All-Atom GPCR-Ligand Simulation via Residual Isometric Latent Flow

Anonymous Authors[1]

Abstract

G-protein-coupled receptors (GPCRs)—primary targets for over one-third of approved therapeutics—rely on intricate conformational transitions to transduce signals. While Molecular Dynamics (MD) is essential for elucidating this transduction process, particularly within ligand-bound complexes, conventional all-atom MD simulation is computationally prohibitive. In this paper, we introduce `GPCRLMD`, a deep generative framework for efficient all-atom GPCR-ligand simulation. `GPCRLMD` employs a Harmonic-Prior Variational Autoencoder (HP-VAE) to first map the complex into a regularized isometric latent space, preserving geometric topology via physics-informed constraints. Within this latent space, a Residual Latent Flow samples evolution trajectories, which are subsequently decoded back to atomic coordinates. By capturing temporal dynamics via relative displacements anchored to the initial structure, this residual mechanism effectively decouples static topology from dynamic fluctuations. Experimental results demonstrate that `GPCRLMD` achieves state-of-the-art performance in GPCR-ligand dynamics simulation, faithfully reproducing thermodynamic observables and critical ligand-receptor interactions.

1. Introduction

Comprising over 800 distinct members, G protein-coupled receptors (GPCRs) form an essential class of seven-transmembrane proteins (Figure 1) that modulate a vast array of biological functions (Hauser et al., 2017; Aranda-García et al., 2025; Conflitti et al., 2025). Their ubiquity and accessibility have made them the cornerstone of modern medicine, with more than one-third of all marketed drugs targeting this protein family (Lorente et al., 2025). Understanding the mechanism of these therapeutics requires characterizing GPCR-ligand complexes, where the dynamic coupling between the ligand and the receptor shapes the conformational landscape necessary for signal transduction (Kapla et al., 2021; Aranda-García et al., 2025).

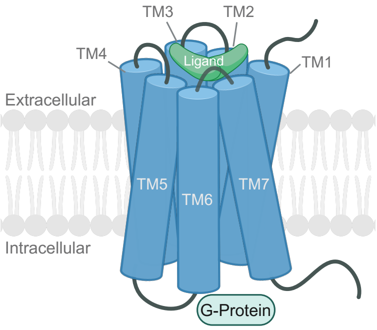

Figure 1. Schematic structure of a ligand-bound GPCR. A GPCR contains seven transmembrane helices (TM1-TM7). It senses ligands and transfers this signal into the cell.

Molecular Dynamics (MD) simulations serve as a critical tool for elucidating the kinetic and thermodynamic mechanisms governing biomolecular function (Karplus & McCammon, 2002; Torrens-Fontanals et al., 2020; Frenkel & Smit, 2023). This approach is particularly indispensable for investigating GPCR-ligand complexes (Ciancetta et al., 2015; Latorraca et al., 2017), as the receptor adopts distinct conformational ensembles dictated by the functional profile of the bound ligand (Zhang et al., 2024; Conflitti et al., 2025). However, conventional MD is limited by its sampling efficiency (Lindorff-Larsen et al., 2011). Relying on step-wise numerical integration to propagate dynamics, adequately exploring the vast, high-dimensional conformational landscape of protein-ligand interactions entails prohibitive computational costs (Newport et al., 2019; Rodríguez-Espigares et al., 2020; Siebenmorgen et al., 2024). This often renders the generation of comprehensive statistical ensembles a significant challenge, even with enhanced sampling methods (Roessner et al., 2025; Li et al., 2025).

Deep generative learning has emerged as a transformative paradigm for accelerating molecular simulations, offering a data-driven shortcut to bypass the high cost of physical integration steps (Feng et al., 2025; Jing et al., 2024b; Lewis et al., 2025). Within the GPCR domain specifically, recent studies have successfully leveraged these models to capture the diverse conformational ensembles of the receptor itself (Sengar et al., 2025a;b). However, modeling the dy-

---

[1] Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Submitted to the AI for Science workshop (ICML 2026).

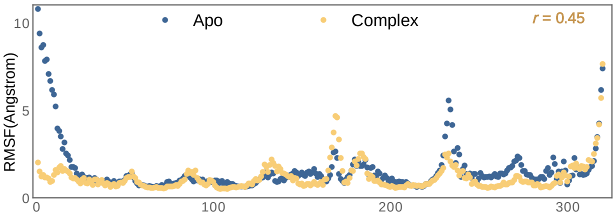

Figure 2. Comparison of Root Mean Square Fluctuation (RMSF) between the apo (ligand-free) and complex (ligand-bound) states. The receptor exhibits significantly different fluctuation profiles when bound to the ligand (Pearson correlation $r = 0.45$), demonstrating that ligand binding substantially modulates receptor dynamics. Results are shown by residue index for PDB ID 6LW5, utilizing trajectory data from Aranda-García et al. (2025).

namic evolution of the GPCR-ligand complex remains a critical unmet need. As shown in Figure 2, the presence of a ligand induces distinct fluctuation patterns compared to the apo state. The introduction of the ligand significantly increases the system's complexity (Aranda-García et al., 2025; Conflitti et al., 2025), creating a coupled dynamic challenge that existing generative frameworks have yet to address efficiently.

In this work, we address this challenge by introducing GPCRLMD, a deep generative framework designed for the trajectory forecasting setting—predicting the temporal evolution of the GPCR-ligand system given an initial conformational state (Jing et al., 2024b; Feng et al., 2025). GPCRLMD uniquely integrates an Atom Anchored Thermal Variational Autoencoder to transform the atomic coordinates of the GPCR-ligand complex into a regularized, isometric latent space. Unlike traditional dimensionality reduction that sacrifices structural detail, this approach preserves full spatial resolution while mapping the molecular distribution onto a tractable manifold governed by physics-informed constraints. Furthermore, by leveraging attention mechanisms to fuse receptor-ligand features, our model can capture the critical non-covalent interactions driving allostery. Subsequently, we employ a Residual Latent Flow model to capture temporal dynamics. By explicitly learning the relative displacement field anchored to the initial structure, our model effectively decouples static topology from dynamic fluctuations, enabling the efficient and parallel generation of future states. Experimental results demonstrate that GPCRLMD generates biophysically realistic and structurally diverse molecular motions that are consistent with known ensemble and kinetic properties.

In summary, our main contributions are highlighted as follows: **Benchmark Curation:** We establish a comprehensive benchmark for GPCR-ligand complex dynamics derived from the GPCRMD dataset (Aranda-García et al., 2025). Unlike existing protocols that focus on protein-only systems, we evaluate model performance in capturing the coupled dynamics of the receptor-ligand interface. **Novel Framework:** We propose GPCRLMD, a deep generative framework for all-atom GPCR-ligand simulation. It comprises: (1) an Atom-Anchored Thermal VAE that maps complexes into an isometric latent space via an atom-centered prior; and (2) Residual Latent Flow Matching. By anchoring latent displacements to the initial Cartesian frame, we establish an explicit mathematical linkage between physical and generative spaces. This decouples static equilibrium from stochastic dynamics, forcing the network to focus purely on transition probabilities rather than reconstructing covalent geometries. **State-of-the-Art Performance:** We comprehensively evaluate GPCRLMD on both ensemble and trajectory prediction benchmarks. The results demonstrate that our method achieves state-of-the-art performance across key fidelity and diversity metrics. Notably, GPCRLMD reproduces the biologically critical transmembrane helix motions characteristic of GPCR dynamics, ensuring high structural plausibility. Demonstrating broader applicability, we extend our framework to general protein-ligand systems and achieve state-of-the-art results on the MISATO (Siebenmorgen et al., 2024) dataset.

2. Related work

Molecular Dynamics Simulation with Generative Models. Generative model-based molecular dynamics (MD) simulation methods (Dao & Rahman, 2025) have attracted significant attention, as they can efficiently produce trajectories or ensembles compared to conventional MD simulations (Han et al., 2024; Lu et al., 2024; Wu & Li, 2023). In particular, protein-centric approaches have achieved impressive performance in capturing conformational ensembles of the receptor itself at varying resolutions (Lu et al., 2024; Lewis et al., 2025; Jing et al., 2024a;b; Lu et al., 2025; Sengar et al., 2025a;b; Shen et al., 2025). However, these methods generally ignore non-protein components and thus cannot be directly applied to protein-ligand complexes. Additionally, while methods like NeuralMD (Liu et al., 2023a) and UnbindingFlow (Li et al., 2025) focus on simulating ligand dynamics, they fail to capture fine-grained protein fluctuations. Although BioMD (Feng et al., 2025) recently attempted to address all-atom protein-ligand simulation, it is constrained by the Cartesian coordinate representation and lacks the necessary GPCR domain knowledge. As a result, it fails to achieve satisfactory performance on GPCR-ligand dynamics (Table 1). Therefore, all-atom GPCR-ligand complex simulation remains an open challenge.

Structure Generation in Latent Space. Mansoor et al. (2024) demonstrate that Variational Autoencoders (VAEs) offer a more suitable representation for generating plausible ensembles compared to Cartesian space. Furthermore, several studies (Geffner et al., 2025; Xu et al., 2023; Samaddar et al., 2025) indicate that latent spaces capture continuous and smooth structural information, thereby enhancing the robustness of generative models in producing diverse and physically valid biomolecular structures (Kong et al., 2024; 2025). While recent approaches (Sengar et al., 2025a;b) have leveraged this to generate GPCR conformation ensembles, modeling receptor-ligand complexes within latent space remains largely under-explored. In this work, we bridge this gap by anchoring latent displacements directly to the initial Cartesian frame. To the best of our knowledge, no existing work in the molecular domain has established such an explicit mathematical linkage between physical Cartesian space and the generative latent space. Due to space constraints, related work on **GPCR Dynamic Simulations** is provided in Appendix A.

3. Preliminaries

Notation. The trajectory of the complex is denoted as $\mathcal{X}_T = [\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{T-1}]$, where $\mathbf{x}_t = [\mathbf{x}_t^{\mathcal{P}}, \mathbf{x}_t^{\ell}] \in \mathbb{R}^{N\times 3}$ represents the Cartesian coordinates of GPCR $\mathcal{P}$ and ligand ℓ at time step t. The corresponding latent trajectory is denoted as $\mathcal{Z}_T = [\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{T-1}]$. In this paper, we define the latent space as *isometric* if the latent representation preserves the same dimensionality as the input structure (i.e., $\mathbf{z}_t \in \mathbb{R}^{N\times 3}$).

3.1. Molecular Dynamics

Molecular Dynamics (MD) is an *in silico* technique that simulates the motion of atoms and molecules under near-physiological conditions. In classical MD, the time evolution of the system is governed by Newton's equations of motion, $m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i$, where m_i and $\mathbf{r}_i$ denote the mass and position of atom i, and $\mathbf{F}_i = -\nabla U(\mathbf{r}_i)$ is the force derived from the potential energy U. This links atomic interactions and system energetics with dynamical behavior. In practice, MD employs numerical integration schemes such as the Verlet algorithm (Verlet, 1967) or stochastic approaches like Langevin dynamics to propagate trajectories that approximate the Boltzmann distribution, thereby enabling the estimation of thermodynamic and kinetic properties. Despite its utility in revealing atomistic mechanisms, MD remains computationally demanding, restricting accessible timescales and system sizes.

3.2. Variational Autoencoder (VAE)

A variational autoencoder (VAE) (Kingma et al., 2019) is a probabilistic generative model that learns to map the original data distribution into a latent space, which is typically assumed to follow a standard Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Unlike a deterministic autoencoder, the VAE introduces a variational posterior $q_\theta(\mathbf{z} \mid \mathbf{x})$ to approximate the true but intractable posterior $p(\mathbf{z} \mid \mathbf{x})$. From a statistical perspective, training a VAE is equivalent to maximizing the evidence lower bound (ELBO) of the marginal likelihood $p(\mathbf{x})$:

$$\log p(\mathbf{x}) \geq \mathbb{E}_{q_\theta(\mathbf{z}|\mathbf{x})}\big[\log p_\phi(\mathbf{x} \mid \mathbf{z})\big] - \mathrm{KL}\big(q_\theta(\mathbf{z} \mid \mathbf{x}) \,\|\, p(\mathbf{z})\big),$$

where $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the prior over the latent variable $\mathbf{z}$, $p_\phi(\mathbf{x} \mid \mathbf{z})$ is the likelihood (decoder), which is usually parameterized as Gaussian distribution, and $\mathrm{KL}(\cdot \,\|\, \cdot)$ denotes the Kullback–Leibler divergence.

3.3. Flow Matching

Flow matching (Lipman et al., 2023) is a class of generative models that learns continuous transformations between probability distributions by parameterizing the dynamics of an ordinary differential equation (ODE). Formally, given a flow $\phi_\tau : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$, its dynamics are described by $\frac{d\phi_\tau(\mathbf{x})}{d\tau} = v(\phi_\tau(\mathbf{x}), \tau)$, where $v : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ is a time-dependent vector field (or velocity). The goal of flow matching is to learn a parameterized vector field v_θ such that the induced flow ϕ_τ transports samples from a simple base distribution (e.g., $\mathcal{N}(\mathbf{0}, \mathbf{I})$) to the target data distribution.

4. Our Method: `GPCRLMD`

Direct modeling of molecular conformations in Cartesian coordinate space often suffers from inherent instability, as even small perturbations can lead to unphysical distortions or violations of steric constraints. To address this, we introduce a latent representation governed by a **physics-inspired atom-anchored prior**, constructing a structured and regularized manifold for generative modeling. Unlike standard approaches that compress diverse geometries into a generic isotropic Gaussian space (Kingma et al., 2019), our design utilizes an *isometric* latent space anchored to atomic coordinates. This provides a stable foundation for probabilistic learning, enabling the flow model to capture smooth conformational transitions while avoiding the irregularities characteristic of raw Cartesian coordinates. A decoder subsequently maps these refined latent variables back to precise three-dimensional positions, ensuring physical plausibility. This design aligns with recent geometric representation learning advancements (Kong et al., 2024; Mansoor et al., 2024; Geffner et al., 2025; Kong et al., 2025), demonstrating the effectiveness of decoupling representation robustness from generative dynamics.

4.1. Atom-Anchored Thermal VAE

Unlike standard autoencoders that provide deterministic point estimates, a Variational Autoencoder (VAE) introduces a probabilistic latent space to capture the intrinsic uncertainty and conformational entropy of molecular systems. However, the standard Gaussian prior used in conventional VAEs is structurally agnostic, forcing spatially distinct atoms to collapse toward a featureless origin. Although recent work considers coarse-grain geometric constraints in the latent space (Kong et al., 2024), this ignores the fundamental all-atom geometric nature of molecules, especially in the dynamics simulation. To overcome this, we propose an Atom Anchored Thermal VAE (AAT-VAE), which imposes

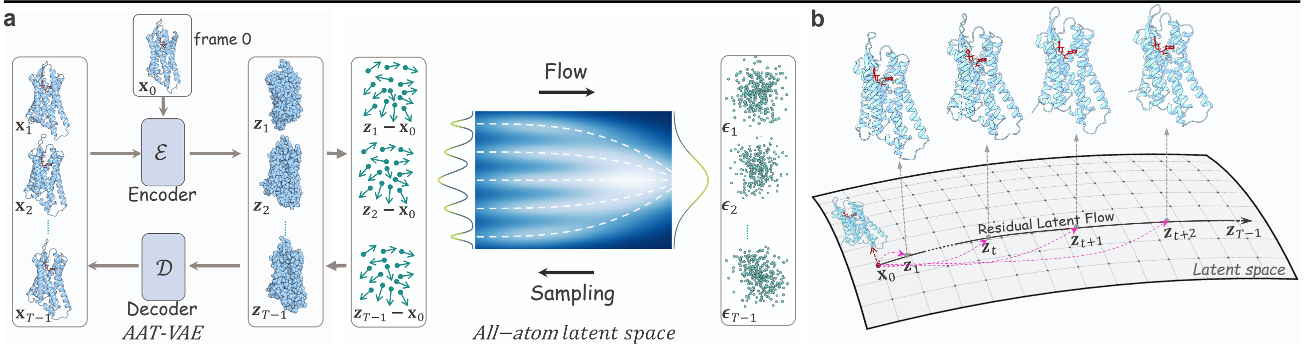

Figure 3. Pipeline of `GPCRLMD`. **a) Framework Overview.** The complete pipeline maps the GPCR-ligand trajectory $\{\mathbf{x}_i\}_{i=1}^{T-1}$ into an isometric latent space via **AAT-VAE**, enabling the Flow Matching network to learn dynamics on a smooth, structure-preserving manifold. **b) Residual Latent Flow.** To effectively decouple the static equilibrium structure from stochastic dynamic fluctuations, the flow models the relative evolution $\mathbf{r}_t = \mathbf{z}_t - \mathbf{x}_0$ anchored to the initial frame $\mathbf{x}_0$. The detailed network architecture can be found in App. Figure 6.

a physically grounded constraint on the all-atom isometric latent space.

All Atom Anchored Thermal Prior. In molecular dynamics (MD), atomic motions are often modeled as fluctuations around equilibrium positions governed by local force fields. Inspired by this, we replace the standard Gaussian prior with an **data-dependent atom-centered prior** that respects the instantaneous geometry of the molecule. Because the generative dynamics in our framework are handled subsequently by the flow matching model, the VAE is tasked with probabilistically regularizing the representation space rather than acting as a standalone zero-shot generator. Thus, for each atom i with coordinate $\mathbf{x}_i$, we formulate the variational autoencoding objective with a spatially-anchored regularization:

$$\mathcal{J}(\mathbf{x}_i) = \mathbb{E}_{q_\theta(\mathbf{z}_i|\mathbf{x}_i,\mathbf{c}_i)}\left[\log p_\phi(\mathbf{x}_i \mid \mathbf{z}_i, \mathbf{c}_i)\right] \tag{1}$$

$$-\mathrm{KL}\big(q_\theta(\mathbf{z}_i \mid \mathbf{x}_i, \mathbf{c}_i) \,\|\, \mathcal{N}(\mathbf{x}_i, \sigma\mathbf{I})\big), \tag{2}$$

where $\mathbf{c}_i$ denotes conditional context (e.g., residue type, atom type, transmembrane helix indices). Crucially, the prior $p(\mathbf{z}_i) = \mathcal{N}(\mathbf{x}_i, \sigma\mathbf{I})$ serves as a harmonic regularization term. This choice is motivated by three key physical considerations:

1. **Harmonic Approximation:** The log-probability of our prior $p(\mathbf{z}_i)$ corresponds to a quadratic penalty term $\|\mathbf{z}_i - \mathbf{x}_i\|^2$. This mimics the harmonic approximation of local potential energy wells used in normal mode analysis (Hinsen, 2005; Zhou et al., 2024), keeping latent codes physically near their atomic centers.
2. **Thermal Fluctuation:** Atoms in MD simulations naturally exhibit thermal noise around mean positions (Hoyt et al., 2010), quantified by Root-Mean-Square Fluctuation (RMSF) (Frenkel & Smit, 2023; Karplus & McCammon, 2002; Dong et al., 2018). This formulation naturally aligns the probabilistic latent noise with the intrinsic stochasticity of atomic vibrations, treating randomness as a physical property.
3. **Geometric Preservation:** By anchoring $\mathbf{x}_i$, we prevent the *posterior collapse* often seen in VAEs and ensure the latent space retains the explicit topology of the protein-ligand complex. This creates a compact, informative manifold that filters high-frequency noise while preserving essential structural motifs.

Training Objectives. We input all frames from one trajectory $\mathcal{X}_T$ into the encoder q_θ. Implementation-wise, the encoder is built upon the Diffusion Transformer (DiT) architecture (details in Sec. 4.3), which employs attention layers to deeply fuse the protein and ligand features, ensuring that the latent codes capture the coupled allosteric effects. The encoder q_θ output the corresponding latent coordinates $\mathcal{Z}$, which are subsequently reconstructed by the decoder p_ϕ as $\hat{\mathcal{X}}_T$. The total VAE objective aggregates the reconstruction error and the regularization:

$$\mathcal{L}_{vae} = \frac{1}{TN} \sum_{t=0}^{T-1} \sum_{i=1}^{N} \Big(\|\hat{\mathbf{x}}_{t,i} - \mathbf{x}_{t,i}\|_2^2 \tag{3}$$

$$+ \lambda_{\mathrm{KL}} \cdot \mathrm{KL}\big(\mathcal{N}(\mathbf{z}_{t,i}; \mu_\theta, \sigma_\theta) \,\|\, \mathcal{N}(\mathbf{x}_{t,i}, \sigma\mathbf{I})\big)\Big) + \mathcal{L}_{aux} \tag{4}$$

where $\mathbf{z}_{t,i}$ is sampled from the encoder distribution. The reconstruction term corresponds to the negative log-likelihood under an isotropic Gaussian decoder. In practice, we use weighted rigid aligned MSE (details in Appendix E.2) to ensure that the MSE loss remains invariant to global rotational and translational differences. The KL term explicitly penalizes deviations from the input geometry, effectively learning a *denoised* and *distinct* representation of the trajectory. Here, λ_{KL} is a trade-off hyperparameter. $\mathcal{L}_{aux}$ represents auxiliary geometric consistency losses (e.g., bond lengths, angles), detailed in Appendix F. To further enhance physical fidelity during long-trajectory sampling, we extend the framework to an Energy-Guided AAT-VAE, as detailed in Appendix E.1.

4.2. Residual Latent Flow Matching

Different from previous latent diffusion or flow approaches that compress molecular structures into low-dimensional semantic spaces (Sengar et al., 2025a;b), our method performs flow matching in an *isometric* all-atom-level latent space. This design preserves the original topological information to the greatest extent while enhancing the robustness of the generative process.

