Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics

Downloaded from: https://research.chalmers.se, 2026-05-12 19:48 UTC

Citation for the original published paper (version of record):
Viguera Diez, J., Schreiner, J., Olsson, S. (2026). Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics. Science advances, 12(15).
http://dx.doi.org/10.1126/sciadv.aed2333

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

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Check for updates

MACHINE LEARNING

Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics

Juan Viguera Diez[1,2], Mathias Schreiner[1], Simon Olsson[1]*

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Understanding the molecular structure, dynamics, and reactivity requires bridging processes that occur across widely separated timescales. Conventional molecular dynamics simulations provide an atomistic resolution, but their femtosecond time steps limit access to the slow conformational changes and relaxation processes that govern chemical function. Here, we introduce a deep generative modeling framework that accelerates sampling of molecular dynamics by four orders of magnitude while retaining physical realism. Applied to small organic molecules and peptides, the approach enables quantitative characterization of equilibrium ensembles and dynamical relaxation processes that were previously only accessible by costly brute-force simulation. The method generalizes across chemical composition and system size, extrapolating to peptides larger than those used for training, and captures chemically meaningful transitions on extended timescales. By expanding the accessible range of molecular motions without sacrificing the atomistic detail, this approach opens opportunities for probing conformational landscapes, thermodynamics, and kinetics in systems central to chemistry and biophysics.

INTRODUCTION

Many of the most important observables in statistical mechanics—such as the stability of a folded protein, the conformational transitions underlying allosteric regulation, or the unbinding rate of a drug from its target—are central to understanding chemical and biological function. These processes span timescales from nanoseconds to seconds, and while they are directly accessible through experiments such as spectroscopy (*1*) and single-molecule techniques (*2*), their atomistic origins are often hidden.

Molecular dynamics (MD) offers a powerful complement to such experiments. By simulating the trajectories of atoms and molecules at an atomic resolution, MD connects the fundamental interatomic forces that govern molecular motion to the statistical behavior observed in bulk. In this way, simulations provide a mechanistic bridge between microscopic physics and macroscopic phenomena (*3*).

Yet, MD comes with a fundamental limitation. To ensure numerical stability, simulations must take time steps small enough to resolve the fastest motions in the system, such as bond and angle vibrations. This requirement restricts MD to femtosecond update steps, even though many processes of chemical and biological interest—protein folding, conformational transitions, and ligand binding—unfold over microseconds to seconds. These processes are typically governed by rare transitions between metastable states (*4*), creating a persistent gap between simulation and experiment that limits our ability to characterize slow molecular processes with statistical confidence (*5*).

This challenge, known as the "sampling problem," continues to inspire a growing array of strategies aiming at accelerating the observation of rare events. Most approaches either bias the underlying dynamics or simulate multiple coupled replicas in parallel, both designed to make infrequent transitions occur more often (*6*). Biasing methods rely on the definition of collective variables that capture the progress of a process of interest (*7*–*9*). However, identifying suitable collective variables for complex, high-dimensional systems remains difficult: Variables that accelerate one process may obscure others, and their design has become a discipline in its own right with numerous options (*10*), including ones derived using machine learning–based strategies (*11*–*13*). Furthermore, given that these methods bias the dynamic behavior of the system, estimation of kinetic properties is only possible under restrictive conditions (*14*). A complementary line of work seeks to increase the integration time step directly, reducing the number of integration steps per unit time. Despite decades of intense research in this direction (*15*–*17*), integration steps remain on the femtosecond scale, leaving even the most efficient simulations orders of magnitude too slow to capture experimentally relevant molecular processes.

A parallel line of progress has focused on harnessing ever-larger computational resources. Specialized compute architectures (*18*, *19*) have achieved continuous millisecond-scale trajectories for small proteins, revealing a mechanistic detail inaccessible to conventional hardware (*20*). Distributed platforms such as Folding@home leverage millions of short trajectories contributed by volunteers (*21*), while modern graphics processing unit (GPU)–based algorithms have brought comparable acceleration to widely used MD engines (*22*–*25*). Together with statistical frameworks such as Markov state models (*26*–*29*), these efforts have enabled the reconstruction of long-timescale kinetics from massive ensembles of short simulations. Yet, all remain bound by the need to generate femtosecond-resolved trajectories, keeping progress tied to extreme computational resources. A conceptually different approach would be to model the effective long-lag dynamics directly without resorting to brute-force sampling or biasing (*30*).

MD simulations function through the numerical integration of the Langevin equation (*31*). As we move along the simulation trajectory, MD generates statistical samples from a transition probability distribution $p(\mathbf{x}_{t+\tau}|\mathbf{x}_t)$ where τ is on the order of femtoseconds, and $\mathbf{x}_t$ and $\mathbf{x}_{t+\tau}$ are points in the phase space. Crucially, this distribution is not ad hoc: It approximates Green's function of the Fokker-Planck equation governing Langevin dynamics, providing the theoretical foundation for viewing MD trajectories as stochastic samples from

[1]Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden. [2]Molecular AI, Discovery Sciences, R&D, AstraZeneca Gothenburg, Pepparedsleden 1, 431 50 Mölndal, Sweden.
*Corresponding author. Email: simonols@chalmers.se

Downloaded from https://www.science.org on April 27, 2026

an underlying probabilistic process (*32*). It follows that analogous transition probability distributions exist for much larger Δt and that these can, in principle, be learned directly for a given molecular system (*33*). Learning such long-lag transition densities offers a direct route to coarse-grained yet statistically faithful dynamics, sidestepping the need for explicit time integration.

Here, we introduce Transferable Implicit Transfer Operators (TITO), a deep generative framework that learns these transition probability distributions across molecular systems. TITO allows us to choose the simulation step size freely, whether to match the characteristic timescales of experiments or to accelerate sampling of slow conformational transitions. Trained on MD data from small molecules and short peptides, TITO simultaneously learns transitions at multiple step sizes, ensuring consistency with the underlying stochastic process. As a result, it preserves key statistical properties such as Boltzmann equilibrium, Markovianity, and relaxation dynamics, suggesting approximate energy conservation and equipartition.

TITO demonstrates quantitative transferability to molecular systems of similar size as in the training data and provides qualitative insights for molecules twice as large. Unlike conventional simulation-based sampling, TITO offers explicit control over the trade-off between accuracy and computational cost, enabling speedups of up to 15,000-fold. By learning effective long-lag dynamics directly, TITO takes a step toward bridging the long-standing gap between atomistic resolution and experimentally relevant timescales. More broadly, it establishes a previously unidentified paradigm for accelerating molecular simulations, with the potential to extend atomistic modeling to processes previously beyond reach.

Transferable Implicit Transfer Operators

At its core, TITO (Fig. 1) learns the effective rules of molecular motion: predicting how atomic configurations evolve over time without explicit time integration. Rather than advancing dynamics through numerous femtosecond scaled steps, TITO draws statistical samples directly from the transition distribution $p(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t)$, capturing how configurations change over a specified lag time Δt, which may be arbitrarily long. Trained across diverse molecular systems and lag times, TITO generalizes both across chemistry and temporal scale.

Training proceeds from reference MD trajectories simulated with a small integration step τ

$$\mathbf{X} = \{\mathbf{x}_\tau, \ldots, \mathbf{x}_{N\tau}\}, \quad \mathbf{x}_{n\tau} \sim p(\mathbf{x}_{n\tau}|\mathbf{x}_{(n-1)\tau}), \quad n = 1, \ldots, N \tag{1}$$

collected across a diverse set of molecules. From these data, the model learns to reproduce the time-integrated transition statistics that would arise if the dynamics were propagated at much larger effective steps $\Delta t = m\tau$, where m is an arbitrary large integer.

We parametrize the transition probability distribution using a continuous normalizing flow (CNF) through the equivariant flow matching (*34*, *35*) objective. A CNF consists of an ordinary differential equation (ODE) and an easy-to-sample "base distribution," p_0, such as a Gaussian (*36*). The velocity field of the ODE is parameterized with a neural network model, which is trained to ensure that the resulting flow transports samples from p_0 to a distribution p_1 closely matching the target data distribution, here, the transition probability distribution. The flow is then the set of all integral paths $\mathbf{x}^T_{t+\Delta t}$, where $T \in [0, 1]$ is the ODE integration time. Throughout this work, superscripts denote ODE integration time, while subscripts indicate MD simulation time.

