中国科学院分子细胞科学卓越创新中心
（生物化学与细胞生物学研究所）
Center for Excellence in Molecular Cell Science, CAS

AlphaFold初级介绍

张少庆

2022/4/3

Credit to online resources.

Mastermind

- **Demis Hassabis** CBE FRS FREng FRSA (1976-).
- Became master of chess at 13.
- Self-taught programming.
- Designed video games.
- B.A. from Cambridge in computer science at 21.
- Game designer again.
- Research on hippocampal amnesia ranked top 10 breakthroughs by the Science magazine (at 31).
- Ph.D. from UCL in cognitive neuroscience at 33.
- Co-founded DeepMind at 34.

Credit: Royal Academy of Engineering

DeepMind

- Founded by Hassabis, Shane Legg & Mustafa Suleyman in 2010.
- Firstly artificial intelligence technology on how to play old games.
- **To create a general-purpose AI useful and effective for almost anything.**
- Invested by many people, including Elon Musk.
- Acquired by Google for £400m in 2014.
- Products: **AlphaGo**, AlphaZero, AlphaStar, **AlphaFold** & AlphaCode.
- Technologies: new AI algorithms, text-to-speech system, etc.

ALPHAGO

A DOCUMENTARY • SPRING 2017

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

Nature, 28 January 2016

Citation: 12573

Mastering the game of Go without human knowledge

David Silver[1]*, Julian Schrittwieser[1]*, Karen Simonyan[1]*, Ioannis Antonoglou[1], Aja Huang[1], Arthur Guez[1], Thomas Hubert[1], Lucas Baker[1], Matthew Lai[1], Adrian Bolton[1], Yutian Chen[1], Timothy Lillicrap[1], Fan Hui[1], Laurent Sifre[1], George van den Driessche[1], Thore Graepel[1] & Demis Hassabis[1]

Nature, 19 October 2017

Citation: 6973

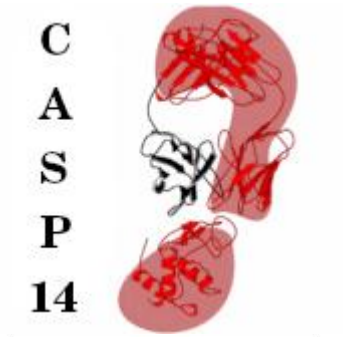
AlphaFold Lead

- **John M. Jumper** (198?-).
- B.S. in physics & math at Vanderbilt.
- Self-taught programming.
- Studied condensed matter physics at Cambridge (1y).
- Worked on protein folding simulation at D. E. Shaw.
- Obtained Ph.D. (2017) with Karl Freed and Tobin Sosnick at Uchicago by predicting folding dynamics.
- Joined DeepMind (2017) and started working on AlphaFold.

Credit: Vanderbilt Univ.

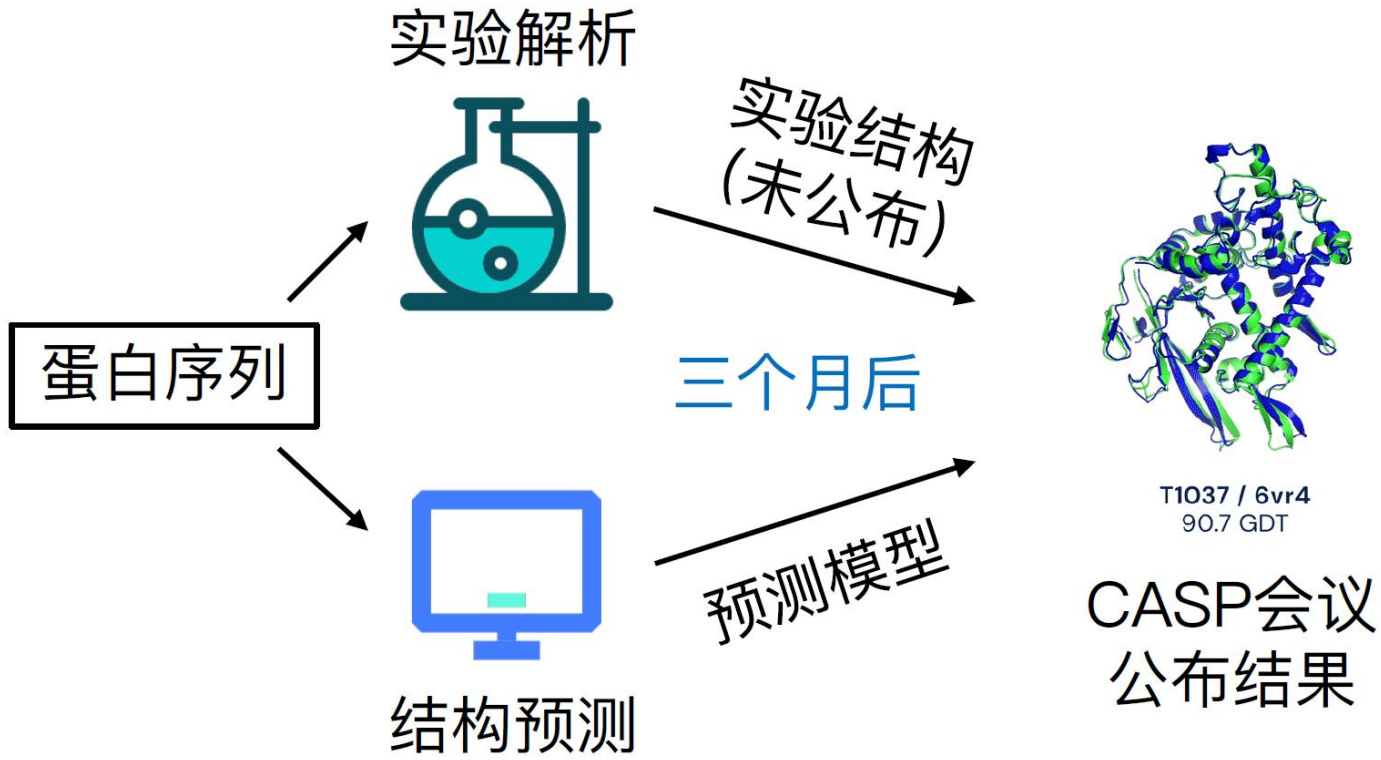
Jumper's publications before joining DeepMind

- Joshua A Riback, Micayla A Bowman, Adam M Zmyslowski, Catherine R Knoverek, John M Jumper, James R Hinshaw, Emily B Kaye, Karl F Freed, Patricia L Clark, Tobin R Sosnick, **Science** 358 (6360), 238-241 (2017).
- JM Jumper, KF Freed, TR Sosnick, **bioRxiv**:169326 (2017).
- JM Jumper, KF Freed, TR Sosnick, **arXiv**:1610.07277 (2016).
- JF Zimmerman, GF Murray, Y Wang, JM Jumper, JR Austin, B Tian, **Nano letters** 15 (8), 5492-5498 (2015).
- MC Baxa, EJ Haddadian, JM Jumper, KF Freed, TR Sosnick, **Proceedings of the National Academy of Sciences** 111 (43), 15396-15401 (2014).
- MP Eastwood, T Chitra, JM Jumper, K Palmo, AC Pan, DE Shaw, **The Journal of Physical Chemistry B** 117 (42), 12898-12907 (2013).
- Yibing Shan, Michael P Eastwood, Xuewu Zhang, Eric T Kim, Anton Arkhipov, Ron O Dror, John Jumper, John Kuriyan, David E Shaw, **Cell** 149 (4), 860-870 (2012).
- David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, Willy Wriggers, **Science** 330 (6002), 341-346.



