TELD: Trajectory-Level Langevin Dynamics for Versatile Constrained Sampling

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

Gemma Roig
Goethe University Frankfurt
CBMM MIT
The Hessian Center for AI
roig@cs.uni-frankfurt.de

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

Abstract

Trajectory Ensemble Langevin Dynamics (TELD) enables sampling of molecular dynamics at the trajectory level by applying Langevin dynamics to entire molecular trajectories rather than individual conformations. TELD formulates a trajectory probability distribution and derives a score function from it, incorporating energetic and dynamic information, which enables gradient-based sampling in the trajectory phase space. By shifting the perspective from conventional conformation-based MD to trajectory-level MD, TELD can impose diverse constraints on entire trajectories, such as fixing the start and end points in two distinct states, thereby sampling only the transition path ensemble. Our implementation leverages automatic differentiation for computing the high-order derivatives needed for the trajectory probability score calculation, making it compatible with differentiable classical force fields and GNN-based neural network potentials. We validate TELD's performance on a molecular system, AIB9, demonstrating its ability to accurately reproduce equilibrium properties, dynamics, and rare event statistics.

1 Introduction

Molecular dynamics (MD) simulations have become an indispensable tool in computational chemistry and biology, offering atomic-level insights into complex molecular processes [1]. However, conventional MD techniques often struggle to sample rare events and long-time-scale phenomena efficiently, limiting our ability to study critical biological processes such as protein folding, ligand binding, and conformational changes.

To address these limitations, various enhanced sampling methods have been developed, including transition path sampling (TPS) [2], metadynamics [3], and umbrella sampling [4], each offering different approaches to overcome energy barriers. In particular, some research has explored a perspective shift from conventional step-by-step MD to sampling in the trajectory space for TPS, where the trajectory is treated as a string or chain of conformations [5, 6, 7, 8], which allows the targeted generation of transition paths and the sampling of the transition path ensemble. Despite recent advancements [9, 10], challenges remain in efficiently sampling high-dimensional molecular systems with these methods, especially for those sampling the trajectory space.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

Recent developments in machine learning and computational methods opened new avenues for improving the sampling speed and accuracy of MD simulations. Graph neural network (GNN)-based potentials have demonstrated remarkable precision in representing complex energy landscapes and learning coarse-grained representations of systems and their dynamics [11, 12, 13]. Concurrently, the development of differentiable classical force fields [14, 15] has enabled the seamless integration of physics-based models with machine learning frameworks. These advances allow for flexible and efficient computation of high-order or non-standard gradients. This capability has paved the way for more sophisticated enhanced sampling schemes that potentially address the limitations of traditional enhanced sampling methods.

TELD applies Langevin dynamics or optimization to entire molecular trajectories, leveraging automatic differentiation with differentiable force fields, in addition to theoretical advancements inspired by geometric deep learning, to overcome sampling limitations of earlier similar approaches. This enables efficient gradient-based sampling in high-dimensional trajectory spaces, allowing exploration of rare events. We detail TELD's foundations and demonstrate its effectiveness on a molecular system.

1.1 Trajectory Probability

The probability $p(X, V)$ of a trajectory of length L in phase space, where $X = \{x_0, ..., x_l, ..., x_L\}$ and $V = \{v_0, ..., v_l, ..., v_L\}$ represent configurations and velocities, respectively, can be decomposed into the probabilities of the initial state and the transition probabilities of going from one configuration to the next [5, 16, 17]:

$$p(X, V) = p(x_0) \cdot p(v_0) \prod_{l}^{L-1} p(x_{l+1}, v_{l+1} | x_l, v_l) \tag{1}$$