In practice, we learn the weights θ of a neural network, $\mathbf{v}_\theta(\mathbf{x}^T_{t+\Delta t}; \mathbf{x}_t, \Delta t, T)$, to match a conditional flow, which approximates the transition probability distribution, by minimizing the conditional flow matching loss

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_t, \mathbf{x}_{t+\Delta t} \sim \mathbf{X}, T \sim \mathcal{U}(0,1)} \left[\left\| \mathbf{v}_\theta(\mathbf{x}^T_{t+\Delta t}; \mathbf{x}_t, \Delta t, T) - (\mathbf{x}^1_{t+\Delta t} - \mathbf{x}^0_{t+\Delta t}) \right\|^2 \right] \tag{2}$$

During training, we sample molecules and lag times jointly, enabling TITO to generalize across both the chemical composition and temporal scale. After training, new trajectories are generated by

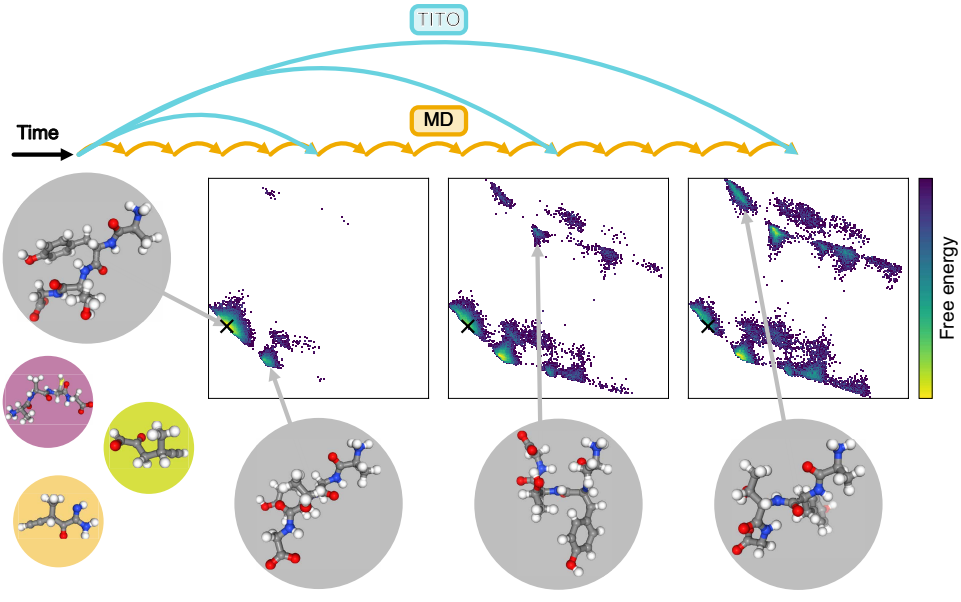

Fig. 1. Transferable Implicit Operators (TITO). A multitimescale surrogate model for MD that is transferable across systems. Starting from an initial condition (black cross), TITO generates MD ensembles for diverse molecules at arbitrary lag times.

Downloaded from https://www.science.org on April 27, 2026

sampling from p_0 and integrating the learned ODE defined by $\mathbf{v}_\theta$. Full model details are provided in the "Model" section.

We train TITO models on two datasets. The first, MDQM9-nc (*37*), contains MD simulations of small organic molecules, while the second, Timewarp (*38*), provides tetrapeptide trajectories. Together, these datasets enable training across a range of molecular sizes and chemistries. We provide details on dataset generation and preprocessing in the "Data" section.

RESULTS

Integrity of the Boltzmann distribution under TITO dynamics in unseen small molecules and peptides

A defining property of molecular systems undergoing Langevin dynamics is convergence to the Boltzmann distribution. In contrast, when a generative model is trained to approximate time-integrated transition probabilities, this guarantee is no longer automatic. The central question is therefore whether TITO preserves physical realism—whether it samples configurations consistent with the Boltzmann distribution—or instead produces unphysical states, analogous to large language models generating text that is fluent but factually incorrect.

To test this, we examined whether the Boltzmann distribution, $\mu \propto \exp\left[-\beta U(\mathbf{x})\right]$, of the potential energy function, U, at inverse temperature β is the invariant measure (i.e., stationary distribution) of the transfer operator implicitly learned by TITO (*39*). Because the learned operator is not directly accessible, we assessed this property numerically. Specifically, we drew initial conditions $\mathbf{x}_0 \sim \mu$ from long unbiased MD simulations, generated trajectories using $p(\mathbf{x}_{\Delta t}|\mathbf{x}_0)$ with TITO, and compared the empirical distributions of $\mathbf{x}_0$ and $\mathbf{x}_{\Delta t}$ using the Jensen-Shannon divergence (JSD) (*40*). Across a broad set of small molecules and tetrapeptides unseen during training, TITO reproduced the Boltzmann distribution obtained from reference MD simulations (Fig. 2). A small fraction of cases exhibited elevated JSD values, indicating discrepancies that could reflect either spurious (hallucinatory) samples or genuine metastable states not explored by reference simulations (fig. S1).

To probe these outliers, we constructed Koopman operator models from long unbiased MD and from TITO trajectories. These models characterize the system's slowest dynamical modes and associated metastable states. We found that TITO generally reproduced relaxation timescales of MD, suggesting that much of the JSD tail arises from minor numerical mismatches. However, a subset of systems displayed substantially slower relaxation times under TITO, as revealed by notably larger VAMP (variational approach to Markov processes) scores (*41*). Because theoretical results show that timescales are bounded from above (*42*), this indicates that TITO sampled metastable states not observed in the MD trajectories.

To evaluate whether these previously unobserved states were physically meaningful, we performed extensive replica exchange (RE) MD simulations (*43*). On average, TITO covered all density regions visited by RE MD, whereas long conventional MD failed to do so in a significant fraction of systems (Fig. 3A). States detected by TITO but absent in long MD were consistently recovered in RE MD (fig. S2), confirming that they correspond to genuine metastable basins. In one representative example, a propiolamide, TITO uncovered a metastable basin absent from long MD but corroborated by both ultralong unbiased MD and RE MD simulations (Fig. 3C and figs. S3 and S4). This metastable basin was not sampled when the model was trained on a training set-length trajectory of this molecule (fig. S5), showing that it emerges from multimolecule training. Although trained only on nanosecond MD data, TITO correctly inferred an exchange timescale between basins on the order of microseconds—consistent with estimates from ultralong trajectories.

Intriguingly, TITO also predicted additional states not observed in either reference method (Fig. 3B). To further investigate these cases, we initialized ensembles of nanosecond-length unbiased MD trajectories from TITO-generated configurations. These simulations exhibited a small but systematic improvement in agreement with RE MD (fig. S6), indicating that TITO samples near-physical configurations capable of relaxing into correct basins under explicit dynamics. We then assessed the stability of these identified states in ensemble simulations and found that the configurations remained metastable (fig. S7), suggesting that they represent physically valid states rather than artifacts of the learned dynamics. Collectively, these findings demonstrate that TITO not only preserves the integrity of the Boltzmann distribution but also uncovers metastable states that would likely remain undetected using conventional methods within practical computational limits.

We further examined whether TITO's generalization correlates with chemical similarity between training and test molecules. Unexpectedly, no such correlation was observed in either of the datasets (figs. S8 and S9 and table S1). This lack of correlation suggests that chemical composition alone provides limited signal for guiding iterative refinement or active learning of generative dynamical models, underscoring the need for alternative strategies to improve generalization.

TITO faithfully reproduces relaxation transients in unseen molecular systems

Next, we investigate whether the dynamics generated by TITO is statistically equivalent to that generated by numerical MD simulations. Given that we here target MD in the NVT ensemble—keeping the number of particles (N), volume (V), and temperature (T) constant during a simulation—the dynamics are stochastic, and consequently, we use statistical tools to quantitatively compare the two approaches.

A stringent test of dynamical fidelity is whether a model can reproduce relaxation processes across systems that differ vastly in their intrinsic timescales. As noted, TITO accurately recapitulates relaxation dynamics in molecules and peptides outside its training set and crucially provides quantitative predictions spanning over three orders of magnitude in characteristic times (fig. S10). Predicted timescales remain faithful with reference to MD and consistent with the variational principle for Markov processes (*42*) across different lag times (fig. S11). This level of agreement indicates that TITO has learned an effective and generalizable representation of the underlying stochastic dynamics rather than merely fitting short-time correlations.

Still, matching timescales alone does not guarantee that the associated motions are physically meaningful. To examine this, we performed extensive TITO simulations and compared the full relaxation transients of slow dynamical modes with those obtained from long, unbiased MD trajectories. Across two orders of magnitude in timescale, the agreement was notable, suggesting that TITO reproduces both the rate and mechanism of the underlying MD (Fig. 4A). This demonstrates that the learned transition operator generalizes dynamically, faithfully capturing the hierarchy of molecular motions that govern relaxation kinetics.

