Accelerated Simulations of Molecular Systems through Learning of Effective Dynamics

Pantelis R. Vlachas, Julija Zavadlav, Matej Praprotnik, and Petros Koumoutsakos*

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations | 🆂🆘 Supporting Information

ABSTRACT: Simulations are vital for understanding and predicting the evolution of complex molecular systems. However, despite advances in algorithms and special purpose hardware, accessing the time scales necessary to capture the structural evolution of biomolecules remains a daunting task. In this work, we present a novel framework to advance simulation time scales by up to 3 orders of magnitude by learning the effective dynamics (LED) of molecular systems. LED augments the equation-free methodology by employing a probabilistic mapping between coarse and fine scales using mixture density network (MDN) autoencoders and evolves the non-Markovian latent dynamics using long short-term memory MDNs. We demonstrate the effectiveness of LED in the Müller−Brown potential, the Trp cage protein, and the alanine dipeptide. LED identifies explainable reduced-order representations, i.e., collective variables, and can generate, at any instant, all-atom molecular trajectories consistent with the collective variables. We believe that the proposed framework provides a dramatic increase to simulation capabilities and opens new horizons for the effective modeling of complex molecular systems.
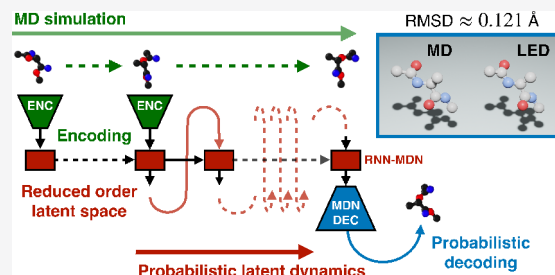
1. INTRODUCTION

Over the last 30 years, molecular dynamics (MD) simulations of biological macromolecules have advanced our understanding of their structure and function.[1] Today MD simulations have become an essential tool for scientific discovery in the fields of biology, chemistry, and medicine. However, they remain hampered by their limited access to time scales of biological relevance for protein-folding pathways, conformational dynamics, and rare-event kinetics.

In order to resolve this bottleneck, two complementary approaches have been pursued. First efforts centered around innovative hardware solutions started with crowd sourcing for compute cycles[2] and have more recently received a boost with the Anton machine[3] enabling remarkable, millisecond-long simulations of biomolecules. Complementary algorithmic efforts aim to advance time scales by systematic coarse graining of the system dynamics. One of the first such studies used the principal component or normal-mode analysis to simulate the conformational changes in proteins.[4−8] Several coarse-graining (CG) methods reduce the complexity of molecular systems by modeling several atoms as a single particle.[9−11] Backmapping techniques[12−14] can be subsequently utilized to recover the atomistic degrees of freedom from a CG representation. Multiscale approaches combine the atomistic and coarse-grained/continuum models[15−17] to augment the accessible time scales, while significant efforts have focused on enhanced sampling techniques.[18−24] Several of these methods exploit the fact that coarse kinetic dynamics on the molecular level are often governed by a few slow collective variables (CVs) (also termed reaction coordinates)[25−28] or by transitions between a few long-lived metastable states.[29,30]

The CVs are typically specified a priori, and their choice crucially impacts the performance and success of the respective sampling methods. Similar to the CG models, the CVs provide a low-order representation of the molecular system, albeit without a particle representation. CVs are of much lower dimensionality than CG models, and retrieving atomistic configurations from CVs is a more challenging problem. While many research efforts have addressed the fine to coarse mapping in CG models, the literature is still scarce on methods to retrieve atomistic configurations from CVs.

Machine-learning (ML) methods,[31,32] exploiting the expressive power of deep networks and their scalability to large data sets, have been used to alleviate the computational burden associated with the simulation of proteins, leading to profound scientific discoveries.[33−35]

The pioneering work in ref 36 utilized neural networks to learn an approximate potential energy surface of density functional theory (DFT) in bulk silicon from quantum mechanical calculations, performing MD simulations with this approximate potential and accelerating the DFT
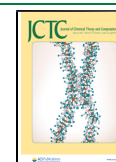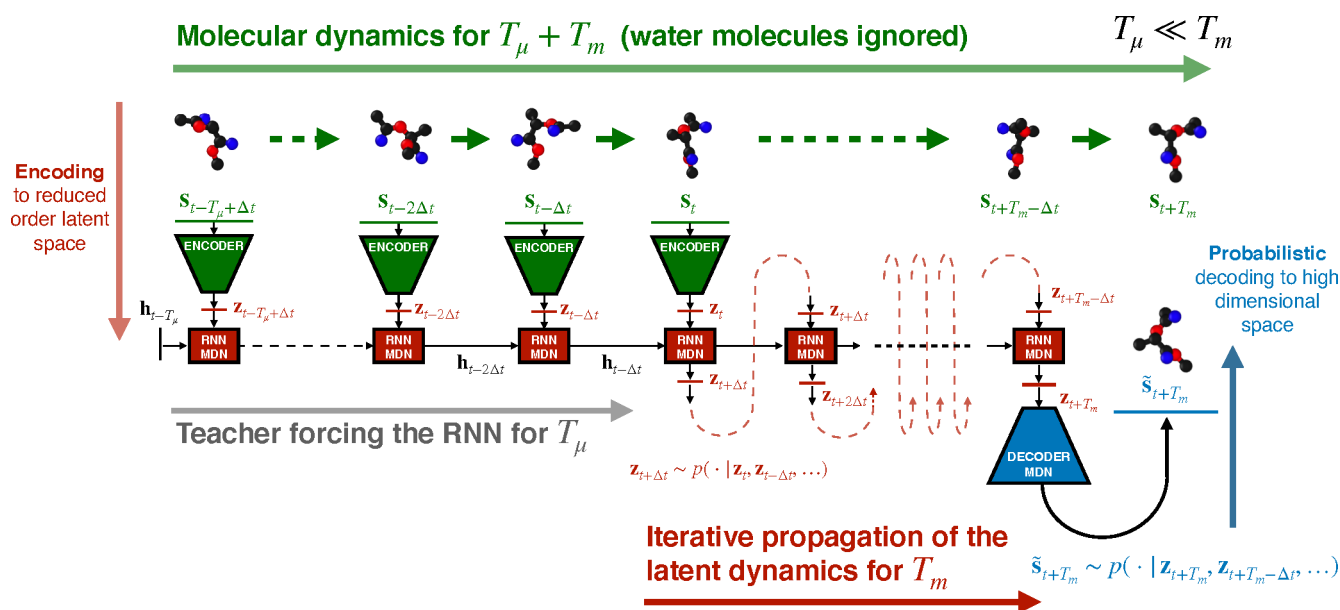
Figure 1. High-dimensional (fine-scale) dynamics s_t are simulated for a short period (T_μ). During this warm-up period, the state s_t is passed through the encoder network. The outputs of the encoder z_t provide the time series input to the LSTM, allowing for the update of its hidden state h_t, thus capturing non-Markovian effects. The output of the LSTM is a parametrization of the probabilistic non-Markovian latent dynamics $p(z_t|h_t)$. Starting from the last latent state z_t, the LSTM iteratively samples $p(z_t|h_t)$ and propagates the low-order latent dynamics up to a total horizon of T_m time units, with $T_m > T_\mu$. The LED decoder may be utilized at any desired time scale to map the latent state z_t back to a high-dimensional representation $s_t \sim p(\cdot|z_t, z_{t-\Delta t}, \ldots)$. Propagation in the low-order space unraveled by LED is orders of magnitude cheaper than evolving the high-dimensional system based on first-principles (molecular dynamics/density functional theory, etc.).

simulations. The field of data-driven learning of potential energy surfaces and force fields is rapidly attracting attention with important recent extensions and applications.[37−45] ML is employed to identify CG models for MD in refs [46−48]. Boltzmann generators are proposed in ref [49] to sample from the equilibrium distribution of a molecular system directly surpassing the need to perform MD.

Early ML methods for the identification of CVs utilized manifold-learning techniques, i.e., diffusion maps,[50−54] while others were based on the variational approach[55] leading to the time-lagged independent analysis (TICA).[56] TICA is based on the Koopman operator theory, suggesting the existence of a latent transformation to an infinite-dimensional space that linearizes the dynamics on average. As a consequence, slow CVs are modeled as linear combinations of feature functions of the state of the protein (atom coordinates or internal structural coordinates). Coarse graining of the molecular dynamics is achieved by discretizing the state space and employing indicator vector functions as features.[55,57−59] Consequently, the feature state dynamics reduce to the propagation law of a Markov state model (MSM). MSMs have been extended to "core set MSMs" in ref [29] employing Markovian milestoning[60] on metastable core sets.

More recently the need for expert knowledge to construct the latent feature functions has been alleviated by learning the latent space using neural networks.[61,62] The dynamics on the latent space are assumed to be linear and Markovian. For example, VAMPnets[61,63] learn nonlinear features of the molecular state with autoencoder (AE) networks. However, they are not generative and cannot recover the detailed configuration of the protein (decoding part). Moreover, the method requires the construction of an MSM to sample the latent dynamics and approximate the time scales of the dynamics. Time-lagged AE networks have been utilized to

identify a reaction coordinate embedding and propagate the dynamics in ref [62], but they are not generative, as the learned mappings are deterministic, while the effective dynamics are assumed to be Markovian.

