Briefings in Bioinformatics, 2025, **26(4)**, bbaf340
https://doi.org/10.1093/bib/bbaf340
Review

Beyond static structures: protein dynamic conformations modeling in the post-AlphaFold era

Xinyue Cui[‡], Lingyu Ge[‡], Xia Chen, Zexin Lv, Suhui Wang, Xiaogen Zhou, Guijun Zhang[*]

College of Information Engineering, Zhejiang University of Technology, 288 Liuhe Road, Hangzhou 310023, Zhejiang, China
[*]Corresponding author. College of Information Engineering, Zhejiang University of Technology, 288 Liuhe Road, Hangzhou 310023, Zhejiang, China.
E-mail: zgj@zjut.edu.cn
[‡]Xinyue Cui and Lingyu Ge contributed equally to this work.

Abstract

The emergence of deep learning, particularly AlphaFold, has revolutionized static protein structure prediction, marking a transformative milestone in structural biology. However, protein function is not solely determined by static three-dimensional structures but is fundamentally governed by dynamic transitions between multiple conformational states. This shift from static to multi-state representations is crucial for understanding the mechanistic basis of protein function and regulation. This review outlines the fundamental concepts of protein dynamic conformations, surveys recent computational advances in modeling these dynamics in the post-AlphaFold era, and highlights key challenges, including data limitations, methodological constraints, and evaluation metrics. We also discuss potential strategies to address these challenges and explore future research directions to deepen our understanding of protein dynamics and their functional implications. This work aims to provide insights and perspectives to facilitate the ongoing development of protein conformation studies in the era of artificial intelligence-driven structural biology.

Keywords: protein structure prediction; dynamic conformations; ensemble; deep learning; molecular dynamics (MD); diffusion model

Introduction

Proteins are the foundation of life processes, with their functions fundamentally dependent on intricate dynamic conformational changes. Many pathological conditions, such as Alzheimer's disease [1], Parkinson's disease [2], and other diseases, stem from protein misfolding [3] or abnormal dynamic conformations. Therefore, systematically elucidating transitions between conformational states is essential for designing conformation-specific drugs and treating diseases.

Since Anfinsen introduced the concept of protein structure prediction in 1972, the field has undergone significant progress, including the development of homology modeling [4, 5], fragment assembly [6, 7], and co-evolution analysis [8–10], which laid a strong foundation for this area. More recently, end-to-end methods, such as AlphaFold series [11, 12] and RoseTTAFold [13], have revolutionized static monomeric protein structure prediction, particularly for single-domain folding, achieving near-perfect accuracy. In 2024, the Nobel Prize in Chemistry was awarded for researching protein static structure prediction, further highlighting the groundbreaking advancements in this field. However, proteins should not be viewed as static entities but as conformational ensembles that mediate various functional states [14]. Although deep learning has made significant progress in protein structure prediction, capturing dynamic conformational changes and sampling conformational space remain challenges in studying protein dynamics [15]. Notably, the 2022 Critical Assessment of Structure Prediction (CASP15) community experiment introduced a dedicated category for predicting multiple conformations for the first time [16], highlighting the growing focus on protein dynamic conformations.

With the increasing focus on modeling protein dynamic conformations, two broad approaches have emerged: experimental and computational methods. Experimental methods such as nuclear magnetic resonance (NMR), cryo-electron microscopy (Cryo-EM), and X-ray crystallography are capable of resolving high-resolution structures. Additionally, methods like fluorescence resonance energy transfer, single-molecule fluorescence microscopy, and hydrogen-deuterium exchange mass spectrometry capture conformational changes and dynamic behaviors of proteins to some extent. However, the practical application of these methods is substantially limited by their dependence on rigorous crystallization conditions and/or the inherent challenges of sparse, ambiguous, and noisy data [17]. Computational methods, such as molecular dynamics (MD) simulations, have provided valuable insights for exploring protein dynamic conformations by directly simulating the physical movements of molecular systems. Moreover, several research [18–22] built on the artificial intelligence (AI) protein structure prediction methods [11–13, 23], such as AlphaFold2 [11], by changing the model input, including multiple sequence alignment (MSA) masking, subsampling, and clustering, to capture different co-evolutionary relationships of proteins and thus generating diverse predicted conformations. Recently, generative models [24–29], leveraging techniques like diffusion and flow matching, have emerged as powerful tools

Received: February 18, 2025. **Revised:** May 29, 2025. **Accepted:** June 20, 2025

for predicting protein multiple conformations. Unlike MSA-based methods, these models transform protein structure prediction into a sequence-to-structure generation through iterative denoising. Some of these methods [27, 28] can effectively predict equilibrium distributions of molecular systems, allowing for the sampling of effectively diverse and functionally relevant structures.

In the post-AlphaFold era, driven by breakthrough advancements in static protein structures, the paradigm of protein research is gradually shifting from static structures to dynamic conformations. In this review, we examine recent advances in computational approaches for modeling protein dynamic conformations, highlighting the importance of integrating experimental and MD simulation data and the critical role of incorporating physical knowledge. Finally, we conclude by critically evaluating current challenges and future directions in this rapidly evolving field.

Protein dynamic conformations

Essential concepts in protein conformational dynamics

Dynamic conformations emphasize a process of protein conformational change over time and space [30], including both subtle fluctuations and significant conformational transitions. Many functional proteins rely on dynamic conformational changes to perform specific biological roles. A representative example can be found in biological systems where enzymes dynamically modulate their conformational states to facilitate catalytic processes, while membrane proteins utilize specific conformational transitions to mediate signal transduction and regulate molecular transport across cellular membranes [31]. Consequently, elucidating the dynamic conformational landscapes of proteins is essential for deciphering their biological functions and underlying regulatory mechanisms. As illustrated in Fig. 1, assuming that the energy function accurately describes the conformational free energy surface of the protein, the dynamic conformations B of a protein usually involves multiple key conformational states (i.e. A), including stable state (a), metastable states (b, c), and transition states (d) between them. It is important to note that the definition of all conformational states depends on the measurement system. Under varying energy landscapes, metastable states can transition into stable states. In the field of protein science, the concepts of dynamic conformations and ensemble exhibit significant overlap in their theoretical definitions and application scopes and are therefore often used interchangeably in most research contexts. In this review, the protein conformation ensemble (as shown in Fig. 1C) represents the collection of independent conformations of proteins in various motion states under certain conditions [32]. It reflects the structural diversity of the protein under thermodynamic equilibrium, capturing the distribution and probabilities of the protein's conformations under given conditions [24]. To some extent, it can also be regarded as the dynamic conformations under specific conditions.

Factors affecting protein dynamic conformations

Dynamic conformations can arise in a variety of situations and can be broadly divided into two categories. The first category is driven by the intrinsic factors of the protein. For example, the presence of disordered regions, which lack α-helices or β-sheets, results in higher flexibility; relative rotations or adjustments between structural domains also facilitate transitions between different conformations. Additionally, proteins such as G Protein-Coupled Receptors (GPCRs), transporters, kinases, and others undergo conformational changes to perform their biological functions (illustrated in Fig. 2A–C). The second category encompasses alternative conformations influenced by external environmental conditions. On the one hand, different conformational states can be triggered by the binding of small ligands or by interactions with other macromolecules [16]. On the other hand, changes in environmental factors such as temperature, pH, and ion concentration can directly impact the stability and conformation of protein. In some cases, proteins may unfold or alter their conformation to adapt to these environmental changes. Additionally, mutations in the primary amino acid sequence may also induce conformational shifts. Whether the changes are caused by the structural properties of the protein itself or induced by different external conditions, proteins can fulfill their diverse functions through conformational transitions. This conformational flexibility is the basis for proteins to perform complex biological activities [37, 38] (Fig. 2D–F).

It is noteworthy that protein dynamic conformations are modulated by both intrinsic properties and external factors. Emerging evidence indicates that dynamic information facilitating conformational transitions may be inherently encoded within the protein sequence itself. A compelling illustration of this phenomenon can be observed in the CASP15 test cases (T1160 and T1161) [16], where multiple distinct conformations were accurately predicted using an AlphaFold-based enhanced sampling approach, independent of external environmental perturbations. These findings strongly suggest that the observed conformational heterogeneity originates from sequence-encoded information, potentially embedded within either the target sequence or the MSA.

