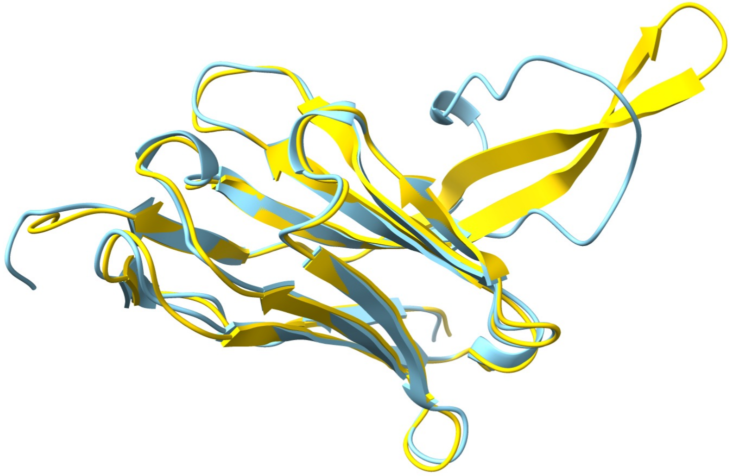
Using AlphaFold predictions for structure determination

CBMS Structural Biology Workbenches,
Sept 2024, Brookhaven National Laboratory, NY

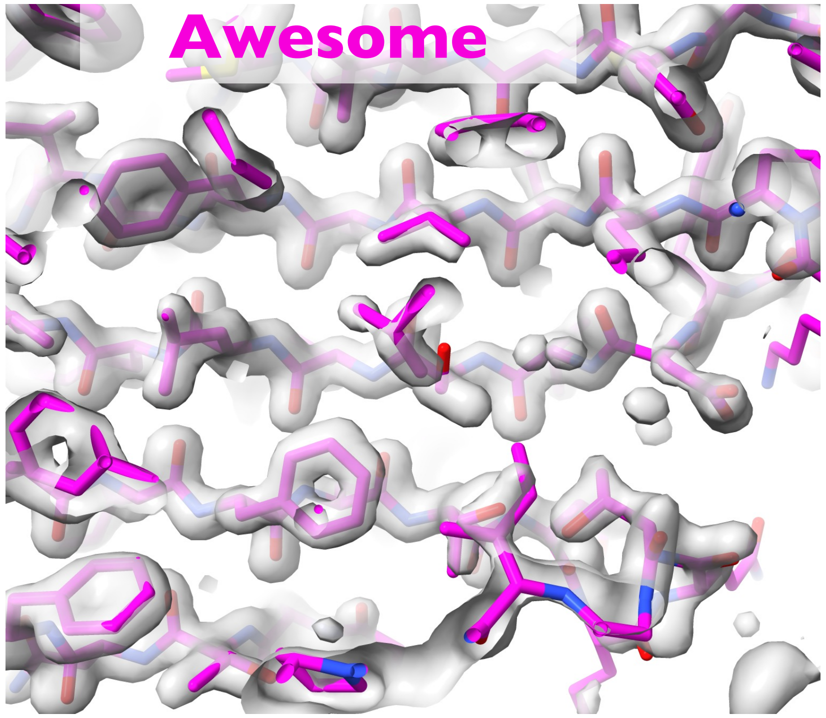


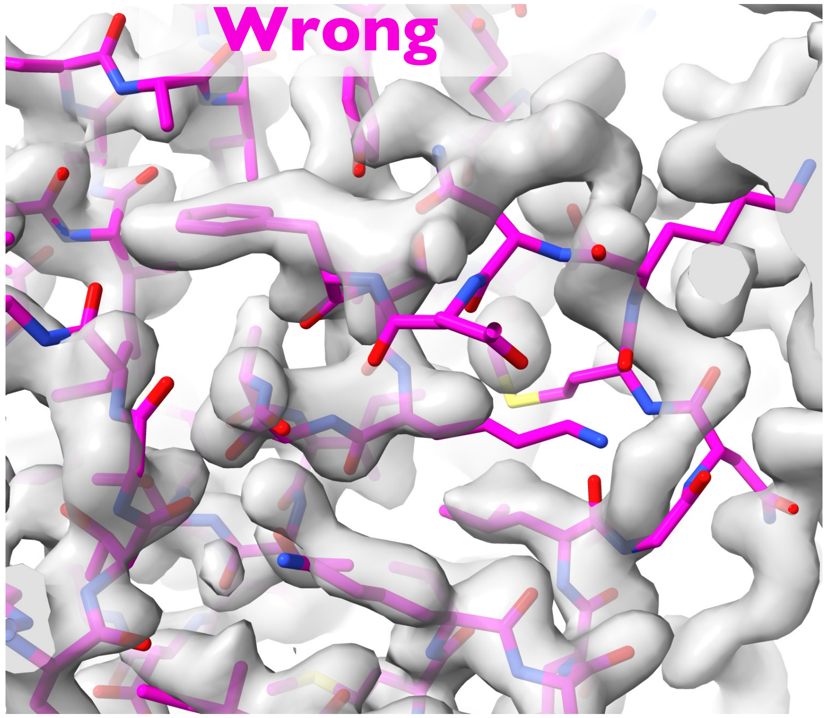
Oleg Sobolev

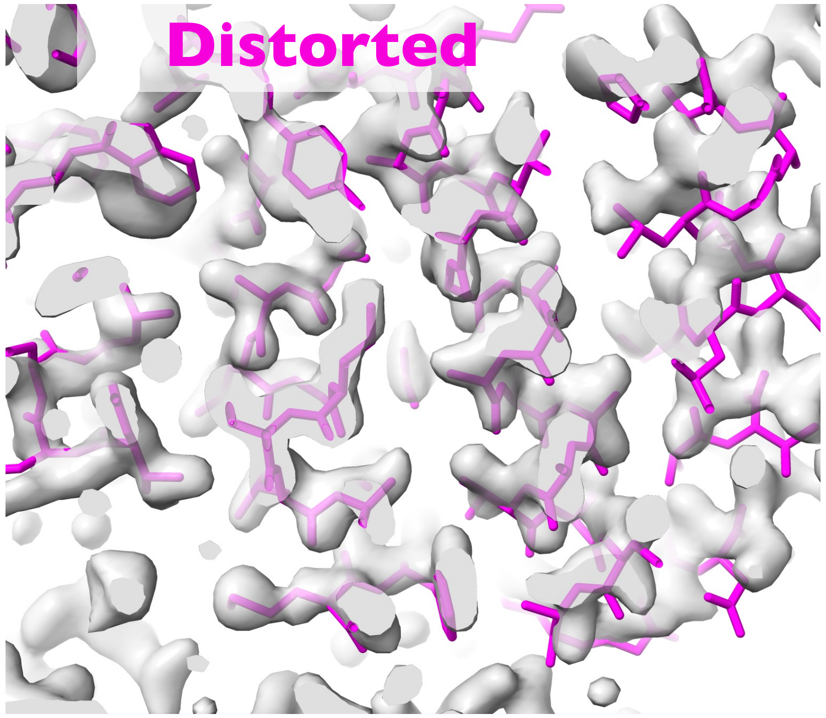
Lawrence Berkeley National Laboratory

AlphaFold predictions are great hypotheses

AlphaFold models can be….