Extensions to generative approaches include refs [64−66]. In ref [64] a deep generative MSM is utilized to capture the long-time-scale dynamics and sample realistic alanine dipeptide configurations. Even though mixture density networks (MDNs) are employed in ref [66] to propagate the dynamics in the latent space, memory effects are not taken into account. The proposed method is based on the autocorrelation loss, which suffers from the dependency on the batch size.[65] In refs [59] and [67] the reweighted autoencoded variational Bayes for enhanced sampling (RAVE) method is proposed that alternates between iterations of MD and a variational AE (VAE) model. RAVE is encoding each time step independently without taking into account the temporal aspect of the latent dynamics. RAVE requires the transition to the high-dimensional configuration space to progress the simulation in time, which can be computationally expensive. In recent work RAVE has been expanded to incorporate a "variational mixture of posteriors"[68] prior in the VAE[69] enhancing its performance.

The works mentioned above imply memoryless (Markovian) latent space dynamics by selecting an appropriate time lag in the master equations.[57,58] The time lag is usually estimated heuristically, balancing the requirements to be large enough so that the Markovian assumption holds and at the same time small enough to ensure that the method samples the configuration space efficiently. We remark that in cases where a protein is interacting with a solvent, only the configuration of the protein is taken into account and not the solvent. This renders the Markovian assumption in the latent dynamics rather unrealistic. This issue is addressed in this work by employing long short-term memory (LSTM)[70] recurrent

neural networks (RNNs) that capture memory effects of the latent dynamics. An LSTM has been used in ref [71] in the form of a language model to learn non-Markovian protein dynamics. The model they propose, however, requires discretization of the state space, is not generative, and does not identify a low-order latent representation of the dynamics.

Here we propose a novel data-driven generative framework that relies on learning the effective dynamics (LED) of molecular systems.[72] LED is founded on the equation-free framework (EFF,)[73] and it enriches it by employing ML methodologies to evolve the latent space dynamics with the mixture density network long short-term memory RNN (MDN-LSTM) and the two-way mapping between coarse and fine scales with mixture density network autoencoders (MDN-AEs).[74] These enrichments are essential in extending the applicability of EFF to molecular systems with stochastic non-Markovian dynamics. We demonstrate the effectiveness of the LED framework in simulations of the Müller−Brown potential (MBP), the Trp cage miniprotein, and the alanine dipeptide in water. LED can accurately capture the statistics and reproduce the free energy landscape from data. Moreover, LED uncovers low-energy metastable states in the free energy projected to the latent space and recovers the transition time scales between them. We find that in simulations of the alanine dipeptide and the Trp cage miniprotein, LED is 3 orders of magnitude faster than the classical MD solver (not including training or data acquisition cost). As a data-driven generative method, LED has the ability to sample novel unseen configurations interpolating the training data and accelerating the exploration of the state space.

2. MATERIALS AND METHODS

The LED framework[72] for molecular systems is founded on the equation-free framework (EFF).[73] It addresses the key bottlenecks of EFF, namely, the coarse to fine mapping and the evolution of the latent space using an MDN-AE and an MDN-LSTM, respectively. An illustration of the LED framework is given in Figure 1.

In the following, the state of a molecule at time t is described by a high-dimensional vector $s_t \in \Omega \subseteq \mathbb{R}^{d_s}$, where $d_s \in \mathbb{N}$ denotes its dimension. The state vector can include the atom positions or their rotation/translation invariant features obtained using for example the Kabsch transform.[75] A trajectory of this system is obtained by an MD integrator, and the state of the molecule after a time step Δt is described by the probability distribution function (PDF):

$$p(s_{t+\Delta t}|s_t) \tag{1}$$

The transition distribution in eq 1 depends on the choice of Δt.

2.1. Mixture Density Network Autoencoder. Here the MDN-AE is utilized to identify the latent (coarse) representation and upscale it probabilistically to the high-dimensional state space. MDNs[32] are neural architectures that can represent arbitrary conditional distributions. The MDN output is a parametrization of the distribution of a multivariate random variable conditioned on the input of the network.

The latent state is computed by $z_t = \mathcal{E}(s_t; w_\mathcal{E})$, where $\mathcal{E}$ is the encoder (a deep neural network) with trainable parameters $w_\mathcal{E}$ and $z_t \in \mathbb{R}^{d_z}$ with $d_z \ll d_s$. Since z_t is a coarse approximation, many states can be mapped to the same z_t. As a consequence, a deterministic mapping $z_t \to s_t$ like the one

used in refs [61] and [62] does not provide the full distribution $p(s_t|z_t)$. Here, an MDN is employed to model the upscaling conditional PDF $p(s_t|z_t)$ described by the parameters $w_{s|z}$. These parameters are the outputs of the decoder with weights $w_\mathcal{D}$ and are a function of the latent representation z_t, i.e.

$$w_{s|z}(z_t) = \mathcal{D}(z_t; w_\mathcal{D}) \tag{2}$$

The state of the molecule can then be sampled from $p(s_t|z_t)$ $:= p(s_t; w_{s|z})$.

Arguably, including the rotation/translation invariant features of the molecule under study in the state s_t encourages the MDN to sample physically meaningful molecular configurations. The state s_t is composed of states representing bond lengths $s_t^b \in \mathbb{R}^{d_s^b}$ and angles $s_t^a \in \mathbb{R}^{d_s^a}$. Initially, the MD data of the bonds are scaled to $[0, 1]$. An auxiliary variable vector $v_t \in \mathbb{R}^{d_s^b}$ is defined to model the distribution of bonds. In particular, $p(v_t|z_t)$ is modeled as a Gaussian mixture model with K_s mixture kernels as

$$p(v_t|z_t) = \sum_{k=1}^{K_s} \pi_v^k(z_t)\mathcal{N}(\mu_v^k(z_t), \sigma_v^k(z_t)) \tag{3}$$

and the mapping $s_t^b = \ln(1 + \exp(v_t))$ is used to recover the distribution of the scaled bond lengths at the output. The functional form of the mixing coefficients $\pi_v^k(z_t)$, the means $\mu_v^k(z_t)$, and the variances $\sigma_v^k(z_t)$ is a deep neural network (decoder $\mathcal{D}$). The distribution of the dihedral angles is modeled with the circular normal (von Mises) distribution, i.e.

$$p(s_t^a|z_t) = \sum_{k=1}^{K_s} \pi_{s^a}^k(z_t)\frac{\exp(\nu_{s^a}^k(z_t)\cos(s_t^a - \mu_{s^a}^k(z_t)))}{2\pi I_0(\nu_{s^a}^k(z_t))} \tag{4}$$

where $I_0(\nu_{s^a}^k)$ is the modified Bessel function of order 0. Here, again the functional form of $\pi_{s^a}^k(z_t)$, $\mu_{s^a}^k(z_t)$, and $\nu_{s^a}^k(z_t)$ is a deep neural network (decoder $\mathcal{D}$).

In total, the outputs of the decoder $\mathcal{D}$ that parametrize $p(s_t| z_t)$ are

$$w_{s|z} = \{\pi_v^k, \mu_v^k, \sigma_v^k, \pi_{s^a}^k, \mu_{s^a}^k, \nu_{s^a}^k\}_{k \in \{1,...,K_s\}} \tag{5}$$

which are all functions of the latent state z_t, which is the decoder input. The MDN-AE is trained to predict the mixing coefficients maximizing the data likelihood

$$\begin{aligned} w_\mathcal{E}, w_\mathcal{D} &= \underset{w_\mathcal{E},w_\mathcal{D}}{\mathrm{argmax}}\, p(s_t|z_t) \\ &= \underset{w_\mathcal{E},w_\mathcal{D}}{\mathrm{argmax}}\, p(s_t; w_{s|z}) \end{aligned} \tag{6}$$

where $w_{s|z} = \mathcal{D}(\mathcal{E}(s_t; w_\mathcal{E}); w_\mathcal{D})$ is the output of the MDN-AE and s_t are the MD data. The details of the training procedure can be found in ref [76].

2.2. Long Short-Term Memory Recurrent Neural Network. The latent dynamics may be characterized by non-Markovian effects, i.e.

$$p(z_{t+\Delta t}|z_t, z_{t-\Delta t}, ...)$$

due to the neglected degrees of freedom (solvent) or the selection of a relatively small time lag Δt.

