



phaser and phasing

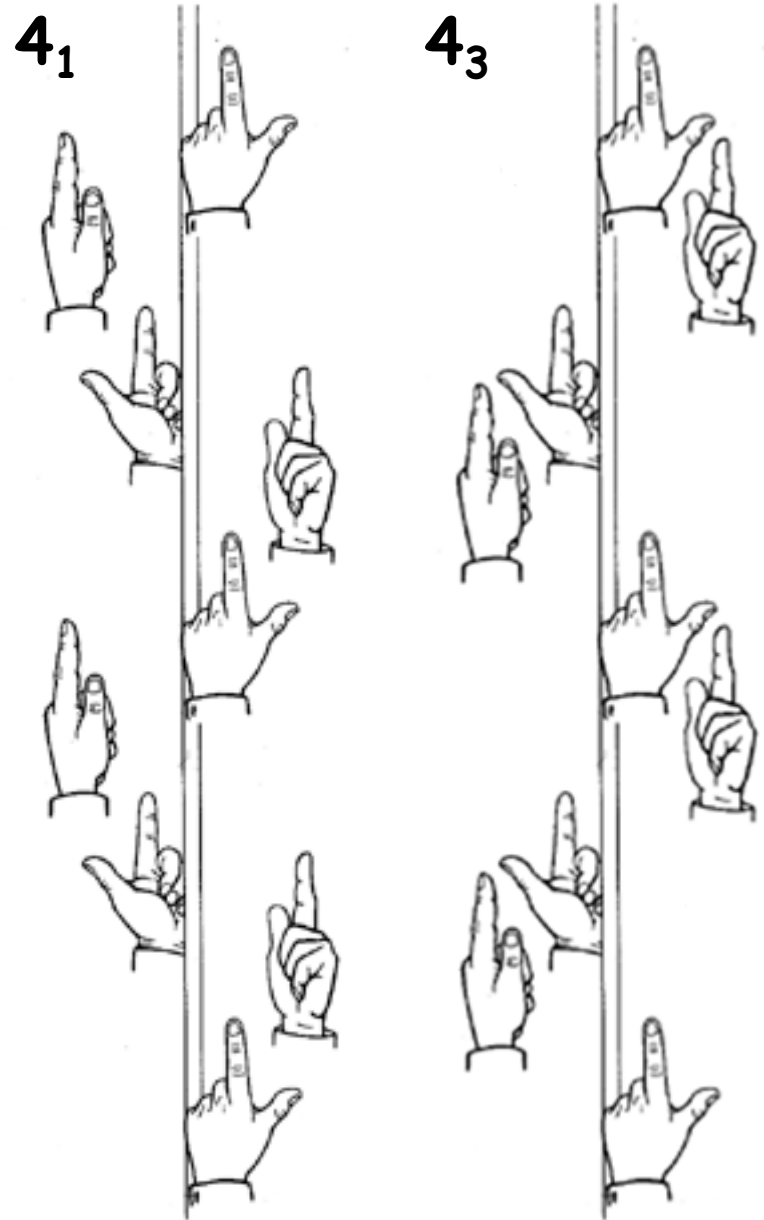
Airlie McCoy

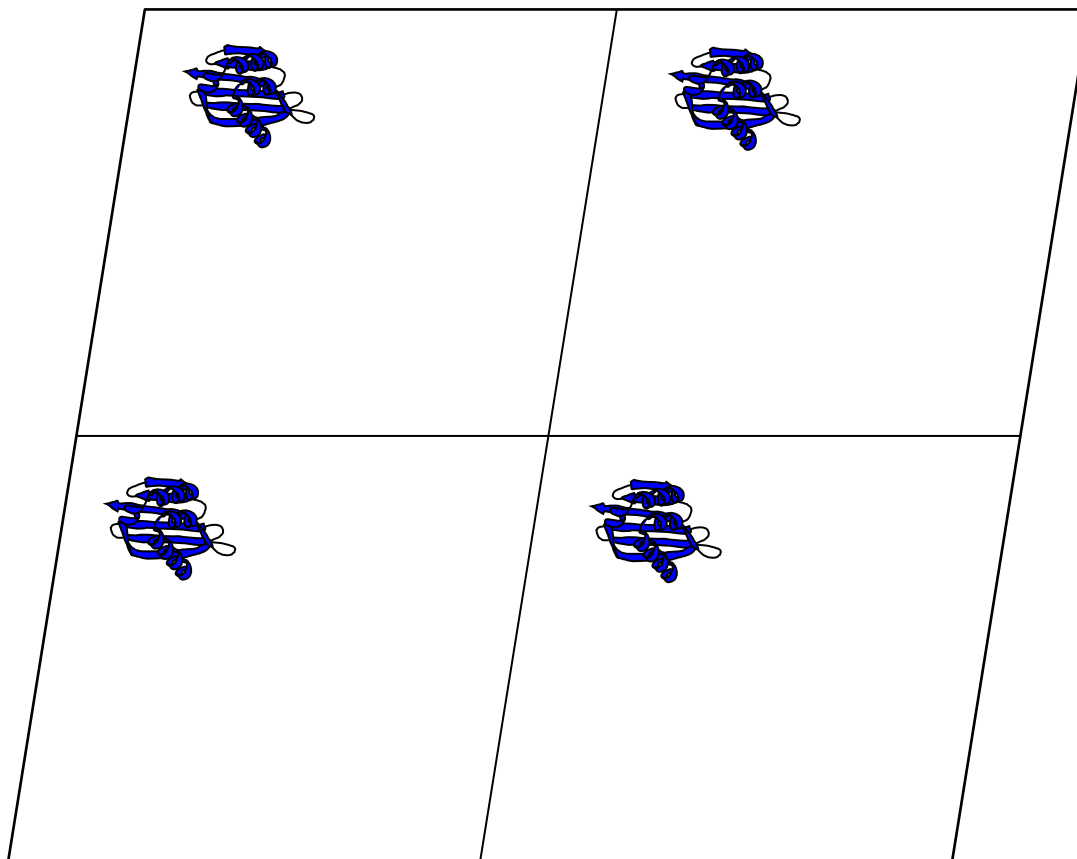


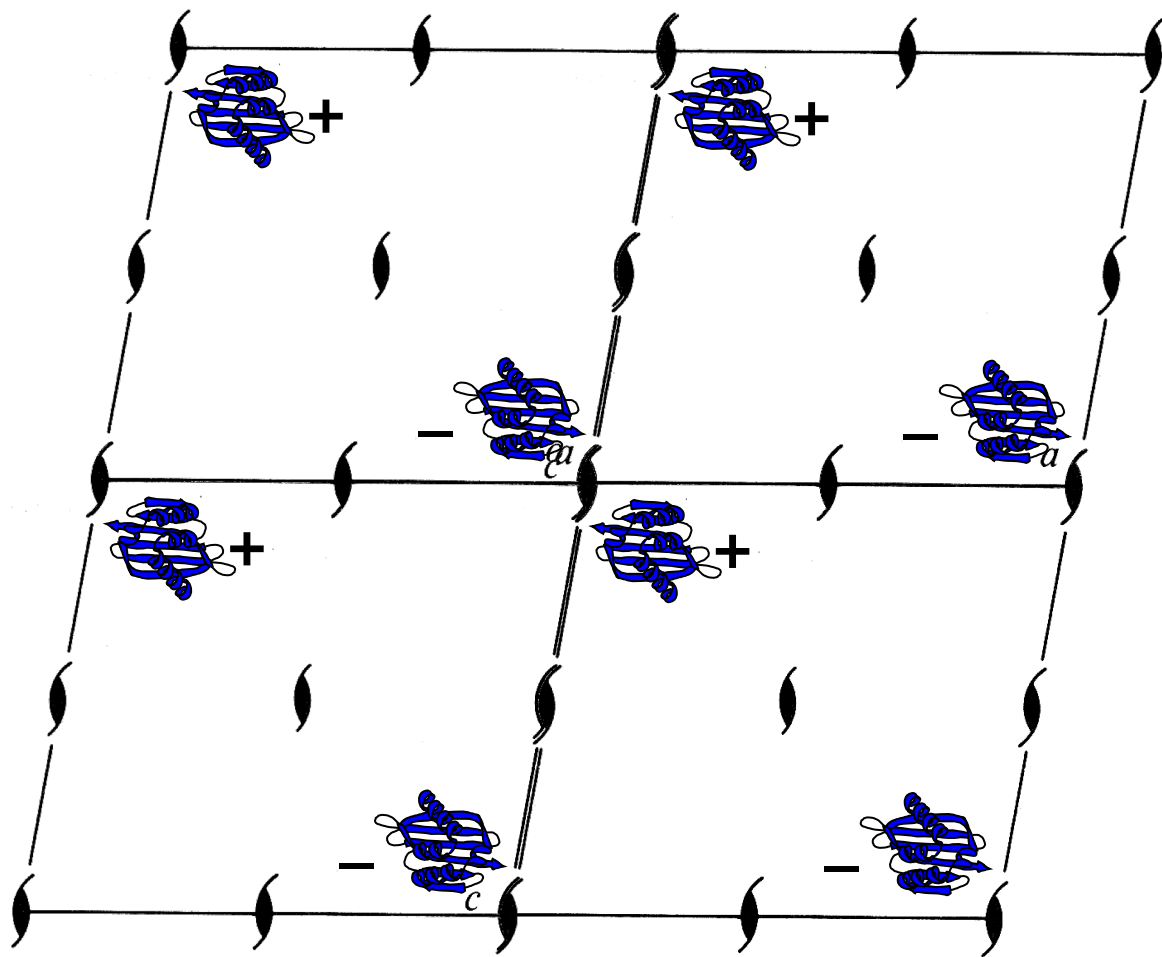
UNIVERSITY OF
CAMBRIDGE

Enantiomorphs

- Space group ambiguities
 - Hand
 - Missing data

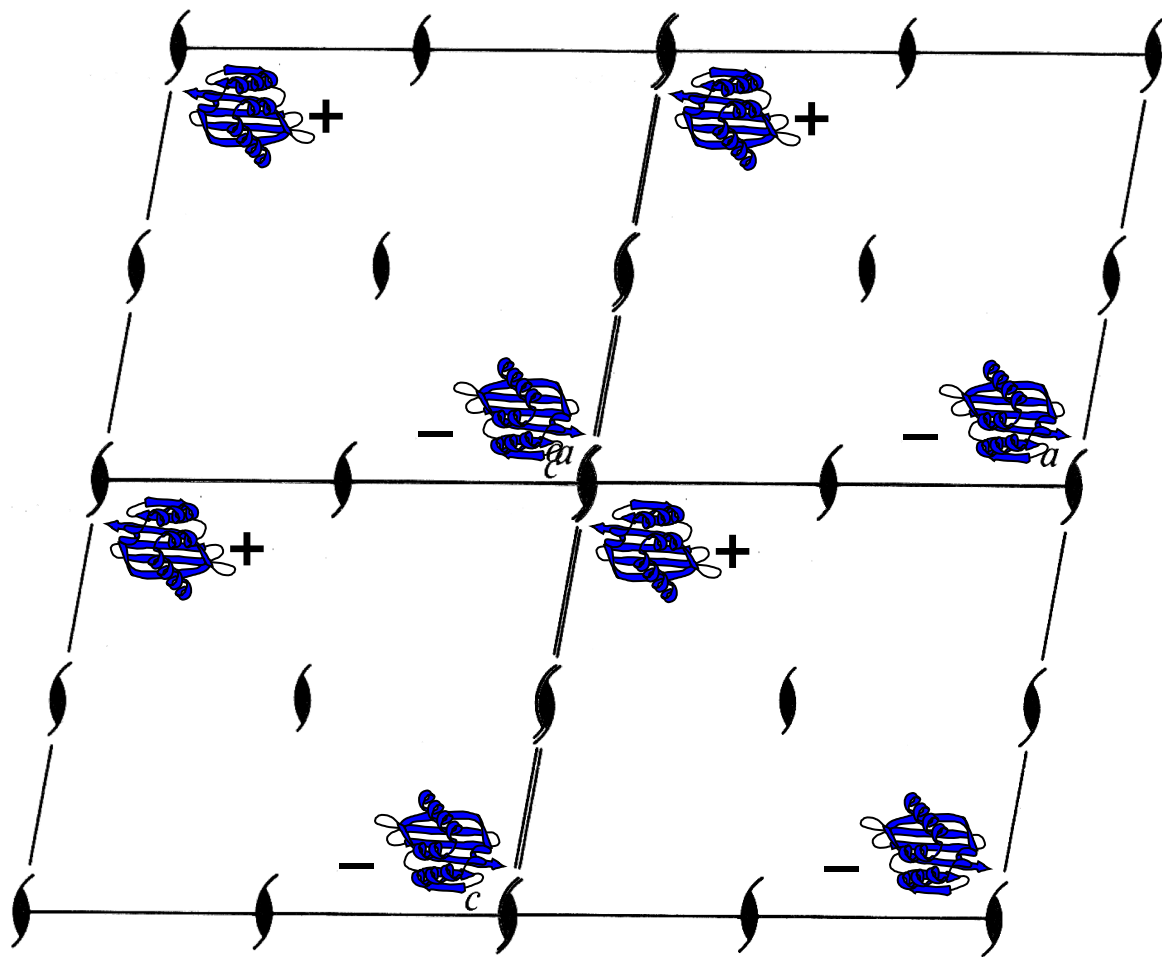


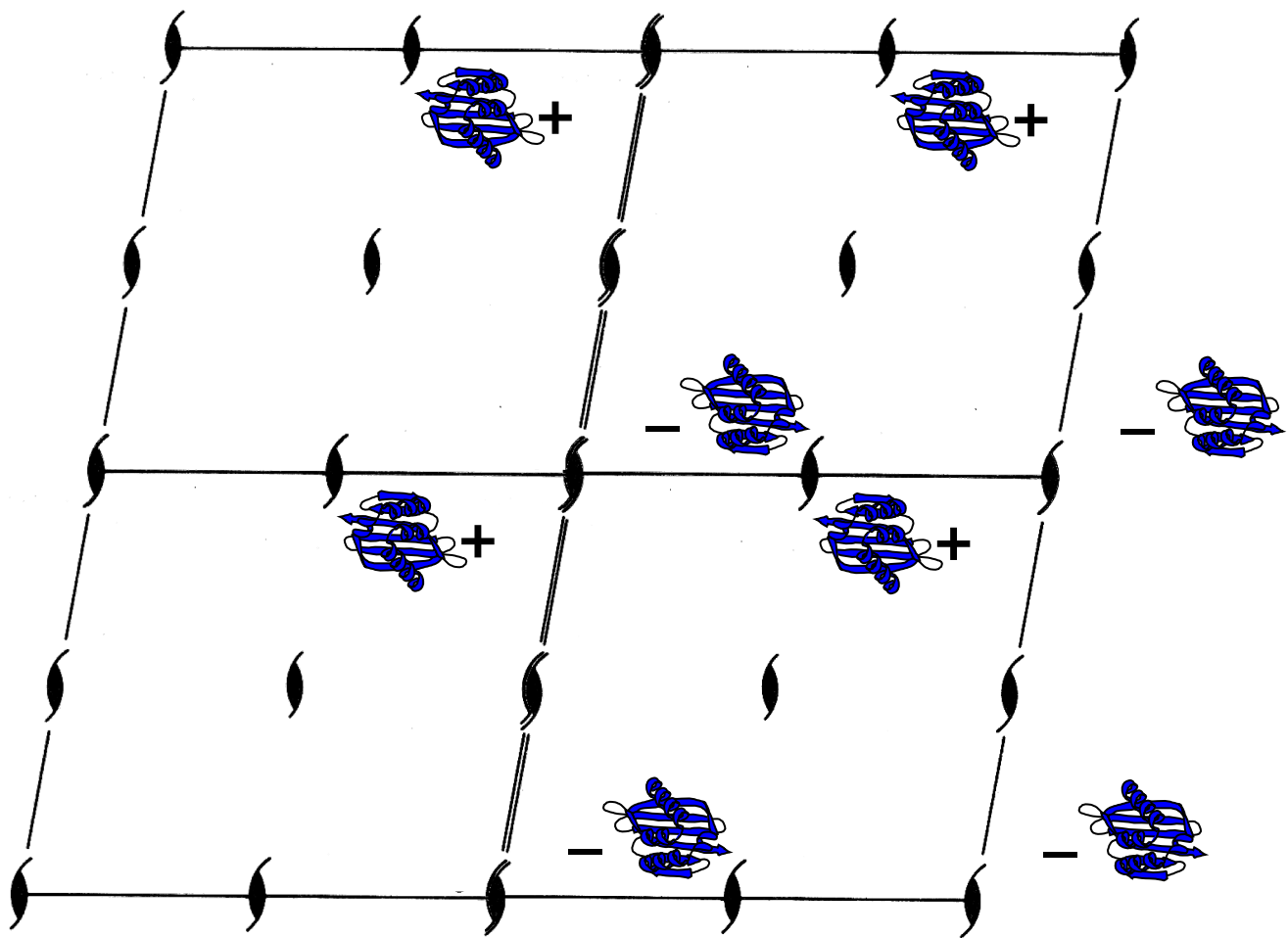


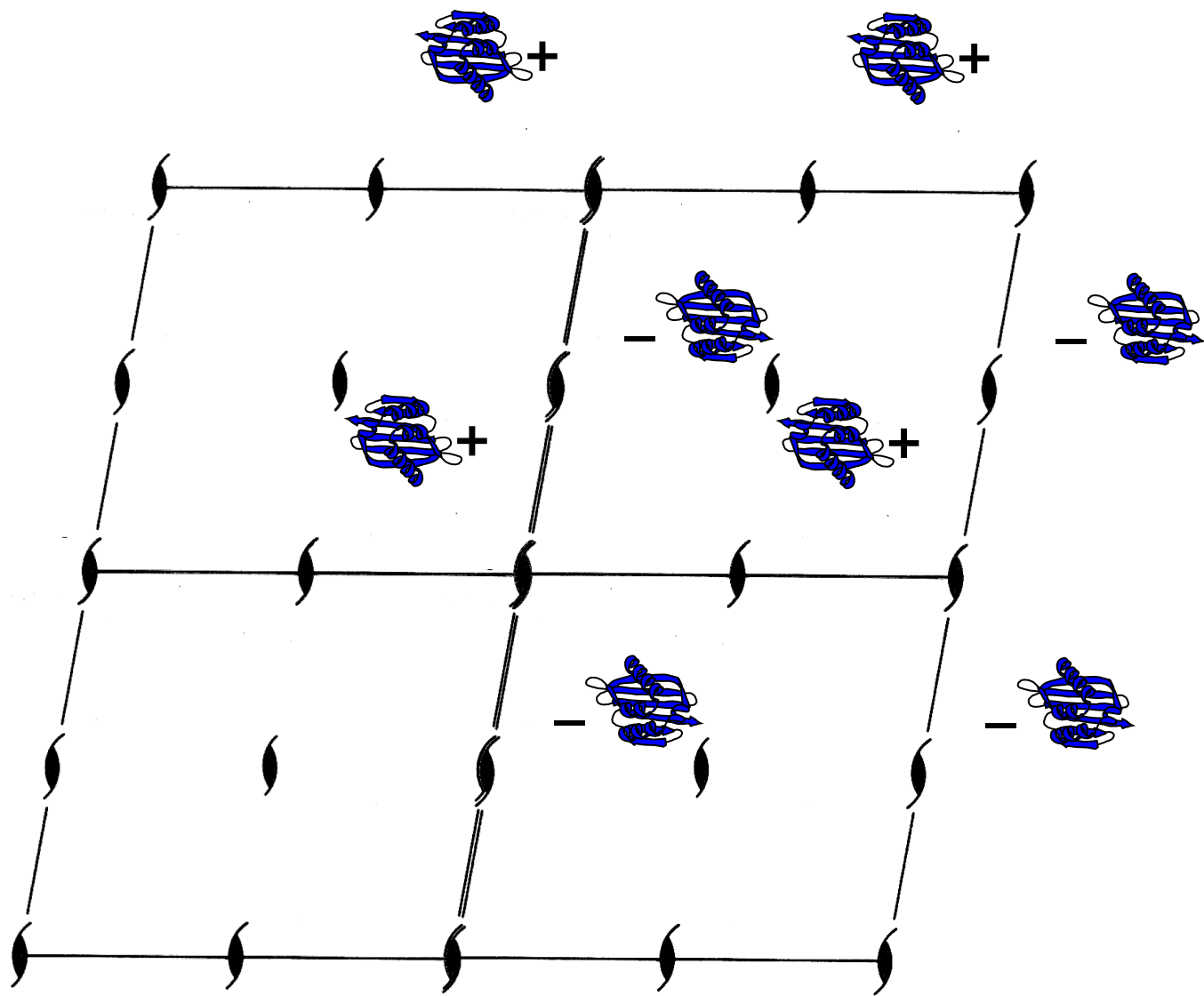


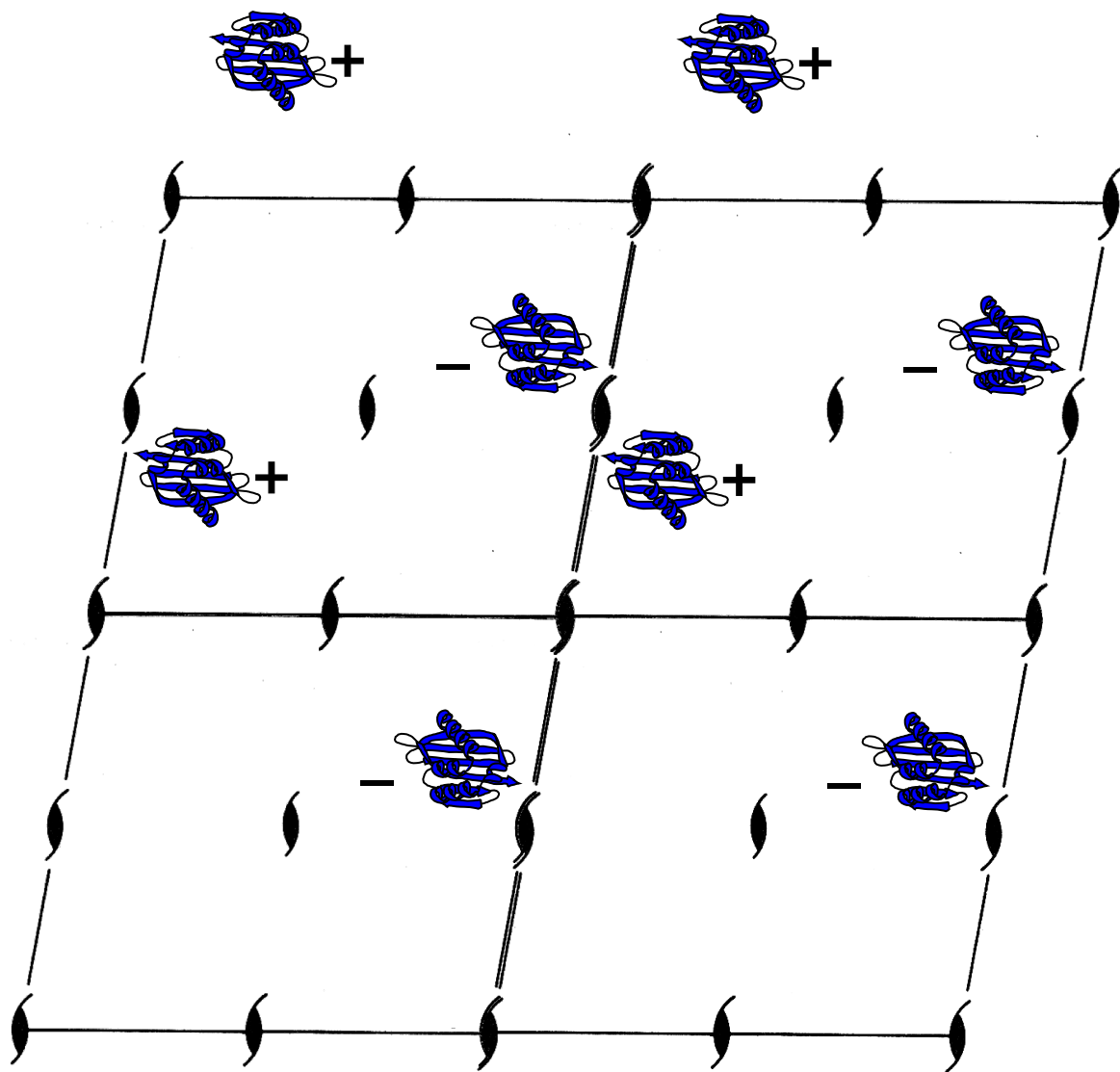
Origins

- Can only find the translation perpendicular to a rotation axis
 - no rotation symmetry, no translation to find!
- If there are multiple symmetry axes of the same order of rotation in a plane then the translation can be defined with respect to any one of these
