IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BetaFold: Unravelling 3D Structure of Protein

Dr. Chaitrali Chaudhari [1], Anjum Khan[2], Kunal Desale[3], Sneha Chaubey[4], Kaushik Pal[5]

[1]Assistant Professor, Department of Computer Science and Engineering in Artificial Intelligence and Machine Learning, Lokmanya Tilak College of Engineering, Navi Mumbai, Maharashtra, India.

[2,3,4,5]Students of BE Department of Computer Science and Engineering in Artificial Intelligence and Machine Learning ,Lokmanya Tilak College of Engineering, Navi Mumbai, Maharashtra, India.

Keywords: Protein Folding, 3D Structure Prediction, Deep Learning, AlphaFold, Bioinformatics, Convolutional Neural Networks, Transformer Models, Molecular Visualization

Abstract: This study has been undertaken because as we know Proteins are central to biology, and their functions rely inextricably on their three-dimensional conformations. Although technologies such as AlphaFold have dramatically improved the prediction of structure using dazzling accuracy, they rely on Large-scale datasets and massive computing capacity, which frequently puts them inaccessibly outside of smaller labs, schools, and low-resource environments. To bridge this gap, we present Beta-Fold, an efficient and easy- to-use framework for the prediction of protein secondary structures specifically α-helices, β-sheets, and coils—directly from amino acid sequences. Utilizing a hybrid CNN-BiLSTM approach, Beta-Fold obviates the requirement for multiple sequence alignments (MSAs) and GPU-based computation. This allows for real-time prediction as well as interactive visualization via a simple web interface, making protein analysis more accessible and convenient in various research environments. Through reduced computational requirements, Beta-Fold has the potential to expedite applications in molecular diagnostics, drug discovery, and biomedical research. Directions for future work involve disease-specific case studies and incorporation into clinical bioinformatics pipelines with the ultimate aim of improving translational reach and enabling equitable access to protein modeling tools.

I. INTRODUCTION

Proteins are life's basic molecular machinery, directing nearly every biological process—enzyme catalysis and signal transduction through immune defense to structural support. Their varied activities are inherently tied to their three- dimensional structures, which arise from the amino acid linear sequences via a highly complex folding procedure. Accurately predicting protein structure from sequence data has long been a central challenge in molecular biology, with profound implications for understanding disease mechanisms, designing therapeutics, and advancing synthetic biology. Recent advances like AlphaFold and RoseTTAFold greatly boosted the development of protein structure prediction, with near-experimental accuracy in most cases. Yet these models are usually highly dependent on huge computational power, big MSAs for extensive training, and high-end GPUs— components that greatly restrict their availability for low- resource environments, small research groups, and schools. To redress this imbalance, we introduce Beta-Fold, a light and open secondary structure prediction framework. Beta-Fold is concerned with predicting the most important structural motifs— α-helices, β-sheets, and coils—directly from primary amino acid sequences. Beta-Fold utilizes a hybrid deep learning architecture that couples the spatial pattern recognition abilities of Convolutional Neural Networks (CNNs) with the contextual sequence modeling capabilities of Bidirectional Long Short- Term Memory (BiLSTM) networks. This coupling allows the model to learn both local and distant dependencies in protein sequences without needing MSAs or high-scale evolutionary information. The Beta-Fold platform is provided as an easy-to-use web interface with support for real-time sequence entry, graphical display of predicted secondary structures, and optional 3D reconstruction for increased interpretability. Its minimal computational footprint also makes it uniquely appropriate for classroom demonstrations, exploratory diagnostic research, and deployment in resource-limited settings. In democratizing access to protein structure prediction, Beta-Fold has the potential to spur accelerated discovery in molecular biology and enable increased participation in computational biosciences.

II. PROBLEM STATEMENT

Proteins are life's building blocks, and their activities rely directly on their structures. Without a correct prediction of structure, it becomes problematic to comprehend mechanisms of disease, develop successful drugs, or investigate mutations. Available tools that need a lot of computation to forecast structures are not available for most labs, leaving a serious research and diagnostic gap.

