PSD '20 -- Xray Lecture 5, 6

Patterson Space, Molecular Replacement and Heavy Atom Isomorphous Replacement

$$\rho(r) = \sum_{h} \left| F(h) \right| e^{i\alpha(h)} e^{-i2\pi h \bullet r}$$

The Phase Problem

We can't measure the phases!

X-ray detectors (film, photomultiplier tubes, CCDs, etc) **can measure only the intensity of the X-rays** (which is the *amplitude squared*), but we need the full wave equations $Ae^{i\alpha}$ for each reflection to do the reverse Fourier transform.

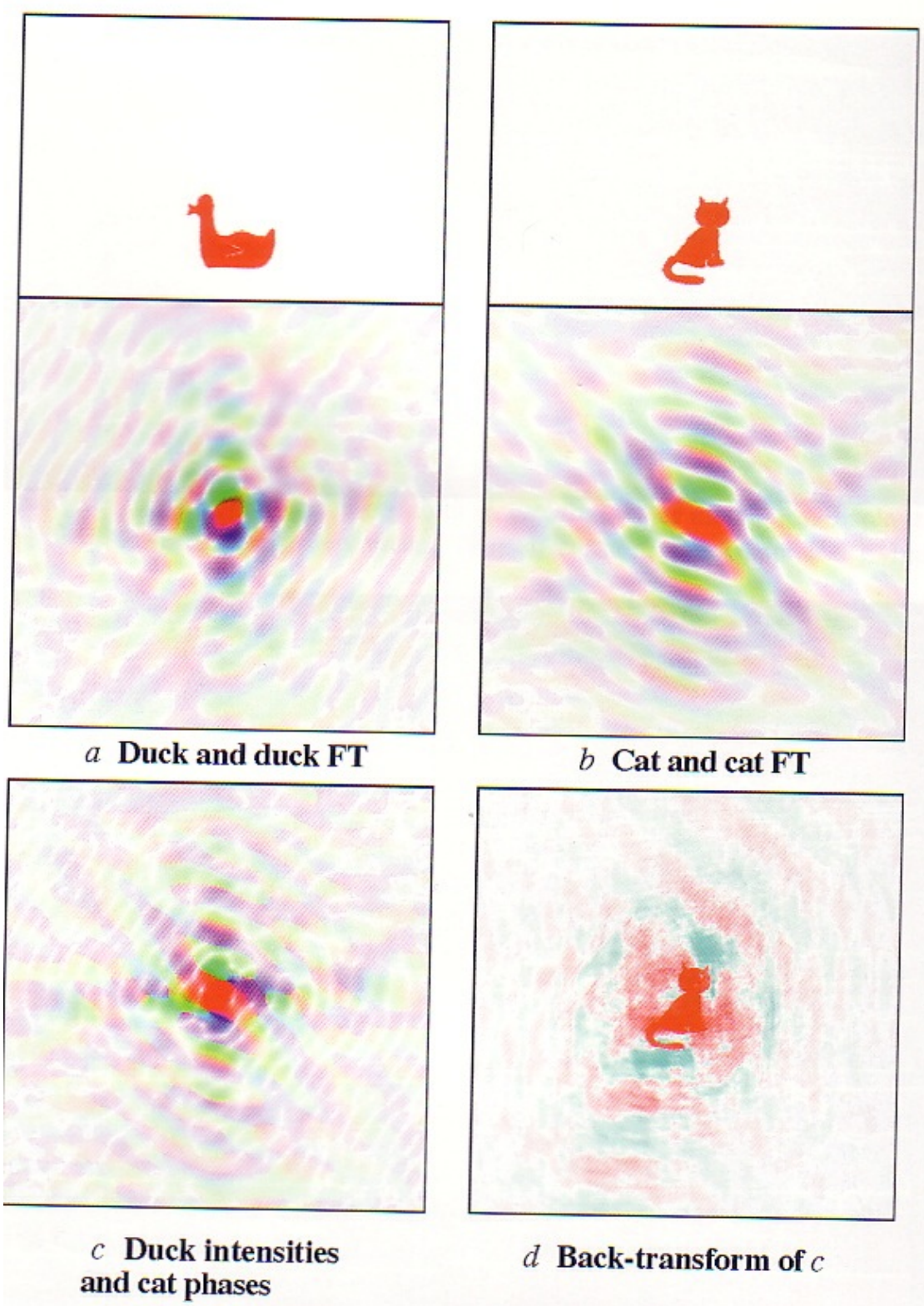
And because it is called the phase "*problem*", the process of getting the phases is called a "solution". That's why we say we "*solved*" the crystal structure, instead of "*measured*" or "*determined*" it.

Phase is more important than amplitude

color=phase angle

darkness=amplitude

Combining the amplitudes of the duck with the phases of the cat, then reversing the FT, we get...the cat.



a **Duck and duck FT**

b **Cat and cat FT**

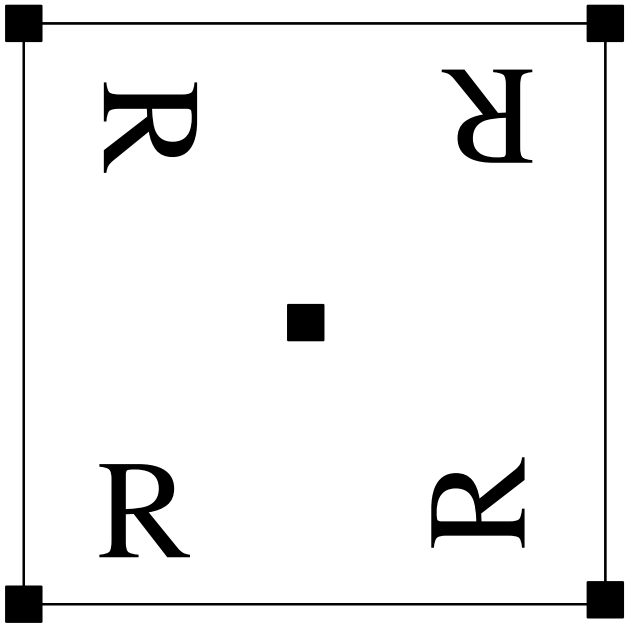
c **Duck intensities and cat phases**

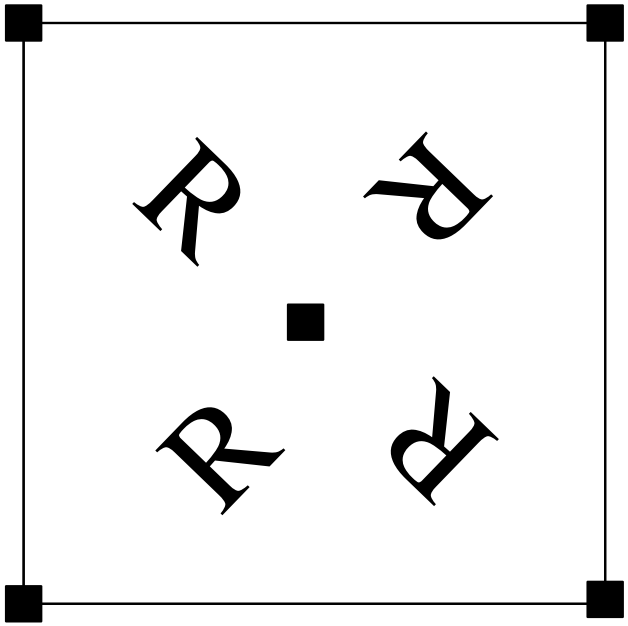
d **Back-transform of** *c*

molecular replacement

If the structure of the molecule is known approximately, then the phases can be calculated.

BUT. We need to know how the molecule is oriented.





Same molecule. Same symmetry. Same cell dimensions.
But... these two crystals are <u>not</u> isomorphous.

We can use *homology* to infer structure

If the protein sequence identity is > 25%, then infer that the sequences are "homologous".

Homologous proteins have similar structures.

How similar is not known until both structures are solved.

Molecular replacement will not work if the structures are too different.

If a homolog of known structure exists, then it can be used to do molecular replacement

The unknown is the *orientation*

The space of all possible *rigidbody* transformations of a molecule has 6 dimensions. 3 angles of rotation (defining a matrix of 9 coefficients), and a 3D translation vector.

$$x' = c_{11}x + c_{21}y + c_{31}z + v_x$$

$$y' = c_{12}x + c_{22}y + c_{32}z + v_y$$

$$z' = c_{13}x + c_{23}y + c_{33}z + v_z$$

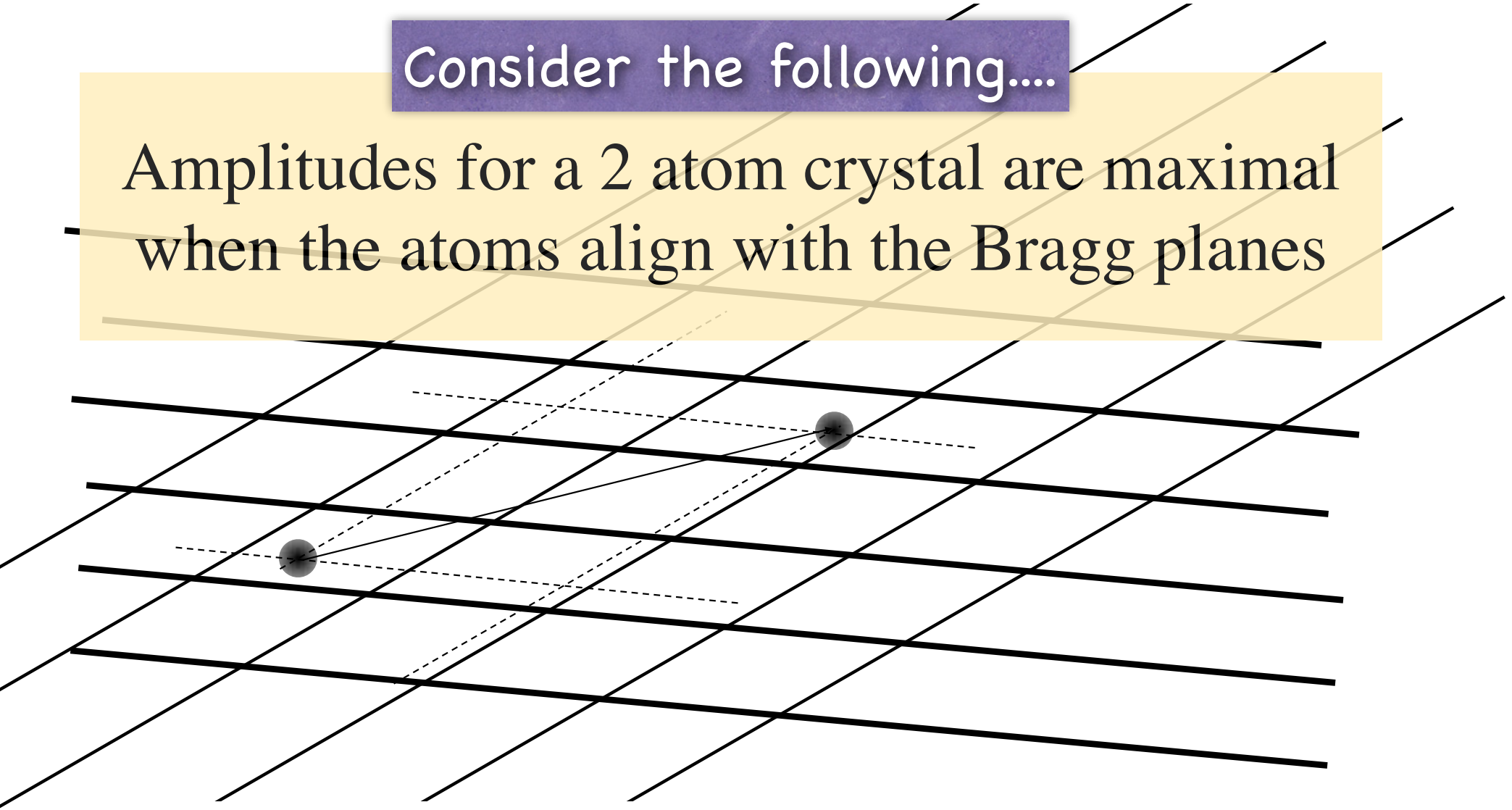
or $x' = \underline{C}x + v$

Therefore, the position of our molecule in the crystal unit cell must be a 6D transformation of its current position. Molecular replacement is the method for finding the angles and vector that define the transformation.

Patterson Space

Consider the following....

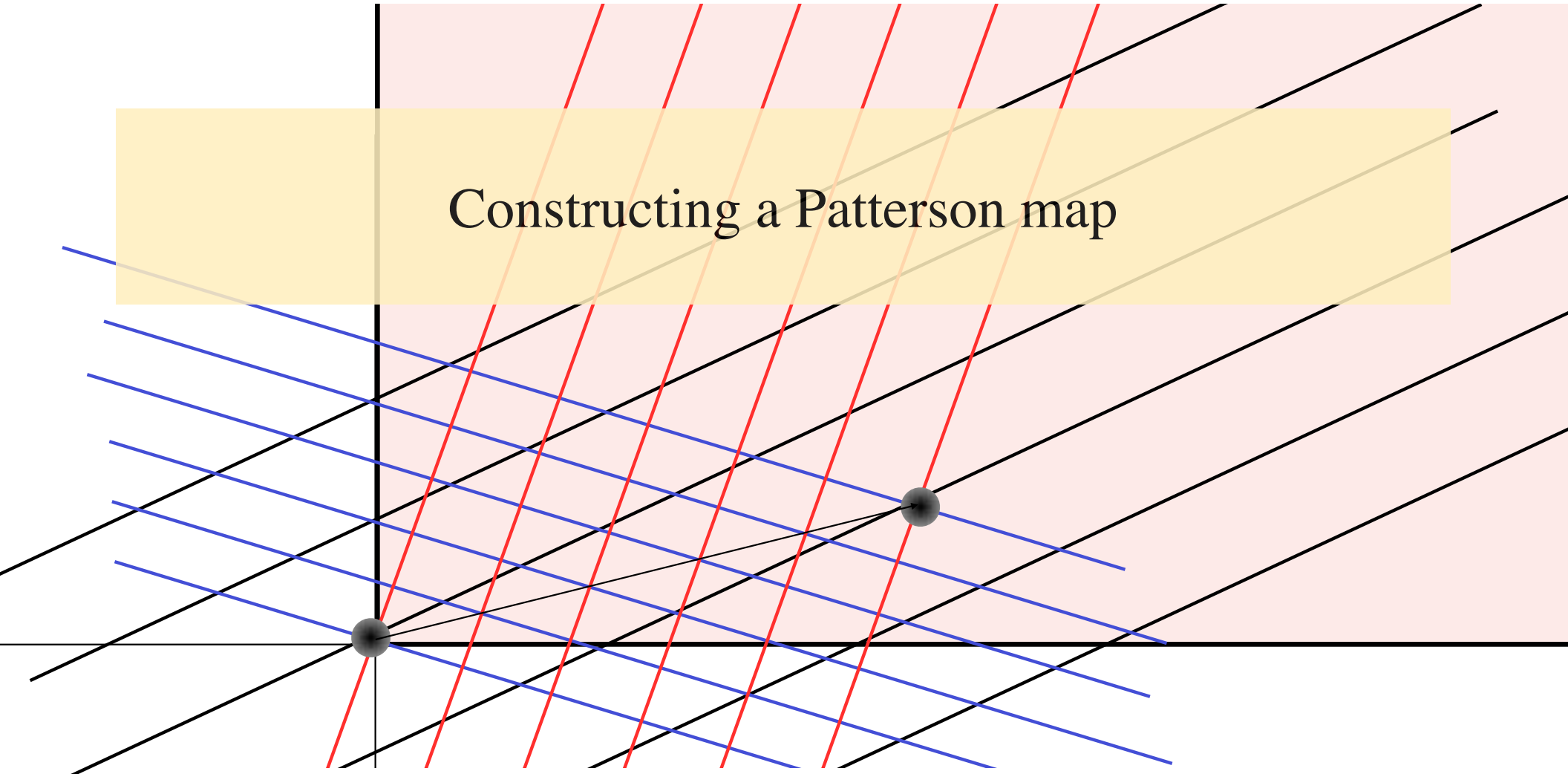
Amplitudes for a 2 atom crystal are maximal when the atoms align with the Bragg planes



This of two symmetry-related heavy atoms....

