ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

AlphaFold: Evolution, Applications, and Challenges

Ziqin He[1,a], Hexin Tan[2,b,*]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China
[2]Department Chinese Medicine Authentication, College of Pharmacy, Naval Medical University (Second Military Medical University), Shanghai, China
[a]hzq_0822@163.com, [b]hexintan@163.com
**Corresponding author*

Abstract: *Protein structure prediction represents a fundamental challenge in biochemistry. Conventional experimental approaches remain constrained by complex sample preparation requirements, substantial costs, and limited capacity for capturing dynamic structural information. These methodological limitations in experimental techniques have motivated the advancement of AI-assisted computational approaches. This review systematically examines the technological evolution of the AlphaFold system across three successive generations, from AlphaFold1's integration of deep learning with evolutionary covariance analysis, to AlphaFold2's revolutionary attention-based Evoformer architecture, culminating in AlphaFold3's diffusion model for multi-molecular complex prediction. Key advancements encompass the diversification of input modalities, expansion of predictive scope, and optimization of computational efficiency. Furthermore, we critically evaluate core application domains spanning fundamental biological research, pharmaceutical discovery and design, biotechnology and synthetic biology applications, and the structural bioinformatics tool ecosystem. Finally, we delineate persistent technical challenges within the field.*

Keywords: *AlphaFold, Protein Structure Prediction, Artificial Intelligence*

1. Introduction

Protein structure prediction remains one of the most significant challenges in biochemistry. Established experimental techniques for structure determination—namely X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) spectroscopy—are persistently confronted by a triad of limitations: demanding sample preparation requirements, restricted access to dynamic structural information, and high costs coupled with low throughput. These constraints have driven the advancement of integrated methodological approaches and artificial intelligence (AI)-assisted methodologies. Recent breakthrough advances in AI have profoundly transformed this field, enabling high-accuracy prediction of protein structures directly from amino acid sequences. This review comprehensively examines the technological evolution of the AlphaFold system from its initial iteration to the third generation. Furthermore, it explores the system's extensive applications across diverse domains, including fundamental biological research, drug discovery and design, biotechnology and synthetic biology, and the structural bioinformatics tool ecosystem. The review demonstrates how AI is fostering a new paradigm of "predictive and experimental synergy" within structural biology. Current limitations and potential future research directions are also discussed.

2. AlphaFold1

AlphaFold1 [1], introduced by DeepMind in 2018, represents the inaugural deep learning-based model for protein structure prediction. Its core innovation lies in the integration of deep learning with evolutionary covariance information. The model initially employs deep residual networks (ResNets) to analyze multiple sequence alignment (MSA) data, effectively extracting co-evolutionary signals between residues. Subsequently, distance geometry methods are applied to convert the predicted distributions of inter-residue distances and dihedral angle probabilities into three-dimensional spatial constraints. Finally, a loss function incorporating physicochemical constraints (e.g., bond lengths, bond angles, van der Waals forces) is optimized to ensure the biological plausibility of the predicted structure. Compared to traditional computational approaches, such as fragment assembly or homology modeling, AlphaFold1's

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

innovations are manifested in three key aspects: (1) It demonstrated for the first time that deep learning can substantially enhance the accuracy of co-evolutionary analysis; (2) It achieved robust transformation from discrete constraints to continuous structures through distance geometry; (3) It explicitly incorporated physical rules within the machine learning framework. The following section provides a detailed analysis of its technical architecture, with Figure 1 offering a visual overview of its core components.

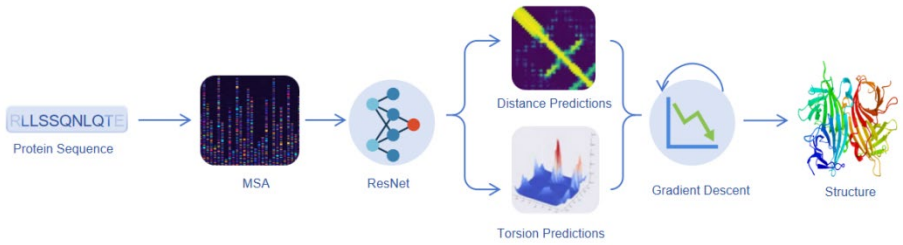

Figure 1 Simplified Architecture Diagram of the AlphaFold1 Algorithm.

