4-2022

Analysis of an Existing Method in Refinement of Protein Structure Predictions using Cryo-EM Images

Maytha Alshammari
Old Dominion University

Jing He
Old Dominion University

Willy Wriggers
Old Dominion University

Jiangwen Sun
Old Dominion University

Analysis of an Existing Method in Refinement of Protein Structure Predictions using Cryo-EM Images

Maytha Alshammari[1], Willy Wriggers[2], Jiangwen Sun[1], Jing He[1]
[1]*Department of Computer Science, Old Dominion University, Norfolk, VA*
[2]*Department of Mechanical and Aerospace Engineering , Old Dominion University, Norfolk, VA*

Abstract

Protein structure prediction produces atomic models from its amino acid sequence. Three-dimensional structures are important for understanding the function mechanism of proteins. Knowing the structure of a given protein is crucial in drug development design of novel enzymes. AlphaFold2 is a protein structure prediction tool with good performance in recent CASP competitions. Phenix is a tool for determination of a protein structure from a high-resolution 3D molecular image. Recent development of Phenix shows that it is capable to refine predicted models from AlphaFold2, specifically the poorly predicted regions, by incorporating information from the 3D image of the protein. The goal of this project is to understand the strengths and weaknesses of the approach that combines Phenix and AlphaFold2 using broader data. This analysis may provide insights for enhancement of the approach.

Method

We downloaded the atomic structure from Protein Data Bank (PDB) as well as high-resolution Cryo-EM images from Electron Microscopy Data Bank (EMDB) of 13 proteins. Then, we generated medium-resolution maps (6Å -12Å) from the high-resolution maps using Situs, a package for the modeling and refinement of biomolecular structures. The high-resolution maps and medium-resolution maps were sharpened using the phenix.local_aniso_sharpen. Next, we chopped the maps as a box around the atomic chain using the tool phenix.map_box. Then, we predicted the protein structure from the amino acid sequence using AlphaFold2, a protein structure prediction tool, which has a good performance in recent CASP competitions. We used the AlphaFold2 that is integrated with Phenix in a Colab Notebook. After that, we rebuilt the AlphaFold2 models with the high-resolution maps and medium-resolution maps using phenix.dock_and_rebuild tool in the Colab Notebook. Finally, we superimposed the AlphaFold2 models and the rebuilt models on the true atomic model using phenix.superpose_pdbs and we calculated the accuracy by comparing the superimposed models, AlphaFold2 and rebuilt, with the true atomic model using phenix. chain_comparison, which identifies how many of the CA atoms in one model match CA atoms in another model.

Data

The 13 proteins' IDs, their maps resolutions, the AlphaFold2 models' accuracy, and the rebuilt models' accuracy are shown in Figure 1 and Table 1.

Results

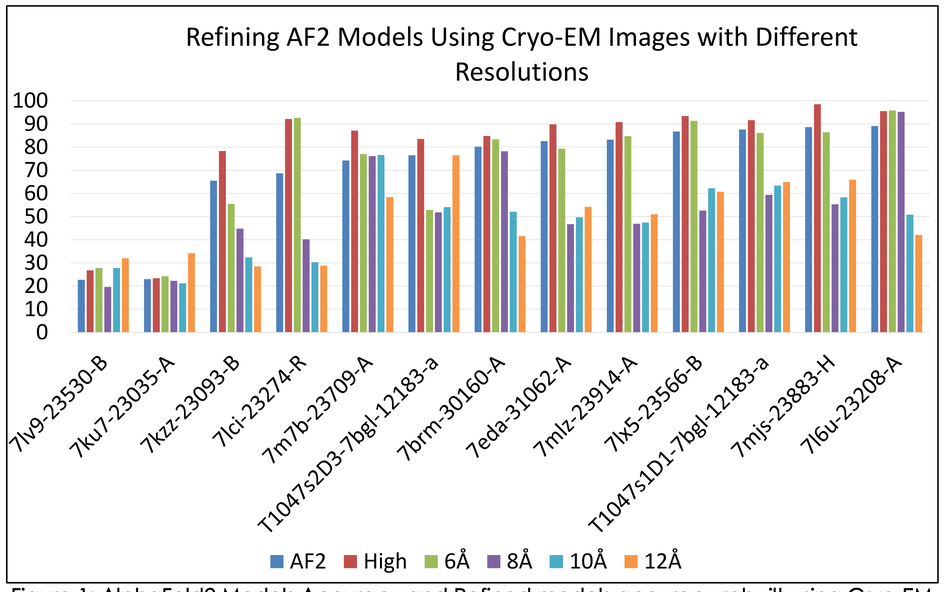

Figure 1: AlphaFold2 Models Accuracy and Refined models accuracy rebuilt using Cryo-EM Images with Different Resolutions.

