
Supplementary information

**Breaking timescales with generative
sampling of conformational transitions**

In the format provided by the
authors and unedited

Supplementary Information for Breaking Timescales with Generative Sampling of Conformational Transitions

Chenyu Tang^{1*†}, Mayank Prakash Pandey^{1*†}, Cheng Giuseppe Chen¹,
Alberto Megías², François Dehez¹, Christophe Chipot^{1,3,4*}

¹*Université de Lorraine, CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), F-54000 Nancy, France.

²Complex Systems Group and Department of Applied Mathematics, Universidad Politécnica de Madrid, Av. Juan de Herrera 6, E-28040 Madrid, Spain.

³Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, USA.

⁴Theoretical and Computational Biophysics Group, Beckman Institute, and Department of Physics, University of Illinois at Urbana-Champaign, Urbana, USA.

*Corresponding author(s). E-mail(s): chenyu.tang@univ-lorraine.fr;
mayank.pandey@univ-lorraine.fr; chipot@illinois.edu;

Contributing authors: cheng-giuseppe.chen@univ-lorraine.fr;
alberto.megias@upm.es; francois.dehez@univ-lorraine.fr;

[†]These authors contributed equally to this work.

Contents

1	Further Examples and Extended Data	S3
1.1	NANMA	S3
1.2	Trialanine	S4
1.3	Binding upon folding of the ribose-binding protein	S5
1.4	Rotary mechanism of the V-ATPase V_o domain	S7
2	Materials and Methods	S10
2.1	Gen-COMPAS framework	S10
2.2	Denoising diffusion probabilistic model (DDPM)	S12
2.2.1	Forward diffusion process	S12
2.2.2	Reverse generation process	S12
2.2.3	Denoising network architecture	S13
2.2.4	Sampling implementation	S14
2.3	Committor function and selecting transition intermediates	S14
2.3.1	Selection of the transition-state ensemble	S14
2.3.2	Construction of the transition-path ensemble	S15
2.4	Free-energy estimation using RiteWeight	S15
2.5	Convergence of Gen-COMPAS	S16
2.5.1	Convergence based on increased sampling	S16
2.5.2	Convergence and uncertainty analysis of FEL estimations	S18
2.6	Committor Validation Analysis	S22
2.6.1	Committor validation for NANMA	S22
2.6.2	Committor Validation for Trp-cage	S26
2.7	Ensemble-based uncertainty estimation of the committor	S31
2.8	TMD validation and reproducibility	S33
2.8.1	Practical guidance for parameter tuning	S33
2.8.2	Validation of TMD parameters	S33
2.8.3	Reproducibility tests	S35
2.9	Molecular systems and the Gen-COMPAS setup	S36
2.9.1	NANMA isomerization in vacuum	S36
2.9.2	Trialanine conformational equilibrium in vacuum	S36
2.9.3	Ribose-binding protein binding-upon-folding	S36
2.9.4	V-ATPase V_o domain rotary transition	S36
2.9.5	Numerical details	S37

1 Further Examples and Extended Data

Here, we discuss the application of the Gen-COMPAS methodology to a diverse set of molecular systems spanning different levels of structural and dynamical complexity. These include prototypical benchmark systems commonly used to evaluate rare-event sampling methods, such as the isomerization of *N*-acetyl-*N'*-methylalaninamide (NANMA) and the conformational equilibrium of trialanine, as well as more complex biomolecular processes. In particular, we examine the binding-upon-folding mechanism of the ribose-binding protein (RBP) and the rotary transition of the V_o domain of a V-type ATPase.

The small peptide systems serve as controlled benchmark models for validating the accuracy of the Gen-COMPAS framework in reconstructing free-energy landscapes, committor maps, transition-state ensembles, and transition pathways. The larger biomolecular systems, on the other hand, test the scalability and generality of the method in realistic environments involving explicit solvent, large conformational changes, and complex interaction networks. Together, these examples demonstrate the ability of Gen-COMPAS to recover both thermodynamic and kinetic information across a wide range of biomolecular transitions.

1.1 NANMA

NANMA is a classical system widely used to validate new sampling and RC discovery methods, which can be conveniently described by the backbone dihedral angles ϕ and ψ (see Figure S1A), which are determined as the important degrees of freedom¹. With the CHARMM22 force field, the molecule exhibits transitions between the $C7_{eq}$ ($\phi = -85^\circ$, $\psi = 75^\circ$) and $C7_{ax}$ ($\phi = 75^\circ$, $\psi = -75^\circ$) conformations. The Gen-COMPAS framework was applied to investigate this transition at $T = 300$ K, and the resulting committor distribution, TPE, and FEL are shown in Figure S1(B–C). The obtained results are consistent with previous computational studies^{1–7}, demonstrating that Gen-COMPAS reliably captures the expected isomerization behavior of this benchmark system, both for the thermodynamic and kinetic properties.

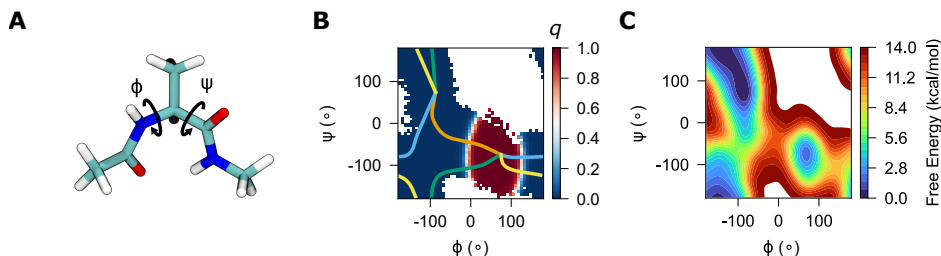


Fig. S1 Gen-COMPAS of NANMA (A) A representative structure of NANMA in vacuum and the two CVs describing the isomerization: dihedral angles ϕ and ψ . (B) The committor and committor-consistent pathways revealing the possible transition pathways by Gen-COMPAS. (C) Free-energy landscape of NANMA isomerization obtained by Gen-COMPAS.

1.2 Trialanine

We also tested the framework on the conformational equilibrium of trialanine in vacuum (Figure S2A), another well-established benchmark system characterized by multiple interconnected pathways between metastable states. The conformational transition is described using the backbone dihedral angles (ϕ_1, ϕ_2, ϕ_3) where the state *A* is defined by ($60^\circ, -70^\circ, 60^\circ$) and the state *B* by ($-70^\circ, 60^\circ, -70^\circ$). The Gen-COMPAS framework, performed at $T = 300$ K, converged within a few iterations, yielding committor distributions, TPE, and FEL (Figure S2B–D) that are in line with earlier computational investigations^{8–10}.

The close agreement with established results for both NANMA and trialanine confirms that Gen-COMPAS performs accurately and robustly on these toy models, validating its reliability for tackling more sophisticated molecular processes involving intricate conformational transitions.

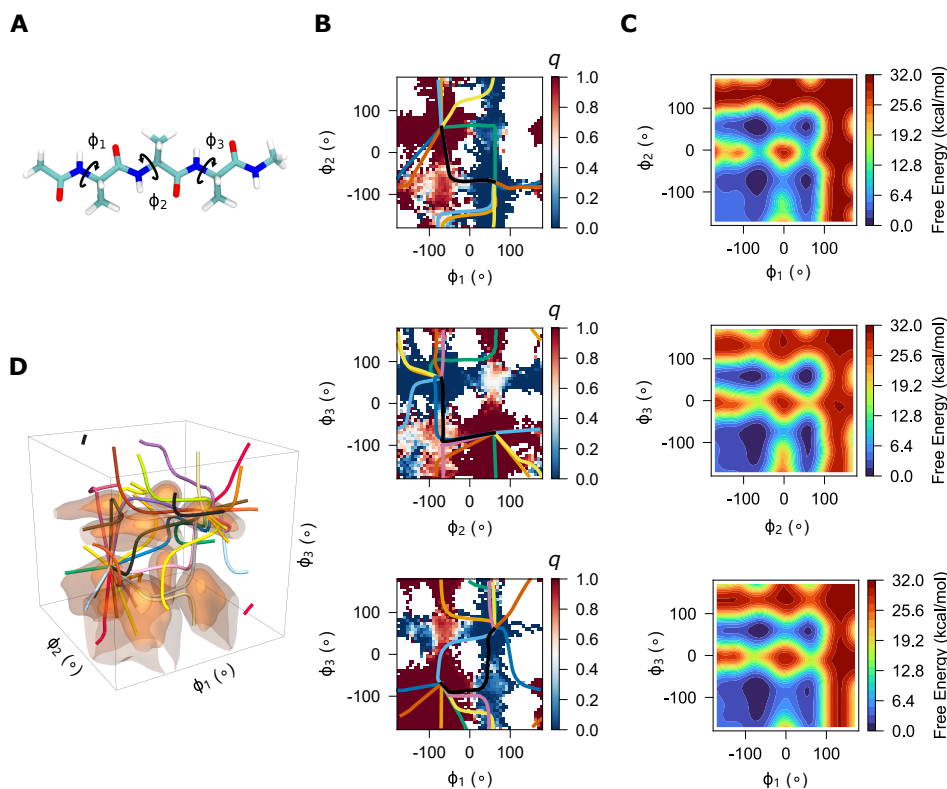


Fig. S2 Gen-COMPAS of Trialanine (A) A representative structure of trialanine in vacuum and the three CVs describing the conformational equilibrium: dihedral angles ϕ_1 , ϕ_2 , and ϕ_3 . (B) The committor and committor-consistent pathways revealing the possible transition pathways by Gen-COMPAS. (C) Free-energy landscapes of trialanine equilibrium obtained by Gen-COMPAS. (D) The committor-consistent pathways and free-energy landscape in 3 dimensions.

