KATEDRA FYZIKÁLNÍ CHEMIE
UNIVERZITY PALACKÉHO V OLOMOUCI

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

elixir CZECH REPUBLIC

eatris-cz

6th Advanced *in silico* Drug Design

KFC/ADD

AlphaFoldology

Machine learning revolution in structural biology

Karel Berka

(w help of Marian Novotny, PřF UK Praha)

UP Olomouc, 30.1.-3.2. 2023

EMBL-EBI

INSTITUTE OF PHYSICS
National academy of Sciences of Ukraine

ÚOCHB AV ČR
IOCB PRAGUE

Outline

- Protein structure prediction
- CASP14
- Alphafold 2 - under the hood
- Basic uses of AF2
- AFDB
- AF2 publicly available servers
- Limitations and challenges - Alphafoldology
- Future - CASP15

elixir CZECH REPUBLIC

NEWS · 30 NOVEMBER 2020

‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

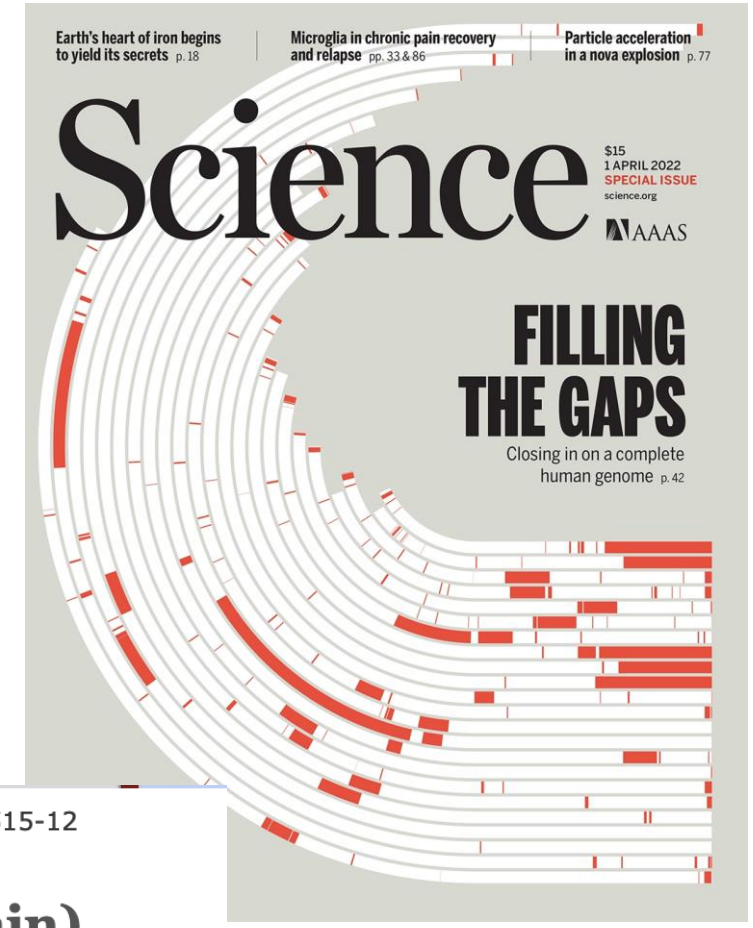
‘The game has changed.’ AI triumphs at solving protein structures

We have been stuck on this one problem – how do proteins fold up – for nearly 50 years. To see DeepMind produce a solution for this, having worked personally on this problem for so long and after so many stops and starts, wondering if we’d ever get there, is a very special moment.

elixir CZECH REPUBLIC

PROFESSOR JOHN MOULT
CO-FOUNDER AND CHAIR OF CASP, UNIVERSITY OF MARYLAND

(almost) Complete human genome



Published: 21 October 2004

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium

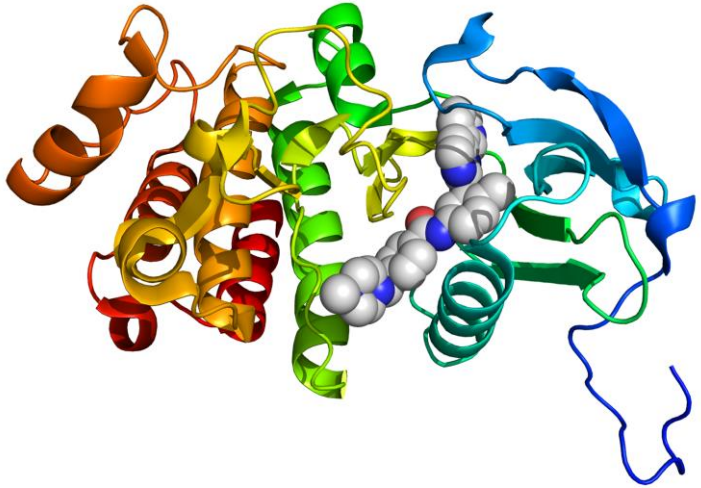
Published online 17 May 2006 | Nature | doi:10.1038/news060515-12

News

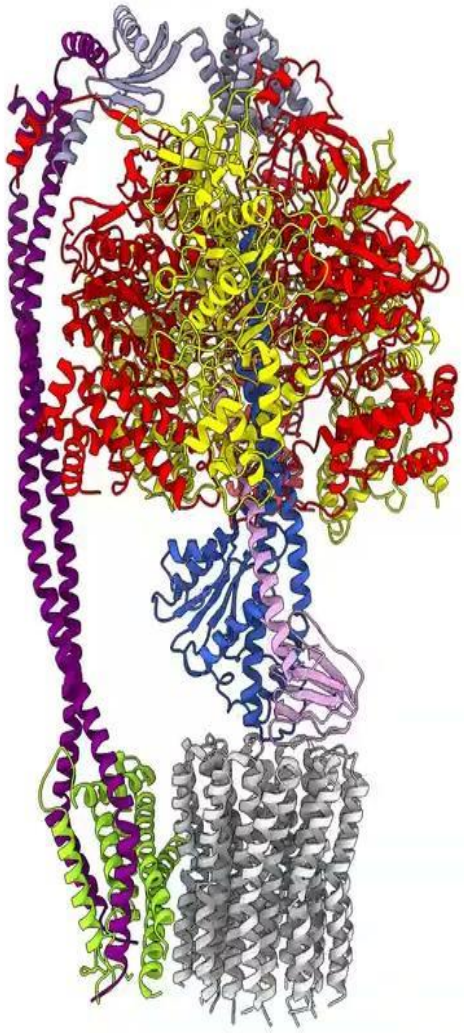
Human genome completed (again)

Scientists today publish the sequence of chromosome 1: the largest and last of the human chromosomes to be done and dusted. *News@nature* finds out what this latest milestone means.

Knowing structure to understand function

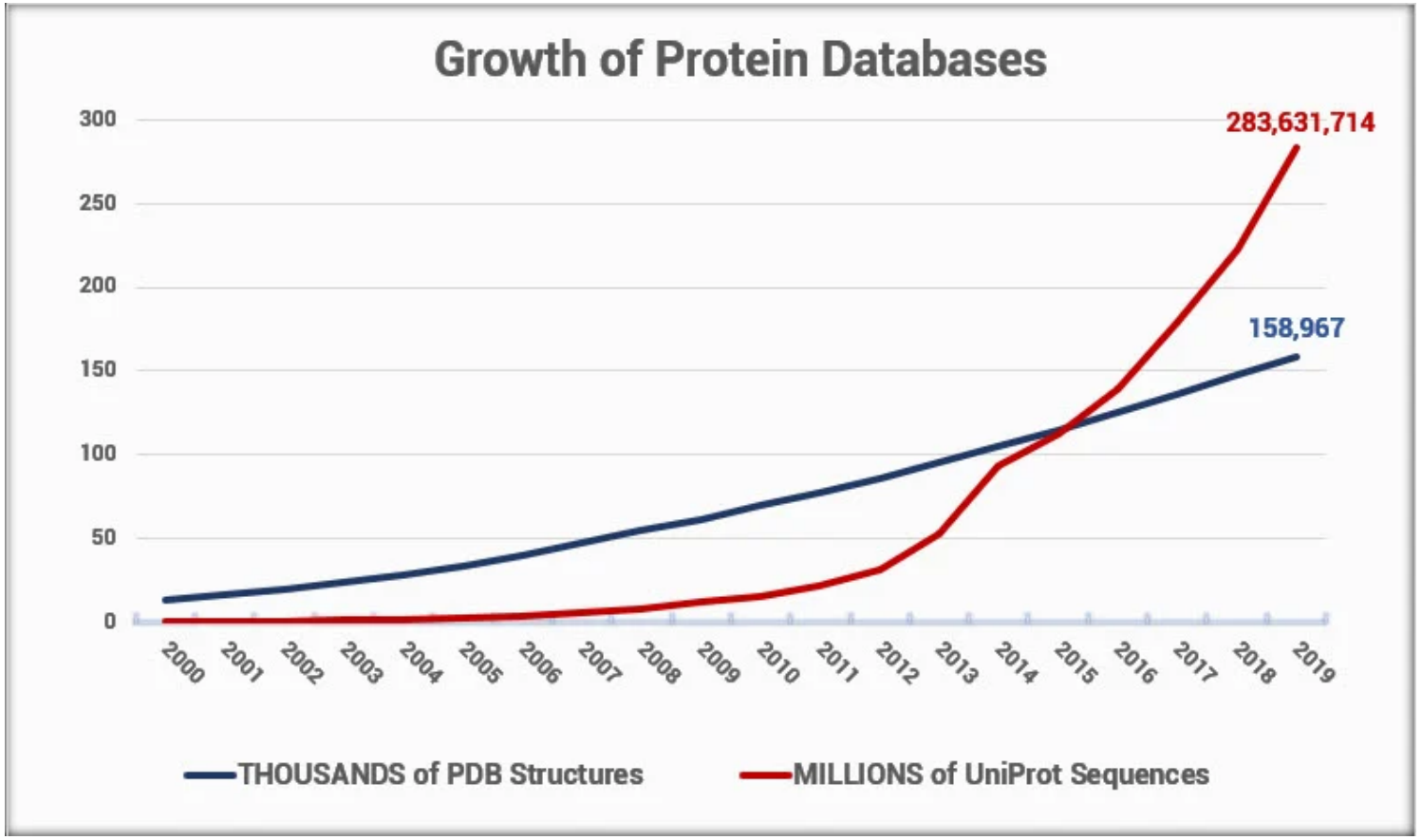


wikipedia/imatinib



Guo et al., 2019

Solving 3D structures is still difficult...



https://www.dnastar.com/blog/structural-biology/why-structure-prediction-matters

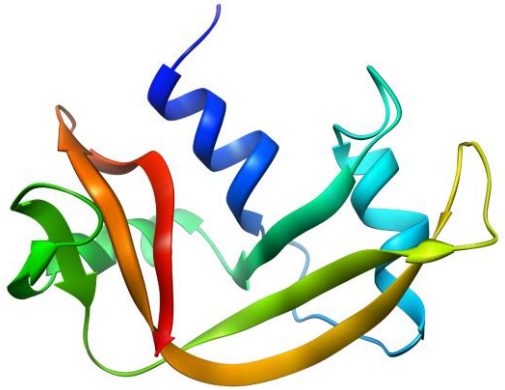
The gap between numbers of experimental structures and sequences is increasing over time

Can we use sequence to predict 3D structure?

- C.B. Anfinsen received Nobel prize in Chemistry (1972) for describing the relationship between sequence and structure

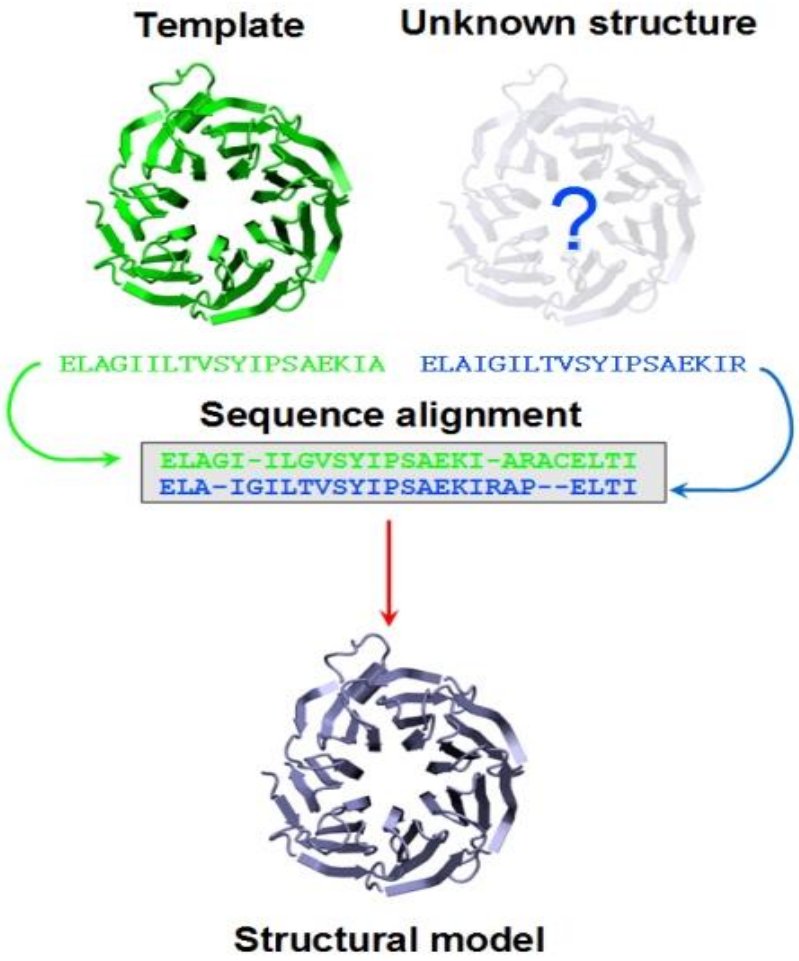
"The native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment."

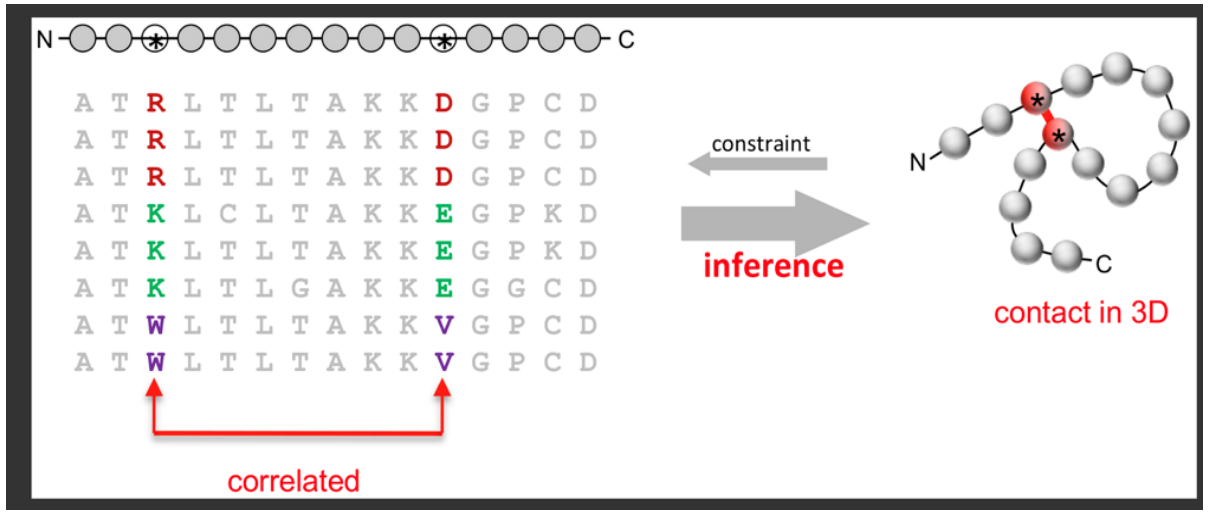
- it shall be possible to predict structure from sequence

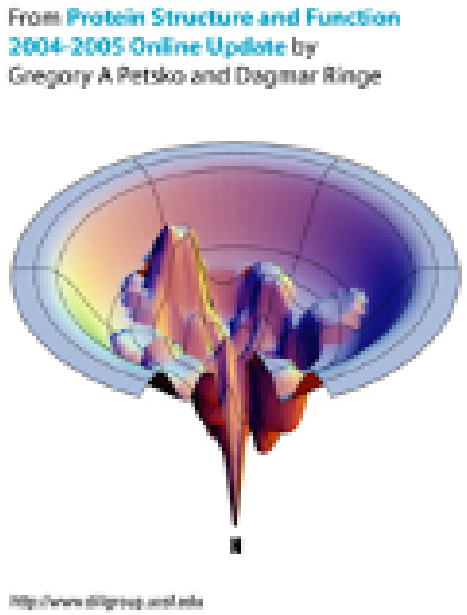


ribonuclease A
Wikipedia

Principles of prediction from sequence





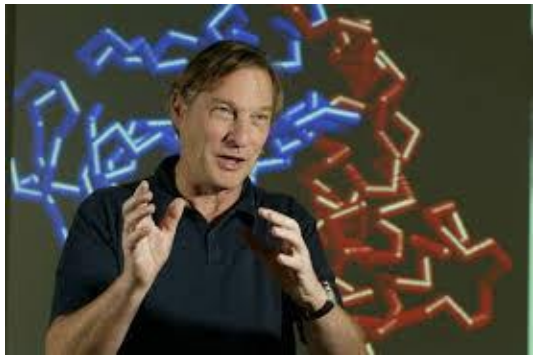


https://www.unil.ch/pmf/en/home/menuinst/technologies/homology-modeling.html

elixir CZECH REPUBLIC

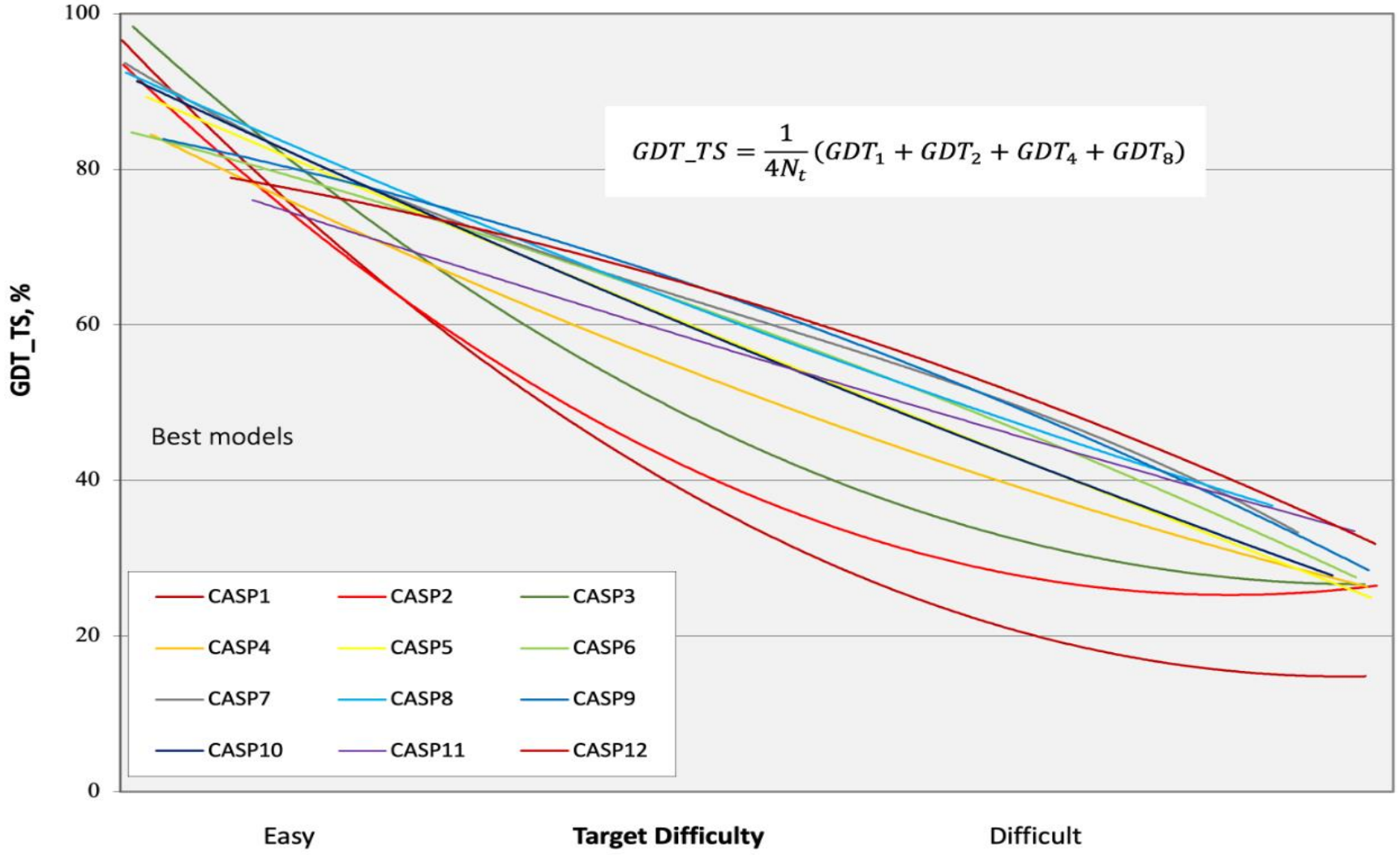
How to move the prediction field forward?

