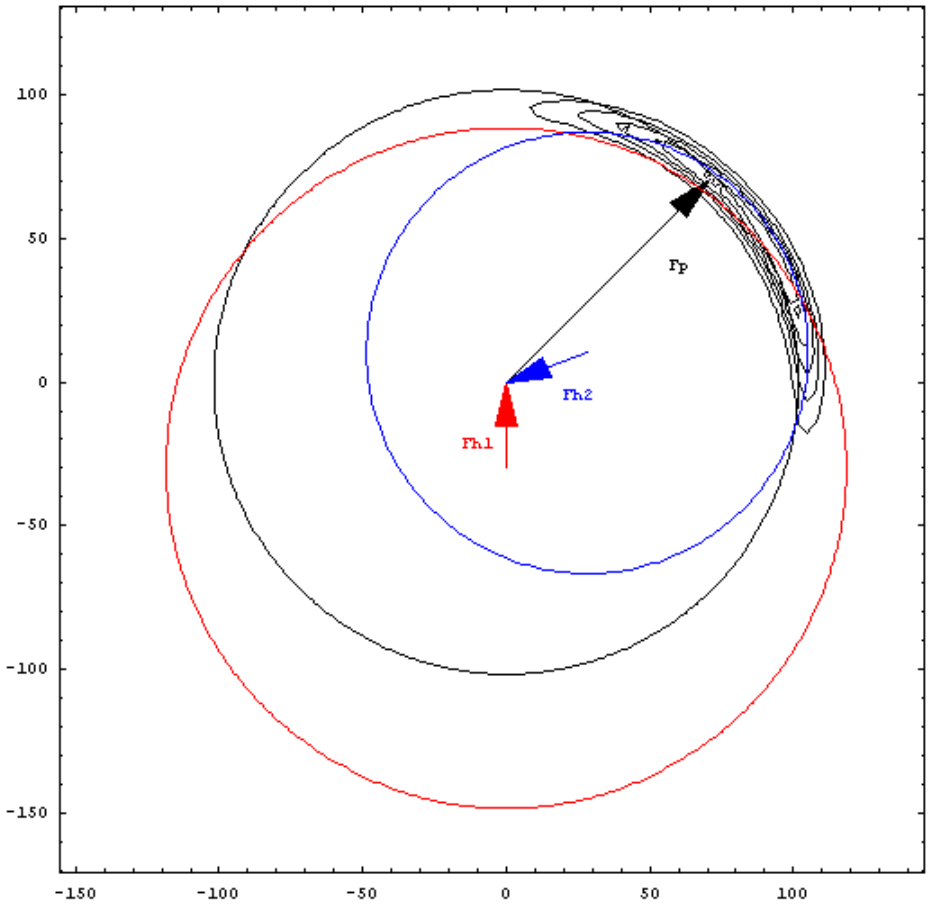
Probabilistic Harker Diagram

- Each circle has an error associated with it to give a distribution
- The total likelihood is the volume under the curve of the product of the distributions



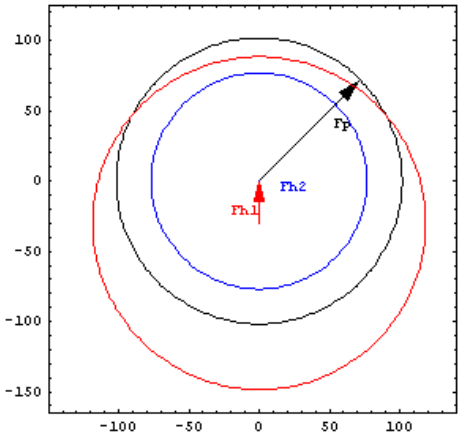
Probabilistic Harker Diagram

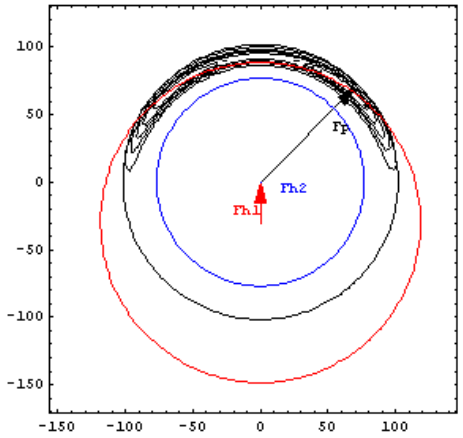
- The final distribution is high only where all three circles overlap



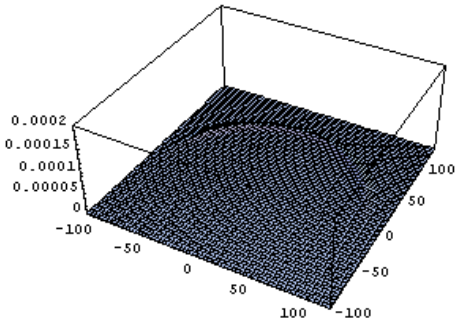
Refining Occupancy

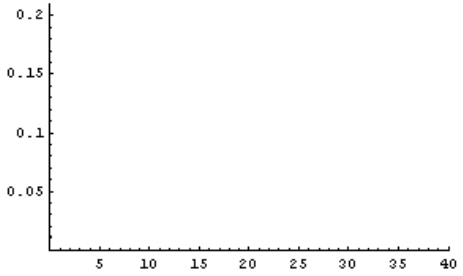
To refine the occupancy of a heavy atom, maximise the likelihood (area under the curve)





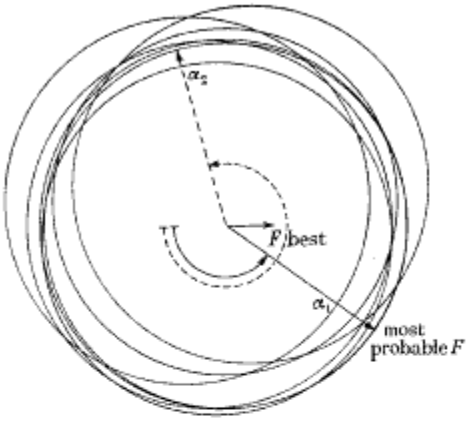
Final refined value is the optimum for ALL reflections (movie shows ONE reflection)



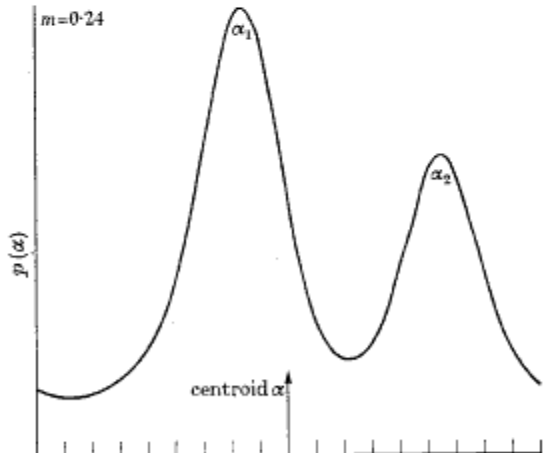


Phase probability

- Each reflection really has a phase probability density function (PDF) rather than a single phase
- This is a complicated mathematical function
  - Requires lots of memory
- Four Hendrickson-Lattman coefficients (A,B,C,D) are used to store this PDF in a compact form



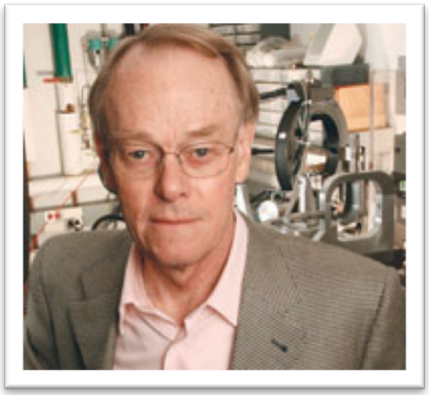
(b)



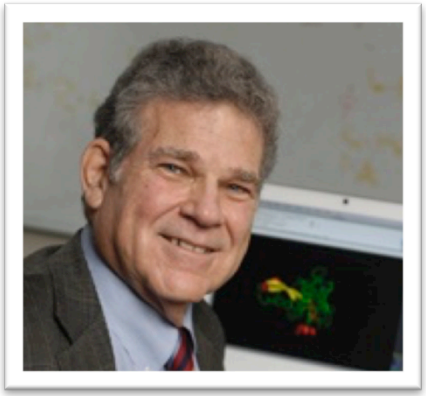
Hendrickson-Lattman Coefficients

$$P(\alpha) \propto \exp\left[A\cos(\alpha) + B\sin(\alpha) + C\cos(2\alpha) + D\sin(2\alpha)\right]$$

- HL coefficients allow for easy combination of phase information from multiple sources
  - the combined PDF is formed simply by adding the A,B,C, and D from the two distributions



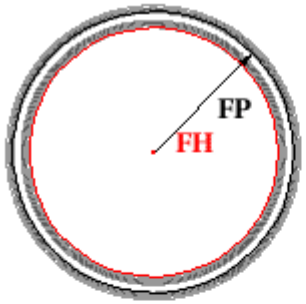
Wayne Hendrickson

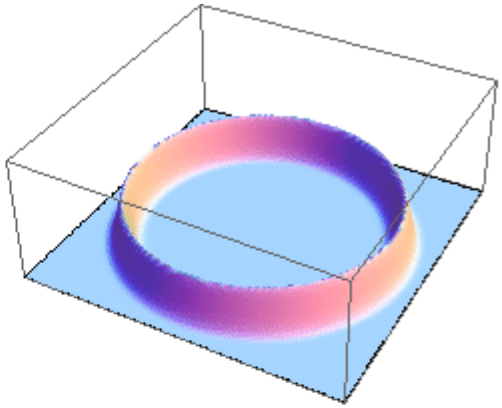


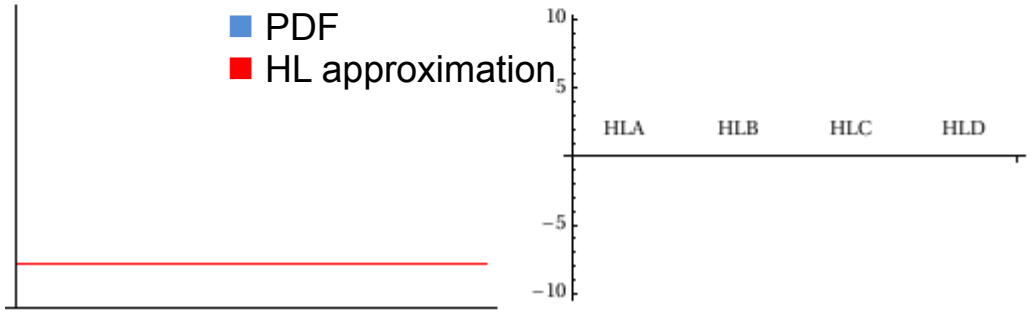
Eaton Lattman

Hendrickson-Lattman Coefficients

HL coefficients as a function of F_H occupancy








MIR
SAD
SIR
SIRAS
MIRAS
MAD
RIP

SAD

- The most popular way of solving structures by experimental phasing (over 70%)
- Can be done with intrinsic S and CuKα X-rays
- SAD phasing theory is very good
- Easy to automate
- Can be very fast
  - Can be done from single dataset
- May need multiple crystals
  - And careful data processing

Anomalous Scattering

A·nom·a·lous

adj.

