

Advancements and Applications of Protein Structure Prediction Algorithms

Ye Chen

Biochemistry, University College London, London, England

zcbty97@ucl.ac.uk

Abstract. Protein structure prediction serves as a foundational aspect of molecular biology, where computational advancements have recently propelled significant increases in prediction accuracy. This paper evaluates traditional protein structure prediction methods, including homology modeling, threading, and Ab Initio techniques, emphasizing the inherent challenges these methods face in accurately modeling novel and highly flexible proteins. With the advent of AI-based models, particularly AlphaFold, the landscape of protein structure prediction has undergone a transformative shift. AlphaFold integrates evolutionary data with sophisticated deep learning algorithms to achieve accuracies close to experimental results. Despite such progress, AlphaFold grapples with challenges in dynamic regions of proteins, such as intrinsically disordered proteins (IDPs), and struggles to accurately predict the structural impacts of genetic mutations. The paper also explores novel emerging methods, such as protein language models, designed to address these specific limitations. These cutting-edge approaches not only enhance the accuracy of protein structure predictions but also broaden the scope of potential applications. Areas such as drug discovery and vaccine development stand to benefit immensely from these advancements, potentially expediting critical breakthroughs. This examination of both established and novel methodologies illuminates the ongoing evolution of protein structure prediction, pointing towards future innovations that may further revolutionize the field.

Keywords: AlphaFold, Intrinsically disordered proteins, Drug discovery.

1. Introduction

Research Background: Protein structure prediction is pivotal in molecular biology, underscored by its critical role in understanding protein function and the implications for health when proteins misfold, potentially leading to diseases like Alzheimer's and cystic fibrosis [1-3]. The foundational hypothesis set by Christian Anfinsen posits that a protein's amino acid sequence dictates its three-dimensional structure, which in turn determines its function [3]. Traditional methods for studying these structures, such as X-ray crystallography and NMR spectroscopy [4, 5], have significantly contributed to our knowledge, albeit with limitations in cost, time, and the resolution of dynamic folding processes. The challenge of keeping pace with the exponential growth of sequence data, juxtaposed with a relatively slower growth in the experimental resolution of structures, emphasizes the necessity for advanced computational approaches [6, 7, 8].

Current Research Status: Despite the vast number of protein sequences available in databases like UniProtKB, a significant gap remains between the sequences known and those with resolved tertiary structures [7, 8]. This disparity is largely due to the complex, costly, and time-intensive nature of traditional experimental methods that often fail to represent natural protein conformations and dynamics effectively [9]. The Levinthal paradox highlights these challenges by questioning how proteins fold so quickly given their conformational freedom [10]. Addressing these gaps, computational methods, including AI-driven algorithms, have emerged as essential tools. These include both template-based approaches, leveraging existing structural data, and template-free methods that predict structures from primary amino acid sequences [9]. The development of deep learning models like AlphaFold represents a significant leap forward, offering predictions with near-experimental accuracy and the potential to transform how proteins are studied and applied in biomedical research.

Research Content of This Paper: This paper delves into the core challenges of protein folding and structure prediction, with a focus on the integration of deep learning technologies in this field. It reviews current methodologies, from homology modeling and threading to advanced AI algorithms like AlphaFold, evaluating their efficacy, limitations, and potential applications in drug discovery and vaccine development. Despite the advances, challenges persist, particularly in accurately predicting structures of complex proteins and multi-protein assemblies [9, 10]. This study aims to assess how combining big data technology with structural biology can not only enhance the predictive accuracy but also streamline drug development processes. Furthermore, it explores the potential of future predictive models to revolutionize personalized medicine and precision therapies, leveraging vast datasets to tailor biomedical solutions to individual genetic profiles. This paper seeks to provide a comprehensive analysis of the current landscape and future directions in protein structure prediction, emphasizing the transformative impact of deep learning and big data in the field.

