Highly Accurate Protein Structure Prediction with AlphaFold

Developed by Google DeepMind and EMBL-EBI

歐任翔 Mark Ou, PhD

Sep 25, 2024

Slide available here.

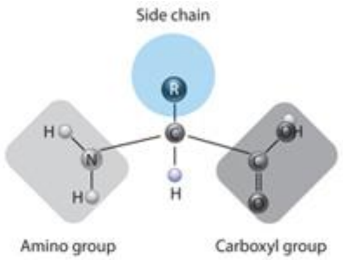
Something About DeepMind

Central mission: solve fundamental problems with AI

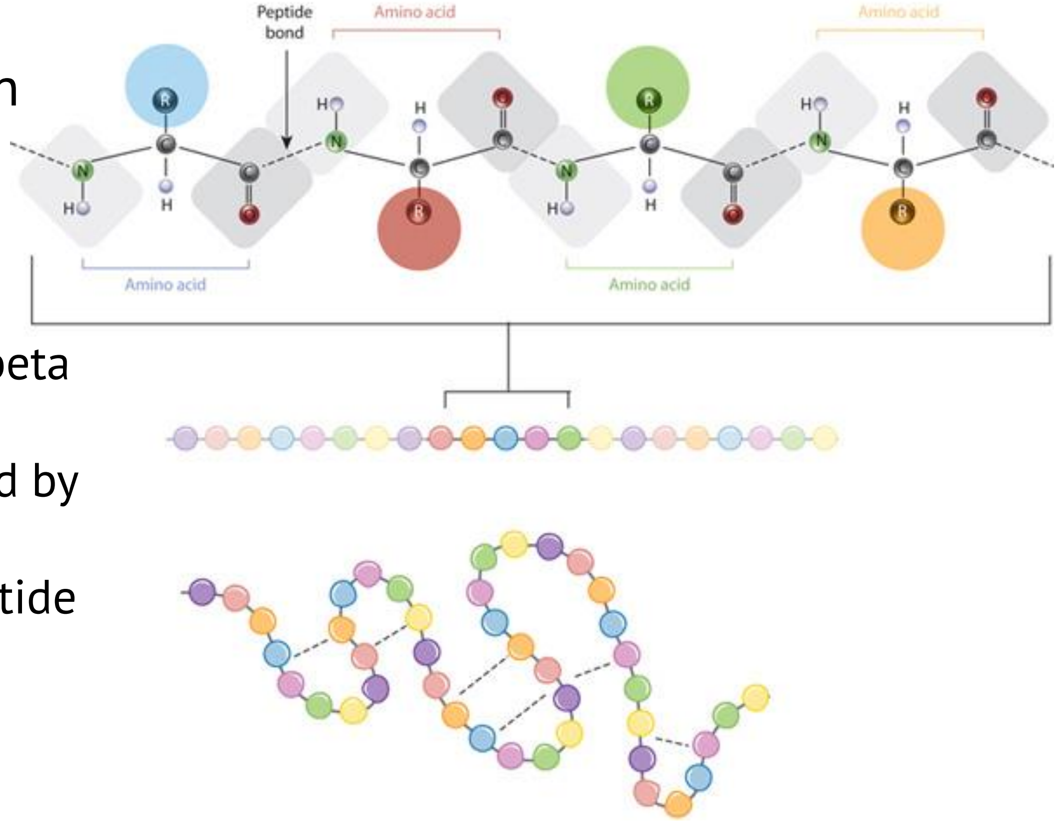
=> Predicting the 3D structure of a protein form its amino acid sequence is one such challenge

https://www.nature.com/articles/s41586-021-03819-2

What are proteins?



- Molecular machines which are essential to life
- Have many functions, from hair to the immune system
- Consist of *chains of amino acids* that fold into a 3D structure
  => The 3D shape is important for a protein’s function

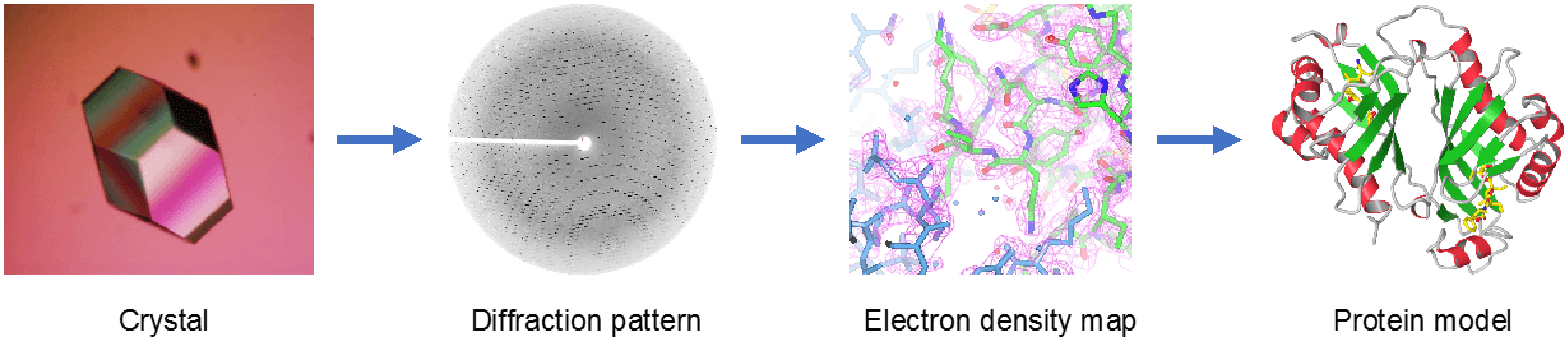


- Protein structures
  - **Primary structure**: Linear sequence of amino acids.
  - **Secondary structure**: Patterns like alpha helices and beta sheets formed by hydrogen bonds.
  - **Tertiary structure**: The 3D shape of the protein formed by side-chain interactions.
  - **Quaternary structure**: Complexes of multiple polypeptide chains or subunits.

Why and how to predict protein structure?

- Experimental structure determination takes months to years
- Structure prediction can provide actionable information faster

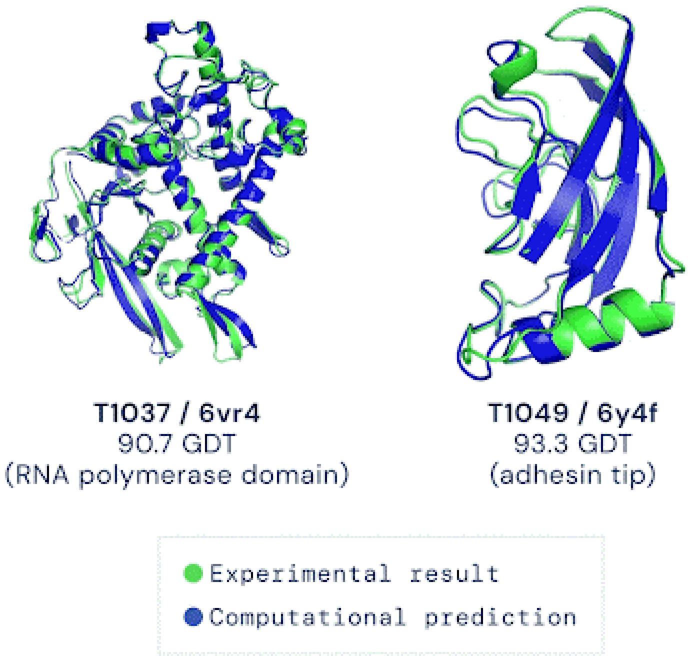
X-ray Crystallography:

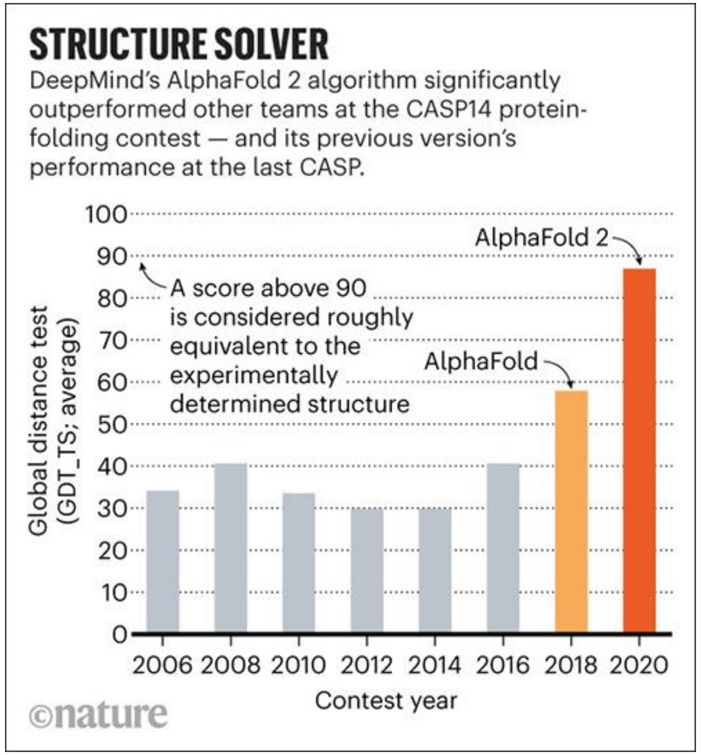


Other methods: NMR spectroscopy, Cryo-EM, EM

How well is the prediction by AlphaFold

- Protein structure prediction community established CASP (Critical Assessment of Protein Structure Prediction)
- CASP assessment involves the prediction of recently solved structures that are not public
- From CASP 14 (2020), AlphaFold is the top-ranked method achieving consistently high accuracy




STRUCTURE SOLVER

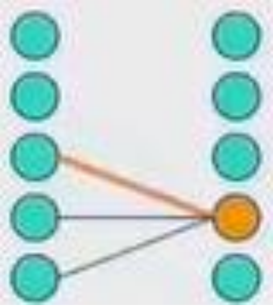
DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.

©nature

How AlphaFold works: INDUCTIVE BIAS FOR DEEP LEARNING MODELS

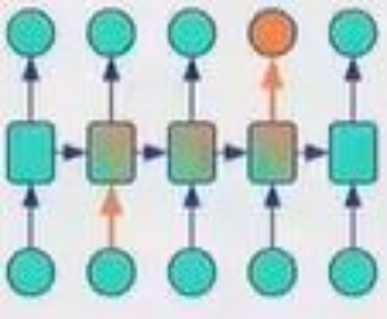
Examples:

卷積神經網路 | Convolutional Networks (CNN)



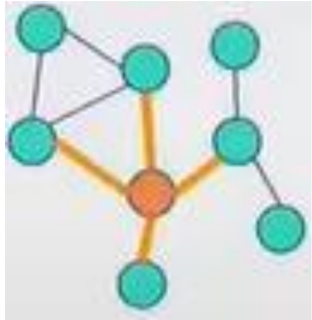
- Used to process structured data with locality (such as sequence data)
- Extract features from local regions in the protein sequence (interaction between amino acids)
- Local sequence segments creates secondary structures

循環神經網路 | Recurrent Networks (RNN)



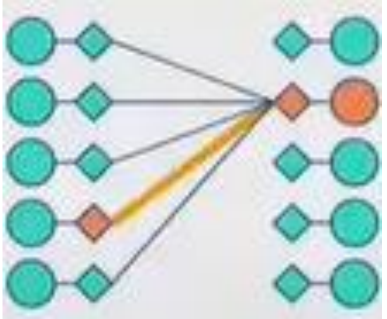
- Used to process sequence data, particular for capturing long-range dependencies.
- Capture interactions between distant amino acid → overall stability and folding
- The 3D structures of a protein is not only determined by adjacent amino acid but also by long-range interactions (hydrogen bonds or hydrophobic interactions).

圖神經網路 | Graph Networks (GNN)



- Used to process graph-structure data.
- Considering amino acids as nodes and edges as their interactions, GNNs help learn the complex relationships between amino acids and convert this into structural information
- Consider global information about the protein structure, not just local fragments.