Advancements in protein dynamic conformations modeling

Databases for dynamic conformations

Datasets for dynamic conformations are the foundation of research on protein dynamic conformations and serve as the prerequisite for the application of deep learning in modeling within this field. High-quality datasets are crucial for understanding and predicting the dynamic behavior of proteins. Advancements in simulation technologies like GROMACS [39], AMBER [40], OpenMM [41], and CHARMM [42] have significantly enhanced the analysis of MD simulation data. These technological advances have played a pivotal role in facilitating the creation of comprehensive databases documenting protein dynamic conformations. Consequently, several specialized MD-generated databases have been established, including the Atlas of Protein Molecular Dynamics (ATLAS), G Protein-Coupled Receptor molecular dynamics database (GPCRmd), and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) proteins database, among others (as detailed in Table 1). The ATLAS database comprises simulations of approximately 2000 representative proteins, covering a vast portion of structural space. GPCRmd focuses on transmembrane proteins of the GPCR family to better understand their mechanisms and identify potential drug targets. Additionally, the SARS-CoV-2 database includes simulation trajectories of coronaviruses, which supports the drug discovery process for COVID-19. All raw data in these databases can be accessed through the corresponding links listed in Table 1. Some datasets are sourced from existing static databases such as CoDNaS 2.0, a comprehensive database of protein diversity in its native state, and PDBFlex, which offers insights into protein

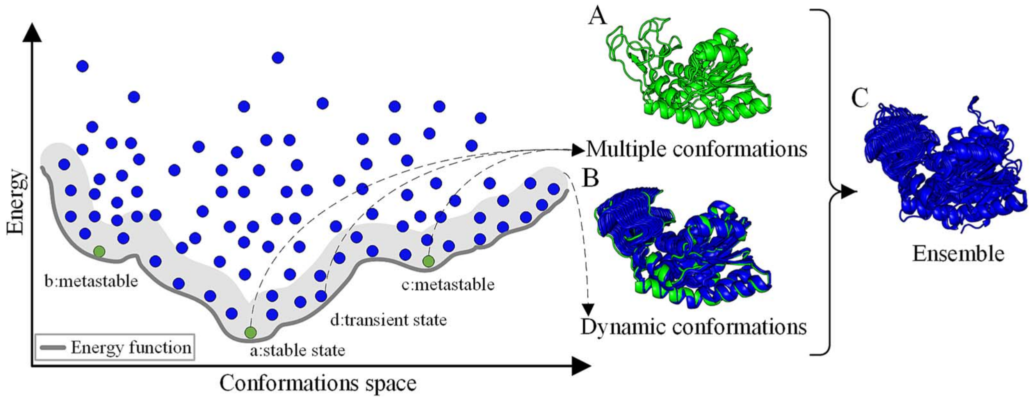

Figure 1. Assuming that the energy function is accurate, the energy relationship corresponding to different states of the protein (stable, metastable, and transient states) can be quantified. The multiple conformations A represent the key conformations adopted by the protein during functional execution, while the dynamic conformations B include the transition states between these key conformations. The ensemble C represents the collection of all possible conformations under given conditions.

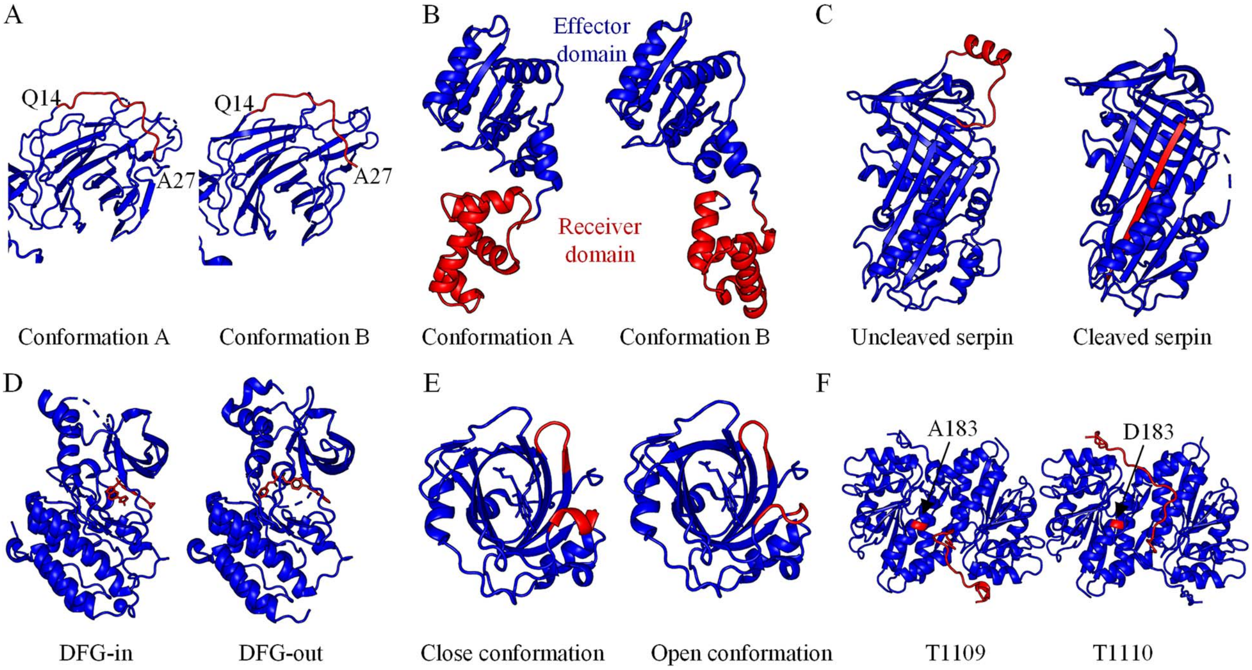

Figure 2. (A) Conformational diversity of the loop in the amino acid 14–27 region of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein [33] (PDB IDs: 6zgeA, 7dddC). (B) The structural composition of *Streptococcus pneumoniae* response regulator spr1814 (PDB ID: 4hyeA). The disappearance of the salt bridge in the receiver domain induces a 74° rotation, transitioning from conformation A to conformation B [34]. (C) The serine protease inhibitor (Serpin) where a loop inserts as a strand in the middle of a β-sheet following proteolytic cleavage [35]. (D) c-Met kinase exhibits active Asp-Phe-Gly (DFG)-in and inactive DFG-out states in response to different small molecule ligands [32]. (E) Nitrophorin 4 (NP4) is in a closed conformation at pH 5.5 (PDB ID: 1x8o) in which NO is tightly bound. At pH 7.5, Asp30 becomes deprotonated, leading to a transition to an open conformation (PDB ID: 1x8n), allowing NO to escape easily [36]. (F) A conformational change due to the mutation of the amino acid at position 183 from alanine to aspartic acid in isocyanide hydratase [16].

structural flexibility. These datasets were obtained by collating and analyzing clusters from the protein data bank (PDB, https://www.rcsb.org/) databases.

In the study of protein dynamic conformations, the extraction of dynamic information from static structural data has become an essential complementary approach, which has alleviated the scarcity of dynamic data to a certain extent. Specifically, by analyzing multiple crystal structures or NMR structures of the same protein from static structural databases (e.g. PDB), researchers can construct conformational ensembles, thereby inferring potential conformational transition pathways [3, 54]. Furthermore, computational methods such as Normal Mode Analysis [55] and Elastic Network Models [56], which are based on static structures, enable the prediction of low-frequency motion modes of proteins in the absence of experimental dynamic data. The dynamic information derived from static data not only provides initial conformations and validation benchmarks for MD simulations but also aids in interpreting experimentally observed dynamic phenomena, particularly in cases where obtaining high-quality dynamic data is challenging (e.g. membrane proteins or large macromolecular complexes). Therefore, the integration of in-depth mining of static structural data with dynamic data offers a more comprehensive framework for understanding protein conformational dynamics.

