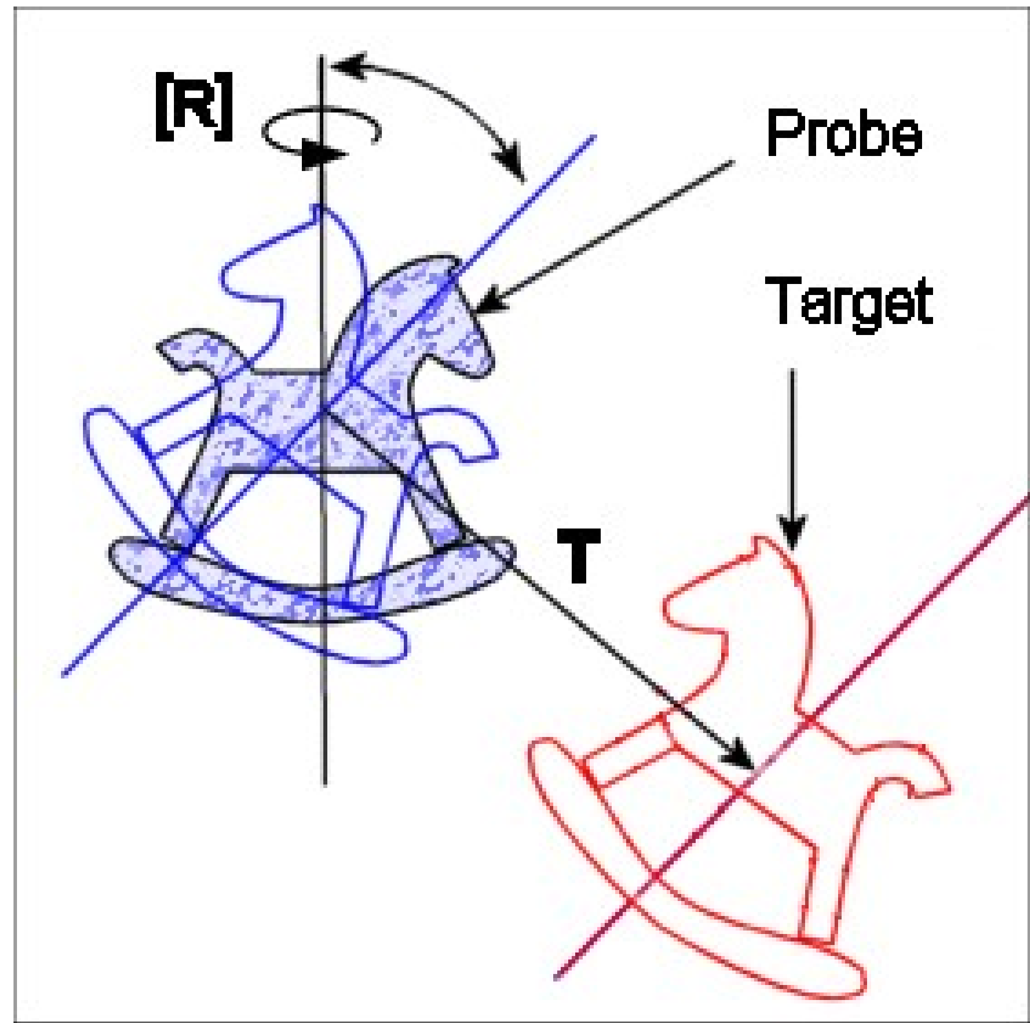
Phasing/1: Molecular Replacement



Molecular replacement

Phasing method that uses a molecular probe (a model protein structure) to obtain an initial estimation of phases.

The probe needs to be correctly located in the unit cell before calculating phases.

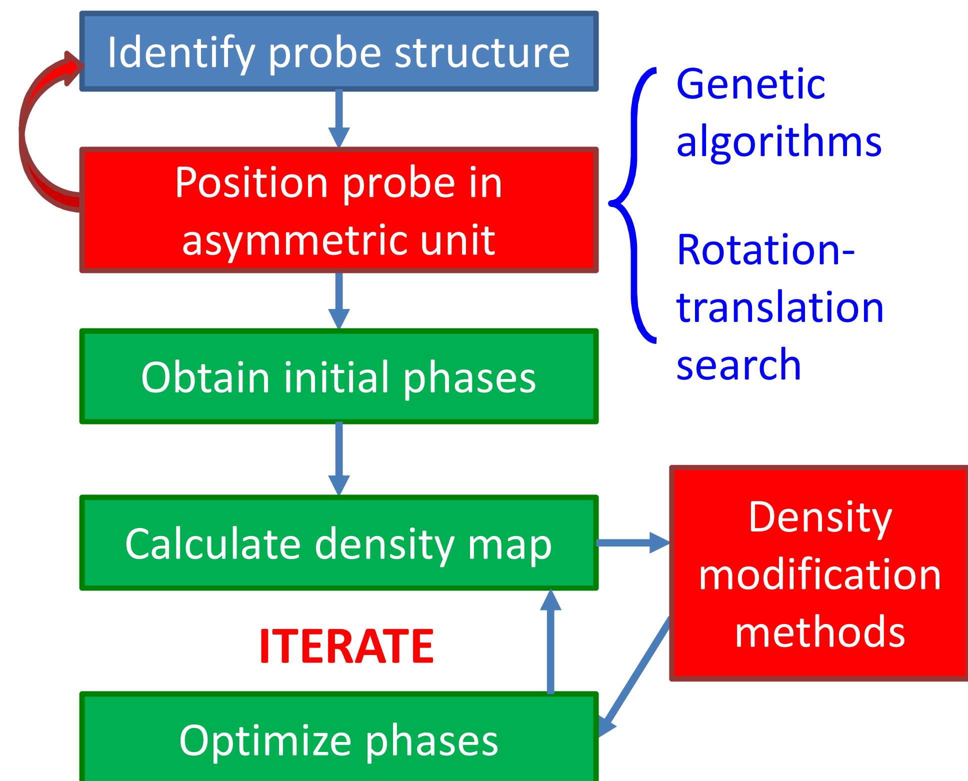
Replacement → relocating a molecular probe