A further requirement is internal consistency across time resolutions. Because TITO predicts time-integrated dynamics at multiple timescales, the resulting transients must be consistent regardless of

Downloaded from https://www.science.org on April 27, 2026

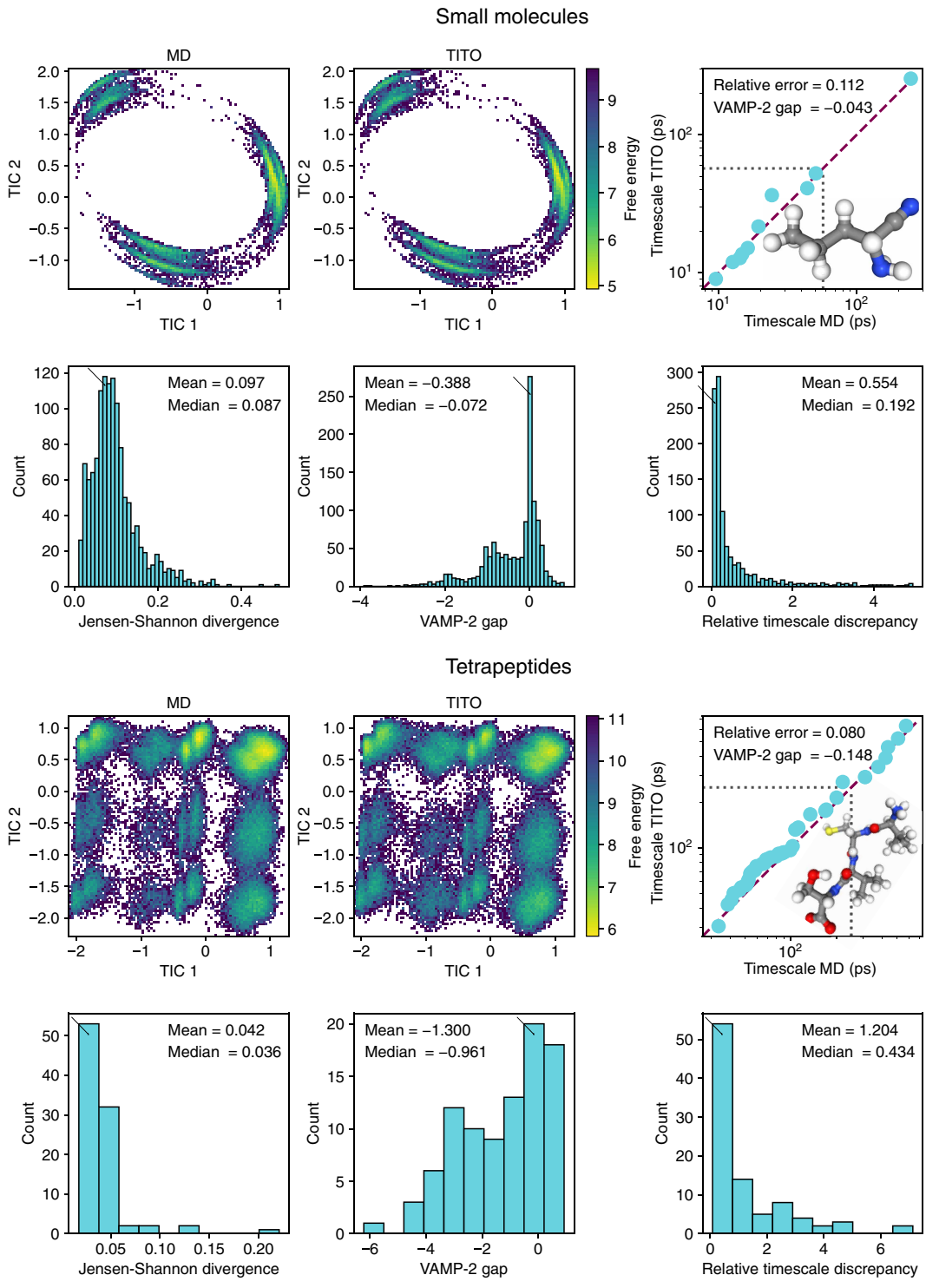

Fig. 2. TITO accurately predicts both thermodynamics and kinetics. Small molecules (top) and tetrapeptides (bottom). Top row: Projection onto the first two time-lagged independent components (TICs) and comparison of VAMP timescales between MD and TITO-generated samples for a representative molecule. Bottom row: Aggregated evaluation across systems: Jensen-Shannon divergence in the TIC subspace (left), VAMP-2 gap (center), and top 10 relative error (right). Black arrows denote the position of the example molecule within each histogram.

Downloaded from https://www.science.org on April 27, 2026

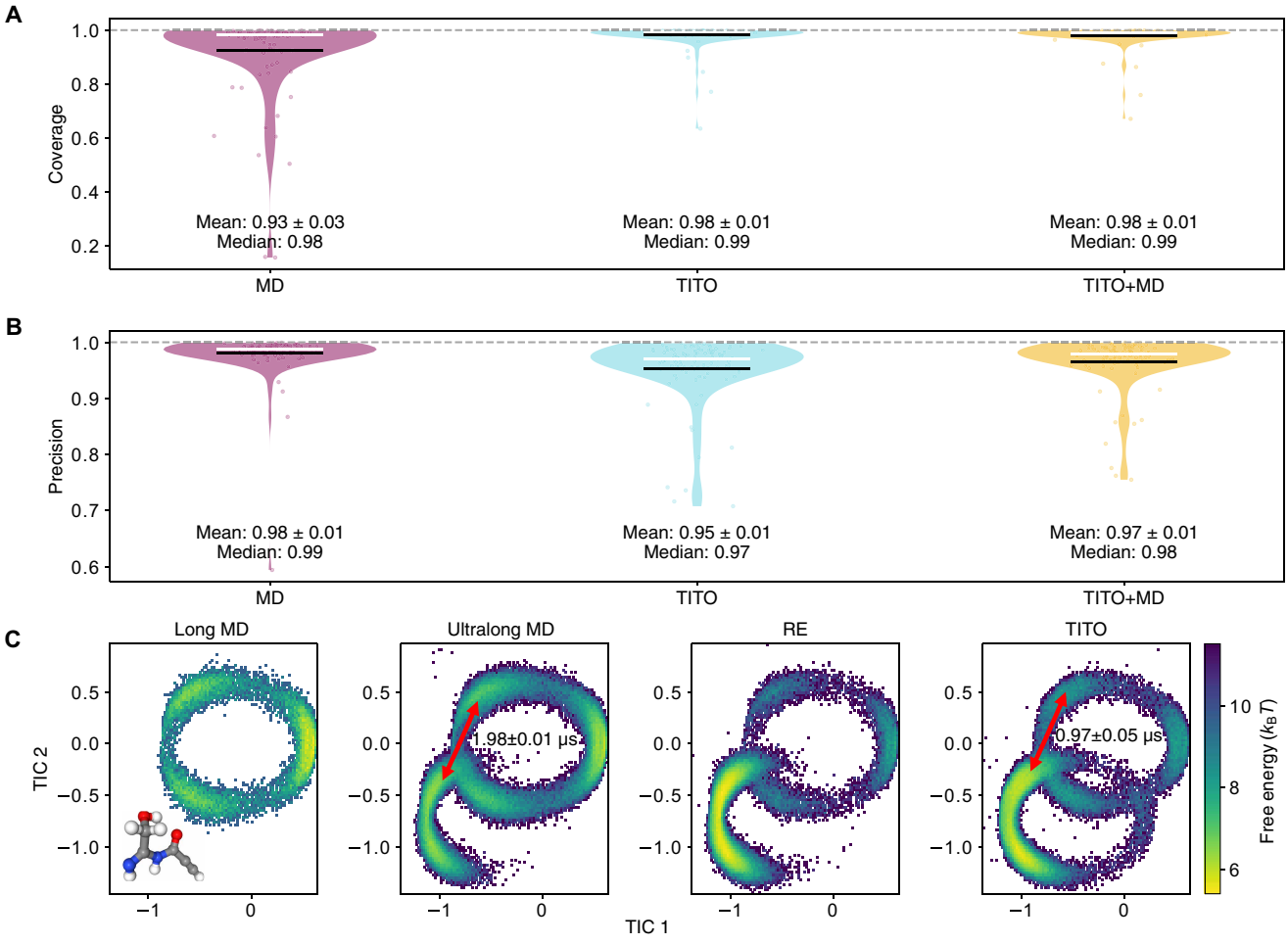

Fig. 3. TITO accurately samples two orders of magnitude slower dynamics than the training data. (**A**) Coverage and (**B**) precision (see the "Evaluation metrics" section) with reference to RE of projections onto the first two TICs for training set–like simulations (MD, 36.5 ns), TITO (32 μs), and an ensemble of ultrashort MD initialized with TITO samples (10 ps per sample). TITO achieves better coverage of the conformational space than MD. (**C**) Example propiolamide test-set molecule: Long MD (36.5 ns) fails to sample the most metastable basin, while TITO (160 μs, initialized from long MD) successfully transitions into the dominant basin, recovers the Boltzmann distribution, and estimates the right order of magnitude for the corresponding VAMP-implied timescale, estimated from ultralong MD (16 ms). Structural insights of this transition and alternative TIC projections are provided in figs. S3 and S4, respectively.

