Path Integral Stochastic Optimal Control for Sampling Transition Paths

Lars Holdijk [*1 2] **Yuanqi Du** [*2] **Ferry Hooft** [3] **Priyank Jaini** [4] **Bernd Ensing** [3 5] **Max Welling** [2]

Abstract

We consider the problem of *Sampling Transition Paths*. Given two metastable conformational states of a molecular system, e.g. a folded and unfolded protein, we aim to sample the most likely transition path between the two states. Sampling such a transition path is computationally expensive due to the existence of high free energy barriers between the two states. To circumvent this, previous work has focused on simplifying the trajectories to occur along specific molecular descriptors called Collective Variables (CVs). However, finding CVs is not trivial and requires chemical intuition. For larger molecules, where intuition is not sufficient, using these CV-based methods biases the transition along possibly irrelevant dimensions. Instead, this work proposes a method for sampling transition paths that consider the entire geometry of the molecules. To achieve this, we first relate the problem to recent work on the Schrödinger bridge problem and stochastic optimal control. Using this relation, we construct a method that takes into account important characteristics of molecular systems such as second-order dynamics and invariance to rotations and translations. We demonstrate our method on the commonly studied Alanine Dipeptide, but also consider larger proteins such as Polyproline and Chignolin.

1. Introduction

Modeling non-equilibrium systems in natural sciences involves analyzing dynamical behaviour that occur with very low probability known as *rare events*, i.e. particular instances of the dynamical system that are atypical. The kinetics of many important molecular processes, such as phase transitions, protein folding, conformational changes, and chemical reactions, are all dominated by these rare events. One way to sample these rare events is to follow the time evolution of the underlying dynamical system using Molecular Dynamic (MD) simulations until a reasonable number of events has been observed. However, this is computationally highly inefficient due to the large time-scales involved, which are typically related to the presence of high energy and/or entropy barriers between the metastable states. Thus, the main problem is: *How can we efficiently sample trajectories that give rise to these rare but interesting transition events?*

Numerous enhanced sampling methods such as steered MD (Jarzynski, 1997), umbrella sampling (Torrie & Valleau, 1977), constrained MD (Carter et al., 1989), transition path sampling (Dellago & Bolhuis, 2009), and many more, have been developed to deal with the problem of *rare events* in molecular simulation. Most of these methods bias the dynamical system with well-chosen geometric descriptors of the transition, so-called *collective variables* (CVs), that allow the system to overcome high-energy transition barriers and sample these rare events. The performance of these enhanced sampling techniques is critically dependent on the choice of these CVs. However, choosing appropriate CVs for all but the simplest molecular systems is fraught with difficulty, as it relies on human intuition, insights about the molecular system, and trial and error.

A key alternative to sampling these rare transition paths is to model an alternate dynamical system that allows sampling these rare trajectories in an optimal manner (Ahamed et al., 2006; Jack, 2020; Todorov, 2009). (Rose et al., 2021) proposes an RL solution to modelling such a system.

In this paper, we consider the problem of sampling *rare* transition paths by developing an alternative dynamical system using *path integral stochastic optimal control* (Kappen, 2005; 2007; Kappen & Ruiz, 2016; Theodorou et al., 2010). Our method models this alternative dynamics of the system by applying an external control policy to each of the atoms in the molecule. In particular, the method learns the external control policy such that it minimizes the amount of external work needed to overcome the lowest barrier and transition the molecular system from an initial meta-stable state to a final one. The method does not require any knowledge

[*]Equal contribution [1]OATML, University of Oxford [2]AMLab, University of Amsterdam [3]Computational Chemistry Group, University of Amsterdam [4]Google Brain [5]AI4Science, University of Amsterdam. Correspondence to: Lars Holdijk <larsholdijk@gmail.com>.

2nd AI4Science Workshop at the 39th International Conference on Machine Learning (ICML), 2022. Copyright 2022 by the author(s).

of CVs to sample these rare trajectories. Furthermore, we draw connections between sampling rare transition paths and the Schrödinger bridge problem (Schrödinger, 1931; 1932). Subsequently, we demonstrate that stochastic optimal control is well suited to solving these problems by extending the work of (Kappen & Ruiz, 2016) for molecular systems by incorporating Hamiltonian dynamics and equivariance constraints in the model.

Our main contributions in this paper are:

- We demonstrate the equivalence between sampling transition paths, the Schrödinger bridge problem, and path integral stochastic optimal control (§2).
- We extend the PICE algorithm (Kappen & Ruiz, 2016) for molecular dynamics by incorporating inductive biases in the form of equivariances and Hamiltonian dynamics (§3).
- Our proposed method does not require any prior knowledge of CVs, which is important for modeling large and complex molecular transitions for which CVs are unknown (§2-3).
- We demonstrate the efficacy of our method on conformational transitions in three molecular systems of varying complexity, namely Alanine Dipeptide, Polyproline and, Chignolin (§4).

2. Preliminaries and Problem Setup

We first define notations, provide key definitions, and introduce a mathematical framework of our main problem. Subsequently, we will briefly discuss various methods that can be used to solve our main problem and their drawbacks which provides motivation for our proposed method.

Consider a system evolving over time with an initial and final configuration $\boldsymbol{x}_0$ and $\boldsymbol{x}_\tau$ respectively. Our interest here lies in sampling trajectories $\boldsymbol{x}(\tau) := \boldsymbol{x}_{0:\tau}$ of length τ between an initial $\boldsymbol{x}_0$ and final $\boldsymbol{x}_\tau$ configuration. While the evolution of the system is continuous, for convenience we discretize time and represent the trajectory as an ordered sequence of states $\boldsymbol{x}(\tau) = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_\tau\}$.

In our applications of interest, these initial ($\boldsymbol{x}_0$) and final ($\boldsymbol{x}_\tau$) configurations are stable states of the molecular system located at local minima on a free energy surface. For example, these stable states can be reactants and products of a chemical reaction, or native and unfolded states of a protein. These stable states are often separated by high free energy barriers. Thus, the system spends the bulk of its time merely fluctuating within these stable states. Sampling the rare transition paths, $\boldsymbol{x}(\tau)$, that bring the system from one stable state to another involves overcoming these free energy barriers. This problem of sampling transition paths is our main interest in this paper.

Formally, let $\pi_0(\boldsymbol{x}_0)$ be the distribution of states $\boldsymbol{x}_0$ and $\pi_i(\boldsymbol{x}_i|\boldsymbol{x}_{i-1})$ be a Markovian transition kernel. Then, we can define the probability of each trajectory by:

$$\pi(\boldsymbol{x}_{0:\tau}) := \pi_0(\boldsymbol{x}_0) \cdot \prod_{i=1}^{\tau} \pi_i(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}), \tag{1}$$

Thus, $\pi(\boldsymbol{x}_{0:\tau})$ defines the original dynamics of the system.