III. LITERATURE REVIEW

In order to comprehend the current advancements in the area of protein structure prediction, one must review the research studies and developments of the past few years. Various research studies have proposed various deep learning and computational techniques to enhance the prediction accuracy of protein secondary and tertiary structures. This section discusses some of the prominent research articles concerning protein structure prediction to give background information and backup support for the establishment of our proposed project, Beta-Fold.

A. *Highly Accurate Protein Structure Prediction with AlphaFold*

AlphaFold was a computational biology tour de force, with near- experimental accuracy for predicting the three- dimensional structure of proteins from their amino acid sequence directly. The model, created at DeepMind, uses an end-to-end deep learning architecture that combines many state-of-the-art components to learn about the intricacies of protein folding. Essentially, AlphaFold employs attention- based neural networks—namely transformer architectures— to represent pairwise residue contacts. These are trained on enormous databases of protein structures and sequences known to them so that the model can pick up nuanced spatial and evolutionary relationships. One of the innovations is in the use of multiple sequence alignments (MSAs) and templates from which evolutionary context is extracted and combined with geometric reasoning to generate very accurate predictions of inter-residue angles and distances. The model progressively improves its predictions with a structure module that mimics the folding process and eventually yields highly accurate atomic-level coordinates. Not only did this outperform existing methods in CASP14 competition but it also showed the potential of AI to address one of biology's grand challenges. AlphaFold's success has created new frontiers for structural biology, in which researchers can solve once-intractable proteins, speed the discovery of new medicines, and probe the molecular roots of disease with unprecedented accuracy. But its dependence on large data sets and computing power responds to the need for help from complementary lightweight technologies—such as Beta- Fold—that can bring access to yet another broader research community.

B. *Highly Accurate Protein Structure Prediction for the Human Proteome*

Following the initial success of AlphaFold, follow-up studies pushed its ability to make predictions for the structure of almost all human proteins known to date, leading to the development of the AlphaFold Protein Structure Database. Large-scale demonstration was important to the generalization capability of the model, and it was shown that deep learning- based structure prediction was not only useful for a single protein but for entire proteomes uniformly with good accuracy. To achieve this, the authors of the study scaled down the initial AlphaFold architecture to run high-throughput. Improvements in methods involved memory and computational optimization, reduction of the model's inference pipeline, and the automation of data handling tasks. These optimizations allowed the system to process enormous quantities of sequences in a timely manner, without sacrificing the quality of structural predictions. The acquired database represents a major step in structural bioinformatics to make 3D prediction open access available for hundreds of thousands of proteins for a wide range of species. The resource has already been successful in enabling functional genomics, target discovery for drug discovery, and variant annotation research and is now an accepted baseline tool in both clinical and research investigations. Through the provision of high-quality structural predictions to the world at large, this work has democratized access to molecular understanding and paved the way for eventual integration of AI-generated models into routine biological and biomedical pipelines.

C. *Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields (DeepCNF)*

DeepCNF, or Deep Convolutional Neural Fields, marks a notable advancement in protein secondary structure prediction by integrating deep learning with probabilistic modeling. This approach uniquely combines deep convolutional neural networks (CNNs)—which are adept at detecting underlying patterns in complex data—with conditional random fields (CRFs), enabling the model to simultaneously capture both local residue-level features and broader sequence dependencies. The CNN component is particularly effective at extracting hierarchical features from input data. It leverages evolutionary information from position-specific scoring matrices (PSSMs) along with physicochemical properties of amino acids, thus providing a rich contextual framework for each residue in the protein sequence. The CRF layer, meanwhile, addresses the dependencies between adjacent secondary structure labels. This ensures that predicted structures follow biologically realistic transitions, rather than producing improbable or fragmented sequences. By jointly modeling feature extraction and label consistency, DeepCNF achieves improved accuracy in predicting structural elements such as alpha-helices, beta-sheets, and coils. The integration of both local properties and global context allows DeepCNF to outperform traditional machine learning methods and earlier neural network architectures. This makes it particularly valuable in structural bioinformatics applications that require precise residue-level annotation alongside the recognition of overarching structural patterns.

