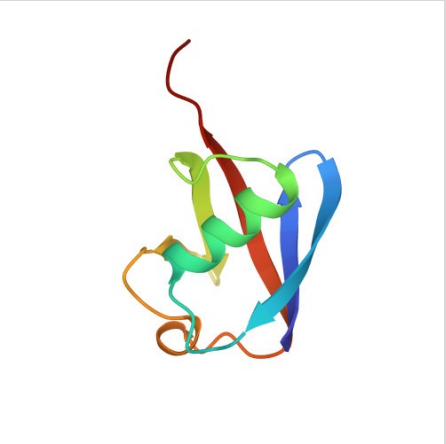
Dive into computational physical chemistry

Lecture 6: Visualizing and predicting protein structures

Glen Hocky
October 17, 2024

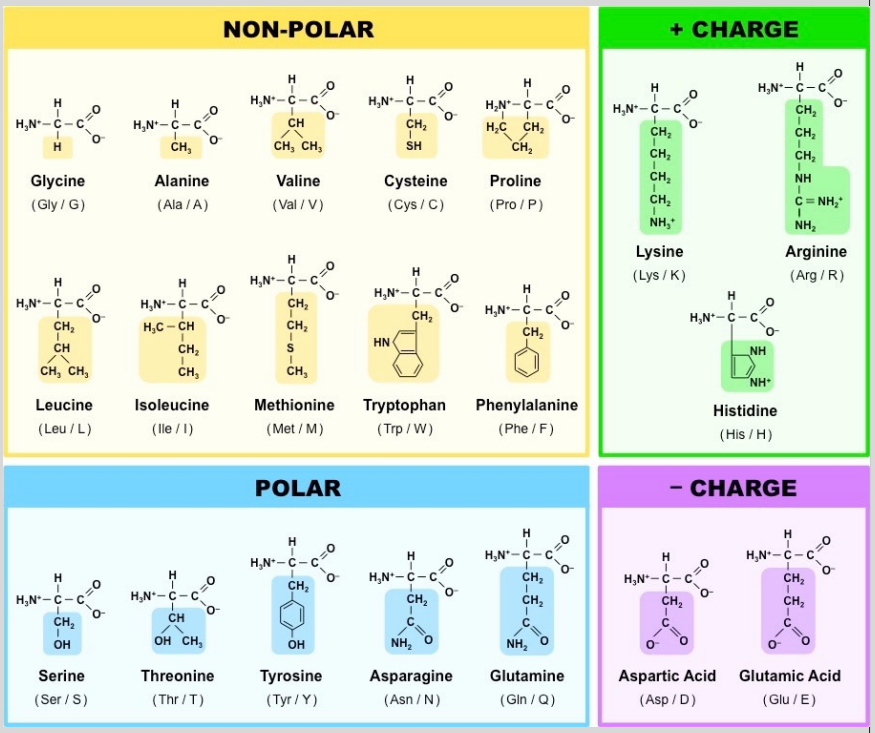
Protein structures

- Proteins do many of the functions in our bodies
- They are formed from chains of amino-acids that “fold” into three dimensional arrangements
- The three-dimensional structure determines the function, in general
- Structures are “Solved” using X-ray crystallography, electron microscopy or NMR, and deposited in the Protein Data Bank (https://www.rcsb.org/)



