

Literature Review

A New Age of Protein Structure Prediction: How AlphaFold Changed the Landscape

Zhang, Samuel (K21120489)

¹ Lambden, Edward

ABSTRACT

AlphaFold revolutionised Protein Structure Prediction (PSP), transitioning from experimental and *ab initio* methods to Deep Learning (DL) methods, achieving atomic-level accuracy PSP. AlphaFold 2 (AF2) introduced the Evoformer, enabling accurate PSP and expanding structural coverage. AlphaFold 3 (AF3)'s transition to diffusion-based modeling enables prediction of novel structures across the biomolecular space at lower computational cost. AF3 shows promising results for Protein-protein Interactions (PPIs) and mutation prediction despite challenges in stereochemistry and complex interactions. This review explores how AlphaFold changed the landscape of PSP.

CONTENTS

I Introduction	4
I.i Approaches to PSD	4
I.ii Experimental PSD	4
I.iii Computational PSP	4
II AF2's Impact on PSP	6
II.i AF2's Release	6
II.ii Architecture of AF2	6
II.iii Achievements and Limitations of AF2	7
III AF3's Impact and Future Outlook	9
III.i Release of AF3	9
III.ii AF3 Comparative Studies	9
III.iii Stereochemical Violation Refinement	10
IV Conclusion	11
References	12
Abbreviations	13
Appendix	15

I. INTRODUCTION

i. Approaches to PSD

Proteins are essential for life, performing roles from structural support to cellular signaling; making them key targets in drug development. Proteins function depends on their three-dimensional structure, making structural understanding crucial for research and drug development.¹

A typical protein can adopt an enormous number of possible conformations, yet in nature, proteins reliably fold into their functional native structure. This contrast between the vast theoretical possibilities and the precise biological outcome has puzzled scientists for generations.² In 1961, Nobel laureate Christian Anfinsen proposed his thermodynamics theory of protein folding: protein structure is determined solely by the thermodynamics of its amino acid sequence and environment alone.³ This implied that *in silico* PSP is possible from sequence alone (Figure 1). Despite this promising insight, the sheer complexity of protein folding made PSP elusive for decades.

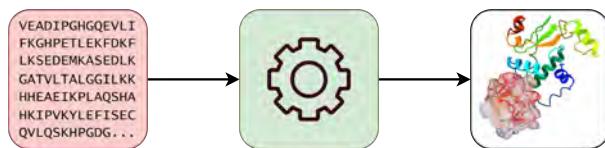


Figure 1. An algorithm that determines total structure information of a protein based on amino acid sequence alone. Created using ChimeraX⁴ and Draw.io⁵.

ii. Experimental PSD

During this time, several breakthroughs in Protein Structure Determination (PSD) emerged through experimental techniques such as X-ray crystallography⁶, Cryogenic Electron Microscopy (Cryo-EM)⁷, and protein NMR⁸.

Protein crystallography was pioneered by Kendrew, J. C. *et al.* while determining the structure of myoglobin.⁶ PSD involves producing, purifying, and crystallising proteins, then fitting the diffracted patterns from an X-ray source aimed at the monocrystalline protein crystals to construct a model in reciprocal (Fourier) space. However, this method has several limitations: structure determination initially took years, only rigid and relatively small proteins could be crystallised, and most importantly, crystallography assumes ligand binding to crystallised proteins is identical to binding in wild-type proteins, a significant oversimplification from

reality.^{9,10}

Cryo-EM involves flash-freezing protein samples in vitreous ice. Millions of electron micrographs of the protein at different orientations are then classified into 2D classes, Fourier transformed, and assembled in 3D Fourier-space before converting into real-space electron density maps. The protein structure is then fitted to these maps. The RCSB Protein Data Bank (PDB)¹¹ recorded a surge in Cryo-EM usage in the 2000s following breakthroughs in direct electron detectors¹² and computational power⁷ shown in Figure 3c.

NMR spectroscopy is used to determine dynamic protein structures in solution. Utilising proton coupling, nuclear magnetic resonance, and nuclear overhauser effect¹³, Kurt Wüthrich pioneered the method of using multidimensional heteronuclear NMR experiments for PSD, earning the 2002 Nobel Prize in Chemistry. Without requiring protein crystallisation or freezing, NMR excels at studying dynamic proteins.⁸

Each experimental method, with their own benefits and limitations (Figure 2), determined approximately 230,000 protein structures recorded in the PDB. With increasing structure released per year (Figure 3a,b).¹¹

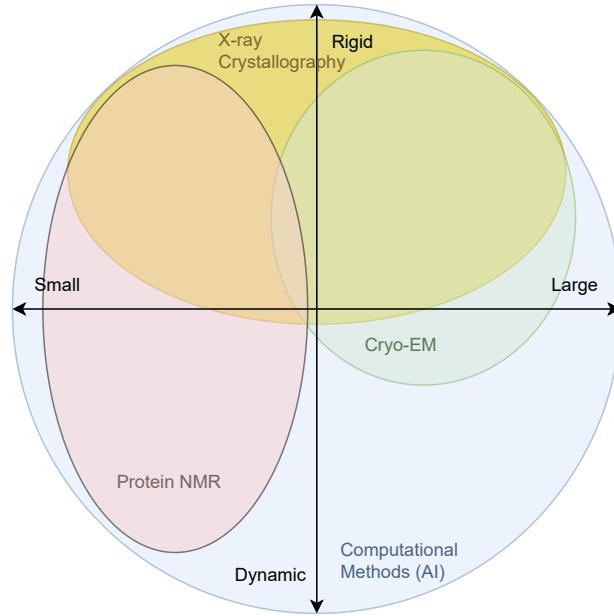


Figure 2. Map of PSD domain. With an optimistic AI covering the entire domain. Created with Draw.io⁵.

iii. Computational PSP

Despite major advancements in experimental methods, only about 0.1% of known protein structures have

