

Review Article

The Role of Artificial Intelligence in Protein Structural and Functional Prediction: Current Status and Future Prospective

Alebachew Molla^{1,*} , Gedif Meseret² 

¹Department of Biotechnology, Wolaita Sodo University, Wolaita Sodo, Ethiopia

²Department of Biology, Wolaita Sodo University, Wolaita Sodo, Ethiopia

Abstract

Artificial intelligence (AI) has transformed the landscape of protein structural and functional prediction, significantly advancing the accuracy and efficiency of these processes. Currently, AI-driven methods, especially deep learning algorithms, enable the prediction of protein 3D structures from amino acid sequences with unprecedented precision. Artificial intelligence (AI) has emerged as a transformative force in the field of protein science, offering powerful tools for the structural and functional prediction of proteins. AI models use vast databases of known protein structures and leverage evolutionary information from multiple sequence alignments or protein language models to infer spatial conformations of proteins. Deep neural networks, convolutional neural networks, and graph-based models enhance prediction accuracy beyond traditional homology or ab initio methods. AlphaFold2's breakthrough in CASP14 demonstrated near-experimental accuracy for many proteins, ushering in a new era of AI-based structural biology. AI-driven protein structure and function prediction tools are democratizing access to complex biological data, making it possible for many research groups to accelerate discovery without expensive and time-consuming experiments. Machine learning models, such as DeepGO-SE, utilize pretrained protein language models alongside biological knowledge and protein interaction networks to predict Gene Ontology functions. These models improve prediction accuracy even for proteins with unknown interactions. This review discusses the latest advancements in AI-driven methodologies, including deep learning models and large language models, highlighting their significant contributions to resolving protein structures, functional annotation, and interaction mapping. The article summarizes current achievements, evaluates the strengths and limitations of AI approaches, and outlines future prospects for integrating AI with experimental data to accelerate discoveries in proteomics and drug discovery.

Keywords

Artificial Intelligence, Deep Learning, Machine Learning, Protein Folding, AlphaFold, Protein Design, Protein Structural Prediction, Protein Functional Prediction

*Corresponding author: alexbm2127@gmail.com (Alebachew Molla)

Received: 31 July 2025; **Accepted:** 11 August 2025; **Published:** 3 September 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Understanding the three-dimensional structure and biological function of proteins is fundamental to deciphering the molecular mechanisms of life, enabling advances in drug discovery, disease diagnosis, and bioengineering [1]. Proteins orchestrate virtually all cellular processes, yet their functional properties are exquisitely tied to complex, dynamic folds that emerge from linear amino acid sequences which excellently encapsulated by Anfinsen's dogma [2].

Historically, experimental techniques like X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) spectroscopy have been the gold standard for protein structural determination [3]. However, these methods remain labor-intensive, costly (often exceeding \$100,000 per structure), and technically constrained by protein size, flexibility, or crystallization challenges, creating a massive gap between the exponentially growing repository of protein sequences (>200 million in UniProt) and experimentally resolved structures (~200,000 in the PDB). This "structure deficit" has long hindered progress in precision medicine and protein design.

Computational approaches emerged to bridge this gap, evolving from rudimentary homology modeling to sophisticated *ab initio* methods, but accuracy remained limited until the integration of artificial intelligence [4]. The recent convergence of deep learning architectures, evolutionary data, and biophysical constraints has triggered a paradigm shift, exemplified by AlphaFold2's solution of the 50-year-old protein folding problem is innovation recognized by the 2024 Nobel Prize in Chemistry [5]. This review comprehensively examines the transformative role of AI in protein structural and functional prediction, analyzing state of the art methodologies (including AlphaFold3, RoseTTAFold, and ESMFold), current applications in biomedicine and biotechnology, persistent limitations, and future paths for scalable, dynamic, and multi-scale molecular modeling.

2. Overview of Protein Structural and Functional Prediction

Proteins exhibit hierarchical structural organization essential to their function: *primary structure* (linear amino acid sequence), *secondary structure* (local folding into α -helices and β -sheets stabilized by hydrogen bonds), *tertiary structure* (three-dimensional folding of a single polypeptide chain), and *quaternary structure* (assembly of multiple polypeptide subunits) [6]. Historically, determining these structures relied exclusively on experimental techniques like *X-ray crystallography*, *NMR spectroscopy*, and later *cryo-electron microscopy (cryo-EM)*. While powerful, these methods faced significant limitations high costs (often >\$100,000 per structure), lengthy timelines

(months to years), technical constraints (e.g., crystallization difficulties, size limits for NMR), and challenges with dynamic or membrane proteins [3]. These bottlenecks catalyzed the development of computational approaches, which evolved from rudimentary *physical energy minimization* in the 1970s to sophisticated *machine learning* driven pipelines.

