Proceedings of ICMMGH 2026 Symposium: Cell and Molecular Biology: Understanding Immune Cell Functions
DOI: 10.54254/2753-8818/2026.PJ30488

Applications of AlphaFold in the Medical Field

Zhi Lin Tan

Shanghai Macduffie Bilingual High School, Shanghai, China
DMNKC@outlook.com

Abstract. Traditional protein structure determination techniques, including X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, are typically time-consuming and labor-intensive. Moreover, these approaches face inherent constraints when dealing with specific protein classes, such as membrane proteins, which pose substantial technical hurdles. This severely hinders in-depth elucidation of the molecular mechanisms underlying life processes and impedes advancements in drug discovery and development. However, the advent of AlphaFold, an AI, has completely altered the landscape. AlphaFold can rapidly and accurately predict a protein's three-dimensional architecture directly from its amino acid sequence alone, representing a paradigm shift in structural biology and broader life sciences. Therefore, through a review of the literature and representative cases, this paper systematically examines the fundamental principles and evolutionary trajectory of AlphaFold. This study highlights its key applications in target discovery, protein design, and genetic mutation analysis in medical research. In addition, it also evaluates the technological strengths of AlphaFold and explores its existing limitations as well as future development directions.

Keywords: AlphaFold, Artificial Intelligence, Structure Prediction, Deep Learning, Biomedical Research

1. Introduction

Understanding the three-dimensional architecture of proteins is pivotal for elucidating their biological functions, as proteins serve as core components of living organisms and primary mediators of biological processes. This understanding provides a theoretical foundation and technical support for studies in diverse subfields of the life sciences, encompassing structure-based drug discovery and design, protein engineering, and the mechanistic analysis of diseases [1,2]. As a powerful tool based on advanced machine learning, AlphaFold (AF) revolutionized protein structure prediction. The application of AF facilitates the accurate and efficient determination of proteins' three-dimensional structures. Jumper et al. demonstrated during the Critical Assessment of Structure Prediction (CASP) 14 that their neural network-based model, AlphaFold 2 (AF2), could achieve atomic-level accuracy in predictions [3]. Building on this, Abramson et al. subsequently introduced AlphaFold 3 (AF3), predicts the structures of not only canonical proteins but also nucleic acids, small molecules, ions, and modified residues [4].

DOI: 10.54254/2753-8818/2026.PJ30488

However, existing research on this innovative technology has predominantly focused on single-dimensional analyses. This paper systematically examines the application and future development trends of AF technology in medical research through a literature review and case studies, and provides theoretical implications for this revolutionary tool in redefining the paradigms of medical research.

2. Fundamentals and advances of AlphaFold

AF1 builds on decades of cumulative research, leveraging large-scale genomic datasets to support protein structure prediction. The inaugural version, developed by DeepMind, operated by training a neural network to predict inter-residue distance scores, integrating these probabilistic predictions into a unified metric for assessing the accuracy of candidate protein structures. At the same time, they also evaluated the similarity between candidate structures and the native structure based on the neural network output [5]. AF2 addressed and substantially resolved the protein folding problem that had perplexed the biological community for five decades, specifically the challenge of accurately predicting three-dimensional structures from amino acid sequences. By employing an attention mechanism, AF2 can process protein sequences along with evolutionary information to directly predict the three-dimensional coordinates of all heavy atoms. At its core, the neural network module designated as the Evoformer processes both Multiple Sequence Alignment (MSA) and inter-residue pair representations. It enables continuous information exchange between these two representations throughout the network, thereby facilitating the direct inference of evolutionary relationships and geometric constraints [3].

The core of the newest model, AF3, is a module called “Pairfotmer”, which replaces the complex Evoformer with a simplified “MSA module”. Additionally, the model incorporates a diffusion module that operates directly on raw atomic coordinates. This approach leads to a key benefit: the algorithm can handle diverse chemical inputs without requiring specialized, molecule-specific output representations [4].