Requirements:

- suitable probe with sufficient structural homology with the target structure: usually a crystallographic structure, but also NMR or computational structures can work;
- good data at resolution < 4 Å.

Issues:

- unsuitable model;
- model bias;
- presence of multiple copies (NCS-related) may hamper phasing with MR.



Probe selection and preparation

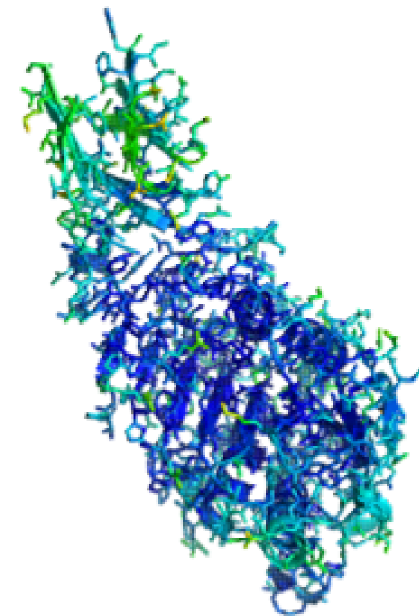
Ideal probe:

- **High homology**: > 30% identity (*but structural homology may be higher than sequence homology!*)
- **Good accuracy**: X-ray structures usually are optimal, but also NMR and computational homology modeling (when done wisely...)

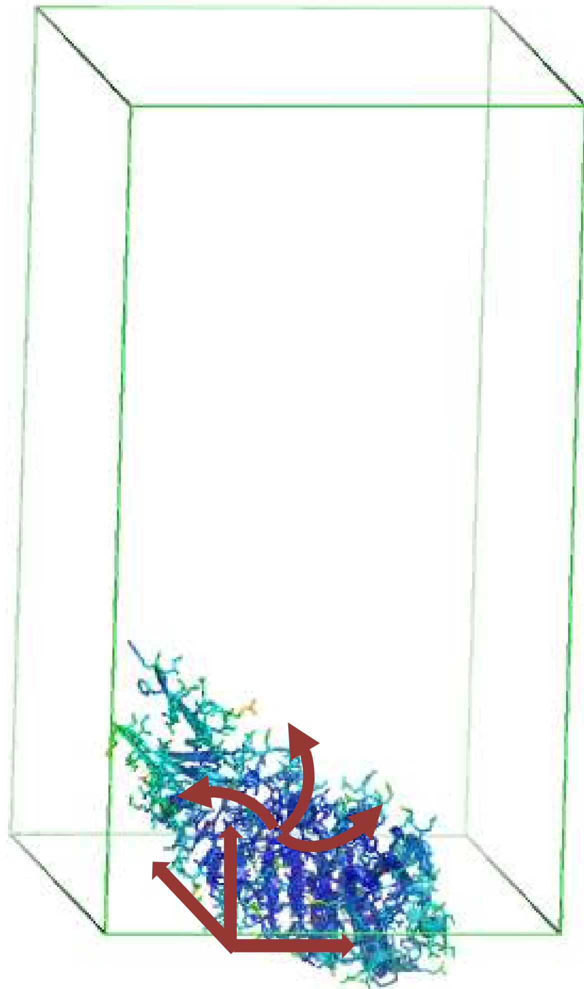
Use bioinformatics tools for alignment, find highest homology model (and test more than one...)

Probe preparation:

- Remove solvent molecules, ions, ligands...
- If the structure is oligomeric, evaluate the interaction surface: if oligomer is stable, use entire oligomer, otherwise select monomer
- Remove parts expected to be different from target (alignment)
- Analyse B-factors of crystallographic model and remove flexible parts
- Sometimes useful to convert to all-Ala



Genetic algorithms



6 variables problem:

3 rotation variables

and 3 translation variables in the unit cell

1) Initial starting population of different rotations and translations of the model.
2) Scoring of the population against data; different scoring methods.
3) Best results are kept for next round and mutations are introduced (small variations of parameters).

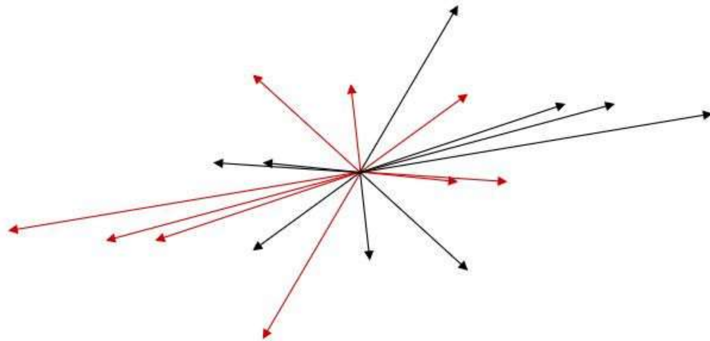
Different algorithms, with different scoring functions.

