(*In silico*) model building

BCH394P/BCH364C

Systems Biology & Bioinformatics

Caitie McCafferty
clmccafferty@utexas.edu

Today's agenda

- Sequence to structure
- Types of *in silico* modeling
  - Comparative modeling (homology modeling)
  - Evolutionary coupling modeling
  - *Ab initio* modeling with neural networks

Anfinsen’s dogma

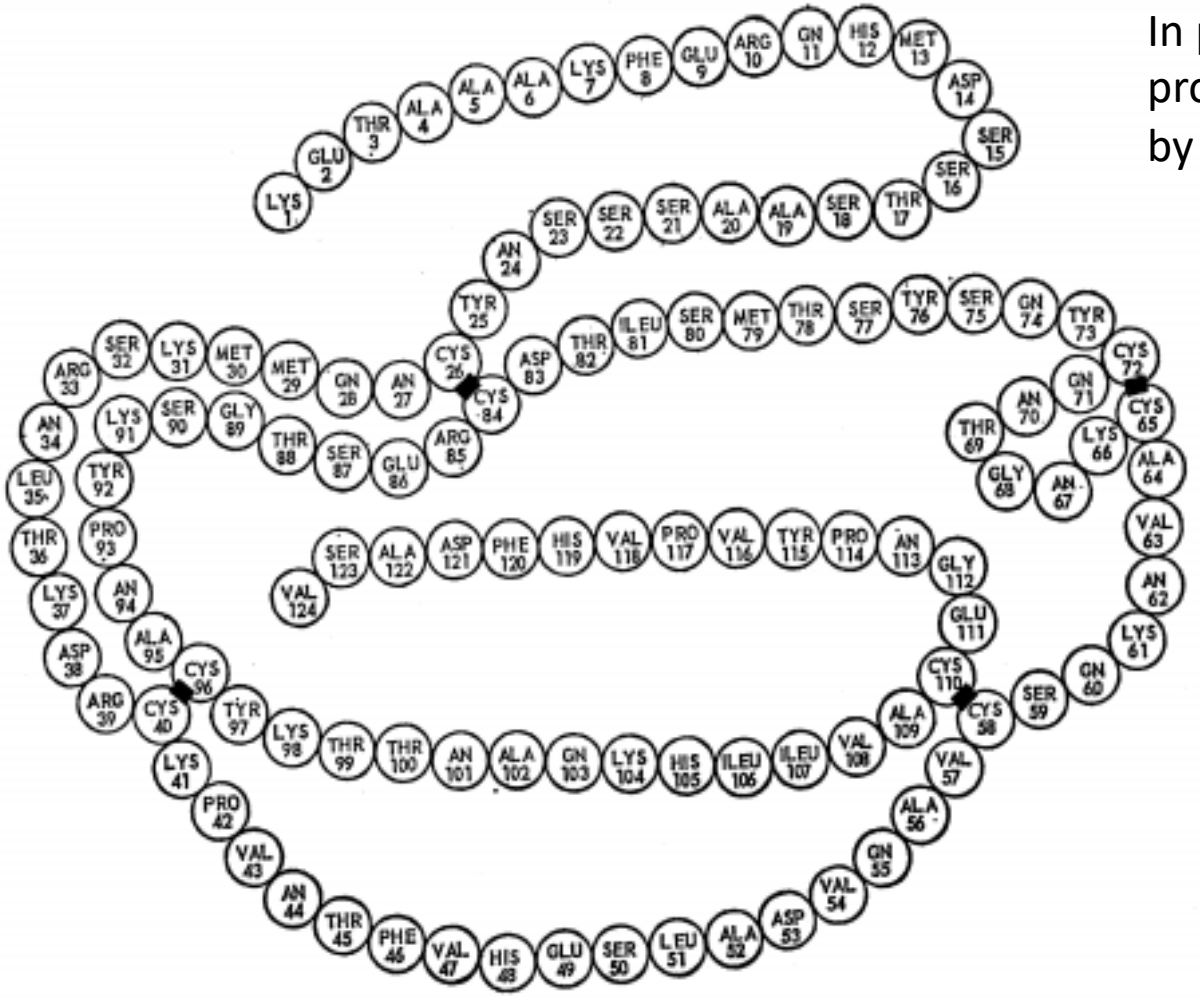


Fig. 1. The amino acid sequence of bovine pancreatic ribonuclease (*50*).

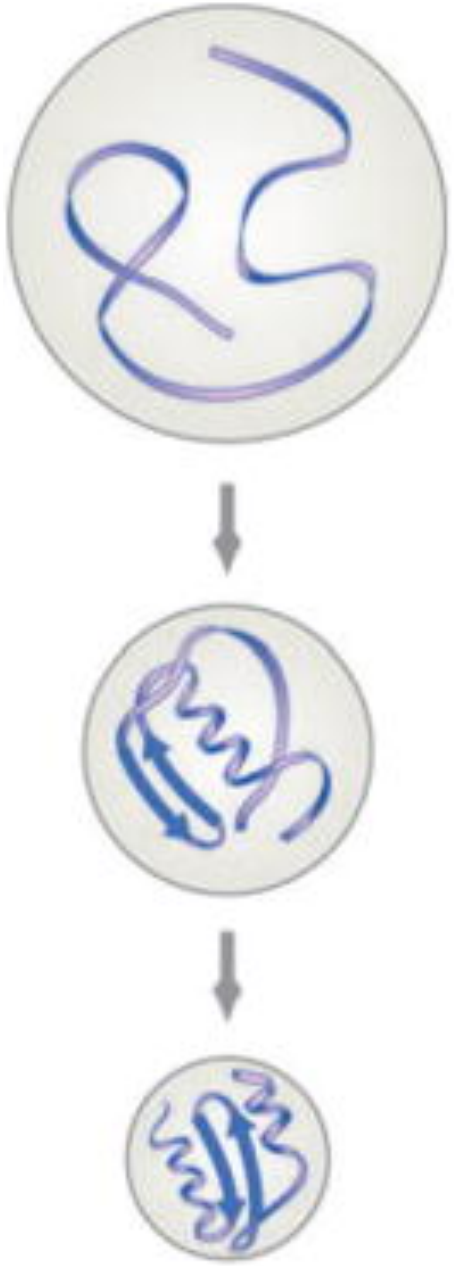
In physiological conditions, a globular protein’s native structure is determined by the protein amino acid sequences.

Is this true?

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*.

Protein folding problem

1. What is the folding code?
2. What is the folding mechanism?
3. Can we predict a native protein structure from its primary, amino acid sequence?



Dill, K. A., et al. (2008). The protein folding problem. *Annu. Rev. Biophys.*

Why might we need to build computational models?