In this paper, we consider the problem of finding an alternative dynamics of the system that enables us to efficiently sample *optimal* trajectories of the original dynamics given by Equation (1). Concretely, we aim to find another parameterized Markovian dynamical system $\pi_\theta(\boldsymbol{x}_{0:\tau})$ which facilitates optimal sampling of rare trajectories as close to $\pi(\boldsymbol{x}_{0:\tau})$ as possible, i.e.

$$\pi^*_\theta(\boldsymbol{x}_{0:\tau}) = \underset{\pi_\theta(\boldsymbol{x}_{0:\tau})}{\arg\min}\, \mathbb{D}_{\mathsf{KL}}\Big(\pi_\theta(\boldsymbol{x}_{0:\tau}) \| \pi(\boldsymbol{x}_{0:\tau})\Big) \tag{2}$$

2.1. Sampling Transition Paths as a Schrödinger Bridge Problem

This problem of sampling transition paths defined above in Equations (1) and (2) can be solved using the Schrödinger Bridge formulation. The Schrödinger Bridge Problem (SBP) as originally posed by Schrödinger (Schrödinger, 1931; 1932) aims to find a stochastic evolution between two distributions – π_0, the source distribution, and π_τ, the target distribution – that is closest to a Brownian motion prior. Formally, let $\mathbb{P}$ be the Brownian prior defined over the space of path measures $\mathcal{D}(\pi_0, \pi_\tau)$ with marginals π_0 and π_τ. Then, the Schrödinger Bridge Problem can be stated as:

$$\mathbb{Q}^* := \underset{\mathbb{Q} \in \mathcal{D}(\pi_0, \pi_\tau)}{\arg\min}\; \mathbb{D}_{\mathsf{KL}}(\mathbb{Q} \| \mathbb{P}) \tag{3}$$

Evidently, this is exactly the formulation of transition path sampling above with initial state $\boldsymbol{x}_0$ distributed as π_0 and the final configuration $\boldsymbol{x}_\tau$ distributed as π_τ.

In recent years, there has been a surge in works proposing computationally efficient methods for solving the Schrödinger Bridge Problem (Vargas et al., 2021; De Bortoli et al., 2021). However, in this work we take a different approach that lends itself better to modelling the chemical nature of our problem by learning a biased force to be added to the system. In the next section, to set the stage for this novel approach, we relate the problem of sampling transition paths as a path integral stochastic optimal control problem and establish an equivalence between learning alternative dynamical systems for sampling rare events, Schrödinger bridge problem, and stochastic optimal control.

2.2. Sampling Transition Paths through Stochastic Optimal Control

The original dynamics of the system, as given in Equation (1), can be reformulated as a stochastic process:

$$\mathrm{d}\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, t)\,\mathrm{d}t + \boldsymbol{G}(\boldsymbol{x}_t, t)\cdot \mathrm{d}\boldsymbol{\varepsilon}_t, \qquad t \in [0, \tau] \quad (4)$$

where $\boldsymbol{f} : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^d$ and $\boldsymbol{G} : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^{d\times d}$ are deterministic functions representing the drift and volatility of the system. The stochastic process $\boldsymbol{\varepsilon}_t$ is a Brownian motion with variance ν.

We approximate the system dynamics described in Equation (4) with another dynamical system by applying an external bias potential (or control) $\boldsymbol{u}(\boldsymbol{x}_t, t) \in \mathbb{R}^d$ to the system that pushes it over the transition state barriers. We write these dynamics as follows:

$$\mathrm{d}\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, t)\,\mathrm{d}t + \boldsymbol{G}(\boldsymbol{x}_t, t) \cdot \Big(\boldsymbol{u}(\boldsymbol{x}_t, t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{\varepsilon}_t\Big),\ t \in [0, \tau] \quad (5)$$

Given a trajectory $\boldsymbol{x}(\tau) = (\boldsymbol{x}_0, \cdots, \boldsymbol{x}_\tau) \in \mathbb{R}^{\tau\times d}$ generated through the SDE in Equation (5), we define the cost of this trajectory under control $\boldsymbol{u}$ following (Kappen, 2007; Theodorou et al., 2010) as:

$$C(\boldsymbol{x}(\tau), \boldsymbol{u}, \boldsymbol{\varepsilon_t}) = \frac{1}{\lambda}\Big(\varphi(\boldsymbol{x}_\tau) + \sum_{t=0}^{\tau} \frac{1}{2} u(\boldsymbol{x}_t, t)^T \boldsymbol{R} u(\boldsymbol{x}_t, t) + u(\boldsymbol{x}_t, t)^T \boldsymbol{R}\boldsymbol{\varepsilon}_t\Big) \quad (6)$$

where φ denotes the terminal cost, λ is a constant and $\boldsymbol{R}$ is the cost of taking action $\boldsymbol{u}$ in the current state and is generally given as a weight matrix for a quadratic control cost. Note that the inclusion of the cost related to the Brownian motion is unusual but needed for relating the cost to a KL-divergence as we will show next. Additionally, the Brownian cost will vanish under expectation (Thijssen & Kappen, 2015).

The goal then is to find the optimal control $\boldsymbol{u}$ that minimizes the expected cost:

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} \mathbb{E}_{\tau, \boldsymbol{\varepsilon_t}}[C(\boldsymbol{x}(\tau), \boldsymbol{u}, \boldsymbol{\varepsilon_t})] \quad (7)$$

where the expectation is taken over trajectories τ sampled using the SDE under control $\boldsymbol{u}$.