3. Typical applications of AlphaFold

3.1. Accelerating drug target discovery and validation

AlphaFold 3 (AF3) transforms drug target identification and validation by predicting complex structures exclusively from protein sequences and ligand Simplified Molecular Input Line Entry System (SMILES) strings. In contrast to conventional molecular docking approaches, which rely on experimentally determined protein structures, AF3's “blind prediction” capability confers a unique advantage in the exploration of novel therapeutic targets. Experimental data from the Posebusters benchmark, which utilized 428 high-quality complex structures not seen during AF3’s training, show a ligand docking success rate (RMSD < 2 Å) of 76.4%, significantly surpassing conventional docking tool Vina (52.3%) [6]. This performance gap signals a paradigm shift in computational structural biology.

However, the truly transformative attribute of AF3 resides not solely in its statistical advantage but in its fundamental reshaping of the research workflow. Data show that AF3's prediction of the binding mode of the clinical inhibitor LGK974 to its target PORCN (RMSD = 1.00 Å) is remarkably consistent with experimental structures [4]. This enables reliable virtual screening and molecular design even before experimental structures become available. As such, AF3 is positioned to directly

DOI: 10.54254/2753-8818/2026.PJ30488

impact and accelerate early-stage drug discovery, particularly in target validation and lead compound optimization.

3.2. Promoting protein design

AF technology has emerged as a core driving force in protein design through two approaches: optimizing existing designs and creating entirely new ones, revolutionizing the entire field of protein research [7,8].

Firstly, for enhancing design through accurate prediction, the utility of AF has been validated in antibody drug optimization. Specifically, using antibody-antigen interactions as an example, AF can predict binding modes with high accuracy and help design antibody drugs with higher affinity and stronger specificity. For instance, in the study of the ISB6 antibody and the mesothelin peptide, the predicted structure (DockQ = 0.85) showed excellent agreement with the experimental structure, revealing the structural details of the binding interface [4]. Based on such structural information, researchers can rationally design antibody drugs with superior performance, markedly reducing the extensive trial-and-error procedures associated with traditional experimental approaches.

Secondly, the application of AF technology has evolved from predicting natural structures to "creating entirely new proteins". Baker et al. employed the AlphaFold toolchain, combining it with deep learning for tool design, and successfully designed a kind of microprotein binders (miBds), which can specifically recognize and eliminate cancer cells within a mere four to six weeks. Experiments then further confirmed the efficacy in eliminating cancer cells [8]. This achievement directly demonstrates AF's remarkable potential in developing novel therapeutic proteins, marking its advancement in steering protein design toward greater complexity and precision.

3.3. Analysis of the process of viral gene mutation

AF3 has achieved remarkable precision in predicting protein-nucleic acid interactions. By accurately predicting the binding structures between viral proteins and biomacromolecules, AF3 provides structural insights into the mechanisms by which viruses evade host humoral immunity. This capability enables a deeper understanding of genetic mutations.

Numerous viral life cycles, including those of SARS-CoV-2 and HIV, rely on specific interactions between viral proteins and nucleic acids. Contemporary research is investigating this through the three-dimensional structure of the spike protein from human coronavirus OC43. When comparing predicted results with experimentally determined structures, the GDT score of 83.1 indicates near-perfect structural equivalence. Similarly, the LDDT result of 83.0 at the atomic level further validates its accuracy [4]. These predictive outputs can directly guide molecular docking, elucidate functional mechanisms, and even inform rational drug design. They enable rapid assessment of how genetic mutations affect spike-antibody binding and identify mutation categories that disrupt hydrogen bonds, salt bridges, or hydrophobic interactions. This provides invaluable insights for deciphering viral immune evasion mechanisms and optimizing antibody drug development to mitigate the effects of viral mutations.

4. Distinct advantages of AlphaFold

4.1. Efficient prediction and scalability

The AF2 and AF3 technologies enable protein structure prediction within a timeframe ranging from minutes to over ten hours upon input of amino acid sequences, with processing durations varying

DOI: 10.54254/2753-8818/2026.PJ30488

according to sequence length. For instance, a 256-amino-acid sequence requires just 5 minutes, whereas a 2,500-amino-acid sequence necessitates 18 hours [3]. These technological advancements have markedly alleviated the time-consuming and labour-intensive nature of protein structure determination. The AF database has predicted over 200 million protein structures, covering nearly all known proteins, thereby substantially expanding the data foundation and application scope of structural bioinformatics.

