

# Phasing by Isomorphous Replacement

Bio5325

Spring 2006

# Phase Determination by the Heavy Atom Method

How do we obtain PHASE information?

How does heavy atom method/isomorphous replacement generate phase information by knowing the coordinates of heavy atoms in the unit cell?

How do we locate heavy atoms?

- Difference Patterson synthesis
- Difference Fourier synthesis

How do anomalous differences produce phase information?

How are phase estimates combined and improved?

# Phase of a Reflection

Familiar analogy: the polynomials  $1, x, x^2, x^3, \dots$  can closely approximate a function if we include the terms  $\alpha_0 1 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \dots$

Polynomials can be designed to form an ORTHONORMAL BASIS consistent with various boundary conditions.

For PERIODIC BOUNDARY conditions, the choice of basis functions are:

$$A_0 + A_1 \cos 1x + iB_1 \sin 1x + A_2 \cos 2x + iB_2 \sin 2x + \dots$$

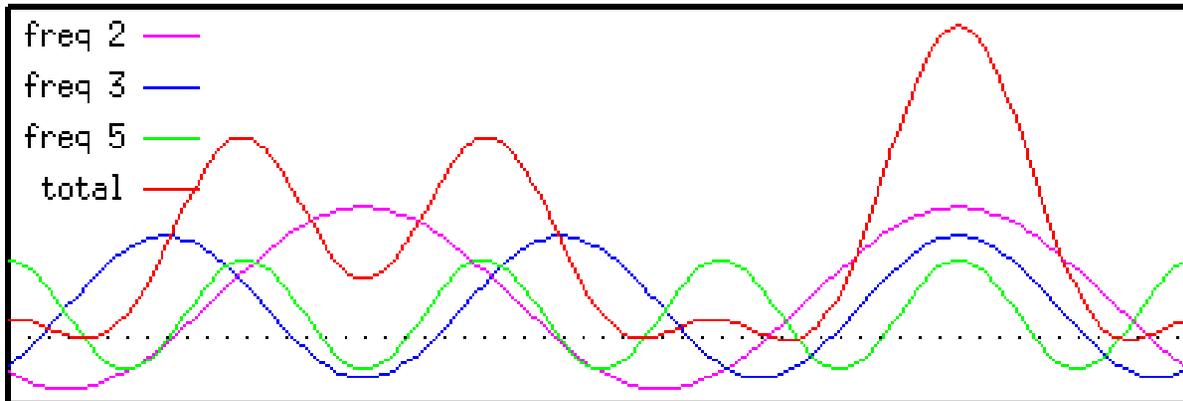
$$A_0 + |F_1| \cos \phi_1 \cos 1x + i|F_1| \sin \phi_1 \sin 1x + |F_2| \cos \phi_2 \cos 2x + i|F_2| \sin \phi_2 \sin 2x + \dots$$

$$A_0 e^{-2\pi i(0)x} + |F_1| e^{i\phi_1} e^{-2\pi i(1)x} + |F_2| e^{i\phi_2} e^{-2\pi i(2)x} + \dots$$

# Phase of a Reflection

$$A_0 e^{-2\pi i(0)x} + |F_1| e^{i\phi_1} e^{-2\pi i(1)x} + |F_2| e^{i\phi_2} e^{-2\pi i(2)x} + \dots$$

Add together waves ( $h = 0, 1, 2, \dots$ ) with amplitude  $|F_h|$  and displace each from origin by  $\phi_h$ .



# The Crystallographic Experiment

Purpose is to reconstruct electron density ( $\rho$ ), which is the 3-D image of the molecule (microscope lens analogy).

$$\rho(x) = \sum_h |F_h| e^{i\phi_h} e^{-2\pi i \vec{h} \cdot \vec{x}}$$

$|F_h|$  is obtained from  $\text{SQRT}(I_h)$ , but we're missing the phase information (offset from the origin). Must guess or estimate the phases.

How do we know the calculated phases are correct?

- Solvent boundary in electron density
- Protein-like features: beta strands and alpha helices
- Density consistent with known amino acid sequence
- Sensible clustering of hydrophobic residues in core of fold
- Model building and refinement leads to a sensible R-factor

# The Heavy Atom Method

Requires 2 or more complete data sets collected from ISOMORPHOUS crystals.

Heavy atom derivatives have a few well ordered metal binding sites. Resulting structure is the native protein plus a discrete number of metal atoms (soak in or incorporate as selenomethionine substitutions).

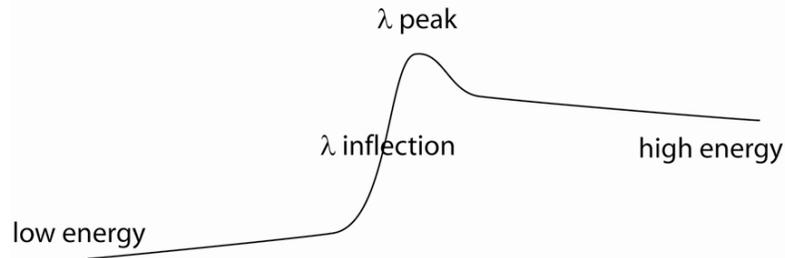
$\{|F_{nat}|\}$ ,  $\{|F_{ph1}|\}$ ,  $\{|F_{ph2}|\}$ , etc.

With heavy atoms of large Z-value, NONISOMORPHISM is limiting for accuracy of phases.

Anomalous scatterers cause differences in  $F_{hkl}$  and  $F_{\overline{hkl}}$  (Friedel pairs) attributable to a small group of atoms.

