PINN-MEP: CONTINUOUS NEURAL REPRESENTATIONS FOR MINIMUM-ENERGY PATH DISCOVERY IN MOLECULAR SYSTEMS

Magnus Petersen
Goethe University Frankfurt
Frankfurt Institute for Advanced Studies
mapetersen@fias.uni-frankfurt.de

Roberto Covino
Goethe University Frankfurt
Frankfurt Institute for Advanced Studies
covino@fias.uni-frankfurt.de

ABSTRACT

Characterizing conformational transitions in physical systems remains a fundamental challenge in the computational sciences. Traditional sampling methods like molecular dynamics (MD) or MCMC often struggle with the high-dimensional nature of molecular systems and the high energy barriers of transitions between stable states. While these transitions are rare events in simulation timescales, they often represent the most biologically significant processes - for example, the conformational change of an ion channel protein from its closed to open state, which controls cellular ion flow and is crucial for neural signaling. Such transitions in real systems may take milliseconds to seconds but could require months or years of continuous simulation to observe even once. We present a method that reformulates transition path generation as a continuous optimization problem solved through physics-informed neural networks (PINNs) inspired by string methods for minimum-energy path (MEP) generation. By representing transition paths as implicit neural functions and leveraging automatic differentiation with differentiable molecular dynamics force fields, our method enables the efficient discovery of physically realistic transition pathways without requiring expensive path sampling. We demonstrate our method's effectiveness on two proteins, including an explicitly hydrated bovine pancreatic trypsin inhibitor (BPTI) system with over 8,300 atoms. Our approach reproduces the same conformational change captured in a landmark millisecond-scale explicit-solvent MD simulation (Shaw et al., 2010), while achieving remarkable computational efficiency, requiring only $\sim 480,000$ force field evaluations compared to the approximately 412 billion evaluations in the original study. This represents a reduction of nearly six orders of magnitude, allowing us to generate the transition pathway in just 15 minutes on a standard GPU rather than weeks on specialized hardware.

1 INTRODUCTION

Understanding molecular conformational changes is fundamental to our understanding of biological processes, from enzyme catalysis to protein folding. Although molecular dynamics (MD) simulations provide unprecedented atomic-level detail of these processes (Dror et al., 2012), capturing rare transition events remains computationally challenging due to the inherent timescale gap between molecular motion and conformational changes. This difficulty arises from the fact that transition rates between stable states are exponentially suppressed by the height of the energy barriers that separate them, as described by transition state theory (Eyring, 1935). As a result, direct simulation methods spend the vast majority of computational effort sampling stable states, making the observation of transitions between them prohibitively expensive.

The search for efficient methods to identify transition pathways between stable states has a rich history in computational physics. Early theoretical work established key principles through the Freidlin-Wentzell (Freidlin & Wentzell, 1998) and Onsager-Machlup (Onsager & Machlup, 1953) functionals, which connect the most probable transition paths to minimizers of specific action functionals in different limiting regimes. Building on these foundations, the minimum-energy path (MEP) between

stable molecular states represents the most probable transition pathway in the limit of high friction. This path reveals the mechanism of conformational change and identifies key transition states and intermediate configurations, and can thus be highly informative about the dynamics of the system. Moreover, the MEP can serve as a scaffold for subsequent sampling of the full transition path ensemble, making it a powerful tool for understanding molecular dynamics. Methods such as the string method (E et al., 2002; 2005; Ren et al., 2005; Maragliano et al., 2006; E et al., 2007; Petersen et al., 2024; Dellago et al., 1998) approximate these paths through a discrete series of molecular configurations, connected by artificial spring forces to ensure continuity or as splines. Although grounded in solid theoretical principles, these approaches face fundamental challenges stemming from their optimization in Cartesian molecular coordinates. The discrete chain of conformations requires periodic reparameterization to maintain even spacing, leading to a complex alternation between energy minimization and geometric constraints. Additional limitations include the difficulty of generating good initial transition guesses and limited parallelization, which restricts the scale of practical applications. While this formulation has proven useful for many small-scale systems, it has not yet fully benefited from modern optimization approaches that could address these challenges.

Concurrent advances in scientific machine learning and molecular simulation have created opportunities for novel approaches to this challenge. Physics-informed neural networks (PINNs) have revolutionized our ability to solve complex physical problems by embedding differential equations directly into neural architectures (Raissi et al., 2019). Additionally, the development of differentiable molecular force fields (Wang et al., 2023) has enabled seamless integration of physical forcefield models used in MD simulations with modern machine learning frameworks, allowing efficient computation of derivatives needed for path optimization.

The last decade has seen numerous approaches emerge to tackle the transition path sampling problem. Traditional shooting methods (Jung et al., 2023; Lazzeri et al., 2023; Bolhuis et al., 2002; Dellago et al., 1998; Laio & Parrinello, 2002; Torrie & Valleau, 1977) rely on expensive molecular dynamics simulations, require initial MD-harvested trajectories to begin with or preexisting knowledge of the transition dynamics, and face inherent tradeoffs between rejection rates and path diversity. More recent machine learning approaches have explored several directions: variational approaches using Doob's h-transform (Du et al., 2024; Das & Limmer, 2019; Singh & Limmer, 2023), stochastic control formulations (Yan et al., 2022; Holdijk et al., 2023), and even attempts to sample full transition trajectories directly using diffusion models (Petersen et al., 2023; Han et al., 2024; Jing et al., 2024). Notably, Du et al. (2024) use a similar neural representation of transition paths but with a different objective, targeting the *full transition path ensemble* via Doob's h-transform. This approach captures additional statistical information at the cost of increased computational expense. Most closely related to our method, Ramakrishnan et al. (2025) employ a continuous neural parameterization of minimum-energy paths inspired by Nudged Elastic Band (NEB), discarding tangential forces in the loss function, regularizing the path velocity, and treating the highest energy point with an additional loss. While both approaches demonstrate the promise of neural network-based path discovery, our method introduces a simpler optimization scheme that scales to large biomolecular systems with explicit solvent while avoiding the higher computational overhead required for either full-path ensemble approaches or NEB-based constraints.

Here, we introduce an approach focused on efficient minimum-energy path discovery as a foundation for understanding molecular transition mechanisms. Our method represents transition paths as continuous, differentiable neural functions that map a progress coordinate to molecular configurations. This representation, inspired by recent advances in neural implicit representations (Mildenhall et al., 2020; Sitzmann et al., 2020), naturally incorporates physical constraints through a physics-informed neural network (PINN) framework. By directly targeting the minimum-energy path through energy minimization, we achieve remarkable computational efficiency, enabling applications to complex biomolecular systems in minutes on standard hardware while extracting the most physically relevant information about transition mechanisms. The resulting pathways can serve as valuable starting points for established transition path sampling methods (Dellago et al., 1998; Bolhuis et al., 2002) when exploration of the complete ensemble is desired.

