cryo-EM Course
Laboratory for BioMolecular Structure (LBMS)
Friday, June 6th 2025

Predicted models for cryo-EM

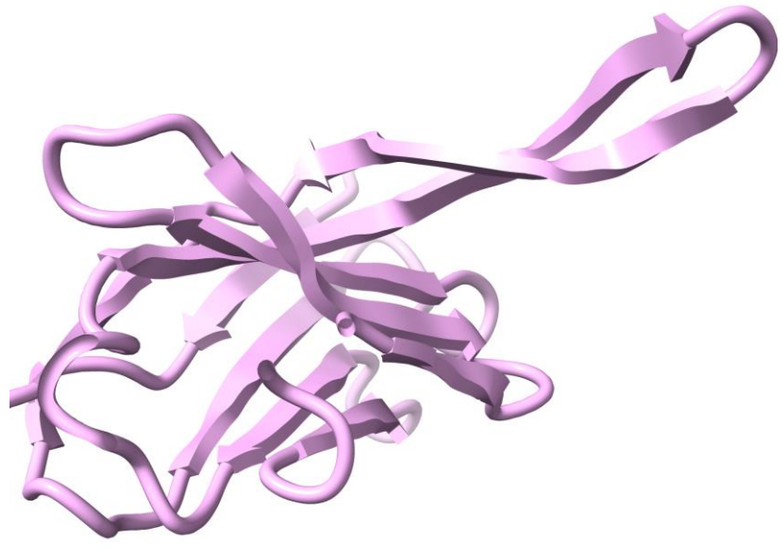
Dorothee Liebschner
Lawrence Berkeley Laboratory

Outline

- About AlphaFold predictions
- Using predicted models in cryo-EM
  1. Docking
  2. Model completion
  3. Reference model restraints
- Automated workflow:
  Iterating prediction and model building

Predicting models with AlphaFold


```
EVQLVESGGGLVQPGGSLRLSCAASGFNIYSSSIHWVRQAPGKGLEWVAYI
.............................F....M........Q........
...................K.........Y....L.......A.........
...............A.................................V......
....................A.....................................
.............................L....V....E.............
.........A.....................................Q.....
```



Sequence

Multiple sequence alignment

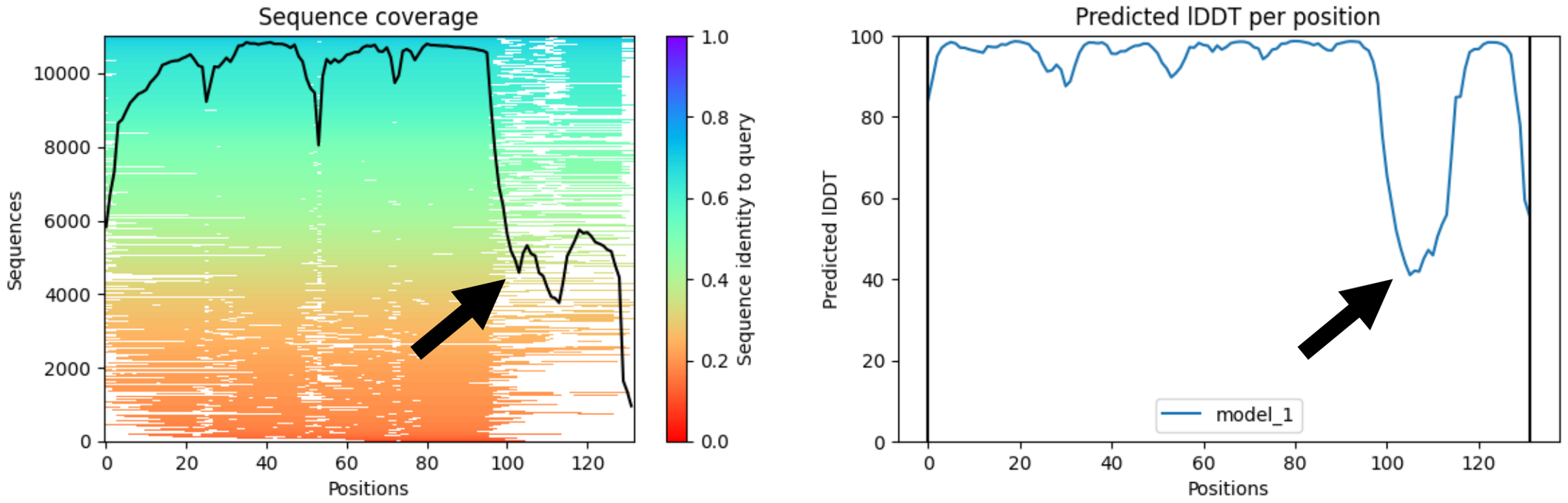
3D prediction

Predictions come with confidence measures

1. pLDDt (predicted Local Distance Difference Test)

- pLDDt identifies where errors are more likely.
- Per-residue confidence measure.
- Scales from 0 – 100 (pLDDt > 90: predicted with high accuracy).

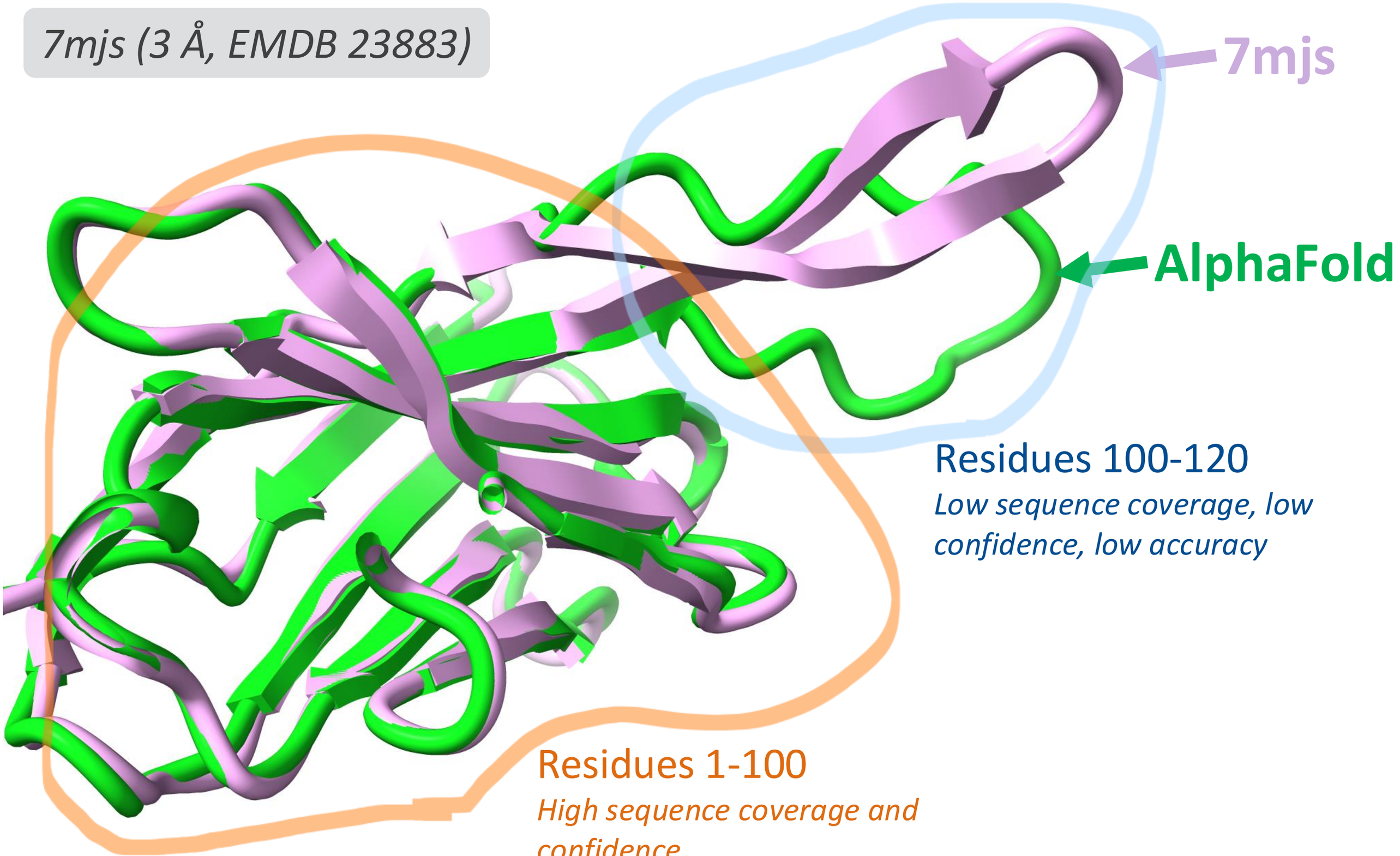
Sequence coverage → **Confidence**



Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319

Predictions come with confidence measures

7mjs (3 Å, EMDB 23883)



Residues 100-120

Low sequence coverage, low confidence, low accuracy

Residues 1-100

High sequence coverage and confidence

Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319

Predictions come with confidence measures

1. pLDDt (predicted Local Distance Difference Test)

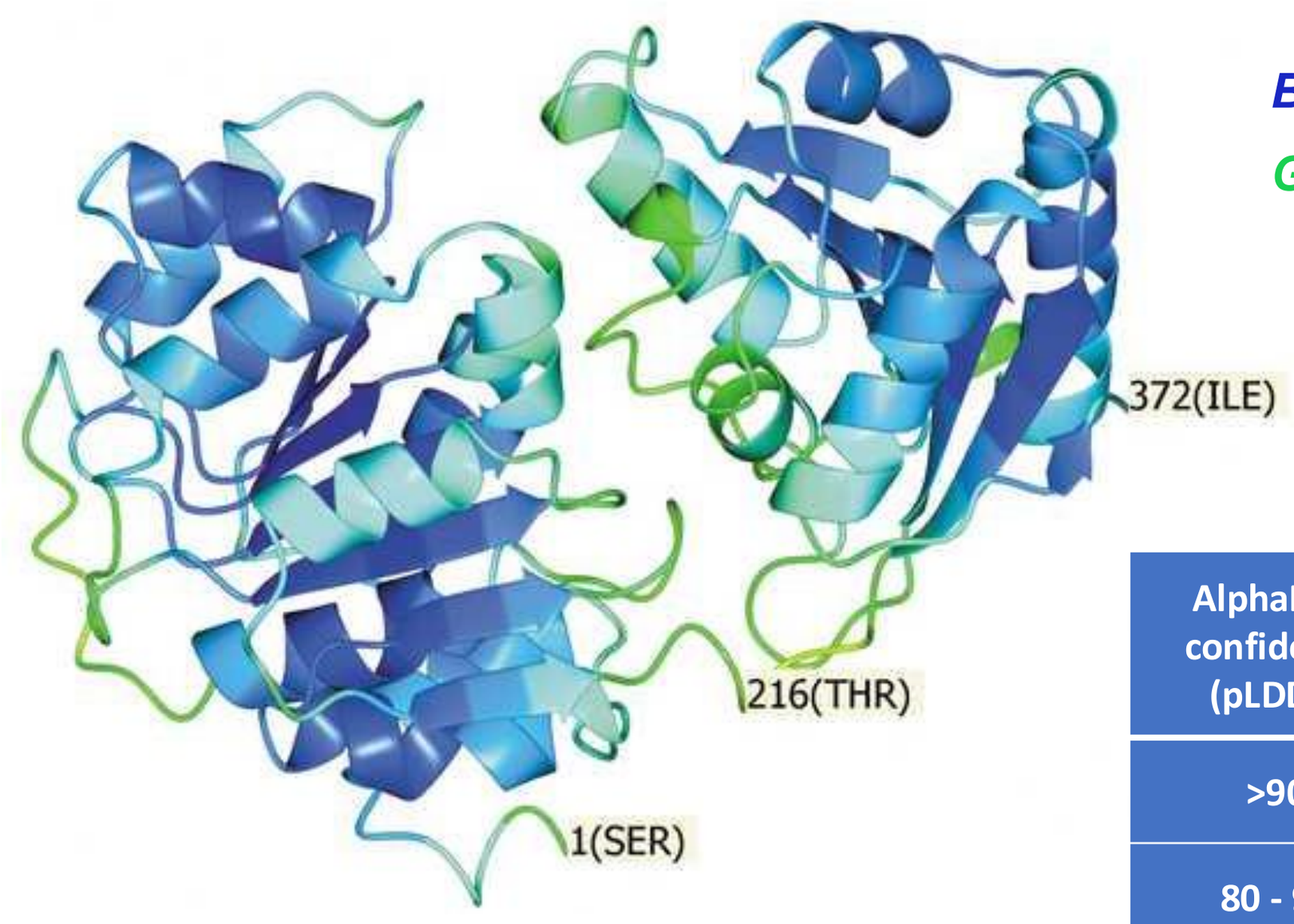
- pLDDt identifies where errors are more likely.
- Per-residue confidence measure.
- Scales from 0 – 100 (pLDDt > 90: predicted with high accuracy).

AlphaFold confidence (pLDDT)	Median prediction error (Å)	Percentage with error over 2 Å
>90	0.6	10
80 - 90	1.1	22
70 - 80	1.5	33
<70	3.5	77

Terwilliger et al. (2023), AlphaFold predictions are valuable hypotheses, and accelerate but do not replace experimental structure determination. Nature Methods 2023: https://doi.org/10.1038/s41592-023-02087-4

Predictions come with confidence measures

1. pLDDt (predicted Local Distance Difference Test)



AlphaFold prediction for RNA helicase
(PDB entry 6L5L)

Blue: pLDDt > 90

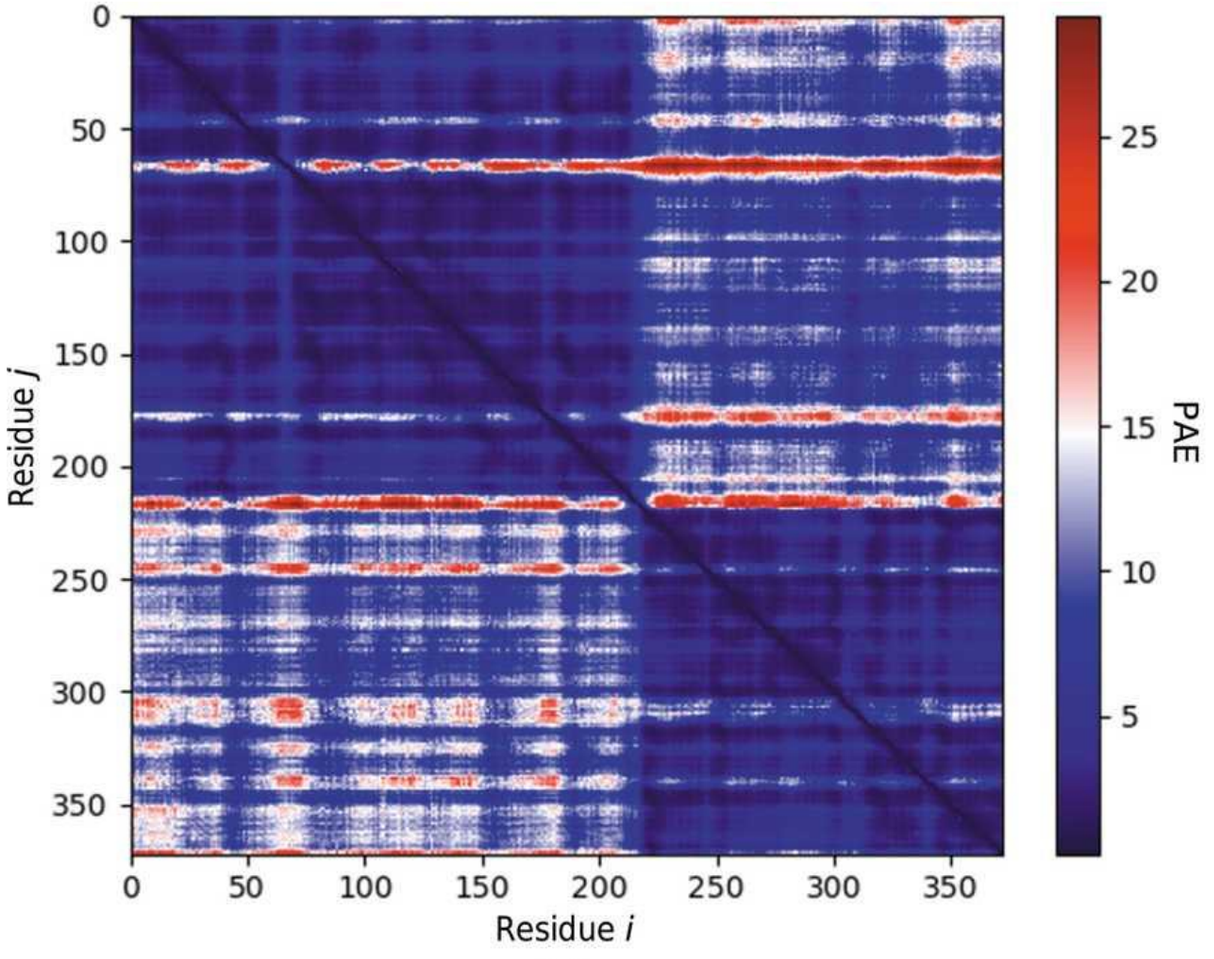
Green: pLDDt 80 - 90

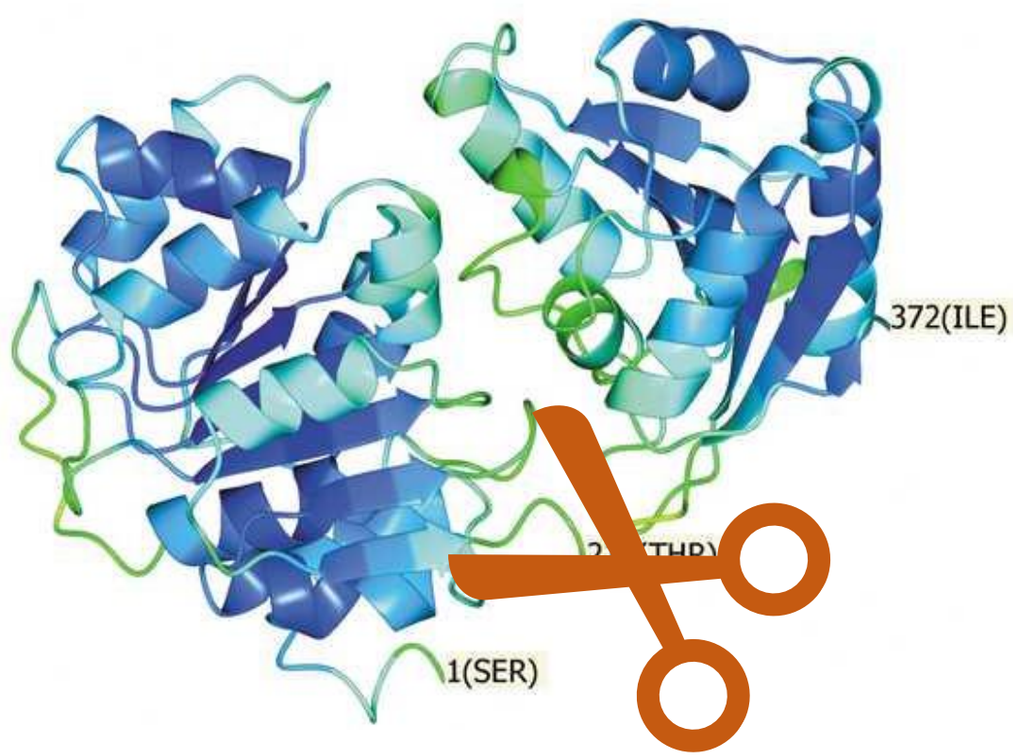
AlphaFold confidence (pLDDT)	Median prediction error (Å)	Percentage with error over 2 Å
>90	0.6	10
80 - 90	1.1	22
70 - 80	1.5	33
<70	3.5	77

Predictions come with confidence measures

2. Predicted aligned error (PAE)

- Certainty of relative positions between two residues.
- Identifies accurately-predicted domains.
- Dark blue: uncertainty in relative positions < 5 Å.





- Suggests 2 domains

Using predicted models: B-factors

Color by pLDDT

Color by B-factor

high pLDDT (high confidence)

low pLDDT (low confidence, uncertain)

high B-factor (disordered, uncertain)

low B-factor (ordered)

Oeffner RD, Croll TI, Millán C , Poon BK, Schlicksup CJ, Read RJ, Terwilliger TC. Acta Cryst. D, 2022 (78):1303-1314; https://doi.org/10.1107/S2059798322010026

Using predicted models: B-factors

high pLDDT (high confidence)

low pLDDT (low confidence, uncertain)

high B-factor (disordered, uncertain)

low B-factor (ordered)

B-factor may be used in downstream calculations, e.g. to calculate weights for docking. Residues with high B-factors are downweighed.

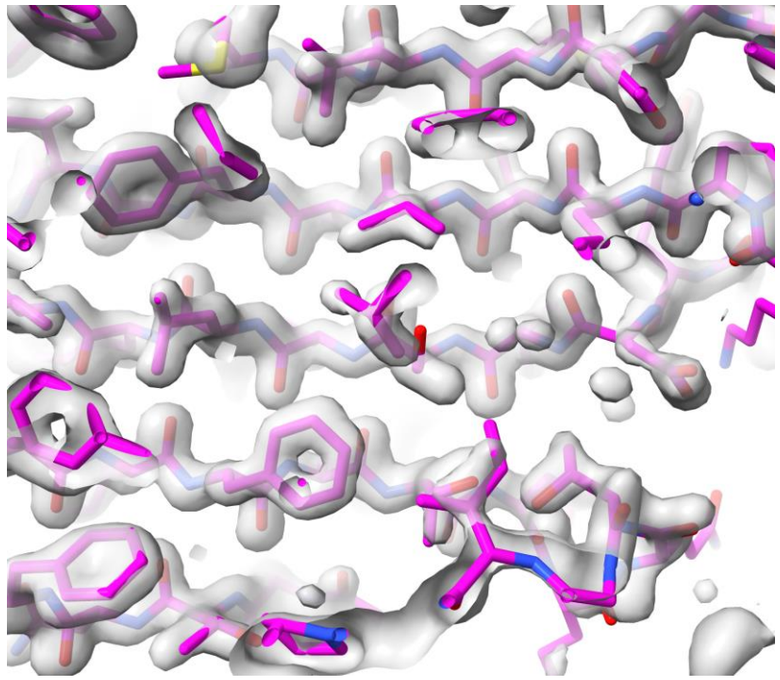
→ Convert pLDDT to pseudo B-factors.

$$\Delta = 1.5 \exp[4(0.7 - \text{pLDDT})] \qquad B = \frac{8\pi^2\Delta^2}{3}$$

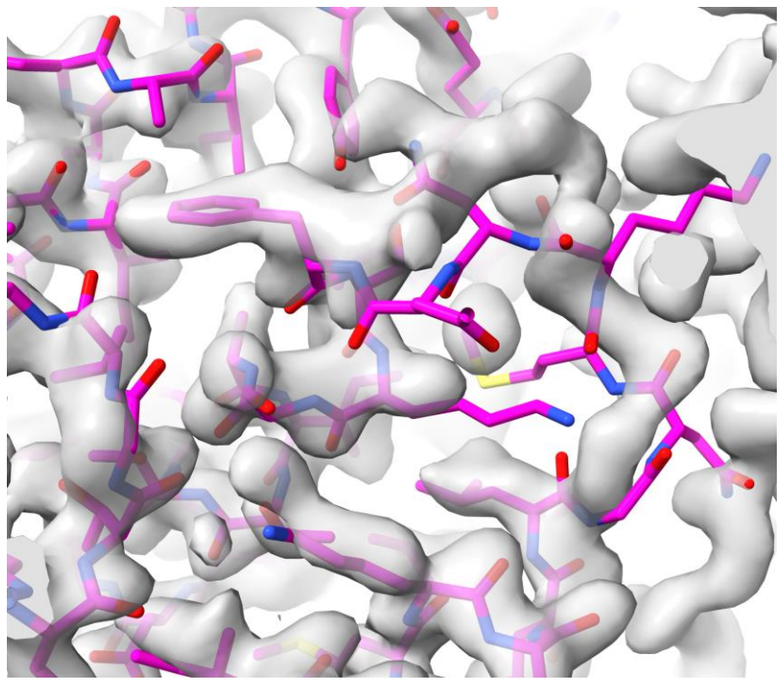
Oeffner RD, Croll TI, Millán C , Poon BK, Schlicksup CJ, Read RJ, Terwilliger TC. Acta Cryst. D, 2022 (78):1303-1314; https://doi.org/10.1107/S2059798322010026

AlphaFold predictions are great hypotheses

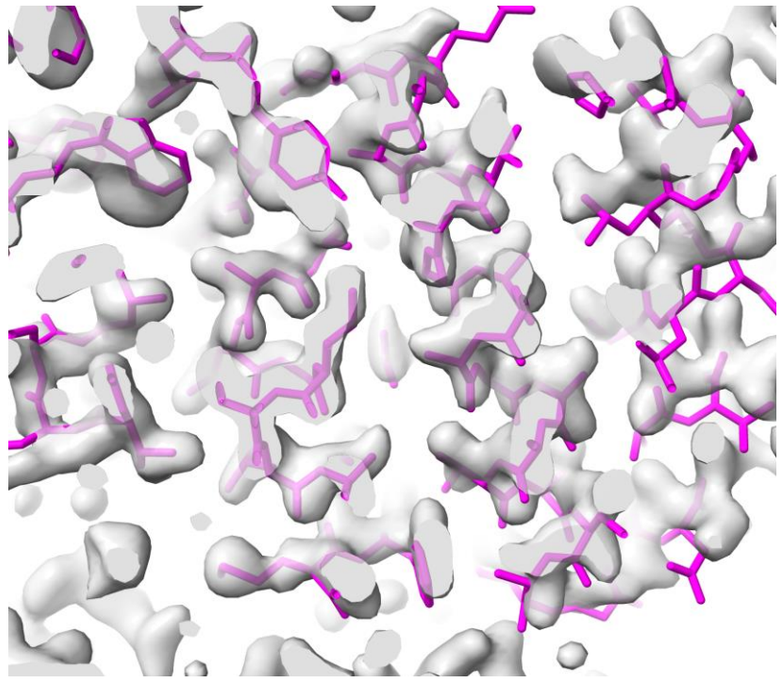
AlphaFold models can be….