 - These are equivalent to different choices of origin
 - Different MR solutions may be on “different origins” and look different when they are really the same

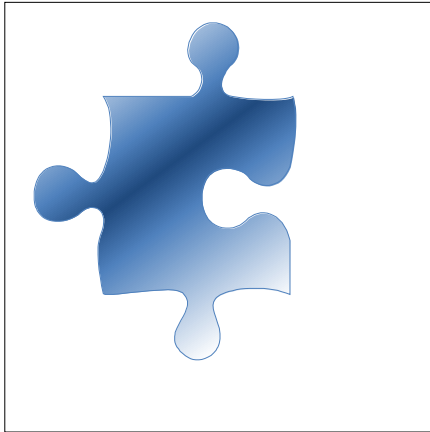






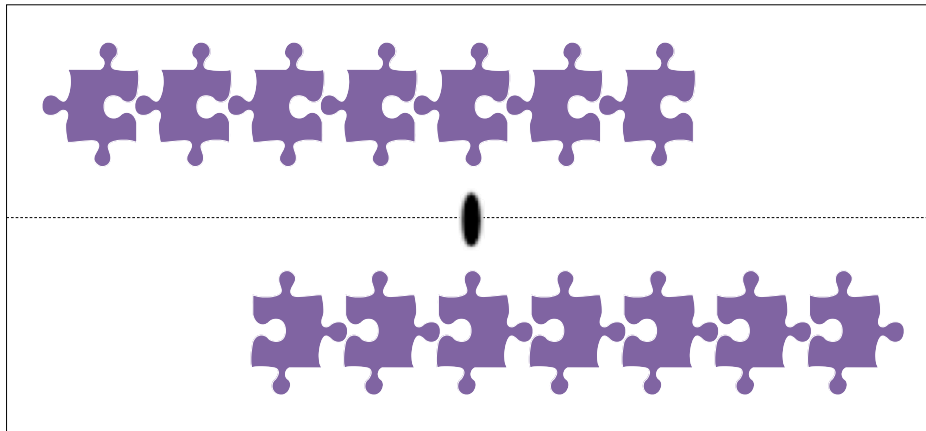


Pathologies



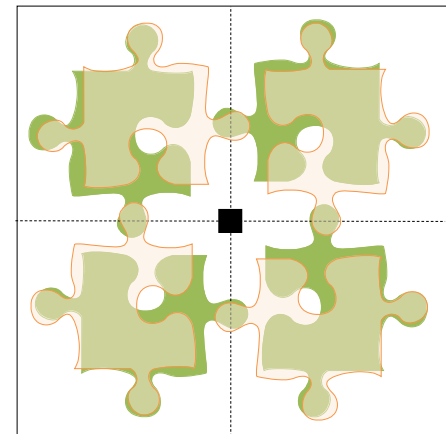
Anisotropy

Anisotropy is corrected by normalization terms refined against Wilson distribution



Translational non-crystallographic symmetry

Twinning



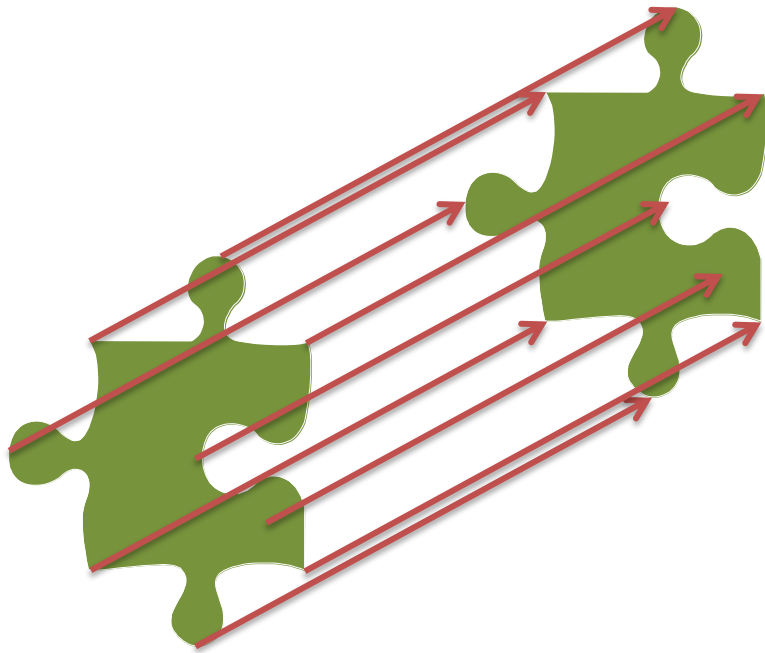
TNCS is corrected by expected intensity terms refined against Wilson distribution