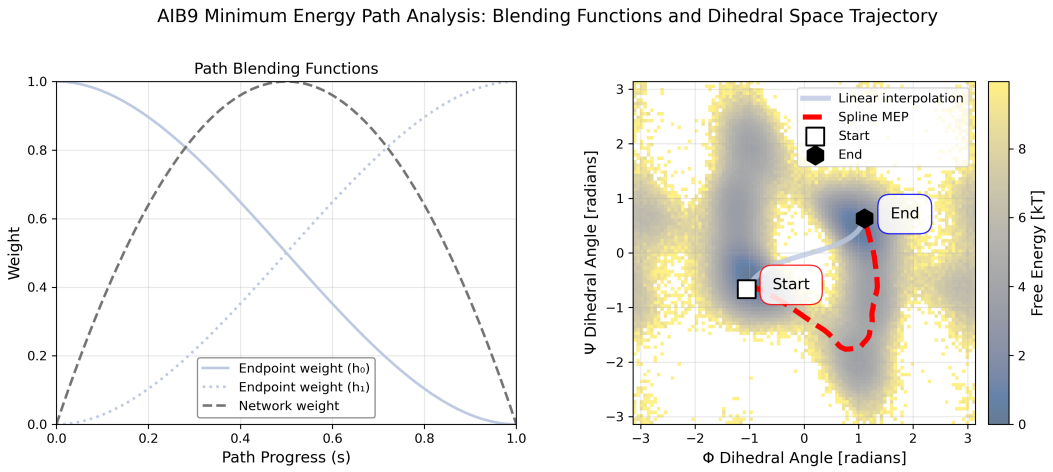

Figure 1: Analysis of our neural path representation method on the AIB9 system. **Left:** The blending functions that control the contribution of different components to the final path. The network weight function (gray, dashed) peaks in the transition region, allowing the neural network to focus on the critical part of the path while ensuring smooth boundary conditions. **Right:** Projection of transition paths onto the ϕ-ψ dihedral angle space, overlaid on the free energy landscape. The neural network path (red, dashed) discovers a physically realistic route through low-energy valleys, while direct linear interpolation (blue) crosses high-energy barriers.

2 Theory and Methods

2.1 Derivation of the Energy-Based Loss Function and Neural Implementation

To our loss function, we lean on alternative formulations of the Onsager-Machlup action functional as derived by Vanden-Eijnden & Heymann (2008); Olender & Elber (1997) in the context of the string methods. We begin with a molecular system evolving under overdamped Langevin dynamics, a well-established model for molecular motion in the high-friction regime characteristic of biomolecular systems:

$$\dot{x}(t) = -\nabla U\big(x(t)\big) + \sqrt{2}\eta(t) \tag{1}$$

$U(x)$ represents the potential energy function, $\eta(t)$ is Gaussian white noise modeling thermal fluctuations, and physical constants are absorbed into the nondimensionalization. To derive the path probability, we discretize the time interval $[0, T]$ into N steps of duration Δt, approximating the dynamics as:

$$x_{i+1} = x_i - \Delta t \nabla U(x_i) + \sqrt{2\Delta t}\xi_i \tag{2}$$

where ξ_i are independent standard Gaussian random variables. The conditional probability of transitioning from x_i to x_{i+1} follows a Gaussian distribution:

To obtain the probability of an entire continuous trajectory, we take the product of these conditional probabilities over all time steps and then take the limit as $\Delta t \to 0$. This yields the path probability:

$$\mathbb{P}[x(t)] \propto \exp\left(-\frac{1}{4}\int_0^T \left\|\dot{x}(t) + \nabla U\big(x(t)\big)\right\|^2 dt\right) \tag{3}$$

The exponent term can be recognized as the Onsager-Machlup action functional, which is defined as:

$$S_{\mathrm{OM}}[x] = \tfrac{1}{4}\int_0^T \left\|\dot{x}(t) + \nabla U\big(x(t)\big)\right\|^2 dt \tag{4}$$

Thus, maximizing the probability of a transition path (subject to fixed endpoints $x(0) = x_A$ and $x(T) = x_B$) is equivalent to minimizing this action functional. The action penalizes paths that deviate from following the deterministic drift $-\nabla U(x)$. To further simplify, we expand the integrand of the action functional:

$$\left\|\dot{x}(t) + \nabla U(x(t))\right\|^2 = \dot{x}(t)\|^2 + 2\dot{x}(t) \cdot \nabla U\big(x(t)\big) + \|\nabla U(x(t))\|^2 \tag{5}$$

Applying the Cauchy-Schwarz inequality to the kinetic and potential terms, we can establish an important relationship:

$$\|\dot{x}(t)\|^2 + \|\nabla U(x(t))\|^2 \geq 2\dot{x}(t) \cdot \nabla U(x(t)) \tag{6}$$

with equality holding if and only if $\dot{x}(t)$ is parallel (or antiparallel) to $\nabla U(x(t))$. When the path is reparameterized so that velocity consistently aligns with the gradient, the action is minimized, yielding the most likely transition pathway. For such pathways where $\dot{x}(t)$ is everywhere proportional to $\nabla U(x(t))$, we can transform the time integral into a purely geometric one. However, to maintain this parallel alignment, the path length itself must be subject to optimization, as a fixed-length path would generally not satisfy the parallelity condition. By defining the arc-length element $ds = \|\dot{x}(t)\|\, dt$ along the spatial curve φ traced out by $x(t)$, we obtain:

$$\int_0^T \|\nabla U(x(t))\| \|\dot{x}(t)\| dt = \int_\varphi \|\nabla U(x)\| ds \tag{7}$$

This geometric reformulation connects directly to the concept of a minimum-energy path (MEP)—the path that follows the potential energy landscape's gradient while avoiding high-energy regions. The MEP represents the most probable transition mechanism in the overdamped limit, making it an ideal target for our optimization.

We represent the transition path as a continuous neural mapping $\varphi_\theta(s)$ for $s \in [0, 1]$ such that $\varphi_\theta(0) = x_A$ and $\varphi_\theta(1) = x_B$, where θ represents the network parameters. By sampling this path at discrete points $\{s_j\}_{j=1}^L$, we can approximate the geometric action by a simple sum of potential energies:

$$\mathcal{L}(\theta) = \sum_{j=1}^{L} U\big(\varphi_\theta(s_j)\big) \tag{8}$$

This discrete approximation serves as our neural network loss function. By minimizing $\mathcal{L}(\theta)$, we drive the path representation to discover low-energy regions connecting stable states, approximating the MEP predicted by the Onsager-Machlup framework.

