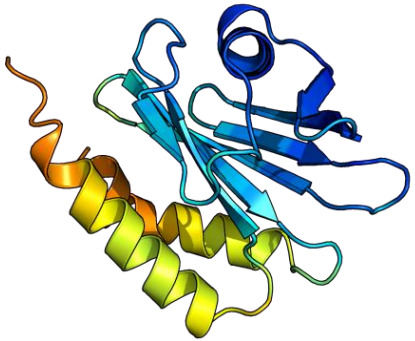
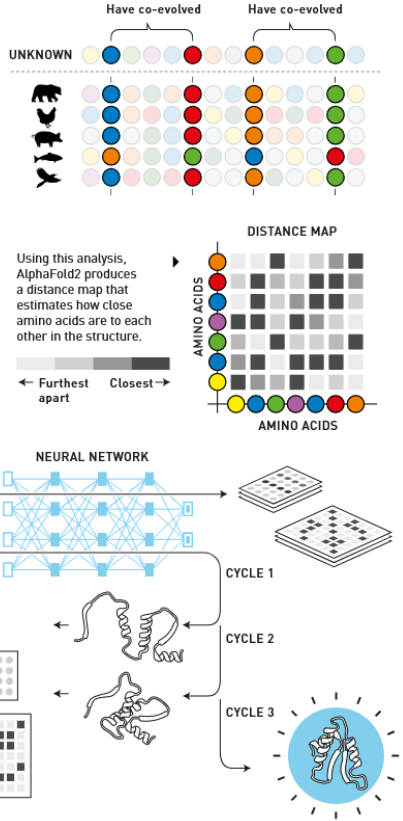
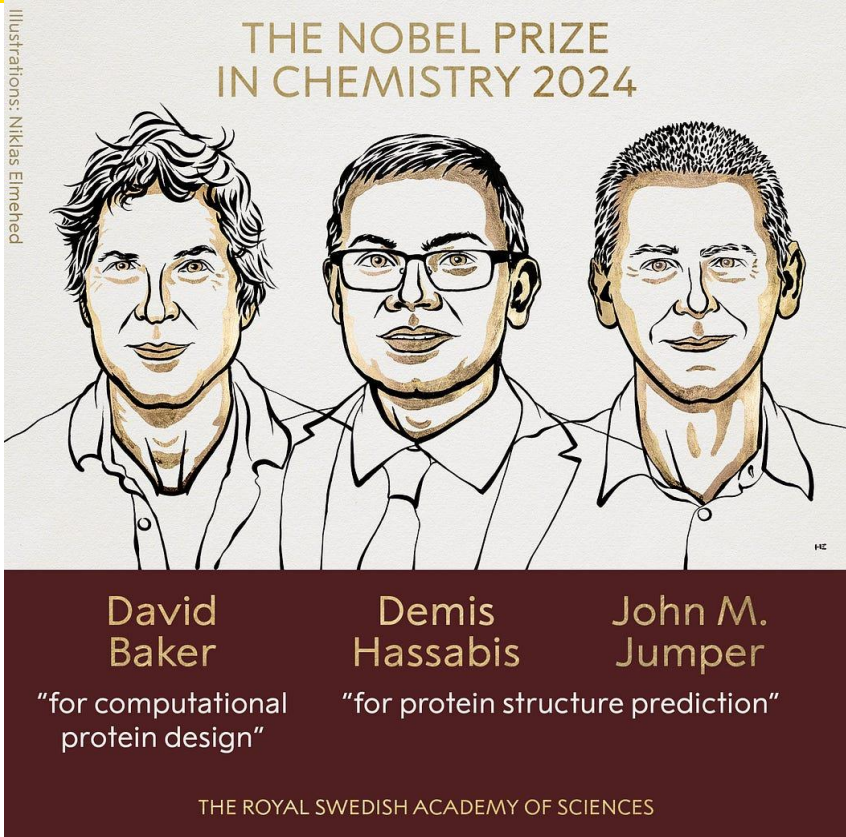
Dr Keiran Rowell – *The Florey* Institute of Neuroscience and Mental Health - 13 November 2025

Easy protein structure prediction with ProteinFold

AI/Deep Learning for biomedical research





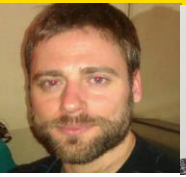


unsw.edu.au/research/facilities-and-infrastructure/find-a-facility/**sbf**

Structural Biology Facility

Developing ProteinFold

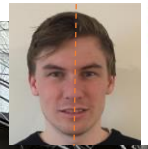
Protein structure prediction from a spreadsheet (NextFlow)

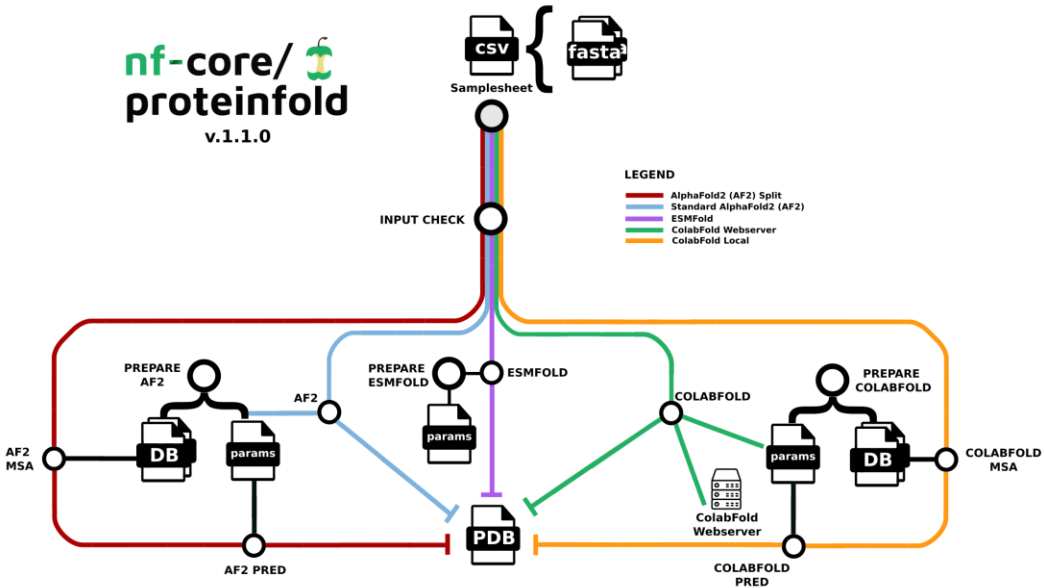


CRG Centre for Genomic Regulation

Australian BioCommons

Structural Biology Facility





Protein prediction structure

Q&A at any time

How did we get here? – 15 mins

- A quick history of protein structure folding – how did we get to AlphaFold?
- Looking under the hood – roughly how does AlphaFold work