Relation to sampling transition paths: Interestingly, the objective in Equation (7) is exactly related to the problem of sampling transition paths as given in Equation (2). As (Kappen & Ruiz, 2016) establish, Equation (5) defines a probability distribution $\pi_{\boldsymbol{u}}\big(\boldsymbol{x}(\tau)\big)$ over trajectories $\boldsymbol{x}(\tau)$ through:

$$\pi_{\boldsymbol{u}}\big(\boldsymbol{x}(\tau)\big) = \prod_{t=0}^{\tau} \mathcal{N}(\boldsymbol{x}_{t+1} | \boldsymbol{\mu}_t, \Sigma_t) \quad (8)$$

with $\boldsymbol{\mu}_s = x_s + f(\boldsymbol{x}_s, s)\,\mathrm{d}t + \boldsymbol{G}(\boldsymbol{x}_s, s)(u(\boldsymbol{x}_s, s)\,\mathrm{d}t)$ and $\Sigma_s = \lambda \boldsymbol{G}(\boldsymbol{x}_s, s)^T \boldsymbol{R} \boldsymbol{G}(\boldsymbol{x}_s, s)$. For different $\boldsymbol{u}$, these distributions are related through the Girsanov Theorem (Cameron & Martin, 1944). Denoting the distribution over trajectories with no control (i.e. $\boldsymbol{u} = 0$) as $\pi_0\big(\boldsymbol{x}(\tau)\big)$, we obtain:

$$\log \frac{\pi_{\boldsymbol{u}}\big(\boldsymbol{x}(\tau)\big)}{\pi_0\big(\boldsymbol{x}(\tau)\big)} = \frac{1}{\lambda}\sum_{t=0}^{\tau} \frac{1}{2}\boldsymbol{u}(\boldsymbol{x}_t, t)^T \boldsymbol{R}\boldsymbol{u}(\boldsymbol{x}_t, t) + \boldsymbol{u}(\boldsymbol{x}_t, t)^T \boldsymbol{R}\boldsymbol{\varepsilon}_t \quad (9)$$

(Kappen & Ruiz, 2016) show that we can use Equation (9) to rewrite the objective in Equation (7) as:

$$\pi_{\boldsymbol{u}^*} = \arg\min_{\pi_{\boldsymbol{u}}} \mathbb{E}_{\boldsymbol{x}(\tau)\sim\pi_{\boldsymbol{u}}}\big[\frac{1}{\lambda}\varphi(\boldsymbol{x}_\tau)\big] + \mathbb{D}_{\mathsf{KL}}\Big(\pi_{\boldsymbol{u}}\big(\boldsymbol{x}(\tau)\big) \| \pi_0\big(\boldsymbol{x}(\tau)\big)\Big) \quad (10)$$

This objective is equivalent to the Schrödinger bridge formulation in Equation (3) where the terminal cost $\varphi(\boldsymbol{x}_\tau)$ is a constant. Indeed, this connection between the Schrödinger Bridge Problem and stochastic optimal control has previously been established by (Chen et al., 2016; Pavon et al., 2021). Here, through the formulations in Equations 2, 3 & 10, we also establish the equivalence between sampling transition paths, Schrödinger bridge problem, and stochastic optimal control. This allows us to utilize solutions for finding the optimal control in Equation (10) for the aforementioned problems.

Optimal Control Policy: (Kappen & Ruiz, 2016) introduced the Path Integral Cross Entropy (PICE) method for solving Equation (10). The PICE method derives an explicit expression for the optimal policy and distribution $\pi_{\boldsymbol{u}^*}$ when $\lambda = \nu\boldsymbol{R}$ given by

$$\pi_{\boldsymbol{u}^*} = \frac{1}{\eta(\boldsymbol{x}, t)}\pi_{\boldsymbol{u}}\big(\boldsymbol{x}(\tau)\big)\exp(-C(\boldsymbol{x}(\tau), \boldsymbol{u})) \quad (11)$$

where $\eta(\boldsymbol{x}, t) = \mathbb{E}_{\tau\sim\pi_{\boldsymbol{u}}}[\exp(-C(\boldsymbol{x}(\tau), \boldsymbol{u})]$ is the normalization constant. PICE, subsequently, minimizes the KL-divergence between the optimal controlled distribution $\pi_{\boldsymbol{u}^*}$ and a parameterized distribution $\pi_{\boldsymbol{u}_\theta}$ using gradient descent as follows:

$$\frac{\partial \mathbb{D}_{\mathsf{KL}}(\pi_{\boldsymbol{u}^*} | \pi_{\boldsymbol{u}_\theta})}{\partial\theta} = -\frac{1}{\eta}\mathbb{E}_{\boldsymbol{x}(\tau)\sim\pi_{\boldsymbol{u}_\theta}}[\exp(-C(\boldsymbol{x}(\tau), \boldsymbol{u}_\theta))\sum_{0}^{T}(\boldsymbol{R}\varepsilon \cdot \frac{\partial \boldsymbol{u}_\theta}{\partial\theta})] \quad (12)$$

Thus, PICE provides an iterative gradient descent method to learn a parameterized policy (or control) $\boldsymbol{u}_\theta$ and subsequently a distribution over paths $\boldsymbol{x}(\tau)$. We can then use

this learned control, $\boldsymbol{u}_\theta$, to approximate the solution for sampling transition paths as well as the Schrödinger bridge problem.

In this section, we have set up our main problem of sampling rare transition paths and established its relationship to both the Schrödinger bridge problem and stochastic optimal control. Subsequently, we discussed an iterative gradient descent based method for solving the optimal control problem. In the next section, we will extend this iterative algorithm to consider the entire geometry of the molecular system by incorporating Hamiltonian dynamics using an augmented state space $\boldsymbol{x}_t$, and symmetries by learning a policy network $\boldsymbol{u}_\theta$.

3. Path Integral Optimal Control for Sampling Transition Paths

We consider a molecule consisting of n atoms with an initial and final configuration $\boldsymbol{r}_0 \in \mathbb{R}^{3\times n}$ and $\boldsymbol{r}_\tau \in \mathbb{R}^{3\times n}$ i.e. we are given a vector defining the 3D positions of each atom in the molecule. Thus, a direct method to sample transition paths $\boldsymbol{r}(\tau)$ for this problem is to learn a control $\boldsymbol{u}_\theta$ acting directly on the positions $\boldsymbol{r}$ of the molecule using the iterative gradient descent method discussed in Section 2. However, the collective behaviour of the atoms and molecules are governed by classical molecular dynamics i.e. Newtonian equations of motion:

$$\mathrm{d}\boldsymbol{r} = \boldsymbol{v}(t)\,\mathrm{d}t, \quad \text{and,} \quad \mathrm{d}\boldsymbol{v} = \boldsymbol{a}(t)\,\mathrm{d}t$$

where $\boldsymbol{v}(t) \in \mathbb{R}^{3\times n}$ is the velocity and $\boldsymbol{a}(t) \in \mathbb{R}^{3\times n}$ is acceleration given by $\boldsymbol{a}(t) = \nabla_{\boldsymbol{r}}\boldsymbol{U}(\boldsymbol{r})\boldsymbol{m}$ where $\boldsymbol{U}(\boldsymbol{r})$ is the potential energy of the system $\boldsymbol{m}$ is the mass. The potential energy of a system is defined by a heavily parameterized sum of pairwise empirical potential functions, such as harmonic bonds, angle potentials, inter-molecular electrostatic and Van der Waals potentials. Here, we compute this using the OpenMM framework (Eastman et al., 2017). As such, we need to adapt the dynamical system defined in Equation (5) to incorporate these molecular dynamics. Intuitively, we achieve this by limiting the control $\boldsymbol{u} \in \mathbb{R}^{3\times n}$ to act as an additive force on the velocity of the system. This change in velocity via the control $\boldsymbol{u}$ ultimately changes the position of the system.