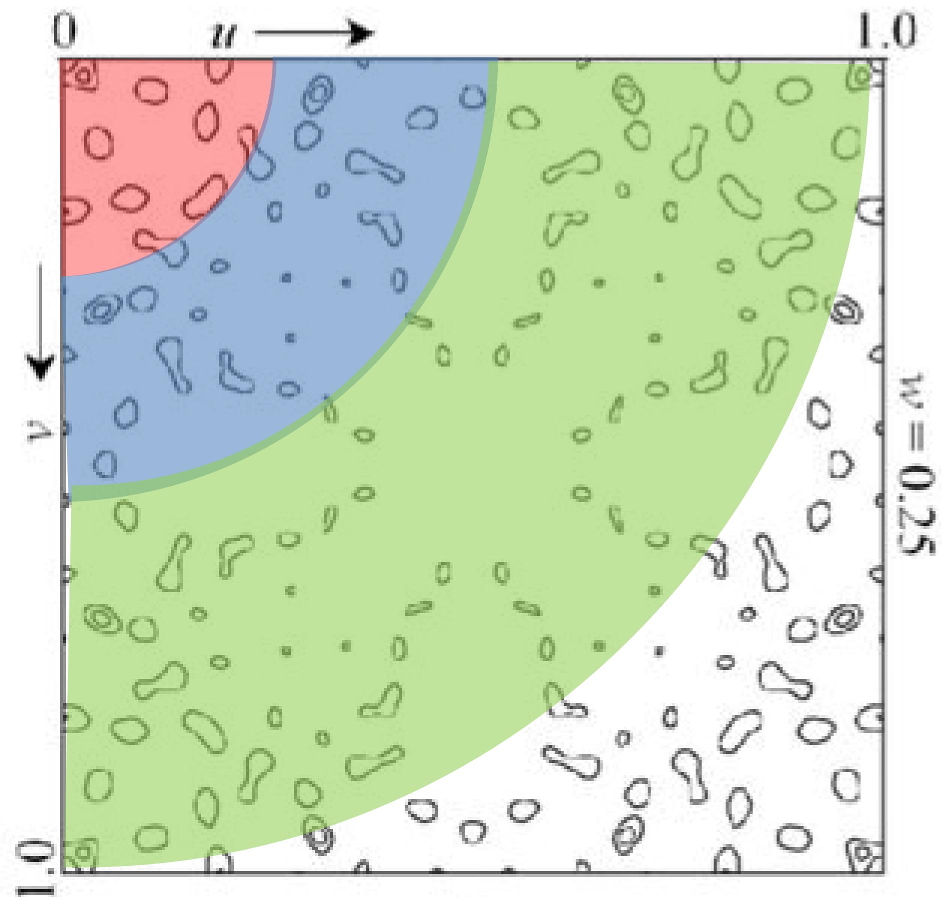
But the method is effective only for small unit cells, otherwise becomes very computationally intensive.

Rotation-translation searches

In the more commonly used MR methods, rotation variables (3) are determined first and results of the rotation undergo a translational search (3 more variables). The dimensionality of the problem is reduced: faster!

Patterson map: peaks of the Patterson map are interatomic vectors.



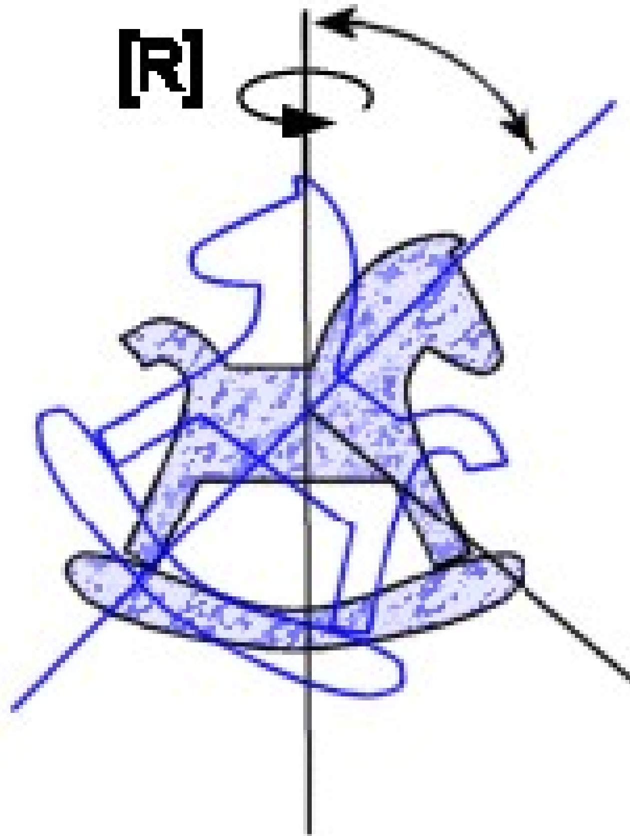


Intramolecular vectors: atoms of the same molecule are closer, intramolecular vectors smaller -> in the Patterson map, peaks closer to the origin (e.g. 4-10 Å)