3. Computational Prediction Methods

Computational prediction methods refer to techniques that use algorithms, mathematical models, and computational tools to analyze data and predict properties, behaviors, or interactions within complex systems [7]. These methods are widely used across various fields such as bioinformatics, drug discovery, and systems biology. Computational prediction methods are broadly classified into three categories as *Template-Based Modeling (TBM)*, *Template-Free (Ab Initio) Modeling*, *Hybrid Approaches* [8].

3.1. Template-Based Modeling

Template-Based Modeling leverages evolutionary relationships by "threading" target sequences onto experimentally solved homologous structures (templates) from databases like the PDB [9]. *Homology modeling* (e.g., MODELLER, SWISS-MODEL) dominated early efforts, achieving moderate accuracy when sequence identity exceeded 30%. *Fold recognition* techniques (e.g., *threading*) extended this to distant homologs by aligning sequences to structural fold libraries [3, 10].

3.2. Template-Free (Ab Initio) Modeling

Template free modeling predicts structures from physical principles without templates, simulating protein folding energetics [11]. Early *ab initio* methods (e.g., Rosetta, fragment assembly) sampled conformational space using Monte Carlo algorithms but struggled with accuracy beyond small proteins due to computational complexity and imperfect energy functions [3].

3.3. Hybrid Approaches

Hybrid approaches integrate TBM and *ab initio* strategies with statistical potentials or machine learning [12]. Tools like *I-TASSER* combined threading with iterative fragment assembly, while *RaptorX* used deep learning to predict contact maps guiding *ab initio* folding [13]. These hybrid frameworks laid the groundwork for the AI revolution, bridging sequence evolution, physical constraints, and pattern recognition to tackle larger, more complex folds [12, 14].

4. Artificial Intelligence Techniques in Protein Structure Prediction

4.1. Role of Machine Learning and Deep Learning

Artificial intelligence (AI) techniques, particularly machine learning (ML) and deep learning (DL), have revolutionized protein structure prediction, a fundamental problem in bioinformatics and structural biology [3]. Protein structure prediction involves determining the 3D arrangement of a protein from its amino acid sequence and is vital for understanding protein function and drug design [15].

Machine learning methods have historically been applied at multiple levels of the protein structure prediction challenge: from 1D sequence-based structural features, 2D spatial relationships, to 3D tertiary structure and 4D quaternary structures of protein complexes [16]. Various ML approaches, including hidden Markov models, support vector machines, Bayesian methods, and clustering, contributed significantly to these efforts [17]. These methods, especially those leveraging co-evolutionary data from multiple sequence alignments, improved predictions for proteins with homologous sequences, but faced limitations when homologs were absent [3].

In contrast, deep learning is a subset of machine learning which has marked a more substantial advance by effectively capturing complex and hierarchical features from protein sequences and evolutionary information [18]. Deep neural networks, convolutional neural networks (CNN), recurrent neural networks (RNN), graph neural networks, and deep residual networks have been developed to predict various structural levels, including secondary and tertiary structures [3]. These DL models learn from large protein databases, integrating sequence data, structural templates, and evolutionary patterns, enabling predictions even when template homologs are not readily available [19].

Notably, state of the art deep learning models such as AlphaFold and RoseTTAFold have set new benchmarks by predicting protein folds with near-experimental accuracy, as demonstrated in critical assessments like CASP14 and CASP15 [6]. AlphaFold, for example, combines multiple sources of data, including protein sequences and structural database information, and uses advanced neural network architectures to directly infer 3D structures from sequence data [20]. This model addresses challenges in template free modeling and significantly closes the gap between known sequences and experimentally determined structures.

Deep learning models operate by initially performing multiple sequence alignments to detect co-evolutionary signals, predicting local structural features such as torsion angles and secondary structures, and assembling 3D models by integrating these predictions with spatial contact information [4]. Optimization techniques, including energy function minimi-

zation and gradient-based methods, refine these structural models. The integration of DL enhances the ability to predict structures of proteins with complex folds and those previously challenging for traditional ML, including certain membrane proteins and intrinsically disordered proteins, although some dynamic and alternative conformations remain difficult to predict reliably [21].

Overall, the role of machine learning and deep learning in protein structure prediction has been transformative. ML laid foundational methods for analyzing sequence-structure relationships, while DL has dramatically improved accuracy and scalability, enabling high-resolution predictive models that impact drug discovery, protein engineering, and biological research [22]. Continuous advancements in AI frameworks and computational power are expected to further refine these predictions and expand their applications.

