Macromolecular Structure Determination using crystallographic and cryo-EM data

What is Phenix?

What is *Phenix*?

- Package for **automated structure solution** (crystallography, cryo-EM)

- Modern programming concepts and new algorithm development

- Designed to be used by **both novices and experienced users**

- Long-term development and **support**

- Why is it called *Phenix*?

  Python **H**ierarchical **EN**vironment for **I**ntegrated **X**tallography

The Phenix Project

Lawrence Berkeley Laboratory

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

University of Cambridge

Randy Read, Airlie McCoy, Alisia Fadini

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker

Duke University

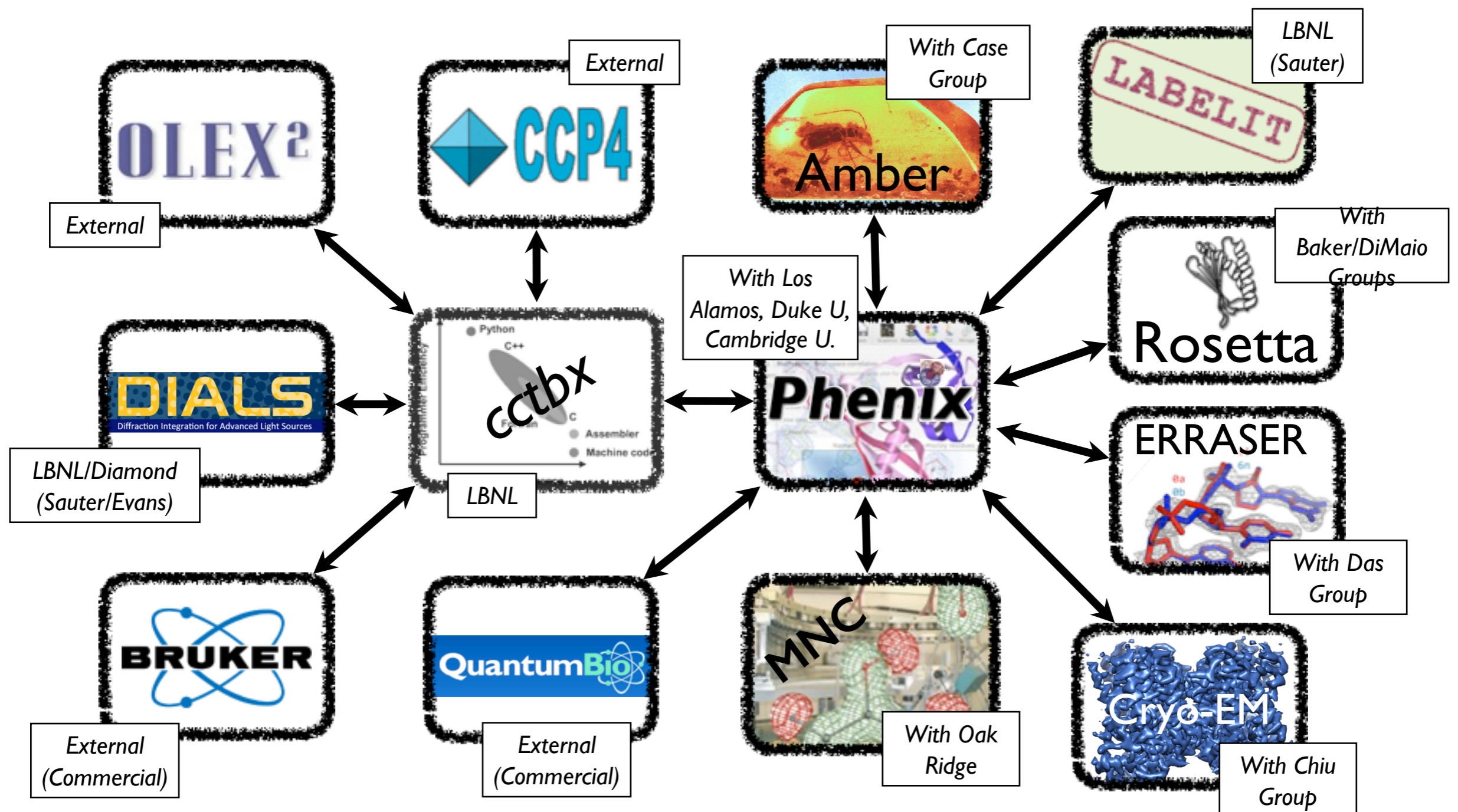Jane & David Richardson, Christopher Williams, Vincent Chen

An NIH/NIGMS funded Program Project

Phenix - a Structural Biology Hub

We have nucleated the development of new computational methods for structural biology

Key Features

- **Python**
  - ➢ Easy scripting of repetitive tasks
  - ➢ Enables rapid prototyping and development

- **Advanced algorithms**
  - ➢ Experimental phasing
  - ➢ Molecular replacement
  - ➢ Automated model building and rebuilding
  - ➢ Structure refinement and validation
  - ➢ Ligand coordinate and restraint generation
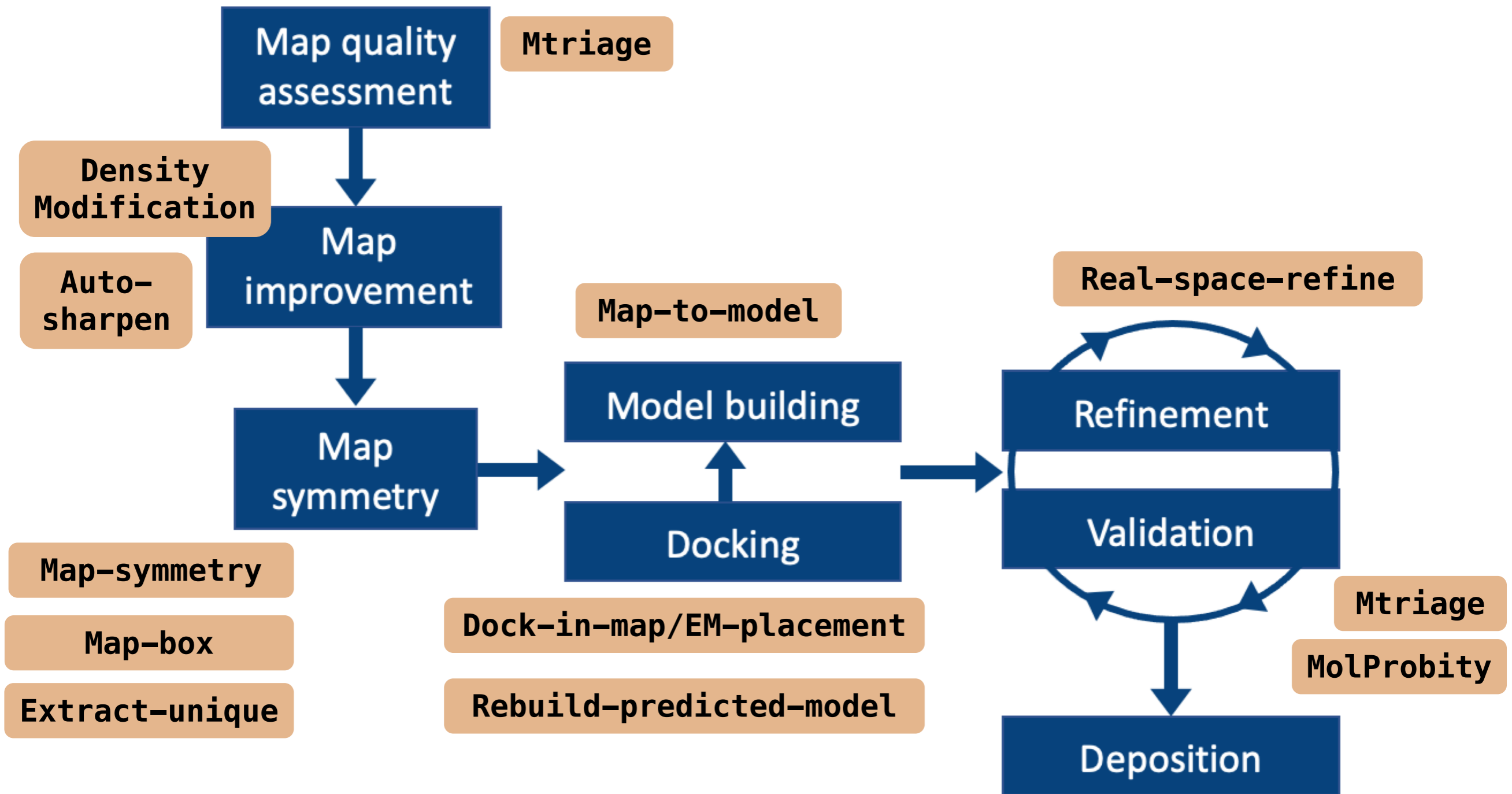
- **Rapid development and bug fixing**

Tools for Crystallography
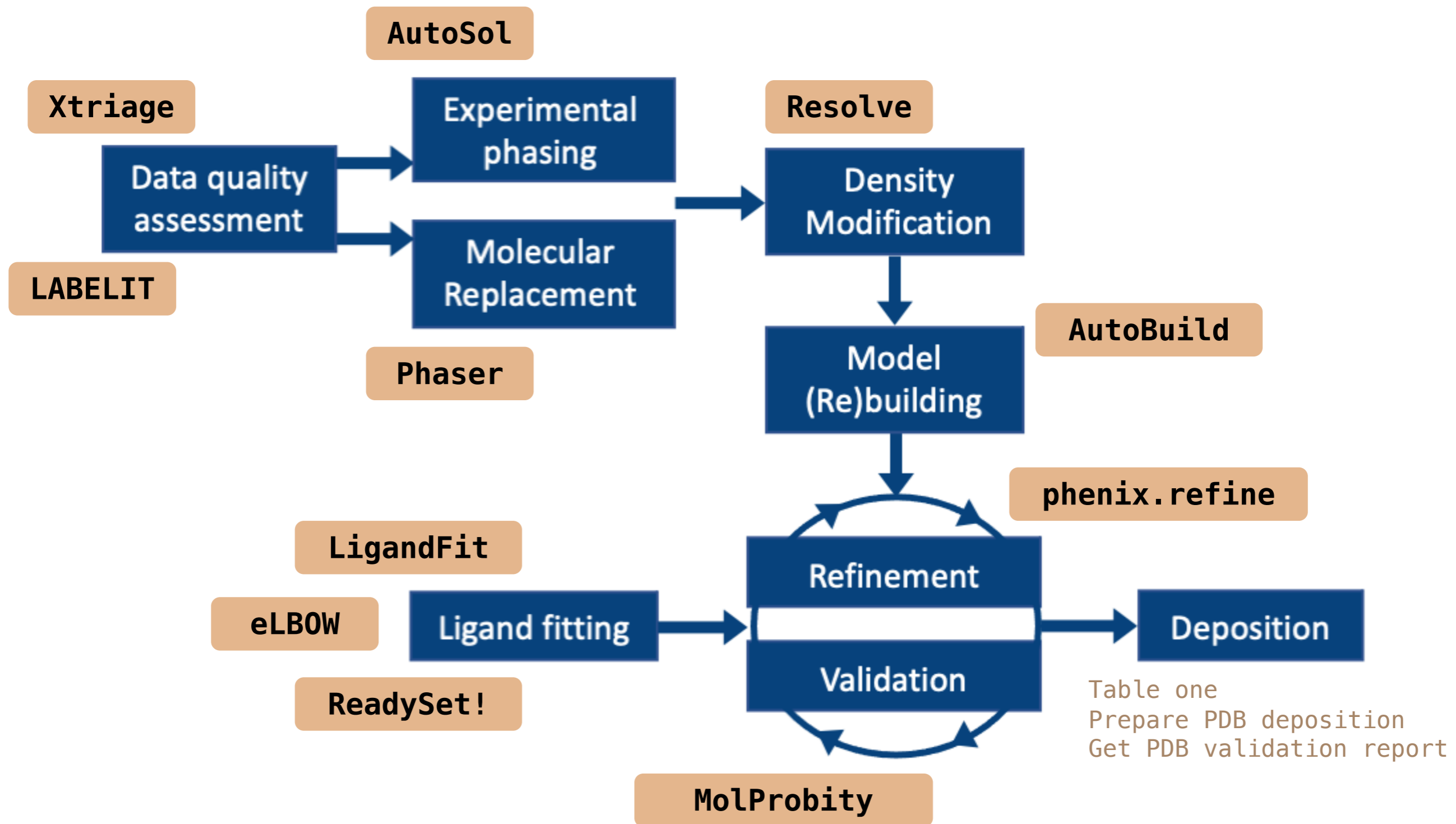
Acta Cryst. 2002, D58:1948-1954   (Phenix)
J. Appl. Cryst. 2002, 35:126-136   (cctbx)
Acta Cryst. 2010, D66: 213-221   (Phenix)
Acta Cryst. 2019 D75:861–877   (Phenix)

Tools for Cryo-EM

Map quality assessment

Mtriage

Density Modification

Auto-sharpen

Map improvement

Map symmetry

Map-symmetry

Map-box

Extract-unique

Map-to-model

Model building

Docking

Dock-in-map/EM-placement

Rebuild-predicted-model

Real-space-refine

Refinement

Validation

Mtriage

MolProbity

Deposition

Acta Cryst. 2002, D58:1948-1954   (Phenix)
J. Appl. Cryst. 2002, 35:126-136   (cctbx)
Acta Cryst. 2010, D66: 213-221   (Phenix)
Acta Cryst. 2019 D75:861–877   (Phenix)

Features

Phenix GUI

Central GUI for job control and to launch new jobs

Coot/PyMOL/ChimeraX integration

- Most results can be opened directly in graphics apps
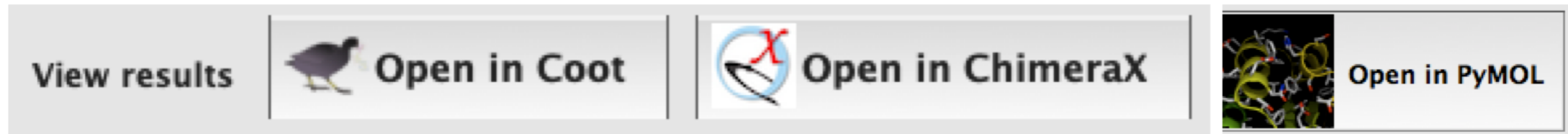
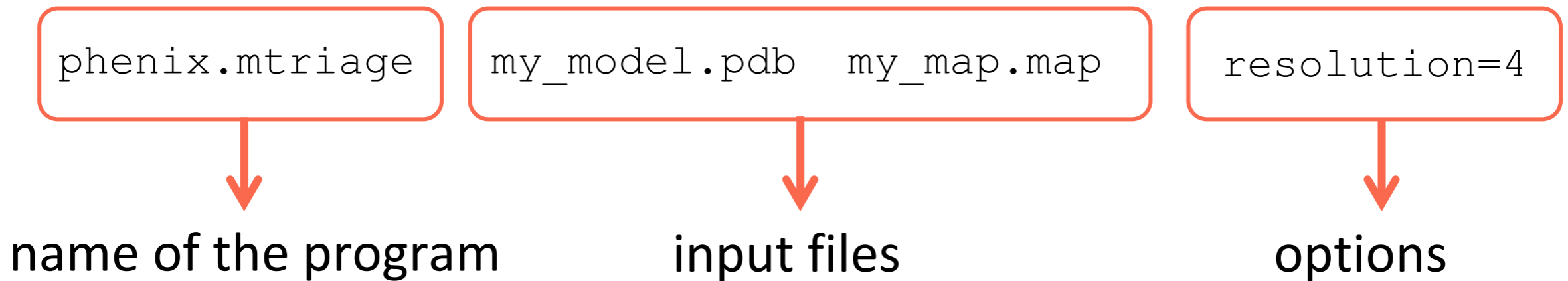- Any PDB file listed in GUI can also be opened
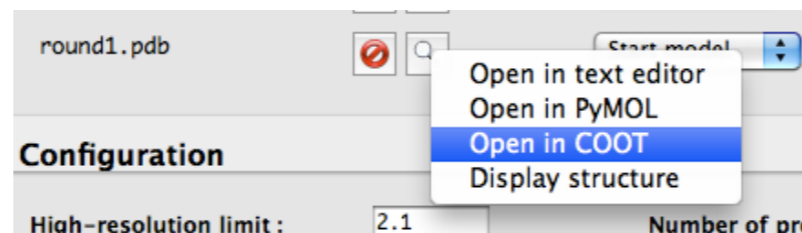
- AutoSol, AutoBuild, and phenix.refine will update Coot continuously while running

- Coot must have Python support (default on Mac)

- Specific paths to executables usually required on Linux

  `Preferences → Graphics → Full path to Coot [...PyMOL]`

Command Line Tools

Run on the terminal

| phenix.mtriage | my_model.pdb   my_map.map | resolution=4 |

name of the program | input files | options

Run in a python script

```python
try:
    easy_run.call("phenix.mtriage my_model.pdb my_map.map")
except Exception as e:
    msg = traceback.format_exc()
    print(msg)
```

Phenix Availability

phenix-online.org

Supported on:

- ○ Linux
- ○ macOS
- ○ Windows

Extensive documentation (online and via GUI)

Nightly builds

Version	Date	Status	Logs	Info
dev-3758	2020-01-22	successful	ci, intel-linux-2.6-x86_64-centos6, mac-intel-osx-x86_64, intel-windows-x86_64	docs; changelog
dev-3753	2020-01-17	successful	ci, intel-linux-2.6-x86_64-centos6, mac-intel-osx-x86_64, intel-windows-x86_64	docs; changelog
dev-3751	2020-01-15	successful	ci, intel-windows-x86_64, intel-linux-2.6-x86_64-centos6, mac-intel-osx-x86_64	docs; changelog
1.17.1-3660	**2019-10-16**	**successful**	**ci, intel-windows-x86_64, intel-linux-2.6-x86_64-centos6, mac-intel-osx-x86_64**	**Official 1.17.1 release; docs; changelog**

Video Tutorials

Dorothee Liebschner, Nigel Moriarty,
Tom Terwilliger, Christopher Schlicksup, Vincent Chen
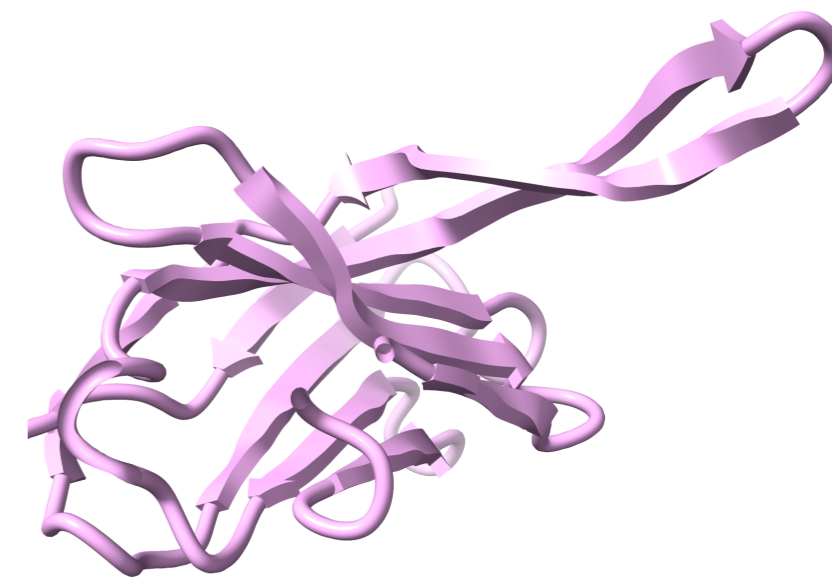
What's new?