Intermolecular vectors: longer vectors connecting -> in the Patterson map peaks further from the origin (e.g. > 10 Å)

*Below 4 Å intra-residue peaks, not informative

Rotation search



1) Calculate Patterson map from experimental intensities (**experimental Patterson**); consider interatomic vectors between 4 Å and an upper limit that depends on the dimension of the protein (about 75% of protein radius)
2) From the model, calculate **theoretical Patterson** map (initial position of the molecule in the cell)
3) Rotate the calculated Patterson map by **small intervals in 3 rotation angles** (Euler angles) and compare with experimental Patterson (i.e. calculate **cross-correlation**)

Best results with globular proteins (intra- and inter-molecular vectors are distinct).

Final result of this step: rotation angles

(usually multiple solutions with relative scoring...)

Translation search

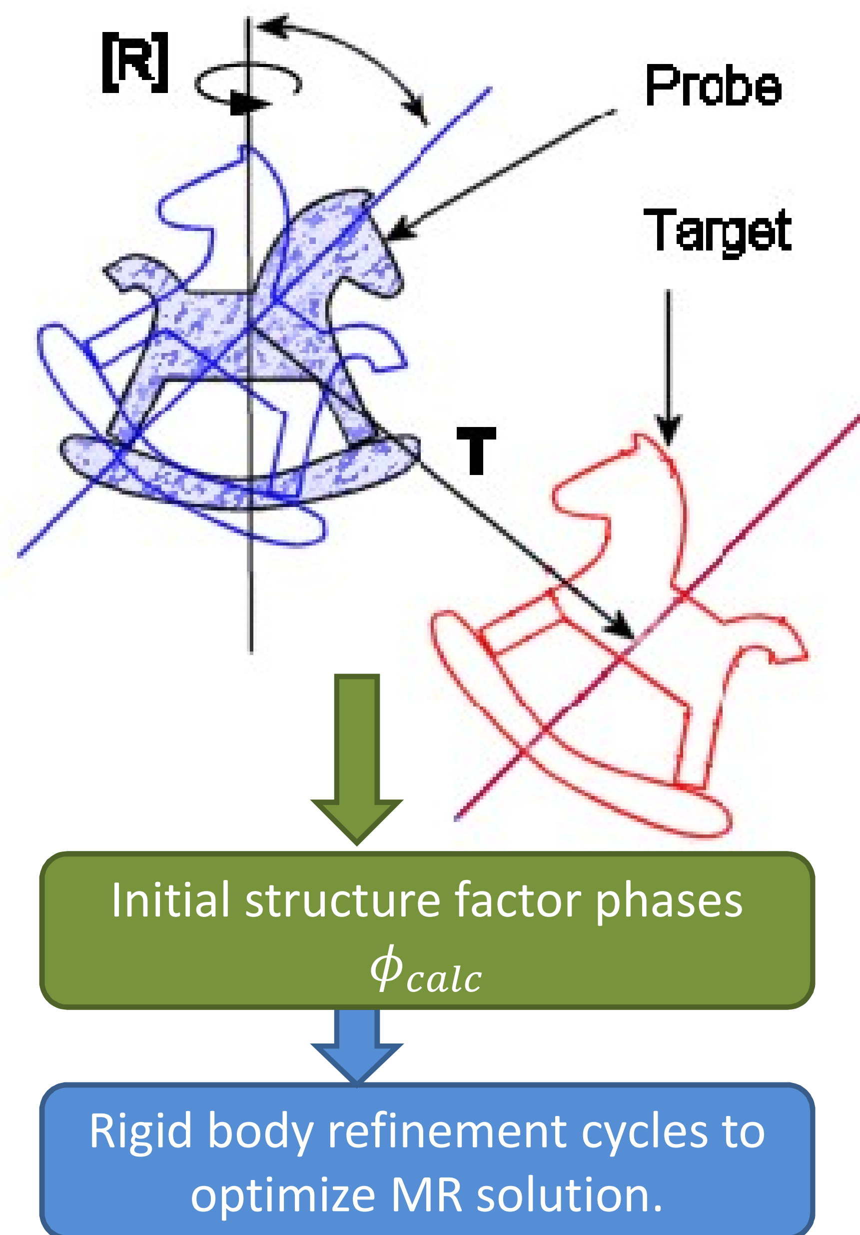
Starting from the rotational solution (or solutions if NCS is present), determine translation vector.

Analysis of the intermolecular vectors of the experimental Patterson map (beyond the limit of the rotational search).

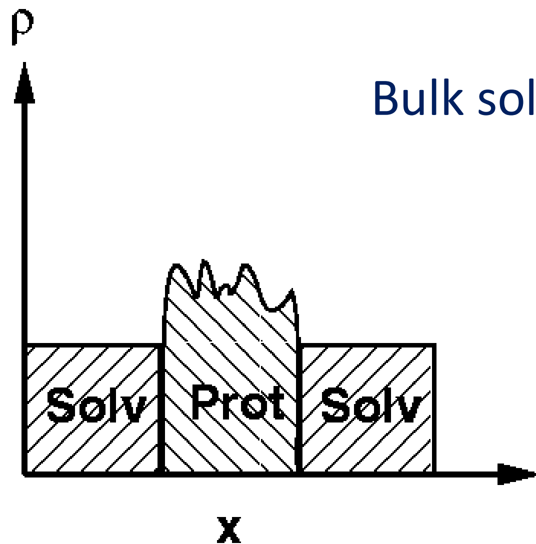
In *non-polar space groups* (i.e. with fixed coordinates for the origin) the first copy of the molecule has to be located with respect to the symmetry elements present in the unit cell.