Table 1. Existing public protein dynamic conformation datasets (update to 27 May 2025)

Name	Data	Number/ trajectories	Time scale	Types	Applications	Database Link	Reference
ATLAS (2023)	MD data	1938/5841	ns	General proteins	Protein dynamics analysis	https://www.dsimb.inserm.fr/ATLAS	[43]
GPCRmd (2020)	MD data	705/2115	ns	GPCR	GPCR functionality and drug discovery	https://www.gpcrmd.org/	[44]
SARS-COV-2 (2024)	MD data	78/300	ns/μs	SARS-CoV-2 proteins	SARS-CoV-2 drug discovery	https://epimedlab.org/trajectories	[45]
COVID-19 (2024)	MD data	318/>10 000	ns/μs	Coronavirus proteins	Dynamics of coronavirus proteins	https://covid.molssi.org/simulations/	[46]
SCoV2-MD (2019)	MD data	252/252	ns/μs	SARS-CoV-2 proteins	Dynamics of SARS-CoV-2 proteins	https://submission.gpcrmd.org/covid19/	[47]
MemProtMD (2015)	MD data	8459/8459	μs	Membrane proteins	Membrane protein folding and stability	https://memprotmd.bioch.ox.ac.uk/	[48]
mdCATH (2024)	MD data	5397/134 925	ns	CATH domains	Protein function, folding, and interactions	https://open.playmolecule.org/mdcath	[49]
MISATO (2023)	MD data	16 972/16 972	ns	General proteins	Structure-based drug discovery	https://zenodo.org/records/7711953	[50]
CoDNaS 2.0 (2016)	PDB	29 148/–	–	General proteins	Conformational diversity analysis	http://ufq.unq.edu.ar/codnas	[51]
CoDNaS-Q (2022)	PDB	3649/–	–	Quaternary structure proteins	Conformational heterogeneity in quaternary structure	https://codnas-q.bioinformatica.org/home	[52]
PDBFlex (2015)	PDB	38 341/–	–	General proteins	Analysis of the intrinsic flexibility of proteins	https://pdbflex.org/	[53]

Methods for predicting protein dynamic conformations

Owing to the time-consuming and resource-intensive nature of experimental approaches in studying protein dynamic conformations, recent breakthroughs in static structure prediction technologies, such as AlphaFold2, have established computational methods for protein dynamic conformations analysis as a forefront research priority in the field of computational biology. These computational approaches address the limitations of experimental data, providing critical insights into the dynamic behaviors and functional mechanisms of proteins. By integrating advanced computational techniques, researchers can systematically explore the conformational space, functional motion patterns, and dynamic interactions of proteins with other biomolecules. The rapid advancements in this field are driving a paradigm shift in structural biology from static structure analysis to dynamic functional studies, thereby opening new research avenues for understanding the molecular basis of life processes. These computational methodologies comprise a suite of advanced techniques, including MD simulations, Monte Carlo (MC) sampling, AlphaFold-based frameworks, and diffusion models, each characterized by unique strengths and tailored applications. MD simulations can provide high temporal resolution of protein dynamic changes by accurately characterizing atomic interactions. However, this method is computationally expensive and highly dependent on the accuracy of the force field. AlphaFold-based methods mainly rely on the co-evolutionary information contained in the sub-MSAs of the input model, so they are limited in capturing protein dynamics. In contrast, generative models can generate diverse conformations but still face limitations in terms of physical accuracy and reliable energy evaluation. This review will focus on recent advancements in these methods, while progress in other related approaches can be found in reference [57].

MD simulation and MC sampling

MD simulation is a computational method used to simulate the motion of atoms and molecules over time under given force fields and initial conditions, and MC sampling is a class of techniques for random sampling of probability distributions. With the advent of powerful computational capabilities, MD simulations can now be performed on time scales ranging from microseconds to milliseconds, allowing the description of the protein's dynamic properties. Currently, various force fields [39–42] have been developed for MD simulations [58]. By applying these force fields, MD simulations can effectively explore the structure and dynamic features of proteins. Additionally, integrating data from MD simulations with information obtained from experimental structures offers a comprehensive view of large-scale protein dynamics and provides valuable insights into representative proteins. It is worth mentioning that MD simulations are usually limited by the accuracy of the force field, computational complexity, high-dimensional sampling, and the challenges of the tendency to fall into local minima (as listed in Fig. 3). To address these challenges, researchers have proposed a variety of solutions. For example, generalized ensemble algorithms [59] and temperature acceleration methods [60] can effectively improve sampling efficiency and help the system overcome high energy barriers on the potential energy surface. In addition, multiscale modeling methods [61] further expand the scope of MD simulations by integrating information at the atomic, amino acid, and secondary structure levels. However, such methods often sacrifice some simulation accuracy while improving computational efficiency. Recently, the emergence of AI^{2}BMD [62], an AI-based ab initio computational molecular system, has marked a breakthrough in protein dynamics simulation. This method not only accurately simulates full-atom large biomolecules with ab initio accuracy [62], but also effectively explores the protein conformational space and reveals the protein folding and unfolding process. Although AI^{2}BMD is faster than

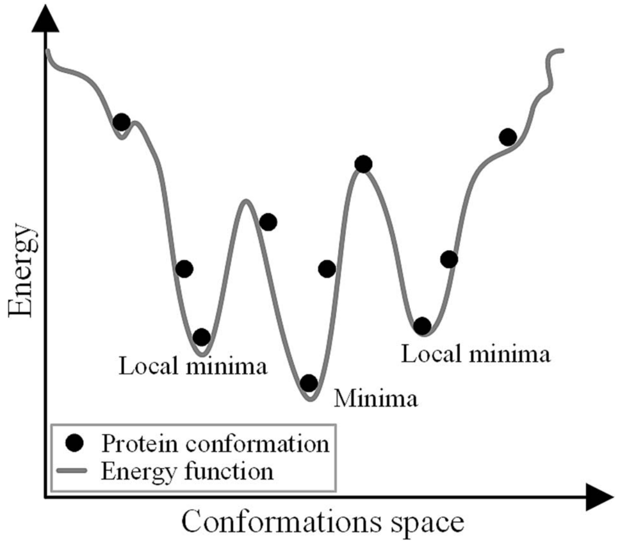

Figure 3. MD simulation trapped in local energy minima diagram. MD simulations are typically limited by factors such as force field accuracy, computational complexity, high-dimensional sampling, and simulation time, making them prone to getting trapped in local minima (i.e. energy values corresponding to conformations that are local minima or in a metastable state).

density functional theory, its computational efficiency is still lower than that of classical force field methods, and it requires higher memory resources. Furthermore, beyond accelerating MD simulation force fields with deep learning, incorporating evolutionary relationships is essential to enhance the exploration of conformational space [63]. In the future, integrating deep learning and evolutionary relationships is expected to become an important direction for developing MD simulations, providing more precise and efficient tools for studying protein dynamic conformations and drug discovery.

Different from MD simulations, MC sampling does not depend on the time evolution process; it is usually combined with Metropolis Markov chain algorithms to explore the conformational space of proteins by accepting or rejecting random moves. MC sampling methods can be used to study the conformational landscape of small proteins [64], including free energy landscapes, transition states, energy barriers, and the stability of individual amino acids at different temperatures, offering new perspectives and methodologies for protein dynamic conformational studies. Based on MC sampling, our group has developed a Metropolis MC sampling method, FoldPAthreader [54], guided by the folding force field. This method leverages the intrinsic relationship between protein evolution and folding. Additionally, we have proposed Pathfinder [3], a method that predicts protein folding pathways by estimating the transition probabilities between metastable states of sampled conformations. These methods offer valuable insights into the mechanisms of protein folding and highlight the complexity of their dynamic conformations. Additionally, evolutionary algorithms based on multiple populations can sample multiple conformational states under the guidance of multiple energy functions [65]. Moreover, conformational sampling using an iterative exploration and exploitation strategy based on multi-objective optimization, geometric optimization, and structural similarity clustering [66, 67] has also shown significant potential in dynamic conformational modeling. Despite the advancements made by MC methods in dynamic conformational exploration, challenges like low sampling efficiency and difficulties in overcoming energy barriers still exist. To overcome these issues, researchers often combine MD, MC, and deep learning techniques to enhance efficiency and prediction accuracy in sampling [68].

