1. Modelling unknown structures with Alphafold2

Go the Alphafold2 website and click connect. You will be allocated GPU usage and some disk space in Google Colab. (You need to be signed in with your SU account). When you are connected, paste your seqeuence in the 'query sequence' line and give it a job name.

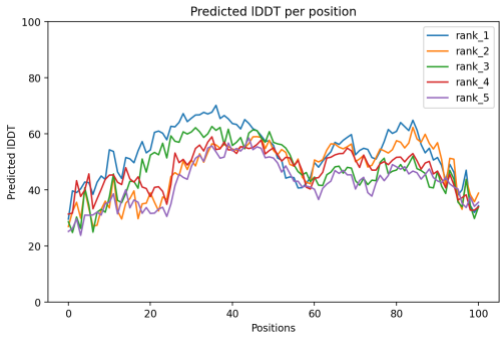
Click **Runtime → Run all**. It should take about 5 min for a monomeric protein of about 100 aa. When a code block is running, it is circled by dashed lines.

When the job is complete, a zipped folder will start downloading automatically.

2. Visualization of predicted structures with ChimeraX

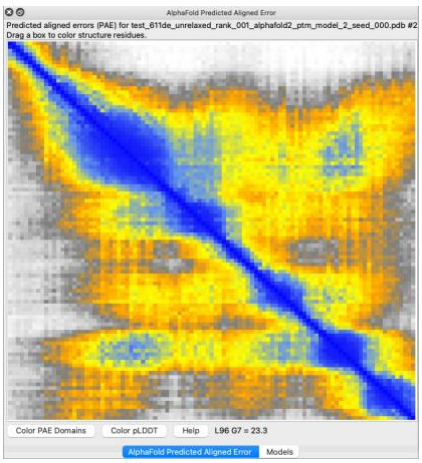
Part 1. Examining AF2 error plots

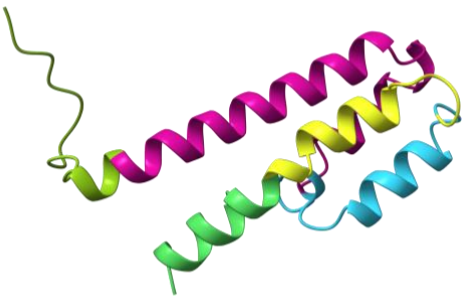
Let's look at the contents of the downloaded file. There are .pdb files named Rank1, Rank2, …etc. Models are ranked based on their average prediction accuracy. Alphafold has two prediction error estimates; Per Residue confidence score (pLDDT) and Predicted Aligned Error (PAE). First look at the **plddt.png** plot. You will see that Rank1 model has higher pLDDT score for almost all residues. Look at the **coverage.png** plot. Do you any sequence coverage from multiple sequence alignment ?



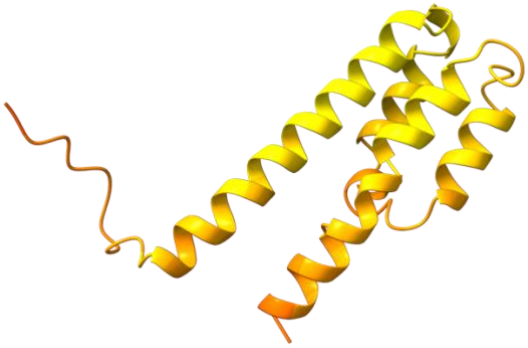
Now let's examine the models in ChimeraX. Drag and drop all your .pdb files into models area. Start with Rank1 model (uncheck all others). Go to **Tools → Structure Prediction → Alphafold Error Plot**. Make sure your Rank1 model is selected. Now click browse and find the .json file for Rank1 model to plot the PAE score. PAE score shows how accurately each residue is positioned *relative to* another. Lower score (blue color) means higher relative position accuracy. Scroll your mouse over the PAE plot; you will see residue pairs and a PAE score at the bottom. Check the position of residue pairs on the structure using selection tool that we learned in previous hands-on. (e.g select #1:10:20). Blue diagonal shows PAE score for each residue relative to itself.

Now click **Color PAE** domains. It clusters the domains into residue pairs with relatively low PAE values.





Now let's choose **Color pLDDT**. It is a colored representation of pLDDT scores (they are actually found in the 'B factor' columns of your PDBs). What can you say about the prediction accuracy of your model ? Compare the four other models as well.

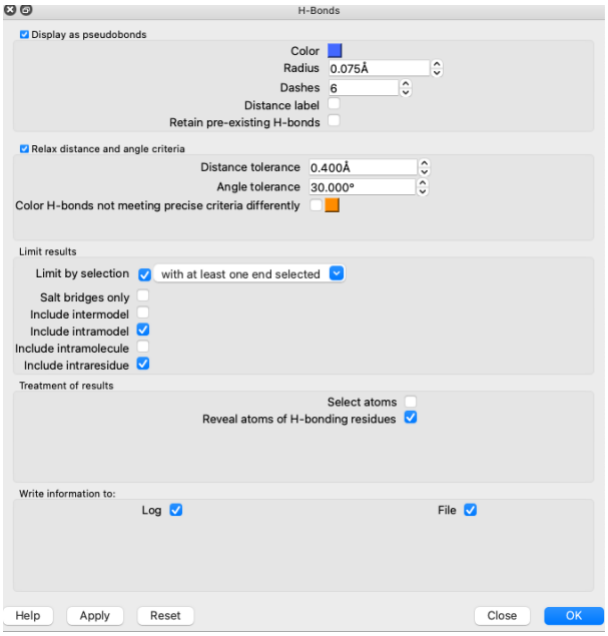


- 100 to 90 – high accuracy expected
- 90 to 70 – backbone expected to be modeled well
- 70 to 50 – low confidence, caution
- 50 to 0 – should not be interpreted, may be disordered

Part 2. Analysis of models with ChimeraX

Sequence viewer: With your Rank1 model active, go to **Tools → Sequence → Sequence Viewer.** Yellow regions correspond to alpha helices, blue regions correspond to beta-strands, white regions correspond to unstructured loops. You can select any region from the sequence and see where it is in the structure.

