Benchmarking AlphaFold-Class Tools on Recently Solved Membrane-Protein Complexes

1. Introduction:

Membrane proteins, critical components of cellular machinery, perform a diverse array of functions essential for life, including transport of molecules, reception of signals from the environment, mediation of ion flow across membranes, and catalysis of biochemical reactions.[1] These proteins constitute a significant portion of the proteome across various organisms and are of particular interest as therapeutic targets for a wide range of diseases.[2] Furthermore, many membrane proteins function as part of larger complexes, where their interactions regulate fundamental biological processes such as signal transduction, immune responses, energy production, and nutrient uptake.[7] The sheer number and functional diversity of membrane proteins and their complexes underscore the critical need for detailed structural information to understand their mechanisms of action and to facilitate the development of targeted therapies.

Despite their biological significance, the experimental determination of three-dimensional structures for membrane proteins and their complexes poses considerable challenges.[1] These proteins are inherently hydrophobic, making them difficult to handle outside their native lipid environment, often leading to aggregation or instability that hinders structural studies.[1] While techniques like X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron cryomicroscopy (cryo-EM) have been instrumental in elucidating protein structures, they can be complex, time-consuming, and expensive, and may not always capture the protein in its most biologically relevant conformation.[1] Consequently, the number of experimentally resolved structures for membrane proteins, and especially their complexes, remains relatively small in comparison to the vast number of known protein sequences.[1] This disparity highlights the critical need for accurate and reliable computational methods capable of predicting these structures directly from their amino acid sequences.

The field of protein structure prediction has experienced a transformative breakthrough with the advent of deep learning-based tools, most notably AlphaFold2, developed by Google DeepMind.[1] This AI system demonstrated an unprecedented ability to predict protein structures with near-experimental accuracy, marking a significant advancement in addressing the long-standing protein folding problem.[16] Following this success, other powerful tools such as RoseTTAFold, from the Baker lab, and OpenFold, an open-source implementation of AlphaFold2, have emerged, offering comparable or enhanced capabilities for protein structure prediction and complex modeling.[24] Furthermore, extensions like AlphaFold-Multimer have been specifically developed to predict the structures of multi-chain protein complexes, providing new opportunities for understanding protein-protein interactions.[7] These AlphaFold-class tools represent a

significant leap in our ability to computationally determine protein structures, including those of complexes, potentially revolutionizing biological research and drug discovery, particularly for challenging systems like membrane proteins.

Given the remarkable progress in protein structure prediction by AlphaFold-class tools, it is crucial to specifically evaluate their performance on membrane protein complexes, which present unique challenges due to their environment and often limited sequence homology.[6] Rigorous benchmarking against recently solved experimental structures of membrane protein complexes is essential to understand the current capabilities and limitations of these computational methods in this specific context.[1] This report aims to benchmark the performance of AlphaFold-class tools, including AlphaFold2, AlphaFold-Multimer, RoseTTAFold, and OpenFold, on membrane protein complexes for which high-resolution experimental structures have been solved and released in the Protein Data Bank between 2022 and 2024. The scope includes an overview of these tools, a discussion of the challenges in modeling membrane protein complexes, an analysis of recent experimental advancements, a critical evaluation of the tools' performance using relevant structural assessment metrics, and a discussion of their limitations and future directions. Case studies of specific recently solved membrane protein complexes will be presented to illustrate the capabilities and shortcomings of these prediction methods in detail.

2. Overview of AlphaFold-Class Protein Structure Prediction Tools:

AlphaFold, developed by Google DeepMind, is an AI system that has significantly advanced the field of protein structure prediction by predicting 3D structures from amino acid sequences.[17] The second iteration, AlphaFold2 (AF2), utilizes a deep neural network architecture, including the Evoformer module for processing multiple sequence alignments (MSAs) and pairwise residue information, and the Structure module, which employs invariant point attention (IPA) to generate geometrically accurate 3D models.[7] AF2 achieved performance comparable to experimental methods in many cases.[17] To address the complexity of protein-protein interactions, AlphaFold-Multimer was developed as an extension of AF2, specifically trained to predict the structures of multi-chain protein complexes using paired MSAs and modified algorithms to capture inter-chain interactions.[7] The latest version, AlphaFold3 (AF3), expands the prediction capabilities to include DNA, RNA, ligands, and ions by employing a diffusion-based architecture and a broader training dataset.[7] The evolution of AlphaFold demonstrates a continuous effort to enhance accuracy and broaden the range of biomolecules that can be modeled, with significant implications for the study of complex and membrane protein systems.

