Google It!

AlphaFold and its impact on obtaining protein structures

Pablo Ricardo Arantes, PhD.

Slides
Jupyter Notebook

https://github.com/pablo-arantes/BIEN-249

SCAN ME

AlphaFold2

Article

Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2

Received: 11 May 2021

Accepted: 12 July 2021

Published online: 15 July 2021

John Jumper[1,4]✉, Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4]✉

The News

Science

nature

Explore our content | Journal information | Subscribe

nature › news › article

NEWS · 30 NOVEMBER 2020

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

Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

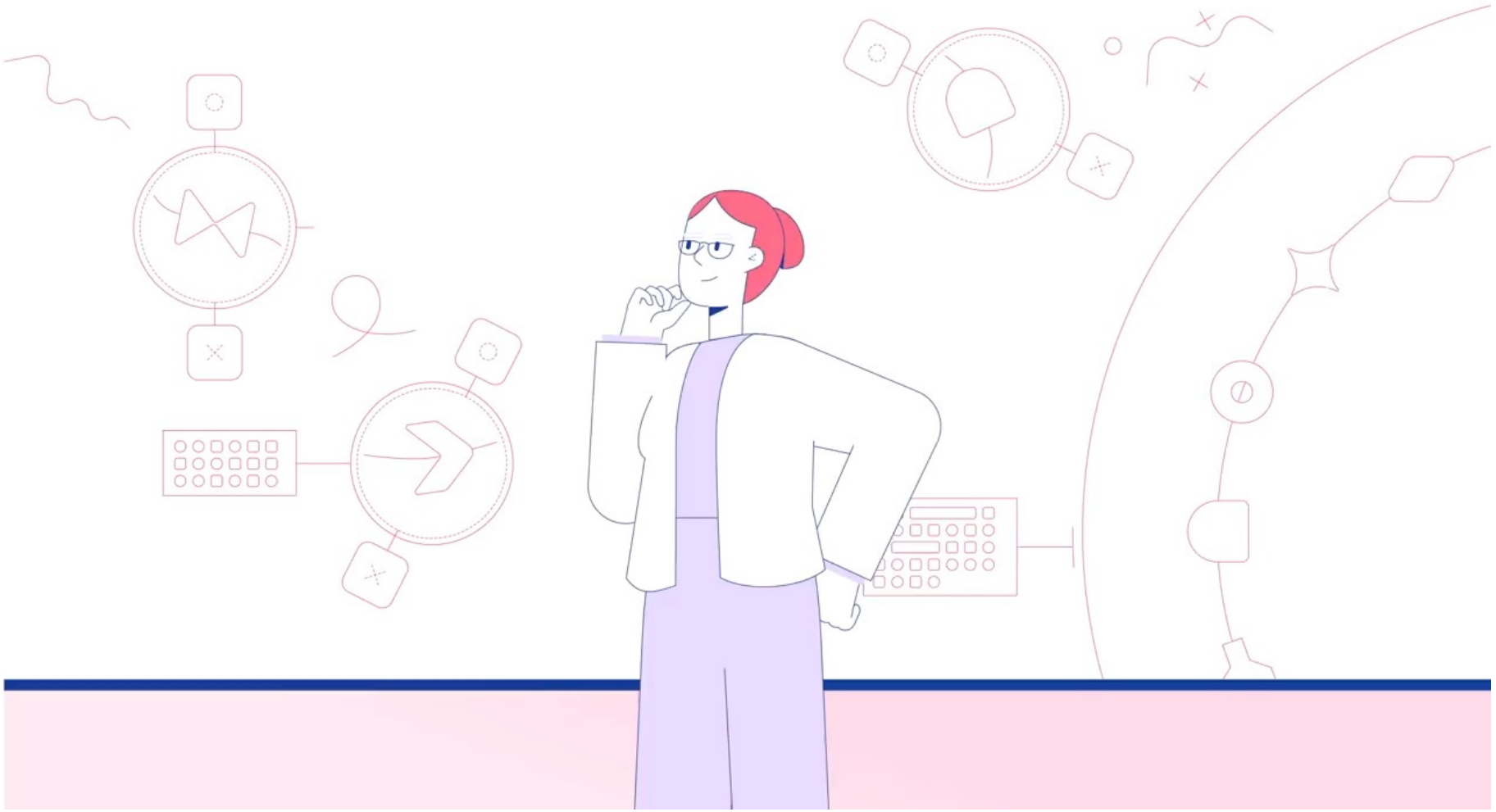
Ewen Callaway

By Robert F. Service | Nov. 30, 2020 , 10:30 AM

https://www.nature.com/articles/d41586-020-03348-4

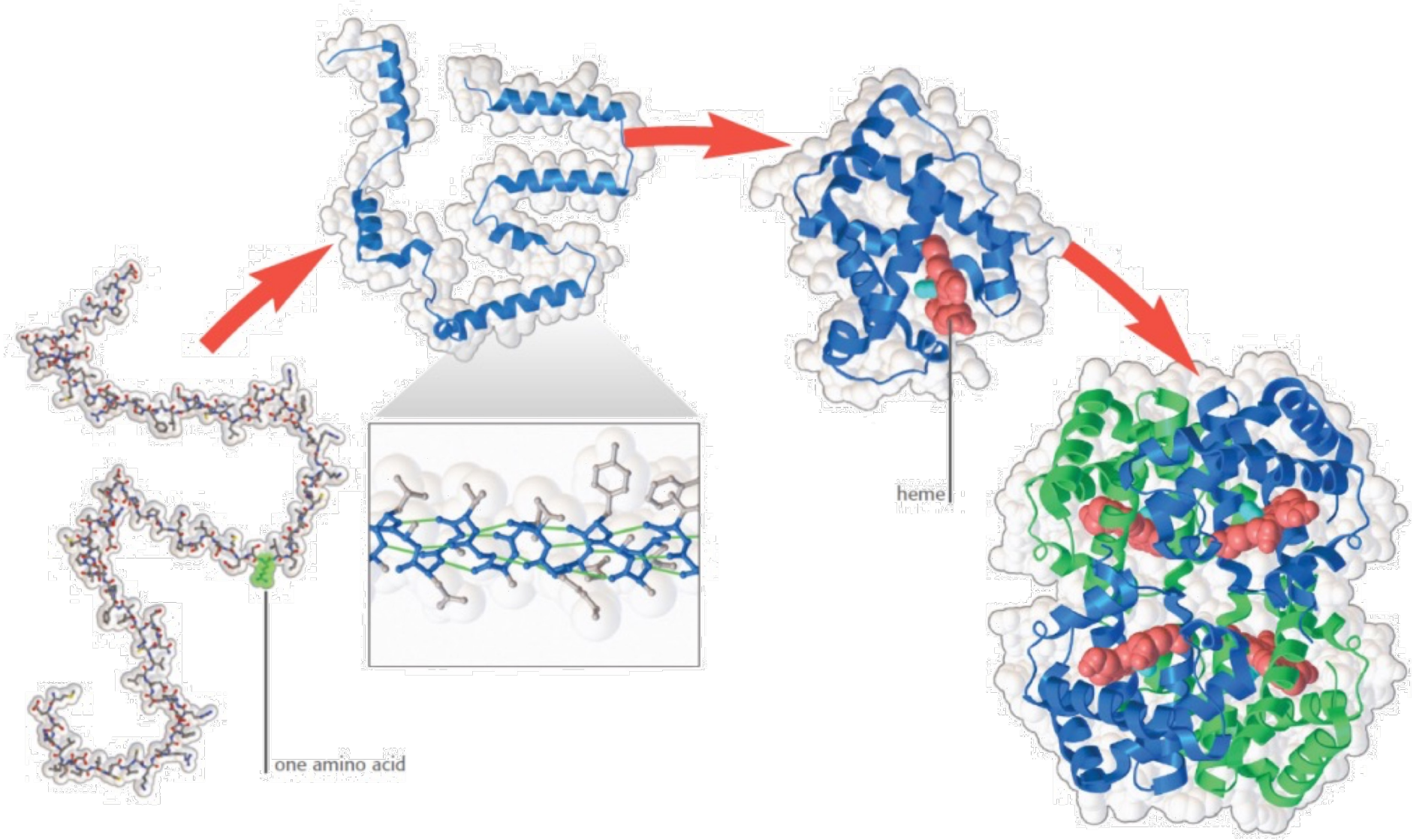
https://www.sciencemag.org/news/2020/11/ai-changed-ai-triumphs-solving-protein-structures

The News



https://youtu.be/KpedmJdrTpY

Proteins



PDB-101: Educational Resources. “What is a protein?”

"Structural Biology is the study of the structure and dynamics of macro-biomolecules, particularly proteins and nucleic acids, and how changes in their structure affect their function. It incorporates principles of molecular biology, biochemistry and biophysics." (Nature, 2019)

Structure = Function

Marcus Vitruvius Pollio (I a.C.)

Folding

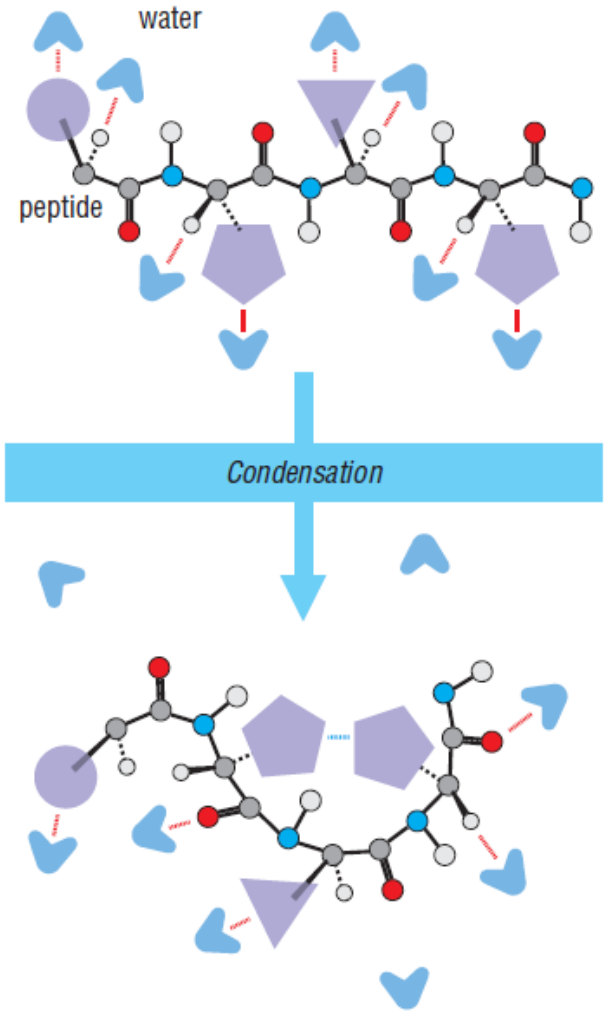
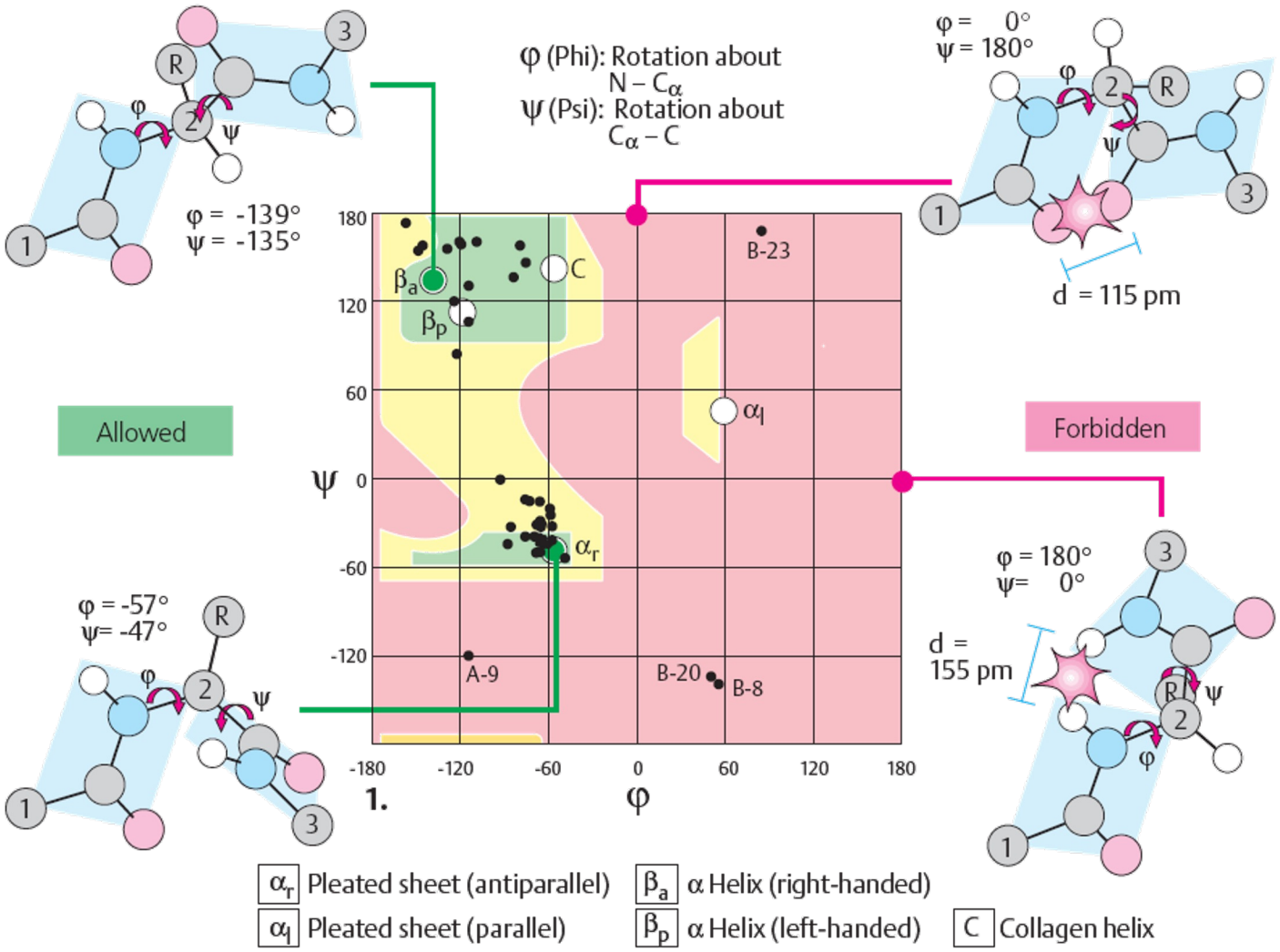


Figure 1-23 Highly simplified schematic representation of the folding of a polypeptide chain in water In the unfolded chain, side-chain and main-chain groups interact primarily with water, even if they are hydrophobic and the interaction is unfavorable. Burying the hydrophobic groups in the interior of a compact structure enables them to interact with each other (blue line), which is favorable, and leaves polar side chains on the surface where they can interact with water (red lines). The polar backbone groups that are buried along with the hydrophobic side chains must make hydrogen bonds to each other (not shown), as bulk water is no longer available.

Ramachandran Plot



The Protein Folding Problem(s)

Thermodynamics

How does structure emerge from the interactions between amino acids?

Folding code

Kinetics

How is structure achieved so quickly?

Folding rate

Computational

How to predict 3D structures starting from amino acids sequence?