The amplitude of a wave F(h k l) scattered by just two atoms depends how similar are their phases. If they have the same phase, any phase it doesn't matter, then the amplitude is greatest.

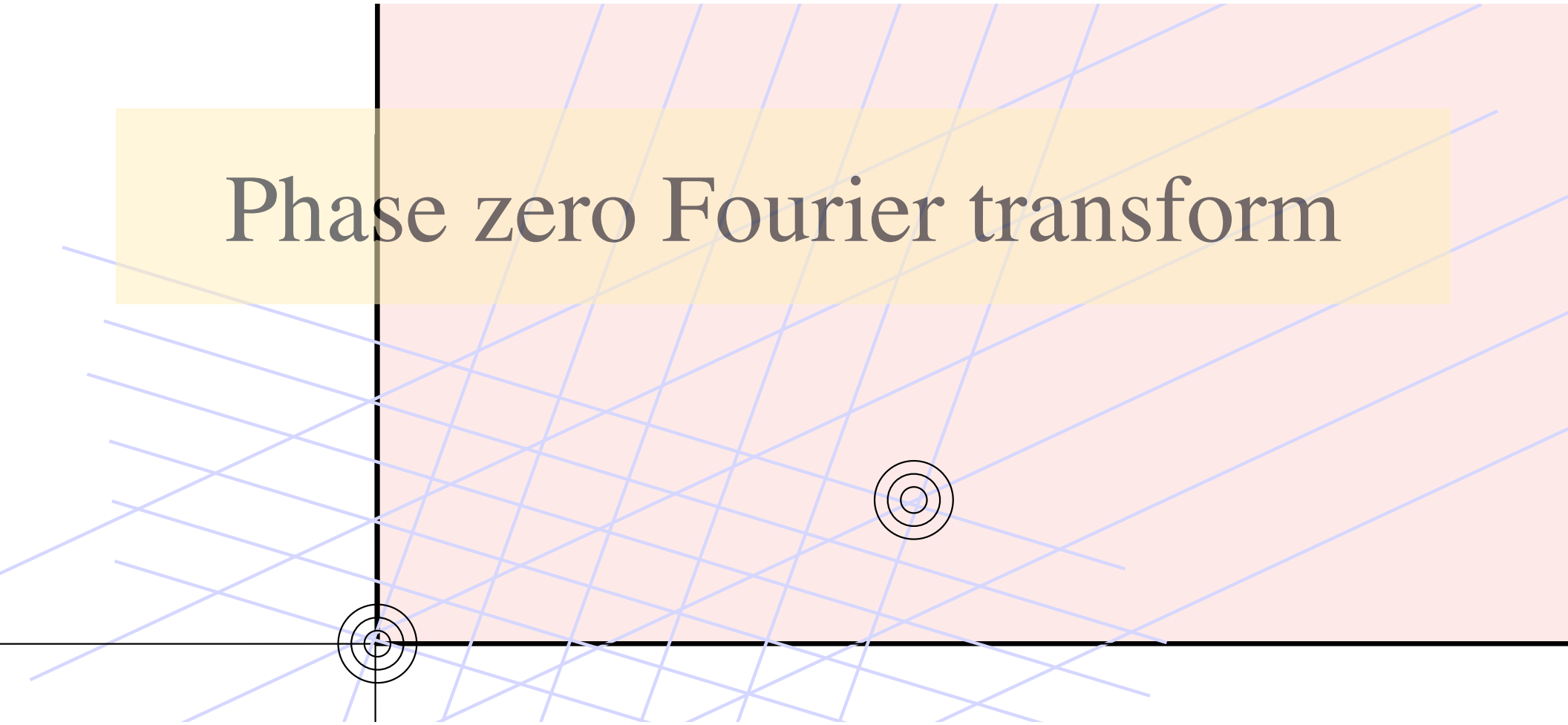
Constructing a Patterson map



If these planes have the largest amplitudes, the atoms must be at the intersection of these Bragg planes.

Does not require phases!!

Phase zero Fourier transform



If we calculate the inverse FT with all phases set to zero, the "wave crests" of electron density all go through the origin.

But the other places where wave crests intersect, are the relative positions of heavy atoms.

The phase zero inverse Fourier is called a Patterson Map

Reverse Fourier Transform vs Patterson Function

amplitude phase

Reverse FT *with* phases

$$\rho(r) = \sum_h |F(h)| e^{i\alpha(h)} e^{-i2\pi h \bullet r}$$

Reverse FT *without* phases

$$\rho(r) = \sum_h |F(h)| e^{-i2\pi h \bullet r}$$

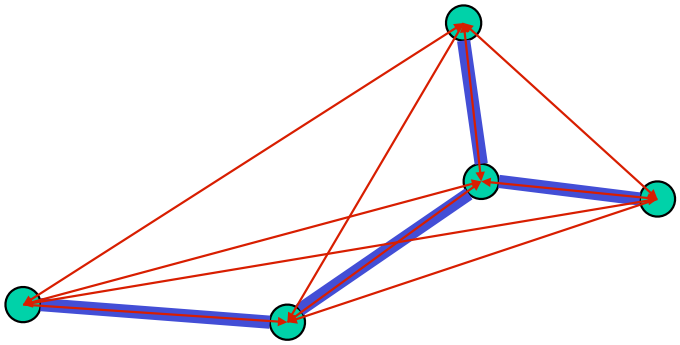
This is the ***Patterson function***

it uses the measured amplitude, no phase

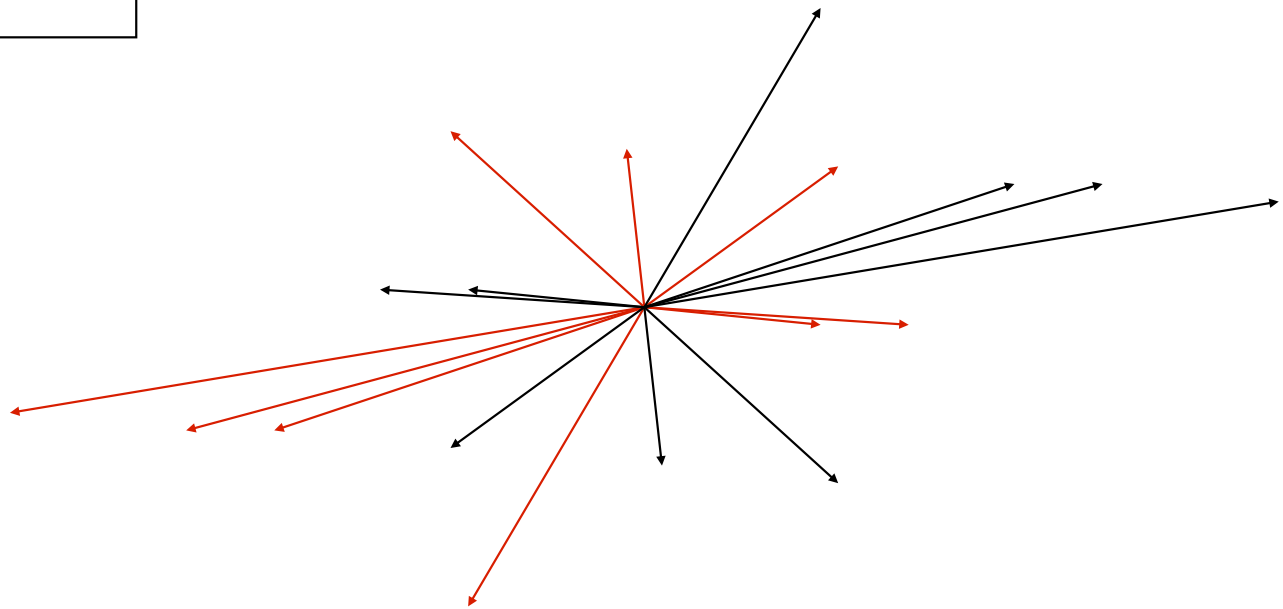
Patterson map represents all inter-atomic vectors

To generate a centro-symmetric projection in 2D, draw all inter-atomic vectors, then move the tails to the origin. The heads are where peaks would be.

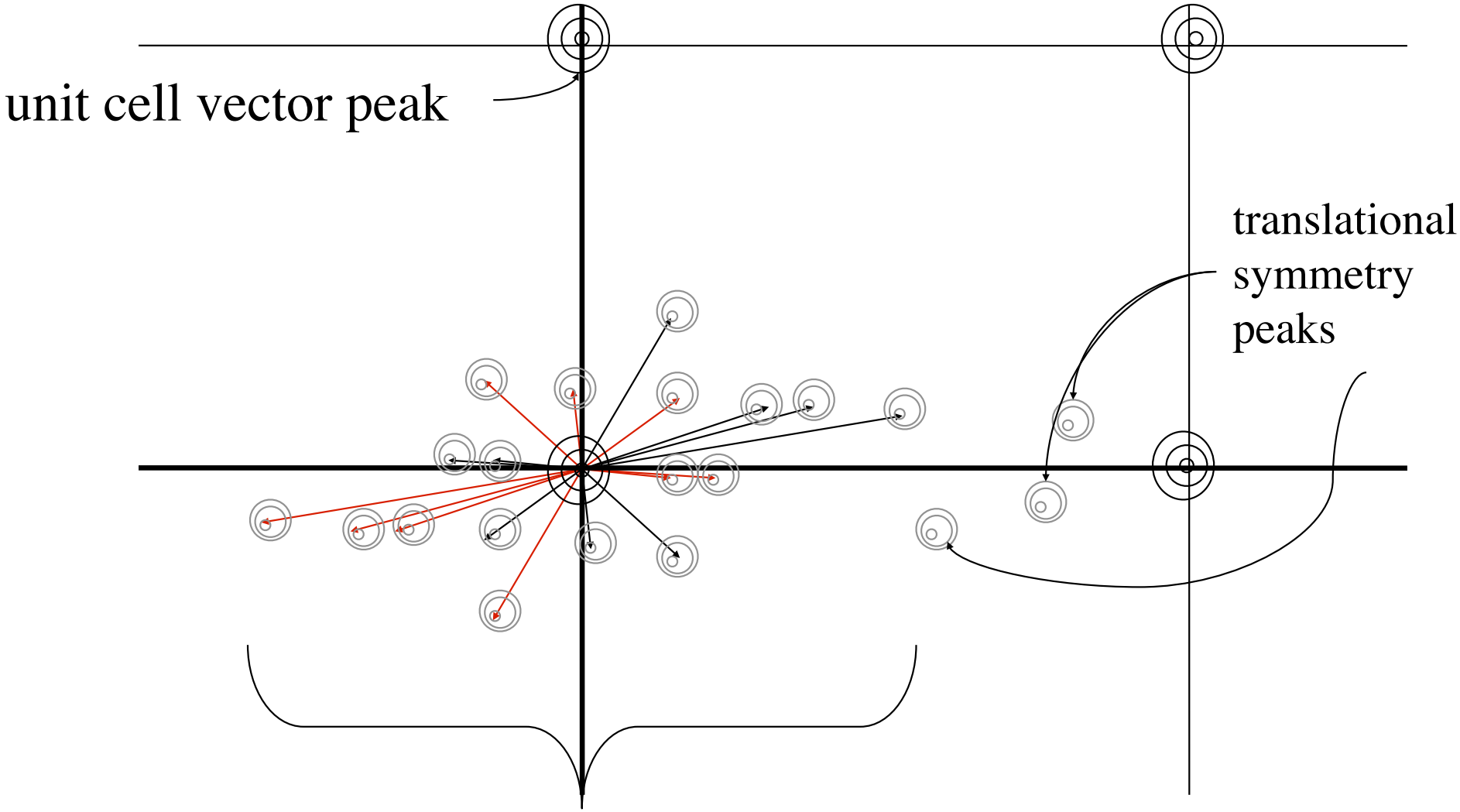
For example, take glycine, 5 atoms (not counting H's)