2.1 Deep integration of evolutionary information

AlphaFold1 innovatively leveraged co-evolutionary information embedded within protein sequences. For MSA enhancement, the HHblits tool was employed to search the Uniclust30 database for homologous sequences. Utilizing a stringent E-value threshold of <0.001 and three iterative search cycles,an MSA matrix comprising over 10,000 homologous sequences was generated, thereby providing a rich dataset for subsequent co-evolutionary signal extraction. Subsequently, the GREMLIN algorithm was applied to compute residue-pair coupling scores. Pseudo-likelihood maximization optimization was also integrated, effectively mitigating overfitting issues associated with limited datasets and enhancing the accuracy of signal extraction. Regarding feature engineering, a 44-channel feature matrix was constructed, incorporating 20 channels for amino acid types, 22 channels for co-evolutionary/conservation features, and 2 channels for positional encodings. Auxiliary features, including sequence entropy and solvent accessibility, were incorporated to enrich the dimensionality of the input features, thereby providing comprehensive information for model learning.

2.2 Geometry-constrained deep learning

AlphaFold1 modified the conventional residual network (ResNet) architecture, implementing a 64-layer residual block design. Each block incorporated 3×3 convolutions, Batch Normalization (BatchNorm), and Rectified Linear Unit (ReLU) activation functions. Skip connections were utilized to mitigate the vanishing gradient problem, enabling deeper network learning. The output layer predicted 64-bin distance distributions (spanning 0–20 Å) and φ/ψ dihedral angles, providing essential geometric information for protein structure construction. A multi-task learning strategy was additionally implemented, designating distance distribution classification as the primary task and dihedral angle regression as the auxiliary task. This strategy facilitated the simultaneous optimization of predictions across different aspects, thereby enhancing overall performance. To ensure biophysical plausibility, the model integrated physicochemical constraints into the training process. High-confidence distance predictions were selected to construct the loss function, which incorporated additional terms derived from the Rosetta force field, including bond length and bond angle constraints. This systematic integration of physical principles guaranteed that predicted structures adhered to fundamental biomolecular rules.

2.3 Structural Optimization Algorithm

AlphaFold1 implemented a sophisticated structure optimization framework comprising two principal components: an advanced gradient descent strategy and a robust structure reliability assessment protocol. The optimization procedure initiated with conformation generation through sampling from a Gaussian random distribution, followed by refinement utilizing the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, which exhibits significantly enhanced convergence characteristics compared to conventional gradient descent approaches. The optimization process incorporated a progressive constraint weighting scheme, with weights systematically increased from 0.1 to 0.5 during the refinement process to gradually enforce physical constraints, while van der Waals repulsion terms

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

were explicitly included to prevent atomic steric clashes.

For quantitative evaluation of structural accuracy, the local Distance Difference Test (lDDT) metric was employed, which assesses the agreement between predicted and reference structures through comprehensive comparison of interatomic distances. Specifically, the metric calculates distance deviations for all Cα atom pairs within a 15Å threshold, with higher computed scores indicating superior structural accuracy.

2.4 The CASP13 Achievements and Limitations of AlphaFold1

Building upon these innovations, AlphaFold1 achieved groundbreaking success in the 13th Critical Assessment of Structure Prediction (CASP) competition. Its successful integration of deep learning with co-evolutionary analysis demonstrated, for the first time, that neural networks could surpass traditional statistical methods (such as GREMLIN). Furthermore, the use of probabilistic distance distributions mitigated discretization errors, while the integration of physicochemical constraints ensured the plausibility of predicted structures. In the CASP13 Free Modeling category, AlphaFold1 demonstrated a substantial advantage, with key metrics outperforming traditional methods, as detailed in Table 1.

Table 1: AlphaFold1 vs Traditional Methods (CASP13).

Metric	AlphaFold1	Traditional Methods (Best in CASP13)
GDT_TS	68.5	42.9
Long-Range Contact Accuracy (β-sheet)	72%	41%
RMSD (Moderate-Difficulty Targets)	2.8 Å	5.6 Å

Despite its groundbreaking advancements in protein structure prediction, AlphaFold1 exhibits several critical limitations. The model's performance exhibits a strong dependence on high-quality multiple sequence alignments (MSAs), being significantly constrained by the quantity and quality of available homologous sequences. Consequently, prediction accuracy diminishes for proteins with low sequence homology—such as orphan proteins or those with unique evolutionary histories—due to insufficient co-evolutionary signals. Furthermore, AlphaFold1 is restricted to predicting static structures; it cannot simulate conformational dynamics under physiological conditions (e.g., allosteric effects, folding intermediates) or model interactions with ligands. The approach also incurs high computational costs, requiring initial prediction of a distance matrix followed by iterative optimization to generate the 3D structure, resulting in a complex and time-consuming process that becomes particularly pronounced for longer proteins. Additionally, predicting multi-chain complexes presents a challenge, as the model was designed specifically for single-chain proteins and cannot readily predict protein-protein complexes or multimeric structures. Finally, limitations exist in modeling long-range interactions, leading to increased prediction errors for distant residue pairs (>20 Å), which can propagate to cause global topological inaccuracies.

