

Quo Vadis Experimental Structural Biology?

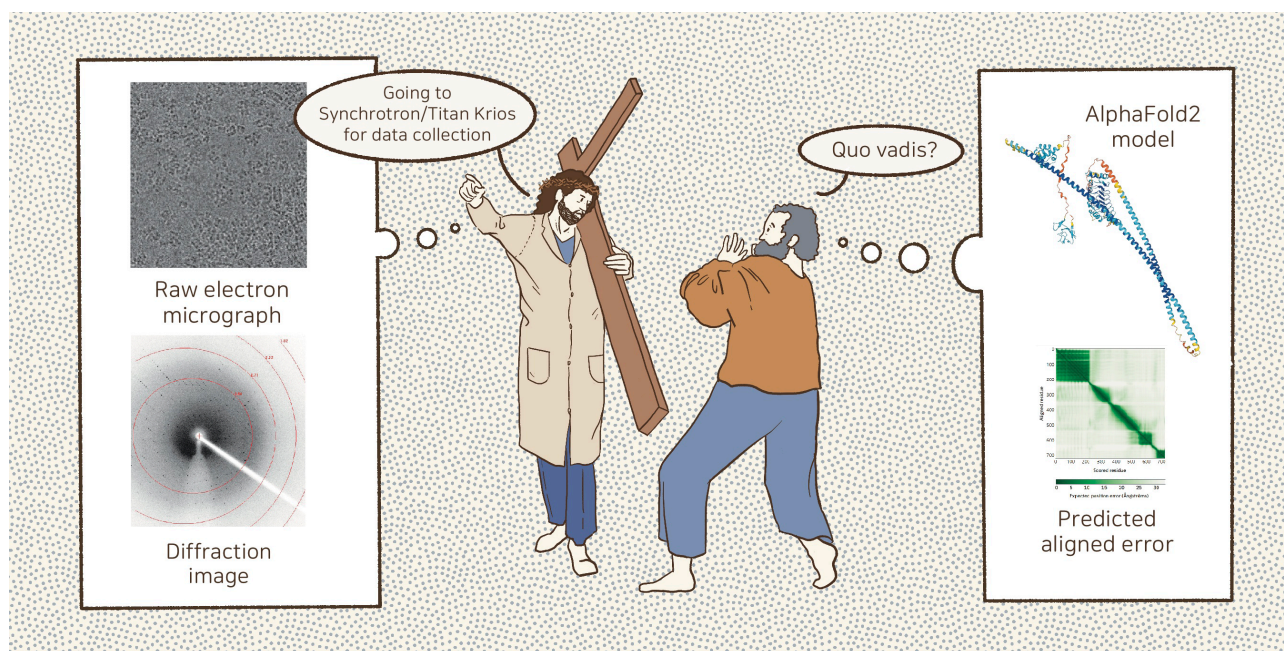
Hyun Kyu Song*

Department of Life Sciences, Korea University, Seoul 02841, Korea

*Correspondence: hksong@korea.ac.kr

<https://doi.org/10.14348/molcells.2023.2197>

www.molcells.org



AlphaFold2 was developed by DeepMind (UK), an artificial intelligence subsidiary of Alphabet Inc. (USA). AlphaFold2 is a deep learning system that is revolutionizing biology by producing accurate structural predictions of target proteins. However, AlphaFold2 still has various limitations. Experimental techniques have played a major role in determining the three-dimensional structures of proteins, and they have coevolved with *in silico* methods for mutual benefit. The figure was envisioned from a 1602 painting, “*Domine, quo vadis?*” by the Italian Baroque painter Annibale Carracci. The representative AlphaFold model on the right is a leucine-rich repeat and sterile alpha motif containing 1 isoform, CRA_a (UniProt: A0A024R870). The raw electron micrograph and diffraction image on the left are unpublished.