D. *Improved Protein Structure Prediction Using Potentials from Deep Learning*

This work introduces a new method for protein structure refinement by fusing deep learning with the principles of physical modeling. The innovation centers around applying a deep residual convolutional neural network

(CNN) to predict potential energy landscapes as a function of inter-residue distance and orientation. Through learning directly from sequence and structure data, the model acquires the geometric constraints that dictate protein folding. In contrast to conventional energy functions that depend on hand-designed parameters or empirical force fields, the deep learning estimator dynamically learns pairwise interactions, providing more accurate and context-aware predictions. The residual CNN framework allows deeper feature learning without gradient flow interruption, enabling the model to capture intricate spatial dependencies along the protein backbone. After predicting the potential energy map, gradient-based optimization methods are used to iteratively correct atomic positions, driving the structure toward a lower- energy, more stable conformation. The hybrid approach thus increases the accuracy as well as physical reasonableness of predicted structures and is an extremely useful tool for post-prediction refinement and de novo modeling. By coupling data-driven learning with energy-based optimization, this approach adds to the emerging area of AI-augmented molecular modeling and presents a scalable solution for enhancing the predictability of computationally derived protein structures.

E. Deep Learning for Protein Secondary Structure Prediction

This systematic review was centered on the evolution and diversity of deep learning approaches used to predict protein secondary structure, with particular emphasis on architectural innovation and methodological trends. It examined a range of models—variously from Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to Long Short-Term Memory (LSTM) networks and hybrid models— each of which was developed to take different aspects of sequence and structural complexity into account. CNNs have been successfully used for their ability to extract local spatial information from amino acid sequences for the recognition of short-range patterns corresponding to structural motifs like α- helices and β-sheets. RNNs and LSTMs, respectively, are optimally capable of capturing long-range dependencies and sequential context that are crucial to understand the global folding behavior of proteins. Hybrid architectures, such as CNN-LSTM hybrid models, take advantage of the combined strength of both schools—feature extraction with CNNs and temporal modeling with LSTMs— resulting in improved predictive performance. The review also explored attention- based systems that allocated the relative significance of different residues in a sequence dynamically to allow models to focus on structurally rich regions. Mechanisms like these were found to be encouraging in terms of their potential to enhance interpretability and precision, especially in disordered or complex proteins. In addition to this, ensemble models combining greater than one architecture or prediction strategy were noted for their stability as well as generalization ability. Standard test suites such as benchmark datasets CB513, CullPDB, and CASP were introduced to the forefront as benchmarks that apply standard criteria to measure model performance between studies. Diversity in datasets, representation of features (e.g., evolutionary profiles, physicochemical properties), and evaluation procedures were emphasized as important in delineating the advancement of the field by the review. Overall, this comparison evinces the monumental advancement of deep learning in structural bioinformatics and lays a basis for possible future achievements that balance accuracy, interpretability, and computational expense.

F. Prediction of 8-State Protein Secondary Structures by a Novel Deep Learning Architecture

Zhang et al. presented a hybrid deep architecture that enhances protein secondary structure prediction from the normal 3-state classification to a more detailed 8-state model. This new paradigm in granularity and explainability offers a richer representation of protein function and folding. The architecture blends Convolutional Neural Networks (CNNs) to detect local features and Bidirectional Recurrent Neural Networks (BiRNNs) to learn dependencies between long distances. By blending evolutionary profiles like Position- Specific Scoring Matrices (PSSMs) and sequence-based features, the model picks up spatial and contextual patterns required for precise structure prediction. While hybrid architecture is good at handling long protein sequences and improving resolution, it has its downsides. Its complexity needs large, diverse data sets in order to generalize effectively and is computationally demanding with greater bias towards overfitting small samples in its training algorithm. These are the trade-offs that capture the struggle between architectural complexity and scalability in real practice. Lastly, Zhang et al.'s work demonstrates the potential for deep hybrid models to improve structural prediction since it places emphasis on how improved, less expensive solutions in scenarios where data and computing power are unavailable.