注意力機制 | Attention Modules

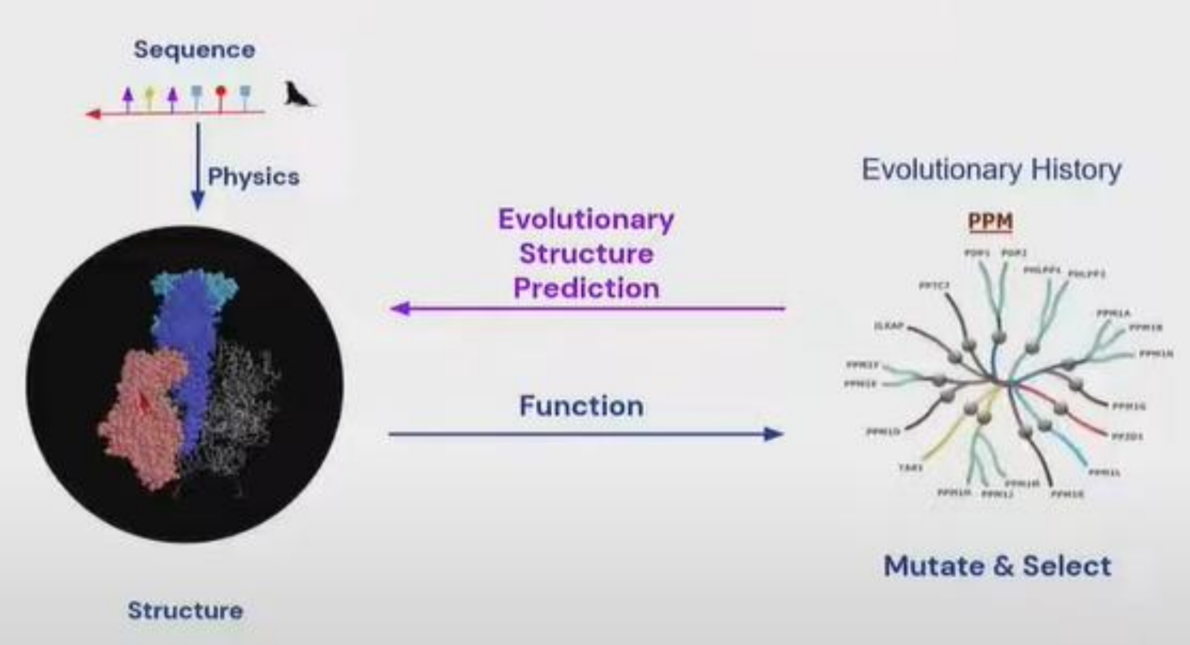


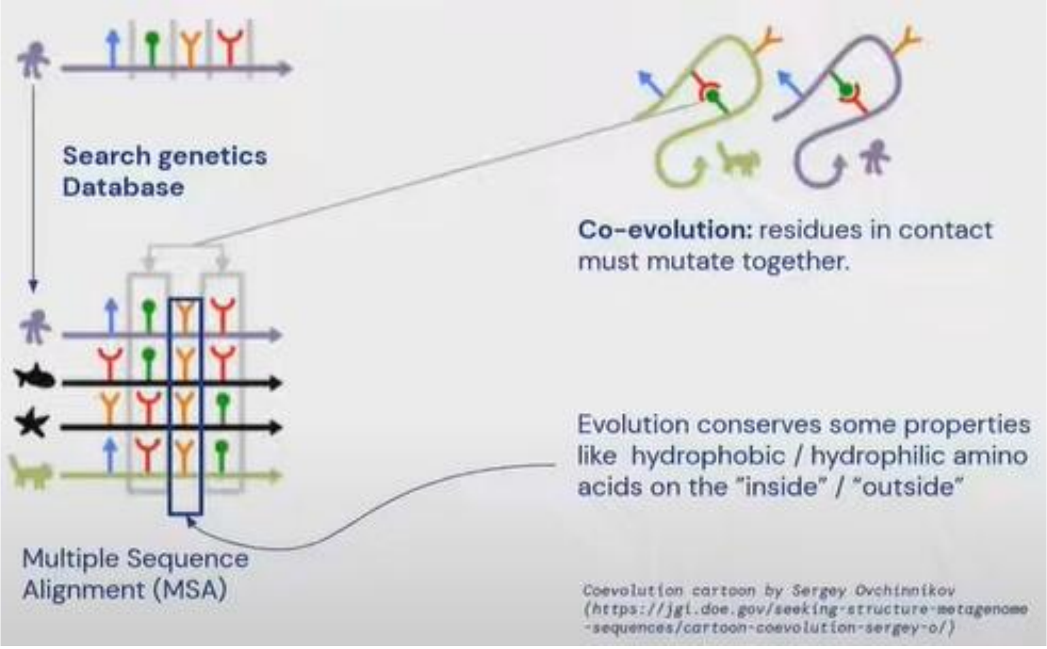
- It allow the model to focus on different part of a sequence based on the importance of each amino acid.
- Certain amino acids are more crucial than others, especially in forming structural cores or functional regions.
- Enable model to flexibly focus more attention on these important amino acids.

Putting protein knowledge into the model

- Physical and geometric insights are built into the network structure, not just a process around it.
- Inductive biases reflect our knowledge of protein physics and geometry
    - De-emphasized sequential order of amino acids
      (any amino acid can talk to any amino acid in that protein)
    - Instead, residues that are close in space need to communicate
    - Iteratively learning a graph of which residuals are close while reasoning over this implicit graph as it is being built

Determining structure from evolution




Search genetics Database
Co-evolution: residues in contact must mutate together.
Evolution conserves some properties like hydrophobic / hydrophilic amino acids on the "inside" / "outside"
Multiple Sequence Alignment (MSA)

Coevolution cartoon by Sergey Ovchinnikov (https://jgi.doe.gov/seeking-structure-metagenome-sequences/cartoon-coevolution-sergey-o/)

The structure needs to be similar to carry out the same function.
→ Given an evolutionally related sequence, we can try to computationally infer the structure.

Considering other important factors

- Structure module
  - End-to-end folding instead of gradient descent
- Noisy student distillation (bootstrapping oneself for a better performance)
  - Make use of unlabeled sequences (didn’t have a known experimental structure)
  - Train AlphaFold on just PDB data → Predict structure on a large set of unlabeled sequences → Train second model where training set is enriched by confidently predicted structure of first model

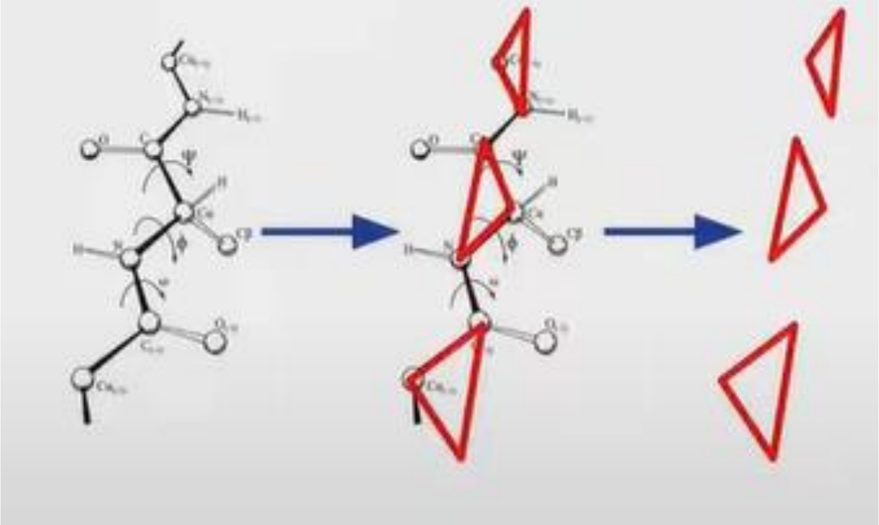
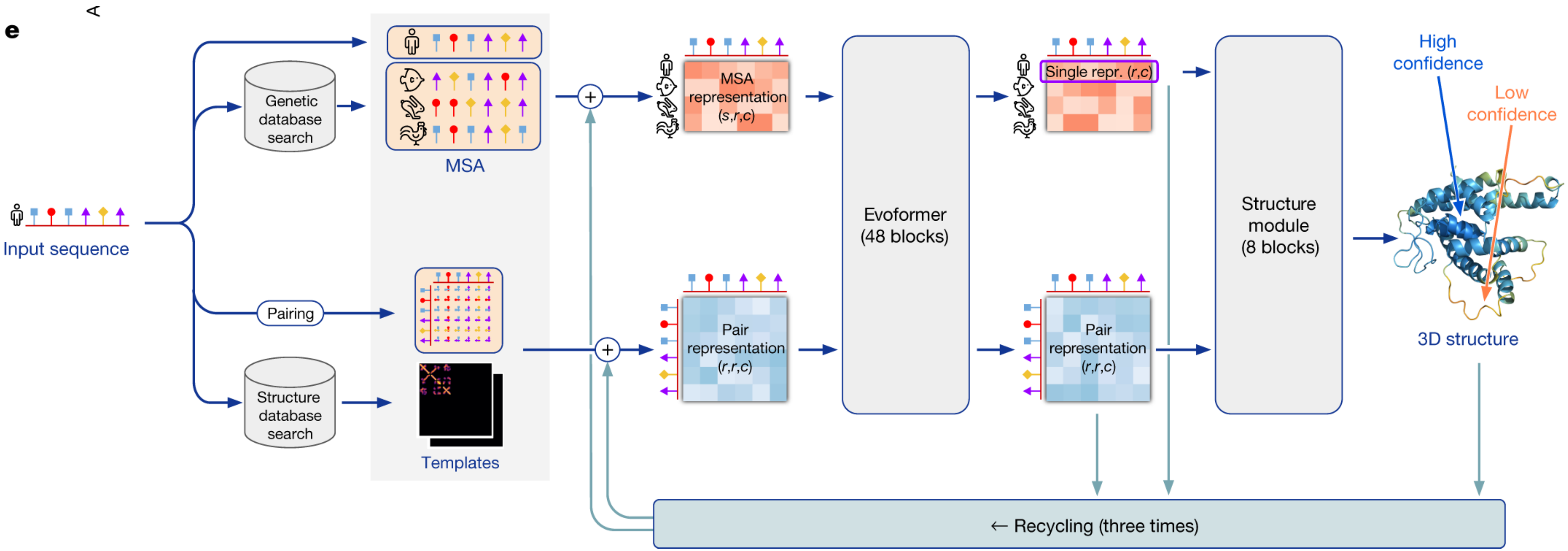


Image: Dcrjsr, vectorised Adam Rędzikowski (CC BY 3.0, Wikipedia)

Overview

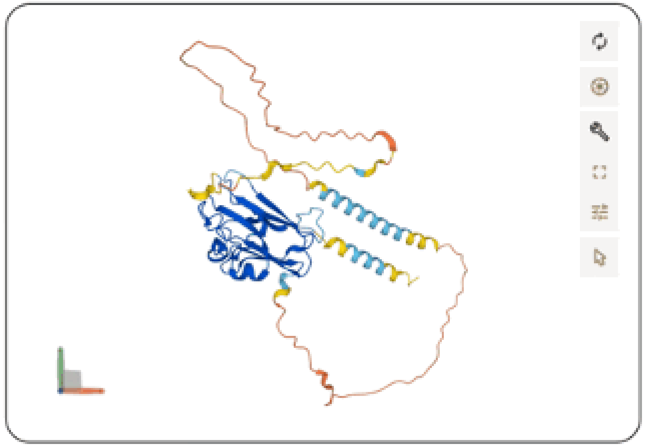

e
A
Input sequence
Genetic database search
MSA
Pairing
Structure database search
Templates
MSA representation (s,r,c)
Pair representation (r,r,c)
Evoformer (48 blocks)
Single repr. (r,c)
Pair representation (r,r,c)
Structure module (8 blocks)
High confidence
Low confidence
3D structure
← Recycling (three times)

How to interpret predictions?

Predicted LDDT (local distance difference test)

- A metric used to assess the accuracy of predicted protein structures by comparing them to the true (experimental) structures
- Specifically, it measures how well the predicted atomic distance, both in backbone and side chains, matches the actual distance in the native structure
- **Ranging 0~100. The higher, the better.**
- Compared with the old criteria RMSD (root mean square deviation), which focuses on the overall alignment, LDDT emphasized the accuracy of local regions within the protein structure.

Ephrin-B2



Model Confidence

🟦 Very high (pLDDT > 90)

🟦 High (90 > pLDDT > 70)

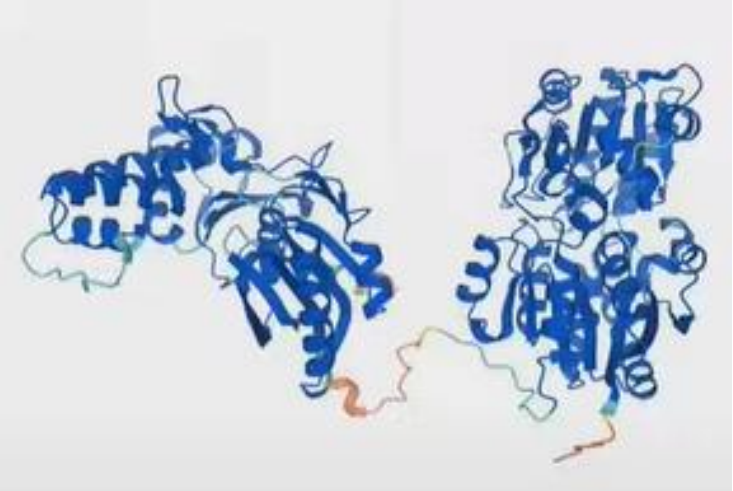
🟨 Low (70 > pLDDT > 50)

🟧 Very low (pLDDT < 50)

- pLDDT < 50: don’t trust the structure or not taking any structure
- pLDDT > 70: May work with the backbone predictions but may not want to trust the side chains in the area
- pLDDT > 90: Reasonable to investigate side chains / active site details

Pitfalls of pLDDT

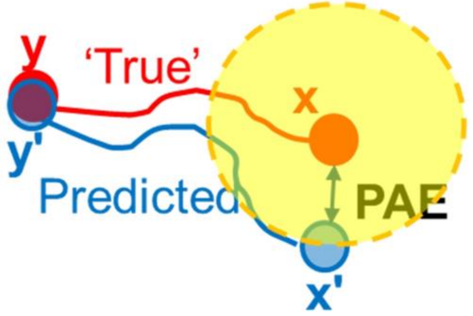
High pLDDT on all domains does NOT imply AlphaFold is confident of their relative positions



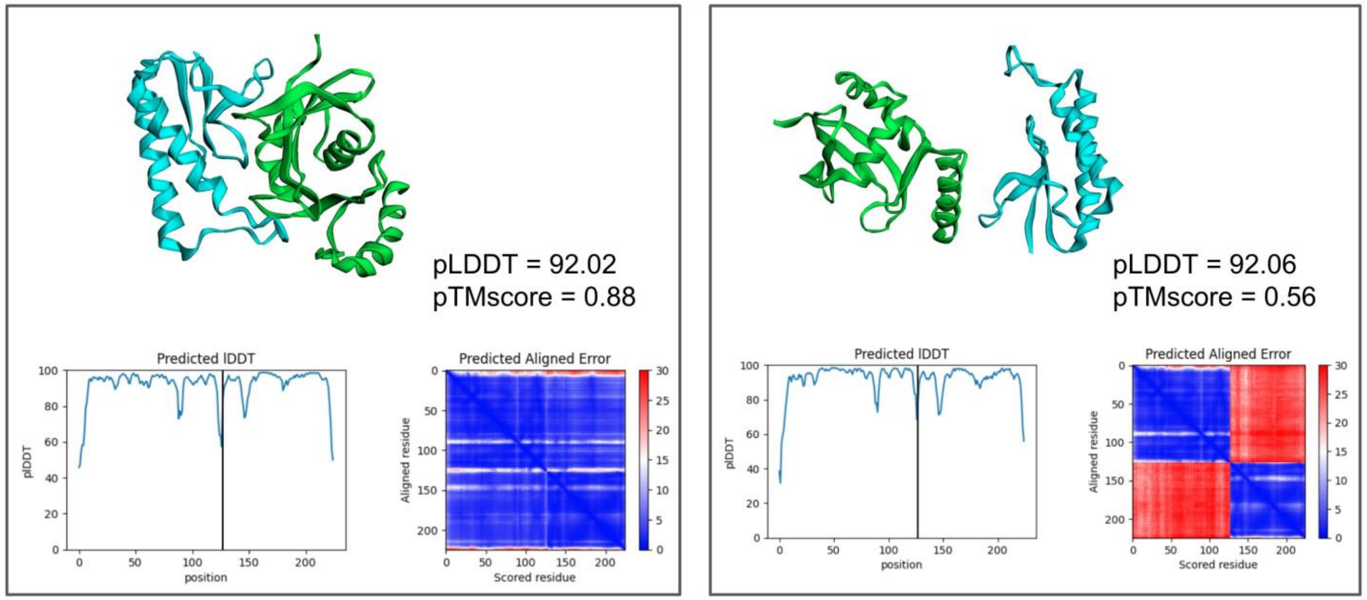
Assessing inter-domain confidence requires a different metric