Crucially, we propose a residual generative paradigm to model the evolution of dynamics. Standard flow matching typically integrates an ODE trajectory from uninformative Gaussian noise $\mathcal{N}(\mathbf{0}, \mathbf{I})$ to the data distribution (Lipman et al., 2023). However, MD simulation is intrinsically an evolution process governing how a system propagates from an initial state $\mathbf{x}_0$, rather than emerging from chaos. Leveraging the property that our latent coordinates maintain the same dimensionality as the original input (dimension is 3), we explicitly leverage the initial frame from Cartesian space as a *structural prior*. We define the flow on the *relative displacement vector* between the current frame t and the initial frame $\mathbf{x}_0$ in the latent space. To the best of our knowledge, no existing work in the molecular domain has established such an explicit *mathematical linkage* between physical Cartesian space and the generative latent space. This formulation effectively *decouples* the static equilibrium structure from the stochastic dynamic fluctuations. By shifting the learning target to the displacement field, the model focuses its capacity on learning the transition probability density and the complex force fields required to cross energy barriers, rather than reconstructing the trivial covalent geometries from scratch (Figure 3b). Formally, we define the flow on the relative latent coordinate space. Let $\mathbf{r}_{t,i} = \mathbf{z}_{t,i} - \mathbf{x}_{0,i}$ denote the **residual vector**. The flow is defined as:

$$\frac{d\phi_\tau(\mathbf{r}_{t,i})}{d\tau} = v(\phi_\tau(\mathbf{r}_{t,i}), \tau), \tag{5}$$

$$\phi_0(\mathbf{r}_{t,i}) = \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \phi_1(\mathbf{r}_{t,i}) = \mathbf{r}_{t,i} = \mathbf{z}_{t,i} - \mathbf{x}_{0,i} \tag{6}$$

where $\tau \in [0, 1]$ represents the flow time. We employ the Rectified Flow framework (Lipman et al., 2023; Liu et al., 2023b; Hertrich et al., 2025) to learn straight paths between the noise and the target displacement. The intermediate state is interpolated as $\mathbf{r}^\tau_{t,i} := \tau\mathbf{r}^1_{t,i} + (1-\tau)\epsilon$. The target velocity is simply $\mathbf{r}^1_{t,i} - \mathbf{r}^0_{t,i} = \frac{\mathbf{r}^1_{t,i} - \mathbf{r}^\tau_{t,i}}{1-\tau}$. We parameterize a neural network v_θ to regress the clean data $\mathbf{r}^1_{t,i}$ (Stark et al., 2023), yielding the velocity form $v_\theta(\mathbf{r}^\tau_{t,i}, \mathbf{c}_i, \tau) = \frac{\mathbf{r}_\theta(\mathbf{r}^\tau_{t,i}, \mathbf{c}_i, \tau) - \mathbf{r}^\tau_{t,i}}{1-\tau}$. The training objective in the latent space is implemented as:

$$\mathcal{L}_{fm} = \frac{1}{N(T-1)} \sum_{t=1}^{T-1} \sum_{i=1}^{N} \|\mathbf{r}_\theta(\mathbf{r}^\tau_{t,i}, \mathbf{c}_i, \tau) - \mathbf{r}^1_{t,i}\|_2^2 \tag{7}$$

This residual formulation offers significant advantages over absolute coordinate prediction. First, it aligns well with the physics of conformational ensembles, where dynamics can be viewed as "pushing" a representation away from a reference state. Second, it improves generative robustness; since the flow learns perturbations around a valid physical structure ($\mathbf{x}_0$), the generated latent representations are less likely to drift outside the valid region of the VAE manifold, ensuring high fidelity in the decoded structures.

Sampling. Given the first frame, the flow model will generate the relative latent coordinate $\mathcal{R}^1 = [\mathbf{r}^1_1, \mathbf{r}^2_1, \cdots, \mathbf{r}^1_{T-1}]$ of the trajectory simultaneously via an ODE-solver, taking the Euler method as an example,

$$\hat{\mathcal{R}}^{\tau+d\tau} = \hat{\mathcal{R}}^\tau + v_\theta(\hat{\mathcal{R}}^\tau, \mathcal{C}, \tau)d\tau, \ \hat{\mathcal{R}}^0 = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \cdots, \boldsymbol{\epsilon}_{T-1}] \tag{8}$$

where $\mathcal{C} = \{\mathbf{c}_i\}_{i=0}^{T-1}$ denotes all the conditional information. After completing the iteration, we input the latent trajectory $\hat{\mathcal{Z}}^1_{1:T-1} = \hat{\mathcal{R}}^1 + \mathbf{x}_0$ into the decoder $\mathcal{D}_\phi$, obtaining the whole trajectory in the Cartesian space, i.e. $\hat{\mathcal{X}}_{1:T} = \mathcal{D}_\phi(\hat{\mathcal{Z}}^1_{1:T-1})$.

4.3. Model Architecture

We develop an Asymmetric Spatio-Temporal AutoEncoder to bridge the training-inference gap. Specifically, we omit temporal modules in the encoder (aligning with inference inputs) and utilize a hierarchical spatial-temporal module within the decoder to strongly enhance temporal relevance modeling. The core backbone of our architecture is built upon the Diffusion Transformer (DiT) (Peebles & Xie, 2023). The detailed schematic is provided in Appendix Figure 6.

Structural Conditioning and Domain Embedding. Inspired by (Feng et al., 2025), the initial conformation ($\mathbf{x}_0$) is processed by a Graph Transformer to extract SE(3)-invariant single ($\mathbf{s}_i$) and pair ($\mathbf{z}_{ij}$) representations to establish a structural prior (alg. 1). To integrate biological priors, we explicitly inject GPCR-specific domain knowledge: (1) Transmembrane (TM) Topology: The TM helix index (TM1–TM7) of each residue is embedded as a learnable vector. (2) Pharmacological Classification: The functional category of the ligand (e.g., agonist, antagonist) is similarly embedded. These domain features are concatenated with the token-level single representations—where a token corresponds to a receptor residue or a ligand atom—serving as the conditioning context for the encoder, decoder, and velocity network.

Spatio-Temporal AAT-VAE. To efficiently model high-resolution dynamics, we employ a hierarchical interaction mechanism inspired by AlphaFold3 (Abramson et al., 2024). (i) **Spatial Encoder:** As detailed in alg. 7, we utilize an `Atom Attention Encoder` to process the multi-frame coordinates. By treating the receptor and ligand as a contiguous sequence of atoms, the sequence-local all-atom attention naturally captures the joint distribution of the complex. This ensures that receptor and ligand structural information are deeply fused in the resulting atom-level latent variables $\mathbf{q}_l$ (subsequently projected to $(\vec{\mu}_l, \vec{\sigma}_l)$). (ii) **Temporal Decoder:** As detailed in alg. 9, the sampled atomic latents $\mathbf{q}_l \sim \mathcal{N}(\vec{\mu}_l, \vec{\sigma}_l)$ are first aggregated into tokens to capture global temporal dependencies

Table 1. Statistical metrics on MD ensemble benchmark of GPCRMD test set (sequence similarity $< 50\%$), where the median across all test targets is reported. The runtime is reported as **GPU second** required per sample, averaged on all test targets.

Metrics / Methods		100 ns		200 ns		500 ns	
		BioMD	GPCRLMD	BioMD	GPCRLMD	BioMD	GPCRLMD
Predicting flexibility	Pairwise RMSD $r \uparrow$	0.45	**0.67**	0.30	**0.62**	0.60	**0.64**
	Global RMSF $r \uparrow$	0.69	**0.76**	0.65	**0.74**	0.57	**0.75**
	Per-target RMSF $r \uparrow$	0.77	**0.81**	0.71	**0.79**	0.60	**0.79**
Distributional accuracy	Root mean $\mathcal{W}_2$-dist. $\downarrow$	2.81	**2.40**	2.87	**2.53**	**2.83**	2.92
	$\hookrightarrow$ Trans. contrib. $\downarrow$	2.43	**2.24**	2.49	**2.34**	**2.51**	2.68
	$\hookrightarrow$ Var. contrib. $\downarrow$	1.41	**0.87**	1.38	**0.97**	1.34	**1.10**
	MD PCA $\mathcal{W}_2$-dist. $\downarrow$	1.30	**1.16**	1.29	**1.16**	1.24	**1.10**
	Joint PCA $\mathcal{W}_2$-dist. $\downarrow$	2.17	**1.92**	2.22	**2.02**	**2.19**	2.41
	% PC-sim $> 0.5 \uparrow$	7.83	**14.78**	1.74	**3.48**	1.04	**1.74**
Ensemble observables	Weak contacts $J \uparrow$	0.11	**0.52**	0.12	**0.51**	0.16	**0.48**
	Transient contacts $J \uparrow$	0.26	**0.30**	0.24	**0.27**	0.20	**0.22**
	Exposed residue $J \uparrow$	0.30	**0.54**	0.29	**0.55**	0.29	**0.53**
	Exposed MI matrix $\rho \uparrow$	0.17	**0.27**	0.15	**0.27**	0.15	**0.24**
Runtime*	GPU sec. per sample	10.0	**2.0**	20.0	**4.0**	50.0	**10.0**

Note: *GPCRLMD achieves faster inference speeds because its latent velocity network has significantly fewer parameters than BioMD's Cartesian velocity network (1.6M vs. 78M). A traditional 500 ns MD simulation requires over 48 hours to complete.

via a Token Temporal Attention Decoder (attention maps are visualized in Appendix Figure 19). Subsequently, an Atom Temporal Attention Decoder reconstructs the fine-grained coordinates $\mathbf{x}_t$ by processing the atomic latents $\mathbf{q}_l$ conditioned on these temporally updated tokens. This hierarchical design ensures that global dynamics (token-level) effectively guide local atomic interactions (atom-level) during reconstruction.

Latent Velocity Network. The velocity network v_θ, responsible for the flow matching process, is distinct from the VAE. It accepts the noisy residual latent coordinates $\mathbf{r}^\tau$, and the conditional single and token representations $\mathbf{s}_i$ and $\mathbf{a}_i$ from the structural conditioner. It utilizes an atomic DiT-based architecture with cross-attention to the conditioned structural tokens to predict the flow velocity in the all-atom latent space. To better utilize the token information from the encoder and bridge the gap between training and inference, we creatively proposed a Stochastic Token Propagator, which actively transfers the structural context from the initial frame tokens ($\mathbf{a}_0$) into the temporal domain by mixing them with time-scaled diffusion noise, as shown in alg. 10,

5. Experimental Results

We primarily evaluate GPCRLMD on the GPCRMD dataset. To further demonstrate its generalization capabilities, we extend our framework to **MISATO** (Siebenmorgen et al., 2024), a comprehensive protein-ligand MD dataset. As detailed in Appendix 7, GPCRLMD achieves state-of-the-art performance on this benchmark.

Data Curation: We constructed our dataset from the GPCRMD database (Rodríguez-Espigares et al., 2020), leveraging the extensive GPCR dynamics sampled by Aranda-García et al. (2025), supplemented by additional test trajectories generated in-house. The raw dataset comprises 703 molecular dynamics trajectories (detailed in Table 6), with each replica spanning 500 ns (2,500 frames). To ensure rigorous generalization and prevent data leakage, we implemented a strict split based on receptor sequence identity ($< 50\%$), resulting in a partition of 531:57:115 trajectories for training, validation, and testing, respectively. Notably, over 95% of the test samples feature ligand scaffolds distinct from those in the training set. Crucially, this training partition constitutes a massive-scale dataset of over 1.32 million distinct conformational states. Unlike static protein datasets, this dense temporal sampling allows our model to learn the continuous dynamical manifold and rare allosteric transitions, rather than merely memorizing discrete crystal structures. **Training Protocol and Strided Sampling:** To efficiently capture long-timescale dynamics within memory constraints, we employ a strided window sampling strategy. During training, we sample sub-trajectories of 50 frames with a temporal stride of 10 frames. This effectively covers a 100ns physical timescale per input sample ($50 \times 10 \times 0.2$ns), enabling the model to learn temporal correlations across significant biological windows while maintaining computational tractability. **Evaluation Metrics:** For quantitative assessment, we adopt the rigorous benchmarking criteria proposed by Jing et al. (2024a), focusing on predicting flexibility, distributional consistency, and ensemble validity. Please refer to Appendix H.1 for details.

5.1. Ensemble Fidelity and Structural Validity

We first assess whether GPCRLMD captures the correct ensemble distribution of the receptor-ligand complex. Because GPCRLMD is, to the best of our knowledge, the first generative framework specifically tailored for GPCR-ligand dynamics, we benchmark it against leading methods from related general domains: BioMD (Feng et al., 2025) (a general protein-ligand dynamics model, retrained on our dataset) and AnewSampling (Wang et al., 2026) (a recent state-of-the-art protein-ligand ensemble generation method, evalu-

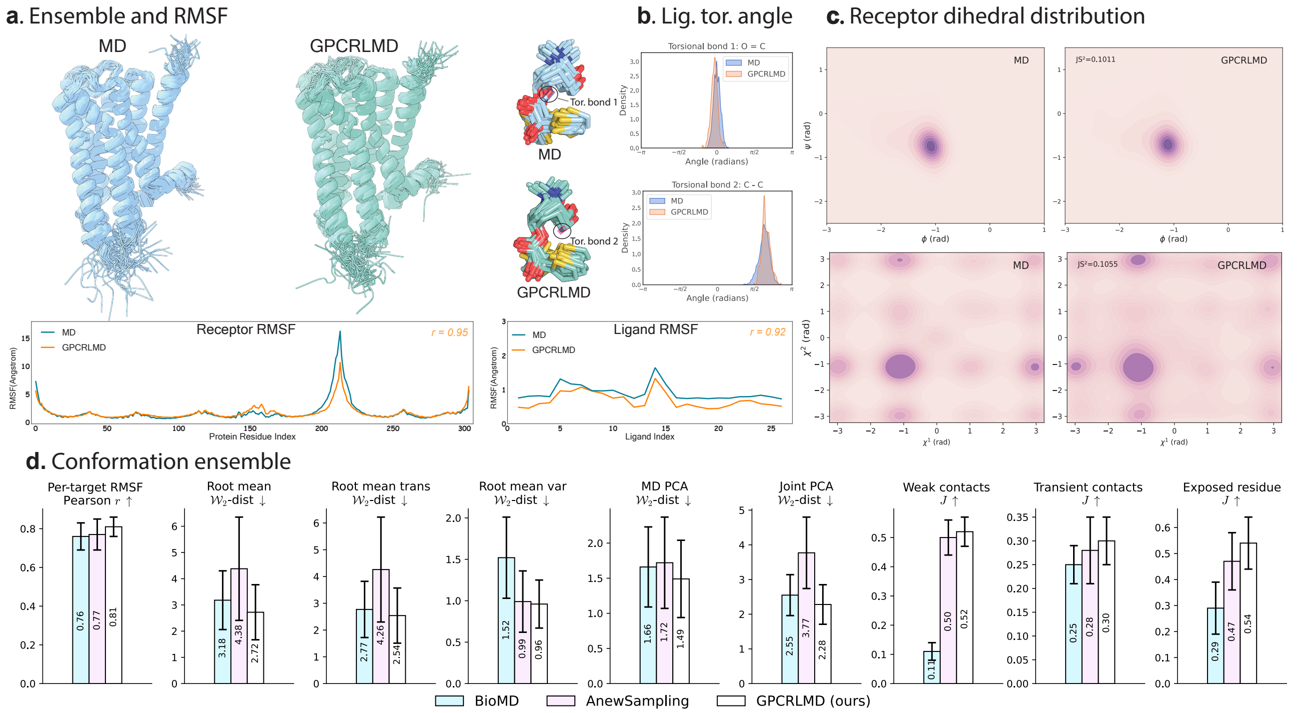

Figure 4. Evaluation of ensemble fidelity and structural validity. **a)** Receptor and ligand RMSF profiles compared to ground-truth MD (PDBID 5DSG). **b)** Ligand torsional angle distributions (PDBID 5DSG). **c)** Coupled backbone (ϕ-ψ) and side-chain (χ^1-χ^2) dihedral distributions (PDBID 5DSG). **d)** Conformation ensemble benchmarks against AnewSampling and BioMD. Metrics are averaged across the test set (250 conformations per sample).

ated via its official API). Comparisons with BioEmu (Lewis et al., 2025) (a SOTA protein ensemble generator, using official checkpoints) are provided in Appendix D.2.

Statistical Benchmarks and Flexibility. As summarized in Table 1, `GPCRLMD` achieves state-of-the-art performance across most metrics for 100ns–500ns simulations. It significantly surpasses the BioMD baseline in *Predicting Flexibility* and *Ensemble Observables*. We also evaluate our conformational ensemble against AnewSampling (Wang et al., 2026); the results in Figure 4d demonstrate that `GPCRLMD` consistently outperforms AnewSampling across all primary metrics. This quantitative success is visually mirrored in the residue-wise RMSF profiles for both the receptor and ligand (Figure 4a; Appendix Figures 13 and 14). The profiles exhibit high similarity to ground-truth MD: `GPCRLMD` correctly maintains structural rigidity in the transmembrane helices while reproducing high diversity in the flexible loops, and it accurately captures the fluctuation patterns of the bound ligand.

Structural Validity (Dihedrals and Torsions). To ensure physical plausibility, we analyze the coupled distribution of backbone (ϕ-ψ) and side-chain (χ^1-χ^2) dihedral angles (Figure 4c), as well as the ligand torsional angle distributions (Figure 4b). `GPCRLMD` maintains high distributional similarity with ground-truth (GT) MD across both protein and ligand degrees of freedom, indicating that the model respects local chemical geometry.

5.2. Complex Dynamics and Interaction Stability

We further evaluate the model's ability to simulate time-evolving biological mechanisms.

Capturing Critical State Transitions. The relative movement between Transmembrane Helix 2 (TM2) and TM6 is a hallmark of GPCR activation (Aranda-García et al., 2025). As shown in Figure 5a, `GPCRLMD` accurately reproduces the distance fluctuations and opening/closing trends of the TM helices. This confirms the model learns the underlying allosteric mechanism rather than just memorizing static coordinates.

Dynamic Contact Stability. Non-covalent interactions are the driving force of complex stability. We compute the dynamic contact map using a dual-cutoff metric to measure interaction frequency over time. As shown in Figure 5b. `GPCRLMD` faithfully reproduces the specific contact patterns observed in the GT MD, maintaining critical receptor-ligand interactions throughout the trajectory.

Ligand Stability and Energy Landscape. We examine the ligand's dynamic stability via RMSD profiles (Figure 5c). The generated ligands exhibit bounded RMSD fluctuations consistent with stable binding. Finally, to assess global dynamics, we project the trajectories onto the Free Energy Landscape (FEL) using TICA. As shown in Figure 5d, the `GPCRLMD` trajectory consistently maps to valid low-energy basins defined by the reference replica MD. This demonstrates that the model recovers essential metastable states

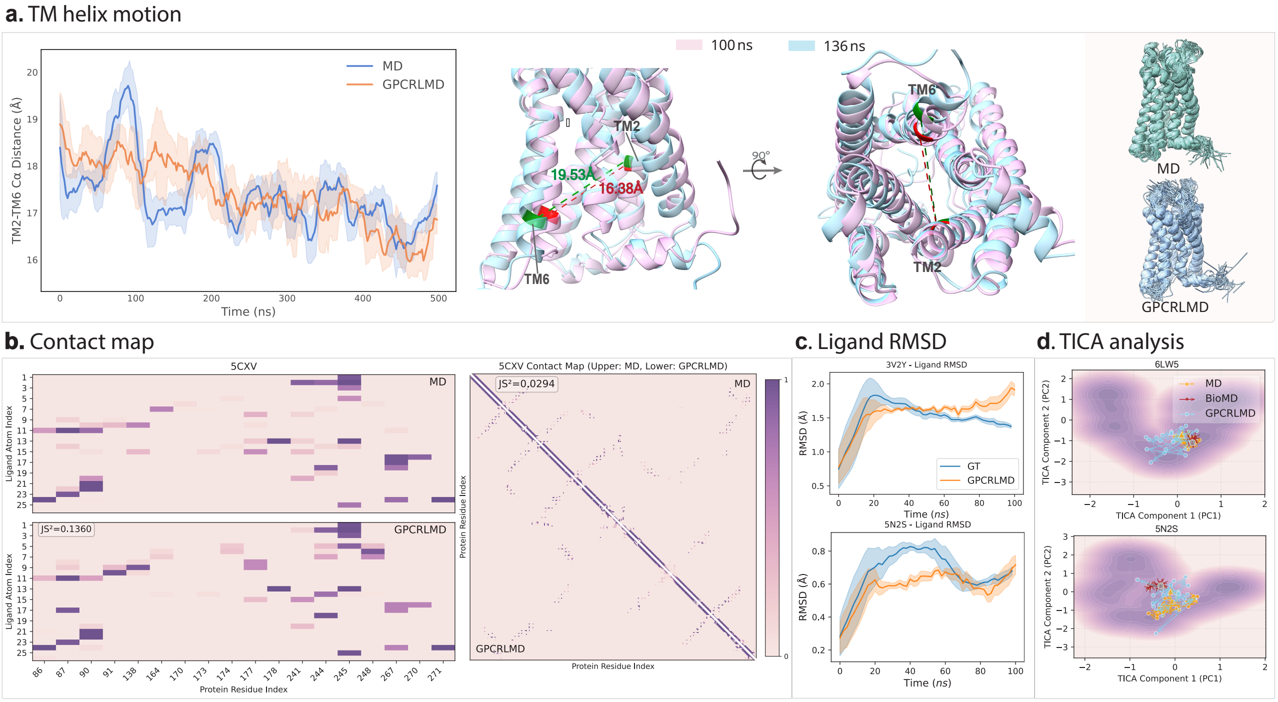

Figure 5. Evaluation of complex dynamics and interaction stability. **a)** Transmembrane (TM) helix motion. Time-evolution of the critical TM2-TM6 distance (PDBID: 6LW5) alongside a structural overlay of the conformational transition. More examples in Figure 8. **b)** Dynamic contact maps. Receptor-ligand and receptor-receptor interaction frequencies calculated using a dual-cutoff metric (details in Appendix H.4). **c)** Ligand stability. Time-dependent Ligand RMSD relative to the initial frame ($t = 0$). **d)** Free Energy Landscape (TICA). Projection of generated and reference trajectories onto the backbone energy surface, visualizing conformational space coverage. More examples and results are shown in Figure 9, 10 and Table 10.

and effectively samples diverse conformations, robustly maintaining physical validity.