Incorporating second order dynamics: Formally, we incorporate the second order dynamics of the system defined above by considering an augmented state space: Let $\boldsymbol{x}_0 := (\boldsymbol{r}_0, \boldsymbol{v}_0) \in \mathbb{R}^{3\times n} \times \mathbb{R}^{3\times n}$ be the initial configuration of the system defining the initial positions and velocities of each atom and $\boldsymbol{x}_\tau := (\boldsymbol{r}_\tau, \boldsymbol{v}_\tau)$ be the final configuration. We, thus, model the dynamical system in Equation (5) as:

$$\underbrace{\begin{pmatrix}\mathrm{d}\boldsymbol{r}_t\\ \mathrm{d}\boldsymbol{v}_t\end{pmatrix}}_{\mathrm{d}\boldsymbol{x}_t} = \underbrace{\begin{pmatrix}\boldsymbol{v}_t\\ -\nabla_{\boldsymbol{r}_t}\boldsymbol{U}(\boldsymbol{r}_t)\end{pmatrix}}_{\boldsymbol{f}(\boldsymbol{x}_t,t)}\mathrm{d}t + \underbrace{\begin{pmatrix}\boldsymbol{0}_{3n}\\ \mathbb{I}_{3n}\end{pmatrix}}_{\boldsymbol{G}(\boldsymbol{x}_t,t)} \cdot \Big(\boldsymbol{u}(\boldsymbol{x}_t,t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{\varepsilon}_t\Big), \tag{13}$$

$t \in [0, \tau]$. We note that due to the choice of $\boldsymbol{G}(\boldsymbol{x}_t, t)$ in Equation (13) above, the additional bias force, $\boldsymbol{u}(\boldsymbol{x}_t, t)$, applied to the system only influences the acceleration and velocity of the atoms. $\mathrm{d}\boldsymbol{r}_t$ is solely influenced by the velocity $\boldsymbol{v}_t$, thus conforming to the classical molecular dynamics of the system.

Invariance to rotations and translations: Secondly, the molecules in consideration are invariant w.r.t. translations and 3D rotations i.e. the molecular orientations achieved along a transition path need to incorporate this equivariance w.r.t. $\mathsf{SE}(3)$ group. For this purpose, we need to make the terminal cost function, $\varphi(\boldsymbol{x}_\tau)$, in Equation (6) to be equivariant. We enforce this by defining the terminal cost as the exponentiated pairwise distance between atoms which is commonly used distance metric (Shi et al., 2021) that is invariant to rotations and translations i.e. $\varphi(\boldsymbol{r}_t) = \exp\sum_{i,j}^{n}\left(d_{ij}(\boldsymbol{r}_t) - d_{ij}(\boldsymbol{r}_\tau)\right)^2$ where $d_{ij}(\boldsymbol{r}_t) = \|(\boldsymbol{r}_t)_i - (\boldsymbol{r}_t)_j\|_2^2$.

Algorithm 1 Training Policy $\boldsymbol{u}_\theta$

Input: $\boldsymbol{r}_0, \boldsymbol{r}_T$: *Initial and target molecular positions*,
$\boldsymbol{U}(\cdot)$: *Potential Energy function*,
$\varphi(\cdot)$: *Terminal cost*,
$\boldsymbol{u}_\theta(\cdot,\cdot)$: *Initial parameterized policy*,
N: *Number of trajectories sampled per update*,
τ: *Time horizon*,
ν: *Variance of Brownian noise*,
$\boldsymbol{R}$: *Control cost matrix*,
μ: *Learning rate*
while *not converged* **do**
  ▷ *Generate trajectories with current policy* $\boldsymbol{u}_\theta$
  $\lambda \leftarrow \boldsymbol{R}\nu$ $n \leftarrow 0$ **while** n ¡ N **do**
    ▷ *Initialize initial trajectory state*
    $(\boldsymbol{r}_{n,t}, \boldsymbol{v}_{n,t}, t) \leftarrow (\boldsymbol{r}_0, \boldsymbol{0}, 0)$ **while** $t < T$ **do**
      ▷ *Sample Brownian noise and action*
      $\varepsilon_{n,t} \sim \mathcal{N}(0,\nu)$ $\boldsymbol{u}_{n,t} \leftarrow \boldsymbol{u}_\theta(\boldsymbol{r}_{n,t}, t)$ ▷ *Update positions and velocity*
      $\Delta\boldsymbol{r}_{n,t} \leftarrow \boldsymbol{v}_{n,t}\cdot\Delta t$ $\Delta\boldsymbol{v}_{n,t} \leftarrow -\big(\nabla_{\boldsymbol{r}_{n,t}}\boldsymbol{U}(\boldsymbol{r}_{n,t}) + \boldsymbol{u}_{n,t} + \varepsilon_{n,t}\big)\cdot\Delta t$ $t \leftarrow t + \Delta t$
    end
    ▷ *Determine trajectory cost and gradient*
    $C_n \leftarrow \frac{1}{\lambda}(\varphi(\boldsymbol{r}_{n,T}) + \sum_{i=0}^{T}\boldsymbol{u}_{n,i}^T\boldsymbol{R}\boldsymbol{u}_{n,i} + \boldsymbol{u}_{n,i}^T\boldsymbol{R}\varepsilon_{n,i})$
    $\Delta\theta_n \leftarrow \exp(-C_n) + \sum_{i=0}^{T}\frac{\partial\boldsymbol{u}_{n,i}}{\partial\theta}$ $n \leftarrow n+1$
  end
  ▷ *Determine gradient normalization and perform policy update*
  $\eta \leftarrow \sum_{i=0}^{N}\exp(-C_i)$ $\theta \leftarrow \theta + \frac{\mu}{\eta}\sum_{i=0}^{N}\Delta\theta_i$
end