Awesome

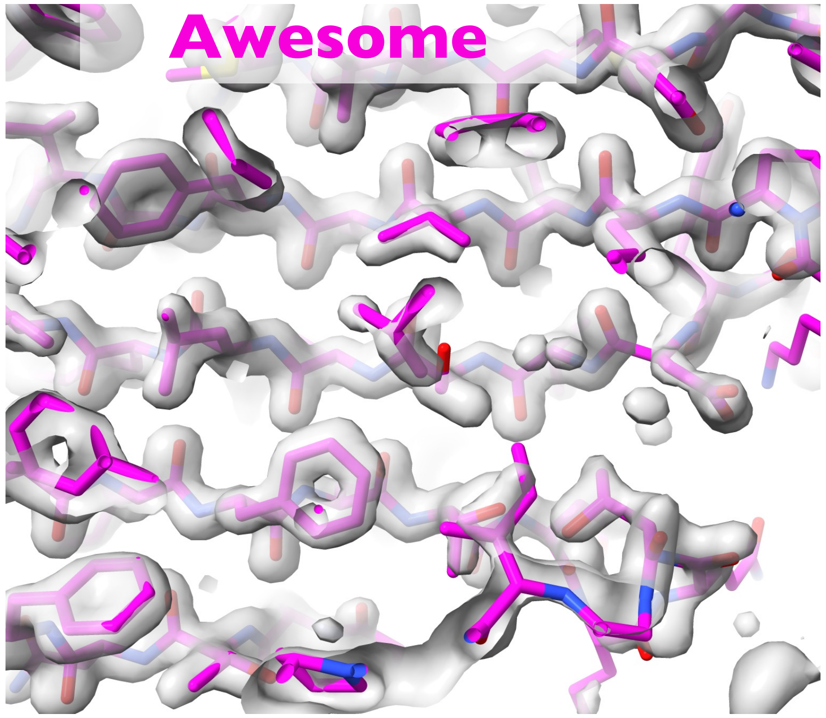

Wrong

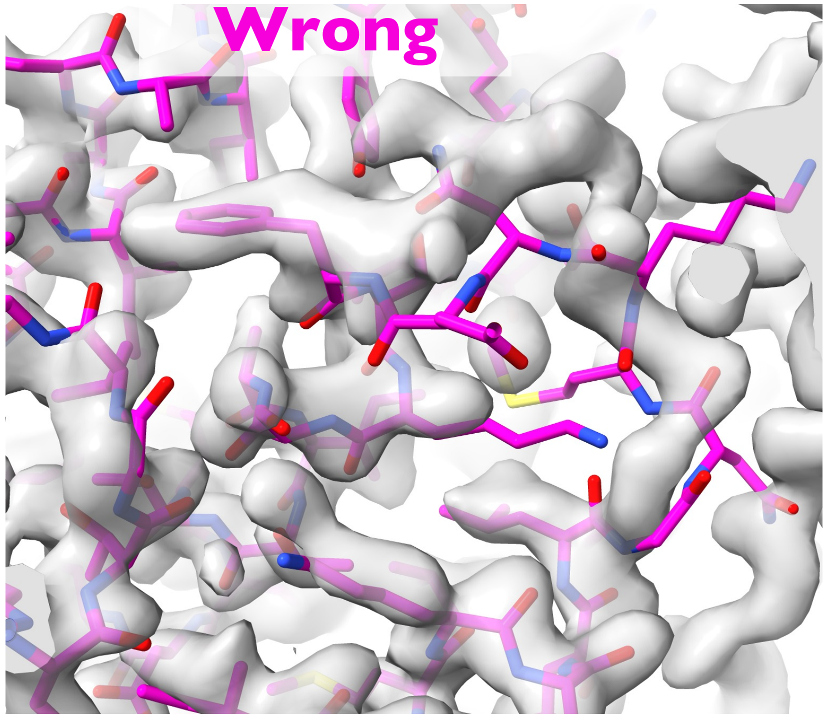

Distorted

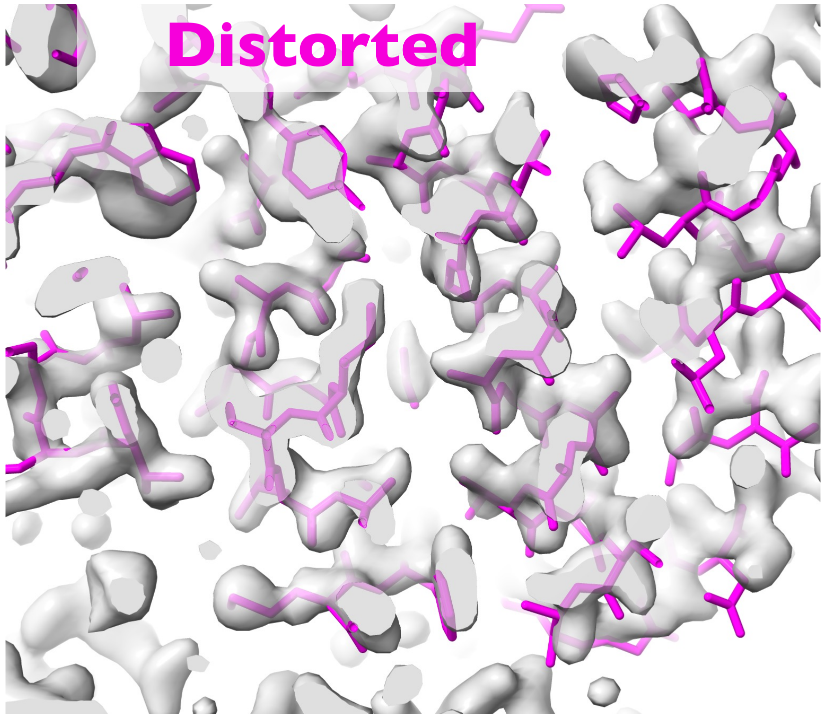
AlphaFold predictions and confidence estimates

