Phases may be calculated given a known homolog structure.
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 not 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 \textit{orientation}

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

\[
\begin{align*}
  x' &= c_{11}x + c_{21}y + c_{31}z + v_x \\
  y' &= c_{12}x + c_{22}y + c_{32}z + v_y \\
  z' &= c_{13}x + c_{23}y + c_{33}z + v_z
\end{align*}
\]

or \( x' = Cx + v \)

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.
Procedure for molecular replacement:

1) Calculate Search Model diffraction data.
2) Calculate the Search Model Patterson map. $P_c$
3) Calculate the observed Patterson map.
4) Rotate one Patterson versus the other, calculating correlation.
5) Rotate coordinates.
6) Re-calculate diffraction data. This time use observed cell dimensions.
7) Translate the rotated Search Model to every position in the unit cell. Calculate the R-factor. Choose the lowest R-factor.
8) Translate the coordinate to the position with the lowest R-factor.

Done. Next step: Refinement.
Calculate Search Model
diffraction data

The search model $|F_c|$'s are used to calculate a Patterson map.

A large P1 unit cell is used because then the Patterson map (the part close to the origin) will have only intramolecular peaks.

P1=no symmetry, not necessarily the same cell dimensions as Fobs
Superposing Patterson maps to solve rotation function

Only the part of the Patterson map within the mask is used.

Ignore short vectors. They contain little shape info.

Ignore long vectors. They are all intermolecular, we want intramolecular.
Rotation Function

Target structure

Search model

Target Patterson

Search model Patterson

rotating Search model Patterson

Matches target Patterson

[C]
Intramolecular not Intermolecular
Intramolecular not Intermolecular

intramolecular vectors rotate around the origin

intermolecular vectors are transformed differently

intermolecular vectors
The Rotation Function

Three angles ($\alpha, \beta, \gamma$) define all possible rigidbody rotations. The solution of the rotation function are the angles that give the highest *Patterson correlation function*.
The correlation between any two functions \( x \) and \( y \) is defined as:

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

\( x \)-bar means the average 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

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

The sums are generally done over \( v \) in a spherical shell of the Patterson map that excludes the huge self-peak \( (v < 4\text{Å}) \) and also excludes long (mostly intermolecular) vectors \( (v > 20\text{Å}) \).

So, \( 4\text{Å} \leq |v| \leq 20\text{Å} \), is a good range for the rotation 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 theasu exists for which:

\[
\rho(r) = \rho(M_{ncs}r + v_{ncs})
\]
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?
(4) The Translation Function

Symmetry related positions for each atom are calculated as follows:

\[ x' = Mx + v \quad (M \text{ is the sym-op matrix and } v \text{ is the sym-op vector}) \]

A translation of the coordinates is:

\[ x' = x + t \]

Symmetry-related, translated coordinates are:

\[ x' = M(x + t) + v = Mx + Mt + v \]
What happens to the phases and amplitudes when we translate?

Amplitudes don’t change.

Phases change depending on the dot product of the translation vector and the scattering vector $S$ (alias $hkl$)

New phase = old phase + $2\pi(hv_a + kv_b + lv_c)$

note: as usual, $\nu$ is in fractional coordinates
Reciprocal space symmetry is transposed

Rotating atoms in real space,

\[ r' = Mr \]

then multiplying by \( h \) to get the phase,

\[ \text{phase} = 2\pi(h \cdot (Mr)) \]

is the same as rotating reciprocal space \textit{the other way}.

\[ h \cdot (Mr) = 2\pi(M^T h \cdot r) \]

You can prove this by writing out the matrix multiplication.
R-factor: How good is the model?

Calculate $F_{\text{calc}}$’s based on the model.

Compute R-factor

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

Depending on the space group, an R-factor of $\sim 55\%$ would be attainable by scaled random data.

The R-factor must be $< \sim 50\%$.

**Note:** It is possible to get a high R-factor for a **correct model**. What kind of mistake would do this?
1. Molecular replacement is the solution of the problem \( r' = Mr + v \) where \( r \) are the model coordinates (from a homolog model) and \( r' \) are the true crystallographic coordinates.

2. The rotation function finds the rotation matrix \( M \).

3. The translation function finds the translation vector \( v \).

4. The rotation function is done in Patterson space.

5. The translation function can be done in reciprocal space because \( F_{\text{calc}} \) can be computed from \( F_{\text{mod}} \) and symmetry.
Model building from a map

(1) The map is calculated using $\alpha_{\text{best}}$.

(2) The map is contoured and displayed.

(3) A “trace” is attempted.
Class exercise:

Tracing an electron density map

sequence: AGDLLEHEIFGMPPAGGA

Can you locate the density above in the sequence?
What can you do if the phases are not good enough?

1. Collect more heavy atom derivative data
2. Try density modification techniques.

Density modification:

- Initial phases
  - Fo’s and (new) phases
  - Map
  - Modified map
  - Fc’s and new phases
Density modification techniques

Solvent Flattening: Make the water part flat.

1. Draw envelope around protein part
2. Set solvent $\rho$ to $<\rho>$ and back transform.
Solvent flattening

Requires that the *protein part* can be distinguished from the *solvent part*. BC Wang’s method: Smooth the map using a 10Å Guassian. Then take the top X% of the map, where X is calculated from the crystal density.
Skeletonization

1. Calculate map.
2. Skeletonize it (draw ridge lines)
3. Prune skeleton so that it is “protein-like”
4. Back transform the skeleton to get new phases.

Protein-like means: (a) no cycles, (b) no islands
Non-crystallographic symmetry

If there are two molecules in the ASU, there is a matrix and vector that rotate one to the other: $\mathbf{M} \mathbf{r}_1 + \mathbf{v} = \mathbf{r}_2$

1. Using Patterson Correlation Function, find $\mathbf{M}$ and $\mathbf{v}$.
2. Calculate map.
3. Set $\rho(r_1)$ and $\rho(r_2)$ to $(\rho(r_1) + \rho(r_2))/2$
4. Back transform to get new phases.
In class: Phenix Lab

• go to Xray Lab instructions.