Modeling Boltzmann weighted structural ensembles of proteins using AI based methods

Akashnathan Aranganathan[a,b], Xinyu Gu[b,c,*], Dedi Wang[e], Bodhi P. Vani[d], Pratyush Tiwary[b,c,f,*]

[a] *Biophysics Program , University of Maryland, College Park, 20742, MD, USA*
[b] *Institute of Physical Science and Technology, University of Maryland, College Park, 20742, MD, USA*
[c] *University of Maryland Institute for Health Computing, Bethesda, 20852, MD, USA*
[d] *Genentech, 1 DNA Way, South San Francisco, 94080, CA, USA*
[e] *Postdoctoral Fellow, Genentech, 1 DNA Way, South San Francisco, 94080, CA, USA*
[f] *Department of Chemistry and Biochemistry, University of Maryland, College Park, 20742, MD, USA*

Abstract

This review highlights recent advances in AI-driven methods for generating Boltzmann-weighted structural ensembles, which are crucial for understanding biomolecular dynamics and drug discovery. With the rise of deep learning models like AlphaFold2, there has been a shift toward more accurate and efficient sampling of structural ensembles. The review discusses the integration of AI with traditional molecular dynamics techniques as well as experiments, the challenges of conformational sampling, and future directions for AI-driven research in structural biology, particularly in drug discovery and protein dynamics.

Keywords: generative AI sampler, AI-assisting traditional sampling,

1. Introduction

Understanding the function of biomolecules, such as globular proteins, intrinsically disordered proteins/regions, nucleic acids, and various associated

*Corresponding author
Email addresses: xgu1997@gmail.com (Xinyu Gu), ptiwary@umd.edu (Pratyush Tiwary)

small molecules, is crucial for biomedicine, human health, and drug discovery. Traditionally, it has been rooted in the "sequence-structure-function paradigm," which focuses on the most frequently observed or *native state* structure. This paradigm led to the involvement of X-ray crystallography, nuclear magnetic resonance (NMR), and more recently, cryo-electron microscopy and other methods in obtaining structures for around 193,000 proteins and 4300 nucleic acids in the RCSB protein databank as of August 2024. However, this perspective is still incomplete, as biomolecules are inherently dynamic and mostly function through an *ensemble* of structures, understood as a set of representative structures with physical or *Boltzmann* weights representing their probabilities (Fig. 1). Examples include G-protein-coupled receptors (GPCRs), kinase-domain proteins, and RNA riboswitches, which can adopt multiple conformations associated with different states, enabling the downstream function of these biomolecules. Uncovering the comprehensive structural ensemble under equilibrium distribution, is critical for discovering cryptic pockets—druggable pockets hidden in the native state but revealed in metastable states[1] and understanding the allosteric effects, where changes in one part of the molecule affect another upon interacting with binding partners. These often lead to the development of more specific and selective drugs, compared to those targeting the orthosteric site in the native state, which tends to be conserved among protein homologs due to the evolutionary nature of the human proteome[2]. This calls for a shift from the "sequence-structure-function" to the "sequence-ensemble-function" paradigm.

While recently attempts have been made to realize this shift through the use of Artificial Intelligence (AI) in structural bioinformatics, traditionally Molecular Dynamics (MD) has been used for studying conformational ensembles by integrating Newton's equations of motion with an optimized potential. However, all-atom MD suffers from a separation of timescales effect where the integration timestep captures the fastest motion (bond vibration, in femtoseconds) and is much smaller than most events of interest. Enhanced or rare event sampling methods sample a modified distribution with a more uniform distribution across metastable states, then reweight to obtain the "true" physical distribution[3]. Despite the development of various methods, reliable enhanced sampling requires significant *a priori* knowledge, computing resources and is not high-throughput.

From a data-based methods perspective, with the availability of an enormous structural database, particularly for proteins, methods such as Al-

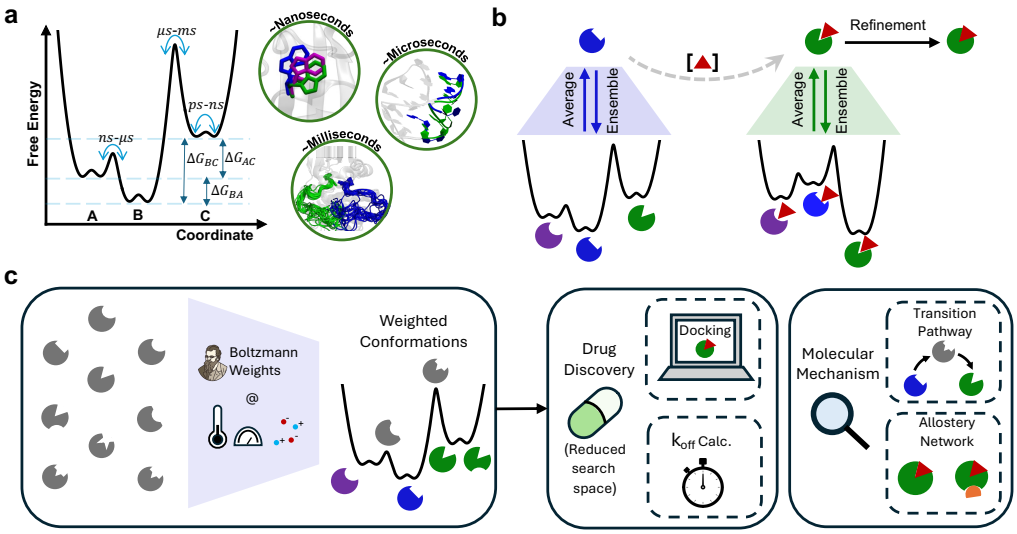

Figure 1: **Underlying distribution of biomolecules dictates their function in a given environmental condition**
a, Shows a representative free energy profile for a biomolecule. It also illustrates the interplay between timescales of biological processes and conformational-metastable states for 1) Trp8 rotomer of Cold Shock Protein[4], 2) Rearrangements of bases in PreQ1 Riboswicth[5], and 3) Activation loop movements for Abl1 kinase[6]. **b**, Illustrates a case where a binding partner (red triangle) triggers the conformational selection towards metastable state for downstream function of a biomolecule. **c** Illustrates the overall importance of Boltzmann weights in drug discovery and in understanding molecular mechanisms.