2.1.1 Neural Representation of minimum-energy Paths

Drawing inspiration from the string method formalism, we propose representing the transition path $\varphi(s)$ using an implicit neural parameterization, similar to the mean parameterization of Du et al. (Du et al., 2024). While our approaches share similar representations, they differ in objectives and computational efficiency (see Appendix B for theoretical connections and performance comparisons, respectively). Specifically, we construct a continuous function mapping a progress coordinate $s \in [0, 1]$ to molecular configurations $\varphi(s) \in \mathbb{R}^{3N}$, where N is the number of atoms:

$$\varphi(s) = h_0(s)x_A + h_1(s)x_B + b(s)f_\theta(s) \tag{9}$$

Here f_θ is a neural network with parameters θ, $h_0(s)$ and $h_1(s)$ are basis functions that satisfy the endpoint constraints $h_0(0) = 1$, $h_0(1) = 0$, $h_1(0) = 0$, and $h_1(1) = 1$, ensuring that $\varphi(0) = x_A$ and

$\varphi(1) = x_B$ regardless of the network's output. While simple linear functions $h_0(s) = (1 - s)$ and $h_1(s) = s$ can be used, other basis functions with appropriate boundary conditions are also effective. The blending function $b(s)$ controls the network's contribution along the path, typically designed to allow maximum influence in the transition region while smoothly vanishing at endpoints (e.g., $b(s) = s(1 - s)$).

2.1.2 Path Optimization

Rather than alternating between evolution and reparameterization steps as in the classical string method, we minimize $\mathcal{L}(\theta)$ directly through stochastic gradient descent. For each batch of paths:

Algorithm 1 Neural minimum-energy Path Optimization

Initialize: $\theta \leftarrow \theta_0$, endpoint configurations $x_A, x_B \in \mathbb{R}^{3N}$, basis functions $h_0(s), h_1(s)$ with $h_0(0) = 1, h_0(1) = 0, h_1(0) = 0, h_1(1) = 1$, and blending function $b(s)$
while not converged **do**
    Sample $\{s_j\}_{j=1}^{L} \sim \mathcal{U}[0, 1]$
    for each s_j in batch **do**
        $\varphi(s_j) = h_0(s_j)x_A + h_1(s_j)x_B + b(s_j)f_\theta(s_j)$
    end for
    Compute $U(\varphi(s_j))$ for each configuration using differentiable force field
    $\mathcal{L}(\theta) = \sum_{j=1}^{L} U(\varphi(s_j))$
    $\theta \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}(\theta)$
end while
return $\varphi(s) = h_0(s)x_A + h_1(s)x_B + b(s)f_\theta(s)$

The implicit neural representation naturally handles reparameterization constraints through its continuous parameterization. Unlike traditional chain-of-states methods, explicit spring-like continuity terms are unnecessary—the inherent continuity of $f_\theta(s)$, combined with smooth blending functions, ensures path continuity without additional constraints.

2.2 Established Approaches to Minimum-Energy Path Finding

Classical approaches to minimum-energy path discovery operate directly in Cartesian atomic coordinates, predating differentiable molecular force fields (Wang et al., 2023) that enable our neural network representation. This direct optimization faces significant challenges as steric clashes create extreme energy barriers that are difficult to overcome with local optimization. The string method and chain-of-states approaches discussed below represent broad families with numerous variations, extensively cataloged in reviews such as "Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events" by E & Vanden-Eijnden (2010). These established techniques illustrate both foundational concepts and limitations that our approach addresses.

2.2.1 String Method

The string method (E et al., 2002; 2005; Ren et al., 2005; Maragliano et al., 2006; E et al., 2007; Ramakrishnan et al., 2025) determines MEPs by evolving a curve in configuration space according to physical forces while maintaining proper parameterization. Given the same endpoints x_A and x_B used in our formulation, the method seeks a curve (string) $\varphi(s)$ parameterized by $s \in [0, 1]$ that connects these states. The MEP satisfies:

$$[\nabla U]^\perp = 0 \tag{10}$$

where $[\nabla U]^\perp = \nabla U - (\nabla U \cdot \hat{\tau})\hat{\tau}$ is the component of the force perpendicular to the unit tangent vector $\hat{\tau} = \varphi'/|\varphi'|$ along the string. This orthogonality condition has an intuitive interpretation: at each point along the MEP, the forces acting on the system are entirely parallel to the path itself, with no perpendicular components trying to push the system away from the path. This represents a ridge or valley in the energy landscape—the system naturally follows the path without being forced sideways.

The simplest dynamics for evolving a curve toward this MEP is given by:

$$\frac{\partial \varphi}{\partial t} = -[\nabla U(\varphi)]^{\perp} + \lambda \hat{\tau} \tag{11}$$

Here $\lambda\hat{\tau}$ is a Lagrange multiplier term enforcing the parameterization constraint, typically equal arc-length parameterization where $|\varphi'(s)| = L$ (constant) for all s. In practice, the string $\varphi(s)$ is discretized into $N+1$ images $\{\varphi_i\}_{i=0}^{N}$ and evolved through a two-step procedure:

1. Evolution step:

$$\dot{\varphi}_i = -\nabla U(\varphi_i) \tag{12}$$

2. Reparameterization step to enforce equal spacing:

$$s_i = \frac{\sum_{j=0}^{i} |\varphi_{j+1} - \varphi_j|}{\sum_{j=0}^{N-1} |\varphi_{j+1} - \varphi_j|} \tag{13}$$

This two-step procedure converges to a discretized MEP satisfying the local orthogonality condition $[\nabla U]^{\perp} = 0$ at each image. Unlike the neural approach presented in this paper, which handles parameterization implicitly, this method requires explicit reparameterization steps that can introduce computational inefficiencies.

2.2.2 Chain-of-States Methods

An alternative perspective views MEP finding as optimization of a chain of states (Pratt, 1986; Elber & Karplus, 1987; Ulitsky & Elber, 1990; Jónsson et al., 1998; Henkelman et al., 2000; Gillilan & Lilien, 2004; Sheppard et al., 2008). Given a sequence $\{x_0, \ldots, x_N\}$ with $x_0 = x_A$ and $x_N = x_B$, the objective function is defined as:

$$E(\{x_i\}_{i=1}^{N-1}) = \sum_{i=1}^{N-1} U(x_i) + \frac{k}{2} \sum_{i=1}^{N} \frac{|x_i - x_{i-1}|^2}{\Delta s^2} \tag{14}$$

where $\Delta s = 1/N$ and k is a spring constant. The spring term enforces continuity of the path while allowing flexibility in the discretization. However, this formulation introduces a well-known limitation: the spring forces tend to cause "corner cutting," where the path takes shortcuts through high-energy regions rather than following the low-energy valleys that represent physically meaningful transitions. The neural representation described earlier avoids this limitation through its natural continuity properties.

