Phenix User Workshop, Pittsburgh Diffraction Conference, October 14th 2023

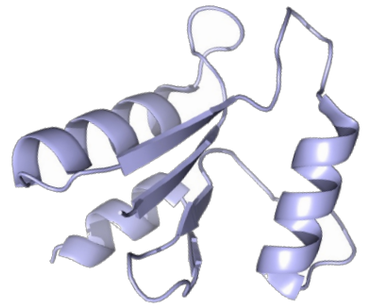
Molecular Replacement

Dorothee Liebschner

Lawrence Berkeley Laboratory

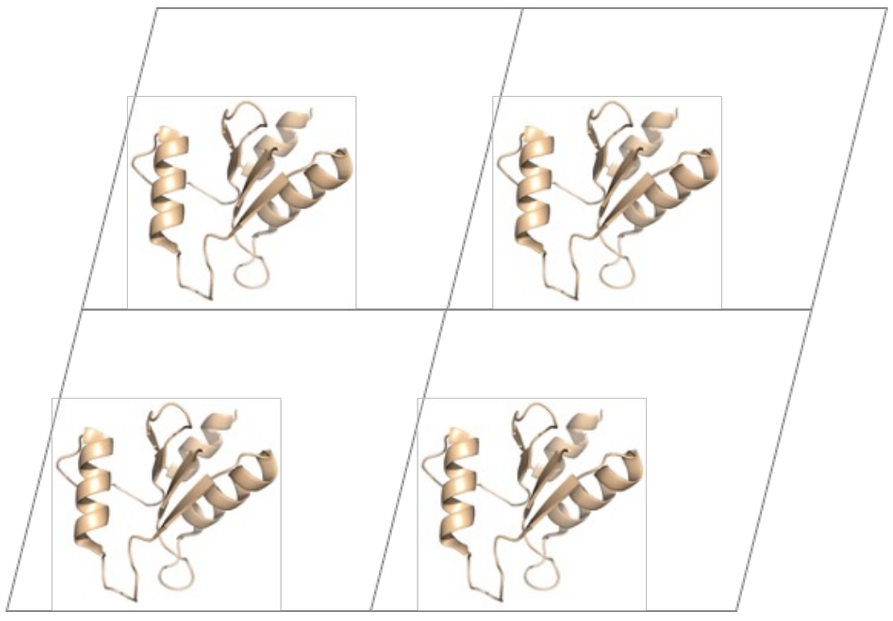
Molecular Replacement (MR)

MR = Use a known molecular model to solve the unknown crystal structure of a related molecule.



Known model

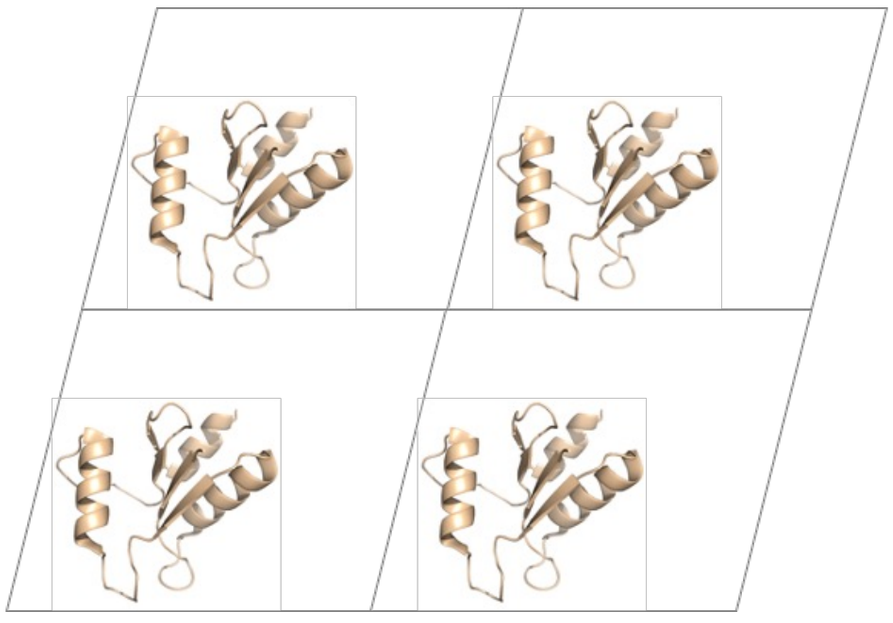
Crystal of unknown structure

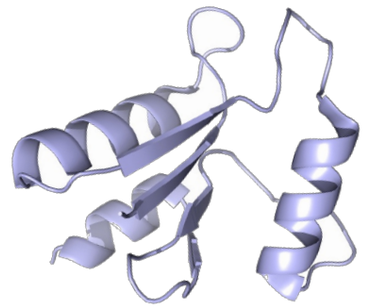


Molecular Replacement (MR)

MR = Use a known molecular model to solve the unknown crystal structure of a related molecule.

Crystal of unknown structure



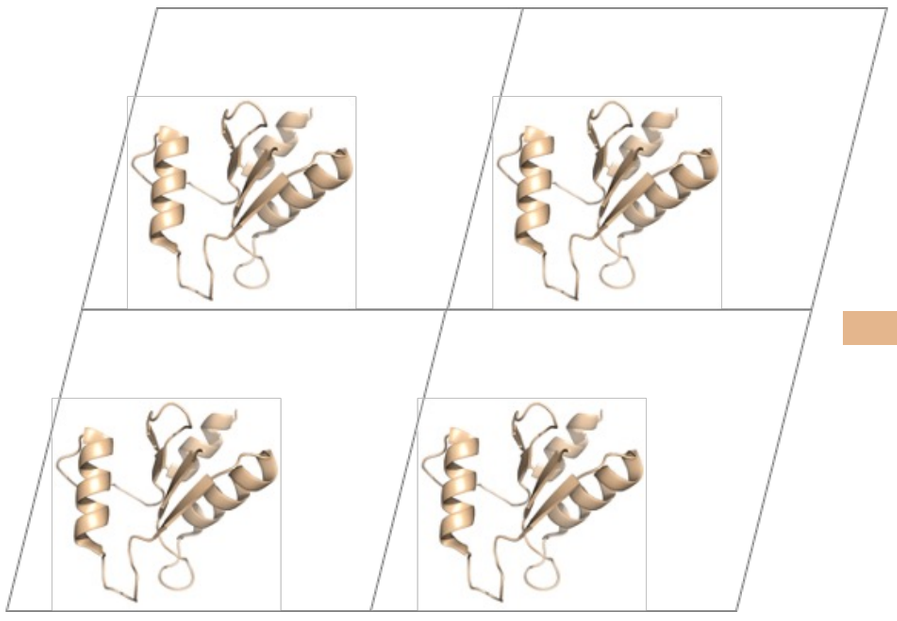


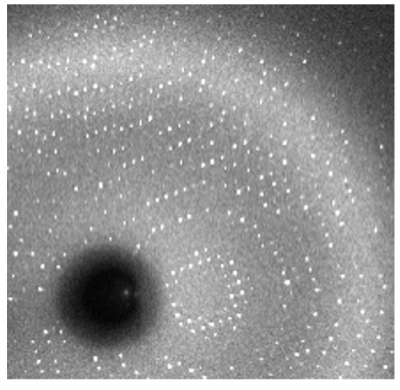
Known model

Known model provides initial estimates of the phases of the unknown structure.

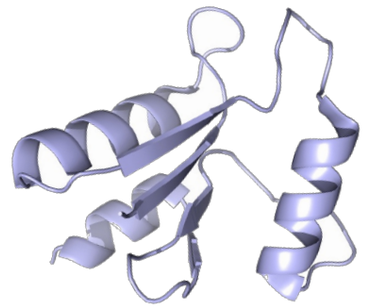
Molecular Replacement (MR)

Crystal of unknown structure



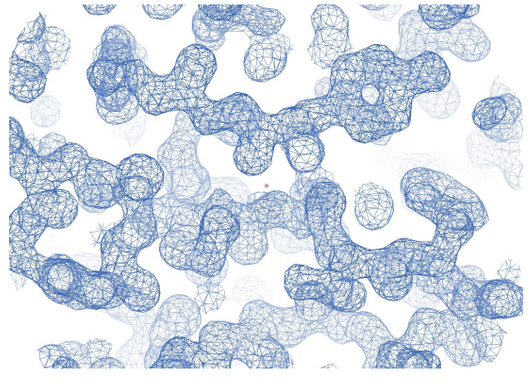


Intensities (hkl)