phaFold2 (AF2)[7], RoseTTAFold2[8] and OpenFold[9] have been developed to solve the protein folding problem. These methods use the transformer model in the natural language processing field to interpret co-evolutionary information in multiple sequence alignments (MSA) to map a primary sequence to its crystal-like structure. Alternative models, such as OmegaFold[10] and ESMFold[11], use Protein Language Models (PLMs) to bypass the requirement of MSA. More recently, AlphaFold3 (AF3)[12] expands its predictive capabilities to complex structures including proteins, nucleic acids, small molecules, ions and more. While these methods exist within the "sequence-structure-function" paradigm, a wide range of methods based on these have been developed to operate from "sequence-ensemble-function" standpoints by modifying the input or prior information of AF2. They include MSA-

subsampling[13] or reducedMSA-AF2 (rMSA-AF2) that reduces the information entering the AF2 by randomly sampling sequences from MSA, AF-cluster that clusters the MSA sequence based on sequence similarity[14], SPEACH_AF[15] that perturbs the MSA with Alanines, and more recently local frustration based MSA clustering/perturbing method[16]. Further, by leveraging the AF2 architecture, the Diffold[17] method uses the diffusion framework to sample heterogeneous conformations. We point to the review article by Sala et al. [18] for details of these and other methods.

However, the majority of biomolecular functions depend on the precise conformational distribution appropriate for the given environmental variables such as temperature, pressure and ion concentration. There is thus a need to obtain not just any distribution but specifically a Boltzmann-weighted distribution of conformations accurate for the environmental conditions. This has been done in many ways, including either by developing directly an AI-based sampler or by using AI to augment enhanced MD. This ensures the system explores conformations with the correct relative probabilities and fluctuations at a given temperature and pressure, following thermodynamic principles. These Boltzmann weights provide insights into the allostery networks and downstream biomolecular function[19], and also reduce the search space of metastable conformations for drug discovery via docking and other applications[20] (Fig.1c). In this mini-review, we will discuss the progress made in generative AI and the influence of AI in assisting traditional methods for biomolecule conformational distribution over the past few years and further outline the key steps we believe the community could be taking to enable sampling of Boltzmann weighted structural ensembles of arbitrary proteins and their complexes.

2. Boltzmann-Weighted Ensemble Sampling with Generative AI

The recent advancements in PLMs and diffusion models have significantly increased the application of generative AI models in predicting biomolecular conformational ensembles with corresponding Boltzmann weights. In particular, AF2 has been an essential element in this progress, particularly for protein molecules, since it has learned a mapping from primary amino acid sequences to contact maps and, subsequently, to Cartesian coordinate space. Leveraging this progress in the past three years to achieve a transferrable architecture, the field of protein conformational ensembles has witnessed a surge in works incorporating the components of AF2 into genera-

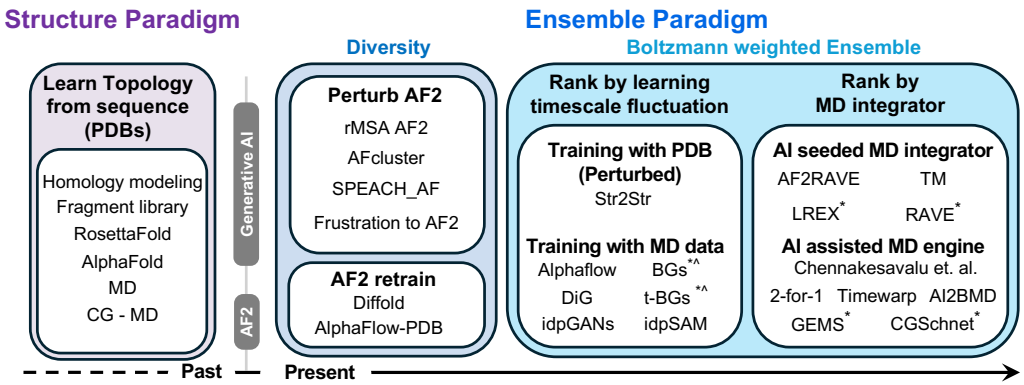

Figure 2: **Paradigm shift to "ensemble-function" realised in computational methods for biomolecules**
Illustrates the emergence of methods in generating Boltzmann-weighted ensemble in the past three years since AF2, generative AI development. * - These methods are not directly influenced by either Generative AI advancements or by utilizing components of AF2. ^ - These methods minimize potential energy through their loss function

tive AI frameworks, including Generative Adversarial Networks (GANs)[21], Flow-based Models (FMs)[22], and Diffusion Models (DMs)[23]. Due to such progress in protein ensemble generation, we explore current protein-specific methods and the methods that can transfer the architecture to all biomolecules in this mini-review. For a detailed assessment of the generative AI methods for biomolecular ensemble sampling, we direct the readers to reviews by Rotskoff [24], Liu et al. [25] and Tiwary et al. [26].

2.1. Generic molecular systems

Before AF2, Boltzmann Generators[27] (BGs) were one of the first methods to use a flow-based generative AI method to obtain Boltzmann reweighted biomolecule samples by minimizing the potential energy in their loss function. BGs training also utilized MD trajectories to enhance the network initialization through a "training-by-example" protocol. Recent updates include equivariant flow matching models[28] and transferrable BGs[29], which prescribes a protocol to account for the topology of the generated molecules and symmetries in the energy function. With this correction, the updated Boltzmann generator architecture can be applied to a wide range of chemically distinct systems due to its inputs in the Cartesian coordinate space.

Despite its elegant protocol, this method requires significant training data to capture essential modes in the underlying probability distribution. Post AF2, new approaches have emerged[30, 31, 32, 33, 34] for protein molecules as AF2's modules can account for the topology and structural representation from the sequence, allowing these methods to work from the primary sequence to lie well within the "sequence-ensemble-function" paradigm.

2.2. Proteins

Following AF2's success with an extensive protein structure database, the Str2Str[30] model implemented heating-annealing training for a score-matching model, allowing moving across energy landscape barriers. The method is trained only on crystal structures and allows for achieving local fluctuations akin to microsecond-long MD simulations. AlphaFlow/ESMFlow [31] is another model that utilizes the AF2 network under a flow-matching framework. This model was trained on the PDB as well as short 100ns long MD datasets to include timescale information in the training set. AlphaFlow/ESMFlow captures local fluctuations compared to the 100-200 ns test set of ATLAS MD by learning nanosecond timescale information in its training set.

Combining some key ideas from AF2 and diffusion models, the Distributional Graphormer (DiG)[32] method uses a score-matching framework on an expressive Graphormer architecture to predict Boltzmann-weighted distributions for proteins. Like AlphaFlow, DiG also uses a PDB and 100-500ns long MD datasets for training. However, unlike previous methods, DiG employs MD potential energy as boundary conditions on its loss function during pretraining, helping it learn the potential energy function. With this setup, DiG outperforms nanosecond-level simulation data and produces distribution comparable with millisecond-long simulations for two SARS-CoV-2 virus proteins.