whether they are generated in a single long step or as a sequence of shorter steps (nested sampling). We find that the relaxation transients remain self-consistent under this test (Fig. 4A), suggesting that the learned transition density satisfies the Chapman-Kolmogorov equation and thus encodes genuinely Markovian dynamics. We remark that even if Chapman-Kolmogorov consistency is not hard-coded into the model architecture, it is implicitly promoted by the Implicit Transfer Operator objective. Because the model is trained to learn conditional distributions at multiple lag times, the loss encourages coherence across compositions of the learned dynamics.

Last, we investigated whether the high fidelity of slow dynamics is achieved at the expense of accuracy in fast, rapidly relaxing modes. Despite operating at time steps orders of magnitude larger than those of bond and angle vibrations, TITO accurately reproduced equilibrium properties (Fig. 4B) and generated conformations with potential energies closely matching those from unbiased MD (Fig. 4C). The main deviation we observed was a slight underestimation of the variance of fast modes, which in turn leads to systematically lower potential energies relative to reference simulations.

Qualitative extrapolation to larger peptide systems

We next examined whether TITO can extrapolate beyond the molecular sizes represented in its training data. This setting provides a stringent test of transferability: A model trained on short peptides must infer effective dynamics at unseen length scales, where both the number of atoms and the hierarchy of internal motions increase substantially. Specifically, we applied a model trained only on tetrapeptides to generate trajectories for penta-, hexa-, hepta-, and octapeptides.

Extrapolation introduces a scale mismatch in the latent base distribution p_0, whose variance depends on system size. To mitigate this, we rescaled the standard deviation of p_0 according to Flory's scaling law for the radius of gyration of random polymers, $\langle R_g \rangle \propto N^{0.688}$ (*44*), where N denotes the number of residues. Guided by this simple physical prior, TITO produced configurations with realistic local geometry and global compactness across peptide lengths; without this correction, stable extrapolation beyond pentapeptides was not achievable.

With the scaling correction in place, TITO approximately recovered the conformational landscapes of larger peptides and reproduced relaxation times qualitatively consistent with explicit MD,

Downloaded from https://www.science.org on April 27, 2026

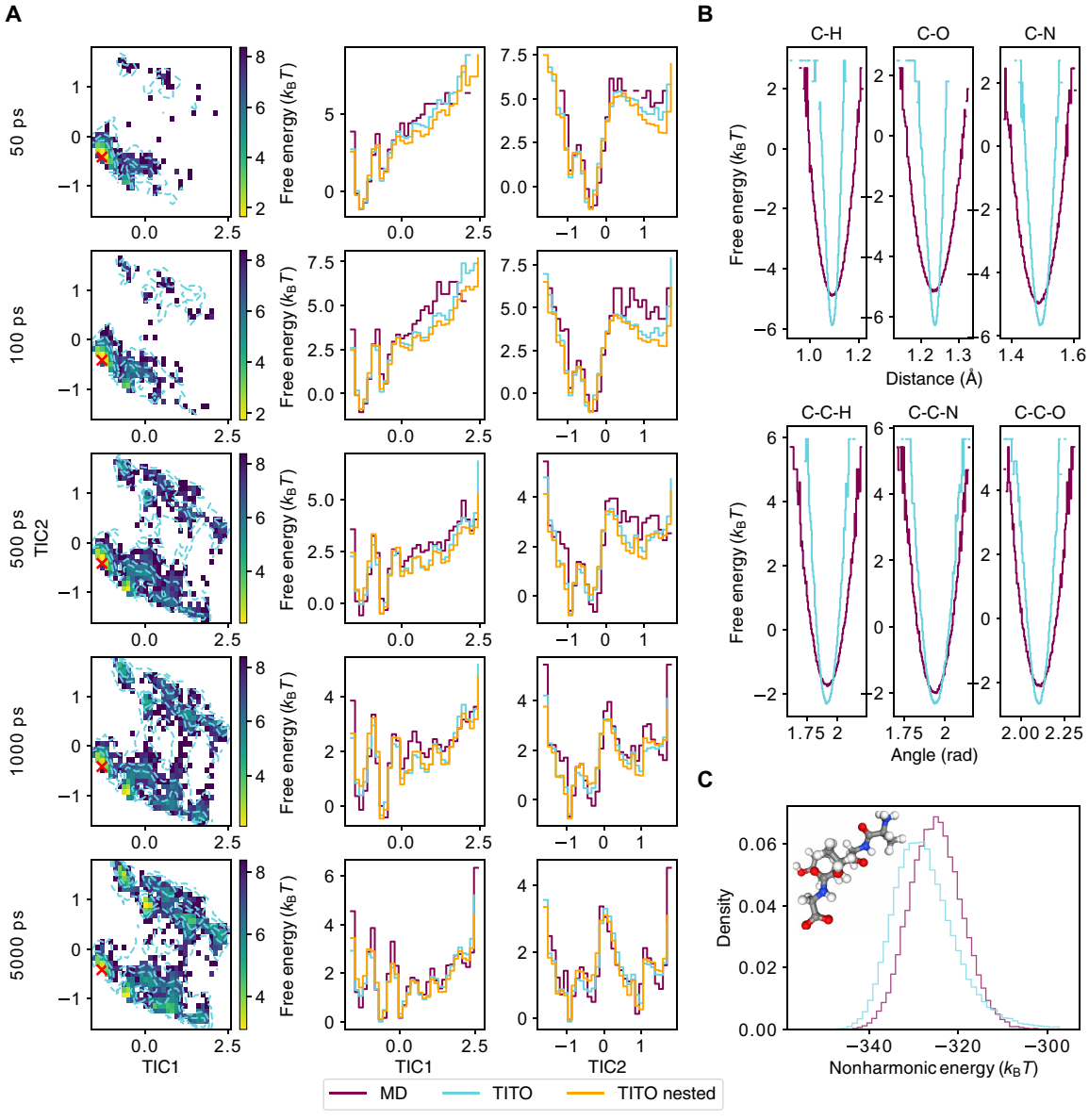

Fig. 4. TITO recapitulates transients, fast vibrations, and potential energies. (**A**) Free energy of the transition probability estimated with TITO for AYTG (test set) at increasing lag times (top to bottom). The first column shows conditional free energies projected onto the first two TICs. Contours represent TITO samples, while two-dimensional histograms correspond to Markov state model estimates from MD data. The red cross marks the simulation's initial state. The second and third columns display marginal distributions along each TIC. Nested samples are generated with five steps. (**B**) Free energy profiles of bond distances (top) and angles (bottom). (**C**) Probability density of the nonharmonic potential energy.

even when the sequence length was doubled relative to the training systems (Fig. 5). For the largest peptides, however, the generated trajectories exhibited mild structural compaction (fig. S12) and a systematic downward drift in potential energy (fig. S13), leading to instability in long nested-sampling runs. These deviations likely arise from cumulative local errors that are amplified at increasing system sizes. Nevertheless, the generated configurations remain physically meaningful and can be readily refined by short low-cost MD equilibrations.

Together, these results demonstrate that TITO captures transferable physical principles sufficient to generalize far beyond its training domain while also delineating the limits of such extrapolation. The observed degradation at large sizes naturally motivates hybrid divide-and-conquer strategies, in which long TITO propagation steps are interleaved with brief MD equilibration phases—akin to hybrid Monte Carlo or multiresolution simulation schemes (*37*, *45*, *46*).

Calibration of simulation accuracy to compute budget

For practical impact, TITO must deliver substantially higher throughput than conventional MD while retaining quantitative accuracy in equilibrium and dynamical properties. Two factors determine the effective simulation throughput: (i) the molecular size, which constrains the number of simulations that can be run in parallel on a GPU with fixed memory, and (ii) the number of ODE solver steps required for each CNF evaluation, which controls the cost per TITO step.

