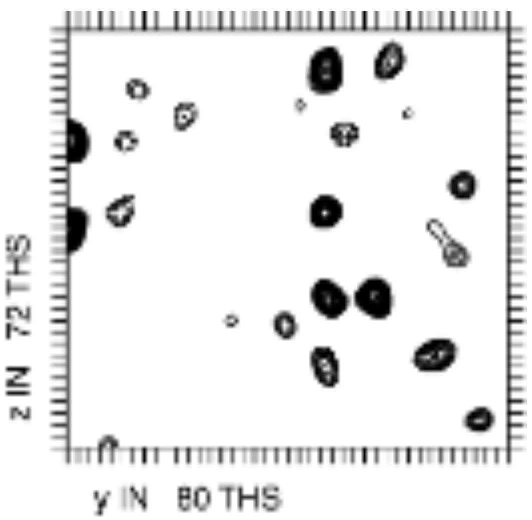
Phenix Software for X-ray Crystallography & cryo-EM Workshop
March 18th 2024

Molecular Replacement

Dorothee Liebschner
Lawrence Berkeley Laboratory

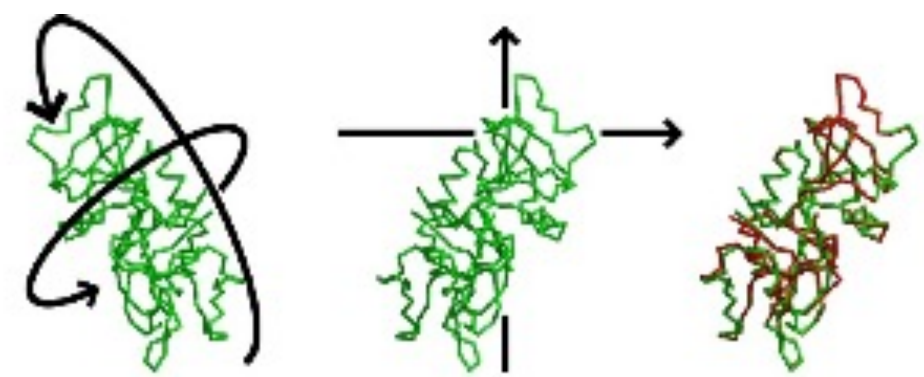
How to recover phases



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

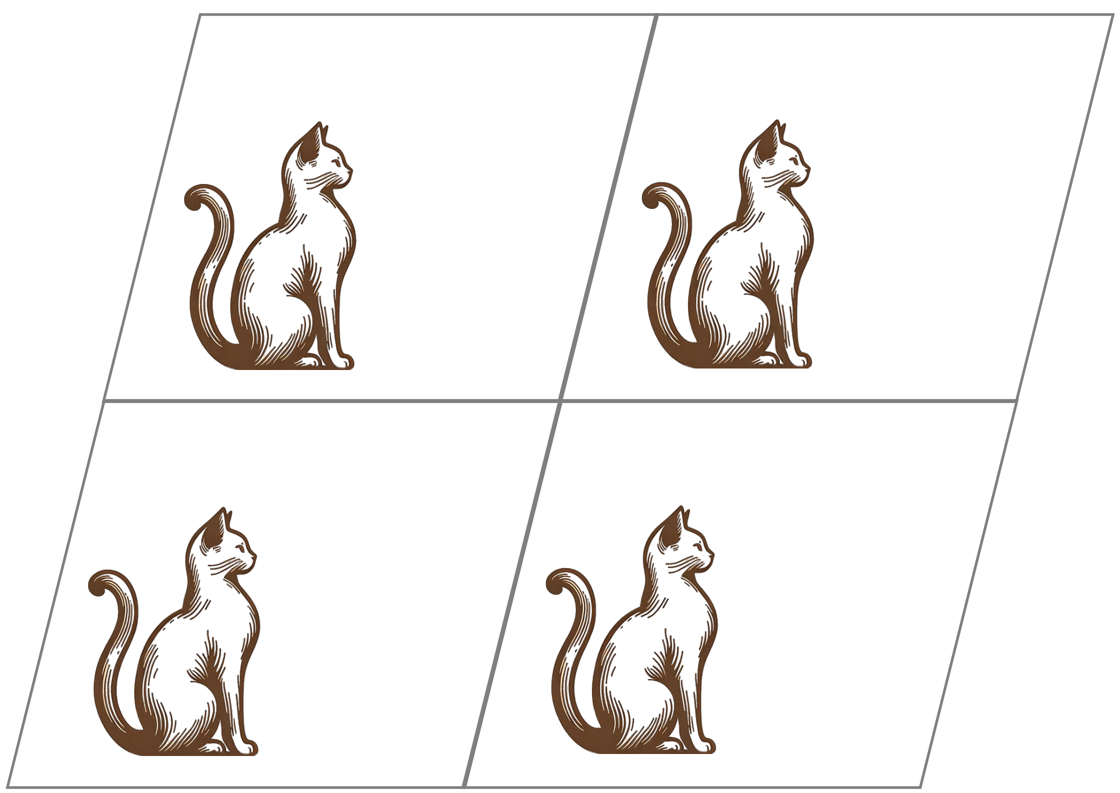
- *Direct Methods*
  Phase relationships can be formulated by assuming the positivity and atomicity of the electron density

Molecular Replacement (MR)

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

Known model

Crystal of unknown structure

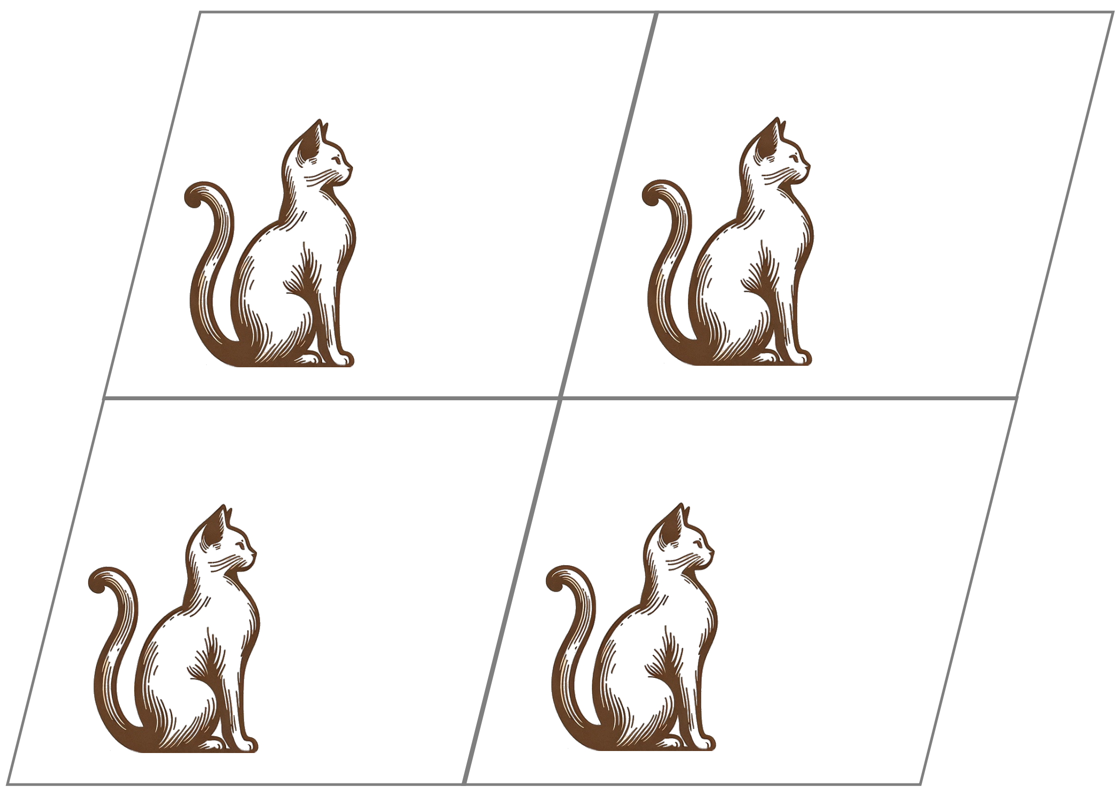


Molecular Replacement (MR)

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

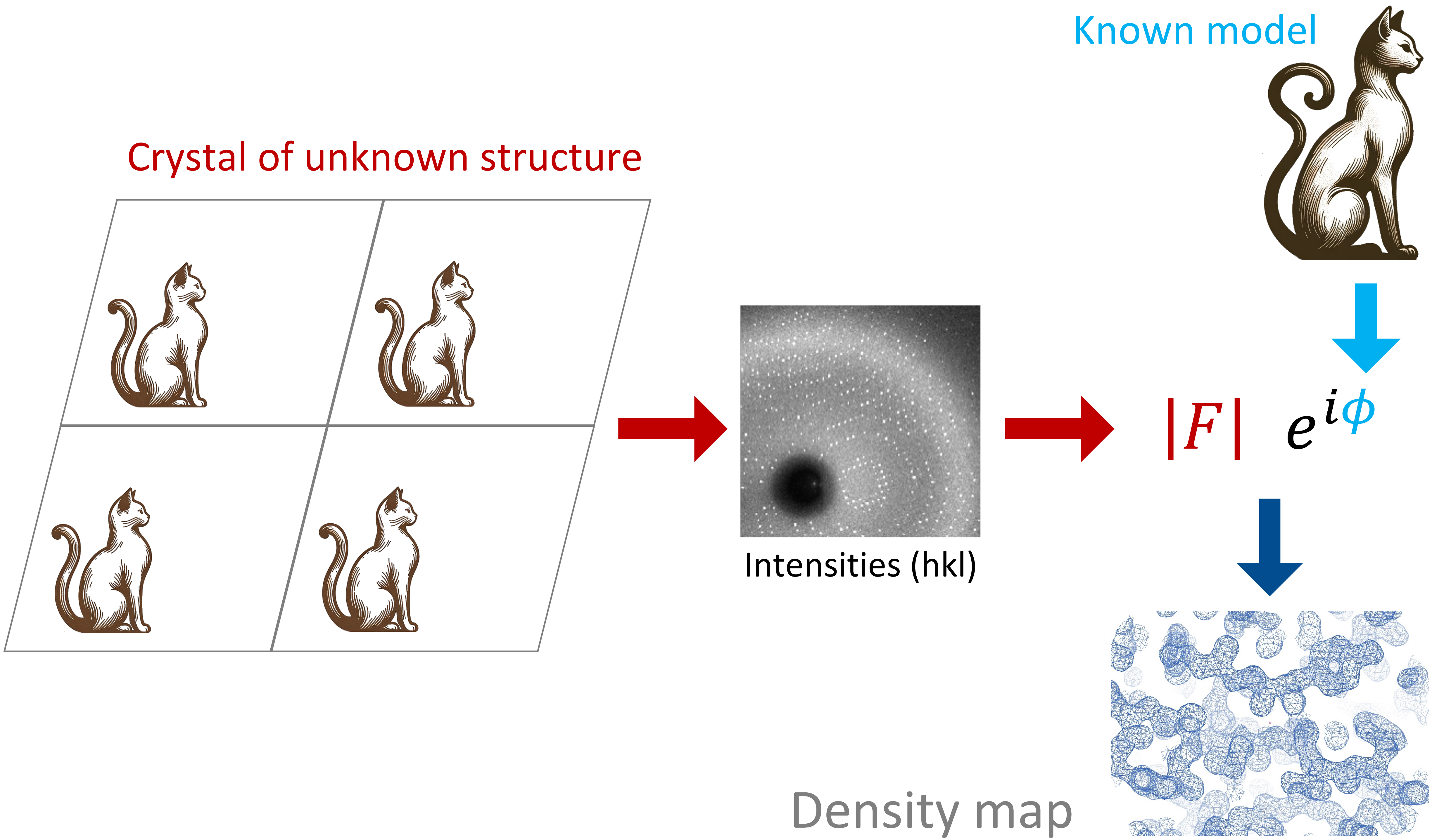
Known model

Crystal of unknown structure

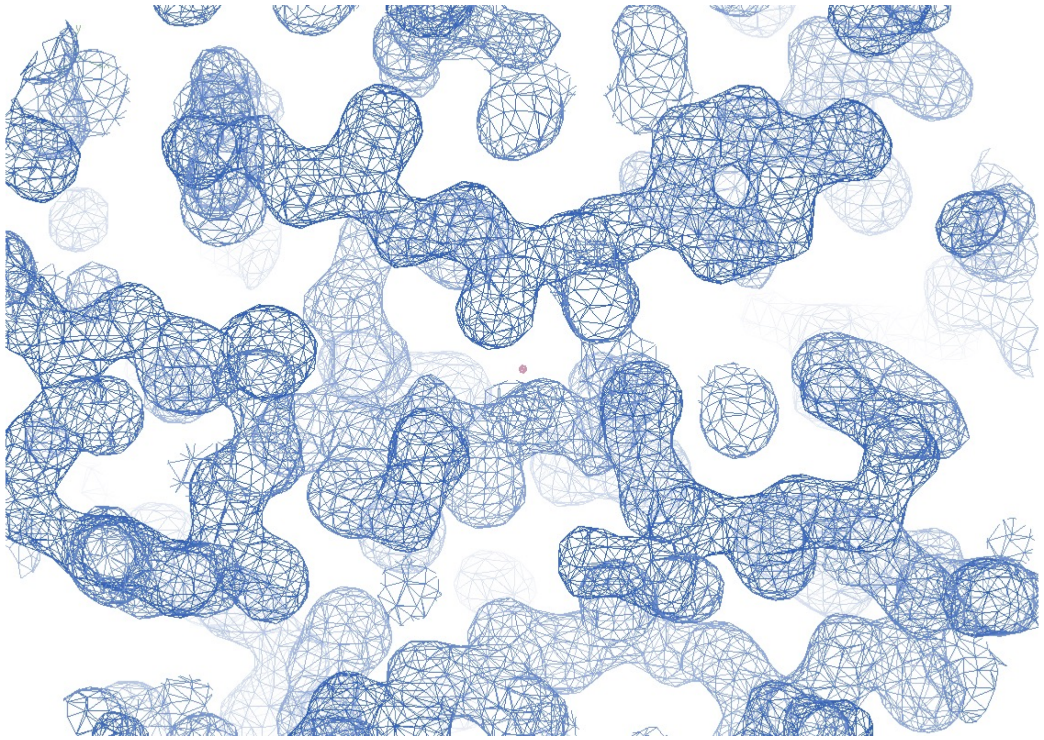


Known model provides initial estimates of the phases of the unknown structure.

