



KATEDRA FYZIKÁLNÍ CHEMIE
UNIVERZITY PALACKÉHO V OLMOUCI



INSTITUTE OF MOLECULAR AND
TRANSLATIONAL MEDICINE



6th Advanced *in silico* Drug Design

KFC/ADD

AlphaFoldology

Machine learning revolution in structural biology

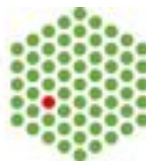
Karel Berka

(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
CR
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

'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.

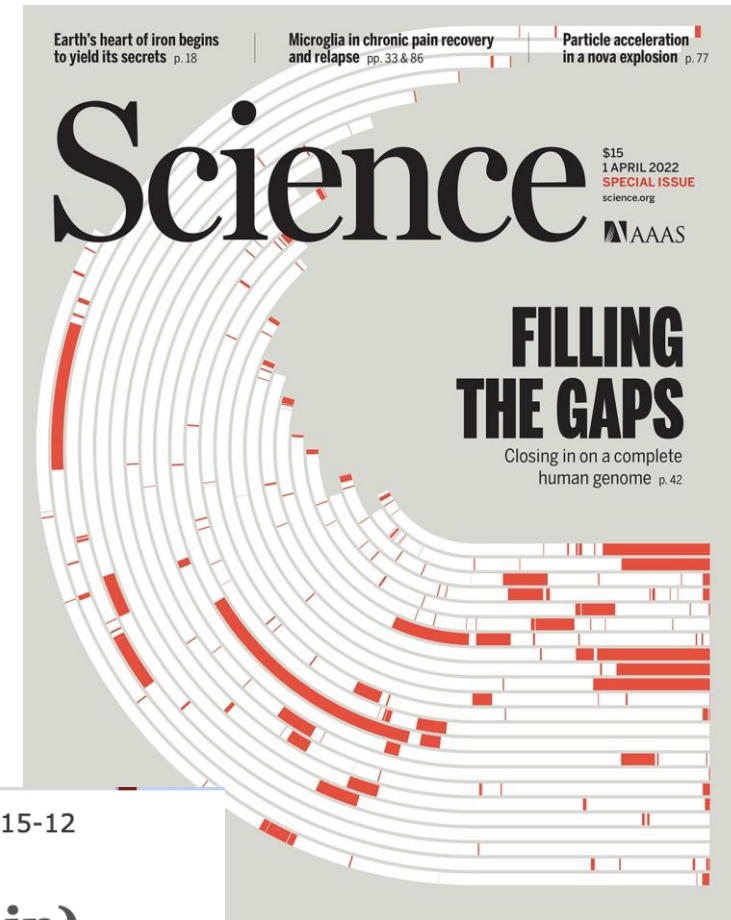
(almost) Complete human genome



[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)

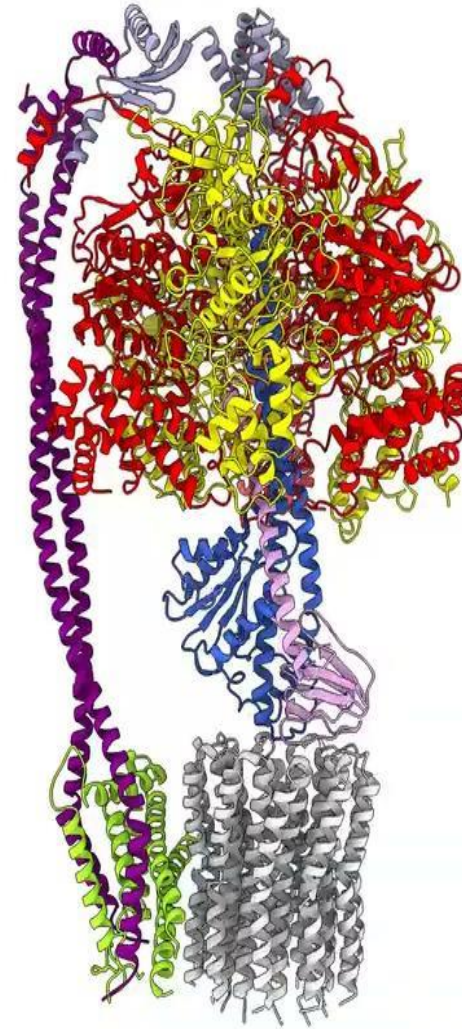
Scientists today publish the sequence of chromosome 1: the largest and last of the human chromosomes to be done and dusted. *News@nature* finds out what this latest milestone means.



Knowing structure to understand function

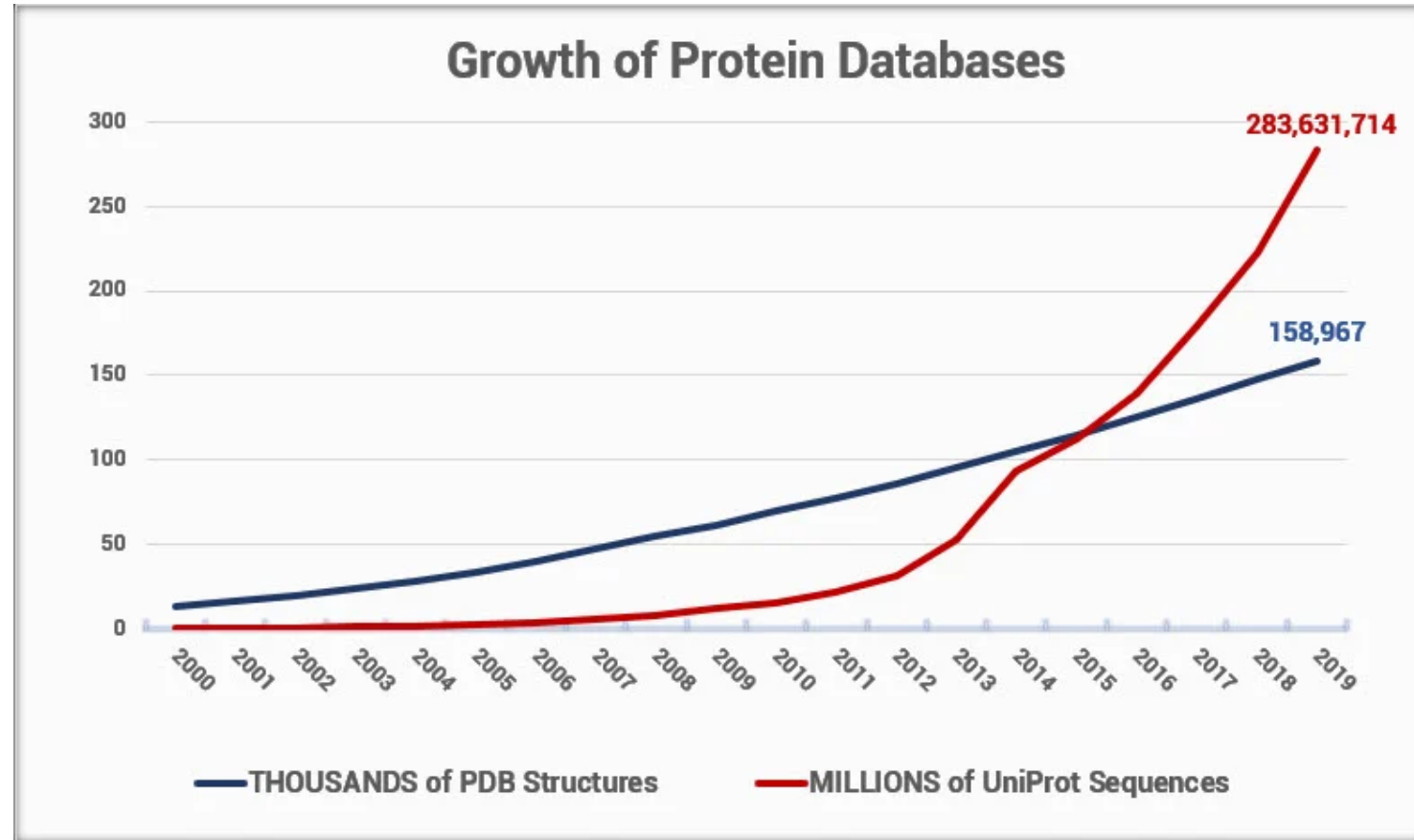


[wikipedia/imatinib](https://en.wikipedia.org/wiki/Imatinib)



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

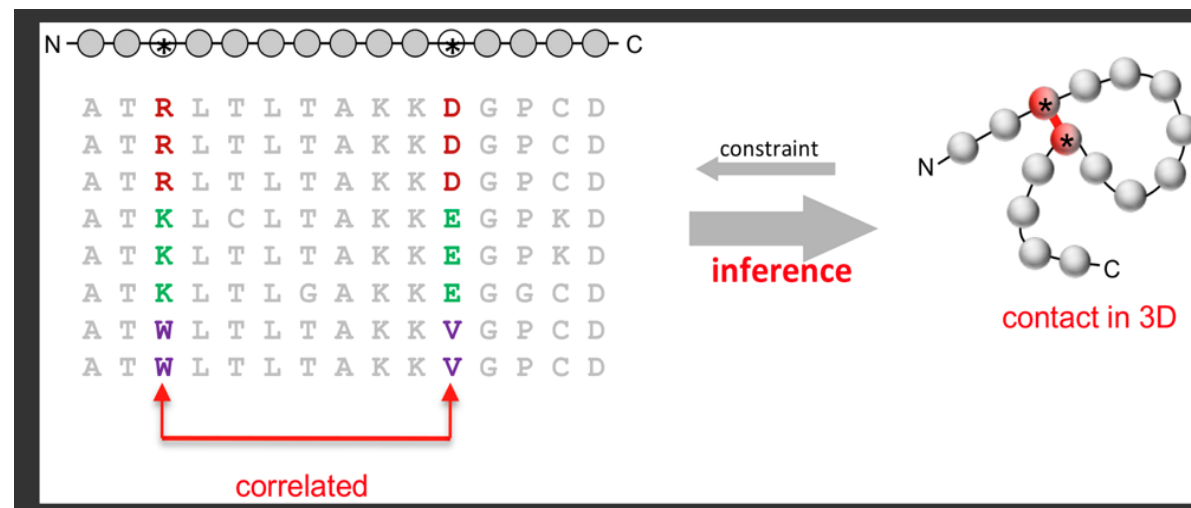
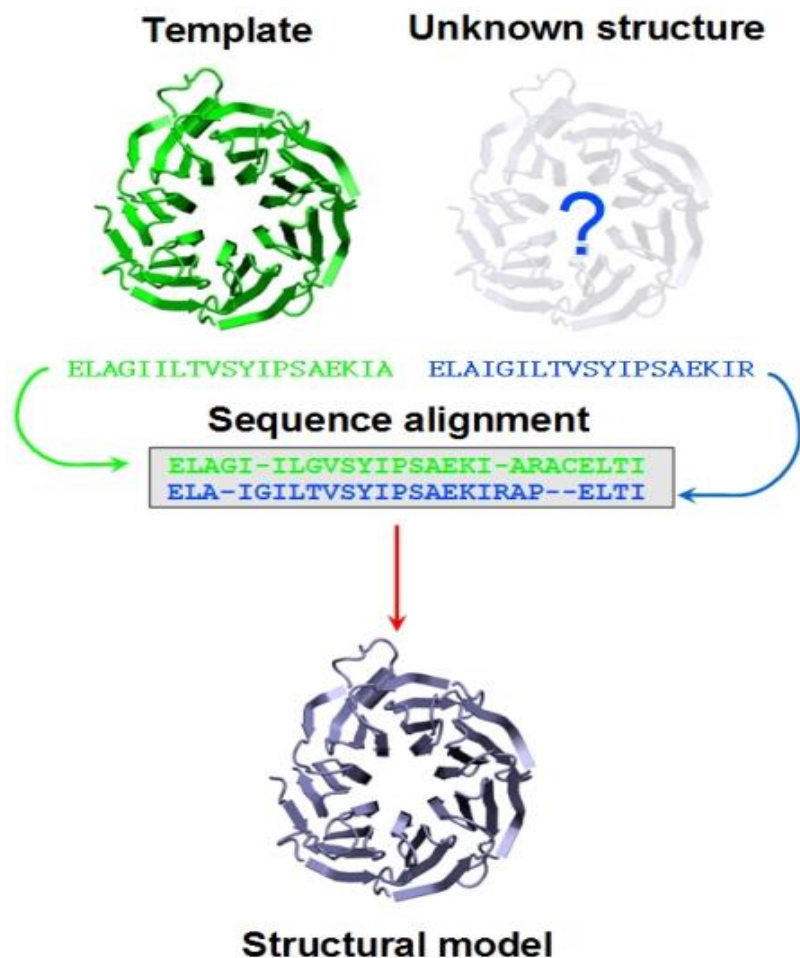
"The native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment."

- it shall be possible to predict structure from sequence

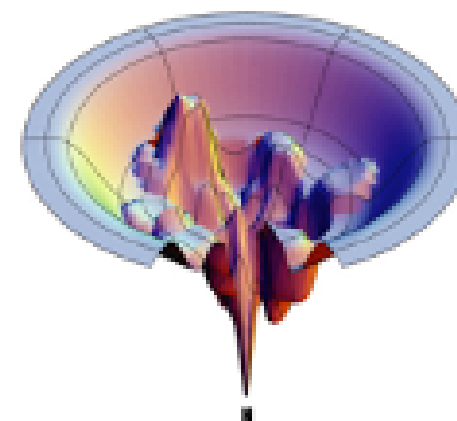


ribonuclease A
Wikipedia

Principles of prediction from sequence



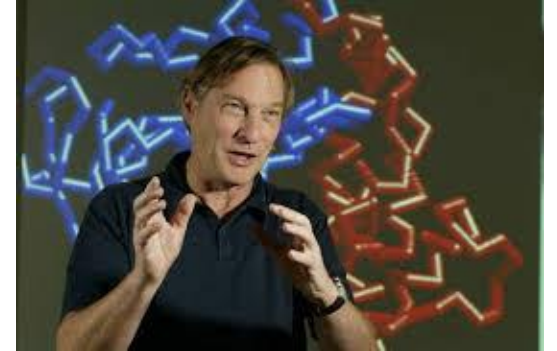
From Protein Structure and Function
2004-2005 Online Update by
Gregory A Petsko and Dagmar Ringe



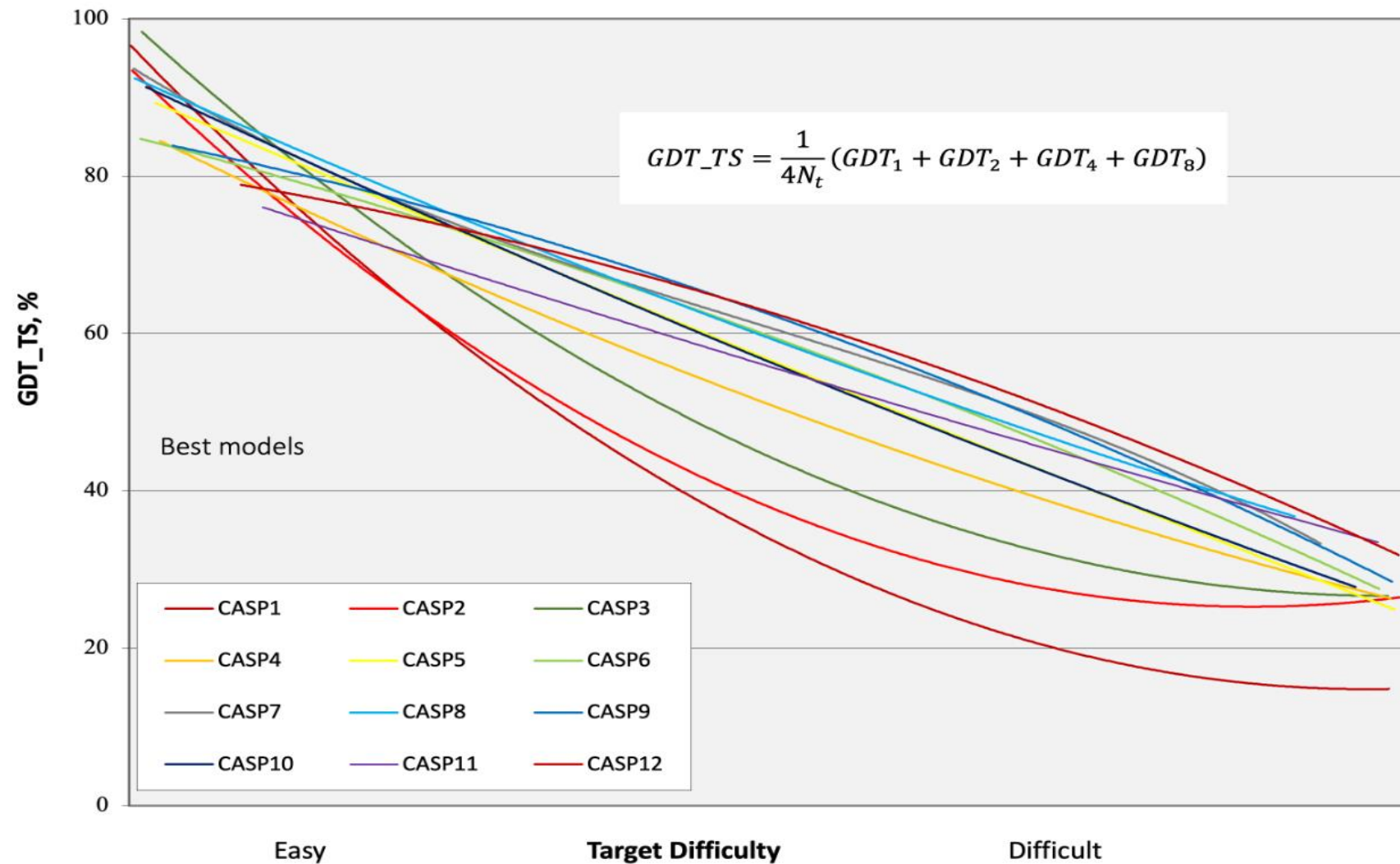
<http://www.pdb.org/pdb>

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?