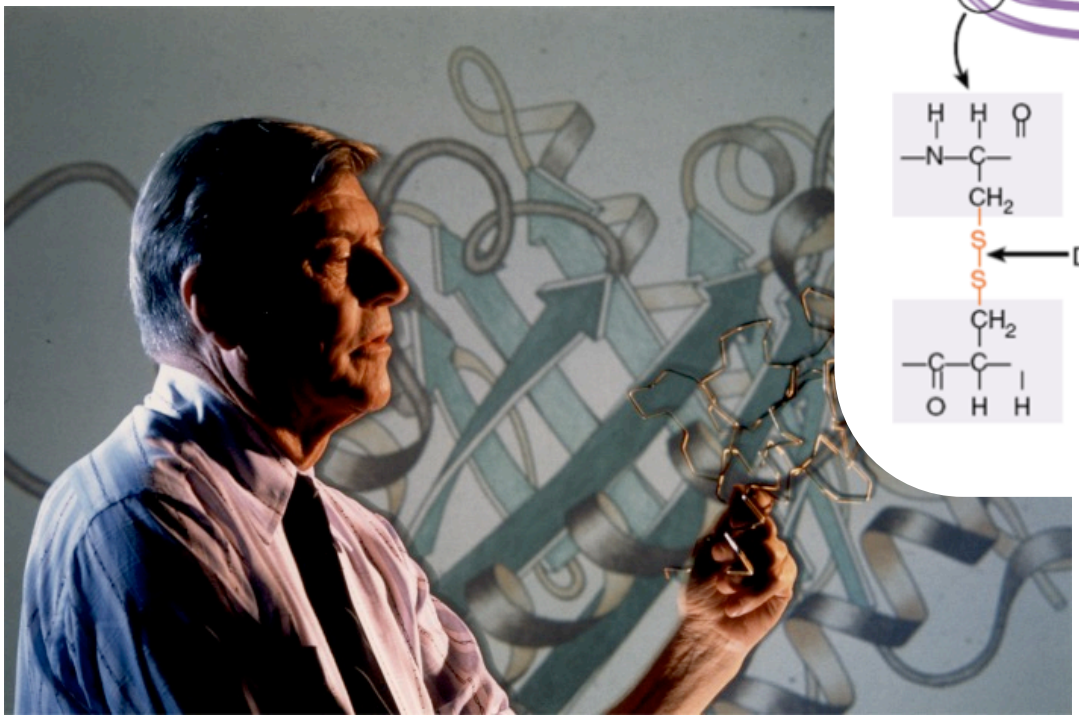
Protein Structure Prediction

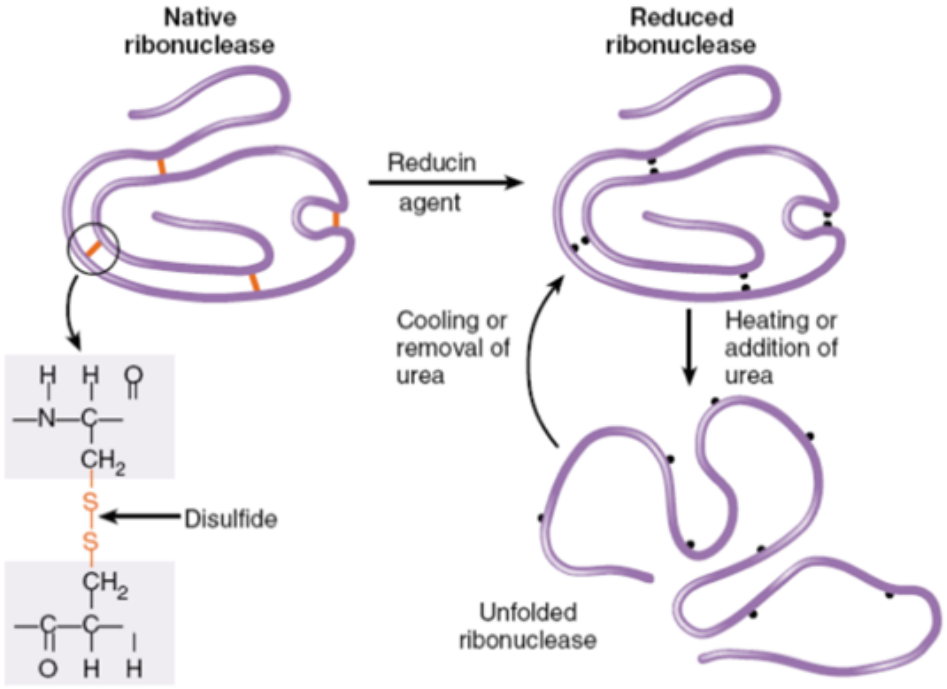
Liu et al. in Kaumaya, *Protein Engineering*, Intech (2012); Dill et al., *Curr Opin Struct Biol*, 17, 342-346 (2007)

The Protein Folding Problem(s)

Anfinsen - 1965
(Nobel - 1972)

Christian Boehmer Anfinsen, Jr.





https://profiles.nlm.nih.gov/spotlight/kk

The Protein Folding Problem(s)

Levinthal - 1968
Levinthal's paradox

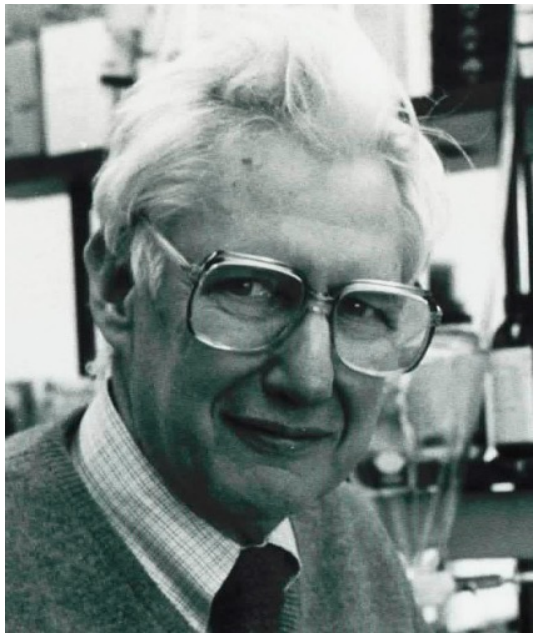
100 amino acids: $3^{100} = 5 \times 10^{47}$ configurations

Sampling 10^{13} per second (3×10^{20} per year)

10^{27} years

Universe: 2×10^{10} years

Cyrus Levinthal



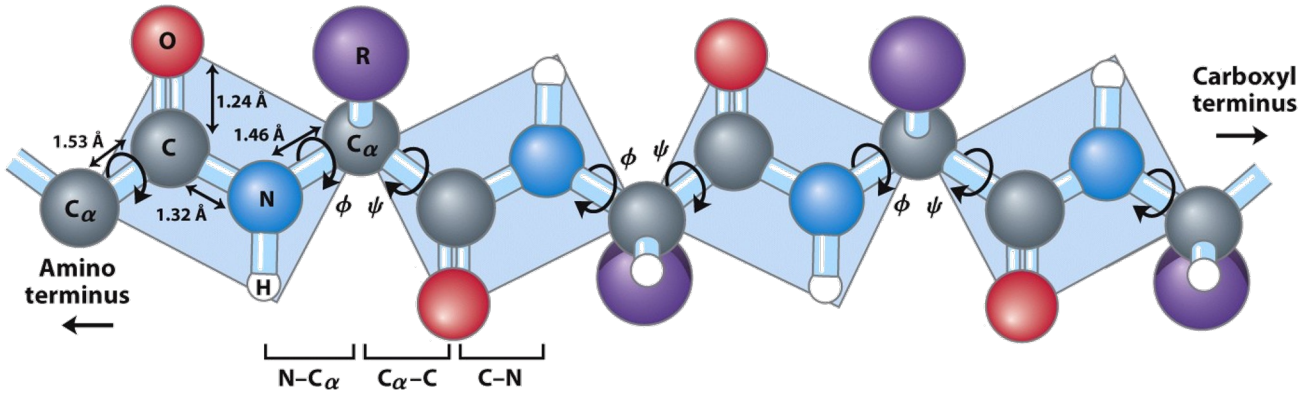
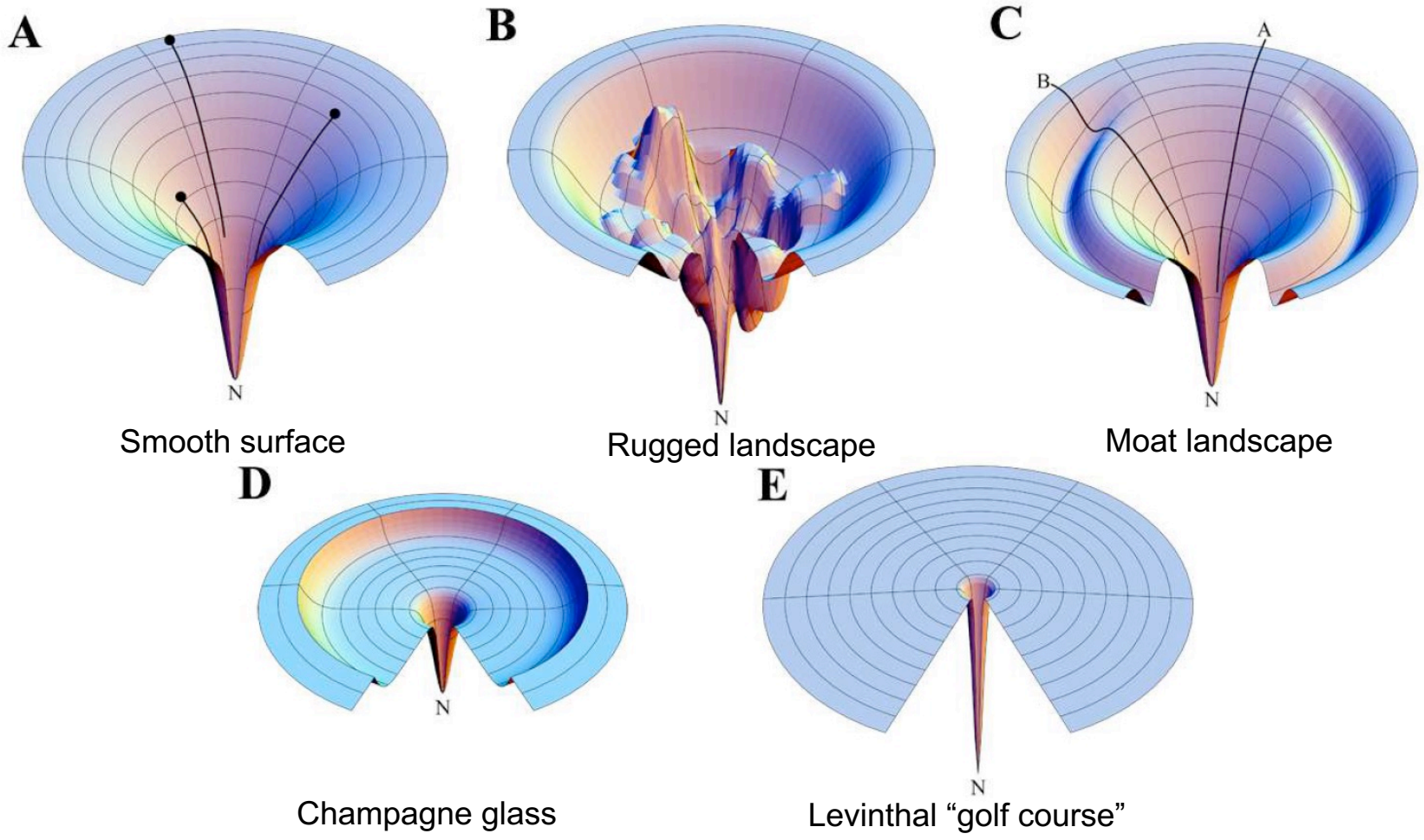


Figure 4-2b
Lehninger Principles of Biochemistry, Fifth Edition

Macagno et al., *Biographical Memoir: Cyrus Levinthal (1922-1990)*, NAS (2014); Zwanzig et al., *PNAS* 89, 20-222 (1992)

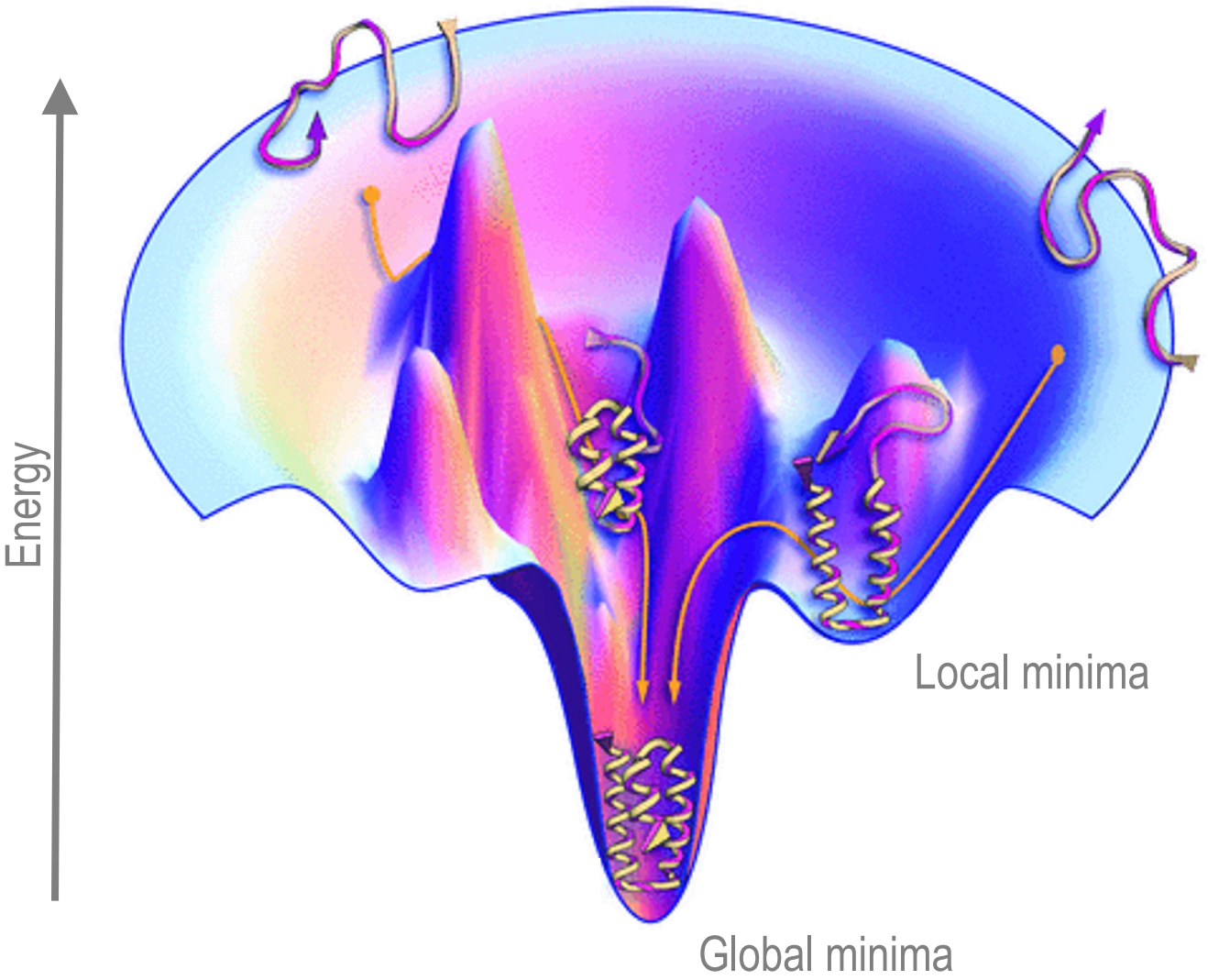
The Protein Folding Problem(s)

Depending on the folding rate, properties, flexibility and the native structure of the proteins, folding energy landscapes are divided into the following types:



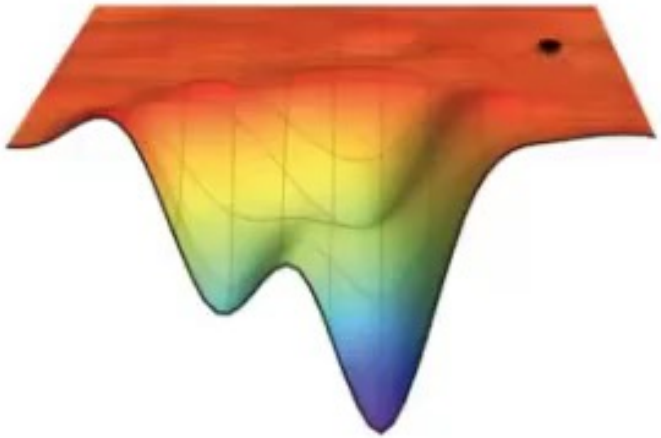
Liu et al. in Kaumaya, *Protein Engineering*, Intech (2012); Dill et al., *Curr Opin Struct Biol*, 17, 342-346 (2007)

The Protein Folding Problem(s)



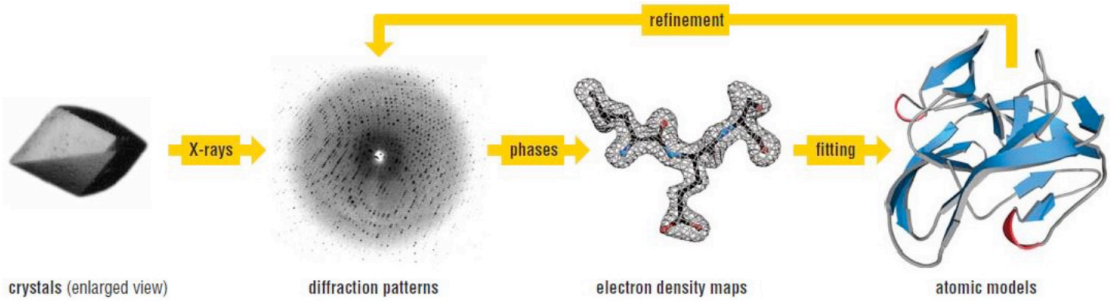
Dill & MacCallum (2012) Science 338:1042-6.