Translational NCS

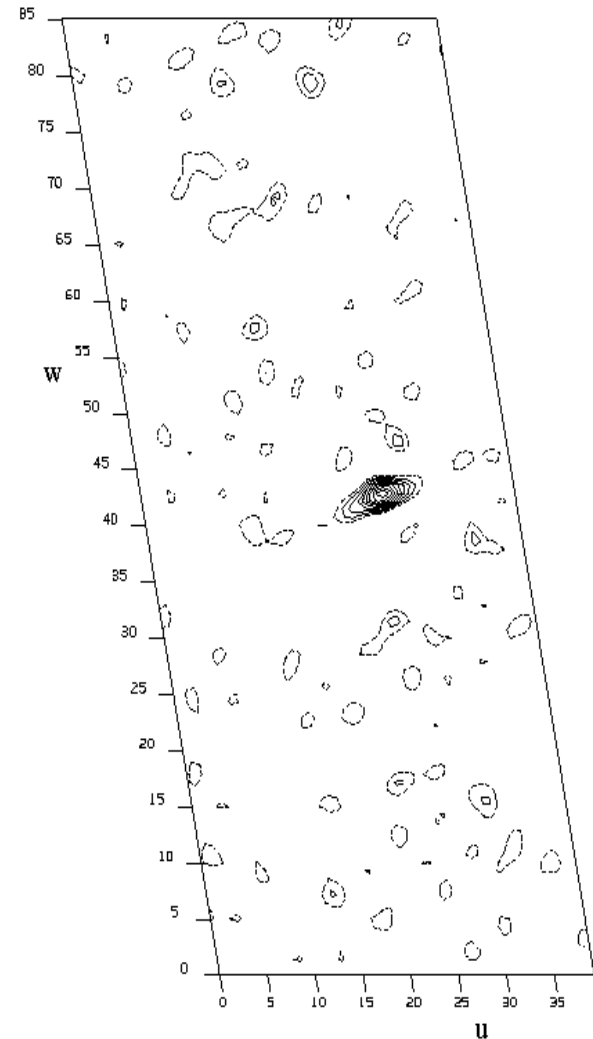
- If TNCS is not accounted for then $TFZ > 8$ does not indicate a correct placement
 - TFZ always higher
 - $TFZ > 12$ can be wrong
- i.e. Assumptions of ML violated
- When TFZ is accounted for the TFZ values are those expected of data without TNCS (?)

Translational NCS

- TNCS indicated by Patterson Peak
- 16% origin @ 5Å

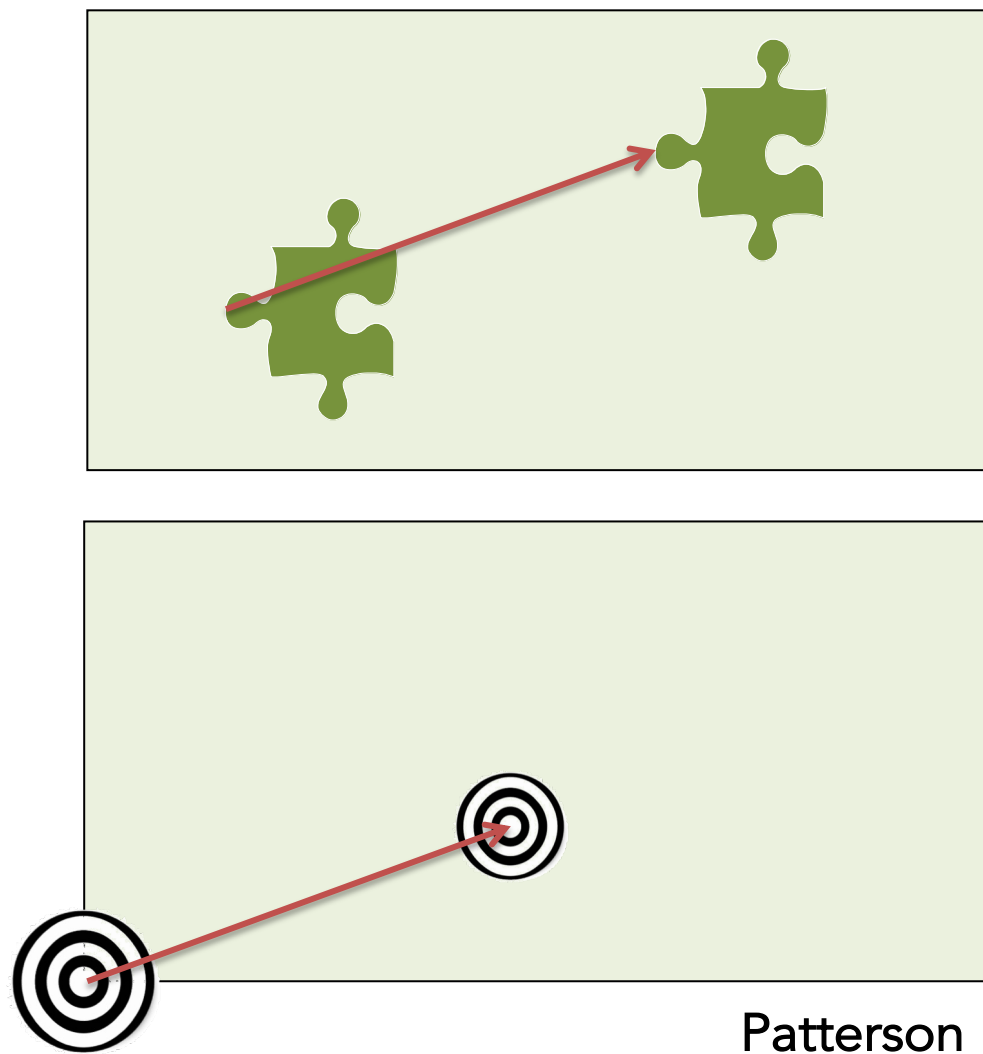


Native Patterson of mouse renin.



TNCS

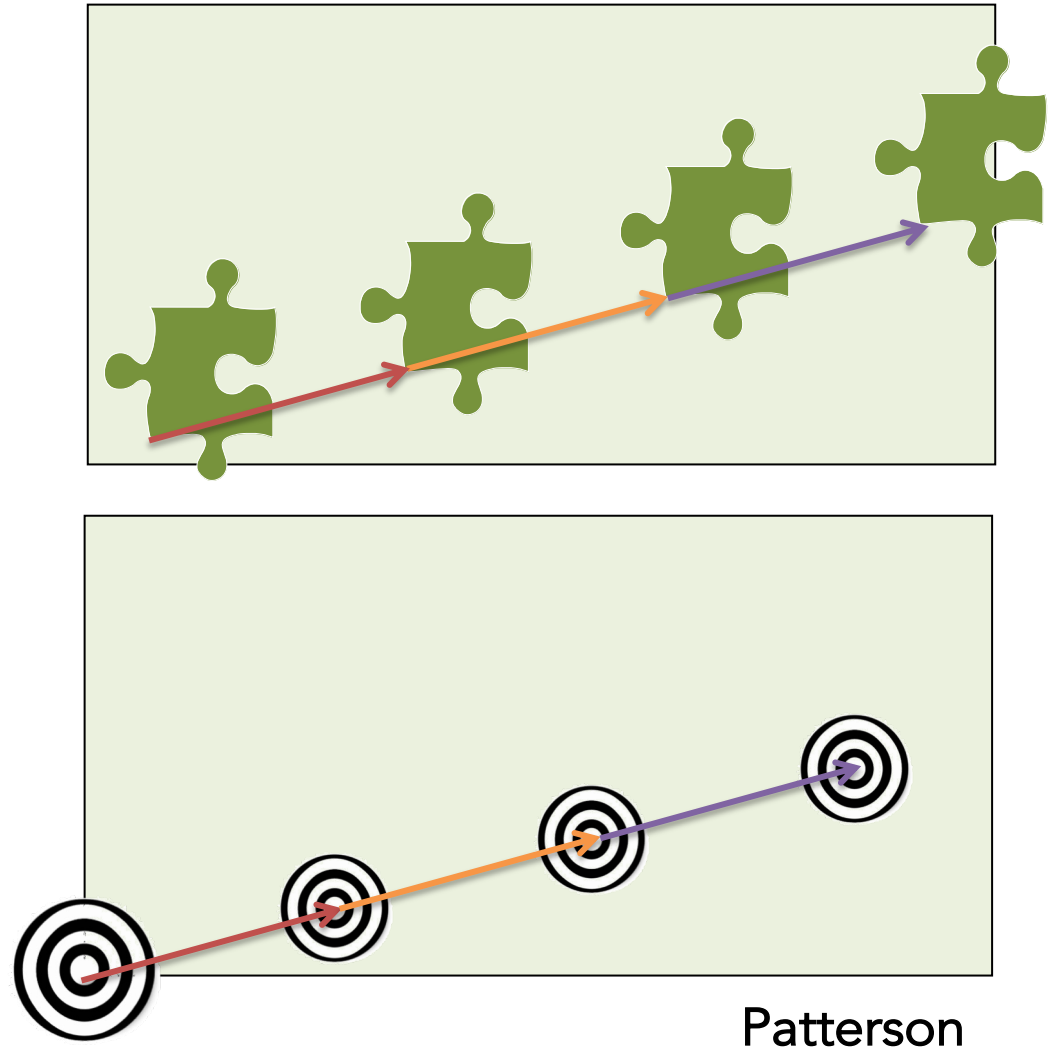
- Molecules related by a vector translation
- Patterson shows a single vector
 - symmetry copies



TNCS:

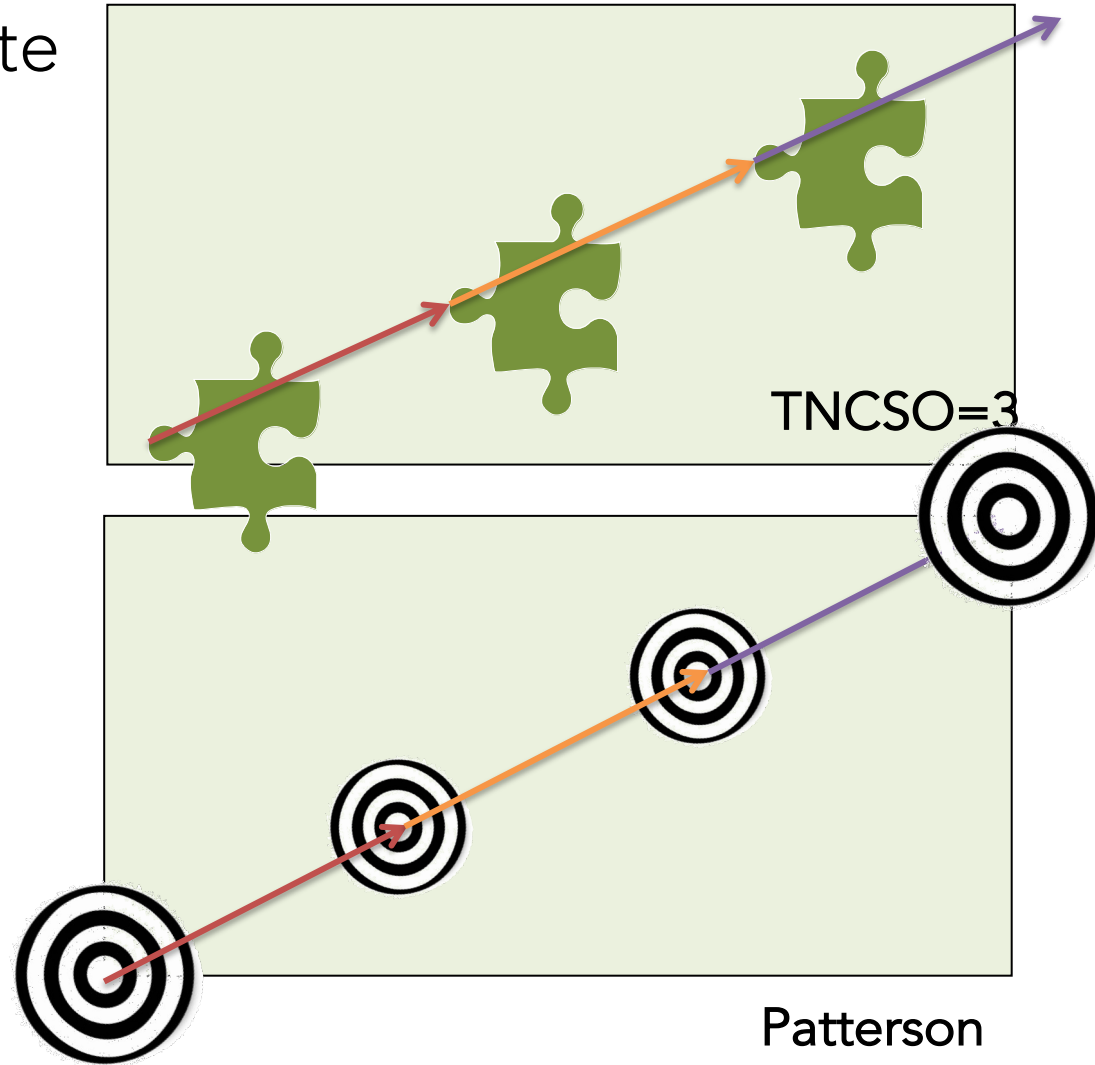
Order

- Molecules may be related by multiples of same vector
 - TNCSO
- Patterson shows a family of vectors
 - and symmetry copies



TNCS: Commensurate

- TNCS is often commensurate (?)
- Molecules at simple fractions of the cell
 - e.g. $(\frac{1}{3}, 0, 0)$