3. Alphafold2

AlphaFold2 [2] achieved a revolutionary breakthrough in protein structure prediction. Its principal innovation resides in the development of the Evoformer module, an architecture based on attention mechanisms. This module dynamically integrates information from MSAs with three-dimensional structural features through an SE(3)-equivariant Transformer framework. Employing an end-to-end training paradigm, AlphaFold2 directly predicts atomic coordinates. A dedicated Structure Module facilitates iterative refinement of the predicted structure, incorporating template information through specialized processing channels to enhance predictive accuracy. The core architecture comprises two essential components: the Evoformer module and the Structure Module. Their collaborative interplay—including the integration of MSA, template information, and iterative refinement—is visually summarized in Figure 2, enabling high-accuracy prediction of protein tertiary structures.

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

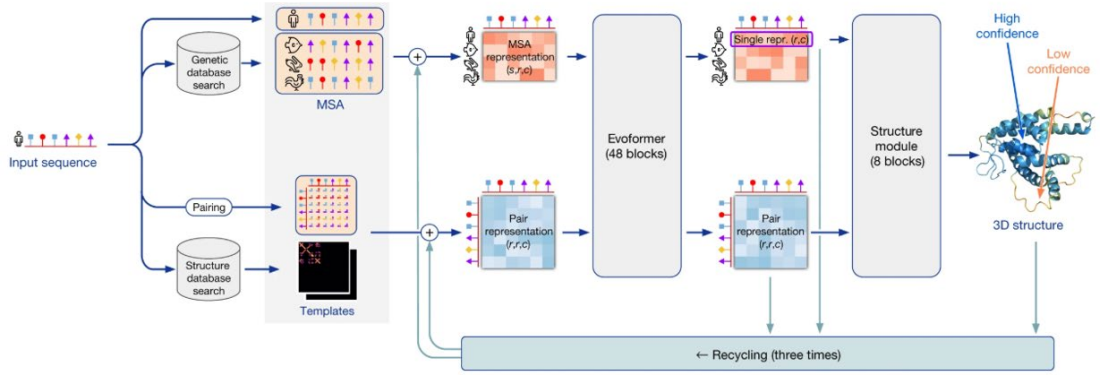

Figure 2: Schematic of AlphaFold2's model architecture [2].

3.1 Evoformer

Serving as the encoder module within AlphaFold2, Evoformer extracts high-order relational features from protein sequences. It processes two primary inputs: the MSA of the target protein and its homologous sequences, and the pairwise representation characterizing amino acid interactions. Evoformer utilizes a deep network consisting of 48 repetitive blocks, incorporating three key computational components: axial attention, outer product mean, and triangular attention. Their functional interplay is illustrated in Figure 3.

(1) An axial attention mechanism that sequentially updates MSA and pairwise representations to reduce computational complexity;

(2) An outer product mean block that extracts pairwise information from MSA to reciprocally update pairwise representations;

(3) A triangular attention operator that incorporates geometric constraints from third-party residues when updating inter-residue distances.

This architecture captures long-range interactions, co-evolutionary relationships, and implicit geometric constraints between residues. The module outputs refined MSA features and enhanced pairwise representations containing high-precision distance distributions and dihedral angles, thereby providing contextual information for the Structure Module.

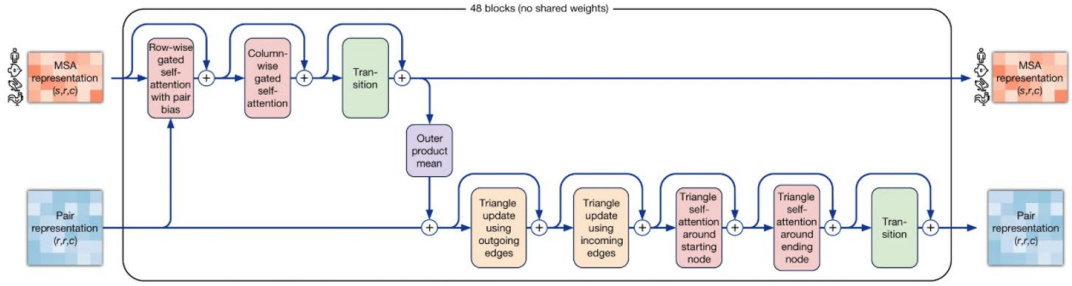

Figure 3: Schematic diagram of the Evoformer module [2].