Software access – 10 mins

- **Code** choice – Different accuracies & speeds
- **Licenses** – restrictions to non-commercial use, and restriction-free software

Pipelines automating research – 15 mins

- **Research pipeline** – different expertise pipelines between facilities
- Providing **easy access** – using Open Source code and a web form

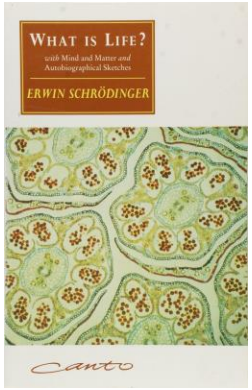
Live demo – 15 mins

- Medical **proteins** – example class of fatty-acid binding proteins predicted by AlphaFold
- ProteinFold – AlphaFold & co. through a sample spreadsheet
- **On-Demand** ProteinFold – automated webform to a university cluster

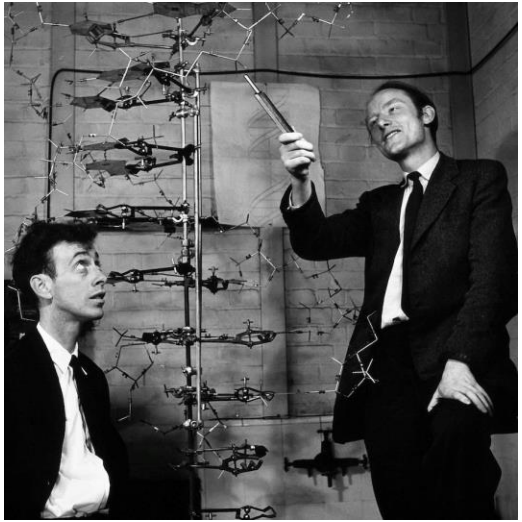
Do it yourself – remaining time

- Running AlphaFold2 in a Google Colab notebook

Models in Chemical Biology - function follows chemical form

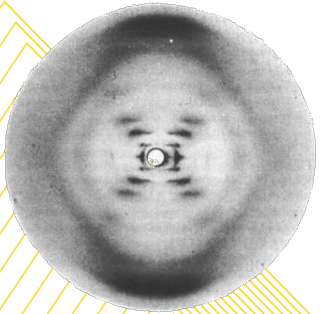


Schrödinger speculates on the molecules of life **1944**

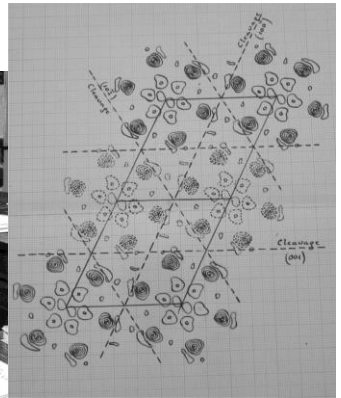


Watson & Crick's DNA double-helix model **1953**

Perutz & Kendrew's model of the 3D structure of a protein (myoglobin) **1957**



Wilkins & Fraklin's X-ray diffraction images of DNA **1953**

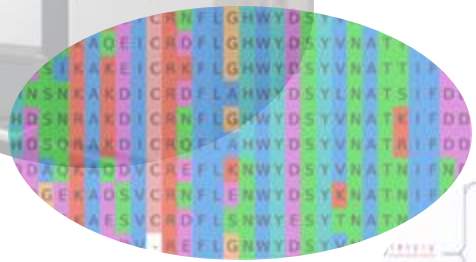


