

A Breakthrough in Protein Structure Prediction: AlphaFold

Emily Wu '28

Proteins are essential to life. Made from long chains of amino acids that fold into 3D structures, tens of thousands of different protein structures exist inside our bodies. Each structure is unique to the protein's function, as small changes in a protein's structure can drastically affect how it works. To name a few, enzymes, antibodies, and collagen are all useful proteins in our bodies that have distinctly different structures. However, in order to properly understand their functions, it is critical to fully and accurately capture their structures.

For 50 years, determining the structure of proteins in the human genome has been an open research problem. Experimental methods like x-ray crystallography, the shining of x-rays through a protein crystal, or cryo-electron microscopy, the freezing of proteins then placed under an electron microscope, have produced some accurate 3D structures of proteins, but they often come with physical limitations and are not versatile enough to be used for billions of structures.

In recent years, however, the development of a deep learning system, called AlphaFold, has been able to bring many breakthroughs towards identifying protein structures. For one, AlphaFold accurately predicts protein structures solely based on the amino acid sequence of a protein. It was created by Google DeepMind and was trained on publicly available data. Some of the most notable open access resources include the Protein Data Bank, Uniprot, and MGnify, which offer numerous amounts of reliable data towards protein sequences and structures.

These data sets have enabled AlphaFold to accurately predict the backbone structures of proteins, which are the main chain of amino acids that determine the protein's overall structure. Once the backbone was predicted correctly, it could also model the side chains with precision. Side chains, which are responsible for a protein's chemical activity and interactions, were

especially hard to predict with previous methods because they are small and chemically diverse. Impressively, AlphaFold is also capable of predicting entirely new proteins that have never been studied before, making it a powerful tool for scientific exploration in the human genome.

Furthermore, AlphaFold was entered into the Critical Assessment of Structure Prediction (CASP14) competition in 2020, which assessed the accuracy of computational methods for predicting protein structures. Not only did AlphaFold win the competition, but it also outperformed its competitors by a wide margin. In the competition, all methods were measured in terms of Å, or Angstroms, for accuracy. The lower the number, the lower the margin of error. For AlphaFold, its predictions achieved an accuracy of 0.96 Å, profoundly better than the two next best-performing methods: 2.8 Å for backbone RMSD and 3.5 Å for all-atom RMSD. AlphaFold's stellar accuracy illustrated a huge scientific achievement—a more accessible, computational method of protein structure prediction.

Additionally, AlphaFold offers a predicted Local Distance Difference Test (pLDDT) score, a confidence report on how trustworthy AlphaFold believes its predictions are on the protein structure. The score ranges from 0-100, with a higher pLDDT score indicating a more reliable prediction than a lower one. The pLDDT score can vary across the entire protein structure, as some areas may be more reliable than others. This offers transparency for scientists who are studying protein structures, and it also guides their experimental methods towards unreliable areas which are in need of research, saving a lot of time and resources.

The AlphaFold Protein Structure Database is an open access resource for anyone interested in protein structure. Offering millions of unique structures, the database offers a 3D image of each protein as well as the pLDDT score of each region. Scientists and students are able

to have a much more rapid, accessible, and accurate version of structural data, compared to performing experimental methods that both offer a degree of unreliability as well as inefficiency.

AlphaFold represents a revolutionary step towards the scientific research of protein structures, overcoming the many limitations of traditional, experimental methods and quickening the progress of scientific exploration. Even beyond the study of the human genome, new discoveries are bound to be made, including discoveries in drugs and disease. This offers hope for many to explore the mysteries of biology, developing further breakthroughs in the unknown.

References

Cleveland Clinic. (n.d.). What are proteins? Definition, types & examples.

<https://my.clevelandclinic.org/health/body/proteins>

EMBL. (2021, July 15). AlphaFold: Using open data and AI to discover the 3D protein universe.

<https://www.embl.org/news/science/alphafold-using-open-data-and-ai-to-discover-the-3d-protein-universe/>

European Bioinformatics Institute. (n.d.). pLDDT: Understanding local confidence.

<https://www.ebi.ac.uk/training/online/courses/alphafold/inputs-and-outputs/evaluating-alphafolds-predicted-structures-using-confidence-scores/plddt-understanding-local-confidence/>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool,

K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A.,

Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ...

Hassabis, D. (2021, July 15). *Highly accurate protein structure prediction with alphafold.*

Nature News. <https://www.nature.com/articles/s41586-021-03819-2>

News-Medical.net. (n.d.). Protein structure determination.

<https://www.news-medical.net/life-sciences/Protein-Structure-Determination.aspx>

Toews, R. (2021, October 3). AlphaFold is the most important achievement in AI ever. Forbes.

<https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever/>