4.2. Use of Sequence Information and Multiple Sequence Alignments

The use of sequence information and multiple sequence alignments (MSAs) is essential in protein structural and functional prediction [23]. MSAs align homologous protein sequences from various species, revealing conserved regions and evolutionary relationships that are critical for understanding protein structure and function [24]. These alignments capture co-evolution signals where spatially close residues in the 3D structure tend to mutate in a correlated way to maintain structural integrity. AI models, such as AlphaFold2, rely heavily on MSAs as input to detect these evolutionary constraints and accurately predict protein folds. By leveraging the evolutionary patterns encoded in MSAs, AI can infer residue contacts and improve the accuracy of 3D structure prediction, even for proteins without close structural homologs [25]. Thus, MSAs form a foundational dataset that integrates sequence information across related proteins to guide and enhance structural biology predictions and protein function inference [26]. This role of MSAs remains pivotal despite emerging alternative approaches, due to their ability to provide rich evolutionary context essential for high-confidence modeling.

4.3. Advances in Accuracy and Challenges with Complex Proteins and Novel Folds

Advances in protein structure prediction accuracy have been remarkable, especially with AI systems like AlphaFold2 and its successor AlphaFold3. AlphaFold2 demonstrated near-experimental accuracy in predicting protein structures, achieving a median backbone RMSD (root-mean-square deviation) of approximately 0.96 Å in CASP14, vastly outperforming competing methods [6]. It produced highly accurate domain and side-chain conformations, even for proteins without close homologs, which was considered a major breakthrough in computational biology [5]. This accuracy has

transformed structural biology, facilitating faster experimental structure determination and advancing drug discovery and protein design [27].

AlphaFold3 introduces architectural advancements such as a diffusion-based module to improve prediction of biomolecular interactions, including complexes of proteins and other molecular types more accurately [6]. It reduces reliance on extensive multiple sequence alignment processing and directly predicts atomic coordinates with enhanced chemical and structural fidelity [28]. This evolution in model design is a major step toward overcoming limitations in predicting novel folds, protein dynamics, and complex assemblies.

Ongoing efforts focus on enhancing flexibility by sampling multiple conformations, incorporating experimental data iteratively to refine models, and enriching predictions with cofactors and ligands through complementary resources [29]. Despite outstanding progress, predicting protein-ligand interactions at docking accuracy levels and fully resolving novel or highly flexible structures remain critical frontiers for future development [30].

5. AI for Functional Prediction of Protein

AI substantially advances functional prediction of proteins by linking their predicted three-dimensional structures to biological roles, which is crucial since protein function is largely determined by structure [8]. With highly accurate structural models from AI methods like AlphaFold, functional sites such as ligand binding pockets, interaction interfaces, and enzymatic active sites can be identified [4]. AI approaches integrate predicted structures with sequence and evolutionary information to detect surface pockets and conserved residues, using techniques like graph-based deep learning and transformer models to predict ligand binding, protein-protein interactions, and enzymatic activity [4].

Moreover, integration with functional annotation tools such as AlphaFill and LegoFill enriches structural predictions by adding biologically relevant ligands, cofactors, and ions drawn from experimental databases, providing context that aids functional interpretation [31]. This combination of precise structural modeling and advanced annotation tools enables detailed and accurate functional predictions, fostering deeper insight into protein biology and supporting applications in drug discovery and protein engineering [32].

6. Databases, Tools and Resources

6.1. Protein Structure Databases and AI-predicted Model Repositories

Protein structure databases and AI-predicted model repositories are fundamental resources that accelerate research in structural biology by providing access to experimentally determined and computationally predicted protein structures [4].

The Protein Data Bank (PDB) remains the gold standard for experimentally resolved protein structures, offering extensively curated models derived from crystallography, NMR, and cryo-EM. It serves as a critical reference for validating AI predictions and understanding protein conformations under physiological conditions [33].

Complementing the PDB is the AlphaFold Protein Structure Database (AlphaFold DB) which contains over 200 million high-accuracy protein structure predictions generated by the AlphaFold AI system [34]. This repository dramatically expands structural coverage beyond experimentally solved proteins, offering predicted models for vast numbers of proteins across multiple species. AlphaFold DB provides confidence metrics such as pLDDT scores to indicate prediction reliability and features advanced visualization tools for easy exploration of predicted structures [35]. It is freely accessible and widely used to fill gaps in structural knowledge, thus supporting applications like drug discovery and protein engineering.

The integration of these resources is exemplified by tools such as the PDBe-KB platform, which allows for direct superposition and comparison of AlphaFold models with corresponding experimental structures from the PDB [36]. This facilitates validation of AI predictions and highlights conformational states captured by experiments versus models. Additional tools like AlphaFind enable structure-based searches across the entire AlphaFold DB, helping researchers discover structurally similar proteins rapidly within this massive dataset [36].

Together, these databases form a comprehensive ecosystem: the PDB provides experimentally validated structures as a foundation, while AlphaFold DB offers AI-predicted models that expand coverage to almost the entire proteome and specialized web tools offer interactivity and cross-comparison [34]. This synergy advances structural and functional protein analysis by merging experimental accuracy with AI-driven breadth and scale [37]. Such resources are indispensable for accelerating biological understanding, hypothesis generation, and translational research in molecular biology and biomedicine.

