Journal of High School Science

Future of protein structure modeling

Wang Y.A

Submitted: September 30, 2024, Revised: version 1, May 17, 2025
Accepted: June 3, 2025

Abstract

Understanding a protein's 3D structure, which is linked to its function, is essential for biological processes, identifying disease mechanisms, and drug development. Traditional experimental techniques like X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (Cryo-EM) have been used to determine precise protein structures. However, these methods are resource-intensive, resulting in only a few thousand structures being deciphered every year. This is in stark contrast to the extensive number of sequences in the UniProt database. To this end, numerous studies have investigated the possibility of integrating template-based and template-free structure prediction approaches. The most notable advancement was through the use of a hybrid approach in 2021 when DeepMind's AlphaFold2 (AF2) introduced a protein structure prediction model with unprecedented accuracy using an end-to-end deep learning algorithm. Despite its widespread impact, AF2 continues to face architectural challenges and struggles with limited available data, preventing it from generating structures consistent with those that result from experimental methods. Recently, several techniques, such as smart sampling methods, and hybrid experimental approaches that enhance Cryo-EM maps with AF2 structures have been developed to address these limitations, surpassing AF2 and traditional methods in many instances. As developments of novel core algorithms allow AF2 to model a wider range of proteins with higher accuracy and hybrid AF2-experimental approaches expedite experimental data collection for AF2, a pure computation based solution to protein structure prediction seems increasingly likely. This review will provide an overview of AF2 and identify promising advancements in protein structure prediction since its introduction.

Keywords

Protein structure prediction, DeepMind, Distogram, AlphaFold, Deep Learning, Hybrid approaches, Computational biology, Cryogenic Electron Microscopy, Interresidue contacts, Coevolutionary data

Yu-Hong Alexander Wang, Central Jersey College Prep Charter School, Somerset, NJ 08873, USA.
alexanderwang3445@gmail.com

https://doi.org/10.64336.001c.140648

1. Introduction

Proteins are macromolecules crucial for various functions and biological processes in the human body and all living organisms. These polymers consist of amino acids linked by peptide bonds, forming polypeptide sequences that naturally fold into stable, complex 3-dimensional tertiary structures (1). Proteins can have multiple unique structures, configurations, or conformations (or states) that provide the shape necessary for carrying out their biological functions, such as transporting oxygen, providing structure and strength to cells, acting as enzymes, offering immune protection, and regulating genetic processes (2). These unique structures fall within a space containing an ensemble of all possible conformations and folds, commonly referred to as the conformational space (3, 4). The folds are determined by the many complex amino acid encoding possibilities, which researchers propose as being several billion or greater (5), of these proteins that enable them to perform a large range of functions across all organisms.

Specifically, the interactions between the 'R' substituent groups in the amino acids are responsible for their ability to fold. Misfolding, or incorrect folding, of proteins into their three-dimensional structures can occur due to the substitution of amino acids, particularly with changes in the chemical properties of their R-groups, ultimately causing changes in folding behavior. For instance, the amino acid cysteine of a polypeptide has a sulfhydryl group that can form a disulfide bridge (critical for structure stability) with another part of its polypeptide chain, but when amino acid valine—which lacks a sulfhydryl group and is nonpolar—replaces cysteine, the polypeptide may tend to misfold. Since protein folds define their biological function, misfolds like these are highly consequential to improper function of proteins in cells, often resulting in disease onset (6, 7).

For example, in Alzheimer's disease (a brain disorder), the loss of memory and brain control is commonly attributed to abnormalities associated with the Amyloid beta protein. Misfolding of the amyloid beta protein can cause it to form oligomers that are toxic to neurons, and these oligomers can also clump together to fabricate fibrils that causes plaques to form between neurons (8). The Tau protein is also involved in this disorder, where the disease causes it to form tangles within brain cells, leading to memory dysfunction (9). As a result of disrupted protein function caused by the many structural changes that accompany the disease, the human body can be significantly harmed. Other canonical examples of its harmful effects include breast cancer, which occurs when breast cells grow uncontrollably and tumors are formed, and spreads quicker with the presence of proteins like HER2 (10); cystic fibrosis, a genetic disease in which the CFTR protein does not fold properly, causing the body to produce excessive mucus and leading to a blockade of the digestive system and other organs (11); and Huntington's disease, a brain disorder where an inherited defective form of the Huntington protein causes it to impair nerve cells, affecting movement and thinking increasingly over time (12).

Characterizing the structure of target proteins allows scientists to identify the function of

disease-related proteins, formulate treatments, and tailor drugs to bind to the targeted areas of the structure. Besides a protein's amino acid sequence (known as the primary structure), proteins can also form secondary, tertiary, and quaternary structures. Studying more complex structures is an integral part for scientists to gain deeper insight, and understand relationships, and interactions between different proteins. Secondary structures, like alpha helices and beta folded sheets, can be arranged to form conformations called motifs, which determine a protein's shape (13). On the other hand, tertiary and quaternary structures determine a protein's 3D function and protein-protein interactions, and as a result, are often favored for studies involving the development of potential drugs for diseases. These two structures are especially important for fighting complex diseases by disrupting harmful protein interactions or stabilizing beneficial conformations, which, therefore, are critical for improving therapeutic outcomes.

1.1 Traditional structure determination techniques

The most widely used structural characterization techniques designed primarily for (but not limited to) determining tertiary structures include X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (cryo-EM). For the purpose of comparing models constructed by these experimental processes, scientists often measure the similarity (the accuracy of one structure relative to another) and the resolution (the level of structural detail or how well the model represents the actual atomic positions) of the generated models. The resolution of experimental structures is calculated using several techniques such as diffraction data for X-ray crystallography (14), coordinate data for NMR (15), image comparisons for cryo-EM (16), and other mathematical techniques. The higher the resolution (indicated by lesser Angstroms, Å), the more precise the positioning of atoms and structural features such as conformation changes are, in a structural model; higher resolution also entails a higher confidence level and allows for better assessment of the structures.

Typically, X-ray crystallography results in atomic-level resolution ranging from 1.0 to 2.0 Å, while NMR and cryo-EM result in coarser resolutions ranging from 2.0 to 3.5 Å, and sometimes these processes produce even lower resolution structures (> 3.5 Å) due to experimental limitations such as the quality of the sample. On the other hand, measuring the similarity between two superimposed models' atomic coordinates is usually calculated by their root mean square deviation (RMSD) (18), where a lower RMSD value corresponds to a higher similarity between structures. Overall, conventional experimental methods offer a high amount of resolution detail in their solved structures that are crucial for future studies or drug development, while it is important to note that some methods may inherently perform better at solving a specific type of structure depending on the protein's size, structure-specific characteristics and complexity.

X-ray crystallography (14) is by far the most favored physical experimental technique for determining protein structures because of its efficiency and accuracy. As of 2024, considering all the 192,074 recorded structures from the Protein Data Bank (PDB) (19)—a collection of the world's largest domain of

atomic-resolution protein structures determined by experimental methods—X-ray crystallography is responsible for 97% of them, including the cases where multiple methods are used side-by-side. The process begins by obtaining a pure protein sample, followed by setting up crystallization trials through several conditions to grow high-quality crystals (17). Then, the crystals are exposed to an X-ray beam, and their diffraction data is collected, which are converted using computational methods into electron density maps that are used to form a final model of the protein structure. Although X-ray crystallography tends to yield high resolution and throughput, it can be time-consuming, often requiring months or even years to obtain a single structure depending on the quality of crystal samples and other experimental conditions. NMR (15) is another common method for determining protein dynamics and interactions. There is a large difference in viable throughput compared to X-ray crystallography, as NMR accounts for only 7% of all the PDB structures, indicating that it is not as suitable or as developed for this process, as X-ray crystallography is for protein structure modeling. NMR spectroscopy involves dissolving a labeled protein sample into a special NMR buffer, exposing it to a strong magnetic field and radiofrequency pulses, and finally converting the resulting resonances into spectrum graphs over several experiment trials; the data from which are subsequently analyzed by computer algorithms to create the final protein structure (20, 21). Advancement in superconducting magnet architecture follows a higher strength in the magnet field, which would capture models of higher complexity (22). Unlike X-ray crystallography experiments that often damage the sample, NMR offers a flexible, nondestructive method, meaning its physical samples can be reused across multiple experiments (23). However, large proteins have a larger number of atoms, requiring more experimental trials to obtain the needed number of spectra to pinpoint the exact molecular positions, which translates to more resource intensive computational work. The isotopic labeling also makes these experiments very costly and time-consuming, leading to a decreased throughput.

Within the field of 3D electron microscopy, cryo-EM (16) has emerged as the most popular and widely used technique in protein structure modeling, currently contributing to 11% of the PDB structures. First, a purified protein sample is prepared and placed into an electron microscopy grid, which is then rapidly frozen in liquid ethane to form vitreous ice (vitrification) (24, 25). Next, using an electron microscope, 2D images from many angles are captured and go through computational post-processing to correct distortions, align particles, and enhance the signal-to-noise ratio (a ratio between the strengths of some desired signal relative to those of background noise) to produce a 3D model. Technological improvements like the development of a direct electron detector (26) allows for higher resolution density maps and new image processing algorithms (27) offer high-speed image processing and improvements in signal-to-noise ratio. Finally, multiple rounds of producing density maps and applying software are applied and weaved into a completed protein structure, precise to often an atomic level resolution. These models are essential for not only capturing the protein's structure, but also its functional sites and conformational

changes. This method is also nondestructive similar to NMR.

Cryo-EM is known for its increasing potential of achieving high resolution due to technological advancements, such as from Yip et al. who tested a new electron microscope and achieved a world-record atomic resolution at 1.54 Å in 2020 (28, 29). Following this advance, new technological and computational advancements will likely allow cryo-EM to compete well with X-ray crystallography models. Cryo-EM's relatively flexible process (compared to the other two traditional methods) enables this technique to model larger proteins or protein complexes (quaternary structures), because [1] it does not require essential steps that limit the size of the protein such as X-ray crystallography having to obtain crystals (which are difficult to obtain for larger proteins) and [2] the advances in cryo-EM technology allow it to reach near-atomic resolution for certain samples, which are beneficial for a analyzing protein-protein interactions of complexes with high precision. Despite this potential, cryo-EM remains expensive and requires extensive computational resources for the image processing and the 3D construction of the protein structure. In addition, consequent to the high demand for computational power, data collection and processing can often be time-consuming and labor-intensive, resulting in less throughput.