RoseTTAFold, developed by the Baker lab, offers another powerful deep learning approach to protein structure prediction.[24] Its unique "three-track" neural network simultaneously considers protein sequences, inter-residue distances and orientations, and 3D coordinates, allowing for collective reasoning about the relationship between a protein's chemical components and its folded structure.[26] RoseTTAFold's accuracy approaches that of AlphaFold2, and it can rapidly generate accurate models of complex

biological assemblies, including protein-protein complexes, directly from sequence information, bypassing the need for separate subunit modeling and docking.[16] Furthermore, RoseTTAFold All-Atom extends these capabilities to model a wider array of biomolecules, including proteins, DNA, RNA, small molecules, and ions, making it particularly useful for studying membrane protein complexes and their interactions with various ligands and cofactors.[25]

OpenFold is an open-source project that provides a faithful and trainable PyTorch-based reproduction of DeepMind's AlphaFold2.[24] Its aim is to democratize access to state-of-the-art protein modeling tools by providing not only the inference code and model parameters but also the full training code under a permissive license.[28] OpenFold closely replicates the features of the original AlphaFold2 inference code, with optimizations for speed and memory efficiency, allowing for faster inference on GPUs and the ability to handle very long protein chains and large complexes.[29] The OpenFold consortium has also developed tools like OpenFold-Multimer for improved complex modeling and OpenFold-SoloSeq for faster predictions without requiring MSAs.[28]

The architectural and algorithmic advancements in these AlphaFold-class tools reflect different strategies for tackling the protein folding problem. AlphaFold2 utilizes the Evoformer and IPA mechanisms [17], while RoseTTAFold employs a three-track network [26], and OpenFold focuses on replicating AlphaFold2 with efficiency improvements.[29] AlphaFold3 introduces a diffusion network and Pairformer architecture.[68]Understanding these differences is crucial for interpreting their predictions and appreciating their relative strengths and weaknesses in modeling various protein structures, including membrane protein complexes.

3. Challenges in Computational Modeling of Membrane Protein Complexes:

Membrane proteins exist within the unique environment of the lipid bilayer, which significantly influences their structure, stability, and interactions.[2] This lipid milieu, with its hydrophobic core and hydrophilic headgroups, creates a partitioning effect that dictates the arrangement of amino acid residues within membrane proteins.[3]Accurately representing this complex environment and its interactions with membrane proteins is essential for reliable structure prediction, a feature not inherently present in standard AlphaFold-class tools.

A significant challenge in modeling membrane protein complexes is representing their interactions with the lipid bilayer and their orientation within it. While AlphaFold2 can predict the overall fold of individual domains, it lacks inherent awareness of the membrane plane, which can lead to inaccuracies in the predicted orientations of transmembrane domains.[6] Tools like the TmAlphaFold database attempt to address this by predicting membrane localization for AlphaFold2 structures using algorithms like TMDET.[6]

Many membrane protein complexes rely on cofactors, ligands, and post-translational modifications for their proper function and stability.[7] AlphaFold2 does not explicitly

model these interactions [7], although AlphaFold3 has expanded capabilities in this area.[7] The absence of explicit modeling can lead to inaccuracies in predicted structures that depend on these components for their native conformation.

Furthermore, AlphaFold-class tools primarily predict static structures and do not inherently capture the dynamic nature of proteins, including conformational changes that are often crucial for the function of membrane protein complexes.[7] Membrane proteins often undergo significant conformational changes during their functional cycles within the membrane environment.[7] The static nature of the predictions might not accurately represent the biologically relevant states of these complexes.