For physical systems in equilibrium $p(x_0)$ and $p(v_0)$ are given by the Boltzmann and Maxwell-Boltzmann distributions, respectively. The transition probability $p(x_{l+1}, v_{l+1}|x_l, v_l)$ can be calculated using the Langevin propagator, discretized using the Euler-Maruyama method:

$$x_{l+1} = x_l + \Delta t v_{l+1} \tag{2}$$

$$v_{l+1} = \alpha v_l + \frac{(1-\alpha)}{M} F(x_l) + \sqrt{\frac{(1-\alpha^2)}{M\beta}} dW \tag{3}$$

where Δt is the time step, α is the friction coefficient, $F(x_l)$ is the force at x_l, M is the particle mass, β is the inverse temperature, and dW is a Wiener process. The propagator can be recast into a probabilistic form, $x_{l+1} \sim \mathcal{N}(\mu_{l+1},\, \sigma^2_{l+1})$ with:

$$\mu_{l+1} = x_l + \Delta t(\alpha v_l + \frac{(1-\alpha)}{M} F(x_l)) \tag{4}$$

$$\sigma_{l+1} = \Delta t \sqrt{\frac{(1-\alpha^2)}{M\beta}} \tag{5}$$

The overall path probability is the product of these transition probabilities for each time step and each particle in the system in parallel. We dropped the particle subscript for clarity. Furthermore, since velocities can be recalculated from the change in position over the time step Δt we only need the initial velocity and the positions to calculate $p(X, V)$.

1.2 Sampling with Score Langevin Dynamics

Given the probability of the trajectory $p(X, V)$, we define the score function as the logarithm of the probability gradient with respect to the trajectory:

$$s(X, V) = \nabla_{X,V} \log p(X, V) \tag{6}$$

$$(X, V)_{k+1} = (X, V)_k + \epsilon s((X, V)_k) + \sqrt{2\epsilon} dW \tag{7}$$

where $(X, V)_k$ represents the full phase space trajectory at iteration k, ϵ is the step size, and dW is a Wiener process in $\mathbb{R}^{(L+1)\times 3N}$. Using this sampling scheme, it is possible to sample from $p(X, V)$ in the limit of $k \to \infty$ and $\epsilon \to 0$.

1.2.1 Physical Interpretation of the Trajectory Score Function

Explicitly expressing the trajectory probability's score function uncovers the forces driving its evolution. The score function can be decomposed into three main components:

$$\nabla_{X,V} \log p(X,V) \ = \nabla_{X,V} \log p(x_0) + \nabla_{X,V} \log p(v_0) + \sum_{l=0}^{L-1} \nabla_{X,V} \log p(x_{l+1}, v_{l+1} | x_l, v_l) \tag{8}$$

Initial configuration and velocity terms can be expanded as:

$$\nabla_{X,V} \log p(x_0) = \beta F(x_0) \hat{e}_{x_0} \tag{9}$$

$$\nabla_{X,V} \log p(v_0) = -\beta M v_0 \hat{e}_{v_0} \tag{10}$$

Where $\hat{e}_{x_0}$ and $\hat{e}_{v_0}$ correspond to the first unit vectors of the components of the trajectory vector X and V. These terms represent forces that constrain the initial position and velocity to evolve according to the Boltzmann and Maxwell-Boltzmann distributions, respectively. The transition probability term can be expanded as follows:

$$\sum_{l=0}^{L-1} \nabla_{X,V} \log p(x_{l+1}, v_{l+1} | x_l, v_l) = \frac{d_{l \to l+1}}{\sigma^2} (-\alpha \Delta t \hat{e}_{v_0} - (1 + \frac{\Delta t (1-\alpha)}{\beta M} \frac{dF(x_l)}{dx_l}) \hat{e}_{x_l} + \hat{e}_{x_{l+1}}) \tag{11}$$

$$+ \sum_{l=0}^{L-1} \frac{d_{l \to l+1}}{\sigma^2} (\alpha \hat{e}_{x_{l-1}} - (1 + \alpha - \frac{\Delta t (1-\alpha)}{\beta M} \frac{dF(x_l)}{dx_l}) \hat{e}_{x_l} + \hat{e}_{x_{l+1}}) \tag{12}$$

Here $d_{l \to l+1}$ is the difference vector between x_{l+1} and the deterministic mean of x_{l+1}, μ_{l+1} and $\hat{e}_{v_0}$, $\hat{e}_{x_{l-1}}$, $\hat{e}_{x_l}$ and $\hat{e}_{x_{l+1}}$ are the unit vectors corresponding to the initial velocity in V and the conformations in X at time index $l-1$, l and $l+1$ respectively. These interactions keep the conformations within the Gaussian distribution described by the probabilistic formulation of the under-damped Langevin propagator described in section 1.1.