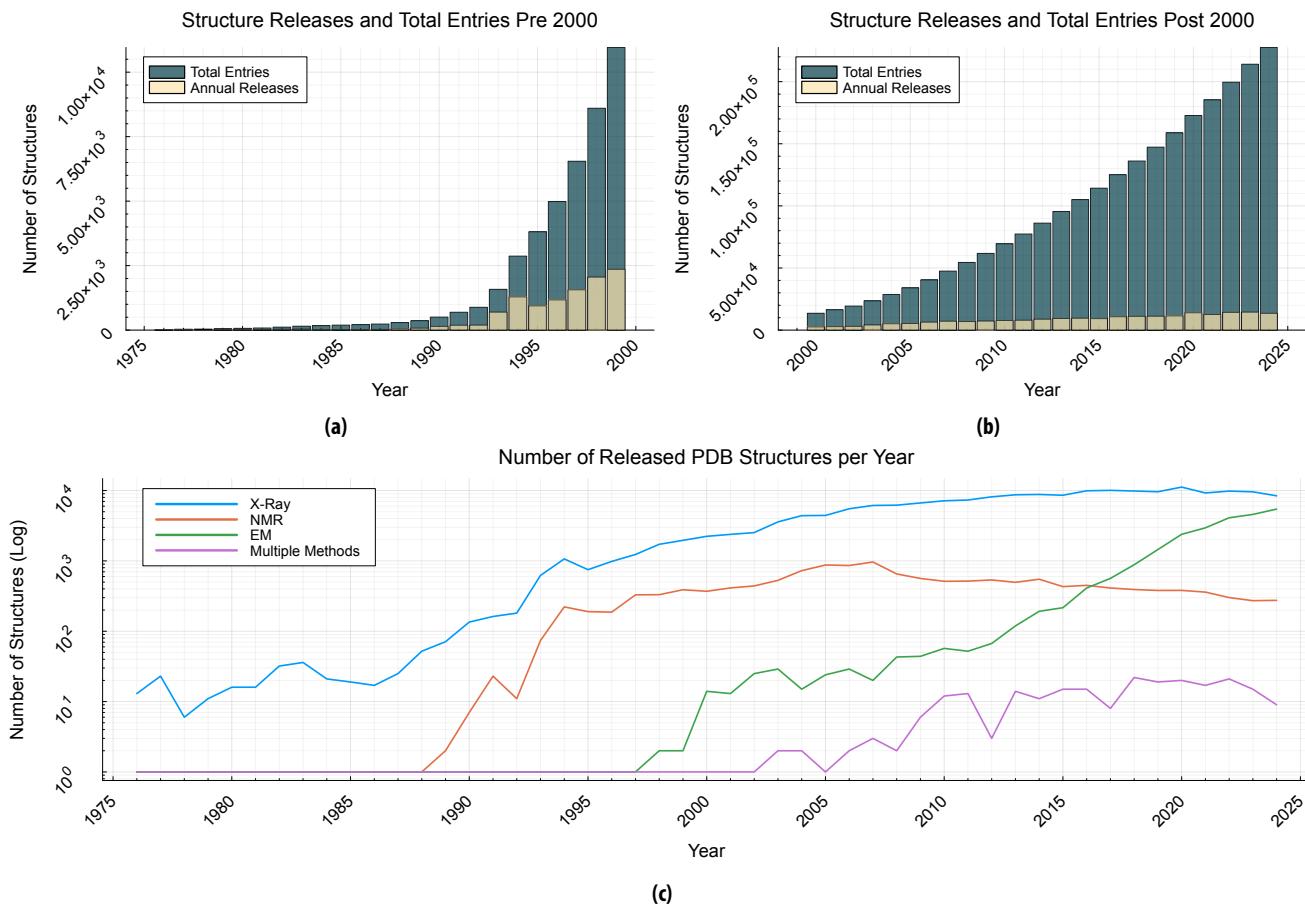


Figure 3. (a) and (b): annual protein structures released to the PDB; (c): structures released by method.¹¹ Created with JuliaLang¹⁴.

been determined. This is compounded by structure elucidation requires months to years, and many proteins cannot be studied in their native dynamic states (Figure 2). Over time, several *in silico* methods have been developed over the decades aimed to expedite PSD by PSP (Figure 4).¹⁵

Most models can be divided into Template-based Modeling (TBM) and Free Modeling (FM).^{15,16} TBM approaches typically predict proteins based on pre-existing structurally similar homologs from PDB; FM mostly uses physics based simulations to determine protein structure, and are also referred as *in silico* or *ab initio* modelling.¹⁷ Although simplistic, the TBM/FM classification has been adopted by the Critical Assessment of Structure Prediction (CASP)¹⁸ competition, a biennial PSP competition created by structural biologist John Moult with the aim of providing a benchmark of the current progress of PSP.

A key technique in TBM is Multiple Sequence Alignment (MSA). By comparing the protein sequence to a database of other protein sequences, structural, evo-

lutionary patterns, domain information can be elucidated from homologs.^{15,19} With generally superior structural accuracy and reduced computational costs, web servers such as SWISS-MODEL²⁰ provides fully automated homology modeling workflow. A general limitation of MSA based TBM is the reduced ability to predict novel, synthetic, and mutant proteins. Even minor changes to a residue's sequence can significantly alter residue charge, allowed Ramachandran angles, flexibility, and overall protein structure. Non-physics-based MSA fails to capture the chemical impact mutations can have on the overall free energy landscape of the protein. This limitation leads to the shift to a more physical approach to PSP, FM.¹⁹

Molecular Dynamics (MD) models like GROMACS²¹, AMBER²², and CHARMM²³, are FM methods using various level of theory for PSP.²⁴ While improvements continue, FM is limited by protein size.¹⁹ *Ab initio* methods like PEP-FOLD3²⁵ and Quark²⁶ achieve $\lesssim 3 \text{ \AA}$ backbone C α Root-Mean-Square Deviation (RMSD) vs experiments for ≤ 20 residues in < 14 hours. However,

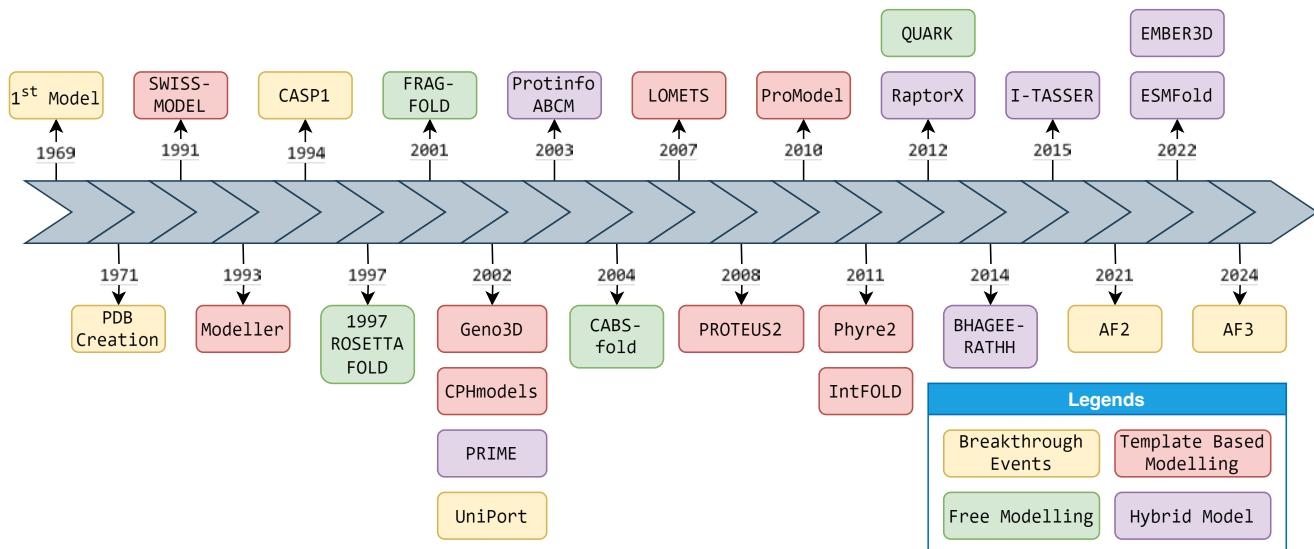


Figure 4. Timeline of key computational models developed over the past 50 years. Adopted from Bertoline *et al.*¹⁵ Created using Draw.io⁵.

computation time increases exponentially for larger proteins and longer sampling. For the 20-residue Trp-cage (PDB:2JOF, Figure 5),²⁷ 208 μs MD with CHARMM²³ takes 287 days using 16 cores and 4 A100 GPUs.^{19,28}



Figure 5. NMR structure of superimposed multiple conformers of 2JOF²⁷ illustrating disordered regions (residue length: 20). Created with ChimeraX⁴.

To address issues FM/TBM possess such as limitations on protein size, homology, and high computational costs, researchers turned to deep learning (DL) to explore new methods (Figure 4).

II. AF2'S IMPACT ON PSP

i. AF2's Release

In the 2021 CASP14 competition¹⁸, Jumper *et al.* entered with AF2, the first PSP algorithm with atomic accuracy.²⁹ AF2 performed significantly better compared to the next best candidate in the TBM, FM/TBM,

and the more difficult FM in the Tertiary Structure (TS) predictions category. Moreover, AF2 scored 244.0 in summed z-scores compared to 90.8 by the next group in Figure 6, while achieving atomic accuracy of RMSD at 95% residue coverage (RMSD₉₅) of $\lesssim 2 \text{ \AA}$ shown in Table 1. This kind of mean accuracy was never seen in previous CASP competitions, marking AF2 a groundbreaking approach to PSP.