CASP

Critical Assessment of protein Structure Prediction, 1994



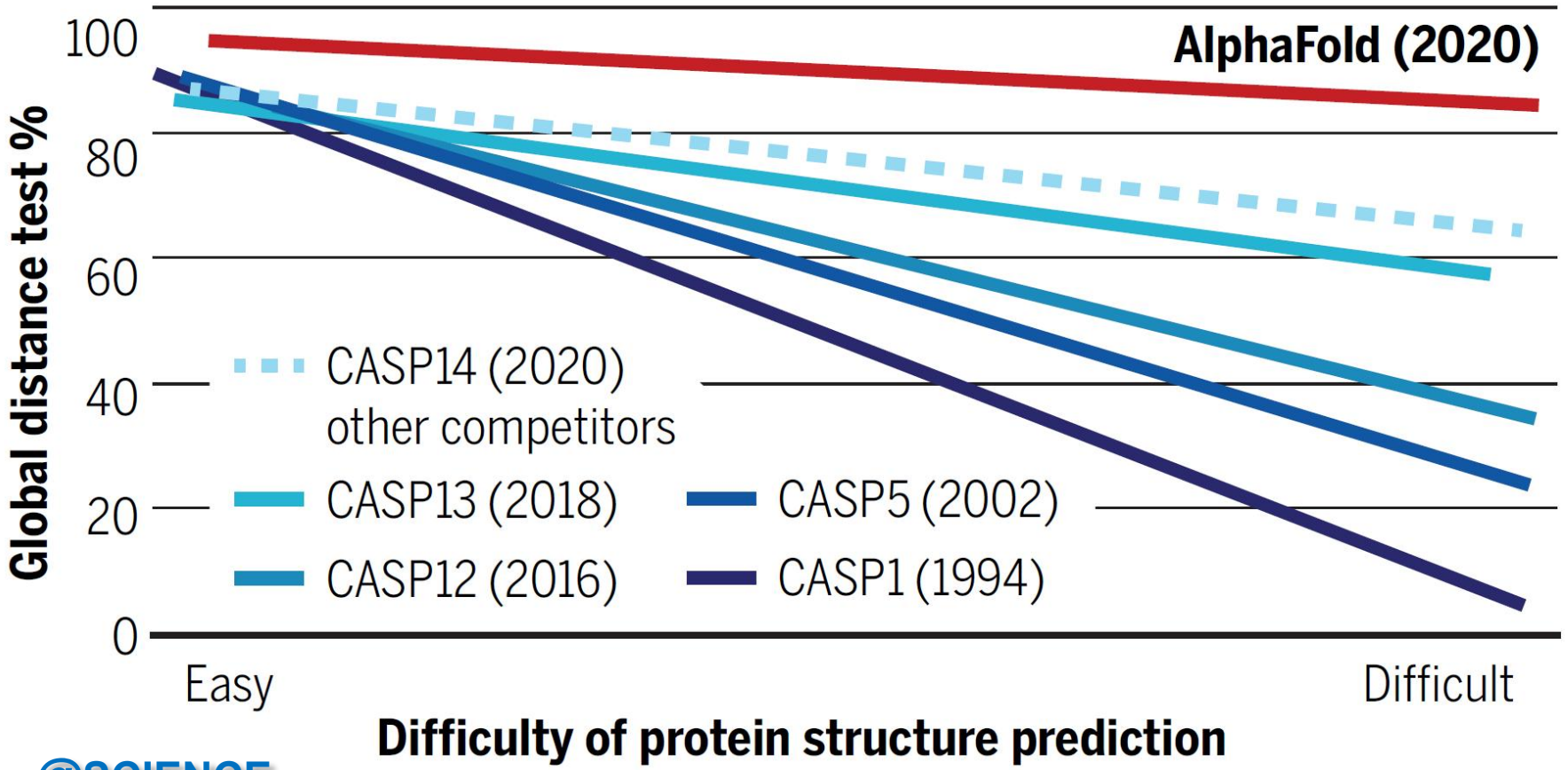
评估计算机预测蛋白质结构方法的准确度

Critical Assessment of protein Structure Prediction (CASP)

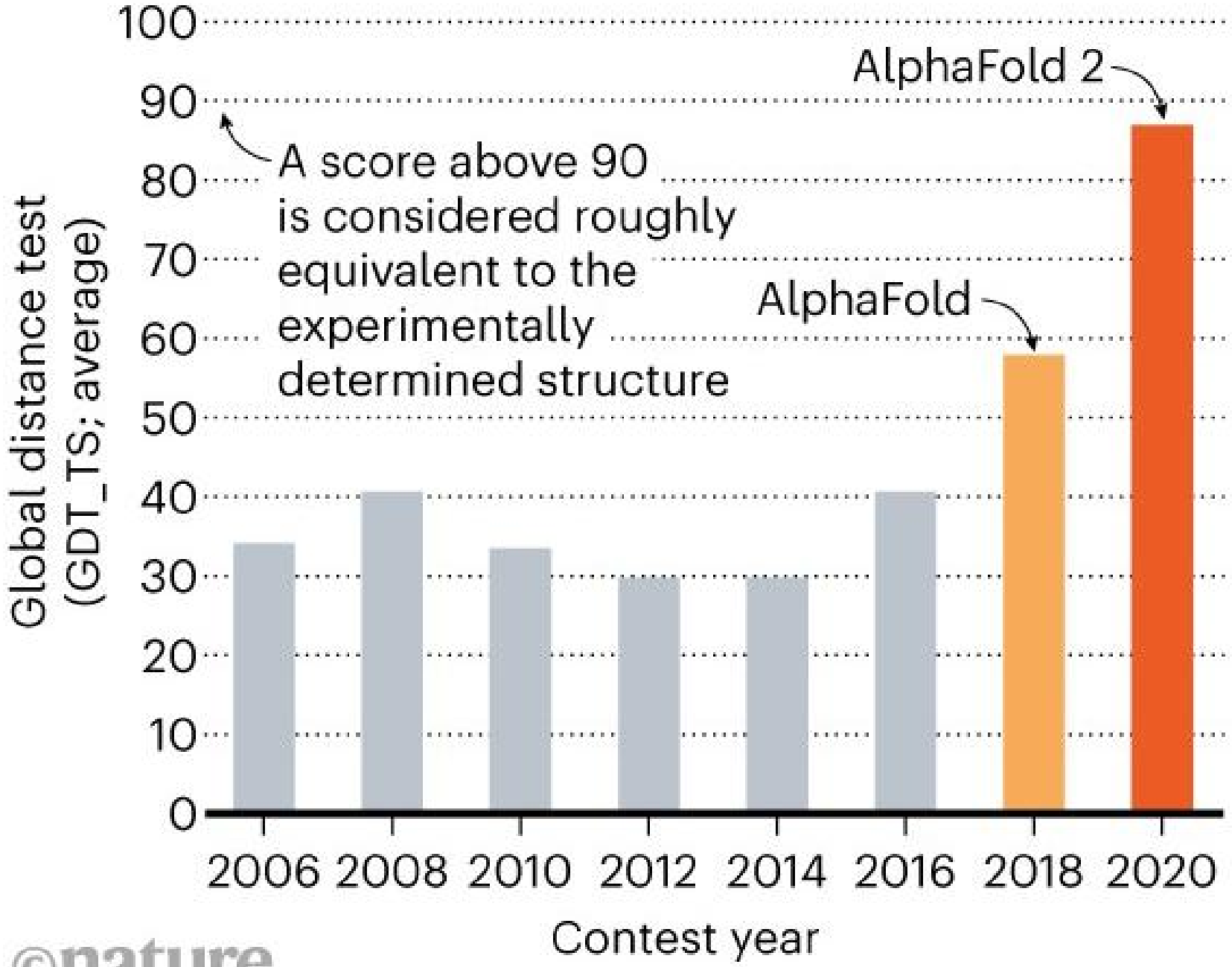
- An international competition on protein structure prediction biennially since 1994
- CASP14 takes place in 2020
- More than 200 groups participated in CASP14, including labs from academia and industry
- Baidu, Tencent and Huawei also participated
- The ultimate winner is the group AlphaFold2 from Google's AI department DeepMind

Getting real

At the Critical Assessment of Protein Structure Prediction (CASP) competition, AlphaFold matched experimental findings on a measure of accuracy.