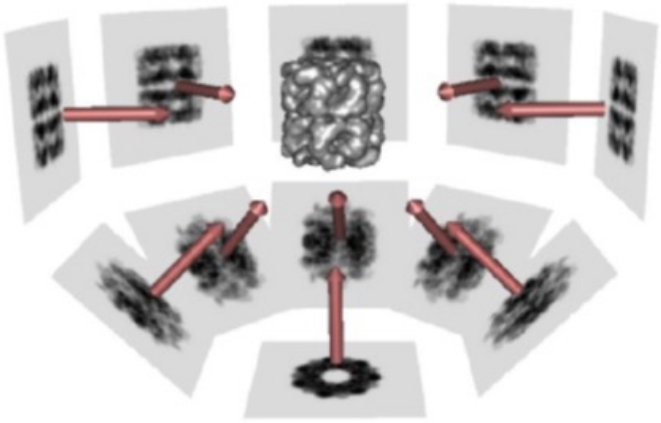
The Protein Folding Problem

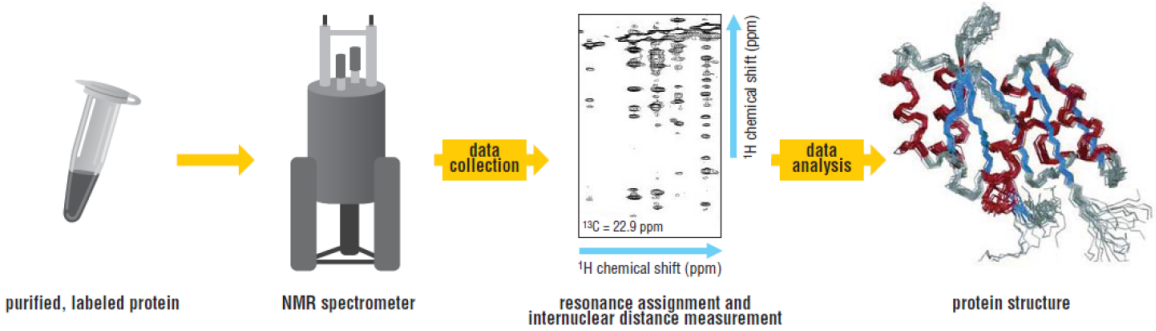


Liu et al. in Kaumaya, *Protein Engineering*, Intech (2012); Dill et al., *Curr Opin Struct Biol*, 17, 342-346 (2007)

Structural Solving







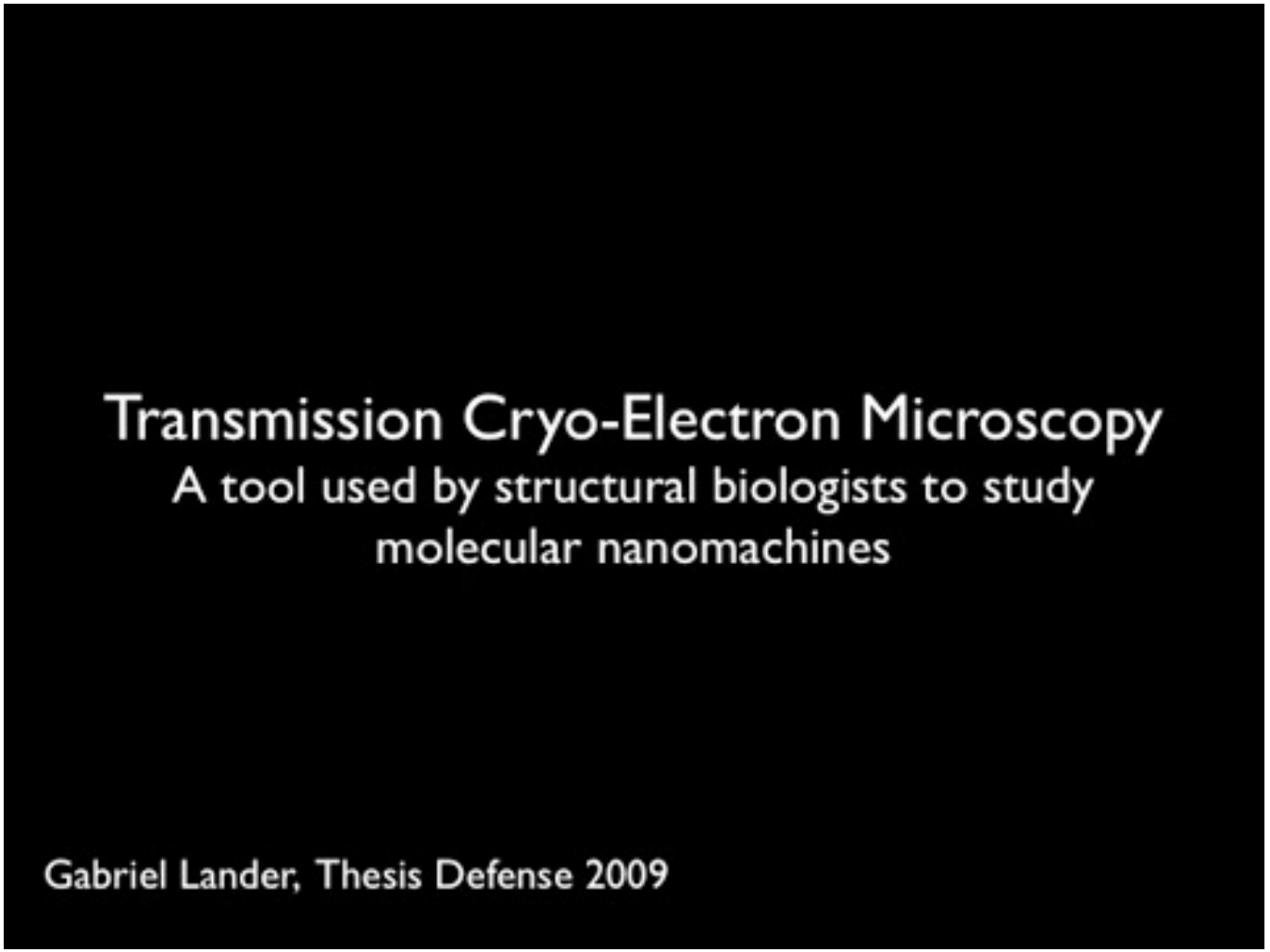
Petsko & Ringe, *Protein Structure and Function,* Blackwell (2004); Nogales & Scheres, *Mol Cell* 58, 677-89 (2015).

Structural Solving

https://youtu.be/uqQlwYv8VQI

Petsko & Ringe, *Protein Structure and Function,* Blackwell (2004); Nogales & Scheres, *Mol Cell* 58, 677-89 (2015).

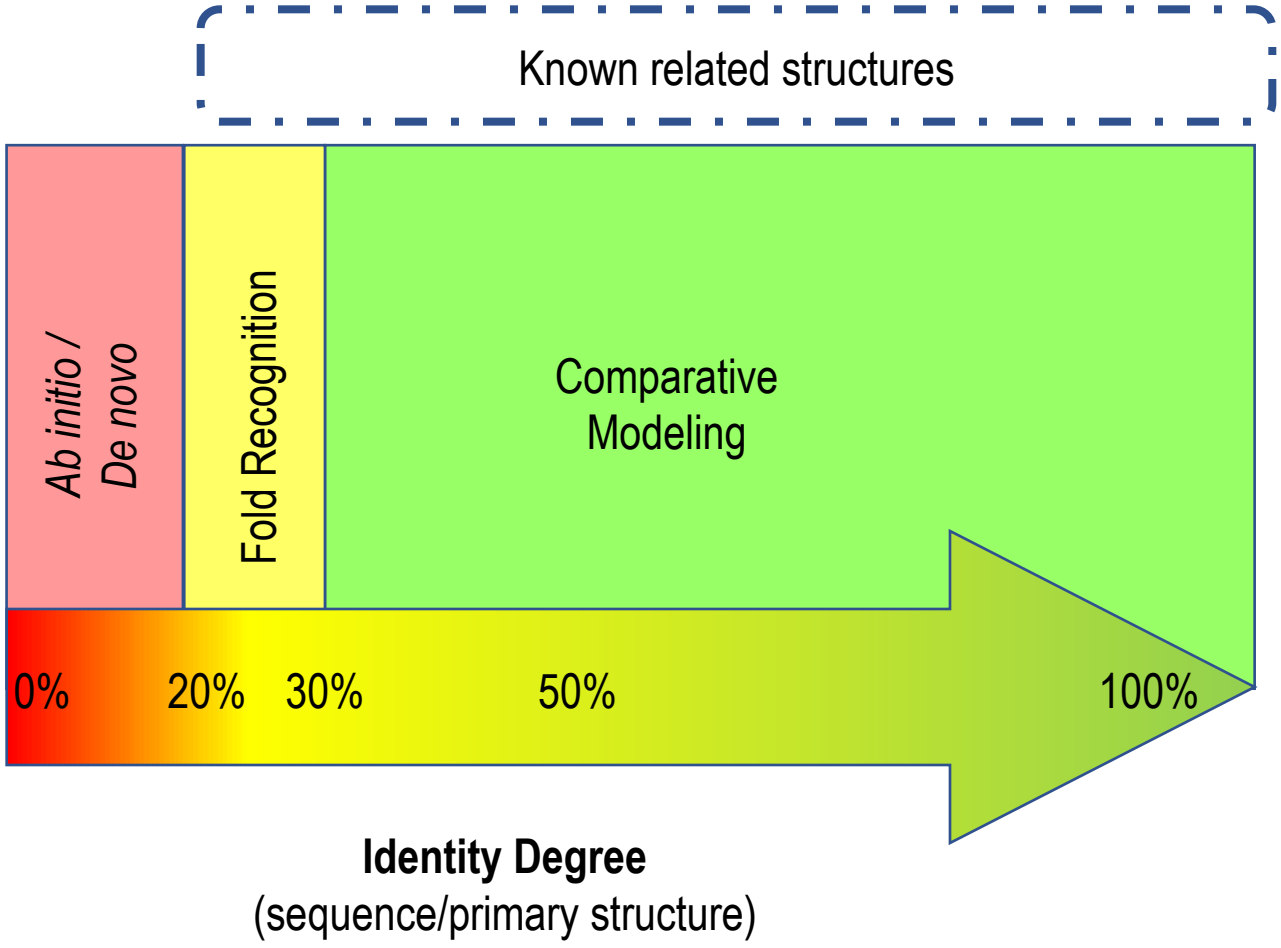
Structural Solving



https://youtu.be/BJKkC0W-6Qk

Petsko & Ringe, *Protein Structure and Function,* Blackwell (2004); Nogales & Scheres, *Mol Cell* 58, 677-89 (2015).

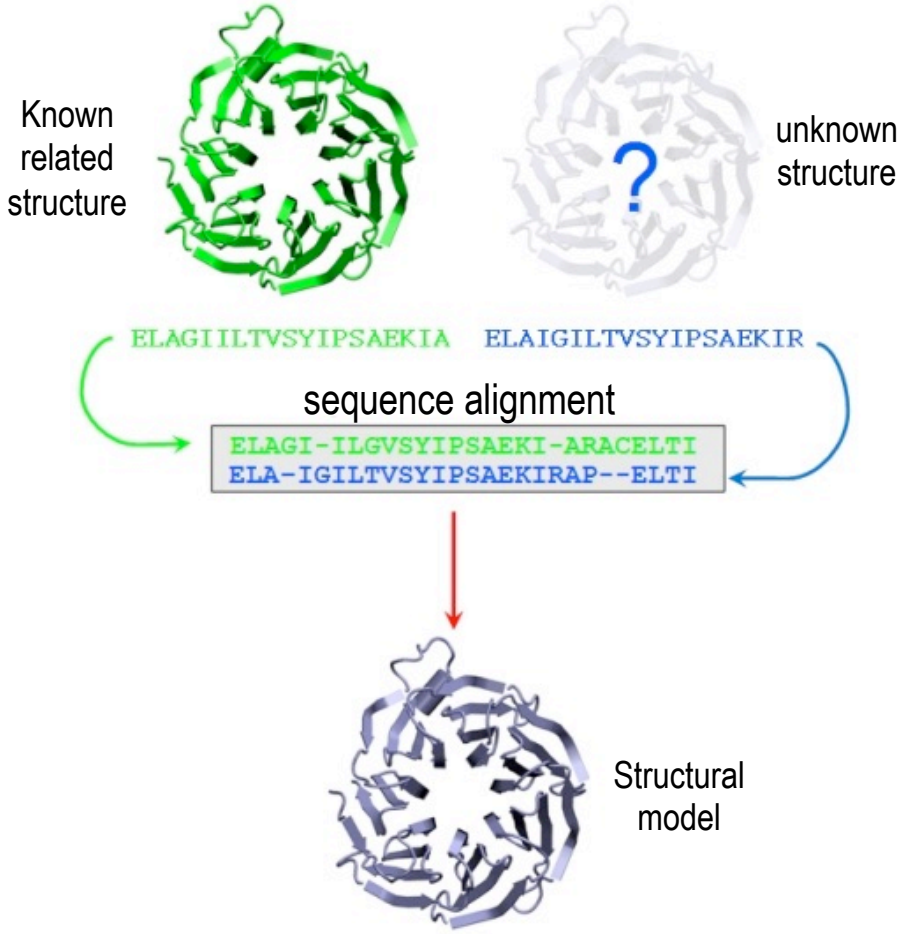
Structural Modeling



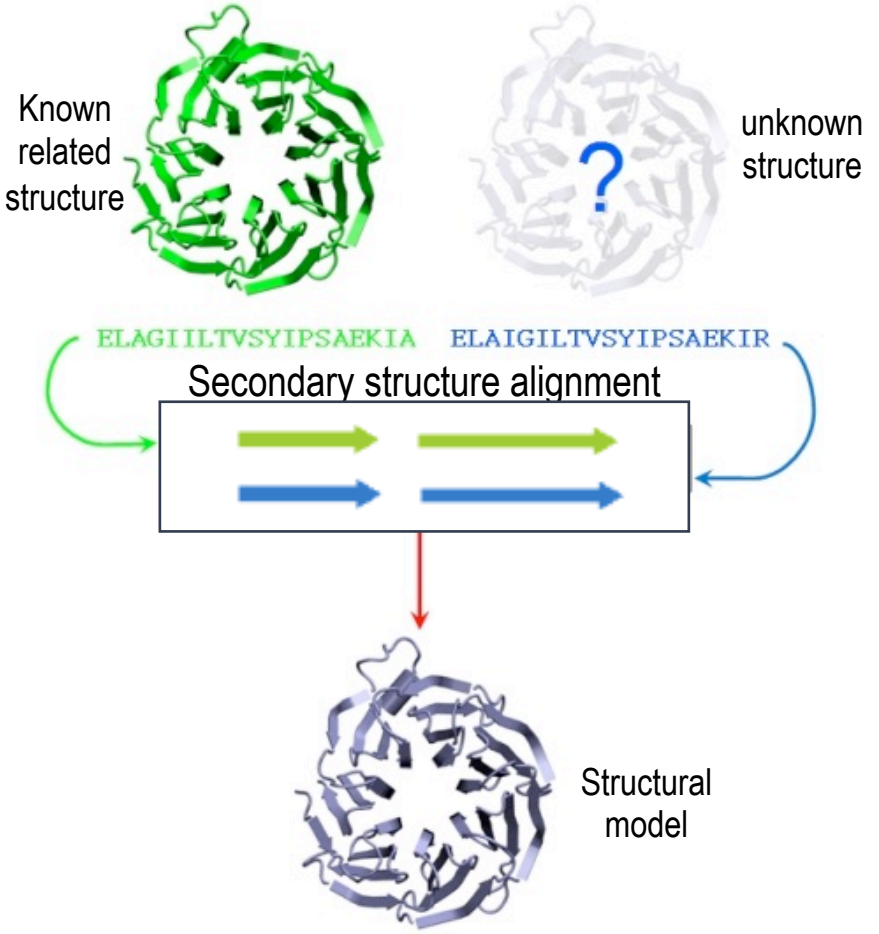
Capriles et al. in Verli, *Bioinformática: Da biologia à flexibilidade molecular*, SBBq (2014).

Structural Modeling

Comparative Modeling



Fold Recognition



Sánchez & Šali, *Methods Mol Biol* 143, 97-129 (2000).

Structural Modeling

SIB Swiss Institute of Bioinformatics | BIOZENTRUM Universität Basel

SWISS-MODEL

Modelling
- myWorkspace
- Automated Mode
- Alignment Mode
- Project Mode

Tools
- Template Identification
- Domain Annotation
- Structure Assessment
- Template Library

Repository
- Search by Sequence
- Search by AC

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists WorldWide.

What's new?

- New automated modeling pipeline with improved hierarchical approach for template selection.
- Increased sensitivity of template detection (sequence to profile search using an adapted HHSearch protocol)
- New tools for model and structure quality assessment: Dfire and Qmean global scores; ProQres residue based assessment scores

SWISS-MODEL Team

Torsten Schwede:	Project Leader
Florian Kiefer:	SWISS-MODEL Repository
Lorenza Bordoli:	Method Development and user support
Konstantin Arnold:	SWISS-MODEL Workspace

References:
When you publish or report results using SWISS-MODEL, please cite the relevant publications:

- Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392.
- Peitsch, M. C. (1995) Protein modeling by E-mail Bio/Technology 13: 658-660.