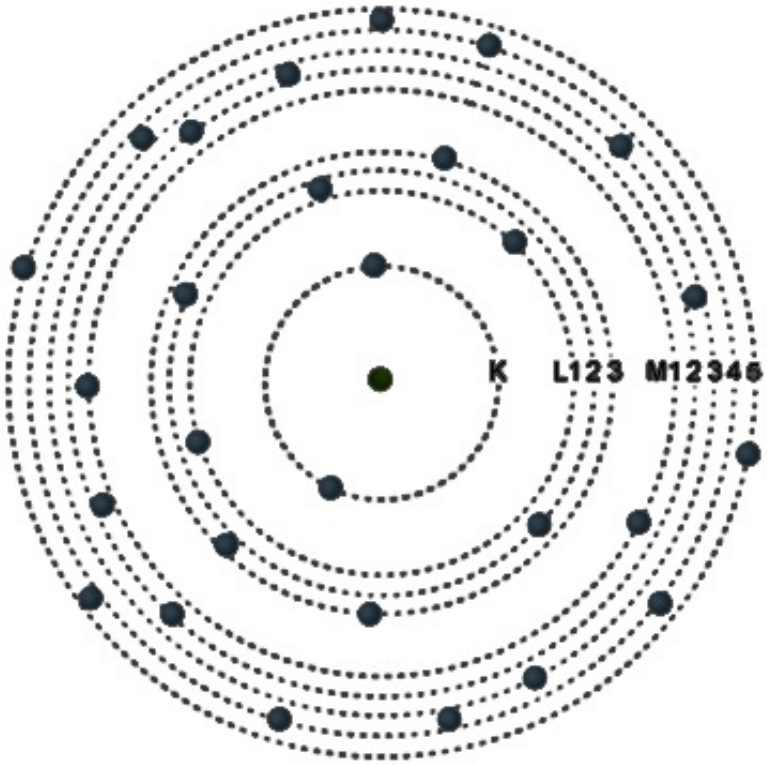
Deviating from the normal or common order, form or rule

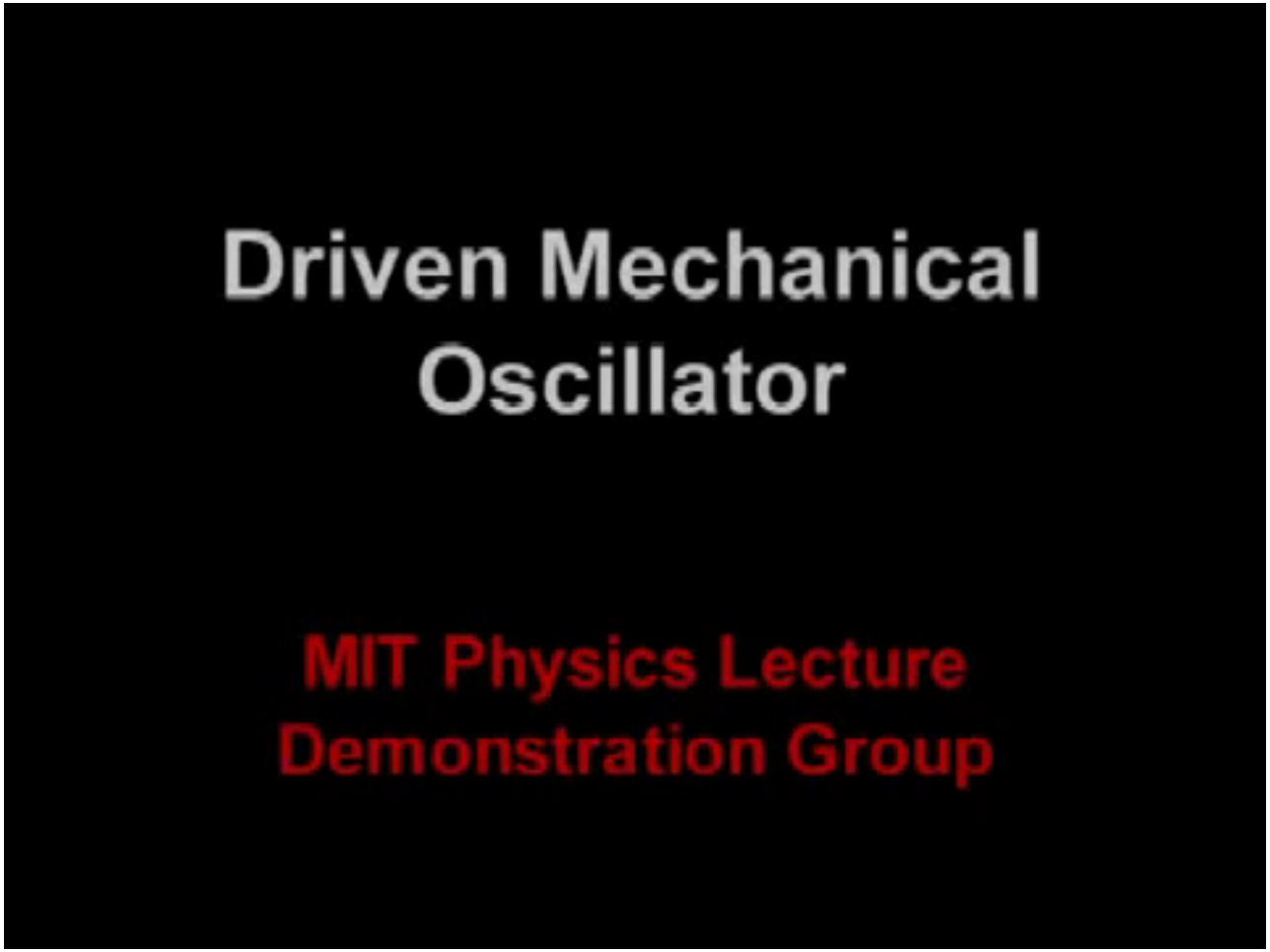
"**Anomalous** scattering" is **absolutely normal** while "**norma**l scattering" occurs only as an ideal, over simplified model, which can be used as a first approximation when studying scattering problems"

IUCR Pamphlet "Anomalous Dispersion of X-rays in Crystallography"
S. Caticha-Ellis (1998)

Anomalous Scattering

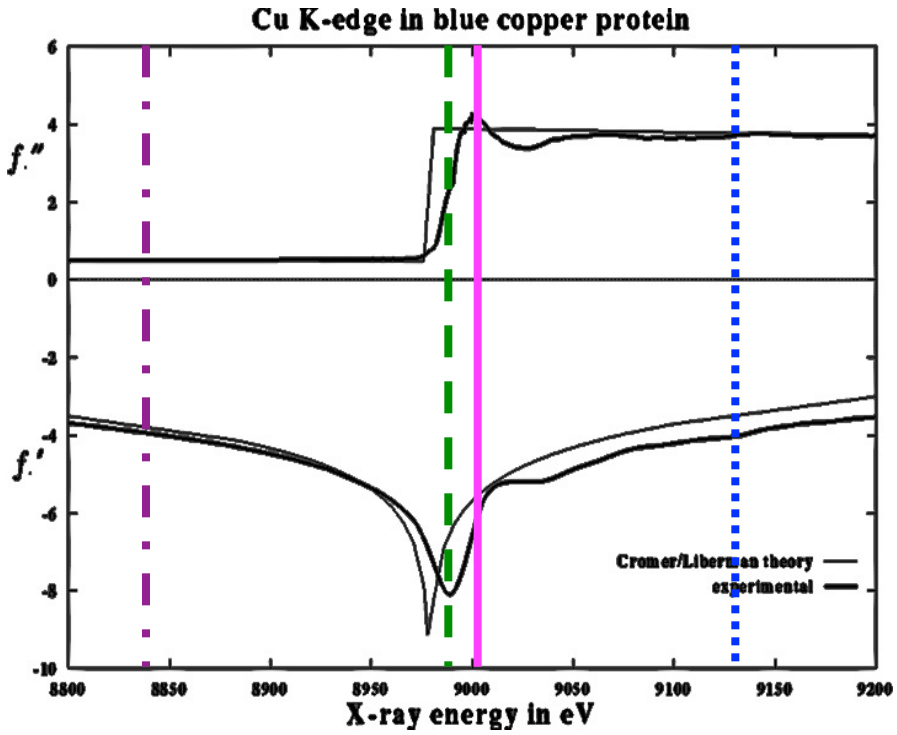
- Anomalous scattering is due to the electrons being tightly bound (particularly in K & L shells)
- In classical terms, the electrons scatter as though they have resonant frequencies