AlphaFold-based method

With the advancement of computational biology, the structural accuracy predicted by machine learning-based methods (such as the AlphaFold series [11, 12] and RoseTTAFold [13]) is generally comparable to that of experimental structures. However, directly using static structures to understand the conformations of biological systems is not straightforward [69]. Recent studies have demonstrated that changing the inputs of AlphaFold2 [11], such as clustering, masking, subsampling MSAs, or adjusting templates, can effectively predict multiple conformations of proteins. MSA clustering methods, such as AF-Cluster [18], group MSAs based on sequence similarity to extract diverse co-evolutionary information, which is then fed into AlphaFold2 for separate predictions of each cluster (Fig. 4), generating multiple conformational models of the target protein. Through clustering, AF-Cluster [18] identifies representative MSAs with distinct sequence features. Tests on fold-switching proteins have demonstrated that this approach effectively captures multiple key substates of proteins involved in biological functions. Moreover, some works have been studied through shallow MSA. Methods like Subsampled AF2 [22] and af2_conformations [21] reduce the depth of input MSAs through a subsampling strategy and then combine AlphaFold2 to generate multiple conformations (as listed in Table 2). To disrupt the co-evolutionary information in MSAs, methods like masking and alanine mutation have proven effective. SPEACH_AF [70] replaces MSA columns with alanine to study the conformational diversity of the target protein. AFsample2 [20] randomly masks columns in the MSA according to a certain ratio to enhance the structural diversity of the models generated by AlphaFold2. AF2-RASS [71] expands the conformational diversity of the predicted structural ensemble by employing random alanine masking and shallow MSA strategies, enabling the study of protein apo-holo conformational states.

These methods extend existing protein structure prediction techniques to facilitate the exploration of protein dynamic conformations. However, the effectiveness of such post-processing methods largely depends on the quality, depth, and co-evolutionary information contained within the sub-MSAs. For proteins with shallow MSAs or limited sequence diversity, the accuracy and reliability of predicted conformations may be compromised. Furthermore, whether the conformations generated by these models are influenced by memory effects from AlphaFold's training process remains to be further investigated. To address the above limitations, integrating protein structure profile information and leveraging the structures generated by automatic template recognition [72] or MD simulations [69] as prior knowledge is expected to enhance the prediction accuracy for proteins with shallow MSA or limited sequence diversity, and further validate the rationality of the model output conformations.

Enhanced sampling algorithms based on AlphaFold2 are currently the most successful and effective methods for studying protein dynamic conformations [16]. By employing strategies to disrupt the coupled co-evolutionary information in MSAs effectively, these methods enable the generation of diverse conformations. They have demonstrated high accuracy in modeling conformational changes induced by single amino acid mutations, transporters, and kinases [16]. However, it should be noted that the optimal depth of the MSA varies depending on the protein,

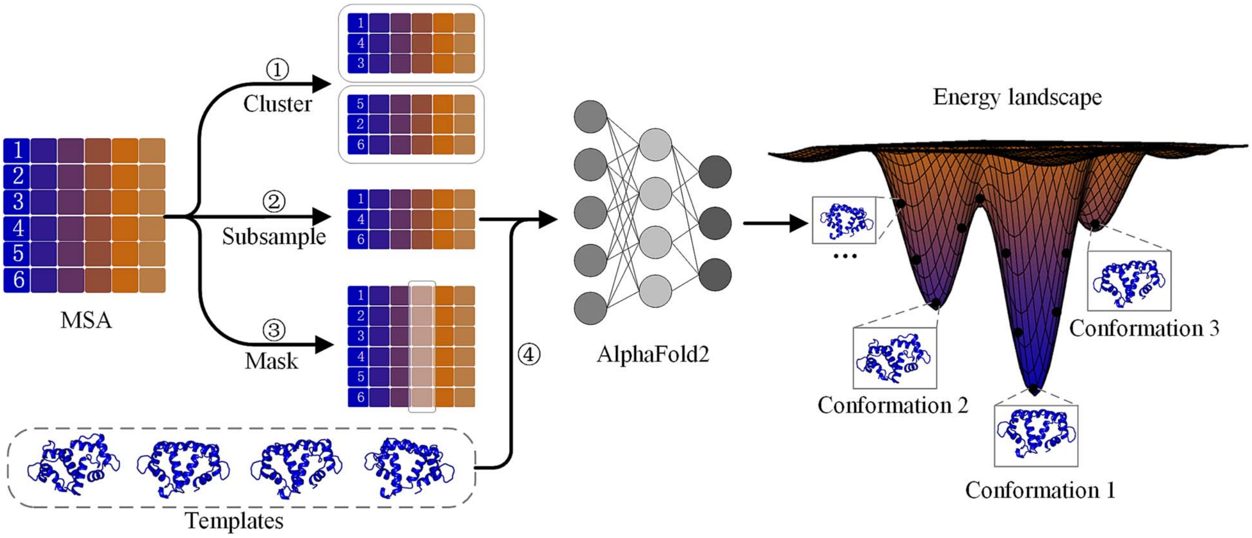

Figure 4. Research categories on protein dynamic conformations based on the AlphaFold2 method. ①, ②, ③, ④ represent the processing workflows of MSA and templates, respectively.

Table 2. AlphaFold-based methods for predicting dynamic conformations.

Name	Strategy	Metrics	Application	Reference
AF-Cluster (2024)	Clustered MSA	RMSD, plDDT, PCA	Predicting alternative conformations of metamorphic proteins	[18]
Subsampled AF2 (2024)	Shallow MSA	RMSD, plDDT	Qualitative prediction of the conformational landscape shaped by mutations or evolution	[22]
SPEACH_AF (2022)	Mutated MSA	TM-score, PCA, RMSF	Sampling of alternative conformations and modeling of membrane protein conformational landscapes	[70]
AFsample2 (2025)	Masked and clustered MSA	TM-score, RMSF	Modeling conformational changes in open-closed states, membrane proteins, and transport proteins	[20]
AF2-RASS (2024)	Shallow and masked MSA	TM-score, RMSD, pLDDT	Predicting conformational diversity of structural ensembles and capturing conformational changes between apo and holo protein forms	[71]
af2_conformations (2022)	Shallow MSA	TM-score, RMSF, plDDT PCA	Modeling alternative conformations of transporters and GPCR	[21]

plDDT, Predicted local Distance Difference Test, a confidence metric for predicted protein structures; PCA, Principal Component Analysis, a dimensionality reduction technique used for analyzing conformational dynamics.

and shallow MSAs may result in complex structures that do not align with the target conformation. Moreover, the generation of the final conformation heavily relies on the dominant co-evolutionary information present in the MSAs. When sub-MSAs include co-evolutionary information from diverse conformations, it may affect the predicted conformation's accuracy and reasonableness (Fig. 5). Whether by changing the MSA or using the template method, these processes often involve multiple iterative cycles and large-scale sampling, which directly affects the prediction efficiency. Optimized computational strategies can be applied to improve efficiency, such as combining enhanced sampling techniques and algorithmic architecture optimization [73]. Moreover, with the ongoing advancements in protein language models [23, 74–76], the exploration of protein dynamic conformations directly from single sequences, independent of MSAs and template-based approaches, may emerge as a promising research direction in the future. It is worth noting that a representative example is ESMFlow [24], which integrates the ESM language model with flow matching techniques, enabling the learning of mappings between protein sequences and their dynamic conformational distributions without relying on MSAs.

Generative models

In recent years, generative models have emerged as a pivotal methodology for investigating dynamic conformational transitions in proteins, a development primarily attributed to the rapid evolution of artificial intelligence technologies and substantial enhancements in computational capabilities. These advanced models demonstrate the capacity to extract intricate conformational distribution patterns from extensive protein structural datasets, thereby enabling the generation of novel and biophysically plausible protein conformational states. Notably, such models exhibit dual functionality, as they not only facilitate accurate prediction of static protein architectures but also simulate dynamic conformational changes under diverse environmental conditions, including but not limited to protein folding pathways, conformational switching mechanisms, and structural adaptations during ligand–receptor interactions. These methods aim to learn patterns from training datasets, identify low-dimensional representations within the high-dimensional space of proteins, and directly generate diverse sets of protein conformations without relying on large-scale sampling operations. However, the efficacy of generative models is predominantly contingent upon the quality of training datasets,