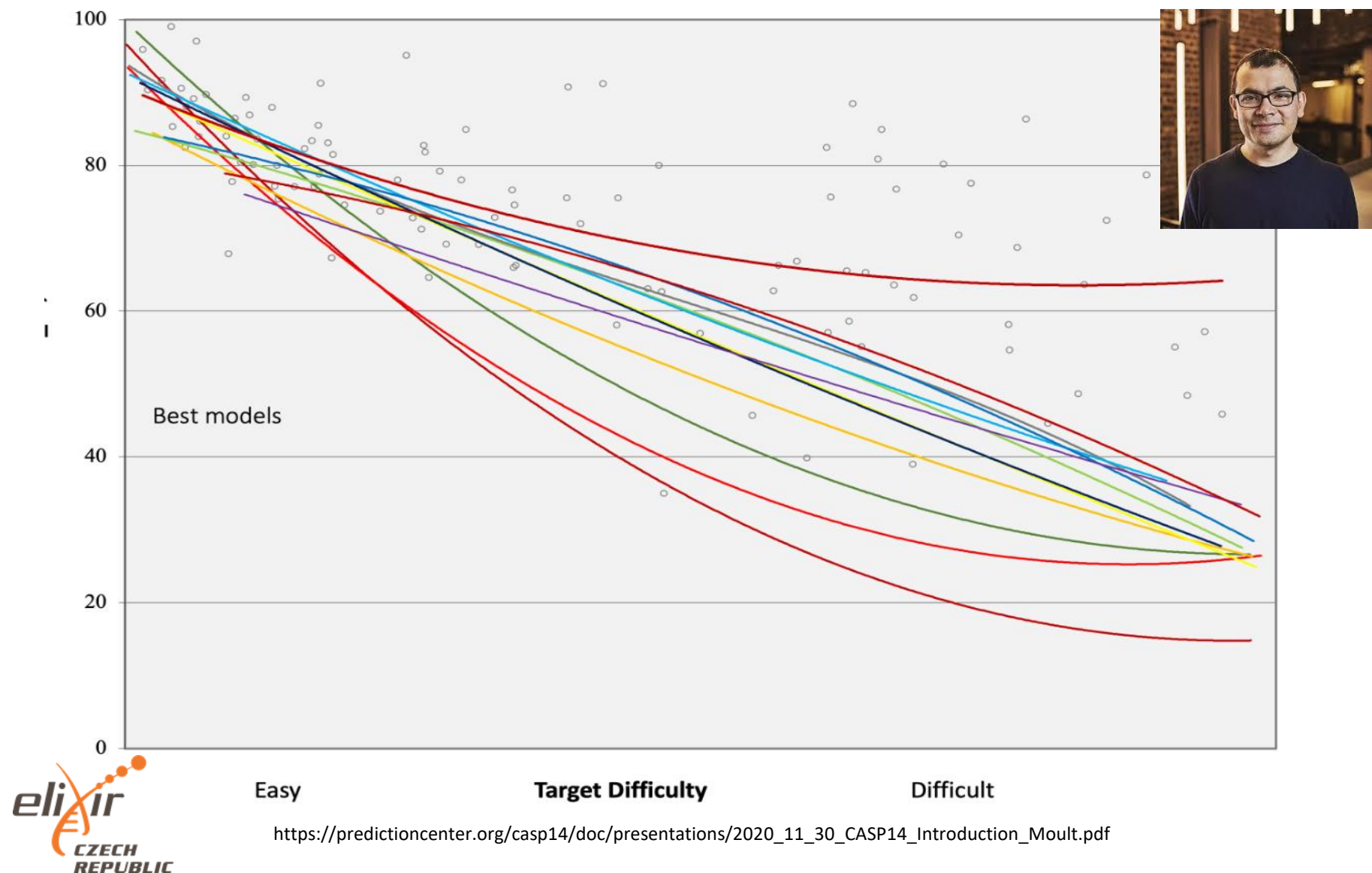


https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf

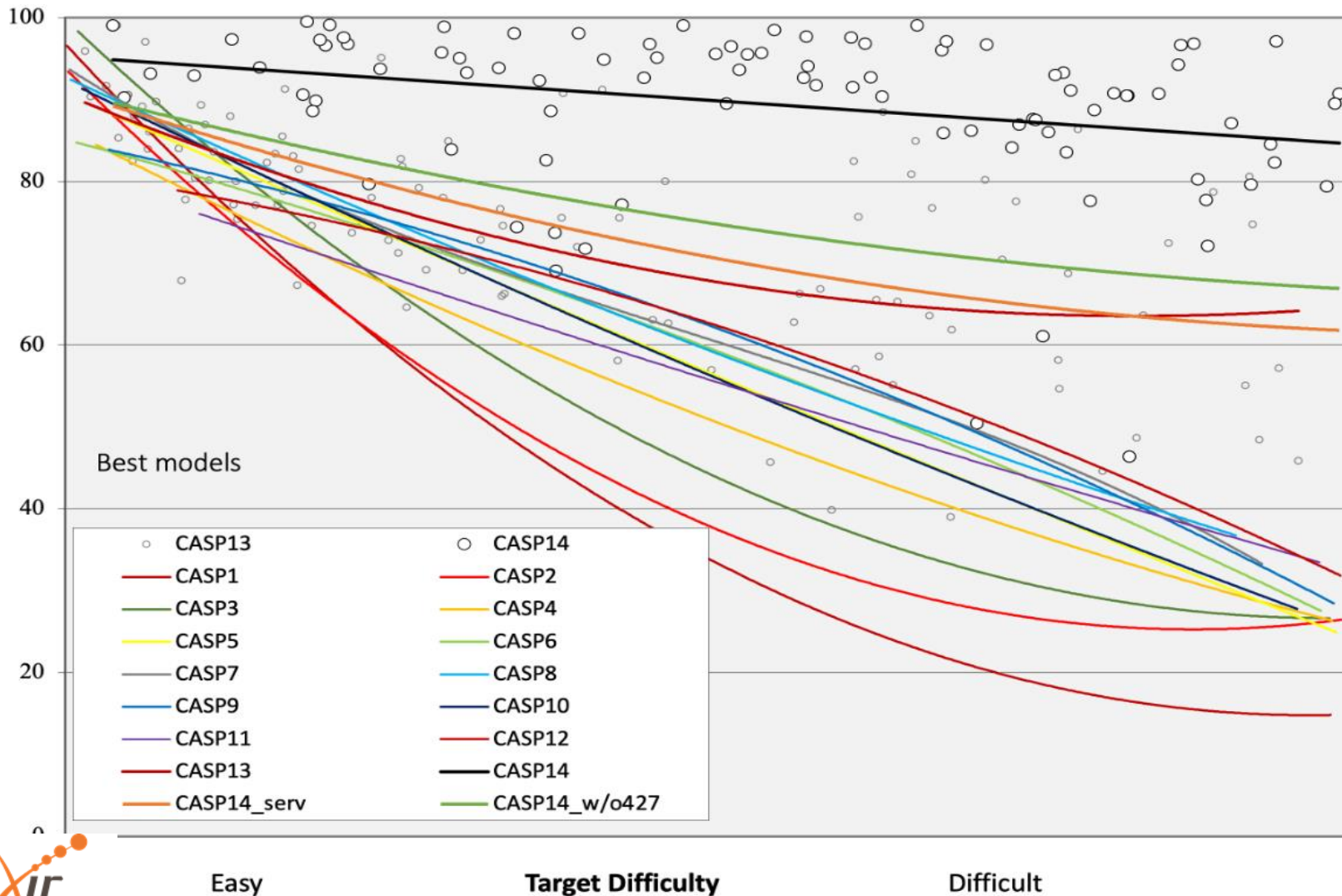
GDT_TS = Global distance test - total score (max 100%)

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...



2020 CASP 14: Alphafold2 wins



2022 CASP15: AlphaFold is basic

#1 **PEZYZFoldings** AF2-based. Diverse MSAs.

Custom, fine-tuned AF2 refinement

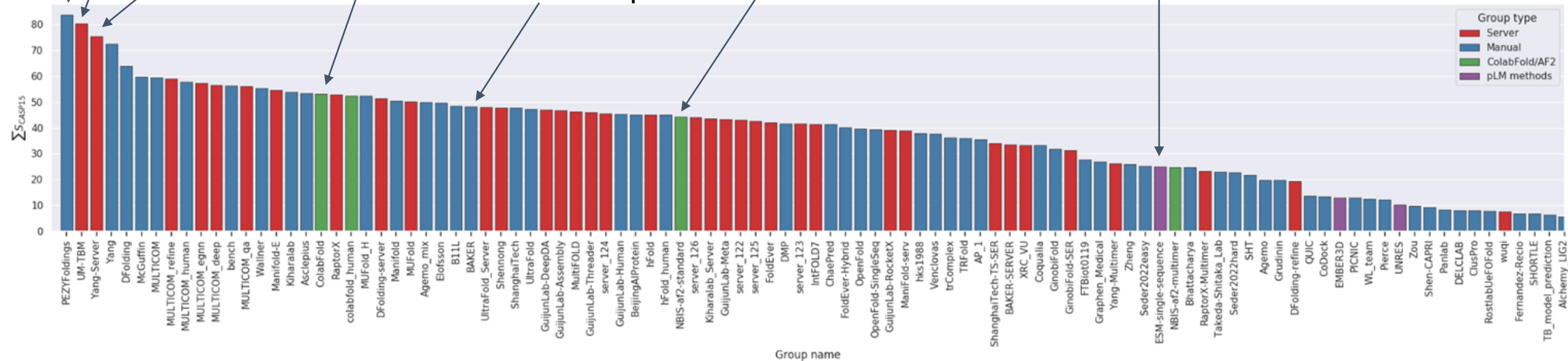
#2 **UM-TBM** Diverse MSAs. Threading then AF2 predictions guide I-TASSER REMC

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

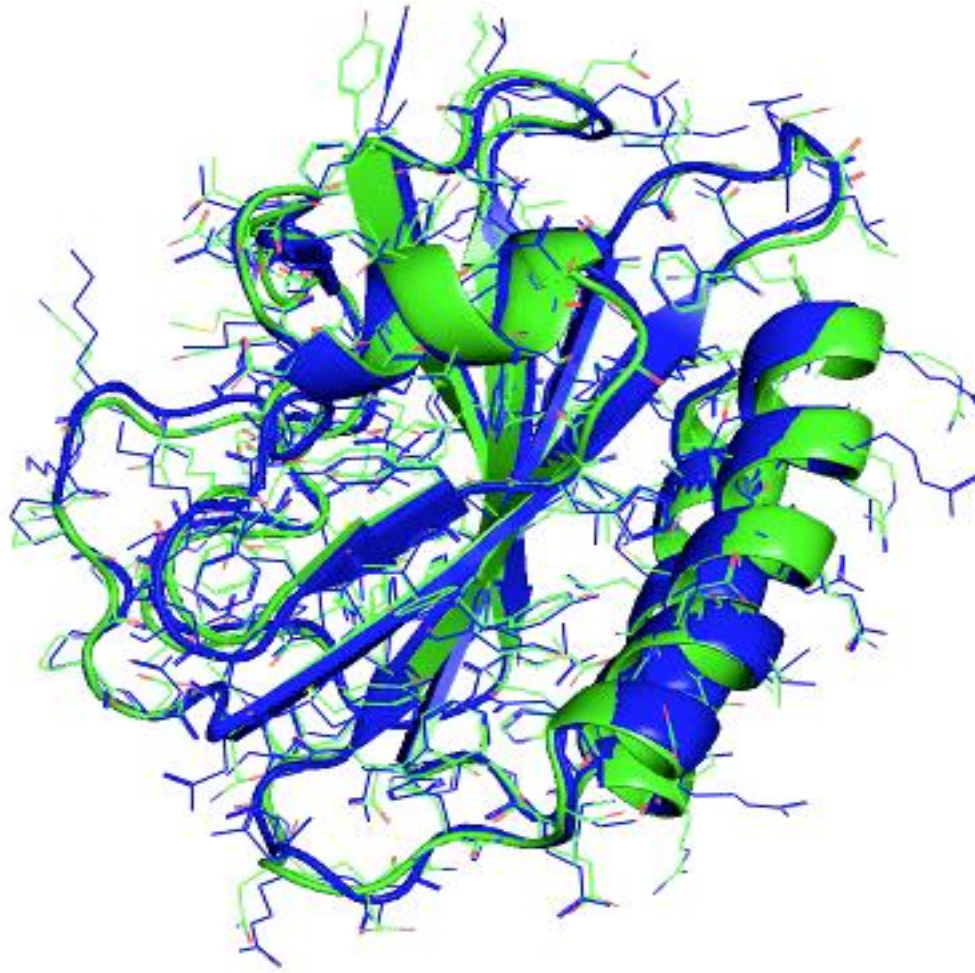
ColabFold and **NBIS-af2-standard**

BAKER top non-AF2 method

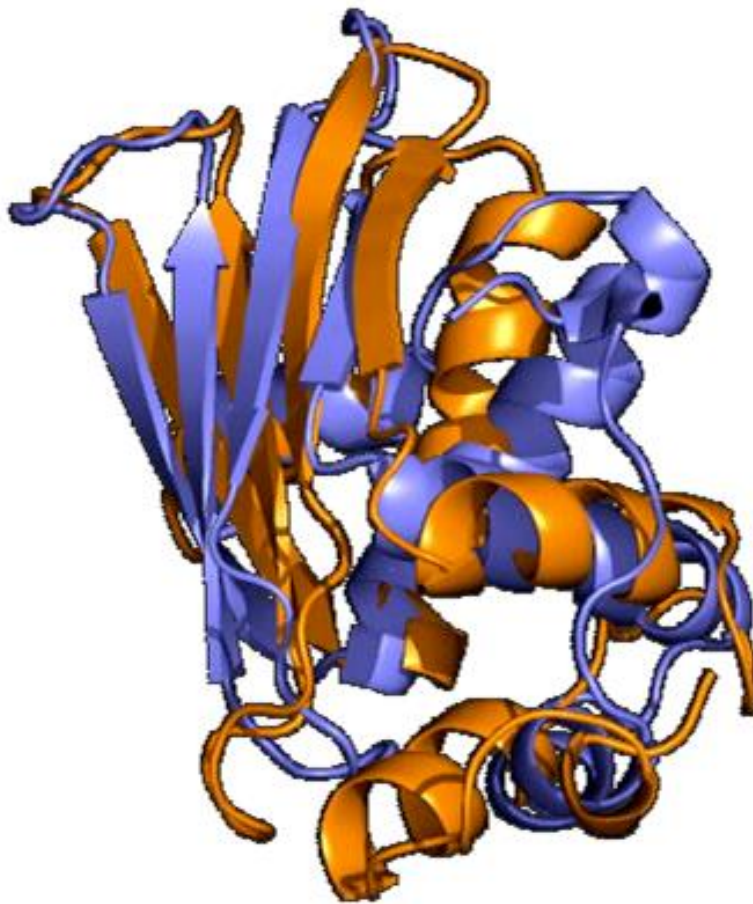
ESM-singlesequence is the top pure **pLM** method by this metric



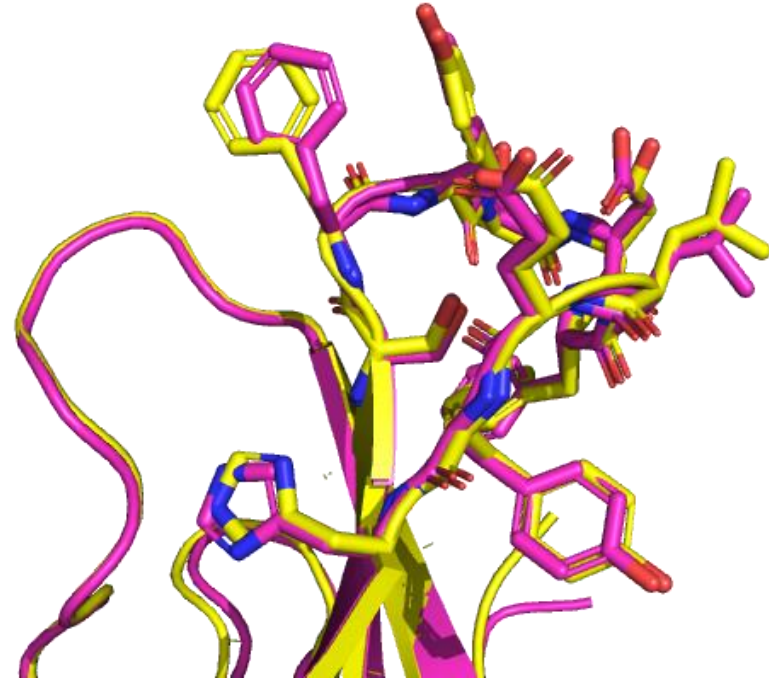
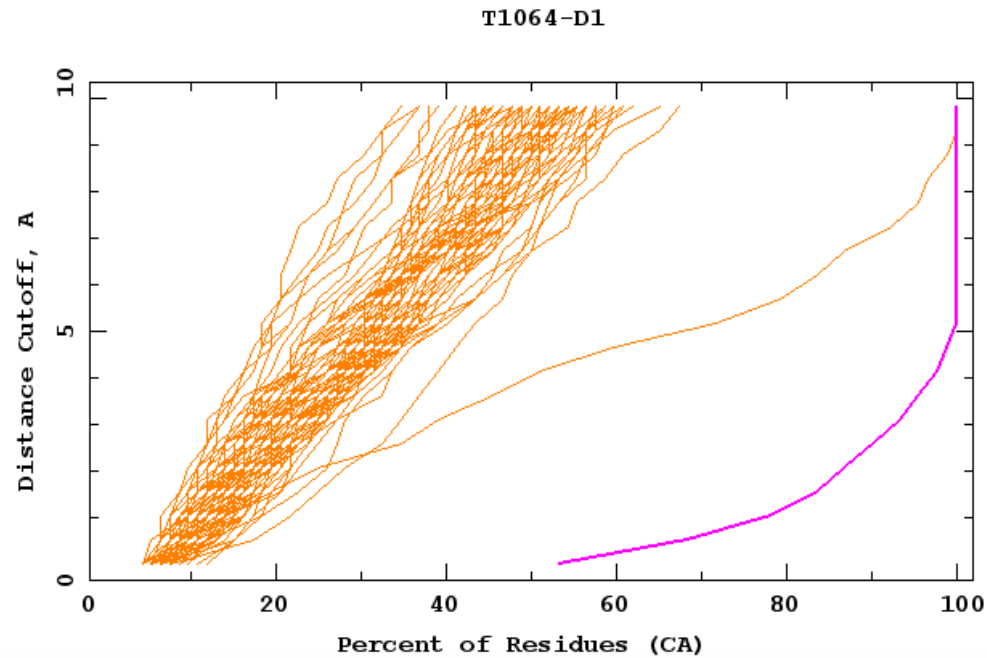
How does good prediction look like?



