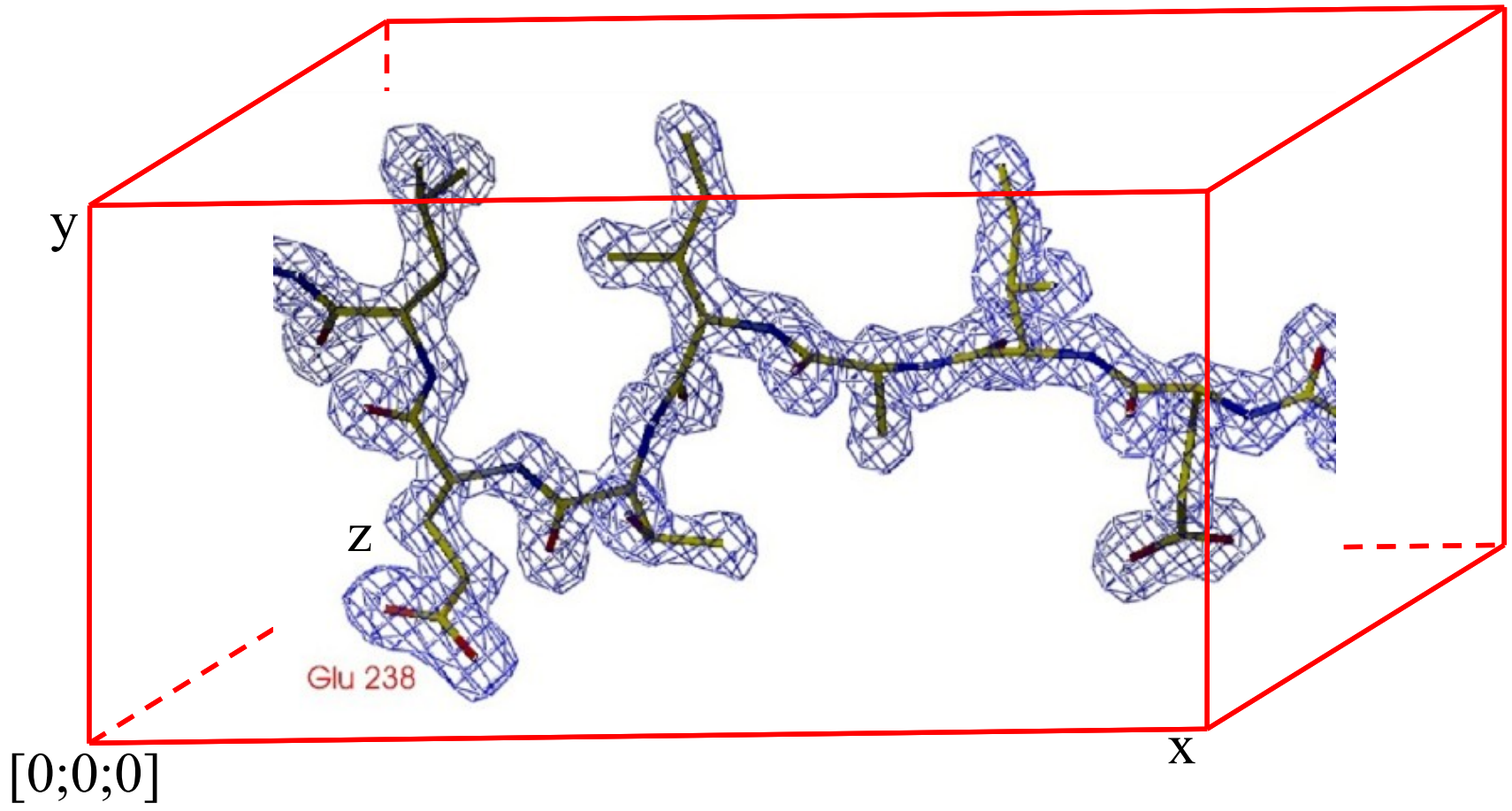
Structural Biology Methods

Fall 2022

Lecture #4

Phase problem

$$\rho(x\,y\,z) = \frac{1}{V}\sum_h\sum_k\sum_l \left|F(h\,k\,l)\right| \exp\left[-2\pi i(hx + ky + lz) + i\,\text{[redacted]}\right]$$



Solving the phase problem by:

Multiple/Single Isomorphous Replacement (MIR/SIR)

- source of phases – intensity differences between data from native and derivative (heavy atom containing) crystals
- Positions of heavy atoms identified from isomorphous difference Patterson maps

Solving the phase problem 3

Multiple/Single-wavelength anomalous diffraction (MAD/SAD)

- source of phases – intensity differences between structure factors due to the presence of atom that specifically interacts with X-rays of a given wavelength
- Positions of heavy atoms identified from anomalous difference Patterson maps

Phase problem

$$\rho(x\ y\ z) = \frac{1}{V}\sum_{hkl} |F(h\,k\,l)| \exp[-2\pi i(hx + ky + lz) + i\,\text{[redacted]}] \qquad (7.1)$$

Patterson function

$$P(u\ v\ w) = \frac{1}{V}\sum_{hkl} |F(h\,k\,l)|^2 \cos[2\pi(h\,u + k\,v + l\,w)]; \qquad (7.2)$$

or, shorter,

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathrm{s}} |F(\mathbf{S})|^2 \cos[2\pi\mathbf{u}\cdot\mathbf{S}] \qquad (7.3)$$

$$P_{1,2}(\mathbf{u}) = \int_V \rho_1(\mathbf{x}) \times \rho_2(\mathbf{x}+\mathbf{u})d\mathbf{x}. \qquad (10.7)$$

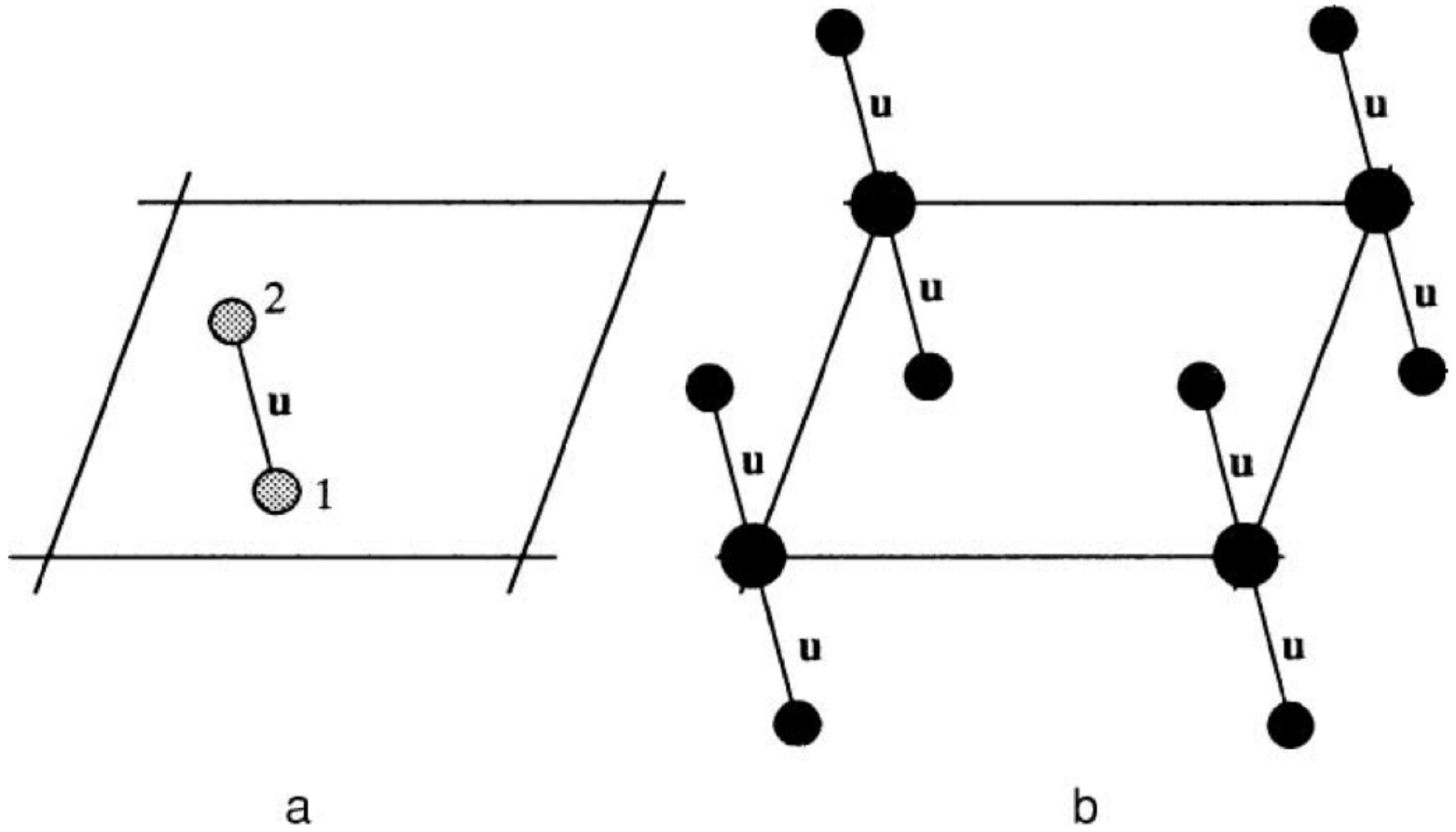


Figure 7.1. **(a)** A two-dimensional unit cell with only two atoms. **(b)** The corresponding Patterson cell.

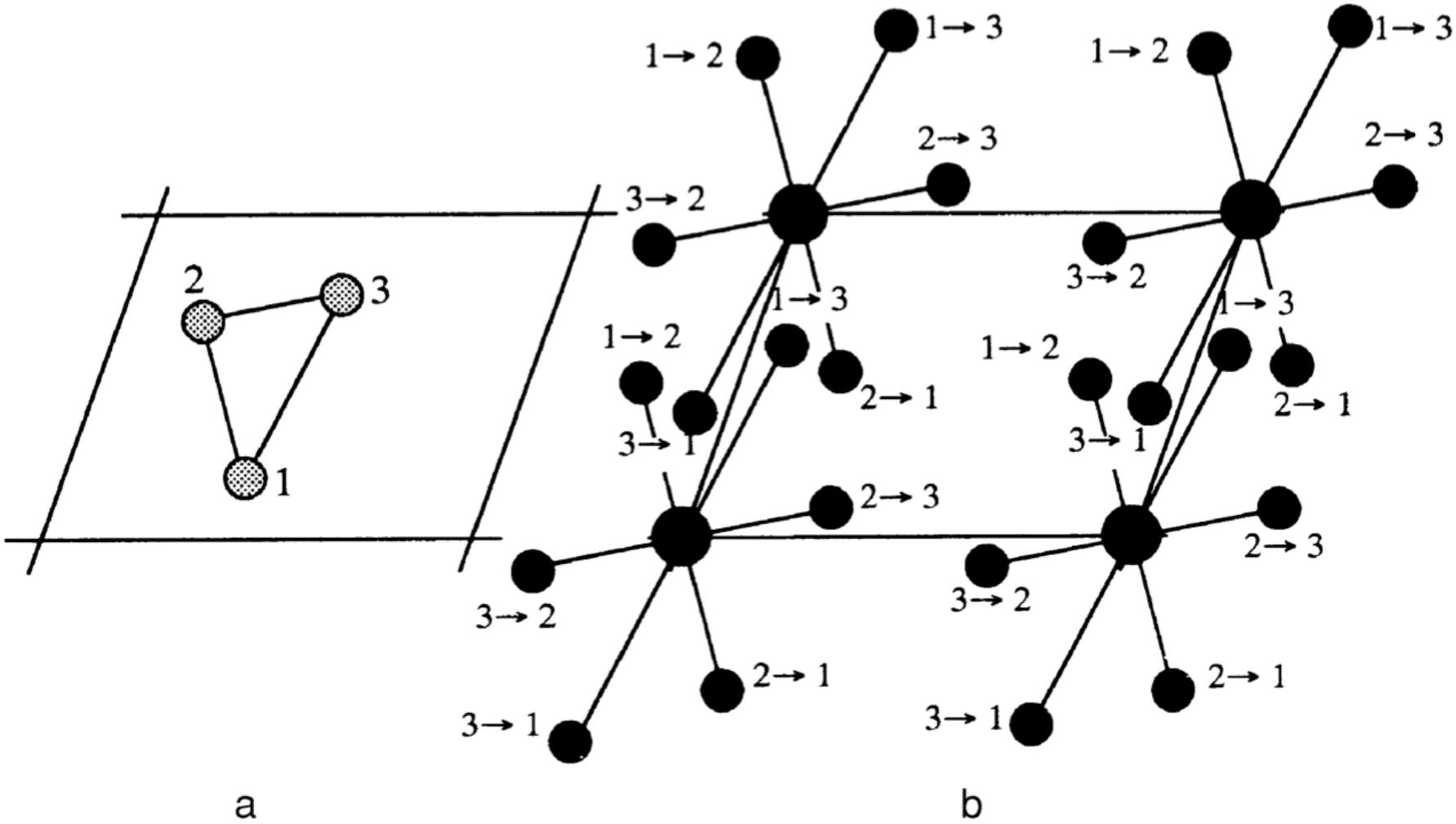


Figure 7.2. (a) A two-dimensional unit cell with three atoms. (b) The corresponding Patterson map. Note the large increase in the number of Patterson peaks compared with Figure 7.1. The total number of peaks is N^2, but the N self-peaks overlap at the origin and, therefore, $N(N-1)$ nonorigin peaks are found in a Patterson map. Because of the centrosymmetry in the map, the number of unique peaks is $[N(N-1)]/2$; in this figure, $1 \rightarrow 2$, $1 \rightarrow 3$, and $2 \rightarrow 3$ are unique peaks.

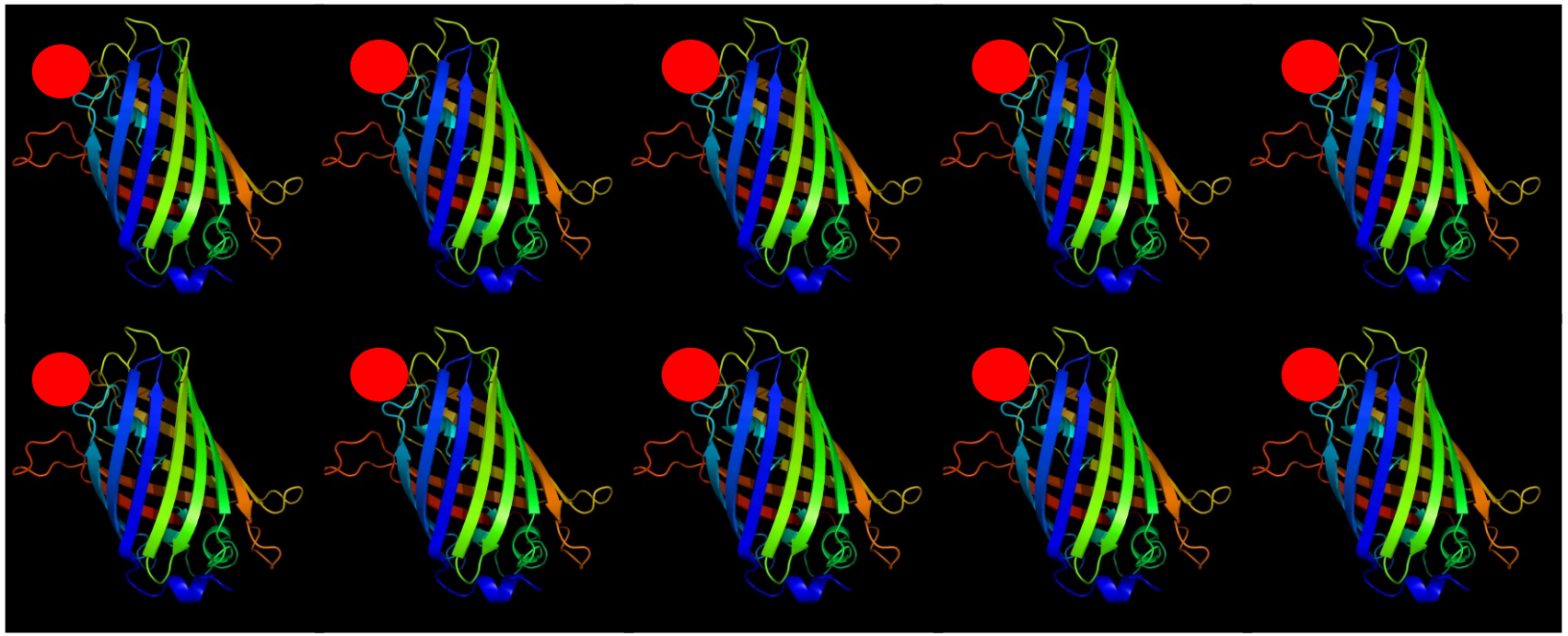
1. The Patterson map has peaks at end points of vectors **u** equal to vectors *between* atoms in the real cell.
2. For every pair of atoms in the real cell, there exists a unique peak in the Patterson map.
3. A Patterson map is always centrosymmetric.

5. Symmetry elements can cause a concentration of peaks in certain lines or planes: "Harker lines" or "Harker planes."

7. In locating Patterson peaks of heavy atoms in the isomorphous replacement method, it is useful to realize that the height of a peak is proportional to the product of the atomic numbers of the atoms that are responsible for the peak.

Patterson map of a macromolecule is a mess!

Isomorphous replacement method



1. Preparation of at least one, but preferably a few heavy-atom-containing derivatives of the protein in the crystalline state. A first check for isomorphism is measuring the cell dimensions.
2. X-ray intensity data must be collected for crystals of the native protein as well as for crystals of the derivatives.
3. Application of the Patterson function for the determination of the heavy atom coordinates.
4. Refinement of the heavy atom parameters and calculation of the protein phase angles.
5. Calculation of the electron density of the protein.