Two classes of tNCS

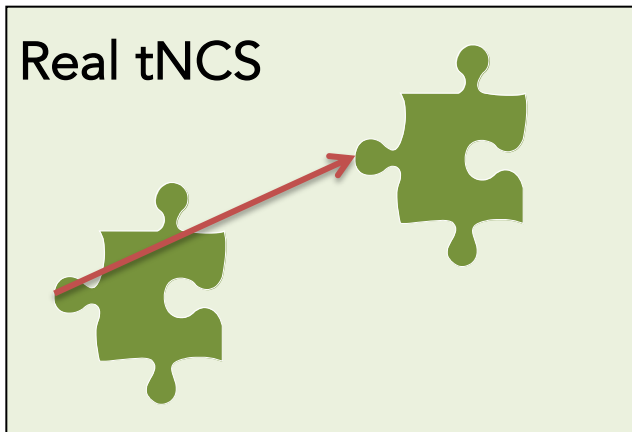
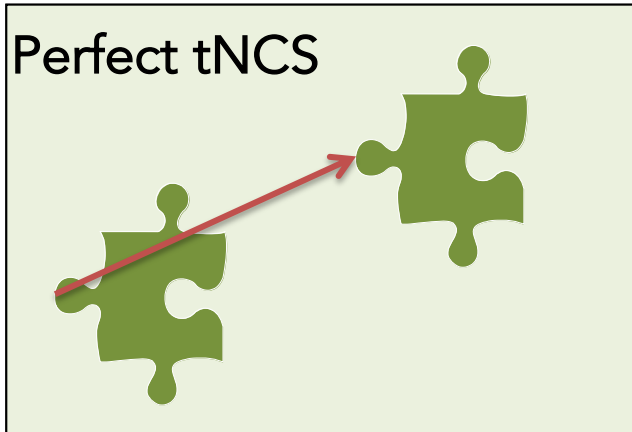
Pairs of molecules related by one vector

- One peak in Patterson
- May be commensurate or not commensurate
- There can be any number of pairs of molecules related by the same TNCS vector

Molecules related by multiples of one vector

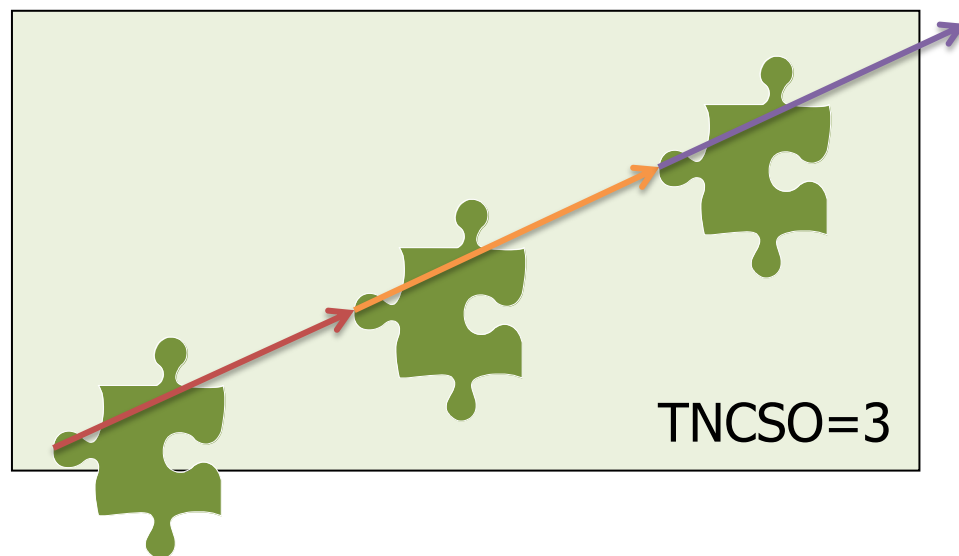
- Peaks in Patterson are multiples of same vector
- May be commensurate or not commensurate
- There can be any number of sets of molecules related by the same TNCS vector

Translational NCS: Pairs



- Three TNCS parameters refined from data alone
 1. Rotation
 2. Translation
 3. RMSD between copies

Translational NCS: $\text{TNC SO} > 2$



- Two TNCS parameters refined from data alone
 1. ~~Rotation~~
 2. Translation
 3. RMSD between copies

The rotational differences between the copies and the resulting intensity modulations become less significant, and the simplification is justified

Other classes of tNCS

These cover the majority of cases (?)

Sets of molecules related by different vectors in different directions

- Multiple peaks in Patterson
- May be commensurate or not commensurate

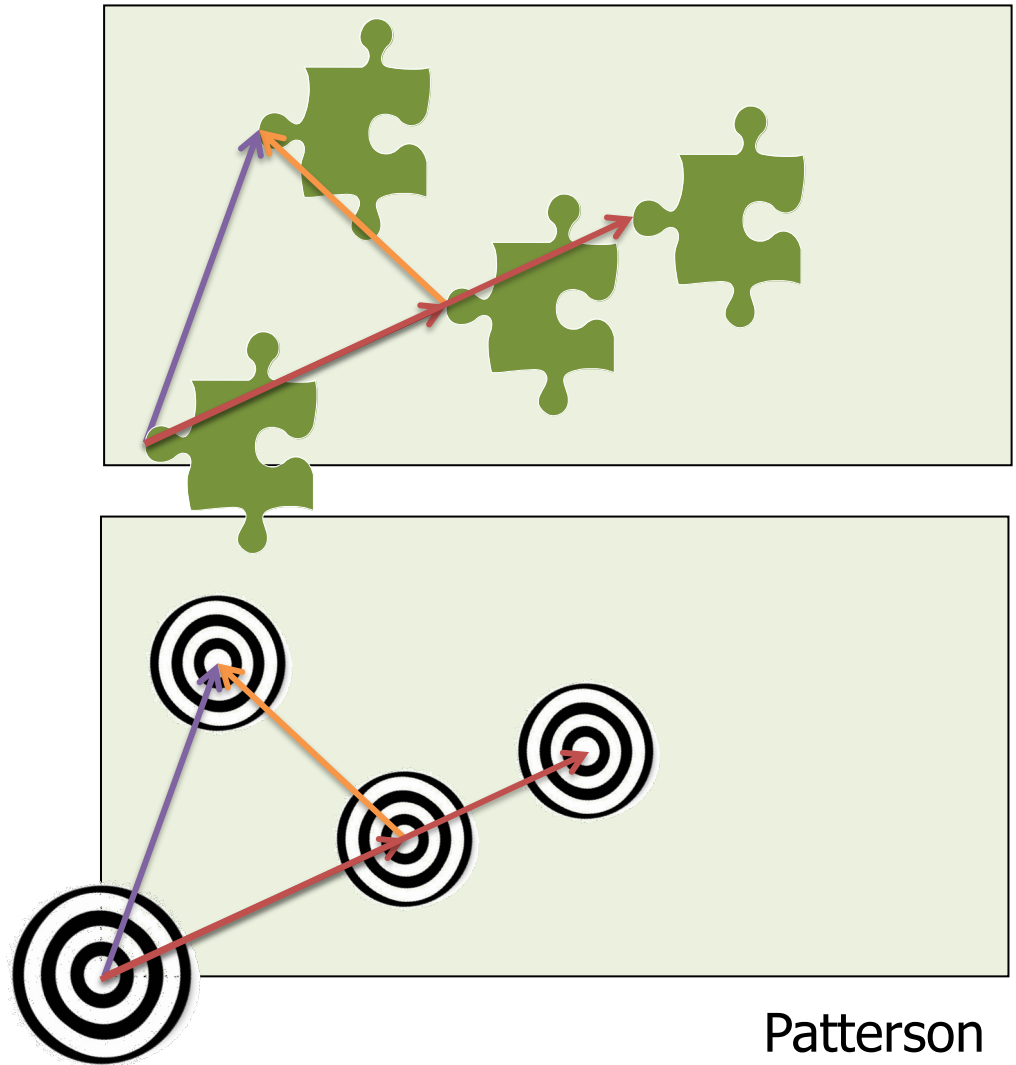
Sets of molecules related by vectors in same direction, different displacements

- Looks like commensurate modulation in Patterson
- Could be an incomplete commensurate modulation

TNCS:

Complex

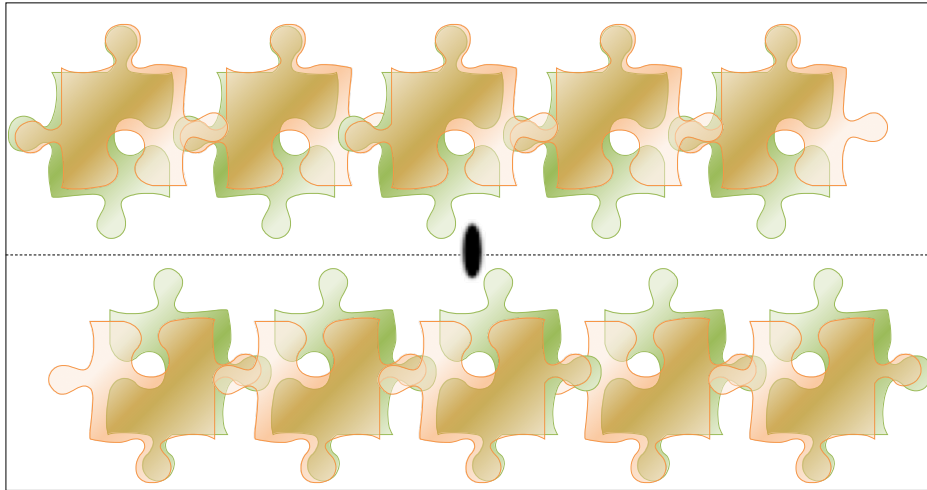
- Can be more complex than multiples of the same vector
- Halving in different cell directions
 - e.g. $(\frac{1}{2}, 0, 0)$, $(0, \frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2}, 0)$



Other classes of TNCS

- Break TNCS down into components that can be dealt with by the two classes
- When complex forms of TNCS are present assume the modulations will not be as strong as the modeled classes and so the exact correction terms will not be as critical in structure solution (?)

Pathologies



Anisotropy +
Translational NCS +
Twinning

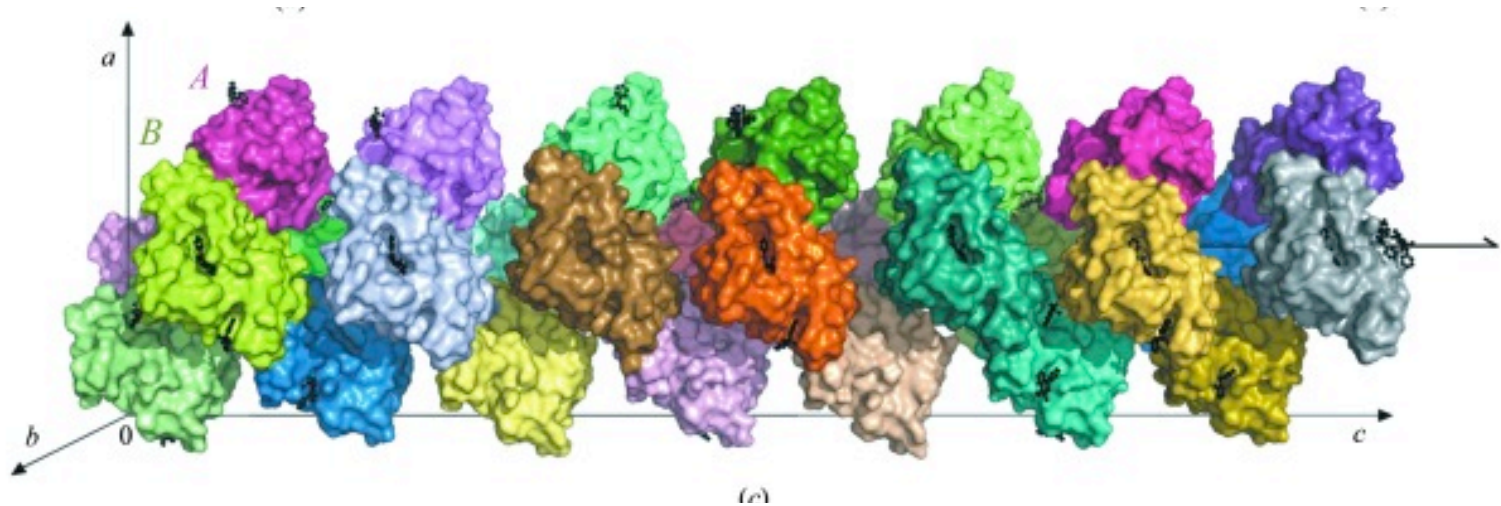
- Translational NCS masks twinning
 - Correcting for translational NCS unmask twinning
- Phaser gives a P-value for there being twinning in the presence of translational NCS

Twin Detection

- Translational NCS masks twinning
 - Has opposite effect on intensity statistics
- Correcting the data for tNCS unmasks twinning
- Phaser generates cumulative intensity plots for centric and acentric reflections after correction for tNCS and anisotropy
- Phaser gives a P-value for there being twinning in the presence of tNCS

Hyp-1 (PR-10)

- Repeats $7/2$ along c^* (sevenfold translational NCS)
- Tetartohedrally twinned
- Solved finding 56 copies of the monomer in P1
 - 28 copies in C2 (true space group)



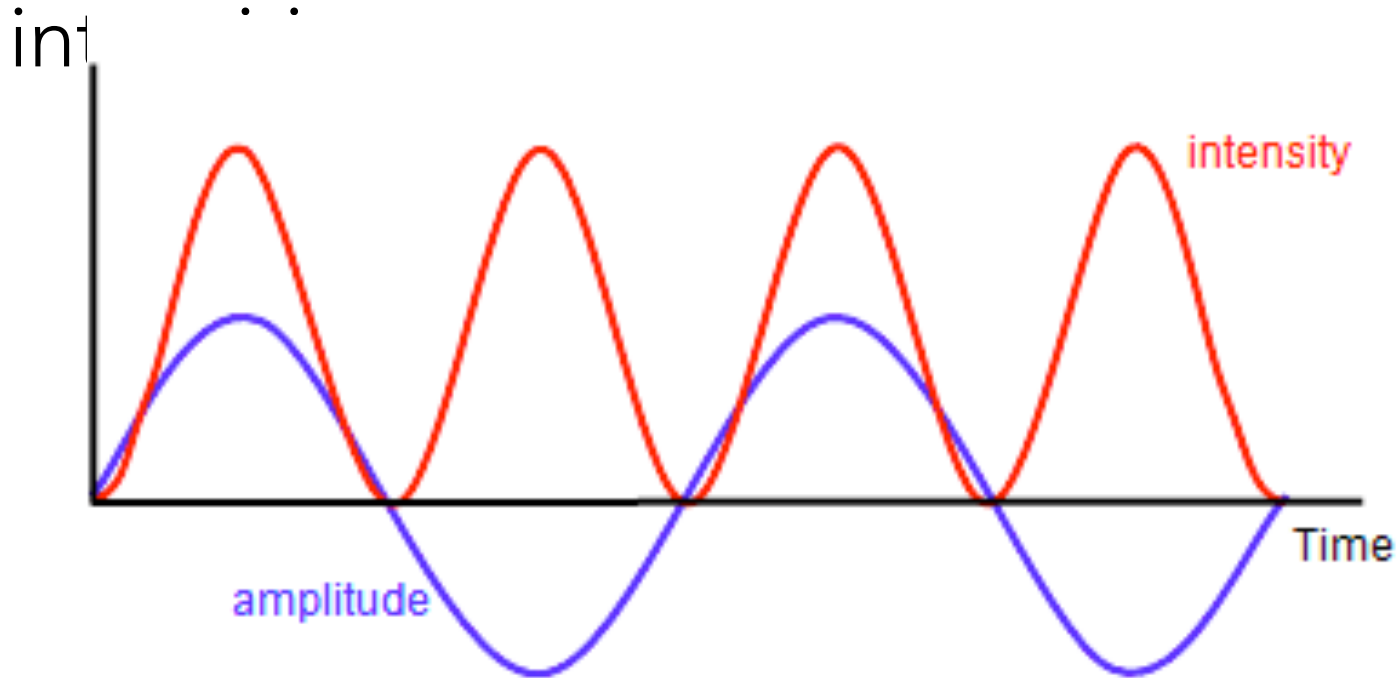
Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold tNCS Sliwiak J, Jaskolski M, Dauter Z, McCoy AJ, Read RJ. Acta Cryst D70, 471 (2014)

