Tom Terwilliger
The New Mexico Consortium
Los Alamos National Laboratory

January 18, 2022
RéNaFoBiS Webinar

How you can benefit from AlphaFold …

… by determining structures more easily

… and getting better models

Presentation available at: phenix-online.org/presentations

Randy Read, Tristan Croll, Claudia Millán, (Cambridge, University), Paul Adams, Billy Poon, Pavel Afonine, Christopher J. Schlicksup (Lawrence Berkeley National Laboratory); Jane Richardson (Duke University)
AlphaFold models …

- are great hypotheses for protein structures

- jump-start structure determination by X-ray and CryoEM

- can be iteratively improved with a density map

- allow a new work-flow for structure determination
AlphaFold models …

are great **hypotheses** for protein structures

jump-start structure determination by X-ray and CryoEM

can be iteratively improved with a density map

allow a new work-flow for structure determination
Machine-learning prediction of protein structure

Training

- Sequence
- Multiple sequence alignment
- 3D structure

\{ 180,000 protein structures in the PDB \}

Prediction

- Sequence
- Multiple sequence alignment

\{ 21 million parameters \}

Focus attention on important relationships

3D prediction

Confidence estimates (plDDT)

\{ 21 million parameters \}
Multiple sequence alignment is key information for AlphaFold.

Multiple sequence alignment

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

All sequences in alignment should be compatible with the right structure

Sequence coverage → Confidence

Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319
Models are accurate where sequence coverage is high

7mjs (3 Å, EMDB 23883)

Residues 1-100
High sequence coverage and confidence

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

Residues 100-120
Low sequence coverage, low confidence, low accuracy

AlphaFold

7mjs
Models are accurate where sequence coverage is high

7mjs (3 Å, EMDB 23883)

High confidence residues can be accurate

Residues 1-100

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

- Only protein
  - No water, ions, covalent modifications, carbohydrates, ligands, DNA, RNA

- Trained on good and poor structures
  - Parameters may systematically include poor geometry

- Little information about residues that are far apart
  - Models may have distortions and incorrect domain relationships
Two interpretations of PDB entry 6te3 … which is right?

AlphaFold

Uncommon rotamer
Poor fit to density
High-confidence \((p\text{DDT} = 0.97)\)

Common rotamer
Fits density

→ AlphaFold models are great hypotheses

ISOLDE rebuilding of 6te3 and slide by Tristan Croll
Local accuracy better than global accuracy

PDB entry 7bgl
(domain 3 of chain a)

AlphaFold
(multimer prediction)

Data from 7bgl, Johnson, S. et al. (2021).
Nat Microbiol 6, 712–721
Local accuracy better than global accuracy

PDB entry 7bgl
(domain 3 of chain a)

→ AlphaFold models are great hypotheses

AlphaFold (multimer prediction)
What can we expect from AlphaFold models?

They are great hypotheses for protein structures

Parts of AlphaFold models are accurate
Parts are completely wrong

The confidence measure is helpful but may not fully reflect accuracy
are great **hypotheses** for protein structures

jump-start structure determination by X-ray and CryoEM (and NMR, cryo-ET, neutron…)

can be iteratively improved with a density map

allow a new work-flow for structure determination
AlphaFold models are great for jump-starting structure determination

*Confidence measure* (pLDDT) allows pruning of worst parts of models

High-confidence parts are often accurate

Better than a homology model: no insertions and deletions in the sequence
Example: Finishing a difficult crystal structure

Repressor – DNA complex, solved with 2.6 Å SeMet SAD data and 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
Cryo-EM (7mjs 3 Å, EMDB 23883)
AlphaFold models …

- are great *hypotheses* for protein structures
- jump-start structure determination by X-ray and CryoEM
- can be iteratively improved with a density map
- allow a new work-flow for structure determination
Iterative AlphaFold prediction and rebuilding

Rebuilt model as a template ...

... can we get a better AlphaFold model?

Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319
The prediction is poor in the loop region

Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319
**AlphaFold prediction with a template**

- **Sequence**
- **Multiple sequence alignment**

```
EVQLVESGGGLVQPGGLSLCAGASGFNYSSSIWVRQAPGKGLEWYI
```

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

**21 million parameters**

**3D prediction**

**Template**

**Backbone angles**

Focus attention on residues that are close

→ Yes, the template improves prediction

→ The new prediction is even better than the template

Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319
Iterative AlphaFold prediction and rebuilding

Note this AlphaFold model is superimposed, not refined

Data from 7mjs, Cater, R.J., et al. (2021). Nature 595, 315–319
Iterative AlphaFold prediction and rebuilding

Rebuild AlphaFold model with density map

Rebuilt model improves next AlphaFold prediction

Iterate to improve model

Terwilliger et al. (2022). Improving AlphaFold modeling using implicit information from experimental density maps. BioRxiv 2022.01.07.475350
AlphaFold models …

are great *hypotheses* for protein structures

jump-start structure determination by X-ray and CryoEM

can be iteratively improved with a density map

allow a new work-flow for structure determination
Workflow for getting the best AlphaFold model

Sequence

Multiple sequence alignment

Templates from PDB

Templates you have created

AlphaFold Colab

AlphaFold prediction

Dock-and-rebuild

Rebuild with density map

ISOLDE

Fix errors

Rebuilt model
Workflow for cryo-EM

Density map

AlphaFold prediction for each chain

Dock-and-rebuild
Rebuild AlphaFold predictions

Model for each chain with full exact sequence

ISOLDE, real-space refine
Fix errors, refine

AlphaFold Colab
AlphaFold with refined model for each chain

Refined model for each chain
Workflow for crystallography

X-ray data

AlphaFold prediction for each chain

Phaser

Molecular replacement

MR model

AutoBuild Automated rebuilding

Autobuilt model

Density map

AlphaFold prediction

Dock-and-rebuild

Rebuild AlphaFold predictions

ISOLDE, Real-space refine

Fix errors, refine

Rebuilt model for each chain with full exact sequence

Refined model for each chain

AlphaFold Colab

AlphaFold with supplied template on each chain
AlphaFold models …

are great **hypotheses** for protein structures

jump-start structure determination by X-ray and CryoEM

can be iteratively improved with a density map

allow a new work-flow for structure determination
The Phenix Project

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

New Mexico Consortium
Los Alamos National Laboratory
Tom Terwilliger, Li-Wei Hung

University of Cambridge
Randy Read, Airlie McCoy, Tristan Croll, Claudia Millán Nebot, Rob Oeffner, Duncan Stockwell

Baylor College of Medicine
Matt Baker, Corey Hyrc

Duke University
Jane & David Richardson, Chris Williams, Vincent Chen

An NIH/NIGMS funded Program Project