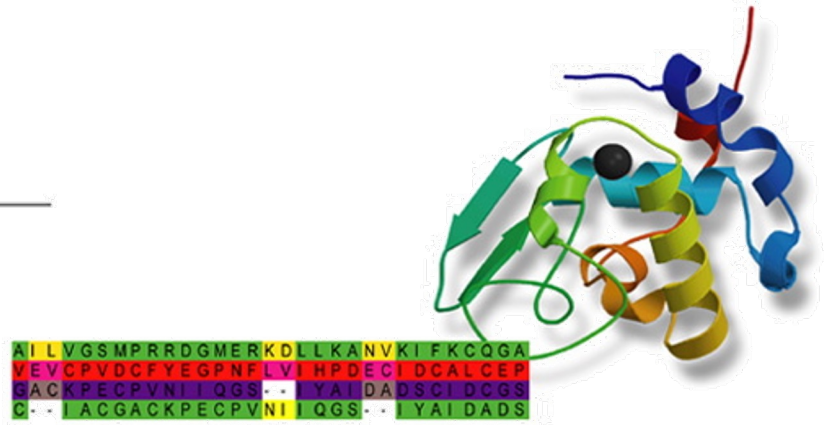
http://swissmodel.expasy.org

Structural Modeling

To main Sali lab pages

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



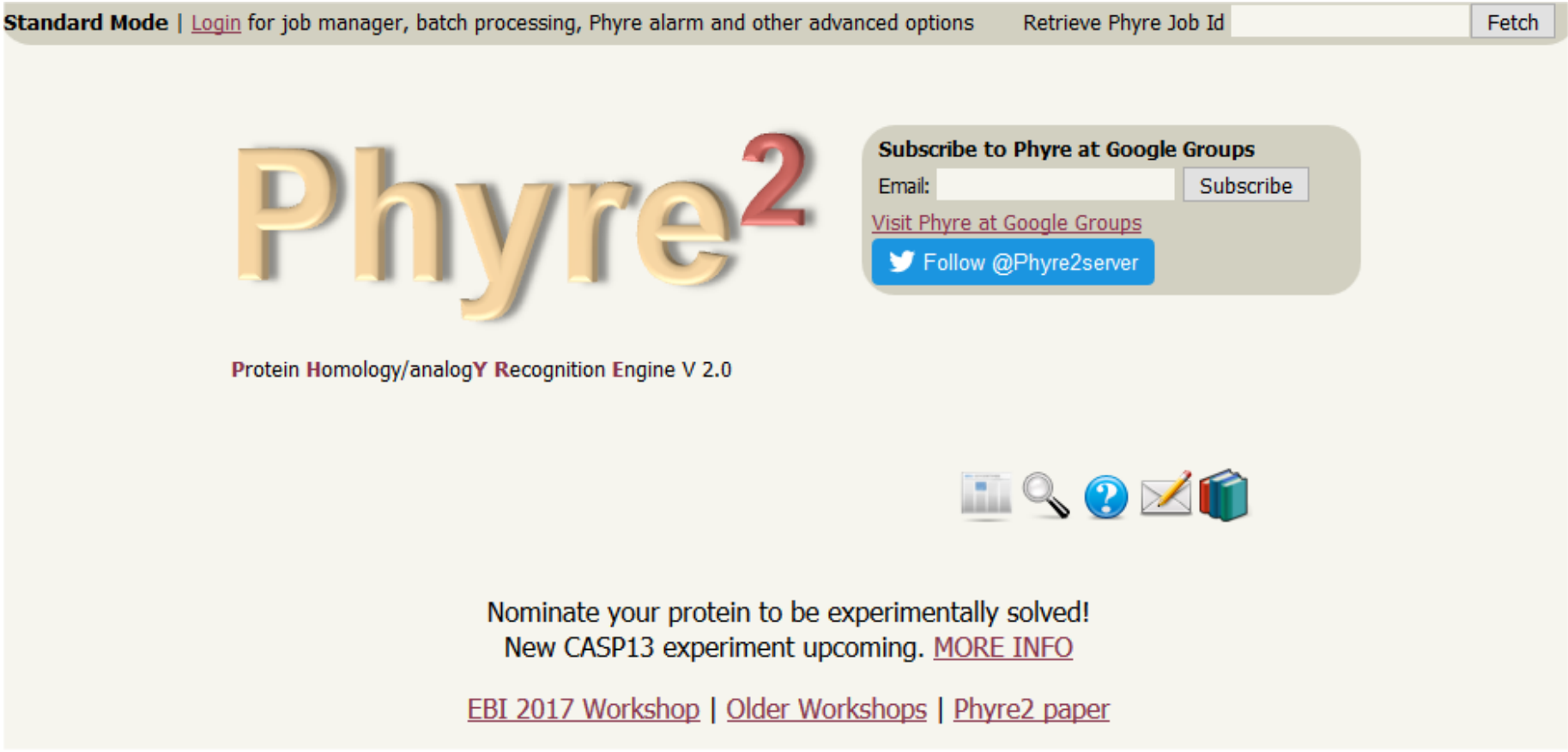
- About MODELLER
- MODELLER News
- Download & Installation
- Release Notes
- Data file downloads
- Registration
- Accelrys licensing
- Discussion Forum

About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is available for download for most Unix/Linux systems, Windows, and Mac.

www.salilab.org/modeller

Structural Modeling



www.sbg.bio.ic.ac.uk/phyre2

Structural Modeling

fold it BETA
Solve Puzzles for Science

SCIENTIFIC AMERICAN™

Sign In | Register 0

Search ScientificAmerican.com

Subscribe | News & Features | Topics | Blogs | Videos & Podcasts | Education | Citizen Science

Technology » News

6 :: Email :: Print

Foldit Gamers Solve Riddle of HIV Enzyme within 3 Weeks

The online game poses protein-folding puzzles, and participants provided insights recently that solved the structure of an enzyme involved in reproduction of HIV

By Michael J. Coren and Fast Company | September 20, 2011

When video gamers armed with the world's most powerful supercomputers take on science and its most vexing riddles, who wins? Sometimes, it's the gamers.

Humans retain an edge over computers when complex problems require intuition

https://fold.it

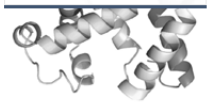
CASP

Critical Assessment of Techniques for Structure Prediction (CASP) competition

Protein Structure Prediction Center

Menu

- Home
- PC Login
- PC Registration
- CASP Experiments
  - CASP14 (2020)
  - *CASP Commons (COVID-19, 2020)*
  - CASP13 (2018)
  - CASP12 (2016)
  - CASP11 (2014)
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)
  - CASP2 (1996)
  - CASP1 (1994)
- Initiatives
- Data Archive
- Proceedings
- CASP Measures
- Feedback
- Assessors
- People
- Community Resources
- Job Fair

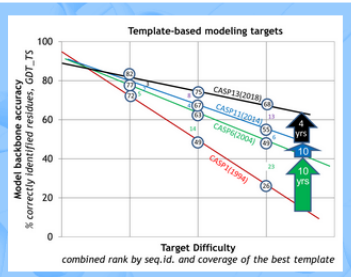


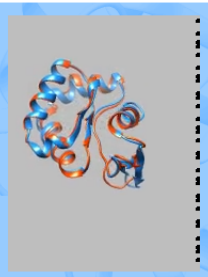
Success Stories From Recent CASPs

template-based modeling | *ab initio modeling* | *contact prediction* | *help structural biologists* | *refinement* | *data-assisted modeling* | >

template-based modeling

Models based on templates identified by sequence similarity remain the most accurate. Over the course of the CASP experiments there have been enormous improvements in this area. However, the overall accuracy improvements that we have seen in the first 10 years of CASP remained unmatched until CASP12 (2016), when a new burst of progress happened [Kryshtafovych et al, 2018]. In two years from 2014 to 2016, the backbone accuracy of the submitted models improved more than in the preceeding 10 years. The next CASP continued the trend [Croll et al, 2019], and the 2014-2018 model accuracy improvement doubled that of 2004-2014 (see the plot). Several factors contributed to this, including more accurate alignment of the target sequence to that of available templates, combining multiple templates, improved accuracy of regions not covered by templates, successful refinement of models, and better selection of models from decoy sets due to improved methods for estimation of model accuracy.





target T0868-D1 (orange)
model 330_2 (blue): GDT_TS=87 best template: 2cw6 (seq.id= 4.2%)

Welcome to the Protein Structure Prediction Center!

Our goal is to help advance the methods of identifying protein structure from sequence. The Center has been organized to provide the means of objective testing of these methods via the process of blind prediction. The Critical Assessment of protein Structure Prediction (CASP) experiments aim at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused.

There have been fourteen previous CASP experiments. The fifteenth experiment is planned to start in Spring 2022. Description of these experiments and the full data (targets, predictions, interactive tables with numerical evaluation results, dynamic graphs and prediction visualization tools) can be accessed following the links:

CASP1 (1994) | CASP2 (1996) | CASP3 (1998) | CASP4 (2000) | CASP5 (2002) | CASP6 (2004) | CASP7 (2006) | CASP8 (2008) | CASP9 (2010) | CASP10 (2012) | CASP11 (2014) | CASP12 (2016) | CASP13 (2018) | CASP14 (2020)

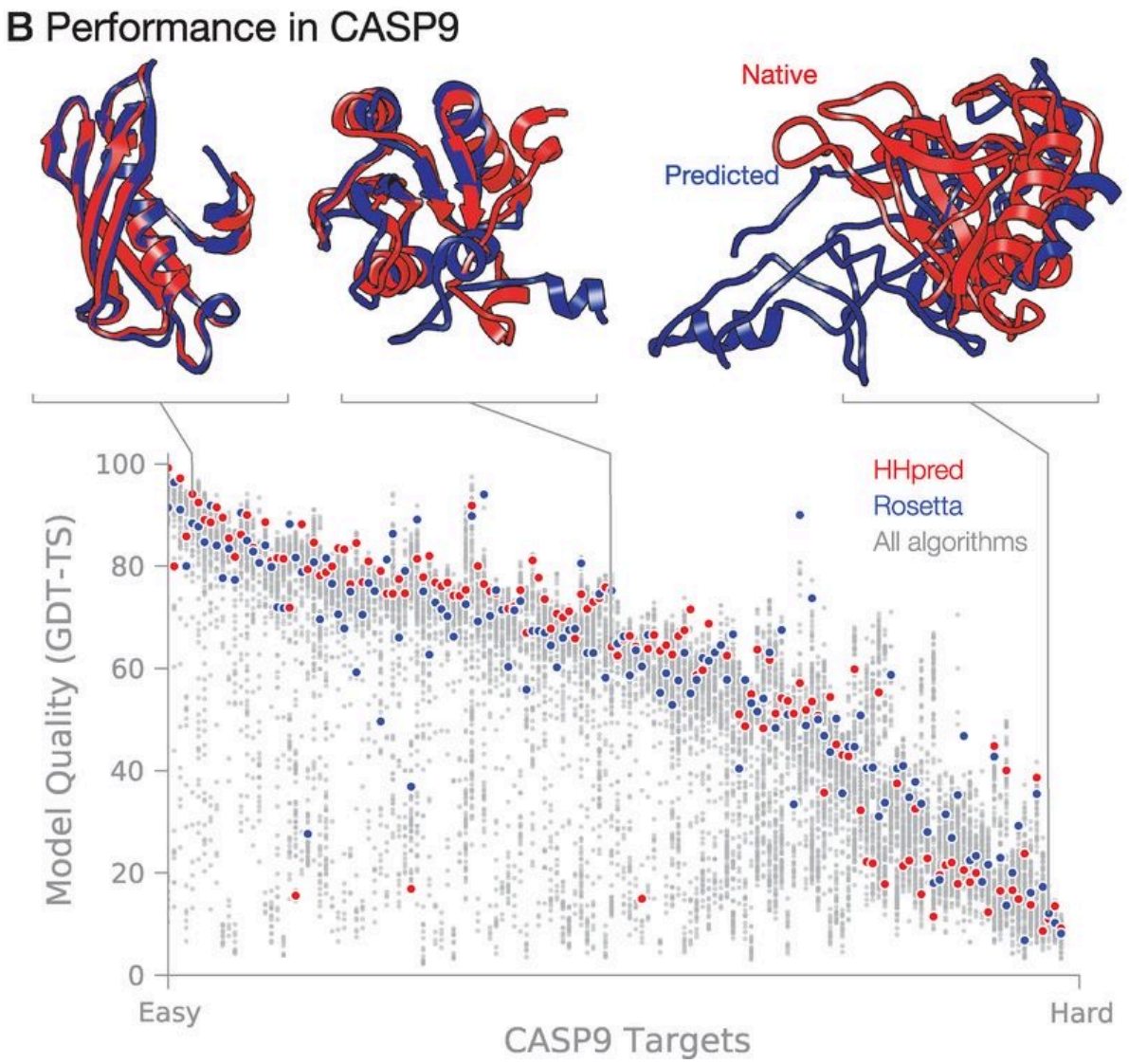
Raw data for the experiments held so far are archived and stored in our data archive.

Details of the experiments have been published in a scientific journal *Proteins: Structure, Function and Bioinformatics*. **CASP proceedings** include papers describing the structure and conduct of the experiments, the numerical evaluation measures, reports from the assessment teams highlighting state of the art in different prediction categories, methods from some of the most successful prediction teams, and progress in various aspects of the modeling.

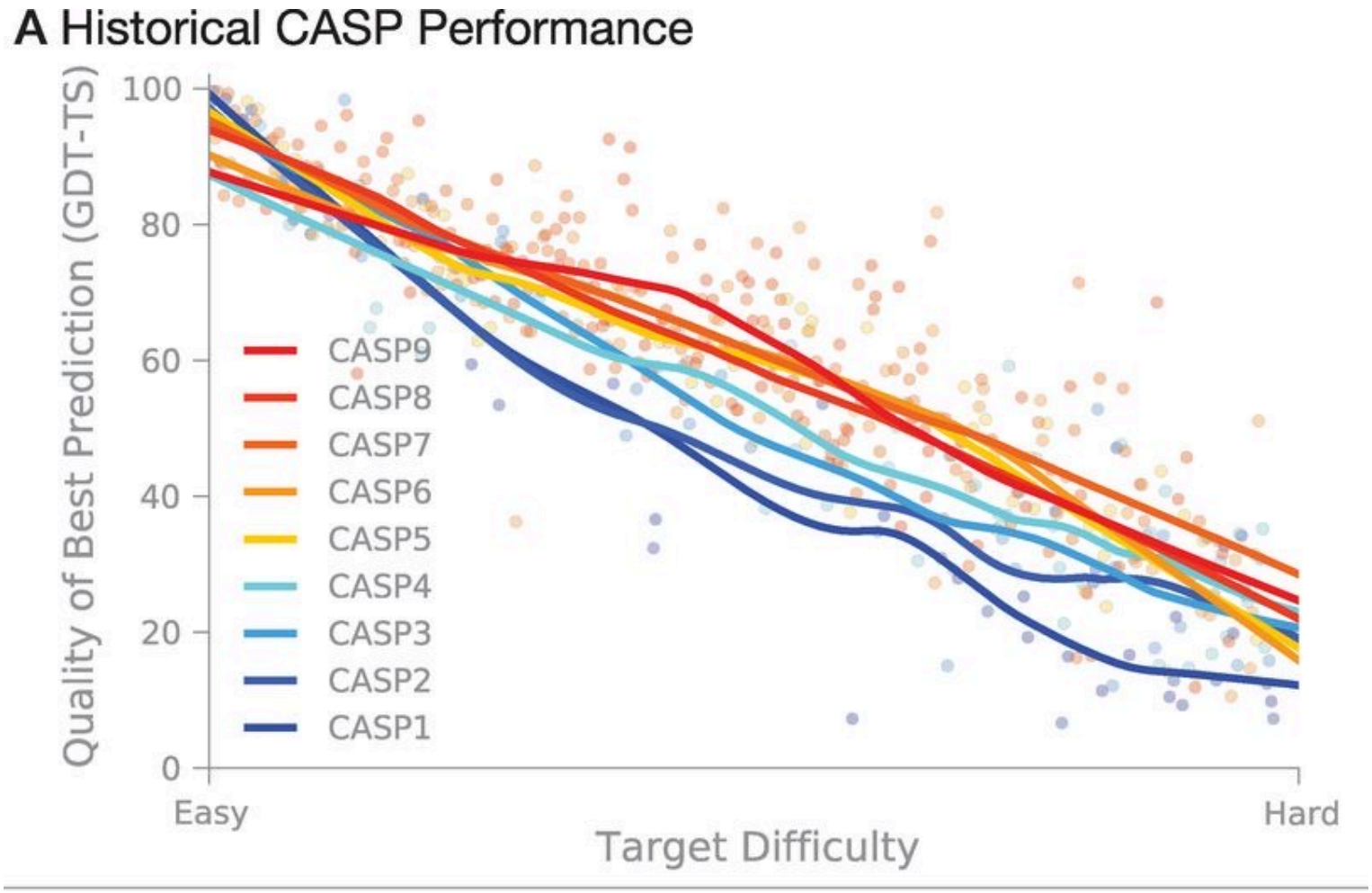
Prediction methods are assessed on the basis of the analysis of a large number of blind predictions of protein structure. Summary of numerical evaluation of the tertiary structure prediction methods tested in the latest CASP experiment can be found on this web page. The main numerical measures used in evaluations, data handling procedures, and guidelines for navigating the data presented on this website are described in [1].

Some of the best performing methods are implemented as fully automated servers and therefore can be used by public for protein structure modeling.

CASP

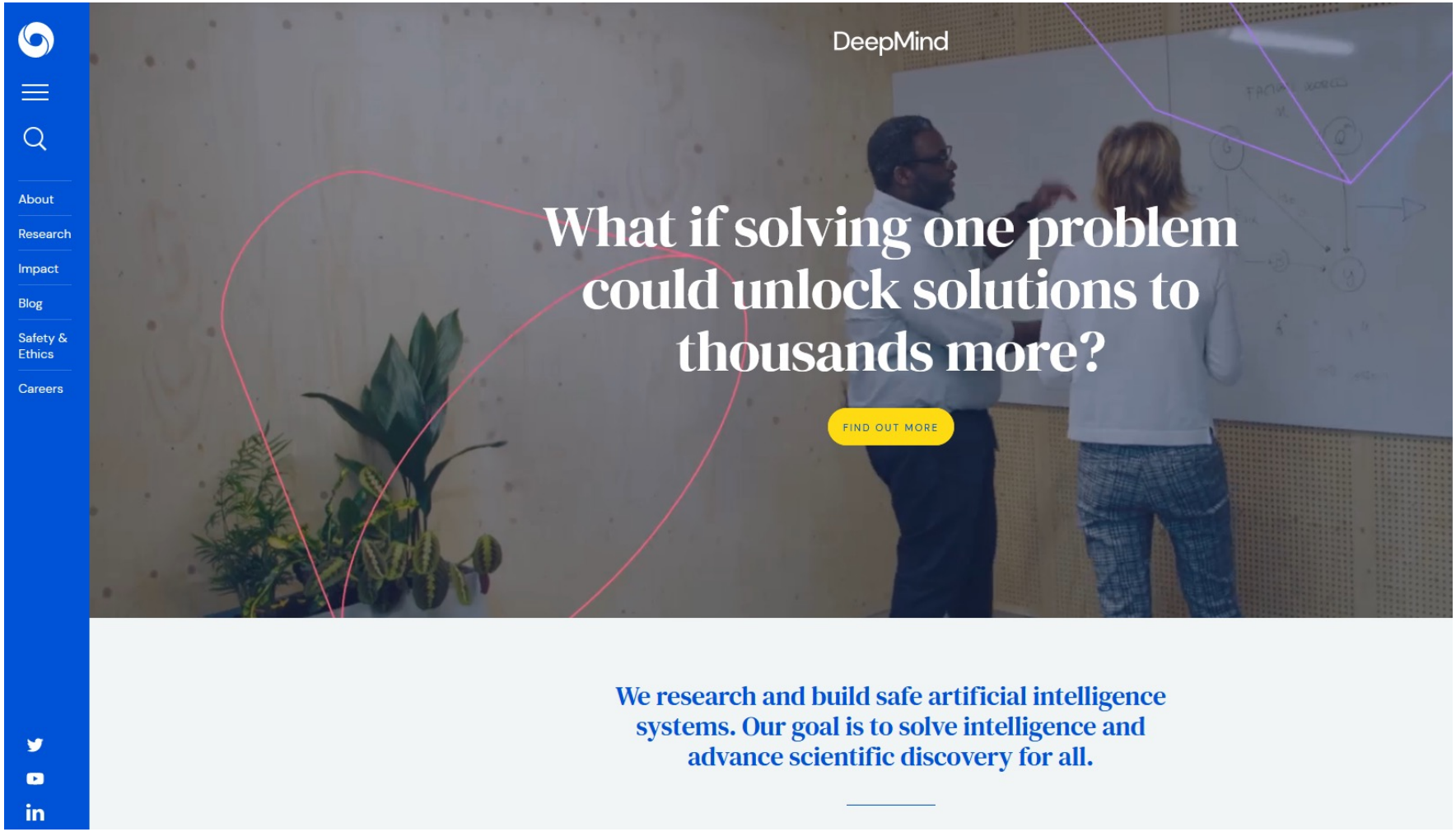


Dill & MacCallum, *Science* 338, 1042-1046 (2012).