Physics inspired policy network ($\boldsymbol{u}_\theta$): The main learnable component of our path integral stochastic optimal control method for sampling transition paths is the policy network $\boldsymbol{u}_\theta$. Following the discussion above and formalized in Equation (13), we can interpret the control $\boldsymbol{u}_\theta$ as an additive bias force applied to the system. In this work, we consider two different design approaches to modelling $\boldsymbol{u}_\theta$. In our first approach, we model $\boldsymbol{u}_\theta$ as a neural network that predicts the bias force on the system in which case the velocity evolves as $\mathrm{d}\boldsymbol{v} = (\nabla_{\boldsymbol{r}_t}\boldsymbol{U}(\boldsymbol{r}_t)+\boldsymbol{u}_{\theta,t})\,\mathrm{d}t$. Alternatively, in our second approach, we model $\boldsymbol{u}_\theta$ as a network predicting the bias potential energy. In this case, the corresponding force, $\boldsymbol{F}(\boldsymbol{r}_t)$,applied to the system is calculated by backpropagating through the network, $\boldsymbol{F}(\boldsymbol{r}_t) := \nabla_{\boldsymbol{r}_t}\boldsymbol{u}_{\theta,t}$. The change in velocity is then given by $\mathrm{d}\boldsymbol{v} = \big(\nabla_{\boldsymbol{r}_t}\boldsymbol{U}(\boldsymbol{r}_t) + \boldsymbol{F}(\boldsymbol{r}_t)\big)\,\mathrm{d}t$. Additionally, $\boldsymbol{u}_\theta$ could be implemented using recent advances in physics inspired equivariant neural networks (Cohen & Welling, 2016; Satorras et al., 2021) that take into account the $\mathsf{SE}(3)$ symmetry of the system. We provide complete details for training the control network $\boldsymbol{u}_\theta$ in Algorithm 1.

4. Experiments

We evaluate our path integral stochastic optimal control method for sampling transition paths with three different molecular systems, namely (i) **Alanine Dipeptide**, a small amino acid with well-studied transition paths, (ii) **Polyproline**, a small protein with two distinct conformations with different helix orientations, and (iii) **Chignolin**, an artificial mini-protein studied to understand the folding process of proteins. We begin by detailing the experimental setup below.

Molecular Dynamics Simulation: We simulate the molecular dynamics following Equation (13). For the drift function $\boldsymbol{f}(\boldsymbol{x}_t, t)$, we evaluate the potential function (and subsequently the gradient) using the OpenMM engine (Eastman et al., 2017). We define the forcefield used by OpenMM for each considered molecular system in their respective following sections. Furthermore, we use the Velocity Verlet with Velocity Randomization (VVVR) integrator (Sivak et al., 2013) within OpenMM at a temperature of $300\,\mathrm{K}$ with a collision rate of $1.0\,\mathrm{ps}^{-1}$ wherein the bias potential, $\boldsymbol{u}(\boldsymbol{x}_t, t)$, given by the policy network is integrated as a custom external force.

Policy Network, $\boldsymbol{u}(\boldsymbol{x}_t, t)$: We implement the policy network as a 6 layer MLP with ReLU activation for all our experiments below. The width of the layers of the policy network is dependent on the number of atoms in the molecule under consideration. We implement all code in Pytorch. We ran the experiments on a single GPU (either an NVIDIA RTX3080 or RTX2080) and our code is available on GitHub.

4.1. Alanine Dipeptide

Alanine Dipeptide is an extensively studied molecule (Tobias & Brooks III, 1992; Rossky & Karplus, 1979; Head-Gordon et al., 1991; Swenson et al., 2018) for developing and testing enhanced sampling methods due to ready availability of its two CVs (ϕ, ψ). The conformation transition for Alanine Dipeptide can thus be understood in terms of these two dihedral angles ϕ and ψ as displayed in Figure 1 A. Prior work has, thus, focused on transforming from the initial configuration (see Figure 1.A) to the final configuration (Figure 1.E) by rotating these CVs. As we discussed previously, a major advantage of our method is that we do not require the knowledge of CVs to sample a transition path. However, in our experiment for Alanine Dipeptide, we will use these CVs to compare the quality of the trajectory sampled by our method.

OpenMM and policy network setup: We use the amber 99sb-ildn force field (Lindorff-Larsen et al., 2010) without any solvent, a time-step of $1.0\,\mathrm{fs}$ for the VVVR integrator and a cutoff of $1\,\mathrm{nm}$ for the Particle Mesh Ewald (PME) method (Essmann et al., 1995). For our experiment, we consider both the design choices for the policy network, $\boldsymbol{u}_\theta(\boldsymbol{x}_t, t)$, discussed in Section 3 i.e. directly predicting the force and predicting the energy. We trained the policy networks for 15,000 roll-outs with a time horizon of $500\,\mathrm{fs}$ each consisting of 16 samples. A gradient update was made to the policy network after each roll-out with a learning rate of 10^{-5}. The Brownian motion has a standard deviation of 0.1.

Baseline and evaluation metrics: We compare our method to MD simulations with extended time-horizon and increased system temperatures to sample transition paths. To our knowledge, there are no fixed quantitative metrics in the literature to compare different methods that sample transition paths. Thus, we introduce here three metrics to evaluate the quality of transition paths: (i) *Expected Pairwise Distance* (EPD) measures the euclidean distance between the final conformation in the trajectory and the target conformation, reflecting the goal of the transition to end in the target state, (ii) *Target Hit Percentage* (THP) assures that the final configuration is also close in terms of CVs by measuring the percentage of trajectories correctly transforming these CVs, and (iii) *Energy Transition Point* (ETP) which evaluates the capacity of each method to find transition paths that cross the high-energy barrier at a low point by taking the maximum potential energy of the molecule along the trajectory. A good trajectory will be one that passes through the minimal high-energy barrier and ETP aims to measure this.

Results: We first visualize the trajectory generated by the *energy prediction* policy in Figure 1. The trajectory in Figure 1 demonstrates that the control policy transforms the

	τ fs	Temp. K	EPD (↓) nm $\times 10^{-3}$	THP (↑) %	ETP (↓) kJ mol^{-1}
Force Prediction	500	300	2.07	41.1 %	0.68
Energy Prediction	500	300	1.25	89.2 %	-5.21
MD w. fixed timescale	500	300	7.92	0%	-
	500	1500	7.47	0%	-
	500	4500	6.33	0%	-
	500	9000	6.82	1.7 %	1019.83
MD w/ fixed timescale	34810	1500	1.88	100%	551.51
	48683	4500	2.01	100%	1647.35

Table 1. Benchmark scores for the proposed method and extended MD baselines. From-left-to-right: Time-horizon τ representing the trajectory length (note that we take one policy step every 1 fs), simulation temperature, Expected Pairwise distance (EPD), Target Hit Percentage (THP), and Energy Transition Point (ETP). ETP can only be calculate when a trajectory reaches the target. All metrics are averaged over 1000 trajectories except for MD w/ fixed timescale which is ran only for 10 trajectories.