Kathleen Lonsdale resolves the structure of benzene **1924**

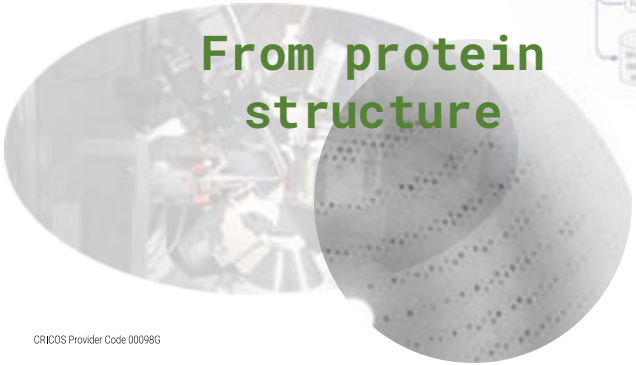
UNSW SYDNEY

Deep Learning trained on experimental data

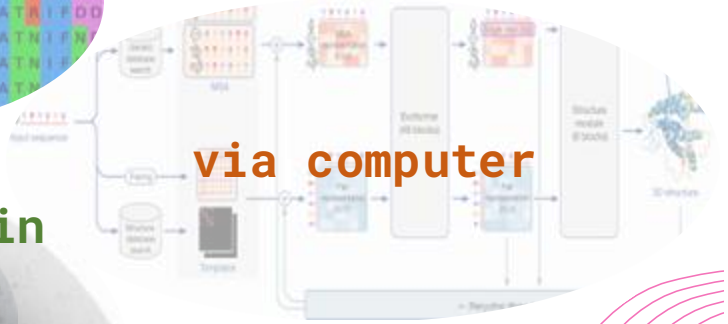
From genetic sequence



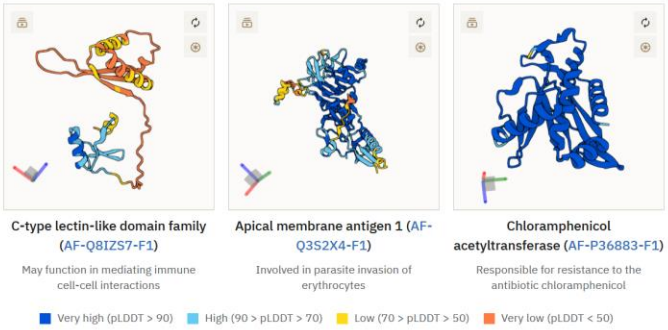
From protein structure



via computer

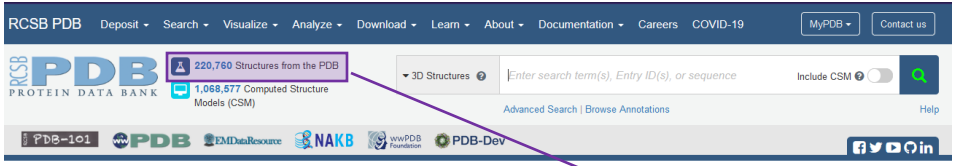


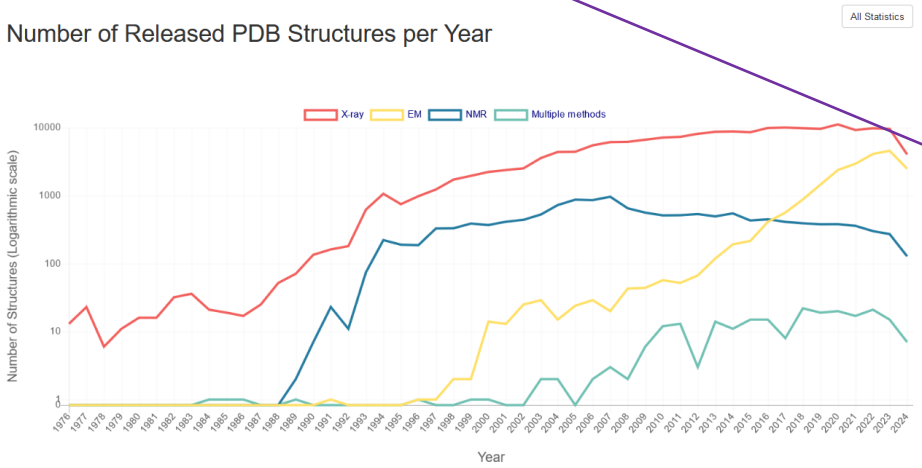
To modelled structure prediction

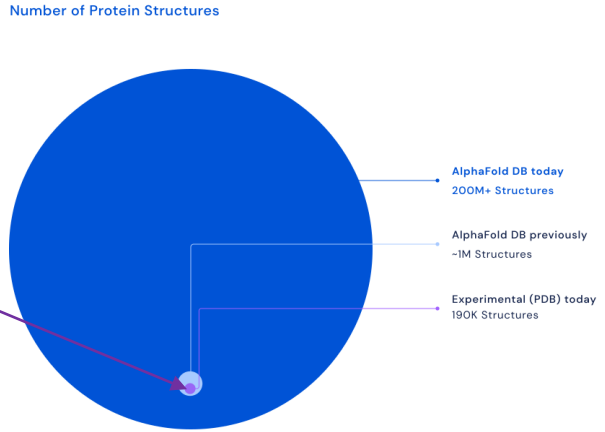


Source: https://www.ebi.ac.uk/training/online/courses/alphafold/inputs-and-outputs/evaluating-alphafolds-predicted-structures-using-confidence-scores/plddt-understanding-local-confidence/

X-rays ruled 20thC – CryoEM & Calcs the 21st







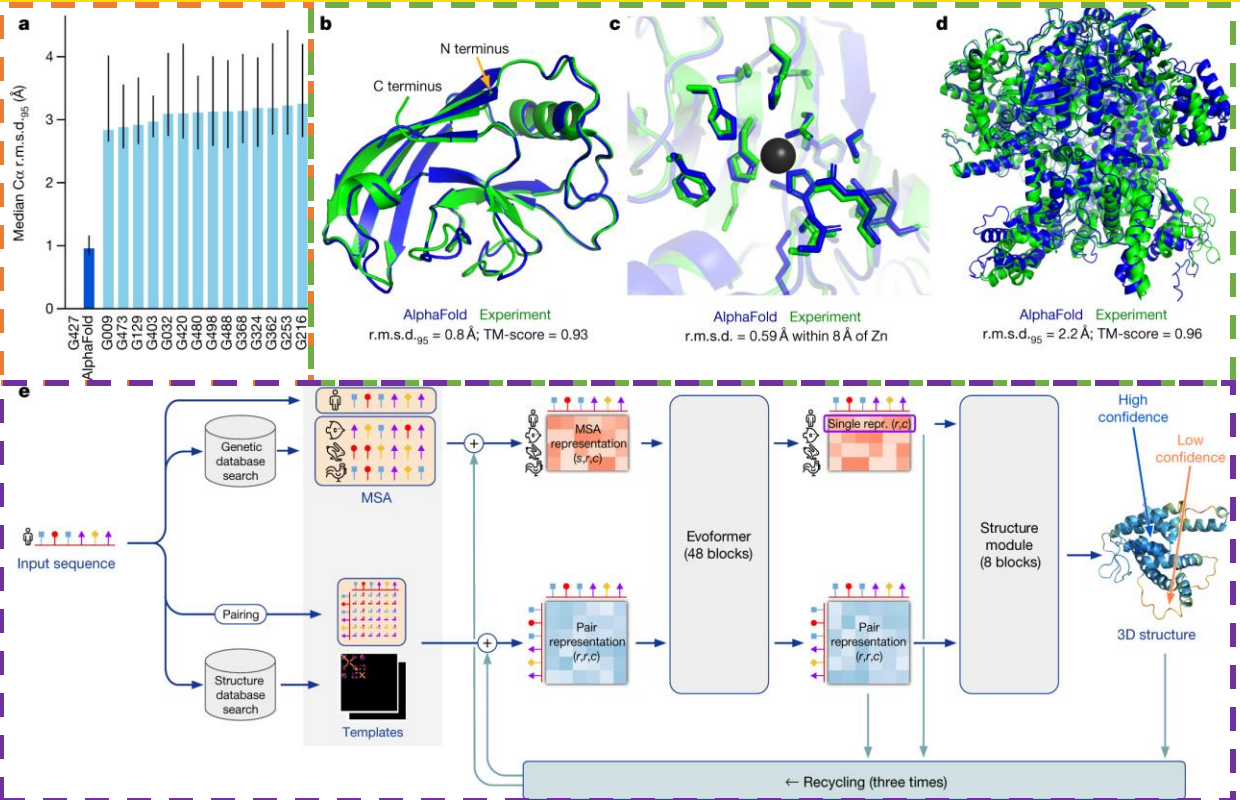
[Protein Data Bank](#) (**PDB**)

Compute power – Moore's law +
DNA Sequencing – super-Moore's law +
Deep Learning – context scaling laws
=
Comp.Struct.Bio. - **1000x** in 2 years

AlphaFold2

A leap forward in computational prediction accuracy

Fig. 1: AlphaFold produces highly accurate structures – from *"Highly accurate protein structure prediction with AlphaFold"* - Nature, **596**, 583-89 (2021).