Predicted aligned error (PAE)

- PAE measures how well AlphaFold predicts the distance between two residues (amino acids) in the protein, along with the uncertainty in the alignment of those residues
- PAE score:
  - values often range from 0 to 30 Å (Ångströms; 10^{-10} m) or more
  - Lower PAE score: high confidence in the relative position between those residues
  - Higher PAE score: greater uncertainty in their relative positioning



casp:H1065 - pdb:7M5F



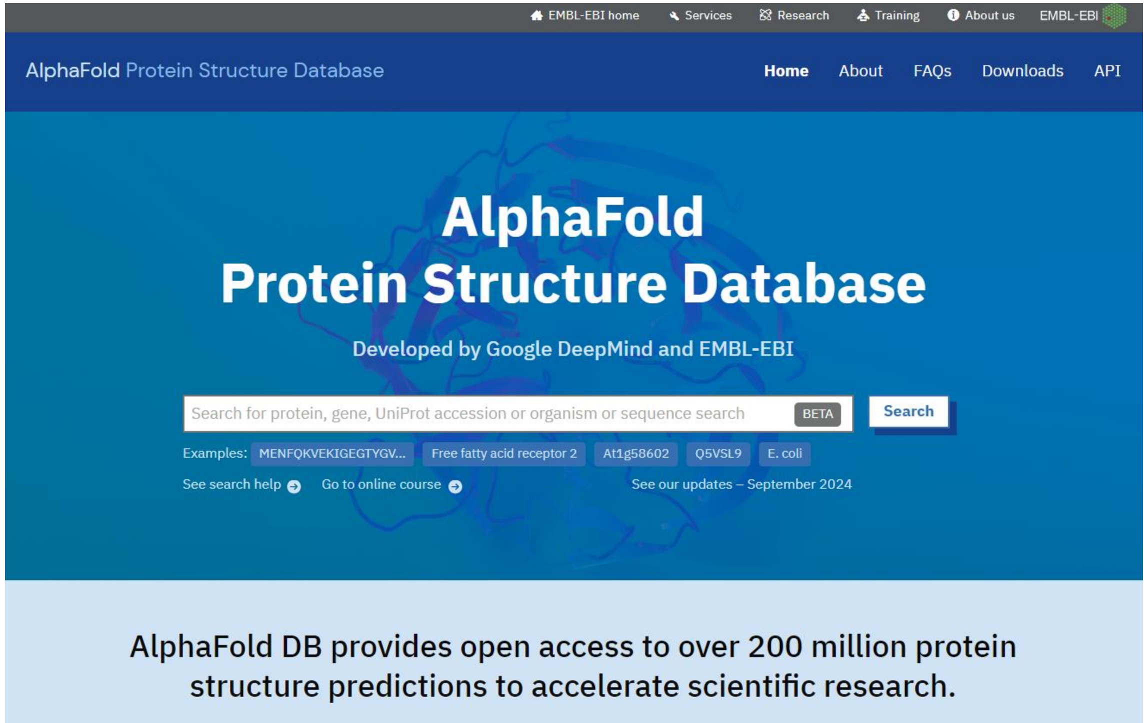
Don't trust how AlphaFold positions these two domains!

Limitations of AlphaFold

- Only accepts the 20 standard amino acids in its input
- (By default) predicts 5 models per run:
  - However, models are generally very similar
  - Usually cannot predict conformational variability in a protein
- AlphaFold was not trained for
  - Predict assemblies (AlphaFold-Multimer was trying to do this)
  - Predict the effects of mutations (What AlphaMissense is trying to do)
  - Predict the binding of ligand molecules (some recent research is trying to achieve this using AlphaFold as the basic)
  - Predict nucleic acid structures

AlphaFold protein structure database

- Website developed and hosted by EMBL-EBI


EMBL-EBI home
Services
Research
Training
About us
EMBL-EBI
AlphaFold Protein Structure Database
Home
About
FAQs
Downloads
API
Search for protein, gene, UniProt accession or organism or sequence search
BETA
Search
Examples: MENFQKVEKIGEGTYGV... Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli
See search help
Go to online course
See our updates – September 2024

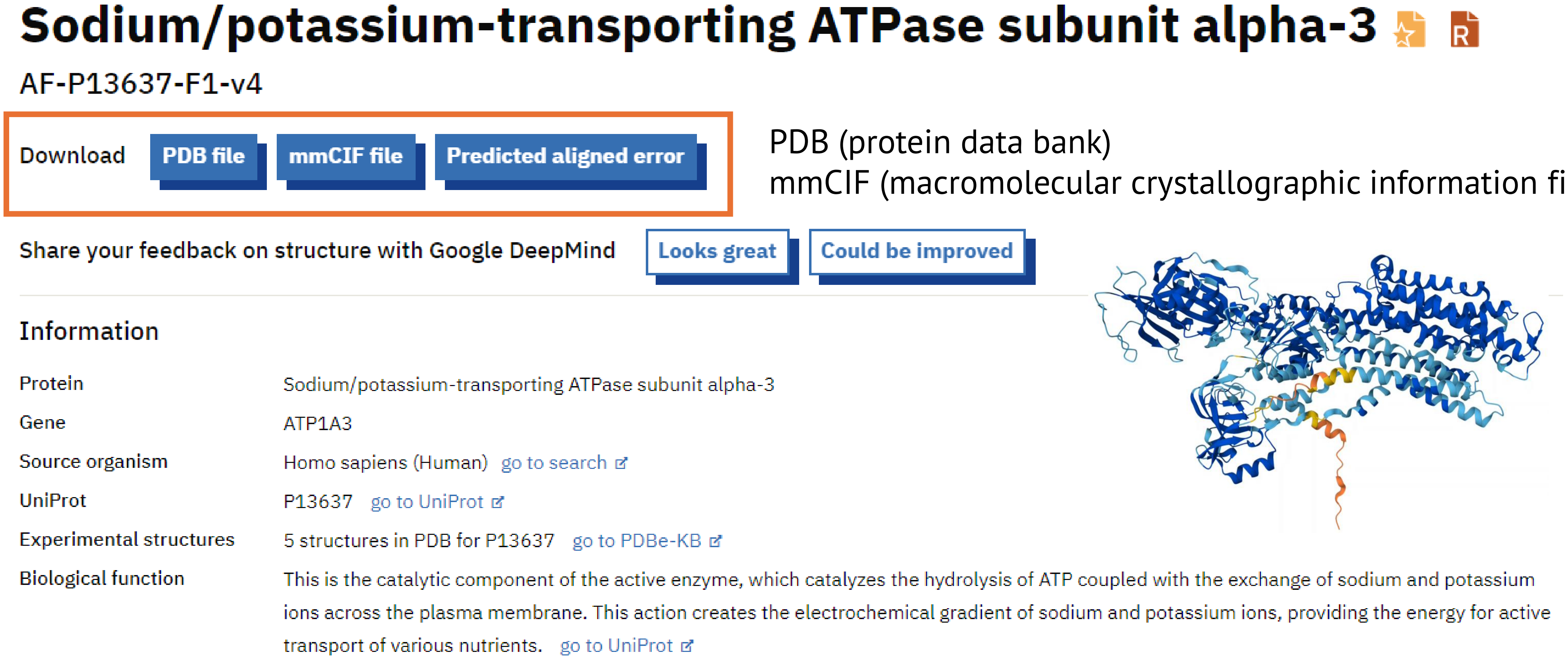
AlphaFold
Protein Structure Database

Developed by Google DeepMind and EMBL-EBI

AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.

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

An example



Sodium/potassium-transporting ATPase subunit alpha-3

AF-P13637-F1-v4

Download PDB file mmCIF file Predicted aligned error

PDB (protein data bank)
mmCIF (macromolecular crystallographic information file)

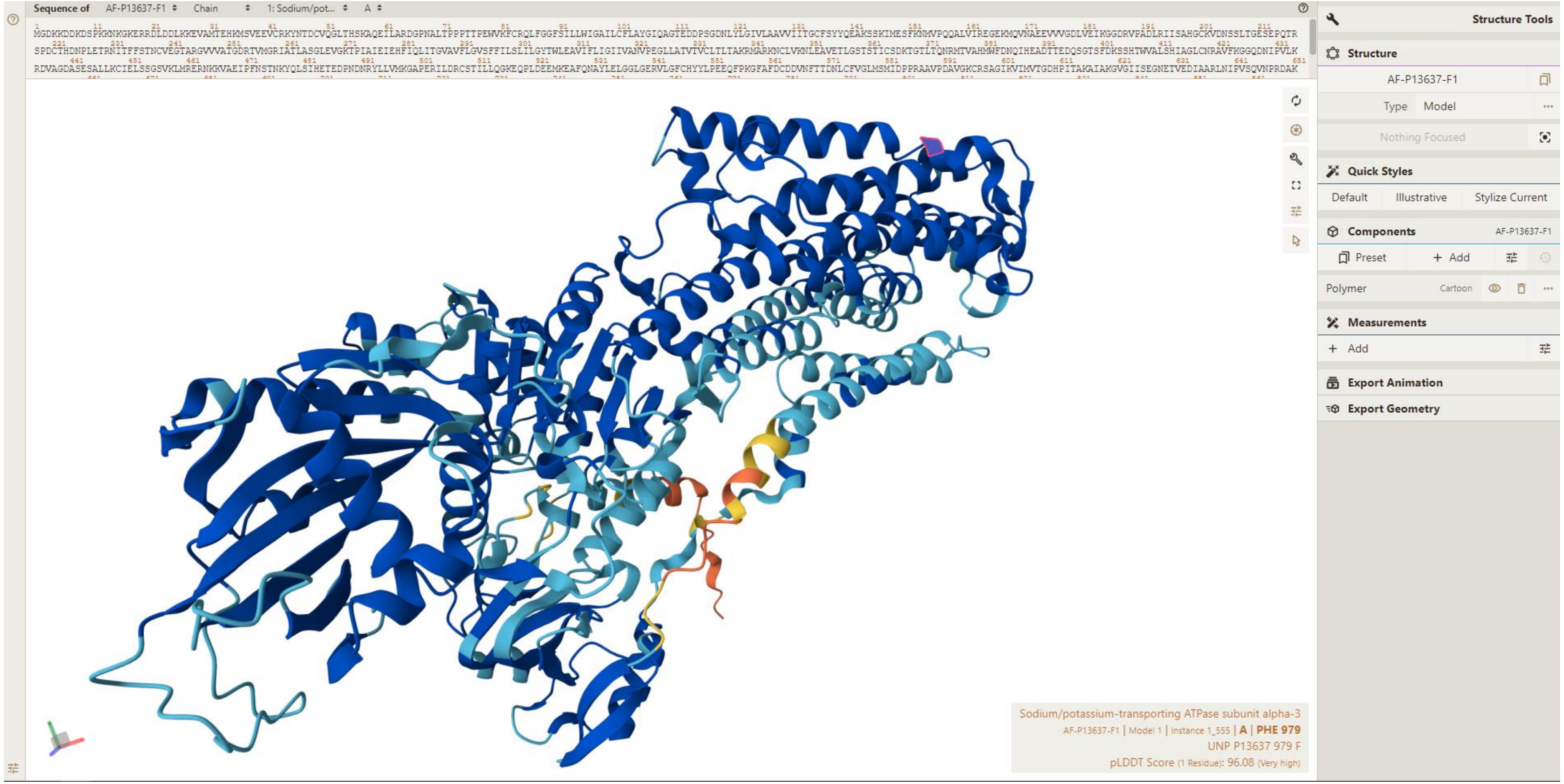
Share your feedback on structure with Google DeepMind Looks great Could be improved

Information

Protein	Sodium/potassium-transporting ATPase subunit alpha-3
Gene	ATP1A3
Source organism	Homo sapiens (Human) go to search
UniProt	P13637 go to UniProt
Experimental structures	5 structures in PDB for P13637 go to PDBe-KB
Biological function	This is the catalytic component of the active enzyme, which catalyzes the hydrolysis of ATP coupled with the exchange of sodium and potassium ions across the plasma membrane. This action creates the electrochemical gradient of sodium and potassium ions, providing the energy for active transport of various nutrients. go to UniProt

https://alphafold.ebi.ac.uk/entry/P13637

Structure viewer



Sequence of AF-P13637-F1 Chain 1: Sodium/pot... A

```
MGDKKDDKDSPKKNKGKERRDLDDLKKEVAMTEHKMSVEEVCRKYNTDCVQGLTHSKAQEILARDGPNALTPPPTTPEWVKFCRQLFGGFSILLWIGAILCFLAYGIQAGTEDDPSGDNLYLGIVLAAVVIITGCFSYYQEAKSSKIMESFKNMVPQQALVIREGEKMQVNAEEVVVGDLVEIKGGDRVPADLRIISAHGCKVDNSSLTGESEPQTR
SPDCTHDNPLETRNITFFSTNCVEGTARGVVVATGDRTVMGRIATLASGLEVGKTPIAIEIEHFIQLITGVAVFLGVSFFILSLILGYTWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKRMARKNCLVKNLEAVETLGSTSTICSDKTGTLTQNRMTVAHMWFDNQIHEADTTEDQSGTSFDKSSHTWVALSHIAGLCNRAVFKGGQDNIPVLK
RDVAGDASESALLKCIELSSGSVKLMRERNKKVAEIPFNSTNKYQLSIHETEDPNDNRYLLVMKGAPERILDRCSTILLQGKEQPLDEEMKEAFQNAYLELGGLGERVLGFCHYYLPEEQFPKGFAFDCDDVNFTTDNLCFVGLMSMIDPPRAAVPDAVGKCRSAGIKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPVSQVNPRDAK
```