Awesome



Wrong

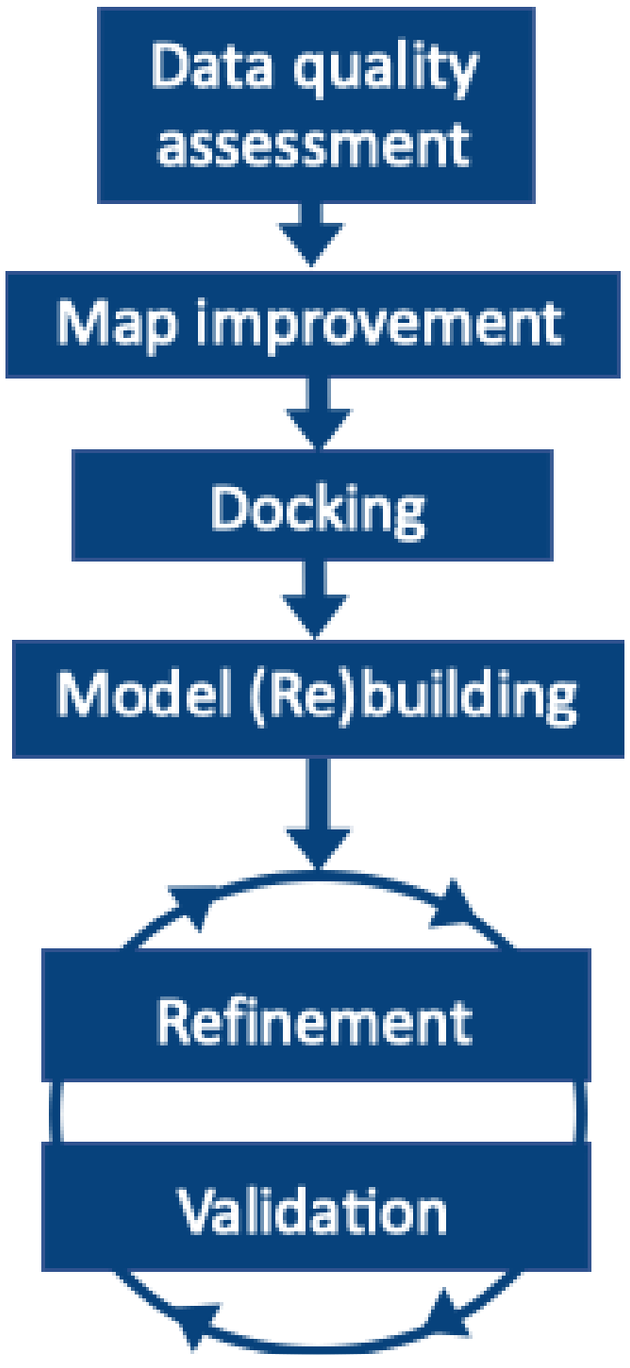


Distorted

Terwilliger et al. (2023), AlphaFold predictions are valuable hypotheses, and accelerate but do not replace experimental structure determination. Nature Methods 2023: https://doi.org/10.1038/s41592-023-02087-4

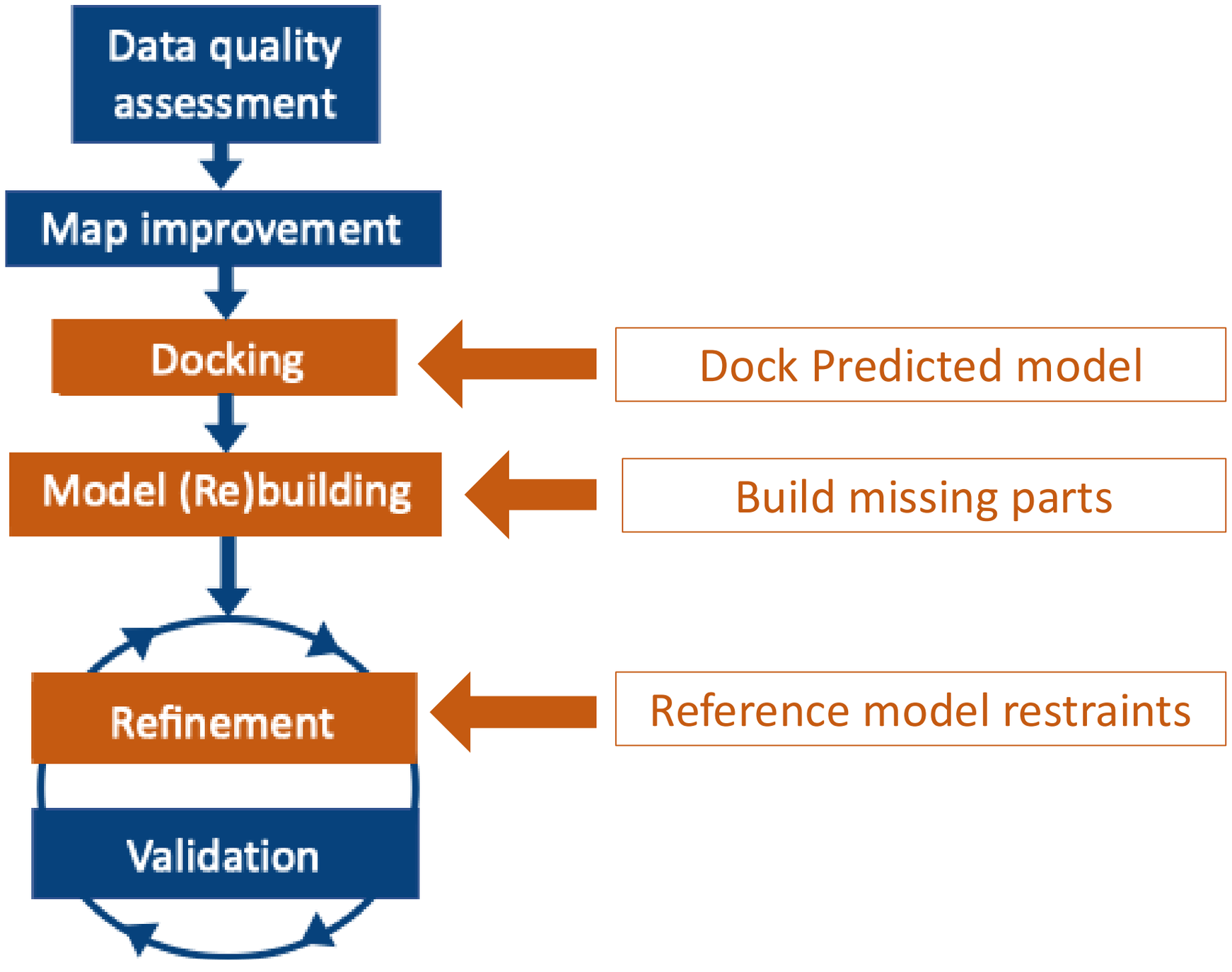
How to use predictions?

Incorporate predictions into the typical cryo-EM workflow.

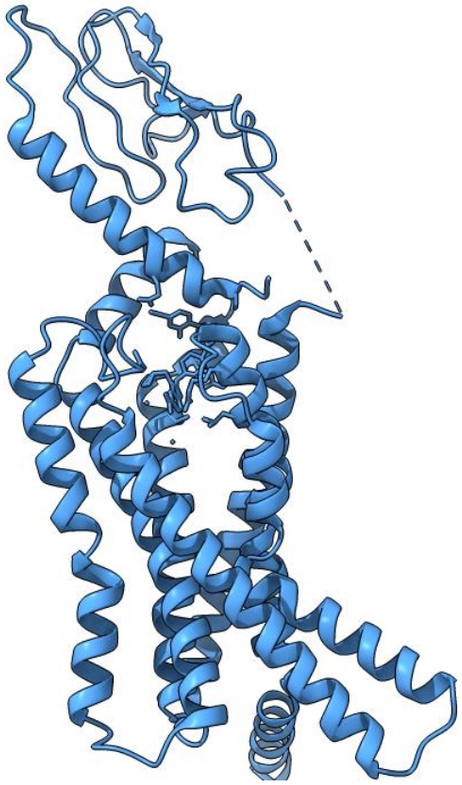


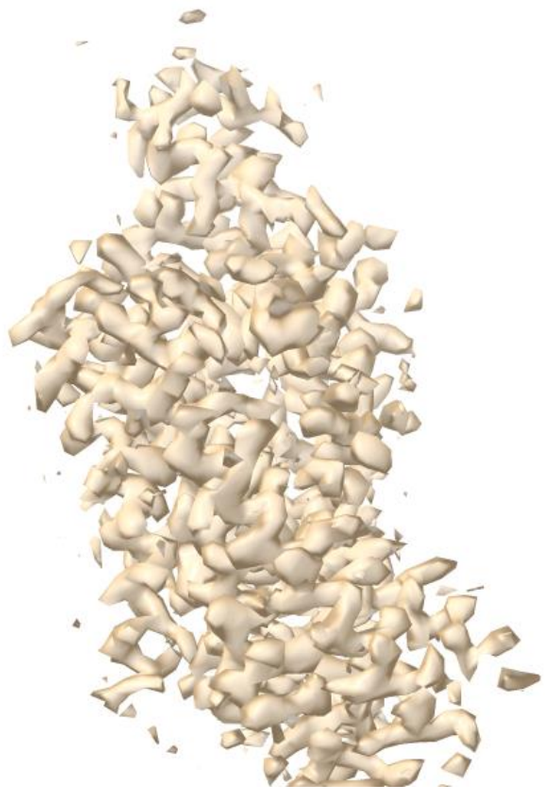
How to use predictions?

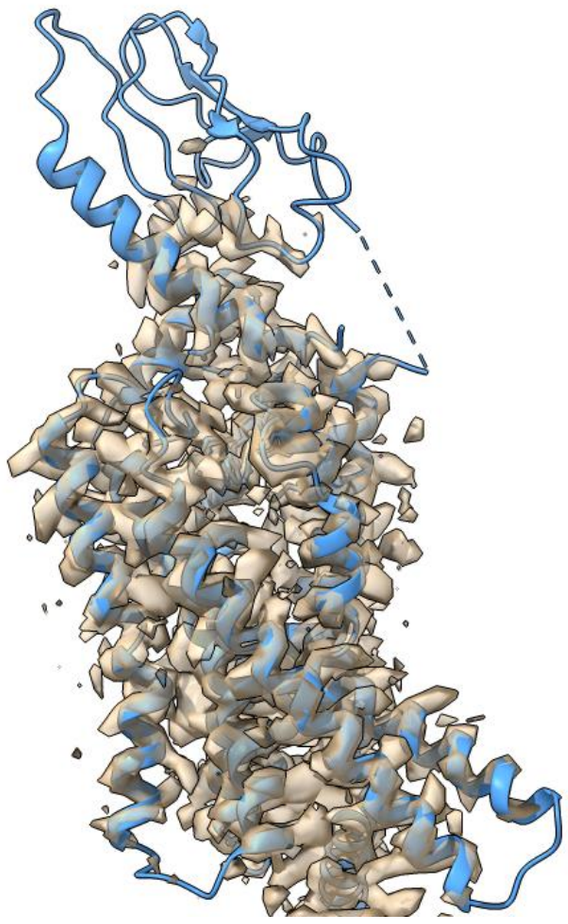
Incorporate predictions into the typical cryo-EM workflow.



1. Use a predicted model for cryo-EM docking



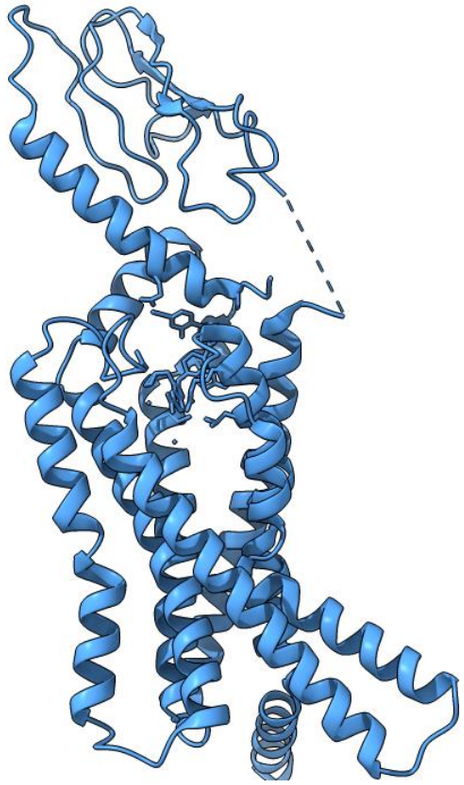




cryo-EM docking

Cryo-EM maps typically lack the necessary resolution and quality for *ab initio* model building.

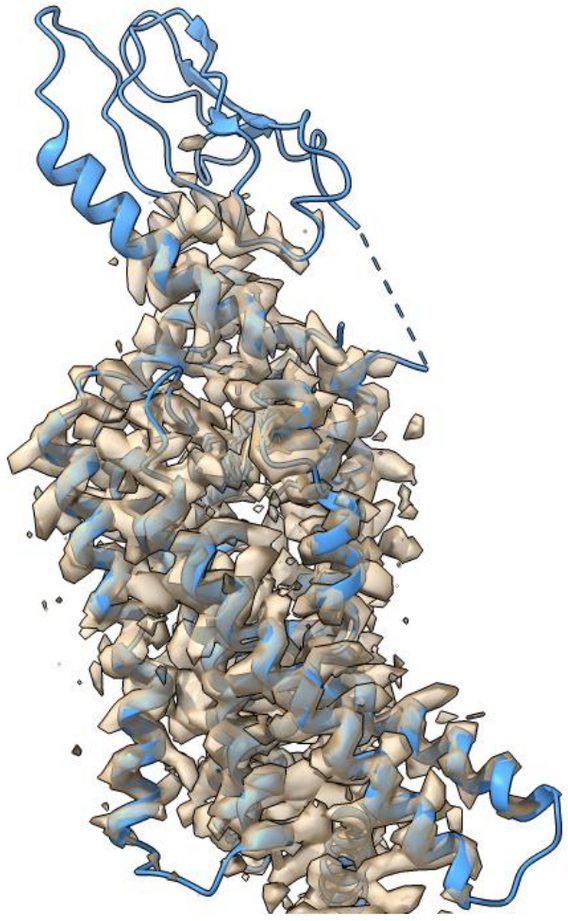
→ dock a pre-existing model into the map.



Model assumed to look like the sample.

Cryo-EM map

Find translation & orientation



Docked model

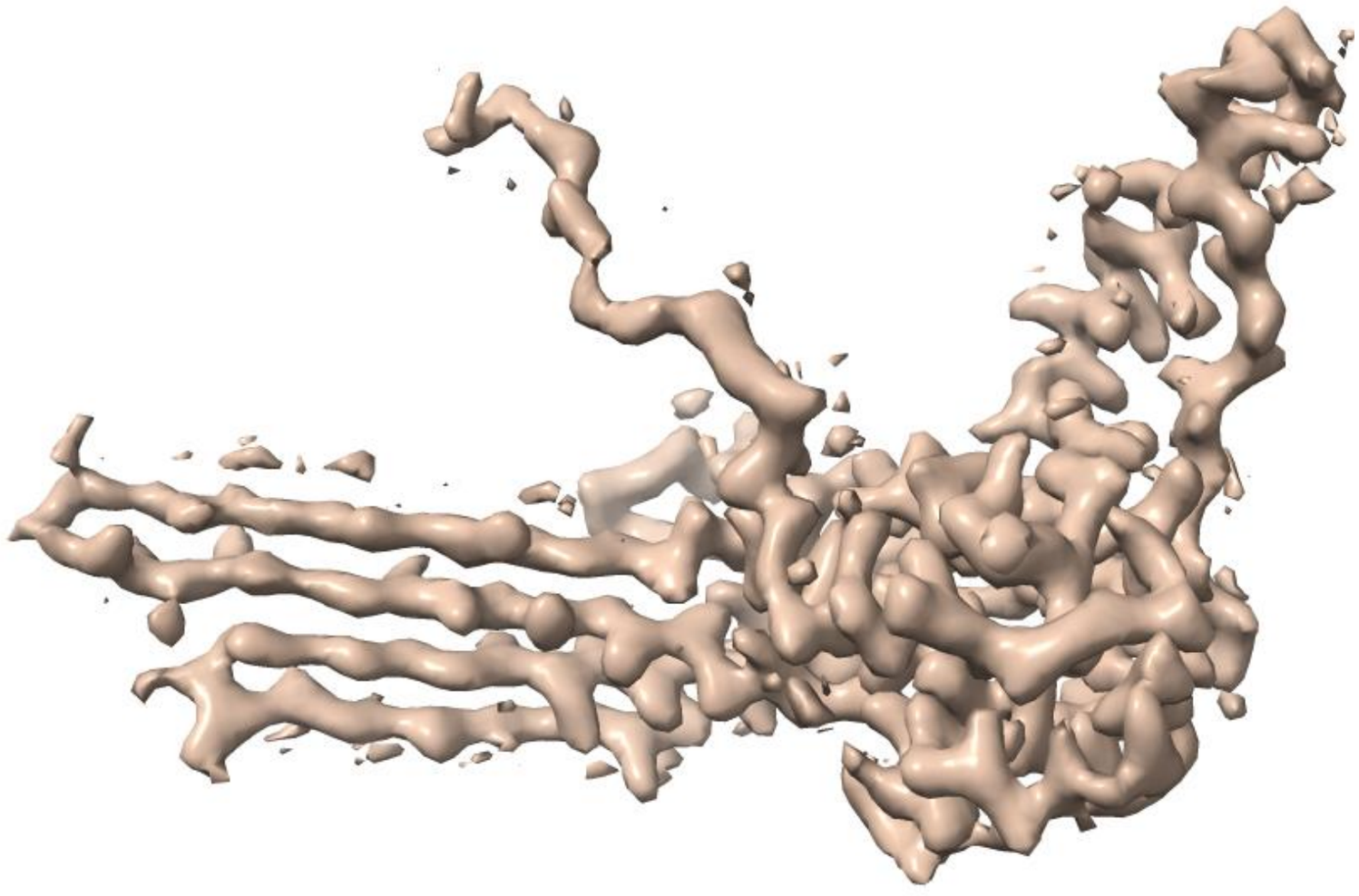
Use a predicted model for cryo-EM docking

Example:

Cryo-EM map (30160 – 7brm)
3.6 Å

```
>chain ' A'
XXXXXXXXXXXXXXXCLTAPPKEAARPTLMPRAQSYKDLTHLPAP
TGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDS
RWFIPLERQGLQNLLNERKIIRAAQENGTVAINNRIPLQSLTAAN
IMVEGSIIGYESNVKSGGVGARYFGIGADTQYQLDQIAVNLRVVN
VSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNE
PVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRH
MS
```

sequence

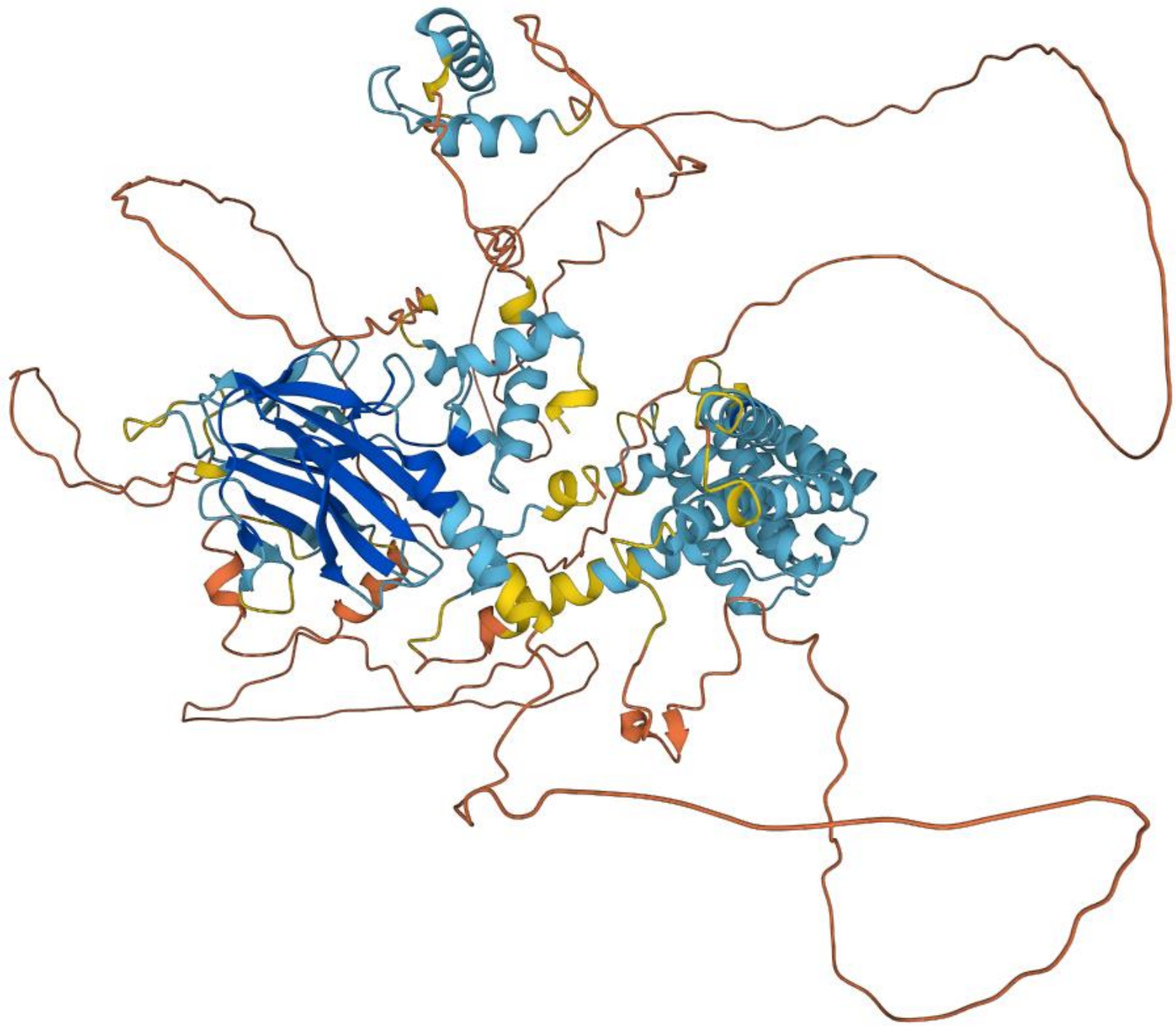


cryo-EM map

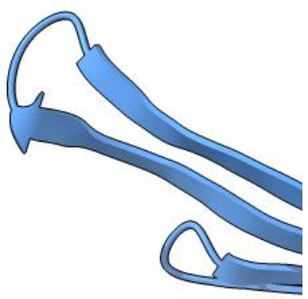
Get a prediction

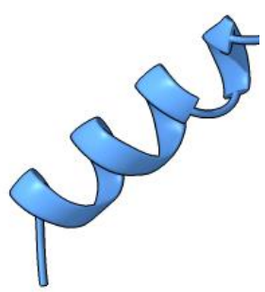
sequence

```
>chain ' A'
XXXXXXXXXXXXXXXCLTAPPKEAARPTLMPRAQSYKDLTHLPAP
TGKIF
RWFIP
IMVEG
VSTGE
PVMLC
MS
```