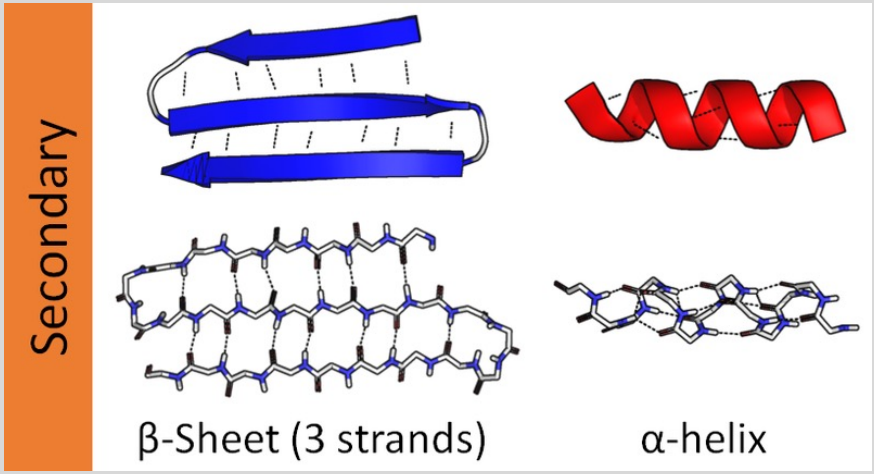
Protein primary sequence

- Proteins are typically composed of sequences of 20 common amino acids
- In PDB, you will see these given as FASTA files, e.g.
  - >1UBQ_1|Chain A|UBIQUITIN|Homo sapiens (9606)
    MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQ
    QRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG



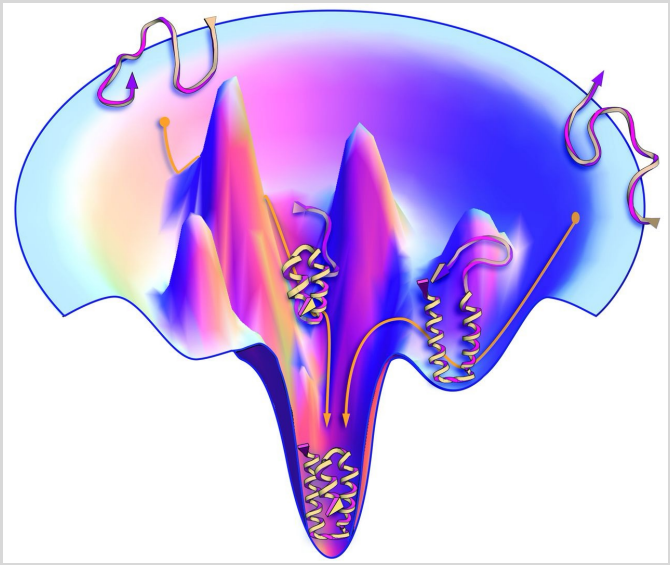
Protein secondary structure

- Proteins have locally defined structures called “secondary structure” that are repeated motifs
- Examples include sheets and helices. Programs like DSSP can assign from a PDB file
- Special cases like metal or ligand bound structures, disulfide bonds add complexity



Tertiary structure: Protein folding problem

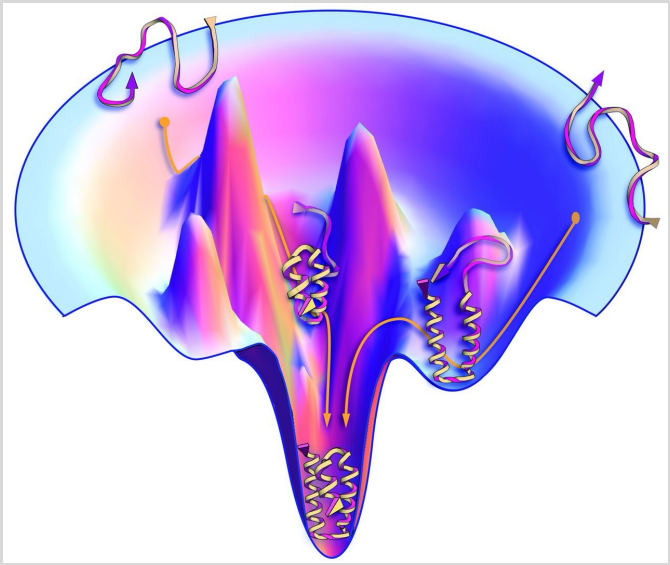
- Proteins have astronomically large number of configurations they can be in, how is the lowest free energy one found?
- Local motifs could produce single/multiple pathways or funnel picture, which makes it much less complicated
- Some protein folding facilitated by chaperones