In *polar space groups* one (or more) coordinates of the origin are arbitrarly assigned. Further NCS molecules need to be related to the first.

Usually more difficult than rotational part, but density modification methods (or independent knowledge of phases) can significantly improve success rate.



Solvent correction



Bulk solvent: in large channels present in the protein structure.

Diffraction data contain an average of all possible positions of solvent molecules in large channels. Electron density (in real space) in these regions is flat.

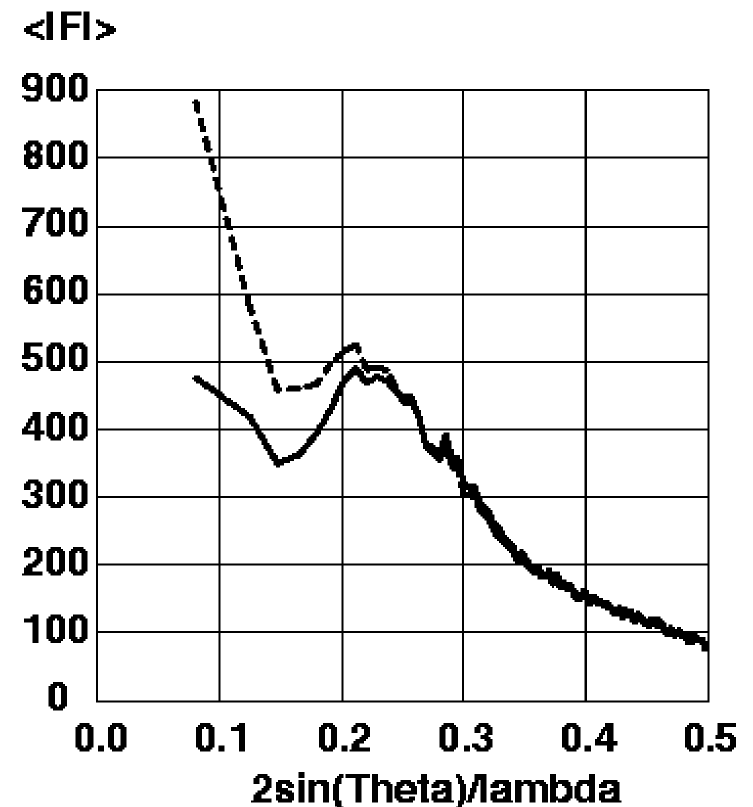
Average value of electron density for solvent (0.33 $e^-/Å^3$) is lower than for protein (0.43 $e^-/Å^3$).

Calculated structure factors (in reciprocal space): from a model that considers the protein in vacuum.

Effect of the solvent negligible at high resolution: dampening of signal due to high B-factor for solvent.

At low resolution, calculated structure factors (dashed line): overestimation of the contrast at protein boundaries, compared to experimental structure factors (solid line).

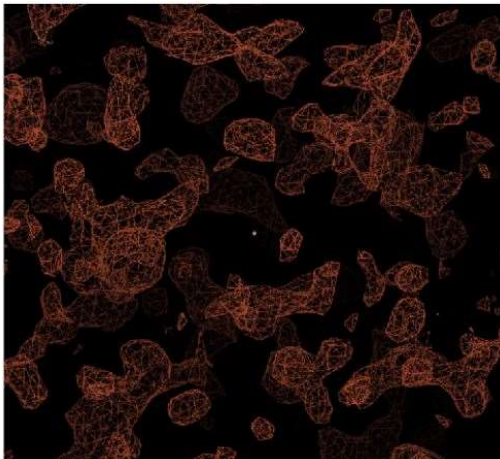
Correction of calculated structure factors improves electron density map obtained by calculated model!



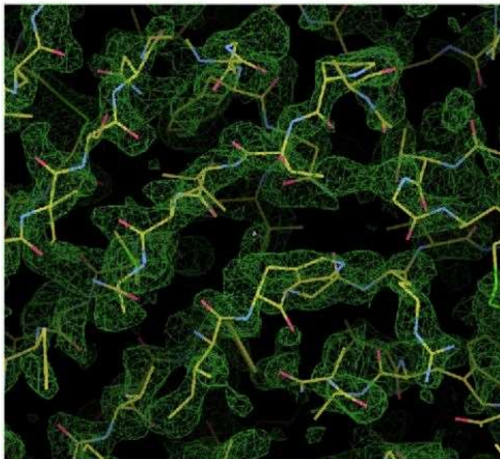
Density modification

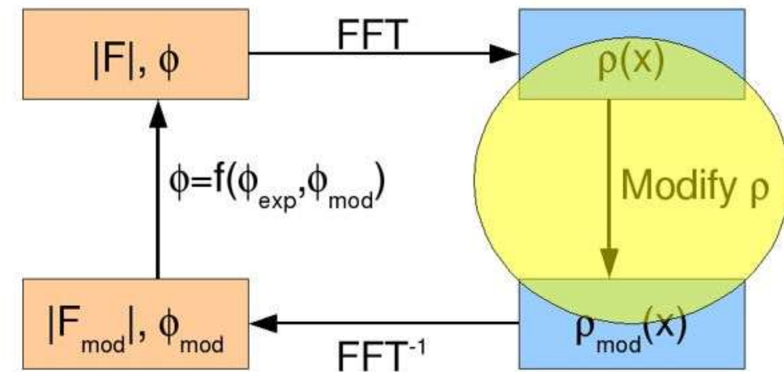
Complementary techniques useful to **improve and/or extend phases** from molecular replacement or experimental phasing.