Table 2. Ablation study comparing `GPCRLMD` with four variants.

Methods	Global RMSF r	Per-target RMSF r	Root mean $\mathcal{W}_2$-dist	Joint PCA $\mathcal{W}_2$-dist	Transient contacts J
`GPCRLMD` (w.all)	**0.76**	**0.81**	**2.40**	**1.92**	**0.30**
w/o Residual Vector	0.71	0.76	4.00	3.53	0.19
w/o AAT Prior (Std. Gauss)	0.75	0.75	3.57	3.18	0.13
w/o Latent Flow	0.71	0.78	2.41	1.97	0.29
w/o Time Dep. AAT Prior	0.72	0.78	2.42	1.95	0.28

5.3. Ablation Study

We conduct an ablation study to quantify the contribution of our core design components (Table 2). We evaluate five variants: (i) w/o Residual Vector: Predicting absolute latent coordinates rather than relative displacements anchored to the initial conformation; (ii) w/o AAT Prior: Replacing the target $\mathbf{x}_{t,i}$ with the Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ to demonstrate the importance of the proposed prior. (iii) w/o Latent Flow: Replacing the residual flow with standard VAE sampling ($\mathbf{z} \sim \mathcal{N}(\mu, \sigma\mathbf{I})$) to generate trajectories directly; (iv) w/o time dependent AAT Prior: Replacing the time-dependent target $\mathbf{x}_{t,i}$ with the static initial frame $\mathbf{x}_{0,i}$ in Eq. 3 to assess the benefit of dynamic structural prior. and (v) Cartesian Space Flow: Applying flow matching directly in Cartesian space (represented by the baseline BioMD). Removing any of these components leads to a significant performance drop, demonstrating that our architecture is tailored specifically for MD simulation rather than being a naive combination of modules. Furthermore, the performance advantage of our full model becomes even more pronounced in long-term trajectory generation, as detailed in Appendix Table 9.

6. Conclusion

We introduced `GPCRLMD`, a generative framework integrating an Atom-Anchored Thermal VAE and Residual Latent Flow for all-atom GPCR-ligand complex simulation. Crucially, `GPCRLMD` challenges the conventional reliance on dimensionality reduction in latent generative models. We demonstrate that performing sampling in a regularized all-atom latent space—rather than in compressed bottlenecks or raw Cartesian coordinates—yields superior structural fidelity and kinetic consistency in dynamics simulation. Consequently, the proposed method simulates dynamics with unprecedented efficiency, compressing hundreds of nanoseconds of trajectories into seconds while achieving SOTA fidelity in capturing ensemble distributions, structural flexibility, and critical interactions. However, because the generated trajectories are anchored to the initial state, the model's performance may be sensitive to the physical plausibility of this starting conformation. To mitigate this, future pipelines could employ a brief energy relaxation step to refine the local structure prior to inference. Looking ahead, extending this framework to general GPCR monomer systems presents a highly promising direction for broader applications.

Impact Statement

This paper presents work aimed at advancing Machine Learning for Science, with a specific focus on accelerating molecular simulation for drug discovery. The potential societal impact of this research is primarily positive, offering tools to expedite the development of GPCR-targeted therapeutics. We do not foresee specific negative consequences that require highlighting here.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630 (8016):493–500, 2024.

Aranda-García, D., Stepniewski, T. M., Torrens-Fontanals, M., García-Recio, A., Lopez-Balastegui, M., Medel-Lacruz, B., Morales-Pastor, A., Peralta-García, A., Dieguez-Eceolaza, M., Sotillo-Nuñez, D., et al. Large scale investigation of gpcr molecular dynamics data uncovers allosteric sites and lateral gateways. *Nature communications*, 16(1):2020, 2025.

Ciancetta, A., Sabbadin, D., Federico, S., Spalluto, G., and Moro, S. Advances in computational techniques to study gpcr–ligand recognition. *Trends in pharmacological sciences*, 36(12):878–890, 2015.

Conflitti, P., Lyman, E., Sansom, M. S., Hildebrand, P. W., Gutiérrez-de Terán, H., Carloni, P., Ansell, T. B., Yuan, S., Barth, P., Robinson, A. S., et al. Functional dynamics of g protein-coupled receptors reveal new routes for drug discovery. *Nature Reviews Drug Discovery*, 24(4):251–275, 2025.

Dao, T. M. and Rahman, T. Deep generative modeling of protein conformations: A comprehensive review. *BioChem*, 5(3):32, 2025.

Dong, Y.-w., Liao, M.-l., Meng, X.-l., and Somero, G. N. Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proceedings of the National Academy of Sciences*, 115(6):1274–1279, 2018.

Feng, B., Zhang, J., Zhang, X., Liu, Z., and Li, Y. BioMD: All-atom generative model for biomolecular dynamics simulation. In *The Fourteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=LQDeJk6NOr`.

Frenkel, D. and Smit, B. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

Geffner, T., Didi, K., Cao, Z., Reidenbach, D., Zhang, Z., Dallago, C., Kucukbenli, E., Kreis, K., and Vahdat, A. La-proteina: Atomistic protein generation via partially latent flow matching. *arXiv preprint arXiv:2507.09466*, 2025.

Han, J., Zhang, J., Nazarova, A. L., Bernhard, S. M., Krumm, B. E., Zhao, L., Lam, J. H., Rangari, V. A., Majumdar, S., Nichols, D. E., et al. Ligand and g-protein selectivity in the κ-opioid receptor. *Nature*, 617(7960): 417–425, 2023.

Han, J., Xu, M., Lou, A., Ye, H., and Ermon, S. Geometric trajectory diffusion models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=OYmms5Mv9H`.

Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B., and Gloriam, D. E. Trends in gpcr drug discovery: new agents, targets and indications. *Nature reviews Drug discovery*, 16(12):829–842, 2017.

Hertrich, J., Chambolle, A., and Delon, J. On the relation between rectified flows and optimal transport. *arXiv preprint arXiv:2505.19712*, 2025.

Hilger, D., Masureel, M., and Kobilka, B. K. Structure and dynamics of gpcr signaling complexes. *Nature structural & molecular biology*, 25(1):4–12, 2018.

Hinsen, K. Normal mode theory and harmonic potential approximations. In *Normal Mode Analysis*, pp. 25–40. Chapman and Hall/CRC, 2005.

Hoyt, J., Trautt, Z., and Upmanyu, M. Fluctuations in molecular dynamics simulations. *Mathematics and Computers in Simulation*, 80(7):1382–1392, 2010.

Jing, B., Berger, B., and Jaakkola, T. Alphafold meets flow matching for generating protein ensembles. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 22277–22303, 2024a.

Jing, B., Stärk, H., Jaakkola, T., and Berger, B. Generative modeling of molecular dynamics trajectories. *Advances in Neural Information Processing Systems*, 37:40534–40564, 2024b.

Kapla, J., Rodríguez-Espigares, I., Ballante, F., Selent, J., and Carlsson, J. Can molecular dynamics simulations improve the structural accuracy and virtual screening performance of gpcr models? *PLoS Computational Biology*, 17(5):e1008936, 2021.

Karplus, M. and McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9):646–652, 2002.

Katritch, V., Cherezov, V., and Stevens, R. C. Structure-function of the g protein–coupled receptor superfamily. *Annual review of pharmacology and toxicology*, 53:531–556, 2013.

Kingma, D. P., Welling, M., et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

Kong, X., Jia, Y., Huang, W., and Liu, Y. Full-atom peptide design with geometric latent diffusion. *Advances in Neural Information Processing Systems*, 37:74808–74839, 2024.

Kong, X., Zhang, Z., Zhang, Z., Jiao, R., Ma, J., Huang, W., Liu, K., and Liu, Y. Unimomo: Unified generative modeling of 3d molecules for de novo binder design. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=KUN7A7Okb6.

Latorraca, N. R., Venkatakrishnan, A., and Dror, R. O. Gpcr dynamics: structures in motion. *Chemical reviews*, 117 (1):139–155, 2017.

Lewis, S., Hempel, T., Jiménez-Luna, J., Gastegger, M., Xie, Y., Foong, A. Y., Satorras, V. G., Abdin, O., Veeling, B. S., Zaporozhets, I., et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, pp. eadv9817, 2025.

Li, M., Zhang, J., Feng, B., Zeng, W., Chen, D., Pan, Z., Li, Y., Liu, Z., and Yang, Y. I. Enhanced sampling, public dataset and generative model for drug-protein dissociation dynamics. *arXiv preprint arXiv:2504.18367*, 2025.

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

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. In *ICLR*, 2023.

Liu, S., Du, W., Ma, Z.-M., Guo, H., and Tang, J. A group symmetric stochastic differential equation model for molecule multi-modal pretraining. In *International Conference on Machine Learning*, pp. 21497–21526. PMLR, 2023a.

Liu, X., Gong, C., and Liu, Q. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023b.

López-Correa, J. M., König, C., and Vellido, A. Gpcr molecular dynamics forecasting using recurrent neural networks. *Scientific reports*, 13(1):20995, 2023.

Lorente, J. S., Sokolov, A. V., Ferguson, G., Schiöth, H. B., Hauser, A. S., and Gloriam, D. E. Gpcr drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery*, pp. 1–22, 2025.

Lu, J., Zhong, B., Zhang, Z., and Tang, J. Str2str: A score-based framework for zero-shot protein conformation sampling. In *The Twelfth International Conference on Learning Representations*, 2024.

Lu, J., Chen, X., Lu, S. Z., Lozano, A., Chenthamarakshan, V., Das, P., and Tang, J. Aligning protein conformation ensemble generation with physical feedback. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=Asr955jcuZ.

Mansoor, S., Baek, M., Park, H., Lee, G. R., and Baker, D. Protein ensemble generation through variational autoencoder latent space sampling. *Journal of Chemical Theory and Computation*, 20(7):2689–2695, 2024.

Newport, T. D., Sansom, M. S. P., and Stansfeld, P. J. The memprotmd database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic acids research*, 47(D1):D390–D397, 2019.

Peebles, W. and Xie, S. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Rodríguez-Espigares, I., Torrens-Fontanals, M., Tiemann, J. K., Aranda-García, D., Ramírez-Anguita, J. M., Stepniewski, T. M., Worp, N., Varela-Rial, A., Morales-Pastor, A., Medel-Lacruz, B., et al. Gpcrmd uncovers the dynamics of the 3d-gpcrome. *Nature Methods*, 17(8):777–787, 2020.

Roessner, R. A., Floquet, N., and Louet, M. Unveiling g-protein-coupled receptor conformational dynamics via metadynamics simulations and markov state models. *Journal of Chemical Information and Modeling*, 65(9): 4630–4642, 2025.

Samaddar, A., Sun, Y., Nilsson, V., and Madireddy, S. Efficient flow matching using latent variables. *arXiv preprint arXiv:2505.04486*, 2025.

Sengar, A., Hariri, A., Probst, D., Barth, P., and Vandergheynst, P. Generative modeling of full-atom protein conformations using latent diffusion on graph embeddings. *arXiv preprint arXiv:2506.17064*, 2025a.

Sengar, A., Zhang, J., Vandergheynst, P., and Barth, P. Beyond ensembles: Simulating all-atom protein dynamics in a learned latent space. In *The Fourteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=AwowReRWXI.

Shen, Y., Wang, L., Yuan, H., Wang, Y., Yang, B., and Gu, Q. Simultaneous modeling of protein conformation and dynamics via autoregression. *arXiv preprint arXiv:2505.17478*, 2025.

Siebenmorgen, T., Menezes, F., Benassou, S., Merdivan, E., Didi, K., Mourão, A. S. D., Kitel, R., Liò, P., Kesselheim, S., Piraud, M., et al. Misato: machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature Computational Science*, pp. 1–12, 2024.

Stark, H., Jing, B., Barzilay, R., and Jaakkola, T. Harmonic prior self-conditioned flow matching for multiligand docking and binding site design. In *NeurIPS 2023 AI for Science Workshop*, 2023. URL https://openreview.net/forum?id=3WF88uMjGz.

Torrens-Fontanals, M., Stepniewski, T. M., Aranda-García, D., Morales-Pastor, A., Medel-Lacruz, B., and Selent, J. How do molecular dynamics data complement static structural data of gpcrs. *International journal of molecular sciences*, 21(16):5933, 2020.

Verlet, L. Computer ”experiments” on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159(1):98–103, 1967.

Wang, Y., Xu, Y., Li, W., Yu, H., Tan, W., Li, S., Huang, Q., Chen, N., Wu, X., Wu, Q., et al. Learning the all-atom equilibrium distribution of biomolecular interactions at scale. *bioRxiv*, pp. 2026–03, 2026.

Wu, F. and Li, S. Z. Diffmd: A geometric diffusion model for molecular dynamics simulations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 5321–5329, 2023.

Xu, M., Powers, A. S., Dror, R. O., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.

y Schnitzler, M. M., Storch, U., Meibers, S., Nurwakagari, P., Breit, A., Essin, K., Gollasch, M., and Gudermann, T. Gq-coupled receptors as mechanosensors mediating myogenic vasoconstriction. *The EMBO journal*, 27(23):3092, 2008.

Zhang, M., Chen, T., Lu, X., Lan, X., Chen, Z., and Lu, S. G protein-coupled receptors (gpcrs): advances in structures, mechanisms and drug discovery. *Signal transduction and targeted therapy*, 9(1):88, 2024.

Zhou, Y., Lopez, G. E., and Giovambattista, N. The harmonic and gaussian approximations in the potential energy landscape formalism for quantum liquids. *Journal of chemical theory and computation*, 20(5):1847–1861, 2024.

Technical Appendices and Supplementary Material

Contents

A. Additional Related Work

GPCR Dynamic Simulations. G-protein coupled receptors (GPCRs) are pivotal in mediating numerous physiological processes (Zhang et al., 2024; Conflitti et al., 2025) by transducing chemical and physical signals, such as hormones, light, and pressure (y Schnitzler et al., 2008). GPCR signaling is driven by complex, ligand-induced conformational transitions (Zhang et al., 2024; Hilger et al., 2018; y Schnitzler et al., 2008). While databases like GPCRMD (Aranda-García et al., 2025) provide reference trajectories, they are limited to specific pre-simulated pairs. Furthermore, recent generative surrogates (López-Correa et al., 2023; Sengar et al., 2025a;b) lack generalizability, being restricted to single-system training that requires re-optimization for each new target. Additionally, their reliance on amino acid-based featurization inherently limits their ability to generalize across diverse GPCR-small molecule systems. Thus, developing a unified framework capable of efficient, zero-shot trajectory synthesis for unseen GPCR-ligand complexes remains a significant open challenge.

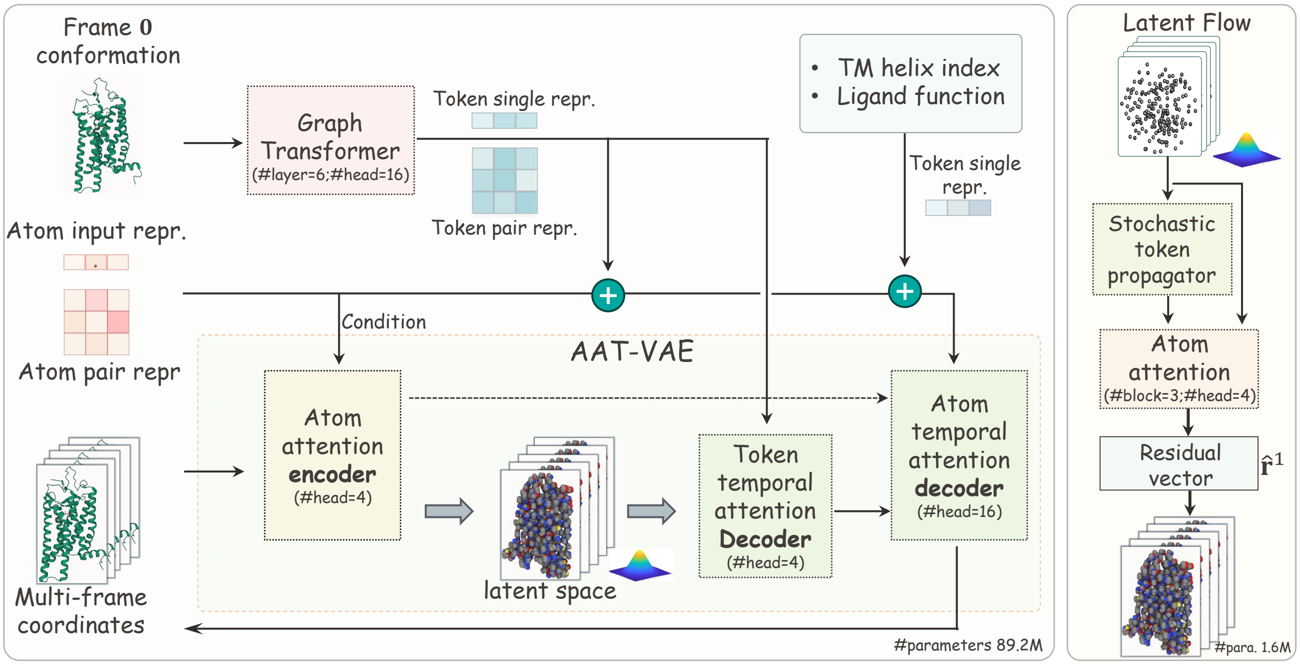

Figure 6. Schematic diagram of the model architecture. **Left**) The whole framework of `GPCRLMD`. The atomic structure information will be encoded in the latent space, where the latent coordinate can be decoded to the original Cartesian space. The frame 0 structure information, sequence/atom type, as well as GPCR Transmembrane helix indices (TM index) and ligand function information are the conditional information involved in both the encoder and decoders. The TM index and ligand function information will be concatenated to the token single representations. The dotted arrow in VAE between the encoder and the decoder indicates information without structure. **Right**). The residual latent flow will input the Gaussian noise and output the relative latent coordinates. To bridge the information gap for future time steps, we employ a Stochastic Token Propagator. This module actively transfers the structural context from the initial frame tokens ($\mathbf{a}_0$) into the temporal domain by mixing them with time-scaled diffusion noise. The details of the velocity network can be seen in algorithm 10.

B. Data Pipeline and Featurization

To specialize our model for GPCR-ligand interactions, we introduce domain-specific features alongside standard protein featurization.

GPCR Transmembrane (TM) Topology. To explicitly encode the receptor's topological structure, we assign a Transmembrane Helix Index to each residue.

- **Categorization:** Residues belonging to the transmembrane helices are labeled as `TM1` through `TM7`. Residues in loop regions or termini (non-TM) and ligand atoms are assigned distinct Non-TM and Ligand tokens, respectively.
- **Encoding:** These discrete categorical indices are projected into continuous feature vectors via a learnable lookup table (implemented as an `Embedding` layer), which is optimized jointly with the model.

Ligand Pharmacological Classification. We incorporate the functional role of the ligand as a global conditioning signal. To handle diverse annotation terminology, we map raw ligand descriptions into four primary pharmacological classes: Agonist, Antagonist, Inverse Agonist, and Unknown. The detailed mapping strategy is as follows:

- **Agonist:** Includes Agonist, Partial Agonist, Allosteric Agonist, and PAM (Positive Allosteric Modulator).
- **Antagonist:** Includes Antagonist, NAM (Negative Allosteric Modulator), and Allosteric Antagonist.
- **Inverse Agonist:** Includes Inverse Agonist.
- **Unknown:** Used for ligands with unclassified or ambiguous functions.

Similar to the structural features, these functional categories are encoded via learnable embeddings and concatenated to the global token representation.

General Protein Features. Beyond domain-specific attributes, we adopt the standard featurization pipeline from AlphaFold3 (Abramson et al., 2024). This includes amino acid identity, relative token indices, and chain identifiers (asym_id), ensuring the model captures fundamental protein geometry and sequence context.

C. More Experimental Details

C.1. Hyper-parameters.

The main hyper-parameters of `GPCRLMD` are shown in Table 3,4,5.

In particular, the parameter σ in the AAT Prior $p(\mathbf{z}_i)$ in Eq. 2 physically controls the width of the local harmonic energy well. In practice, its optimal value scales with the dataset's conformational diversity: smaller values are suited for highly stable unbiased MD, while larger values should be used for enhanced sampling or highly flexible systems. For our benchmark, we empirically found $\sigma = 16$ strikes an optimal balance.

Table 3. Training hyperparameters

Hyperparameters	**Values**
Batch Size	2
Frames Num	50
EMA decay	0.999
Temporal stride	10
Learning Rate	1×10^{-4}
Optimizer	Adam (weight decay = 0.)