SAD Phasing

Airlie McCoy

Interference

- X-ray detectors only detect X-ray

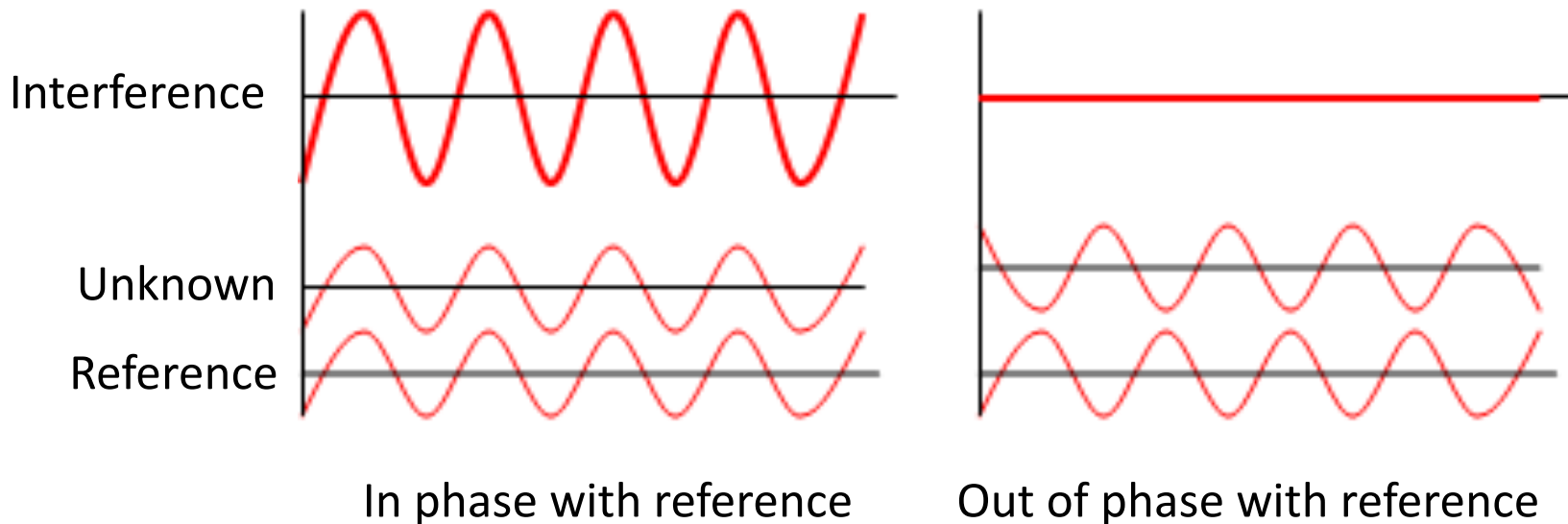


$$E \propto A^2$$

*because they respond to
the energy of the wave*

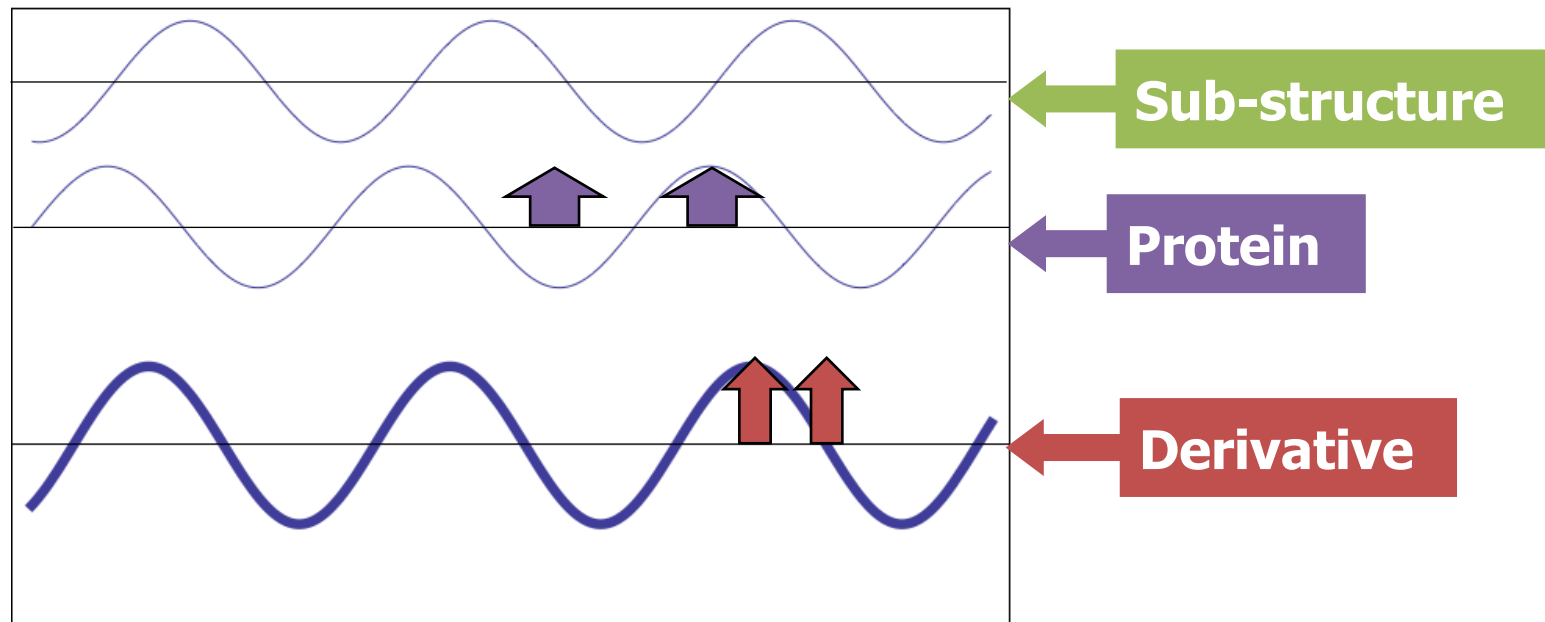
Reference Wave

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

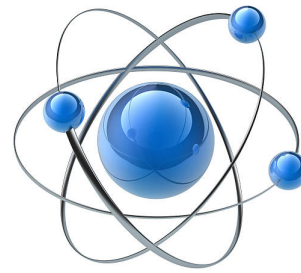


Interference and Phase

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



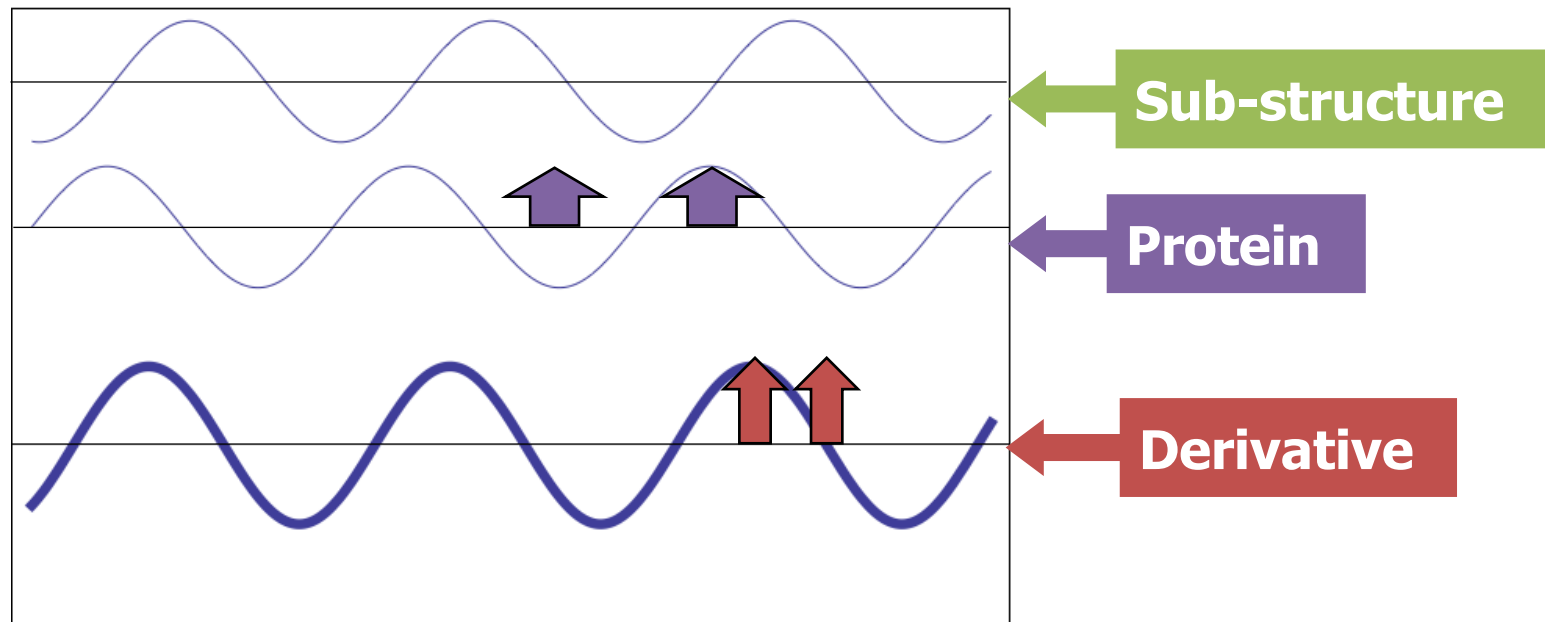
Phassade



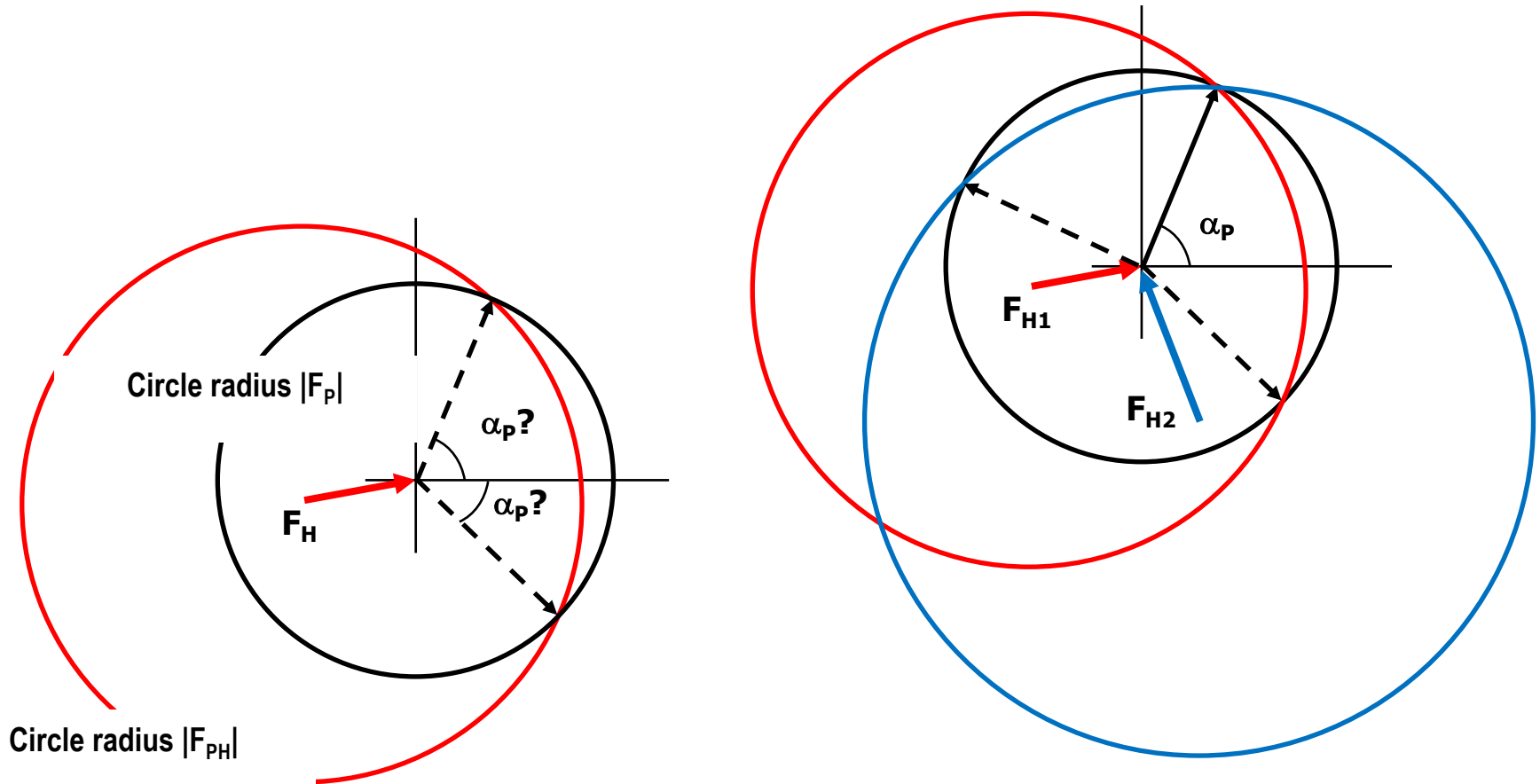
- Single Atom MR has inspired a new way to find the SAD substructure
- Phassade is an FFT-substructure search
 - based on the SAD likelihood target
 - Very fast
- Phassade avoids any requirement to estimate F_A
 - automatically takes account of the effects of both real and imaginary scattering in the substructure
- Particularly good for substructures in which the
 - signal comes from mainly from real scattering
 - anomalous scattering contributions are very large

