A Brief Introduction to AlphaFold

Vamsi Varanasi
Tuva AI and Courant Institute, NYU

AI Reasoning in Theoretical Physics Workshop
Aspen Center for Physics

June 10, 2026

AlphaFold is one of the great triumphs of specialist AI models in science, and its development can teach us many lessons about how AI can be used to great success in other fields. Here, I provide a brief overview of the protein folding problem, the state of the field before AlphaFold, and a bare-bones sketch of the model before discussing limitations/opportunities of AlphaFold in its own domain and takeaways that may transfer to other physical problems.

1 The problem

Proteins are the primary "machines" of biology: enzymes catalyze nearly every reaction in the cell, hemoglobin ferries oxygen, antibodies recognize pathogens, motor proteins like myosin and kinesin generate force, and collagen and keratin give tissues their structure. A protein is a chain of $\mathcal{O}(100)$ amino acid molecules drawn from an alphabet of 20 "residues," which, with few exceptions, folds into a functional form ("structure") through electrostatic and chemical interactions between residues.

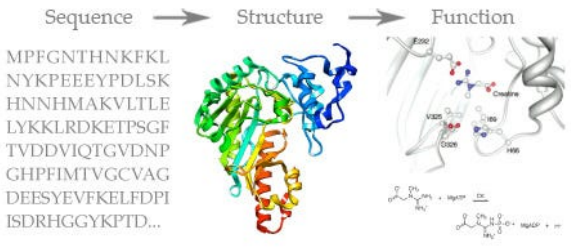


Image credit: Apoorva Srinivasan, Cedars-Sinai.

To first order, structure determines function: an enzyme's active site is a pocket shaped to grip one specific substrate, and a receptor's cleft is shaped to its ligand. However, biological molecules are subject to mutational drift, to the extent that two proteins sharing similar same folds and function ("homologs") can have as little as 20% of their residues in common [1]; across evolution, structure is conserved roughly 3 to 10 times more tightly than sequence [2]. The protein folding problem is thus twofold:

- given a sequence, can we predict the structure? ("folding")
- given a structure, can we write down a sequence that folds into it? ("inverse folding" / "de novo protein design")

The 2024 Nobel in Chemistry went to Jumper and Hassabis of Google DeepMind for the former (AlphaFold2 [3]) and Baker for the latter. Here, we focus on AlphaFold and the forward folding problem. Inverse folding, which fills in amino acids with desired properties onto a given structural backbone, uses entirely different machine learning techniques (e.g. RFdiffusion [4] for backbone construction, ProteinMPNN [5] for sequence selection).

Anfinsen won the 1972 Nobel Prize in Chemistry for showing that the forward protein folding problem is well conditioned in that the fold is primarily determined by sequence alone. He denatured ribonuclease, unfolding the chain and breaking its disulfide bonds, then removed the denaturant and watched it spontaneously refold to full catalytic activity, with no external machinery or template [6], conjecturing that the folded state sits at a (not always unique) free-energy minimum [7]. But with 100 amino acids and coarsely $\mathcal{O}(10)$ configurations available per residue (backbone angles), the search space is enormous, $\sim 10^{100}$.

Random search would take longer than the age of the universe, yet biology folds proteins to a reliable, reproducible structure in microseconds ("Levinthal's Paradox" [8]). Combined, these lines of reasoning indicate that the physics of folding follows rules set by interactions between residues, guiding the chain down a funnel into the low-energy state [9]. The game is to learn this underlying physics.

2 Protein folding before AlphaFold

The protein folding problem maps neatly to deep learning: fixed inputs, combinatorial logic, differentiable outputs (coordinates of residues). DeepMind further chose to tackle protein folding because of an uncommon confluence of community factors: availability of open, clean community data, decades of structural and evolutionary insight, and an objective benchmark to judge success. Below, I'll detail each, and provide a brief overview of pre-AlphaFold models.