Dill and coworkers

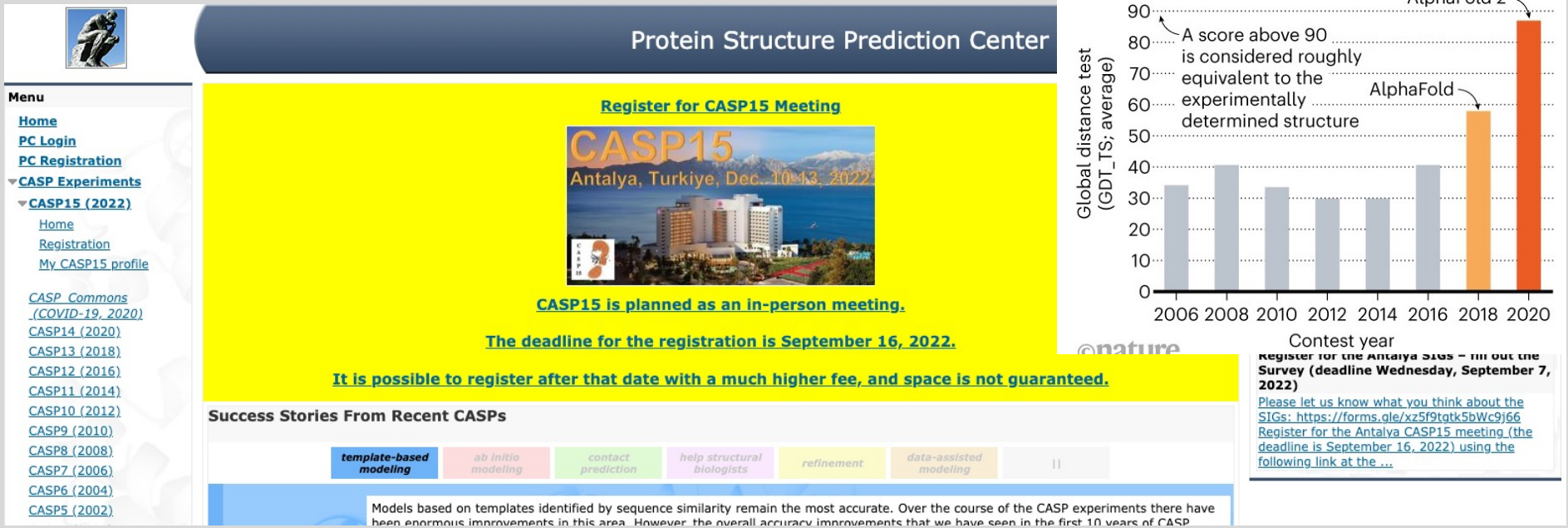
Tertiary structure: Protein structure prediction problem

- How can we take sequence and predict final structure
  - MD – could work, but real folding can be way too slow
  - Physical modeling/coarse-grained MD. Can try to make a lower resolution model that folds faster: this sort of works!
  - Information based modeling. Similar idea, but based on data from PDB: works better
  - Homology modeling: works great when data available
  - New frontier: deep learning