The worst prediction of Alphafold 2 in CASP 14



Side chain predictions– orf8 covid19



GDT_TS= 87

Also good
so how does it work?

AlphaFold2

- under the hood

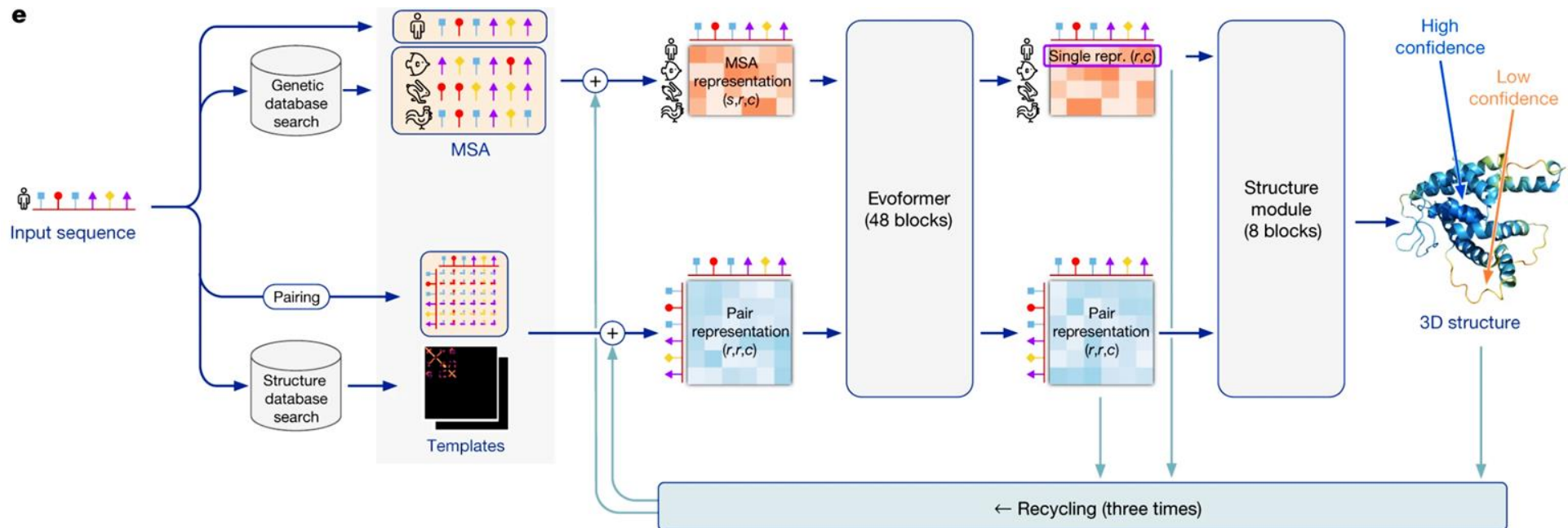
AlphaFold2

Input: sequence

extended by MSA + structural templates

Evoformer and Structure modules (w MD simulation)

pLDDT - predicted local confidence prediction



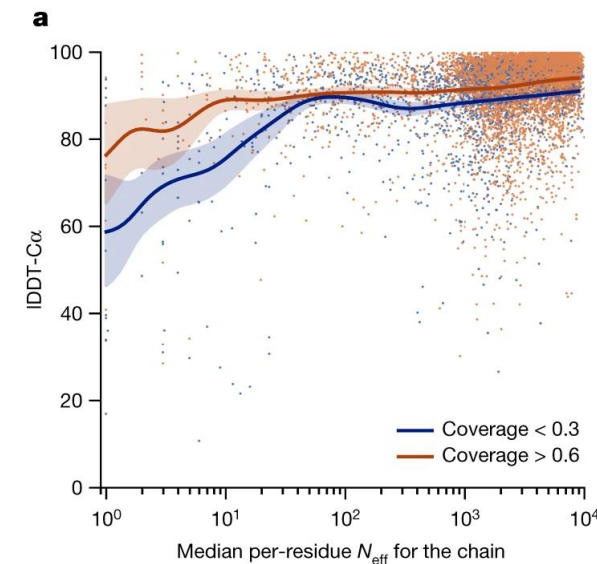
MSA - multiple sequence alignment

using standard tools - jackhmmer, HHBlits

- sequence DBs:
 - *UniRef90*
 - *UniClust30* = for sequence self-distillation
- metagenomicsDBs - to fully cover classes underrepresented 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**

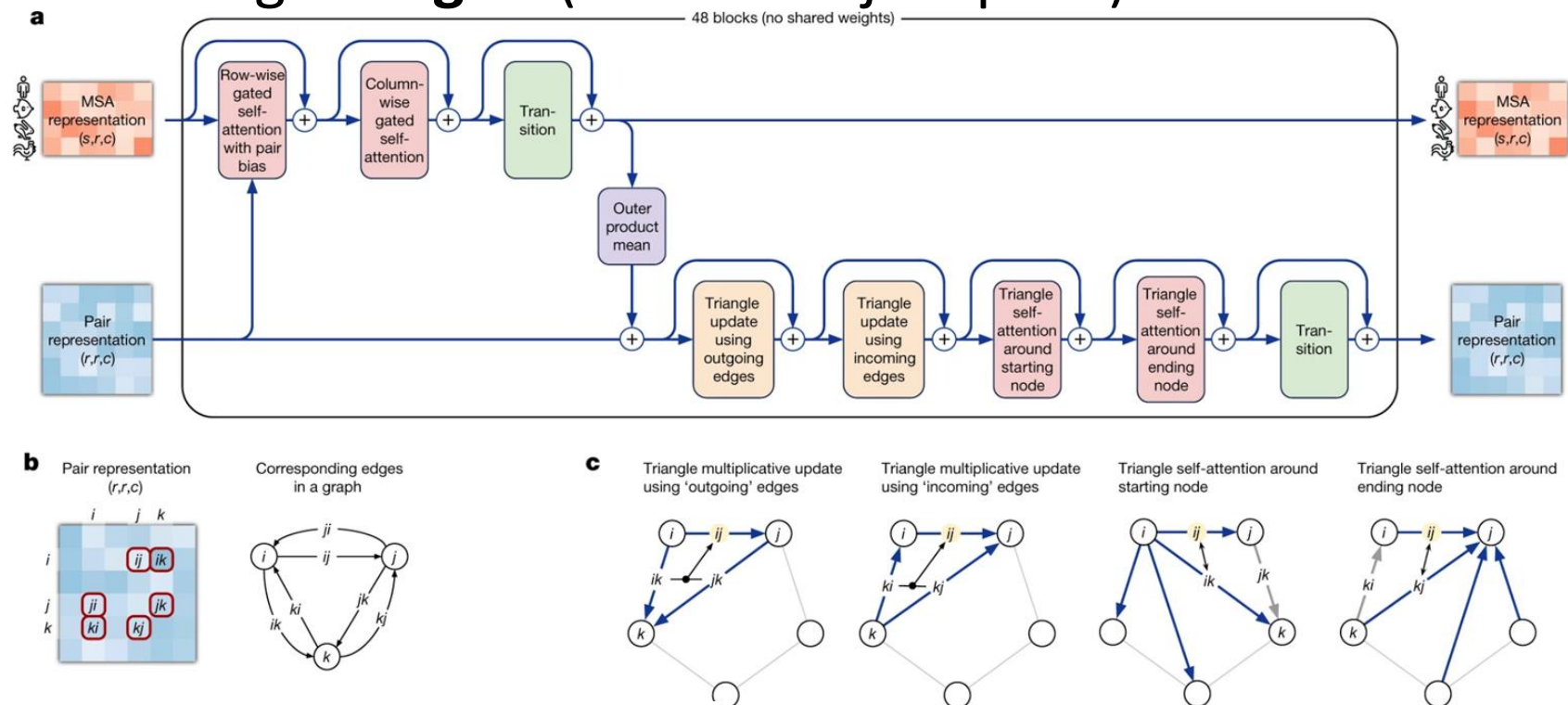
- predict 350000 structures from UniRef30 using trained network
- filter to high confidence subset
- then train again from scratch with mixture of PDB and UniRef30

=> 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

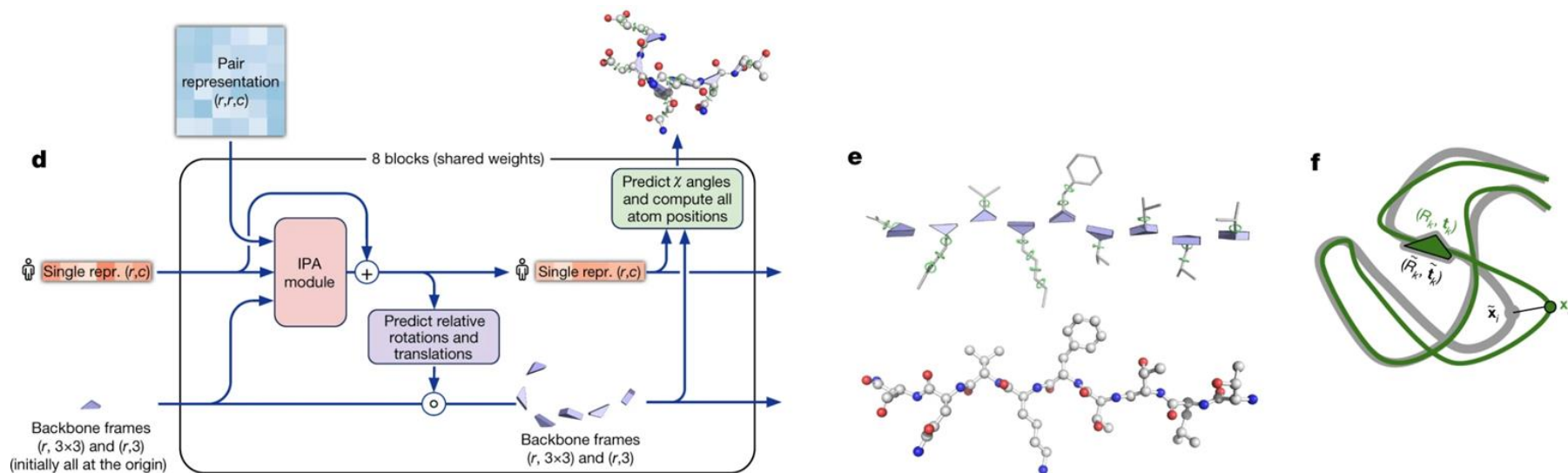
EvoFormer module

- **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)



Structure module

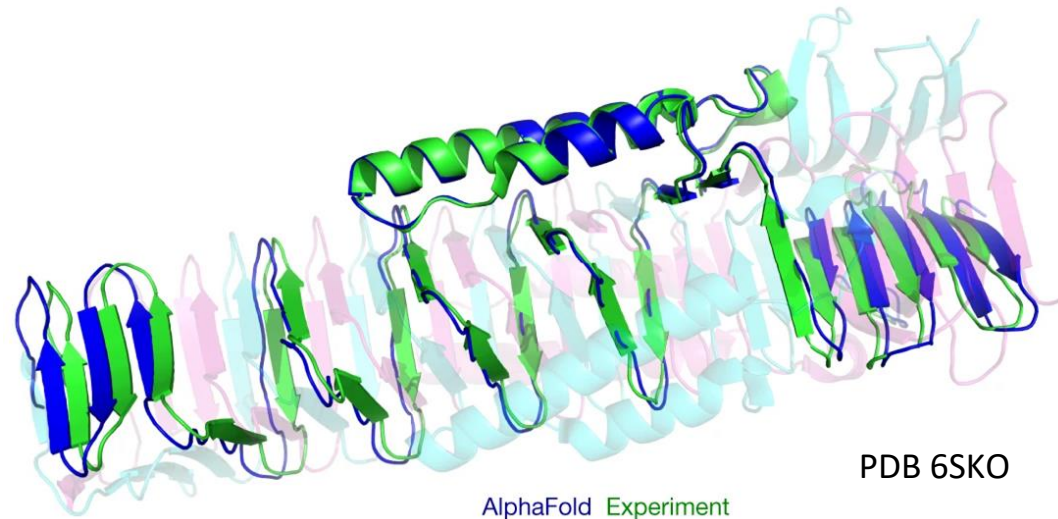
- prioritize backbone positions+orientations
 - **residue gas** - free floating rigid body rotations and translation
 - updates
 - IPA (invariant point attention) - neural activations only in rigid 3D
 - equivariant update using updated activations
- later fix backbone geometry
 - **avoid loop closure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



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 intertwined



AlphaFoldDB

EMBL-EBI Services Research Training About us EMBL-EBI

AlphaFold Protein Structure Database Home About FAQs Downloads

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help: AlphaFold DB search help

AlphaFold DB provides open access to protein structure predictions for the human proteome and 20 other key organisms to accelerate scientific research.

"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
<i>Arabidopsis thaliana</i>	<i>Arabidopsis</i>	UP000006548 	27,434	Download (3,678 MB)
<i>Caenorhabditis elegans</i>	Nematode worm	UP000001940 	19,694	Download (2,626 MB)
<i>Candida albicans</i>	<i>C. albicans</i>	UP000000559 	5,974	Download (974 MB)

Compressed prediction files for global health proteomes:

Species	Common Name	Reference Proteome	Predicted Structures	Download
<i>Ajellomyces capsulatus</i>	<i>Ajellomyces capsulatus</i>	UP000001631 	9,199	Download (1,351 MB)
<i>Brugia malayi</i>	<i>Brugia malayi</i>	UP000006672 	8,743	Download (1,274 MB)
<i>Campylobacter jejuni</i>	<i>C. jejuni</i>	UP000000799 	1,620	Download (173 MB)
<i>Cladophialophora carrionii</i>	<i>Cladophialophora carrionii</i>	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

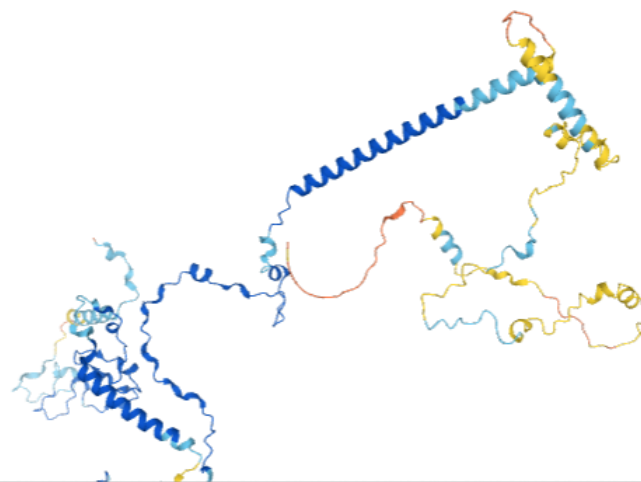
Model Confidence:

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

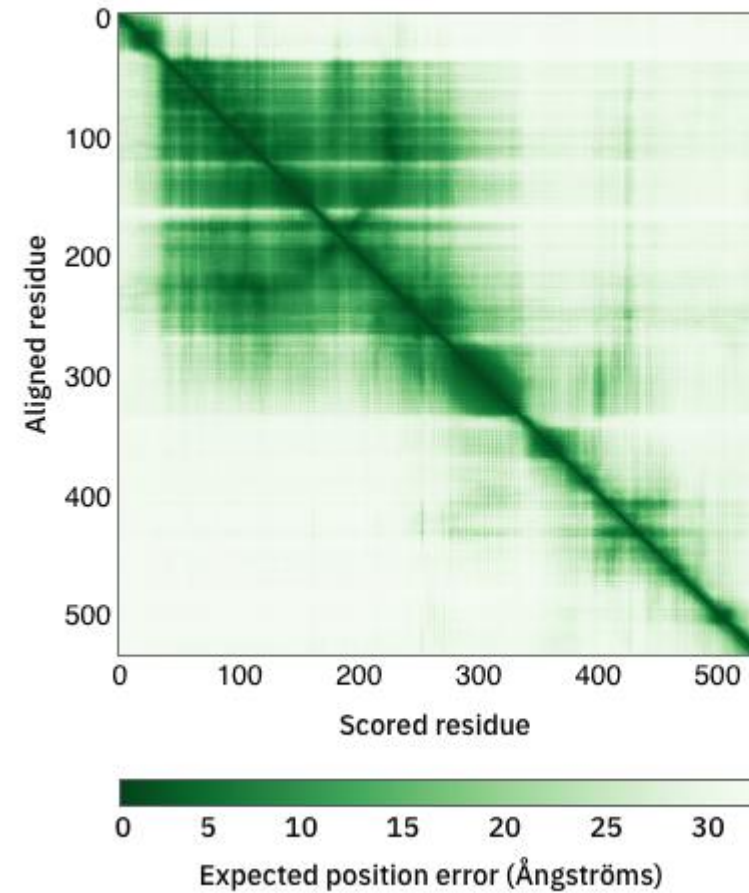
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Sequence of AF-Q13573-... 1: SNW do... A

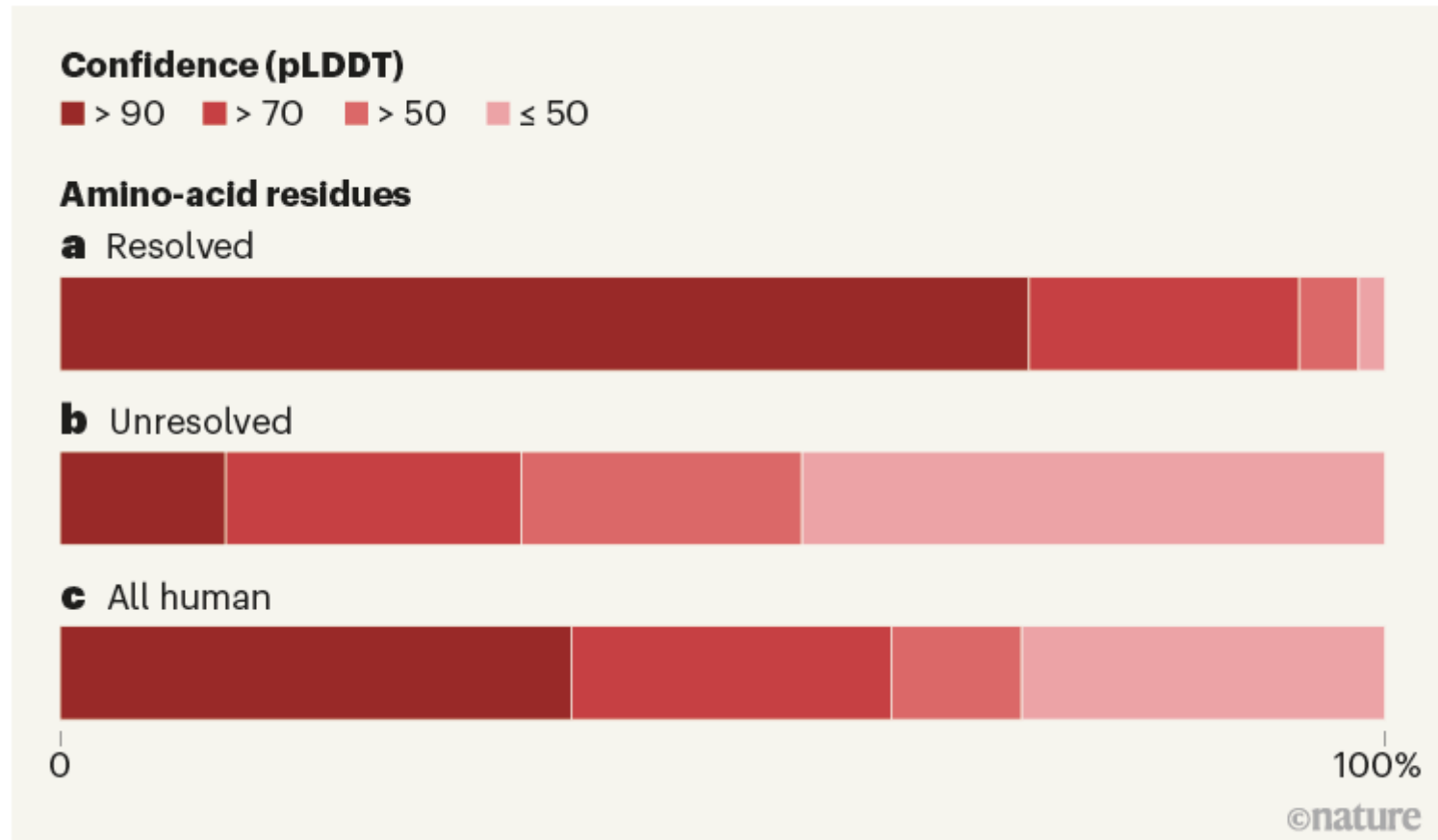
1 11 21 31 41 51 61 71 81 91 101 111 121
MALTSFLPAPTQLSQDLEAEKARSQSRQTSLVSSRREPPPYGYRKGWIPRLLEDGFGGAFPEIHVAQYPLDMGRKKKMSNALAIQVDSEGKI KYDAIARQQGSKDKVIYSKYTDLVPKEV
131 141 151 161 171 181 191 201 211 221 231 241
MNADDPDLQRPDEEAIKEITEKTRVALEKSVSQKVAAMPVRAADKLAPQYIRYTPSQQGVAFNSGAKQVRVIRMVEMQKDPMEPPRFKINKKI PRGPPSPAPVMHSPSRKMTVKEQQEWKIP
251 261 271 281 291 301 311 321 331 341 351 361 371
PCISNWKNAKGYTIPLDKRLAADGRGLQTVHINENFAKLAEALYIADRKAREAVEMRAQVERKMAQKEKEKHEEKLREMAQKARERRAGIKTHVEKEDGEARERDEIRHDRKERQHDRNLSRA



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 [IDDT-C \$\alpha\$ 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	MGLEALVPLAMIVAIFLLLVDLMRHQRWAARYPPGGLPLPLPGLGNLLHVDFQNTPYCFDQ
A0A087X1C5	CP2D7_HUMAN	1	MGLEALVPLAMIVAIFLLLVDLMRHQRWAARYPPGGLPLPLPGLGNLLHVDFQNTPYCFDQ

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	61	LRRRFGDVFSLQLAWTPVVVLNGLAAVREAMVTRGEDTADRPPAPIYQVLGFGPRSQ---
A0A087X1C5	CP2D7_HUMAN	61	LRRRFGDVFSLQLAWTPVVVLNGLAAVREAMVTRGEDTADRPPAPIYQVLGFGPRSQGVI

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	118	-----GRPFRPNGLLDK
A0A087X1C5	CP2D7_HUMAN	121	LSRYGPAWREQRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFADQAGRPFPRPNGLLDK

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	130	AVSNVIASLTCGRREFYDDPRFLRLDLAQEGSKESGFLREVLNAVVPVLPHPALAGKV
A0A087X1C5	CP2D7_HUMAN	181	AVSNVIASLTCGRREFYDDPRFLRLDLAQEGSKESGFLREVLNAVVPVLPHPALAGKV

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	190	LRFOKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAKKEKAKGSPESFNDENLRIVVG
A0A087X1C5	CP2D7_HUMAN	241	LRFOKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAKKEKAKGSPESFNDENLRIVVG

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	250	NLFLAGMVTTLTTLAWGLLLMILHLDVQ-----LRVQQEIDDVIGQV
A0A087X1C5	CP2D7_HUMAN	301	NLFLAGMVTTSSTTLAWGLLLMILHLDVQRGRVSPGCPVIGTHVCPVRVQQEIDDVIGQV

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	292	RRPEMGDAQHMPYTTAVIHEVQHFQDIVPLGVTHMTSRDIEVQGFRIKGTTLITNLSSV
A0A087X1C5	CP2D7_HUMAN	361	RRPEMGDAQHMPCTTAVIHEVQHFQDIVPLGVTHMTSRDIEVQGFRIKGTTLITNLSSV

A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	352	LKDEAVWEKPFRLFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFFFTSLQ
A0A087X1C5	CP2D7_HUMAN	421	LKDEAVWKKPFRLFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFFFTSLQ

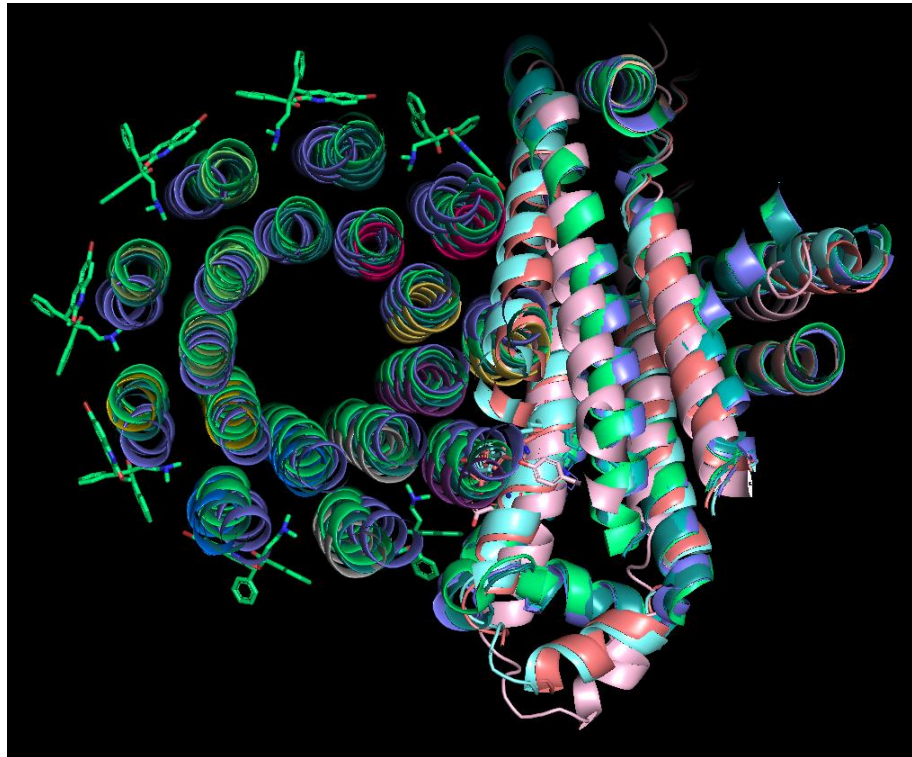
A0A1B0GTQ1	A0A1B0GTQ1_HUMAN	412	HFSFSVAAGQPRPSHSRVVSFLVTPSPYELCAVPR
A0A087X1C5	CP2D7_HUMAN	481	HFSFSVAAGQPRPSHSRVVSFLVTPSPYELCAVPR



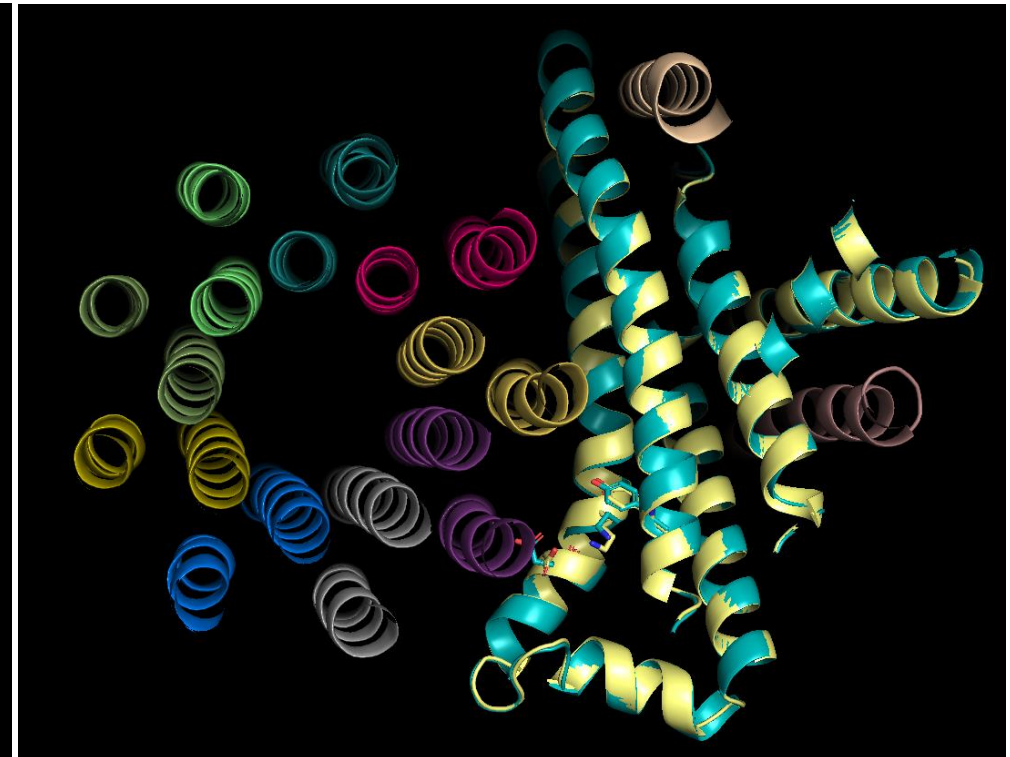
Structure can be only as good as its sequence

AlphaFoldDB contains monomers

- *Myc* ATPsynthase



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

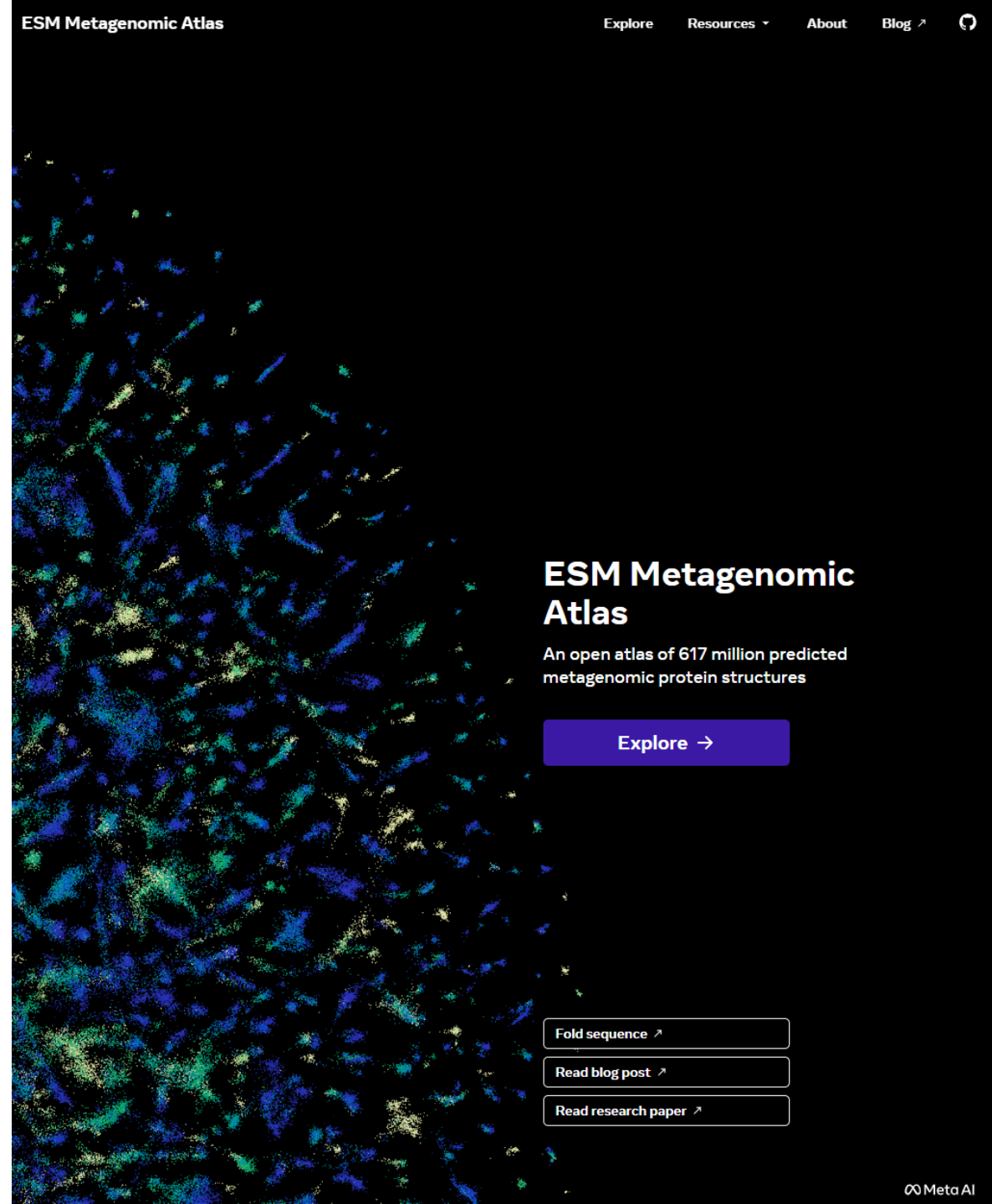


SwissModel
over *M. smegmatis* ATPsynthase
(PDBID: 7njp)