2.3. Intrinsically Disordered Proteins/Regions (IDP/IDR)

Unlike globular proteins, IDPs/IDRs are dynamically heterogeneous, interconverting between distinct ensembles of states[35]. Understanding their underlying distribution requires MD trajectories to account for timescale information. IdpGAN [33], developed using GANs architecture, generates coarse-grained conformational ensembles of IDPs/IDRs. To improve transferability, Janson and Feig [34] developed idpSAM, a latent diffusion model using transformer architecture. With an increased training set, idpSAM

shows promise as a transferrable all-atom ensemble generator while trained on coarse-grained(CG) data. DynamICE [36] is another model that uses a recurrent neural network, specifically the Long-Short Term Memory(LSTM) network, to incorporate experimental prior in ensemble generation. While these models work on predicting ensembles, ALBATROSS [37] is another bidirectional LSTM-based deep learning network [38] that directly predicts the ensemble dimensions, such as radius of gyration and end to end distance, for a given sequence of IDP/IDR.

2.4. Summary of Generative AI samplers

Altogether, following the paradigm shift from structure to ensemble (Fig.2), generative AI has significantly advanced in predicting Boltzmann-weighted protein conformational distribution. In spite of this current progress (Fig. 3), the field of protein conformational distribution faces two major challenges: 1) a timescale-rich, well-designed training dataset and 2) an evaluation process and metrics. Current works show how nanosecond-level information for diverse proteins could be used to extrapolate to different protein systems. While extrapolation to different systems of varied length scales is manageable due to the current growth in AI, extrapolation to unseen states is highly questionable. As seen in the case of DiG, this depends on the training set, learning architecture and using forcefields in pre-training to overcome the free energy barriers for sampling metastable states. This leads to the question of how to evaluate these weights. Currently, all the methods mentioned use various probability distribution-based metrics to compare with different lengths of MD simulations. However, reliable benchmark datasets, such as experimentally determined Boltzmann ensembles, and standardized metrics, analogous to LDDT or TM-scores in the case of static structures evaluation, are still lacking. Achieving such ambitious desirables in generative AI for Boltzmann weighted ensemble requires a well-curated training dataset to avoid test data leakage, with timescale-rich information and standardized evaluation metrics to extrapolate in both length-scale and timescale.

3. AI-assisted traditional structural sampling methods

AI samplers are known for their efficient sampling, but they struggle to extrapolate beyond training data and cannot estimate dynamic, time-dependent properties like transition pathways or rates. In contrast, traditional methods such as molecular simulations and cryo-EM follow physical

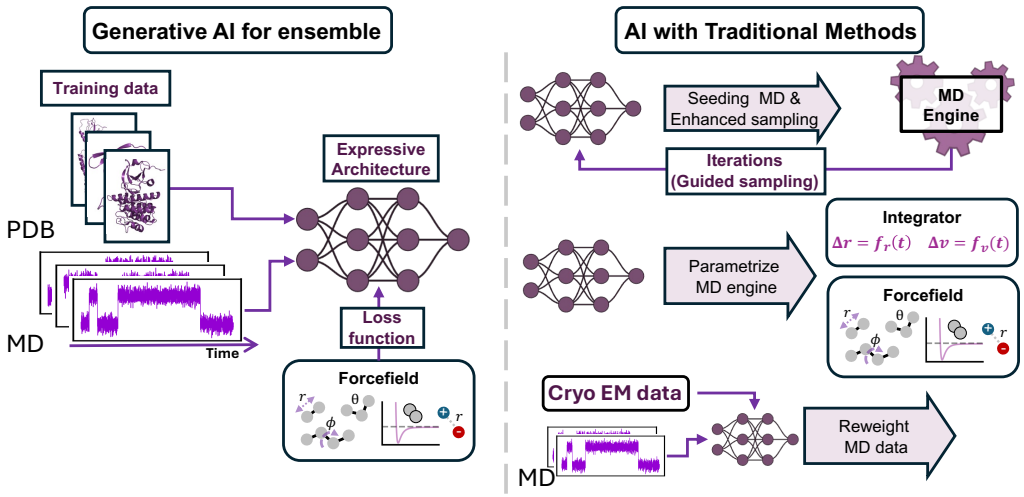

Figure 3: **Summary of current attempts in generating Boltzmann weighted ensemble.**
Illustrates the overall summary of all the attempts made in either using generative AI or augmenting traditional methods in obtaining Boltzmann-weighted ensemble.

laws and can extrapolate, though they face challenges in accelerating sampling and 3D reconstruction. AI methods can assist here, as structures from AI samplers can directly serve as templates or seeds [39]. Additionally, AI-derived scores like AF2 pLDDT provide valuable prior information [40, 41] for traditional methods. Recently, new AI models have been developed to integrate more deeply with these methods. In this section, we review a range of such methods.

3.1. AI for enhanced sampling

Enhanced sampling methods have long been developed to harness the extrapolation power of physics-based simulations, by overcoming free energy barriers and accelerating sampling. Typical enhanced sampling workflow usually involves three key stages: 1. propose possible states/structures, 2. identify reaction coordinates or low-dimensional manifolds for moving between these states, and 3. perform enhanced sampling. For each of these stages, AI methods have been incorporated.

Generating possible structures: While generating a Boltzmann ensemble is complicated, achieving structural diversity is relatively straightforward with generative AI methods. These diverse structures can effectively initial-

ize molecular simulations, enabling generating structures from discontinuous regions of wide conformational spaces. Case studies demonstrate that unbiased MD simulations seeded by rMSA-AF2 can construct Markov State Models Markov State Models with metastable states containing cryptic ligand pockets[42].

Learning reaction coordinate: AI methods for dimensionality reduction that incorporate time information improve the capture of dynamic properties from time series data and help identify reaction coordinates for slow processes, with pioneering examples including time-lagged independent component analysis (TICA)[43] using a linear neural network architecture and VAMPnet [44] employing non-linear architecture. Iterative applications of these methods combined with enhanced sampling can optimize reaction coordinates and enhance convergence on thermodynamic properties, such as Boltzmann weights. For example, the Reweighted Autoencoded Variational Bayes for Enhanced Sampling (RAVE) method [45] iterates between time-lagged autoencoder-type ML models learning low-dimensional reaction coordinates [46, 47] and enhanced sampling schemes like metadynamics. Reweighting biased samples and correcting bias effects for the time series have been implemented for this iterative biasing protocol [48]. An example of AI-integrated enhanced sampling workflow for general proteins is the AF2RAVE protocol, which combines rMSA-AF2 with the RAVE method to systematically explore metastable states and rank structures using Boltzmann weights [4, 49].

