



The transformative power of transformers in protein structure prediction

Bernard Moussard¹, Anand B. Swarna², and D. B. Bhattacharya

Edited by Axel Brunger, Stanford University, Stanford, CA; received March 1, 2023

Transformer neural networks have revolutionized structural biology with the ability to predict protein structures at unprecedented high accuracy. Here, we report the predictive modeling performance of the state-of-the-art protein structure prediction methods built on transformers for 69 protein targets from the recently concluded 15th Critical Assessment of Structure Prediction (CASP15) challenge. Our study shows the power of transformers in protein structure modeling and highlights future areas of improvement.

protein structure predictions | neural networks | transformers | deep learning

In 2020, DeepMind's AlphaFold2 method (1) made a major scientific breakthrough in predicting the three-dimensional (3D) structure of proteins from the amino acid sequences. AlphaFold2 demonstrated remarkable performance in the 14th edition of the biennial Critical Assessment of Structure Prediction (CASP14) meeting, attaining better predictive modeling accuracy than ever before (2). Notably, DeepMind released the source code of AlphaFold2 with an open-source license along with an extensive database of predicted protein structures (3). Meanwhile, the scientific community has been investigating ways to explore, exploit, and extend AlphaFold2 to further improve the state of the art of protein structure prediction for advancing structural biology research (4–6).

The core of the AlphaFold2 framework consists of transformer neural networks powered by the attention mechanism (7), a sequence transduction model originally proposed to draw global dependencies between input and output. What makes the transformer architecture particularly powerful is its ability to model long-range relationships in the input sequence beyond their sequential neighborhoods. AlphaFold2 harnessed the power of transformers by formulating a “two-track” neural architecture, named Evoformer, to jointly embed the evolutionary and spatial relationships between the amino acids using multiple sequence alignment (MSA) and pairwise representations, followed by a structure module based on equivariant transformer architecture with invariant point attention. RoseTTAFold (5) extended the two-track architecture of AlphaFold2 by adding a third track operating in 3D coordinate space using SE (3)-transformer (8). The key principle of the three-track RoseTTAFold network is to process MSA, pairwise, and coordinate information simultaneously, with accuracies approaching those of AlphaFold2.

It is worth noting that both AlphaFold2 and RoseTTAFold rely on the availability of evolutionary information in the form of MSAs, which are not always as abundant as these methods demand, such as with orphan proteins or rapidly evolving proteins. Protein language models (PLMs) aim to overcome this limitation by pretraining a family of transformer-based models of protein sequences. For example, OmegaFold (4) introduced a transformer-based PLM called OmegaPLM to learn single- and pairwise-residue embeddings as features that are fed into a geometry-inspired transformer block named Geoformer. ESMFold (6) trained a massive PLM called ESM-2 based on an encoder-only transformer architecture (7) and leverages the evolutionary information captured by the model using a modified Evoformer block of AlphaFold2 to remove its dependence on MSAs. Unlike AlphaFold2 or RoseTTAFold, neither OmegaFold nor ESMFold rely on MSAs or use known structures as templates.

Despite their methodological differences, the state-of-the-art methods for protein structure prediction have two things in common: 1) Their underlying neural network architectures use transformers as building blocks, and 2) their source codes are freely available with open-source license. That is, an open-source ecosystem of transformer-based neural networks is having a transformative effect in the field of protein structure prediction.

Given these recent advances, a natural question arises: how far have we progressed in predicting protein structures since the breakthrough of AlphaFold2 in 2020? It is widely accepted by the structure prediction community that blind structure prediction tests are needed to objectively assess the performance of new protein structure prediction methods. Fortunately, the recently concluded CASP15 meeting provides an excellent testbed to evaluate these emerging methods over the same set of proteins. Here, we benchmark the predictive modeling performance of AlphaFold2, RoseTTAFold, ESMFold, and OmegaFold on a set

Author contributions: D.B. designed the experiment of Computer Science, Blacksburg, VA 24061.

Author contributions: D.B. designed the experiment of Computer Science, Blacksburg, VA 24061.

The authors declare no competing financial interests. Copyright © 2023 the Author(s). This open access article is distributed under a Creative Commons Attribution License 4.0.

¹B.M. and R.R. contributed equally.

²To whom correspondence may be addressed: dbhattacharya@vt.edu.

This article contains supporting data for the main text. <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2303499120/-/DCU-supplemental>

Published July 31, 2023.

whereas the PLM-based methods essentially generated random predictions (GDT-TS score < 11). Finally, for the large multidomain target T1154 with a complex topology, all methods performed poorly; even the top-performing AlphaFold2 attained a low GDT-TS score of 24.3. That is, despite yielding near-perfect predictive accuracy for small proteins with simple topology, state-of-the-art methods are still far from perfect for medium to large proteins with complex topology.

In terms of correct overall topology prediction, quantified by the percentage of template modeling score (TM-score) > 0.5 (11), all methods performed quite well (Fig. 1E) with AlphaFold2 attaining the highest performance close to 80%. RoseTTAFold attained the second-best performance with just over 70%. That is, MSA-based methods attained better correct overall topology prediction performance compared to PLM-based approaches. We investigated whether the accuracy of MSA-based methods is related to the availability of evolutionary information by analyzing the MSA depth measured by the logarithm of the normalized number of effective sequences (Neff) against the local-distance difference test (IDDT) metric (12). The results show (Fig. 1F) that both AlphaFold2 and RoseTTAFold are somewhat MSA dependent, with RoseTTAFold exhibiting more MSA dependence. We also examined the relationship between protein size and IDDT metric for all four methods. Despite the apparent trend of a decline in prediction accuracy as the protein size increases (Fig. 1G), the accuracy is in fact largely size independent when the few targets with length more than 750 are discarded (Fig. 1G, *Inset*).