Method	Backbone-accuracy (RMSD ₉₅)	All-Atom-accuracy (RMSD ₉₅)
AlphaFold	0.96 Å (95% CI: 0.85-1.16 Å)	1.5 Å (95% CI: 1.2-1.6 Å)
Next Best	2.8 Å (95% CI: 2.7-4.0 Å)	3.5 Å (95% CI: 3.1-4.2 Å)

Table 1. Comparison of AF2 and the next best method in CASP14.^{18,30}

ii. Architecture of AF2

AF2 uses DL-based neural networks trained on the evolutionary, physical, and stereochemical constraints of protein structures.^{30,32} This expanded the traditional TBM/FM categories with DL methods.^{15,19,29}

AF2 has two main stages (Figure 7). The first stage combines the TBM-based MSA module and FM-based Pair Representation (PR) module. MSA searches databases like PDB for homologs, creating an $N_{\text{seq}} \times N_{\text{res}}$ matrix to identify evolutionarily related proteins and conserved regions. PR encodes chemical features between residue pairs in $N_{\text{res}} \times N_{\text{res}}$ matrices, then operates with pre-trained templates using triangular attention.²⁹

The key innovation Evoformer allows rapid infor-

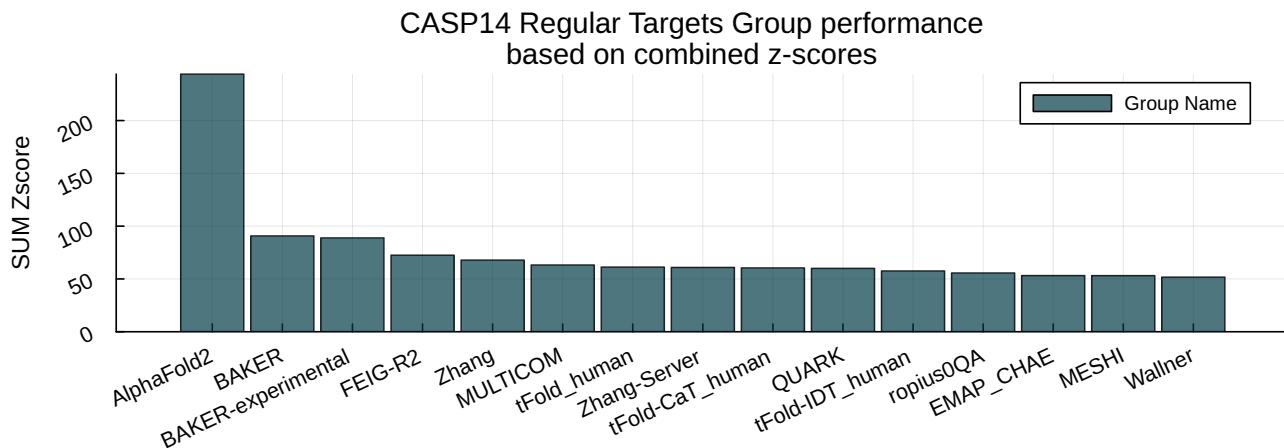


Figure 6. AF2 CASP14 Performance in summed z-score across all TS categories.^{18,30} The z-score measures how well predicted structure aligns with the template. Higher z-score indicates more statistically significant overlap between the predicted structure to template structure compared to a set of random Monte-Carlo structures.³¹ Created with JuliaLang¹⁴.

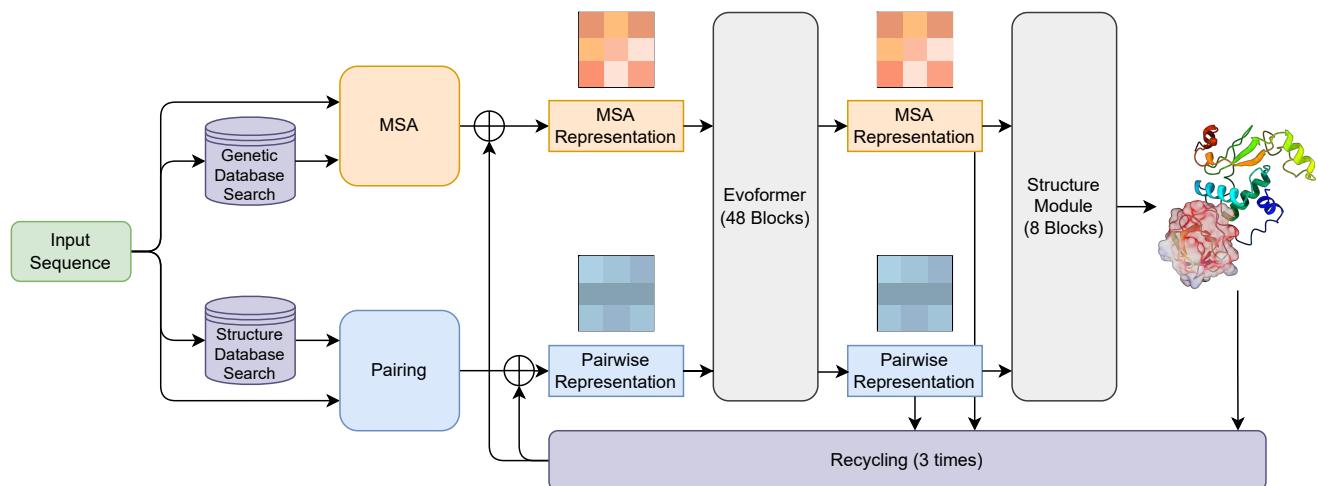


Figure 7. Simplified architecture of AF2 adopted from Jumper *et al.*²⁹ Created with Draw.io⁵.

mation exchange between MSA and PR Evoformer employs variations of attention ranging from row-wise attention for MSA processing, and computationally cheaper triangular attention for PR processing.²⁹ The attention mechanism used for Transformers pioneered by Vaswani *et al.*³³ is a DL algorithm that embeds key-value pairs in a multidimensional embedding space using vectors, capturing relationships between different tokens. This embedding space contains all the information and relations between different concepts.³⁴ Originally developed for large language models and machine translation^{34–36}, attention in the Evoformer allows the model to dynamically focus on relevant relationships within the protein sequence and structure through MSA and PR.²⁹

The Structure Module transforms Evoformer's out-

put into 3D coordinates using Invariant Point Attention (IPA), which determines relative positions while maintaining geometric consistency. It predicts backbone and side chain positions using per-residue rotations and translations, enabling long-range awareness. This approach achieved 0.96 Å RMSD₉₅ accuracy at CASP14, even for proteins without close homologs.^{18,29,30}

iii. Achievements and Limitations of AF2

PPIs are crucial for understanding biological processes from cell signaling and gene regulation to immune responses and disease mechanisms. Determining individual protein structures alone is not enough to comprehend PPIs.³⁹ The initial release of AF2 was extensively used to predict PPIs through methods that incorporated multiple AF2-predicted domains together. Benchmark studies showed that AF2 had more ac-

Category	Before AlphaFold	After AlphaFold	Improvement
General Structural Coverage	48%	76%	+28%
Dark Proteome	26%	10%	-16%
Disease Mutations (ClinVar)	69%	94%	+25%
Oncogenic Mutations	88%	91%	+3%

Table 2. Impact of AF2 on structural coverage of the human proteome. The dark proteome represents protein regions with no structural coverage or predicted features. Disease mutations refer to pathogenic variants in ClinVar³⁷, while oncogenic mutations include cancer-associated variants.^{15,38}

curate heteromeric interface predictions compared to other computational methods, with some predictions showing near-native structures.^{40,41} However, AF2 struggled with accurate antibody-antigen complex predictions.⁴¹ Jumper *et al.* addressed this with AlphaFold-Multimer, which achieved significant improvements in heteromeric interface (+27%) and flexible linker prediction (+14%) compared to AF2.⁴²

Working in partnership with European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), AF2 released over 200 million protein structures, 1,000 times more than that of PDB's experimentally determined structures.^{29,43} Table 2 shows how much AF significantly reduced the sequence-structure gap in the human proteome, with even greater structural coverage improvements observed in other organisms (81–85% for bacterial proteomes).^{15,38}

