Genome Publications
https://doi.org/10.61096/978-81-990998-7-6_8

Chapter 8

Sequence-Based Approaches: Homology Modeling and Protein Structure Prediction

Gaddala Roja
Assistant Professor, Department of Pharmacology, TGSWRPC (W) Mahabubabad, Telangana 506101

Dr. Sangu Vineela
Associate Professor, Department of Pharmaceutics, TGSWRPC (W) Mahabubabad Telangana 506101

Maloth Swathi
Assistant Professor, Department of Pharmaceutics, TGSWRPC (W) Mahabubabad Telangana 506101

Abstract: Accurate prediction of protein three-dimensional structures remains one of the most fundamental challenges and triumphs of modern computational biology. Sequence-based approaches such as homology modeling, threading, and ab initio prediction have enabled the rational design of drug targets even when experimental structures are unavailable. Homology modeling exploits evolutionary relationships between proteins to generate structural models using known templates, whereas ab initio and deep learning methods exemplified by AlphaFold2 and AlphaFold3 have revolutionized predictive accuracy through data-driven learning of residue–residue relationships. This chapter provides a comprehensive exposition of the principles, algorithms, and workflows underlying protein structure prediction, emphasizing template-based and template-free approaches, validation metrics, and practical tutorials using software such as MODELLER and SWISS-MODEL. It also discusses emerging deep learning paradigms, benchmark assessments, applications in drug discovery, and the integration of structure prediction with docking and dynamics simulations to enhance rational design. Limitations related to data quality, model bias, and conformational flexibility are critically analyzed, and future perspectives highlight hybrid AI–physics models and quantum-enhanced prediction strategies that promise to further redefine the field.

Keywords: Homology modeling, AlphaFold, protein structure prediction, threading, MODELLER

Citation: Gaddala Roja, Sangu Vineela, Maloth Swathi. Sequence-Based Approaches: Homology Modeling and Protein Structure Prediction. *Comprehensive Approaches in Computer-Aided Drug Design: QSAR, Docking, Screening, Homology, Pharmacophore and AI-Driven Insights.* Genome Publication. 2025; Pp 93-103. https://doi.org/10.61096/978-81-990998-7-6_8

8.0 INTRODUCTION

The Centrality of Structure Prediction in Modern Drug Design

The three-dimensional structure of a protein defines its biochemical function, interaction specificity, and suitability as a drug target. Determining these structures experimentally through X-ray crystallography, cryo-electron microscopy (cryo-EM), or nuclear magnetic resonance (NMR) is both resource-intensive and time-consuming, often taking months to years per protein. Computational structure prediction provides an indispensable complement by rapidly generating approximate models from sequence data, which can be refined and validated through iterative in silico and experimental workflows. In computer-aided drug design (CADD), such models underpin virtual screening, docking, molecular dynamics simulations, and pharmacophore generation. Sequence-based modeling methods rely on the fundamental principle that proteins with similar amino acid sequences tend to adopt similar three-dimensional conformations. Homology modeling capitalizes on this evolutionary conservation by using experimentally resolved "template" structures to infer the shape of unknown "target" proteins. Threading and ab initio methods, on the other hand, do not depend on close homologues and are therefore invaluable when sequence identity falls below the twilight zone (typically <25%) [1]. These methods, combined with advances in artificial intelligence, now enable atomic-level predictions even for previously intractable proteins.

Over the past decade, developments such as AlphaFold2 (2021) and AlphaFold3 (2024) have transformed the landscape by achieving near-experimental accuracy in global protein folding benchmarks such as CASP (Critical Assessment of Structure Prediction) [2]. Despite these breakthroughs, classical homology modeling remains essential in contexts requiring control over template selection, ligand-bound conformations, and comparative modeling of protein families. Moreover, sequence-based approaches extend beyond monomeric proteins to encompass multimeric complexes, intrinsically disordered regions, and protein–protein or protein–nucleic acid interactions, thus forming a foundation for rational drug discovery pipelines.