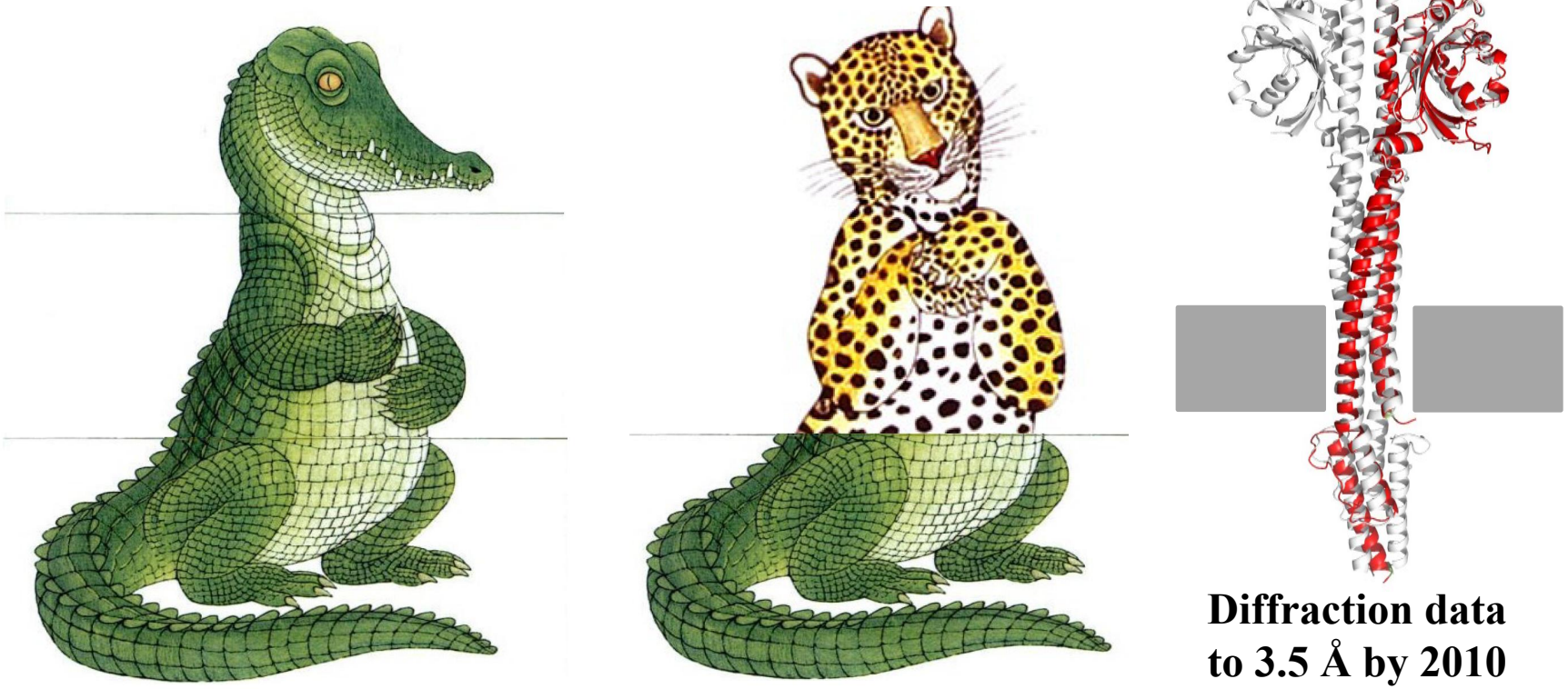


@SCIENCE


100
90
80
70
60
50
40
30
20
10
0
Global distance test
(GDT_TS; average)
A score above 90
is considered roughly
equivalent to the
experimentally
determined structure
AlphaFold 2
AlphaFold
2006
2008
2010
2012
2014
2016
2018
2020
Contest year

Membrane protein structures are correctly predicted AlphaFold2!

Functional histidine kinase chimera

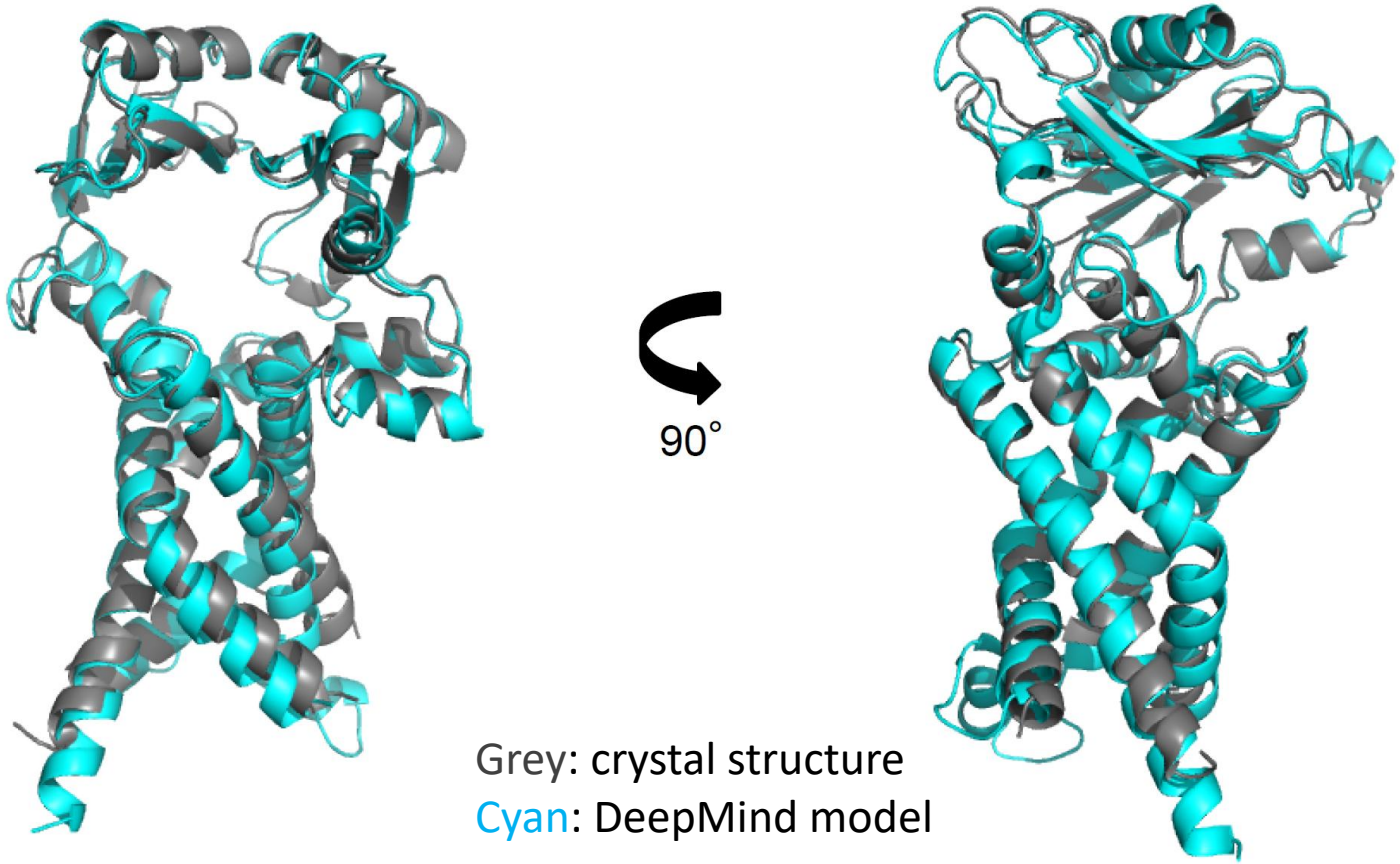


Diffraction data to 3.5 Å by 2010

by **Andrei N. Lupas** @Max Planck Institute

Another membrane protein structure correctly predicted by AlphaFold2

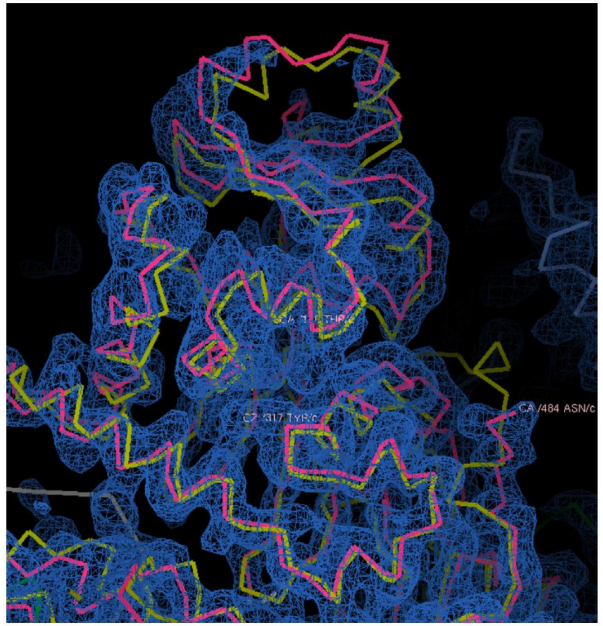
Inner membrane reductase FoxB



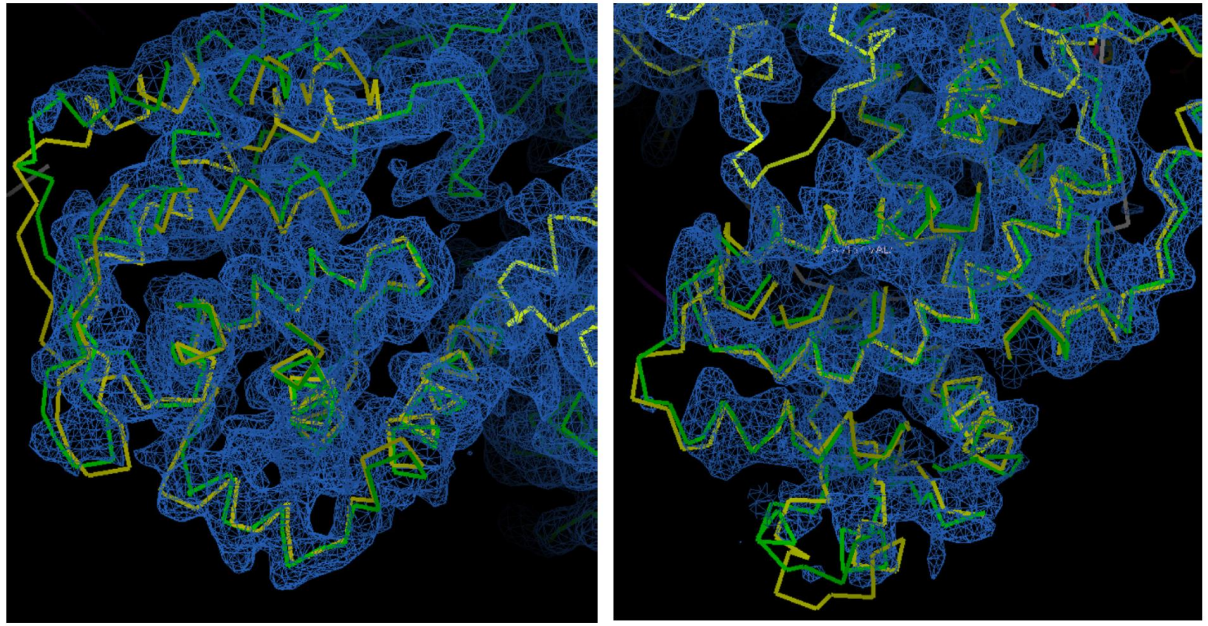
Grey: crystal structure
Cyan: DeepMind model

DeepMind models used for cryoEM

gp105



gp226



DeepMind model (yellow) is fitted into the cryoEM map.

