Beni-Suef University Journal of Basic and Applied Sciences

LETTER TO THE EDITOR

Open Access

AlphaFold-latest: revolutionizing protein structure prediction for comprehensive biomolecular insights and therapeutic advancements

Henrietta Onyinye Uzoeto[1,2], Samuel Cosmas[3], Toluwalope Temitope Bakare[4,5] and Olanrewaju Ayodeji Durojaye[6,7,8*]

Abstract

Breakthrough achievements in protein structure prediction have occurred recently, mostly due to the advent of sophisticated machine learning methods and significant advancements in algorithmic approaches. The most recent version of the AlphaFold model, known as "AlphaFold-latest," which expands the functionalities of the groundbreaking AlphaFold2, is the subject of this article. The goal of this novel model is to predict the three-dimensional structures of various biomolecules, such as ions, proteins, nucleic acids, small molecules, and non-standard residues. We demonstrate notable gains in precision, surpassing specialized tools across multiple domains, including protein–ligand interactions, protein–nucleic acid interactions, and antibody–antigen predictions. In conclusion, this AlphaFold framework has the ability to yield atomically-accurate structural predictions for a variety of biomolecular interactions, hence facilitating advancements in drug discovery.

Keywords Alphafold-latest, Alphafold2, Google DeepMind, AlphaFold-multimer, CASP

*Correspondence:
Olanrewaju Ayodeji Durojaye
lanredurojaye@mail.ustc.edu.cn
[1] Department of Therapy and Applied Science, Federal University of Allied Health Sciences, Enugu, Enugu State, Nigeria
[2] Department of Biological Sciences, Coal City University, Emene, Enugu State, Nigeria
[3] Department of Biochemistry, University of Nigeria, Nsukka, Enugu State, Nigeria
[4] Department of Nursing Sciences, Babcock University, Ilishan-Remo, Ogun State, Nigeria
[5] Nursing Department, The Alpha Assisted Reproductive Klinic, Ikoyi, Lagos State, Nigeria
[6] MOE Key Laboratory of Membraneless Organelle and Cellular Dynamics, Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230027, Anhui, China
[7] School of Life Sciences, Department of Molecular and Cell Biology, University of Science and Technology of China, Hefei, China
[8] Department of Chemical Sciences, Coal City University, Emene, Enugu State, Nigeria

Dear Editor,

Recent advances in protein structure prediction have witnessed groundbreaking developments, largely propelled by the emergence of advanced machine learning techniques and substantial improvements in algorithmic approaches [1–3]. Key highlights in this dynamic field include the release of AlphaFold2 by DeepMind which marked a watershed moment in protein structure prediction [3]. Leveraging deep learning methodologies, AlphaFold demonstrated remarkable accuracy, often rivaling experimental techniques such as X-ray crystallography and cryo-electron microscopy [3]. This success has paved the way for subsequent iterations, such as AlphaFold-latest, with expanded capabilities covering a broader range of biomolecular interactions.

The field has also seen a diversification of approaches beyond traditional methods. Integrating evolutionary

information, co-evolution analysis, and deep learning, these approaches capitalize on large-scale genomic data and advancements in computational power [4]. Such diversity allows for more robust predictions across various protein classes and interaction types. Recent models exhibit enhanced accuracy in predicting intricate protein structures. They showcase improved generalization capabilities, allowing accurate predictions even for proteins with low homology to training data. This addresses a longstanding challenge in the field, making structural predictions applicable to a more extensive range of biological entities [5].

Language model has become a household phrase. Yet, what many may not realize is the profound impact language models have had within the realm of protein research [6, 7]. Among these advancements stand the illustrious AlphaFold2, renowned for its uncanny ability to accurately predict protein structures solely from amino acid sequences, employing an attention-based transformer architecture [8]. This transformative architecture, pioneered by Google in 2017, marked a paradigm shift in natural language processing, enabling machines to understand and process language with unprecedented nuance and context. Indeed, the sequence–structure–function paradigm of proteins lies at the very heart of molecular biology, serving as the linchpin for understanding biological mechanisms. By harnessing language models borrowed from the domain of computer science, we gain a powerful lens through which to explore the intricate relationship between protein sequences, structures, and functions [6].

In the domain of protein sciences, language models serve a dual purpose: They excel in protein representation and facilitate protein design. By encoding proteins into a format intelligible to machines, language models pave the way for enhanced understanding of their structural and functional properties. Moreover, they empower researchers to venture into the realm of protein design, leveraging computational prowess to engineer proteins with tailored functionalities for a myriad of applications [9]. In essence, the marriage of language models and protein sciences heralds a new era of discovery and innovation, where the boundaries between disciplines blur, and the vast potential of interdisciplinary collaboration unfolds. As we continue to unlock the secrets encoded within the language of proteins, the transformative impact of language models will undoubtedly continue to reverberate throughout the scientific community and beyond [6, 9].