Downloaded from https://www.science.org on April 27, 2026

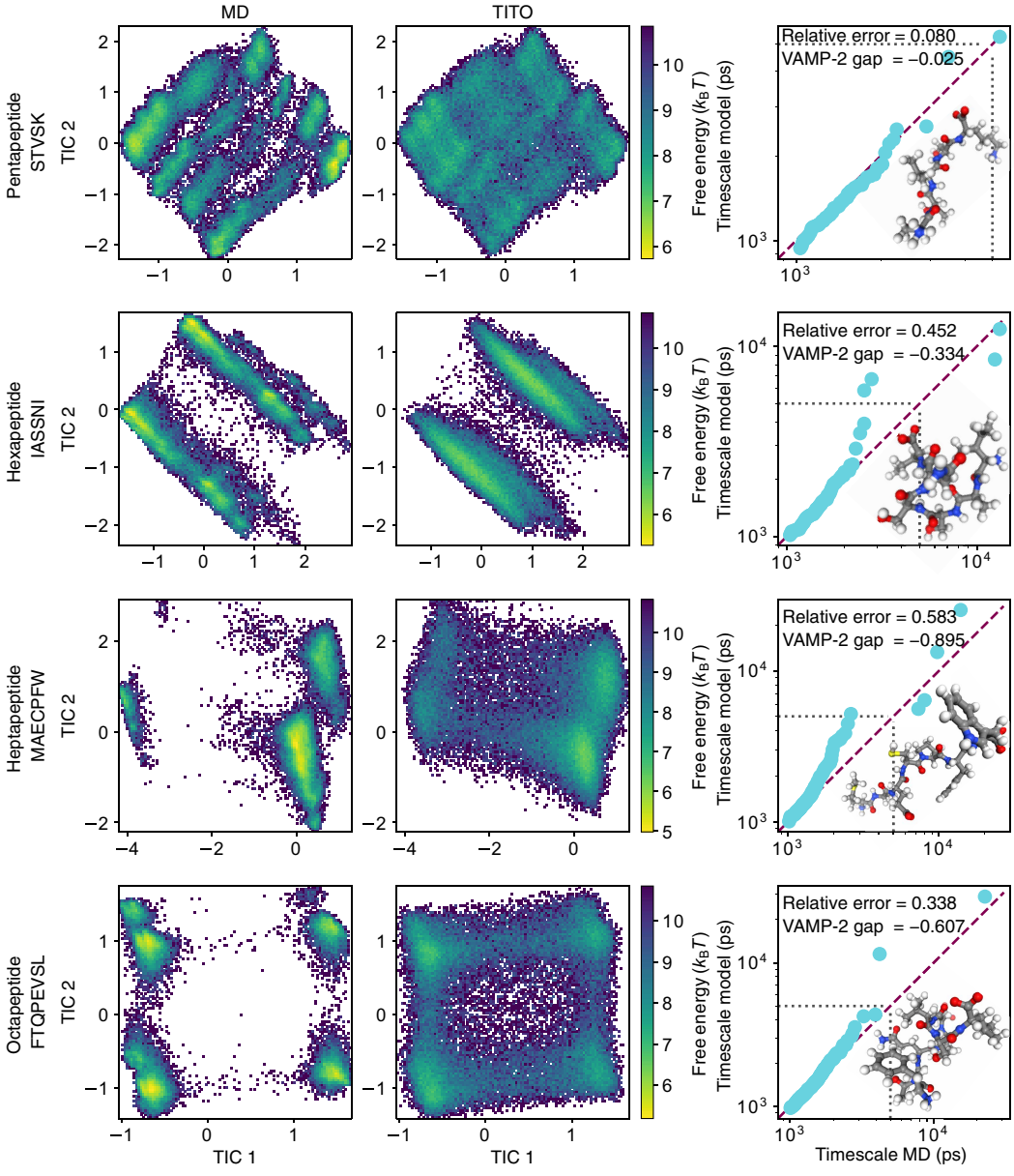

Fig. 5. Extrapolation to larger systems. Free energy landscape and VAMP timescales of a TITO model trained on tetrapeptides and performing 5-ns single-step sampling for penta-, hexa-, hepta-, and octapeptides.

We find that equilibrium properties can be reproduced at comparatively low computational budgets, whereas accurate estimation of relaxation timescales requires additional solver steps and, hence, higher cost (fig. S14). This trade-off implies that the compute budget can be calibrated to match the target application, for example, prioritizing thermodynamic ensemble generation versus reproducing kinetic observables measured in experiment.

To quantify achievable throughput, we report the maximum simulation time reached on a single GPU. As shown in Table 1, TITO enables ~10 ms of physical simulation time per day of computation, representing a four-orders-of-magnitude improvement relative to standard unbiased MD simulations using the same resources. These results suggests that TITO can be tuned flexibly: Users may trade simulation fidelity against throughput depending on the level of accuracy required. Furthermore, we emphasize that these gains can potentially be even larger if the model is trained to predict larger time steps, paving the way to study ultraslow processes in biology and materials science.

DISCUSSION

We introduce TITO, a chemically generalizable generative model that reproduces MD at a fraction of the computational cost of traditional simulations. TITO quantitatively recovers both the equilibrium probabilities of molecular configurations and the rates and mechanisms of conformational exchange across diverse chemistries from

Downloaded from https://www.science.org on April 27, 2026

Table 1. Maximum simulation time throughput. Estimates using one NVIDIA A100 80-gigabyte GPU for 1 day.

Systems	Method	Simulation throughput per day in one GPU
Small molecules	MD	3.5 μs
	TITO	11.3 ms
Tetrapeptides	MD	0.67 μs
	TITO	10.3 ms

small molecules to peptides. In essence, it delivers the fidelity near that of MD at the cost of sampling a deep generative model.

TITO achieves this by learning the statistics of time-integrated dynamics directly from simulation data, allowing propagation over arbitrarily long lag times without explicit numerical integration at a femtosecond resolution. This formulation yields an acceleration of up to four orders of magnitude in the quantitative characterization of equilibrium states and relaxation kinetics at compute cost parity. Furthermore, TITO retains predictive power beyond its training regime, qualitatively reproducing thermodynamic and kinetic behavior in peptides up to twice the size of those used for training, highlighting its ability to extrapolate across molecular size.

TITO differs fundamentally from dominant paradigms in generative models of MD, Boltzmann Generators, which aim to quantitatively sample the independent equilibrium samples from the Boltzmann distribution (*47*), and Boltzmann Emulators, which sacrifice quantitative alignment with MD to boost efficiency and scaling (*37*, *48*, *49*). These methods, and in particular, their transferable variants (*37*, *49*, *50*), are rapidly becoming a viable complement to MD simulations when equilibrium properties are the target of investigation. However, these approaches cannot capture dynamic properties such as rates or mechanisms.

Other methods that aim to predict the dynamics provide mainly qualitative insights and at fixed timescales (*38*, *51*–*54*) or enable much smaller time steps (*55*, *56*). Recent work explored multitimescale training in small proteins (*57*). To our knowledge, TITO is the first framework to achieve physically realistic, multitimescale sampling with demonstrated transferability across both chemical composition and molecular size. Other complementary strategies, including machine learning–infused path-sampling (*58*) strategies or latent space simulators (*56*, *59*, *60*), show promise in scaling to larger systems, but generalization remains an open challenge requiring careful modeling and calibration for every specific process of interest.

Machine-learned interatomic potentials (*61*, *62*) and coarse-grained force fields (*63*, *64*) have similarly shown impressive strides toward general purpose transferability. These models can guarantee realistic physical dynamics depending on the integration strategy chosen. So, while they might boost the accuracy over current force fields and coarse-grained models, they still rely on iterative numerical integration with tiny time steps, making their computational footprint substantial. TITO instead offers a paradigm shift: bypassing iterative integration together.

Despite these advances, important limitations remain. At present, TITO is restricted to implicit solvent representations and system sizes of at most a few hundred atoms. Extending the method to explicitly solvated biomolecules with tens to hundreds of thousands of degrees of freedom will require innovations in neural architectures (*65*–*68*), such as equivariant transformers (*69*), linear attention (*70*) or Euclidean fast attention (*65*), and/or hierarchical strategies such as coarse graining (*71*). In addition, periodic boundary conditions—essential for realistic modeling of solvated systems—are not yet supported. While our experiments show promising extrapolation to larger and chemically distinct molecules, generalization performance still depends on the chemical similarity between target and training systems. Achieving broad chemical coverage will necessitate larger and more diverse training datasets. Last, TITO is presently limited to a single thermodynamic state (NVT ensemble at room temperature). Extending thermodynamic transferability, including across temperatures, pressures, or even force fields, would enable the study of the influence of thermodynamic perturbations on stationary and dynamic properties of molecular systems. Promising prior work has demonstrated that thermodynamic transferability can be achieved by additional conditioning variables through learned embeddings of the thermodynamic state (*72*, *73*), which motivates its applicability to kinetic modeling as well.