Due to the limitations of these experimental methods, discovering the exact 3D structure of these proteins remains an exceptionally challenging task. Consequently, the 192,074 (mostly tertiary) structures in the PDB (19) are nearly incomparable with the 245,482,527 sequences recorded in the UniProt database (30), the world's leading collection of known primary protein structures. In addition to low efficiency in tertiary and quaternary structure determination, high-throughput sequencing technology has been advancing the number of sequences discovered per year, and therefore, the gap between structures and primary sequences has been growing wider. To put in perspective, the ratio of PDB structures to UniProt structures was 2% in 2004, 0.7% in 2010, and currently 0.08% in 2024. At this rate, it would take longer than the age of the entire universe to solve all possible protein structures (31). This wide gap between structure determination and sequence identification suggests a need for a new approach that can consistently generate convenient, high-resolution structures at high efficiency and a low cost, which would ultimately revolutionize the entirety of drug discovery.

1.2 Protein structure prediction

While fast-growing technology and computer algorithms continue to revolutionize these physical determination methods, the big question in structural biology emerges: is it possible to develop a pure computer system to predict higher-level structures using only something as simple as primary sequences?

Investigations along this idea began in 1973, when the relationship between the 3D tertiary structure and the 1D amino acid sequence of a protein was first established by Anfinsen et al. landmark study (32). They discovered that the native structure of a protein is the protein's thermodynamically most stable structure, and is "folded" from only the amino acid sequence without kinetic interference (33). This marked the beginning of the protein folding problem

and protein structure prediction as a whole. Since then, scientists have explored this direction by predicting the corresponding structure(s) from a protein's amino acid sequences using many computational methods such as energy models and sampling procedures. Recently, with the advent of AI, these methods were replaced by machine learning approaches such as deep learning. The design of automated predictive approaches allows these programs to negate the time and cost issues associated with experimental methods.

Protein structure prediction can generally be dissected into two categories: comparative modeling (34) and de novo methods (35). Comparative modeling, also known as homology or template-based modeling, uses templates of one or more solved structures of homologous proteins—proteins that share similar amino acid sequences—and uses either sequence alignments or sequence profiles to determine appropriate residue positions (36-39). This approach is based on Anfinsen's work (32), leveraging the established relationship between a protein's amino acid sequence and its native conformation to deduce that proteins with similar sequences have similar structures. On the other hand, de novo modeling, or ab initio ("from the beginning" in Latin) or free modeling, does not rely on templates to predict structures and rather on fundamental physics principles and Anfinsen's thermodynamic hypothesis (40), which posits that the most favorable native structure corresponds to the one with lowest free energy. This information allows these free modeling methods to sample different conformations and use algorithms to compare their free energies to determine the structure at local minimum energy states.

The Critical Assessment of protein Structure Prediction (CASP) competitions are hosted biennially with a goal of "establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused" (41). During a CASP assessment, unpublished solved protein structures are split into three groups: template-based modeling (TBM), Free-modeling (FM), and a hybrid TBM/FM. Structures tested for TBM share some degree of structural similarity (allowing templates to be used), whereas those in the FM category are not closely related to recorded structures (thereby encouraging a free modeling-orientated approach due to no available homologous templates). Any structure that falls somewhere in between those two categories (with mild templates available) is placed in the TBM/FM category. Discerning predictions with these two main types of structures (TBM and FM) ensures emphasis on research for methods that predict well for both (1) commonly solved proteins, which are those with many homologous structures in the PDB, and (2) "orphan" proteins with very limited homology in the PDB. The main section of CASP assesses group performances across tertiary structures, while there is also a side competition for quaternary structures (crucial for capturing complex protein-protein interactions), revealing a contemporary focus on tertiary structures, as protein complexes may be too far of a leap before finding an adequate solution for tertiary structure prediction.

The CASP experiment primarily uses the Global Distance Test – Total Score (GDT_TS) (42) metric to evaluate each structure produced by the prediction models. Given a predicted structure, the α-carbon atom positioning is compared to that of an experimentally solved structure and the percentages of atoms that are within distances of 1 Å, 2 Å, 4 Å, and 8 Å are calculated and averaged to finally yield the GDT_TS, ranging from 0 (no similarity) to 100 (perfect prediction). This method is more robust than many other similarity metrics such as residue coverage, which only accounts for one distance, and allows for a consistent measure of the overall structural quality of predictions. GDT_TS is also favored over RMSD, which can often be overly sensitive to poor modeling in small regions, and therefore offers a poorer indicator of overall structural similarity.

This review will provide a comprehensive overview of the breakthrough of protein structure prediction after CASP12, recent developments and techniques under tertiary and quaternary structure prediction, as well as noteworthy paths for the future of protein structure prediction.

2. The breakthrough in protein structure prediction

2.1 Computational methods

Predicting tertiary FM structures is challenging compared to template-based predictions since there are no references available, essentially predicting structures from scratch. More coevolutionary methods that use multiple sequence alignments (MSAs) were introduced to CASP11 predictions in both categories, but its success was limited due to the structures having scarce PDB sequence similarity (43). This trace of MSAs in CASP11 increased residue-residue contact prediction accuracy, with the result that many models that included co-evolutionary methods outperformed other methods. Previous FM techniques struggled with reliance on shallow sequence data, whereas MSAs could extract more evolutionary data. Even without the use of PDB templates, MSAs and co-evolutionary methods alone allowed the models to perform well by applying the MSA patterns and sequence relationships to predict residue contacts and structures (44). Thus, in CASP12, deeper MSAs found more use and resulted in greater success across both FM and TBM models. Average contact precision increased from 27% in CASP11 to 47% in CASP12, and models were found to be more successful with increased homologous sequences added to the PDB, reaching up to 90% to 100% precision in many cases (45).

This result showed that with the help of sequencing technology and more structures recorded in the PDB, MSAs could bring promising potential to protein structure prediction when utilized to their maximum. Because the discovery of the power of MSAs also amplified predictions within the intermediate section (TBM/FM), many pure FM and TBM projects shifted toward a composite approach capable of coupling MSAs with template-based methods for template-available structures to generate more accurate predictions; while still performing well in the absence of templates. For example, parent template usage among FM predictions have been prevalent among consistent top performers like the Baker-Rosetta server and

Zhang-Servers during CASP11, using 47% and 17% templates, respectively (39, 46).

With the advent of deep learning algorithms, deeper MSAs could be extracted to reveal detailed evolutionary relationships, forming better predictions. It follows that an end-to-end deep learning, hybrid TBM/FM method will be able to precisely predict a wide range of protein structures even with high complexity or challenges faced by traditional threading methods. As protein structure prediction models move closer to a perfect GDT_TS score of 100, prediction models become ever closer as a widespread replacement for experimental processes. Even though experimental structure determination methods are not completely accurate, they have provided feasible and important structures for studies over many decades. Therefore, a goal of sustaining a GDT_TS near 100 for prediction models will likely be enough to entail a large breakthrough in structural biology.

By the time of CASP13 (2018), all of the top performing models were utilizing deep neural networks and many incorporated a hybrid FM/TBM approach to produce accurate predictions— MSAs paired with deep learning techniques had become a popular approach for many of the protein structure prediction models. However, a number of models have also found success of using a convolutional neural network to predict structures and contacts with almost no use of evolutionary information. Rather than using sequence inputs, models can use supervised training of neural networks with sequence profiles to extract information from fragments or assemblies of fragments to make contact predictions (47). For example, one of the top performing groups in this competition, Zhang et al., worked with protein profiles with low sequence depth (and only fed the neural network a single sequence as input) but still reported high contact prediction accuracy (48, 49). Another non-MSA centralized approach entered in the competition was one by Park et al.'s ROSETTA, which was refined to output more accurate predictions. ROSETTA used co-evolutionary information to inform its models, but due to the size of the search space, refining a near-native structure was difficult since there were more ways to decrease its quality than to improve it. To counteract this, they introduced an approach that thoroughly searched for the low energy states in one neighborhood of the starting model in order to avoid straying too far and creating totally inaccurate predictions. Although this model did not perform the best in this competition, it was capable of avoiding largely incorrect structures when the starting model or core regions started out closer to the correct (low energy) structure (50).

2.2 DeepMind's AlphaFold

Presented by the Google DeepMind group, AlphaFold was the winner and overall best performing model at CASP13. Similar to many other top performing models, AlphaFold employed a deep learning framework that used MSAs to extract co-evolutionary information from the primary sequences into 3D structural predictions. However, the core of AlphaFold's success lay within its use of attention mechanisms that allowed it to efficiently utilize deeper MSAs, leading to more powerful insights from the structures of related proteins. Other techniques that allowed it to succeed against the other models were its refinement through physical and geometric constraints and

its large-scale training across the PDB using a correspondingly large dataset (51).

AlphaFold's architecture also incorporates a composite approach of using AI with both comparative modeling and de novo modeling approaches, similar to many other models. Inspired by Anfinsen's study on the relationship between sequences and structures, the convolutional neural network behind AlphaFold is trained from template structures in the PDB and analyzes homologous sequences for MSAs (as represented by the filter layer shown in Figure 1). To predict structures, AlphaFold relies on analyzing MSAs of the target protein's sequence to predict the distances between amino acid residues and generates an initial distogram, a 3D data structure model that represents distances of amino acid residues (52). Predicting using the residue distances proved more effective than contact positions, as it conveyed more information, and therefore, AlphaFold employs 200+ convolution blocks to process the residue data. AlphaFold inspects both local and non-local residue contacts, where non-local residue contacts are especially crucial for accurate predictions: Noivirt-Brik et. al noted that non-local residue contacts are significantly more evolutionarily and structurally conserved than local ones (53). Subsequently, the prediction network estimates a probability distribution of backbone torsion angles, which is then optimized along with the initial model using gradient descent to generate the final protein structure (51).