Received December 30, 2022; accepted January 2, 2023; published online January 27, 2023

eISSN: 0219-1032

©The Korean Society for Molecular and Cellular Biology.

©This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

In 2020, AlphaFold2 protein structure prediction was presented at the 14th meeting of the Critical Assessment of Structure Prediction (CASP14) and fundamentally changed the structural biology field (Callaway, 2020; Jumper et al., 2021). AlphaFold2's accurate performance has greatly impacted structural biology and medicine research. The AlphaFold2 neural network algorithm is based on deep learning processes that use prior knowledge of protein structures and multiple sequence alignments to accurately predict protein structures from the sequence data (Jumper et al., 2021). The deep learning system enabled the establishment of an AlphaFold database larger than the Protein Data Bank, which contains protein structures obtained using experimental techniques such as X-ray crystallography, cryogenic electron microscopy, and nuclear magnetic resonance spectroscopy. DeepMind and EMBL-EBI have made over 200 million predicted protein structures publicly available (<https://alphafold.ebi.ac.uk>) by including them in the UniProt database. Protein structures obtained using experimental techniques have historically been preferred over *in silico* predicted structures, mainly because of their accuracy. However, AlphaFold2 predictions are very reliable, as assessed by the structural biology community (Akdel et al., 2022). The experimental determination of protein structures is extremely painstaking and can take several years to complete, depending on the characteristics of the target molecules. In contrast, AlphaFold2 can generate highly accurate structures quickly (depending on the protein size). AlphaFold2 does not require expensive sample preparation or huge facilities such as the synchrotron facility, the Titan Krios microscope, or powerful magnets. Therefore, are experimental techniques in structural biology a thing of the past?

In the view of experimental structural biologists, the answer is no, as AlphaFold2 predictions still have limitations (Callaway, 2022). An important limitation is that although AlphaFold2 can accurately predict static protein structures, it cannot predict dynamic states, which are physiologically relevant. The native structures of many proteins are metastable and not necessarily the most energetically stable structures (Ghosh and Ranjan, 2020). For example, kinases fold into inactive conformations without signals but change into active conformations in response to signals to perform their roles in signal transduction (Huse and Kuriyan, 2002). Serpin protease inhibitors fold into different conformational states in cells, namely cleaved, active, and latent conformations (Ghosh and Ranjan, 2020). Depending on the reaction steps, the proteolytic chamber of the self-compartmentalizing caseinolytic protease shows conformational diversity (Kim et al., 2022). Depending on cellular conditions, such as pH, an autophagic receptor formed a filamentous assembly (Kwon et al., 2018), and cytoskeletal proteins were found to polymerize and depolymerize dynamically (Goodson and Jonasson, 2018). Generally, multiprotein complexes (including fibrous assemblies) are not accurately predicted by AlphaFold2. An important limitation is that the structures of protein complexes with nonprotein ligands (e.g., small compounds with therapeutic potential or binding partners, such as DNA or RNA) are not well predicted by AlphaFold2, which was developed and trained for protein structure determination but not for

docking (Jumper et al., 2021). Furthermore, protein modifications such as phosphorylation, glycosylation, lipidation, acetylation, and methylation affect the accuracy of AlphaFold2 predictions (Callaway, 2023). Like experimental techniques, the AlphaFold2 structures of intrinsically disordered proteins are mostly inaccurate. Furthermore, a single-point mutation can drastically affect protein folding, but the AlphaFold2 predicted structures of the mutated and wild-type proteins are often very similar. To overcome these limitations, numerous approaches are ongoing (Varadi and Velankar, 2022) and have been introduced in the recent CASP15 (Callaway, 2023). These efforts continuously strengthen artificial intelligence-based structure predictions. Perhaps one day, sequence-to-structure prediction will be straightforward for all situations, including all the components in entire organelles or even cells. However, given the complexity of biology, this may be some time away.