Intriguingly, we find that TITO's generalization performance shows no clear relationship to the chemical similarity between training and test systems. This observation challenges the prevailing assumption that broader chemical coverage alone ensures generalization. Instead, it suggests that the structure and diversity of training data—how well they represent relevant dynamical motifs and energy landscapes—may be more critical than sheer data volume. In this view, progress may hinge less on scaling to ever-larger simulation datasets and more on carefully curated, mechanistically diverse benchmarks that capture the essential physics of MD.

In summary, TITO establishes a previously unidentified paradigm for transferable generative modeling of MD, unifying thermodynamic sampling and dynamical prediction in a singular generative surrogate. By enabling accelerated and chemically transferable estimation of stationary and dynamic properties—such as free energies and rates—TITO paves the way toward practical deep learning–based acceleration of molecular simulations.

MATERIALS AND METHODS

Data

We use three datasets covering different regions of the chemical space:

1) Small molecules: The MDQM9-nc dataset (*37*) contains MD simulations for 12,530 small noncyclic molecules from the QM9 dataset (*74*). Simulations are performed in a vacuum at room temperature using the generalized Amber force field (*75*). The simulation time is dependent on the molecule size with a median sampling time of 36.5 ns. We perform extra 1-μs RE simulations across eight temperatures (300, 400, 500, 600, 700, 800, 900, and 1000 K). The average exchange rate is 58%.

2) Tetrapeptides: The Timewarp dataset (*38*) contains MD simulations for tetrapeptides. It contains two tetrapeptide subdatasets,

Downloaded from https://www.science.org on April 27, 2026

large, which contains 1457 training molecules, and huge, with 92 larger molecules. We use large as training data and huge as the test set. The simulations are performed in implicit water and at room temperature. Simulation times are 50 ns for training set molecules and 1 μs for test set molecules.

3) Larger systems: We performed 1-μs simulation of penta-, hexa-, hepta-, and octapeptides with the same simulation parameters as in the Timewarp dataset. For each peptide length, six sequences were randomly sampled on the basis of vertebrate amino acid frequencies.

Model

TITO uses equivariant optimal transport flow matching to parameterize the transition probability. Flow matching (*34*) provides an efficient framework for training CNFs by aligning a learnable velocity field with the optimal transport velocity field between an easy-to-sample base distribution p_0 and a target distribution p_1. Rather than directly minimizing a divergence between the generated distribution and p_1, flow matching constructs intermediate states along an interpolation between p_0 and p_1. The neural ODE vector field is then trained to predict the conditional displacement between these paired states. We use linear interpolants (*76*)

$$\mathbf{x}^T = (1-T)\mathbf{x}^0 + T\mathbf{x}^1, \;\; T \in [0,1] \tag{3}$$

with target velocity

$$\mathbf{v}^T = \frac{d\mathbf{x}^T}{dT} = \mathbf{x}^1 - \mathbf{x}^0 \tag{4}$$

Molecular distributions live in a space with inherent symmetries, such as rotational and permutational invariances of atomic coordinates. Equivariant optimal transport (*35*) incorporates these symmetries to construct shorter paths by aligning sample pairs, $\mathbf{x}^0$ and $\mathbf{x}^1$, along their symmetry orbits. In practice, this minimization is approximated sequentially: optimal permutation by solving a linear sum assignment problem (*77*), followed by optimal superpositioning through solving a Procrustes problem (*78*).

For the velocity field model, $\mathbf{v}_\theta$, we use a modified SE3-ITO architecture (*33*) enriched with edge features encoding interaction types between atoms. Specifically, we distinguish single, double, triple, and through-space bonds or interactions. As in SE3-ITO, we assume a complete interaction graph where every atom interacts with all others, with bonded interactions prioritized according to the order listed above. Model training and inference parameters for different experiments are included in tables S2 and S3, respectively.

Evaluation metrics

Jensen-Shannon divergence (JSD)

The Jensen-Shannon divergence provides a symmetric measure of similarity between two probability distributions. Given two distributions p and q, the JSD is defined as

$$\mathrm{JSD}(p\|q) = \frac{1}{2}D_{\mathrm{KL}}(p\|m) + \frac{1}{2}D_{\mathrm{KL}}(q\|m) \tag{5}$$

where $m = \frac{1}{2}(p+q)$ and D_{KL} denotes the Kullback-Leibler divergence. Unlike the KL divergence, the JSD is bounded between 0 and 1.

Free energy

The free energy (negative log-likelihood) of a subspace of the conformational space $\Omega_i \in \Omega$ is

$$F(\Omega_i) = -k_B T \log[p(\Omega_i)] \;\text{ with }\; p(\Omega_i) = \int_{\Omega_i} p(\mathbf{x})\,d\mathbf{x} \tag{6}$$

Coverage and precision

Coverage and precision quantify the degree of overlap between the probability distributions generated by two sampling methods. Given two methods, m_1 and m_2, the coverage of m_1 with respect to m_2 is defined as

$$\mathrm{COV}_{m_1,m_2} = \int_{\Omega_{p_{m_1}\cap p_{m_2}}} p_{m_2}(\mathbf{x})\,d\mathbf{x} \tag{7}$$

where p_{m_i} denotes the probability mass sampled from method m_i, and $\Omega_{p_{m_1}\cap p_{m_2}} = \{\mathbf{x}: p_{m_1}(\mathbf{x}) > \delta \text{ and } p_{m_2}(\mathbf{x}) > \delta\}$. Precision is defined analogously as

$$\mathrm{PRE}_{m_1,m_2} = \int_{\Omega_{p_{m_1}\cap p_{m_2}}} p_{m_1}(\mathbf{x})\,d\mathbf{x} \tag{8}$$

Intuitively, coverage measures how much of the probability mass of m_2 is captured by m_1, while precision measures the fraction of m_1's probability mass supported by m_2. In practice, we estimate these quantities from empirical histograms, defining $\Omega_{p_{m_1}\cap p_{m_2}}$ as the set of discrete states where both methods have nonzero counts.

Variational approach for Markov processes (VAMP)

The VAMP framework provides a principled method for evaluating the quality of dynamical models on the basis of the variational principle of conformation dynamics. We used the following VAMP-based metrics:

Implied timescales. The eigenvalues of the estimated transfer operator were used to compute implied timescales, defined as

$$t_i = -\frac{\tau}{\ln\sigma_i} \tag{9}$$

where σ_i is the ith singular value of the Koopman operator approximation, and τ is the lag time.

Relative timescale discrepancy. We define the relative timescale discrepancy as

$$\epsilon_{\mathrm{rel}} = \frac{1}{N}\sum_i^N \frac{\left|t_i^{\mathrm{MD}} - t_i^{\mathrm{TITO}}\right|}{t_i^{\mathrm{MD}}} \tag{10}$$

where t_i^{MD} and t_i^{TITO} are implied timescales predicted with MD and TITO, respectively, and are sorted in decreasing order. We use $N = 10$ implied timescales throughout this work.

VAMP-2 score. The VAMP-2 score is the squared Frobenius norm of the singular value spectrum

$$\mathrm{VAMP\text{-}2}^{(k)} = \sum_{i=1}^{k} \sigma_i^2 \tag{11}$$

Higher scores indicate that the model captures slow dynamical modes.

VAMP-gap. We define the VAMP-gap as the difference in VAMP-2 scores between TITO and the MD

$$\mathrm{VAMP\text{-}gap} = \mathrm{VAMP\text{-}2\text{-}score}_{\mathrm{MD}} - \mathrm{VAMP\text{-}2\text{-}score}_{\mathrm{TITO}} \tag{12}$$

Negative VAMP-gaps indicate that TITO predicts slower dynamics and vice versa. Together, these evaluation metrics provide complementary insights: The JSD and free energy excess measure how well the model reproduces equilibrium distributions, while VAMP metrics assess the model's fidelity in capturing slow dynamical processes and metastability.

