



Generative molecular dynamics

Downloaded from: <https://research.chalmers.se>, 2026-01-31 16:52 UTC

Citation for the original published paper (version of record):

Olsson, S. (2026). Generative molecular dynamics. *Current Opinion in Structural Biology*, 96.
<http://dx.doi.org/10.1016/j.sbi.2025.103213>

N.B. When citing this work, cite the original published paper.



Generative molecular dynamics

Simon Olsson

Understanding biomolecular function depends on bridging experimental observables with models that capture structural, stationary, and dynamical properties. Molecular dynamics (MD) simulations, in principle provide a bridge, but *the sampling problem* remains a fundamental roadblock toward this goal. In this mini-review, I outline recent progress in the area of Generative MD (GenMD)—an approach where generative AI (GenAI) is used to mimic the statistical distributions resulting from MD simulations, which are inaccessible using current numerical algorithms. Here, I highlight a few exemplars of GenMD and then outline open problems and current limitations.

Addresses

Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Gothenburg, SE-41296, Sweden

Corresponding author: Olsson, Simon (simonols@chalmers.se)

Current Opinion in Structural Biology 2026, **96**:103213

This review comes from a themed issue on **Folding, Binding and Protein Design (2026)**

Edited by **John Christodoulou** and **Michele Vendruscolo**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online xxx

<https://doi.org/10.1016/j.sbi.2025.103213>

0959-440X/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Understanding how proteins function requires characterizing not only their structures but also their dynamics. Advances in data acquisition and data analysis now let us extract flexibility from X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy data, usually through structural models with multiple conformations—*ensembles* [1–4]. Complementary techniques, including NMR relaxation [5] and single-molecule Förster resonance energy transfer (smFRET) [6], further probe conformational exchange and its associated timescales. Yet, experimental observables remain incomplete and indirect, reflecting low-dimensional projections or ensemble averages such as distances or chemical shifts. Interpreting these data in structural and dynamical terms therefore requires some degree of modeling.

Molecular dynamics (MD) simulations provide a powerful complement, yielding atomically detailed models of structure and motion. In principle, MD grants access to both the configurational ensemble and its kinetics, linking microscopic mechanisms to experimental observables [7]. In practice, however, MD is limited by the accuracy of force fields and the finite timescales accessible through current computational techniques. Even with modern hardware and enhanced sampling algorithms [8], exhaustive exploration of conformational space remains challenging, especially complex systems involving multiple slow processes [9,10].

Many of the enhanced sampling techniques—along with techniques to learn collective variables [11,12]—are effective in certain situations where low-dimensional representations of the configurational space can be found. Yet, fairly limiting assumptions are required to close the picture and extract dynamics [68]. This creates an opportunity for *generative models*—machine learning frameworks that can efficiently learn and sample complex, high-dimensional distributions—without necessarily relying on dimension reduction.

Finally, artificial intelligence (AI) models such as AlphaFold [13] have revolutionized sequence-to-structure prediction and are being engineered to also predict flexibility [14–17].

In this review, we discuss a new frontier: *Generative Molecular Dynamics* (GenMD), the use of generative AI to emulate or replace costly numerical simulations at a fraction of the computational cost. This direction complements ongoing advances in machine-learned interatomic potentials, and enhanced sampling techniques, and by directly mimicking the statistical behavior of MD simulations, it has the potential to bridge the gap between accurate physical models and complex biophysical experiments.

Molecular dynamics—an informal overview

MD simulations of biomolecular systems involve the numerical integration of the *Langevin equation* [18], which describes the time evolution of all-atomic positions and their velocities in a potential $U(\bullet)$ and at a fixed thermodynamic state. The potential, or *force field*, is derived under the Born–Oppenheimer approximation and thereby describes the effective microscopic interactions between atomic nuclei. The thermodynamic

state specifies the macroscopic control parameters, typically particle number N , temperature T , and volume V (or pressure p). In other words, MD simulations are a physical model of the time-dependent behavior of molecules, including proteins and other biomolecules.

Each realization of the Langevin equation corresponds to a *trajectory* in phase space (Figure 1). However, experimental observables rarely correspond to single trajectories—they represent *ensemble averages* over all possible realizations of molecular motion. To connect simulations to measurable quantities, we thus move from the stochastic dynamics of individual trajectories to the evolution of the probability distribution $p_t(x, q)$ over positions, x , and momenta, q . This evolution is governed by the *Fokker–Planck equation*, which describes how an ensemble of trajectories evolve from an initial condition $p_0(x, q)$ through phase space under the combined effects of deterministic forces and stochastic thermal noise. For simplicity, we ignore the momenta q in the remainder of this mini-review.

For long times, $t \rightarrow \infty$, $p_t(x)$ approaches the Boltzmann distribution, $\mu(x)$, ensuring that long-time averages along trajectory reproduce ensemble averages. This connection—between the stochastic dynamics of atoms and the statistical properties of ensembles—is what makes MD such a powerful bridge between microscopic physics and macroscopic observables (Figure 1).