Isomorphous replacement method

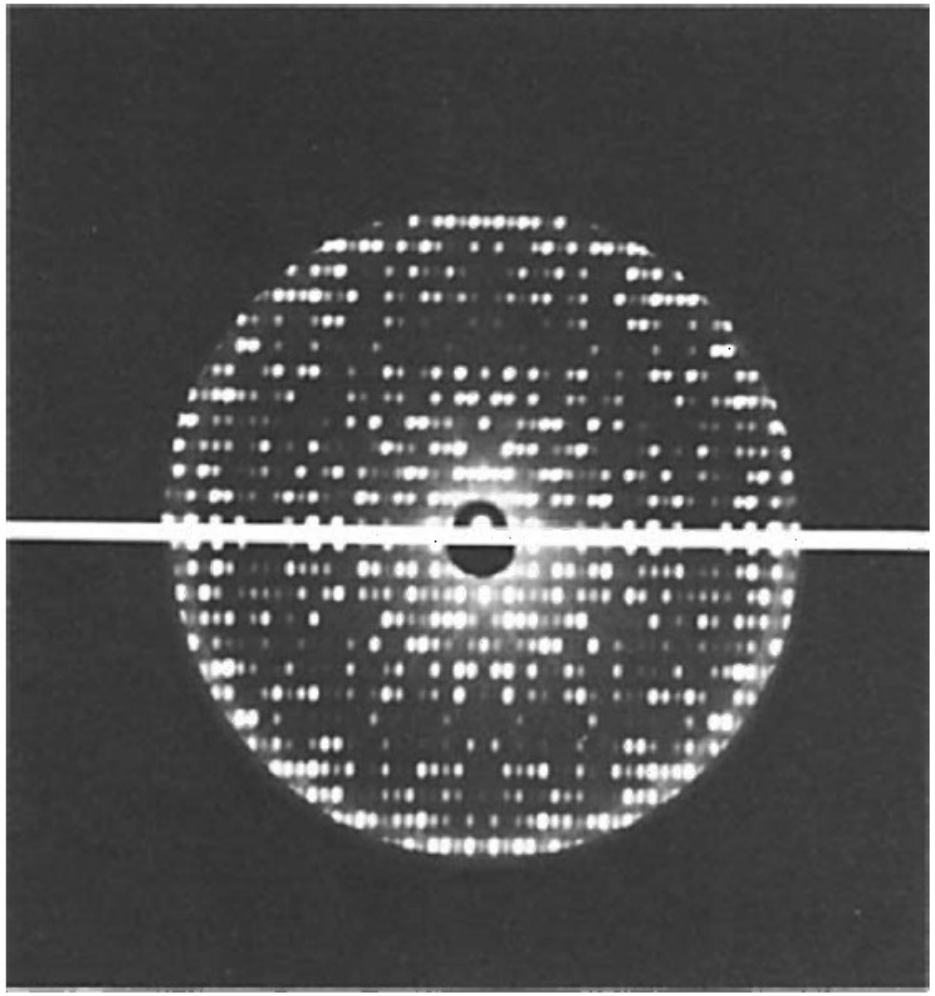
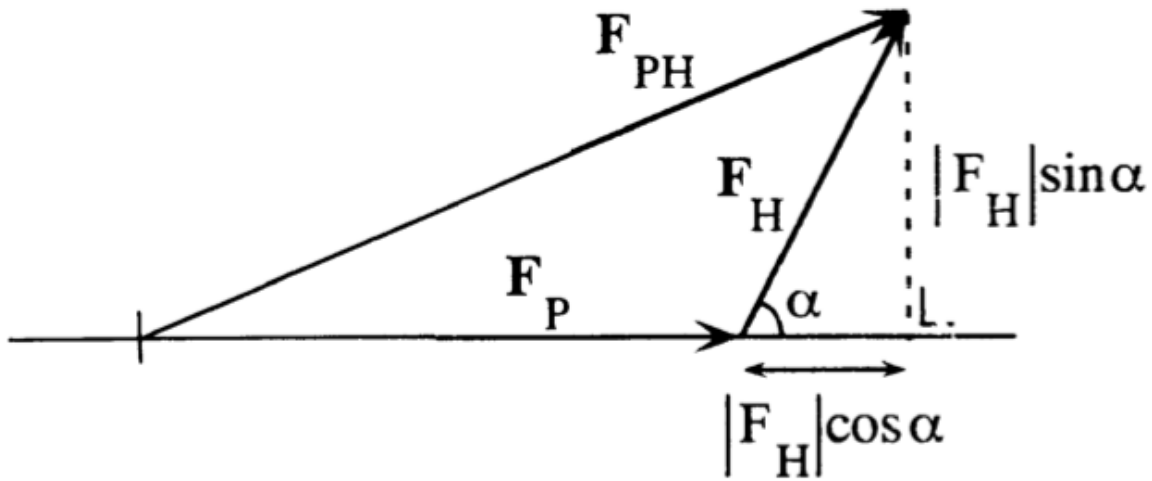


Figure 7.6. A comparison of the diffraction photographs of the same reciprocal lattice plane for a native papain crystal and a heavy atom derivative in which one mercury atom was attached to each protein molecule. Appreciable differences in intensity between corresponding diffraction spots can be seen.

Figure 7.8. Structure factors in the isomorphous replacement method for noncentric reflections; the horizontal direction of $\mathbf{F}_P$ is arbitrary.



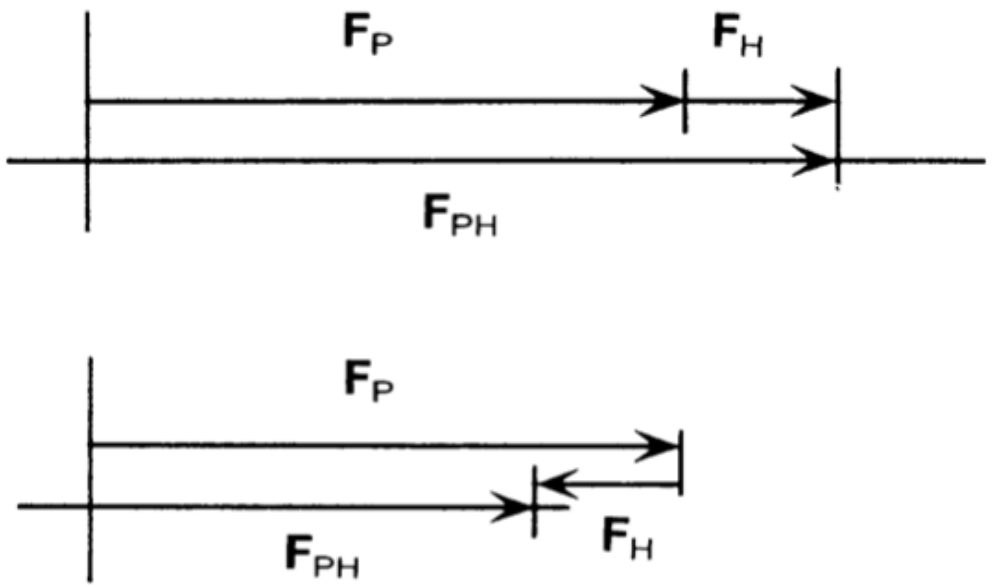
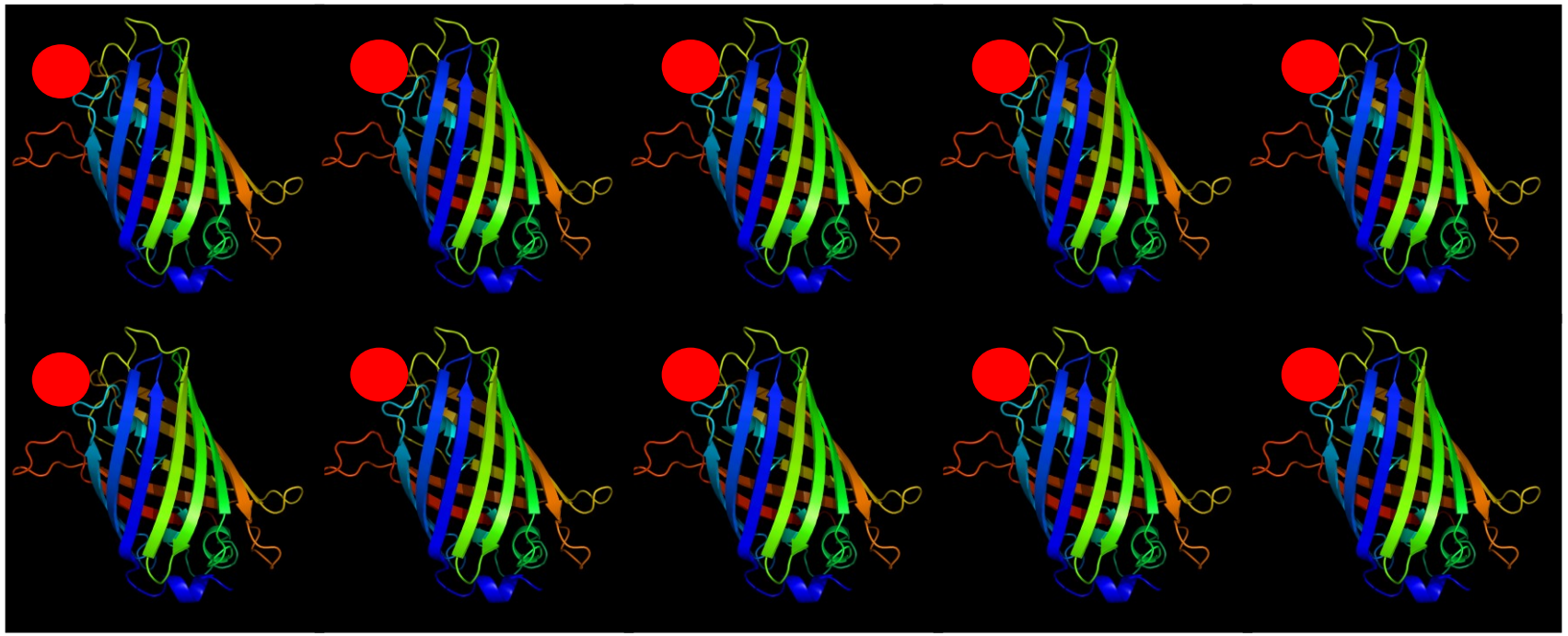


Figure 7.7. Structure factors in the isomorphous replacement method for centric reflections. $\mathbf{F}_P$ is for the protein, $\mathbf{F}_{PH}$ is for the derivative, and $\mathbf{F}_H$ is for the heavy atom contribution.

Isomorphous replacement method



Determination of heavy atom positions from centric reflections (centro-symmetric projections)

$$\mathbf{F}_{\mathrm{PH}} = \mathbf{F}_P + \mathbf{F}_H$$

is simplified to

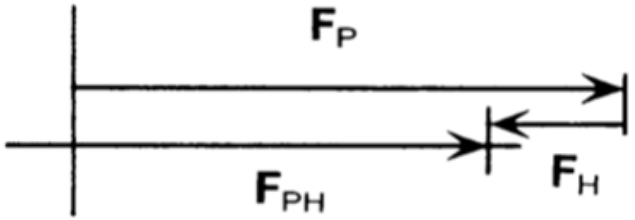
$$|F_{\mathrm{PH}}| = |F_P| \pm |F_H|\,,$$
$$|F_H| = |F_{\mathrm{PH}}| - |F_P|$$

or

$$|F_H| = |F_P| - |F_{\mathrm{PH}}|$$

and

$$|F_H|^2 = (|F_{\mathrm{PH}}| - |F_P|)^2\,.$$



We have made the assumption that F_{PH} and F_P have the same sign, either both positive or both negative. With this assumption, the Patterson summation with the coefficients $(|F_{PH}| - |F_P|)^2$ will give a Patterson map of the heavy atom arrangement in the unit cell. For the majority of the reflections, the assumption will be true, because, in general, F_H will be small compared with F_P and F_{PH}. If, however, F_P is small, F_{PH} could have the opposite sign and F_H would be $F_P + F_{PH}$. Fortunately, this does not occur often enough to distort the Patterson map seriously.

$$P(u\,v\,w) = \frac{1}{V}\sum_{hkl} |F(h\,k\,l)|^2 \cos[2\pi(h\,u + k\,v + l\,w)]$$

$$P(u\,v\,w) = \frac{1}{V}\sum_{h} (\Delta\,|F|_{\text{iso}})^2 \cos[2\pi(hu + kv + lw)]$$

Scale!

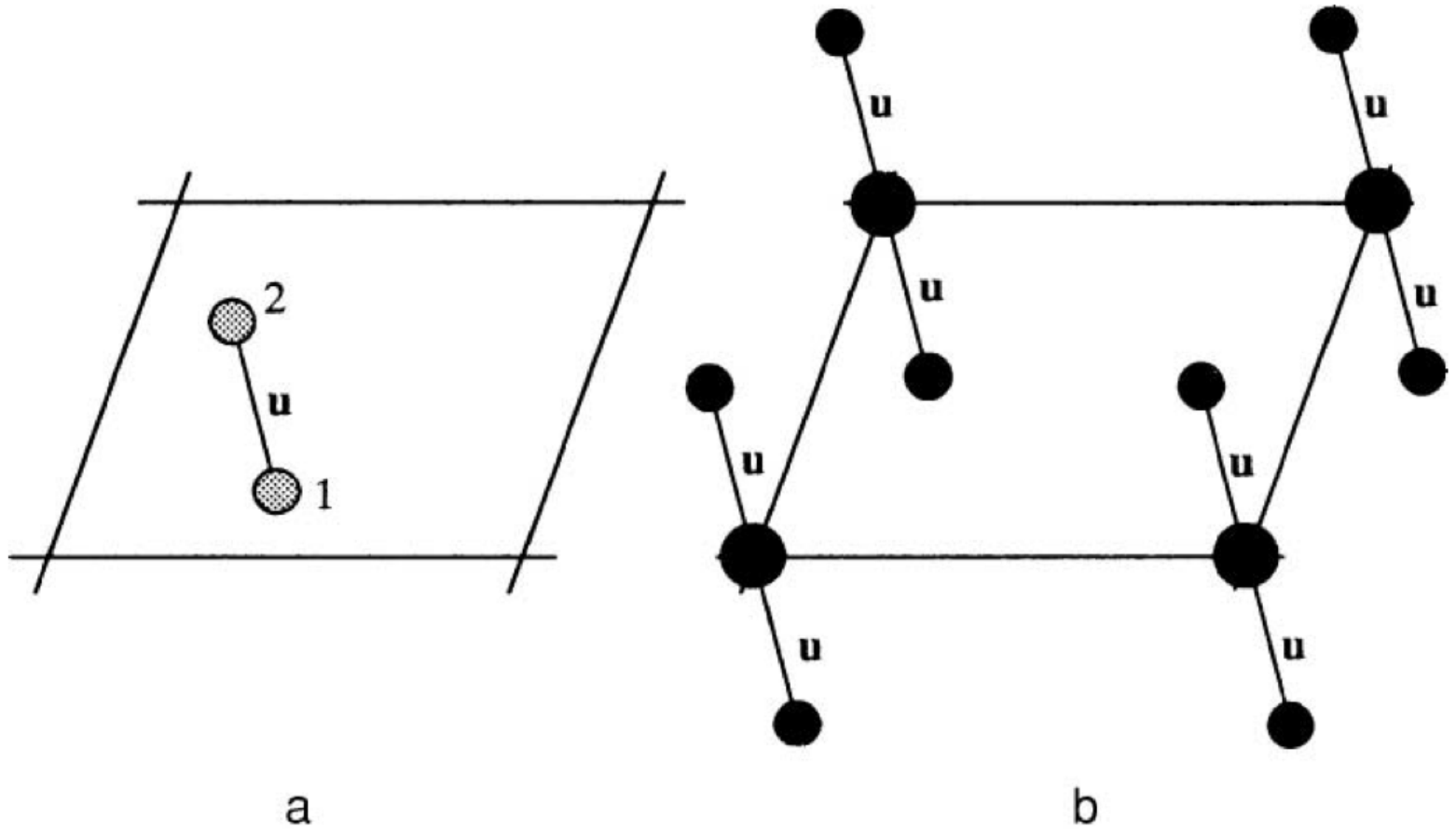


Figure 7.1. **(a)** A two-dimensional unit cell with only two atoms. **(b)** The corresponding Patterson cell.

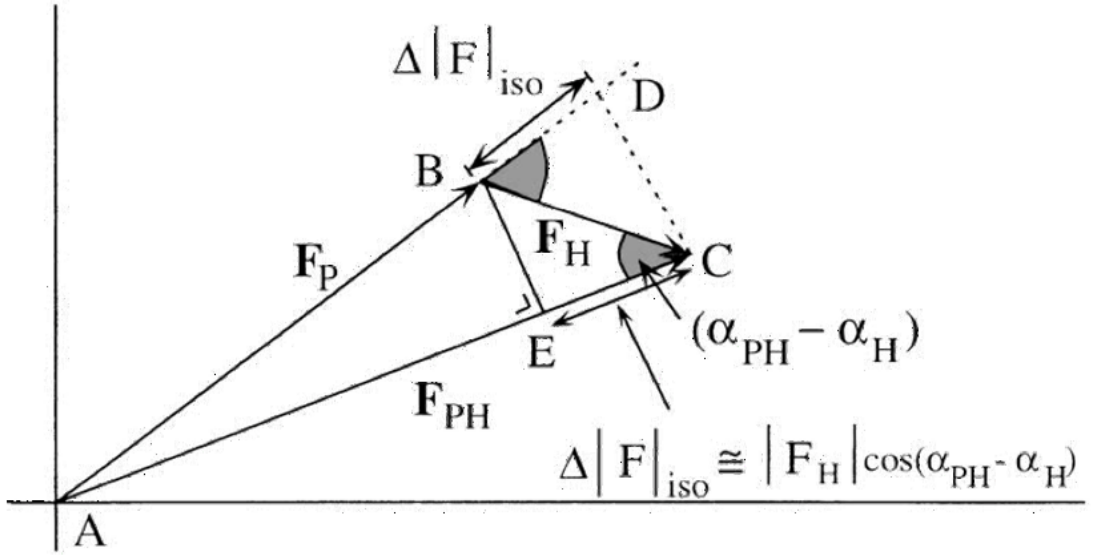
Determination of heavy atom positions from acentric reflections

$$\Delta |F|_{iso} = |F_{PH}| - |F_P| .$$

We will now see that the coordinates of the heavy atoms can generally be derived from a Patterson map calculated with $(\Delta|F|_{iso})^2$. The triangle ABC in Figure 7.9 expresses the vector sum: $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$. However, for the time being, only the lengths of $\mathbf{F}_{PH}(|F_{PH}|)$ and that of $\mathbf{F}_P(|F_P|)$ are known, but not their directions. For $\mathbf{F}_H$, both the length and direction are unknown.