Residue-specific confidence (pLDDT) identifies where errors are more likely

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



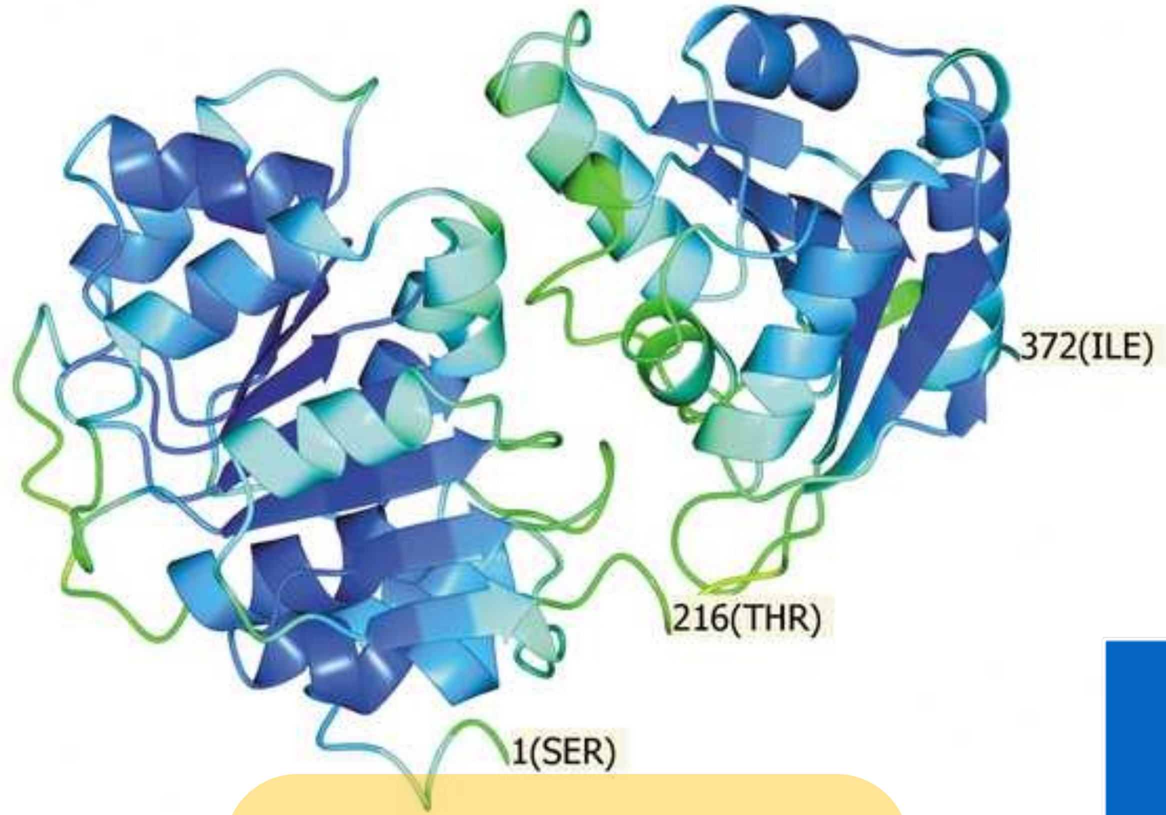




Terwilliger et al. (2024), AlphaFold predictions are valuable hypotheses, and accelerate but do not replace experimental structure determination. Nature Methods 21, 110-116.

AlphaFold confidence measure

(pLDDT, Predicted difference distance test)



Confidence:

Blue: > 90

Green: 80 - 90

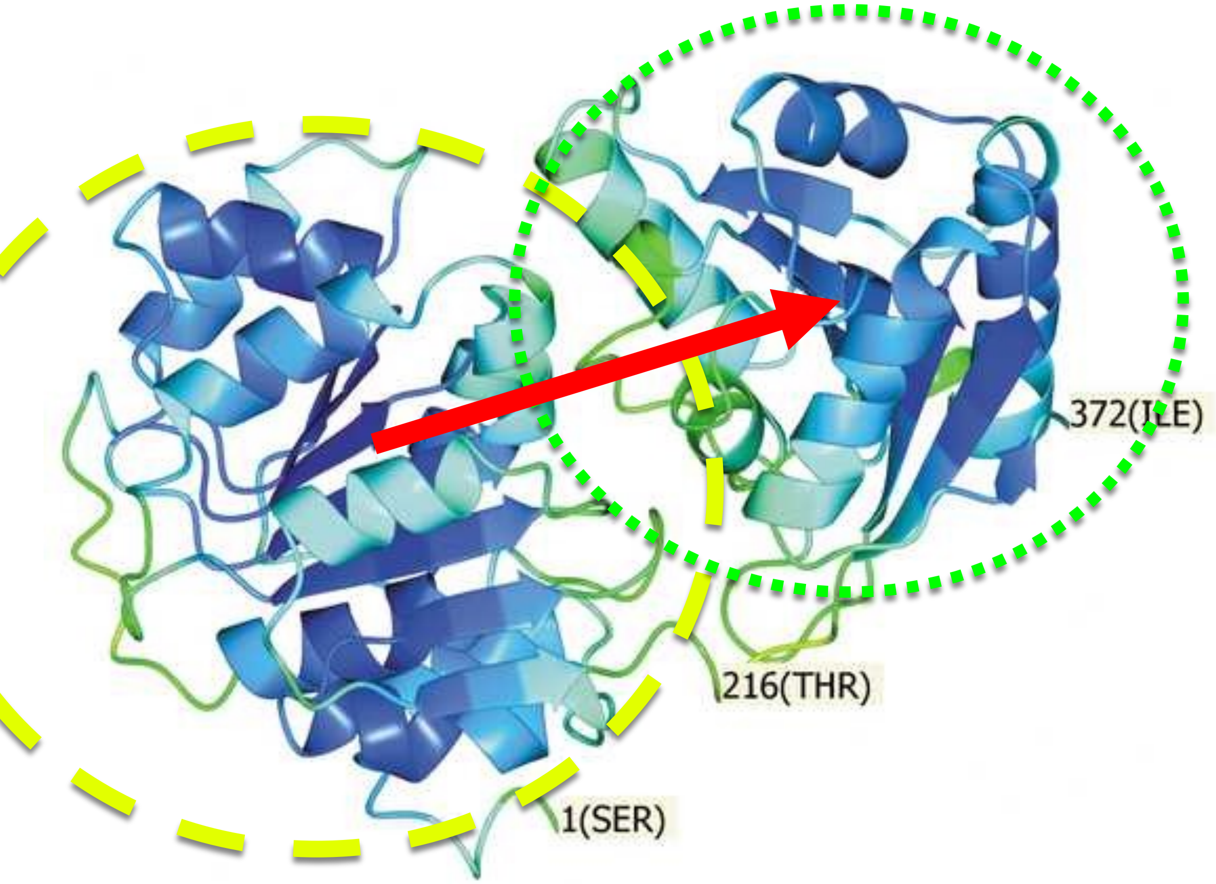
AlphaFold confidence (pLDDT)	Median prediction error (Å)	Percentage with error over 2 Å
>90	0.6	10
80 - 90	1.1	22
70 - 80	1.5	33
<70	3.5	77