Molecular Replacement (MR)

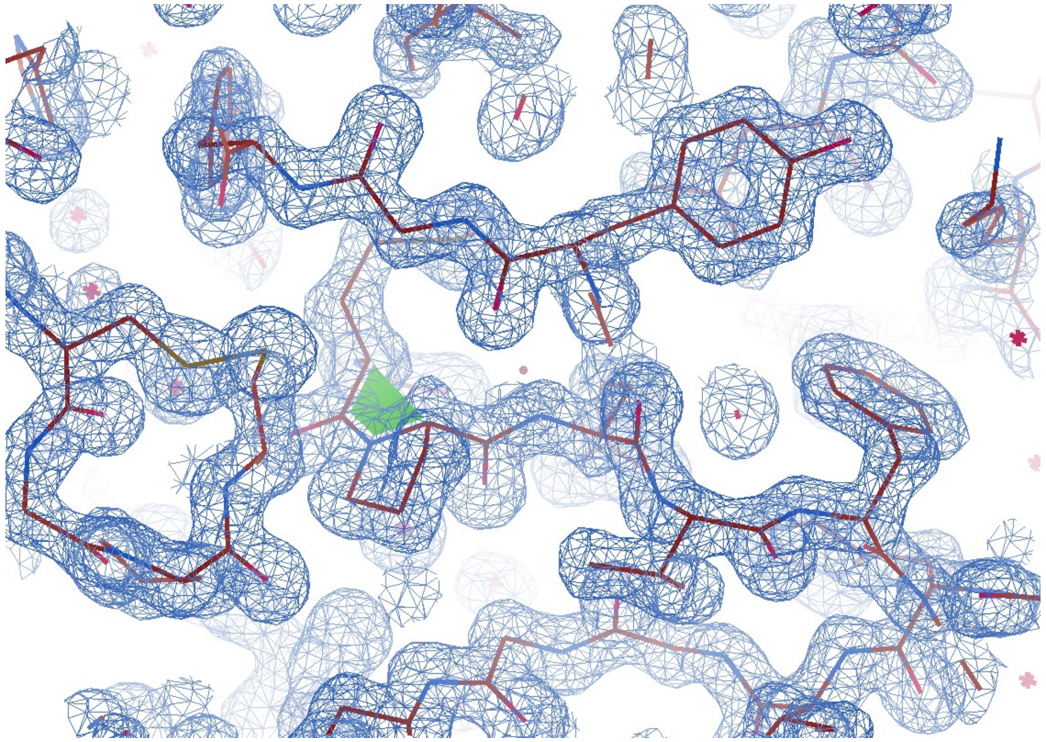


Molecular Replacement (MR)



If we know the density…

Molecular Replacement (MR)



If we know the density…

… then we can
determine the structure

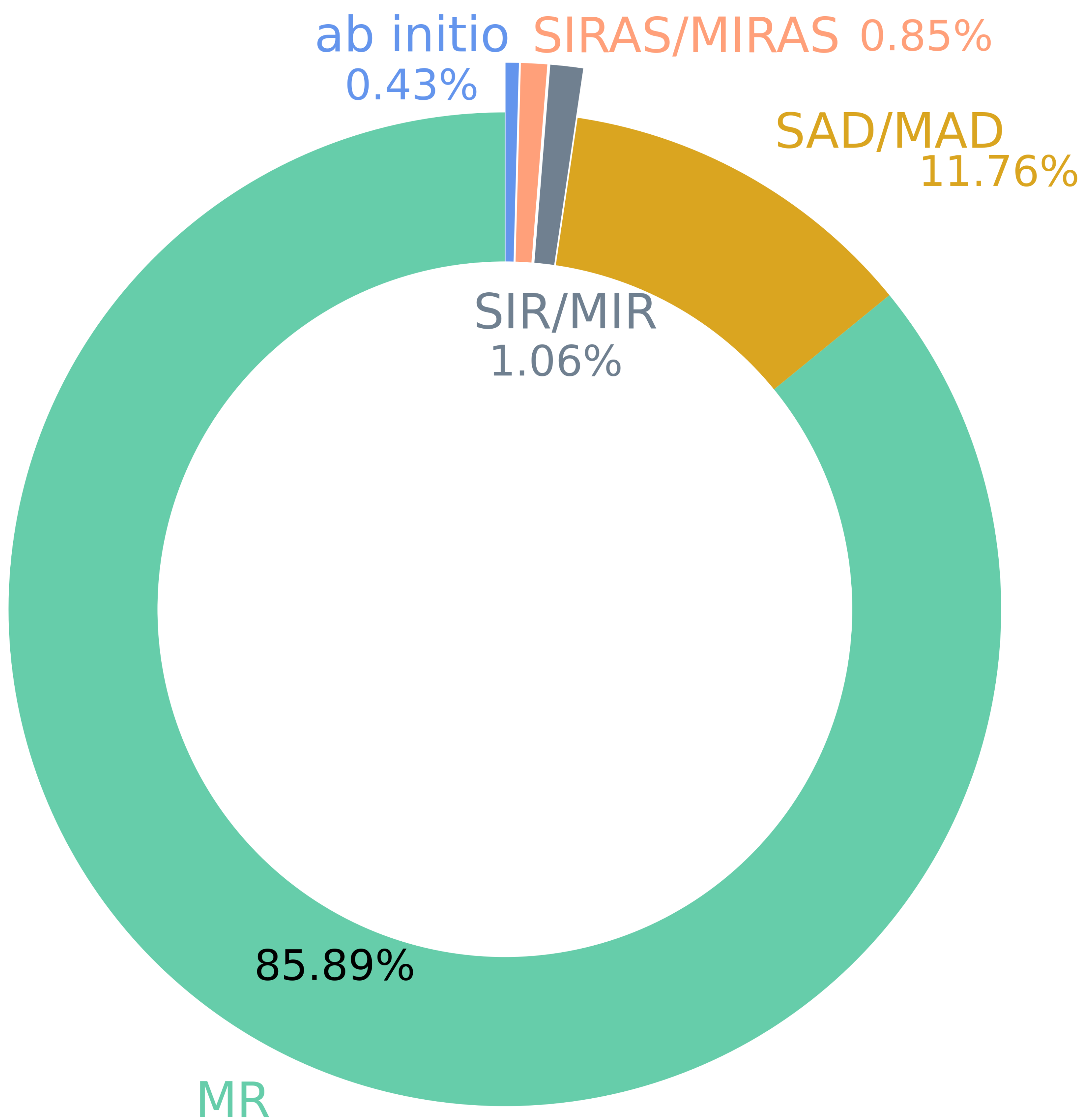
Phasing methods in the PDB

MR method:

- Fast
- Cheap
- Highly automated

Known structure:

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



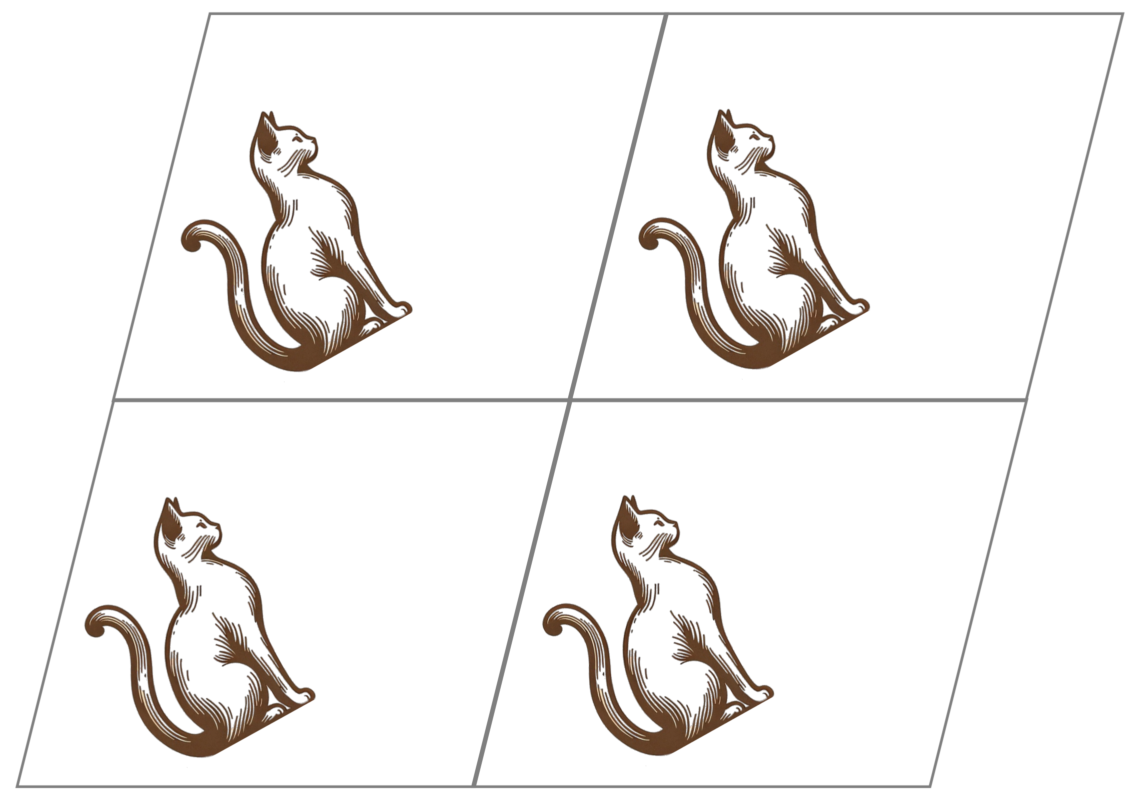
Note: Not all models in the PDB have (correct) info.

Molecular replacement: Approach

Try to match the known model with the unknown structure.

Search model

Crystal of unknown structure



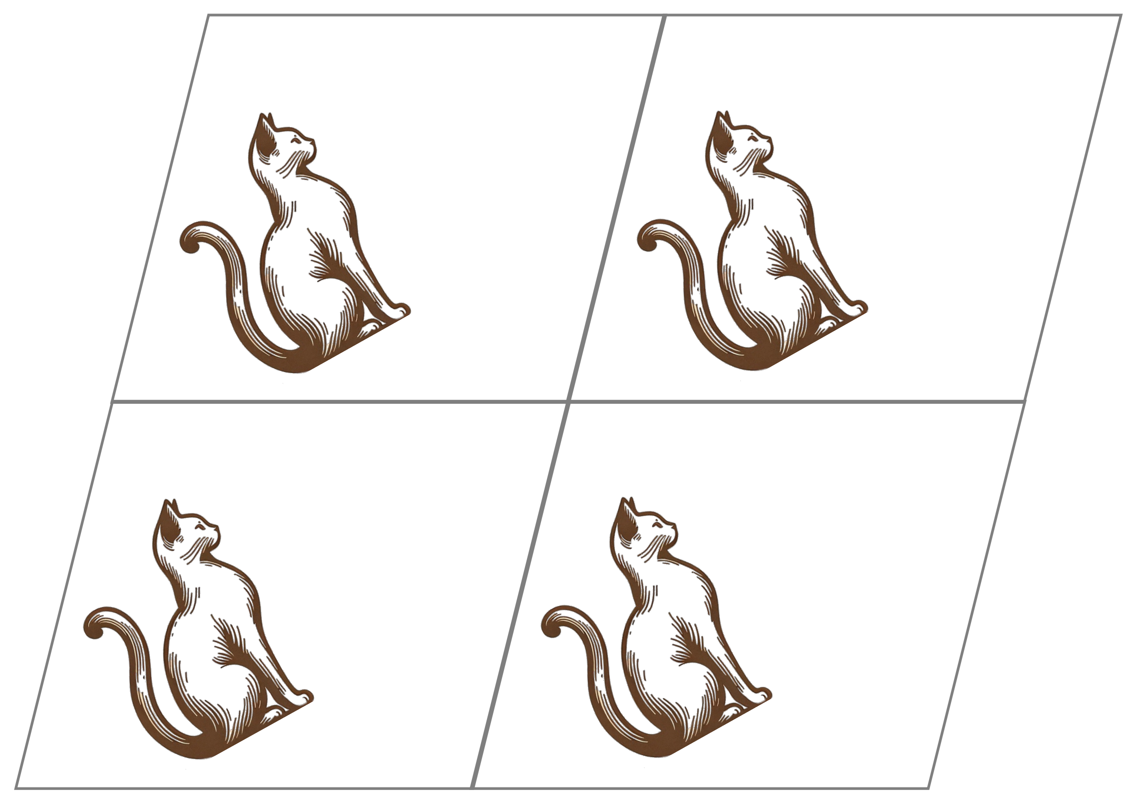
Find the rotation and translation of the search model so that it matches the unknown structure.

The search model

- Finding a suitable search model is critical step in MR.
- Should provide a high proportion of the scattering from the target structure with high accuracy (low r.m.s.d.).

Not similar to the target

Crystal of unknown structure



What can be used as search model?

Search model should be similar to the target structure.

1) **Homologue structure**

If the sequence is similar → structures may be similar

- Sequence comparison search
- Prune regions of large sequence diversity
- Truncate side-chains

Search model

Unknown structure

What can be used as search model?

Search model should be similar to the target structure.

1) Homologue structure

If the sequence is similar → structures may be similar

- Sequence comparison search
- Prune regions of large sequence diversity
- Truncate side-chains

Search model

Unknown structure

What can be used as search model?

Search model should be similar to the target structure.

2) **Predicted structure (AlphaFold2, RosettaFold)**
   Can be very accurate.
   - Distortions
   - Incorrect domain relationships

   → Remove low pLDDT regions, split into domains

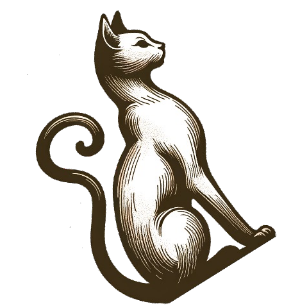
Predicted model

Unknown structure

Molecular replacement: Scoring

Compare observed and calculated diffraction.

Poor score



Good score

Different approaches:

- Patterson function
- Maximum-likelihood Methods (Phaser)

MR Scoring: Maximum Likelihood Method

"For any postulated orientation and position of the model, what is the probability of obtaining the structure amplitudes that we observe?"