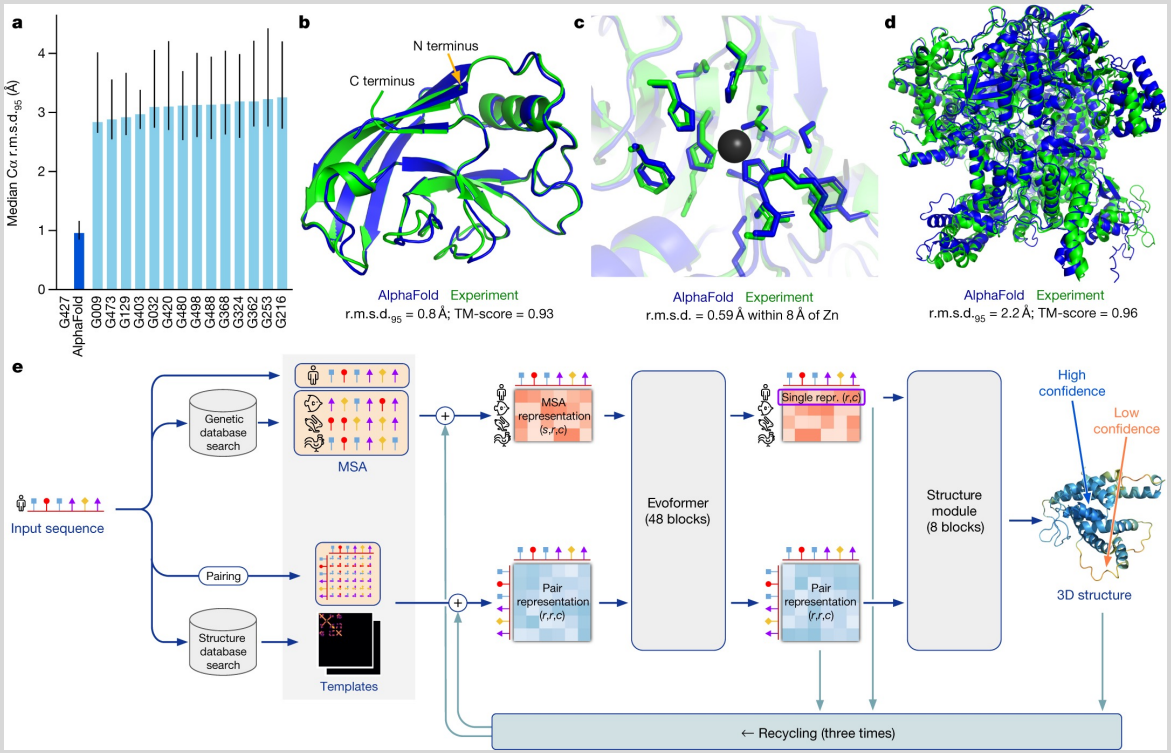


Dill and coworkers

How do we know if it's working?