2. Traditional Protein Structure Prediction Methods

Protein folding is an inherently complex process. Even though most current computer modeling techniques have achieved high experimental accuracy and vast data storage capabilities, no algorithm has yet been able to perfectly predict the natural folding of proteins [11]. Traditional protein folding structure prediction techniques are primarily based on three modeling approaches: homology modeling, threading/fold recognition, and the Ab Initio method. Both homology modeling and threading/fold recognition rely on known protein structure templates, while the Ab Initio method does not depend on pre-existing templates. Instead, it predicts possible folding structures directly from the protein sequence by combining energy functions with conformational sampling. The underlying principle of these methods is based on the relationship between protein evolution and folding. Homology modeling and threading methods predict the target protein structure by assuming conservation in the amino acid sequence or fold during the protein's evolutionary process. In contrast, the Ab Initio method simulates the physical process of protein folding to predict the target protein structure [12].

2.1. Homology modeling

Homology modeling uses known protein structures as templates to predict the structure of target proteins with high sequence similarity. This modeling approach assumes that similar sequences in homologous proteins will fold into similar structures and that structural evolution tends to be more conserved than sequence evolution, especially in functionally critical core regions [13]. This means that during the evolutionary process of homologous proteins, structurally conserved regions (SCRs) are more likely to be preserved, while variable loop regions or residues with greater flexibility tend to evolve independently. If certain residues are similar in terms of size and hydrophobicity, they are likely to replace one another during evolution. Even if parts of the sequence change throughout evolution, the resulting three-dimensional structure remains largely similar [14].

When the target protein and template protein share more than 25% sequence similarity, they are likely to be evolutionarily homologous. Homology modeling fragments and combines the SCRs of the

template protein with the variable loop regions or active side chains of the target protein. The core structure of the target protein is built using conserved regions from the known template, while the variable regions are adjusted and optimized flexibly. The final fitted three-dimensional model can be superimposed on the native model using specific structural comparison programs, and its accuracy is evaluated by calculating the root-mean-square deviation (RMSD) of the C α atoms. The workflow for homology modeling includes the following steps: 1) searching for homologous template sequences in databases; 2) aligning the target sequence with the template sequence; 3) generating the conserved core regions (i.e., the backbone structure); 4) modeling the loops and side-chain regions; 5) optimizing the model following energy minimization; and 6) evaluating the model.

When a homologous template is available, homology modeling is the preferred method for predicting protein structure. In biomedical research, homology modeling assists scientists in predicting how mutations at different sites affect protein structures, thereby shedding light on potential disease mechanisms. For example, human tyrosinase plays a crucial role in melanin synthesis, and mutations at its active site can lead to albinism. However, the crystal structure of human tyrosinase is not available in the Protein Data Bank. To address this, Mubashir Hassan and his team used homology modeling in 2017 to predict and analyze eight mutations in the active binding region of tyrosinase and assess their impact on the structural stability of the enzyme [15]. The six histidine residues that form the active site of human tyrosinase bind to copper atoms to generate catalytic activity, and this active region is largely conserved across species. The crystal structure of tyrosinase from *Bacillus megaterium**, a bacterium with high sequence similarity to the target protein, was selected as a template. The study involved replacing residues at the active site (H180N, H202Q, H202R, H211R, H363Y, H367R, H367Y, and H390D), and molecular dynamics (MD) simulations were used to predict the effects of these mutations on the structure in a force field. The results showed that mutations at residues such as Q/R202 and Y/R363 were more likely to affect the stability and folding of the protein, thereby interfering with melanin biosynthesis pathways and leading to melanin-related disorders. The quality and accuracy of the model were evaluated using several online servers, such as MolProbity, ERRAT, and ProSA. Since neuromelanin in the brain is linked to neurodegeneration and Parkinson's disease, the mutated structures of human tyrosinase have also been explored as therapeutic targets for identifying potential inhibitors.