At convergence, both formulations yield a discretized MEP satisfying $[\nabla U]^{\perp} = 0$ at each image, the same condition that PINN-MEP targets through energy minimization, but with the advantage of continuous representation and gradient-based optimization.

3 Experiments

We evaluate our method on two molecular systems: the AIB9 peptide and Bovine Pancreatic Trypsin Inhibitor (BPTI). These systems represent different scales of complexity and serve to validate our method's ability to discover physically meaningful transition paths. In Appendix A.1, we compare and benchmark our method to the related approach outlined in (Du et al., 2024), and in Appendix B, we present comprehensive ablation studies analyzing the impact of different embedding strategies, interpolation methods, network architectures, and regularization approaches.

3.1 AIB9 Transition Path Discovery

We first examine the AIB9 system, a 9-residue artificial protein with 129 atoms that exhibits two well-defined metastable states, making it an ideal test case for transition path methods. This system serves as an effective bridge between abstract models and complex proteins: it is small enough for extensive reference simulations yet complex enough to exhibit realistic conformational changes; its energy landscape is well-characterized; and its transitions are easily visualized through dihedral angle projections. Unlike artificial toy problems, AIB9 allows direct validation with actual biomolecular force fields. We use the AMBER ff15ipq-m force field for protein mimetics (Bogetti et al., 2020) implemented in DMFF (Wang et al., 2023), and we generate multiple minimum-energy paths between the system's two metastable states.

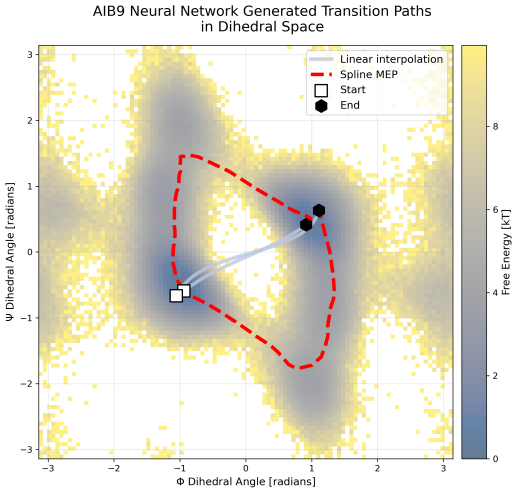

Figure 2: AIB9 free energy surface projected onto central residue dihedral angles (ϕ, ψ). Two distinct minimum-energy paths were found with different random seeds, demonstrating the method's ability to find multiple distinct transition pathways.

Figure 2 shows the discovered transition paths projected onto the ϕ-ψ dihedral angle space of the central residue. The method identifies multiple physically plausible pathways between the states, also observed with long MD simulations. For each path, different conformations of the states were picked as start and end points; this, along with the neural network's random initialization, leads to the discovery of multiple different transitions.

4 BPTI Conformational Change Pathway

Bovine Pancreatic Trypsin Inhibitor (BPTI) is a clinically significant protein used as an anticoagulant in medical procedures. Beyond its therapeutic applications, BPTI has become a fundamental model system for studying protein dynamics due to its moderate size (58 residues) and complex conformational changes involving disulfide bond rearrangements.

To investigate BPTI's conformational transitions, we analyze five key states from a landmark molecular dynamics (MD) simulation performed by D.E. Shaw Research on their Anton supercomputer (Shaw et al., 2010). These publicly available snapshots provide rare insight into BPTI's transition mechanisms. Full details of our BPTI system preparation, explicit solvent shell construction, and hybrid neural network architecture for handling water molecules are provided in Appendix C.

We generate a transition MEP connecting the initial and final BPTI states from the D.E. Shaw simulation. The intermediate snapshots from the D.E. Shaw simulation serve as validation points; they were not used as inputs but represent physically realized conformations from the full trajectory. To analyze this path, we compute two collective variables matching those used in the original D.E. Shaw study: (1) the disulfide torsion angle between Cys14 and Cys38, calculated using the dihedral angle formed by CB14–SG14–SG38–CB38 atoms, and (2) the backbone RMSD of residues 4-54 after mean-centering. Remarkably, our optimized path passes through all intermediate snapshots, indicating that it successfully discovers the same transition mechanism as the computationally intensive explicit-solvent MD simulation.

To ensure a meaningful comparison with the Anton simulation, we matched their simulation conditions by incorporating explicit water molecules in our model. We generated water shells extending 14Å from any protein atom for both the initial and final BPTI conformations. To handle water molecules during transitions, we pair proximal water molecules between start and end states, then model their movement using simple parameterized splines, rather than directly incorporating all water coordinates into the neural network. This hybrid approach allows us to include explicit solvent effects while keeping computational costs low. The specific details of our water shell construction and matching algorithm are described in Appendix C.