C.2. Data Curation

The training, validation, and test datasets are summarized in Table 6. All training and validation samples, as well as a subset of 19 test samples, were collected from the public GPCRMD dataset contributed by the *GPCRMD community*. For the remaining 96 test samples, we adopted the same MD configurations (e.g., temperature and pressure) as GPCRMD (Rodríguez-Espigares et al., 2020) and generated a 500 ns (2,500 frames) trajectory for each system (protocol details refers to sec. C.5).

Table 4. Flow matching hyperparameters

Hyperparameters	**Values**
Sampling step in Flow	10
Velocity network blocks Num	3
Velocity network head Num	4

Table 5. AAT-VAE hyperparameters

Hyperparameters	**Values**
Dim of TM index embeddings	8
Dim of ligand function embeddings	8
Atom attn encoder blocks Num	3
Atom attn encoder head Num	4
Token temporal attn decoder blocks Num	3
Token temporal attn decoder head Num	4
Atom temporal attn decoder blocks Num	6
Atom temporal attn decoder head Num	16
σ in AAT Prior $p(\mathbf{z}_i)$ in Eq. 2	16
Latent dim in VAE	3

Generating these test samples required approximately 5,000 GPU hours in total, highlighting that the production of a single traditional 500 ns all-atom MD trajectory takes over 50 hours. Notably, these 96 newly generated test samples share $< 30\%$ sequence similarity and feature distinct ligand scaffolds compared to the training set.

Table 6. Data summary across Train, Validation, and Test sets. The test set contains the samples with ¡ 50% sequence similarity compared to the training set. Also, over 95% of the test samples feature ligand scaffolds distinct from those in the training set.

Statistical variable	**Train**	**Val**	**Test**
# Traj	531	57	115
# Simulation time for each traj	500 ns	500 ns	500 ns
# Frames Num	1,327,500	142,500	287,500
# PDB id	177	19	115
Receptor sequence length range	$276 \sim 501$	$259 \sim 327$	$253 \sim 372$
Averaged receptor Sequence length	317.9	304.4	293.87

C.3. Baselines.

Since we are the first model that is specifically designed for the GPCR-ligand complex dynamics, we use three SOTA general protein dynamics simulation models to demonstrate the SOTA performance of our model. We select **AnewSampling** (Wang et al., 2026), a recent state-of-the-art protein-ligand ensemble generation method, and **BioMD** (Feng et al., 2025), a model designed for protein-ligand complex trajectory generation. We also select **BioEmu** (Lewis et al., 2025), a well-known general method for protein-only conformation ensemble.

For BioEmu, we use the released checkpoint to sample the receptor ensemble on our test set, ignoring the ligand. For BioMD, we retrain the model from scratch on our training set and report the results on our test set. For AnewSampling, we evaluate via its official API.

We omit comparisons to recent GPCR-related generative models (Sengar et al., 2025a;b) as their architectures are typically restricted to single-system modeling. These methods lack the generalizability to handle unseen complexes without retraining.

C.4. Training Details.

We train `GPCRLMD` in two separate stages. In the first stage, we optimize only the VAE (encoder and decoder, etc, except for the latent flow velocity network) using the KL-divergence loss together with the auxiliary losses. In the second stage, we freeze the VAE parameters and optimize only the latent velocity network for latent flow matching.

The total number of parameters in `GPCRLMD` is 90.8M, of which the latent flow network accounts for 1.6M. We use the Adam optimizer to train the model on 16 H100 GPUs. Training the VAE takes approximately 6 days for 80,000 steps. For latent flow matching, training takes approximately 2 days for 50,000 steps.

For the baseline BioMD, the total number of parameters is 89.2M, and the velocity network required for sampling occupies 78M parameters, meaning that it requires more computation time to generate the final trajectory compared with `GPCRLMD`.

Long-Trajectory Generation Paradigm. As illustrated in Figure 7, once the initial batch of frames is generated from the starting frame $\mathbf{x}_0$, we can leverage `GPCRLMD` to extend the trajectory. In this process, the final frame of the current iteration serves as the initial frame for the subsequent iteration, allowing for the continuous generation of long-trajectory simulations.

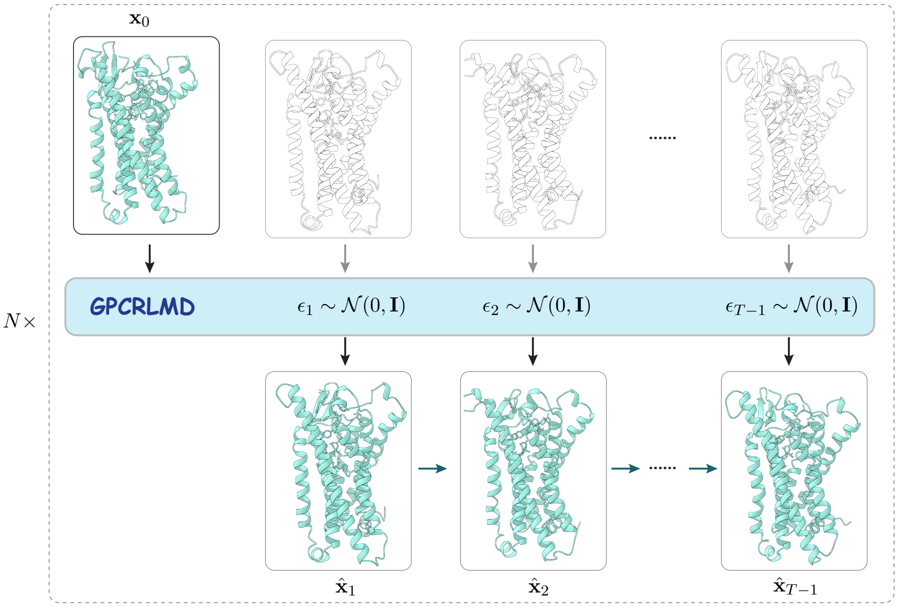

Figure 7. Schematic diagram of the trajectory generative process. The proposed model, `GPCRLMD`, utilizes an initial frame to generate the subsequent $T-1$ frames in parallel. This process is repeated N times to obtain $N \times (T-1)$ samples, where the final frame of the previous iteration serves as the initial frame for the next.

C.5. Simulation Protocol

The all-atom molecular dynamics (MD) simulations for the 96 GPCR-ligand test samples are conducted using a dedicated Python-based simulation pipeline. The systems are parameterized with the **CHARMM36** force field (July 2020 version). Following initial solvation and neutralization, the relaxation of each system began with an energy minimization of 5,000 steps.

Following minimization, the systems undergo a 40 ns equilibration phase in the NPT (constant pressure) ensemble at a physiological temperature of 310 K and a pressure of 1 atm. The temperature is controlled using a **Langevin thermostat** with a damping coefficient of 1 ps^{-1}, while the pressure is maintained using a **Monte Carlo barostat**. The equilibration utilizes a 2 fs time step, with trajectory frames save every 0.05 ns (25,000 integration steps).

During this phase, flat-bottom harmonic position restraints with a width of 2 Å were applied along the three axes to the protein, co-solved ligand molecules, HOMOLWAT, and co-solved waters. The restraint strength followed a specific schedule:

- **0–20 ns:** Constant at 1 kcal/mol/Å^2.
- **20–30 ns:** Linearly tapered from 1 to 0 kcal/mol/Å^2.
- **30–40 ns:** Completely removed (0 kcal/mol/Å^2) to allow for final system relaxation.

Upon successful equilibration, production simulations were conducted in the NVT (constant volume) ensemble for 500 ns per sample. For the production runs, the integration time step was increased to 4 fs, and the Langevin thermostat damping coefficient was reduced to 0.1 ps^{-1}. Production trajectories were saved every 0.2 ns (50,000 integration steps).

Finally, the resulting trajectories were wrapped and aligned around the receptor using the `MDAnalysis` module (version 2.1.0) to ensure data consistency for downstream structural analyses.

D. Additional Experimental Results

D.1. Experimental Results on MISATO

To definitively prove that our architecture `GPCRLMD` is inherently generalizable to universal protein-ligand systems (by simply omitting these domain-specific token embeddings), we train `GPCRLMD` from scratch on MISATO (Siebenmorgen et al., 2024), a massive universal machine learning dataset containing 16,972 simulated protein-ligand complexes (10 ns each).

Setting. Following the BioMD protocol, we trained our model from scratch on the MISATO (Siebenmorgen et al., 2024) training set and evaluated it on 1,031 test targets (protein length ≤ 800, ligand size ≤ 100 heavy atoms).

Baselines. (1) BioMD: A concurrent SOTA framework designed for general protein-ligand dynamics. (2) DenoisingLD, GNNMD, NeuralMD-ODE, NeuralMD-SDE, VerletMD: SOTA models predicting binding dynamics, proposed and implemented by Liu et al. (2023a).

Results. As shown in Table 7, `GPCRLMD` significantly outperforms baseline methods on 1,031 test samples across main RMSF metrics. Remarkably, despite being primarily designed for GPCR systems, it surpasses even the concurrent SOTA general protein-ligand MD model, BioMD, in terms of ligand RMSF metrics. These strong empirical results validate that our core physics-informed architecture inherently generalizes well beyond the GPCR domain.

Table 7. Results on the MISATO test set (1031 samples). Mean ± std values are reported. Values closer to those of the reference (in parentheses) indicate better performance.

Method	RMSF Correlation ↑		RMSF Value (closer to Ref.)	
	Ligand	Protein	Ligand (3.375)	Protein (3.301)*
DenoisingLD (Liu et al., 2023a)	-0.0290	—	$> 10^{12}$	—
GNNMD (Liu et al., 2023a)	-0.0103	—	0.2165	—
NeuralMD-ODE (Liu et al., 2023a)	0.3405	—	0.3220	—
NeuralMD-SDE (Liu et al., 2023a)	0.3405	—	0.3220	—
VerletMD (Liu et al., 2023a)	0.3356	—	0.3226	—
BioMD (Feng et al., 2025)	0.4861	0.5945	0.5369	0.5177
GPCRLMD (ours)	**0.6268 ± 0.1381**	0.5831 ± 0.4228	**3.2194 ± 2.2493**	**1.2990 ± 0.35**

Note: * RMSF values for reference trajectories are given in parentheses. Baseline results are copied from the BioMD (Feng et al., 2025) paper.

D.2. Comparison with BioEmu

To rigorously assess the quality of the generated receptor ensembles, we benchmark `GPCRLMD` against BioEmu, a state-of-the-art protein-only generative model. As shown in Table 8, `GPCRLMD` demonstrates superior performance across most kinetic and computational metrics.

Capturing Conformational Flexibility. `GPCRLMD` exhibits a consistently stronger correlation with ground-truth flexibility profiles. In the *Predicting Flexibility* task, `GPCRLMD` achieves a high Pairwise RMSD correlation of $\mathbf{0.68}$, outperforming BioEmu (0.58). Similarly, our method surpasses BioEmu in reproducing both global ($\mathbf{0.77}$ vs. 0.70) and per-target ($\mathbf{0.82}$ vs. 0.76) RMSF profiles. This indicates that `GPCRLMD` more accurately models the magnitude and localization of atomic fluctuations during GPCR activation.

Distributional Fidelity and Observables. In terms of distributional accuracy, GPCRLMD achieves the lowest Root Mean $\mathcal{W}_2$ distance ($\mathcal{W}_2 = \mathbf{2.36}$) compared to BioEmu (3.22), suggesting a more precise recovery of the equilibrium Boltzmann distribution. This advantage extends to specific biophysical observables; our model excels in recovering weak contacts ($J = \mathbf{0.52}$) and exposed residue patterns ($J = \mathbf{0.54}$), outperforming BioEmu (0.45 and 0.47, respectively). While BioEmu shows a slight edge in capturing high-similarity principal components (% PC-sim > 0.5), GPCRLMD overall better preserves the subtle internal packing and surface geometry required for realistic protein dynamics.

Computational Efficiency. Beyond accuracy, GPCRLMD offers a transformative advantage in efficiency. While BioEmu requires substantial computational overhead ($\approx$ 1800 GPU seconds per sample) due to its denoising sampling process, GPCRLMD generates equivalent ensembles in merely $\mathbf{10.0}$ seconds. This represents a $\mathbf{180}\times$ speedup, making GPCRLMD uniquely scalable for high-throughput virtual screening campaigns where rapid ensemble generation is critical.

Importance of Ligand-Aware Modeling. Ultimately, this comparison rigorously validates the necessity of explicit ligand modeling. The substantial performance gap between GPCRLMD and BioEmu across all receptor-specific metrics provides a critical biophysical insight: general protein models that ignore the ligand fail to capture the specific conformational rearrangements, such as induced fit, driven by binding events. By explicitly modeling the entire complex, GPCRLMD successfully recovers the functional ligand-bound receptor ensemble—a task where "ligand-blind" baselines fundamentally struggle.

Table 8. Statistical metrics on the MD ensemble benchmark using the GPCRMD test set (sequence similarity $< 50\%$). We compare AnewSampling, BioEmu, BioMD, and GPCRLMD on receptor conformational ensemble tasks, reporting the median metric across all test targets. Each method generates 250 conformations per sample. To ensure a fair comparison with BioEmu (a protein-only model), metrics for BioMD, AnewSampling and GPCRLMD are calculated solely on the receptor, excluding ligand coordinates. Runtime is reported as GPU seconds required per sample, averaged across all targets. The best results are highlighted in **bold**.

Metrics / Methods		250 frames			
		BioMD	BioEmu	AnewSampling	GPCRLMD
Predicting flexibility	Pairwise RMSD $r \uparrow$	0.49	0.58	**0.68**	**0.68**
	Global RMSF $r \uparrow$	0.72	0.70	0.76	**0.77**
	Per-target RMSF $r \uparrow$	0.78	0.76	0.78	**0.82**
Distributional accuracy	Root mean $\mathcal{W}_2$-dist. $\downarrow$	2.71	3.22	3.71	**2.36**
	$\hookrightarrow$ Trans. contrib. $\downarrow$	2.40	3.04	3.61	**2.18**
	$\hookrightarrow$ Var. contrib. $\downarrow$	1.39	1.00	0.86	**0.86**
	MD PCA $\mathcal{W}_2$-dist. $\downarrow$	1.29	1.18	1.35	**1.16**
	Joint PCA $\mathcal{W}_2$-dist. $\downarrow$	2.15	2.65	3.23	**1.91**
	% PC-sim $> 0.5 \uparrow$	8.70	**17.39**	13.16	15.65
Ensemble observables	Weak contacts $J \uparrow$	0.11	0.45	0.49	**0.52**
	Transient contacts $J \uparrow$	0.26	0.30	0.29	**0.30**
	Exposed residue $J \uparrow$	0.29	0.47	0.49	**0.54**
	Exposed MI matrix $\rho \uparrow$	0.17	0.27	**0.31**	0.27
Runtime	GPU sec. per sample	50	1800	-	**10.0**

D.3. RMSF Comparison.

Figure 13 and Figure 14 demonstrate that our method achieves high RMSF similarity with the reference MD, particularly for the receptor. Notably, GPCRLMD accurately captures the distinct dynamical profiles of the GPCR structure: it maintains the requisite structural stability (low flexibility) of the seven transmembrane helices while successfully reproducing the high conformational diversity of the loop regions. Qualitative visualizations of randomly selected cases from the test set are provided in Figure 15.

D.4. Ablation Study

We conduct an additional ablation study to verify the effectiveness of our designed components for long trajectory generation (500 ns), as shown in Table 9.

Table 9. Ablation study on the test set (115 samples) at **500 ns**, with protein sequence similarity $< 50\%$.

Metrics / Methods		500 ns			
		`GPCRLMD`	w/o Time Dep. AAT	w/o AAT Prior (Std. Gauss)	w/o Residual Vector
Predicting flexibility	Pairwise RMSD $r \uparrow$	0.64	0.60	**0.77**	0.39
	Global RMSF $r \uparrow$	**0.75**	0.71	0.71	0.61
	Per-target RMSF $r \uparrow$	**0.79**	0.76	0.70	0.71
Distributional accuracy	Root mean $\mathcal{W}_2$-dist. $\downarrow$	**2.92**	3.14	6.18	6.40
	$\hookrightarrow$ Trans. contrib. $\downarrow$	**2.68**	2.82	5.95	6.19
	$\hookrightarrow$ Var. contrib. $\downarrow$	**1.10**	1.33	1.69	1.55
	MD PCA $\mathcal{W}_2$-dist. $\downarrow$	**1.10**	1.10	1.40	1.36
	Joint PCA $\mathcal{W}_2$-dist. $\downarrow$	**2.41**	2.52	5.87	6.10
	% PC-sim $> 0.5 \uparrow$	**1.74**	1.74	0.00	0.87
Ensemble observables	Weak contacts $J \uparrow$	**0.48**	0.48	0.36	0.08
	Transient contacts $J \uparrow$	**0.22**	0.19	0.03	0.07
	Exposed residue $J \uparrow$	**0.53**	0.51	0.37	0.14
	Exposed MI matrix $\rho \uparrow$	**0.24**	0.15	0.13	0.04

D.5. TICA Analysis

We assess kinetic mode recovery using Time-lagged Independent Component Analysis (TICA). The ground-truth free energy landscape is established using the combined trajectories of MD Replicas 2 and 3. Because certain test targets lack multi-replica sampling, we specifically curated a subset of 17 systems (PDB: 4EA3, 4XNV, 5C1M, 3V2Y, 5CXV, 4XNW, 4ZUD, 3ODU, 3PBL, 3RZE, 5DSG, 4Z36, 4Z34, 4DJH, 3VW7, 6LW5, 5N2S) that possess three independent replicas; these datasets are also publicly available via the GPCRMD database. The TICA model is fitted on these high-resolution trajectories ($1.0\mu s$, 5000 frames, stride $\Delta t = 0.2$ns, lag time $\tau = 4$ns) to identify the slowest relaxation modes. This pre-calculated landscape serves as the background density in our visualizations.

For evaluation, we project three distinct sets of trajectories onto this fixed TICA space: (1) the unseen Replica 1 (serving as the MD baseline), (2) BioMD, and (3) `GPCRLMD`. All projected trajectories are normalized to a 500 ns duration (250 frames, 2 ns stride) for fair comparison (Figure 9,10). The quantitative fidelity is measured using the Jensen-Shannon Divergence (JSD) between the TICA distribution of each generated trajectory and the ground-truth reference distribution (Replicas 2+3).

Results. As shown in Table 10, `GPCRLMD` consistently achieves the lowest Jensen-Shannon Divergence (JSD) scores compared to BioMD, outperforming even the held-out MD baseline (Replica 1) in three out of four metrics. This indicates that `GPCRLMD` does not merely memorize a single trajectory path but effectively samples the broader, biologically plausible conformational ensemble defined by the extensive ground-truth simulations. By covering a more comprehensive region of the valid energy landscape, `GPCRLMD` demonstrates superior generative diversity and thermodynamic fidelity.

These quantitative findings are corroborated by the TICA projections in Figures 9 and 10. Visual inspection reveals that `GPCRLMD` achieves high conformational state recovery across most test cases. Crucially, the model demonstrates the capacity to sample more diverse conformations than the individual reference trajectory segments. By effectively populating broader low-energy basins, `GPCRLMD` captures a more comprehensive view of the equilibrium thermodynamic distribution.

Table 10. Jensen-Shannon divergence (JSD) on TICA projections (500ns simulation). The table reports the divergence between the generated ensembles and the reference distribution on the first TICA component (TICA-0) and the joint first two components (TICA-0,1). Lower values indicate better performance. We report the mean and variance for all samples in the test set.

Method	**Backbone + Sidechain**		**Backbone Only**	
	TICA-0 (JSD $\downarrow$)	**TICA-0,1 (JSD $\downarrow$)**	**TICA-0 (JSD $\downarrow$)**	**TICA-0,1 (JSD $\downarrow$)**
MD (500ns)	0.50 ± 0.06	0.66 ± 0.02	0.44 ± 0.10	**0.62 ± 0.06**
BioMD	0.56 ± 0.08	**0.64 ± 0.04**	0.51 ± 0.10	0.63 ± 0.04
GPCRLMD (Ours)	**0.44 ± 0.05**	**0.64 ± 0.03**	**0.41 ± 0.10**	0.63 ± 0.04

D.6. Salt Bridge Analysis

Salt bridges are critical electrostatic interactions that anchor the ligand within the binding pocket. We specifically measure the distance between the ligand's cationic nitrogen and the conserved aspartate residue ($D^{3.32}$) on Transmembrane Helix 3 (Katritch et al., 2013; Han et al., 2023). Physiologically, the side chain of Asp$^{3.32}$ contains a carboxylate group that is deprotonated (negatively charged), while the ligand's amine group is typically protonated (positively charged). This electrostatic attraction creates a stable salt bridge essential for receptor activation. The results in Figure 11 demonstrate that conditional on the presence of this salt bridge in the initial conformation ($\mathbf{x}_0$), `GPCRLMD` effectively preserves the interaction within the canonical range (< 4 Å) throughout the trajectory. Conversely, if the interaction is absent initially (as seen in the 5CXV case), the model faithfully propagates the non-interacting state.

D.7. Verifying Isometric Topology Preservation

For each trajectory frame in the test set, we computed the pairwise Euclidean distance matrix for all atoms in both Cartesian space and latent spaces. We observe a strong positive Pearson correlation of 0.80. The slight variance from 1.0 is physically expected, reflecting the localized thermal fluctuations intentionally introduced by the AAT Prior.

D.8. Case Study for Long Trajectory Generation

Our framework naturally extends simulations via continuous iterative sampling. As shown in the figure at `https://anonymous.4open.science/r/GPCRLMD/assets/6Z10_1us.png`, extending the trajectory for 6Z10 captures a large-scale transition (TM2-TM6 distance shifting from 17.5 Åto 13.8 Å).