Improvements

- More than 2x as accurate as anyone else
- The global fold reliable when compared to very expensive experiments
- Set the algorithm architecture that is that standard today

Alpha'Fold' programs – the way they 'fold' is very strange

No (explicit) Physics or Chemistry

Large Language Models (*e.g.* GPT) learns context from pure text tokens

EvoFormers learns 1D -> 2D -> 3D from sequence and solved structures in Protein Data Bank (**PDB**)

PDB ID: 7RDT

PDB ID: 7B3A

Supplementary Videos:
Ahdritz, G., Bouatta, N., Floristean, C. *et al.* OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat Methods* **21**, 1514–1524 (2024). https://doi.org/10.1038/s41592-024-02272-z

AlphaFold2

Ingenious algorithm design made this breakthrough

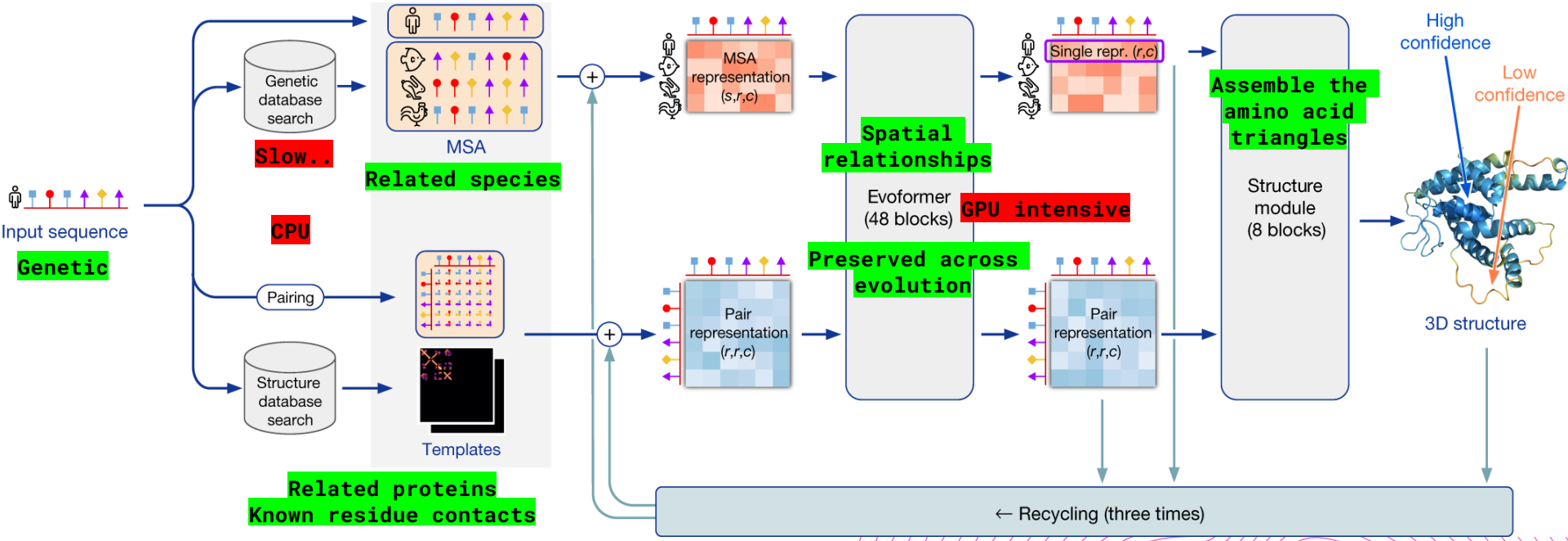


Fig. 1: AlphaFold produces highly accurate structures – from *"Highly accurate protein structure prediction with AlphaFold"* - Nature, **596**, 583-89 (2021).

Alpha'Fold'-*style* software

The largest Big Tech and academic groups in world dedicated serious expertise & resources into developing these programs.

They are mostly free and open source (academic, non-commercial work).
If you develop the skills to run the code on local hardware.

Google DeepMind — github.com/google-deepmind/**alphafold**
github.com/google-deepmind/**alphafold3**

Facebook AI Research — github.com/facebookresearch/**esm**

ByteDance — github.com/**bytedance**/Protenix

github.com/aqlaboratory/**openfold**

MIT | JAMEEL CLINIC — github.com/jwohlwend/**boltz**

github.com/sokrypton/**ColabFold**

INSTITUTE FOR Protein Design, UNIVERSITY of WASHINGTON — github.com/baker-laboratory/**RoseTTAFold-All-Atom**

UNSW SYDNEY

Please cite!
- Code
- Papers
- Clusters
- Facilities

Where to run?