2.2. Threading/Fold recognition

Similar to homology modeling, threading (or fold recognition) primarily focuses on predicting the static folded state of a protein based on existing structural libraries, rather than simulating the dynamic folding process. While homology modeling assumes that similar amino acid sequences fold into similar three-dimensional structures, amino acid sequence alone is not the only determinant of protein folding. In nature, the number of unique folding structures is limited, and even in cases where there is low sequence identity, proteins can still fold into highly similar three-dimensional structures [16].

When sequence identity drops below 25%, threading offers higher accuracy than homology modeling. Threading assumes that the folding patterns of proteins are more conserved than their sequences, meaning that folding structures remain more stable throughout evolution. Unlike homology modeling, which relies on high sequence similarity, threading is applied to protein sequences with little similarity to known structures. By "threading" the target sequence into different folding structure templates, the method evaluates whether each possible fold matches the target sequence.

Residue environment classification is the core scoring criterion in threading methods. The 3D profile method defines different environment classes based on the secondary structure of amino acid residues and their interactions with the surrounding environment, such as solvent accessibility and whether the residue is located in a core or surface region. This approach translates the three-dimensional spatial position of a residue into a one-dimensional string. The residue environment classification is then used to evaluate whether the sequence fits appropriately into the template structure. If the same amino acid residue is located in different environments, the matching score may decrease [17]. By considering the environmental class of each residue, threading avoids producing

unreasonable alignment results. For example, hydrophobic residues should be buried within the protein core and not exposed to solvents. The basic workflow of the threading method includes the following steps: 1) selecting a template protein from the database; 2) designing a scoring function based on the template and template sequence; 3) "threading" the one-dimensional sequence through the known template structure library; and 4) evaluating and validating the fit.

When used together, threading and homology modeling can cover a wider range of sequence similarities. This combined approach has been applied to predict the structure and function of hypothetical proteins in *Mycoplasma hyopneumoniae* (*M. hyopneumoniae*). In databases, some proteins encoded by this bacterium's genes lack inferred functions, including those critical for its biological activities, such as ATP and NAD synthetase activities. The study suggests that seven specific proteins (YP_287866, YP_287786, YP_287675, YP_287559, YP_288024, YP_287971, and YP_288034) are involved in metabolism and transcription processes in *M. hyopneumoniae*. When predicting the structure of the N-terminal region of YP_287866, the crystal structure of *Staphylococcus aureus* nicotinamide mononucleotide adenylyltransferase was used as a template. Even though the sequence identity between the two proteins was not high, the high score obtained in the threading method was due to the similarity in their folded topological structures [18].

2.3. *Ab initio* method

Although template-based modeling provides high accuracy, in many cases, the target protein lacks a suitable template for reference, especially for novel and unknown proteins. The *Ab Initio* method offers a reference model for proteins that are difficult to analyze using X-ray crystallography or NMR. Given the high cost and time-consuming nature of experimental structure determination, the rough models predicted by the *Ab Initio* method serve as a guide for subsequent, more precise measurements. The *Ab Initio* method is based on Anfinsen's thermodynamic hypothesis, which posits that the native folding state of a protein has the lowest free energy. This method simulates the physical interactions at the molecular level—such as van der Waals forces, hydrogen bonds, and hydrophobic effects—allowing for an extensive search of the conformational space until the three-dimensional fold with the global minimum free energy is found [19]. However, the vast computational demands of the *Ab Initio* method limit its applicability to long protein sequences.

One optimization approach is to search for similar sequence fragments and model them separately. These fragments are then assembled into a complete model based on simulated force fields and interactions, thereby reducing the computational burden. The length and number of fragments used influence the model's accuracy, and the fragments themselves can be based on known structures from databases [20]. It is important to note that while the *Ab Initio* method involves physical simulation of the protein folding process, its goal is to predict the native state of the protein, rather than to trace the path it takes to achieve correct folding.