Advancements extend beyond single-chain protein structures to encompass protein–protein interactions, protein–ligand binding, and nucleic acid interactions. Models such as AlphaFold-multimer and AlphaFold-latest showcase the ability to predict structures of complexes, including non-protein elements such as nucleic acids and small molecules [10]. This expansion broadens the scope for studying the intricacies of biomolecular assemblies. The integration of experimental data, such as cryo-EM maps and NMR (nuclear magnetic resonance) data, with computational predictions further refines and validates models. This synergistic approach enhances the accuracy and reliability of predicted structures, providing a more comprehensive understanding of protein conformations [10]. The high accuracy achieved by recent models holds significant implications for drug discovery. Virtual screening and structure-based drug design benefit from reliable predictions of protein–ligand interactions. This has the potential to expedite the identification of drug candidates and streamline the drug development process [11].

Several recent advances emphasize open science and collaborative efforts. Initiatives like the Critical Assessment of Structure Prediction (CASP) provide a platform for evaluating and comparing different models. Open-sourcing models and datasets foster transparency and accelerate progress across the scientific community [4]. These recent strides in protein structure prediction reflect a transformative era in computational biology. These advances not only push the boundaries of prediction accuracy but also open up new possibilities for understanding the complexities of biomolecular interactions, accelerating drug discovery, and contributing to a more comprehensive knowledge of cellular processes. The integration of diverse methodologies and ongoing collaborative efforts position the field for continued breakthroughs in the coming years [12].

Further in this piece, we highlight the progress on the latest iteration of the AlphaFold model, termed "AlphaFold-latest," which extends the capabilities of the groundbreaking AlphaFold2 [13]. This new model aims to predict the 3D structures of a wide range of biomolecules, including proteins, nucleic acids, small molecules, ions, and modified residues. This development highlights significant improvements in accuracy, outperforming specialized tools in various categories, such as protein–ligand interactions, protein–nucleic acid interactions, and antibody–antigen predictions, while the outcome indicates the potential for achieving atomically-accurate structure predictions for diverse biomolecular interactions within the AlphaFold framework. This expanded scope is crucial for understanding the full complexity of biological systems [13].

The reported performance of AlphaFold-latest is highlighted across various benchmarks, demonstrating superior accuracy compared to the previous models and specialized tools. The model excels in ligand docking, protein–protein interactions, and interactions involving

nucleic acids [13]. The results showcase the model's ability to predict the 3D structures of different biomolecular entities, including proteins, nucleic acids, ligands, and modified residues, while also emphasizing the model's generalizability and potential applications in diverse scientific domains. AlphaFold-latest's success in predicting therapeutically relevant structures, including covalently bound ligands and structures with unique folds, underscores its versatility. The ability to make accurate predictions for challenging drug targets suggests practical implications for drug design and therapeutic interventions [13]. Following this segment is some insights into the realm of biomolecular discovery, guided by the remarkable capabilities of AlphaFold-latest and its unprecedented precision in predicting protein structures and its potential in therapeutic advancements.

Unprecedented accuracy AlphaFold-latest, the latest iteration of DeepMind's groundbreaking protein folding algorithm, stands as a pinnacle of achievement in predictive accuracy. By leveraging a sophisticated blend of deep learning and evolutionary principles, AlphaFold-latest has demonstrated unparalleled precision in predicting protein structures from their amino acid sequences. This level of accuracy transcends previous limitations, offering researchers a reliable blueprint of protein structures with unprecedented fidelity [13].

Expeditious insights Traditional methods of experimental protein structure determination, such as X-ray crystallography and cryo-electron microscopy, are often time-consuming and resource-intensive. In contrast, AlphaFold-latest expedites the process by swiftly generating accurate structural models, thereby accelerating the pace of biomolecular research. This rapid turnaround time empowers scientists to glean insights into the structure–function relationships of proteins more efficiently, unlocking a deeper understanding of their biological roles and mechanisms of action [13].

Facilitating drug discovery The ability to accurately predict protein structures has profound implications for drug discovery and development. With AlphaFold-latest, researchers can now elucidate the three-dimensional architectures of key drug targets with unprecedented precision. This enables rational drug design, wherein therapeutics can be tailored to interact more effectively with their target proteins, thereby enhancing efficacy and minimizing off-target effects. Additionally, AlphaFold-latest aids in the identification of druggable binding sites and facilitates virtual screening of small molecule compounds, expediting the process of drug candidate selection [13].