https://www.youtube.com/watch?v=aZNnwQ8HJHU

Wavelength Choice



PEAK: |f”| is large

INFLECTION: |f’| is large

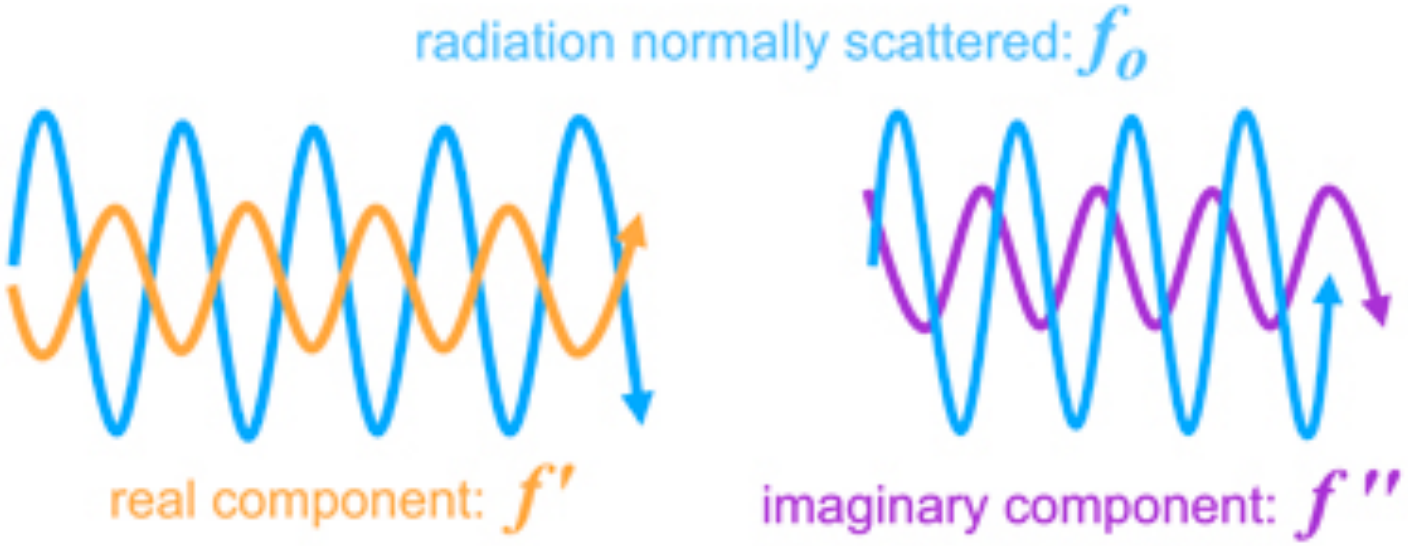
REMOTE: low energy

REMOTE: high energy

Warning
Radiation
Damage
INCREASES
at peak

Anomalous Scattering

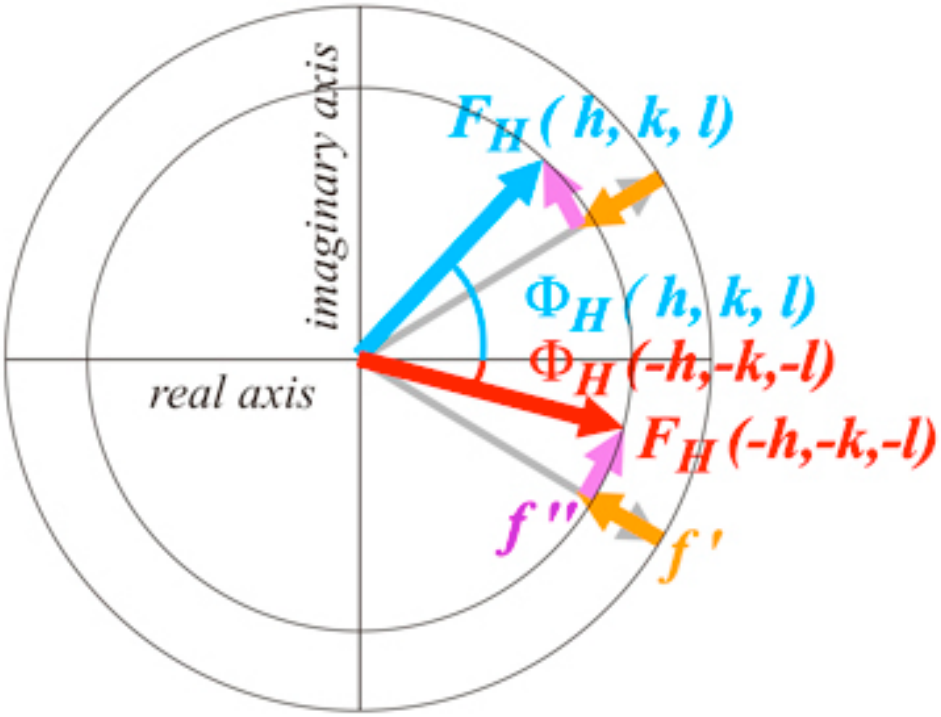
$$f = f_o + f' + if''$$



The real component (f ') is 180 degrees out of phase with the normally scattered radiation.

The imaginary component (f ") is 90 degrees out of phase.