3.2 Structure Module

Functioning as the decoder module within AlphaFold2, the Structure Module transforms the abstract features output by the Evoformer into precise three-dimensional atomic coordinates. This transformation is achieved via a network comprising eight repetitive blocks based on invariant point attention (IPA). IPA ensures translational and rotational invariance by representing amino acid coordinates as centroid-relative vectors and computing attention using relative positional vectors, enabling the model to focus on local geometric features. During training, a "recycling" mechanism feeds intermediate coordinate predictions from the module's output back into the Evoformer, establishing a recurrent cycle. This cycle approximates the dynamic process of protein folding and facilitates error correction. Concurrently, the Frame-Aligned Point Error (FAPE) loss function emphasizes local structural accuracy. Geometric

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

constraints, including bond lengths and bond angles, are incorporated to enforce structural and chemical plausibility. The module ultimately outputs a reliable structure containing the 3D coordinates of all atoms and per-residue confidence estimates (pLDDT values), with its internal mechanism (including IPA and recycling) detailed in Figure 4 .

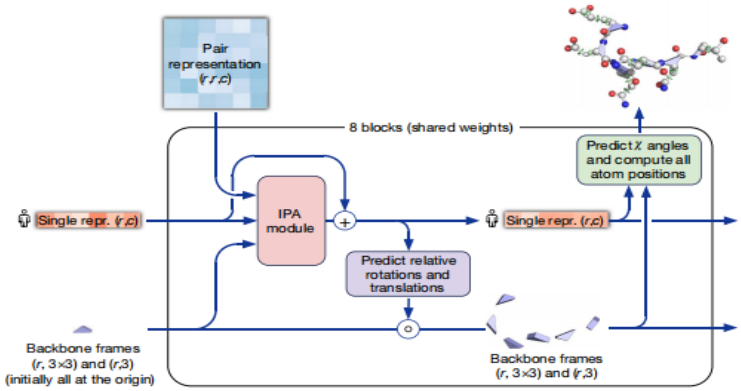

Figure 4: Schematic diagram of the Structure module [2].

3.3 AlphaFold2's Synergistic Architecture and CASP14 Breakthrough

The synergistic logic between AlphaFold2's two principal modules is realized through an iterative refinement mechanism and biologically inspired design. The Evoformer and Structure modules engage in multiple cycles of interaction—progressing from capturing global evolutionary signals to generate coarse backbone scaffolds in early stages, to cooperatively optimizing local structural details such as side-chain conformations and hydrogen-bond networks in later stages. This process accomplishes a structural refinement trajectory from coarse-grained to fine-grained representations. Within this framework, Evoformer's axial and triangular attention mechanisms simulate patterns of amino acid co-evolution, while the Structure module's IPA and recycling mechanism incorporate principles analogous to energy minimization in molecular dynamics, thereby integrating evolutionary patterns with physicochemical constraints.

The architectural innovation is further demonstrated by its fully differentiable, end-to-end learning process, overcoming the pipeline limitations inherent in AlphaFold1. Notably, Evoformer extends the Transformer self-attention mechanism from one-dimensional sequences to two-dimensional matrices to capture higher-order interdependencies. Concurrently, the integration of a "noisy student" self-distillation strategy achieves synergistic advancement through algorithmic innovation and data augmentation, establishing a benchmark for subsequent biological sequence modeling.

These collective technological innovations underpin AlphaFold2's exceptional performance. AlphaFold2 demonstrated a revolutionary breakthrough in CASP14, achieving atomic-level prediction accuracy. Of the 53 targets evaluated, over 90% of its predictions exhibited close proximity to experimentally determined structures, with an overall median GDT_TS of 92.4. For 14 challenging free-modeling (FM) targets, the median GDT_TS reached 88.5, particularly showcasing high congruence with experimental results in predicting complex loop regions. The model outperformed all other competing methods by a substantial margin, with its prediction accuracy even surpassing that of certain experimental techniques. Nature has recognized this achievement as "one of the most significant breakthroughs in 21st-century biology," fundamentally transforming the landscape of protein structure research.

4. AlphaFold3

AlphaFold3[3] represents the most recent iteration of the protein structure prediction artificial intelligence model developed by DeepMind. This system extends beyond the high-accuracy prediction capabilities characteristic of the AlphaFold series to encompass interactions between proteins and other biomolecules. Consequently, it serves as a robust tool for fields including drug discovery, gene therapy, and synthetic biology. The advent of AlphaFold3 signifies substantial progress toward the fundamental objective of predicting biological processes directly from sequence data. Building upon the AlphaFold2 framework, AlphaFold3 achieves multidimensional technological advancements. Its expanded network

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

architecture—including modules for multi-molecular complex prediction—is visualized in Figure 5.