How does the stereochemistry of the predicted structural models compare with the experimental structures? The Ramachandran plots (13), capturing the ϕ , ψ backbone torsion angle distributions (Fig. 1H), and the MolProbity score (14) distributions (Fig. 1I), quantifying the estimated deviations of the predicted structural models from their ideal crystallographic resolutions, reveal that the stereochemistry of the MSA-based methods AlphaFold2 and RoseTTAFold are closer to experimental observations than that of the PLM-based methods ESMFold and OmegaFold. The PLM-based methods in particular exhibit noticeably lower stereochemical qualities having physically unrealistic local structural regions, limiting their applicability to calculations based on the predicted whole chain coordinates such as electrostatic potential, accessibility, etc.

With respect to the side-chain positioning, measured by the global distance calculation for side-chains (GDC-SC) metric (15), there is considerable room for improvement (Fig. 1J). While the most accurate method is still AlphaFold2, which outperformed the other methods for more than 80% of the cases, the mean GDC-SC score of AlphaFold2 fell short of 50. It is interesting to note that PLM-based methods ESMFold and OmegaFold outperformed MSA-based RoseTTAFold for a large majority of cases and attained higher mean GDC-SC scores. The case study

presented in Fig. 1D also underscores the low GDC-SC scores for almost all targets. Improved side-chain packing algorithms will undoubtedly benefit MSA- and PLM-based methods alike.

Discussion

Despite the dramatic progress in protein structure prediction with transformer-based neural networks trailblazed by AlphaFold2, our study reveals two lingering issues. First, accurate prediction of large multidomain proteins with complex topology remains challenging. For some of such targets, none of the methods were able to predict even the correct overall topology with appropriate domain orientations. Adjusting the transformer model architecture by emulating the principles of interprotein interactions for modeling intraprotein domain–domain interactions may be necessary for large multidomain proteins. Second, the accuracy of side-chain positioning remains low for all methods. Structure prediction algorithms will benefit from accurate side-chain packing, possibly by incorporating side-chain optimization in the neural architecture.

The rise of PLMs clearly offers some unique advantages by being applicable even for proteins that do not have identifiable homologous sequences in the current sequence databases and bypasses the computational overhead of MSA searching. However, the performance of ESMFold and OmegaFold in our benchmarking raises the question: have PLM-based methods already eliminated the need for evolutionary information? Probably not yet. But PLM can be effective for small proteins with simple topology. Combining the strengths of MSA and PLM may be a promising avenue for further progress, broadening the applicability of predictive protein modeling in structural biology.

Materials and Methods

We considered 69 single-chain protein targets from CASP15 having length less than 3,000 residues with available experimental coordinates. We obtained the sequences of these target proteins from the CASP15 website and predicted their structures using publicly-available versions of AlphaFold2, RoseTTAFold, OmegaFold, and ESMFold (*SI Appendix*) and subsequently evaluated the predictive modeling performance using standard evaluation metrics including GDT-TS (9), TM-score (11), IDDT (12), MolProbity (14), and GDC-SC (15) (*SI Appendix*).

Data, Materials, and Software Availability. Predicted protein structural models in Protein Data Bank (PDB) format and metrics calculated to benchmark the predictive modeling performance have been deposited in <https://github.com/Bhattacharya-Lab/CASP15> (10.5281/zenodo.7682977) (16). All other data are included in the article and/or [supporting information](#).

ACKNOWLEDGMENTS. This work was partially supported by the National Institute of General Medical Sciences (R35GM138146 to D.B.) and the National Science Foundation (DBI2208679 to D.B.).

1. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
2. J. Pereira *et al.*, High-accuracy protein structure prediction in CASP14. *Proteins: Struct. Funct. Bioinf.* **89**, 1687–1699 (2021).
3. M. Varadi *et al.*, AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
4. R. Wu *et al.*, High-resolution de novo structure prediction from primary sequence. *BioRxiv* [Preprint] (2022). <https://doi.org/10.1101/2022.07.21.500999> (Accessed 3 January 2023).
5. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
6. Z. Lin *et al.*, Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
7. A. Vaswani *et al.*, “Attention is all you need” in *Advances in Neural Information Processing Systems*, I. Guyon *et al.*, Eds. (Curran Associates, Inc., 2017), vol. **30**, pp. 5998–6008.
8. F. Fuchs, D. Worrall, V. Fischer, M. Welling, Se (3)-transformers: 3d roto-translation equivariant attention networks. *Adv. Neural Inform. Process. Syst.* **33**, 1970–1981 (2020).
9. A. Zemla, LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31**, 3370–3374 (2003).
10. L. N. Kinch *et al.*, CASP9 target classification. *PROTEINS: Struct. Funct. Bioinf.* **79**, 21–36 (2011).
11. J. Xu, Y. Zhang, How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
12. V. Mariani, M. Biasini, A. Barbato, T. Schwede, IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).
13. G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
14. V. B. Chen *et al.*, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D: Biol. Crystallogr.* **66**, 12–21 (2010).
15. D. A. Keedy *et al.*, The other 90% of the protein: Assessment beyond the C α s for CASP8 template-based and high-accuracy models. *Proteins: Struct. Funct. Bioinf.* **77**, 29–49 (2009).
16. B. Moussad, R. Roche, D. Bhattacharya, Data for CASP15 performance benchmarking of the state-of-the-art protein structure prediction methods (v1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8153087>. Accessed 17 July 2023.