



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 **E**Nvironment for **I**ntegrated **X**tallography

The 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

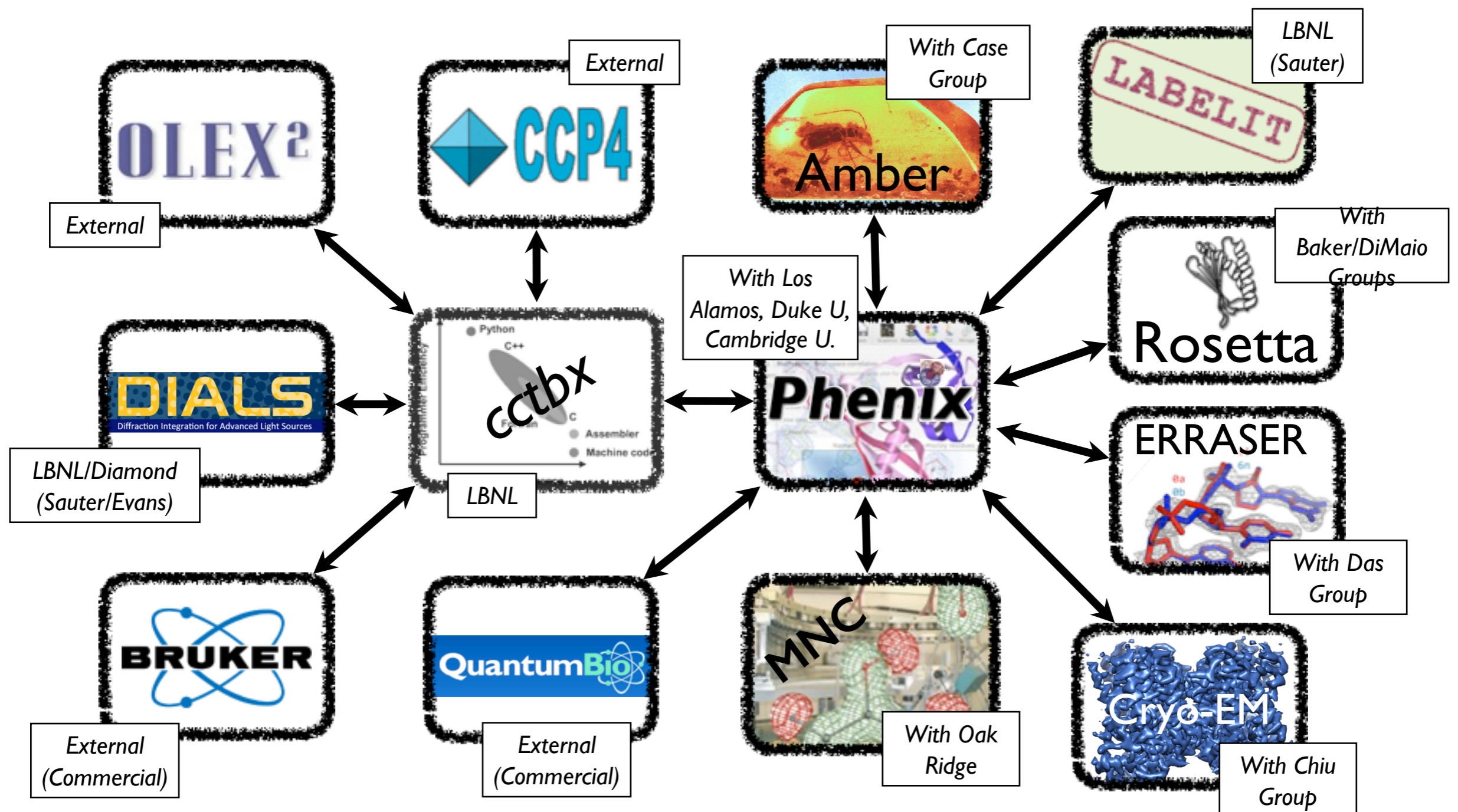


An NIH/NIGMS funded
Program Project

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

Phenix - a Structural Biology Hub

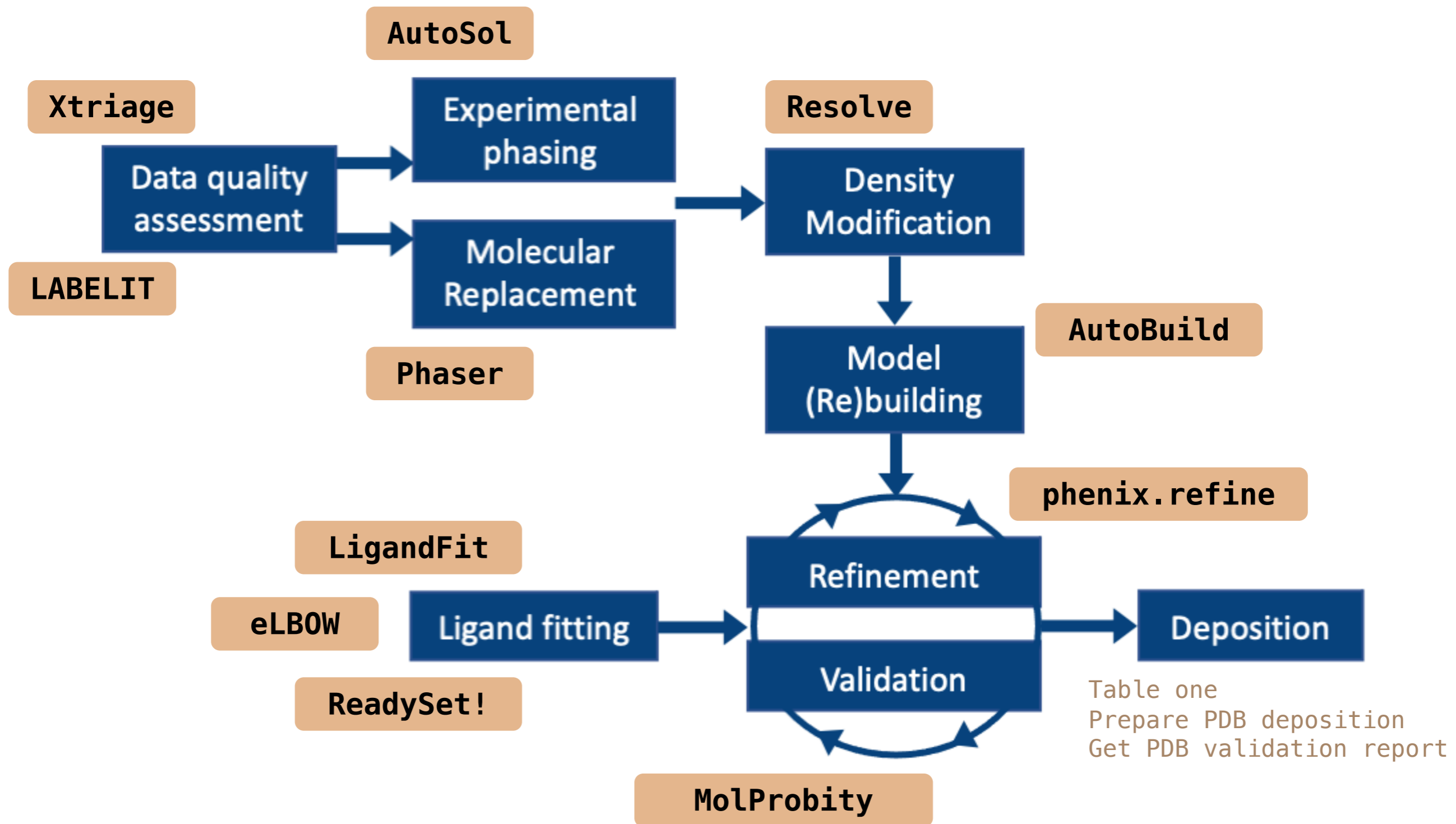
We have nucleated the development of new computational methods for structural biology



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**

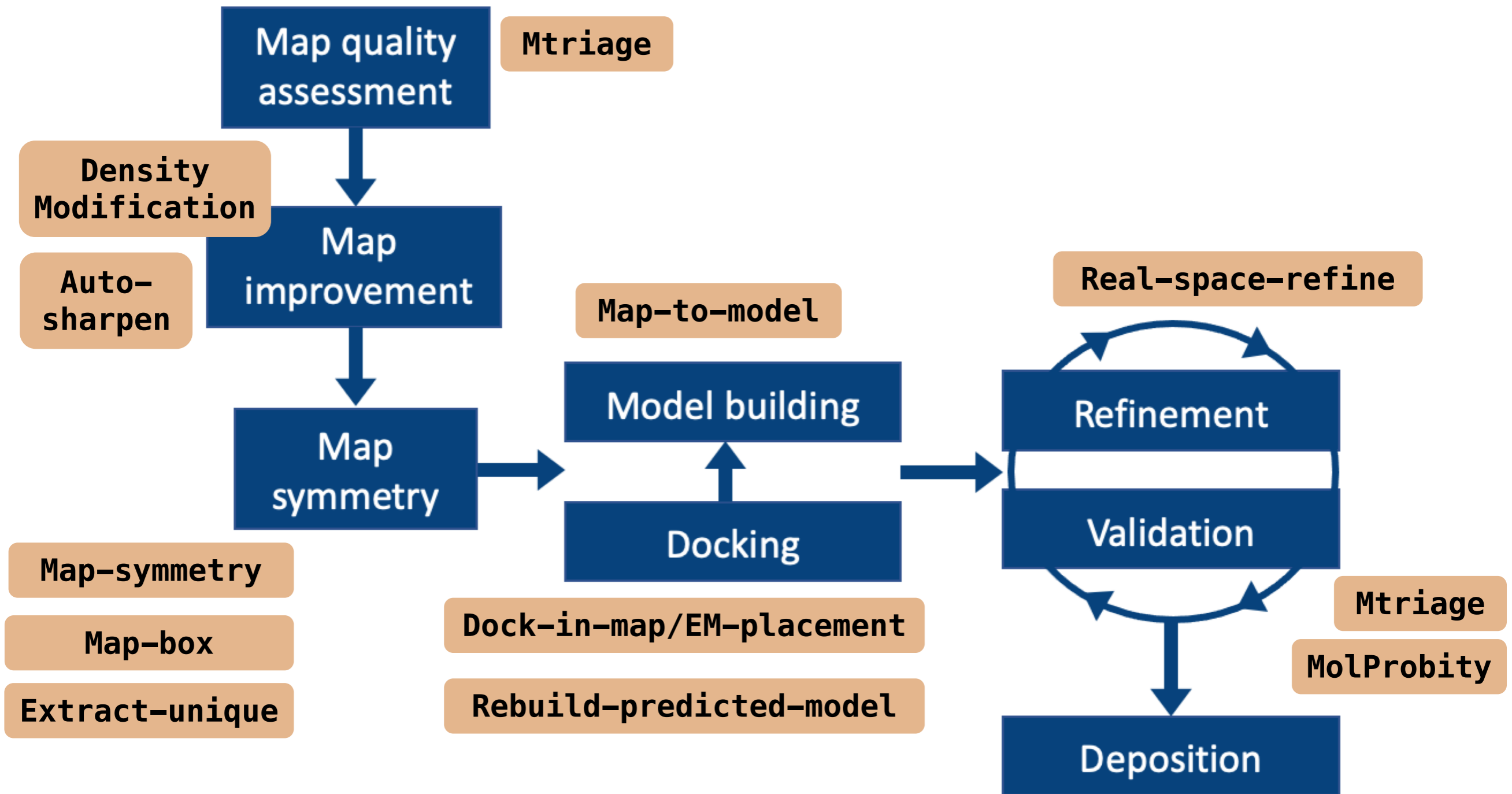
Tools for Crystallography



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

(Phenix)
(cctbx)
(Phenix)
(Phenix)

Tools for Cryo-EM



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

(Phenix)
 (cctbx)
 (Phenix)
 (Phenix)

Features

Phenix GUI

Central GUI for job control and to launch new jobs

The screenshot displays the PHENIX GUI interface. At the top, there is a title bar with the text "PHENIX home" and a series of icons for various functions: Quit, Preferences, Help, Citations, Reload last job, Coot, PyMOL, KiNG, Other tools, and Ask for help. Below the title bar, there are two tabs: "Actions" and "Job history".

The main interface is divided into two main sections. On the left, there is a "Projects" section with a "Show group:" dropdown menu set to "All groups" and a "Manage..." button. Below this are buttons for "Select", "Delete", "New project", and "Settings". A table lists the projects:

ID	Last modified	# of jobs	R-free
✓ test	Sep 14 2021 02:2...	529	0.0971

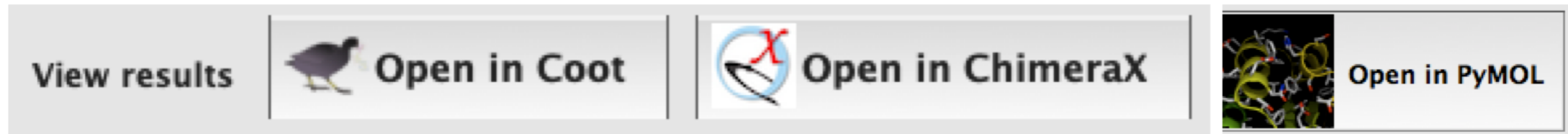
On the right, there is a "Favorites" section with a list of categories and sub-items:

- AlphaFold (predicted models)
- Crystals: Data analysis and manipulation, Validation and map-based comparisons, Experimental phasing, Molecular replacement, Maps (create, manipulate, compare), Enhanced maps (Polder, FEM, density-modified...), Model building, Refinement, Ligands
- Cryo-EM: Map analysis, symmetry, manipulation, Validation and map-based comparisons, Map improvement, Docking, model building and rebuilding, Refinement
- Models: Superpose, search, compare, analyze symmetry, Modification, minimization and dynamics
- PDB Deposition
- Program search

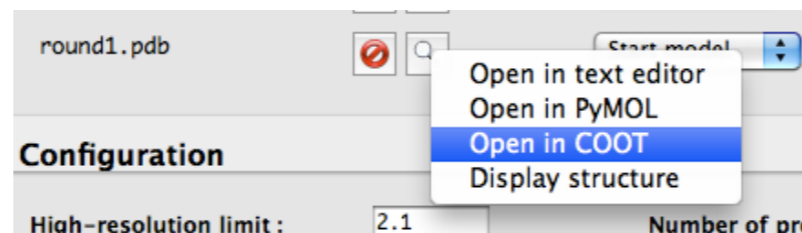
At the bottom, there is a "Current directory:" field with the path "/Users/dcliebschner/Desktop/Projects/test" and a "Browse..." button. The status bar at the very bottom shows "PHENIX version dev-svn-000" on the left and "Project: test" on the right.

Coot/PyMOL/ChimeraX integration

- Most results can be opened directly in graphics apps



- Any PDB file listed in GUI can also be opened

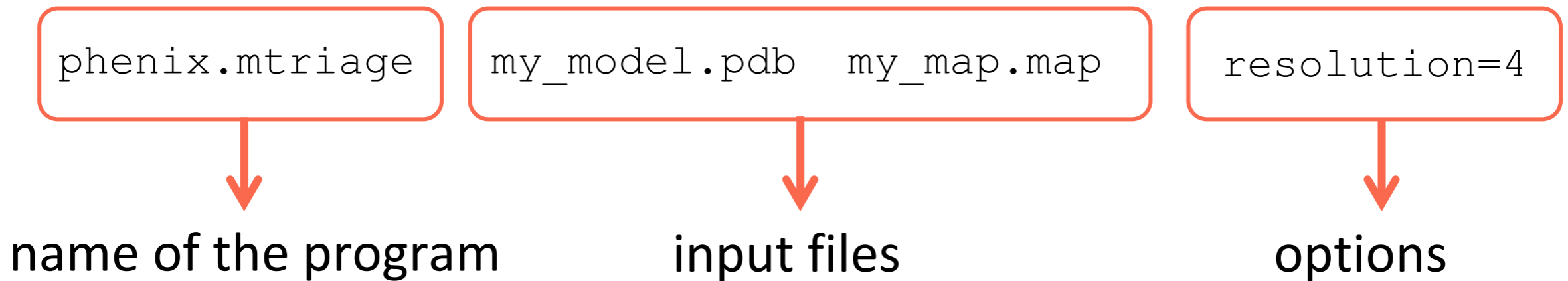


- AutoSol, AutoBuild, and phenix.refine will update Coot continuously while running
- Coot must have Python support (default on Mac)
- Specific paths to executables usually required on Linux