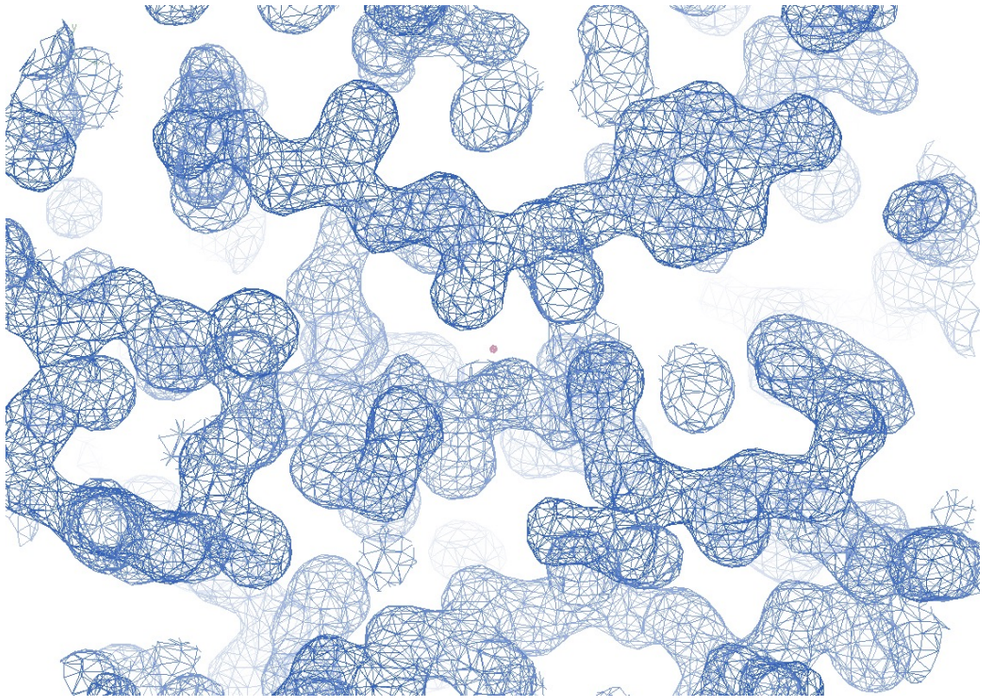
Known model

$|F|\ e^{i\phi}$



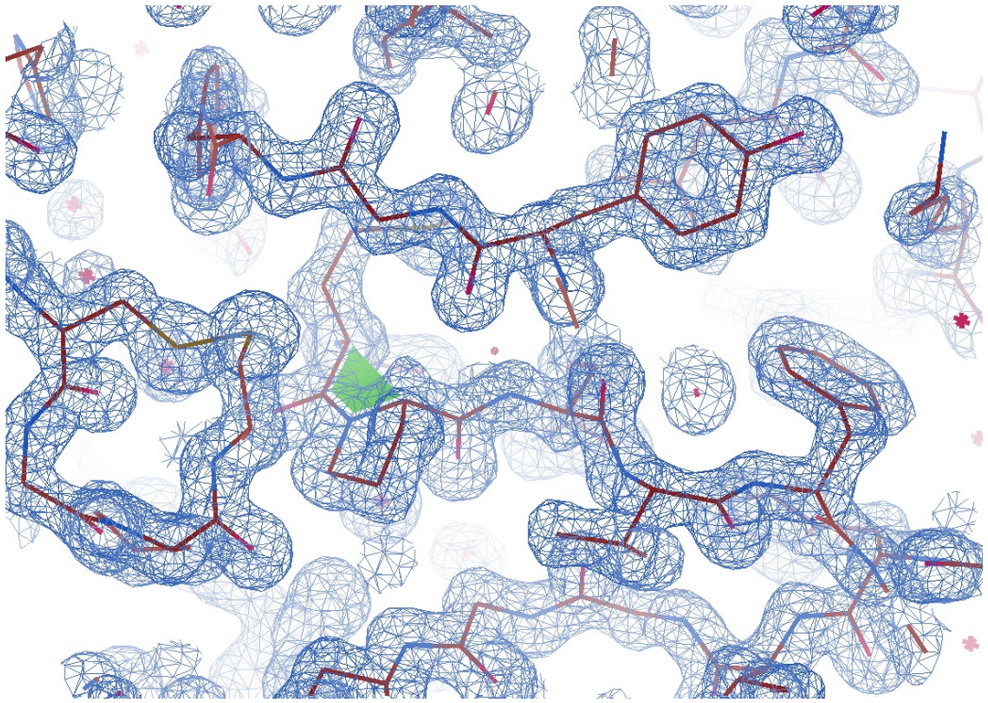
Density map

Molecular Replacement (MR)



If we know the density…

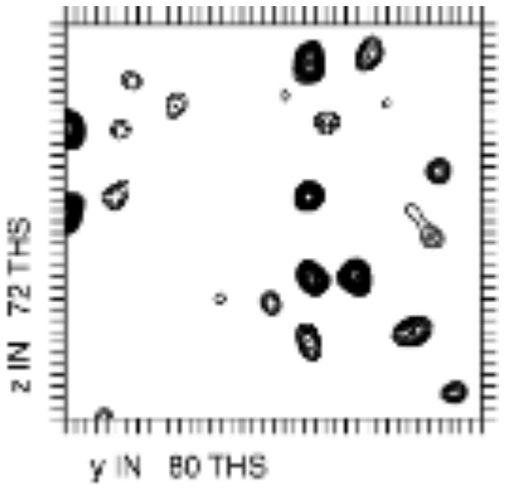
Molecular Replacement (MR)



If we know the density…

… then we can
determine the structure

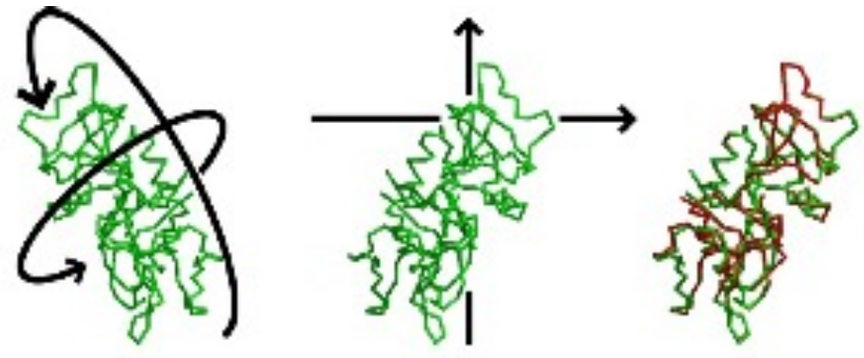
How to recover phases



Experimentally

Exploit the properties of a few special atoms:

- Anomalous scattering
- A large number of electrons



Computationally

- *Molecular Replacement (MR)*
  A previously known structure provides initial phase estimates for a new structure

- *Direct Methods*
  Phase relationships can be formulated by assuming the positivity and atomicity of the electron density

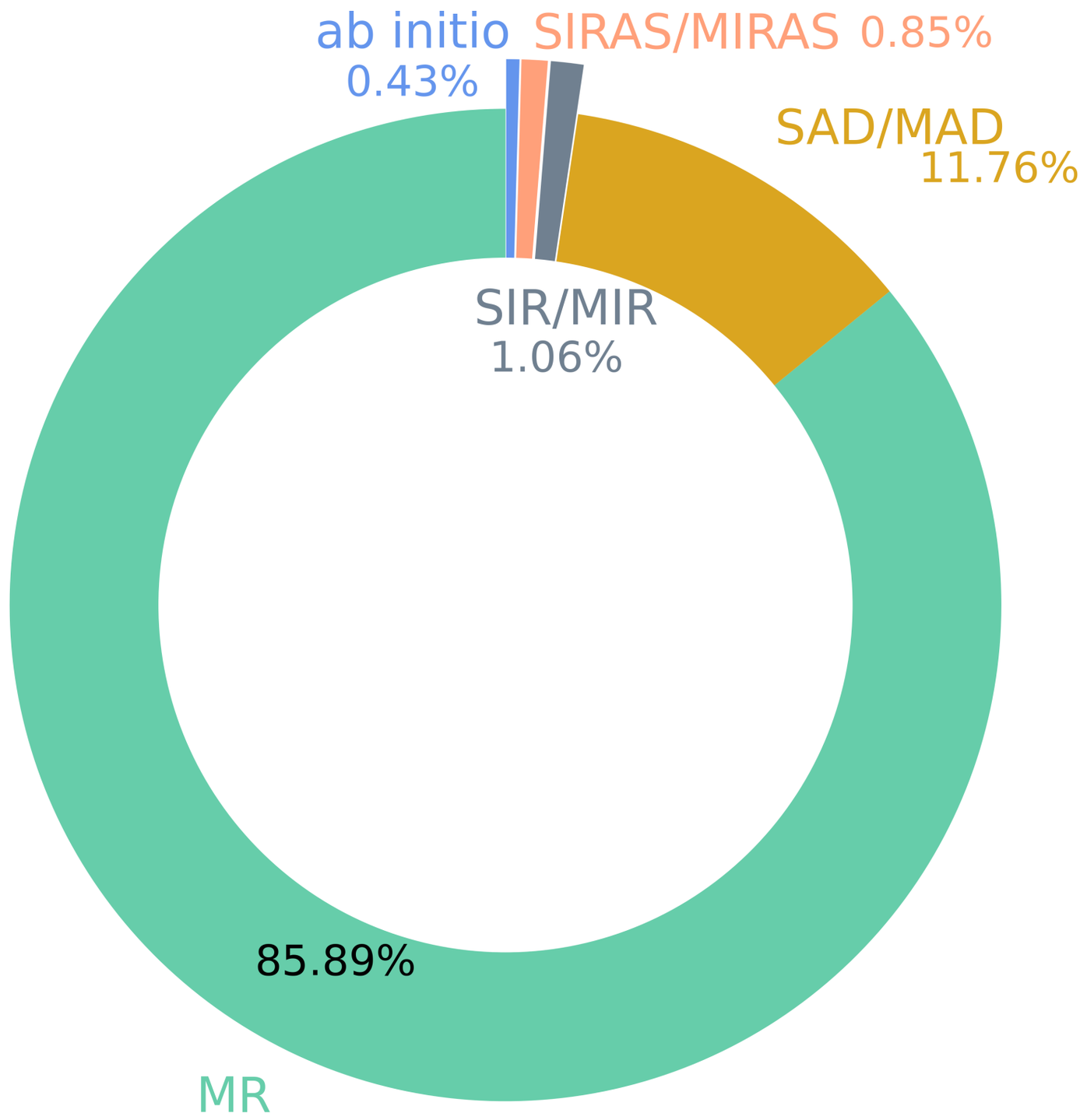
Phasing methods in the PDB

MR method:

- Fast
- Cheap
- Highly automated

Known structure:

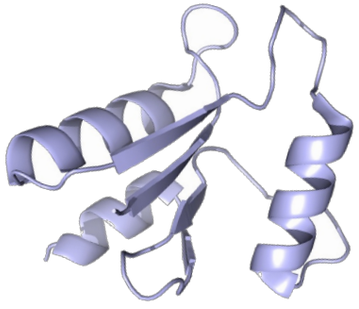
- Number of known structures increases (PDB)
- Predicted models



Note: Not all models in the PDB have (correct) info.