DeepMind solves 50-year-old ‘grand challenge’ with protein folding A.I.

PUBLISHED MON, NOV 30 2020•10:30 AM EST | UPDATED MON, NOV 30 2020•10:07 PM EST

12-14-20

DeepMind’s latest AI breakthrough could turbocharge drug discovery

Creating new drugs is notoriously slow and expensive. AlphaFold’s ability to understand proteins could help researchers move much faster.

The AlphaFold2 Method Paper: A Fount of Good Ideas

-**Mohammed AlQuraishi**, Columbia University
https://moalquraishi.wordpress.com/2021/07/25/the-alphafold2-method-paper-a-fount-of-good-ideas/

"What is impressive about the AF2 effort is not top-notch hardware, but top-notch software and ML engineering that renders brute-force scaling unnecessary. **This part I suspect will be hardest for academia to replicate**, **because it is less about increased national investment in computing resources** (which would undoubtedly help) **and more about in-house professional software engineering capacity, a much taller order.**

… on the whole, their culture appears to be one of **computational frugality**, in the best sense of the word, and quite possibly **the biggest compliment** I can give the AF2 team on an achievement that is remarkable in so many other ways."

AlphaFold

Improved protein structure prediction using potentials from deep learning

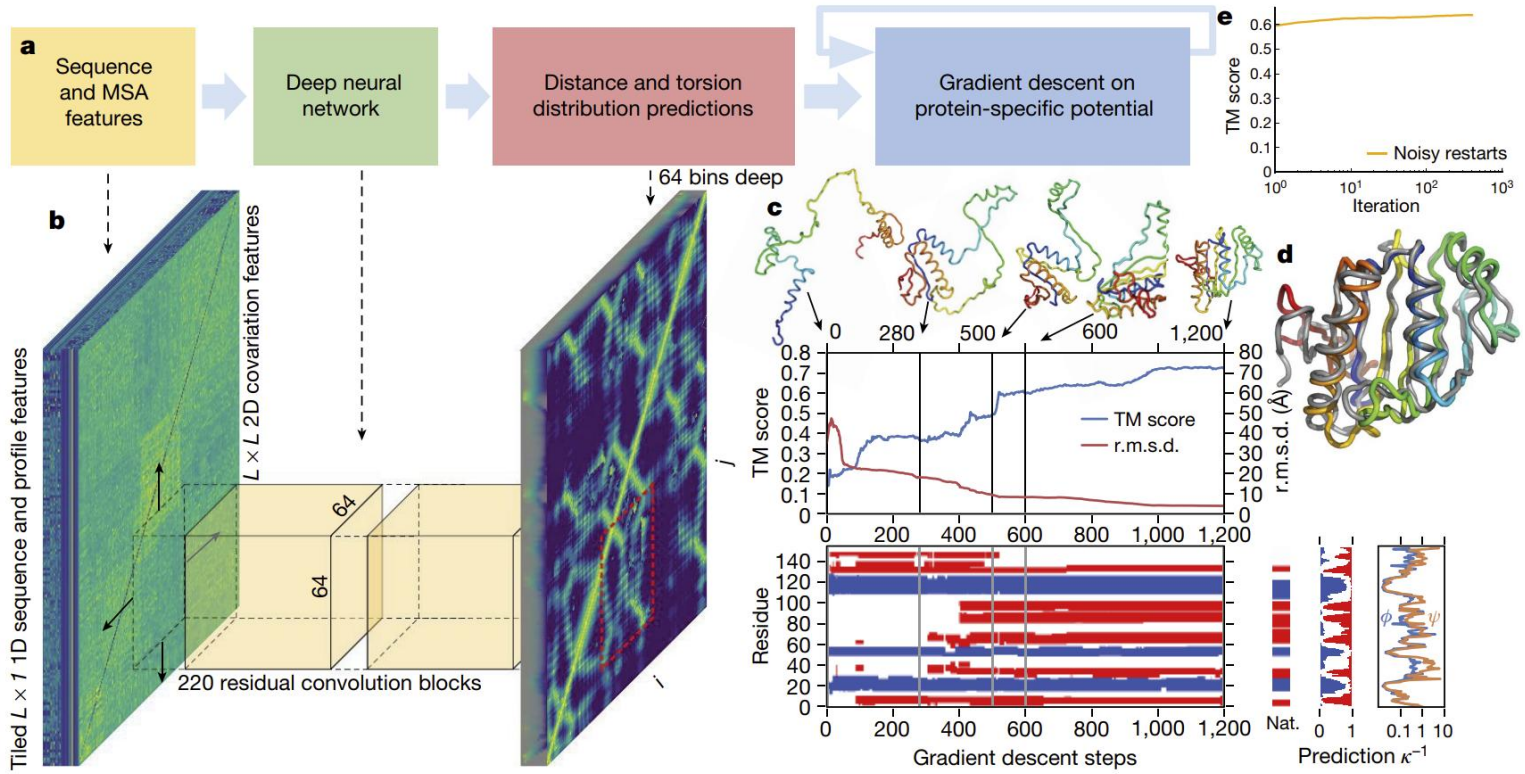
https://doi.org/10.1038/s41586-019-1923-7