- transparent competition
- provide an “environment” for communication and exchange of experience
- develop metrics for careful examination of predicted structures



- **CASP** – critical assessment of protein structure prediction
- once in two years since 1994
- compare with experimentally solved structures

How to compare structures?



https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf

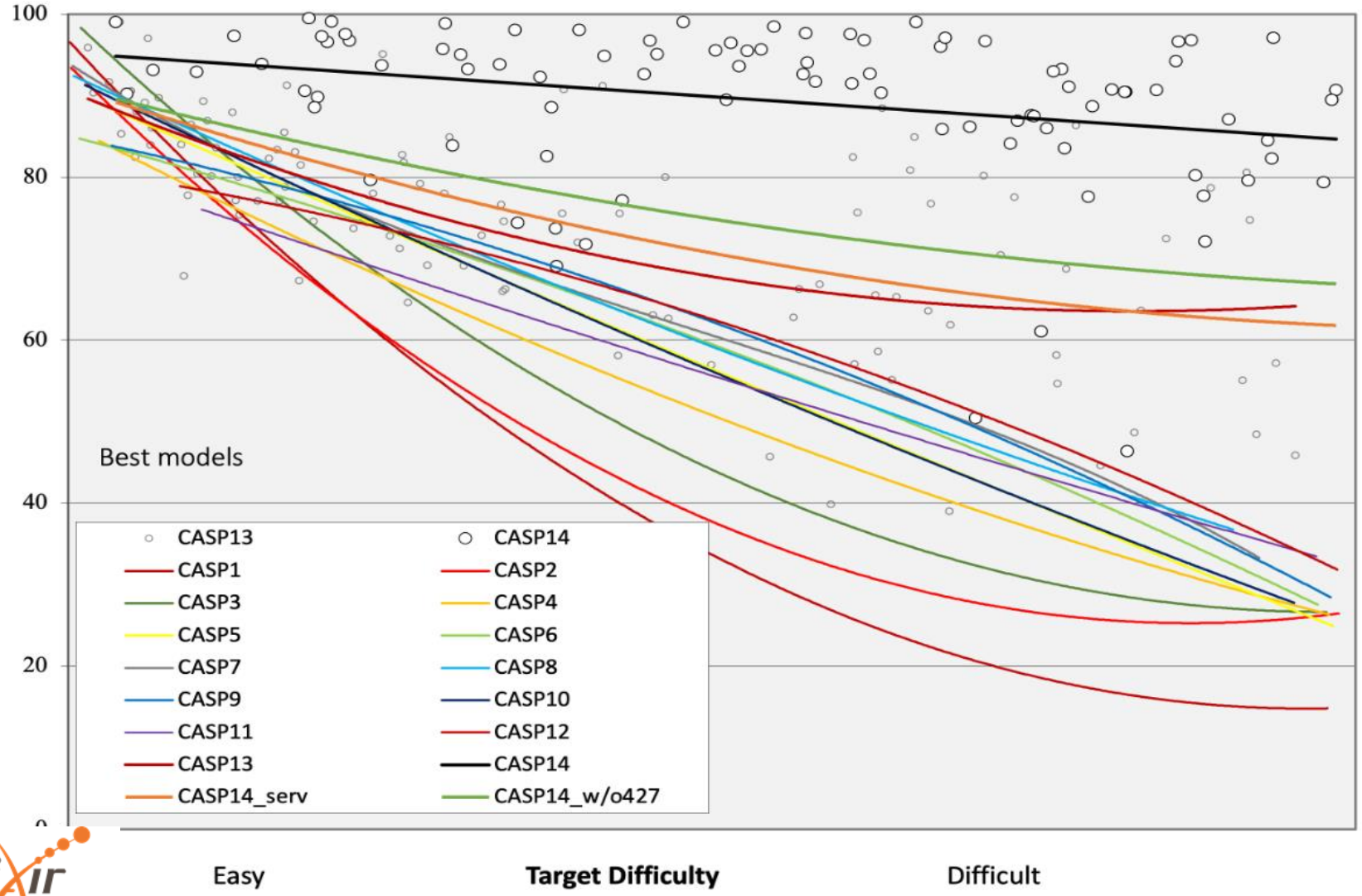
GDT_TS = Global distance test - total score (max 100%)

The conventional GDT_TS total score in CASP is the average result of cutoffs at 1, 2, 4, and 8 Å falling within experimental position

2018 CASP 13: AlphaFold enters...

https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf

2020 CASP 14: Alphafold2 wins



https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf

2022 CASP15: AlphaFold is basic

#1 PEZYFoldings AF2-based. Diverse MSAs. Custom, fine-tuned AF2 refinement

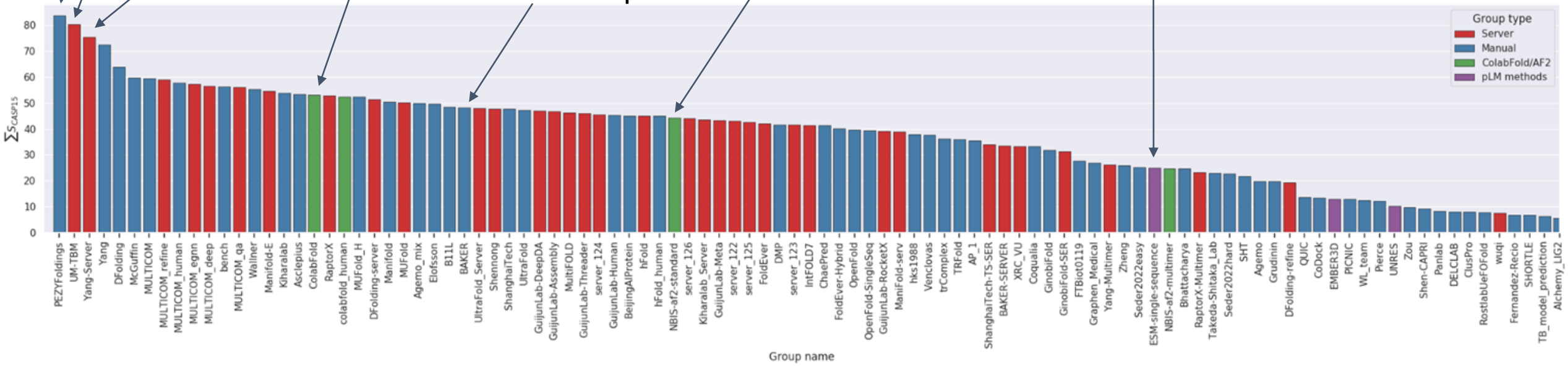
#2 UM-TBM Diverse MSAs. Threading then AF2 predictions guide I-TASSER REMC

#3 Yang-Server Diverse MSAs. AF2 predictions fed to trRosettaX2

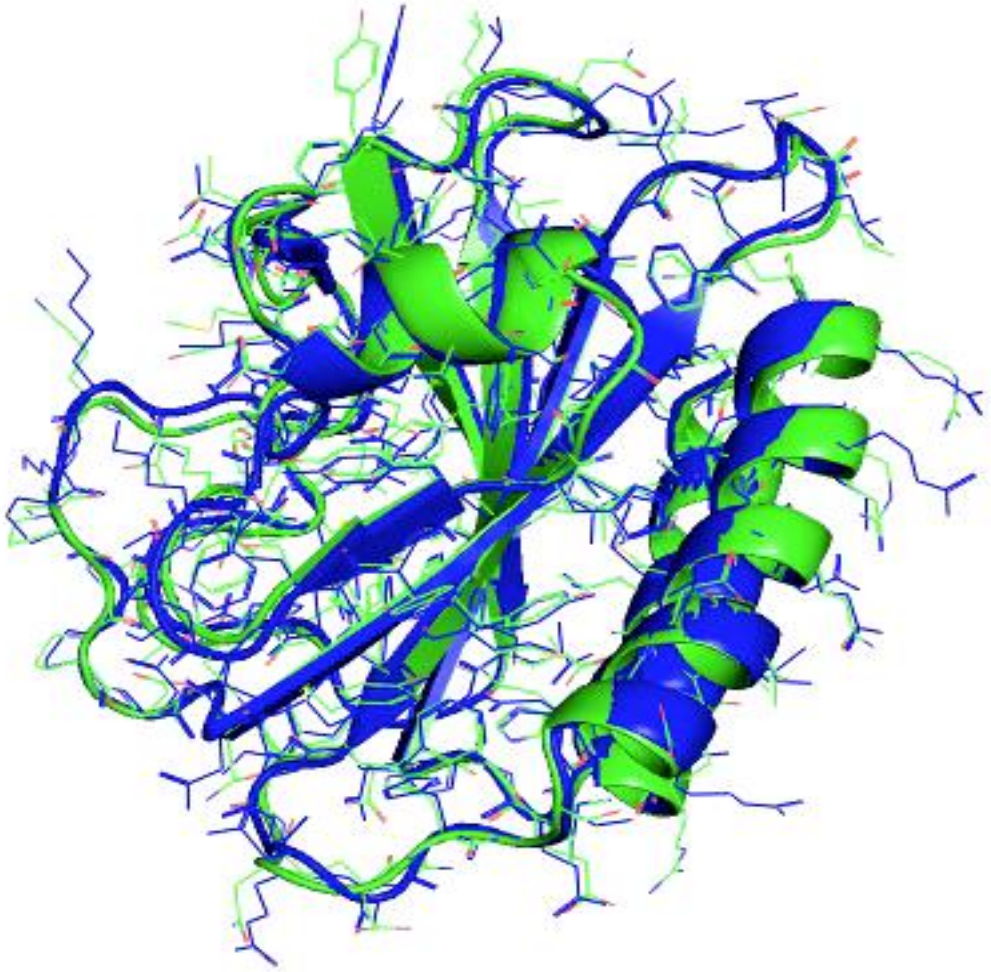
ColabFold and NBIS-af2-standard

BAKER top non-AF2 method

ESM-singlesequence is the top pure pLM method by this metric

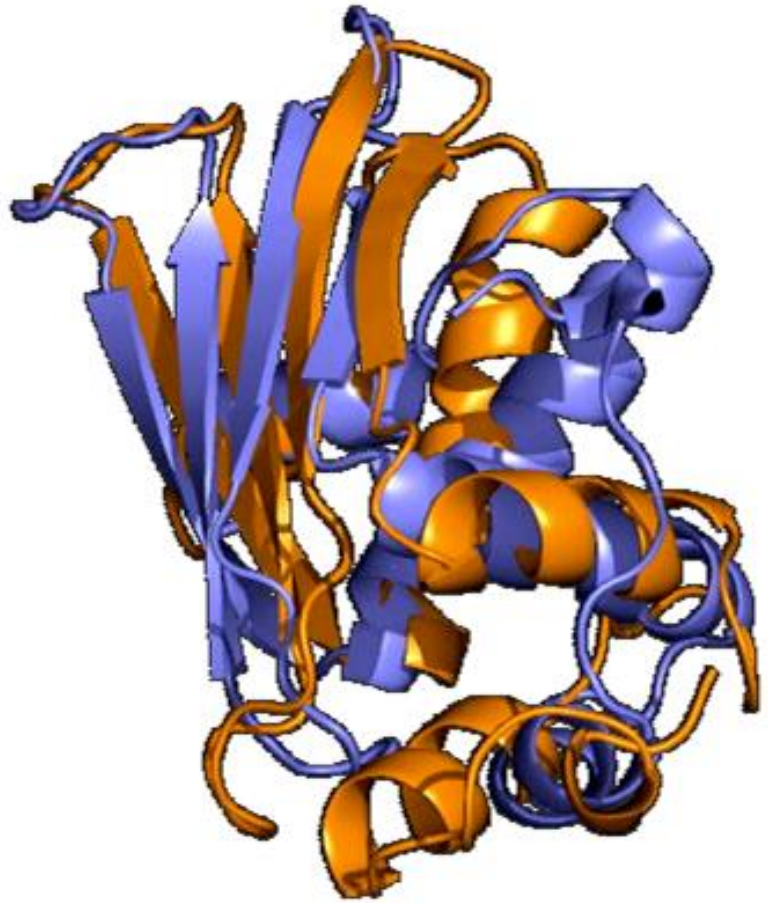


How does good prediction look like?



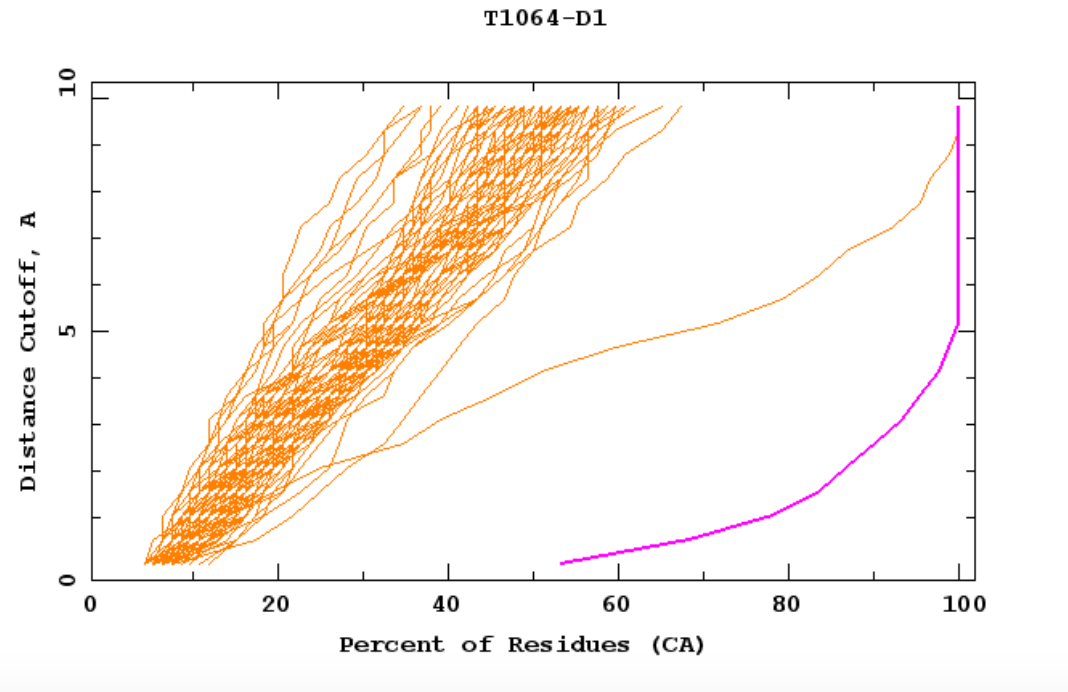
GDT_TS = 96.5

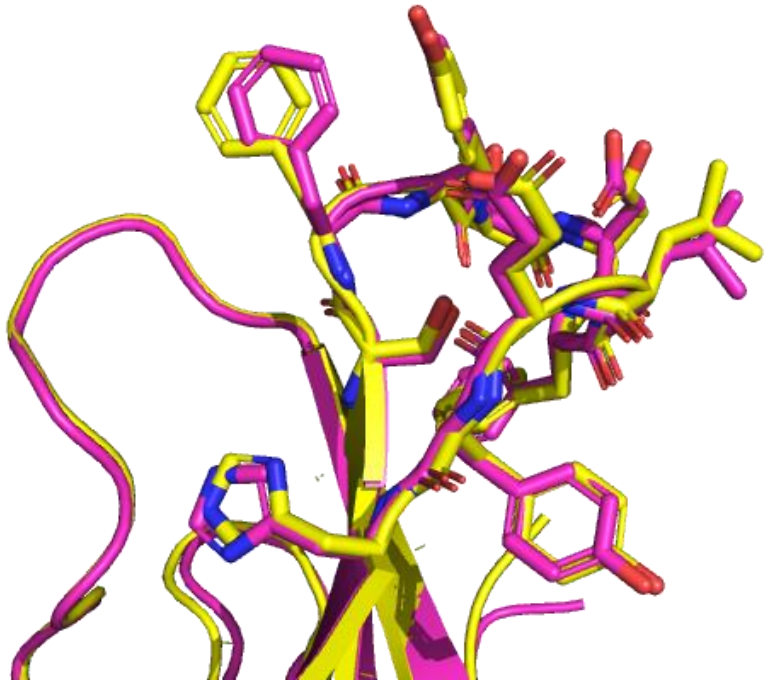
The worst prediction of Alphafold 2 in CASP 14



GDT_TS = 44.6

Side chain predictions– orf8 covid19





GDT_TS= 87

Also good
so how does it works?