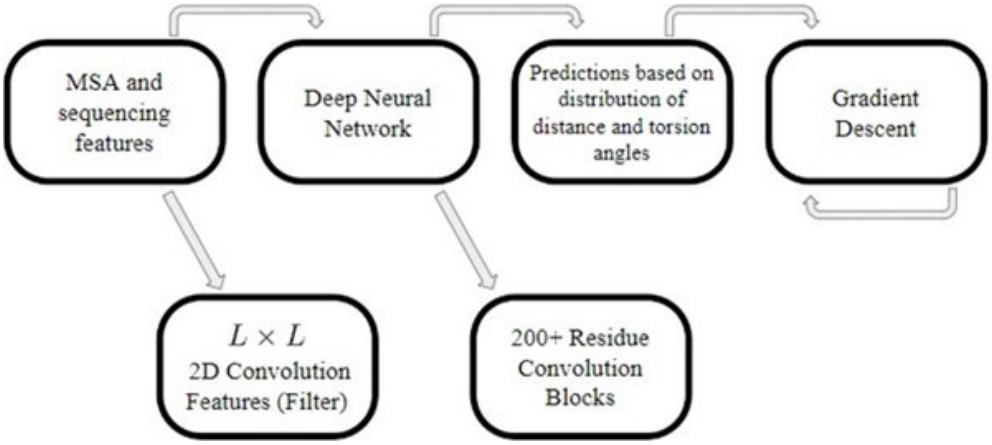

Figure 1. AlphaFold's architecture.

With more emphasis on de novo techniques, AlphaFold placed first in both the FM and TBM/FM sections, while coming in at second place for TBM structures. Using the sequences provided in the competition, AlphaFold generated precise structures for 24 of the 43 free modeling domains, compared to the next best method only scoring high accuracy on 14 of the 43 domains. AlphaFold received a GDT_TS of 58.9 while the second and third best performing models scored 52.5 and 52.4, respectively. AlphaFold also held a summed z-score (a standardized measure of GDT_TS across all groups' performances) of 52.8

compared to 36.6 for the next best group, the largest difference seen so far between the first and second places. Albeit increasing almost twice as much as the second-best performing model from CASP12, this score was only a slight boost in prediction success. Nonetheless, AlphaFold presented a promising path in protein structure prediction, and provided an important reference for future CASP14 competitors. Figure 2 below shows the spike in GDT_TS of the best performing model across the FM category, from 41 in CASP12 to 56 in CASP13 (41, 54).

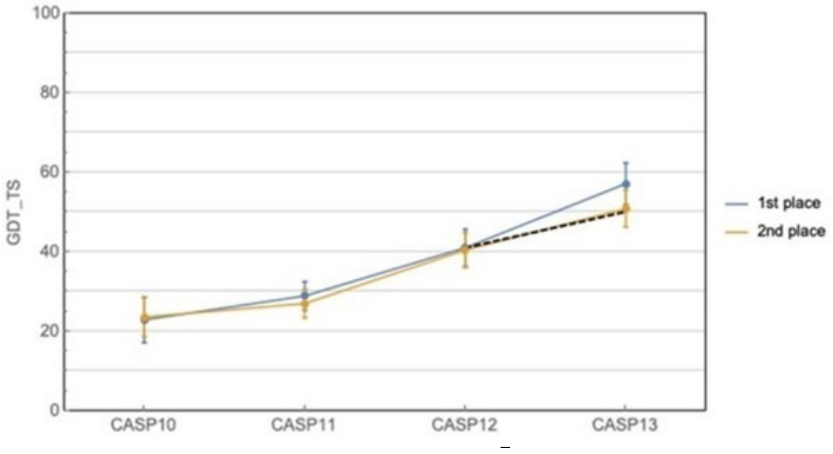

Figure 2. Progression of only FM models in GDT_TS in CASP10-14. Image obtained under license from Oxford University Press.

2.3 AlphaFold2: An accurate Deep-Learning system

Capitalizing on the success of AlphaFold, the DeepMind group introduced AlphaFold2 (AF2) during the 14th CASP contest. Unlike AlphaFold, AF2 increases the use of MSAs to flesh out more co-evolutionary information from the amino acid sequences and reduces its dependence on homologous structures, evolving it into an end-to-end approach that performs well even without templates. Due of this shift, AF2 is considered more of a de novo method with little influence of comparative modeling. Since the nature of FM structures usually cause them to be more difficult to predict, AF2's shift away from template-based modeling makes its discovery even more impactful. Skolnick et al. attribute AF2's increased generalization ability to other features in its improved attention mechanism within its architecture that allow it to utilize the PDB library of single-domain structures to the best of its ability (55). Jumper et al. stated that "AlphaFold greatly improves the accuracy of structure prediction by incorporating novel neural network architectures and training procedures based on the evolutionary, physical and geometric constraints of protein structure" (56).

AF2 takes in the input of an amino acid sequence and attempts to find homologs from

sequence databases to create an MSA representation; it also searches for structure templates in structure databases such as the PDB to create a pair representation. Although these representations are common to homology or template-based modeling, AF2 utilizes the MSAs in a different way by extracting co-evolution information from them through its unique transformer modules: Evoformer and the structure module (56, 57). These are shown in Figure 3.

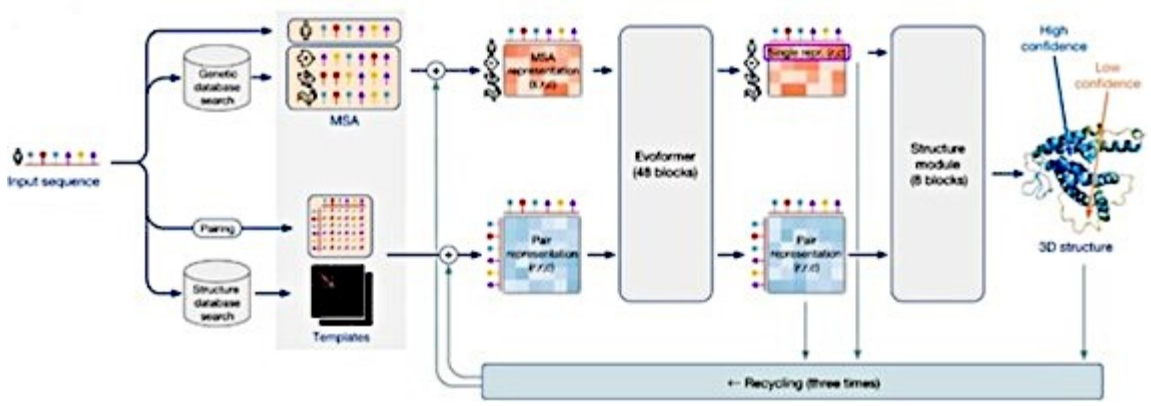

Figure 3. The architecture of AlphaFold2. Image adapted from the 2021 AlphaFold2 article by Jumper et al. (56). This work is licensed under CC BY 4.0 (http://creativecommons.org/licenses/by/4.0/).

One essential feature of AF2 is its Evoformer module, which integrates data from both MSAs and structural templates. This module contains two specialized transformers—one for the MSA data and another for the pair representations of protein residues. These transformers engage in continuous communication, allowing data to easily flow from one transformer to the other. As a result, the MSA transformer can exploit pair representation data to enhance its attention mechanism to more easily detect residue interactions. On the other hand, the pair representation transformer applies triangles of residues and imposes geometric constraints like triangle inequality that lead to further sophistication. Meanwhile, the structure module utilizes an attention mechanism to adjust the protein backbone's orientation, ensuring precise positioning of the amino acid side chains. Subsequently, local refinement and optimization through gradient descent are applied to produce the highly accurate 3D model (56).

The results were remarkable. AF2 proved to be a highly accurate and successful system, achieving an unprecedented median GDT_TS of 89.42 for predicting CASP14's experimental backbone protein structures. Almost half (51/111) of those targets were above GDT_TS of 90 and within those, >90 GDT_TS targets, more than half (26/51) scored greater than 95 GDT_TS (the highest two being 99.07 and 99.06). AF2's near-perfect GDT_TS score revolutionizes the field of structural biology, proving that computer systems are indeed capable of achieving

consistent experimental-like precision. AF2 greatly outperformed both its predecessor and other methods at CASP14 across all three categories, scoring a record-high median RMSD of 0.8 Å, which is significantly lower than the 2.8 Å held by the next best approach. AF2's dominance in the competition can be alternatively demonstrated by its standardized z-score of 244.0, far surpassing the second-place group's 90.8 (56), as shown in Figure 4.

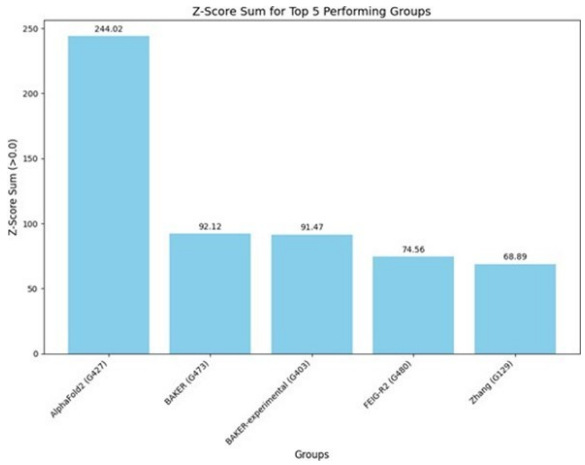

Figure 4. Top 5 performing groups for CASP14 across regular targets. AlphaFold2 (Group 427) scored the highest with a large margin to BAKER (Group 473), the second highest group. Rankings are based on the sum of z-scores of GDT_TS compared to the competition mean. Image created using data from the CASP14 public webpage (41).

These results reveal that the use of MSAs may be the better path forward than other pure fragment-based and sequence profiling methods.