Accurate predicted models

EVQLVESGGGLVQPGGSLRLSCAASGFN**I**YSSS**I**HWVRQAPGKGLEWVAYI
.................................**F**...**M**..........Q..........
...................K...........**Y**...**L**......A..........
..................A.............................V.....
.................A............................
...................**L**...**V**....E..........
.................A...............................Q.....

Sequence
Multiple sequence alignment

→

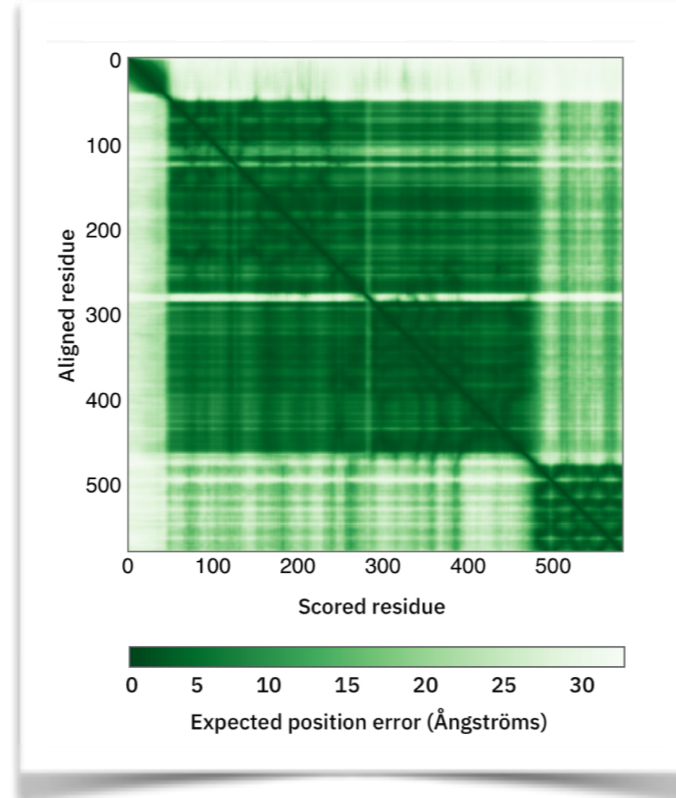3D prediction

New tools for predicted models in Phenix

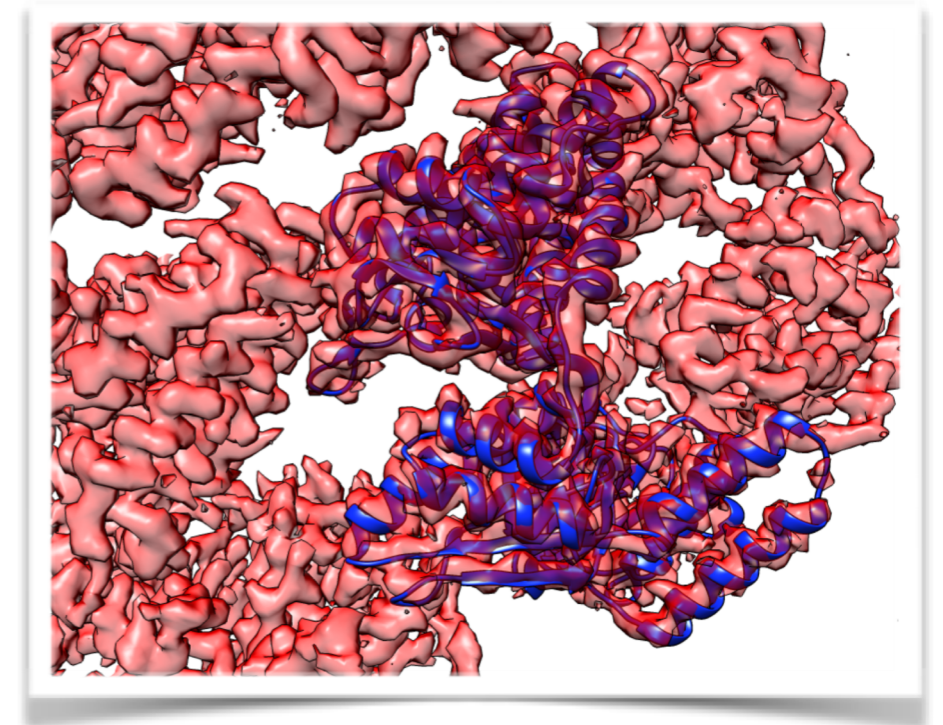>chain ' A'
XXXXXXXXXXXXXXXXCLTAPPKEAARPTLMPRAQSYKDLTHLPAP
TGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDS
RWFIPLERQGLQNLLNERKIIRAAQENGTVAINNRIPLQSLTAAN
IMVEGSIIGYESNVKSGGVGARYFGIGADTQYQLDQIAVNLRVVN
VSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNE
PVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRH
MS
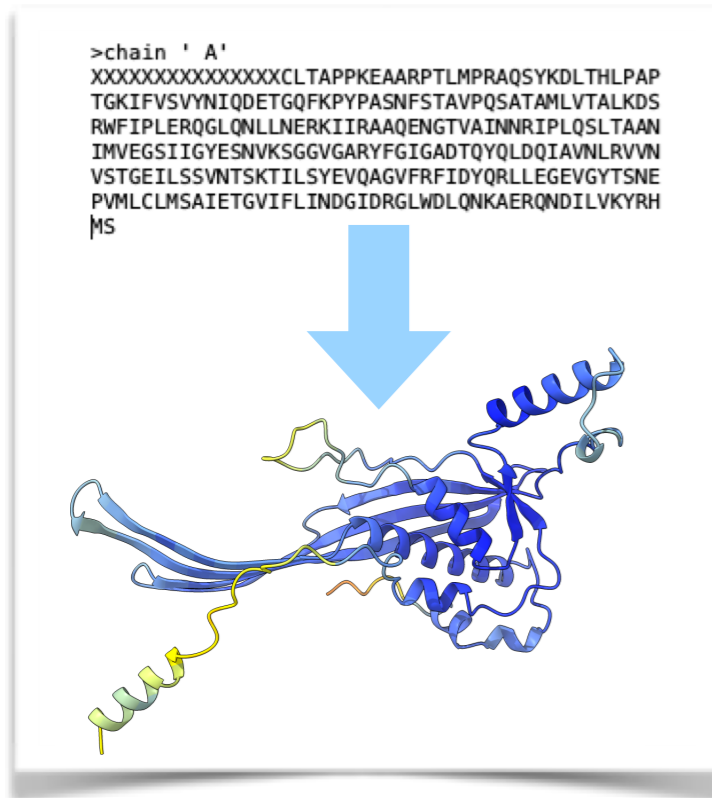
AlphaFold model prediction

(Phenix server, no need to have AF installed locally)

Process predicted model
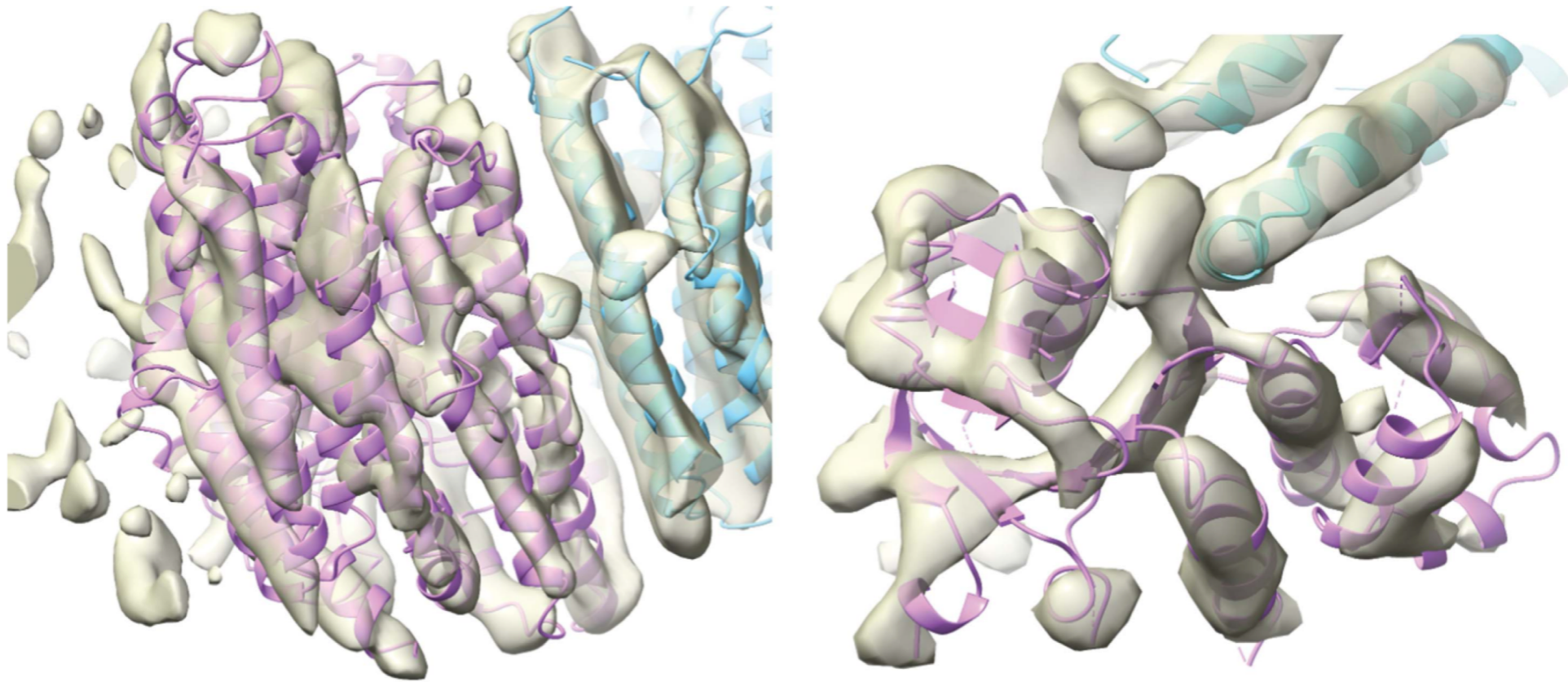
Predict and Build

(Iterative AlphaFold prediction, docking, and rebuilding)

Fully automatic!

Likelihood-based EM docking

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

$$LLG(\mathbf{E}_{mean}; \mathbf{E}_C) = \frac{2}{1 - D_{obs}^2 \sigma_A^2} D_{obs} \sigma_A E_{mean} E_C \cos(\Delta\varphi)$$
$$- \frac{D_{obs}^2 \sigma_A^2 (E_{mean}^2 + E_C^2)}{1 - D_{obs}^2 \sigma_A^2} - \ln(1 - D_{obs}^2 \sigma_A^2).$$

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

Barbed wire analysis

- Automatically select the most promising regions of an AlphaFold prediction
- Annotate a prediction to help you make informed decisions about it.
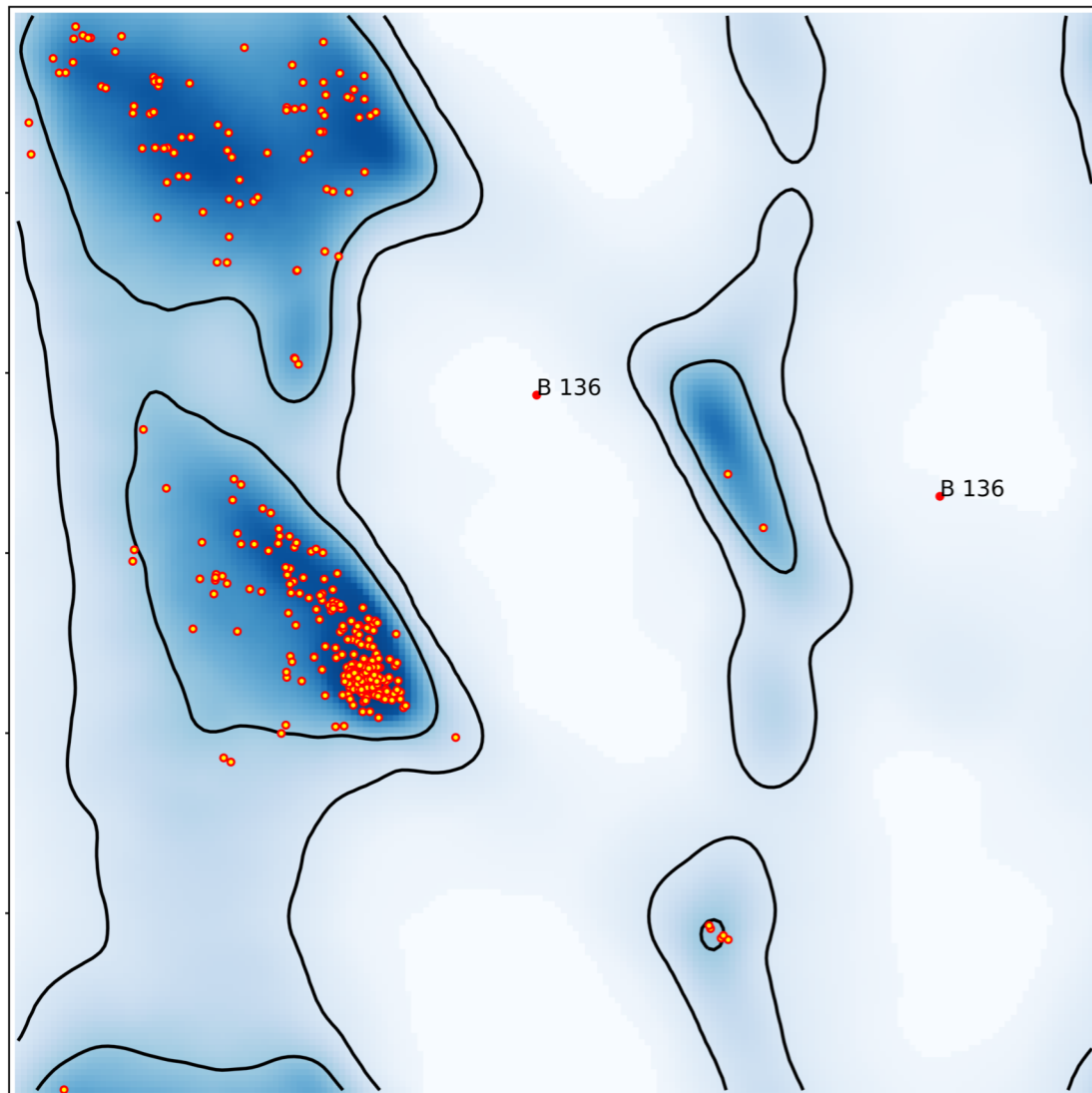
Uniprot **P53076**
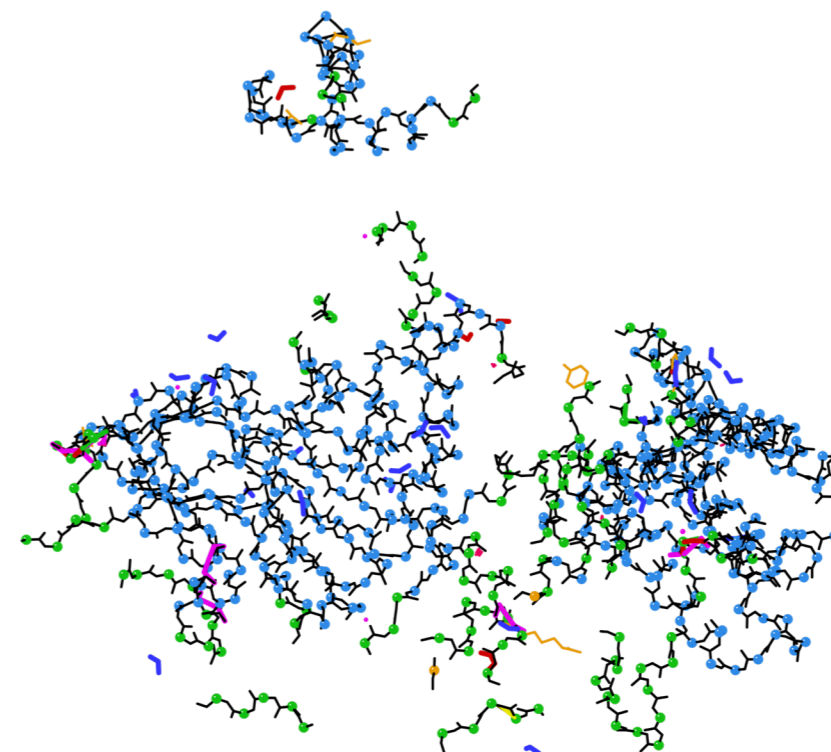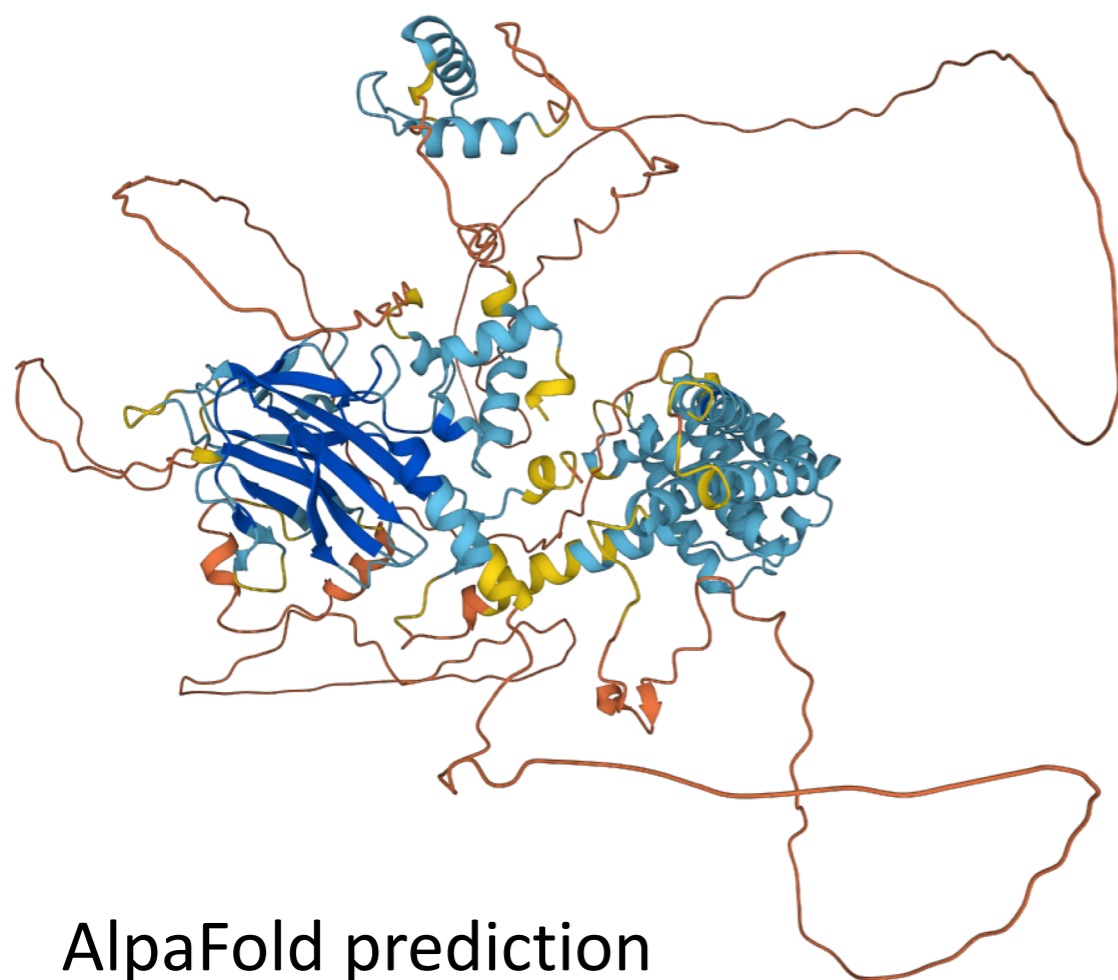
AlpaFold prediction