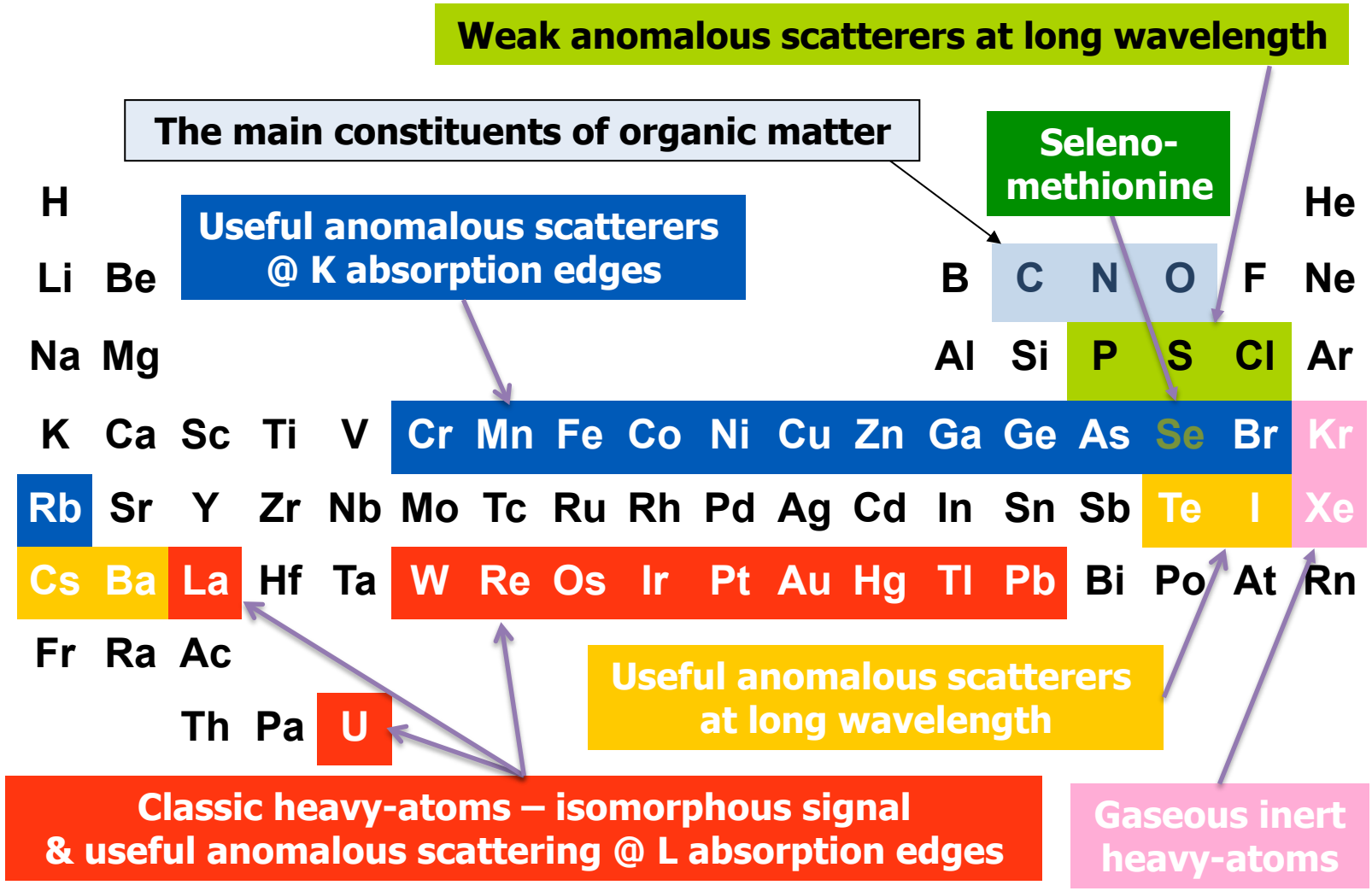
Anomalous Scattering



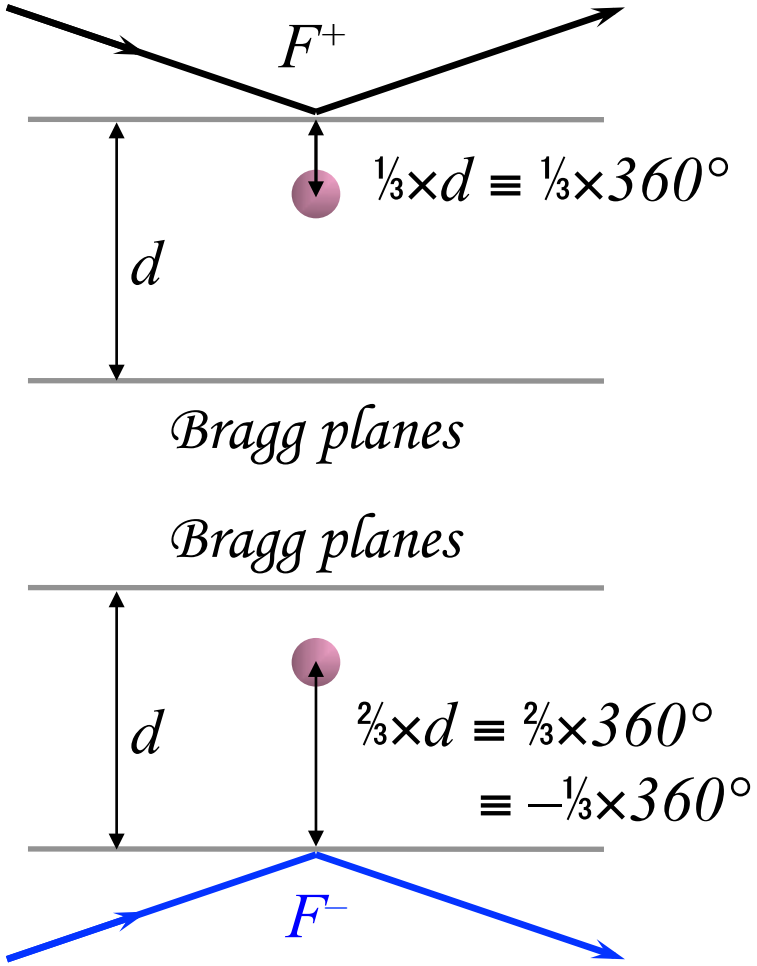
$$I_H(h, k, l) = I_H(-h,-k,-l)$$

$$\Phi_H(h, k, l) \neq -\Phi_H(-h,-k,-l)$$

Anomalous Scatterering

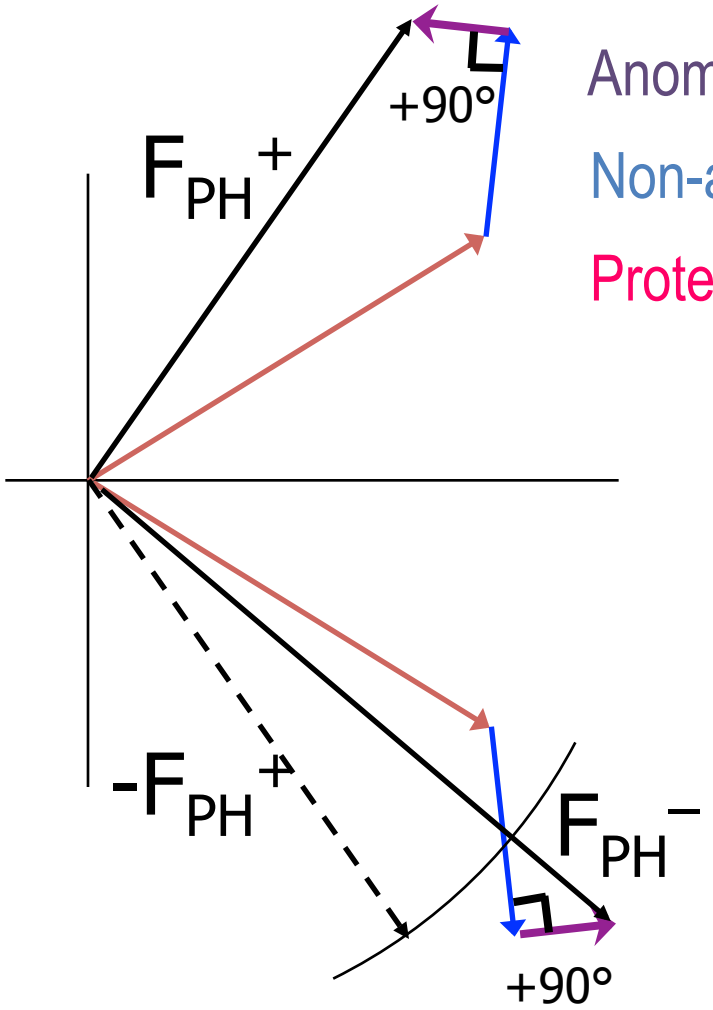


SAD Harker Diagram



- Phase of non-anomalous component depends on the atom's location between Bragg planes
- Anomalous scattering component is always advanced by 90°
  - irrespective of position of atom between Bragg planes

SAD Harker Diagram



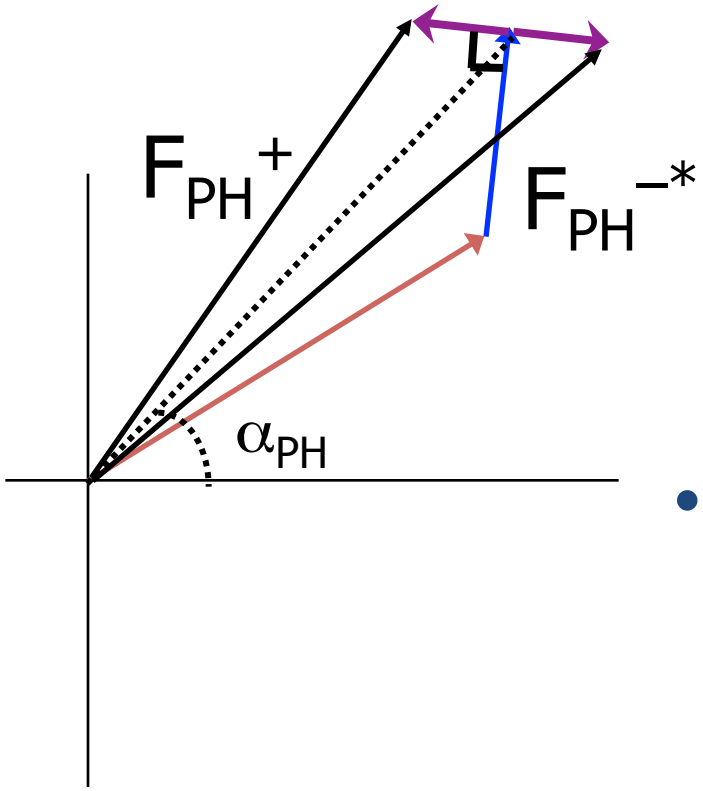
Anomalous part of heavy atom

Non-anomalous part of heavy atom

Protein (non-anomalous scattering)

- Non-anomalous and anomalous scattering components sum differently in the Friedel mates F_{PH}^{+} and F_{PH}^{-}
  - $F_{PH}^{+}$and F_{PH}^{-} have different intensities

SAD Harker Diagram



Anomalous part of heavy atom

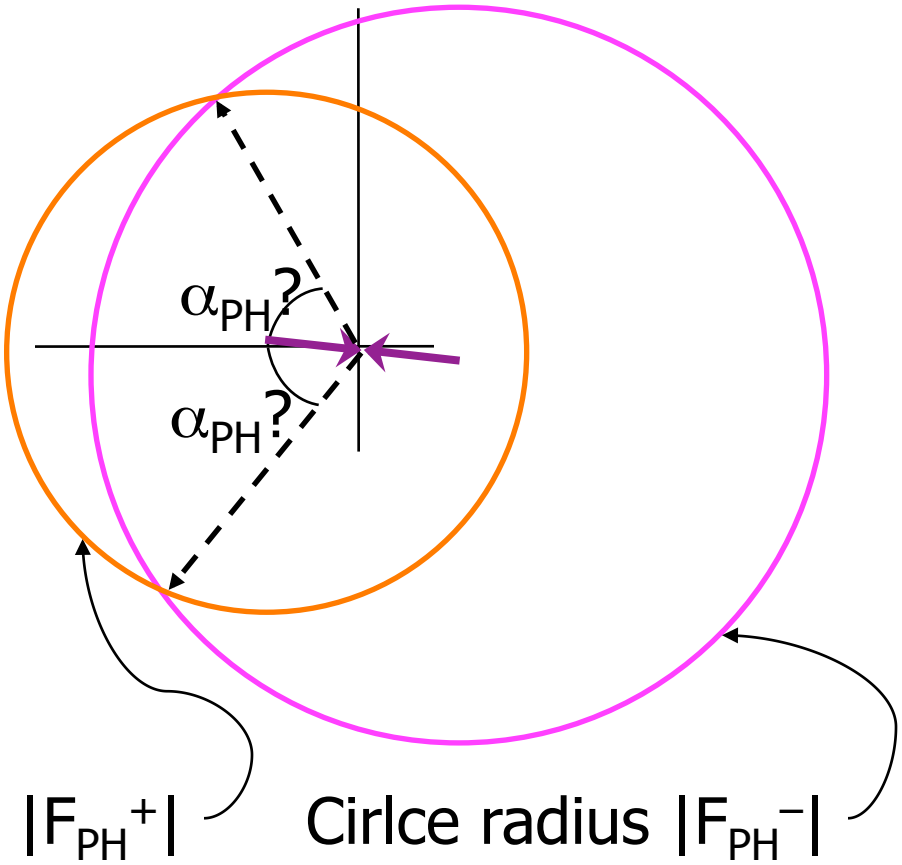
Non-anomalous part of heavy atom

Protein (non-anomalous scattering)

- It is easier/usual to use the complex conjugate of all Friedel F^{-}
  - i.e. all structure factors for (-h,-k,-l)
  - Phase F_{PH} (α_{PH})

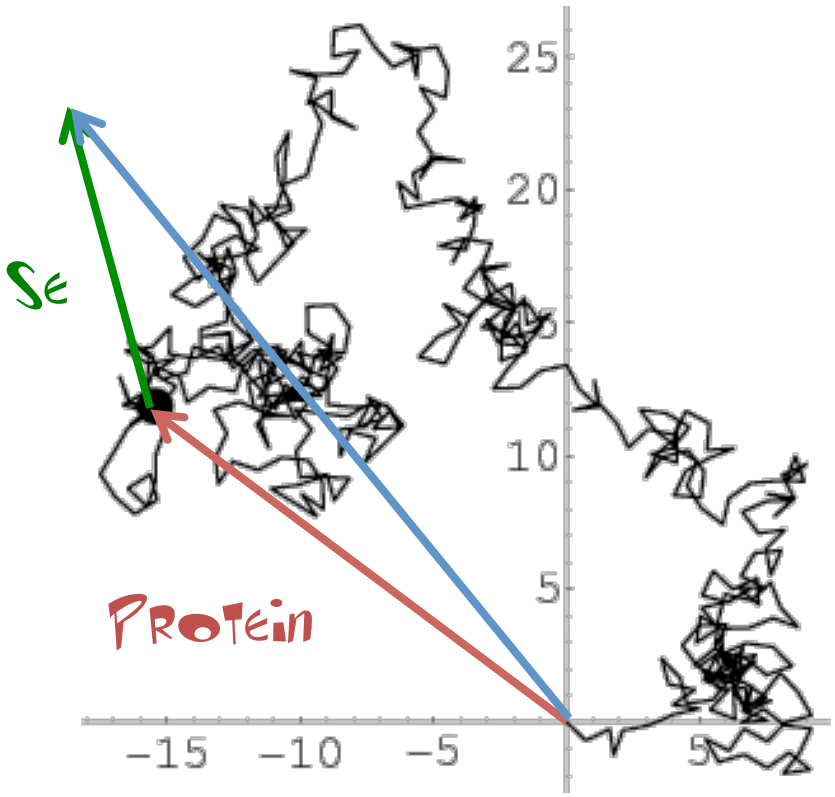
SAD Harker Diagram

- There is a phase ambiguity for FPH given by the two places where the two circles with centres displaced by the anomalous scattering vectors overlap