Nonetheless, the relatively accurate AlphaFold2 predictions have changed structural biology and related fields in a positive way. For initial model building with low-resolution experimental data from X-ray crystallography or cryogenic electron microscopy, AlphaFold2 can be greatly advantageous. Many successful examples of using AlphaFold2 models to obtain the phases of X-ray data with molecular replacement, model building for uninterpretable electron density, and biochemical experiments based on AlphaFold2 models (Cramer, 2021; Kleywegt and Velankar, 2022) are available. Although the accuracy of AlphaFold2 is not yet high enough for docking studies, rational drug design based on AlphaFold2's predicted models will likely expand enormously in the near future (Varadi and Velankar, 2022). To fully understand biology, various hybrid and integrated approaches are inevitable, and therefore experimental structural biologists can focus on other biochemical, biophysical, and cell biology experiments. Experimental structural biologists will still be needed to validate the structure of uncertainly predicted models. Artificial intelligence-based structural biology is one of the biggest trends, and experimental structural biologists may have to tread a thorny path. This situation reminded me of a scene from the Apocrypha Acts of Peter.

Peter: Domine, quo vadis? / Jesus: Romam vado iterum crucifigi.

When Peter asked Jesus where he was going, Jesus replied that he was going to Rome to be crucified again. Of course, it is not directly comparable, but the notion is that structure determination using experimental techniques is extremely difficult compared with computational prediction. It is my view that experimental and prediction methods will continue to coevolve to answer complex biological questions.

ACKNOWLEDGMENTS

This work is supported by grants from the National Research Foundation of Korea (2020R1A2C3008285 and 2021M3A9I4030068).

CONFLICT OF INTEREST

The author has no potential conflicts of interest to disclose.

ORCID

Hyun Kyu Song <https://orcid.org/0000-0001-5684-4059>

REFERENCES

- Akdel, M., Pires, D.E.V., Pardo, E.P., Janes, J., Zalevsky, A.O., Meszaros, B., Bryant, P., Good, L.L., Laskowski, R.A., Pozzati, G., et al. (2022). A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* *29*, 1056-1067.
- Callaway, E. (2020). 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* *588*, 203-204.
- Callaway, E. (2022). What's next for AlphaFold and the AI protein-folding revolution. *Nature* *604*, 234-238.
- Callaway, E. (2023). After AlphaFold: protein-folding contest seeks next big breakthrough. *Nature* *613*, 13-14.
- Cramer, P. (2021). AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* *28*, 704-705.
- Ghosh, D.K. and Ranjan, A. (2020). The metastable states of proteins. *Protein Sci.* *29*, 1559-1568.
- Goodson, H.V. and Jonasson, E.M. (2018). Microtubules and microtubule-associated proteins. *Cold Spring Harb. Perspect. Biol.* *10*, e022608.
- Huse, M. and Kuriyan, J. (2002). The conformational plasticity of protein kinases. *Cell* *109*, 275-282.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* *596*, 583-589.
- Kim, L., Lee, B.G., Kim, M., Kim, M.K., Kwon, D.H., Kim, H., Brotz-Oesterhelt, H., Roh, S.H., and Song, H.K. (2022). Structural insights into ClpP protease side exit pore-opening by a pH drop coupled with substrate hydrolysis. *EMBO J.* *41*, e109755.
- Kleywegt, G.J. and Velankar, S. (2022). Whither structural biologists? *IUCr* *9*(Pt 4), 399-400.
- Kwon, D.H., Park, O.H., Kim, L., Jung, Y.O., Park, Y., Jeong, H., Hyun, J., Kim, Y.K., and Song, H.K. (2018). Insights into degradation mechanism of N-end rule substrates by p62/SQSTM1 autophagy adapter. *Nat. Commun.* *9*, 3291.
- Varadi, M. and Velankar, S. (2022). The impact of AlphaFold Protein Structure Database on the fields of life sciences. *Proteomics* *2022* Nov 16 [Epub]. <https://doi.org/10.1002/pmic.202200128>