Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ37012

AlphaFold 3: Architecture and Applications in Protein–Ligand Complex Prediction

Kaichun Tang

Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China
1250030411@student.must.edu.mo

Abstract. The interaction between proteins and ligands is the core mechanism of biological activities and drug development. In traditional experimental methods, there are a series of problems such as long research cycles, high costs, and limited precision. In the rapid way of artificial intelligence (AI) era, AI tools have driven the progress of biological science. Therefore, this article focuses on elaborating on the new AI tool Alphafold to analyze the three-dimensional spatial structure of proteins. Especially Alphafold3(AF3), which adds diffusion algorithms, confidence assessment units, and the Pairformer module to the original AI model and simplifies the calculation steps, has achieved a revolutionary leap from single atom or protein homologous sequence pairing output to position pairing output between nucleic acids - proteins and peptide residues. It has realized the all-round design of protein-ligand complexes throughout the process. At the same time, this article systematically reviews the technical architecture and core algorithm modules of AlphaFold3, and looks forward to the application progress in the construction of protein-ligand complex structures, including the improvement in the confidence of pocket-type ligands paired with G protein-coupled receptors compared to AF2. The conclusion indicates that AF3 can be widely applied in biological fields such as life health and medical technology.

Keywords: AlphaFold3, protein–ligand complexes, deep learning, drug discovery

1. Introduction

Proteins are the main bearers of life. The realization of biological life functions relies on the dynamic recognition between molecules, and it is the interactome that reveals the intricate network of relationships among life forms. In the 21st century, targeted drugs have become an important direction in the development of new drugs. However, the key to targeted drug development lies in the difficulty of finding target structures and the ease of off-target ligand interactions. Therefore, clear and specific protein-ligand structures are particularly important. People use deep neural networks to accurately predict the distance potential energy between residues, thereby describing the three-dimensional shape of proteins [1, 2], which has become a mainstream method. Just the emergence of Alphafold has precisely promoted the development of this scientific goal.

This review will focus on the core principles and technical architecture of AlphaFold3, including its decoder, encoder, diffusion module, and Pairformer algorithm module for the adaptation of atomic-level input, and its application in practical cases of biomolecular conformational studies and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ37012

drug design, demonstrating the huge potential of AlphaFold3 in empowering protein-ligand complexes.

2. The historical development of AlphaFold

People could obviously know the bumpy road of AI since many people even famous scientist misunderstood the neutral network. Here is the long history starting in2018 ,. this system adopts the latest generative neural network architecture and uses the gradient descent algorithm, significantly simplifying the computational steps and winning the first place in the International Protein Structure Prediction Competition (CASP13) that year. However, this model may have experienced a significant decline in prediction accuracy when lacking homologous sequence alignment information, and may be limited by its modular design when dealing with more extensive structural features [3], making its application scope relatively limited.

Another turning point subsequently followed in2021.People could see DeepMind introduced an innovative method in 2021 for extracting MSA of proteins using an innovative approach, which can perform precise predictions even in the absence of homologous sequences, significantly promoting the development of bioinformatics. AlphaFold significantly improved the accuracy of structure prediction by integrating a new neural network architecture and training methods based on the evolutionary characteristics, physical properties, and geometric constraints of protein structures [3]. In the 14th CASP, AlphaFold2 demonstrated far superior prediction capabilities compared to other competitors, achieving atomic-level accuracy. Scientists successfully predicted over 200 million protein structures with its high-precision prediction ability, far exceeding the total of all 100,000 known proteins. Moreover, AlphaFold has been fully validated through experiments and has demonstrated its wide application value, applicable to molecular replacement studies and the interpretation of low-temperature electron maps [3]. Although AlphaFold2 still has shortcomings in predicting the three-dimensional structure of small molecule templates, multi-chain complexes, flexible regions, and dynamic processes, it has shown great applicability in structural biology.

AlphaFold has indeed stood at the base of the previous amazing master work, you could observe the achievement - integrating multiple protein algorithms, demonstrating the possibility of achieving precise docking for ligand proteins. From AlphaFold–Multimer revealing disordered proteins, to ESMFold achieving rapid predictions without multiple sequence alignment, to RoseTTAFold full-atom biological modeling. This marks that research has entered the era of full-atom modeling, capable of uniformly predicting and designing complex assemblies including proteins, nucleic acids, small molecules, metals, and their covalent modifications [3, 4], providing more powerful tools for structural biology and drug design.