Sodium/potassium-transporting ATPase subunit alpha-3
AF-P13637-F1 | Model 1 | Instance 1_555 | **A | PHE 979**
UNP P13637 979 F
pLDDT Score (1 Residue): 96.08 (Very high)

Structure Tools

Structure

AF-P13637-F1

Type: Model

Nothing Focused

Quick Styles

Default | Illustrative | Stylize Current

Components AF-P13637-F1

Preset | + Add

Polymer: Cartoon

Measurements

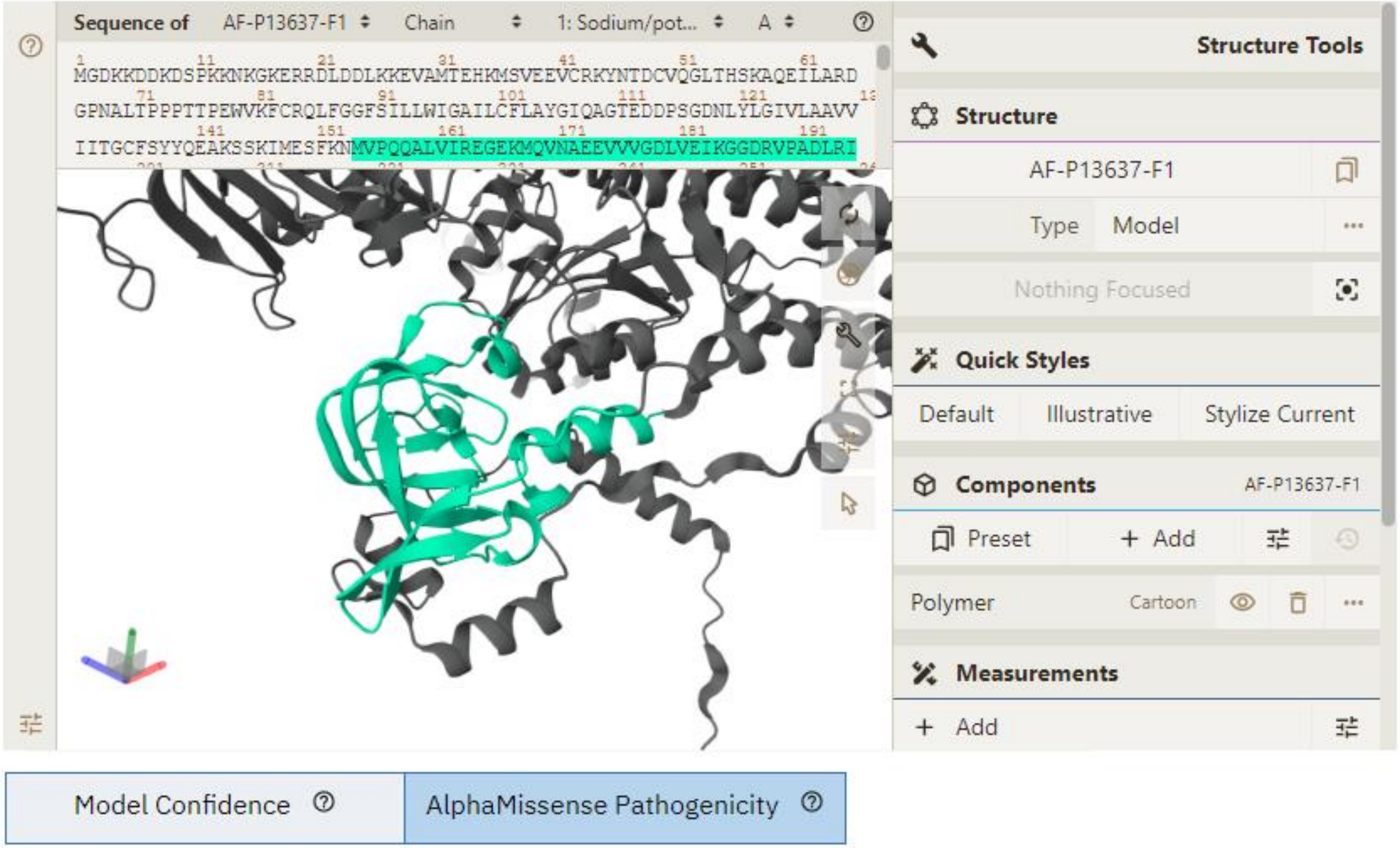
+ Add

Export Animation

Export Geometry

Predicted aligned error (PAE)

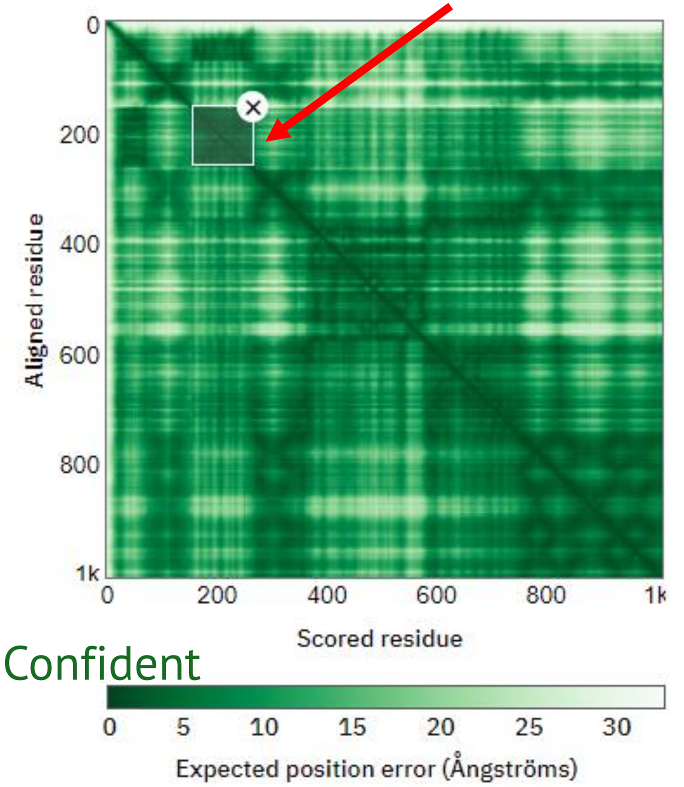
Assess relative domain positions.



Sequence of AF-P13637-F1 Chain 1: Sodium/pot... A

```
1         11        21        31        41        51        61
MGDKKDDKDSPKKNKGKERRDLDDLKKEVAMTEHKMSVEEVCRKYNTDCVQGLTHSKAQEILARD
    71        81        91        101       111       121
GPNALTPPPTTPEWVKFCRQLFGGFSILLWIGAILCFLAYGIQAGTEDDPSGDNLYLGIVLAAVV
       141       151       161       171       181       191
IITGCFSYYQEAKSSKIMESFKNMVPQQALVIREGEKMQVNAEEVVVGDLVEIKGGDRVPADLRI
```

The heatmap on the website is interactive!



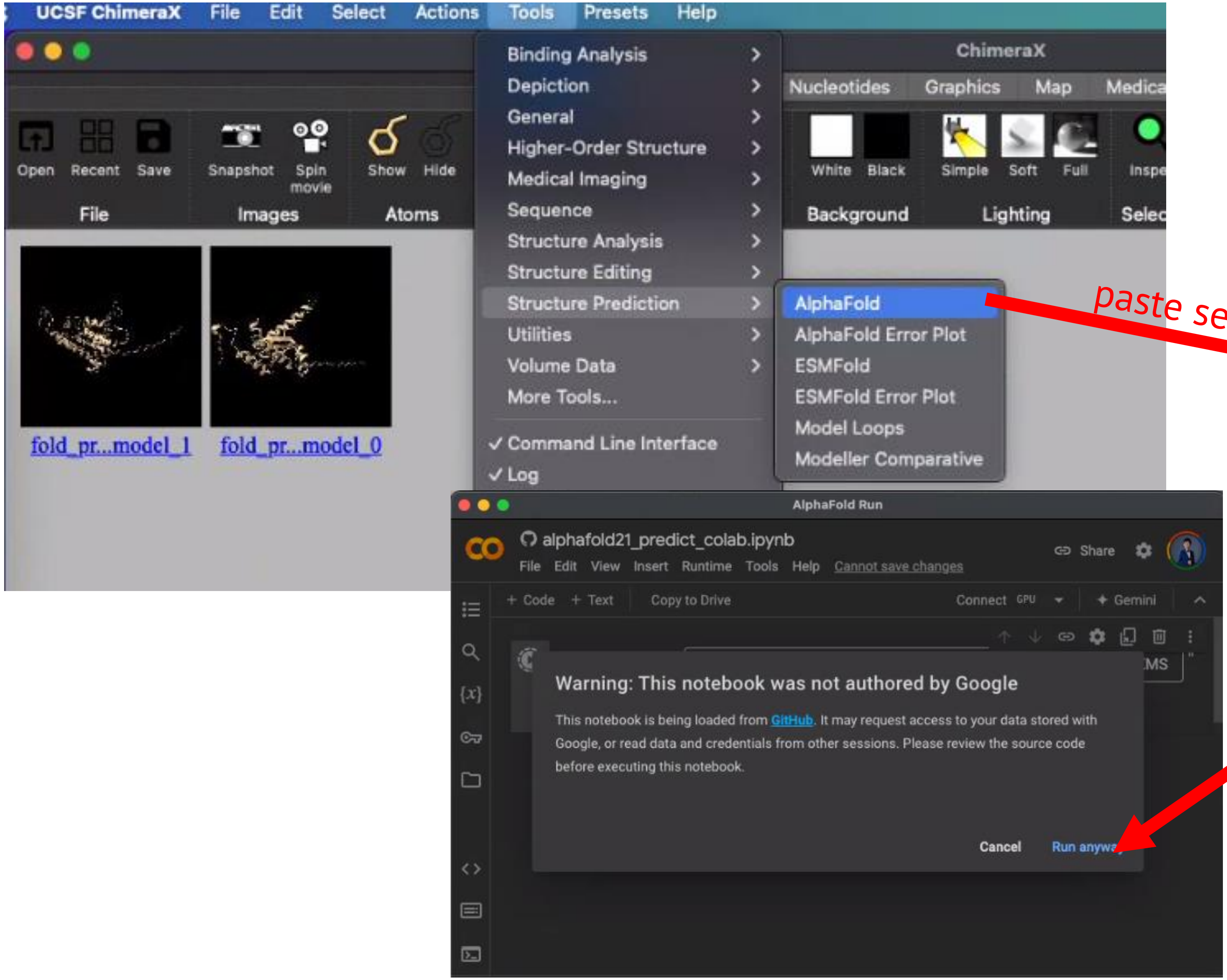
Predicted aligned error (PAE)

Predict own data

Main ways of accessing predicted protein structures from AlphaFold:

- The open-source code is publicly accessible at https://github.com/google-deepmind/alphafold
  - Total control over predictions
  - Need large storage space and a modern GPU
- Interactive Google Colab notebooks: https://bit.ly/alphafoldcolab
  - More limited in terms of configuration
  - Easier to use, harder to break

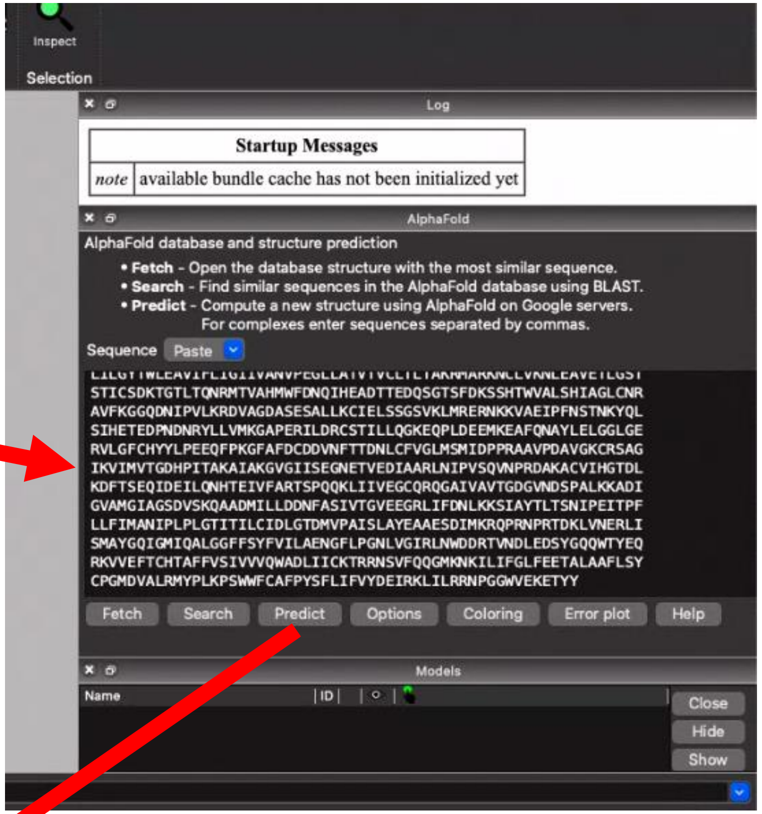
UCSF ChimeraX + Colab (google) + AlphaFold



Warning: This notebook was not authored by Google

This notebook is being loaded from GitHub. It may request access to your data stored with Google, or read data and credentials from other sessions. Please review the source code before executing this notebook.