Move each vector to the origin

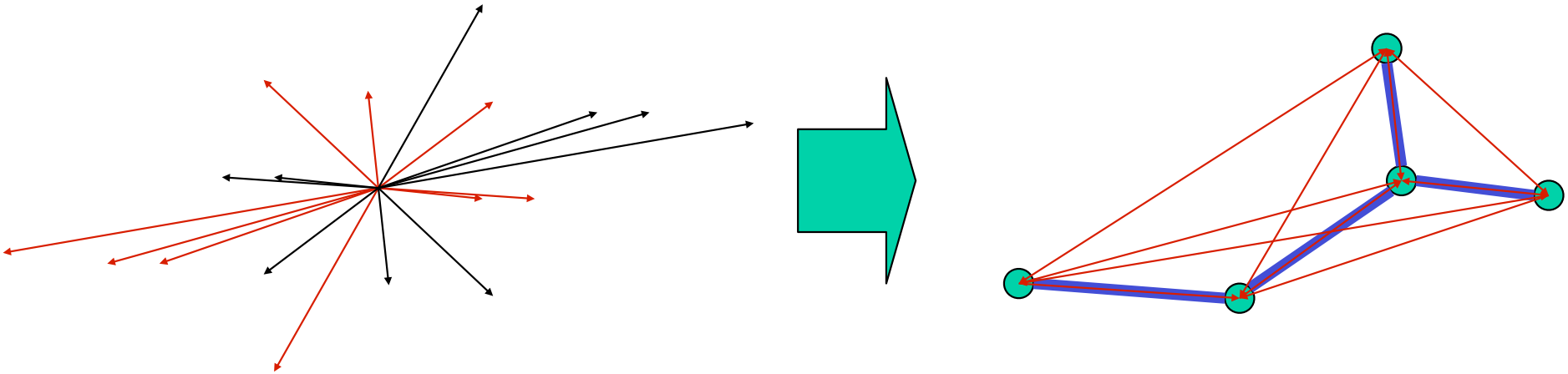


Patteron map for Gly in P1



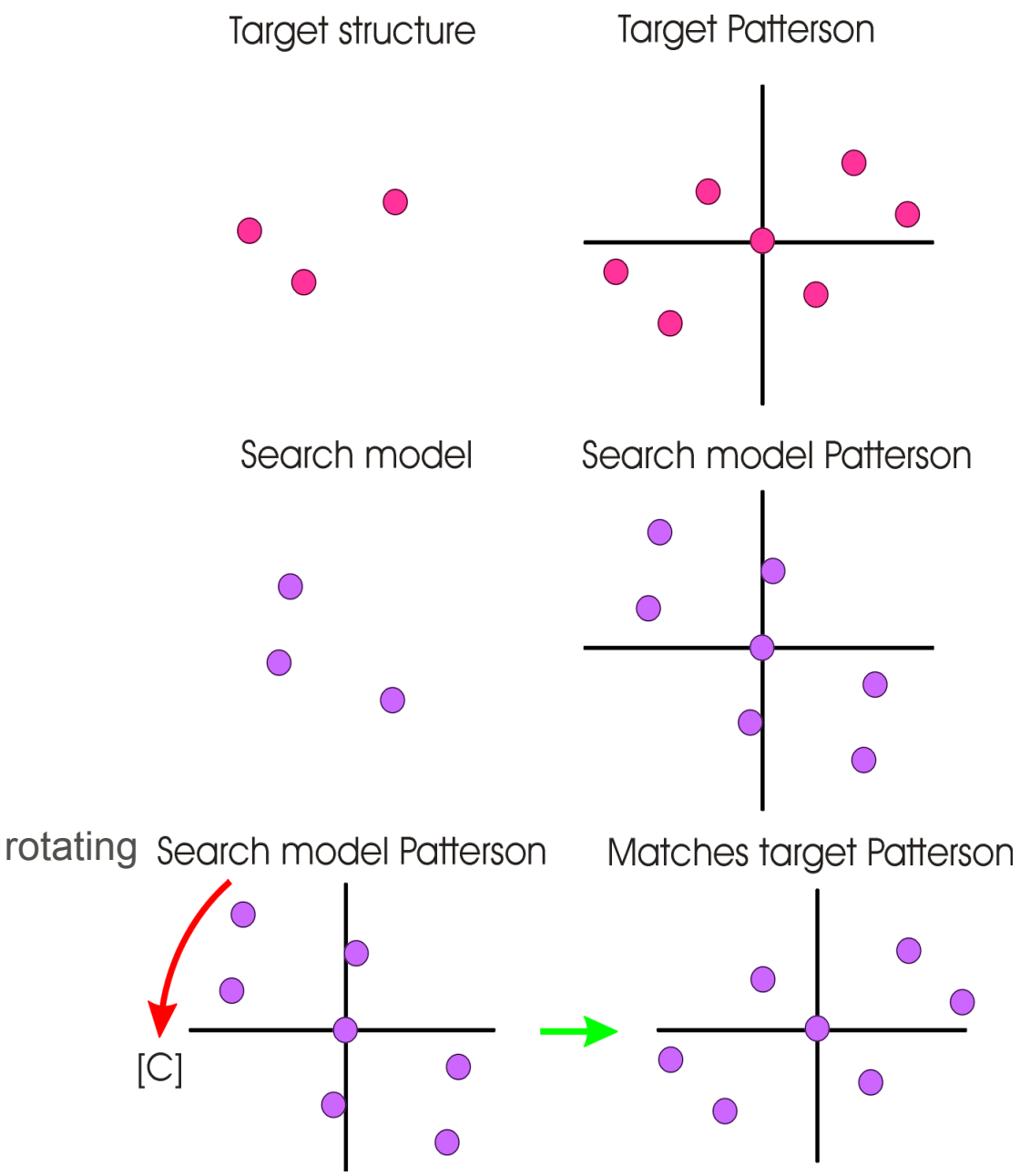
Can you reassemble glycine from this?

For small molecules, vector/geometry problem can be solved...

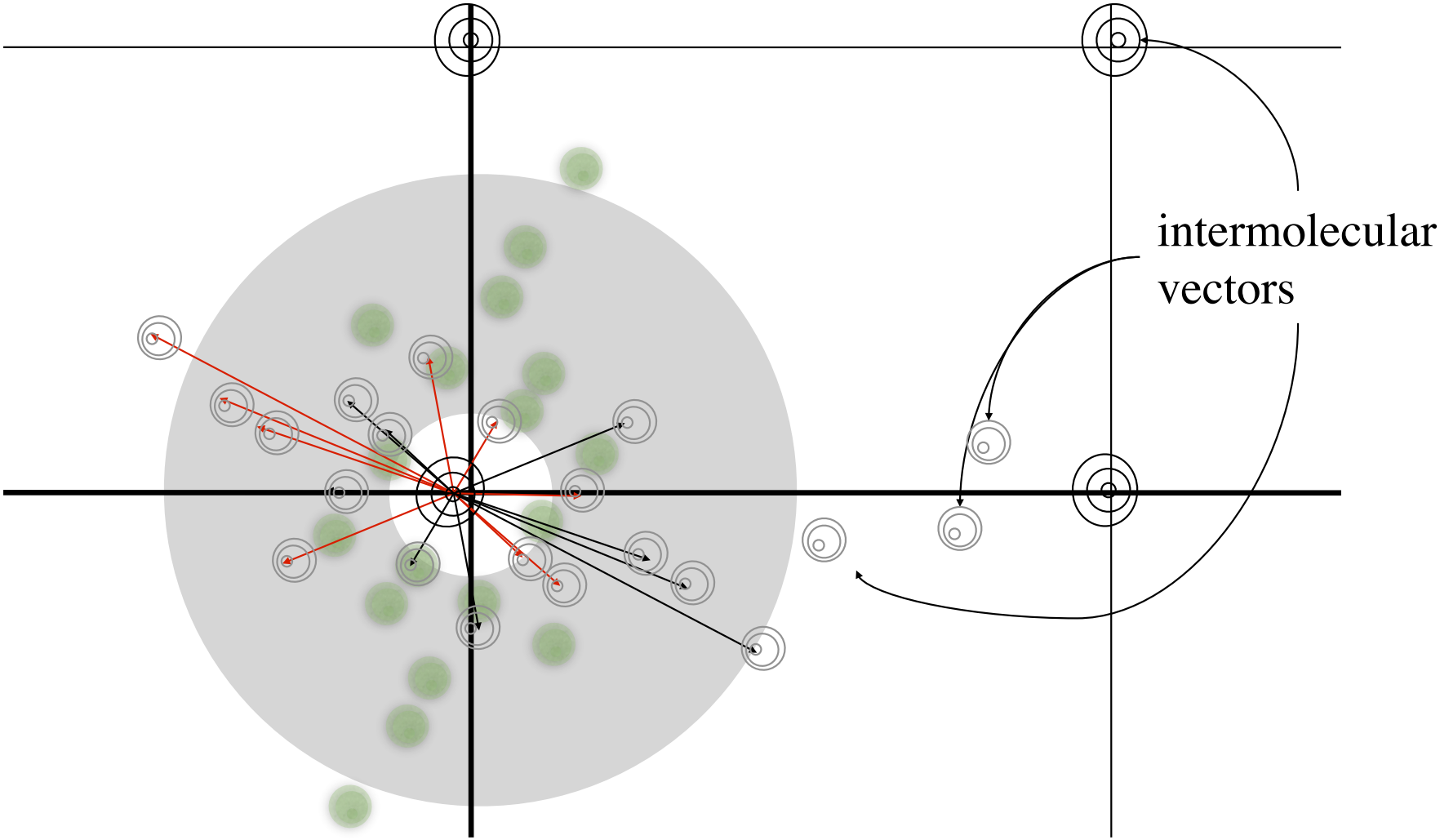


...if you know the stereochemistry (bond lengths, angles) of the molecule.

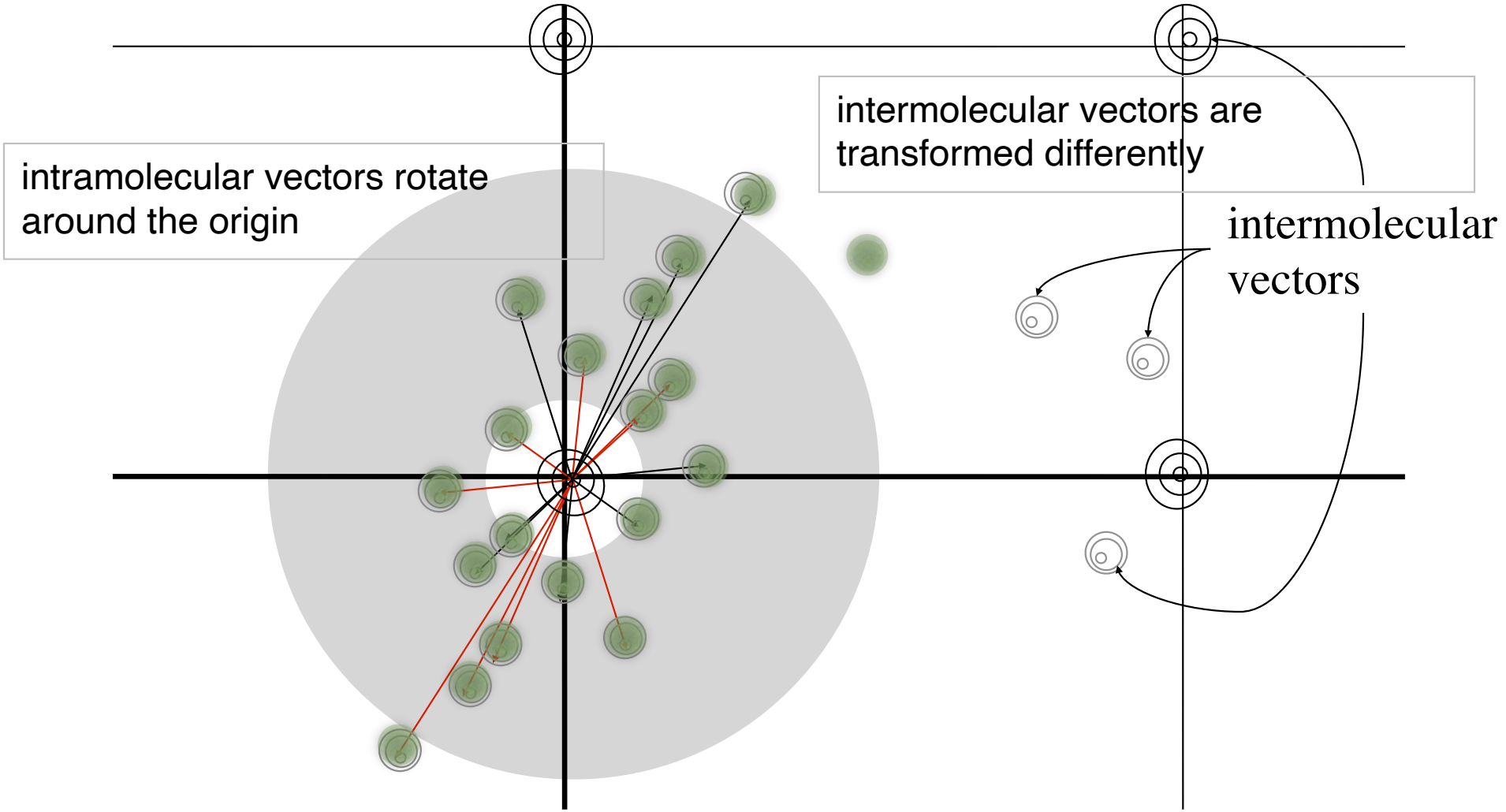
Rotation Function



Rotating the Patterson



Rotating the Patterson



The correlation function

The correlation between any two functions *x* and *y* is defined as:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

x and y are functions with the same range. x-bar means the mean value of the function *x*

If the correlation is perfect, *r=1.000*

If the anti-correlation is perfect, *r=-1.000*

If there is no correlation, *r* is close to zero.

Patterson correlation function

$$r = \frac{\sum \left(P_o(v) - \overline{P}_o\right)\left(P_{\text{mod}}(v) - \overline{P}_{\text{mod}}\right)}{\sqrt{\sum \left(P_o(v) - \overline{P}_o\right)^2 \sum \left(P_{\text{mod}}(v) - \overline{P}_{\text{mod}}\right)^2}}$$

= correlation between observed Patterson and rotated, model Patterson

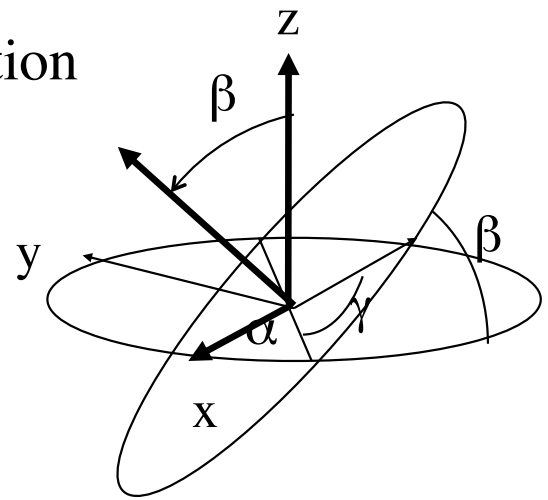
$\left(P_{\text{mod}}(v) - \overline{P}_{\text{mod}}\right)$ = rotated model Patterson, at v, mean corrected.