2.4 Singular predictive prowess of AlphaFold2

The results of the end-to-end deep learning system are unarguably impressive and transformative. The local amino acid side chain rearrangements allow AF2 to refine its models to high-resolution structures, comparable to experimental resolution (55). As a predictive AI method, it has a much higher throughput compared to the physical methods. This throughput can be further refined with algorithm developments like ParaFold, which separates the CPU and GPU to enable large-scale structure predictions, improving GPU utilization and overall throughput (58).

AF2 could also provide other useful information about protein-protein interactions similar to the experimental methods. Guo et al. found that AF2 unintentionally reveals information about protein dynamics like residue flexibility, something that NMR succeeds at when explicitly interrogated (59). This information may be attributed to AF2's transformer modules in which attention mechanisms capture long-range interactions between residues. In fact, AF2 can also extend to a wide range of capabilities that match or even out-compete the methods designed specifically for doing that task. In 2023, McDonald et al. benchmarked AF2's accuracy in peptide prediction across 588 peptide structures. AF2 excelled in predicting α-helical, β-hairpin, and disulfide-rich peptides, so well, that it performed the same, if not

better, than alternative methods developed specifically for peptide structure prediction (60).

Yin et al. showed that AF2 could also predict heterodimeric protein complexes more accurately than traditional protein-protein docking methods, which is most likely due to AF2's use of MSAs (61). Two other studies by Edich M. et al. and Akdel M. et al. found other advantages that likely are consequences of AF2's use of MSAs or attention mechanisms that make it such an effective deep learning model; [1] AF2's unparalleled accuracy can be utilized for de novo protein design and in conjunction with Cryo-EM to interpret its data with an atomic model, [2] AF2 can identify structural features uncommonly seen in the PDB, [3] AF2 surpasses dedicated tools in predicting protein disorder and complexes, and [4] AF2's models have the same diverse applicability as experimentally determined structures (62, 63).

2.5 Limitations of AlphaFold2

Yet, many of the aforementioned studies have also found some of AF2's inefficiencies with MSAs being biased towards standard proteins with rich sequence data, thereby causing nonstandard proteins to be poorly represented. Specifically, AF2 reported difficulty modeling certain complexes such as antibody-antigen interactions (61), struggled to accurately predict specific aspects of peptide structures, such as Φ/Ψ angles and disulfide bond patterns, and sometimes displayed inaccurate accuracy estimates (60, 64). Also, it faced shortcomings in predicting cofactors, other ligands, and post-translational modifications (62).

In addition, many limitations with AF2's predictions such as conformational variability and fold flexibility are also caused by the MSA reliance evident in AF2's architecture. For example, a study by Chakravarty et al. evaluated AF2's bias in modeling single conformations—specifically, that the evolutionary data from MSAs represents the most probable conformation rather than the entire ensemble of possible structures—finding that AF2 accurately predicts one experimentally determined conformation for fold-switching proteins 94% of the time, while failing to capture a second distinct conformation (65). Similarly, Saldaño et al. demonstrated this bias among proteins that can exist in different forms, specifically those with 'apo' (unbound) and 'holo' (bound) conformers called apo-holo pairs. They discovered that AF2 predicts the holo form of the protein ~ 70% of the time and that its performance decreases as the protein's conformational diversity increases (66).

Chakravarty et al. also noted that regions with high sequence conversations in MSAs correspond to high-confidence predictions and that low sequence conservations correspond to low-confidence ones. This observation supports their result that AF2 using solely MSAs may not be sufficient to predict all conformations of proteins with heterogeneous structures. This limitation was further exemplified by AF2's difficulty in accurately predicting intrinsically disordered proteins (IDPs), which have low sequence conservation, resulting in low confidence and accuracy (65).

Similar to IDPs, loop structures have low sequence conservation, which presents another

challenge for AF2. Predicting these regions could be particularly useful in designing drug targets due to their high involvement in protein-protein interactions. Stevens et al. demonstrated this limitation, showing that as the number of residues in these loop structures increased, the accuracy of AF2's models decreased (67).

3. New algorithm developments and implementations of AlphaFold2

AF2 has shown the ability of surpassing conventional methods dedicated to many protein structure-related tasks. Despite that, limitations still exist, such as challenges in modeling regions with low sequence conservation, multiple conformations, and complexes/quaternary structures, all of which are likely caused by AF2's architectural reliance on MSAs. Recently; to address these challenges; many novel algorithm improvements and AF2 variants have emerged, aiming to increase accuracy in modeling many different structures and proteins.

3.1 Enhancements in secondary and tertiary structure prediction with AlphaFold2 variants

Although AF2's emphasis is on predicting tertiary structure, its model is also applicable for secondary structure predictions. Gulsevin et al. explored AF2's computational ability for modeling peptide structures in a benchmark of peptide structures with 16-60 amino acid chains. Comparing to the structures they experimentally determined, they reported that AF2 was able to achieve low RMSD values of ~ 3 Å. Despite shortcomings such as in predicting disulfide bonds, this benchmark displayed a potential for highly successful secondary structure predictions with the AF2 model (68).

Subsequent to this study, Szelogowski introduced ProteinINN as a Transformer-based model for end-to-end protein structure prediction from amino acid sequences. Similar to AF2, ProteiNN is a deep learning network that includes a Transformer model with an Attention mechanism called the Multi-Head Self-Attention to capture the relationships between amino acids in the sequence, utilizing MSA as an essential part of its model. But rather than combining this with geometry modeling to form predictions, the key to ProteiNN's predictions is its predictions on residue angles of each amino acid through sine and cosine outputs. Other notable differences are the smaller size in its training dataset (SideChainNet) than that for AF2—which trains on broader datasets such as the PDB—and a less MSA-heavy approach through its Transformers. With a RMSE loss of 0.448, it excelled "in predicting secondary structures, such as α-helices and βsheets", as noted by the authors. Success in secondary structure prediction generates useful insights to a protein's local backbone conformation, providing valuable understanding to its function and behavior across different (including native) states (69).

In tertiary structure prediction, researchers have focused on enhancing AF2's residue contact prediction accuracy and reducing its MSA bias. Sawhney et al. experimented with whether AF2's residue contact prediction accuracy could be increased by manually extracting features from the 3D structure. Using an inter-helical residue contact dataset, they tested the standard AF2's accuracy, achieving 83% average precision. They then added a procedure in which they extracted features from certain neighborhoods of residue

pairs in the atomic structure of the protein. This significantly improved the residue contact prediction when testing the precision, increasing from AF2's 83% to 91.9% accuracy, thereby demonstrating the underlying amount of detail that residue neighborhoods could contribute. Although AF2 uses MSA input to generate 3D structures, this approach delivered an accurate end-to-end improvement where almost all sequences (98%) sought an increase in classification performance, including those with lower sequence conservation. However, the authors described potential overfitting, and an incapability to generalize across proteins the model had not seen before, as challenges (70).

Additionally, smart sampling methods have emerged as a promising approach to address native conformational bias. AF2 biases a single conformation due to its incorporation of free modeling techniques that favor the native state. Smart sampling methods strive to induce AF2 to consider other conformational states that can help determine the biological function of the protein (71, 72), providing valuable insights to drug discovery. Manual adjustments to AF2's pipeline also play a crucial role in addressing these conformational biases.

Del Alamo et al. directly addressed AF2's considerable MSA reliance that caused it to bias towards single, homogenous conformations across most of its generated proteins. To directly decrease MSA reliance, they altered AF2's pipeline to reduce its depth of input MSAs through stochastic subsampling, leading to accurate models of multiple conformations being generated. Out of a sample of eight transporters and G-protein-coupled receptors tested, they found that every protein had at least one highly accurate conformational model produced. They also found that the models' suggested conformational fluctuations correlated with the behavior of actual conformational dynamics. However, this AF2 model still struggled to capture multiple conformations of long Loop regions despite decreasing the model's reliance on MSAs. Predicting states of proteins in the training data using this alternate model resulted in similar biased, single-conformational models generated by the base AF2. Nevertheless, this study presented an important approach to changing AF2's deep learning pipeline for identifying different conformational states, suggesting an AF2 pathway for accurately modeling native-like alternative states of proteins (73).

Lastly, in 2023, a sampling method was paired with AF2 to create AF-Cluster, a model developed by Wayment-Steele et al. to predict the effects of point mutations. In this study, the investigators clustered MSAs by sequence similarity, allowing AF2 to sample and predict alternative conformational states of metamorphic (fold-switching) proteins. AF-Cluster could also detect novel alternative states across proteins without prior knowledge of their fold-switching nature and also predict the effects of point mutations. However, this model of AF2 reported a correlation between deviations from the ground truth structure and regions of high local disorder, suggesting that there are still additional limitations that arise such as modeling those high local disorder regions. The authors of this project point out that developing these methods, along with experimental methods, could have a significant impact in predicting protein energy landscapes, which would further provide valuable

information on the biological function of proteins (74).

3.2 AlphaFold2 for quaternary structure prediction

AF2 often encounters challenges in accurately modeling protein-protein complexes, which is critical for understanding protein-protein interactions and developing novel therapeutics to combat serious diseases. A year after the release of AF2 in 2021, Google DeepMind extended the capabilities of AF2 to AlphaFold-Multimer (AFM) (75), an AI system specially designed to predict multimeric structures.

To test the strengths of AFM, Johansson-Åkhe et al. benchmarked AFM against previously used methods for peptide-protein docking. In a testing dataset of 112 complexes, AFM achieved a respectable native structure similarity score for more than half (66) of the peptide-protein complexes predicted (DockQ) (76), a measurement of the quality of docking models from 0 to 1, of ≥ 0.23) and almost half of those (25) were highly accurate (DockQ $\geq$ 0.8). This performance was outstanding compared to the other two methods, which accessed $\leq$ 47 accepted models and output only four or eight highly accurate models. This result provided evidence that AFM could be viable in predicting peptide-protein interactions.