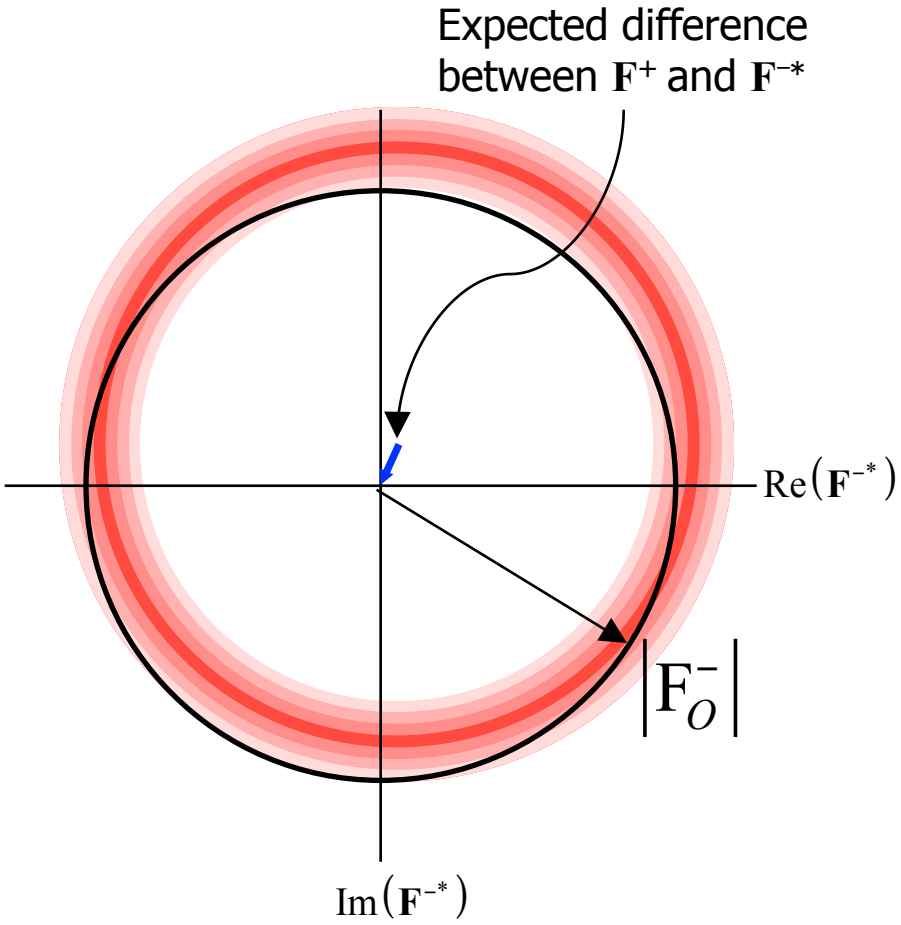
Partial Structure

- But there is more phase information!
- The anomalous scatterer is also a part of the structure
  - Gives additional phase information
  - Begins to break phase ambiguity



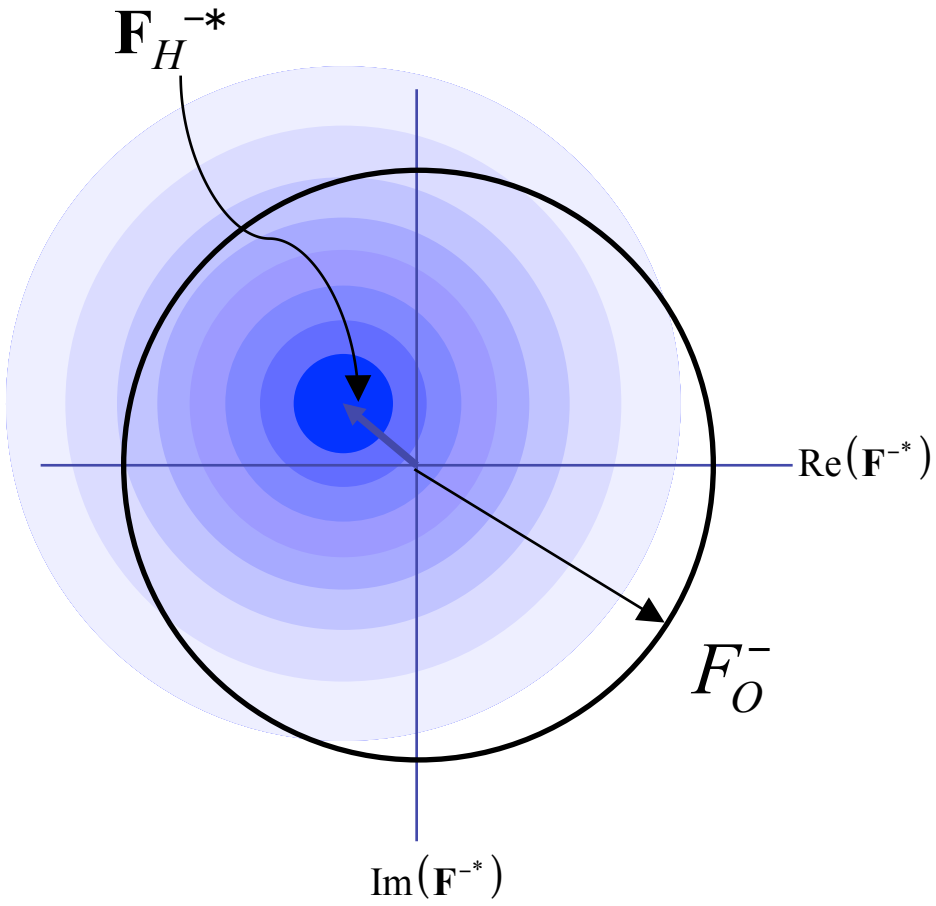
SAD Phasing

- Primarily anomalous scattering
- Gives a narrow probability distribution



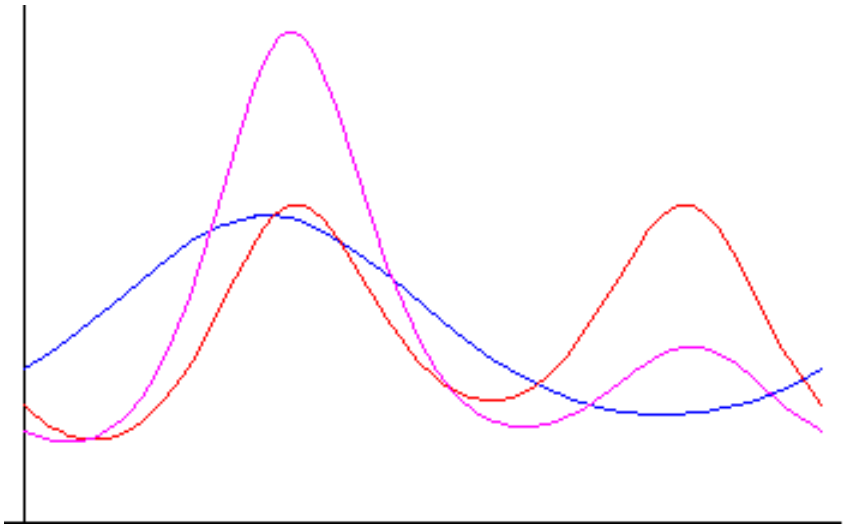
SAD Phasing

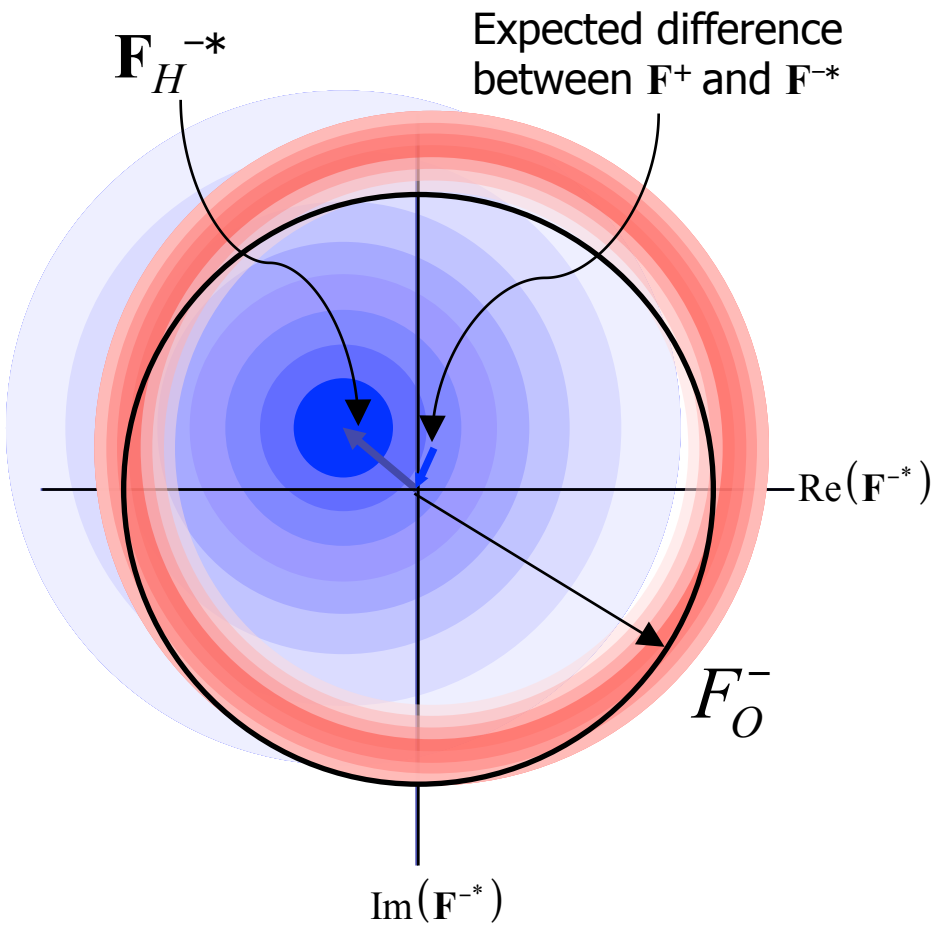
- Primarily Normal scattering
- Gives a broad probability distribution



