AlphaFold Colab

This Colab notebook allows you to easily predict the structure of a protein using a slightly simplified version of AlphaFold v2.3.2.

Differences to AlphaFold v2.3.2

In comparison to AlphaFold v2.3.2, this Colab notebook uses **no templates (homologous structures)** and a selected portion of the BFD database. We have validated these changes on several thousand recent PDB structures. While accuracy will be near-identical to the full AlphaFold system on many targets, a small fraction have a large drop in accuracy due to the smaller MSA and lack of templates. For best reliability, we recommend instead using the full open source AlphaFold, or the AlphaFold Protein Structure Database.

This Colab has a small drop in average accuracy for multimers compared to local AlphaFold installation, for full multimer accuracy it is highly recommended to run AlphaFold locally. Moreover, the AlphaFold-Multimer requires searching for MSA for every unique sequence in the complex, hence it is substantially slower. If your notebook times-out due to slow multimer MSA search, we recommend either using Colab Pro or running AlphaFold locally.

Please note that this Colab notebook is provided for theoretical modelling only and caution should be exercised in its use.

The **PAE file format** has been updated to match AFDB. Please see the AFDB FAQ for a description of the new format.

Citing this work

Any publication that discloses findings arising from using this notebook should cite the AlphaFold paper.

Licenses

This Colab uses the AlphaFold model parameters which are subject to the Creative Commons Attribution 4.0 International (CC BY 4.0) license. The Colab itself is provided under the Apache 2.0 license. See the full license statement below.

More information

You can find more information about how AlphaFold works in the following papers:

- AlphaFold methods paper
- AlphaFold predictions of the human proteome paper
- AlphaFold-Multimer paper

FAQ on how to interpret AlphaFold predictions are here.

If you have any questions not covered in the FAQ, please contact the AlphaFold team at alphafold@deepmind.com.

Get in touch

We would love to hear your feedback and understand how AlphaFold has been useful in your research. Share your stories with us at alphafold@deepmind.com.

Setup

Start by running the 2 cells below to set up AlphaFold and all required software.

1. Install third-party software

Please execute this cell by pressing the *Play* button on the left to download and import third-party software in this Colab notebook. (See the acknowledgements in our readme.)

Note: This installs the software on the Colab notebook in the cloud and not on your computer.

Show code

```
100%                                  100/100 [elapsed: 01:52 remaining: 00:00]
```

2. Download AlphaFold

Please execute this cell by pressing the *Play* button on the left.

Show code

```
100%                                  100/100 [elapsed: 02:57 remaining: 00:00]
Running with Tesla V100-SXM2-16GB GPU
```

Making a prediction

Please paste the sequence of your protein in the text box below, then run the remaining cells via *Runtime* > *Run after*. You can also run the cells individually by pressing the *Play* button on the left.

Note that the search against databases and the actual prediction can take some time, from minutes to hours, depending on the length of the protein and what type of GPU you are allocated by Colab

(see FAQ below).

3. Enter the amino acid sequence(s) to fold ⬇️

Enter the amino acid sequence(s) to fold:

- If you enter only a single sequence, the monomer model will be used (unless you override this below).
- If you enter multiple sequences, the multimer model will be used.

sequence_1: " MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVY "

sequence_2: " Insert text here "

sequence_3: " Insert text here "

sequence_4: " Insert text here "

sequence_5: " Insert text here "

sequence_6: " Insert text here "

sequence_7: " Insert text here "

sequence_8: " Insert text here "

sequence_9: " Insert text here "

sequence_10: " Insert text here "

sequence_11: " Insert text here "

sequence_12: " Insert text here "

sequence_13: " Insert text here "

sequence_14: " Insert text here "

sequence_15: " Insert text here "

sequence_16: " Insert text here "

sequence_17: " Insert text here "

sequence_18: " Insert text here "

sequence_19: " Insert text here "

sequence_20: " Insert text here "

Select this checkbox to run the multimer model for a single sequence. For proteins that are monomeric in their native form, or for very large single chains you may get better accuracy and memory efficiency by using the multimer model.

Due to improved memory efficiency the multimer model has a maximum limit of 4000 residues, while the monomer model has a limit of 2500 residues.

use_multimer_model_for_monomers: ☐

Show code

```
Using the single-chain model.
```

4. Search against genetic databases

Once this cell has been executed, you will see statistics about the multiple sequence alignment (MSA) that will be used by AlphaFold. In particular, you'll see how well each residue is covered by similar sequences in the MSA.