Other considerations: modeling and resolutions based on your need

As with any experiment, keep your intended application in mind…

- do you want to examine a ligand binding site (high resolution)
- or maybe a residue neighborhood (medium resolution)
- or domain boundary definition or even topology classification (low resolution)

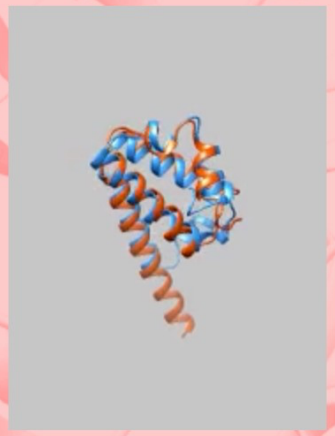
CASP competition to evaluate structure prediction

Protein Structure Prediction Center

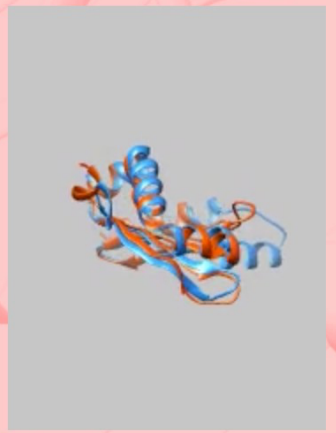
Success Stories From Recent CASPs

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

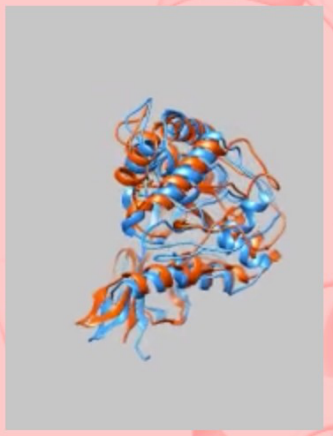
ab initio modeling



CASP7: T0283-D1
model 321_1: GDT_TS=75



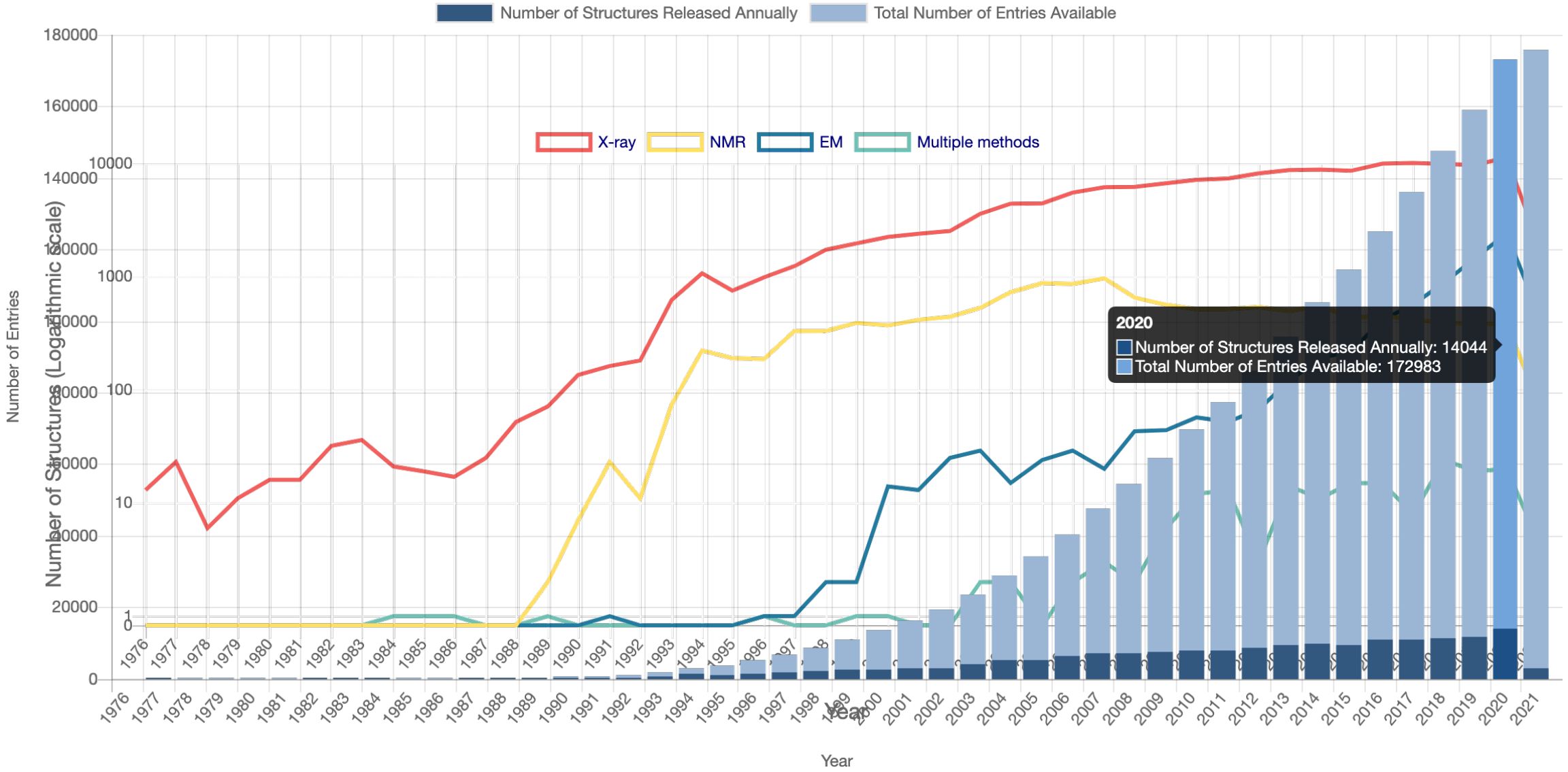
CASP9: T0581-D1
model 170_1: GDT_TS=71



CASP11: T0806-D1
model 064_1: GDT_TS=61

CASP12: T0866-D1
model 325_5: GDT_TS=81

Comparative Modeling: some numbers



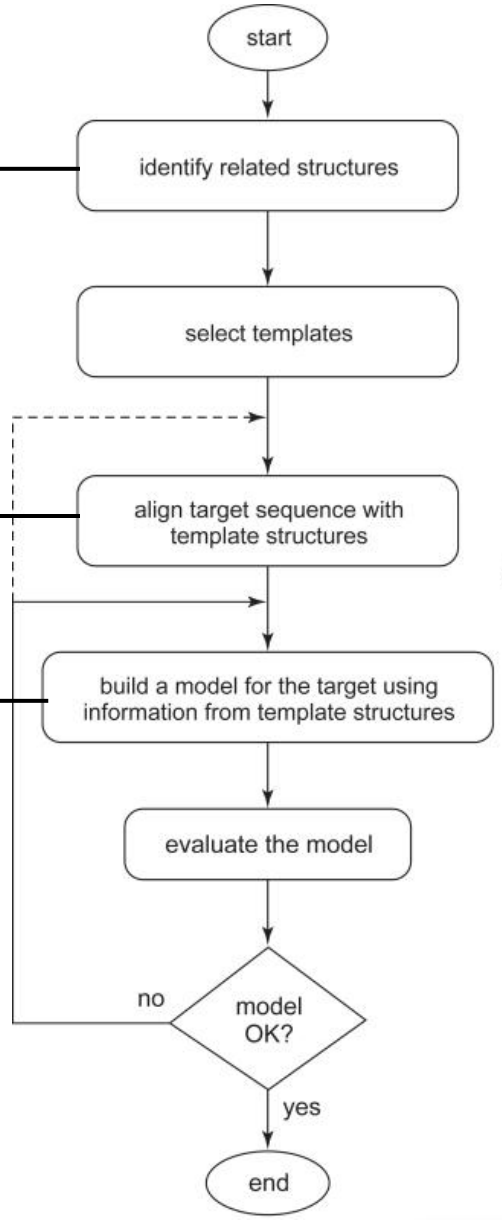
Comparative modeling workflow

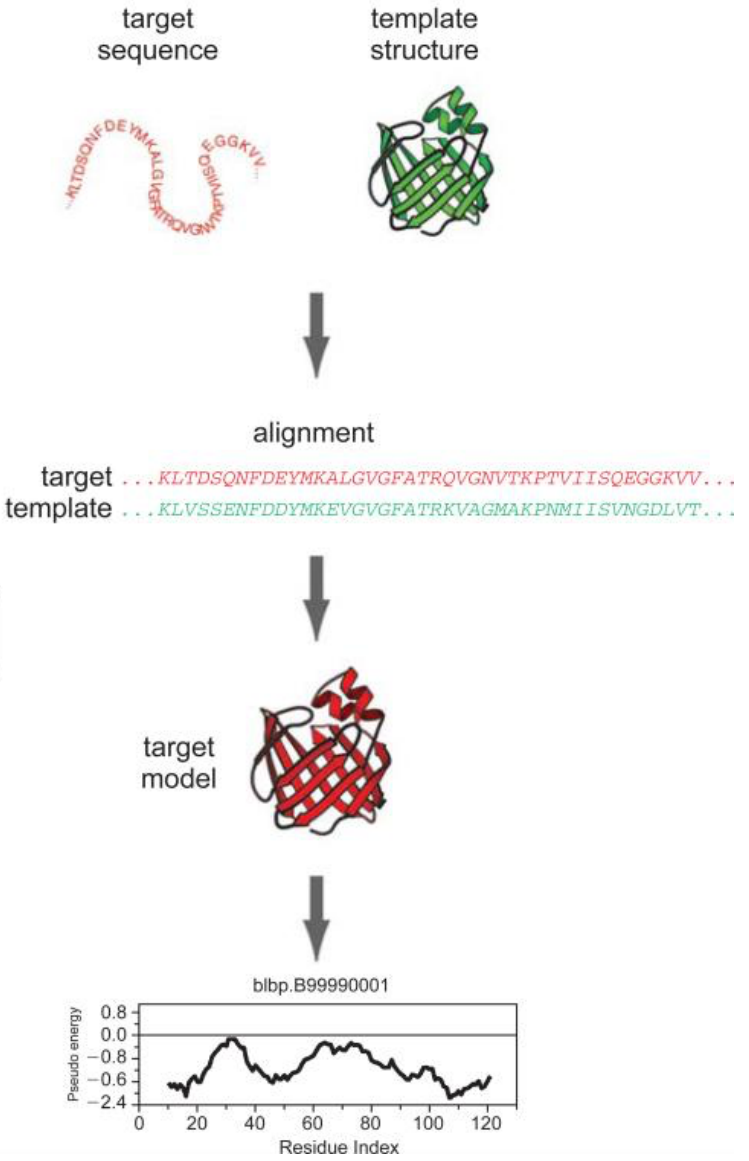
- Sequence-based methods (BLAST)
- Profile-based methods
- HMM methods

Target sequence is aligned to templates using PSSM or HMM approaches

Copy coordinates from template to assemble the backbone of the model

- Energy functions to simulate loops
- Side-chain modeling using rotamer libraries
- Energy minimization to resolve clashes





Eswar, N., Webb, B., Marti-Renom, M. A., ... & Sali, A. (2006). *Current protocols in bioinformatics*.

I-TASSER for comparative modeling

Zhang, Y. (2007). *Proteins: Structure, Function, and Bioinformatics*.

Zhang Lab

UNIVERSITY OF MICHIGAN

Home | Research | Services | Publications | People | Teaching | Job Opening | Facilities | News | Forum | Lab Only

Online Services
- I-TASSER
- QUARK
- LOMETS
- COACH
- COFACTOR
- MetaGO
- MUSTER
- SEGMER
- FG-MD
- ModRefiner
- REMO
- DEMO
- SPRING
- COTH
- BSpred
- SVMSEQ
- ANGLOR
- BSP-SLIM
- SAXSTER
- ThreaDom
- ThreaDomEx
- EvoDesign
- GPCR-I-TASSER
- BindProf
- BindProfX
- ResQ
- IonCom
- STRUM
- DAMpred

I-TASSER

Protein Structure & Function Predictions

(The server completed predictions for 506544 proteins submitted by 119641 users from 144 countries or regions)
(The template library was updated on 2019/10/03)

