Alphafold opens the doors to de-orphanizing secreted proteins

Shruthi Viswanath[1,*]

[1]National Center for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India

*Corresponding Author: shruthiv@ncbs.res.in (SV)

NOTE: This is the accepted version without the type-setting by the journal. The final published article is at doi: 10.1016/j.cels.2024.10.010

Abstract

Danneskiold-Samsøe and co-workers have developed an *in silico* screening pipeline based on Alphafold2 for identifying single-pass transmembrane receptors for secreted peptides that play important roles in cell-cell signaling. Their approach can be used to de-orphanize a diverse range of ligands. The overall strategy can be valuable in screening for weak and transient interactions.

Main text

Cell-cell communication and signaling occur through interactions between membrane-bound receptors and secreted peptides acting as signaling molecules. Although more than two thousand known secreted peptides (ligands) exist in humans, only a tiny fraction of these have known receptors. Similarly, a large fraction of human receptors do not have a known ligand. Experimental identification of ligand-receptor pairs is challenging and time-consuming since these interactions are weak and transient and involve membrane proteins. However, characterizing these interactions is essential for understanding the mechanistic details of cell-cell communication as well as for probing their therapeutic potential[1].

In the study by Danneskiold-Samsøe and co-workers[1], the authors establish the use of Alphafold2 as a rapid tool to screen interactions between extracellular peptides and membrane receptors[2]. Alphafold and related methods have revolutionized structural biology by enabling accurate proteome-wide characterization of protein structures, facilitating improved structure determination for protein-protein complexes, and large assemblies[3]. Alphafold has been established as a tool for screening protein-protein interactions (identifying protein pairs that interact), enabling the identification of new interactions in proteome-wide screens[4,5]. Tools such as AlphaPulldown have been developed for performing *in silico* screens using Alphafold[6]. The interface predicted template matching (interface pTM) score is commonly used as the metric for determining if two proteins interact[4]. Previous studies have shown that Alphafold2 can be used to determine structures of protein-peptide complexes, with careful choice of input sequence fragments and multiple sequence alignment strategies[7,8]. This is one of the first studies to demonstrate the use of Alphafold for screening protein-peptide interactions.

In [1], the authors used Alphafold2 to de-orphanize secreted peptide ligands by identifying single-pass transmembrane receptors that bind to them. They first constructed a library of around one thousand single-pass human transmembrane receptors distributed across a range of tissue types and functions. Similarly, a library of secreted proteins containing close to two thousand entries was constructed. The choice of ligand sequence fragment: secreted or full-length peptide, did not affect the prediction accuracy significantly. In contrast, the choice of receptor sequence fragment mattered: using the extra-cellular sequence alone improved the prediction accuracy over using the full-length sequence. It is important to account for the membrane topology while constructing the input to Alphafold2, as it is blind to the membrane topology, and can sometimes predict structures with interlinked membranous, extracellular, and intracellular domains[1]. An earlier smaller-scale study using Alphafold2 for screening

ligands for membrane-bound receptors also used membrane topology as an additional filter for post-processing structures[9].

A key strength of the study is the benchmarking of Alphafold2 performance on a large number of structurally and functionally diverse receptor-ligand pairs, with careful attention paid to ensure no overlap with the Alphafold2 training set. The approach is limited by the computational time required for screening with large libraries with Alphafold2. Future improvements to the pipeline can look for ways to make it more efficient, for example, by pre-computing the MSA for each receptor[9]. Another limitation is the low accuracy on longer input sequences, which can perhaps be addressed by sequence fragmentation strategies[7,8]. The current approach by [1] is of broad applicability and can be readily used for de-orphanizing a range of secreted ligands such as tyrosine kinase ligands and hormones. It can also be extended to multi-pass membrane proteins such as G-protein coupled receptors, which hold significant pharmaceutical interest.