To efficiently sample the correct conformation among numerous folding patterns, the *Ab Initio* method requires a function capable of high-efficiency conformational sampling. This sampling function is typically based on algorithms like simulated annealing or Monte Carlo, which are designed to search for global optimal solutions. Since the stability of a protein's native structure is only 5-10 kcal/mol higher than its denatured state, there are many possible conformations between the denatured and native states. Therefore, folding model predictions must account for a variety of energy considerations, especially those between residues. For example, hydrophobic interactions drive non-polar residues to aggregate in polar solvent environments, while polar residues are influenced by van der Waals forces and hydrogen bonding, leading to either attraction or repulsion. The typical workflow of the *Ab Initio* method includes the following steps: 1) determining the amino acid sequence of the target protein; 2) searching the conformational space; 3) evaluating the energy function; 4) finding the conformation with the lowest free energy and optimizing the search results; and 5) outputting the final folded structure.

The boundary between template-based and template-free modeling is often blurred, as the inter-residue interactions referenced in *Ab Initio* methods can also be inferred from the evolutionary

correlations within protein sequences. Thomas A. Hopf and his team successfully applied a hybrid Ab Initio method to blindly predict the three-dimensional structures of 11 known transmembrane proteins. Since more than a quarter of human proteins contain membrane structures, accurate prediction of membrane protein structures has significant medical implications. A detailed understanding of conformational changes and active sites in membrane proteins can aid the field of molecular chemistry in identifying better drug targets. Compared to purely homology modeling or Ab Initio methods, the prediction approach developed in this experiment, known as EVfold-membrane, does not rely on known protein structure templates or purely physical-chemical laws. Instead, it uses statistical amino acid co-evolution relationships to predict protein structures. By incorporating evolutionary constraints, this method significantly reduces the enormous computational burden typically associated with conformational searches in traditional Ab Initio methods. The final predicted models showed high consistency with actual crystal structures, and the prediction accuracy was even higher for regions critical to protein function than for the overall protein structure [21].

3. AI-Based Protein Structure Prediction Methods

3.1. AlphaFold

Traditional methods such as homology modeling, threading, and Ab Initio have laid the groundwork for computer-based drug design through protein structure prediction. To encourage the development of protein structure prediction algorithms, the Critical Assessment of Structure Prediction (CASP) initiative was established in 1994. This event provides participants with proteins whose three-dimensional structures are known but not publicly disclosed, allowing research teams worldwide to test and evaluate their prediction algorithms. In 2020, AlphaFold, developed by DeepMind, demonstrated remarkable accuracy in the CASP14 assessment, far surpassing other methods. AlphaFold's precision in both backbone and all-atom predictions was significantly superior to previous structure prediction algorithms. The model has since been widely validated using recently published PDB structures, offering excellent references for experimental structure determination, protein function analysis, and proteome-wide predictions [22].

AlphaFold integrates the strengths of traditional prediction methods by combining evolutionary information and deep learning to predict the tertiary structure of proteins. This method uses neural networks to predict inter-residue distances, thereby inferring structural information, and employs co-evolutionary constraints and gradient descent algorithms to optimize the predicted protein structure [23]. The input to the AlphaFold network includes the primary amino acid sequence of the target protein as well as a multiple sequence alignment (MSA) of homologous proteins. The network's core utilizes an innovative Evoformer module, which effectively integrates MSA information and residue-pair data to infer both spatial and evolutionary relationships in proteins. The Evoformer module overcomes the limitations of traditional methods, such as dependence on templates or complex physical calculations, allowing AlphaFold to handle more complex protein structure problems and achieve accuracy that approaches experimentally determined structures. The main workflow for AlphaFold's protein structure prediction includes: 1) inputting the amino acid sequence of the target protein; 2) constructing MSA information, including alignment sequences from homologous proteins; 3) using neural network modules to generate high-precision three-dimensional structures based on the input information; 4) iteratively refining the predicted structure; and 5) outputting the predicted three-dimensional coordinates of the target protein.

