EXPLORATION-NFT: REINFORCING PROTEIN DYNAMICS GENERATORS FOR PERSISTENT CONFORMATIONAL EXPLORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

The scarcity of long-timescale molecular dynamics (MD) trajectories restricts protein dynamics generators to short-horizon supervision, biasing autoregressive rollouts toward local free-energy basins. We present Exploration-NFT, a backbone-agnostic forward-process reinforcement framework for persistent conformational exploration. Same-condition rollout groups define a shared, non-whitened time-lagged independent component analysis space; segment-wise new-cell increments provide temporal exploration rewards, while structural validity is regularized independently. DiffusionNFT maps these rewards to denoising updates without modifying native samplers or evaluating their likelihoods. We evaluate Exploration-NFT with BioKinema and DyneTrion on 100-ns ATLAS and 10-μs CATH1 and fast-folding benchmarks. Across both backbones and all three benchmarks, post-training improves valid diversity or reference-MD coverage while maintaining all reported local-validity scores above 99.96%. For BioKinema, ATLAS diversity increases $2.24\times$, and CATH1 and fast-folding coverage increases by 47% and 37%, respectively. At the maximum-coverage fast-folding checkpoint, all reported thermodynamic and kinetic discrepancies also decrease. Improvements beyond the training sequence-length range and at unseen temporal intervals further demonstrate transfer across conditions. These results establish forward-process reinforcement as an effective mechanism for expanding the conformational support of protein dynamics generators.

1 INTRODUCTION

Proteins function through distributions of conformations and transitions rather than single static structures. Atomistic molecular dynamics (MD) offers a mechanistic description of these motions, but rare events and slow relaxation make broad sampling expensive. Learned generative models amortize this sampling cost: equilibrium samplers produce independent configurations (Noé et al., 2019), while recent diffusion, flow, and autoregressive models generate ensembles or temporally ordered protein trajectories (Jing et al., 2024a;b; Shen et al., 2025; Yu et al., 2026; Cheng et al., 2026a; Feng et al., 2026). Their long-horizon use nevertheless creates a distinct problem. Because long, well-equilibrated trajectories are scarce, these models predominantly learn conditional distributions from short windows. Such a transition model can accurately reproduce frequent intra-basin fluctuations while underestimating rare basin-exit events. Recursive rollout repeatedly compounds this stay-in-basin bias, so local forecast accuracy need not translate into long-horizon conformational coverage; feeding generated structures back as context further amplifies the distribution shift.

Existing approaches mitigate this mode-trapping problem by modifying the sampling process. Classical enhanced sampling introduces explicit biasing potentials, often along prescribed collective variables (Laio & Parrinello, 2002). More recently, Cheng et al. (2026b) learn a history-dependent score bias that steers a frozen dynamics emulator away from previously visited conformations; environment-support regularization and score-based refinement limit departure from the learned manifold. This inference-time strategy demonstrates that biasing a generative space can accelerate rare-state discovery. We pursue the complementary objective of encoding persistent exploration into the dynamics generator itself through online post-training, while retaining its native sampler.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Such post-training requires temporally resolved credit. Initial displacement or terminal diversity can be dominated by an early excursion and therefore cannot distinguish sustained discovery from subsequent stagnation. Moreover, multiplying exploration by a binary validity mask conflates two objectives: a newly reached region provides no exploration signal if the corresponding frame is even mildly imperfect.

We address these problems with Exploration-NFT, a backbone-adaptable online reinforcement framework. For each initial conformation, the collection policy samples a group of trajectories. Their most variable Cα distance features define a shared two-dimensional TICA projection (Pérez-Hernández et al., 2013). We preserve the scale of this projection rather than whitening every group, place a fixed hard grid over it, and assign each segment the number of cells newly discovered by that trajectory. The credit is self-referenced and temporally resolved, and coverage is monotone only at the level of cumulative visited support; trajectories remain free to move forward, backward, or dwell. A separate invalid-frame penalty controls geometry.

Protein dynamics generators do not necessarily expose a common tractable likelihood for their native sampler. We consequently optimize the reward with the forward-process DiffusionNFT construction (Zheng et al., 2026): rollout remains unchanged, whereas rewarded clean frames are re-corrupted using the backbone's pretraining process for a contrastive denoising update. A minimal adapter provides sampling, corruption, and diffusion-prediction operations without redefining the model's temporal transition. This sampler–optimizer separation permits forward-process reinforcement for models with intractable or inaccessible sampling likelihoods, including autoregressive world-model rollouts (Bardhan et al., 2026).

Our contributions are threefold:

- We introduce a group-relative reward for persistent conformational exploration, based on segment-wise new-cell discovery in a group-fitted, non-whitened slow-coordinate space, with structural validity optimized as a separate term.
- We formulate a forward-process post-training interface that retains a dynamics generator's native sampler while assigning separate credit to temporal segments inside jointly generated or autoregressive blocks.
- We evaluate whether support expansion remains locally valid, enters MD-supported regions, preserves their populations and kinetics, and transfers across length and temporal interval. Matched BioKinema and DyneTrion experiments separate improvements due to the reinforcement objective from behavior specific to one sampler architecture.

2 Related Work

Learned biomolecular ensembles and dynamics. Boltzmann generators learn equilibrium samplers with statistical-mechanical reweighting (Noé et al., 2019), whereas AlphaFlow generates protein ensembles by flow matching around a structure predictor (Jing et al., 2024a). BioEmu scales equilibrium-ensemble generation and introduces the CATH benchmarks used here (Lewis et al., 2025). Temporally ordered models include the SE(3)-equivariant MDGen (Jing et al., 2024b), the pretrained variational bridge PVB (Yu et al., 2026), the autoregressive ConfRover (Shen et al., 2025), the multiscale SE(3) model DyneTrion (Cheng et al., 2026a), and the all-atom diffusion forecaster BioKinema (Feng et al., 2026). Our framework preserves these generators and post-trains their diffusion policies to expand visited conformational support. Learning Implicit Bias addresses long-horizon mode trapping through a history-conditioned score correction and refinement module attached to a frozen emulator (Cheng et al., 2026b). In contrast, Exploration-NFT updates the generator through online exploration rewards without altering its native sampling procedure.

Collective variables and molecular kinetics. TICA identifies directions with high time-lagged autocorrelation and underlies many Markov-state-model pipelines (Pérez-Hernández et al., 2013); VAMPnets learn related kinetic representations end to end (Mardt et al., 2018). We use TICA differently during training and evaluation. The training basis is fitted to a same-condition rollout group solely to define relative exploration credit. Evaluation instead freezes a basis fitted only to independent MD, so a generated ensemble cannot improve its score by rotating its own coordinates.

Unlike metadynamics (Laio & Parrinello, 2002), our grid reward is neither a physical bias potential nor an equilibrium reweighting procedure.

Reinforcement learning for diffusion models. DDPO and DPOK formulate reverse denoising as a policy and apply policy-gradient updates (Black et al., 2024; Fan et al., 2023). Flow-GRPO converts deterministic flow sampling into a stochastic process to obtain transition likelihoods (Liu et al., 2025). These formulations require an explicit sampling transition, whereas protein-dynamics backbones employ heterogeneous churn, conditioning, and block-autoregressive procedures. DiffusionNFT instead contrasts rewarded clean samples under the forward process and is compatible with black-box solvers (Zheng et al., 2026); Persistent World Models extends this idea to multi-step autoregressive generation (Bardhan et al., 2026). Our contribution is complementary: a protein-specific exploration objective, segment-level credit within jointly generated blocks, and an evaluation that separates support, equilibrium weighting, local geometry, and kinetics.