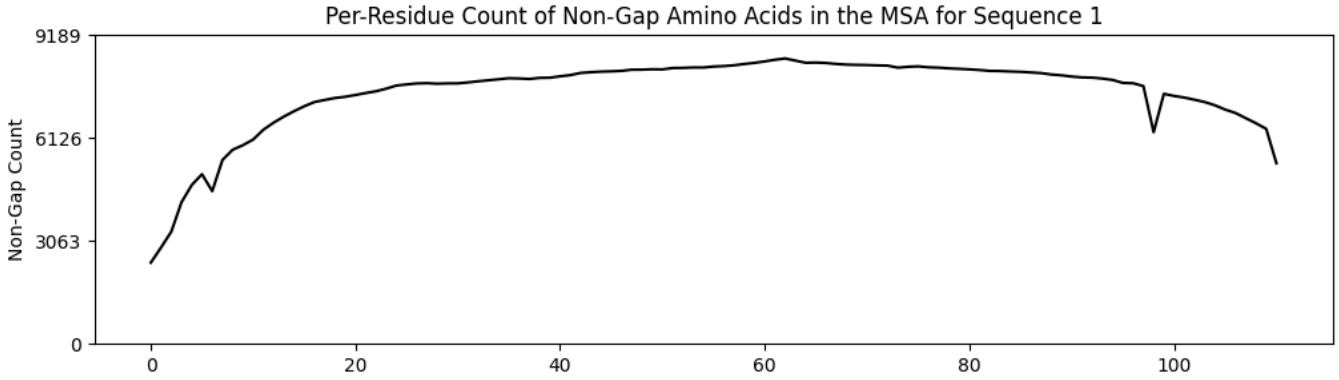
Show code

```
Getting MSA for all sequences
```

```
Searching mgnify: 100%  199/199 [elapsed: 35:47 remaining: 00:00]
```

```
7858 unique sequences found in uniref90 for sequence 1
1137 unique sequences found in smallbfd for sequence 1
421 unique sequences found in mgnify for sequence 1

9191 unique sequences found in total for sequence 1
```



5. Run AlphaFold and download prediction

Once this cell has been executed, a zip-archive with the obtained prediction will be automatically downloaded to your computer.

In case you are having issues with the relaxation stage, you can disable it below. Warning: This means that the prediction might have distracting small stereochemical violations.

run_relax: ☑

Relaxation is faster with a GPU, but we have found it to be less stable. You may wish to enable GPU for higher performance, but if it doesn't converge we suggested reverting to using without GPU.