AlphaFold prediction for RNA helicase

(PDB entry 6i5i)

Oeffner et al. (2022). Acta Cryst. D78, 1303-1314

PAE matrix (Predicted aligned error)



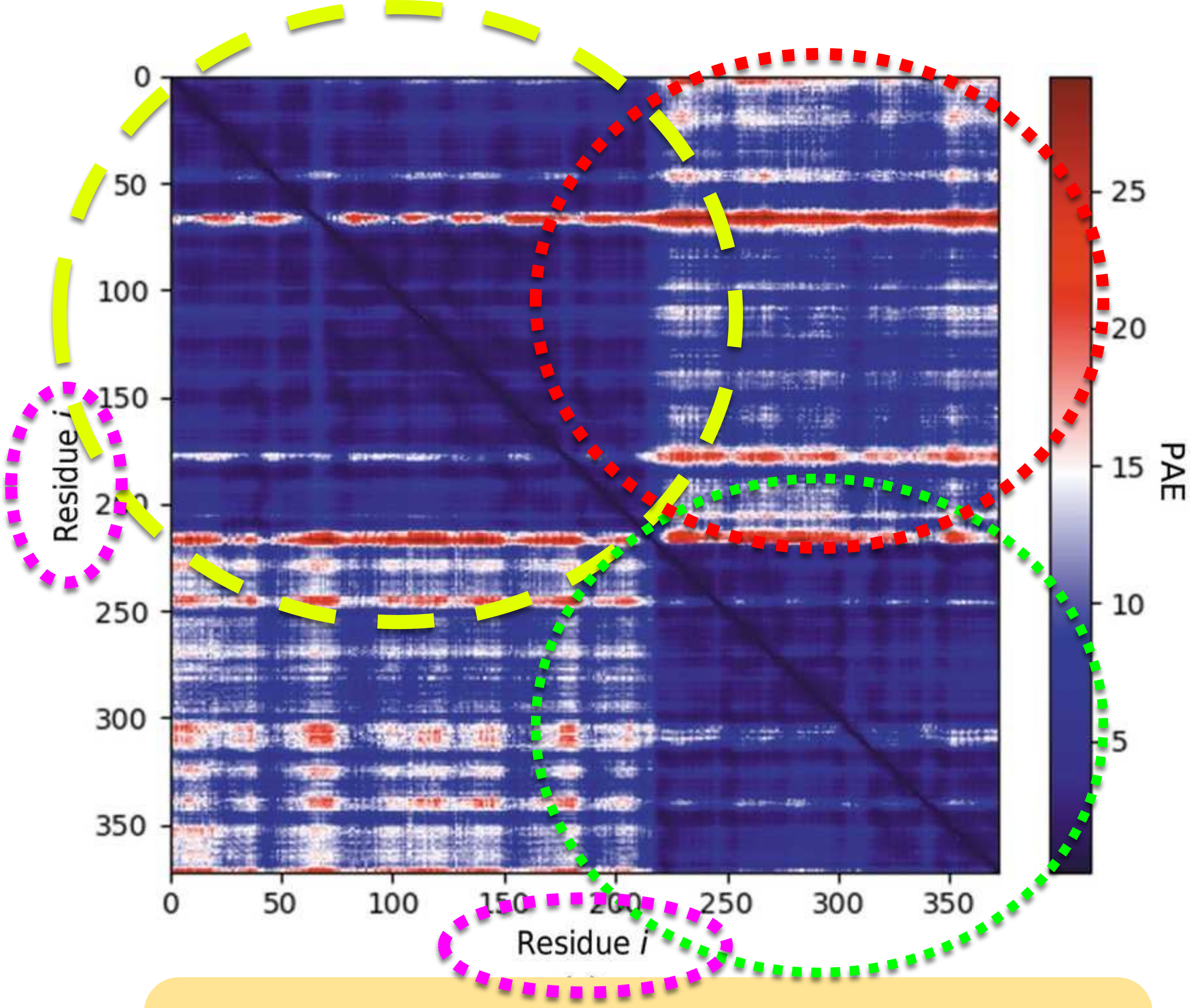
Confidence:

Blue: > 90

Green: 80 - 90

AlphaFold prediction for RNA helicase
(PDB entry 6i5i)