E. Model Architecture

E.1. Energy-Guided AAT-VAE for Long Trajectory Prediction

We extend the AAT-VAE to an energy-guided formulation to enhance the physical fidelity of long-term trajectory predictions.

Energy-Augmented Posterior. Standard generative models capture global conformational variability but often violate low-entropy geometric constraints, such as chemically valid side-chain bond lengths and angles. These constraints are naturally expressed as coordinate-space energies rather than through a Gaussian decoder. To address this, we incorporate a physics-based potential $E_{\text{sc}}(\mathbf{x})$ directly into the inference process. We define a corrected conditional distribution (energy-augmented posterior):

$$\tilde{p}(\mathbf{x} \mid \mathbf{z}) \propto \mathcal{N}(\mathbf{x}; g(\mathbf{z}), \sigma^2 I) \exp(-\beta E_{\text{sc}}(\mathbf{x})), \tag{9}$$

where $g(\mathbf{z})$ is the raw decoder output. Finding the most likely structure $\mathbf{x}^*$ under this distribution corresponds to a Maximum A Posteriori (MAP) estimate:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left(\frac{1}{2\sigma^2} |\mathbf{x} - g(\mathbf{z})|^2 + \beta E_{\text{sc}}(\mathbf{x}) \right). \tag{10}$$

This formulation acts as a proximal operator, $\text{prox}_{\mu E}(g(\mathbf{z}))$ with $\mu = \sigma^2/\beta$. It effectively applies a likelihood correction that enforces physical plausibility while preserving the learned manifold of the generative model.

Test-Time Optimization. In our specific implementation, we employ a local structural potential $E_{\text{sc}}(\mathbf{x})$ that regularizes side-chain bond stretching and bond angles, using the initial frame $\mathbf{x}_0$ as a reference topology to define the equilibrium geometry. We integrate this minimization directly into the sampling process. After generating the trajectory via flow matching:

$$\hat{\mathcal{R}}^0 = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \cdots, \boldsymbol{\epsilon}_{T-1}] \tag{11}$$

$$\hat{\mathcal{R}}^{\tau+1} = \hat{\mathcal{R}}^{\tau} + v_\theta(\hat{\mathcal{R}}^{\tau}, \mathcal{C}, \tau) d\tau, \tag{12}$$

we apply K steps of gradient-based energy minimization to the final prediction $\hat{\mathcal{R}}^1$:

$$\hat{\mathcal{R}}^1_{k+1} = \hat{\mathcal{R}}^1_k - \alpha \nabla E_{\text{sc}}(\hat{\mathcal{R}}^1_k), k = 0, 2, ..., K-1, \tag{13}$$

where α is the step size. This ensures the final trajectory adheres to local chemical constraints without requiring retraining of the flow network. In practice, we define the side-chain geometric energy as:

$$E_{\text{sc}}(\hat{\mathbf{x}}_t) := \frac{2}{|\mathcal{E}_{\text{bond}}|} \sum_{(i,j)\in\mathcal{E}_{\text{bond}}} (\hat{d}_{ij}^t - d_{ij}^0)^2 + \frac{1}{|\mathcal{E}_{\text{angle}}|} \sum_{(i,j,k)\in\mathcal{E}_{\text{angle}}} \left|\cos(\hat{\theta}_{ijk}^t) - \cos(\theta_{ijk}^0)\right|,$$

where d_{ij}^0 and $\hat{d}_{ij}^t$ represent the bond lengths of side-chain bond (i, j) in the initial frame ($t = 0$) and the predicted frame t, respectively. Similarly, θ_{ijk}^0 and $\hat{\theta}_{ijk}^t$ denote the bond angles defined by side-chain atoms (i, j, k) in frame 0 and frame t, respectively. The cosine terms are calculated following Eq. 21.

E.2. Weighted Rigid Aligned MSE

To resolve the global rotational and translational invariance of the generated structures while emphasizing the structural accuracy of the ligand, we follow AF3 (Abramson et al., 2024) to employ a weighted Mean Squared Error (MSE) loss with rigid alignment. Let $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^N$ and $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}_i\}_{i=1}^N$ denote the ground-truth and predicted coordinates for a complex with N atoms. We assign an importance weight w_i to each atom to prioritize the ligand geometry:

$$w_i = \begin{cases} 1 + \lambda_{\text{ligand}} & \text{if atom } i \in \text{Ligand} \\ 1 & \text{if atom } i \in \text{Protein} \end{cases} \tag{14}$$

where λ_{ligand} is a hyperparameter balancing the contribution of ligand atoms (we use $\lambda_{\text{ligand}} = 10$ in training `GPCRLMD`). Before computing the error, we compute the optimal rigid transformation $(R^*, \mathbf{t}^*)$ that aligns the ground truth to the prediction by minimizing the weighted RMSD:

$$(R^*, \mathbf{t}^*) = \underset{R\in SO(3),\mathbf{t}\in\mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^N w_i \left\|\hat{\mathbf{x}}_i - (R\mathbf{x}_i + \mathbf{t})\right\|^2, \tag{15}$$

$$\mathbf{x}_i^{\text{GT-aligned}} = R^*\mathbf{x}_i + \mathbf{t}^*, \tag{16}$$

where $\mathbf{x}_i^{\text{GT-aligned}}$ represents the aligned ground-truth coordinates. The weighted rigid alignment loss is then defined as the mean squared discrepancy between the predicted and aligned coordinates:

$$\mathcal{L}_{\text{align-mse}} = \frac{1}{\sum w_i} \sum_{i=1}^N w_i \left\|\hat{\mathbf{x}}_i - \mathbf{x}_i^{\text{GT-aligned}}\right\|^2. \tag{17}$$

This formulation ensures that the model learns the precise internal geometry of the ligand and its relative pose within the binding pocket, decoupled from the global frame of reference.

F. Auxiliary Losses

By enforcing local geometric constraints within the decoder, the model learns a compact latent representation that filters out redundant local fluctuations and retains the essential conformational features of the trajectory.

Ligand geometric center loss. To stabilize the global placement of the ligand and prevent spurious rigid translations, we align the predicted and reference geometric centers of ligand atoms. Let $\mathbf{x}_t^\ell = \{\mathbf{x}_t^{\ell,i}\}_{i=1}^{N_\ell}$ and $\hat{\mathbf{x}}_t^\ell = \{\hat{\mathbf{x}}_t^{\ell,i}\}_{i=1}^{N_\ell}$ denote ground-truth and predicted ligand coordinates at step t. The geometric center is

$$C(\mathbf{x}_t^\ell) = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \mathbf{x}_t^{\ell,i}, \qquad C(\hat{\mathbf{x}}_t^{\ell,i}) = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \hat{\mathbf{x}}_t^{\ell,i},$$

and the loss is the mean-squared discrepancy

$$\mathcal{L}_{\text{center}} = \left\|C(\hat{\mathbf{x}}_t^\ell) - C(\mathbf{x}_t^\ell)\right\|_2^2.$$

This term softly anchors the ligand's global position while remaining agnostic to its internal geometry.

Collision loss. To penalize steric clashes, we define a collision loss between protein–ligand atoms and within ligand atoms. Let $\mathbf{x}_t^\ell$ and $\mathbf{x}_t^\mathcal{P}$ denote ligand and protein atom coordinates at step t, and $\hat{\mathbf{x}}_t^\ell$, $\hat{\mathbf{x}}_t^\mathcal{P}$ their predictions. We compute predicted distances

$$d_{ij}^{PL} = \|\hat{\mathbf{x}}_t^{\mathcal{P},i} - \hat{\mathbf{x}}_t^{\ell,j}\|_2, \quad d_{ij}^{L} = \|\hat{\mathbf{x}}_t^{\ell,i} - \hat{\mathbf{x}}_t^{\ell,j}\|_2,$$

and corresponding ground-truth minimal distances

$$d_{ij}^{PL,gt} = \min_t \|\mathbf{x}_t^{\mathcal{P},i} - \mathbf{x}_t^{\ell,j}\|_2, \quad d_{ij}^{LL,gt} = \min_t \|\mathbf{x}_t^{\ell,i} - \mathbf{x}_t^{\ell,j}\|_2.$$

Protein–ligand and ligand–ligand thresholds are set as

$$\zeta_{ij}^{PL} = \min\Big(0.9\, d_{ij}^{PL,gt},\ \zeta_{pl}\Big) \qquad \zeta_{ij}^{LL} = \min\Big(0.9\, d_{ij}^{LL,gt},\ \zeta_{ll}\Big),$$

where $\zeta_{pl} = 3.0\,\text{Å}$ and $\zeta_{ll} = 2.0\,\text{Å}$.

The collision loss is then defined as

$$\mathcal{L}_{\text{collision}} = \sum_{i,j} \mathbf{1}\big(d_{ij}^{PL} < \zeta_{ij}^{PL}\big)\ (\zeta_{ij}^{PL} - d_{ij}^{PL})^2 + \sum_{i \neq j} \mathbf{1}\big(d_{ij}^{LL} < \zeta_{ij}^{LL}\big)\ (1 - b_{ij})\,(\zeta_{ij}^{LL} - d_{ij}^{LL})^2,$$

where $\mathbf{1}(\cdot)$ represents the indicator function and b_{ij} is the ligand bond mask to exclude bonded pairs.

Ligand bond loss. To preserve ligand bond lengths, we penalize deviations between predicted and ground-truth bonded atom distances. Let $\mathcal{B}$ denote the set of bonded atom pairs according to the ligand bond mask. For each bond $(i,j) \in \mathcal{B}$, we compute the predicted and ground-truth distances

$$d_{ij}^{\ell} = \|\hat{\mathbf{x}}_t^{\ell,i} - \hat{\mathbf{x}}_t^{\ell,j}\|_2, \quad d_{ij}^{\ell,gt} = \|\mathbf{x}_t^{\ell,i} - \mathbf{x}_t^{\ell,j}\|_2.$$

The bond loss is then defined as the mean squared deviation:

$$\mathcal{L}_{\text{bond}} = \frac{1}{|\mathcal{B}|} \sum_{(i,j)\in\mathcal{B}} \Big(d_{ij}^{\ell} - d_{ij}^{\ell,gt}\Big)^2.$$

Side-chain bond length loss. To ensure physical plausibility and robustness against outliers in the predicted structure, we constrain bond lengths using a robust error function. Given the ground-truth coordinates $\mathbf{x}$ and predicted coordinates $\hat{\mathbf{x}}$, for each bond in the side-chain bond index set $\mathcal{E}_{\text{bond}}$, we compute the Euclidean bond lengths $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ and $\hat{d}_{ij} = \|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|_2$. The loss is minimized using the Huber loss with a threshold $\delta = 0.05$:

$$\mathcal{L}_{\text{sc-bond}} = \frac{1}{|\mathcal{E}_{\text{bond}}|} \sum_{(i,j)\in\mathcal{E}_{\text{bond}}} H_\delta(\hat{d}_{ij} - d_{ij}), \tag{18}$$

where $H_\delta(\cdot)$ applies a quadratic penalty for small errors and a linear penalty for large errors:

$$H_\delta(e) = \begin{cases} 0.5e^2 & \text{if } |e| \leq \delta \\ \delta(|e| - 0.5\delta) & \text{otherwise.} \end{cases} \tag{19}$$

Side-chain bond angle loss. To maintain correct local stereochemistry while avoiding numerical instabilities associated with angular periodicity, we define the loss directly on the cosine of the bond angles. For each angle triplet $(i,j,k) \in \mathcal{E}_{\text{angle}}$, we compute the cosine via the dot product of normalized bond vectors $\mathbf{v}_{ji}$ and $\mathbf{v}_{jk}$:

$$\cos\theta_{ijk} = \frac{\mathbf{v}_{ji}}{\|\mathbf{v}_{ji}\|} \cdot \frac{\mathbf{v}_{jk}}{\|\mathbf{v}_{jk}\|}. \tag{20}$$

We minimize the L_1 distance between the predicted and ground-truth cosine values:

$$\mathcal{L}_{\text{sc-angle}} = \frac{1}{|\mathcal{E}_{\text{angle}}|} \sum_{(i,j,k)\in\mathcal{E}_{\text{angle}}} \left| \widehat{\cos\theta}_{ijk} - \cos\theta_{ijk} \right|. \tag{21}$$

Sidechain torsion loss To ensure the model captures correct side-chain rotameric states while circumventing the periodicity issues inherent in direct angular regression, we optimize the cosine of the torsion angles. Let $\mathcal{T}$ be the set of side-chain torsion angles, where each angle is defined by an ordered sequence of four atoms (i, j, k, l). We define the bond vectors $\mathbf{v}_1 = \mathbf{x}_j - \mathbf{x}_i$, $\mathbf{v}_2 = \mathbf{x}_k - \mathbf{x}_j$, and $\mathbf{v}_3 = \mathbf{x}_l - \mathbf{x}_k$. The normal vectors to the planes defined by these bonds are given by $\mathbf{n}_1 = \mathbf{v}_1 \times \mathbf{v}_2$ and $\mathbf{n}_2 = \mathbf{v}_2 \times \mathbf{v}_3$. The cosine of the torsion angle ϕ corresponds to the dot product of the normalized plane normals:

$$\cos \phi_{ijkl} = \frac{\mathbf{n}_1}{\|\mathbf{n}_1\|} \cdot \frac{\mathbf{n}_2}{\|\mathbf{n}_2\|}. \tag{22}$$

The loss is defined as the mean absolute difference (L_1 loss) between the predicted and ground-truth cosine values:

$$\mathcal{L}_{\text{sc-torsion}} = \frac{1}{|\mathcal{T}|} \sum_{(i,j,k,l)\in\mathcal{T}} \left| \widehat{\cos\phi}_{ijkl} - \cos\phi_{ijkl} \right|. \tag{23}$$

Backbone torsion loss To ensure the recovery of accurate secondary structure elements (such as α-helices and β-sheets), we enforce geometric constraints on the backbone torsion angles ϕ and ψ. Similar to the side-chain formulation, we mitigate angular periodicity issues by operating directly on the cosine of the angles. Let $\mathcal{T}_{\text{bb}}$ be the set of backbone torsion angles defined by the consecutive backbone atoms (e.g., $C_{i-1} - N_i - C_{\alpha,i} - C_i$ for ϕ and $N_i - C_{\alpha,i} - C_i - N_{i+1}$ for ψ). We minimize the L_1 distance between the predicted and ground-truth cosine values:

$$\mathcal{L}_{\text{bb-torsion}} = \frac{1}{|\mathcal{T}_{\text{bb}}|} \sum_{\theta\in\mathcal{T}_{\text{bb}}} \left| \widehat{\cos\theta} - \cos\theta \right|. \tag{24}$$

The cosine values are computed via the dot product of the normalized normal vectors of the planes defined by the atom quadruplets.

For both the backbone and side-chain torsion losses, we exclusively use the cosine of the angles rather than computing both the sine and cosine. Empirically, we found that relying solely on the cosine facilitates easier and more stable training convergence. This simplification does not introduce adverse effects, as the torsion loss serves as an auxiliary objective while the primary MSE loss drives the overall parameter optimization.

Smooth LDDT loss. To directly optimize the local structural accuracy while maintaining differentiability, we employ a smoothed version of the Local Distance Difference Test (LDDT) score, as suggested in AlphaFold 3 (Abramson et al., 2024). Standard LDDT relies on a non-differentiable step function, and the smooth version approximates it using a sigmoid function.

Let d_{ij} and $\hat{d}_{ij}$ be the ground-truth and predicted Euclidean distances between atoms i and j. We define a binary mask m_{ij} that is 1 if the pair exists in the ground truth and the distance d_{ij} is within the inclusion radius R_{inc}, and 0 otherwise. Following AlphaFold 3, we set $R_{\text{inc}} = 15\text{Å}$.

For each valid pair, we compute the absolute distance error $e_{ij} = |\hat{d}_{ij} - d_{ij}|$. The smooth score contribution is the average of sigmoid activations over four thresholds $\mathcal{T} = \{0.5, 1.0, 2.0, 4.0\}$Å:

$$s_{ij} = \frac{1}{4} \sum_{\tau\in\mathcal{T}} \sigma(\tau - e_{ij}), \tag{25}$$

where $\sigma(\cdot)$ is the sigmoid function. This term smoothly approximates the fraction of thresholds satisfied by the prediction. The final loss is defined as:

$$\mathcal{L}_{\text{smooth-LDDT}} = 1 - \frac{\sum_{i,j} m_{ij} s_{ij}}{\sum_{i,j} m_{ij} + \epsilon}. \tag{26}$$

Final Objective.

$$\mathcal{L}_{aux} = 0.5 \cdot \mathcal{L}_{\text{center}} + \mathcal{L}_{\text{bond}} + \mathcal{L}_{\text{collision}} + 6 \cdot \mathcal{L}_{\text{sc-angle}} + 6 \cdot \mathcal{L}_{\text{sc-torsion}} + \mathcal{L}_{\text{sc-bond}} + 1.5 \cdot \mathcal{L}_{\text{bb-torsion}} + 3 \cdot \mathcal{L}_{\text{smooth-LDDT}} \tag{27}$$

Finally, the AAT-VAE loss is

$$\mathcal{L}_{vae} = \mathcal{L}_{\text{aligned-mse}}(\lambda_{\text{ligand}} = 10) + 0.5 \cdot \min(1.0, \text{step}/500) \cdot \text{KL}(\|) + \mathcal{L}_{aux} \tag{28}$$

For the flow matching loss, we also consider applying a higher weight to the ligand.

$$\mathcal{L}_{fm} = \frac{1}{N(T-1)} \sum_{t=1}^{T-1} \sum_{i=1}^{N} \omega_i \|\mathbf{r}_\theta(\mathbf{r}_{t,i}^{\tau}, \mathbf{c}_i, \tau) - \mathbf{r}_{t,i}^{1}\|_2^2 \tag{29}$$

where $\omega_i = 25$ if atom i belongs to ligand, otherwise $\omega_i = 1$.

G. Training and Inference Algorithms

The main training and inference procedure is shown in algorithm 1.

Algorithm 1: Main Loop of `GPCRLMD`

```
Input: $\{\mathbf{f}^*\}$ , $\{\vec{\mathbf{x}}_{0,l}\}$, $N_{cycle} = 4$, $c_s = 384$, $c_z = 128$:
    # Construct the token embeddings, including GPCR-ligand specific features;
    $\{\mathbf{s}_i^{\text{inputs}}\} \leftarrow$ InputFeatureEmbedder($\{\mathbf{f}^*\}$);
    $\mathbf{s}_i^{\text{init}} \leftarrow \text{LinearNoBias}(\mathbf{s}_i^{\text{inputs}})$;
    $\mathbf{z}_{ij}^{\text{init}} \leftarrow \text{LinearNoBias}(\mathbf{s}_i^{\text{inputs}}) + \text{LinearNoBias}(\mathbf{s}_j^{\text{inputs}})$;
    $\{\mathbf{z}_{ij}\}, \{\mathbf{s}_i\} \leftarrow 0, 0$;
    foreach $c \in \{1, \ldots, N_{cycle}\}$ do
        $\mathbf{z}_{ij} \leftarrow \mathbf{z}_{ij}^{\text{init}} + \text{LinearNoBias}(\text{LayerNorm}(\mathbf{z}_{ij}))$;
        $\{\mathbf{z}_{ij}\}, \{\mathbf{s}_i\} \leftarrow$ GraphTransformer($\{\vec{\mathbf{x}}_{0,l}\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{s}_i^{\text{inputs}}\}$);
        $\mathbf{s}_i \leftarrow \mathbf{s}_i^{\text{init}} + \text{LinearNoBias}(\text{LayerNorm}(\mathbf{s}_i))$;
    if TrainAATVAE :
        $\mathcal{L}_{vae} \leftarrow$ TrainAATVAE($\{\vec{\mathbf{x}}_l\}$, $\mathbf{t}$, $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i\}$, $\{\mathbf{z}_{ij}\}$);
        return $\mathcal{L}_{vae}$
    elif TrainLatentFlow :
        $\mathcal{L}_{fm} \leftarrow$ TrainLatentFlow($\{\vec{\mathbf{x}}_l\}$, $\mathbf{t}$, $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i\}$, $\{\mathbf{z}_{ij}\}$);
        return $\mathcal{L}_{fm}$
    else:
        # Inference;
        traj_list=[$\{\vec{\mathbf{x}}_{0,l}\}$] ;
        $\{\vec{\mu}_l, \vec{\sigma}_l\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\} \leftarrow$ Encoder($\{\vec{\mathbf{x}}_{0,l}\}$, $\mathbf{t}$, $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i\}$, $\{\mathbf{z}_{ij}\}$);
        $\{\vec{\mathbf{z}}_{t,l}^{\text{pred}}\} \leftarrow$ SampleLatentFlow($\{\vec{\mathbf{x}}_{0,l}\}, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{s}_i\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}$);
        traj_list $\leftarrow$ Decoder($\{\vec{\mathbf{z}}_{t,l}^{\text{pred}}\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}, \beta_{ij} = 0$);
        return traj_list
```