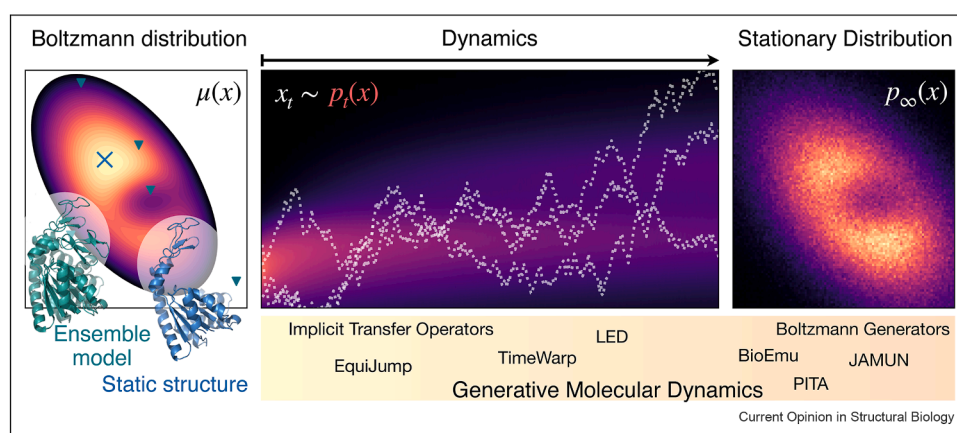
The bridge to experiments comes directly from sampling these statistical distribution: given a *forward model* f , the stationary *ensemble-averaged* observable, such as NMR scalar couplings or free energies, can be computed as averages $O_f = \int f(x)\mu(x) dx$. Similarly, in spectroscopies,

such as NMR and smFRET, we can measure dynamic observables which directly probe evolution of ensembles of trajectories, through an observable probe such as distances or chemical shifts. We can express these experiments as time-correlations between forward model(s) f and g : $O_{fg}^t = \iint p_t(x_0)\mu(x_0)f(x_0)g(x_t) dx_0 dx_t$. In practice, these averages are calculated using statistical samples from equilibrium and the trajectory distributions—thus, generating independent statistical samples from these distributions is how MD is connected to experiments.

In practice, however, the gap between the microsecond to second timescales we probe in experiments, and the femtosecond scale time steps in MD simulations, make direct, quantitative comparisons elusive. This limitation known as the *sampling problem*—and as the name implies, is fundamentally one of probability: drawing, or rather, generating samples from a probability distribution, and using these to compute averages.

The challenge motivates a complementary perspective: instead of relying on explicit time integration, we can attempt to *model* the underlying statistical structure of molecular motion directly, with deep generative models. Broadly, generative approaches to molecular simulation can be organized around two goals. The first is to learn to generate from the *Boltzmann distribution* itself—that is, the equilibrium probability density over molecular configurations—enabling efficient sampling of unbiased ensembles without explicit trajectories. With equilibrium samples, we can calculate stationary observables such as free energies, including protein stability and binding affinities. The second approach is to learn the *time-dependent evolution* of the system, capturing the dynamical transitions between states approximating the transition

Figure 1



Connections among structural biology, molecular dynamics (MD), and generative molecular dynamics (GenMD). Left: The Boltzmann distribution, $\mu(x)$, represents the target equilibrium ensemble, with experimental structures (blue cross) and ensemble models (teal triangles) superimposed. Center: The time-dependent density, $p_t(x)$ (ignoring momenta), illustrates the ensemble of possible MD trajectories evolving through a configuration space. Right: An empirical sample from stationary distribution, $p_\infty(x) \approx \mu(x)$. The probability densities illustrated in the center and right panels are what GenMD aims to target. Below: Representative methods spanning the dynamic-to-stationary spectrum of GenMD.

probability function $p(x_t|x_0) = p_t(x)$ with the point-mass initial condition $p_0(x) = \delta(x_0 - x)$ and a large time, t . Sampling from the transition probability function, corresponds to generating long timescale solutions to the Langevin or Fokker–Planck equation. This perspective allows us to compute dynamic observables, such as binding and unbinding rates. Consequently, these two viewpoints—modeling equilibrium structure versus dynamical evolution—offer complementary routes toward bridging the microscopic realism of molecular dynamics with the data-driven, statistical efficiency and scalability of modern machine learning (ML).

Generative molecular dynamics

The philosophy of generative molecular dynamics (GenMD) is to acknowledge the limitations of numerical simulation schemes and instead learn *surrogate models* that reproduce the statistics needed to compute quantities of interest. As such, these models are not *exactly* faithful to the physical model they approximate—whether it be the potential energy surface or the underlying stochastic dynamics, e.g. Langevin dynamics. Consequently, GenMD offers at best a dramatic speedup in the quantitative computation of ensemble averages, and at worst, becomes a computationally heavy nonsense generator.

GenMD methods are implemented using a growing family of deep generative architectures, which train neural networks to transform samples from an easy-to-sample reference distribution (e.g. a high-dimensional normal) into samples from a difficult target distribution (e.g. a Boltzmann distribution) [19]. These models are typically trained either from data, obtained through simulations, or by interacting directly with a potential energy function.