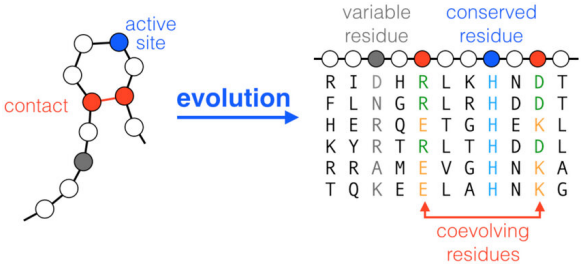
2.1 The data: PDB and MSAs

Experimentally measuring protein structure is painstaking (historically by X-ray crystallography and NMR, more recently made more efficient by cryo-EM). Indeed, at least half a dozen Chemistry Nobels have been awarded for determining the structures of individual proteins and molecular machines.[1] Since 1971, biologists have aggregated $\mathcal{O}(10^5)$ structures into the Protein Data Bank (PDB) [10], forming the core training set for AlphaFold and any other structure prediction models. While impressive, this number alone is not enough to train a general model. Furthermore, the PDB rarely has substantial coverage of structures for homologous sequences (different sequences with the same folds), exactly the problem we're trying to solve.

Multiple sequence alignments (MSAs) address both problems by linking sequences where we understand the structure to sequences where we don't, leveraging the statistical properties of evolution. Sequence data is far easier to obtain than structure, with $\mathcal{O}(10^8)$ sequences in reference databases like UniRef [11] and $\mathcal{O}(10^9)$ in larger metagenomic sets. Given a query sequence, one can build a profile hidden Markov model and scan the reference database with it, pulling in and aligning by residue every sequence the model scores as a likely homolog (one related to the query by a chain of substitutions and insertions or deletions). Iterating the search broadens the profile to reach more divergent relatives. This lets us extrapolate from the measured structures to a much larger dataset of evolutionarily related sequences—in other words, we encode the sequences that we think we can understand in terms of the sequence we know.

2.2 Evolutionary insights

Recall that different homologs perform the same function, but differ dramatically in sequence due to mutational divergence. However, only some mutations can be tolerated; mutate a structurally or chemically important residue and the protein is no longer functional, and the organism would be selected against. Biophysicists had discovered two important classes of conserved elements between homologs. The first order conserved residues, which are often required for the chemical activity of an enzyme, are the same across all sequences in the MSA. The second order correlations are for conserved *contacts*, interactions between residues that are structurally important. A mutation to one residue in a contact must be accompanied by a corresponding mutation to the



Coevolved and conserved residues can be inferred from an MSA [12].

[1] 1962 (hemoglobin and myoglobin, the first ever protein structures), 1988 (photosynthetic reaction center), 2003 (ion channels and aquaporins), 2006 (RNA polymerase), 2009 (the ribosome), and 2012 (G-protein-coupled receptors).

other residue in such a way that preserves the physical and chemical interaction of the contact. These evolutionary traces imprint the MSAs with inferrable correlations, from which we can back out structural information.

In physics language, if we take a sequence σ such that $\sigma_i \sim \mathcal{A}$ is the residue at i drawn from our alphabet $\mathcal{A}$, then we find the maximum entropy guess is

$$P(\sigma) \propto \exp\Big(\sum_i h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i, \sigma_j)\Big),$$

a Potts model with fields h_i (the log frequency of each residue at site i, in the independent-site limit) and couplings J_{ij}. Inferring these couplings from the correlations in the MSA, while disentangling direct couplings from indirect (transitive) ones, and reading contacts off the largest J_{ij}, is the method of "direct coupling analysis" (DCA) [13–15]. In going from sequence to MSA to contacts, DCA is a precursor to the methods of AlphaFold, which improves upon this method by inferring distances and orientations from MSAs instead of just contacts (e.g., geometry instead of topology).