DOI: 10.54254/2753-8818/2026.HZ37012

3. Breakthroughs in the technical principles of AlphaFold3 and demonstration of its core architecture

3.1. Overview of the alphaFold3 core framework

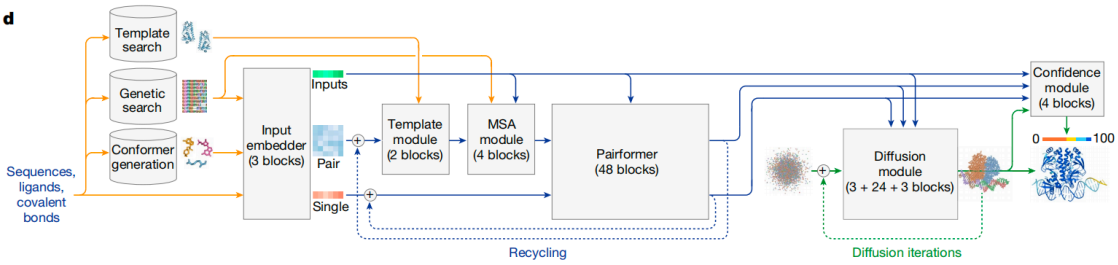


Figure 1. The overall architecture of AlphaFold3 [3]

The whole framework can be divided into about 3 parts, following are input and template base, algorithm part which proved efficiency a lot and last output and confidence part.AF3 (Figure 1) has undergone comprehensive updates in the input module and template library search and generation to be applicable to more small molecules and multi-chain complexes. In terms of the database, the model is trained on a large-scale biological molecule structure data to learn the key features of molecular interactions. Then, using a multi-scale modeling strategy, the 1D, 2D, and 3D multi-scale modeling is adopted to handle the fusion sequence and atomic sequence, process the chemical bond information between atoms, and incorporate stereoscopic chemical constraints such as chirality to improve overall accuracy. Additionally, the model introduces graph neural networks to handle the complex relationships between molecules, enhancing the prediction ability for protein complexes and nucleic acids in tree-like relationships between atoms and atoms, amino acid pairs, and proteins and proteins. Even to further improve robustness and generalization, AF3 utilizes cross-distillation technology, generating pseudo-label data through high-performance models for training, thereby alleviating the problem of insufficient data [3, 5].

The algorithm part is trying a nearly innovative way. For what is the most vital part —The input processing module realizes the unified representation transformation of heterogeneous biological molecule data. It breaks away from the traditional single protein sequence limitation, integrates multimodal chemical information of proteins, nucleic acids (DNA/RNA), and small molecule ligands, and extracts features from homologous template retrieval in the protein database (PDB) and co-evolution signals from multiple sequence alignments to complete the mapping of molecular chemical graphs to atomic-level initial coordinates and chemical bond topological structures.

The encoder module (Encoder Module) undertakes the task of modeling geometric space molecular interaction constraints. The template module injects prior spatial geometric constraints, and then the MSA module converts sequence alignment information into residue-interaction features, capturing long-range evolutionary constraints. The core algorithm Pairformer unit integrates the high-dimensional features of the MSA sequence and individual protein through a 48-layer iterative attention mechanism.

The core innovation of the diffusion module (Diffusion Module) (Figure 2) replaces the previous iterative structural module and models the atomic three-dimensional coordinates as a Gaussian noise distribution. During training, the same structure undergoes triangular attention mechanism transformation in different scales of local and global data, and the resulting feature values are used

DOI: 10.54254/2753-8818/2026.HZ37012

to add noise and diffuse to generate, and are supplemented by 20 steps of reinforcement learning to gradually form a clear image [3].