AlphaFold model



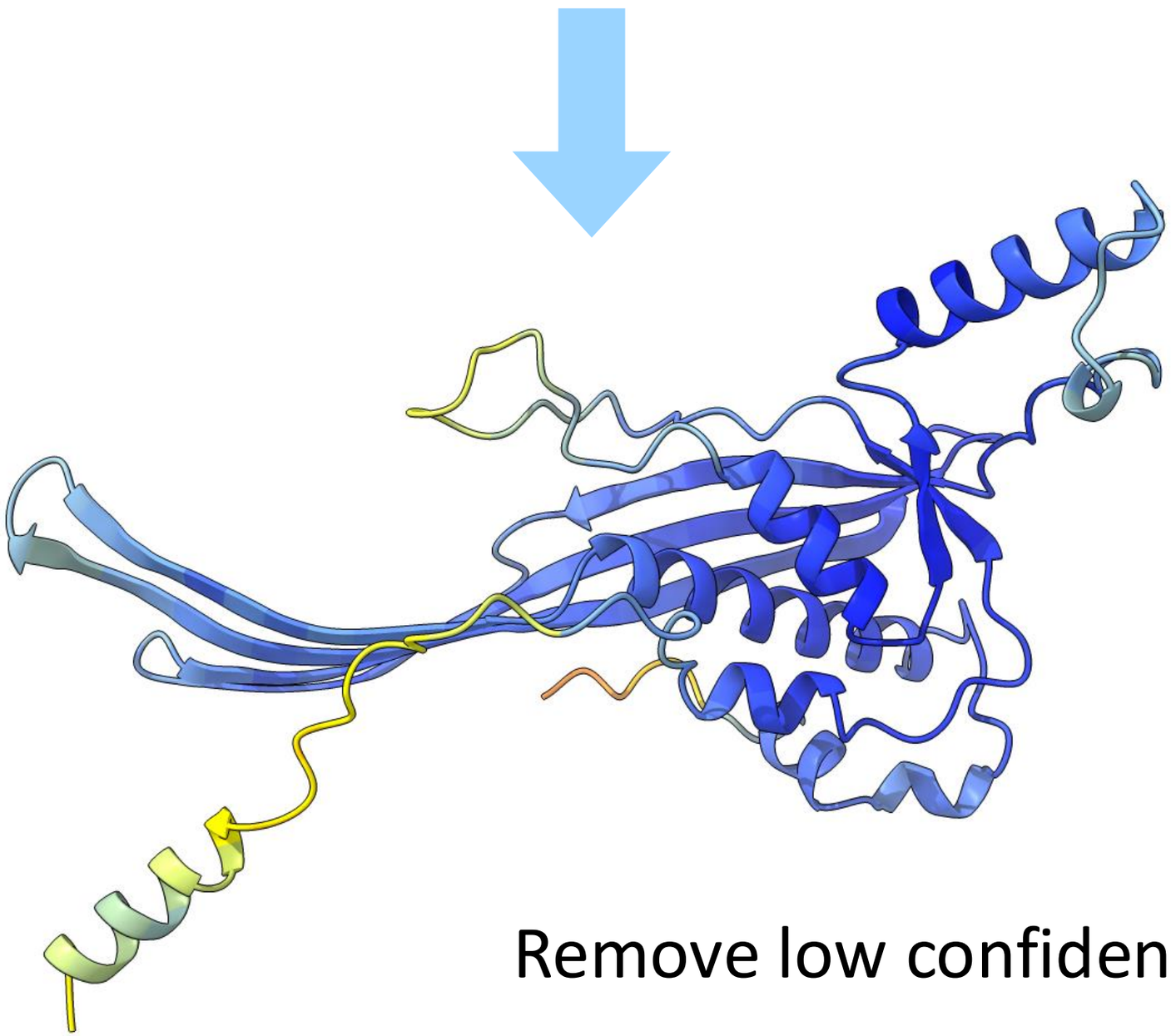


Process prediction

sequence

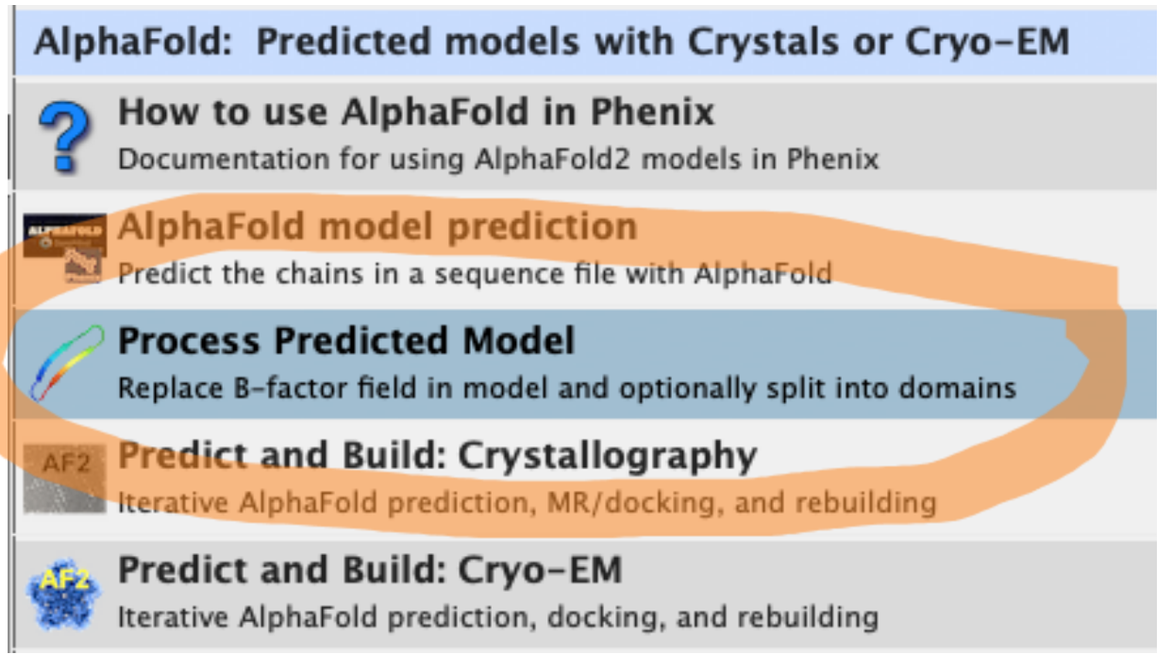
```
>chain ' A'
XXXXXXXXXXXXXXXXCLTAPPKEAARPTLMPRAQSYKDLTHLPAP
TGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDS
RWFIPLERQGLQNLLNERKIIRAAQENGTVAINNRIPLQSLTAAN
IMVEGSIIGYESNVKSGGVGARYFGIGADTQYQLDQIAVNLRVVN
VSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNE
PVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRH
MS
```

AlphaFold model



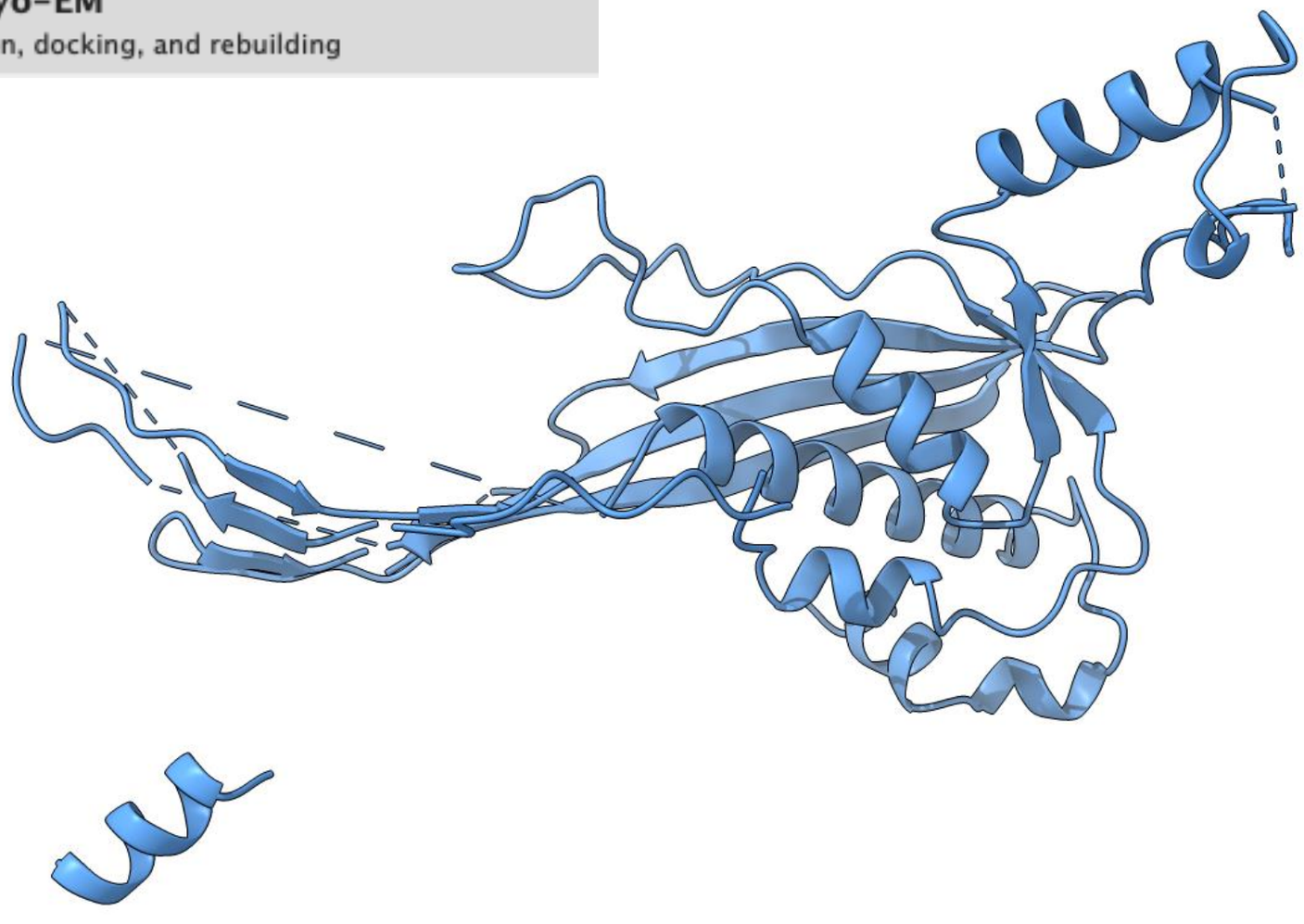
Remove low confidence parts

Process prediction



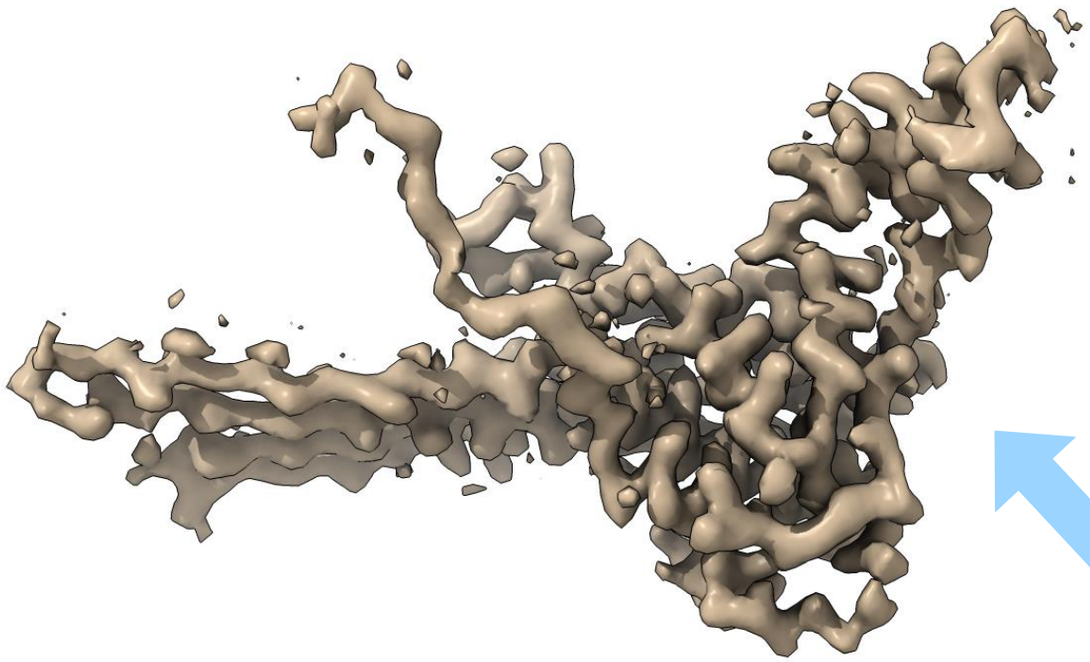
phenix.process_predicted_model

Processed AlphaFold model

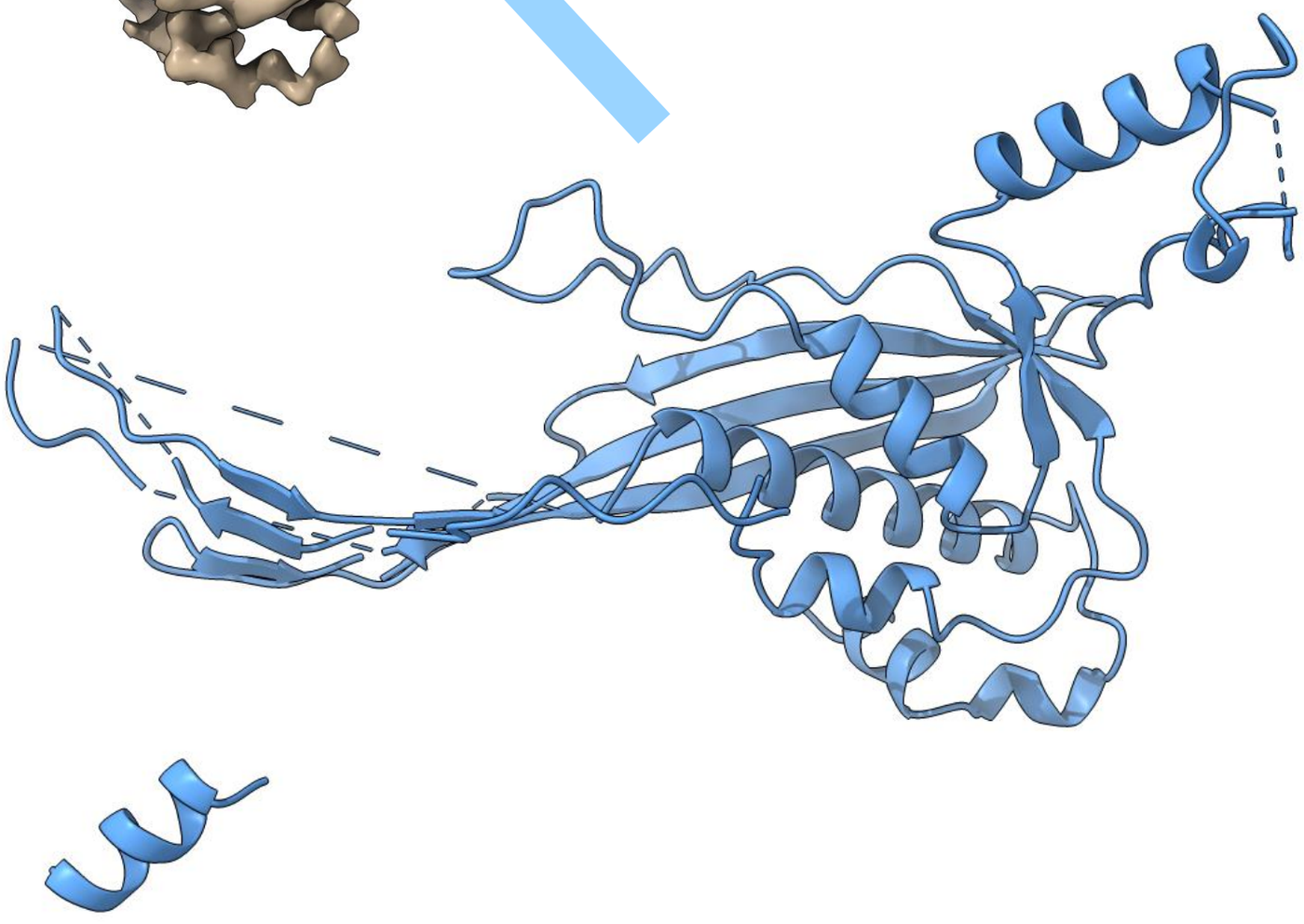


Dock processed model

cryo-EM map

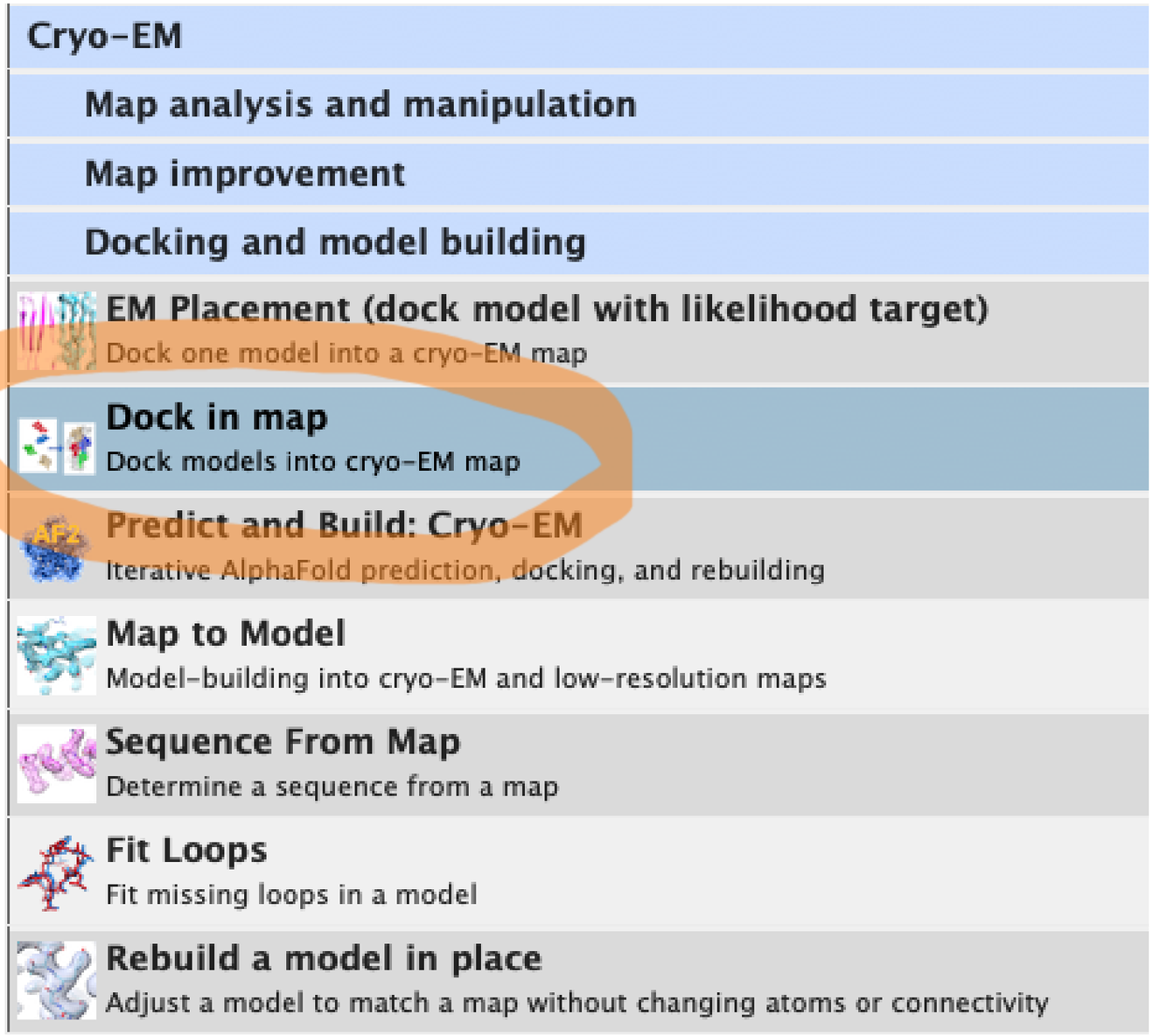


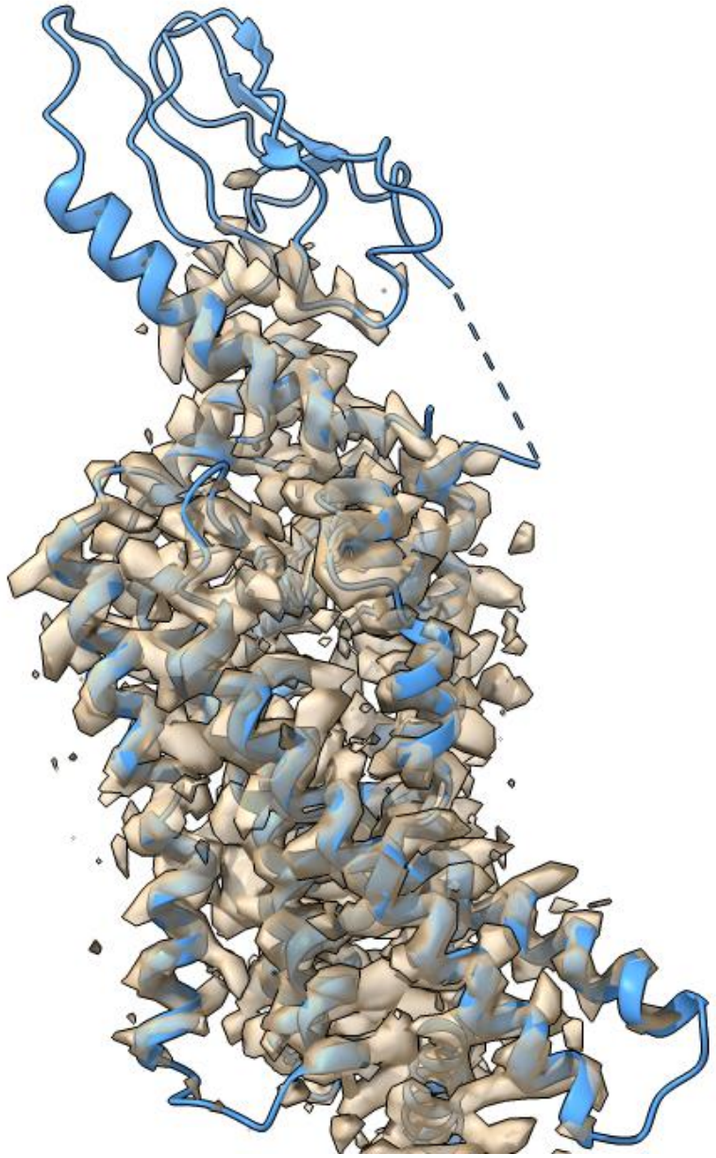
Processed AlphaFold model



Docking in Phenix

- Dock-in-map (T. Terwilliger) – phenix.dock_in_map



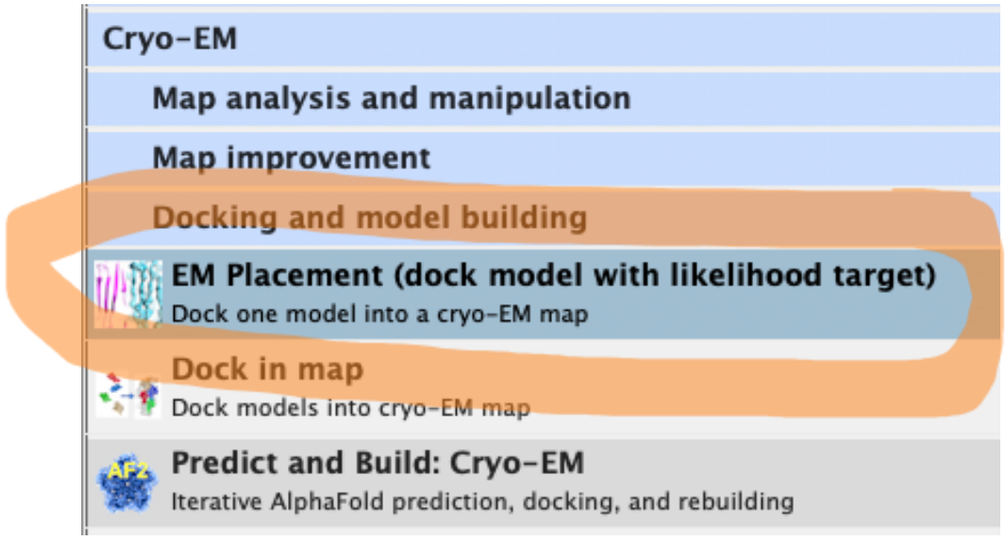


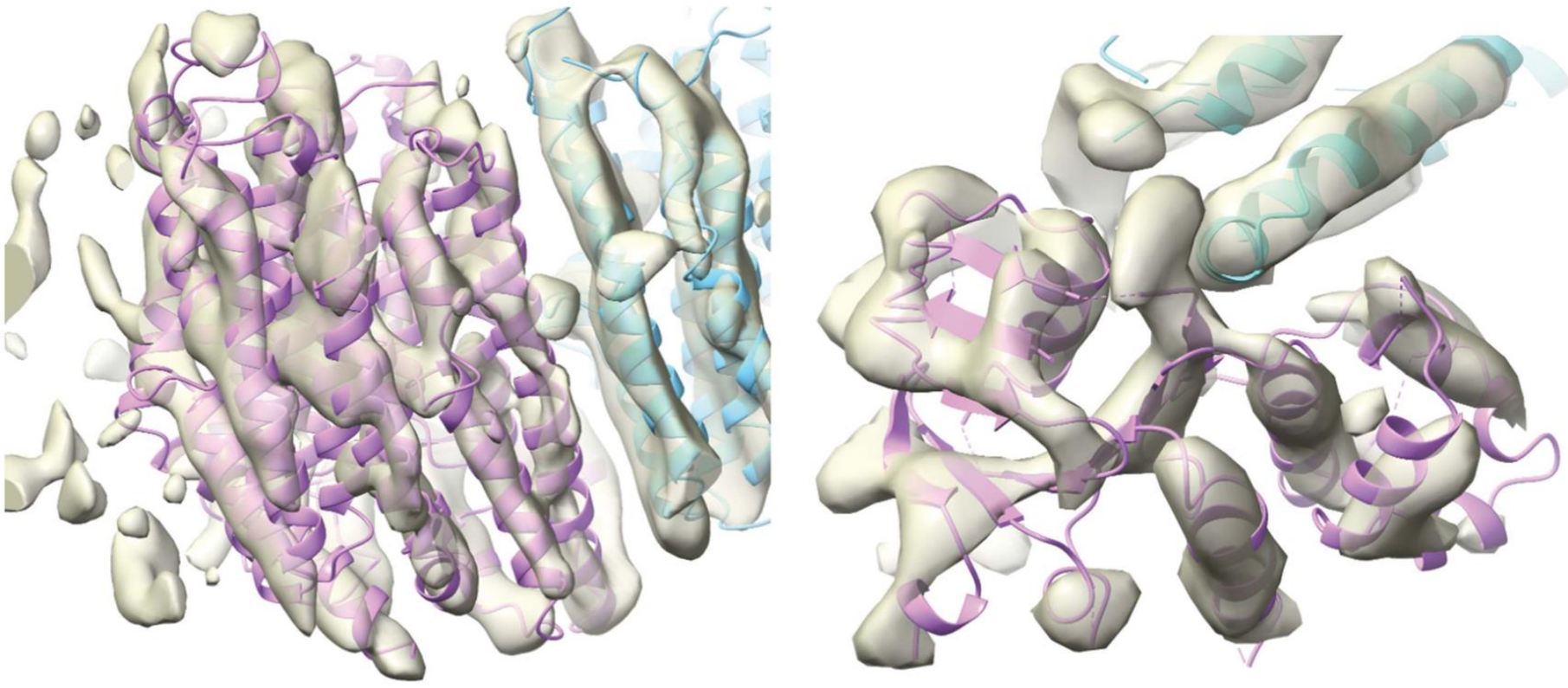
Docking in Phenix

Likelihood-based EM docking:

- Use likelihood scores to dock a model into a map
- Works at low resolution (8.5 Å)

$$\mathrm{LLG}(\mathbf{E}_{\mathrm{mean}};\mathbf{E}_{\mathrm{C}}) = \frac{2}{1-D_{\mathrm{obs}}^2\sigma_{\mathrm{A}}^2} D_{\mathrm{obs}}\sigma_{\mathrm{A}}E_{\mathrm{mean}}E_{\mathrm{C}}\cos(\Delta\varphi) - \frac{D_{\mathrm{obs}}^2\sigma_{\mathrm{A}}^2(E_{\mathrm{mean}}^2+E_{\mathrm{C}}^2)}{1-D_{\mathrm{obs}}^2\sigma_{\mathrm{A}}^2} - \ln(1-D_{\mathrm{obs}}^2\sigma_{\mathrm{A}}^2).$$

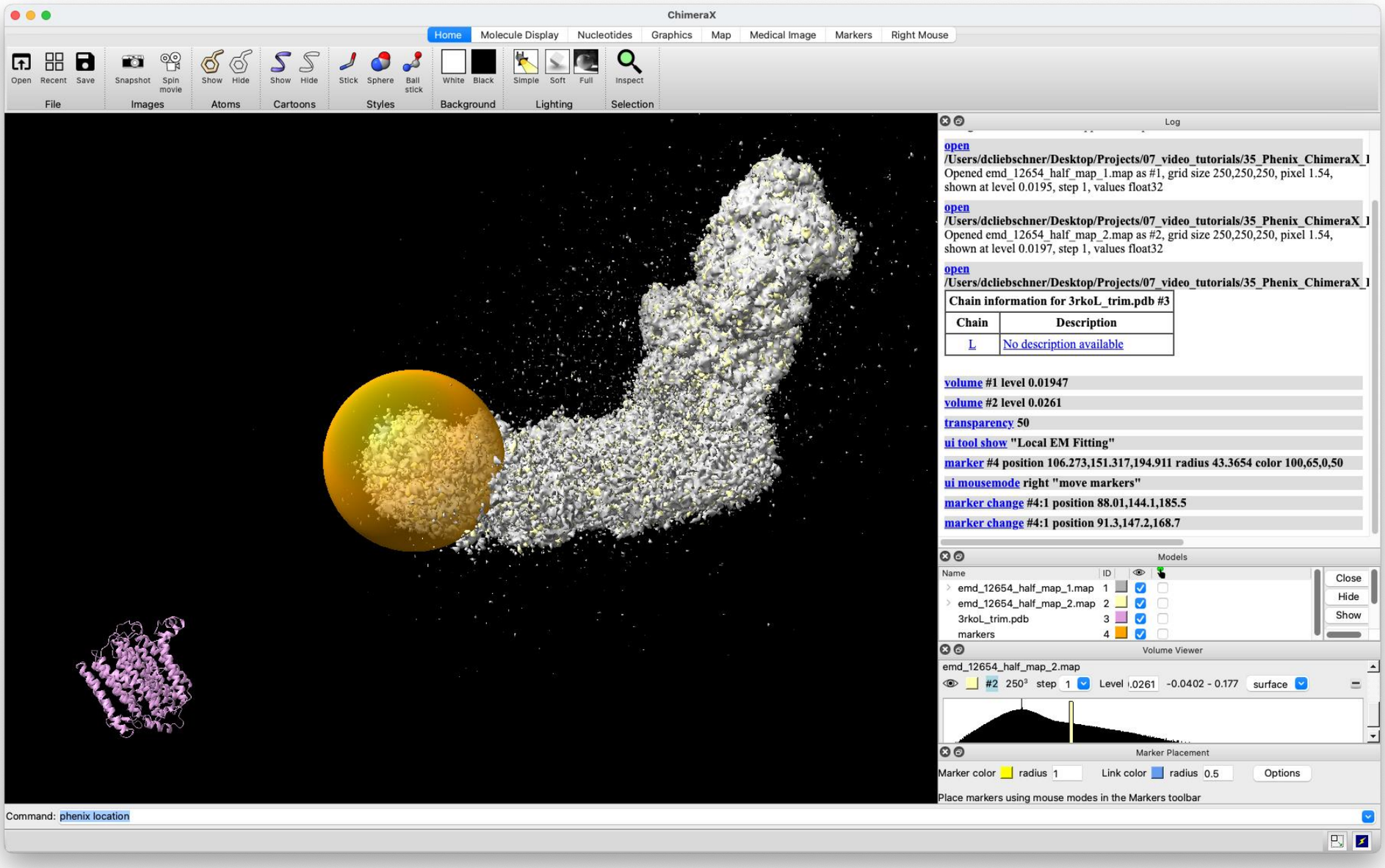




Read RJ, Millán C, McCoy AJ, Terwilliger TC. Likelihood-based signal and noise analysis for docking of models into cryo-EM maps. Acta Cryst. D 2023 271–80.
Millán C, McCoy AJ, Terwilliger TC, Read RJ. Likelihood-based docking of models into cryo-EM maps. Acta Cryst. D 2023 Apr 1;79(Pt 4):281–9.

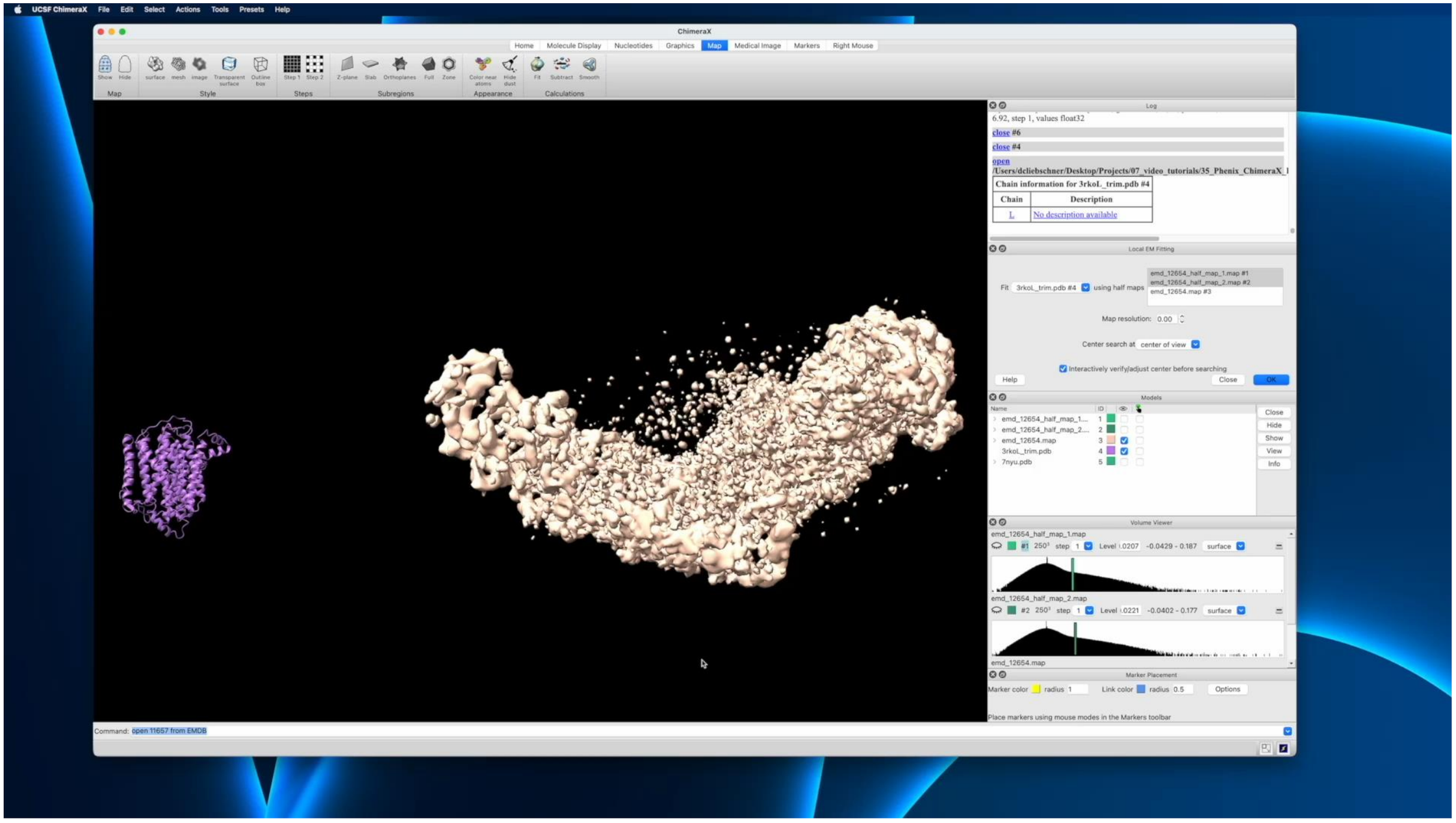
Docking in Phenix/ChimeraX

- Likelihood-based docking can be done via ChimeraX.
- Can select the region into which the model should be docked.

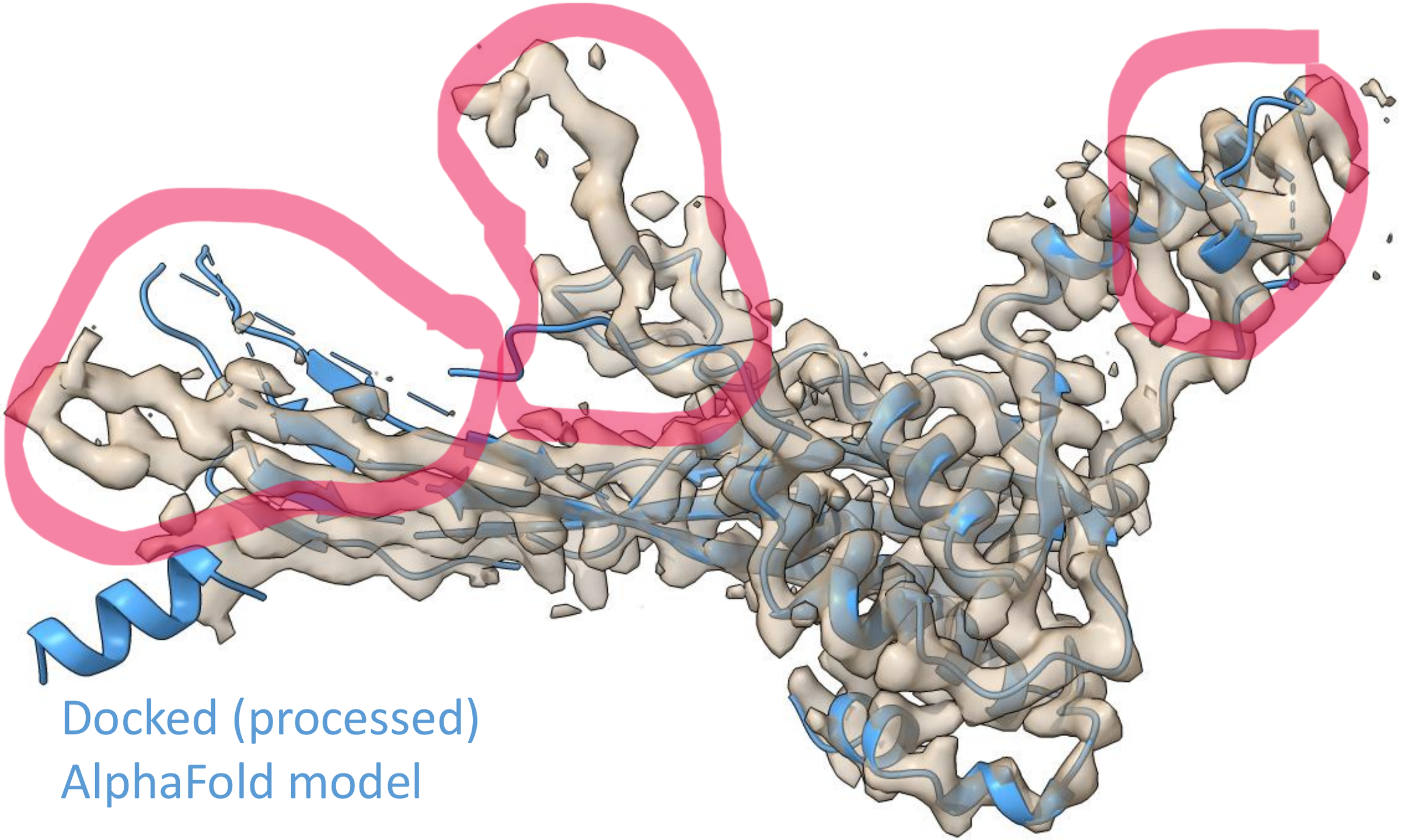


Docking in Phenix/ChimeraX

- Likelihood-based docking can be done via ChimeraX.
- Can select the region into which the model should be docked.