1.3 Binding upon folding of the ribose-binding protein

The RBP is a periplasmic binding protein (PBP), essential for bacterial nutrient uptake, mediating high-affinity ribose recognition and delivery. Like other PBPs, the RBP undergoes a large conformational change between open (apo, or ligand-free) and closed (holo, or ligand-bound) forms^{11,12}. Ribose binding is tightly coupled to folding of flexible regions, making the RBP a model for binding-upon-folding processes that illustrate how local disorder drives recognition¹³. Capturing this process is computationally challenging, as it requires simultaneous sampling of large-scale motions, local loop ordering, and ligand docking across a rugged FEL and long timescales.

Gen-COMPAS overcomes these challenges, reconstructing the full binding-upon-folding pathway of the RBP. It captures transitions from the disordered open state, through partially folded intermediates, to the final ribose-bound, closed state (Fig. S3A), revealing the cooperative interplay between ligand binding and protein folding. The binding of ribose to the RBP follows two cooperative pathways (Fig. S3B). In terms of twist-angle rearrangements, both pathways proceed similarly, indicating a conserved motion. On the other hand, for hinge-bending rearrangements, the pathways diverge: one follows a stepwise induced-fit mechanism, whereby ligand binding precedes protein closure^{13,14}, while the other exhibits simultaneous binding and folding.

Quantitatively, Gen-COMPAS directly estimates the committor function to identify the TSE, where ribose is interacting with the RBP and the protein is not yet fully closed (Fig. S3B). Projected FELs along retrospectively defined CVs for ligand position and interdomain angles reveal distinct open and closed basins, separated by barriers consistent with the expected gating motions¹¹ (Fig. S3C). Put together, these analyses unite thermodynamics and kinetics into a coherent mechanistic map of the binding-upon-folding process. By resolving the transition pathways atomistically and quantifying their energetics, Gen-COMPAS offers a general strategy to dissect complex coupled folding-binding events, with broad relevance to other PBPs or intrinsically disordered proteins.

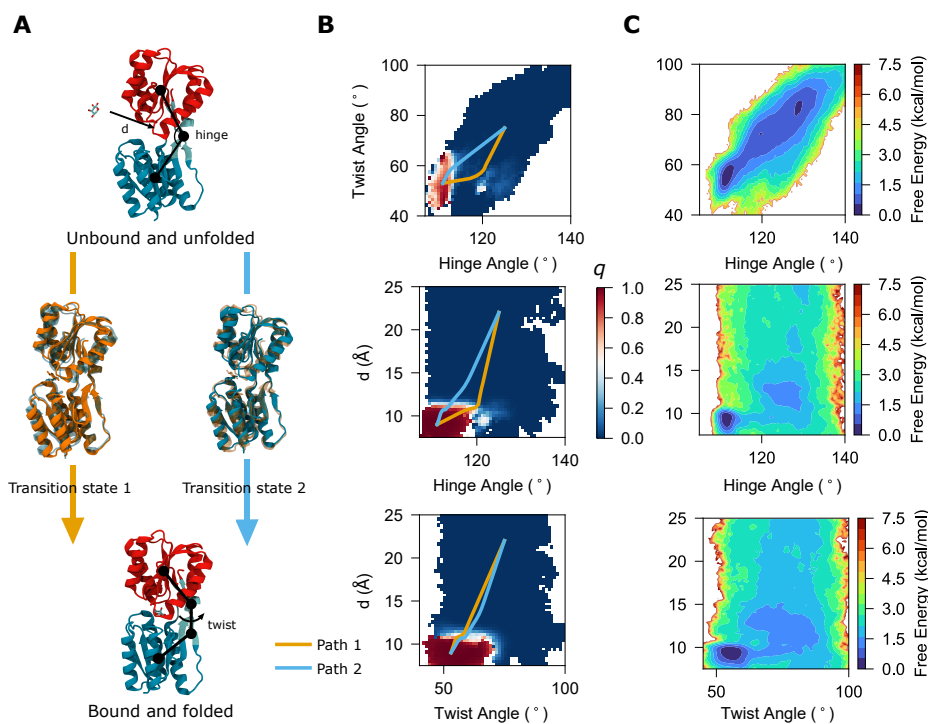


Fig. S3 Gen-COMPAS of RBP binding-upon-folding process (A) Representative structures of RBP-ribose unbound state and bound state, the transition states, and pathways for the RBP binding-upon-folding process. The three collective variables describing the process: d , hinge, and twist¹¹. (B) The committor and committor-consistent pathways revealing the two distinct binding-folding mechanism by Gen-COMPAS. (C) Free-energy landscapes obtained by Gen-COMPAS showing the binding-upon-folding mechanism.

1.4 Rotary mechanism of the V-ATPase V_o domain

Vacuolar ATPases (V-ATPases) are rotary proton pumps that couple ATP hydrolysis in the soluble V_1 motor to proton translocation through the membrane-embedded V_o domain. The V_o sector contains the c-ring rotor and the a-subunit stator, whose relative motion drives proton transport across the membrane. During operation, conserved acidic residues on the c-subunits interact with charged residues on the C-terminal domain of the a-subunit, forming and breaking salt bridges and hydrogen bonds that gate proton transfer and stabilize successive configurations of the rotor–stator interface. Structural and computational studies have shown that this process proceeds through a sequence of metastable intermediates that partition the rotation of the c-ring into discrete steps, reflecting the coordinated rearrangement of electrostatic and hydrogen-bonding interactions at the interface¹⁵. Capturing these transitions computationally is challenging because the relevant rearrangements involve both the collective motion of the rotor and the localized restructuring of interfacial contact networks.

Gen-COMPAS elucidates these transitions by reconstructing the rotary pathway of the V-ATPase V_o domain using structural descriptors that directly monitor the rearrangement of key rotor–stator interactions. Instead of imposing an explicit rotational coordinate, we characterize the transition between two endpoint states (A and B), corresponding to adjacent rotational configurations of the rotor, using the RMSDs of interfacial residues involved in the electrostatic network. Specifically, we track the heavy-atom RMSDs of E108 of c'' , E137 of $c(1)$, and E789, R735, and R799 of the a_{CT} domain relative to the endpoint structures. These descriptors report on the formation and disruption of salt bridges and hydrogen-bond networks between the c-ring and the a-subunit, which define the mechanistic progression of the rotary step.

Using these variables, Gen-COMPAS reconstructs the transition pathway connecting the two rotational states and identifies three metastable intermediates along the reaction coordinate (Fig. S4A). These intermediates correspond to configurations wherein distinct subsets of rotor–stator contacts are stabilized, consistent with the electrostatic gating mechanism proposed in previous experimental and computational studies¹⁵. The reconstructed pathway reveals a sequential rearrangement of the interaction network: contacts associated with the initial state gradually weaken while new interactions characteristic of the subsequent rotational state form, stabilizing intermediate conformations of the c-ring relative to the a-subunit.

Quantitatively, the FEL determined by Gen-COMPAS along the RMSD descriptors reveals multiple metastable basins separated by barriers that correspond to successive rearrangements of the rotor–stator interface (Fig. S4B). The dominant barrier is estimated to be approximately 5–6 kcal/mol, while the end-state free-energy difference remains small, $\Delta G_{AB} < 1$ kcal/mol. This is semi-quantitatively consistent with the conventional-sampling study of Roh et al.¹⁵, which reported a main barrier of approximately 8 kcal/mol and likewise found $\Delta G_{AB} < 1$ kcal/mol for the same V_o rotary step. Thus, both calculations support a nearly isoenergetic adjacent-state transition separated by a moderate rotary barrier.

We utilized the committor function along this transition to identify the transition pathway and transition-state ensemble (Fig. S4C). Configurations lying near different metastable states along the pathway correspond to partially reorganized hydrogen-bond and salt-bridge networks (Fig. S4D), reflecting the cooperative exchange of residue contacts that drives the rotary step. One quantitative difference from Roh et al. is that the basin labeled M3 in our FEL lies close in free energy to the initial state A, whereas the analogous region in the previous calculation is approximately 5 kcal/mol above A. This difference, together with the lower apparent barrier in

the Gen-COMPAS profile, may reflect the high degeneracy of the projected RMSD space and the presence of multiple reaction channels. For example, different combinations of interfacial hydrogen bonds and salt bridges may project onto similar RMSD values while corresponding to partially distinct microscopic pathways. Because Gen-COMPAS samples an ensemble of generated transition pathways, these channels may overlap in the low-dimensional projection, effectively averaging over several routes and stabilizing intermediate basins such as M3.

The comparison with the F-type ATP synthase c-ring study of Blanc et al.¹⁶ should be interpreted more qualitatively. Their work considers a different rotary motor and constructs free-energy profiles along the c-ring rotation angle for different protonation states, explicitly linking the rotational landscape to proton transfer. Unlike the isolated membrane sector, F_o , of an F-type ATP synthase, which acts as a passive proton pore, a free V_o domain, in its auto-inhibited state, does not conduct protons along a pH gradient under physiological conditions. Nevertheless, it provides a useful point of comparison because both systems exhibit discrete metastable rotary substates stabilized by interfacial electrostatic interactions, and both highlight the importance of coupled rearrangements between the rotor and stator during directional rotation. In contrast to the F-type system, where protonation-state changes and water-wire-mediated proton transfer play a central role in reshaping the free-energy profile, the present V_o calculation focuses on the structural rotary transition between adjacent states defined by the rotor–stator contact network. We therefore do not treat the absolute free energies from the two systems as directly interchangeable benchmarks. Instead, the comparison emphasizes a shared physical motif: rotary ATP-ase or synthase motors can generate stepped FELs through sequential, interface-localized rearrangements, while the detailed thermodynamic and kinetic landscapes remain system-specific.