Annotated and pruned prediction

Sunday, August 27th 5:55 pm, Christopher Williams, *Identification and Valdiation of low-pLDDT regions in AF2 predictions* (A011)
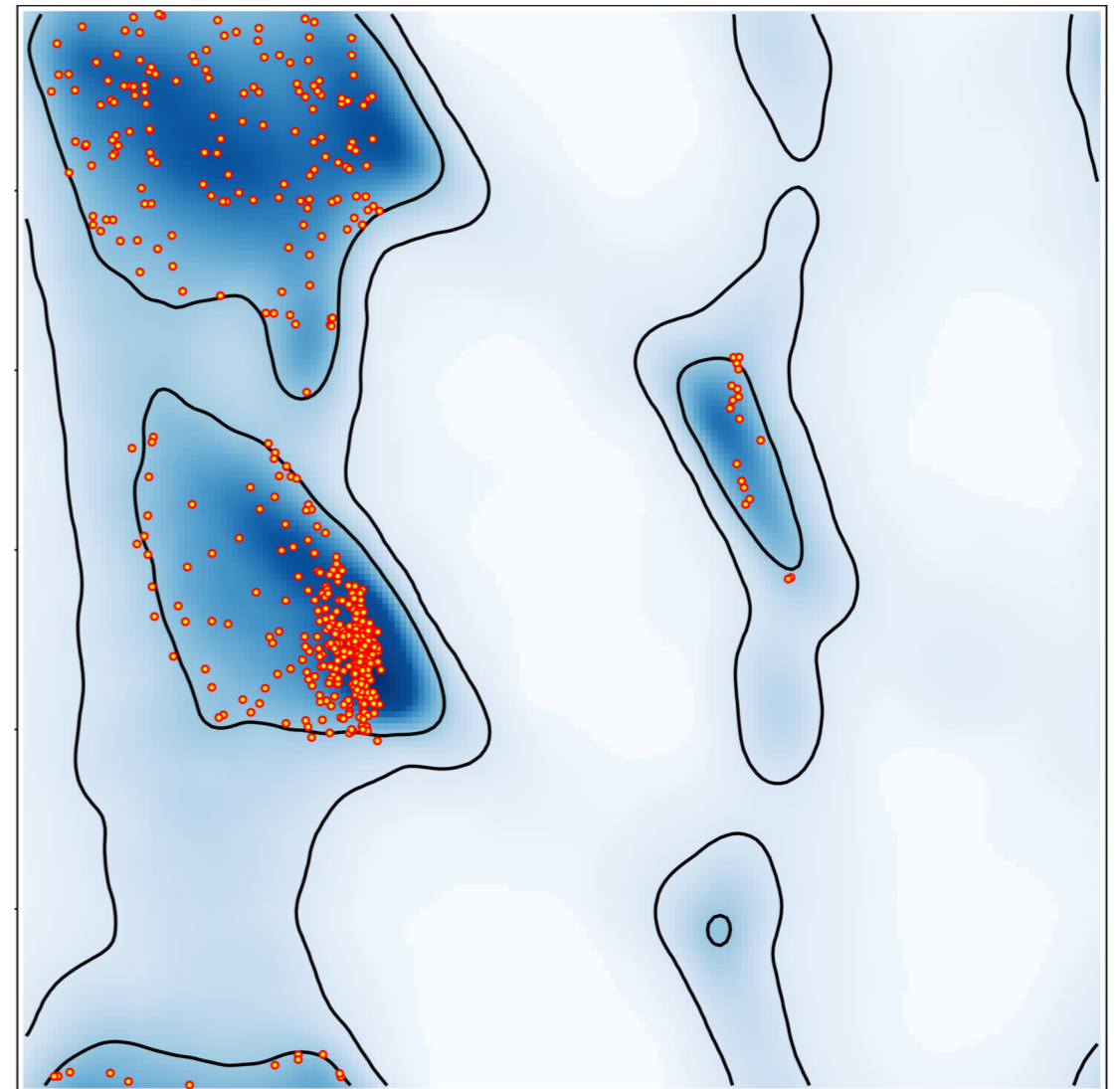
Two Ramachandran distributions

Model A

Favored	97.8 %
Allowed	1.95 %
Outliers	0.25 %

Model B

Favored	96.2 %
Allowed	3.8 %
Outliers	0.0 %

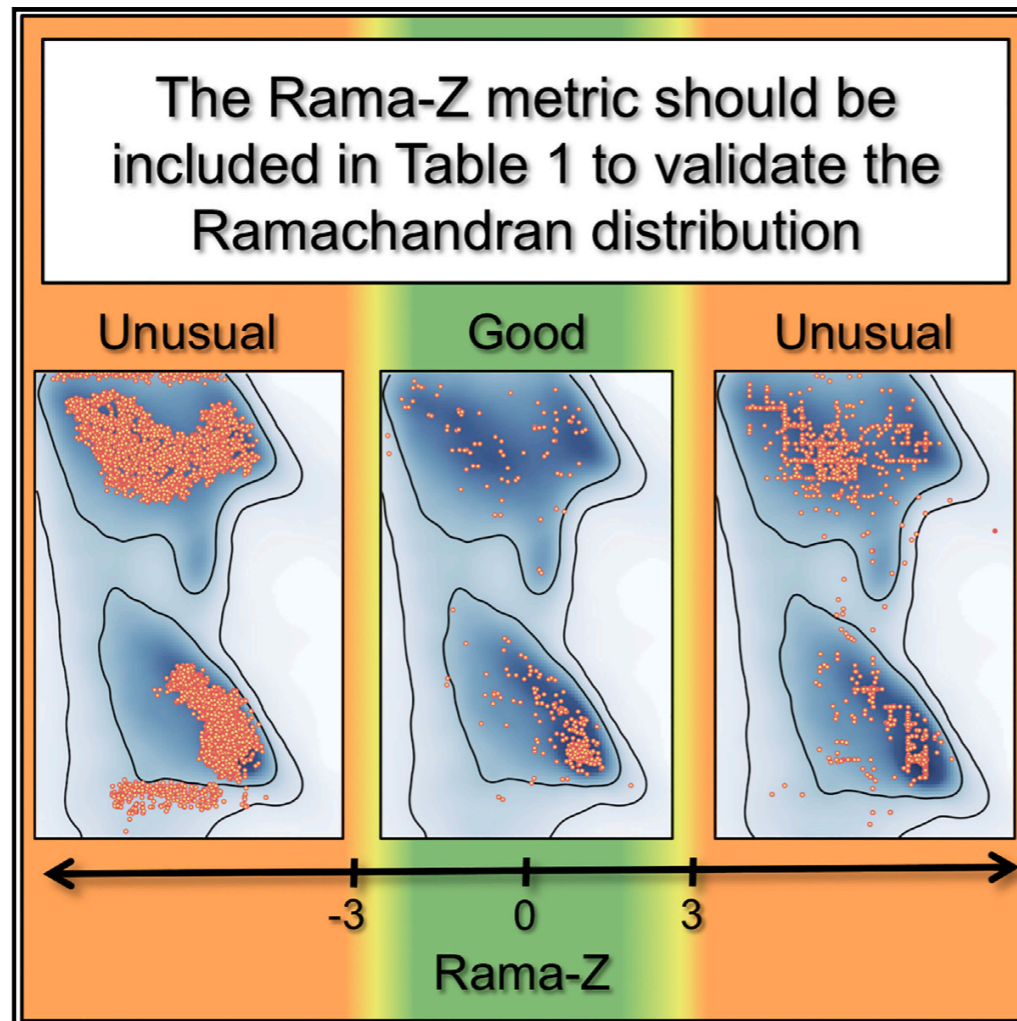# Global Ramachandran Score

A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry

Graphical Abstract

Authors

Oleg V. Sobolev, Pavel V. Afonine,
Nigel W. Moriarty,
Maarten L. Hekkelman,
Robbie P. Joosten,
Anastassis Perrakis, Paul D. Adams

Correspondence

osobolev@lbl.gov (O.V.S.),
r.joosten@nki.nl (R.P.J.)

In Brief

Counting the number of Ramachandran outliers is not sufficient for protein backbone validation. Sobolev et al. revisited the underutilized Ramachandran *Z* score. The authors describe its reimplementation in Phenix and PDB-REDO and showcase its utility. They advocate including it in the validation reports provided by the Protein Data Bank.

Global Ramachandran Score

Model A

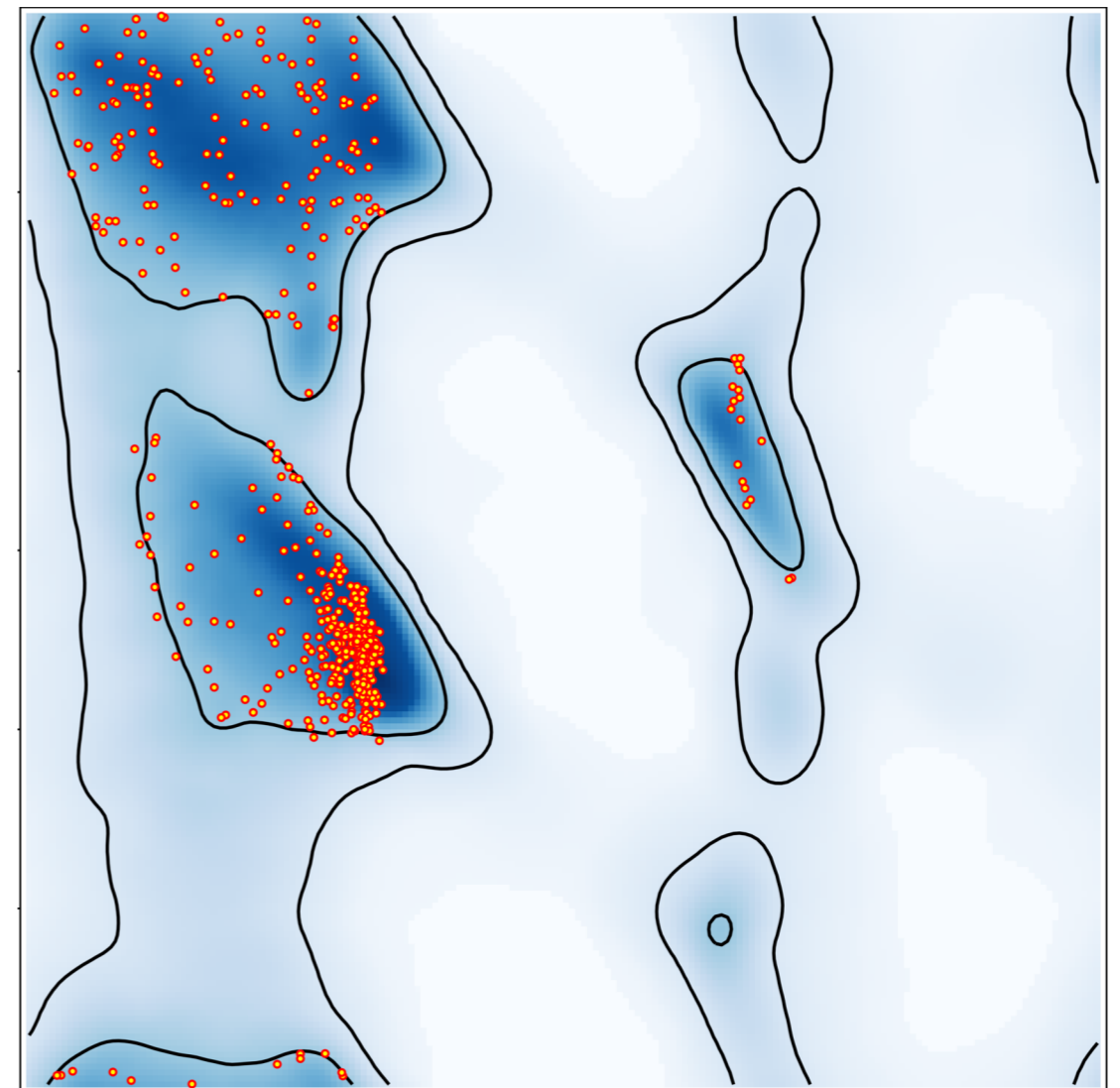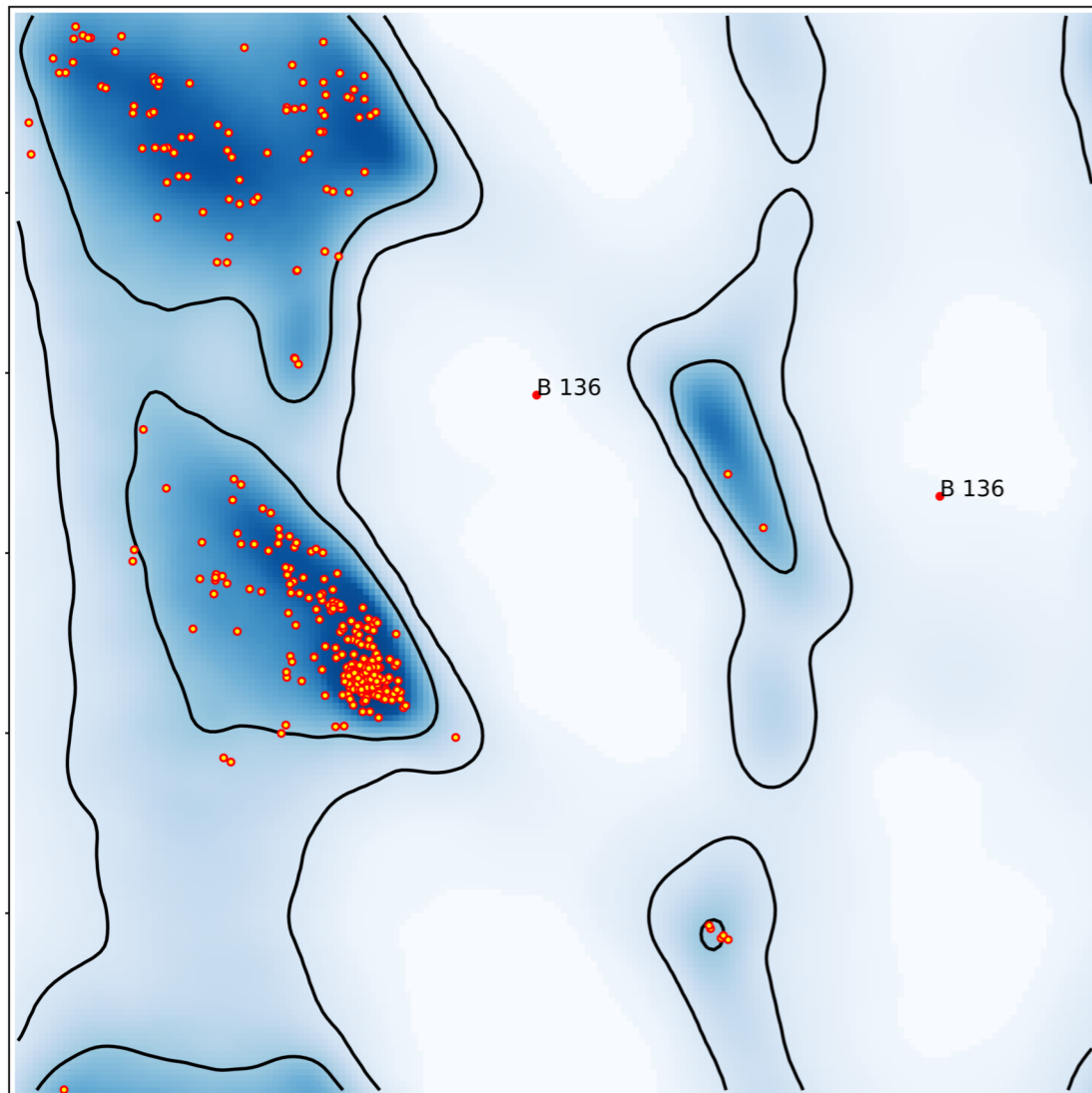| Favored	97.8
Allowed	1.95
Outliers	0.25

Rama z-score

-0.19   -4.08

Model B

Favored	96.2
Allowed	3.8
Outliers	0.0

Structure validation with H-bond parameters

Y   (A = O, N, S)

$R_{H...A}$

Θ_2

$R_{D...A}$ Θ_1 H

D   (D = O, N, S)