In the same study, the investigators also explored the use of a forced sampling method that was tested for its effect on conformational output; i.e. random perturbation. By randomly perturbing its neural network weights, the AFM model was forced to sample more alternative states from the conformational space. This sampling technique increased the number of acceptable structures that the model generated from 66 to 75 and additionally improved the overall quality of the models. This approach—along with the previously discussed sampling methods—highlights the effect that sampling techniques can have on enhancing secondary, tertiary, and quaternary structure prediction among AF2-based prediction models such as AFM (77).

Given that quaternary structures are more complex than tertiary structures, Liu et al. sought to improve the model by deepening MSA and structure template usage. They developed an updated version of MULTICOM (78), a quaternary structure prediction system that builds on the AFM model. MULTICOM first predicts each subunit of a multimer using an in-house system similar to AF2. Then, MSAs from the subunits are concatenated in different ways depending on the type (hetero-multimers vs homo-multimers), generating both paired and unpaired MSAs. MULTICOM also uses structural templates, which are concatenated based on PDB data. Subsequently, a customized version of AFM takes pre-generated MSAs and templates and accepts them as input to generate tertiary structural predictions. Extensive sampling is performed to produce more predictions, up to 145 predictions for hetero-multimers. Next, a method called Foldseek Structure Alignment-Based Multimer Structure Generation (FSAMG) then uses Foldseek (79)—an algorithm for decreasing computation time for searching through protein structure databases—to align those predicted structures with known complexes in the PDB and structures from AlphaFoldDB (80) (a collection of AlphaFold's predictions), and additional predictions are built by pairing sequence

alignments of structural hits. Both sequence alignment-based and structure alignment-based predictions are then ranked based on confidence scores. Finally, the top-ranked models undergo further processing and refinement using Foldseek to improve accuracy, yielding MULTICOM's accurate predictions of a complex. The Foldseek structure alignment -based method is essential to this model, as it finds similar structures for hard targets that sequence alignment methods previously used by AFM could not, leading to deeper MSAs and more templates for AFM to use to generate better predictions.

MULTICOM was a contestant in CASP15, in which it placed 7th overall. It achieved an average Template Modeling (TM) score of ~ 0.76, which is 5.3% higher than approximately 0.72 by the stand-alone AFM model (41, 78). The average of MULTICOM's 5 best models achieved a TM score of around 0.8, which was 8% higher than the approximate 0.74 score of standard AFM. While MULTICOM, similar to the standard AFM model, still struggles with larger complexes, it demonstrated the potential of the Foldseek-based method, outperforming the commonly used sequence alignment-based multimer structure generation by achieving more quality predictions for many targets. In the same year, Liu et al. created a different version of MULTICOM using similar methods tailored to AF2 for tertiary structures instead of for complexes. Subsequently, they entered this version in the tertiary category of CASP15 and ranked 7th out of 132 groups, proving successful against AF2 with an average TM score of 0.92 compared to AF2's 0.85 (41, 81).
In 2024, a new benchmark by Lee et al. found that AFM performs well for interfaces between folded regions and short linear motifs, but struggles in predicting interactions from domain-motif pairs and binding specificities, especially mutations which are predicted at accuracies close to random. To increase prediction accuracy across these instances, they further investigated the use of protein fragmentation strategy - a method for predicting domain-motif interfaces. They applied this strategy to predict likely disease-related novel interfaces with high confidence and confirmed these using experimental methods. Since longer disordered regions were found to generally decrease prediction accuracy, they noted that protein fragmentation could increase the probability of obtaining high-confidence interface predictions by breaking those into smaller fragments. In addition to MULTICOM's success with introducing new metric systems, these authors also suggested that combining different metrics could enhance the discrimination of better performing structural models, and contribute to an increase in the overall accuracy (82).

4. The Future of AlphaFold2

4.1 Efficient data collection through hybrid experimental techniques

There are additional aspects to consider for improving the accuracy of an AI model such as AF2. While this paper has focused on the machine learning algorithm itself, the quality and quantity of the training data are also consequential to a model's generalization ability, and studying improvements in training data collection can offer promising results.
AF2's bias for certain proteins is often the result of the bias present in the current training dataset, including many proteins commonly

studied and desired by scientists. One way to reduce this bias is by incorporating more diverse data, thereby allowing AF2 to capture currently underrepresented proteins more effectively and more intricate features of protein structures - leading to improved generalization abilities. OpenFold, a version of AF2 trained from scratch to achieve the same accuracy, has been developed to address the issue of the limited size and diversity of the training dataset by disabling the template feature of AF2 and fine-tuning the training process to attain the original accuracy. OpenFold does particularly well in generalizing regions absent in the training dataset, especially relevant to AF2's significant limitation of data scarcity (83). However, given the difficulty of developing a method that accounts for all kinds of proteins and the significance of the biased and common proteins in drug delivery, this approach may not be ideal for current research. In fact, preserving a degree of bias in AF2's architecture could make it easier for scientists to focus on more direct changes and aid accurate predictions of high-priority proteins.

An alternative path is to refine AF2's predictions by focusing on increasing efficiency in data collection, which would, in the long run, increase the quantity of available structures for the convolutional neural network to train on. Experimental methods are currently responsible for AF2's data collection, as they provide highly reliable structures for the PDB. Experimentally determined structures are also the reference for predictive methods' similarity metrics. However, these physical methods remain time-consuming and inefficient, sometimes taking 1-2 years for complex structures. To eventually obtain a system capable of generalizing to any protein, uncommonly studied or structures currently difficult to discern due to experimental constraints will need to be experimentally solved. Thus, it is important to establish improvements to experimental techniques such as equipment technology or AI algorithms for computer postprocessing. Scientists have recently started to develop hybrid methods - a synthesis of experimental methods and AF2. Complementing AF2 with experimental methods can yield more accurate structures faster, utilizing predicted models as hypotheses to determine the correct structures. Over time, this can decrease the model's reliance on MSAs or other biases that have been formed due to a lack of training data.

In 2021, the CASP team utilized X-ray crystallography and cryo-EM to determine structures for the CASP14 competition. Kryshtafovych et al. reported the results of experimental solutions to four of the protein structures that were used for the competition. When the team evaluated AF2's precision and ability to pair refinements to structures generated by traditional physical methods, AF2 turned out to be highly successful in fine-tuning these processes' outputs. The AR9 nvRNAP enzyme, for example, had models constructed by both cryo-EM reconstruction and several X-ray crystallography maps with a resolution of 3.8 Å. The team aligned and fitted AF2's models into electron density maps using superposition and docking techniques, followed by real-space refinement. The resolution doubled to 1.9 Å, demonstrating that AF2 models were significantly insightful in guiding misplaced residues into correct assignments (84).

Cryo-EM density maps have been found to work particularly well with AF2. In 2022, Terwilliger et al. tested the effect of changing AF2 to use its initial model paired with a cryo-EM density map to rebuild a new model using the Phenix-Refine tool. This was then used as a template, and iterated across three steps before producing a final model. The results demonstrated that the final model performed better than either AF2 or cryo-EM used alone, which suggested that data from the density maps contains essential information that improves AF2 models (85).

In 2024, Chen et al. created DeepTracer-Refine. Unlike the previous two studies that relied on manual refinement; which is time-consuming and labor-intensive; this model offers an automated method that aligns residues from AF2 predictions with structures generated by DeepTracer, a de novo map-to-model method that converts cryo-EM density maps into structures. Depending on the resolution and quality of the density maps, this approach can achieve > 90% residue coverage. In a subset of 36 cryo-EM density maps with lower resolutions of 3.6–5 Å, the DeepTracer consistently performed better than AF2's models for average residue coverage, achieving 60.4% coverage to AF2's 43.9%. Overall, DeepTracer performed better when AF2 performed worse (<80% residue coverage), but more importantly, it significantly outperformed AF2 by 90% to 77.8% average residue coverage and 0.667 to 0.707 average Integrated Dual Disorder Treatment (IDDT), a confidence measurement. Additionally, it out-performed the Phenix-Refine tool used in the previous study in average residue coverage (90% to 88.4%) but presented with a lower average IDDT (0.707 to 0.728) (86).

4.2 Emerging approaches to improve predictive modeling

While experimental methods can be improved through constantly improving algorithms and technology, the potential of end-to-end computer systems brought by the development of AF2 shows that these physical methods may be an integral part of a transition to a standardized protein structure prediction model that replaces those traditional methods. This final prediction model should have architectural pieces of common threads between successful methods across different kinds of prediction studies (different kinds of proteins and regions). To improve the capacity for generalization of prediction models with architecture similar to AF2, techniques like protein fragmentation or increasing MSA and template usage are worth investigating, specifically for quaternary structure prediction. For tertiary structure prediction, smart sampling methods have demonstrated success and potential. On the other hand, since experimental methods are currently the only reliable way for protein structure data collection, their processes, which are limited by time and cost, will need to be improved as well. Therefore, techniques to improve AF2's algorithms, although promising, may be insufficient to significantly improve AF2's accuracy on their own, especially on expanded unseen proteins.

Instead; exploring hybrid techniques, which complement prediction models with experimental data, could provide a more feasible way to enhance physical methods. With improved physical methods, data

collection for protein structure prediction models will become more efficient, which, over time, allows those models to predict structures more confidently. Thus, hybrid methods could be instrumental to the transition from experimental methods into a fully reliable and standardized protein structure prediction model. For now, rather than treating AF2's predicted models as de facto ground truth structures, it may be wiser to use them as testable hypotheses, complemented with experimental techniques to provide insights toward more accurate models (87). Hybrid methods continue to hold promising advancements ahead and are expected to outperform current physical methods across all protein sizes at a much lower cost and time. Cryo-EM, particularly, has found great success pairing its density maps with AF2 to remarkably improve its model, even without the need for manual adjustments.

Although hybrid experimental methods may take longer than predictive methods, they are still much faster than traditional physical techniques due to the reduction in the amount of experimentation required. Exploring this path obtains synergies from both experimental methods and predictive methods: hybrid methods are faster than experimental methods, while being more precise and resulting in a higher level of detail than those of computational methods. These benefits are highly important when discussing structures as models for the process and purpose of drug delivery. Thus, hybrid techniques may likely be the optimal solution going forward for protein structure prediction and modeling, and may be an important step leading scientists to significantly accelerate the drug discovery process, boost understanding in protein function and stability, and potentially solve the protein prediction problem.

