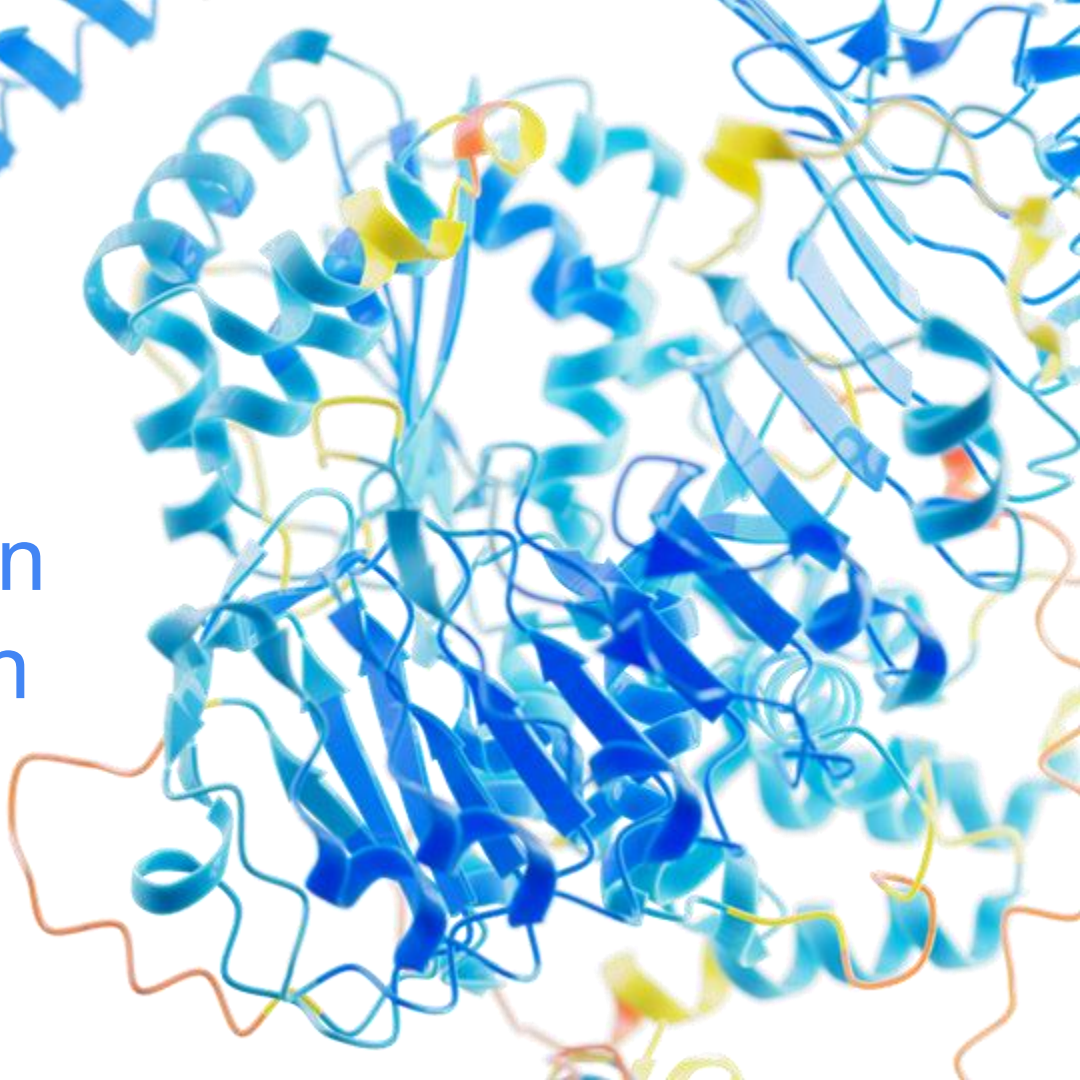


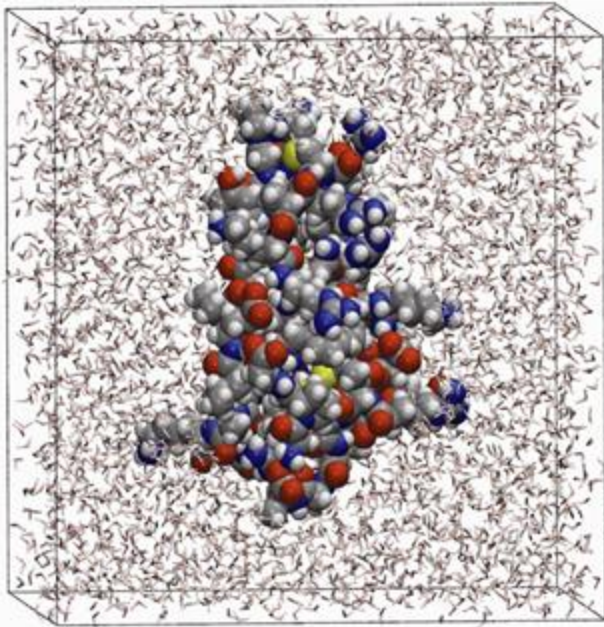
Building chemical and biological intuition into protein structure prediction

John Jumper

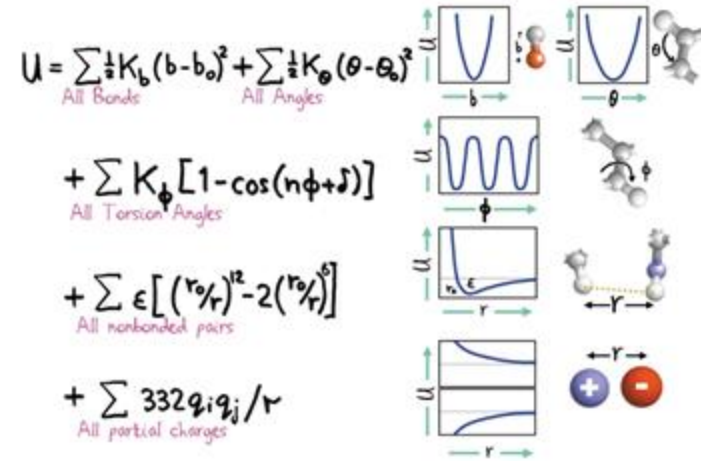
Google DeepMind



Proteins and simulation




Shaw et al, 2009



Levitt, M., 2001. The birth of computational structural biology. Nature Structural & Molecular Biology, 8(5), p.392.

Protein Data Bank

Biological Assembly 1



2SRC