Galaxy Aus

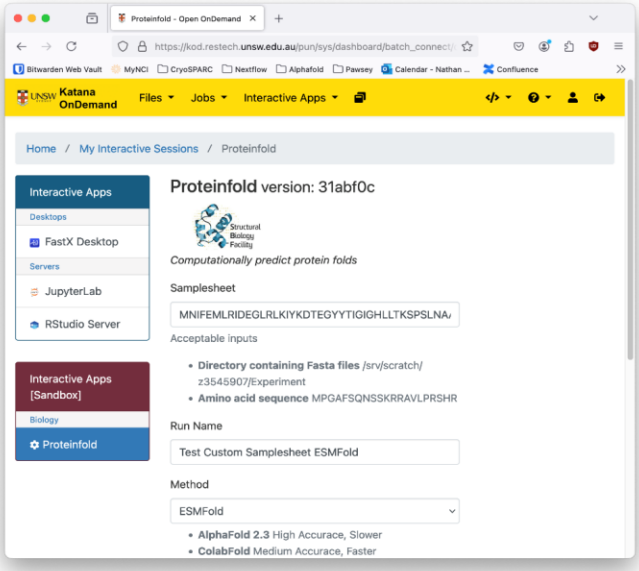
Australian BioCommons

ABOUT ACTIVITIES SERVICES TRAINING & EVENTS DOMAINS NEWS CONTACT HELP

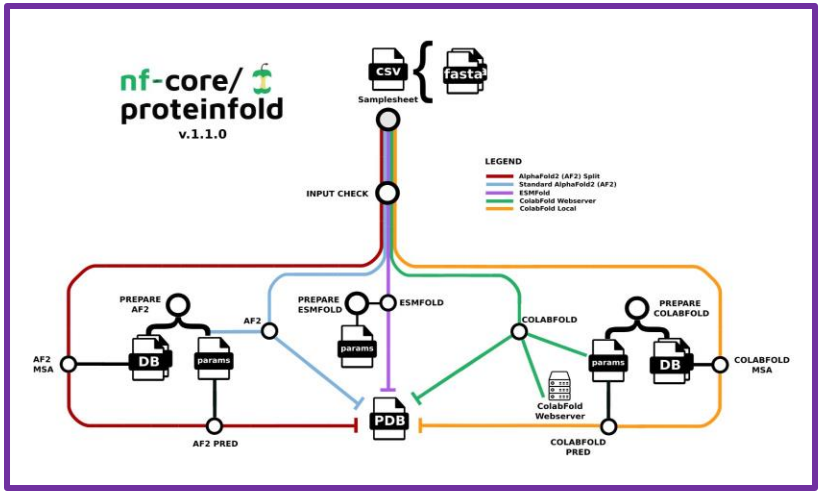
Australian AlphaFold Service

AlphaFold is an artificial intelligence (AI) system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy that is competitive with experimental methods (see Jumper *et al. Nature* 2021).

Uni



proteinfold



UNSW SYDNEY

All open code!

AlphaFold2:
(Google Deepmind)
github.com/google-deepmind/alphafold

Galaxy:
(Cam Hyde, Grace Hall)
github.com/usegalaxy-au/tools-au/tree/master/tools/alphafold

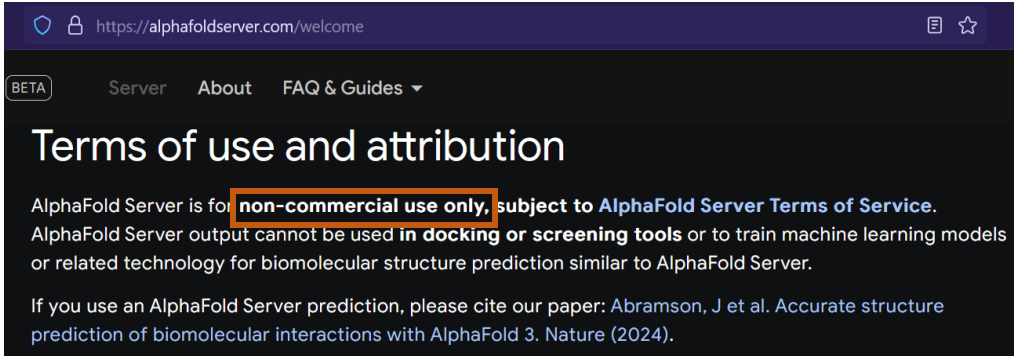
Aus Service:
(Aus BioCommons)
biocommons.org.au/alphafold

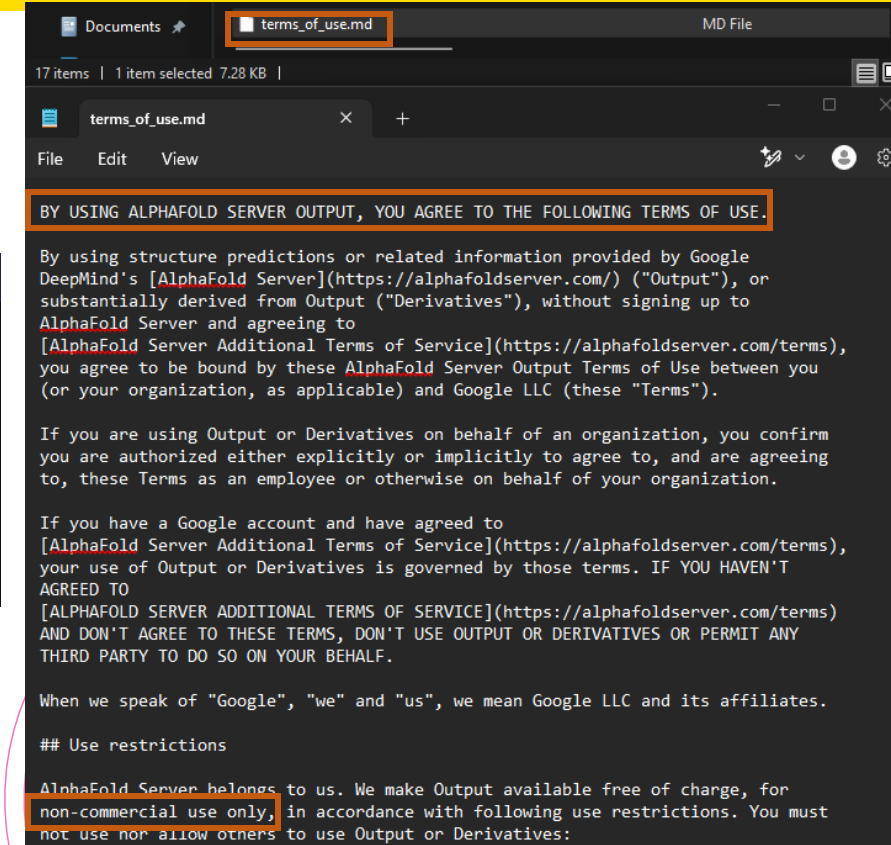
ProteinFold:
(nf-core, CGR, SBF)
nf-co.re/proteinfold

OnDemand Web Form:
(SBF devs)
github.com/Australian-Structural-Biology-Computing/ood-proteinfold

Licensing restrictions

Think before you input!





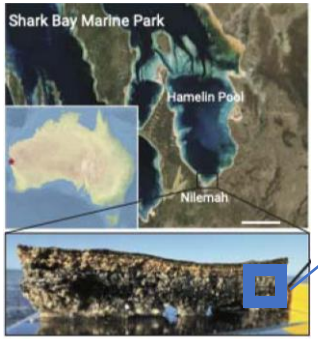
Folding Entire Proteomes

A pipeline of experts

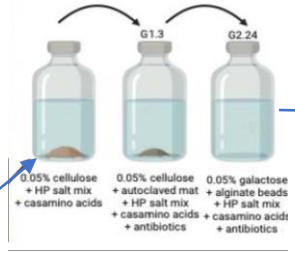
An Asgard archaeon from a modern analog of ancient microbial mats:

Nobs; S., Johnson; M. D., Williams; T. J., Meltzer; J., Vázquez-Campos; X., MacLeod; F. I., Rowell; K., Pitt; M., Paul; B., Shepherd; D. C., Michie; K. A., Duggin; I., G. Ghosal; D., Burns; B. P.