3 METHOD

3.1 PROBLEM AND BACKBONE INTERFACE

Let a pretrained diffusion dynamics generator define a native rollout operator $\mathcal{S}_\theta$, a forward corruption $q_\sigma(x_\sigma \mid x)$, and a denoising prediction $D_\theta(x_\sigma, \sigma, c)$. The native sampler may be SDE-, churn-, or block-autoregressive; Exploration-NFT neither replaces it nor requires its transition likelihood. Given an initial conformation x_0 and condition c, a collection policy $\bar{\theta}$ generates a group

$$\tau_g = \mathcal{S}_{\bar{\theta}}(x_0, c) = (x_{g,1}, \ldots, x_{g,T}), \qquad g = 1, \ldots, G. \tag{1}$$

All trajectories share the same start and conditioning and are divided into K contiguous, non-overlapping segments $\mathcal{T}_k$. Autoregressive contexts may differ after the first generated block, so the resulting segment-wise group baseline is a trajectory-level Monte Carlo comparison at a common start and horizon, not a same-state action baseline. Training requires only three adapter calls—native rollout, native forward corruption, and the pretraining denoiser—and the reward depends only on generated coordinates and topology. It does not access reference MD or labeled transition paths.

3.2 GROUP-FITTED SLOW COORDINATES

For same-chain residue pairs (i, j) with $|i - j| \geq 4$, let

$$d_{g,t,(i,j)} = \left\| r^{\mathrm{C}\alpha}_{g,t,i} - r^{\mathrm{C}\alpha}_{g,t,j} \right\|_2 . \tag{2}$$

We rank pairs by standard deviation over all $G(T+1)$ frames, including the shared start, retain the top M, and set $u_{g,t} = d_{g,t}/\sqrt{M}$. We do not standardize each feature: per-group variance normalization would make a contracted group appear artificially broad. Lagged pairs are formed only within each trajectory. At lag $\ell = 1$, we estimate symmetric covariances C_{00} and $C_{0\ell}$ and solve

$$C_{0\ell} v_q = \lambda_q (C_{00} + \rho I) v_q. \tag{3}$$

The eigensystem is solved in FP64 after removing numerically null covariance directions. We retain two directions and Euclidean-orthonormalize them, but do not whiten their projected variance. The projected coordinate is $z_{g,t} \in \mathbb{R}^2$ and the trajectory-relative coordinate is $\tilde{z}_{g,t} = z_{g,t} - z_{g,0}$. Appendix B gives the covariance estimators, trace-scaled ridge ρ, and numerical safeguards.

3.3 PERSISTENT COVERAGE REWARD

With a fixed grid spacing Δ, a projected frame occupies

$$b_{g,t} = \left\lfloor \frac{\tilde{z}_{g,t}}{\Delta} + \frac{1}{2} \right\rfloor \in \mathbb{Z}^2. \tag{4}$$

Let $N_{g,k}$ be the number of distinct cells visited from frame zero through the end of segment k. Cumulative and incremental coverage are

$$C_{g,k} = \frac{N_{g,k} - 1}{T}, \qquad \Delta C_{g,k} = C_{g,k} - C_{g,k-1}. \tag{5}$$

Thus a segment receives credit only for cells not visited earlier by the same trajectory. Revisiting and backward motion are permitted but receive zero incremental coverage credit.

Coverage is not gated by geometry: a newly visited frame that fails a local validity criterion still contributes an exploration signal. Instead, let $P_{g,k}^{\text{invalid}}$ be the segment's fraction of frames that fail local bond or clash criteria. The reward is

$$r_{g,k} = \Delta C_{g,k} - \lambda_{\text{inv}} P_{g,k}^{\text{invalid}}. \tag{6}$$

For each segment, rewards are standardized across the G candidates and mapped to an optimality target:

$$A_{g,k} = \frac{r_{g,k} - \mu_k}{\sigma_k + 10^{-6}}, \qquad o_{g,k} = \frac{1}{2} + \frac{\text{clip}(A_{g,k}, -c, c)}{2c}. \tag{7}$$

Segments with insufficient group contrast are skipped. Because later new cells are rarer, fixed weights w_k are calibrated from pretrained step-0 rollouts and applied after standardization; the exact rule is in Appendix B. This weighting changes temporal credit without redefining the underlying reward.

3.4 Forward-process policy update

Let $D_\theta(x_\sigma, \sigma, c)$ be the current denoiser and $D_{\bar{\theta}}$ an exponential-moving-average collection policy. The adapter samples x_σ from the backbone's original forward corruption and uses its clean pretraining target y. Following DiffusionNFT, positive and negative predictions are

$$D^+ = (1 - \beta)\,\text{sg}(D_{\bar{\theta}}) + \beta D_\theta, \qquad D^- = (1 + \beta)\,\text{sg}(D_{\bar{\theta}}) - \beta D_\theta. \tag{8}$$

Here sg denotes stop-gradient. For $s \in \{+, -\}$, the scale-normalized reconstruction term is

$$L_{g,k}^s = \frac{\text{MSE}_{g,k}(D^s, y)}{\text{sg}[\text{MAE}_{g,k}(D^s, y)] + \epsilon}. \tag{9}$$

The segment loss is

$$L_{g,k}^{\text{NFT}} = \frac{c}{\beta}\left[o_{g,k} L_{g,k}^+ + (1 - o_{g,k}) L_{g,k}^-\right]. \tag{10}$$

The factor c/β cancels the scale introduced by Equation 7, keeping the initial reward-directed gradient on the standardized-advantage scale. One optimizer step minimizes the weighted mean over active segments and candidates, plus a masked denoiser-space penalty to the frozen pretrained policy D_{ref}:

$$L = \frac{\sum_k w_k m_k \, G^{-1} \sum_g L_{g,k}^{\text{NFT}}}{\sum_k w_k m_k} + \lambda_{\text{ref}} \left\| D_\theta - D_{\text{ref}} \right\|_{\text{mask}}^2. \tag{11}$$

Here m_k indicates that segment k has sufficient contrast. Segmenting the loss never changes a backbone's sampling block: frames can be generated jointly, while disjoint slices of their reconstruction loss receive different optimality targets. Algorithm 1 summarizes the full procedure.

Algorithm 1 Exploration-NFT post-training

Require: Pretrained policy θ_0, conditions $\mathcal{D}$, native sampler $\mathcal{S}$, corruption q, and reward/update hyperparameters

Set $\theta \leftarrow \theta_0$, $\bar{\theta} \leftarrow \theta_0$, $\theta_{\text{ref}} \leftarrow \theta_0$; calibrate fixed w_k and contrast floors η_k from step-0 rollouts
for each optimizer step **do**
  Sample $(x_0, c) \sim \mathcal{D}$ and $\{\tau_g\}_{g=1}^{G} \sim \mathcal{S}_{\bar{\theta}}(x_0, c)$
  Select the M most variable Cα-pair distances; fit one shared, non-whitened 2D TICA basis to the group
  for $g = 1, \ldots, G$ and $k = 1, \ldots, K$ **do**
    Compute new-cell increment $\Delta C_{g,k}$ and $r_{g,k} = \Delta C_{g,k} - \lambda_{\text{inv}} P_{g,k}^{\text{invalid}}$
  Normalize rewards within each segment; set $o_{g,k}$ by Equation 7 and skip segments with $\sigma_k < \eta_k$
  Re-corrupt clean rollout frames with q and compute $D_\theta, D_{\bar{\theta}}, D_{\text{ref}}$
  Update θ once using Equation 11; then update $\bar{\theta}$ by EMA

4 Experimental Setup

Backbones and post-training. We test two pretrained dynamics generators with different architectures and native samplers: all-atom BioKinema (Feng et al., 2026) and SE(3)-equivariant DyneTrion (Cheng et al., 2026a). Both are post-trained on the same 940 ATLAS (Vander Meersche et al., 2024) training proteins of at most 300 residues; their MD trajectories are not accessed by the reward. The 39 ATLAS validation proteins are evaluated in fixed, step-0-calibrated reward coordinates. Shared Exploration-NFT settings are $G = 8$, $K = 10$, $M = 128$, $\Delta = 0.03$, $\lambda_{\text{inv}} = 0.5$, $c = 2$, $\beta = 0.1$, and $\lambda_{\text{ref}} = 10^{-3}$. We use Adam with learning rate 10^{-6} and gradient clipping at 1.0.