Even though AlphaFold's prediction accuracy is close to that of experimental models, it is not meant to operate independently of laboratory-based protein measurement techniques. The models predicted by AlphaFold still require experimental validation, and they can complement traditional methods to improve the efficiency and accuracy of predictions. In 2022, Dylan P. Noone and colleagues combined cryo-electron microscopy (Cryo-EM), mass spectrometry, and AlphaFold's structural prediction methods to determine the complete structural model of the long pentraxin PTX3. PTX3 has demonstrated antiviral properties in mouse models of COVID-19 infection. As an important

part of innate immune recognition, PTX3 lacked high-resolution structural data from laboratory measurements. The C-terminal domain of PTX3 had been observed through Cryo-EM, but the flexibility of the N-terminal region reduced the resolution of the structure. In this experiment, AlphaFold was used to model the N-terminal region, which was then combined with the known C-terminal structure to create a complete model of PTX3. This model revealed the octameric structure of PTX3, consisting of a core domain made up of eight PTX domains and a flexible N-terminal region composed of two coiled-coil tetramers. These structural details help to understand the multifunctionality of PTX3 in complement activation, antiviral activity, and its role in the extracellular matrix [24].

During the subsequent COVID-19 pandemic, AlphaFold models were used to study the pathogenic proteins of SARS-CoV-2 in search of potential vaccine targets. SARS-CoV-2, the causative agent of COVID-19, is a coronavirus, and vaccine development against it has primarily focused on the spike protein, which mediates viral infection. Variations in the spike protein's structure may influence the virus's infectivity and transmissibility, and vaccines need to account for stable mutations in the spike protein [25]. Similar structure-based reverse vaccinology approaches have been applied in the development of vaccines for viruses such as influenza, human immunodeficiency virus (HIV), and respiratory syncytial virus (RSV). In vaccine design, AlphaFold's core algorithm involves a global analysis of sequence covariance. In protein sequences, covariance analysis is often used to evaluate alignment results across multiple sequences, inferring whether two amino acid residues co-evolve. At the heart of the host-pathogen struggle is covariance. The structures predicted by AlphaFold must keep pace with the evolutionary changes in the pathogen to prevent immune evasion. Since viral surface molecules are constantly changing, understanding how these changes occur in response to environmental factors through AlphaFold models is even more crucial than simply predicting static structures.

3.2. Limitations of AlphaFold

Despite its achievements, AlphaFold still has some limitations, such as its insensitivity to mutations within the input amino acid sequence [26]. Missense mutations, caused by base pair substitutions, are common in nature and can alter a protein's conformation, stability, and resistance, such as in the formation of antibody-antigen complexes. These structural disruptions are often associated with diseases. Predicting the structural impact of point mutations using AlphaFold requires inputting the mutated sequence into the database. However, there is no specific database for mutations that cause structural disruptions, and AlphaFold's initial input is primarily based on evolutionarily constrained homologous models. As a result, the accuracy of AlphaFold's predictions for protein structures after missense mutations is limited.

AlphaFold utilizes multiple sequence alignment (MSA) to analyze the conservation and variation across homologous protein sequences during the prediction process. Typically, orthologous gene sequences exhibit a degree of conservation, but certain regions may display high variability and conformational flexibility, such as intrinsically disordered proteins (IDPs) and loops. AlphaFold sometimes predicts disordered regions as more structured helices or underestimates their confidence scores. Many proteins in living organisms exhibit such unstable structures when isolated, and IDPs play crucial biological roles, particularly in signal transduction, regulation, and protein-protein interactions. Due to their flexibility, IDPs can bind to various molecules and play important roles in cellular regulation. AlphaFold treats IDPs as unstructured regions, leading to an overlap between disordered regions and domains with low confidence scores. This can result in missing important structural elements [27]. Additionally, loop regions often correspond to missing sections in protein structures from databases. These loops tend to be located on the protein surface, making them prone to involvement in protein-protein interactions. AlphaFold frequently overpredicts loop regions, especially when the loops exceed 20 residues in length [28]. Consequently, AlphaFold has limited accuracy in predicting structurally flexible regions that evolve over time.