Here the LSTM cell architecture[70] is utilized to evolve the nonlinear and non-Markovian latent dynamics. The propagation in the LSTM is given by

$$\mathbf{h}_t, \mathbf{c}_t = \mathcal{R}(\mathbf{z}_t, \mathbf{h}_{t-\Delta t}, \mathbf{c}_{t-\Delta t}; \mathbf{w}_{\mathcal{R}}) \tag{7}$$

where the hidden-to-hidden recurrent mapping $\mathcal{R}$ takes the form

$$\mathbf{g}_t^f = \sigma_f(W_f[\mathbf{h}_{t-\Delta t}, \mathbf{z}_t] + \mathbf{b}_f)$$

$$\mathbf{g}_t^i = \sigma_i(W_i[\mathbf{h}_{t-\Delta t}, \mathbf{z}_t] + \mathbf{b}_i)$$

$$\tilde{\mathbf{c}}_t = \tanh(W_c[\mathbf{h}_{t-\Delta t}, \mathbf{z}_t] + \mathbf{b}_c)$$

$$\mathbf{c}_t = \mathbf{g}_t^f \odot \mathbf{c}_{t-\Delta t} + \mathbf{g}_t^i \odot \tilde{\mathbf{c}}_t$$

$$\mathbf{g}_t^z = \sigma_h(W_h[\mathbf{h}_{t-\Delta t}, \mathbf{z}_t] + \mathbf{b}_h)$$

$$\mathbf{h}_t = \mathbf{g}_t^z \odot \tanh(\mathbf{c}_t) \tag{8}$$

where $\mathbf{g}_t^f, \mathbf{g}_t^i, \mathbf{g}_t^z \in \mathbb{R}^{d_h}$ are the gate vector signals (forget, input, and output gates), $\mathbf{z}_t \in \mathbb{R}^{d_z}$ is the latent input at time t, $\mathbf{h}_t \in \mathbb{R}^{d_h}$ is the hidden state, $\mathbf{c}_t \in \mathbb{R}^{d_h}$ is the cell state, while W_f, W_i, W_c, $W_h \in \mathbb{R}^{d_h \times (d_h + d_z)}$ are weight matrices and $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_c, \mathbf{b}_h \in \mathbb{R}^{d_h}$ biases. The symbol $\odot$ denotes the elementwise product. The activation functions σ_f, σ_i, and σ_h are sigmoids. The dimension of the hidden state d_h (number of hidden units) controls the capacity of the cell to encode history information. The set of trainable parameters of the LSTM are

$$\mathbf{w}_{\mathcal{R}} = \{\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_c, \mathbf{b}_h, W_f, W_i, W_c, W_h\} \tag{9}$$

An illustration of the information flow in an LSTM cell is given in Figure 2. The cell state can encode the history of the latent state evolution and capture non-Markovian effects.
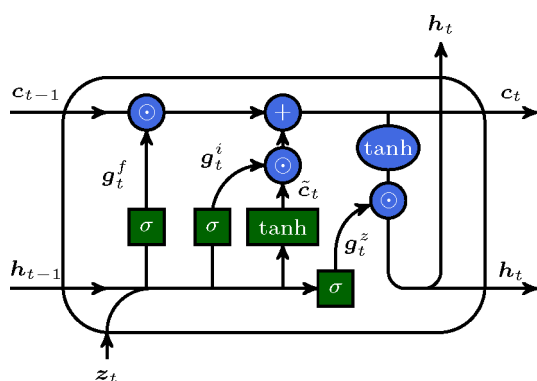

Figure 2. Information flow in an LSTM cell.

2.3. Mixture Density LSTM Network.
The LSTM captures the history of the latent state, and the non-Markovian latent transition dynamics are expressed as

$$p(\mathbf{z}_{t+\Delta t} | \mathbf{z}_t, \mathbf{z}_{t-\Delta t}, \ldots) = p(\mathbf{z}_{t+\Delta t} | \mathbf{h}_t) \tag{10}$$

where are $\mathbf{h}_t$ given in eq 7. A second MDN is used to model the conditional distribution $p(\mathbf{z}_{t+\Delta t} | \mathbf{h}_t)$ of the latent transition dynamics. This MDN is conditioned on the hidden state of the LSTM $\mathbf{h}_t$ and implicitly conditioned on the history, i.e., $p(\mathbf{z}_{t+\Delta t} | \mathbf{z}_t, \mathbf{z}_{t-\Delta t}, \ldots) := p(\mathbf{z}_{t+\Delta t}; \mathbf{w}_{z|h})$, so it can capture non-Markovian dynamics. The distribution $p(\mathbf{z}_{t+\Delta t} | \mathbf{h}_t)$ is modeled as a Gaussian mixture with K_z mixture kernels

$$p(\mathbf{z}_{t+\Delta t} | \mathbf{h}_t) = \sum_{k=1}^{K_z} \pi_z^k(\mathbf{h}_t) \mathcal{N}(\boldsymbol{\mu}_z^k(\mathbf{h}_t), \boldsymbol{\sigma}_z^k(\mathbf{h}_t)) \tag{11}$$

with parameters $\mathbf{w}_{z|h}$ given by

$$\mathbf{w}_{z|h}(\mathbf{h}_t) = \{\pi_z^k(\mathbf{h}_t), \boldsymbol{\mu}_z^k(\mathbf{h}_t), \boldsymbol{\sigma}_z^k(\mathbf{h}_t)\} \tag{12}$$

that are a function of $\mathbf{h}_t$. These parameters are the outputs of the neural network $\mathcal{Z}(\mathbf{h}_t; \mathbf{w}_{\mathcal{Z}})$, with trainable weights $\mathbf{w}_{\mathcal{Z}}$, and are a function of the hidden state, i.e.

$$p(\mathbf{z}_{t+\Delta t} | \mathbf{h}_t) := p(\mathbf{z}_{t+\Delta t}; \mathbf{w}_{z|h})$$

$$\mathbf{w}_{z|h}(\mathbf{h}_t) = \mathcal{Z}(\mathbf{h}_t; \mathbf{w}_{\mathcal{Z}}) \tag{13}$$

The weights of the LSTM $\mathbf{w}_{\mathcal{R}}$ and the latent MDN $\mathbf{w}_{\mathcal{Z}}$ are trained to output the parameters $\mathbf{w}_{z|h}$ that maximize the likelihood of the latent evolution

$$\mathbf{w}_{\mathcal{R}}, \mathbf{w}_{\mathcal{Z}} = \underset{\mathbf{w}_{\mathcal{R}}, \mathbf{w}_{\mathcal{Z}}}{\mathrm{argmax}} p(\mathbf{z}_{t+\Delta t}; \mathbf{w}_{z|h}) \tag{14}$$

where $\mathbf{w}_{z|h}$ is defined in eq 13, and $\mathbf{h}_t$ appearing in eq 13 is defined in eq 7. During the training phase, the MD trajectory data $\mathbf{s}_t$ are provided at the input of the trained MDN-AE $\mathbf{z}_t = \mathcal{E}(\mathbf{s}_t; \mathbf{w}_{\mathcal{E}})$. The encoder outputs the latent dynamics $\mathbf{z}_t$ that are used to update the hidden state of the LSTM and optimize its weights according to eq 14. In contrast to the linear operator utilized in MSMs, the recurrent functional form in eq 7 can be nonlinear and incorporate memory effects via the hidden state of the LSTM.

2.4. Learning Effective Dynamics.
The LED framework can be employed to accelerate MD simulations and enable more efficient exploration of the state space and uncovering of novel protein configurations. The networks in LED are trained on trajectories from MD simulations in two phases. First, the MDN-AE provides a reduced-order representation, maximizing the data likelihood (ref 72). The MDN-AE is trained with backpropagation[77] using the adaptive stochastic optimization method Adam.[78] Adding a pretraining phase fitting the kernels $\boldsymbol{\mu}^k$, $\boldsymbol{\sigma}^k$ of the MDN-AE to the data and fixing them during MDN-AE training led to better results. Next, the MDN-LSTM is trained to forecast the latent space dynamics (the MDN-AE weights are considered fixed) to maximize the latent data likelihood. MDN-LSTM is trained with backpropagation through time (BPTT)[79] with the Adam optimizer.

The LED propagates the computationally inexpensive dynamics on its latent space. Starting from an initial state from a test data set (unseen during training), a short time history T_μ of the state evolution is utilized to warm up the hidden state of the LED. The MDN-LSTM is used to propagate the latent dynamics for a time horizon $T_m \gg T_\mu$. High-dimensional state configurations can be recovered at any time instant by using the probabilistic decoder part of MDN-AE. We find that the LED framework can accelerate MD simulations by 3 orders of magnitude.