Dill & MacCallum, *Science* 338, 1042-1046 (2012).

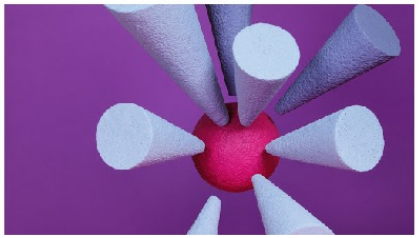
What about Google?



https://deepmind.com

What about Google?

WaveNet: A generative model for raw audio



Giving doctors a headstart on acute kidney injury

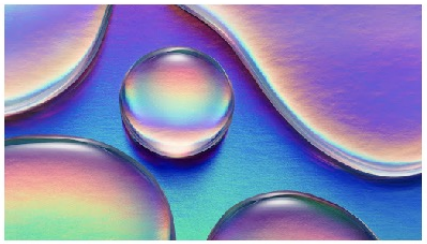
More accurately identifying breast cancer

AlphaStar plays StarCraft II at Grandmaster level

AlphaZero: Shedding new light on chess, shogi, and Go

DQN: Human-level control of Atari games

A neural network with dynamic memory



AlphaGo defeats Lee Sedol in the game of Go

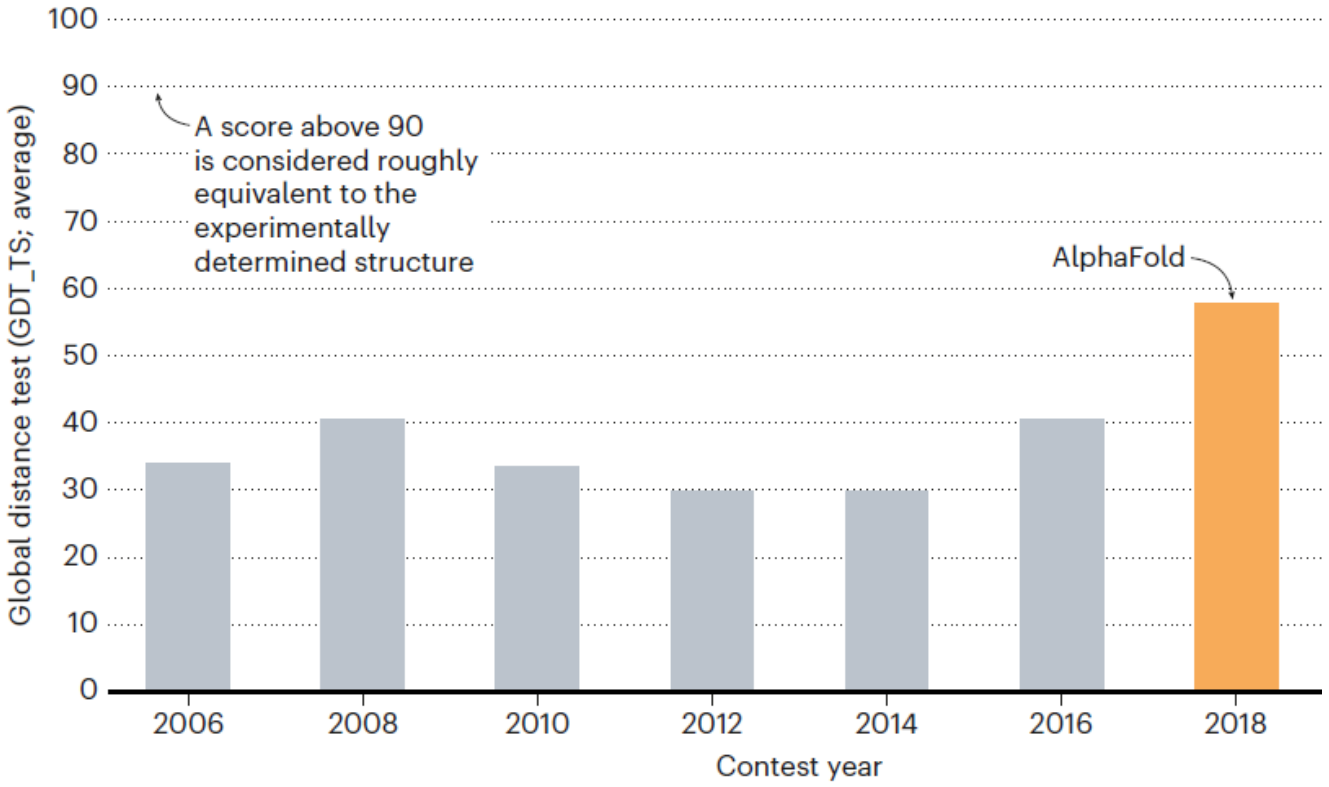
GQN: Neural scene representation and rendering

https://deepmind.com/impact

AlphaFold2

STRUCTURE SOLVER

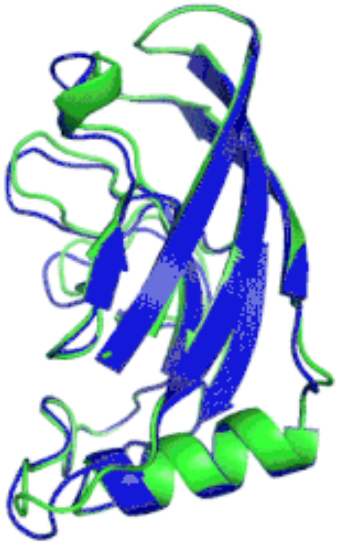
DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



Callaway, *Nature* 588, 203-204 (2020)

AlphaFold2

T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

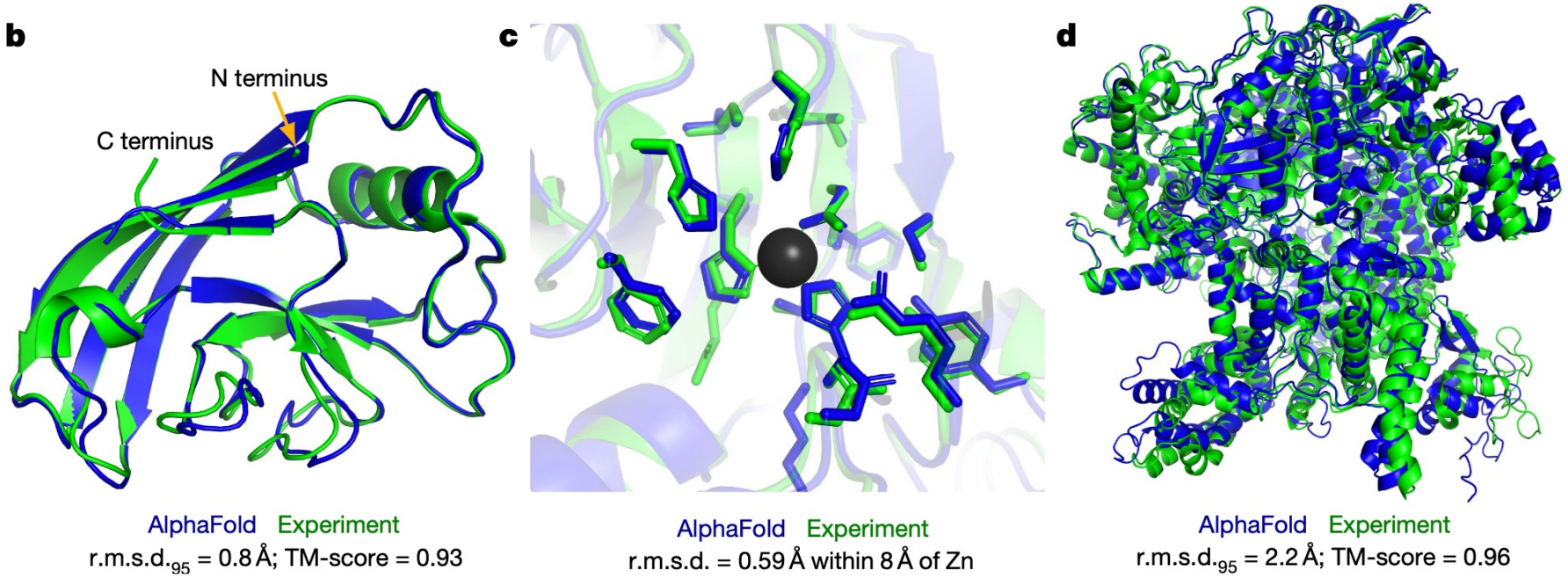


T1049 / 6y4f
93.3 GDT
(adhesin tip)

● Experimental result
● Computational prediction

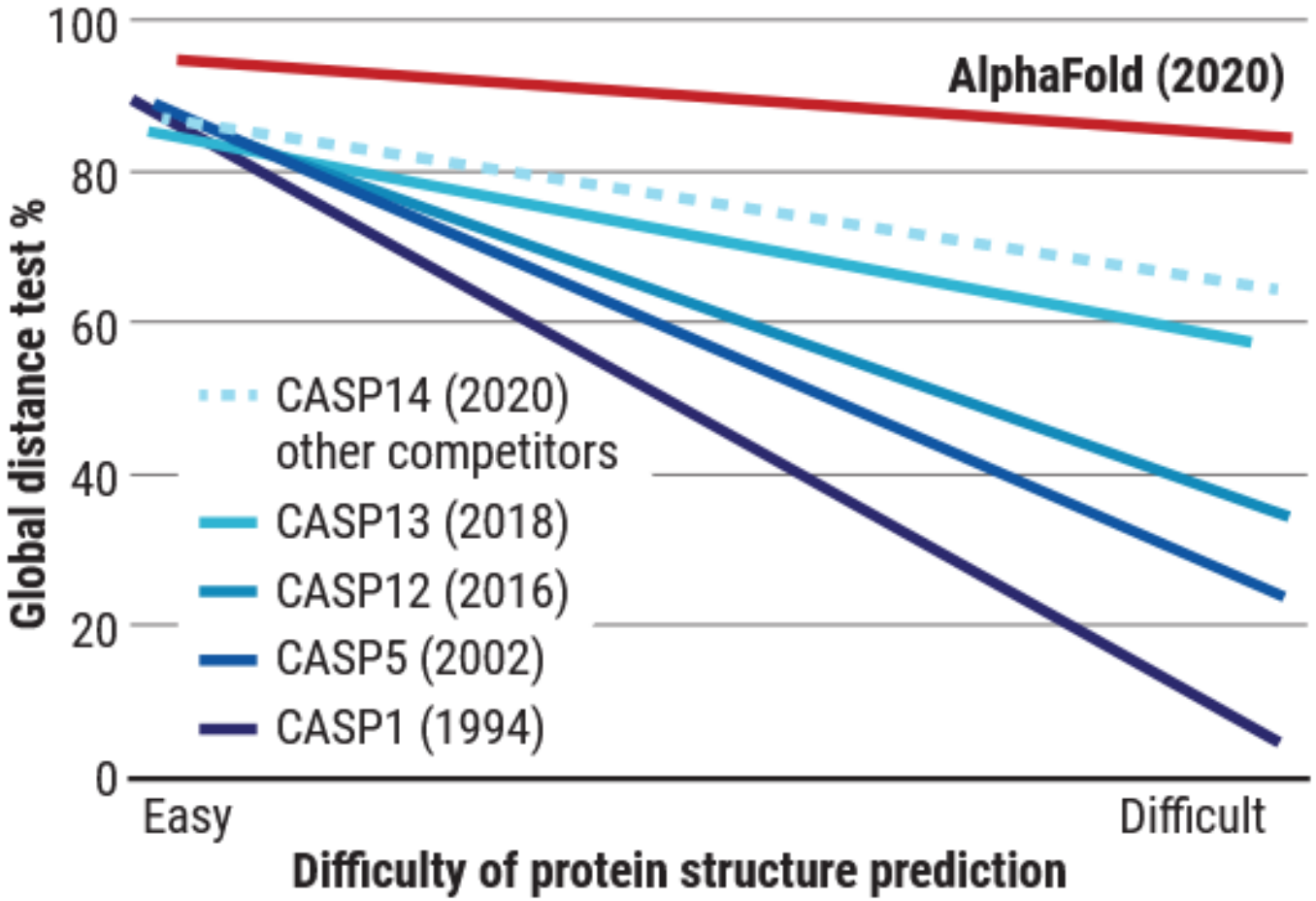
https://www.theverge.com/2020/12/1/21754310/deepmind-alphafold-ai-protein-folding-casp-competition

AlphaFold2



Callaway, *Nature* 588, 203-204 (2020)

AlphaFold2

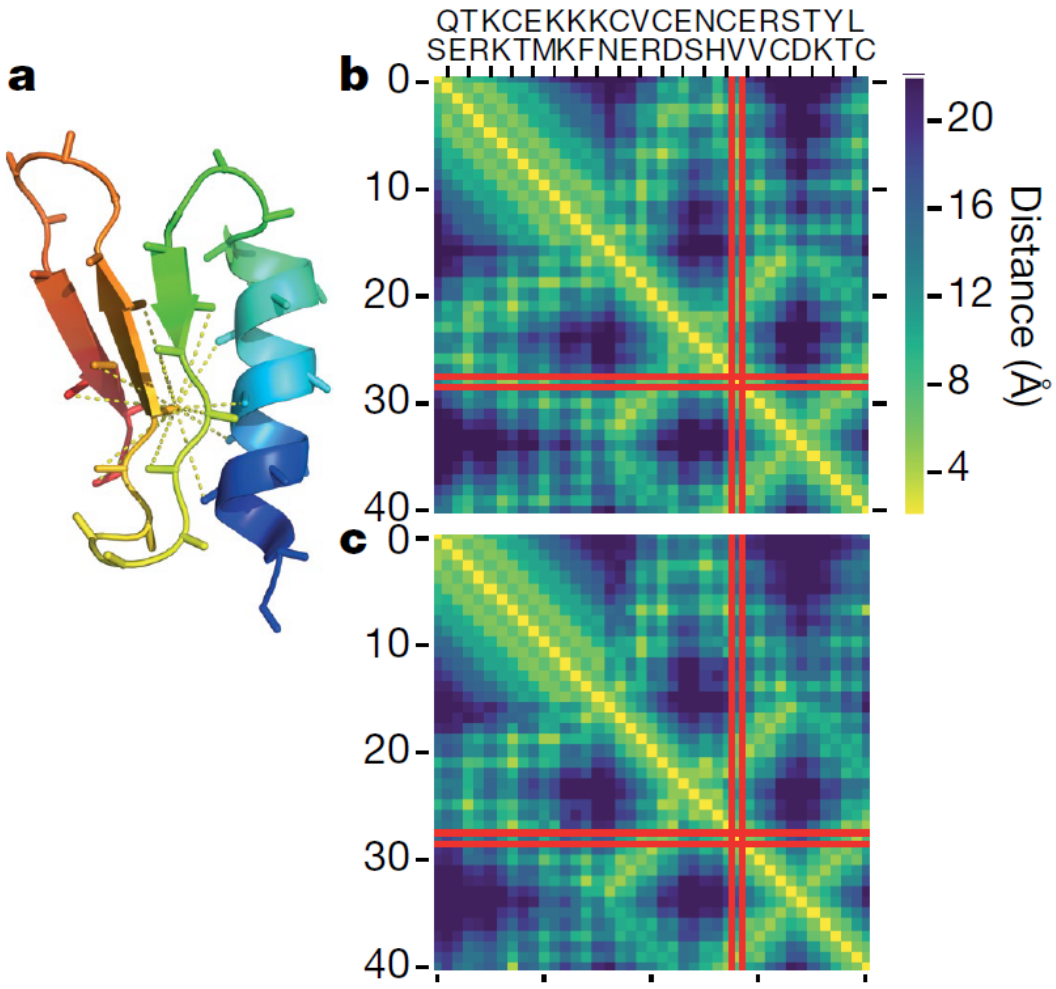


Service, Science (2020) 10.1126/science.abf9367

AlphaFold2

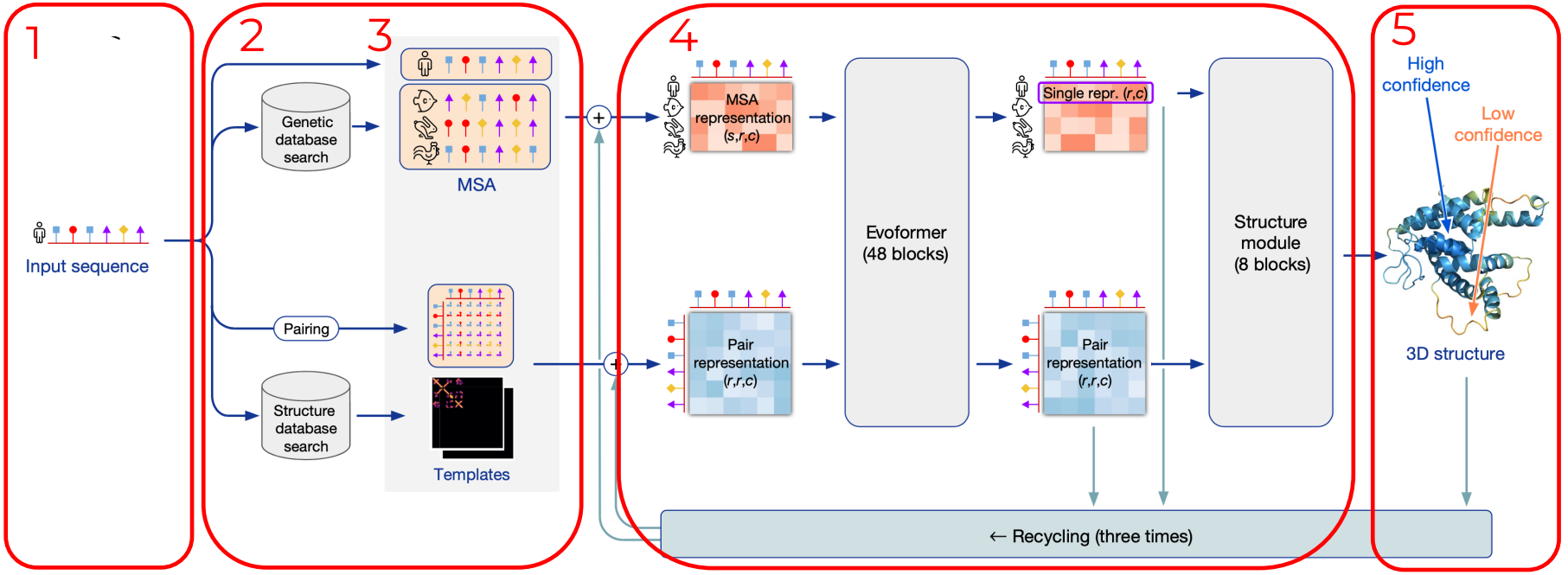
https://youtu.be/gg7WjuFs8F4

AlphaFold2



Senior et al., *Nature* 577, 706-710 (2020).