6.2. Software and Platforms Enabling AI-driven Predictions

Software and platforms enabling AI-driven protein structure predictions encompass a range of advanced tools integrating deep learning with bioinformatics to accurately model protein 3D structures from sequences [38]. Notable platforms include AlphaFold2 and its updated version AlphaFold3, developed by DeepMind, which deliver near-experimental accuracy and support complex predictions including protein complexes and interactions [20]. AlphaFold is accessible both as open-source software for local use and through web servers like the AlphaFold Protein Structure Database, offering extensive repositories of predicted models [36]. The Ro-

seTTAFold platform uses deep neural networks for rapid and accurate structure predictions and is available as open-source code and via associated web interfaces [39].

Commercial solutions such as NovaFold AI utilize AlphaFold2's algorithms within integrated environments to facilitate user-friendly prediction and visualization workflows, targeting difficult cases like membrane proteins and multi-domain structures [40]. Additional tools leverage AI in various prediction approaches, including I-TASSER, Robetta, and trRosetta, which combine machine learning with homology modeling, threading, and ab initio predictions [13, 41-43]. These platforms typically provide web servers and downloadable packages to accommodate diverse research needs.

Moreover, complementary software like AlphaFill and LegoFill enhance functional annotation by incorporating ligands and cofactors into predicted structures, enabling functional insights [31, 38]. The ecosystem of AI-driven protein prediction tools thus spans from foundational deep learning models to integrated platforms and annotation tools, collectively advancing biological research through high-accuracy, scalable, and accessible protein modeling solutions [4].

6.3. Role of Open Data and Collaborative Resources

Open data and collaborative resources play a pivotal role in accelerating advances in protein structural biology by providing widespread, unrestricted access to both experimentally determined and AI-predicted protein structures [33, 44]. The AlphaFold Protein Structure Database (AlphaFold DB), a prime example, offers over 200 million high-accuracy predicted protein models freely available to the global scientific community under an open Creative Commons license [34]. This open access enables researchers across disciplines to use comprehensive structural data for hypothesis generation, drug discovery, protein engineering, and functional annotation without barriers posed by proprietary data or limited experimental coverage.

Collaborative integration with established databases like the Protein Data Bank (PDB), UniProt, and InterPro further enriches the utility of these resources by linking predicted structures to experimental evidence, sequence annotations, and functional classifications, fostering data interoperability and broader biological insights [45]. Platforms hosting these data also provide user-friendly interfaces, bulk downloads and programmatic access that support diverse research workflows and large-scale computational analyses [46].

The communal availability of such vast data catalyzes innovation by enabling reproducibility, cross-validation, and comparative studies, while also lowering entry barriers for labs lacking experimental infrastructure. This collaborative ecosystem nurtures rapid progress in understanding protein function, dynamics and interactions, ultimately driving

breakthroughs in medicine, biotechnology, and fundamental biology. Open data repositories and cooperative resources democratize access to protein structural knowledge, underpinning a new era of data-driven biological discovery and translational research.

7. Current Status and Achievements

7.1. Milestones in AI Based Prediction Accuracy

AlphaFold2 set a new standard by achieving near-experimental accuracy at CASP14 (2020), with a median backbone RMSD of approximately 0.96 Å and all-atom RMSD around 1.5 Å [6]. This accuracy vastly outperformed competing methods, which typically had backbone RMSDs near 2.8 Å or higher. AlphaFold2's architecture integrated evolutionary, physical, and geometric constraints into a deep learning model that jointly embeds multiple sequence alignments (MSAs) and pairwise residue features, enabling precise 3D coordinate predictions solely from amino acid sequences [47].

AlphaFold3 was released recently and basic for this field. It introduces a diffusion-based architecture that improves prediction of protein complexes, including proteins, nucleic acids, small molecules, ions, and modified residues [20]. AlphaFold3 achieves substantially higher accuracy across a broad range of biomolecular interaction types compared to specialized tools and reduces dependence on extensive MSA processing by using a simplified "pairformer" module [48]. It also directly predicts atomic coordinates with enhanced chemical and stereochemical fidelity, advancing modeling of protein dynamics and interactions.

7.2. Successful Applications in Drug Discovery, Genomics and Synthetic Biology

The availability of high-accuracy protein structural models at proteome scale (over 200 million structures in the AlphaFold Protein Structure Database) has transformed multiple fields [6]. In drug discovery, precise structural models facilitate identification of ligand binding sites and protein-protein interaction interfaces, accelerating rational drug design and optimization [13]. Genomics benefits from structural annotation of uncharacterized proteins, aiding functional insights and variant interpretation at scale. Synthetic biology and protein engineering leverage AI models to design novel proteins with bespoke functions, improve enzyme catalytic efficiency, and develop biosensors [49].