Modeling the Boltzmann distribution

Boltzmann generators (BGs) [20] are an early and influential class of GenMD. In this approach, a model distribution $\rho(x)$ is trained to approximate $\mu(x)$ using samples obtained either from molecular simulations or directly from the potential energy function. A key advantage of BGs is that they enable the *exact* computation of ensemble averages through importance sampling. Specifically, any observable ensemble average $\langle A(x) \rangle_\mu$ can be evaluated as a weighted average over model-generated samples, with weights

$$w(x) \propto \frac{\exp(-\beta U(x))}{\rho(x)}, \langle A(x) \rangle_\mu \approx \frac{\sum_{i=1}^N w(x_i) A(x_i)}{\sum_{i=1}^N w(x_i)}. \quad (1)$$

This reweighting procedure systematically corrects for modeling errors in the surrogate $\rho(x)$, ensuring unbiased estimates of observables corresponding to the Boltzmann distribution specified by a potential energy model, $U(\cdot)$, and the thermodynamic state, β .

The efficiency and practical usefulness of BGs ultimately depend on how much they reduce the cost of generating a sample from μ compared to state-of-the-art MD [21]. If this cost is low, the up-front expense of training and generating data can quickly be recovered. In practice, efficiency is governed by three factors: the cost of sampling, the computation of $\rho(x)$, and how well a neural network can approximate $\mu(x)$ for diverse molecular systems. Finding generative architectures that satisfy these criteria has proven challenging. Recent work improves generalization across related chemistries, though BGs remain limited to small peptides [22,23,69].

Can we be more pragmatic?

If we regard BGs as an ideal to aspire to, we may ask which corners can be cut while still making meaningful progress. Recognizing that most classical potential energy models are themselves approximate, we might accept that perfect faithfulness to these models is unnecessary and that it may suffice to sample from a distribution that merely approximates their Boltzmann statistics. Viewed this way, we can exploit highly scalable generative architectures while freeing ourselves from the costly computation of $\rho(x)$ and the subsequent reweighting step. Models that follow this more pragmatic strategy are termed *Boltzmann Emulators* (BEs) [24].

Unsurprisingly, BEs have attracted considerable attention and achieved notable success within only a few years. One prominent example is Microsoft's *BioEmu* model [25], which adopts an AlphaFold2-inspired architecture to parameterize a diffusion-based generative model. BioEmu generates coarse-grained, approximately Boltzmann-weighted samples across a broad range of globular proteins, capturing multiple conformational states. Coupled with a fine-tuning strategy against large-scale experimental data on protein thermal stability, BioEmu can be calibrated to predict the destabilizing impact of mutations with remarkable accuracy. Nevertheless, there are several cases where prediction accuracies drop dramatically, such as detection of cryptic pockets, suggesting that these models still have a far way to go.

BioEmu is not alone: a large number of BEs are continuously presented in the literature, differing in both their scope and their representation of molecular configurations within the neural architecture [26,27]. One promising strategy that we have been exploring adopts a divide-and-conquer approach [28,29]. First, a BE generates an ensemble of configurations that spans the metastable states of a molecule; second, an ensemble of short MD simulations is launched from these configurations, combining the best of both paradigms, and offers a pragmatic path forward for scalable molecular simulation, possibly combined with new

reweighing and guiding strategies inspired by the Markov state modeling community [30–38].

Beyond the methods discussed here, there are numerous other AI- and ML-based methods to target this problem, recently reviewed more comprehensively by Rotskoff [39]; Aranganathan et al. [40]; Zhu et al. [41]; and Janson and Feig [42]—for machine-learned interatomic potentials, several excellent reviews are available [43,44].

Modeling the dynamics

The immediate impulse when modeling molecular dynamics with generative models might be to take inspiration from video and text generative models and to model entire trajectories. Indeed, several works have taken this approach [45–47]. However, we know the generative process of MD is Markovian and therefore, the joint distribution of a trajectory is specified by an initial condition, x_0 , and the transition probability density $p(x_t | x_0)$.

Consequently an alternative perspective is to model the transition probability density $p(x_t | x_0)$, directly. This object encodes the probabilistic evolution, and corresponds to solutions to the Langevin equations at long time horizons. This approach enables prediction of dynamic and stationary observables, including binding and unbinding rates and free energies, and can unravel molecular mechanisms of action with speedups of several orders of magnitude [48–54]. Further, in contrast to BGs and BEs, this approach can be trained directly on MD simulation data using arbitrary time strides. Because it does not rely on an independent and identically distributed (i.i.d.) Boltzmann sampling assumption, this can be done without introducing statistical bias. A related idea is *latent space simulation*, which simultaneously learns a latent space and a propagator on that space [55–57].

One interesting strategy is to investigate the mathematical structures of the solutions to the Langevin and Fokker–Planck equations, and come up with training strategies and architectures that exploit these structures. Implicit Transfer Operators (ITOs) do this [48]. ITO uses the Markovian operator—the Transfer Operator—of equilibrium MD that encodes the solutions to the Fokker–Planck equation. In this context, the operator eigenfunctions are independent of t , and are linearly combined with their corresponding t -dependent eigenvalues to encode the solutions. This simple mathematical structure tells us that we should be able to learn a single model to predict the probability of x_t given x_0 for any t . We have shown that this approach can both work on coarse-grained protein representations [48], and can be combined with BEs to boost data efficiency [50]. More recently, we also show that these

models can generalize across different chemical systems [49] with all-atom resolutions. Remarkably, these models can faithfully predict microsecond dynamics on unseen molecules even if the model only ever saw nanosecond-scale dynamics during training. Related methods can also be used as Boltzmann generators, when combined with a Monte Carlo accept and reject scheme as explored in *TimeWarp* Klein et al. [51].

