





West Coast Structural Biology Workshop, March 2023

Using predicted models in Phenix



Dorothee Liebschner

Lawrence Berkeley Laboratory













The Phenix Project





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



Example for a prediction (POMGNT2)

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

(Example from Phenix AlphaFold MR tutorial)



Predictions come with confidence measures

1. pLDDt (predicted Local Distance Difference Test)

- per-residue confidence measure
- scales from 0 100
- pLDDt > 90: predicted with high accuracy



Predictions come with confidence measures

2. Predicted aligned error (PAE)

- Relative domain positions \rightarrow Large-scale topology
- Dark green: the predicted relationship between this pair of residues is likely to be accurate





- Suggests 3 domains
- Mutual configuration is not clear



Predicting a model in Phenix

	AlphaFold model prediction (Project: presentation)	Preferences	Pelp Run	Abort Save	Help	Phenix Server					
	Input/Output Prediction Settings					4 Þ					
	PredictModel: Predict protein chains with AlphaFold Job title :										
Seque	nce (template, multiple sequ	lence	alignn	nent)							
	Required: Sequence file Ontional: Templates MSA					- 11					
	File pathFormatData typeQ/Users/dcliebschner/Desktop/Projects/15_confSequence	file	>			18					
	Add file Remove file Modify file data type										
	Composition (load a sequence file and/or paste and edit sequencesthe composition show	n 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 : EISEVQLVESGGGLVQPGGSLRLSCAASGFNIYSSSIHWVRQAPGKGLEWVAY	IYSYSGYTSYADSV	KGRFTISADTSKNTA	Copies : 1	•						
	Add another Delete last										
	Project Directory: /Users/dcliebschner/Desktop/Projects/15_conference	es-presentatio	ns/2023_03b_WCS	BW/presentation	٩						
	O Idle		Project: pres	entation							

Do I need to install AlphaFold?

Phenix Server or Colab for running AlphaFold

e e AlphaFold	model predictio	on (Project: presentation)	Preferences	P Help H	Sun Abort	Save	Co Help	Phenix Server		
Input/Output Prediction Se	ettings							4		
Prediction strategy										
Number of models :	5	✓ Include templates from PDB	🗸 Use MSA	Skip	all MSA afte	er first cyc	le			
Prediction Server										
Prediction Server : Phe	enixServer 📀	Allow precalculated results	✓ Allow precalc	ulated MSA	🗹 Stop	if internet	not available	Ru		
Additional prediction inputs (Note: this GUI is only for prediction, not docking or model-building)										
Predict and Build	Input Files	Prediction Control	All parameters							

Run AlphaFold on

- Phenix Server
- Colab

	Phenix prediction server status					
	SERVER STATUS: UP					
	RUNNING JOBS: 2 WAITING JOBS: 0					
	WAIT TIME: 0 min					
This server is used in PredictModel and PredictAndBuil Colab can be used as an alternative						
	✓ ОК					
	Composition tipad a s					



model

Predicted models are great hypotheses...

7s5L, 1.58 Å, X-ray diffraction



Predictions can be very accurate.

Terwilliger TC et al. AlphaFold predictions: great hypotheses but no match for experiment. BioRxiv. 2022



...but (still) no match for experiment

7s5L, 1.58 Å, X-ray diffraction





AlphaFold model

Some parts may be inaccurate

(even when predicted with high confidence)

Parts may be distorted

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



Predicted models can disagree globally and locally with experimental structures.

Only ~40% of residues in the human proteome are modeled with high confidence (pLDDt).



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: Cryo-EM map (30160 – 7brm)





cryo-EM map (3.6 Å)

sequence



Get a prediction



>chain ' A' XXXXXXXXXXXXXXXXXCLTAPPKEAARPTLMPRAQSYKDLTHLPAP TGKIFVSVYNIQDETGQFKPYPASNFSTAVPQSATAMLVTALKDS RWFIPLERQGLQNLLNERKIIRAAQENGTVAINNRIPLQSLTAAN IMVEGSIIGYESNVKSGGVGARYFGIGADTQYQLDQIAVNLRVVN VSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNE PVMLCLMSAIETGVIFLINDGIDRGLWDLQNKAERQNDILVKYRH MS

AlphaFold model





Process prediction



AlphaFold model

Remove low confidence parts

11sk







Dock processed model





Dock processed model



Some parts don't fit into the map



Dock processed model



Some parts don't fit into the map \rightarrow fit loops and rebuild



Fit loops and rebuild



This model is close to the deposited model.