Predicted Proteome: github.com/keiran-rowell-unsw/Loki-ASV2_in_silico



Sampling



Culturing

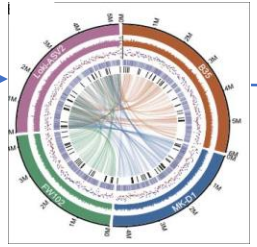
Genome Sequencing

UTS

UNSW Ramaciotti Centre for Genomics

Oxford Nanopore Technologies Gridion

Genome Bioinformatics

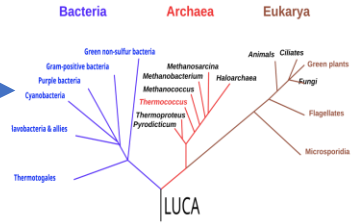


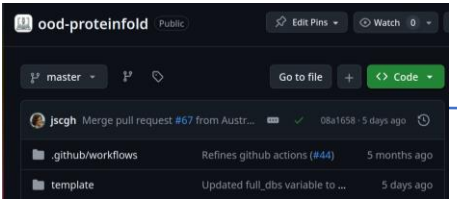
HPC

4544 proteins

Reports

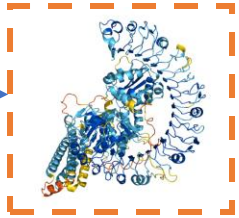
Struct. Bio. Analysis



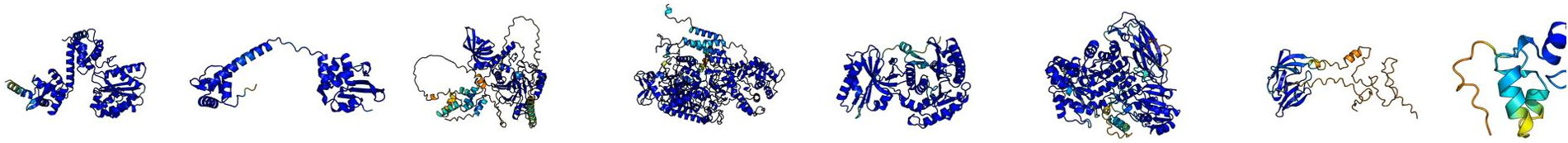


Software Dev.

Struct. Pred.



Struct. Search

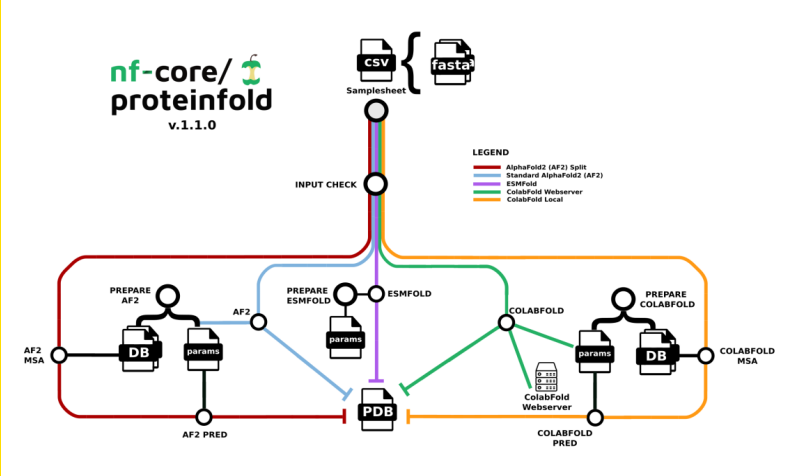


Example set: Fatty acid-binding proteins (FABPs, blood-brain barrier transport)

#	Gene	Tissue	UniProt (sequence)	Protein Data Bank (crystal)	AlphaFoldDB (calculation)
1	FABP1	Liver	P07148	2F73	AF-P07148-F1
2	FABP2	Intestine	P12104	3AKM	AF-P12104-F1
3	FABP3	Muscle (Heart)	P05413	1G5W	AF-P05413-F1
4	FABP4	Adipose	P15090	1TOU	AF-P15090-F1
5	FABP5	Epidermal	Q01469	1B56	AF-Q01469-F1
6	FABP6	Small intestine	P51161	5L8I	AF-P51161-F1
7	FABP7	Brain	O15540	1FDQ	AF-O15540-F1
8	PMP2	PNS	P02689	2WUT	AF-P02689-F1
9	FABP9	-	Q0Z7S8	4A60	AF-Q0Z7S8-F1
-	-	-	-	-	-
11	→ FAB3	Fish (?)	-	-	-
12	FABP12	Retina	A6NFH5	-	AF-A6NFH5-F1

Source: Wikipedia & Scientific databases – Fatty acid binding protein – family members

Live Demo – ProteinFold terminal vs Web



Running the pipeline

The typical commands for running the pipeline on AlphaFold2, Colabfold and ESMFold modes are shown below.

AlphaFold2 regular can be run using this command:

```
nextflow run nf-core/proteinfold \
      --input samplesheet.csv \
      --outdir <OUTDIR> \
      --mode alphafold2 \
      --alphafold2_db <null (default) | DB_PATH> \
      --full_dbs <true/false> \
      --alphafold2_model_preset monomer \
      --use_gpu <true/false> \
      -profile <docker/singularity/.../institute>
```

To run the AlphaFold2 that splits the MSA calculation from the model inference, you can use the `--alphafold2_mode split_msa_prediction` parameter, as shown below:

Proteinfold version: 0.1.6a-1-g77aceb9

Structural Biology Facility

Computationally predict protein structures

Samplesheet

/srv/scratch/USER/test_run_files

Acceptable inputs

- **Directory containing Fasta files:** `/srv/scratch/z3141592/my_experiment`
- **Amino acid sequence:** `NLYIQWLKDGGPSSGRPPPS`

Run Name

test_run

Method

Alphafold2

- **AlphaFold2.3** High Accuracy, Slower - Paper
- **ESMFold** Medium/Low Accuracy, Fastest (No Evolutionary Sequence Calculations) - Paper
- **RoseTTAFold-All-Atom** High Accuracy, Slower; optimised for atomic-level modeling - Paper