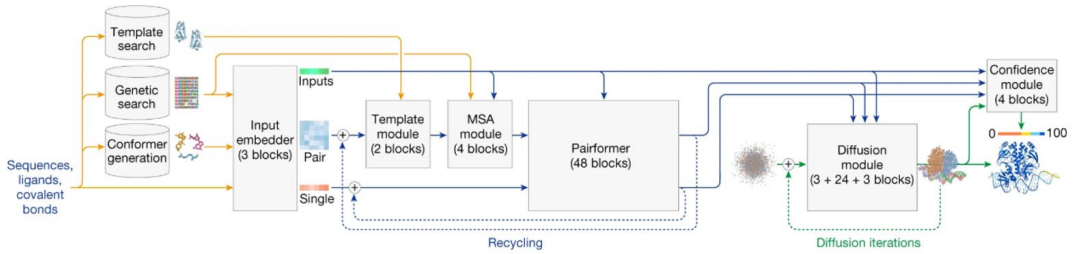

Figure 5: Schematic diagram of AlphaFold 3's network architecture [3].

4.1 Pairformer

AlphaFold3 replaces AlphaFold2's Evoformer module with the Pairformer module, fundamentally transforming the processing of MSAs. This module eliminates explicit maintenance of MSA representations, retaining only pairwise representations. Through architectural streamlining, computational complexity is substantially reduced—manifested by decreasing the number of MSA blocks from 48 to 4. Evolutionary features are extracted via weighted averaging, thereby circumventing the high dependency on homologous sequence count inherent in the conventional Evoformer.

Concurrently, Pairformer directly focuses on modeling spatial relationships between residues, eliminating complex operations such as the outer product mean. It enhances learning of interaction patterns involving non-protein molecules (e.g., nucleic acids, ligands) through refined attention mechanisms. Furthermore, unaligned MSA sequences employ a more compact representation, increasing the number of processable homologous sequences per chain by approximately threefold. This significantly broadens the scope of captured evolutionary information, particularly enhancing predictive capabilities for complex systems such as protein complexes.

These modifications—including streamlined MSA processing and enhanced attention mechanisms—simultaneously simplify the model architecture and improve its adaptability to sparse evolutionary data (Figure 6), establishing a foundation for multimodal molecular modeling.

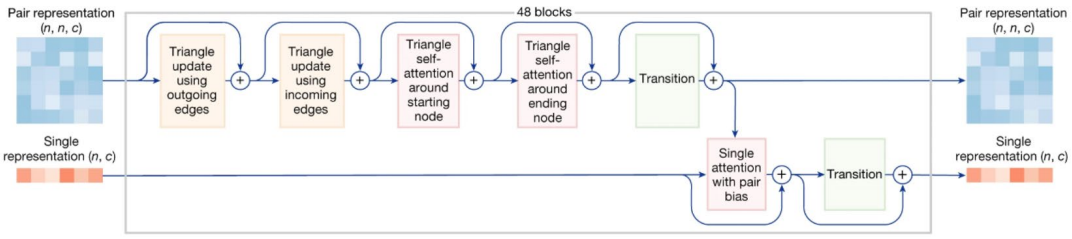

Figure 6: Schematic diagram of the Pairformer module [3].

4.2 Diffusion Model

Tables must appear inside the designated margins.AlphaFold3 departs from AlphaFold2's iterative refinement paradigm based on invariant point attention and physical constraints, introducing instead an all-atom diffusion generative model. This innovation constitutes a fundamental paradigm shift in structure prediction. The model operates through a forward diffusion process that incrementally perturbs ground-truth atomic coordinates with Gaussian noise, progressively degrading the molecular structure. Subsequently, a reverse denoising process trains the model to recover atomic coordinates from noise. Critically, minimizing the mean squared error (MSE) alone suffices for the model to learn inherent geometric constraints, eliminating the need for predefined explicit physical rules governing bond lengths, bond angles, or similar parameters.

This architecture inherently captures physicochemical principles, such as chemical bonding and van der Waals forces, while enabling flexible handling of non-standard chemical components, including metal ions and modified residues. By dynamically modulating noise intensity, the model progressively refines structures across multiple scales: early iterations prioritize coarse-grained construction of global

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

topology (e.g., protein folding scaffolds), while later stages optimize local details such as side-chain conformations and hydrogen-bond networks. This approach eliminates reliance on post-processing steps. The diffusion model's structure—incorporating noise modulation and multi-scale refinement (Figure 7)—enables genuine end-to-end generation of full atomic coordinates directly from sequence data, significantly enhancing predictive accuracy for complex conformations and dynamic regions.