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

Command Line Tools

Run on the terminal



Run in a python script

```
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

Phenix

A comprehensive software package for macromolecular structure determination using crystallographic (X-ray, neutron and electron) and electron cryo-microscopy data.

NEW: Phenix with AlphaFold models

- Trim, weight, create domains and use for molecular replacement
- Trim, dock into cryo-EM maps and fill in gaps
- Reference models for refinement

[Learn more](#)

Cryo-EM map and superposed, refined AlphaFold model

Download | Getting Started | Workshops & Tutorials | Documentation

Help | Developers | National Resource | Industrial Consortium

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

www.youtube.com/channel/UCcdl0hfHngWAZLJWynxPQWg/videos

YouTube IT Search

Phenix Tutorials

560 subscribers

SUBSCRIBE 560

HOME VIDEOS PLAYLISTS CHANNELS ABOUT

Uploads PLAY ALL SORT BY

- real_space_refine Tutorial** (5:27)
How to run real-space-refine
763 views • 7 months ago
CC
- Secondary Structure Restraints Tutorial** (6:23)
How to use secondary structure restraints
399 views • 8 months ago
CC
- Multiple refinement strategies Tutorial** (5:46)
How to use multiple refinement strategies and...
281 views • 8 months ago
CC
- Planning a SAD experiment Tutorial** (6:00)
Simulate a SAD experiment with...
483 views • 1 year ago
CC
- Map-to-model Tutorial** (5:33)
Automatic map interpretation with map_to_model
1.3K views • 1 year ago
CC
- Scale-and-merge Tutorial** (7:31)
Scaling and merging anomalous data
387 views • 1 year ago
CC
- Automated map sharpening Tutorial** (6:05)
- Ligandfit Tutorial** (5:48)
- Wilson plots and space group identification phenix.xtriage** (8:23)
- Twinning phenix.xtriage** (6:08)
- Translational NCS phenix.xtriage** (4:55)
- Checking data quality with Xtriage** (6:49)

Dorothee Liebschner, Nigel Moriarty,
Tom Terwilliger, Christopher Schlicksup, Vincent Chen

What's new?

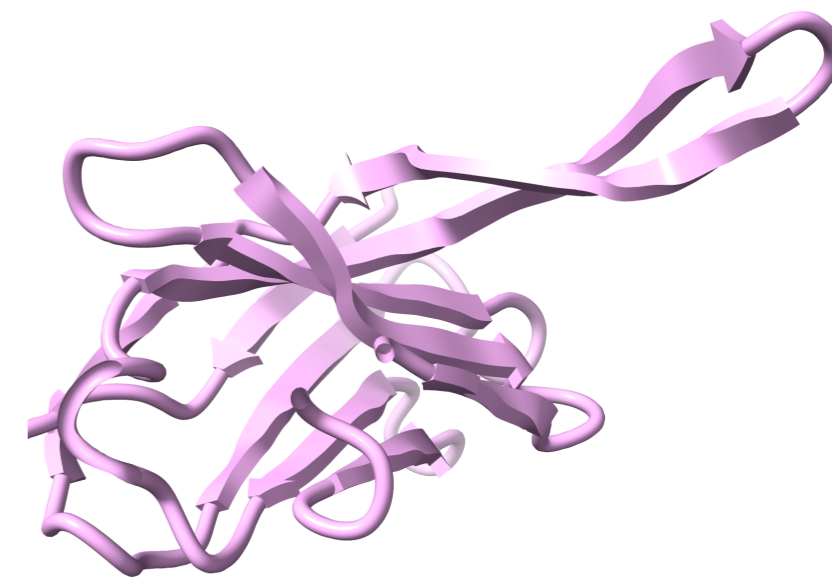
Accurate predicted models



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

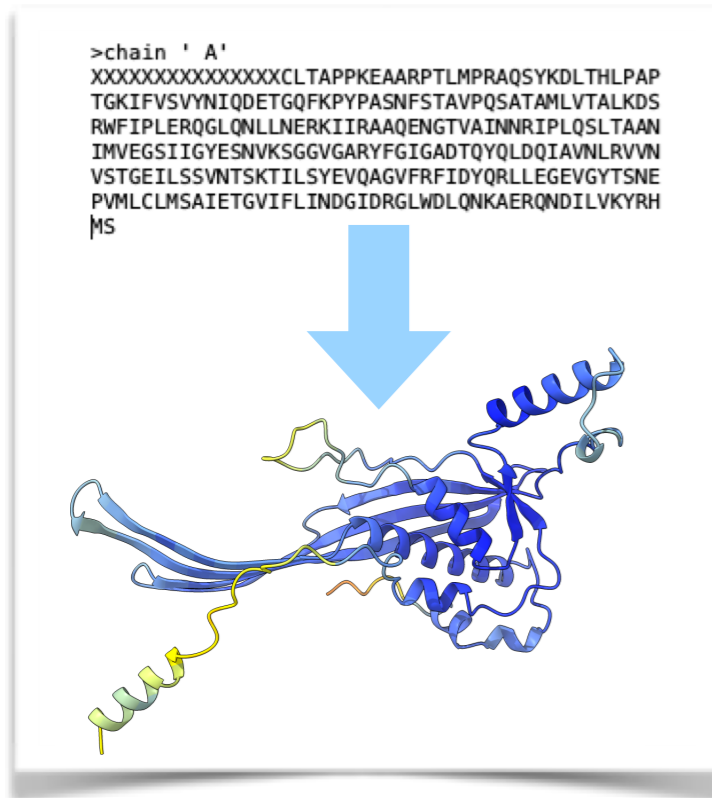
Sequence

Multiple sequence alignment



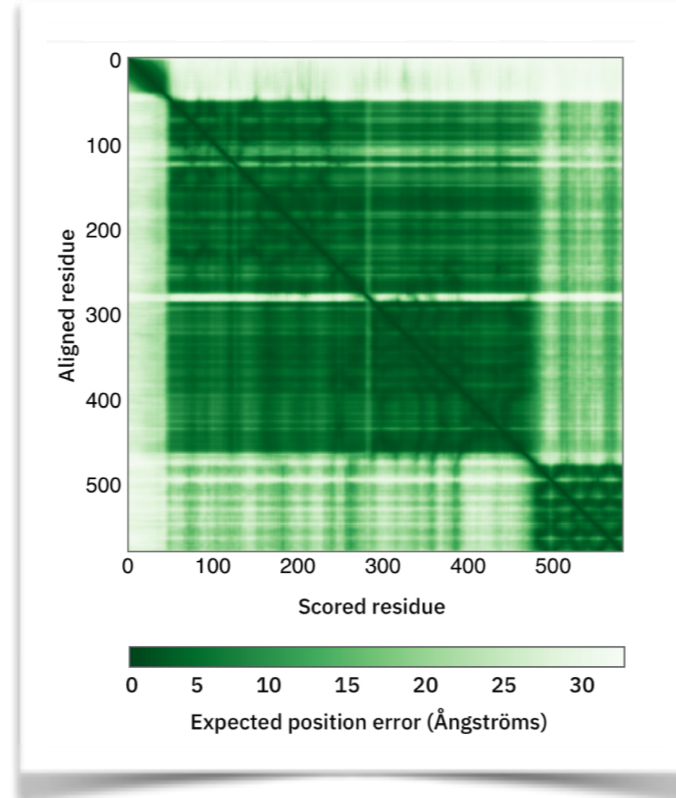
3D prediction

New tools for predicted models in Phenix

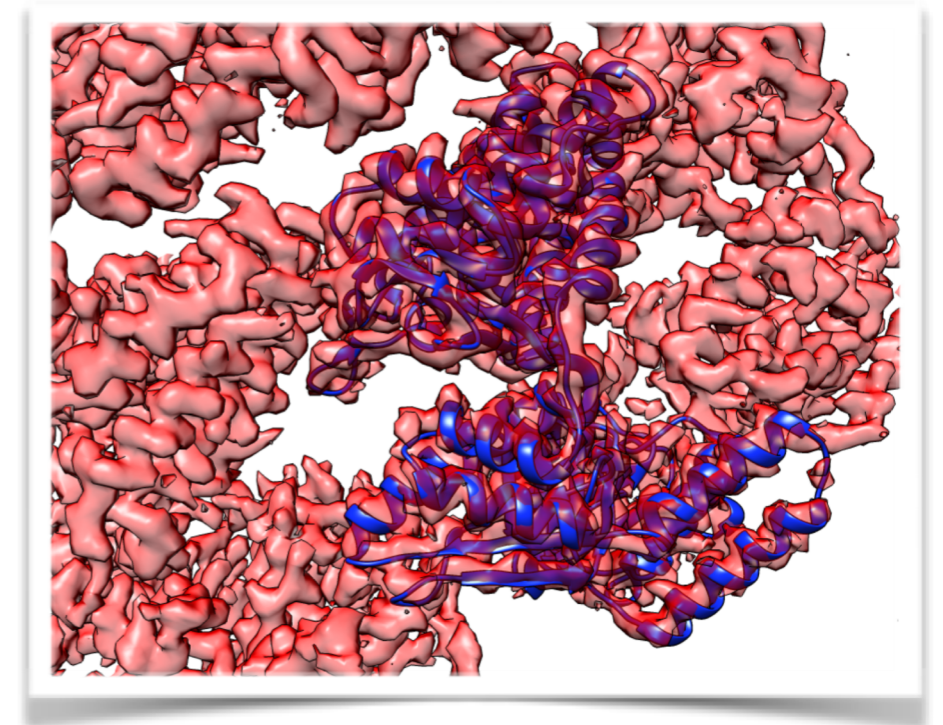


AlphaFold model prediction

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



Process predicted model



Predict and Build

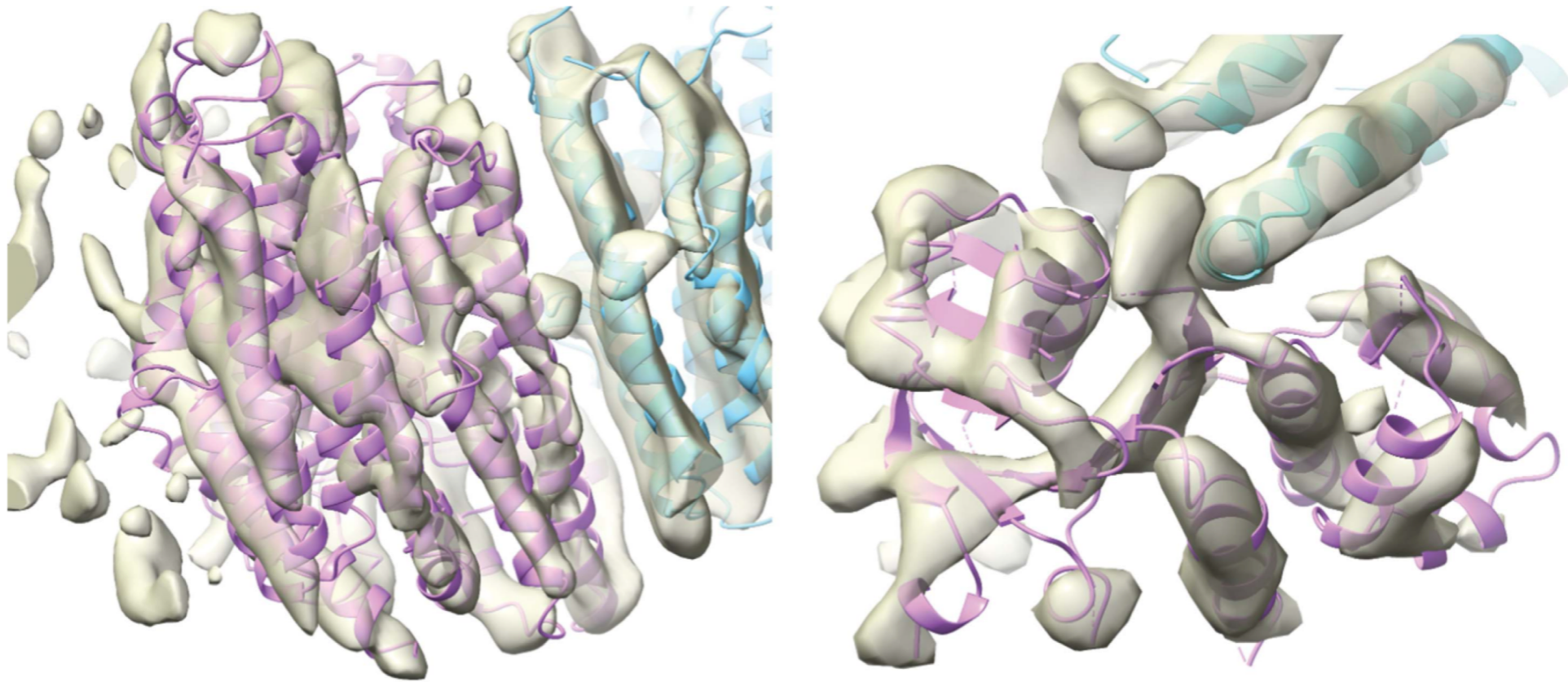
(Iterative AlphaFold prediction, docking, and rebuilding)

Fully automatic!

Likelihood-based EM docking