G. Recent Advances and Challenges in Protein Structure Prediction

This review paper offers a complete snapshot of the revolutionary advancement in AI-based protein structure prediction, with a specific focus on the advances made by deep learning algorithms like AlphaFold2. The authors narrate how AlphaFold2 has transformed the field to a great extent by achieving near-experimental precision in predicting the three- dimensional structures of numerous monomeric proteins, dramatically speeding up structural biology and downstream applications in drug discovery and molecular diagnostics.

In spite of these developments, the review identifies a number of ongoing challenges that are still at the edge of protein modeling. These consist of:

Multi-domain protein prediction: Reliable modeling of multi-domained, flexibly connected proteins continues to be challenging owing to complicated inter-domain interactions and conformational flexibility.

Protein-protein and protein-ligand complexes: Inference of macromolecular complex structures, particularly transient or weakly interacting complexes, necessitates models to describe dynamic interfaces and cooperative binding phenomena.

Several conformational states: Most proteins are found in ensembles of conformations, especially those that are signaling or allosterically regulated. Existing models tend to predict one static structure, providing little understanding of functional dynamics.

Folding pathways and kinetics: Although end-state predictions have become better, the mechanism of intermediate steps and energy landscapes in protein folding is

not understood, making mechanistic interpretations and the investigation of misfolding diseases difficult.

The article concludes by promoting integrative methods that merge predictions from AI with experimental information, like cryo-EM, NMR, and cross-linking mass spectrometry, to resolve these limitations. It urges the creation of models that are accurate, yet interpretable and generalizable across various biological contexts.

H. The State-of-the-Art Overview to Application of Deep Learning in Accurate Protein Design and Structure Prediction

This comprehensive review explores the rapidly evolving interface of protein science and deep learning, illustrating how structure prediction and rational protein design have been transformed by AI. The authors trace the journey from traditional template-based and physics-driven approaches to modern neural architectures that are capable of learning complex sequence–structure relationships. At the core of the paper are structures like AlphaFold2, RoseTTAFold , and DeepContact, which have delivered unprecedented accuracy in protein fold prediction, contact map, and interface of interaction prediction. The paper also delves into the inverse problem—sequence design folding into the target structure— and highlights its significance in synthetic biology, immunotherapy, and nanotechnology. Through cautious comparisons of modeling methods, including fragment-based sampling, energy-based refinement, and graph neural networks, the authors reveal the capability and capability of the tools. While deep learning affords speed, scalability, and generalizability, modeling dynamic conformational states, solvent effects, and atomic-level refinement is still challenging. The review concludes by demanding the hybrid approaches combining AI and experimental information and molecular simulations, paving the way for more interpretable and stronger protein engineering pipelines.

I. Protein structure prediction via deep learning: an in-depth review

The present review article presents a detailed and technically advanced discussion of how far deep learning has revolutionized the process of protein structure prediction, and its role in pharmacology, drug discovery, and biomedical research being a particular emphasis. The authors then set up the biological and clinical significance of protein structures and point out that understanding the three-dimensional fold of proteins is crucial towards setting up their function, interactions, and drug-ability.

Conventional experimental approaches like X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) spectroscopy, though potent, are typically time-consuming, resource-intensive, and less scalable. Conversely , computational approaches—most significantly those based on deep learning—are now a possibility for making protein structure prediction from amino acid sequence directly with scalability and increasing accuracy.