Together, these analyses provide a consistent thermodynamic and kinetic description of the V_o rotary mechanism. By resolving the sequence of salt-bridge and hydrogen-bond rearrangements and the associated metastable states, Gen-COMPAS reproduces the key mechanistic features previously inferred from conventional enhanced-sampling approaches. This comparison also motivates future, more exhaustive sampling of the complete rotation cycle, where pathway-resolved analyses may clarify how multiple microscopic contact-rearrangement states contribute to the apparent free-energy profile of rotary molecular motors.

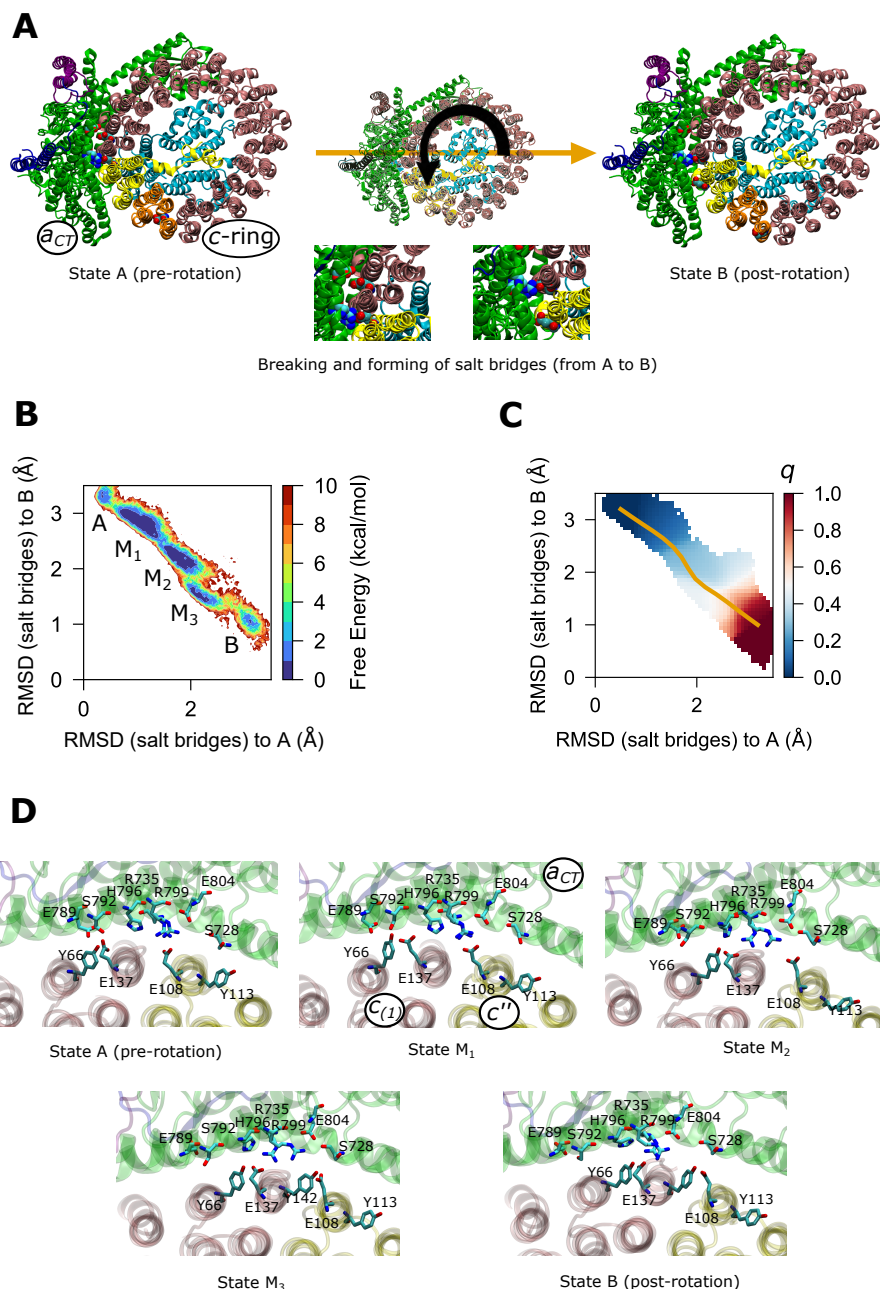


Fig. S4 Gen-COMPAS of V-ATPase V_0 domain rotation (A) Representative structures and rotation of the V-ATPase V_0 , and the forming and breaking of salt bridges controlling the transition. (B) Free-energy landscapes obtained by Gen-COMPAS showing the 5 metastable states. (C) The committor and committor-consistent pathways revealing the rotary mechanism by Gen-COMPAS. (D) Structures extracted from the Gen-COMPAS simulations along the rotary pathway, and some of the polar residues related to the rotation near the a_{CT} :c-ring interface.

2 Materials and Methods

2.1 Gen-COMPAS framework

Algorithm S1 summarizes the iterative Gen-COMPAS workflow used to sample the transition region between two metastable states A and B . The procedure begins with short unbiased molecular dynamics simulations initiated from the endpoint states to generate an initial dataset. A diffusion-based generative model is trained on these configurations and used to propose candidate intermediate conformations connecting the two states.

Because a committor model is not yet available during the first iteration, the generated intermediates are clustered to obtain a representative set of target structures. Targeted MD (TMD) simulations are then performed from both A and B toward these targets, converging to physically accessible regions of conformational space. Unbiased MD trajectories launched from these configurations provide the first sampling of the transition region and constitute the initial training data for the next iteration.

From the second iteration onward, the accumulated simulation data are used to train both the generative model and a committor predictor that estimates the probability of reaching state B before state A . Generated structures with committor values close to $q \approx 1/2$ are selected as targets near the separatrix. TMD simulations refine these structures, and unbiased trajectories initiated from the resulting configurations further enrich the dataset. Repeating this feedback loop progressively focuses sampling around the transition region and yields sufficient data to identify transition states, reconstruct transition pathways, and approximate the underlying FELs projected onto a posteriori collective variables (CVs).

Algorithm S1 Iterative Gen-COMPAS workflow for sampling transition regions.

Require: Metastable states A (reactant) and B (product)

Ensure: Transition-state ensemble, committor map, transition pathways, approximate FEL

- 1: Run short unbiased MD simulations (1–2 ns) starting from A and B
 - 2: Collect configurations into initial dataset $\mathcal{D}_0$
 - 3: Train an initial diffusion generative model on $\mathcal{D}_0$
 - 4: Generate candidate intermediate conformations $\{\mathbf{x}_i\}$ connecting A and B
 - 5: Cluster the generated conformations to obtain representative targets $\mathcal{T}_0$
 - 6: **for all** $\mathbf{x}^* \in \mathcal{T}_0$ **do**
 - 7: Run targeted MD (TMD) from A toward $\mathbf{x}^*$
 - 8: Run targeted MD (TMD) from B toward $\mathbf{x}^*$
 - 9: Identify configurations converging to the physical neighborhood of $\mathbf{x}^*$
 - 10: **end for**
 - 11: Launch unbiased MD trajectories from the converged configurations
 - 12: Riteweight: collect new trajectories into frames $\mathcal{D}_1$ and weights $\mathcal{W}_1$
 - 13: $k \leftarrow 1$
 - 14: **while** convergence not reached **do**
 - 15: Train diffusion generative model on $\mathcal{D}_k$
 - 16: Train committor predictor $q_k(\mathbf{x})$ on $\mathcal{D}_k$ and $\mathcal{W}_k$
 - 17: Generate candidate intermediate conformations $\{\mathbf{x}_i\}$ connecting A and B
 - 18: Evaluate committor values $q_k(\mathbf{x}_i)$
 - 19: Select targets $\mathcal{T}_k = \{\mathbf{x}_i \mid q_k(\mathbf{x}_i) \approx 1/2\}$ near the separatrix
 - 20: **for all** $\mathbf{x}^* \in \mathcal{T}_k$ **do**
 - 21: Run targeted MD (TMD) from A toward $\mathbf{x}^*$
 - 22: Run targeted MD (TMD) from B toward $\mathbf{x}^*$
 - 23: Identify configurations converging to the physical neighborhood of $\mathbf{x}^*$
 - 24: **end for**
 - 25: Launch unbiased MD trajectories from configurations near the separatrix
 - 26: Riteweight: collect new trajectories into frames $\Delta\mathcal{D}_k$ and weights $\Delta\mathcal{W}_k$
 - 27: $\mathcal{D}_{k+1} \leftarrow \mathcal{D}_k \cup \Delta\mathcal{D}_k$, $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \cup \Delta\mathcal{W}_k$
 - 28: $k \leftarrow k + 1$
 - 29: **end while**
 - 30: Identify transition-state ensemble from structures with $q(\mathbf{x}) \approx 1/2$
 - 31: Project committor onto interpretable CVs defined post-simulation
 - 32: Extract candidate transition pathways connecting A and B
 - 33: Approximate the FELs from cumulative sampling
-

2.2 Denoising diffusion probabilistic model (DDPM)

2.2.1 Forward diffusion process

The forward-diffusion process gradually adds noise to a native protein structure over a series of time steps, $t = 1, \dots, t_T$, following a variance-preserving schedule. Given a set of Cartesian coordinates, i.e., data distribution, $\mathbf{x}_0 \sim p(\mathbf{x}_0)$, a forward Markov process is defined to generate random variables, i.e., random sets of coordinates, $\mathbf{x}_1, \dots, \mathbf{x}_{t_T}$, according to the transition kernels $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, such that,

$$p(\mathbf{x}_1, \dots, \mathbf{x}_{t_T}|\mathbf{x}_0) = \prod_{t=1}^{t_T} p(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \quad (\text{S1})$$

where $\beta_t \in (0, 1)$ for $t = 1, \dots, t_T$ are hyperparameters setting the variances of the Gaussian-noise addition process, i.e., the variance schedule. From a practical point of view, the noised state at time step t is defined as,

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_{t-1} + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \quad (\text{S2})$$

where $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$ is the cumulative product of the noise schedule $\alpha_t = 1 - \beta_t$. To maintain translational equivariance, all coordinates are centered by subtracting their center of mass before the noise is applied. We employed a cosine-based variance schedule for β_t , which we found to yield stable and efficient training.