What do we sacrifice?

In moving toward data-driven surrogates of molecular dynamics, we inevitably step away from the explicit physical models that have anchored molecular simulation for decades. Classical molecular dynamics is grounded in equations of motion derived from a potential energy function: forces following from gradients, kinetic, and potential energies are clearly defined, and the integration scheme ensures approximate conservation laws and detailed balance under specified thermodynamic conditions. These properties give trajectories physical meaning and guarantee that long-time statistics correspond to a well-defined ensemble.

With GenMD, by contrast, these mechanistic foundations are replaced by learned statistical transformations optimized to reproduce target distributions or transition statistics. In such models, there is no explicit potential energy function; energy and forces are implicit, at best. As a result, detailed balance—and even Markovianity—are not formally guaranteed. This is particularly true because most learned representations omit momenta altogether or rely on coarse-grained descriptions, both consistent with a Mori–Zwanzig-type projection in which memory effects become unavoidable. Nevertheless, effective strategies are beginning to emerge to mitigate these issues [58].

This raises a deeper question: how can we be confident that we can discover something under such a phenomenological, data-driven paradigm? If the generative process itself is learned rather than derived from physical principles, can it reveal nontrivial phenomena not explicitly encoded in the training data? In traditional physics-based models, we know that simple microscopic laws can give rise to complex emergent behavior. In learned models, these laws are replaced by a complex interplay between architectural inductive biases, model training procedures, and the available training data. I argue that such models can exhibit complex and nontrivial behavior even without explicit physical constraints; however, whether this behavior can consistently translate into genuine scientific insight remains an open question. AlphaFold provides an optimistic example from structural biology. Ultimately, however, the value of GenMD will rest on its ability to generate nontrivial, testable hypotheses that are inaccessible through

other methodologies—or to enable inquiry at scales that are currently impractical.

Ultimately many may remain hesitant: we gain flexibility and computational speed but risk losing interpretability and physical grounding. Balancing these trade-offs, developing more robust testing strategies and correcting models when they violate basic principles will be important. However, I argue that the acceptable balance between speed, fidelity, and interpretability must ultimately depend on the scientific questions being addressed at a given time. Some questions might still be answered even if the tools are imperfect right now.

Where do we go from here?

In the long term, it seems inevitable that we will converge toward learned models that faithfully reproduce the statistical behavior of MD, effectively side-stepping explicit numerical simulations in most biophysical and structural biology applications. Yet, this vision remains distant. My reasoning is as follows: we already know the statistical quantities we aim to reproduce, and numerical algorithms provide a principled—if computationally expensive—route to generate them. Decades of progress in algorithms and hardware have delivered, at best, only a few orders of magnitude of speedup, and while impressive, the complexity of the systems we study in structural biology has grown even faster.

Closing the gap between theoretical descriptions and experimentally observed phenomena will require alternative strategies. GenMD currently stands out as a promising route: one that replaces direct numerical simulation with learned statistical surrogates, trading exactness for scalability in a controlled way, and, in the future, with rigorous statistical testing and experimental data integration.

While these methods show some ability to generalize and extrapolate to larger systems, they are by no means perfect drop-in replacements for MD yet. They suffer from similar error accumulation issues, albeit less severely, as the latent space simulators [55,57], when applied to systems very different from, and in particular much larger than, those seen during training [49]. Most methods consider only a single thermodynamic state; however, recent progress also shows promise in generalizing across thermodynamic states as discussed by Dibak et al. [59]; Moqvist et al. [60]; Janson et al. [61]; Herron et al. [62]; and Qiu et al. [63].

The naive, data-centric approach to resolve these problems would favor scaling training data volumes, model size, and training time, including both simulation and experimental data. This strategy has been celebrated in

other domains including in large language models, lately improvement in model performance in these domains have started plateauing. More data will likely be necessary but it is unclear whether it is sufficient or practical. Going forward, better neural architectures and balancing smart data acquisition techniques with rigorous testing of models might prove to be a productive direction. Beyond this, encouraging transparency and clearly delineating the generalization scope of a given GenMD model, providing well-calibrated confidence scores of samples not unlike those of AlphaFold, and developing hybrid methods which integrate traditional MD with GenMD methods will all have their place. Further, the design of benchmarks that probe a model's ability to reproduce emergent phenomena, including: long-time dynamical behavior, slow relaxation processes, and out-of-distribution dynamical regimes, beyond those explicitly present in the training data, will be essential. Finally, grounding or verification strategies, which test whether the generated samples are faithful to the underlying physical dynamics, satisfy detailed balance or other constraints will likely be significant. However, we must not become impatient: classical MD force fields have undergone decades of development and still fails to be quantitatively predictive and generalizable to arbitrary systems—similarly, simulation algorithms which are both practical and satisfy basic constraints such as equipartition and detailed balance took decades to emerge. In comparison, GenMD is in its infancy and relies on data from classical MD—for now.

