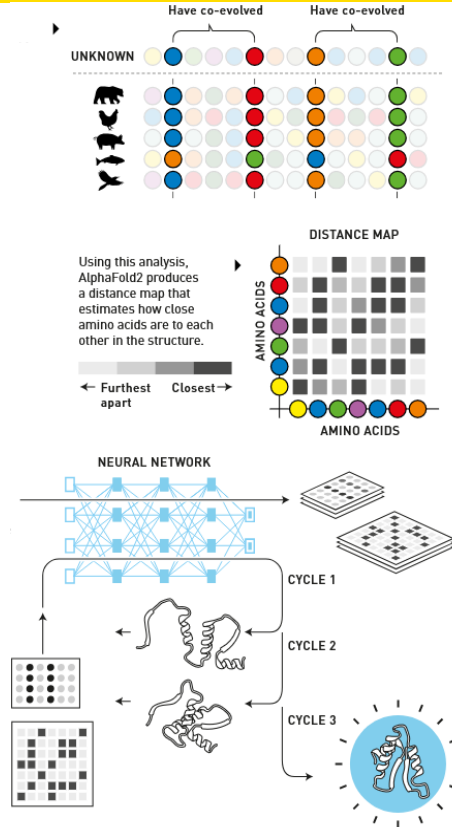
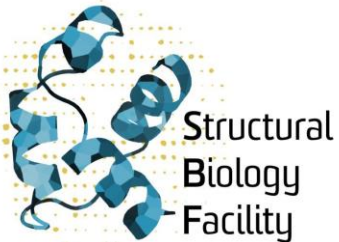
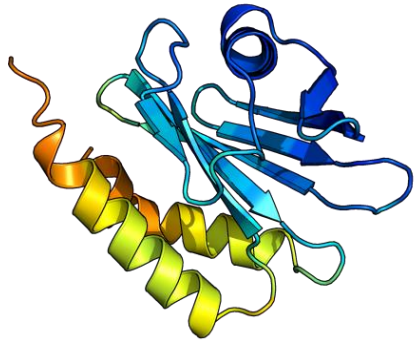


Protein Structure Prediction

Source: <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>



Illustrations: Niklas Elmehed

THE NOBEL PRIZE IN CHEMISTRY 2024



David
Baker

"for computational
protein design"

Demis
Hassabis

"for protein structure prediction"

John M.
Jumper

THE ROYAL SWEDISH ACADEMY OF SCIENCES

Tutorial structure

Q&A (a few mins) between each section (yellow slides)

How did we get here? – 30 mins

- A quick [history](#) of protein structure folding – how did we get to [AlphaFold](#)?
- Looking [under the hood](#) – roughly how does AlphaFold [work](#)
- An explosion of [code choice](#) – awareness of 'right tool' for the [job](#)

Biotech applications, and access – 30 mins

- Folding [all proteins](#) in an organism
- Resolving [previously unsolved crystal structures](#)
- Building [previously unseen giant assemblies](#)
- Access methods - [licence restrictions](#) on code

Break! – 10 mins

Do it yourself – 1 hr

- Running [AlphaFold2](#) in a [Google Colab notebook](#)

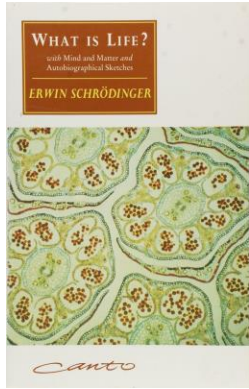
Live demo – 30 mins

- [ProteinFold](#) – X'Fold' on a high performance computer terminal
- On-Demand ProteinFold – automated webform through UNSW's [Katana OnDemand](#)

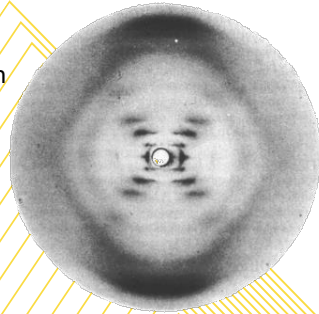
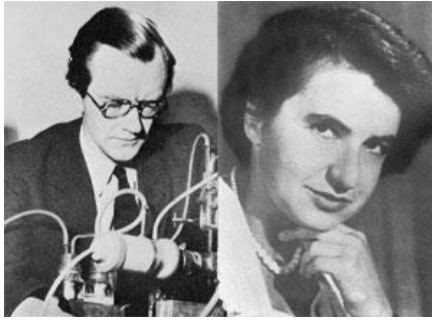


A brief history..

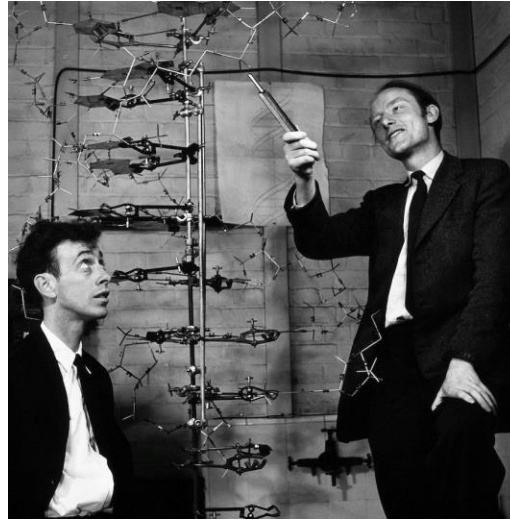
Models in Chemical Biology - function follows chemical form



Schrödinger speculates on
the molecules of life
1944



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



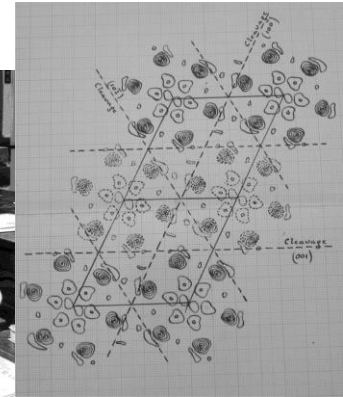
Watson & Crick's DNA double-
helix model 1953