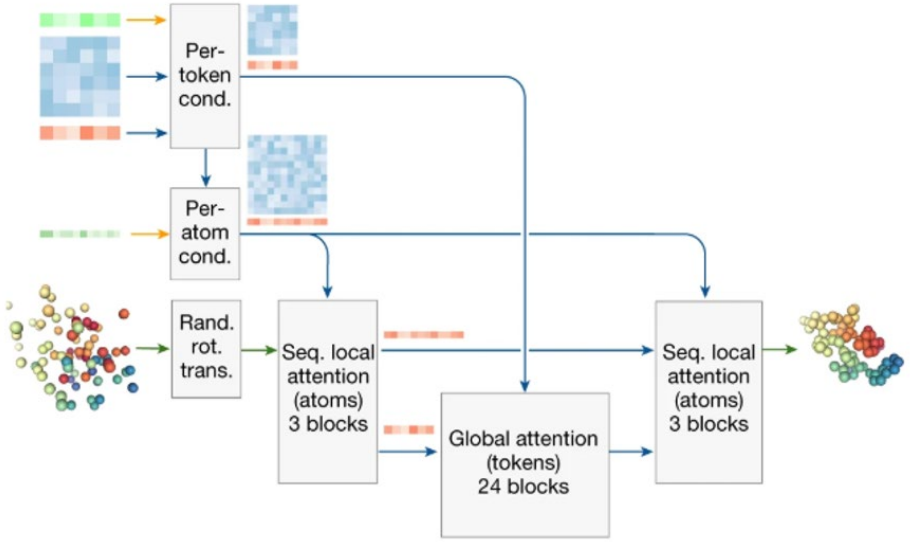

Figure 7: Schematic diagram of the diffusion module [3].

4.3 Atomic-Level Unified Representation Framework

AlphaFold3 represents a significant departure from AlphaFold2's rigid, amino acid-centric structural representation framework. Instead, it adopts an atom-centered modeling paradigm where each atom is treated as a discrete entity. This model directly predicts the three-dimensional coordinates of individual atoms, thereby eliminating the constraints imposed by rigid backbone assumptions and side-chain parameterization. Consequently, it naturally captures the relative positions and dynamic interactions between atoms (such as hydrogen bonding and hydrophobic effects) through attention mechanisms, enabling the precise characterization of molecular chemical structure details.

Furthermore, AlphaFold3 extends the token type system. Beyond the 20 standard amino acids, it incorporates eight distinct nucleic acid tokens, ligand atom tokens, and modified residue tokens. Each token type encodes specific chemical properties, including charge and hydrophobicity. This enhanced representation enables the model to differentiate between diverse molecular types—such as proteins, DNA, RNA, and ligands—and facilitates the unified modeling of various interaction modes (protein-protein, protein-nucleic acid, protein-ligand) via an improved geometry-informed attention mechanism.

This novel framework overcomes the limitations inherent in traditional methods for modeling non-protein molecules. It achieves generalized predictive capabilities for complex biomolecular complexes within a single network architecture, thereby establishing itself as a powerful framework for investigating fundamental biological processes such as post-translational modifications and molecular binding mechanisms.

5. AlphaFold's Evolution

Examining the technological evolution of AlphaFold reveals three defining characteristics: diversification of input information, expansion of predictive scope, and enhanced biophysical plausibility within the algorithmic architecture. The progression from an initial version reliant solely on single protein sequences, to a breakthrough iteration incorporating evolutionary biology information, and subsequently to the current third-generation model capable of handling multi-component biomolecular systems, corresponds directly to the expansion of structural biology's research scope. As illustrated in Table 2, which systematically compares key parameters across the three generations, these advancements manifest not only as incremental improvements in prediction accuracy but also as a qualitative shift in

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

predictive capability—from isolated proteins to the molecular interaction networks fundamental to biological processes. It is noteworthy that this transformative progress has been consistently accompanied by innovations in algorithmic paradigms: from residual networks incorporating geometric constraints, to the Evoformer integrating attention mechanisms, and most recently to the introduction of diffusion models. Breakthroughs in computational methodology have continuously advanced the boundaries of predictive capacity.

Table 2: Evolution of AlphaFold Models.