8.1 Theoretical Foundations of Protein Structure Prediction

Protein structure prediction involves translating linear amino acid sequences into accurate three-dimensional coordinates. The theoretical challenge lies in navigating the astronomically large conformational space often referred to as Levinthal's paradox where a typical 100-residue protein could theoretically adopt more than 10^100 conformations [3]. Sequence-based algorithms circumvent exhaustive conformational searches by leveraging evolutionary and statistical constraints derived from known structures.

Three major paradigms govern protein structure prediction:

1. Template-Based Modeling (TBM):

Includes homology modeling and threading. These methods assume that structural templates exist in the Protein Data Bank (PDB) that share evolutionary ancestry with the target. Sequence alignment, template fitting, and model refinement collectively yield the predicted structure.

2. Template-Free (Ab Initio) Modeling:

Used when no suitable template is available. It constructs structures from physicochemical principles or through data-driven fragment assembly. Deep learning methods such as trRosetta and AlphaFold have dramatically improved this domain [4].

3. Hybrid and Deep Learning Approaches:

Combine physical modeling, statistical potentials, and neural networks to capture residue–residue contact maps, torsion angles, and long-range interactions. These approaches integrate sequence co-

evolution data and attention-based networks to refine predictions beyond classical methods [5]. Each approach balances accuracy, computational cost, and interpretability. While homology models are relatively fast to generate and validate, ab initio and deep learning methods demand greater computational resources but deliver higher precision for challenging targets.

8.2 Homology Modeling: Concepts and Workflow

Homology modeling, also called comparative modeling, is grounded in the evolutionary assumption that proteins sharing high sequence similarity will exhibit similar structural folds. The method comprises a well-defined pipeline of steps template identification, sequence alignment, model building, loop and side-chain refinement, and model validation each crucial for the accuracy of the final structure.

1. Template Identification

The first step involves searching for homologous structures within databases such as the Protein Data Bank (PDB) using BLASTp or HHpred. Generally, a sequence identity above 40% ensures high-confidence modeling, while models with 25–40% identity require careful refinement and validation [6].

2. Sequence Alignment

Accurate alignment between the target and template sequences ensures the correct transfer of conserved structural motifs. Tools like Clustal Omega, MUSCLE, and PROMALS3D integrate evolutionary and structural information to optimize alignment quality [7].

3. Model Building

Once alignment is established, modeling tools such as MODELLER, SWISS-MODEL, or MOE reconstruct the three-dimensional coordinates of the target by copying the backbone and side-chain geometries from the template. MODELLER, for instance, employs spatial restraints derived from the alignment and uses optimization algorithms based on the satisfaction of spatial restraints to generate the model [8].

4. Loop and Side-Chain Refinement

Loops often correspond to variable regions or insertions not conserved between the target and template. Refinement using molecular dynamics (MD)-based sampling or rotamer libraries improves geometry and stabilizes energetically favorable conformations. Tools such as RosettaLoop or ModRefiner are widely employed for this purpose [9].

5. Model Validation

Model quality is assessed using statistical and stereochemical measures. Popular validation tools include PROCHECK (Ramachandran analysis), Verify3D (residue–environment compatibility), ProSA (Z-score analysis), and MolProbity (geometry and clash assessment). Models with >90% residues in favored Ramachandran regions are generally considered reliable [10]. The homology modeling workflow's efficiency has led to its integration into automated platforms like SWISS-MODEL, I-TASSER, and Phyre2, which can generate and refine models in a few hours.

8.3 Template-Free and Deep Learning Approaches

While template-based methods depend on structural homologues, template-free (ab initio) and deep learning models aim to predict structures solely from sequence information. Historically, ab initio methods relied on fragment assembly guided by energy functions derived from statistical potentials, but their accuracy was limited for proteins larger than 150 residues. The advent of machine learning transformed this field by enabling networks to infer spatial constraints directly from multiple sequence alignments (MSAs).