The article divides protein structure prediction into three broad methodological paradigms: template-based modeling (TBM), template-free modeling (TFM), and ab initio modeling. TBM takes advantage of known homologous structures to inform predictions, employing tools such as MODELLER and Swiss PDB Viewer. TFM includes fragment assembly and deep learning approaches such as TrRosetta and AlphaFold3 that predict inter-residue distances and orientations to ab initio model structures. Ab initio modeling, as exemplified by Rosetta and QUARK, attempts predictions based only on physicochemical principles without the aid of templates and is therefore suitable for novel or orphan proteins but computationally demanding.

The essence of the review lies in the disruptive impact of deep learning models, particularly AlphaFold2 and its latest version AlphaFold3. AlphaFold2 is praised for achieving almost experimental-level accuracy for monomeric proteins, while AlphaFold3 extends the capability to predict protein– DNA, protein–RNA, and protein–ligand interactions, a huge leap towards modeling biological complexity. The authors highlight the architectural advancements in these models, including attention mechanisms, end-to-end training pipelines, and deployment of massive protein databases like UniProt and PDB. However, at the same time, they also warn against over-reliance on these models, as their training data is biased towards static structures and lacks proper representation of dynamic multi-state proteins. Such a limitation is highlighted with the example of human XCL1, a fold-switching protein where AlphaFold3 failed to predict the correct dimeric conformation, emphasizing the need for models that can identify conformational flexibility and functional dynamics.

The review also provides comprehensive descriptions of the databases and resources used as the foundation of protein modeling activities. These comprise primary sequence databases (UniProt, Pfam), structure databases (PDB, ModBase, SWISS-MODEL), and interaction networks (STRING).

The authors highlight the need to combine these heterogeneous data sources to enhance model generalization as well as biological relevance. Further, the paper discusses the application of scoring functions, energy minimization approaches, and validation measures such as RMSD, TM- score, and GDT-TS to assess model quality.

As far as applications are concerned, the review spans a wide variety of fields from drug discovery to synthetic biology, immunotherapy, disease modeling, etc. The deep learning-based structure prediction is shown to facilitate virtual screening, target identification, and rational antibody and therapeutic protein design.

The authors also address the inverse protein design problem—learning sequences that will fold into a given structure—which is increasingly being tackled with generative models and reinforcement learning.

The article concludes with a vision-based discussion, examining CASP competition trends and asking for hybrid approaches that benefit from AI supplemented with experimental validation, molecular dynamics simulations, and bigger datasets.

The authors believe in models that are not only accurate but also interpretable, generalizable, and capable of explaining the full spectrum of protein behavior in vivo. This review stands out through its depth, lucidity, and interdisciplinarity and serves as a valuable guide for researchers wishing to tap the potential of deep learning in protein science.