PAE matrix identifies accurately-predicted domains

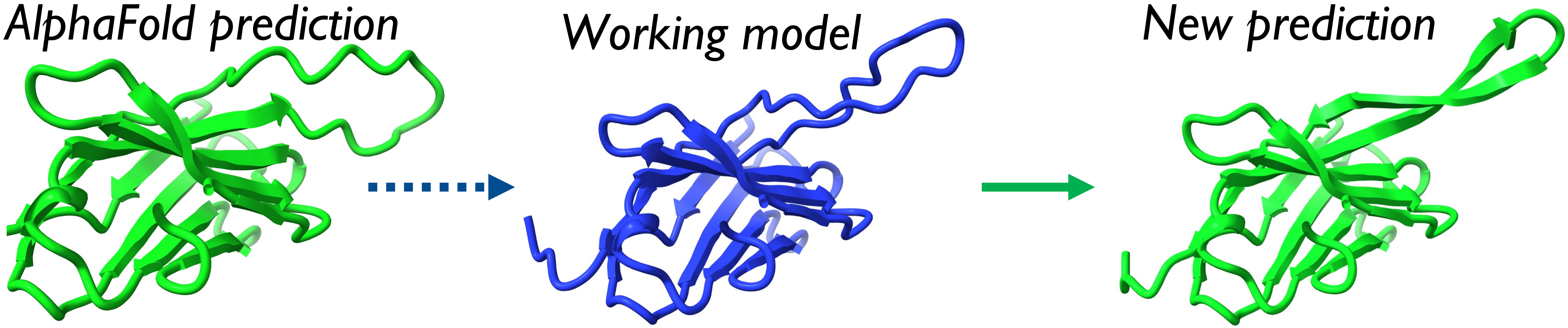


Dark blue: uncertainty in relative positions < 5 Å

Using your best model as a template in AlphaFold prediction

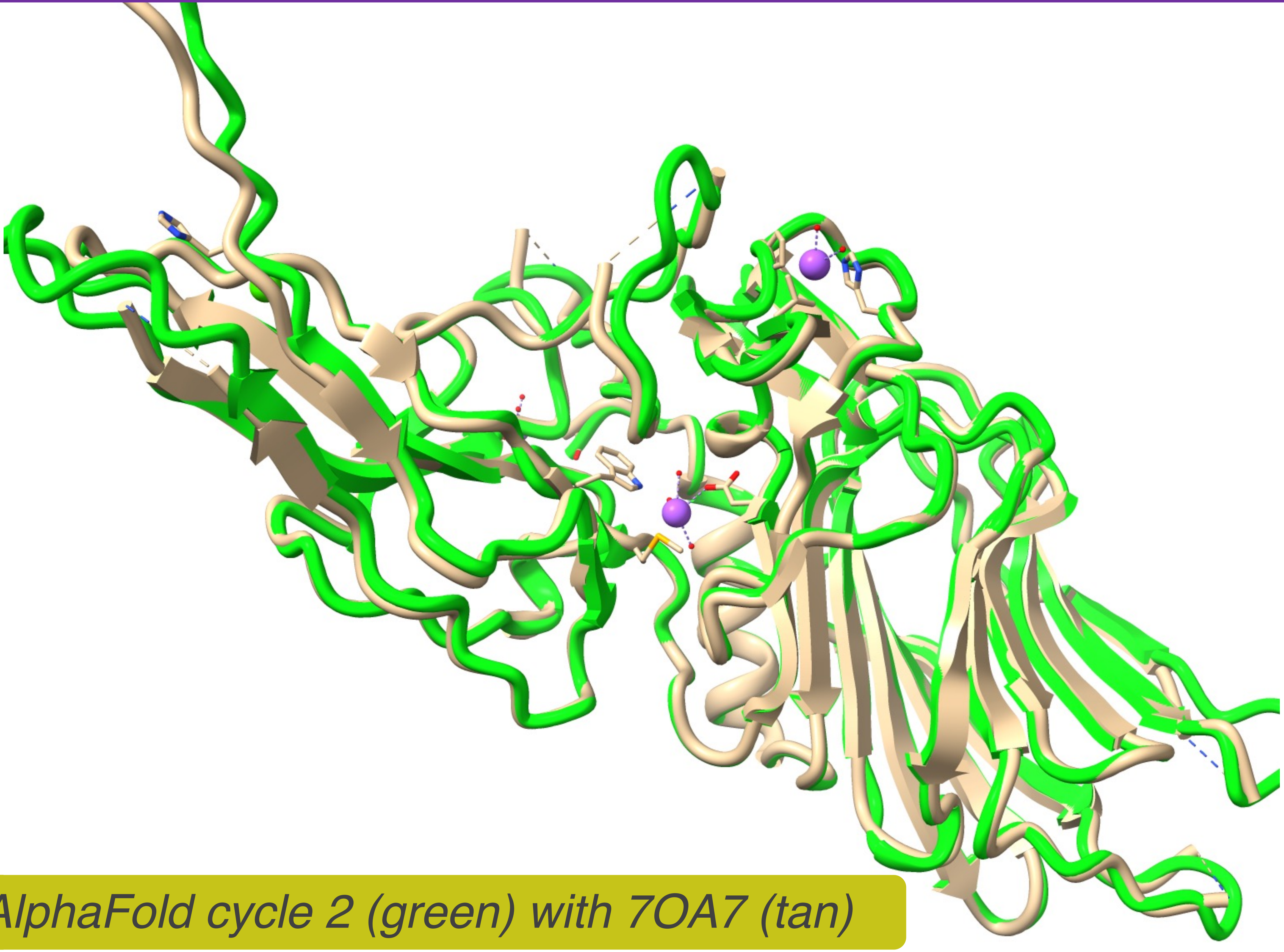
Why?

Because your new prediction might be better than your model ...and better than your original AlphaFold prediction