CRYSTAL STRUCTURE OF HUMAN TYROSINE-PROTEIN KINASE C-SRC, IN COMPLEX WITH AMP-PNP

DOI: 10.2210/pdb2SRC/pdb

Classification: TYROSINE-PROTEIN KINASE
Organism(s): Homo sapiens
Expression System: Spodoptera frugiperda
Mutation(s): Yes

Deposited: 1998-12-29 Released: 1999-07-22
Deposition Author(s): Xu, W., Oishi, A., Lee, M., Eck, M.J., Harrison, S.C.

Experimental Data Snapshot

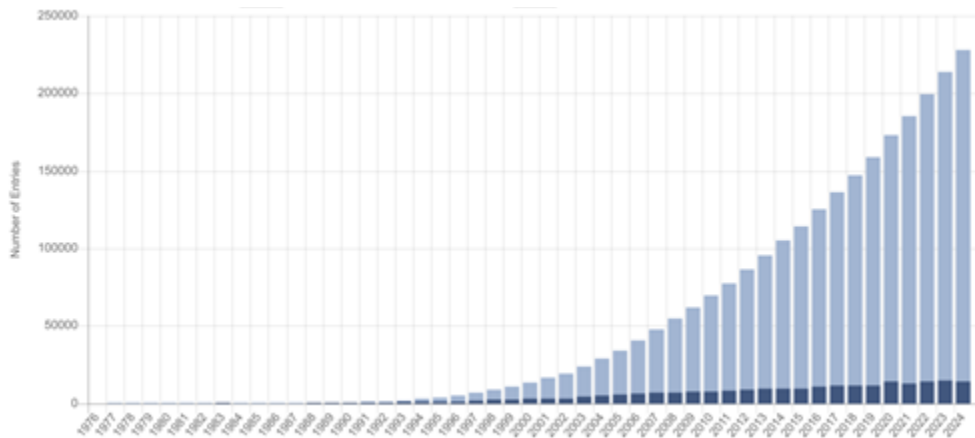
Method: X-RAY DIFFRACTION
Resolution: 1.50 Å
R-Value Free: 0.281
R-Value Work: 0.226
R-Value Observed: 0.226