$1.4\text{Å} \leq R_{H...A} \leq 3.0\text{Å}$ $\Theta_1 \geq 120°$

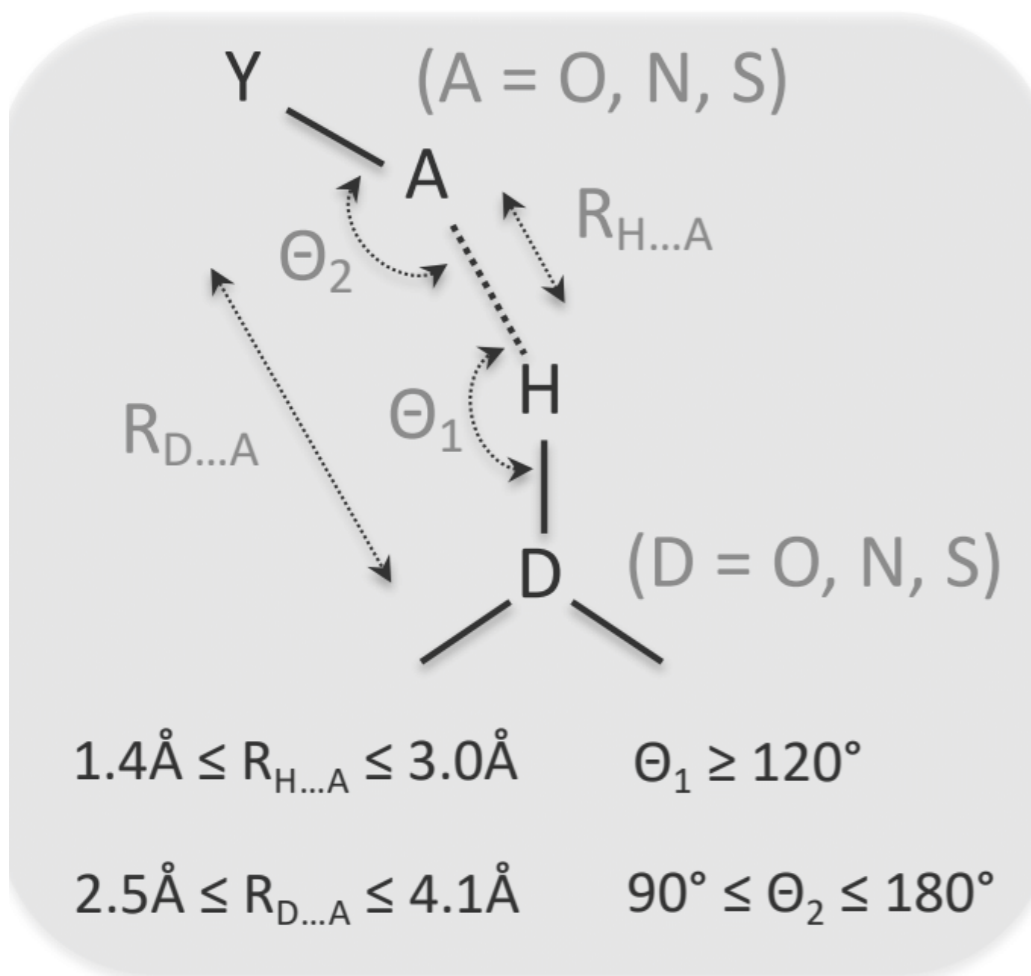$2.5\text{Å} \leq R_{D...A} \leq 4.1\text{Å}$ $90° \leq \Theta_2 \leq 180°$
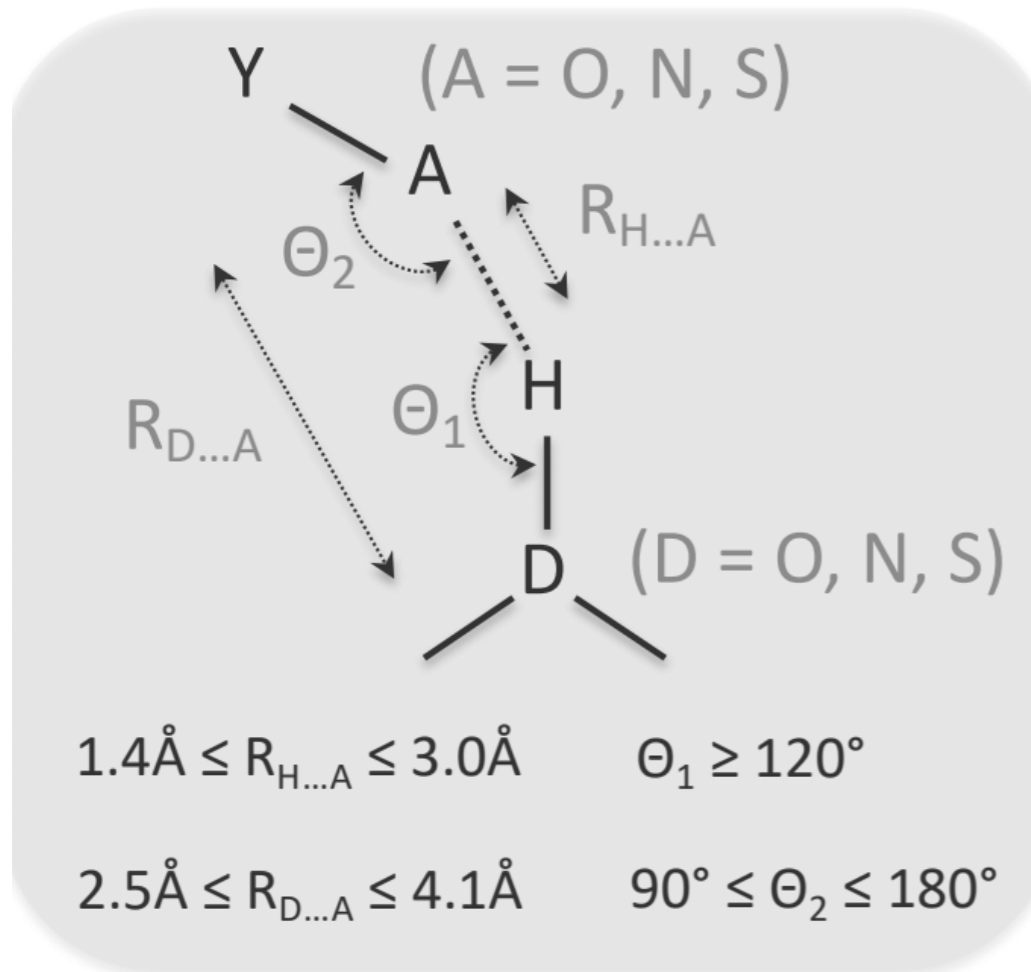
Validation tool based on H-bond parameter distribution of high-resolution PDB models.
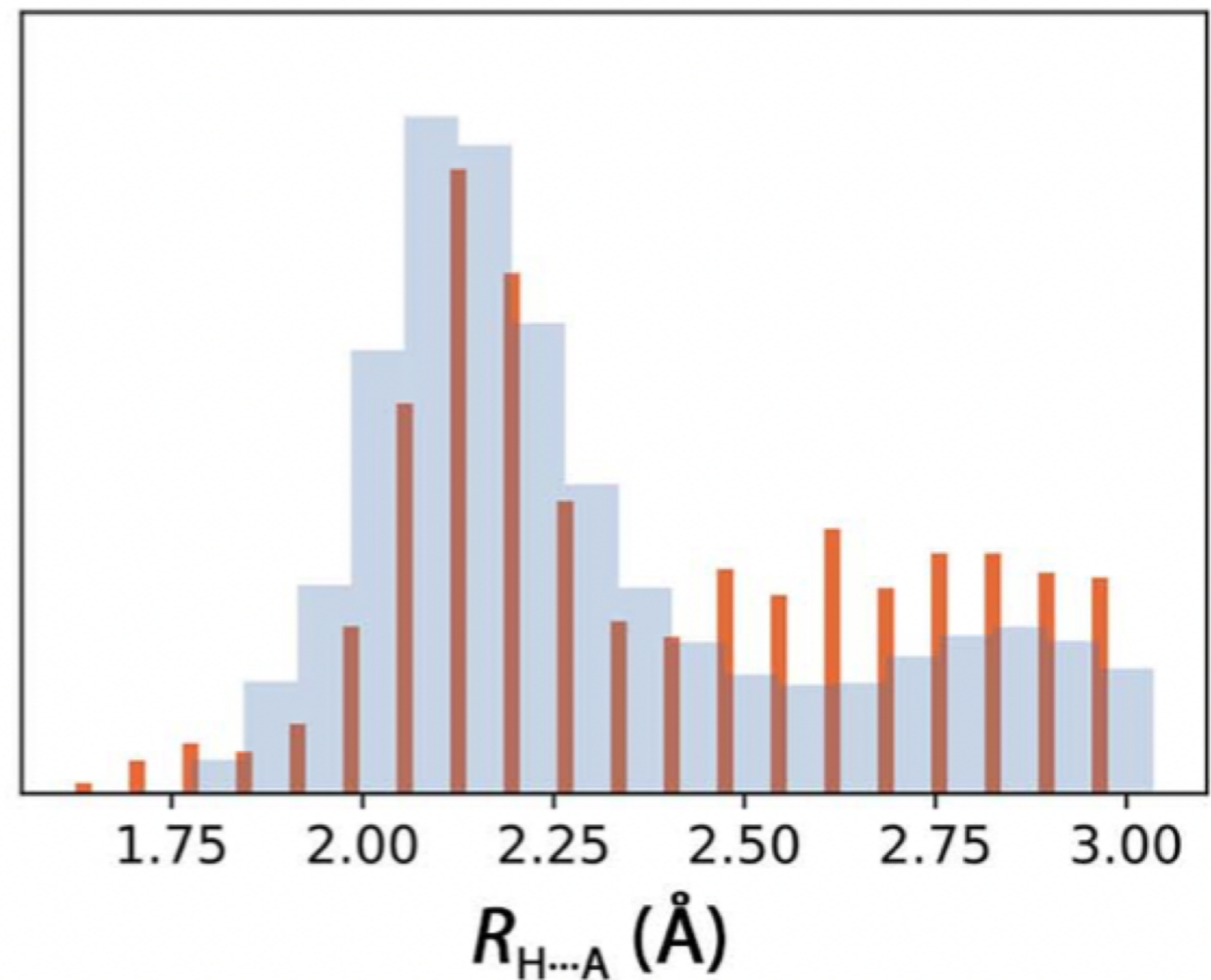
Grey: reference distribution

Red: 6mdo

$R_{H...A}$ (Å)

Afonine et al. (2023). *Acta Cryst. D***79**, 684–693

Structure validation with H-bond parameters

$(A = O, N, S)$

$R_{H\ldots A}$

$R_{D\ldots A}$

Θ_2

Θ_1

$(D = O, N, S)$

$1.4\text{Å} \leq R_{H\ldots A} \leq 3.0\text{Å}$ $\Theta_1 \geq 120°$

$2.5\text{Å} \leq R_{D\ldots A} \leq 4.1\text{Å}$ $90° \leq \Theta_2 \leq 180°$

A model not matching the tabulated distributions is not necessarily wrong
→ Inspect model to find explanation

Grey: reference distribution
Red: 5j1f

$R_{H\cdots A}$ (Å)

Afonine et al. (2023). *Acta Cryst. D***79**, 684–693
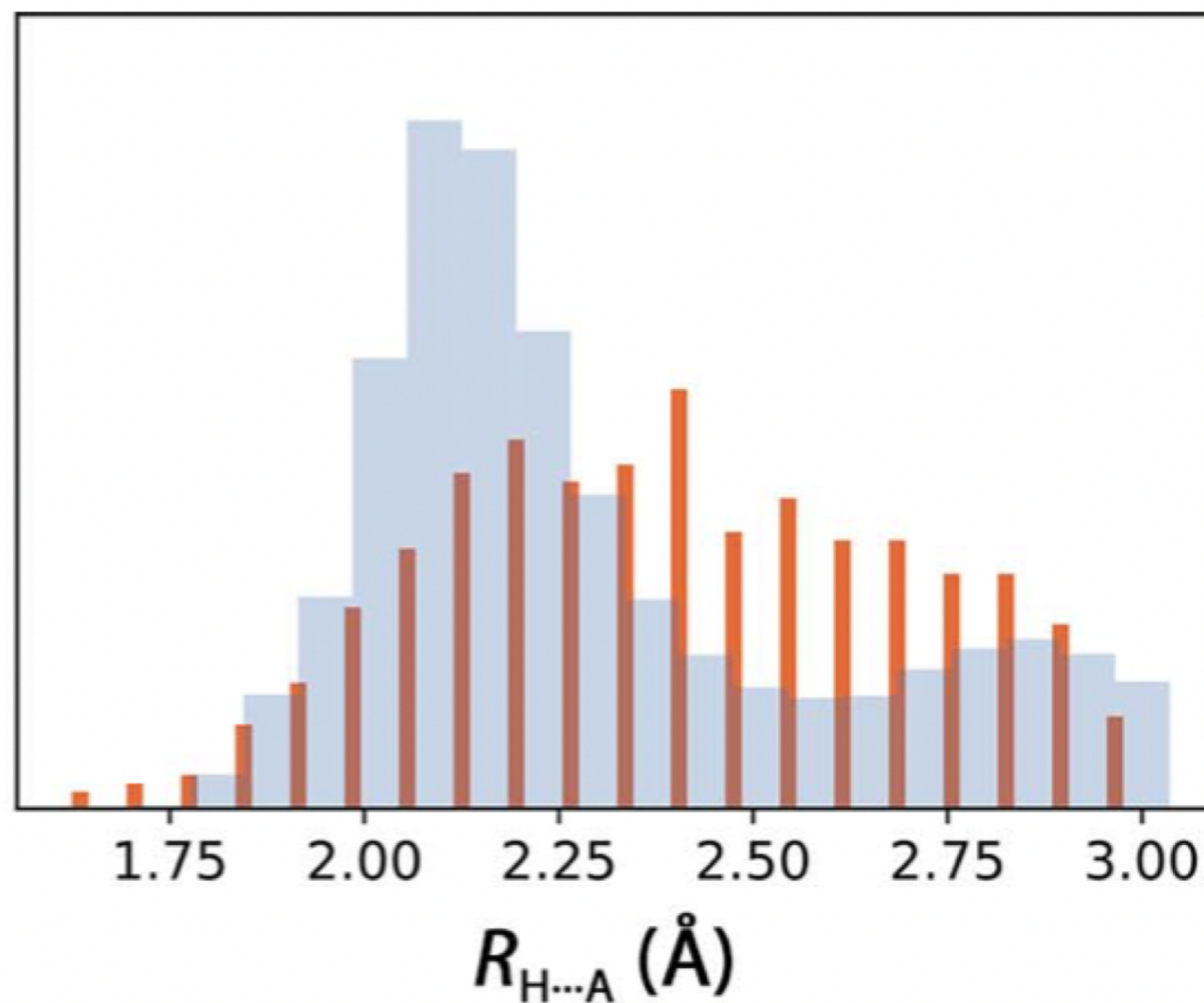
QMR – quantum mechanical restraints

- Ligands need restraints (description of chemical structure) for refinement.
- Restraint generators often ignore chemical variability & specific binding interactions.
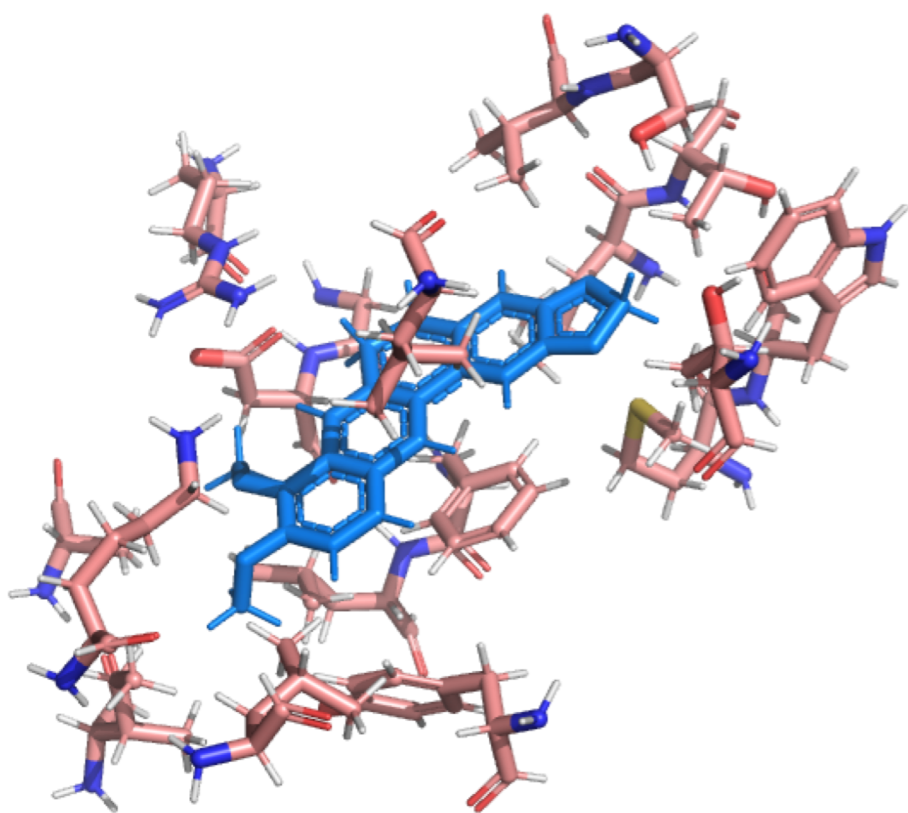

BER in 3vw2

QMR approach:
- Minimize the ligand geometry in the binding pocket.
- Use minimized ligand geometry as targets for restraints.