B
E
F
O
R
E



A
F
T
E
R





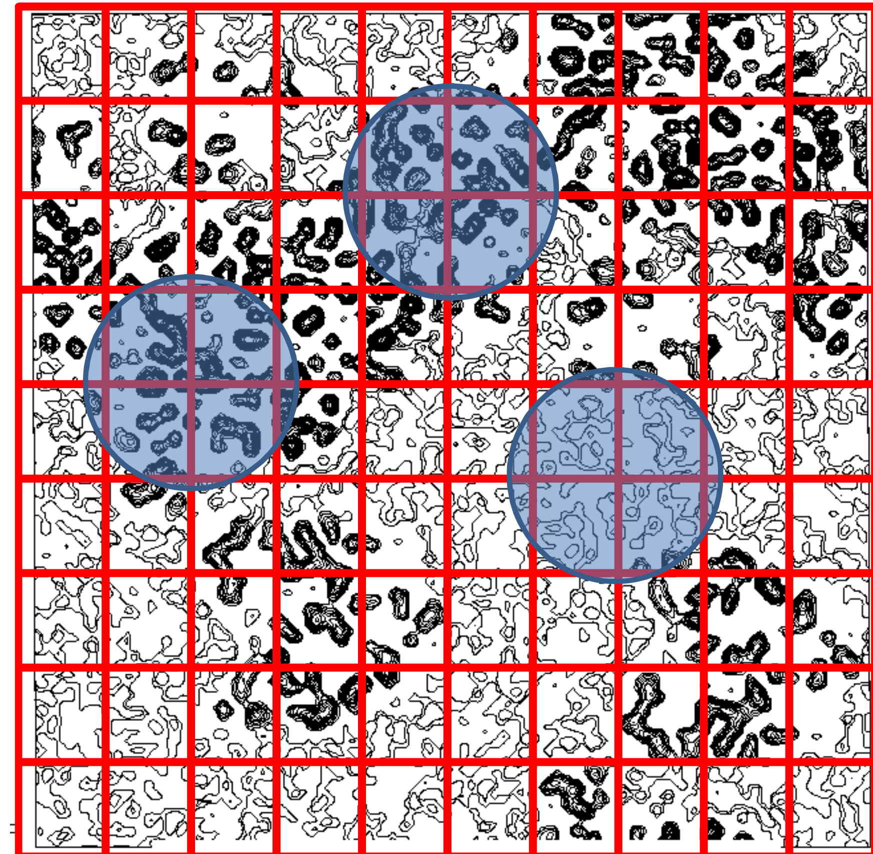
Significant improvement of the electron density (and of the Patterson function) that are crucial for phasing of complex cases.

- Solvent flattening or solvent flipping
- Histogram matching
- Density averaging for NCS
- Phase extension (FreeLunch, beyond resolution limit of the data!)

Solvent flattening or flipping

1) MR solution: phases are calculated.
2) Initial electron density map is obtained from F_{obs} and ϕ_{calc}.

Consider the average flat value of electron density in solvent-occupied volumes.

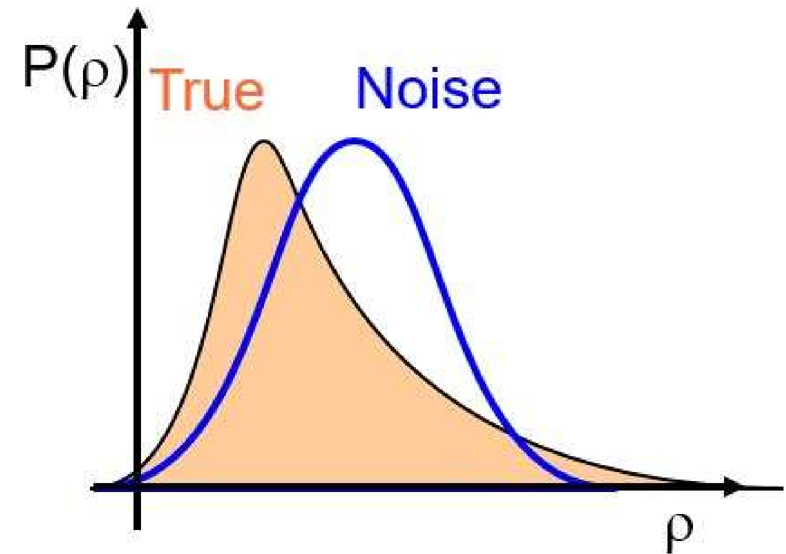


3) Idenfity the solvent occupied volumes: in coarse 3D grid, integrate electron density in a sphere around each grid point
4) Average electron density above a defined threshold is assigned to protein, otherwise solvent
5) In solvent regions, average electron density is introduced in the model (flattening) or electron density value is flipped to negative values (flipping).
6) Calculate correction of phases, ϕ_{corr}
7) Calculate improved electron density map after phase modification.
8) Iterate.

