Using Artificial Intelligence (AI) to Predict A Structure of Protein Complex

Yiqing Zang

Academy of Mathematics and Science, Fort Hays State University

Research Mentor: Dr. Masakatsu Watanabe (Chemistry)

Abstract

Proteins play pivotal roles in essential life processes and elucidating their three-dimensional (3D) structures is crucial for understanding their functions. AlphaFold2, an advanced artificial intelligence-based method developed by Google DeepMind, has emerged as a promising tool for predicting protein structures. In this study, we evaluated the predictive capabilities of AlphaFold2. Our findings highlight AlphaFold2's efficacy in providing valuable insights into protein structure prediction, albeit with certain limitations. While AlphaFold2 represents a significant advancement in the field, its utility is best realized when integrated with complementary experimental approaches. Consequently, combining the strengths of AlphaFold2 with experimental validation remains essential for achieving a comprehensive understanding and precise characterization of protein. structures.

Introduction

The advent of AlphaFold2 has revolutionized structural biology. This AI-based technology provides unprecedented accuracy and speed in predicting protein structures, which can greatly shorten the time expense. AlphaFold2 has demonstrated remarkable accuracy in predicting protein structures. In this study, we aimed to explore the accuracy of AlphaFold2 by comparing predicted protein oligomer structures to experimental structures and then evaluating its precision by analyzing scoring functions, such as pLDDT of AlphaFold2, and Root Mean Square Deviations (RMDs). The results of this study manifest that AlphaFold2 has a board application prospect, considering its high accuracy. However, it has some limitations: AlphaFold2 seems to give a better result regarding hetero oligomer structure than homo oligomer structure. Nevertheless, it still has limitations. The AI-based models do not yet take the presence of ligands, covalent modifications, or environmental factors into their prediction algorithm and take protein-protein interactions in a limited way[1,2].

Results

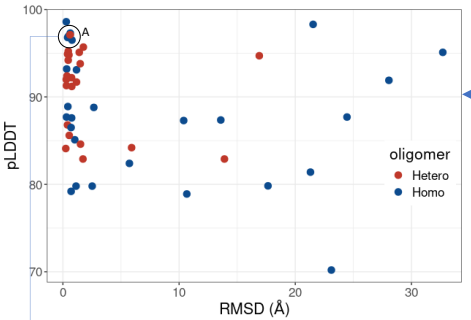
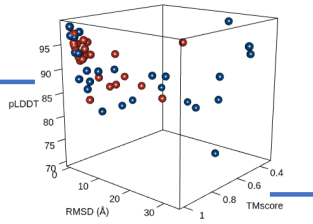


Fig. 1. 3D graph of pLDDT vs. TMscore vs. RMSD.



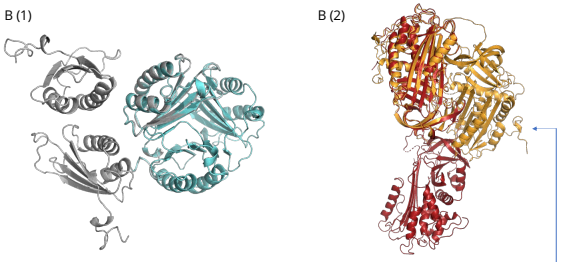


Fig. B. The aligned comparison of PDB and predicted protein 3D structure with a low TMscore value. (1) PDB ID: 8DBB, RMSD = 21.535Å, pLDDT = 98.3, Tmscore = 0.332; (2) PDB ID: 1ANT, RMSD = 23.106Å, pLDDT = 70.2, Tmscore = 0.465)

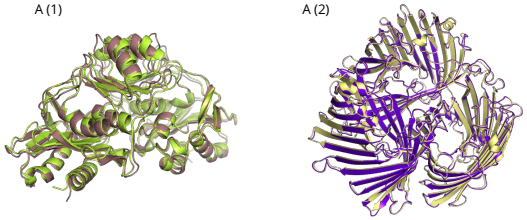


Fig. A. The aligned comparison of PDB and predicted protein 3D structure with a high pLDDT value. (1) PDB ID: 1DK4 – RMSD = 0.780Å, pLDDT = 96.5, TMscore = 0.981; (2) PDB ID: 1A0S RMSD = 0.408Å, pLDDT = 96.8, TMscore = 0.997.

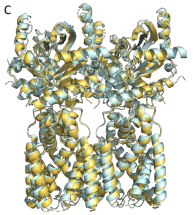
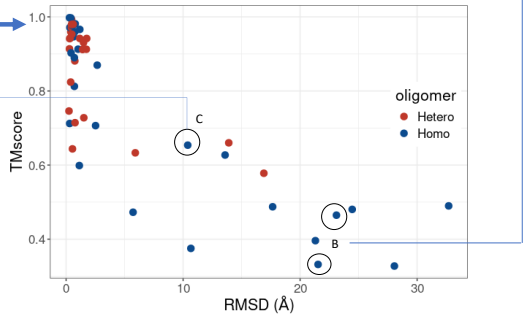


Fig. C. The aligned comparison of PDB and predicted protein 3D structure with a medium TMscore. PDB ID7SII: RMSD= 10.396Å, pLDDT:= 87.3, Tmscore = 0.654)



*In all the 3D protein structures displayed, darker color is the experimental structure, lighter color is the predicted structure.

Methodology

- Selected 52 protein sequences from PDB for this study:
  - Proteins with more than 100 residues and without DNA or RNA are included. Here are the classifications of studied protein:

Dimer	Trimer	Tetramer	Homo	Hetero
28	23	1	27	25

- PyMOL was used to compare predicted structures with experimental structures in the Protein Data Bank (PDB) and evaluate RMSDs between the structures.
- An R script was created to calculate TMscore to compare with pLDDT value of AlphaFold2.

Conclusion

- Results show that that AlphaFold2 provides a reliable prediction in a majority of cases. However, there are some limitations.
- We also found that AlphaFold2 gives a better prediction to hetero oligomer structure than to homo oligomer structure.
- AlphaFold2 serves as a promising tool in determining protein structure, but it can never give 100% accurate structure prediction. However, the structures will be very useful hypotheses.
- AlphaFold with other computational methods could improve the structure predictions.

Future Study

- More experiments on homo and hetero oligomer protein structure predictions are needed to make the comparison.
- Improve the ability of AlphaFold2 by involving more protein-protein interactions and environmental factors considerations.
- AlphaFold2 can be applied to discover new drugs, predict the effect of genetic variants, etc.. . With the development of artificial intelligence technology, it has a broad application prospect.

Reference

1. Terwilliger, T.C., Liebschner, D., Croll, T.I. *et al.* AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat Methods* **21**, 110–116 (2024). https://doi.org/10.1038/s41592-023-02087-4
2. Raphaelle Versini, Sujith Sritharan, Burcu Aykac Fas, Thibault Tubiana, Sana Zineb Aimeur, Julien Henri, Marie Erard, Oliver Nüsse, Jessica Andreani, Marc Baaden, Patrick Fuchs, Tatiana Galochkina, Alexios Chatzigoulas, Zoe Cournia, Hubert Santuz, Sophie Sacquin-Mora, and Antoine Taly*Journal of Chemical Information and Modeling* **2024** *64* (1), 26-41. https://doi.org/10.1021/acs.jcim.3c01361

Acknowledgements

- Academy of Mathematics and Science at FHSU
- FHSU Department of Chemistry