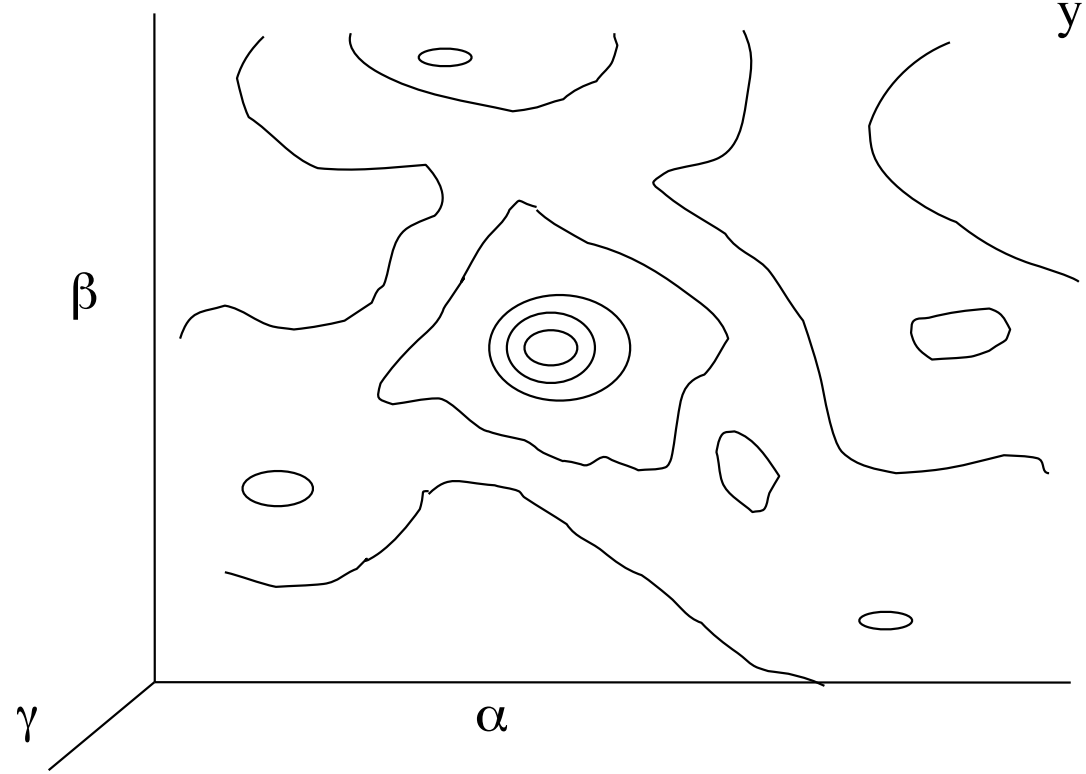
The sums Σ are done over v in a spherical shell of the Patterson map that excludes the self-peak and the very long vectors.

Typically, $4\text{Å} \leq |v| \leq 20\text{Å}$, is a good range for the rotation function.

The Rotation Function

Three angles (α,β,γ) define all possible rigidbody rotations. The solution of the rotation function are the angles that give the highest *Patterson correlation function*.





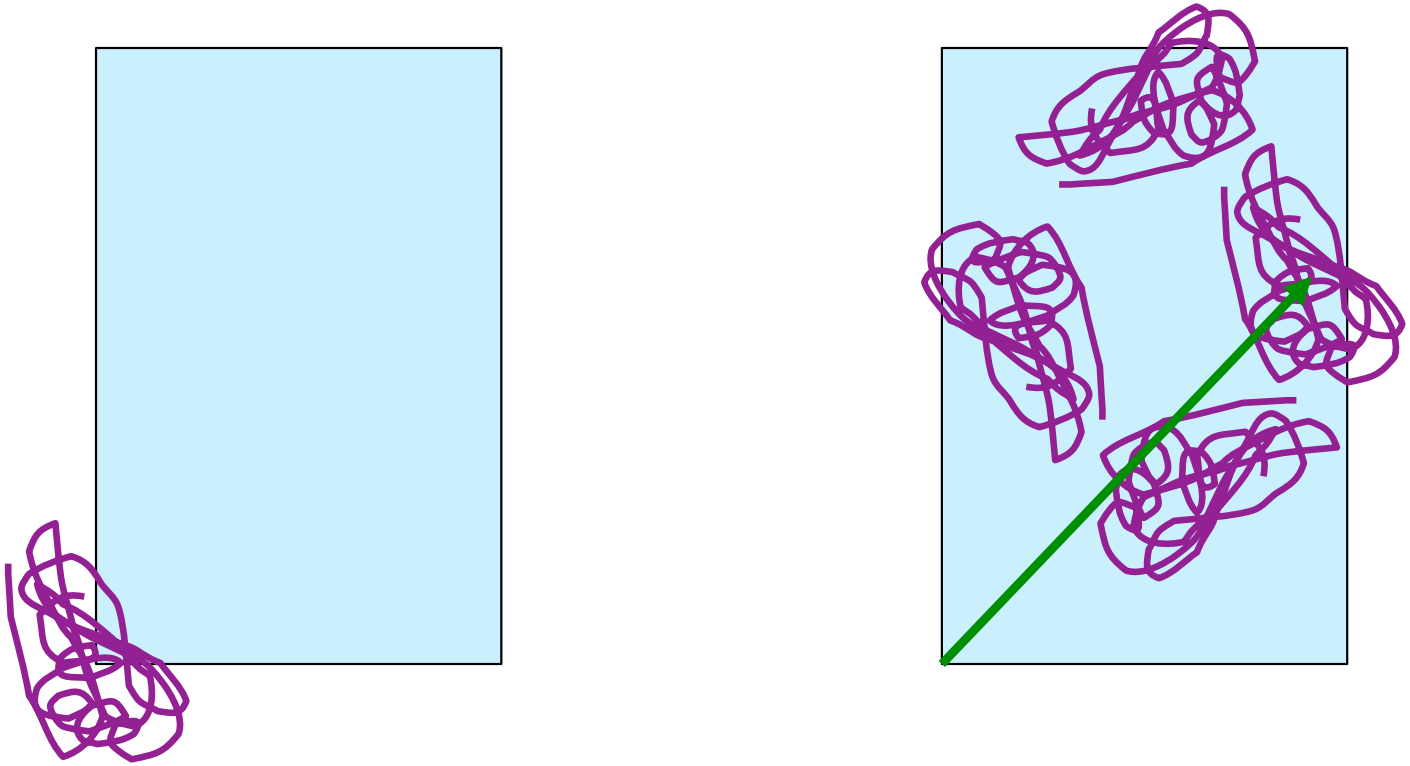
Non-crystallographic symmetry can be detected using the Self Rotation Function

If the native Patterson is rotated against itself and the correlation (r) is calculated, the result (call the "Self Rotation Function") will have at a non-symmetry-related position only if the asymmetric unit has NON-CRYSTALLOGRAPHIC SYMMETRY (NCS).

NCS means that an envelope of the asu exists for which:

$$\rho(r) = \rho(\underline{M}_{ncs} r + v_{ncs})$$

The Translation Function



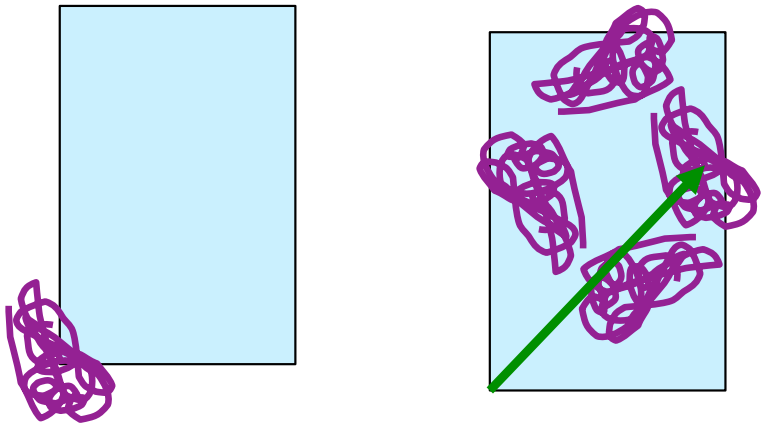
The model is oriented correctly with respect to the cell axes, but it is still at the origin. We need a translation vector (green) to translate the model to its position in the crystal unit cell relative to the origin.

How do we know which vector to use?

The Translation Function

1. Translate model
2. Generate symmetry
3. Check for good packing.
4. Calculate R-factor

===> keep translation with best R-factor



A translation of the coordinates is:

$x' = x + t$

Symmetry positions are calculated using symops (M, v)

$x' = Mx + v$ (M is the matrix and v is the vector)

Combining, we can express each translation as:

$x' = M(x + t) + v = Mx + Mt + v$

The Translation Function

Equation to calculate F_{calc} give a translation vector *t*.

$$\mathbf{F_{calc}(h)} = \sum_{\underline{Z}} \sum_{g} f_g(\mathbf{h}) e^{i2\pi \mathbf{h} \bullet (\underline{Z}(\mathbf{r}_g + \mathbf{t}))}$$

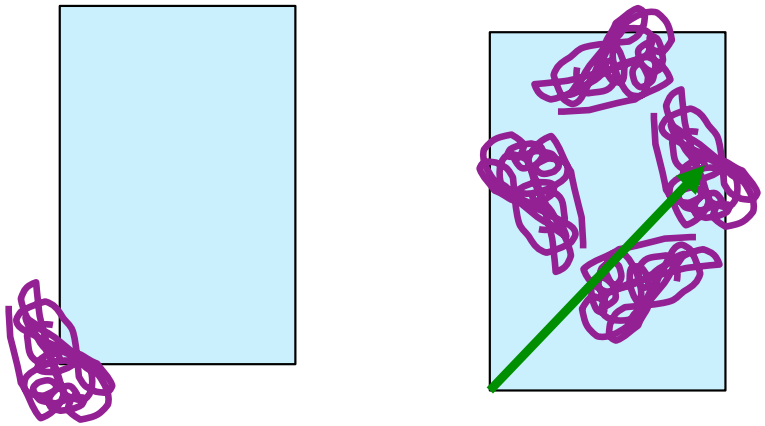
where

g = all atoms,

$\underline{Z}$ = all symops,

$\mathbf{r}_g$ = fractional coordinates of atom g,

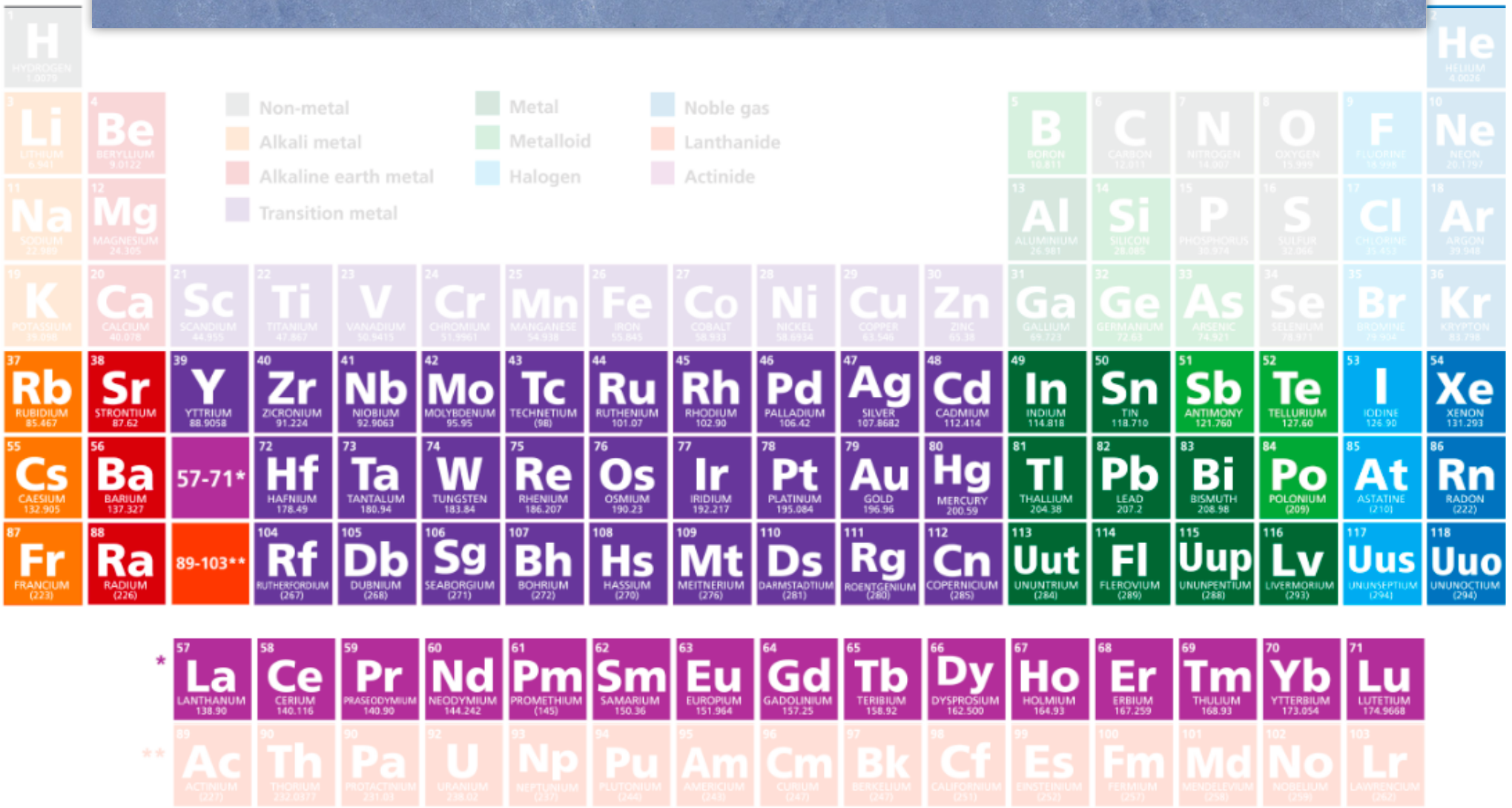
t = translation vector.