In Figure 7.9, $CE = |F_H| \cos(\alpha_{PH} - \alpha_H)$. In general, $\alpha_P - \alpha_{PH}$ is small, because for most reflections, $|F_H| \ll |F_P|$ and $|F_{PH}|$. Therefore, $CE \cong \Delta|F|_{iso}$ and the result is

$$\Delta |F|_{iso} = |F_H| \cos(\alpha_{PH} - \alpha_H). \qquad (7.20)$$



The result is that a Patterson summation with $(\Delta|F|_{\text{iso}})^2$ as the coefficients will in fact be a Patterson summation with coefficients $|F_H|^2 \cos^2(\alpha_{\text{PH}} - \alpha_{\text{H}})$. Because

$$\cos^2(\alpha_{\text{PH}} - \alpha_{\text{H}}) = \frac{1}{2} + \frac{1}{2}\cos 2(\alpha_{\text{PH}} - \alpha_{\text{H}}),$$

we obtain

$$|F_H|^2 \cos^2(\alpha_{\text{PH}} - \alpha_{\text{H}}) = \frac{1}{2}|F_{\text{H}}|^2 + \frac{1}{2}|F_{\text{H}}|^2 \cos 2(\alpha_{\text{PH}} - \alpha_{\text{H}}).$$

Because the angles α_{PH} and α_H are not correlated, the second term on the right-hand side will contribute only noise to the Patterson map. However, the first term, $\frac{1}{2}|F_{\text{H}}|^2$, will give the Patterson function for the heavy atom structure on half of the scale.

$$P(u\,v\,w) = \frac{1}{V}\sum_h (\Delta\,|F|_{\text{iso}})^2 \cos[2\pi(hu + kv + lw)]$$

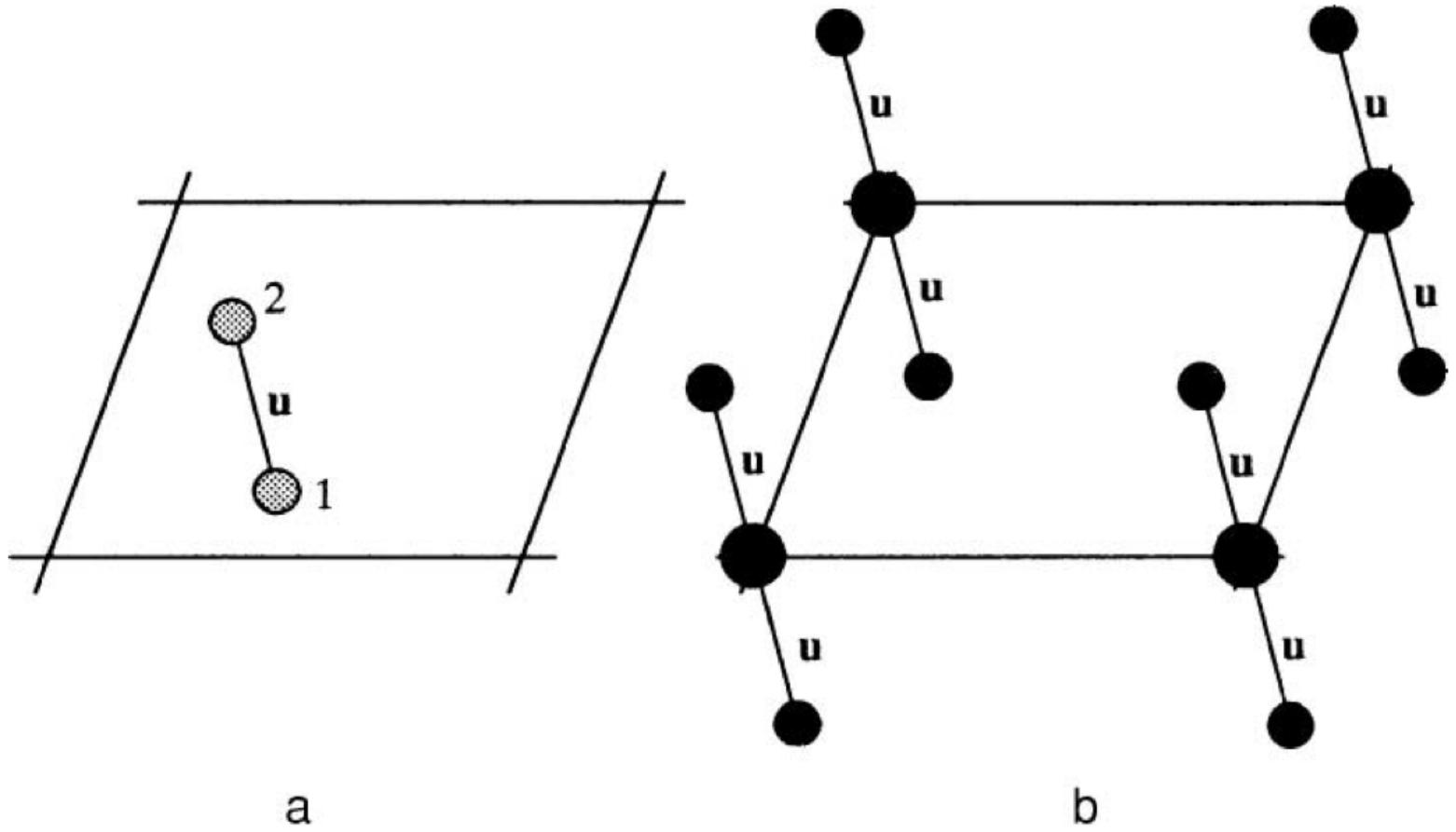
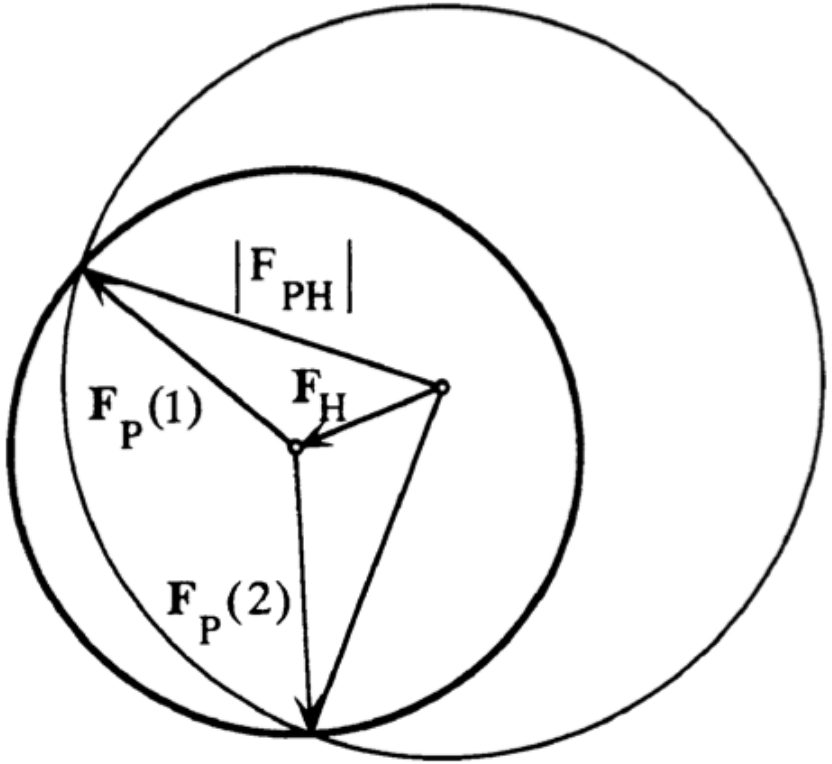
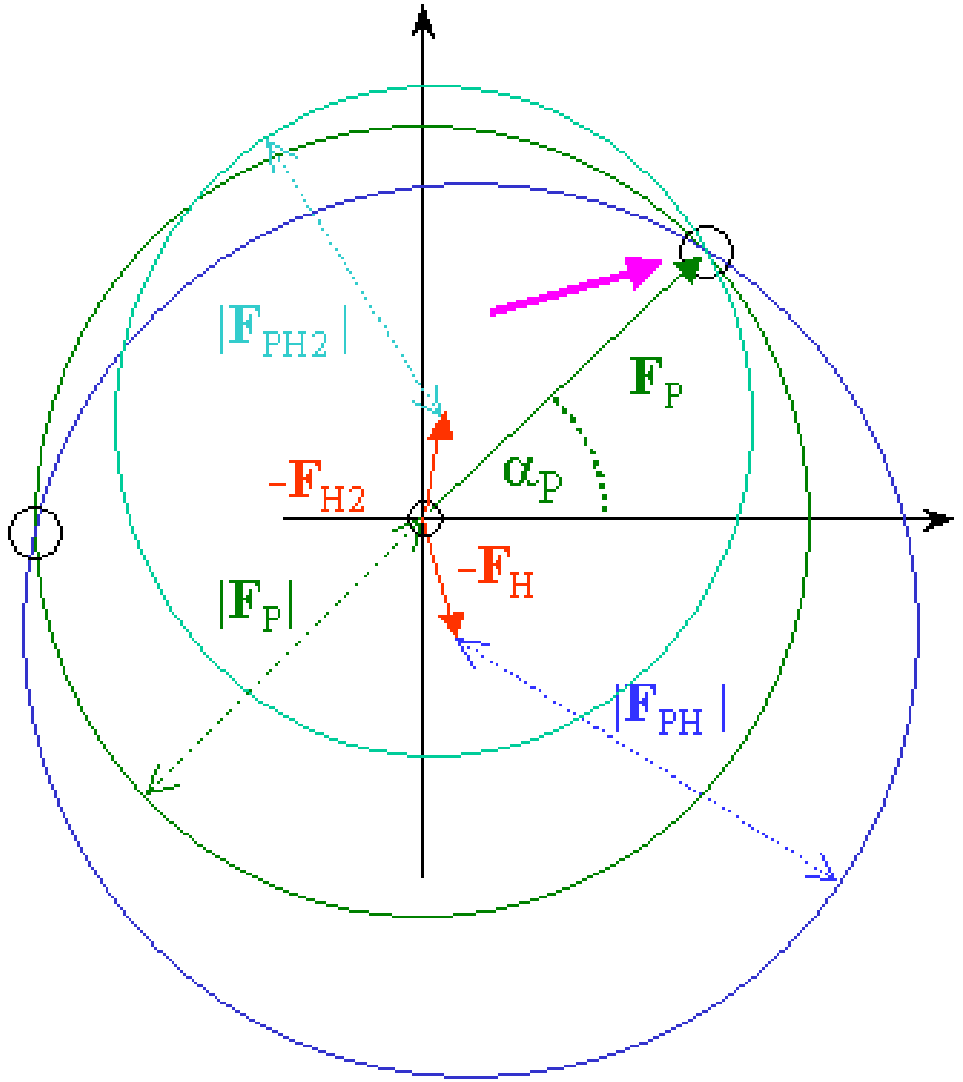


Figure 7.1. **(a)** A two-dimensional unit cell with only two atoms. **(b)** The corresponding Patterson cell.

Determination of protein phase angles

Figure 7.17. Harker construction for protein phase determination. In the isomorphous replacement method, each heavy atom derivative gives two possibilities for the protein phase angle α_P, corresponding to the two vectors $\mathbf{F}_P(1)$ and $\mathbf{F}_P(2)$.

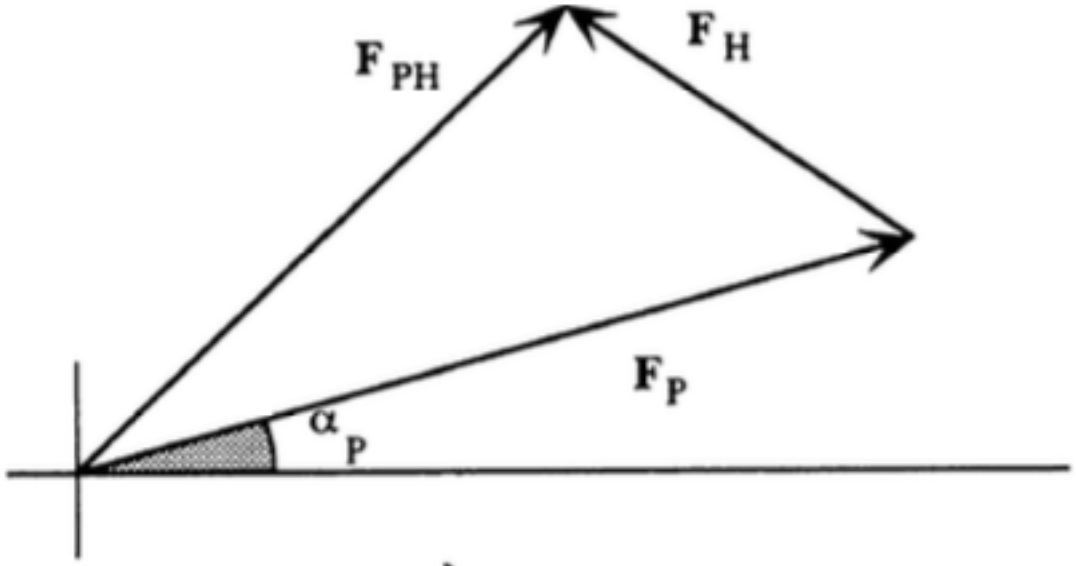




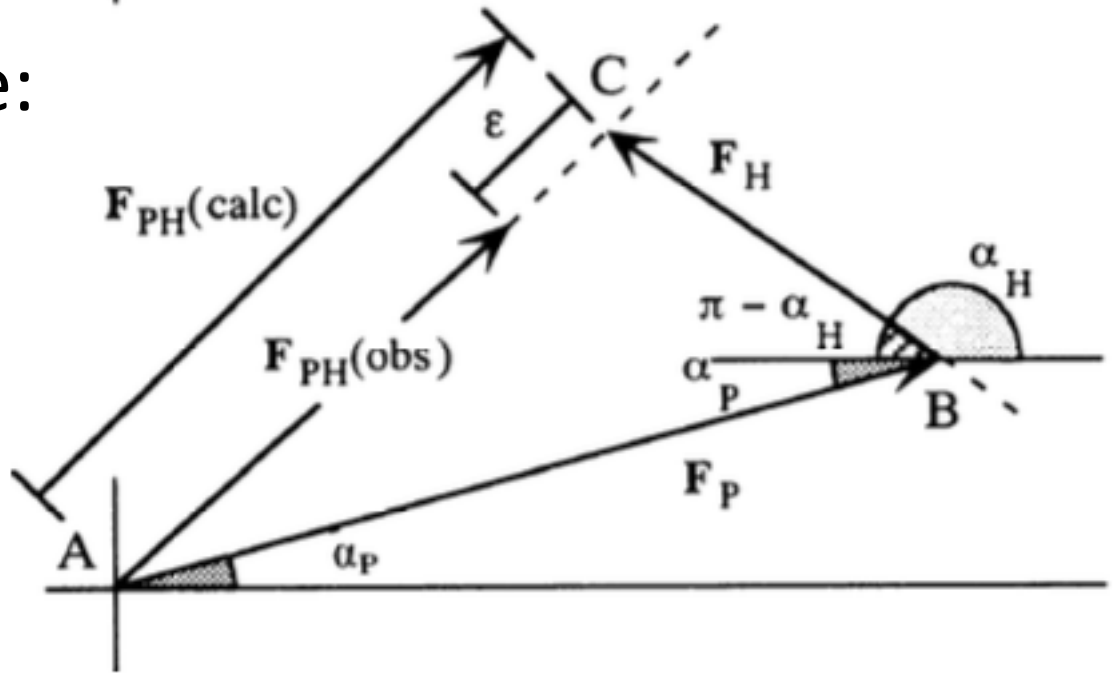
$$\rho(x\,y\,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F(h\,k\,l)\right|\exp\left[-2\pi i(hx+ky+lz)+i\,\right]$$