Finally, scaling to systems which are of interest in structural biology is also critical for these methods to truly have impact. A significant road block is that current architectures are deep and work with high dimensional feature spaces and dense interaction graphs. These three factors lead to large memory and computational footprints, limiting both the size and the throughput of the leading GenMD models. Moving forward, architectures which more carefully balance expressivity against the costs are needed [64], and the community will likely benefit from classic ideas from computational chemistry and physics [65,66] and long-context ideas from large language models [67].

CRedit authorship contribution statement

Simon Olsson wrote and edited manuscript, prepared figures, and secured funding.

Declaration of competing interest

There are no competing interests to disclose.

Acknowledgements

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. This work is further supported by Chalmers Academic

Excellence Program. I thank Juan Viguera Diez and Christopher Kolloff for feedback on an early draft.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

- Hoff SE, Thomasen FE, Lindorff-Larsen K, Bonomi M: **Accurate model and ensemble refinement using cryo-electron microscopy maps and bayesian inference.** *PLoS Comput Biol* July 2024, **20**, e1012180, <https://doi.org/10.1371/journal.pcbi.1012180>. ISSN 1553-7358.
- Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature x-ray crystallography.** *Proc Natl Acad Sci* Sept. 2011, **108**:16247–16252, <https://doi.org/10.1073/pnas.1111325108>. ISSN 1091-6490.
- Bottaro S, Lindorff-Larsen K: **Biophysical experiments and biomolecular simulations: a perfect match?** *Science* July 2018, **361**:355–360, <https://doi.org/10.1126/science.aat4010>. ISSN 1095-9203.
- Bonomi M, Heller GT, Camilloni C, Vendruscolo M: **Principles of protein structural ensemble determination.** *Curr Opin Struct Biol* Feb. 2017, **42**:106–116, <https://doi.org/10.1016/j.sbi.2016.12.004>. ISSN 0959-440X.
- Palmer III AG: **Nmr probes of molecular dynamics: overview and comparison with other techniques.** *Annu Rev Biophys Biomol Struct* June 2001, **30**:129–155, <https://doi.org/10.1146/annurev.biophys.30.1.129>. ISSN 1545-4266.
- Schuler B: **Single-molecule fret of protein structure and dynamics - a primer.** *J Nanobiotechnol* 2013, **11**(Suppl 1):S2, <https://doi.org/10.1186/1477-3155-11-S1-S2>. ISSN 1477-3155.
- Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F: **Markov models of molecular kinetics: generation and validation.** *J Chem Phys* May 2011, **134**, 174105, <https://doi.org/10.1063/1.3565032>.
- Hénin J, Lelièvre T, Shirts MR, Valsson O, Delemotte L: **Enhanced sampling methods for molecular dynamics simulations.** *Living J. Comput. Mol. Sci* 2022, **4**, <https://doi.org/10.33011/livecoms.4.1.1583>.
- Olsson S, Noé F: **Dynamic graphical models of molecular kinetics.** *Proc Natl Acad Sci* July 2019, **116**:15001–15006, <https://doi.org/10.1073/pnas.1901692116>.
- Hempel T, Olsson S, Noé F: **Markov field models: scaling molecular kinetics approaches to large molecular machines.** *Curr Opin Struct Biol* Dec. 2022, **77**, 102458, <https://doi.org/10.1016/j.sbi.2022.102458>. ISSN 0959-440X.
- Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F: **Identification of slow molecular order parameters for markov model construction.** *J Chem Phys* July 2013, **139**, <https://doi.org/10.1063/1.4811489>. ISSN 1089-7690.
- Wang D, Tiwary P: **State predictive information bottleneck.** *J Chem Phys* Apr. 2021, **154**, <https://doi.org/10.1063/5.0038198>. ISSN 1089-7690.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D: **Highly accurate protein structure prediction with alphafold.** *Nature* July 2021, **596**:583–589, <https://doi.org/10.1038/s41586-021-03819-2>. ISSN 1476-4687.
- Ma P, Li D, Brüscheweiler R: **Predicting protein flexibility with <sc>sc>alphafold</sc>.** *Proteins: Struct, Funct, Bioinf* Feb. 2023, **91**:847–855, <https://doi.org/10.1002/prot.26471>. ISSN 1097-0134.
- Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Hömberger M, Ovchinnikov S, Colwell L, Kern D: **Predicting multiple conformations via sequence clustering and alphafold2.** *Nature* Nov. 2023, **625**:832–839, <https://doi.org/10.1038/s41586-023-06832-9>. ISSN 1476-4687.
- Vani BP, Aranganathan A, Wang D, Tiwary P: **Alphafold2-rave: from sequence to boltzmann ranking.** *J Chem Theor Comput* May 2023, **19**:4351–4354, <https://doi.org/10.1021/acs.jctc.3c00290>. ISSN 1549-9626.
- Jing B, Berger B, Jaakkola T: **Alphafold meets flow matching for generating protein ensembles.** In Proceedings of the 41st international conference on machine learning, *ICML'24*. JMLR.org; 2024.
- Langevin P: **Sur la théorie du mouvement brownien.** *C R Acad Sci* 1908, **146**:530–533.
- John R, Herron L, Tiwary P: **A survey of probabilistic generative frameworks for molecular simulations.** 2024.
- Noé F, Olsson S, Köhler J, Wu H: **Boltzmann generators: sampling equilibrium states of many-body systems with deep learning.** *Science* Sept. 2019, **365**, <https://doi.org/10.1126/science.aaw1147>. ISSN 1095-9203.
- Gloy JF, Hollowflow S, Olsson: **Efficient sample likelihood evaluation using hollow message passing.** In *The thirty-ninth annual conference on neural information processing systems*; 2025. <https://openreview.net/forum?id=KYIIC6sLhw>; 2025.
- Akhound-Sadeh T, Lee J, Bose J, Bortoli VD, Doucet A, Bronstein MM, Beaini D, Ravanbakhsh S, Neklyudov K, Tong A: **Progressive inference-time annealing of diffusion models for sampling from boltzmann densities.** In *The thirty-ninth annual conference on neural information processing systems*; 2025. <https://openreview.net/forum?id=vf2GHcxzMV>; 2025.
- Tan CB, Bose J, Lin C, Klein L, Bronstein MM, Tong A: **Scalable equilibrium sampling with sequential boltzmann generators.** In *Frontiers in probabilistic inference: learning meets sampling*; 2025.
- This paper introduces an improved way to learn Boltzmann generators. The author demonstrates scaling to small peptides and miniproteins.
- Jing B, Corso G, Chang J, Barzilay R, Jaakkola T: **Torsional diffusion for molecular conformer generation.** In Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A. *Advances in neural information processing systems*, **35**. Curran Associates, Inc.; 2022:24240–24253. In https://proceedings.neurips.cc/paper_files/paper/2022/file/994545b2308bbbc97e3e687ea9e464f-Paper-Conference.pdf.
- Lewis S, Hempel T, Jiménez-Luna J, Gastegger M, Xie Y, Foong AYK, Satorras VG, Abidin O, Veeling BS, Zaporozhets I, Chen Y, Yang S, Foster AE, Schneuing A, Nigam J, Barbero F, Stimper V, Campbell A, Yim J, Liene M, Shi Y, Zheng S, Schulz H, Munir U, Sordillo R, Tomioka R, Clementi C, Noé F: **Scalable emulation of protein equilibrium ensembles with generative deep learning.** *Science* Aug. 2025, **389**.
- This paper introduces BioEmu, a method to emulate Boltzmann distributions of single-chain proteins. The method predicts conformational ensembles of conformationally diverse proteins and predicts fold destabilization upon mutation.
- Chennakesavalu S, Rotskoff GM: **Data-efficient generation of protein conformational ensembles with backbone-to-side-chain transformers.** *J Phys Chem B* Feb. 2024, **128**: 2114–2123, <https://doi.org/10.1021/acs.jpcc.3c08195>. ISSN 1520-5207.
- Daigavane A, Vani BP, Davidson D, Saremi S, Rackers J, Jamun J, Kleinhenz: **Bridging smoothed molecular dynamics and score-based learning for conformational ensembles.** 2024.
- Viguera Diez J, Romeo Atance S, Engkvist O, Olsson S: **Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics.** *Mach Learn: Sci Technol* Apr. 2024, **5**, 025010, <https://doi.org/10.1088/2632-2153/ad3b64>. ISSN 2632-2153.
- Havens A, Miller BK, Yan B, Domingo-Enrich C, Sriram A, Wood B, Levine D, Hu B, Amos B, Karrer B, Fu X, Liu G-H,