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

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



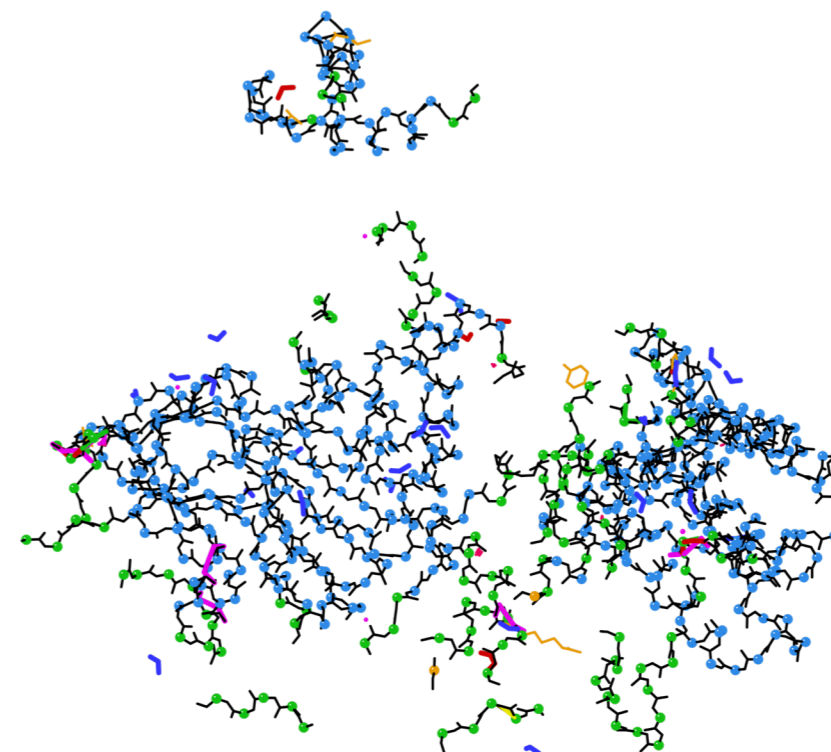
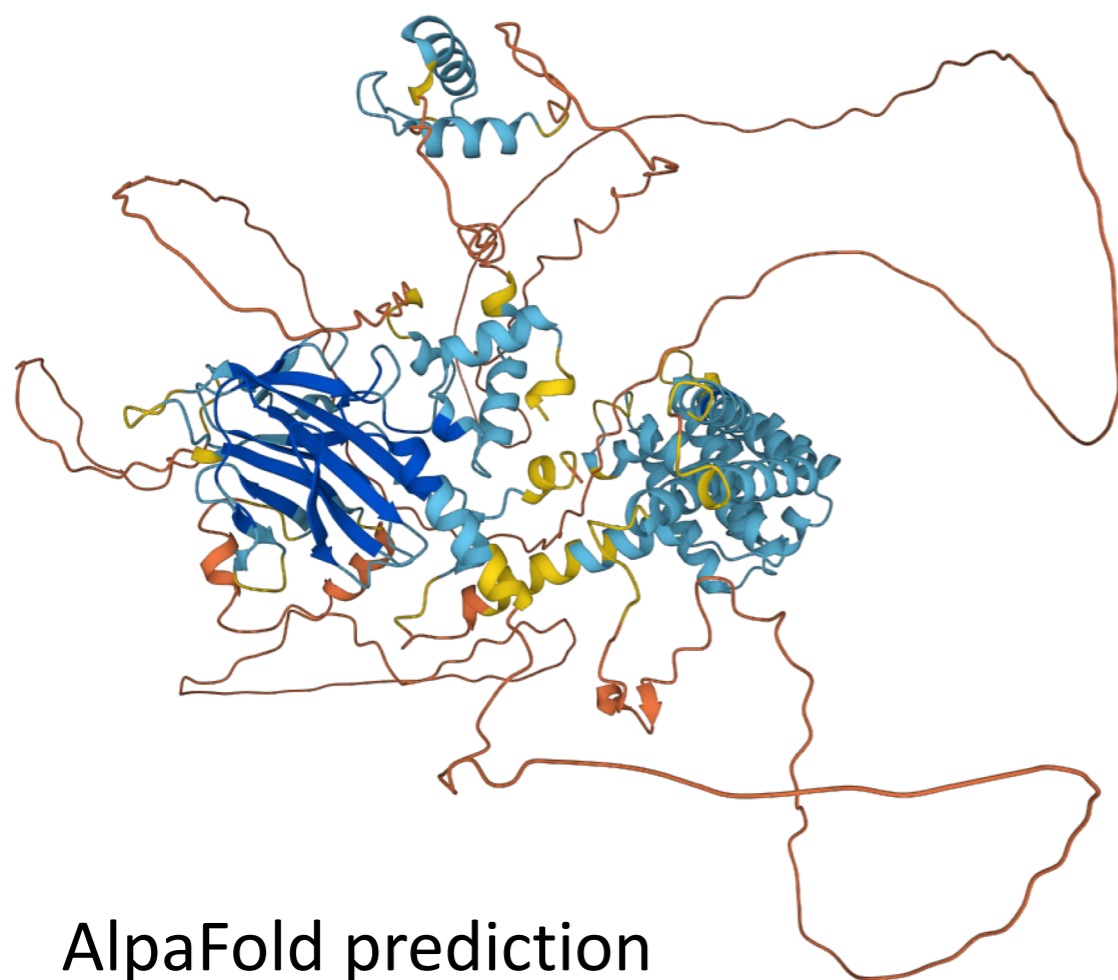
Read RJ, Millán C, McCoy AJ, Terwilliger TC. Likelihood-based signal and noise analysis for docking of models into cryo-EM maps. *Acta Crystallogr D Struct Biol.* 2023 Apr 1;79(Pt 4):271–80.

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**

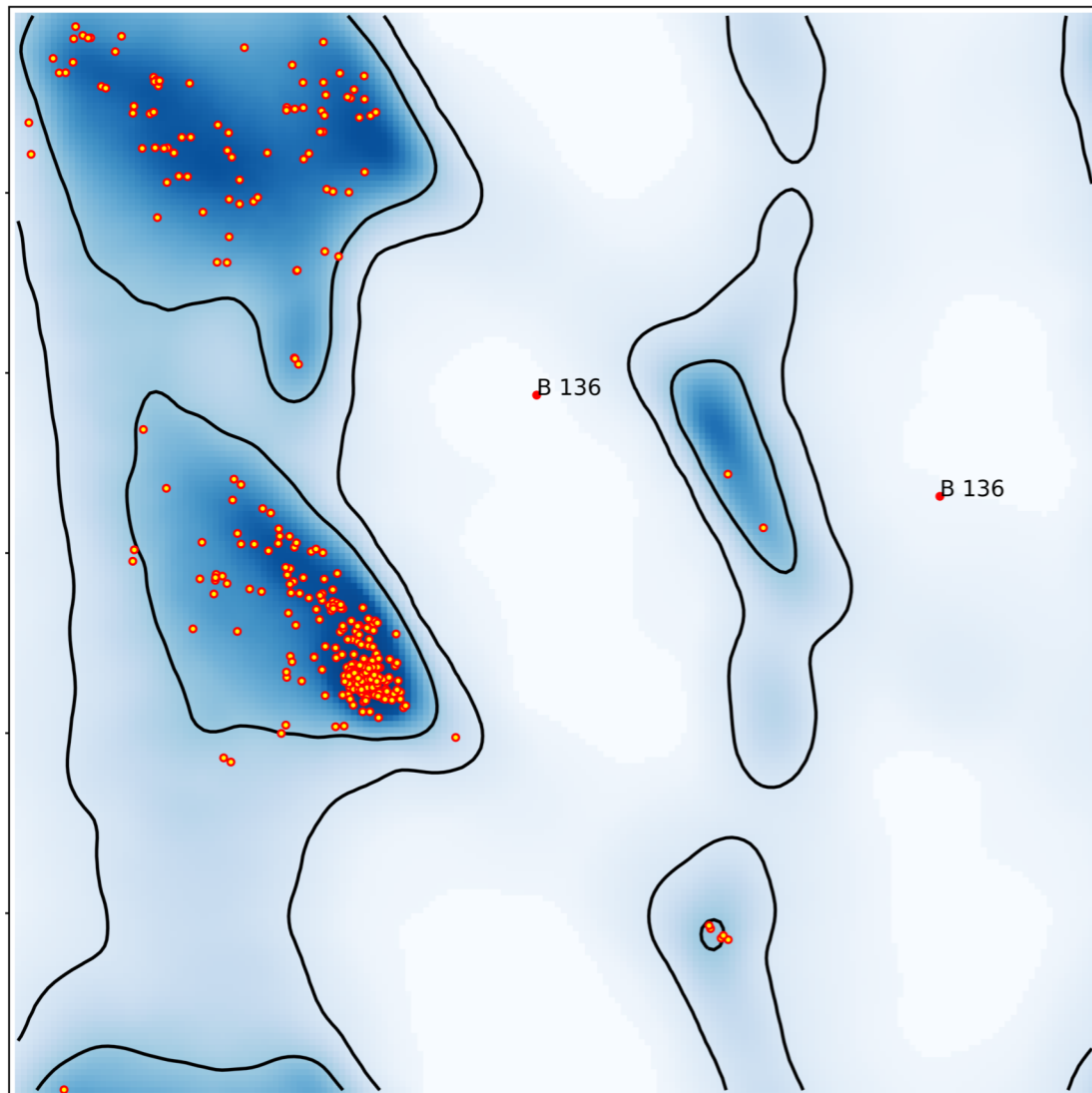


Sunday, August 27th 5:55 pm, Christopher Williams, *Identification and Validation of low-pLDDT regions in AF2 predictions (A011)*

Two Ramachandran distributions

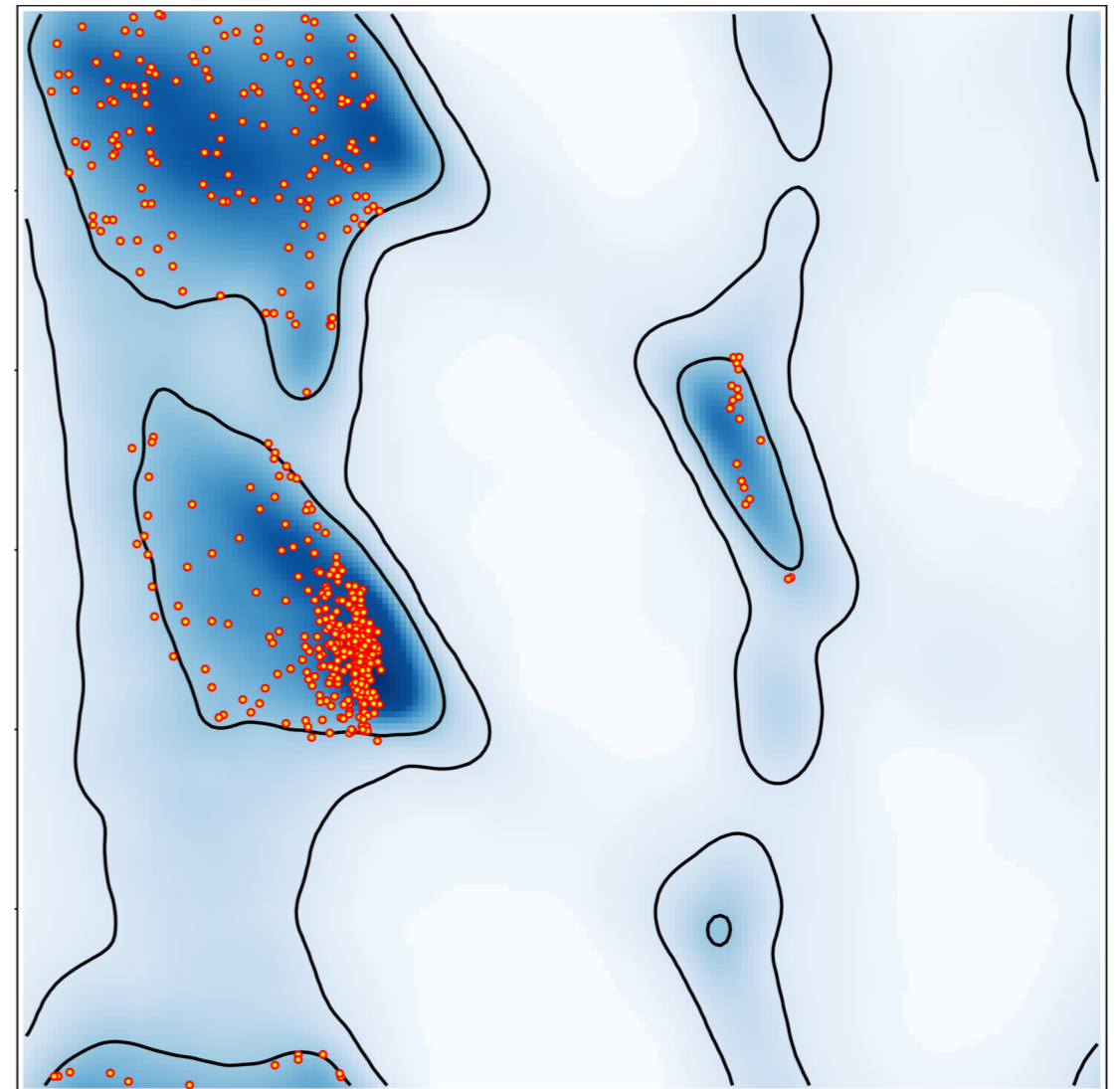
Model A

Favored	97.8 %
Allowed	1.95 %
Outliers	0.25 %



Model B

Favored	96.2 %
Allowed	3.8 %
Outliers	0.0 %



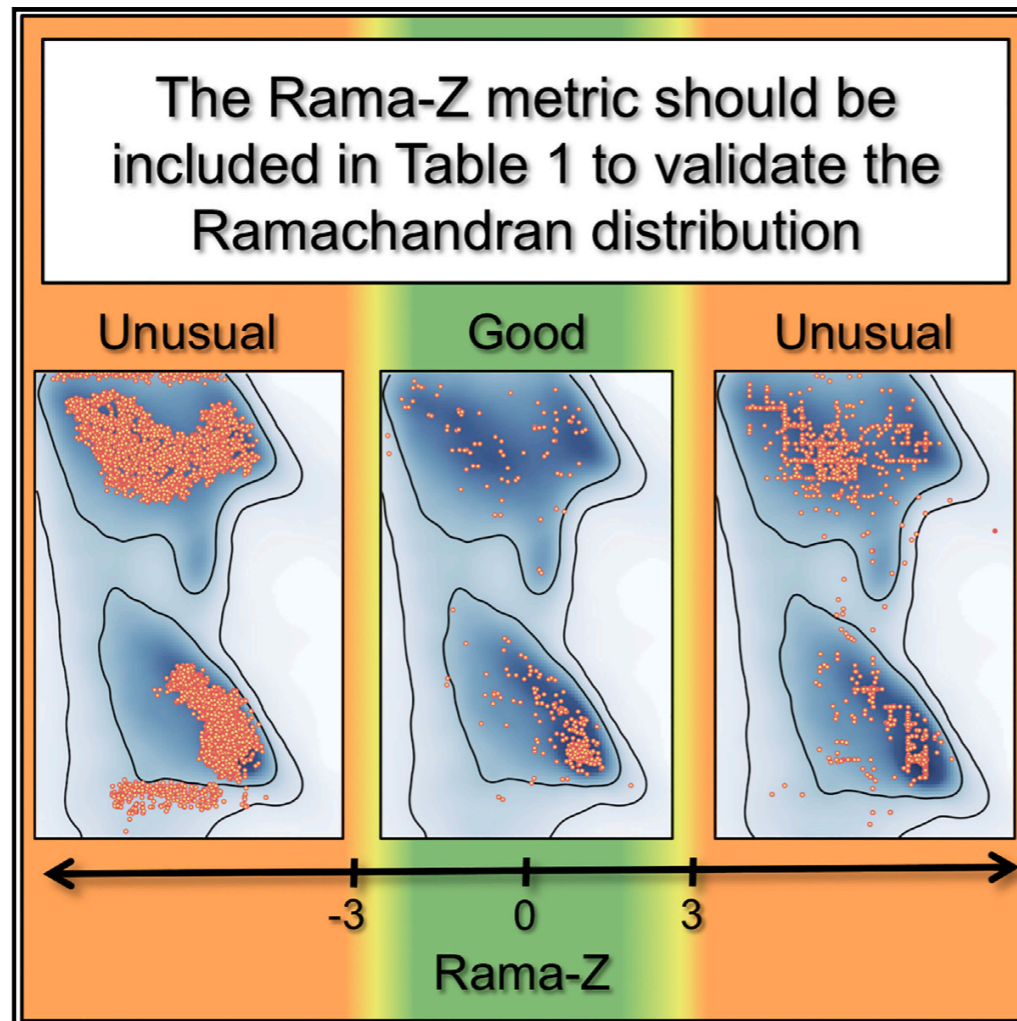
Global Ramachandran Score

Resource

Structure

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 reimplementa-tion 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

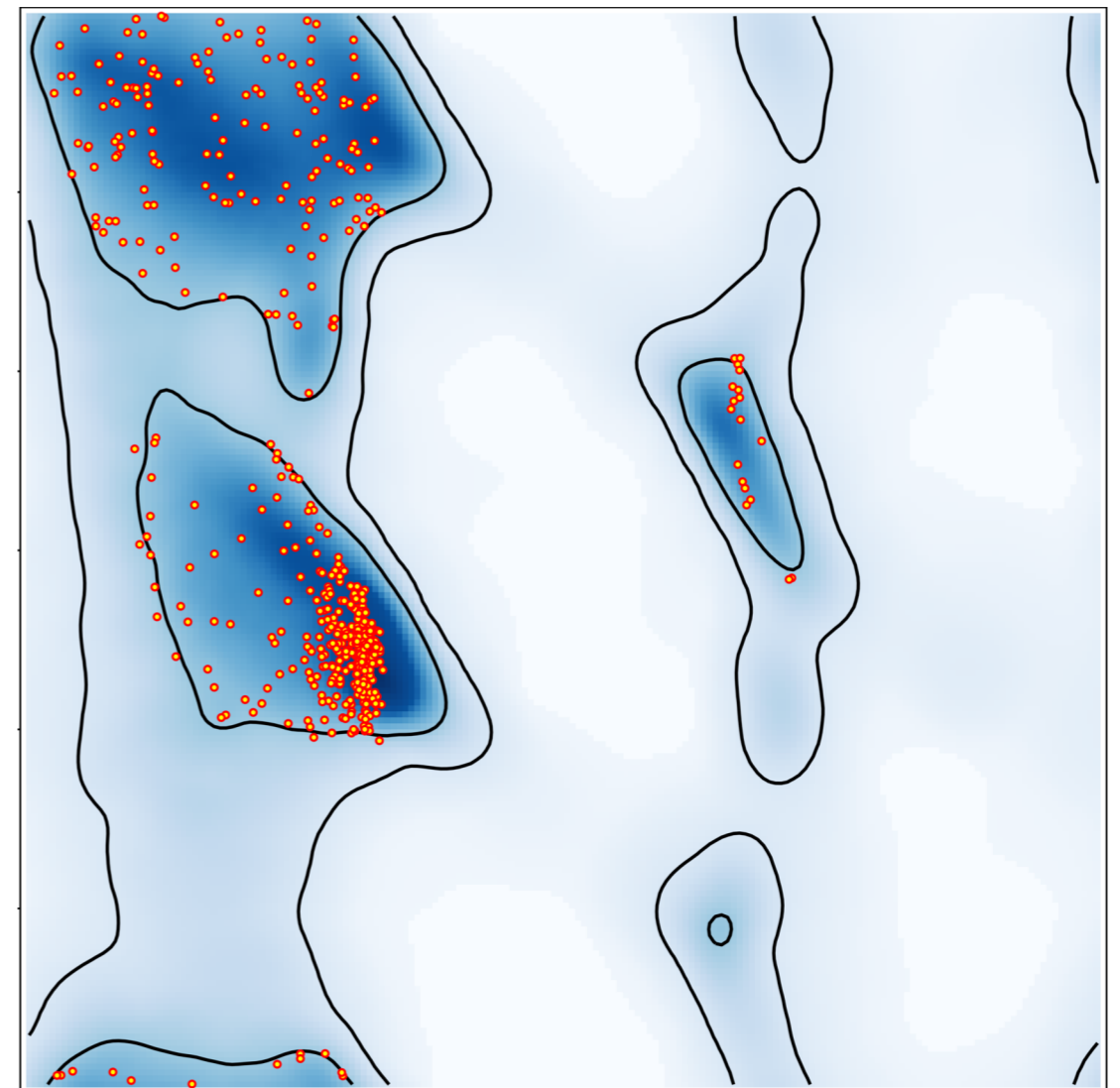
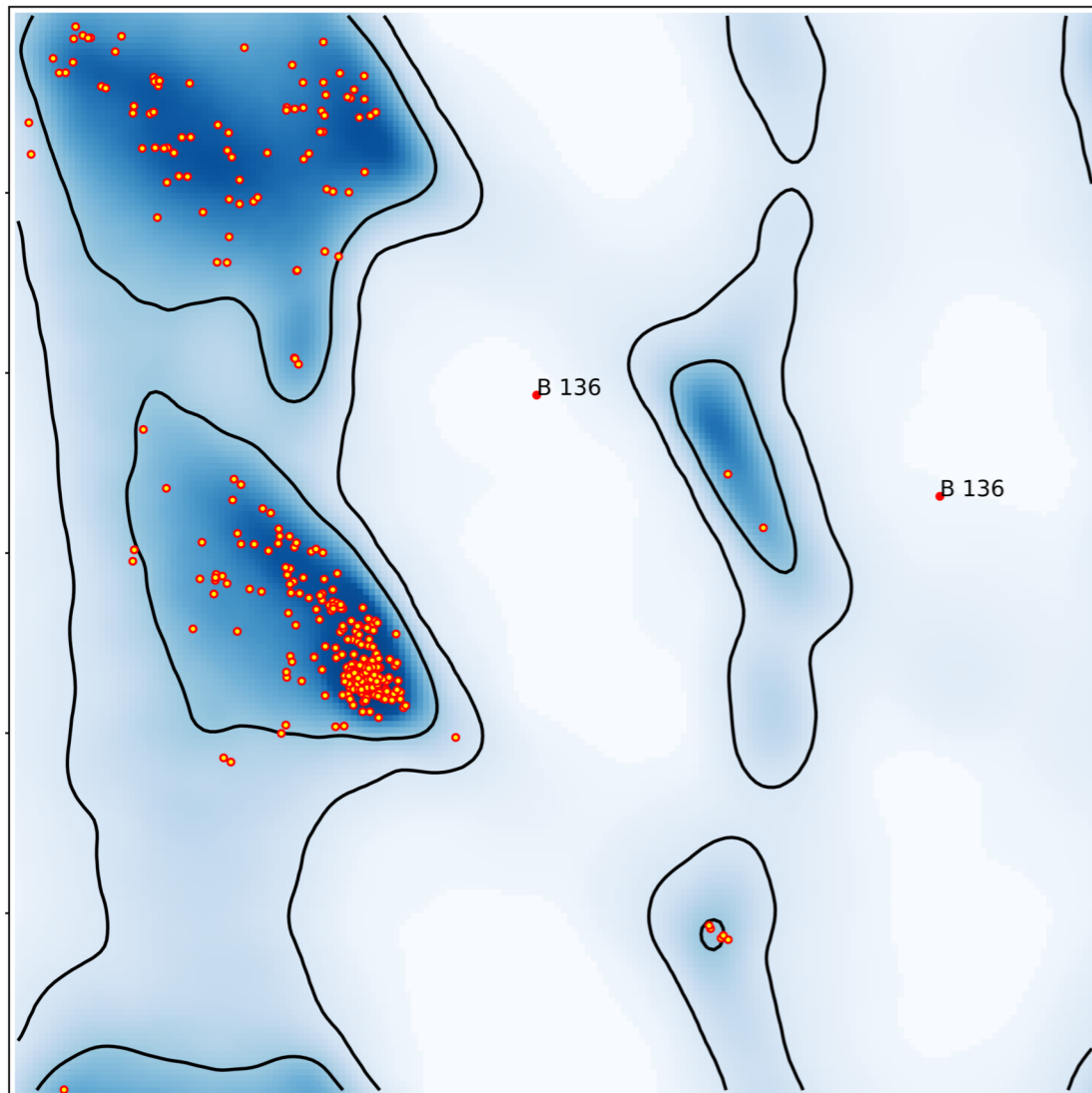
Rama z-score

-0.19

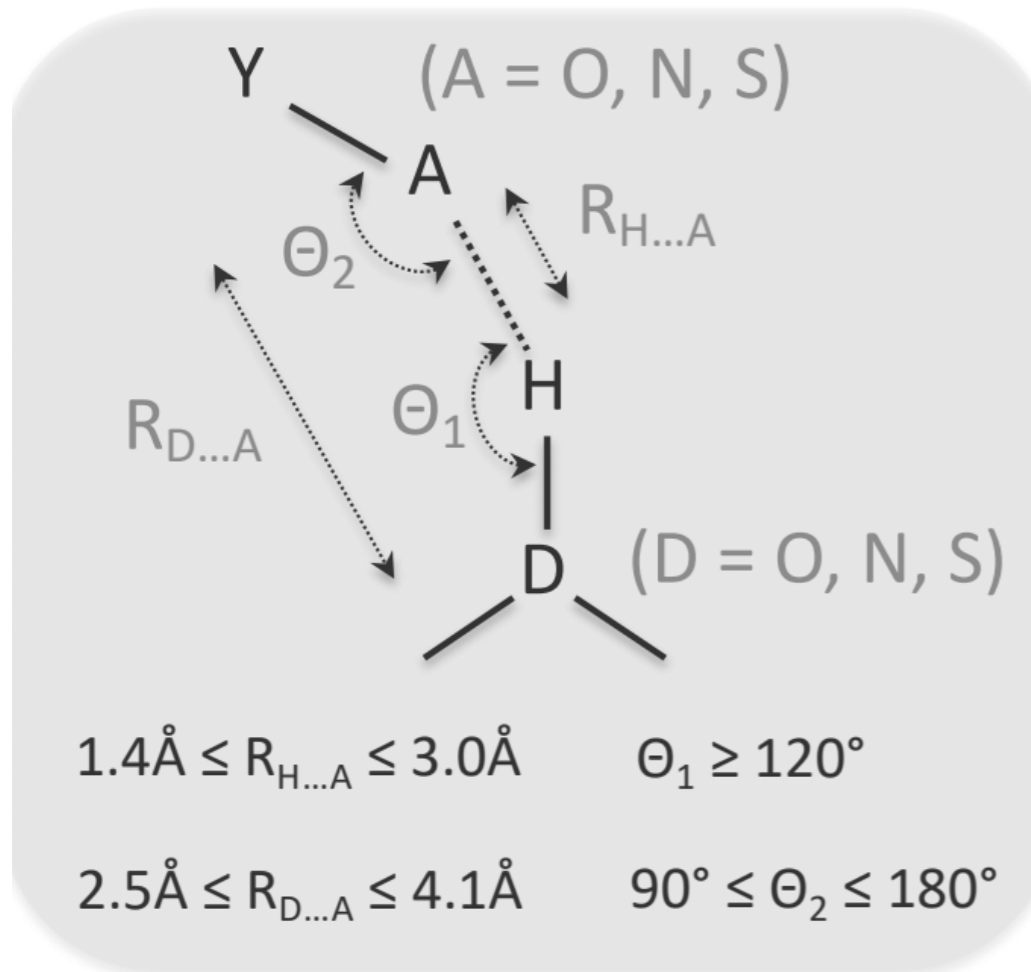
-4.08

Model B

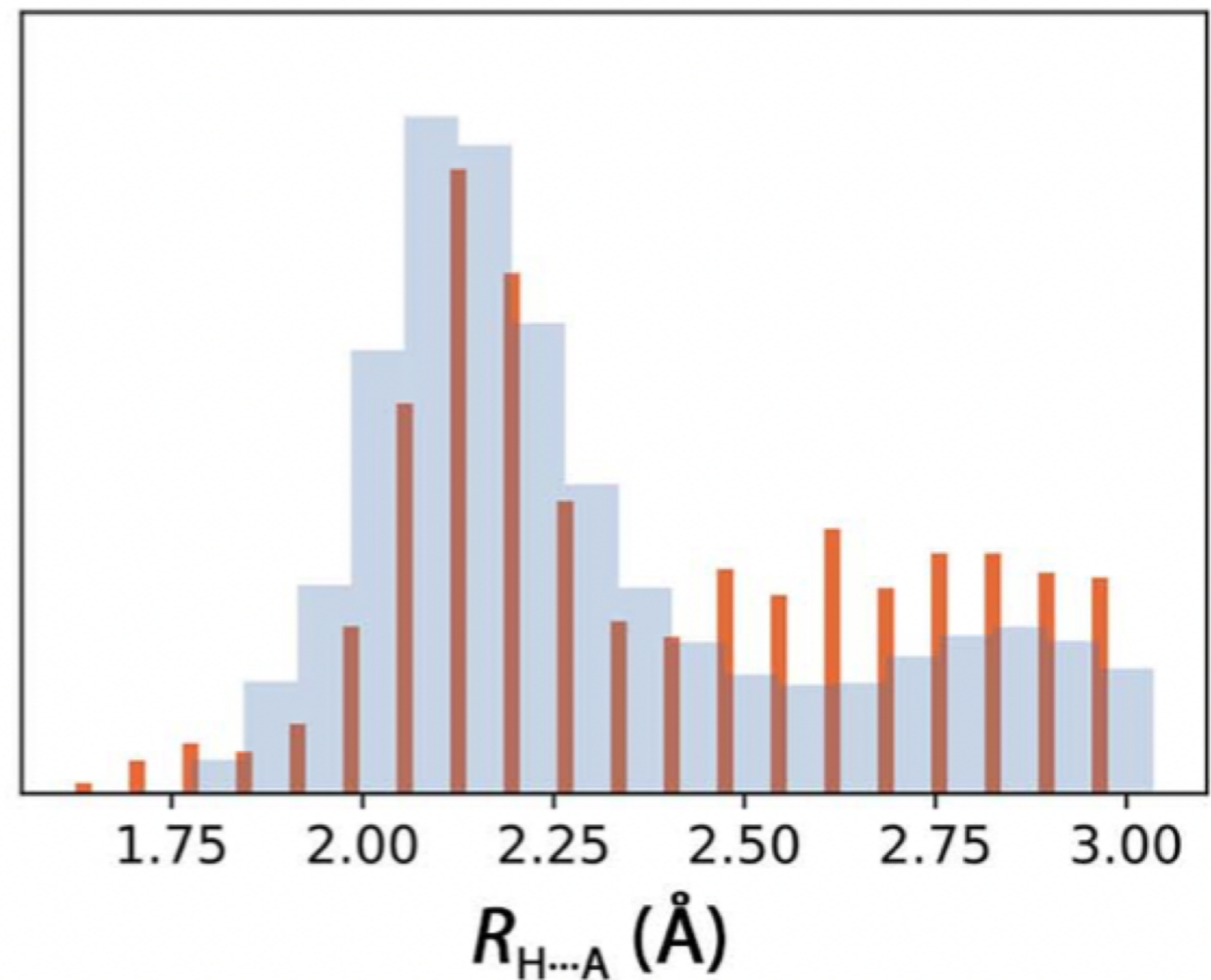
Favored	96.2
Allowed	3.8
Outliers	0.0



Structure validation with H-bond parameters



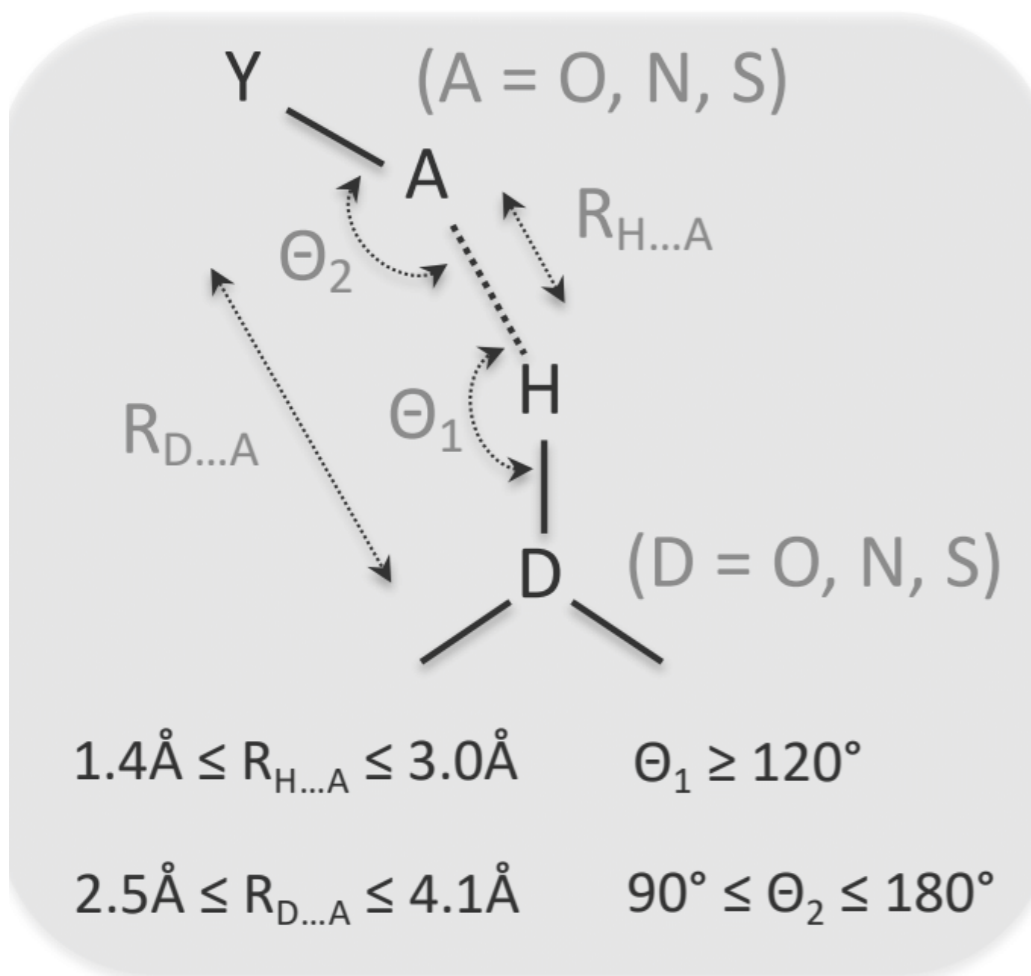