Interference and Phase

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

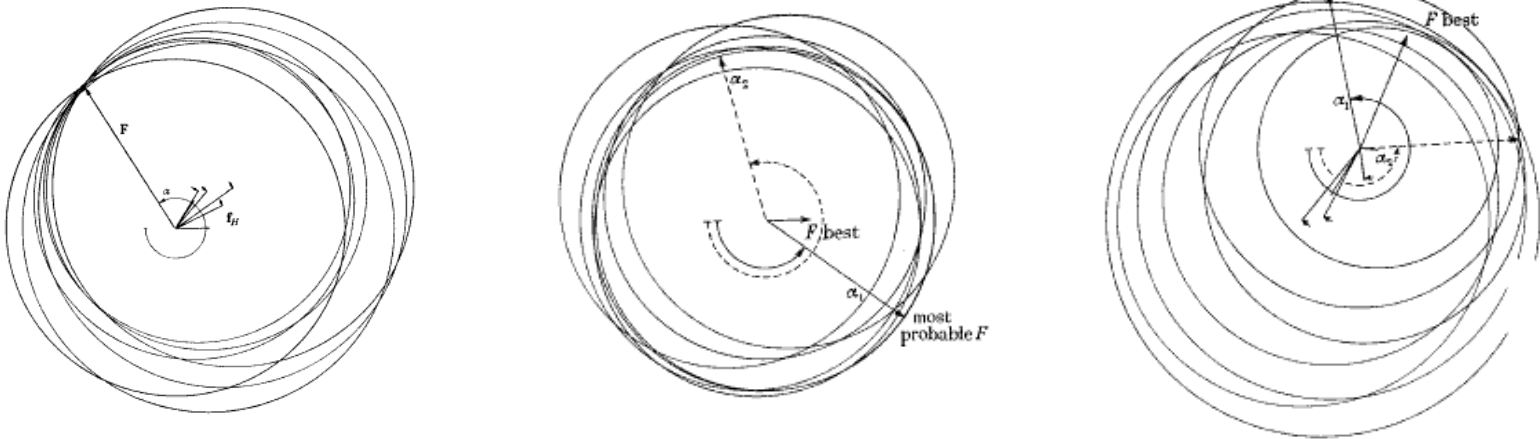


Harker Diagrams



Reality...

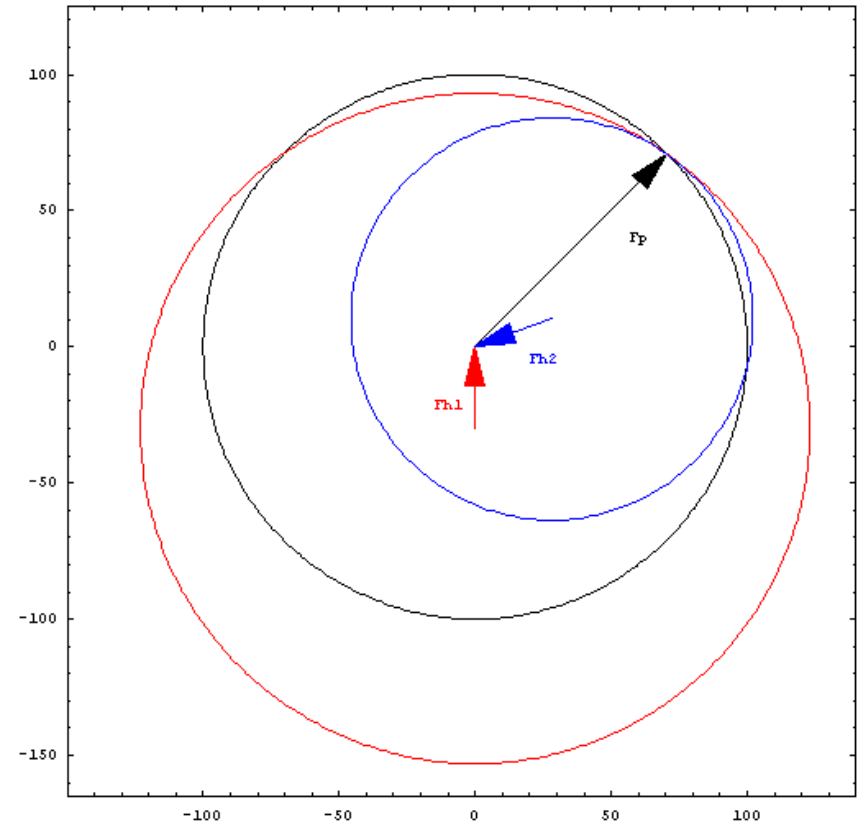
- Some real Harker diagrams from the phasing of haemoglobin with 6 derivatives



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

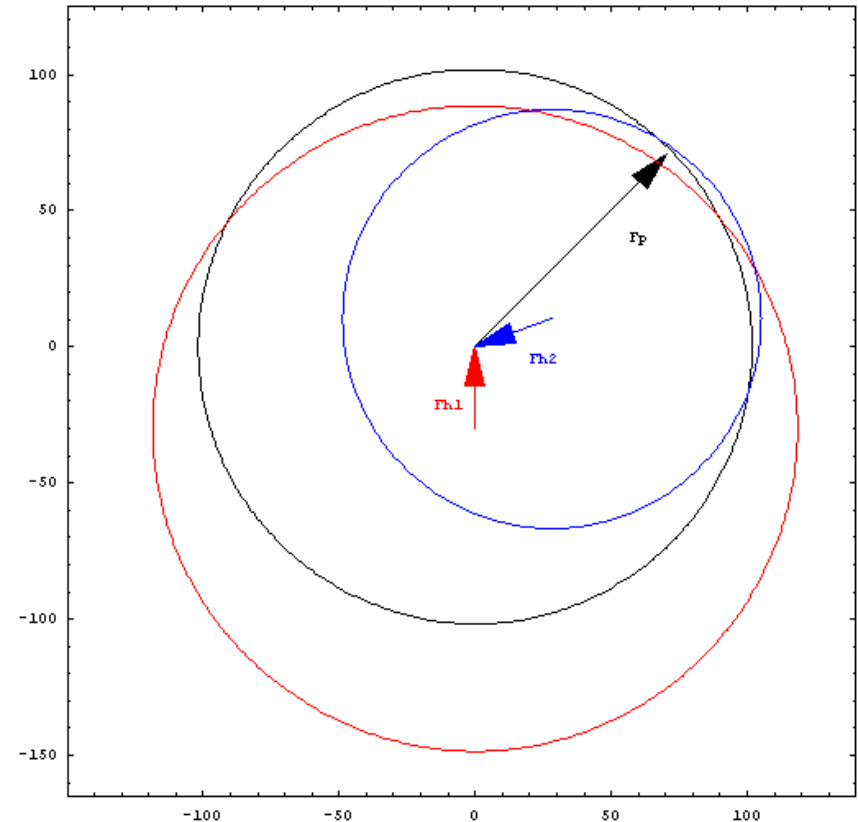
Harker construction

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



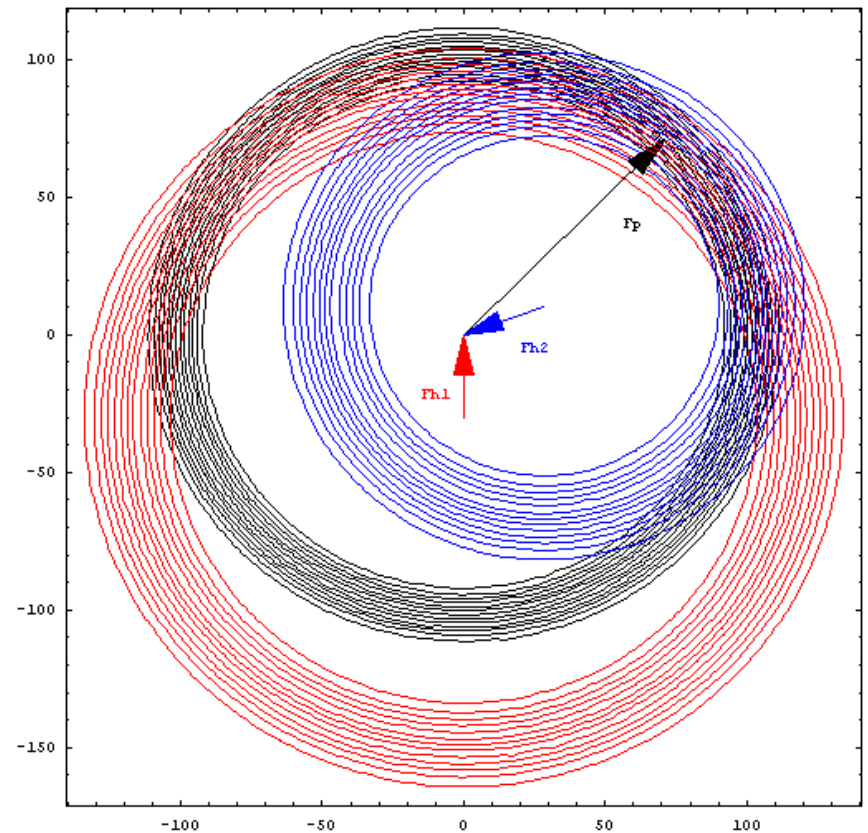
Harker construction

- 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



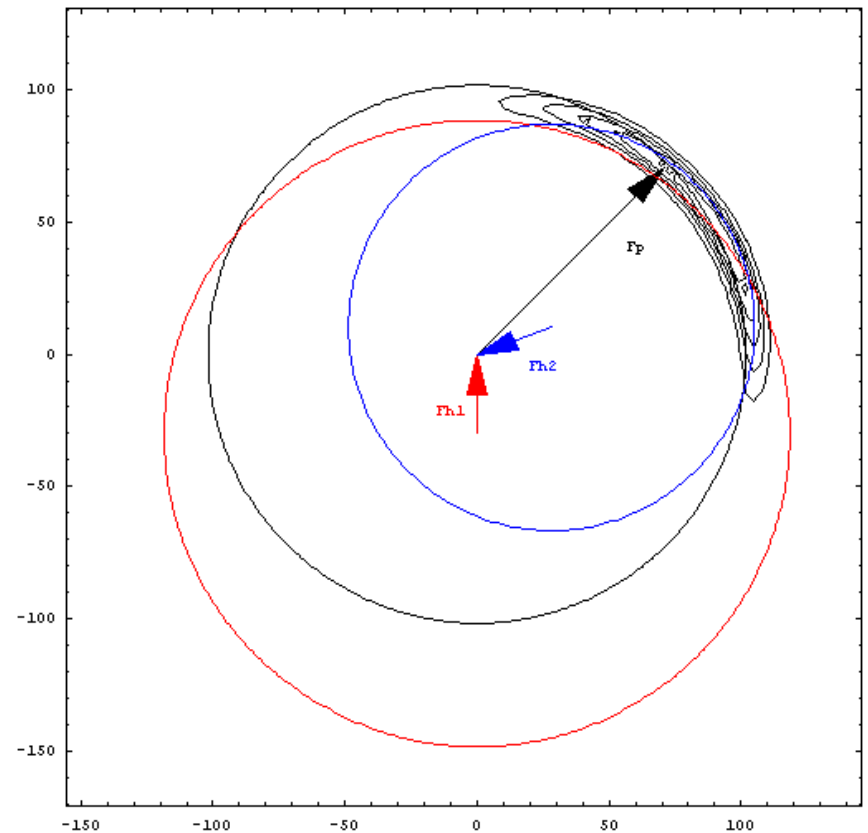
Probabilistic Harker Diagram

- 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



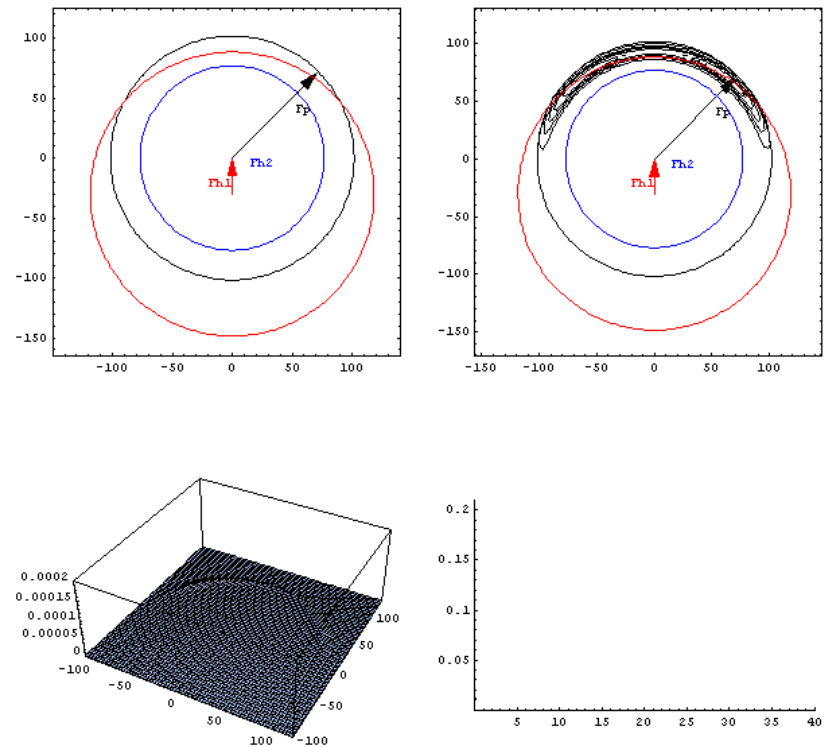
Probabilistic Harker Diagram

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



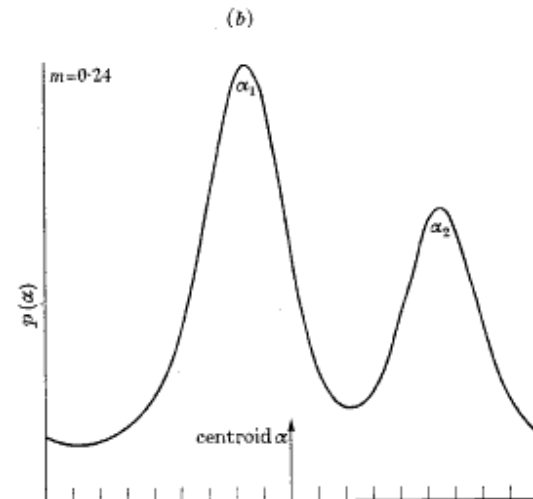
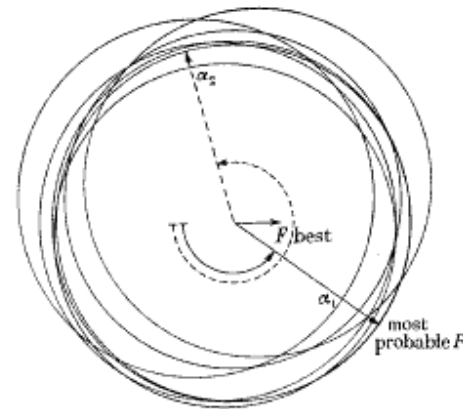
Refining Occupancy