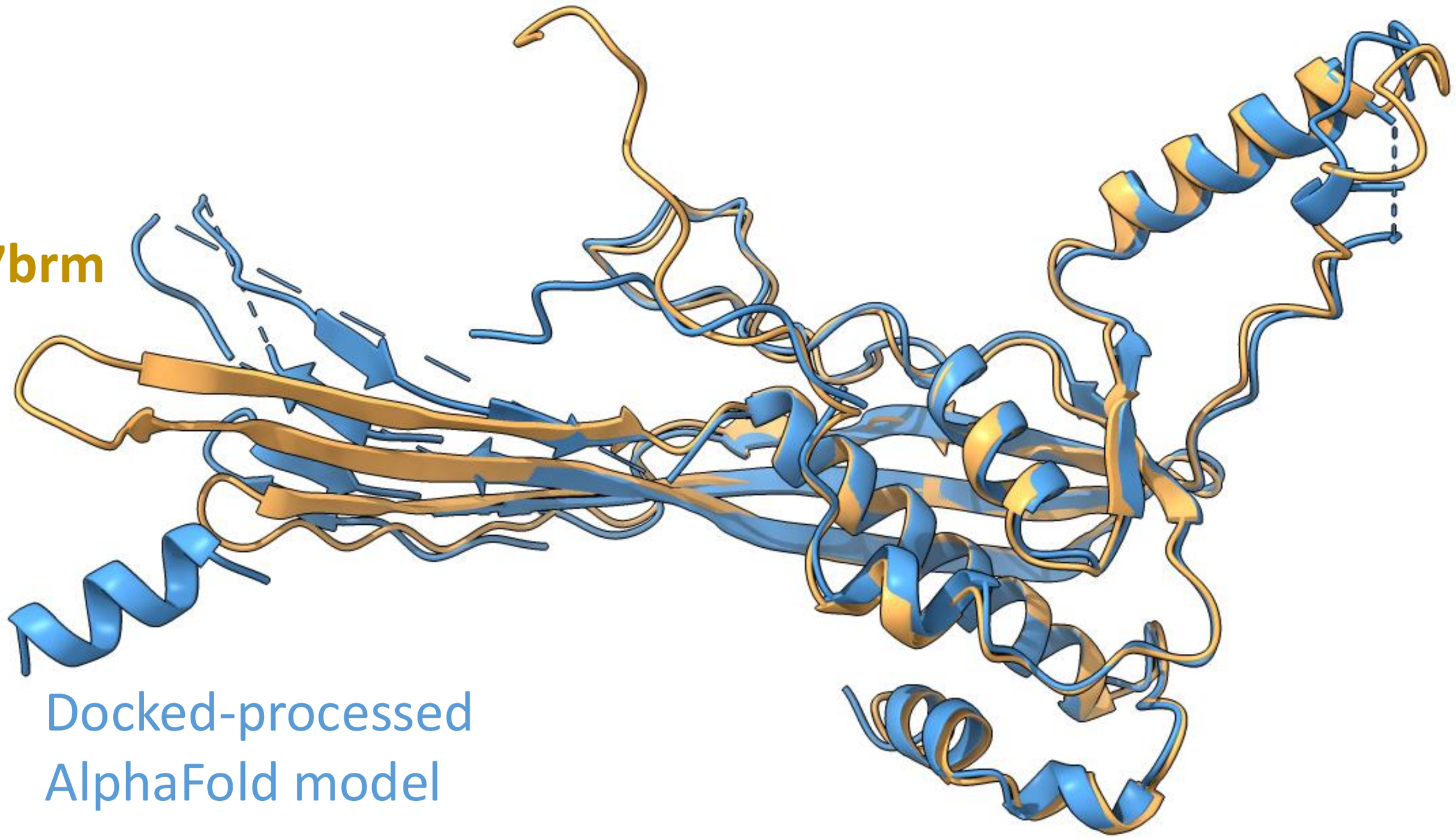
Dock processed model



Docked (processed)
AlphaFold model

Some parts don’t fit into the map

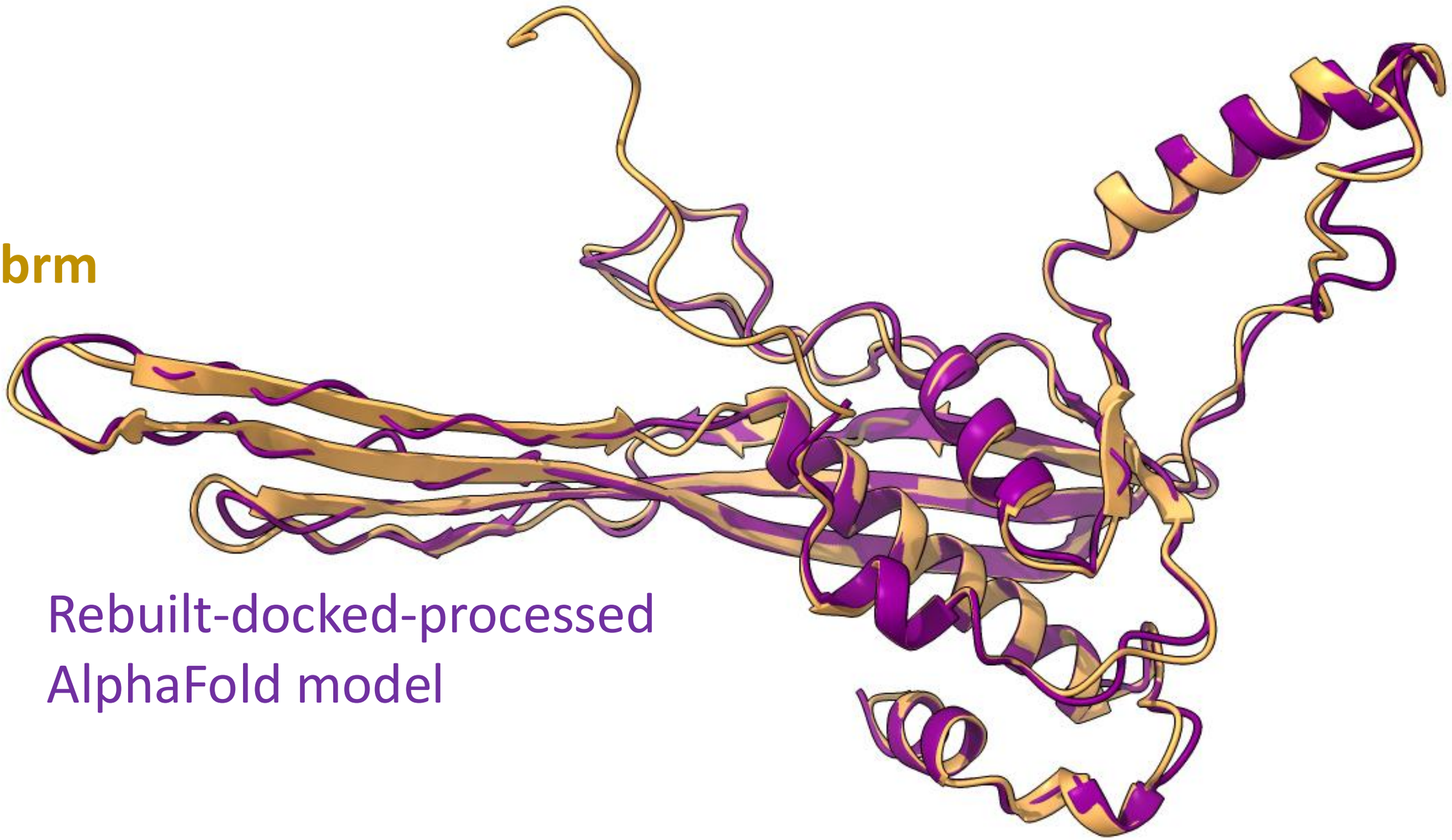
Dock processed model



Some parts don’t fit into the map → fit loops and rebuild

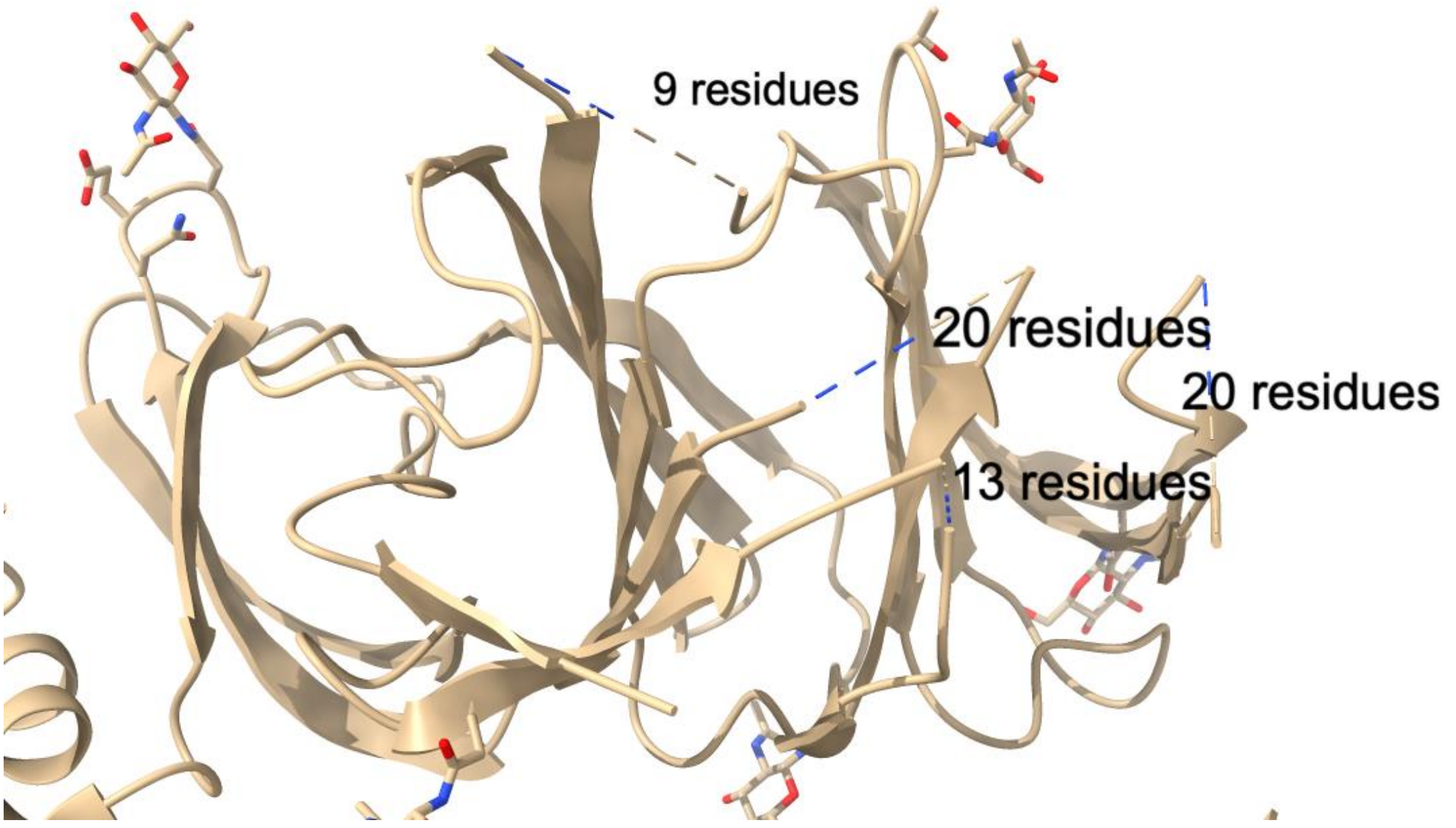
Fit loops and rebuild

phenix.fit_loops



Next step: real space refinement

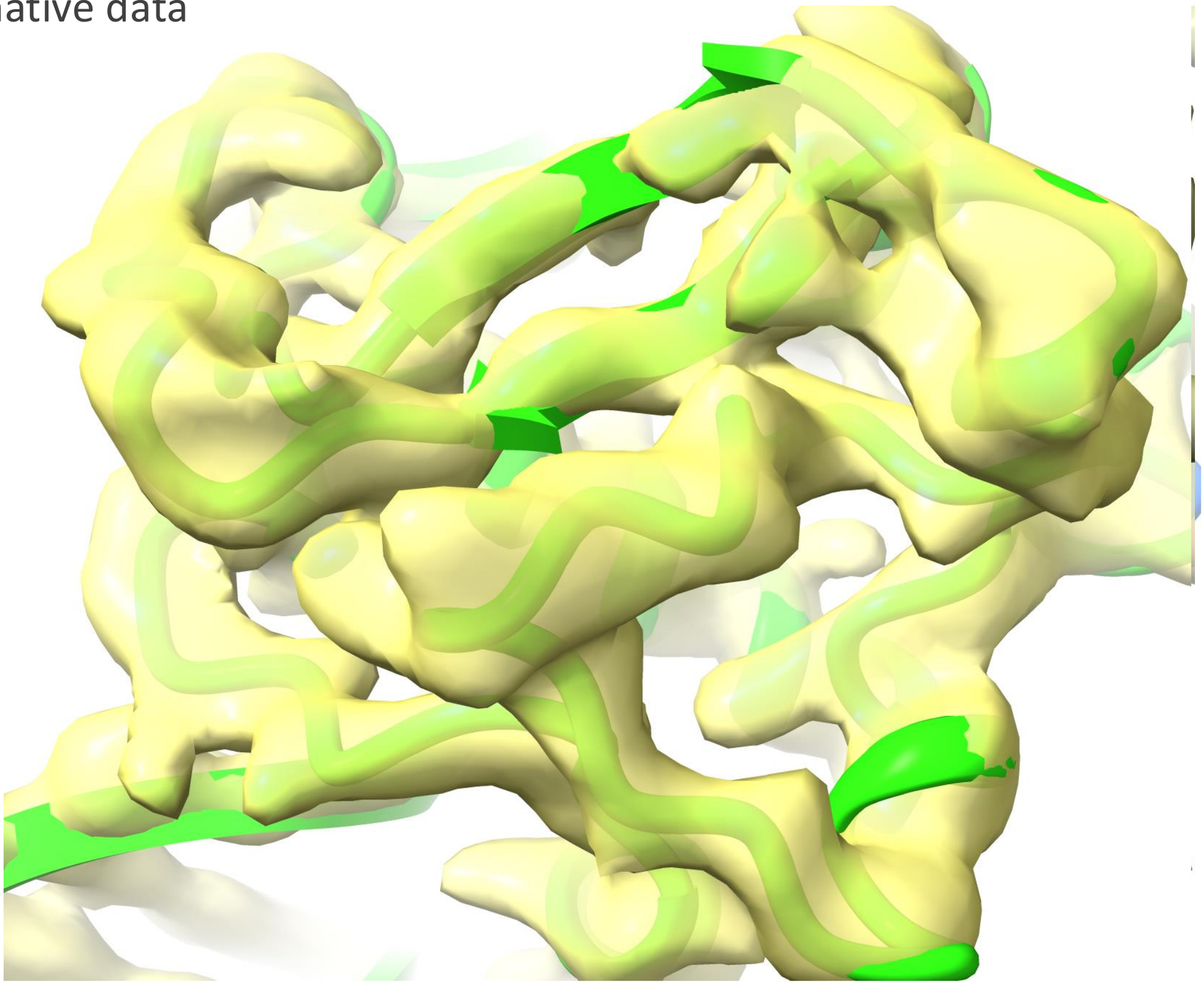
2. Use a predicted model to complete your structure



Use predicted model as hypothesis for missing parts.

Can AF predictions help if the structure is already solved?

Repressor-DNA complex, solved with 2.6 Å SeMet SAD data & refined against 3.1 Å native data



Before AlphaFold, R/Rfree = 0.27/0.29

*AlphaFold model: A **hypothesis** about this structure*

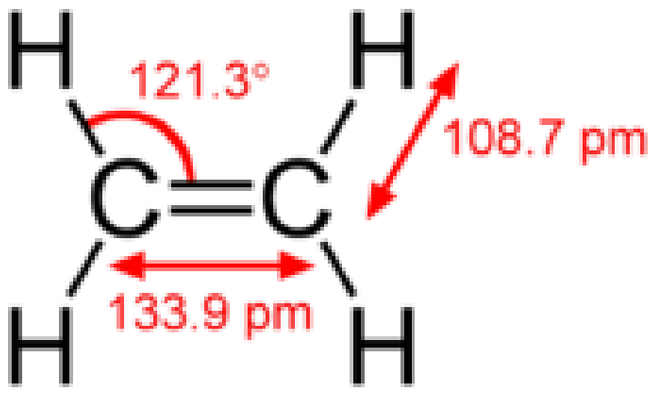
After AlphaFold, R/Rfree = 0.21/0.24
(it was a good hypothesis)

Jamie Wallen, Western Carolina University

3. Use an Alphafold model for reference model restraints

Restraints: *a priori* knowledge

- Restraints increase the number of observations.
- Restraints modify the target function by creating relationships between independent parameters.
- **Example**: restrained bond lengths



- the coordinates of the two atoms are independent
- restraint keeps their distance within a certain target value
- imposes a penalty if it deviates too much.

Reference model Restraints

Concept

- Use a related model to generate a set of torsion restraints.
- Restrain each torsion angle in the working model to the corresponding torsion angle in the reference model.
- Allows for structural differences.

When to use

Low resolution (worse than 3Å).

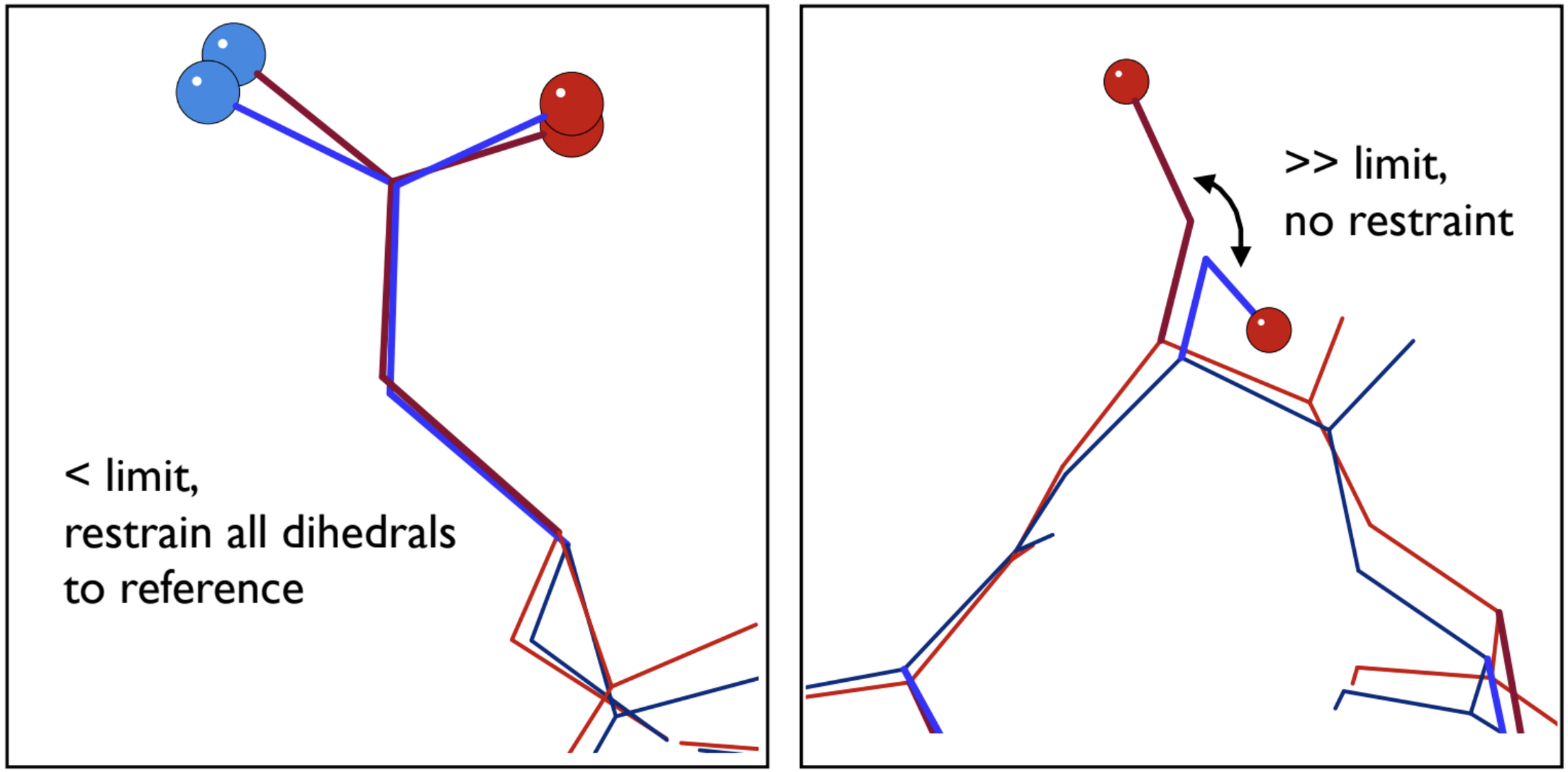
If no high resolution homologue available, could use AF model for reference model (AF models have good geometry).

Headd JJ et al., 2012, *Acta Cryst.* **D68**:381-390

Reference model Restraints

The 'limit' parameter

default: limit = 15.0°

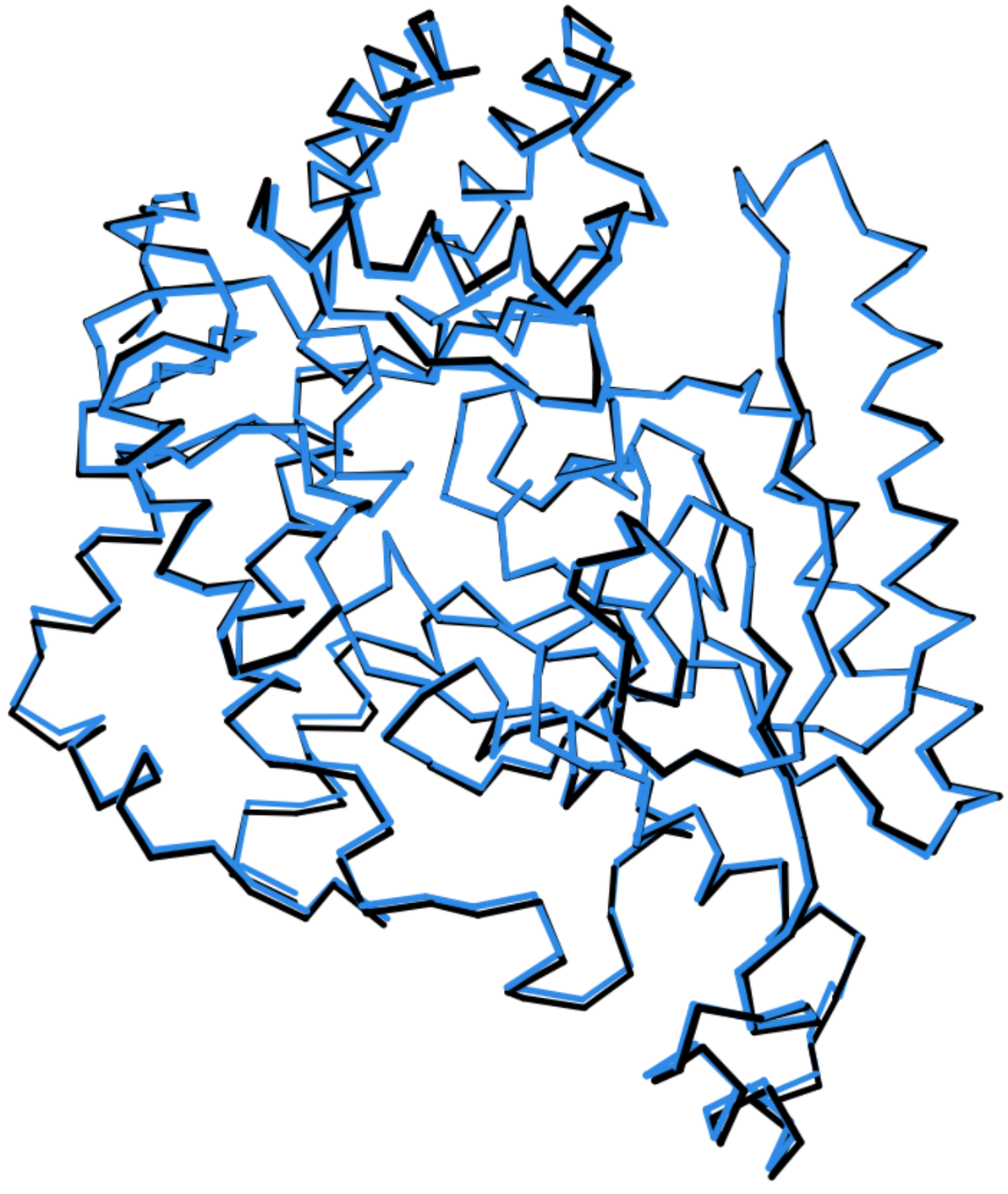


Headd JJ et al., 2012, *Acta Cryst.* **D68**:381-390

Reference model Restraints: example

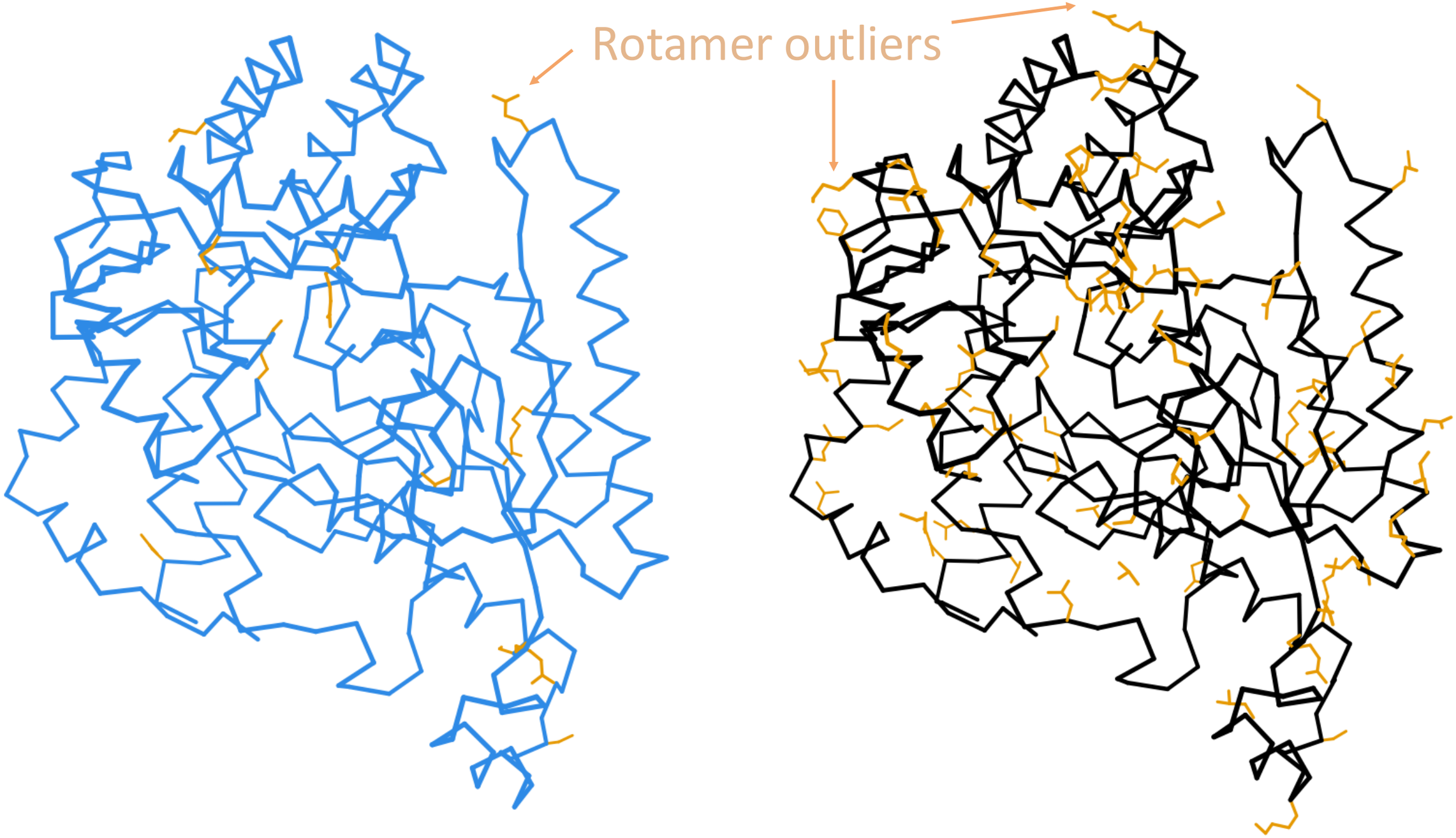
1GTX: 3.0 Å

1OHV: 2.3 Å



4-aminobutyrate-aminotransferase

Reference model Restraints: example



1OHV: 2.3 Å

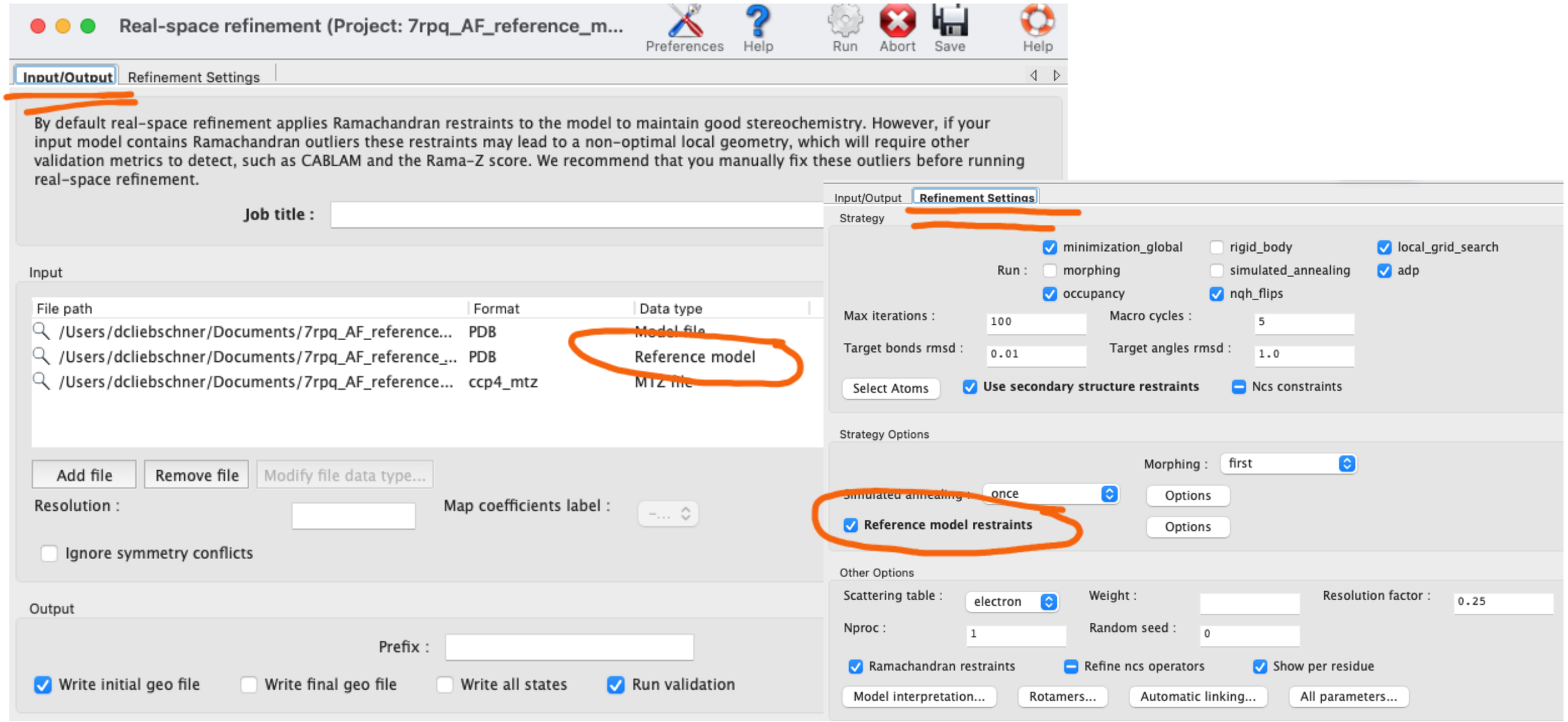
1GTX: 3.0 Å

4-aminobutyrate-aminotransferase

Reference model Restraints

How to use

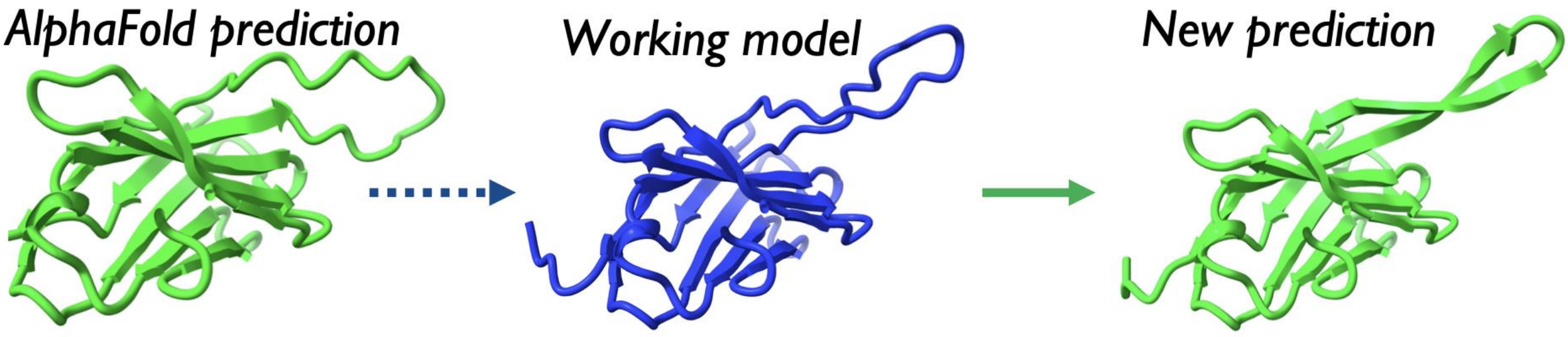
Supply a reference model in phenix.refine; check the corresponding box.
(Oleg Sobolev: working on finding reference automatically)



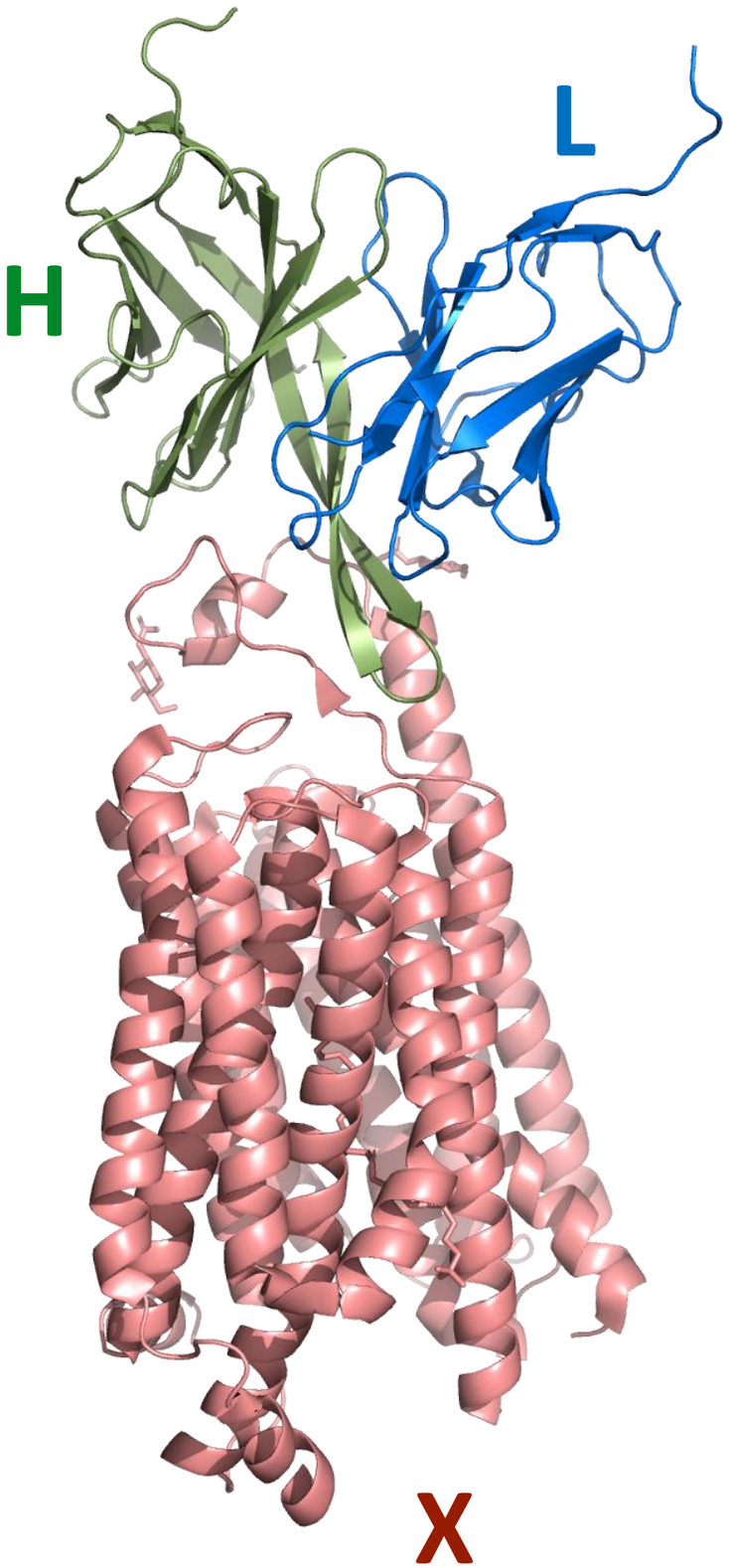
Use your working model to get a new AlphaFold prediction

Why?