AlphaFold2
- under the hood

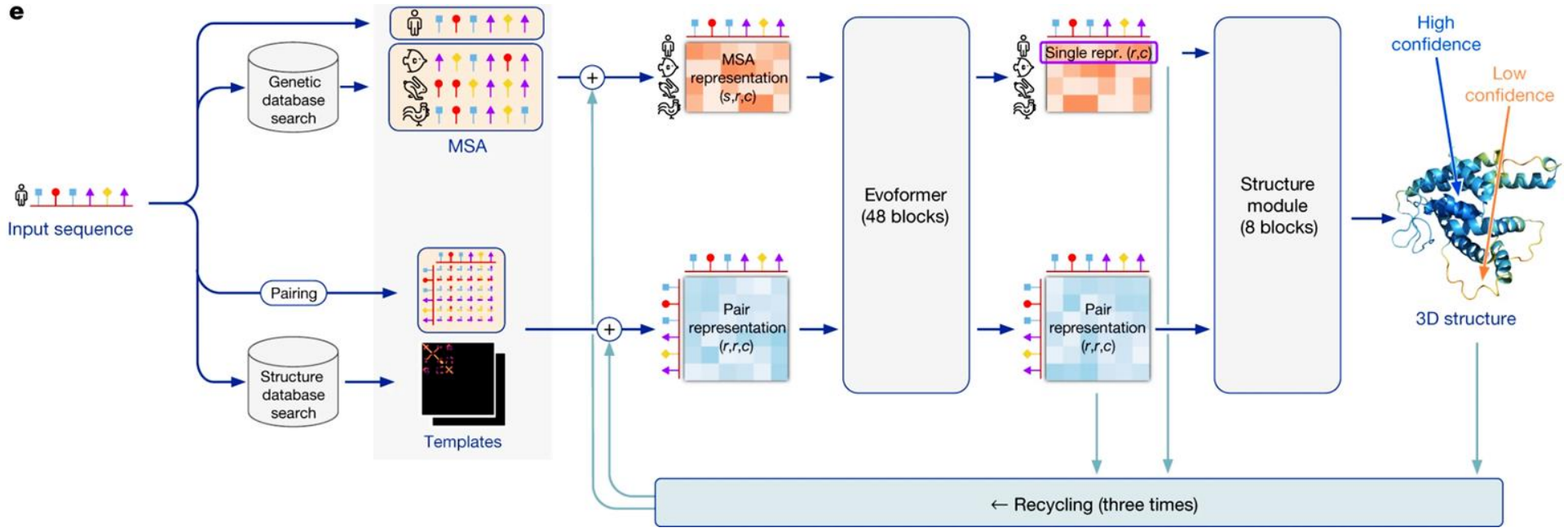
AlphaFold2

Input: sequence

extended by MSA + structural templates

Evoformer and Structure modules (w MD simulation)

pLDDT - predicted local confidence prediction



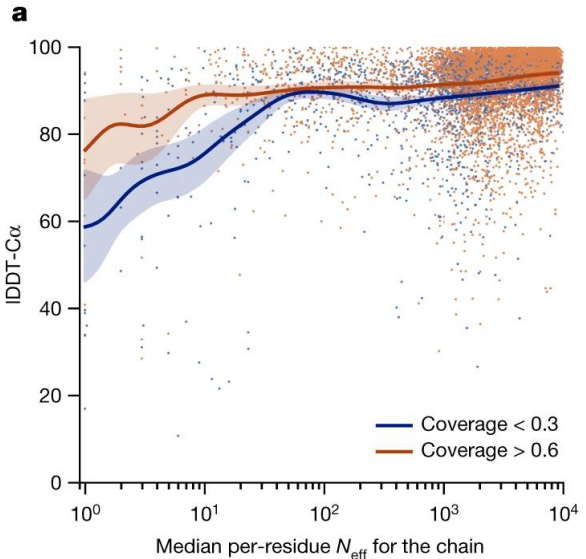
https://www.nature.com/articles/s41586-021-03819-2

MSA - multiple sequence alignment

using standard tools - jackhmmer, HHBlits

- sequence DBs:
  - *UniRef90*
  - *UniClust30* = for sequence self-distilation
- metagenomicsDBs - to fully cover classes underepresented in UniRef90
  - *Big Fantastic database (BFD)* = 66M protein families from 2.2G protein sequences
  - clustered *MGnify*

needed at least 30 sequences per MSA otherwise quality deteriorated ->



https://www.nature.com/articles/s41586-021-03819-2

Training

PDB database + PDB70 clusters

1. training db:

- 40% identity clusters, crop to 258 residues, batches by 128 per Tensor processing unit (TPU)

2. enhance accuracy by **noisy student self-distillation**

- predict 350000 structures from UniRef30 using trained network
- filter to high confidence subset
- then train again from scratch with mixture of PDB and UniRef30

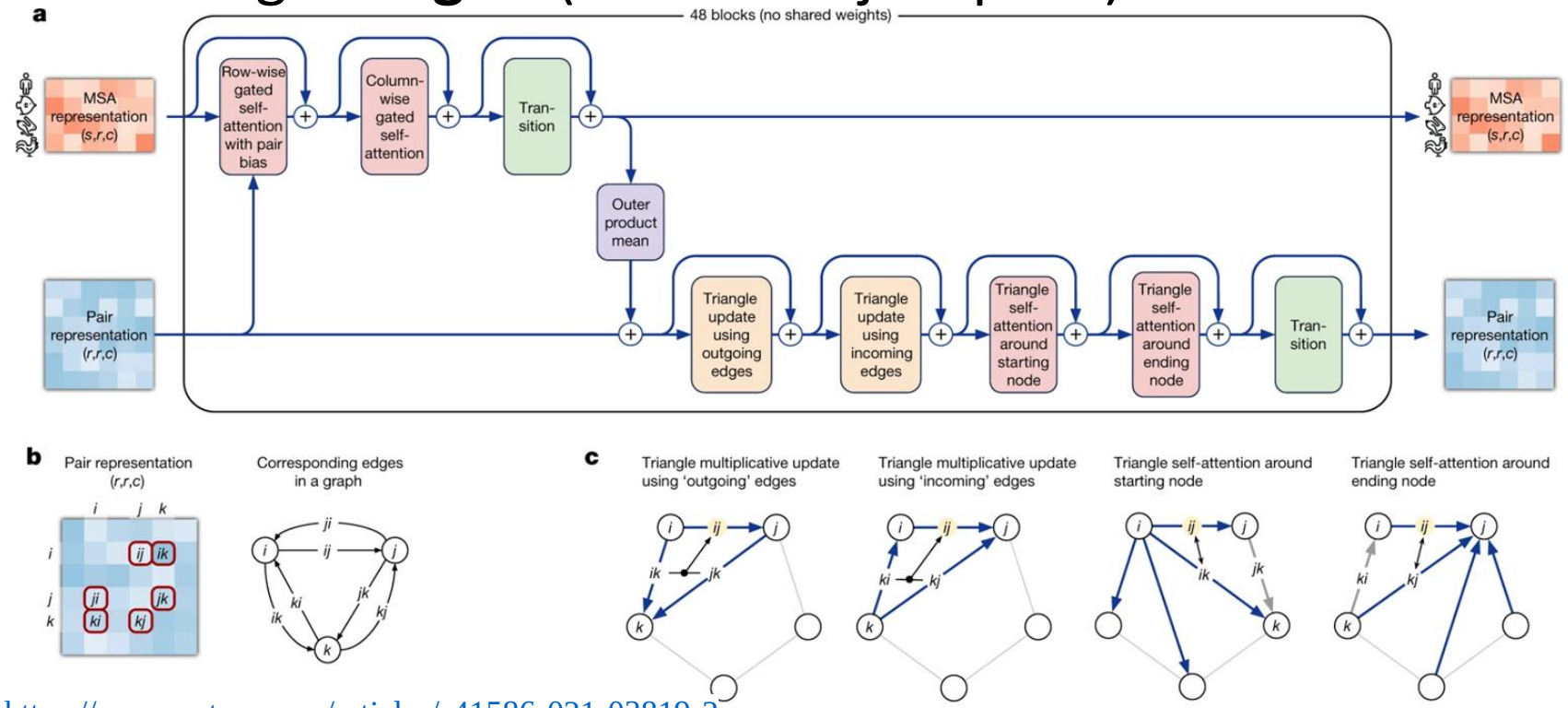
=> effective use of unlabeled sequence data

3. randomly mask or mutate individual residues from MSA using BERT (bidirectional encoder representations from Transformers => to predict masked elements within MSA

https://www.nature.com/articles/s41586-021-03819-2

EvoFormer module

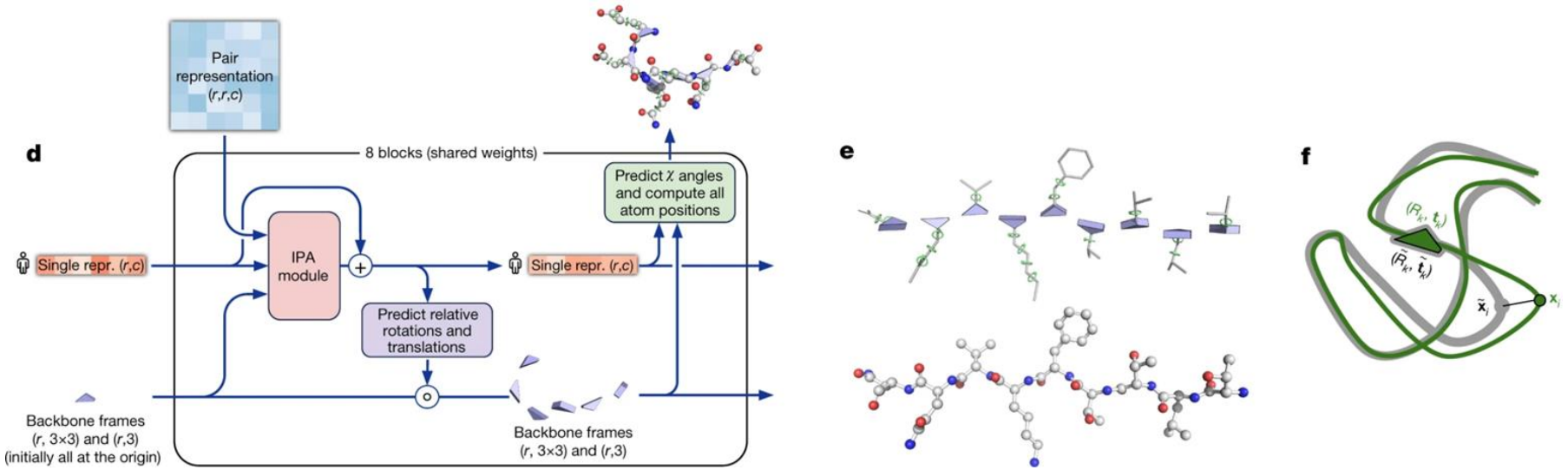
- **mixing** MSA and pairs via updates
- graph inference problem in 3D space
  - edges = residues in proximity
  - updates per each block (48 blocks) separately (AF1 updated all network at once)
- using **triangles** (instead of just pairs)



https://www.nature.com/articles/s41586-021-03819-2

Structure module

- prioritize backbone positions+orientations
  - **residue gas** - free floating rigid body rotations and translation
  - updates
    - IPA (invariant point attention) - neural activations only in rigid 3D
    - equivariant update using updated activations
- later fix backbone geometry
  - **avoid loop closure problem**
- sidechain final refinement:
  - OpenMM with Amber 99sb forcefield

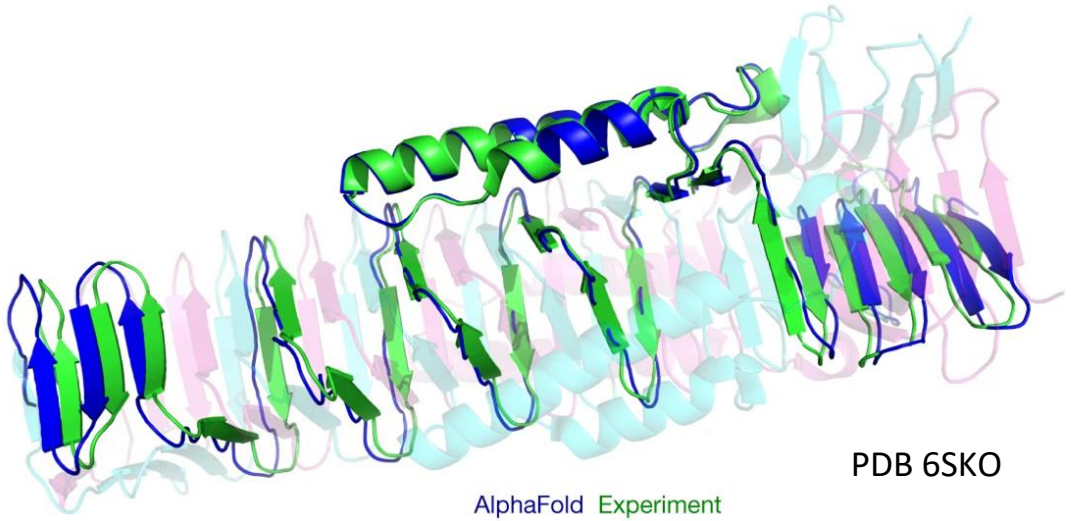


https://www.nature.com/articles/s41586-021-03819-2

Effect of cross-chain contacts

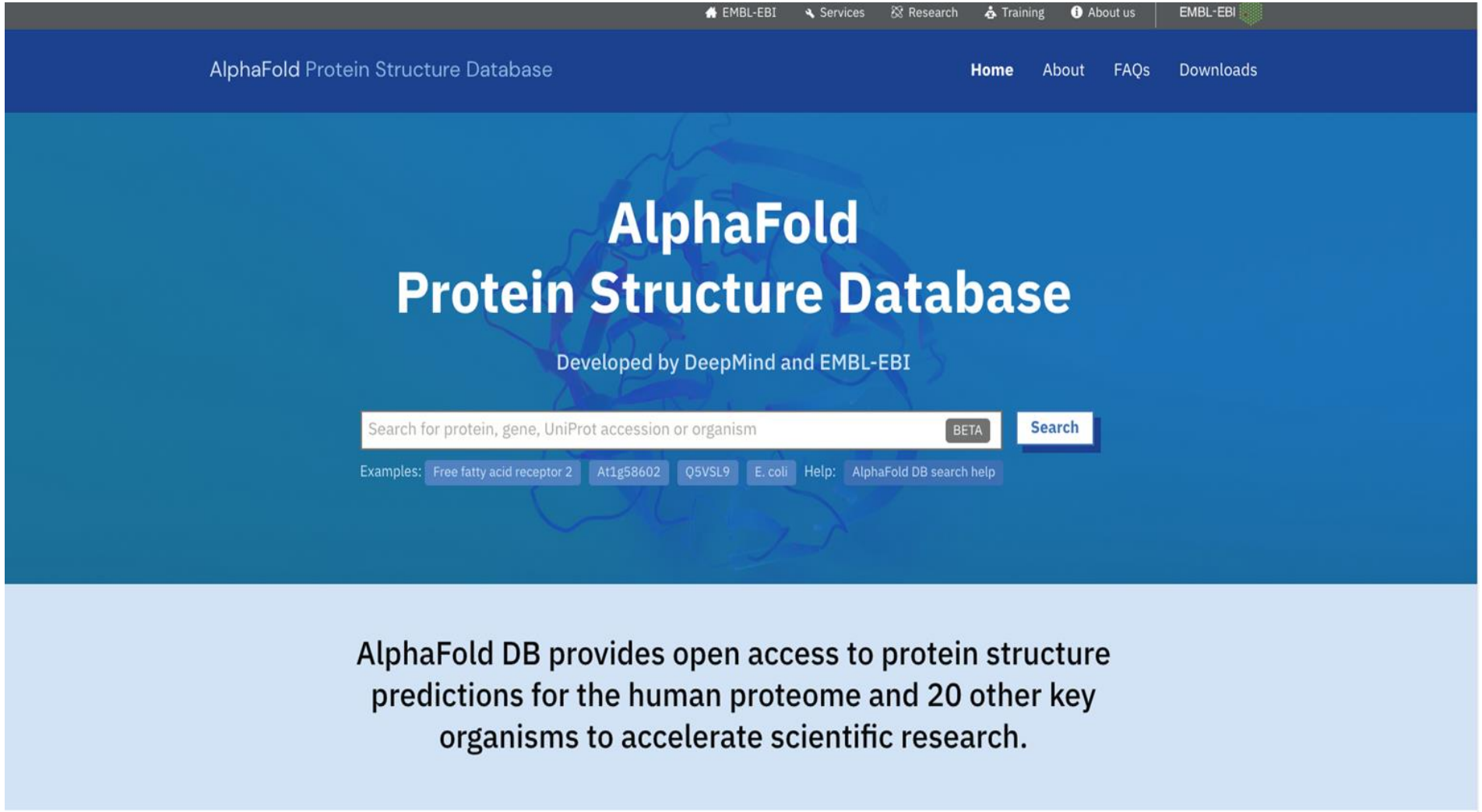
prediction is worse for heterotropic contacts (large complexes where 3D structure is dictated by other chains in complex)

homotropics yields high-accuracy even when chains are intertwinned



PDB 6SKO

AlphaFoldDB



"This will be one of the most important datasets since the mapping of the Human Genome."
Professor Ewan Birney
EMBL Deputy Director General and EMBL-EBI Director

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

Complete structures of proteomes

AlphaFold DB currently provides predicted structures for the 48 organisms listed below, as well as the majority of Swiss-Prot.

Compressed prediction files for model organism proteomes:

Species	Common Name	Reference Proteome	Predicted Structures	Download
Arabidopsis thaliana	*Arabidopsis*	UP000006548	27,434	Download (3,678 MB)
Caenorhabditis elegans	Nematode worm	UP000001940	19,694	Download (2,626 MB)
Candida albicans	*C. albicans*	UP000000559	5,974	Download (974 MB)

Compressed prediction files for global health proteomes:

Species	Common Name	Reference Proteome	Predicted Structures	Download
Ajellomyces capsulatus	Ajellomyces capsulatus	UP000001631	9,199	Download (1,351 MB)
Brugia malayi	Brugia malayi	UP000006672	8,743	Download (1,274 MB)
Campylobacter jejuni	*C. jejuni*	UP000000799	1,620	Download (173 MB)
Cladophialophora carrionii	Cladophialophora carrionii	UP000094526	11,170	Download (1,716 MB)

Compressed prediction files for Swiss-Prot:

File type	Predicted Structures	Download
Swiss-Prot (CIF files)	542,380	Download (36,896 MB)
Swiss-Prot (PDB files)	542,380	Download (26,935 MB)

SNW domain-containing protein 1

AlphaFold structure prediction

Download **PDB file** **mmCIF file** **Predicted aligned error**

Information

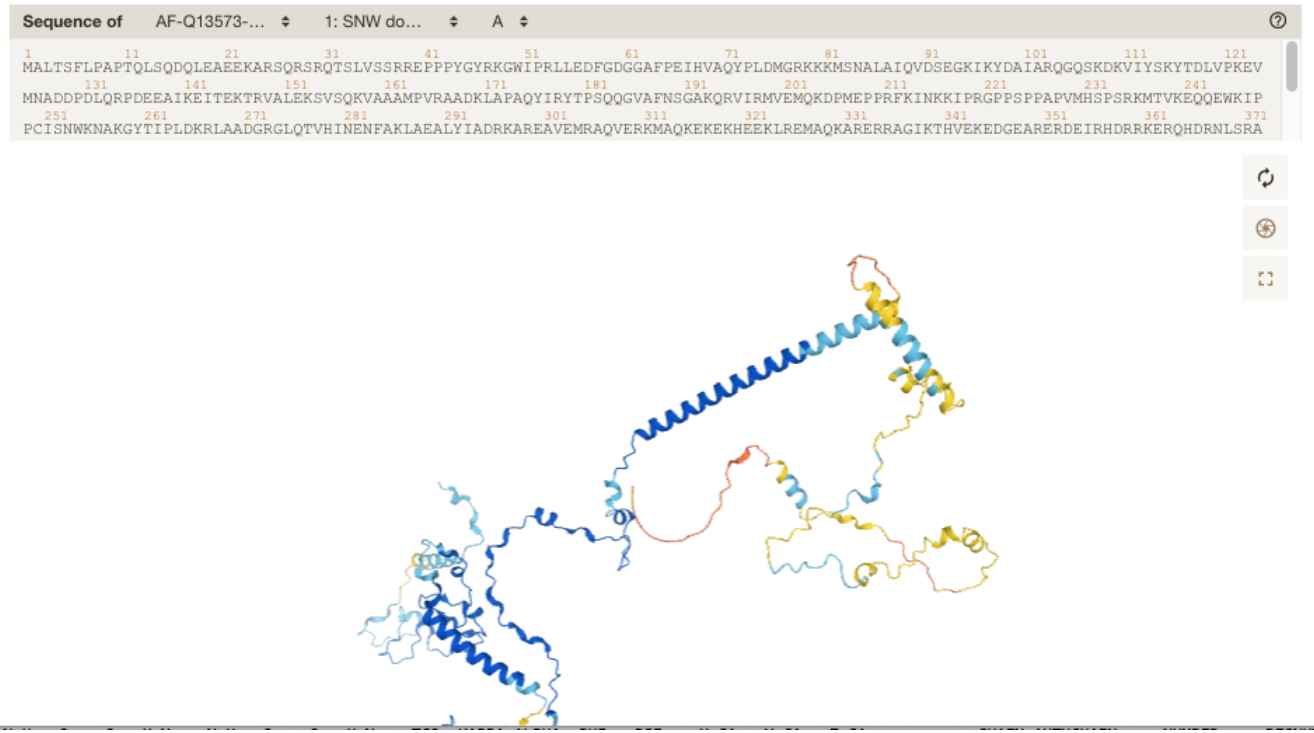
Protein	SNW domain-containing protein 1
Gene	SNW1
Source organism	Homo sapiens go to search
UniProt	Q13573 go to UniProt
Experimental structures	17 structures in PDB for Q13573 go to PDBe-KB
Biological function	(Microbial infection) Proposed to be involved in transcriptional activation by EBV EBNA2 of CBF-1/RBPJ-repressed promoters. go to UniProt