Algorithm 2: TrainAATVAE

```
Input: $\{\vec{\mathbf{x}}_l\}$, $\mathbf{t}$, $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i^{\text{trunk}}\}$, $\{\mathbf{z}_{ij}^{\text{trunk}}\}$:
    # Data Augmentation for exempting SE(3) Equivalent Model design (Algorithm 19 in AF3) ;
    $\{\vec{\mathbf{x}}_l\} \leftarrow \text{CentreRandomAugmentation}(\{\vec{\mathbf{x}}_l\})$;
    $(\{\vec{\mu}_l, \vec{\sigma}_l\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}) \leftarrow$ Encoder($\{\vec{\mathbf{x}}_l\}, \tau, \{\mathbf{f}^*\}, \{\mathbf{s}_i^{\text{inputs}}\}, \{\mathbf{s}_i^{\text{trunk}}\}, \{\mathbf{z}_{ij}^{\text{trunk}}\}$);
    $\{\mathbf{q}_l\} \leftarrow$ Reparameterization($\{\vec{\mu}_l, \vec{\sigma}_l\}$);
    $\{\hat{\vec{\mathbf{x}}}_l\} \leftarrow$ Decoder($\{\mathbf{q}_l\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}, \beta_{ij} = 0$);
    $\mathcal{L}_{vae} = \text{MSE}(\{\hat{\vec{\mathbf{x}}}_l\}, \{\vec{\mathbf{x}}_l\}) + \{\text{KL}(\mathcal{N}(z; \vec{\mu}_l, \vec{\sigma}_l) \| \mathcal{N}(\mathbf{x}_l, \sigma\mathbf{I}))\}$;
    $\mathcal{L}_{aux} = \text{GeometricConstraint}(\{\hat{\vec{\mathbf{x}}}_l\}, \{\vec{\mathbf{x}}_l\})$;
    return: $\mathcal{L}_{vae} + \mathcal{L}_{aux}$
```

```
Algorithm 3: TrainLatentFlow
Input: $\{\vec{\mathbf{x}}_l\}$, $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i^{\text{trunk}}\}$, $\{\mathbf{z}_{ij}^{\text{trunk}}\}$:
    # Data Augmentation for exempting SE(3) Equivalent Model design (Algorithm 19 in AF3) ;
    $\{\vec{\mathbf{x}}_l\} \leftarrow$ CentreRandomAugmentation($\{\vec{\mathbf{x}}_l\}$);
    $(\{\vec{\mu}_l, \vec{\sigma}_l\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}) \leftarrow$ Encoder($\{\vec{\mathbf{x}}_l^{\tau}\}, \tau, \{\mathbf{f}^*\}, \{\mathbf{s}_i^{\text{inputs}}\}, \{\mathbf{s}_i^{\text{trunk}}\}, \{\mathbf{z}_{ij}^{\text{trunk}}\}$);
    $\mathbf{q}_l \leftarrow \vec{\mu}_l$;
    $\{\mathbf{q}_l$.stop_grad()$\}, \{\mathbf{s}_i$.stop_grad()$\}, \{\mathbf{z}_{ij}$.stop_grad()$\}$;
    $\{\mathbf{a}_i^{\text{skip}}$.stop_grad()$\}, \{\mathbf{c}_l^{\text{skip}}$.stop_grad()$\}, \{\mathbf{p}_{lm}^{\text{skip}}$.stop_grad()$\}$;
    $\mathcal{L}_{fm} \leftarrow$ LatentFlowMatching($\{\mathbf{q}_l\}, \{\vec{\mathbf{x}}_{0,l}\}, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{s}_i\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}$;
    return: $\mathcal{L}_{fm}$
```

```
Algorithm 4: LatentFlowMatching
Input: $\{\vec{\mathbf{z}}_l\}$,$\{\vec{\mathbf{x}}_{0,l}\}$, $\{\mathbf{a}_i^{\text{skip}}\}$, $\{\mathbf{s}_i^{\text{trunk}}\}$,$\{\mathbf{c}_l^{\text{skip}}\}$, $\{\mathbf{p}_{lm}^{\text{skip}}\}$:
    # Independent noise levels ;
    $\tau \sim (\mathcal{U}(0,1), \mathcal{U}(0,1), \cdots, \mathcal{U}(0,1))$;
    $\{\vec{\mathbf{r}}_l^0\} \sim \mathcal{N}(\vec{0}, \mathbf{I}_3)$ ;
    # Calculate the residual vector;
    $\vec{\mathbf{r}}_l = \vec{\mathbf{z}}_l - \vec{\mathbf{x}}_{0,l}$ ;
    $\{\vec{\mathbf{r}}_l^{\tau}\} = \tau\{\vec{\mathbf{r}}_l\} + (1-\tau)\{\vec{\mathbf{r}}_l^0\}$;
    $\{\hat{\vec{\mathbf{r}}}_l^1\} \leftarrow$ LatentVelocityNetwork($\{\vec{\mathbf{r}}_l^{\tau}\}, \tau, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{s}_i^{\text{trunk}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}$);
    $\mathcal{L}_{flow} = \text{MSE}(\{\vec{\mathbf{r}}_l\}, \{\hat{\vec{\mathbf{r}}}_l^1\})$;
    return: $\mathcal{L}_{flow}$
```

H. Evaluation

H.1. Main quantitative metrics

We adopt the evaluation criteria proposed by Jing et al. (2024a), which can be grouped as follows:

- **Flexibility correlation (↑):** Pearson correlation coefficient r computed for pairwise RMSD, global RMSF, and per-target RMSF.
- **Distributional accuracy:** Root mean of the 2-Wasserstein distance (W_2-dist) along with its translation and variance components (↓), molecular dynamics (MD) PCA W_2-dist (↓), joint PCA W_2-dist (↓), and the proportion of samples achieving PC-sim > 0.5 (↑).
- **Ensemble observables (↑):** Jaccard similarity J for weak contacts, transient contacts, and exposed residues, together with the Spearman correlation ρ of the exposed mutual information (MI) matrix.

JS Divergence. We use the JS Divergence to measure the distribution distinction in this paper. The Jensen-Shannon Divergence between two probability vectors p and q is defined as,

$$\frac{D(p \parallel m) + D(q \parallel m)}{2}$$

where m is the pointwise mean of and p and q, and D is the Kullback-Leibler divergence.

This routine will normalize p and q if they do not sum to 1.0.

We use JSD or JS^2 to represent the JS Divergence throughout this paper.

Algorithm 5: SampleLatentFlow

```
Input: $\{\vec{\mathbf{x}}_{0,l}\}$, $\{\mathbf{a}_i^{\text{skip}}\}$, $\{\mathbf{s}_i^{\text{trunk}}\}$, $\{\mathbf{c}_l^{\text{skip}}\}$, $\{\mathbf{p}_{lm}^{\text{skip}}\}$:
    $\vec{\mathbf{r}}_l^0 \sim \mathcal{N}(\vec{0}, \mathbf{I}_3)$;
    $d\tau = 0.1$;
    foreach $\tau$ in $\{0, 0.1, 0.2, ..., 0.9\}$ do
        $\{\vec{\hat{\mathbf{r}}}_l^1\} \leftarrow$ LatentVelocityNetwork($\{\vec{\mathbf{r}}_l^\tau\}, \tau, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{s}_i^{\text{trunk}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}$);
        $\vec{\mathbf{r}}_l^{\tau+d\tau} \leftarrow \vec{\mathbf{r}}_l^\tau + d\tau \cdot (\frac{\vec{\hat{\mathbf{r}}}_l^1 - \vec{\hat{\mathbf{r}}}_l^\tau}{1-\tau})$;
    return: $\{\vec{\mathbf{r}}_l^1 + \vec{\mathbf{x}}_{0,l}\}$
```

Algorithm 6: InputFeatureEmbedder

```
Input: $\{\mathbf{f}^*\}$:
    # Construct the GPCR-ligand complex specific features. $c = 8 + 8$;
    $\mathbf{a}_i^{\text{gpcr}} \leftarrow$ concat($\mathbf{f}_i^{\text{TM\_index}}, \mathbf{f}_i^{\text{lig\_func}}$);
    # Concatenate the per-token features.;
    $\mathbf{s}_i \leftarrow$ concat($\mathbf{a}_i^{\text{gpcr}}, \mathbf{f}_i^{\text{restype}}$);
    return: $\{\mathbf{s}_i\}$
```

H.2. TICA Analysis

To evaluate the kinetic fidelity of the generated ensembles, we employ Time-lagged Independent Component Analysis (TICA) (Jing et al., 2024b). TICA identifies collective variables (slow modes) by maximizing their time-lagged autocorrelation:

$$\text{maximize } \rho_i = \text{corr}\big(y_i(t), y_i(t+\tau)\big). \tag{30}$$

H.3. Fidelity metrics

We use dihedral angle coupling distributions to verify the physical Fidelity of the conformational ensemble.

- Side chain torsional angles χ_1, χ_2.
- Backbone torsional angle ϕ and ψ.

H.4. Dual-cutoff contact map calculation

To characterize stable ligand-protein interactions and mitigate high-frequency noise arising from thermal fluctuations at the cutoff boundary, we employ a dual-cutoff hysteresis scheme for contact definition. Let $d_{ij}^{(t)}$ denote the Euclidean distance between a protein heavy atom i and a ligand heavy atom j at frame t. The binary atomic contact state $s_{ij}^{(t)}$ is determined recursively to enforce stability:

$$s_{ij}^{(t)} = \begin{cases} 1 & \text{if } d_{ij}^{(t)} \leq r_{\text{on}} \\ 0 & \text{if } d_{ij}^{(t)} \geq r_{\text{off}} \\ s_{ij}^{(t-1)} & \text{otherwise (hysteresis region),} \end{cases} \tag{31}$$

with thresholds set to $r_{\text{on}} = 3.5\text{Å}$ and $r_{\text{off}} = 5.0\text{Å}$ and $s_{ij}^{(0)} = 0$. We aggregate these atomic states to the residue level: a residue R_k is considered in contact with ligand atom j at frame t if any of its constituent heavy atoms are in contact:

$$c_{kj}^{(t)} = \bigvee_{i \in R_k} s_{ij}^{(t)}. \tag{32}$$

Algorithm 7: Encoder

```
Input: $\{\vec{\mathbf{x}}_l\}$, $\mathbf{t}$, $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i^{\text{trunk}}\}$, $\{\mathbf{z}_{ij}^{\text{trunk}}\}$, $\sigma_{\text{data}} = 16$, $c_{\text{atom}} = 128$, $c_{\text{atompair}} = 16$, $c_{\text{token}} = 768$:
    $\{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\} \leftarrow$ VAEConditioning($\{\mathbf{f}^*\}, \{\mathbf{s}_i^{\text{inputs}}\}, \{\mathbf{s}_i^{\text{trunk}}\}, \{\mathbf{z}_{ij}^{\text{trunk}}\}, \sigma_{\text{data}}$);
    # Sequence-local Atom Attention and aggregation to coarse-grained tokens (following Ag. 5 in AF3), and we also add frame-index ($\mathbf{t}$) embeddings before AtomTransformer into $\mathbf{q}_l$ ;
    $\{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{q}_l\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\} \leftarrow$ AtomAttentionEncoder($\{\mathbf{f}^*\}, \{\vec{\mathbf{x}}_l\}, \{\mathbf{s}_i\}, \mathbf{t}, \{\mathbf{z}_{ij}\}, c_{\text{atom}}, c_{\text{atompair}}, c_{\text{token}}$);
    # Linear heads to get the mean and variane of VAE with atom-level structure-relevant variable $\mathbf{q}_l$;
    $\{\vec{\mu}_l, \vec{\sigma}_l\} \leftarrow$ MeanVarianceHeads($\{\mathbf{q}_l\}$);
    return: $\{\vec{\mu}_l, \vec{\sigma}_l\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\mathbf{a}_i^{\text{skip}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}$
```

Algorithm 8: VAEConditioning

```
Input: $\{\mathbf{f}^*\}$, $\{\mathbf{s}_i^{\text{inputs}}\}$, $\{\mathbf{s}_i^{\text{trunk}}\}$, $\{\mathbf{z}_{ij}^{\text{trunk}}\}$, $\sigma_{\text{data}}$, $c_z = 128$, $c_s = 384$:
    # Pair conditioning (AF3) ;
    $\mathbf{z}_{ij} \leftarrow$ concat([ $\mathbf{z}_{ij}^{\text{trunk}}$, RelativePositionEncoding($\{\mathbf{f}^*\}$) ]);
    $\mathbf{z}_{ij} \leftarrow$ LinearNoBias(LayerNorm($\mathbf{z}_{ij}$));
    foreach $b \in \{1, 2\}$ do
        $\mathbf{z}_{ij}$ += Transition($\mathbf{z}_{ij}$, $n = 2$);
    # Single conditioning;
    $\mathbf{s}_i \leftarrow$ concat([ $\mathbf{s}_i^{\text{trunk}}$, $\mathbf{s}_i^{\text{inputs}}$ ]);
    $\mathbf{s}_i \leftarrow$ LinearNoBias(LayerNorm($\mathbf{s}_i$));
    foreach $b \in \{1, 2\}$ do
        $\mathbf{s}_i$ += Transition($\mathbf{s}_i$, $n = 2$);
    return: $\{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}$
```

The final contact map represents the interaction frequency over the trajectory of length T:

$$P_{kj} = \frac{1}{T} \sum_{t=1}^{T} c_{kj}^{(t)}. \tag{33}$$

Similarly, the calculation of dual-cutoff contact maps between amino acids can be performed simply by replacing the ligand with amino acids.

Algorithm 9: Decoder

```
Input: $\{\mathbf{q}_l\}$, $\{\mathbf{s}_i\}$, $\{\mathbf{z}_{ij}\}$, $\{\mathbf{c}_l^{\text{skip}}\}$, $\{\mathbf{p}_{lm}^{\text{skip}}\}$, $\{\beta_{ij}\}$, $N_{\text{block}} = 3$, $N_{\text{head}} = 4$:
    $\{\mathbf{q}_l\} \leftarrow \text{Linear}(\{\mathbf{q}_l\})$;
    # Aggregate Latent atom representations to token representation.;
    $\mathbf{a}_i \leftarrow \text{AggregateAtomToToken}(\{\mathbf{q}_l\})$;
    # Full self-attention on token level;
    $\mathbf{a}_i += \text{LinearNoBias}(\text{LayerNorm}(\mathbf{a}_i))$;
    $\{\mathbf{a}_i\} \leftarrow \text{TemporalDiffusionTransformer}(\{\mathbf{a}_i\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \beta_{ij} = 0, N_{\text{block}} = 6, N_{\text{head}} = 16)$;
    $\mathbf{a}_i \leftarrow \text{LayerNorm}(\mathbf{a}_i)$;
    # Broadcast token activations to atoms and run Sequence-local Atom Attention;
    $\{\vec{\mathbf{x}}_l^{\text{out}}\} = \text{AtomTemporalAttentionDecoder}(\{\mathbf{a}_i\}, \{\mathbf{q}_l\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\})$ ;
    return: $\{\vec{\mathbf{x}}_l^{\text{out}}\}$
```

Algorithm 10: LatentVelocityNetwork

```
Input: $\{\mathbf{r}_i^\tau\}$,$\tau$,$\{\mathbf{a}_i\}$, $\{\mathbf{s}_i\}$, $\{\mathbf{c}_l^{skip}\}$, $\{\mathbf{p}_{lm}^{\text{skip}}\}$, $N_{\text{block}} = 3$, $N_{\text{head}} = 4$:
    # Stochastic Token Propagator is used to transfer the structural context from the initial frame tokens ($\mathbf{a}_0$) into the temporal domain ;
    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$;
    $\{\mathbf{a}_{t>0,i}\} \leftarrow \{\tau \cdot \mathbf{a}_{0,i} + 0.1 \cdot (1 - \tau) \cdot \epsilon\}$;
    $\mathbf{a}_i \leftarrow \mathbf{a}_{\{t\},i}$;
    $\mathbf{a}_i \leftarrow \text{LayerNorm}(\mathbf{a}_i + \text{LinearNoBias}(\text{LayerNorm}(\mathbf{s}_i)))$;
    $\mathbf{r}_l = \text{LinearNoBias}(\mathbf{r}_l^\tau)$;
    $\mathbf{r}_l \leftarrow 0.1 \cdot \text{LinearNoBias}(\mathbf{a}_{\text{tok\_idx(l)}}) + \mathbf{r}_l$;
    # Cross attention transformer from AF3.;
    $\{\mathbf{r}_l\} = \text{AtomTransformer}(\{\mathbf{r}_l\}, \{\mathbf{c}_l^{skip}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}, N_{\text{block}} = 3, N_{\text{head}} = 4)$;
    # Map to positions update;
    $\mathbf{r}_l^{\text{update}} = \text{LinearNoBias}(\text{LayerNorm}(\mathbf{r}_l))$;
    return: $\{\mathbf{r}_l^{\text{update}}\}$
```

Algorithm 11: AtomTemporalTransformer

```
Input: $\{\mathbf{q}_l\}, \{\mathbf{c}_l\}, \{\mathbf{p}_{lm}\}, N_{\text{block}} = 3, N_{\text{head}}, N_{\text{queries}} = 32, N_{\text{keys}} = 128, \mathcal{S}_{\text{subset centres}} = \{15.5, 47.5, 79.5, \dots\}$:
    $\{\mathbf{q}_l\} = \text{TemporalDiffusionTransformer}(\{\mathbf{q}_l\}, \{\mathbf{c}_l\}, \{\mathbf{p}_{lm}\}, \{\beta_{ij} = 0\}, N_{\text{block}}, N_{\text{head}})$;
    return: $\{\mathbf{q}_l\}$
```

Algorithm 12: AtomTemporalAttentionDecoder

```
Input: $\{\mathbf{a}_i\}, \{\mathbf{q}_l^{\text{skip}}\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}$:
    # Broadcast per-token activiations to per-atom activations and add the skip connection;
    $\mathbf{q}_l = \text{LinearNoBias}(\mathbf{a}_{\text{tok\_idx}(l)}) + \mathbf{q}_l^{\text{skip}}$;
    # Cross attention temporal transformer.;
    $\{\mathbf{q}_l\} = \text{AtomTemporalTransformer}(\{\mathbf{q}_l\}, \{\mathbf{c}_l^{\text{skip}}\}, \{\mathbf{p}_{lm}^{\text{skip}}\}, N_{\text{block}} = 3, N_{\text{head}} = 4)$;
    # Map to positions update.;
    $\mathbf{x}_l = \text{LinearNoBias}(\text{LayerNorm}(\mathbf{q}_l))$;
    return: $\{\mathbf{x}_l\}$
```

Algorithm 13: TemporalDiffusionTransformer

```
Input $\{\mathbf{a}_i\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\beta_{ij}\}, N_{\text{block}} = 6, N_{\text{head}} = 16$:
    for $n \in [1, \ldots, N_{\text{block}}]$ do
        # Algorithm 24 in AF3;
        $\{\mathbf{b}_i\} \leftarrow \text{AttentionPairBias}(\{\mathbf{a}_i\}, \{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{\beta_{ij}\}, N_{\text{head}})$;
        # Multihead attention on time-level, following Feng et al. (2025);
        $\{\mathbf{b}_i\} \leftarrow \text{TemporalAttention}(\{\mathbf{a}_i + \mathbf{b}_i\})$;
        # Algorithm 25 in AF3;
        $\mathbf{a}_i \leftarrow \mathbf{b}_i + \text{ConditionedTransitionBlock}(\mathbf{a}_i, \mathbf{s}_i)$;
    return $\{\mathbf{a}_i\}$
```

Algorithm 14: TemporalAttention

```
Input: Single representation $c_s$ of shape $(T, B, N, c_s)$, where $T$ is time, $B$ is batch size, and $N$ is residues.:
    # Permute dimensions to make time the sequence axis for attention;
    $c_s' \leftarrow \text{Permute}(c_s, \text{dims} = (2, 1, 0, 3))$;
    # Project to Query, Key, Value for each residue independently;
    $Q \leftarrow \text{Linear}_Q(c_s')$;
    $K \leftarrow \text{Linear}_K(c_s')$;
    $V \leftarrow \text{Linear}_V(c_s')$;
    # Calculate scaled dot-product attention scores across time;
    $d_k \leftarrow$ dimension of $K$;
    $\text{logits} \leftarrow (Q \cdot K^T)/\sqrt{d_k}$;
    $\text{weights} \leftarrow \text{Softmax}(\text{logits}, \text{dim} = -1)$;
    # Apply attention weights to values;
    $\text{output}' \leftarrow \text{weights} \cdot V$;
    # Permute back to the original dimension order;
    $\mathbf{o} \leftarrow \text{Permute}(\text{output}', \text{dims} = (2, 1, 0, 3))$;
    return: $\mathbf{o}$;
```

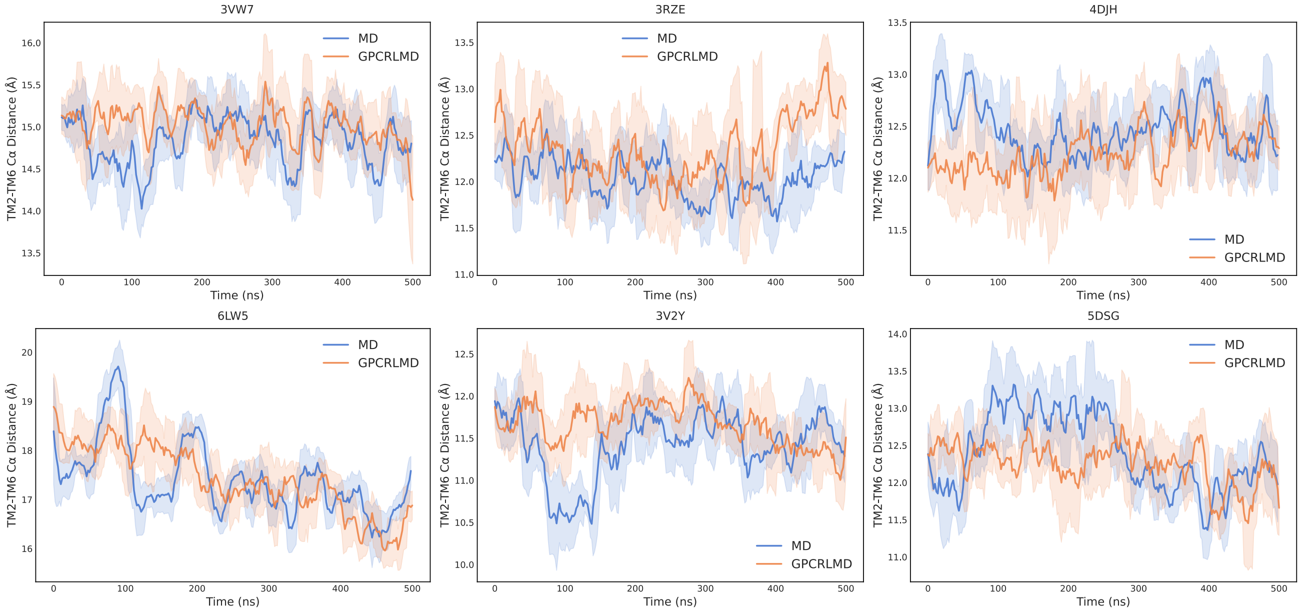

Figure 8. TM2-TM6 distance. For GPCR class A, B1, and C, we calculate the distance between residue 2×46 and 6×37. For class F, we calculate distance $2 \times 46 - 6 \times 31$ (Aranda-García et al., 2025). Our model keeps a similar movement tendency to the MD reference.