Insights into disease mechanisms Many diseases, ranging from cancer to neurodegenerative disorders, are rooted in aberrant protein function. By accurately predicting protein structures, AlphaFold-latest provides invaluable insights into the molecular underpinnings of disease. Researchers can elucidate how mutations alter protein structures and functions, unraveling the intricate mechanisms driving pathogenesis. This deeper understanding of disease mechanisms lays the groundwork for the development of targeted therapies and precision medicine approaches [13].

Empowering structural biology AlphaFold-latest democratizes access to structural biology insights, making advanced computational techniques accessible to researchers worldwide. Its open-access framework and user-friendly interface empower scientists from diverse backgrounds to explore protein structures and interrogate biomolecular phenomena with unprecedented granularity. This democratization of structural biology catalyzes collaboration and innovation, fostering a vibrant scientific community poised to tackle the most pressing challenges in biology and medicine [13].

The ongoing development of AlphaFold-latest holds promise for the future of computational structural biology. The model's current capabilities, especially in predicting diverse biomolecular interactions, open avenues for advancing research in understanding biological processes and designing novel therapeutics [13]. Future research could focus on refining the model further and extending its applications. Continuous refinement of AlphaFold-latest could involve addressing specific challenges, such as improving accuracy in large complexes or enhancing predictions for specific classes of biomolecules. Expansion to cover additional types of interactions or structural features may broaden its utility. Integrating experimental data into the model training process could also enhance accuracy and reliability. Combining computational predictions with experimental results may provide a more comprehensive understanding of biomolecular structures and interactions (Fig. 1).

1 Future perspective

Looking ahead, recent strides in protein structure prediction signal a transformative era poised to reshape the landscape of biomolecular research and therapeutic development. Fueled by advancements in machine learning techniques and algorithmic approaches, the field is witnessing unprecedented progress, propelled by seminal innovations such as DeepMind's AlphaFold2 [3].

The release of AlphaFold2 marked a monumental milestone in protein structure prediction, revolutionizing the field with its remarkable accuracy. By harnessing the power of deep learning methodologies, AlphaFold2 overcame previous limitations, rivaling experimental techniques such as X-ray crystallography and cryo-electron microscopy in predictive precision [3, 14]. This

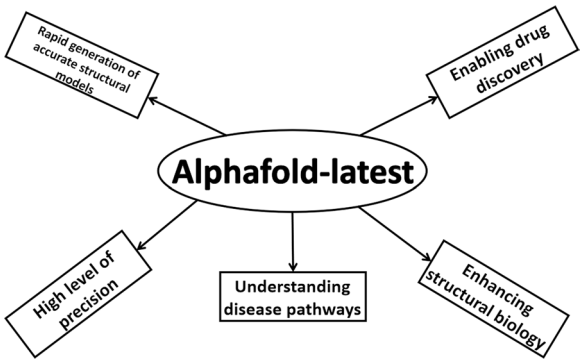

Fig. 1 Chat illustrating potential functionalities of AlphaFold-latest which represents a watershed moment in the field of protein structure prediction, catalyzing a paradigm shift that holds profound implications for comprehensive biomolecular insights and therapeutic advancements

breakthrough laid the foundation for subsequent iterations, including the highly anticipated AlphaFold-latest, which promises expanded capabilities across a broader spectrum of biomolecular interactions [13].

Moreover, the landscape of protein structure prediction has evolved to embrace a diverse array of methodologies beyond traditional approaches. Integrating evolutionary information, co-evolution analysis, and deep learning, these innovative strategies capitalize on vast genomic data and computational resources, enabling more robust predictions across diverse protein classes and interaction types [13, 15].

Crucially, recent models exhibit enhanced accuracy and generalization capabilities, addressing long-standing challenges in the field and making structural predictions applicable to a wider range of biological entities. This progress extends beyond single-chain protein structures to encompass complex biomolecular assemblies, including protein–protein interactions, protein–ligand binding, and nucleic acid interactions [16]. Furthermore, the integration of experimental data, such as cryo-EM maps and NMR data, with computational predictions has refined and validated models, providing a more comprehensive understanding of protein conformations. This synergistic approach enhances the accuracy and reliability of predicted structures, paving the way for accelerated drug discovery efforts and therapeutic interventions [17].

The accurate prediction of protein–ligand interactions has direct implications for drug discovery. Future applications may involve leveraging AlphaFold-latest for virtual screening of potential drug candidates, optimizing lead compounds, and designing molecules with specific binding properties. Furthermore, as the model continues to advance, there is potential for tailoring predictions based on individual genetic variations. Personalized medicine approaches could benefit from accurately predicting how specific individuals respond to certain drug molecules or therapies. Beyond applications in drug discovery, AlphaFold-latest can contribute to fundamental biological research by providing detailed structural insights into diverse biomolecular interactions. This includes unraveling the intricacies of cellular processes and pathways.