3. RESULTS

The LED framework is tested in three systems, single-particle Langevin dynamics using the two-dimensional MBP, the Trp cage miniprotein, and the alanine dipeptide, widely adopted as benchmarks for molecular dynamics modeling.[55,61,62,66,80]

3.1. Müller–Brown Potential (MBP).
The Langevin dynamics of a particle in the MBP are characterized by the stochastic differential equation
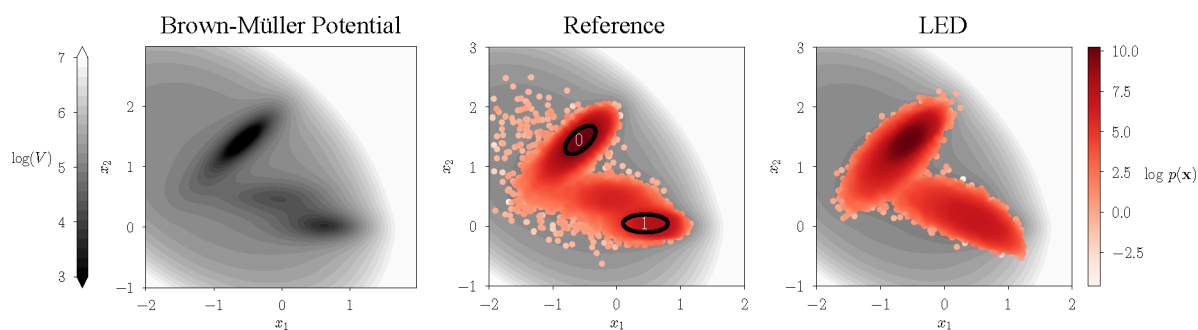
Figure 3. From left to right: the Müller−Brown potential, a scatter plot of the joint state distribution computed from reference data (with annotation of two long-lived metastable states), and the same scatter plot obtained by LED sampled trajectories.
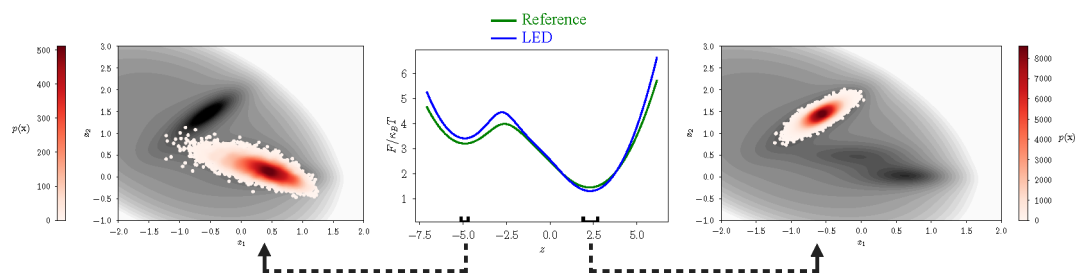

Figure 4. Middle: Free energy profile projected on the latent space learned by the LED encoder, i.e., $F = -\kappa_B T \ln p(z_t)$. The free energy profile computed by LED (propagation of the latent dynamics with LED) matches closely the one from the reference data. Quantitatively, the root-mean-square error is 0.74 $\kappa_B T$. LED recovers two low-energy regions that are mapped to the two long-lived metastable states (left and right) in the two-dimensional state space $s_t \in \mathbb{R}^2$.

$$m\ddot{x}(t) = -\nabla V(x(t)) - \gamma\dot{x}(t) + \sqrt{2k_B T}\, R(t) \qquad (15)$$

where $x \in \mathbb{R}^2$ is the position, $\dot{x}$ is the velocity, $\ddot{x}$ is the acceleration, $V(x)$ is the MBP (defined in the Supporting Information), k_B is Boltzmann's constant, T is the temperature, γ is the damping coefficient, and $R(t)$ a delta-correlated stationary Gaussian process with zero-mean. The nature of the dynamics is affected by the damping coefficient γ. Low damping coefficients lead to an inertial regime. High damping factors lead to a diffusive regime (Brownian motion) with less prominent memory effects. Here, a damping $\gamma = 1$ is considered, along with $k_B T = 15$.

The equations are integrated with the velocity Verlet algorithm with time step $\delta t = 10^{-2}$, starting from 96 initial conditions randomly sampled uniformly from $x \in [-1.5, 1.2] \times [-0.2, 2]$ until $T = 10^4$, after truncating an initial transient period of $\tilde{T} = 10^3$. The data are subsampled keeping every 50th data point to create the training and testing data sets for LED. The coarse time step of LED is $\Delta t = 0.5$. We use 32 initial conditions for training, 32 for validation, and all 96 for testing. LED is trained with a one-dimensional reduced order latent representation $z_t \in \mathbb{R}$. The reader is referred to the Supporting Information for further information regarding the MBP parametrization of ref 80 and hyperparameters of LED.

The MBP is shown in Figure 3, along with a density scatter plot of the joint distribution of the MBP states computed from the testing data and LED. The potential has three minima. The potential value around the middle one, however, is more than 1 order of magnitude higher than the other two and lies very close to the one on the right (Figure 3, left and middle). The joint distribution reveals two dominant long-lived metastable states that correspond to the two dominant low-energy regions. The local minimum in the middle is not clearly distinguishable. The LED learns to transition probabilistically between the metastable states, mimicking the dynamics of the system and reproducing the state statistics. We note, however, that the LED cannot distinguish the region of the local minimum in the middle as a separate metastable state (Figure 3, right). A "weak" metastable state region (local minimum but high potential compared to other nearby minima) that lies close to a dominant one of much lower potential value can be missed (or be absorbed to the nearby regions).

The free energy projected on the latent space, i.e., $F = -\kappa_B T \log p(z_t)$, is plotted in Figure 4. The free energy profile of the trajectories sampled from LED matches closely the one from the reference data with a root-mean-square error between the two free energy profiles of ≈ 0.74 $\kappa_B T$. LED reveals two minima in the free energy profile. As noted before, the third local minimum of the BMP is absorbed by the dominant one close to it. Even though LED cannot distinguish the third local minimum, the projected free energy profile is reproduced. Utilizing the LED decoder, the latent states in these regions are mapped to their image in the two-dimensional state representation $s_t \in \mathbb{R}^2$ (here corresponding to $x_t \in \mathbb{R}^2$) in Figure 4. LED is mapping the low-energetic regions in the free energy profile to the long-lived metastable states in the two-dimensional space of the MBP.

Next, we evaluate the LED framework in reproducing the transition times between the long-lived states. In LED, metastable states can be defined on either the reduced order latent space $z_t \in \mathbb{R}$ or the state space $s_t \in \mathbb{R}^2$ (as the decoder can map any latent state to a state space). In the following, two metastable states are defined as ellipses on the state space (Table S1) and depicted in Figure 3. The time scales will vary depending on the definition of the metastable states in the phase space. The distribution of transition times computed from LED trajectories is compared with the transition time
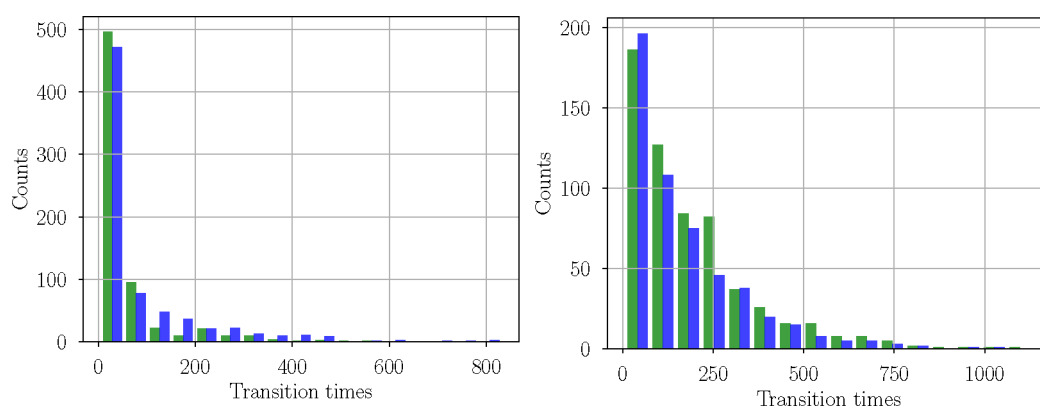
Figure 5. Distribution of the transition times learned by LED (blue), computed from sampled trajectories, matching the original fine-scale transition times of the MBP dynamics (green). Left: Histogram of $T_{0\to1}$. Mean $T_{0\to1}$ of MD trajectories is 61; mean $T_{0\to1} = 91$ for LED. Right: Histogram of $T_{1\to0}$. Mean $T_{1\to0}$ of MD trajectories is 188; mean $T_{1\to0} = 164$ for LED. LED has learned to propagate the effective dynamics (a one-dimensional latent state z) and capture the non-Markovian effects.
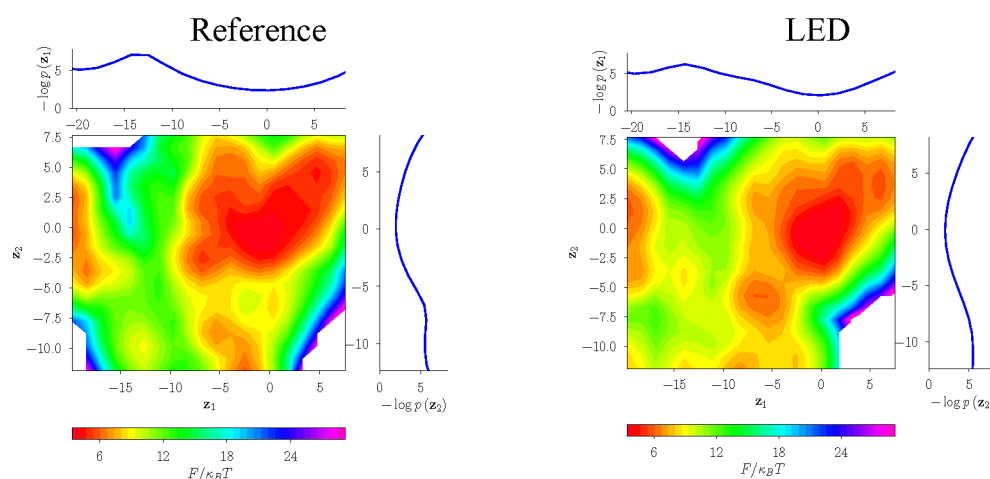

Figure 6. Free energy projection on the latent space $F = -\kappa_B T \log p(z_t)$, with $z_t \in \mathbb{R}^2$. Left: MD data projected to the LED latent space. Right: Free energy of trajectories sampled from LED. LED is capturing the free energy profile.

distribution from the test data in Figure 5. LED captures quantitatively the transition time distributions, and the mean values are close to each other. In the Supporting Information, we also report the transition times obtained with metastable states definition on the latent space. This approach has the benefit of not requiring the prior knowledge about the metastable states in the state space. In conclusion, LED is capturing the joint state distribution on the MBP and matching the time scales of the system.