paste sequence



Startup Messages

note available bundle cache has not been initialized yet

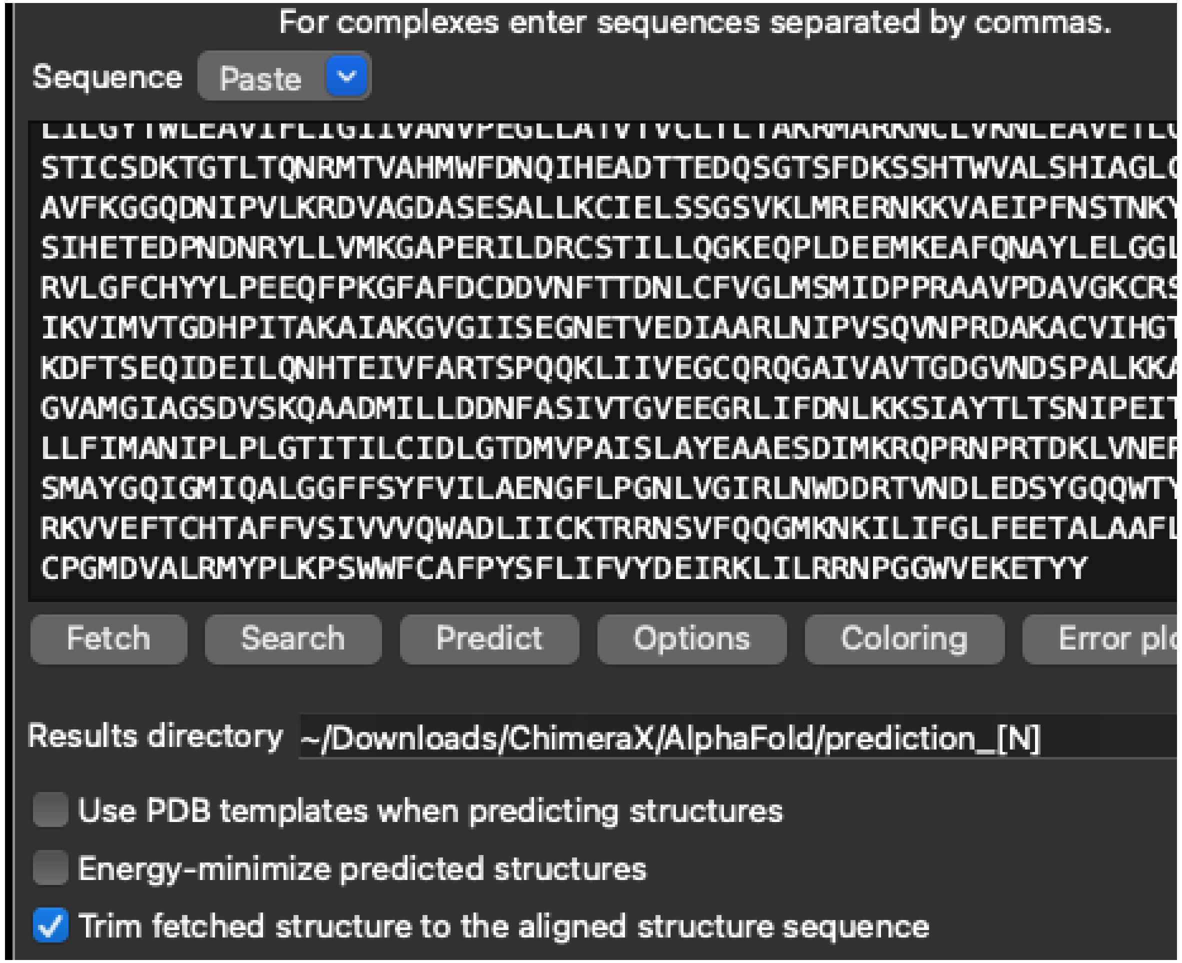
AlphaFold database and structure prediction

- **Fetch** - Open the database structure with the most similar sequence.
- **Search** - Find similar sequences in the AlphaFold database using BLAST.
- **Predict** - Compute a new structure using AlphaFold on Google servers. For complexes enter sequences separated by commas.

```
LILGYTWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKRMARKNCLVKNLEAVETLGST
STICSDKTGTLTQNRMTVAHMWFDNQIHEADTTEDQSGTSFDKSSHTWVALSHIAGLCNR
AVFKGGQDNIPVLKRDVAGDASESALLKCIELSSGSVKLMRERNKKVAEIPFNSTNKYQL
SIHETEDPNDNRYLLVMKGAPERILDRCSTILLQGKEQPLDEEMKEAFQNAYLELGGLGE
RVLGFCHYYLPEEQFPKGFAFDCDDVNFTTDNLCFVGLMSMIDPPRAAVPDAVGKCRSAG
IKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPVSQVNPRDAKACVIHGTDL
KDFTSEQIDEILQNHTEIVFARTSPQQKLIIVEGCQRQGAIVAVTGDGVNDSPALKKADI
GVAMGIAGSDVSKQAADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEITPF
LLFIMANIPLPLGTITILCIDLGTDMVPAISLAYEAAESDIMKRQPRNPRTDKLVNERLI
SMAYGQIGMIQALGGFFSYFVILAENGFLPGNLVGIRLNWDDRTVNDLEDSYGQQWTYEQ
RKVVEFTCHTAFFVSIVVVQWADLIICKTRRNSVFQQGMKNKILIFGLFEETALAAFLSY
CPGMDVALRMYPLKPSWWFCAFPYSFLIFVYDEIRKLILRRNPGGWVEKETYY
```

- Sequence data from UniProt: https://www.uniprot.org/uniprotkb/P13637/entry#sequences
- Use commas to separate if more than one sequences

Options



For complexes enter sequences separated by commas.

Sequence: Paste

```
LILGYTWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKRMARKNCLVKNLEAVETL
STICSDKTGTLTQNRMTVAHMWFDNQIHEADTTEDQSGTSFDKSSHTWVALSHIAGL
AVFKGGQDNIPVLKRDVAGDASESALLKCIELSSGSVKLMRERNKKVAEIPFNSTNK
SIHETEDPNDNRYLLVMKGAPERILDRCSTILLQGKEQPLDEEMKEAFQNAYLELGG
RVLGFCHYYLPEEQFPKGFAFDCDDVNFTTDNLCFVGLMSMIDPPRAAVPDAVGKCR
IKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPVSQVNPRDAKACVIHG
KDFTSEQIDEILQNHTEIVFARTSPQQKLIIVEGCQRQGAIVAVTGDGVNDSPALKK
GVAMGIAGSDVSKQAADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEI
LLFIMANIPLPLGTITILCIDLGTDMVPAISLAYEAAESDIMKRQPRNPRTDKLVNE
SMAYGQIGMIQALGGFFSYFVILAENGFLPGNLVGIRLNWDDRTVNDLEDSYGQQWT
RKVVEFTCHTAFFVSIVVVQWADLIICKTRRNSVFQQGMKNKILIFGLFEETALAAF
CPGMDVALRMYPLKPSWWFCAFPYSFLIFVYDEIRKLILRRNPGGWVEKETYY
```

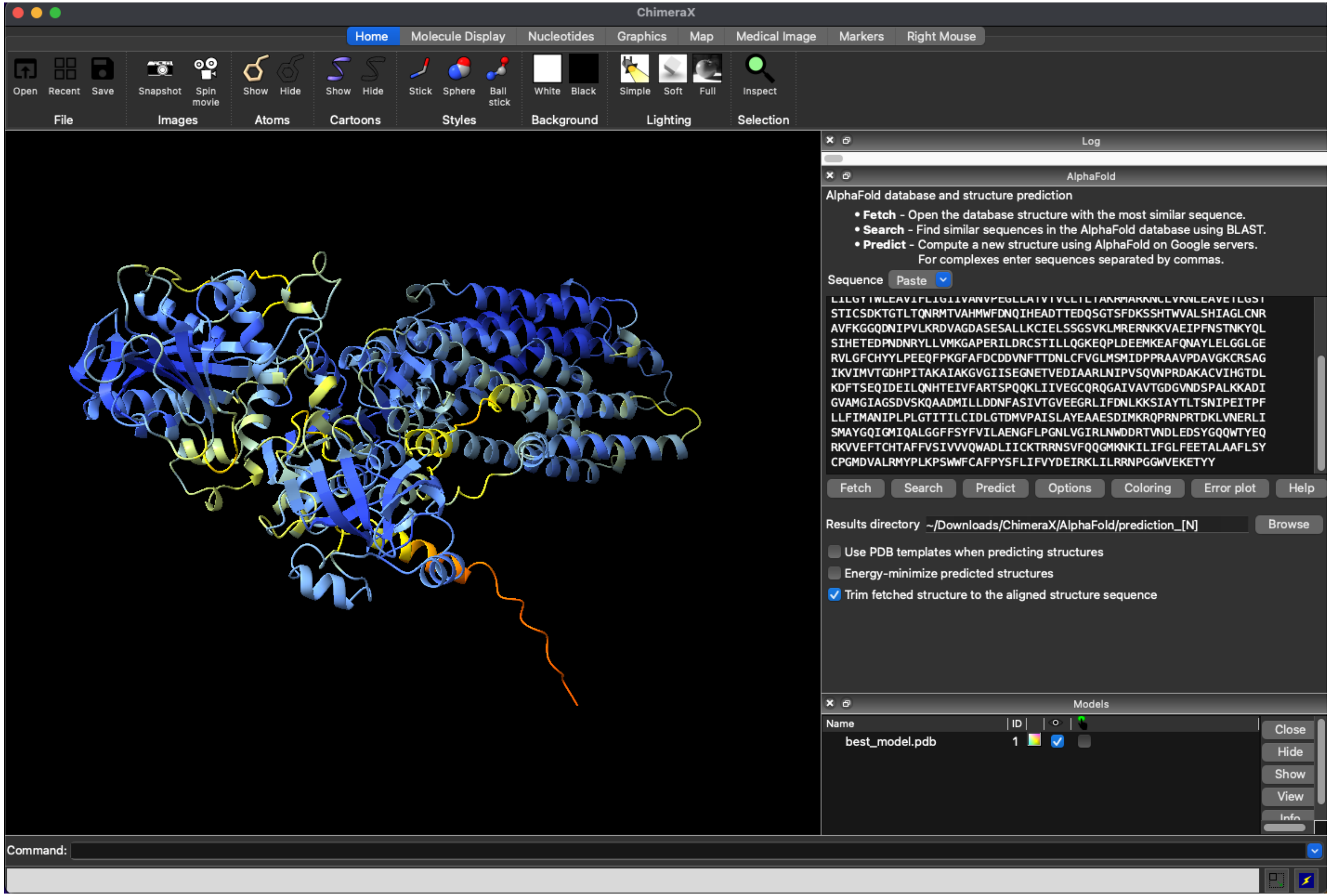
Results directory: ~/Downloads/ChimeraX/AlphaFold/prediction_[N]

□ Use PDB templates when predicting structures

□ Energy-minimize predicted structures

☑ Trim fetched structure to the aligned structure sequence

Result



AlphaFold database and structure prediction

- **Fetch** - Open the database structure with the most similar sequence.
- **Search** - Find similar sequences in the AlphaFold database using BLAST.
- **Predict** - Compute a new structure using AlphaFold on Google servers. For complexes enter sequences separated by commas.

Sequence Paste

```
LILGYTWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKRMARKNCLVKNLEAVETLGST
STICSDKTGTLTQNRMTVAHMWFDNQIHEADTTEDQSGTSFDKSSHTWVALSHIAGLCNR
AVFKGGQDNIPVLKRDVAGDASESALLKCIELSSGSVKLMRERNKKVAEIPFNSTNKYQL
SIHETEDPNDNRYLLVMKGAPERILDRCSTILLQGKEQPLDEEMKEAFQNAYLELGGLGE
RVLGFCHYYLPEEQFPKGFAFDCDDVNFTTDNLCFVGLMSMIDPPRAAVPDAVGKCRSAG
IKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPVSQVNPRDAKACVIHGTDL
KDFTSEQIDEILQNHTEIVFARTSPQQKLIIVEGCQRQGAIVAVTGDGVNDSPALKKADI
GVAMGIAGSDVSKQAADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEITPF
LLFIMANIPLPLGTITILCIDLGTDMVPAISLAYEAAESDIMKRQPRNPRTDKLVNERLI
SMAYGQIGMIQALGGFFSYFVILAENGFLPGNLVGIRLNWDDRTVNDLEDSYGQQWTYEQ
RKVVEFTCHTAFFVSIVVVQWADLIICKTRRNSVFQQGMKNKILIFGLFEETALAAFLSY
CPGMDVALRMYPLKPSWWFCAFPYSFLIFVYDEIRKLILRRNPGGWVEKETYY
```

Results directory ~/Downloads/ChimeraX/AlphaFold/prediction_[N]