(forget about what happened with the residues)

Liebschner D., Moriarty, N.W., Poon, B. K. & Adams, P. D., In situ ligand restraints from quantum-mechanical methods. *Acta Cryst. D*. **79**, 100–110, (2023)

Run Phenix tools through ChimeraX

Local EM fitting
(EM placement)
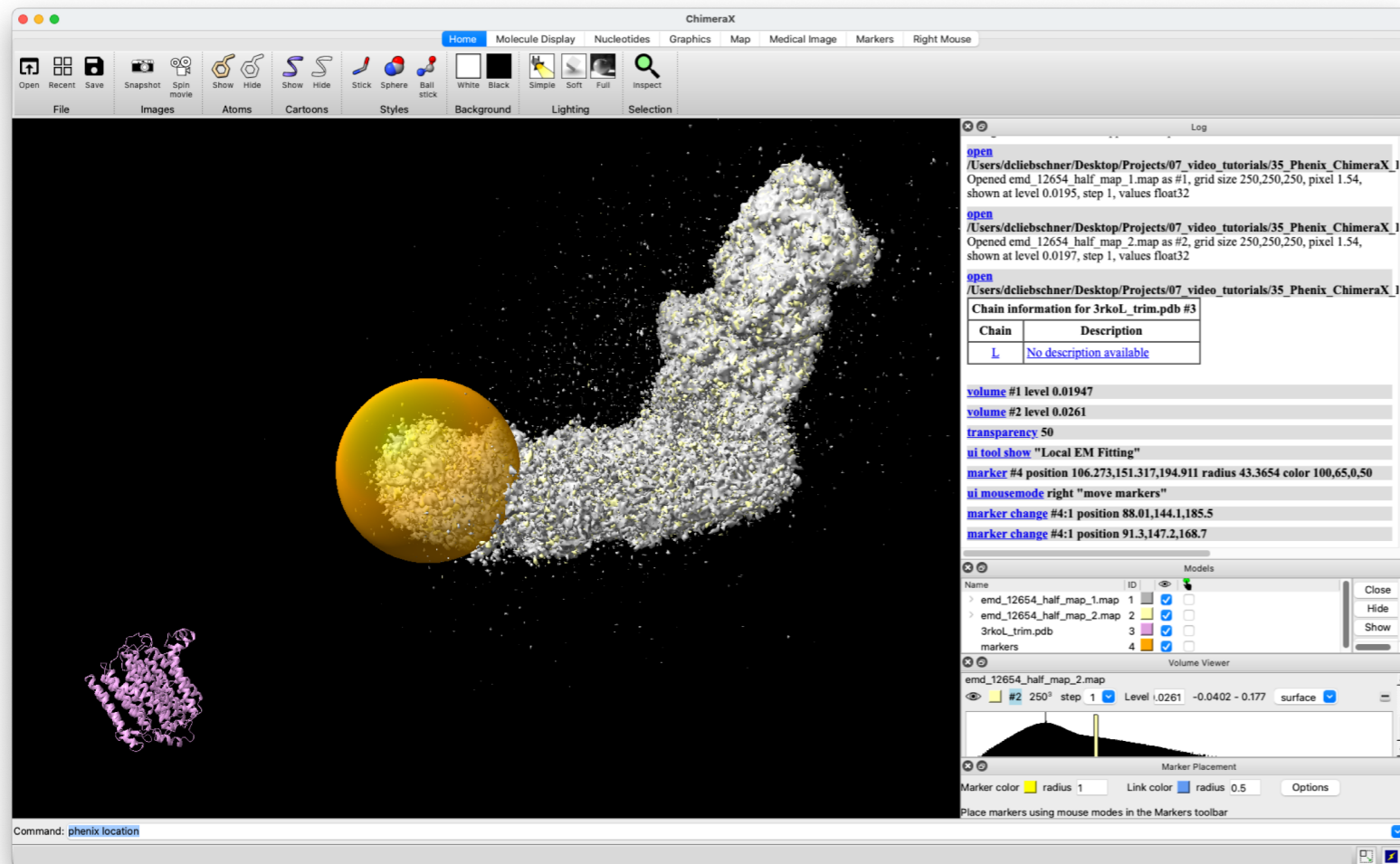
Automated water building
(phenix.douse)
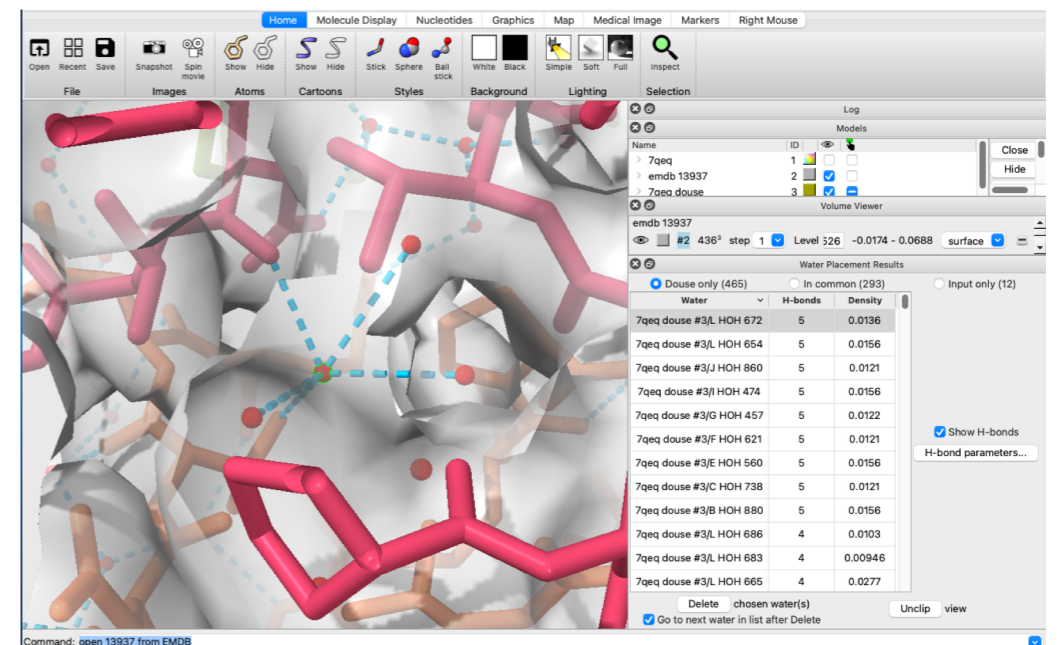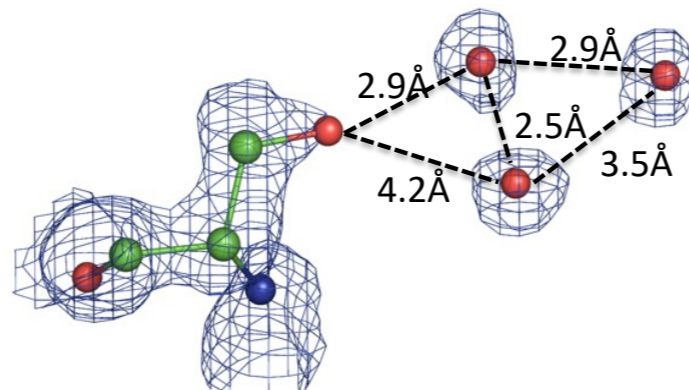
"Demo"

How to use predictions?

Incorporate predictions into the "conventional" structure determination workflow.

Millán C. et al.. Assessing the utility of CASP14 models for molecular replacement. Proteins. 2021 Dec;89(12):1752–69.
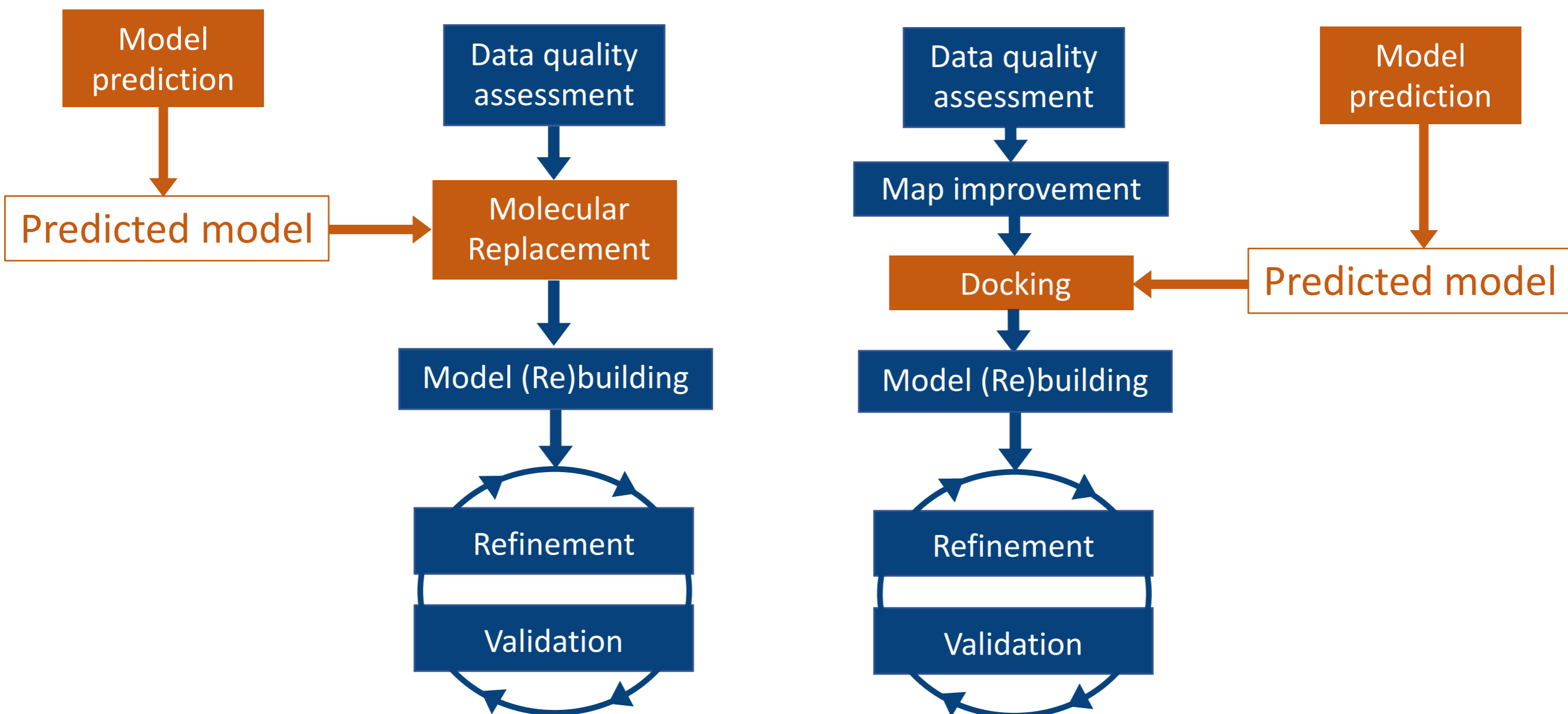
Incorporate predictions into the "conventional" structure determination workflow.
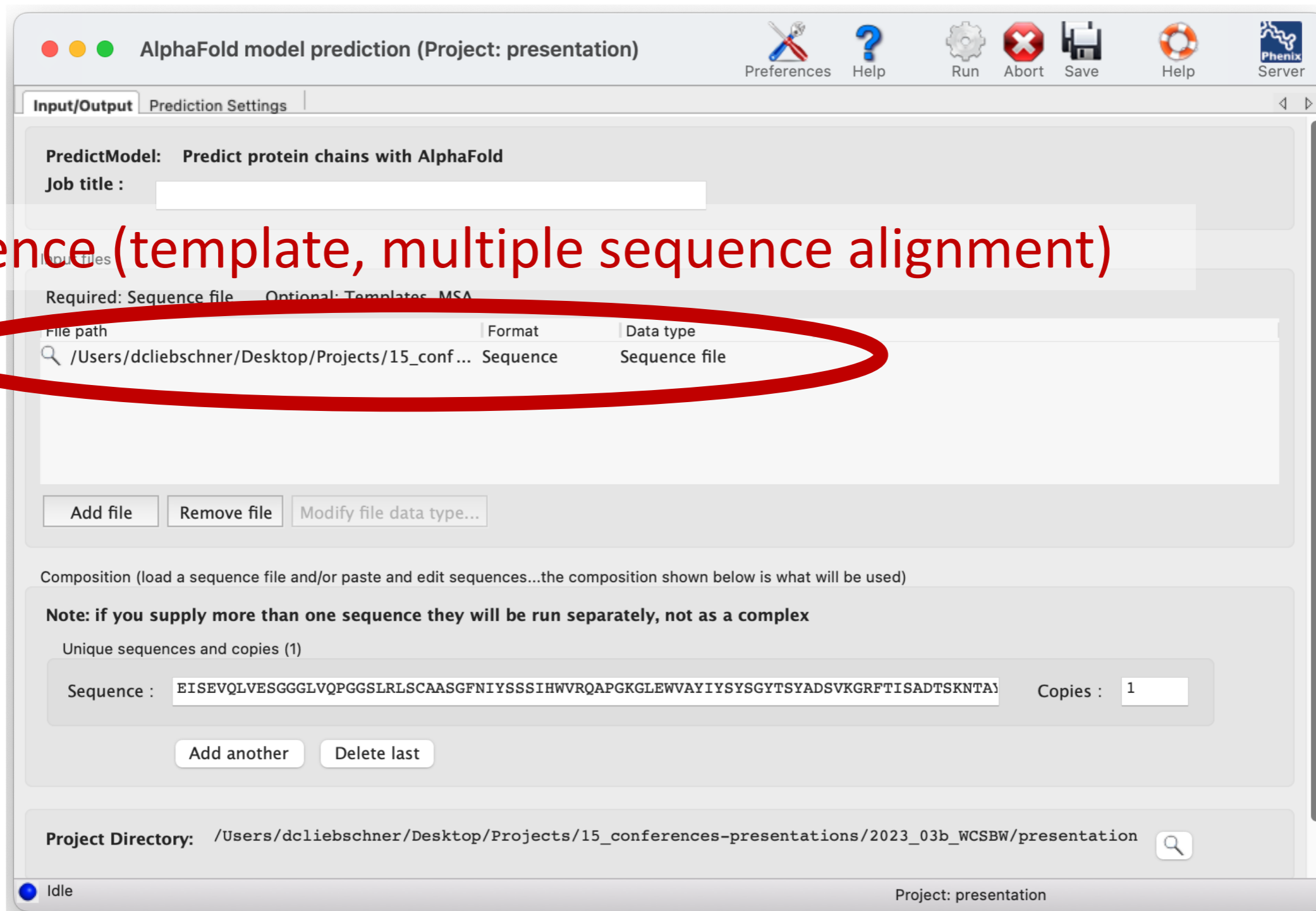
Example:
Diffraction data 7e9L (bovine protein of POMGNT2)

```
>POMGNT2_BOVINE soluble expression construct
GAPPAPALRIDYPKALQIL
TEGGTHMVCTGRTHTDRLCRFKWLCYSSEAEEFIFFHGNASVMLPSLGSRRFQPALLDLS
TVEDHNTQYFNFVELPAAALRFMPKPVFVPDVALIANRFNPDNLMHVFHDDLLPLFYTLR
QFPGLAREARLFFMEGWGEGAHFDLYKLLSPKQPLLRAQLKALGRLLCFSHAFVGLSKVT
TWYQYGFVQPQGPKANILVSGNEIRQFAHFLMEKLNVSQAGGPLGEEYILVFSRTQNRLI
LNEAELLLALAQEFQMKTVTVSLEDHAFADVVRLVSNASMLVSMHGAQLVTALFLPRGAA
VVELFPYAVNPDHYTPYKTLATLPGMDLQYIAWQNTMPENTVTHPERPWDQGGIAHLDRA
EQARILQSREVPRHLCCRNPEWLFRIYQDTKVDIPSLIQTIRRVVKGHPGPRKQKWTVSL
YPGKVREARCQASVQGASEARLSVSWQIPWNLKYLKVREVKYEVWLQEQGENTYVPYMLA
LQNHTFTENIKPFTTYLVWIRCIFNKTLLGPFADVLVCST
```

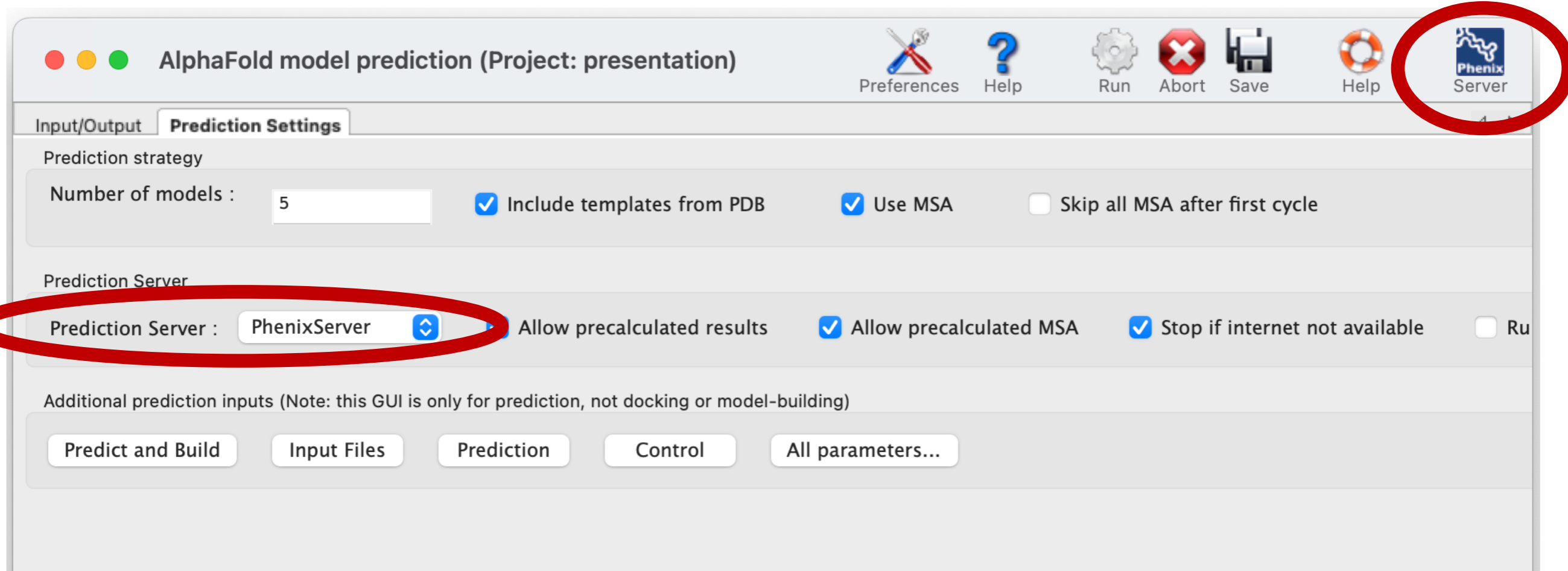Sequence (human protein)   Diffraction data (2.1 Å)
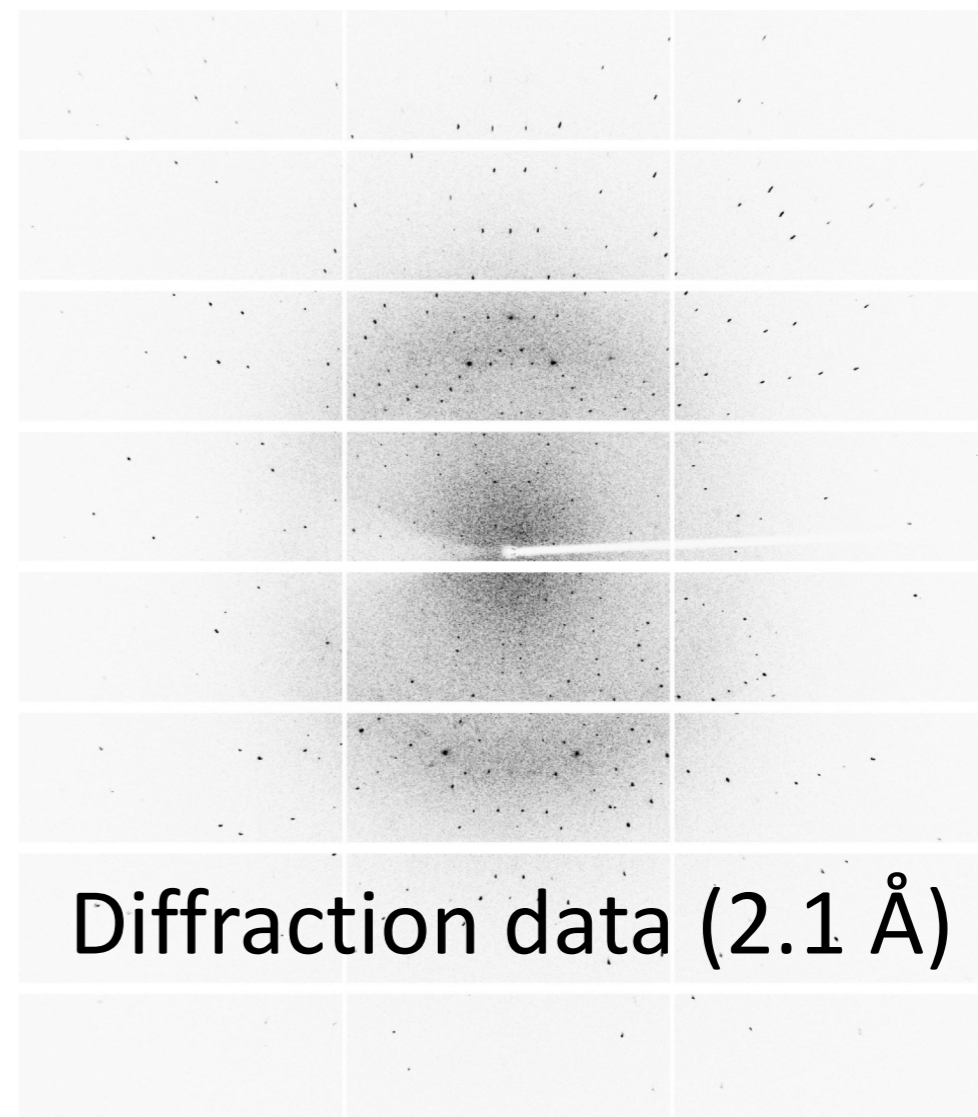
Predicting a model in Phenix