Improving AlphaFold prediction using partial models as templates

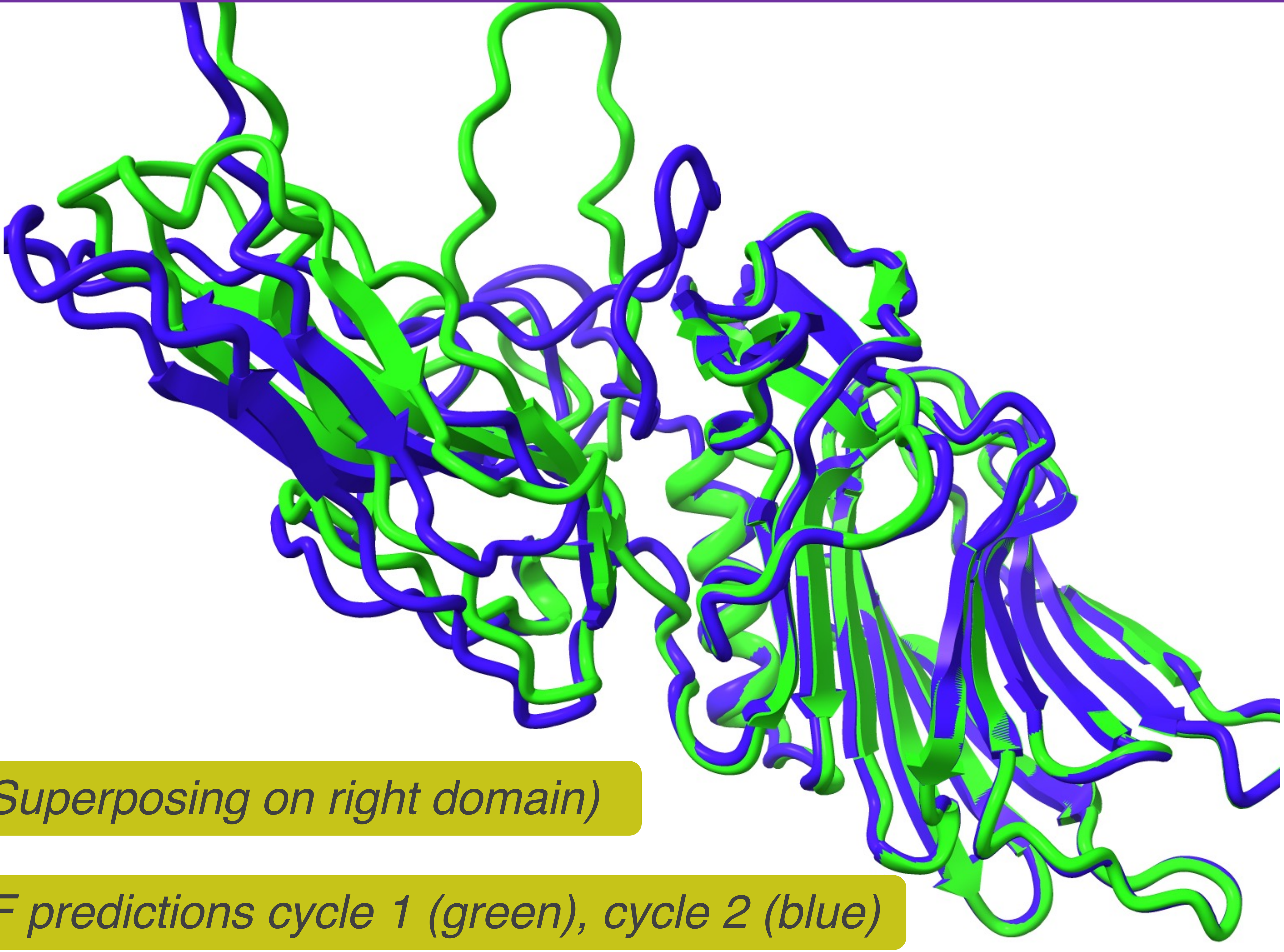
(X-ray crystallography)



AlphaFold cycle 2 (green) with 7OA7 (tan)

Improving AlphaFold prediction using partial models as templates

(X-ray crystallography)

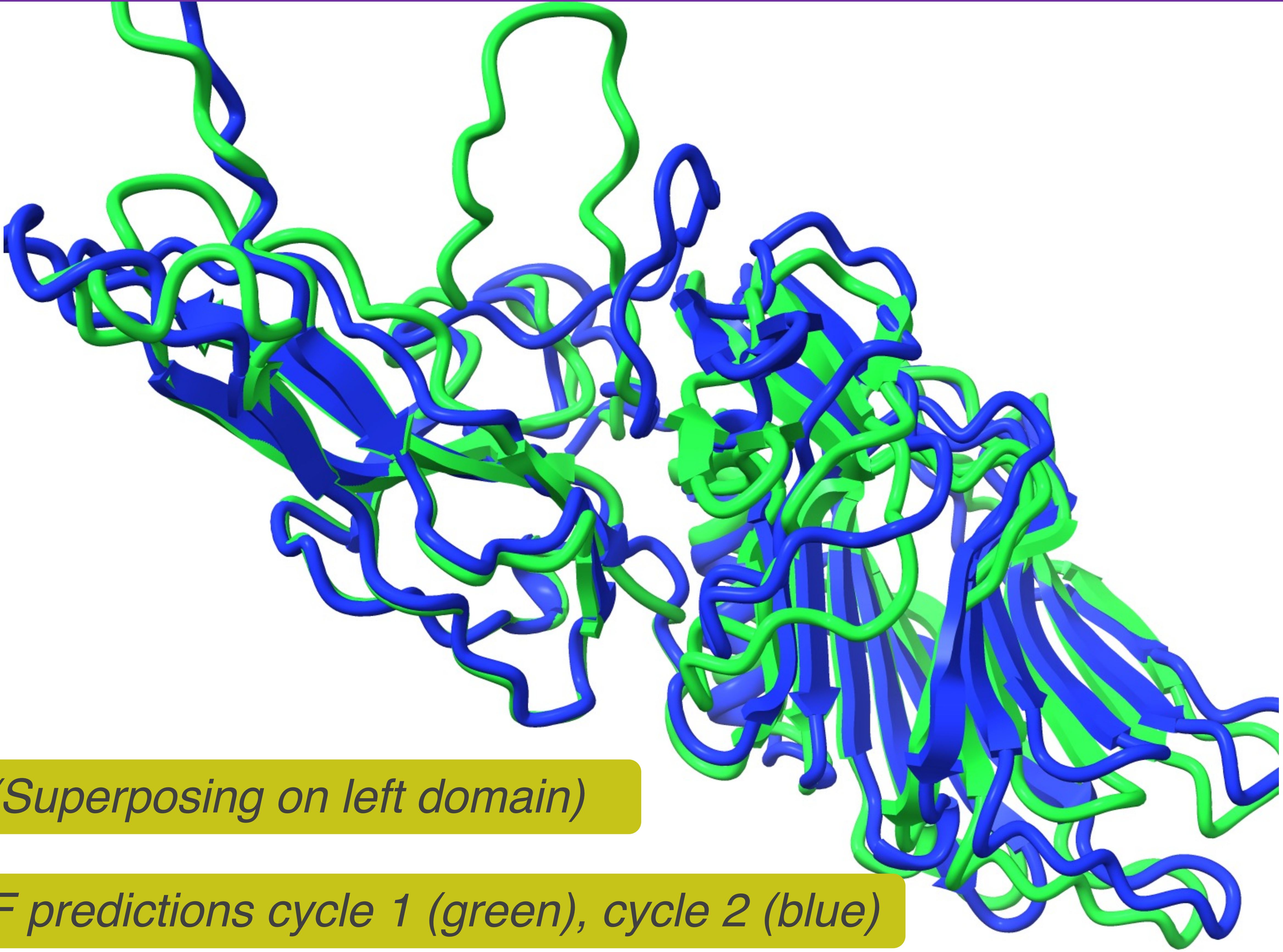


(Superposing on right domain)

AF predictions cycle 1 (green), cycle 2 (blue)

Improving AlphaFold prediction using partial models as templates

(X-ray crystallography)



(Superposing on left domain)

AF predictions cycle 1 (green), cycle 2 (blue)

Strategy for structure determination in the AlphaFold era

1. Predict your structure

→ *Design your experiment based on predicted models*
(choose experimental approach, consider trimming at domain boundaries