Explicitly models errors
- Experimental uncertainties
- r.m.s. coordinate error of the search model

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

Maximum Likelihood Scoring in Phaser

LLG = Log Likelihood gain

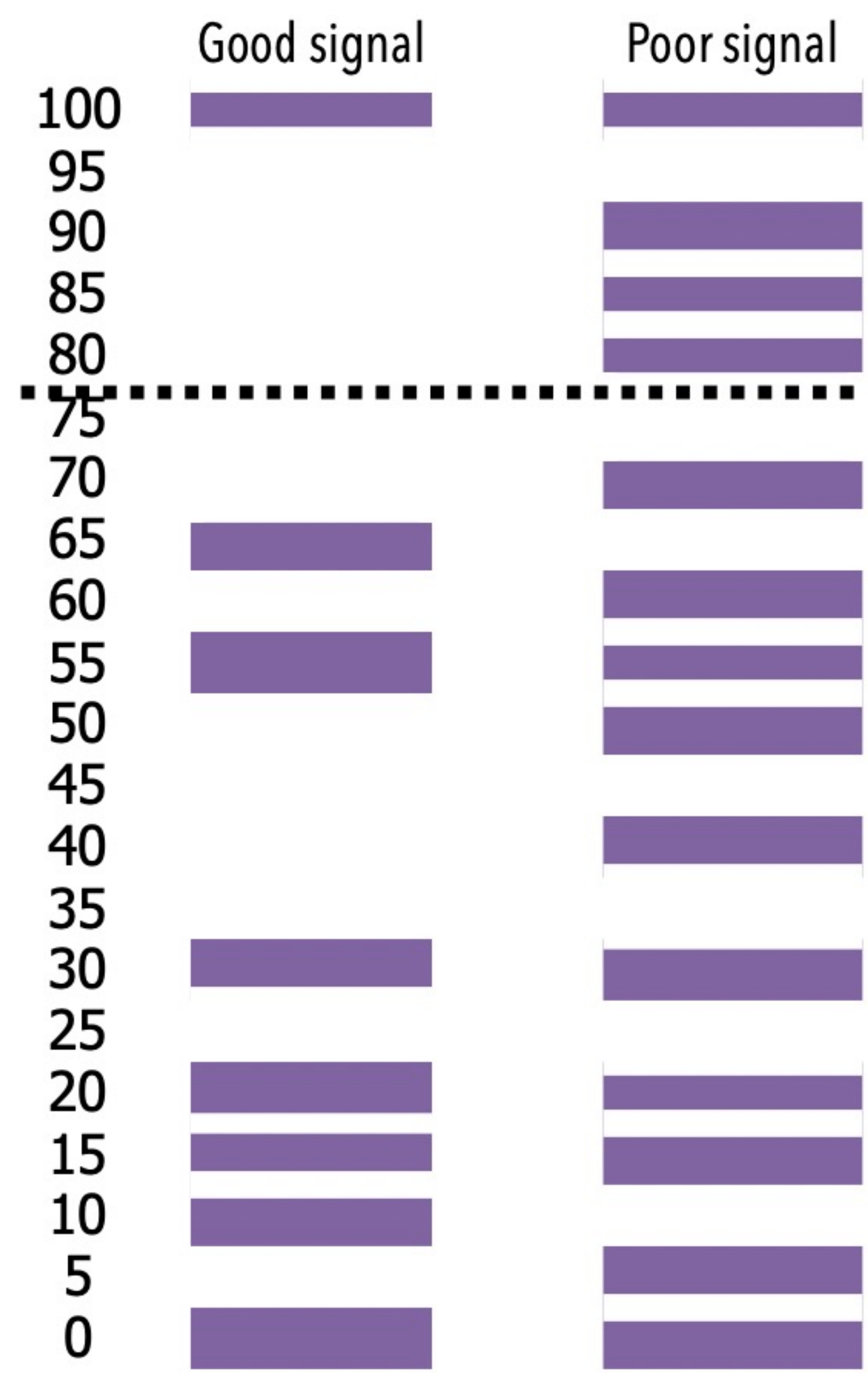
→ It measures how much better the data can be predicted with the search model than with a random distribution of the same atoms.

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

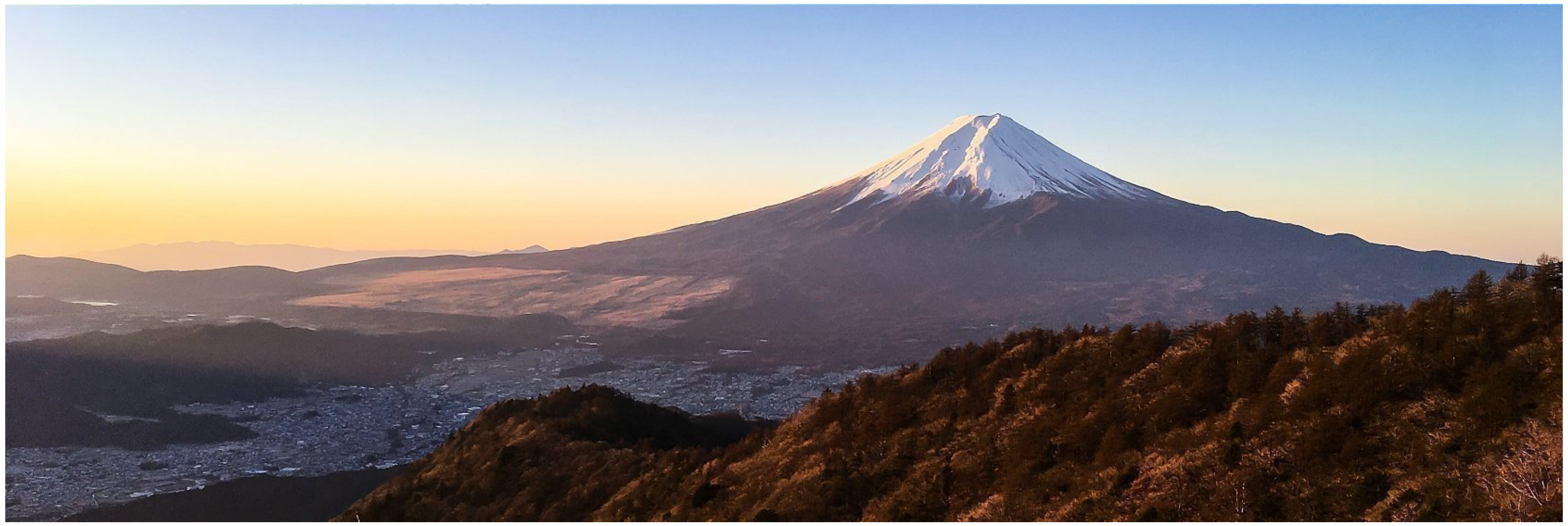
Maximum Likelihood Scoring in Phaser

Select solutions that are over 75% of the difference between the top peak and the mean.

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

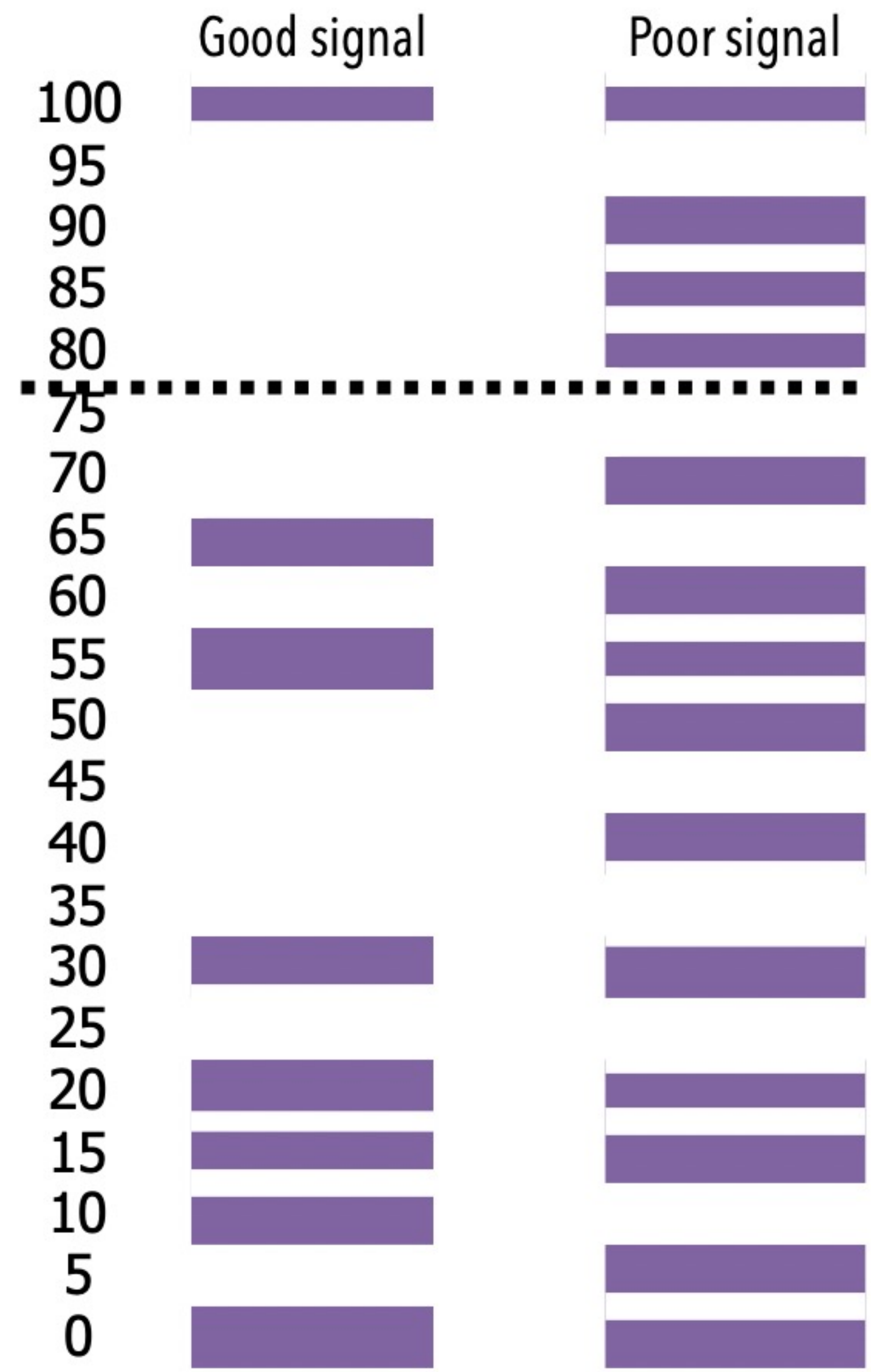


Maximum Likelihood Scoring in Phaser



Good signal

Poor signal



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

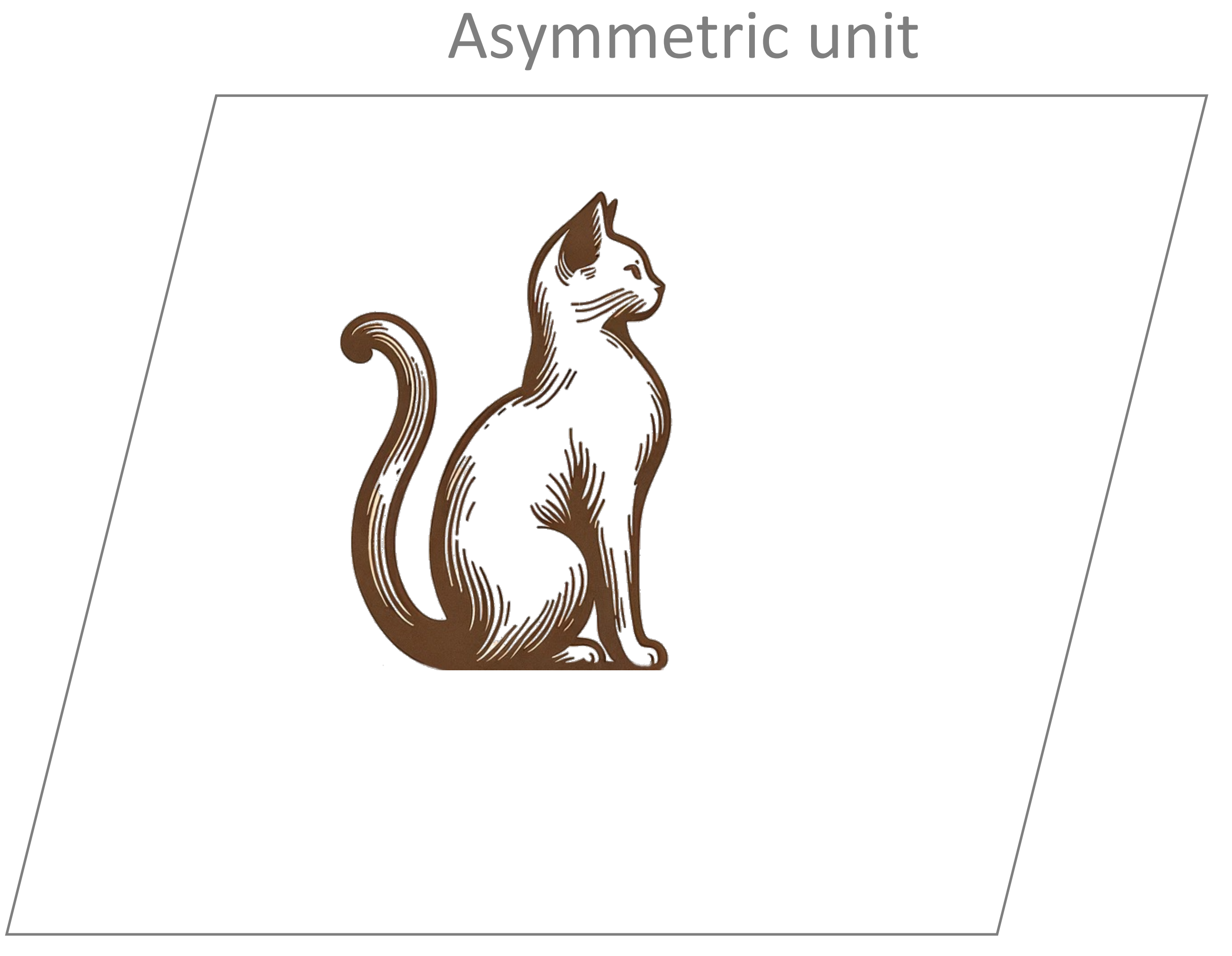
What is needed to run Phaser MR

- Reflection data
- Search model
- Error estimation of the search model
  - Homologue: sequence identity
  - Predicted model: r.m.s.d. (1.0 Å)
- ASU content
- Twinning
- (tNCS)