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database BioLiP. I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments. It was also ranked as the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate structural and function predictions using state-of-the-art algorithms. Please report problems and questions at I-TASSER message board and our developers will study and answer the questions accordingly. (>> More about the server ...)

[**Queue**] [**Forum**] [**Download**] [**Search**] [**Registration**] [**Statistics**] [**Remove**] [**Potential**] [**Decoys**] [**News**] [**Annotation**] [**About**] [**FAQ**]

Our server is undergoing maintenance. You can submit jobs normally during the maintenance period, but they will get queued much slower than usual. We apologize for any inconvenience this may cause.

I-TASSER On-line Server (View an example of I-TASSER output):

Copy and paste your sequence below ([10, 1500] residues in FASTA format). Click here for a sample input:

Or upload the sequence from your local computer:

Choose File no file selected

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click here if you do not have a password)

ID: (optional, your given name of the protein)

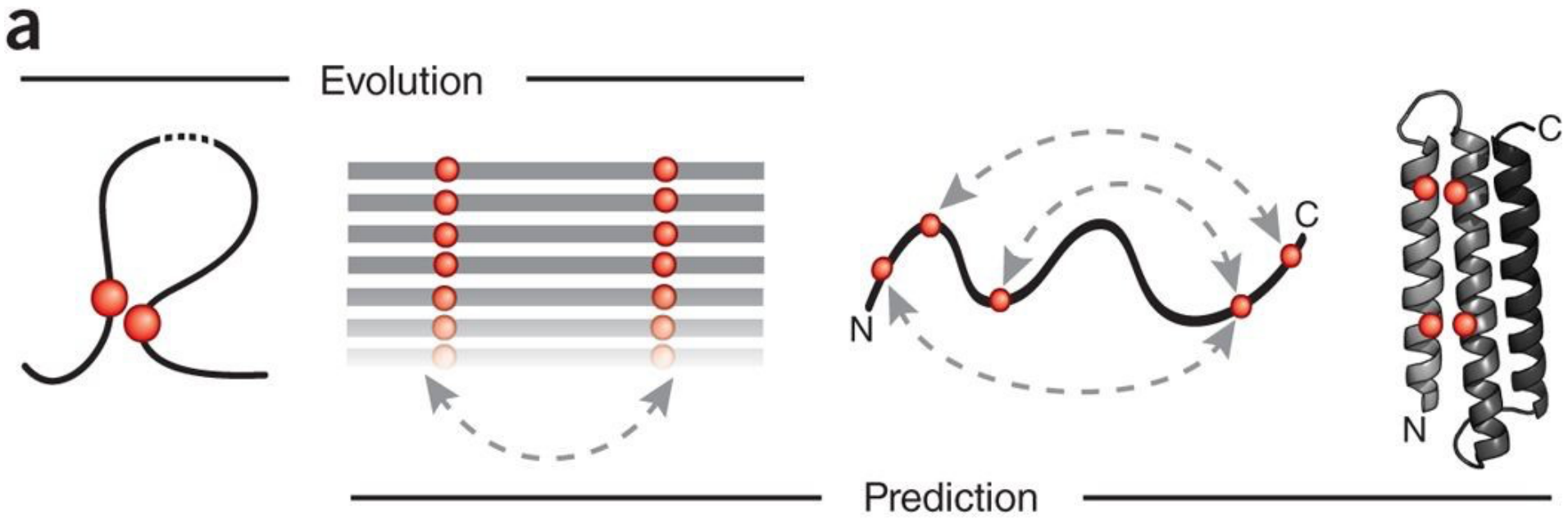
"ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments"

Comparative modeling limitations

Continued reading on comparative modeling

- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., ... & Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current protocols in protein science*, *50*(1), 2-9.
- Song, Y., DiMaio, F., Wang, R. Y. R., Kim, D., Miles, C., Brunette, T. J., ... & Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure*, *21*(10), 1735-1742.
- Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug discovery today*, *14*(13-14), 676-683.
- Larsson, P., Wallner, B., Lindahl, E., & Elofsson, A. (2008). Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Science*, *17*(6), 990-1002.
- Bower, M. J., Cohen, F. E., & Dunbrack Jr, R. L. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of molecular biology*, *267*(5), 1268-1282.
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, *39*(suppl_2), W29-W37.

Evolutionary coupling-based modeling



"The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated."