wwPDB Validation

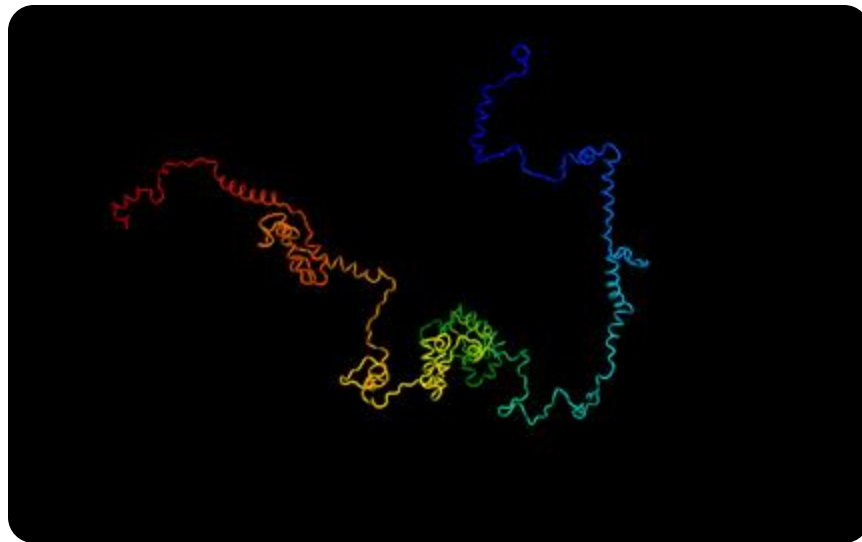
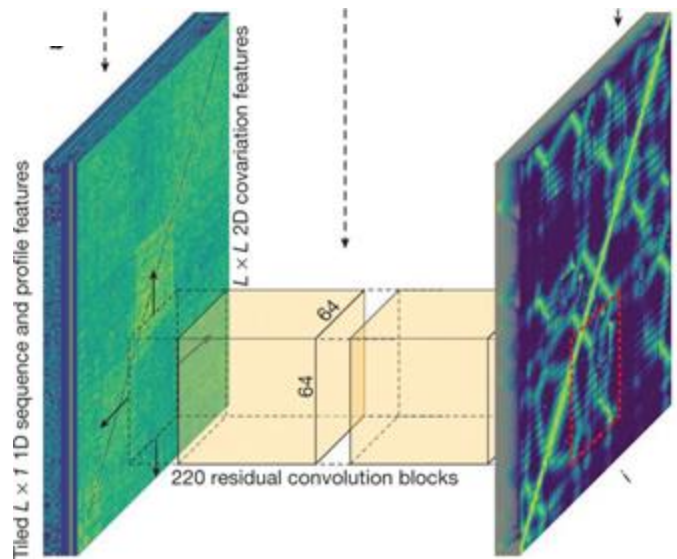
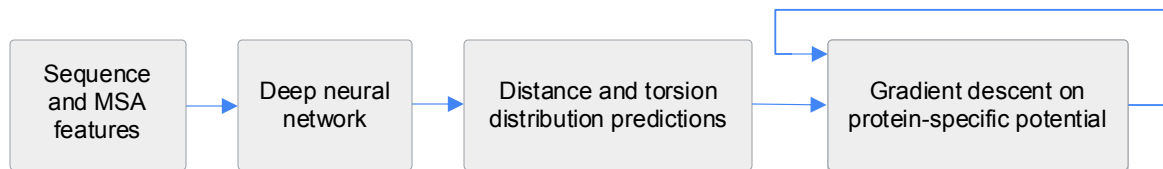
Metric	Percentile Ranks	Value
Cis-peptide		18
Ramachandran outliers		7.2%
Sekelahn outliers		12.4%

3D View: Structure | Ligand Interaction

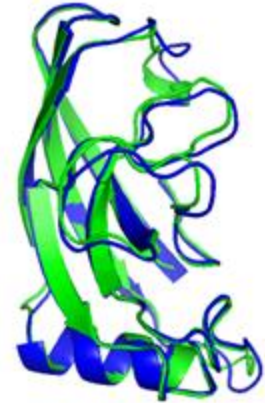
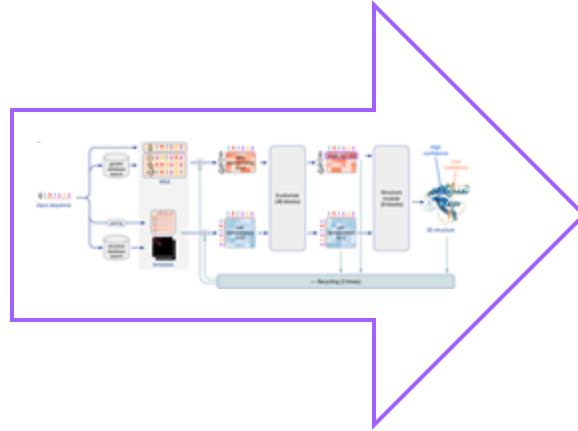
Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A1