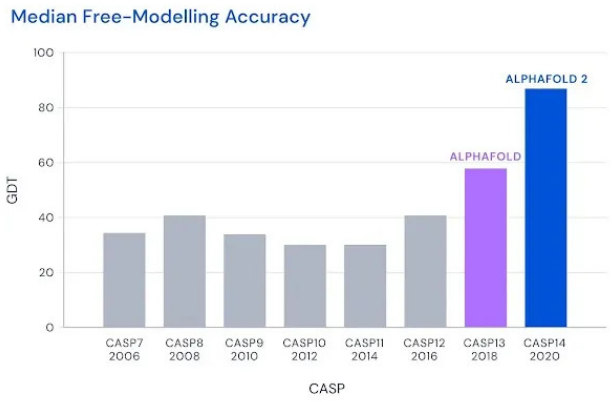
3 The benchmark: CASP and prior efforts

Held every two years since 1994, the open Critical Assessment of Structure Prediction (CASP) competition is the field's yardstick. Predictors are handed sequences whose structures have been measured but not released. Submissions are made blindly and are scored once structures are revealed. Given the difficulty of measuring protein structures, the importance of this competition infrastructure already existing should not be understated.

Without going into too much detail, the approaches over the two decades preceding AlphaFold fell into a few camps:

- DCA-based inference of contacts from correlated residues in an MSA, as above [13–15]
- Template matching: deform the structure of a known homolog [16], or assemble the fold from fragments of known structures [17]
- Direct simulation of whole-protein folding via molecular dynamics [18]

As shown, these approaches were insufficient. However, CASP12 saw the emergence of RaptorX [19], a convolutional neural network-based approach that took in DCA couplings and predicted contact maps—the bridge between DCA and AlphaFold.



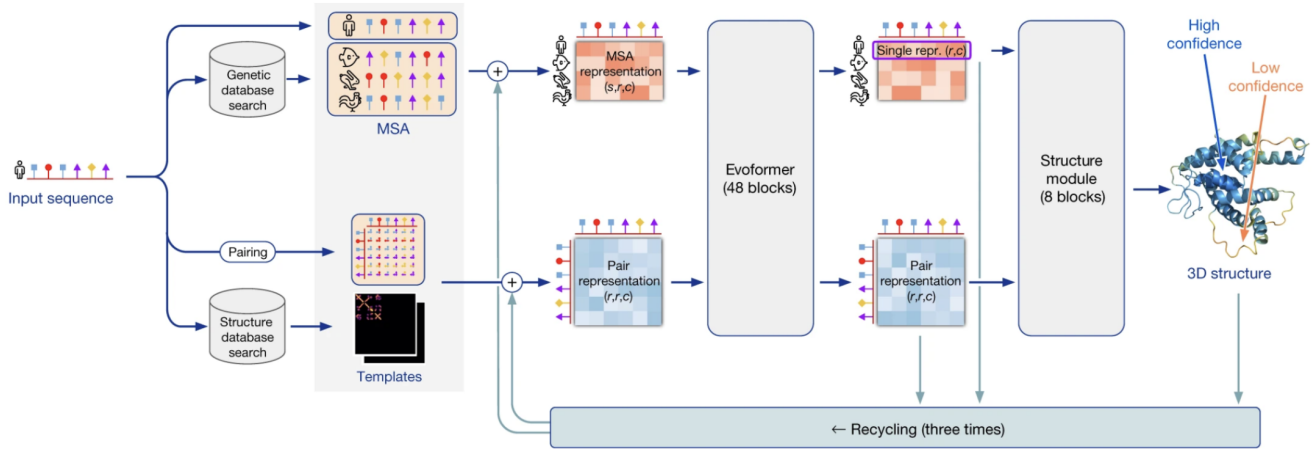
CASP before and after AlphaFold. Image credit: Siddhant Rai.

4 AlphaFold

AlphaFold completed the arc of using deep learning to predict protein folding—Euclidean distances between residues at the atomic scale—from sequences via MSAs. The first AlphaFold (CASP13) predicted a distribution over inter-residue distances from the MSA with a deep convolutional network, then folded the chain by gradient descent to satisfy those distances, a two-stage pipeline [20]. AlphaFold 2 replaced this with a single end-to-end model: it reasons jointly over the MSA and a pairwise representation of the

structure using attention, then reads out atomic coordinates directly, removing the separate folding step and jumping to near-experimental accuracy [3], winning CASP14.