7.3. Limitations and Gaps

Despite extraordinary progress, several challenges remain. AlphaFold models typically predict a single protein conformation and struggle with intrinsically disordered regions,

dynamic states, and conformational flexibility critical for many functions [34]. Membrane proteins, post-translational modifications, and multi-chain complexes still pose prediction difficulties, though AlphaFold3's enhanced architecture addresses some gaps. Prediction of protein-ligand interactions at docking accuracy, fully accurate modeling of protein dynamics and comprehensive prediction of novel folds or highly flexible proteins are ongoing research frontiers [41, 50].

8. Future Perspectives

Future perspectives in AI-driven protein structural and functional prediction, particularly following the breakthroughs by AlphaFold2 and AlphaFold3, focus on expanding the capabilities of AI models beyond static protein structures to encompass dynamic, complex, and interaction-rich biological systems. The key future directions include Modeling Protein Complexes and Biomolecular Interactions, Integration with Experimental Data, Addressing Limitations in Dynamic and Flexible Regions, Leveraging Proprietary Pharmaceutical Data for Drug Discovery, Open Source and Democratization of AI Tools, Enhanced Chemical and Structural Fidelity and Expanding Functional Annotation Integration.

8.1. Modeling Protein Complexes and Biomolecular Interactions

AlphaFold3 marked a major advance by enabling accurate prediction of molecular complexes, such as protein-protein, protein-ligand, and protein-nucleic acid assemblies. Future work aims to refine this capacity with improved modeling of multi-component complexes, including transient interactions and conformational variability, which are essential for cellular functions and drug targeting.

8.2. Integration with Experimental Data

Combining AI predictions with experimental structural biology techniques like cryo-electron microscopy and NMR spectroscopy is expected to enhance model accuracy and validation. Iterative workflows integrating AI and experimental data will improve understanding of protein dynamics, folding pathways, and functional states beyond static snapshots.

8.3. Addressing Limitations in Dynamic and Flexible Regions

Current AI models often predict a single predominant conformation and struggle with intrinsically disordered or flexible regions. Future advances will focus on capturing protein dynamics, multiple conformations, and post-translational modifications to better represent biological reality.

8.4. Leveraging Proprietary Pharmaceutical Data for Drug Discovery

The shortage of diverse protein-drug interaction data in public repositories limits AI's ability to predict drug binding accurately. Consortia of pharmaceutical companies are developing AI models trained on their proprietary structural data to enhance drug-target interaction predictions, though these will initially be restricted to member companies, highlighting a growing tension between open science and commercial interests.

8.5. Open Source and Democratization of AI Tools

While AlphaFold3's initial release limited code availability for commercial use, recent moves have made it accessible to academic researchers. Parallel efforts to develop open-source reproductions like OpenFold3 will democratize access, enabling broader innovation in biotechnology and drug development.

8.6. Enhanced Chemical and Structural Fidelity

New AI architectures employ diffusion-based methods and refined neural network modules to predict atomic coordinates with higher stereochemical accuracy, improving prediction of ligand binding sites, metal ion coordination, and enzyme active sites, thereby facilitating more precise functional annotation and drug design.

8.7. Expanding Functional Annotation Integration

Future tools will increasingly combine structure prediction with functional annotation databases and AI-driven pipelines to link sequence, structure, and function comprehensively. This will accelerate discovery of novel protein functions and design of tailored biomolecules.

9. Conclusion

AI has profoundly transformed the field of protein structural and functional prediction by delivering unprecedented accuracy, speed, and scale in modeling protein 3D structures directly from amino acid sequences. Landmark breakthroughs with models like AlphaFold2 and the recent AlphaFold3 have achieved near-experimental precision, vastly expanding structural coverage across millions of proteins spanning diverse species. This revolution has accelerated hypothesis generation, drug discovery, functional annotation and novel protein design to effectively bridging the gap between vast protein sequence data and detailed structural knowledge.

The future of AI-driven protein structural biology envisions

holistic models capable of representing dynamic, multi-component biomolecular systems with enhanced chemical fidelity and interpretability. Democratization of AI tools through open-source initiatives, combined with collaborative data sharing, will foster innovation across biotechnology, synthetic biology and medicine. The synergy between AI predictions and diverse experimental and functional datasets promises deeper biological understanding and accelerated translational applications.

Generally, AI has revolutionized protein structural and functional prediction by overcoming traditional barriers and providing transformative tools that reshape biomedical research and biotechnology. While significant challenges persist, ongoing methodological advances and resource development are poised to drive the field toward comprehensive, accurate and functionally rich protein models used to unlocking new frontiers in science and medicine.