Downloaded from https://www.science.org on April 27, 2026

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S14
Tables S1 to S3

REFERENCES

1. A. G. Palmer III, NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 129–155 (2001).
2. B. Schuler, Single-molecule FRET of protein structure and dynamics—A primer. *J. Nanobiotechnol.* **11** (Suppl. 1), S2 (2013).
3. B. Leimkuhler, C. Matthews, *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods* (Springer International Publishing, 2015), 10.1007/978-3-319-16375-8.
4. M. Karplus, J. A. McCammon, Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).
5. C. Schütte, F. Noe, E. Meerbach, P. Metzner, C. Hartmann, "Conformation dynamics," in *Proceedings of the 6th International Congress on Industrial and Applied Mathematics* (International Council for Industrial and Applied Mathematics, 2009), pp. 297–336.
6. J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, L. Delemotte, Enhanced sampling methods for molecular dynamics simulations [Article v1.0]. *Living J. Comput. Mol. Sci* **4**, 1583 (2022).
7. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566 (2002).
8. G. Torrie, J. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
9. H. Grubmüller, Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E* **52**, 2893–2906 (1995).
10. D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).
11. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
12. P. Tiwary, B. J. Berne, Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2839–2844 (2016).
13. W. Chen, A. L. Ferguson, Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **39**, 2079–2102 (2018).
14. A. F. Voter, Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **78**, 3908–3911 (1997).
15. M. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).
16. K. A. Feenstra, B. Hess, H. J. C. Berendsen, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).
17. B. Leimkuhler, C. Matthews, Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Chem. Phys.* **138**, 174102 (2013).
18. D. E. Shaw, P. J. Adams, A. Azaria, J. A. Bank, B. Batson, A. Bell, M. Bergdorf, J. Bhatt, J. A. Butts, T. Correia, R. M. Dirks, R. O. Dror, M. P. Eastwood, B. Edwards, A. Even, P. Feldmann, M. Fenn, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, M. Gorlatova, B. Greskamp, J. Grossman, J. Gullingsrud, A. Harper, W. Hasenplaugh, M. Heily, B. C. Heshmat, J. Hunt, D. J. Ierardi, L. Iserovich, B. L. Jackson, N. P. Johnson, M. M. Kirk, J. L. Klepeis, J. S. Kuskin, K. M. Mackenzie, R. J. Mader, R. McGowen, A. McLaughlin, M. A. Moraes, M. H. Nasr, L. J. Nociolo, L. O'Donnell, A. Parker, J. L. Peticolas, G. Pocina, C. Predescu, T. Quan, J. K. Salmon, C. Schwink, K. S. Shim, N. Siddique, J. Spengler, T. Szalay, R. Tabladillo, R. Tartler, A. G. Taube, M. Theobald, B. Towles, W. Vick, S. C. Wang, M. Wazlowski, M. J. Weingarten, J. M. Williams, K. A. Yuh, "Anton 3: Twenty microseconds of molecular dynamics simulation before lunch," in *SC21: International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE, 2021).
19. I. Ohmura, G. Morimoto, Y. Ohno, A. Hasegawa, M. Taiji, MDGRAPE-4: A special-purpose computer system for molecular dynamics simulations. *Philos. Trans. A Math. Phys. Eng. Sci.* **372**, 20130387 (2014).
20. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341–346 (2010).
21. M. Shirts, V. S. Pande, Screen savers of the world unite! *Science* **290**, 1903–1904 (2000).
22. P. Eastman, R. Galvelis, R. P. Peláez, C. R. A. Abreu, S. E. Farr, E. Gallicchio, A. Gorenko, M. M. Henry, F. Hu, J. Huang, A. Krämer, J. Michel, J. A. Mitchell, V. S. Pande, J. P. Rodrigues, J. Rodriguez-Guerra, A. C. Simmonett, S. Singh, J. Swails, P. Turner, Y. Wang, I. Zhang, J. D. Chodera, G. De Fabritiis, T. E. Markland, OpenMM 8: Molecular dynamics simulation with machine learning potentials. *J. Phys. Chem. B* **128**, 109–116 (2023).
23. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).
24. M. J. Harvey, G. Giupponi, G. D. Fabritiis, ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).
25. D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan, K. M. Merz, AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191 (2023).
26. W. C. Swope, J. W. Pitera, F. Suits, Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* **108**, 6571–6581 (2004).
27. N.-V. Buchete, G. Hummer, Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **112**, 6057–6069 (2008).
28. J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
29. G. R. Bowman, V. S. Pande, F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (Springer, 2014), 10.1007/978-94-007-7606-7.
30. S. Olsson, Generative molecular dynamics. *Curr. Opin. Struct. Biol.* **96**, 103213 (2026).
31. P. Langevin, Sur la théorie du mouvement brownien. *C. R. Acad. Sci.* **146**, 530–533 (1908).
32. H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer Series in Synergetics (Springer Berlin Heidelberg, ed. 2, 1996), http://dx.doi.org/10.1007/978-3-642-61544-3.
33. M. Schreiner, O. Winther, S. Olsson, "Implicit transfer operator learning: Multiple time-resolution models for molecular dynamics," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine, Eds. (Curran Associates Inc., 2023), vol. 36, pp. 36449–36462, https://proceedings.neurips.cc/paper_files/paper/2023/file/7274ed909a312d4d869cc328ad1c5f04-Paper-Conference.pdf.
34. Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, M. Le, "Flow matching for generative modeling," in *The Eleventh International Conference on Learning Representations* (International Conference on Learning Representations, 2023), https://openreview.net/forum?id=PqvMRDCJT9t.
35. L. Klein, A. Krämer, F. Noe, "Equivariant flow matching," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine, Eds. (Curran Associates Inc., 2023), vol. 36, pp. 59886–59910, https://proceedings.neurips.cc/paper_files/paper/2023/file/bc827452450356f9f558f4e4568d553b-Paper-Conference.pdf.
36. R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett, Eds. (Curran Associates Inc., 2018), vol. 31, https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.
37. J. Viguera Diez, S. Romeo Atance, O. Engkvist, S. Olsson, Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics. *Mach. Learn. Sci. Technol.* **5**, 025010 (2024).
38. L. Klein, A. Foong, T. Fjelde, B. Mlodozeniec, M. Brockschmidt, S. Nowozin, F. Noe, R. Tomioka, "Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine, Eds. (Curran Associates Inc., 2023), vol. 36, pp. 52863–52883, https://proceedings.neurips.cc/paper_files/paper/2023/file/a598c367280f9054434fdcc227ce4d38-Paper-Conference.pdf.
39. J. V. Diez, M. J. Schreiner, O. Engkvist, S. Olsson, "Boltzmann priors for Implicit Transfer Operators," in *The Thirteenth International Conference on Learning Representations* (International Conference on Learning Representations, 2025), https://openreview.net/forum?id=pRCOZIlZdT.
40. J. Lin, Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).
41. H. Wu, F. Noé, Variational approach for learning Markov processes from time series data. *J. Nonlinear Sci.* **30**, 23–66 (2020).
42. F. Nüske, B. G. Keller, G. Pérez-Hernández, A. S. J. S. Mey, F. Noé, Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **10**, 1739–1752 (2014).
43. D. J. Earl, M. W. Deem, Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **7**, 3910–3916 (2005).
44. P. Flory, *Principles of Polymer Chemistry* (Cornell Univ. Press, 1953), https://books.google.se/books?id=CQ0EbEkT5R0C.
45. J. P. Nilmeier, G. E. Crooks, D. D. L. Minh, J. D. Chodera, Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1009–E1018 (2011).
46. G. A. Ross, E. Russell, Y. Deng, C. Lu, E. D. Harder, R. Abel, L. Wang, Enhancing water sampling in free energy calculations with grand canonical Monte Carlo. *J. Chem. Theory Comput.* **16**, 6061–6076 (2020).
47. F. Noé, S. Olsson, J. Köhler, H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019).