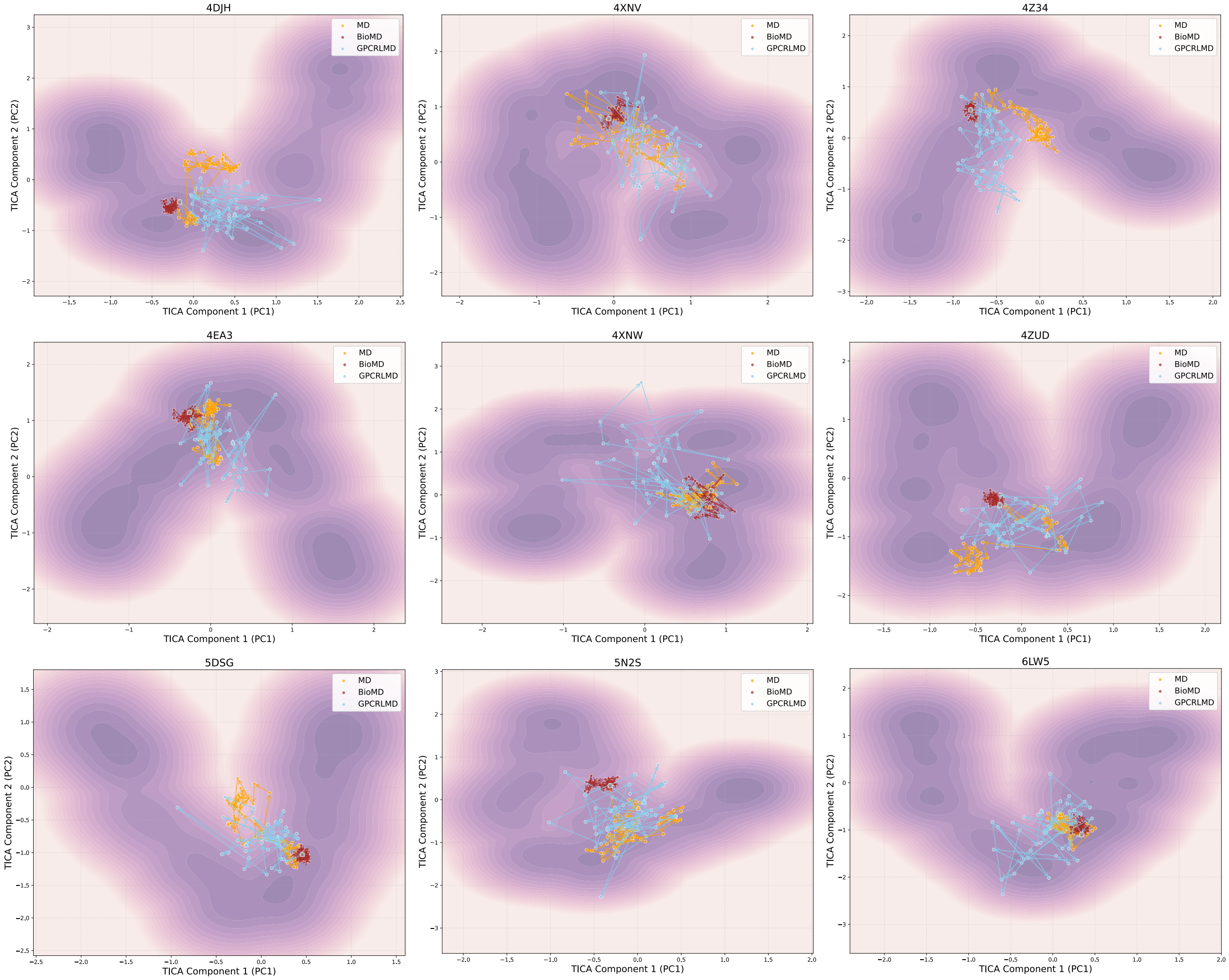

Figure 9. TICA Visualization of the 500ns trajectories. The TICA features only contain the backbone. Baseline MD is a 500ns trajectory, downsampling by 2ns to get 250 frames to fairly align the generative method BioMD and `GPCRLMD`. The pink background indicates the density of the ground-truth conformation distribution from the MD references (replica 2 and replica 3, total 1.0 μs). `GPCRLMD` shows high recovery of conformational states in most cases, sampling more diverse conformations. The quantization results (JSD) on all test samples are shown in the Table 10.

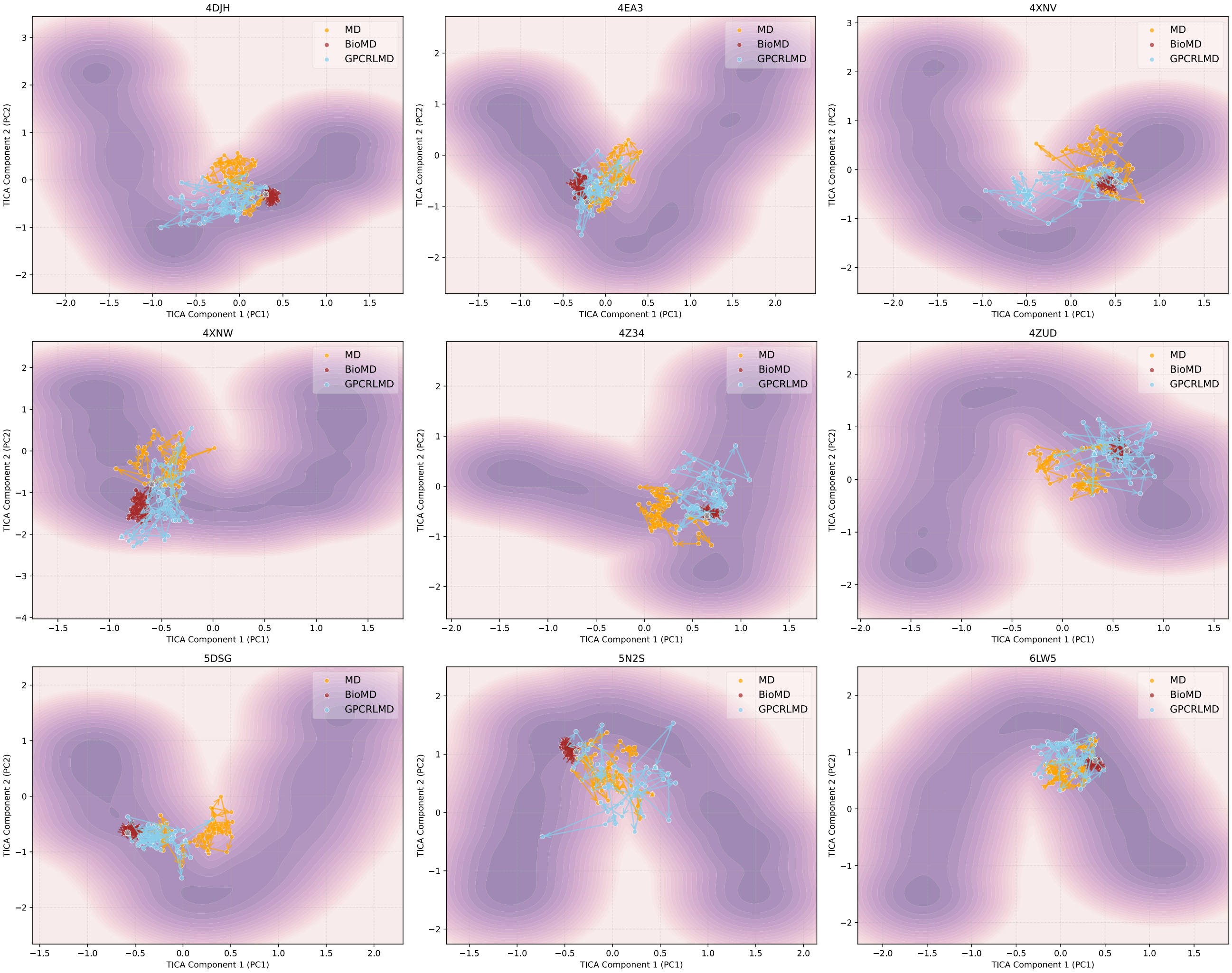

Figure 10. TICA Visualization of the 500ns trajectories. Baseline MD is a 500 ns trajectory, downsampled by 2 ns to obtain 250 frames to fairly align the generative methods BioMD and `GPCRLMD`. The TICA features include the backbone and side chain. The pink background indicates the density of the ground-truth conformation distribution from the MD references (replica 2 and replica 3, total 1.0 μs). `GPCRLMD` shows high recovery of conformational states in most cases, sampling more diverse conformations. The quantization results (JSD) on all test samples are shown in the Table 10.

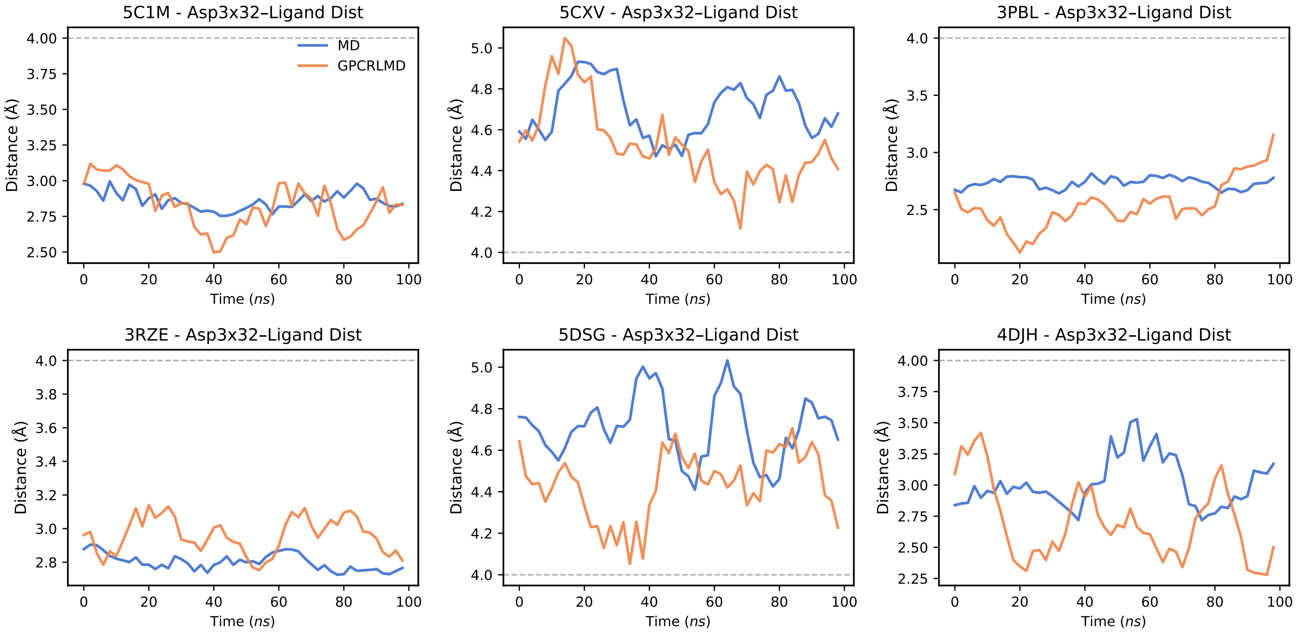

Figure 11. **Salt bridge stability analysis.** We plot the distance between the carboxylate group of residue Asp$^{3.32}$ and the closest nitrogen atom of the ligand over time. A distance < 4.0 Å (dashed grey line) indicates the existence of a salt bridge. The results indicate that `GPCRLMD` effectively maintains the salt bridge structure only if it is present in the initial conformation; otherwise, the distance fluctuates outside the interaction threshold, consistent with the starting structure.

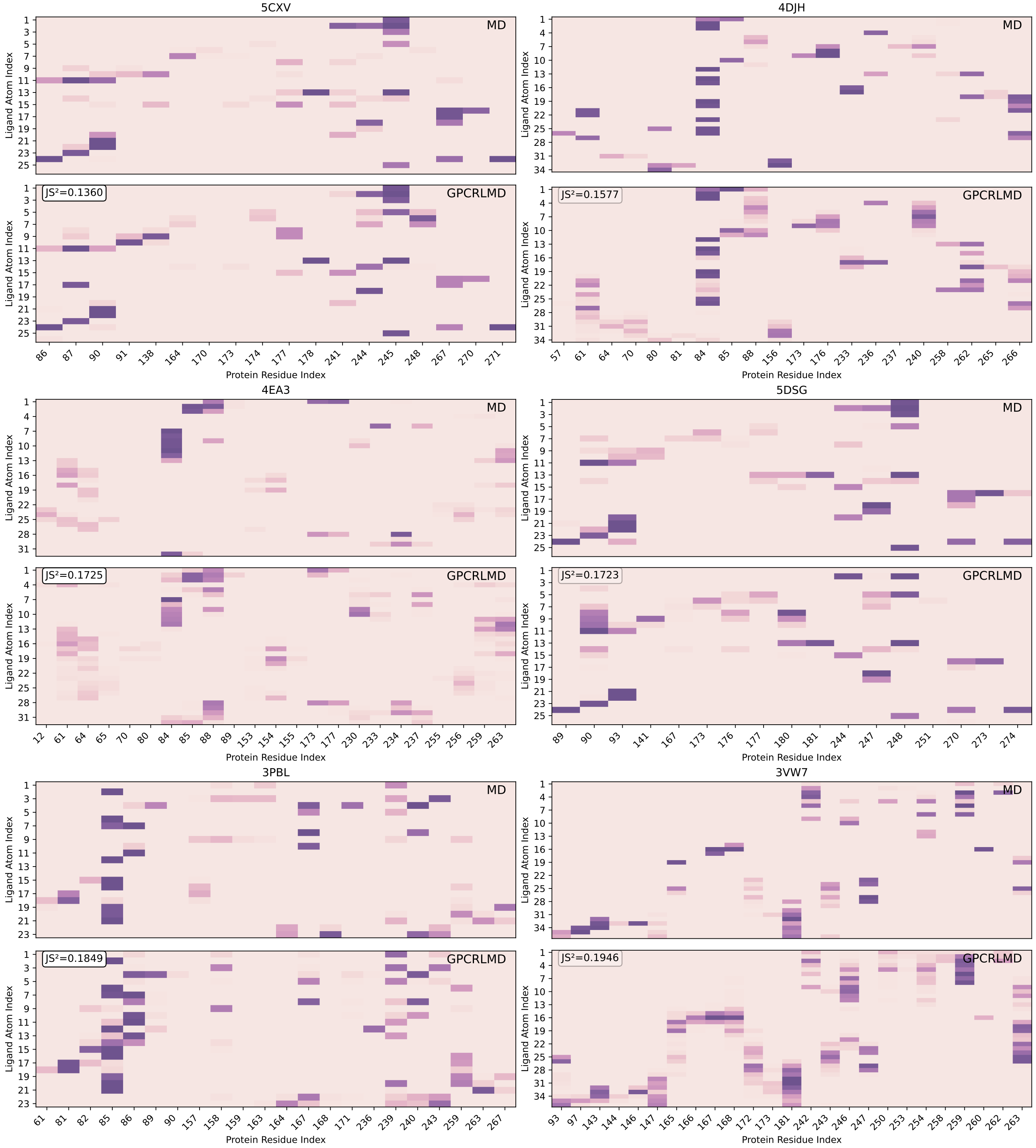

Figure 12. The contact map between ligand and receptor. We compare the 100ns trajectory between the reference MD and `GPCRLMD` prediction, and the results indicate that our approach is highly consistent with the ligand-receptor interactions presented by Ground Truth MD.

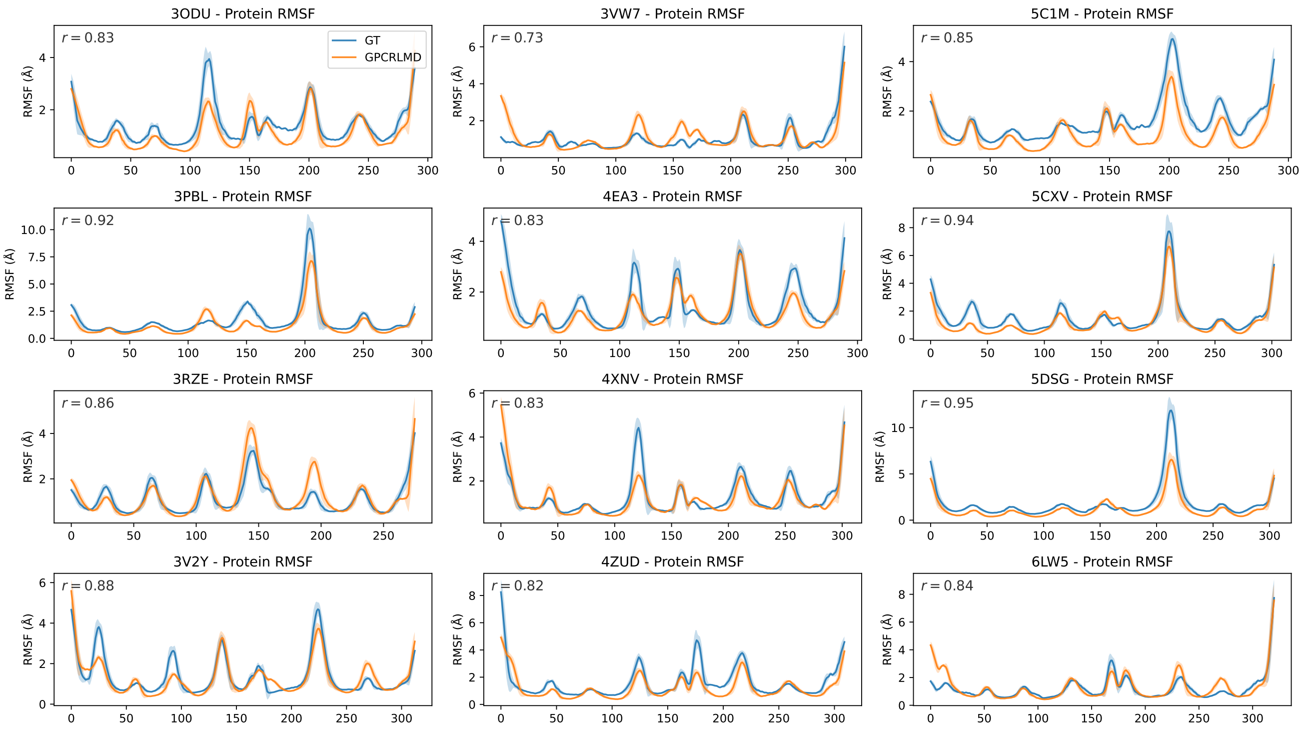

Figure 13. RMSF visualization of GPCR receptor on the test set. This results show that our method achieves high RMSF similarity with the original MD, especially for the receptor. Notably, our method can maintain the low flexibility in the seven helix region while reproducing high diversity on the loop area.