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

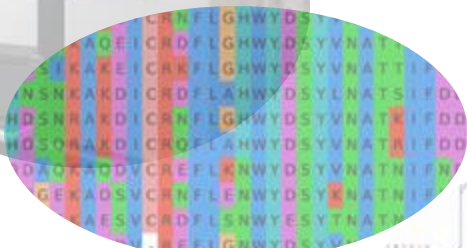


Kathleen Lonsdale resolves the
structure of benzene 1924

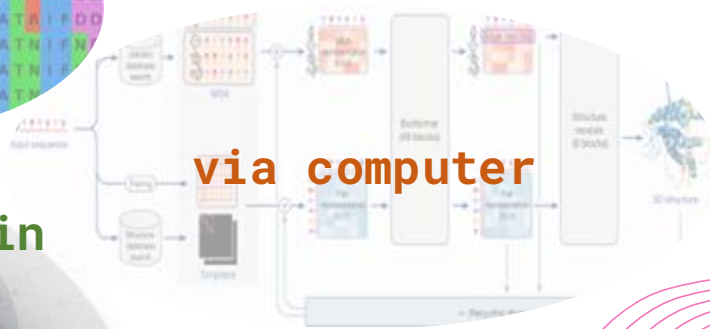


Deep Learning trained on experimental data

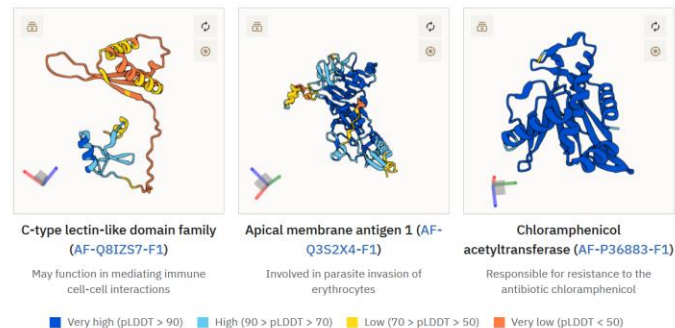
From genetic
sequence



From protein
structure

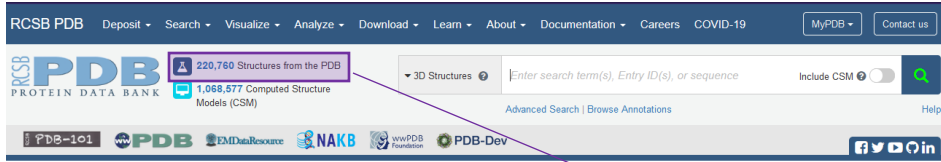


To modelled
structure prediction

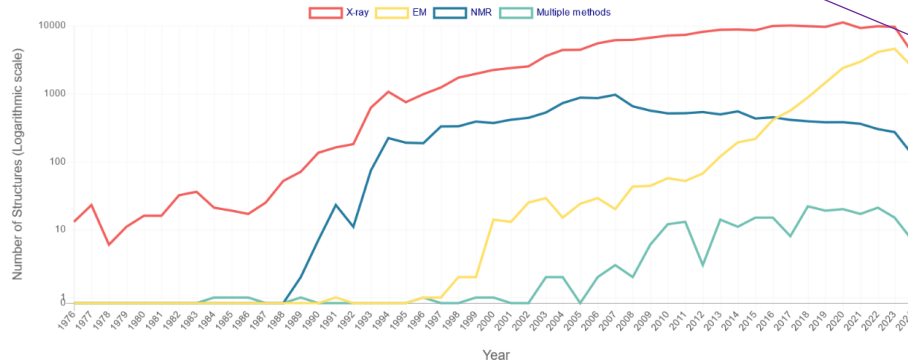


Source: <https://www.ebi.ac.uk/training/online/courses/alphafold/inputs-and-outputs/evaluating-alphafolds-predicted-structures-using-confidence-scores/plddt-understanding-local-confidence/>

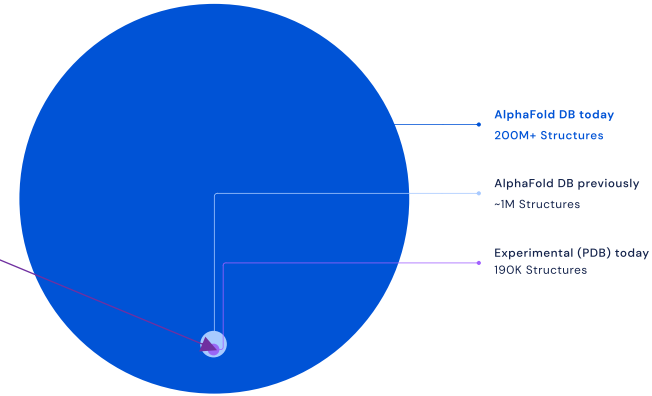
X-rays ruled 20thC – CryoEM & Calcs the 21st



Number of Released PDB Structures per Year



Number of Protein Structures



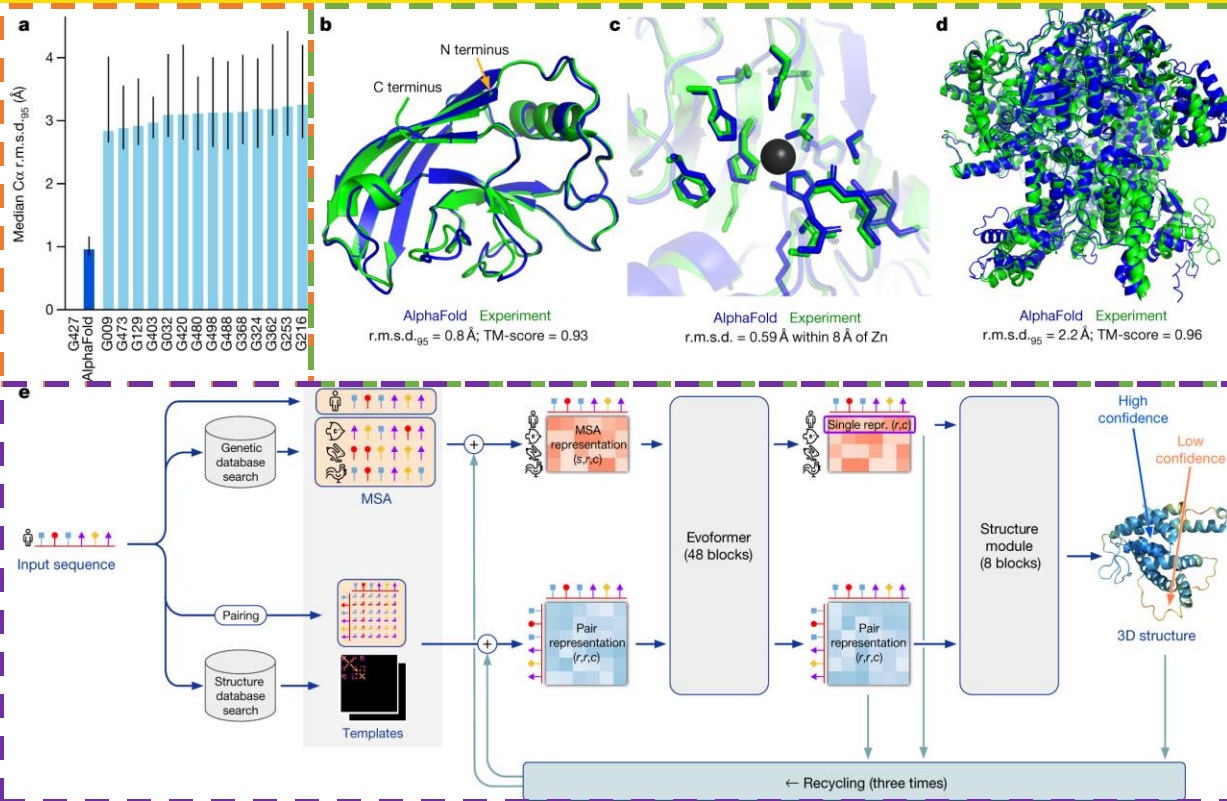
Protein Data Bank (PDB)

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



Under the hood..