# MAD Phasing

Collection of anomalous scattering data at specific wavelengths where heavy atoms scatter strongly. This is a Multiwavelength Anomalous Diffraction experiment.



For anomalous scattering, isomorphism is perfect. However, anomalous signal is small and requires accurate intensity measurements.

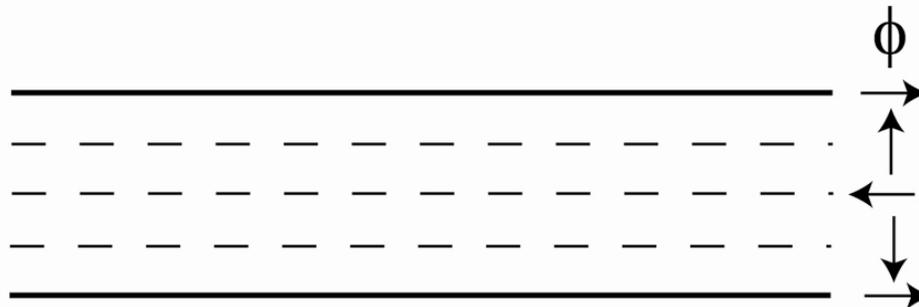
Anomalous signal increases with resolution, but diffraction intensity decreases, resulting in lower accuracy measurements at high angles of diffraction.

# How is Phase Info Obtained from $\{|F_{\text{nat}}|\}$ , $\{|F_{\text{ph}}|\}$ ?

Each reflection corresponds to a particular set of parallel planes of atoms that dissect the axes of the unit cell into an integer number of pieces ( $h, k, l$ ).

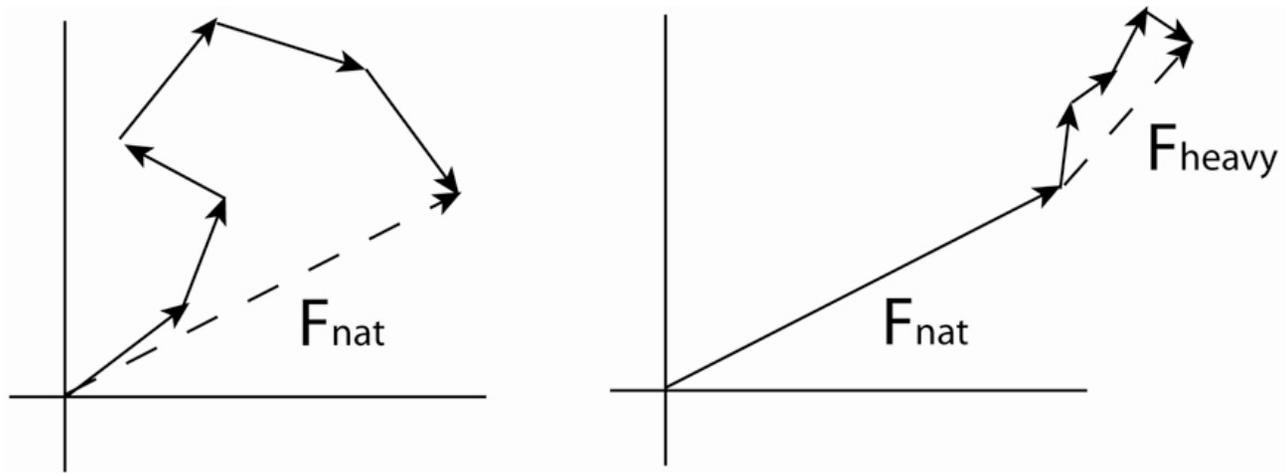
Atoms lying on the plane through the origin all contribute a scattering vector with a phase  $\phi = 0$ .

Out of plane atoms contribute scattering vectors with a phase offset:



# How is Phase Info Obtained from $\{|F_{\text{nat}}|\}$ , $\{|F_{\text{ph}}|\}$ ?

On the scale of the Fourier waves, the placement of atoms is approximately random, so the  $F_{\text{nat}}(h)$  is a weighted sum of a large number of random steps in the complex plane:



In a heavy atom derivative, the heavy atom contribution is a second random walk that adds to  $F_{\text{nat}}$ , representing a few sites where  $\rho_{\text{(PH)}} > \rho_{\text{(NAT)}}$ .

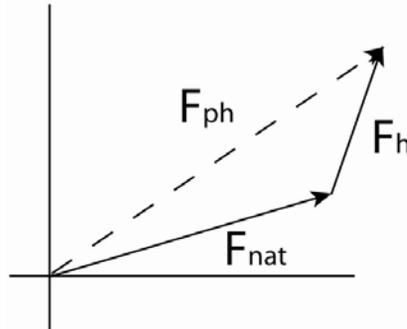


# Heavy Atom Phasing

Idea: If we could determine where the heavy atoms are located then we could calculate the Fourier transform of the heavy atom model (structure of just these atoms).

$$F_H(h) = F_{HEAVY}(h) = \sum_{j=\text{atoms}} f_j e^{-B_j S^2} e^{2\pi i \vec{h} \cdot \vec{x}}$$