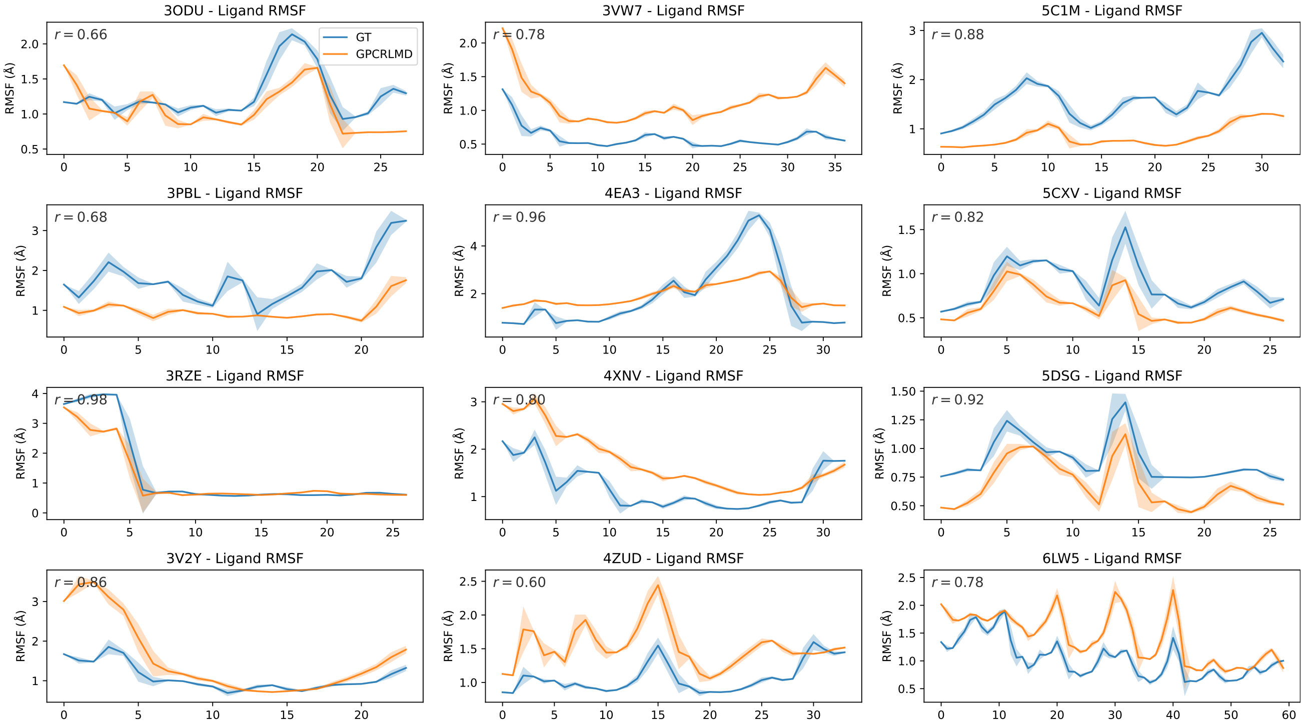

Figure 14. RMSF visualization of ligand on the test set.

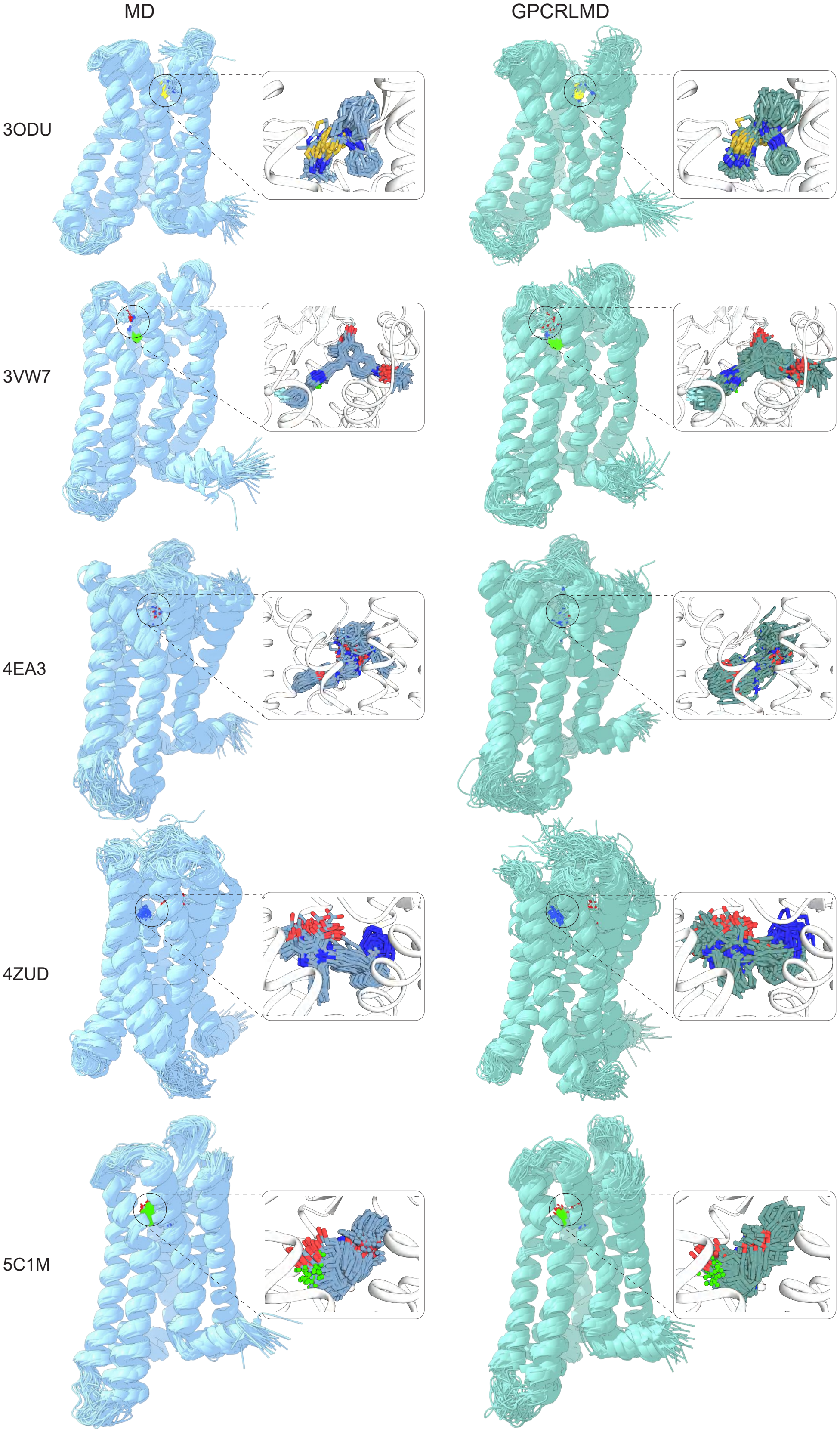

Figure 15. Ensemble visualization. We compare `GPCRLMD` with the MD reference (100ns).

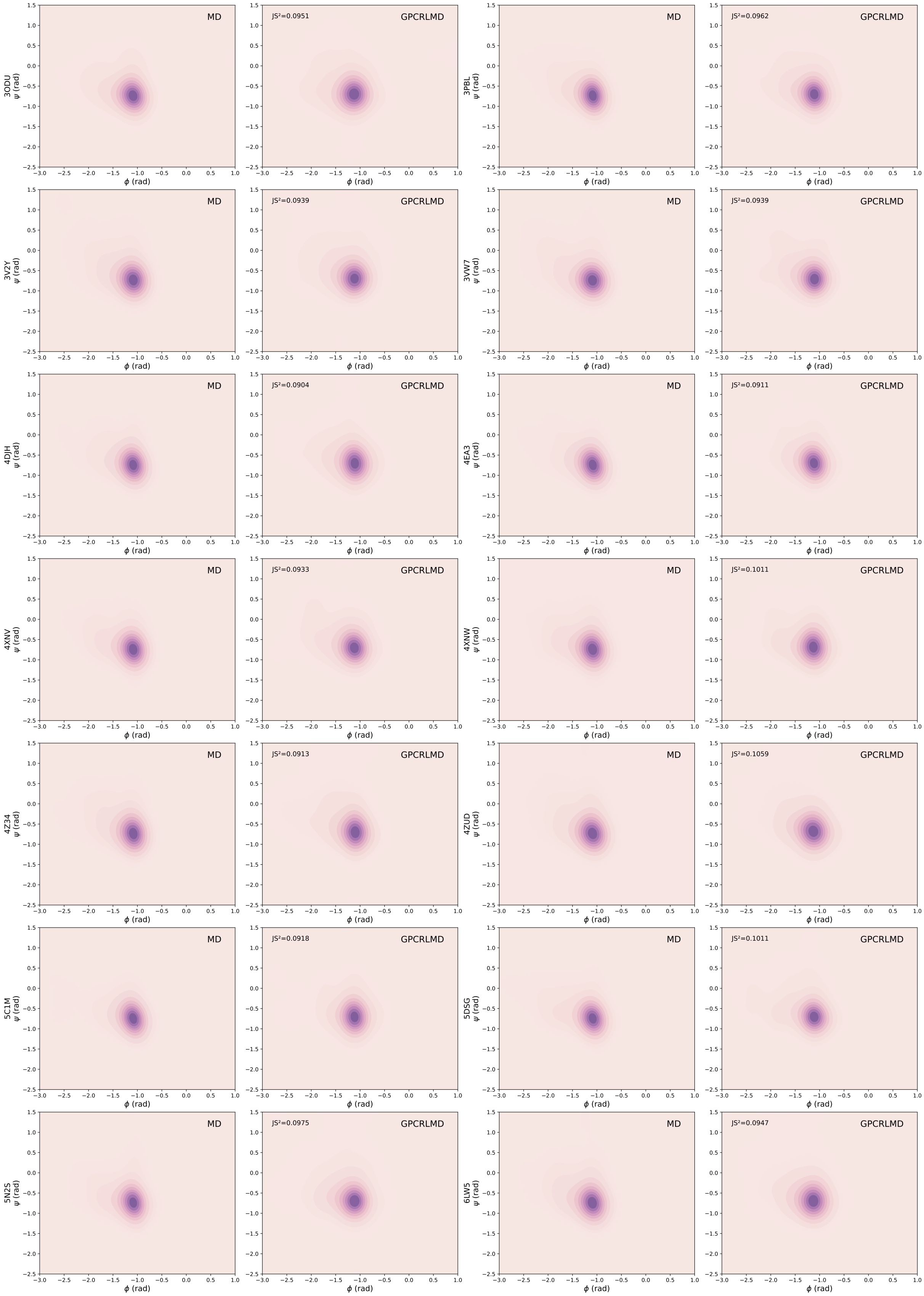

Figure 16. Backbone dihedral angles ψ, ϕ coupling distribution visualization. Our method generates a receptor ensemble with a high degree of similarity in backbone angular coupling compared to the reference MD.

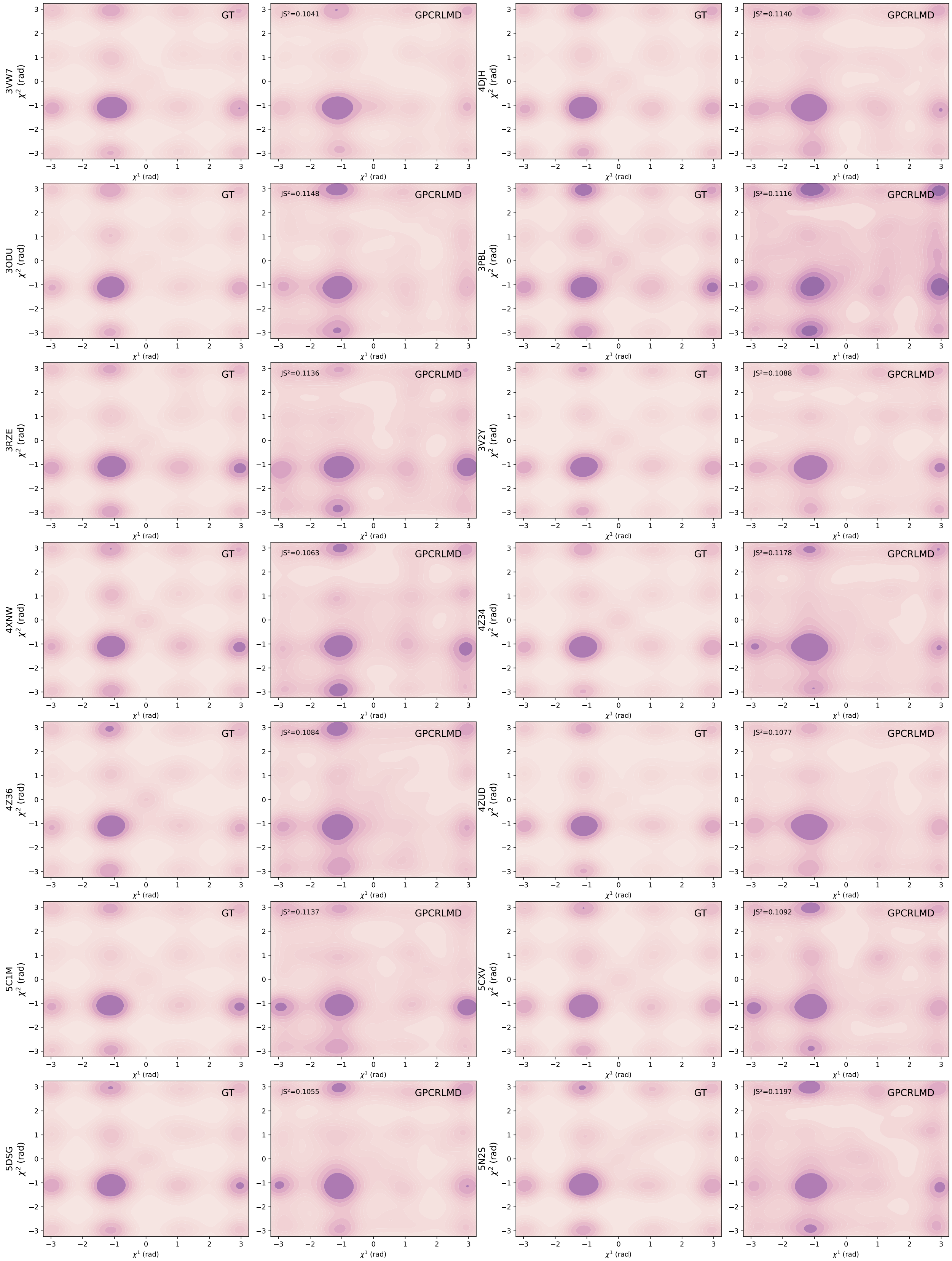

Figure 17. Side chain torsional angles χ^1, χ^2 distribution visualization. Our method generates a receptor ensemble with a high degree of similarity in side-chain angular coupling compared to the reference MD.

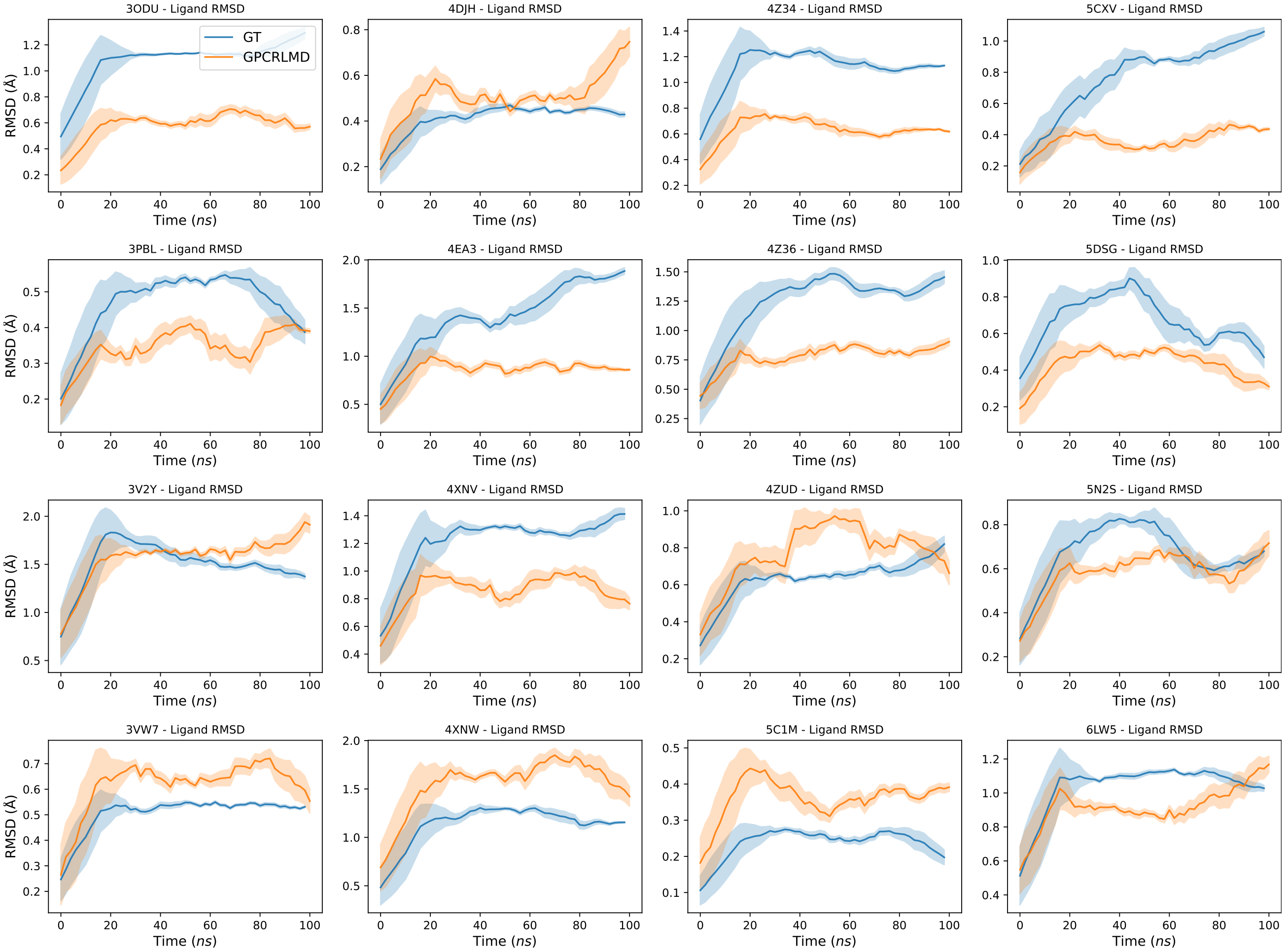

Figure 18. The ligand RMSD change curve over time. We calculate the RMSD between the first frame (initial ligand conformation) and t-th frame in the MD and prediction trajectory, respectively. The results indicate that our method can maintain the same pattern of reference MD in about one-third of the samples and also shows higher stability than reference MD in the remaining samples.

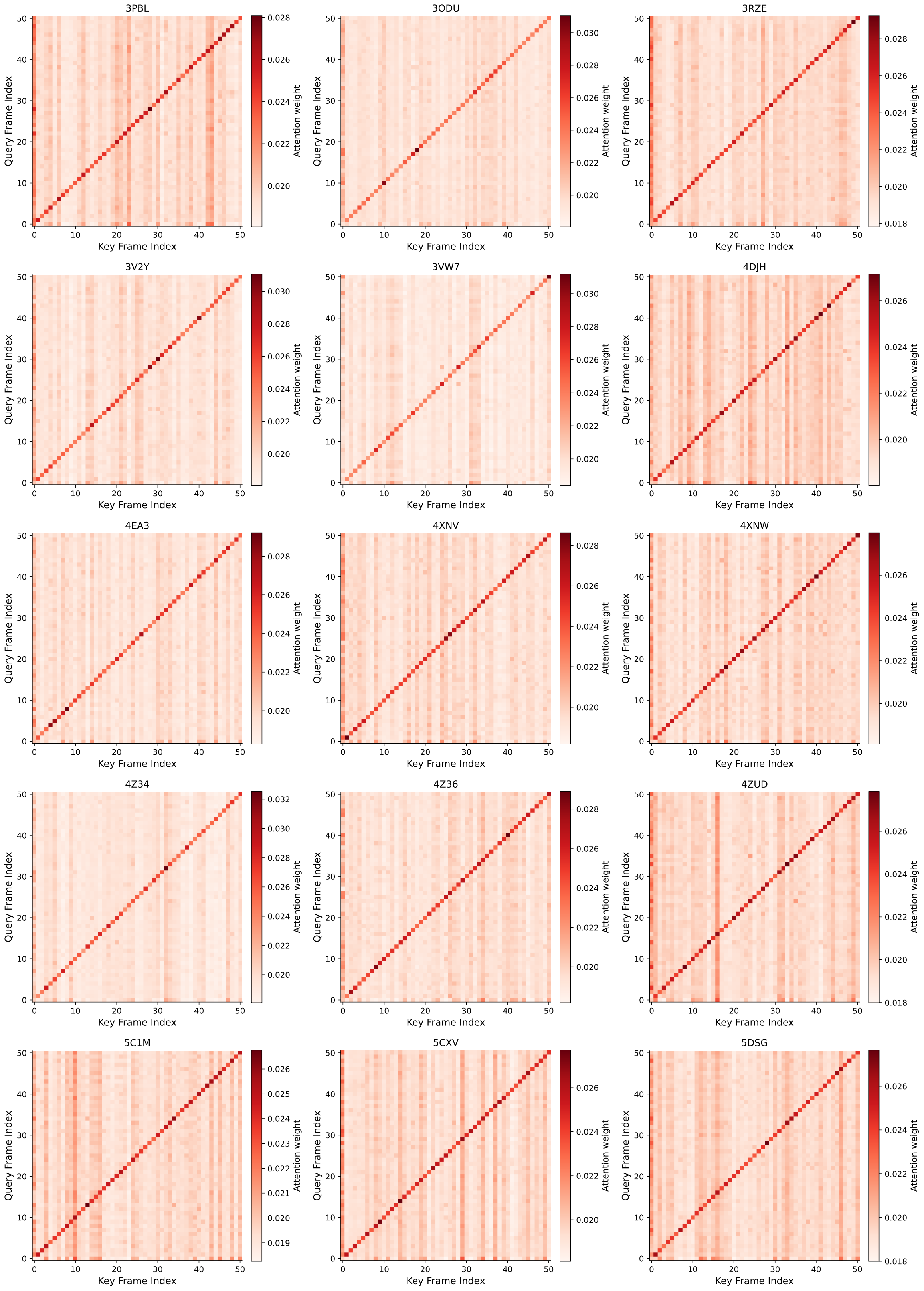

Figure 19. The attention map in the token temporal attention decoder. We focus on the last-layer temporal attention, which captures the most task-relevant and long-range dependencies across frames.