AlphaFold 1



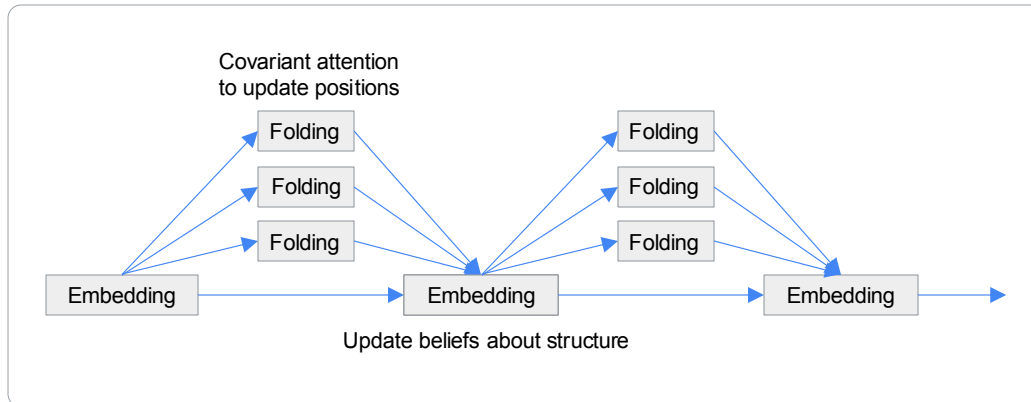
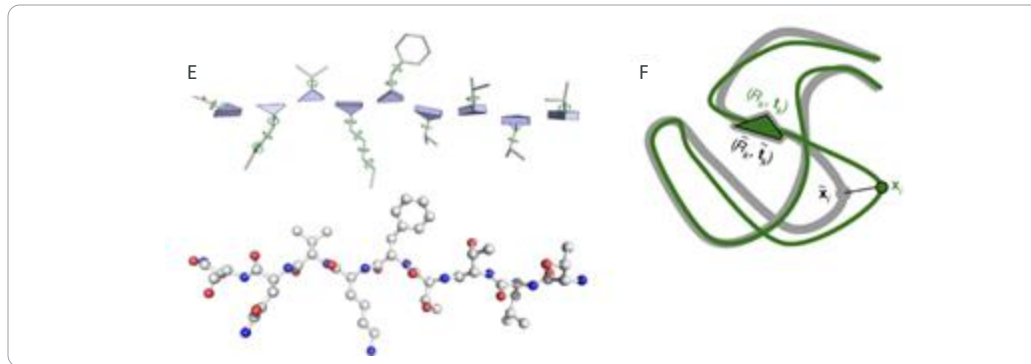
SIFSYITESTGTPSNATYT
YVIERWDPETSGILNPCYG
WPVCYVTVNHKHTVNGTGG
NPAFQIARIEKLRTLAEVR
DVVLKNRSFPIEGQT THRG
PSLNSNQECVGLFYQPNS
GISPRGKLLPGSLCGIAPP
PVHHHHHH



Working without evolutionary information

After AlphaFold1, we decided to spend time working on structure prediction without evolutionary information