relax_use_gpu: ☐

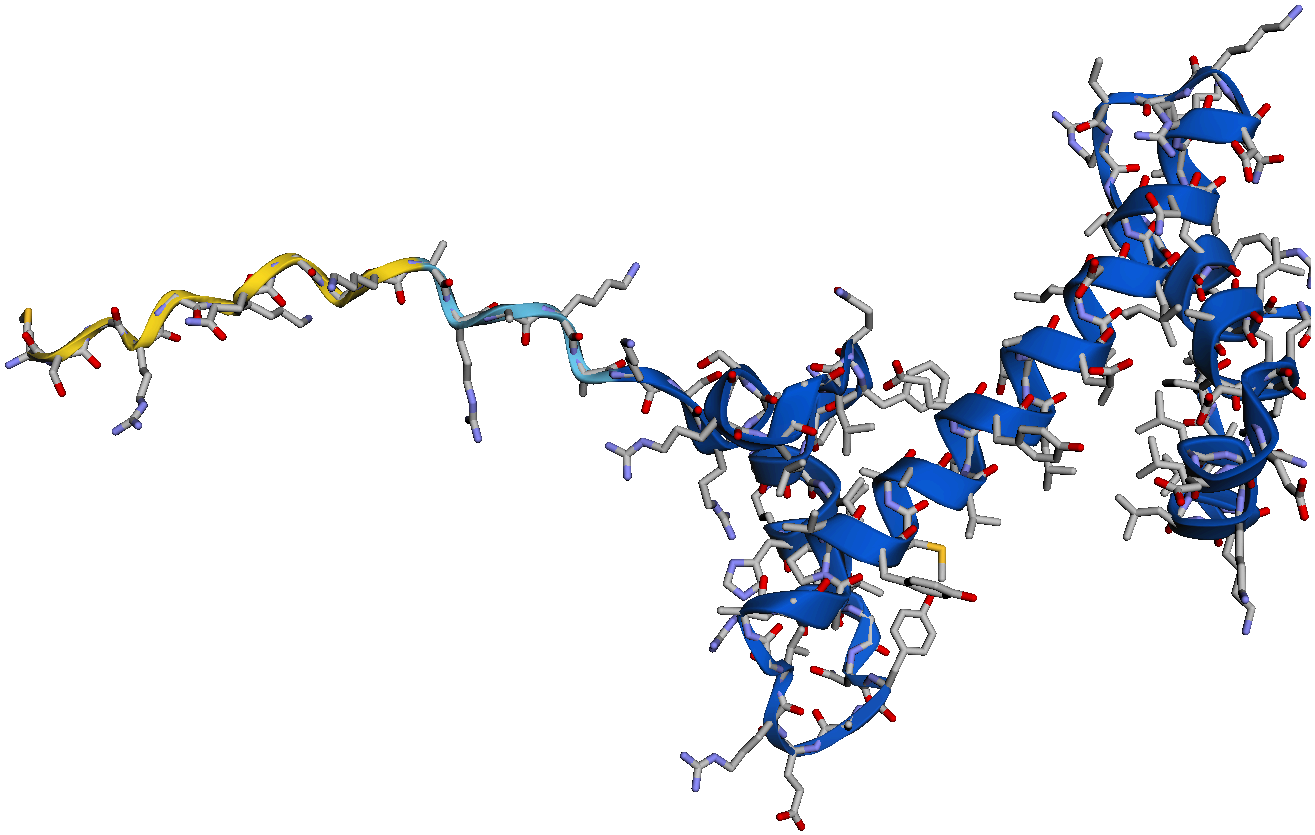
The multimer model will continue recycling until the predictions stop changing, up to the limit set here. For higher accuracy, at the potential cost of longer inference times, set this to 20.

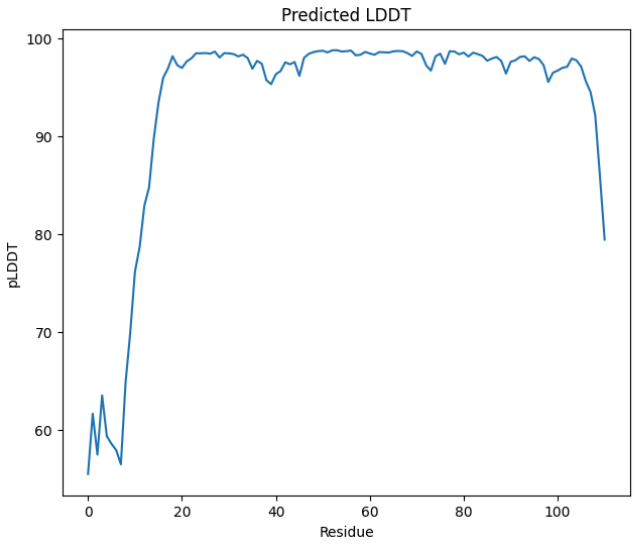
multimer_model_max_num_recycles: 3

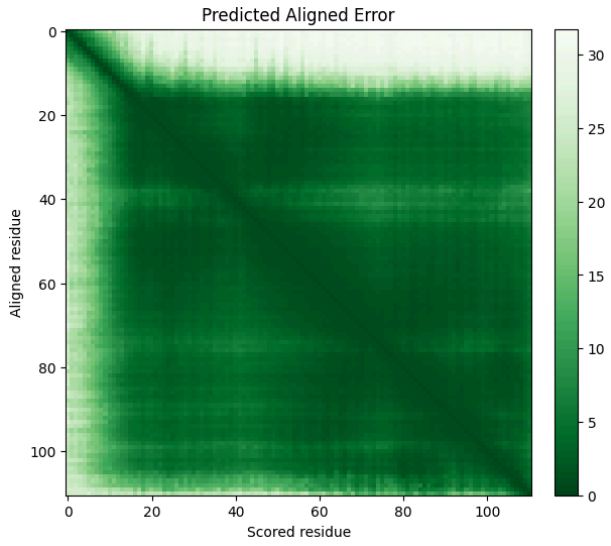
Show code

AMBER relaxation: 100% 7/7 [elapsed: 21:07 remaining: 00:00]

```
WARNING:root:Warning: importing 'simtk.openmm' is deprecated.  Import 'openmm' i
```







Interpreting the prediction

In general predicted LDDT (pLDDT) is best used for intra-domain confidence, whereas Predicted Aligned Error (PAE) is best used for determining between domain or between chain confidence.