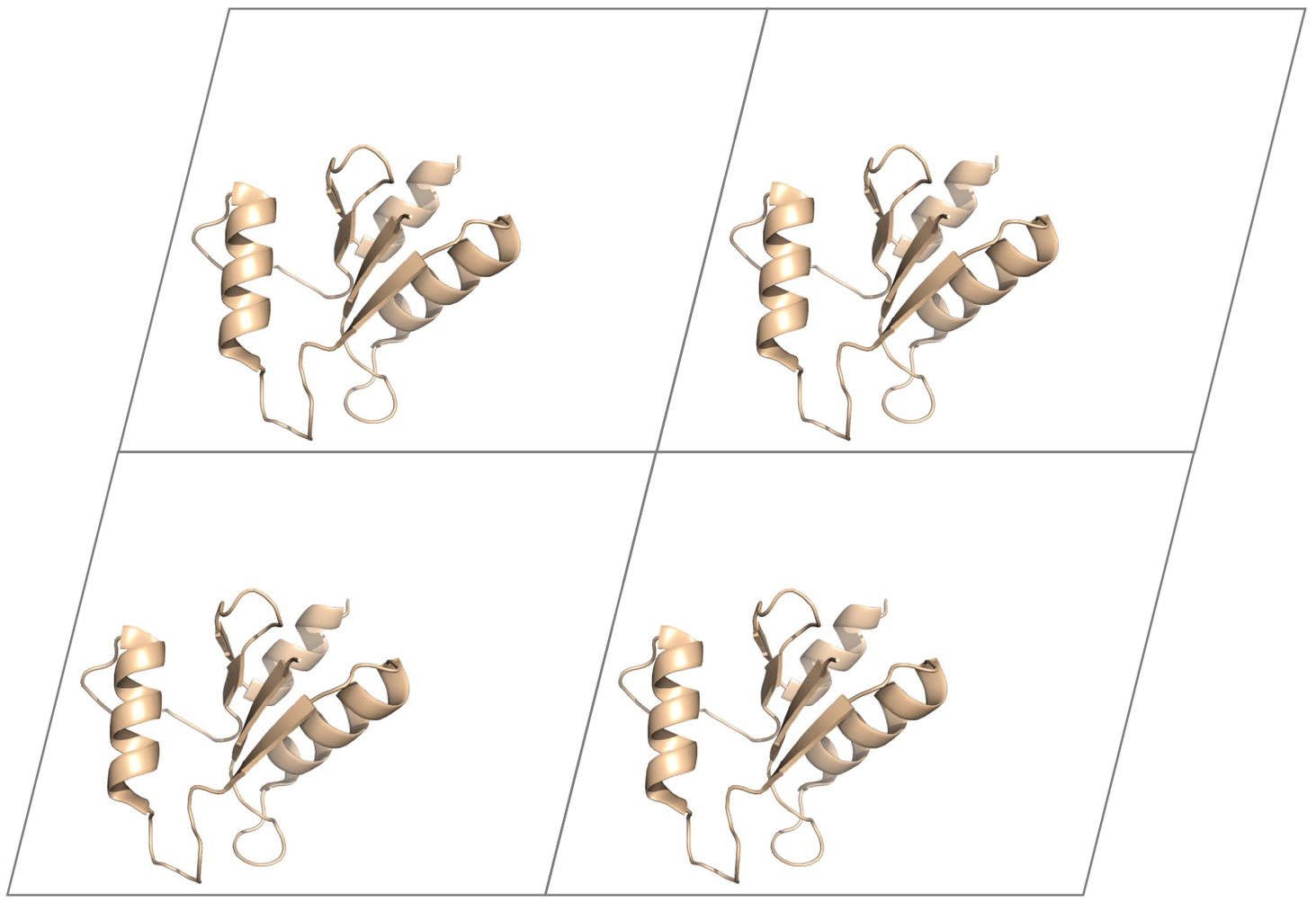
Molecular replacement: Approach

Try to match the known model with the unknown structure.