2. Solve your structure

→ *Cryo-EM or X-ray MR with trimmed predicted model, SAD*

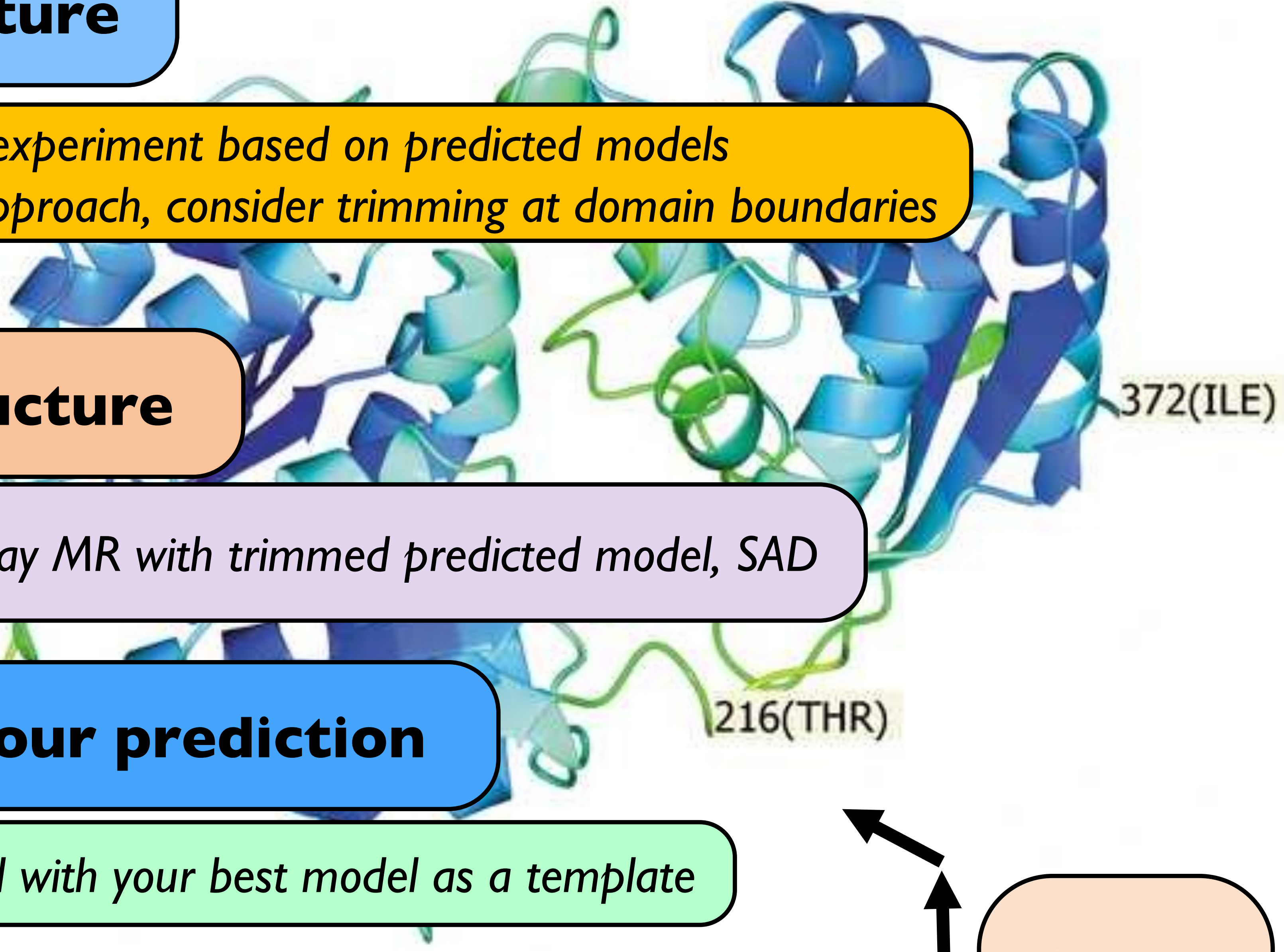
3. Update your prediction

→ *Run AlphaFold with your best model as a template*

4. Improve your structure

→ *Use your new predictions as hypotheses*

Iterate



Phenix AlphaFold prediction server

Available from the Phenix GUI

Predicts structures of protein chains

(one at a time)