Roughly, AlphaFold generates an MSA from a sequence, learns correlations in residues from correlations in homologous sequences, and translates them into distances leveraging known structural data. The AlphaFold architecture is fascinating and well documented elsewhere, for example in these slides from Amy Lu[2]. For a detailed and intuitive explanation of the (latest) AlphaFold 3 architecture, I recommend Elana Simon and Jake Silberg's *The Illustrated AlphaFold*, an outstanding visual walkthrough of the model and its design choices.[3]. For our purposes, I'll focus on the salient aspects of AlphaFold 2's architecture that enabled it to win CASP 2020.



The architecture of AlphaFold 2. [3]

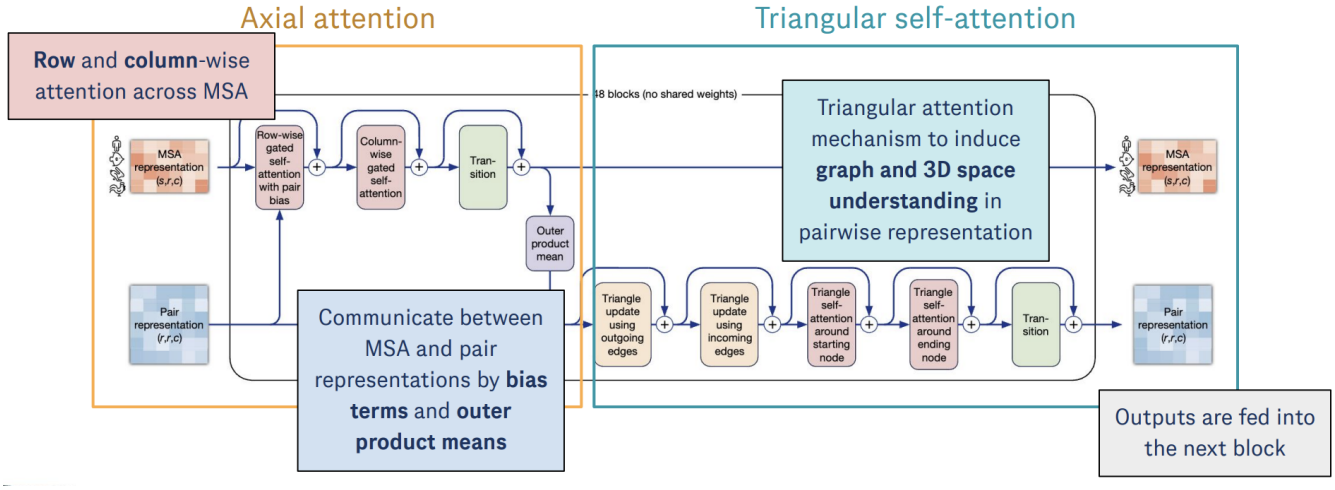
Upon receiving an input sequence, AlphaFold 2 generates an MSA and searches for any structures of any homologs. From these, it creates and maintains two representations of each sequence: an MSA representation, which learns evolutionary information, and a pair representation, which learns how residues interact with each other. Each of these representations can be thought of as a vector of $\mathcal{O}(100)$ entries assigned to each (homolog, residue) entry in the MSA case and (residue i, residue j) in the pair case. These representations are refined by two main modules: the Evoformer, the core attention layer of the model, and the structure module, which converts the pair representation into an actual ball-and-stick geometry. Outputs from each prediction run are recycled back to improve performance. Loss is a complicated function dominated by a frame-aligned metric based on local deviations between truth and prediction.

The Evoformer is a bespoke, attention-based layer that both learns the evolutionary correlations within the MSA and translates them into structural pairing information. It is broken up into two sectors: the axial attention sector and the triangular self-attention sector. The axial attention sector learns the evolutionary correlations of the MSA, in the mold of DCA but accounting for higher order interactions. Row-wise attention learns the intra-sequence relationships between residues; column-wise attention learns the evolutionary relationships of each residue between homologs. Critically, the attention mechanism here is looking for relationships between residues, which is exactly the combinatorial problem we're trying to solve—an "inductive bias", that is, the model is built to make predictions about relationships between residues without being told *a priori* that this is what it should look for. By aligning the inductive biases of the model with the dynamics of the problem we want to solve, we hope to bake in some physics into the model without being overly prescriptive.