3D viewer

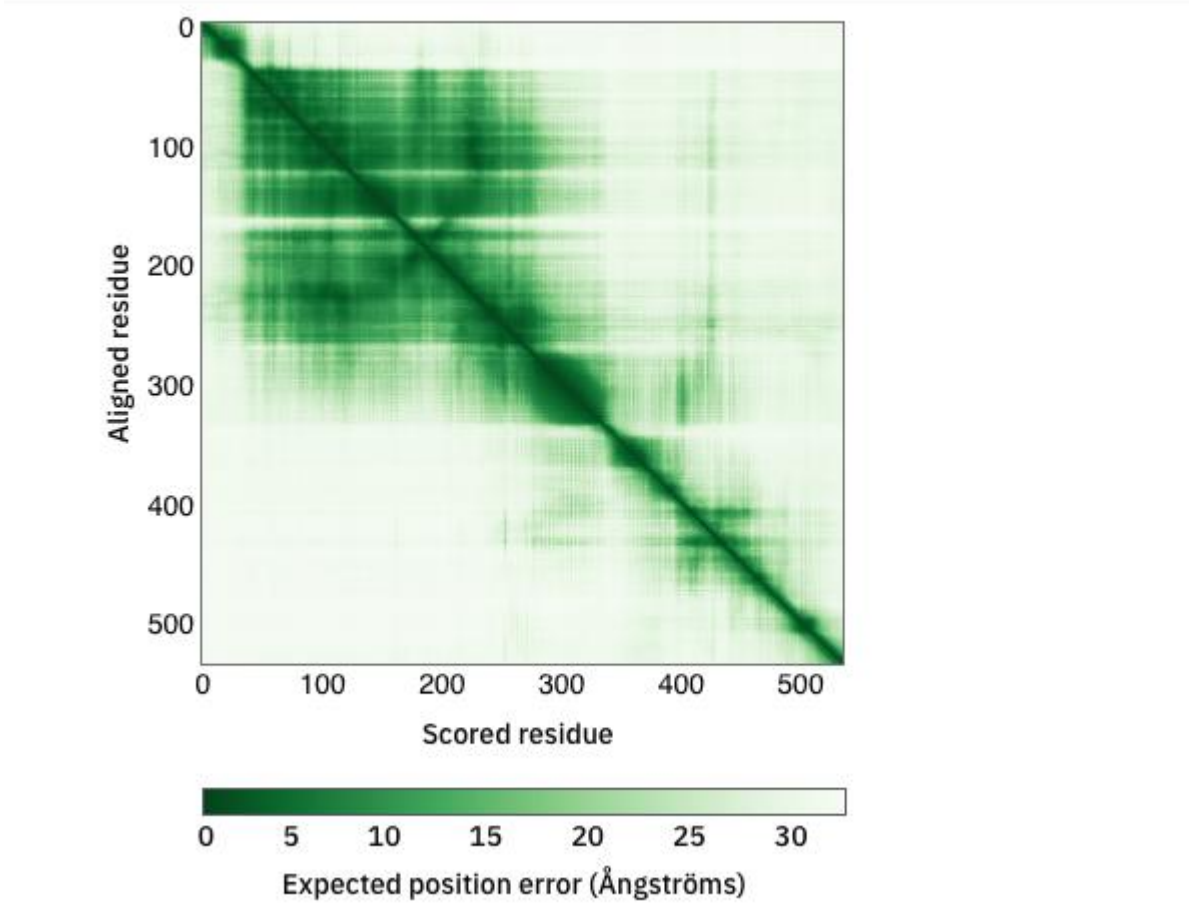
Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

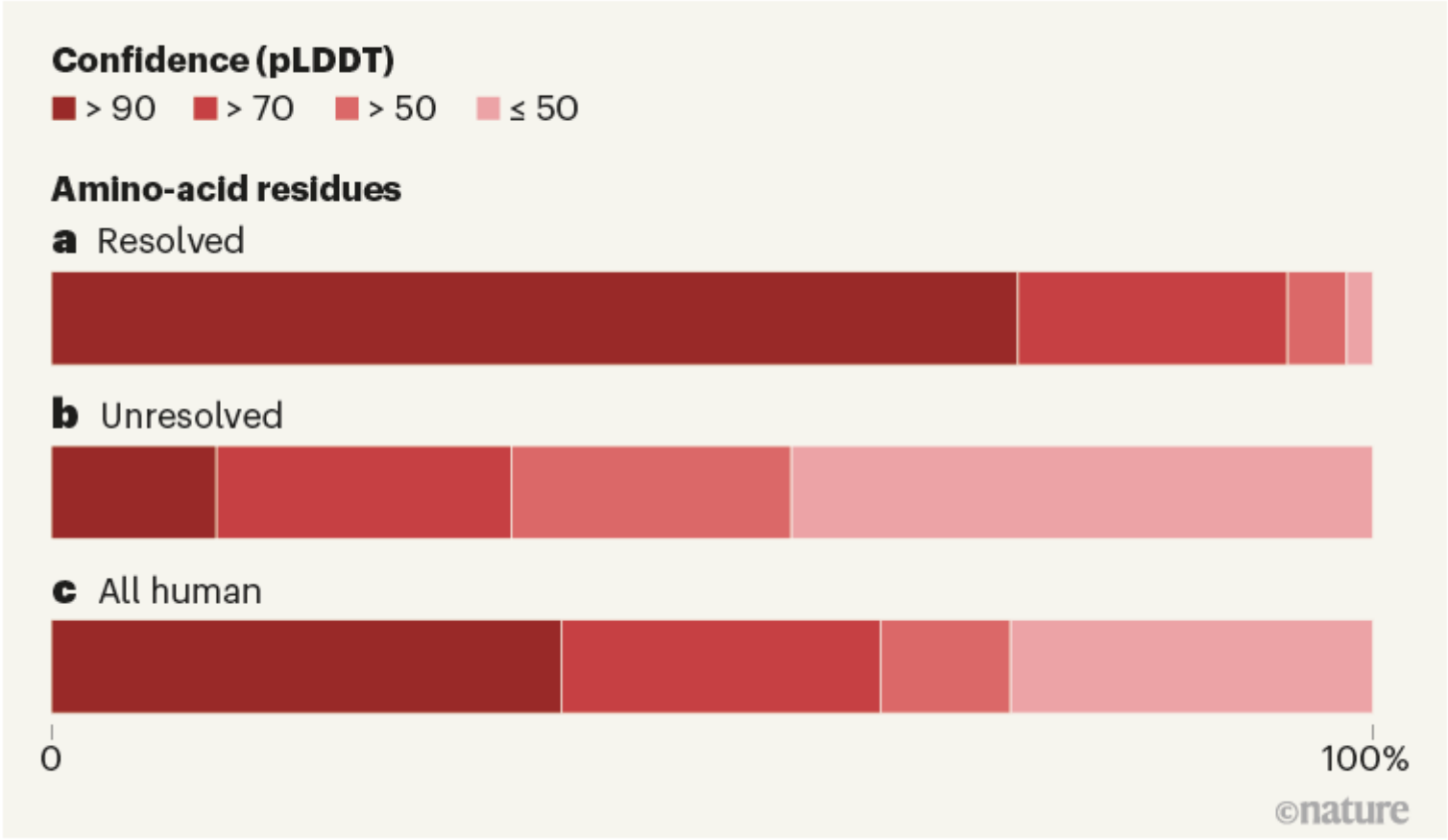
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.



Alphafold tells you where is it right!



How good are the predictions of human proteins?

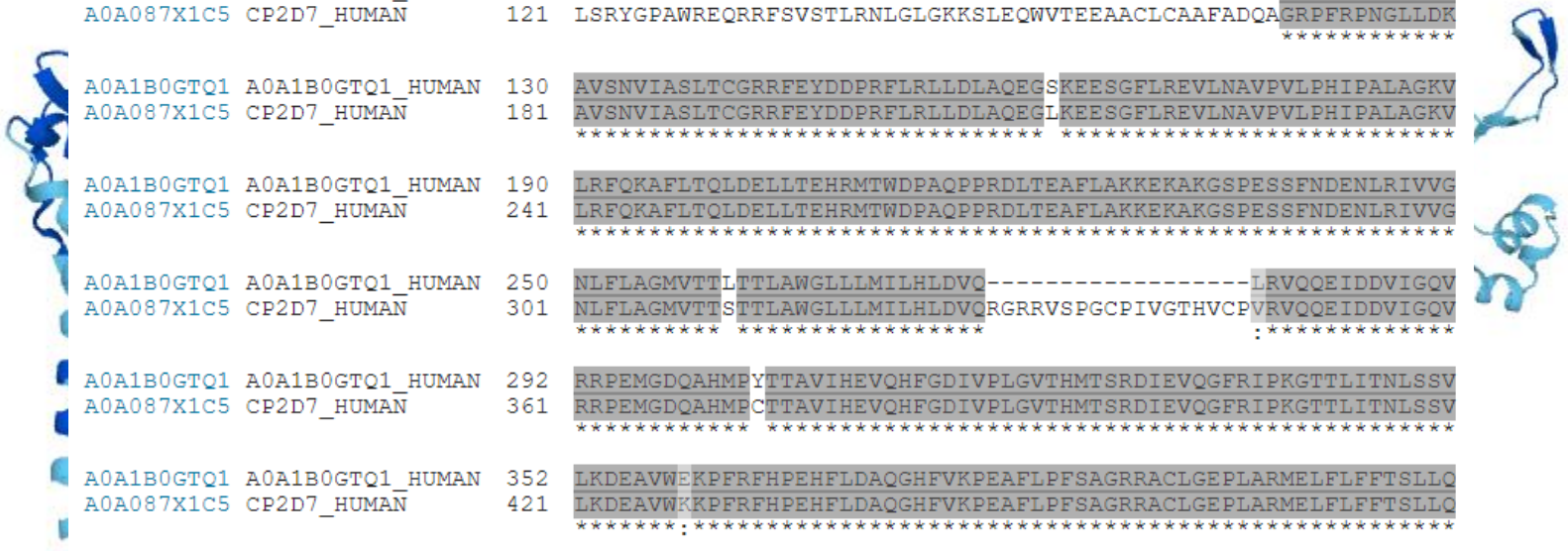


pLDDT - per-residue estimate of its confidence on a scale from 0 - 100
model's predicted score on the lDDT-Cα metric (local superposition-free score for comparing protein structures and models using distance difference tests).

But one still needs to be careful…

e.g. putative human cytochrome P450 2C7

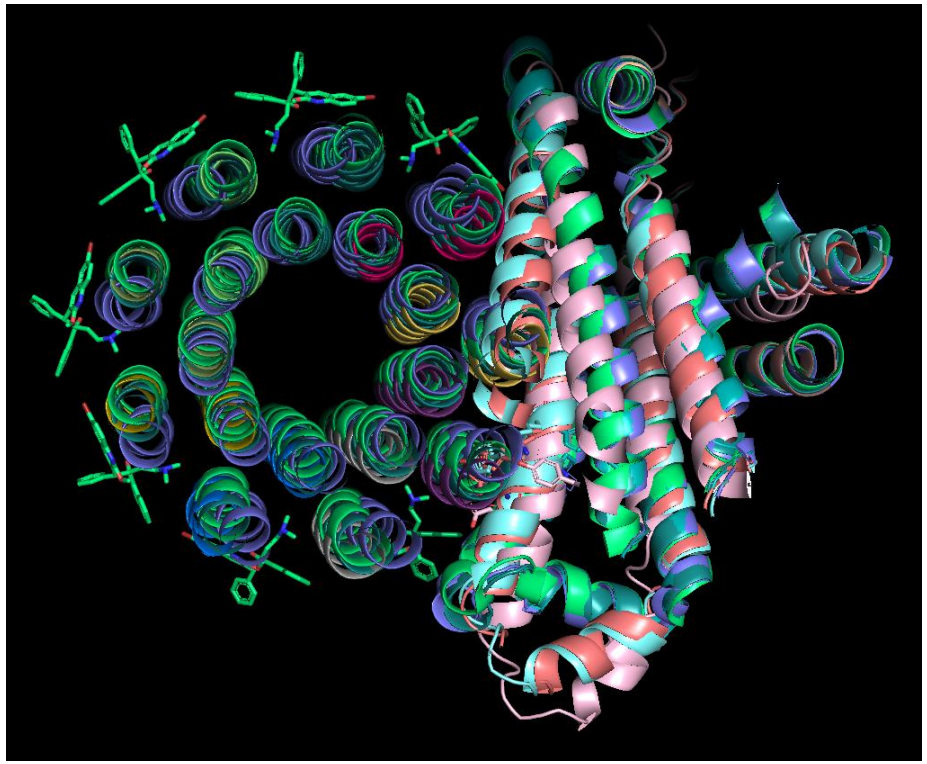
```
A0A1B0GTQ1 A0A1B0GTQ1_HUMAN    1  MGLEALVPLAMIVAIFLLLVDLMHRHQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ
A0A087X1C5 CP2D7_HUMAN         1  MGLEALVPLAMIVAIFLLLVDLMHRHQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ
                                  ************************************************************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN   61  LRRRFGDVFSLQLAWTPVVVLNGLAAVREAMVTRGEDTADRPPAPIYQVLGFGPRSQ---
A0A087X1C5 CP2D7_HUMAN        61  LRRRFGDVFSLQLAWTPVVVLNGLAAVREAMVTRGEDTADRPPAPIYQVLGFGPRSQGVI
                                  ************************************************************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  118  ----------------------------------------------GRPFRPNGLLDK
A0A087X1C5 CP2D7_HUMAN       121  LSRYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFADQAGRPFRPNGLLDK
                                                                                    ************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  130  AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGSKEESGFLREVLNAVPVLPHIPALAGKV
A0A087X1C5 CP2D7_HUMAN       181  AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLPHIPALAGKV
                                  ********************************* ****************************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  190  LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAKKEKAKGSPESSFNDENLRIVVG
A0A087X1C5 CP2D7_HUMAN       241  LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAKKEKAKGSPESSFNDENLRIVVG
                                  ************************************************************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  250  NLFLAGMVTTLTTLAWGLLLMILHLDVQ------------------LRVQQEIDDVIGQV
A0A087X1C5 CP2D7_HUMAN       301  NLFLAGMVTTSTTLAWGLLLMILHLDVQRGRRVSPGCPIVGTHVCPVRVQQEIDDVIGQV
                                  ********** ***************                      :**************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  292  RRPEMGDQAHMPYTTAVIHEVQHFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSV
A0A087X1C5 CP2D7_HUMAN       361  RRPEMGDQAHMPCTTAVIHEVQHFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSV
                                  ************ ************************************************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  352  LKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQ
A0A087X1C5 CP2D7_HUMAN       421  LKDEAVWKKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQ
                                  *******:****************************************************

A0A1B0GTQ1 A0A1B0GTQ1_HUMAN  412  HFSFSVAAGQPRPSHSRVVSFLVTPSPYELCAVPR
A0A087X1C5 CP2D7_HUMAN       481  HFSFSVAAGQPRPSHSRVVSFLVTPSPYELCAVPR
                                  ********************************
```



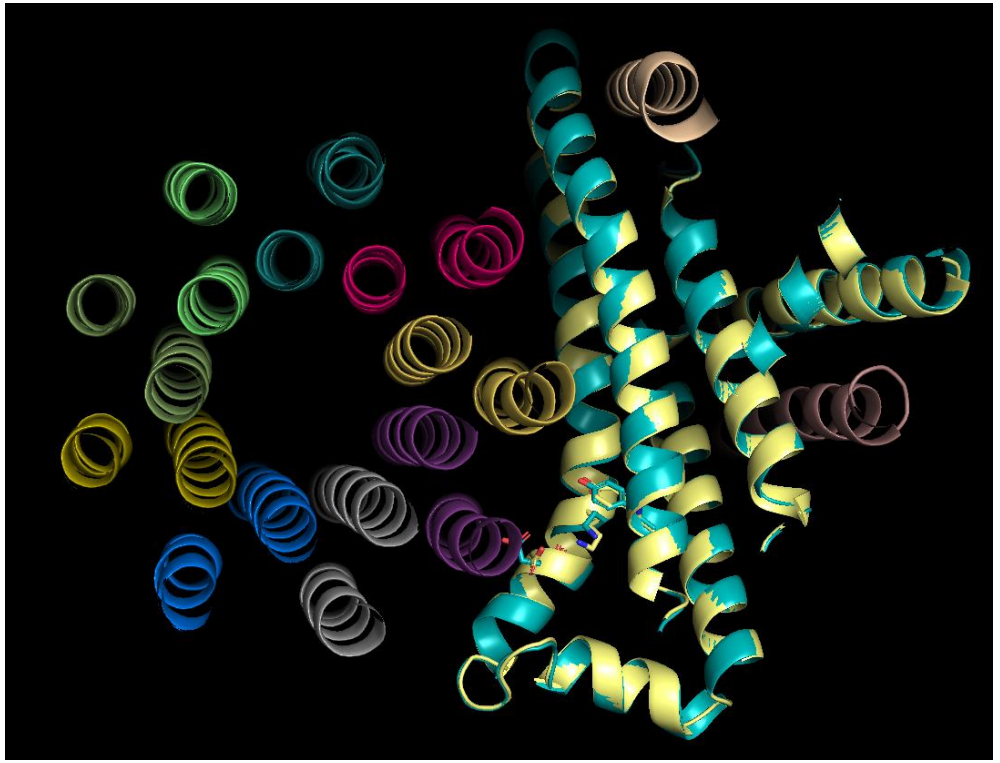
Structure can be only as good as its sequence

