

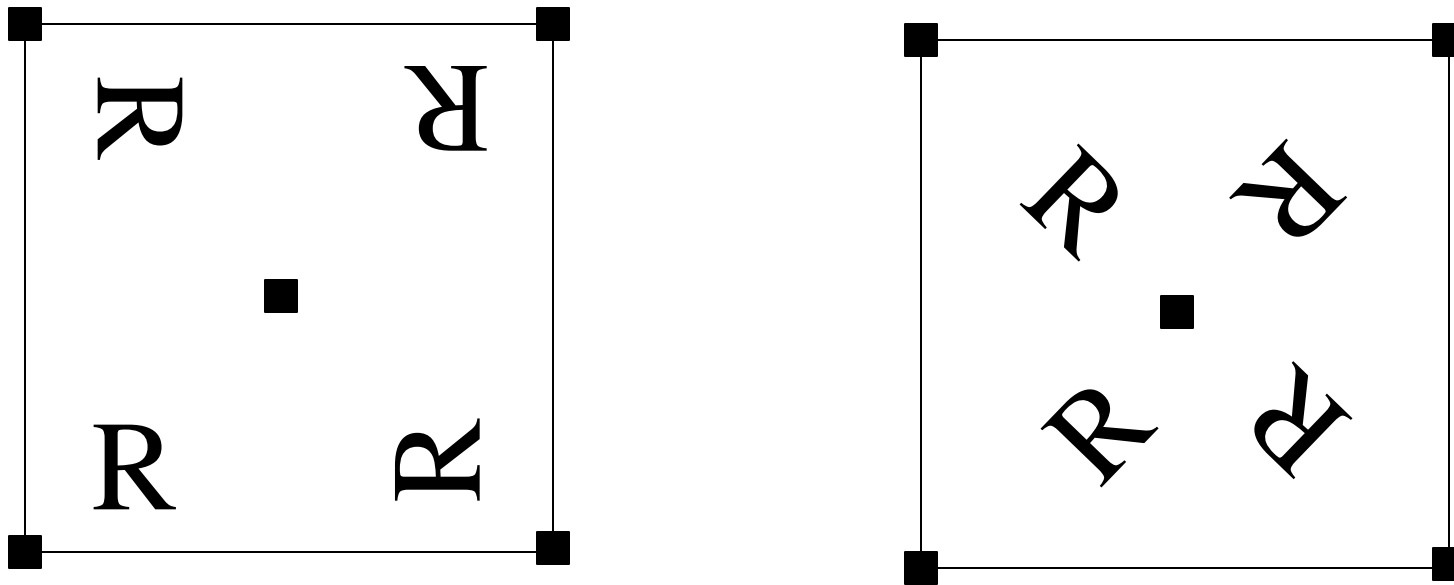
# Protein Structure Determination- Lecture 9 Molecular Replacement

Phases may be calculated given a known 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.



The diffraction patterns of these two crystals are not the same.

# We can use *homology* to infer structure

Protein sequences tell us whether or not the protein structures are likely to be the same. **If the sequence similarity is > 25%, then we say the sequences are "homologous", meaning they evolved from the same common ancestor, and they therefore must have similar structures.**

*How similar is not known* until both structures are solved.

Molecular replacement will not work if the structures are too different. It is used to solve structures of “close homologs” or even the same molecule in a non-isomorphous crystal.

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

# 6-dimensional search space

Every possible rigidbody transformation of a molecule can be described using 6 parameters. 3 angles of rotation (defining a matrix of 9 coefficients), and a vector of translation (3 values). i.e.