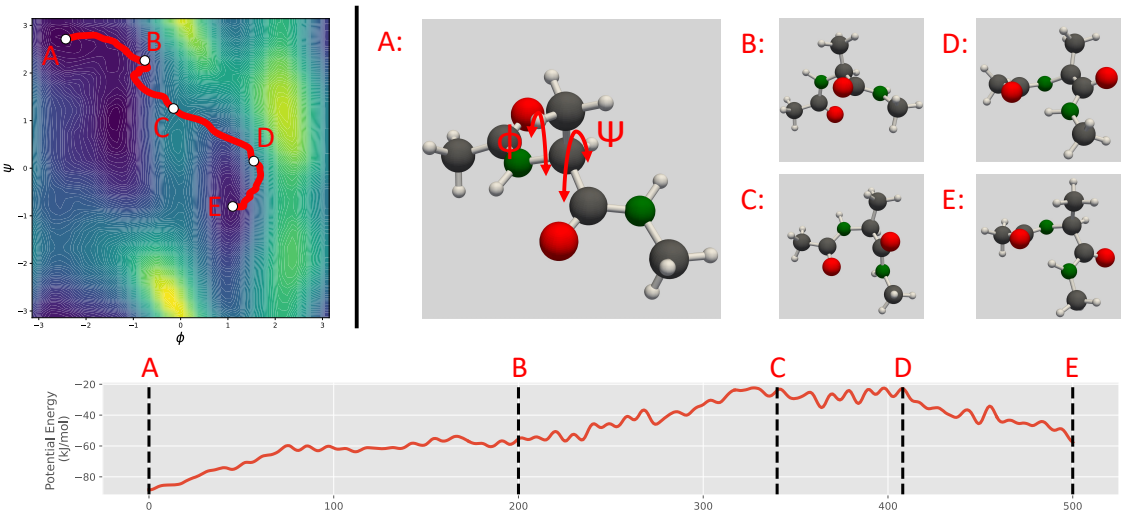

Figure 1. Visualization of a trajectory sampled with the proposed method. *Left:* The sampled trajectory projected on the free energy landscape of Alanine Dipeptide as a function of two CVs *Right:* Conformations along the sampled trajectory: A) starting conformation showing the CV dihedral angles, B-D) intermediate conformations with C being the highest energy point on the trajectory, and E) final conformation, which closely aligns with the target conformation. *Bottom:* Potential energy during transition. Letters represent the same configurations in the transition.

molecule from the initial position (A) to the final position (E) by transitioning over the barrier with the least energy at (C). Interestingly, the trajectory follows the expected transitions in the CVs without them being explicitly specified e.g. the transition path visualized on the left in Figure 1 shows that the molecule first rotates the dihedral angle associated with CV ϕ in (A $\rightarrow$ B), then gradually rotates along both ψ and ϕ in (B $\rightarrow$ C $\rightarrow$ D), and finally rotates ψ in (D $\rightarrow$ E) to reach the final configuration. As expected, we observe that the potential energy goes up during the transition until it reaches the top of the energy barrier (C). After this point, the molecule settles down in its new low-energy state.

Next, we compare the performance of the trajectories sampled using the *force* and *energy* predicting policy networks with MD simulations on the metrics introduced before in Section 4. We find that the trajectories generated by both the policy networks outperform the MD baselines, but the more physics-aligned energy predicting policy performs best under our metrics. This policy network consistently reaches the target conformation both in terms of full geometry and the CVs orientation. Furthermore, our policy network generates these trajectories in a significantly shorter time than temperature enhanced MD simulations without a fixed timescale. When we do limit MD to run for the same timescale as the proposed method, we found that, in contrast to the proposed method, temperature enhanced MD simulations are unable to generate successful trajectories.

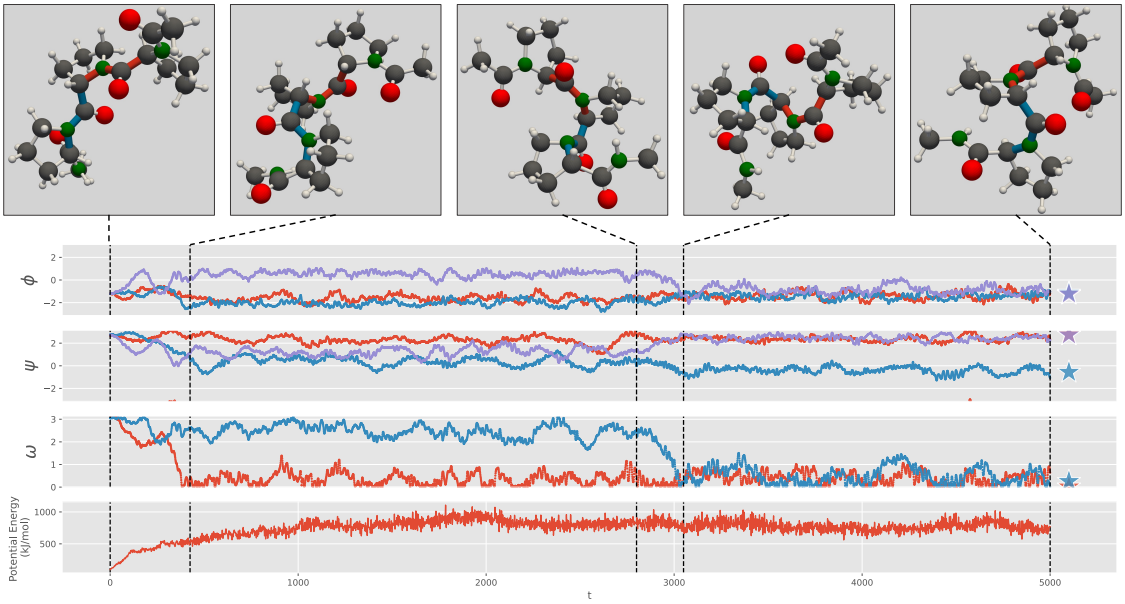

Figure 2. Visualization of the Polyproline transformation from PP-II to PP-I. *From-top-to-bottom* 5 stages of the transition, ψ CVs, ϕ CVs, ω CVs, and Potential Energy. For the CVs multiple instances of the same dihedral angles can be found in a single molecule. Stars indicate target CV states. Colored bonds represent the bonds involved in the ω CV.