AlphaFoldDB contains monomers

- *Myc* ATPsynthase



AlphaFoldDB models for
P63655, P9WPV6, P9WPV7
over *M.smegmatis* ATPsynthase

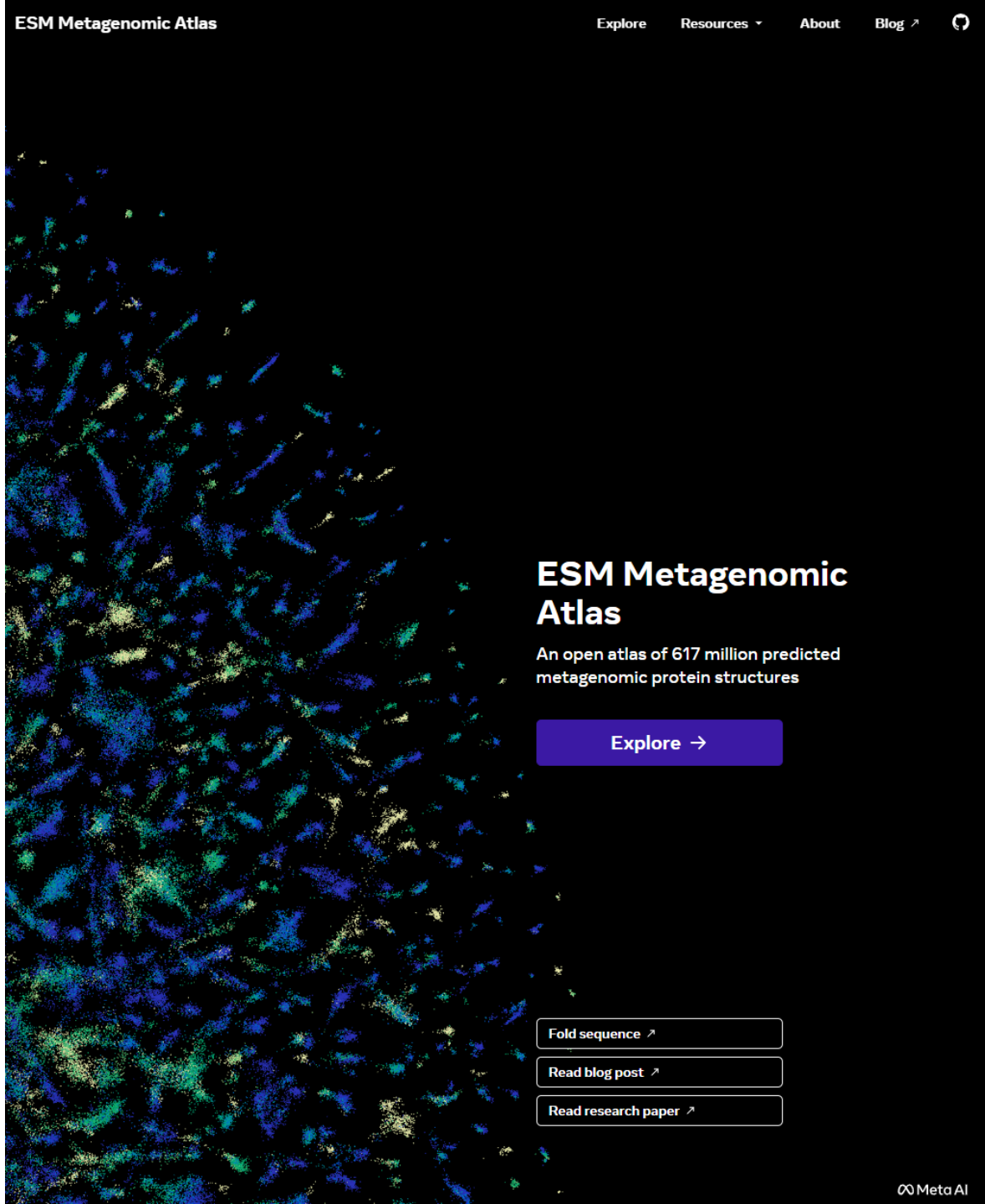


SwissModel
over *M.smegmatis* ATPsynthase
(PDBID: 7njp)

ESM Metagenomic Atlas

- Larger dataset of metagenomic proteins – 600M+ structures
- Quicker folding – using large language models (similar to ChatGPT-3)
- Enables folding on the fly
  - https://esmatlas.com/resources?action=fold
- Lower quality
- Esp. problem in domain decomposition

https://esmatlas.com/



Where to run AlphaFold

AlphaFold in Google Colab

Github enabled JupyterNotebooks running in Google Colab environment

limitation in size (timing)
start also from Chimera

Repozitář: sokrypton/ColabFold

Větev: main

Cesta

AlphaFold2.ipynb

AlphaFold2_complexes.ipynb

RoseTTAFold.ipynb

batch/AlphaFold2_batch.ipynb

Mirdita M, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. *bioRxiv*, 2021 https://doi.org/10.1101/2021.08.15.456425

https://colab.research.google.com/github/sokrypton/ColabFold/

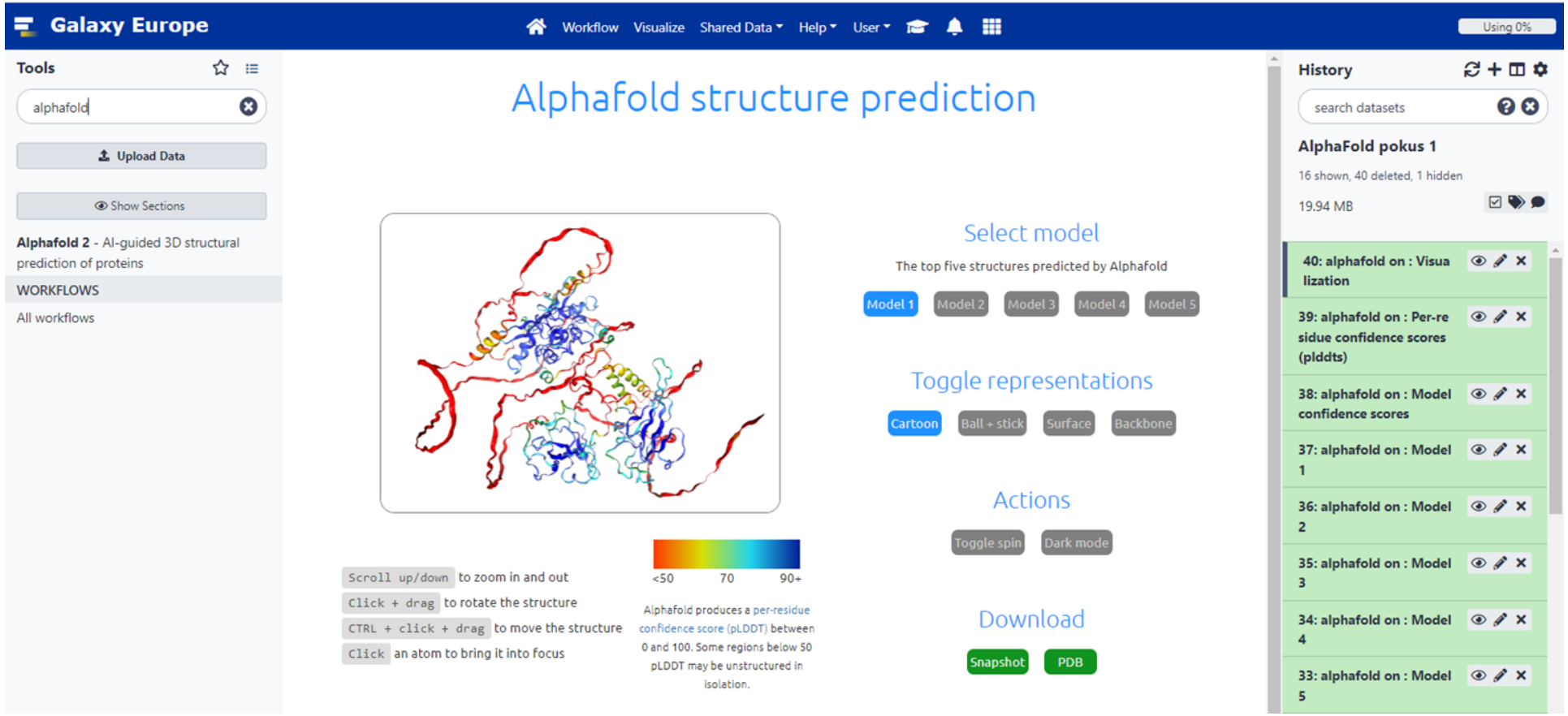
Alphafold 2 on ELIXIR CZ

- Alphafold “needs” GPU to run -> not many people have it on their PC
- Alphafold has been installed on Elixir CZ hardware
- Alphafold (Multimer) in the newest version 2.2.0 is accessible through Metacentrum
- speed is dependent on size of predicted protein (complex)

https://wiki.metacentrum.cz/wiki/AlphaFold

AlphaFold in UseGalaxy.eu

e.g. dimer Nucleocapsid protein from SARS-CoV-2



trick - dimerization fake as long disordered poly-N chain

https://usegalaxy.eu/tool_runner?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fgalaxy-australia%2Falphafold2%2Falphafold%2F2.1.2%2Bgalaxy0

Limitations

Alphafold is just a start...

- use Alphafold ideas for development of their own 3D structure predictions - RoseTTAfold
- prediction of designed proteins
- **tools** for molecular replacement
- **tools** for interpretation of cryoEM

...

As of 30.1.2023

elixir CZECH REPUBLIC

MrParse:

Finding homologues in the PDB and the EBI AlphaFold database for Molecular Replacement and more

Adam J. Simpkin, Jens M. H. Thomas, Ronan M. Keegan, Daniel J. Rigden

doi: https://doi.org/10.1101/2021.09.02.458604

Are structural biologists and bioinformaticians on the job market?

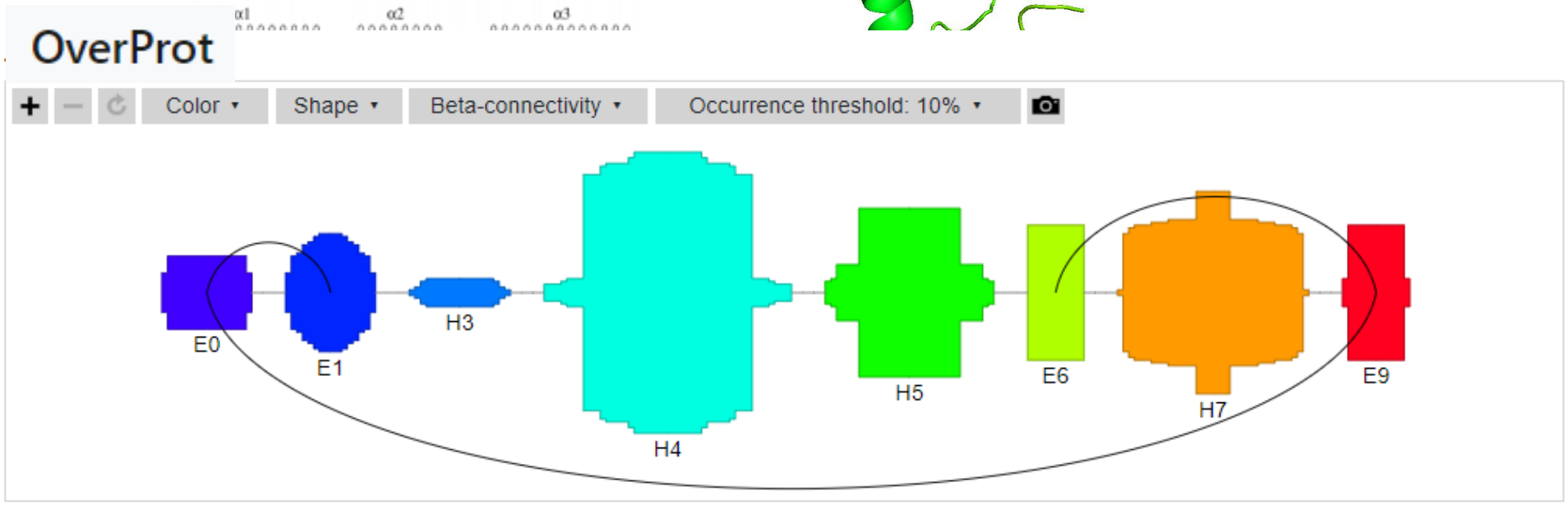
Not yet as:

- Alphafold can not do **point mutations** - design of functions
- Alphafold can not be used for **drug design**
- Alphafold can not do **conformational changes** or **dynamics**
- Alphafold can not do **multiprotein complexes** – interactions
- Alphafold can not do effects of **post-translational protein modifications**
- Alphafold can not do **ligand effects**
- Alphafold can not predict good **orphan sequences**
- Alphafold can not tell much about **folding process**
- **or can it?**

elixir CZECH REPUBLIC

Alphafold can do **point-mutations effects** Fold-switching proteins

GA77



https://overprot.ncbr.muni.cz/

- only one α helix is retained 100% (out of 50 structures),
- other α helices have minimal presence up to 72% (shorter in GB structures)
- β sheets presence change from 24% to 48%

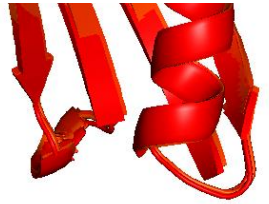
A minimal sequence code for switching protein structure and function

Patrick A. Alexander, Yanan He, Yihong Chen, +1, and Philip N. Bryan Authors Info & Affiliations

December 15, 2009 | 106 (50) 21149-21154 | https://doi.org/10.1073/pnas.0906408106

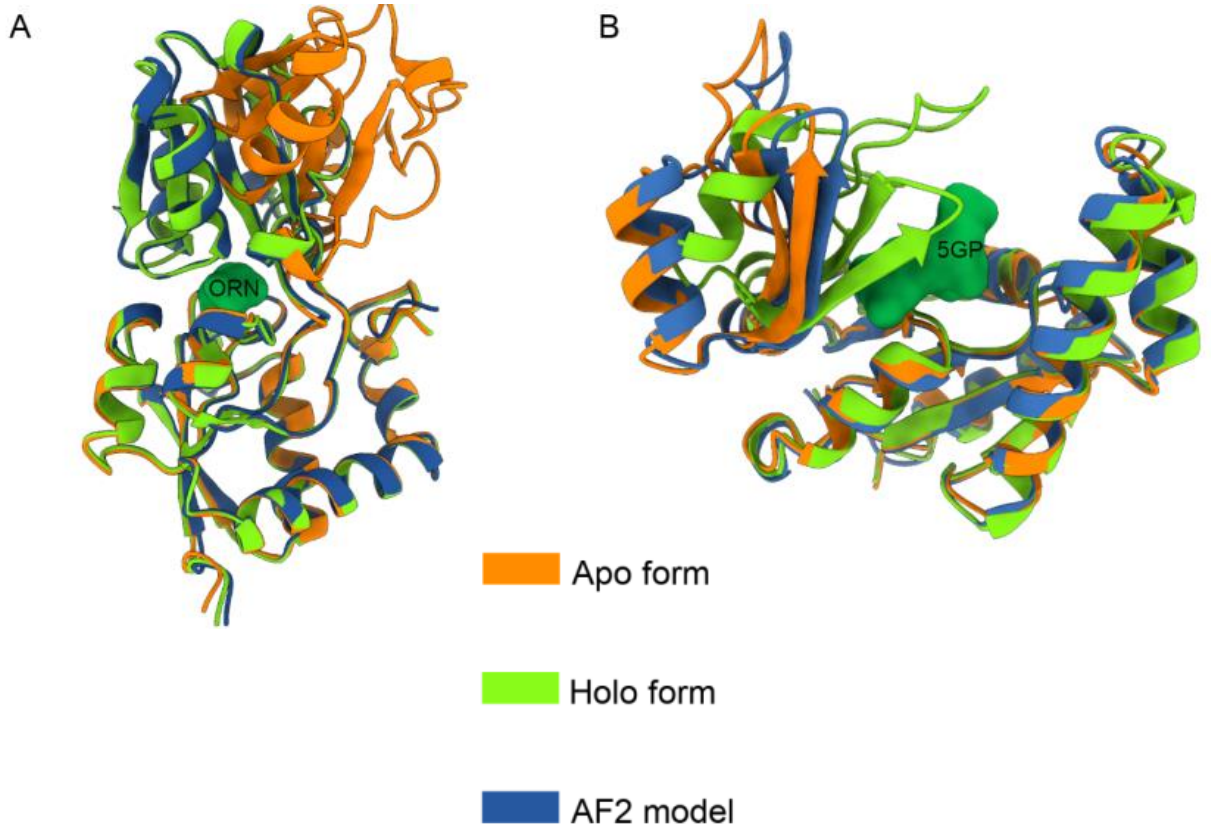
GB98 models shows mix between
3α to α+β

own calculations



AlphaFold for **drug design**?

- AlphaFold2 predicts **holo** protein in 70% => it can be used for drug designing
- pLDDT values in a single 3D model could be used to infer local conformational change linked to ligand binding transitions.
- locally AlphaFold2 can be there -
  but it needs validation
  (as always)



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

Impact of protein conformational diversity on AlphaFold predictions

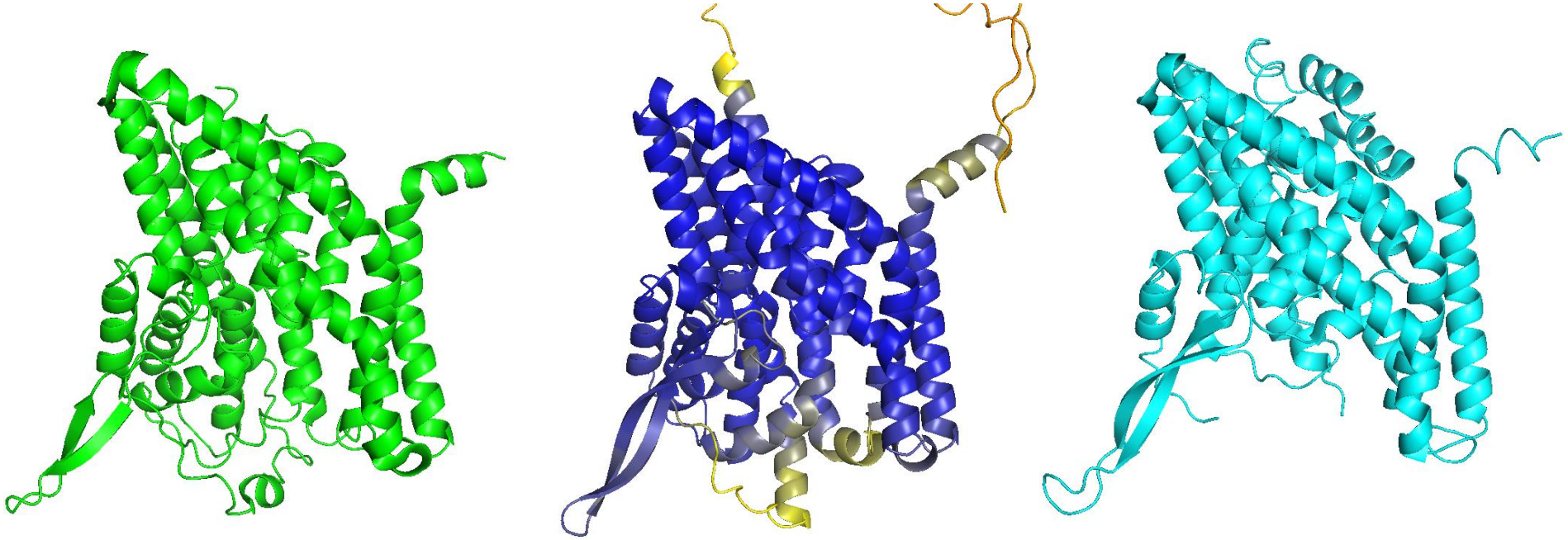
Tadeo Saldaño, Nahuel Escobedo, Julia Marchetti, Diego Javier Zea, Juan Mac Donagh, Ana Julia Velez Rueda, Eduardo Gonik, Agustina García Melani, Julieta Novomisky Nechcoff, Martín N. Salas, Tomás Peters, Nicolás Demitroff, Sebastian Fernandez Alberti, Nicolas Palopoli, Maria Silvina Fornasari, Gustavo Parisi

doi: https://doi.org/10.1101/2021.10.27.466189

Alphafold can predict **dynamics**

pLDDT shows flexibility

SLC1A5



6mp6
Outward-Facing

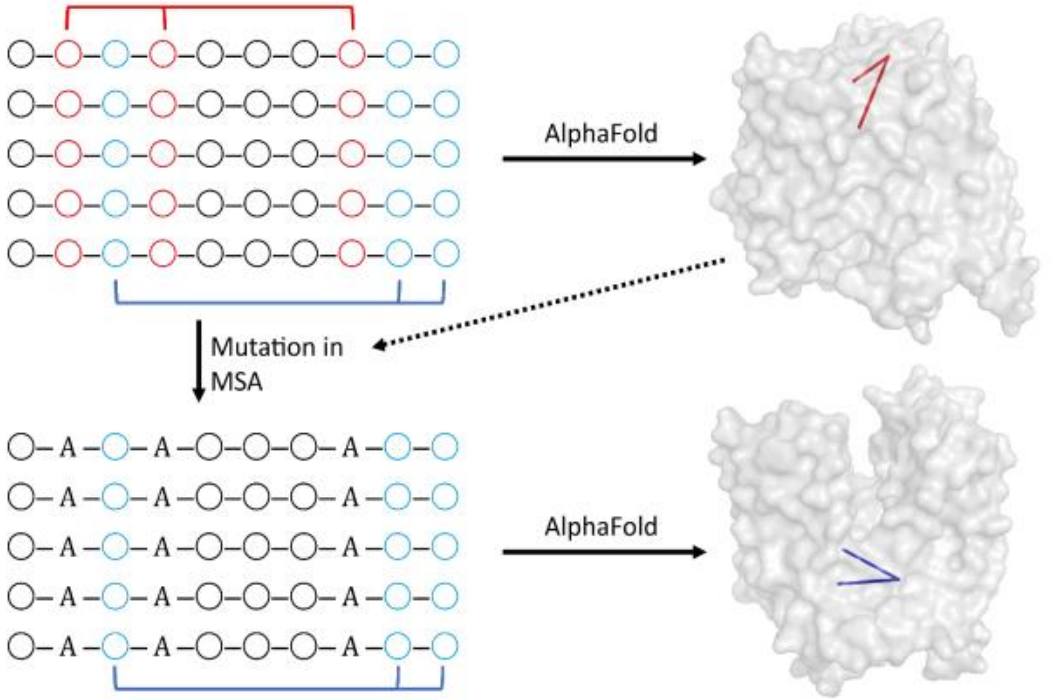
AlphaFoldDB
similar to OF

6rvx
Inward-Facing

lower pIDDT values show flexible regions

Alphafold can do **conformational changes**

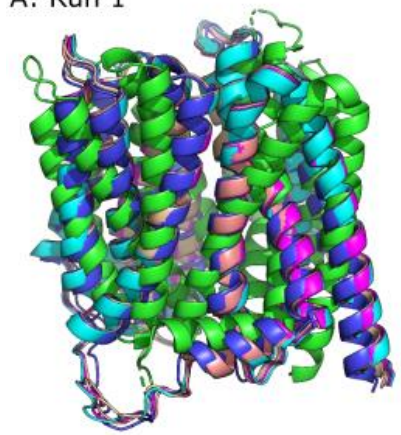
manipulation with MSA allows selection of multiple conformers via mutation of contact points in MSA