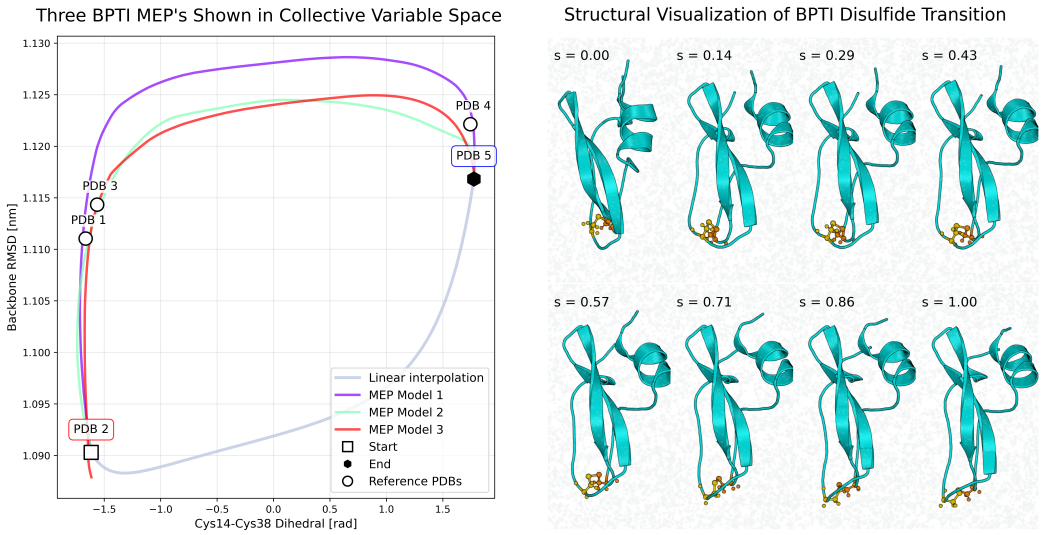

Figure 3: Analysis of our discovered BPTI transition pathway. (a) Evolution of BPTI dynamics projected onto two collective variables matching the original D.E. Shaw study: the Cys14-Cys38 disulfide torsion angle (formed by CB14–SG14–SG38–CB38 atoms) versus the mean-centered backbone RMSD of residues 4-54. Reference structures from the D.E. Shaw simulation are marked with circles. Our optimized path (dashed red line) passes through all intermediate snapshots, demonstrating that it discovers the same transition mechanism observed in the millisecond-scale MD simulation. (b) Structural visualization of the Cys14-Cys38 disulfide bond conformation along our optimized transition path. Each panel represents a different value of the progress coordinate s, showing how the disulfide bond rotates during the transition. This rotation is a key feature of the BPTI conformational change and represents the highest energy barrier in the transition pathway.

The system was modeled using the AMBER99sb-ILDN force field (Lindorff-Larsen et al., 2010) with the TIP3P water model (Jorgensen et al., 1983; Neria et al., 1996). Each minimum-energy path was computed in approximately 15 minutes on a single A6000 GPU, compared to the D.E. Shaw simulation, which required approximately 1.3 million node-hours on their specialized Anton supercomputer to generate the 1.03 millisecond trajectory capturing the same conformational changes. A detailed computational performance comparison is provided in Appendix C.6.

5 Conclusion

We introduced a novel approach to molecular transition path discovery that leverages physics-informed neural networks and differentiable force fields to enable continuous representation of conformational changes. By reformulating minimum-energy path discovery as a neural optimization problem, we demonstrated the method's effectiveness on both the small AIB9 peptide and the explicitly solvated BPTI protein, showing its ability to discover physically meaningful transition pathways across different scales of complexity.

Our approach achieves remarkable computational efficiency, reducing computational requirements by approximately six orders of magnitude compared to traditional MD simulations for the BPTI system. These advances make previously intractable conformational transition studies accessible on standard computational resources.

While our current implementation shows promising results, several directions for improvement remain. A key limitation is the simple feedforward architecture used to represent the transition paths. Future work should explore neural architectures that better respect the structural organization of protein systems. Ramakrishnan et al. (2025) investigate non-uniform sampling schemes along the path, among other innovative design decisions that might further aid convergence. Additionally, extending the method to handle larger multi-domain proteins represents a promising frontier.

REFERENCES

Anthony T. Bogetti, Hannah E. Piston, Jeremy M. G. Leung, Chino C. Cabalteja, Darian T. Yang, Alex J. DeGrave, Karl T. Debiec, David S. Cerutti, David A. Case, W. Seth Horne, and Lillian T. Chong. A twist in the road less traveled: The AMBER ff15ipq-m force field for protein mimetics. *The Journal of Chemical Physics*, 153(6):064101, August 2020. ISSN 0021-9606. doi: 10.1063/5.0019054. URL https://doi.org/10.1063/5.0019054.

Peter Bolhuis, David Chandler, Christoph Dellago, and Phillip Geissler. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annual review of physical chemistry*, 53:291–318, February 2002. doi: 10.1146/annurev.physchem.53.082301.113146.

Avishek Das and David T. Limmer. Variational control forces for enhanced sampling of nonequilibrium molecular dynamics simulations. *The Journal of Chemical Physics*, 151(24):244123, December 2019. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.5128956. URL http://arxiv.org/abs/1909.03589. arXiv:1909.03589 [cond-mat].

Christoph Dellago, Peter G. Bolhuis, Félix S. Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. *The Journal of Chemical Physics*, 108(5):1964–1977, February 1998. ISSN 0021-9606. doi: 10.1063/1.475562. URL https://doi.org/10.1063/1.475562.

Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics*, 41 (1):429–452, 2012. doi: 10.1146/annurev-biophys-042910-155245. URL https://doi.org/10.1146/annurev-biophys-042910-155245. _eprint: https://doi.org/10.1146/annurev-biophys-042910-155245.

Yuanqi Du, Michael Plainer, Rob Brekelmans, Chenru Duan, Frank Noé, Carla P. Gomes, Alán Aspuru-Guzik, and Kirill Neklyudov. Doob's Lagrangian: A Sample-Efficient Variational Approach to Transition Path Sampling, December 2024. URL http://arxiv.org/abs/2410.07974. arXiv:2410.07974 [cs].

Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61:391–420, 2010. ISSN 1545-1593. doi: 10.1146/annurev.physchem.040808.090412.

Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(5):052301, August 2002. doi: 10.1103/PhysRevB.66.052301. URL https://link.aps.org/doi/10.1103/PhysRevB.66.052301. Publisher: American Physical Society.

Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Finite Temperature String Method for the Study of Rare Events. *The Journal of Physical Chemistry B*, 109(14):6688–6693, April 2005. ISSN 1520-6106. doi: 10.1021/jp0455430. URL https://doi.org/10.1021/jp0455430. Publisher: American Chemical Society.

Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *The Journal of Chemical Physics*, 126(16):164103, April 2007. ISSN 0021-9606. doi: 10.1063/1.2720838. URL https://doi.org/10.1063/1.2720838.

R. Elber and M. Karplus. A method for determining reaction paths in large molecules: Application to myoglobin. *Chemical Physics Letters*, 139(5):375–380, January 1987. ISSN 0009-2614. doi: 10.1016/0009-2614(87)80576-6. URL https://www.sciencedirect.com/science/article/pii/0009261487805766.

Henry Eyring. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*, 3 (2):107–115, February 1935. ISSN 0021-9606. doi: 10.1063/1.1749604. URL https://doi.org/10.1063/1.1749604.

M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, NY, 1998. ISBN 978-1-4612-6839-0 978-1-4612-0611-8. doi: 10.1007/978-1-4612-0611-8. URL http://link.springer.com/10.1007/978-1-4612-0611-8.

Richard E. Gillilan and Ryan H. Lilien. Optimization and dynamics of protein-protein complexes using B-splines. *Journal of Computational Chemistry*, 25(13):1630–1646, October 2004. ISSN 0192-8651. doi: 10.1002/jcc.20088.

Jiaqi Han, Minkai Xu, Aaron Lou, Haotian Ye, and Stefano Ermon. Geometric Trajectory Diffusion Models, October 2024. URL http://arxiv.org/abs/2410.13027. arXiv:2410.13027 [cs].