3.2. Trp Cage. The Trp cage is considered a prototypical miniprotein for the study of protein folding.[66] The protein is simulated with MD[81] with a time step $\delta t = 1$ fs, up to a total time of $T = 100$ ns. The data are subsampled at $\Delta t = 0.1$ ps, creating a trajectory with $N = 10^6$ samples. The data are divided into 248 sequences of 4000 samples ($T = 400$ ps each). The first 96 sequences are used for training (corresponding to 38.4 ns) and the next 96 sequences for validation, while all data is used for testing.

The protein positions are transformed into rototranslational invariant features (internal coordinates), composed of bonds, angles, and dihedral angles, leading to a state with dimension $d_s = 456$. LED is trained with a latent space $z_t \in \mathbb{R}^2$, i.e., $d_z = 2$. LED is tested by starting from the initial condition in each of the 248 test sequences, iteratively propagating the latent space

to forecast $T = 400$ ps. For more information on the hyperparameters of LED, refer to the Supporting Information.

The projection of MD trajectory data to LED latent space is illustrated in Figure 6, left, in the form of the free energy, i.e., $F = -\kappa_B T \log p(z_t)$, with $z_t = (z_1, z_2)^T \in \mathbb{R}^2$. The free energy on the latent space computed from trajectories sampled from LED is given in Figure 6 on the right. LED successfully captures the three metastable states of the Trp cage miniprotein, while being 3 orders of magnitude faster compared to the MD solver. Quantitatively, the two profiles agree up to an error margin of approximately 22.5 $\kappa_B T$. The Supporting Information provides additional results on the agreement of the marginal state distributions (Figure S2) and realistic samples of the protein configuration sampled from LED (Figure S3).

3.3. Alanine Dipeptide. The alanine dipeptide is often used as the testing ground for enhanced sampling methods.[82] LED is evaluated in learning and propagating the dynamics of alanine dipeptide in water. The molecule is simulated with MD,[81] and the same data acquisition procedure with the Trp cage in Section 3.2 is used ($\delta t = 1$ fs, $N = 10^6$, $\Delta t = 0.1$ ps, 96 trajectories for training, 96 for validation). LED is tested by starting from the initial condition in each of the total 248 test trajectories, iteratively propagating the latent dynamics to
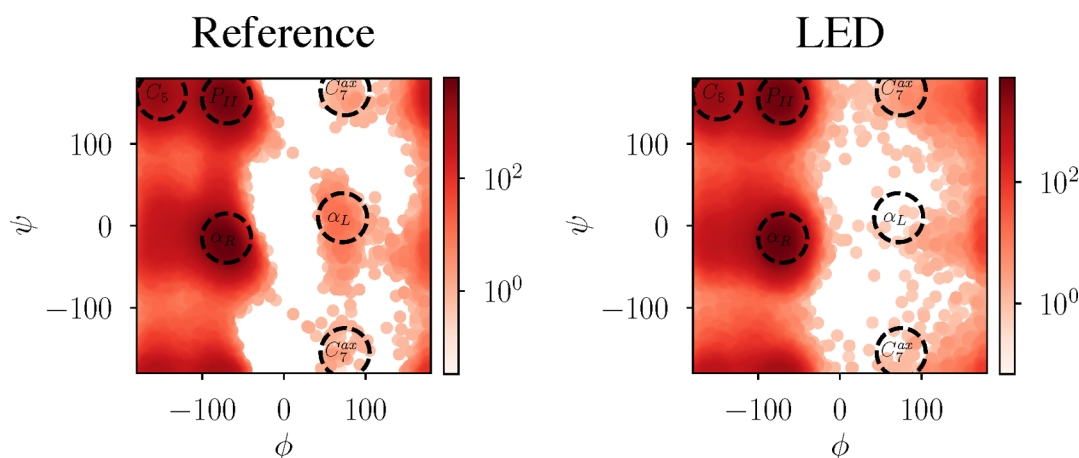
Reference   LED

Figure 7. Ramachandran plot of the alanine dipeptide, i.e., space spanned by two backbone dihedral angles (ϕ, ψ). Scatter plots are colored based on the joint density of (ϕ, ψ). Left: Test data. Right: LED trajectories. We observe five energetically favorable metastable states denoted with $\{C5, P_{II}, \alpha_R, \alpha_L, C_7^{ax}\}$. LED captures the three dominant metastable states $\{C5, P_{II}, \alpha_R\}$. The states $\{\alpha_L, C_7^{ax}\}$ are rarely observed in the training data.
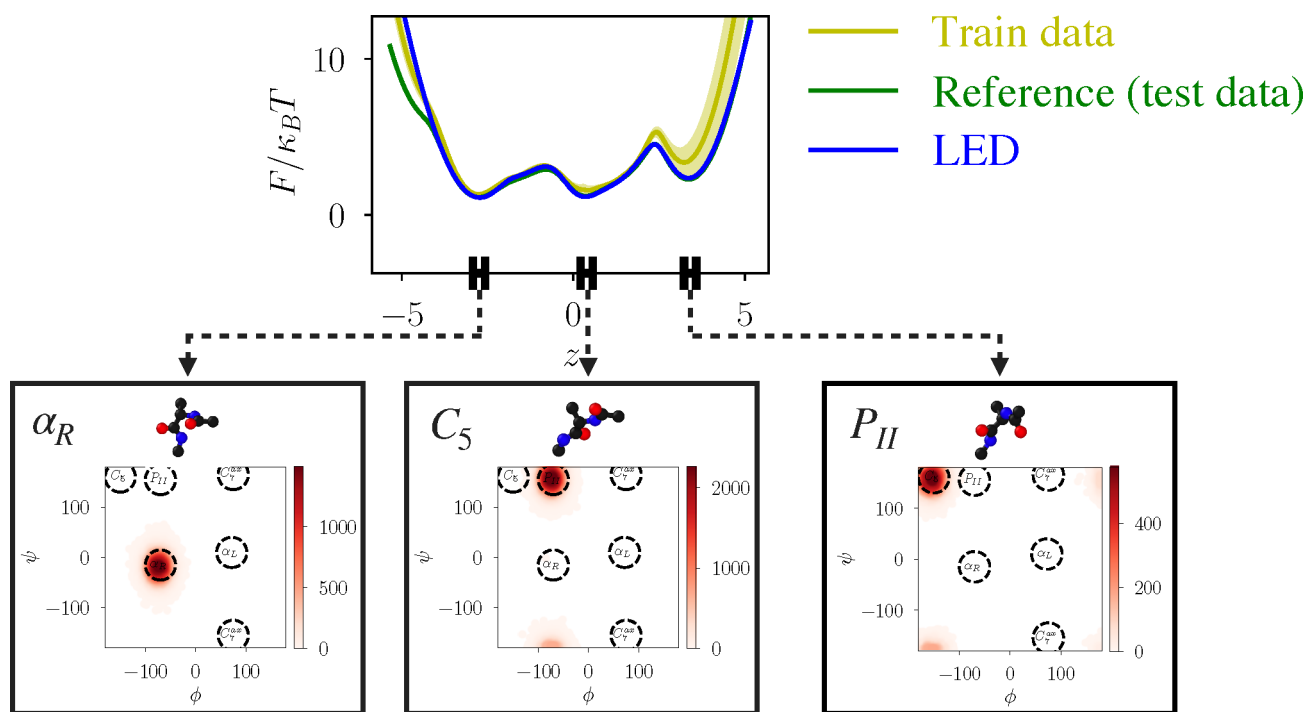

Figure 8. Plot of the free energy profile projected on the latent state learned by the LED, i.e., $F = -\kappa_B T \ln p(z_t)$. The latent free energy profile of MD trajectories is compared with the latent free energy profile of trajectories sampled from LED. The two profiles agree up to a root-mean-square error of $0.25 \, \kappa_B T$. Utilizing the LED decoder, the low-energy regions in the latent space (close to the minima) can be mapped to the corresponding protein configurations and metastable states in the Ramachandran plot. The LED uncovers the three dominant metastable states $\{C_S, P_{II}, \alpha_R\}$ in the free energy surface (minima). The LED captures the free energy profile and the dominant metastable states while being computationally 3 orders of magnitude cheaper than MD.

forecast $T = 400$ ps. This testing is repeated 5 times with a different random seed, producing data equivalent to $T = 496$ ns in total.