LmrP transporter

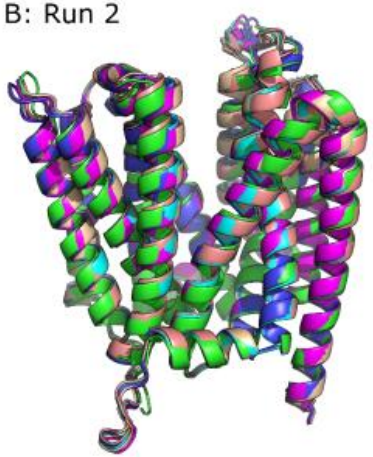
default

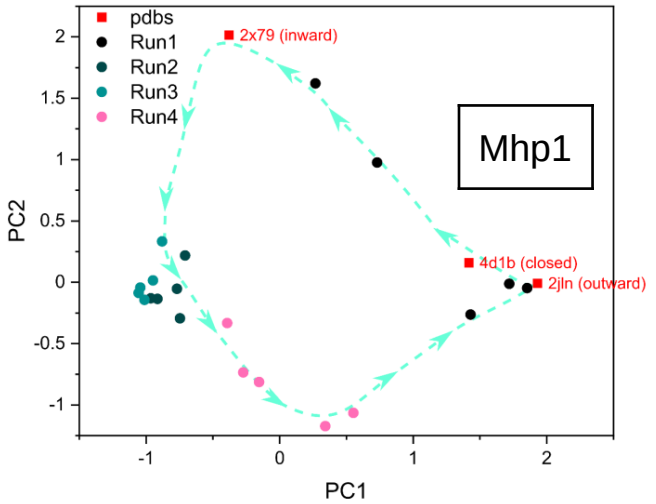
A: Run 1



after mutation on interface

B: Run 2





Modeling Alternate Conformations with Alphafold2 via Modification of the Multiple Sequence Alignment

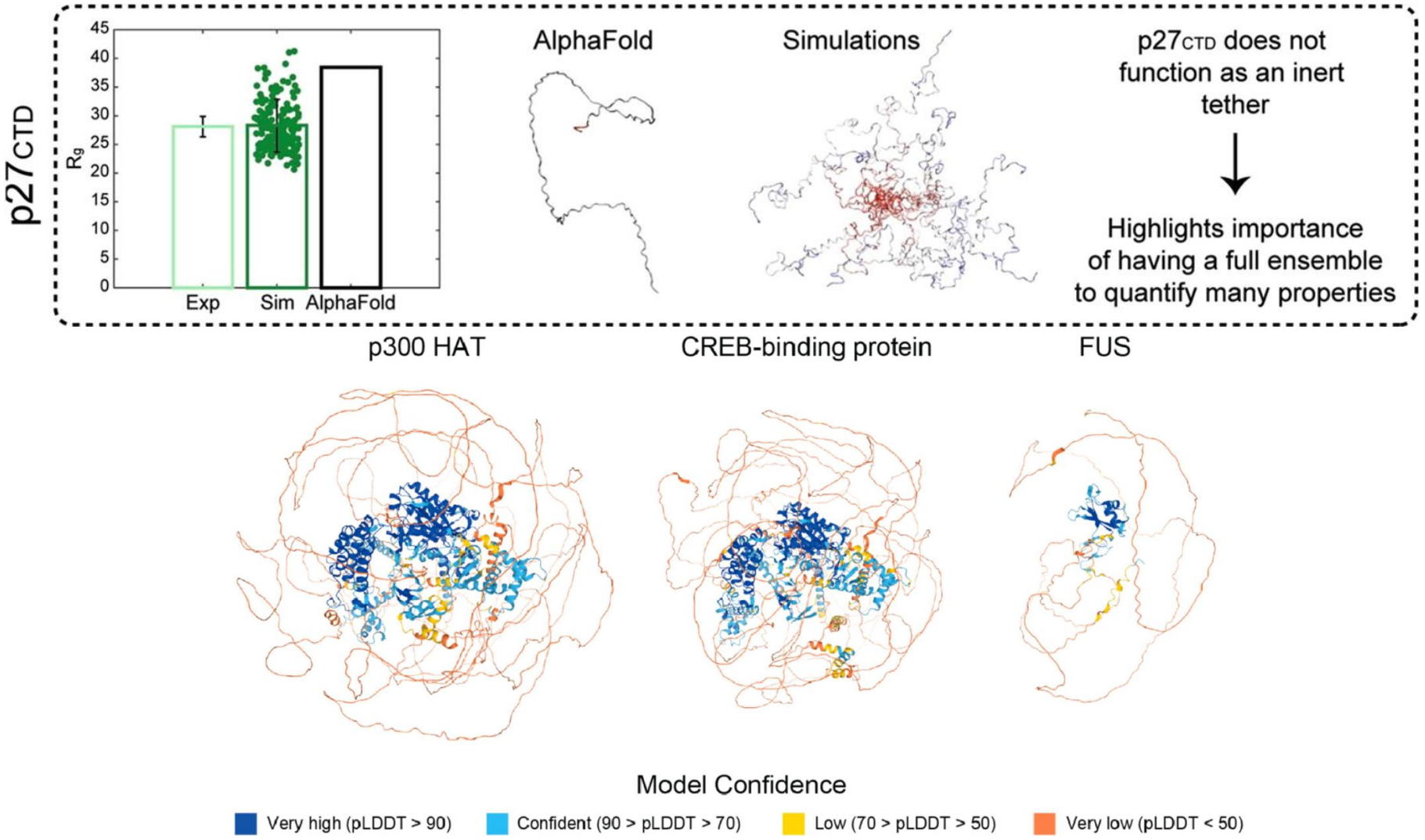
Richard A. Stein, Hassane S. Mchaourab

doi: https://doi.org/10.1101/2021.11.29.470469

bioRχiv

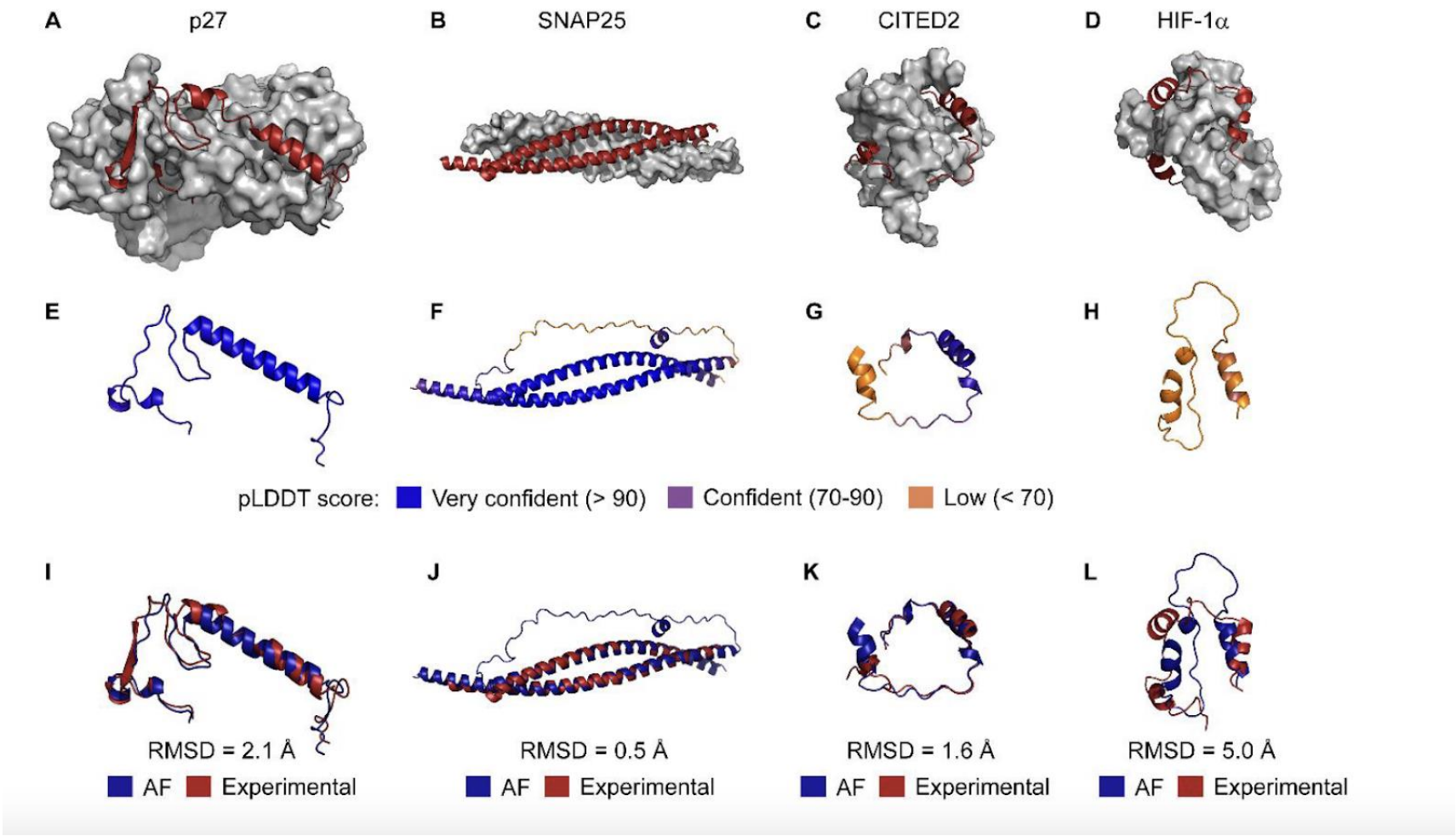
THE PREPRINT SERVER FOR BIOLOGY

AlphaFold and Intrinsically Disordered Proteins



Ruff KM, Pappu RV ,AlphaFold and Implications for Intrinsically Disordered Proteins, J Mol Biol. 2021, https://doi.org/10.1016/j.jmb.2021.167208

AlphaFold and **Intrinsically Disordered Proteins**

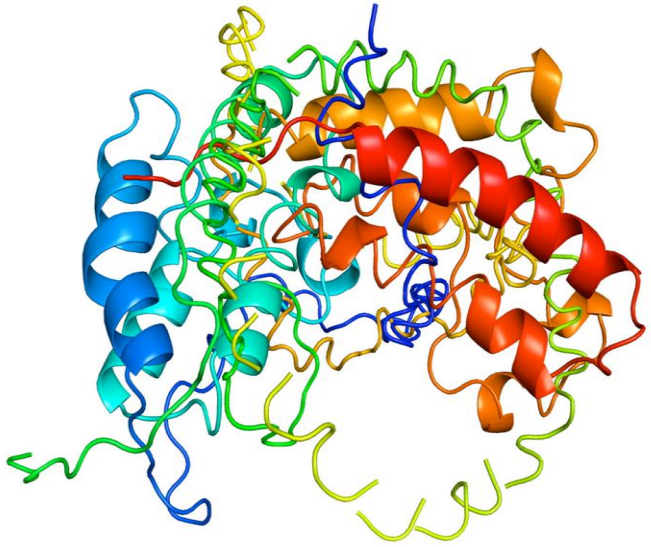


Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2

T. Reid Alderson, Iva Pritišanac, Alan M. Moses, Julie D. Forman-Kay

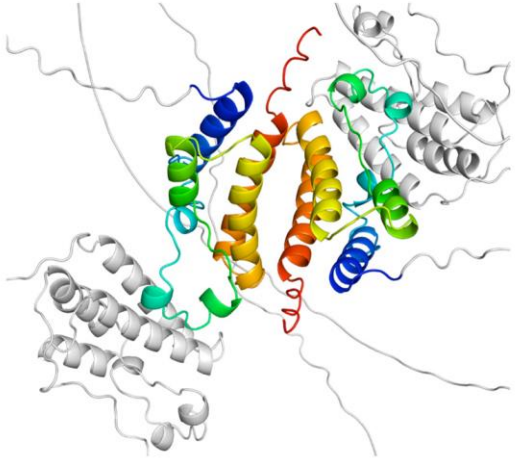
doi: https://doi.org/10.1101/2022.02.18.481080

Alphafold can do **multiprotein complexes**

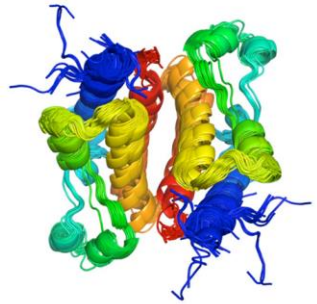


Alphafold-Multimer v2 reproduces dimer of Bromodomains BD2 of BET proteins observed in crystal structures

AF2mult_v2 homodimer of BRD2_HUMAN Bromodomain B2 in rainbow; BD1 in gray



ProtCID cluster of dimers of BD2 domains of human BRD2, BRD3, BRD4, mouse BRDT
http://dunbrack2.fccc.edu/protcid



https://twitter.com/RolandDunbrack/status/1502818748868317188

DeepMind

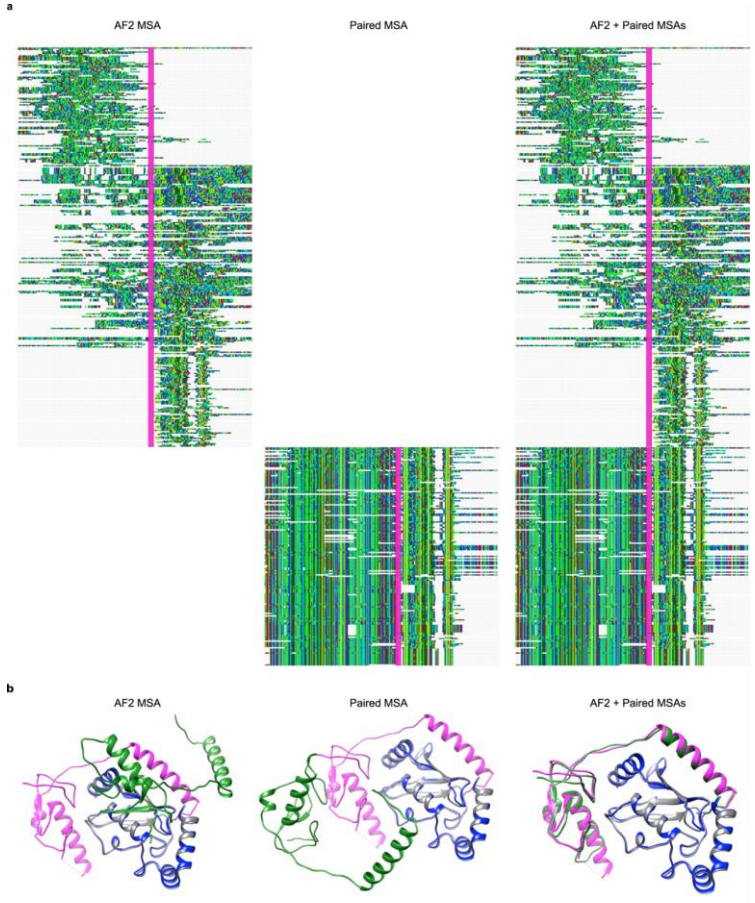
2022-03-10

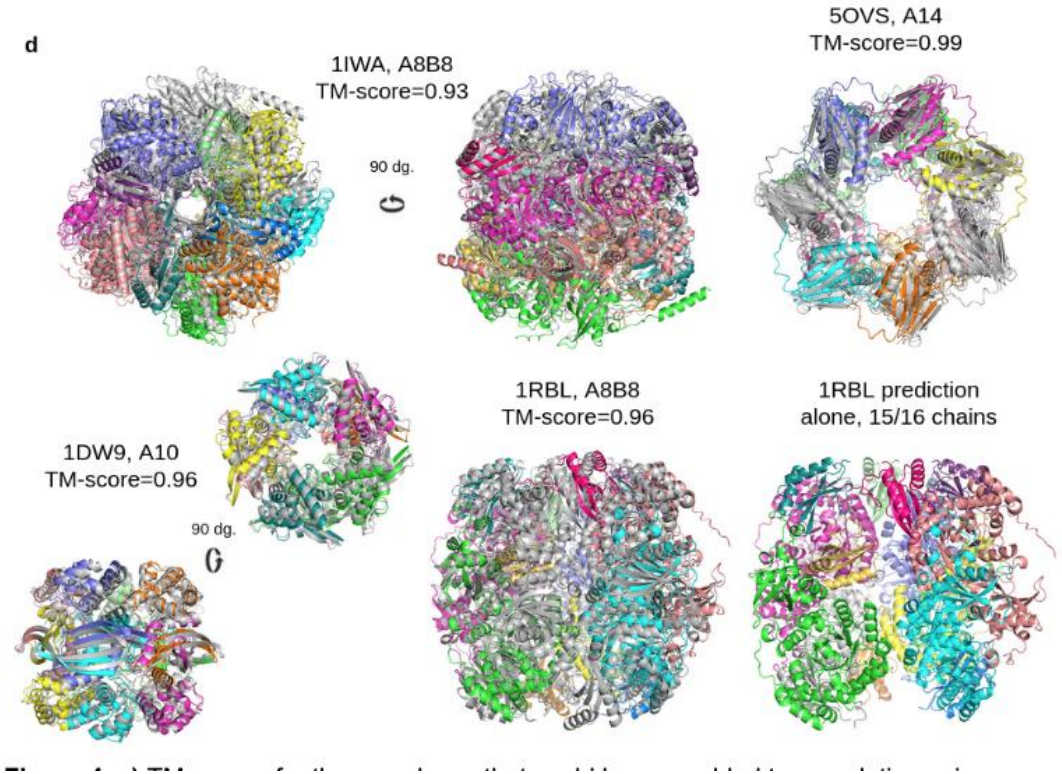
Protein complex prediction with AlphaFold-Multimer