- Chen RTQ: *Adjoint sampling: highly scalable diffusion samplers via adjoint matching*. 2025. <https://arxiv.org/abs/2504.11713>.
30. Kolloff C, H  ppe T, Angelis E, Schreiner MJ, Bauer S, Dittadi A, Olsson S: *Minimum-excess-work guidance*. 2025.
 31. Wu H, Paul F, Wehmeyer C, No   F: **Multiensemble markov models of molecular thermodynamics and kinetics**. *Proc Natl Acad Sci* May 2016, **113**, <https://doi.org/10.1073/pnas.1525092113>. ISSN 1091-6490.
 32. Trendelkamp-Schroer B, Wu H, Paul F, No   F: **Estimation and uncertainty of reversible markov models**. *J Chem Phys* Nov. 2015, **143**, <https://doi.org/10.1063/1.4934536>. ISSN 1089-7690.
 33. Olsson S, Wu H, Paul F, Clementi C, No   F: **Combining experimental and simulation data of molecular processes via augmented markov models**. *Proc Natl Acad Sci* July 2017, **114**:8265–8270, <https://doi.org/10.1073/pnas.1704803114>. ISSN 1091-6490.
 34. Kolloff C, Olsson S: **Rescuing off-equilibrium simulation data through dynamic experimental data with dynammo**. *Mach Learn: Sci Technol* Dec. 2023, **4**, 045050, <https://doi.org/10.1088/2632-2153/ad10ce>. ISSN 2632-2153.
 35. Wu H, Mey ASJS, Rosta E, No   F: **Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states**. *J Chem Phys* Dec. 2014, **141**, <https://doi.org/10.1063/1.4902240>. ISSN 1089-7690.
 36. Stelzl LS, Hummer G: **Kinetics from replica exchange molecular dynamics simulations**. *J Chem Theor Comput* July 2017, **13**:3927–3935, <https://doi.org/10.1021/acs.jctc.7b00372>. ISSN 1549-9626.
 37. Chodera JD, Swope WC, No   F, Prinz J-H, Shirts MR, Pande VS: **Dynamical reweighting: improved estimates of dynamical properties from simulations at multiple temperatures**. *J Chem Phys* June 2011, **134**, <https://doi.org/10.1063/1.3592152>. ISSN 1089-7690.
 38. Keller BG, Bolhuis PG: **Dynamical reweighting for biased rare event simulations**. *Annu Rev Phys Chem* June 2024, **75**: 137–162, <https://doi.org/10.1146/annurev-physchem-083122-124538>. ISSN 1545-1593.
 39. Rotskoff GM: **Sampling thermodynamic ensembles of molecular systems with generative neural networks: will integrating physics-based models close the generalization gap?** *Curr Opin Solid State Mater Sci* June 2024, **30**, 101158, <https://doi.org/10.1016/j.cossms.2024.101158>. ISSN 1359-0286.
 40. Aranganathan A, Gu X, Wang D, Vani BP, Tiwary P: **Modeling boltzmann-weighted structural ensembles of proteins using artificial intelligence–based methods**. *Curr Opin Struct Biol* Apr. 2025, **91**, 103000, <https://doi.org/10.1016/j.sbi.2025.103000>. ISSN 0959-440X.
 41. Zhu K, Trizio E, Zhang J, Hu R, Jiang L, Hou T, Bonati L: **Enhanced sampling in the age of machine learning: algorithms and applications**. *Chem Rev* Oct. 2025, <https://doi.org/10.1021/acs.chemrev.5c00700>. ISSN 1520-6890.
 42. Janson G, Feig M: **Generation of protein dynamics by machine learning**. *Curr Opin Struct Biol* Aug. 2025, **93**, 103115, <https://doi.org/10.1016/j.sbi.2025.103115>. ISSN 0959-440X.
 43. Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, Sch  tt KT, Tkatchenko A, M  ller K-R: **Machine learning force fields**. *Chem Rev Mar. 2021*, **121**:10142–10186, <https://doi.org/10.1021/acs.chemrev.0c01111>. ISSN 1520-6890.
 44. Jacobs R, Morgan D, Attarian S, Meng J, Shen C, Wu Z, Xie CY, Yang JH, Artrith N, Blaiszik B, Ceder G, Choudhary K, Csanyi G, Cubuk ED, Deng B, Drautz R, Fu X, Godwin J, Honavar V, Isayev O, Johansson A, Kozinsky B, Martiniani S, Ong SP, Poltavsky I, Schmidt K, Takamoto S, Thompson AP, Westermayr J, Wood BM: **A practical guide to machine learning interatomic potentials – status and future**. *Curr Opin Solid State Mater Sci* Mar. 2025, **35**, 101214, <https://doi.org/10.1016/j.cossms.2025.101214>. ISSN 1359-0286.
 45. Jing B, Stark H, Jaakkola T, Berger B: **Generative modeling of molecular dynamics trajectories**. In *The thirty-eighth annual conference on neural information processing systems*; 2024.