AF2's impact extends beyond *in silico* PSP, it is actively used by experimentalists to expedite PSD. In crystallography, both the amplitude and the phase of diffracted spots are needed. Experimentally, only the amplitude can be measured directly.⁴⁶ Phase information from homologs of the protein in study is often used as a close approximation known as Molecular Replacement (MR).⁴⁶ When homologs are unavailable, additional phase-determining Multi-wavelength Anomalous Diffraction (MAD) or Single-wavelength Anomalous Diffraction (SAD) experiments are needed. With the advancements of AF2's PSP providing models for MR, the phasing problem for crystallography has significantly improved.⁴⁶

Flower *et al.* investigated the efficacy of using AF2 for MR for a SARS-CoV-2 protein (ORF8, PDB:7JTL, at 2.04 Å).^{44,47} SAD was used to obtain phase information as close homologs for MR were not available. Comparing the experimentally determined structure (ORF8, PDB:7JTL) with AF2's prediction (T1064TS427_1-D1) (Figure 8), AF2's PSP was of sufficient accuracy to yield a correct MR solution. AF2 not only passes one of the more stringent structural tests, it also paints AF2 as a

formidable tool for experimentalists using crystallography and by extension Cryo-EM.⁴⁷

Despite the impressive accuracy of AF2, limitations persists. Details such as loop regions, side-chain confirmations, and disordered regions that are critical for function are still inaccurate as seen in Figure 8.⁴⁷

Beyond accuracy limitations, AF2 faces significant challenges in predicting Intrinsically Disordered Regions (IDRs) and Intrinsically Disordered Proteins (IDPs), which constitute approximately 30% of the human proteome.⁴⁸ These IDRs/IDPs are particularly important in drug development as they play crucial roles in cellular signaling, transcription regulation, and PPIs networks.⁴⁹ Ruff *et al.* demonstrated that regions with low prediction confidence in AF2 strongly correlate with IDRs. A benchmark study of AF2's performance on protein loop regions revealed that while the model accurately predicts short loops of 10 residues or less ($\bar{x}_{\text{RMSD}} < 0.33 \text{ Å}$), its accuracy significantly decreases for longer sequences ($\bar{x}_{\text{RMSD}} > 2.04 \text{ Å}$ for 20 residues).⁵⁰

Although AF2 excels at predicting single protein conformers, it struggles to accurately capture both apo and holo forms of proteins, exhibiting a 70% bias towards the holo form.⁵¹ Scardino *et al.* investigated AF2's PSP efficacy for High-throughput Screening (HTS) in protein docking applications. They concluded that AF2-predicted structures are not accurate enough for docking-based virtual screening compared to experimentally determined PDB structures, as even small side chain variations such as that of Figure 8 in AF2 predictions significantly impact docking performance. Additional refinements are needed before AF2 can be reliably used for HTS in protein docking.⁵² Furthermore, AF2 shows limited accuracy for membrane proteins⁵³ due to both the additional complexity of the lipidome and insufficient membrane protein structures in the PDB training dataset.

The biggest limitations of AF2 is the inability to predict novel structures, PPIs, mutant proteins, complexes with ligands, DNA/RNA protein complexes due to the

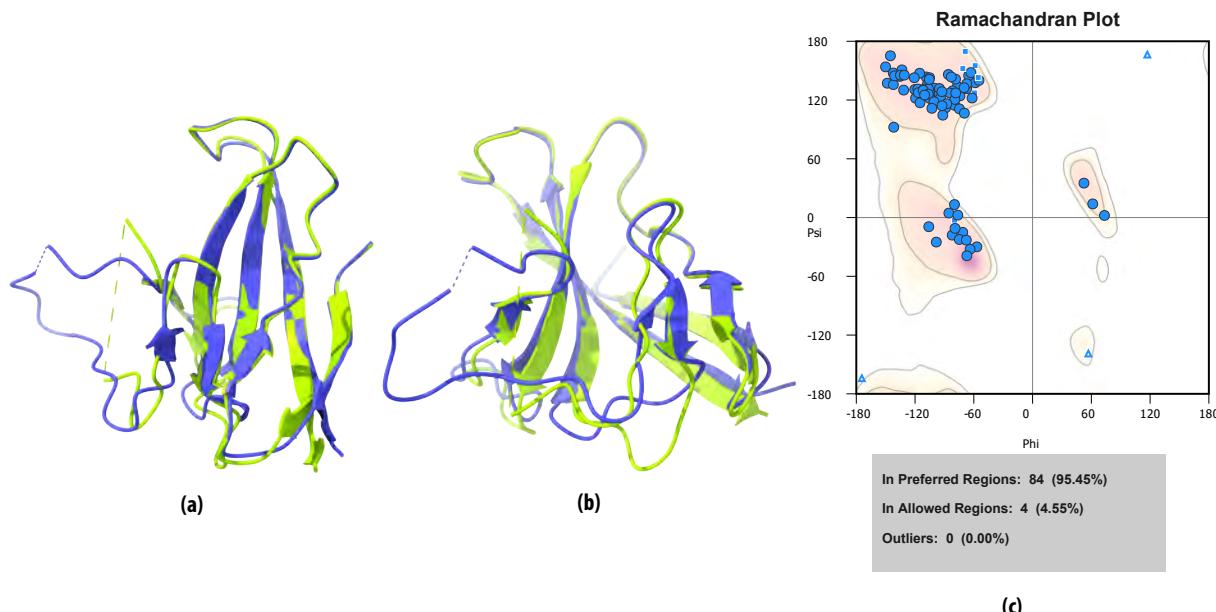


Figure 8. (a) and (b): Structural alignment of experimentally determined CoV-2 ORF8 crystal structure (PDB:7JTL)⁴⁴ in green with AF2 PSP (T1064TS427_1-D1)^{18,29} in blue. Created with ChimeraX⁴. (c): Ramachandran plot analysis of (T1064TS427_1-D1), showing well behaved C_α dihedral angles using WinCoot⁴⁵.

heavy reliance of MSA.²⁹ AF2 also suffers from high computational cost as numbers of residue increases for PSP.¹⁵ The release of AF3 addressed many limitations of AF2, once more changing the landscape of PSP.

III. AF3'S IMPACT AND FUTURE OUTLOOK

i. Release of AF3

AF3, released by Google DeepMind in May 2024, represents a major advancement in PSP. Its developers, John Jumper and Demis Hassabis, were awarded the Nobel Prize in Chemistry later that year for their groundbreaking work.^{54,55} AF3 simplifies AF2's Evolver with the Pairformer module, prioritising PR over MSA processing for more accurate novel structure prediction. It employs a diffusion-based DL approach to iteratively refine uncertain structures. Operating directly on atomic coordinates without rotational frames or specialised bonding patterns, AF3 can handle arbitrary chemical components and accurately predict structures across the broader biomolecular space, including PPIs, protein-ligand binding, and protein-nucleic acid interactions.⁵⁴

The diffusion module (Figure 9) simplifies stereochemistry and bonding, allowing iterative denoising and refinement of a protein's tertiary, secondary, and primary structures, akin to a text to image gener-

ator. Despite occasional chirality and clash issues, AF3 outperforms specialised tools in most categories.⁵⁴ Challenges remain in predicting multiple conformations and complex interactions involving significant changes or dynamic systems.^{56,57}

ii. AF3 Comparative Studies

Recent investigations have evaluated AF3's PPIs prediction capabilities against specialised algorithms, notably the MultiTask-Topological Laplacian (MT-TopLap) model⁵⁶. MT-TopLap, developed by Wee *et al.*, implements Topological Deep Learning (TDL)⁵⁸ to analyse PPIs and predict mutation-induced binding free energy changes using structural data. The comparative assessment used SKEMPI 2.0, the most comprehensive PPIs database containing 317 protein-protein complexes with 8,330 documented mutations⁵⁹.