Fragment Assembly and Ab Initio Modeling:

Classical approaches such as Rosetta's ab initio method assemble short peptide fragments (typically 3–9 residues) derived from known structures and combine them through Monte Carlo sampling to minimize global energy [11]. Despite their computational cost, these methods remain useful for modeling flexible or disordered regions absent in template-based approaches.

Deep Learning Revolution

The introduction of AlphaFold2 in 2021 by DeepMind marked a paradigm shift. AlphaFold2 employs a transformer-based architecture that integrates attention mechanisms to learn residue–residue distance distributions and orientation constraints from MSAs [12]. It uses end-to-end learning to directly predict atomic coordinates, achieving near-experimental accuracy in CASP14 benchmarks with an average global distance test (GDT_TS) score exceeding 90. AlphaFold3, released in 2024, extended this success by modeling protein–ligand and protein–nucleic acid complexes, leveraging joint embeddings for multi-chain systems and incorporating diffusion-based sampling [13].

Comparative Assessment

Although AlphaFold2/3 outperform traditional methods in fold prediction, their accuracy is contingent upon high-quality MSAs and well-populated structural databases. In contrast, homology models offer interpretability and control over template choice, ligand state, and conformational context. Moreover, while AlphaFold predictions may excel in global topology, loop regions and binding site conformations sometimes require further MD refinement before use in docking or virtual screening workflows [14].

8.4 Step-by-Step Homology Modeling Tutorial (MODELLER and SWISS-MODEL)

The growing accessibility of computational tools has democratized protein modeling, enabling researchers to predict target structures even in the absence of crystallographic data. Two widely adopted tools MODELLER and SWISS-MODEL represent the most commonly used pipelines for comparative modeling. Both implement the canonical five-step procedure described earlier but differ in user control, automation level, and integration with web-based resources.

8.4.1 MODELLER Tutorial

MODELLER, developed by Šali and Blundell, uses a spatial restraint approach to generate three-dimensional models based on sequence–structure alignment. It optimizes the model by satisfying geometric and stereochemical constraints derived from template coordinates [15]. It provides a Python interface and command-line workflow that allows high customizability.

8.4.2 SWISS-MODEL Tutorial

SWISS-MODEL is an automated web-based platform maintained by the Swiss Institute of Bioinformatics. It provides an accessible interface integrating template identification, alignment, modeling, and validation within a single pipeline [16]. It is especially useful for researchers lacking programming experience.

8.4.3 Hybrid Workflows

Many researchers combine both tools: an initial SWISS-MODEL run to obtain a baseline model and MODELLER for advanced refinement and loop rebuilding. Further, combining Rosetta Relax or

short molecular dynamics (MD) simulations improves side-chain packing and hydrogen bonding accuracy, yielding models suitable for structure-based drug design [17].

8.5 Model Refinement and Validation Metrics

A model's scientific value depends not only on geometric plausibility but also on its reliability for downstream applications such as docking or pharmacophore modeling. Therefore, comprehensive validation using multiple metrics is essential.

8.5.1 Structural Geometry and Stereochemistry

1. **Ramachandran Plot (PROCHECK):** Evaluates backbone ϕ and ψ angles. Ideally, more than 90% of residues should fall within favored regions.
2. **Bond Lengths and Angles (MolProbity):** Identifies steric clashes and deviations from ideal geometry.
3. **Rotamer Analysis:** Assesses side-chain conformations relative to known rotamer libraries.
4. **Clashscore:** Quantifies van der Waals overlaps, guiding refinement steps.

8.5.2 Energy and Statistical Potentials

1. **DOPE Score (MODELLER):** Estimates atomic-level pseudo-energy; lower values indicate more favorable conformations.
2. **ProSA Z-score:** Compares overall model energy with that of experimentally determined structures; values within the typical native protein range suggest plausibility.
3. **Verify3D and ERRAT:** Evaluate the compatibility of each residue with its 3D environment and detect non-random errors.