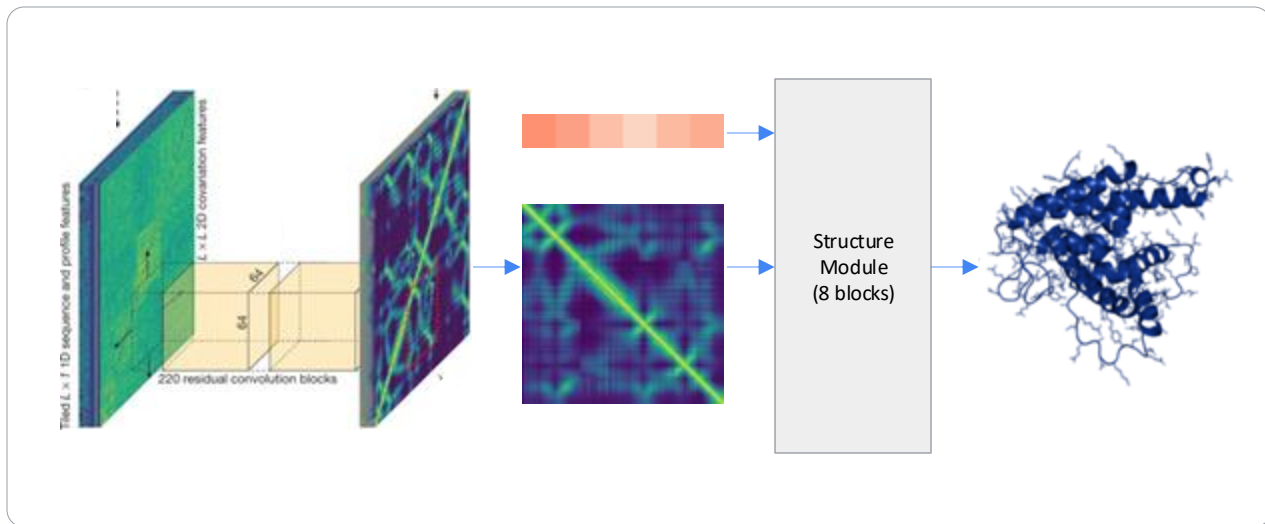
- Goal was to figure out how to build physics into the network
- Also forced us to throw away our highly tuned systems



Fusing direct structure prediction and evolution

AlphaFold1 distograms into the structure module (~October 2018)

Was around 2 GDT worse than published AlphaFold1 system, but strongly suggested end-to-end structure prediction would work

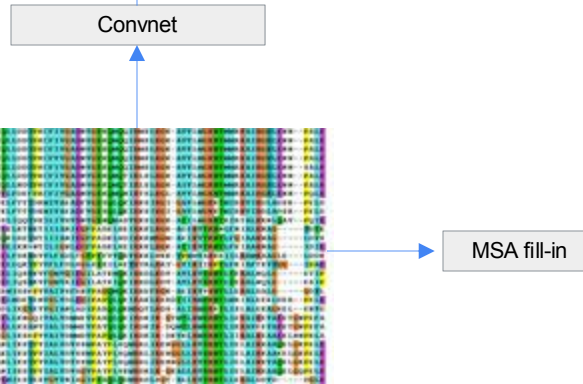
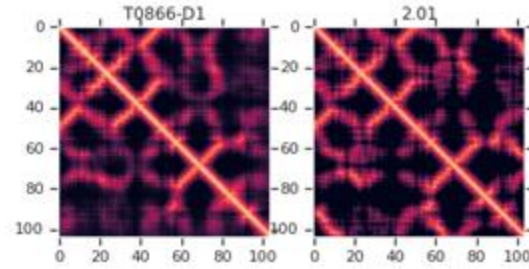