Histogram matching

Probability distribution of electron density values:

- Gaussian distribution for noise
- Non-Gaussian distribution for protein



Histogram matching:

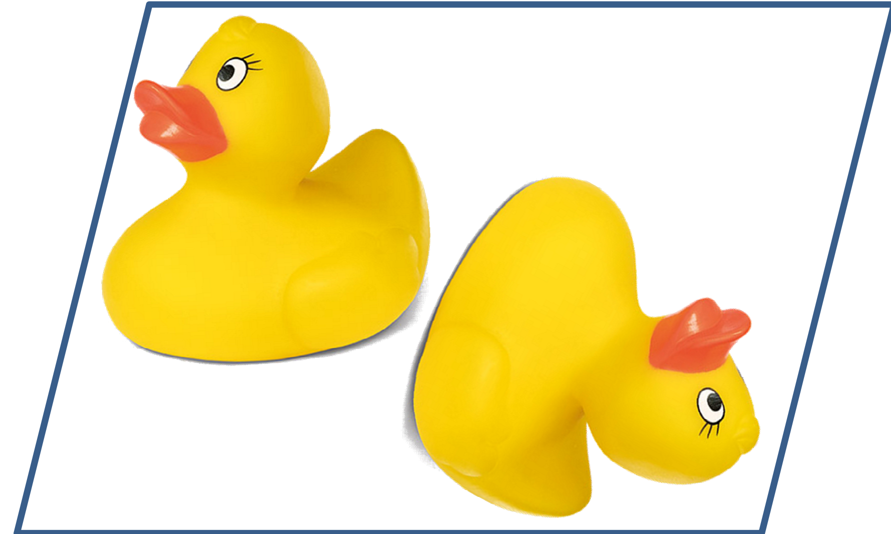
Translation of probability maximum to a value expected for protein,

Sharpening of peaks in electron density map to match electron density distribution expected for protein

...taking into account solvent content and resolution...

NCS transformations

When number of molecules in unit cell determined from Matthews coefficient is larger than number of molecules expected from space group multiplicity, the presence of **non-crystallographic symmetry (NCS)** is likely.



Each NCS operation can be described as a rotation + a translation of the original protein model:

$$\boldsymbol{Protein' = Rot(\varphi\psi\kappa) \cdot Protein + Trans(t_x t_y t_z)}$$

