Jack Michaels

CS279

Comparison of Modern Deep-Learning Protein Structure Prediction Software

Introduction

This paper compares the accuracy between three modern deep learning protein structure prediction softwares: AlphaFold 2[3], RoseTTAFold[4], and ESMFold[5]. To do so, we consider three evaluation metrics: minimum Root-Mean-Square Deviation (RMSD), average RMSD, and both maximum and average template modeling score (TM-score[7]) to understand each model's general performance and best-case performance. Analysis is conducted on ten statistically distinct protein sequences to improve data distribution and thus better inform our evaluation metrics. The statistically distinct proteins chosen from UniProt[6] include: human hemoglobin (P68871), chicken lysozyme (P00698), human ferritin (P02792), ubiquitin (P62068), green fluorescent protein (P42212), beta-lactamase (P14488), zinc finger protein (Q9NTW7), amyloid beta peptide (P05067), ribonuclease H (O60930), and ATP synthase (P19483). This paper finds that AlphaFold 2 retains its title as the most effective protein structure prediction model while RoseTTAFold comes in a close second.

Background

Accurately predicting the structure of proteins is crucial for understanding their underlying functionality. Disease protection, pharmaceutical manufacturing, and even increased biological understanding all fall once we are able to accurately and consistently predict the three dimensional structure of proteins. Historically, expensive and time consuming methods like x-ray crystallography[1] or nuclear magnetic resonance (NMR)[2] were used to experimentally determine the three dimensional structure of these biomolecules. While these methodologies guarantee

accuracy, they sacrifice precious resources monetarily and temporally. With the advent of modern deep learning advancements, we can instead use our previously collected data to train efficient and accurate protein structure prediction tools. Testing the efficacy of these models is imperative to guarantee their safe usage which ultimately allows quicker and more effective drug discovery and better research into serious disease prevention. This paper offers insights into the performance characteristics of each model to better inform which is more effective.

Methods

Sequence data is taken from UniProt.org. The models chosen for this paper are AlphaFold 2, RoseTTAFold, and ESMFold and are run on Google Colab[8,9,10]. These models were chosen for their performance characteristics and architectural uniqueness. Specifically, AlphaFold 2 uses a transformer based approach with iterative refinement, RoseTTAFold uses both convolutional neural networks and transformers, and ESMFold uses evolutionary scale modeling (ESM) which is grounded primarily on transformers though incorporates evolutionary information to improve efficacy. By using architecturally distinct models we can better investigate which architectural choices lead to effective models.

The evaluation metrics chosen for this paper include: RMSD and TM-score. RMSD will almost directly calculate the distance between each atom from its true, experimentally determined position. To better investigate the performance of each model we take the minimum RMSD and the average RMSD. Minimum RMSD will inform us on the potential of a model while average RMSD informs us on the consistency of the model. Finally, TM-score better informs us on the global similarity between protein structures. RMSD and TM-score evaluation were run online[11] using experimental and generated PDBs.

The chosen proteins represent a statistically diverse set of possible proteins. Each protein offers a unique challenge to the models. For instance, human hemoglobin is a tetrameric protein and green fluorescent protein has its unique barrel structure.

Results and Analysis

Our results are almost certainly conclusive: AlphaFold 2 is the dominant protein structure prediction software. As seen in Table 1, AlphaFold 2 nearly always outperformed other architectures on the range of proteins provided. Interestingly enough, we see that RoseTTAFold nears the predictive capabilities of AlphaFold 2 while ESMFold comes in a close third. On naive RMSD, AlphaFold 2 performs outstandingly well on P42212, the green fluorescent protein, bumping up its average performance and maximal performances drastically. Green fluorescent protein is common in the literature, suggesting that it may be present in AlphaFold's training set which may have led to its high accuracy. On the other hand, the more general TM-score may offer an analysis of which model is able to perform the best on general protein structure. AlphaFold 2 handedly outperforms ESMFold with RoseTTAFold following in a close second as well. It is worth noting the efficiency of ESMFold; both AlphaFold 2 and RoseTTAFold are computationally demanding in comparison. Given this, in a situation where time becomes a factor, ESMFold may be a viable alternative (at the cost of accuracy) for quick results.

	ESMFold	AlphaFold 2	RoseTTAFold
Minimum RMSD	0.988	**0.482**	0.956
Average RMSD	11.408	10.5758	**10.4498**
Maximum TM-score	0.827	**0.9346**	0.9170
Average TM-score	0.37175	**0.41187**	0.40364

Table 1: Evaluation Metrics for Various Protein Structure Prediction Models

Given the iterative approach AlphaFold 2 takes to determine protein structure, how much harder the model was trained, and the computational resources required to run the model in comparison

to the others, it's reasonable to realize that AlphaFold 2 is the best performing protein structure predictor. It is worth noting that these metrics are not comprehensive (as mentioned above with regards to time taken), though the initial results suggest that AlphaFold 2 is most effective.

Next steps include: running on a more diverse set of proteins, analyzing based on more metrics (perhaps more 'local' metrics instead of the global we used), and testing a more diverse set of architectures. Implementation wise, I could imagine reverse engineering a protein sequence that performs poorly on these models and then testing based on those sequences.

Bibliography

1. Ilari, A., Savino, C. (2008). Protein Structure Determination by X-Ray Crystallography. In: Keith, J.M. (eds) Bioinformatics. Methods in Molecular Biology™, vol 452. Humana Press. https://doi.org/10.1007/978-1-60327-159-2_3
2. Acton, T. B.; Xiao, R.; Anderson, S.; Aramini, J.; Buchwald, W. A.; Ciccosanti, C.; Conover, K.; Everett, J.; Hamilton, K.; Huang, Y. J., In *Fragment-Based Drug Design: Tools, Practical Approaches, and Examples*; Methods in Enzymology, Vol. 493; Kuo, L. C., Ed.; Academic Press, 2011; pp 21– 60.
3. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
4. Minkyung Baek *et al.*,Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**,871-876(2021).DOI:10.1126/science.abj8754
5. Zeming Lin *et al.* ,Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**,1123-1130(2023).DOI:10.1126/science.ade2574
6. *Uniprot website fallback message*. UniProt. (n.d.). https://www.uniprot.org/uniprot
7. Zhang, Y. and Skolnick, J. (2004), Scoring function for automated assessment of protein structure template quality. Proteins, 57: 702-710. https://doi.org/10.1002/prot.20264
8. Google. (n.d.). *ESMFold*. Google Colab. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/ESMFold.ipynb

9. Google. (n.d.). *RoseTTAFold*. Google Colab. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/RoseTTAFold.ipynb
10. Google. (n.d.). *AlphaFold 2*. Google Colab. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=kOblAo-xetgx
11. *TM-Score: Quantitative Assessment of similarity between protein structures*. Zhang Lab. (n.d.). https://zhanggroup.org/TM-score/