Because your new prediction might be better than your model



Example: Fab heavy chain



7MJS

Single-Particle Cryo-EM Structure of Major Facilitator Superfamily Domain containing 2A in complex with LPC-18:3

PDB DOI: https://doi.org/10.2210/pdb7MJS/pdb
EM Map EMD-23883: EMDB EMDataResource

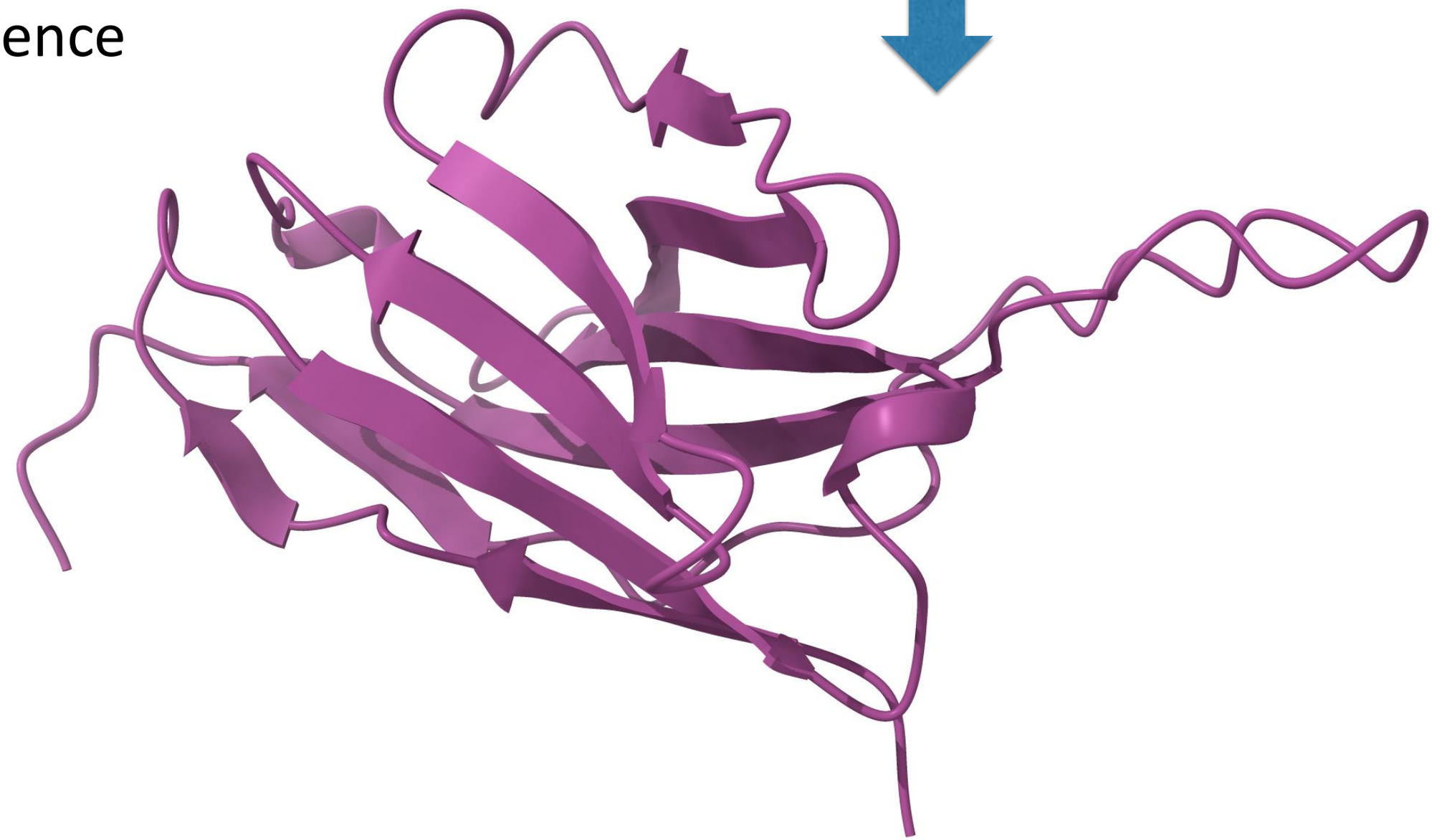
3.03 Å resolution

A loop that interacts with other chains is not correctly predicted.

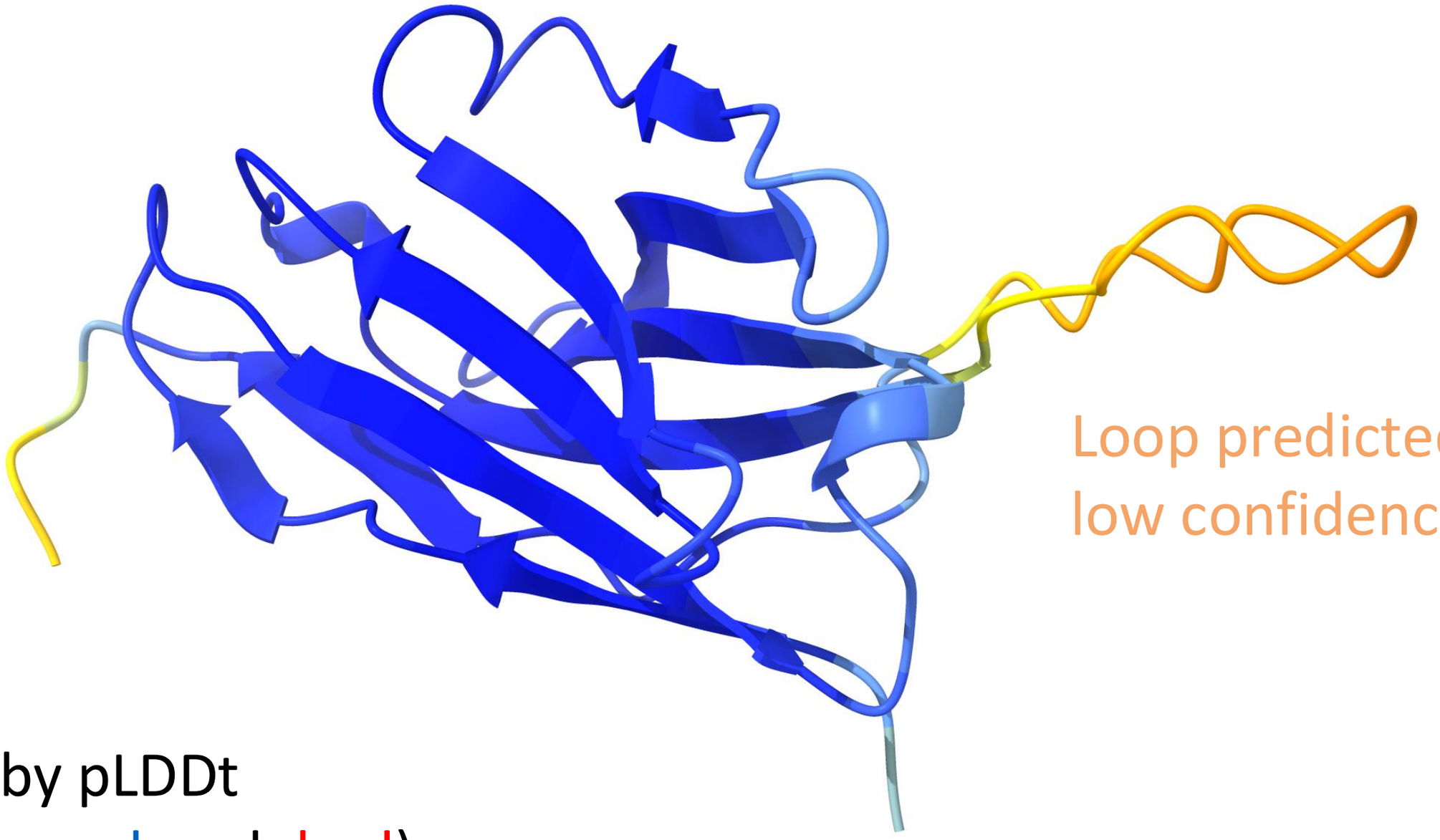
AF2 prediction of chain H

EVQLVESGGGLVQPGGSLRLSCAASGFNIYSSSIHWVRQ
APGKGLEWVAYIYSYSGYTSYADSVKGRFTISADTSKNTA
YLQMNSLRAEDTAVYYCARSLEYLYSSGYQYKWATGLDY
WGQGTLVTVSSAST

Sequence



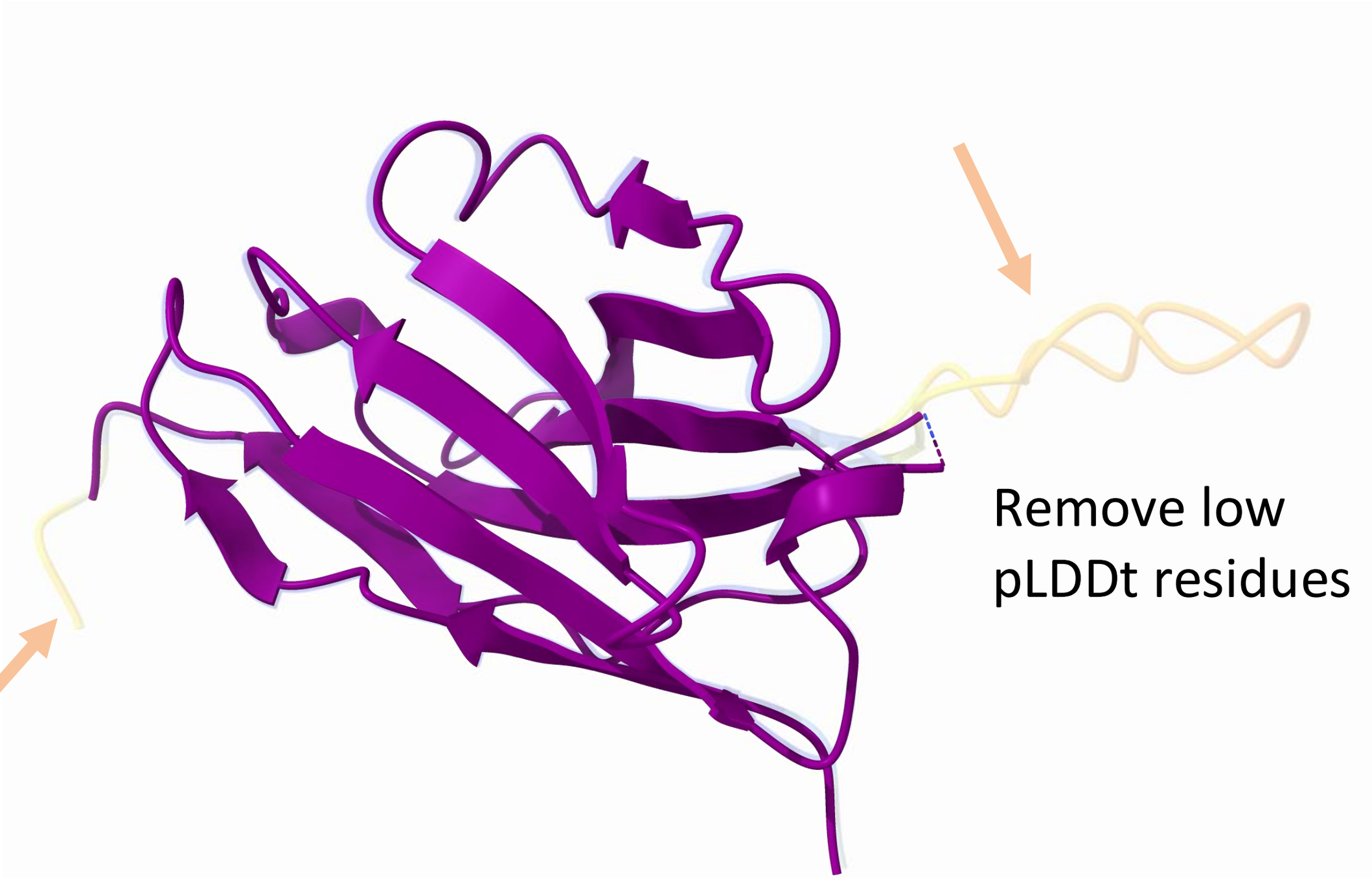
AF2 prediction of chain H



Loop predicted with low confidence

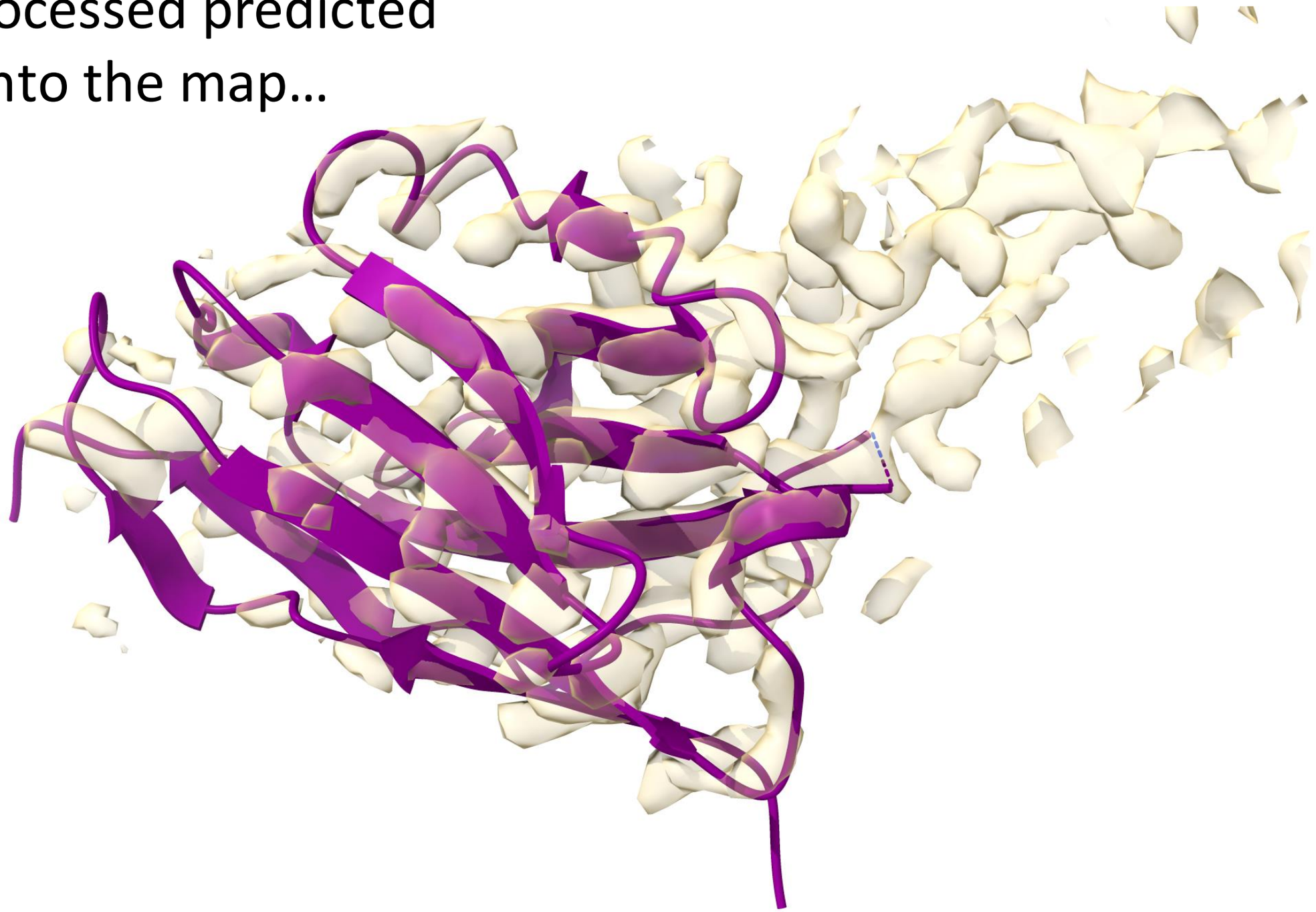
Color by pLDDt
(blue: good, red: bad)