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 1

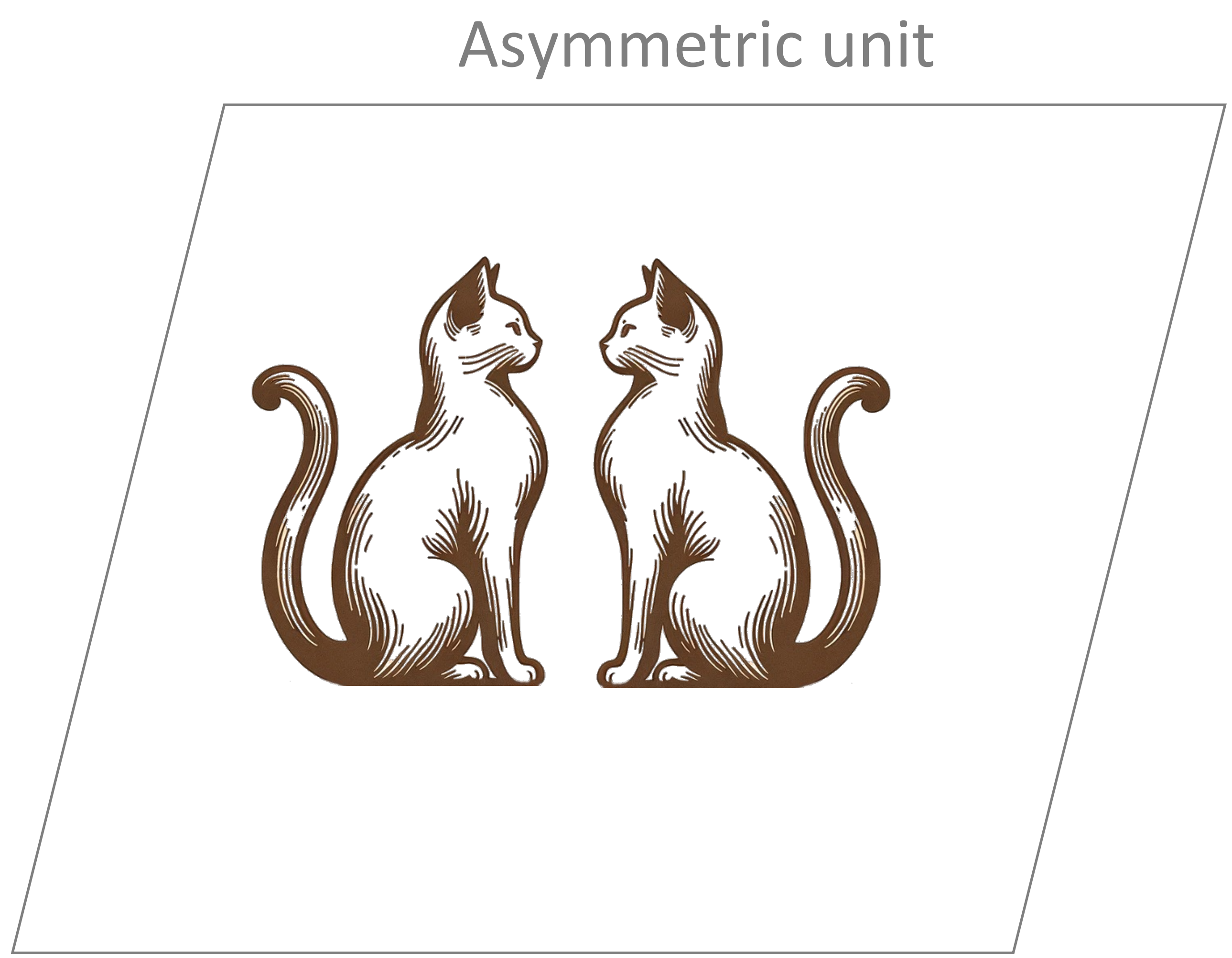


One copy in the asymmetric unit.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 2

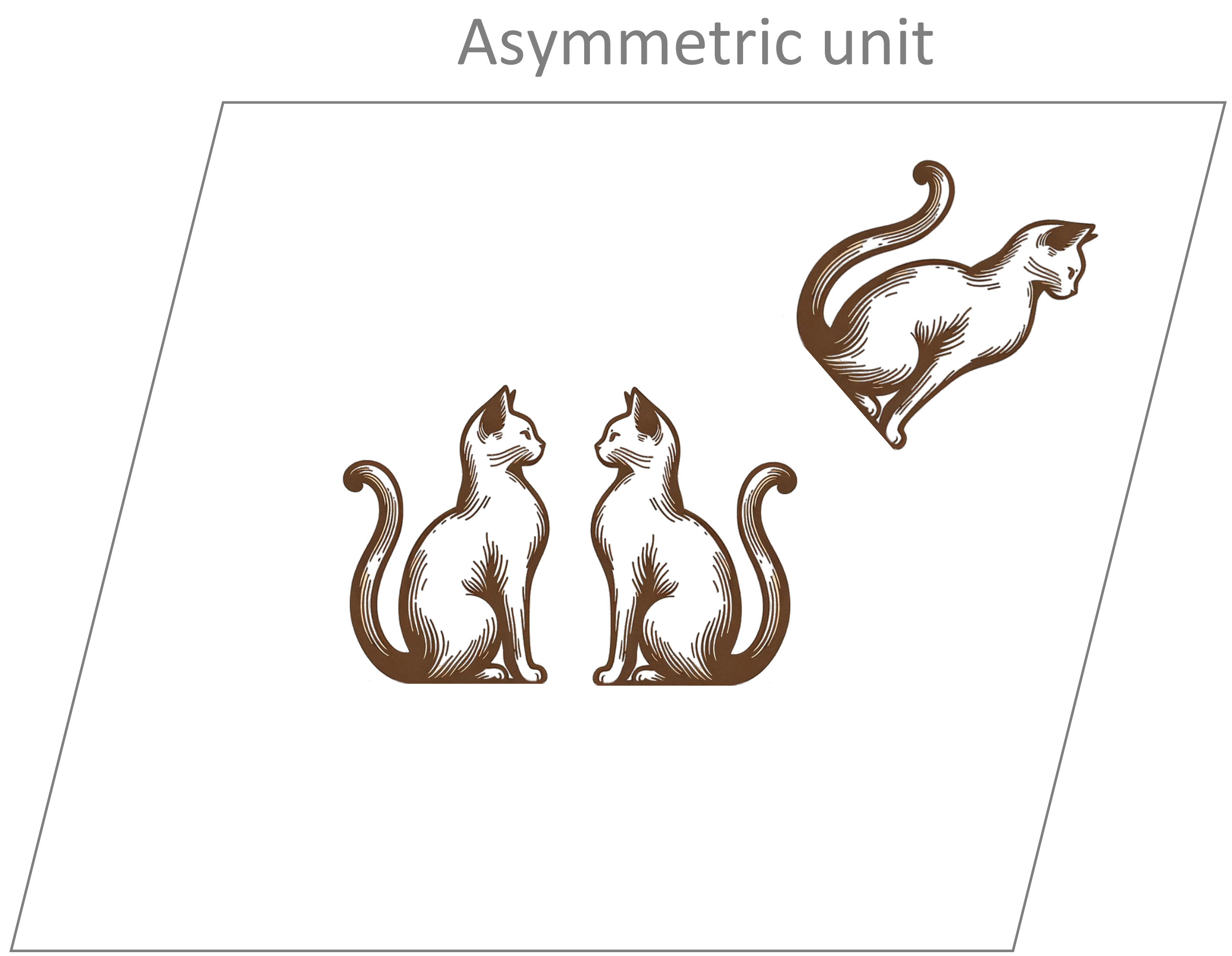


Two copies in the asymmetric unit, rotational symmetry.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 3

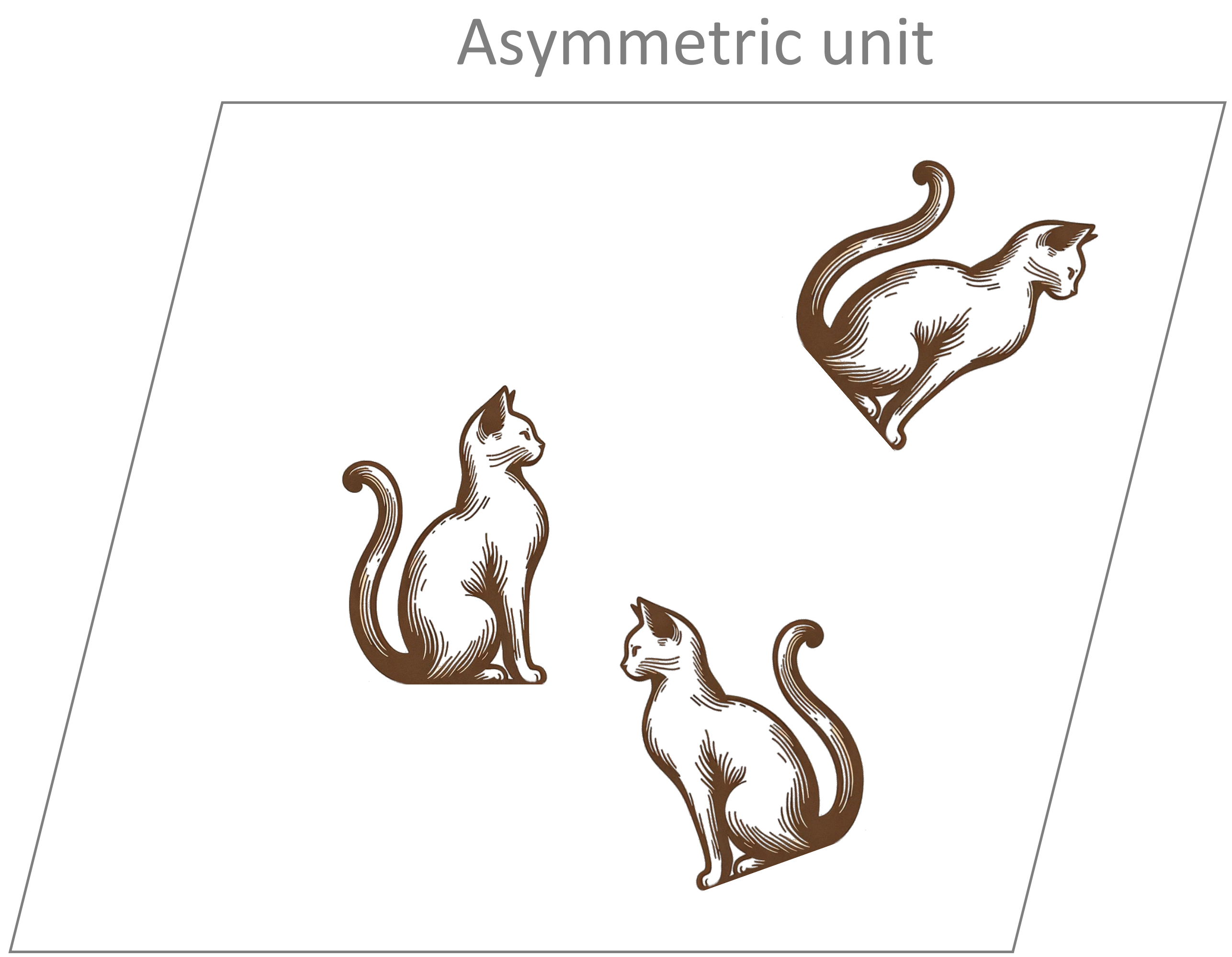


Three copies in the asymmetric unit, two in rotational symmetry, one unrelated.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 3

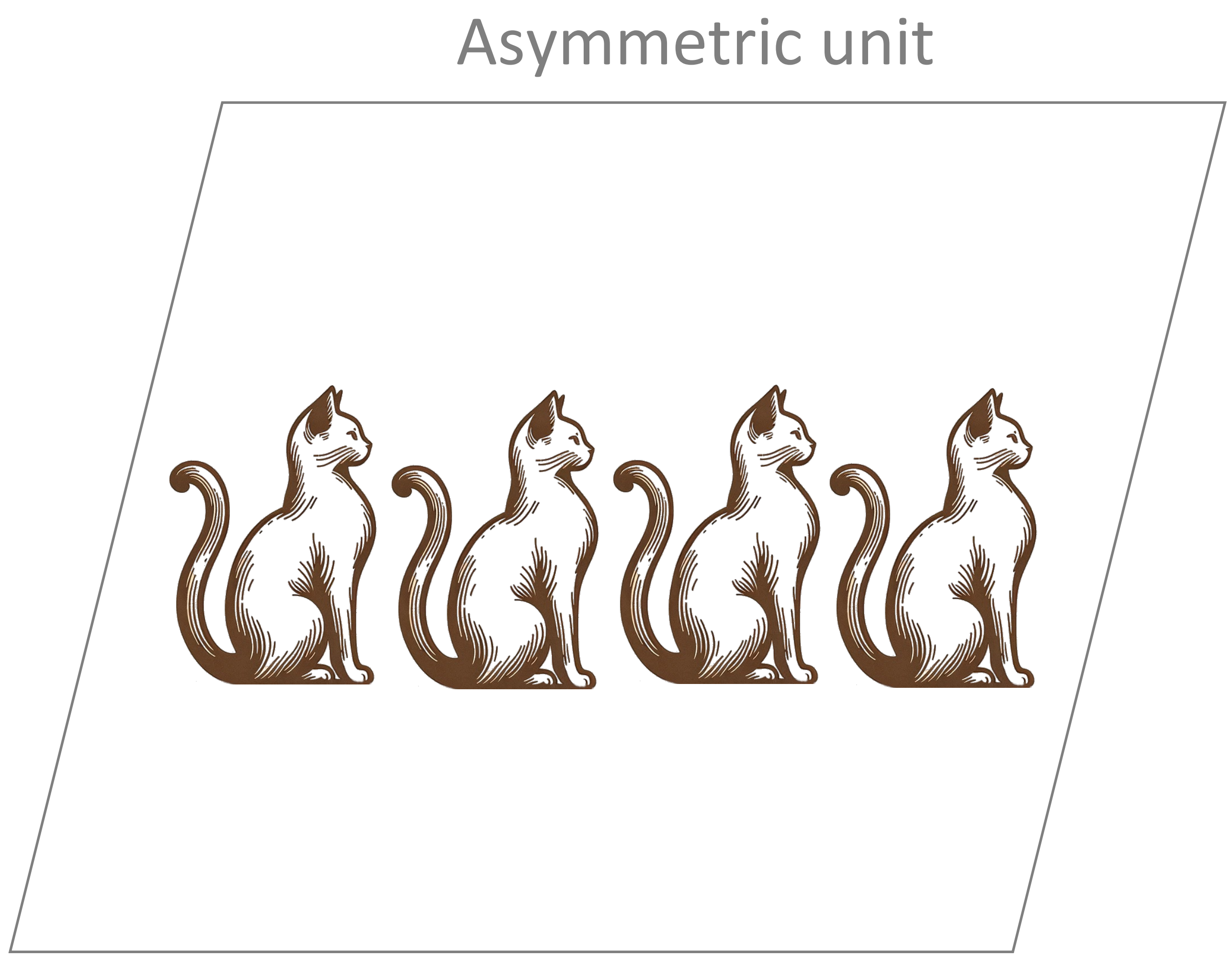


Three copies in the asymmetric unit.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