New modeling approaches have been developed to address some of AlphaFold's limitations. Protein language models, built upon the foundation of AlphaFold, have shown promising potential in biotechnological applications. The core concept of protein language models is to draw an analogy between common amino acid sequences and words in a natural language, with the entire protein sequence resembling a sentence. By applying this analogy, researchers utilize techniques from natural language processing (NLP), such as autoregressive models, bidirectional models, and masked models, to analyze, model, and predict protein structure and function. The innovation of these language models lies in the neural network architecture they employ, specifically the Transformer model. This model uses an attention mechanism to infer dependencies between "contextual" sequences. Unlike AlphaFold, which relies heavily on MSA and co-evolutionary information from homologous proteins, protein language models explore tertiary structure using only the primary sequence as the starting point. This avoids some of the biases that arise in AlphaFold's predictions due to its dependence on MSA and co-evolution data. Additionally, these language models offer greater efficiency in structure retrieval [29].

4. Conclusion

This paper has explored the pivotal advancements AlphaFold has brought to protein structure prediction, acknowledging its profound impact on the accuracy of predicting complex protein structures. Despite its achievements, AlphaFold's limitations, particularly its dependence on multiple sequence alignments and challenges with intrinsically disordered proteins and point mutations, indicate areas requiring further enhancement. The exploration into novel computational models such as protein language models has demonstrated potential to address some of these shortcomings by prioritizing primary sequence data and reducing biases associated with co-evolutionary analysis.

The field of structural biology stands at a promising juncture where the fusion of experimental methods with AI-driven technologies, including and extending beyond AlphaFold, will be instrumental. Future research should focus on integrating these advanced computational models with traditional experimental techniques, creating a synergistic approach that enhances our understanding of protein dynamics, interactions, and the effects of mutations at a granular level. Such integration will not only refine the accuracy of structural predictions but also expand the potential applications in drug design and the broader understanding of biological mechanisms. Continuing to develop and refine AI-driven approaches, while addressing their current limitations, will significantly propel the capabilities of predictive models, fostering innovations in precision medicine and therapeutic interventions.

References

- [1] Nassar, R., Dignon, G. L., Razban, R. M., & Dill, K. A. (2021). The protein folding problem: The role of theory. *Journal of Molecular Biology*, 433(20), 167126.
- [2] Bertoline, L. M. F., Lima, A. N., Krieger, J. E., & Teixeira, S. K. (2023b). Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*, 3, 1120370.
- [3] Al-Janabi, A. (2022). Has DeepMind's AlphaFold solved the protein folding problem? *BioTechniques*, 72(3), 73–76.
- [4] Zimmer, M. (2020, December 2). AI makes huge progress predicting how proteins fold – one of biology's greatest challenges – promising rapid drug development.
- [5] Waman, V. P., Sen, N., Varadi, M., Daina, A., Wodak, S. J., Zoete, V., Velankar, S., & Orengo, C. (2020). The impact of structural bioinformatics tools and resources on SARS-CoV-2 research and therapeutic strategies. *Briefings in Bioinformatics*, 22(2), 742–768.
- [6] De Almeida Paiva, V., De Souza Gomes, I., Monteiro, C. R., Mendonça, M. V., Martins, P. M., Santana, C. A., Gonçalves-Almeida, V., Izidoro, S. C., De Melo-Minardi, R. C., & De Azevedo Silveira, S. (2022). Protein structural bioinformatics: An overview. *Computers in Biology and Medicine*, 147, 105695.