Wee *et al.* revealed that AF3-predicted structures, when analysed using MT-TopLap, achieved a Pearson correlation coefficient of 0.86 compared to 0.88 for experimental PDB structures (Table 3). This performance surpasses several established TDL methods that require experimental structural data as input. Notably, AF3 accomplishes this using only residue sequence information. However, the study also identified limitations: AF3's predictions exhibited an 8.6% increase in RMSE compared to experimental structures, with complexes such as 4YH7^{54,59} showing significant structural deviations from crystallographic data (Figure 10) due

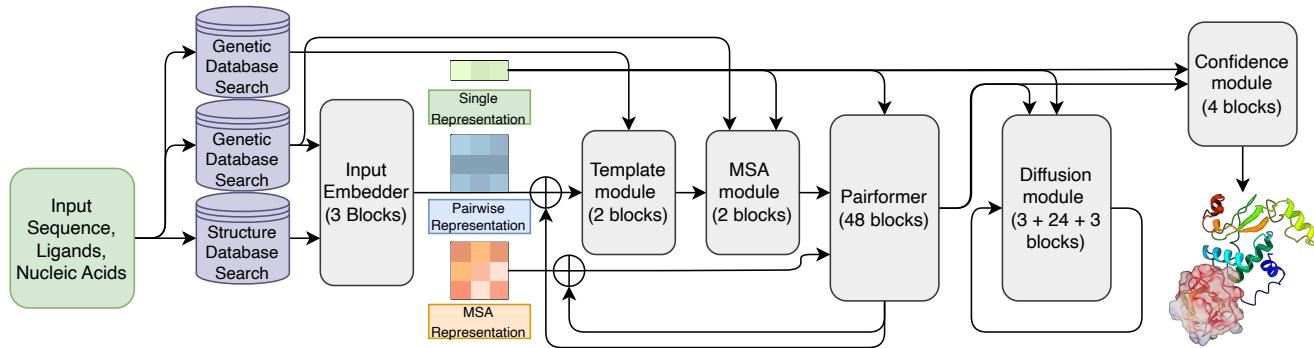


Figure 9. The biggest architectural change in AF3 is the vastly reduced emphasis on MSA, allowing most chemical information to be processed through Pairformer. The new diffusion based module reduces computational cost while increasing compatibility with wider biomolecular molecules, directly generating all atom coordinates without additional geometric reasoning.⁵⁴ Adopted from Abramson *et al.*⁵⁴ Created with Draw.io.⁵

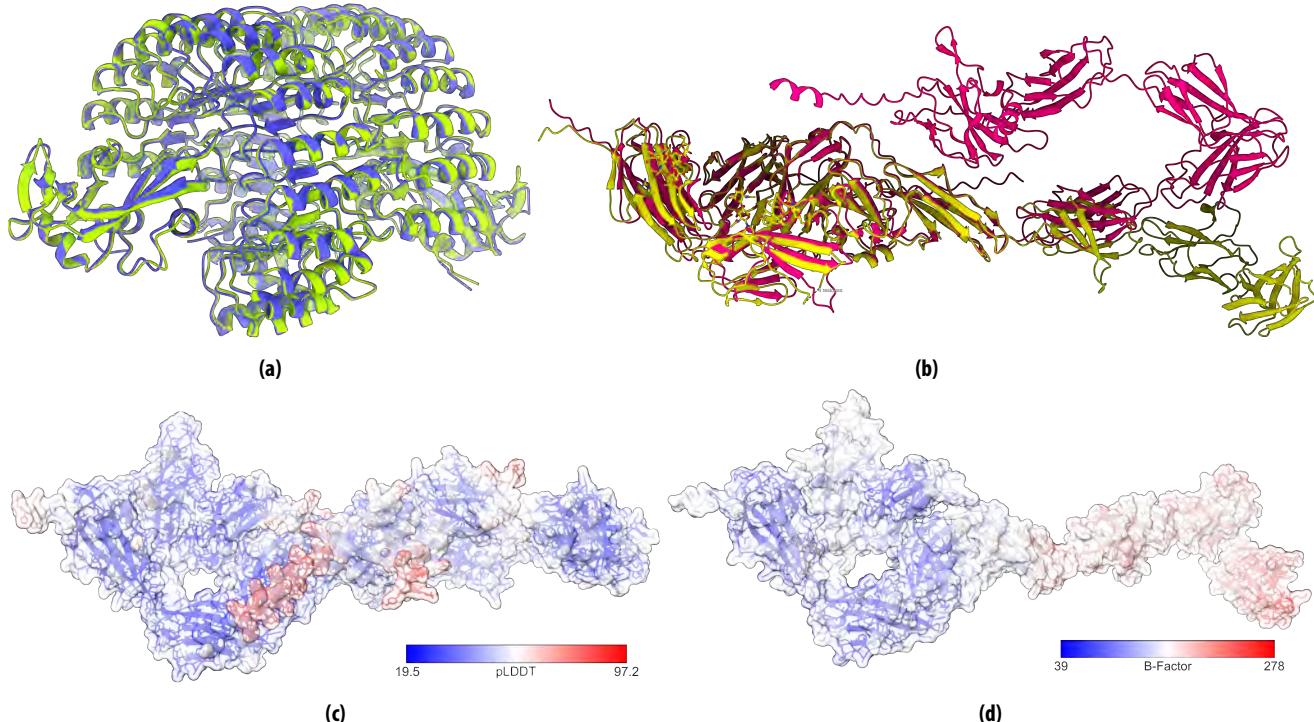


Figure 10. (a) Structural alignment of SKEMPI 2.0⁵⁹ mutant 1A4Y (blue) with AF3-predicted structure^{54,56} (green) demonstrating high alignment accuracy. (b) Structural comparison of mutant 4YH7⁵⁹ (yellow) with AF3 prediction^{54,56} (magenta) showing significant deviations in IDRs. (c) AF3's Predicted Local Distance Difference Test (pLDDT), a per-residue measure of local confidence test (from 0-100), with crystallographic B-factors (d) demonstrating strong correlation between regions of high uncertainty and elevated B-factors. Created in ChimeraX⁴.

to IDRs.

AF3's structural prediction accuracy is demonstrated through complementary metrics (Table 4). The RMSD value of 1.61 Å indicates high structural similarity to experimental structures.⁶² Average Interface Predicted Template Modeling Score (ipTM) and predicted Template Modeling Score (pTM) scores surpass confidence thresholds⁶³, with 71.6% of complexes achieving high-confidence interface predictions and 98.7% producing acceptable overall structures.⁵⁶ While higher RMSE

suggests room for improvement, AF3 achieves reliable accuracy for most applications, though experimental validation may still benefit high-precision studies.⁵⁶

iii. Stereochemical Violation Refinement

However, despite these promising results, AF3 is prone to stereochemical hallucination^{54,64}. Two main types of violations were observed: 4.4% of AF3's Pose-Busters benchmark structures violates chirality, and steric clashing were also observed.⁵⁴ In contrast to

Method ^{54,56,59}	Pearson-Correlation	RMSE (kcal/mol)
MT-TopLap (PDB structures)	0.88	0.937 ± 0.018
MT-TopLap (AF3 structures)	0.86	1.025 ± 0.015
TopLapNetGBT ⁶⁰	0.87	-
TopLapNet ⁶⁰	0.87	-
TopNetGBT ⁶⁰	0.87	-
TopNet ⁶⁰	0.86	-
TopLapGBT ⁶⁰	0.85	-
TopGBT ⁶⁰	0.85	-
LapNetGBT ⁶⁰	0.83	-
mCSM-PPI2 ⁶¹	0.82	-

Table 3. Comparative analysis of protein-protein binding free energy prediction methods. The evaluation encompasses both experimentally determined (PDB) and AF3-predicted structures, highlighting AF3's competitive performance despite relying solely on sequence information.

Metric	Value	Description
Average RMSD	1.61 Å	Structural alignment between AF3 and PDB structures
Average ipTM	0.803	Interface prediction accuracy
Average pTM	0.847	Overall structure prediction accuracy
High ipTM (≥ 0.8)	71.6%	% complexes with confident predictions
High pTM (≥ 0.5)	98.7%	% complexes with acceptable overall structure
RMSE Increase	8.6%	% difference compared to PDB structures

Table 4. Key performance metrics of AF3-predicted protein-protein complexes.^{54,56,60,61} ipTM scores evaluate interface prediction accuracy, while pTM scores assess overall structure quality. RMSD measures structural deviation from experimental structures.