Process AF2 prediction

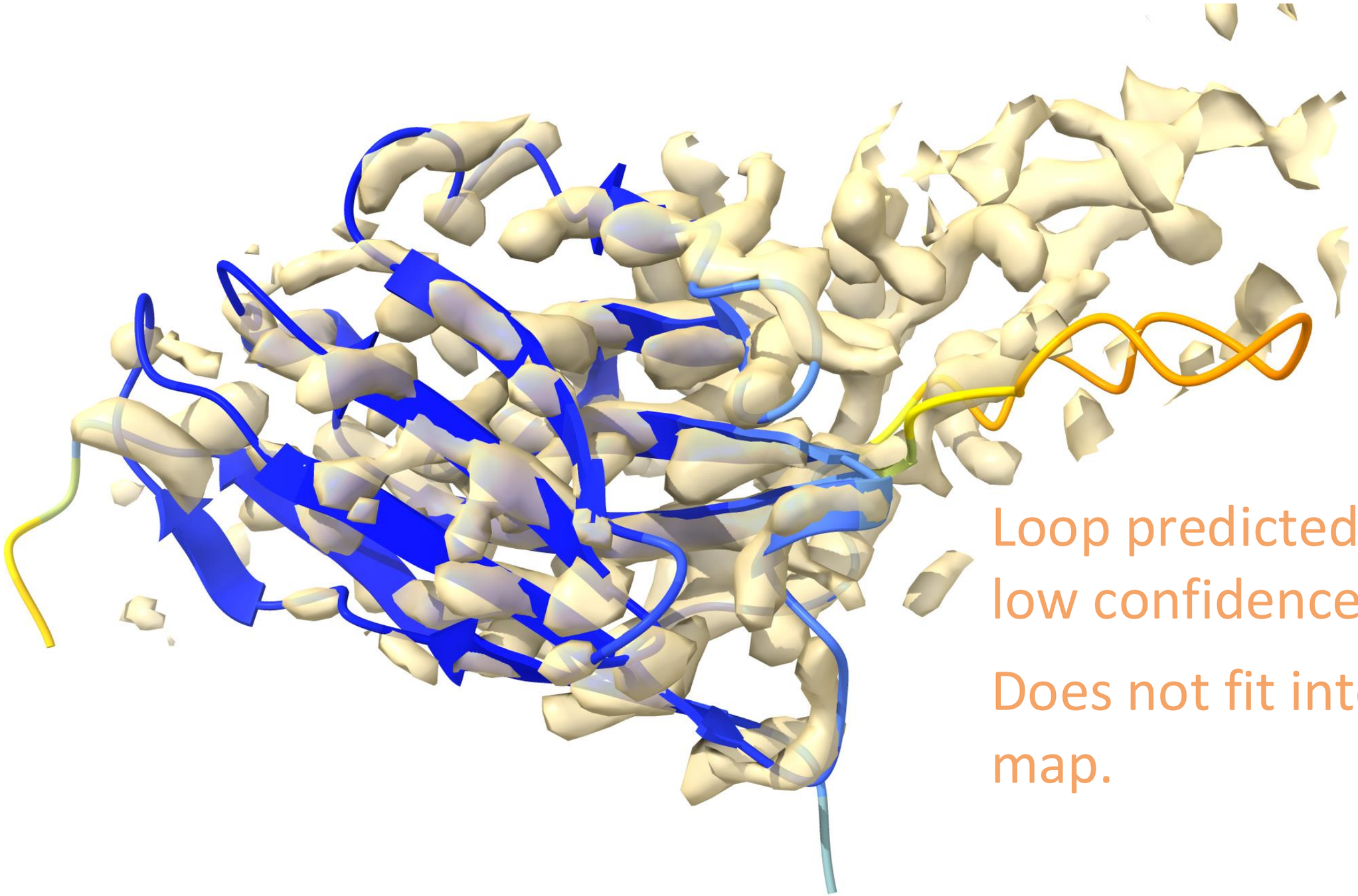


Dock processed model into the map

Dock processed predicted model into the map…



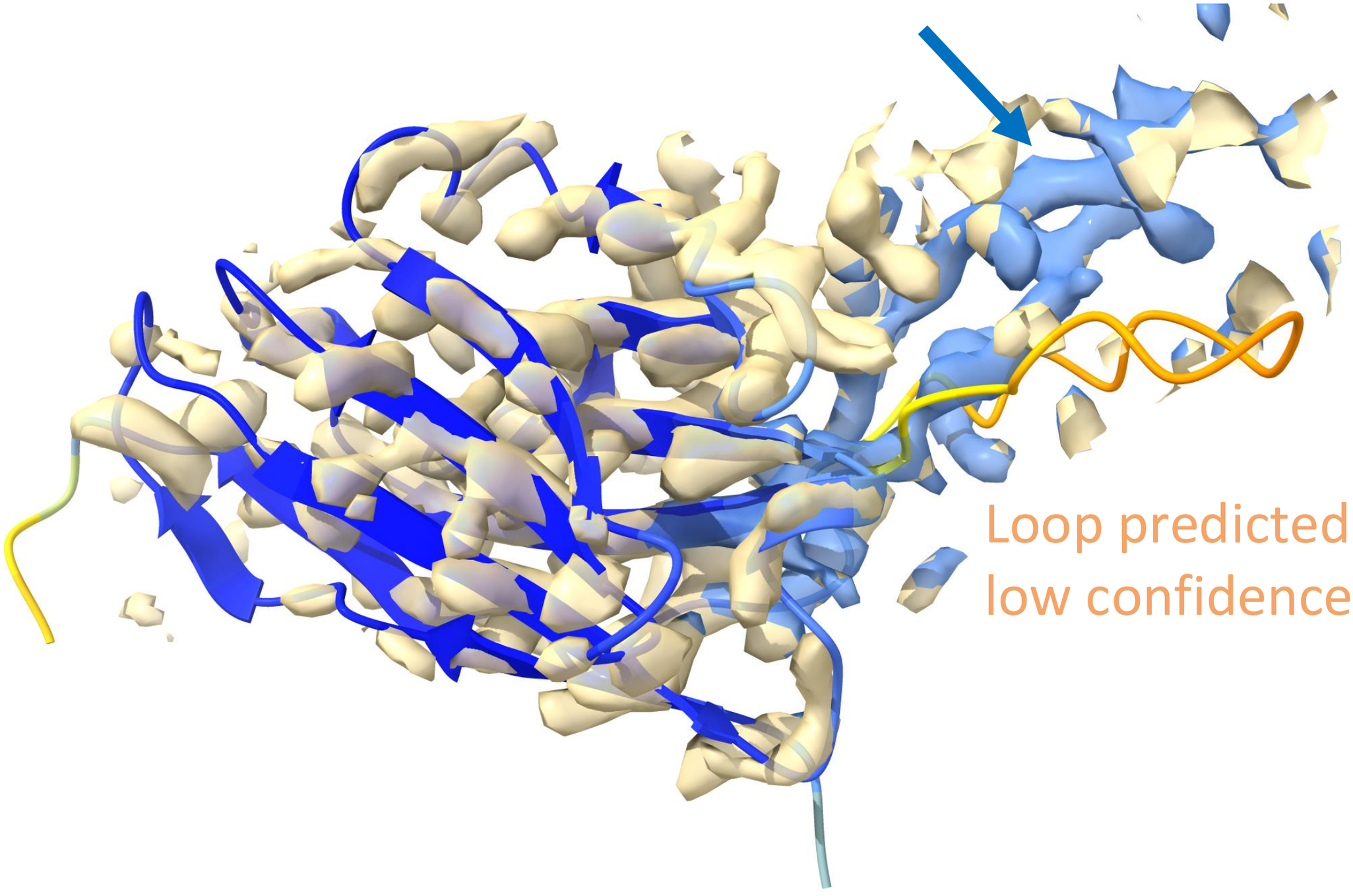
AF2 prediction of chain H



Loop predicted with low confidence

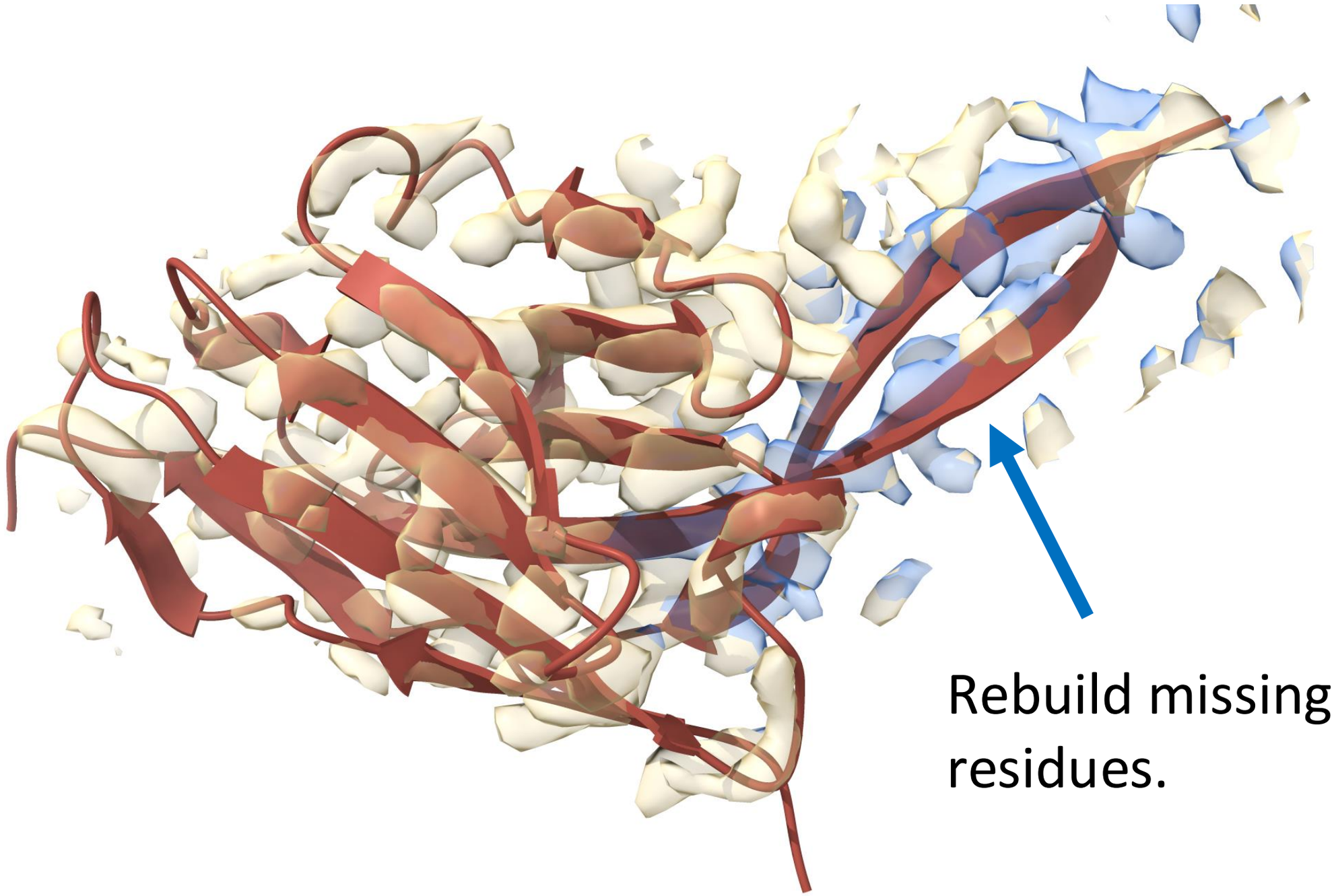
Does not fit into map.

AF2 prediction of chain H



Loop predicted with low confidence

Dock and rebuild model



Rebuild missing residues.

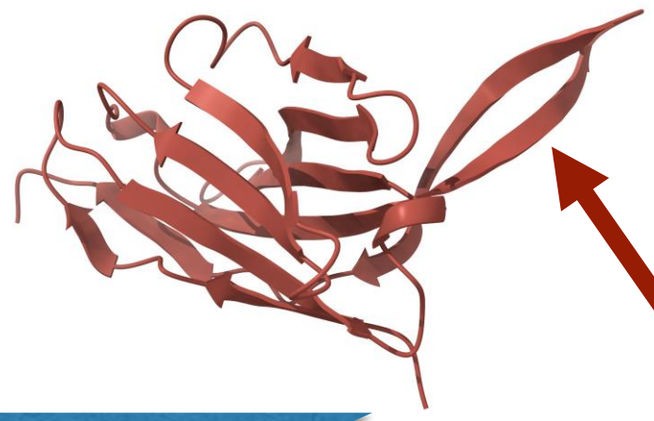
“predicted-processed-docked-rebuilt” model

Make a new prediction

Sequence

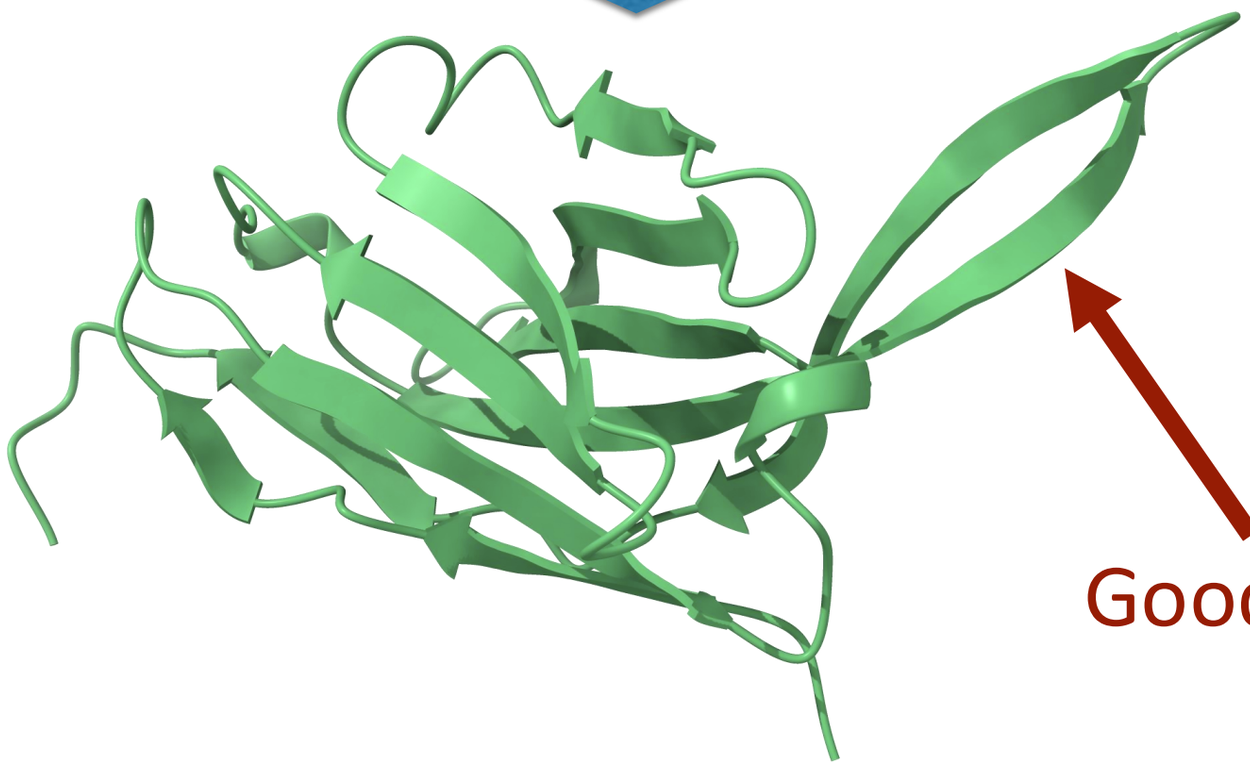
EVQLVESGGGLVQPGGSLRLSCAASGFNIYSSSIHWVRQ
APGKGLEWVAYIYSYSGYTSYADSVKGRFTISADTSKNTA
YLQMNSLRAEDTAVYYCARSLEYLYSSGYQYKWATGLDY
WGQGTLVTVSSAST