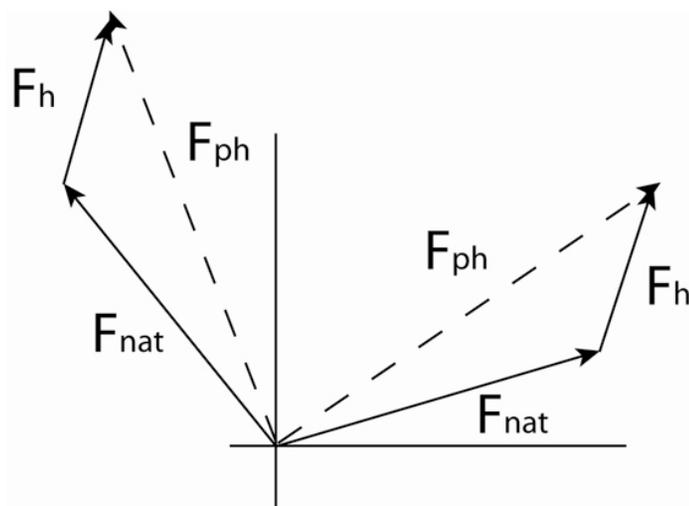
If  $|F_{\text{nat}}|$ ,  $|F_{\text{ph}}|$ , and  $|F_h|$  are known, there is 1 unique triangle having sides of those lengths (vector addition):



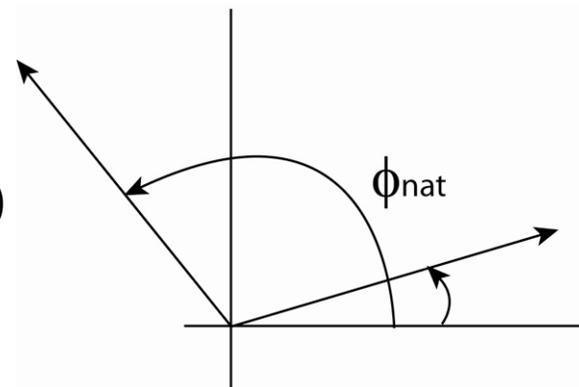
Triangle is flat for centric reflections ( $\phi = 0, 180$ )

# Heavy Atom Phasing

There are only 2 ways to orient the triangle so that its  $|F_h|$  side maintains the (known) phase of the heavy atom contribution  $\phi_h$ :



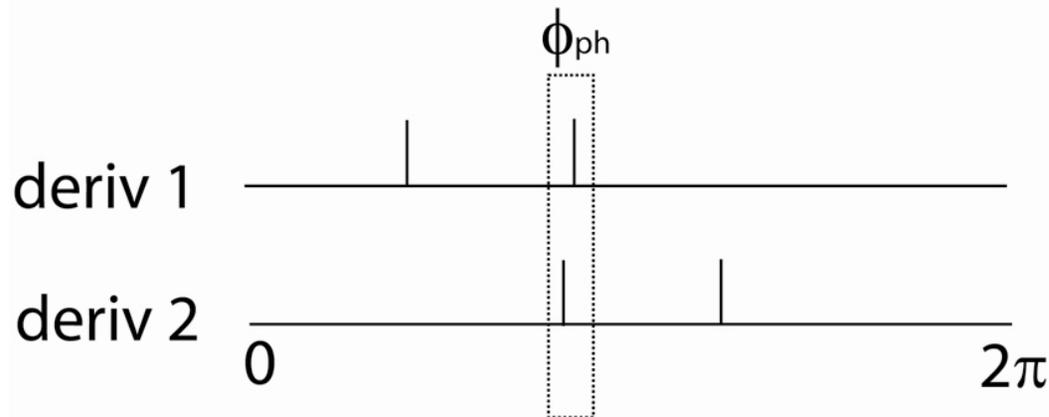
This gives 2 possibilities for the native phase ( $\phi_{nat}$ )



# Phase Improvement

For a single derivative (SIR) can take the centroid phase, which is average of the 2 possible phase choices.

For more than 1 derivative (MIR), might expect different pairs of 2-fold ambiguous choices for each derivative.



# Phase Improvement

**Density Modification:** change the calculated density in sensible ways then back transform (Fourier synthesis) to obtain modified (more accurate) phases that can be subsequently applied to observed  $F(h)$ 's to improve the electron density map.

- Add definition to boundary between protein and solvent, remove spurious density in solvent region.
- Modify density values assigned as protein envelope to reflect values typical of the %solvent and resolution.
- Calculate average density of multiple independent copies of the protein—apply noncrystallographic symmetry to superimpose molecules then calculate average values .

# Locating Heavy Atoms

**Patterson function:** The Patterson function  $P(u,v,w)$  is a convolution of the electron density at positions  $\mathbf{x}$  and  $(\mathbf{x}+\mathbf{u})$ , that is, separated by vector  $\mathbf{u}$ .

$$P(uvw) = \int_{\substack{\text{vol. of} \\ \text{unitcell}}} \rho(xyz)\rho(x+u, y+v, z+w) d v$$

The Patterson function corresponds to a Fourier series squared amplitudes (proportional to measured intensities), phases of zero, and a center of symmetry.

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}^2| e^{-2\pi i(hu+kv+lw)}$$

Electron density calculated with a Patterson function corresponds to a map of INTERATOMIC VECTORS all referenced to a common origin. For  $N$  atoms there are  $(N^2-N)$  interatomic vectors—Patterson space is crowded!

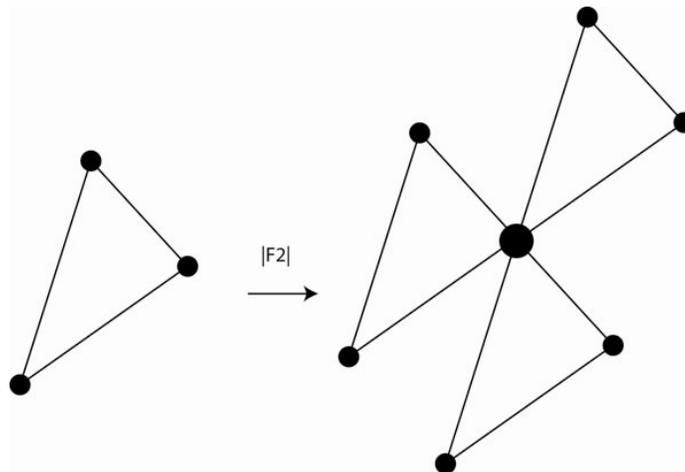
# Locating Heavy Atoms

Patterson vectors pile up (generate strong density) in peaks resulting from superposition of molecules by rotational symmetry. Peaks resulting from crystallographic symmetry are located on the **Harker sections** specified for each space group (except P1=Triclinic).