Therefore, for the known low-confidence regions of proteins of interest, going forward, it is possible to utilize a hybrid approach of applying cryo-EM testing of those specific low-confidence regions (such as mutations and loop structures) and leaving the rest to be quickly yet accurately predicted by AF2. With all the hybrid transformations occurring in the biocomputational field, this process could be an efficient solution that utilizes both methods in terms of their accuracies and efficiencies, allowing experimental methods to aid AF2 in its path to collecting more training data and subsequently obtaining 100% ground-truth prediction accuracy. Note that AF2 would be required to be highly trained and biased, carefully ensuring its ability to consistently predict the non-low-confidence regions to the best of its ability; i.e. at near ground-truth accuracies.

4.3 Overarching trends

With its unique iterative system and quality metric system, MULTICOM has shown a promising direction for an AF2-based model, in its obvious improvement from AF2-based de novo predictions. During the 2024 CASP16, MULTICOM ranked first in the average protein complex prediction without stoichiometry information (TM-score of 0.752). Most models were AF2-based, with MULTICOM scoring 88.69% average median predictions (compared with other top models 88-90%). On the other hand, the ROSETTA model is also controversial in its de novo modeling approach through a focus on optimization. Since most top-scoring models retain AF2's main properties, and due to AF2's

scale of breakthrough in CASP14, it is likely in the upcoming years for there to arise several significant competing models to the AF2 program. As AF2 has brought the computational biology field to an unprecedented stage leaving only about 10% accuracy of predictions to be corrected, AF2-based model architecture increasingly figures in adapting the path forward. In addition, AF2's breakthrough alongside the CASP contest have inspired many research programs to become interested in this field, forming them into an improved new state-of-the-art focus. Different strategies discussed in this review may be utilized to construct a model evolving AF2 into a deeper and more comprehensive model.

While the trend of the computational biology field progresses toward a final model, other paths to innovation like computer technology advancement in the conventional structure determination methods such as x-ray crystallography and cryo-EM become temporary fixes. With a significant lack of diverse protein structure data, the demand for generalization is growing. While an increased use of MSAs has shown to lead to better predictions (especially in the transition from the introduction of MSAs in AlphaFold1 to the deep MSA usage in AF2), the bias must be accounted for, through implementing a series of changes, whether it be to the architecture, algorithm-level mechanics, or the incorporation of other methods to increase the generalizability of the model**.**

4.4 Building a final model

Model generalizability is a *sine qua non* for improving overall predictions, specifically for proteins uncommon to contemporary studies. Therefore, it is important to take into account successful de novo methods. One of these includes random perturbation, random adjustments of the model's weights, notably introduced by Johansson-Åkhe et al. as a prominent aspect to improving generalizability of the AlphaFold2 model.

Next, it may be possible to combine the strengths of template-based and free modeling through an automatic 'bias modification switch' where there is a specific threshold which, through multiple accurate metrics of determining bias, chooses either a free modeling or a mixed approach depending on the commonality of the protein. This would allow for scientists to interact with the end result of the program, depending on their scientific wants and research goals. An improvement of overall metrics for the model's confidence levels (either replacing or in addition to pLDDT) for particular regions could be beneficial, paired with an iterative deduction of the best generated model. The MULTICOM program utilizes a blend of aforementioned approaches, providing a good direction to study in the upcoming years. This improvement would allow for better generalizability as a result of a more accurate overall and specific evaluations of the protein models in an unbiased manner.

The trade-off of added duration of the prediction program in return for more accuracy as demonstrated by MULTICOM is worth it, due to the time still being in hours or days, as opposed to the expenses, months or years, or other resources required by experimental processes. Some of AF2's weak prediction regions like protein-protein complexes have shown to be as, or more accurate, than other

programs dedicated to its modeling, like traditional docking approaches. This is a typical result of supervised learning and other AI systems, where unintended outcomes result without its specificity in the program. If specificity is provided, such as utilizing multiple algorithms for separate regions and applying new confidence evaluation metrics to compare the best models produced by either a new-algorithm-less or a new-algorithm-included process, allowing for the program to optimize and choose over the iterations the highest-accuracy model produced. In this incorporation, there should be region-specific or protein-specific algorithms for at least the following regions; (not accurately predicted by AF2) protein-protein complexes, loop structures, intrinsic disordered domains, high local disordered regions, and low sequence conservation regions. Keeping these additions and the concurrent use of MSAs could continue to provide accurate predictions in more common proteins, increasing the efficiency of template and data collection. This path of highly accurate prediction for the sake of data collection could be a solution; while simultaneously pursuing another more difficult path of finding an improved model focusing on generalization. Considering AF2's unintended result and power to generalize and the near-infinite variety of proteins possible, the former would be only a temporary solution to increase its capabilities, while the main focus remains in altering other central aspects of the program, particularly its core architecture.

4.5 The core model

Considering today's technology paired with the improvements seen from AF2 and previous models, it may be possible to achieve a generalized model without bias while retaining a high prediction accuracy. This goal may however require the creation of another breakthrough of a purely de novo model, by changing the core system. While the modification or addition of new hyperparameters will improve the model's effectiveness and usability within specific scientific contexts, the decisive change may occur by changing the transformers. AF2 constitutes the Evoformer transformers, the pair and MSA transformer, as well as its structure module. Since fewer templates were used by AF2 than by AF1, transformer improvement may eventually allow the model to function without the need of templates, in order to create a viable de novo prediction system.

It is plausible to alter the transformer modules such that the model's attention mechanism can improve. Since the flow of information between AF2's transformer modules is largely responsible for its predictions, there can be an introduction of a more closely connected model with extra layers while being efficient, for better back-and-forth communication extrapolation of information, perhaps done by adding a third transformer addressing the bias brought by MSAs, as well as by retaining yet adjusting current transformers, so as to avoid the risk of removing successful aspects of AF2.

For example, an algorithm by Han and Lu in 2016 on an alternating back-propagation algorithm (88) demonstrated a possible transformer add-on feature that could introduce many advantages. The algorithm offers a framework to the generator network model, utilizing a convolutional neural network to map latent factors to observed data

like images, video, and sound. Through its back-propagation, the program "alternates" between inferring latent factors by Langevin dynamics or gradient descent and updating parameters given the inferred factors.

The program is particularly effective in its ability to generalize in the presence of incomplete training data, which is prevalent in the PDB. With proper implementation and re-adjustments to structure modeling, this change could allow for coarse predictions to be refined into higher resolution, cover missing data scenarios such as low-confidence regions, and therefore increase prediction over the range of conformational variability accounted by the program's modeling.

While it is of importance to discover new transformers for the system, some, and if not many, features may be retained. One of these is the importance of both local and non-local residues (53). Adding non-local residues to local ones in the model building process has improved prediction metrics. This technique will serve to ensure greater accuracy, traditionally coupled with physics and mathematical principles, such as the triangle inequality as thresholds for the purpose of correct positioning and angles of the predicted residues.

4.6 Other improvements

The program's data weights should be re-evaluated to test whether the model's average prediction accuracies could improve by increasing representation of low-representation protein families in the training data, altering weights and adding new layers and iterations that allows the program to focus on more unique proteins and therefore correct the bias defect of the current model. This is much alike the gain in accuracy experienced by incorporations within MULTICOM's system. This could improve the predictability for certain regions which are, as of now, poorly represented by the model.

Other adjustments may include AlphaFold2's self-supervised learning from AlphaFoldDB, where the model improves its predictions based on the collection of previous predictions. MSA is undoubtedly needed in the final model as a prediction aspect, but more information could be obtained from MSAs. The success of MSAs comes from its ability to garner co-evolutionary information from the primary sequence, for it has been largely successful in its dominant popularization and usage among many models adapting this approach in the early 2010s. Thus, there is likely another method of analysis for a different aspect of the protein sequence for which its intricacies are able to be exploited for co-evolutionary inferences, with a notable example being direct coupling analysis, as examined in a AF2 study conducted by Caredda et al. in 2024 (89). MSA may hence be integrated with direct coupling analysis, with consequent gains in prediction. Furthermore, parts of ROSETTA's optimization architecture and its use of gradient descent could also be incorporated in the construction of the final model, since it has shown to perform well in free modeling settings (90).

An implicit assumption in AF2 and other models is that the protein solvent surroundings are primarily water at prevalent intracellular pH. It is now known that the crowded milieu of the cell contains myriad other molecules and polymers, biomolecular condensates, at

different redox potentials and pH. Protein folding and hence protein function may be altered as a result of not only primary structure mutations, but also proteins' cellular solvent composition and properties. Models should therefore be able to predict structures based on minimum free energy not only in 'standard' water solvents but also in altered and/or diseased cellular solvent environments.

5. Conclusion

Leveraging a deep learning architecture, AF2 has emerged as an important breakthrough in the realm of computational methods for protein structure prediction—AI models before AF2's development had not been remotely as successful as it was, partly due to its greater emphasis on MSAs, thereby revealing the usefulness and potential of MSAs for accurate predictions. AF2 revolutionized the field by proving that these AI models could indeed succeed with high accuracy in predicting protein structures, offering an alternative that can largely address the limitations encountered by conventional experimental techniques. By significantly reducing the time and cost, AF2 provides a faster and more efficient approach to modeling and understanding the complex structures of protein molecules.

However, significant challenges remain, such as improvement needed for multimeric protein prediction, predicting the structural impact of point mutations, and modeling regions with low sequence conservation. Often offset by AF2's architectural bias of MSAs, these challenges prevent its structures from becoming as consistently modeled as by experimental methods. To combat these problems, several novel heuristic advancements and AlphaFold variants have been recently developed, resembling ongoing efforts to maximize AF2's use of the limited training data in the PDB. These studies reveal potential in sampling techniques, protein fragmentation, and better evaluation systems. Another as yet unexplored challenge is to predict protein folding in altered and/or diseased protein cellular solvent environments.