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/>

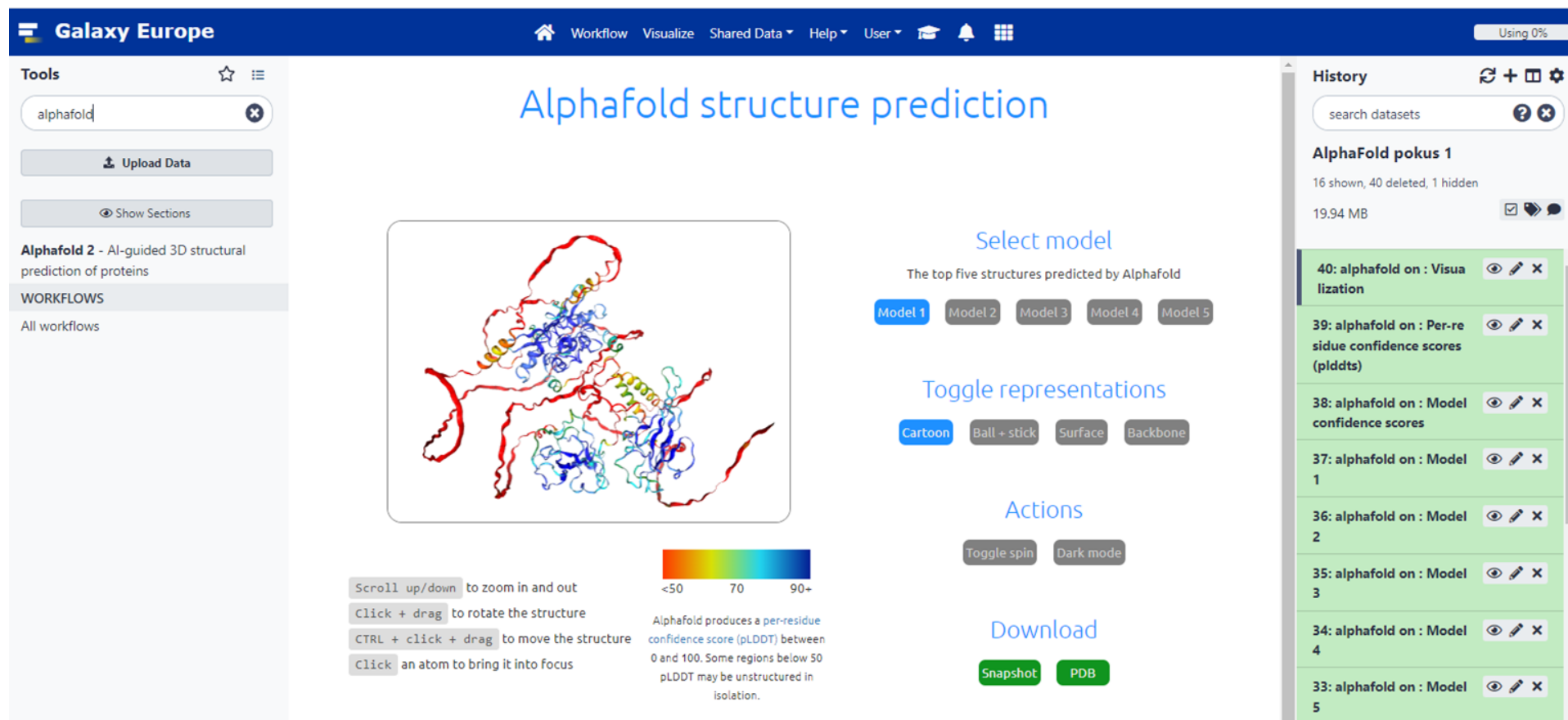
Alphafold 2 on ELIXIR CZ

- 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



The screenshot shows the Galaxy Europe interface for AlphaFold structure prediction. The main panel displays a 3D ribbon model of a protein dimer. A color scale at the bottom indicates the per-residue confidence score (pLDDT) from <50 (red) to 90+ (blue). A legend explains that AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100, with regions below 50 pLDDT being unstructured in isolation.

On the left, the 'Tools' panel shows 'AlphaFold 2 - AI-guided 3D structural prediction of proteins' under 'WORKFLOWS'. The 'History' panel on the right lists previous jobs, including 'AlphaFold pokus 1' and several jobs labeled 'alphafold on : Visualization', 'alphafold on : Per-residue confidence scores (pLDDTs)', and 'alphafold on : Model confidence scores'.

Navigation and interaction instructions are provided at the bottom left:

- Scroll up/down to zoom in and out
- Click + drag to rotate the structure
- CTRL + click + drag to move the structure
- Click an atom to bring it into focus

Additional controls include 'Select model' (Model 1 to Model 5), 'Toggle representations' (Cartoon, Ball + stick, Surface, Backbone), 'Actions' (Toggle spin, Dark mode), and 'Download' (Snapshot, PDB).

trick - dimerization fake as long disordered poly-N chain

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

...



Europe PMC

Search life-sciences

alphafold

[Advanced search](#)

Search only

Type [?](#)

- ☐ Research articles (2 747)
- ☐ Reviews (654)
- ☐ Preprints (469)

Free full text [?](#)

- ☐ Free to read (3 447)
- ☐ Free to read & use (3 289)

Date

- ☐ 2023 (227)
- ☐ 2022 (2 945)
- ☐ 2021 (593)

[Custom date range](#) ▶

As of 30.1.2023

MrParse:

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



MrParse Analysis

Version: 0.2.1

MrParse: a program to find and analyse search models for crystallographic Molecular Replacement. The program is being developed by [Dan Rigden's group](#) at the University of Liverpool.

MrParse is currently under development and we are keen to make it as useful to the community as possible. If you have any suggestions for it's development, or ideas on how we could improve it, please [get in touch](#).

IKL Info

Name	Resolution	Space Group	Has NCS?	Has Twinning?	Has Anisotropy?
7dry-sf	1.44	P41212	false	false	true

Experimental structures from the PDB

Name	PDB	Resolution	Region	Range	Length	eLLG	Mol. Wt.	eRMSD	Seq. Ident.
2cvi_B_1	2cvi	1.50	1	158-230	71	43.5	8676	1.085	0.31

Visualisation of Regions



Sequence Based Predictions



Structure predictions from the EBI AlphaFold database

Name	model	Date Made	Region	Range	Length	Avg. pLDDT	H-score	Seq. Ident.
Q12362_1	Q12362	01-JUL-21	1	2-180	177	90.15	85	0.41
P87241_1	P87241	01-JUL-21	1	4-176	171	91.55	85	0.38

Visualisation of Regions



Are structural biologists and bioinformaticians on the job market?

Not yet as:

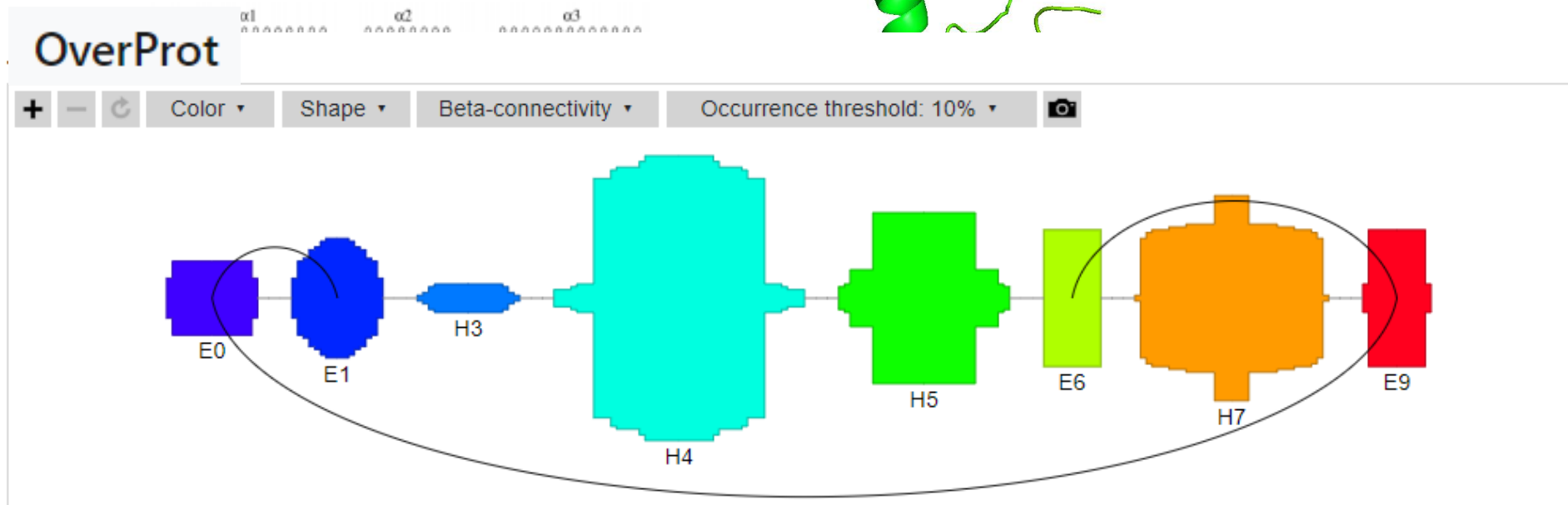
- Alphafold can not do **point mutations** - design of functions
- Alphafold can not be used for **drug design**
- Alphafold can not do **conformational changes** or **dynamics**
- Alphafold can not do **multiprotein complexes** – interactions
- Alphafold can not do effects of **post-translational protein 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?**

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%

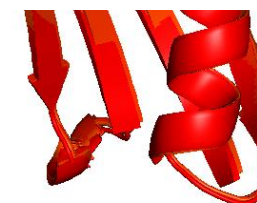
A minimal sequence code for switching protein structure and function

Patrick A. Alexander, Yanan He, Yihong Chen, , and Philip N. Bryan  [Authors Info & Affiliations](#)

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

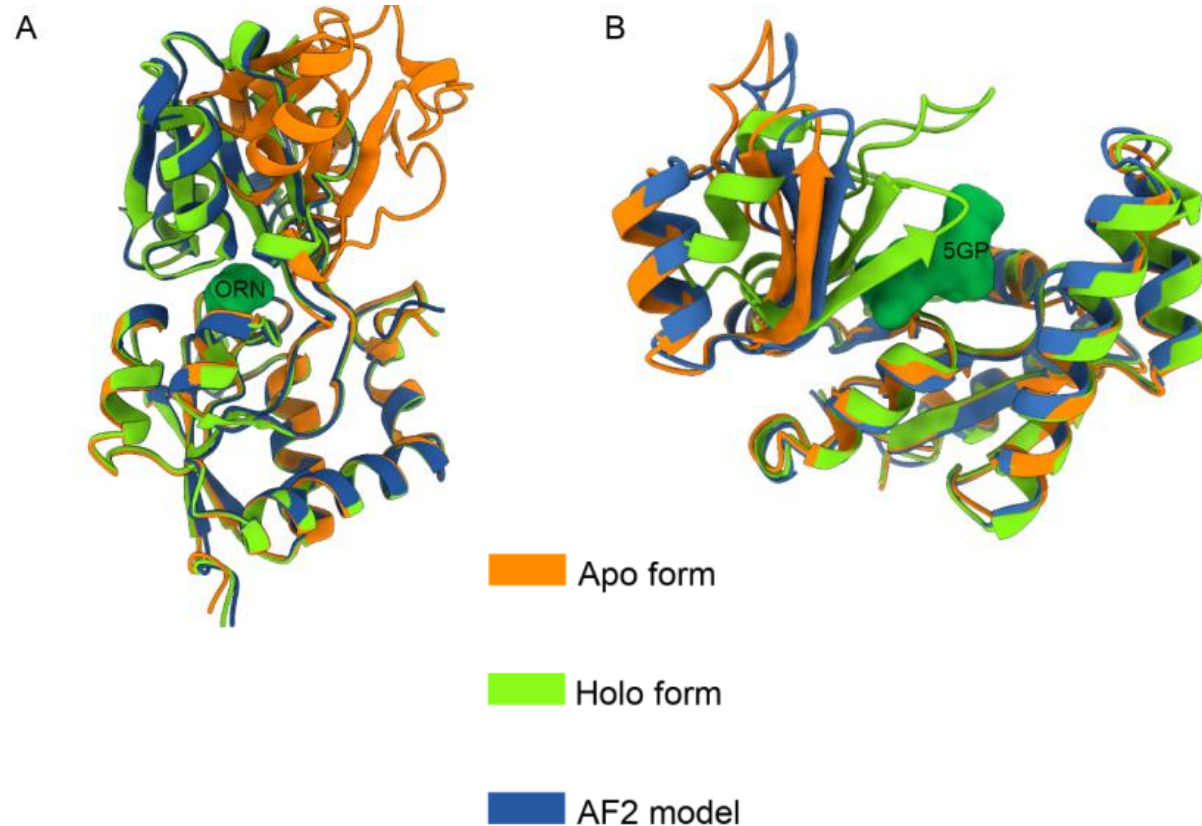
GB98 models shows mix between 3α to $\alpha+\beta$

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)



