Using AlphaFold predictions for structure determination

Phenix Workshop
March 7, 2024  Berkeley, CA

Tom Terwilliger
The New Mexico Consortium
Los Alamos National Laboratory
AlphaFold predictions are great hypotheses

AlphaFold models can be....
## AlphaFold predictions and confidence estimates

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

<table>
<thead>
<tr>
<th>AlphaFold confidence (pLDDT)</th>
<th>Median prediction error (Å)</th>
<th>Percentage with error over 2 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;90</td>
<td>0.6</td>
<td>10</td>
</tr>
<tr>
<td>80 - 90</td>
<td>1.1</td>
<td>22</td>
</tr>
<tr>
<td>70 - 80</td>
<td>1.5</td>
<td>33</td>
</tr>
<tr>
<td>&lt;70</td>
<td>3.5</td>
<td>77</td>
</tr>
</tbody>
</table>

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 prediction for RNA helicase
(PDB entry 6i5i)

<table>
<thead>
<tr>
<th>AlphaFold confidence (pLDDT)</th>
<th>Median prediction error (Å)</th>
<th>Percentage with error over 2 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;90</td>
<td>0.6</td>
<td>10</td>
</tr>
<tr>
<td>80 - 90</td>
<td>1.1</td>
<td>22</td>
</tr>
<tr>
<td>70 - 80</td>
<td>1.5</td>
<td>33</td>
</tr>
<tr>
<td>&lt;70</td>
<td>3.5</td>
<td>77</td>
</tr>
</tbody>
</table>

**PAE matrix** (Predicted aligned error)

AlphaFold prediction for RNA helicase
(PDB entry 6i5i)

Confidence:
- **Blue**: > 90
- **Green**: 80 - 90

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 (Cryo-EM)

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

---

**Strategy for structure determination in the AlphaFold era**
**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)*

- **Working structure and map**
  - Molecular replacement

- **Rebuilt model and optimized map**
  - Density modification and autobuilding
  - 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)

**ResolveCryoEM, LocalAnisoSharpen** (map improvement)

**EMPlacement, DockInMap** (Docking of single, multiple chains)

**DockAndRebuild** (Morphing and rebuilding)

**RealSpaceRefine** (Refinement)

**Phaser-MR** (Molecular replacement)

**AutoBuild** (Density modification and rebuilding)

**Phenix.refine** (Refinement)

**PredictAndBuild** (Prediction and structure determination)
An NIH/NIGMS funded Program Project