Graeme Henkelman, Blas P. Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*, 113(22):9901–9904, December 2000. ISSN 0021-9606. doi: 10.1063/1.1329672. URL https://doi.org/10.1063/1.1329672.

Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Bernd Ensing, and Max Welling. Stochastic Optimal Control for Collective Variable Free Sampling of Molecular Transition Paths, July 2023. URL http://arxiv.org/abs/2207.02149. arXiv:2207.02149 [physics, q-bio].

Bowen Jing, Hannes Stärk, Tommi Jaakkola, and Bonnie Berger. Generative Modeling of Molecular Dynamics Trajectories, September 2024. URL http://arxiv.org/abs/2409.17808. arXiv:2409.17808 [q-bio].

William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983. ISSN 0021-9606. doi: 10.1063/1.445869. URL https://doi.org/10.1063/1.445869.

Hendrik Jung, Roberto Covino, A. Arjun, Christian Leitold, Christoph Dellago, Peter G. Bolhuis, and Gerhard Hummer. Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nature Computational Science*, 3(4):334–345, April 2023. ISSN 2662-8457. doi: 10.1038/s43588-023-00428-z. URL https://www.nature.com/articles/s43588-023-00428-z. Number: 4 Publisher: Nature Publishing Group.

Hannes Jónsson, Greg Mills, and Karsten W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, pp. 385–404. WORLD SCIENTIFIC, June 1998. ISBN 978-981-02-3498-0. doi: 10.1142/9789812839664_0016. URL https://www.worldscientific.com/doi/abs/10.1142/9789812839664_0016.

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, October 2002. doi: 10.1073/pnas.202427399. URL https://www.pnas.org/doi/10.1073/pnas.202427399. Publisher: Proceedings of the National Academy of Sciences.

Gianmarco Lazzeri, Hendrik Jung, Peter G. Bolhuis, and Roberto Covino. Molecular free energies, rates, and mechanisms from data-efficient path sampling simulations, July 2023. URL http://arxiv.org/abs/2307.11240. arXiv:2307.11240.

Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, June 2010. ISSN 0887-3585. doi: 10.1002/prot.22711. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2970904/.

Luca Maragliano, Alexander Fischer, Eric Vanden-Eijnden, and Giovanni Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *The Journal of Chemical Physics*, 125(2):024106, July 2006. ISSN 0021-9606. doi: 10.1063/1.2212942. URL https://doi.org/10.1063/1.2212942.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, August 2020. URL http://arxiv.org/abs/2003.08934. arXiv:2003.08934 [cs].

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action Matching: Learning Stochastic Dynamics from Samples. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 25858–25889. PMLR, July 2023. URL https://proceedings.mlr.press/v202/neklyudov23a.html. ISSN: 2640-3498.

Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A Computational Framework for Solving Wasserstein Lagrangian Flows, July 2024. URL http://arxiv.org/abs/2310.10649. arXiv:2310.10649 [cs].

Eyal Neria, Stefan Fischer, and Martin Karplus. Simulation of activation free energies in molecular systems. *The Journal of Chemical Physics*, 105(5):1902–1921, August 1996. ISSN 0021-9606. doi: 10.1063/1.472061. URL https://doi.org/10.1063/1.472061.

Roberto Olender and Ron Elber. Yet another look at the steepest descent path. *Journal of Molecular Structure: THEOCHEM*, 398-399:63–71, June 1997. ISSN 0166-1280. doi: 10.1016/S0166-1280(97)00038-9. URL https://www.sciencedirect.com/science/article/pii/S0166128097000389.

L. Onsager and S. Machlup. Fluctuations and Irreversible Processes. *Physical Review*, 91(6):1505–1512, September 1953. doi: 10.1103/PhysRev.91.1505. URL https://link.aps.org/doi/10.1103/PhysRev.91.1505. Publisher: American Physical Society.

M. Petersen, G. Roig, and R. Covino. TELD: Trajectory-Level Langevin Dynamics for Versatile Constrained Sampling, December 2024. URL https://scholar.google.com/citations?view_op=view_citation&hl=en&user=yWSSx4wAAAAJ&citation_for_view=yWSSx4wAAAAJ:9yKSN-GCB0IC.

Magnus Petersen, Gemma Roig, and Roberto Covino. DynamicsDiffusion: Generating and Rare Event Sampling of Molecular Dynamic Trajectories Using Diffusion Models. October 2023. URL https://openreview.net/forum?id=pwYCCq4xAf.

Lawrence R. Pratt. A statistical method for identifying transition states in high dimensional problems. *The Journal of Chemical Physics*, 85(9):5045–5048, November 1986. ISSN 0021-9606. doi: 10.1063/1.451695. URL https://doi.org/10.1063/1.451695.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, February 2019. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Kalyan Ramakrishnan, Lars L. Schaaf, Chen Lin, Guangrun Wang, and Philip Torr. Implicit Neural Representations for Chemical Reaction Paths, February 2025. URL http://arxiv.org/abs/2502.15843. arXiv:2502.15843 [cs].

Weiqing Ren, Eric Vanden-Eijnden, Paul Maragakis, and Weinan E. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *The Journal of Chemical Physics*, 123(13):134109, October 2005. ISSN 0021-9606. doi: 10.1063/1.2013256.

David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*, 330(6002):341–346, October 2010. doi: 10.1126/science.1187409. URL https://www.science.org/doi/full/10.1126/science.1187409. Publisher: American Association for the Advancement of Science.

Daniel Sheppard, Rye Terrell, and Graeme Henkelman. Optimization methods for finding minimum energy paths. *The Journal of Chemical Physics*, 128(13):134106, April 2008. ISSN 0021-9606. doi: 10.1063/1.2841941. URL https://doi.org/10.1063/1.2841941.

Aditya N. Singh and David T. Limmer. Variational deep learning of equilibrium transition path ensembles. *The Journal of Chemical Physics*, 159(2):024124, July 2023. ISSN 0021-9606, 1089-7690. doi: 10.1063/5.0150278. URL http://arxiv.org/abs/2302.14857. arXiv:2302.14857 [physics].

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit Neural Representations with Periodic Activation Functions, June 2020. URL http://arxiv.org/abs/2006.09661. arXiv:2006.09661 [cs].

G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, February 1977. ISSN 0021-9991. doi: 10.1016/0021-9991(77)90121-8. URL https://www.sciencedirect.com/science/article/pii/0021999177901218.

Alexander Ulitsky and Ron Elber. A new technique to calculate steepest descent paths in flexible polyatomic systems. *The Journal of Chemical Physics*, 92(2):1510–1511, January 1990. ISSN 0021-9606. doi: 10.1063/1.458112. URL https://doi.org/10.1063/1.458112.