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

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

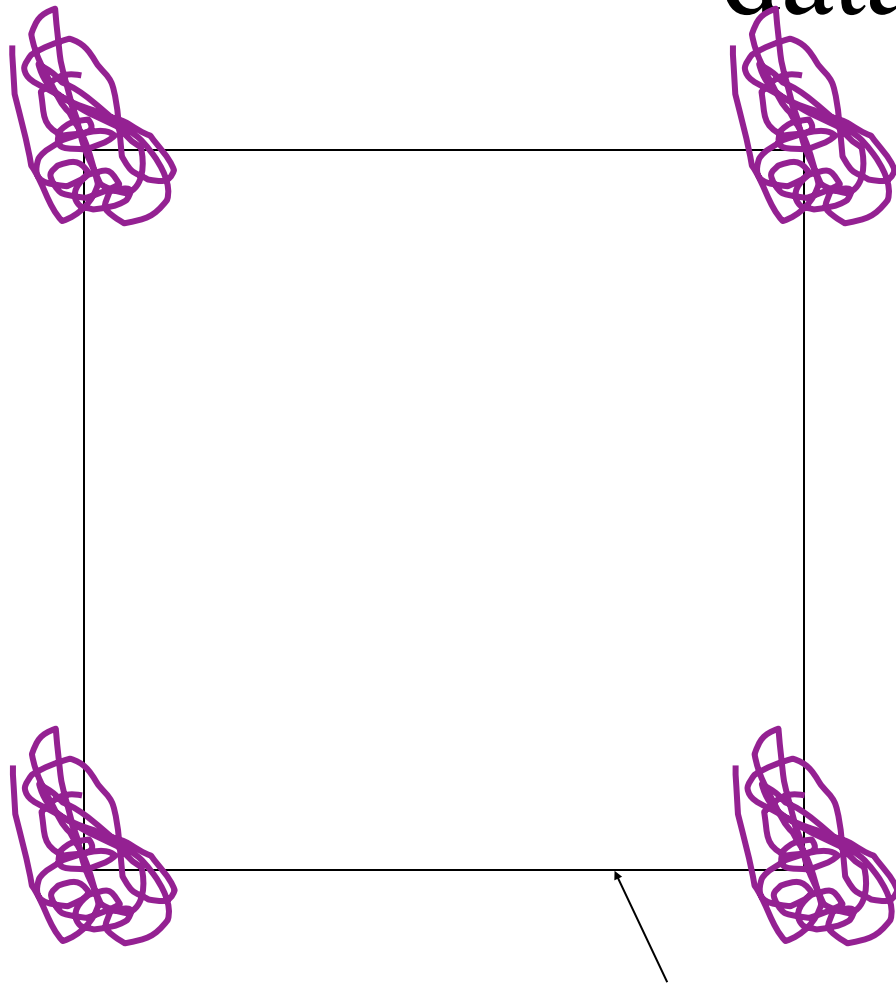
$$z' = c_{13}x + c_{23}y + c_{33}z + v_z$$

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 fake diffraction data for the model, using a large P1 unit cell.  $F_c$
- (2) Calculate the Patterson map.  $P_c$
- (3) Calculate the Patterson map of the observed crystal data ( $P_{obs}$  or  $P_o$ ).
- (4) Rotate one Patterson versus the other. Find the rotation with the maximum correlation.
- (5) Re-calculate structure factors ( $F_{model}$ ) for a P1 unit cell of the same cell dimensions as the true crystal unit cell (isomorphous except for symmetry).
- (6) Translate the P1 cell origin to every position in the unit cell, then sum the 'syms' (hkl's related by space group symmetry) to get the  $F_{calc}$ 's. Calculate the R-factor between  $F_{obs}$  and  $F_{calc}$ .
- (7) The position with the lowest R-factor (if < 50%) is the solution.

# (1) Calculate Fake diffraction data

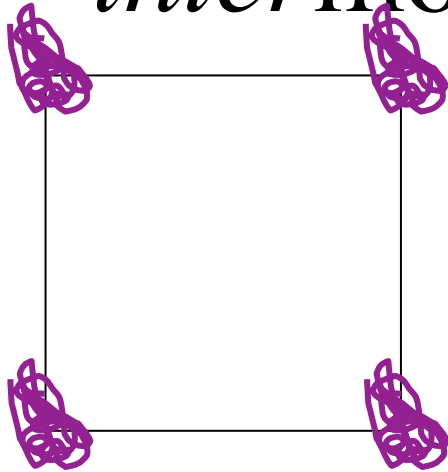


The  $|F_c|$ 's are used to calculate the Patterson map.