The “lack of closure" error

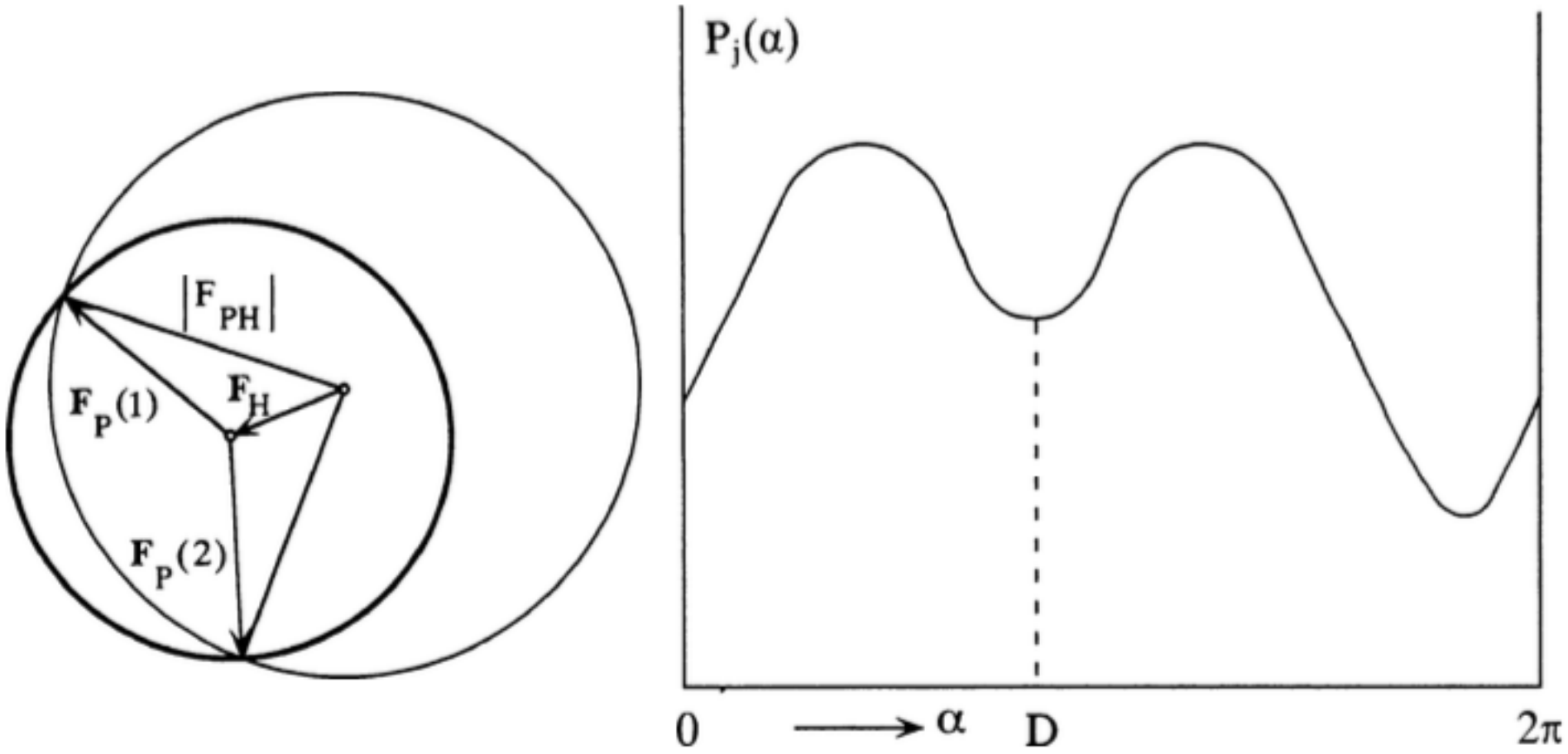
Ideal:

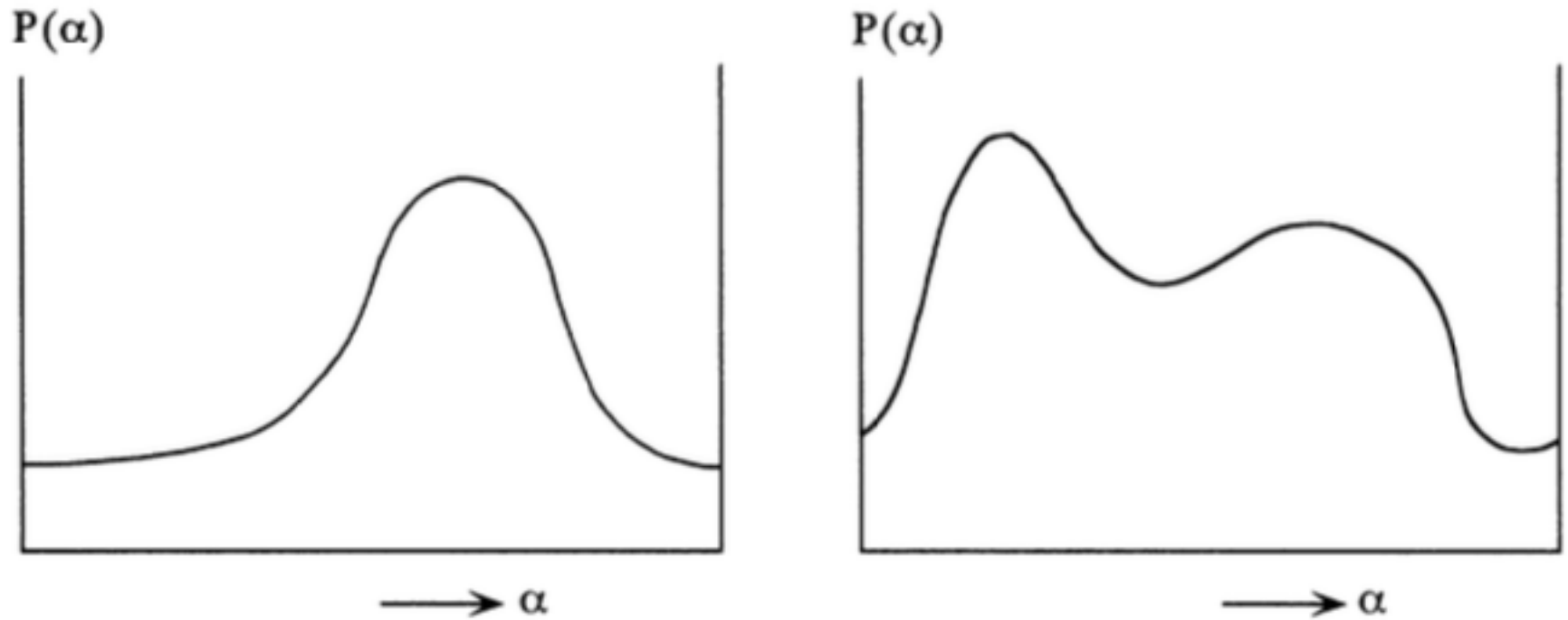


Real life:



Phase probability

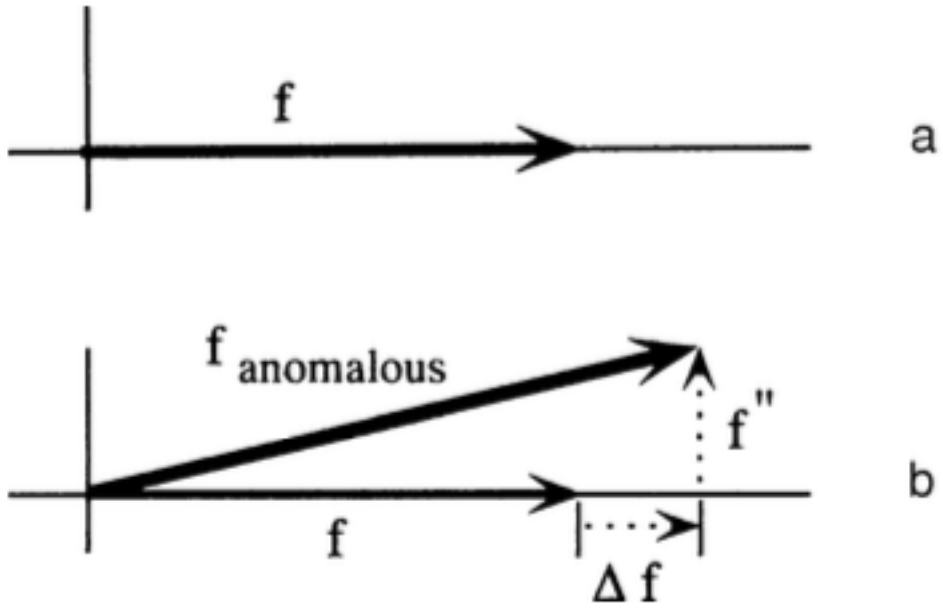


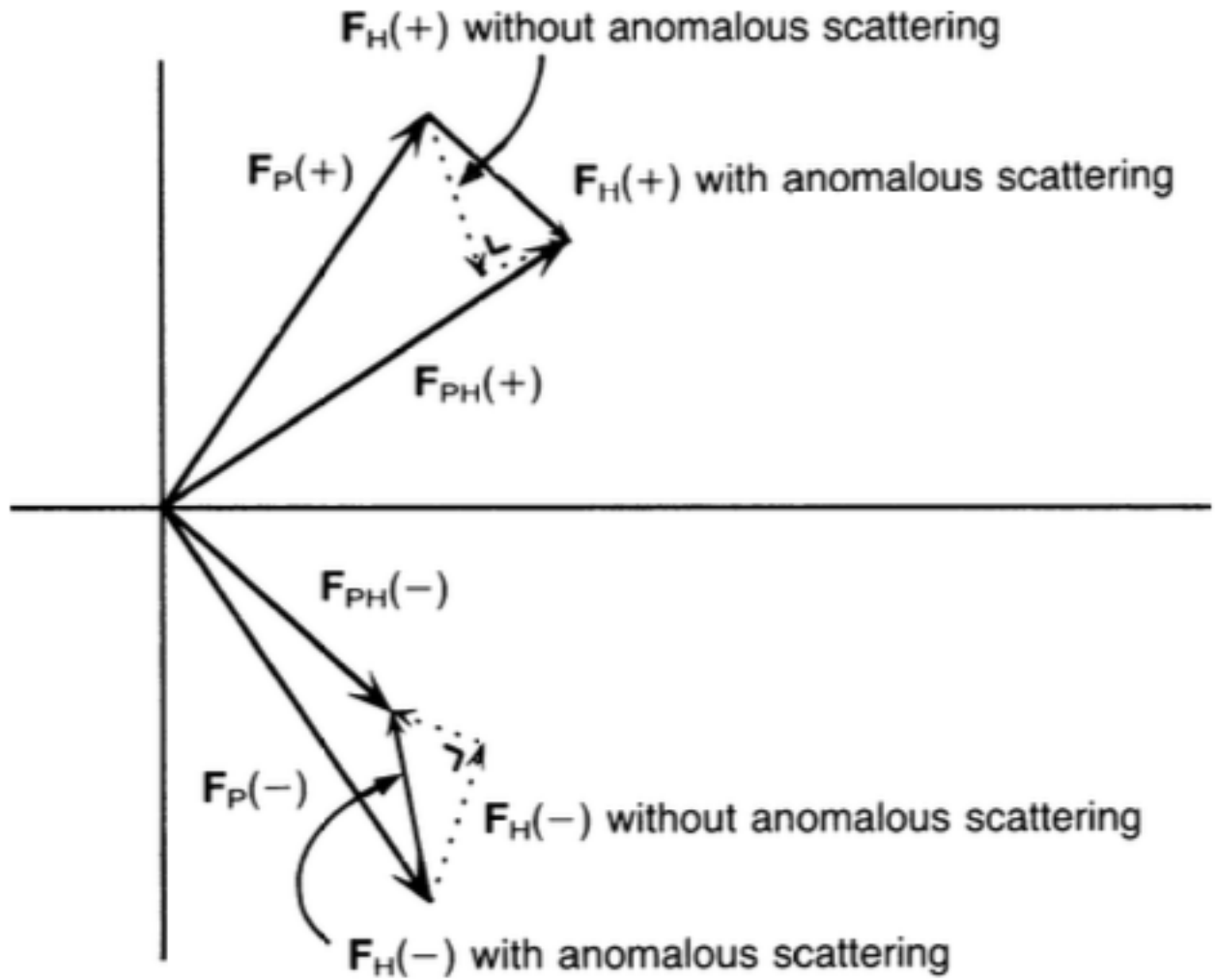


$$P_{hkl}(\alpha) = \prod_{j=1}^{n} P_j(\alpha_{hkl})$$

Anomalous scattering

Figure 7.11. The atomic scattering factor for a completely free electron (a) and for a bound electron (b). The anomalous contribution consists of two parts: a real part Δf and an imaginary part if''. The direction of the primary beam is pointing to the left; it has a 180° phase difference with f.




$F_H(+)$ without anomalous scattering
$F_P(+)$
$F_H(+)$ with anomalous scattering
$F_{PH}(+)$
$F_{PH}(-)$
$F_P(-)$
$F_H(-)$ without anomalous scattering
$F_H(-)$ with anomalous scattering

The anomalous Patterson map

The heavy atom contribution to the structure factor consists of a normal part, $\mathbf{F}_H$, and an anomalous part, $\mathbf{F}_{H''}$. In Figure 7.13 this is drawn for a reflection $(h\,k\,l)$ and for $(\bar{h}\,\bar{k}\,\bar{l})$. However, for convenience, the structure factors for $(\bar{h}\,\bar{k}\,\bar{l})$ have been reflected with respect to the horizontal axis. It can be derived that a Patterson summation with the coefficients $(\Delta|F|_{ano})^2$ can be approximated by a summation with the coefficients [redacted]
This will give a Patterson map of the anomalous scatterers (the heavy atoms, see belc

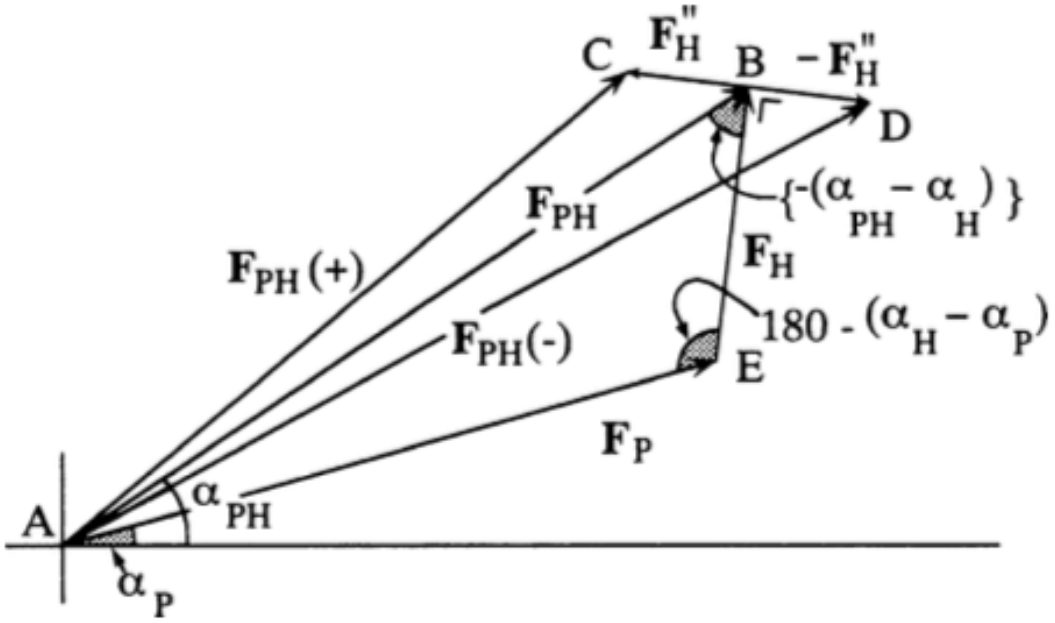


Figure 7.13. In this drawing, the structure factors $\mathbf{F}_P(-)$, $\mathbf{F}_{PH}(-)$, $\mathbf{F}_H(-)$, and $\mathbf{F}''_H(-)$ have been reflected with respect to the horizontal axis and combined with the structure factors for the reflection $(h\,k\,l)$. Note that α_H is the phase angle for the nonanomalous part of $\mathbf{F}_H$.

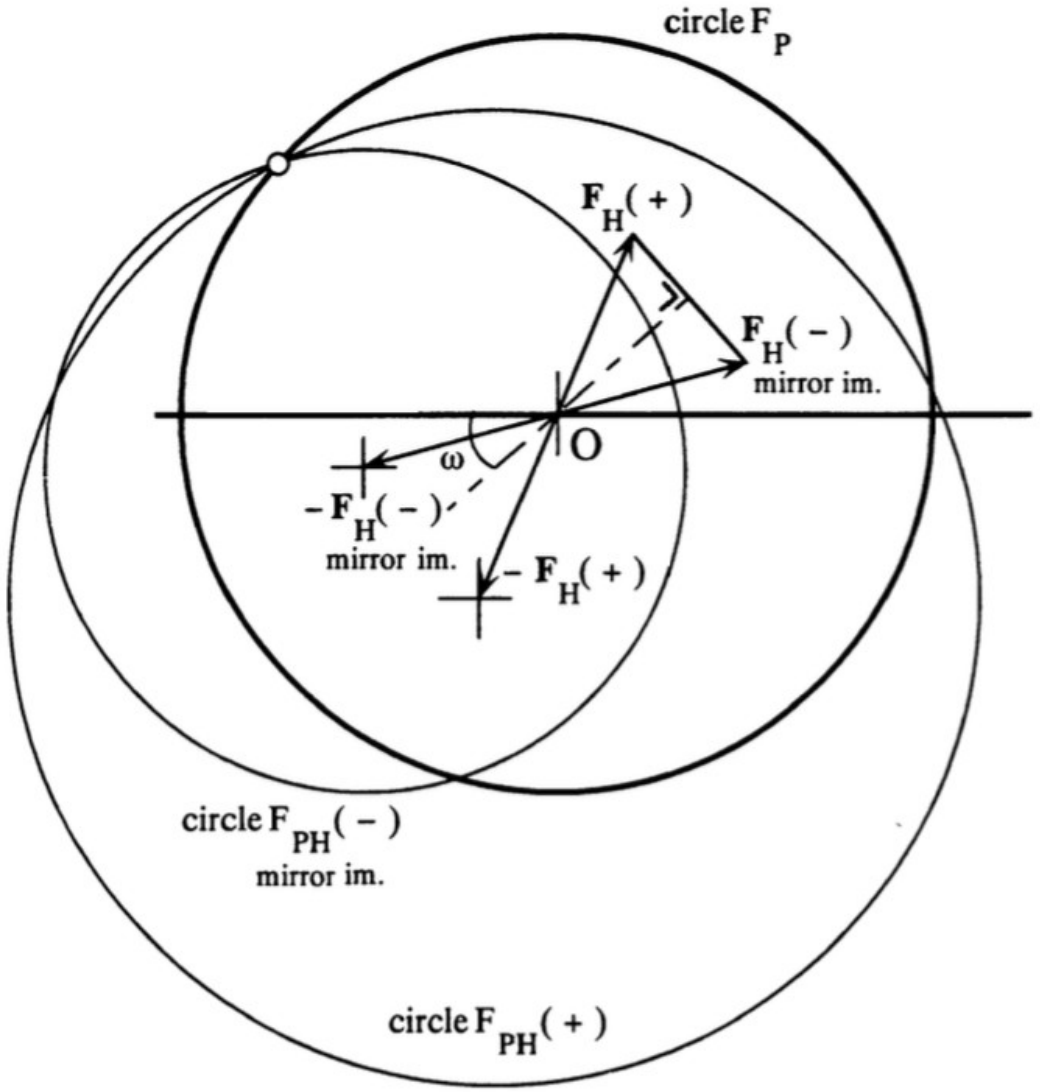


Figure 9.2. This figure gives the same information as Figure 9.1. The difference is that the vector [redacted] (mirror image with respect to the horizontal axis). The consequence is a different position for the $F_{PH}(-)$ circle. The advantage of this drawing is that the three circles have one common point of intersection. The dashed line indicates the direction of the nonanomalous scattering part of the heavy atoms.