Abbreviations

AI	Artificial Intelligences
CASP	Critical Assessment of Techniques for Protein Structure Prediction
CNN	Convolutional Neural Networks
Cryo-EM	Cryo-Electron Microscopy
DL	Deep Learning
ML	Machine Learning
Msas	Multiple Sequence Alignments
NMR	Nuclear Magnetic Resonance
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
RMSD	Root-Mean-Square Deviation
RNN	Recurrent Neural Networks
TBM	Template-Based Modeling
TTA	Three-Track Attention

Author Contributions

Alebachew Molla: Conceptualization and framing, Comprehensive literature collection and critical analysis, Writing the original draft, Reviewing and editing, Providing expert interpretation.

Gedif Meseret: Reviewing and editing, Providing expert interpretation. The authors read and approved the final manuscript.

Funding

This review received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] V. M. Prabantu, N. Naveenkumar, and N. Srinivasan, Influence of Disease-Causing Mutations on Protein Structural Networks, *Front. Mol. Biosci.*, vol. 7, no. March, pp. 1-11, 2021, <https://doi.org/10.3389/fmolb.2020.620554>
- [2] M. N. Gupta and V. N. Uversky, Protein structure-function continuum model: Emerging nexuses between specificity, evolution, and structure, *Protein Sci.*, vol. 33, no. 4, pp. 1-36, 2024, <https://doi.org/10.1002/pro.4968>
- [3] Y. Meng *et al.*, Protein structure prediction via deep learning: an in-depth review, *Front. Pharmacol.*, vol. 16, no. April, pp. 1-20, 2025, <https://doi.org/10.3389/fphar.2025.1498662>
- [4] L. Chen *et al.*, AI-Driven Deep Learning Techniques in Protein Structure Prediction, *Int. J. Mol. Sci.*, vol. 25, no. 15, pp. 1-21, 2024, <https://doi.org/10.3390/ijms25158426>
- [5] L. Wang *et al.*, Overview of AlphaFold2 and breakthroughs in overcoming its limitations, *Comput. Biol. Med.*, vol. 176, 2024, <https://doi.org/10.1016/j.combiomed.2024.108620>
- [6] J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold, *Nature*, vol. 596, no. 7873, pp. 583-589, 2021, <https://doi.org/10.1038/s41586-021-03819-2>
- [7] M. Thafar, A. Bin Raies, S. Albaradei, M. Essack, and V. B. Bajic, Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities, *Front. Chem.*, vol. 7, no. November, pp. 1-19, 2019, <https://doi.org/10.3389/fchem.2019.00782>
- [8] S. Zhang, K. Liu, Y. Liu, X. Hu, and X. Gu, The role and application of bioinformatics techniques and tools in drug discovery, *Front. Pharmacol.*, vol. 16, no. February, pp. 1-12, 2025, <https://doi.org/10.3389/fphar.2025.1547131>
- [9] K. Yan, X. Fang, Y. Xu, and B. Liu, Protein fold recognition based on multi-view modeling, *Bioinformatics*, vol. 35, no. 17, pp. 2982-2990, 2019, <https://doi.org/10.1093/bioinformatics/btz040>
- [10] R. T. Maia, Protein structure prediction by computational homology modeling: a brief explanation, *Int. J. Mol. Biol. Open Access*, vol. 7, no. 1, pp. 118-120, 2024, <https://doi.org/10.15406/ijmboa.2024.07.00180>
- [11] S. Dhingra, R. Sowdhamini, F. Cadet, and B. Offmann, A glance into the evolution of template-free protein structure prediction methodologies, *Biochimie*, vol. 175, pp. 85-92, 2020, <https://doi.org/10.1016/j.biochi.2020.04.026>
- [12] E. Küküktopcu, E. Cemek, B. Cemek, and H. Simsek, Hybrid Statistical and Machine Learning Methods for Daily Evapotranspiration Modeling, *Sustain.*, vol. 15, no. 7, pp. 1-15, 2023, <https://doi.org/10.3390/su15075689>
- [13] W. Zheng *et al.*, Deep-learning-based single-domain and multidomain protein structure prediction with D-I-TASSER, *Nat. Biotechnol.*, 2025, <https://doi.org/10.1038/s41587-025-02654-4>
- [14] N. Khetrpal, Cognition meets assistive technology: Insights from load theory of selective attention, *Handb. Res. Hum. Cogn. Assist. Technol. Des. Access. Transdiscipl. Perspect.*, pp. 96-108, 2010, <https://doi.org/10.4018/978-1-61520-817-3.ch006>