Hydrogen bonds: Select your Rank1 model by typing select #1 in the command line. Go to **Tools → Structure Analysis → H-bonds.** Visualize the H-bonds with the below criteria. This will show the estimated H-bonds in the whole structure. Also save the H bond information to a file named **Hbonds.dat**. Look at the donor-acceptor distances listed in this document. To delete H-bond representation, go to Molecule Display → Hide H-bonds

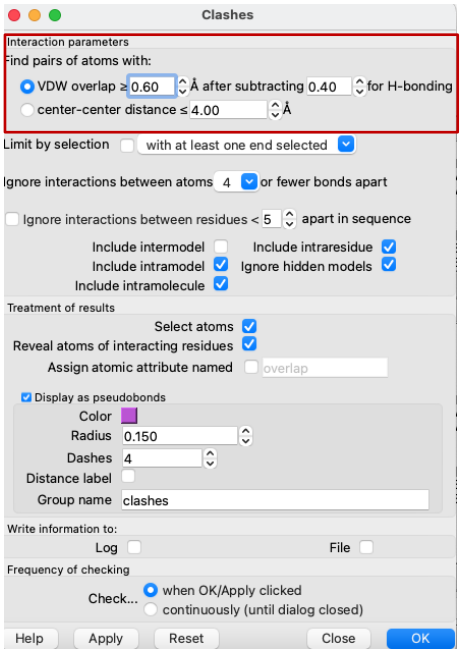


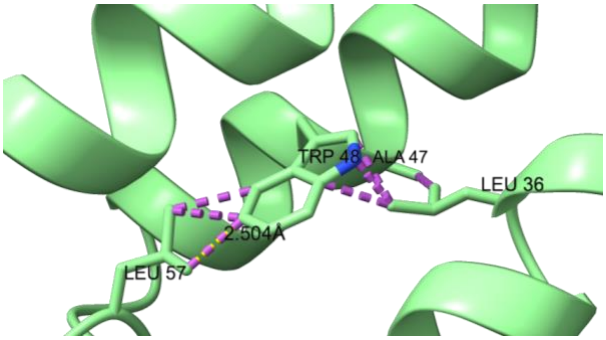
Compare two models: Go to **Tools → Structure Analysis → Matchmaker**. Select the 'Reference structure' as you Rank1, and any other model as 'Structures to match'. To see the sequence alignment, check 'Show pairwise sequence alignment(s)'

Fit is iterated until the most similar parts superimpose and the dissimilar parts stand out. Pairs used in the final fit are shown with light orange boxes on the sequence alignment.

Finding clashes and contacts: Go to **Tools → Structure Analysis → Clashes.** Contacts and clashes are distinguished by interaction parameters. VDW overlap means the sum of the vdw radii minus the distance between their centers. Zero overlap means the sphere are touching. Negative overlap means they are seperated by a distance. Positive overlap means they are penetrating. Find the clashes in your models by applying below criteria.

- ***clashes*** – unfavorable interactions where atoms are too close together; close contacts
- ***contacts*** – all kinds of direct interactions: polar and nonpolar, favorable and unfavorable, including clashes





Side chain clashes

Solvent Excluded Area (SES): Type in the command line `surface #1 enclose #1 probeRadius 1.4`. Now click Surfaces → Show**.** Change the probe radius to 3.0 and look at the new surface.

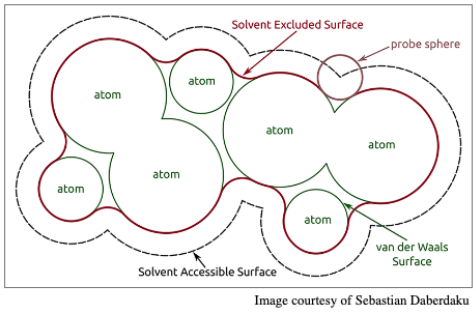
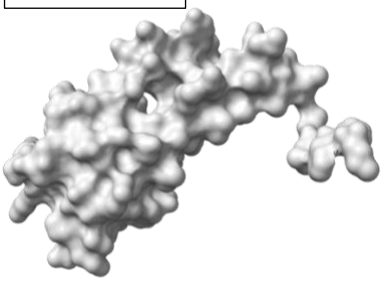
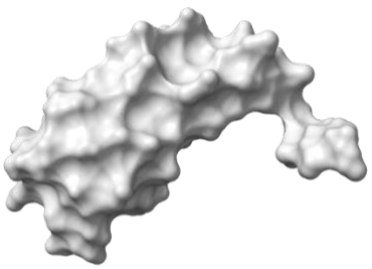


Image courtesy of Sebastian Daberdaku

Probe radius 1.4



Probe radius 3.0



You can also calculate Solvent Accessible Area (SASA) of the protein by typing `measure sasa` but ChimeraX does not visualize such a surface.