Validation tool based on H-bond parameter distribution of high-resolution PDB models.



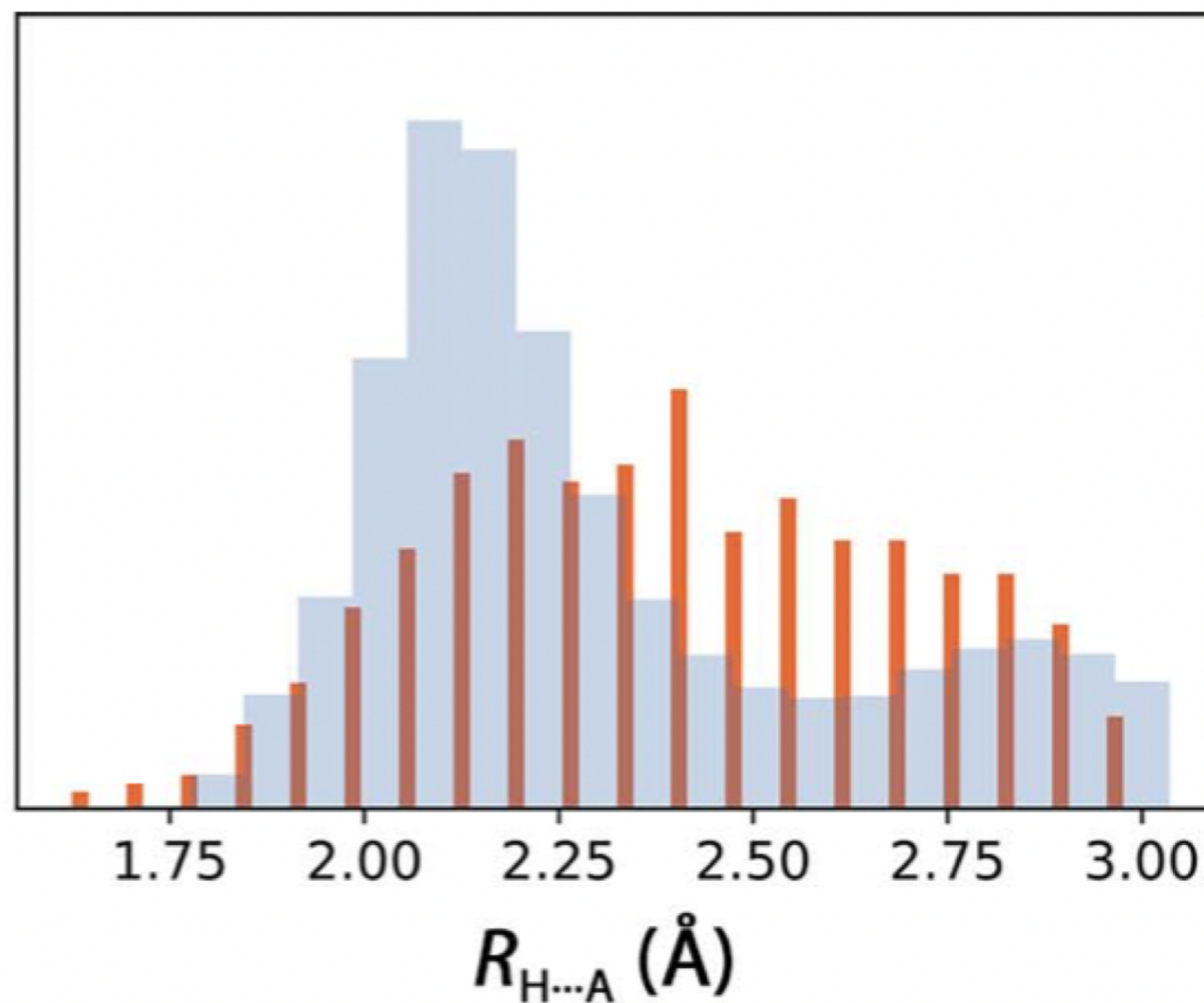
Grey: reference distribution

Red: 6mdo

Structure validation with H-bond parameters



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

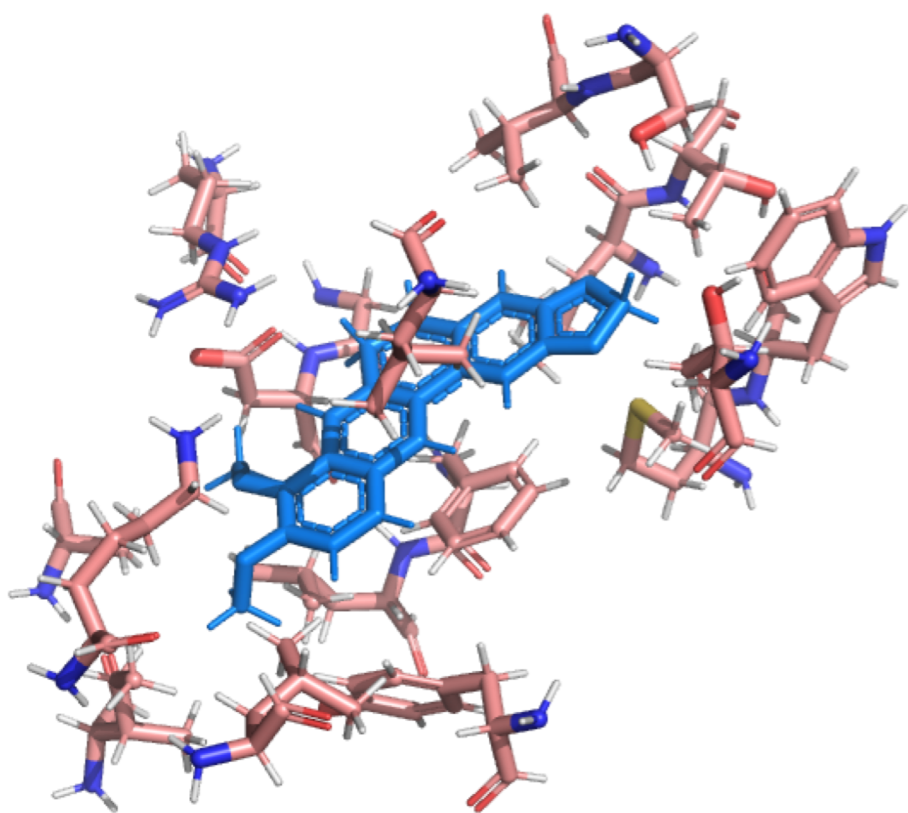


Grey: reference distribution

Red: 5j1f

QMR – quantum mechanical restraints

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



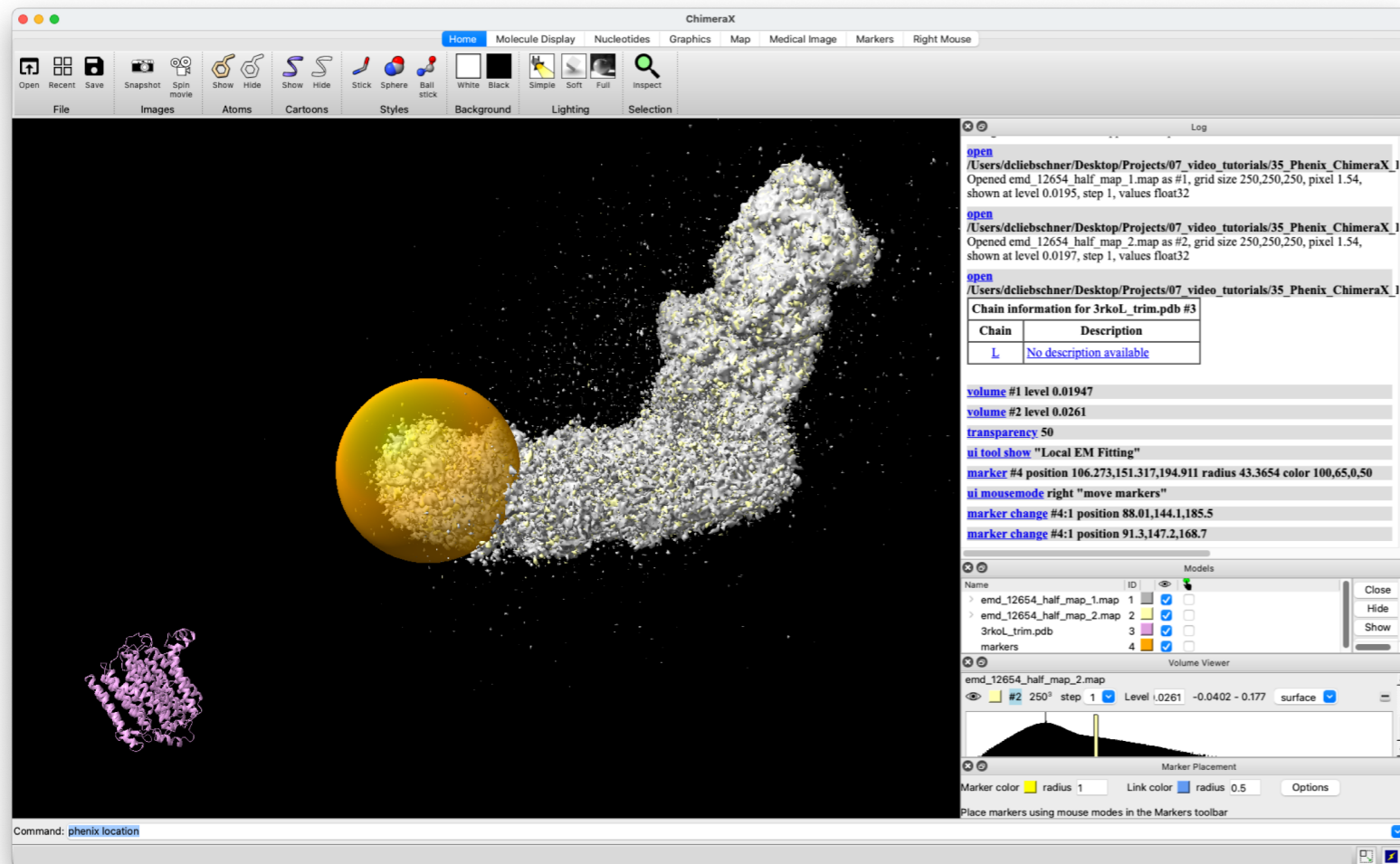
BER in 3vw2

QMR approach:

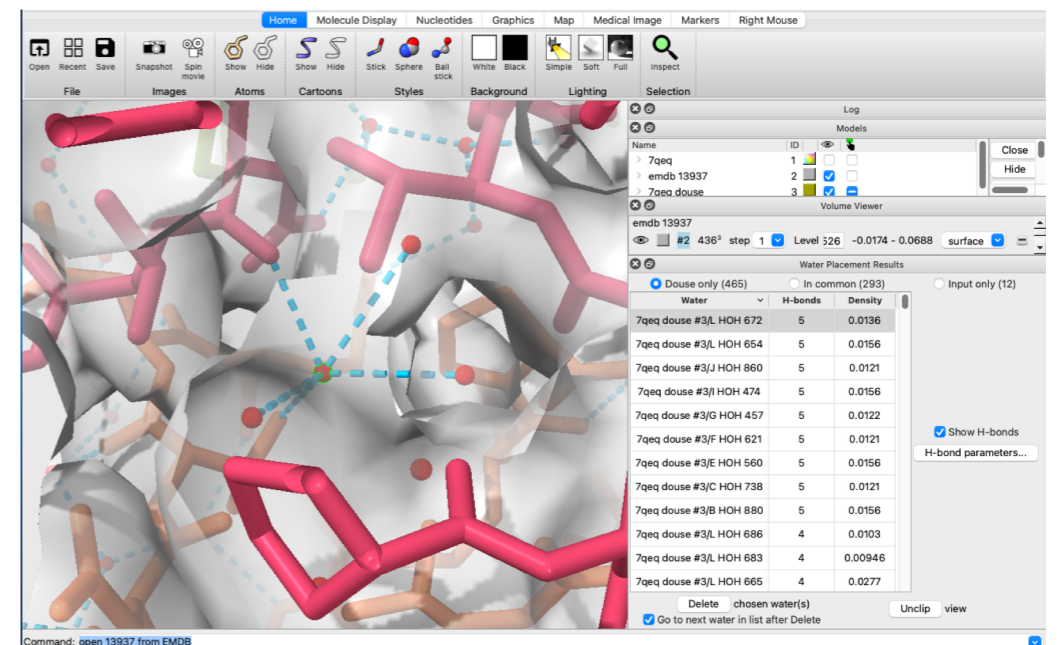
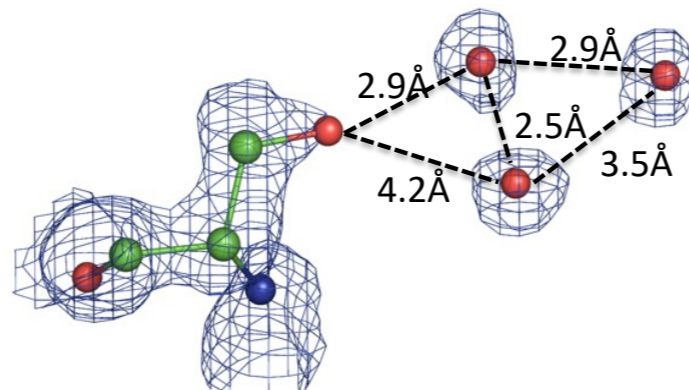
- Minimize the ligand geometry in the binding pocket.
- Use minimized ligand geometry as targets for restraints.
(forget about what happened with the residues)

Run Phenix tools through ChimeraX

Local EM fitting
(EM placement)



Automated water building
(phenix.douse)



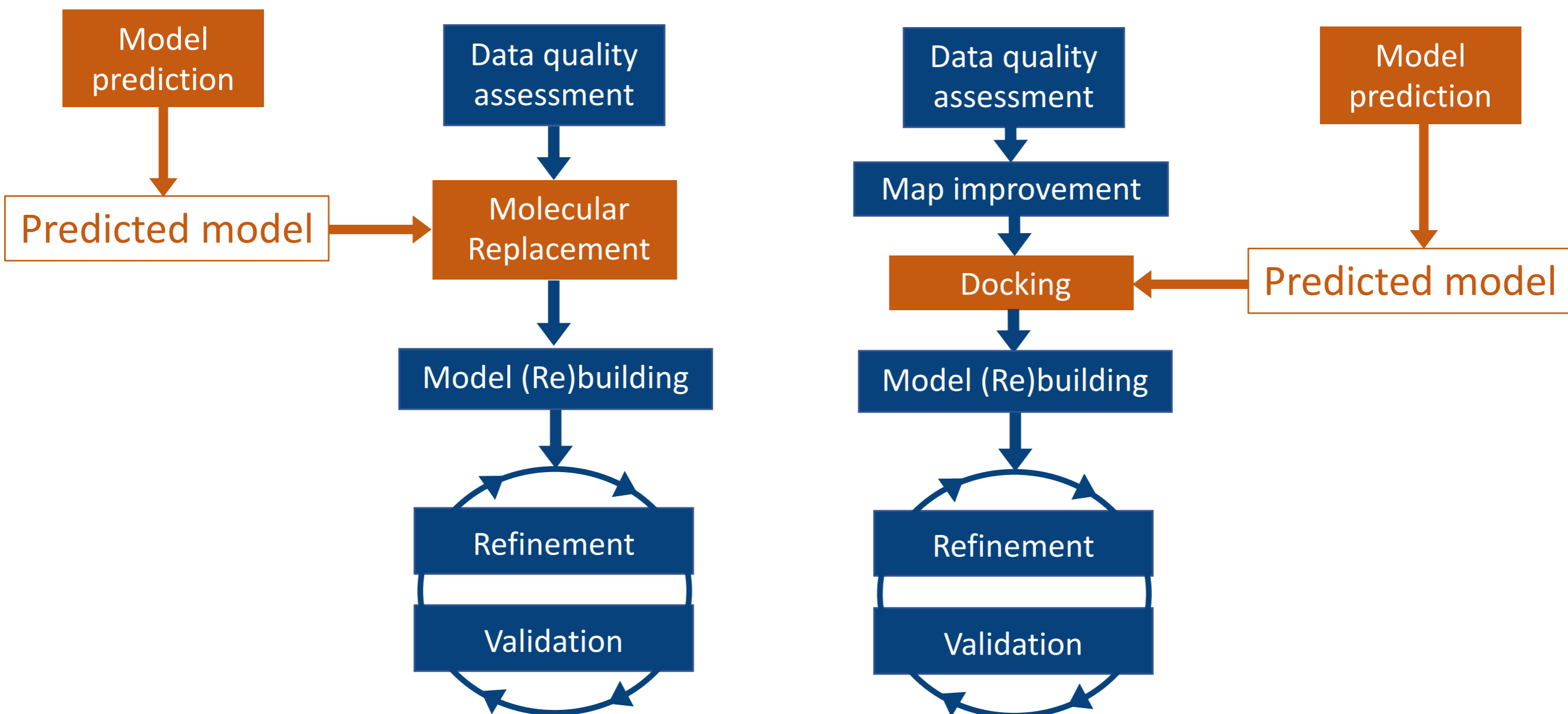
“Demo”

How to use predictions?

Incorporate predictions into the “conventional” structure determination workflow.

Crystallography

Cryo-EM



How to use predictions? (X-ray)

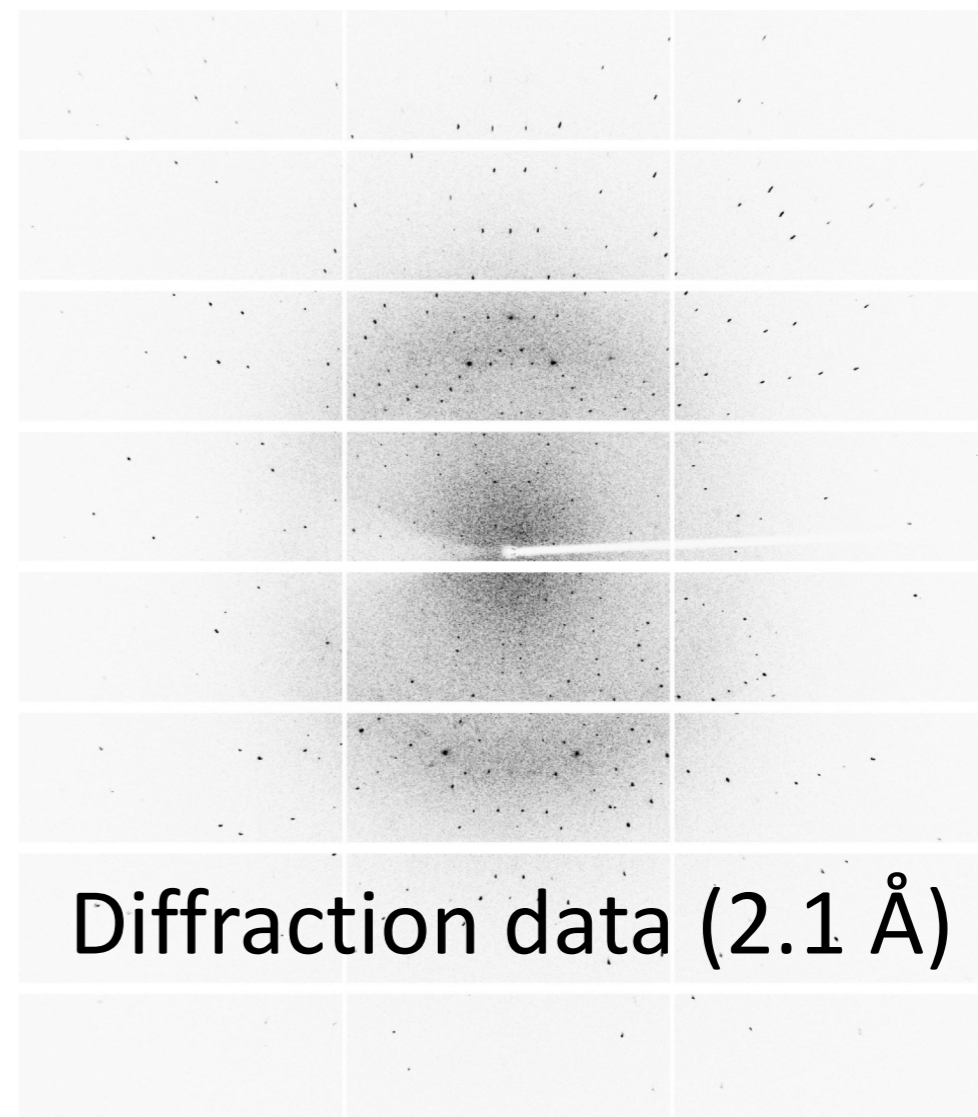
Incorporate predictions into the “conventional” structure determination workflow.

Example:

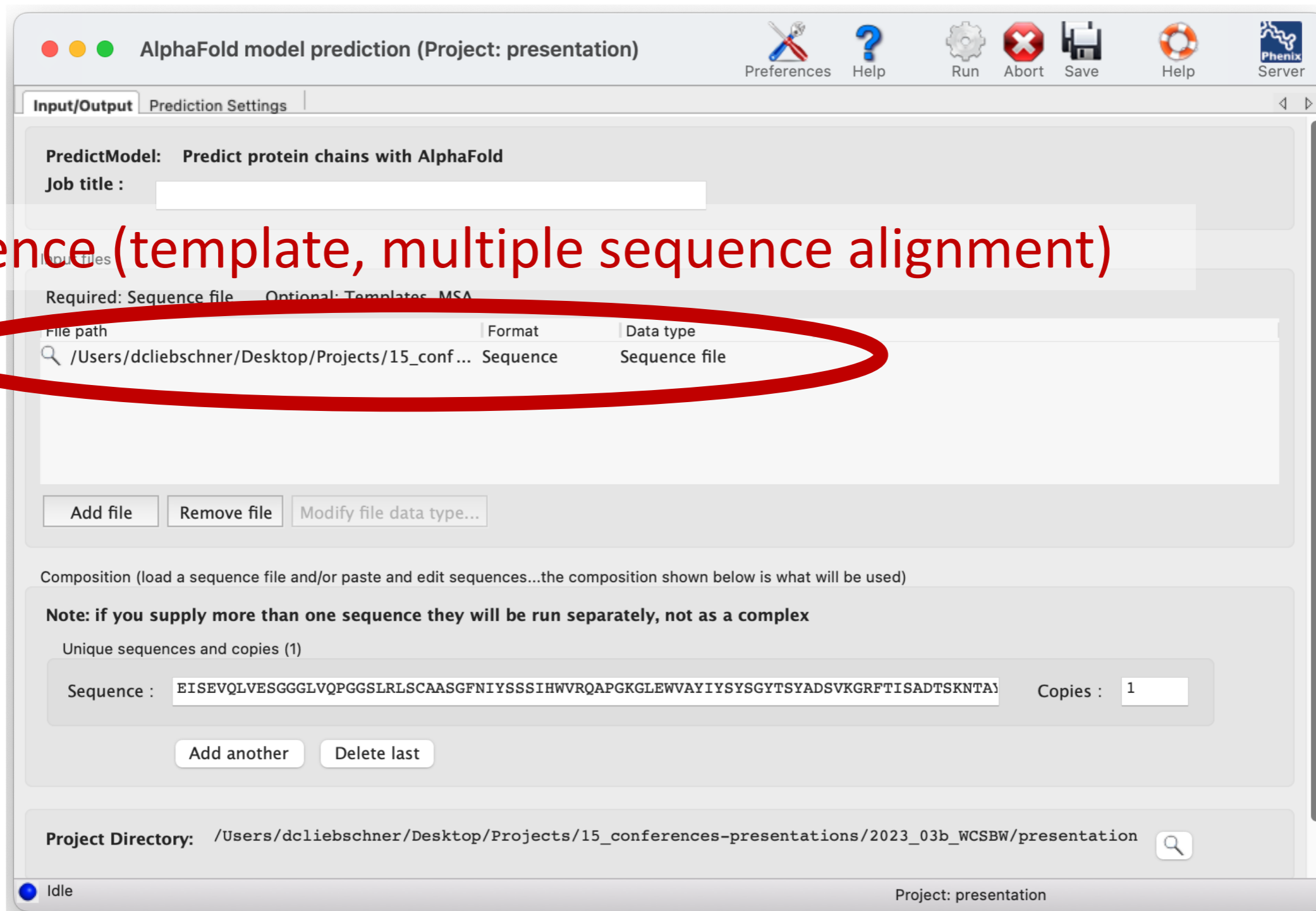
Diffraction data 7e9L (bovine protein of POMGNT2)