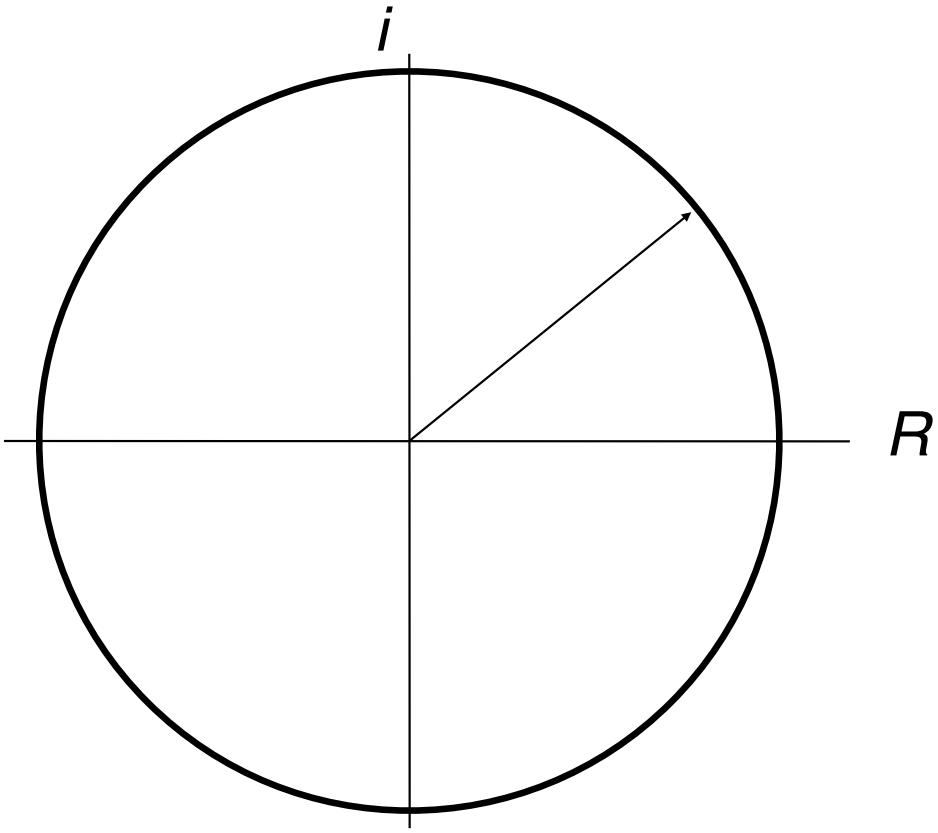
Equation to calculate R-factor given F_{calc} as a function of **t**.

$$R(t) = \frac{\sum_{h} \left\| F_{obs}(h) \right| - k \left| F_{calc}(h,t) \right\|}{\sum_{h} \left| F_{obs}(h) \right|}$$

Heavy Atom Methods



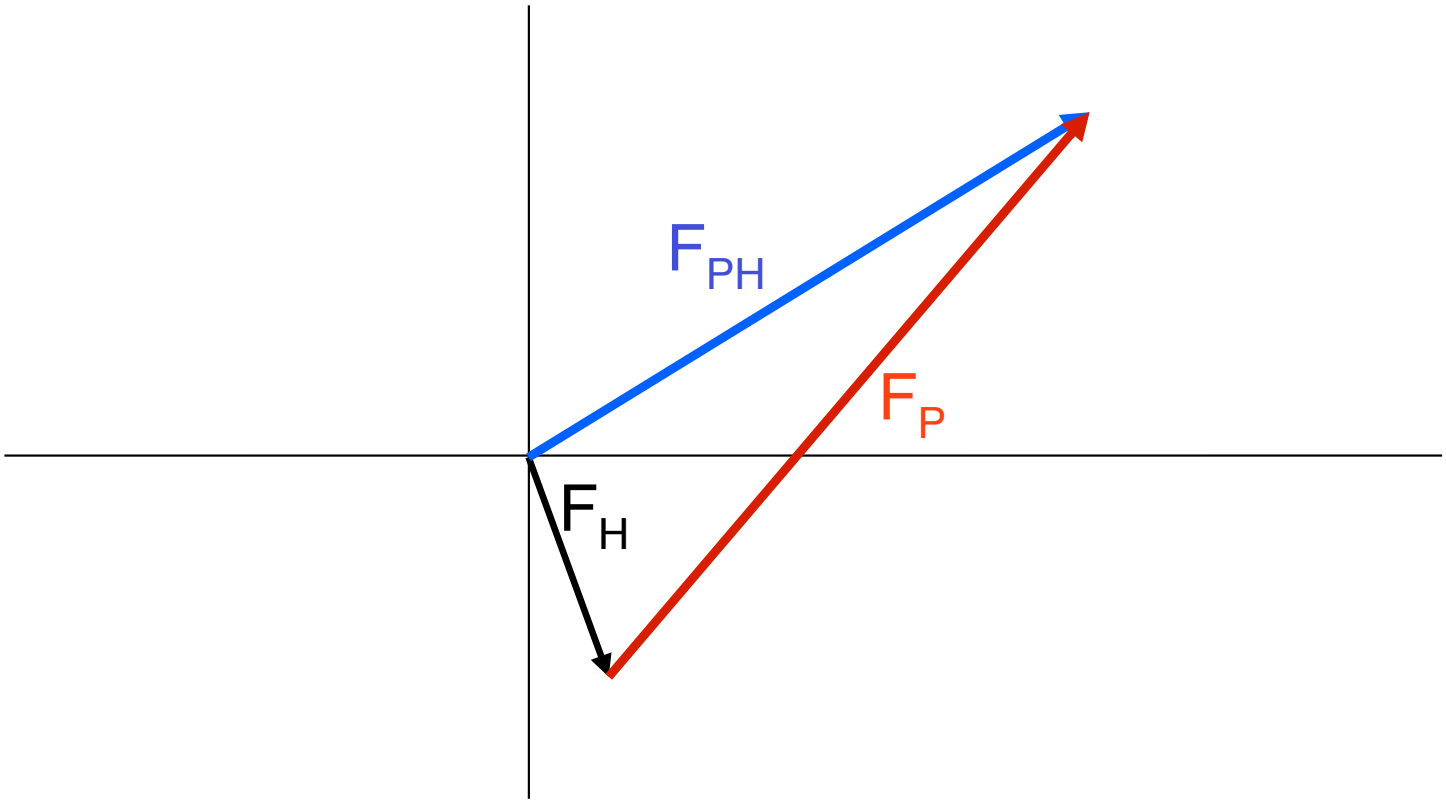
We can represent a structure factor of *known amplitude* and *unknown phase* as a circle in Argand space.



Radius of the circle is the amplitude. The true *F* lies somewhere on the circle.

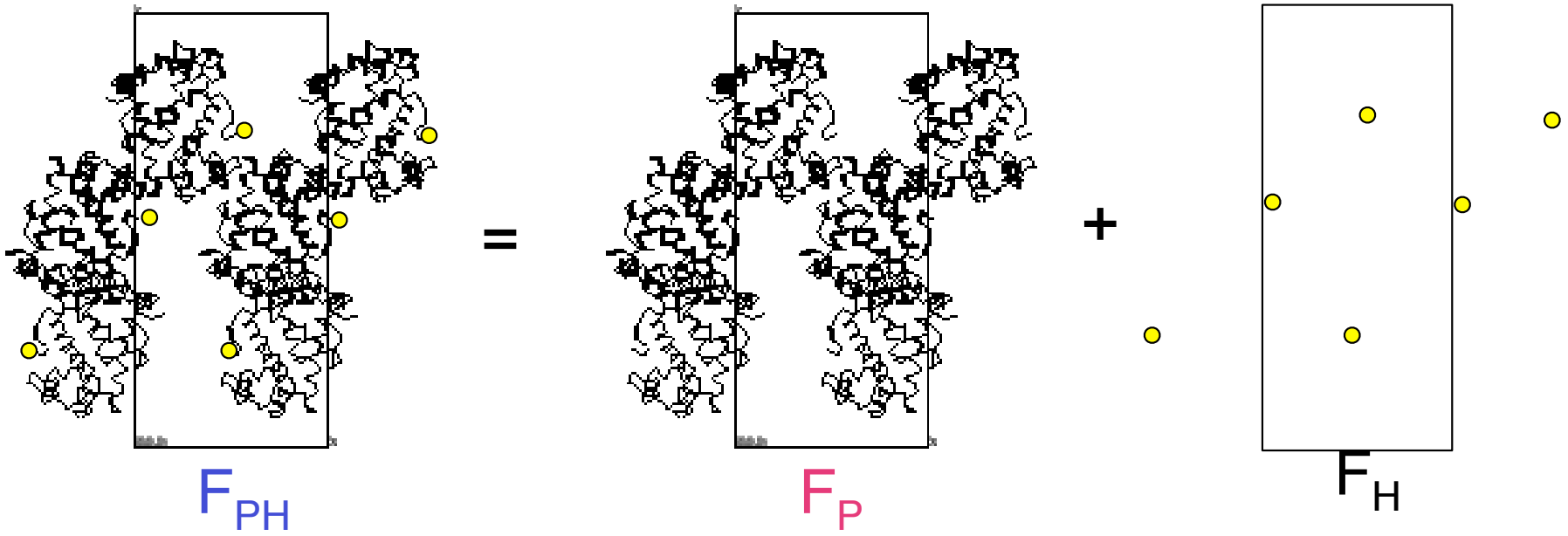
Heavy atom isomorphous replacement

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



Heavy atom isomorphous replacement

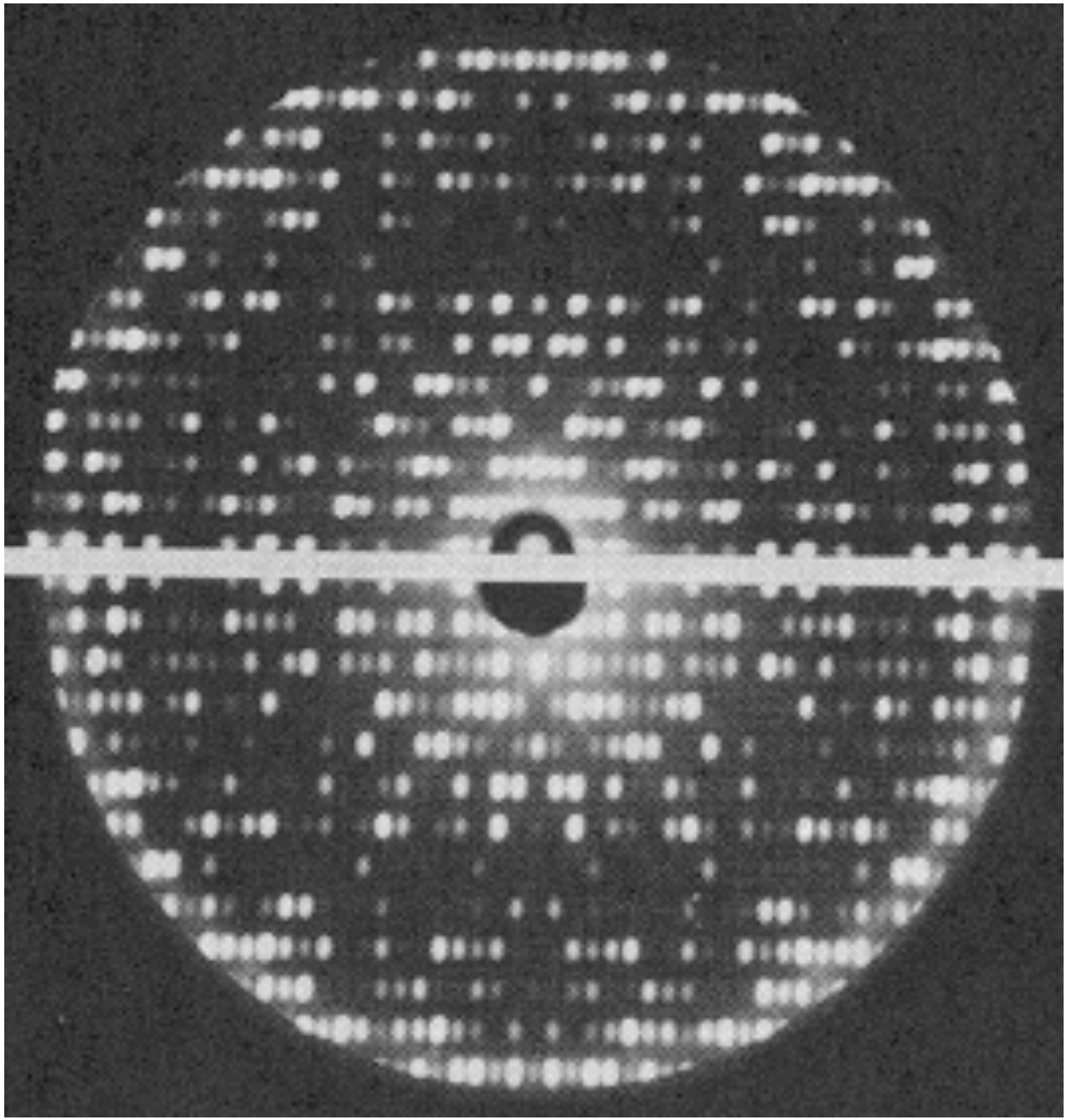
=Turning proteins into small molecules by soaking in heavy atoms



The Fourier transform (i.e. diffraction pattern) of a heavy atom derivitive is the vector sum of the transforms of the protein and the heavy atoms.