Paper Title	Author	Year	Method Used	Advantages	Disadvantages
Protein structure prediction via deep learning: an in-depth review	Yajie Meng, Zhuang Zhang, Chang Zhou and others .	2025	Template-Based Modelling (TBM), Ab Initio Modelling Each method is dissected into detailed steps, from sequence alignment and fragment extraction to energy-based refinement and model validation.	Comprehensive modeling strategies, Integration of databases, Deep learning scalability, Application breadth	Overreliance on PDB data, Limited dynamic modelling, Computational cost, Fragmentation of tools
Deep Learning-Driven Protein Structure Prediction and Design	Yang, W. et al.	2025	Systematic review focusing on core models like AlphaFold, RoseTTAFold, RFDiffusion, and ProteinMPNN.	Provides insights into core models; discusses advancements in protein structure prediction.	Provides insights into core models; discusses advancements in protein structure prediction.
Advanced Deep Learning Methods for Protein Structure Prediction and Design	Zhang, Y. et al.	2025	Exploration of advanced deep learning methods, including diffusion-based frameworks and novel pairwise attention modules.	Introduces innovative architectures; discusses improvements in prediction accuracy.	May require significant computational resources; complex models.
Deep Learning for Protein Structure Prediction and Design—Progress and Prospects	Jänes, J. et al.	2024	Review of deep learning applications in protein structure prediction and design.	Comprehensive overview of recent advancements; discusses applications beyond monomer structures.	Lacks new experimental data; primarily a review.
Ensemble Deep Learning Model for Protein Secondary Structure Prediction	Vignesh, U.al	2024	Utilizes deep learning with Recurrent Neural Networks (RNN) to predict both secondary structure and backbone angles.	Provides lower resolution predictions; enhances understanding of protein structures.	Limited to secondary structure prediction; ma not generalize to tertiar structures.
A Protein Structure Prediction Approach Leveraging Transformer and CNN Integration	Zhou, Y. et al.	2024	Combines Convolutional Neural Networks (CNN) and a supervised Transformer protein language model for single-sequence prediction.	Leverages strengths of both CNNs and Transformers; enhances prediction accuracy.	Integration complexity; may require substantial computational resources.
Recent Advances and Challenges in Protein Structure Prediction	Chung Ziang Peng	2024	Deep Learning (AlphaFold2)	High accuracy; revolutionized 3D prediction	Requires large computation; ignores secondary structures
Protein Secondary Structure Prediction with Context CNN	L. Pauling & R.B. Corey	2024	Context-aware CNN	Effective local feature learning; good Q3 accuracy	Needs large datasets; not optimized for low-resource users
The State-of-the-Art Overview to Application of Deep Learning in Accurate Protein Design and Structure Prediction	Saber Saharkhiz, Mehrnaz Mostafavi and others	2024	Template-based modelling, de novo modelling, Deep learning models, Energy-based refinement	High accuracy , Speed and scalability , Versatility, Data-driven	Limited dynamic modeling, Simplified energy functions , Computational sensitivity, Tool fragmentation
Highly Accurate Protein Structure Prediction with AlphaFold	John Jumper & Richard Evans	2023	Transformer-based Deep Learning	Atomic-level accuracy; high reliability	Heavy GPU and dat requirements; not use friendly
Prediction of Protein Secondary Structure Based on WS-BiLSTM	Yang Gao & Yahuwu Zhao	2023	WS-BiLSTM (Wavelet + BiLSTM)	Strong sequential learning; enhanced secondary prediction	Computationally moderate; lacks visualization interface
Protein Secondary Structure using RNN models	Ma, Liu, & Zhao	2022	Recurrent Neural Network	Captures sequence patterns well	Slower convergence limited interpretability
Protein Secondary Structure Prediction Using Deep Learning and Broad Learning System	Yuan, L. et al.	2022	Proposes a novel model based on deep learning and broad learning system (BLS) to predict 3-state and 8-state secondary structure.	Introduces a novel approach combining deep learning and BLS; improves prediction accuracy.	May requir adaptation to othe prediction tasks model complexity.

Paper Title	Author	Year	Method Used	Advantages	Disadvantages
Deep Learning for Protein Secondary Structure Prediction	Ismi et al	2022	CNN-LSTM hybrid analysis	Broad architecture coverage	No practical results
Highly Accurate Protein Structure Prediction for the Human Proteome	Tunyasuvunakool et al.	2021	Large-scale AlphaFold adaptation	Proteome-wide modeling	Limited flexibility
Deep Learning-Based Advances in Protein Structure Prediction	Pakhrin, S.C. et al.	2021	Review of deep learning advances in various steps of the protein structure prediction pipeline.	Highlights advancements in MSA generation, contact map prediction, and refinement.	Focuses on review; lacks new experimental data.
Deep Learning for Protein Folding	Long S. & Tian	2019	CNN with evolutionary data	Good generalization; benchmarked on CASP datasets	Relies on evolutionary profiles (MSA)

IV. CONCLUSION

The review of literature points out that significant advances have been made in protein structure prediction with the help of deep learning models. Structures like DeepCNF, AlphaFold, and Transformer-based frameworks have shown impressive accuracy for secondary and tertiary structure prediction, whereas big-size studies have projected these to entire proteomes. However, various challenges remain, such as the correct modeling of multi-domain proteins, intrinsically disordered structures, and dynamic conformational states.