- 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

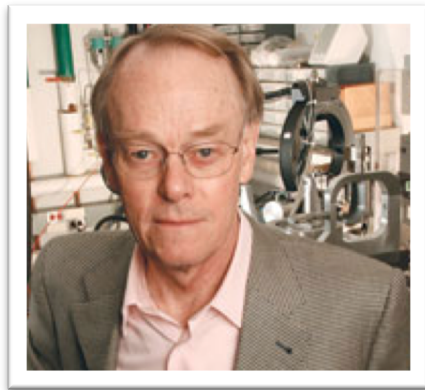
- Each reflection has a phase probability density function (PDF) rather than a single phase
- This is a complicated mathematical function
 - Requires lots of memory
- Hendrickson-Lattman coefficients are used to store the PDF in a compact form



Hendrickson-Lattman Coefficients

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

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



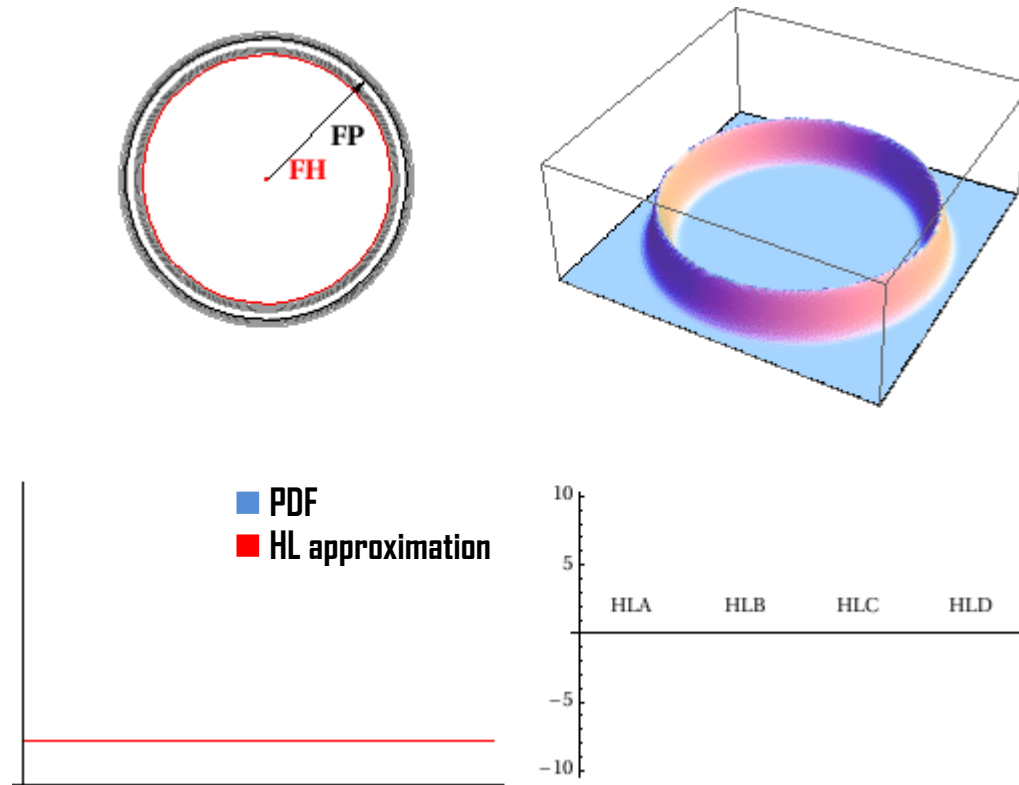
Wayne Hendrickson



Eaton Lattman

Hendrickson-Lattman Coefficients

HL coefficients as a function of F_H occupancy



A purple starburst shape with a white shadow.

MIR

An orange starburst shape with a white shadow.

SAD

A red starburst shape with a white shadow.

SIR

A green starburst shape with a white shadow.

SIRAS

A pink starburst shape with a white shadow.

MIRAS

A blue starburst shape with a white shadow.

MAD

A dark blue starburst shape with a white shadow.

RIP

SAD

- The most popular way of solving structures by experimental phasing (over 70%)
- Can be done with intrinsic S and CuK α X-rays
- SAD phasing theory is 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

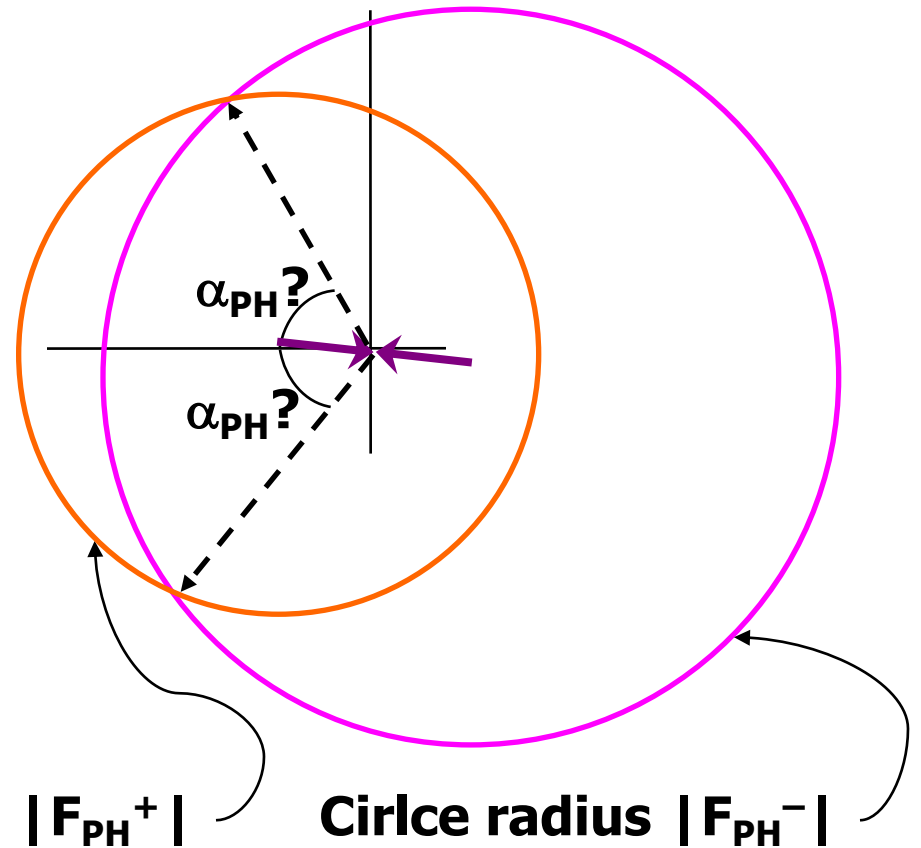
"Anomalous scattering" is **absolutely normal** while "normal scattering" occurs only as an ideal, over simplified model, which can be used as a first approximation when studying scattering problems"

IUCR Pamphlet "Anomalous Dispersion of X-rays in Crystallography"