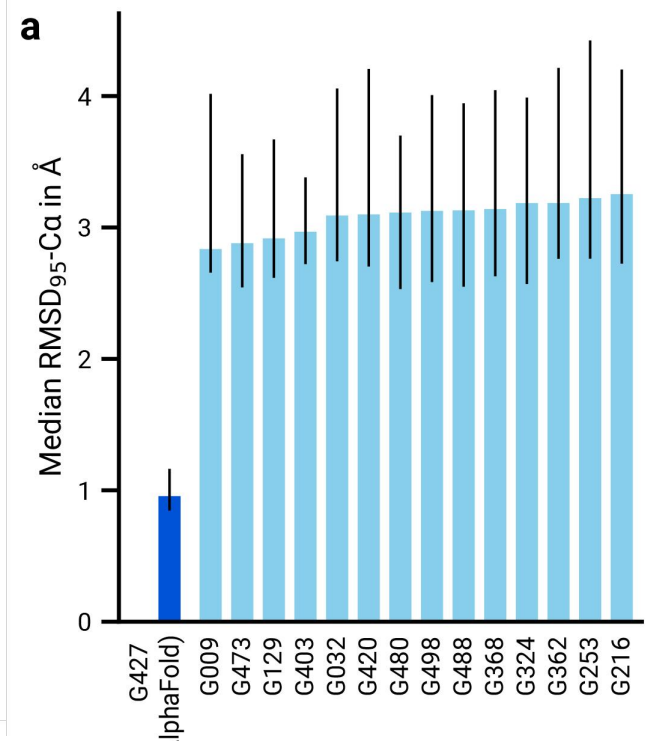
Received: 2 April 2019

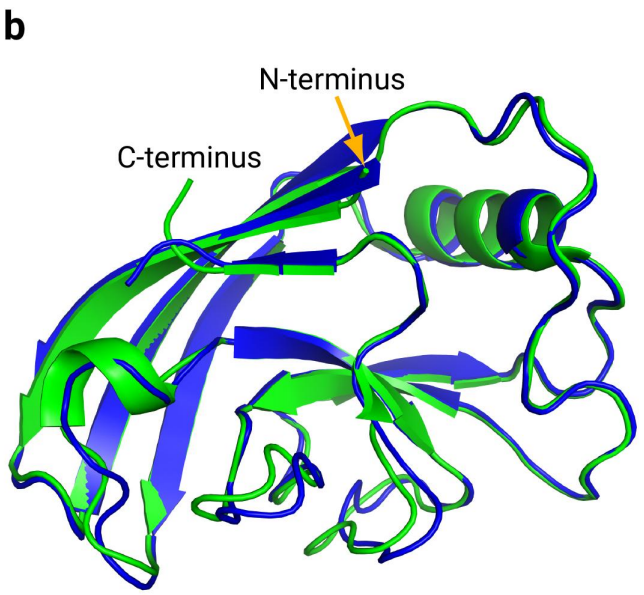
Accepted: 10 December 2019

Published online: 15 January 2020

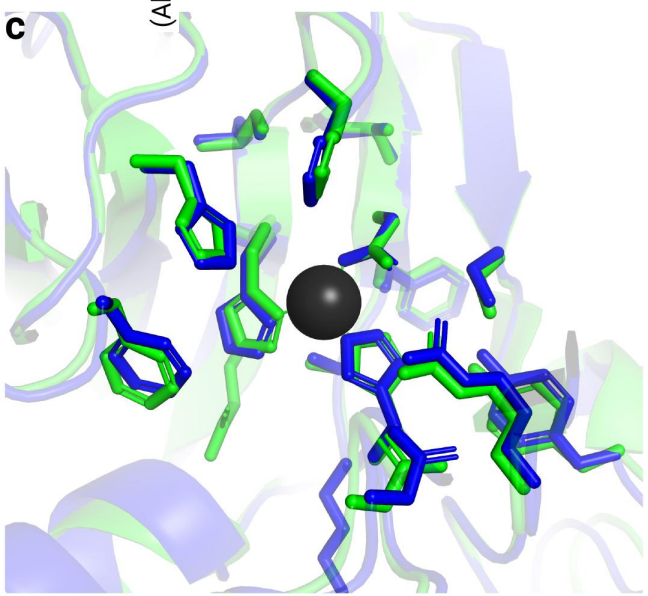
Andrew W. Senior[1,4]*, Richard Evans[1,4], John Jumper[1,4], James Kirkpatrick[1,4], Laurent Sifre[1,4], Tim Green[1], Chongli Qin[1], Augustin Žídek[1], Alexander W. R. Nelson[1], Alex Bridgland[1], Hugo Penedones[1], Stig Petersen[1], Karen Simonyan[1], Steve Crossan[1], Pushmeet Kohli[1], David T. Jones[2,3], David Silver[1], Koray Kavukcuoglu[1] & Demis Hassabis[1]



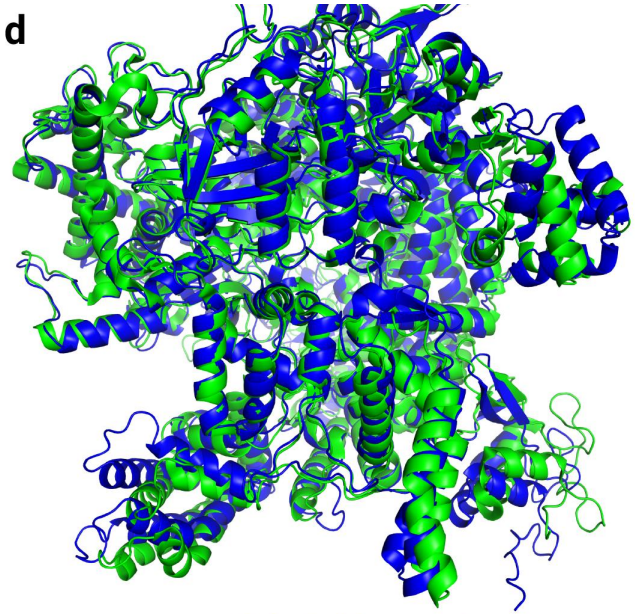




T1049 – AlphaFold / experiment
$RMSD_{95}$: 0.8 Å, TM-score: 0.93



T1056 (6yj1) – AlphaFold / experiment
RMSD: 0.59 Å within 8 Å of Zn

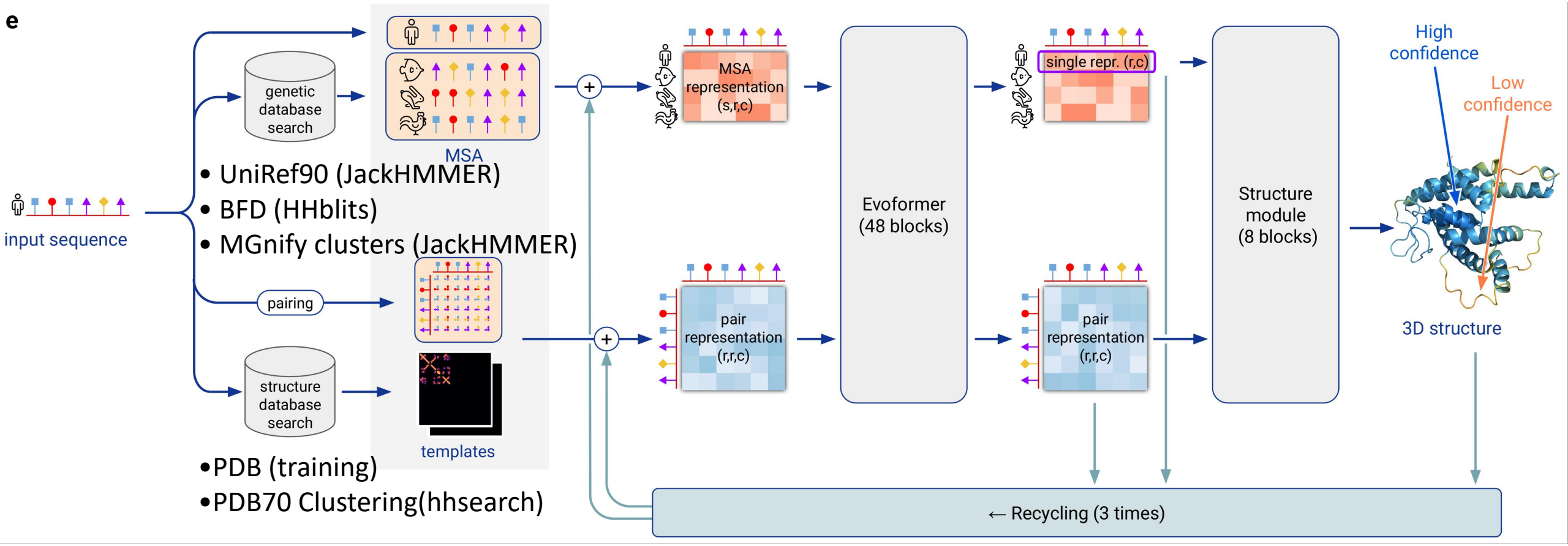


T1044 – AlphaFold / experiment
$RMSD_{95}$: 2.2 Å, TM-score: 0.96

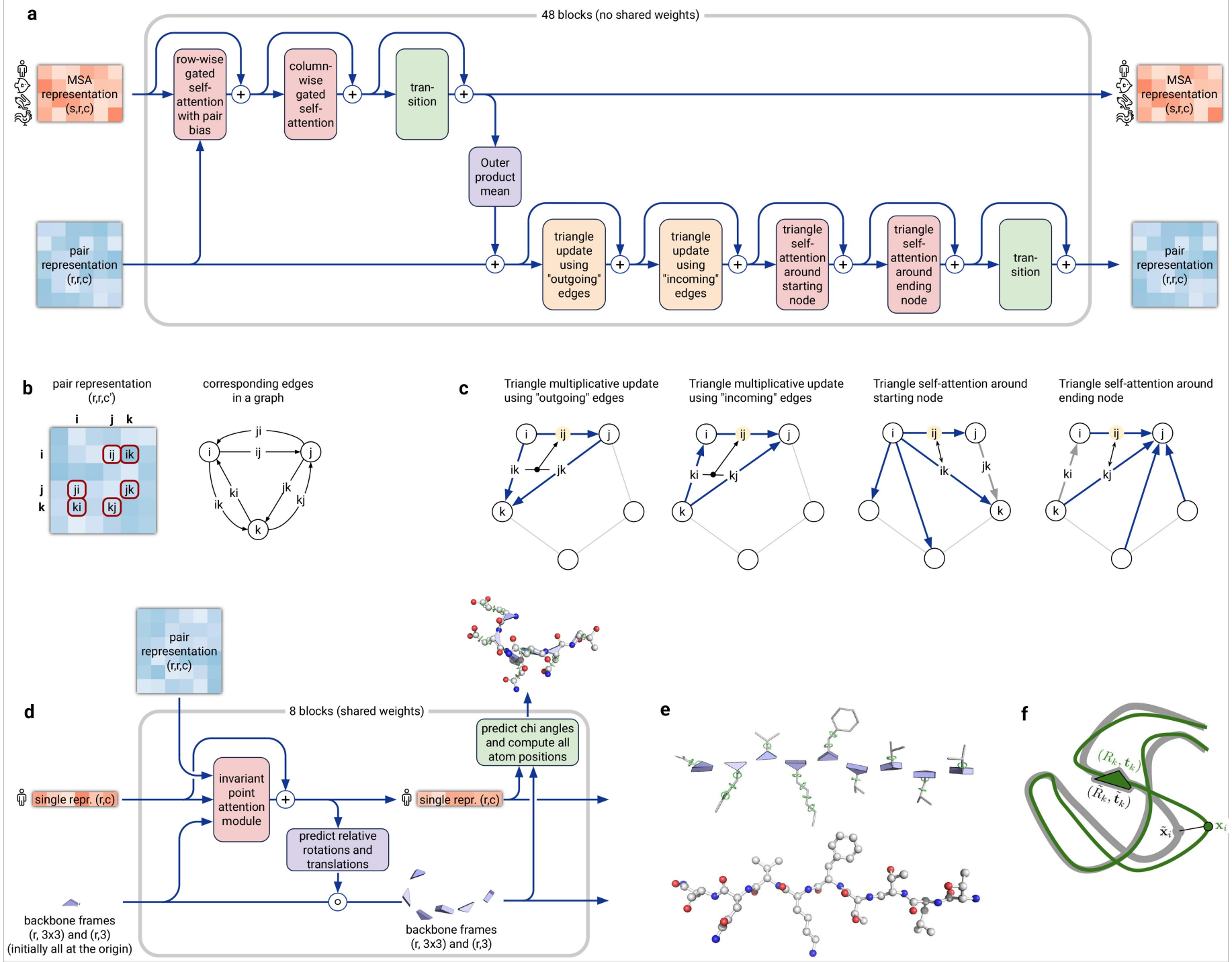
AlphaFold2

Highly accurate protein structure prediction with AlphaFold, Nature, 15 July 2021.
https://doi.org/10.1038/s41586-021-03819-2