X'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)

0000

PDB ID: 7RDT

0000

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/
s41592-024-02772-z](https://doi.org/10.1038/s41592-024-02772-z)

AlphaFold2 architecture

ingenious **algorithm design** made this breakthrough

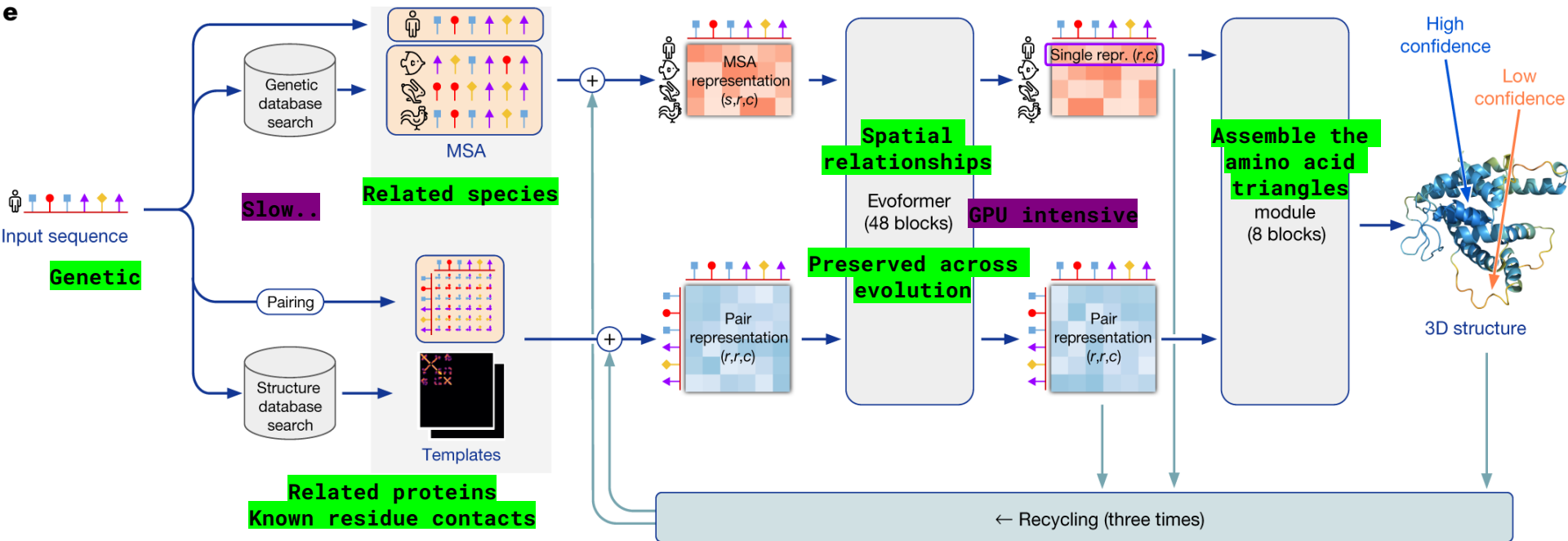
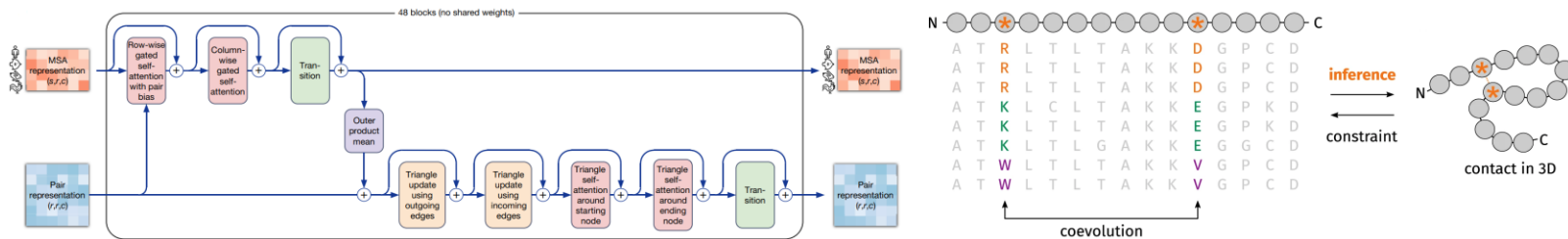


Fig. 1: AlphaFold produces highly accurate structures – from "Highly accurate protein structure prediction with AlphaFold" - Nature, 596, 583-89 (2021).

Code choice..

X'Fold' programs - co-evolution DeepLearning

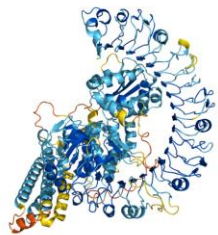
DeepLearning: No folding physics – Transformers – 1D sequence -> 3D structure



High Performance Computing: Not for domestic computers. Use your uni's **HPC** cluster.

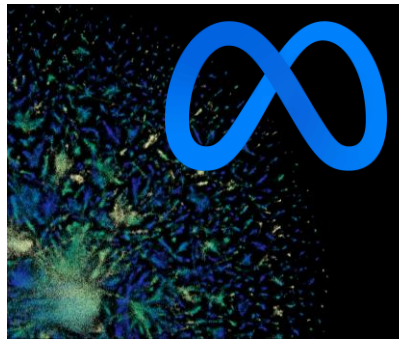
'Evo' formers

(Alpha|Open|RoseTTA|Ab)Fold



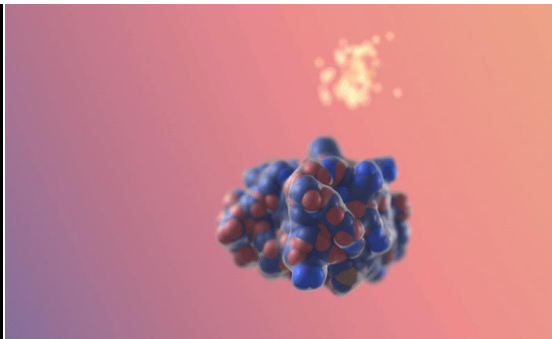
Protein-LLMs

ESMFold, ProtTrans



GenAI - Diffusion

RFDiffusion, Chroma



Tutorial only on **protein folding** - X'Fold' programs

The **largest Big Tech** and **academic groups** in world have dedicated serious expertise and 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.



<https://github.com/google-deepmind/alphafold>

<https://github.com/google-deepmind/alphafold3>

Facebook AI Research

<https://github.com/facebookresearch/esm>



<https://github.com/bytedance/Proteinix>



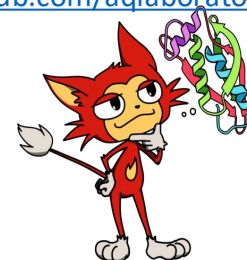
<https://github.com/jwohlwend/boltz>



<https://github.com/baker-laboratory/RoseTTAFold-All-Atom>



<https://github.com/aqlaboratory/openfold>



<https://github.com/sokrypton/ColabFold>

Let's do
Biology!