Downloaded from https://www.science.org on April 27, 2026

48. B. Jing, G. Corso, J. Chang, R. Barzilay, T. Jaakkola, "Torsional diffusion for molecular conformer generation," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, A. Oh, Eds. (Curran Associates Inc., 2022), vol. 35, pp. 24240–24253, https://proceedings.neurips.cc/paper_files/paper/2022/file/994545b2308bbbbc97e3e687ea9e464f-Paper-Conference.pdf.
49. S. Lewis, T. Hempel, J. Jiménez-Luna, M. Gastegger, Y. Xie, A. Y. K. Foong, V. G. Satorras, O. Abdin, B. S. Veeling, I. Zaporozhets, Y. Chen, S. Yang, A. E. Foster, A. Schneuing, J. Nigam, F. Barbero, V. Stimper, A. Campbell, J. Yim, M. Lienen, Y. Shi, S. Zheng, H. Schulz, U. Munir, R. Sordillo, R. Tomioka, C. Clementi, F. Noé, Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* **389**, eadv9817 (2025).
50. L. Klein, F. Noé, "Transferable Boltzmann generators," in *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, C. Zhang, Eds. (Curran Associates Inc., 2024), vol. 37, pp. 45281–45314, https://proceedings.neurips.cc/paper_files/paper/2024/file/5035a409f5798e188079e236f437e522-Paper-Conference.pdf.
51. B. Jing, H. Stark, T. Jaakkola, B. Berger, "Generative modeling of molecular dynamics trajectories," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems* (Neural Information Processing Systems, 2024), https://openreview.net/forum?id=yRRCH1OsGW.
52. F. L. Thiemann, T. Reschützegger, M. Esposito, T. Taddese, J. D. Olarte-Plata, F. Martelli, Force-free molecular dynamics through autoregressive equivariant networks. arXiv:2503.23794 [physics.comp-ph] (2025).
53. A. dos Santos Costa, I. Mitnikov, F. Pellegrini, A. Daigavane, M. Geiger, Z. Cao, K. Kreis, T. Smidt, E. Kucukbenli, J. Jacobson, EquiJump: Protein dynamics simulation via SO(3)-equivariant stochastic interpolants. arXiv:2410.09667 [cs.LG] (2024).
54. M. H. Murtada, Z. F. Brotzakis, M. Vendruscolo, MD-LLM-1: A large language model for molecular dynamics. arXiv:2508.03709 [q-bio.BM] (2025).
55. F. Bigi, S. Chong, A. Kristiadi, M. Ceriotti, "FlashMD: Long-stride, universal prediction of molecular dynamics," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems* (Neural Information Processing Systems, 2025), https://openreview.net/forum?id=ogZu06NgQs.
56. P. R. Vlachas, J. Zavadlav, M. Praprotnik, P. Koumoutsakos, Accelerated simulations of molecular systems through learning of effective dynamics. *J. Chem. Theory Comput.* **18**, 538–549 (2021).
57. A. dos Santos Costa, M. Ponnapati, D. Rubin, T. Smidt, J. Jacobson, Accelerating protein molecular dynamics simulation with DeepJump. arXiv:2509.13294 [q-bio.BM] (2025).
58. H. Jung, R. Covino, A. Arjun, C. Leitold, C. Dellago, P. G. Bolhuis, G. Hummer, Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nat. Comput. Sci.* **3**, 334–345 (2023).
59. H. Sidky, W. Chen, A. L. Ferguson, Molecular latent space simulators. *Chem. Sci.* **11**, 9459–9467 (2020).
60. D. Wang, Y. Wang, L. Evans, P. Tiwary, From latent dynamics to meaningful representations. *J. Chem. Theory Comput.* **20**, 3503–3513 (2024).
61. A. Kabylda, J. T. Frank, S. Suárez-Dou, A. Khabibrakhmanov, L. Medrano Sandonas, O. T. Unke, S. Chmiela, K.-R. Müller, A. Tkatchenko, Molecular simulations with a pretrained neural network and universal pairwise force fields. *J. Am. Chem. Soc.* **147**, 33723–33734 (2025).
62. I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Cărare, M. Ceriotti, S. Chong, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, E. Fako, F. Falcioni, A. C. Ferrari, J. L. A. Gardner, M. J. Gawkowski, A. Genreith-Schriever, J. George, R. E. A. Goodall, J. Grandel, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. H. Ho, S. Hofmann, C. Holm, J. Jaafar, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, P. Kourtis, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, C. Lin, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. A. M. Rosset, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, C. Sutton, V. Svahn, T. D. Swinburne, J. Tilly, C. van der Oord, S. Vargas, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, T. Wolf, F. Zills, G. Csányi, A foundation model for atomistic materials chemistry. *J. Chem. Phys.* **163**, 184110 (2025).
63. N. E. Charron, K. Bonneau, A. S. Pasos-Trejo, A. Guljas, Y. Chen, F. Musil, J. Venturin, D. Gusew, I. Zaporozhets, A. Krämer, C. Templeton, A. Kelkar, A. E. P. Durumeric, S. Olsson, A. Pérez, M. Majewski, B. E. Husic, A. Patel, G. De Fabritiis, F. Noé, C. Clementi, Navigating protein landscapes with a machine-learned transferable coarse-grained model. *Nat. Chem.* **17**, 1284–1292 (2025).
64. M. Majewski, A. Pérez, P. Thölke, S. Doerr, N. E. Charron, T. Giorgino, B. E. Husic, C. Clementi, F. Noé, G. De Fabritiis, Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **14**, 5739 (2023).
65. J. T. Frank, S. Chmiela, K.-R. Müller, O. T. Unke, Euclidean fast attention: Machine learning global atomic representations at linear cost. arXiv:2412.08541 [cs.LG] (2024).
66. R. Irwin, A. Tibo, J. P. Janet, S. Olsson, SemlaFlow—Efficient 3D molecular generation with latent attention and equivariant flow matching. arXiv:2406.07266 [cs.LG] (2025).
67. A. Caruso, J. Venturin, L. Giambagli, E. Rolando, F. Noé, C. Clementi, Extending the RANGE of graph neural networks: Relaying attention nodes for global encoding. arXiv:2502.13797 [physics.comp-ph] (2025).
68. E. Qu, A. S. Krishnapriyan, "The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains," in *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, C. Zhang, Eds. (Curran Associates Inc., 2024), vol. 37, pp. 139030–139053, https://proceedings.neurips.cc/paper_files/paper/2024/file/fad8e1915f66161581bb127ccf01092e-Paper-Conference.pdf.
69. F. Fuchs, D. Worrall, V. Fischer, M. Welling, "SE(3)-transformers: 3D Roto-translation equivariant attention networks," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin, Eds. (Curran Associates Inc., 2020), vol. 33, pp. 1970–1981, https://proceedings.neurips.cc/paper_files/paper/2020/file/15231a7ce4ba789d13b722cc5c955834-Paper.pdf.
70. A. Katharopoulos, A. Vyas, N. Pappas, F. Fleuret, "Transformers are RNNs: Fast autoregressive transformers with linear attention," in *Proceedings of the 37th International Conference on Machine Learning*, H. Daumé III, A. Singh, Eds. (PMLR, 2020), vol. 119, pp. 5156–5165, https://proceedings.mlr.press/v119/katharopoulos20a.html.
71. J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. de Fabritiis, F. Noé, C. Clementi, Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci.* **5**, 755–767 (2019).
72. S. Moqvist, W. Chen, M. Schreiner, F. Nüske, S. Olsson, Thermodynamic interpolation: A generative approach to molecular thermodynamics and kinetics. *J. Chem. Theory Comput.* **21**, 2535–2545 (2025).
73. L. Herron, K. Mondal, J. S. Schneekloth, P. Tiwary, Inferring phase transitions and critical exponents from limited observations with thermodynamic maps. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2321971121 (2024).
74. M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
75. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general Amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
76. M. Albergo, N. M. Boffi, E. Vanden-Eijnden, Stochastic interpolants: A unifying framework for flows and diffusions. *J. Mach. Learn. Res.* **26**, 1–80 (2025).
77. D. F. Crouse, On implementing 2D rectangular assignment algorithms. *IEEE Trans Aerosp Electron Syst* **52**, 1679–1696 (2016).
78. W. Kabsch, A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923 (1976).

Acknowledgments

Funding: This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation (WASP AI-MLX Professorship to S.O.). Model training and inference were made possible by an allocation on the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre hosted by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) (project: Berzelius-2025-189, to S.O.), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. S.O. acknowledges funding from the Chalmers Academic Excellence Program. **Author contributions:** S.O.: Writing—original draft, conceptualization, writing—review and editing, methodology, resources, funding acquisition, data curation, validation, supervision, software, project administration, and visualization. M.S.: Conceptualization, investigation, methodology, data curation, validation, formal analysis, and software. J.V.D.: Writing—original draft, conceptualization, investigation, writing—review and editing, methodology, funding acquisition, data curation, validation, supervision, formal analysis, software, project administration, and visualization. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. Code is available on the public repository: https://github.com/olsson-group/tito and https://zenodo.org/records/19187907. The repository also includes instructions to retrieve data used to train the model and model weights used to generate the figures presented in the manuscript. This study did not generate new materials.

Submitted 20 October 2025
Accepted 9 March 2026
Published 8 April 2026
10.1126/sciadv.aed2333

Downloaded from https://www.science.org on April 27, 2026

ScienceAdvances

Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics

Juan Viguera Diez, Mathias Schreiner, and Simon Olsson

Sci. Adv. **12** (15), eaed2333. DOI: 10.1126/sciadv.aed2333

View the article online
https://www.science.org/doi/10.1126/sciadv.aed2333
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org on April 27, 2026

Use of this article is subject to the Terms of service

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).