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

Alphafold can do **multiprotein complexes**

Alphafold 2.2.0 - multimer





Article | Open Access | Published: 10 March 2022

Improved prediction of protein-protein interactions using AlphaFold2

Patrick Bryant, Gabriele Pozzati & Arne Elofsson

Nature Communications **13**, Article number: 1265 (2022) | Cite this article

6092 Accesses | **27** Altmetric | Metrics

New Results

Follow this preprint

Predicting the structure of large protein complexes using AlphaFold and sequential assembly

Patrick Bryant, Gabriele Pozzati, Wensi Zhu, Aditi Shenoy, Petras Kundrotas, Arne Elofsson

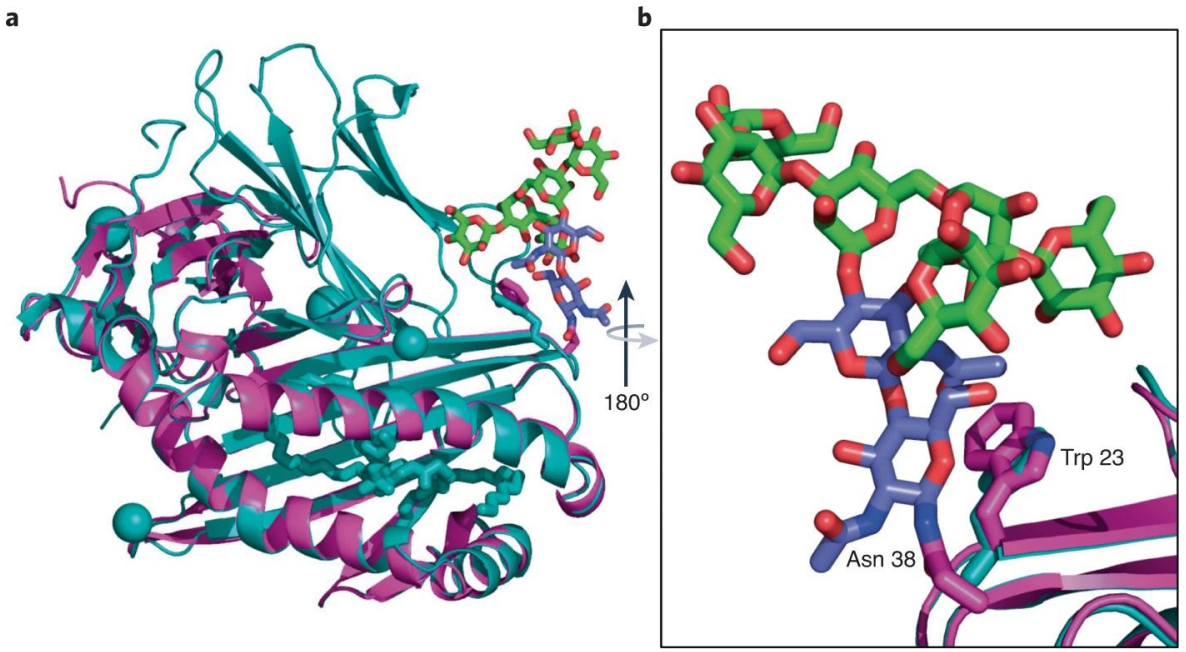
doi: https://doi.org/10.1101/2022.03.12.484089

This article is a preprint and has not been certified by peer review [what does this mean?].

Alphafold can not do effects of

post-translational protein modifications

(by itself)



Correspondence | Published: 29 October 2021

The case for post-predictional modifications in the AlphaFold Protein Structure Database

Haroldas Bagdonas, Carl A. Fogarty, Elisa Fadda & Jon Agirre

Nature Structural & Molecular Biology **28**, 869–870 (2021) | Cite this article

Alphafold can be filled with **ligands**

αfill

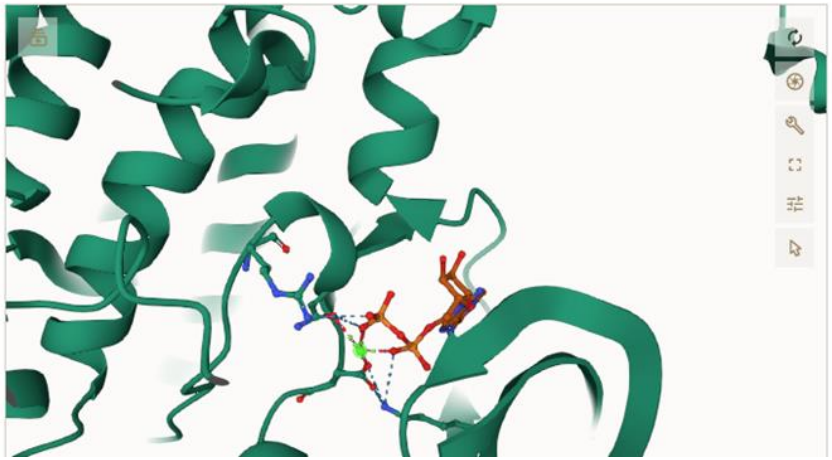
NKI Research | Biochemistry | Perrakis group

Home Structures Compounds Model About Download

P12931

Proto-oncogene tyrosine-protein kinase Src

Structure file	https://alphafill.eu/v1/aff/P12931
Metadata	https://alphafill.eu/v1/aff/P12931/json
Original AlphaFold model	https://alphafold.ebi.ac.uk/entry/P12931



35% identity | 40% identity | 50% identity | 60% identity | 70% identity

Compound	PDB-ID	Global RMSd	Asym	Local RMSd	Show
ADP	6f3f.A	1.54	B	0.45	☑
AGS -> ATP	3dqw.A	6.78 ?	I	1.38 ?	☐
AMP	3dqx.A	6.02 ?	H	0.57	☐
MG	6f3f.A	1.54	C	0.10	☑

New Results

Follow this preprint

AlphaFill: enriching the AlphaFold models with ligands and co-factors

Maarten L. Hekkelman, Ida de Vries, Robbie P. Joosten, Anastassis Perrakis

doi: https://doi.org/10.1101/2021.11.26.470110

Alphafold can describe **folding process** to some level

Table 2.
Performance of the structure predictors at identifying the secondary structure interactions present in an intermediate

	RoseTTA Fold	trRosetta	RaptorX	DMPfold	EVfold	SAINT2	Rosetta	Random
200 Decoys								
Accuracy	0.453	0.534	0.495	0.489	**0.540**	—	—	0.502
F1-score	0.222	0.169	0.110	0.026	**0.307**	—	—	0.252
Jaccard	0.052	0.052	0.052	0.052	0.052	—	—	**0.094**
AUROC	0.441	0.503	0.502	0.492	**0.530**	—	—	0.498

https://doi.org/10.1093/bioinformatics/btab881

Alphafold can describe **folding process** to some level

Was Anfinsen right?

New Results

Follow this preprint

Previous

Posted March 24, 2022.

Download PDF

Print/Save Options

Data/Code

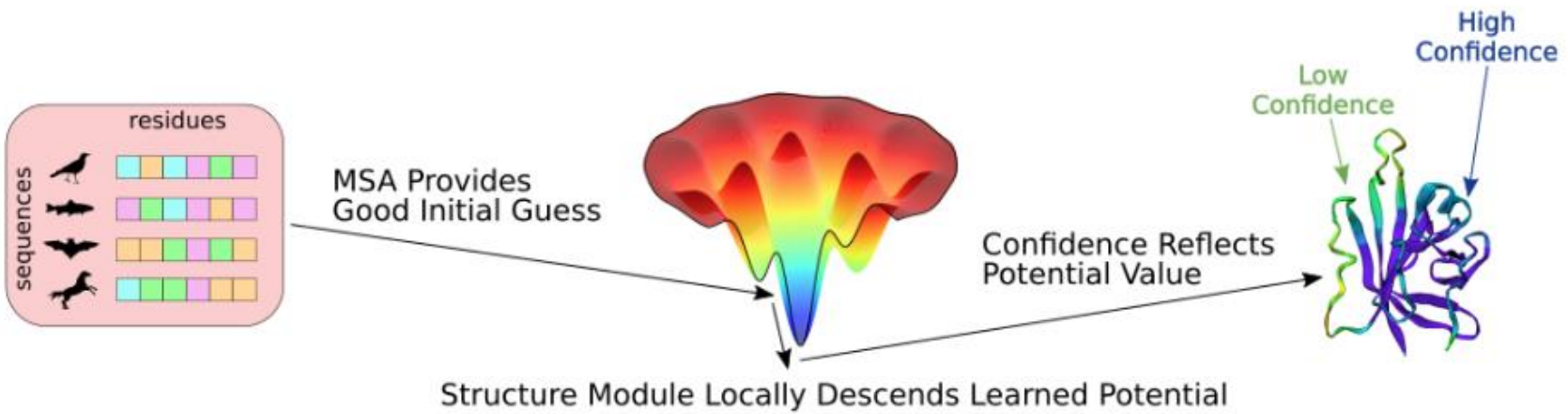
Revision Summary

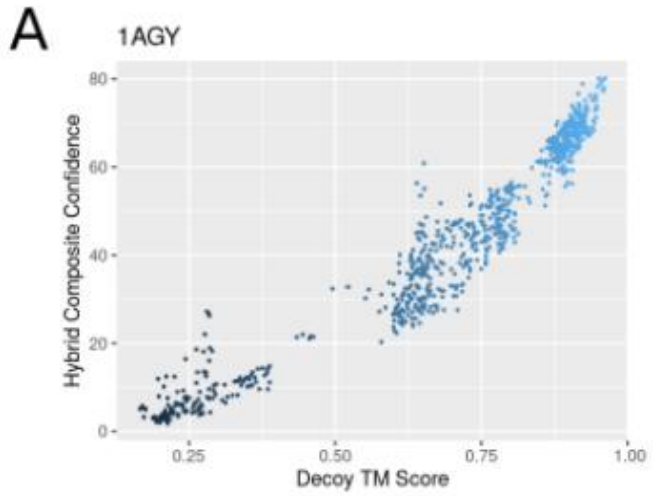
State-of-the-Art Estimation of Protein Model Accuracy using AlphaFold

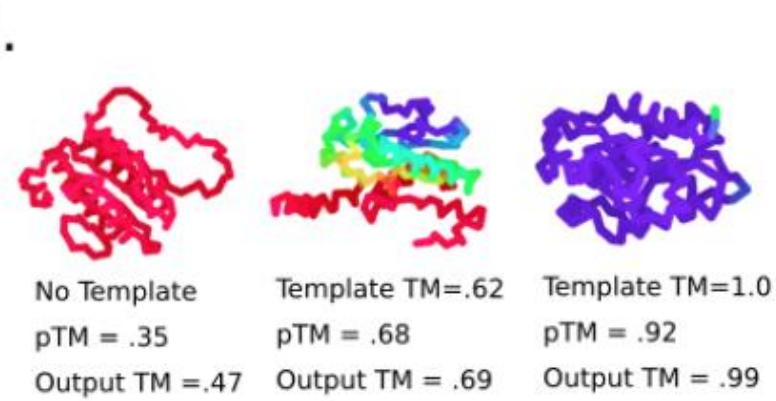
James P. Roney, Sergey Ovchinnikov

doi: https://doi.org/10.1101/2022.03.11.484043

This article is a preprint and has not been certified by peer review [what does this mean?].



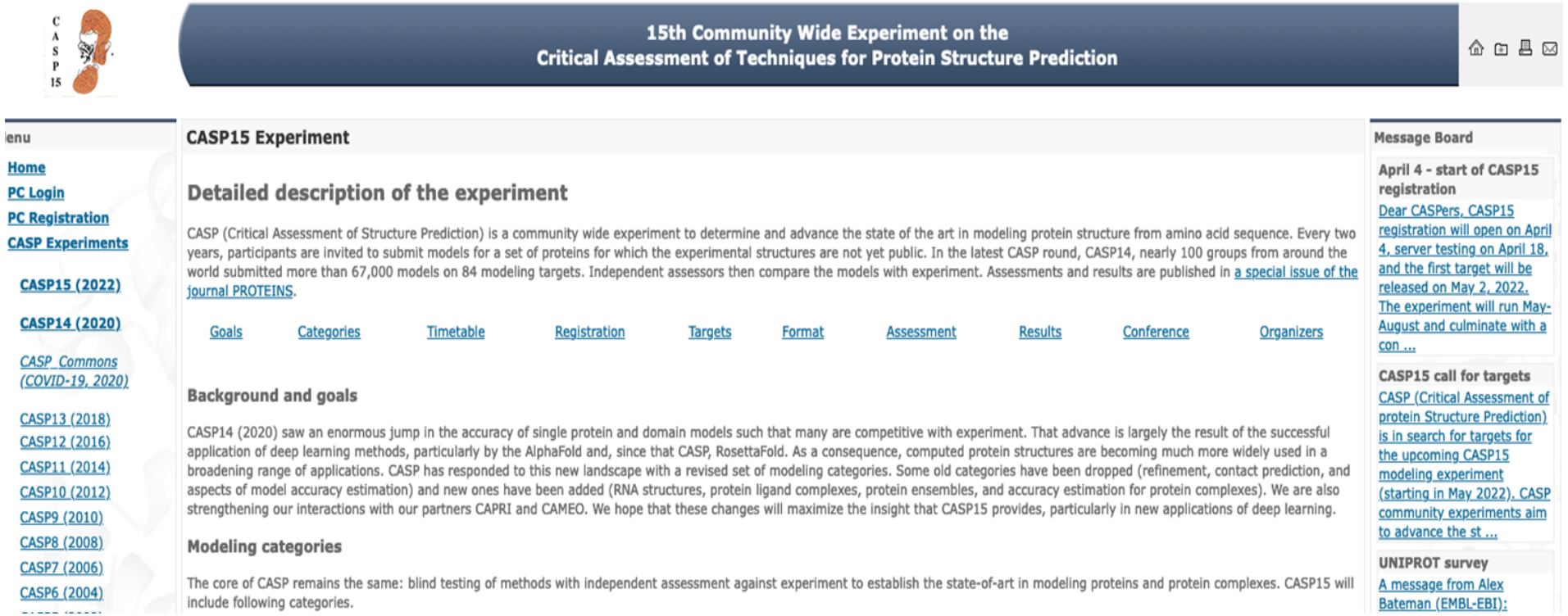




Summary

- Alphafold2 made a huge leap in **prediction accuracy**
- Role of **open science and publicly available data** can not be overstated
- **CASP competition** was a driver of the change
- Alphafold is **publicly available** and can be run from many places including ELIXIR CZ
- Alphafold has **inspired many tools and uses** already
- Alphafold **limits** are yet to be fully described, but we learning more each day
- Alphafold is **important tool** to structural biologist/ bioinformatician toolbox enhancing our capabilities

CASP is back!



elixir CZECH REPUBLIC

Thank you for your attention.

Any questions?