Enhanced sampling schemes: AI methods have been used to design new enhanced sampling strategies. For instance, data-driven free energy or bias potential estimators can be used for adaptive biased simulations[50], while reinforcement learning models can probe adaptive sampling initialization or bias deposition as policy selection problems[51, 52]. Additionally, AI models can reduce the computational cost of traditional enhanced sampling methods. For example, in the learned replica exchange method (LREX)[53], Boltzmann generator facilitate replica exchange by mapping high-temperature replica configurations to target temperatures, eliminating the need for intermediate replicas. Thermodynamic Map methods (TM)[54, 55] show promise in using diffusion models combined with MD simulations to extrapolate Boltzmann weights of structural ensembles under conditions not present in the MD training sets.

Advances in AI methods for enhanced sampling have been carefully discussed in recent reviews [56, 57]. While specific methods for each stage of

the enhanced sampling workflow depend on the system and available tools, establishing a general guideline for selecting the most suitable approach at each stage would be highly beneficial. For that purpose, benchmarking these methods with reliable dataset based on unified metrics for sampling efficiency and performance (eg. accuracy and convergence of sampled Boltzmann distribution) is essential. Transferability remains a challenge for AI-augmented enhanced sampling methods, as both MD and AI models often require rerunning for new systems, except in specific cases such as homologous proteins or mutants. In practical applications, prior knowledge and manual tuning are typically inevitable. Looking ahead, interpretable AI and reinforcement learning may play a role in streamlining this process by treating traditional manual tuning as a policy problem.

3.2. AI for Molecular Dynamics surrogates and force-fields

Recent works have integrated AI more deeply into the MD simulation process itself, beyond just enhancing sampling methods. Schreiner et al.[58] implemented an Implicit Transfer Operator (ITO) using denoising diffusion probabilistic models with a SE(3) equivariant architecture to learn surrogates of MD simulations at multiple time resolutions, accurately capturing stochastic dynamics across various time scales. Another example is Timewarp[59], a normalizing flow-based generative model by Klein et al., which learns to make large time steps (10^5-10^6 fs) to simulate MD, achieving a 10^5-fold acceleration over traditional MD simulations. Timewarp demonstrates transferability between different molecular systems and generalizes well to unseen small peptides (2-4 amino acids) at all-atom resolution.

AI has also been employed to accelerate MD simulations by learning coarse-grained (CG) models. This involves using AI to group atoms into larger particles or "beads" [60], develop effective CG force fields such as CGSchNet[61] and 2-for-1 methods[62], and backmap CG configurations to recover atomistic molecular details such as the DiAMoNDBack method[63]. By compromising system resolution in a systematic manner, these AI-powered CG models significantly reduce computational costs, more efficiently sample conformational space and enable the study of larger systems and longer timescales, such as the application of the CALVADOS model [64, 65] to liquid-liquid phase separation for intrinsically disorder proteins.

One advantage of AI-powered MD engines is their transferability across different systems without the need for retraining. However, to truly claim a model is transferable and generalized, its performance must be demonstrated

on large, novel test systems. In future endeavors, improving transferability will require incorporating more high-quality MD data from diverse systems across different temperatures, pressures, or other environmental parameters, into the training set. Moreover, the impact of different CG resolutions transferability and accuracy has not been investigated, and it remains unclear which resolutions would be optimal.

Additionally, AI has been used to facilitate the development of quantum-mechanical quality force fields for accurate *ab initio* MD simulations of large molecular systems containing millions of atoms, making previously prohibitive simulations feasible. These AI-powered force fields, such as AI2BMD[66] and GEMS[67], allow researchers to explore conformational space and detailed interatomic interactions for processes such as folding and unfolding of chignolin, providing a level of detail often unattainable with conventional force fields. For practical applications of AI-driven *ab initio* quality MD on biological systems and process, overcoming the challenge of large timescales will be the next hurdle in the future.

3.3. AI for Boltzmann Ensembles from Cryo-EM

While so far in this section we have focused on AI methods interfacing with molecular simulations, their impact has been significant also in enhancing experimental structural ensemble determination, particularly in cryo-electron microscopy (cryo-EM). Since biomolecules are flash-frozen before Cryo-EM imaging, sampling of each particle follows the Boltzmann distribution prior to freezing. However, reconstructing a 3D ensemble from millions of noisy 2D images is challenging. Each 2D image is a randomly oriented "planar slice" of a biomolecule in a specific conformation, and the low signal-to-noise ratio complicates both 3D reconstruction and free energy landscape estimation.

Both traditional and AI-driven approaches for reconstructing Boltzmann-weighted structural ensembles from single-particle cryo-EM have been systematically reviewed recently [68]. Here, we highlight a few applications of AI in this field. For 3D reconstruction of heterogeneous conformations, VAE models, such as CryoDRGN [69] and TomoDRGN [70], have been widely used, taking cryo-EM/ET 2D images as encoder inputs and 3D density map as decoder outputs. The synthetic 2D images from 3D density map outputs are compared with input images for the VAE reconstruction loss. While GAN models, like Multi-cryoGAN [71] establish a competition between physics reconstruction simulator with a discriminator neural network that distinguishes

between experimental 2D cryo-EM images and the ones synthesized from the 3D reconstruction.

For extracting Boltzmann weights of conformations from Cryo-EM datasets, Bayesian frameworks are often employed to reweight computationally modeled structural ensembles, typically derived from MD simulations and potentially from generative AI in the future. The ensemble reweighting process involves recovering the posterior probability of the weights for each conformational state, given the observed cryo-EM images[72, 73].

4. Outlook

We conclude this review with our outlook on what we perceive as some exciting avenues and open challenges. Numerous reweighting methods have been developed to integrate experimental data from X-ray crystallography, FRET, NMR, and cryo-EM into molecular dynamics simulations. These methods help capture the long-timescale behaviors observed experimentally, addressing sampling limitations in MD and producing more accurate Boltzmann weights[74, 75]. Looking ahead, we expect AI-driven samplers for Boltzmann ensembles to better incorporate diverse experimental priors and high-quality MD data into input channels, model architectures, and training sets. Comprehensive benchmark datasets and metrics are still needed to evaluate these ensembles and their Boltzmann weights. It is crucial to consider not only a model's transferability but also its generalizability, particularly its performance on test sets that include critical point mutations and long-timescale dynamics in large systems.