8.5.3 Consensus Quality Assessment

Modern practice advocates multi-criteria evaluation. Integrating QMEAN, ProSA, and PROCHECK outputs ensures comprehensive validation. Discrepancies such as acceptable geometry but poor energy profiles signal the need for localized refinement (e.g., flexible loop minimization).

8.5.4 Loop Refinement and Dynamics Optimization

Loops contribute significantly to ligand binding, especially in enzymes and receptors. Dedicated algorithms (e.g., MODELLER's loopmodel, RosettaLoop, GalaxyRefine) or short molecular dynamics (10–50 ns) in GROMACS can relax strained geometries and enhance local accuracy [18]. In many cases, docking performance improves markedly after such refinement, demonstrating the functional importance of post-model optimization.

8.6 Comparative Assessment: Homology, Threading, and Deep Learning Models

The growing diversity of structure prediction methods warrants systematic comparison to guide methodological choice. Table 8.1 summarizes the core characteristics of three main paradigms.

Table 8.1. Comparative Evaluation of Structure Prediction Approaches

Method	Template Requirement	Typical Accuracy (Cα RMSD)	Strengths	Limitations	Representative Tools
Homology Modeling	High (≥25% sequence identity)	1–3 Å (moderate identity)	Fast, interpretable, customizable	Dependent on template; limited for novel folds	MODELLER, SWISS-MODEL, MOE
Threading (Fold Recognition)	Moderate (distant homology)	2–5 Å	Detects remote homologues, uses fold libraries	Lower precision; limited flexibility	Phyre2, I-TASSER, RaptorX
Ab Initio / Deep Learning	None	≤1 Å (AlphaFold2/3)	High accuracy, suitable for novel folds and complexes	Requires large MSAs; limited for disordered regions	AlphaFold2/3, trRosetta, RoseTTAFold

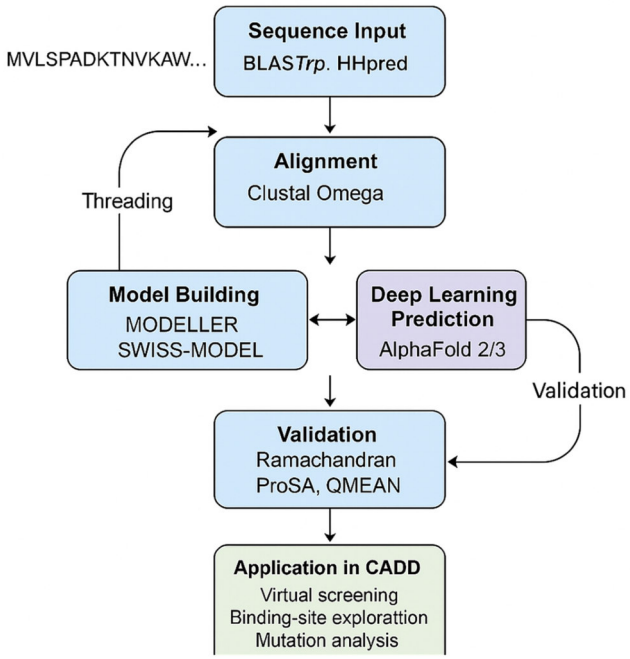

Figure 1

8.6.1 Homology vs Threading

Threading bridges the gap between homology modeling and ab initio prediction. It identifies compatible folds even at low sequence identity by matching target sequences to known structural templates using profile–profile or energy-based alignments. Programs such as I-

TASSER and Phyre2 combine threading with ab initio assembly to improve model accuracy in difficult cases [19].