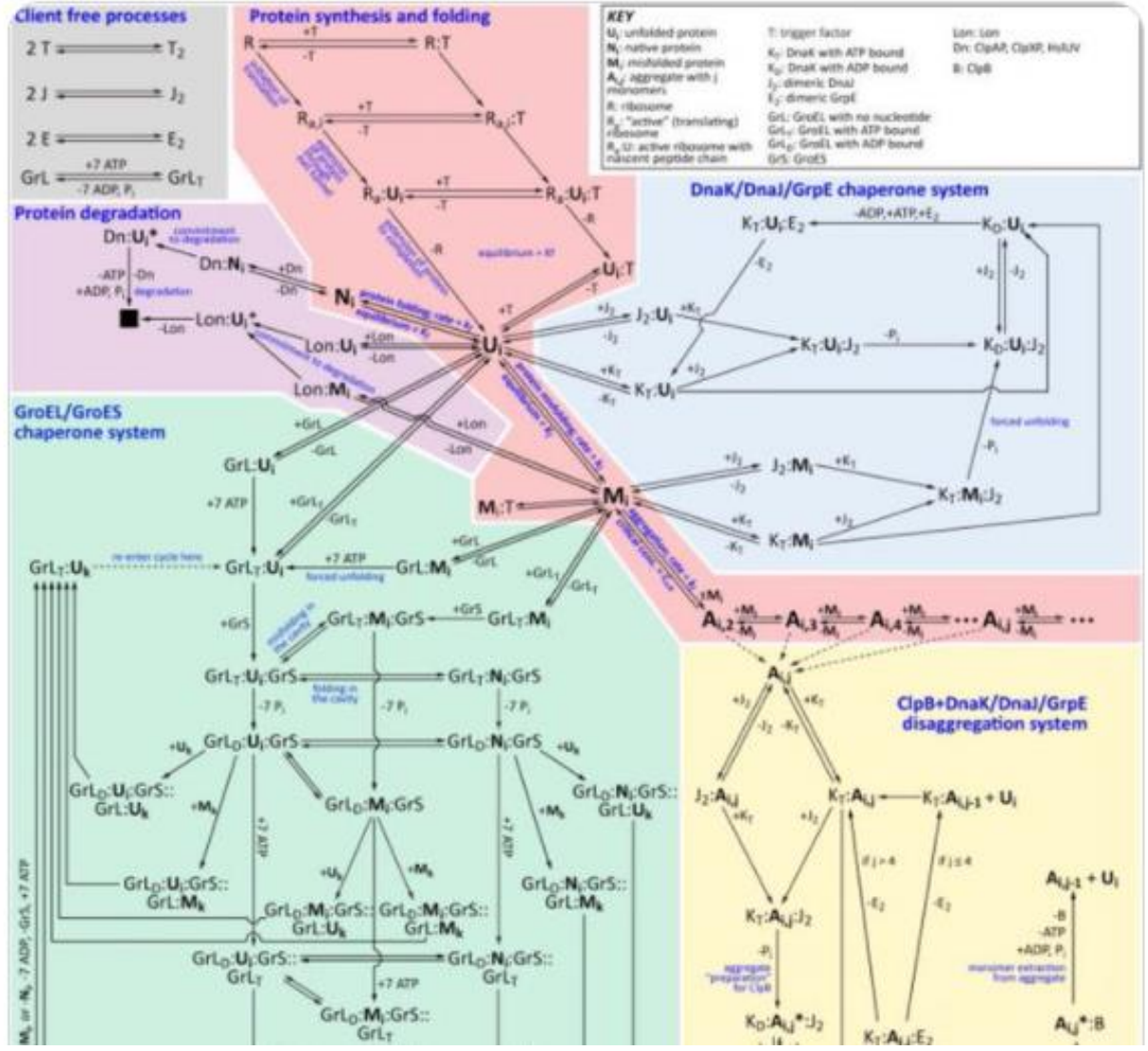
MINISTRY OF EDUCATION, YOUTH AND SPORTS

CZECH REPUBLIC

Kresten Lindorff-Larsen
@LindorffLarsen

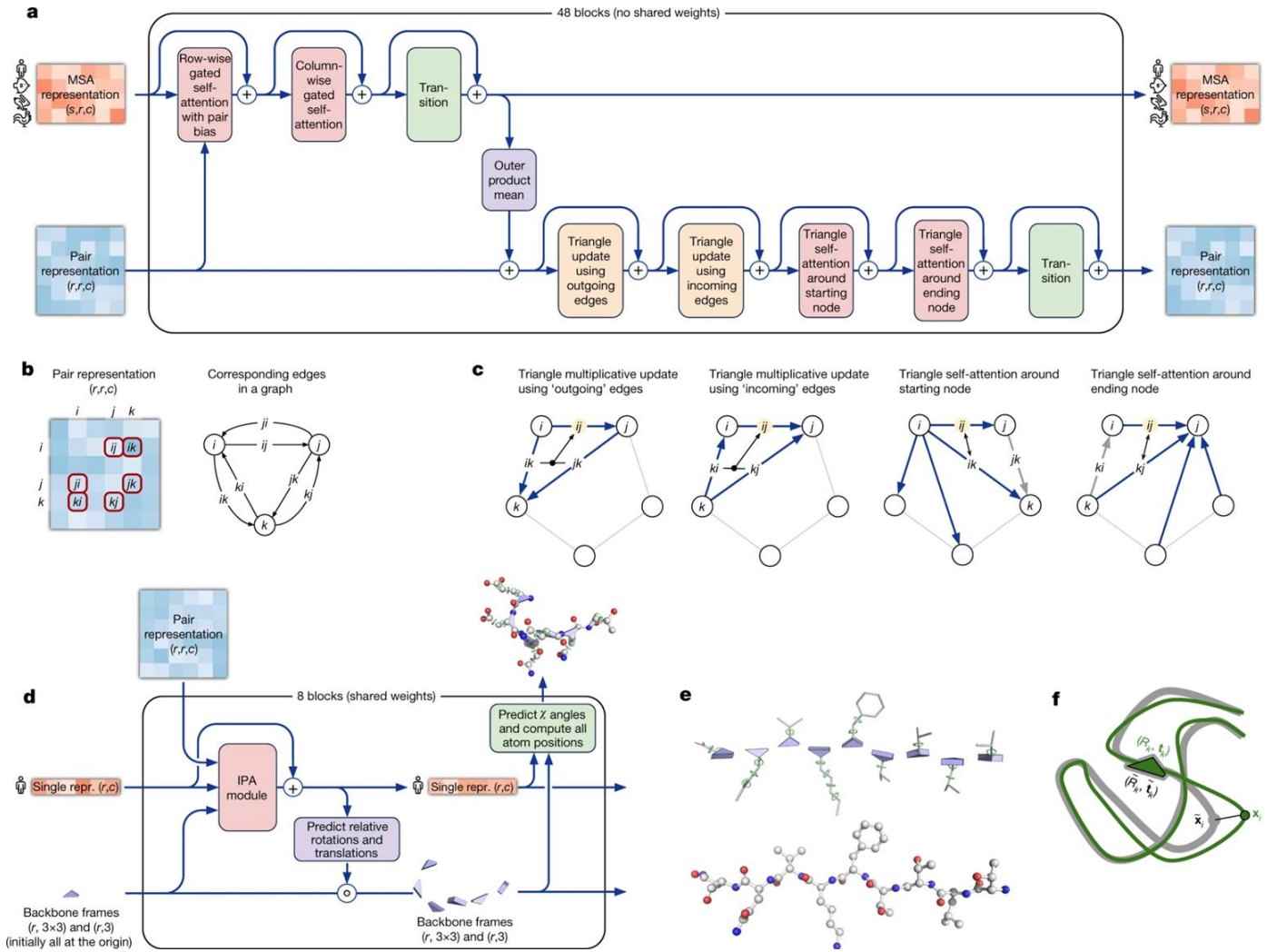
Tell me again how the folding problem has been solved
doi.org/10.1016/j.jmb.... doi.org/10.1016/j.celr...

Přeložit Tweet

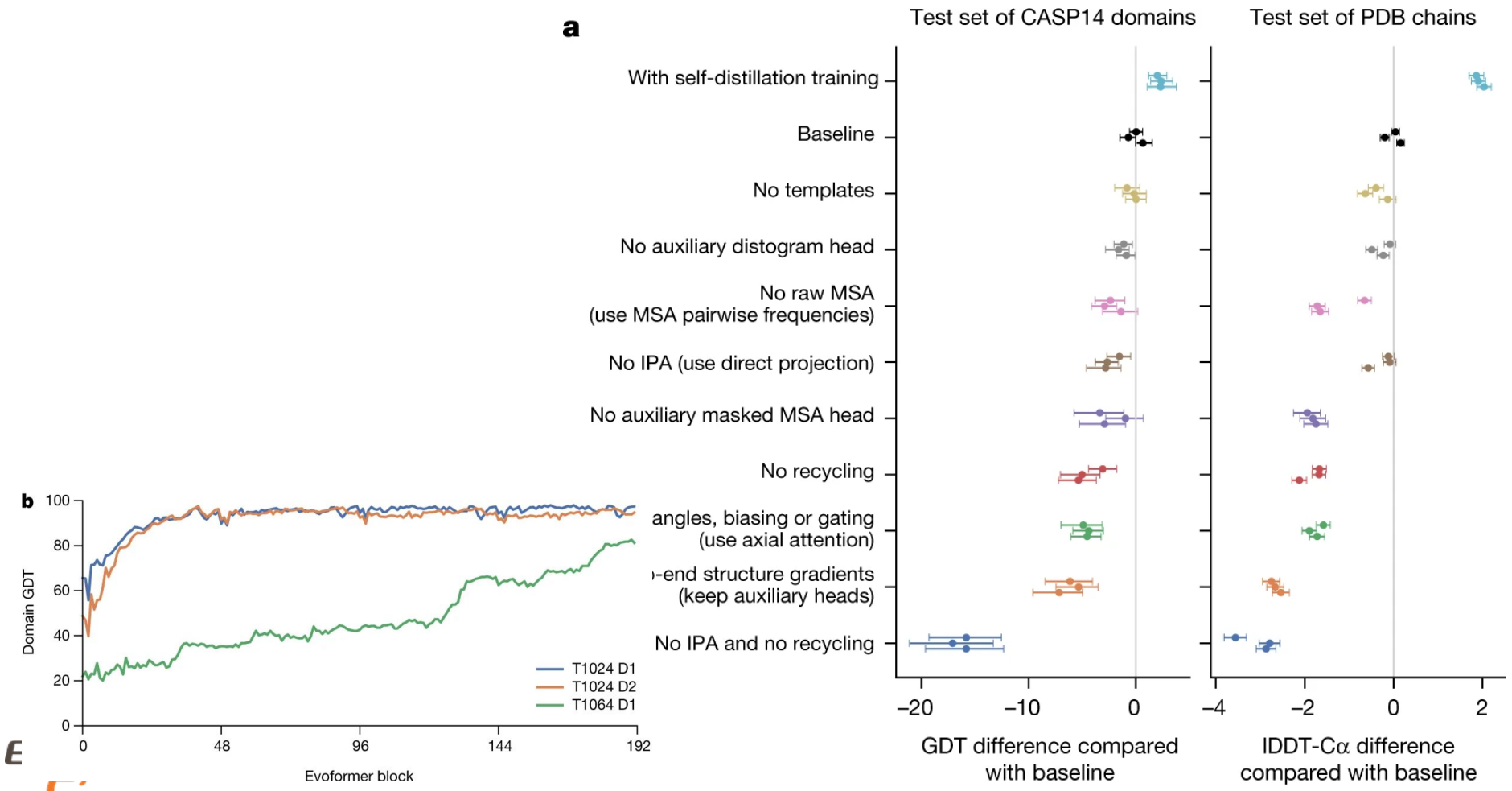


Extra slides

Architectural details.



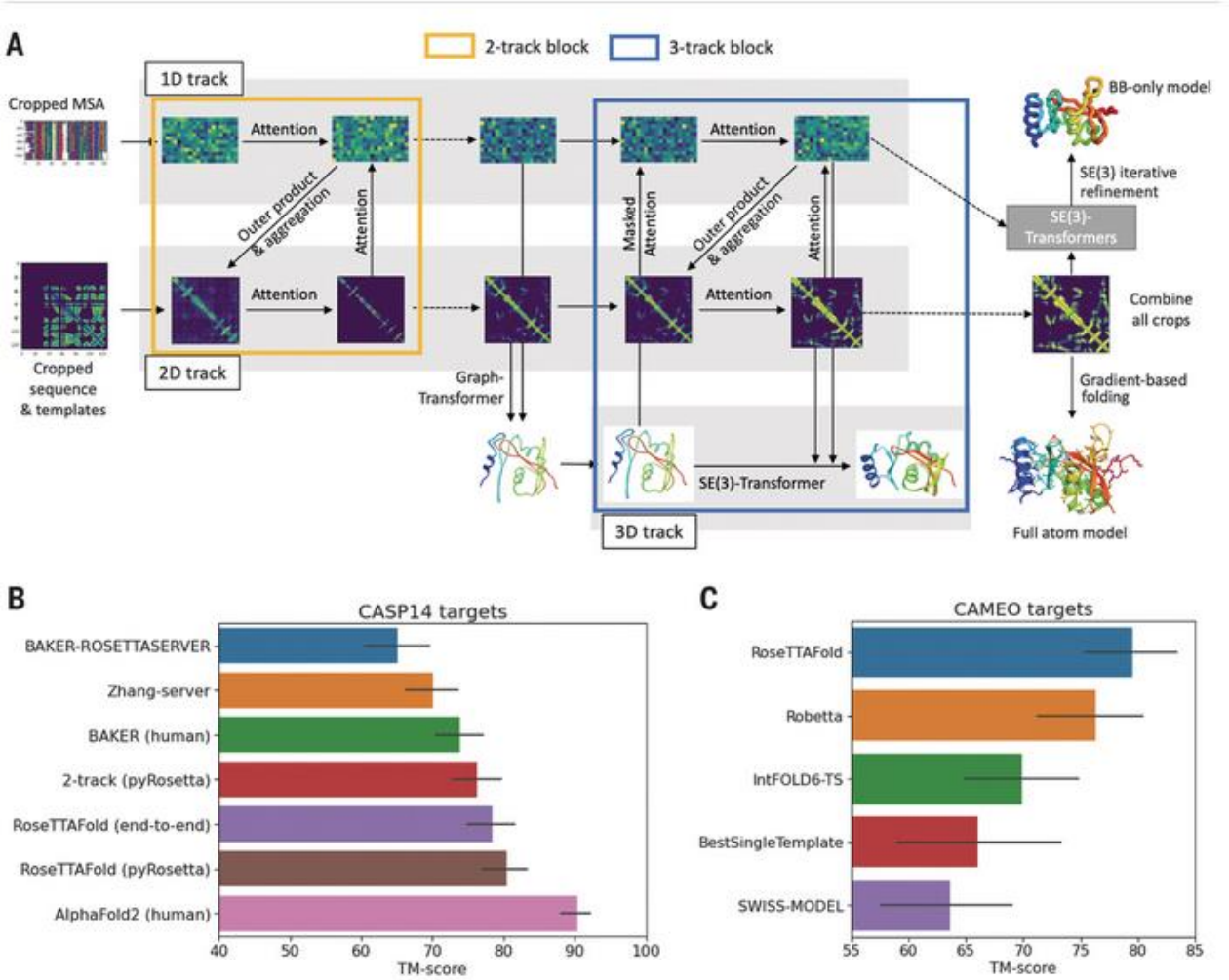
Interpreting the neural network



depth of neural network - it is usually quick, but for challenging targets it can be quite deep

https://www.nature.com/articles/s41586-021-03819-2

Accurate prediction of protein structures and interactions using a three-track neural network



https://www.science.org/doi/full/10.1126/science.abj8754

Alphafold can work with **orphan sequences**

Single-sequence protein structure prediction using language models from deep learning

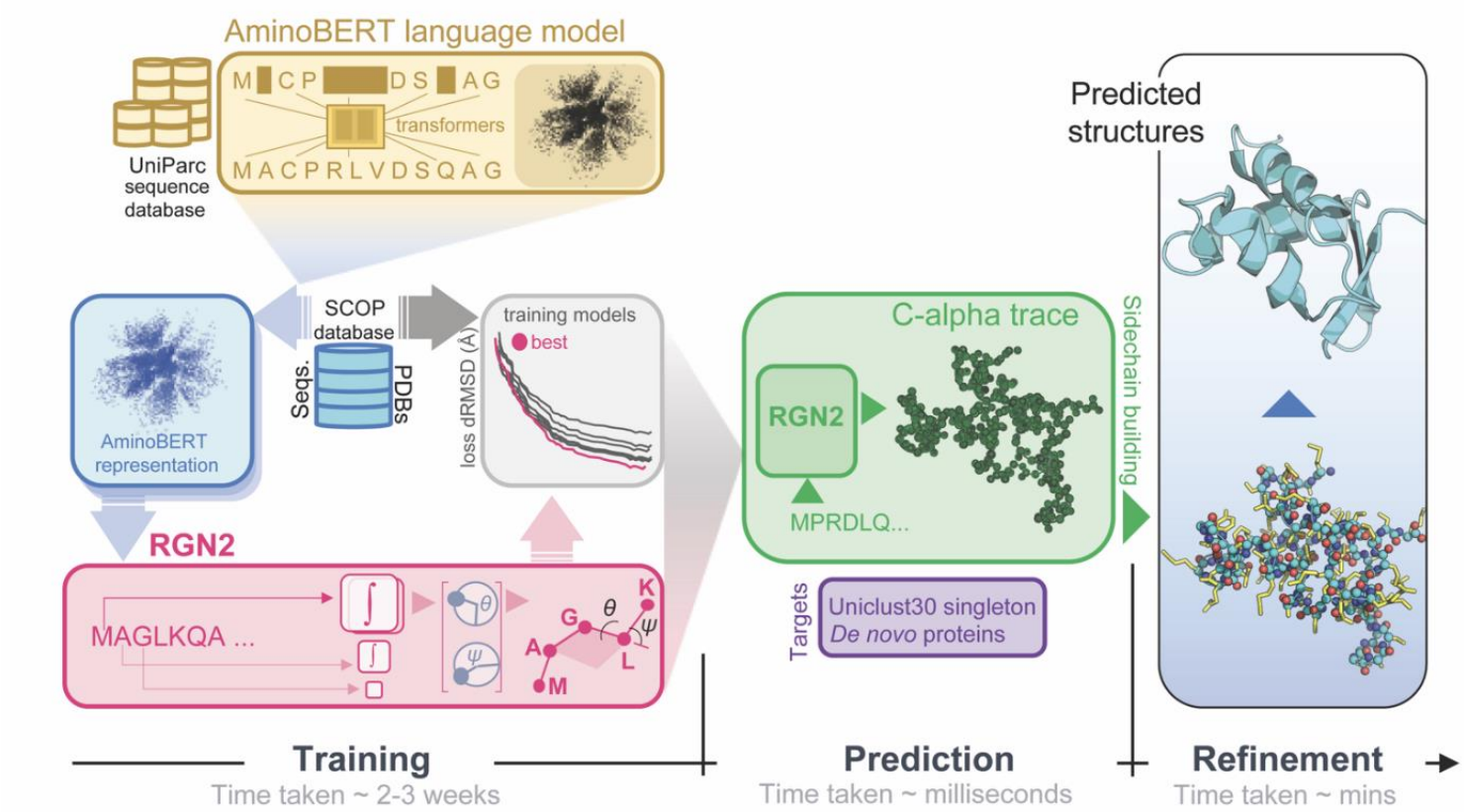


Figure 1. Organization and application of RGN2. RGN2 combines a Transformer-based protein language model (AminoBERT) with a recurrent geometric network that utilizes Frenet-Serret frames to generate the backbone structure of a protein. Placement of side chain atoms and refinement of hydrogen-bonded networks are subsequently performed using the Rosetta energy function.

https://www.youtube.com/watch?v=eobc7cMMpeY&feature=youtu.be

USING ALPHAFOLD FOR RAPID AND ACCURATE FIXED BACKBONE PROTEIN DESIGN

Lewis Moffat
Department of Computer Science
University College London
Gower St, London WC1E 6BT
lewis.moffat@cs.ucl.ac.uk

Joe G. Greener
Department of Computer Science
University College London
Gower St, London WC1E 6BT
j.greener@ucl.ac.uk

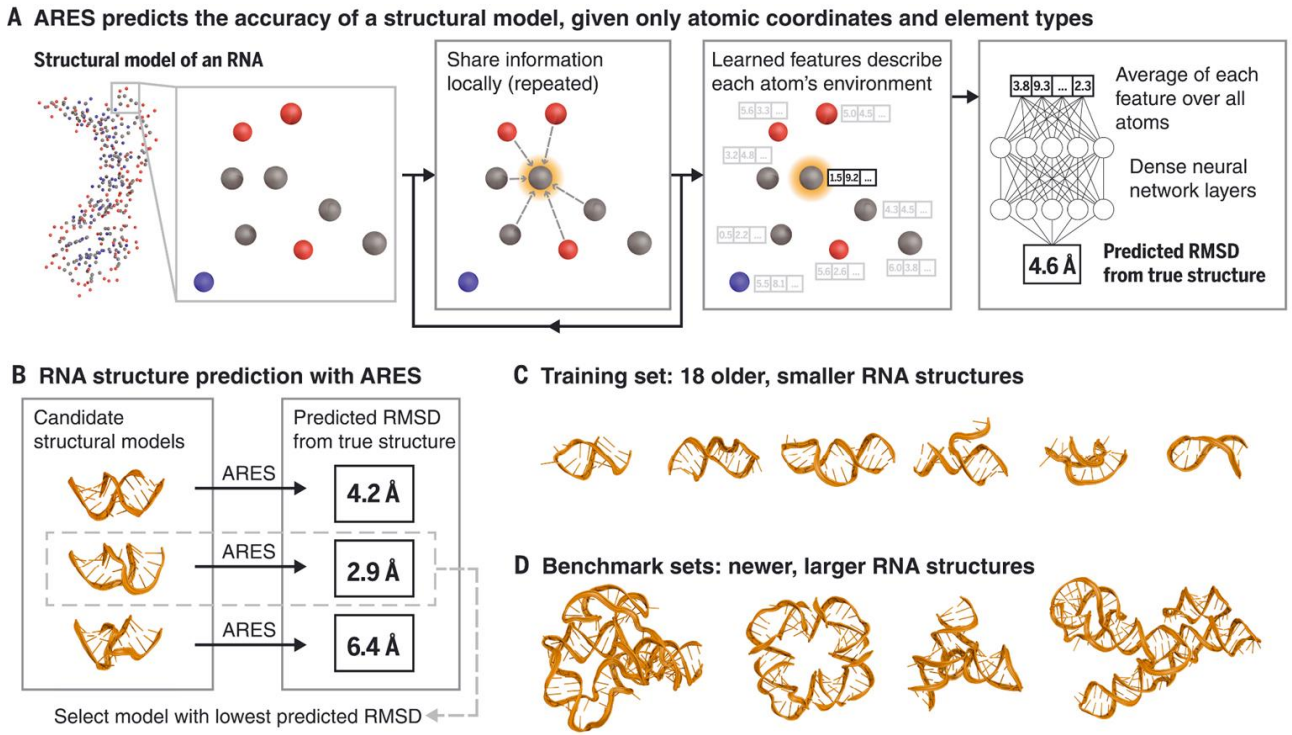
David T. Jones*
Department of Computer Science
University College London
Gower St, London WC1E 6BT
d.t.jones@ucl.ac.uk

ABSTRACT

The prediction of protein structure and the design of novel protein sequences and structures have long been intertwined. The recently released AlphaFold has heralded a new generation of accurate protein structure prediction, but the extent to which this affects protein design stands yet unexplored. Here we develop a rapid and effective approach for fixed backbone computational protein design, leveraging the predictive power of AlphaFold. For several designs we demonstrate that not only are the AlphaFold predicted structures in agreement with the desired backbones, but they are also supported by the structure predictions of other supervised methods as well as *ab initio* folding. These results suggest that AlphaFold, and methods like it, are able to facilitate the development of a new range of novel and accurate protein design methodologies.

*To whom correspondence should be addressed

Geometric deep learning of RNA structure



elixir CZECH REPUBLIC

https://www.science.org/doi/10.1126/science.abe5650