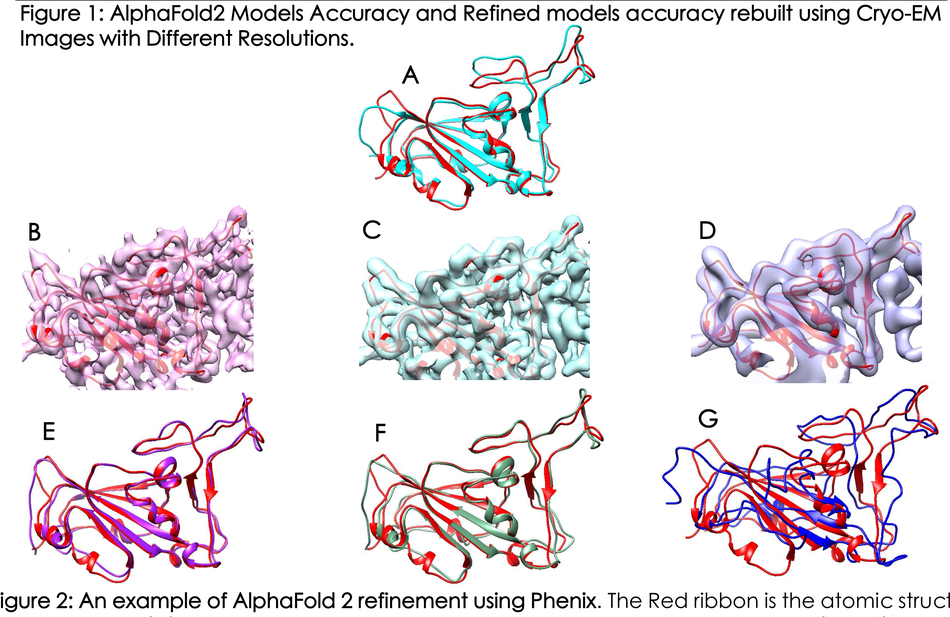

Figure 2: An example of AlphaFold 2 refinement using Phenix. The Red ribbon is the atomic structure of 7lx5 chain B. (A) The Cyan ribbon is the AF2 Predicted model. Accuracy is 86.7. (B,C,D)The Cryo-EM Image, EMDB 23566 (3.44Å, 6Å, 8Å respectively), used in the refinement. (E) The Purple ribbon is the rebuilt model with 3.44Å map. Accuracy is 93.4 (F) The Green ribbon is the rebuilt model with 6Å map. Accuracy is 91.3. (G) The Blue ribbon is the rebuilt model with 8Å map. Accuracy is 52.6.

Case[a]	Resolution[b]	AF2[c]	Refining[d]				
			High	6Å	8Å	10Å	12Å
7lv9-23530-B	4.5Å	22.7	26.8	27.8	19.6	27.8	32
7ku7-23035-A	3.4Å	23	23.4	24.2	22.3	21.2	34.2
7kzz-23093-B	3.42Å	65.5	78.3	55.5	44.8	32.4	28.5
7lci-23274-R	2.9Å	68.7	92.1	92.6	40.2	30.3	28.8
7m7b-23709-A	2.95Å	74.2	87.1	77	76.1	76.6	58.4
T1047s2D3-7bgl-12183-a	2.2Å	76.5	83.5	52.9	51.8	54.1	76.5
7brm-30160-A	3.6Å	80.2	84.8	83.3	78.2	52.1	41.6
7eda-31062-A	2.78Å	82.6	89.8	79.3	46.7	49.7	54.2
7mlz-23914-A	3.7Å	83.2	90.8	84.7	46.9	47.4	51
7lx5-23566-B	3.44Å	86.7	93.4	91.3	52.6	62.2	60.7
T1047s1D1-7bgl-12183-a	2.2Å	87.6	91.6	86.1	59.4	63.4	64.9
7mjs-23883-H	3.03Å	88.6	98.5	86.4	55.3	58.3	65.9
7l6u-23208-A	3.3Å	89.1	95.5	95.8	95.2	50.8	42.1

Table 1: Data and results. [a]Protein IDs (PDB ID_ EMDB ID_Chain ID). CASP cases IDs (CASP Target ID_ PDB ID_ EMDB ID_Chain ID). [b]The Resolution (electron density).[c]AlphaFold2 models' accuracy. [d]Refined models' accuracy rebuilt using Cryo-EM Images with Different Resolutions.

Conclusion

We examined the approach that combines Phenix and AlphaFold2 using broader data, and we found that refining AlphaFold2 models with high-resolution Cryo-EM Images can enhance the models' accuracy. Also, using the 6Å maps increase the models' accuracy in most cases. However, AlphaFold2 models' accuracy decreased, in general, when 8-12Å maps were used. In future work, we will analyze the results and enhance the approach.

References

[1] Terwilliger, Thomas C., et al. "Improving AlphaFold modeling using implicit information from experimental density maps." bioRxiv (2022).
[2] Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." *Nature* 596.7873 (2021): 583-589.
[3] Wriggers, Willy, Ronald A. Milligan, and J. Andrew McCammon. "Situs: a package for docking crystal structures into low-resolution maps from electron microscopy." Journal of structural biology 125.2-3 (1999): 185-195.

Acknowledgement

This work is supported by NIH R01-GM062968 and the scholarship to Maytha Alshammari by the government of Saudi Arabia.