AlphaFold2



Senior et al., *Nature* 577, 706-710 (2020); Jumper et al., *Nature*, Published online 15jul2021; DOI: 10.1038/s41586-021-03819-2

AlphaFold2

Steps:

1. An amino acid sequence is provided as an input to the AlphaFold algorithm,

2. Some data pre-processing is carried out to generate the backbone atoms contained in the amino acid sequence. The atoms are initially assigned random coordinates.

3. Search a database of structures & templates to:
3a. Find structures that correspond to the amino acid sequence aka Multiple Sequence Alignment (MSA). This helps to detect parts of the amino acid sequence that are more likely to change and gather relationships between these “mutations”.
3b. Identify proteins that may have similar structure based on existing templates, and construct an initial representation of structures for pairwise comparison. This model represents which amino acids are likely to be in contact with each other.

4. The MSA and pairwise structures (3a. & 3b.) are passed through a deep learning model called “Evoformer” — which has 48 blocks of neural networks to exchange information between the amino acid sequence & pairwise structural characterisation of the geometry, this helps to refine the model over several iterations. Next comes another deep learning module called “structure module” which has 8 blocks of neural networks to construct the final 3D protein structure of the input sequence.

Rerun steps 3 & 4, three times

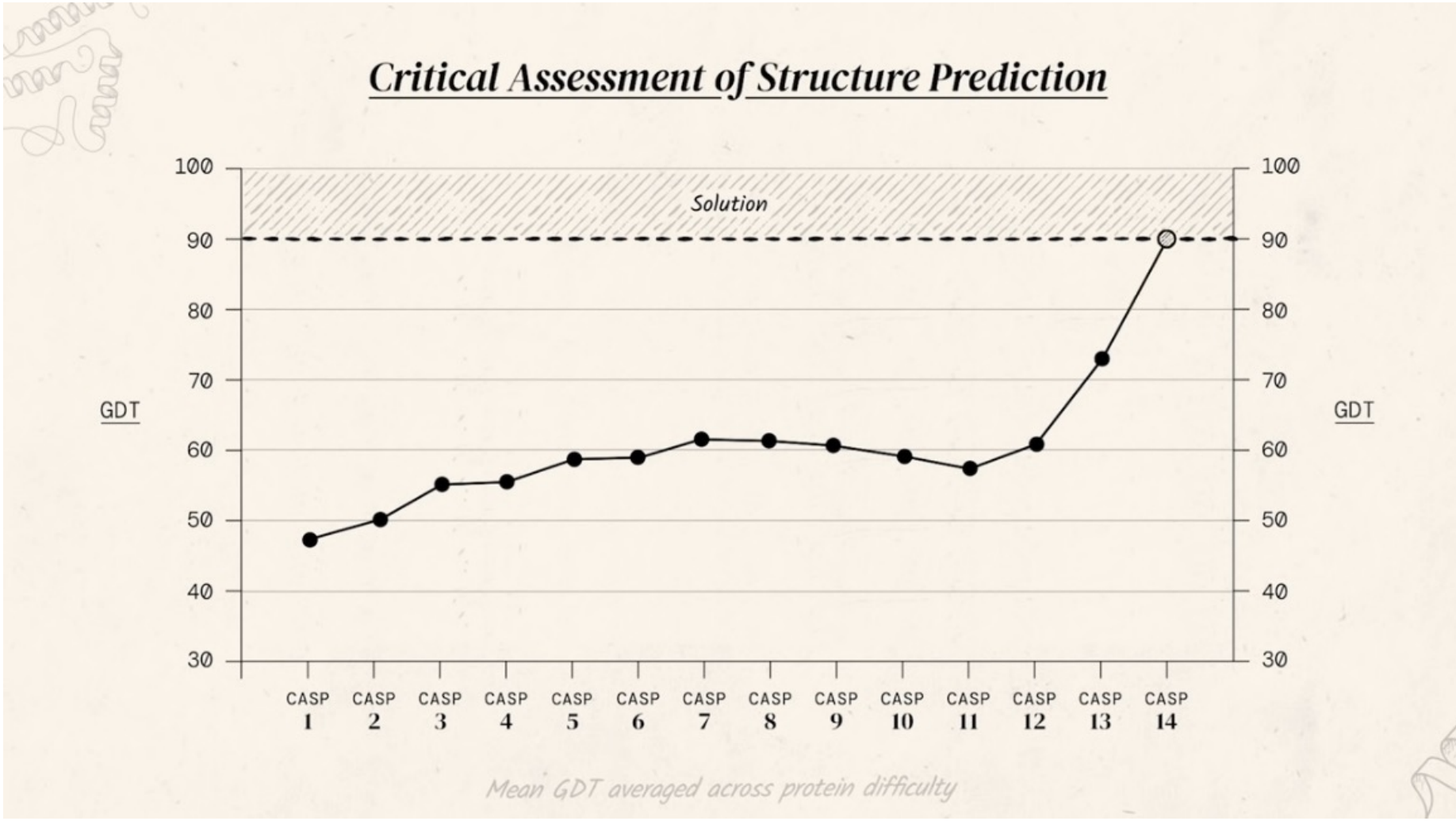
5. The final output result is a refined folded protein structure with the corresponding coordinates for each atom in the protein.

Senior et al., *Nature* 577, 706-710 (2020); Jumper et al., *Nature*, Published online 15jul2021; DOI: 10.1038/s41586-021-03819-2

AlphaFold2

https://youtu.be/gg7WjuFs8F4

AlphaFold2



https://youtu.be/gg7WjuFs8F4

Potential

Basic Science

Drug Design

Protein Engineering

https://youtu.be/KpedmJdrTpY

Are we done?

"Being able to investigate the shape of proteins quickly and accurately has the potential to revolutionize life sciences. Now that the problem has been largely solved for single proteins, the way is open for development of new methods for determining the shape of protein complexes – collections of proteins that work together to form much of the machinery of life, and for other applications."

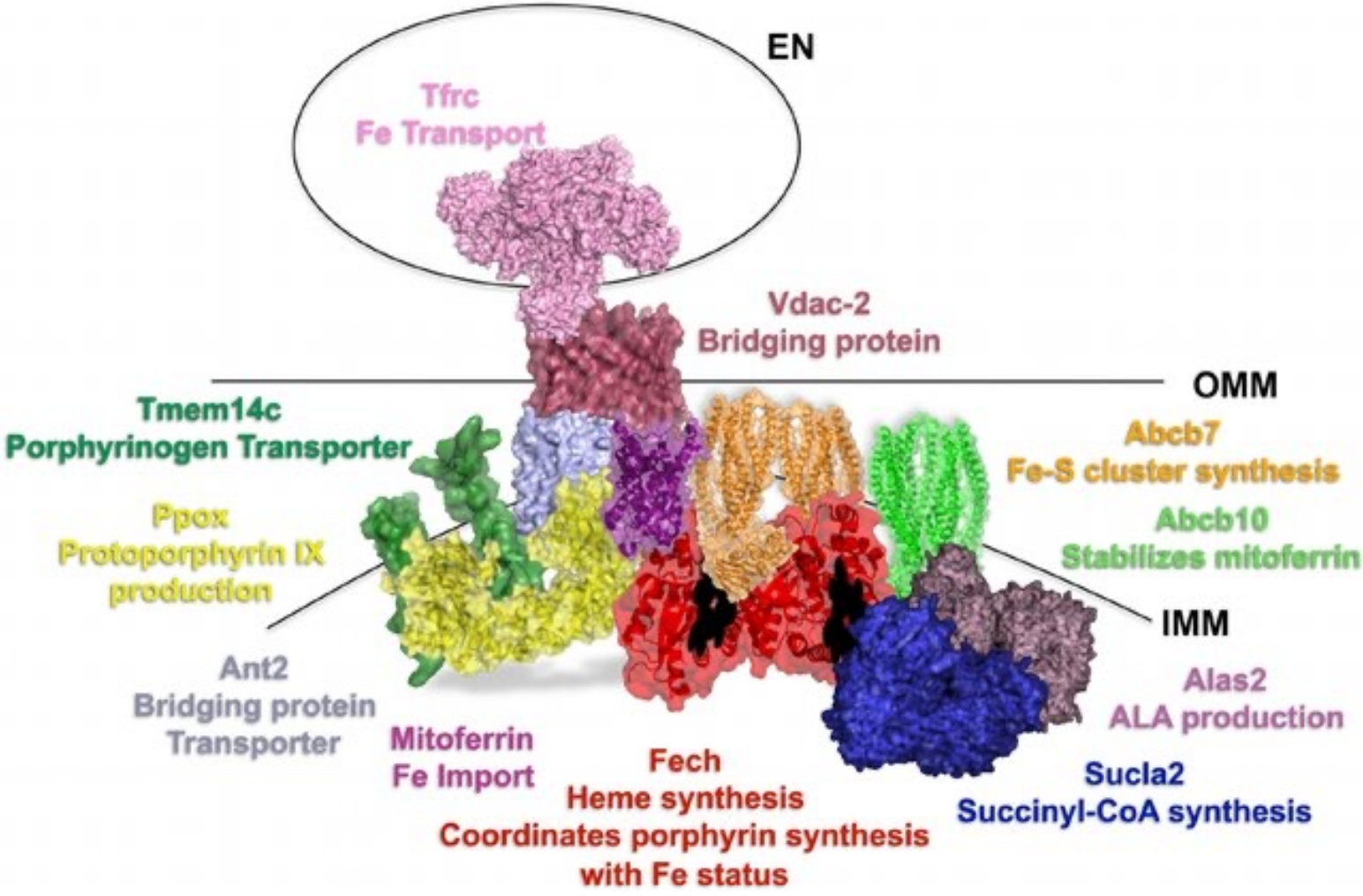
https://predictioncenter.org/casp14/doc/CASP14_press_release.html

Wrap up & future outlook

- We have built a system that confidently predicts accurate structures for most proteins - and knows when it is wrong
- As for CASP13[1,2], we'll publish a peer-reviewed paper
- We're also working on providing broad access to our work
- Lots of exciting work ahead for the field: Complexes, conformational change etc

[1] Senior, A. W., et al. "Improved protein structure prediction using potentials from deep learning." Nature 577.7792 (2020): 706-710.
[2] Senior, A. W., et al. "Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13)." Proteins 87.12 (2019): 1141-1148.

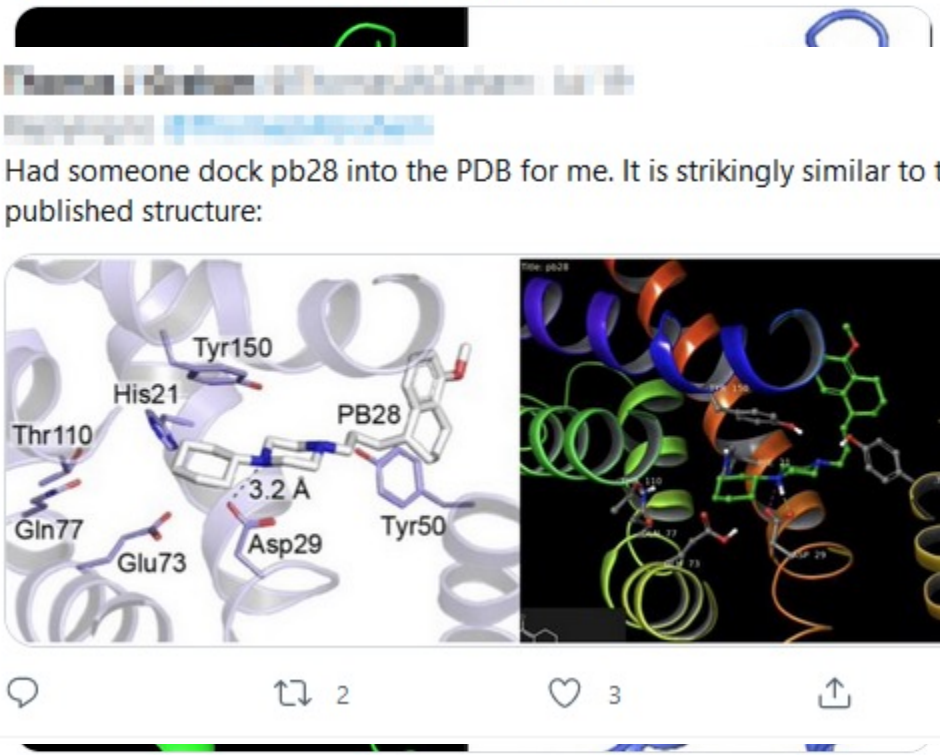
Protein Complexes/Oligomers



Medlock et al., *PLoS One* 10, e0135896 (2015).

Are we done?

I'm pretty impressed with AlphaFold2. Predicted structure of sigma2/tmem97 on left vs. unpublished x-ray structure on right. Keep in mind, there are no good templates for homology models.



Had someone dock pb28 into the PDB for me. It is strikingly similar to the published structure:

2 3

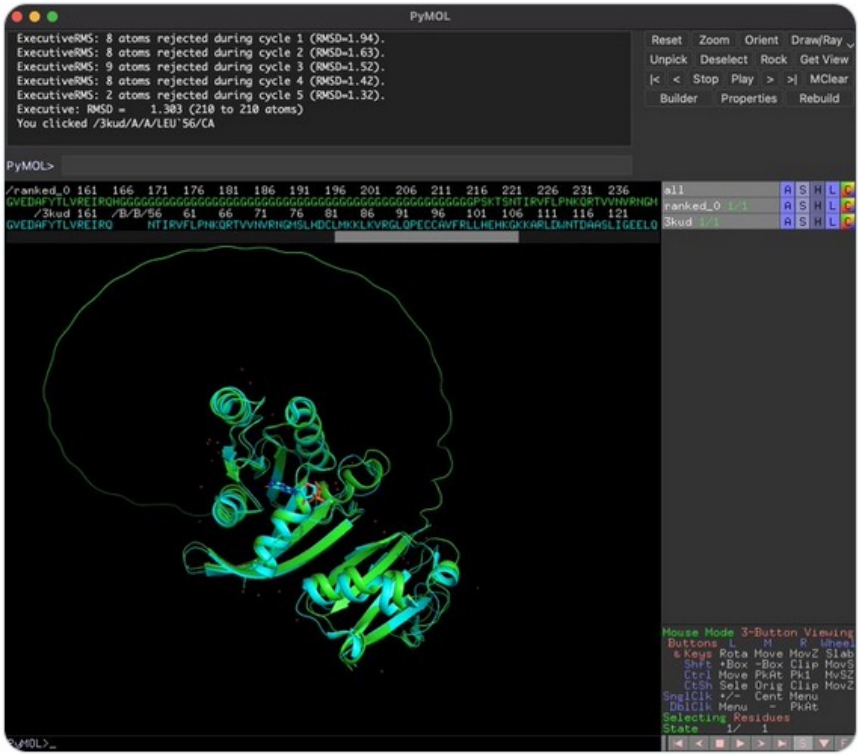
1:37 AM · Jul 19, 2021 · Twitter Web App

Are we done?