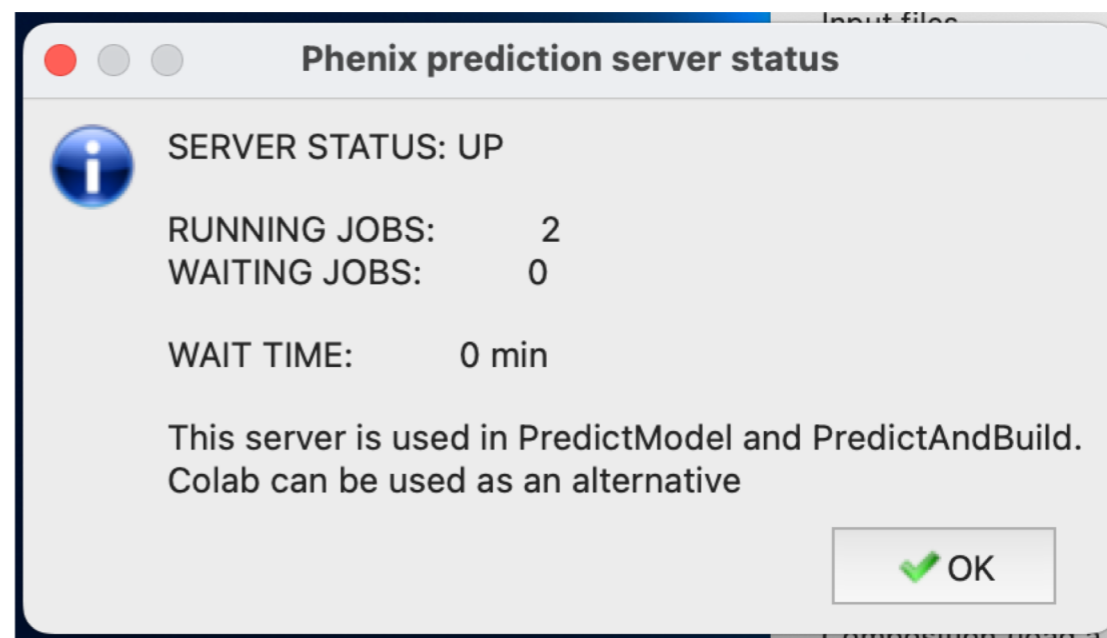**Sequence (template, multiple sequence alignment)**

Do I need to install AlphaFold?

Phenix Server for running AlphaFold


No need for a local AlphaFold installation

Get a prediction

sequence

>POMGNT2_BOVINE soluble expression construct
GAPPAPALRIDYPKALQIL
TEGGTHMVCTGRTHTDRLCRFKWLCYSSEAEEFIFFHGNASVMLPSLGSRRFQPALLDLS
TVEDHNTQYFNFVELPAAALRFMPKPVFVPDVALIANRFNPDNLMHVFHDDLLPLFYTLR
QFPGLAREARLFFMEGWGEGAHFDLYKLLSPKQPLLRAQLKALGRLLCFSHAFVGLSKVT
TWYQYGFVQPQGPKANILVSGNEIRQFAHFLMEKLNVSQAGGPLGEEYILVFSRTQNRLI
LNEAELLLALAQEFQMKTVTVSLEDHAFADVVRLVSNASMLVSMHGAQLVTALFLPRGAA
VVELFPYAVNPDHYTPYKTLATLPGMDLQYIAWQNTMPENTVTHPERPWDQGGIAHLDRA
EQARILQSREVPRHLCCRNPEWLFRIYQDTKVDIPSLIQTIRRVVKGHPGPRKQKWTVSL
YPGKVREARCQASVQGASEARLSVSWQIPWNLKYLKVREVKYEVWLQEQGENTYVPYMLA
LQNHTFTENIKPFTTYLVWIRCIFNKTLLGPFADVLVCST

EBI AlphaFold database
or
Run AlphaFold prediction

AlphaFold
model
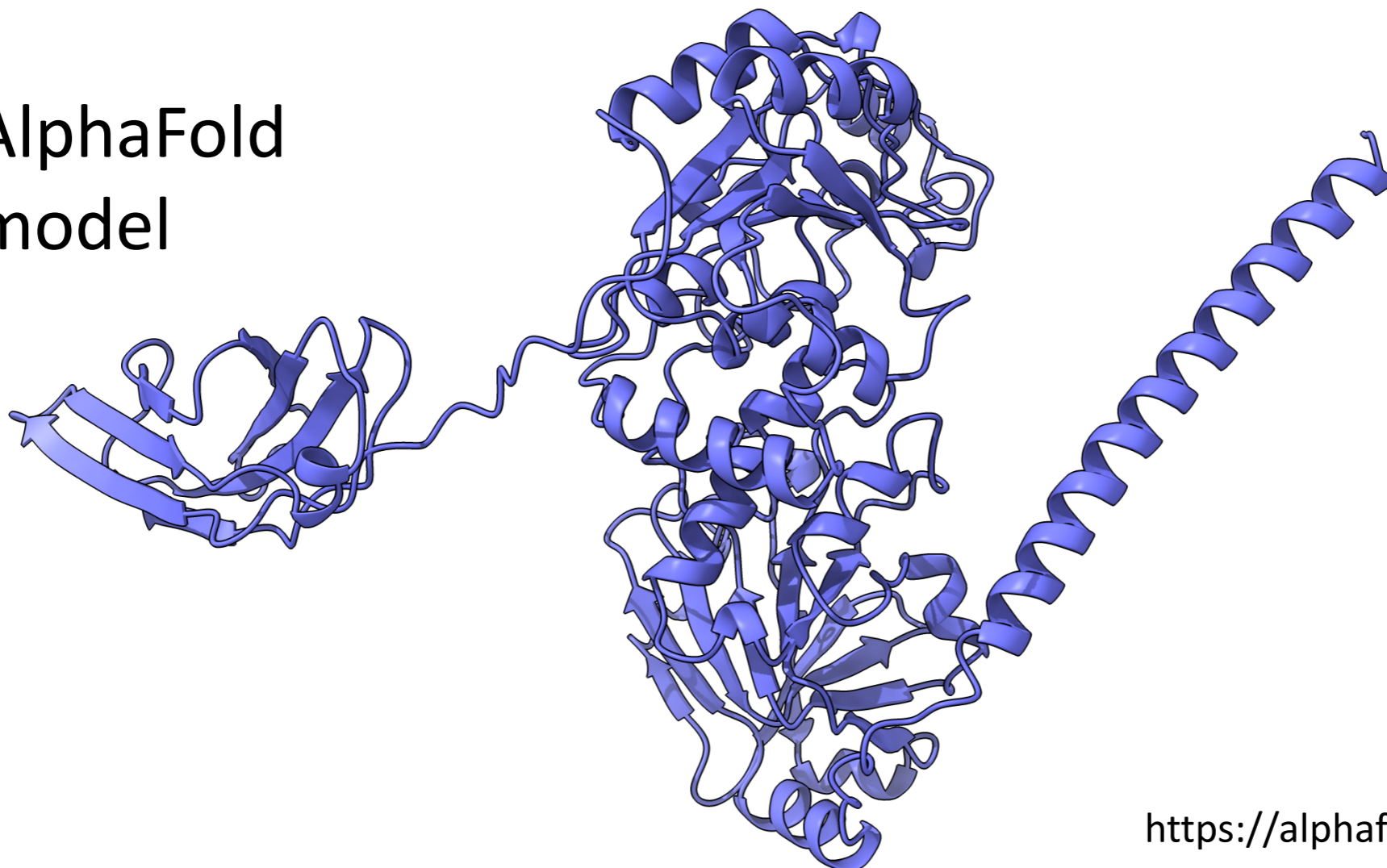
https://alphafold.ebi.ac.uk/entry/Q8NAT1

Process prediction

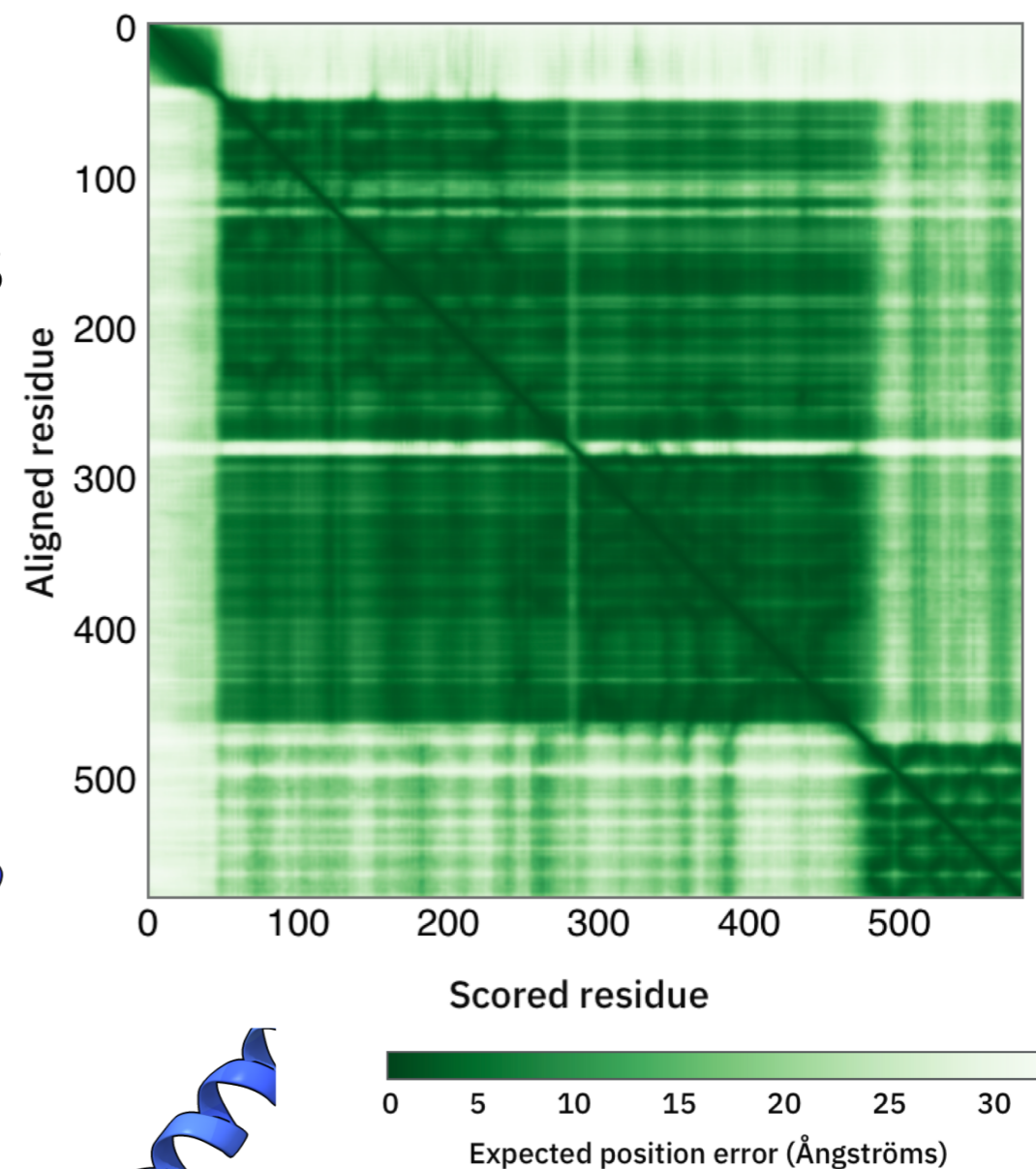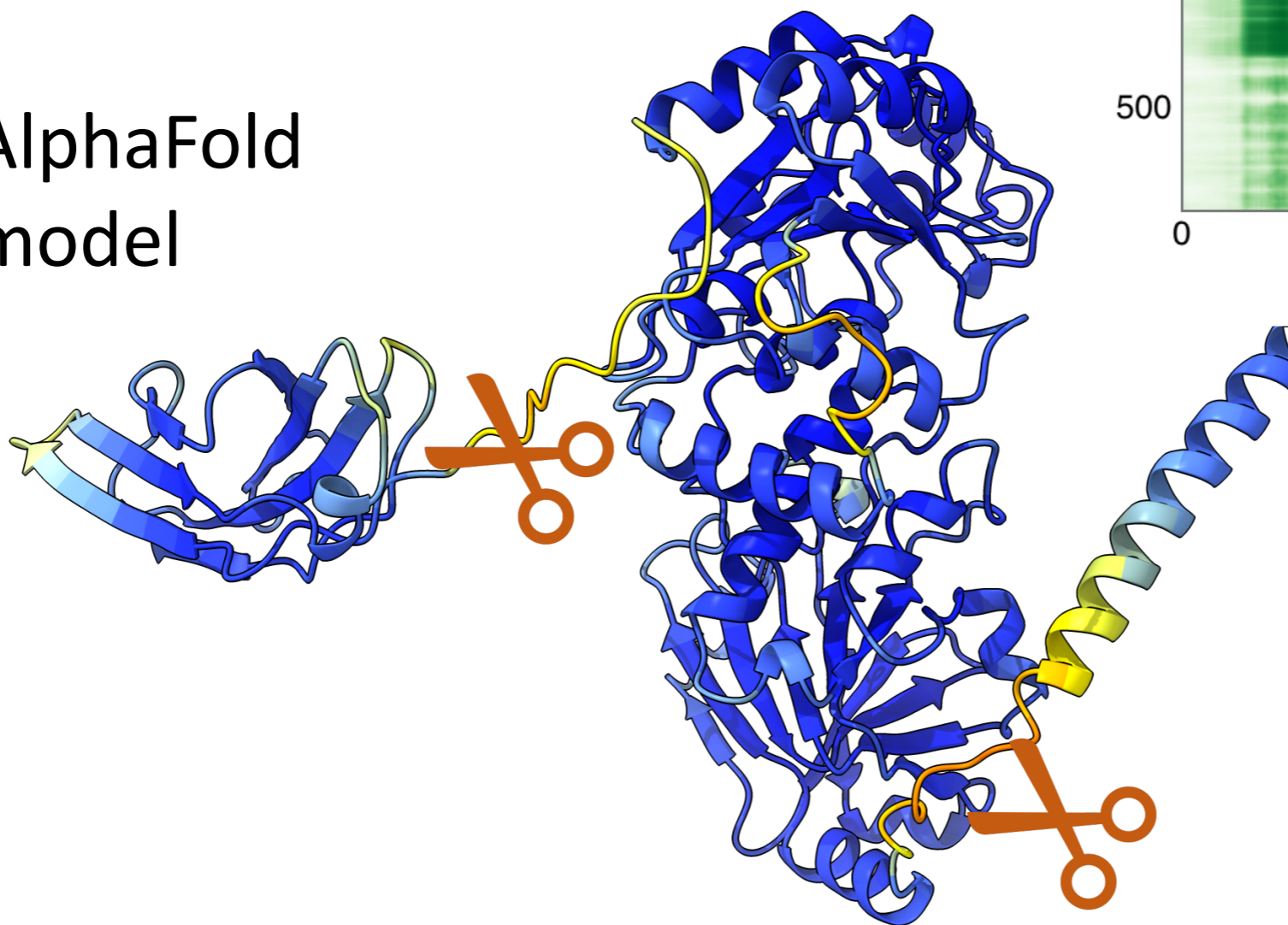- Remove low confidence parts.
- Break model into domains.

AlphaFold
model

Process prediction

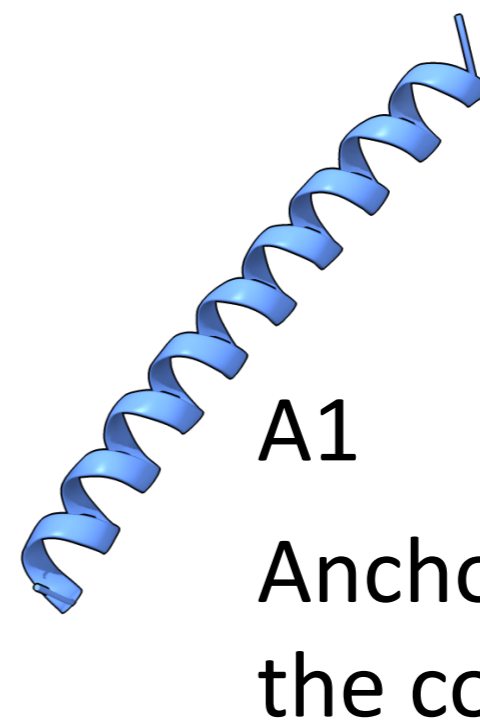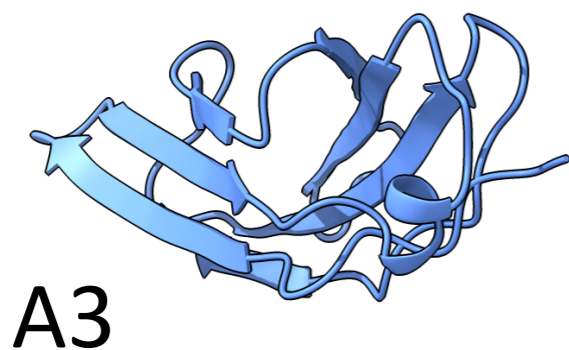- Remove low confidence parts
- Break model into domains.