Focus
on **easiest**
AlphaFold2
method

X'Fold' programs – each has its place

Computing is all about **trade-offs**:

speed ✂ accuracy ✂ cost

Algorithms are also about **hand-offs**:

💾 Storage 🤝 📁 memory 🤝 🏢 Processors 🤝 ⌚ run time

COMPUTERS ARE FAST

Programming is all about **re-use**:

Code is crystallized labour that can be run x**billions**/second

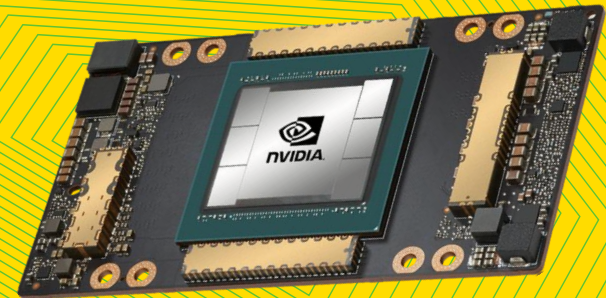
Researchers adapt & script the **last-mile** for **their usage**

The computing stack is also about **access**:

Programmers 🤝 pipetters | customisability ✂ ease-of-use

Data processing ✂ restricted graphical interfaces

Free & Open Source licences ✂ recoup development cost



\$10,000s a pop in the AI arms race
10s/100s/1,000s **structures/day**

Let's do Biology!

Focus on **easiest-to-use**
AlphaFold2 method



UNSW
SYDNEY

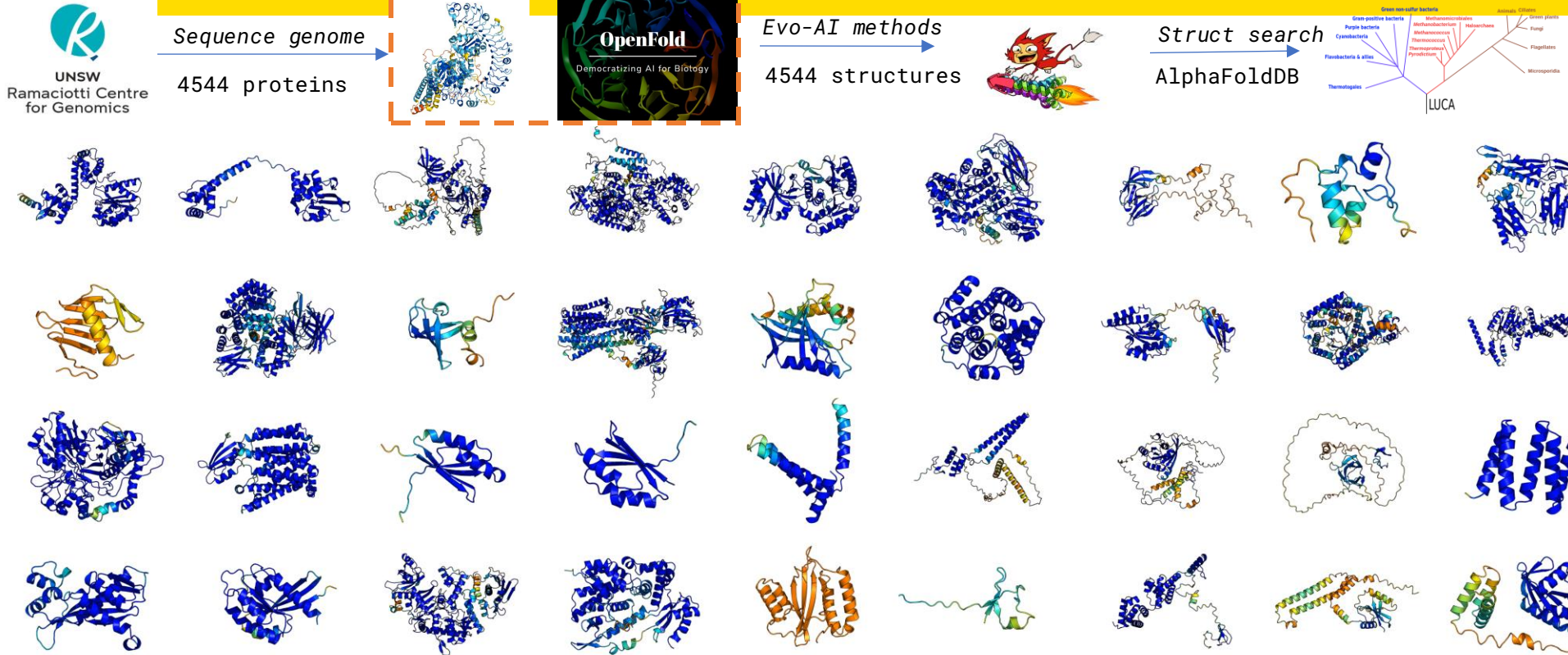


BioTech applications..

Folding Entire Proteomes

Predict a molecular inventory of your organism
(my predictions from bioinformatician's closed genome)

A mix of AlphaFold2, ESMFold and AlphaFold3



Clarifying Experiments

Using computer models on unsolved structure problems



STRUCTURAL
BIOLOGY

ISSN: 2059-7983

Volume 78 | Part 1 | January 2022 | Pages 1-13
<https://doi.org/10.1107/S2059798321012122>

OPEN ACCESS

Cited by 121

Part of a special issue

Part of a special issue

Implications of *AlphaFold2* for crystallographic phasing by molecular replacement

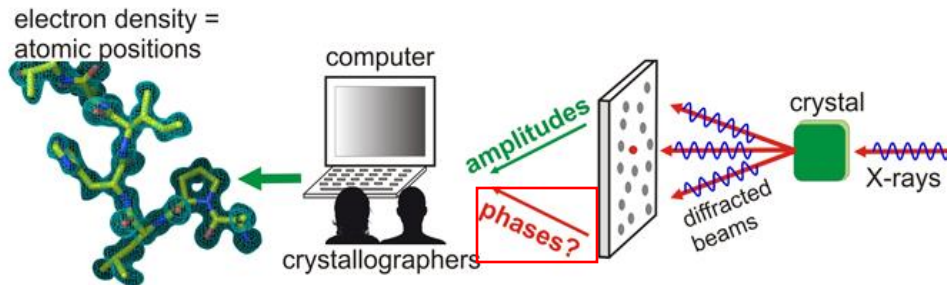
Airlie J. McCoy,^{a*} Massimo D. Sammito^a and Randy J. Read^a

^aDepartment of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY, United Kingdom

*Correspondence e-mail: ajm201@cam.ac.uk

Edited by D. J. Rigden, University of Liverpool, United Kingdom (Received 18 May 2021; accepted 13 November 2021)

The *AlphaFold2* results in the 14th edition of Critical Assessment of Structure Prediction (CASP14) showed that accurate (low root-mean-square deviation) *in silico* models of protein structure domains are on the horizon, whether or not the protein is related to known structures through high-coverage sequence similarity. As highly accurate models become available, generated by harnessing the power of correlated mutations and deep learning, one of the aspects of structural biology to be impacted will be methods of phasing in crystallography. Here, the data from CASP14 are used to explore the prospects for changes in phasing methods, and in particular to explore the prospects for molecular-replacement phasing using *in silico* models.