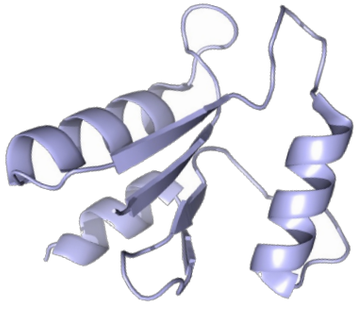
Search Model

Crystal of unknown structure

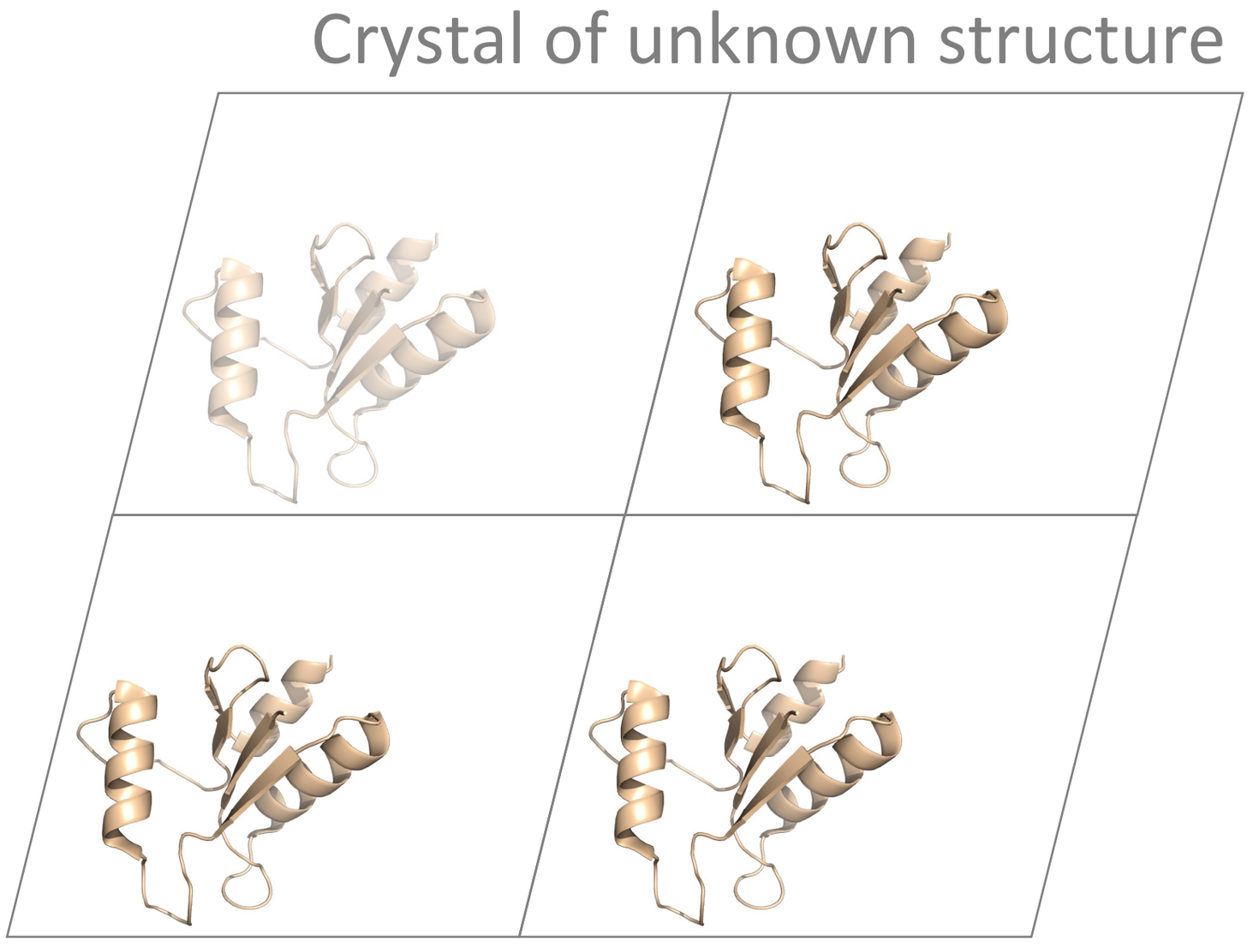


Molecular replacement: Approach

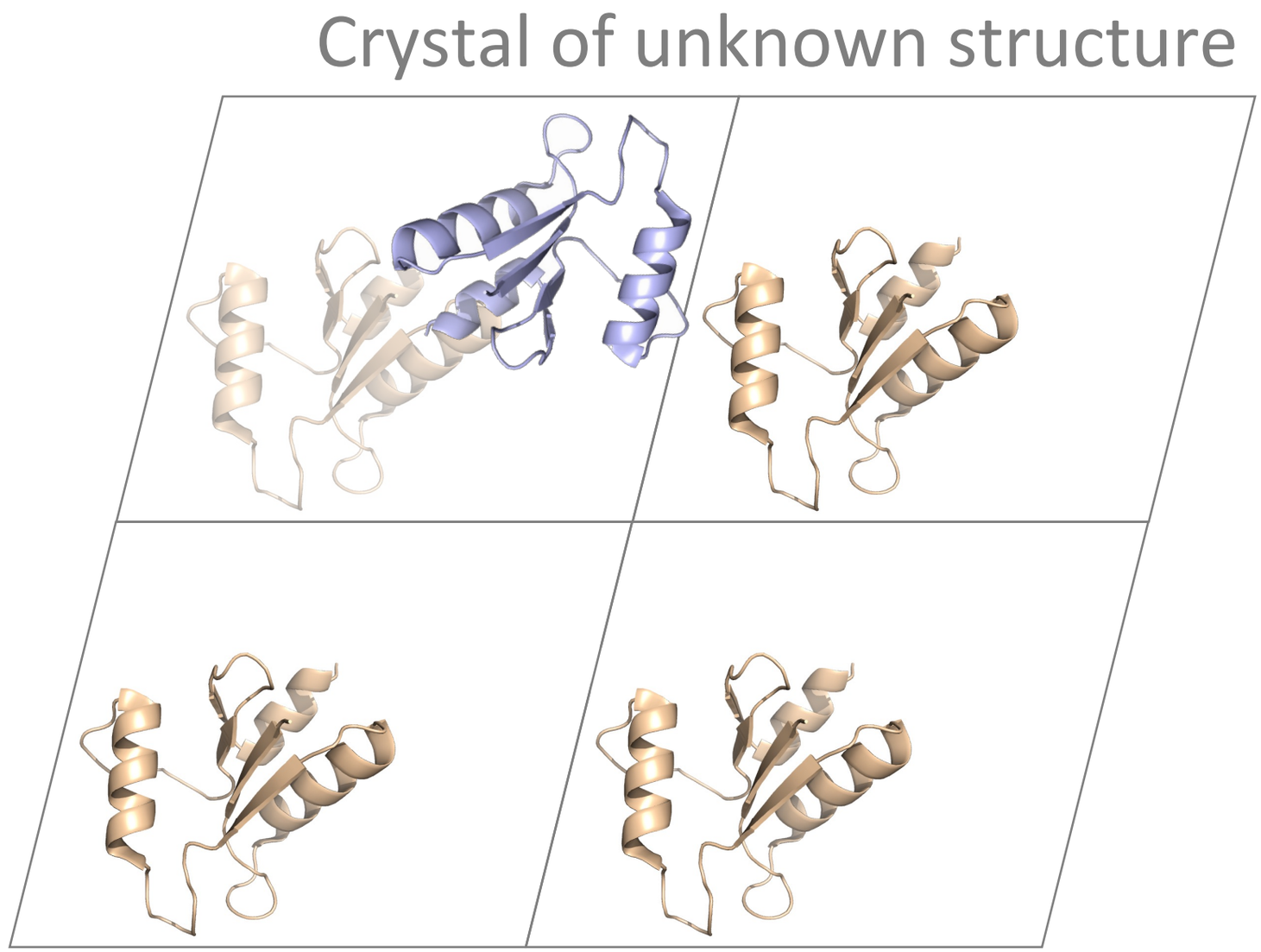
- Try all possible positions and orientations of the model
- Find where the predicted diffraction best matches the observed diffraction.