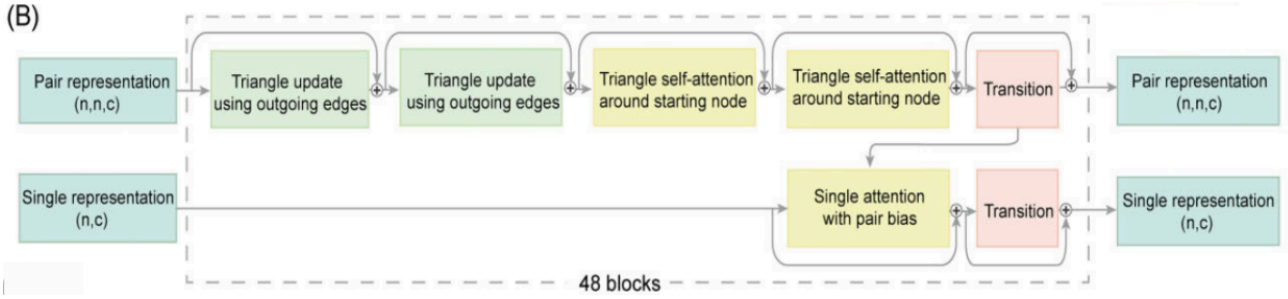


Figure 2. Working mechanism of the diffusion module in AlphaFold3 [3, 5]

3.2. Supporting experimental structure determination and accelerating protein structure prediction

The current mainstream experimental method for protein structure determination is in situ structural analysis using cryo-electron microscopy. Through the classification and distribution of two-dimensional spatial electron density, atomic sequences and amino acid sequences are constructed in three-dimensional space. However, the complexity of the experiment and the massive amount of data required often result in the need to spend several months to detect the structure of a protein and incur high costs. At this opportune time, AlphaFold, based on all approximately 10,000 known human proteins in the protein database, has already predicted over 200 million types [6]. And through pre-training to automatically match the corresponding algorithm from the obtained electron density and arrangement, and to reason independently, it can quickly construct the atomic model of the protein and has achieved an accuracy approaching the experimental level in a very short time.

Coincidentally, the revolutionary change of AlphaFold3 lies in the greatly enriched experimental objects. The objects have now expanded from the previously only predictable single-chain proteins to the current full biological molecule systems of proteins, nucleic acids, and small molecules. The algorithm has changed from the original RoseTTA design method to the RFdiffusion diffusion model. It can set up specific small molecule protein experiments for validation and activity, and has added a residue confidence update module. Its success rate has increased by three orders of magnitude compared to traditional methods [7]. Especially in the targeted drug design for micro-binding proteins, the strategy for neutralizing the target protein has shown excellent efficacy in animal models. And more evidence shows that AF3 has an accuracy of approximately 80% in the prediction of small molecules [8]. This is sufficient to demonstrate the significant role of AF3 in drug research.

4. Application cases and performance comparison analysis

AlphaFold3 is equipped with powerful AI prediction algorithms and deep learning models, and it shows broad application prospects in the field of life and health. It covers multiple purposes: whether it is drug design and discovery, exploration of disease mechanisms, or plant biological immune reagents, or innovative research and development of agricultural production and chemical nanomaterials, AlphaFold3 is fully empowering technological development [9].

DOI: 10.54254/2753-8818/2026.HZ37012

4.1. Analysis of typical application cases

Development of GPCR drugs G protein-coupled receptors (GPCRs) are one of the most important membrane protein targets in drug development, involved in many physiological processes. However, due to their inherent flexibility and the ability to interact with different ligands, many structures remain unresolved. Traditional methods often suffer from limitations in modeling GPCR-molecule complexes due to the flexibility and conformational diversity of membrane proteins. Research results show that although there are still significant differences from experimental results in ligand binding, AlphaFold3 can significantly reduce prediction errors when predicting GPCR-ligand complexes compared to AlphaFold2, especially in terms of Cα-RMSD data, demonstrating better performance [10]. For example, in the modeling of the β2-adrenergic receptor (β2AR) complex with a small molecule agonist, AlphaFold3 successfully captured the key conformational changes of the ligand binding pocket, and the prediction results have reached atomic-level accuracy.

4.2. Performance comparison with traditional methods

Based on the viewpoints of this paper, AF3 can predict the protein-ligand complex structure from scratch in the analysis of new target structures, providing a structural basis for subsequent drug design. It can support the prediction of various biological molecule complexes such as protein-protein, protein-nucleic acid, protein-peptide, and covalent ligands in the prediction of multiple types of complexes, covering large molecular interaction scenarios that are difficult for traditional docking to handle. In the deep analysis of the binding mode, it can target the experimental verification of active molecules, enabling the prediction of the binding posture and pocket conformation of the experimental object in three-dimensional space. For example, AlphaFold3 performs excellently in the receptor framework prediction of GPCR-small molecule ligand complex experimental structures, with an average root mean square deviation (RMSD) of Cα of only 0.98 Å [9]. In these aspects, AF3 has significant advantages.