- [15] V. A. Jisna and P. B. Jayaraj, Protein Structure Prediction: Conventional and Deep Learning Perspectives, *Protein J.*, vol. 40, no. 4, pp. 522-544, 2021, <https://doi.org/10.1007/s10930-021-10003-y>
- [16] M. Torrisi, G. Pollastri, and Q. Le, Deep learning methods in protein structure prediction, *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 1301-1310, 2020, <https://doi.org/10.1016/j.csbj.2019.12.011>
- [17] N. A. E. Venanzi, A. Basciu, A. V. Vargiu, A. Kiparissides, P. A. Dalby, and D. Dikicioglu, Machine Learning Integrating Protein Structure, Sequence, and Dynamics to Predict the Enzyme Activity of Bovine Enterokinase Variants, *J. Chem. Inf. Model.*, vol. 64, no. 7, pp. 2681-2694, 2024, <https://doi.org/10.1021/acs.jcim.3c00999>
- [18] J. Cui, S. Yang, L. Yi, Q. Xi, D. Yang, and Y. Zuo, Recent advances in deep learning for protein-protein interaction: a review, *BioData Min.*, vol. 18, no. 1, 2025, <https://doi.org/10.1186/s13040-025-00457-6>
- [19] J. Jänes and P. Beltrao, Deep learning for protein structure prediction and design-progress and applications, *Mol. Syst. Biol.*, vol. 20, no. 3, pp. 162-169, 2024, <https://doi.org/10.1038/s44320-024-00016-x>
- [20] J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold3, *Nature*, vol. 630, no. 8016, pp. 493-500, 2024, <https://doi.org/10.1038/s41586-024-07487-w>
- [21] Y. Zhang, Protein structure prediction and protein design, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9096, no. 734, pp. 194-207, 2015, <https://doi.org/10.1016/j.sbi.2021.01.007.Deep>
- [22] J. Durairaj, D. de Ridder, and A. D. J. van Dijk, Beyond sequence: Structure-based machine learning, *Comput. Struct. Biotechnol. J.*, vol. 21, pp. 630-643, 2023, <https://doi.org/10.1016/j.csbj.2022.12.039>
- [23] Y. Si and C. Yan, Protein complex structure prediction powered by multiple sequence alignments of interologs from multiple taxonomic ranks and AlphaFold2, *Brief. Bioinform.*, vol. 23, no. 4, pp. 1-13, 2022, <https://doi.org/10.1093/bib/bbac208>
- [24] M. Shatsky, R. Nussinov, and H. J. Wolfson, Protein Structure Prediction, *Protein Struct. Predict.*, no. 1, pp. 125-146, 2008, <https://doi.org/10.1007/978-1-59745-574-9>
- [25] L. M. S. Das C Hansen KC and Tyler JK, Wray, and L. M. S. Das C Hansen KC and Tyler JK, HHS Public Access, *Physiol. Behav.*, vol. 176, no. 3, pp. 139-148, 2017, <https://doi.org/10.1002/prot.25405.Protein>
- [26] C. Zhang *et al.*, The Historical Evolution and Significance of Multiple Sequence Alignment in Molecular Structure and Function Prediction, *Biomolecules*, vol. 14, no. 12, 2024, <https://doi.org/10.3390/biom14121531>
- [27] N. Borkakoti and J. M. Thornton, AlphaFold2 protein structure prediction: Implications for drug discovery, *Curr. Opin. Struct. Biol.*, vol. 78, 2023, <https://doi.org/10.1016/j.sbi.2022.102526>
- [28] Z. Lin *et al.*, Language models of protein sequences at the scale of evolution enable accurate structure prediction, *bioRxiv*, p. 2022. 07. 20. 500902, 2022.
- [29] B. M *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network, *Yearb. Paediatr. Endocrinol.*, vol. 373, no. 6557, pp. 871-876, 2022, <https://doi.org/10.1530/ey.19.15.15>
- [30] J. Jiménez, S. Doerr, G. Martínez-Rosell, A. S. Rose, and G. De Fabritiis, DeepSite: Protein-binding site predictor using 3D-convolutional neural networks, *Bioinformatics*, vol. 33, no. 19, pp. 3036-3042, 2017, <https://doi.org/10.1093/bioinformatics/btx350>
- [31] M. L. Hekkelman, I. de Vries, R. P. Joosten, and A. Perrakis, AlphaFill: enriching AlphaFold models with ligands and cofactors, *Nat. Methods*, vol. 20, no. 2, pp. 205-213, 2023, <https://doi.org/10.1038/s41592-022-01685-y>
- [32] P. Bryant, A. Kelkar, A. Guljas, C. Clementi, and F. Noé, Structure prediction of protein-ligand complexes from sequence information with Umol, *Nat. Commun.*, vol. 15, no. 1, pp. 1-12, 2024, <https://doi.org/10.1038/s41467-024-48837-6>
- [33] S. K. Burley *et al.*, Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students, *Biomolecules*, vol. 12, no. 10, 2022, <https://doi.org/10.3390/biom12101425>
- [34] M. Varadi *et al.*, AlphaFold Protein Structure Database in 2024, *Nucleic Acids Res.*, vol. 52, no. November 2023, pp. D368-D375, 2024.
- [35] M. A. Pak *et al.*, Using AlphaFold to predict the impact of single mutations on protein stability and function, *PLoS One*, vol. 18, no. 3 March, pp. 1-9, 2023, <https://doi.org/10.1371/journal.pone.0282689>
- [36] M. Varadi *et al.*, PDBe and PDBe-KB: Providing high-quality, up-to-date and integrated resources of macromolecular structures to support basic and applied research and education, *Protein Sci.*, vol. 31, no. 10, pp. 1-10, 2022, <https://doi.org/10.1002/pro.4439>
- [37] O. Kovalevskiy, J. Mateos-Garcia, and K. Tunyasuvunakool, AlphaFold two years on: Validation and impact, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 121, no. 34, pp. 1-6, 2024, <https://doi.org/10.1073/pnas.2315002121>
- [38] T. T. Ogunjobi *et al.*, Bioinformatics tools in protein analysis: Structure prediction, interaction modelling, and function relationship, *Eur. J. Sustain. Dev. Res.*, vol. 9, no. 3, p. em0298, 2025, <https://doi.org/10.29333/ejosdr/16340>
- [39] Y. Gao, H. Wang, J. Zhou, and Y. Yang, An easy-to-use three-dimensional protein-structure-prediction online platform 'DPL3D' based on deep learning algorithms, *Curr. Res. Struct. Biol.*, vol. 9, no. December 2024, p. 100163, 2025, <https://doi.org/10.1016/j.crstbi.2024.100163>
- [40] H. Zhang *et al.*, AlphaFold2 in biomedical research: facilitating the development of diagnostic strategies for disease, *Front. Mol. Biosci.*, vol. 11, no. July, pp. 1-16, 2024, <https://doi.org/10.3389/fmolb.2024.1414916>