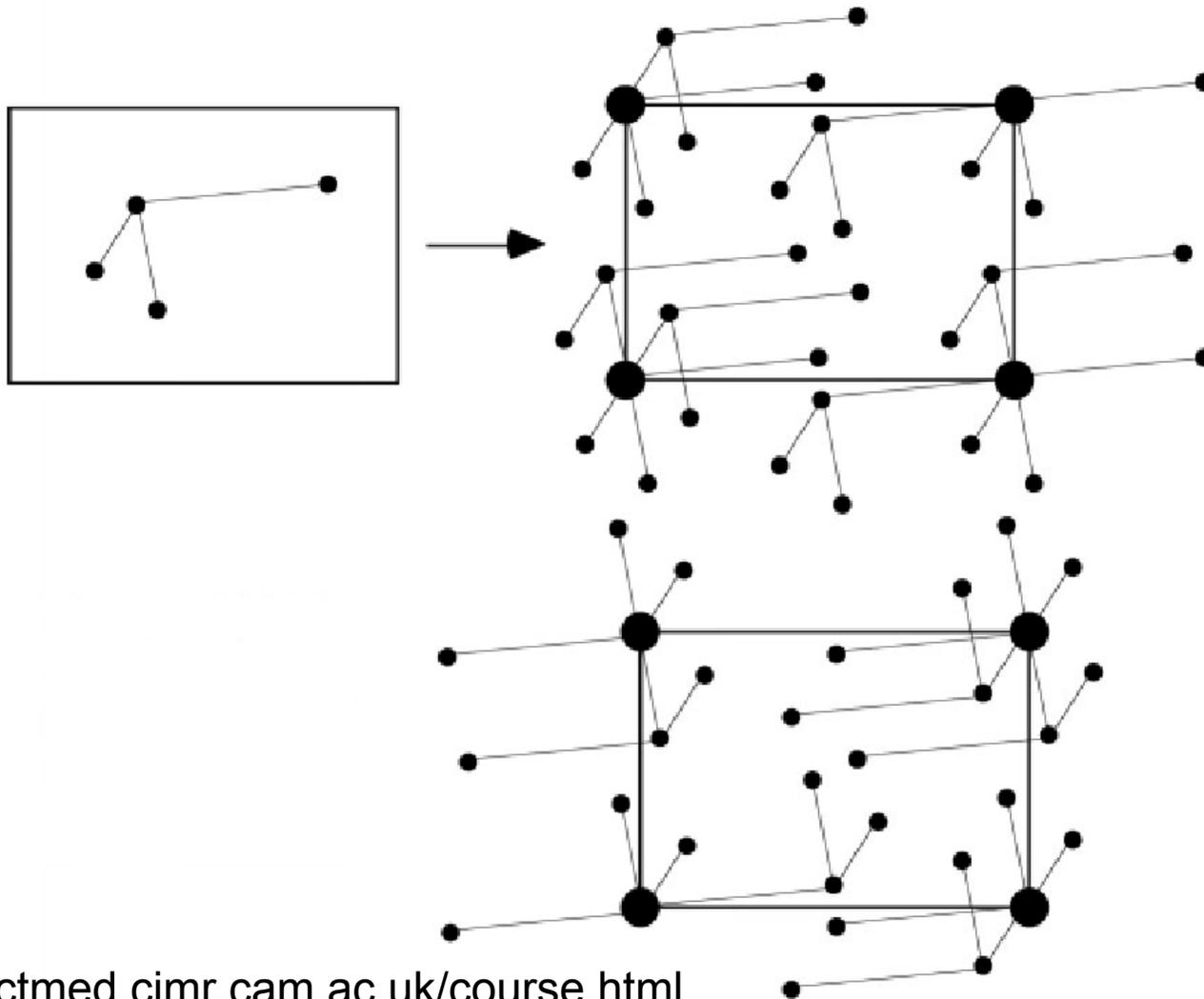
Patterson space is CENTROSYMMETRIC, reflecting the contributions of pairs of vectors ( $a \rightarrow b$ ,  $b \leftarrow a$ ) for all atoms.

Translational components of symmetry are absent in Patterson vector space. For example, space groups P2 and P2<sub>1</sub> have identical symmetry in Patterson space.