Fully automatic: predict and build

Predict and Build (Project: 7brm)	Preferences	? Help	Run	S Abort	Save	C Help	Phenix Server
Input/Output Prediction and Building Settings							4
PredictAndBuildCryoEM: Predict, dock and rebuild AlphaFold models Job title :	into a cryo-EM map						
Input files							
Required: Sequence file and 2 half maps or one full map							
Optional: Scaffold (replaces docked model), Template for prediction, Predic	ted model (can be RNA)	/DNA)					
Tip: If running all at once does not work, first density modify your map with extract unique part with MapBox, then run PredictAndBuild one chain at a t You can use ApplyNCS to apply symmetry and "Combine PDB files" to merge	h ResolveCryoEM and ime. e.						
File path			Sea		nce		
<pre>//Users/dcliebschner/Desktop/Projects/15_conf CCP4 map Map // Ma</pre>	o file		JUY	uc	ince		
\sim /Users/dcliebschner/Desktop/Projects/15_conf Sequence Seq	uence file		Ma	n			
Add file Remove file Modify file data type							
Contents of B-value field for supplied predicted protein chains : plddt	0						
Experiment							
High-resolution limit : Caracteria Asymmetric map	😑 Wrapping	Scatterin	ig table :	elect	ron 📀		
Composition (load a sequence file and/or paste and edit sequencesthe composition	n shown below is what will	be used)				
Unique sequences and conies (1)							

Will perform all the steps automatically.



New approach: Iterate prediction and model building





Improving prediction





Improving prediction



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



Improving prediction



• New prediction can be even better than the template

Terwilliger TC et al. Improved AlphaFold modeling with implicit experimental information. Nat Methods. 2022 Nov;19(11):1376–82.



Iterate with Predict and Build

•••	Predict and B	Build (Proje	ct: 7brm)		Preferences	?	Run Abort	Save	C Help	Phenix Server
Input/Output	Prediction and	Buildina Settir	PredictAndBuild	dCryoEM_17						< ▷
Prediction st	rategy									
Number of	models : 1		Include ter	nplates from PDB	🗸 Use MSA	Skip	all MSA aft	er first cyc	le	
Rebuilding st	hate					Numb	er of	:		
Cycles :	10	Number	of processors :	4 🗘 🕻	Refine on (no r		tion	/bui	Iding	
Pause	after docking a	nd rennemer	nt to check results				/		0	
✓ Update unique sequences										
Prediction S										
Prediction	Server : Phenix	xServer 🔇	Allow pre	calculated results	Allow precale	culated MSA	🗸 Stop	if internet	not available	1
Carry on pre	vious run (Restore	a completed ru	In from Job history to	fill this in automatically	/. You can also selec	t the CarryOn s	ubdirectory i	n a Predict_	_and_Build_xx	directory)
Carry-on d	lirectory :				Bro	owse Q	Co	ntinue this	s previous rur	n
Advanced in	puts									
Predict a	nd Build In	put Files	Box info	Processing	Search	Building				
Predictio	n (Control	All parameters							

Further reading/material

Documentation:

https://www.phenixonline.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:

😑 🕒 YouTube

>

- Overview: AlphaFold and Phenix
- <u>Processing a predicted model</u>
 - ssed predicted model in a cryo-EM map
 - ked predicted model in a cryo-EM map
 - cing and rebuilding a predicted model in a cryo-EM map pping parts of models



Phenix Tutorials

Video tutorials

AlphaFold changes everything: Incorporating predicted models in X-ray and Cryo-EM structure determination Presentation available at: January 18, 2022 phenix-online.org/presentation RéNaFoBiS Webinar How you can benefit from Tom Terwilliger AlbhaFold ... The New Mexico Consortium Los Alamos National Laboratory ... by determining structures more easily Randy Read, Tristan Croll, Claudia Millár (Cambridge, University), Paul Adams, Billy oon, Pavel Afonine, Christopher J. Schlicksup کری Phenix Lawrence Berkeley National Laboratory); Jane Richardson (Duke University) UNIVERSITY OF CAMBRIDGE Los Alamos BERKELEY New Mexico 🚥 🏟 🖬 🗖 🖸 Tom Terwilliger: AlphaFold changes everything

Search

Q I

Tom Terwilliger: AlphaFold changes everything



Amino acid sequence

1) Multiple sequence alignment (MSA)

- \rightarrow find AA more likely to mutate
- → find similar structures ("templates")
- ightarrow guess which AA are likely to be in contact with each other

Residues that **co-vary** are probably close in 3D structure

2) Neural networks:

Which parts of the MSA are more informative? Build a model