Protein Complexes/Oligomers

[redacted]

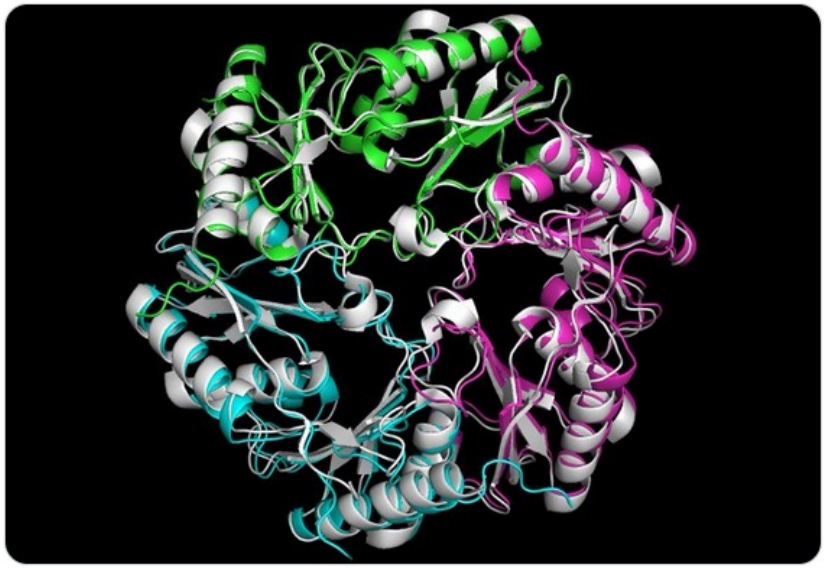
AlphaFold2 can also predict heterocomplexes. All you have to do is input the two sequences you want to predict and connect them with a long linker.



7:07 AM · Jul 19, 2021 · Twitter Web App

[redacted]

Homooligomeric prediction in #alphafold works a little too good. So far worked on nearly every case we (me & [redacted]) tried. Going beyond dimers! Seems @deepmind accidentally "solved" the homooligomeric prediction problem (w/ MSA input) 😂



4:52 PM · Jul 21, 2021 · Twitter Web App

Science

RESEARCH ARTICLES

Cite as: M. Baek *et al.*, *Science*
10.1126/science.abj8754 (2021).

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

Minkyung Baek[1,2], Frank DiMaio[1,2], Ivan Anishchenko[1,2], Justas Dauparas[1,2], Sergey Ovchinnikov[3,4], Gyu Rie Lee[1,2], Jue Wang[1,2], Qian Cong[5,6], Lisa N. Kinch[7], R. Dustin Schaeffer[6], Claudia Millán[8], Hahnbeom Park[1,2], Carson Adams[1,2], Caleb R. Glassman[9,10], Andy DeGiovanni[12], Jose H. Pereira[12], Andria V. Rodrigues[12], Alberdina A. van Dijk[13], Ana C. Ebrecht[13], Diederik J. Opperman[14], Theo Sagmeister[15], Christoph Buhlheller[15,16], Tea Pavkov-Keller[15,17], Manoj K. Rathinaswamy[18], Udit Dalwadi[19], Calvin K. Yip[19], John E. Burke[18], K. Christopher Garcia[9,10,11,20], Nick V. Grishin[6,21,7], Paul D. Adams[12,22], Randy J. Read[8], David Baker[1,2,23]*

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [3]Faculty of Arts and Sciences, Division of Science, Harvard University, Cambridge, MA 02138, USA. [4]John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA 02138, USA. [5]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA. [6]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA. [7]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA. [8]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. [9]Program in Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA. [10]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305, USA. [11]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [12]Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [13]Department of Biochemistry, Focus Area Human Metabolomics, North-West University, 2531 Potchefstroom, South Africa. [14]Department of Biotechnology, University of the Free State, 205 Nelson Mandela Drive, Bloemfontein 9300, South Africa. [15]Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50, 8010 Graz, Austria. [16]Medical University of Graz, Graz, Austria. [17]BioTechMed-Graz, Graz, Austria. [18]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada. [19]Life Sciences Institute, Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC, Canada. [20]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [21]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX, USA. [22]Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA. [23]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

***Corresponding author. Email: dabaker@uw.edu**

First release: 15 July 2021

Article

Highly accurate protein structure prediction for the human proteome

https://doi.org/10.1038/s41586-021-03828-1

Received: 11 May 2021

Accepted: 16 July 2021

Published online: 22 July 2021

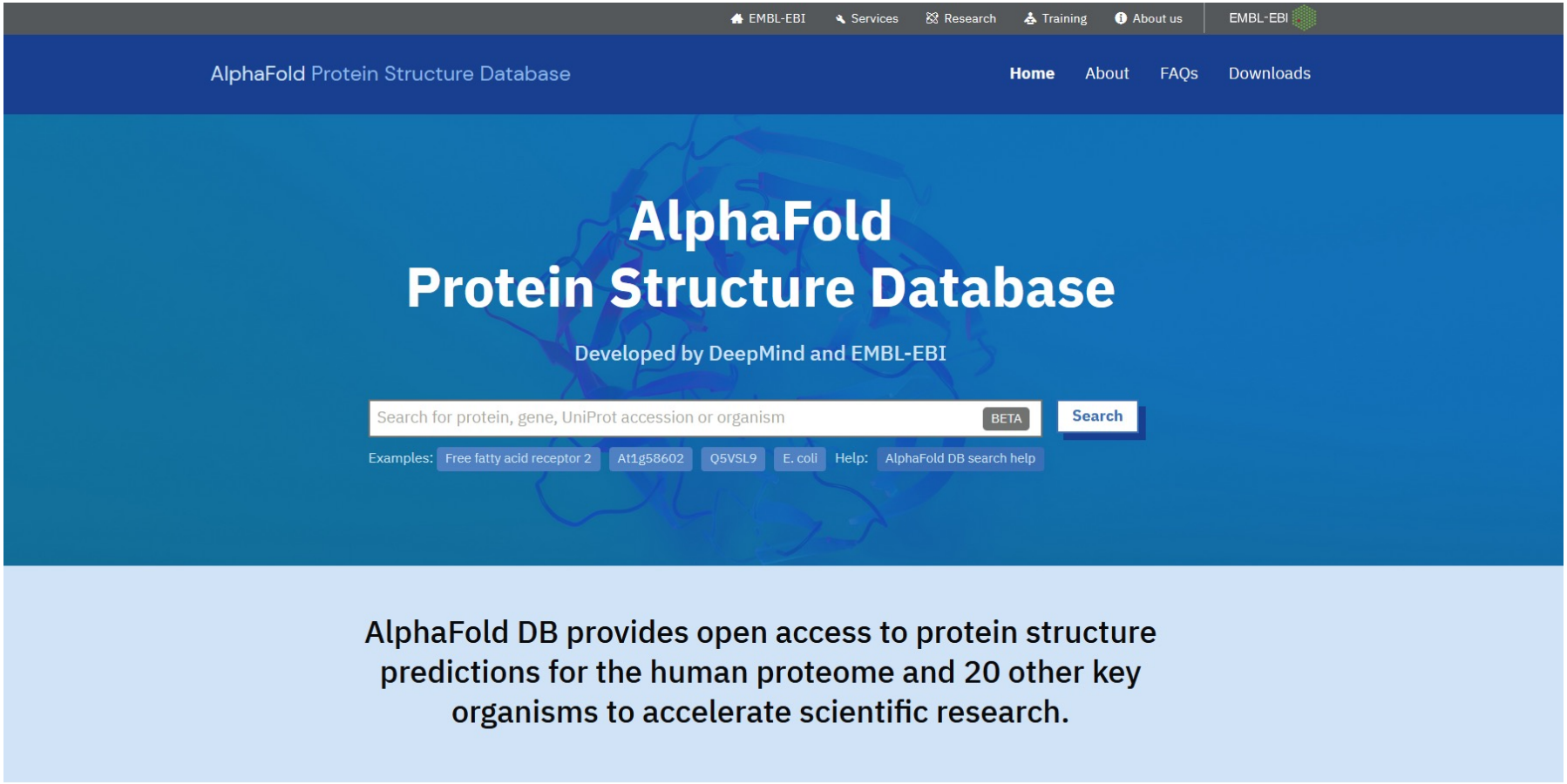
Open access

Check for updates

Kathryn Tunyasuvunakool[1]✉, Jonas Adler[1], Zachary Wu[1], Tim Green[1], Michal Zielinski[1], Augustin Žídek[1], Alex Bridgland[1], Andrew Cowie[1], Clemens Meyer[1], Agata Laydon[1], Sameer Velankar[2], Gerard J. Kleywegt[2], Alex Bateman[2], Richard Evans[1], Alexander Pritzel[1], Michael Figurnov[1], Olaf Ronneberger[1], Russ Bates[1], Simon A. A. Kohl[1], Anna Potapenko[1], Andrew J. Ballard[1], Bernardino Romera-Paredes[1], Stanislav Nikolov[1], Rishub Jain[1], Ellen Clancy[1], David Reiman[1], Stig Petersen[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Ewan Birney[2], Pushmeet Kohli[1], John Jumper[1,3]✉ & Demis Hassabis[1,3]✉

Protein structures can provide invaluable information, both for reasoning about biological processes and for enabling interventions such as structure-based drug development or targeted mutagenesis. After decades of effort, 17% of the total residues in human protein sequences are covered by an experimentally determined structure[1]. Here we markedly expand the structural coverage of the proteome by applying the state-of-the-art machine learning method, AlphaFold[2], at a scale that covers almost the entire human proteome (98.5% of human proteins). The resulting dataset covers 58% of residues with a confident prediction, of which a subset (36% of all residues) have very high confidence. We introduce several metrics developed by building on the AlphaFold model and use them to interpret the dataset, identifying strong multi-domain predictions as well as regions that are likely to be disordered. Finally, we provide some case studies to illustrate how high-quality predictions could be used to generate biological hypotheses. We are making our predictions freely available to the community and anticipate that routine large-scale and high-accuracy structure prediction will become an important tool that will allow new questions to be addressed from a structural perspective.

Are we done?



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

Are we done?

RCSB PDB | Deposit | Search | Visualize | Analyze | Download | Learn | More | Documentation | Careers | MyPDB | Contact us

RCSB PDB PROTEIN DATA BANK

201,196 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM

Advanced Search | Brow

PDB-101 | wwPDB | EMDataResource | NUCLEIC ACID DATABASE | wwPDB Foundation

To include/exclude Computed Structure Models (CSMs) in the search results, toggle the "Include CSM" switch on/off.

New: More Computed Structure Models

Welcome

Deposit

Search

Visualize

Analyze

Download

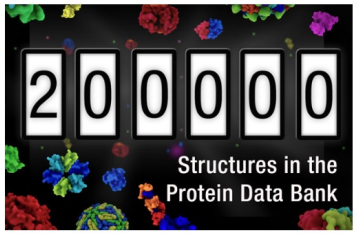
Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

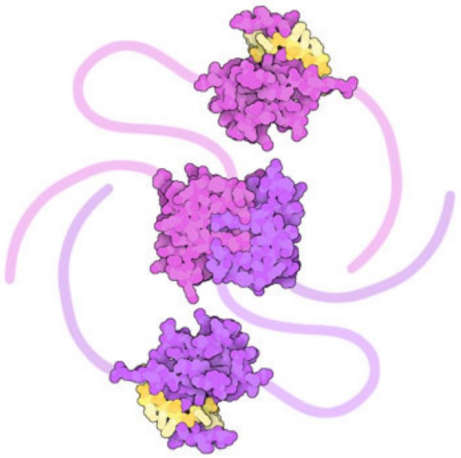
Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive

Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.



February **Molecule of the Month**



SARS-CoV-2 Nucleocapsid and Home Tests

https://www.rcsb.org/

Are we done?

README.md

ColabFold - v1.5.1

```
+ 04Feb2023: v1.5.0 - ColabFold updated to use AlphaFold v2.3.1!
+ 06Feb2023: v1.5.1 - fixing --save-all/--save-recycles option
```

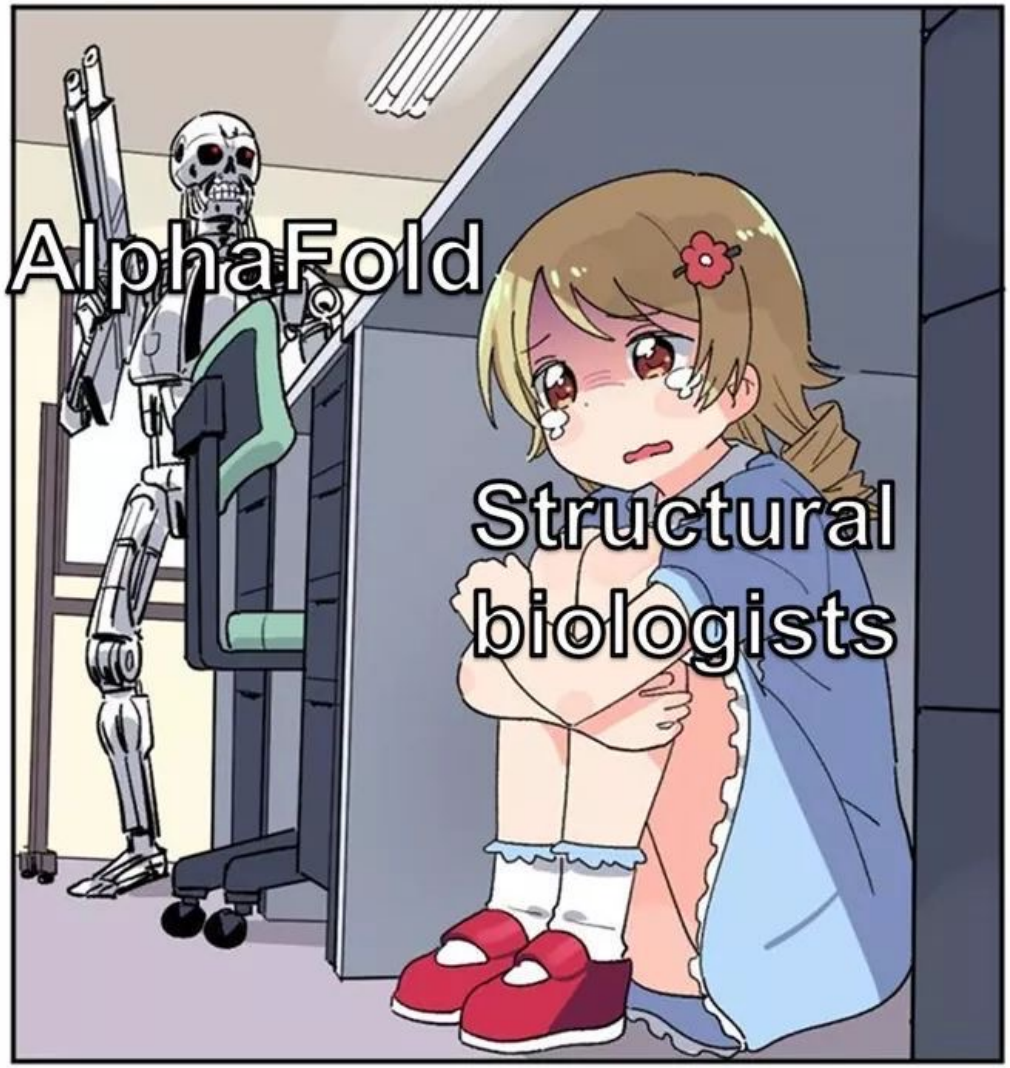
For details of what was changed in v1.5, see change log!

Making Protein folding accessible to all via Google Colab!

Notebooks	monomers	complexes	mmseqs2	jackhmmer	templates
AlphaFold2_mmseqs2	Yes	Yes	Yes	No	Yes
AlphaFold2_batch	Yes	Yes	Yes	No	Yes
RoseTTAFold	Yes	No	Yes	No	No
AlphaFold2 (from Deepmind)	Yes	Yes	No	Yes	No
ESMFold	Yes	Maybe	No	No	No

https://github.com/sokrypton/ColabFold

Are we done?



Are we done?

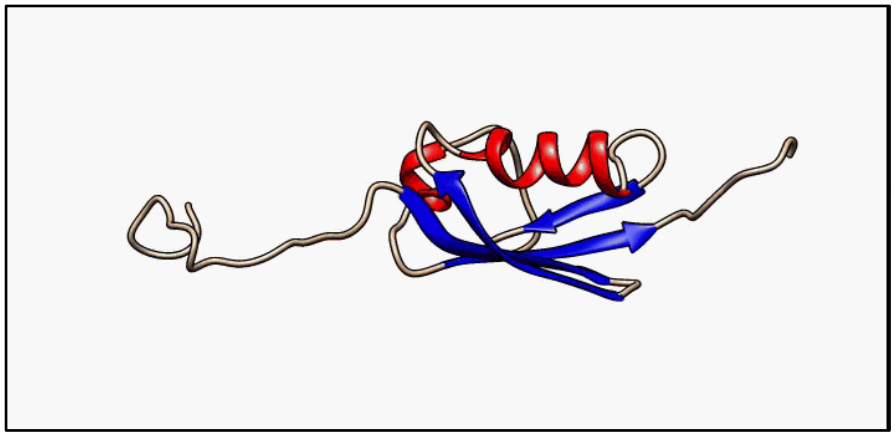
Current limitations of the prediction method

Although the availability of predicted 3D models for the known “protein universe” is an exciting prospect with huge impact, there are nevertheless limitations to the AlphaFold method and resource, some of which may be addressed in the future:

- Many proteins function as complexes with other proteins, nucleic acids (DNA or RNA) or ligands. AlphaFold does not currently predict 3D structures for protein-protein or protein-DNA/RNA/ligand complexes. In some cases, the single-chain prediction may correspond to the structure adopted in a complex. In other cases (especially if the protein is structured only upon binding partner molecules) the missing context from surrounding molecules may lead to an uninformative prediction.
- Proteins are dynamic systems and adopt different structures depending on their environment or state within a functional cycle. Where a protein is known to have multiple conformations AlphaFold will usually only produce one of them. This leaves open many interesting questions about the conformational dynamics of proteins, crucial for understanding biological function, and this will remain a very active area of research.
- For regions that are intrinsically disordered or unstructured in isolation, AlphaFold is expected to produce a low-confidence prediction and the predicted structure will have an extended, ribbon-like appearance. AlphaFold may be of use as a tool for identifying such regions, but the prediction makes no statement about the relative likelihood of different conformations (in biophysical terms: it is not a sample from the Boltzmann distribution). Furthermore, AlphaFold does not claim to predict the “folding pathway”.
- AlphaFold has not been trained or validated for predicting the effect of mutations. In particular, it is not expected to capture the effect of point mutations that destabilise a protein.
- Ligands are not included in the structures since AlphaFold does not make any predictions about any of the non-protein components that are often observed in experimental structures (such as cofactors, metals, ligands including drug-like molecules, ions, carbohydrates and other post-translational modifications).
- As with experimental structures, predicted structures may (or may not) lead to hypotheses about the function of the protein and the mechanism underlying that function, but such hypotheses then have to be tested by further experimentation.