While there are possibly several relevant applications of Boltzmann-weighted ensembles in biology and associated fields, for the sake of brevity here we highlight one important area, namely structure-based discovery. AI-generated static structures, especially from AF2, show promise for drug discovery [76, 77]. Refining ligand-free AF2 structures with flexible docking or MD improves small molecule docking by addressing ligand-induced fit effects [78, 79]. However, conformational diversity is critical for drug discovery, especially in cryptic pocket and antibody design [80]. Boltzmann weights are key for identifying metastable states and selecting favorable holo (ligand-bound) structures. AF2RAVE-generated structures have demonstrated success in kinase inhibitor studies by proposing metastable, druggable states [6]. Moving forward, we believe AI-driven Boltzmann ensembles offer exciting potential in

Method	System	Input	Architecture	Accuracy	Trans.
Generative AI-Sampler					
BGs[27]	universal	Cartesian Coords	Flow-based	enhanced MD	×
transferrable-BGs [29]	universal	Cartesian Coords	Equivariant-Flow	dipeptide MD (50ns)	✓
Str2Str [30]	proteins	Cartesian Coords	Score-based	fast folding proteins (1μs MD)	✓
AlphaFlow-MD [31]	proteins	AA seq, MSA	Flow-matching	ATLAS MD (100ns)	✓
ESMFlow-MD [31]	proteins	AA seq	Flow-matching	ATLAS MD (100ns)	✓
DiG-proteins [32]	proteins	AA seq	score-based	ms-long MD	✓
idpGAN [33]	IDP	AA seq	GANs	ms-long MD	✓
idpSAM [34]	IDP	AA seq	Latent Diffusion	MCMC	✓
AI-accelerating MD					
RAVE [45]	universal	initial structure	VAE	enhanced MD	×
AF2RAVE [4]	proteins	AA seq, MSA	rMSA-AF2, RAVE	enhanced MD	×
TM [55]	universal	initial structure	score-based	ms-long MD	×
LREX [53]	universal	initial structure	Flow-matching	enhanced MD	×
AI-assisted MD engines/toolkits					
Chennakesavalu et. al.[60]	biomolecules	all-atom structures and forces from MD	GNN, Normalizing flow	μs-long MD	×
DiAMoNDBack [63]	proteins	CG structure	DDPM	all-atom, diverse no BW	✓
Timewarp [59]	peptides	AA seq	Flow-matching	μs-long MD	✓
2-for-1 [62]	proteins	CG MD	score-based	ms-long MD	×
CGSchNet [61]	proteins	AA seq	GNN	μs-long MD	✓
AI2BMD [66]	proteins	initial structure	GNN	*ab initio*	✓
GEMS [67]	biomolecules	initial structure	SpookyNet	*ab initio*, NMR	✓
AI for Cryo-EM reconstruction					
Multi-cryoGAN [71]	biomolecules	2D images	GAN	diverse, no BW	×
CryoDRGN [69]	biomolecules	2D images, pose	VAE	diverse, no BW	×
TomoDRGN [70]	biomolecules	tilt series images	VAE	diverse, no BW	×
ensemble-reweighting [72]	biomolecules	MD, cryo-EM data	Bayesian inference	ms-long MD	×

Table 1: Representative examples of AI methods for generating structural ensembles. Unless otherwise specified as 'no BW', the ensembles generated by the methods above are capable of extracting Boltzmann weights. Transferability of architectures (Trans.) specifies whether the model requires retraining when applied to different systems.

allosteric drug design, understanding biomolecular dynamics and interactions, exploring the proteome and more.

In conclusion, AI-driven methods for generating Boltzmann-weighted structural ensembles have advanced but continue to face significant challenges. Interdisciplinary collaboration is essential to overcome these challenges, improving biomolecular understanding and accelerating therapeutic discovery.

5. Papers of interest

Papers of special interest(*) or outstanding interest (**)

[32]** generative AI model to sample the Boltzmann distribution, by incorporating the physical potential into the loss function of a score-matching framework on a Graphormer architecture.

[29]* transferable Boltzmann Generator, based on a continuous normalizing flows is demonstrated on general dipeptides.

[34]* a transferrable all-atom generator for IDPs, trained on coarse-grained (CG) data. It employs an autoencoder architecture combined with a DDPM model matching the latent space distribution to a prior.

[4]** combined rMSA-AF2 with the Reweighted Autoencoded Variational Bayes for Enhanced Sampling method to extrapolate heterogeneous conformations and assign Boltzmann weights.

[55]* a score matching model to extrapolate thermodynamic properties like Boltzmann weights for structure ensembles at temperatures unseen in the training MD dataset.

[62]* demonstrated that the score function of diffusion model trained on CG MD can approximate a force field for CG MD simulation.

[59]* a Markov chain Monte Carlo algorithm targeting the Boltzmann distribution using a conditional normalising flow as a proposal distribution. Its transferability has been demonstrated on small peptides.

[67]* constructed a machine-learned force field at quantum-mechanical accuracy that can capture long-range interactions in large molecules.

[72]** Bayesian framework reweighting MD ensembles based on Cryo-EM images.

6. Acknowledgments

This work was supported by NIH/NIGMS under award number R35GM142719 (P.T.). A.A. was supported by NCI-UMD Partnership for Integrative Cancer

Research. P.T. is an investigator at the University of Maryland-Institute for Health Computing, which is supported by funding from Montgomery County, Maryland and The University of Maryland Strategic Partnership: MPowering the State, a formal collaboration between the University of Maryland, College Park and the University of Maryland, Baltimore. We thank Suemin Lee and Lukas Herron for their supports and helpful discussions on illustration design.

References

[1] V. Oleinikovas, G. Saladino, B. P. Cossins, F. L. Gervasio, Understanding cryptic pocket formation in protein targets by enhanced sampling simulations, Journal of the American Chemical Society 138 (2016) 14257–14263. URL: https://doi.org/10.1021/jacs.6b05425. doi:10.1021/jacs.6b05425. arXiv:https://doi.org/10.1021/jacs.6b05425, pMID: 27726386.

[2] A. Hadzipasic, C. Wilson, V. Nguyen, N. Kern, C. Kim, W. Pitsawong, J. Villali, Y. Zheng, D. Kern, Ancient origins of allosteric activation in a ser-thr kinase, Science 367 (2020) 912–917. URL: http://dx.doi.org/10.1126/science.aay9959. doi:10.1126/science.aay9959.