- Sewall Wright

Marks, D. S., Hopf, T.A., & Sander, C. (2012). *Nature biotechnology*.

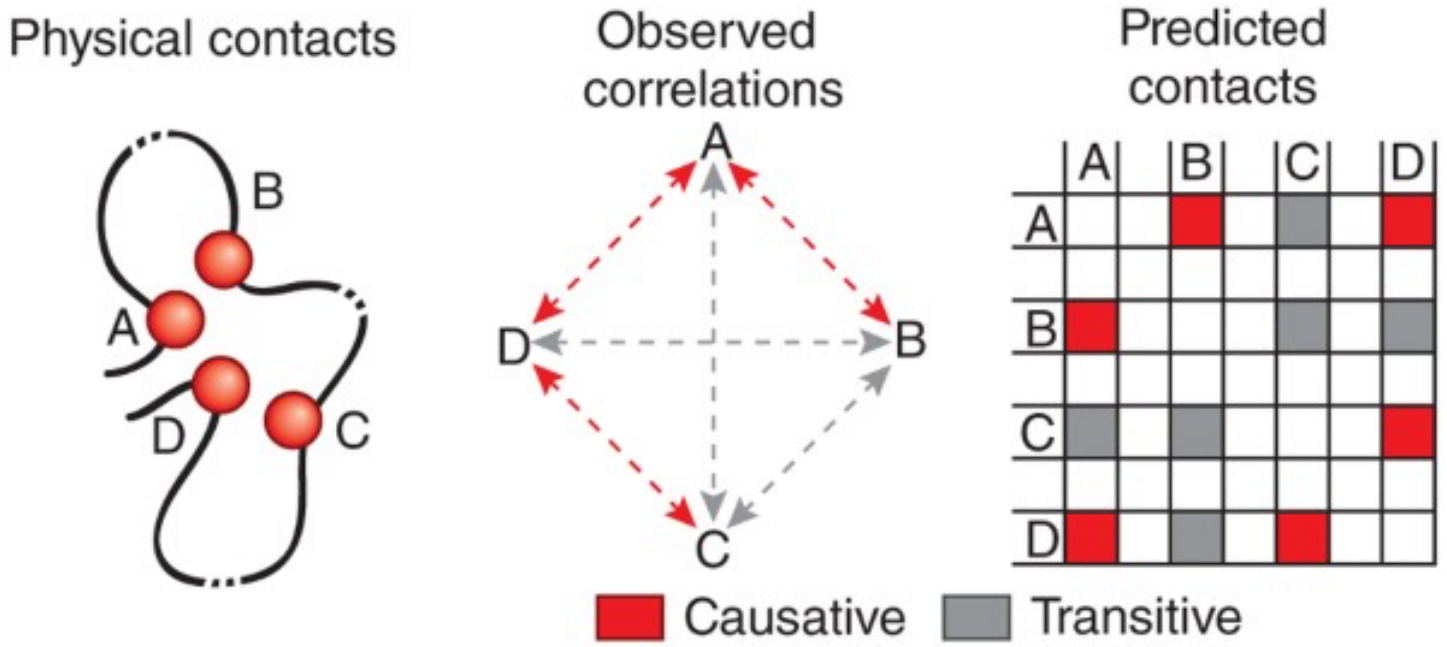
Finding evolutionary covariation

Local statistical models i.e. mutual information

- Assume pairs of residues are statistically independent of other pairs of residues
- Issue: cooperative interaction exist and are crucial in folding

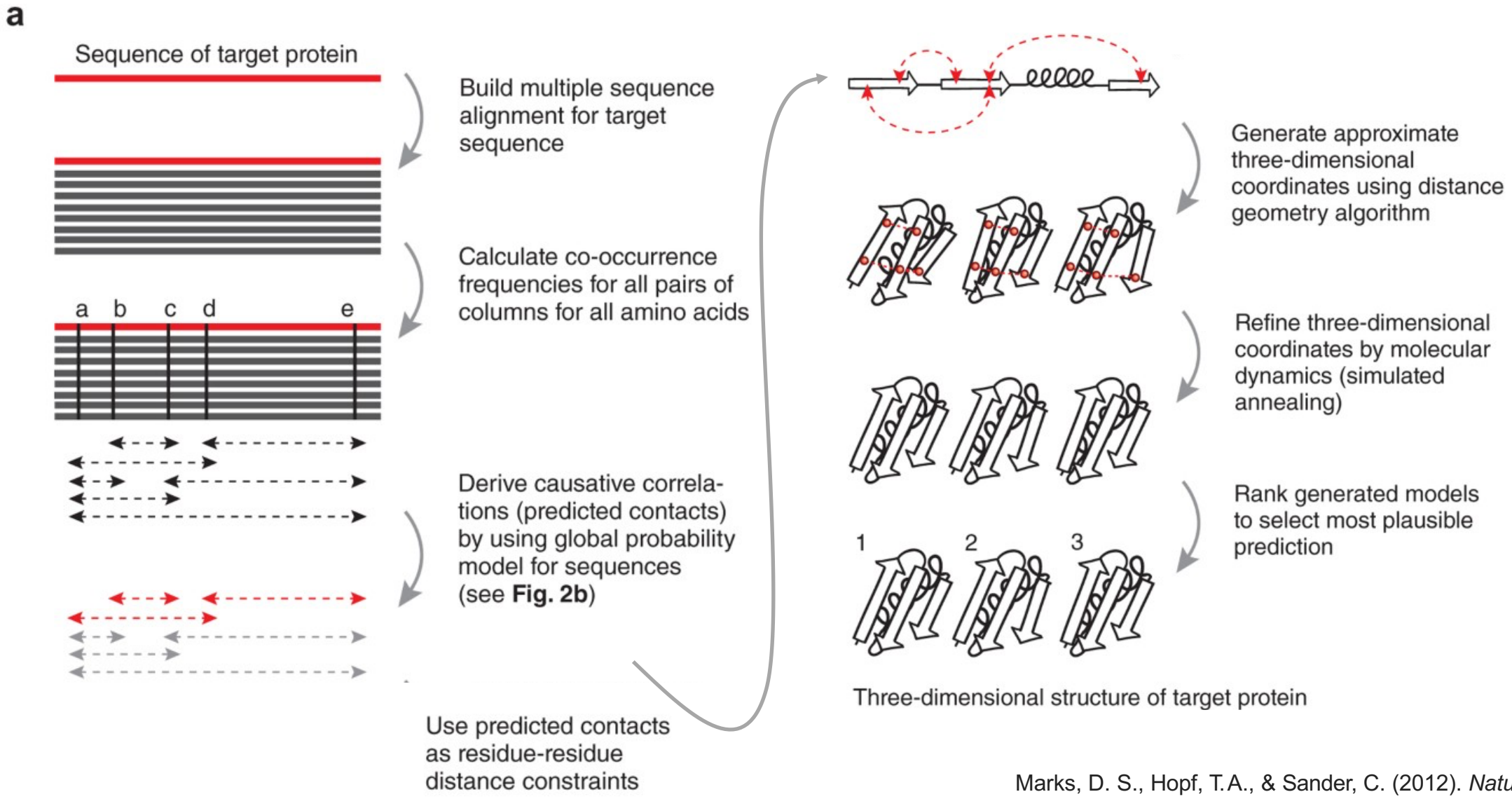
Global statistical models i.e. maximum entropy

- Correlated pairs of residues are dependent on each other
- Given all pair correlations, which best explain all the others (going from correlation to causation)