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

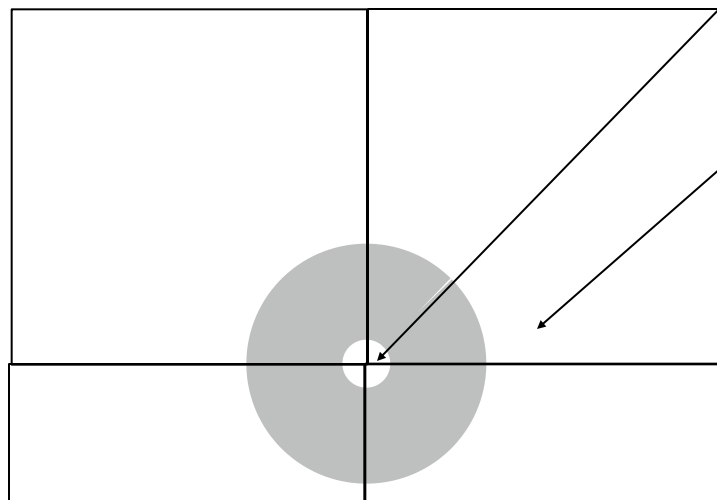
# *Intramolecular versus intermolecular Patterson peaks*



Only the part of the Patterson map within the shaded region is used.

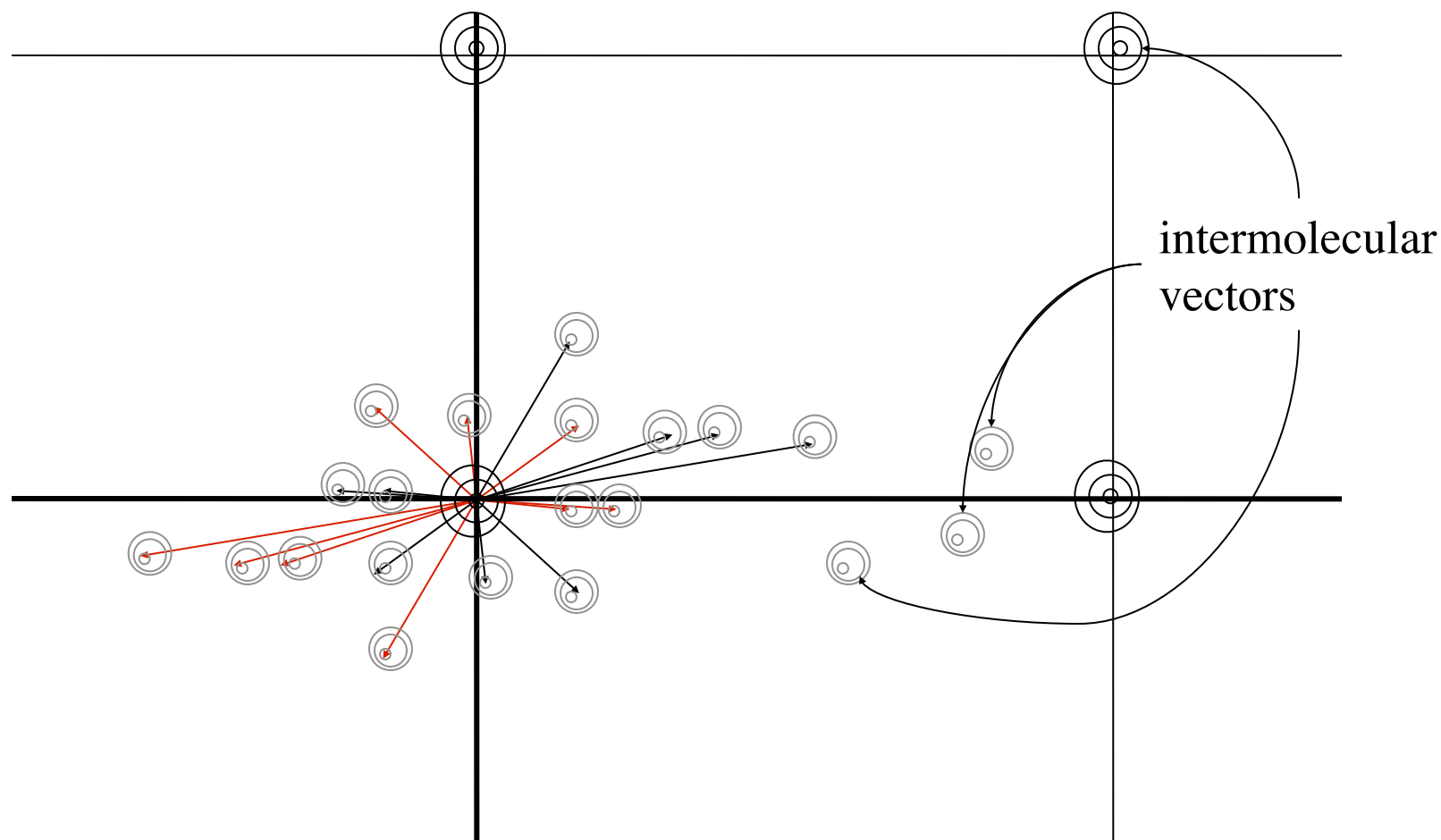
Short vectors are blocked out because they contain little shape info.