Eric Vanden-Eijnden and Matthias Heymann. The geometric minimum action method for computing minimum energy paths. *The Journal of Chemical Physics*, 128(6):061103, February 2008. ISSN 0021-9606. doi: 10.1063/1.2833040. URL https://doi.org/10.1063/1.2833040.

Xinyan Wang, Jichen Li, Lan Yang, Feiyang Chen, Yingze Wang, Junhan Chang, Junmin Chen, Wei Feng, Linfeng Zhang, and Kuang Yu. DMFF: An Open-Source Automatic Differentiable Platform for Molecular Force Field Development and Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation*, 19(17):5897–5909, September 2023. ISSN 1549-9618. doi: 10.1021/acs.jctc.2c01297. URL https://doi.org/10.1021/acs.jctc.2c01297. Publisher: American Chemical Society.

Jiawei Yan, Hugo Touchette, and Grant M. Rotskoff. Learning nonequilibrium control forces to characterize dynamical phase transitions. *Physical Review E*, 105(2):024115, February 2022. ISSN 2470-0045, 2470-0053. doi: 10.1103/PhysRevE.105.024115. URL http://arxiv.org/abs/2107.03348. arXiv:2107.03348 [cond-mat].

A Appendix

A.1 Connection between Doob's Lagrangian and Minimum-Energy Path Optimization

We elaborate on the theoretical connections between our energy-based optimization approach and the framework of Doob's h-transform for transition path sampling (Du et al., 2024). This comparison is particularly relevant as both methods employ neural networks to represent molecular state transitions—either as a minimum-energy path in our case or the full transition path ensemble in theirs. Notably, the parametrization of the mean trajectory in Du et al. closely resembles our MEP representation, differing primarily in our use of blending functions and spatial embeddings for the neural network component.

Doob's Lagrangian formulation considers the following action functional:

$$S = \min_{q,v} \int_0^T dt \int dx \, q_{t|0,T}(x) \langle v_{t|0,T}(x), G_t v_{t|0,T}(x) \rangle \tag{15}$$

This functional quantifies the cost of controlling a stochastic process to achieve desired endpoint conditions, subject to the energy landscape of the system. Intuitively, it measures the amount of "effort" needed to steer the dynamics from starting state A to target state B, with smaller values indicating more probable transition paths.

In this formulation, $q_{t|0,T}(x)$ is the probability density at time t given boundary conditions, parameterized as a Gaussian $\mathcal{N}(x|\mu_{t|0,T}, \Sigma_{t|0,T})$, and $v_{t|0,T}(x)$ is the control vector field satisfying $v_{t|0,T}(x) = \frac{1}{2}(G_t)^{-1}(u_{t|0,T}(x) - b_t(x))$, with $b_t(x) = -\nabla V(x)$ representing the reference drift for overdamped dynamics.

This action functional describes a Wasserstein Lagrangian flow as formalized in (Neklyudov et al., 2023) and (Neklyudov et al., 2024). In this framework, one can learn stochastic dynamics from samples, thereby learning a process that defines a time-dependent density evolving from initial to final states.

The key difference with our PINN-MEP approach is that we minimize the action of a single deterministic path rather than a full distribution of paths. Mathematically, this corresponds to working with a single trajectory through configuration space rather than an ensemble of possible trajectories. This simplification is reflected in our loss function, which contains one fewer expectation than the full Doob's Lagrangian.

A.2 Comparison with Doob's Lagrangian Method

.

We provide a direct comparison between our PINN-MEP approach and the method of Du et al. in terms of computational efficiency, as evaluated by the number of force-field/energy function calls, and the maximum energies observed in the paths, corresponding to the peak of the potential barrier. To provide a rigorous comparison, we conducted benchmarks using experimental conditions identical to those reported by Du et al. for alanine dipeptide: AMBER14 force field, 300K temperature. Table **??** presents these results alongside those reported in Du et al. (2024). The only exception is that we do not evaluate trajectory probability, as the assumption of optimization over a non-physical path length makes the calculation of this quantity impossible.

Table 1: Comparison of Doob's Lagrangian and PINN-MEP approach on alanine dipeptide

Method	**States**	**# Evaluations**	**Max Energy**	**MinMax Energy**
MCMC (variable length)	CV	21.02M	740.70 $\pm$ 695.79	52.37
MCMC*	CV	1.29B	288.46 $\pm$ 128.31	60.52
MCMC (variable length)	relaxed	187.54M	412.65 $\pm$ 334.70	26.97
MCMC	relaxed	$>$ 10B	N/A	N/A
MCMC (variable length)	exact	$>$ 10B	N/A	N/A
MCMC	exact	$>$ 10B	N/A	N/A
Doob's (Cartesian)	exact	38.40M	726.40 $\pm$ 0.07	726.18
Doob's (Cartesian, 2 Mixtures)	exact	51.20M	709.38 $\pm$ 162.37	513.72
Doob's (Cartesian, 5 Mixtures)	exact	51.20M	541.26 $\pm$ 278.20	247.96
Doob's (Internal)	exact	38.40M	-14.62 $\pm$ 0.02	-14.67
Doob's (Internal, 2 Mixtures)	exact	51.20M	-15.38 $\pm$ 0.14	-15.54
Doob's (Internal, 5 Mixtures)	exact	51.20M	-15.50 $\pm$ 0.31	-15.95
PINN-MEP (Linear)	exact	15700	-67.22 $\pm$ 3.55	-70.78

PINN-MEP requires approximately 3261× fewer energy function evaluations (15.7K vs. 51.20M) compared to the Doob's Lagrangian approach while achieving a 4.44 times lower maximum energy barrier. The lower number of required evaluations is likely due to the computation of the loss, which requires sampling their path $q_{t|0,T}$ over both T, as we do as well, but also the width of the path. Thus, the tradeoff of sacrificing full ensemble characterization is expected to come with computational efficiency.

A.3 Experimental Details

For our alanine dipeptide minimum-energy path (MEP) experiments, we used a neural network architecture optimized through systematic hyperparameter evaluation.