AlphaFold
model

Process prediction

- Remove low confidence parts.
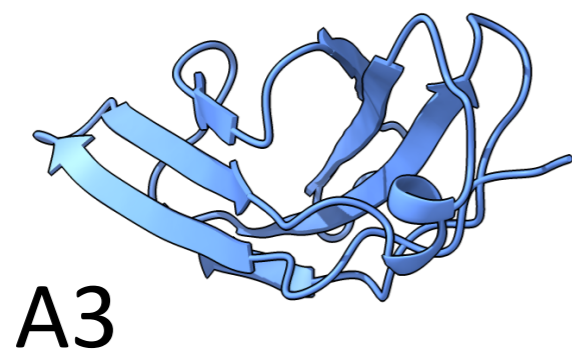- Break model into domains.

Processed AlphaFold model

A2

A3

A1

Anchor helix not in the construct

Process prediction

- Remove low confidence parts.
- Break model into domains.

Processed
AlphaFold
model

A2

A3
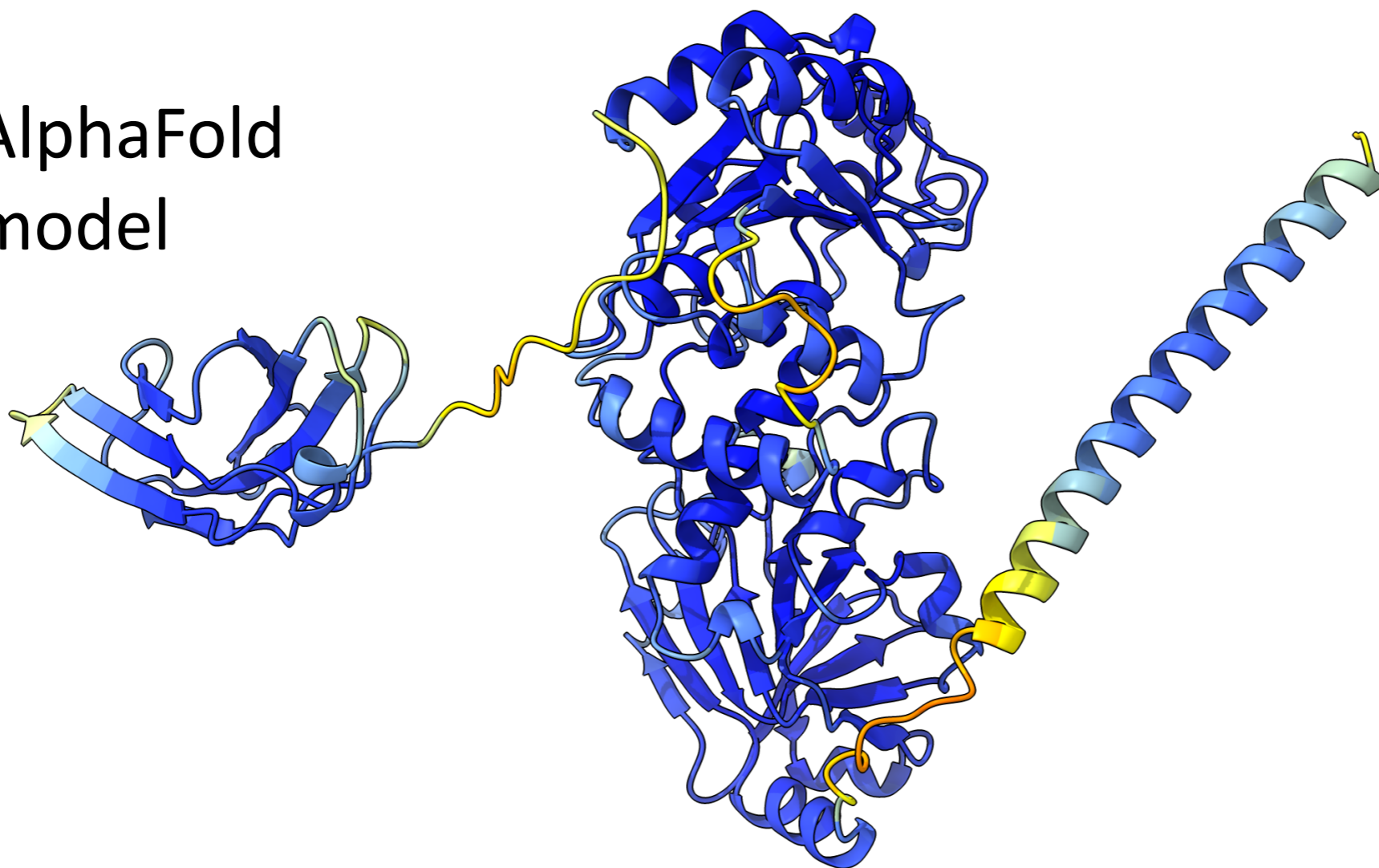
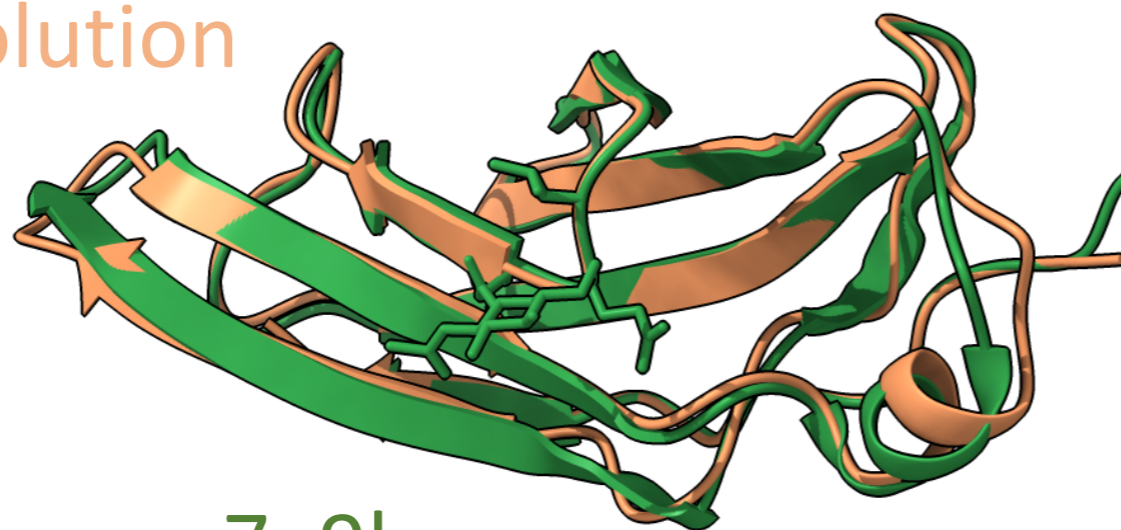Do MR search with 2
components

Molecular replacement

Phaser has found 1 MR solution(s).

Top LLG:	4528.201
Top TFZ:	34.6
Spacegroup:	P 21 21 21

MR solution

7e9l

Molecular replacement

Phaser has found 1 MR solution(s).

Top LLG: 1884.960
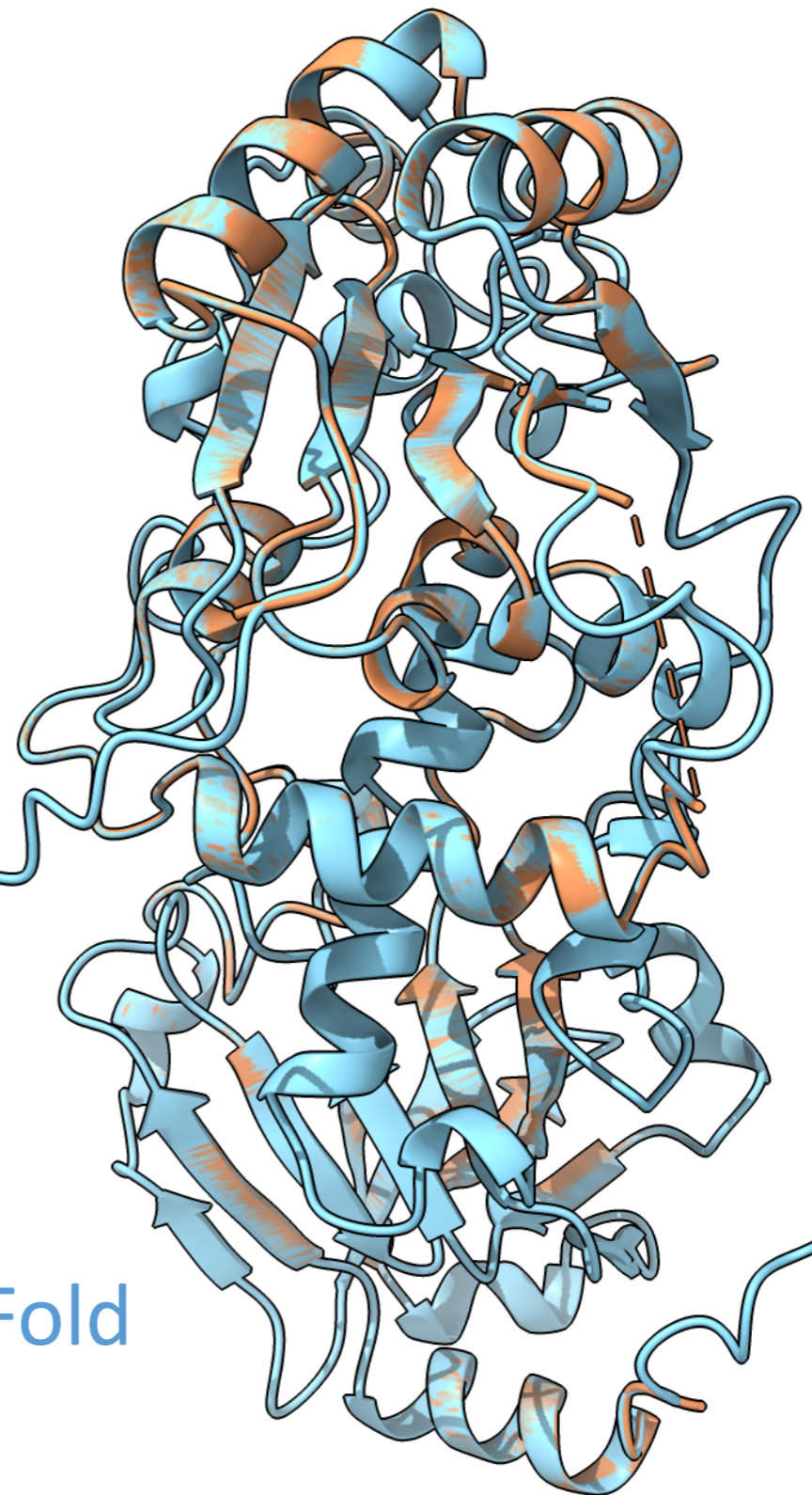
Top TFZ: 39.1

Spacegroup: P 21 21 21

If using AF2 model "as is"
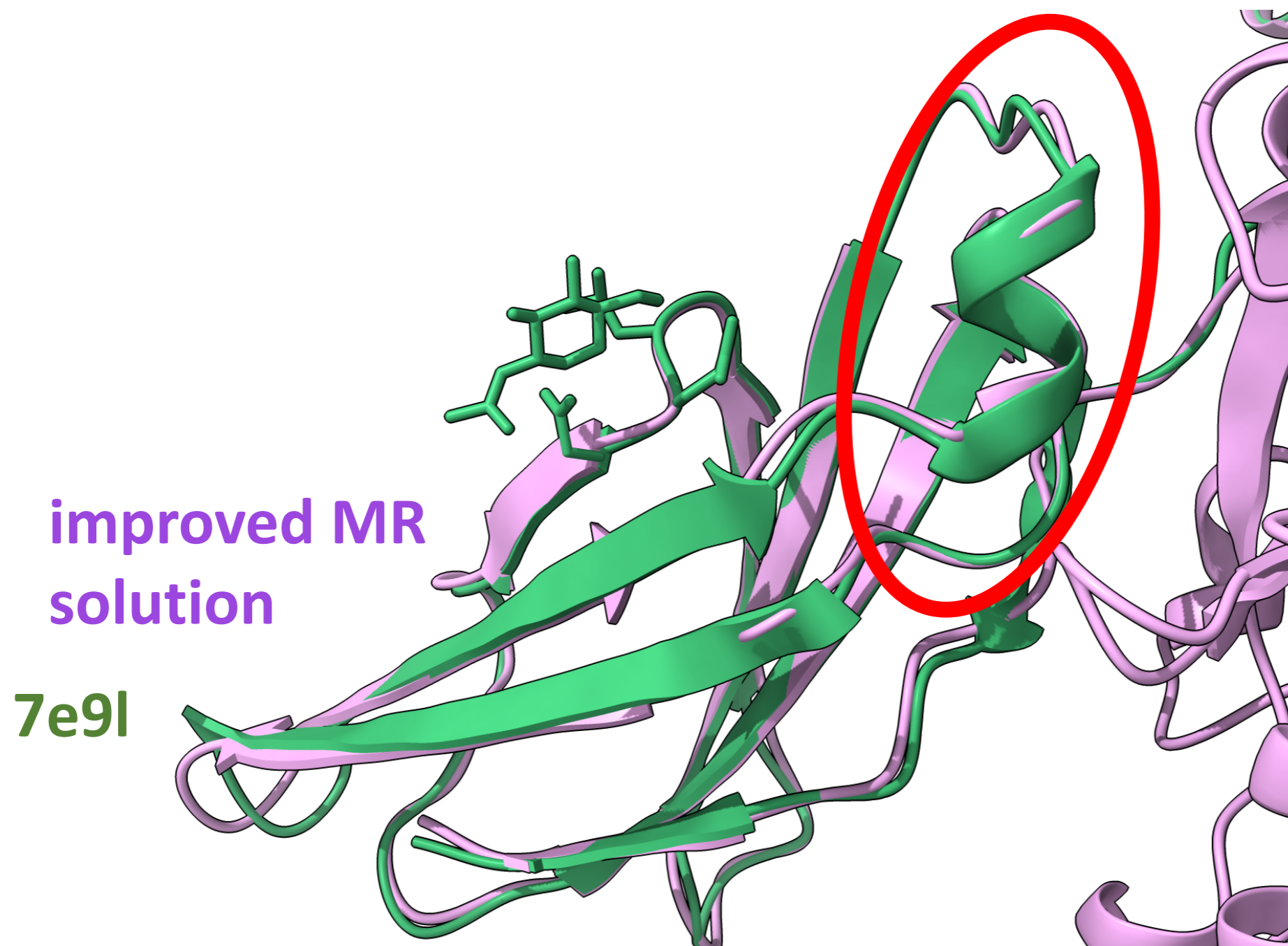
MR solution
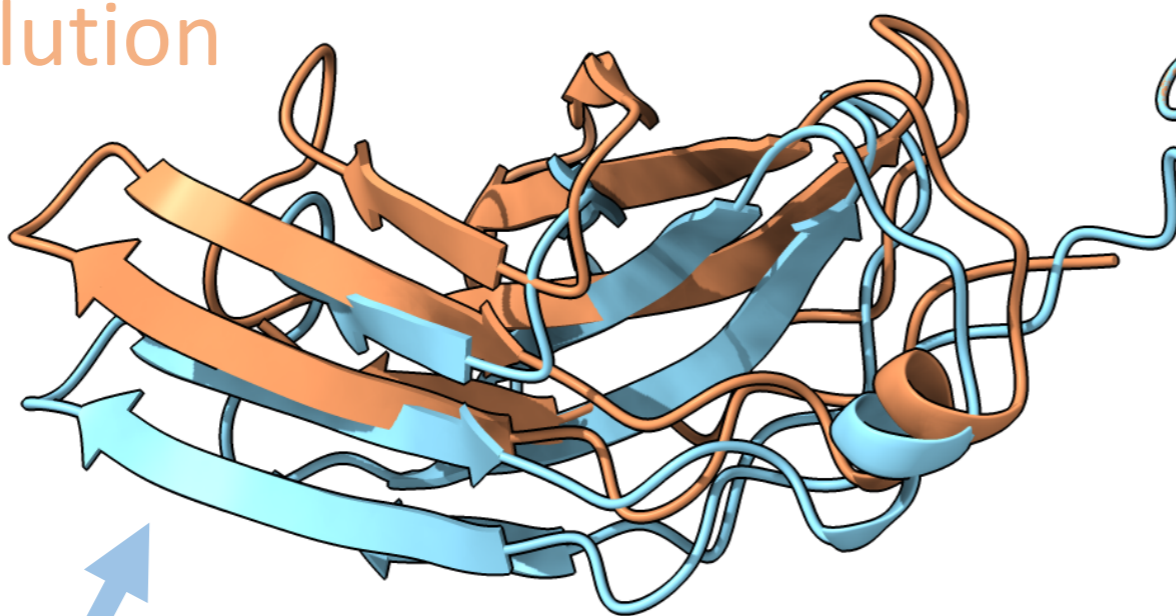
Domain is shifted
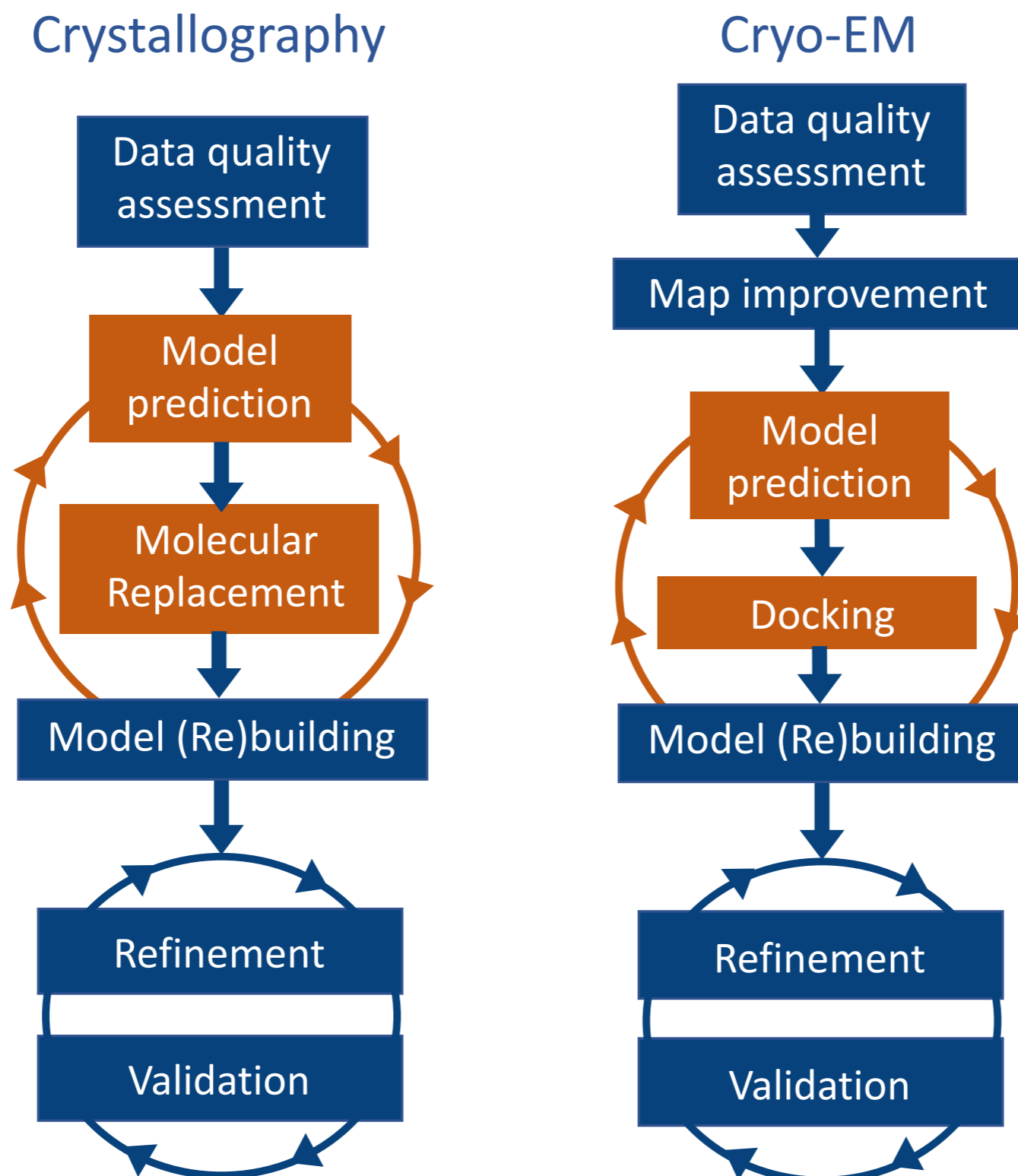
AlphaFold model