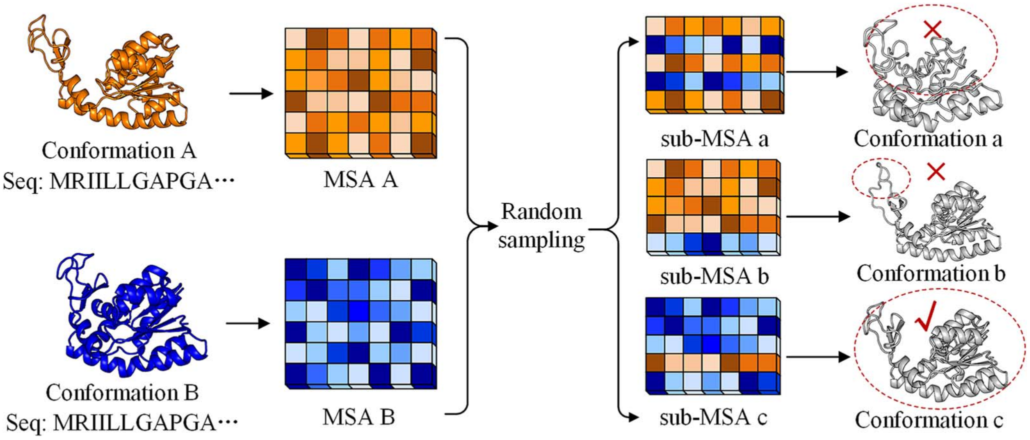

Figure 5. Protein conformations are predicted using random sampling to generate different sub-MSAs. While random sampling can predict some potentially correct conformations, its accuracy is limited by the co-evolutionary relationships contained in the sub-MSAs. If multiple co-evolutionary conformations corresponding to different conformations dominate in a given sub-MSA, the predicted conformation may deviate from all target conformations (as shown in conformations a and b in the figure).

among which MD simulation data constitutes a fundamental source and plays an indispensable role in model development and performance optimization. By combining MD data, the generative models based on traditional transformer [15] or GAN [77] networks can efficiently generate physically reasonable protein conformations, significantly improving prediction speed and efficiency. Leveraging these recent technological advancements, diffusion models have emerged as a novel and powerful computational framework for protein dynamic conformations prediction, demonstrating superior performance in modeling complex biomolecular transitions and energy landscapes.

The basic principle of diffusion models is to perturb the data distribution to a simple Gaussian distribution through a forward stochastic differential equation, and then generate the target conformation through a reverse process. Compared with the previous diffusion-based structure generation method of directly adding Gaussian noise [78], EigenFold [26] proposes a diffusion process based on elastic potential energy, which makes the denoising process more aligned with physical laws (Table 3). However, its effectiveness still depends on the accuracy of the energy function. To further improve the physical rationality of the generated conformations, ConfDiff [25] and ExEndiff [79] use physical information guidance and experimental data guidance, respectively, which significantly improves the generation speed and ensures the generated conformation distribution is closer to the Boltzmann distribution. In addition, IDPFold [80] generates intrinsically disordered protein conformations that are more aligned with physical principles through a multi-step noise addition and removal process. To address the lack of high-quality training and test datasets for equilibrium sampling of proteins, and the almost complete absence of relevant benchmark datasets, methods like BioEmu [28] and DiG [27] combine the strengths of different datasets to provide an efficient and low-cost approach for large-scale studies of protein equilibrium distributions and conformational changes. Compared with MD simulations, BioEmu significantly reduces computational costs, making it a promising alternative for exploring protein dynamics on a broader scale. However, its training relies on high-quality MD simulation data and experimental measurement data. For proteins with limited data availability, the model's performance may be constrained. Based on diffusion models. AlphaFLOW [24] introduces a flow-matching technique that fine-tunes high-precision monomer predictors such as AlphaFold [11] and ESMFold [23] within a customized flow-matching framework. This enables protein structure generation conditioned on sequences, significantly enhancing the accuracy and efficiency of conformation generation. Another flow-based generative model, P2DFlow [29], integrates invariant point attention modules and SE(3) equivariant graph neural network modules. By using perturbed ESMFold-predicted structures as priors, P2DFlow opens up new research directions for protein dynamic conformations prediction.

Although the above generative models represent a significant advancement in protein dynamic conformations modeling, they also have limitations. On the one hand, the performance of generative models heavily depends on the conformational distribution in the training data. However, existing static protein databases, such as the AlphaFold database (AFDB, https://alphafold.ebi.ac.uk/) and the PDB (https://www.rcsb.org/), contain relatively limited dynamic information, making it difficult to fully capture the dynamic behavior of proteins in living organisms. On the other hand, generative models may predict hallucinatory proteins, which often have plausible-looking structures [12], but fail to perform their intended functions under experimental conditions. To overcome this limitation, a comprehensive integration strategy that integrates static structural databases, MD simulation data, and experimental observations, while systematically incorporating fundamental physical constraints and knowledge distillation into generative models, may provide a robust solution to current challenges.

Challenges in protein dynamic conformations

The present works highlight the considerable potential of AlphaFold-based approaches and generative models in predicting protein dynamic conformations. Nevertheless, several significant challenges persist in this research field, particularly concerning (i) the limited availability of high-quality experimental and computational data, (ii) the need for improved prediction accuracy in modeling complex conformational transitions, and (iii) the absence of standardized evaluation protocols. These limitations are particularly pronounced in simulating

Table 3. Generative model-based methods for predicting dynamic conformations.

Name	Network architecture	Training set	Reference
AlphaPPImd (2024)	Transformer	MD simulation	[15]
idpGAN (2023)	GAN	MD simulation	[77]
EigenFold (2023)	Diffusion model	PDB	[26]
ConfDiff (2024)	Diffusion model	PDB	[25]
ExEnDiff (2024)	Pretrained model and Diffusion model	NMR, cryo-EM, SAXS	[79]
IDPFold (2024)	Diffusion model	PDB, NMR, MD simulation	[80]
BioEmu (2024)	Diffusion model	AFDB, MD simulation	[28]
DiG (2024)	Diffusion model	PDB	[27]
AlphaFLOW (2024)	Pretrained model and flow matching	PDB, MD simulation	[24]
P2DFlow (2024)	Pretrained model and flow matching	MD simulation	[29]

physiologically relevant conformational dynamics and developing robust multi-scale modeling frameworks that can bridge different temporal and spatial resolutions. Future advancements in protein dynamic conformations research are anticipated to emerge from several key developments: the synergistic integration of experimental and computational simulation data to mitigate experimental noise and enhance data reliability, the development of physics-aware machine learning frameworks that incorporate fundamental biophysical principles, substantial improvements in force field parameterization and energy function accuracy, and the establishment of robust, standardized evaluation metrics. These combined efforts are expected to drive significant breakthroughs in understanding and predicting protein conformational dynamics at multiple temporal and spatial scales.

Establishing a high-quality dataset for protein dynamic conformational transitions

Currently, the methods for studying the dynamic conformations of proteins mainly rely on static databases and MD simulation data. Static databases like PDB and AFDB mainly record the single stable conformation of proteins under specific conditions, typically representing their lowest energy states, which provides valuable information for understanding the basic structure of proteins. However, these data cannot fully demonstrate the dynamic behavior and conformational change process of proteins in living organisms. MD simulations can observe dynamic phenomena such as local flexibility, conformational transitions, and changes in the binding sites of proteins. However, existing MD simulation datasets (such as ATLAS and GPCRmd) usually only simulate movements at the millisecond timescale and may not fully capture all relevant protein conformations. Nevertheless, MD simulations require significant computational resources and time, and the choice of simulation parameters and the accuracy of the force field can directly affect the reliability of the simulation results. Notwithstanding the significant progress achieved by the state-of-the-art methodology AI^{2}BMD [62], which has demonstrated unprecedented capabilities in computational efficiency and prediction accuracy, the scientific community still faces substantial limitations in accessing comprehensive, large-scale MD datasets that are publicly available and sufficiently annotated for diverse research applications. To compensate for the lack of dynamic data, transfer learning techniques can be used to effectively integrate static data (such as PDB and AFDB) with dynamic data (such as MD simulation data), thereby providing more information about protein dynamic conformations. Cryo-EM data could also be incorporated, but the consistency between different data sources still requires further exploration. In the future, with the advancement of computational power and simulation technologies, obtaining large-scale MD simulation datasets on longer time scales and building a protein dynamic conformations dataset that includes a collection of known representative conformations will be crucial.

Physics-guided deep learning model

The protein dynamic conformations modeling methods based on AlphaFold primarily depend on changing the model inputs, with the most common being modifications to the MSA. These methods utilize different MSAs to get conformational diversity while also being constrained by dominant co-evolutionary information in the MSA. In addition, the quality of protein conformations generated by the generative model is largely affected by the conformational distribution in the training data. If the conformational distribution covered by the training data contains only a few conformational states or even non-functional states, the model may learn a biased distribution, which will further affect the quality of the final generated protein dynamic conformations. Existing diffusion-based generative models often fail to fully consider the physical priors of proteins, leading to generated conformations that deviate from the target states or generate hallucinatory proteins. For example, as shown in Fig. 6, the apo (PDB ID: 4ake) and holo (PDB ID: 2eck) conformations of *Escherichia coli* adenylate kinase during functional activation have an RMSD of 6.95 Å. The dynamic conformations generated by the diffusion model (Fig. 6A) show an opposite trend in the distribution of structural deviations from the conformations generated by the theoretically correct method (Fig. 6B), that is, the conformations generated by the diffusion model deviate from both the apo and holo states. This may be due to the biased distribution learned by the diffusion model during training. To address the above issues, a potential strategy is to integrate protein language models with physical force fields to decode the dynamic information embedded in amino acid sequences. This approach would provide richer prior knowledge for conditional diffusion models, thereby better balancing rationality and accuracy in dynamic conformations generation.