A.3.1 Network Architecture

- **Model type**: Transition MLP with sinusoidal time embedding
- **Network depth**: 10 hidden layers

- **Network width**: $8\times$ width factor (scaled relative to the number of atoms)
- **Time embedding**: Sinusoidal position encoding with maximum period of 100,000
- **Activation function**: ReLU

A.3.2 Training Parameters

- **Optimizer**: Adam
- **Learning rate**: 0.0004
- **Batch size**: 1
- **Frames per batch**: 8
- **Spring constant** (k): 0.0
- **Interpolation method**: Linear

B Ablation Studies

In this section, we present a series of ablation studies examining the impact of various components of our PINN-MEP approach. These experiments help validate our design choices and provide insights into the importance of different architectural and training elements. We conduct these ablation studies on the AIB9 system because Alanine dipeptide as a benchmark system, is too simple to provide meaningful insights into the performance of our method. For each ablation study, we report the lowest maximum energy of the path, the mean and standard deviation of the energy, and the mean number of force field evaluations to reach the lowest maximum energy. We run each setting in the ablations five times for a maximum of 150000 batches, with a batch size of 1 and 32 frames per batch. For experiments without a spring constant between adjacent sampled frames, the batch size and frames are treated identically. Due to the nature of physical force fields, small deviations in the path can lead to large changes in energy, leading to large variations in the energy of the path, leading to similar settings having very different energies. This can limit the ability to compare the performance of different settings, however, we find that the results are still directionally informative.

B.1 Embedding Strategy

We compared sinusoidal time embeddings with different maximum periods against simple linear projections of the progress variable. Table 2 shows that sinusoidal embeddings with a period of 10^6 generally outperformed other configurations.

Table 2: Impact of embedding strategy on MEP generation

Embedding Type	Max Period	Mean Energy	Best Energy	Evaluations
Sinusoidal	10^6	5832.29 ± 7006.09	-33.80	142400
Sinusoidal	10^4	4939.20 ± 2836.56	1388.48	122100
Sinusoidal	10^3	9508.61 ± 7297.49	1273.18	106900
Sinusoidal	10^2	5851.25 ± 4618.35	819.25	129600
Linear (None)	–	9567.01 ± 7578.20	303.63	127600

B.2 Spring Regularization

We tested the effect of chain-of-states-like spring constant regularization, as shown in Table 3. Interestingly, we found that a small amount of spring regularization ($k = 10^{-4}$) provided the best results, suggesting that a minimal amount of explicit path continuity enforcement can be beneficial when combined with our neural representation, however, this is not a requirement for the method to work and the additional hyperparameter does not necessarily be worth including.

Table 3: Impact of spring constant regularization

Spring Constant k	Mean Energy	Best Energy	Evaluations
10^{-4}	368.78 $\pm$ 416.68	-278.09	130300
10^{-5}	8862.90 $\pm$ 9138.59	-155.85	120600
0.0	8200.08 $\pm$ 15576.30	240.85	116500
10^{-3}	1156.70 $\pm$ 428.50	791.59	120400
10^{-2}	7440.13 $\pm$ 3339.87	4657.28	92300
10^{-1}	45567.30 $\pm$ 5015.78	40609.10	64700

B.3 Network Architecture

We investigated the effect of network depth and width on transition path quality. Table 4 shows that an 8-layer network with a width factor of 4 achieved the best balance between expressivity and computational efficiency, with wider networks also performing well.

Table 4: Network architecture ablation results

Depth	Width Factor	Mean Energy	Best Energy	Evaluations
8	4	2746.08 $\pm$ 5381.57	-168.17	112300
8	6	6337.10 $\pm$ 6284.13	-278.09	79300
8	2	10396.20 $\pm$ 14039.40	-278.09	113500
6	6	6189.19 $\pm$ 7667.11	334.22	112800
6	4	9371.15 $\pm$ 15277.50	240.85	111800
6	2	9872.01 $\pm$ 14229.30	152.36	122500
4	6	13242.00 $\pm$ 8432.26	6013.58	94000
4	4	16300.90 $\pm$ 5768.37	8295.52	132100
4	2	24916.70 $\pm$ 14933.30	12697.60	78700

C Appendix: BPTI Modeling and Computational Details

C.1 System Preparation

We used the structures from the D.E. Shaw Research BPTI simulation of BPTI (Shaw et al., 2010) as reference conformations. The initial and final states for our minimum-energy path optimization were selected as the second snapshot (index 1) and the second-to-last snapshot (index -2) from the provided reference structures, as they were the most separated in the CVs determined in the original study, named the backbone RMSD and the disulfide torsion angle.

C.2 Explicit Solvent Shell Construction

For the BPTI system, we developed a sophisticated approach to handle explicit water molecules efficiently:

- **Initial water box generation**: Both start and end configurations were solvated with TIP3P water molecules using a padding distance of 2.0 nm and energy minimized for 15,000 steps.
- **Water shell extraction**: Rather than using the entire water box, we extracted spherical shells of water molecules within 1.4 nm radius of any protein atom for both the start and end configurations. This significantly reduced the system size while maintaining the crucial first hydration shell.
- **Radius optimization**: We employed a binary search algorithm to determine optimal shell radii that would capture an identical number of water molecules in both start and end configurations, ensuring consistent system sizes throughout the transition.

- **Water molecule matching**: We applied the Hungarian algorithm to optimally pair water molecules between the start and end states based on spatial proximity, minimizing the overall displacement of water molecules during the transition, which ensures that the intial guess for the MEP does not have very large energies stemming from overlapping molecules.

C.3 Hybrid Neural Network Architecture

For the BPTI system with explicit solvent, we developed a specialized hybrid transition model to handle the larger system size efficiently:

- **Protein representation**: Neural network with sinusoidal time embedding for the protein atoms (58 residues, 892 atoms)
- **Water representation**: Parameterized spline interpolation between matched water molecules
- **Control points**: 5 control points for the spline interpolation
- **Water weight**: 0.1 (relative weight of water molecule contributions to the loss)
- **Clamp threshold**: 1×10^{-4} (minimum/maximum energy clipping threshold)

C.4 Training Parameters

The BPTI MEP was optimized using the following training parameters:

- **Optimizer**: Adam
- **Learning rate**: 1×10^{-4}
- **Batch size**: 1
- **Frames per batch**: 16
- **Total iterations**: 30,000
- **Random seed**: 1

C.5 Force Field Parameters

We used the following force field parameters to match the D.E. Shaw simulation conditions:

- **Protein force field**: AMBER99sb-ILDN (amber99sbildn.xml)
- **Water force field**: TIP3P (tip3p.xml)

C.6 Computational Performance

For the BPTI system with explicit solvent (approximately 11,000 atoms total):

- **Training time**: Approximately 15 minutes on a single NVIDIA A6000 GPU
- **Force field evaluations**: Approximately 480,000 evaluations (30,000 iterations x 16 frames per batch)
- **Trajectory generation**: 64 frames for the final MEP visualization and analysis

This computational effort is approximately six orders of magnitude less, in node hours, than the original D.E. Shaw Research simulation on their specialized Anton supercomputer, which required approximately 1.3 million node-hours to generate the 1.03 millisecond trajectory capturing the same conformational changes.