The correlations in the MSA representation learned by the axial representation layers then get transduced into the pair representation by the triangular self-attention sector, the rest of the Evoformer block, which learns the 3D representation of the protein in terms of relationships between residues. This, too, follows the inductive bias of the model. The pair representation can be thought of as a graph, with the ij^{th} entry

[2] https://amyx.lu/data/alphafold.pdf
[3] https://elanapearl.github.io/blog/2024/the-illustrated-alphafold/

representing the edge between residues i and j. To update the edge (i, j), triangular self-attention attends over every third residue k and folds in all three sides of the triangle (i, j, k): the edge (i, j) supplies the query, (i, k) the key and value, and the closing edge (j, k) an additive bias on the attention score. Stacking this over all triples lets information propagate around the graph through closed triangles. The motivation is geometric: if z_{ij} encodes the relationship between two residues, loosely a distance, then the three sides of a triangle cannot be chosen independently. The network does not impose a literal triangle inequality on these vectors; it only forces each update to respect triangle topology, nudging the pair representation toward a geometry that can actually be embedded in three dimensions. That consistency is learned rather than enforced, since the pair representation must ultimately decode into real coordinates under the structure module.



The Evoformer [3]. Callouts added by Amy Lu.

The structure module represents each residue as a rigid local frame and learns rotations and translations that convert the Evoformer's single and pair representations into 3D coordinates. The Evoformer does most of the global reasoning over evolutionary and pairwise geometry, while the structure module turns that learned geometry into an equivariant all-atom structure. AlphaFold 3 [21] substantially redesigns this coordinate-generation stage around a diffusion-based architecture that reasons at the atom level across proteins, nucleic acids, ligands, ions, and modifications. I'll defer to the aforementioned architecture resources for a more detailed description of the structure module.

At inference time, the model reports two self-assessed confidence metrics, each trained as a minor auxiliary term in the loss: the predicted local distance difference test (pLDDT), a per-residue estimate of how accurate the local distances around that residue are, and the predicted aligned error (PAE), the expected error in residue j's position when truth and prediction are aligned on residue i. These are a one-body and a two-body quantity respectively, echoing the first- and second-order structure of the Potts model: the single-site fields h_i (conservation) and the pairwise couplings J_{ij} (coevolution) from earlier.

5 Limitations and horizons

AlphaFold 3 moved from protein structure prediction to biomolecular complex prediction, modeling proteins, RNA, DNA, ligands, and their interactions. These higher-order structural problems are far from solved. Major areas of opportunity include:

- Proteins with limited evolutionary information, including orphan, de novo, and synthetic proteins.
- Conformational dynamics, structural ensembles, allostery, and intrinsically disordered regions.
- Large multimolecular assemblies, including viral capsids, ribosomes, and transient cellular complexes.
- Cotranslational folding, assembly pathways, and other kinetic effects.

- Protein–RNA, RNA–RNA, and other nucleic-acid-mediated interactions, where rugged energy landscapes and alternative secondary structures complicate prediction.

Several active research directions aim to address these limitations. Competitive models such as Chai-1 and Boltz seek to extend and democratize biomolecular structure prediction, while protein language models such as ESMFold and ESM3 reduce dependence on deep multiple sequence alignments and increasingly integrate structure generation with sequence modeling. More broadly, large biological foundation models trained directly on evolutionary sequence data may provide a path toward predicting not only static structures, but also dynamics, function, and design.

6 Takeaways

Why was AlphaFold so successful? Firstly, the ingredients for a successful machine learning model were readily available: enough data, a differentiable loss function, and an objective benchmark that itself requires substantial coordination infrastructure. Much of the physics of folding, especially the interplay between structure and evolutionary trajectory, had been understood through substantial effort by the biology community. This understanding both informed model development and data expansion. The combinatorics was roughly what was left to generalize.