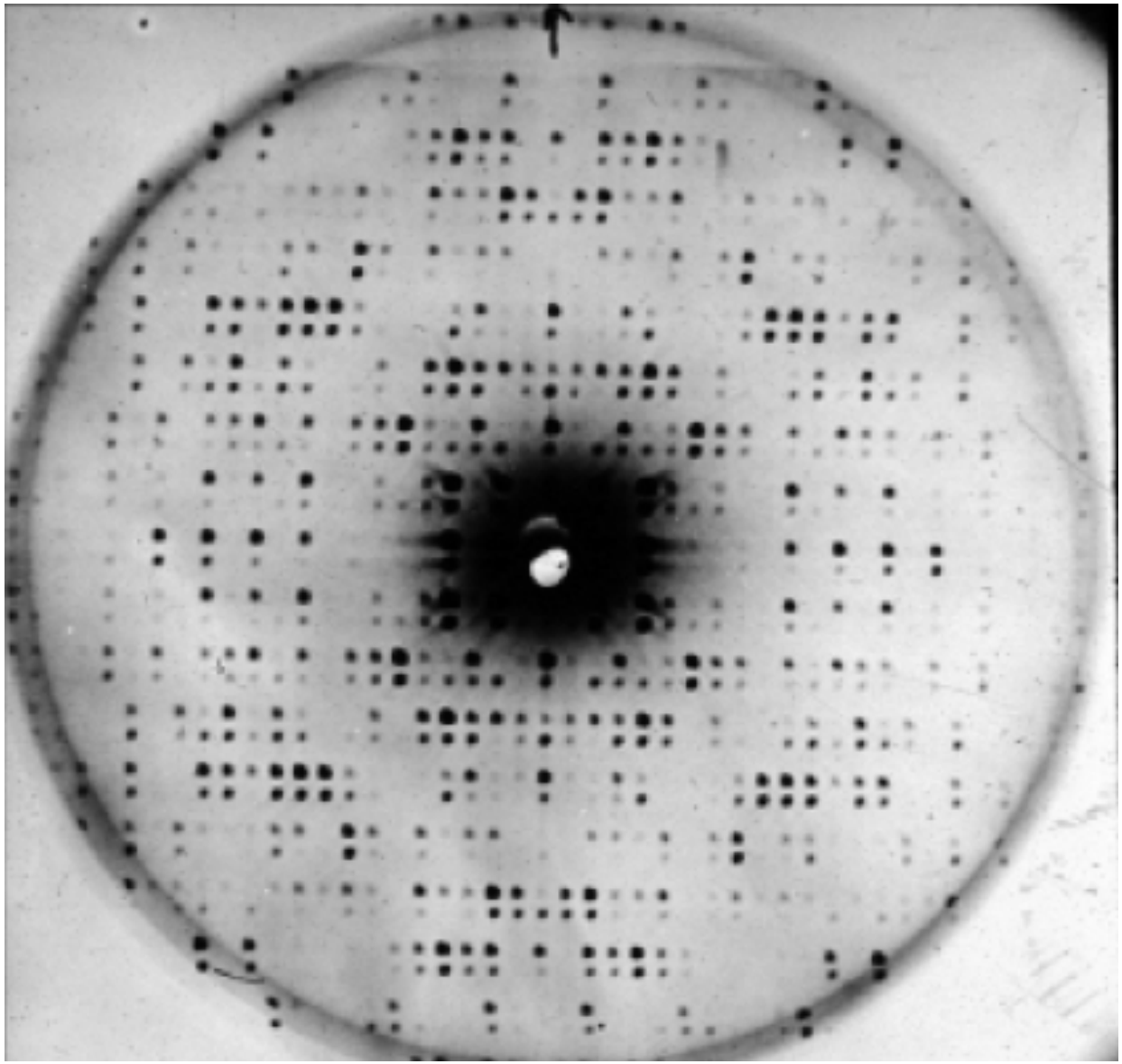
NOTE: protein and protein-heavy atom crystals must be *isomorphous*.

Comparing parent and heavy atom data sets.



The upper and lower images are two precession photographs, showing the l=0 level of reciprocal space. The upper is the original protein crystal. The lower is after soaking in heavy atom solution. Note changes in intensity.

Two protein diffraction patterns superimposed and shifted vertically relative to one another. One is from native bovine β-lactoglobulin and the other is from a crystal soaked in a mercury-salt solution. Note the intensity changes for certain reflections and the identical unit cells (spacing of the spots) suggesting isomorphism. (Photograph courtesy of Professor Lindsay Sawyer.)



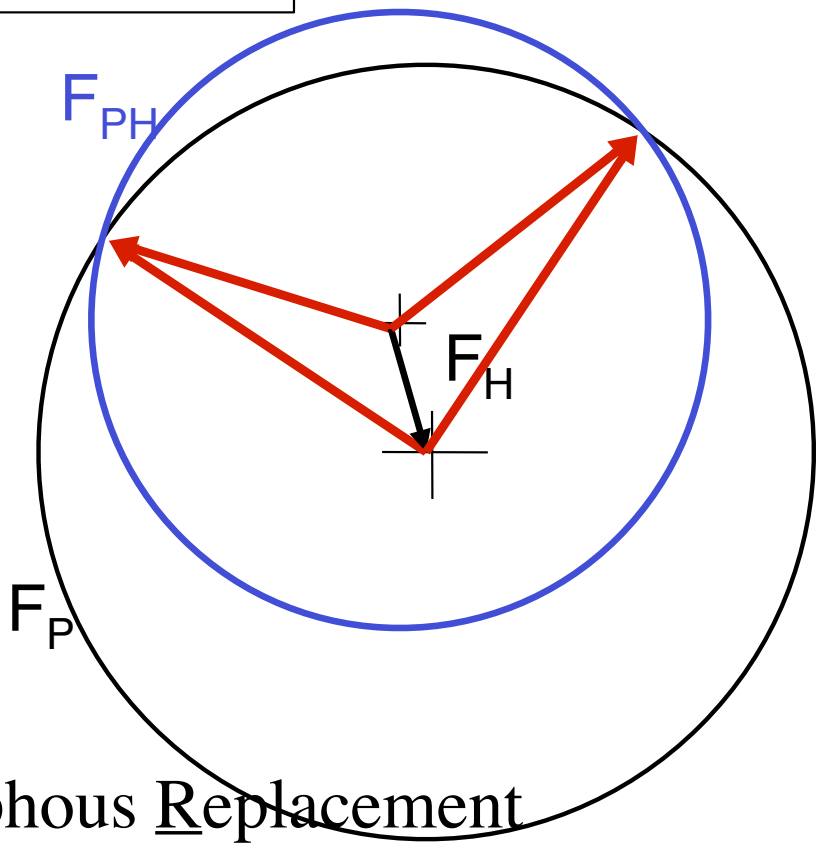
https://openi.nlm.nih.gov/detailedresult.php?img=PMC2852296_d-66-00325-fig5&req=4

Harker diagram method for discovering phase from amplitudes.

We know only amplitude

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

For the heavy atom F_H, we know both amplitude and phase

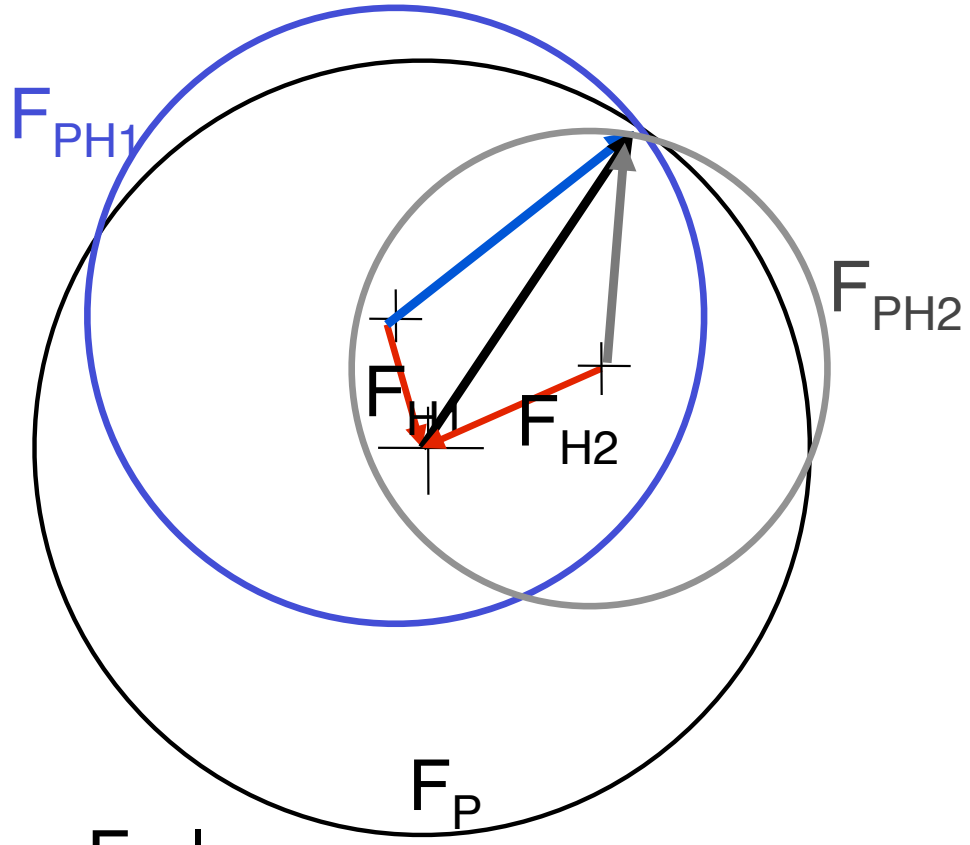


One heavy atom (SIR): There are two ways to make the vector sums add up.

SIR = Single Isomorphous Replacement

Two heavy atom derivatives (MIR), produce unambiguous phases

Multiple Isomorphous Replacement



$|F_P + F_{H1}| = |F_{PH1}|$

$|F_P + F_{H2}| = |F_{PH2}|$

Or

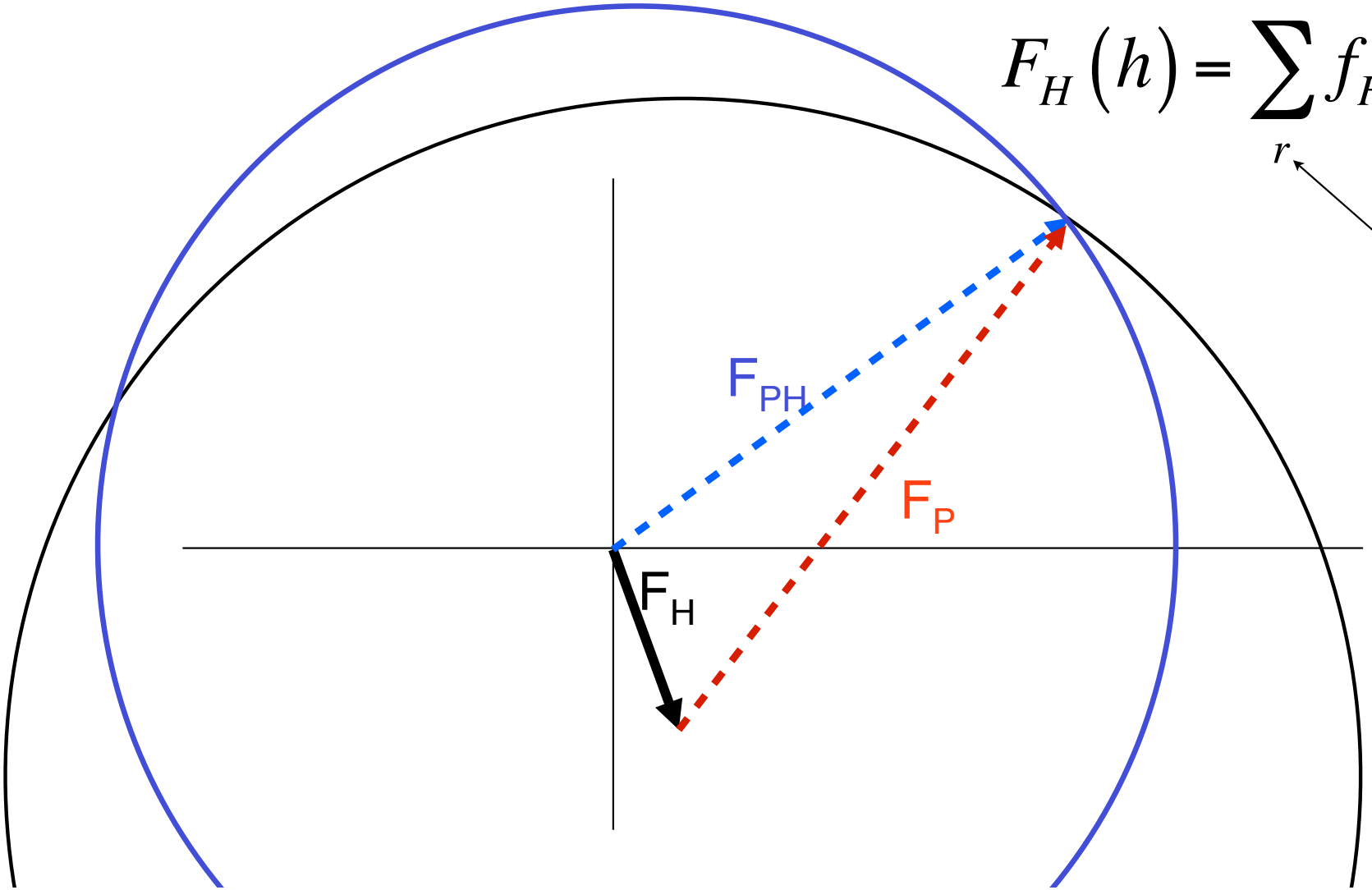
$|F_P| = |F_{PH1} - F_{H1}| = |F_{PH2} - F_{H2}|$

- Phases are more important than amplitudes.
- We can't measure phases, only amplitudes.
- By adding heavy atoms, we change the amplitudes by a significant amount.
- If we know the contribution of the heavy atom, we can solve for the phase of the protein.
- SIR = single isomorphous replacement, gives an ambiguous phase.
- MIR = multiple isomorphous replacement, gives an unambiguous answer.

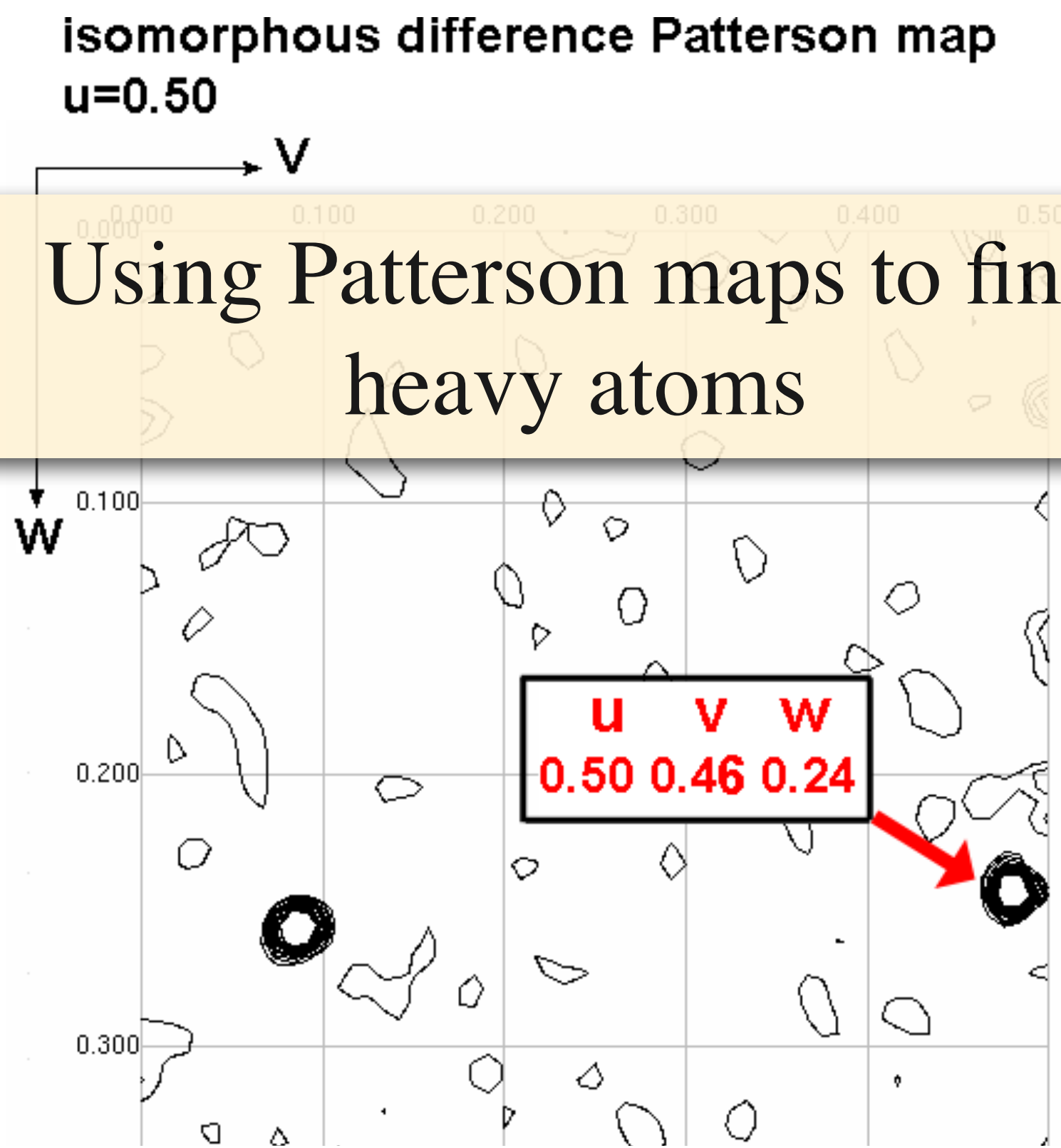
If we can locate the heavy atom, then we can calculate its contribution to F_{PH}

$$F_H(h) = \sum_r f_H e^{i2\pi h \bullet r}$$

r ← real-space locations of heavy atoms, r.