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>

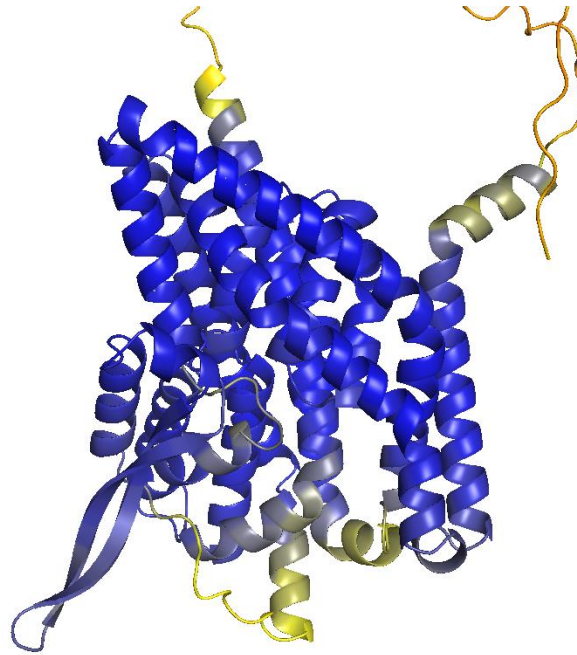
AlphaFold can predict **dynamics**

pLDDT shows flexibility

SLC1A5



6mp6
Outward-Facing



AlphaFoldDB
similar to OF

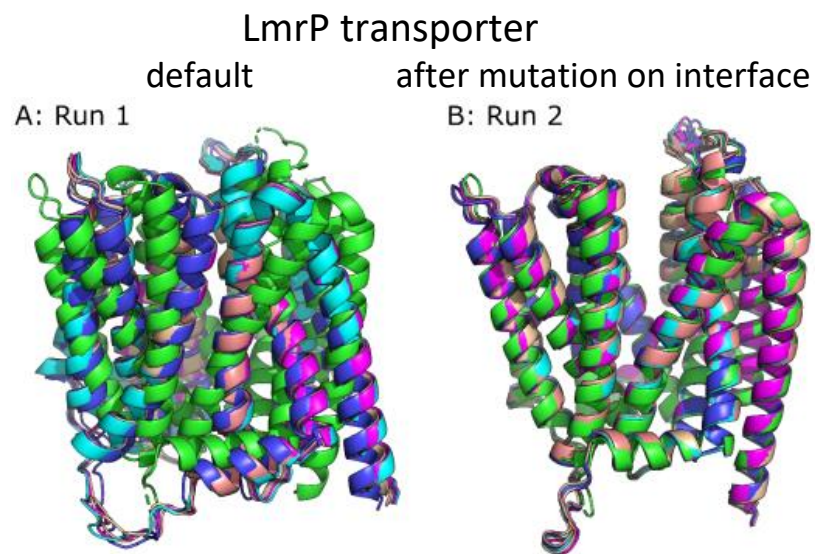


6rvx
Inward-Facing

lower pLDDT values show flexible regions

Alphafold can do **conformational changes**

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



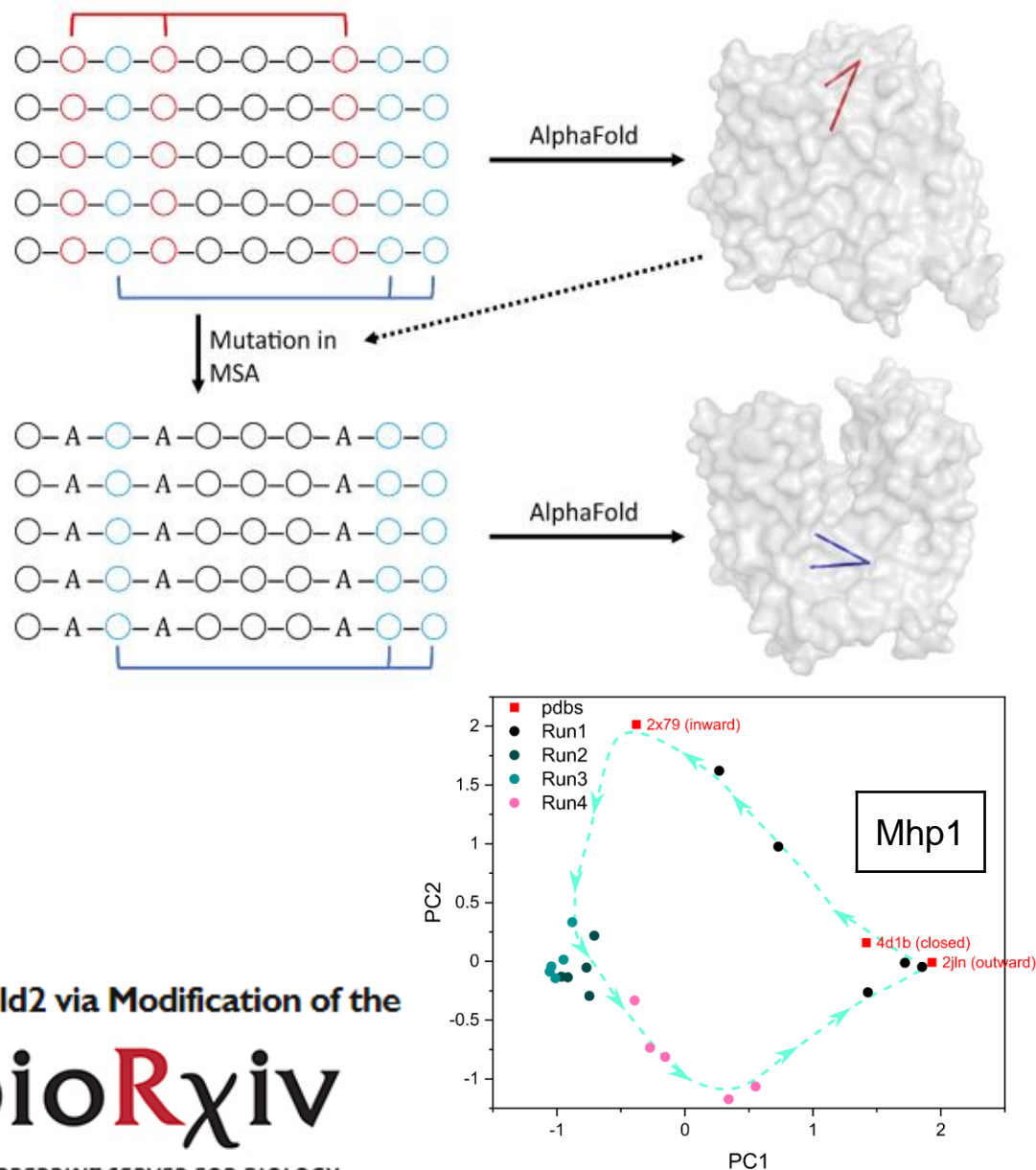
Modeling Alternate Conformations with Alphafold2 via Modification of the Multiple Sequence Alignment

Richard A. Stein, Hassane S. Mchaourab

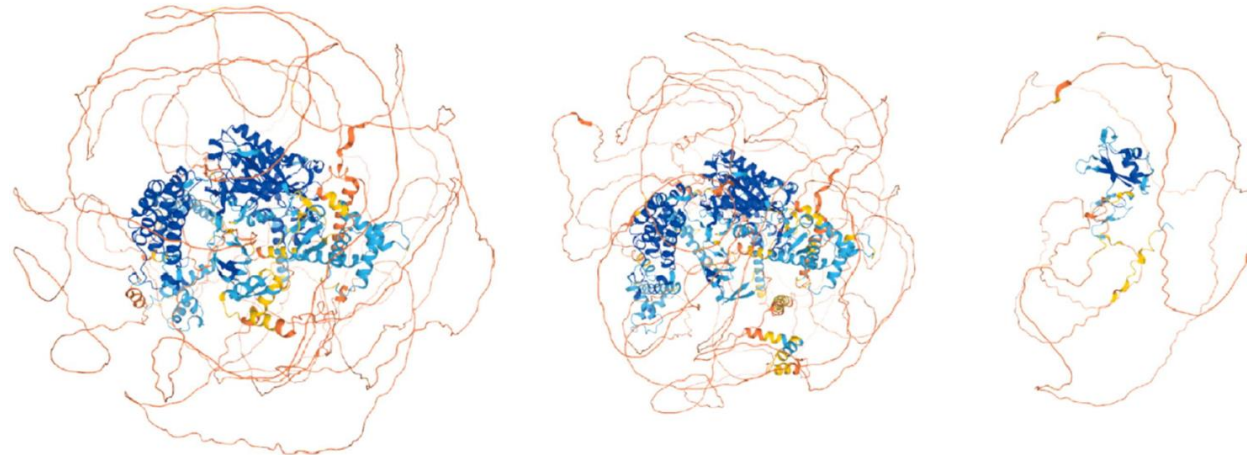
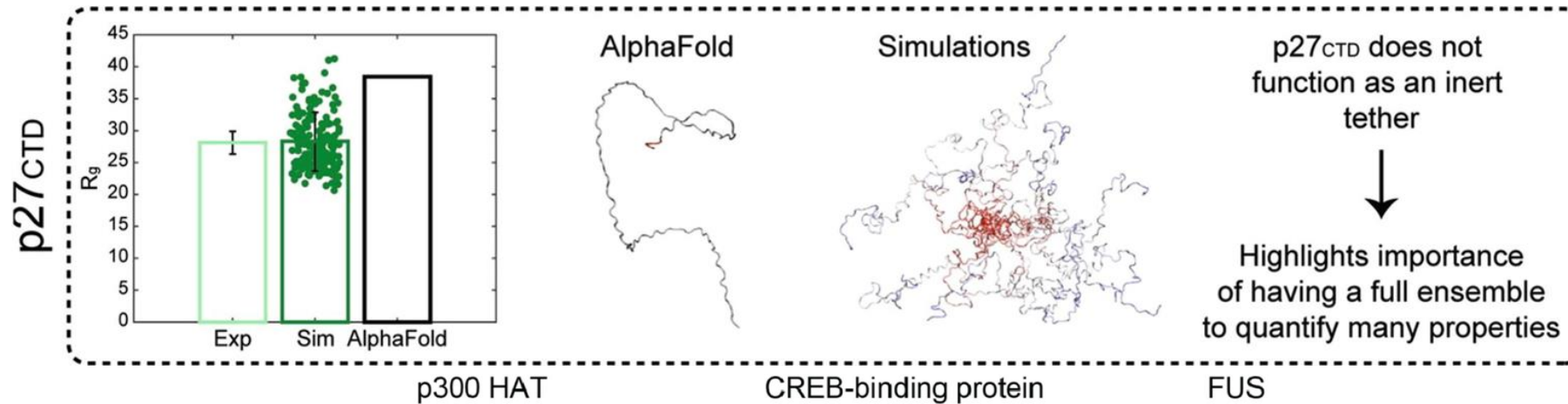
doi: <https://doi.org/10.1101/2021.11.29.470469>

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY



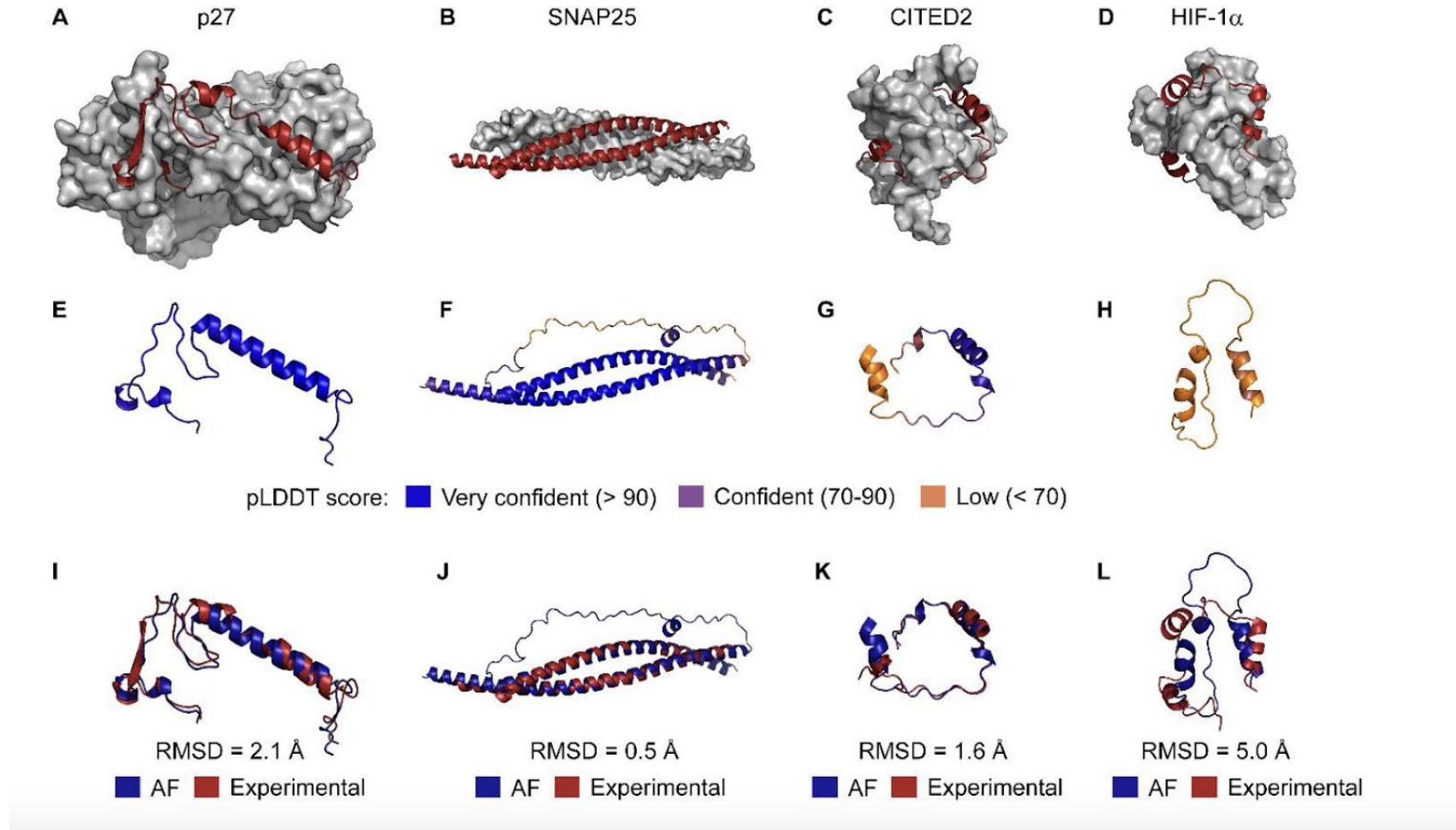
AlphaFold and Intrinsically Disordered Proteins



Model Confidence

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

AlphaFold and Intrinsically Disordered Proteins

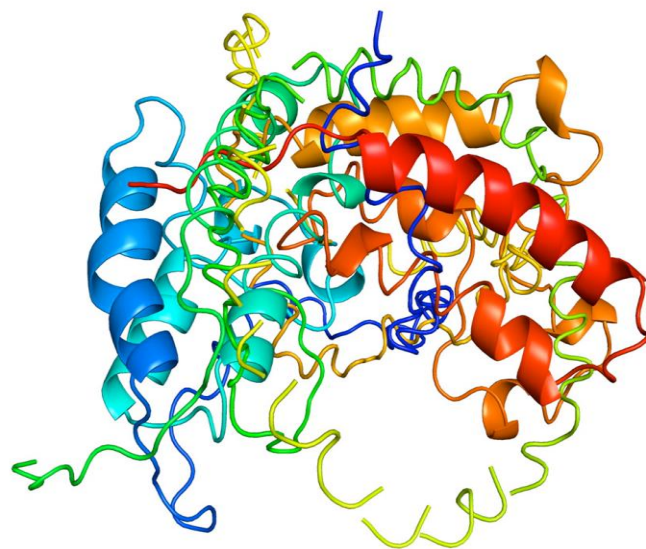


Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2

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

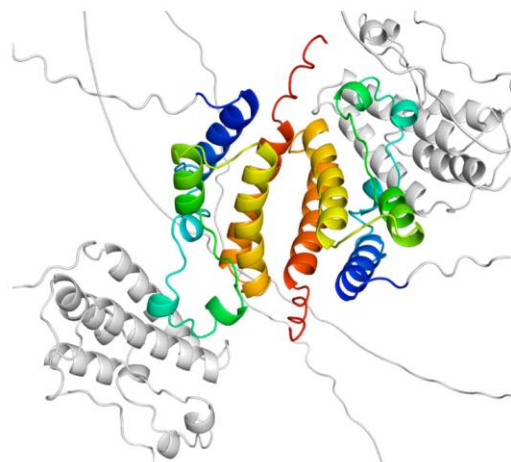
doi: <https://doi.org/10.1101/2022.02.18.481080>

AlphaFold can do multiprotein complexes

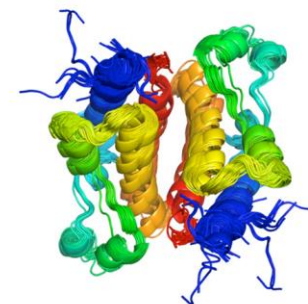


AlphaFold-Multimer v2 reproduces dimer of Bromodomains BD2 of BET proteins observed in crystal structures

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>

bioRxiv preprint doi: <https://doi.org/10.1101/2021.10.04.463034>; this version posted March 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



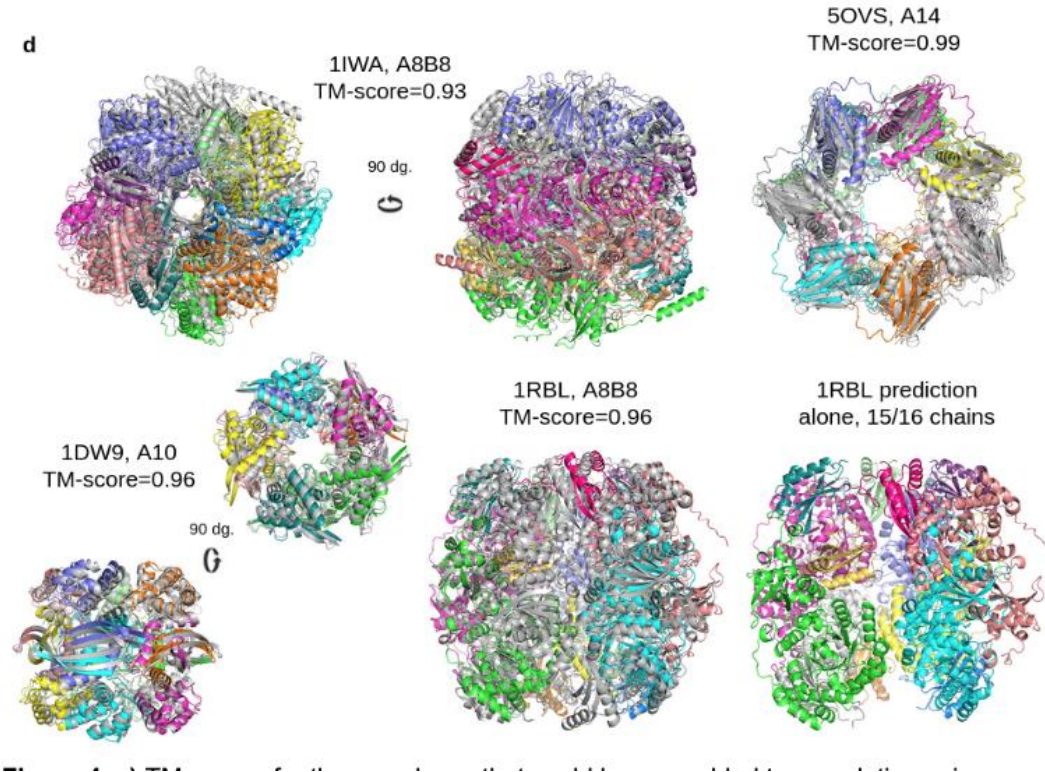
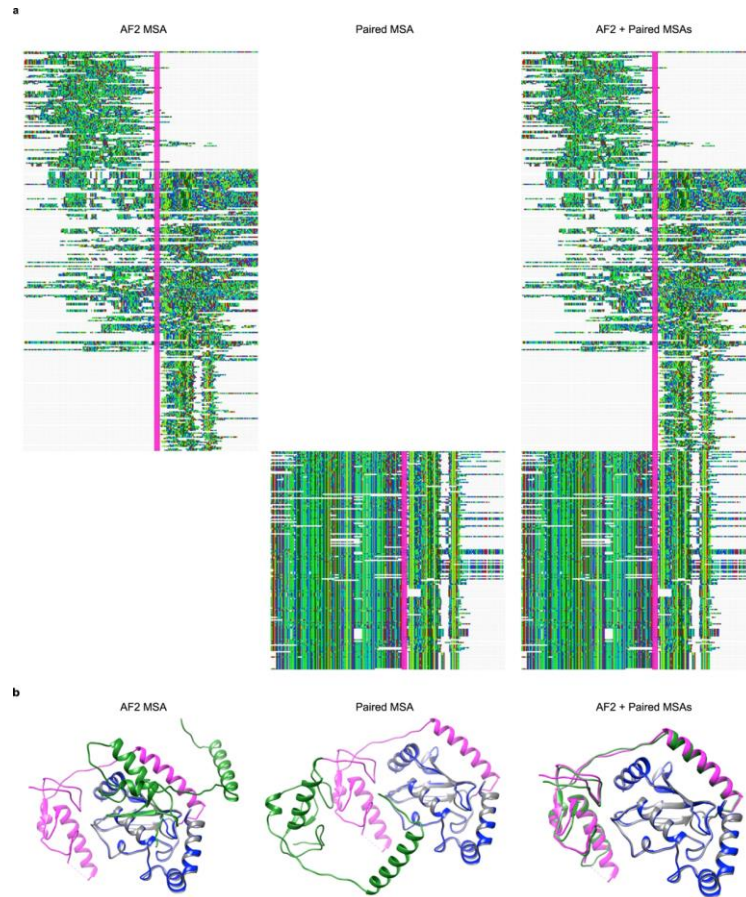
2022-03-10