1.2.2 Rotational Invariant Transition Probability

The probability $\log p(x_{l+1} | x_l, v_l)$, for $l = 0$ and $\log p(x_{l+1} | x_l, x_{l-1})$, for $l \neq 0$ are not invariant under arbitrary rotations of its input conformations.

$$p(R_1 x_{l+1} | R_2 x_l, R_3 x_{l-1}) \neq p(x_{l+1}, x_l | x_{l-1}) \quad \text{for } R_{1,2,3} \in SO(3) \tag{13}$$

This lack of rotational invariance can lead to fictitious energy barriers in the sampling landscape, as the probability of transitions depends on the absolute orientation of the molecular system at each point in time. This can be interpreted as the trajectory's conformations sterically clashing with their temporal neighbors, see equation 11, thereby drastically slowing down sampling. Previous trajectory-level sampling methods did not encounter this, as they only studied single-particle systems. To address this issue, we reformulate the transition probability as rotationally invariant. We achieve this by aligning the conformations x_{l-1} and x_{l+1} with the frame x_l before computing the probability:

$$p_{\text{inv}}(x_{l+1} | x_l, x_{l-1}) = p(\mathcal{A}_l(x_{l+1}), x_l, \mathcal{A}_l(x_{l-1})) \tag{14}$$

where $\mathcal{A}_l$ is an alignment operator that rotates a given frame to best match the frame x_l found via the Kabsch algorithm. The resulting invariant transition probability satisfies the following conditions:

$$p_{\text{inv}}(R_1 x_{l+1} | R_2 x_l, R_3 x_{l-1}) = p_{\text{inv}}(x_{l+1} | x_l, x_{l-1}) \quad \text{for all } R_{1,2,3} \in SO(3) \tag{15}$$

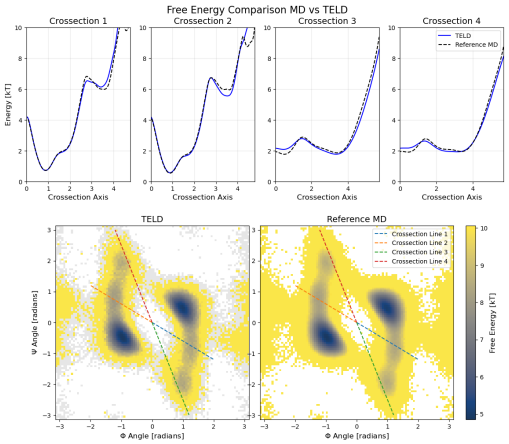
By formulating the transition probability, we ensure that the sampling depends only on the relative changes after alignment. This approach results in a smoother energy landscape for trajectory sampling, at the cost of the rotational alignment of the conformations, as conformations can now rotate freely to each other, which one can deal with via rotational alignment in post-processing.

2 Experiments

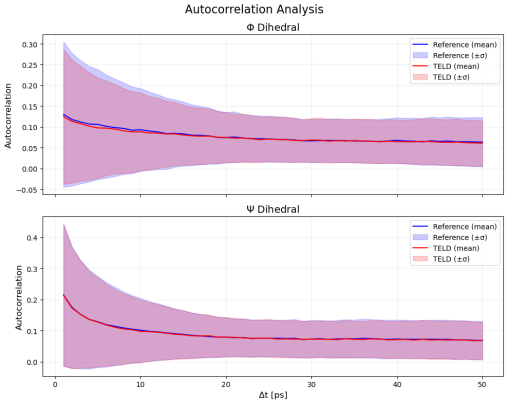
We evaluate Trajectory Ensemble Langevin Dynamics (TELD) on the AIB9 system, a 9-residue artificial protein comprising 129 atoms. AIB9 serves as an ideal test system due to its well-characterized conformational landscape with two distinct metastable states, making it a benchmark for transition sampling methods. We compare TELD to conventional MD simulations at 500K, using the AMBER ff15ipq-m force field for protein mimetics [18] implemented in DMFF [15].