Starting over on using evolutionary data

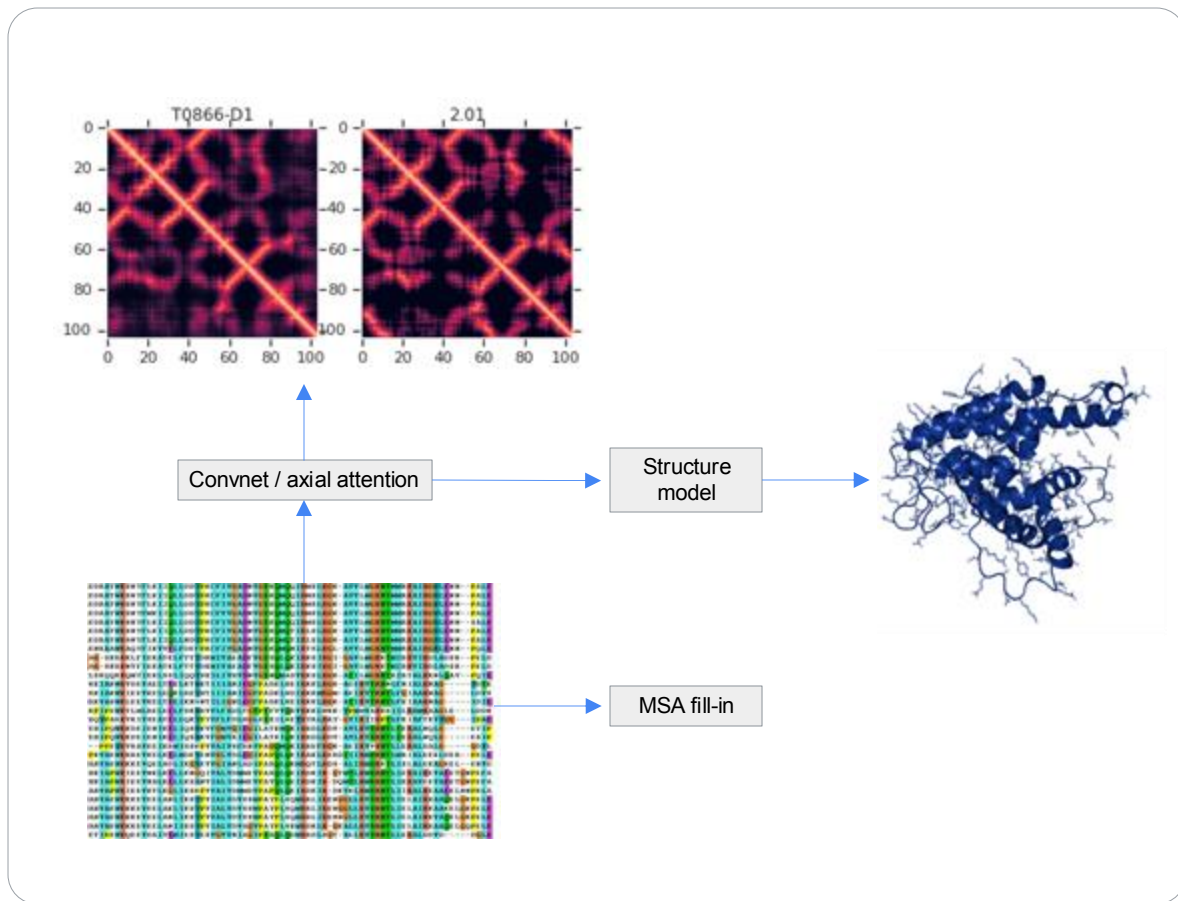
Taught networks to “fill-in” evolutionary data as a generalization of co-evolution (rawBERT models)

Used axial attention as a meaningful alternation between processes per-sequence and processing across sequences

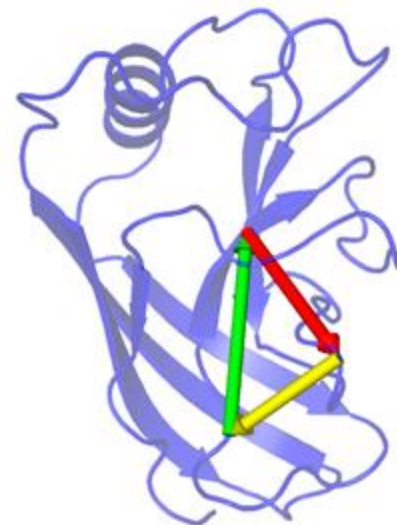
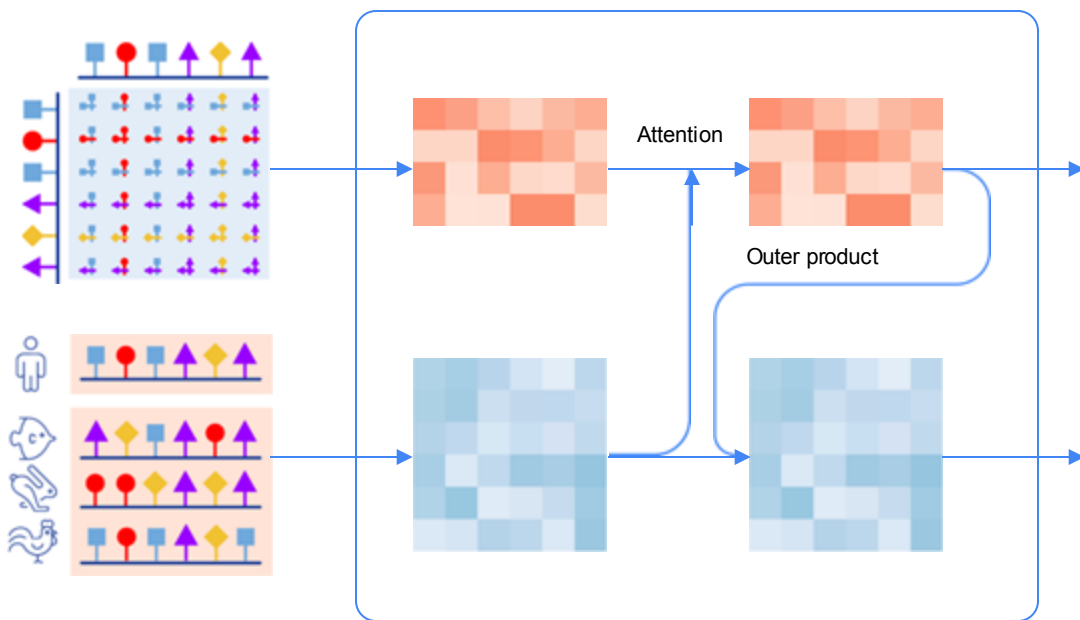
Almost immediately better than our earlier work because the neural network operations were meaningfully aligned to the scientific problem