Can use a template to guide the prediction

You do not need an MSA (multiple sequence alignment) if you supply a template

The template should not be an AlphaFold model

Many thanks for AlphaFold, ColabFold scripts, and the MMseqs2 server for MSAs

Process predicted model

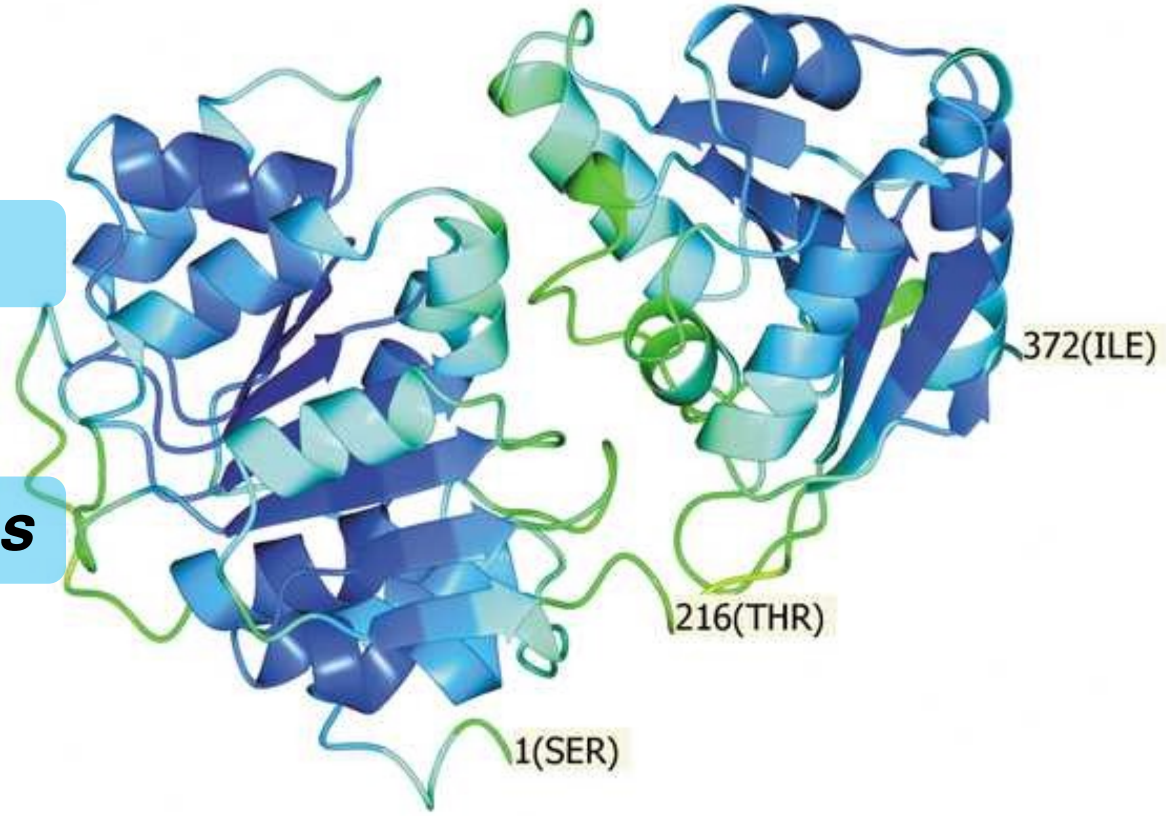
Convert pLDDT to B-value

Trim low-confidence parts of model

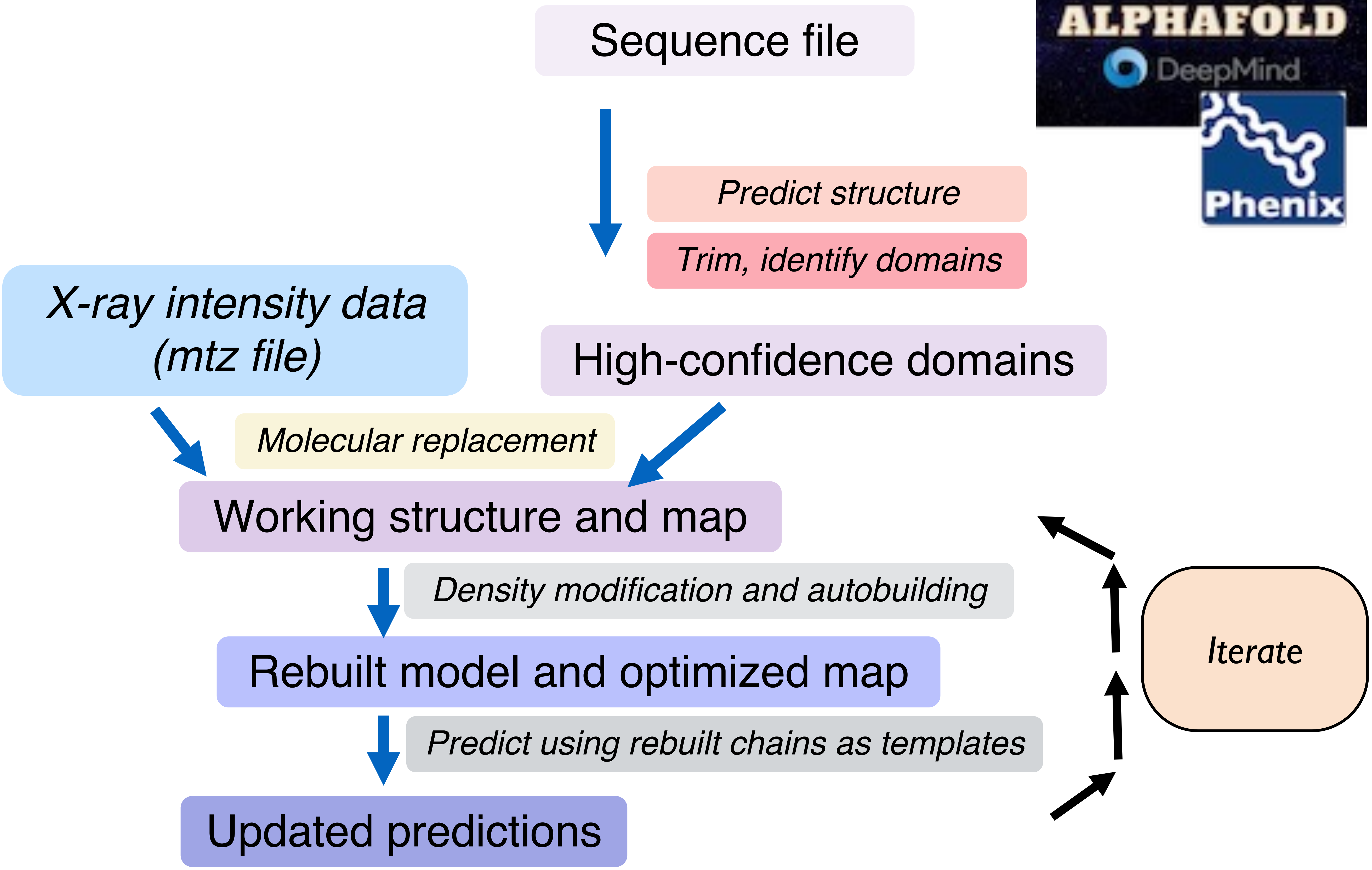
Identify high-confidence domains

Compact high-confidence regions

Groupings of residues with low PAE values



X-ray structure determination with AlphaFold



Sequence file

Predict structure

Trim, identify domains

X-ray intensity data (mtz file)

High-confidence domains

Molecular replacement

Working structure and map

Density modification and autobuilding

Rebuilt model and optimized map

Predict using rebuilt chains as templates

Updated predictions

Iterate

Input and output from structure determination with AlphaFold

Input

Experimental data (maps or X-ray data)

Contents of asymmetric unit (sequence file)

Output

Rebuilt model
Optimized map

Docked predicted models

Map and model ready for next steps

Useful as high-quality reference models

Phenix tools for structure determination with AlphaFold

PredictModel (Predict with AlphaFold)

ProcessPredictedModel (Trim and identify domains)

AlphaFold models

ResolveCryoEM, LocalAnisoSharpen (map improvement)

EMPlacement, DockInMap (Docking of single, multiple chains)

DockAndRebuild (Morphing and rebuilding)

RealSpaceRefine (Refinement)

Cryo-EM

Phaser-MR (Molecular replacement)

AutoBuild (Density modification and rebuilding)

Phenix.refine (Refinement)

X-ray

PredictAndBuild (Prediction and structure determination)

Full automation

The Phenix Project

Lawrence Berkeley Laboratory

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

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, Vincent Chen, Michael Prisant, Christopher Williams,

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