Protein complex prediction with AlphaFold-Multimer

Richard Evans^{1*}, Michael O'Neill^{1*}, Alexander Pritzel^{1*}, Natasha Antropova^{1*}, Andrew Senior¹, Tim Green¹, Augustin Žídek¹, Russ Bates¹, Sam Blackwell¹, Jason Yim¹, Olaf Ronneberger¹, Sebastian Bodenstein¹, 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](#)

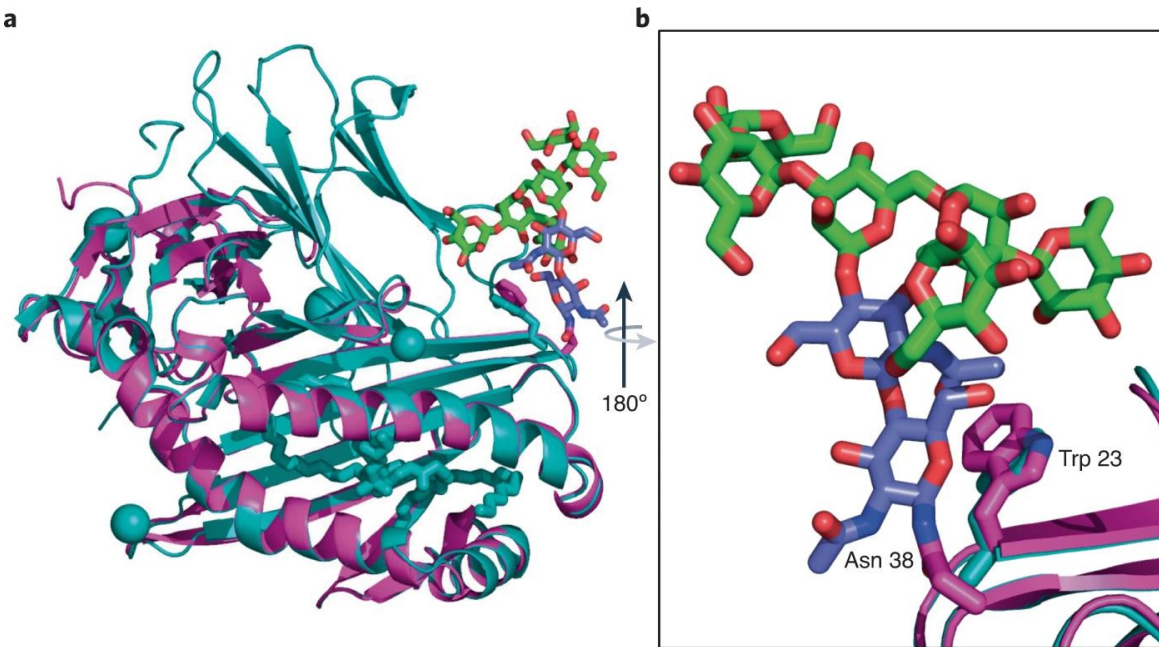
Predicting the structure of large protein complexes using AlphaFold and sequential assembly

 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) ^a



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](#)

10k Accesses | 2 Citations | 151 Altmetric | [Metrics](#)

AlphaFold can be filled with **ligands**



NKI Research | Biochemistry | Perrakis group

[Home](#) [Structures](#) [Compounds](#) **Model** [About](#) [Download](#)

P12931

Proto-oncogene tyrosine-protein kinase Src

Structure file

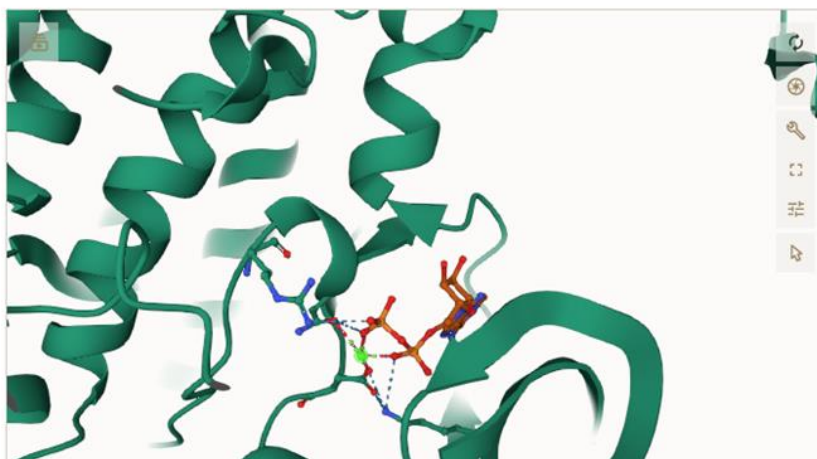
<https://alphafill.eu/v1/aff/P12931>

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	<input checked="" type="checkbox"/>
AGS -> ATP	3dqw.A	6.78	? I	1.38	<input type="checkbox"/>
AMP	3dqx.A	6.02	? H	0.57	<input type="checkbox"/>
MG	6f3f.A	1.54	C	0.10	<input checked="" type="checkbox"/>

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
<i>200 Decoys</i>								
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

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

Was Anfinsen right?

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

[Follow this preprint](#)

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?].

[Previous](#)

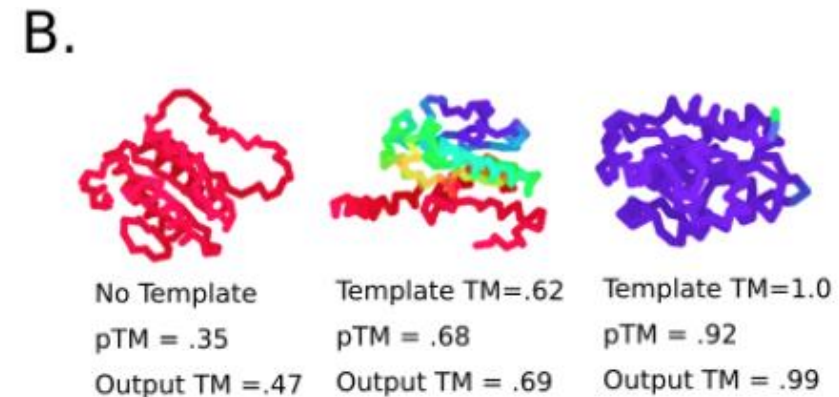
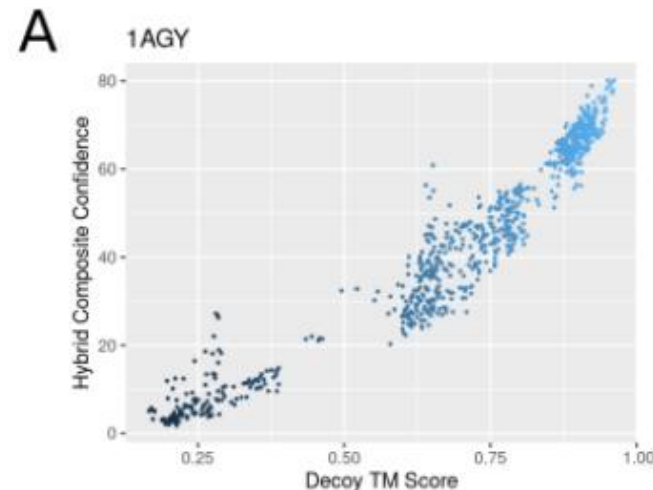
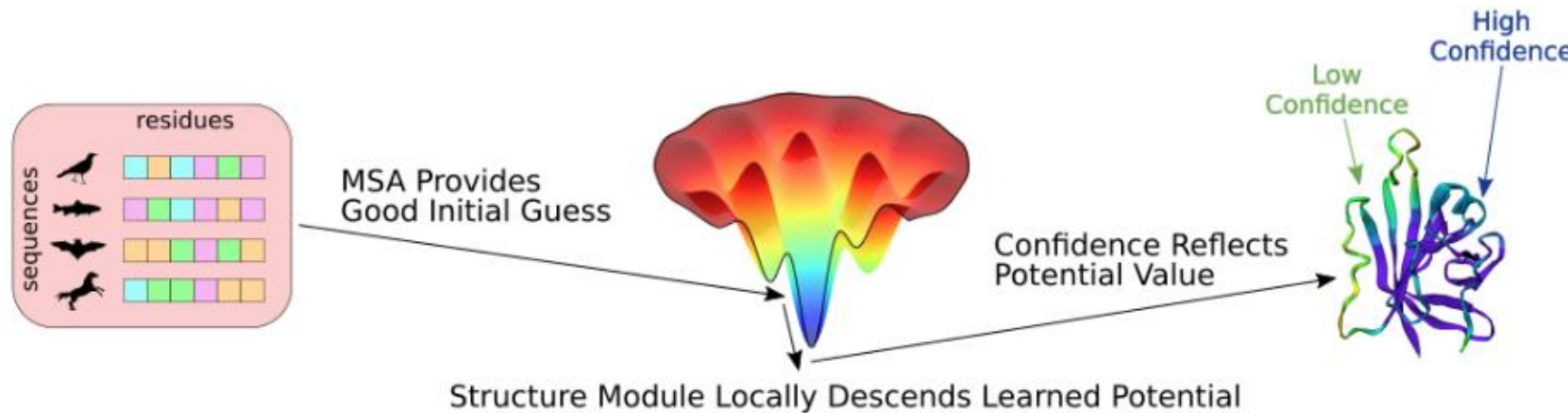
Posted March 24, 2022.

[Download PDF](#)

[Print/Save Options](#)

[Data/Code](#)


[Revision Summary](#)







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!



15th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction



Menu

- [Home](#)
- [PC Login](#)
- [PC Registration](#)
- [CASP Experiments](#)
- [CASP15 \(2022\)](#)
- [CASP14 \(2020\)](#)
- [CASP Commons \(COVID-19, 2020\)](#)
- [CASP13 \(2018\)](#)
- [CASP12 \(2016\)](#)
- [CASP11 \(2014\)](#)
- [CASP10 \(2012\)](#)
- [CASP9 \(2010\)](#)
- [CASP8 \(2008\)](#)
- [CASP7 \(2006\)](#)
- [CASP6 \(2004\)](#)

CASP15 Experiment

Detailed description of the experiment

CASP (Critical Assessment of Structure Prediction) is a community wide experiment to determine and advance the state of the art in modeling protein structure from amino acid sequence. Every two years, participants are invited to submit models for a set of proteins for which the experimental structures are not yet public. In the latest CASP round, CASP14, nearly 100 groups from around the world submitted more than 67,000 models on 84 modeling targets. Independent assessors then compare the models with experiment. Assessments and results are published in [a special issue of the journal PROTEINS](#).

[Goals](#)[Categories](#)[Timetable](#)[Registration](#)[Targets](#)[Format](#)[Assessment](#)[Results](#)[Conference](#)[Organizers](#)

Background and goals

CASP14 (2020) saw an enormous jump in the accuracy of single protein and domain models such that many are competitive with experiment. That advance is largely the result of the successful application of deep learning methods, particularly by the AlphaFold and, since that CASP, RosettaFold. As a consequence, computed protein structures are becoming much more widely used in a broadening range of applications. CASP has responded to this new landscape with a revised set of modeling categories. Some old categories have been dropped (refinement, contact prediction, and aspects of model accuracy estimation) and new ones have been added (RNA structures, protein ligand complexes, protein ensembles, and accuracy estimation for protein complexes). We are also strengthening our interactions with our partners CAPRI and CAMEO. We hope that these changes will maximize the insight that CASP15 provides, particularly in new applications of deep learning.

Modeling categories


The core of CASP remains the same: blind testing of methods with independent assessment against experiment to establish the state-of-art in modeling proteins and protein complexes. CASP15 will include following categories.

Message Board

April 4 - start of CASP15 registration
[Dear CASPers, CASP15 registration will open on April 4, server testing on April 18, and the first target will be released on May 2, 2022. The experiment will run May-August and culminate with a con ...](#)

CASP15 call for targets
[CASP \(Critical Assessment of protein Structure Prediction\) is in search for targets for the upcoming CASP15 modeling experiment \(starting in May 2022\). CASP community experiments aim to advance the st ...](#)

UNIPROT survey
[A message from Alex Bateman \(EMBL-EBI\):](#)



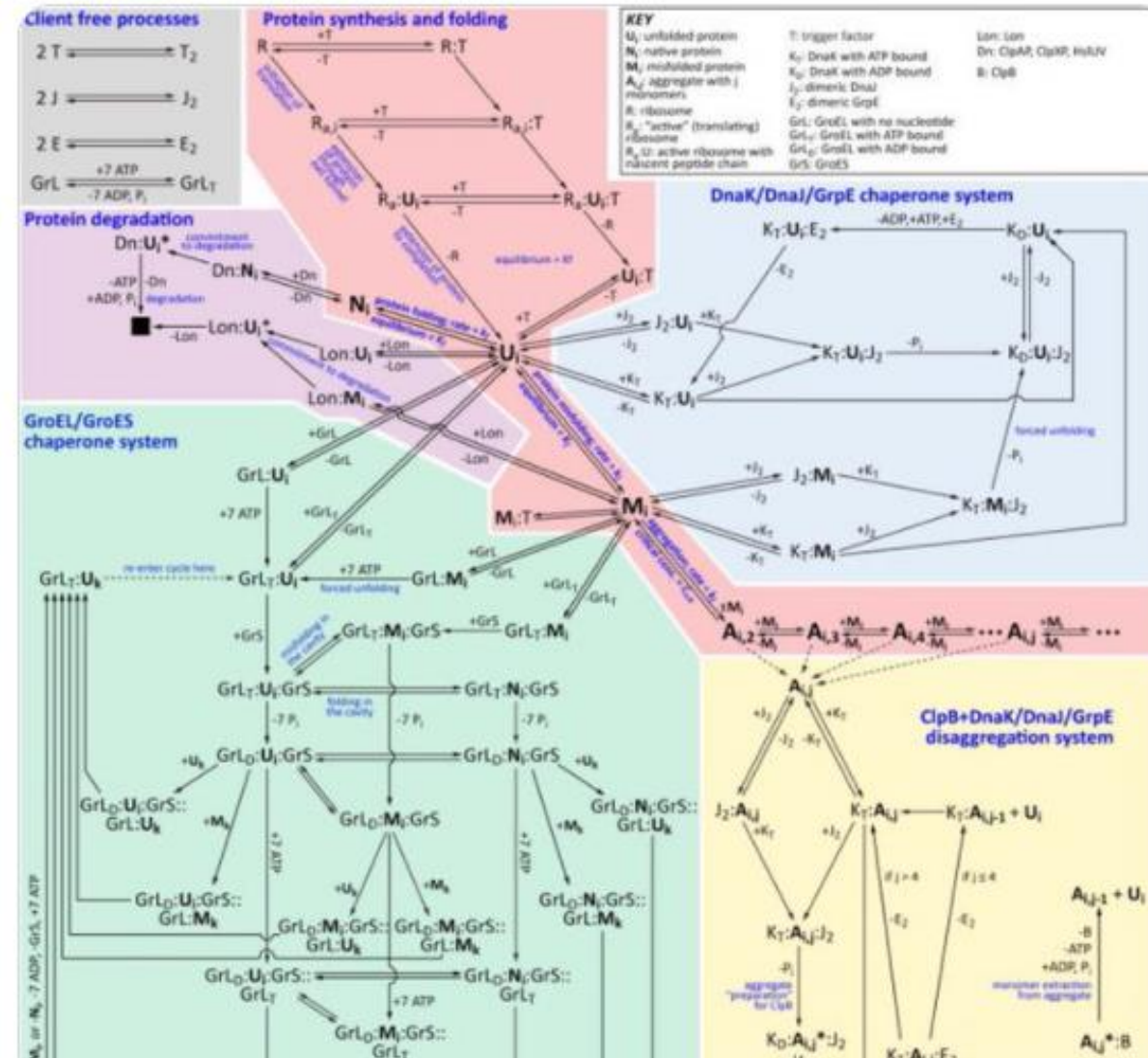
Thank you for your
attention.

Any questions?