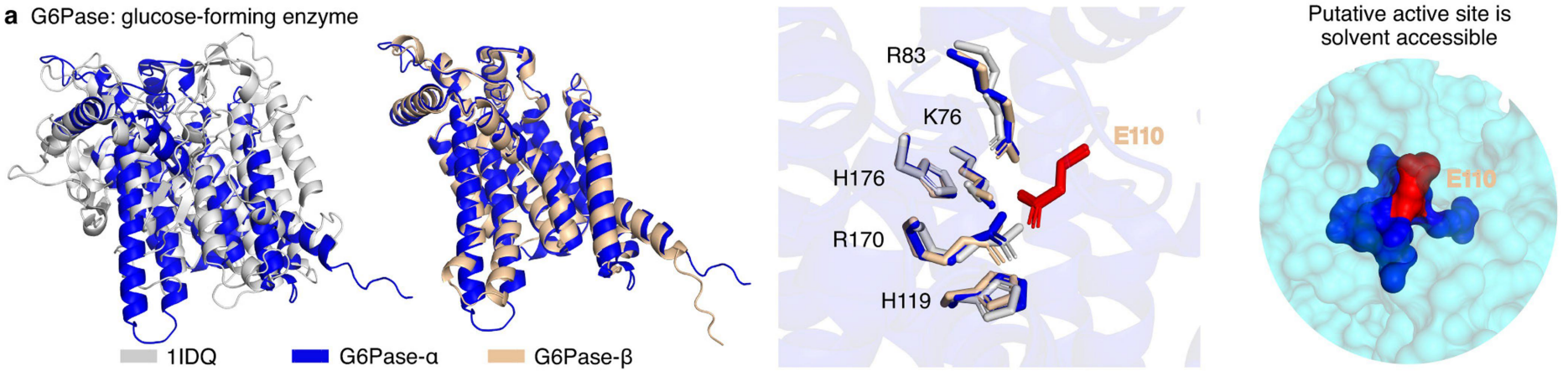
John Jumper ✉, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli & Demis Hassabis ✉ -

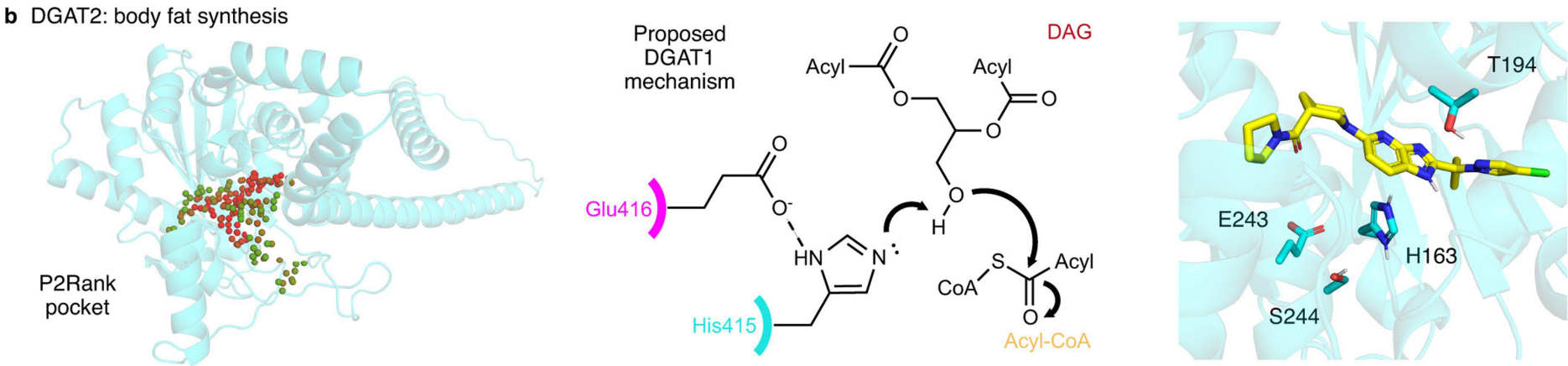


Highly accurate protein structure prediction with AlphaFold,
Nature, https://doi.org/10.1038/s41586-021-03819-2

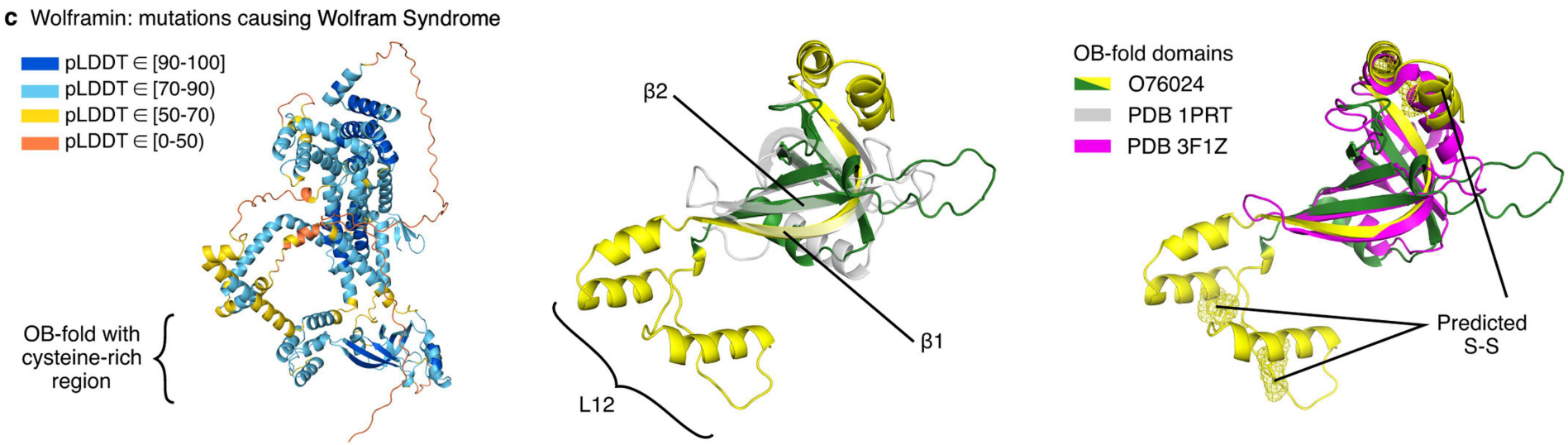


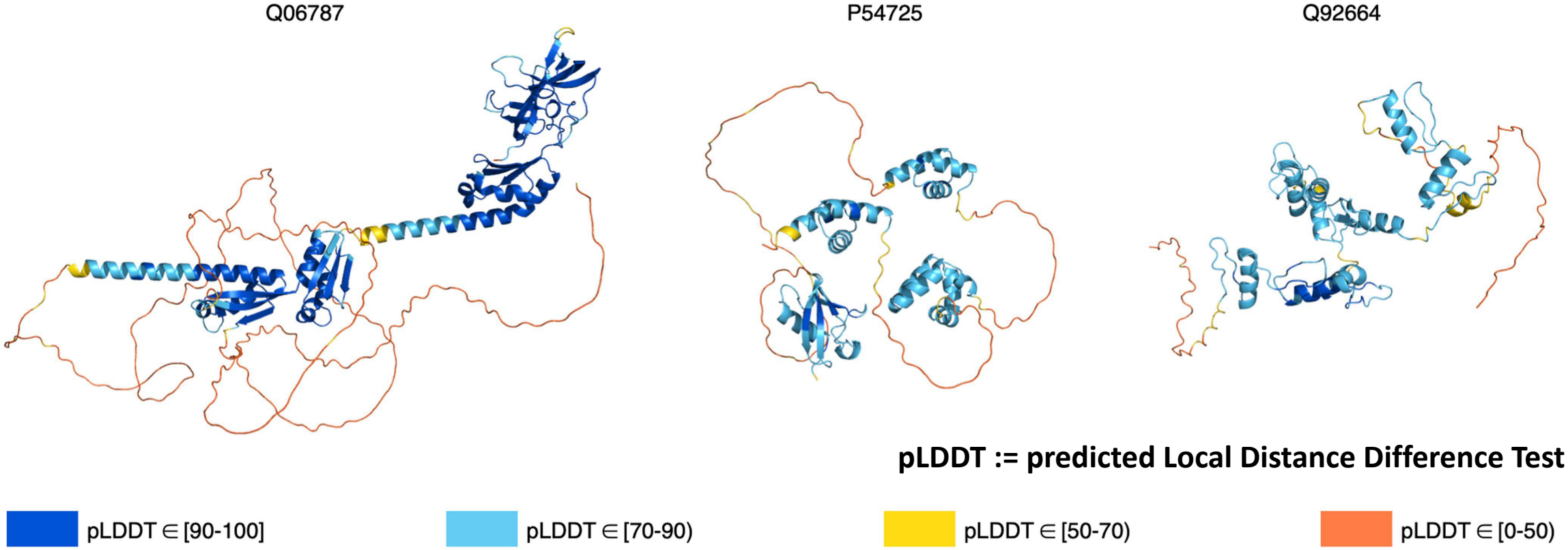
Highly accurate protein structure prediction with AlphaFold, Nature, https://doi.org/10.1038/s41586-021-03819-2