Solving the phase problem by:

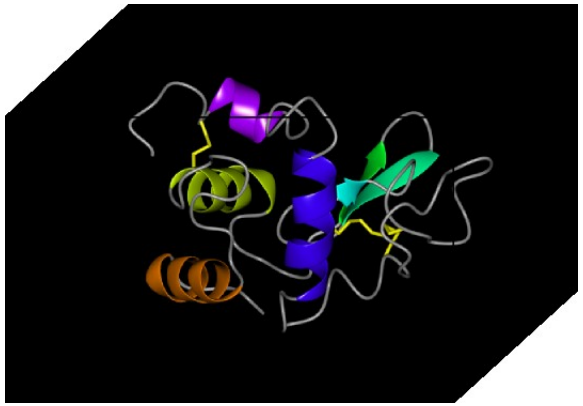
Molecular replacement

1. source of initial phases is a model
2. the model is oriented and positioned to obtain the best agreement with the x-ray data
3. phases are calculated from the model
4. The calculated phases are combined with the experimental data

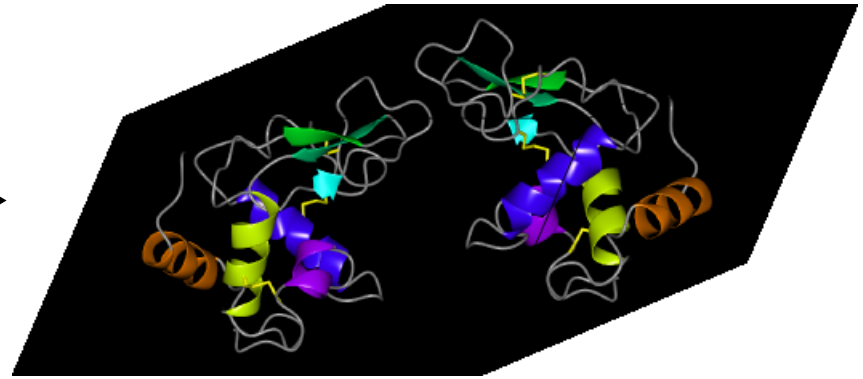
Molecular **R**eplacement was invented by
Michael **R**ossmann

Molecular Replacement

Known crystal structure



New crystal structure

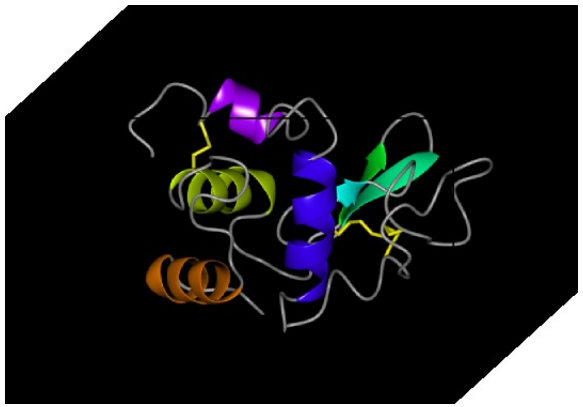


Given:
- Crystal structure of a homologue
- New X-ray data

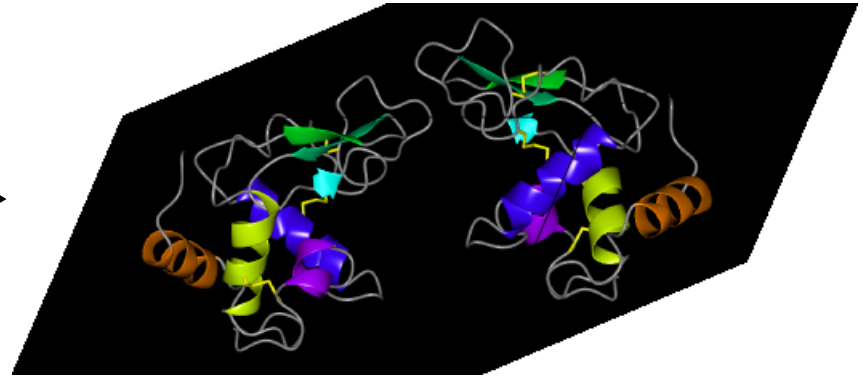
Determine:
- The new crystal structure

MR Technique

Known crystal structure



New crystal structure



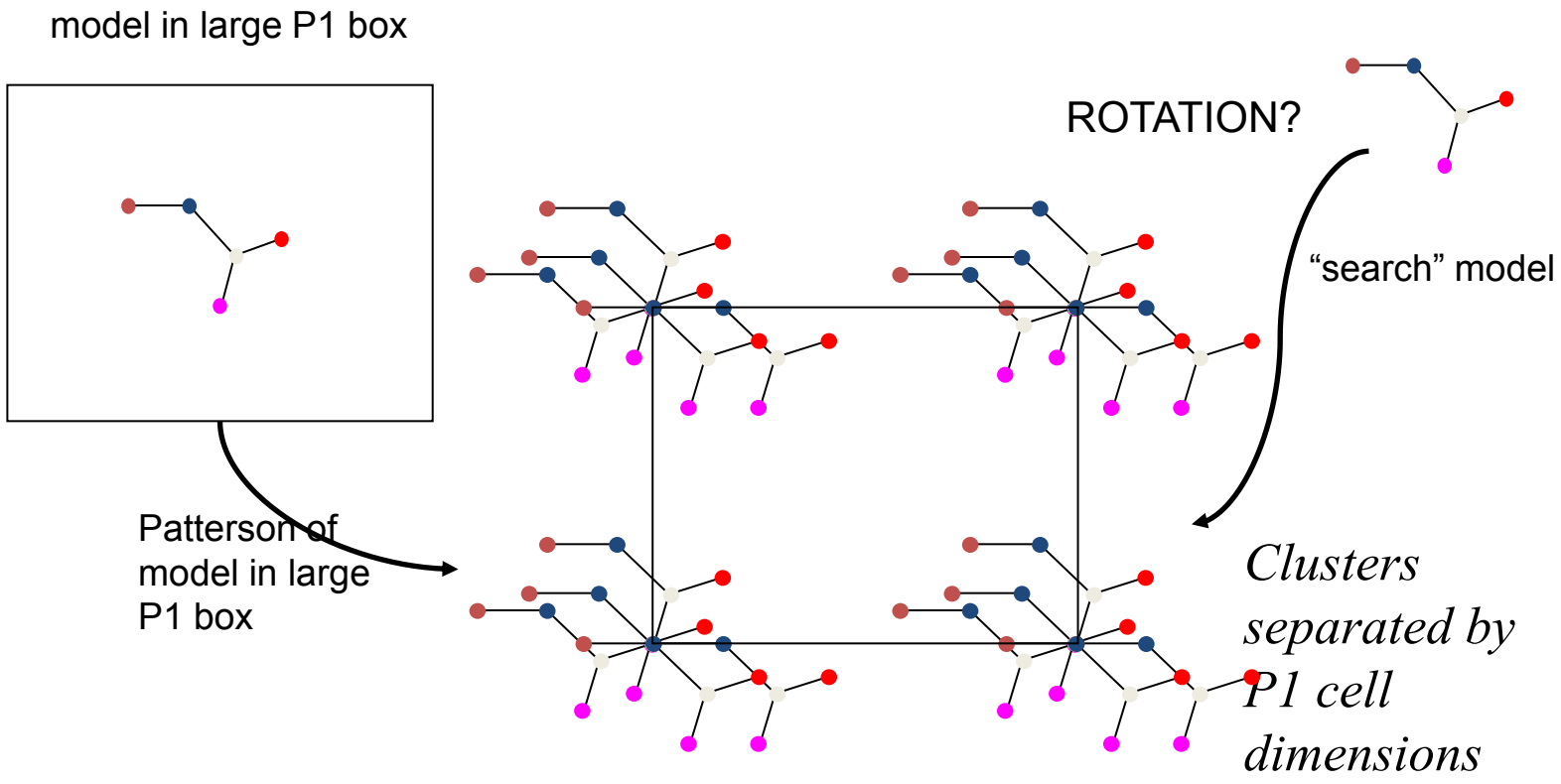
Method:
- 6-dimensional global optimisation
  - one 6-d search for each molecule in the AU
    >> split further to orientation + translation searches = 3 + 3?

Required:
- Scoring
  - the match between the data and (incomplete) model
  - ideally: the highest score = the correct model

ROTATION FUNCTION

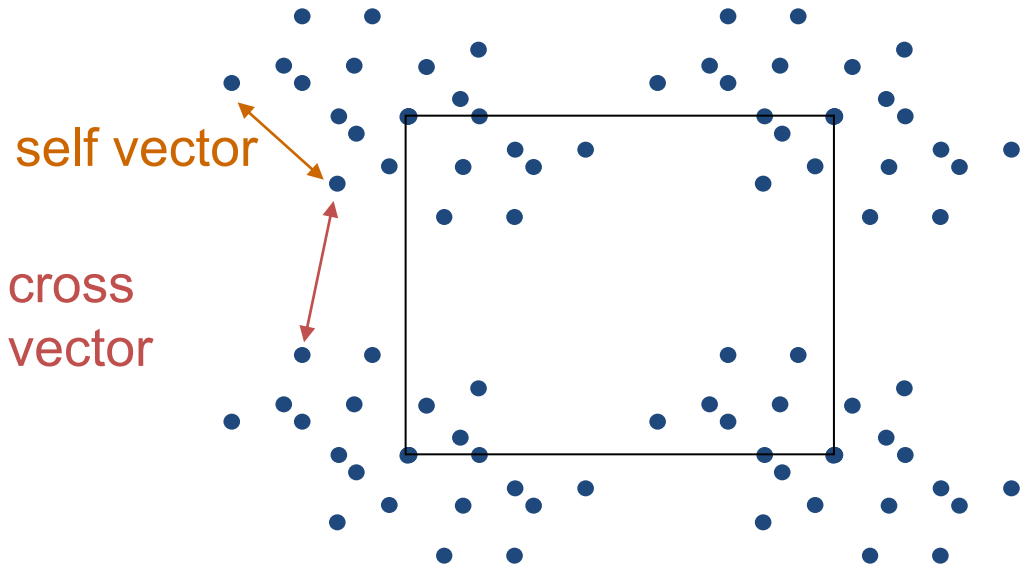
First, consider the model Patterson

We put the model in a large P1 box and calculate the Patterson from the structure factors of the model in the P1 box.



ROTATION FUNCTION

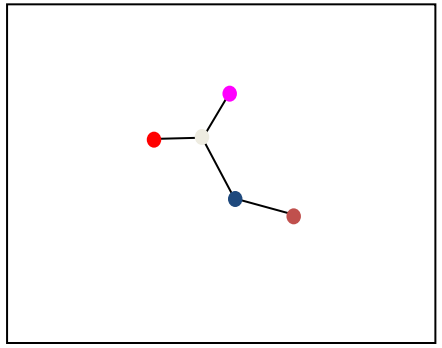
The *Patterson* of our unknown structure contains self-vectors and cross-vectors, but because the cell was large, the self-vectors and cross vectors are well separated from one another.



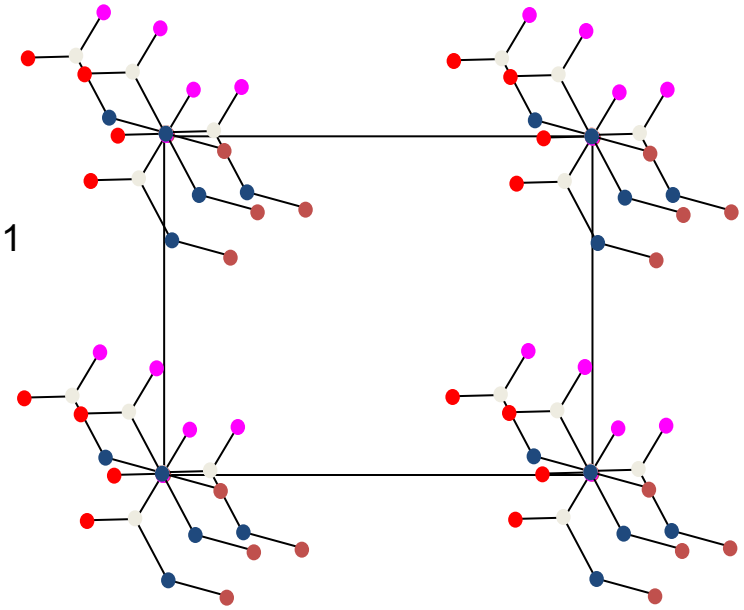
ROTATION FUNCTION

Just as we generated the Patterson for our model in the first orientation, we can generate the Patterson for the model in *any* orientation in any sized box.

model in same large P1 box
in different orientation

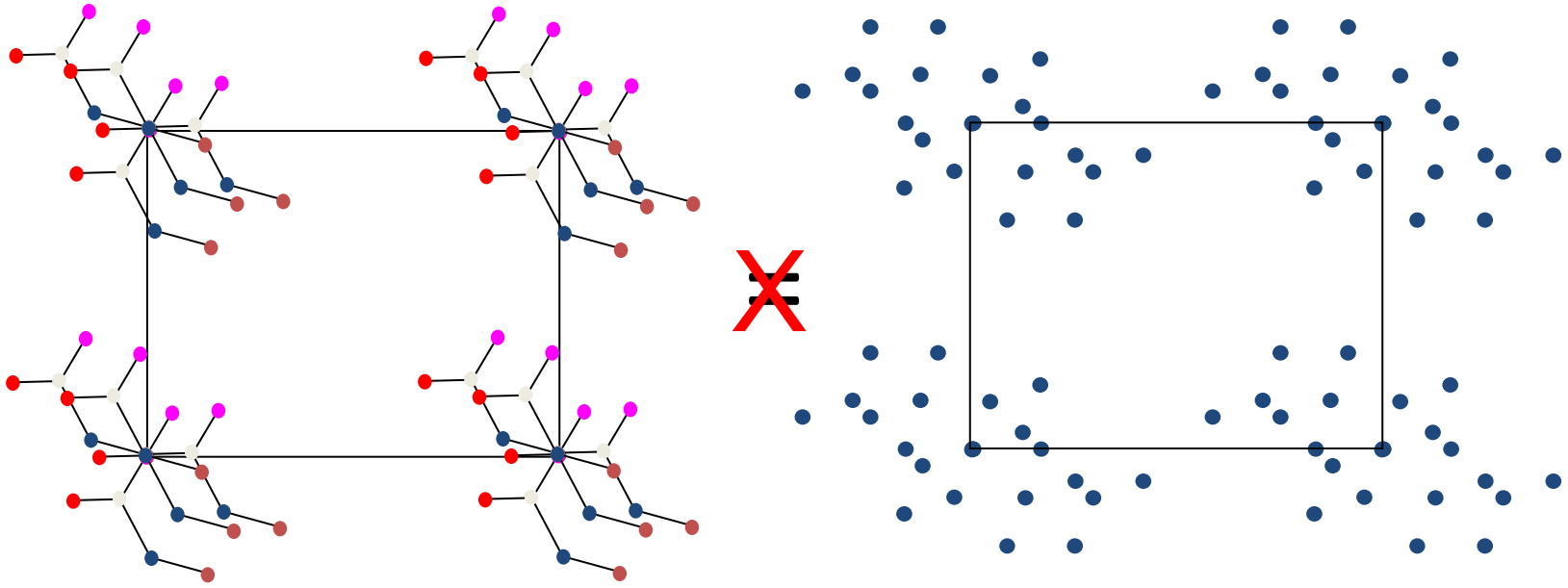


Patterson of
model in large P1
box
in different
orientation



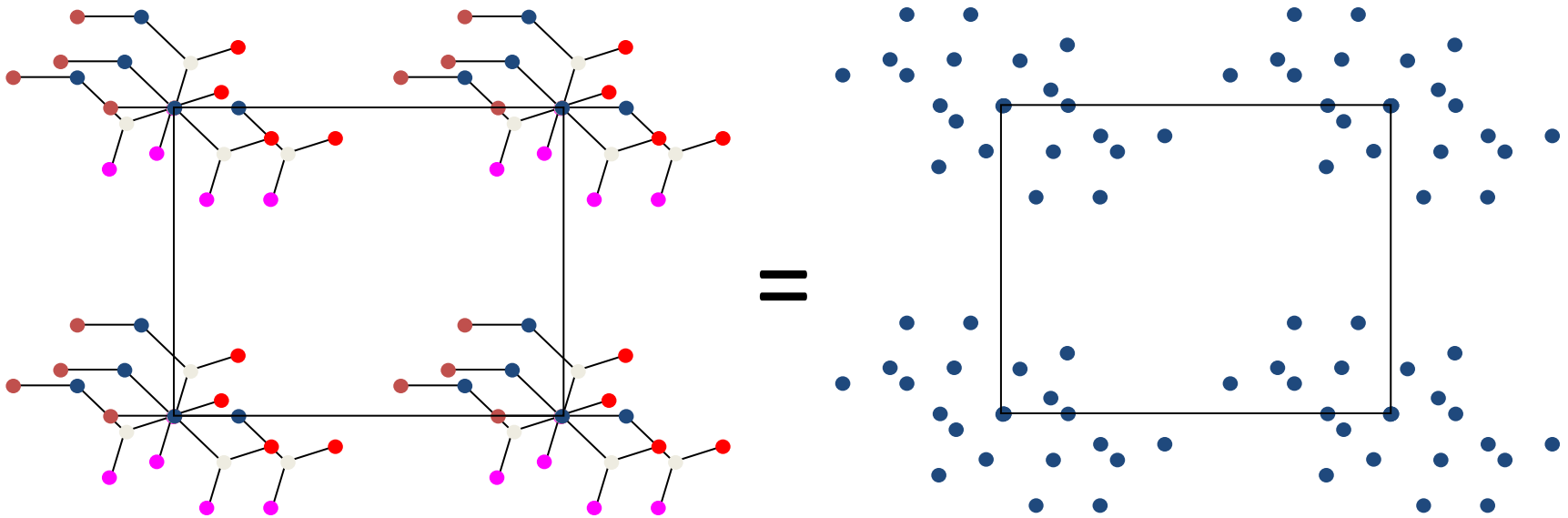
ROTATION FUNCTION

When the models are in different orientations the Pattersons will not match one another.