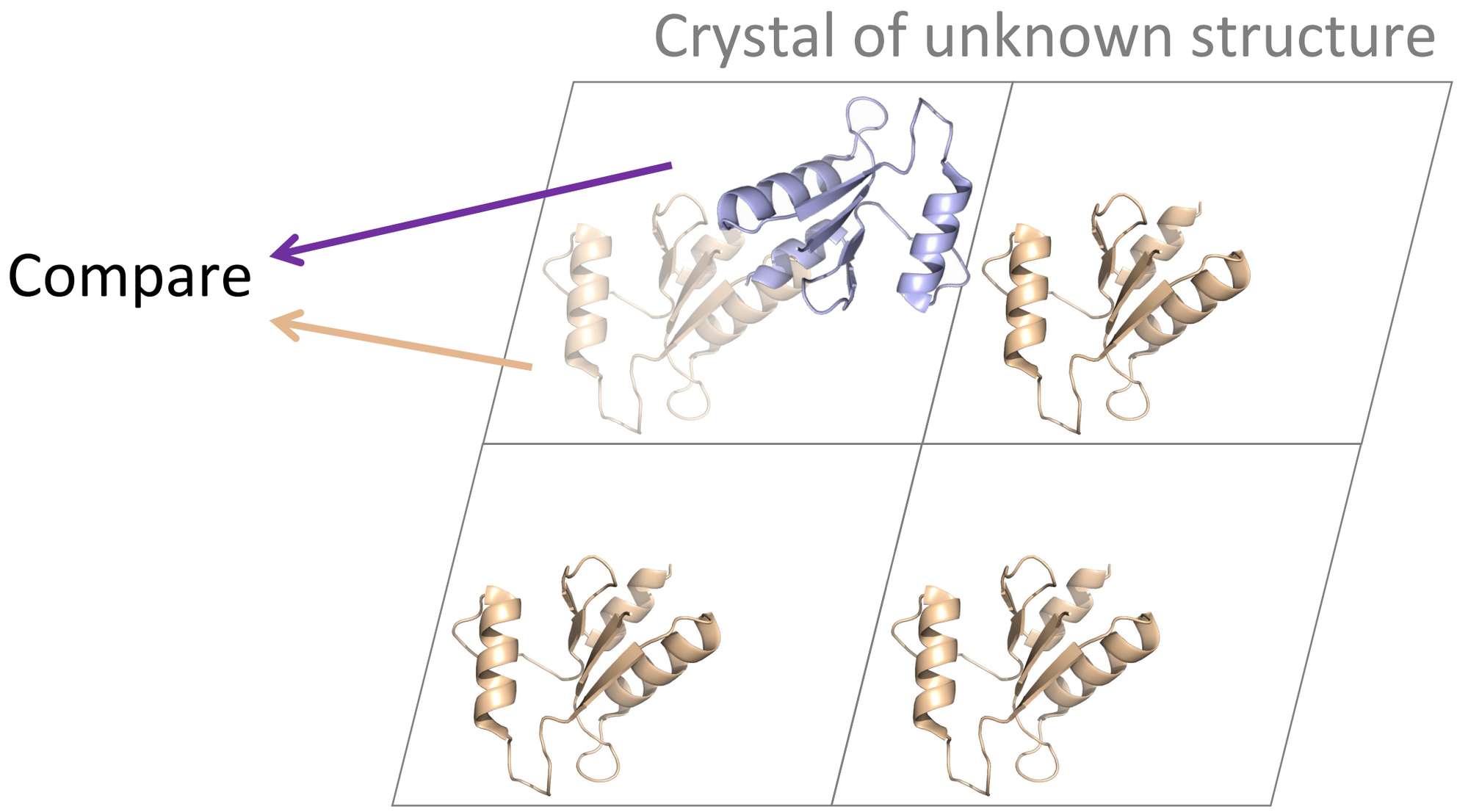
Search Model



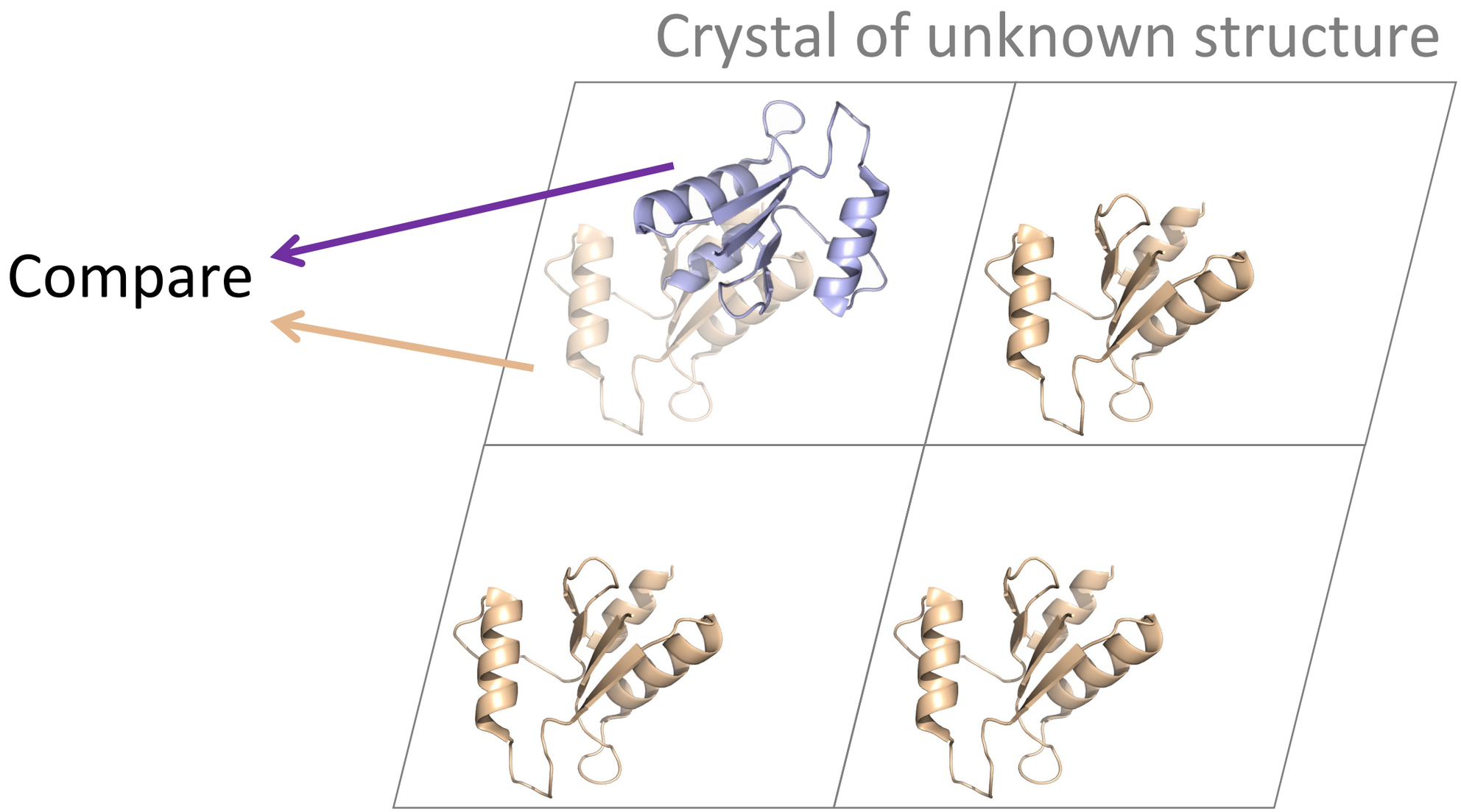
Molecular replacement: Approach



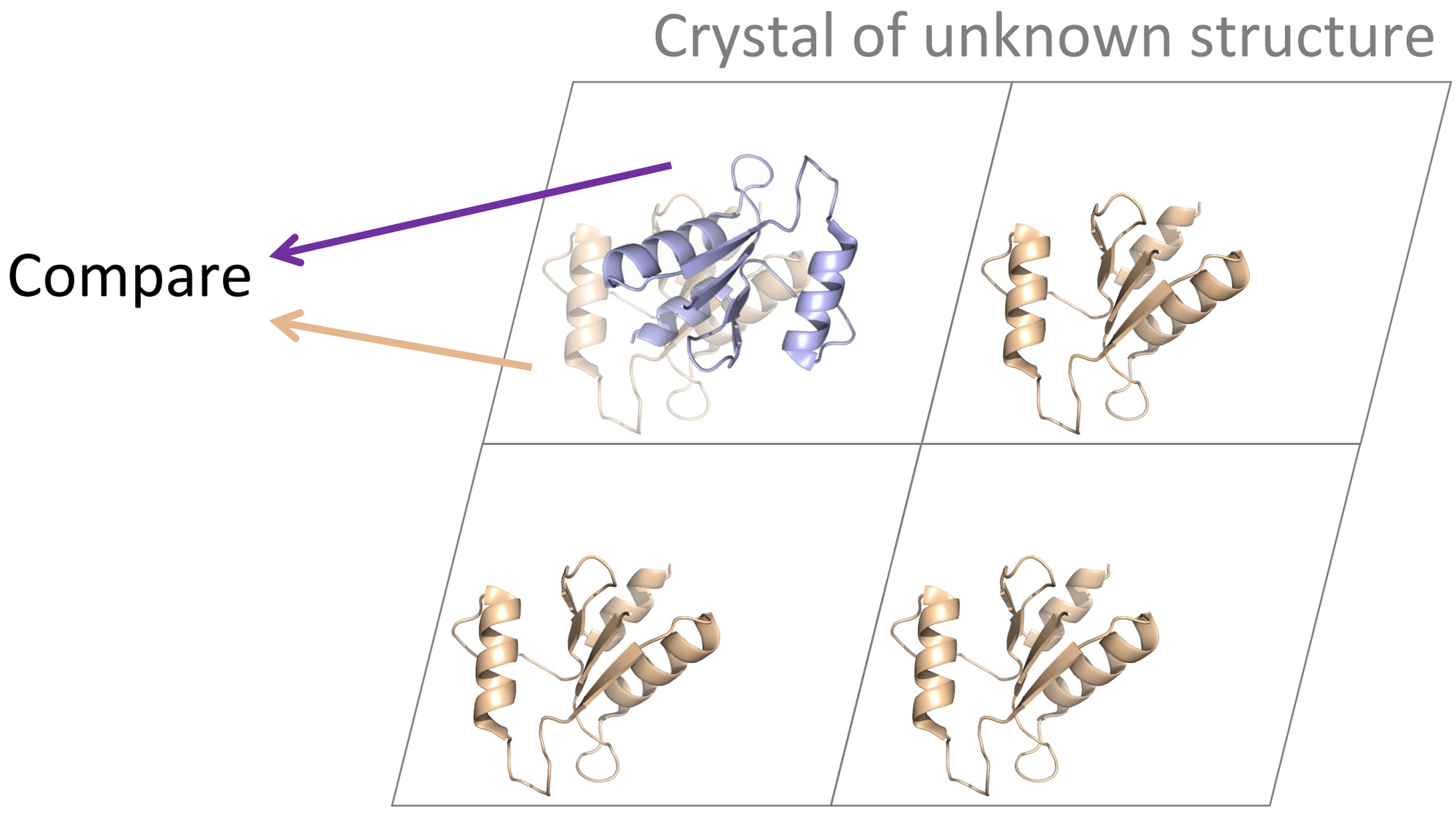
Molecular replacement: Approach



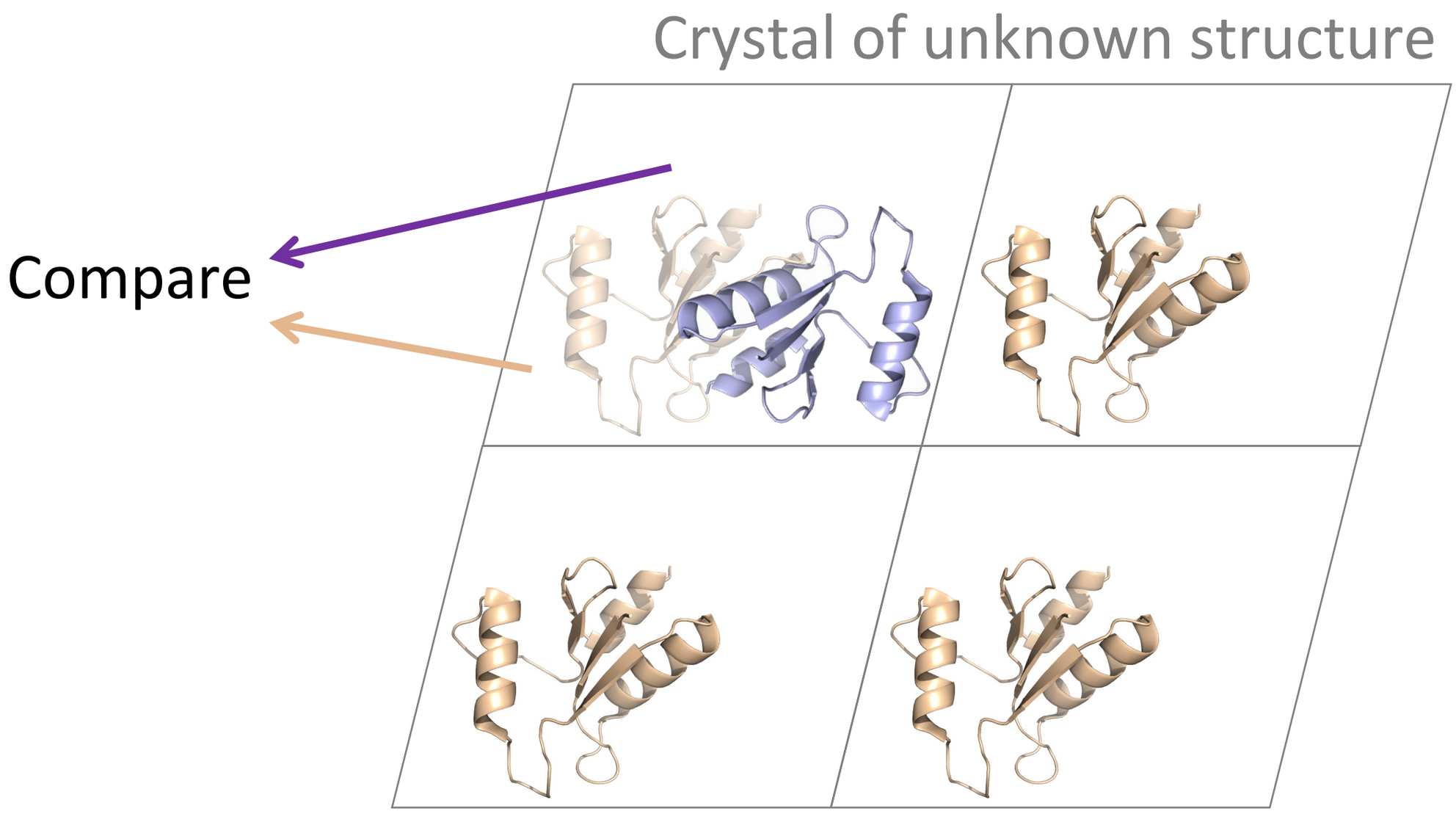
Molecular replacement: Approach



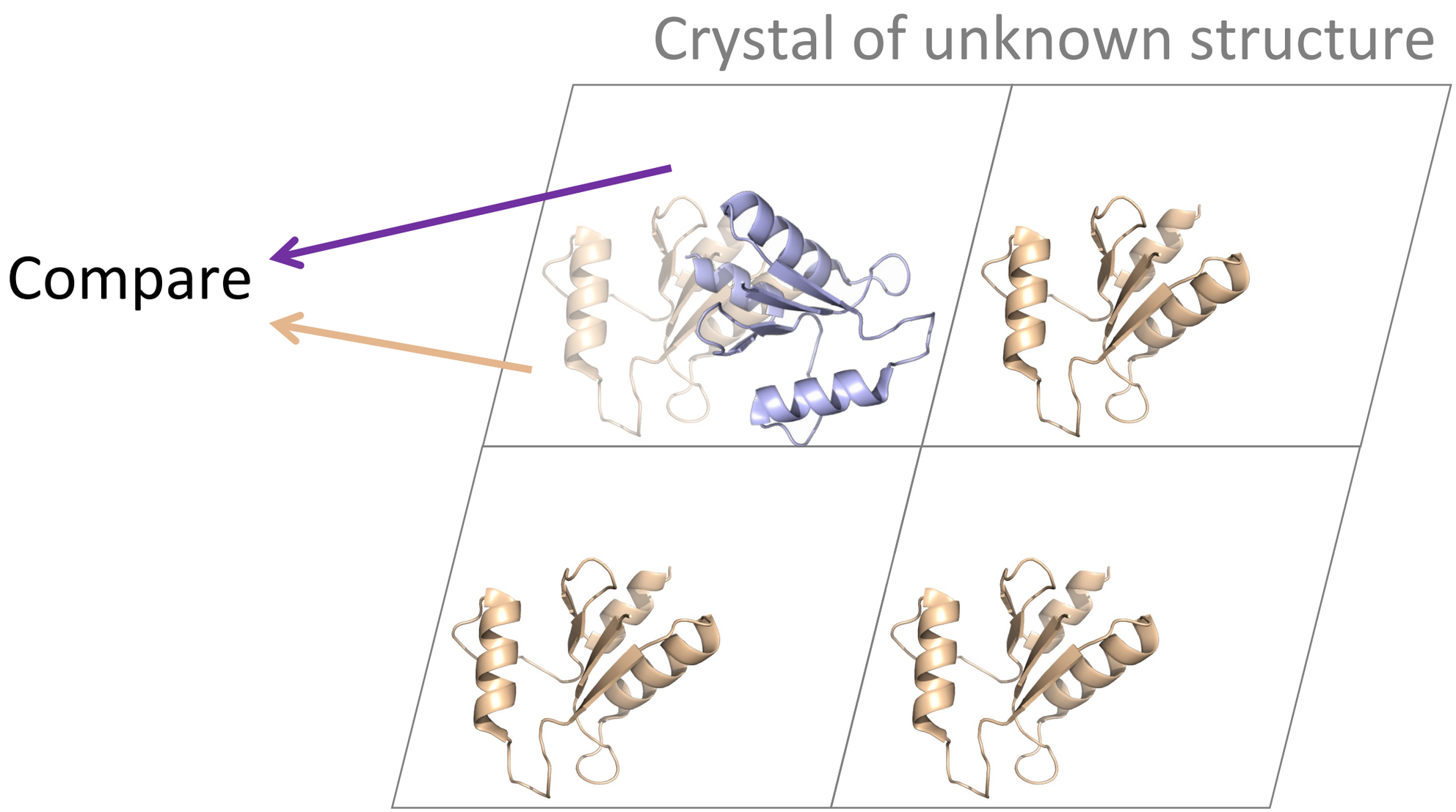