```
>POMGNT2_BOVINE soluble expression construct
GAPPAPALRIDYPKALQIL
TEGGTHMVCTGRTHDRLCRFKWLCYSSEAEFFIFFHGNASVMLPSLGSRRFQPALLDLS
TVEDHNTQYFNFVELPAAALRFMPKPVFVPDVALIANRFNPDNLMHVFHDDLLPLFYTLR
QFPGLAREARLFFMEGWGEGAHFDLYKLLSPKQPLLRAQLKALGRLLCFSHAFVGLSKVT
TWYQYGFVQPQGPKANILVSGNEIRQFAHFLMEKLNVSQAGGPLGEEYILVFSRTQNRLI
LNEAELLALAQEFQMKTVTVSLEDHAFADVRLVSNASMLVSMHGAQLVTALFLPRGAA
VVELFPYAVNPDHYTPYKTLATLPGMDLQYIAWQNTMPENTVTHPERPWDQGGIAHLDR
EQARILQSREVPRLCCRNPEWLFRIYQDTKVDIPSLIQTIRRVVKGHPGPRKQKWTVSL
YPGKVREARQCASVQGASEARLSVSWQIPWNLKYLKVREVKYEVWLQEQQENTYVPYMLA
LQNHTFTENIKPFTTYLVWIRCI FNKTL LGPFADVLCST
```

Sequence (human protein)



Predicting a model in Phenix



AlphaFold model prediction (Project: presentation)

PredictModel: Predict protein chains with AlphaFold

Job title :

Required: Sequence file Optional: Templates MSA

File path	Format	Data type
/Users/dcliebschner/Desktop/Projects/15_conf...	Sequence	Sequence file

Add file Remove file Modify file data type...

Composition (load a sequence file and/or paste and edit sequences...the composition shown below is what will be used)

Note: if you supply more than one sequence they will be run separately, not as a complex

Unique sequences and copies (1)

Sequence : EISEVQLVESGGGLVQPGGSLRLSCAASGFNIYSSSIHWVRQAPGKLEWVAYIYSYSGYTSYADSVKGRFTISADTSKNTA Copies : 1

Add another Delete last

Project Directory: /Users/dcliebschner/Desktop/Projects/15_conferences-presentations/2023_03b_WCSBW/presentation

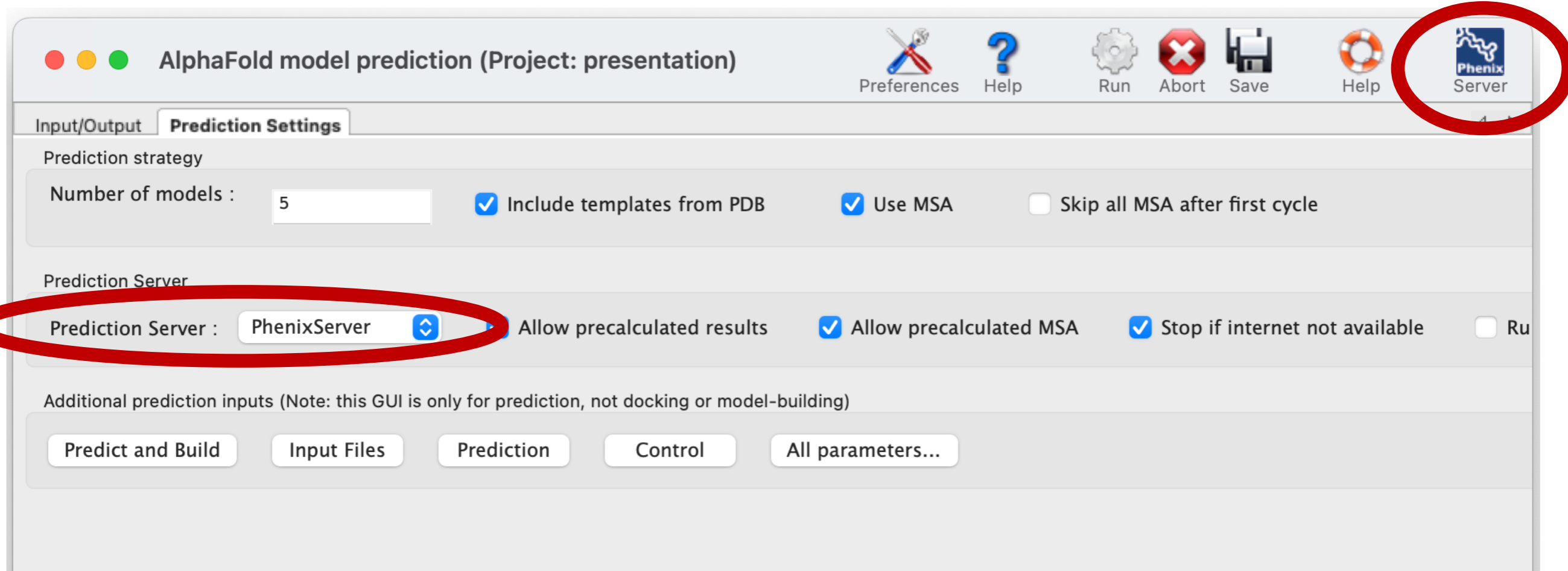
Idle Project: presentation

Sequence (template, multiple sequence alignment)

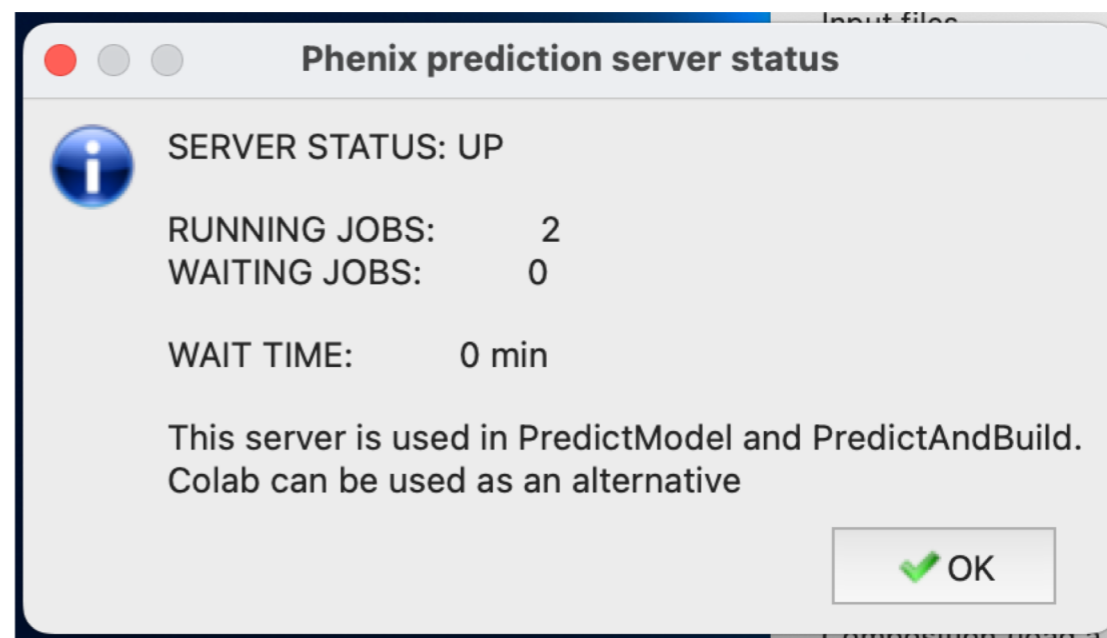
Do I need to install AlphaFold?



Phenix Server for running AlphaFold



No need for a local
AlphaFold installation



Get a prediction

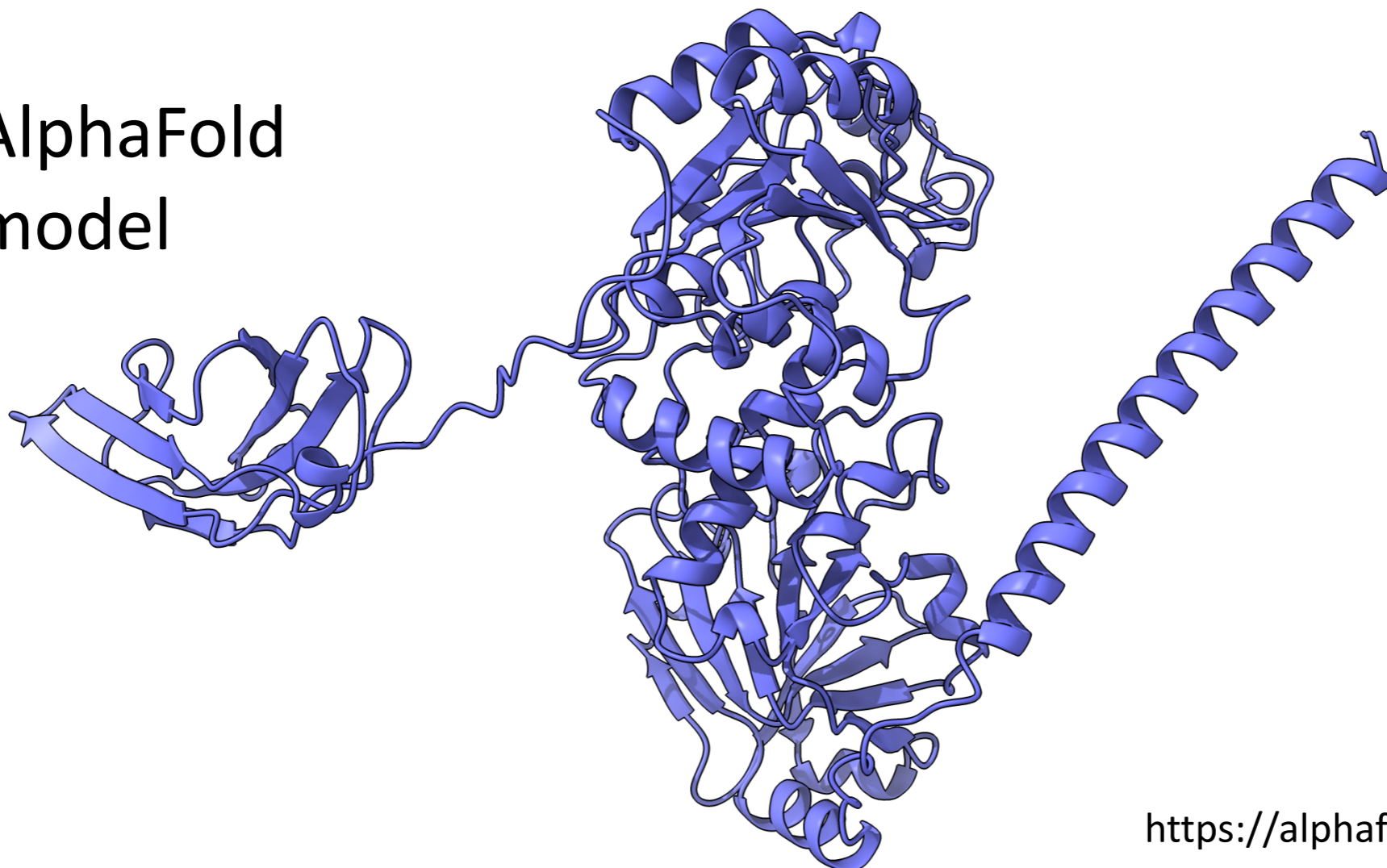
sequence