Mode

Monomer

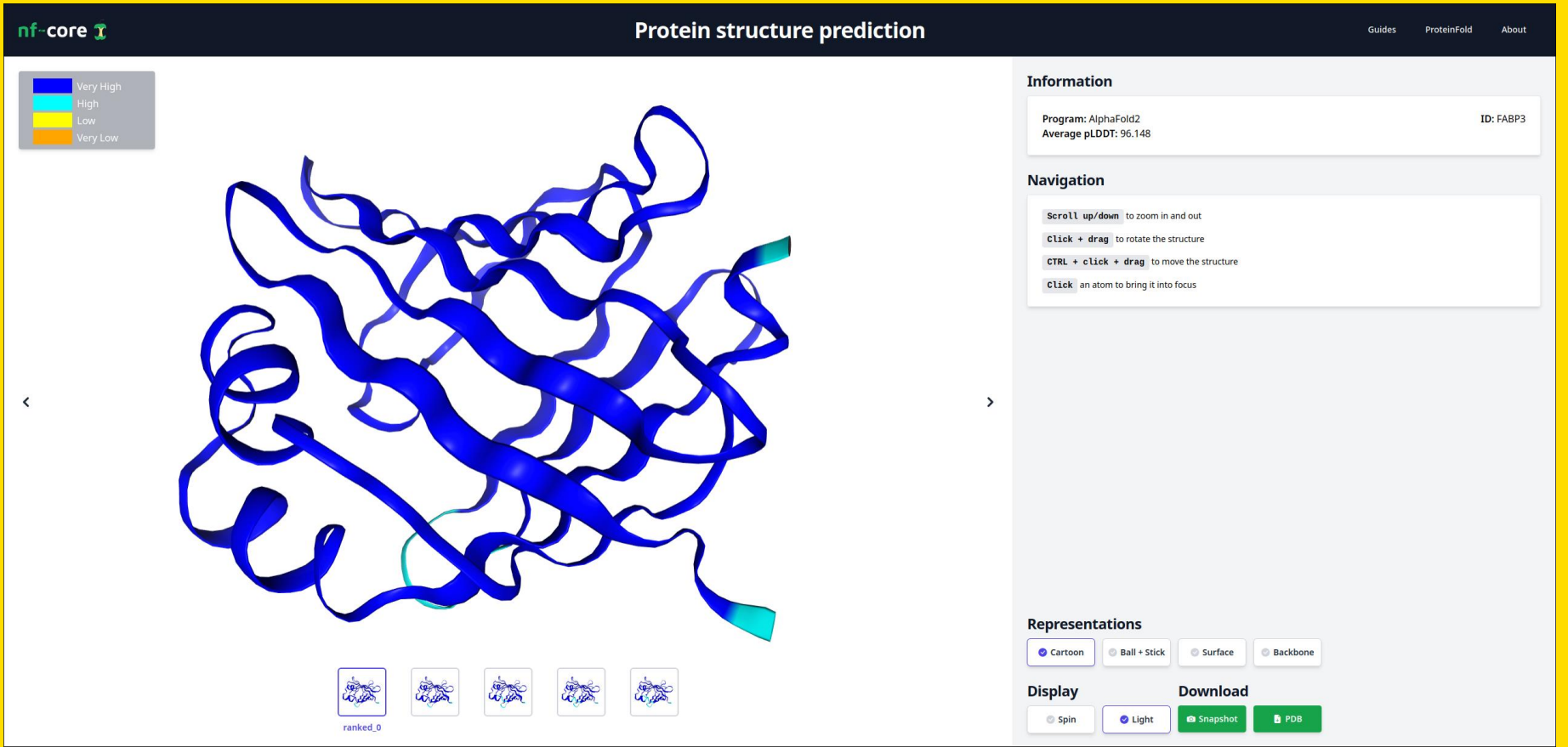
Email Address

k.rowell@unsw.edu.au

Launch

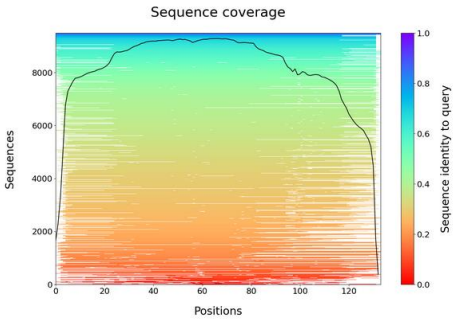
* The Proteinfold session data for this session can be accessed under the data root directory.