Molecular replacement: Approach



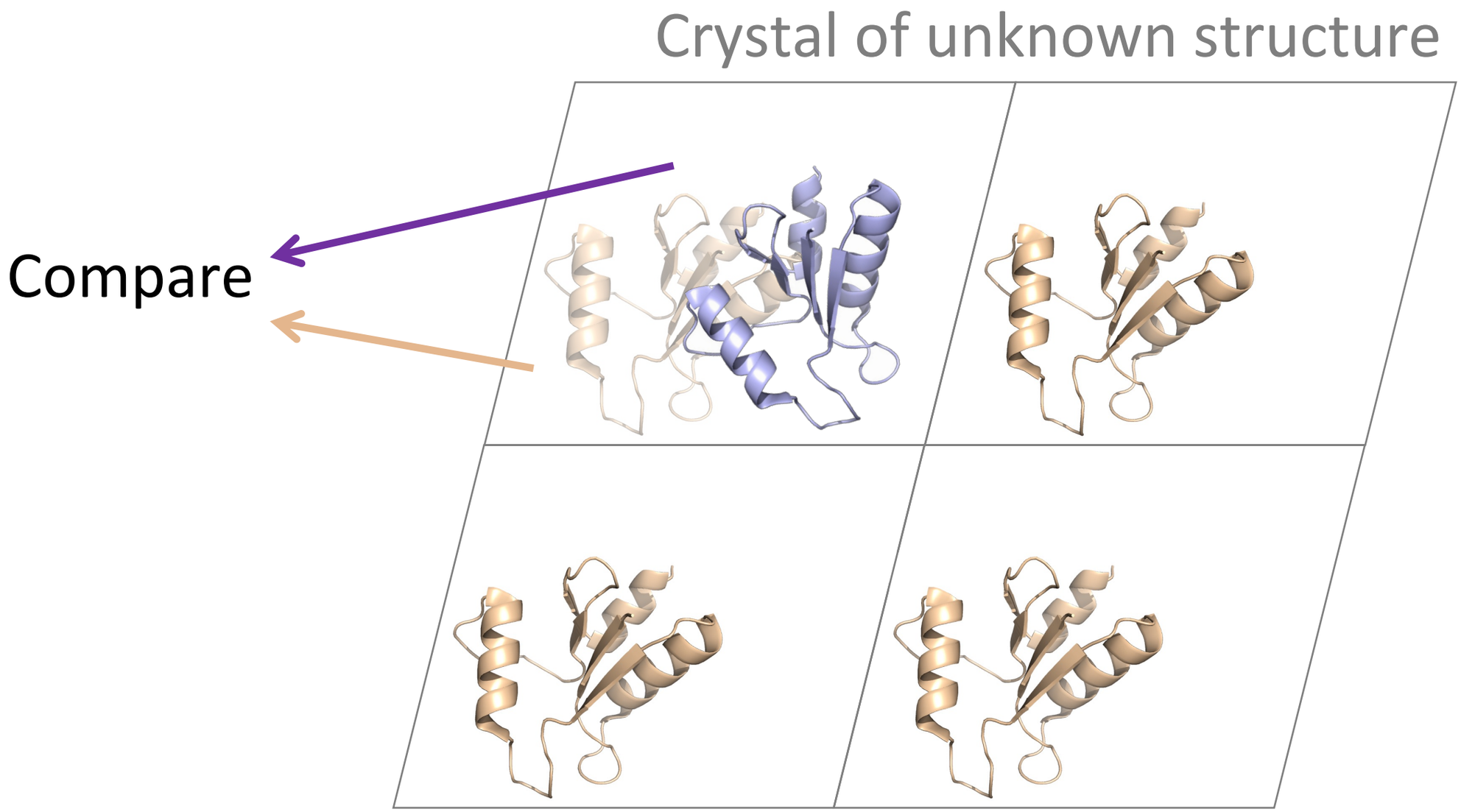
Molecular replacement: Approach



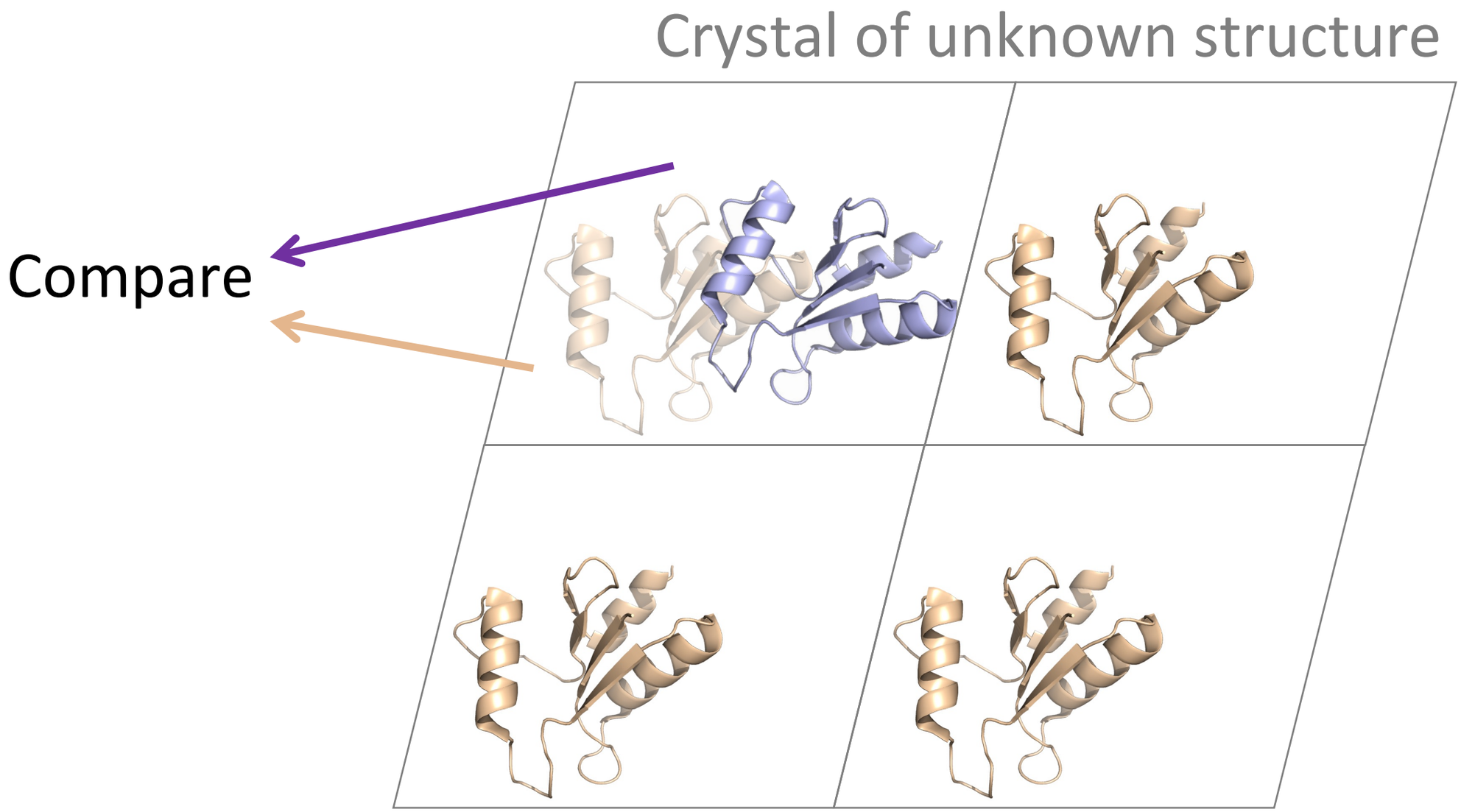
Molecular replacement: Approach



Molecular replacement: Approach



Molecular replacement: Approach



Challenges of MR

1) How to choose a search model (how to modify it)