Highly accurate protein structure prediction for the human proteome, Nature, 22 July 2021. https://doi.org/10.1038/s41586-021-03828-1





- Regions with **pLDDT >= 90** are expected to be **modelled to high accuracy**. These should be suitable for any application that benefits from high accuracy (e.g. characterising binding sites).
- Regions with **70 <= pLDDT < 90** are expected to **be modelled well** (a generally good backbone prediction).
- Regions with **50 <= pLDDT < 70** are **low confidence** and should be treated with caution.
- The 3D coordinates of regions with **pLDDT < 50** often has **a reasonably strong predictor of disorder**, i.e. it suggests such a region is either unstructured in physiological conditions or only structured as part of a complex.

Highly accurate protein structure prediction for the human proteome,
Nature, 2021. https://doi.org/10.1038/s41586-021-03828-1

Running AlphaFold on your own

https://github.com/deepmind/alphafold

Hardware required:

- Download **genetic databases** - Total: ~ 2.2 TB (download: 428 GB)
- Download **model parameters** (3.5 GB)

- **Computational resources (i.e. GPU)** required:

  On 1 16-GB V100, 256 AA for 4.8 min, 384 AA for 9.2 min, 2500 AA for 18 hr
- Large proteins need more computer memories:

  The limit for 1 V100 is about 1300 AA. 4 V100s is optimal for 2500 AA

Martin Steinegger(AlphaFold Coauthor) and **Sergey Ovchinnikov** realized AlphaFold on **Google Colaboratory**

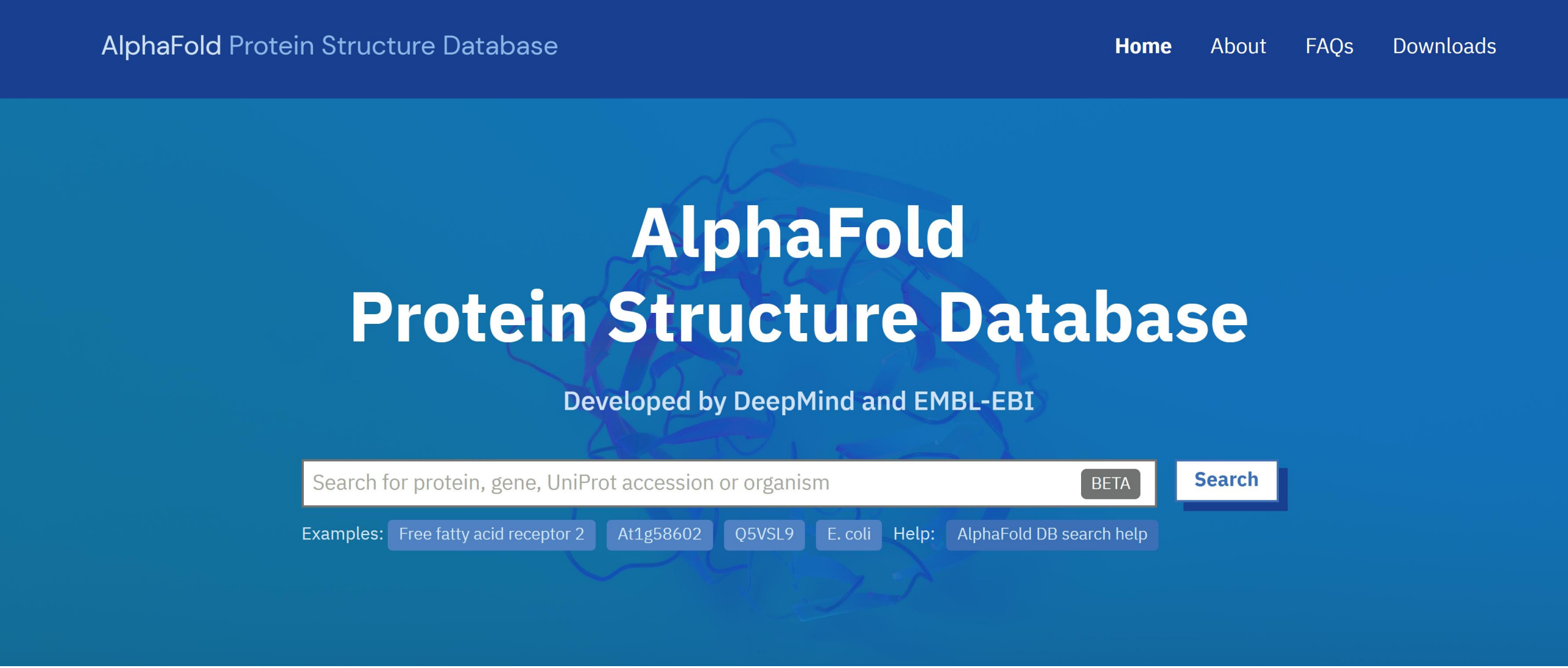
https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb

or

colabfold.com

79 AA for 7 min

Video tutorial
https://www.bilibili.com/video/BV1kY411W7Tu



Highly accurate protein structure prediction for the human proteome,
Nature, 2021. https://doi.org/10.1038/s41586-021-03828-1

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

Species	Common Name	Reference Proteome	Predicted Structures	Download
Arabidopsis thaliana	*Arabidopsis*	UP000006548	27,434	Download (3642 MB)
Caenorhabditis elegans	Nematode worm	UP000001940	19,694	Download (2601 MB)
Candida albicans	*C. albicans*	UP000000559	5,974	Download (965 MB)
Danio rerio	Zebrafish	UP000000437	24,664	Download (4141 MB)
Dictyostelium discoideum	*Dictyostelium*	UP000002195	12,622	Download (2150 MB)
Drosophila melanogaster	Fruit fly	UP000000803	13,458	Download (2174 MB)
Escherichia coli	*E. coli*	UP000000625	4,363	Download (448 MB)
Glycine max	Soybean	UP000008827	55,799	Download (7142 MB)
Homo sapiens	Human	UP000005640	23,391	Download (4784 MB)
Leishmania infantum	*L. infantum*	UP000008153	7,924	Download (1481 MB)
Methanocaldococcus jannaschii	*M. jannaschii*	UP000000805	1,773	Download (171 MB)
Mus musculus	Mouse	UP000000589	21,615	Download (3547 MB)
Mycobacterium tuberculosis	*M. tuberculosis*	UP000001584	3,988	Download (421 MB)
Oryza sativa	Asian rice	UP000059680	43,649	Download (4416 MB)
Plasmodium falciparum	*P. falciparum*	UP000001450	5,187	Download (1132 MB)
Rattus norvegicus	Rat	UP000002494	21,272	Download (3404 MB)
Saccharomyces cerevisiae	Budding yeast	UP000002311	6,040	Download (960 MB)
Schizosaccharomyces pombe	Fission yeast	UP000002485	5,128	Download (776 MB)
Staphylococcus aureus	*S. aureus*	UP000008816	2,888	Download (268 MB)
Trypanosoma cruzi	*T. cruzi*	UP000002296	19,036	Download (2905 MB)
Zea mays	Maize	UP000007305	39,299	Download (5014 MB)

< 2700 AA

Great expectations–the potential impacts of AlphaFold DB

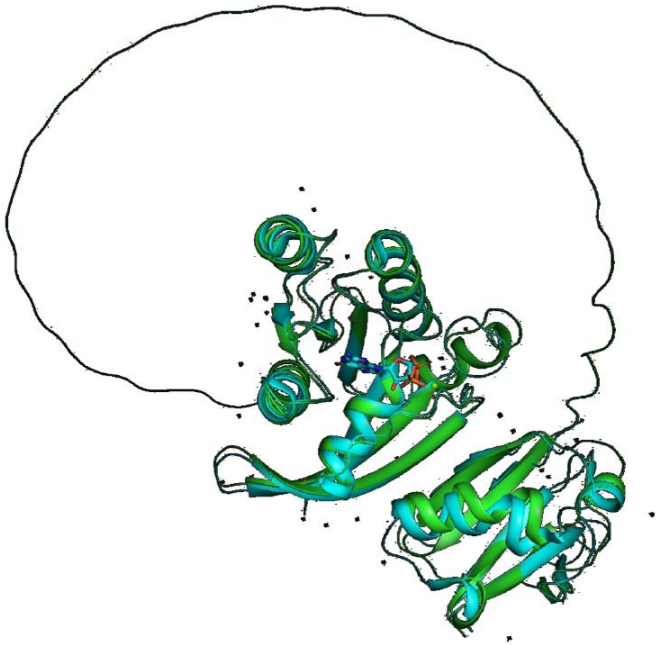
https://www.embl.org/news/science/alphafold-potential-impacts/