However, AF3 is not perfect and also contains many challenges. This includes the fact that traditional molecular docking can quickly narrow the range of candidate molecules in the initial screening and enrichment of millions of compounds in large-scale screening, which is still the mainstream method. It also includes the extremely high demand for computing power, as protein three-dimensional modeling and analysis often require multiple GPUs and a large amount of computing power. It also includes many issues such as its own weak explanatory power. But in any case, AF3 is undoubtedly a phenomenal work. I believe that scientific research should combine the advantages of both and complete it step by step, which will surely have great achievements.

5. Conclusion

This article traces the advancing and explosive history of AlphaFold and elaborates on the core algorithms and structure of AF3. The author think AF3 is totally a new template of biology working method since AF3 has broken away from AlphaFold2's strong reliance on multiple sequence alignment at the underlying architecture level, constructed an end-to-end generation framework centered on the diffusion Transformer, and achieved unified modeling of different types of biological molecules. In the specific application at the biological level, it demonstrates a comprehensive breakthrough from "copying nature" to "customized as needed" in protein structure design, outlining a grand blueprint for the protein-ligand field. Finally, by comparing AF3 with traditional molecular docking, it highlights the powerful functional attributes of AF3, providing

DOI: 10.54254/2753-8818/2026.HZ37012

revolutionary tools for structural biology and drug development. However, it points out that there are still limitations in its physical interpretation, initial protein selection through enrichment, and computational power requirements, which need to be complemented by traditional methods such as molecular dynamics simulation.

In summary, the emergence of AlphaFold3 marks a new stage for structural biology. Despite the challenges, with the optimization of core algorithms, the continuous enrichment of database templates, and the breakthroughs in the boundaries of biological structure, future structure prediction tools will surely become more complete. The author believes that with the continuous progress of technology, humanity will enjoy the benefits brought by these achievements in the near future, promoting the further development of life science and medical research, and ultimately benefiting the entire human society.

References

[1] Machado, M., Bautista-Hérnandez, I., Gómez-García, R., Silva, S., & Costa, E. M. (2025). Bioactive food proteins: Bridging nutritional and functional benefits with sustainable protein sources. *Foods, 14*(17), 3035. https://doi.org/10.3390/foods14173035

[2] Akif Uzman. (2003). Molecular biology of the cell (4th ed.): Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Biochemistry and Molecular Biology Education, 31*(4), 212-214. https://doi.org/10.1002/bmb.2003.494031049999

[3] Abramson, J., Adler, J., Dunger, J., Evans, R., Pritzel, A., Green, T., … & Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature, 630*(8016), 493-500.

[4] Bertoline, L. M. F., Lima, A. N., Krieger, J. E., & Teixeira, S. K. (2023). Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics, 3*:1120370. https://doi.org/10.3389/fbinf.2023.1120370

[5] 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. https://doi.org/10.1038/s41586-021-03819-2

[6] Bertoni, M., Varadi, M., Magana Rodriguez, U., Velankar, S., & Steinegger, M. (2024). AlphaFold protein structure database in 2024: Providing structure coverage for over 200 million protein sequences. *Nucleic Acids Research, 52*(D1), D374-D382. https://doi.org/10.1093/nar/gkad1032

[7] Krokidis, M. G., Koumadorakis, D. E., Lazaros, K., Ivantsik, O., Exarchos, T. P., Vrahatis, A. G., Kotsiantis, S., & Vlamos, P. (2025). AlphaFold3: An overview of applications and performance insights. *International Journal of Molecular Sciences, 26*(8), 3671. https://doi.org/10.3390/ijms26083671

[8] Gong, W. B. (2024). The major breakthroughs in AlphaFold structural prediction and their impact and challenges on protein research. *Advances in Biochemistry and Biophysics, 51*(12), 3073-3083.

[9] Wang, S. Y., & Li, M. (2025). Application scenarios and ethical risk governance of AlphaFold3 in the field of life sciences and health. *Medicine & Philosophy, 46*(19), 32-36.

[10] Shen, S. Y., Li, J. R., Wang, Y. S., et al. (2025). An update for AlphaFold3 versus experimental structures: Assessing the precision of small molecule binding in GPCRs. *Acta Pharmacologica Sinica, 46*, 3355–3364. https://doi.org/10.1038/s41401-025-01617-4