2.2.2 Reverse generation process

Structure generation is performed by a neural network (NN) with parameters θ that learns to reverse the noising process. The reverse process starts from pure noise $\mathbf{x}_{t_T} \sim \mathcal{N}(0, \mathbf{I})$, i.e., $p(\mathbf{x}_{t_T}) = \mathcal{N}(\mathbf{x}_{t_T}; 0, \mathbf{I})$, and the model iteratively denoises the structure over t_T steps. At each reverse step, the model, conditioned on the current time step t , predicts the denoised structure according to $p_\theta(\mathbf{x}_t|\mathbf{x}_{t+1})$. This is characterized, again, by another Markov process, such that,

$$p_\theta(\mathbf{x}_0, \dots, \mathbf{x}_{t_T}|\mathbf{x}_0) = p(\mathbf{x}_{t_T}) \prod_{t=1}^{t_T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t),$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)). \quad (\text{S3})$$

Here, we use the standard criteria $\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t) = \sigma_t^2\mathbf{I}$, where $\sigma_t^2 = \beta_t$, and

$$\boldsymbol{\mu}_\theta = \frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t} \beta_t f_\theta(\mathbf{x}_t, t) + \frac{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t, \quad (\text{S4})$$

such that $\hat{\mathbf{x}}_0 = f_\theta(\mathbf{x}_t, t)$ is a prediction of the initial structure, with f_θ the generative function. During the denoising process, the aim is to learn f_θ according to the prediction of the initial structure imposing the recovery of the ground truth.

Additionally, a scale factor on σ_t is considered for stochastic sampling, enabling deterministic (DDIM-like¹⁷) sampling when set to zero. Our generation pipeline uses progressively reduced noise scaling across iterative refinement cycles—initial rounds employ high noise scales (see table S3) to encourage structural diversity, while later rounds progressively reduce noise scales toward zero, producing increasingly deterministic, high-accuracy structures after model retraining and physical refinement.

The generative function f_θ was trained to minimize the mean squared error between the predicted clean coordinates, $\hat{\mathbf{x}}_0$, and the true native structure, $\mathbf{x}_0$, from the training dataset, encoded in the following loss function,

$$\mathcal{L}(\theta) = \mathbb{E}_{t, \mathbf{x}_0 \sim p(\mathbf{x}_0), \epsilon \sim \mathcal{N}(0, \mathbf{I})} \left[\|f_\theta(\sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1 - \alpha_t}\epsilon, t) - \mathbf{x}_0\|_2^2 \right], \quad (\text{S5})$$

where $\|\cdot\|_2$ refers to the Euclidean norm.

2.2.3 Denoising network architecture

The core of the reverse process is designed to be driven by a hierarchical graph neural network (GNN), which predicts $\hat{\mathbf{x}}_0$ from $\hat{\mathbf{x}}_t$ and t . The process is based on different steps that are described as follows: First, initial node features $\mathbf{h}^{(0)} \in \mathbb{R}^{B \times N \times D}$ —where B , N and D represent the batch size, the number of nodes and the embedding dimension, respectively—are constructed by summing learned embeddings for atom-type and residue identity, $\text{Embed}_{\text{res}}$ (residue ID) and $\text{Embed}_{\text{atom}}$ (atom type), with a linear projection of the input coordinates $\hat{\mathbf{x}}_t$, via the action of the matrix $\mathbf{W}_{\text{coord}}$. This combined result is then passed through a so-called layer normalization operation. Subsequently, a sinusoidal embedding of time step t , $\text{Embed}(t)$, is projected through a multilayer perceptron (MLP) at time step t , MLP_t , and added to the normalized features, $\mathbf{h}'$, to create the final input features, $\mathbf{h}^{(0)}$, conditioning the network on the noise level. This two-step transformation can be expressed as follows,

$$\mathbf{h}' = \text{LayerNorm}(\text{Embed}_{\text{res}}(\text{residue ID}) + \text{Embed}_{\text{atom}}(\text{atom type}) + \mathbf{W}_{\text{coord}}\hat{\mathbf{x}}_t), \quad (\text{S6a})$$

$$\mathbf{h}^{(0)} = \mathbf{h}' + \text{MLP}_t(\text{Embed}(t)). \quad (\text{S6b})$$

Next, a dynamic k-nearest neighbor (k-NN) graph, $\mathcal{G}_{\text{kNN}}$, is constructed based on the current coordinates $\hat{\mathbf{x}}_t$. This graph is combined with a static graph of covalent bonds, $\mathcal{G}_{\text{bonds}}$, from the molecular topology, to define the complete interaction network, $\mathcal{G}_{\text{kNN}} \cup \mathcal{G}_{\text{bonds}}$.

For heterogeneous systems, including ligand–protein complexes, all selected atoms from the different molecular components are included in the same coordinate tensor and graph representation. Thus, ligands are not generated separately and then docked into the protein; ligand and protein coordinates are denoised jointly. The dynamic k-NN graph captures local nonbonded contacts across molecular components, while the residue-level attention layers capture longer-range conformational coupling. This joint representation is important for systems such as nAChR, where ligand occupancy and receptor rearrangement are coupled.

For the propagation of information, a hierarchical message-passing policy is used. The model utilizes a dual-level—atom and residue—processing architecture. For the atom-level processing, continuous-filter convolutional layers (SchNet-style)¹⁸ update node features by aggregating messages from neighbors. The message from atom j to atom i is modulated by a

filter network through a MLP, $\text{MLP}_{\text{filter}}$, operating on the interatomic distance d_{ij} as follows,

$$\mathbf{h}_i^{(l+1)} = \text{SchNetLayer}\left(\mathbf{h}_i^{(l)}, \{\mathbf{h}_j^{(l)} \cdot \text{MLP}_{\text{filter}}(d_{ij}) \mid j \in \mathcal{N}(i)\}\right), \quad (\text{S7})$$

where $\mathcal{N}(i)$ denotes the neighbor atom indices to atom i . Moreover, for the residue-level processing, atom features within each residue, r , are averaged to create residue-level features, $\mathbf{h}_r = \frac{1}{|r|} \sum_{i \in r} \mathbf{h}_i$. These features are processed through multi-head self-attention layers¹⁹ to capture long-range interactions,

$$\mathbf{h}_r^{\text{attn}} = \text{MHA}(\mathbf{h}_r, \mathbf{h}_r, \mathbf{h}_r), \quad (\text{S8})$$

The updated residue features are then mapped back to atom features and combined with the existing atom-level features, yielding a final set of refined features, $\mathbf{h}^{(\text{final})}$. The latter set is passed through a final MLP to predict the clean, centered coordinates,

$$\hat{\mathbf{x}}_0 = \text{MLP}_{\text{out}}(\mathbf{h}^{(\text{final})}). \quad (\text{S9})$$

2.2.4 Sampling implementation

For inference, we initialize a set of random Gaussian noises, i.e., $\mathcal{N}(0, \mathbf{I})$, with dimensions matching the target molecular system. The trained model is applied iteratively over the defined time steps to denoise this initial state into a final, coherent molecular structure. The resulting Cartesian coordinates are restored to physical space by reversing the centering and normalization steps (during the forward diffusion process). All generated structures are output in a PDB format towards subsequent analysis.

2.3 Commitor function and selecting transition intermediates

2.3.1 Selection of the transition-state ensemble

Among all isocommittor surfaces, the hypersurface defined by $q = 1/2$, known as the separatrix, plays a particular role. The set of conformations lying on the separatrix represents the collection of states from which the system has equal probability to commit to A or B . It, therefore, constitutes an optimal dynamical dividing surface between the two basins of interest. Within TPT, this separatrix provides a rigorous definition of the TSE: The set of critical molecular structures that mediate the transformation between the reactant and the product. Identifying the TSE is central to understanding molecular reactivity, since these conformations encode the essential structural motifs of the barrier-crossing event.

In practice, the Gen-COMPAS framework identifies the TSE in two stages. First, conformations consistent with the separatrix are generated, ensuring statistical fidelity to the $q = 1/2$ hypersurface within a neighborhood $q \in (1/2 - \varepsilon, 1/2 + \varepsilon)$, with $2\varepsilon < 1$. Second, these conformations are refined using the TMD procedure, which extracts a physically interpretable subset of representative structures. This iterative approach converges toward a meaningful reconstruction of the TSE, yielding direct insight into the microscopic structures responsible for enabling the transition.