The adapter preserves each backbone's established rollout contract. BioKinema uses two native 50-frame direct/churn blocks at 10 ns per frame, 20 denoising steps, and BF16 autocast. DyneTrion uses ten native 16-frame SDE blocks at 1 ns per frame and 50 denoising steps. Thus rollout length and sampling details are backbone configurations rather than definitions of our algorithm. Appendix A reports algorithm hyperparameters separately from these adapter settings.

Short-horizon ATLAS test. We evaluate all 81 proteins in BioKinema's public ATLAS test set, starting from the first-replica structure and generating 100 frames at 1 ns intervals. Following Cheng et al. (2026b), diversity is the mean pairwise $1 - \text{lDDT}_{\text{C}\alpha}$ (Mariani et al., 2013). Div@τ retains only frames whose Cα–Cα, C–N, and clash-free fractions are all at least τ; mDiv-mean averages Div@.98, Div@.99, and Div@1.00. DyneTrion is evaluated on the 68 proteins within its 500-residue inference limit. We also stratify BioKinema results by sequence length.

Long-horizon MD-referenced tests. BioEmu CATH1 contains 17 domains with 21 independent reference runs per protein (Lewis et al., 2025). The fast-folding test contains 12 proteins with long-timescale reference MD (Lindorff-Larsen et al., 2011). For each protein and checkpoint, we generate one 1,000-frame trajectory at 10 ns per frame (10 μs). The native MSA-enabled sampler, initial structure, noise scale 1.75, and random seed 101 are matched. A two-dimensional TICA basis, evaluation grid, and 12 microstates are fitted only to reference MD, using 300 K for CATH1 and the recorded simulation temperature for each fast-folding protein.

All ensemble and kinetic metrics exclude a frame unless all three local validity fractions are at least 98%. Tables report unweighted protein-level means. RMWD measures Gaussian 2-Wasserstein discrepancy between aligned heavy-atom ensembles; coverage is the fraction of reference TICA cells within 4 kcal/mol of the free-energy minimum that are visited; occupancy JSD compares MD and generated populations over MD-fitted microstates. AutoCor-MAE, Transition JSD, and ITS-log-MAE compare TICA autocorrelation, joint state transitions, and the leading three implied time scales at 10–200 ns lags. Exact definitions, replica-boundary handling, and missing-value rules appear in Appendix C.

Table 1: Held-out ATLAS short-horizon evaluation (mean). BioKinema uses all 81 proteins; DyneTrion uses the 68 proteins within its 500-residue limit. Each trajectory contains 100 frames over 100 ns. mDiv-mean averages Diversity@.98, Diversity@.99, and Diversity@1.00.

Model	CA (%) ↑	CN (%) ↑	Clash-Free (%) ↑	Diversity@.98 ↑	mDiv-mean ↑
DyneTrion	99.99	99.98	100.00	0.184	0.178
DyneTrion-NFT	99.99	99.98	100.00	**0.199**	**0.192**
BioKinema	100.00	100.00	100.00	0.150	0.150
BioKinema-NFT (300)	100.00	100.00	100.00	0.227	0.227
BioKinema-NFT (500)	100.00	99.99	100.00	0.277	0.277
BioKinema-NFT (800)	100.00	100.00	100.00	**0.336**	**0.336**

Table 2: CATH1 long-horizon evaluation (17 proteins; mean). Each trajectory contains 1,000 frames over 10 μs. Lower is better except validity and coverage.

	Structure validity			Thermodynamics			Kinetics		
Model	CA (%) ↑	CN (%) ↑	Clash-Free (%) ↑	RMWD (Å) ↓	Coverage ↑	Occupancy JSD ↓	AutoCor-MAE ↓	Transition JSD ↓	ITS-log-MAE ↓
DyneTrion	99.99	99.98	100.00	**4.988**	0.636	**0.473**	0.638	**0.567**	4.285
DyneTrion-NFT	99.99	99.98	100.00	6.459	**0.662**	0.527	**0.635**	0.618	**4.180**
BioKinema	100.00	100.00	100.00	1.654	0.525	0.332	0.911	**0.434**	5.046
BioKinema-NFT (300)	99.99	99.99	100.00	**1.620**	0.610	**0.314**	**0.910**	0.475	4.744
BioKinema-NFT (500)	99.98	99.98	100.00	1.903	0.681	0.351	0.916	0.537	4.710
BioKinema-NFT (800)	99.98	99.98	100.00	2.562	**0.772**	0.405	0.919	0.611	**4.583**

5 RESULTS

5.1 SHORT-TIME DIVERSITY INCREASES WITHOUT LOSS OF LOCAL VALIDITY

Table 1 evaluates whether reinforcement broadens the generated ensemble without increasing unstructured coordinate variation. For BioKinema, Diversity@.98 rises monotonically by 51.2%, 84.1%, and 123.5% at steps 300, 500, and 800. Every checkpoint also improves all 81 paired proteins, excluding the possibility that the mean increase is dominated by a small subset of unusually flexible targets. The effect is therefore systematic across the test distribution.

The validity metrics distinguish physically admissible support expansion from apparent diversity induced by structural distortion. CA and CN scores remain above 99.99% before rounding, clash-free scores round to 100.00%, and gated and ungated diversity are equal. Thus the additional pairwise structural variation is carried by physically admissible frames. This is consistent with the intended functional separation: the coverage reward alters the global conformations reached, whereas the pretrained denoising prior, reference regularization, and explicit invalidity term preserve local geometry.

Cross-architecture transfer is evaluated with DyneTrion, whose equivariance, sampler, rollout blocks, and denoising schedule differ from BioKinema. Its Diversity@.98 still rises from 0.184 to 0.199 and its improvement survives the strictest validity threshold. The smaller gain shows that the reward supplies a transferable optimization signal, but not a backbone-independent effect size: the magnitude depends on the pretrained support and on the backbone's response to forward-process updates. Length-stratified results below test whether this behavior is confined to the post-training size range.

5.2 LONG-TIME SUPPORT EXPANDS, BUT THERMODYNAMICS AND KINETICS DIVERGE

The CATH1 results expose a hierarchy of learned properties. Coverage records whether a low-free-energy MD cell is visited at least once; RMWD measures the location and spread of the generated ensemble; occupancy JSD measures the statistical weight assigned to visited states; and the kinetic metrics further depend on temporal ordering. Success at each level is necessary but not sufficient for success at the next.

At step 300, BioKinema coverage increases from 0.525 to 0.610 while RMWD, occupancy JSD, AutoCor-MAE, and ITS-log-MAE also decrease. The initial support gains therefore overlap MD-supported regions that the pretrained model undersamples, rather than reflecting unrestricted excursions. Continued training separates support from calibration. Coverage reaches 0.772 at step

Table 3: Fast-folding long-horizon evaluation (12 proteins; mean). Each trajectory contains 1,000 frames over 10 μs.

	Structure validity			Thermodynamics			Kinetics		
Model	CA (%) ↑	CN (%) ↑	Clash-Free (%) ↑	RMWD (Å) ↓	Coverage ↑	Occupancy JSD ↓	AutoCor-MAE ↓	Transition JSD ↓	ITS-log-MAE ↓
DyneTrion	100.00	99.99	100.00	6.635	0.624	0.445	**0.506**	0.525	**2.871***
DyneTrion-NFT	100.00	99.99	100.00	**6.584**	**0.631**	**0.425**	0.556	**0.501**	2.986*
BioKinema	99.99	99.99	100.00	6.041	0.576	0.496	0.732	0.557	3.772
BioKinema-NFT (300)	99.98	99.98	100.00	5.064	0.688	0.455	0.724	0.543	3.430
BioKinema-NFT (500)	99.97	99.97	100.00	4.572	0.746	0.438	**0.695**	**0.539**	3.194
BioKinema-NFT (800)	99.96	99.96	100.00	**4.087**	**0.791**	**0.429**	0.701	0.551	**3.075**

800 and improves for all 17 proteins, but RMWD, occupancy JSD, and Transition JSD rise. This behavior follows directly from the objective: a cell earns credit upon first discovery regardless of its equilibrium mass, and revisiting order carries no reward. Near-constant jointly valid-frame rates exclude local structural degradation as the explanation. Step 300 simultaneously improves support and ensemble agreement on CATH1, whereas step 800 maximizes coverage at the expense of distributional calibration.