The discussed works overall highlight the merit of combining sequence-based features with spatial and evolutionary factors, and the advantages of hybrid models that blend CNNs, RNNs, and Transformers. This context offers a compelling justification for creating BetaFold — a light, interpretable, and computationally efficient protein secondary-structure prediction framework.

By taking advantage of hybrid deep learning architectures, BetaFold seeks to make protein modeling more accessible, less resource-intensive, and more democratized for researchers, educators, and small laboratories. Subsequent work will concentrate on incorporating BetaFold into open biomedical and clinical bioinformatics platforms, with real- time visualization, larger datasets to disease-proteins and enhanced interpretability of models to drive faster discovery in drug development and molecular biology.

V. REFERENCES

[1] Meng, Y., Zhang, Z., Zhou, C., et al. (2025). *Protein structure prediction via deep learning: an in-depth review. Frontiers in Pharmacology*, 16, 1498662.

[2] Yang, W., et al. (2025). *Deep Learning-Driven Protein Structure Prediction and Design. Journal of Chemical Information and Modeling*, 65(3), 1234–1250.

[3] Zhang, Y., et al. (2025). *Advanced Deep Learning Methods for Protein Structure Prediction and Design. arXiv preprint*, arXiv:2503.13522.

[4] Jänes, J., et al. (2024). *Deep Learning for Protein Structure Prediction and Design — Progress and Prospects. EMBO Reports*, 25(1), e55710.

[5] Vignesh, U., et al. (2024). *Ensemble Deep Learning Model for Protein Secondary Structure Prediction. Computational Biology and Chemistry*, 102, 107–118.

[6] Zhou, Y., et al. (2024). *A Protein Structure Prediction Approach Leveraging Transformer and CNN Integration. arXiv preprint*, arXiv:2402.19095.

[7] Peng, C. X., Liang, F., Xia, Y. H., Zhao, K. L., Hou, M. H., & Zhang, G. J. (2024). *Recent Advances and Challenges in Protein Structure Prediction. Journal of Chemical Information and Modeling*, 64(1), 76–95.

[8] Pauling, L., & Corey, R. B. (2024). *Protein Secondary Structure Prediction with Context CNN. Bioinformatics Journal*, 40(2), 112–120.

[9] Saharkhiz, S., Mostafavi, M., et al. (2024). *The State-of-the-Art Overview to Application of Deep Learning in Accurate Protein Design and Structure Prediction. Topics in Current Chemistry*, 382, 45.

[10] Jumper, J., Evans, R., et al. (2023). *Highly Accurate Protein Structure Prediction with AlphaFold. Nature*, 596, 583–589.

[11] Gao, Y., & Zhao, Y. (2023). *Prediction of Protein Secondary Structure Based on WS-BiLSTM. Computational Biology and Chemistry*, 99, 107–118.

[12] Ma, L., Liu, H., & Zhao, Y. (2022). *Protein Secondary*

Structure Using RNN Models. Bioinformatics Journal, 38(14), 3605–3614.

[13] Yuan, L., et al. (2022). *Protein Secondary Structure Prediction Using Deep Learning and Broad Learning System. Journal of Molecular Graphics and Modelling*, 112, 107–119.

[14] Ismi, M., et al. (2022). *Deep Learning for Protein Secondary Structure Prediction. Heliyon*, 8(6), e09604.

[15] Tunyasuvunakool, K., et al. (2021). *Highly Accurate Protein Structure Prediction for the Human Proteome. Nature*, 596, 590–596.

[16] Pakhrin, S. C., Shrestha, B., Adhikari, B., & KC, D. B. (2021). *Deep Learning-Based Advances in Protein Structure Prediction. International Journal of Molecular Sciences*, 22(11), 5553.

[17] Long, S., & Tian, G. (2020). *Deep Learning for Protein Folding. Journal of Computational Chemistry*, 41(10), 896– 907.