[3] O. Valsson, P. Tiwary, M. Parrinello, Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint, Annual Review of Physical Chemistry 67 (2016) 159–184. URL: http://dx.doi.org/10.1146/annurev-physchem-040215-112229. doi:10.1146/annurev-physchem-040215-112229.

[4] B. P. Vani, A. Aranganathan, D. Wang, P. Tiwary, Alphafold2-rave: From sequence to boltzmann ranking, Journal of chemical theory and computation 19 (2023) 4351–4354.

[5] J. L. Jenkins, J. Krucinska, R. M. McCarty, V. Bandarian, J. E. Wedekind, Comparison of a preq1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation, Journal of Biological Chemistry 286 (2011) 24626–24637. URL: http://dx.doi.org/10.1074/jbc.M111.230375. doi:10.1074/jbc.m111.230375.

[6] X. Gu, A. Aranganathan, P. Tiwary, Empowering alphafold2 for protein conformation selective drug discovery with alphafold2-rave, ArXiv (2024).

[7] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with alphafold, Nature 596 (2021) 583–589.

[8] M. Baek, I. Anishchenko, I. R. Humphreys, Q. Cong, D. Baker, F. DiMaio, Efficient and accurate prediction of protein structure using rosettafold2, BioRxiv (2023) 2023–05.

[9] G. Ahdritz, N. Bouatta, C. Floristean, S. Kadyan, Q. Xia, W. Gerecke, T. J. O'Donnell, D. Berenberg, I. Fisk, N. Zanichelli, B. Zhang, A. Nowaczynski, B. Wang, M. M. Stepniewska-Dziubinska, S. Zhang, A. Ojewole, M. E. Guney, S. Biderman, A. M. Watkins, S. Ra, P. R. Lorenzo, L. Nivon, B. Weitzner, Y.-E. A. Ban, S. Chen, M. Zhang, C. Li, S. L. Song, Y. He, P. K. Sorger, E. Mostaque, Z. Zhang, R. Bonneau, M. AlQuraishi, Openfold: retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization, Nature Methods 21 (2024) 1514–1524. URL: http://dx.doi.org/10.1038/s41592-024-02272-z. doi:10.1038/s41592-024-02272-z.

[10] R. Wu, F. Ding, R. Wang, R. Shen, X. Zhang, S. Luo, C. Su, Z. Wu, Q. Xie, B. Berger, J. Ma, J. Peng, High-resolutionde novostructure prediction from primary sequence, bioRxiv (2022). URL: http://dx.doi.org/10.1101/2022.07.21.500999. doi:10.1101/2022.07.21.500999.

[11] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379 (2023)

1123–1130. URL: http://dx.doi.org/10.1126/science.ade2574. doi:10.1126/science.ade2574.

[12] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, , J. M. Jumper, Accurate structure prediction of biomolecular interactions with alphafold 3, Nature (2024) 1–3.

[13] D. Del Alamo, D. Sala, H. S. Mchaourab, J. Meiler, Sampling alternative conformational states of transporters and receptors with alphafold2, Elife 11 (2022) e75751.

[14] H. K. Wayment-Steele, A. Ojoawo, R. Otten, J. M. Apitz, W. Pitsawong, M. Hömberger, S. Ovchinnikov, L. Colwell, D. Kern, Predicting multiple conformations via sequence clustering and alphafold2, Nature 625 (2024) 832–839.

[15] R. A. Stein, H. S. Mchaourab, Speach_af: Sampling protein ensembles and conformational heterogeneity with alphafold2, PLOS Computational Biology 18 (2022) e1010483.

[16] X. Guan, Q.-Y. Tang, W. Ren, M. Chen, W. Wang, P. G. Wolynes, W. Li, Predicting protein conformational motions using energetic frustration analysis and alphafold2, Proceedings of the National Academy of Sciences 121 (2024) e2410662121.

[17] J. Fan, Z. Li, E. Alcaide, G. Ke, H. Huang, E. Weinan, Accurate conformation sampling via protein structural diffusion, bioRxiv (2024). URL: http://dx.doi.org/10.1101/2024.05.20.594916. doi:10.1101/2024.05.20.594916.

[18] D. Sala, F. Engelberger, H. Mchaourab, J. Meiler, Modeling conformational states of proteins with alphafold, Current Opinion in Structural Biology 81 (2023) 102645.

URL: http://dx.doi.org/10.1016/j.sbi.2023.102645. doi:10.1016/j.sbi.2023.102645.

[19] R. Nussinov, Y. Liu, W. Zhang, H. Jang, Cell phenotypes can be predicted from propensities of protein conformations, Current Opinion in Structural Biology 83 (2023) 102722.

[20] A. S. Kamenik, I. Singh, P. Lak, T. E. Balius, K. R. Liedl, B. K. Shoichet, Energy penalties enhance flexible receptor docking in a model cavity, Proceedings of the National Academy of Sciences 118 (2021) e2106195118.

[21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, MIT Press, Cambridge, MA, USA, 2014, p. 2672–2680.

[22] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, J. Mach. Learn. Res. 22 (2021).

[23] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 6840–6851.

[24] G. M. Rotskoff, Sampling thermodynamic ensembles of molecular systems with generative neural networks: Will integrating physics-based models close the generalization gap?, Current Opinion in Solid State and Materials Science 30 (2024) 101158. URL: http://dx.doi.org/10.1016/j.cossms.2024.101158. doi:10.1016/j.cossms.2024.101158.

[25] Z. H. Liu, M. Tsanai, O. Zhang, J. Forman-Kay, T. Head-Gordon, Computational methods to investigate intrinsically disordered proteins and their complexes, 2024. URL: https://arxiv.org/abs/2409.02240. arXiv:2409.02240.

[26] P. Tiwary, L. Herron, R. John, S. Lee, D. Sanwal, R. Wang, Generative artificial intelligence for computational chemistry: a roadmap to predicting emergent phenomena, 2024. URL: https://arxiv.org/abs/2409.03118. doi:10.48550/ARXIV.2409.03118.

[27] F. Noé, S. Olsson, J. Köhler, H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, Science 365 (2019) eaaw1147.

[28] L. Klein, A. Krämer, F. Noé, Equivariant flow matching, 2023. URL: https://arxiv.org/abs/2306.15030. doi:10.48550/ARXIV.2306.15030.

[29] L. Klein, F. Noé, Transferable boltzmann generators, 2024. URL: https://arxiv.org/abs/2406.14426. doi:10.48550/ARXIV.2406.14426.