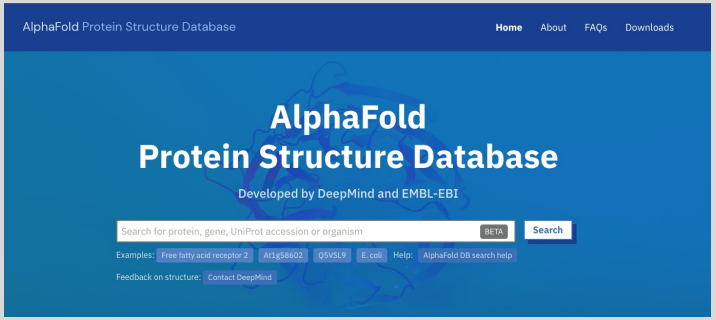
Alphafold2 general architecture

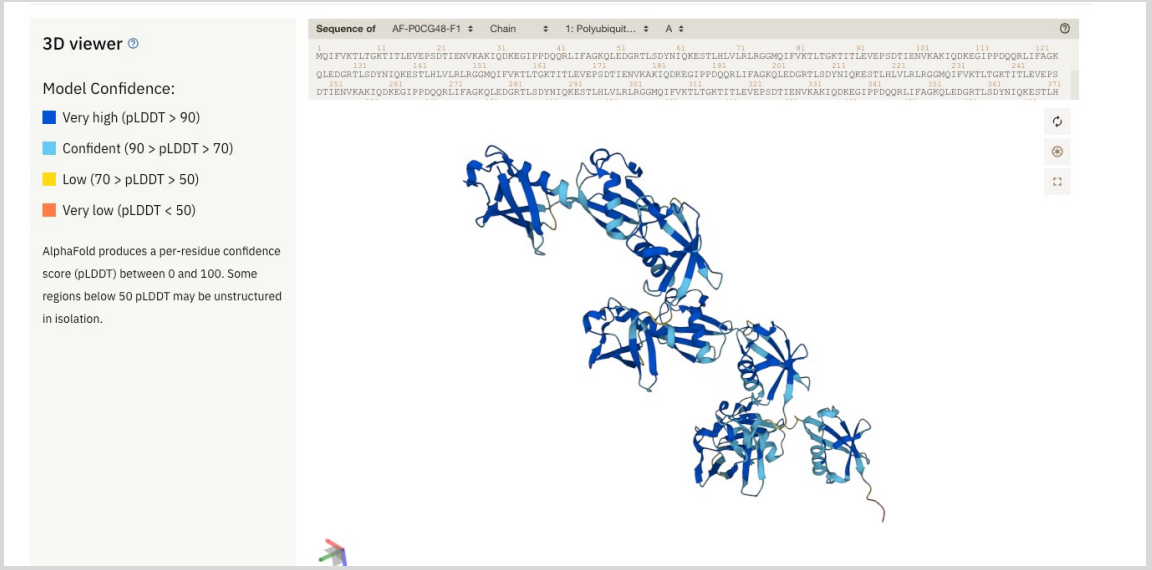


- Generate and use “Multiple Sequence Alignment” to help with pair contacts (see e.g. Ranganathan, Nature 2005)
- Use templates (doesn’t add much)
- Use MD refinement (doesn’t add much)
- Simultaneously predict confidence (pLDDT score) of each residue

Jumper et al. Nature, 2021; See also Baek et al, Science 2021, RoseTTAFold

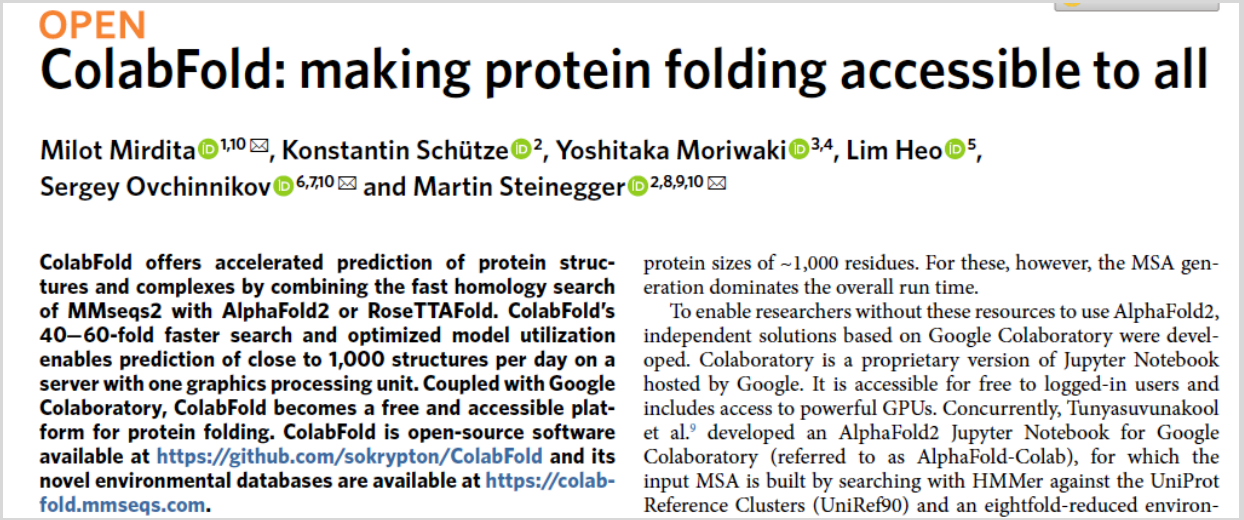
Alphafold2 database





Tunyasuvunakool et al. Nature, 2021

ColabFold



OPEN

ColabFold: making protein folding accessible to all

Milot Mirdita[1,10] ✉, Konstantin Schütze[2], Yoshitaka Moriwaki[3,4], Lim Heo[5], Sergey Ovchinnikov[6,7,10] ✉ and Martin Steinegger[2,8,9,10] ✉