2.1 Validation of Free Energy and Dynamics

We validated TELD against classical MD simulations of the AIB9 system by examining both conformational distributions and dynamical properties. Starting with 3200 trajectories of length L = 96, where each time step corresponds to one picosecond, from classical MD, we evolve them using TELD for 5000 steps and analyze the resulting statistics. Figure 1 shows this comparison through the free energy surface projected onto the central residue's ϕ and ψ dihedral angles, along with their autocorrelation functions. The agreement between methods confirms that TELD preserves both the equilibrium conformational distributions and the system's intrinsic dynamics.



(a) Free energy surface of AIB9 from TELD (left) and classical MD (right).



(b) Auto-correlation functions for ϕ and ψ angles from TELD and classical MD.

Figure 1: Validation of TELD against classical MD shows preservation of both equilibrium distributions (left) and dynamical properties (right) for the AIB9 system.

2.2 Transition Path Generation & Ensemble Sampling

TELD enables sampling of transitions through a two-step approach that addresses a key challenge: generating the initial transition path. First, we construct a preliminary trajectory by sampling one conformation from each metastable state and replicating each $\frac{L}{2}$ times to create the initial and final segments. Although initially discontinuous, these segments are connected by minimizing $-\log p(X, V)$ using the Adam optimizer [19] while maintaining fixed endpoints. This optimization naturally discovers minimum-energy paths connecting the segments. By computing the probabilities of the generated transitions, we can identify and select the most physically feasible paths, as shown in Figure 2. The optimized trajectory can then be evolved through the score sampling method described in Section 1.2. As seen in figure 3, the sampled transi-

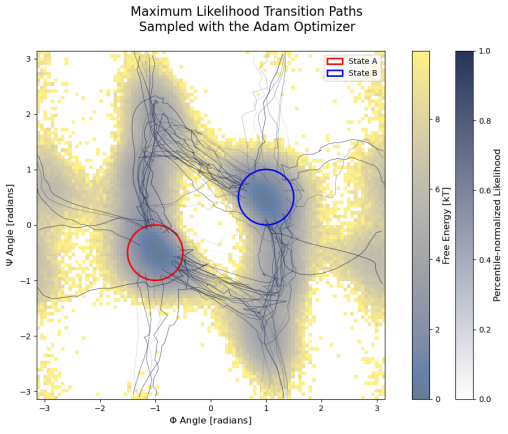


Figure 2: Initial transition paths obtained by optimizing $\log p(X, V)$ subject to start and end state constraints using the Adam optimizer.

tions show reasonable agreement with MD reference data, though some discrepancies in the free energy profiles are observed, likely due to discretization effects from the relatively large trajectory-level timestep ($dt = 0.01$ ps) and TELD stepsize. These trajectories can also serve as input to conventional TPS methods, which typically require an initial transition path.

Table 1: Comparison of transition generation efficiency between conventional MD performed in OpenMM and TELD on the AIB9 system. Measurements were performed on a single A6000 GPU at 500K. A transition is counted when the system moves between the defined metastable states.

Method	Transitions per GPU-hour
Conventional MD	9.8
TELD	310.3

3 Conclusion

We have introduced Trajectory Ensemble Langevin Dynamics (TELD), which advances trajectory-level simulations through modern tools like automatic differentiation and parallelization, as well as theoretical advancements inspired by geometric deep learning to aid in sampling speed. This approach allows for diverse constraints during sampling, with TELD successfully reproducing equilibrium properties, dynamics, and rare event statistics on a model molecule. While TELD shows a 32-fold improvement in transition path generation compared to MD, it has notable limitations: trajectories remain more correlated to their initial conditions than in traditional TPS shooting moves, and the computational cost, though parallelizable, is substantial due to the high-dimensional nature of trajectory space.