Construction of a comprehensive evaluation system

We lack standardized benchmarks and metrics for evaluating dynamic conformations, which remains a significant challenge in current research. Existing evaluation metrics, such as root mean square deviation (RMSD) and template modeling score (TM-score) [81], primarily focus on static structural similarity and do not fully capture the dynamic changes among conformations or their

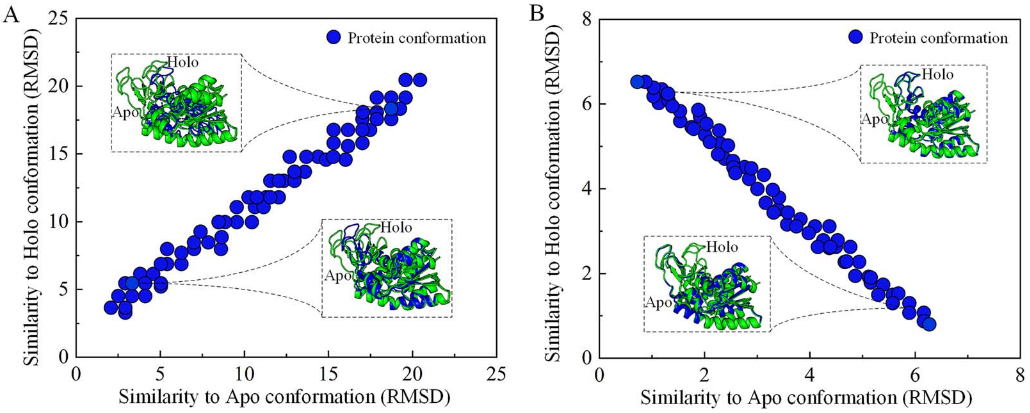

Figure 6. (A) The crystal structure of the *E. coli* adenylate kinase was generated using the diffusion model, showing the structural relationship between apo and holo states and the predicted intermediate. The horizontal and vertical axes represent the RMSD (Å) between the generated conformation and the target conformation. (B) Conformational distribution predicted by the correct method. The generated conformations are located within a reasonable RMSD range between the apo and holo states, indicating that accurate intermediate conformations are being sampled.

time autocorrelation. To better assess dynamic conformations, researchers have introduced new metrics like root mean square fluctuation (RMSF) [24], distance RMSD (dRMSD) [7], and root mean Wasserstein distance (RMWD) [24]. RMSF quantifies the fluctuation of each atom relative to its average position, thereby representing time-averaged structural changes. dRMSD calculates the root mean square of the differences in distances between all pairs of atoms across two conformational sets, but their reliability needs to be built based on the real conformational distribution. RMWD measures the similarity between overall protein conformational distributions, providing a more comprehensive perspective for dynamic structural comparisons. Some methods evaluate model performance by comparing generated conformations with the conformational distribution obtained from MD simulations. However, MD simulation data are usually limited to a short timescale, whereas the conformational motions associated with protein functionality often occur on the millisecond scale or even longer, potentially extending to seconds. Thus, while MD simulations provide valuable insights, using them as an accurate standard for true protein conformational motion requires further validation. Beyond evaluation metrics, accurate model selection is crucial for studying dynamic conformations. However, existing model quality assessment (MQA) methods [82–90] primarily focus on selecting the lowest-energy static structures, and their ability to select dynamic conformations needs further improvement. Therefore, future research should not only focus on establishing a comprehensive set of evaluation criteria that can accurately measure both the dynamics and the accuracy of protein dynamic conformations but also on developing high-precision MQA methods to select key conformations from diverse conformations.

Conclusion

Recent advancements in machine learning methods have broken through the bottleneck in protein structure research, allowing the prediction accuracy of static monomeric structures to reach levels comparable to experimental results. However, in the post-AlphaFold era, focusing only on static single structures can no longer meet the need for a deeper understanding of protein functions. Therefore, how to shift from static single-structure prediction to the revelation of dynamic conformations distribution has become a key step in deeply understanding the functions of proteins and other biological molecules. Currently, research on protein dynamic conformations has made significant progress in areas such as MD simulations, MC sampling methods, AlphaFold-based techniques, and generative models. However, the field still faces many challenges, particularly in data acquisition, method optimization, evaluation metric development, and multi-scale modeling, which are crucial for improving prediction accuracy.

Looking to the future, achieving precise prediction and in-depth understanding of the dynamic behavior of proteins requires coordinated advancement from multiple dimensions. First, it is fundamental to construct more comprehensive and high-quality dynamic datasets, which not only help to portray the conformational changes of proteins under different physiological conditions but also provide rich references for model training. Second, it is crucial to deeply mine the evolutionary information of the existing protein data. Advanced deep learning techniques, particularly language models, can be employed to extract protein sequence diversity and functional characteristics from large-scale datasets. Furthermore, it is important to further optimize the existing force field model to enhance its ability to portray the free energy surface of protein conformation, and at the same time, to establish a more reasonable and unified evaluation system to promote the scientific and reliable prediction methods.

Key Points

- Protein dynamic conformations are crucial to revealing the mechanism of life function.
- With the breakthrough of deep learning in resolving the static structure of proteins and the improved availability of experimental data, it has become possible to study proteins' dynamic conformations.
- Data fusion strategies can provide more information about protein dynamic conformations.
- Language models are expected to reveal protein diversity and functional characteristics from existing data.
- Diffusion models guided by physical prior knowledge are the focus of future research.
- Challenges remain in the availability of data, the effectiveness of prediction methods, and the standardization of evaluation metrics.

Conflict of interest

None declared.

Funding

This work was supported by the National Key R&D Program of China (2022ZD0115103), the National Nature Science Foundation of China (62173304, 62203389), the ‘Pioneer’ and ‘Leading Goose’ R&D Program of Zhejiang (2025C01190), the Zhejiang Province High-level Talent Special Support Program (2023R5248), and The Fundamental Research Funds for the Provincial Universities of Zhejiang (RF-C2024006).

Data availability

No new data were generated or analyzed in support of this research.