Long vectors are blocked out because they are all intermolecular (symmetry) and therefore depend on both orientation and translation.

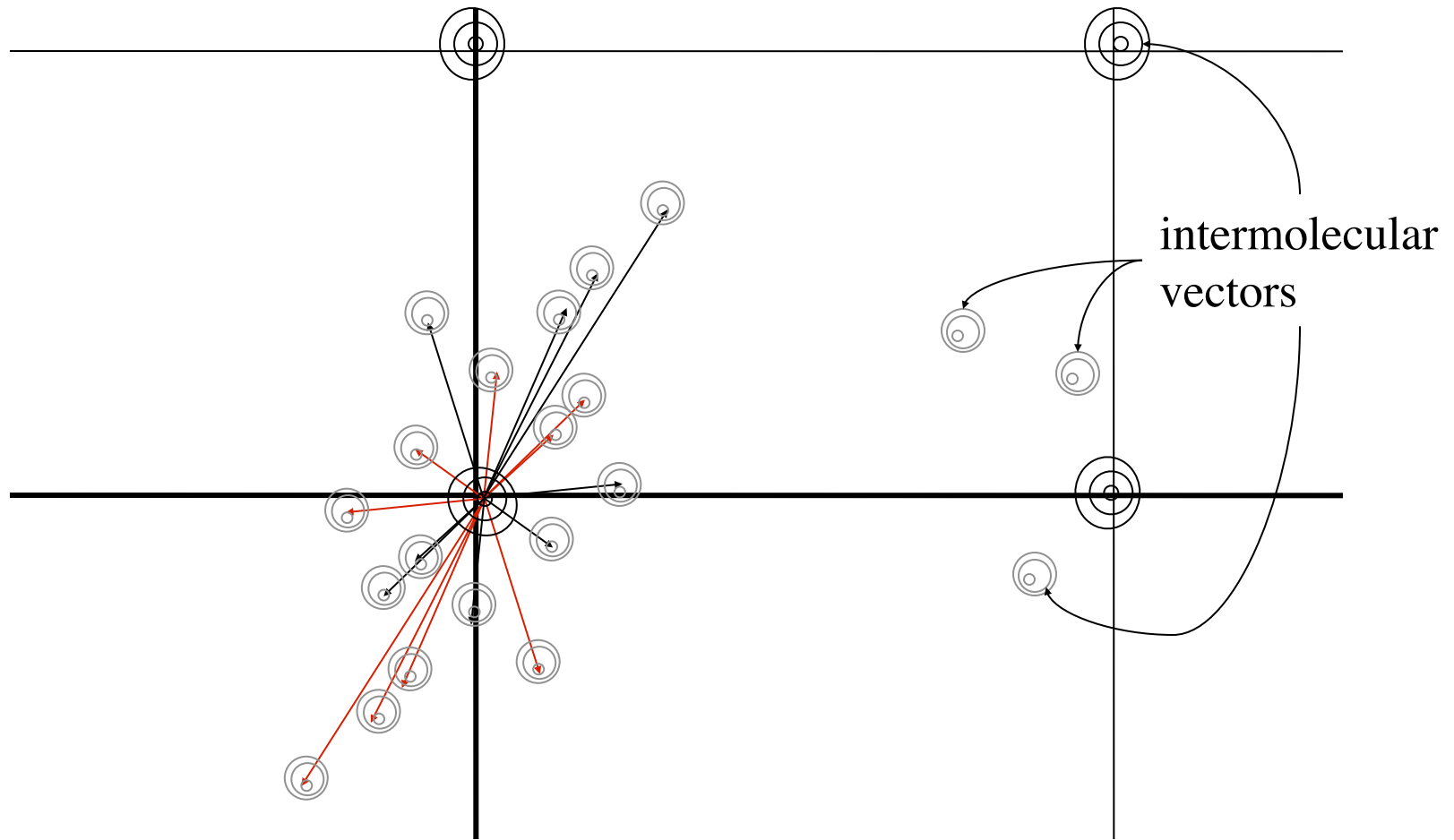


Patterson space

# Rotated Patterson map for Gly



# Rotated Patterson map for Gly



intramolecular vectors  
rotate around the origin

intermolecular vectors are  
transformed differently

## (2) The Patterson map of the crystal

The Patterson map represents all atom-atom vectors, translated to the origin.