“Impossible” X-ray microscope

Source: [LibreTexts Chemistry](https://www.libretexts.org/Bookshelves/Chemistry/Book%3A_Structural_Resolution/1.7%3A_Structural_Resolution)

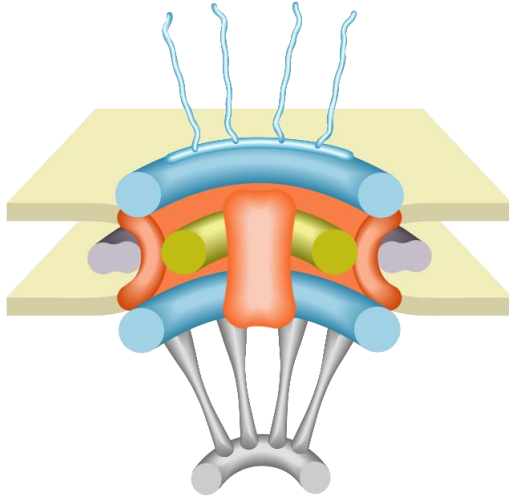
1.7: Structural Resolution

(accessed March 2025)

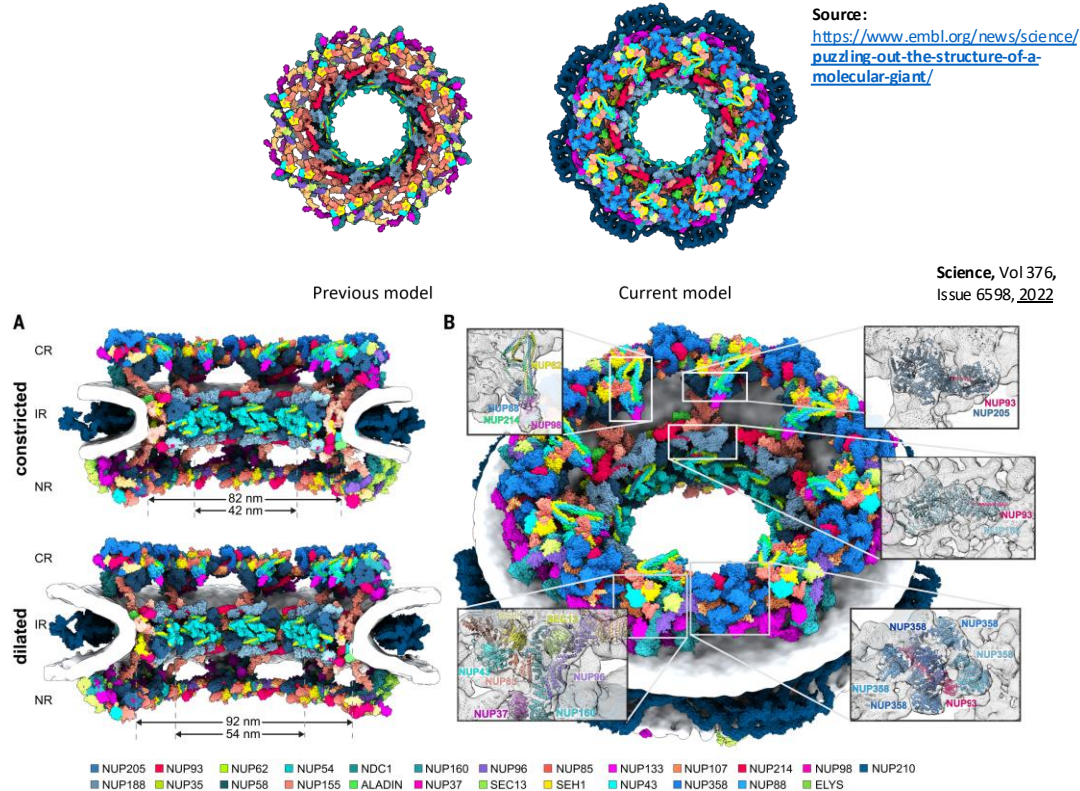
Assembly Building from AlphaFold + Electron Microscopy

Building the "monstorous maw"

Nuclear Pore Complex



Transport of large
molecules to cell



Accessing software.. and licence restrictions...