It just so happened that the inductive biases of the transformer, then a new, substantially more powerful type of model, matched the physics of protein folding—the attention mechanism assigns weights to each residue interacting with every other residue and thus learns the ways they may interact. This is interesting for several reasons. First, AlphaFold is a nuanced answer to the "bitter lesson"—encoding some physics into your model is helpful at a high level of abstraction, but then you want to go from data to output and trust backpropagation instead of overly curating your model. This is especially true when the amount of data we have is on the order of the protein folding problem—enough that the right model can make good predictions, but not so much that any sufficiently powerful model can be trained to make good predictions. Second, many hard combinatorial problems in physics are the result of many bodies interacting. Thus, bespoke attention mechanisms like the Evoformer might find wide utility, especially when we are in this kind of data regime.

AlphaFold did not win by discarding physics, nor by hand-coding all of it. It put the right abstractions into the architecture: MSAs for evolutionary statistics, pair representations for residue-residue structure, triangle updates for geometric consistency, recycling for iterative refinement, and an equivariant structure module for coordinates. Then it let backpropagation learn the details. This is the useful lesson for physics: when data are large but not internet-scale, the right inductive bias may be the difference between a model that merely fits and one that generalizes.

References

[1] Cyrus Chothia and Arthur M. Lesk. The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*, 5(4):823–826, 1986. doi: 10.1002/j.1460-2075.1986.tb04288.x.

[2] Kristoffer Illergård, David H. Ardell, and Arne Elofsson. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics*, 77(3):499–508, 2009. doi: 10.1002/prot.22458.

[3] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

[4] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023. doi: 10.1038/s41586-023-06415-8.

[5] Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J. Ragotte, et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022. doi: 10.1126/science.add2187.

[6] Christian B. Anfinsen, Edgar Haber, Michael Sela, and Frederick H. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences*, 47(9):1309–1314, 1961. doi: 10.1073/pnas.47.9.1309.

[7] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973. doi: 10.1126/science.181.4096.223.

[8] Cyrus Levinthal. How to fold graciously. In Peter Debrunner, John C. M. Tsibris, and Eckard Münck, editors, *Mössbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois*, pages 22–24. University of Illinois Press, Urbana, IL, 1969.

[9] Ken A. Dill and Hue Sun Chan. From levinthal to pathways to funnels. *Nature Structural Biology*, 4 (1):10–19, 1997. doi: 10.1038/nsb0197-10.

[10] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235.

[11] Baris E. Suzek, Hongzhan Huang, Peter McGarvey, Raja Mazumder, and Cathy H. Wu. UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23(10):1282–1288, 2007. doi: 10.1093/bioinformatics/btm098.

[12] Simona Cocco, Christoph Feinauer, Matteo Figliuzzi, Rémi Monasson, and Martin Weigt. Inverse statistical physics of protein sequences: A key issues review. *Reports on Progress in Physics*, 81(3): 032601, 2018. doi: 10.1088/1361-6633/aa9965.

[13] Martin Weigt, Robert A. White, Hendrik Szurmant, James A. Hoch, and Terence Hwa. Identification of direct residue contacts in protein-protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1):67–72, 2009. doi: 10.1073/pnas.0805923106.

[14] Faruck Morcos, Andrea Pagnani, Bryan Lunt, Arianna Bertolino, Debora S. Marks, Chris Sander, Riccardo Zecchina, José N. Onuchic, Terence Hwa, and Martin Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49):E1293–E1301, 2011. doi: 10.1073/pnas.1111471108.

[15] Debora S. Marks, Lucy J. Colwell, Robert Sheridan, Thomas A. Hopf, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. Protein 3d structure computed from evolutionary sequence variation. *PLoS ONE*, 6(12):e28766, 2011. doi: 10.1371/journal.pone.0028766.

[16] Andrej Šali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993. doi: 10.1006/jmbi.1993.1626.

[17] Kim T. Simons, Charles Kooperberg, Enoch Huang, and David Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *Journal of Molecular Biology*, 268(1):209–225, 1997. doi: 10.1006/jmbi.1997.0959.

[18] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011. doi: 10.1126/science.1208351.

[19] Sheng Wang, Siqi Sun, Zhen Li, Renyu Zhang, and Jinbo Xu. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Computational Biology*, 13(1):e1005324, 2017. doi: 10.1371/journal.pcbi.1005324.

[20] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020. doi: 10.1038/s41586-019-1923-7.

[21] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024. doi: 10.1038/s41586-024-07487-w.