```
>POMGNT2_BOVINE soluble expression construct  
GAPPAPALRIDYPKALQIL  
TEGGTHMVCTGRTHDRLCRFKWLCSSEAEFFIFFHGNASVMLPSLGSRRFQPALLDLS  
TVEDHNTQYFNFVELPAAALRFMPKPVFVDPVALIANRFNPDNLMHVFDLLPLFYTLR  
QFPGLAREARLFFMEGWGEGAHFDLYKLLSPKQPLLRAQLKALGRLLCFSHAFVGLSKVT  
TWYQYGFVQPQGPKANILVSGNEIRQFAHFLMEKLNVSQAGGPLGEEYILVFSRTQNRLI  
LNEAELLALAQEFQMKTVTVSLEDHAFADVRLVSNASMLVSMHGAQLVTALFLPRGAA  
VVELFPYAVNPDHYTPYKTLATLPGMDLQYIAWQNTMPENTVTHPERPWDQGGIAHLDR  
EQARILQSREVPRLCCRNPEWLFRIYQDTKVDIPSLIQTIIRRVVKGHPGPRKQKWTVSL  
YPGKVREARQASVQGASEARLSVSWQIPWNLKYLKVREVKYEVWLQEQENTYVPMYLA  
LQNHTFTENIKPFTTYLVWIRCFNKTLLGPFADVLCST
```



EBI AlphaFold database
or
Run AlphaFold prediction

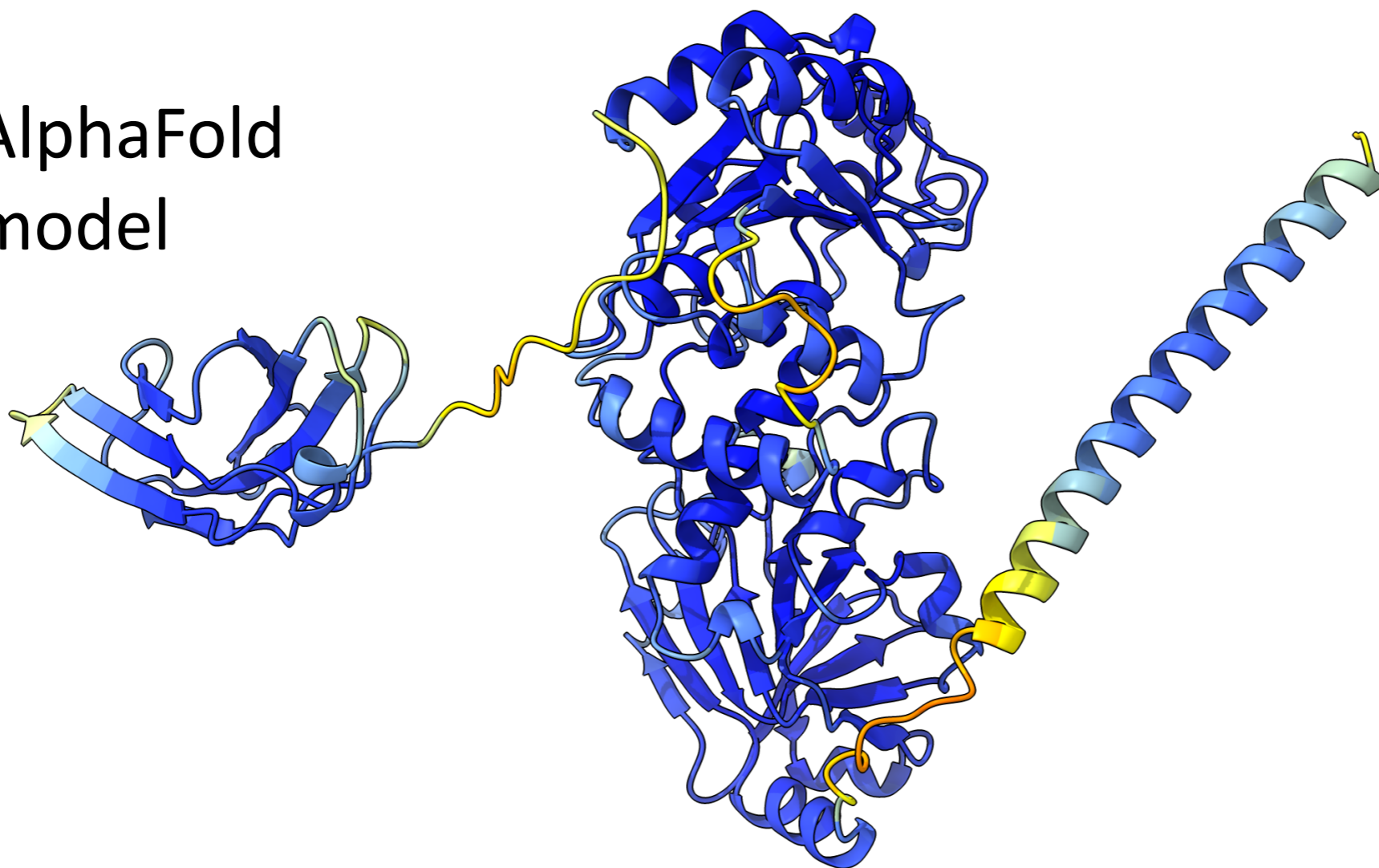
AlphaFold
model



Process prediction

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

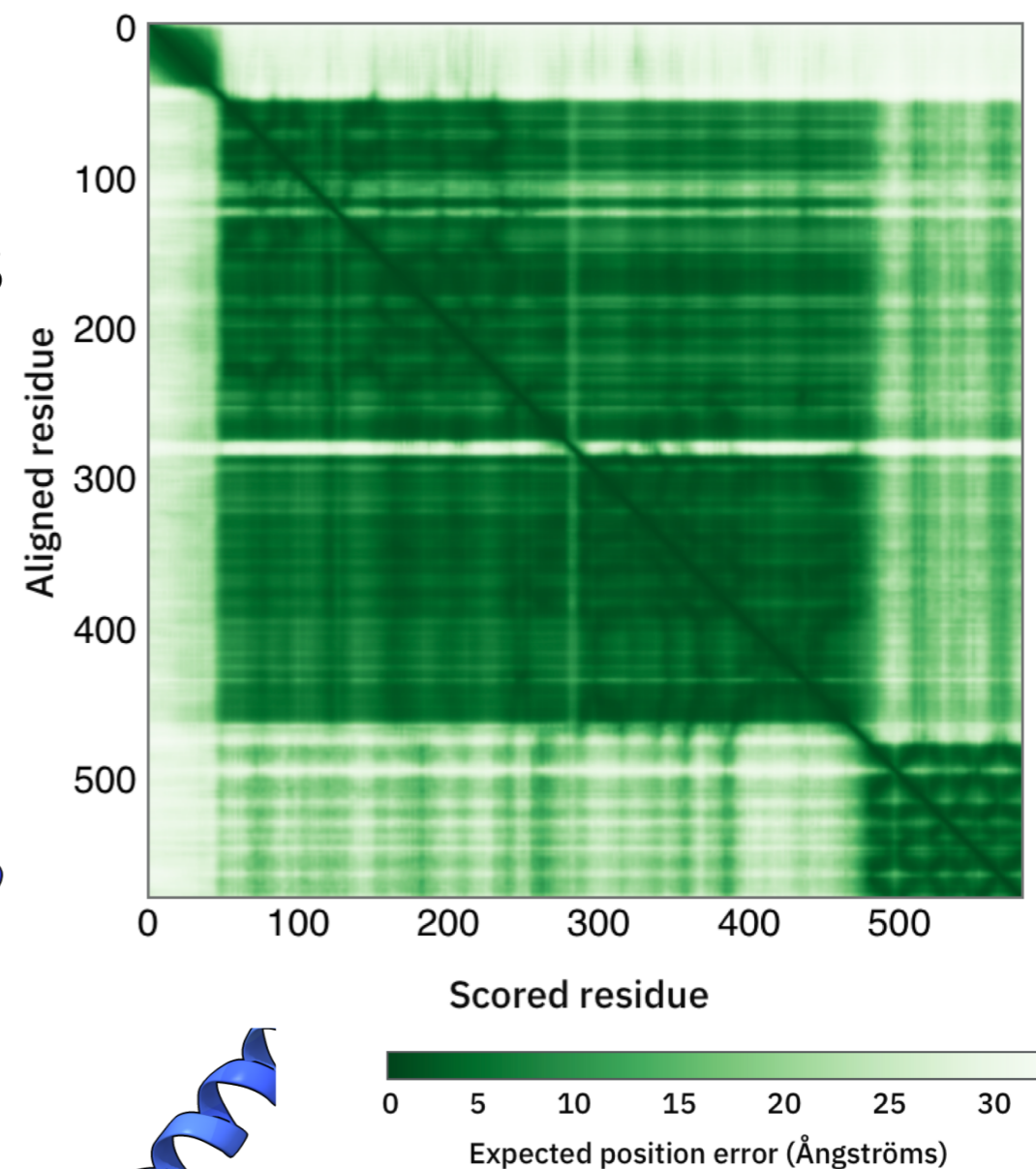
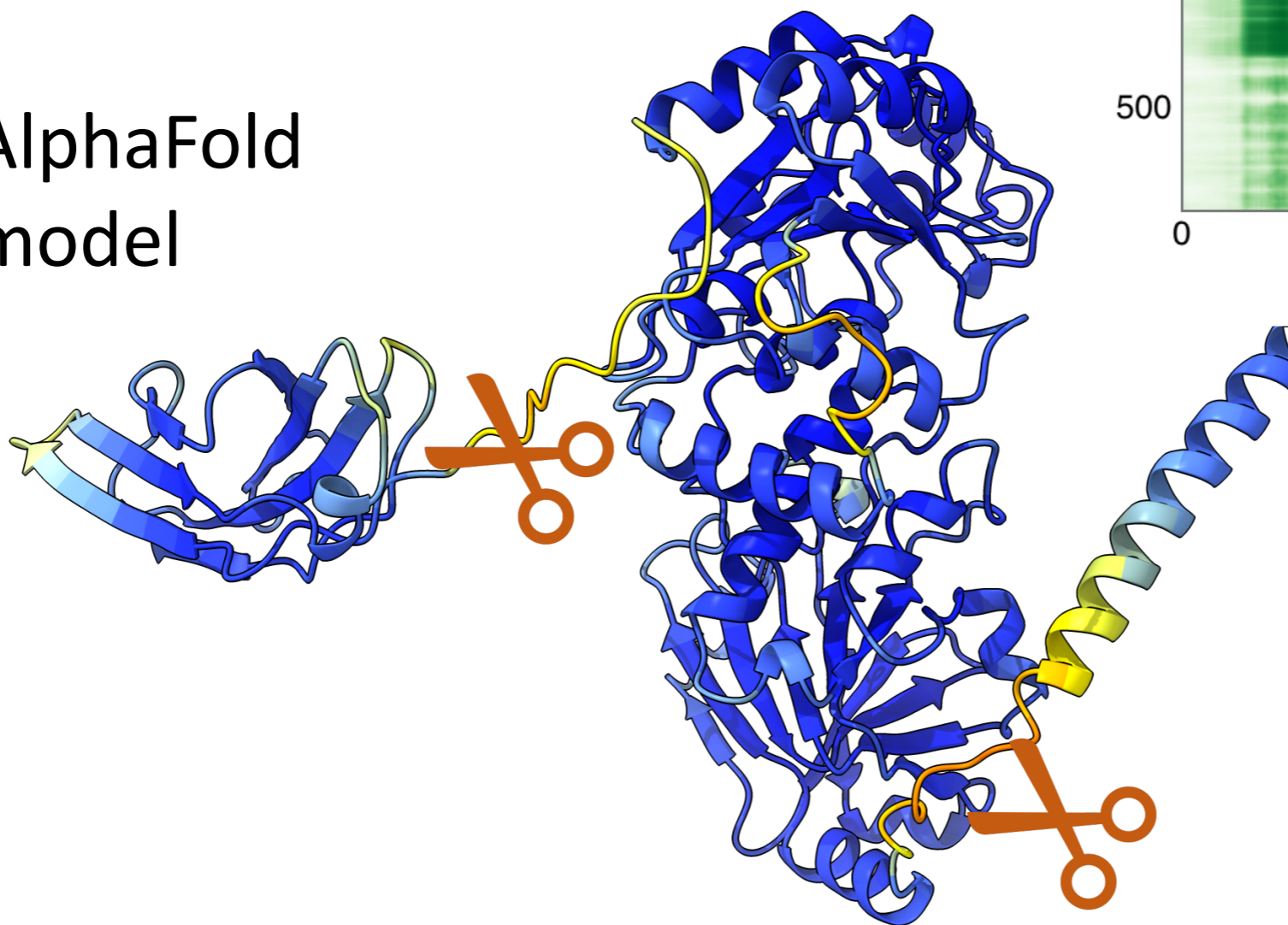
AlphaFold
model



Process prediction

- Remove low confidence parts
- Break model into domains.

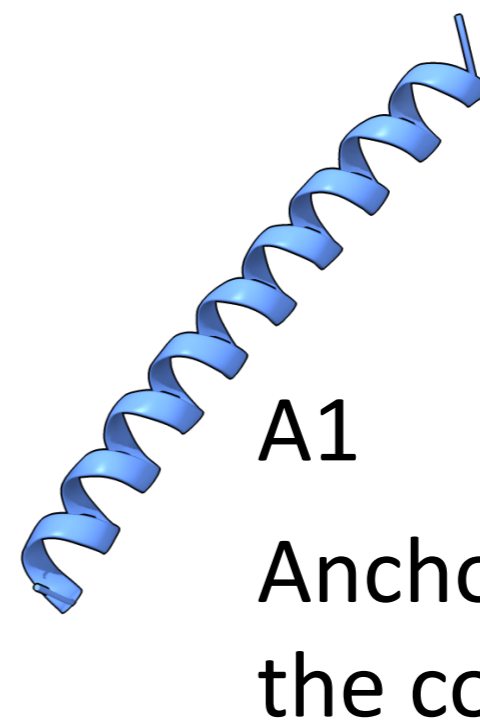
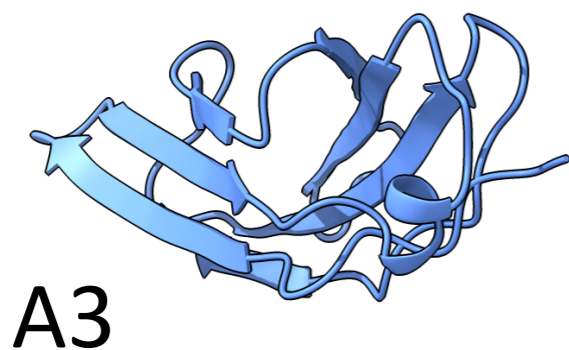
AlphaFold
model



Process prediction

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

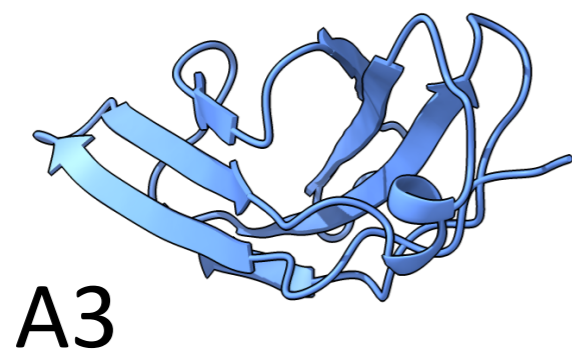
Processed
AlphaFold
model



Process prediction

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

Processed
AlphaFold
model



Do MR search with 2
components

Molecular replacement

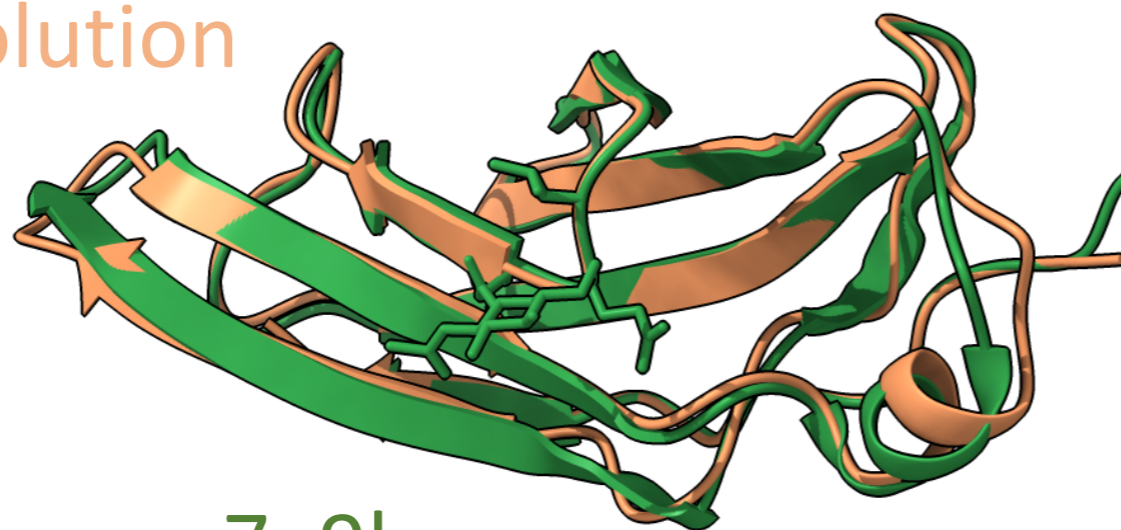
Phaser has found 1 MR solution(s).

Top LLG: 4528.201

Top TFZ: 34.6

Spacegroup: P 21 21 21

MR
solution



Molecular replacement

Phaser has found 1 MR solution(s).

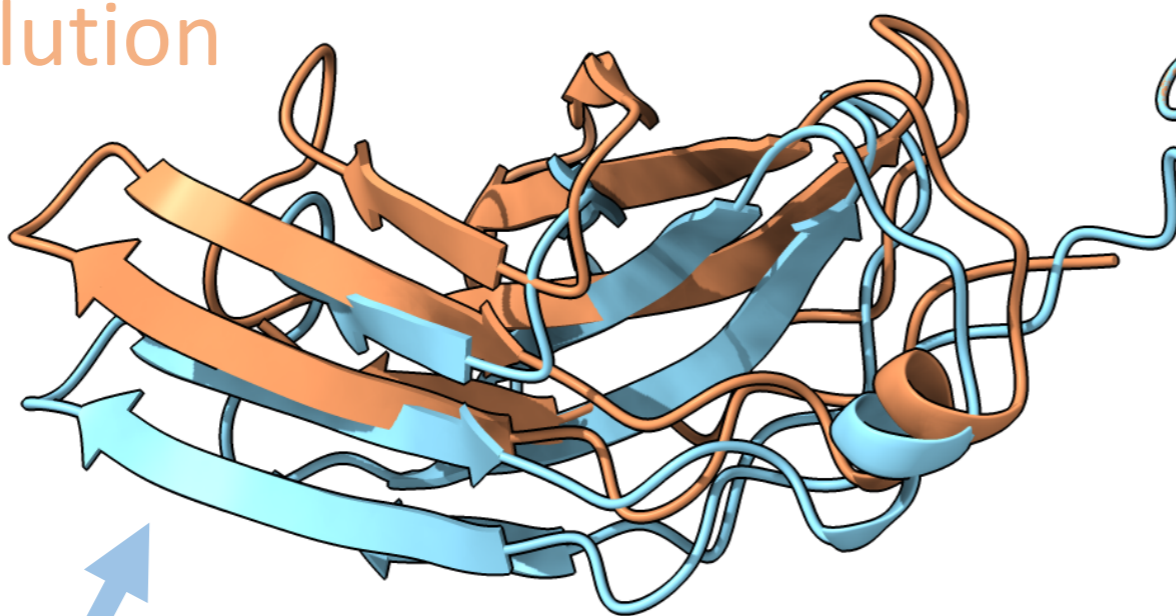
Top LLG: 1884.960

Top TFZ: 39.1

Spacegroup: P 21 21 21

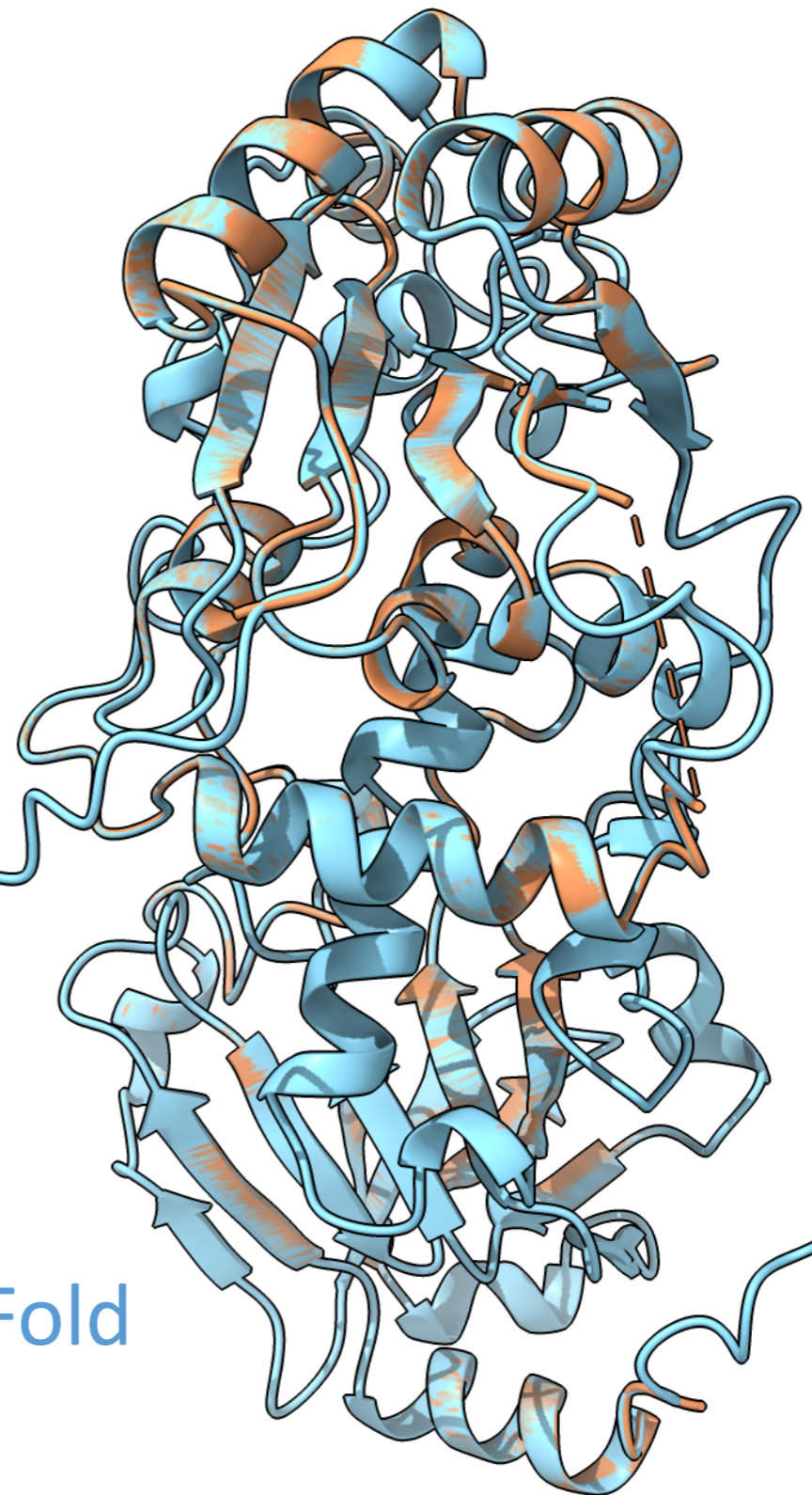
If using AF2
model “as is”