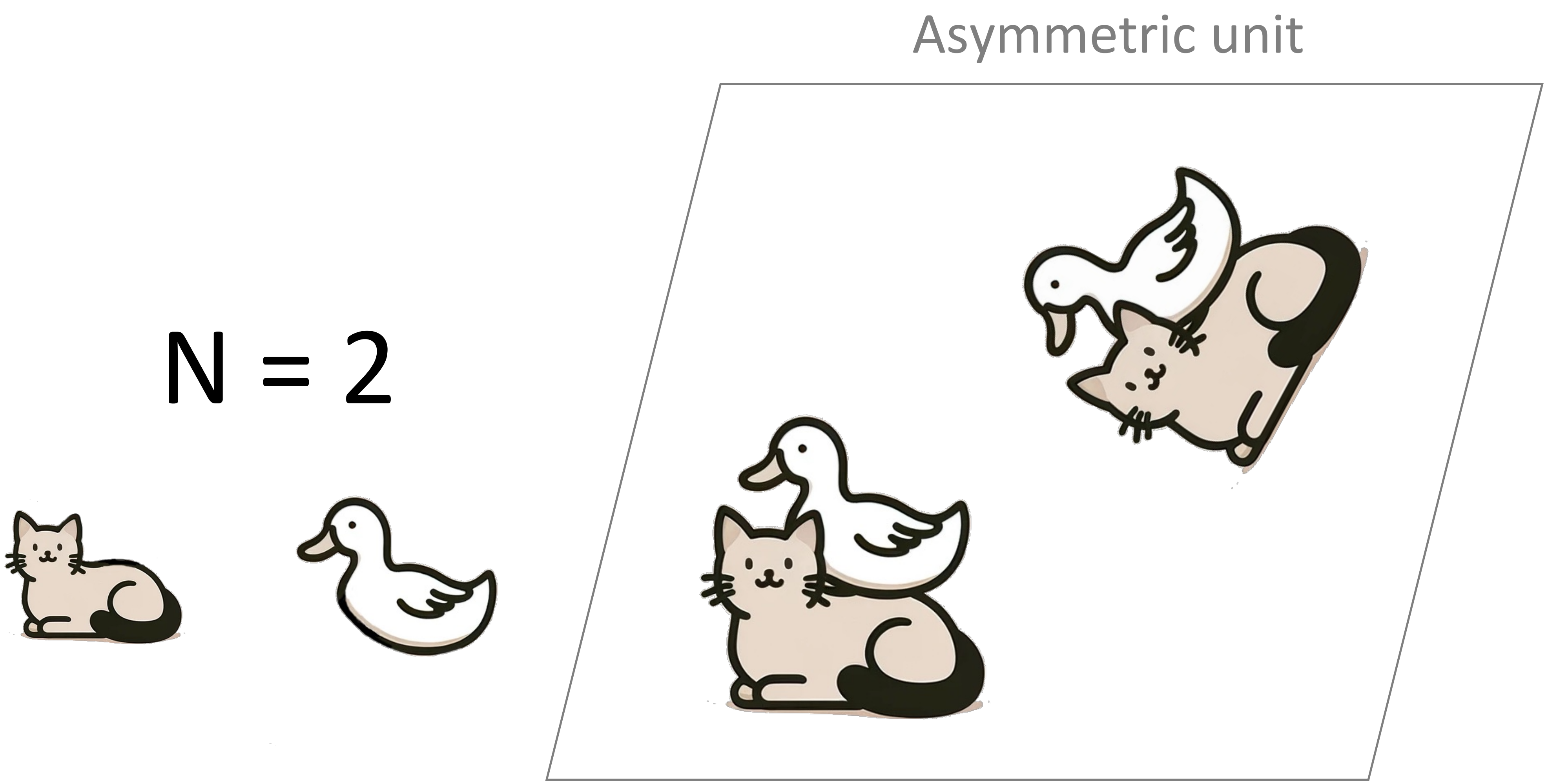
N = 4



Four copies in the asymmetric unit, translational non-crystallographic symmetry.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?
- Different domains?



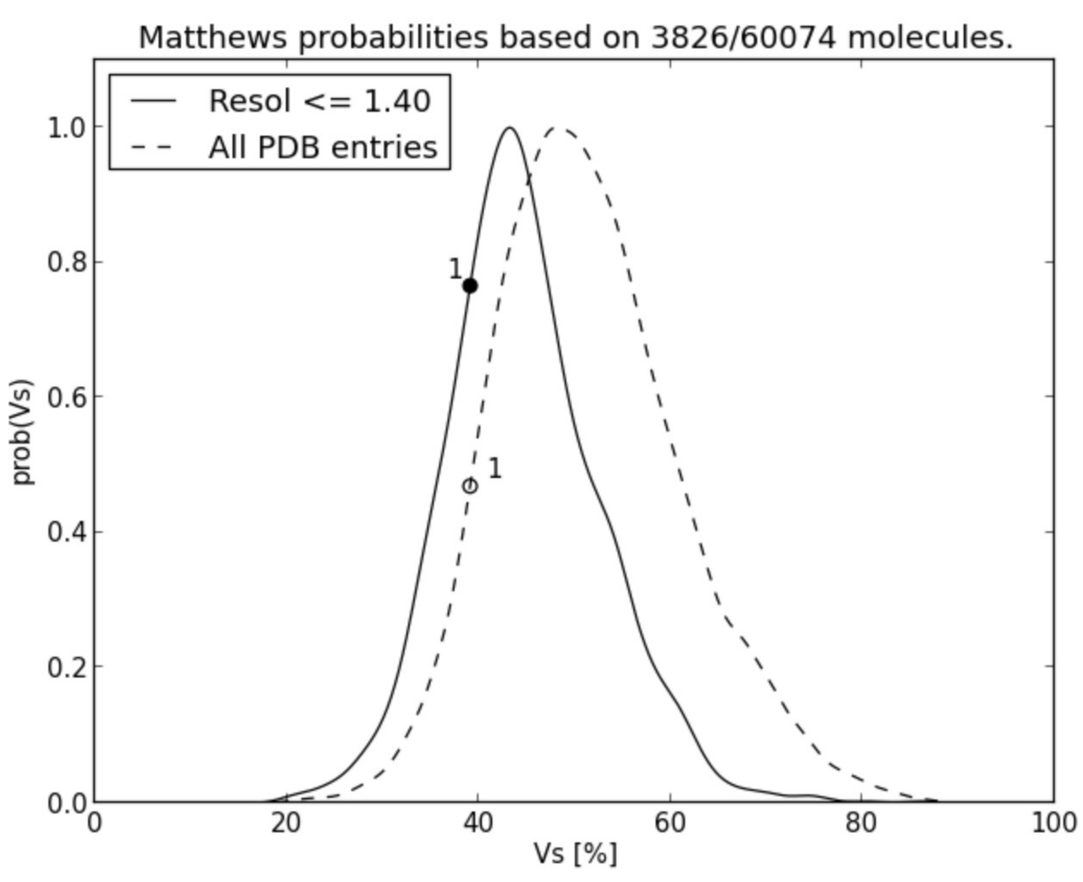
If mutual orientation is unknown, search for the domains separately.

Estimating the number of molecules in the ASU

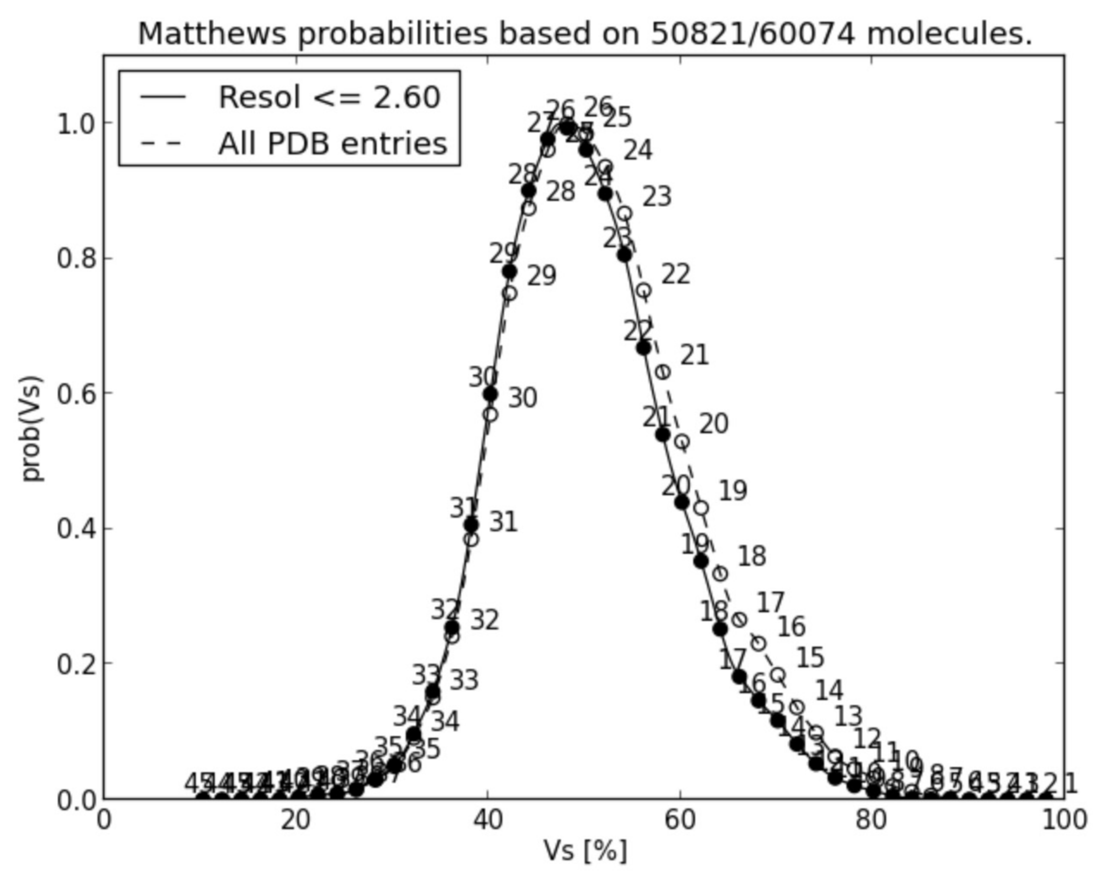
Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

Compare your value of V_M with histograms for known structures (from PDB) → choose most probable.