Parameter	AlphaFold1	AlphaFold2	AlphaFold3
Input	Protein sequence + MSA	Protein sequence + MSA	Protein sequence + MSA + other biomolecules
Output	Single-chain tertiary structure	Single-chain/homomeric structures	Multi-molecular complexes
Computational Efficiency	Medium	High	Optimized
Core Algorithm	ResNet + Geometric constraints	Evoformer + Structure module	Diffusion model + joint embedding representation
Confidence Metrics	IDDT	pLDDT + PAE	pLDDT + PAE + IPC
Limitations	No complexes	Ineffective for non-protein components	Still imperfect for modified residues
References	[1]	[2]	[3]

6. Core Applications of AlphaFold

6.1 Basic Biological Research

In the context of human proteome characterization, Tunyasuvunakool et al. [4] employed AlphaFold to predict the human proteome, thereby addressing significant knowledge gaps in experimental data, particularly for targets such as orphan proteins that lack established characterization methods. Sommer et al. [5] conducted structure-guided transcriptomic analysis to investigate functional differences among isoforms, offering novel structural insights into biological functional diversity. Regarding evolutionary and conservation studies, Bordin et al. [6] performed comparative analyses across 21 model organisms using AlphaFold, elucidating evolutionary patterns in protein structures and enhancing our understanding of biological evolution.Within virology research, Hu et al. [7] integrated AlphaFold2 predictions with crystallographic experiments, leading to the identification of a novel fold in rotavirus spike protein. Nomburg et al. [8] leveraged these advancements to discover ancient immune evasion mechanisms in viruses and phages, providing novel conceptual frameworks for virology and immunology. In the field of membrane protein research, Xiao et al. [9] utilized AlphaFold2 to predict the conformational states of 69 mutants of E. coli MFS transporters, elucidating the structural basis of their transport mechanisms. For the study of protein dynamics, Wayment-Steele et al. [10] developed AF-Cluster, which successfully predicted the multiple conformational states of the denatured protein KaiB. Collectively, these studies address critical gaps in structural biology data while providing novel perspectives for understanding the functional diversity of proteins.

6.2 Drug Discovery and Design

In the field of target identification, the AlphaFold Database (AFDB) [11] provides millions of protein structures through AlphaFold, significantly accelerating virtual screening workflows and enabling efficient prioritization of potential drug targets. The HProteome-BSite tool [12] facilitates precise drug targeting by predicting protein binding sites, thereby supporting the development of highly selective therapeutics. Regarding molecular docking optimization, Wong et al. [13] demonstrated accelerated antibiotic discovery through AlphaFold-predicted target structures, substantially reducing development timelines.Within antiviral drug development, Yang et al. [14] provided structural insights into SARS-CoV-2 variant proteins, delivering mechanistic evidence for mutational effects. Notably, Zeng et al. [15] designed a hemagglutinin stem vaccine (B60-Stem-8070) using AlphaFold2, which demonstrated superior immunogenicity compared to conventional antigens. Concurrently, Ibrahim et al. [16] elucidated structural features of the ATG8/LC3 family, informing rational design strategies for proteolysis-targeting chimeras (PROTACs). Collectively, these applications streamline drug development pipelines, enhance design precision, and advance personalized medicine approaches.

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

6.3 Biotechnology and Synthetic Biology

In enzyme engineering, AlphaFold-predicted protein structures enable targeted modifications of enzymatic active sites, thereby facilitating optimization of industrially relevant properties such as thermostability. This enhancement improves functional utility and operational efficiency in industrial biocatalytic processes. For instance, Wayment-Steele et al. [17] identified novel conformational states in the Mycobacterium tuberculosis oxidoreductase Mpt53, providing structural foundations for enzymatic enhancement.Within de novo protein design, integrating AlphaFold with computational tools (e.g., Rosetta) supports the engineering of proteins with novel functional capabilities, expanding their biomedical and biocatalytic applications. Jendrusch et al. [18] developed the AlphaDesign framework, which leverages AlphaFold's predictive capacity to design proteins with predetermined structures and functions. Complementarily, Goverde et al. [19] implemented neural network inversion of AlphaFold, establishing a methodology for direct amino acid sequence generation from target structures, circumventing iterative sequence optimization.

6.4 Structural Bioinformatics Tool Ecosystem

Regarding databases and resources, the AFDB employs an open-access model, providing comprehensive protein structural data to the global research community. This initiative significantly facilitates academic collaboration and accelerates scientific progress. Viral protein structural libraries continuously expand with AlphaFold integration, extending their applicability and establishing critical infrastructure for virology research.Among auxiliary tools, Foldseek [20] serves as a rapid structural search tool that outperforms conventional methods (e.g., DALI) in computational efficiency for protein structural alignment and retrieval. Furthermore, Foldseek cluster [21] enables identification of uncharacterized structural families by analyzing unannotated clusters within its 2.3 million groupings, thereby advancing discovery in structural bioinformatics.Additionally, Ma et al. [22] demonstrated enhanced functional annotation accuracy using AlphaFold-predicted structures, while Hu et al. [23] optimized the Evoformer module for functional prediction applications. The integration of these resources not only expedites global research collaboration but also provides foundational support for standardized and automated methodologies in structural biology.