2) Speed of the calculations

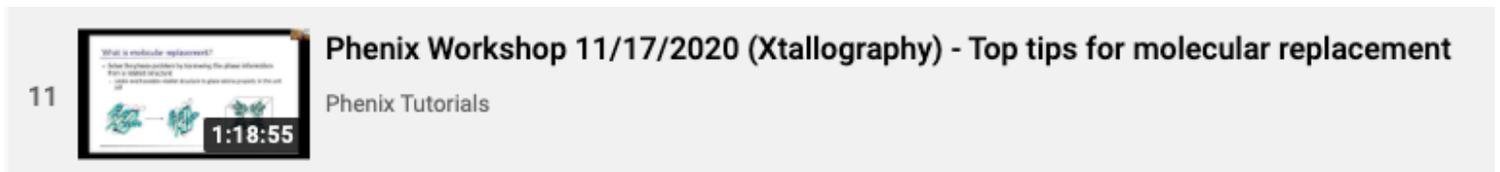
3) How to score each orientation and translation (how to differentiate signal from noise?)

4) How to do the rotations/translations

Evans, P.; McCoy, A. An Introduction to Molecular Replacement. *Acta Cryst. D* **2008**, *64* (1), 1–10

Dodson, E. Introduction to Molecular Replacement: A Time Perspective. *Acta Crystallogr D Struct Biol* **2021**, *77* (7), 867–879

McCoy, A. J.; Sammito, M. D.; Read, R. J. Implications of AlphaFold2 for Crystallographic Phasing by Molecular Replacement. *Acta Cryst D* **2022**, *78* (1), 1–13



1) The search model

- Cristical step in MR.

- Should provide a high proportion of the scattering from the target structure with high accuracy (low r.m.s.d.)

  - Homologue structures

    Low r.m.s.d. → high sequence identity
    (sequence comparison search)
    Prune regions of large sequence diversity
    Truncate side-chains

- Predicted structures
  (Remove low pLDDT regions, split into domains)

2) Computational tricks to improve the speed

An exhaustive search is very slow even on modern computers.

Each molecule needs six parameters to define its orientation and position.

Example:

3 angles (0–360°, 0–180°, 0–360°) at 2.5° intervals → 1.5×10^6 grid points
3 translations in a 100 × 100 × 100 Å cell at 1 Å intervals → 10^6 grid points

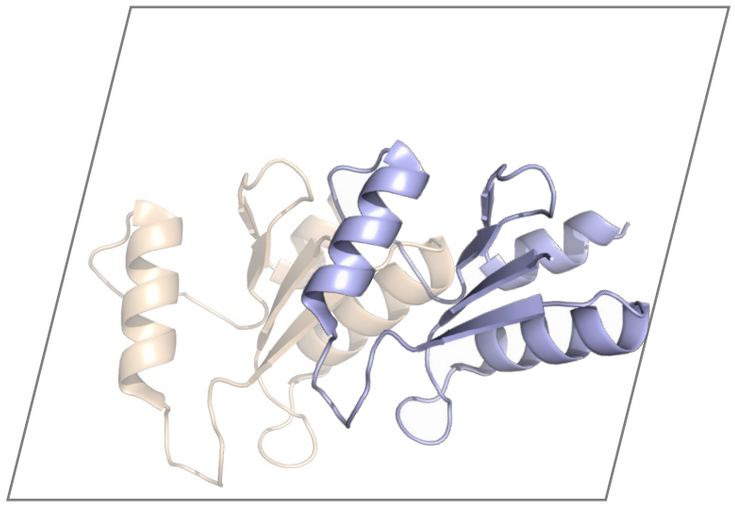
Total search of 1.5×10^{12} points

→ Separate the two searches: Do ration first, then the translation

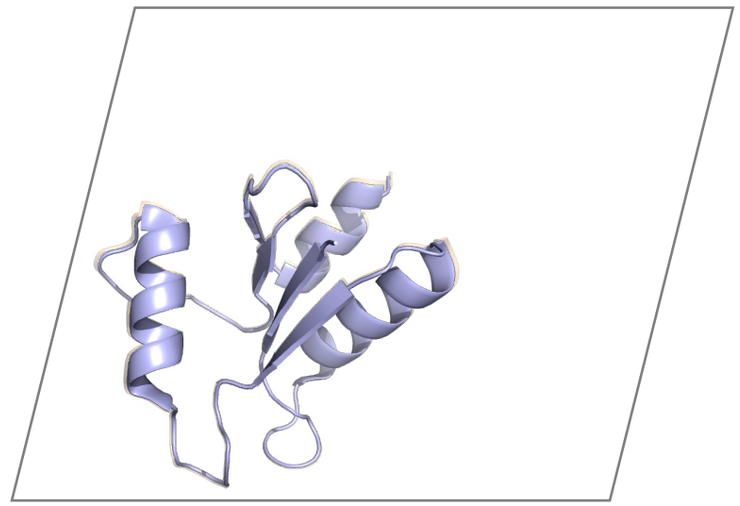
2.5×10^6 points per rotation solution

3) The scoring function

Compare observed and calculated diffraction.



Poor score



Good score

Different approaches:

- Patterson function (vector map)
- Maximum-likelihood Methods (“for any postulated orientation and position of the model, what is the probability of obtaining the structure amplitudes that we observe?”)

3) The scoring function: ML Method

Explicitly models errors (experimental σF and r.m.s. coordinate error of the model)

→ Likelihood methods are more robust and generally give clearer solutions in difficult cases

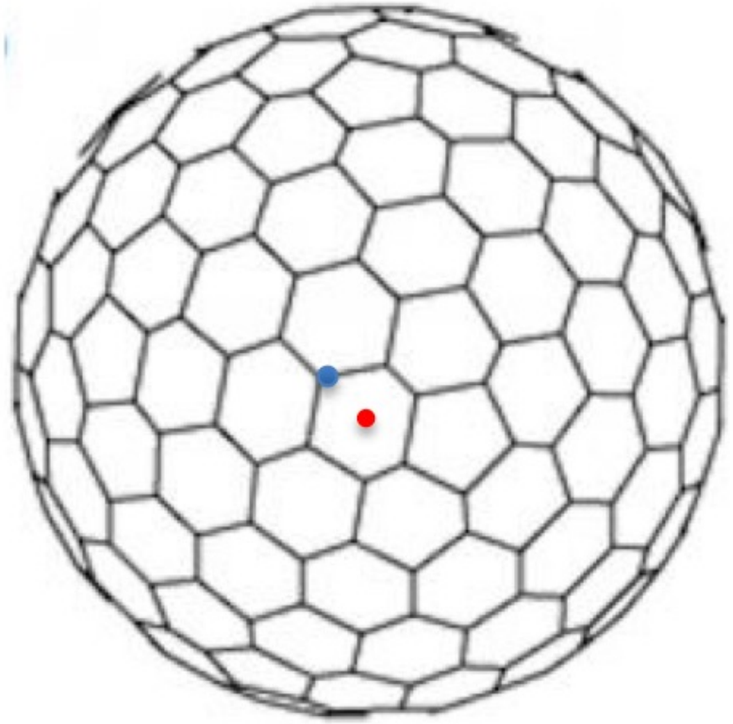
3) The scoring function: ML Method

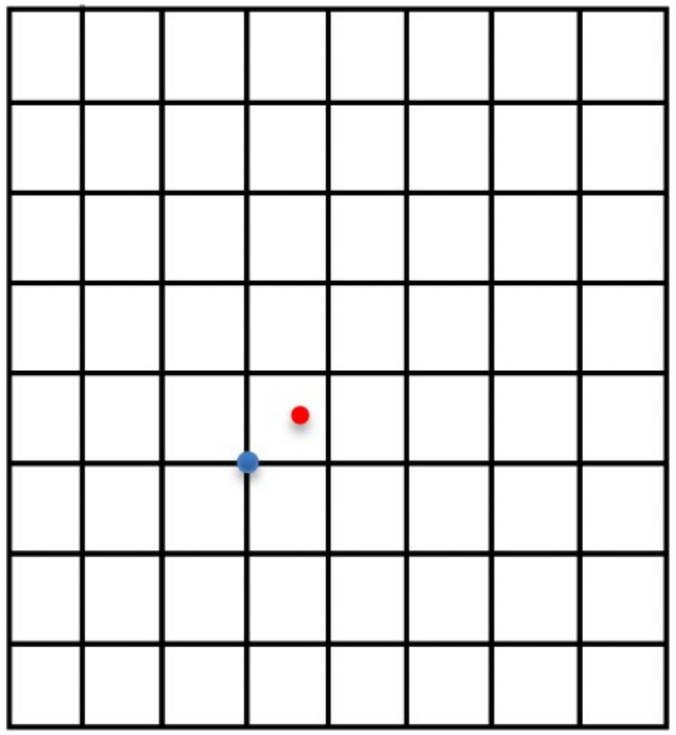
- Place model at orientations and calculate probability of each being correct
- Place model at points and calculate probability of each being correct
- The scoring function is the LLG
- Do packing analysis to see if there are clashes

3) The scoring function: ML Method

Optimize orientation and position away from grid search Locations.