- The authors propose a generative modeling framework for MD trajectories. The approach allows for conditioning on initial and final frames (or system parts) to perform a range of different tasks. They validate this approach on tetrapeptide simulations and show preliminary scaling toward protein-monomer systems.
46. Petersen M, Roig G, Covino R: *Dynamicsdiffusion: generating and rare event sampling of molecular dynamic trajectories using diffusion models*. 2024. <https://openreview.net/forum?id=kX1YUi8ff>.
 47. Murtada MH, Brotzakis ZF, Vendruscolo M: *Md-llm-1: a large language model for molecular dynamics*. 2025.
 48. Schreiner M, Winther O, Olsson S: **Implicit transfer operator learning: multiple time-resolution models for molecular dynamics**. In Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S. *Advances in neural information processing systems*, **36**. Curran Associates, Inc.; 2023:36449–36462.
- This paper introduced Implicit Transfer Operator learning, an approach to learn a multiple time-step molecular dynamics surrogate model. In principle, this method allows for the simulation of MD at any timestep. The authors demonstrate quantitative prediction across a broad range of time-scales.
49. Diez JV, Schreiner M, Olsson S: *Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics*. 2025.
- This paper introduces Transferable Implicit Transfer Operator learning, showing that Implicit Transfer Operator models can be learned in a transferable way. The paper illustrates excellent performance on small organic molecules and peptides, predicting thermodynamic and kinetics quantitatively including exchange rate orders of magnitude slower than the training data has been observed. Qualitative extrapolation to larger systems is also demonstrated. <https://doi.org/10.48550/ARXIV.2510.07589>.
50. Diez JV, Schreiner MJ, Engkvist O, Olsson S: **Boltzmann priors for implicit transfer operators**. In *The thirteenth international conference on learning representations*; 2025.
- This paper introduces a strategy to combine Boltzmann emulators with Implicit Transfer Operators. The strategy systematically improves the data efficiency, allowing for improved estimation of MD surrogates with less data.
51. Klein L, Foong A, Fjelde T, Mlodozieniec B, Brockschmidt M, Nowozin S, Noe F, Timewarp R Tomioka: **Transferable acceleration of molecular dynamics by learning time-coarsened dynamics**. In Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S. *Advances in neural information processing systems*, **36**. Curran Associates, Inc.; 2023:52863–52883.
- This paper introduces TimeWarp, a normalizing flow-based model that allows for the sampling of solutions to MD with large timesteps. The authors show that the method can be used as a Boltzmann generator by using it as a Monte Carlo move and introducing a Metropolis-style acceptance criterion.
52. dos Santos Costa A, Mitnikov I, Pellegrini F, Daigavane A, Geiger M, Cao Z, Kreis K, Smidt T, Kucukbenli E, Jacobson J: *Equijump: protein dynamics simulation via so(3)-equivariant stochastic interpolants*. 2024. <https://arxiv.org/abs/2410.09667>.
 53. dos Santos Costa A, Ponnampati M, Rubin D, Smidt T, Jacobson J: *Accelerating protein molecular dynamics simulation with deep-jump*. 2025. <https://arxiv.org/abs/2509.13294>.
 54. Vlachas PR, Zavadlav J, Praprotnik M, Koumoutsakos P: **Accelerated simulations of molecular systems through learning of effective dynamics**. *J Chem Theor Comput* Dec. 2021, **18**:538–549, <https://doi.org/10.1021/acs.jctc.1c00809>. ISSN 1549-9626.
 55. Sidky H, Chen W, Ferguson AL: **Molecular latent space simulators**. *Chem Sci* 2020, **11**:9459–9467, <https://doi.org/10.1039/D0SC03635H>. ISSN 2041-6539.
 56. Wu H, Mardt A, Pasquali L, Noe F: *Deep generative markov state models*. 2019. <https://arxiv.org/abs/1805.07601>.
 57. Tsai S-T, Fields E, Xu Y, Kuo E-J, Tiwary P: **Path sampling of recurrent neural networks by incorporating known physics**. *Nat Commun* Nov. 2022, **13**, <https://doi.org/10.1038/s41467-022-34780-x>. ISSN 2041-1723.
 58. Liu B, Cao S, Boysen JG, Xue M, Huang X: **Memory kernel minimization-based neural networks for discovering slow**