4. Recent Advancements in Experimental Structure Determination of Membrane Protein Complexes (2022-2024):

The period between 2022 and 2024 has seen significant progress in experimentally determining the structures of membrane protein complexes, largely driven by advancements in cryo-electron microscopy (cryo-EM).[1]Cryo-EM has become a powerful technique for resolving high-resolution structures of large and challenging membrane protein systems without requiring crystallization.[12] X-ray crystallography, with continued improvements in membrane protein biochemistry, also remains a valuable method for obtaining atomic-level details.[1] Techniques like solution NMR and HDX-MS provide complementary insights into the dynamics and interactions of these complexes within lipid environments.[1]

Experimental efforts have successfully elucidated the structures of various types of membrane protein complexes, including transporters like the *Trypanosoma cruzi* MscS channel [93] and human GAT3 [93], receptors such as GPCRs [89], ion channels [92], and enzymes like components of photosynthetic machinery.[89]Notably, studies have also revealed higher-order transient structures (HOTS) formed by membrane proteins, suggesting a novel organizational principle.[96]

Key recently solved membrane protein complex structures (2022-2024) with their PDB IDs and experimental methods include:

PDB ID	Protein Name(s)	Organism	Experimental Method	Release Date (if available)
9BGS	MscS	*Trypanosoma cruzi*	Cryo-EM	May 2025
9CP4	GAT3, 9D5 heavy chain, 9D5 light chain	*Homo sapiens*, *Mus musculus*	Cryo-EM	May 2025
9E7H	BchN, BchB	Unknown	Cryo-EM	May 2025
9NH8	CHD1, Nucleosome components	*Xenopus laevis*, *Homo sapiens*	Electron Microscopy	May 2025

9IWY	RIP3 kinase domain	*Mus musculus*	X-ray Crystallography	May 2025

This growing number of high-quality experimental structures provides a valuable resource for benchmarking the accuracy of computational prediction methods for membrane protein complexes.

5. Benchmarking AlphaFold-Class Tools on Recently Solved Membrane Protein Complexes:

Evaluating the accuracy of AlphaFold-class tools in predicting membrane protein complex structures requires the use of appropriate benchmarking metrics. These include the TM-score for assessing global structural similarity [8], DockQ and interface RMSD for evaluating the quality of protein-protein interfaces [8], and the confidence scores pLDDT and PAE provided by AlphaFold for local and global confidence estimation.[14] For AlphaFold-Multimer, the ipTM score specifically assesses the predicted interface quality.[56] A comprehensive evaluation should consider a combination of these metrics to assess different aspects of prediction accuracy.

Existing benchmark studies have provided insights into the performance of AlphaFold and AlphaFold-Multimer on various protein complexes, including some membrane proteins.[56] AlphaFold-Multimer has shown success in predicting near-native models for a portion of heterodimeric complexes but has faced challenges with certain types, such as antibody-antigen complexes.[56] Comparisons between AlphaFold and RoseTTAFold suggest that their performance can vary depending on the specific protein family and the availability of structural templates.[32] Specialized tools like Rosetta-MPDock have also been developed and benchmarked for transmembrane protein complex docking, sometimes showing improved results when used in conjunction with AlphaFold-Multimer.[9]

Initial studies utilizing recently solved structures (2022-2024) are beginning to emerge. One study evaluating AlphaFold 3 on protein-protein complexes showed good correlation for binding free energy changes but also indicated limitations in overall structural accuracy for some cases.[59] Rosetta-MPDock has also been benchmarked on a dataset including structures solved within this timeframe.[9] These early findings suggest that while progress is being made, a dedicated benchmark specifically focused on a diverse set of recently solved membrane protein complexes is still needed for a more detailed understanding of the current capabilities of AlphaFold-class tools in this challenging area.

The performance of AlphaFold-class tools on membrane protein complexes can vary depending on specific characteristics, such as the availability of sequence homologs and the presence of flexible loop regions at the interaction interface.[7] For example, AlphaFold2 has been reported to struggle with membrane proteins lacking sufficient evolutionary information.[7] AlphaFold-Multimer has shown limitations in modeling complexes with loop-mediated interactions.[7] However, increasing the number of prediction cycles can sometimes improve the accuracy for multi-chain complexes.[47]