ColabFold offers accelerated prediction of protein structures and complexes by combining the fast homology search of MMseqs2 with AlphaFold2 or RoseTTAFold. ColabFold's 40–60-fold faster search and optimized model utilization enables prediction of close to 1,000 structures per day on a server with one graphics processing unit. Coupled with Google Colaboratory, ColabFold becomes a free and accessible platform for protein folding. ColabFold is open-source software available at https://github.com/sokrypton/ColabFold and its novel environmental databases are available at https://colabfold.mmseqs.com.

protein sizes of ~1,000 residues. For these, however, the MSA generation dominates the overall run time.

To enable researchers without these resources to use AlphaFold2, independent solutions based on Google Colaboratory were developed. Colaboratory is a proprietary version of Jupyter Notebook hosted by Google. It is accessible for free to logged-in users and includes access to powerful GPUs. Concurrently, Tunyasuvunakool et al.[9] developed an AlphaFold2 Jupyter Notebook for Google Colaboratory (referred to as AlphaFold-Colab), for which the input MSA is built by searching with HMMer against the UniProt Reference Clusters (UniRef90) and an eightfold-reduced environ-

Note- community figured out how to study protein-protein interactions with AF2
https://github.com/sokrypton/ColabFold

Alphafold 3

- Has ligands and supports protein-protein interactions natively
- Uses “diffusion model” to generate structures (train on rotations of protein to account for invariances)
- Controversy – code not (yet) available
- https://alphafoldserver.com/

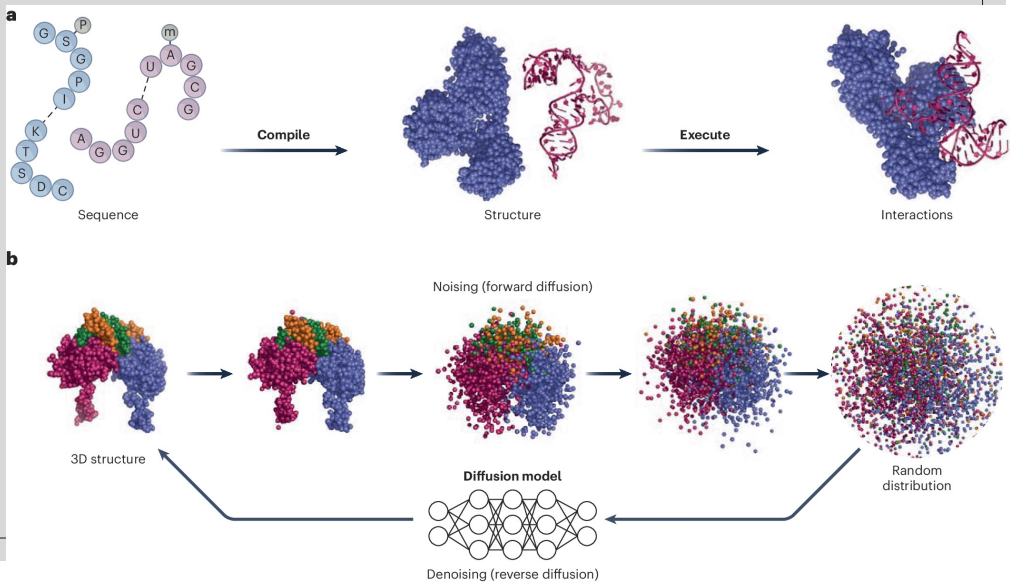
Article

Accurate structure prediction of biomolecular interactions with AlphaFold 3

https://doi.org/10.1038/s41586-024-07487-w
Received: 19 December 2023
Accepted: 29 April 2024
Published online: 8 May 2024
Open access
Check for updates