- collective variables of biomolecular dynamics.** *Nat Comput Sci* June 2025, **5**:562–571, <https://doi.org/10.1038/s43588-025-00815-8>. ISSN 2662-8457.
59. Dibak M, Klein L, Krämer A, Noé F: **Temperature steerable flows and boltzmann generators.** 2022. <https://arxiv.org/abs/2108.01590>.
 60. Moqvist S, Chen W, Schreiner M, Nüske F, Olsson S: **Thermodynamic interpolation: a generative approach to molecular thermodynamics and kinetics.** *J Chem Theor Comput* Feb. 2025, **21**:2535–2545, <https://doi.org/10.1021/acs.jctc.4c01557>. ISSN 1549-9626.
 61. Janson G, Jussupow A, Feig M: **Deep generative modeling of temperature-dependent structural ensembles of proteins.** Mar. 2025, <https://doi.org/10.1101/2025.03.09.642148>.
 62. Herron L, Mondal K, Schneekloth JS, Tiwary P: **Inferring phase transitions and critical exponents from limited observations with thermodynamic maps.** *Proc Natl Acad Sci* Dec. 2024, **121**, <https://doi.org/10.1073/pnas.2321971121>. ISSN 1091-6490.
 63. Qiu Y, John R, Herron L, Tiwary P: **Latent thermodynamic flows: unified representation learning and generative modeling of temperature-dependent behaviors from limited data.** 2025.
 64. Irwin R, Tibo A, Janet JP, Olsson S: **Semlaflow – efficient 3d molecular generation with latent attention and equivariant flow matching.** 2025. <https://arxiv.org/abs/2406.07266>.
 65. Cheng B: **Latent ewald summation for machine learning of long-range interactions.** *npj Comput Mater* Mar. 2025, **11**, <https://doi.org/10.1038/s41524-025-01577-7>. ISSN 2057-3960.
 66. Kosmala A, Gasteiger J, Gao N, Günnemann S: **Ewald-based long-range message passing for molecular graphs.** In *International conference on machine learning (ICML)*; 2023.
 67. Frank JT, Chmiela S, Müller K-R, Unke OT: **Euclidean fast attention: machine learning global atomic representations at linear cost.** 2024. <https://arxiv.org/abs/2412.08541>.
 68. Hyperdynamics A F Voter: **Accelerated molecular dynamics of infrequent events.** *Phys Rev Lett* May 1997, **78**:3908–3911, <https://doi.org/10.1103/PhysRevLett.78.3908>. ISSN 1079-7114.
 69. Klein L, Noé F: **Transferable boltzmann generators.** In Globerson A, Mackey L, Belgrave D, Fan A, Paquet U, Tomczak J, Zhang C. *Advances in neural information processing systems*, **37**. Curran Associates, Inc.; 2024:45281–45314.

This is the first paper to rigorously demonstrate transferable estimation of Boltzmann generators.