SAD Phasing

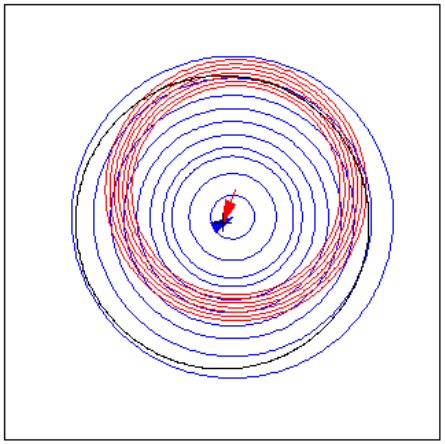
- Likelihood is proportional to the product of the two distributions (magenta) under the black circle

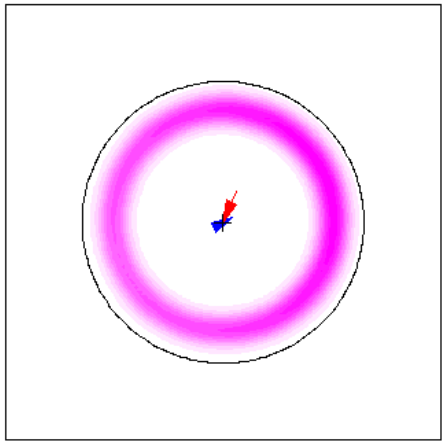




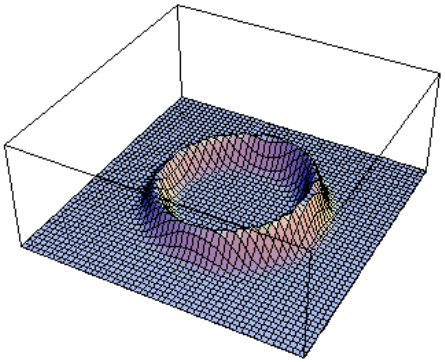
Refining Occupancy

To refine the occupancy of a heavy atom, maximise the SAD likelihood

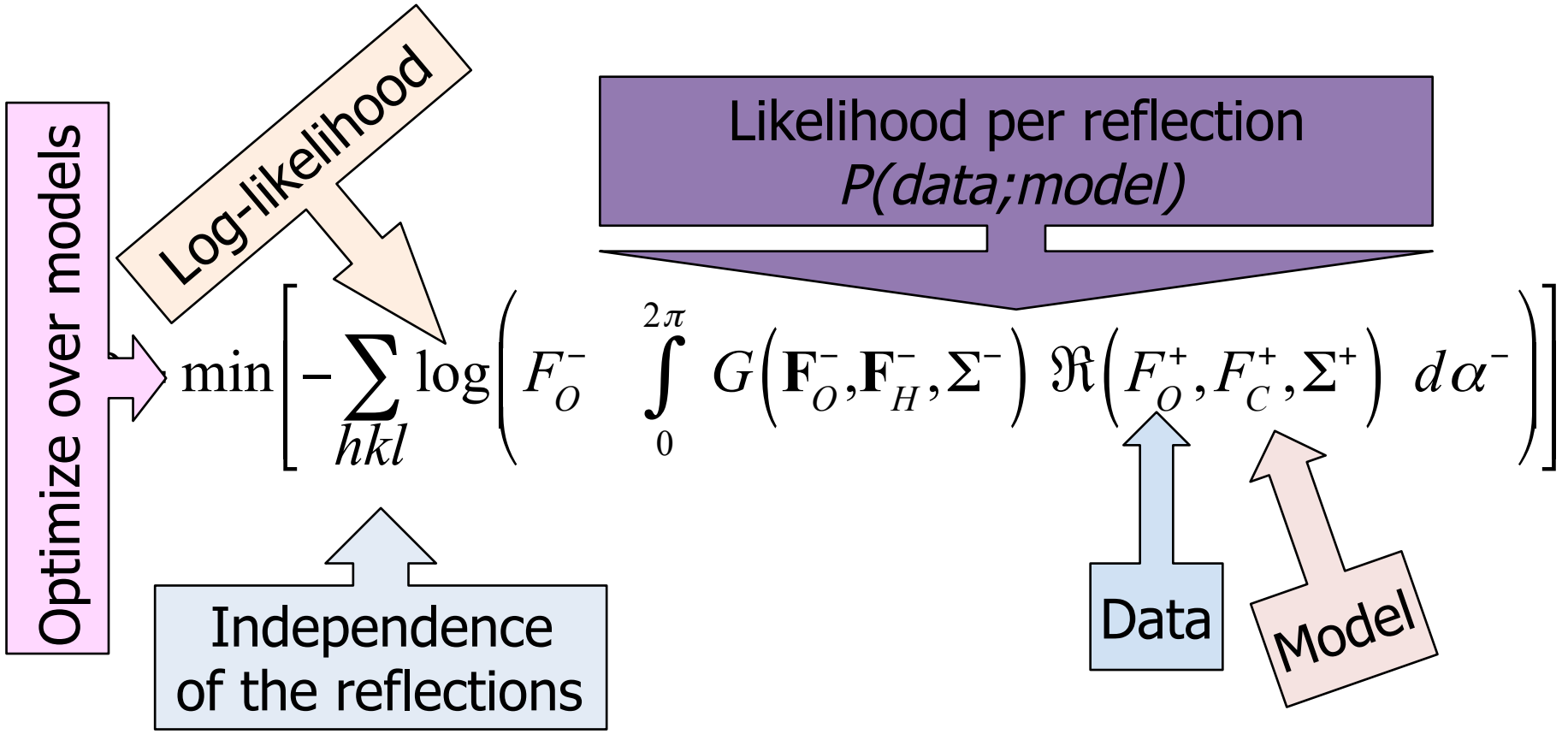




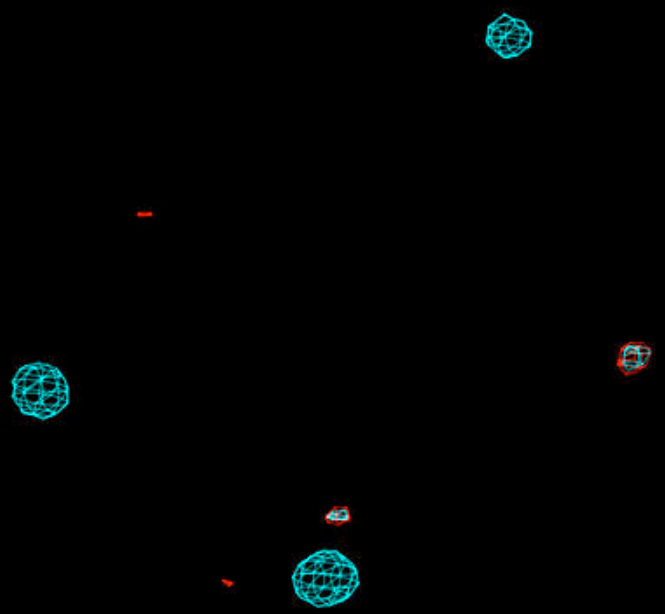
Final refined value is the optimum for ALL reflections (movie shows ONE reflection)



Correlated SAD Likelihood Function

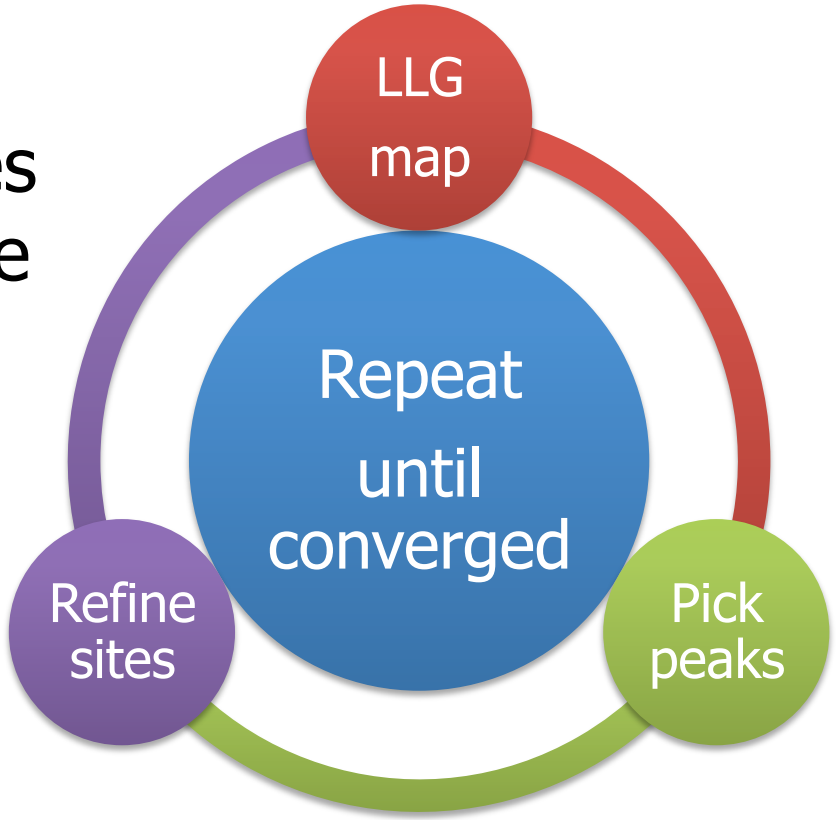


Log-likelihood gradient maps



Completion of sub-structure

- LLG maps are very sensitive
- Inclusion of minor sites **greatly improves** the phases
- Could include all intrinsic sulphurs
- Also finds bound halides



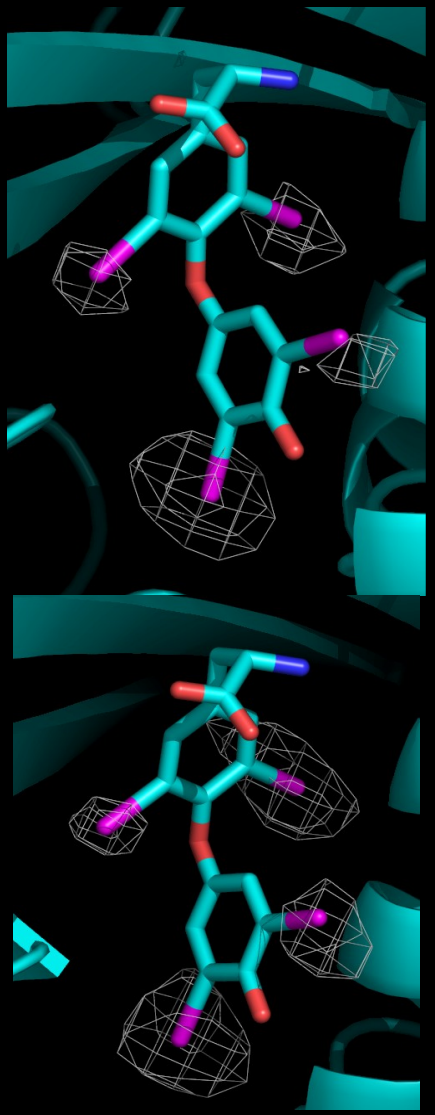
Thyroxine binding globulin

Where does thyroxine bind?