2.3.2 Construction of the transition-path ensemble

While the TSE provides a structural snapshot of the barrier-crossing event, a complete mechanistic picture requires knowledge of the transition pathways themselves. The set of all such reactive trajectories defines the TPE, which contains richer dynamical information than the TSE—it does not only identify the critical transition structures, but also captures the diversity of pathways connecting A to B , including possible intermediate metastable states, parallel routes, and dominant mechanisms.

Since Gen-COMPAS simultaneously estimates the committor and the TSE, it naturally provides the ingredients to reconstruct the TPE. Following recent methodologies²⁰, reactive trajectories can be generated by integrating along the gradient of the committor, starting from representative TSE conformations. This procedure yields a collection of committor-consistent trajectories, each corresponding to a possible reactive pathway.

To interpret the diversity of these trajectories, clustering techniques can be applied to group them into a smaller set of dominant pathways²⁰. This reduction provides a compact, albeit comprehensive description of the TPE, highlighting the main representative reaction mechanisms without losing essential dynamical richness.

2.4 Free-energy estimation using RiteWeight

The trajectories generated by the Gen-COMPAS framework originate from ensembles that are not sampled from the equilibrium stationary distribution. In particular, many trajectories are initiated from TMD endpoints or from structures selected near the separatrix. Consequently, the resulting ensemble of configurations is biased toward the transition region and does not directly obey the Boltzmann distribution. To recover thermodynamic observables, including FELs, it is necessary to reconstruct the stationary probability distribution associated with the underlying unbiased dynamics.

To this end, we employ the Randomized Iterative Trajectory Reweighting (RiteWeight) algorithm²¹. RiteWeight estimates the stationary distribution from collections of trajectories that may be far from equilibrium sampling by iteratively reweighting trajectory segments based on kinetic information extracted from the simulation data. The method relies on transition pairs (or trajectory fragments), defined as two points in configurational space separated by a fixed time lag, τ , which encode the unbiased dynamical transitions. These transition pairs form the statistical foundation used to infer the steady-state distribution of configurations. Importantly, the algorithm can utilize trajectories of arbitrary length, including very short fragments, and does not require local equilibration within discrete states.

Operationally, RiteWeight constructs a sequence of discrete-state transition models using repeated random clusterings of configuration space. At iteration k , the configurations are partitioned into clusters $\{C_i^{(k)}\}$, which define a discrete transition matrix $T^{(k)}$ computed from the weighted transition pairs between clusters. The stationary distribution $\pi^{(k)}$ associated with this transition matrix is then obtained from the corresponding eigenvector equation,

$$\pi^{(k)} = \pi^{(k)} T^{(k)}, \quad \sum_i \pi_i^{(k)} = 1. \quad (\text{S10})$$

The stationary probabilities $\pi_i^{(k)}$ represent the equilibrium weights of the clusters for the given discretization. These cluster weights are then used to update the weights of the trajectory fragments. In the simplest formulation of RiteWeight, the updated weight of a fragment originating from cluster $C_i^{(k)}$ is obtained by rescaling the total weight of fragments within that cluster so that it matches $\pi_i^{(k)}$, while preserving the relative weights of fragments within the cluster.

The defining feature of the algorithm is that a new random clustering is generated at every iteration. This repeated randomization allows the algorithm to progressively adjust the weights of individual configurations across different cluster partitions, effectively refining the stationary distribution beyond the resolution of any single discretization. The iterative procedure converges toward a self-consistent stationary distribution that satisfies the transition statistics of the trajectory data. Because cluster boundaries are repeatedly randomized, the resulting probability density approaches a quasi-continuous estimate of the stationary distribution in configuration space, thereby minimizing the discretization bias typically associated with Markov state models (MSMs) constructed from a single clustering.

Once converged weights $w(x)$ are obtained for all sampled configurations x , thermodynamic observables can be estimated directly. Specifically, the probability density, $P(\xi)$, along any CV, $\xi(x)$, is computed from the weighted ensemble,

$$P(\xi) \propto \sum_{x \in \mathcal{D}} w(x) \delta(\xi(x) - \xi), \quad (\text{S11})$$

from which the free energy, $F(\xi)$, is determined,

$$F(\xi) = -k_B T \ln P(\xi) + C. \quad (\text{S12})$$

This approach is particularly well suited for Gen-COMPAS simulations because the framework generates an extensive number of short unbiased trajectories initiated from non-equilibrium ensembles. RiteWeight leverages the kinetic information contained in these trajectories to reconstruct the stationary distribution without requiring prohibitively long equilibrium simulations. As a result, the method enables powerful reweighted estimates of FELs and other equilibrium observables from the aggregated trajectory ensemble produced by the Gen-COMPAS workflow. The global accuracy of these estimates still depends on trajectory connectivity and sampling coverage, but the resulting FELs are sufficient for identifying stable local basins, transition regions, and pathway-level free-energy organization when the degradation analysis remains within a reasonable range.

2.5 Convergence of Gen-COMPAS

2.5.1 Convergence based on increased sampling

To evaluate convergence of both the committor function and the sampling of the TSE, we have analyzed the evolution of the k-distance distribution of the structures sampled from the TSE over successive Gen-COMPAS iterations.

At each iteration, all accumulated structures are combined, and the pairwise RMSDs between conformations are computed after Kabsch alignment²². For each structure, we calculate the average RMSD to its k -nearest neighbors (k -distance), yielding a k -distance distribution that quantifies the local density of sampling in the conformational space. Smaller k -distances correspond to densely explored regions, while larger values indicate the discovery of new conformations.

Convergence is assessed by comparing k -distance distributions between consecutive iterations using the Wasserstein distance of order 1²³, which measures how much the structural distribution changes as sampling is accrued. A large Wasserstein distance implies that the current iteration explores previously unexplored regions of conformational space, while a small and stable distance indicates that new samples are structurally similar to those already obtained. In this particular case, let us denote $P_{\ell-1}$ the empirical distribution of k -distances of the $(\ell - 1)$ -th iteration with observations $d_1^{(\ell-1)}, \dots, d_n^{(\ell-1)}$ and P_ℓ of the ℓ -th iteration with measurements of the k -distances $d_1^{(\ell)}, \dots, d_n^{(\ell)}$. Then, the Wasserstein distance of order 1 defined on the empirical distributions of k -distances between consecutive iterations is defined as,

$$W_1^\ell = W_1(P_{\ell-1}, P_\ell) = \inf_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^n \left| d_i^{(\ell-1)} - d_{\sigma(i)}^{(\ell)} \right|, \quad (\text{S13})$$

where S_n denotes the set of permutations of n -elements, where P_0 is the distribution deduced from clustering during the 0-th iteration of the Gen-COMPAS framework.

Once the Wasserstein distance between successive k -distance distributions falls below a predefined threshold and remains stable, convergence is considered to have been attained. This condition signifies that the TSE reaches a steady structural distribution. As a consequence, the committor function constructed from this distribution achieves statistical convergence. For all the systems investigated here, we performed at least three iterations from which we check the convergence, setting the threshold to 1.0. The Wasserstein distances for each iteration of the systems examined in this work are shown in Table S1.

Iterations	1 st	2 nd	3 rd	4 th
NANMA isomerization	8.03	6.38	1.63	0.33
Trialanine equilibrium	7.53	2.99	0.42	–
Trp-cage folding	2.20	0.19	0.21	–
RBP binding-upon-folding	0.29	0.12	0.14	–
Apo-state AAC conformational change	2.04	0.84	0.33	–
ADP-bound AAC conformational change	3.31	1.38	0.83	–
Vo domain rotation	1.77	0.58	0.49	–
nAChR activation	2.02	1.26	0.25	–

Table S1 Convergence of Gen-COMPAS: W_1^ℓ of the TSE after each iteration.

2.5.2 Convergence and uncertainty analysis of FEL estimations

To evaluate the robustness of the FELs obtained from the Gen-COMPAS simulations, we performed systematic convergence and uncertainty analyses employing trajectory-level bootstrapping. Because the free-energy estimation relies on trajectory ensembles generated from non-equilibrium initial conditions that are subsequently reweighted using RiteWeight²¹, it is essential to verify that the resulting FELs are statistically stable and converged with respect to sample size.

Bootstrap estimation of statistical uncertainty.

We estimated statistical uncertainties in the FEL using a bootstrap resampling procedure performed at the trajectory level. Let $\{\mathcal{T}_i\}_{i=1}^N$ denote the set of N trajectories generated during the Gen-COMPAS sampling procedure. Each trajectory contributes a set of configurations x with their associated RiteWeight weights $w(x)$. For a given bootstrap replicate, we construct a resampled trajectory ensemble by drawing N trajectories with replacement from the original set. The configurations and weights belonging to the selected trajectories are concatenated to form a new bootstrap dataset for subsequent free-energy estimation.

For each bootstrap sample, we compute the probability distribution along the chosen CVs, ξ , using a weighted histogram,

$$P_b(\xi) = \frac{1}{Z_b} \sum_{x \in \mathcal{D}_b} w(x) \delta(\xi(x) - \xi), \quad (\text{S14})$$

where $\mathcal{D}_b$ denotes the bootstrap dataset and Z_b is a normalization constant. The corresponding free-energy surface is obtained as

$$F_b(\xi) = -k_B T \ln P_b(\xi) + C. \quad (\text{S15})$$

Repeating this procedure over many bootstrap replicates yields a distribution of FEL estimates for each bin of the CV space. From these samples we compute the standard deviation for each grid point. (see Fig. S5) This check provides a direct estimate of the statistical uncertainty associated with the FEL reconstructed from the Gen-COMPAS trajectory ensemble using RiteWeight.

Convergence with respect to trajectory accumulation.