References

1. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer’s disease at 25 years. *EMBO Mol Med* 2016;**8**:595–608. https://doi.org/10.15252/emmm.201606210
2. Kalia LV, Kalia SK, Lang AE. Disease-modifying strategies for Parkinson’s disease. *Mov Disord* 2015;**30**:1442–50. https://doi.org/10.1002/mds.26354
3. Huang Z, Cui X, Xia Y. *et al.* Pathfinder: protein folding pathway prediction based on conformational sampling. *PLoS Comput Biol* 2023;**19**:e1011438. https://doi.org/10.1371/journal.pcbi.1011438
4. Zhou X, Zheng W, Li Y. *et al.* I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat Protoc* 2022;**17**:2326–53. https://doi.org/10.1038/s41596-022-00728-0
5. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010;**5**:725–38. https://doi.org/10.1038/nprot.2010.5
6. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 2004;**32**:W526–31. https://doi.org/10.1093/nar/gkh468
7. Simons KT, Bonneau R, Ruczinski I. *et al.* Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins Struct Funct Bioinf* 1999;**37**:171–6. https://doi.org/10.1002/(SICI)1097-0134(1999)37:3+<171::AID-PROT21>3.0.CO;2-Z
8. Wang S, Sun S, Li Z. *et al.* Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol* 2017;**13**:e1005324. https://doi.org/10.1371/journal.pcbi.1005324
9. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-and structure-rich era. *Proc Natl Acad Sci* 2013;**110**:15674–9. https://doi.org/10.1073/pnas.1314045110
10. Tress ML, Valencia A. Predicted residue–residue contacts can help the scoring of 3D models. *Proteins Struct Funct Bioinf* 2010;**78**:1980–91. https://doi.org/10.1002/prot.22714
11. Jumper J, Evans R, Pritzel A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
12. Abramson J, Adler J, Dunger J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493–500. https://doi.org/10.1038/s41586-024-07487-w
13. Baek M, DiMaio F, Anishchenko I. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;**373**:871–6. https://doi.org/10.1126/science.abj8754
14. Degiacomi MT. Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure* 2019;**27**:1034–40. https://doi.org/10.1016/j.str.2019.03.018
15. Wang J, Wang X, Chu Y. *et al.* Exploring the conformational ensembles of protein–protein complex with transformer-based generative model. *J Chem Theory Comput* 2024;**20**:4–80. https://doi.org/10.1021/acs.jctc.4c00255
16. Kryshtafovych A, Montelione GT, Rigden DJ. *et al.* Breaking the conformational ensemble barrier: ensemble structure modeling challenges in CASP15. *Proteins Struct Funct Bioinf* 2023;**91**:1903–11. https://doi.org/10.1002/prot.26584
17. Gaalswyk K, Muniyat MI, MacCallum JL. The emerging role of physical modeling in the future of structure determination. *Curr Opin Struct Biol* 2018;**49**:145–53. https://doi.org/10.1016/j.sbi.2018.03.005
18. Wayment-Steele HK, Ojoawo A, Otten R. *et al.* Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* 2024;**625**:832–9. https://doi.org/10.1038/s41586-023-06832-9
19. Wallner B. AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics* 2023;**39**:btad573.
20. Kalakoti Y, Wallner B. AFsample2 predicts multiple conformations and ensembles with AlphaFold2. *Commun Biol* 2025;**8**:373.
21. Del Alamo D, Sala D, McHaourab HS. *et al.* Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife* 2022;**11**:e75751. https://doi.org/10.7554/eLife.75751
22. Monteiro da Silva G, Cui JY, Dalgarno DC. *et al.* High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat Commun* 2024;**15**:2464.
23. Lin Z, Akin H, Rao R. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123–30. https://doi.org/10.1126/science.ade2574
24. Jing B, Berger B, Jaakkola T. AlphaFold meets flow matching for generating protein ensembles. *Proceedings of the 41st International Conference on Machine Learning*. *Proceedings of Machine Learning Research: PMLR* 2024;**235**:22277–303. https://doi.org/10.1038/s41409-025-02671-6
25. Wang Y, Wang L, Shen Y. *et al.* Protein conformation generation via force-guided SE(3) diffusion models. *Proceedings of the 41st International Conference on Machine Learning*. *Proceedings of Machine Learning Research: PMLR* 2024;**47**:56835–59. https://doi.org/10.1080/0886022X.2025.2520903
26. Jing B, Erives E, Pao-Huang P. *et al.* EigenFold: generative protein structure prediction with diffusion models. *arXiv preprint* 2023;arXiv:2304.02198, 2023. https://doi.org/10.48550/arxiv.2304.02198
27. Zheng S, He J, Liu C. *et al.* Predicting equilibrium distributions for molecular systems with deep learning. *Nat Mach Intell* 2024;**6**:558–67. https://doi.org/10.1038/s42256-024-00837-3
28. Lewis S, Hempel T, Jiménez-Luna J. *et al.* Scalable emulation of protein equilibrium ensembles with generative deep learning. *bioRxiv* 2024;**2024**:2024.12.05.626885. https://doi.org/10.1101/2024.12.05.626885
29. Jin Y, Huang Q, Song Z. *et al.* P2DFlow: a protein ensemble generative model with SE(3) flow matching. *J Chem Theory Comput* 2025;**21**:3288–96. https://doi.org/10.1021/acs.jctc.4c01620
30. Yang Q, Tang C. On the necessity of an integrative approach to understand protein structural dynamics. *J Zhejiang Univ Sci B* 2019;**20**:496–502. https://doi.org/10.1631/jzus.B1900135

31. Wu F, Jin S, Jin X. *et al.* Pre-training of equivariant graph matching networks with conformation flexibility for drug binding. *Adv Sci* 2022;**9**:2203796.
32. Lu W, Zhang J, Huang W. *et al.* DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nat Commun* 2024;**15**:1071. https://doi.org/10.1038/s41467-024-45461-2
33. Wong SWK, Liu ZJP. Conformational variability of loops in the SARS-CoV-2 spike protein. *Proteins: Struct Funct Bioinf* 2021;**90**: 691–703.
34. Park AK, Moon JH, Oh JS. *et al.* Crystal structure of the response regulator spr1814 from Streptococcus pneumoniae reveals unique interdomain contacts among NarL family proteins. *Biochem Biophys Res Commun* 2013;**434**:65–9. https://doi.org/10.1016/j.bbrc.2013.03.065
35. Perrakis A, Sixma TK. AI revolutions in biology: the joys and perils of AlphaFold. *EMBO Rep* 2021;**22**:e54046. https://doi.org/10.15252/embr.202154046
36. Russo NVD, Estrin DA, Marti MA. *et al.* pH-dependent conformational changes in proteins and their effect on experimental pKas: the Case of Nitrophorin 4. *PLoS Comput Biol* 2012;**8**:e1002761. https://doi.org/10.1371/journal.pcbi.1002761
37. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature* 2007;**450**:964–72. https://doi.org/10.1038/nature06522
38. Jaenicke R, Böhm G. The stability of proteins in extreme environments. *Curr Opin Struct Biol* 1998;**8**:738–48. https://doi.org/10.1016/S0959-440X(98)80094-8
39. Schmid N, Eichenberger AP, Choutko A. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* 2011;**40**:843–56. https://doi.org/10.1007/s00249-011-0700-9
40. Salomón-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *WIREs Comput Mol Sci* 2013;**3**:198–210.
41. Eastman PK, Swails JM, Chodera JD. *et al.* OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* 2017;**13**:e1005659.
42. Brooks BR, Brooks CL, MacKerell AD. *et al.* CHARMM: the biomolecular simulation program. *J Comput Chem* 2009;**30**: 1545–614. https://doi.org/10.1002/jcc.21287
43. Vander Meersche Y, Cretin G, Gheeraert A. *et al.* ATLAS: protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Res* 2024;**52**:D384–92. https://doi.org/10.1093/nar/gkad1084
44. Rodríguez-Espigares I, Torrens-Fontanals M, Tiemann JK. *et al.* GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat Methods* 2020;**17**:777–87. https://doi.org/10.1038/s41592-020-0884-y
45. Liang J, Pitsillou E, Hung A. *et al.* A repository of COVID-19 related molecular dynamics simulations and utilisation in the context of nsp10-nsp16 antivirals. *J Mol Graph Model* 2024;**126**:108666. https://doi.org/10.1016/j.jmgm.2023.108666
46. Beltrán D, Hospital A, Gelpí JL. *et al.* A new paradigm for molecular dynamics databases: the COVID-19 database, the legacy of a titanic community effort. *Nucleic Acids Res* 2023;**52**:D393–403. https://doi.org/10.1093/nar/gkad991
47. Torrens-Fontanals M, Peralta-García A, Talarico C. *et al.* SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions. *Nucleic Acids Res* 2021;**50**:D858–66. https://doi.org/10.1093/nar/gkab977
48. Newport TD, Sansom MSP, Stansfeld PJ. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res* 2018;**47**:D390–7. 10.1093/nar/gky1047
49. Mirarchi A, Giorgino T, De Fabritiis G. mdCATH: a large-scale MD dataset for data-driven computational biophysics. *Scientific Data* 2024;**11**:1299. https://doi.org/10.1038/s41597-024-04140-z
50. Siebenmorgen T, Menezes F, Benassou S. *et al.* MISATO: machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nat Comput Sci* 2024;**4**:367–78. https://doi.org/10.1038/s43588-024-00627-2
51. Monzon AM, Rohr CO, Fornasari MS. *et al.* CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. *Database* 2016;**2016**:baw038. https://doi.org/10.1093/database/baw038
52. Escobedo N, Tunque Cahui RR, Caruso G. *et al.* CoDNaS-Q: a database of conformational diversity of the native state of proteins with quaternary structure. *Bioinformatics* 2022;**38**:4959–61. https://doi.org/10.1093/bioinformatics/btac627
53. Hrabe T, Li Z, Sedova M. *et al.* PDBFlex: exploring flexibility in protein structures. *Nucleic Acids Res* 2016;**44**:D423–8. https://doi.org/10.1093/nar/gkv1316
54. Zhao K, Zhao P, Wang S. *et al.* FoldPAthreader: predicting protein folding pathway using a novel folding force field model derived from known protein universe. *Genome Biol* 2024;**25**:152. https://doi.org/10.1186/s13059-024-03291-x
55. Bauer JA, Bauerová-Hlinková V. Extracting the dynamic motion of proteins using normal mode analysis. *Methods Mol Biol* 2022;**2449**:213–31. https://doi.org/10.1007/978-1-0716-2095-3_9
56. Park SW, Lee BH, Kim MK. Elastic network model: a coarse-grained approach to the study of biomolecular dynamics. *Multiscale Sci Eng* 2023;**5**:104–18. https://doi.org/10.1007/s42493-024-00097-8
57. Gupta A, Singh A, Ahmad N. *et al.* Experimental techniques to study protein dynamics and conformations. In: Tripathi T, Dubey VK, ScienceDirect (eds.). *Advances in Protein Molecular and Structural Biology Methods* 2022, London, UK: Academic Press, an imprint of Elsevier. 181–97.
58. Shukla R, Tripathi T. Molecular dynamics simulation of protein and protein–ligand complexes. In: Singh DB (ed). *Computer-Aided Drug Design*. Singapore: Springer Singapore. 2020, 133–61.
59. Itoh SG, Okumura H, Okamoto Y. Generalized-ensemble algorithms for molecular dynamics simulations. *Mol Simul* 2007;**33**: 47–56. https://doi.org/10.1080/08927020601096812
60. Abrams C, Bussi G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* 2013;**16**:163–99. https://doi.org/10.3390/e16010163
61. Sawade K, Peter C. Multiscale simulations of protein and membrane systems. *Curr Opin Struct Biol* 2022;**72**:203–8. https://doi.org/10.1016/j.sbi.2021.11.010
62. Wang T, He X, Li M. *et al.* Ab initio characterization of protein molecular dynamics with AI^{2}BMD. *Nature* 2024;**635**:1019–27. https://doi.org/10.1038/s41586-024-08127-z
63. Palayam M, Yan L, Nagalakshmi U. *et al.* Structural insights into strigolactone catabolism by carboxylesterases reveal a conserved conformational regulation. *Nat Commun* 2024;**15**:6500.
64. Heilmann N, Wolf M, Kozlowska M. *et al.* Sampling of the conformational landscape of small proteins with Monte Carlo methods. *Sci Rep* 2020;**10**:18211. https://doi.org/10.1038/s41598-020-75239-7
65. Peng C, Zhou X, Liu J. *et al.* Multiple conformational states assembly of multidomain proteins using evolutionary algorithm based on structural analogues and sequential homologues. *Fundam Res* 2024;In press. https://doi.org/10.1016/j.fmre.2024.05.003