Report – the structure

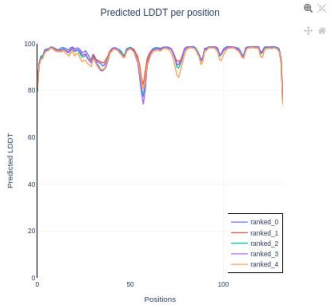


Report – the metrics

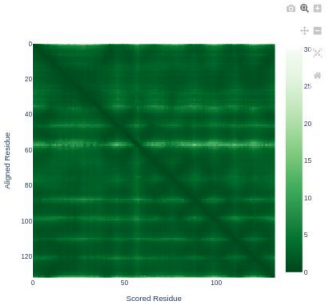
Sequence Coverage – MSA



Residue confidence - pLDDT

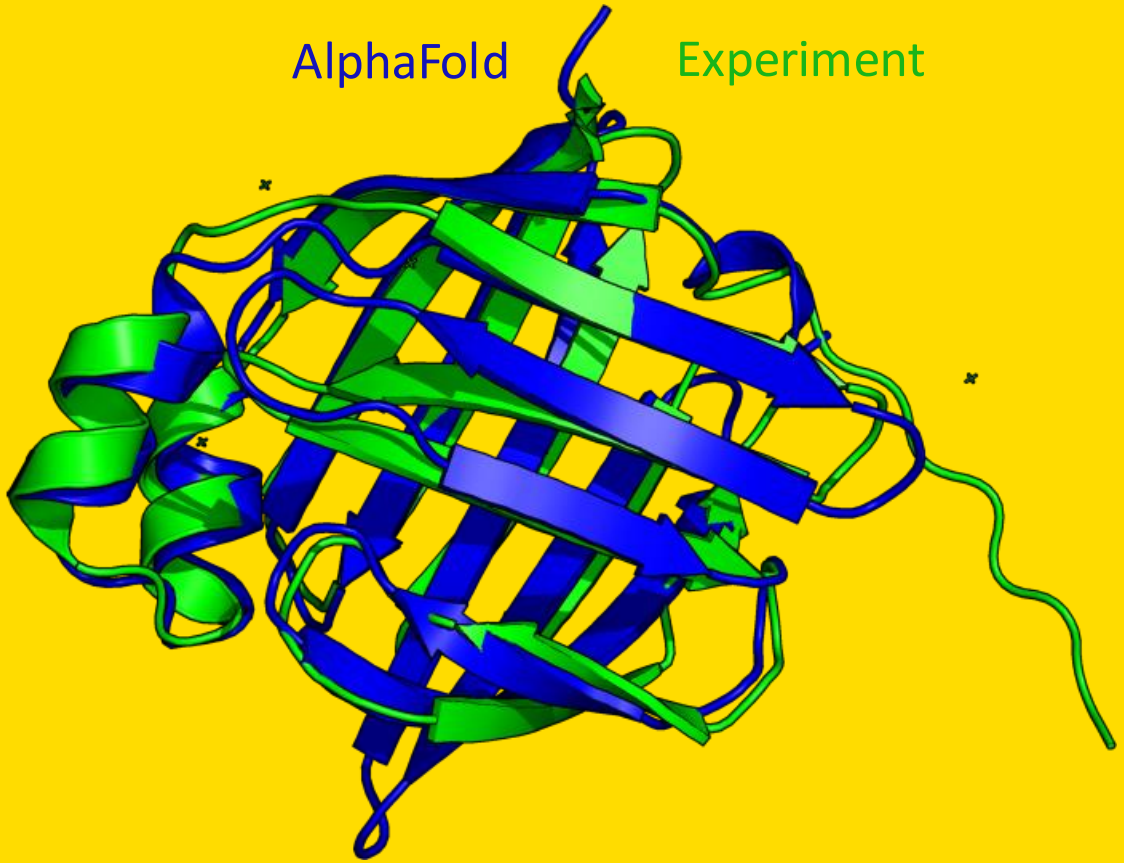


Residue-pair alignment error - PAE

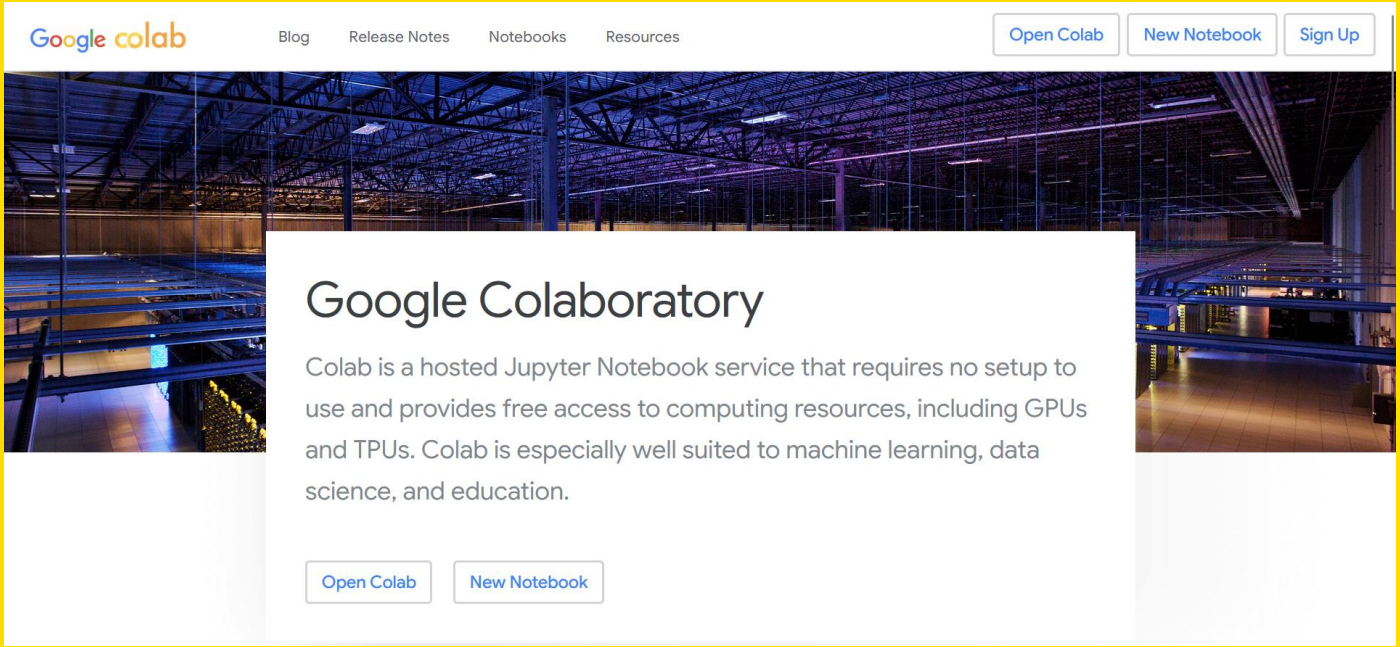


Comparison – prediction vs crystallised

AlphaFold Experiment



ColabFold – running AlphaFold2 on Colab



unsw.to/colabfold

Self-learning – spend 3 hrs with the pros!

https://www.ebi.ac.uk/training/online/courses/alphafold/

ONLINE TUTORIAL

AlphaFold

A practical guide

Enter course

Mark as favourite

Time to complete:
3 hours

This course includes:
- Activities
- Quizzes
- Videos

Written by:
Paulyna Gabriela Magana Gomez
Oleg Kovalevskiy

Last reviewed:
December 2024

Proteins are essential components of life, predicting their 3D structure enables researchers to get an insight into its function and role. AlphaFold is an artificial intelligence (AI) system, developed by Google DeepMind, that predicts a protein's 3D structure based on its primary amino acid sequence. It regularly achieves accuracy competitive with experiment.

Course overview | Course contents | Getting started | Competencies

Feedback and help

Who is this course for?

This tutorial is aimed at researchers who are interested in using AlphaFold2 to predict protein structures and integrate these predictions into their projects. An undergraduate-level knowledge of protein structure and structural biology would be an advantage.

The content of this course provides an understanding of the fundamental concepts behind AlphaFold2, how users can run protein predictions and how AlphaFold2 has been used to enhance research.

Throughout the course there may be terms used you are unfamiliar with. If so, please review the Glossary of terms of help.

This training module on AlphaFold2 has been developed in collaboration with Google DeepMind.

Google DeepMind