Automatically rebuild fragments

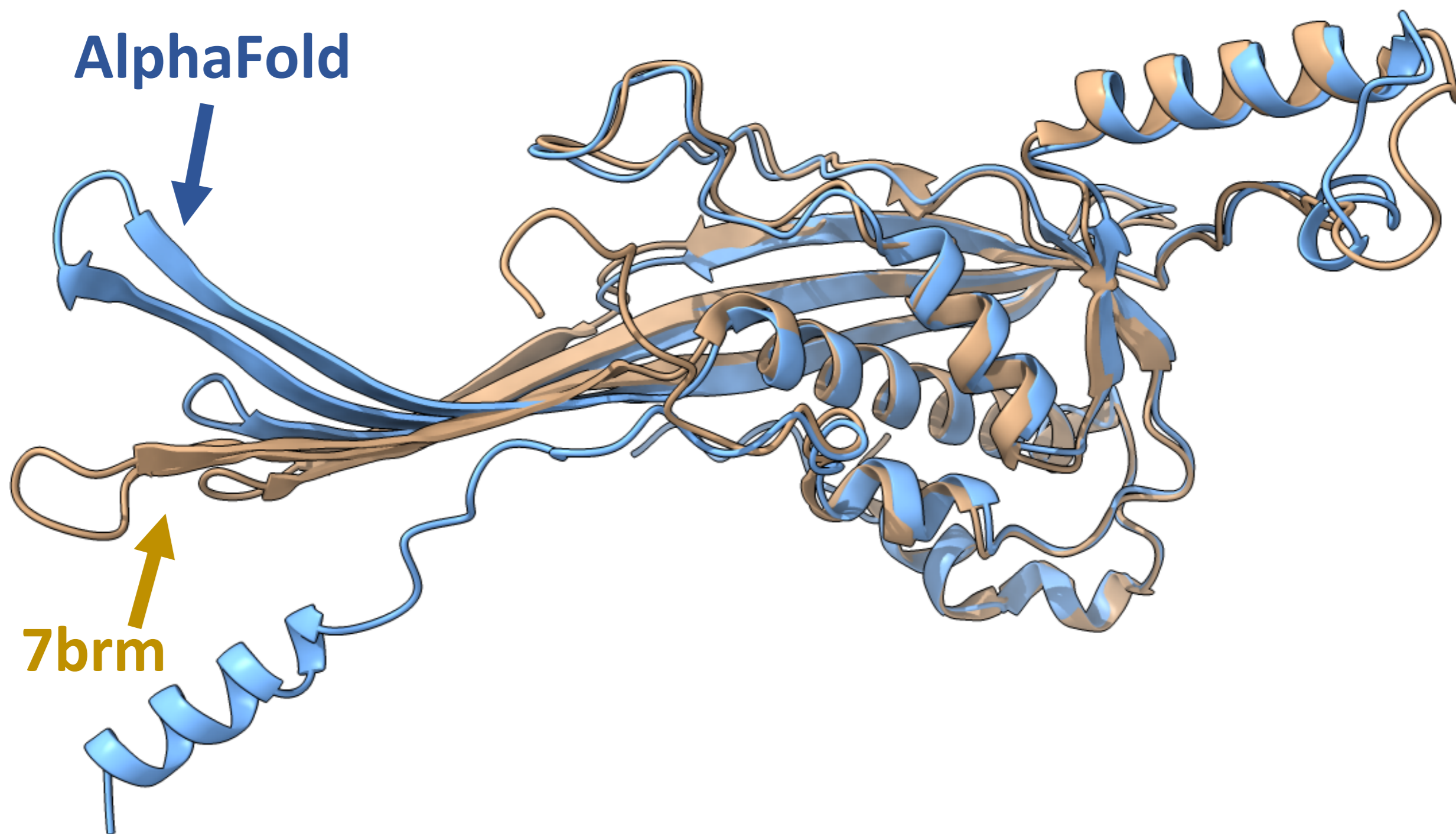Fix deviating part automatically with phenix.fit_loops.
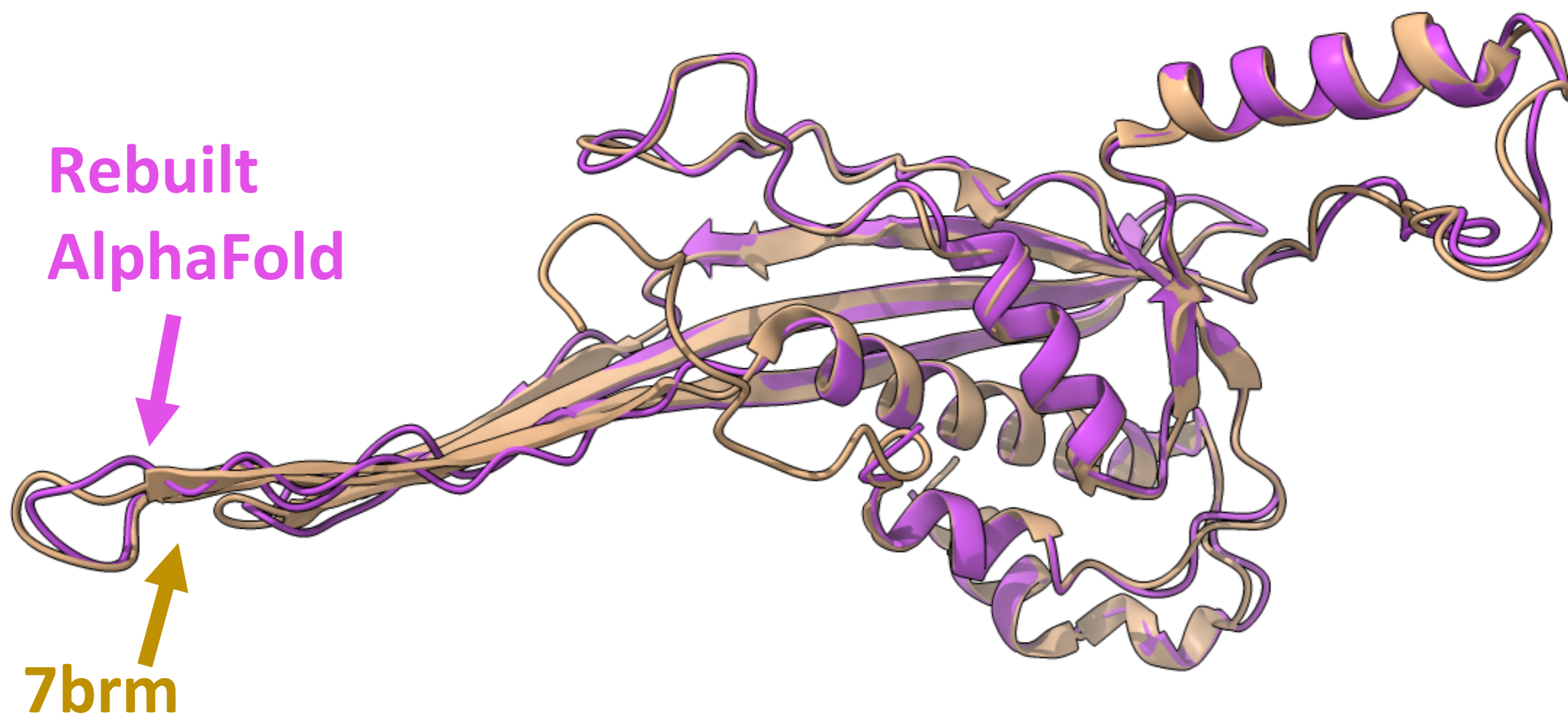
Automatically rebuild fragments

improved MR solution

7e9l

Using predicted models

New approach: Iterate prediction and model building

Improving prediction (cryo-EM)

Improving prediction

Rebuilt AlphaFold
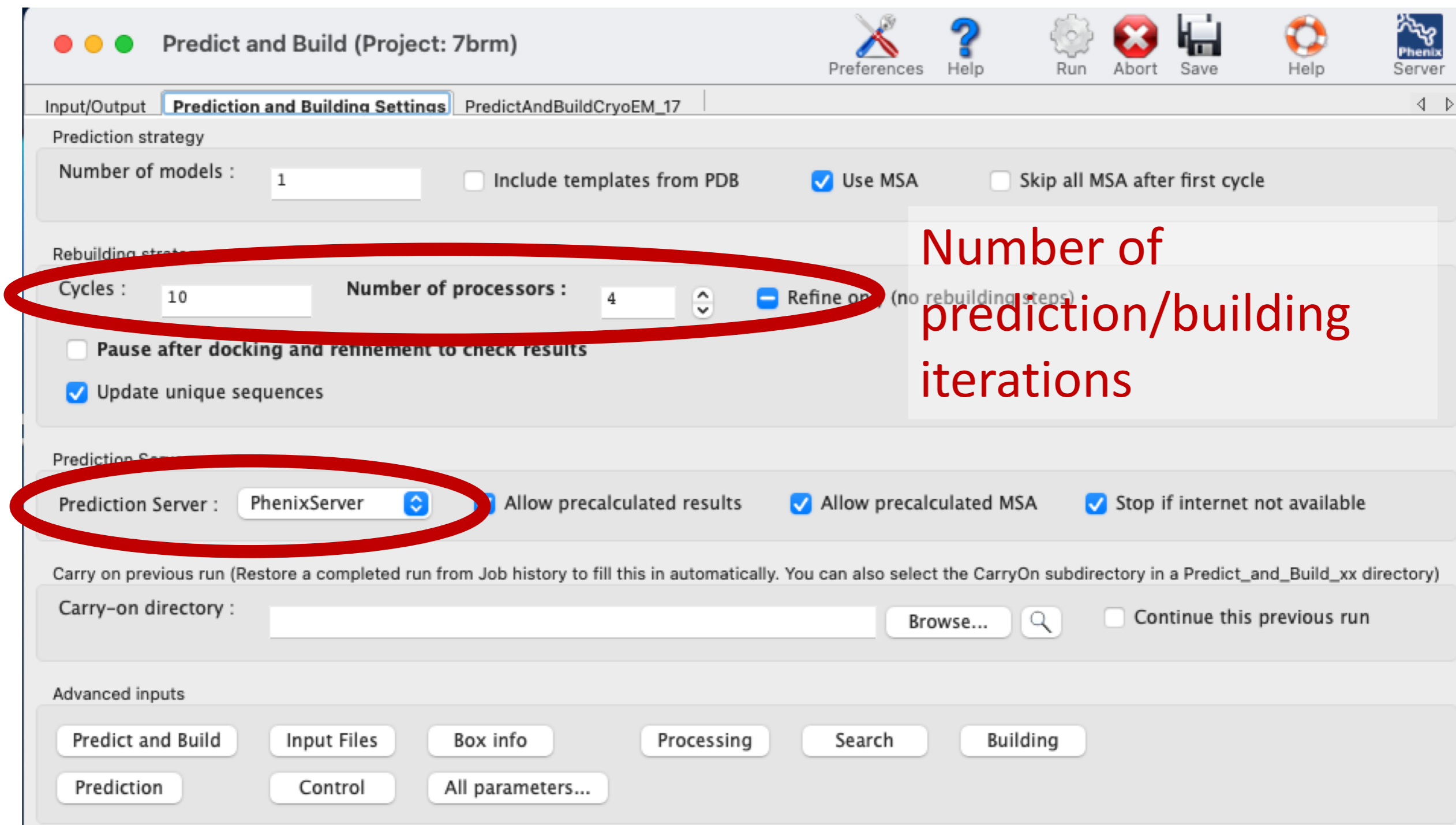
7brm

What happens if we use the rebuilt model as template for AlphaFold?

Improving prediction

Sequence

Sequence
Template (rebuilt model)

AlphaFold

Good sheet

Bad sheet

Better sheet
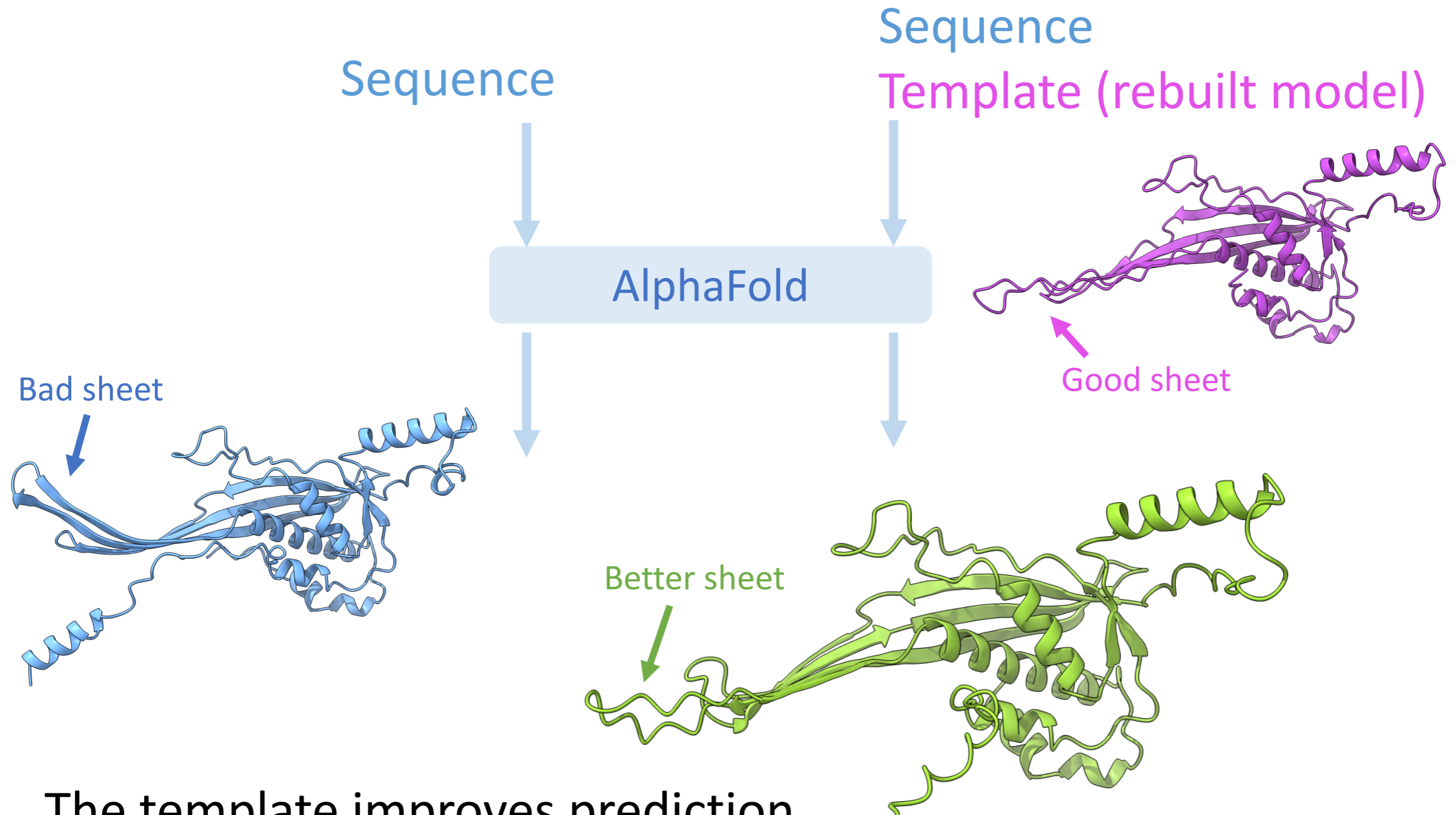
- The template improves prediction
- New prediction can be even better than the template

Iterate with Predict and Build

Talks of Phenix team members at the IUCr 2023

Wednesday, August 23rd

- 9:00 am, Tom Terwilliger, *AlphaFold changes everything (and nothing)* (Keynote 1)

- 1:30 pm, Oleg Sobolev, Global Ramachandran Score (A020)

- 2:40 pm, Randy Read, Likelihood-based docking (A020)

Sunday, August 27th

- 5:55 pm, Christopher Williams, *Identification and Valdiation of low-pLDDT regions in AF2 predictions* (A011)

Monday, August 28th

- 9:00 am, Airlie McCoy, Trekkin' through reciprocal space with Phaser (Keynote 31)

- 2:20 pm, Dorothee Liebschner, *Using the PDB and EMDB for testing new algorithms* (A023)