Few possibilities



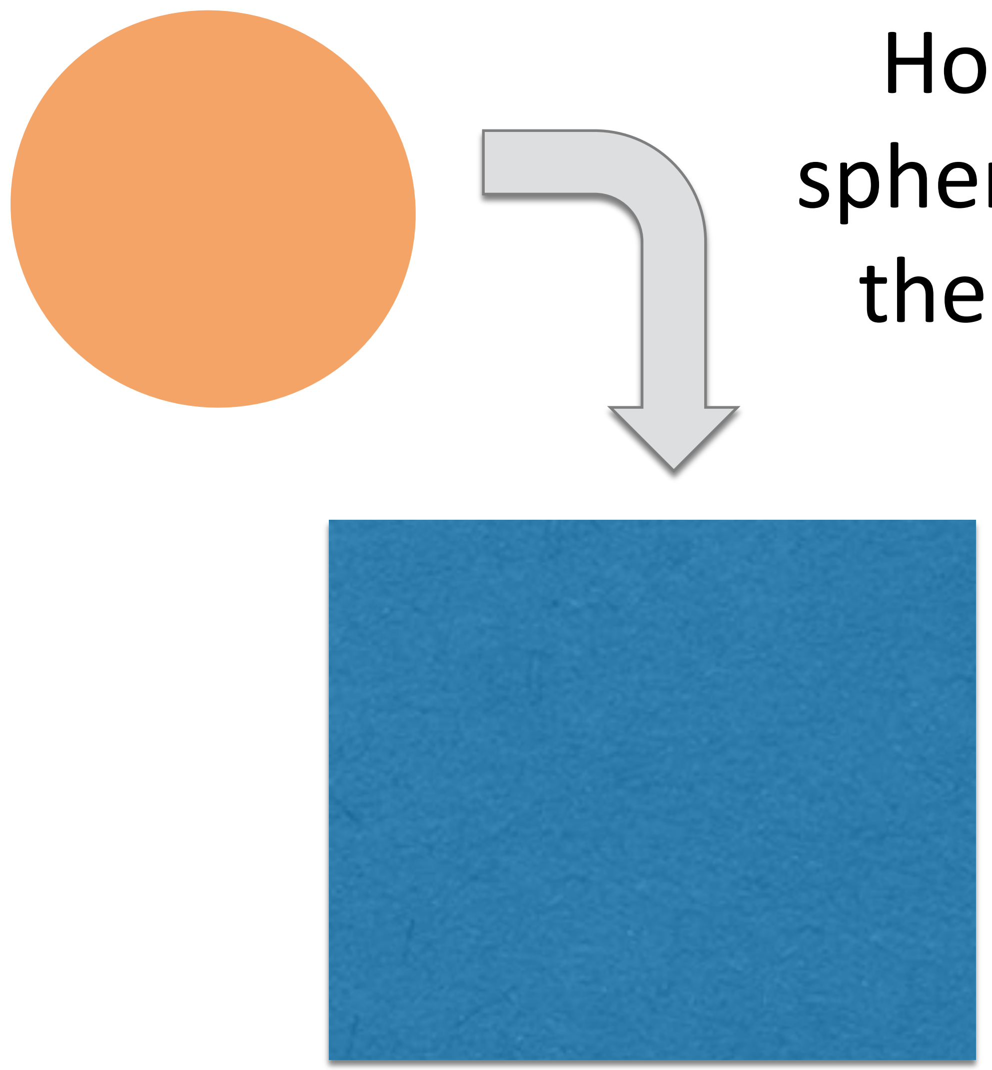
Many possibilities



Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

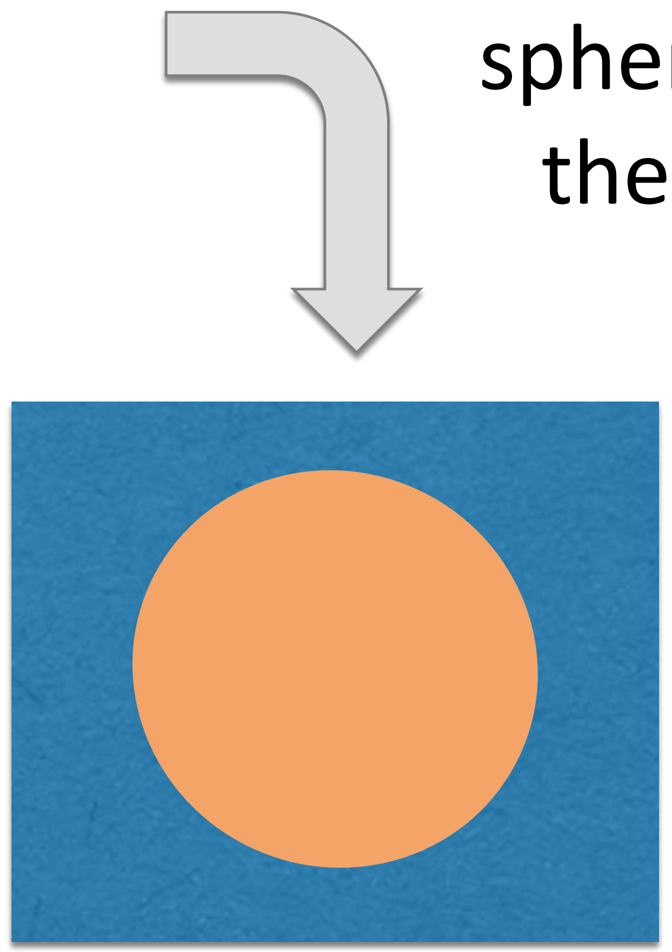
How many spheres fit into the square?



Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

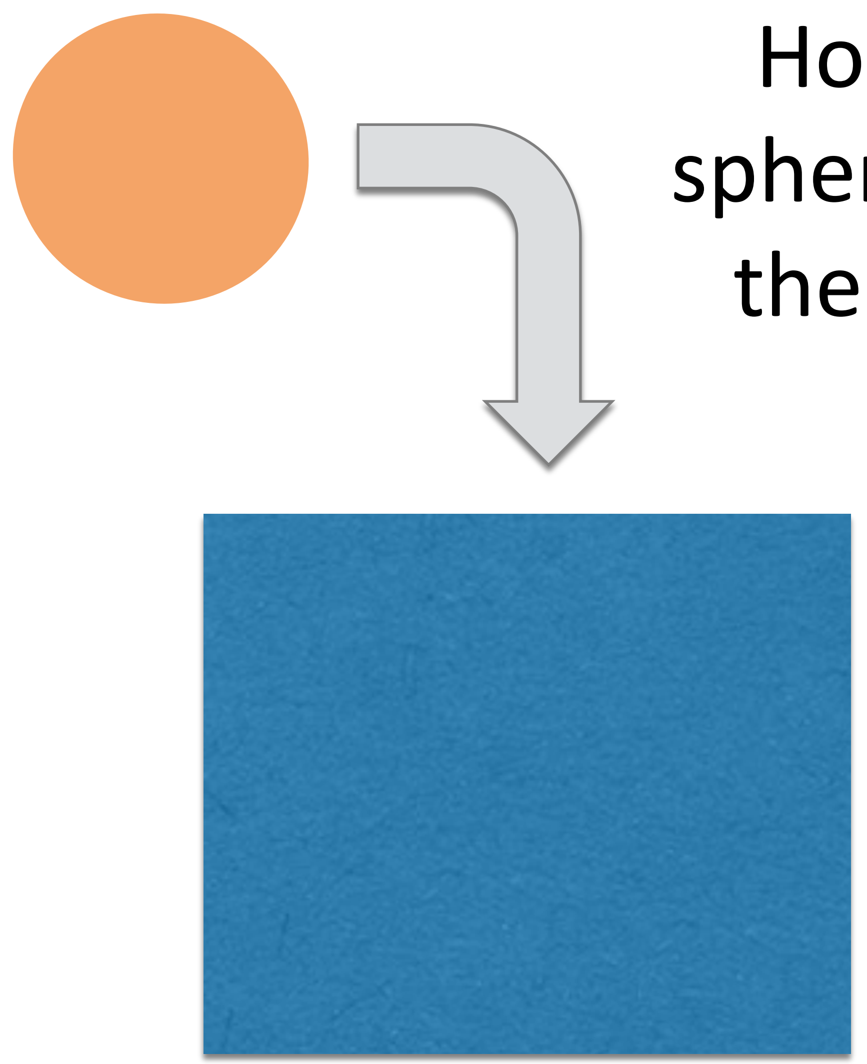


Clearly only one sphere

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

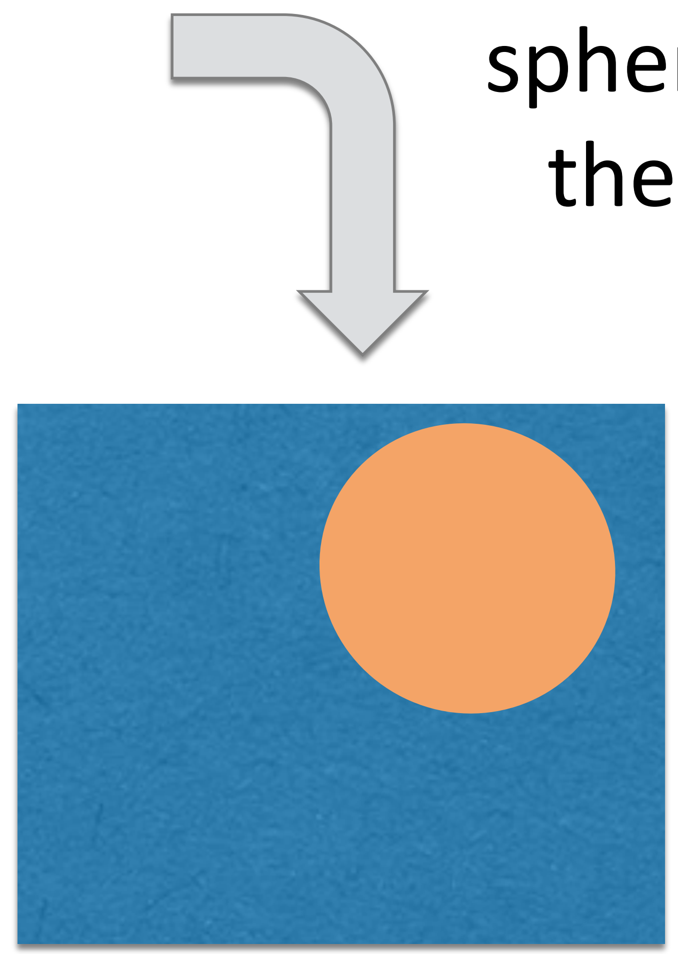
How many spheres fit into the square?



Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

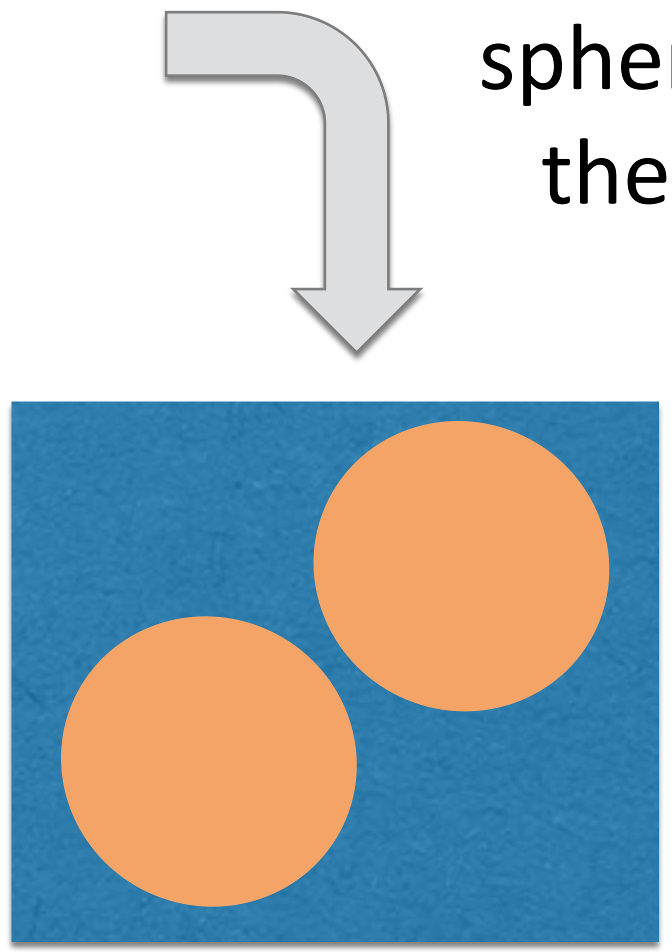


There could be only one…

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

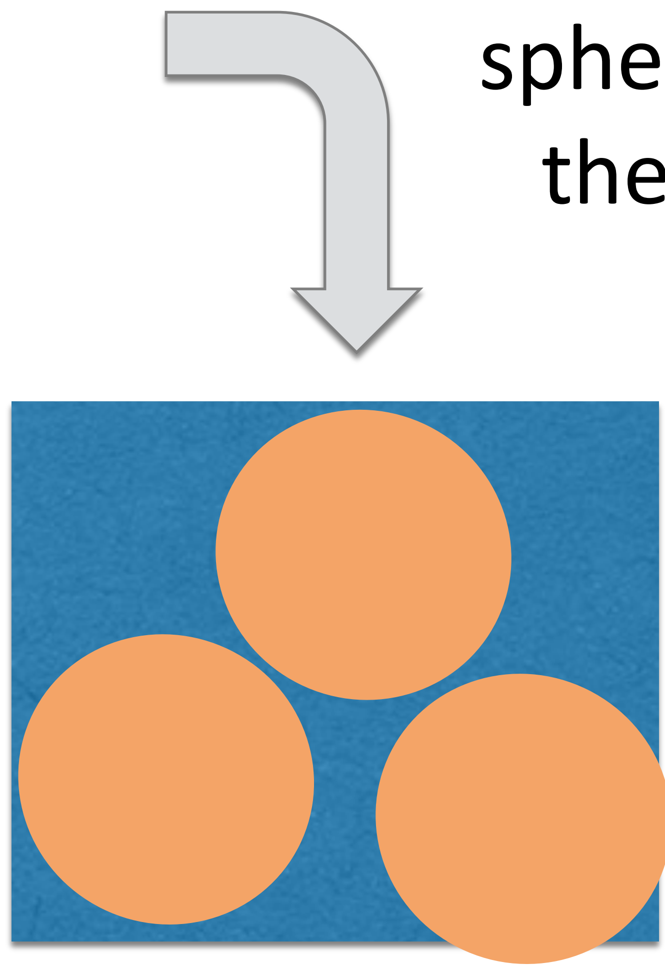


... or maybe two

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

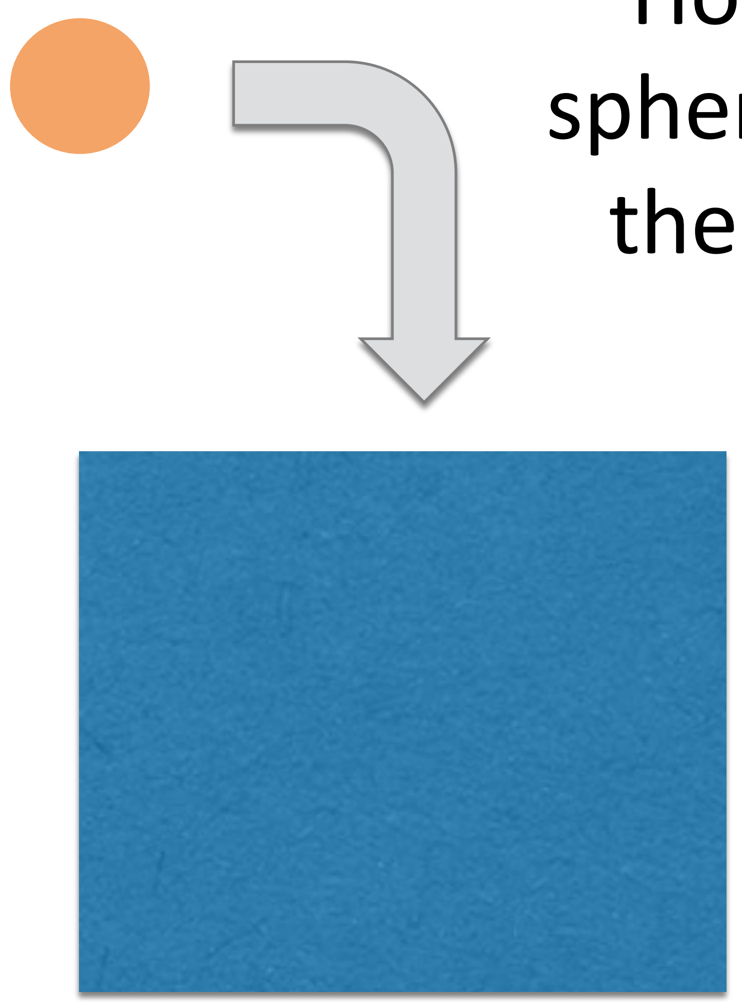


...but not three

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

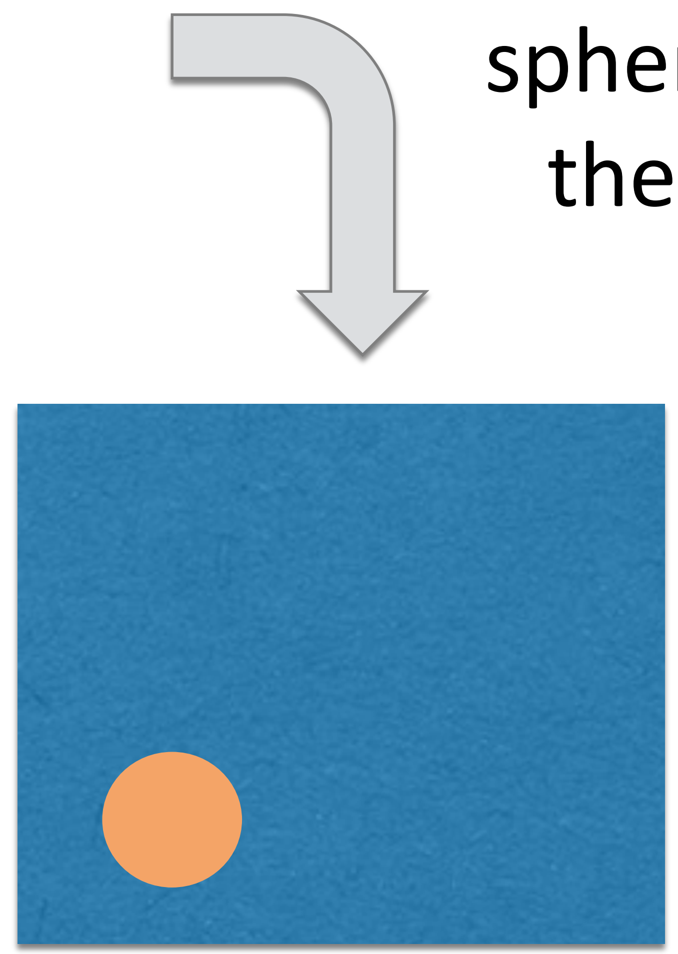
How many spheres fit into the square?



Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

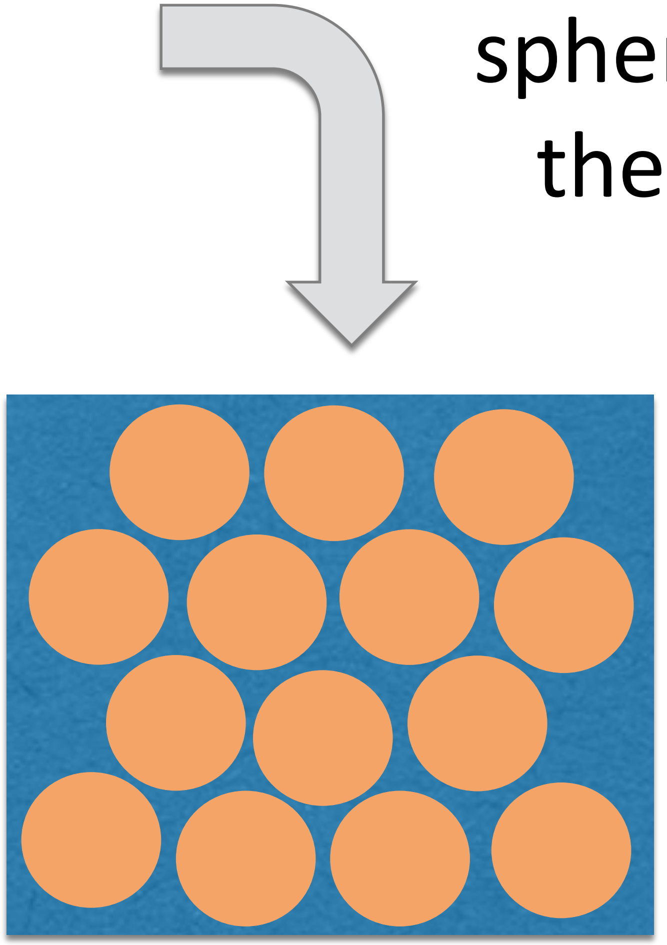


Only one

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

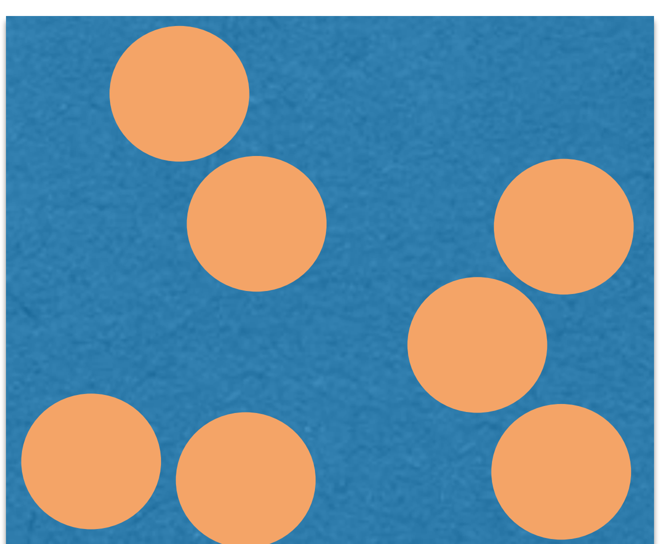


Many (14)

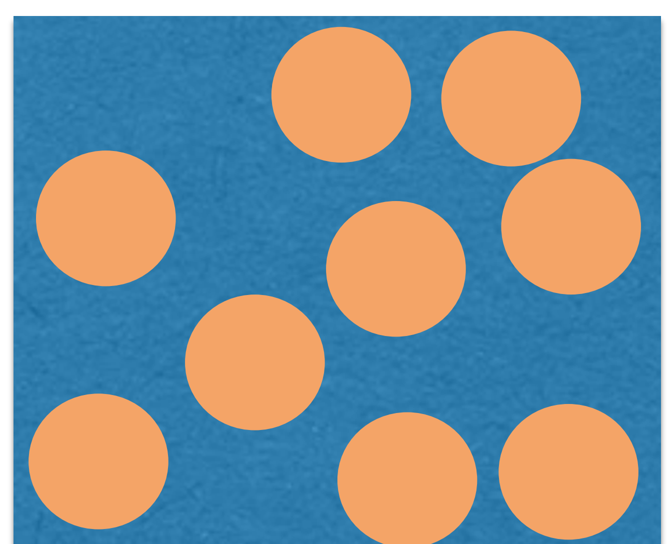
Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?



7



9

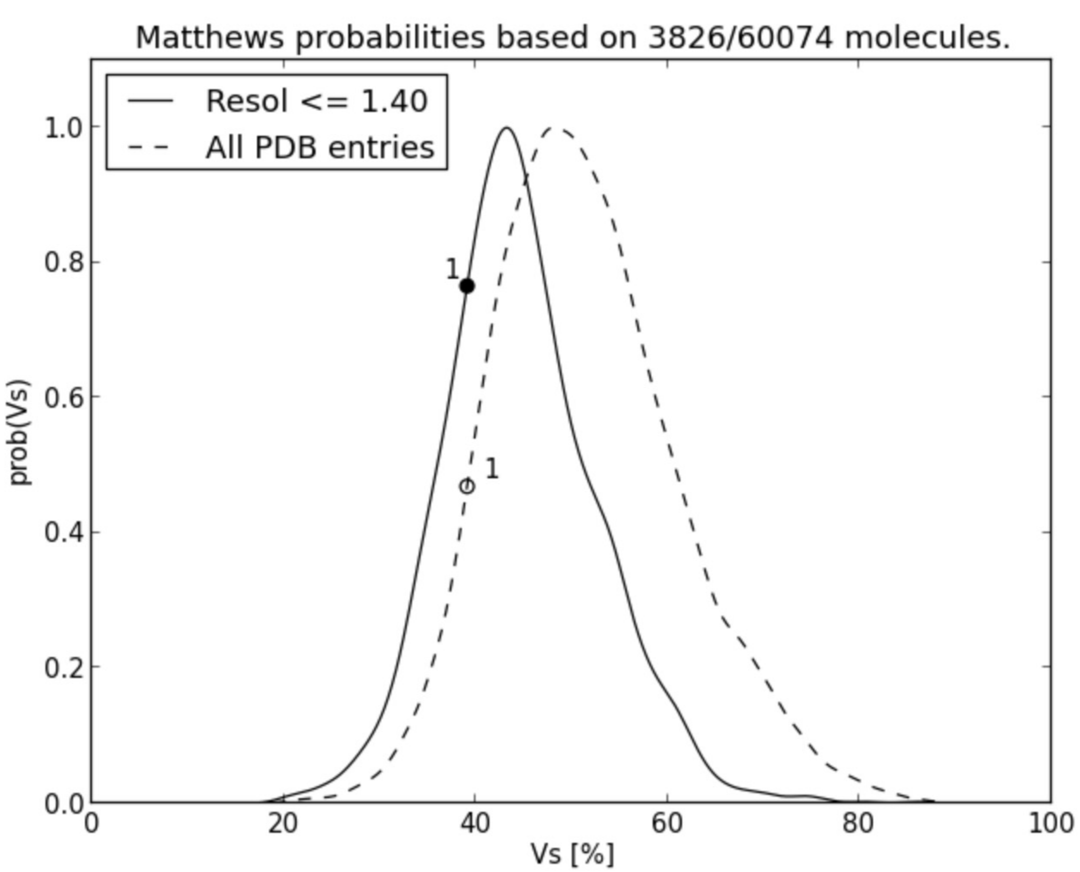
Most likely: something in between

Estimating the number of molecules in the ASU

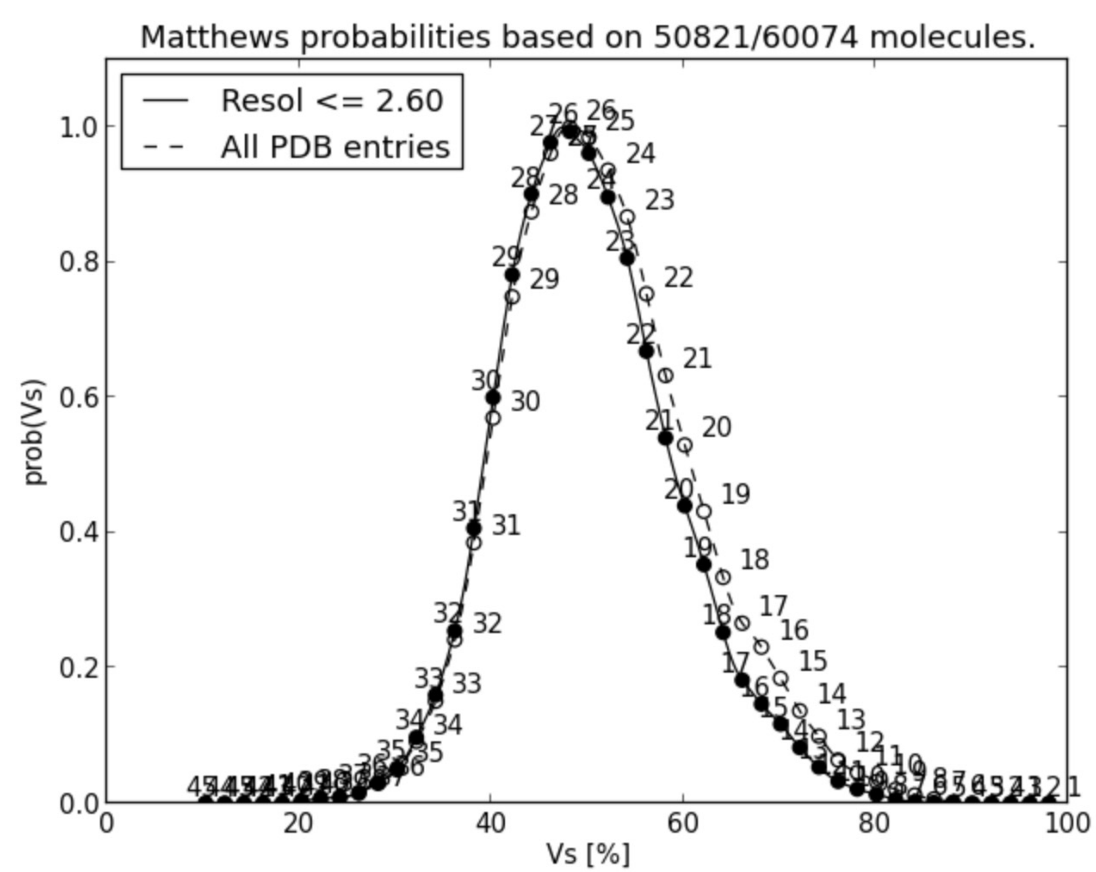
Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

Compare your value of V_M with histograms for known structures (from PDB) → choose most probable.