6. Case Studies:

A detailed analysis of the recently solved membrane protein complex structures listed in Section 4 would provide valuable insights into the current capabilities of AlphaFold-class tools. For instance, predicting the structure of the *Trypanosoma cruzi* MscS channel (9BGS) using AlphaFold-Multimer would allow for an assessment of its ability to model a homoheptameric transmembrane channel, evaluating the accuracy of the predicted oligomeric state, the arrangement of transmembrane helices, and the confidence scores in these regions. Similarly, predicting the structure of the human GAT3 complex (9CP4) with its inhibitor and antibody fragments using AlphaFold-Multimer or RoseTTAFold All-Atom would test the tools' performance on a heterotrimeric membrane protein complex involving a transporter, a small molecule, and antibodies. The accuracy of modeling the inhibitor binding site and the relative positioning of the antibody fragments would be key evaluation points. Analyzing the prediction for the BchN-BchB complex (9E7H) with bound Pchlide would assess the ability to model a multi-subunit membrane enzyme complex with a ligand. For the CHD1-nucleosome complex (9NH8), AlphaFold3's performance in predicting protein-nucleic acid interactions and the overall architecture of this assembly could be evaluated. Finally, predicting the structure of the RIP3 kinase domain in complex with its inhibitor (9IWY) using AlphaFold2 or RoseTTAFold would test the tools' ability to model a kinase domain and the binding mode of a small molecule. By comparing the predicted structures with the experimental data using metrics like TM-score, interface RMSD, and analyzing the confidence scores (pLDDT, PAE, ipTM), a detailed evaluation of the strengths and weaknesses of each tool for different types of membrane protein complexes can be achieved. Visual comparisons of the superimposed structures would further enhance this analysis.

7. Limitations of AlphaFold-Class Tools for Membrane Protein Complex Modeling:

AlphaFold-class tools, while representing a significant advancement, still face limitations when applied to membrane protein complexes. Their performance is highly dependent on the availability of sufficient homologous sequences in the training databases, which might be limited for certain membrane protein families or proteins from less-studied organisms.[45] The training data itself might also be biased towards soluble proteins, potentially affecting the accuracy of predictions for membrane proteins with their unique structural features dictated by the lipid environment.[6]

A significant limitation is the lack of explicit modeling of the interactions between membrane proteins and the surrounding lipid bilayer in standard implementations.[6] While AlphaFold3 has expanded its capabilities, accurately representing the complex and dynamic interactions within the membrane remains a challenge. This can affect the predicted orientation of transmembrane domains and the overall stability of the complex in a physiological context.

Furthermore, these tools primarily predict static structures and do not inherently capture the dynamic nature of membrane proteins, which often undergo significant conformational changes during their functional cycles.[5]This static representation can be

a major limitation when studying the mechanisms of membrane protein complexes involved in dynamic processes.

Finally, while confidence scores like pLDDT, PAE, and ipTM provide valuable estimates of prediction accuracy, their reliability might vary for different classes of proteins, including membrane protein complexes.[6]Researchers should interpret these scores with caution and ideally validate predictions with experimental data.

8. Future Directions and Recommendations:

Future research should focus on enhancing the accuracy of AlphaFold-class tools for membrane protein complex modeling by incorporating membrane-specific information into their training and prediction processes.[103] This could involve integrating data on lipid interactions and the physicochemical properties of the membrane environment. Refining training datasets to include a more representative set of membrane protein complex structures is also crucial. Developing methods to better capture protein dynamics and conformational changes within membranes would be a significant advancement. Exploring the full potential of AlphaFold3's expanded capabilities for modeling interactions with lipids and other membrane components could also lead to improved accuracy.

Researchers using these tools for membrane protein complexes should interpret confidence scores judiciously and validate predictions with experimental data whenever possible. Comparing predictions from different tools can provide a more robust assessment. Utilizing resources like the TmAlphaFold database can aid in evaluating predicted membrane localization. Integrating these tools with molecular dynamics simulations and docking methods can provide a more comprehensive understanding of membrane protein complex structures and functions.[9]

9. Conclusion:

AlphaFold-class tools represent a revolutionary advancement in protein structure prediction, offering a powerful approach to modeling challenging systems like membrane protein complexes. While these tools have shown significant promise, limitations remain, particularly in capturing the complexities of the membrane environment, protein dynamics, and interactions with lipids and other molecules. Future efforts should focus on addressing these limitations through enhanced algorithms, refined training data, and integration with other computational and experimental methods. Continued benchmarking on recently solved structures will be essential for guiding progress and ultimately advancing our understanding of the intricate world of membrane protein biology and facilitating the development of new therapeutic strategies.