In addition to estimating statistical uncertainty, we also evaluated the convergence of the FEL as a function of the number of trajectories included in the analysis. Specifically, we constructed a reference FEL $F_{\text{final}}(\xi)$ using the full trajectory dataset. We then generated multiple random permutations of the trajectory set and progressively accumulated trajectories following each permutation order. After including the first k trajectories in a given permutation, we recomputed the FEL $F_k(\xi)$ using the corresponding weighted configurations.

To quantify the difference between the partial and final estimates, we computed the root-mean-square error (RMSE)

$$\text{RMSE}(k) = \sqrt{\left\langle [F_k(\xi) - F_{\text{final}}(\xi)]^2 \right\rangle_{\Omega}}, \quad (\text{S16})$$

where the average $\langle \cdot \rangle_{\Omega}$ is taken over bins with probability density above a small threshold relative to the maximum probability. This masking procedure restricts the comparison to regions of configuration space that are sufficiently sampled.

Repeating this trajectory accumulation procedure for multiple random permutations yields a mean convergence curve and associated standard deviation as a function of the number of trajectories. The resulting curves show that the FEL stabilizes as additional trajectories are included, with the RMSE approaching a plateau once sufficient sampling of the transition region is achieved.

Practical implementation.

In practice, the analysis is performed using the RiteWeight-derived frame weights associated with all unbiased shooting trajectories generated by Gen-COMPAS. For each trajectory, the corresponding CVs and frame weights are reconstructed from the simulation outputs and used to build weighted histograms in one or two dimensions. The FEL is computed as,

$$F(\boldsymbol{\xi}) = -k_{\text{B}}T \ln P(\boldsymbol{\xi}), \quad (\text{S17})$$

optionally followed by Gaussian smoothing in CV space to reduce discretization noise.

The bootstrap procedure typically employs hundreds of trajectory resamples, while the convergence analysis averages over multiple random permutations of the trajectory order. Together, these analyses provide quantitative estimates of both the statistical uncertainty and the convergence behavior of the reconstructed FELs. The resulting confidence intervals and convergence curves quantify statistical stabilization with respect to the available Gen-COMPAS sampling. For the larger systems, not every curve reaches the strictest plateau expected of a fully converged global equilibrium FEL; nevertheless, the residual degradation relative to the final estimate remains within a reasonable range for mechanistic interpretation, and the local basin topology and transition regions are stable in the uncertainty check. We therefore use these FELs as powerful mechanistic maps. As we would strongly suggest, treating absolute global free-energy differences between weakly connected basins should be cautious. Crucially, both uncertainty and convergence checks were performed for every system examined herein (see Fig.S5 and Fig.S6).

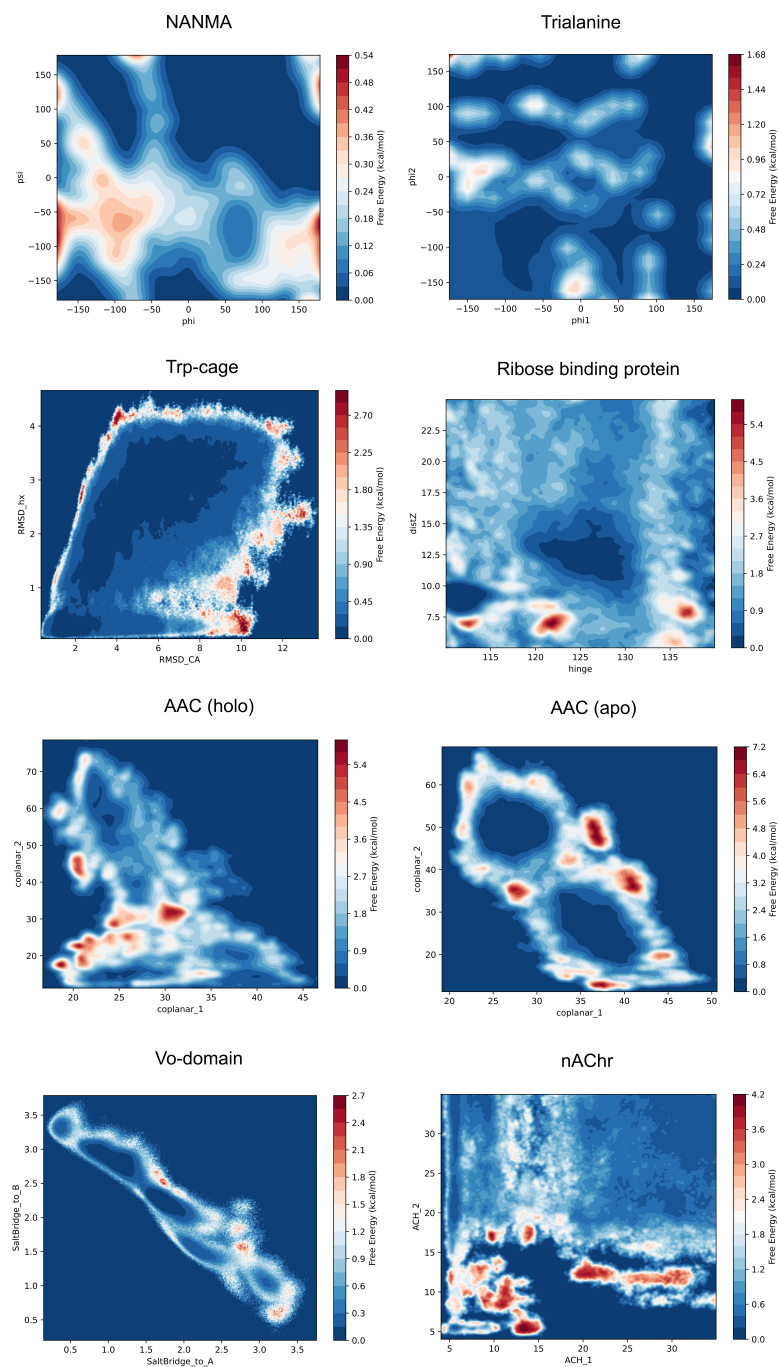


Fig. S5 Standard deviation of FEL estimation The uncertainty estimation of the FEL estimation from GenCOMPAS by checking the standard deviation of the weighted histogram projected on two dimensions.

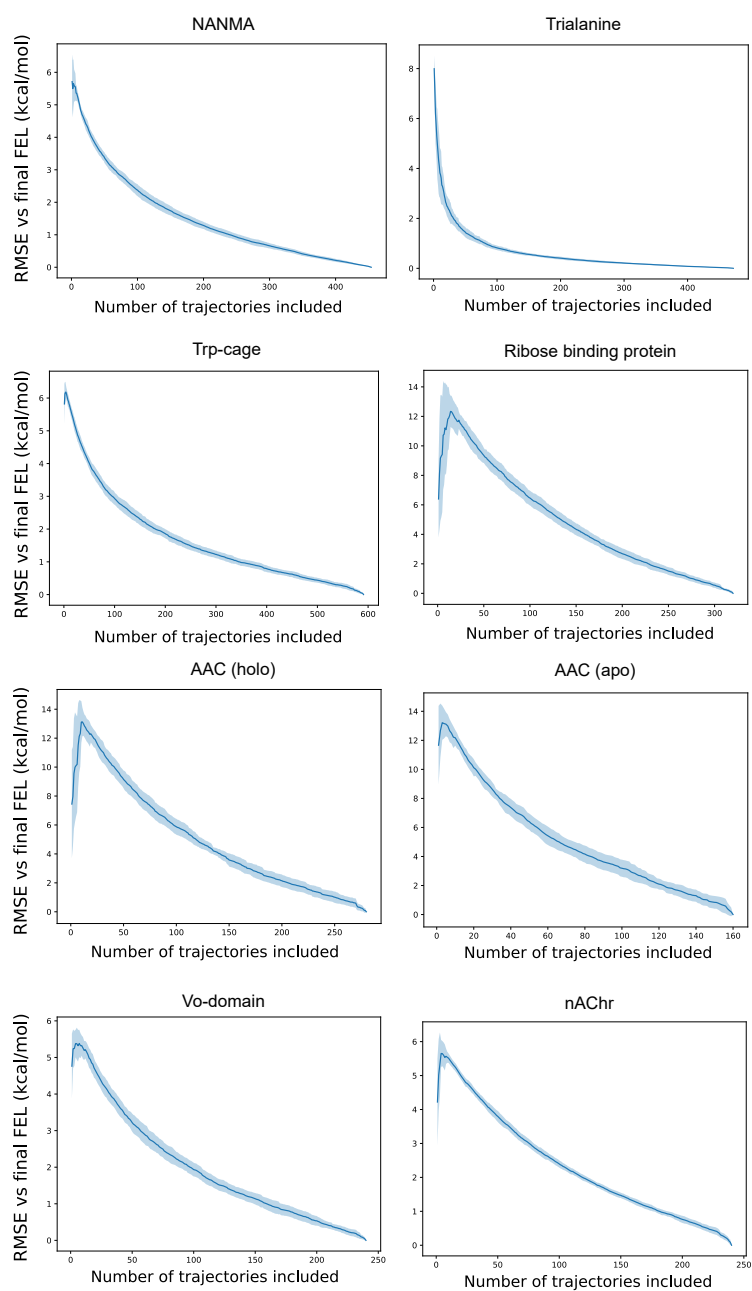


Fig. S6 Convergence of FEL estimation The convergence check based on comparing the FEL estimation with respect to the final estimation as including more unbiased trajectories from Gen-COMPAS.