4.2. Overcomes dependence on homologous sequences

One of the most profound contributions of AlphaFold (AF), particularly AF2, to structural biology is the provision of novel tools for predicting orphan sequences. These sequences, also termed orphan proteins, denote protein sequences that are challenging to retrieve from gene databases or possess only a limited number of homologous sequences. Traditional sequence alignment tools like BLAST cannot find enough high-quality and sufficiently similar sequences to construct multiple sequence alignments (MSAs), and early prediction models were severely limited by MSAs.

To mitigate AF2's suboptimal predictive performance under conditions of homologous sequence scarcity, researchers developed the EvoGen meta-generative model. This model has been validated to generate calibrated or synthetic homologous sequences, substantially enhancing AF2's performance in scenarios with limited data. Currently, AF2 can yield reliable predictions even with a single protein sequence, thereby markedly improving its accuracy in orphan sequence analysis [9]. This advancement substantially reduces AF2's reliance on traditional multiple sequence alignment (MSA) methods, thereby enhancing its utility in high-throughput research applications.

4.3. Broad applicability

AF3's most revolutionary capability resides in its breakthrough beyond protein-centric modelling, enabling the unified simulation of diverse biomolecules. By treating any input molecule as a collection of atoms for direct three-dimensional coordinate prediction, the model can output the Cartesian coordinates of all constituent atoms. By accurately predicting protein-DNA and RNA interactions, it can handle large-scale complexes comprising thousands of residues, such as ribosomes. Furthermore, AF3 accurately predicts various covalent modifications, including phosphorylation and glycosylation, while providing precise modelling for nucleic acids, small molecules, and ions [4].

5. Discussion

5.1. Challenges

5.1.1. Inability to capture dynamic conformational changes

Limited predictive power for structural flexibility is one of the primary barriers to AF's application in drug discovery. Although the current version of AF can generate multiple models, it remains constrained in its ability to capture the full dynamic trajectory of conformational transitions, as its training data is predominantly derived from static crystal structures [10]. A significant deviation was observed by Wei Lu et al. between the predicted ligand binding mode from rigid docking using the AF-generated apolipoprotein structure and the experimental co-crystal structure [6]. Thus, despite AF's high-accuracy models, neglecting the dynamic properties of proteins remains a fundamental challenge in predicting binding poses.

DOI: 10.54254/2753-8818/2026.PJ30488

5.1.2. Mutations and Post-Translational Modifications (PTMs)

When modelling protein behaviour under non-physiological or chemical stress conditions, AF exhibits a critical limitation: AF2 cannot capture how missense mutations alter the protein's energy landscape and consequently its final 3D conformation [11]. For instance, when mutations cause protein misfolding, AF2's predicted structures and pLDDT results show no significant differences compared to the wild type, indicating that its predictions depend more on evolutionary information than physical principles [12]. Furthermore, recent adversarial tests demonstrate that AF3 may generate non-physical structures (such as atomic overlaps and neglecting electrostatic repulsion), indicating its over-reliance on statistical patterns rather than a genuine grasp of physical principles [13].

Moreover, at present, AF still struggles to handle structural perturbations caused by non-standard chemical modifications such as phosphorylation and glycosylation [14]. Thus, when utilizing AF to support drug design or functional studies, experimental structural biology methods should be employed for validation.

5.1.3. Intermolecular interactions

Although AF technology has attained substantial accuracy, its precision in predicting critical ligand-binding events remains inadequate to support precision drug design. Through a systematic assessment, Xu et al. demonstrate that AF3-predicted structures of GPCR-peptide complexes commonly suffer from two issues: poor ligand heavy-atom RMSD and the presence of severe steric clashes. Notably, the model generates severe assembly errors in predictions of Class C GPCR dimers [14]. Furthermore, it is susceptible to hallucinations, frequently converting intrinsically disordered regions into false ordered structures [4].

Equally critically, non-physical atomic collisions, where predicted ligand atoms and receptor proteEqually critically, non-physical atomic collisions, in which predicted ligand and receptor protein atoms exhibit spatially unreasonable overlap, are frequently observed [10]. Additionally, predicted ligand-binding pockets may undergo shrinkage. These discrepancies may lead to misinterpretation of the protein's actual function [15].