The dipeptide positions are transformed into rototranslational invariant features (internal coordinates), composed of bonds, angles, and dihedral angles, leading to a state with dimension $d_s = 24$. In order to demonstrate that LED can uncover the dynamics in a drastically reduced order latent space, the dimension of the latter is set to 1 ($d_z = 1$), i.e., $z_t \in \mathbb{R}$. For more information on the hyperparameters of LED, refer to the Supporting Information.

The metastable states of the dynamics are represented in terms of the energetically favored regions in the state space of two backbone dihedral angles, ϕ and ψ, i.e., the Ramachandran space[83] plotted in Figure 7. Specifically, previous works consider five low-energy clusters, i.e., $\{C5, P_{II}, \alpha_R, \alpha_L, C_7^{ax}\}$. The trained LED is qualitatively reproducing the density in the Ramachandran plot in Figure 7, identifying the three dominant low-energy metastable states $\{C5, P_{II}, \alpha_R\}$. LED, however, fails to capture the state density on the less frequently observed states in the training data $\{\alpha_L, C_7^{ax}\}$. The marginal distributions of the trajectories generated by LED match the ground-truth ones (MD data) closely, as depicted in Figure S4.
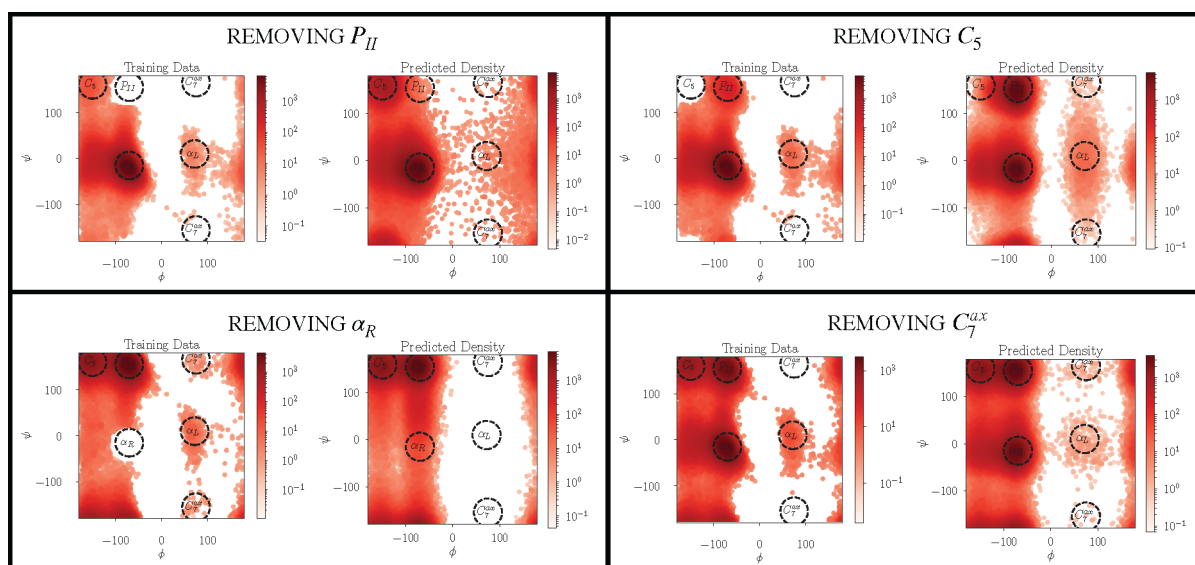
Figure 9. LED is trained in four scenarios hiding data that lie closer than 40° to one of the metastable states $\{P_{II}, C_5, \alpha_R, C_7^{ax}\}$ each time. LED can successfully generate novel probable configurations close to the metastable states $\{P_{II}, C_5, \alpha_R, C_7^{ax}\}$. Due to the limited training data, however, capturing the state density in the Ramachandran plot is challenging.

Even though LED is propagating a one-dimensional latent state, once trained, it can reproduce the statistics while being 3 orders of magnitude faster than the MD solver.

The free energy is projected to the latent space, i.e., $F = -\kappa_B T \ln(p(z_t))$, and plotted in Figure 8. The free energy projection computed from MD trajectories (train and test data) is compared with the one computed from trajectories sampled from LED. We estimate the mean free energy profile and the associated standard error of the mean (SEM) using 96 splits of the data. The free energy profile of LED agrees with the reference (test data) up to a root-mean-square error of 0.25 $\kappa_B T$. The profile estimated from the train data exhibits a slightly higher error of 1.22 $\kappa_B T$ and higher SEM. Note that LED unravels three dominant minima in the latent space. These low-energy regions correspond to metastable states of the dynamics.

The Ramachandran space (ϕ, ψ) is frequently used to describe the long-term behavior and metastable states of the system.[62,84] The latent encoding of the LED is evaluated based on the mapping between the latent space and the Ramachandran space. Utilizing the MDN decoder, the LED can map the latent state z to the respective rototranslational invariant features (bonds and angles) and regions in the Ramachandran plot. As illustrated in Figure 8, the LED is mapping the three low-energy regions in the latent space to the three dominant metastable states in the Ramachandran plot $\{C_5, P_{II}, \alpha_R\}$.

Next, we evaluate LED's effectiveness in unraveling novel configurations of the protein (state-space) absent from the training data. For this purpose, we create four different small data sets composed of trajectories of the protein, each one not including one of the metastable states $\{C5, P_{II}, \alpha_R, C_7^{ax}\}$. This is done by removing any state that lies closer than 40° to the metastable states' centers. In this way, we guarantee that LED has not seen any state close to the metastable state missing from the data. Note that in this case, the LED is not trained on a single large MD trajectory but on small trajectories that are not temporally adjacent. We end up with four data sets, each one consisting of approximately 800 trajectories of length $T =$

50 ps (500 steps of 0.1 ps). Each data set covers approximately 40 ns protein simulation time. These data sets are created to evaluate the effectiveness of LED in generating truly novel configurations for faster exploration of the state space. We do not care at this point for accurate reproduction of the statistics due to the minimal data used for training. In Figure 9, we plot the Ramachandran plots of the training data along with the ones obtained by analyzing the trajectories of the trained LED models in each of the four cases. We observe that the LED can unravel the metastable states P_{II}, C_5, C_7^{ax}, and α_R, even though they were not part of the training data. However, by removing states that lie close to the metastable state α_R, the LED cannot capture the α_R and C_7^{ax} metastable states. This is because the LED is trained on only a small subset of the training data set, and the transitions to these metastable states are rare.

The dynamics learned by LED are evaluated according to the mean first-passage times (MFPTs) between the dominant metastable states. The MFPT is the average time scale to reach a final metastable state, starting from any initial state. The MFPTs are computed a posteriori from trajectories sampled from the LED and the MD test trajectories, using the PyEMMA software.[85] The metastable states considered here are given in the Supporting Information.

As a reference for the MFPTs, we consider an MSM fitted to the MD data (test data set). The reference MFPTs agree with previous literature.[84,86−88] The time lag of the MSM is set to $\Delta t_{MSM} = 10$ ps to ensure the necessary Markovianity of the dynamics. This time lag is 2 orders of magnitude larger than the time step of LED. Fitting an MSM with a time lag of $\Delta t_{MSM} = 1$ ps on the MD data results in very high errors ($\approx 85\%$ on average) in the computation of MFPTs. This emphasizes the need for non-Markovian models that can reproduce the system's dynamics and statistics independent of the selection of the time lag.

The MFPTs of trajectories sampled from LED are estimated with an MSM with a time lag $\Delta t_{MSM} = 10$ ps. We consider 1232 trajectories sampled from LED, split them into 32 groups, and report the mean MFPT and the associated standard error of the mean (SEM). Note that the LED is

Table 1. Mean First-Passage Times (MFPTs) between the Metastable States of Alanine Dipeptide in Water (in ns)[a]

MFPT (ns)	MSM−10 ps on MD data	MSM−1 ps on MD data		MSM−10 ps on MD train data		MSM−10 ps on LED−0.1 ps data	
	reference	MFPT	error (%)	MFPT ± SEM	error (%)	MFPT ± SEM	error (%)
$T_{C_S \to P_{II}}$	0.112	0.017	84	0.123 ± 0.013	9	0.094 ± 0.002	16
$T_{C_S \to \alpha_R}$	0.096	0.014	86	0.107 ± 0.012	11	0.093 ± 0.003	4
$T_{P_{II} \to C_S}$	0.238	0.038	84	0.242 ± 0.021	2	0.218 ± 0.005	8
$T_{P_{II} \to \alpha_R}$	0.098	0.014	86	0.109 ± 0.012	11	0.093 ± 0.003	5
$T_{\alpha_R \to C_S}$	0.247	0.038	85	0.251 ± 0.021	1	0.232 ± 0.005	6
$T_{\alpha_R \to P_{II}}$	0.124	0.018	86	0.134 ± 0.014	8	0.110 ± 0.002	11
average relative error			85.01%		6.99%		8.41%

[a]MFPTs are estimated by fitting MSMs with different time lags (10 and 1 ps) on trajectories generated by MD or the LED framework. The average relative error is given for reference.

operating on a time step $\Delta t = 0.1$ ps. The MFPTs are identified with a low average relative error of 8.41%. The results on the MFPTs are summarized in Table 1. LED captures very well the transitions that are dominant in the data, e.g., $T_{P_{II} \to C_S}$ or $T_{\alpha_R \to C_S}$. In contrast, LED exhibits higher MFPT errors in transitions that are less dominant in the training data.