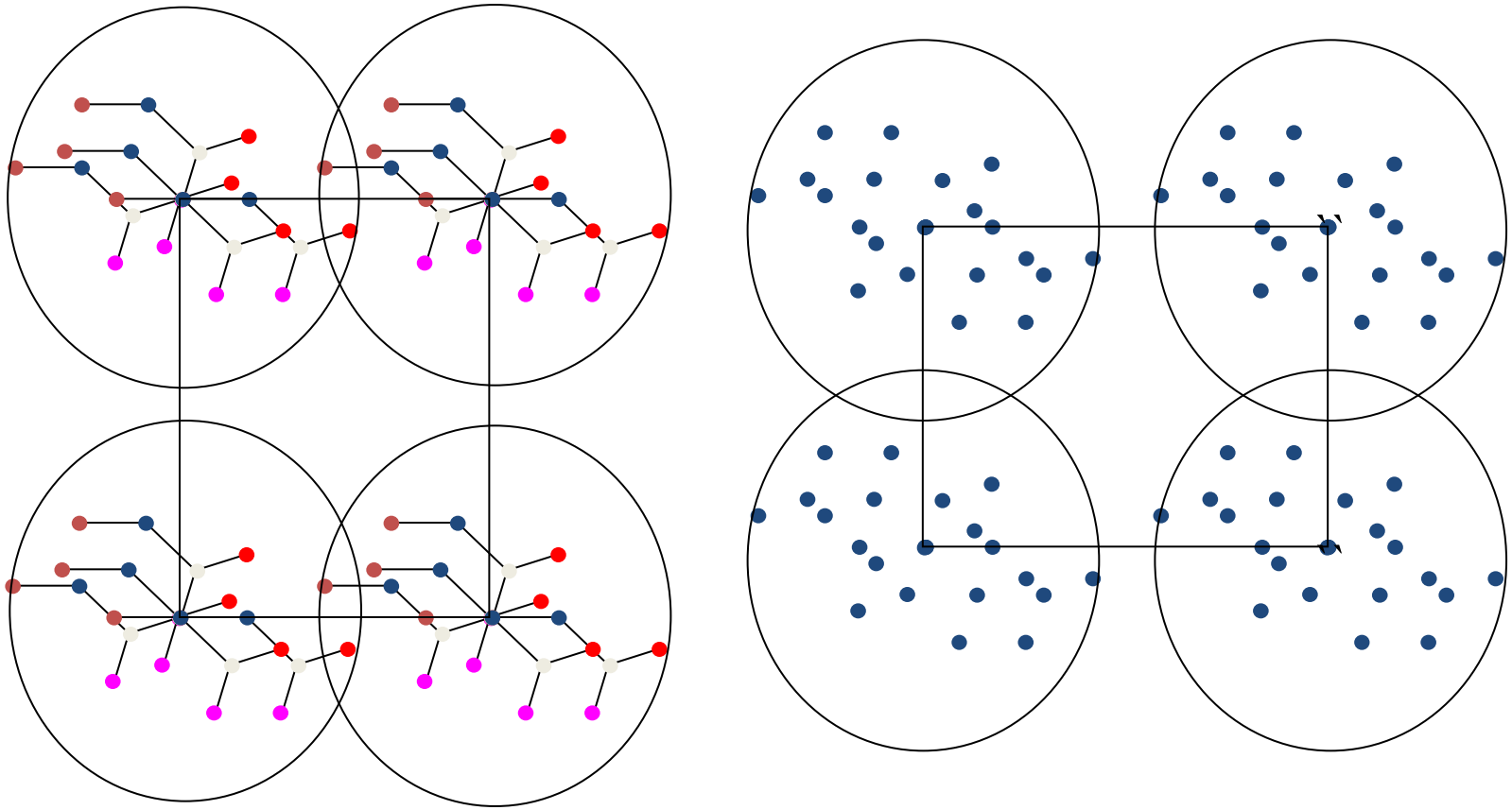
ROTATION FUNCTION

However, when the second model is in the same orientation parts of the Pattersons *will* match one another, and we can “solve” the rotation function for the model.



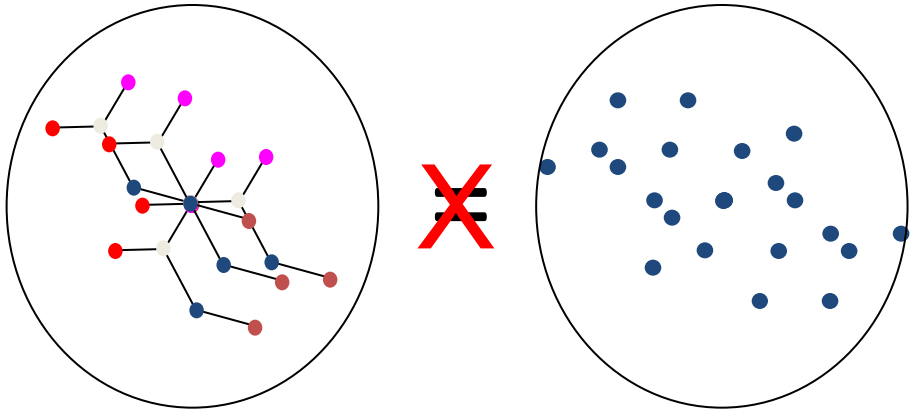
ROTATION FUNCTION

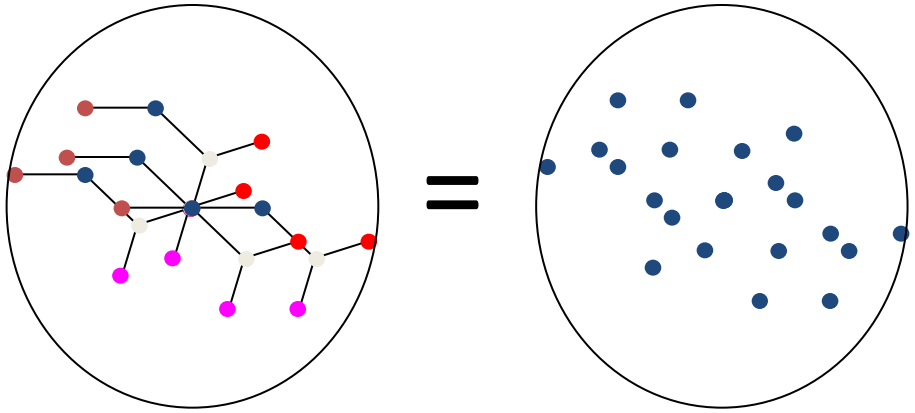
If the model were in a different sized box, the Patterson of the intramolecular vectors, which are located in a sphere centred on the origin, can be overlaid. We can cut out the peaks corresponding to the inter-molecular vectors from each Patterson and just compare the central parts of the Pattersons.



ROTATION FUNCTION

Now, the Pattersons of the intra-molecular vectors will match when the model is in the correct orientation.





Translation function

$$P_{1,2}(\mathbf{u}) = \int_V \rho_1(\mathbf{x}) \times \rho_2(\mathbf{x}+\mathbf{u}) d\mathbf{x}. \qquad (10.7)$$

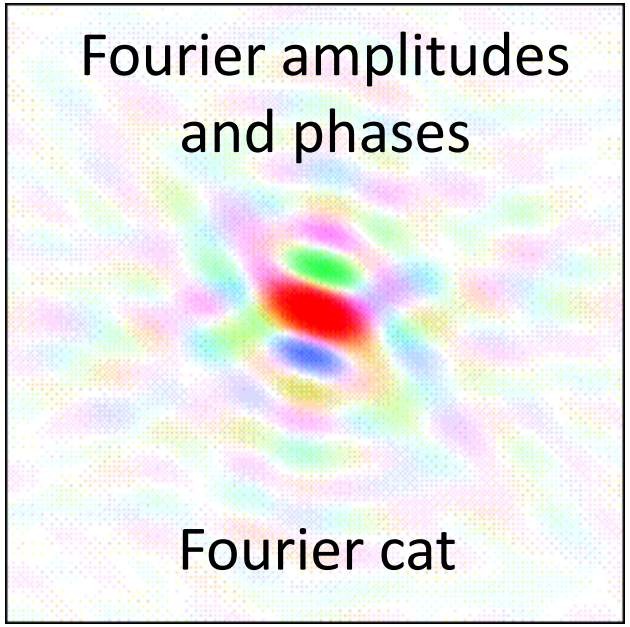
$$T(\mathbf{t}) = \int_V P_{1,2}(\mathbf{u}, \mathbf{t}) \times P(\mathbf{u})\, d\mathbf{u}$$

$$R = \frac{\sum_{hkl} \big|\,|F(\text{obs})| - k|F(\text{calc})|\,\big|}{\sum_{hkl} |F(\text{obs})|}.$$

Real space cat

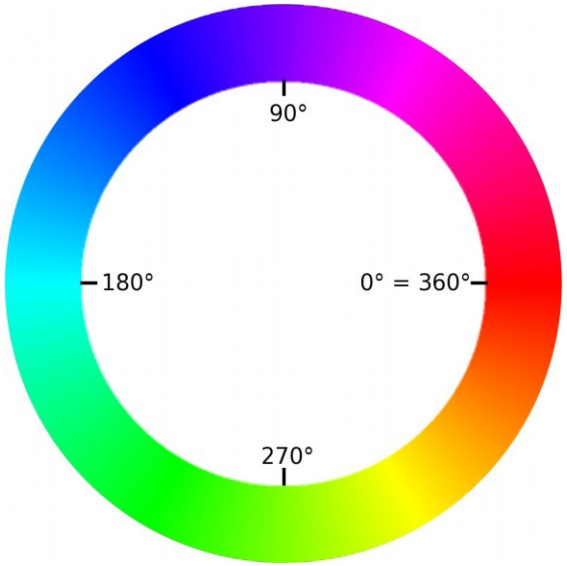
Fourier transform

Fourier amplitudes and phases



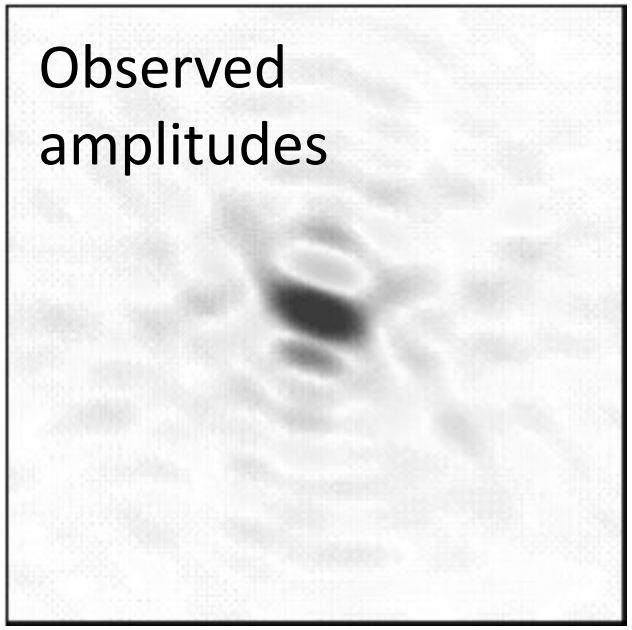
Fourier cat

Circular rainbow scale of phases



Linear intensity scale of amplitude size

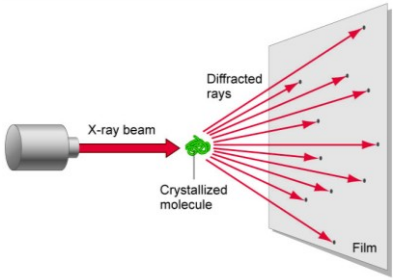
Observed amplitudes



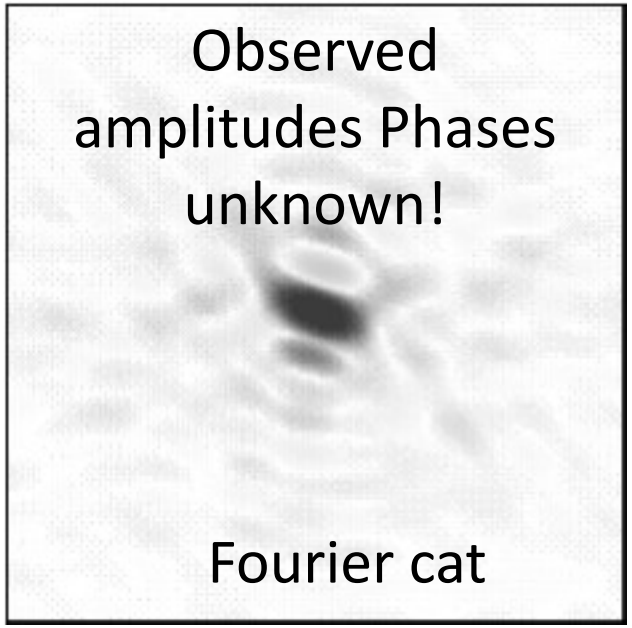
Unknown structure, unknown orientation

Cat

Diffraction experiment



Observed amplitudes Phases unknown!



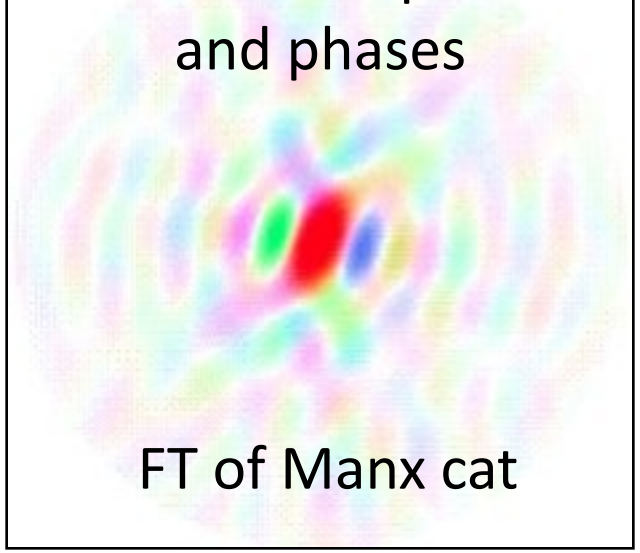
Fourier cat

Known structure

Manx cat

Fourier transform

Calculated amplitudes and phases



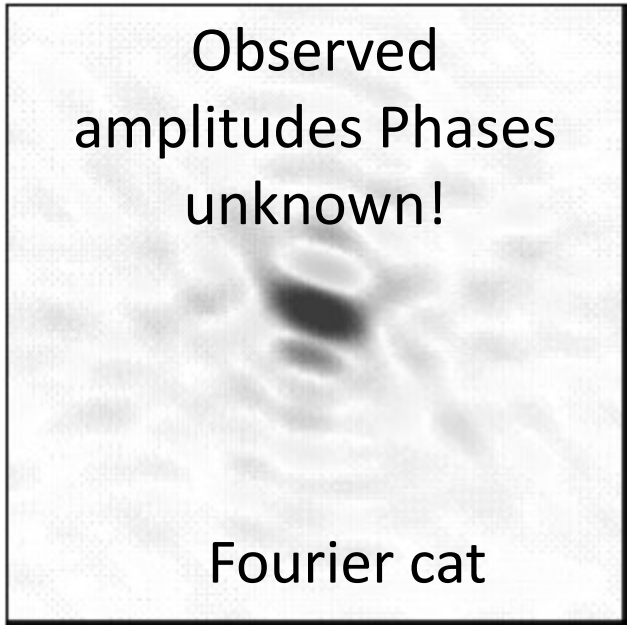
FT of Manx cat

Wrong orientation!

Known structure

Manx cat

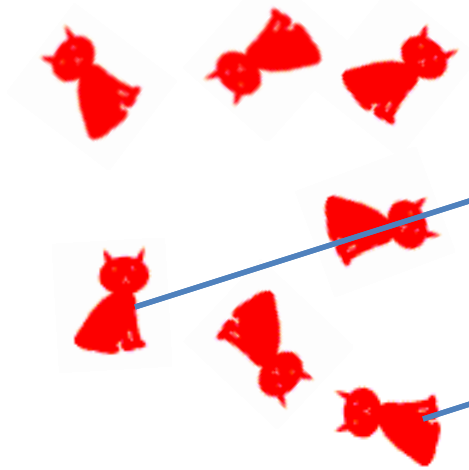
Observed amplitudes Phases unknown!