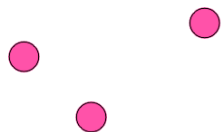
Included in this mess are vectors *within* molecules (this is what we want to detect), and vectors *between* symmetry-related molecules (these are considered noise to the Rotation Function).

Both *intramolecular* and *intermolecular* vectors exist in  $Z$  copies, oriented according to the rotational symmetry within the cell.  $Z$  is the number of symmetry operators in the space group.

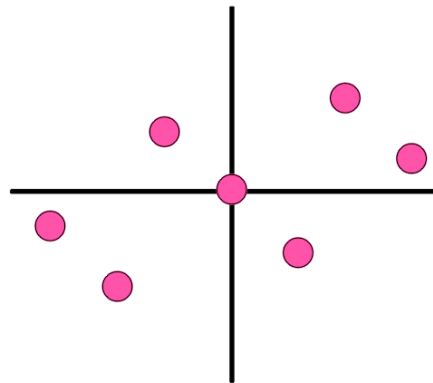
If there is more than one molecule in the asymmetric unit, then there are  $n \cdot Z$  copies of the intramolecular vectors.

Therefore, there are  $n \cdot Z$  correct solutions to the Rotation Function.

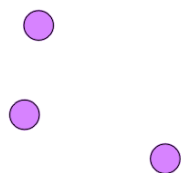
Target structure