Please see the AlphaFold methods paper, the AlphaFold predictions of the human proteome paper, and the AlphaFold-Multimer paper as well as our FAQ on how to interpret AlphaFold predictions.

FAQ & Troubleshooting

- How do I get a predicted protein structure for my protein?
  - Click on the *Connect* button on the top right to get started.
  - Paste the amino acid sequence of your protein (without any headers) into the "Enter the amino acid sequence to fold".
  - Run all cells in the Colab, either by running them individually (with the play button on the left side) or via *Runtime* > *Run all.* Make sure you run all 5 cells in order.
  - The predicted protein structure will be downloaded once all cells have been executed. Note: This can take minutes to hours - see below.
- How long will this take?
  - Downloading the AlphaFold source code can take up to a few minutes.
  - Downloading and installing the third-party software can take up to a few minutes.
  - The search against genetic databases can take minutes to hours.
  - Running AlphaFold and generating the prediction can take minutes to hours, depending on the length of your protein and on which GPU-type Colab has assigned you.
- My Colab no longer seems to be doing anything, what should I do?
  - Some steps may take minutes to hours to complete.
  - If nothing happens or if you receive an error message, try restarting your Colab runtime via *Runtime* > *Restart runtime*.

    - If this doesn't help, try resetting your Colab runtime via *Runtime* > *Factory reset runtime*.
- How does this compare to the open-source version of AlphaFold?
    - This Colab version of AlphaFold searches a selected portion of the BFD dataset and currently doesn't use templates, so its accuracy is reduced in comparison to the full version of AlphaFold that is described in the AlphaFold paper and Github repo (the full version is available via the inference script).
- What is a Colab?
    - See the Colab FAQ.
- I received a warning "Notebook requires high RAM", what do I do?
    - The resources allocated to your Colab vary. See the Colab FAQ for more details.
    - You can execute the Colab nonetheless.
- I received an error "Colab CPU runtime not supported" or "No GPU/TPU found", what do I do?
    - Colab CPU runtime is not supported. Try changing your runtime via *Runtime* > *Change runtime type* > *Hardware accelerator* > *GPU*.
    - The type of GPU allocated to your Colab varies. See the Colab FAQ for more details.
    - If you receive "Cannot connect to GPU backend", you can try again later to see if Colab allocates you a GPU.
    - Colab Pro offers priority access to GPUs.
- I received an error "ModuleNotFoundError: No module named ...", even though I ran the cell that imports it, what do I do?
    - Colab notebooks on the free tier time out after a certain amount of time. See the Colab FAQ. Try rerunning the whole notebook from the beginning.
- Does this tool install anything on my computer?
    - No, everything happens in the cloud on Google Colab.
    - At the end of the Colab execution a zip-archive with the obtained prediction will be automatically downloaded to your computer.
- How should I share feedback and bug reports?
    - Please share any feedback and bug reports as an issue on Github.

Related work

Take a look at these Colab notebooks provided by the community (please note that these notebooks may vary from our validated AlphaFold system and we cannot guarantee their accuracy):

- The ColabFold AlphaFold2 notebook by Sergey Ovchinnikov, Milot Mirdita and Martin Steinegger, which uses an API hosted at the Södinglab based on the MMseqs2 server (Mirdita

et al. 2019, Bioinformatics) for the multiple sequence alignment creation.

License and Disclaimer

This is not an officially-supported Google product.

This Colab notebook and other information provided is for theoretical modelling only, caution should be exercised in its use. It is provided ‘as-is’ without any warranty of any kind, whether expressed or implied. Information is not intended to be a substitute for professional medical advice, diagnosis, or treatment, and does not constitute medical or other professional advice.

AlphaFold Code License