AF2's use of residue specific frames and side-chain torsion angles, AF3's more generalised approach of using diffusion model (denoising random coordinates directly operating on raw atom coordinates) and the lack of built-in stereochemical rules requires the model to implicitly learn all stereochemical relationships from training data alone. This presents particular challenges for stereochemistry.

Incorrect residue side chain stereochemistry at key binding sites can be catastrophic for drug design even if structure is mostly accurate.⁶⁵ AF3 team tried methods such as using AF2 to train AF3 for adopting correct stereochemistry, implemented penalties to ranking model predictions, generating multiple structures of the same residue sequence and selecting for best stereochemical validity.⁵⁴ However, these fixes cannot resolve the inherent hallucination problem diffusion models possess.

Other methods aimed to improve stereochemical

by reintroducing TBM back to AF3, and adding strict stereochemical constraints or validation layers reintroduces the issue of longer computational time and homology dependency.^{64,66,67} Liu *et al.* used a mixture of the proposed improvements and developed PreStoi, a "novel approach integrating AF3 predictions, homologs template data, and template-based stoichiometry adjustment"⁶⁴. PreStoi entered CASP16 in December of 2024 under the team name MULTICOM and performed amicably in many categories, and outcompeted all other predictors in stoichiometry prediction methods.^{64,68}

The shift from PSP (AF2) to general molecular structure prediction (AF3) highlights the challenge of balancing flexibility, speed, and stereochemical validity. While current workarounds are reasonable, integrating explicit chemical knowledge while compromising speed with DL will likely be necessary for another breakthrough.

IV. CONCLUSION

AlphaFold has brought about paradigm shifts in structural biology that extends far beyond PSP. While AF2 achieved the first breakthrough in atomic-level accuracy for PSP, AF3's diffusion-based approach and reduced reliance on MSA has fundamentally transformed the possibilities in structural biology and drug discovery. Despite ongoing challenges with stereochemical validation and complex molecular interactions, the rapid advancement of new predictors inspired by AlphaFold has united experimentalists and computational biologists in accelerating PSD with unprecedented accuracy and efficiency. The future outlook of PSP using DL methods is truly exciting, with the potential to revolutionise drug discovery and other fields.⁶⁹

Word count: 2998

REFERENCES

1. Staker, B. L.; Buchko, G. W.; Myler, P. J., Recent contributions of structure-based drug design to the development of antibacterial compounds, *Current opinion in microbiology*, **2015**, 27, 133–138, DOI: [10.1016/j.mib.2015.09.003](https://doi.org/10.1016/j.mib.2015.09.003).
2. Dill, K. A.; MacCallum, J. L., The protein-folding problem, 50 years on, *Science (New York, N.Y.)*, **2012**, 338 (6110), 1042–1046, DOI: [10.1126/science.1219021](https://doi.org/10.1126/science.1219021).
3. Kresge, N.; Simoni, R. D.; Hill, R. L., The Thermodynamic Hypothesis of Protein Folding: the Work of Christian Anfinsen, *Journal of Biological Chemistry*, **2006**, 281 (14), e11–e13, DOI: [10.1016/S0021-9258\(19\)56522-X](https://doi.org/10.1016/S0021-9258(19)56522-X).
4. Meng, E. C.; Goddard, T. D.; Pettersen, E. F.; Couch, G. S.; Pearson, Z. J.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Tools for structure building and analysis, *Protein Science*, **2023**, 32 (11), e4792, DOI: [10.1002/pro.4792](https://doi.org/10.1002/pro.4792).
5. JGraph, draw.io **2021**.
6. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis, *Nature*, **1958**, 181 (4610), 662–666, DOI: [10.1038/181662a0](https://doi.org/10.1038/181662a0).
7. Rosenthal, P. B.; Henderson, R., Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy, *Journal of Molecular Biology*, **2003**, 333 (4), 721–745, DOI: [10.1016/j.jmb.2003.07.013](https://doi.org/10.1016/j.jmb.2003.07.013).
8. Palmer, A. G.; Patel, D. J., Kurt Wüthrich and NMR of Biological Macromolecules, *Structure*, **2002**, 10 (12), 1603–1604, DOI: [10.1016/S0969-2126\(02\)00915-2](https://doi.org/10.1016/S0969-2126(02)00915-2).
9. Davis, A. M.; St-Gallay, S. A.; Kleywegt, G. J., Limitations and lessons in the use of X-ray structural information in drug design, *Drug discovery today*, **2008**, 13 (19–20), 831–841, DOI: [10.1016/j.drudis.2008.06.006](https://doi.org/10.1016/j.drudis.2008.06.006).
10. Acharya, K. R.; Lloyd, M. D., The advantages and limitations of protein crystal structures, *Trends in Pharmacological Sciences*, **2005**, 26 (1), 10–14, DOI: [10.1016/j.tips.2004.10.011](https://doi.org/10.1016/j.tips.2004.10.011).
11. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank, *Nucleic Acids Research*, **2000**, 28 (1), 235–242, DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).
12. Bammes, B. E.; Rochat, R. H.; Jakana, J.; Chen, D.-H.; Chiu, W., Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency, *Journal of Structural Biology*, **2012**, 177 (3), 589–601, DOI: [10.1016/j.jsb.2012.01.008](https://doi.org/10.1016/j.jsb.2012.01.008).
13. Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M., NMR-Based Methods for Protein Analysis, *Analytical Chemistry*, **2021**, 93 (4), 1866–1879, DOI: [10.1021/acs.analchem.0c03830](https://doi.org/10.1021/acs.analchem.0c03830).
14. Bezzanson, J.; Edelman, A.; Karpinski, S.; Shah, V. B., Julia: A fresh approach to numerical computing, *SIAM Review*, **2017**, 59 (1), 65–98, DOI: [10.1137/14100671](https://doi.org/10.1137/14100671).
15. Bertoline, L. M. F.; Lima, A. N.; Krieger, J. E.; Teixeira, S. K., Before and after AlphaFold2: An overview of protein structure prediction, *Frontiers in Bioinformatics*, **2023**, 3, DOI: [10.3389/fbinf.2023.1120370](https://doi.org/10.3389/fbinf.2023.1120370).
16. Gromiha, M. M.; Nagarajan, R.; Selvaraj, S., Protein Structural Bioinformatics: An Overview, In Ranganathan, S.; Grabskov, M.; Nakai, K.; Schönbach, C., editors, *Encyclopedia of Bioinformatics and Computational Biology*, pp. 445–459, Academic Press Oxford, **2019**, DOI: [10.1016/B978-0-12-809633-8.20278-1](https://doi.org/10.1016/B978-0-12-809633-8.20278-1).
17. Pearce, R.; Zhang, Y., Toward the solution of the protein structure prediction problem, *The journal of biological chemistry*, **2021**, 297 (1), 100870, DOI: [10.1016/j.jbc.2021.100870](https://doi.org/10.1016/j.jbc.2021.100870).
18. Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J., Critical assessment of methods of protein structure prediction (CASP)—Round XV, *Proteins: Structure, Function, and Bioinformatics*, **2023**, 91 (12), 1539–1549, DOI: doi.org/10.1002/prot.26617.
19. Doga, H.; Raubenolt, B.; Cumbo, F.; Joshi, J.; DiFilippo, F. P.; Qin, J.; Blankenberg, D.; Shehab, O., A Perspective on Protein Structure Prediction Using Quantum Computers, *Journal of Chemical Theory and Computation*, **2024**, 20 (9), 3359–3378, DOI: [10.1021/acs.jctc.4c00067](https://doi.org/10.1021/acs.jctc.4c00067).
20. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A.; Rempfer, C.; Bordoli, L.; et al., SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Research*, **2018**, 46 (W1), W296–W303, DOI: [10.1093/nar/gky427](https://doi.org/10.1093/nar/gky427).
21. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, *Journal of Chemical Theory and Computation*, **2008**, 4 (3), 435–447, DOI: [10.1021/ct700301q](https://doi.org/10.1021/ct700301q).
22. Case, D. A.; Cheatham, Thomas E. r.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs, *Journal of computational chemistry*, **2005**, 26 (16), 1668–1688, DOI: [10.1002/jcc.20290](https://doi.org/10.1002/jcc.20290).
23. Hwang, W.; Austin, S. L.; Blondel, A.; Boittier, E. D.; Boresch, S.; Buck, M.; Buckner, J.; Cafisch, A.; Chang, H.-T.; Cheng, X.; et al., CHARMM at 45: Enhancements in Accessibility, Functionality, and Speed, *The Journal of Physical Chemistry B*, **2024**, 128 (41), 9976–10042, DOI: [10.1021/acs.jpcb.4c04100](https://doi.org/10.1021/acs.jpcb.4c04100).
24. Dorn, M.; e Silva, M. B.; Buriol, L. S.; Lamb, L. C., Three-dimensional protein structure prediction: Methods and computational strategies, *Computational Biology and Chemistry*, **2014**, 53, 251–276, DOI: [10.1016/j.combiolchem.2014.10.001](https://doi.org/10.1016/j.combiolchem.2014.10.001).
25. Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P., PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex, *Nucleic acids research*, **2016**, 44 (W1), W449–54, DOI: [10.1093/nar/gkw329](https://doi.org/10.1093/nar/gkw329).
26. Xu, D.; Zhang, Y., Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field, *Proteins*, **2012**, 80 (7), 1715–1735, DOI: [10.1002/prot.24065](https://doi.org/10.1002/prot.24065).
27. Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H., The Trp-cage: optimizing the stability of a globular miniprotein, *Protein Engineering, Design and Selection*, **2008**, 21 (3), 171–185, DOI: [10.1093/protein/gzm082](https://doi.org/10.1093/protein/gzm082).
28. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E., How Fast-Folding Proteins Fold, *Science*, **2011**, 334 (6055), 517–520, DOI: [10.1126/science.1208351](https://doi.org/10.1126/science.1208351).
29. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al., Highly accurate protein structure prediction with AlphaFold, *Nature*, **2021**, 596 (7873), 583–589, DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2).
30. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al., Applying and improving AlphaFold at CASP14, *Proteins: Structure, Function, and Bioinformatics*, **2021**, 89 (12), 1711–1721, DOI: [10.1002/prot.26257](https://doi.org/10.1002/prot.26257).
31. Pearson, W. R.; Lipman, D. J., Improved tools for biological sequence comparison., *Proceedings of the National Academy of Sciences*, **1988**, 85 (8), 2444–2448, DOI: [10.1073/pnas.85.8.2444](https://doi.org/10.1073/pnas.85.8.2444).
32. Bouatta, N.; Sorger, P.; AlQuraishi, M., Protein structure prediction by AlphaFold2: are attention and symmetries all you need?, *Acta Crystallographica Section D*, **2021**, 77 (8), 982–991, DOI: [10.1107/S2059798321007531](https://doi.org/10.1107/S2059798321007531).
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I., Attention Is All You Need **2023**, DOI: [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762).
34. Bahdanau, D.; Cho, K.; Bengio, Y., Neural Machine Translation by Jointly Learning to Align and Translate **2016**, DOI: [10.48550/arXiv.1409.0473](https://doi.org/10.48550/arXiv.1409.0473).
35. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y., Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation **2014**, DOI: [10.48550/arXiv.1406.1078](https://doi.org/10.48550/arXiv.1406.1078).
36. Sutskever, I.; Vinyals, O.; Le, Q. V., Sequence to sequence with Neural Networks **2014**, DOI: doi.org/10.48550/arXiv.1409.3215.
37. Landrum, M. J.; Lee, J. M.; Riley, G. R.; Jang, W.; Rubinstein, W. S.; Church, D. M.; Maglott, D. R., ClinVar: public archive of relationships among sequence variation