DyneTrion exhibits the same separation in a different parameterization. Reinforcement increases its coverage and slightly improves autocorrelation and implied-time-scale errors, but does not improve RMWD, occupancy, or transition JSD. The consistent support expansion across backbones identifies the robust effect of the reward; the different secondary responses show that population and kinetic calibration remain properties of the learned transition operator, rather than automatic consequences of visiting additional states.

For the DyneTrion pair, fast-folding ITS-log-MAE is reported on the 11 proteins with finite values for both models; NuG2 has insufficient comparable active slow modes.

Fast-folding proteins present the complementary regime in which the pretrained BioKinema ensemble is substantially narrower than reference MD. Here increased support and distributional fidelity reinforce one another (Table 3): coverage rises at every checkpoint, while RMWD and occupancy JSD decrease monotonically. Steps 300 and 500 improve both coverage and RMWD for every one of the 12 proteins, so the aggregate trend is again a paired effect rather than an average dominated by a subset. The concurrent increase in coverage and reductions in RMWD and occupancy JSD exclude unstructured dispersion from the initial conformation as a sufficient explanation. Together with at least 99.96% local validity, these results show that the additional support overlaps reference-MD basins and remains structurally admissible.

The temporal metrics explain how this support is acquired. Autocorrelation and Transition JSD are best at step 500, while coverage and ensemble agreement continue improving through step 800. Over the same checkpoints, the 10-ns generated-to-MD displacement ratio changes from 0.874 to 1.426, 1.819, and 2.221. Reinforcement therefore reverses the pretrained model's underestimated short-lag displacement, but prolonged optimization increasingly obtains coverage through excessive motion. Step 500 provides the most balanced kinetic agreement, including the lowest autocorrelation and transition errors, whereas step 800 maximizes support. This ordering demonstrates that checkpoint selection depends on whether the downstream objective prioritizes kinetic fidelity or conformational coverage.

DyneTrion starts from higher fast-folding coverage and a lower implied-time- scale error than BioKinema, and therefore exhibits a smaller initial support deficit. Its NFT checkpoint yields correspondingly smaller but coherent ensemble changes: RMWD and occupancy JSD decrease, coverage and transition agreement improve, and local validity is unchanged, whereas autocorrelation and implied time scales do not improve. Together with CATH1, this result shows that Exploration-NFT consistently expands visited support, while the attainable gain and its kinetic consequences depend on the pretrained transition operator.

Figure 1 provides a structural interpretation of the aggregate metrics. For 1FME, successive checkpoints extend the pretrained distribution along the elongated MD manifold rather than perpendicular to it. GTT acquires most of its additional support by step 300 and subsequently saturates, whereas NTL9 continues to expand across a broad slow-coordinate region through step 500. The learned response is therefore protein dependent rather than a uniform increase in displacement, and coverage

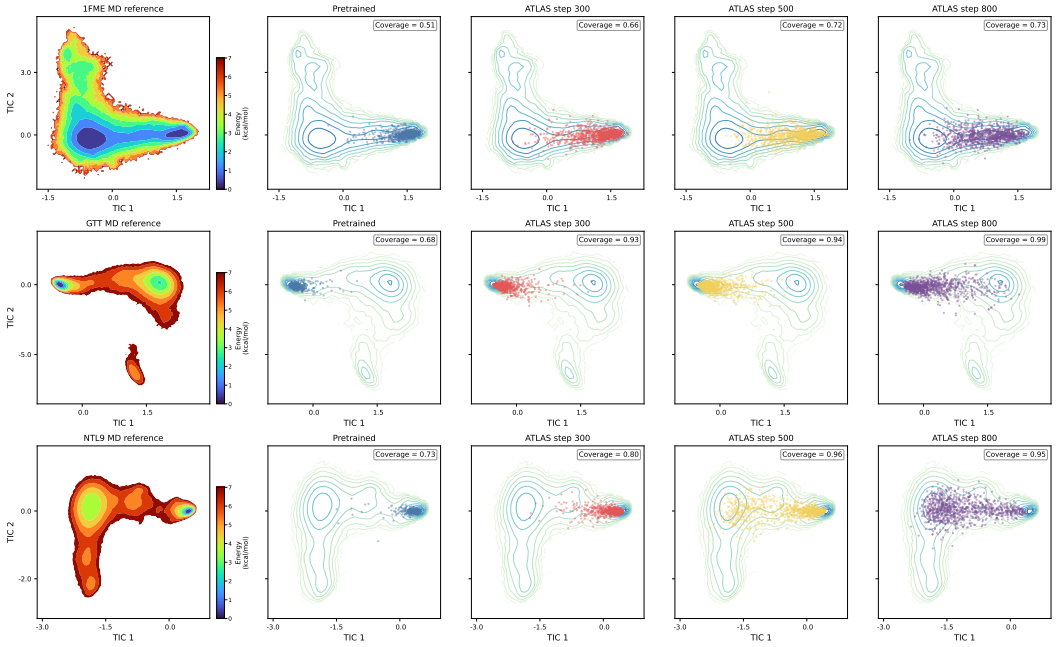

Figure 1: Reference-MD TICA landscapes for three representative fast-folding proteins: 1FME (top), GTT (middle), and NTL9 (bottom). Within each row, columns show MD, pretrained BioKinema, and Exploration-NFT steps 300, 500, and 800 in the same frozen, MD-fitted basis. Only frames passing all three 98% validity gates are overlaid.

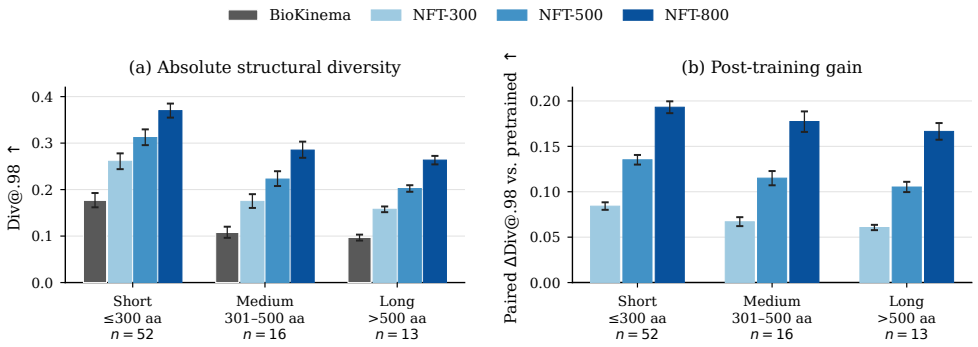

Figure 2: Length-stratified Div@.98 on 81 held-out ATLAS proteins. Short proteins ($\leq$ 300 residues) match the post-training support; medium (301–500) and long ($>$ 500) proteins test length transfer. Bars are unweighted means and error bars are 95% protein-bootstrap intervals.

gains diminish after the accessible grid becomes saturated. The full 12-protein set in Appendix F shows the same pattern of manifold-aligned, protein-dependent expansion.

5.3 Generalization across protein length and temporal interval

Figure 2 stratifies the 81 proteins into 52 short, 16 medium, and 13 long cases. At step 800, Div@.98 rises from 0.177 to 0.370 for short proteins, from 0.108 to 0.285 for medium proteins, and from 0.097 to 0.263 for long proteins. Every checkpoint improves every protein in all three strata, including the 29 medium and long targets excluded by the training length filter. This transfer is consistent with the reward construction: it selects a fixed number of within-protein distance features and measures motion relative to each protein's own start, rather than learning a residue-indexed target landscape. The

larger relative gains above 300 residues should not be interpreted as greater intrinsic efficacy on long proteins, because those strata have lower pretrained diversity; the paired improvement nevertheless demonstrates that the learned exploration behavior is not restricted to the training length regime over the tested 39–724-residue range.