Where to run? Galaxy, Uni compute cluster, proteinfold

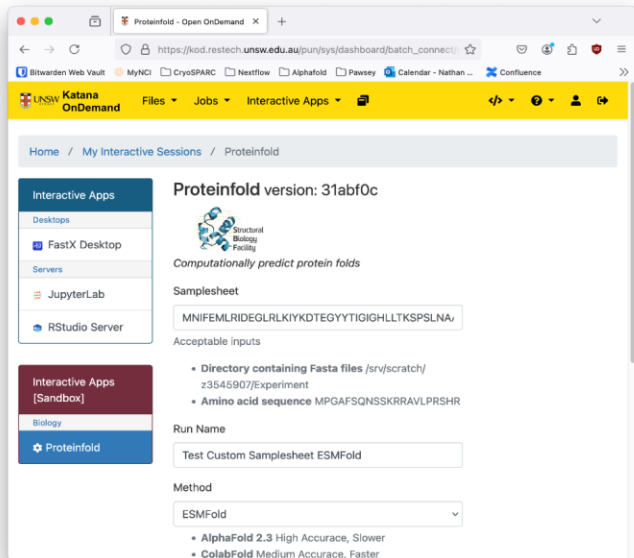


ABOUT ACTIVITIES SERVICES TRAINING & EVENTS DOMAINS NEWS CONTACT HELP

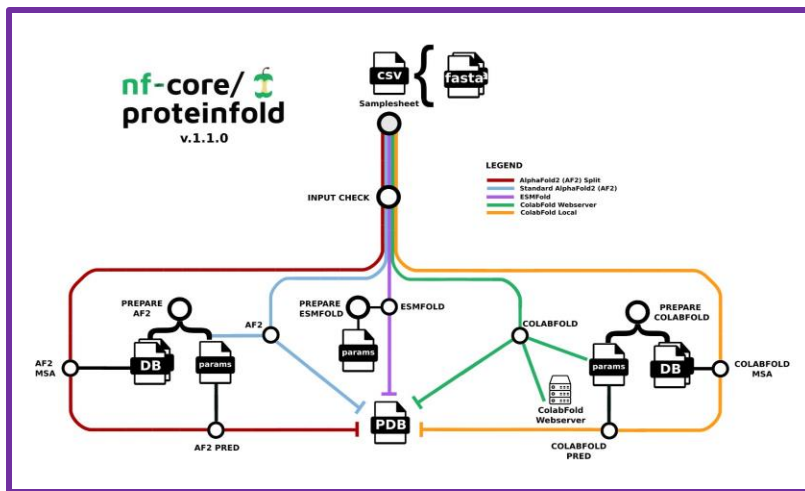


Australian AlphaFold Service

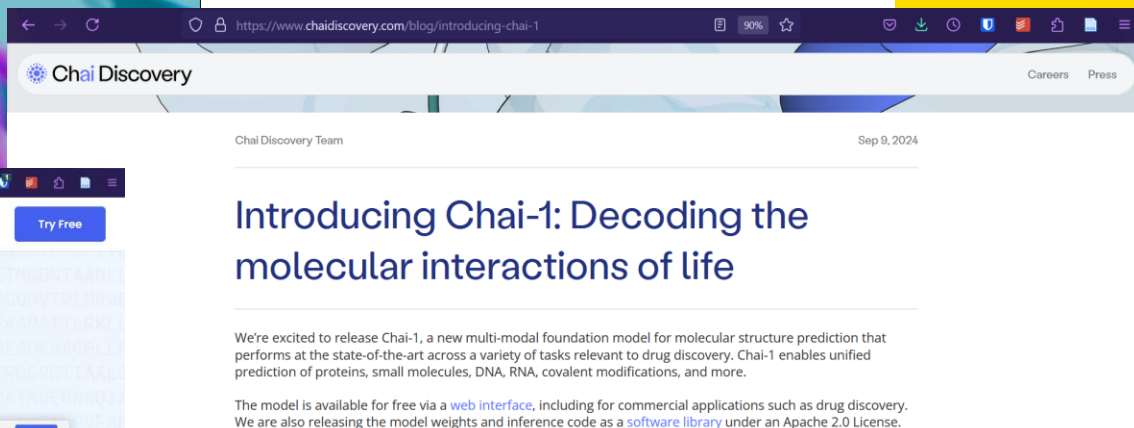
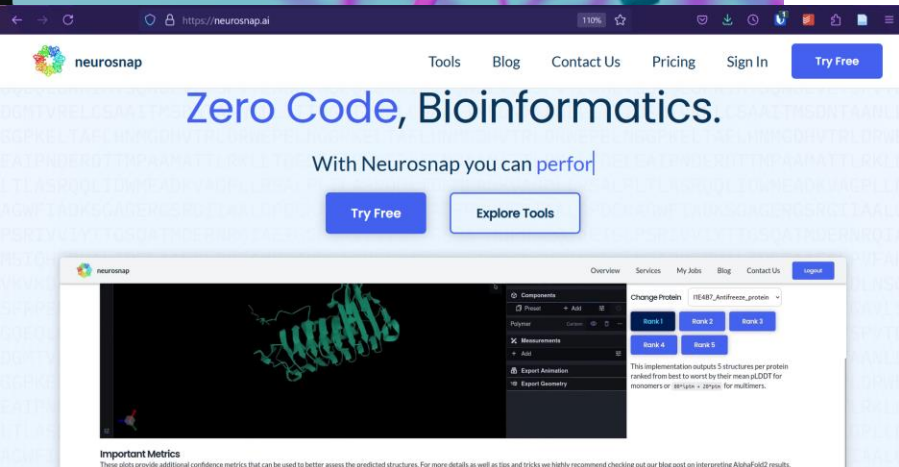
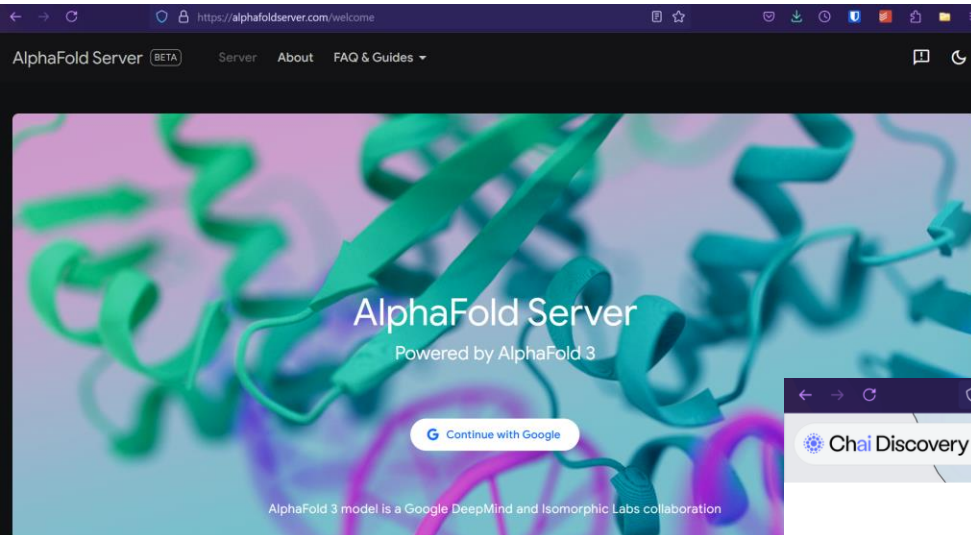
AlphaFold is an artificial intelligence (AI) system developed by [DeepMind](#) that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy that is competitive with experimental methods (see [Jumper et al. Nature 2021](#)).



The screenshot shows the Proteinfold web interface. The top navigation bar includes links for Files, Jobs, Interactive Apps, and a user profile. The main content area is titled "Proteinfold version: 31abf0c" and "Computationally predict protein folds". It features a "Samplesheet" section with a text input field containing the sequence "MNIFEMLRIDGLRLKIYKDTEGYTIGIHLTKSPSLNA". Below this, there are "Acceptable inputs" listed: "Directory containing Fasta files /srv/scratch/z3545907/Experiment" and "Amino acid sequence MPGAFSQNSKRRVLRSHR". The "Run Name" field is set to "Test Custom Samplesheet ESMFold". The "Method" dropdown is set to "ESMFold". A sidebar on the left lists "Interactive Apps" including Desktop, FastX Desktop, Servers, JupyterLab, and RStudio Server. The "Proteinfold" app is highlighted.



Dedicated web forms - portals built by companies



Licensing restrictions

Think before you input!

https://alphafoldserver.com/welcome

BETA Server About FAQ & Guides ▾

Terms of use and attribution

AlphaFold Server is for **non-commercial use only**, subject to [AlphaFold Server Terms of Service](#). AlphaFold Server output cannot be used **in docking or screening tools** or to train machine learning models or related technology for biomolecular structure prediction similar to AlphaFold Server.

If you use an AlphaFold Server prediction, please cite our paper: Abramson, J et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature (2024).

Documents terms_of_use.md MD File

17 items | 1 item selected 7.28 KB |

terms_of_use.md

File Edit View

BY USING ALPHAFOLD SERVER OUTPUT, YOU AGREE TO THE FOLLOWING TERMS OF USE.

By using structure predictions or related information provided by Google DeepMind's [\[AlphaFold Server\]\(https://alphafoldserver.com/\)](#) ("Output"), or substantially derived from Output ("Derivatives"), without signing up to [AlphaFold Server](#) and agreeing to [\[AlphaFold Server Additional Terms of Service\]\(https://alphafoldserver.com/terms\)](#), you agree to be bound by these [AlphaFold Server Output Terms of Use](#) between you (or your organization, as applicable) and Google LLC (these "Terms").

If you are using Output or Derivatives on behalf of an organization, you confirm you are authorized either explicitly or implicitly to agree to, and are agreeing to, these Terms as an employee or otherwise on behalf of your organization.