7. Challenges and Future Directions

While the AlphaFold series has achieved remarkable success in protein structure prediction, its development encounters persistent challenges. These include significant difficulties in predicting complex protein-protein interactions, limited capability for modeling dynamic processes such as conformational changes and folding pathways, substantial computational resource requirements with associated time inefficiencies, and the necessity for experimental validation coupled with interdisciplinary interpretation of predictions. Furthermore, acquisition and annotation of high-quality structural data—particularly for specialized protein classes—presents ongoing obstacles.Future development will prioritize: (i) enhancing prediction accuracy for protein interactions with diverse molecular partners; (ii) advancing protein dynamics simulations through hybrid deep learning/physical modeling approaches; (iii) facilitating deeper integration with experimental methods (e.g., X-ray crystallography) to establish a computational-experimental synergy; (iv) expanding applications across drug discovery, agricultural science, and materials research; and (v) improving computational efficiency and technical accessibility via architectural optimization and user-friendly tool development.

References

[1] Senior, Andrew W., et al. "Improved protein structure prediction using potentials from deep learning." Nature 577.7792 (2020): 706-710.
[2] Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." nature 596.7873 (2021): 583-589.
[3] Abramson, Josh, et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3." Nature 630.8016 (2024): 493-500.
[4] Tunyasuvunakool, Kathryn, et al. "Highly accurate protein structure prediction for the human proteome." Nature 596.7873 (2021): 590-596.
[5] Sommer, Markus J., et al. "Structure-guided isoform identification for the human transcriptome." Elife 11 (2022): e82556.

ISSN 2616-5775 Vol. 8, Issue 7: 1-10, DOI: 10.25236/AJCIS.2025.080701

[6] Bordin, Nicola, et al. "AlphaFold2 reveals commonalities and novelties in protein structure space for 21 model organisms." Communications biology 6.1 (2023): 160.
[7] Hu, Liya, et al. "Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography." Communications Biology 5.1 (2022): 419.
[8] Nomburg, Jason, et al. "Birth of protein folds and functions in the virome." Nature 633.8030 (2024): 710-717.
[9] Xiao, Qingjie, et al. "Utilization of AlphaFold2 to predict MFS protein conformations after selective mutation." International Journal of Molecular Sciences 23.13 (2022): 7235.
[10] Wayment-Steele, Hannah K., et al. "Predicting multiple conformations via sequence clustering and AlphaFold2." Nature 625.7996 (2024): 832-839.
[11] Varadi, Mihaly, et al. "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models." Nucleic acids research 50.D1 (2022): D439-D444.
[12] Sim, Jiho, Sohee Kwon, and Chaok Seok. "HProteome-BSite: predicted binding sites and ligands in human 3D proteome." Nucleic Acids Research 51.D1 (2023): D403-D408.
[13] Wong, Felix, et al. "Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery." Molecular systems biology 18.9 (2022): e11081.
[14] Yang, Qiangzhen, et al. "Structural comparison and drug screening of spike proteins of ten SARS-CoV-2 variants." Research (2022).
[15] Zeng, Dian, et al. "A hemagglutinin stem vaccine designed rationally by AlphaFold2 confers broad protection against influenza B infection." Viruses 14.6 (2022): 1305.
[16] Ibrahim, Tarhan, et al. "AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs." PLoS biology 21.2 (2023): e3001962.
[17] Wayment-Steele, Hannah K., et al. "Predicting multiple conformations via sequence clustering and AlphaFold2." Nature 625.7996 (2024): 832-839.
[18] Jendrusch, Michael, Jan O. Korbel, and S. Kashif Sadiq. "AlphaDesign: A de novo protein design framework based on AlphaFold." Biorxiv (2021): 2021-10.
[19] Goverde, Casper A., et al. "De novo protein design by inversion of the AlphaFold structure prediction network." Protein Science 32.6 (2023): e4653.
[20] Van Kempen, Michel, et al. "Fast and accurate protein structure search with Foldseek." Nature biotechnology 42.2 (2024): 243-246.
[21] Barrio-Hernandez, Inigo, et al. "Clustering predicted structures at the scale of the known protein universe." Nature 622.7983 (2023): 637-645.
[22] Ma, Wenjian, et al. "Enhancing protein function prediction performance by utilizing AlphaFold-predicted protein structures." Journal of Chemical Information and Modeling 62.17 (2022): 4008-4017.
[23] Hu, Mingyang, et al. "Exploring evolution-aware &-free protein language models as protein function predictors." Advances in Neural Information Processing Systems 35 (2022): 38873-38884.