Thyroxine contains 4 iodine atoms

Two molecules in asymmetric unit
d_{min}=2.8Å
λ=0.979Å
f'' ≈ $3e^-$

Phaser LLG map @5.5σ

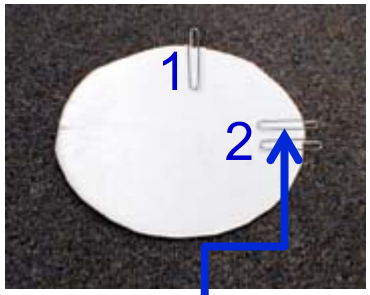


Zhou *et al*. (2006). *PNAS* **103**: 13321

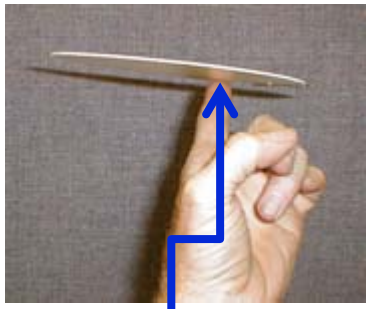
Calculating Electron Density

- ML function is good for refining the parameters, but what phase should be used in the electron density equation?
  - Have to pick one phase
- **We want the phase that gives the electron density with the lowest rms error**
  - Parseval's theorem relates the rms error in real space to the rms error in reciprocal space and vice versa
- This phase (the "**best phase**") is the probability-weighted average of all the phases
  - It is not the "most probable phase"

- Cut the centre out of a polystyrene foam plate
- Balance the disk on your finger
  - The centre of mass is at the centre
- Now put 3 paperclips on the edge of the disc
  - 2 together
  - 1 a distance away
- The balancing point is **between** the 3 paperclips
  - Not on the 2 paperclips

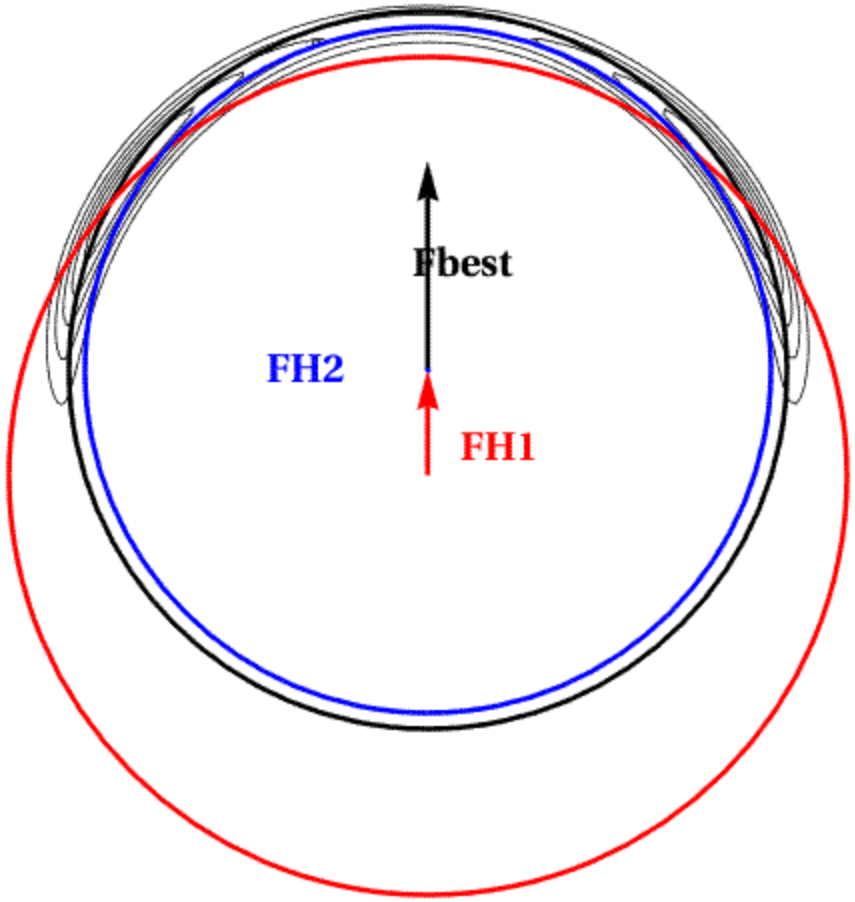


Most Probable Structure Factor



Best Structure Factor

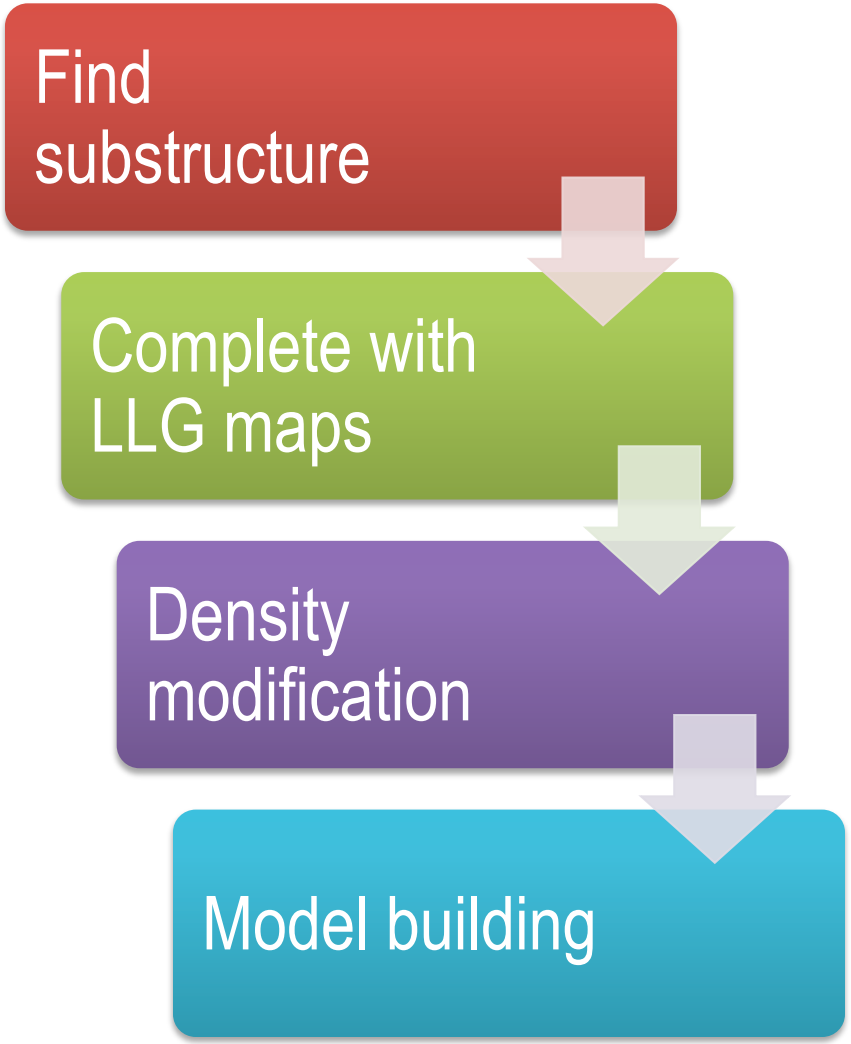
- F_{best} has a lower |F| amplitude than F_{obs}
- The reduction in F_{obs} to give F_{best} is expressed as the "figure of merit" (m)
  - **0<m <1:** F_{best} lies inside the F_{obs} circle
  - **m=1 :** Perfect phase information
  - **m=0:** No phase information
  - The higher the average value of the figure of merit, the better