- and human phenotype, *Nucleic Acids Research*, **2013**, 42 (D1), D980–D985, DOI: [10.1093/nar/gkt113](https://doi.org/10.1093/nar/gkt113).
38. Porta-Pardo, E.; Ruiz-Serra, V.; Valentini, S.; Valencia, A., The structural coverage of the human proteome before and after AlphaFold, *PLOS Computational Biology*, **2022**, 18 (1), 1–17, DOI: [10.1371/journal.pcbi.1009818](https://doi.org/10.1371/journal.pcbi.1009818).
 39. Liddington, R. C., *Structural Basis of Protein-Protein Interactions*, , pp. 3—14, Humana Press Totowa, NJ, **2004**, DOI: [10.1385/1-59259-762-9:003](https://doi.org/10.1385/1-59259-762-9:003).
 40. Bryant, P.; Pozzati, G.; Elofsson, A., Improved prediction of protein-protein interactions using AlphaFold2, *Nature Communications*, **2022**, 13 (1), 1265, DOI: [10.1038/s41467-022-28865-w](https://doi.org/10.1038/s41467-022-28865-w).
 41. Yin, R.; Feng, B. Y.; Varshney, A.; Pierce, B. G., Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants, *Protein Science*, **2022**, 31 (8), e4379, DOI: [10.1002/pro.4379](https://doi.org/10.1002/pro.4379).
 42. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al., Protein complex prediction with AlphaFold-Multimer, *bioRxiv*, **2022**, DOI: [10.1101/2021.10.04.463034](https://doi.org/10.1101/2021.10.04.463034).
 43. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al., Highly accurate protein structure prediction for the human proteome, *Nature*, **2021**, 596 (7873), 590–596, DOI: [10.1038/s41586-021-03828-1](https://doi.org/10.1038/s41586-021-03828-1).
 44. Flower, T. G.; Buffalo, C. Z.; Hooy, R. M.; Allaire, M.; Ren, X.; Hurley, J. H., Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein, *Proceedings of the National Academy of Sciences*, **2021**, 118 (2), e2021785118, DOI: [10.1073/pnas.2021785118](https://doi.org/10.1073/pnas.2021785118).
 45. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot, *Acta Crystallographica Section D*, **2010**, 66 (4), 486—501, DOI: [10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493).
 46. Hendrickson, W. A., Facing the phase problem, *IUCrJ*, **2023**, 10 (5), 521—543, DOI: [10.1107/S2052252523006449](https://doi.org/10.1107/S2052252523006449).
 47. Flower, T. G.; Hurley, J. H., Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8, *Protein science: a publication of the Protein Society*, **2021**, 30 (4), 728–734, DOI: [10.1002/pro.4050](https://doi.org/10.1002/pro.4050).
 48. Necci, M.; Piovesan, D.; Hoque, M. T.; Walsh, I.; Iqbal, S.; Vendruscolo, M.; Sormanni, P.; Wang, C.; Raimondi, D.; Sharma, R.; et al., Critical assessment of protein intrinsic disorder prediction, *Nature Methods*, **2021**, 18 (5), 472–481, DOI: [10.1038/s41592-021-01117-3](https://doi.org/10.1038/s41592-021-01117-3).
 49. Ruff, K. M.; Pappu, R. V., AlphaFold and Implications for Intrinsically Disordered Proteins, *Journal of Molecular Biology*, **2021**, 433 (20), 167208, DOI: [10.1016/j.jmb.2021.167208](https://doi.org/10.1016/j.jmb.2021.167208).
 50. Stevens, A. O.; He, Y., Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction, *Biomolecules*, **2022**, 12 (7), DOI: [10.3390/biom12070985](https://doi.org/10.3390/biom12070985).
 51. Saldaño, T.; Escobedo, N.; Marchetti, J.; Zea, D. J.; Mac Donagh, J.; Velez Rueda, A. J.; Gonik, E.; García Melani, A.; Novomisky Nechcoff, J.; Salas, M. N.; et al., Impact of protein conformational diversity on AlphaFold predictions, *Bioinformatics*, **2022**, 38 (10), 2742–2748, DOI: [10.1093/bioinformatics/btac202](https://doi.org/10.1093/bioinformatics/btac202).
 52. Scardino, V.; Di Filippo, J. I.; Cavasotto, C. N., How good are AlphaFold models for docking-based virtual screening?, *iScience*, **2023**, 26 (1), 105920, DOI: [10.1016/j.isci.2022.105920](https://doi.org/10.1016/j.isci.2022.105920).
 53. Azzaz, F.; Yahi, N.; Chahinian, H.; Fantini, J., The Epigenetic Dimension of Protein Structure Is an Intrinsic Weakness of the AlphaFold Program, *Biomolecules*, **2022**, 12 (10), DOI: [10.3390/biom12101527](https://doi.org/10.3390/biom12101527).
 54. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3, *Nature*, **2024**, 630 (8016), 493–500, DOI: [10.1038/s41586-024-07487-w](https://doi.org/10.1038/s41586-024-07487-w).
 55. O'Leary, K., AlphaFold gets an upgrade (and a Nobel), *Nature Medicine*, **2024**, 30 (12), 3393–3393, DOI: [10.1038/s41591-024-03392-x](https://doi.org/10.1038/s41591-024-03392-x).
 56. Wee, J.; Wei, G.-W., Evaluation of AlphaFold 3's Protein-Protein Complexes for Predicting Binding Free Energy Changes upon Mutation, *Journal of Chemical Information and Modeling*, **2024**, 64 (16), 6676–6683, DOI: [10.1021/acs.jcim.4c00976](https://doi.org/10.1021/acs.jcim.4c00976).
 57. Desai, D.; Kantliwala, S. V.; Vybhavi, J.; Ravi, R.; Patel, H.; Patel, J., Review of AlphaFold 3: Transformative advances in drug design and therapeutics, *Cureus*, **2024**, 16 (7), e36346, DOI: [10.7759/cureus.63646](https://doi.org/10.7759/cureus.63646).
 58. Cang, Z.; Wei, G.-W., TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions, *PLOS Computational Biology*, **2017**, 13 (7), 1–27, DOI: [10.1371/journal.pcbi.1005690](https://doi.org/10.1371/journal.pcbi.1005690).
 59. Jankauskaitė, J.; Jiménez-García, B.; Dapkūnas, J.; Fernández-Recio, J.; Moal, I. H., SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation, *Bioinformatics*, **2018**, 35 (3), 462–469, DOI: [10.1093/bioinformatics/bty635](https://doi.org/10.1093/bioinformatics/bty635).
 60. Chen, J.; Qiu, Y.; Wang, R.; Wei, G.-W., Persistent Laplacian projected Omicron BA.4 and BA.5 to become new dominating variants, *Comput. Biol. Med.*, **2022**, 151 (Pt A), 106262, DOI: [10.1016/j.combiomed.2022.106262](https://doi.org/10.1016/j.combiomed.2022.106262).
 61. Rodrigues, C. H. M.; Myung, Y.; Pires, D. E. V.; Ascher, D. B., mCSM-PPI2: predicting the effects of mutations on protein–protein interactions, *Nucleic Acids Research*, **2019**, 47 (W1), W338–W344, DOI: [10.1093/nar/gkz383](https://doi.org/10.1093/nar/gkz383).
 62. Carugo, O., How root-mean-square distance (r.m.s.d.) values depend on the resolution of protein structures that are compared, *Journal of Applied Crystallography*, **2003**, 36 (1), 125–128, DOI: [10.1107/s0021889802020502](https://doi.org/10.1107/s0021889802020502).
 63. Xu, J.; Zhang, Y., How significant is a protein structure similarity with TM-score = 0.5?, *Bioinformatics*, **2010**, 26 (7), 889–895, DOI: [10.1093/bioinformatics/btq066](https://doi.org/10.1093/bioinformatics/btq066).
 64. Liu, J.; Neupane, P.; Cheng, J., Accurate Prediction of Protein Complex Stoichiometry by Integrating AlphaFold3 and Template Information, *bioRxiv*, **2025**, DOI: [10.1101/2025.01.12.632663](https://doi.org/10.1101/2025.01.12.632663).
 65. Deller, M. C.; Rupp, B., Models of protein–ligand crystal structures: trust, but verify, *Journal of Computer-Aided Molecular Design*, **2015**, 29 (9), 817–836, DOI: [10.1007/s10822-015-9833-8](https://doi.org/10.1007/s10822-015-9833-8).
 66. Gadde, N.; Dodamani, S.; Altaf, R.; Kumar, S., Leveraging AlphaFold 3 for Structural Modeling of Neurological Disorder-Associated Proteins: A Pathway to Precision Medicine, *bioRxiv*, **2024**, DOI: [10.1101/2024.11.18.624211](https://doi.org/10.1101/2024.11.18.624211).
 67. Pan, L.; Wang, A.; Sang, R.; Lu, X.; Chen, W.; Ma, Y.; Deng, F., AlphaFold 3 sheds insights into chemical enhancer-induced structural changes in Cas12a RNPs, *Health Nanotechnology*, **2025**, 1 (1), 1, DOI: [10.1186/s44301-024-00003-z](https://doi.org/10.1186/s44301-024-00003-z).
 68. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Topf, M., 16th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction **2024**, [Accessed 02-02-2025].
 69. Pitt, W. R.; Bentley, J.; Boldron, C.; Colliandre, L.; Esposito, C.; Frush, E. H.; Kopec, J.; Labouille, S.; Meneyrol, J.; Pardoe, D. A.; et al., Real-World Applications and Experiences of AI/ML Deployment for Drug Discovery, *Journal of Medicinal Chemistry*, **2025**, 68 (2), 851–859, DOI: [10.1021/acs.jmedchem.4c03044](https://doi.org/10.1021/acs.jmedchem.4c03044).