2.6 Committor Validation Analysis

2.6.1 Committor validation for NANMA

We separate the validation of the learned committor into tests with different levels of physical stringency. NANMA is sufficiently inexpensive that direct shooting can be performed, and we therefore use it for a full validation hierarchy: direct shooting from raw diffusion-generated structures, direct shooting from structures extracted from TMD trajectories, a bidirectional TMD hysteresis test using both predicted and sampled committors, and comparison against an unbiased NANMA trajectory. Throughout this section, the sampled committor is the first-hitting probability of basin B before basin A, estimated as

$$q_{\text{sample}} = \frac{n_B}{n_A + n_B}, \quad (\text{S18})$$

where n_A and n_B are the numbers of shooting trajectories that first reach basins A and B, respectively.

Direct shooting from raw diffusion-generated structures.

As a stringent stress test of the generative component, we first initiated shooting trajectories directly from raw diffusion-generated NANMA structures, including structures produced during intermediate Gen-COMPAS iterations. Using the VCN-learned committor based on a Z-matrix description of the heavy atoms (see Methods), we selected structures with predicted committor values of $q = 0.3, 0.5$, and 0.7 within a tolerance of $\Delta q = 0.05$. For each selected configuration, we performed 1,000 unbiased shooting simulations with randomized initial velocities drawn from the Maxwell–Boltzmann distribution at $T = 300$ K.

The results are reported in Figs. S7 and S8. This test is intentionally severe because raw diffusion-generated structures have not yet been propagated under the atomistic force field through the TMD protocol. They may therefore retain local geometric inconsistencies, reconstructed degrees of freedom not explicitly represented by the generative model, or slow orthogonal coordinates that are not yet dynamically relaxed. We consequently interpret these figures as a stress test of the raw generative proposals, not as the most relevant validation of the structures used for production Gen-COMPAS analysis.

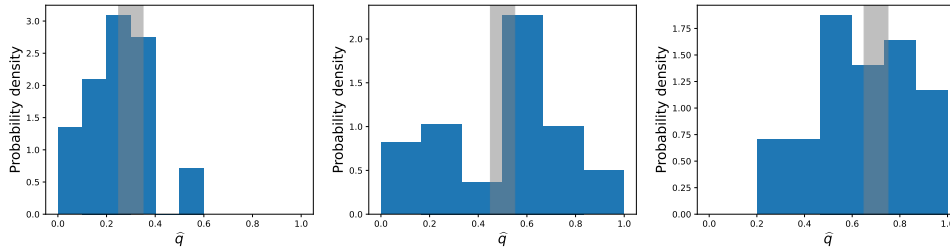


Fig. S7 Committor validation analysis for Gen-COMPAS in NANMA from diffusion-generated structures 1. Histograms of the sampled committor values for selected configurations within $q = 0.3 \pm 0.05$ (left), $q = 0.5 \pm 0.05$ (middle), $q = 0.7 \pm 0.05$ (right) according to the VCN learnings. The vertical grey band correspond to the learned committor interval of selection.

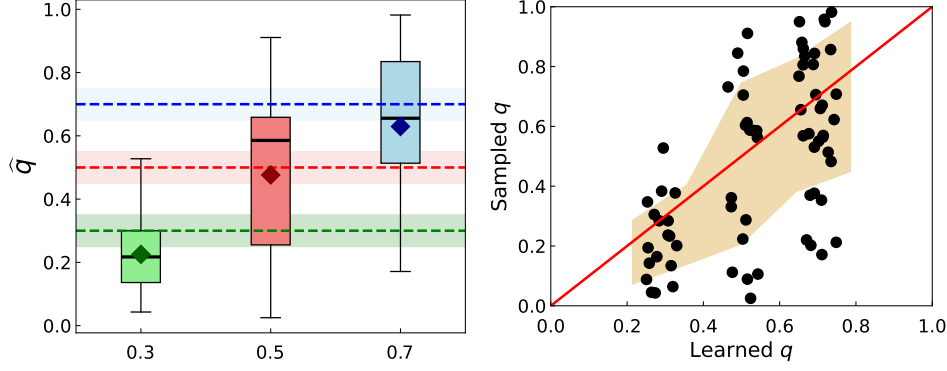


Fig. S8 Committor validation analysis for Gen-COMPAS in NANMA from diffusion-generated structures
2. Box and whisker plots for each histogram is represented against the interval in which each set of structures are selected according to the VCN learnings (left). The mean values for each set are represented with diamonds. We also plot the Learned q vs. Sampled q plot (right) from which a clear correlation can be observed.

Direct shooting from TMD-extracted structures and bidirectional TMD validation.

Because Gen-COMPAS uses diffusion-generated structures as TMD targets rather than as final physical transition states, the validation most directly relevant to the workflow must use structures extracted from TMD trajectories (excluding the last frame to prevent data leakage). We therefore selected TMD-extracted NANMA structures from predicted committor slices centered at $q \approx 0.1, 0.3, 0.5, 0.7$, and 0.9 , and launched unbiased shooting trajectories from these structures.

The resulting sampled committor means are $0.174, 0.359, 0.485, 0.617$, and 0.799 for the five predicted committor slices, respectively (Fig. S9A–G). The pointwise predicted-versus-sampled comparison gives a Pearson correlation coefficient of 0.848 . The validation remains statistically broad, as expected for independent shooting from transition-region configurations, but the sampled committor increases systematically with the predicted committor slice and the $q \approx 0.5$ slice is centered close to 0.5 .

We also tested whether the TMD protocol introduces a systematic bidirectional hysteresis in committor space. For each diffusion-generated target, two TMD trajectories were initiated from basins A and B. Let x_A^{TMD} and x_B^{TMD} denote the corresponding endpoint structures. We computed both the predicted and sampled committor differences,

$$\delta q_{\text{pred}} = q_{\text{pred}}(x_A^{\text{TMD}}) - q_{\text{pred}}(x_B^{\text{TMD}}), \quad (\text{S19})$$

$$\delta q_{\text{samp}} = q_{\text{sample}}(x_A^{\text{TMD}}) - q_{\text{sample}}(x_B^{\text{TMD}}). \quad (\text{S20})$$

If the learned committor masked a TMD hysteresis effect, one would expect δq_{pred} to be narrowly centered near zero while δq_{samp} remained broad or systematically shifted. Instead, both distributions are centered near zero across the predicted committor slices (Fig. S9H–I). This does not prove that RMSD-based TMD is universally free of hysteresis, but it shows that, for NANMA under the present protocol, the two steering directions yield kinetically consistent endpoint structures within shooting uncertainty.

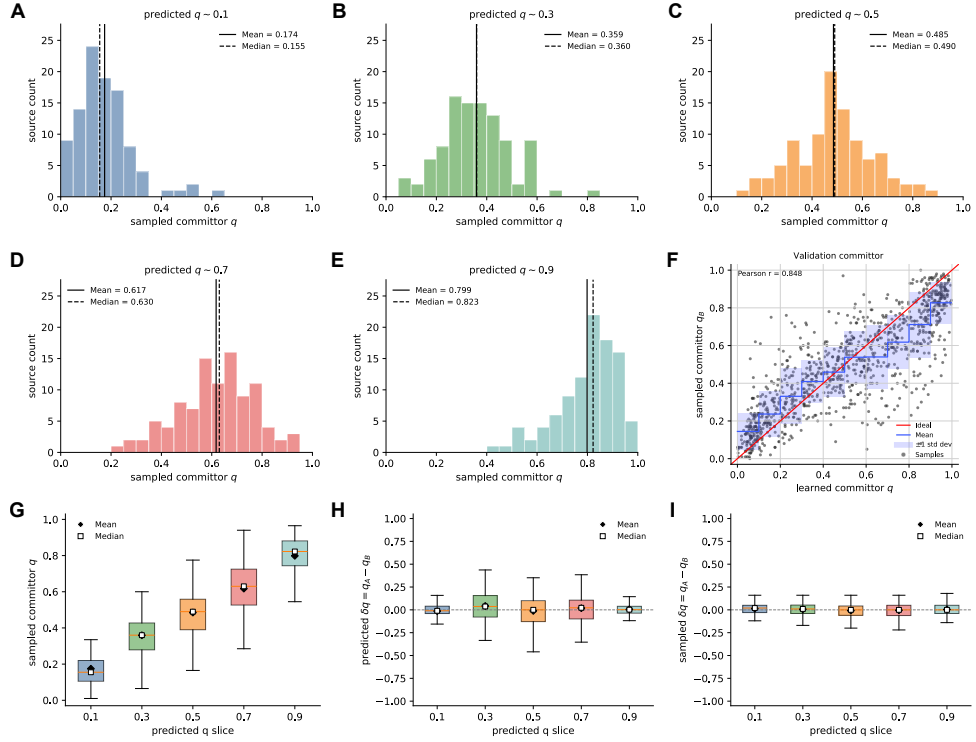


Fig. S9 Committor validation for NANMA from TMD-extracted structures. (A–E) Distributions of sampled committor values for configurations selected from predicted committor slices centered at $q \approx 0.1, 0.3, 0.5, 0.7$, and 0.9 , respectively, with configurations extracted from TMD trajectories. Solid and dashed vertical lines indicate the mean and median sampled committors. The corresponding means are 0.174, 0.359, 0.485, 0.617, and 0.799. (F) Pointwise comparison between predicted committor and sampled committor p_B obtained from unbiased shooting. The Pearson correlation coefficient is 0.848. (G) Box plots of the sampled committor distributions for each predicted committor slice. (H) Distribution of the bidirectional predicted-committor difference, $\delta q_{\text{pred}} = q_A - q_B$. (I) Distribution of the bidirectional sampled-committor difference, δq_{samp} . Both bidirectional distributions are centered near zero, indicating no systematic committor offset between the two TMD steering directions for NANMA.