[30] J. Lu, B. Zhong, Z. Zhang, J. Tang, Str2str: A score-based framework for zero-shot protein conformation sampling, 2023. URL: https://arxiv.org/abs/2306.03117. doi:10.48550/ARXIV.2306.03117.

[31] B. Jing, B. Berger, T. Jaakkola, Alphafold meets flow matching for generating protein ensembles, 2024. URL: https://arxiv.org/abs/2402.04845. doi:10.48550/ARXIV.2402.04845.

[32] S. Zheng, J. He, C. Liu, Y. Shi, Z. Lu, W. Feng, F. Ju, J. Wang, J. Zhu, Y. Min, H. Zhang, S. Tang, H. Hao, P. Jin, C. Chen, F. Noé, H. Liu, T.-Y. Liu, Predicting equilibrium distributions for molecular systems with deep learning, Nature Machine Intelligence 6 (2024) 558–567. URL: http://dx.doi.org/10.1038/s42256-024-00837-3. doi:10.1038/s42256-024-00837-3.

[33] G. Janson, G. Valdes-Garcia, L. Heo, M. Feig, Direct generation of protein conformational ensembles via machine learning, Nature Communications 14 (2023). URL: http://dx.doi.org/10.1038/s41467-023-36443-x. doi:10.1038/s41467-023-36443-x.

[34] G. Janson, M. Feig, Transferable deep generative modeling of intrinsically disordered protein conformations, bioRxiv (2024). URL: http://dx.doi.org/10.1101/2024.02.08.579522. doi:10.1101/2024.02.08.579522.

[35] A. S. Holehouse, B. B. Kragelund, The molecular basis for cellular function of intrinsically disordered protein regions, Nature Reviews Molecular Cell Biology 25 (2023) 187–211. URL: http://dx.doi.org/10.1038/s41580-023-00673-0. doi:10.1038/s41580-023-00673-0.

[36] O. Zhang, M. Haghighatlari, J. Li, Z. H. Liu, A. Namini, J. M. C. Teixeira, J. D. Forman-Kay, T. Head-Gordon, Learning to evolve structural ensembles of unfolded and disordered proteins using experimental solution data, The Journal of Chemical Physics 158 (2023). URL: http://dx.doi.org/10.1063/5.0141474. doi:10.1063/5.0141474.

[37] J. M. Lotthammer, G. M. Ginell, D. Griffith, R. J. Emenecker, A. S. Holehouse, Direct prediction of intrinsically disordered protein conformational properties from sequence, Nature Methods 21 (2024) 465–476. URL: http://dx.doi.org/10.1038/s41592-023-02159-5. doi:10.1038/s41592-023-02159-5.

[38] D. Griffith, A. S. Holehouse, Parrot is a flexible recurrent neural network framework for analysis of large protein datasets, eLife 10 (2021). URL: http://dx.doi.org/10.7554/eLife.70576. doi:10.7554/elife.70576.

[39] J. Ohnuki, K.-i. Okazaki, Integration of alphafold with molecular dynamics for efficient conformational sampling of transporter protein nark, The Journal of Physical Chemistry B (2024).

[40] J. P. Roney, S. Ovchinnikov, State-of-the-art estimation of protein model accuracy using alphafold, Physical Review Letters 129 (2022). URL: http://dx.doi.org/10.1103/PhysRevLett.129.238101. doi:10.1103/physrevlett.129.238101.

[41] A. Jussupow, V. R. Kaila, Effective molecular dynamics from neural network-based structure prediction models, Journal of Chemical Theory and Computation 19 (2023) 1965–1975.

[42] A. Meller, S. Bhakat, S. Solieva, G. R. Bowman, Accelerating cryptic pocket discovery using alphafold, Journal of Chemical Theory and Computation 19 (2023) 4355–4363.

[43] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for markov model construction, The Journal of chemical physics 139 (2013).

[44] A. Mardt, L. Pasquali, H. Wu, F. Noé, Vampnets for deep learning of molecular kinetics, Nature communications 9 (2018) 5.

[45] J. M. L. Ribeiro, P. Bravo, Y. Wang, P. Tiwary, Reweighted autoencoded variational bayes for enhanced sampling (rave), The Journal of chemical physics 149 (2018).

[46] Y. Wang, J. M. L. Ribeiro, P. Tiwary, Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics, Nature communications 10 (2019) 3573.

[47] D. Wang, P. Tiwary, State predictive information bottleneck, The Journal of Chemical Physics 154 (2021).

[48] Y. Wang, P. Tiwary, Understanding the role of predictive time delay and biased propagator in rave, The Journal of Chemical Physics 152 (2020).

[49] B. P. Vani, A. Aranganathan, P. Tiwary, Exploring kinase asp-phe-gly (dfg) loop conformational stability with alphafold2-rave, Journal of Chemical Information and Modeling (2023).

[50] R. Galvelis, Y. Sugita, Neural network and nearest neighbor algorithms for enhancing sampling of molecular dynamics, Journal of chemical theory and computation 13 (2017) 2489–2500.

[51] Z. Shamsi, K. J. Cheng, D. Shukla, Reinforcement learning based adaptive sampling: Reaping rewards by exploring protein conformational landscapes, The Journal of Physical Chemistry B 122 (2018) 8386–8395.

[52] L. Zhang, H. Wang, et al., Reinforced dynamics for enhanced sampling in large atomic and molecular systems, The Journal of chemical physics 148 (2018).

[53] M. Invernizzi, A. Krämer, C. Clementi, F. Noé, Skipping the replica exchange ladder with normalizing flows, The Journal of Physical Chemistry Letters 13 (2022) 11643–11649.

[54] Y. Wang, L. Herron, P. Tiwary, From data to noise to data for mixing physics across temperatures with generative artificial intelligence, Proceedings of the National Academy of Sciences 119 (2022) e2203656119.

[55] L. Herron, K. Mondal, J. S. Schneekloth, P. Tiwary, Inferring phase transitions and critical exponents from limited observations with thermodynamic maps, Proceedings of the National Academy of Sciences 121 (2024) e2321971121. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2321971121. doi:10.1073/pnas.2321971121.

[56] Y. Wang, J. M. L. Ribeiro, P. Tiwary, Machine learning approaches for analyzing and enhancing molecular dynamics simulations, Current opinion in structural biology 61 (2020) 139–145.

[57] S. Mehdi, Z. Smith, L. Herron, Z. Zou, P. Tiwary, Enhanced sampling with machine learning, Annual Review of Physical Chemistry (2024). doi:https://doi.org/10.1146/annurev-physchem-083122-125941.