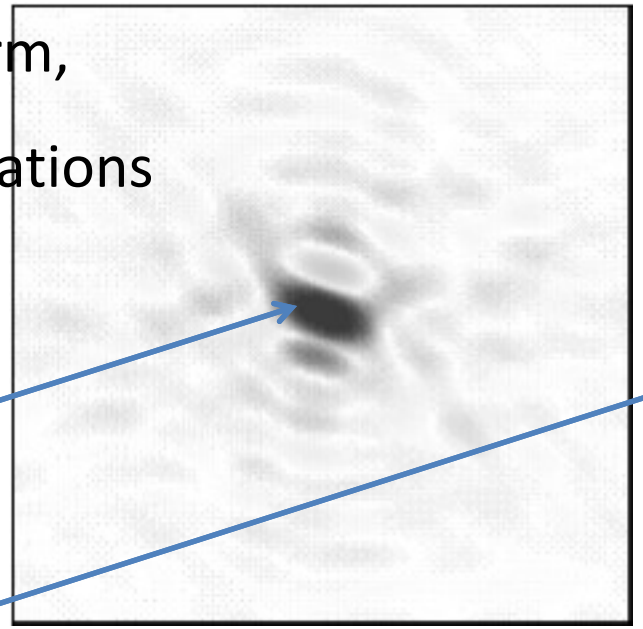


Fourier cat

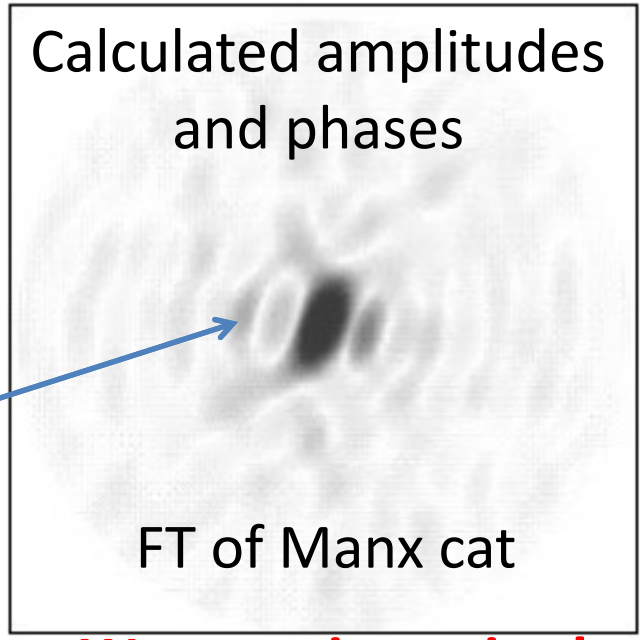
Fourier transform,

try different orientations





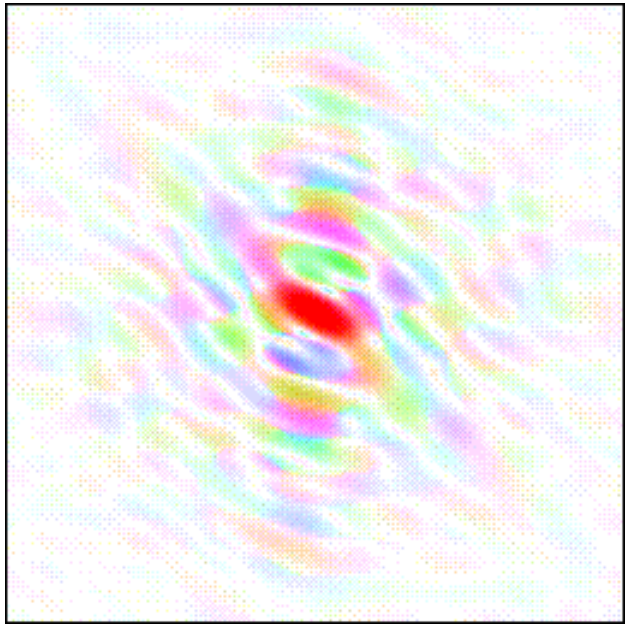
Calculated amplitudes and phases



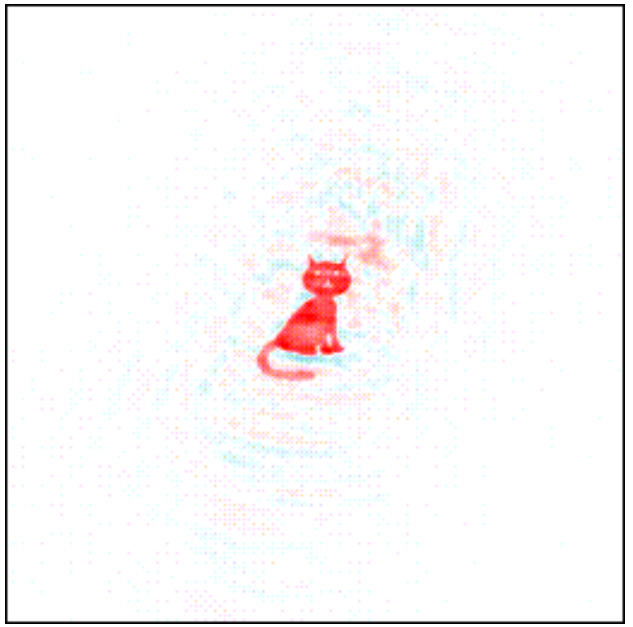
FT of Manx cat

Wrong orientation!

Observed **amplitudes** (tailed cat), calculated **phases** (Manx cat)



Inverted Fourier transform



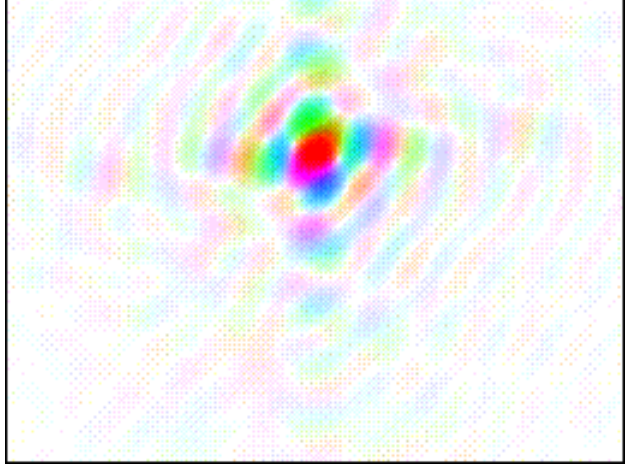
Even the tail becomes visible!

Model Bias

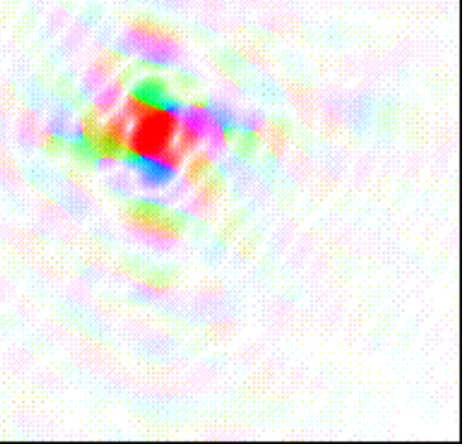
Duck

Fourier transform

Fourier transform of duck

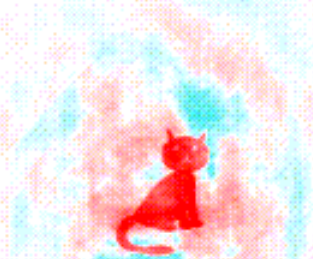


Duck amplitudes + cat phases

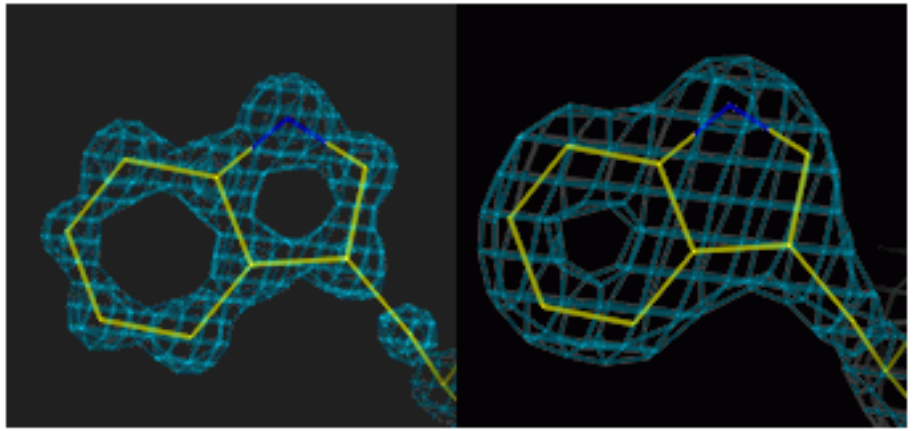


Inverted Fourier transform

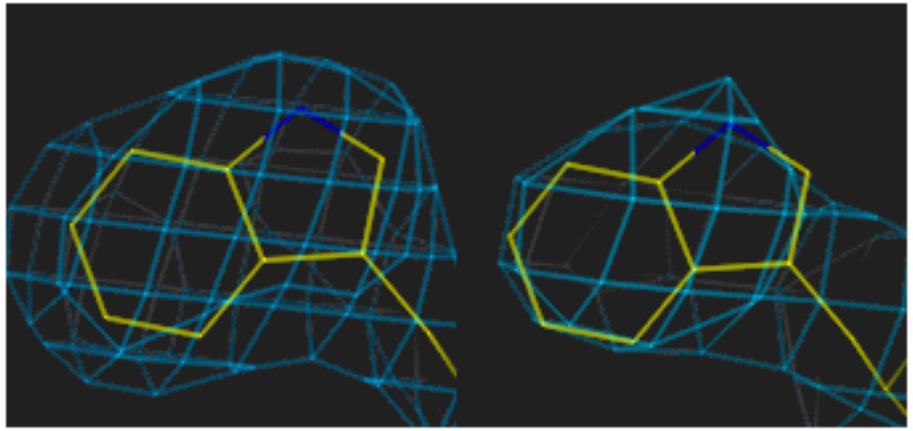
Looks like a cat!!



Model building & resolution



1.0Å 2.5Å



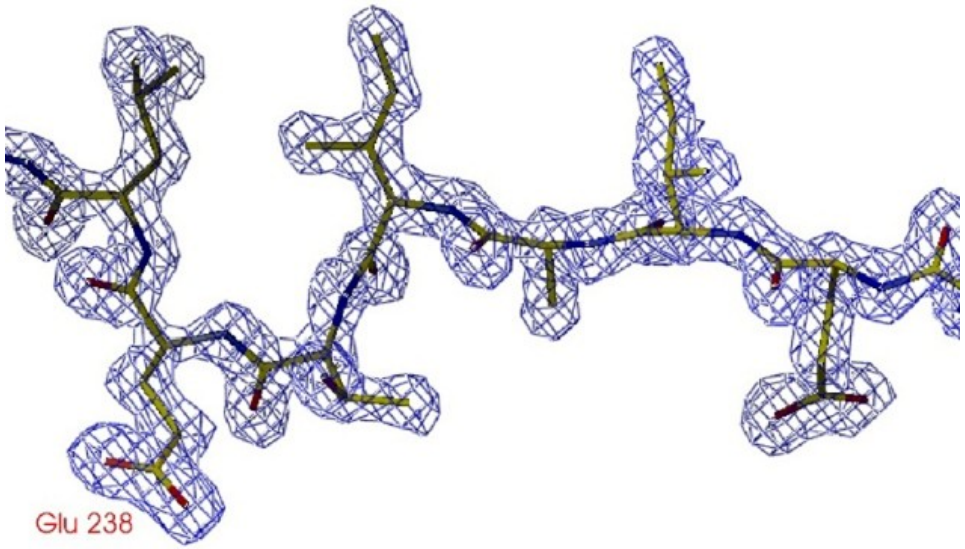
3.0Å 4.0Å

Phase improvement

When to use:

1. The structure is partially known.
2. The protein molecules distinguish themselves as relatively high regions of electron density and their boundaries can be estimated. The electron density between them is then set to a constant value or adjusted otherwise.
3. Noncrystallographic symmetry within the asymmetric unit is present. As in method 2, molecular boundaries must then be determined and the solvent region modified. Moreover, the density of all molecules (or subunits of a molecule) related by noncrystallographic symmetry is averaged.
4. Correct protein electron density maps have a characteristic frequency distribution for the values of the electron density (histogram matching).

Refinement of the Model Structure



$$R = \frac{\sum_{hkl} \| F_{obs} | - k | F_{calc} \|}{\sum_{hkl} | F_{obs} |} \times 100\%$$

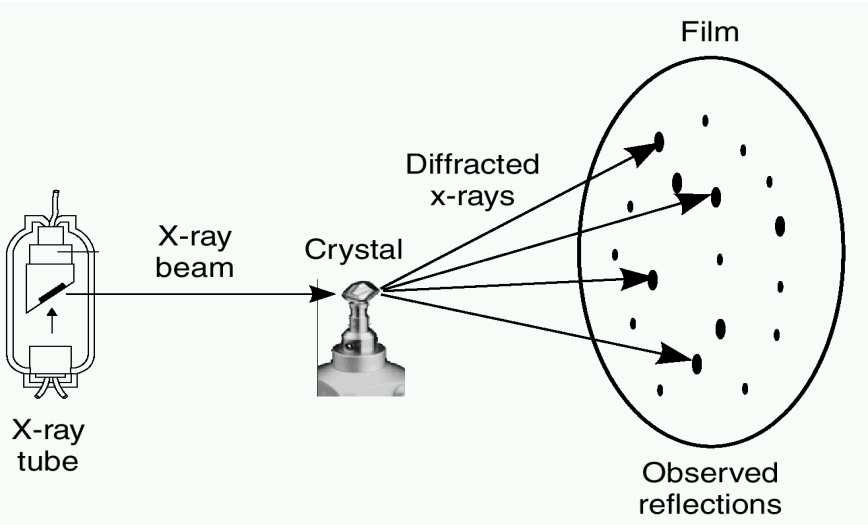
Parameter to measurement ratio – x, y, z, B (anisotropic B-> 8 parameters)

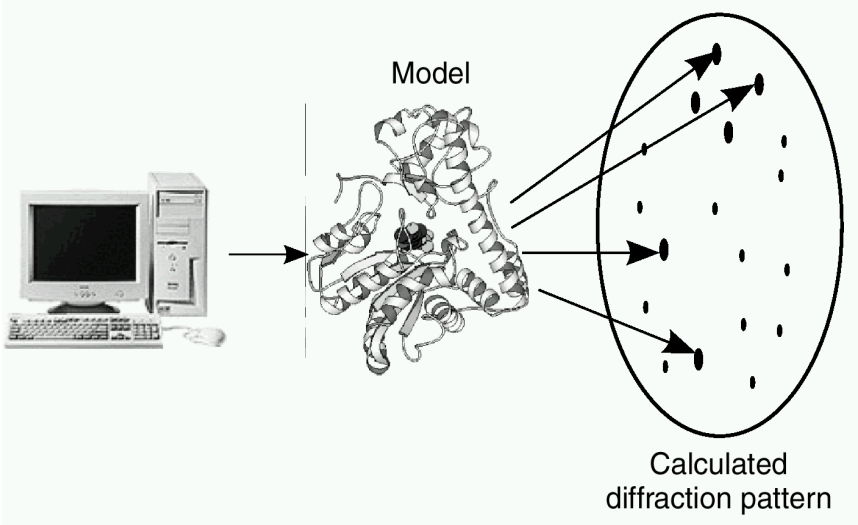
- other data than |Fobs|: stereochemical data (bond lengths and angles)
  solvent flattening
  NCS

Constrains x Restrains

- They are taken as rigid and only dihedral angles can be varied. In this case, the geometry and the refinement are called *constrained*. This effectively reduces the number of parameters to be refined. In the application of this method, it is difficult to move small parts of the structure to a "best fit" position because many angular motions are involved.
- If, on the other hand, the stereochemical parameters are allowed to vary around a standard value, controlled by an energy term, the refinement is called *restrained*. The atomic coordinates are the variables and the restraints are on bond lengths, bond angles, torsion angles, and van der Waals contacts. Restraints are "observations" because a penalty is included for disagreement with a restraint. This allows an easy movement of small parts of the structure, but it is difficult to move large parts (e.g., an entire molecule or domain).