Simultaneously, scientists have started to explore hybrid techniques - complementing experimental approaches like cryo-EM and X-ray crystallography with AF2-generated structures. Refined models utilizing the particular combination of Cryo-EM electron density maps and AF2 models yielded better results than individual models. As these strategies evolve, more training data will evolve, and, when paired with ongoing heuristic advancements, this will lead to a unified end-to-end AI framework capable of efficiently predicting a wide range of protein structures with low cost, ultimately accelerating advancements in drug discovery and molecular biology. Further exploration into automated refinement processes to experimental data such as DeepTracer-Refine and continued development of task-specific AF2 variants like AFM will be crucial for overcoming current obstacles in protein structure prediction and cultivating an understanding of protein dynamics and interactions. Since AF2 is still missing core mechanics to ensure consistent ground-truth accuracy predictions, new upcoming models should; while retaining much of its main methodology; utilize new methods to dissect co-evolutionary information, iteration techniques and better metric systems combined with the use of task-specific algorithms, and

other core architectural changes to the transformers to address region-specific prediction failures.

Acknowledgments

The author would like to thank Jerry Yan for his guidance, advice, and for assisting with the review's structure and scope of ideas. The author would also like to thank John Klich for providing insights into the startup research process and comments for revision.

6. References

1. Voet D, Voet JG, Pratt CW. Fundamentals of Biochemistry: Life at the Molecular Level. Biochemistry and Molecular Biology Education, 36(4), 319-20, 2008, https://doi.org/10.1002/bmb.20198

2. Pearce R, Zhang Y. Toward the solution of the protein structure prediction problem. J Biol. Chem., 297(1), 100870: 1-23, 2021, https://doi.org/10.1016/j.jbc.2021.100870

3. Afonine PV, Gobet A, Moissonnier L, Martin J, Poon BK, Chaptal V. Conformational space exploration of cryo-EM structures by variability refinement. Biochim. Biophys. Acta Biomembr., 1865(4), 184133: 1-7, 2023, https://doi.org/10.1016/j.bbamem.2023.184133

4. Wang T, He X, Li M, Shao B, Liu TY. AIMD-Chig: Exploring the conformational space of a 166-atom protein Chignolin with *ab initio* molecular dynamics. Sci. Data, 10, article 549: 1-12, 2023, https://doi.org/10.1038/s41597-023-02465-9

5. Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. Nature methods, 10 (3):186-187, 2013, https://doi.org/10.1038/nmeth.2369

6. Shrivastava S. Protein folding. Introduction to Biomolecular Structure and Biophysics: Basics of Biophysics. 1st Eds. Springer-Verlag New York, LLC, USA, 2017, https://doi.org/10.1007/978-981-10-4968-2_2

7. Jaenicke R. Stability and folding of domain proteins. Prog. Biophys. Mol. Biol., 71(2):155241, 1999, https://doi.org/10.1016/s0079-6107(98)00032-7

8. Hillen H. The beta amyloid dysfunction (BAD) hypothesis for Alzheimer's disease. Front. Neurosci., 13(1154): 1-10, 2019, https://doi.org/10.3389/fnins.2019.01154

9. Medeiros R, Baglietto-Vargas D, LaFerla FM. The role of tau in Alzheimer's disease and related disorders. CNS Neurosci. Ther., 17(5): 514-524, 2011, https://doi.org/10.1111/j.1755-5949.2010.00177.x

10. Schechter AL, Stern DF, Vaidyanathan L, Decker SJ, Drebin JA, Greene MI, et al. The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. Nature, 312(5994): 513-516, 1984, https://doi.org/10.1038/312513a0

11. National Heart, Lung, and Blood Institute. Cystic Fibrosis. 2023, https://www.nhlbi.nih.gov/health/cystic-fibrosis

12. Schulte J, Littleton JT. The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. Curr. Trends Neurol., 5: 65-78, 2011, https://pmc.ncbi.nlm.nih.gov/articles/PMC3237673/

13. Stollar EJ, Smith DP. Uncovering protein structure. Essays Biochem., 64(4): 649-680, 2020, https://doi.org/10.1042/ebc20190042

14. Glusker JP. X-Ray Crystallography of Proteins. Methods Biochem. Anal., 37: 1-72, 1994, https://doi.org/10.1002/9780470110584.ch1

15. Cavanagh J., Fairbrother WJ, Palmer AG III, Rance M, Skelton NJ. Protein NMR spectroscopy: principles and practice. 2nd Eds. Academic press, Cambridge, MA, USA, 2007, https://doi.org/10.1016/B978-0-12-164491-8.X5000-3

16. Cheng Y. Single-particle cryo-EM at crystallographic resolution. Cell, 161(3): 450-457, 2015, https://doi.org/10.1016/j.cell.2015.03.049

17. Smyth MS, Martin JH. X-Ray Crystallography. Mol. Pathol., 53(1): 8-14, 2000, https://doi.org/10.1136/mp.53.1.8

18. Kufareva I, Abagyan R. Methods of protein structure comparison. Methods Mol. Biol., 857: 231-57, 2012, https://doi.org/10.1007/978-1-61779-588-6_10

19. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat GN, Weissig H et al. The protein data bank. Nucleic Acids Res., 28(1):235-242, 2000, https://doi.org/10.1093/nar/28.1.235

20. Bovey, FA, Mirau PA, Nuclear Magnetic Resonance Spectroscopy. 2nd. Academic Press, cambridge, MA, USA, 1988, https://doi.org/10.1016/b978-0-08-091699-6.50011-8

21. Zardecki C, Dutta S, Goodsell DS, Lowe R, Voigt M, Burley SK. PDB-101: Educational resources supporting molecular explorations through biology and medicine. Protein Sci., 31(1): 129-140, 2022, https://doi.org/10.1002/pro.4200

22. Caceres-Cortes J, Falk B, Mueller L, Dhar TG. Perspectives on nuclear magnetic resonance spectroscopy in drug discovery research. J Med. Chem., 67(3): 1701-1733, 2024, https://doi.org/10.1021/acs.jmedchem.3c02389

23. Krishnan W, Rupp B. Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy. In eLS , John Wiley & Sons, Ltd (Ed.)., Hoboken, NJ, USA, 2012, https://doi.org/10.1002/9780470015902.a0002716.pub2

24. Danev R, Yanagisawa H, Kikkawa M. Cryo-electron microscopy methodology: current aspects and future directions. Trends Biochem. Sci., 44(10): 837-848, 2019, https://doi.org/10.1016/j.tibs.2019.04.008

25. Passmore LA, Russo CJ. Specimen preparation for high-resolution cryo-EM. Methods Enzymol., 579:51-86, 2016, https://doi.org/10.1016/bs.mie.2016.04.011

26. McMullan G, Faruqi AR, Henderson R. Direct electron detectors. Methods Enzymol., 579: 1-17, 2016, https://doi.org/10.1016/bs.mie.2016.05.056

27. Nogales E, Scheres SHW. Cryo-EM: a unique tool for the visualization of macromolecular complexity. Mol. Cell, 58(4):677-689, 2015, https://doi.org/10.1016/j.molcel.2015.02.019

28. Kato T, Makino F, Nakane T, Terahara N, Kaneko T, Shimizu Y, et al. The 1.54 Å resolution structure of apoferritin by CRYOARM300 with ColdFEG. Microsc. Microanal., 2:998-999, 2019, https://doi.org/10.6019/empiar-10248

29. Yip KM, Fischer N, Paknia E, Cahri A, Stark H. Atomic-resolution protein structure determination by cryo-EM. Nature, 587(7832): 157-161, 2020, https://doi.org/10.1038/s41586-020-2833-4

30. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res., 51: D523-D531, 2023, https://doi.org/10.1093/nar/gkac1052

31. Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. PNAS, 89(1): 20-22, 1992, https://doi.org/10.1073/pnas.89.1.20

32. Anfinsen CB. Principles that govern the folding of protein chains. Science, 181(4096): 223-230, 1973, https://doi.org/10.1126/science.181.4096.223

33. Nussinov R, Zhang M, Liu Y, Jang H. AlphaFold, artificial intelligence (AI), and allostery. J Phys. Chem. B, 126(34): 6372-6383, 2022, https://doi.org/10.1021/acs.jpcb.2c04346

34. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol. Biol., 234(3): 779-815, 1993, https://doi.org/10.1006/jmbi.1993.1626

35. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. Protein structure prediction by global optimization of a potential energy function. PNAS, 96(10): 5482-5485, 1999, https://doi.org/10.1073/pnas.96.10.5482

36. Altschul SF. Basic local alignment search tool. J Mol. Biol., 215(3): 403-410, 1990, https://doi.org/10.1006/jmbi.1990.9999

37. Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: detection of distantly related proteins. PNAS, 84(13): 4355-4358, 1987, https://doi.org/10.1073/pnas.84.13.4355

38. Altschul SF. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25(17): 3389-3402, 1997, https://doi.org/10.1093/nar/25.17.3389

39. Dukka BKC. Recent advances in sequence-based protein structure prediction. Brief. Bioinform., 18(6): 1021-1032, 2017, https://doi.org/10.1093/bib/bbw070

40. Kresge N, Simoni RD, Hill RL. The thermodynamic hypothesis of protein folding: the work of Christian Anfinsen. J Biol. Chem., 281(14): e11–e13, 2016, https://doi.org/10.1016/s0021-9258(19)56522-x

41. CASP Team. CASP: Critical Assessment of Protein Structure Prediction. 2024, https://predictioncenter.org/

42. Zemla A, Venclovas Č, Moult J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. Proteins: Struct. Funct. Bioinf., 37(S3): 22-29, 1999, https://doi.org/10.1002/(sici)1097-0134(1999)37:3+%3C22::aid-prot5%3E3.0.co;2-w

43. Kryshtafovych A, Monastyrskyy B, Fidelis K. CASP 11 statistics and the prediction center evaluation system. Proteins: Struct. Funct. Bioinf., 84:15-19, 2016, https://doi.org/10.1002/prot.25005