End-to-end structure prediction

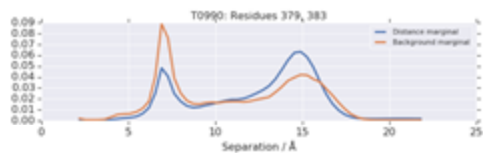
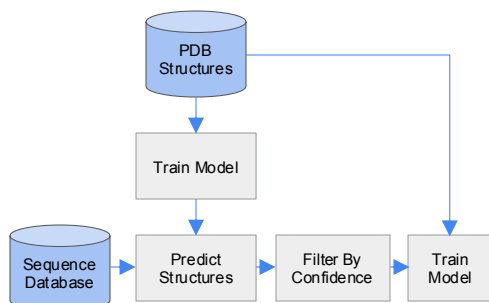


Building physics and biology into our network blocks

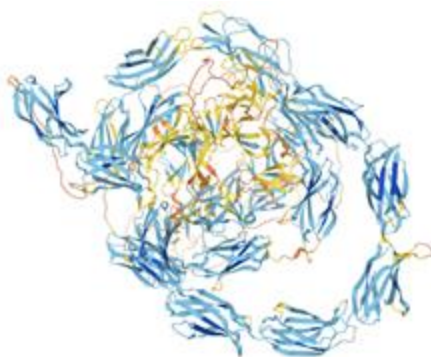


Confidence measures in three acts

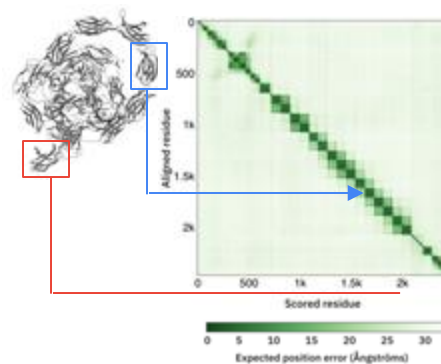
Self distillation



pLDDT



PAE

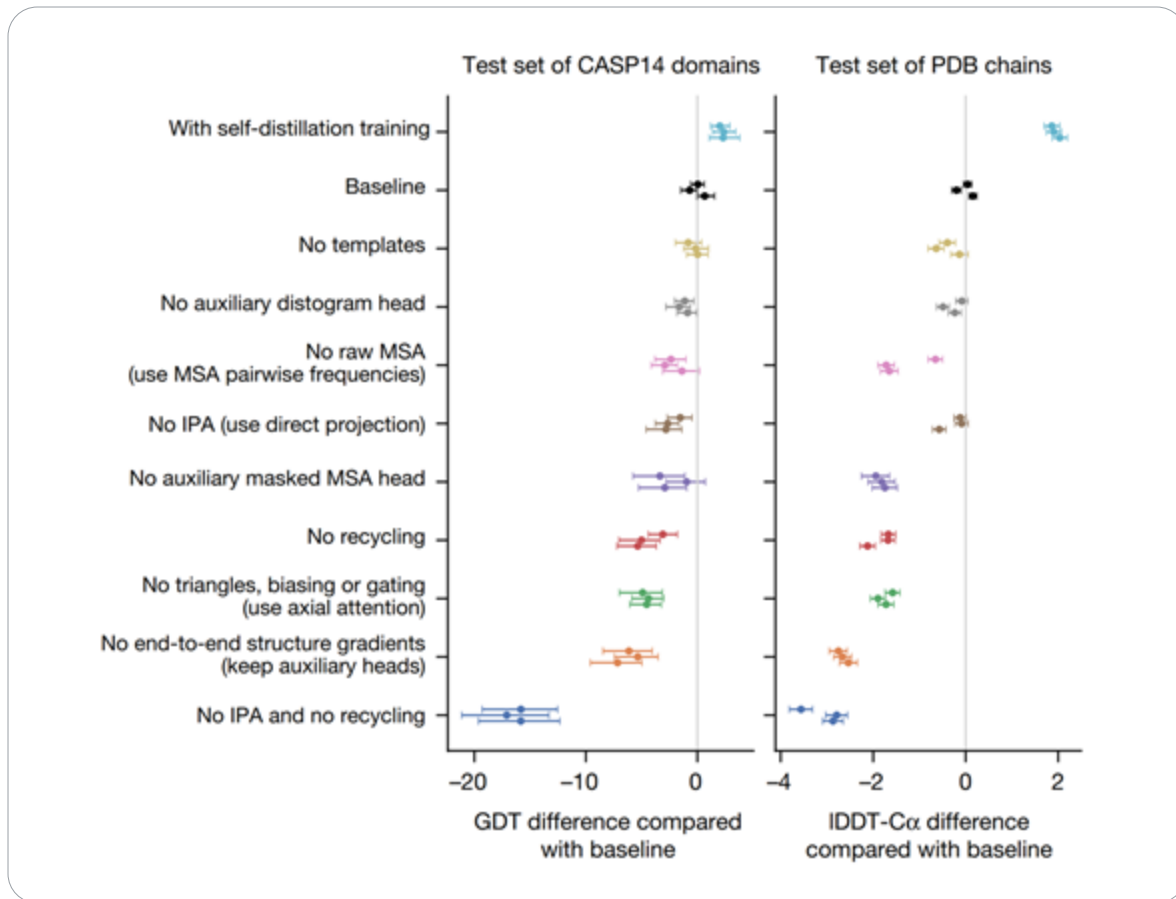


Which parts mattered? All of it

Don't have time to describe the many other ideas that contributed:

- Data engineering
- Recycling
- Relative positional coding
- Extra MSA evoformer
- Weights initialization
- Metagenomic data
- Template embedding
- Gated updates
- Violation losses

and many more!

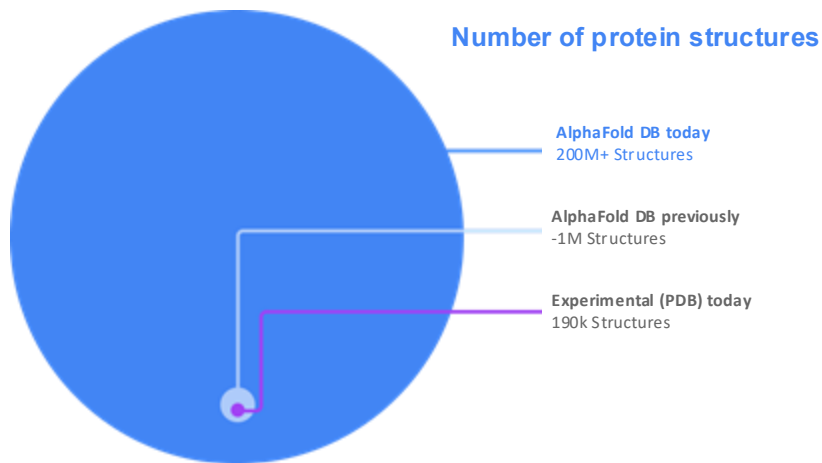


AlphaFold Protein Structure Database

In partnership with EMBL-EBI

Predictions across nearly all of UniProt protein sequences over 200M predictions

Closes the gap between sequencing and structure determination



AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

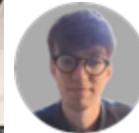
Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [Arl13b02](#) [COX5B](#) [E. coli](#) [Help](#) [AlphaFold DB search help](#)

AlphaFold 2 team



Acknowledgements

AlphaFold 1

Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu & Demis Hassabis

AlphaFold 2

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

Human Proteome

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

The wider team at
DeepMind



Our AFDB collaborators at EMBL-
EBI



The CASP
community



PDB & the experimental
biology community