[58] M. Schreiner, O. Winther, S. Olsson, Implicit transfer operator learning: multiple time-resolution surrogates for molecular dynamics, arXiv preprint arXiv:2305.18046 (2023).

[59] L. Klein, A. Foong, T. Fjelde, B. Mlodozeniec, M. Brockschmidt, S. Nowozin, F. Noé, R. Tomioka, Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics, Advances in Neural Information Processing Systems 36 (2024).

[60] S. Chennakesavalu, D. J. Toomer, G. M. Rotskoff, Ensuring thermodynamic consistency with invertible coarse-graining, The Journal of Chemical Physics 158 (2023).

[61] N. E. Charron, F. Musil, A. Guljas, Y. Chen, K. Bonneau, A. S. Pasos-Trejo, J. Venturin, D. Gusew, I. Zaporozhets, A. Krämer, et al., Navigating protein landscapes with a machine-learned transferable coarse-grained model, arXiv preprint arXiv:2310.18278 (2023).

[62] M. Arts, V. Garcia Satorras, C.-W. Huang, D. Zugner, M. Federici, C. Clementi, F. Noé, R. Pinsler, R. van den Berg, Two for one: Diffusion models and force fields for coarse-grained molecular dynamics, Journal of Chemical Theory and Computation 19 (2023) 6151–6159.

[63] M. S. Jones, K. Shmilovich, A. L. Ferguson, Diamondback: Diffusion-denoising autoregressive model for non-deterministic backmapping of $c\alpha$ protein traces, Journal of Chemical Theory and Computation 19 (2023) 7908–7923.

[64] G. Tesei, T. K. Schulze, R. Crehuet, K. Lindorff-Larsen, Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties, Proceedings of the National Academy of Sciences 118 (2021). URL: http://dx.doi.org/10.1073/pnas.2111696118. doi:10.1073/pnas.2111696118.

[65] G. Tesei, A. I. Trolle, N. Jonsson, J. Betz, F. E. Knudsen, F. Pesce, K. E. Johansson, K. Lindorff-Larsen, Conformational ensembles of the human intrinsically disordered proteome, Nature 626 (2024) 897–904. URL: http://dx.doi.org/10.1038/s41586-023-07004-5. doi:10.1038/s41586-023-07004-5.

[66] T. Wang, X. He, M. Li, Y. Wang, Z. Wang, S. Li, B. Shao, T.-Y. Liu, Ai2bmd: efficient characterization of protein dynamics with ab initio accuracy, bioRxiv (2023) 2023–07.

[67] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, J. T. Berryman, et al., Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments, Science Advances 10 (2024) eadn4397.

[68] W. S. Tang, E. D. Zhong, S. M. Hanson, E. H. Thiede, P. Cossio, Conformational heterogeneity and probability distributions from single-particle cryo-electron microscopy, Current Opinion in Structural Biology 81 (2023) 102626.

[69] E. D. Zhong, T. Bepler, B. Berger, J. H. Davis, Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks, Nature methods 18 (2021) 176–185.

[70] B. M. Powell, J. H. Davis, Learning structural heterogeneity from cryo-electron sub-tomograms with tomodrgn, Nature Methods (2024) 1–12.

[71] H. Gupta, T. H. Phan, J. Yoo, M. Unser, Multi-cryogan: Reconstruction of continuous conformations in cryo-em using generative adversarial networks, in: European Conference on Computer Vision, Springer, 2020, pp. 429–444.

[72] W. S. Tang, D. Silva-Sánchez, J. Giraldo-Barreto, B. Carpenter, S. M. Hanson, A. H. Barnett, E. H. Thiede, P. Cossio, Ensemble reweighting using cryo-em particle images, The Journal of Physical Chemistry B 127 (2023) 5410–5421.

[73] T. Wlodarski, J. O. Streit, A. Mitropoulou, L. Cabrita, M. Vendruscolo, J. Christodoulou, Cryoensemble-a bayesian approach for reweighting biomolecular structural ensembles using heterogeneous cryo-em maps, bioRxiv (2023) 2023–11.

[74] S. Bottaro, T. Bengtsen, K. Lindorff-Larsen, Integrating molecular simulation and experimental data: a bayesian/maximum entropy reweighting approach, Structural bioinformatics: methods and protocols (2020) 219–240.

[75] C. Kolloff, S. Olsson, Rescuing off-equilibrium simulation data through dynamic experimental data with dynammo, Machine Learning: Science and Technology 4 (2023) 045050.

[76] J. Lyu, N. Kapolka, R. Gumpper, A. Alon, L. Wang, M. K. Jain, X. Barros-Álvarez, K. Sakamoto, Y. Kim, J. DiBerto, K. Kim, I. S. Glenn, T. A. Tummino, S. Huang, J. J. Irwin, O. O. Tarkhanova, Y. Moroz, G. Skiniotis, A. C. Kruse, B. K. Shoichet, B. L. Roth, Alphafold2 structures guide prospective ligand discovery, Science (2024) eadn6354.

[77] F. Ren, X. Ding, M. Zheng, M. Korzinkin, X. Cai, W. Zhu, A. Mantsyzov, A. Aliper, V. Aladinskiy, Z. Cao, S. Kong, X. Long, B. H. M. Liu, Y. Liu, V. Naumov, A. Shneyderman, I. V. Ozerov, J. Wang, F. W. Pun, D. A. Polykovskiy, C. Sun, M. Levitt, A. Aspuru-Guzik, A. Zhavoronkov, Alphafold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel cdk20 small molecule inhibitor, Chemical Science 14 (2023) 1443–1452.

[78] H. Guterres, S.-J. Park, W. Jiang, W. Im, Ligand-binding-site refinement to generate reliable holo protein structure conformations from apo structures, Journal of chemical information and modeling 61 (2020) 535–546.

[79] Y. Zhang, M. Vass, D. Shi, E. Abualrous, J. M. Chambers, N. Chopra, C. Higgs, K. Kasavajhala, H. Li, P. Nandekar, H. Sato, E. B. Miller, M. P. Repasky, S. V. Jerome, Benchmarking refined and unrefined alphafold2 structures for hit discovery, Journal of Chemical Information and Modeling 63 (2023) 1656–1667.

[80] M. L. Fernández-Quintero, N. D. Pomarici, A.-L. M. Fischer, V. J. Hoerschinger, K. B. Kroell, J. R. Riccabona, A. S. Kamenik, J. R. Loeffler, J. A. Ferguson, H. R. Perrett, et al., Structure and dynamics guiding design of antibody therapeutics and vaccines, Antibodies 12 (2023) 67.