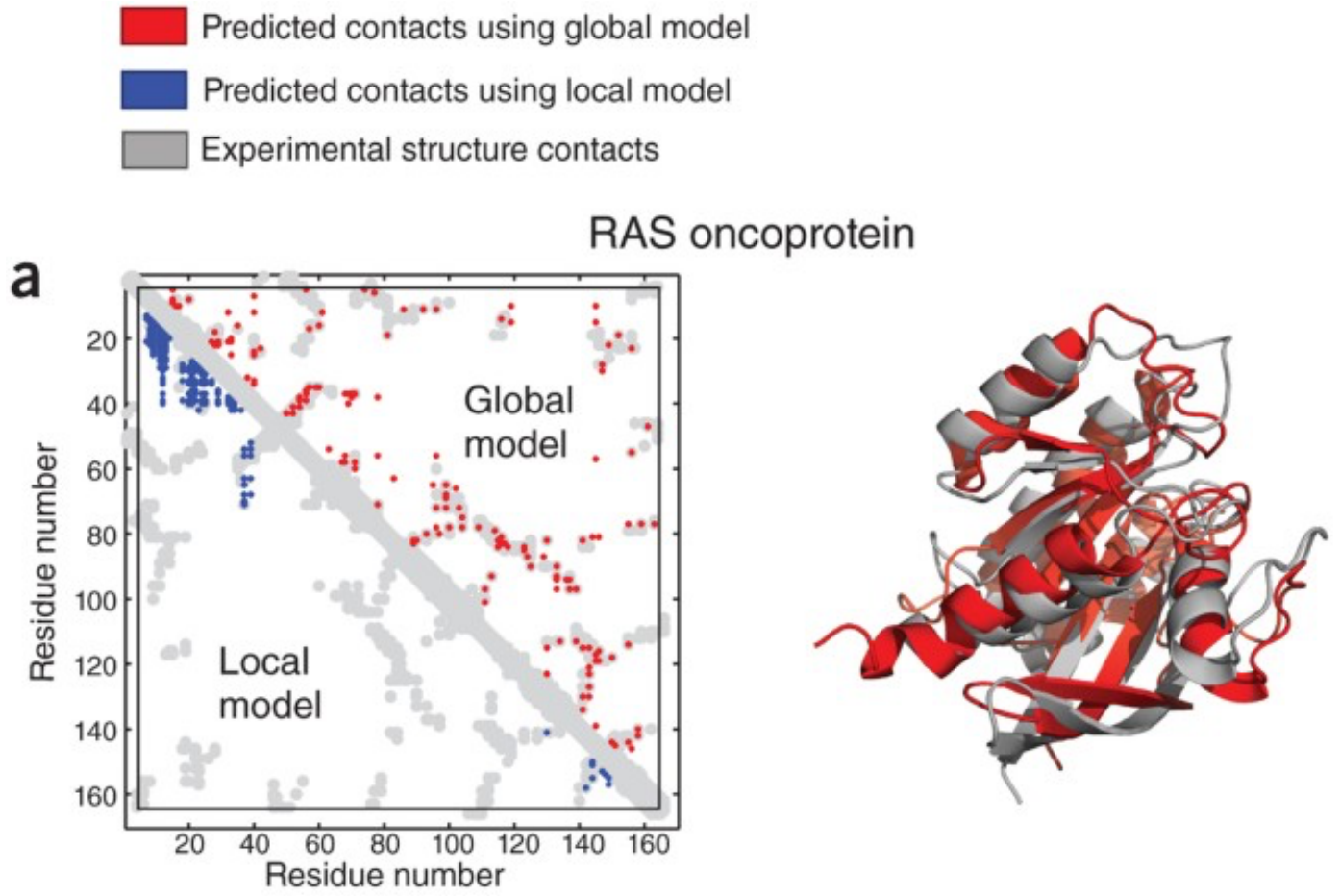
Marks, D. S., Hopf, T.A., & Sander, C. (2012). *Nature biotechnology*.

Detecting evolutionary couplings



Marks, D. S., Hopf, T.A., & Sander, C. (2012). *Nature biotechnology*.

Maximum entropy models for selecting global residue contacts



EVcouplings

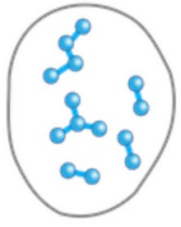
Welcome to EVolutionary Couplings!

The EVolutionary Couplings server provides functional and structural information about proteins derived from the evolutionary sequence record using methods from statistical physics.

SUBMIT JOB >

Webserver and resources

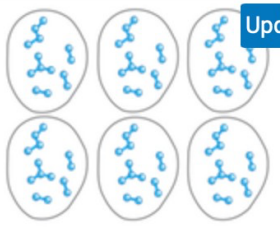
Evolutionary couplings can be used to predict many interesting aspects about protein and RNA molecules from sequence alone. Here is what we have worked on so far:



EVcouplings server

Compute evolutionary couplings from sequence alignments and predict 3D structure for your protein of interest. This webserver allows to run former EVcouplings, EVmutation, EVfold and EVcomplex jobs.

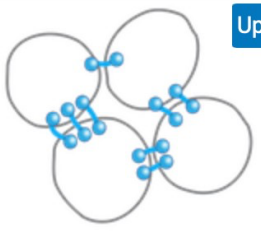
GO



Models

Precomputed evolutionary couplings and 3D models for thousands of experimentally unsolved proteins.

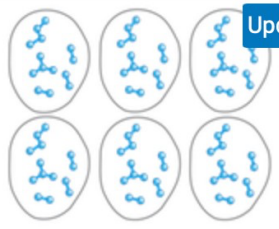
GO



EVcomplex submissions

Predict interacting residues in protein complexes from sequence covariation for your complex of interest.

GO



3Dseq

Data for in-vitro experimental evolution.