LED identifies the dominant MFPTs successfully by utilizing a very small amount of training data (38.4 ns for training and 38.4 ns validation) and propagating the latent dynamics on a reduced order space ($d_z = 1$). LED trajectories are 3 orders of magnitude cheaper to obtain compared to MD data. As multiple trajectories (here 1232) can be sampled from LED at a fraction of the computational cost of MD, the SEM in the estimation of the MFPTs is small.

At the same time, MSM fitting is a relatively fast procedure once the clustering based on the metastable states is obtained. In contrast, a careless selection of the time lag in the MSM that fails to render the dynamics Markovian, (e.g., $\Delta t = 1$ ps) leads to a surrogate model that fails to capture the system time scales. This emphasizes the need to model non-Markovian effects with LED in the case of limited data sampled at a high frequency (small time steps Δt). A more informative selection of the time lag may alleviate this problem, rendering the dynamics Markovian as in the reference MSM. Still, the consequent subsampling of the data can lead to omissions of effects whose time scales are smaller than the time lag. As a consequence, the heuristic selection of the time lag is rendering the modeling process error-prone. Table S12 provides additional results on the MFPTs estimated based on the metastable state definition in the latent space of LED (without prior knowledge).

If the metastable states are known a priori or a latent state is available (e.g., from a trained autoencoder or an other method to unravel collective variables), the MFPTs are represented in, and can be computed directly from, the training data (Table 1, eight splits of the trajectories from the train data set considered). The SEM, however, is large due to the limited amount of data. We note that LED is not expected to reproduce accurately transitions that are not present in the training data.

LED utilizes 76 ns of MD data for training, generated with the MD solver in approximately 10 days. The total training time of LED is approximately 20 h. The trained model can generate 12 μs of MD data per day, while with MD we can generate approximately 7.8 ns per day. In order to acquire the $T = 496$ ns total data used in this case for statistics with MD,

we would need approximately 64 days. By training the LED (20 h) on data acquired from MD (10 days) and then sampling multiple trajectories (4 h) in parallel, we can acquire the same amount of data in 11 days. A realistic speed-up estimation, taking into account data acquisition and training, is thus 3. For a larger protein, this speed-up is expected to be higher.

4. DISCUSSION

This work presents a data-driven framework (LED) to learn and propagate the effective dynamics of molecular systems resulting in dramatically accelerated MD simulations. The LED maximizes the data likelihood for a continuous reduced-order latent representation. The nonlinear dynamics are propagated in the latent space, and the memory effects are captured through the hidden state of the LSTM. Moreover, the method is generative, and the decoder part of the MDN-AE can be employed to sample high-dimensional configurations on any desired time scales. This is in contrast to previous state-of-the-art methods based on the Markovian assumption on the latent state or on the minimization of the autocorrelation or the variational loss in the data. The latter take into account the error in the long-term equilibrium statistics explicitly to capture the system time scales but suffer from a dependency on the batch size.[65]

The encoder of LED is analogous to the coarse-graining model design, while the decoder is implicitly learning a backmapping to atomistic configurations. The LED automates the dimensionality reduction often associated with the empirical a priori selection of collective variables in molecular simulations.[23,62] At the same time the MDN-LSTM propagates the dynamics on the latent space in a form that is comparable to nonlinear, non-Markovian metadynamics.[82]

The effectiveness of LED is demonstrated in simulations of three systems. In Langevin dynamics using BMP, LED recovers the free energy landscape in the latent space, identifies two low-energetic states corresponding to the long-lived metastable states of the potential, and captures the transition times between the metastable states. In the Trp cage miniprotein, LED captures the free energy projection on the latent space and unravels three metastable states. Finally, for the system of alanine dipeptide in water, LED captures the configuration statistics of the system accurately while being 3 orders of magnitude faster than MD solvers. Moreover, it identifies three low-energetic regions in the free energy profile projected to the one-dimensional latent state that corresponds to the three dominant metastable states $\{\alpha_R, C_S, P_{II}\}$. LED also captures the

dominant mean first-passage times in contrast to the MSM operating on the same time scale, owing to the non-Markovian latent propagation in the latent state with the MDN-LSTM. Furthermore, we showcase how our framework is capable of unraveling novel protein configurations interpolating on the training data.

We note that the speed-up achieved by LED depends on the MD solver used, the dimensionality, and the complexity of the protein under study. While extrapolating estimates to systems not yet tested requires caution, our evidence suggests that the computationally efficient propagation in the latent space of the LED will always provide dramatic accelerations over molecular simulations. Further acceleration can be accomplished by coupling LED and MD solver in different parts of the domain for faster exploration of the state space.

We believe that LED paves the way for faster exploration of the conformational space of molecular systems. Future research efforts will target the application of LED to larger proteins and its capabilities in uncovering the metastable states in the free energy profile.

■ ASSOCIATED CONTENT

ⓈＩ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00809.

> Additional MD simulation details, definitions of metastable states, LED hyperparameter tuning as well as final hyperparameters, marginal state distributions, and MFPTs based on metastable state definition in the latent space of LED (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Petros Koumoutsakos − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States;* ⓞ orcid.org/0000-0001-8337-2122; Email: petros@seas.harvard.edu

Authors

Pantelis R. Vlachas − *Computational Science and Engineering Laboratory, ETH, Zurich CH-8092, Switzerland*

Julija Zavadlav − *Professorship of Multiscale Modeling of Fluid Materials, TUM School of Engineering and Design, Technical University of Munich, 85748 Garching bei München, Germany; Munich Data Science Institute, Technical University of Munich, 85748 Munich, Germany;* ⓞ orcid.org/0000-0002-4495-9956

Matej Praprotnik − *Laboratory for Molecular Modeling, National Institute of Chemistry, SI-1001 Ljubljana, Slovenia; Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia;* ⓞ orcid.org/0000-0003-0825-1659

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.1c00809

Notes

The authors declare no competing financial interest.
Code and data to reproduce the findings of this study will be made openly available in the public repository https://github.com/cselab/LED-Molecular upon publication. The PyEMMA

software package[85] is employed in the current work for MSM fitting and MFPT estimation.

■ REFERENCES

(1) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646.

(2) Shirts, M.; Pande, V. S. Screen Savers of the World Unite! *Science* **2000**, *290*, 1903−1904.

(3) Shaw, D.; Dror, R.; Salmon, J.; Grossman, J.; Mackenzie, K.; Bank, J.; Young, C.; Deneroff, M.; Batson, B.; Bowers, K.; et al. *Proceedings of the conference on high performance computing networking, storage and analysis. Millisecond-scale molecular dynamics simulations on Anton*, 2009.

(4) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* **1996**, *100*, 2567−2572.

(5) Brooks, B.; Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 6571−6575.

(6) Skjaerven, L.; Martinez, A.; Reuter, N. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 232−243.

(7) Praprotnik, M.; Janežič, D. Molecular Dynamics Integration Meets Standard Theory of Molecular Vibrations. *J. Chem. Inf. Model.* **2005**, *45*, 1571.

(8) Ichiye, T.; Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205−217.

(9) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(10) Zavadlav, J.; Arampatzis, G.; Koumoutsakos, P. Bayesian selection for coarse-grained models of liquid water. *Sci. Rep.* **2019**, *9*, 1.

(11) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the Representability Problem and the Physical Meaning of Coarse-Grained Models. *J. Chem. Phys.* **2016**, *145*, 044108.

(12) Pezeshkian, W.; König, M.; Wassenaar, T. A.; Marrink, S. J. Backmapping triangulated surfaces to coarse-grained membrane models. *Nat. Commun.* **2020**, *11*, 1.

(13) Stieffenhofer, M.; Wand, M.; Bereau, T. Adversarial Reverse Mapping of Equilibrated Condensed-Phase Molecular Structures. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 045014.

(14) Hess, B.; León, S.; van der Vegt, N.; Kremer, K. Long time atomistic polymer trajectories from coarse grained simulations: Bisphenol-A polycarbonate. *Soft Matter* **2006**, *2*, 409.