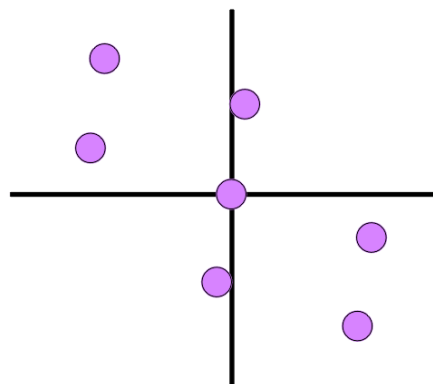
Target Patterson



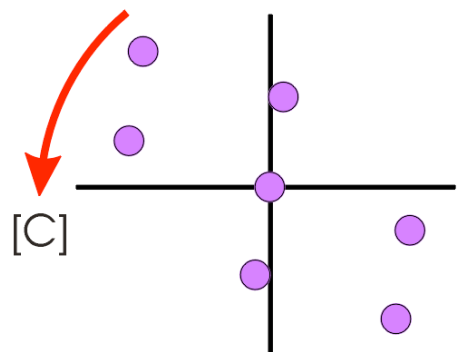
Search model



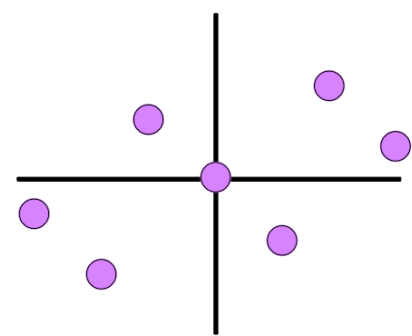
Search model Patterson



Search model Patterson

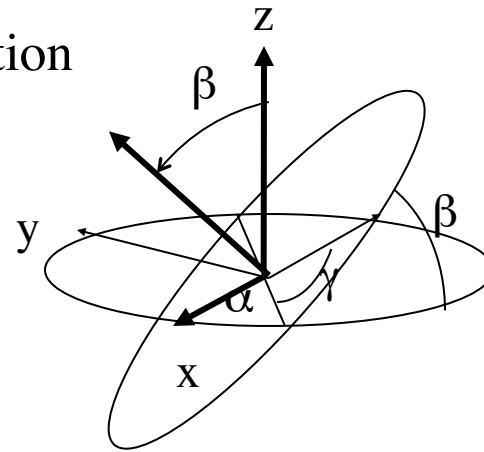
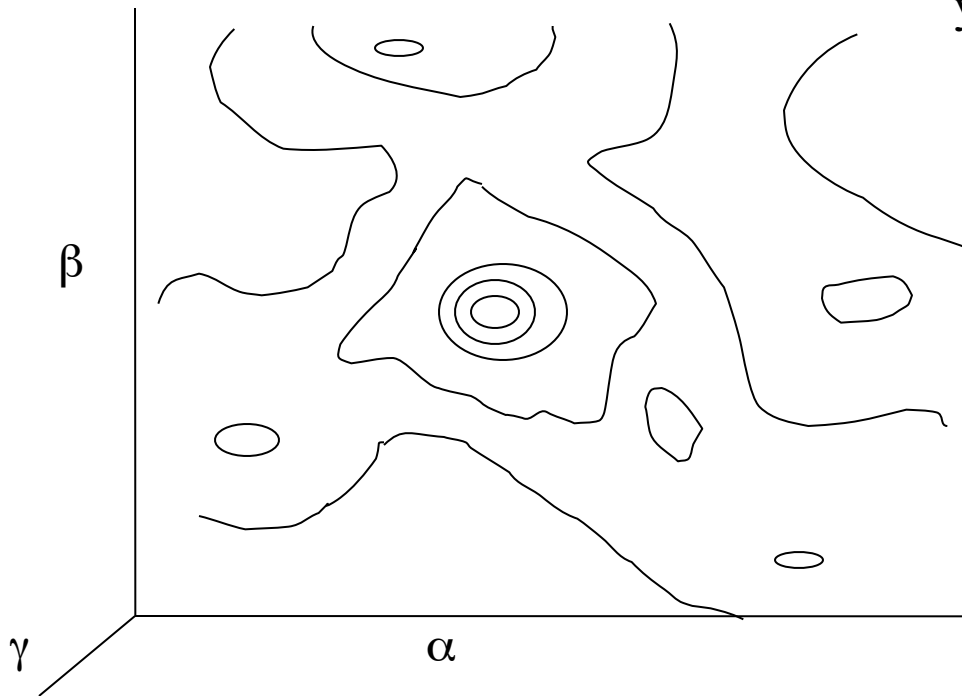


Matches target Patterson



# (3) 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.



# Correlation, defined

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

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

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

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

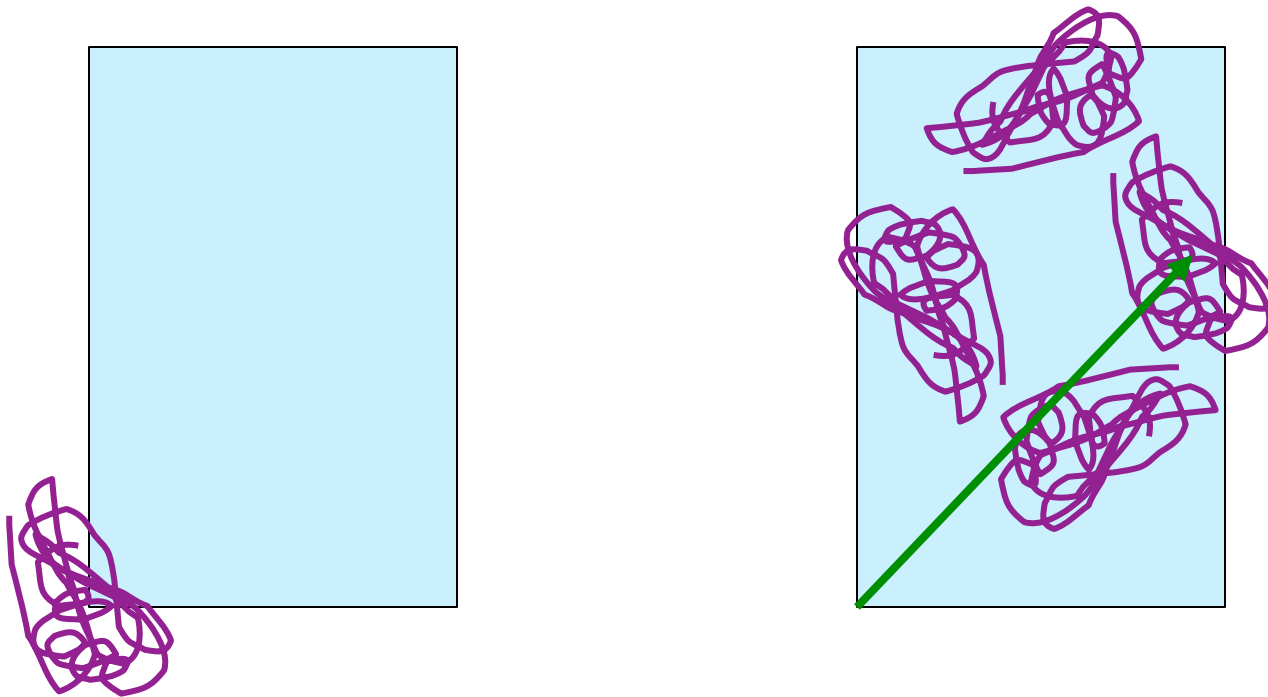
# Non-crystallographic symmetry can be detected using the Self