GO

Evolutionary modeling limitations

Continued reading on EC modeling

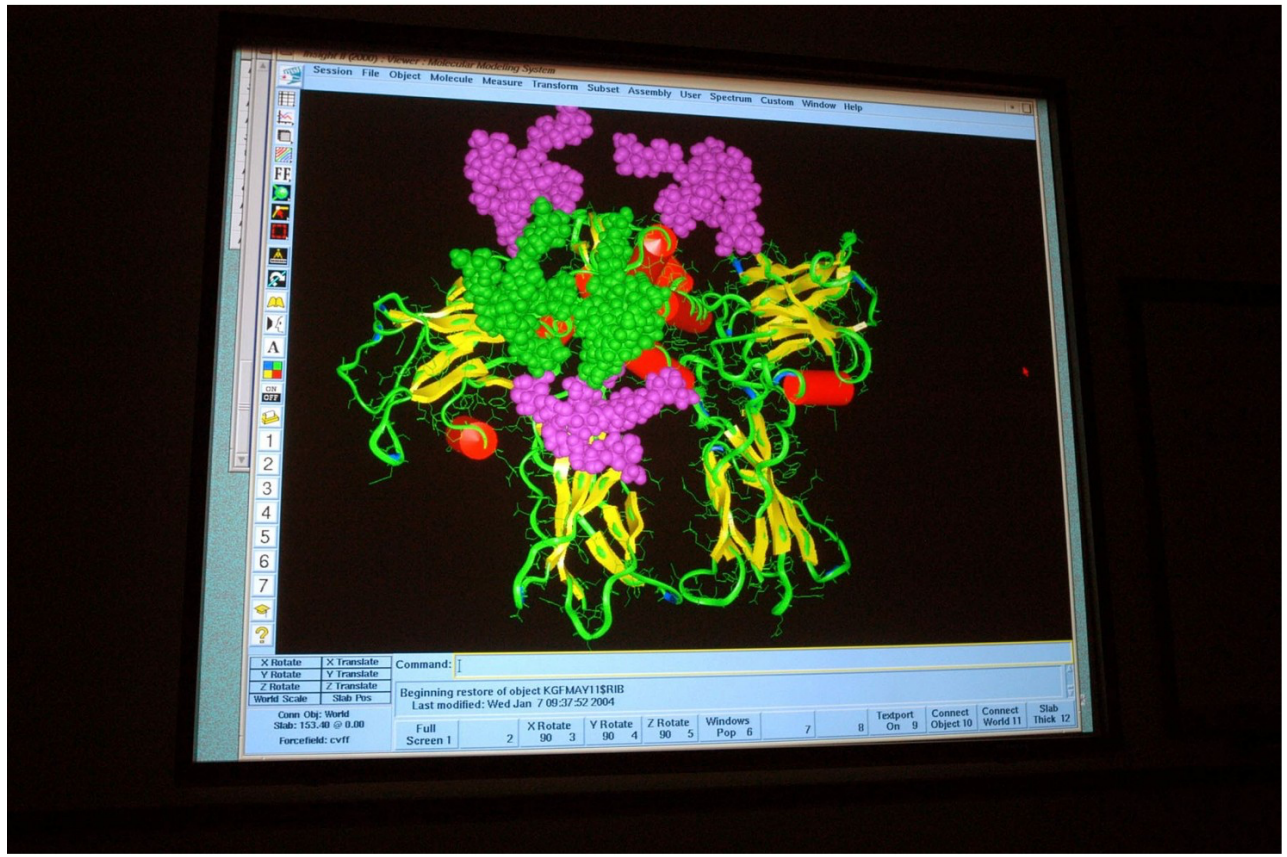
- Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., & Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. *PloS one*, *6*(12), e28766.
- Tang, Y., Huang, Y. J., Hopf, T. A., Sander, C., Marks, D. S., & Montelione, G. T. (2015). Protein structure determination by combining sparse NMR data with evolutionary couplings. *Nature methods*, *12*(8), 751-754.
- Hopf, T. A., Colwell, L. J., Sheridan, R., Rost, B., Sander, C., & Marks, D. S. (2012). Three-dimensional structures of membrane proteins from genomic sequencing. *Cell*, *149*(7), 1607-1621.
- Kamisetty, H., Ovchinnikov, S., & Baker, D. (2013). Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-and structure-rich era. *Proceedings of the National Academy of Sciences*, *110*(39), 15674-15679.
- Ovchinnikov, S., Park, H., Varghese, N., Huang, P. S., Pavlopoulos, G. A., Kim, D. E., ... & Baker, D. (2017). Protein structure determination using metagenome sequence data. *Science*, *355*(6322), 294-298.

How one scientist coped when AI beat him at his life's work

A Harvard biologist on his journey from melancholy to acceptance.

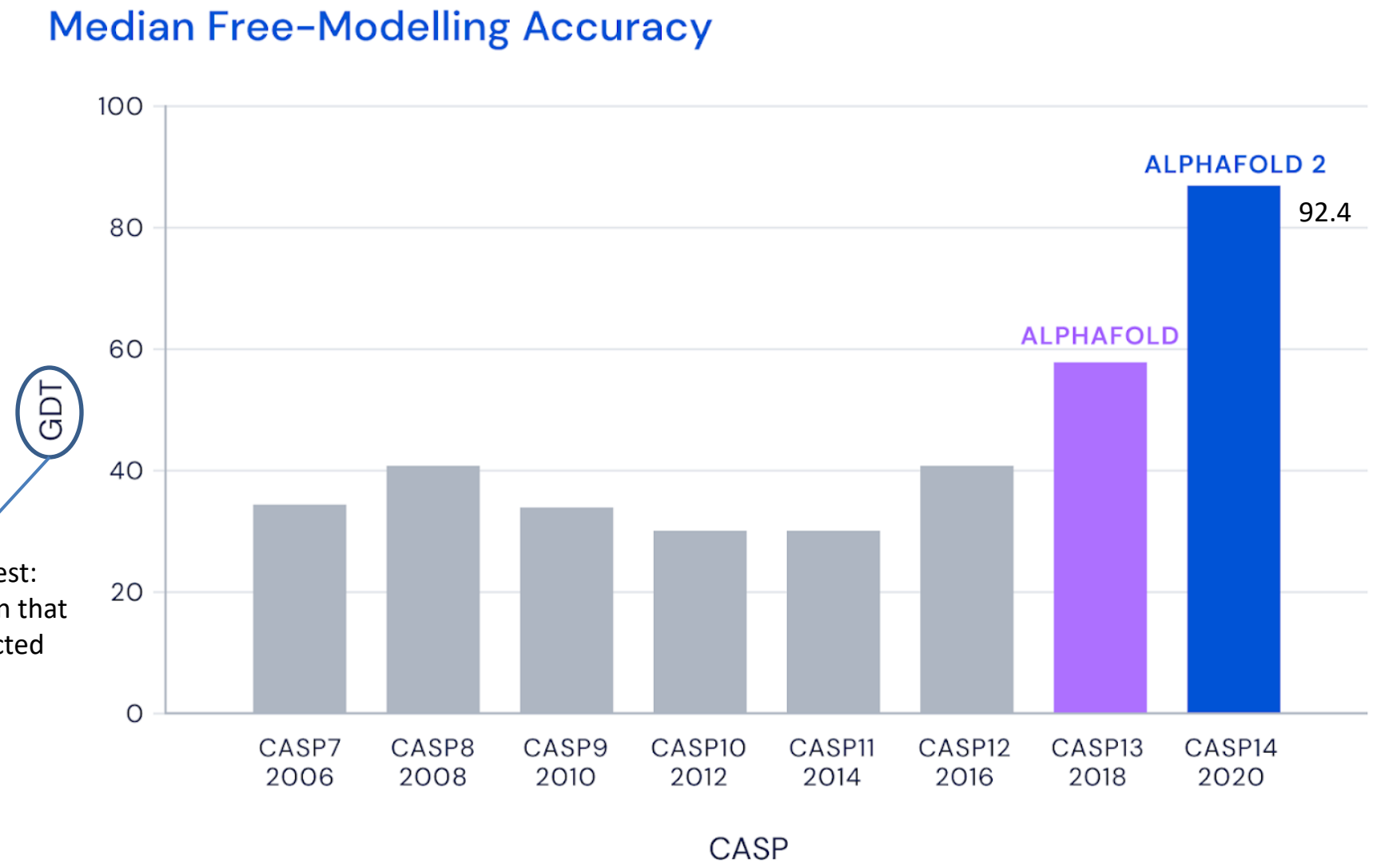
By Sigal Samuel | Feb 15, 2019, 4:10pm EST

SHARE



A protein bound to a receptor is shown in the Molecular Graphics Lab in California. | Ann Johansson/Corbis via Getty Images

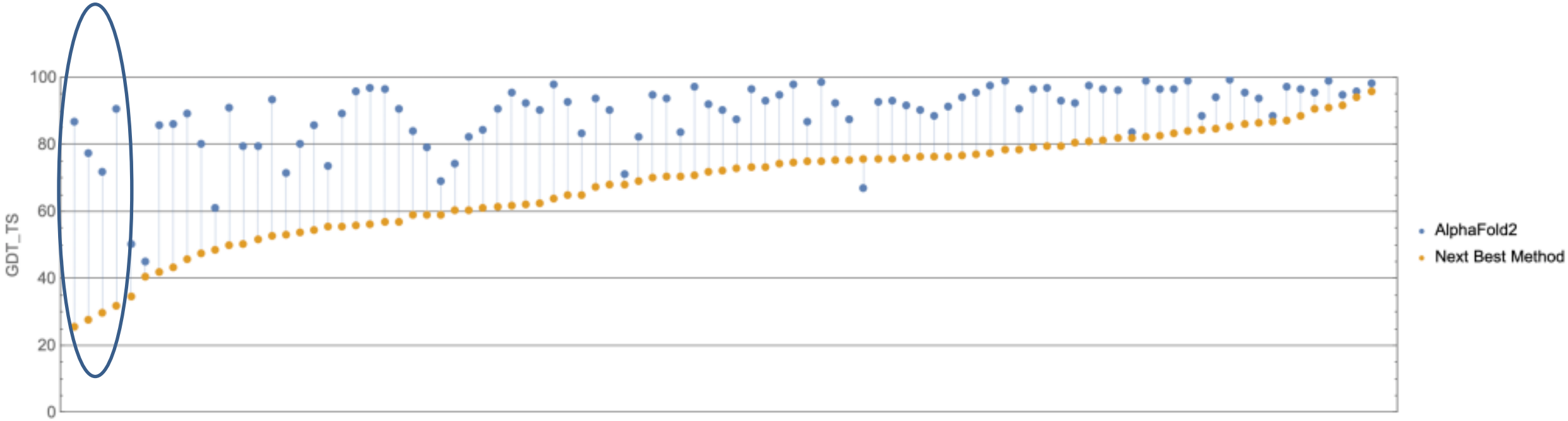
Best team in CASP performance over the years