$\varphi\psi\kappa$ = Euler angles

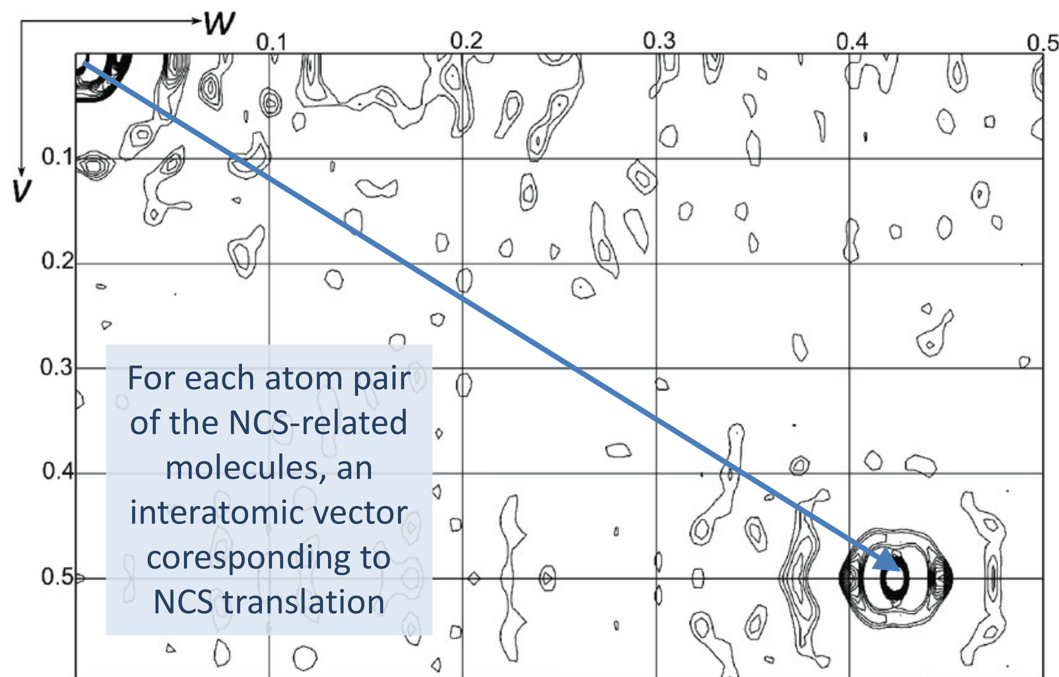
$t_x t_y t_z$ = translation vector

NCS transformations and Patterson maps

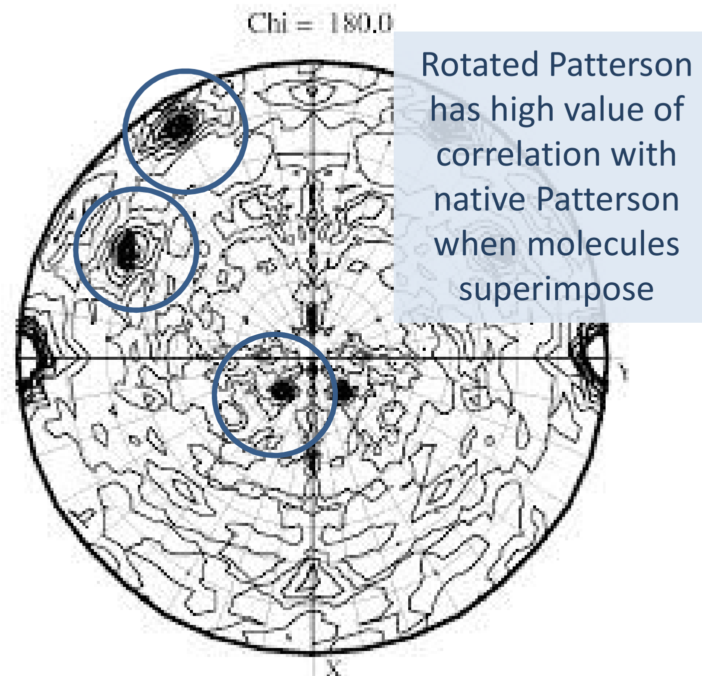
Non-crystallographic symmetry (NCS) elements can be determined from **Patterson map** and **Patterson self-rotation map**.

Patterson map:
autoconvolution of the electron density
$P(\boldsymbol{u}) = \rho(\boldsymbol{x}) \otimes \rho(-\boldsymbol{x})$

Patterson self-rotation map:
Correlation map between Patterson native map and rotated Patterson map
$PSF(\varphi\psi\kappa) = P(\boldsymbol{u}) \otimes [Rot(\varphi\psi\kappa) \cdot P(\boldsymbol{u})]$



Translational symmetry elements:
peaks in the native Patterson map



Rotational symmetry elements:
peaks in the Patterson self-rotation map

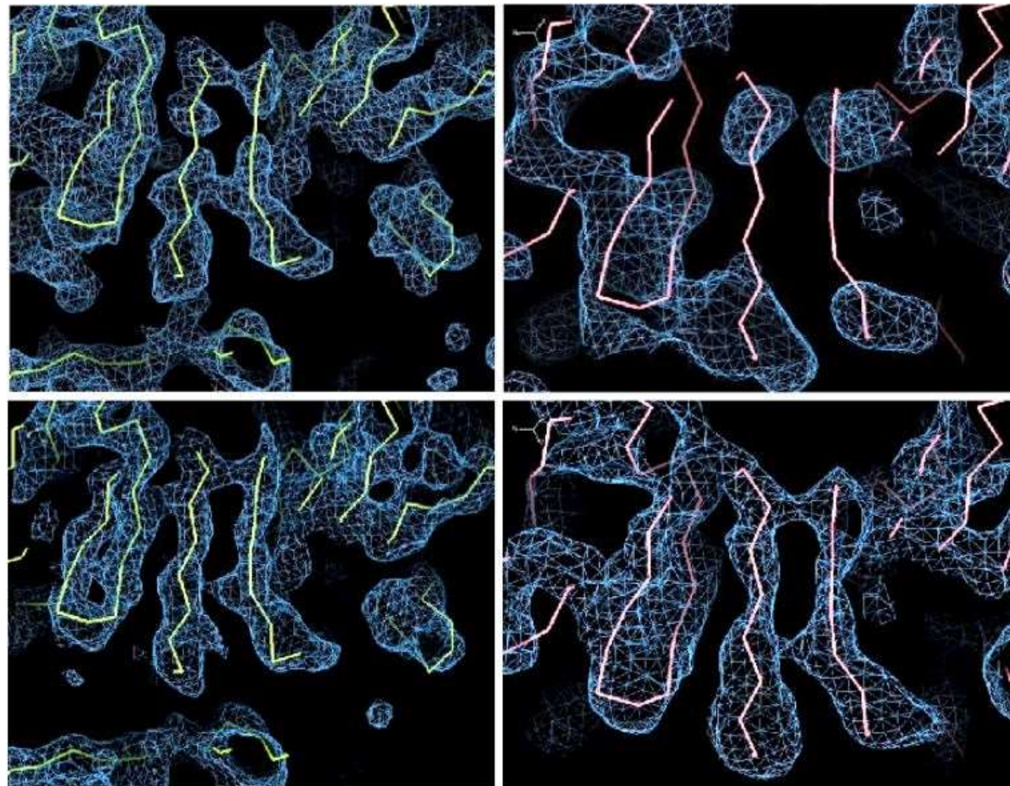
Density averaging for NCS

Molecules related by non-crystallographic symmetry elements should have the same electron density.

Therefore, electron density can be averaged between all NCS-related copies of the molecule.

1) Define regions of map to be averaged
2) Average electron density
3) Recalculate corrected phases
4) Calculate improved map

Density averaging can be used also for cross-crystal averaging.



BEFORE

AFTER