□ Use PDB templates when predicting structures

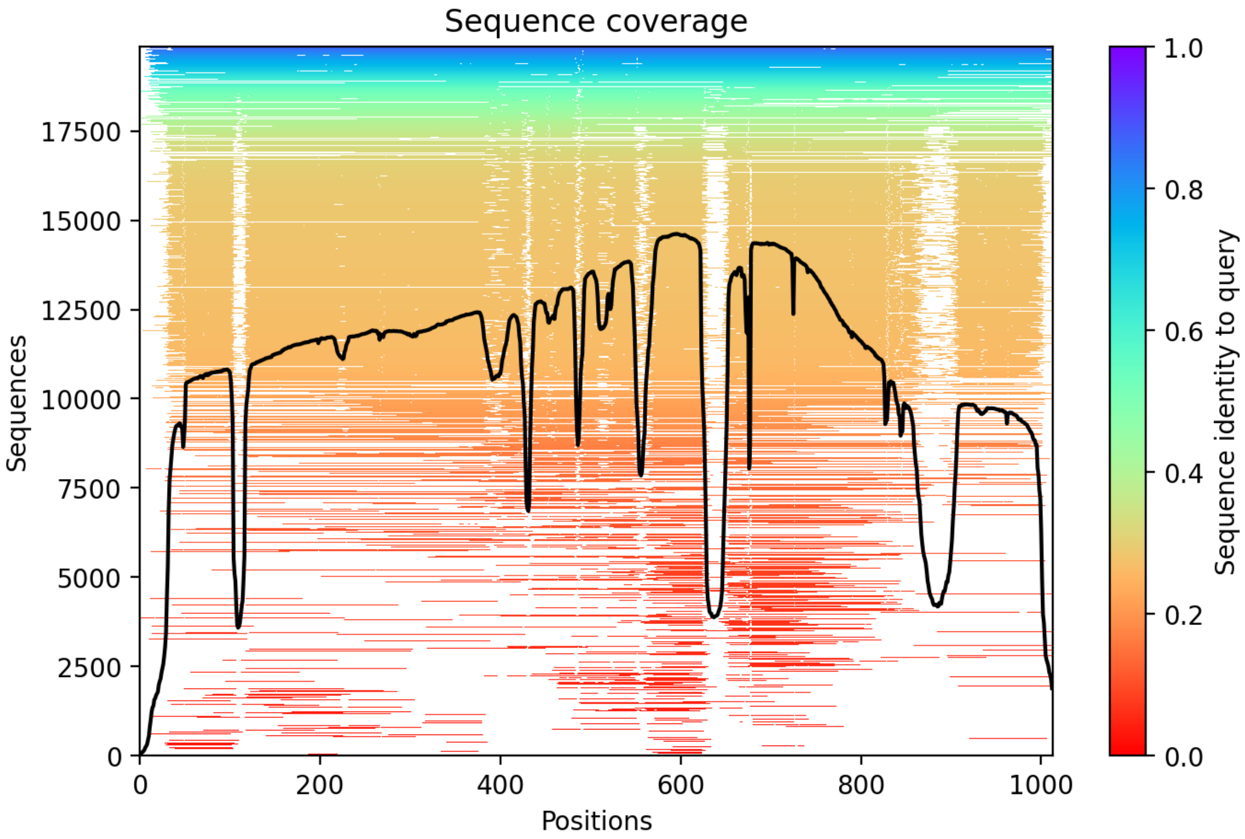
□ Energy-minimize predicted structures

☑ Trim fetched structure to the aligned structure sequence

best_model.pdb 1 ☑

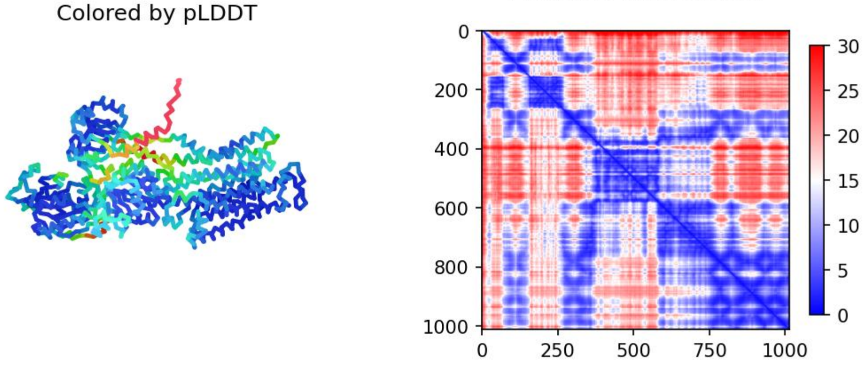
Results available here: [prediction]

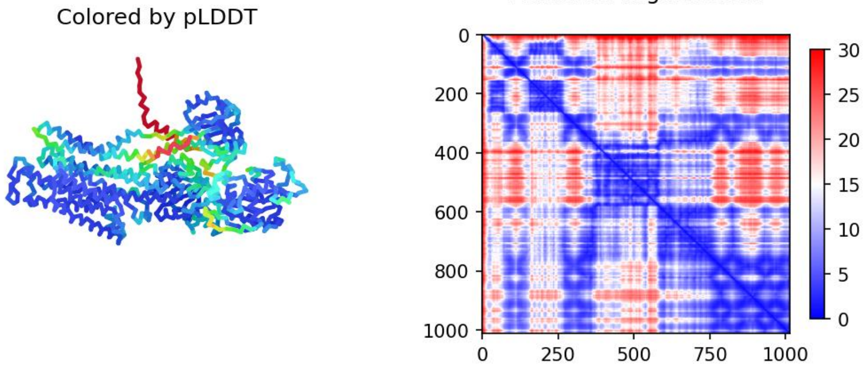
Sequence coverage

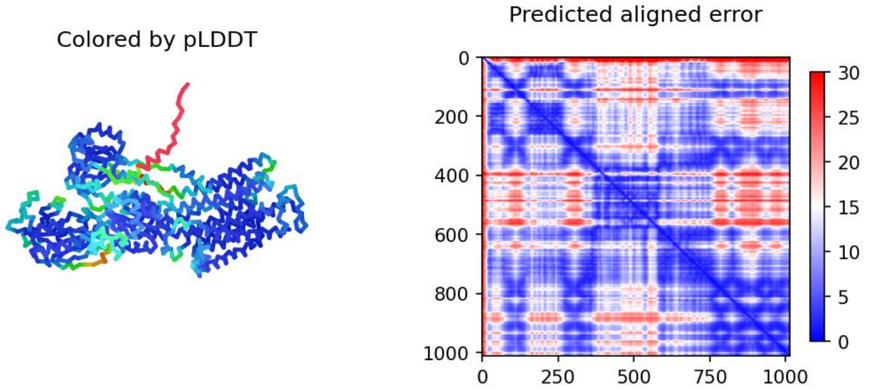


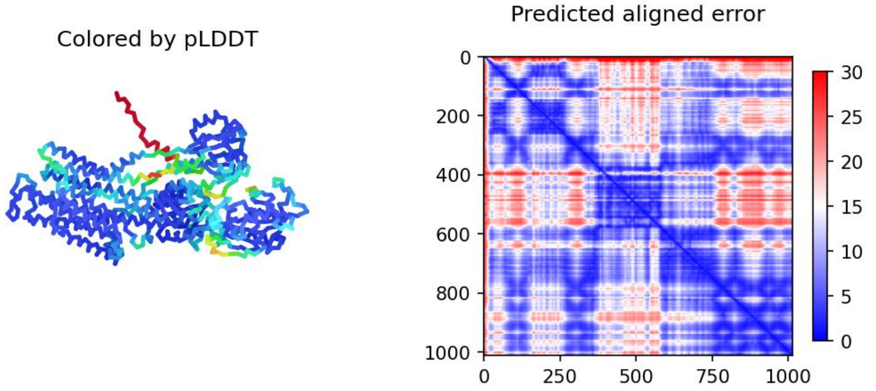
- It refers to how well the target protein sequence is covered by other sequences in the multiple sequence alignment.
- High coverage:
  - Many sequences in the MSA align to the target sequence at each position
  - Informed by a large amount of evolutionary data

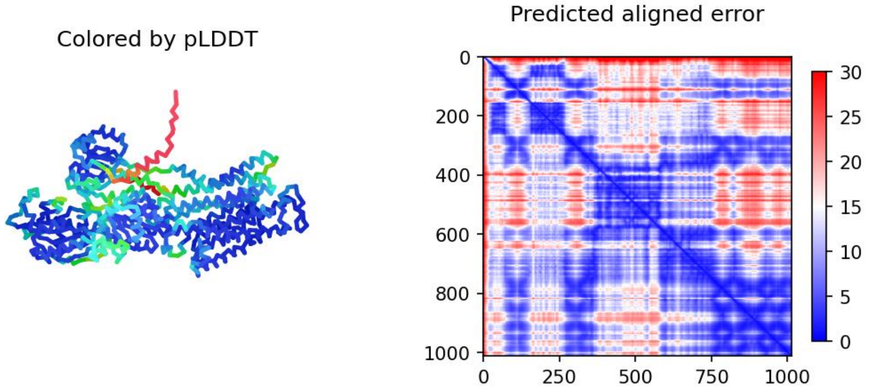
Different predictions



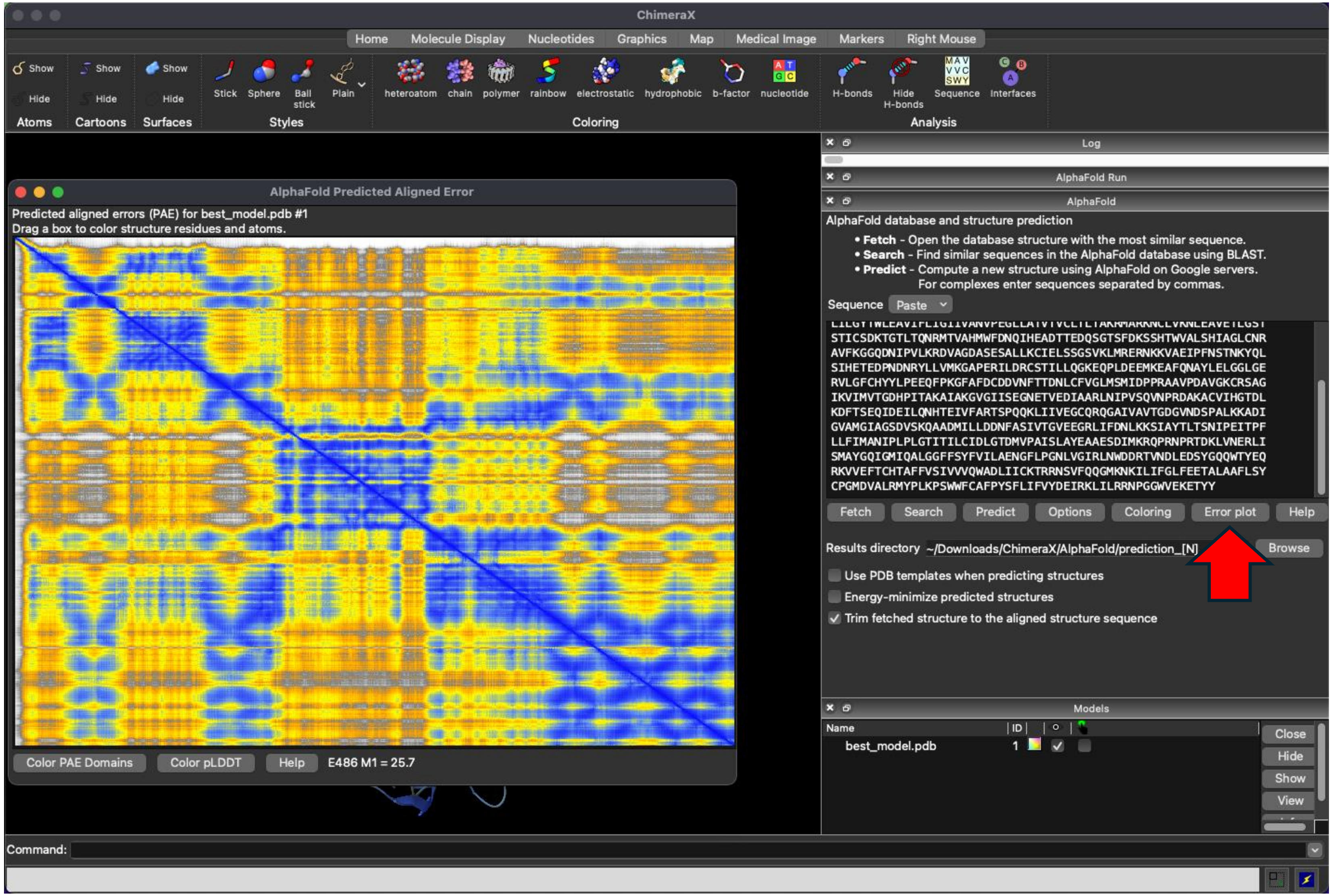








Show error plot in ChimeraX



Predicted aligned errors (PAE) for best_model.pdb #1
Drag a box to color structure residues and atoms.

E486 M1 = 25.7

AlphaFold database and structure prediction

- **Fetch** - Open the database structure with the most similar sequence.
- **Search** - Find similar sequences in the AlphaFold database using BLAST.
- **Predict** - Compute a new structure using AlphaFold on Google servers. For complexes enter sequences separated by commas.

Sequence Paste

```
LILGYIWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKRMARKNCLVKNLEAVETLGST
STICSDKTGTLTQNRMTVAHMWFDNQIHEADTTEDQSGTSFDKSSHTWVALSHIAGLCNR
AVFKGGQDNIPVLKRDVAGDASESALLKCIELSSGSVKLMRERNKKVAEIPFNSTNKYQL
SIHETEDPNDNRYLLVMKGAPERILDRCSTILLQGKEQPLDEEMKEAFQNAYLELGGLGE
RVLGFCHYYLPEEQFPKGFAFDCDDVNFTTDNLCFVGLMSMIDPPRAAVPDAVGKCRSAG
IKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPVSQVNPRDAKACVIHGTDL
KDFTSEQIDEILQNHTEIVFARTSPQQKLIIVEGCQRQGAIVAVTGDGVNDSPALKKADI
GVAMGIAGSDVSKQAADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEITPF
LLFIMANIPLPLGTITILCIDLGTDMVPAISLAYEAAESDIMKRQPRNPRTDKLVNERLI
SMAYGQIGMIQALGGFFSYFVILAENGFLPGNLVGIRLNWDDRTVNDLEDSYGQQWTYEQ
RKVVEFTCHTAFFVSIVVVQWADLIICKTRRNSVFQQGMKNKILIFGLFEETALAAFLSY
CPGMDVALRMYPLKPSWWFCAFPYSFLIFVYDEIRKLILRRNPGGWVEKETYY
```

Results directory ~/Downloads/ChimeraX/AlphaFold/prediction_[N]