A significant fraction of eukaryotic proteins do not have an assigned cellular function. Since proteins perform their functions by interacting with other proteins and biomolecules, methods that elucidate these protein interactions can provide insights into protein function. De-orphanizing proteins is an exciting application of Alphafold, which is now a state-of-the-art method for predicting protein-protein interactions[3]. The study by Danneskiold-Samsøe and co-workers[1] and other studies on screening for protein-protein interactions with Alphafold[4,5] suggests a promising approach for de-orphanizing proteins based on their interactions. A candidate set of interacting pairs in a sub-cellular neighborhood can be obtained by filtering proteins by sub-cellular localization, or by using information from high-throughput experiments such as mass spectrometry and genetic screens. The input sequences of the pairs can be delineated based on their domains, and Alphafold metrics such as the interface pTM and interface PAE (predicted aligned error) can be used to identify potential interactions efficiently. This approach will be especially valuable in cases where the interactions are weak, transient, and challenging to characterize experimentally, such as those involved in signal transduction pathways.

The newest version of Alphafold, Alphafold3, holds exciting promise in this direction. The computational time for predictions has decreased significantly, longer input sequences can be used, and small molecules, DNA, and RNA can now be modeled[10]. This implies that the receptor and ligand can be modeled as oligomers if they are known to form them. Co-factors can also be used to provide additional context for predicting interactions. With these new advances in Alphafold, it is also tempting to speculate if protein-small molecule interactions can be screened using similar approaches. Future research will hopefully provide more insights in this direction.

Acknowledgments

The author thanks lab members Shreyas Arvindekar, Kartik Majila, Muskaan Jindal, and Omkar Golatkar for their useful comments on the manuscript.

This work has been supported by the following grants: Department of Atomic Energy (DAE) TIFR grant RTI 4006, Department of Science and Technology (DST) SERB grant SPG/2020/000475, and Department of Biotechnology (DBT) BT/PR40323/BTIS/137/78/2023 from the Government of India to S.V.

Declarations of Interest

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT to refine the wording. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

1. Danneskiold-Samsøe, N.B., Kavi, D., Jude, K.M., Nissen, S.B., Wat, L.W., Coassolo, L., Zhao, M., Santana-Oikawa, G.A., Broido, B.B., Garcia, K.C., et al. (2023). AlphaFold2 enables accurate deorphanization of ligands to single-pass receptors. Preprint, https://doi.org/10.1101/2023.03.16.531341 https://doi.org/10.1101/2023.03.16.531341.

2. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

3. Kovalevskiy, O., Mateos-Garcia, J., and Tunyasuvunakool, K. (2024). AlphaFold two years on: Validation and impact. Proc. Natl. Acad. Sci. U.S.A. *121*, e2315002121. https://doi.org/10.1073/pnas.2315002121.

4. Gao, M., Nakajima An, D., Parks, J.M., and Skolnick, J. (2022). AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. Nat Commun *13*, 1744. https://doi.org/10.1038/s41467-022-29394-2.

5. Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., Zhang, J., Ness, T.J., Banjade, S., Bagde, S.R., et al. (2021). Computed structures of core eukaryotic protein complexes. Science *374*, eabm4805. https://doi.org/10.1126/science.abm4805.

6. Yu, D., Chojnowski, G., Rosenthal, M., and Kosinski, J. (2023). AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. Bioinformatics *39*, btac749. https://doi.org/10.1093/bioinformatics/btac749.

7. Bret, H., Gao, J., Zea, D.J., Andreani, J., and Guerois, R. (2024). From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. Nat Commun *15*, 597. https://doi.org/10.1038/s41467-023-44288-7.

8. Lee, C.Y., Hubrich, D., Varga, J.K., Schäfer, C., Welzel, M., Schumbera, E., Djokic, M., Strom, J.M., Schönfeld, J., Geist, J.L., et al. (2024). Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. Molecular Systems Biology *20*, 75–97. https://doi.org/10.1038/s44320-023-00005-6.

9. Teufel, F., Refsgaard, J.C., Kasimova, M.A., Deibler, K., Madsen, C.T., Stahlhut, C., Grønborg, M., Winther, O., and Madsen, D. (2023). Deorphanizing Peptides Using Structure Prediction. J. Chem. Inf. Model. *63*, 2651–2655. https://doi.org/10.1021/acs.jcim.3c00378.

10. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.