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:

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

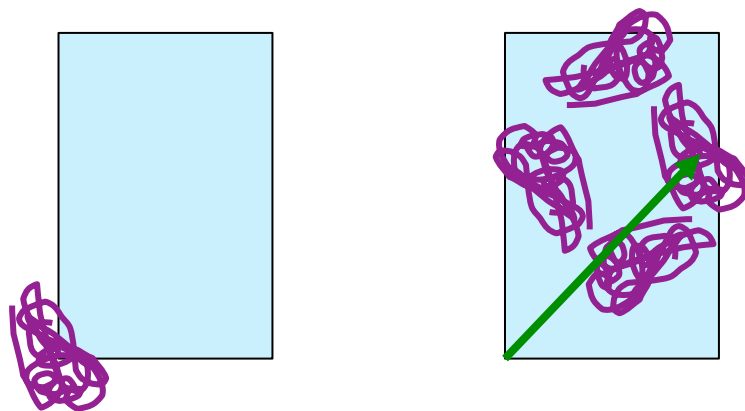
# (4) The Translation Function



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?

# (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:

$$x' = x + t$$

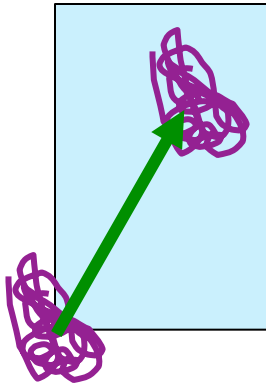
Symmetry-related, translated coordinates are:

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

# 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  $hkl$ )



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

note:  $v$  is in fractional coordinates

# Calculating $F_{\text{calc}}$ using symmetry

all atoms  
all equivs

$$F_{\text{calc}}(\vec{h}) = \sum_{Z=\text{symops}} \sum_{g=\text{atoms}} f_g(\vec{h}) e^{i2\pi\vec{h}\cdot(Z(r_g))}$$

where  $Z(\mathbf{r}) = M\mathbf{r} + \mathbf{v}$ . Let us define  $F_{\text{mod}}$ :

$$F_{\text{mod}}(\vec{h}) = \sum_{g=\text{atoms}} f_g(\vec{h}) e^{i2\pi\vec{h}\cdot r_g}$$

$F_{\text{calc}}$  is, therefore, just  $F_{\text{mod}}$  summed over the symmetry operators  $Z$ .

$$F_{\text{calc}}(\vec{h}) = \sum_Z F_{\text{mod}}(Z(\vec{h})) e^{i2\pi\vec{h}\cdot(Z(r_g))}$$

↑  
syms

# Reciprocal space symmetry

Rotating atoms in real space,

$$\mathbf{r}' = \mathbf{M}\mathbf{r}$$

then multiplying by  $\mathbf{h}$  to get the phase,

$$\text{phase} = 2\pi(\mathbf{h} \cdot (\mathbf{M}\mathbf{r}))$$

is the same as rotating reciprocal space *the other way*.

$$\mathbf{h} \cdot (\mathbf{M}\mathbf{r}) = 2\pi(\underline{\mathbf{M}}^T \mathbf{h} \cdot \mathbf{r})$$

You can prove this by writing out the matrix multiplication.

# Conclusions, summary

- **Molecular replacement** is the solution of the problem  $r' = \underline{M}r + v$  where  $r$  are the model coordinates (from a homolog model) and  $r'$  are the true crystallographic coordinates.
- The **rotation function** finds the rotation matrix  $\underline{M}$ .
- The **translation function** finds the translation vector  $v$ .
- The **rotation function** is done in Patterson space.
- The **translation function** can be done in reciprocal space because  $F_{\text{calc}}$  can be computed from  $F_{\text{mod}}$  and symmetry.

# Problems with the MR method:

## *Phase bias*

Molecular replacement solutions may be “suspicious” due to the possibility of phase bias. Parts of the model may be wrong, but the map may not show this.

We have already discussed ways to detect/correct errors:

real space R-factor

omit maps

B-factors