BioKinema also supports 0.1–10 ns temporal conditioning although post-training uses only 10 ns. We evaluate the step-500 policy zero-shot over a matched 1 μs horizon. Validity uses all native frames; other metrics use common physical times 10, 20, . . ., 1,000 ns so shorter intervals do not receive more evaluation samples. Table 5 reports that every protein improves coverage at every interval, accompanied by lower RMWD and occupancy JSD. The gain is therefore not an artifact of evaluating more frames at shorter strides. Its larger magnitude at 0.1 ns is instead consistent with a per-transition exploratory bias accumulating over more native transitions within the same physical horizon. Transition JSD changes slightly in the opposite direction, indicating that the learned support expansion transfers across temporal conditioning, whereas kinetic calibration does not transfer automatically. The full metric table is in Appendix D.

6 Limitations

First, each model–protein condition uses one rollout seed. Protein-level coverage trends are consistent, but independent trajectory replicates are needed for precise uncertainty. Second, the training TICA basis is estimated from only the sampled group. Avoiding whitening preserves scale and evaluation uses an independent MD basis, yet finite-group statistics and grid resolution remain exploitable. Third, visited support specifies neither equilibrium populations nor kinetic calibration. CATH1 exposes this distinction: persistent optimization expands support while eventually worsening occupancy and transition statistics. Future objectives should constrain kinetics or replay the pretrained distribution while preserving exploration credit. Finally, the architectural study covers two generators; DyneTrion uses a single post-training checkpoint, so broader claims require more backbones and seeds.

7 Conclusion

Exploration-NFT addresses a specific consequence of short-horizon supervision: a locally accurate transition model may systematically underproduce rare exits when recursively rolled out. Group-fitted slow coordinates turn continued discovery into segment-level credit, while forward-process optimization preserves the generator's native sampler and local structural prior. Across BioKinema and DyneTrion, this intervention expands valid conformational support at both 100-ns and 10-μs horizons and transfers across sequence length and temporal conditioning. Agreement among coverage, RMWD, and occupancy on fast-folding proteins further shows that reinforcement can recover MD-supported regions rather than induce unstructured coordinate dispersion. More broadly, the results identify conformational support, equilibrium weighting, and kinetic calibration as separate levels of a learned dynamics model and provide a practical mechanism for directly optimizing the first.

Reproducibility Statement

The accompanying anonymized artifacts include reward, adapter, training, rollout, and evaluation implementations; executable configurations; fixed protein/MSA manifests; checkpoint fingerprints; and per-protein metric outputs. Appendix A records method and adapter settings, while Appendix C specifies metric formulas, validity masks, reference pools, temporal pairing, aggregation, and missing-value handling.

Ethics Statement

Generated trajectories are approximate model outputs, not experimental evidence. In particular, our results demonstrate that improved conformational coverage can coexist with miscalibrated kinetics. Scientific or therapeutic conclusions based on these trajectories should be validated against MD simulations and/or experiments and should report structural, thermodynamic, and kinetic diagnostics rather than a single exploration score.

AI USE STATEMENT

Generative AI tools were used to aid and polish manuscript writing; retrieve, summarize, and organize related literature; and support research ideation and execution.

REFERENCES

Jai Bardhan, Patrik Drozdik, Josef Sivic, and Vladimir Petrik. Persistent robot world models: Stabilizing multi-step rollouts via reinforcement learning. *arXiv preprint arXiv:2603.25685*, 2026. URL https://arxiv.org/abs/2603.25685.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=pBN8CgrDhE.

Kaihui Cheng, Zhiqiang Cai, Peng Tu, Yisong Yao, Limei Han, Libo Wu, Siyu Zhu, Tzuhsiung Yang, and Yuan Qi. DyneTrion: A spatio-temporally coherent generative emulator for protein dynamics across timescales. *arXiv preprint arXiv:2607.15309*, 2026a. URL https://arxiv.org/abs/2607.15309.

Kaihui Cheng, Zhiqiang Cai, Wenkai Xiang, Zhihang Hu, Siyu Zhu, Tzuhsiung Yang, and Yuan Qi. Learning implicit bias in generative spaces for accelerating protein dynamics emulation. *arXiv preprint arXiv:2606.01833*, 2026b. URL https://arxiv.org/abs/2606.01833.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3497.

Bin Feng, Jiying Zhang, Xinni Zhang, He Cao, Ming Zhang, Patrick Barth, Zijing Liu, and Yu Li. Physically grounded generative modeling of all-atom biomolecular dynamics. *bioRxiv*, 2026. doi: 10.64898/2026.02.15.705956.

Bowen Jing, Bonnie Berger, and Tommi Jaakkola. AlphaFold meets flow matching for generating protein ensembles. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 22277–22303, 2024a. URL https://proceedings.mlr.press/v235/jing24a.html.

Bowen Jing, Hannes Stärk, Tommi Jaakkola, and Bonnie Berger. Generative modeling of molecular dynamics trajectories. In *Advances in Neural Information Processing Systems*, volume 37, 2024b. doi: 10.52202/079017-1282.

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

Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew Y. K. Foong, Victor García Satorras, Osama Abdin, Bastiaan S. Veeling, Iryna Zaporozhets, Yaoyi Chen, Soojung Yang, Adam E. Foster, Arne Schneuing, Jigyasa Nigam, Federico Barbero, Vincent Stimper, Andrew Campbell, Jason Yim, Marten Lienen, Yu Shi, Shuxin Zheng, Hannes Schulz, Usman Munir, Roberto Sordillo, Ryota Tomioka, Cecilia Clementi, and Frank Noé. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761): eadv9817, 2025. doi: 10.1126/science.adv9817.

Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011. doi: 10.1126/science.1208351.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-GRPO: Training flow matching models via online RL. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-1362.

Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. VAMPnets for deep learning of molecular kinetics. *Nature Communications*, 9:5, 2018. doi: 10.1038/s41467-017-02388-1.

Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013. doi: 10.1093/bioinformatics/btt473.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019. doi: 10.1126/science.aaw1147.

Guillermo Pérez-Hernández, Fabian Paul, Toni Giorgino, Gianni De Fabritiis, and Frank Noé. Identification of slow molecular order parameters for markov model construction. *The Journal of Chemical Physics*, 139(1):015102, 2013. doi: 10.1063/1.4811489.

Yuning Shen, Lihao Wang, Huizhuo Yuan, Yan Wang, Bangji Yang, and Quanquan Gu. Simultaneous modeling of protein conformation and dynamics via autoregression. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-3666.

Yann Vander Meersche, Gabriel Cretin, Aria Gheeraert, Jean-Christophe Gelly, and Tatiana Galochkina. ATLAS: Protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research*, 52(D1):D384–D392, 2024. doi: 10.1093/nar/gkad1084.

Ziyang Yu, Wenbing Huang, and Yang Liu. Unified biomolecular trajectory generation via pretrained variational bridge. In *International Conference on Learning Representations*, 2026. URL `https://arxiv.org/abs/2602.07588`.

Kaiwen Zheng, Huayu Chen, Haotian Ye, Haoxiang Wang, Qinsheng Zhang, Kai Jiang, Hang Su, Stefano Ermon, Jun Zhu, and Ming-Yu Liu. DiffusionNFT: Online diffusion reinforcement with forward process. In *International Conference on Learning Representations*, 2026. URL `https://arxiv.org/abs/2509.16117`.

A Implementation Details

Table 4: Exploration-NFT hyperparameters shared by the BioKinema and DyneTrion experiments. Backbone architecture, native sampling, temporal resolution, denoising steps, and numerical precision are intentionally excluded.