A 2D Patterson:

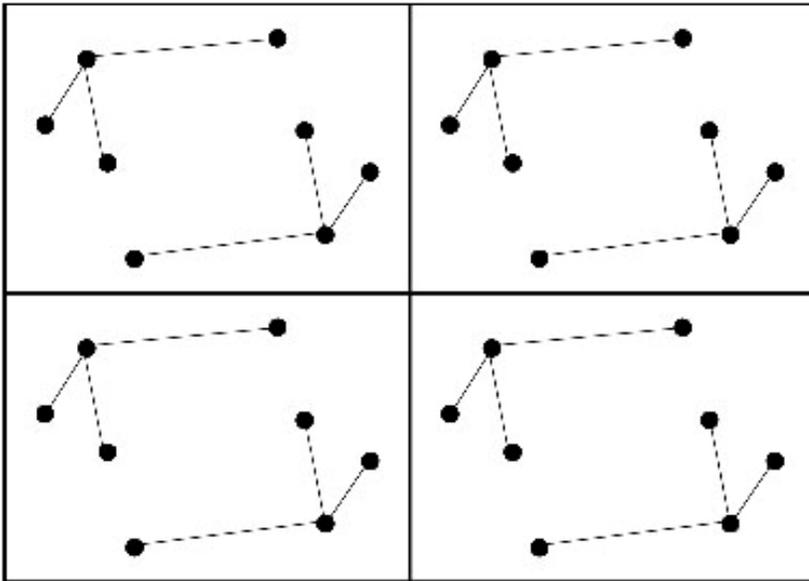


# Patterson Map of a Molecule

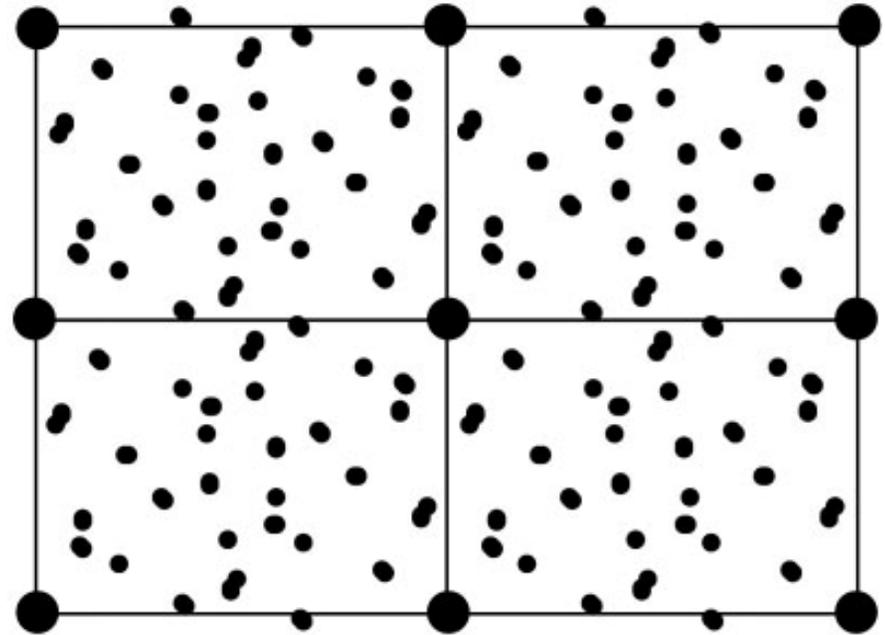


# Patterson Map of a Crystal

Crystal (real space)



Patterson function (vector space)



# Locating Heavy Atoms

To simplify the search for heavy atoms, we use a DIFFERENCE PATTERSON function with coefficients  $(\Delta F)^2 = (|F_p| - |F_{ph}|)^2$

The difference Patterson is a map of interatomic vectors with amplitudes contributed by the heavy atoms alone.

Crystal nonisomorphism and other systematic errors in the x-ray data will contribute to noise in the difference Patterson map.

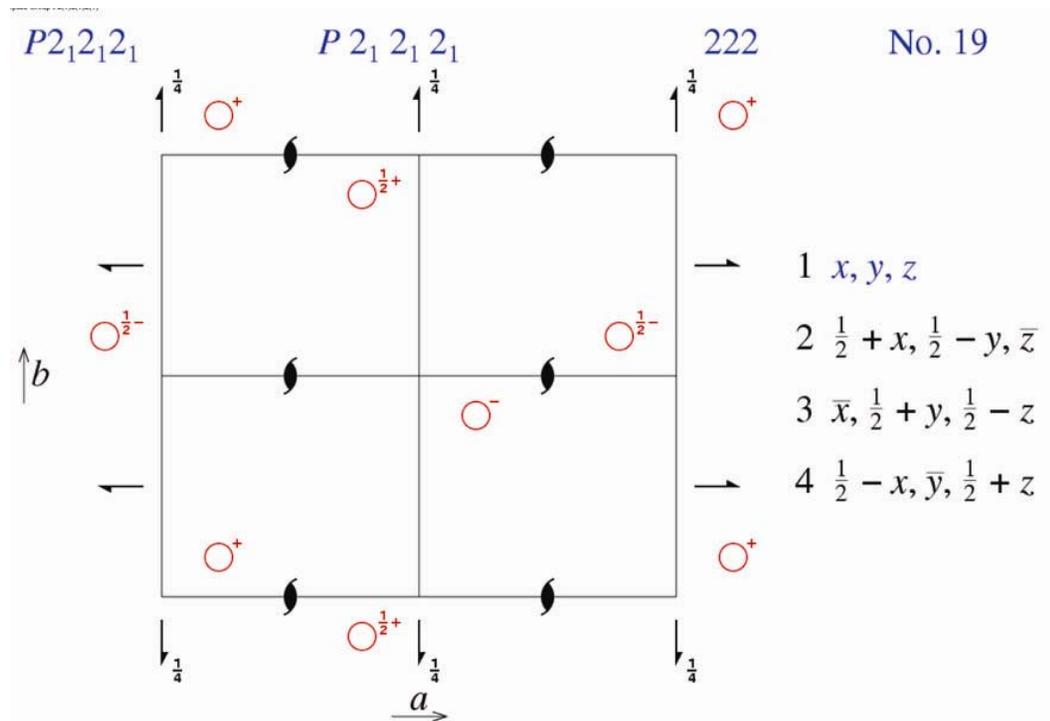
# Finding Heavy Atoms

A protein was crystallized in space group 19 ( $P2_12_12_1$ ) with the following symmetry operators:

1.  $x, y, z$
2.  $-x+1/2, -y, z+1/2$
3.  $x+1/2, -y+1/2, -z$
4.  $-x, y+1/2, -z+1/2$

Harker vector equations:

- 1.-2. =  $2x+1/2, 2y, 1/2$
- 1.-3. =  $1/2, 2y+1/2, 2z$
- 1.-4. =  $2x, 1/2, 2z+1/2$



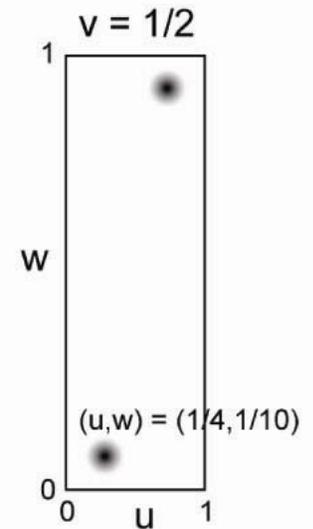
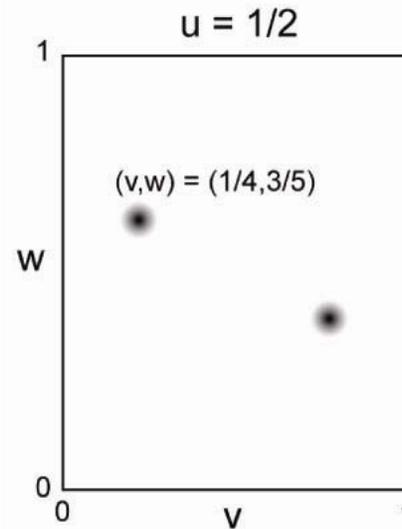
# Finding Heavy Atoms

A heavy atom derivative was prepared and :

1.  $x, y, z$
2.  $-x+1/2, -y, z+1/2$
3.  $x+1/2, -y+1/2, -z$
4.  $-x, y+1/2, -z+1/2$

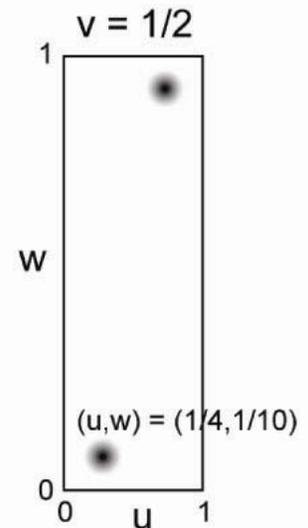
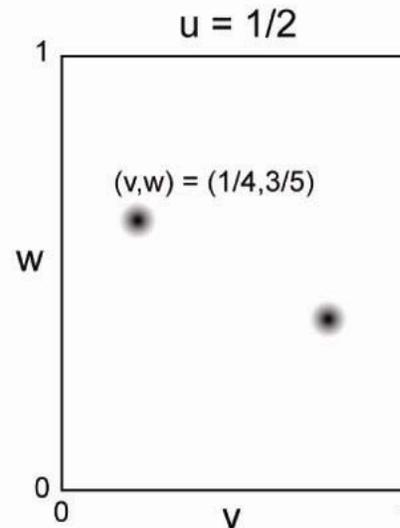
Harker vector equations:

- 1.-2. =  $2x-1/2, 2y, 1/2$
- 1.-3. =  $1/2, 2y-1/2, 2z$
- 1.-4. =  $2x, 1/2, 2z-1/2$



# Finding Heavy Atoms

What are the real space coordinates of the heavy atom(s)?



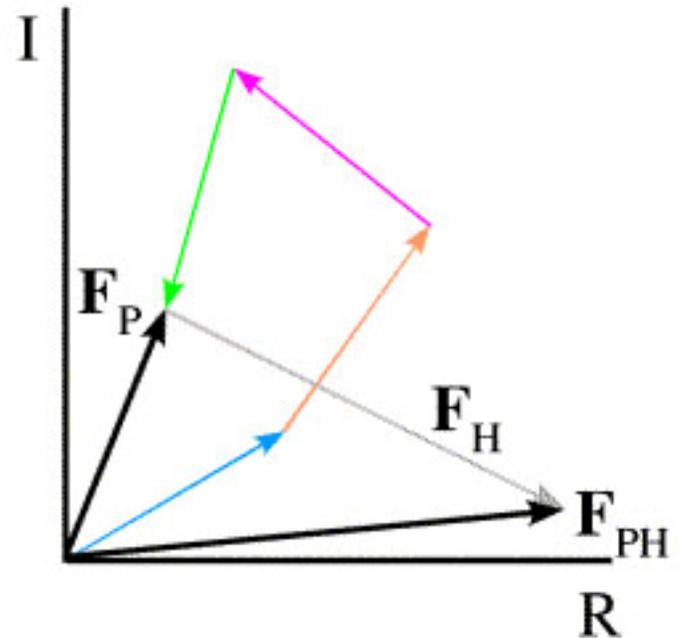
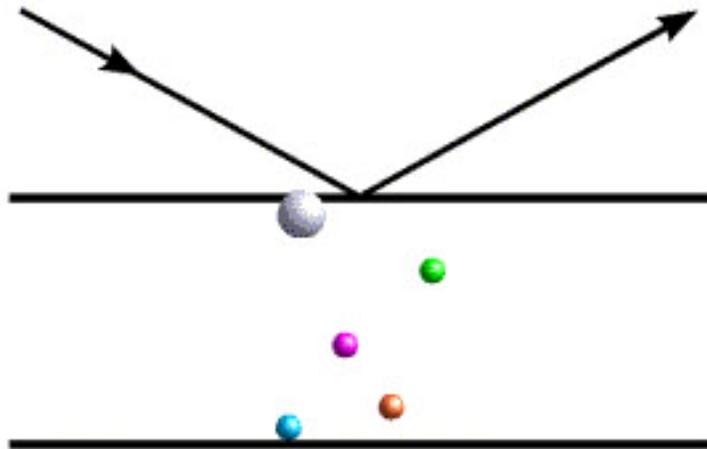
Harker vector equations:

$$(0.5, 0.25, 0.6) = 1/2, 2y-1/2, 2z ; \quad y = (+/-) 0.125 \quad z = (+/-) 0.3$$

$$(0.25, 0.5, 0.1) = 2x, 1/2, 2z-1/2 ; \quad x = (+/-) 0.125 \quad z = (+/-) 0.3$$

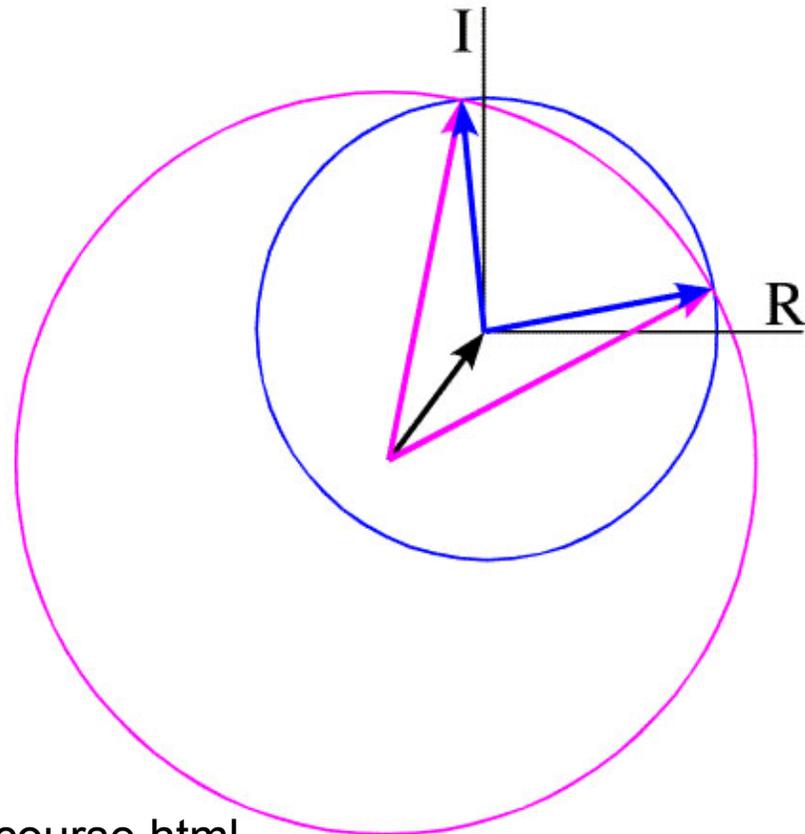
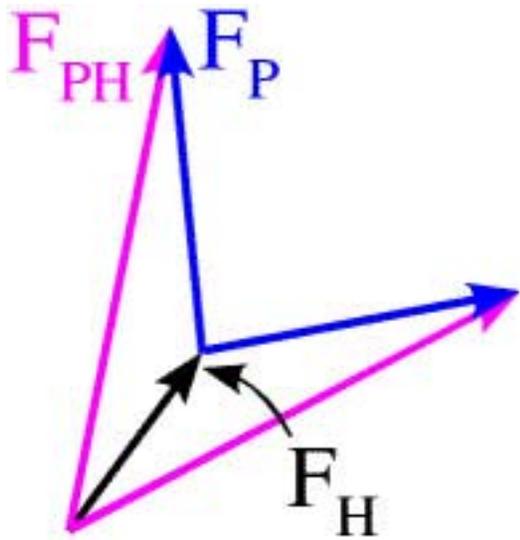
# Isomorphous Replacement

A heavy atom contributes disproportionately to the scattering recorded at each reflection:



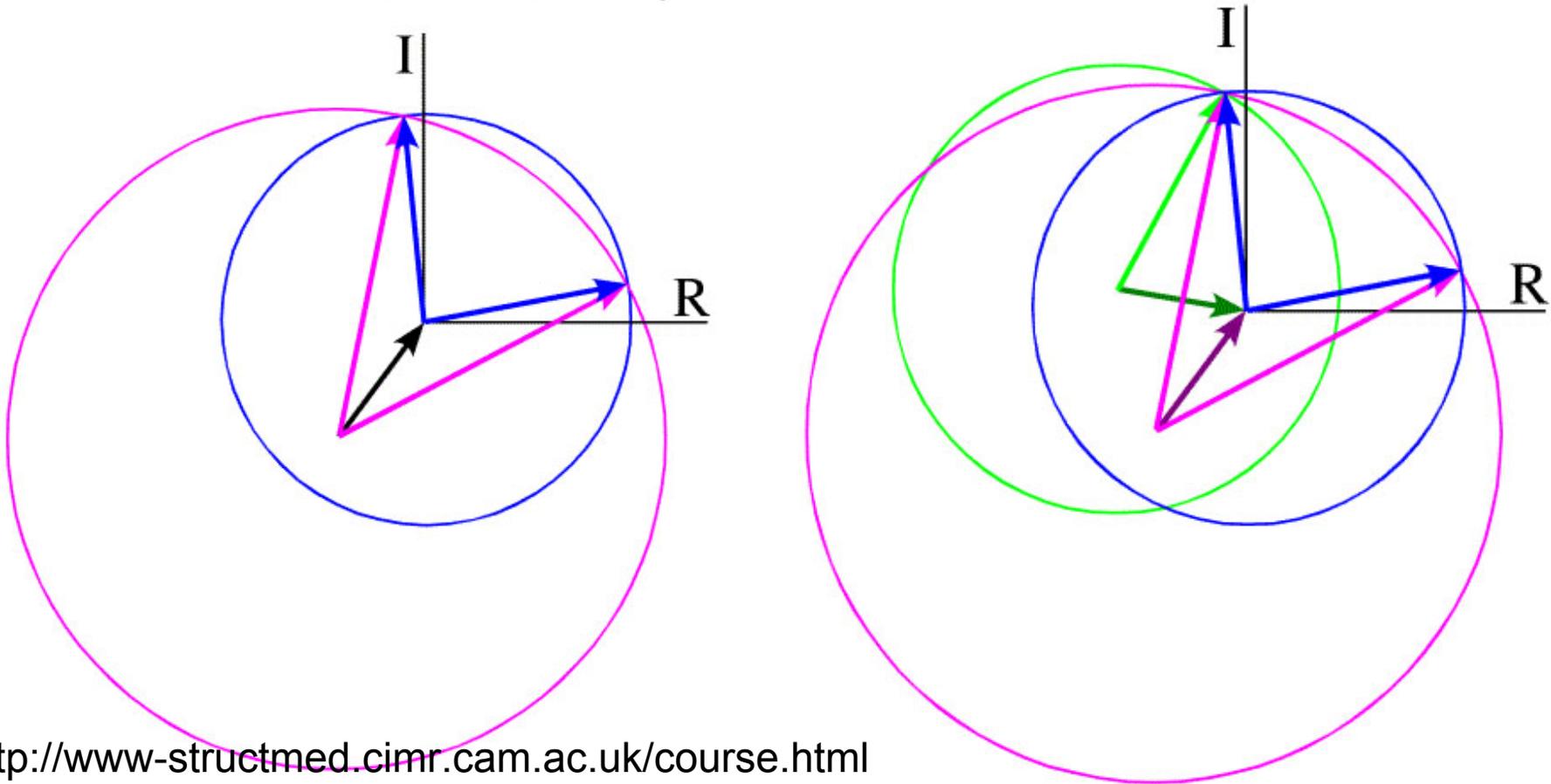
# Estimating Protein Phases

Once we locate the heavy atom(s), we know the length and orientation of  $F_H$ , and the lengths of  $F_P$  and  $F_{PH}$ :



# Multiple Isomorphous Replacement

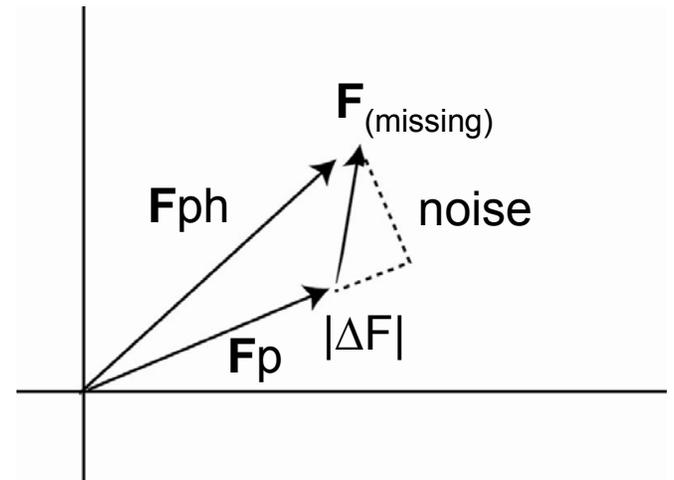
The 2-fold ambiguity in the phase of each reflection can be eliminated by comparing multiple heavy atom derivatives



# Difference Fourier Synthesis

$F_{(\text{missing})}$  represents atomic contributions from:

- Minor sites in derivative #1 that haven't been included
- Sites in derivative #2 that we need to find and position with same choice of origin as for derivative #1
- Other errors in the heavy atom model



$$\Delta F = \left( |F_{ph}| - |F_{nat}| \right) e^{i\phi_{nat}}$$

# Difference Fourier Synthesis

What we want is a map of the true heavy atom sites in derivative #2:

$$\Delta\rho(x)_{true} = \sum_h \left| F_{mis \sin g} \right| e^{i\phi_{mis \sin g}} e^{-2\pi i \vec{h} \cdot \vec{x}}$$

What we settle for is the projection of  $\mathbf{F}_{missing}$  onto  $\mathbf{F}_{nat}$ :

$$\Delta\rho(x)_{true} = \sum_h \left( \left| F_{ph} \right| - \left| F_{nat} \right| \right) e^{i\phi_{mis \sin g}} e^{-2\pi i \vec{h} \cdot \vec{x}}$$

# Working with Experimental Phases

Anomalous scattering is another source of phase information that we won't have time to discuss.

The phases calculated from each heavy atom derivative are improved by heavy atom parameter refinement ( $x, y, z$ , occupancy, and B-factor). We are refining the heavy atom model, taking into account the protein phases estimated from multiple sources.

Our phase estimates contain errors causing incomplete closure of the "phase triangle." The FIGURE OF MERIT corresponds to the cosine of the lack of closure error.

Typically, an experimentally phased electron density map is calculated with each reflection weighted according to its figure of merit.