5.2. Future directions

Contemporary AF models, particularly AF3, have exhibited significant advancements in predicting the structures of protein-protein complexes. Future refinements aimed at enhancing accuracy should focus increasingly on capturing the dynamic behaviour of biological systems. More precisely, a core developmental direction resides in predicting conformational changes and transient dynamic processes. Investigating the transitions between protein active and inactive states is critical, as it directly clarifies disease mechanisms and enables the rational design of allosteric drugs.

Furthermore, enhancing the accuracy of side-chain modelling, especially at key functional sites such as protein-protein interaction interfaces and ligand-binding pockets, is indispensable. Lastly, like many generative AI models, AF3 is prone to hallucinations, requiring stronger physicochemical constraints to ensure predicted structures comply with fundamental physical laws [16].

6. Conclusion

This study illustrates that AF has emerged as a revolutionary force in the field of medical research. It has been demonstrated that AF achieves high predictive accuracy, laying a robust foundation for

DOI: 10.54254/2753-8818/2026.PJ30488

protein design and precision medicine. It continues to serve as a fundamental tool for reliable protein structure determination. Despite this, for complex scenarios, including protein dynamics, complex assembly and ligand interaction, the prediction results of AF still need to be carefully validated and optimized with experimental data.

This study also comes with some limitations. The analysis was based exclusively on published literature and secondary data, rather than data obtained through empirical research utilizing AF. Consequently, it does not permit in-depth exploration of the alignment of the results of AF with data from diverse experimental methods. Future research could integrate empirical studies to further refine AF's applicability and explore potential optimization strategies.

References

[1] Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., ... & Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. Nature, 96(7873), 590-596.
[2] Isert, C., Atz, K., & Schneider, G. (2023). Structure-based drug design with geometric deep learning. Current Opinion in Structural Biology, 79, 102548.
[3] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. nature, 596(7873), 583-589.
[4] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., ... & Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493-500.
[5] Senior, A., Jumper, J., Hassabis, D., & Kohli, P. (2022, July 28). AlphaFold: Using AI for scientific discovery. DeepMind Blog. https: //deepmind.google/discover/blog/alphafold-using-ai-for-scientific-discovery/
[6] AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput.
[7] Yang, Z., Zeng, X., Zhao, Y., & Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduction and Targeted Therapy, 8(1), 115.
[8] Anand, N., & Achim, T. (2022). Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. arXiv: 2205.15019.
[9] Zhang, J., Liu, S., Chen, M., Chu, H., Wang, M., Wang, Z., ... & Gao, Y. Q. (2023). Unsupervisedly prompting AlphaFold2 for accurate few-shot protein structure prediction. Journal of Chemical Theory and Computation, 19(22), 8460-8471.
[10] He, X. H., Li, J. R., Shen, S. Y., & Xu, H. E. (2025). AlphaFold3 versus experimental structures: assessment of the accuracy in ligand-bound G protein-coupled receptors. Acta Pharmacologica Sinica, 46(4), 1111-1122.
[11] Lu, W., Zhang, J., Huang, W., Zhang, Z., Jia, X., Wang, Z., ... & Zheng, S. (2024). DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. Nature Communications, 15(1), 1071.
[12] Buel, G. R., & Walters, K. J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure?. Nature structural & molecular biology, 29(1), 1-2.
[13] Masters, M. R., Mahmoud, A. H., & Lill, M. A. (2025). Investigating whether deep learning models for co-folding learn the physics of protein-ligand interactions. Nature Communications, 16(1), 8854.
[14] HE, X., LI, J., & XU, H. (2025). Why AlphaFold cannot replace structural biology: An exploration of the accuracy of AI-based structure prediction. Science & Technology Review, 43(2), 14-21.
[15] Shen, S. Y., Li, J. R., Wang, Y. S., Li, S. N., Xu, H. E., & He, X. H. (2025). An update for AlphaFold3 versus experimental structures: assessing the precision of small molecule binding in GPCRs. Acta Pharmacologica Sinica, 1-10.
[16] Anishchenko, I., Pellock, S. J., Chidyausiku, T. M., Ramelot, T. A., Ovchinnikov, S., Hao, J., ... & Baker, D. (2021). De novo protein design by deep network hallucination. Nature, 600(7889), 547-552.