Quantity	Value
Group size	8
Temporal reward segments	10
TICA lag / retained dimensions	1 generated frame / 2
Cα-distance features	top 128 by group standard deviation
Minimum sequence separation	4 residues
TICA ridge	10^{-4}, trace-scaled
Hard-grid spacing	0.03
Invalidity coefficient λ_{inv}	0.5
Per-frame validity threshold	98%
Advantage stabilizer / clipping c	10^{-6} / 2
Contrast floor	$\max(10^{-4}, 0.1 s_k^{(0)})$
Late-segment exponent / cap	0.5 / 3.0
Forward-process mixing β	0.1
Frozen-reference coefficient λ_{ref}	0.001

Table 4 is restricted to quantities that define Exploration-NFT. Model checkpoints, sampler controls, block lengths, denoising steps, frame intervals, and mixed precision belong to the backbone adapter and are not algorithm hyperparameters.

Backbone adapters. BioKinema is initialized from its ATLAS+MISATO+MDposit checkpoint. A training rollout contains two native direct/churn blocks of 50 generated frames, with 10 ns per frame, 20 denoising steps, noise scale 1.75, and BF16 autocast. Its diffusion module is trainable while the sequence/MSA-conditioning trunk stays frozen. The collection-policy EMA ramps from 0 to 0.9

during the first 100 steps. DyneTrion uses its native SDE sampler, ten 16-frame blocks at 1 ns per frame, 50 denoising steps, and BF16 forward replay; its temporal denoising pathway is trainable and the collection EMA is capped at 0.5. In both cases, all 940 training proteins have at most 300 residues, one group is collected per data-parallel rank, Adam uses learning rate 10^{-6}, and gradient norm is clipped at 1.0. TICA fitting and the generalized eigensolve run on CPU in FP64.

Training proteins are length-bucketed so ranks within a distributed step have similar compute while bucket order is shuffled. Forward replay uses candidate micro-batches of one to bound activation memory. Fine-tuned inference reconstructs the complete pretrained backbone and loads the updated denoising-module state.

B Training Reward and Optimization Details

B.1 TICA estimation

Let $\bar{u}_{g,t} = u_{g,t} - \mu$ denote a feature vector centered by the mean of all frames in the rollout group. Lagged observations never cross candidate boundaries. With lag ℓ, the symmetric covariance estimates are

$$C_{00} = \frac{1}{2G(T+1-\ell)} \sum_{g,t} \left(\bar{u}_{g,t} \bar{u}_{g,t}^{\mathsf{T}} + \bar{u}_{g,t+\ell} \bar{u}_{g,t+\ell}^{\mathsf{T}} \right). \tag{12}$$

$$C_{0\ell} = \frac{1}{2G(T+1-\ell)} \sum_{g,t} \left(\bar{u}_{g,t} \bar{u}_{g,t+\ell}^{\mathsf{T}} + \bar{u}_{g,t+\ell} \bar{u}_{g,t}^{\mathsf{T}} \right). \tag{13}$$

The trace-scaled regularizer used in every experiment is

$$\rho = 10^{-4} \frac{\mathrm{tr}(C_{00})}{M}. \tag{14}$$

Before solving Equation 3, covariance eigenvectors whose eigenvalue is at most 10^{-8} times the largest eigenvalue are removed. Ridge is added inside the retained subspace. We reject a group if the effective rank is below two, the generalized eigenvalues violate their theoretical range beyond 10^{-6}, or the relative eigensolver residual exceeds 10^{-10}. A QR step Euclidean-orthonormalizes the two generalized eigenvectors. This fixes a reproducible coordinate scale without setting the observed variance of every group to one.

B.2 Physical-validity penalty

For training, let B be the number of consecutive within-chain residue bonds. At frame t, the bond-valid fractions are

$$q_t^{\mathrm{CA}} = \frac{1}{B} \sum_{r=1}^{B} \mathbf{1}\left(\|x_{t,r,\mathrm{CA}} - x_{t,r+1,\mathrm{CA}}\|_2 < 4.5 \text{ Å} \right). \tag{15}$$

$$q_t^{\mathrm{CN}} = \frac{1}{B} \sum_{r=1}^{B} \mathbf{1}\left(\|x_{t,r,\mathrm{C}} - x_{t,r+1,\mathrm{N}}\|_2 < 2.0 \text{ Å} \right). \tag{16}$$

For clashes, only N, CA, C, CB, and O are considered. Atom pairs are nonlocal if they belong to different chains or their residues are separated by at least three positions. A residue is marked clashing if any such present atom pair containing it is at most 1.0 Å apart. With R residues,

$$q_t^{\text{res-free}} = 1 - \frac{1}{R} \sum_{r=1}^{R} \mathbf{1}(\text{residue } r \text{ is clashing or non-finite}). \tag{17}$$

The training-frame indicator is

$$v_t = \mathbf{1}\big(q_t^{\mathrm{CA}} \geq 0.98,\ q_t^{\mathrm{CN}} \geq 0.98,\ q_t^{\text{res-free}} \geq 0.98\big)\,, \tag{18}$$

and $P_{g,k}^{\mathrm{invalid}} = 1 - |\mathcal{T}_k|^{-1} \sum_{t \in \mathcal{T}_k} v_{g,t}$. This residue-level clash gate is specific to the training penalty. The clash-free percentage displayed in evaluation tables is the non-clashing *candidate atom-pair* fraction defined in Appendix C.1; the two must not be conflated.

B.3 Frozen temporal calibration

Let $\mu_k^{(0)}$ be the mean $\Delta C_{g,k}$ and let $s_k^{(0)}$ be the median within-group reward standard deviation, both measured on pretrained step-0 validation rollouts. The raw difficulty factor is

$$a_k = \left(\frac{\mu_1^{(0)} + 10^{-4}}{\mu_k^{(0)} + 10^{-4}} \right)^{1/2}. \tag{19}$$

We make difficulty nondecreasing with horizon, cap it at three, and normalize the mean weight to one:

$$\widetilde{w}_k = \min\left(3, \max_{j \leq k} a_j\right), \qquad w_k = \frac{\widetilde{w}_k}{K^{-1} \sum_{j=1}^{K} \widetilde{w}_j}. \tag{20}$$

Segment k is skipped when its current reward standard deviation is below

$$\eta_k = \max\left(10^{-4}, 0.1 s_k^{(0)}\right). \tag{21}$$

The w_k and η_k are frozen for the run. Applying w_k after the per-segment z-score is essential: multiplying $r_{g,k}$ before standardization would cancel out.

B.4 Normalized forward-process loss

Let Ω be the valid coordinate mask for a candidate/segment and $e^s = D^s - y$. The two normalized reconstruction terms are

$$L_{g,k}^{s} = \frac{|\Omega|^{-1} \sum_{j \in \Omega} (e_j^s)^2}{\mathrm{sg}\left(|\Omega|^{-1} \sum_{j \in \Omega} |e_j^s|\right) + 10^{-5}}, \qquad s \in \{+, -\}. \tag{22}$$

Each adapter reuses the corruption distribution with which its backbone was pretrained. BioKinema samples either one noise level per frame and candidate or one shared level per candidate, each with probability $1/2$, and then applies its EDM corruption; the history frame receives the model's original 4×10^{-4} noise. DyneTrion uses its original diffusion forward process. The frozen-reference loss in Equation 11 compares current and pretrained denoiser outputs on the same noisy input. Current, collection, and reference networks have identical adapter-visible structure; collection weights are updated only after an optimizer step.

C Evaluation Metric Definitions

Fast-folding reference MD is stored at 0.2 ns per frame. Coverage uses the complete projection, whereas occupancy and kinetic metrics subsample each independent MD run to 10 ns. CATH1 reference MD is stored at 10 ns and all 10,221 frames from 21 official runs per protein are used. No lagged pair is formed across independent-run boundaries. Structure-validity percentages use all generated frames; the other metrics use the common validity mask below.

All ensemble and kinetic metrics use a protein-specific TICA representation fitted exclusively to reference MD. Its features are Cα distances

$$d_{t,ij} = \|x_{t,i,\mathrm{CA}} - x_{t,j,\mathrm{CA}}\|_2, \qquad |i-j| \geq 8, \tag{23}$$

and $z_t = \mathrm{TICA}_{\mathrm{MD}}(d_t)_{1:2}$. Generated data never refit the features, TICA basis, density grid, or microstate centers.