4.2. Polyproline Helix

Polyproline is a helix-shaped protein structure that consists of repeating proline residues. Polyproline helix can form two different conformations namely *Polyproline-I* (PP-I) and *Polyproline-II helix* (PP-II) (Moradi et al., 2009; 2010). These conformations can be distinguished by their respective helix rotation. PP-I forms a compact right-handed helix due to its peptide bonds having cis-isomers while PP-II has trans-isomer peptide bonds and forms a left-handed helix. Furthermore, the backbone of the polyproline helix also contains two different dihedral angles. We will refer to these peptide bonds and dihedral-angles as the ω, ϕ and ψ CVs respectively. Polyproline can have varying lengths due to its repeated structure. In our experiment, we consider the polyproline trimer with 3 proline residues transitioning from PP-II to PP-I.

OpenMM and policy network setup: We initialize OpenMM with the amber protein.ff14SBonlysc forcefield and gbn2 as the implicit solvent forcefield. The VVVR integrator had a timestep of $2.0\,\mathrm{fs}$ and a cutoff of $5\,\mathrm{nm}$ for PME. The proposed method was ran for a total of $10.000\,\mathrm{fs}$ (resulting in 5,000 policy steps). The policy network was trained over 500 rollouts with 25 samples each using a learning rate of 3×10^{-5} and a standard deviation of 0.1 for the Brownian motion.

Results: We visualize the transformation of the three collective variables (ω, ϕ, ψ) as well as the corresponding potential energy of the conformation in Figure 2 for a sampled transition path from our trained policy network. The ω CV admits the biggest change for the transition from PP-I going from $180°$ to $0°$. We observe that the transition path sampled by our method aligns with the expected changes in CVs in spite of our method not containing any knowledge about these CVs. Figure 2 shows that the peptide bonds transition from a trans-isomer to a cis-isomer state at steps 450 and 3,000. We notice the biggest changes in CVs at these steps in Figure 2. We also note that in addition to the change in the peptide bonds, the final conformation differs from the initial in one of the ψ-dihedral angles. Technically, PP-I has ψ-dihedral angles similar to PP-II, but as a result of the inherent noise of MD our target conformation was sampled with a slight rotation here as well. Interestingly, our method successfully learned to sample transition paths terminating in a similar perturbed state. This indicates that our proposed method is resilient to target states not having a minimal-energy configuration.

4.3. Chignolin

Chignolin is a small β-hairpin protein constructed artificially to study protein folding mechanisms (Honda et al., 2004; Seibert et al., 2005). Due to its small size, its folding process is easier to study than larger scale proteins while being similar enough to shed light on this complex process. In contrast to Alanine Dipeptide and Polyproline, there is no agreement on the transition mechanism describing the folding of Chignolin. Both the CVs involved (Satoh et al., 2006; Paissoni & Camilloni, 2021), as well as the sequence of steps (Harada & Kitao, 2011; Satoh et al., 2006; Suenaga

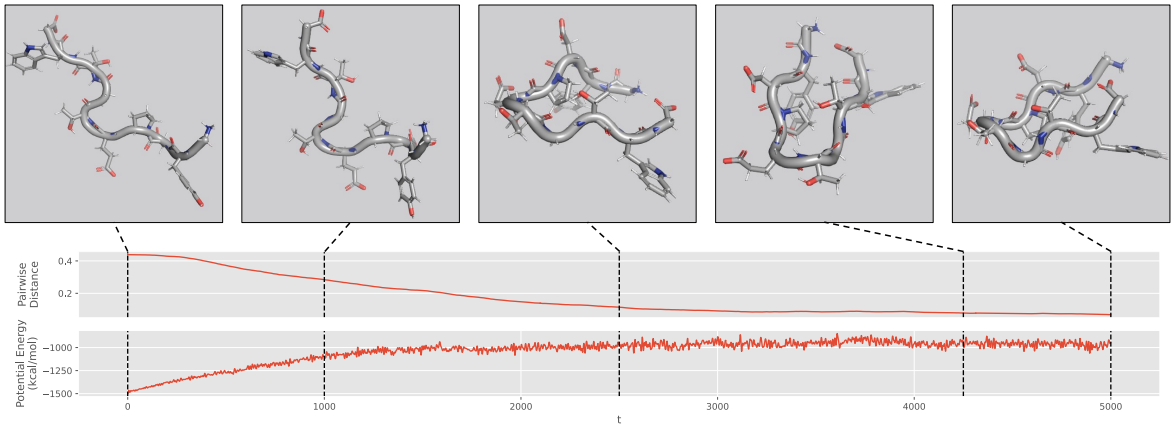

Figure 3. Visualization of the Chignolin folding process. *Top:* 5 stages of the folding process, *Middle:* Pairwise distance wrt to the target conformation of the molecule, *Bottom:* Potential Energy.

et al., 2007; Enemark et al., 2012) describing the folding process have multiple different interpretations. Thus, methods that do not require prior knowledge of CVs are particularly useful to study this protein.

OpenMM and policy network setup: To sample transition paths between the folded and unfolded state of the Chignolin protein, we initialize OpenMM using the same forcefield and VVVR integrator as for Polyproline with the exception that we sample a new force from our policy network every $1.0\,\text{fs}$. We do this 5000 times for each rollout for a total time horizon of $5000\,\text{fs}$. Note that the typical folding time of Chignolin is recorded to be $0.6\,\mu\text{s}$ (Lindorff-Larsen et al., 2011). The policy network is trained over 500 rollouts of 16 samples with a learning rate of 1×10^{-4} and standard deviation of 0.05 for the Brownian motion.

Results: In Figure 3, we visualize the transition of Chignolin at 5 different timesteps during the transition path. We observe that to transition the protein from its low energy unfolded state to the folded conformation, the proposed method guides the protein into a region of higher energy. This increase is initially more steep (0→1500) than in the later stages. Additionally, most of the finer-grained folding (2500→4000) occurs with a high potential energy before settling into the lower-energy folded state. We notice that for the restricted folding time we use in our experiments ($5000\,\text{fs}$ vs $0.6\,\mu\text{s}$), the molecule does not end at the final configuration but reaches close to it as shown by the plot on pairwise distance. Furthermore, the learned policy network is able to transition through the high energy transition barrier in this restricted time. We do not encounter this for molecules with a shorter natural transition time (as illustrated by the potential energy of Alanine Dipeptide in Figure 1).

5. Discussion, Limitations, and Future Work

In this work, we proposed a path integral stochastic optimal control method for the problem of sampling rare transition paths for molecular systems that incorporates the Hamiltonian dynamics and equivariance of the system. In passing, we showed an equivalence between the problem of sampling transition paths, stochastic optimal control, and the Schrödinger bridge problem. We empirically tested our method on three different molecular systems of varying sizes and demonstrated that it was able to sample transition paths on the full geometry of the system without biasing along CVs.