S. Caticha-Ellis (1998)

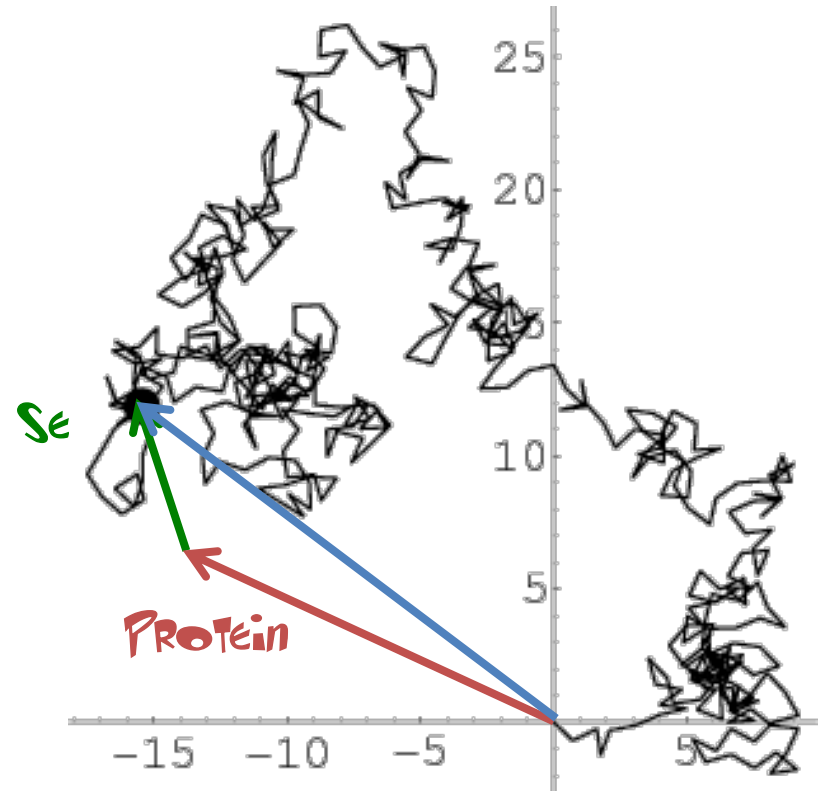
SAD Harker Diagram

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



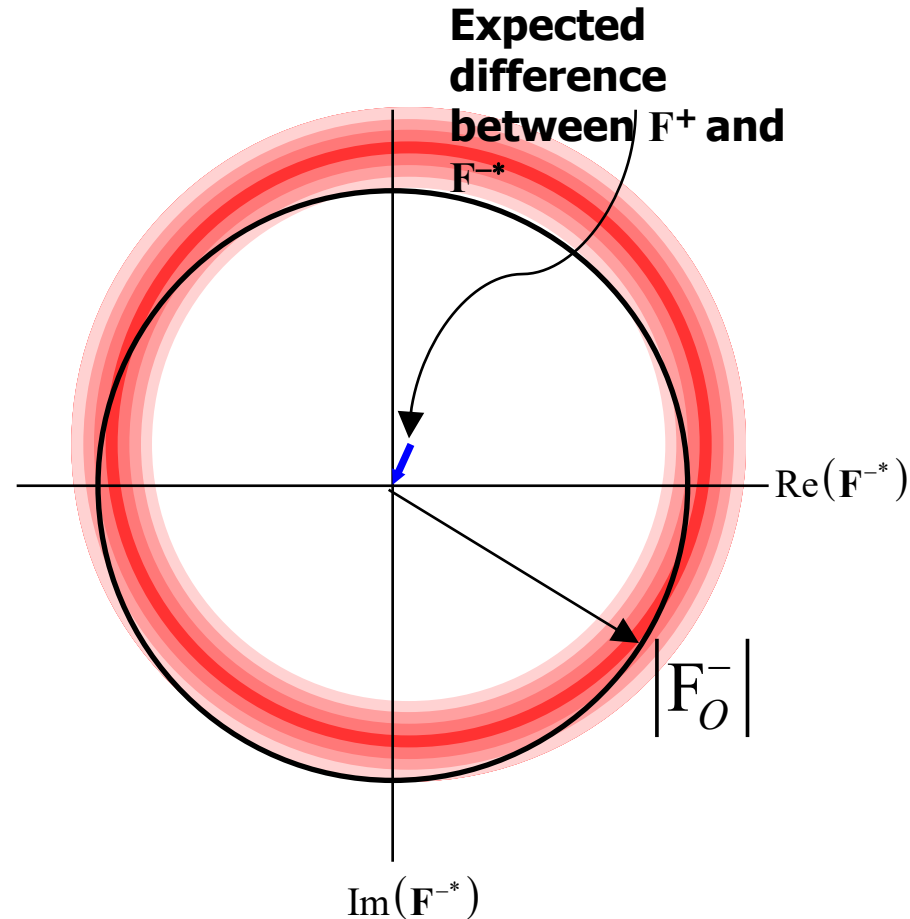
Partial Structure

- 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



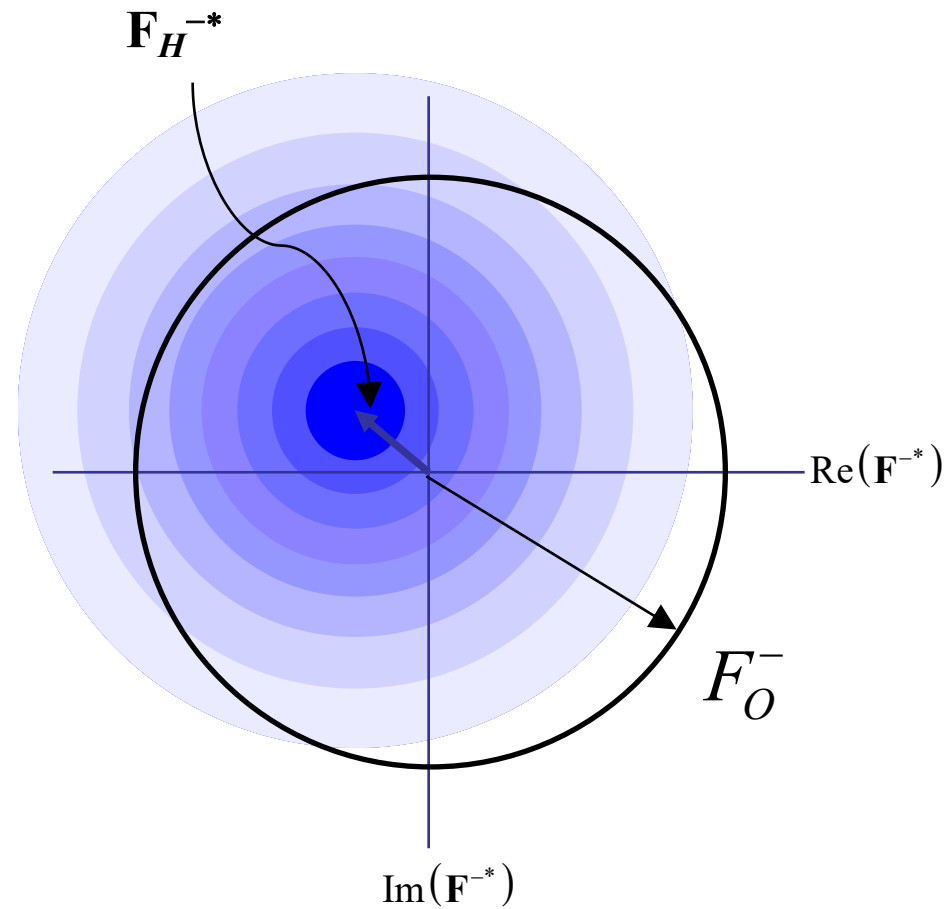
SAD Phasing

- Primarily anomalous scattering
- Gives a narrow probability distribution



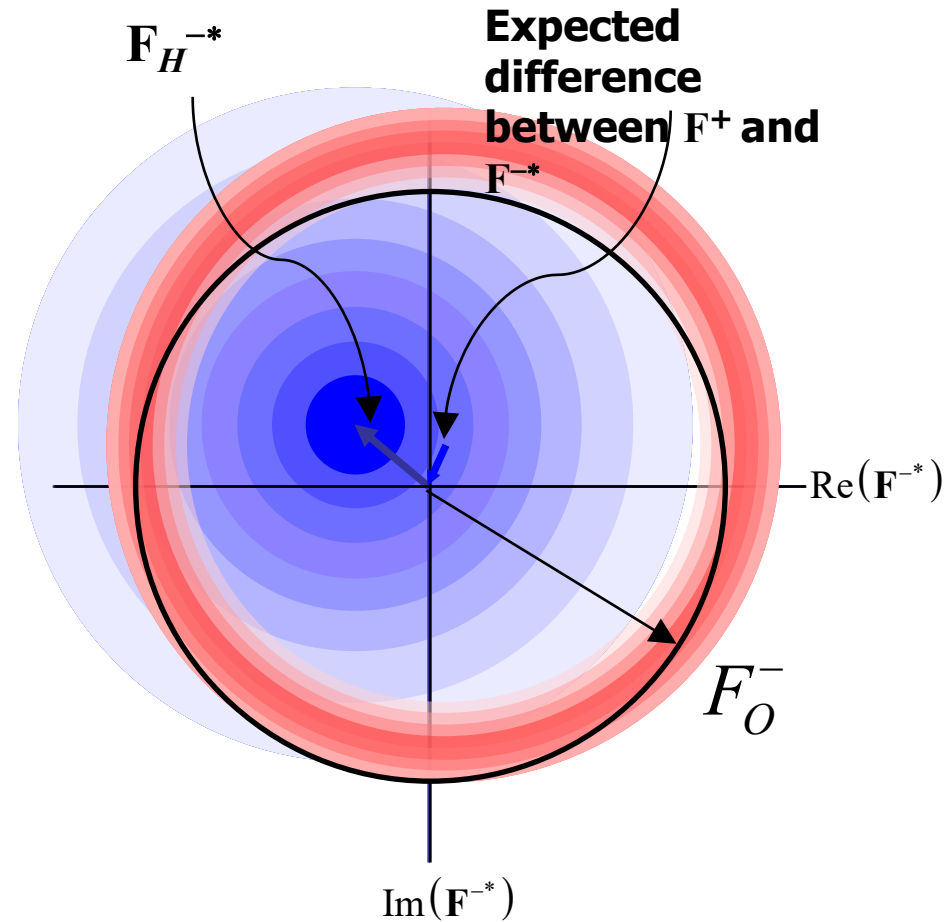
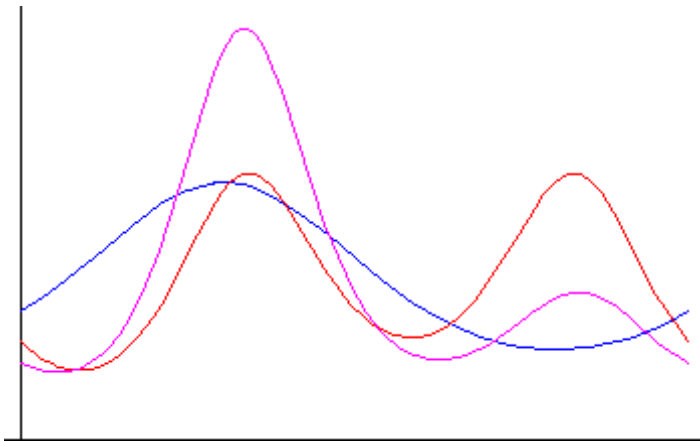
SAD Phasing

- Primarily Normal scattering
- Gives a broad probability distribution



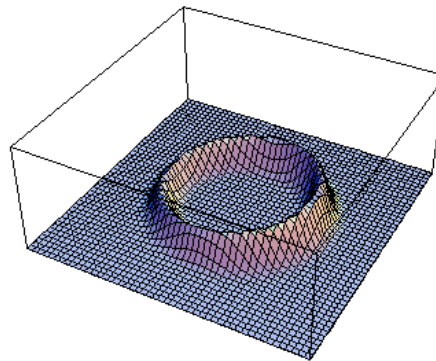
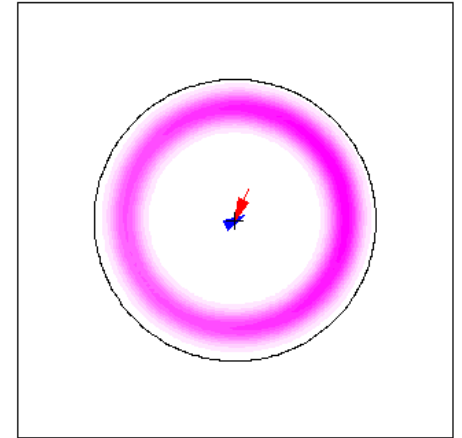
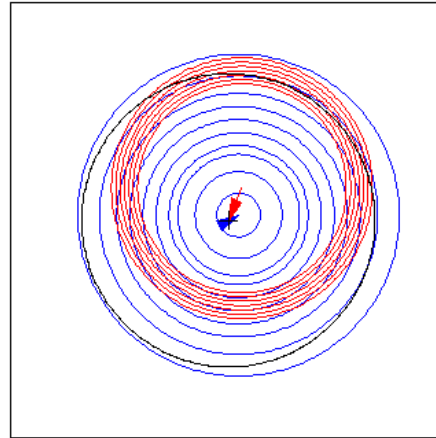
SAD Phasing

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



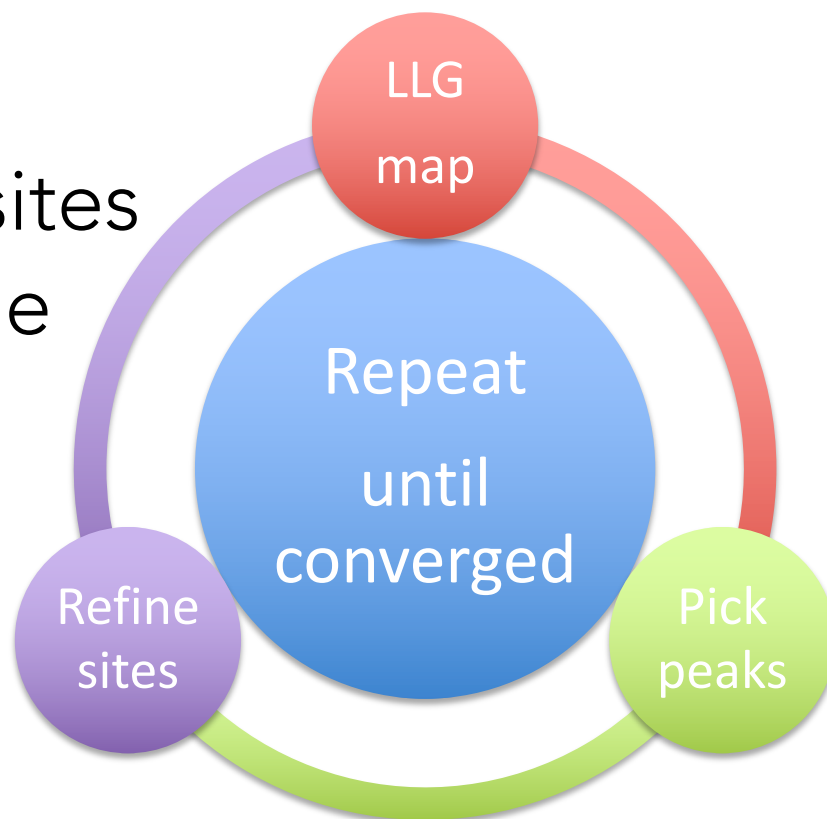
Refining Occupancy

- 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



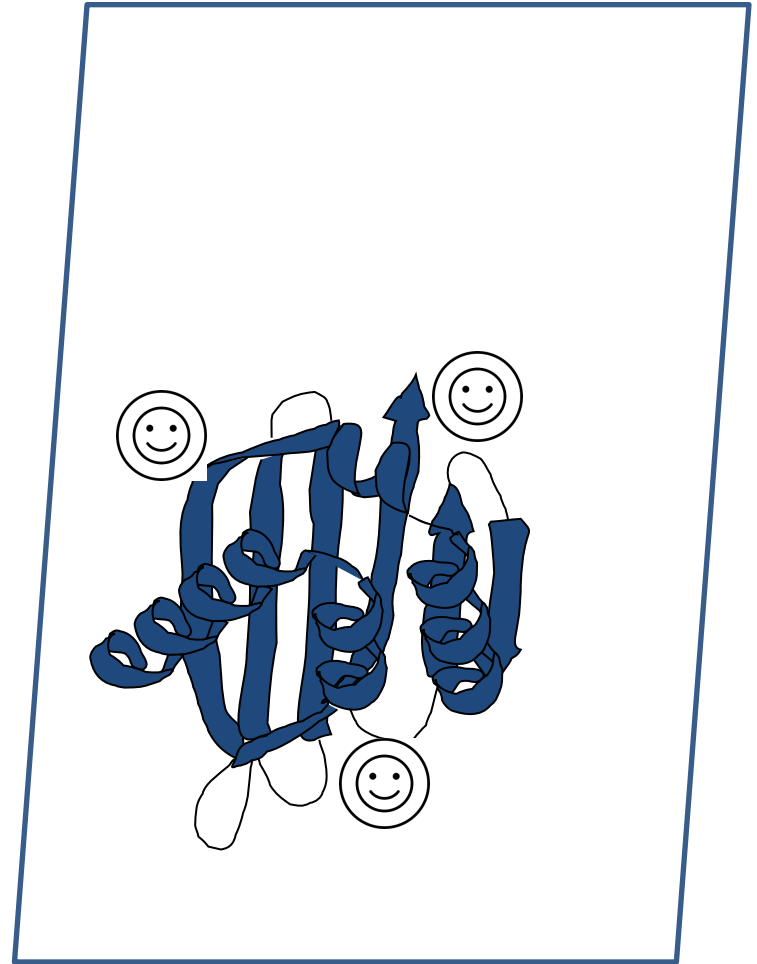
Completion of sub-structure

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



Locating Sites

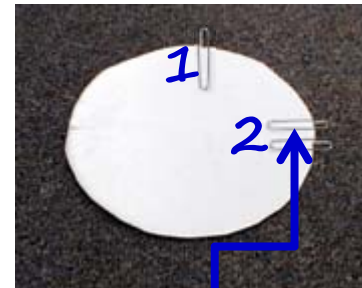
- Log-likelihood
Gradient maps
- Can find very weak sites
- Helps to remove model bias when the model is very poor



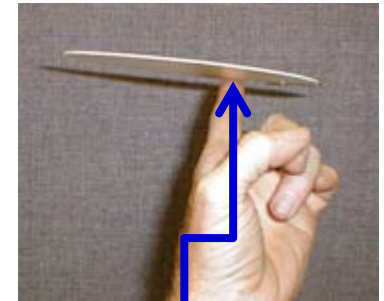
Calculating Electron Density

- 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

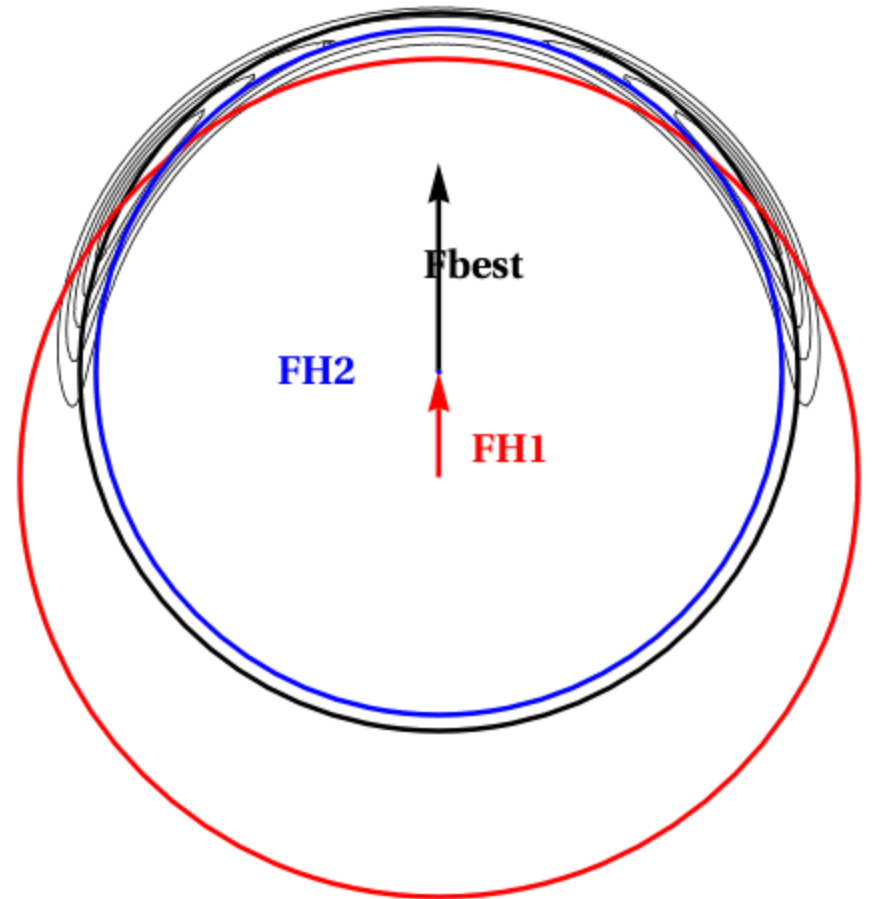


*Most
Probable
Structure
Factor*



*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



Summary

- Fourier Transforms
 - and the phase problem
- SigmaA
 - rmsd
 - partial structure
- phaser
 - LLG and solution probability
 - expected LLG
 - Model selection criteria
 - Use Intensities
- Origins
- Translational non-crystallographic symmetry
- SAD phasing
 - Harker diagrams
 - Best phase
 - HL coefficients