□ Use PDB templates when predicting structures

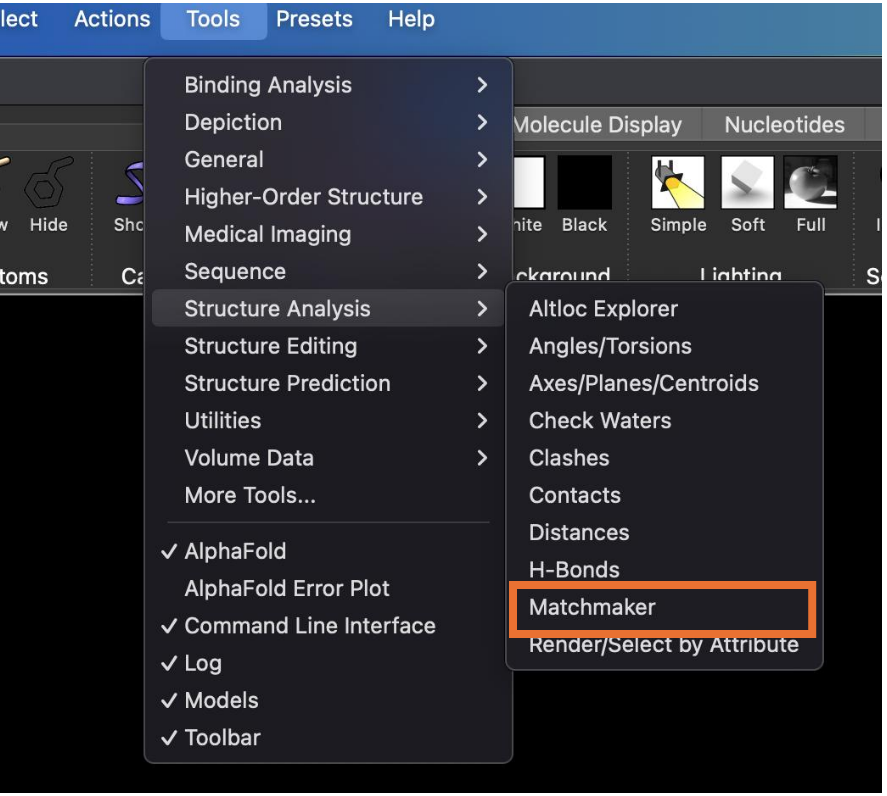
□ Energy-minimize predicted structures

☑ Trim fetched structure to the aligned structure sequence

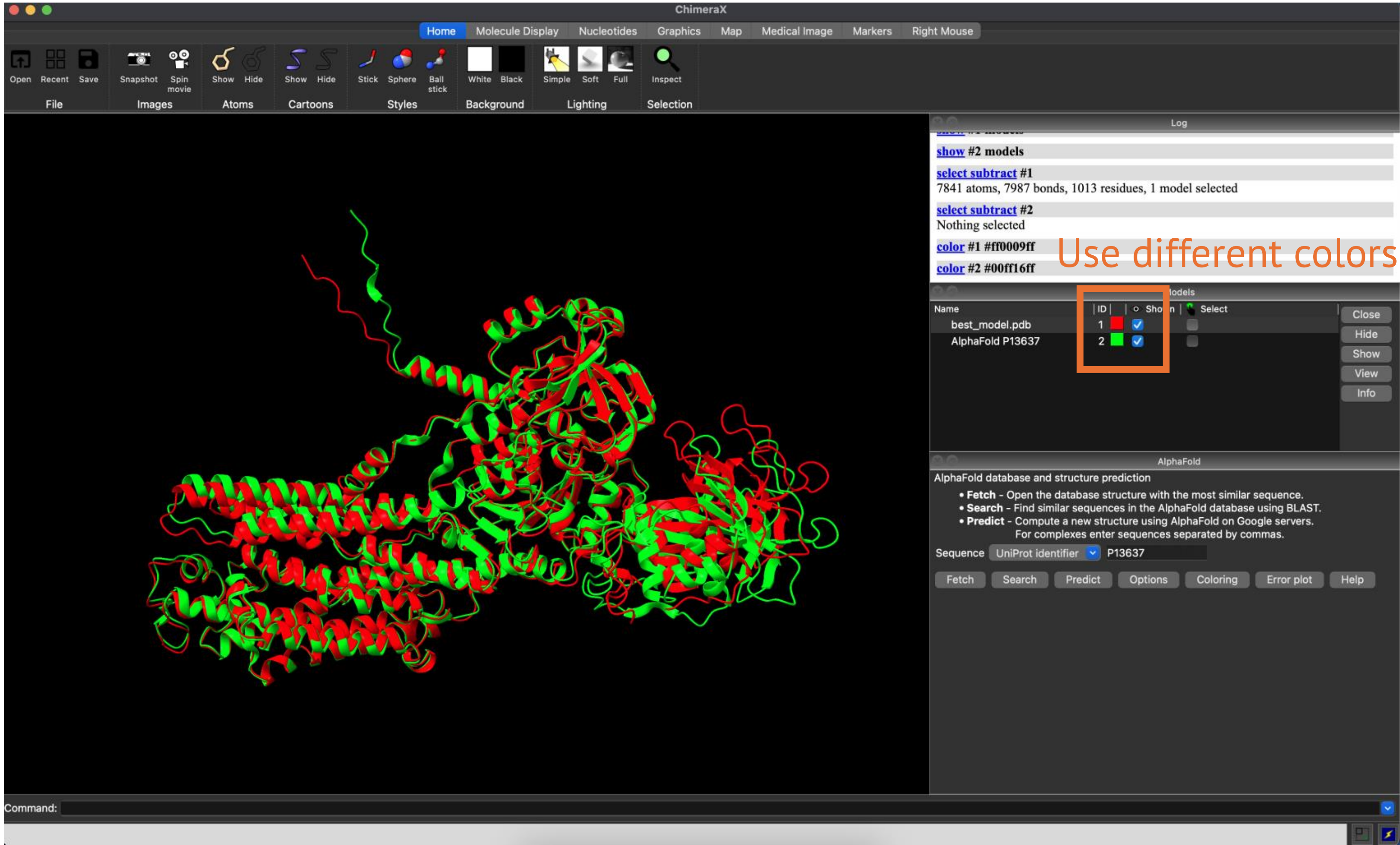
best_model.pdb 1 ✓

https://www.cgl.ucsf.edu/chimerax/

Align the predicted structure with the experimental result



Use matchmaker



```
show #2 models
select subtract #1
7841 atoms, 7987 bonds, 1013 residues, 1 model selected
select subtract #2
Nothing selected
color #1 #ff0009ff
color #2 #00ff16ff
```

Name	ID		Shown	Select
best_model.pdb	1	🟥	☑	□
AlphaFold P13637	2	🟩	☑	□

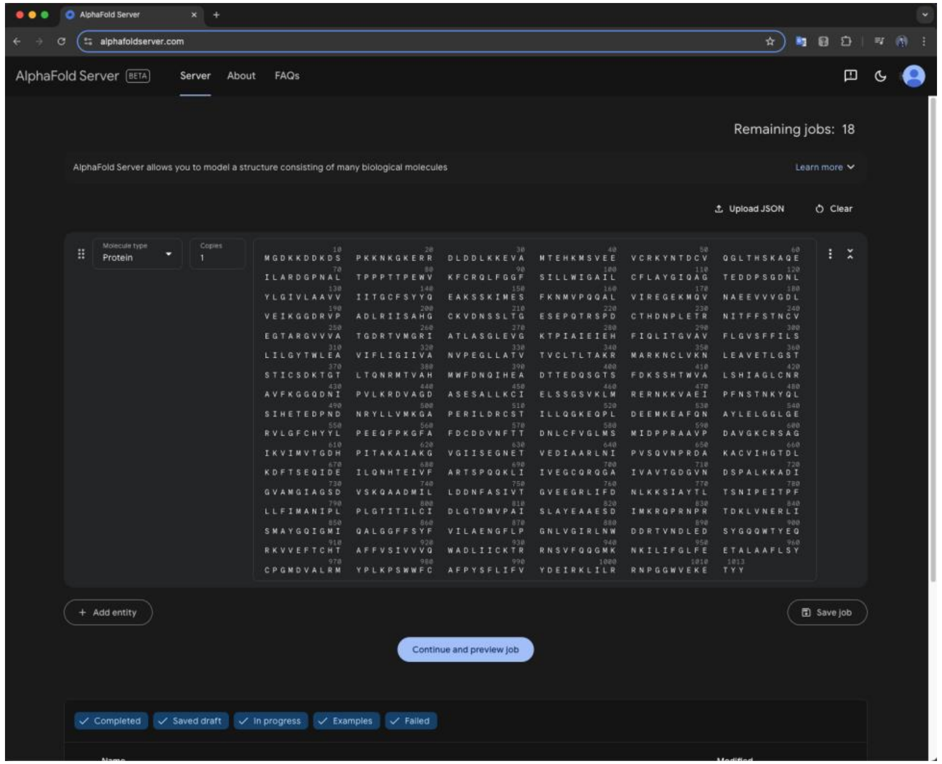
AlphaFold database and structure prediction

- **Fetch** - Open the database structure with the most similar sequence.
- **Search** - Find similar sequences in the AlphaFold database using BLAST.
- **Predict** - Compute a new structure using AlphaFold on Google servers. For complexes enter sequences separated by commas.

Sequence: UniProt identifier P13637

Use different colors

An alternative for Colab: AlphaFold server

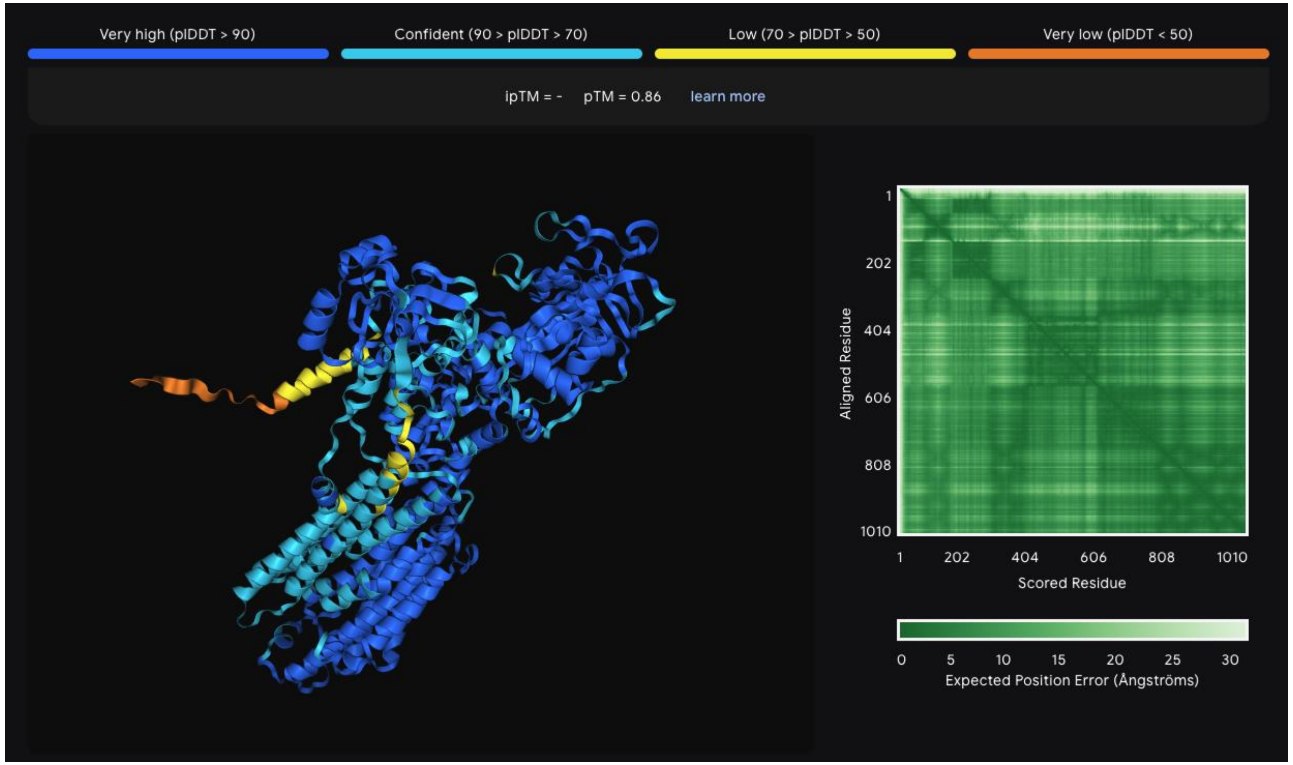


Remaining jobs: 18

AlphaFold Server allows you to model a structure consisting of many biological molecules