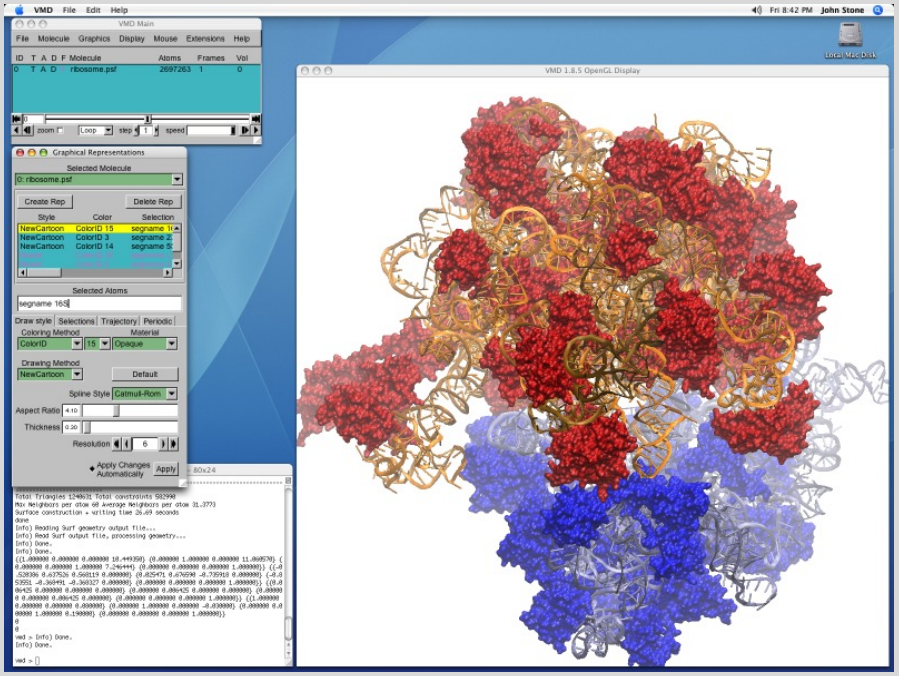
Josh Abramson[1,7], Jonas Adler[1,7], Jack Dunger[1,7], Richard Evans[1,7], Tim Green[1,7], Alexander Pritzel[1,7], Olaf Ronneberger[1,7], Lindsay Willmore[1,7], Andrew J. Ballard[1], Joshua Bambrick[2], Sebastian W. Bodenstein[1], David A. Evans[1], Chia-Chun Hung[2], Michael O’Neill[1], David Reiman[1], Kathryn Tunyasuvunakool[1], Zachary Wu[1], Akvilė Žemgulytė[1], Eirini Arvaniti[3], Charles Beattie[3], Ottavia Bertolli[3], Alex Bridgland[3], Alexey Cherepanov[4], Miles Congreve[4], Alexander I. Cowen-Rivers[3], Andrew Cowie[3], Michael Figurnov[3], Fabian B. Fuchs[3], Hannah Gladman[3], Rishub Jain[3], Yousuf A. Khan[3,5], Caroline M. R. Low[4], Kuba Perlin[3], Anna Potapenko[3], Pascal Savy[4], Sukhdeep Singh[3], Adrian Stecula[4], Ashok Thillaisundaram[3], Catherine Tong[4], Sergei Yakneen[4], Ellen D. Zhong[3,6], Michal Zielinski[3], Augustin Žídek[3], Victor Bapst[1,8], Pushmeet Kohli[1,8], Max Jaderberg[2,8]✉, Demis Hassabis[1,2,8]✉ & John M. Jumper[1,8]✉

The introduction of AlphaFold 2[1] has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein



Protein structure visualization

Many options with different features: **VMD,** PyMol, Chimera, ChimeraX, Mol*, …



Today

1. Compare results from collabfold with alphafold3
2. Learn to visualize and analyze structures using VMD

https://github.com/hockyg/comp-lab-class-2024/blob/main/Week7/Assignment.md