One observed limitation of the proposed method is that for molecules with long natural transition times, we observe the transitions to not converge to the configuration of minimal energy after crossing the high-energy transition barrier. We hypothesize that this is due to the method operating on a reduced time horizon (e.g. $5000\,\text{fs}$ instead of $0.6\,\mu\text{s}$ in the case of Chignolin), or due to the terminal control cost function not requiring the velocity to be zero at the end of the transition. Nevertheless, we note that the method is successful in transitioning the molecules over the high energy barriers as exemplified by the known CVs changing appropriately.

There are many exciting directions for future work. Our experiments showed that the molecules transitioned along with the CVs correctly in spite of not having any information about the CVs. It will be interesting to see if we can infer these CVs from the learned policy and dynamics of the systems. Secondly, our method can have implications for training diffusion models within a fixed time-scale by additionally learning the control policy to transform one distribution into another.

References

Ahamed, T. I., Borkar, V. S., and Juneja, S. Adaptive importance sampling technique for Markov chains using stochastic approximation. *Operations Research*, 54(3): 489–504, 2006.

Cameron, R. H. and Martin, W. T. Transformations of weiner integrals under translations. *Annals of Mathematics*, pp. 386–396, 1944.

Carter, E., Ciccotti, G., Hynes, J. T., and Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters*, 156(5):472–477, 1989.

Chen, Y., Georgiou, T. T., and Pavon, M. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

Cohen, T. and Welling, M. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

De Bortoli, V., Thornton, J., Heng, J., and Doucet, A. Diffusion Schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34, 2021.

Dellago, C. and Bolhuis, P. G. Transition path sampling and other advanced simulation techniques for rare events. *Advanced computer simulation approaches for soft matter sciences III*, pp. 167–233, 2009.

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7): e1005659, 2017.

Enemark, S., Kurniawan, N. A., and Rajagopalan, R. β-Hairpin forms by rolling up from C-terminal: Topological guidance of early folding dynamics. *Scientific Reports*, 2 (1):1–6, 2012.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

Harada, R. and Kitao, A. Exploring the folding free energy landscape of a β-hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method. *The Journal of Physical Chemistry B*, 115(27):8806–8812, 2011.

Head-Gordon, T., Head-Gordon, M., Frisch, M. J., Brooks III, C. L., and Pople, J. A. Theoretical study of blocked glycine and alanine peptide analogs. *Journal of the American chemical society*, 113(16):5989–5997, 1991.

Honda, S., Yamasaki, K., Sawada, Y., and Morii, H. 10 residue folded peptide designed by segment statistics. *Structure*, 12(8):1507–1518, 2004.

Jack, R. L. Ergodicity and large deviations in physical systems with stochastic dynamics. *The European Physical Journal B*, 93(4):1–22, 2020.

Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

Kappen, H. J. Path integrals and symmetry breaking for optimal control theory. *Journal of statistical mechanics: theory and experiment*, 2005(11):P11011, 2005.

Kappen, H. J. An introduction to stochastic control theory, path integrals and reinforcement learning. In *AIP conference proceedings*, volume 887, pp. 149–181. American Institute of Physics, 2007.

Kappen, H. J. and Ruiz, H. C. Adaptive importance sampling for control and inference. *Journal of Statistical Physics*, 162(5):1244–1266, 2016.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

Moradi, M., Babin, V., Roland, C., Darden, T. A., and Sagui, C. Conformations and free energy landscapes of polyproline peptides. *Proceedings of the National Academy of Sciences*, 106(49):20746–20751, 2009.

Moradi, M., Babin, V., Roland, C., and Sagui, C. A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *The Journal of chemical physics*, 133(12):09B614, 2010.

Paissoni, C. and Camilloni, C. How to determine accurate conformational ensembles by metadynamics metainference: a chignolin study case. *Frontiers in molecular biosciences*, 8:491, 2021.

Pavon, M., Trigila, G., and Tabak, E. G. The Data-Driven Schrödinger Bridge. *Communications on Pure and Applied Mathematics*, 74(7):1545–1573, 2021.

Rose, D. C., Mair, J. F., and Garrahan, J. P. A reinforcement learning approach to rare trajectory sampling. *New Journal of Physics*, 23(1):013013, 2021.

Rossky, P. J. and Karplus, M. Solvation. A molecular dynamics study of a dipeptide in water. *Journal of the American Chemical Society*, 101(8):1913–1937, 1979.

Satoh, D., Shimizu, K., Nakamura, S., and Terada, T. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS letters*, 580(14):3422–3426, 2006.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. *arXiv preprint arXiv:2102.09844*, 2021.

Schrödinger, E. *Über die umkehrung der naturgesetze*. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u . . . , 1931.

Schrödinger, E. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. In *Annales de l'institut Henri Poincaré*, volume 2, pp. 269–310, 1932.

Seibert, M. M., Patriksson, A., Hess, B., and Van Der Spoel, D. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *Journal of molecular biology*, 354(1):173–183, 2005.

Shi, C., Luo, S., Xu, M., and Tang, J. Learning gradient fields for molecular conformation generation. In *International Conference on Machine Learning*, pp. 9558–9568. PMLR, 2021.

Sivak, D. A., Chodera, J. D., and Crooks, G. E. Using nonequilibrium fluctuation theorems to understand and correct errors in equilibrium and nonequilibrium simulations of discrete Langevin dynamics. *Physical Review X*, 3(1):011007, 2013.

Suenaga, A., Narumi, T., Futatsugi, N., Yanai, R., Ohno, Y., Okimoto, N., and Taiji, M. Folding dynamics of 10-residue β-hairpin peptide chignolin. *Chemistry–An Asian Journal*, 2(5):591–598, 2007.

Swenson, D. W., Prinz, J.-H., Noe, F., Chodera, J. D., and Bolhuis, P. G. OpenPathSampling: A Python framework for path sampling simulations. 1. Basics. *Journal of chemical theory and computation*, 15(2):813–836, 2018.

Theodorou, E., Buchli, J., and Schaal, S. A generalized path integral control approach to reinforcement learning. *The Journal of Machine Learning Research*, 11:3137–3181, 2010.

Thijssen, S. and Kappen, H. Path integral control and state-dependent feedback. *Physical Review E*, 91(3):032104, 2015.

Tobias, D. J. and Brooks III, C. L. Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results. *The Journal of Physical Chemistry*, 96(9):3864–3870, 1992.

Todorov, E. Efficient computation of optimal actions. *Proceedings of the national academy of sciences*, 106(28): 11478–11483, 2009.

Torrie, G. M. and Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23 (2):187–199, 1977.

Vargas, F., Thodoroff, P., Lamacraft, A., and Lawrence, N. Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021.