C.1 Structure validity and common frame mask

Evaluation uses the same CA and CN bond fractions as above. Let $\mathcal{P}$ be all present pairs among N, CA, C, O, and CB that come from different chains or residues separated by more than two positions. The displayed candidate atom-pair clash-free fraction is

$$q_t^{\text{pair-free}} = \frac{1}{|\mathcal{P}|} \sum_{(a,b)\in\mathcal{P}} \mathbf{1}\big(\|x_{t,a} - x_{t,b}\|_2 > 1.0\ \text{Å}\big). \tag{24}$$

The reported CA, CN, and Clash-Free percentages are

$$\mathrm{CA}\% = \frac{100}{T}\sum_t q_t^{\mathrm{CA}}, \quad \mathrm{CN}\% = \frac{100}{T}\sum_t q_t^{\mathrm{CN}}, \quad \mathrm{ClashFree}\% = \frac{100}{T}\sum_t q_t^{\text{pair-free}}. \tag{25}$$

A frame enters ensemble and kinetic evaluation only if

$$m_t = \mathbf{1}\Big[\mathrm{finite}(x_t),\ q_t^{\mathrm{CA}} \geq 0.98,\ q_t^{\mathrm{CN}} \geq 0.98,\ q_t^{\text{pair-free}} \geq 0.98\Big]. \tag{26}$$

Invalid frames are removed without compressing the physical time axis. A lagged pair is legal only when both original endpoints are valid.

C.2 Short-horizon structural diversity

For a validity threshold τ, let $\mathcal{V}_\tau$ contain finite frames whose three fractions are all at least τ. For frames i and j, let $\mathcal{A}_j$ contain ordered Cα pairs separated by less than 15 Å in target frame j. The evaluation uses

$$\mathrm{lDDT}_{\mathrm{C}\alpha}(i,j) = \frac{1}{4|\mathcal{A}_j|} \sum_{(r,s)\in\mathcal{A}_j} \sum_{\delta\in\{0.5,1,2,4\}\ \text{Å}} \mathbf{1}\big(|d_{rs}^i - d_{rs}^j| < \delta\big). \tag{27}$$

The diversity score is

$$\mathrm{Div@}\tau = \frac{2}{|\mathcal{V}_\tau|(|\mathcal{V}_\tau|-1)} \sum_{\substack{i<j \\ i,j\in\mathcal{V}_\tau}} \big[1 - \mathrm{lDDT}_{\mathrm{C}\alpha}(i,j)\big]. \tag{28}$$

It is zero if fewer than two frames survive. Div@.00 uses every finite frame, and

$$\text{mDiv-mean} = \frac{1}{3} \sum_{\tau\in\{0.98,0.99,1.00\}} \mathrm{Div@}\tau. \tag{29}$$

C.3 THERMODYNAMIC AND ENSEMBLE METRICS

Root mean Wasserstein distance. Common heavy atoms are expressed in Å and every frame is Kabsch-aligned to the initial MD structure. Let X be 1,000 reference frames sampled with replacement and let Y contain every valid generated frame. For atom a, let $\mu_{E,a}$ and $\Sigma_{E,a}$ be its three-dimensional population mean and covariance in ensemble E. Its Gaussian 2-Wasserstein term is

$$W_{2,a}^2(X,Y) = \|\mu_{X,a} - \mu_{Y,a}\|_2^2 + \operatorname{tr}\left[\Sigma_{X,a} + \Sigma_{Y,a} - 2\left(\Sigma_{X,a}^{1/2}\Sigma_{Y,a}\Sigma_{X,a}^{1/2}\right)^{1/2}\right]. \tag{30}$$

For N matched atoms,

$$\mathrm{RMWD}(X,Y) = \left(\frac{1}{N}\sum_{a=1}^{N} W_{2,a}^2(X,Y)\right)^{1/2}. \tag{31}$$

Low-free-energy TICA coverage. MD and valid generated projections are independently resampled with replacement to 10^6 points and perturbed by $\mathcal{N}(0, 0.25^2)$ in each TICA coordinate. A 50×50 grid is defined from MD only. For MD density ρ_j^{MD} in cell j,

$$F_j - F_{\min} = -k_B T \log \frac{\rho_j^{\mathrm{MD}}}{\max_l \rho_l^{\mathrm{MD}}}, \qquad k_B = 0.0019872036\ \mathrm{kcal\,mol^{-1}\,K^{-1}}. \tag{32}$$

With $\mathcal{L} = \{j : F_j - F_{\min} \le 4\ \mathrm{kcal/mol}\}$,

$$\mathrm{Coverage} = \frac{|\{j \in \mathcal{L} : \rho_j^{\mathrm{gen}} > 0\}|}{|\mathcal{L}|}. \tag{33}$$

Occupancy JSD. Twelve KMeans centers are fitted to reference-MD TICA coordinates sampled at 10 ns. Assigning every valid frame to its nearest center gives normalized occupancies $p = p_{\mathrm{MD}}$ and $q = p_{\mathrm{gen}}$. For $u = (p+q)/2$, we report the Jensen–Shannon distance

$$\mathrm{Occupancy\ JSD} = \left[\frac{1}{2}D_{\mathrm{KL}}(p\|u) + \frac{1}{2}D_{\mathrm{KL}}(q\|u)\right]^{1/2}, \tag{34}$$

where $D_{\mathrm{KL}}(a\|b) = \sum_{i:a_i>0} a_i \log(a_i/b_i)$. The column name uses the common abbreviation "JSD," but the implementation reports the square root of the divergence.

C.4 KINETIC METRICS

All kinetic metrics use lags $\mathcal{T} = \{10, 20, 50, 100, 200\}$ ns. For an ensemble composed of independent runs r, let

$$\mathcal{I}_{E,\ell} = \{(r,t) : t + \ell < T_{E,r},\ m_{E,r,t} = m_{E,r,t+\ell} = 1\} \tag{35}$$

be its legal endpoint pairs. The MD mask is one and generated masks follow Equation 26.

AutoCor-MAE. For TICA coordinate $d \in \{1, 2\}$,

$$\mathrm{ACF}_{E,d}(\ell) = \frac{\sum_{(r,t)\in\mathcal{I}_{E,\ell}} (z_{r,t,d} - \bar{z}_{E,d})(z_{r,t+\ell,d} - \bar{z}_{E,d})}{\sqrt{\sum_{(r,t)\in\mathcal{I}_{E,\ell}} (z_{r,t,d} - \bar{z}_{E,d})^2 \sum_{(r,t)\in\mathcal{I}_{E,\ell}} (z_{r,t+\ell,d} - \bar{z}_{E,d})^2}}. \tag{36}$$

We then compute

$$\text{AutoCor-MAE} = \frac{1}{10} \sum_{\ell \in \mathcal{T}} \sum_{d=1}^{2} \left| \text{ACF}_{\text{MD},d}(\ell) - \text{ACF}_{\text{gen},d}(\ell) \right|. \tag{37}$$

Transition JSD. Let $s_{E,r,t} \in \{1, \ldots, 12\}$ be the MD-fitted microstate. The globally normalized directed joint-transition distribution is

$$P^{(\ell)}_{E,ij} = \frac{1}{|\mathcal{I}_{E,\ell}|} \sum_{(r,t) \in \mathcal{I}_{E,\ell}} \mathbf{1}(s_{E,r,t} = i,\ s_{E,r,t+\ell} = j). \tag{38}$$

Transition JSD is the mean Jensen–Shannon distance between $\text{vec}\, P^{(\ell)}_{\text{MD}}$ and $\text{vec}\, P^{(\ell)}_{\text{gen}}$ over the five lags.