(15) Werder, T.; Walther, J. H.; Koumoutsakos, P. Hybrid atomistic-continuum method for the simulation of dense fluid flows. *J. Comput. Phys.* **2005**, *205*, 373.

(16) Ayton, G. S.; Noid, W. G.; Voth, G. A. Multiscale modeling of biomolecular systems: In serial and in parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192.

(17) Praprotnik, M.; Delle Site, L.; Kremer, K. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545.

(18) Huber, T.; Torda, A. E.; Gunsteren, W. F. V. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695.

(19) Voudouris, C. Guided Local Search—An illustrative example in function optimization. *BT Technol. J.* **1998**, *16*, 46.

(20) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236.

(21) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.

(22) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, *118*, 7762.

(23) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.

(24) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; et al. High-Resolution Mining of SARS-CoV-2 Main Protease Conformational Space: Supercomputer-Driven Unsupervised Adaptive Sampling. *Chem. Sci.* **2021**, *12*, 4889−4907.

(25) Peters, B.; Trout, B. L. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* **2006**, *125*, 054108.

(26) Stamati, H.; Clementi, C.; Kavraki, L. E. Application of nonlinear dimensionality reduction to characterize the conformational landscape of small peptides. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 223−235.

(27) Bittracher, A.; Banisch, R.; Schütte, C. Data-driven computation of molecular reaction coordinates. *J. Chem. Phys.* **2018**, *149*, 154103.

(28) Bonati, L.; Rizzi, V.; Parrinello, M. Data-driven collective variables for enhanced sampling. *J. Phys. Chem. Lett.* **2020**, *11*, 2998.

(29) Schütte, C.; Noé, F.; Lu, J.; Sarich, M.; Vanden-Eijnden, E. Markov state models based on milestoning. *J. Chem. Phys.* **2011**, *134*, 204105.

(30) Bittracher, A.; Koltai, P.; Klus, S.; Banisch, R.; Dellnitz, M.; Schütte, C. Transition manifolds of complex metastable systems. *J. Nonlinear Sci.* **2018**, *28*, 471−512.

(31) Michie, D. Memo" functions and machine learning. *Nature* **1968**, *218*, 19−22.

(32) Bishop, C. M. *Pattern recognition and machine learning*; Springer: New York, 2006; pp 128.

(33) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361.

(34) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547.

(35) Schütt, K. T., Chmiela, S., von Lilienfeld, O. A., Tkatchenko, A., Tsuda, K., Müller, K.-R., Eds. *Machine Learning Meets Quantum Physics*; Springer: Cham, 2020.

(36) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(37) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(38) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 1.

(39) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet-A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(40) Rowe, P.; Deringer, V. L.; Gasparotto, P.; Csányi, G.; Michaelides, A. An accurate and transferable machine learning potential for carbon. *J. Chem. Phys.* **2020**, *153*, 034702.

(41) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3*, e1701816.

(42) Imbalzano, G.; Anelli, A.; Giofré, D.; Klees, S.; Behler, J.; Ceriotti, M. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *J. Chem. Phys.* **2018**, *148*, 241730.

(43) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; Von Lilienfeld, O. A.; Tkatchenko, A.; Muller, K.-R. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404−3419.

(44) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; Von Lilienfeld, O. A. Prediction errors of molecular machine learning models lower than hybrid DFT error. *J. Chem. Theory Comput.* **2017**, *13*, 5255−5264.

(45) Cheng, B.; Engel, E. A.; Behler, J.; Dellago, C.; Ceriotti, M. Ab initio thermodynamics of liquid and solid water. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 1110.

(46) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. DeePCG: Constructing coarse-grained models via deep neural networks. *J. Chem. Phys.* **2018**, *149*, 034101.

(47) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; De Fabritiis, G.; Noé, F.; Clementi, C. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci.* **2019**, *5*, 755−767.

(48) Durumeric, A. E.; Voth, G. A. Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *J. Chem. Phys.* **2019**, *151*, 124110.

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

(50) Rohrdanz, M. A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.* **2011**, *134*, 124116.

(51) Coifman, R. R.; Kevrekidis, I. G.; Lafon, S.; Maggioni, M.; Nadler, B. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Model. Simul.* **2008**, *7*, 842−864.

(52) Preto, J.; Clementi, C. Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics. *Phys. Chem. Chem. Phys.* **2014**, *16*, 19181−19191.

(53) Zheng, W.; Rohrdanz, M. A.; Clementi, C. Rapid exploration of configuration space with diffusion-map-directed molecular dynamics. *J. Phys. Chem. B* **2013**, *117*, 12769−12776.

(54) Boninsegna, L.; Gobbo, G.; Noé, F.; Clementi, C. Investigating molecular kinetics by variationally optimized diffusion maps. *J. Chem. Theory Comput.* **2015**, *11*, 5947−5960.

(55) Nuske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S.; Noé, F. Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **2014**, *10*, 1739−1752.

(56) Pérez-Hernández, G.; Noé, F. Hierarchical time-lagged independent component analysis: computing slow modes and reaction coordinates for large molecular systems. *J. Chem. Theory Comput.* **2016**, *12*, 6118.

(57) Buchete, N.-V.; Hummer, G. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057.

(58) Noé, F.; Doose, S.; Daidone, I.; Löllmann, M.; Sauer, M.; Chodera, J. D.; Smith, J. C. Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 4822−4827.

(59) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **2018**, *149*, 072301.

(60) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, 194101.

(61) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 1.

(62) Wehmeyer, C.; Noé, F. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **2018**, *148*, 241703.

(63) Chen, W.; Sidky, H.; Ferguson, A. L. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019**, *150*, 214114.

(64) Wu, H.; Mardt, A.; Pasquali, L.; Noé, F. Deep generative markov state models. *NeurIPS Proc.* **2018**, 3975−3984.

(65) Hernández, C. X.; Wayment-Steele, H. K.; Sultan, M. M.; Husic, B. E.; Pande, V. S. Variational encoding of complex dynamics. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2018**, *97*, 062412.

(66) Sidky, H.; Chen, W.; Ferguson, A. L. Molecular latent space simulators. *Chem. Sci.* **2020**, *11*, 9459.

(67) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Past−future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nat. Commun.* **2019**, *10*, 1−8.

(68) Tomczak, J.; Welling, M. VAE with a VampPrior. *AISTATS* **2018**, 1214−1223.

(69) Wang, D.; Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **2021**, *154*, 134111.

(70) Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735.

(71) Tsai, S.-T.; Kuo, E.-J.; Tiwary, P. Learning molecular dynamics with simple language model built upon long short-term memory neural network. *Nat. Commun.* **2020**, *11*, 1−11.

(72) Vlachas, P. R.; Arampatzis, G.; Uhler, C.; Koumoutsakos, P. Learning the Effective Dynamics of Complex Multiscale Systems. *arXiv preprint arXiv:2006.13431* **2020**.

(73) Kevrekidis, I. G.; Gear, C. W.; Hyman, J. M.; Kevrekidid, P. G.; Runborg, O.; Theodoropoulos, C.; et al. Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis. *Commun. Math. Sci.* **2003**, *1*, 715−762.

(74) Bishop, C. Mixture density networks. *Neural Computing Research Group Report: NCRG/94/004*; Aston University: Birmingham, U.K., 1994.

(75) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32*, 922.

(76) Vlachas, P. R.; Pathak, J.; Hunt, B. R.; Sapsis, T. P.; Girvan, M.; Ott, E.; Koumoutsakos, P. Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Netw.* **2020**, *126*, 191.

(77) Rumelhart, D. E.; McClelland, J. L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*; MIT Press: Cambridge, MA, 1987; pp 318362

(78) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.

(79) Werbos, P. J. Generalization of backpropagation with application to a recurrent gas market model. *Neural Netw.* **1988**, *1*, 339.

(80) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta* **1979**, *53*, 75−93.

(81) Guzman, H. V.; Tretyakov, N.; Kobayashi, H.; Fogarty, A. C.; Kreis, K.; Krajniak, J.; Junghans, C.; Kremer, K.; Stuehn, T. ESPResSo++ 2.0: Advanced methods for multiscale molecular simulation. *Comput. Phys. Commun.* **2019**, *238*, 66−76.

(82) McCarty, J.; Parrinello, M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* **2017**, *147*, 204109.

(83) Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, *7*, 95−99.

(84) Trendelkamp-Schroer, B.; Noé, F. Efficient estimation of rare-event kinetics. *Phys. Rev. X* **2016**, *6*, 011009.

(85) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.

(86) Jang, H.; Woolf, T. B. Multiple pathways in conformational transitions of the alanine dipeptide: an application of dynamic importance sampling. *J. Comput. Chem.* **2006**, *27*, 1136.

(87) Chekmarev, D. S.; Ishida, T.; Levy, R. M. Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous and discrete-state kinetic models. *J. Phys. Chem. B* **2004**, *108*, 19487.

(88) Wang, H.; Schütte, C.; Ciccotti, G.; Site, L. D. Exploring the conformational dynamics of alanine dipeptide in solution subjected to an external electric field: A nonequilibrium molecular dynamics simulation. *J. Chem. Theory Comput.* **2014**, *10*, 1376.