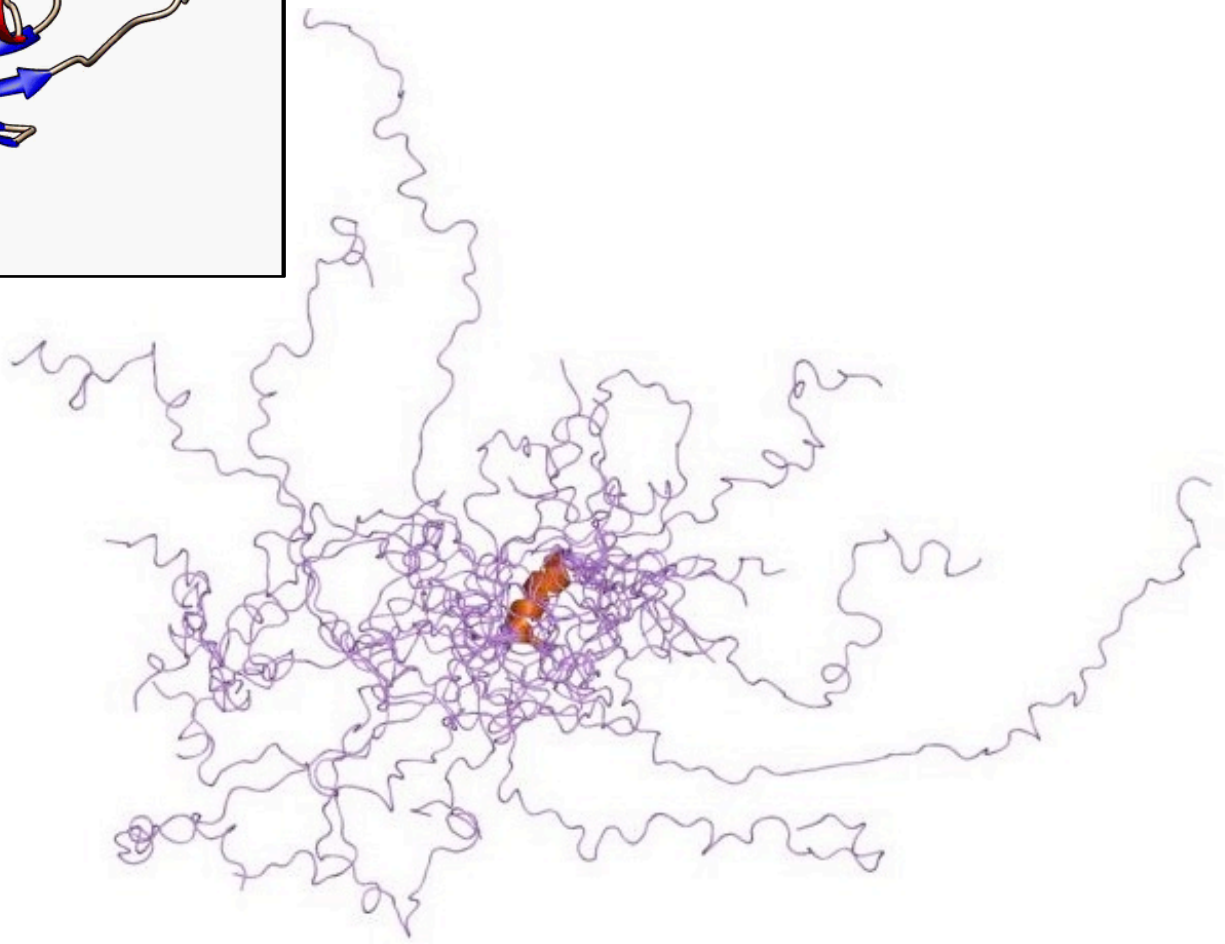
https://www.ebi.ac.uk/about/news/opinion/alphafold-potential-impacts

Are we done?



IDPs

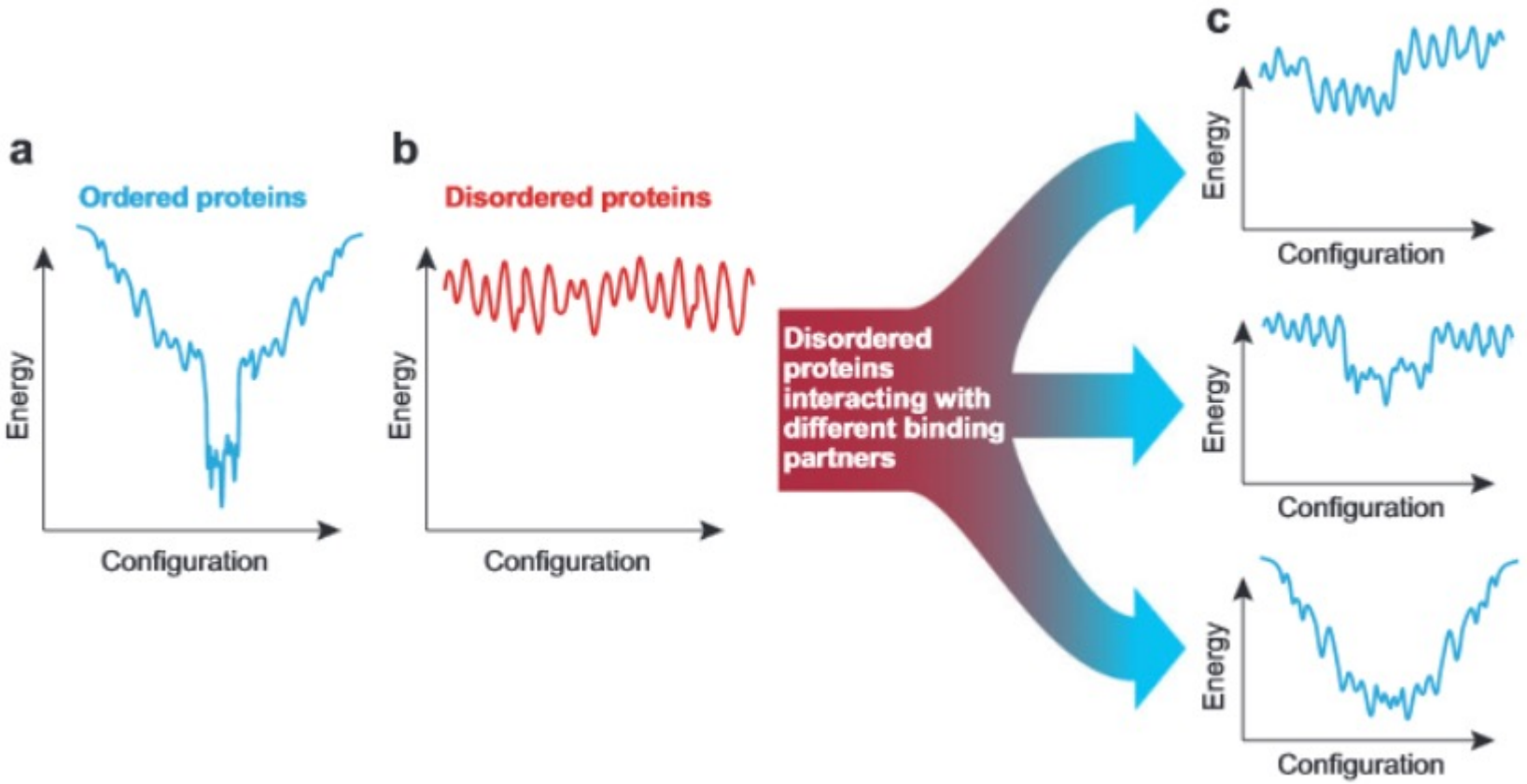
intrinsically disordered proteins



Song et al., *Biochemistry* 45, 15633-15643 (2006)

Are we done?

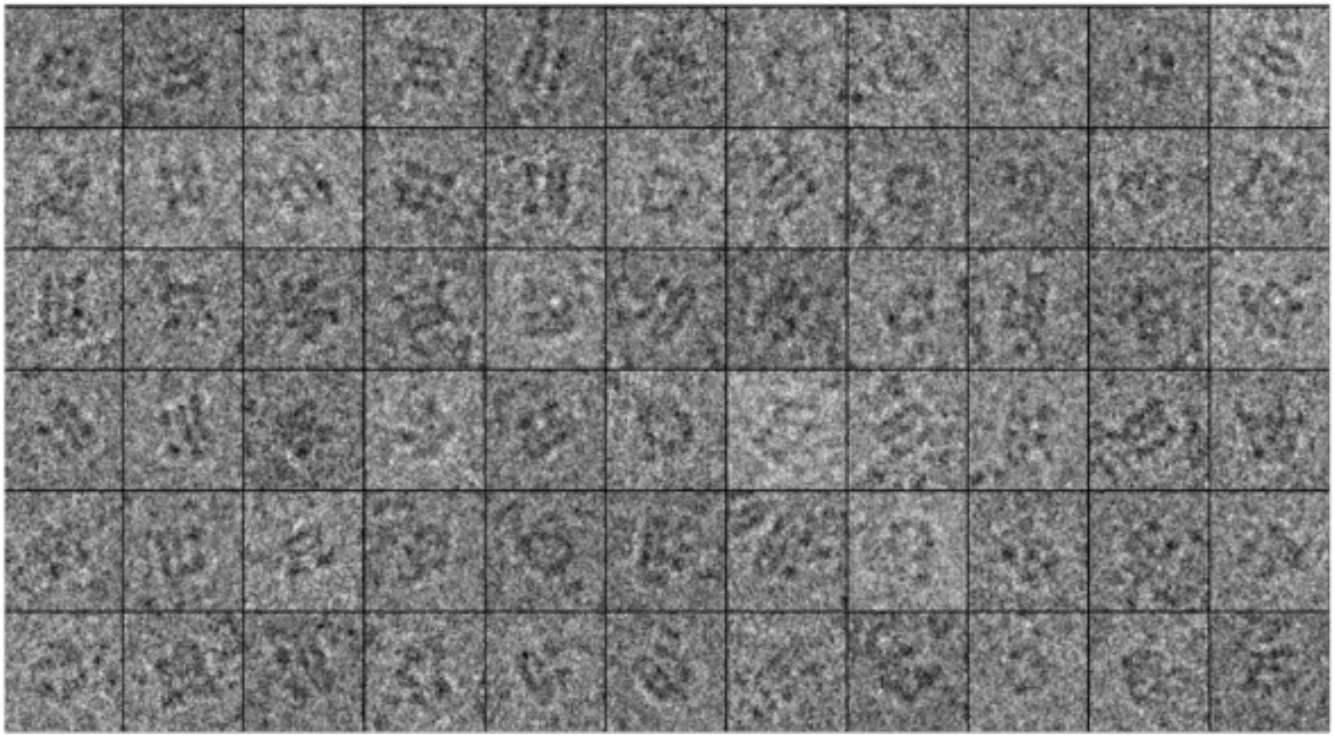
IDPs intrinsically disordered proteins



Uversky et al., *Annu Rev Biophys* 37, 215-246 (2008).

What about the future?

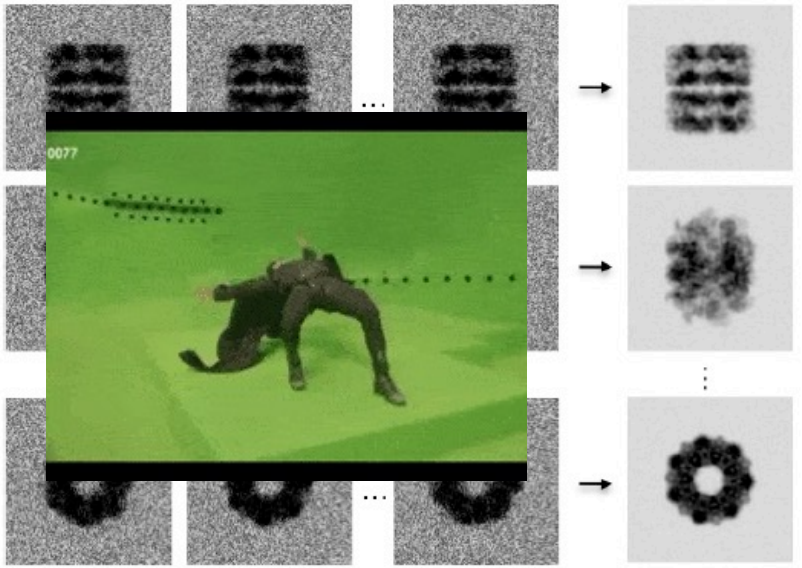
CryoEM



https://hms.harvard.edu/news/future-molecular-visualization

What about the future?

CryoEM



Nogales & Scheres, *Mol Cell* 58, 677-689 (2015).

Current status

Callaway, *Nature* 525, 172-174 (2015).

Try it!

https://colab.research.google.com/github/pablo-arantes/BIEN-249/blob/main/Lecture2_BIEN249.ipynb

Lecture2_BIEN249.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Hello there!

This is a Jupyter notebook to predict protein structure and complex using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see the instructions, checkout the ColabFold GitHub and read their manuscript.

This notebook is part of BIEN 249 class of bioengineering course at University of California, Riverside.

Acknowledgments

- I would like to thank the OpenMM team for developing an excellent and open source engine.
- We would like to thank the AlphaFold team for developing an excellent model and open sourcing the software.
- Söding Lab for providing the computational resources for the MMseqs2 server
- Credit to Sergey Ovchinnikov (@sokrypton), Milot Mirdita (@milot_mirdita) and Martin Steinegger (@thesteinegger) for their fantastic ColabFold
- A Making-it-rain team, **Pablo R. Arantes** (@pablitoarantes), **Marcelo D. Polêto** (@mdpoleto), **Conrado Pedebos** (@ConradoPedebos) and **Rodrigo Ligabue-Braun** (@ligabue_braun).
- Also, credit to David Koes for his awesome py3Dmol plugin.
- For related notebooks see: Making-it-rain

SCAN ME

Hen egg-white lysozyme (Monomer) =
KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITAS
VNCAKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRGCRL
tRNA-specific adenosine deaminase (Complex) =
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMC
AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD:MSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVV
FGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD

Acknowledgements

https://palermolab.com/

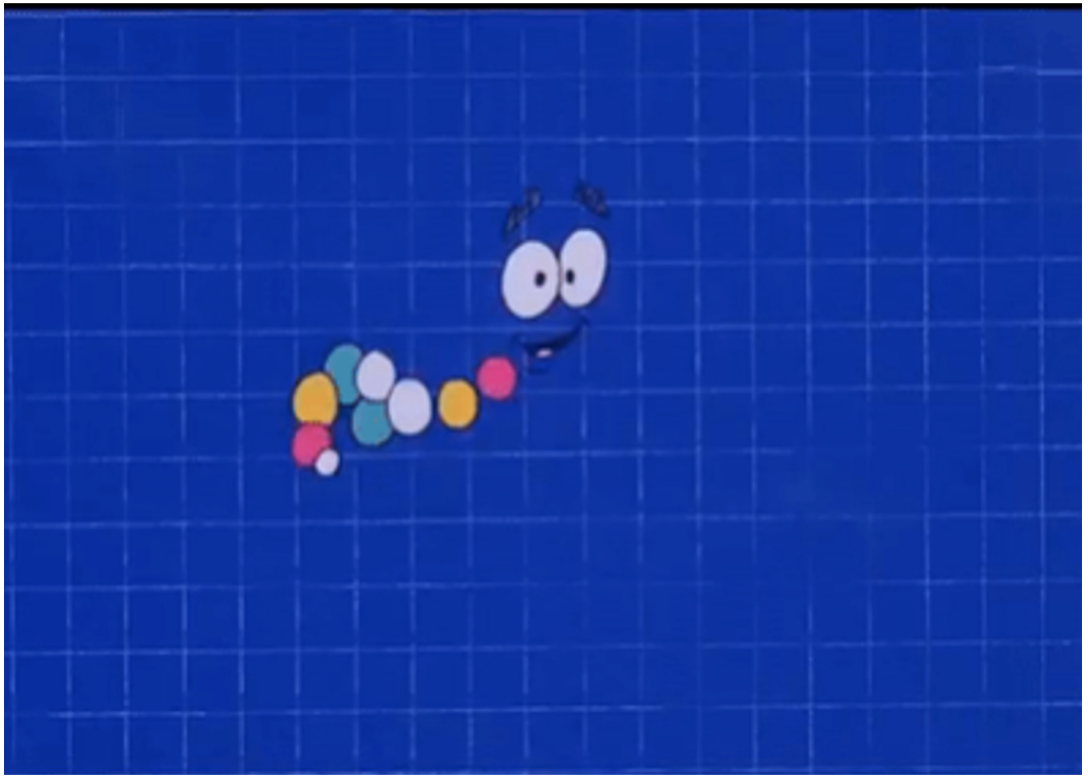
Department of Bioengineering
University of California Riverside
223 Materials Science & Engineering
900 University Ave. | Riverside, CA 92521

Acknowledgements

.website

SCAN ME

pablitoarantes@gmail.com
pabloa@ucr.edu
https://pablo-arantes.github.io



twitter

SCAN ME

Obrigado (Thank you)!

Google It!

AlphaFold and its impact on obtaining protein structures

Pablo Ricardo Arantes, PhD.