MR
solution



Domain is shifted

AlphaFold
model

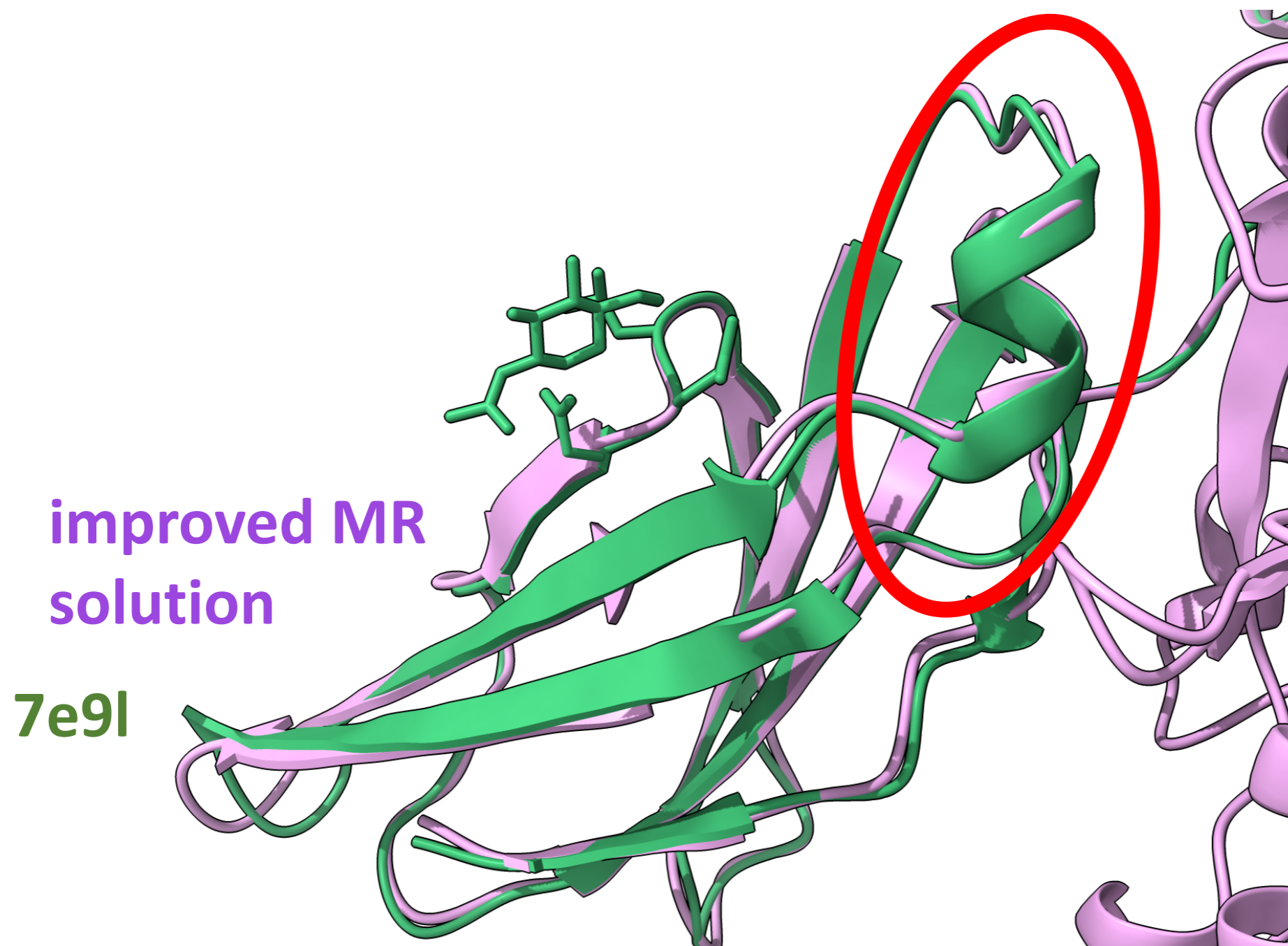


Automatically rebuild fragments



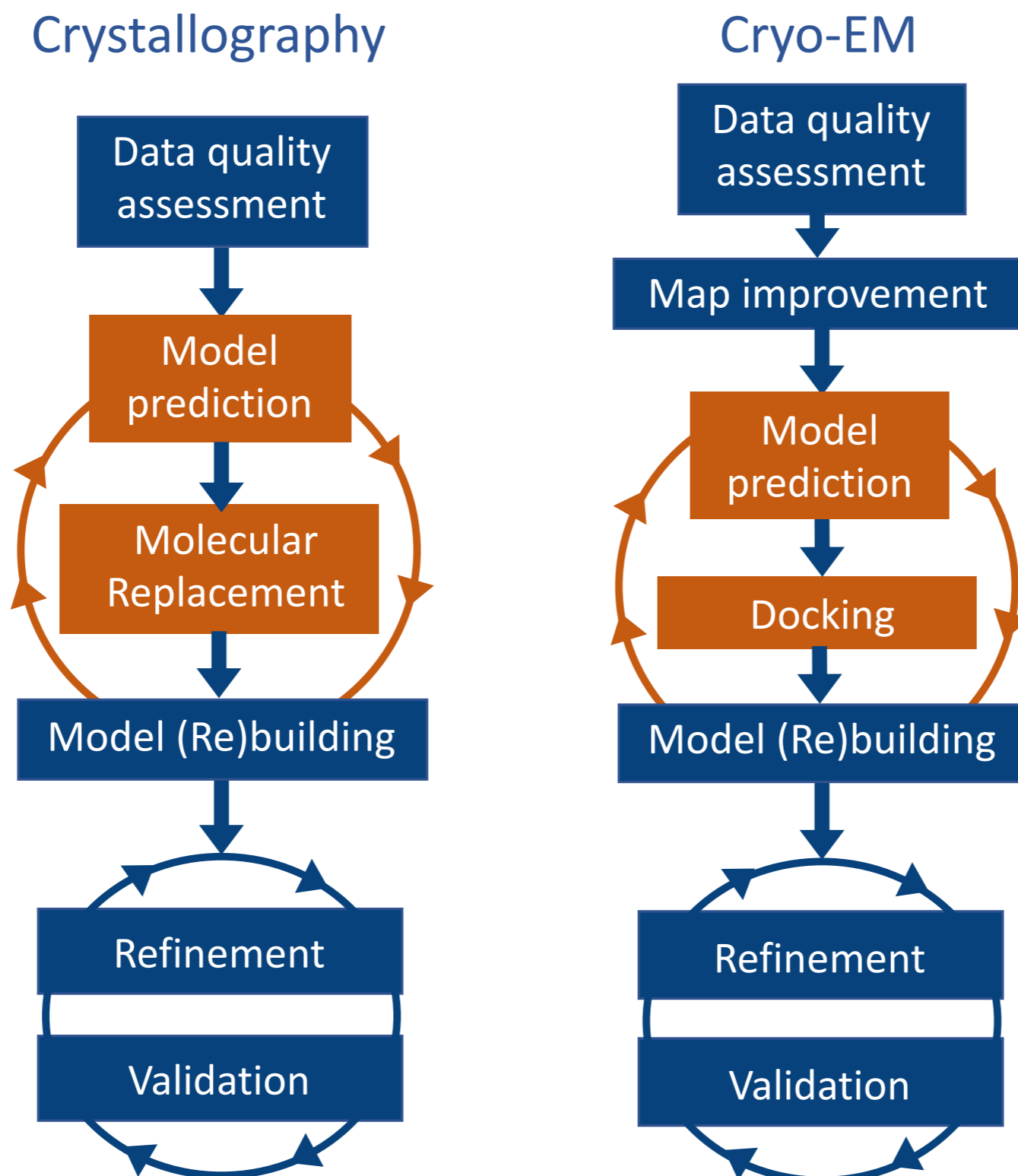
Fix deviating part automatically with `phenix.fit_loops`.

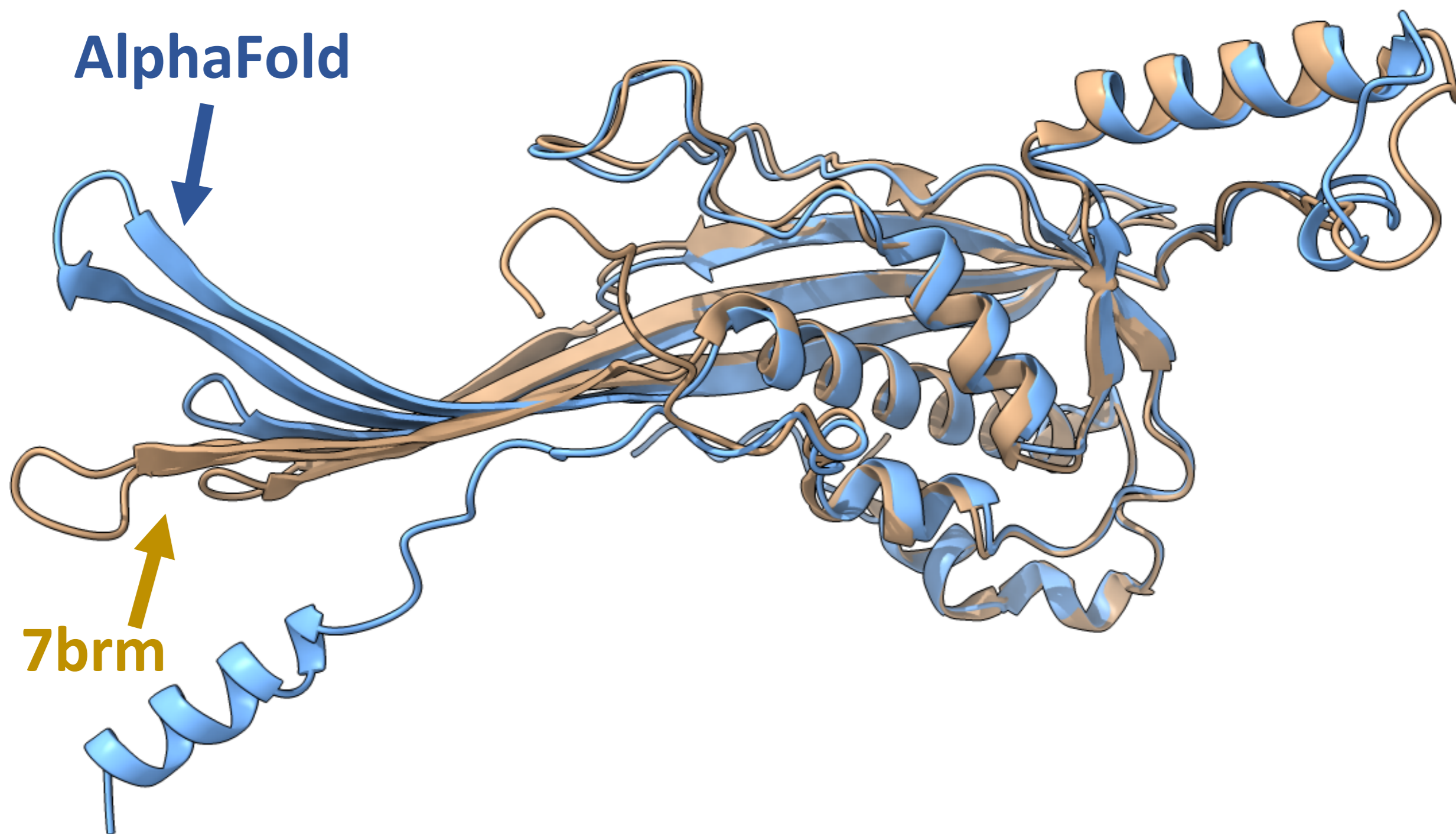
Automatically rebuild fragments



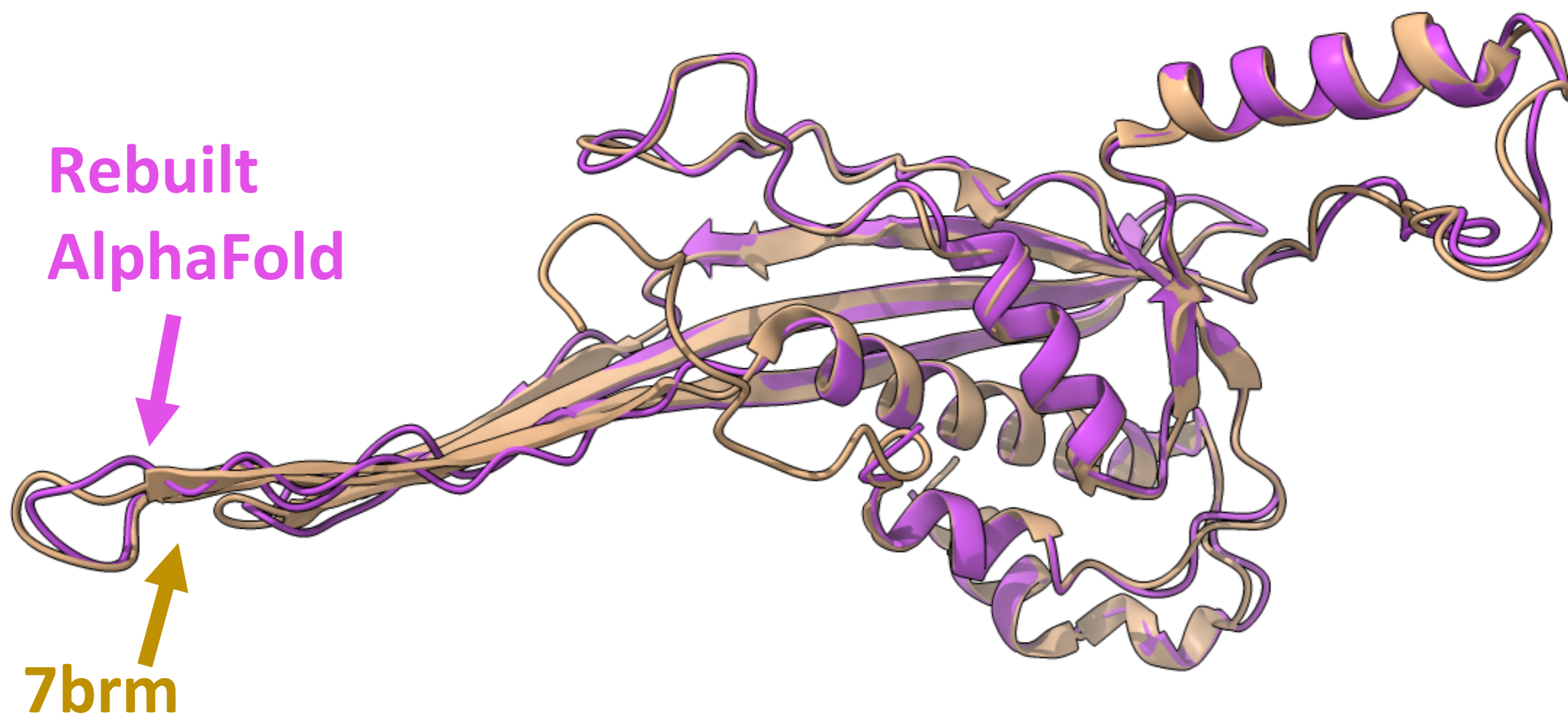
Using predicted models

New approach: Iterate prediction and model building



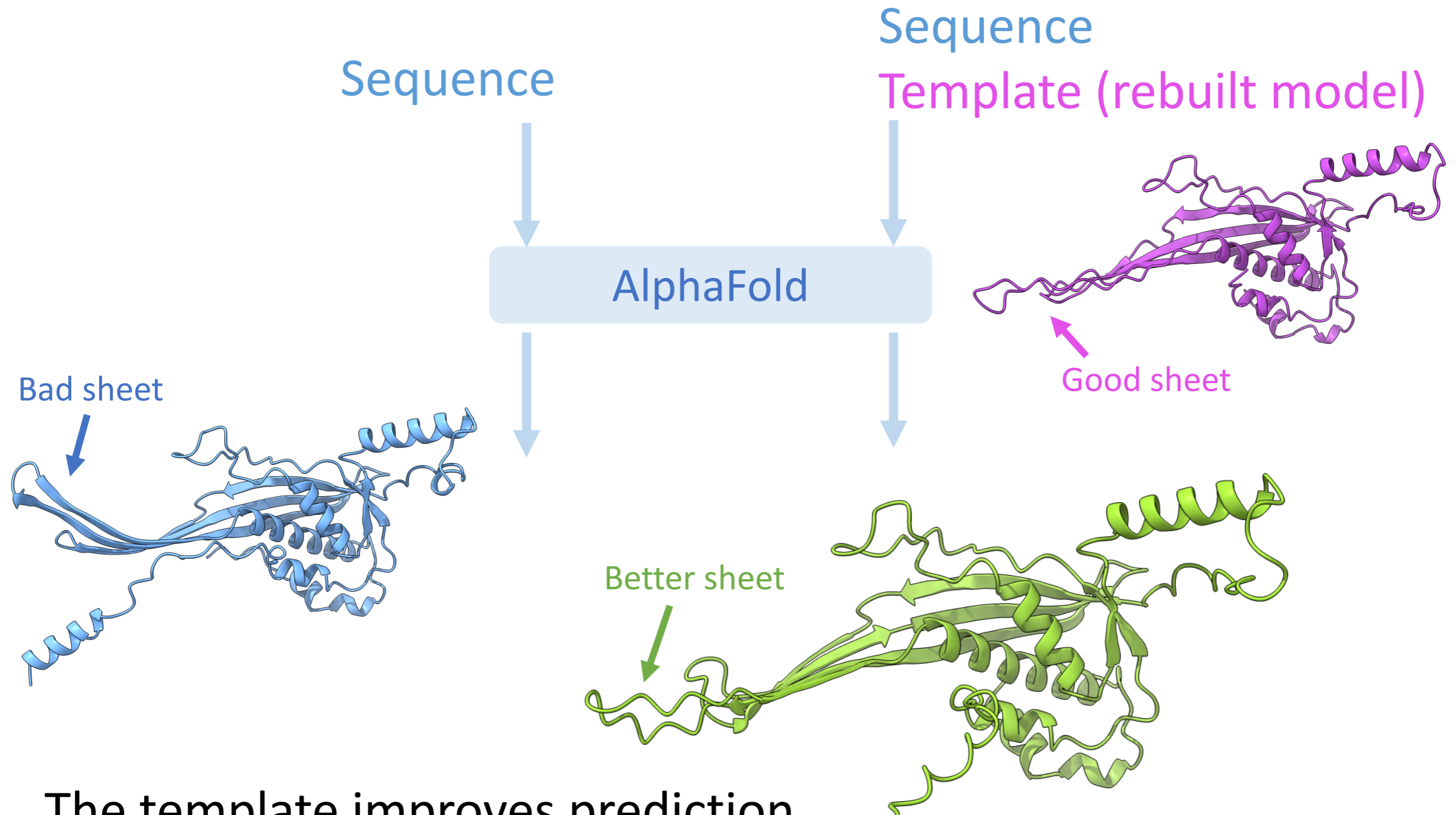


Improving prediction



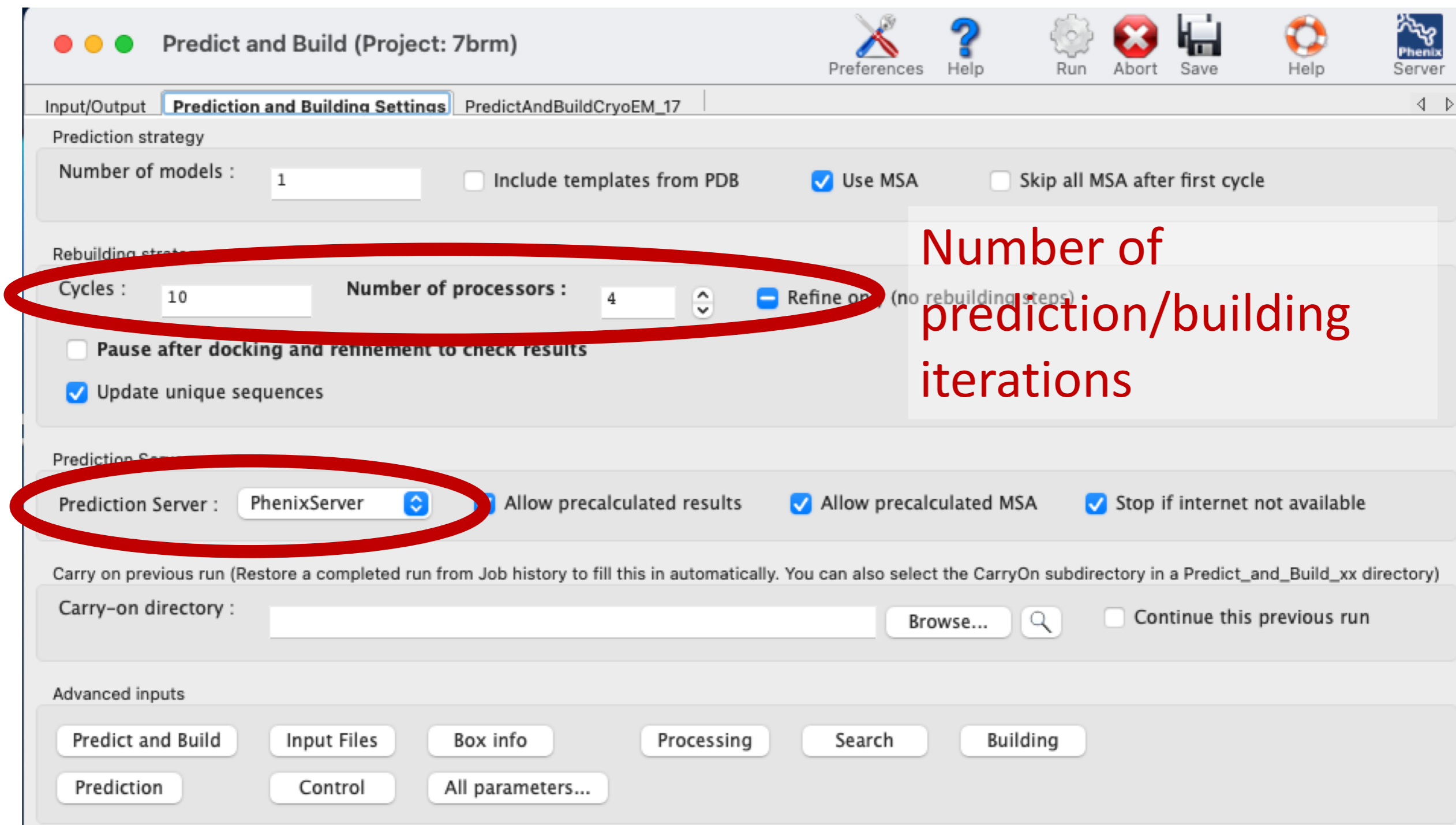
What happens if we use the rebuilt model as template for AlphaFold?

Improving prediction



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

Iterate with Predict and Build



Predict and Build (Project: 7brm)

Input/Output Prediction and Building Settings PredictAndBuildCryoEM_17

Prediction strategy

Number of models : 1 Include templates from PDB Use MSA Skip all MSA after first cycle

Rebuilding strategy

Cycles : 10 Number of processors : 4 Refine on (no rebuilding steps)

Pause after docking and refinement to check results

Update unique sequences

Prediction Server

Prediction Server : PhenixServer Allow precalculated results Allow precalculated MSA Stop if internet not available

Carry on previous run (Restore a completed run from Job history to fill this in automatically. You can also select the CarryOn subdirectory in a Predict_and_Build_xx directory)

Carry-on directory : Browse... Continue this previous run

Advanced inputs

Predict and Build Input Files Box info Processing Search Building

Prediction Control All parameters...

Number of prediction/building iterations

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 Validation 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)