Patterson Space and Heavy Atom Isomorphous Replacement

MIR = Multiple heavy atom Isomorphous Replacement phasing

SIR = Single heavy atom Isomorphous Replacement phasing
**Bragg planes**

- **Reflection**
  - synonymous with scattering vector
  - defines location of crystal planes
  - exists in reciprocal space
  - is wave of electron density
  - is part of structure factor

- **Structure factor**
  - determines variability
  - is measured in fractions of density var.
  - exists in reciprocal space
  - has reciprocal units of lattice vectors in reciprocal space
  - transforms into unit cell

- **Electron density**
  - is located at x y z
  - is part of Bragg planes
  - transforms into crystal planes

- **Crystal planes**
  - defines location of scattering vector
  - is located at a b c
  - exists in real space

- **Unit cell**
  - origins define index
  - exists in reciprocal space

- **Scattering vector**
  - defines location of Bragg planes
  - exists in reciprocal space
  - corresponds to a specific set of h k l (Miller indices)

- **Reciprocal space**
  - lattice vectors in
  - a* b* c*

**Concepts have structure**
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 $A e^{i\alpha}$ 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.
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 derivative 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.
Harker diagram method for discovering phase from amplitudes.

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

For the heavy atom $F_H$, we know only amplitude.

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

$|F_P + F_H| = |F_{PH}|$
Two heavy atom derivatives (MIR), 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 answer.

MIR = multiple isomorphous replacement, gives an unambiguous answer.
If we can locate the heavy atom, then we can calculate its contribution to $F_{PH}$

$$F_H(h) = \sum f_wh e^{i2\pi h \cdot r}$$

real-space locations of heavy atoms, $r$. 
Using Patterson maps to find heavy atoms
Amplitudes for a 2 atom crystal

The amplitude of a wave $F(h \ k \ l)$ scattered by just 2 atoms depends on the distance distance between them in the $(h \ k \ l)$ direction. Ignoring phase, what happens to the amplitude as the two atoms both slide the same distance, any direction? (nothing)

What happens if the two atoms move closer to each other or father apart? (it changes)
Largest amplitudes are when atoms are in-phase with the Bragg planes.

If two atoms are in-phase, they have the largest amplitude.

If both atoms are the same distance to the Bragg plane, amplitude is maximized.
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

Reverse FT with phases
\[ \rho(r) = \sum_{h} |F(h)| e^{i\alpha(h)} e^{-i2\pi h \cdot r} \]

Reverse transform without phases
\[ \rho(r) = \sum_{h} |F(h)| e^{-i2\pi h \cdot r} \]

This is the **Patterson function**

It uses the measured amplitude, **no phase**
The Patterson map is the centrosymmetric projection

Using observed amplitudes, but setting all phases to 0 creates a centro-symmetric image of the molecule.
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.
Patterson peaks generated by symmetry operations found are on *Harker sections*
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 \textit{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, \( \text{P}2_1 \)

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.
Do this at home: make a Patterson map

Using the Escher Web Sketch:
Set the space group to P4mm
Place one atom at (0.4, 0.25)
Draw the Patterson vectors.
Place a second atom (different color) at (0.2, 0.1)
Draw the Patterson vectors.
Solving a simple 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.
Two heavy atom derivatives (MIR), unambiguous phases

\[ |F_p + F_{H1}| = |F_{PH1}| \]
\[ |F_p + F_{H2}| = |F_{PH2}| \]

Or

\[ |F_p| = |F_{PH1} - F_{H1}| = |F_{PH2} - F_{H2}| \]
In class exercise: 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 diameters (scale doesn’t matter)

Offset the PH1 circle from the P circle by $-F_{H1}$

Offset the PH2 circle from the P circle by $-F_{H2}$

Find the intersection of the circles

$$F_{H1} = 7.8 \quad \alpha_{H1} = 155^\circ$$

$$F_{H2} = 11.0 \quad \alpha_{H1} = 9^\circ$$
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:

\[ |F_{ph}| = |F_p| \pm |F_h| \quad \text{....is exact*.} \]

The ± is + if the phase of \( F_p \) and \( F_h \) 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 $\alpha$.

width of circle is $1\sigma$ deviation, derived from data collection statistics.
Review

• What units is Patterson space in?
• What kind of symmetry does Patterson space always have?
• What does a peak in Patterson space mean?
• How is symmetry used to solve a real space position from a Patterson space position?
• Where are the Harker sections in P2_12_12_1?
• What is the equation for subtracting two data sets (F_P, F_PH) to get another data set (F_H)?
• What does “solving” a Patterson mean?
• Why can’t we measure phases experimentally?
• What is a Harker diagram? How do you solve it?