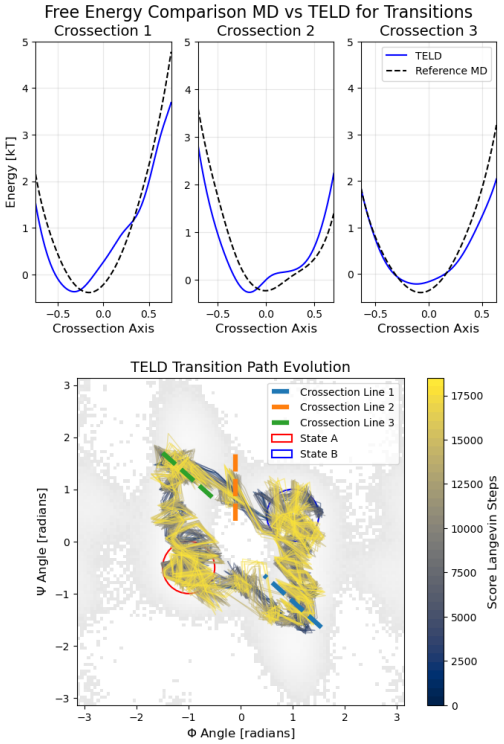


Figure 3: Comparison of transition path distributions between TELD-based sampling of transitions and a long classical MD trajectory (top). The time evolution of two selected transition paths (bottom).

Future work should focus on several key directions: (1) applying TELD to more complex molecular systems such as protein folding and ligand binding, (2) rigorously comparing its performance with established enhanced sampling techniques to fully realize its potential in studying biologically relevant timescales and processes, (3) exploring the integration of neural network potentials to further accelerate sampling, and (4) investigating more sophisticated discretization schemes beyond the current Euler-Maruyama approach, such as BAOAB [20] which might lower discretization errors as seen in figure 3. A particularly promising direction would be to represent transition pathways through implicit neural representations trained as physics-informed neural networks (PINNs) using TELD's trajectory probability as a physics-informed loss function. This could enable continuous-time representations of transitions while maintaining physical consistency.

References

[1] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: A computational microscope for molecular biology. 41(1):429–452. _eprint: https://doi.org/10.1146/annurev-biophys-042910-155245.

[2] Peter Bolhuis, David Chandler, Christoph Dellago, and Phillip Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. 53:291–318.

[3] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. 99(20):12562–12566. Publisher: Proceedings of the National Academy of Sciences.

[4] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. 23(2):187–199.

[5] Christoph Dellago, Peter G. Bolhuis, Félix S. Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. 108(5):1964–1977.

[6] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. 66(5):052301. Publisher: American Physical Society.

[7] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. 126(16):164103.

[8] Silvio A Beccara, Tatjana Škrbić, Roberto Covino, and Pietro Faccioli. Dominant folding pathways of a WW domain. 109(7):2330–2335.

[9] 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. 3(4):334–345. Number: 4 Publisher: Nature Publishing Group.

[10] Gianmarco Lazzeri, Hendrik Jung, Peter G. Bolhuis, and Roberto Covino. Molecular free energies, rates, and mechanisms from data-efficient path sampling simulations.

[11] Cheol Woo Park, Mordechai Kornbluth, Jonathan Vandermause, Chris Wolverton, Boris Kozinsky, and Jonathan P. Mailoa. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. 7(1):1–9. Number: 1 Publisher: Nature Publishing Group.

[12] Brooke E. Husic, Nicholas E. Charron, Dominik Lemm, Jiang Wang, Adrià Pérez, Maciej Majewski, Andreas Krämer, Yaoyi Chen, Simon Olsson, Gianni de Fabritiis, Frank Noé, and Cecilia Clementi. Coarse graining molecular dynamics with graph neural networks. 153(19):194101. Publisher: American Institute of Physics.

[13] Zhiheng Li, Geemi P. Wellawatte, Maghesree Chakraborty, Heta A. Gandhi, Chenliang Xu, and Andrew D. White. Graph neural network based coarse-grained mapping prediction. 11(35):9524–9531.

[14] Samuel S. Schoenholz and Ekin D. Cubuk. JAX, m.d.: A framework for differentiable physics.

[15] 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. 19(17):5897–5909. Publisher: American Chemical Society.

[16] S. Kieninger and B. G. Keller. Path probability ratios for langevin dynamics—exact and approximate. 154(9):094102.

[17] Michael Plainer, Hannes Stärk, Charlotte Bunne, and Stephan Günnemann. Transition path sampling with boltzmann generator-based MCMC moves.

[18] 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. 153(6):064101.

[19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization.