"The initial release of AlphaFold DB contains over **350,000 structures** which will eventually increase to an estimated **130 million 3D models** (around 700 times more than currently in the PDB). The functionality of the website will also continuously be improved and extended. " **Currently 992,316 structures.**

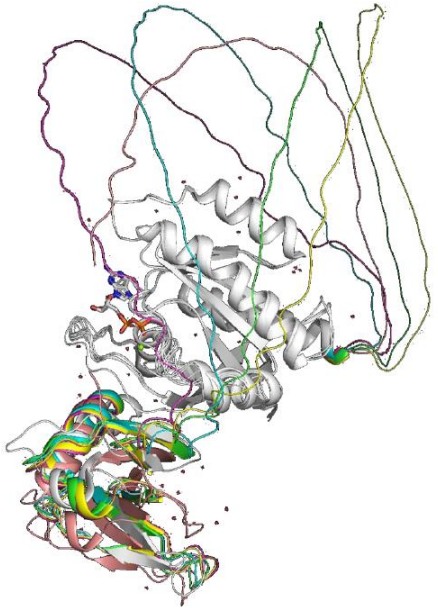
Limitations:

- No prediction for 3D structures for protein-protein or protein-DNA/RNA/ligand **complexes**.
- No predictions for conformational **dynamics** of proteins.
- Low-confidence prediction for **disordered regions** and no claim to predict the 'folding pathway'.
- Not trained or validated for predicting the effect of **mutations**.
- No structure prediction for **cofactors, metals, ligands** including drug-like molecules, ions, carbohydrates and other post-translational modifications.

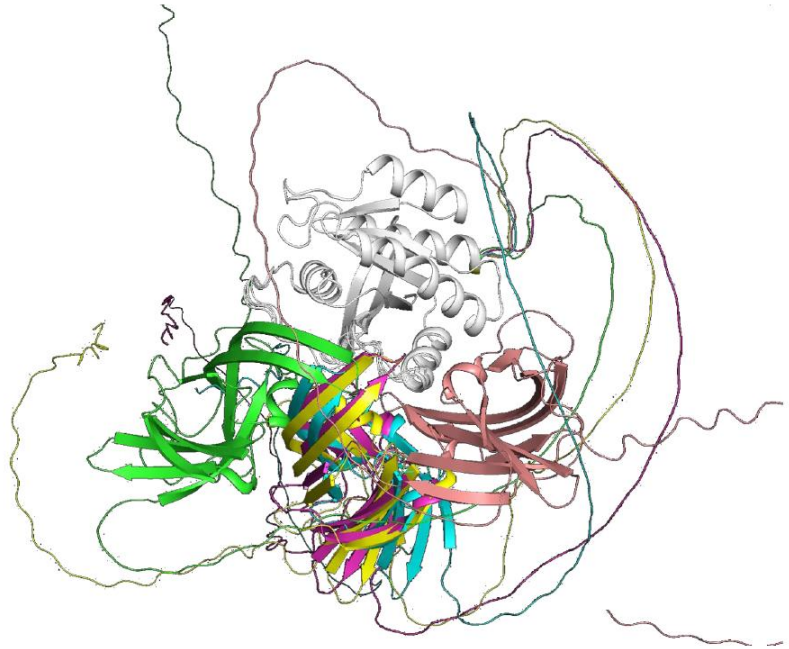
AlphaFold for protein complexes



Adding a long linker



Complex formed



Complex not formed

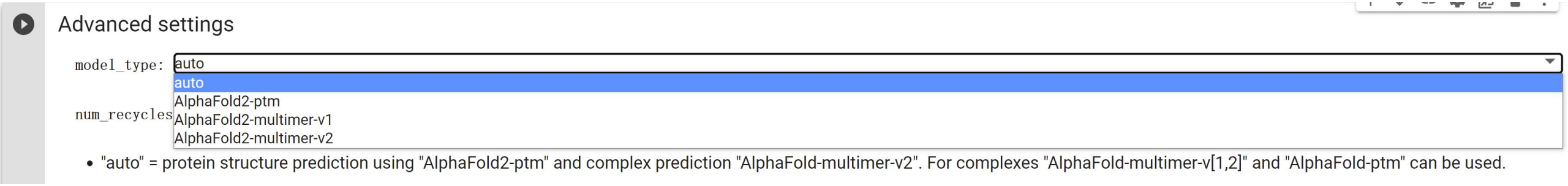
https://twitter.com/thesteinegger/status/1415737015639764995?s=05

Protein complexes

Mode 1

colabfold.com

Query_sequence: Sequence1:Sequence2



Mode 2

Protein complex prediction with AlphaFold-Multimer

bioRxiv 2021.10.04.463034

Richard Evans[1*], Michael O'Neill[1*], Alexander Pritzel[1*], Natasha Antropova[1*], Andrew Senior[1], Tim Green[1], Augustin Žídek[1], Russ Bates[1], Sam Blackwell[1], Jason Yim[1], Olaf Ronneberger[1], Sebastian Bodenstein[1], Michal Zielinski[1], Alex Bridgland[1], Anna Potapenko[1], Andrew Cowie[1], Kathryn Tunyasuvunakool[1], Rishub Jain[1], Ellen Clancy[1], Pushmeet Kohli[1], John Jumper[1*] and Demis Hassabis[1*]

https://cosmic-cryoem.org/tools/alphafoldmultimer/

Mode 3

Homooligomers

https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb

Wide acceptance

- AlphaFold2 paper cited 2127 times on Google Scholar.
- AlphaFold2 Proteome paper cited 413 times.
- ColabFold print cited 144 times.
- AlphaFold-multimer preprint cited 83 times.
- Many labs reproducing AlphaFold2 for various purposes.
- Applications in structures of proteins and protein complexes.

Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization

PNAS, 119 (11) e2122954119 (2022)

AlphaFold Accelerates Artificial Intelligence Powered Drug Discovery: Efficient Discovery of a Novel Cyclin-dependent Kinase 20 (CDK20) Small Molecule Inhibitor

arXiv:2201.09647

AlphaDrug?

AlphaDesign?

AlphaDynamics?

AI works by DeepMind

QUANTUM CHEMISTRY

Science. 2021 Dec 10;374(6573):1385-1389.

Pushing the frontiers of density functionals by solving the fractional electron problem

James Kirkpatrick[1*†], Brendan McMorrow[1†], David H. P. Turban[1†], Alexander L. Gaunt[1†], James S. Spencer[1], Alexander G. D. G. Matthews[1], Annette Obika[1], Louis Thiry[2], Meire Fortunato[1], David Pfau[1], Lara Román Castellanos[1], Stig Petersen[1], Alexander W. R. Nelson[1], Pushmeet Kohli[1], Paula Mori-Sánchez[3], Demis Hassabis[1], Aron J. Cohen[1,4*]

Magnetic control of tokamak plasmas through deep reinforcement learning

https://doi.org/10.1038/s41586-021-04301-9

Received: 14 July 2021

Accepted: 1 December 2021

Published online: 16 February 2022

Open access

Jonas Degrave[1,3], Federico Felici[2,3✉], Jonas Buchli[1,3✉], Michael Neunert[1,3], Brendan Tracey[1,3✉], Francesco Carpanese[1,2,3], Timo Ewalds[1,3], Roland Hafner[1,3], Abbas Abdolmaleki[1], Diego de las Casas[1], Craig Donner[1], Leslie Fritz[1], Cristian Galperti[2], Andrea Huber[1], James Keeling[1], Maria Tsimpoukelli[1], Jackie Kay[1], Antoine Merle[2], Jean-Marc Moret[2], Seb Noury[1], Federico Pesamosca[2], David Pfau[1], Olivier Sauter[2], Cristian Sommariva[2], Stefano Coda[2], Basil Duval[2], Ambrogio Fasoli[2], Pushmeet Kohli[1], Koray Kavukcuoglu[1], Demis Hassabis[1] & Martin Riedmiller[1,3]

Nature. 2022 Feb;602(7897):414-419.

Restoring and attributing ancient texts using deep neural networks

https://doi.org/10.1038/s41586-022-04448-z

Received: 16 August 2021

Accepted: 19 January 2022

Published online: 9 March 2022

Yannis Assael[1,6✉], Thea Sommerschield[2,3,6✉], Brendan Shillingford[1], Mahyar Bordbar[1], John Pavlopoulos[4], Marita Chatzipanagiotou[4], Ion Androutsopoulos[4], Jonathan Prag[5] & Nando de Freitas[1]

Nature. 2022 Mar;603(7900):280-283.