If you have a Google account and have agreed to [\[AlphaFold Server Additional Terms of Service\]\(https://alphafoldserver.com/terms\)](#), your use of Output or Derivatives is governed by those terms. IF YOU HAVEN'T AGREED TO [\[ALPHAFOLD SERVER ADDITIONAL TERMS OF SERVICE\]\(https://alphafoldserver.com/terms\)](#) AND DON'T AGREE TO THESE TERMS, DON'T USE OUTPUT OR DERIVATIVES OR PERMIT ANY THIRD PARTY TO DO SO ON YOUR BEHALF.

When we speak of "Google", "we" and "us", we mean Google LLC and its affiliates.

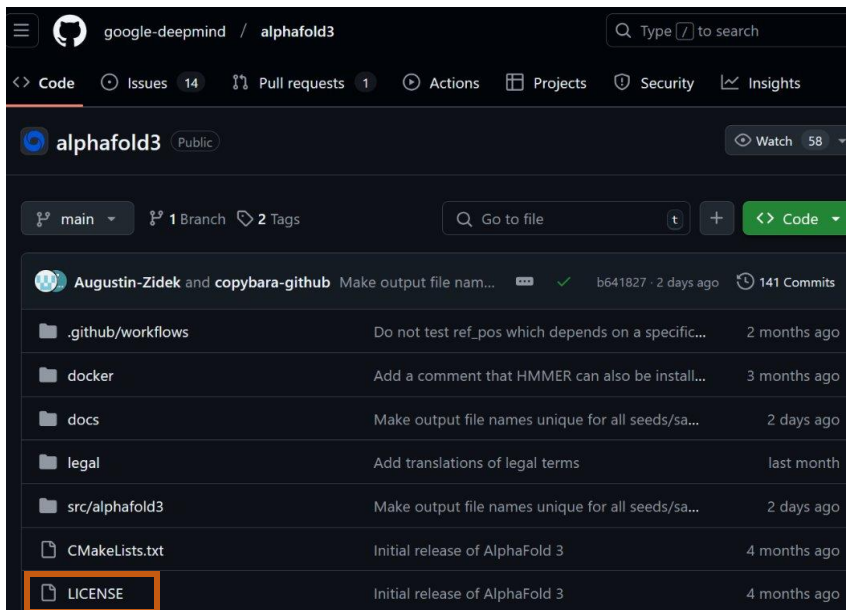
Use restrictions

AlphaFold Server belongs to us. We make Output available free of charge, for **non-commercial use only**, in accordance with following use restrictions. You must not use nor allow others to use Output or Derivatives:

Licensing restrictions

find the LICENSE file summary

Restricted



google-deepmind / alphafold3

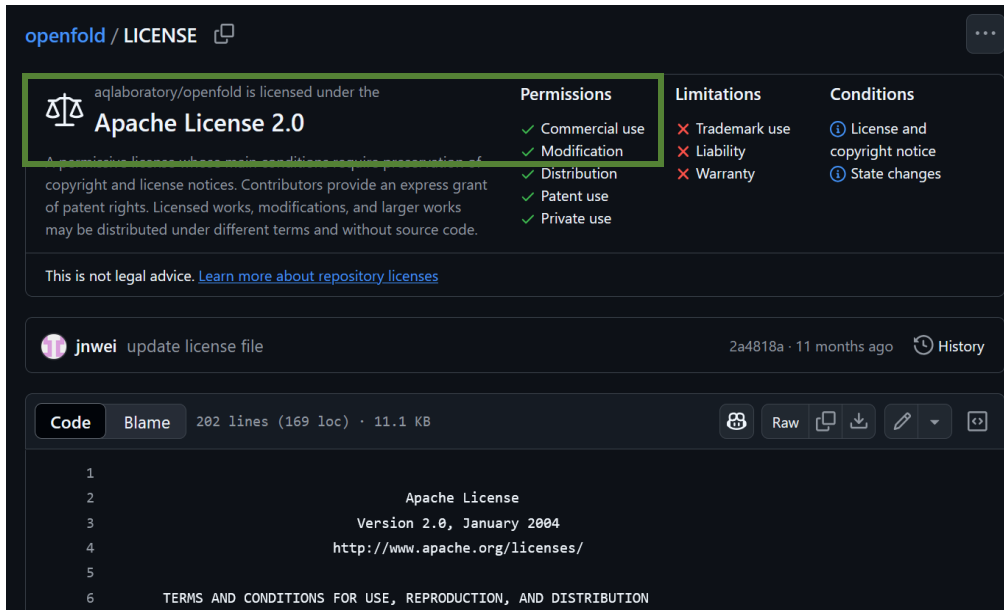
Code Issues 14 Pull requests 1 Actions Projects Security Insights

alphafold3 Public Watch 58

main 1 Branch 2 Tags Go to file

File	Description	Updated
.github/workflows	Do not test ref_pos which depends on a specific...	2 months ago
docker	Add a comment that HMMER can also be install...	3 months ago
docs	Make output file names unique for all seeds/sa...	2 days ago
legal	Add translations of legal terms	last month
src/alphafold3	Make output file names unique for all seeds/sa...	2 days ago
CMakeLists.txt	Initial release of AlphaFold 3	4 months ago
LICENSE	Initial release of AlphaFold 3	4 months ago

Unrestricted



openfold / LICENSE

aqalaboratory/openfold is licensed under the **Apache License 2.0**

A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

This is not legal advice. [Learn more about repository licenses](#)

Permissions	Limitations	Conditions
<ul style="list-style-type: none"> ✓ Commercial use ✓ Modification ✓ Distribution ✓ Patent use ✓ Private use 	<ul style="list-style-type: none"> ✗ Trademark use ✗ Liability ✗ Warranty 	<ul style="list-style-type: none"> ① License and copyright notice ① State changes