8.6.2 Deep Learning vs Classical Modeling

Deep learning models, particularly AlphaFold2, outperform classical techniques in CASP benchmarks by integrating co-evolutionary features and attention-based residue embeddings [20]. However, the interpretability and custom control inherent in homology modeling remain valuable. For example, in ligand-bound conformations or mutant modeling, human-guided template selection can outperform automated AI predictions that are agnostic to biological context.

8.6.3 Hybrid Modeling Strategies

An emerging best practice is hybrid modeling, which combines AI predictions with physics-based refinement. AlphaFold-generated structures are refined by Rosetta Relax, MD simulations, or MODELLER restraints, producing experimentally plausible conformations. Such integration ensures both global and local accuracy crucial for active-site geometry in docking or pharmacophore applications [21].

8.7 Applications of Protein Structure Prediction in Drug Discovery

Homology modeling and modern AI-driven structure prediction play transformative roles in computer-aided drug design (CADD), bridging the gap between genomic information and molecular therapeutics. The capacity to predict reliable protein structures enables the rational exploration of binding sites, understanding of disease-related mutations, and the design of ligands with enhanced specificity.

8.7.1 Structure-Based Virtual Screening and Docking

The earliest and most widespread application of homology models is in virtual screening and docking. When experimental structures are unavailable, validated homology models can substitute as receptor templates to predict ligand binding orientations and affinities. For instance, models of G-protein-coupled receptors (GPCRs) once elusive to crystallography enabled the identification of inverse agonists and allosteric modulators well before experimental structures were determined [22]. Studies have demonstrated that docking results using high-quality homology models often achieve comparable enrichment factors to crystal structures, provided the binding site residues are accurately modeled. Refinement through MD simulations or side-chain optimization further improves predictive performance. Software such as AutoDock Vina, Glide, and GOLD can readily interface with these modeled receptors, expanding structure-based screening coverage across the proteome.

8.7.2 Mutation Analysis and Protein Engineering

Homology and AI-based models also underpin mutational analysis, providing insight into the structural consequences of amino acid substitutions. Predictive tools like FoldX, DynaMut, and Rosetta ddG evaluate how mutations alter stability, flexibility, or binding affinity. Such studies aid in interpreting disease-associated variants and designing stable protein mutants for therapeutic or industrial applications [23]. Moreover, by analyzing predicted conformational shifts, researchers can engineer enzyme variants with improved substrate affinity, altered cofactor specificity, or enhanced thermostability. Integration with directed evolution experiments accelerates protein engineering cycles.

8.7.3 Modeling Protein–Protein and Protein–Ligand Interactions

Sequence-based approaches are equally valuable for modeling protein–protein interactions (PPIs), where one partner's structure may be unknown. Homology-derived complexes or AlphaFold-Multimer predictions provide templates for docking studies using HADDOCK or ClusPro, enabling the identification of hot-spot residues involved in complex formation [24]. In ligand design, hybrid models combining homology modeling and AlphaFold predictions have proven particularly effective in kinase, protease, and GPCR systems. For instance, AlphaFold2-generated models of SARS-CoV-2 spike and NSP proteins during the COVID-19 pandemic facilitated immediate virtual screening of antiviral candidates when no experimental structures were available [25].

8.7.4 Pharmacophore and QSAR Model Derivation

Predicted structures can be used to identify pharmacophoric features such as hydrogen-bond donors, hydrophobic pockets, or charged regions. These features guide pharmacophore-based virtual screening or QSAR model generation. The integration of sequence-based modeling with ligand-based design allows the identification of active-site constraints and scaffold hopping strategies [26].

8.7.5 Functional Annotation and Target Discovery

In the genomic era, many open reading frames (ORFs) encode proteins with unknown functions. Homology and threading models can infer functional annotations by mapping predicted structures onto known enzyme folds or binding-site topologies. Tools like I-TASSER and Phyre2 automatically suggest potential enzymatic or receptor functions, supporting large-scale drug target discovery in neglected or emerging pathogens [27].

