

Here i indexes the atoms, m_i is the mass of atom i , $U(x^p)$ is the potential energy, γ is a friction parameter, and dB_t is a standard Brownian motion process. Starting from an initial state $x(0)$, simulating Equation (2), along with the relationship $dx^p = x^v dt$, yields values of $x(t)$ that are distributed according to the Boltzmann distribution as $t \rightarrow \infty$. Standard MD libraries discretise this SDE with a timestep Δt , which must be chosen to be $\sim 1\text{fs} = 10^{-15}\text{s}$ for stability. Unfortunately, many biomolecules contain metastable states separated by energy barriers that can take milliseconds of MD simulation time ($\sim 10^{12}$ sequential integration steps) to cross, rendering this approach infeasible. To overcome this, prior work has produced an array of *enhanced sampling methods*, such as coarse graining (Clementi, 2008; Kmiecik et al., 2016) and metadynamics (Laio & Parrinello, 2002). However, these methods require domain knowledge specific to each molecular system to implement effectively.

Furthermore, standard MD simulations do not transfer information between molecular systems: for each system studied, a new MD simulation must be performed. This represents a wasted opportunity: many molecular systems exhibit closely related dynamics, and simulating one system should yield information relevant to similar systems. In particular, proteins, being comprised of linear sequences of 20 kinds of amino acids, are prime candidates to study this kind of transferability. We propose *Timewarp*, a general, transferable enhanced sampling method which uses a *normalising flow* (Rezende & Mohamed, 2015) as a proposal distribution for a Markov chain Monte Carlo (MCMC) method targeting the Boltzmann distribution. Our main contributions are:

1. We define an asymptotically unbiased MCMC algorithm targeting the Boltzmann distribution using a conditional normalising flow as a proposal distribution with a Metropolis-Hastings (MH) correction step.
2. We implement a transformer-based flow which acts on all-atom Cartesian coordinates. As it does not use predefined collective variables, it can be applied to general molecules without additional domain knowledge.
3. We produce a dataset of MD trajectories of hundreds of small peptides to train the flow model.
4. We demonstrate *transferability* by showing wall-clock acceleration of MD sampling on small peptides (2-4 amino acids) unseen during training.
5. We show that when deployed *without* the MH correction, Timewarp can be used to explore metastable states of new peptides much faster than MD.

2. Related work

There has recently been a surge of interest in deep learning on molecules. *Boltzmann generators* (Noé et al., 2019; Köhler et al., 2021) use flows to sample from the Boltzmann distribution. There are two ways to generate samples with

Boltzmann Generators: (i) Produce i.i.d. samples from the flow and use statistical reweighting to debias expectation values. (ii) Use the Boltzmann generator in an MCMC framework (Dibak et al., 2022), as in Timewarp. As Boltzmann generators rely on internal coordinates, they do not generalise to multiple proteins, unlike Timewarp. Recently, Xu et al. (2022) proposed *GeoDiff*, a diffusion model that predicts molecular conformations from a molecular graph. Like Timewarp, GeoDiff works in Cartesian coordinates and generalises to unseen molecules. However, GeoDiff was not applied to proteins, but small molecules, and does not target the Boltzmann distribution.

Markov state models (MSMs) (Prinz et al., 2011; Swope et al., 2004; Husic & Pande, 2018) are another related technique. MSMs work by running many short MD simulations, which are used to define a discrete state space, along with an estimated transition probability matrix. Similarly to Timewarp, MSMs estimate the transition probability between the state at a time t and the time $t + \tau$, where $\tau \gg \Delta t$. Recent work has applied deep learning to MSMs, leading to *VAMPnets* (Mardt et al., 2018) and *deep generative MSMs* (Wu et al., 2018), which replace the MSM data-processing pipeline with deep networks. In contrast to Timewarp, these models are not transferable: new MD simulations have to be performed, and new networks trained, for each molecular system. Furthermore, MSMs model the dynamics in a coarse-grained, discrete state space, rather than in the all-atom coordinate representation as with Timewarp.

There has been much previous work on *neural adaptive samplers* (Song et al., 2017; Levy et al., 2018; Li et al., 2021), which use deep generative models as proposal distributions. *A-NICE-MC* (Song et al., 2017) uses a volume-preserving flow trained using a likelihood-free adversarial method. Other methods use objective functions designed to encourage exploration. The entropy term in our objective function is inspired by Titsias & Dellaportas (2019). In contrast to these methods, Timewarp focuses on *generalising* to new molecular systems without retraining the network.

3. Method

Consider the distribution of $x(t + \tau)$ induced by an MD simulation of Equation (2) for a time $\tau \gg \Delta t$, starting from $x(t)$. We denote this conditional distribution by $\mu(x(t + \tau)|x(t))$. Timewarp uses a deep probabilistic model to approximate $\mu(x(t + \tau)|x(t))$ (see Figure 1). Once trained, the model is used in an MCMC method to sample from the Boltzmann distribution.

3.1. Conditional normalising flows

We fit a *conditional normalising flow*, $p_\theta(x(t + \tau)|x(t))$, to $\mu(x(t + \tau)|x(t))$, where θ are learnable parameters. Nor-