jnwei update license file 2a4818a · 11 months ago History

Code Blame 202 lines (169 loc) · 11.1 KB

```

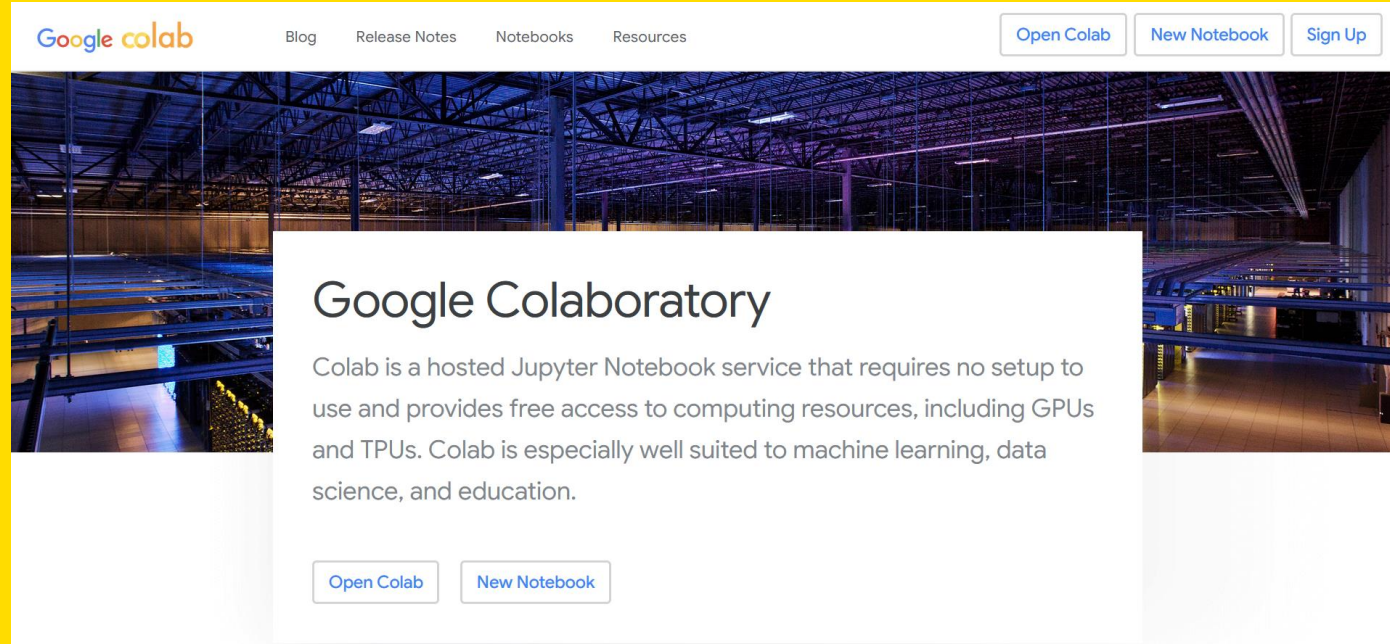
1
2           Apache License
3           Version 2.0, January 2004
4           http://www.apache.org/licenses/
5
6  TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION
  
```

Break!



Get your laptops ready to run your own structure prediction :)

ColabFold – running AlphaFold2 on Colab



The screenshot shows the Google Colaboratory website. At the top, there is a navigation bar with the Google Colab logo and links for Blog, Release Notes, Notebooks, and Resources. On the right side of the navigation bar are three buttons: 'Open Colab', 'New Notebook', and 'Sign Up'. The main content area features a large background image of a server room. Overlaid on this image is a white box containing the text 'Google Colaboratory' in a large, bold font. Below this, a paragraph describes Colab as a hosted Jupyter Notebook service that requires no setup and provides free access to computing resources, including GPUs and TPUs. At the bottom of this white box are two buttons: 'Open Colab' and 'New Notebook'.

Google colab

Blog Release Notes Notebooks Resources

Open Colab New Notebook Sign Up

Google Colaboratory

Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. Colab is especially well suited to machine learning, data science, and education.

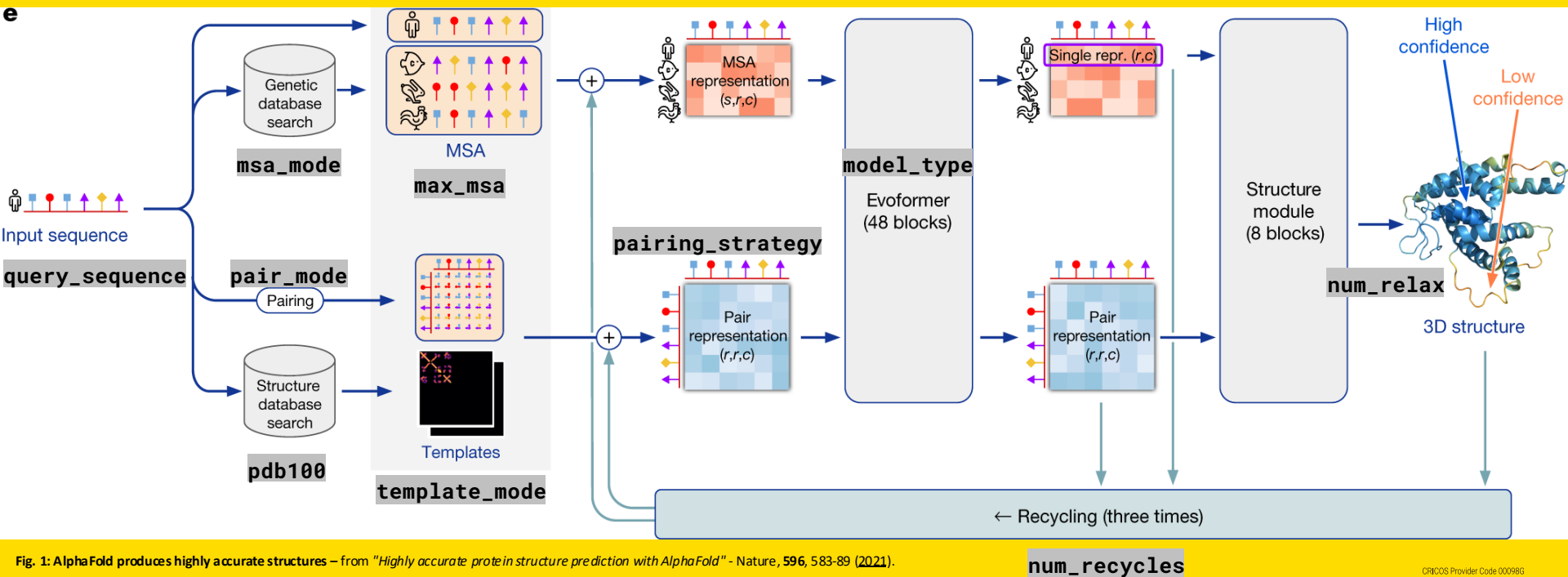
Open Colab New Notebook

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

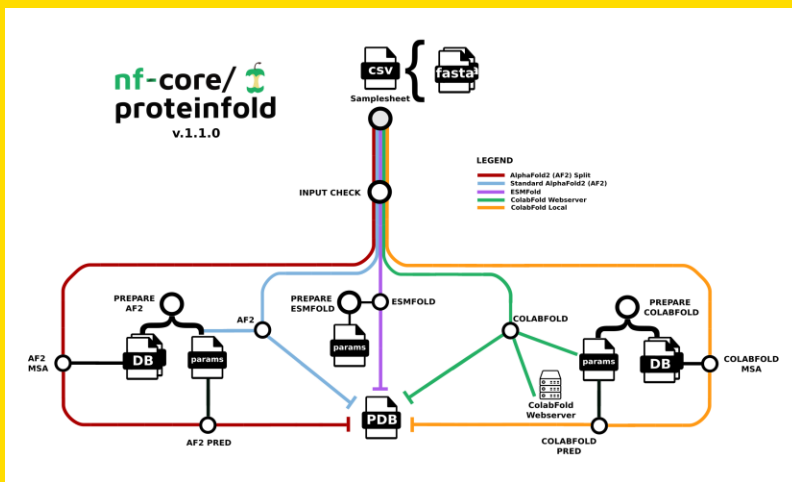
ColabFold – running AlphaFold2 on Colab

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

Architecture compared to `input_options` (mostly default)



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



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: NLVYIQILKDGPPSSGRPPPS

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](#)
 - **HelixFold3** High Accuracy, Slower (Restricted AlphaFold3 method) - [Pre-print](#)
- WARNING: Non-commercial use only!**

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](#).


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

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.