[20] Benedict Leimkuhler and Charles Matthews. Rational construction of stochastic numerical methods for molecular sampling. 2013(1):34–56.

A Appendix

A Adaptive Step Size Calculation

The step size ϵ in equation (7) is computed adaptively to maintain a constant score-to-noise ratio (SNR) between the score function and the Wiener noise across different regions of the trajectory space. The computation follows these steps:

1. Calculate score function magnitude:

$$\|s(X,V)\| = \sqrt{\frac{1}{3N(L+1)} \sum_{n=1}^{N} \sum_{l=0}^{L} \sum_{i=1}^{3} (s_{n,l,i}(X,V))^2} \tag{16}$$

where N is the number of particles, L is the trajectory length, and $s_{n,l,i}$ are the components of the score function for particle n at time step l in dimension i.

2. Calculate noise magnitude:

$$\|dW\| = \sqrt{3N(L+1)} \tag{17}$$

where dW is the Wiener process with the same dimensionality as the trajectory.

3. Compute step size to maintain target SNR:

$$\epsilon = 2 \left(\text{SNR} \cdot \frac{\|dW\|}{\|s(X,V)\|} \right)^2 \tag{18}$$

This ensures that $\text{SNR} = \frac{\epsilon \|s(X,V)\|}{\sqrt{2\epsilon}\|dW\|}$ remains constant throughout sampling, balancing exploration and exploitation based on local score magnitudes. The factor of 2 comes from the standard form of the Langevin equation as shown in equation (7).

B System Parameters and State Definitions

Table 2: TELD Simulation Parameters

Parameter	Value
Software	DMFF (JAX-based)
Force Field	AMBER ff15ipq-m
Temperature	500 K
Trajectory Time Step	1 ps
Trajectory Length Steps (L)	96
Batch Size	64
Number of Batches	50
Total Trajectories	3200
Score Noise Ratio	0.03
Integration Steps	5000
Friction Coefficient	1.0 ps^{-1}
Nonbonded Method	NoCutoff

These parameters were used to generate the equilibrium ensemble shown in Figure 1, which compares TELD's free energy surfaces and autocorrelation functions with classical MD. The relatively large batch size and number of batches ensure adequate sampling of the conformational space, while the

integration steps and noise ratio were tuned to achieve stable sampling. We use DMFF, a differentiable implementation of the AMBER force field, which enables efficient computation of the trajectory probability gradients needed for TELD.

Table 3: Classical MD Parameters

Parameter	Value
Software	OpenMM 8.1
Force Field	AMBER ff15ipq-m
Total Simulation Length	2.7 ms
Integration Timestep	1 fs
Trajectory Saving Interval	1 ps
Temperature	500 K
Friction Coefficient	1.0 ps^{-1}
Total Integration Steps	2.7×10^{9}
Constraint Method	HBonds
Nonbonded Method	NoCutoff
Integrator	LangevinMiddle

Table 4: Metastable State Definitions

State	ϕ Range (rad)	ψ Range (rad)
A	[-1.0 ± 0.5]	[-0.5 ± 0.5]
B	[1.0 ± 0.5]	[0.5 ± 0.5]

C Transition Path Generation Details

Table 5: Initial Transition Path Generation Parameters

Parameter	Value
Trajectory Time Step	0.01 ps
Trajectory Length Steps (L)	38
Initial Paths	124
Initial Optimization Steps	4000
Interpolation Steps	2
Secondary Optimization Steps	1500
Fixed Endpoint Weight	1.0
Central Region Weight	300.0
Weight Fraction	1/4
Learning Rate (Initial)	2×10^{-3}
Learning Rate (Secondary)	1×10^{-3}

These parameters were used to generate the initial transition paths shown in Figure 2. We first optimize short trajectories connecting the two metastable states, then interpolate and refine them to obtain smooth transitions suitable for subsequent sampling. Each conformation in the trajectory is coupled to its potential energy through a weight parameter: a high weight (300.0) in the central region of the trajectory, covering 1/4 of the total length (Weight Fraction), strongly biases the optimization towards finding minimum energy paths in the transition region, while the endpoints remain more loosely constrained with a weight of 1.0.