Template



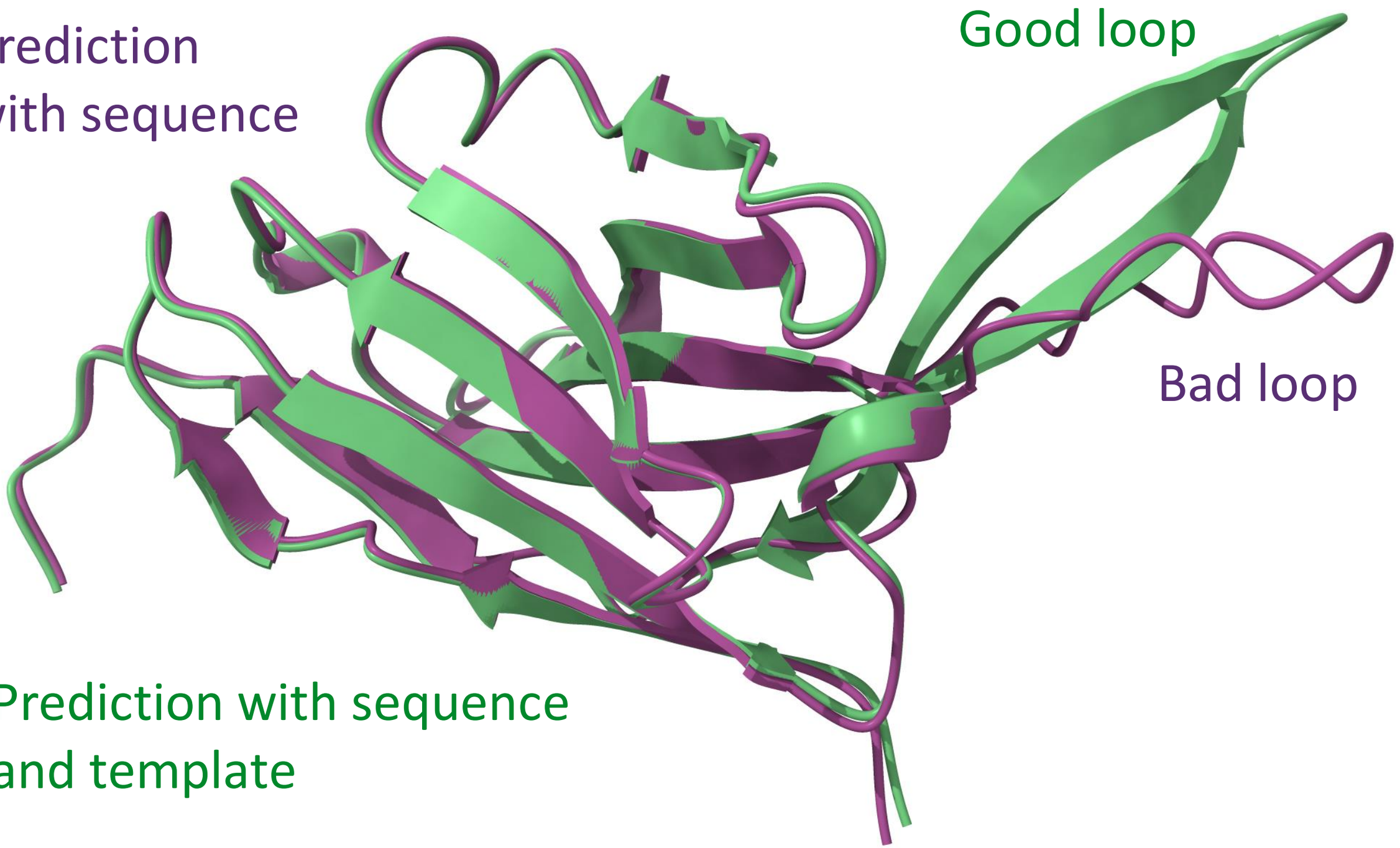
Good loop

AlphaFold



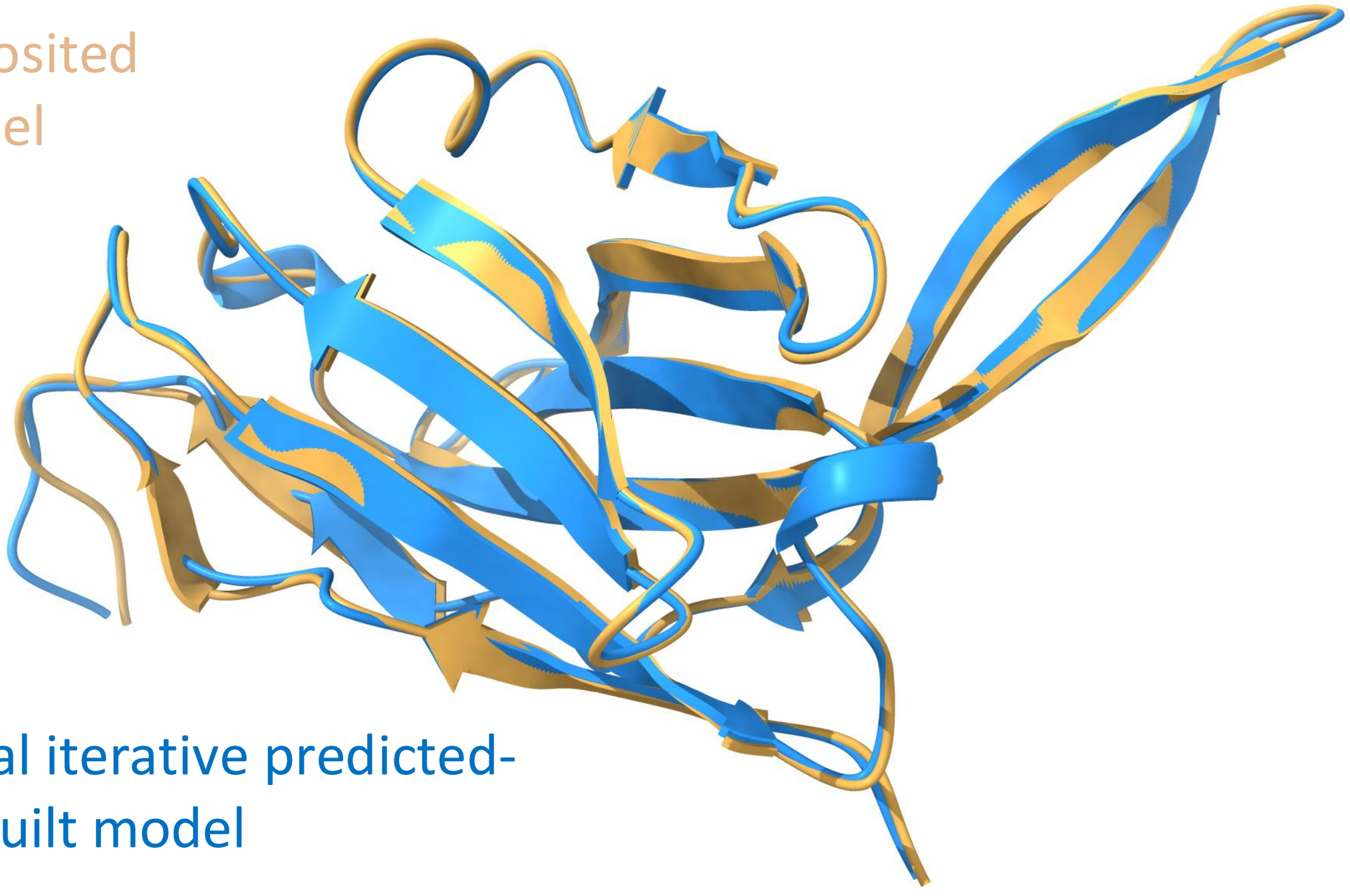
Good loop

Using a template improves prediction



Iterate prediction and rebuilding

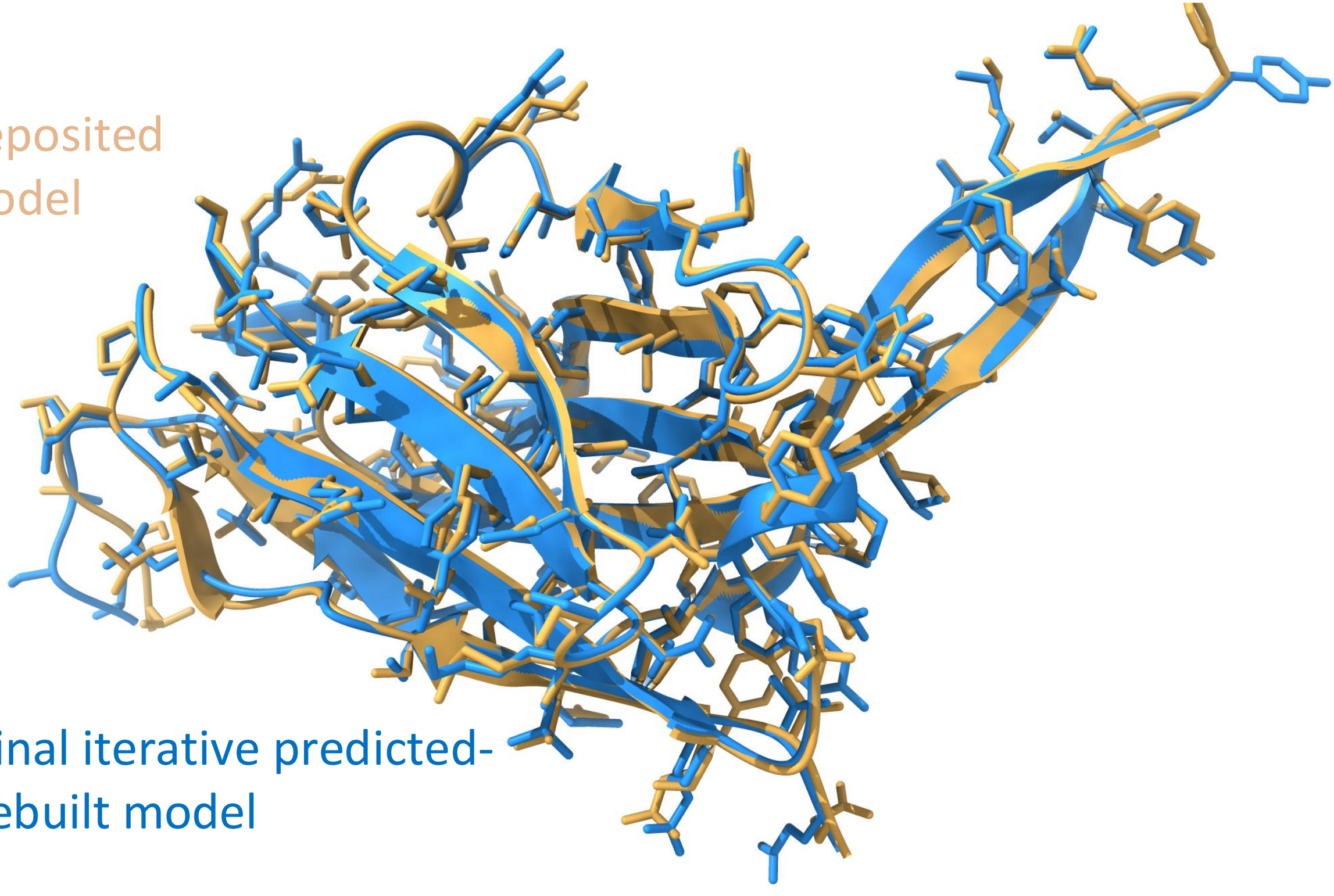
Deposited model



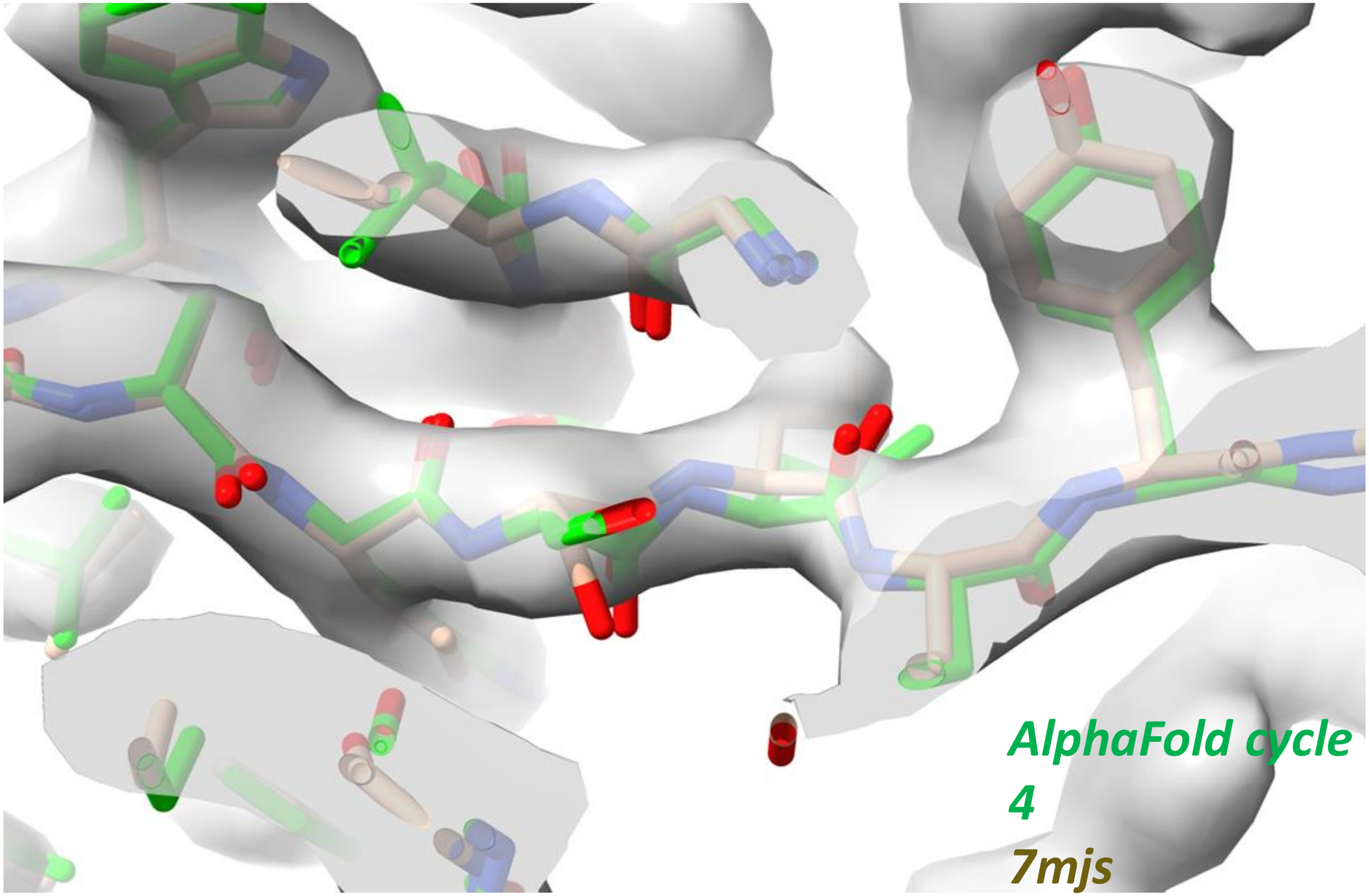
Final iterative predicted-rebuilt model

Iterate prediction and rebuilding

Deposited model



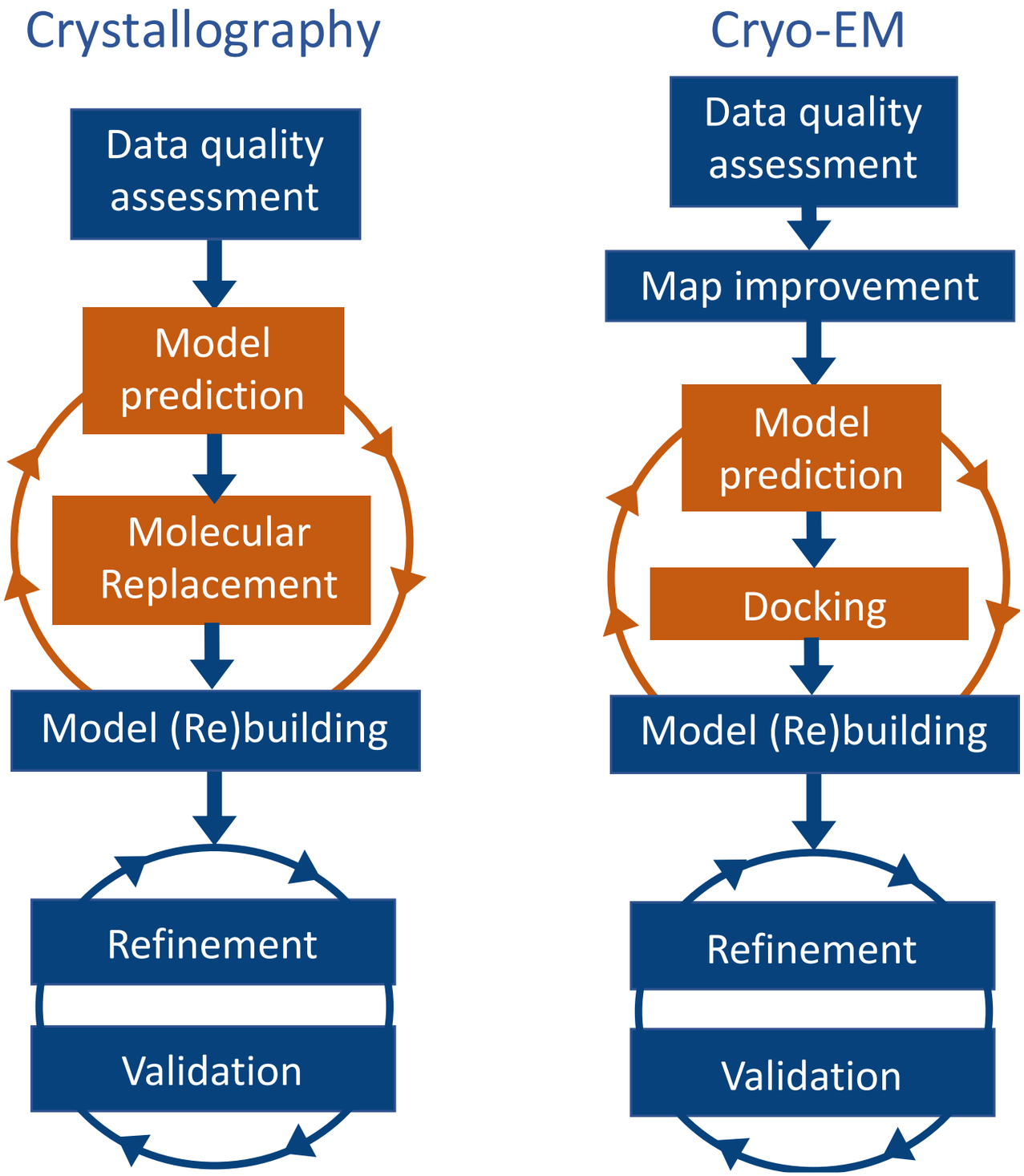
Final iterative predicted-rebuilt model



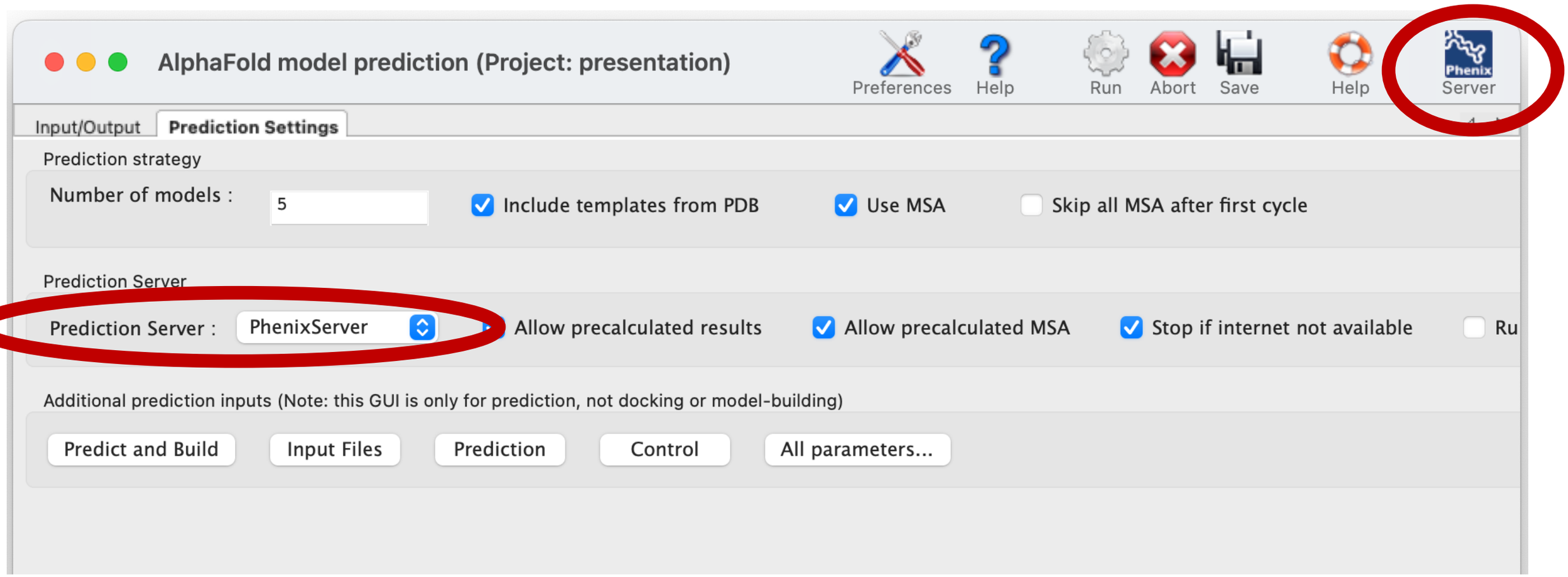
Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319

Using predicted models

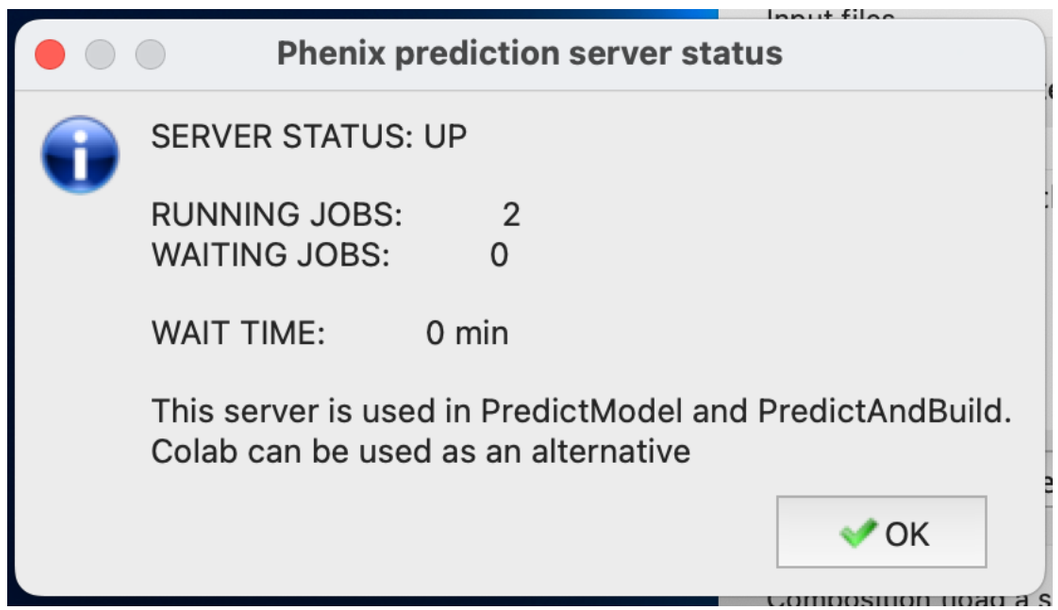
Updated approach: Iterate prediction and model building



Phenix Server for running AlphaFold



No need for a local AlphaFold installation



Process predicted model

Process Predicted Model (Project: 7rpq_AF_reference_model)

Preferences Help Run Abort Help

Configure

ProcessPredictedModel: Prepare predicted model for structure determination

Replace values in B-factor field with estimated B values.
Optionally remove low-confidence residues and split into domains.

Inputs: Model file (PDB, mmCIF)

Job title :

Predicted model

Input

Note: The B-value field in your model can be pLDDT, RMSD, or B

Predicted model : Browse...

Contents of B-value field for input models : plddt

Optional input files

Output

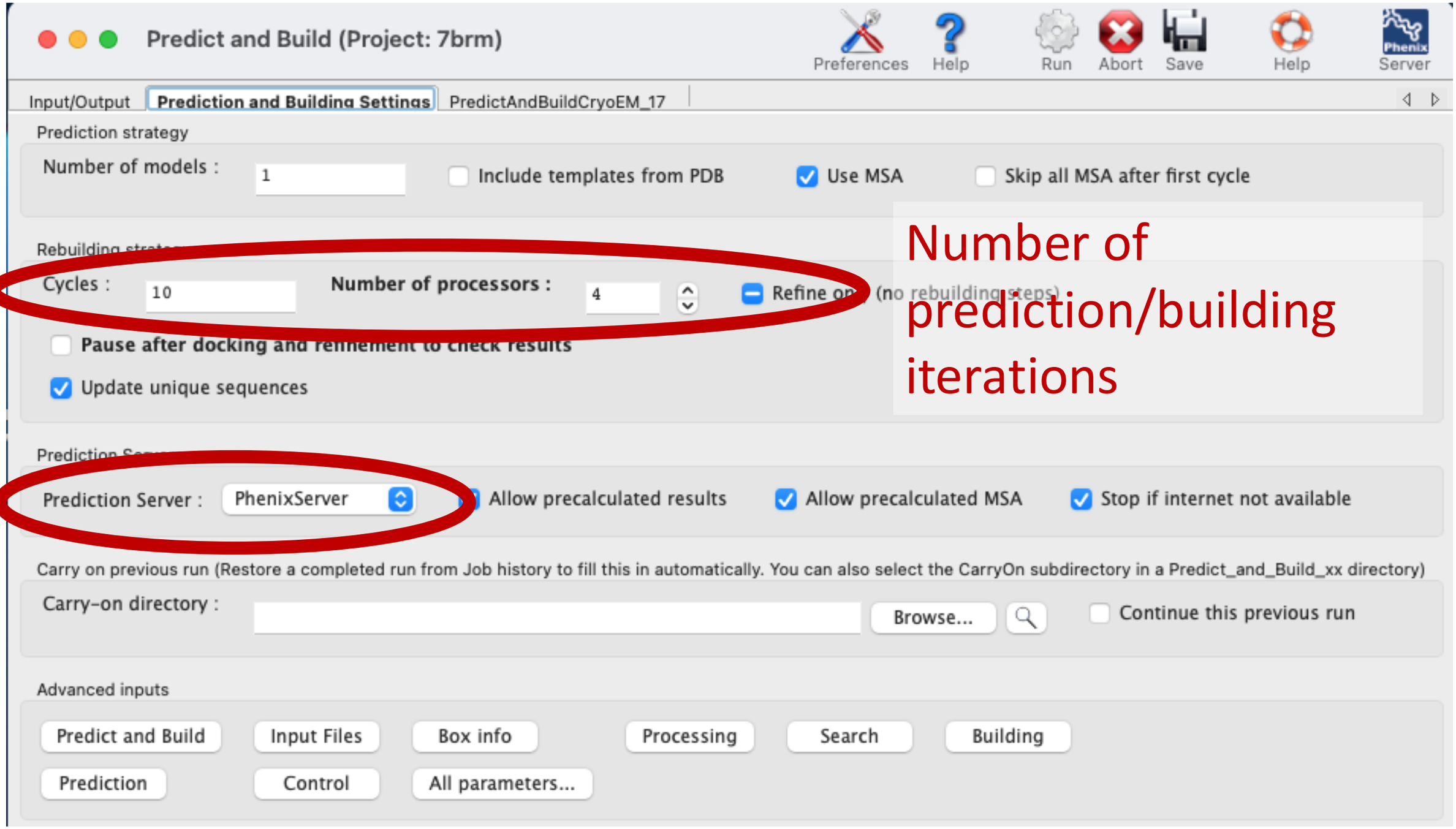
Output file prefix (optional) :

Options

Remove low-confidence residues
Split model into compact regions
Maximum output B : 999.0
Remove hydrogen
Use only single-letter chain ID
Maintain continuous chain

Processing options All parameters...

Iterate with Predict and Build



Fully automatic – AF prediction, processing, building, refinement.

Strategy for structure determination

1. Predict your structure

Design your experiment accordingly

(choose experimental approach, consider trimming at domain boundaries)

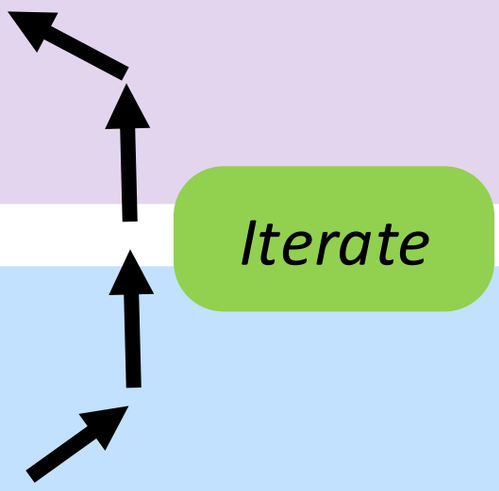
2. Solve your structure

Cryo-EM: docking

X-ray: MR; SAD

3. Update your prediction

Run AlphaFold again with your best model as a template

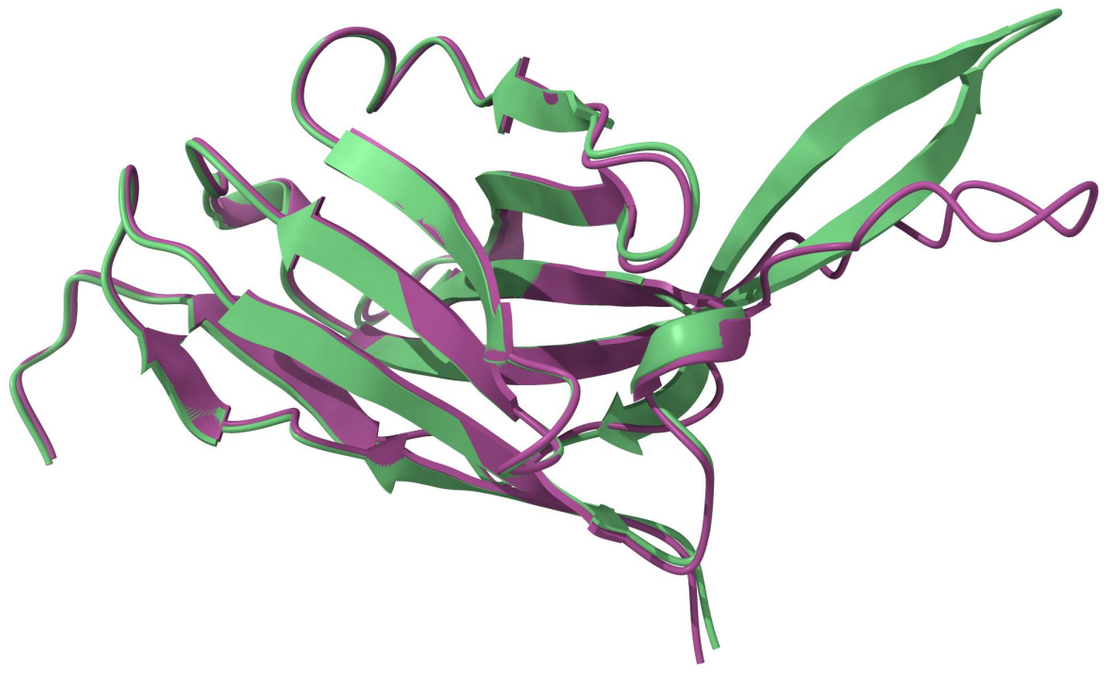


4. Improve your structure

Use your new prediction as hypothesis, rebuild parts

Summary

- AlphaFold models are great hypotheses.
- Can be used for cryo-EM docking (need to interpret the confidence measures), model completion, reference model restraints.
- Iterating prediction and model building can lead to improved models.
- Still need experiment to get a model that best explains the data.



Further reading/material

Documentation:
https://www.phenix-online.org/documentation/

Working with AlphaFold2, RoseTTAFold and other predicted models

You can use the predicted models from AlphaFold and other prediction software in Phenix. Using these models can be very helpful in structure determination because the models can be very accurate over much of their length and the models come with accuracy estimates that allow removal of poorly-predicted regions.

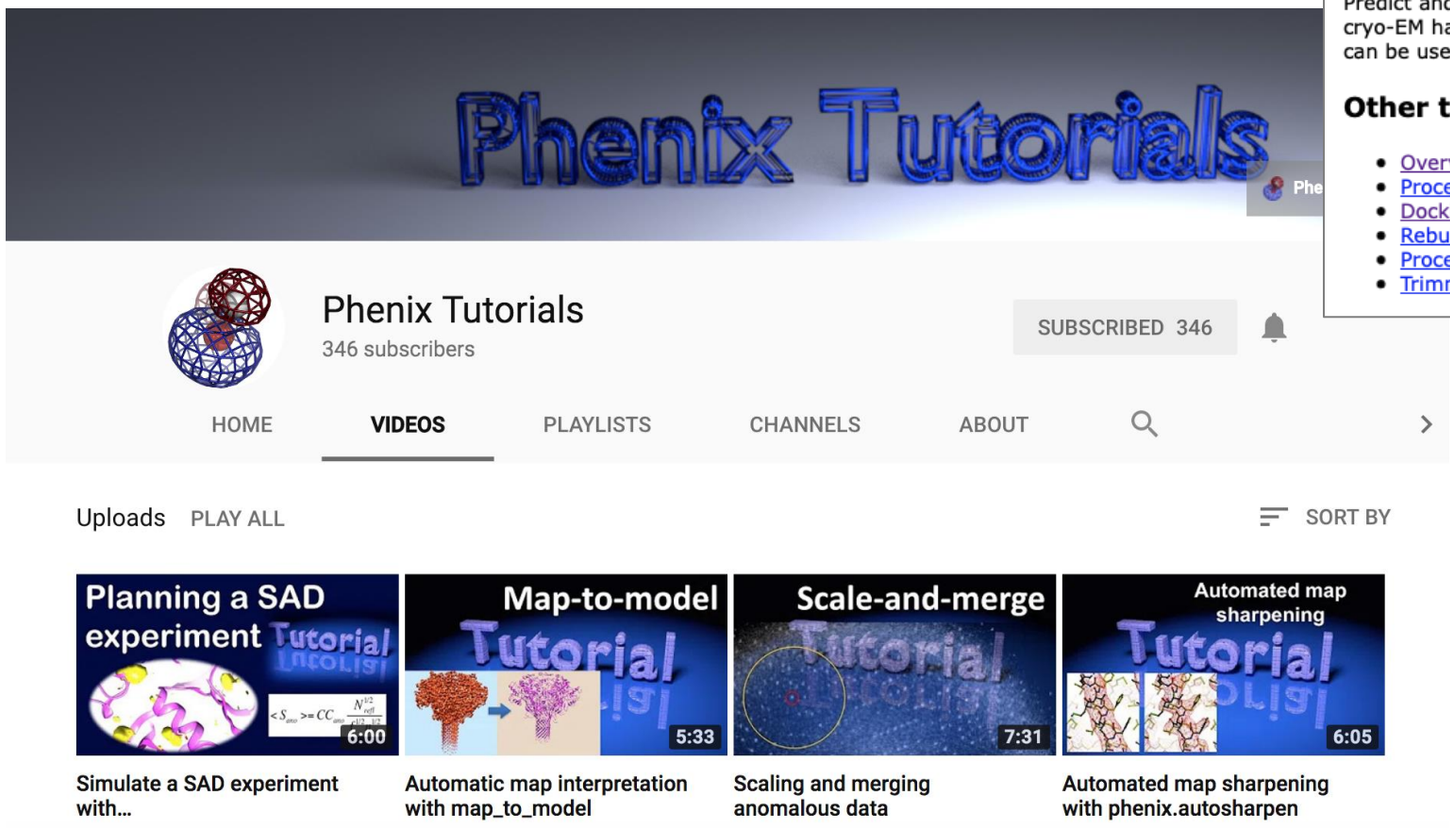
How to use predicted models in Phenix

Use Predict_and_build to incorporate predicted models in the structure solution workflow.

Predict_and_build generates predicted models and uses them to solve an X-ray structure by MR or to interpret a cryo-EM map. The tool then carries out iterative model rebuilding and prediction to improve the models. The iterative procedure allows creating more accurate predicted models than can be obtained with a simple prediction.
Predict and build can automatically generate a fairly accurate model starting from just a sequence file and either cryo-EM half-maps or X-ray data. Additionally, it provides morphed versions of unrefined predicted models that can be useful as reference models for refinement.

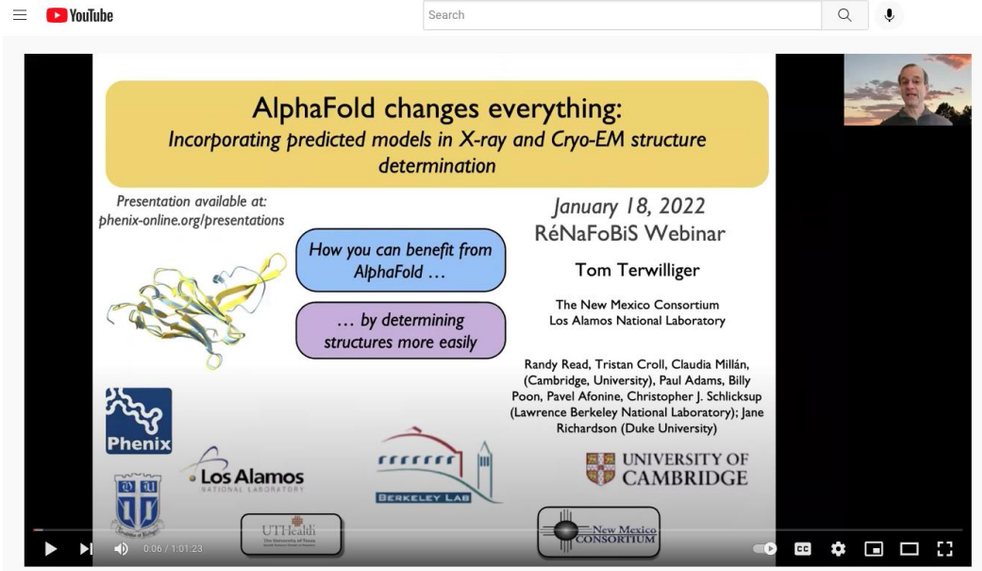
Other tools for using predicted models:

- Overview: AlphaFold and Phenix
- Processing a predicted model
- Docking a processed predicted model in a cryo-EM map
- Rebuilding a docked predicted model in a cryo-EM map
- Processing, docking and rebuilding a predicted model in a cryo-EM map
- Trimming overlapping parts of models



Video tutorials

https://www.youtube.com/c/phenixtutorials



Tom Terwilliger: AlphaFold changes everything

Tom Terwilliger: AlphaFold changes everything

https://youtu.be/ugMPYdPo8Bc?feature=shared

The Phenix Project

Lawrence Berkeley Laboratory

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

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker

University of Cambridge

Randy Read, Airlie McCoy, Alisia Fadini

Duke University

Jane 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