- [41] X. Zhou *et al.*, I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction, *Nat. Protoc.*, vol. 17, no. 10, pp. 2326-2353, 2022, <https://doi.org/10.1038/s41596-022-00728-0>
- [42] W. Wang *et al.*, trRosettaRNA: automated prediction of RNA 3D structure with transformer network, *Nat. Commun.*, vol. 14, no. 1, 2023, <https://doi.org/10.1038/s41467-023-42528-4>
- [43] Z. Peng, W. Wang, H. Wei, X. Li, and J. Yang, Improved protein structure prediction with trRosettaX2, AlphaFold2, and optimized MSAs in CASP15, *Proteins Struct. Funct. Bioinforma.*, vol. 91, no. 12, pp. 1704-1711, 2023, <https://doi.org/10.1002/prot.26570>
- [44] X. Qiu, H. Li, G. Ver Steeg, and A. Godzik, Advances in AI for Protein Structure Prediction: Implications for Cancer Drug Discovery and Development, *Biomolecules*, vol. 14, no. 3, pp. 1-16, 2024, <https://doi.org/10.3390/biom14030339>
- [45] T. U. Consortium, UniProt: the Universal Protein Knowledgebase in 2025 *Nucleic Acids Res.* 53: D0-D0 (2025), no. November 2024, pp. 609-617, 2025.
- [46] M. Blum *et al.*, InterPro: The protein sequence classification resource in 2025, *Nucleic Acids Res.*, vol. 53, no. D1, pp. D444-D456, 2025, <https://doi.org/10.1093/nar/gkae1082>
- [47] L. M. F. Bertoline, A. N. Lima, J. E. Krieger, and S. K. Teixeira, Before and after AlphaFold2: An overview of protein structure prediction, *Front. Bioinforma.*, vol. 3, no. February, pp. 1-8, 2023, <https://doi.org/10.3389/fbinf.2023.1120370>
- [48] M. G. Krokidis *et al.*, AlphaFold3: An Overview of Applications and Performance Insights, *Int. J. Mol. Sci.*, vol. 26, no. 8, pp. 1-19, 2025, <https://doi.org/10.3390/ijms26083671>
- [49] S. Kim *et al.*, Multidisciplinary approaches for enzyme biocatalysis in pharmaceuticals: protein engineering, computational biology, and nanoarchitectonics, *EES Catal.*, vol. 2, no. 1, pp. 14-48, 2024, <https://doi.org/10.1039/d3ey00239j>
- [50] M. Xue, B. Liu, S. Cao, and X. Huang, FeatureDock for protein-ligand docking guided by physicochemical feature-based local environment learning using transformer, *npj Drug Discov.*, vol. 2, no. 1, pp. 1-12, 2025, <https://doi.org/10.1038/s44386-025-00005-6>