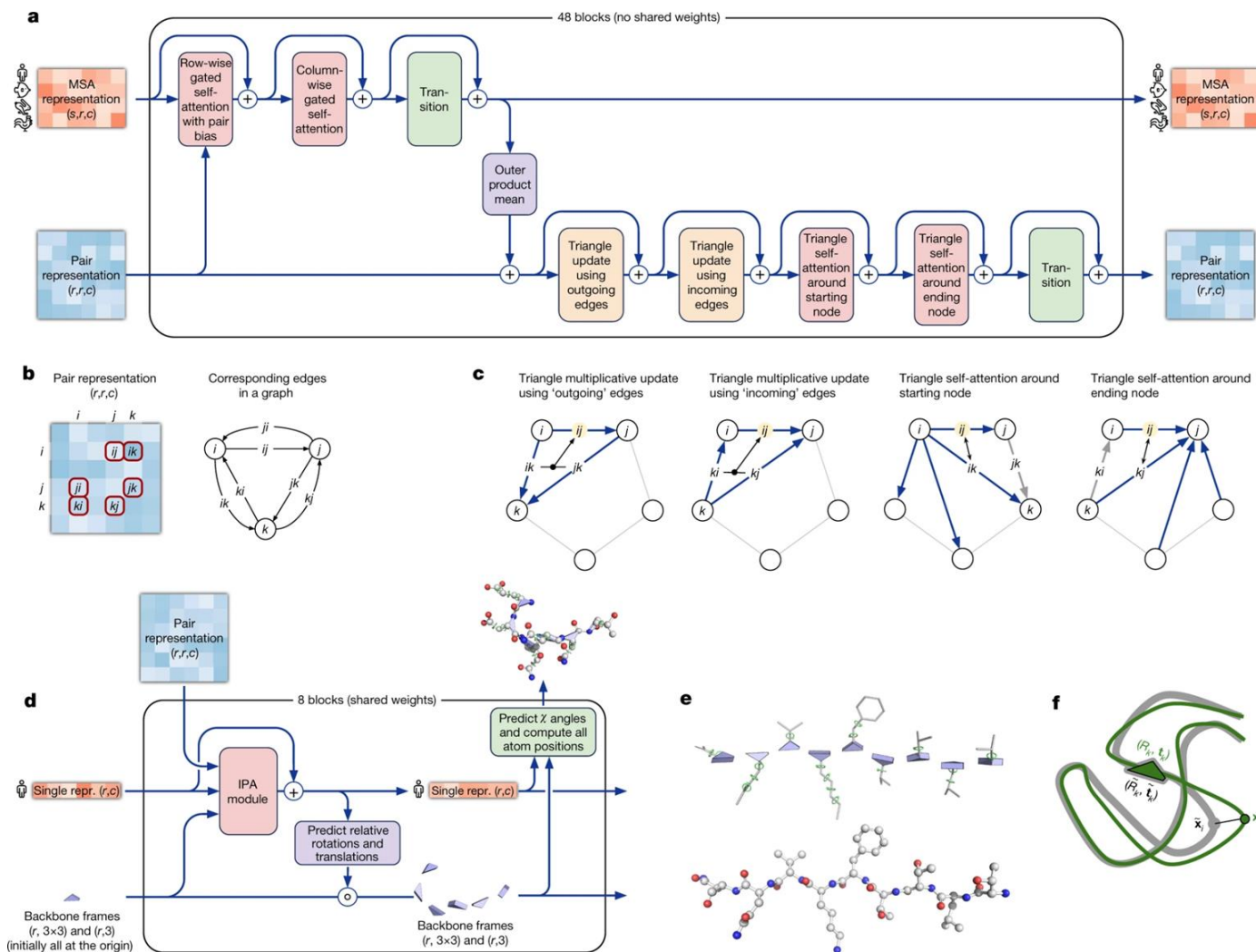


[Přeložit Tweet](#)

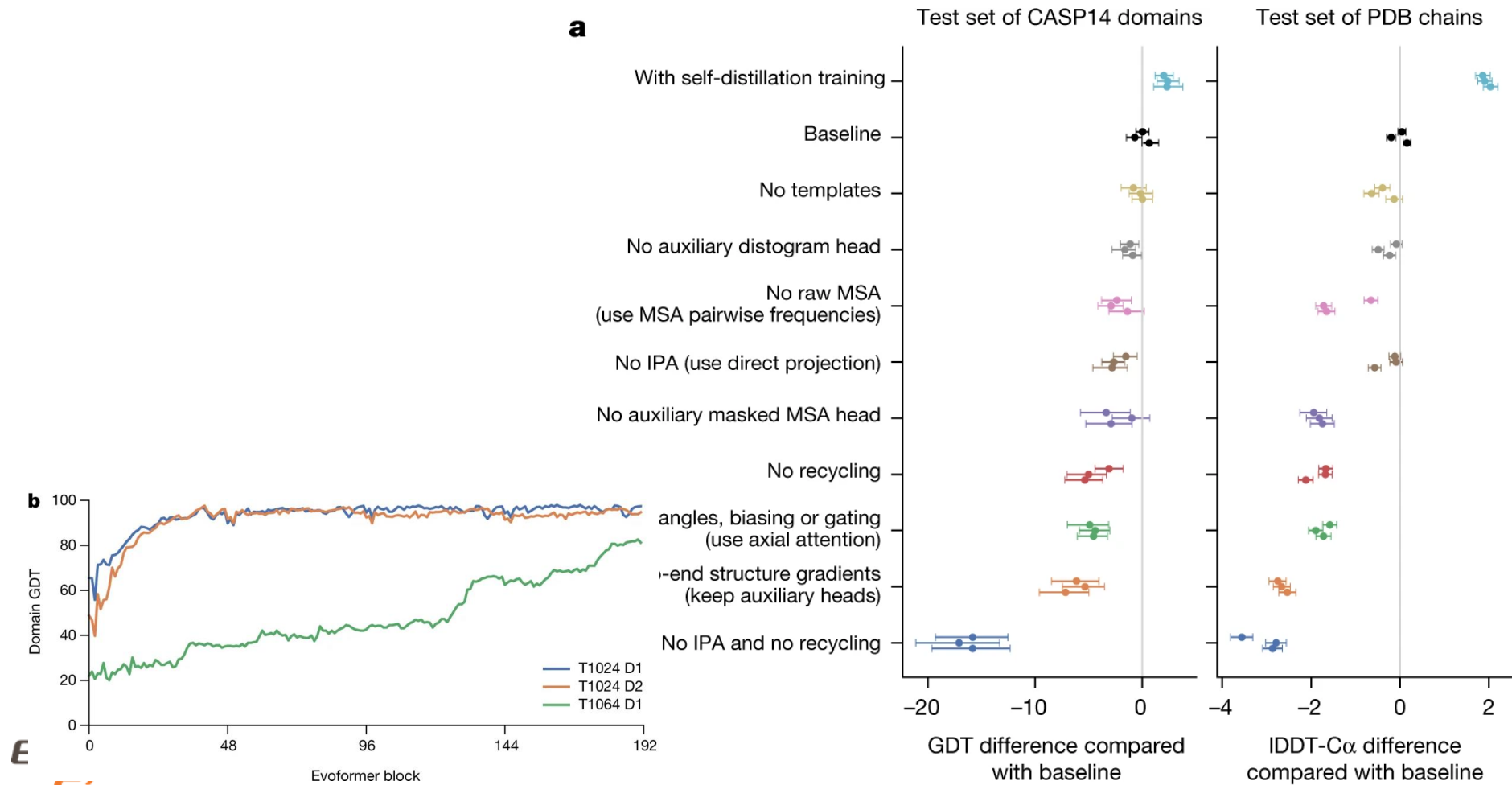


Extra slides

Architectural details.

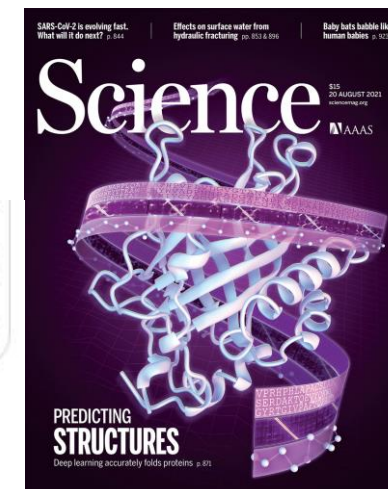
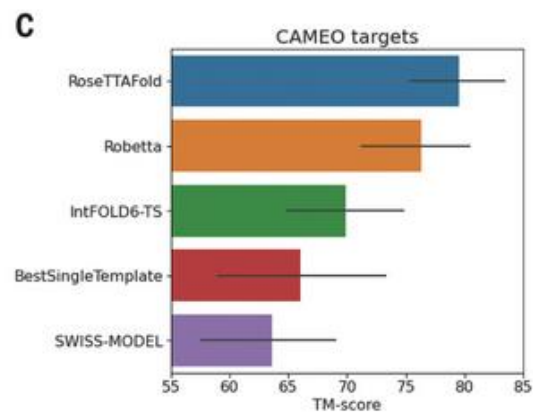
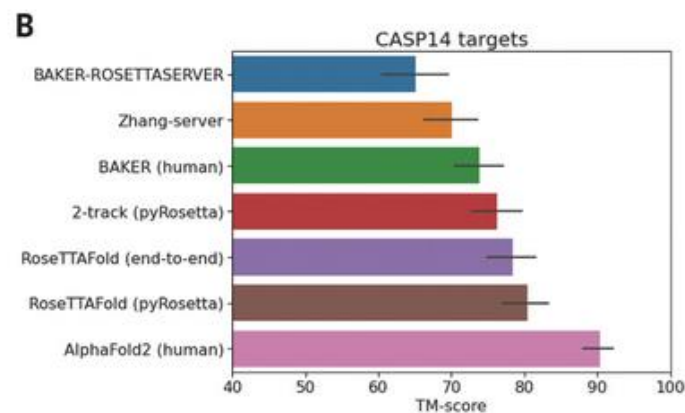
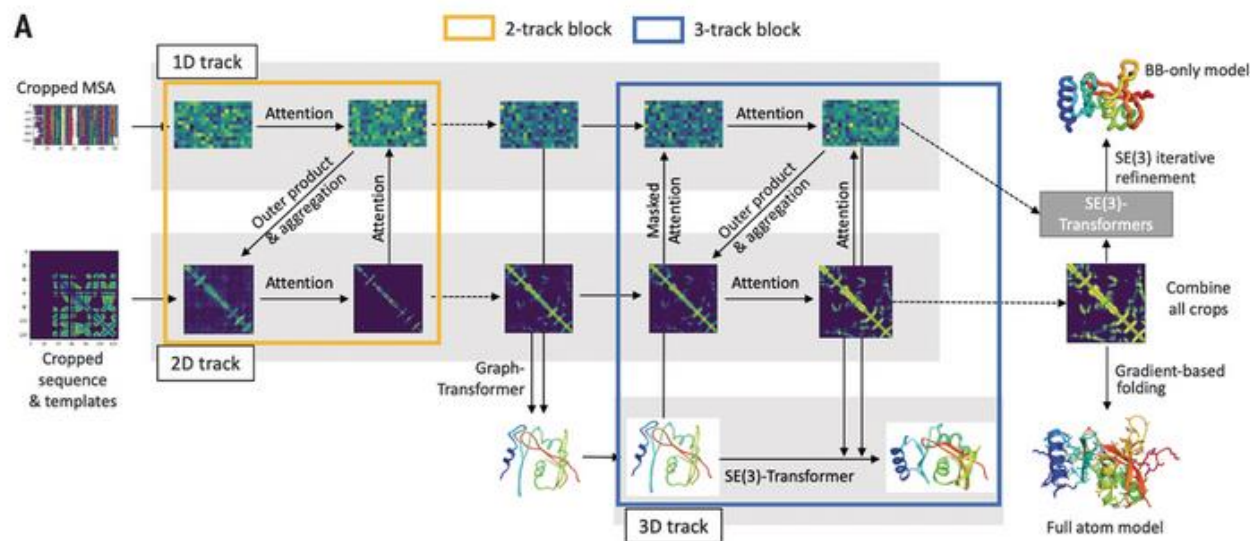


Interpreting the neural network



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

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



Alphafold can work with **orphan sequences**

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

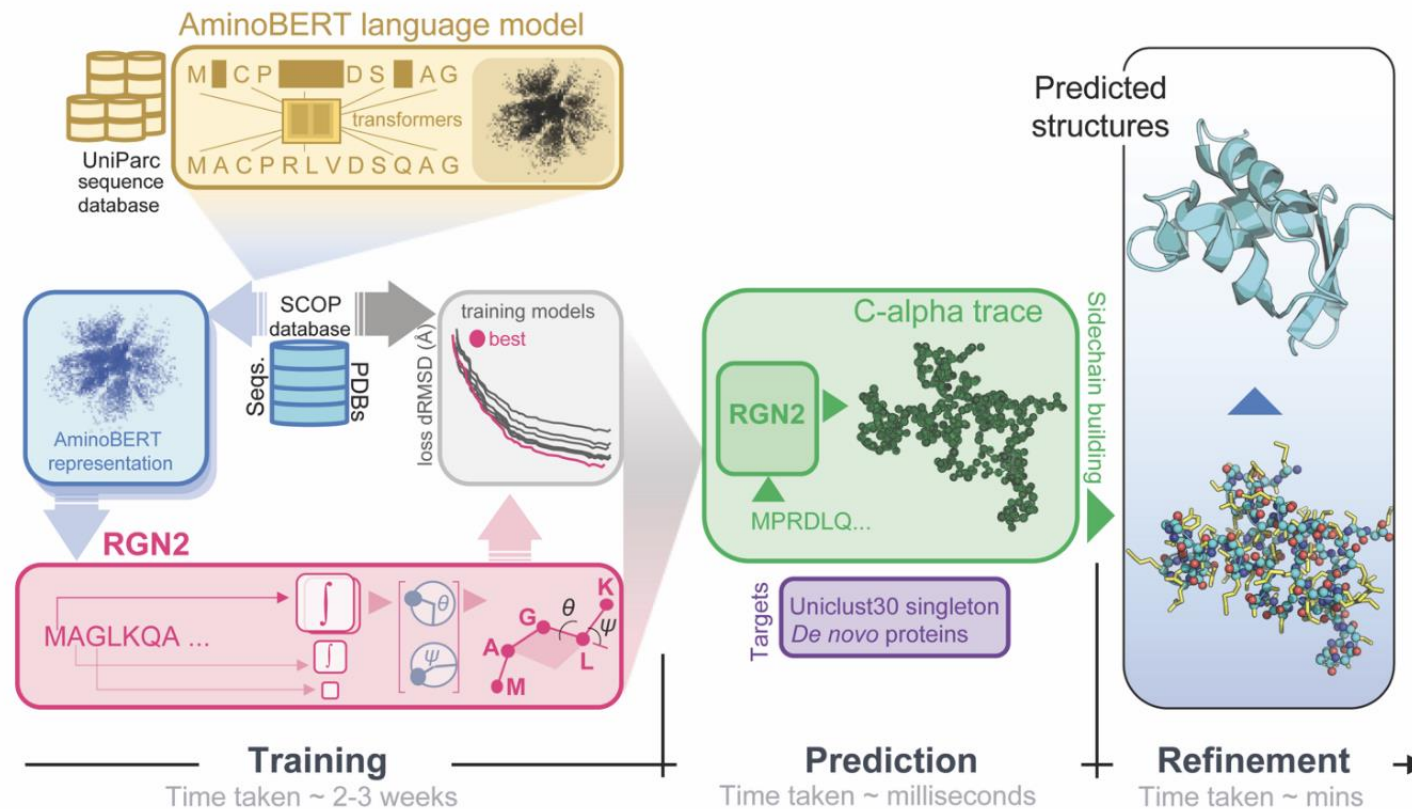



Figure 1. Organization and application of RGN2. RGN2 combines a Transformer-based protein language model (AminoBERT) with a recurrent geometric network that utilizes Frenet-Serret frames to generate the backbone structure of a protein. Placement of side chain atoms and refinement of hydrogen-bonded networks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


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***

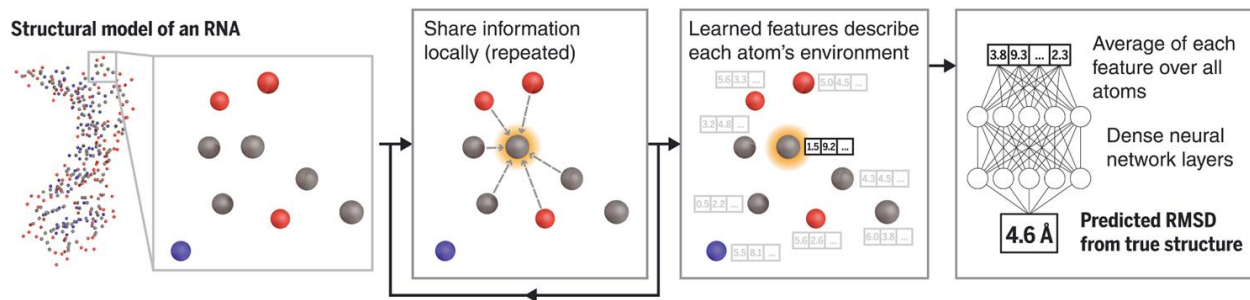
Department of Computer Science
University College London
Gower St, London WC1E 6BT
d.t.jones@ucl.ac.uk

ABSTRACT

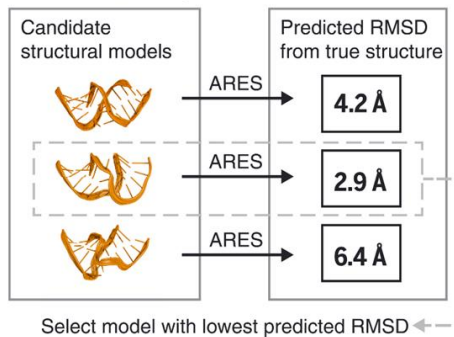
The prediction of protein structure and the design of novel protein sequences and structures have long been intertwined. The recently released AlphaFold has heralded a new generation of accurate protein structure prediction, but the extent to which this affects protein design stands yet unexplored. Here 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.

Geometric deep learning of RNA structure

A ARES predicts the accuracy of a structural model, given only atomic coordinates and element types



B RNA structure prediction with ARES



C Training set: 18 older, smaller RNA structures



D Benchmark sets: newer, larger RNA structures