New approaches

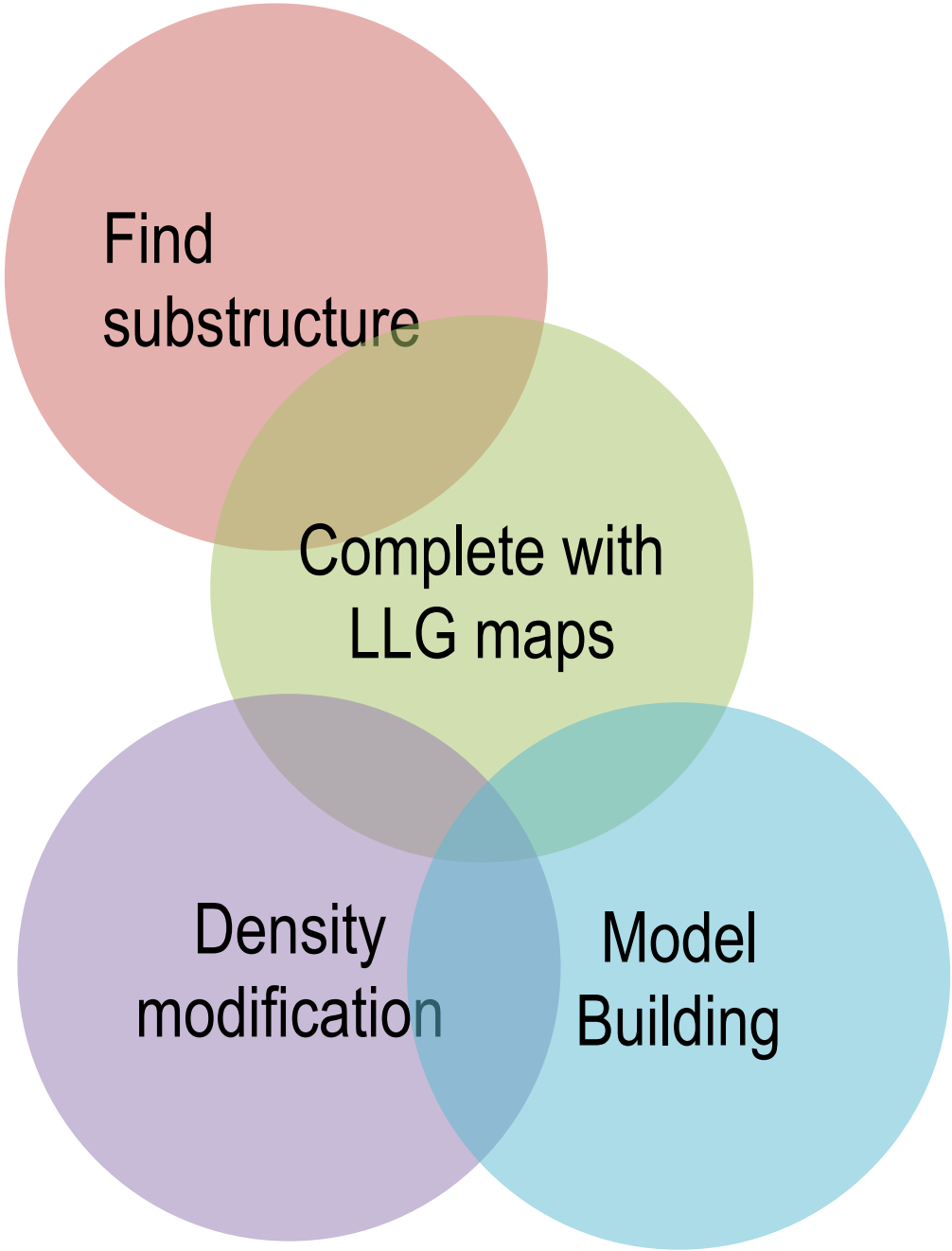
The pathway of structure solution

- Historically, there has been a linear progression through structure solution
- You had to be sure each step is correct before progressing to the next
- When signal is low you cannot be sure (of anything)

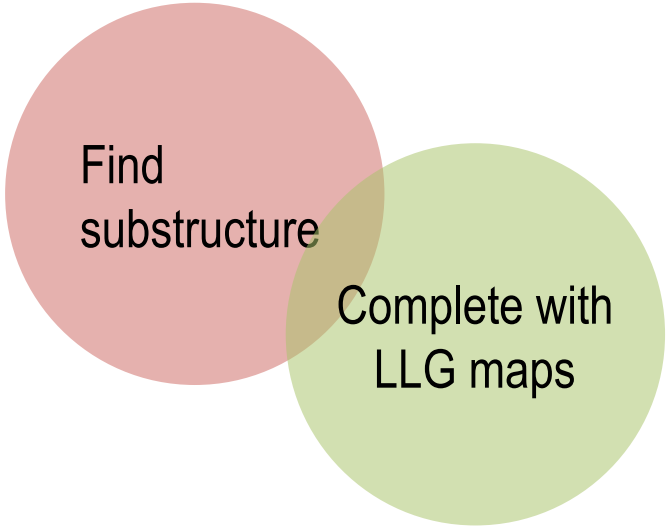


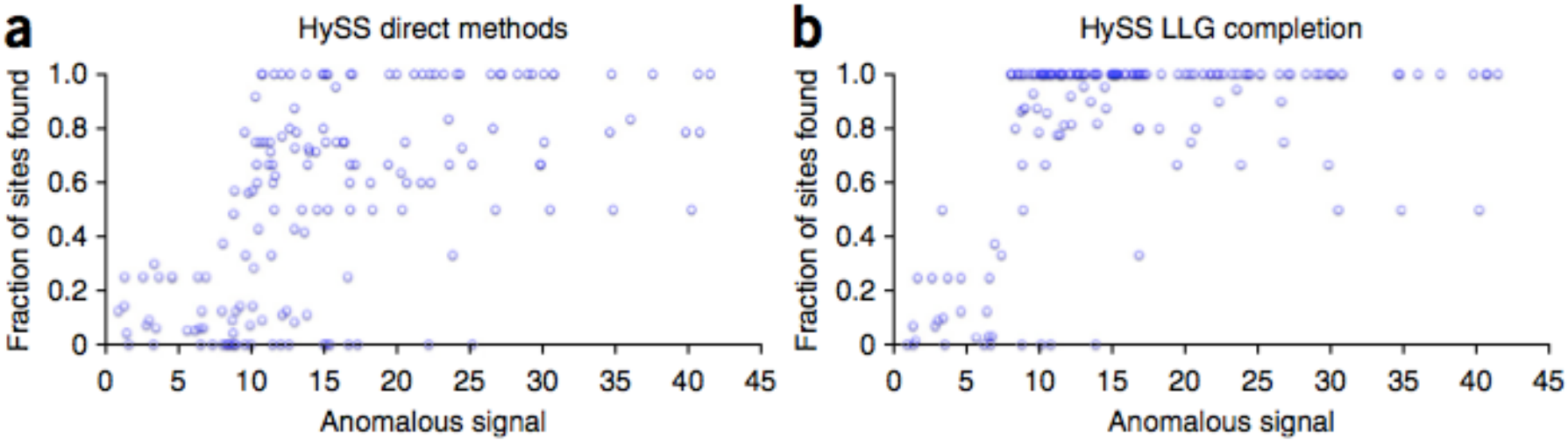
New approaches

- Take multiple possibilities for each step and uses subsequent steps to distinguish correct from incorrect solutions
- Enables structure solution when signal is low

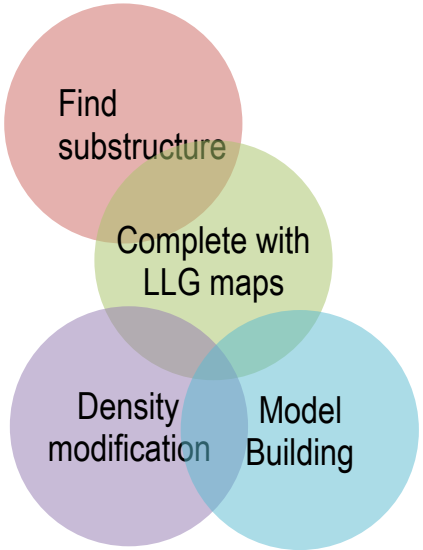


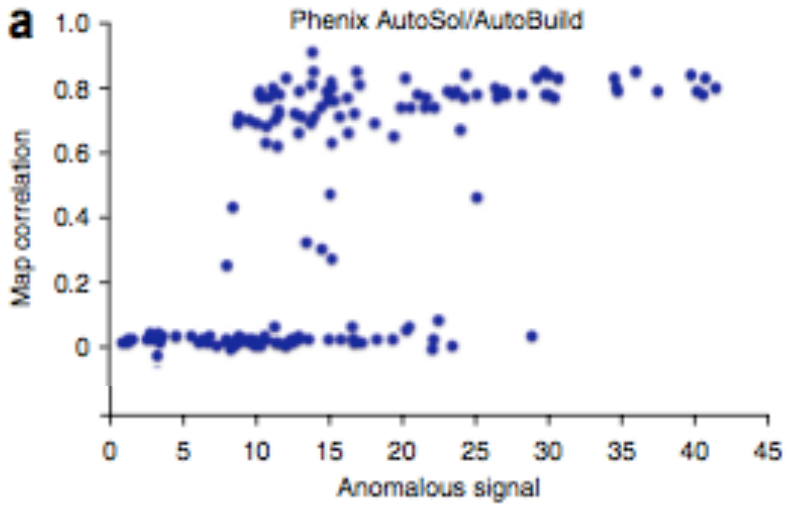
Find phasing substructure

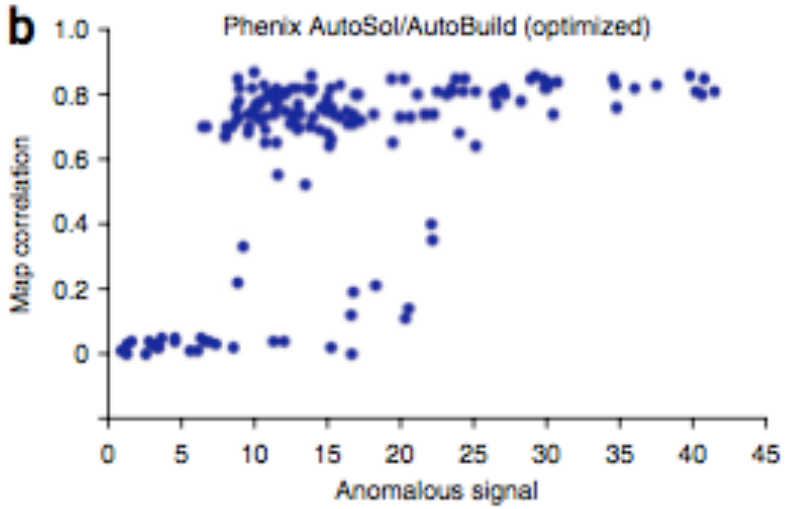





Find substructure
Complete with LLG maps
Density modification
Model building


Find substructure
Complete with LLG maps
Density modification
Model Building


a
Phenix AutoSol/AutoBuild
Map correlation
1.0
0.8
0.6
0.4
0.2
0
0 5 10 15 20 25 30 35 40 45
Anomalous signal


b
Phenix AutoSol/AutoBuild (optimized)
Map correlation
1.0
0.8
0.6
0.4
0.2
0
0 5 10 15 20 25 30 35 40 45
Anomalous signal

The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Youval Dar, Nat Echols, Nigel Moriarty, Nader Morshed, Ian Rees, Oleg Sobolev

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Cambridge University

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner, Richard Mifsud

Duke University

Jane & David Richardson, Chris Williams, Bryan Arendall, Bradley Hintze

An NIH/NIGMS funded Program Project