ITS-log-MAE. For each physical lag τ, let $C^{(\tau)}_E$ be the raw transition-count matrix. We form reversible counts $\widetilde{C} = C + C^{\mathsf{T}}$, remove inactive states, retain the connected component with the largest transition mass, and row-normalize:

$$K^{(\tau)}_{E,ij} = \frac{\widetilde{C}^{(\tau)}_{E,ij}}{\sum_j \widetilde{C}^{(\tau)}_{E,ij}}. \tag{39}$$

For the first three non-stationary real eigenvalues satisfying $0 < \lambda^{(\tau)}_{E,i} < 1 - 10^{-10}$,

$$t^{(\tau)}_{E,i} = -\frac{\tau}{\log \lambda^{(\tau)}_{E,i}}. \tag{40}$$

Let $\mathcal{Q}$ be lag–mode pairs for which both MD and generated timescales are finite and positive. Then

$$\text{ITS-log-MAE} = \frac{1}{|\mathcal{Q}|} \sum_{(\tau,i) \in \mathcal{Q}} \left| \log \frac{t^{(\tau)}_{\text{gen},i}}{t^{(\tau)}_{\text{MD},i}} \right|. \tag{41}$$

Missing modes are not imputed. If $\mathcal{Q}$ is empty, the protein-level metric is undefined. This occurs for pretrained Lambda in the fast-folding test, so its aggregate ITS-log-MAE uses 11 rather than 12 proteins.

Motion diagnostic. The lag-resolved mean TICA displacement and its ratio are

$$\Delta_E(\ell) = \frac{1}{|\mathcal{I}_{E,\ell}|} \sum_{(r,t) \in \mathcal{I}_{E,\ell}} \|z_{E,r,t+\ell} - z_{E,r,t}\|_2, \qquad \text{MotionRatio}(\ell) = \frac{\Delta_{\text{gen}}(\ell)}{\Delta_{\text{MD}}(\ell)}. \tag{42}$$

The ideal ratio is one. Values above one denote motion that is too large in the fixed MD-fitted coordinates, not necessarily a physically faster rate for each individual transition.

C.5 Aggregation and missing values

Every displayed entry is the arithmetic mean of finite protein-level values. Proteins are not weighted by sequence length, MD duration, number of valid frames, or number of transitions. AutoCor-MAE and Transition JSD require all five lags. ITS-log-MAE permits missing lag–mode pairs but records their count and is undefined only if no common pair remains. Machine-readable outputs retain all per-protein values, valid-frame counts, and finite-value counts.

D TEMPORAL-INTERVAL TRANSFER

For the 0.1, 1, and 10 ns conditions, native generated frame counts are 10,000, 1,000, and 100 over the same 1 μs horizon. Validity uses every native frame. To prevent shorter intervals from receiving more ensemble samples, every other metric uses the common physical times 10, 20, . . ., 1,000 ns.

Table 5: Zero-shot temporal-interval transfer on 12 fast-folding proteins (mean). BioKinema+Exploration-NFT is checkpoint 500, trained only at 10 ns.

		Structure validity			Thermodynamics			Kinetics		
Interval	Model	CA (%) ↑	CN (%) ↑	Clash-Free (%) ↑	RMWD (Å) ↓	Coverage ↑	Occupancy JSD ↓	AutoCor-MAE ↓	Transition JSD ↓	ITS-log-MAE ↓
0.1 ns	BioKinema	100.00	100.00	100.00	5.903	0.510	0.504	0.704	**0.580**	3.756
	+Exploration-NFT (500)	99.98	99.98	100.00	**4.364**	**0.790**	**0.437**	**0.600**	0.592	**2.521**
1 ns	BioKinema	100.00	100.00	100.00	6.442	0.514	0.532	0.701	**0.602**	4.110
	+Exploration-NFT (500)	99.99	99.98	100.00	**5.043**	**0.671**	**0.485**	**0.674**	0.607	**3.133**
10 ns	BioKinema	100.00	100.00	100.00	6.484	0.513	0.504	**0.792**	**0.575**	4.193
	+Exploration-NFT (500)	99.98	99.98	100.00	**5.243**	**0.658**	**0.469**	0.793	0.590	**3.629**

The larger coverage gain at 0.1 ns may arise because the post-trained transition is applied many more times over a fixed physical horizon. It should not be interpreted as interval-independent kinetic calibration: Transition JSD worsens slightly in all three matched comparisons.

E LAG-RESOLVED MOTION DIAGNOSTICS

Table 6: Generated-to-MD mean TICA-displacement ratio on fast-folding proteins. The ideal value is one.

Model	10 ns	20 ns	50 ns	100 ns	200 ns
BioKinema	0.874	0.822	0.755	0.705	0.642
+Exploration-NFT (300)	1.426	1.344	1.234	1.159	1.060
+Exploration-NFT (500)	1.819	1.730	1.593	1.490	1.387
+Exploration-NFT (800)	2.221	2.095	1.935	1.813	1.684

F ALL FAST-FOLDING TICA LANDSCAPES

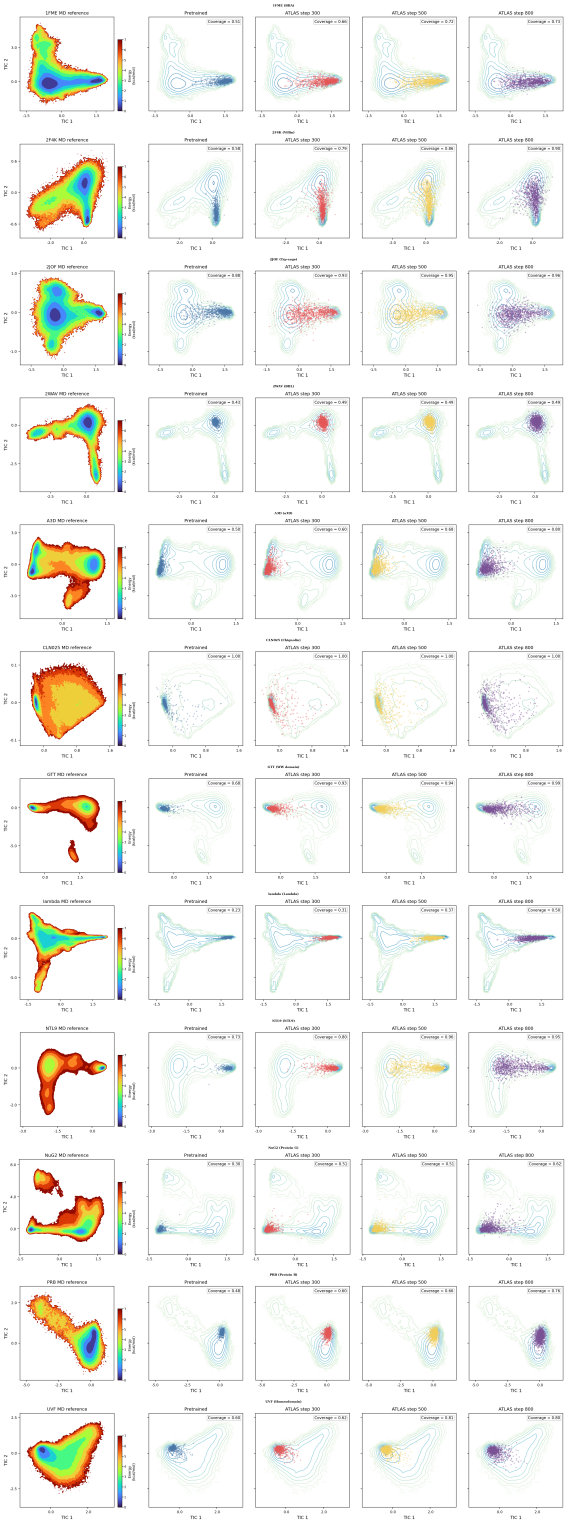

Figure 3: Complete fast-folding TICA comparison. Columns show reference MD, pretrained BioKinema, and Exploration-NFT steps 300, 500, and 800. Each row uses its own MD-fitted basis and recorded reference temperature. Generated frames must pass all three 98% validity gates.