R-factor, R_{free} factor





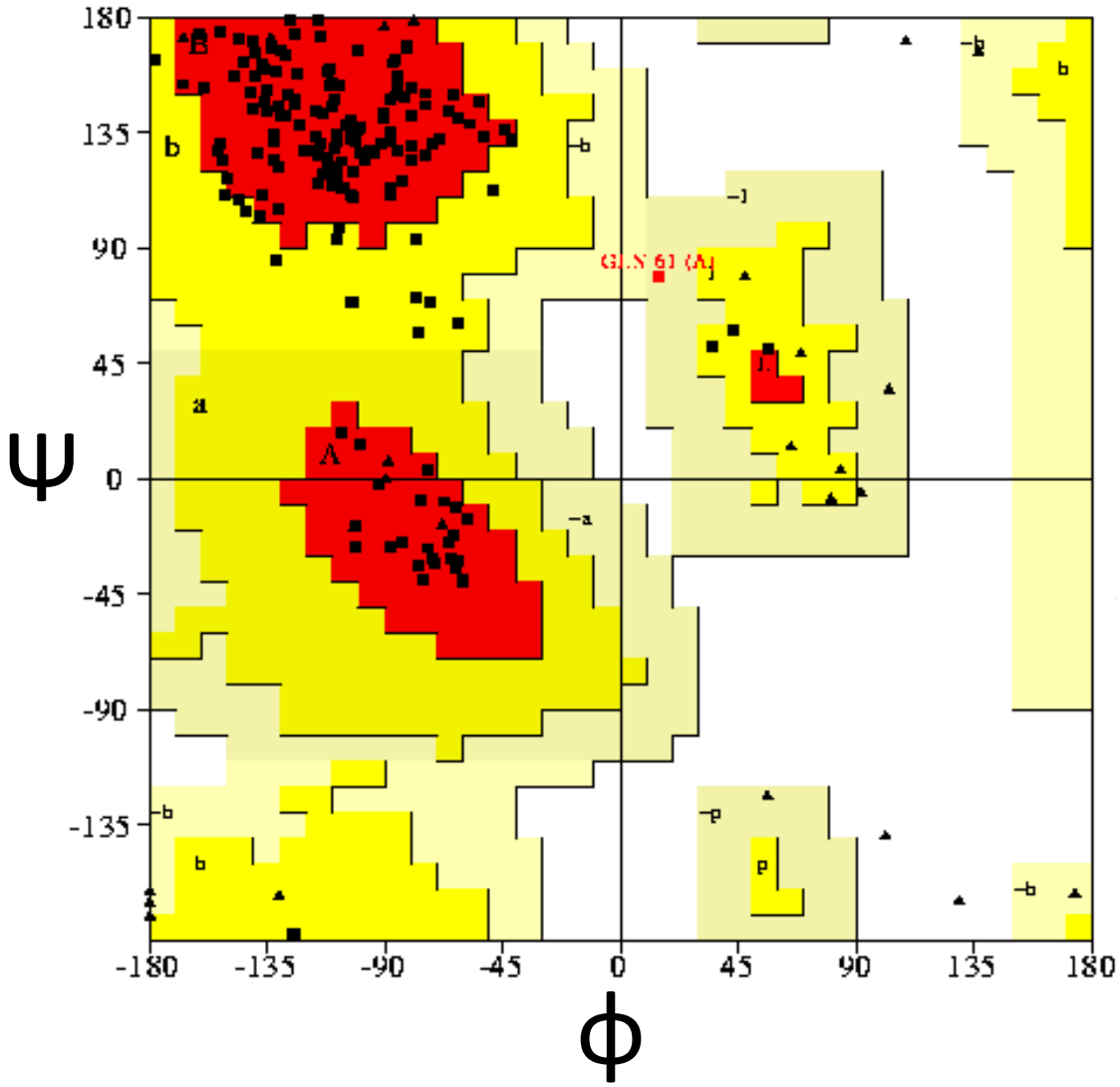
R-factor

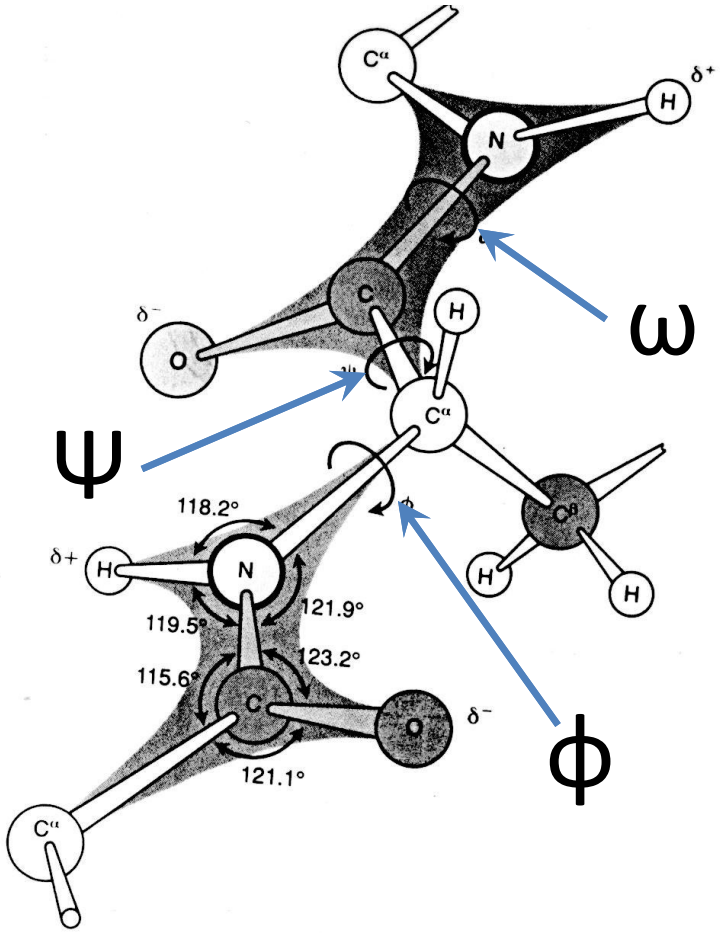
$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

R_{free} factor

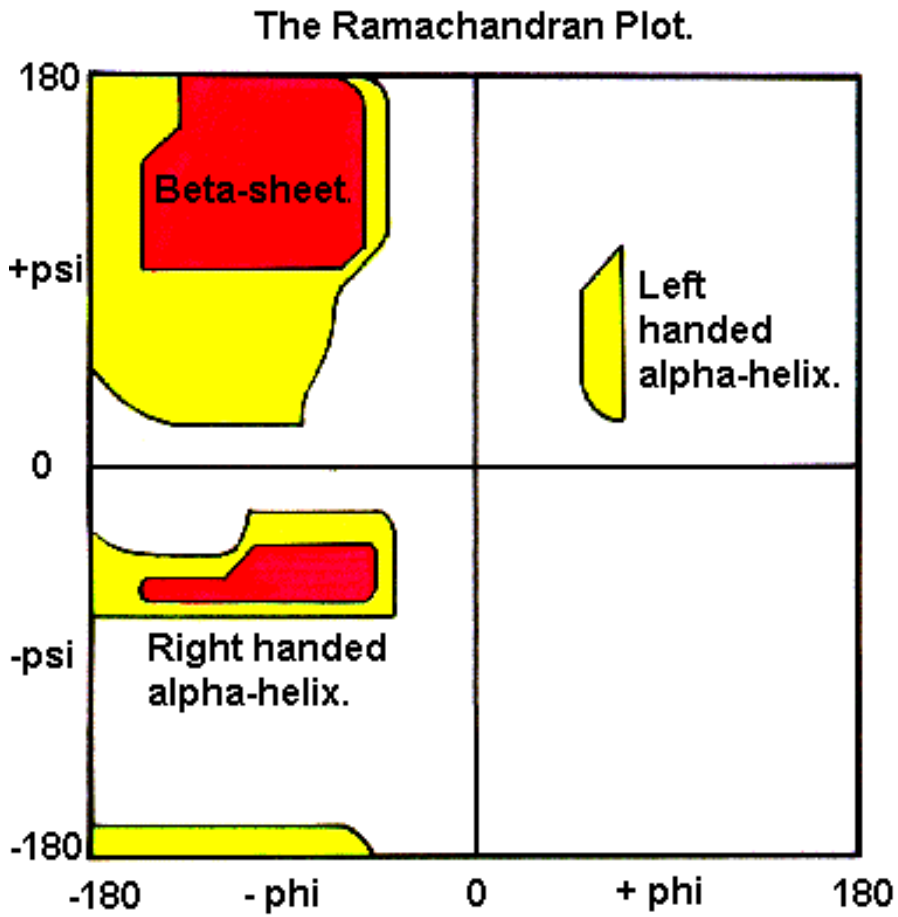
$$R_{free} = \frac{\sum_{hkl \subset T} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl \subset T} |F_{obs}|}$$

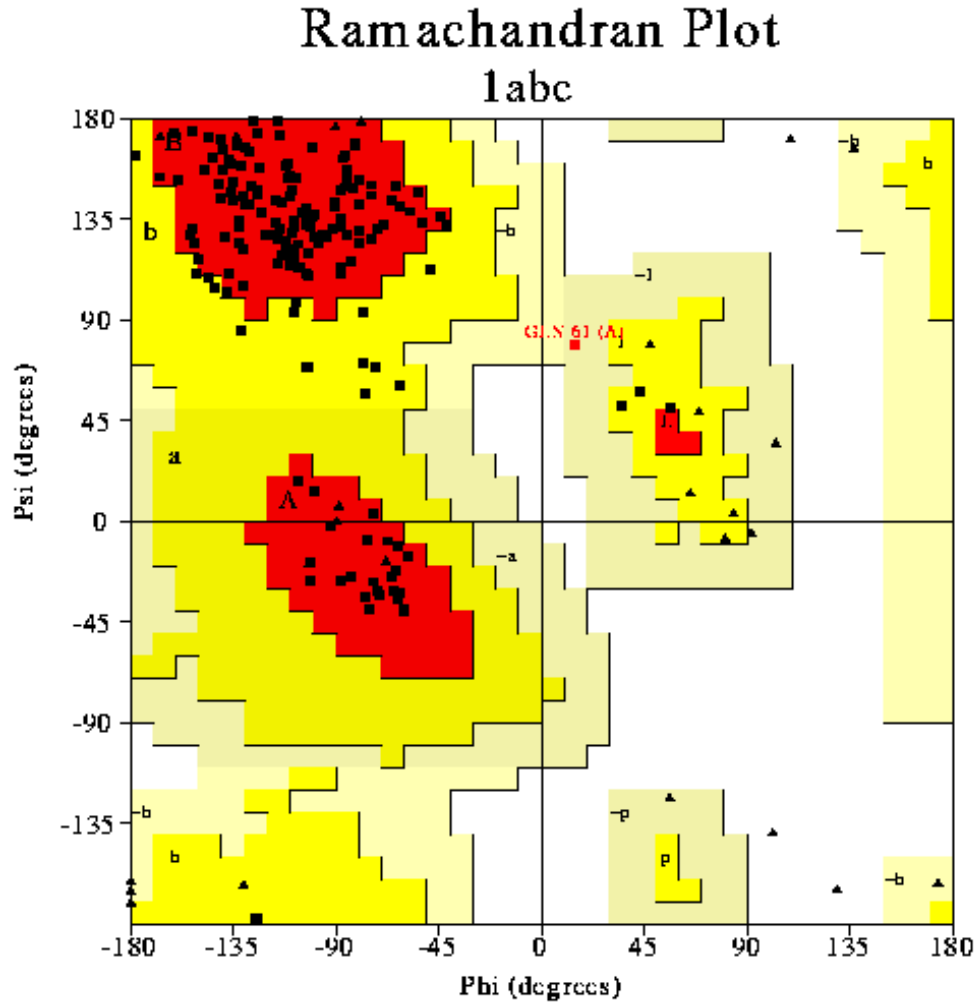
Ramachandran plot





The Ramachandran Plot



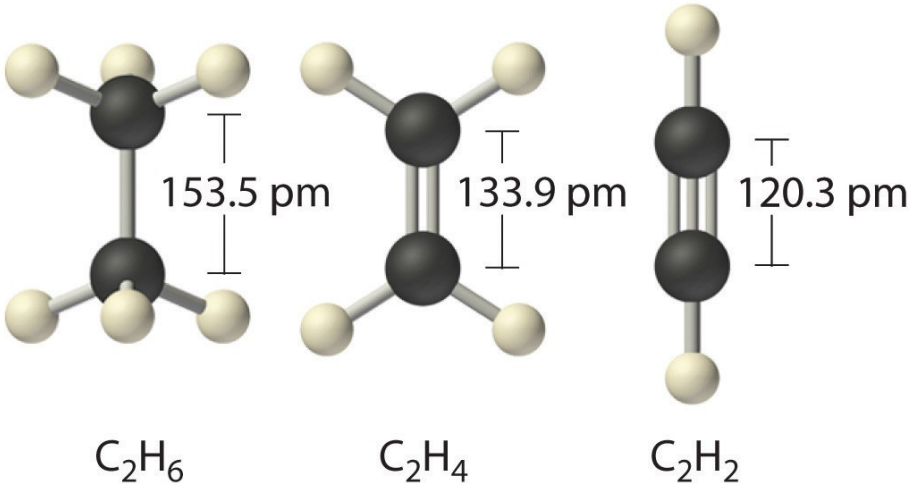


Plot statistics

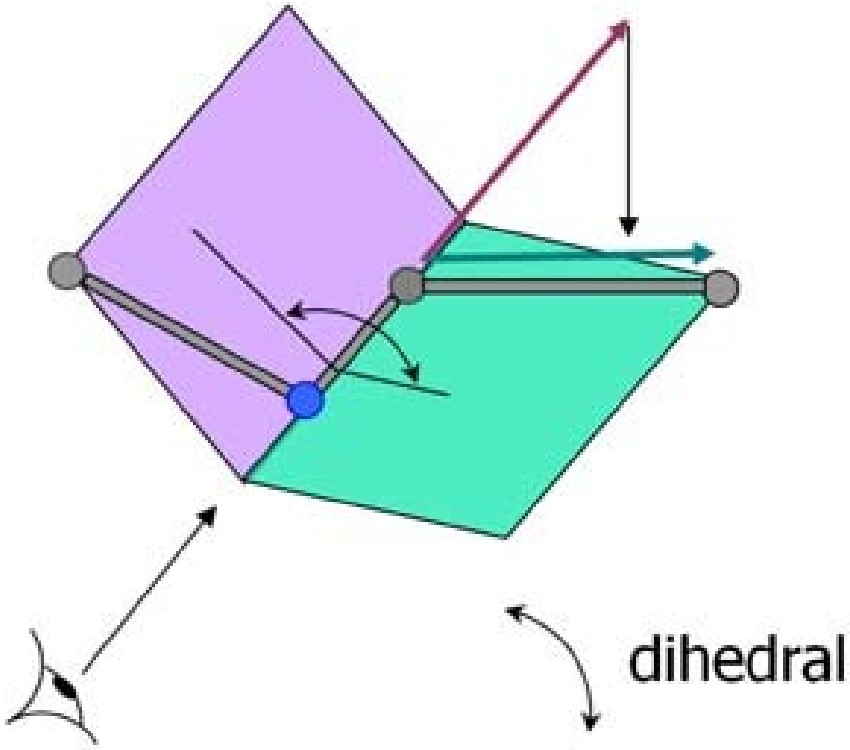
Residues in most favoured regions [A,B,L]	143	89.9%
Residues in additional allowed regions [a,b,l,p]	15	9.4%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.6%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	159	100.0%
Number of end-residues (excl. Gly and Pro)	5	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	15	
Total number of residues	205	

Geometry and stereochemistry

Bond lengths



Dihedral angles



$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}.$$

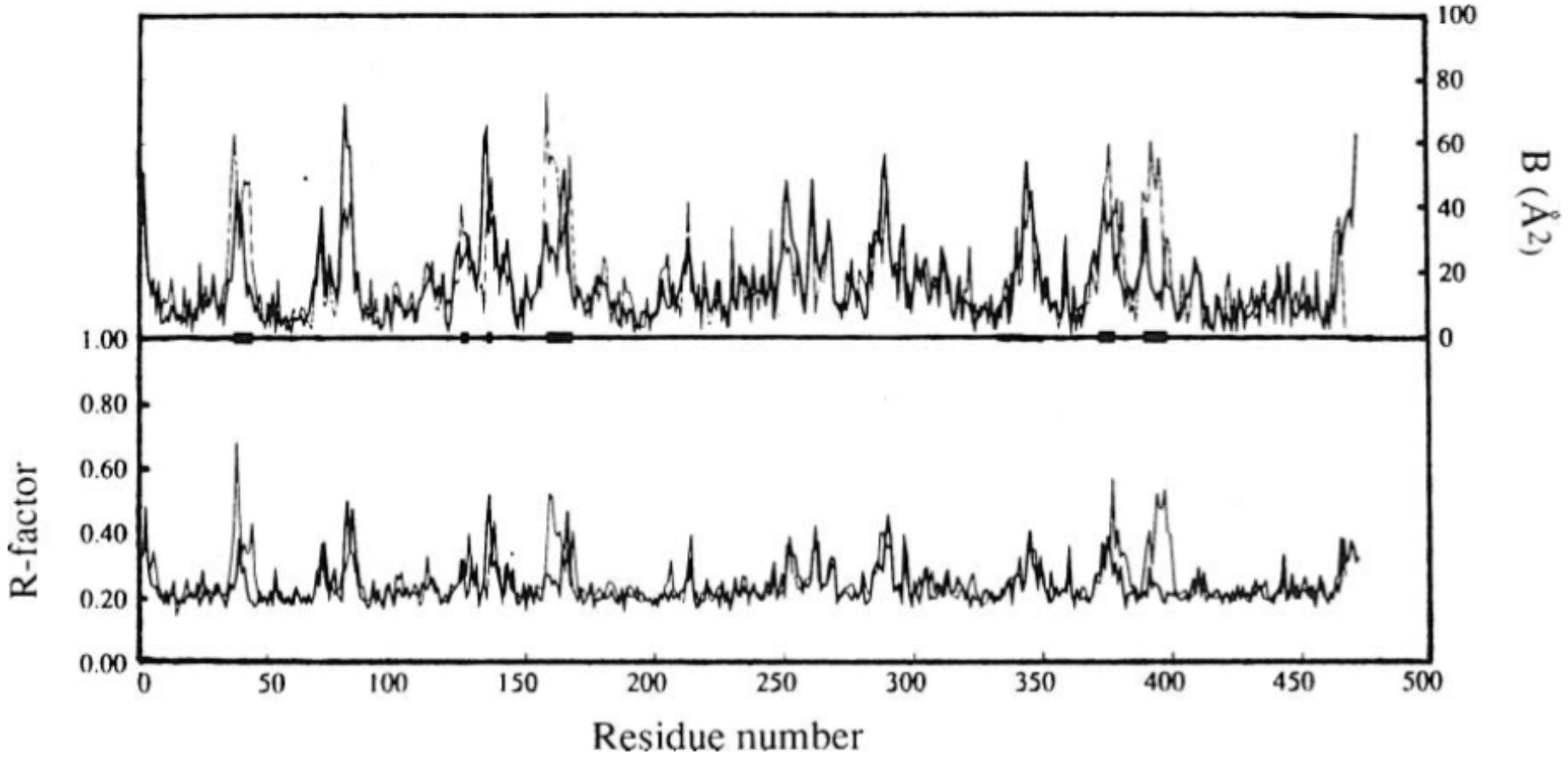


Figure 15.1. Real-space *R*-factor (lower panel) and average *B*-factor (upper panel) of the *Azotobacter vinelandii* lipoamide dehydrogenase. Misplaced loops are indicated by a thin line, and after their correction, they are indicated by a thick line. Note the correspondence between the *R*-factor and the *B*-factor at the problem sites. (Courtesy of Dr. Andrea Mattevi.)

$$R_{\text{real space}} = \frac{\sum |\rho_{\text{obs}} - \rho_{\text{calc}}|}{\sum |\rho_{\text{obs}} + \rho_{\text{calc}}|}$$

Quality of diffraction data

R-Factor for Comparing the Intensity of Symmetry-Related Reflections

$$R_{\mathrm{sym}}(I) = \frac{\sum_{hkl}\sum_{i} |I_i(h\,k\,l) - \overline{I(h\,k\,l)}|}{\sum_{hkl}\sum_{i} I_i(h\,k\,l)}$$

for n independent reflections and i observations of a given reflection. $\overline{I(h\,k\,l)}$ is the average intensity of the i observations.