Global distance test: fraction of protein that is correctly predicted

https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

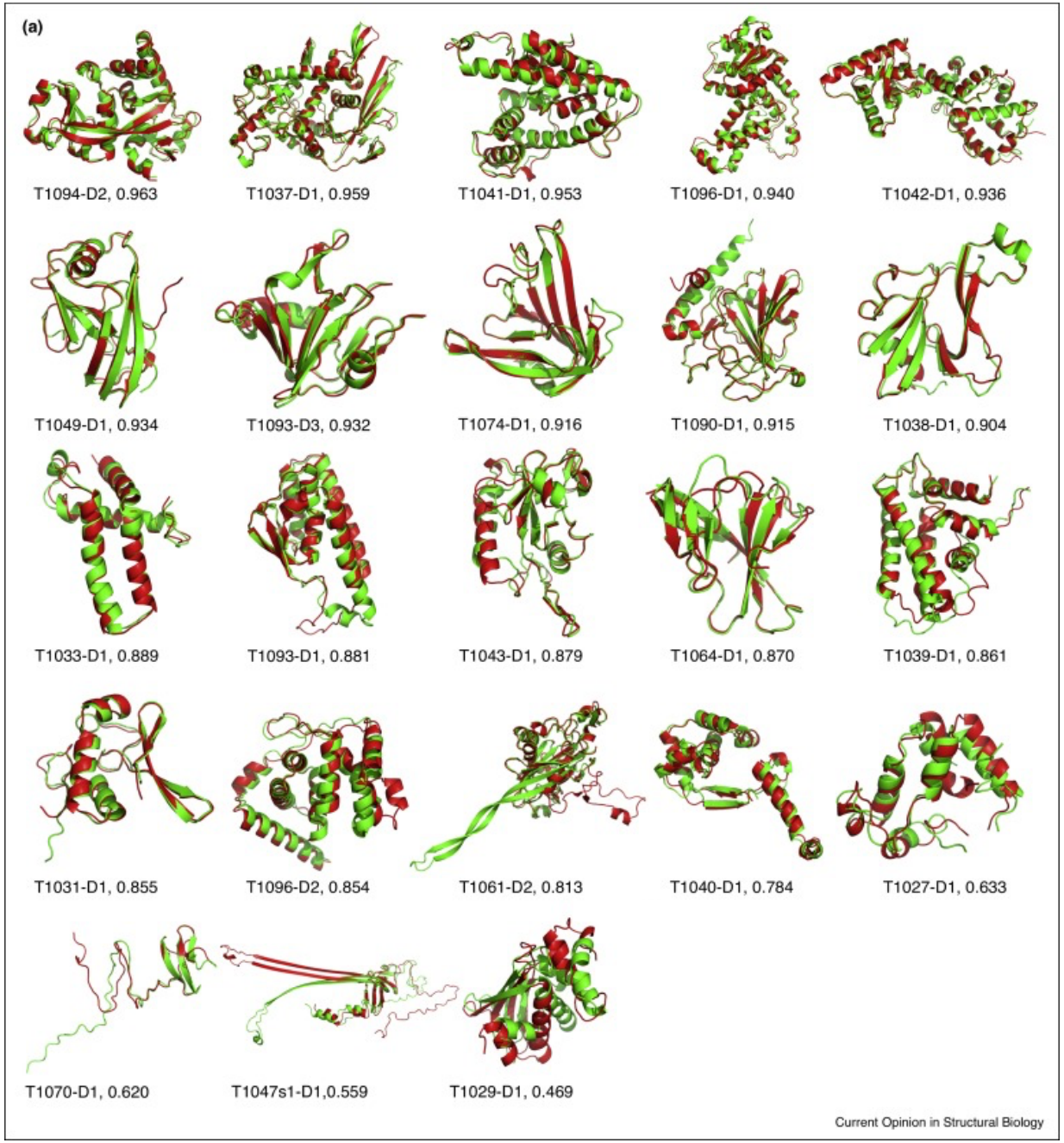
How much better did AF2 do over other methods?