- [7] RCSB Protein Data Bank. (2024). PDB statistics. RCSB Protein Data Bank. Retrieved from <https://www.rcsb.org/stats>
- [8] The UniProt Consortium. (2024). UniProt: The universal protein knowledgebase. UniProt. Retrieved from <https://www.uniprot.org/uniprotkb/statistics>
- [9] Bertoline, L. M. F., Lima, A. N., Krieger, J. E., & Teixeira, S. K. (2023c). Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*, 3, 1120370.
- [10] Ferina, J., & Daggett, V. (2019). Visualizing protein folding and unfolding. *Journal of Molecular Biology*, 431(8), 1540–1564.
- [11] Outeiral, C., Nissley, D. A., & Deane, C. M. (2022). Current structure predictors are not learning the physics of protein folding. *Bioinformatics*, 38(7), 1881–1887.
- [12] Yuan, X., Shao, Y., & Bystroff, C. (2003). AB initio protein structure prediction using pathway models. *Comparative and Functional Genomics*, 4(4), 397–401.
- [13] Agnihotry, S., Pathak, R. K., Singh, D. B., Tiwari, A., & Hussain, I. (2022). Protein structure prediction. In Elsevier eBooks (pp. 177–188).
- [14] Wang R., Zhu J., Wang S., Wang T., Huang J., Zhu X. Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking. *International Journal of Multimedia Information Retrieval*, 2024, 13(4): 39.
- [15] Hassan, M., Abbas, Q., Raza, H., Moustafa, A. A., & Seo, S. (2017). Computational analysis of histidine mutations on the structural stability of human tyrosinases leading to albinism insurgence. *Molecular BioSystems*, 13(8), 1534–1544.
- [16] Bhattacharya, S., Roche, R., Shuvo, M. H., Moussad, B., & Bhattacharya, D. (2023). Contact-assisted threading in low-homology protein modeling. *Methods in Molecular Biology*, 41–59.
- [17] Singh, D. B., & Pathak, R. K. (2020). Computational approaches in drug designing and their applications. In Springer Protocols Handbooks/Springer Protocols (pp. 95–117).
- [18] Da Fonsêca, M. M., Zaha, A., Caffarena, E. R., & Vasconcelos, A. T. R. (2011). Structure-based functional inference of hypothetical proteins from *Mycoplasma hyopneumoniae*. *Journal of Molecular Modeling*, 18(5), 1917–1925. <https://doi.org/10.1007/s00894-011-1212-3>
- [19] Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. *Science*, 294(5540), 93–96.
- [20] Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., & Wang, R. (2024). A Review of Key Technologies for Emotion Analysis Using Multimodal Information. *Cognitive Computation*, 1-27.
- [21] Hopf, T. A., Colwell, L. J., Sheridan, R., Rost, B., Sander, C., & Marks, D. S. (2012). Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell*, 149(7), 1607–1621.
- [22] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
- [23] Corum, M. R., Venkannagari, H., Hryc, C. F., & Baker, M. L. (2024). Predictive modeling and cryo-EM: A synergistic approach to modeling macromolecular structure. *Biophysical Journal*, 123(4), 435–450.
- [24] Zhu, X., Huang, Y., Wang, X., & Wang, R. (2023). Emotion recognition based on brain-like multimodal hierarchical perception. *Multimedia Tools and Applications*, 1-19.
- [25] Higgins, M. K. (2021). Can we AlphaFold our way out of the next pandemic? *Journal of Molecular Biology*, 433(20), 167093.
- [26] Buel, G. R., & Walters, K. J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure? *Nature Structural & Molecular Biology*, 29(1), 1–2.

- [27] Ruff, K. M., & Pappu, R. V. (2021). AlphaFold and implications for intrinsically disordered proteins. *Journal of Molecular Biology*, 433(20), 167208.
- [28] Stevens, A. O., & He, Y. (2022). Benchmarking the accuracy of AlphaFold 2 in loop structure prediction. *Biomolecules*, 12(7), 985.
- [29] Ruffolo, J. A., & Madani, A. (2024). Designing proteins with language models. *Nature Biotechnology*, 42(2), 200–202.