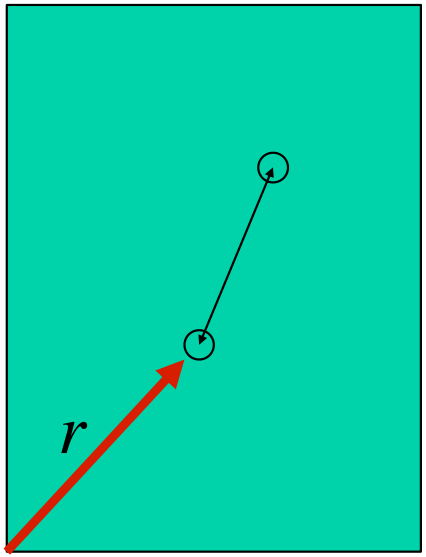
Using Patterson maps to find heavy atoms



"Patterson space"

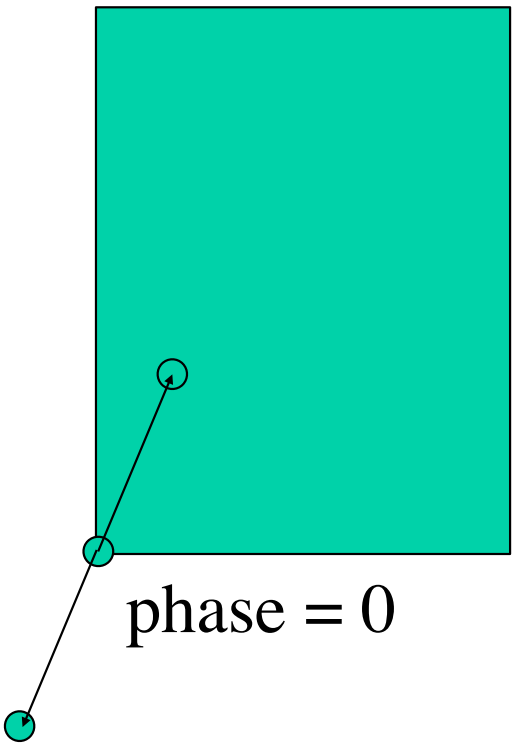
The Patterson map is the centrosymmetric projection

real space



phase=α

Patterson space

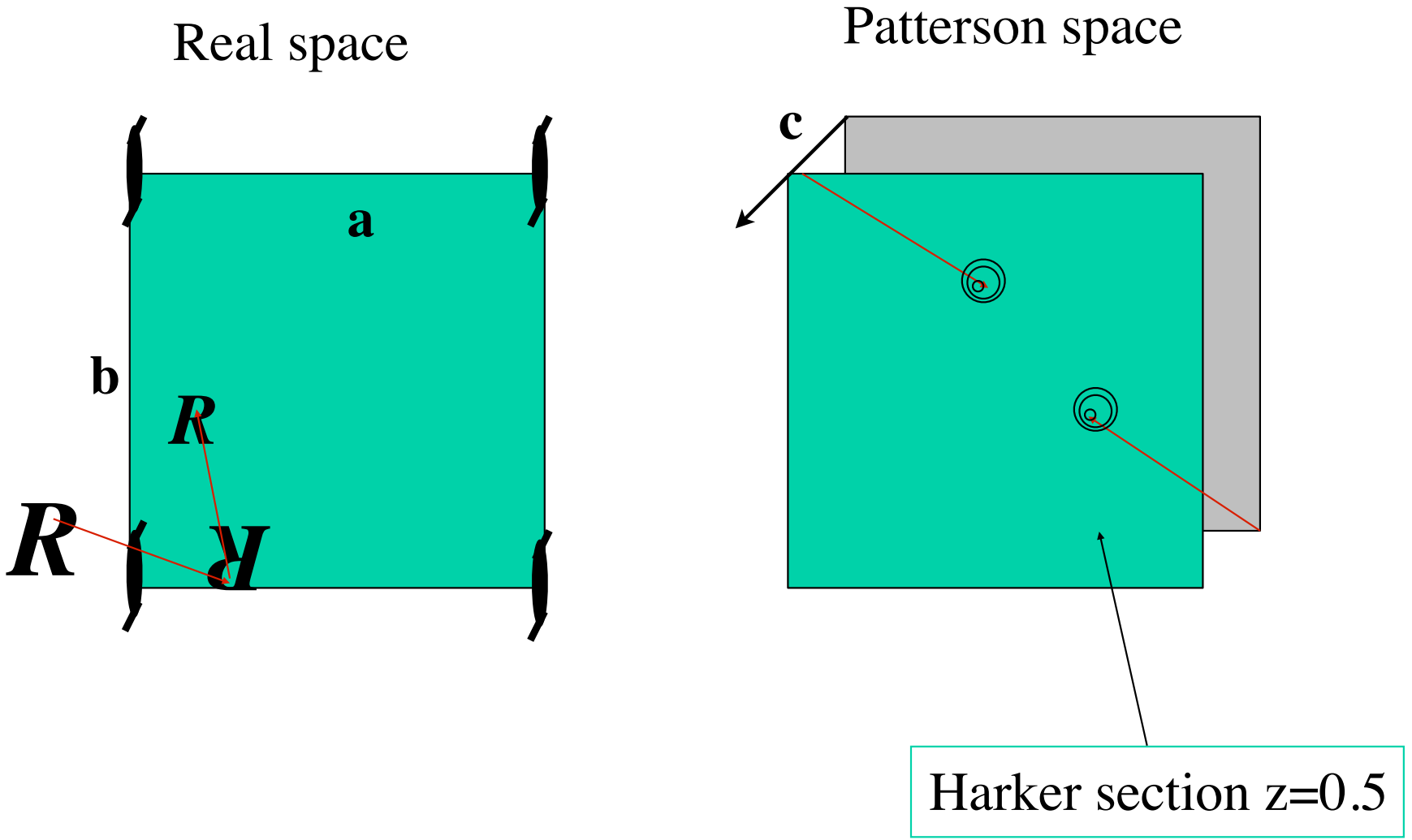


phase = 0

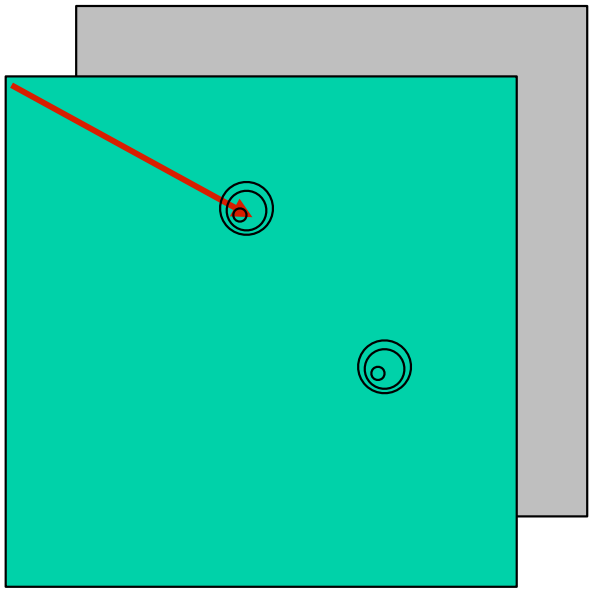
Using observed amplitudes, but setting all phases to 0 creates a centro-symmetric image of the molecule.

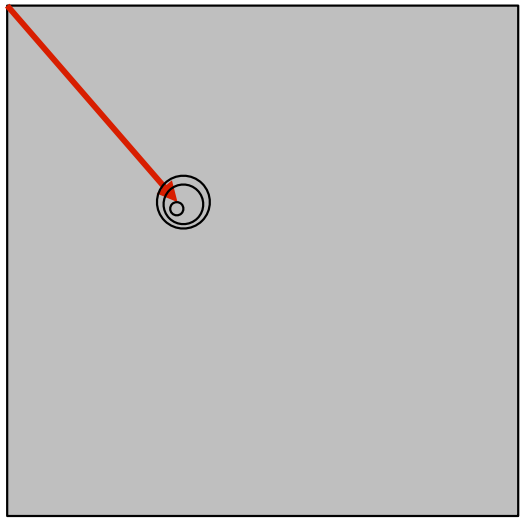
Patterson peaks generated by symmetry operations found are on *Harker sections*

$P2_1$

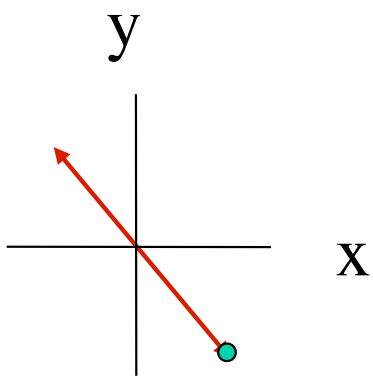


Harker sections tell us the location of atoms relative to the cell axes





If operator is a 2-fold, divide this vector by two to get the XY coordinates.



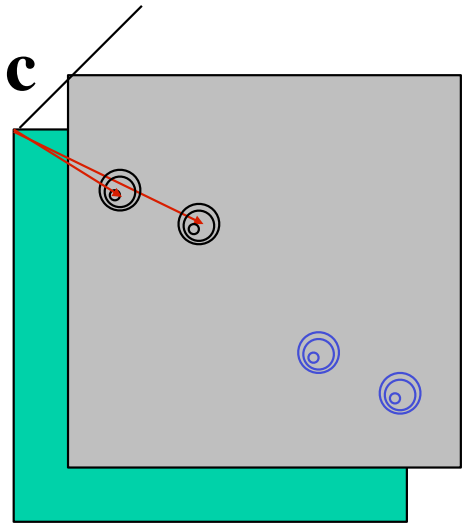
(The Z position is found on other sections)

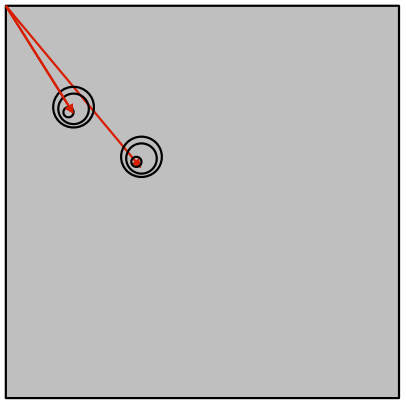
Non-Harker sections tell us inter-atomic vectors not related by symmetry

If there is more than one atom in the *asu*, you can get the vector between them by searching for peaks in non-Harker sections of the Patterson. (like the glycine example)

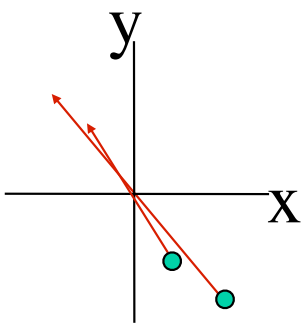
Then, combining knowledge from Harker sections (giving absolute positions) and non-Harker sections (giving relative positions) we can get the atomic coordinates.