97 total targets

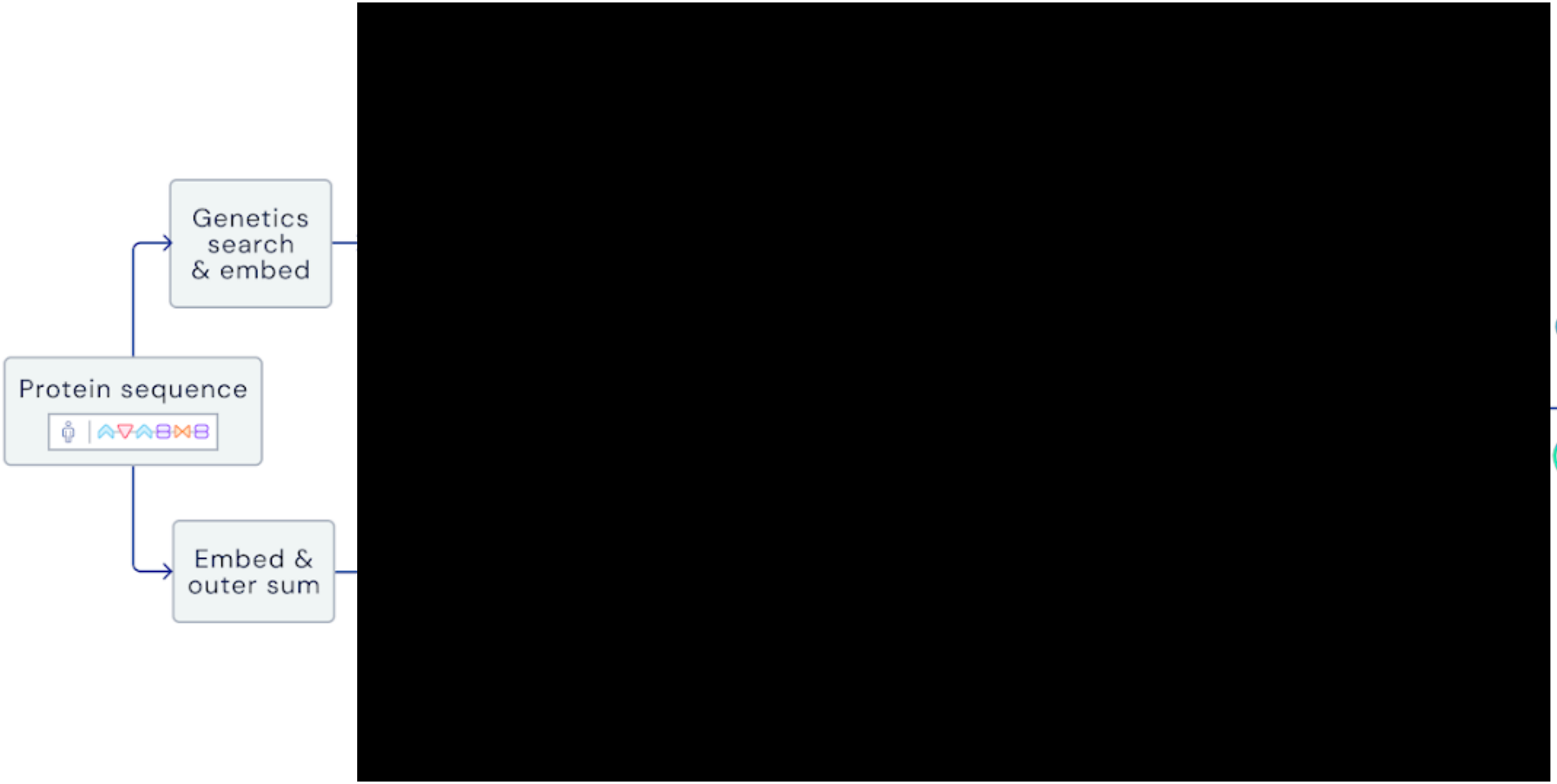
https://moalquraishi.wordpress.com

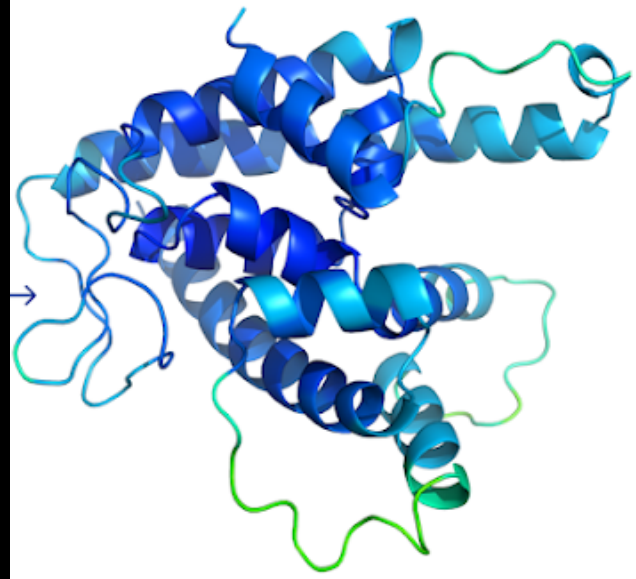
AlphaFold2 at CASP14



Pearce, R., & Zhang, Y. (2021). *Current Opinion in Structural Biology*.

This is exciting but how does it work?!





3D structure

https://deepmind.com/

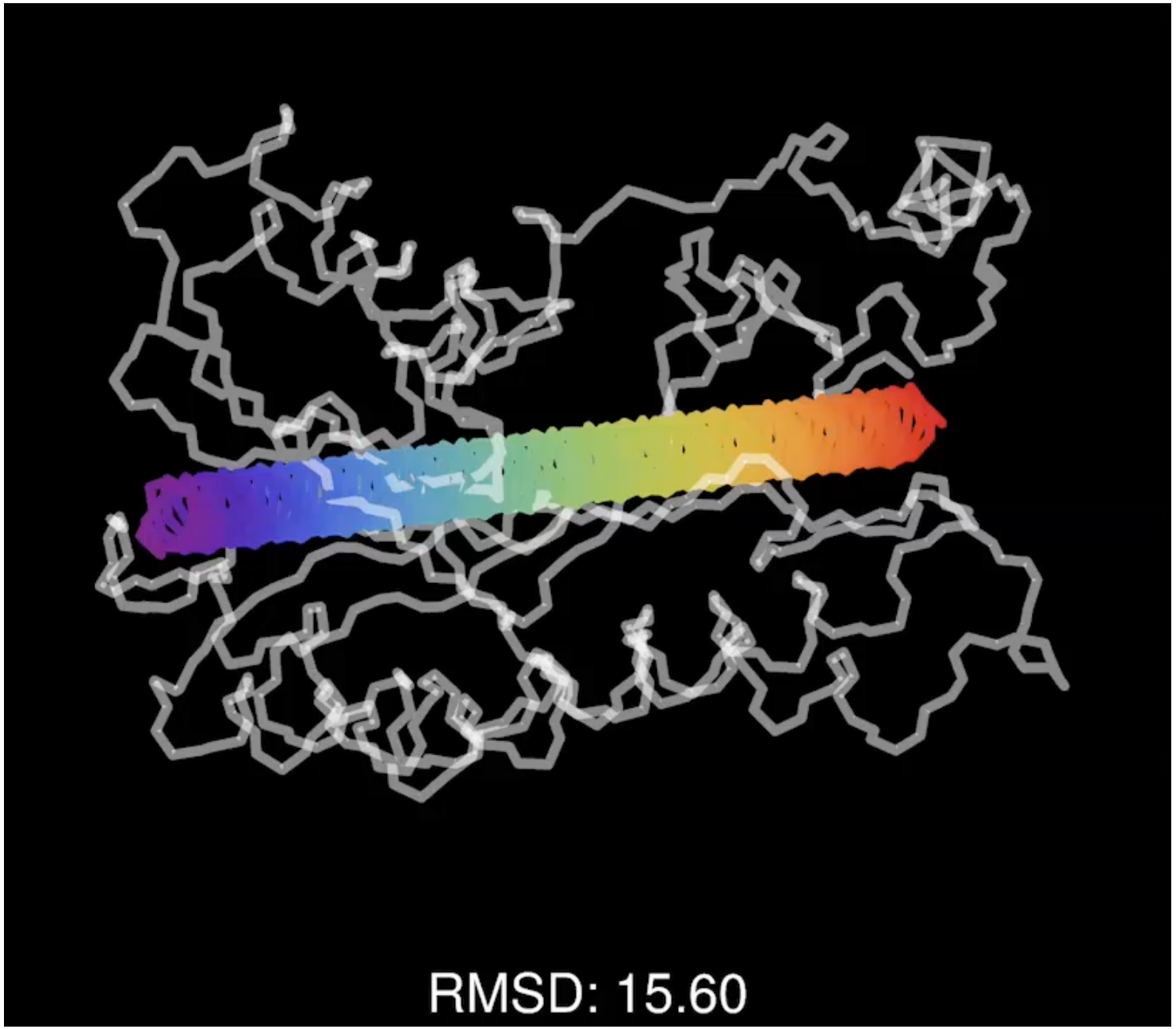
“attention-based neural network system, trained end-to-end”

Attention-based neural network

- Enhancing important parts of the input data and fades out the rest, in the case of the MSA attends to sequences it determines are more important

Trained end-to-end

- Allows for parameters to be tuned from beginning to end to optimize final structure, limits the amount of clashes than can occur from methods that output all coordinates at the same time, needs to be self-consistent



Ab initio NN modeling limitations

Continued reading on *Ab initio* modeling

- AlQuraishi, M. (2019). End-to-end differentiable learning of protein structure. *Cell systems*, *8*(4), 292-301.
- Pearce, R., & Zhang, Y. (2021). Deep learning techniques have significantly impacted protein structure prediction and protein design. *Current Opinion in Structural Biology*, *68*, 194-207.
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706-710.
- AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics*, *35*(22), 4862-4865.
- Xu, D., & Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function, and Bioinformatics*, *80*(7), 1715-1735.

Modeling servers

- Comparative modeling
  - https://zhanglab.ccmb.med.umich.edu/I-TASSER/
- Evolutionary modeling
  - https://evcouplings.org/job
- *Ab initio* modeling
  - http://bioinf.cs.ucl.ac.uk/psipred/

Thank you!

Questions?

How can we evaluate models?

- Root-mean-square deviation (RMSD) → for very similar structures
- Molecular dynamics simulation
- Ramachandra Plot

Ramachandran Plot