8.8 Limitations, Challenges, and Critical Considerations

Despite spectacular progress, sequence-based structure prediction remains constrained by several biological, computational, and methodological factors. Recognizing these limitations is essential for responsible interpretation and application in drug discovery.

8.8.1 Template Dependence and Structural Bias

Classical homology modeling is inherently limited by the availability and quality of template structures. PDB entries are biased toward soluble, crystallizable proteins, underrepresenting membrane proteins and intrinsically disordered proteins. Consequently, homology models for these classes exhibit reduced reliability. Even small errors in backbone alignment or loop orientation can cascade into significant inaccuracies in predicted binding sites [28].

8.8.2 Dynamics and Conformational Flexibility

Static models whether generated by homology or deep learning represent only a single conformation of a dynamic protein. Many enzymes undergo large-scale conformational transitions (e.g., open and closed states) that influence ligand binding and catalysis. AI models such as AlphaFold2 capture average conformations but often fail to represent alternate functional states, underscoring the need for integration with molecular dynamics or enhanced sampling methods [29].

8.8.3 Data Quality and Annotation Errors

Deep learning models inherit biases and inaccuracies from their training datasets. Misannotated PDB entries, redundant structures, or missing cofactors can propagate systematic errors

into predictions. Furthermore, sequence alignments with incorrectly defined domain boundaries can lead to distorted folds or misplaced secondary structures.

8.8.4 Limited Modeling of Complexes and Ligands

Although AlphaFold3 now handles protein–ligand and protein–nucleic acid complexes, challenges remain for multi-protein assemblies with flexible interfaces or transient interactions. Ligand modeling often requires complementary docking or quantum mechanical refinement to achieve precise geometries suitable for drug discovery [30].

8.8.5 Interpretability and Reproducibility

The opaque nature of deep learning architectures raises concerns about Interpretability and scientific reproducibility. Whereas homology modeling workflows are transparent and customizable, AI-generated predictions often function as "black boxes." Standardized evaluation metrics and FAIR (Findable, Accessible, Interoperable, Reusable) data practices are critical for maintaining reproducibility in structural predictions [31].

8.9 Future Perspectives and Outlook

The future of protein structure prediction is being redefined by hybrid approaches that integrate artificial intelligence, physics-based simulations, and quantum computing. The trajectory from homology modeling to deep learning exemplifies how data and algorithmic advances can overcome historical limitations in biophysical modeling.

8.9.1 Integration with Molecular Dynamics and Quantum Mechanics

Next-generation workflows increasingly combine AlphaFold-predicted structures with molecular dynamics (MD) simulations for conformational sampling and quantum mechanics/ molecular mechanics (QM/MM) calculations for accurate energetics. This integration captures both static and dynamic aspects of molecular recognition, offering a more complete representation of protein behavior [32].

8.9.2 Multi-State and Allosteric Modeling

Emerging algorithms aim to predict multiple functional states of proteins, including allosteric transitions and post-translationally modified forms. Hybrid methods that incorporate normal mode analysis and Markov state modeling promise to elucidate conformational landscapes beyond single-structure predictions, crucial for designing allosteric modulators and covalent inhibitors [33].

8.9.3 Quantum-Inspired and Cloud-Integrated Prediction Platforms

Quantum-inspired optimization techniques, such as variational quantum eigensolvers (VQE), are being explored to solve protein folding problems using energy landscape minimization. Simultaneously, cloud-based AI frameworks like AlphaFold Cloud, ESMFold, and RoseTTAFold2 enable large-scale proteome modeling and real-time collaboration across distributed research networks [34].

8.9.4 Toward Explainable and Sustainable AI in Structural Biology

The next frontier involves explainable AI (XAI) models capable of elucidating the reasoning behind predicted interactions and folds. Moreover, as large-scale computations raise environmental

concerns, sustainable strategies such as energy-efficient GPU utilization and green data centers are increasingly emphasized in computational drug design [35].