ABBREVIATIONS

AF2 AlphaFold 2. [3](#), [6](#), [7](#), [8](#), [9](#), [11](#), [15](#)

AF3 AlphaFold 3. [3](#), [9](#), [10](#), [11](#), [15](#)

CASP Critical Assessment of Structure Prediction. [5](#), [6](#), [7](#), [11](#)

Cryo-EM Cryogenic Electron Microscopy. [4](#), [8](#)

DL Deep Learning. [3](#), [6](#), [7](#), [9](#), [11](#)

FM Free Modeling. [5](#), [6](#)

HTS High-throughput Screening. [8](#)

IDPs Intrinsically Disordered Proteins. [8](#)

IDRs Intrinsically Disordered Regions. [8](#), [10](#)

IPA Invariant Point Attention. [7](#)

ipTM Interface Predicted Template Modeling Score. [10](#), [11](#)

MAD Multi-wavelength Anomalous Diffraction. [8](#)

MD Molecular Dynamics. [5](#), [6](#)

MR Molecular Replacement. [8](#)

MSA Multiple Sequence Alignment. [5](#), [6](#), [7](#), [9](#), [10](#), [11](#)

MT-TopLap MultiTask-Topological Laplacian. [9](#), [11](#)

PDB Protein Data Bank. [4](#), [5](#), [6](#), [8](#), [9](#), [11](#)

pLDDT Predicted Local Distance Difference Test. [10](#)

PPIs Protein-protein Interactions. [3](#), [7](#), [8](#), [9](#)

PR Pair Representation. [6](#), [7](#), [9](#)

PSD Protein Structure Determination. [4](#), [5](#), [8](#), [11](#)

PSP Protein Structure Prediction. [3](#), [4](#), [5](#), [6](#), [8](#), [9](#), [11](#), [15](#)

pTM predicted Template Modeling Score. [10](#), [11](#)

RMSD Root-Mean-Square Deviation. [5](#), [6](#), [10](#), [11](#)

RMSE Root-Mean-Square Error. [9](#), [10](#), [11](#)

SAD Single-wavelength Anomalous Diffraction. [8](#)

TBM Template-based Modeling. [5](#), [6](#), [11](#)

TDL Topological Deep Learning. [9](#)

TS Tertiary Structure. [6](#), [7](#)