Simple case: 2 atoms, $P2_1$

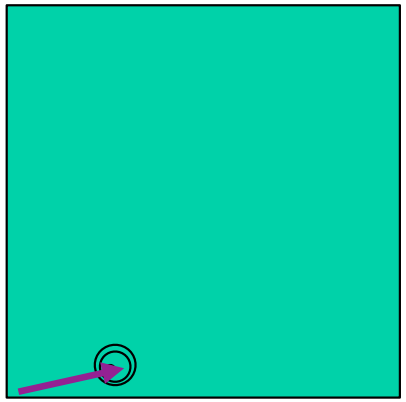


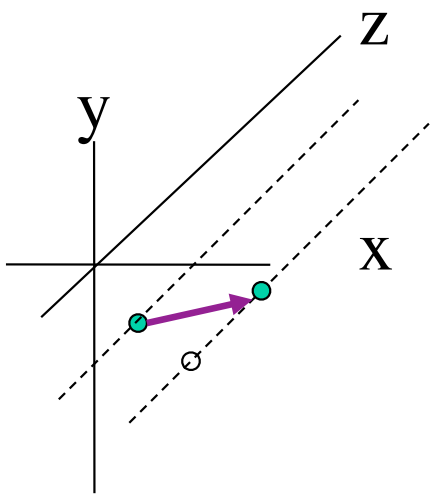


In a Harker section



The xy-position is found relative to the 2-fold axis, for each atom



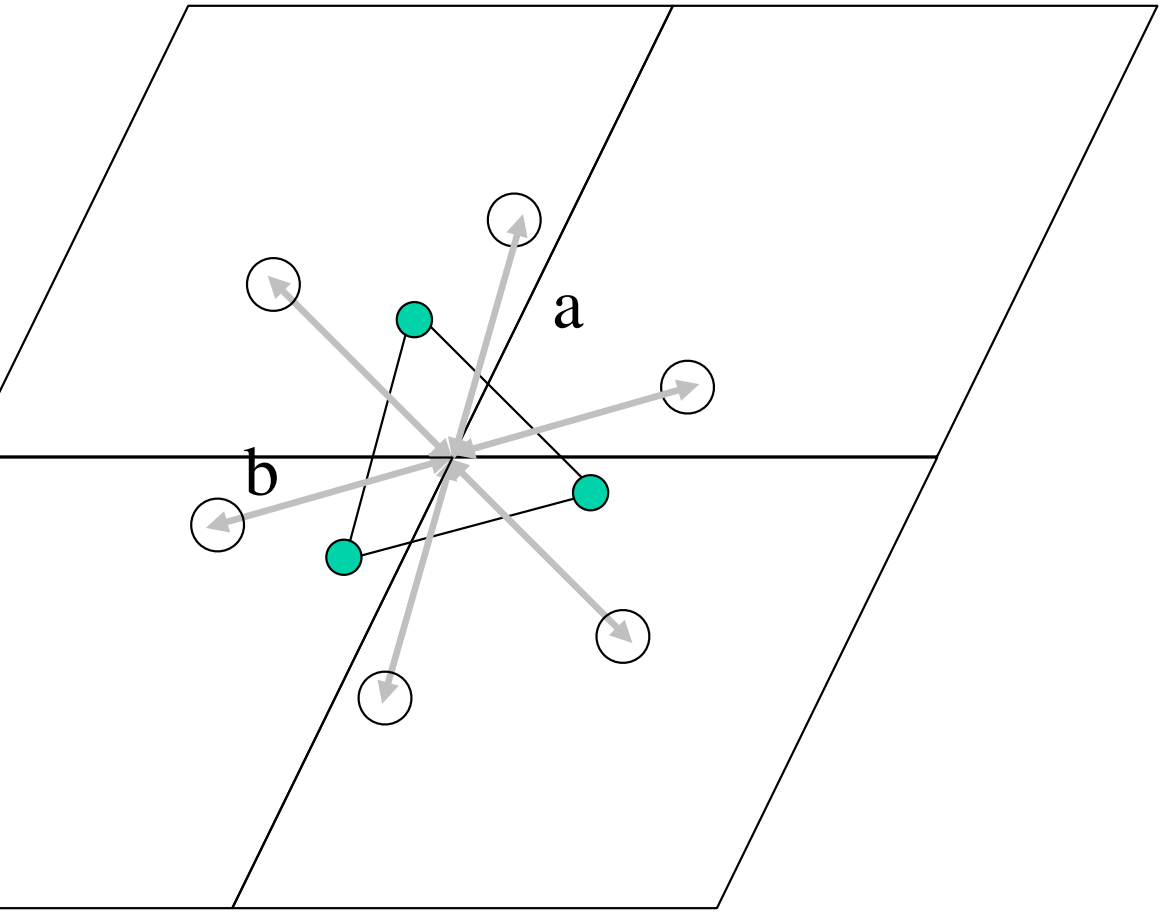


In a non-Harker section

The relative Z position is found for one atom relative to the other.

- A Patterson map can be calculated without needing to know the phases.
- A Patterson map shows the vectors between heavy atoms.
- By considering symmetry, we can locate the heavy atoms, sometimes uniquely.
- If we know where the heavy atoms are, then we can calculate the scattering factors F_H
- If we know F_H, then we can calculate the phases.

Solving a heavy atom Patterson



Patterson peaks are large circles.

Space group is $P3_1$

Where are the 3 heavy atoms?

(1) Draw a trigonal unit cell

(2) Heavy atoms are related by a 3-fold screw, so... draw an equilateral triangle around the origin such that side are Patterson peaks.

(3) Estimate the coordinates of the triangle vertices.

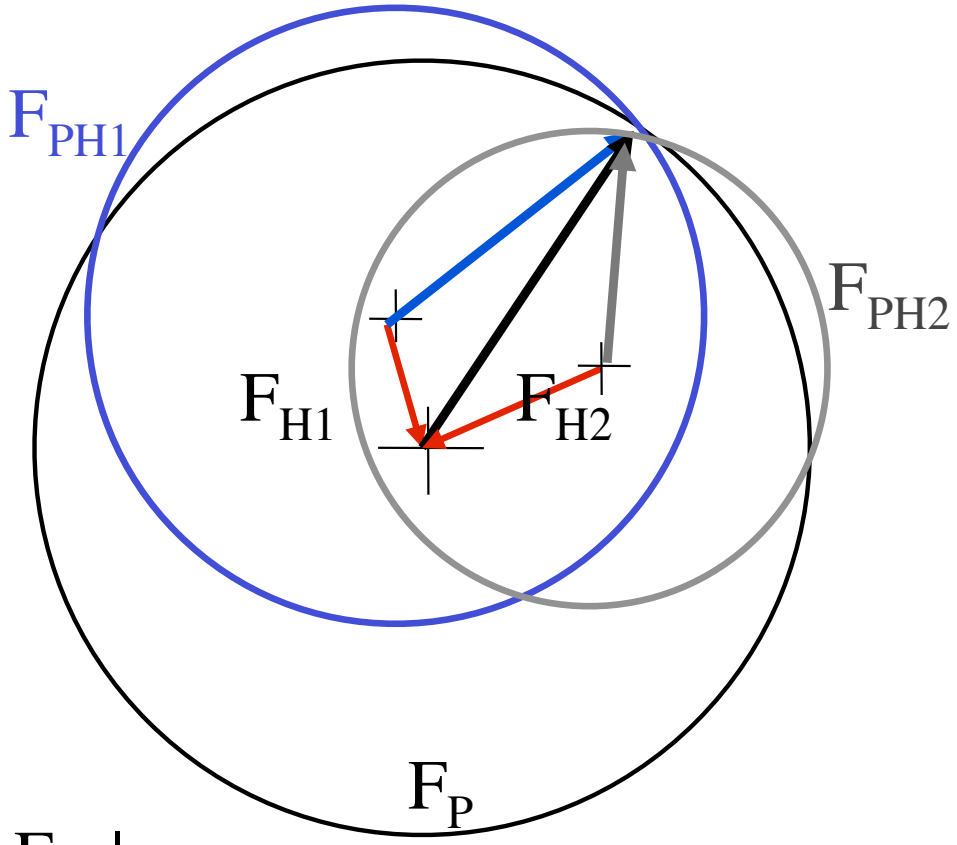
Harker peaks are vectors between symmetry-related heavy atoms.

$$P = v - \underline{M}v$$

Finding and using symmetry operators

- Google International Tables of Crystallography (http://it.iucr.org/)
- Volume A
- Choose space group.
- Coordinates.

Two heavy atom derivatives (MIR), unambigous phases



$|F_P + F_{H1}| = |F_{PH1}|$

$|F_P + F_{H2}| = |F_{PH2}|$

Or

$|F_P| = |F_{PH1} - F_{H1}| = |F_{PH2} - F_{H2}|$

Exercise 4: Solve the phase problem for one F using two F_H's

$|F_P| = 29.0$

$|F_{PH1}| = 26.0$

$|F_{PH2}| = 32.0$

Draw three circles with the three radii (scale doesn't matter)

Offset the PH1 circle from the P circle by -FH1

Offset the PH2 circle from the P circle by -FH2

Find the intersection of the circles.

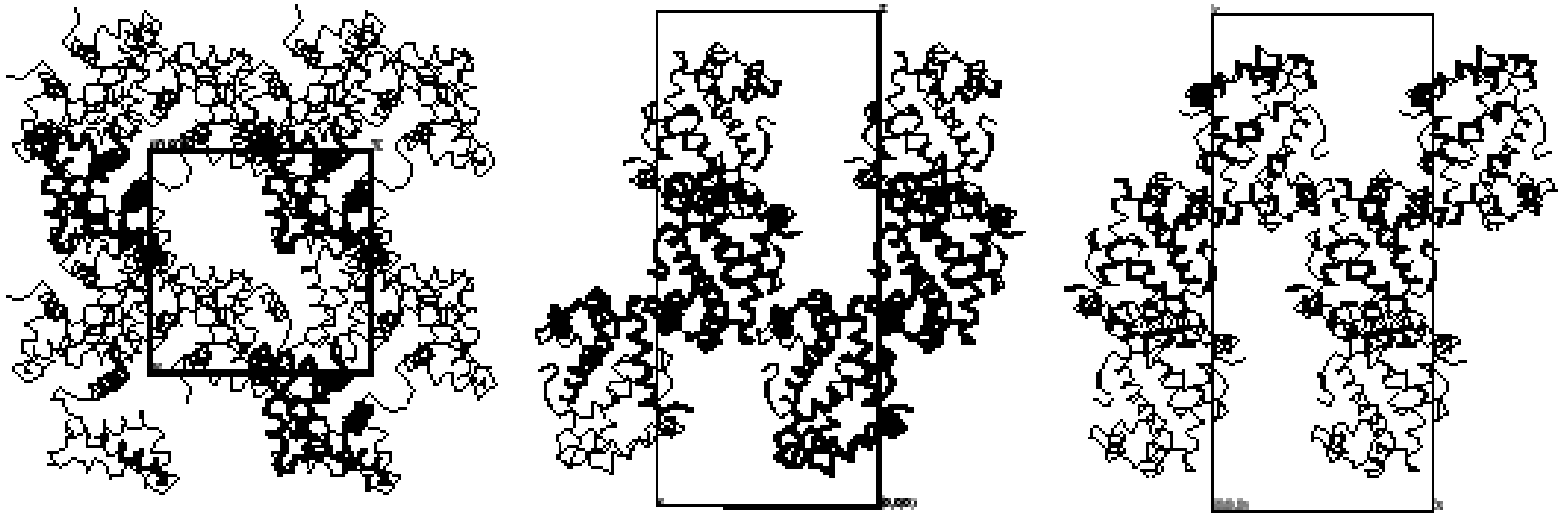
$F_{H1} = 7.8 \quad \alpha_{H1} = 155°$

$F_{H2} = 11.0 \quad \alpha_{H1} = 9°$

Additional topics...

- Crystal packing,
- centric reflections,

Crystal packing



Protein crystal packing interactions are salt-bridges and H-bonds mostly. These are much weaker than the hydrophobic interactions that hold proteins together. This means that (1) *protein crystals are fragile*, and (2) proteins in crystals are probably *not* significantly distorted from their native conformations.

The special usefulness of Centric reflections

•If the crystal has *centrosymmetric symmetry*, all reflections are **centric**, requiring phase = 0° or 180°

•If a non-centric space group has **2**-fold, **4**-fold or **6**-fold rotational symmetry, then the reflections in the ***0-plane*** **are centric**. (Because the projection of the density is centrosymmetric)

For **centric reflections**:

$|F_{ph}| = |F_p| \pm |F_h|$ ....is *exact**.

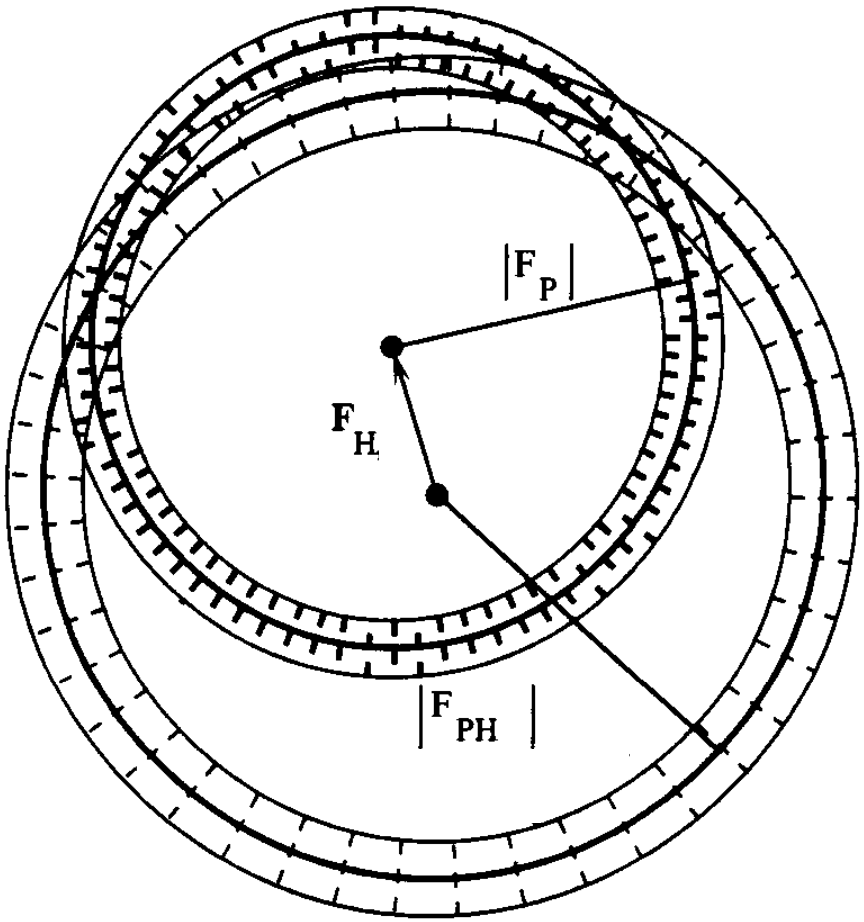
The ± is + if the phase of Fp and Fh are both 0, or both 180, otherwise -.

*assuming perfect scaling.

Initial phases

Phases are not measured exactly because amplitudes are not measured exactly.

Error bars on F_P and F_{PH} create a distribution of possible phase values α.



width of circle is 1σ deviation, derived from data collection statistics.

Review questions

1. What equation is represented by a Harker diagram?
2. Why is there more anomalous signal at high resolution?
3. What is MAD?
4. What are the three circles in a MAD Harker?
5. What are centric reflections?
6. How do we find the figure of merit (m)?
7. Where is the figure of merit applied in the reverse FT?
8. What gives rise to a phase distribution?

Review questions

9. What does homology mean, in terms of the structure?
10. Are crystals with the same cell dimensions and symmetry and the same contents isomorphous?
11. How to we compare the intramolecular peaks between two Pattersons?
12. What is the rotation function?
13. What is the translation function?
14. What is the R-factor?

Review questions

15. Given a symmetry operator, can you find the equivalent position of a point?
16. What units is Patterson space in?
17. What kind of symmetry does Patterson space always have?
18. What does a peak in Patterson space mean?
19. How is symmetry used to solve a real space position from a Patterson space position?
20. Where are the Harker sections in $P2_12_12_1$?
21. What is the equation for subtracting two data sets (F_P, F_{PH}) to get another data set (F_H)?
22. What does "solving" a Patterson mean?
23. Why can't we measure phases experimentally?
24. What is a Harker diagram? How do you solve it?