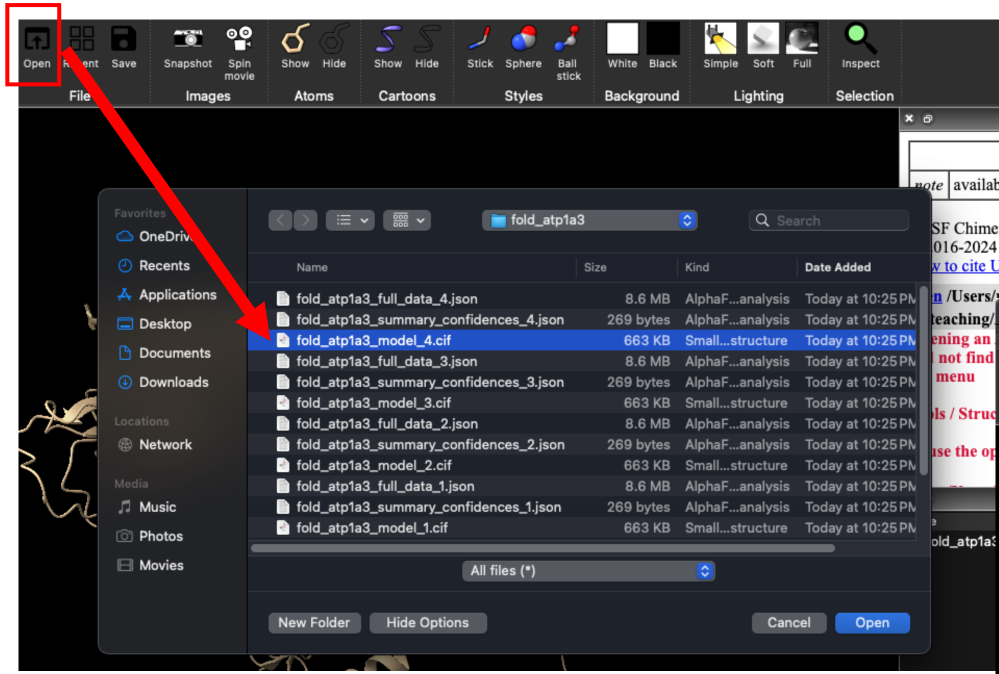
Molecule type: Protein
Copies: 1

```
MGDKKDDKDS PKKNKGKERR DLDDLKKEVA MTEHKMSVEE VCRKYNTDCV QGLTHSKAQE
ILARDGPNAL TPPPTTPEWV KFCRQLFGGF SILLWIGAIL CFLAYGIQAG TEDDPSGDNL
YLGIVLAAVV IITGCFSYYQ EAKSSKIMES FKNMVPQQAL VIREGEKMQV NAEEVVVGDL
VEIKGGDRVP ADLRIISAHG CKVDNSSLTG ESEPQTRSPD CTHDNPLETR NITFFSTNCV
EGTARGVVVA TGDRTVMGRI ATLASGLEVG KTPIAIEIEH FIQLITGVAV FLGVSFFILS
LILGYTWLEA VIFLIGIIVA NVPEGLLATV TVCLTLTAKR MARKNCLVKN LEAVETLGST
STICSDKTGT LTQNRMTVAH MWFDNQIHEA DTTEDQSGTS FDKSSHTWVA LSHIAGLCNR
AVFKGGQDNI PVLKRDVAGD ASESALLKCI ELSSGSVKLM RERNKKVAEI PFNSTNKYQL
SIHETEDPND NRYLLVMKGA PERILDRCST ILLQGKEQPL DEEMKEAFQN AYLELGGLGE
RVLGFCHYYL PEEQFPKGFA FDCDDVNFTT DNLCFVGLMS MIDPPRAAVP DAVGKCRSAG
IKVIMVTGDH PITAKAIAKG VGIISEGNET VEDIAARLNI PVSQVNPRDA KACVIHGTDL
KDFTSEQIDE ILQNHTEIVF ARTSPQQKLI IVEGCQRQGA IVAVTGDGVN DSPALKKADI
GVAMGIAGSD VSKQAADMIL LDDNFASIVT GVEEGRLIFD NLKKSIAYTL TSNIPEITPF
LLFIMANIPL PLGTITILCI DLGTDMVPAI SLAYEAAESD IMKRQPRNPR TDKLVNERLI
SMAYGQIGMI QALGGFFSYF VILAENGFLP GNLVGIRLNW DDRTVNDLED SYGQQWTYEQ
RKVVEFTCHT AFFVSIVVVQ WADLIICKTR RNSVFQQGMK NKILIFGLFE ETALAAFLSY
CPGMDVALRM YPLKPSWWFC AFPYSFLIFV YDEIRKLILR RNPGGWVEKE TYY
```



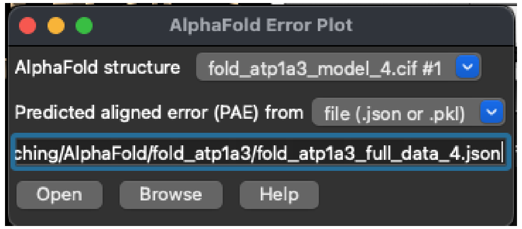
ipTM = - pTM = 0.86 learn more

Open the results from the AlphaFold server in ChimeraX



Name	Size	Kind	Date Added
fold_atp1a3_full_data_4.json	8.6 MB	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_summary_confidences_4.json	269 bytes	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_model_4.cif	663 KB	Small...structure	Today at 10:25 PM
fold_atp1a3_full_data_3.json	8.6 MB	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_summary_confidences_3.json	269 bytes	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_model_3.cif	663 KB	Small...structure	Today at 10:25 PM
fold_atp1a3_full_data_2.json	8.6 MB	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_summary_confidences_2.json	269 bytes	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_model_2.cif	663 KB	Small...structure	Today at 10:25 PM
fold_atp1a3_full_data_1.json	8.6 MB	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_summary_confidences_1.json	269 bytes	AlphaF...analysis	Today at 10:25 PM
fold_atp1a3_model_1.cif	663 KB	Small...structure	Today at 10:25 PM

Open protein structure



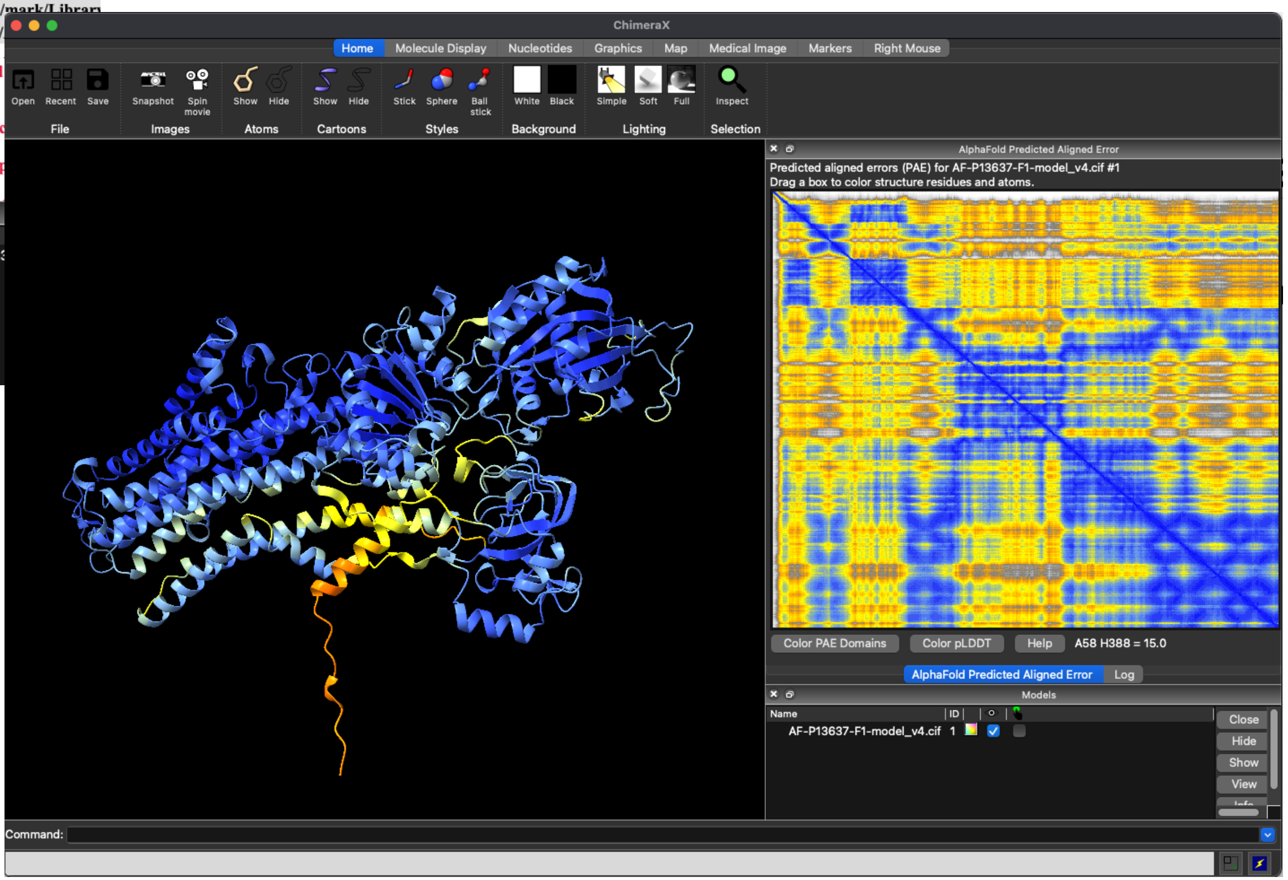
AlphaFold Error Plot

AlphaFold structure: fold_atp1a3_model_4.cif #1

Predicted aligned error (PAE) from: file (.json or .pkl)

ching/AlphaFold/fold_atp1a3/fold_atp1a3_full_data_4.json

Error plot info was safe as json format



Predicted aligned errors (PAE) for AF-P13637-F1-model_v4.cif #1
Drag a box to color structure residues and atoms.

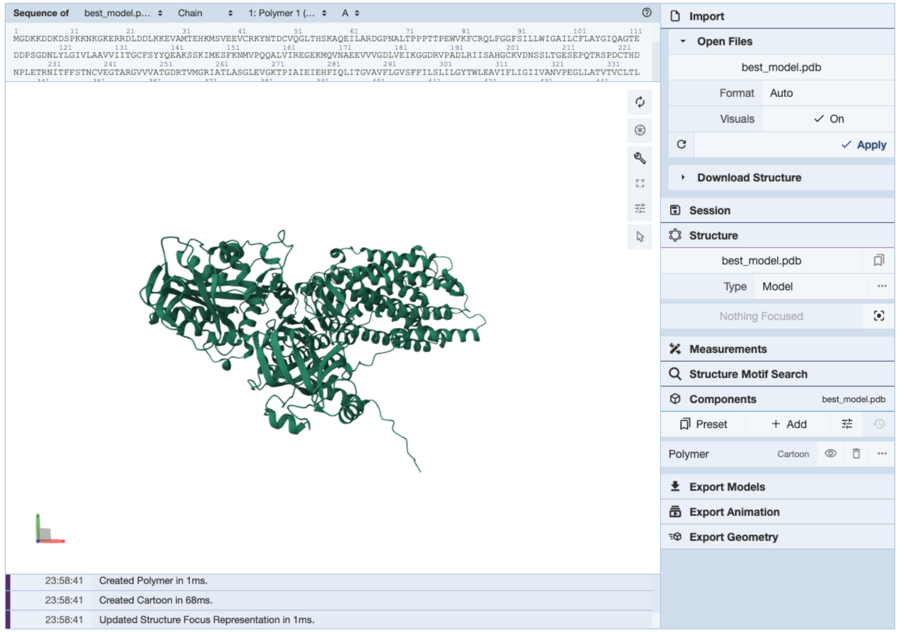
A58 H388 = 15.0

Name	ID
AF-P13637-F1-model_v4.cif	1

Results available [HERE]

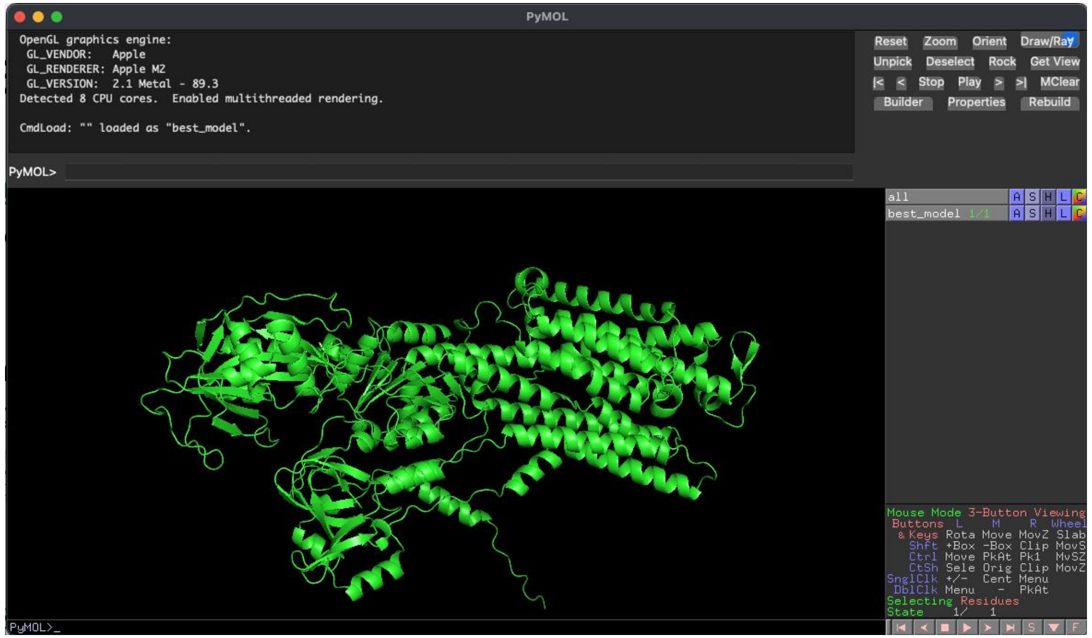
Other visualization tools

- PyMOL: https://github.com/schrodinger/pymol-open-source
  - The open-source version has same functions as the commercial version
  - Installed in conda environment: `conda install conda-forge::pymol-open-source`
- Mol* 3D Viewer
  - Hosted by RCSB protein data bank
  - https://www.rcsb.org/3d-view



```
1         11        21        31        41        51        61        71        81        91        101       111
MGDKKDDKDSPKKNKGKERRDLDDLKKEVAMTEHKMSVEEVCRKYNTDCVQGLTHSKAQEILARDGPNALTPPPTTPEWVKFCRQLFGGFSILLWIGAILCFLAYGIQAGTE
       121       131       141       151       161       171       181       191       201       211       221
DDPSGDNLYLGIVLAAVVIITGCFSYYQEAKSSKIMESFKNMVPQQALVIREGEKMQVNAEEVVVGDLVEIKGGDRVPADLRIISAHGCKVDNSSLTGESEPQTRSPDCTHD
    231       241       251       261       271       281       291       301       311       321       331
NPLETRNITFFSTNCVEGTARGVVVATGDRTVMGRIATLASGLEVGKTPIAIEIEHFIQLITGVAVFLGVSFFILSLILGYTWLEAVIFLIGIIVANVPEGLLATVTVCLTL
```

```
23:58:41   Created Polymer in 1ms.
23:58:41   Created Cartoon in 68ms.
23:58:41   Updated Structure Focus Representation in 1ms.
```



```
OpenGL graphics engine:
 GL_VENDOR:   Apple
 GL_RENDERER: Apple M2
 GL_VERSION:  2.1 Metal - 89.3
Detected 8 CPU cores.  Enabled multithreaded rendering.

CmdLoad: "" loaded as "best_model".
```

```
PyMOL>
```

```
Mouse Mode 3-Button Viewing
 Buttons  L    M    R  Wheel
  & Keys Rota Move MovZ Slab
    Shft +Box -Box Clip MovS
    Ctrl Move PkAt Pk1  MvSZ
    CtSh Sele Orig Clip MovZ
SnglClk +/-  Cent Menu
 DblClk Menu  -   PkAt
Selecting Residues
State    1/   1
```

```
PyMOL>_
```