The scoring function is the LLG





3) The scoring function: ML Method

LLG = Log Likelihood gain

Difference between the likelihood of the model and the likelihood calculated from a Wilson distribution.

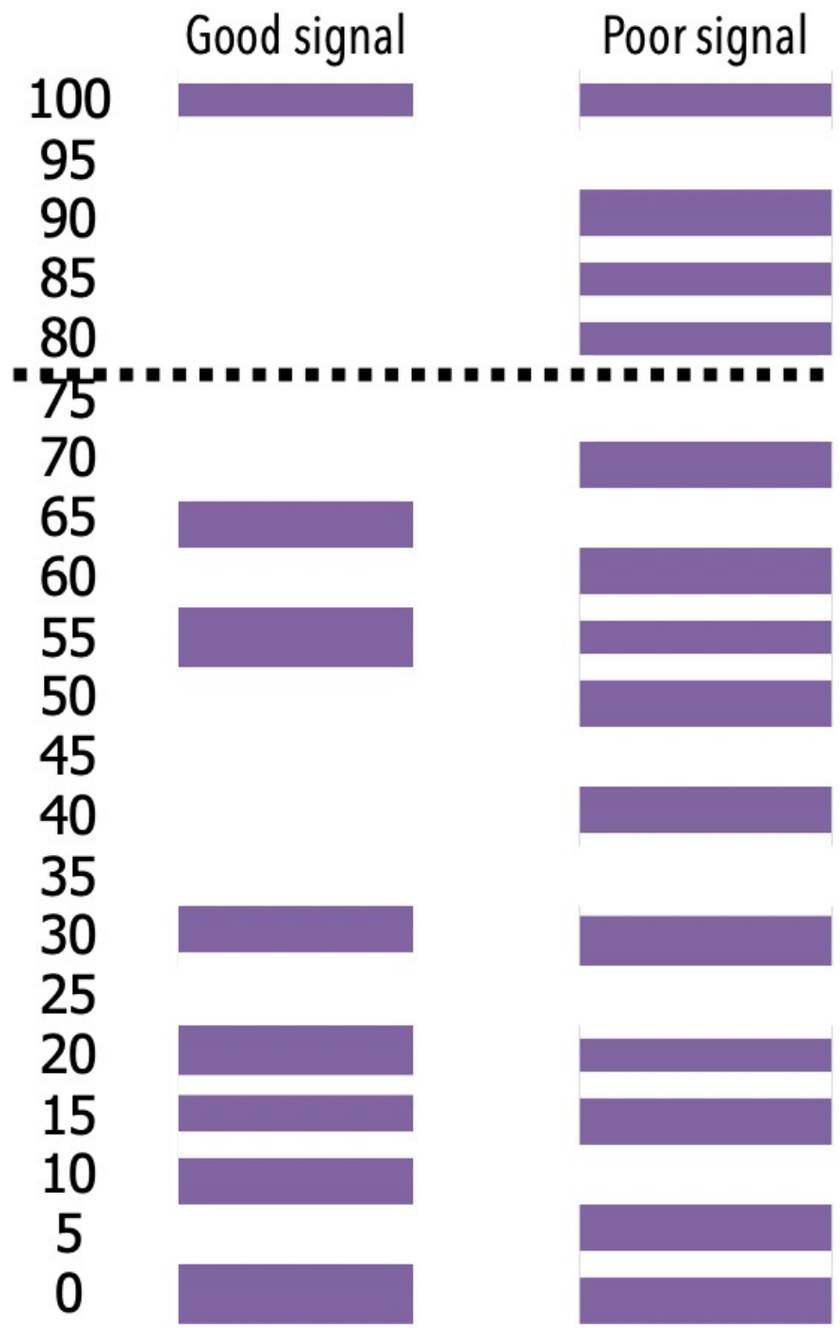
→ it measures how much better the data can be predicted with the search model than with a random distribution of the same atoms.

TF-Z = how many standard deviations your solution is above the mean (the higher the better).

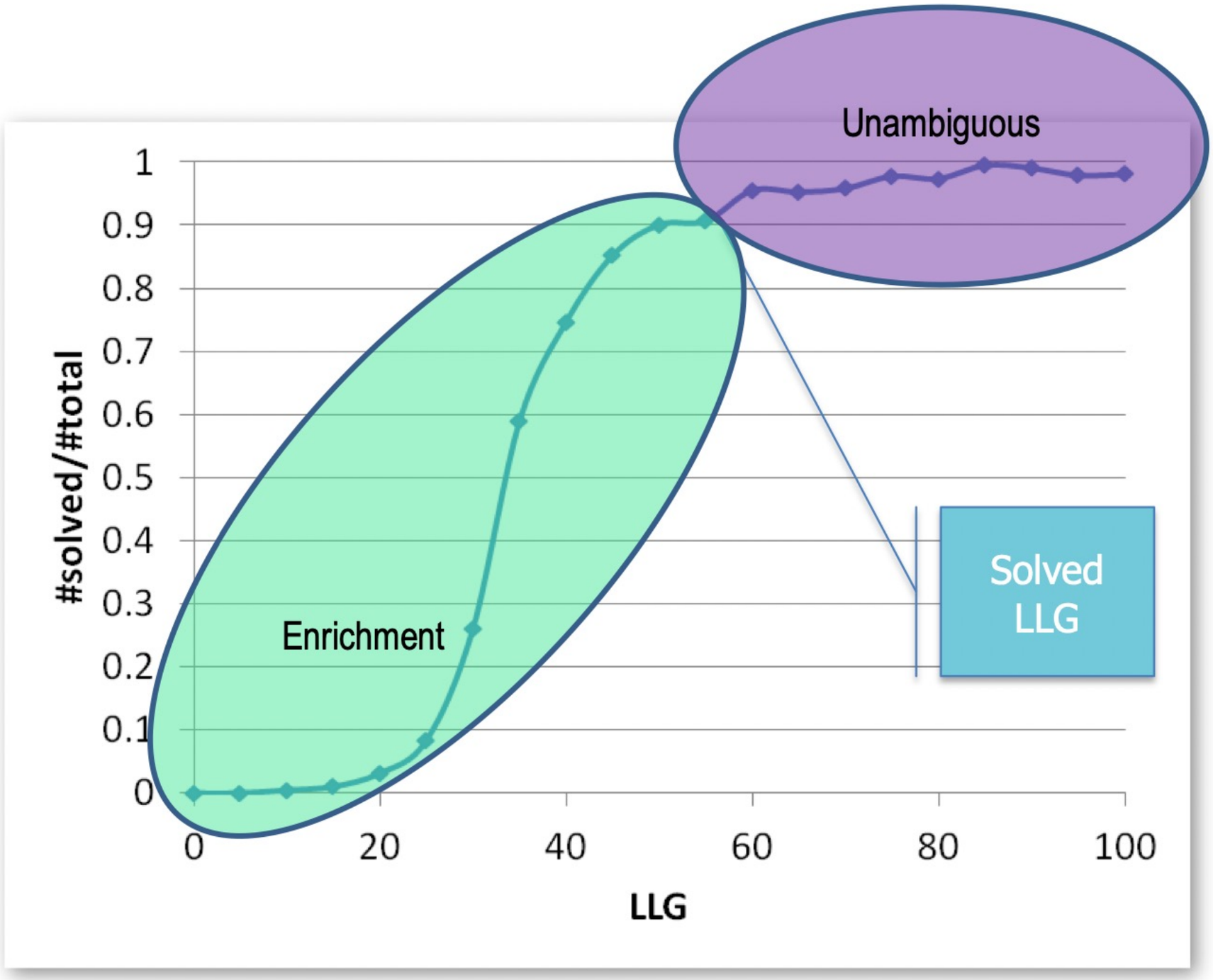
3) The scoring function: ML Method

Solutions over 75% of the difference between the top peak and the mean are selected

- Good signal, few potential solutions
- Poor signal, many potential solutions



3) The scoring function: ML Method



Plot of LLG versus success in structure solution

Database of over 23000 MR problems

R.D. Oeffner

3) The scoring function: ML Method

TF Z-score	LLG score	Solved?
< 5	< 25	no
5 - 6	25 - 36	unlikely
6 - 7	36 - 49	possibly
7 - 8	49 - 64	probably
> 8	> 64	definitely

4) Search strategies

Need to describe rotations and translations which move the coordinates into a new frame

- Translations

- Rotations (Rotation matrix is inconvenient)

- Different angle conventions
  - Polar angles
  - Eulerian angles
  - Lattman angles

Summary

- Choose and prepare your search model carefully (even predicted models!)
- Know how many molecules/domains you need to place
- MR is successful if LLG > 64, TF-Z > 8