44. Marks DS, Colwell LJ, Sheridan R, Hopf, TA, Pagnani A, Zecchina R, et al. Protein 3D structure computed from evolutionary sequence variation. PloS ONE, 6(12): e28766, 2011, https://doi.org/10.1371/journal.pone.0028766

45. Schaarschmidt J, Monastyrskyy B, Kryshtafovych A, Bonvin AM. Assessment of contact predictions in CASP12: co-evolution and deep learning coming of age. Proteins: Struct. Funct. Bioinf., 86: 51-66, 2018, https://doi.org/10.1002/prot.25407

46. Kinch LN, Li W, Monastyrskyy B, Kryshtafovych A, Grishin N. Evaluation of free modeling targets in CASP11 and ROLL. Proteins: Struct. Funct. Bioinf., 84: 51-66, 2016, https://doi.org/10.1002/prot.24973

47. Kandathil SM, Greener JG, Jones DT. Prediction of interresidue contacts with DeepMetaPSICOV in CASP13. Proteins: Struct. Funct. Bioinf., 87(12): 1092-1099, 2019, https://doi.org/10.1002/prot.25779

48. Li Y, Hu J, Zhang C, Yu DJ, Zhang Y. ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. Bioinformatics, 35(22): 4647-4655, 2019, https://doi.org/10.1093/bioinformatics/btz291

49. Shrestha R, Fajardo E, Gil N, Fidelis K, Kryshtafovych A, Monastyrskyy B, et al. Assessing the accuracy of contact predictions in CASP13. Proteins: Struct. Funct. Bioinf., 87(12): 1058-1068, 2019, https://doi.org/10.1002/prot.25819

50. Park H, Lee GR, Kim DE, Anishchenko I, Cong Qian, Baker D. High-accuracy refinement using Rosetta in CASP13. Proteins: Struct. Funct. Bioinf., 87(12): 1276-1282, 2019, https://doi.org/10.1002/prot.25784

51. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature, 577(7792): 706-710, 2020, https://doi.org/10.1038/s41586-019-1923-7

52. Bertoline LMF, Lima AN, Krieger JE, Teixeira SK. Before and after AlphaFold2: An overview of protein structure prediction. Front. Bioinform., 3: 1120370, 2023, https://doi.org/10.3389/fbinf.2023.1120370

53. Noivirt-Brik O, Hazan G, Unger R, Ofran Y. Non-local residue–residue contacts in proteins are more conserved than local ones. Bioinformatics, 29(3): 331-337, 2013, https://doi.org/10.1093/bioinformatics/bts694

54. CASPTeam CASP13 Results. 2024, https://predictioncenter.org/casp13/results.cgi?view=tb-sel.

55. Skolnick J, Gao M, Zhou H, Singh S. AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. J Chem. Inf. Model., 61(10): 4827-4831, 2021, https://doi.org/10.1021/acs.jcim.1c01114

56. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 596: 583-589, 2021, https://doi.org/10.1038/s41586-021-03819-2

57. Yang Z, Zeng X, Zhao Y, Chen R. AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduct. Target Ther., 8:115, 2023, https://doi.org/10.1038/s41392-023-01381-z

58. Zhong B, Su X, Wen M, Zuo S, Hong L, Lin J. ParaFold: paralleling AlphaFold for large-scale predictions. International Conference on High Performance Computing in Asia-Pacific Region Workshops, 1-9, 2022, https://doi.org/10.1145/3503470.3503471

59. Guo HB, Perminov A, Bekele S, Kedziora G, Farajollahi S, Varaljay V, et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. Scientific Reports, 12(1): 10696, 2022, https://doi.org/10.1038/s41598-022-14382-9

60. McDonald EF, Jones T, Plate L, Meiler J, Gulsevin A. Benchmarking AlphaFold2 on peptide structure prediction. Structure, 31(1): 111-119, 2023, https://doi.org/10.1016/j.str.2022.11.012

61. Yin R, Feng BY, Varshney A, Pierce BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. Protein Sci., 31(8): e4379, 2022, https://doi.org/10.1002/pro.4379

62. Edich M, Briggs DC, Kippes O, Gao Y, Thorn A. The impact of AlphaFold2 on experimental structure solution. Faraday Discussions, 240: 184-195, 2022, https://doi.org/10.1039/d2fd00072e

63. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. Nat. Struct. Mol. Biol., 29(11): 1056-1067, 2022, https://doi.org/10.1038/s41594-022-00849-w

64. Takei Y, Ishida T. How to select the best model from AlphaFold2 structures? bioRxiv, 2022-04, 2022, https://doi.org/10.1101/2022.04.05.487218

65. Chakravarty D, Porter LL. AlphaFold2 fails to predict protein fold switching. Protein Sci., 31(6): e4353, 2022, https://doi.org/10.1002/pro.4353

66. Saldaño T, Escobedo N, Marchetti J, Zea DJ, Donagh JM, Rueda AJV, et al. Impact of protein conformational diversity on AlphaFold predictions. Bioinformatics, 38(10): 2742-2748, 2022, https://doi.org/10.1093/bioinformatics/btac202

67. Stevens AO, He Y. Benchmarking the accuracy of AlphaFold 2 in loop structure prediction. Biomolecules, 12(7): 985, 2022, https://doi.org/10.3390/biom12070985

68. Gulsevin A, Meiler J. Benchmarking peptide structure prediction with AlphaFold2. bioRxiv, 2022-02, 2022, https://doi.org/10.1101/2022.02.17.480937

69. Szelogowski D. Deep Learning for Protein Structure Prediction: Advancements in Structural Bioinformatics. bioRxiv, 2023-04, 2023, https://doi.org/10.1101/2023.04.26.538026

70. Sawhney A, Li J, Liao L. Improving AlphaFold Predicted Contacts for Alpha-Helical Transmembrane Proteins Using Structural Features. Int. J Mol. Sci., 25(10): 5247, 2024, https://doi.org/10.3390/ijms25105247

71. Stein A, Fowler DM, Hartmann-Petersen R, Lindorff-Larsen K. Biophysical and mechanistic models for disease-causing protein variants. Trends Biochem. Sci., 44(7): 575-588, 2019, https://doi.org/10.1016/j.tibs.2019.01.003

72. Wang Z, Moult J. SNPs, protein structure, and disease. Hum. Mutat., 17(4): 263-270, 2001, https://doi.org/10.1002/humu.22

73. Alamo DD, Sala D, Mchaourab HS, Meiler J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. eLife, 11: e75751, 2022, https://doi.org/10.7554/elife.75751

74. Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Hömberger M, et al. Predicting multiple conformations via sequence clustering and AlphaFold2. Nature, 625(7996): 832-839, 2024, https://doi.org/10.1038/s41586-023-06832-9

75. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021-10, 2021, https://doi.org/10.1101/2021.10.04.463034

76. Basu S, Wallner B. DockQ: a quality measure for protein-protein docking models. PloS ONE, 11(8): e0161879, 2016, https://doi.org/10.1371/journal.pone.0161879

77. Johansson-˚Akhe I, Wallner B. Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. Front. Bioinform., 2: 959160, 2022, https://doi.org/10.3389/fbinf.2022.959160

78. Liu J, Guo Z, Wu T, Roy RS, Quadir F, Chen C, et al. Enhancing alphafold-multimer-based protein complex structure prediction with MULTICOM in CASP15. Commun. Biol., 6(1): 1140, 2023, https://doi.org/10.1038/s42003-023-05525-3

79. Kempen MV, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat. Biotechnol., 42(2): 243-246, 2024, https://doi.org/10.1038/s41587-023-01773-0

80. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res., 50: D439D444, 2022, https://doi.org/10.1093/nar/gkab1061

81. Liu J, Guo Zhiye, Wu T, Roy RS, Chen C, Cheng J. Improving AlphaFold2-based protein tertiary structure prediction with MULTICOM in CASP15. Commun. Chem., 6(1): 188, 2023, https://doi.org/10.1038/s42004-023-00991-6

82. Lee CY, Hubrich D, Varga JK, Schäfer C, Welzel M, Schumbera E, et al. Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. Mol. Syst. Biol., 20(2): 75-97, 2024, https://doi.org/10.1038/s44320-023-00005-6

83. Ahdritz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, et al. OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nat. Methods, 21(8): 1514-1524, 2024, https://doi.org/10.1038/s41592-024-02272-z

84. Kryshtafovych A, Moult J, Albrecht R, Chang GA, Chao K, Fraser A, et al. Computational models in the service of X-ray and cryo-electron microscopy structure determination. Proteins: Struct. Funct. Bioinf., 89(12): 1633-1646, 2021, https://doi.org/10.1002/prot.26223

85. Terwilliger TC, Poon BK, Afonine PV, Schlicksup CJ, Croll TI, Millán C, et al. Improved AlphaFold modeling with implicit experimental information. Nat. Methods, 19(11): 1376-1382, 2022, https://doi.org/10.1038/s41592-022-01645-6

86. Chen J, Zia A, Luo A, Meng H, Wang F, Hou J, et al. Enhancing cryo-EM structure prediction with DeepTracer and AlphaFold2 integration. Brief. Bioinform., 25(3): bbae118, 2024, https://doi.org/10.1093/bib/bbae118

87. Agarwal V, McShan AC. The power and pitfalls of AlphaFold2 for structure prediction beyond rigid globular proteins. Nat. Chem. Biol., 20(8): 950-959, 2024, https://doi.org/10.1038/s41589-024-01638-w

88. Han, Tian, Lu Y, Zhu SC, Wu YN. Alternating back-propagation for generator network. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1), 2017, https://doi.org/10.1609/aaai.v31i1.10902

89. Caredda F, Pagnani A. Direct coupling analysis and the attention mechanism. BMC Bioinformatics, 26(1): 41, 2025, https://doi.org/10.1186/s12859-025-06062-y

90. Wei S. Rosetta energy approximation using a machine learning approach. Johns Hopkins Univ, Masters Thesis, Baltimore, MD, USA, 2024, https://graylab.jhu.edu/MastersTheses/SenWei2024EApprox.pdf