In conclusion, the progress on AlphaFold-latest represents a significant leap forward in the field of computational structural biology. The model's accuracy and expanded scope offer exciting possibilities for both applied and fundamental research, with the potential to impact drug development, personalized medicine, and our understanding of complex biological systems. Indeed, the future of computational structural biology holds great promise, with initiatives like AlphaFold-latest poised to push the boundaries of prediction accuracy and open new frontiers in understanding biomolecular interactions. As refinement and expansion continue upon these groundbreaking advancements, the integration of diverse methodologies and collaborative efforts will drive further breakthroughs, propelling the field toward new horizons of discovery and innovation.

Abbreviations

CASP Critical Assessment of Structure Prediction
NMR Nuclear magnetic resonance

Acknowledgements

Not applicable.

Author contributions

All authors contributed equally to the writing of this article. All authors have read and approved the manuscript.

Funding

Authors received no funding for this project from any organization.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Authors declare no competing interest.

Received: 16 December 2023 Accepted: 13 May 2024
Published online: 17 May 2024

References

1. Bryant P, Pozzati G, Zhu W, Shenoy A, Kundrotas P, Elofsson A (2022) Predicting the structure of large protein complexes using AlphaFold and

Monte Carlo tree search. Nat Commun 13(1):6028. https://doi.org/10.1038/s41467-022-33729-4
2. Durojaye OA, Yekeen AA, Idris MO, Okoro NO, Odiba AS, Nwanguma BC (2024) Investigation of the MDM2-binding potential of de novo designed peptides using enhanced sampling simulations. Int J Biol Macromol 26:131840. https://doi.org/10.1016/j.ijbiomac.2024.131840
3. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2
4. Liu J, Guo Z, Wu T, Roy RS, Quadir F, Chen C, Cheng J (2023) Enhancing AlphaFold-multimer-based protein complex structure prediction with MULTICOM in CASP15. Commun Chem 6(1):188. https://doi.org/10.1101/2023.05.16.541055
5. O'Reilly FJ, Graziadei A, Forbrig C, Bremenkamp R, Charles K, Lenz S, Elfmann C, Fischer L, Stülke J, Rappsilber J (2023) Protein complexes in cells by AI-assisted structural proteomics. Mol Syst Biol 19(4):e11544. https://doi.org/10.15252/msb.202311544
6. Huang T, Li Y (2023) Current progress, challenges, and future perspectives of language models for protein representation and protein design. Innovation (Camb) 4(4):100446. https://doi.org/10.1016/j.xinn.2023.100446. PMID:37485078;PMCID:PMC10362512
7. Vu MH, Akbar R, Robert PA, Swiatczak B, Sandve GK, Greiff V, Haug DTT (2023) Linguistically inspired roadmap for building biologically reliable protein language models. Nat Mach Intell 5(5):485–496
8. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D (2021) Highly accurate protein structure prediction for the human proteome. Nature 596(7873):590–596. https://doi.org/10.1038/s41586-021-03828-1
9. Unsal S, Atas H, Albayrak M, Turhan K, Acar AC, Doğan T (2022) Learning functional properties of proteins with language models. Nat Mach Intell 4(3):227–245
10. Zhu W, Shenoy A, Kundrotas P, Elofsson A (2023) Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. Bioinformatics 39(7):btad424. https://doi.org/10.1093/bioinformatics/btad424
11. Johansson-Åkhe I, Wallner B (2022) Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. Front Bioinform 26(2):959160. https://doi.org/10.3389/fbinf.2022.959160
12. Chen B, Xie Z, Qiu J, Ye Z, Xu J, Tang J (2023) Improved the heterodimer protein complex prediction with protein language models. Brief Bioinform 24(4):221. https://doi.org/10.1093/bib/bbad221
13. Google DeepMind AlphaFold Team and Isomorphic Labs Team. Performance and structural coverage of the latest, in-development AlphaFold model. https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/a-glimpse-of-the-next-generation-of-alphafold/alphafold_latest_oct2023.pdf. Accessed 25 Nov 2023
14. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Garcia KC, Grishin NV, Adams PD, Read RJ, Baker D (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science 373:871–876
15. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749
16. Hekkelman ML, de Vries I, Joosten RP, Perrakis A (2023) AlphaFill: enriching AlphaFold models with ligands and cofactors. Nat Methods 20:205–213
17. Holcomb M, Chang Y-T, Goodsell DS, Forli S (2023) Evaluation of AlphaFold2 structures as docking targets. Protein Sci 32:e4530

Publisher's Note