Few possibilities



Many possibilities



Key points

1) MR: Use a previously known structure to solve a new structure

2) Known structure can be a homologue or a predicted model

3) Known structures may need to be modified

4) Estimate the ASU content

5) Be aware of data pathologies (twinning, tNCS)

MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

1) check Matthews coefficient

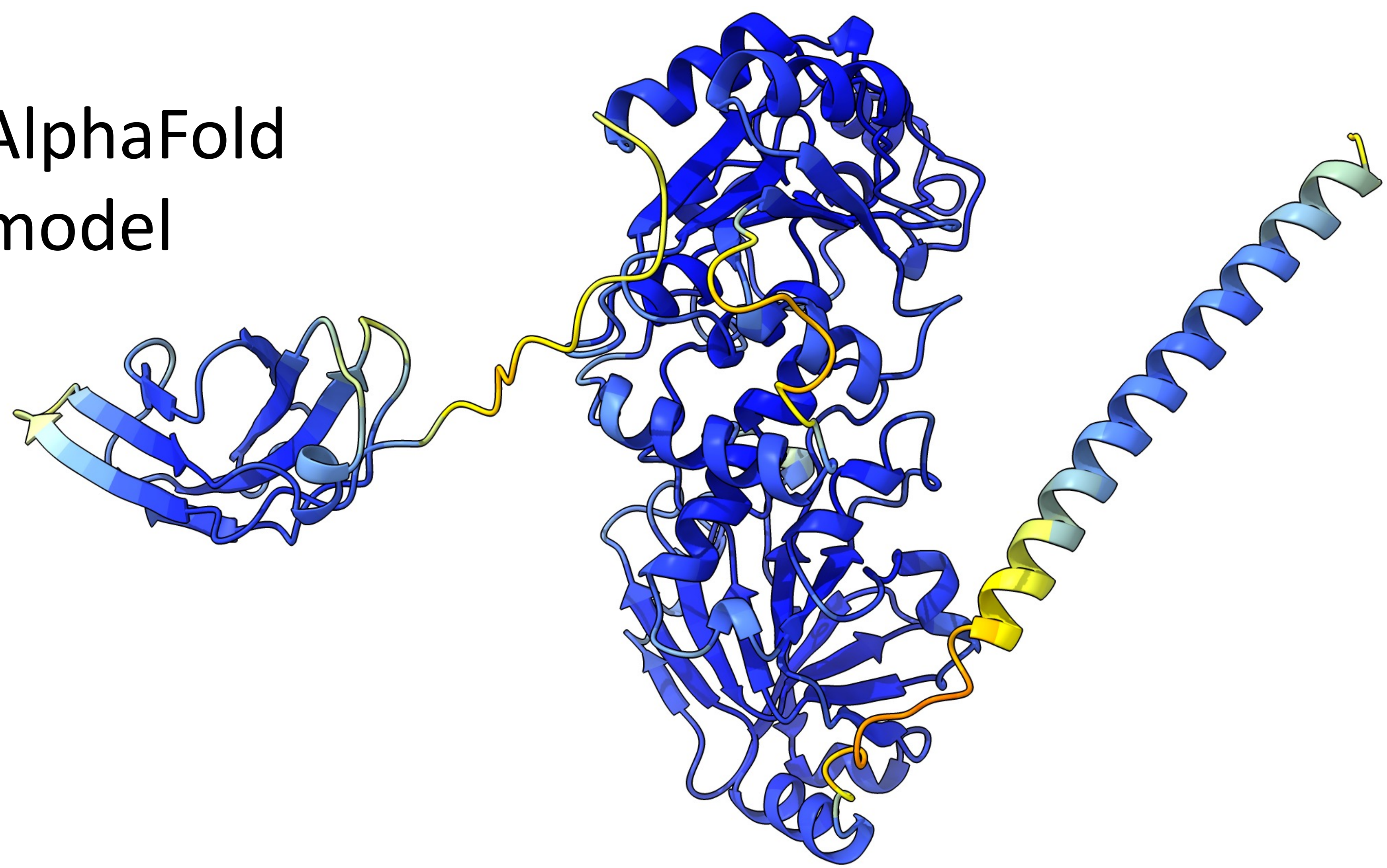
2) process predicted model

3) run Phaser molecular replacement

MR with an AlphaFold model

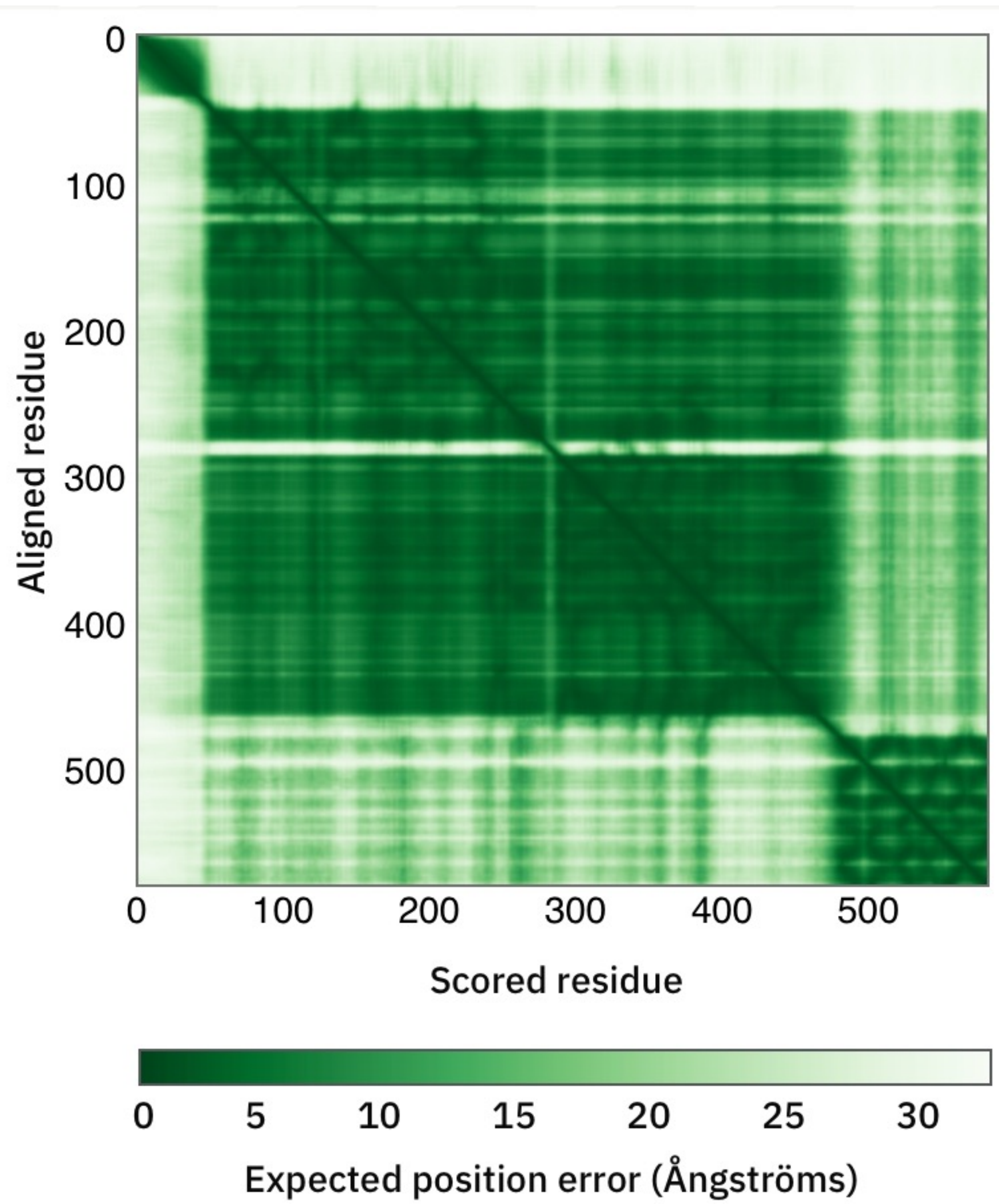
POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

AlphaFold model



MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2



Predicted alignment error (PAE) plot

→ Dark green: the predicted relationship between a pair of residues is likely to be accurate

→ Suggests 3 domains

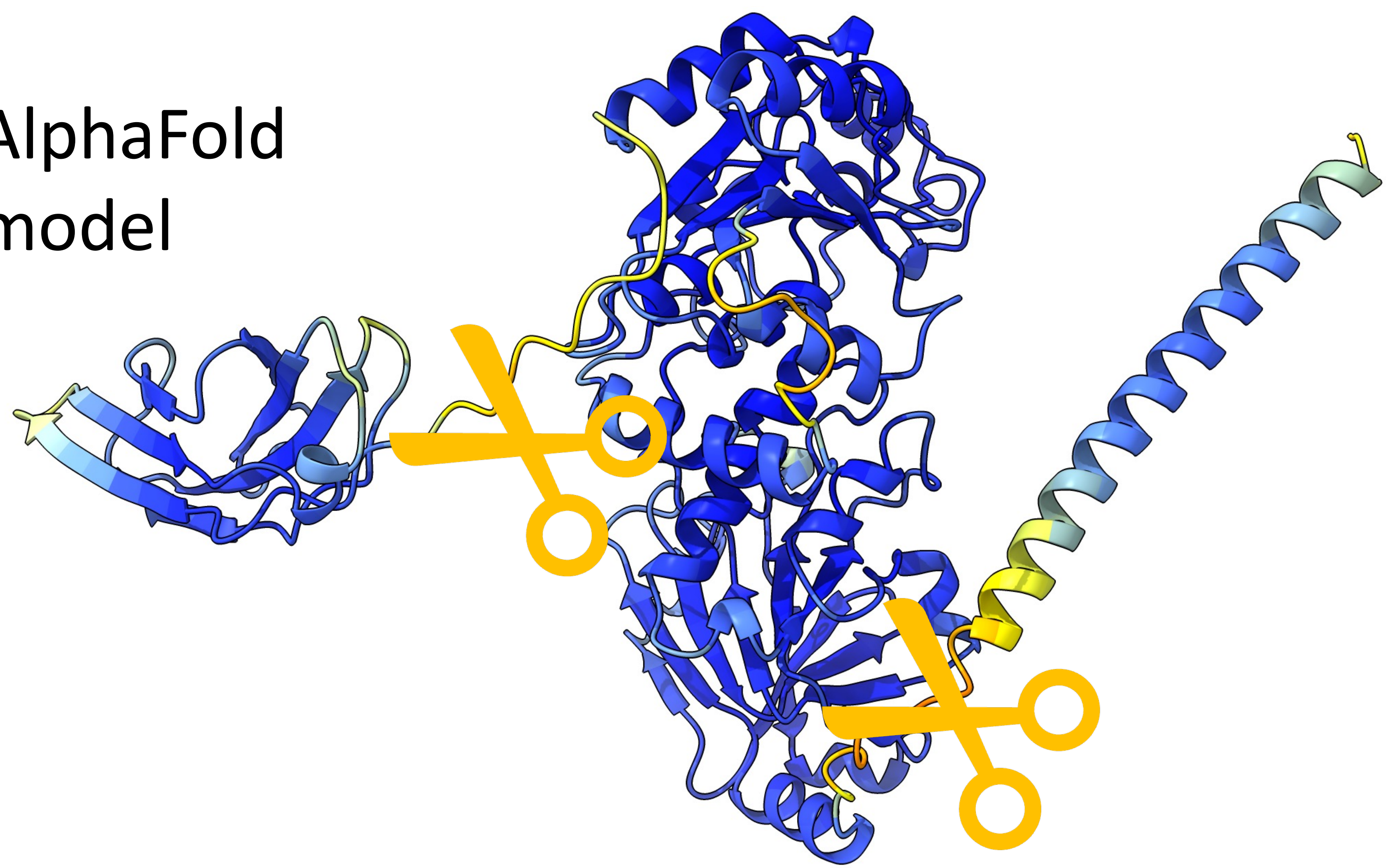
→ Mutual configuration is not so clear

MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

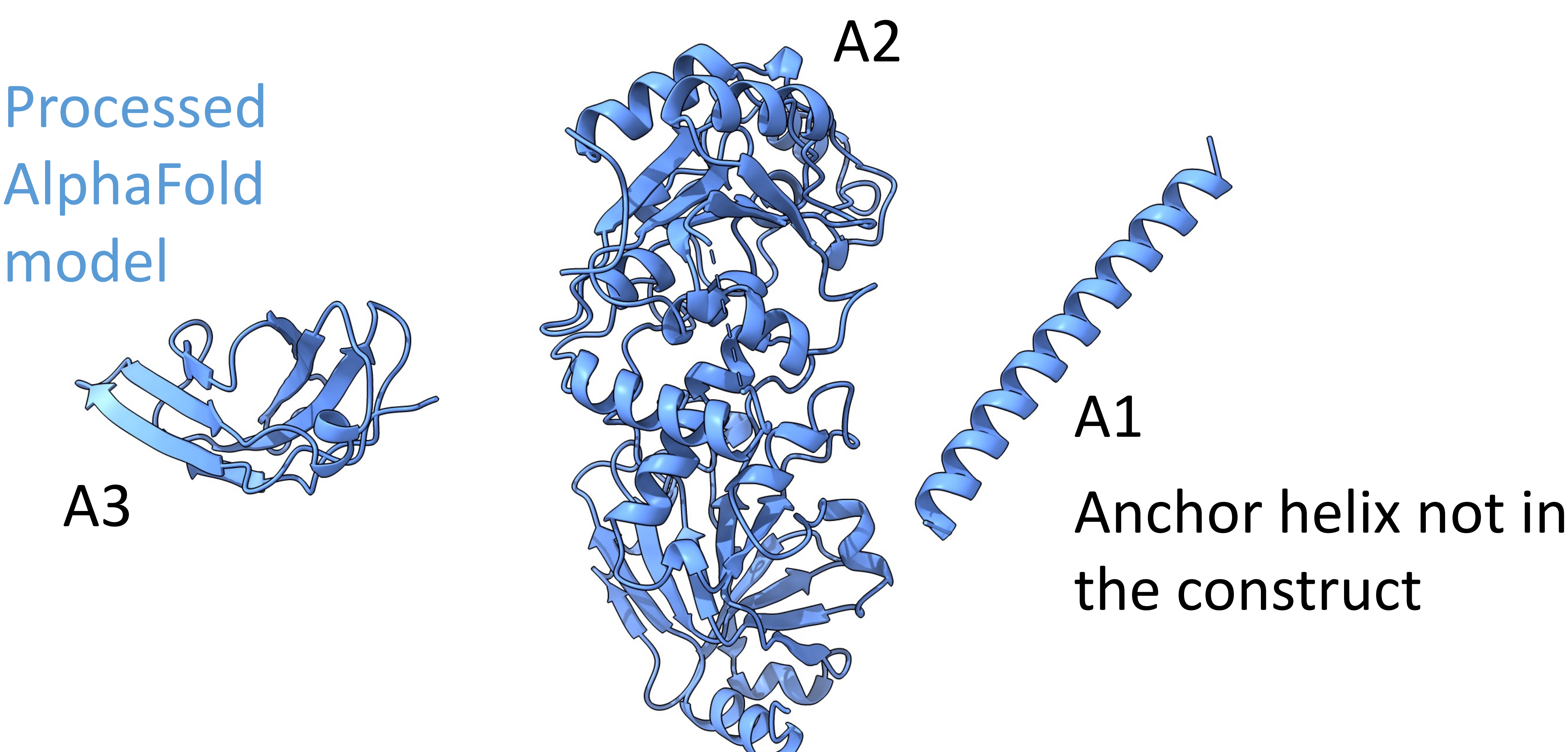
→ Remove low pLDDT regions, split into domains

AlphaFold model



MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2



The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Dorothee Liebschner, Nigel Moriarty, Billy Poon, Christopher Schlicksup, Oleg Sobolev

BERKELEY LAB

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

New Mexico CONSORTIUM

Los Alamos NATIONAL LABORATORY

UTHealth

Matt Baker

UTHealth The University of Texas Health Science Center at Houston

University of Cambridge

Randy Read, Airlie McCoy, Alisia Fadini

UNIVERSITY OF CAMBRIDGE

Duke University

Jane & David Richardson, Christopher Williams, Vincent Chen

An NIH/NIGMS funded Program Project

Liebschner D, *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Cryst. 2019 **D75**:861–877