8.9.5 Future of Homology Modeling in the AI Era

While deep learning has surpassed classical modeling in accuracy, homology modeling remains indispensable for interpretability, user control, and comparative evolutionary studies. In many cases, hybrid pipelines combining AI predictions, human expertise, and physical refinement yield the most trustworthy models. As AI evolves toward open and modular platforms, homology modeling will likely persist as a complementary, scientifically transparent foundation for rational drug design.

REFERENCES

1. Marti-Renom MA, et al. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000;29:291–325.
2. Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–589.
3. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science*. 2012;338(6110):1042–1046.
4. Yang J, et al. The I-TASSER Suite: Protein structure and function prediction. *Nat Methods*. 2015;12(1):7–8.
5. Baek M, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021;373(6557):871–876.
6. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004;57(4):702–710.
7. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(W1):W320–W324.
8. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993;234(3):779–815.
9. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J*. 2011;101(10):2525–2534.
10. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures. *Nucleic Acids Res*. 2007;35(Web Server issue):W407–W410.
11. Rohl CA, et al. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004;383:66–93.
12. Evans R, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. 2021.
13. DeepMind. AlphaFold3 Technical Report. *Nature*. 2024 (in press).
14. Lin YR, et al. Comparative assessment of AlphaFold and experimental structures for docking accuracy. *J Chem Inf Model*. 2023;63(12):3534–3545.
15. Webb B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol*. 2017;1654:39–54.
16. Waterhouse A, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296–W303.
17. Roy A, et al. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5(4):725–738.
18. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res*. 2013;41(W1):W384–W388.

19. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction, and analysis. *Nat Protoc*. 2015;10(6):845–858.
20. Kryshtafovych A, et al. CASP14 assessment of machine learning methods for protein structure prediction. *Proteins*. 2021;89(12):1633–1647.
21. Lupas AN, et al. Hybrid modeling strategies combining AI and physics-based refinement. *Front Drug Discov*. 2024;4:125–139.
22. Ballesteros JA, Palczewski K. G protein-coupled receptor structure and activation mechanisms. *Mol Pharmacol*. 2021;100(3):228–245.
23. Rodrigues CHM, et al. DynaMut: predicting the impact of mutations on protein conformation, flexibility, and stability. *Bioinformatics*. 2018;34(19):3399–3406.
24. Lensink MF, et al. Docking and scoring protein complexes: CAPRI challenges and recent advances. *Proteins*. 2022;90(5):1094–1108.
25. Senior AW, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;577(7792):706–710.
26. Lill MA, Danielson ML. Computer-aided drug design platform using structure-based and pharmacophore methods. *J Med Chem*. 2011;54(15):5159–5170.
27. Zhang C, et al. Functional annotation of proteins using structure-based prediction. *Front Bioinform*. 2022;2:860423.
28. Khafizov K, et al. Challenges in modeling membrane proteins: limitations and solutions. *Biochim Biophys Acta Biomembr*. 2019;1861(12):183050.
29. Hollingsworth SA, Dror RO. Molecular dynamics simulation for allosteric mechanism elucidation. *Nat Rev Mol Cell Biol*. 2018;19(5):288–304.
30. Jumper J, et al. AlphaFold3: integrating ligands and nucleic acids. *Nature*. 2024 (forthcoming).
31. Wilkinson MD, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.
32. Warshel A, et al. QM/MM modeling of enzyme catalysis: progress, challenges, and future prospects. *Chem Rev*. 2022;122(13):11605–11647.
33. McCammon JA. Markov state models and free energy landscapes in protein folding. *Annu Rev Phys Chem*. 2021;72:181–203.
34. Cao Y, et al. Quantum-inspired algorithms for molecular modeling. *npj Quantum Inf*. 2023;9:11.
35. Green AI Consortium. Sustainable practices in large-scale AI computation for biosciences. *Front Comput Biosci*. 2024;1:102–117.