66. Hou M, Jin S, Cui X. *et al.* Protein multiple conformation prediction using multi-objective evolution algorithm. *Interdiscip Sci* 2024;**16**:519–31. https://doi.org/10.1007/s12539-023-00597-5
67. Cui X, Xia Y, Hou M. *et al.* M-DeepAssembly: enhanced DeepAssembly based on multi-objective multi-domain protein conformation sampling. *BMC Bioinformatics* 2025;**26**:120.
68. Jin Y, Johannissen LO, Hay S. Predicting new protein conformations from molecular dynamics simulation conformational landscapes and machine learning. *Proteins: Struct Funct Bioinf* 2020;**89**:915–21. https://doi.org/10.1002/prot.26068
69. Casadevall G, Duran C, Estévez-Gay M. *et al.* Estimating conformational heterogeneity of tryptophan synthase with a template-based Alphafold2 approach. *Protein Sci* 2022;**31**:e4426. https://doi.org/10.1002/pro.4426
70. Stein RA, McHaourab HS. SPEACH_AF: sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLoS Comput Biol* 2022;**18**:e1010483. https://doi.org/10.1371/journal.pcbi.1010483
71. Raisinghani N, Parikh V, Foley B. *et al.* AlphaFold2-based characterization of apo and Holo protein structures and conformational ensembles using randomized alanine sequence scanning adaptation: capturing shared signature dynamics and ligand-induced conformational changes. *Int J Mol Sci* 2024;**25**:12968. https://doi.org/10.3390/ijms252312968
72. Sala D, Hildebrand PW, Meiler J. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Front Mol Biosci* 2023;**10**:1121962. https://doi.org/10.3389/fmolb.2023.1121962
73. Hu Y, Yang H, Li M. *et al.* Exploring protein conformational changes using a large-scale biophysical sampling augmented deep learning strategy. *Adv Sci* 2024;**11**:e2400884. https://doi.org/10.1002/advs.202400884
74. Zhuo L, Chi Z, Xu M. *et al.* ProtLLM: an interleaved protein-language LLM with protein-as-word pre-training. *arXiv preprint* 2024;arXiv:2403.07920. https://doi.org/10.48550/arXiv.2403.07920
75. Meier J, Rao R, Verkuil R. *et al.* Language models enable zero-shot prediction of the effects of mutations on protein function. *Adv Neural Inf Process Syst* 2021;**34**:29287–303.
76. Hsu C, Verkuil R, Liu J. *et al.* Learning inverse folding from millions of predicted structures. *International Conference on Machine Learning, PLMR* 2022;**162**:8946–70.
77. Janson G, Valdes-Garcia G, Heo L. *et al.* Direct generation of protein conformational ensembles via machine learning. *Nat Commun* 2023;**14**:774. https://doi.org/10.1038/s41467-023-36443-x
78. Wu KE, Yang KK, van den Berg R. *et al.* Protein structure generation via folding diffusion. *Nat Commun* 2024;**15**:1059. https://doi.org/10.1038/s41467-024-45051-2
79. Liu Y, Yu Z, Lindsay RJ. *et al.* ExEnDiff: an experiment-guided diffusion model for protein conformational Ensemble generation. *PRX Life* 2025;**3**:023013. https://doi.org/10.1101/2024.10.04.616517
80. Zhu J, Li Z, Zheng Z. *et al.* Precise generation of conformational ensembles for intrinsically disordered proteins via fine-tuned diffusion models. *bioRxiv* 2024;**2024**:2024.05.05.592611. https://doi.org/10.1101/2024.05.05.592611
81. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins: Struct Funct Bioinf* 2004;**57**:702–10. https://doi.org/10.1002/prot.20264
82. Guo S, Liu J, Zhou X. *et al.* DeepUMQA: ultrafast shape recognition-based protein model quality assessment using deep learning. *Bioinformatics* 2022;**38**:1895–903. https://doi.org/10.1093/bioinformatics/btac056
83. Liu D, Zhang B, Liu J. *et al.* Assessing protein model quality based on deep graph coupled networks using protein language model. *Brief Bioinform* 2024;**25**:bbad420.
84. Hiranuma N, Park H, Baek M. *et al.* Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat Commun* 2021;**12**:1340. https://doi.org/10.1038/s41467-021-21511-x
85. Wang Z, Eickholt J, Cheng J. MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. *Bioinformatics* 2010;**26**:882–8. https://doi.org/10.1093/bioinformatics/btq058
86. Wang J, Wang W, Shang Y. *et al.* New heuristic methods for protein model quality assessment via two-stage machine learning and hierarchical ensemble. In: Ceballos C, Xu R (eds.). *2022 IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI)*. *IEEE*. Piscataway, NJ, USA: IEEE. 2022, 84–90. https://doi.org/10.1080/0886022X.2025.2520903
87. McGuffin LJ. The ModFOLD server for the quality assessment of protein structural models. *Bioinformatics* 2008;**24**:586–7. https://doi.org/10.1093/bioinformatics/btn014
88. Uziela K, Wallner B. ProQ2: estimation of model accuracy implemented in Rosetta. *Bioinformatics* 2016;**32**:1411–3. https://doi.org/10.1093/bioinformatics/btv767
89. Igashov I, Olechnovič K, Kadukova M. *et al.* VoroCNN: deep convolutional neural network built on 3D Voronoi tessellation of protein structures. *Bioinformatics* 2021;**37**:2332–9. https://doi.org/10.1093/bioinformatics/btab118
90. Ye L, Wu P, Peng Z. *et al.* Improved estimation of model quality using predicted inter-residue distance. *Bioinformatics* 2021;**37**:3752–9. https://doi.org/10.1093/bioinformatics/btab632

Briefings in Bioinformatics, 2025, **26(4)**, bbaf340
https://doi.org/10.1093/bib/bbaf340
Review