Comparison against an independent unbiased NANMA trajectory.

Finally, to validate the VCN under configurations sampled from unbiased dynamics, we compared the predicted committor values against exhaustive shooting analyses conducted along an unbiased NANMA trajectory comprising 460,000 steps. For each individual configuration, at each discrete step, we followed the shooting protocol described above. The results are shown in Fig. S10. The sampled and learned committor probabilities agree closely, with a correlation coefficient of 0.9999668. This confirms that the Z-matrix VCN can recover the committor accurately on dynamically relaxed NANMA configurations.

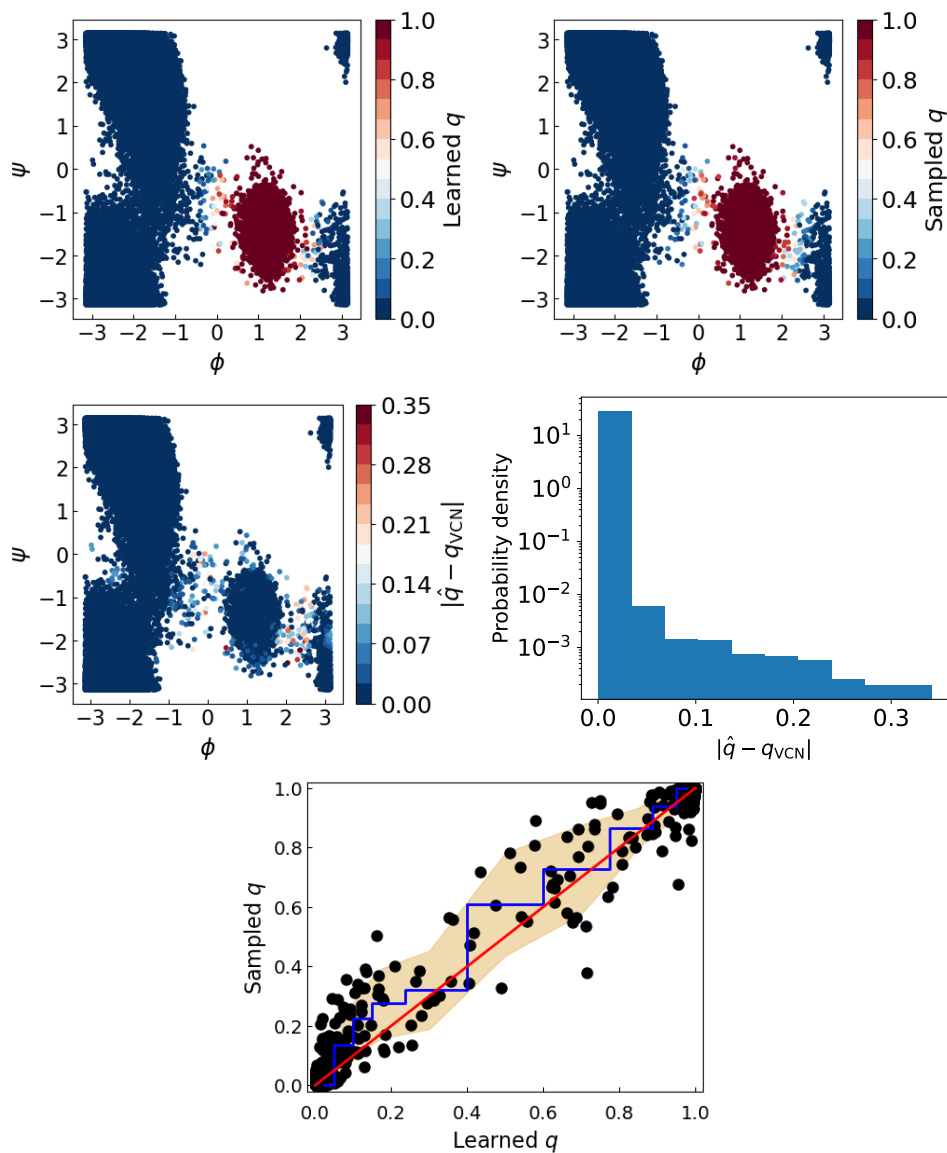


Fig. S10 Committor validation analysis for an unbiased trajectory of NANMA. Analysis of the learned committor probability q for NANMA showing (top left) learned values with the VCN using the Z-matrix representation, q_{predict} , projected into the (ϕ, ψ) two-dimensional space in radians, (top right) sampled committor values, q_{sample} , projected into the (ϕ, ψ) two-dimensional space in radians, (middle left) the absolute error, $|q_{\text{sample}} - q_{\text{predict}}|$, (middle right) the distribution of the absolute errors, and (bottom) learned vs. sampled values to infer the correlation between VCN outcomes and the estimates from the shooting analysis.

2.6.2 Committor Validation for Trp-cage

Direct shooting for Trp-cage.

To directly validate the Gen-COMPAS-predicted committor for Trp-cage, we performed an independent aimless-shooting test (Fig. S11). We selected 100 configurations from an independent Gen-COMPAS run that was not used for training (Run 2 in Fig. S17), with the selected structures spanning the full predicted committor range $q_{\text{pred}} \in [0, 1]$. From each configuration, we launched 20 unbiased trajectories. For each shooting trajectory, the initial velocities were independently sampled from the Maxwell–Boltzmann distribution using different random seeds, so that the shots from the same structure differed only by their stochastic velocity initialization. Each trajectory was propagated for up to 50 ns or until it first reached basin *A* or basin *B*. Because this is a finite-time shooting test, trajectories that did not reach either basin within 50 ns were treated as unresolved.

To reduce finite-time censoring bias, we retained only structures for which more than 50% of the 20 launched trajectories committed to either basin, resulting in $n = 91$ validation points. For each retained structure, the sampled committor was computed as $q_{\text{sample}} = N_B / (N_A + N_B)$, where N_A and N_B are the numbers of resolved trajectories committing to basins *A* and *B*, respectively. The resulting q_{sample} values agree well with q_{pred} over the full committor range, with Pearson correlation coefficient $r = 0.898$ and $R^2 = 0.806$. The uncertainty of each sampled committor estimate was quantified using 90% Wilson confidence intervals. This direct shooting test therefore supports the quantitative accuracy of the Gen-COMPAS committor for Trp-cage, while explicitly accounting for the finite-horizon nature of the validation.

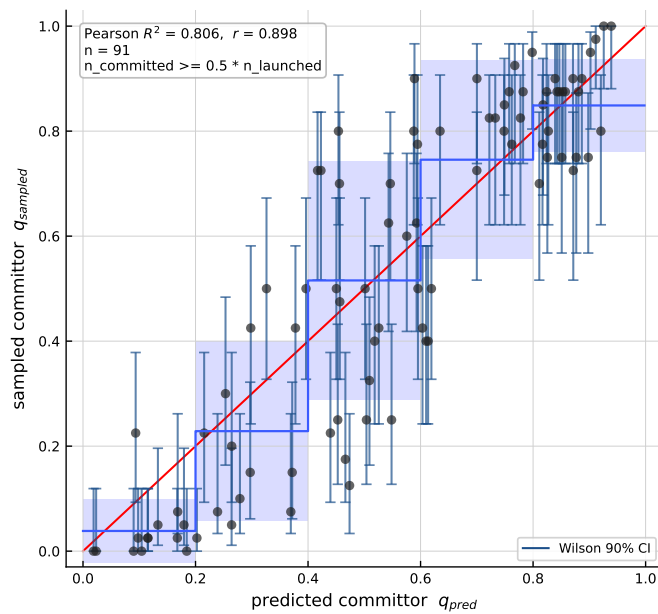


Fig. S11 Direct shooting validation of the Gen-COMPAS-predicted committor for Trp-cage. Predicted versus sampled committor. Structures from an independent Gen-COMPAS run were validated by unbiased shooting trajectories ($n = 91$). Error bars denote 90% Wilson confidence intervals. The strong linear correlation between q_{pred} and q_{sample} ($r = 0.898$, $R^2 = 0.806$) demonstrates good committor prediction accuracy.

Comparison with DESRES and empirical separatrix shooting.

We performed independent checks of the Trp-cage Gen-COMPAS committor using the long unbiased DESRES trajectory. These test data are completely distinct from the Gen-COMPAS training data and was generated with different force-field parameters, making the comparison a stringent assessment of dynamical transferability.

The check combines two complementary checks. First, we evaluated the Gen-COMPAS Z-matrix committor model on DESRES configurations and projected the predicted committor, q_{pred} , onto the two-dimensional CV space defined by RMSD_{CA} and RMSD_{HX} . In the same CV space, we constructed a DESRES first-arrival committor by assigning each frame according to the first basin reached by its future trajectory,

$$q_{\text{AB}}^{\text{DESRES}}(\mathcal{B}) = \frac{N_B(\mathcal{B})}{N_A(\mathcal{B}) + N_B(\mathcal{B})}, \quad (\text{S21})$$

where $\mathcal{B}$ is a CV-space bin, and $N_A(\mathcal{B})$ and $N_B(\mathcal{B})$ count frames in that bin whose future trajectory first reaches the unfolded and folded basins, respectively. The side-by-side maps show that the Gen-COMPAS-predicted separatrix coincides with the DESRES intermediate-committor region, with minimal deviation and no qualitative displacement toward either stable basin (Fig. S12A–B).

Second, we performed an empirical shooting analysis from the separatrix predicted by the Gen-COMPAS model. Specifically, we selected DESRES frames with $q_{\text{pred}} \in [0.45, 0.55]$ and followed the unbiased DESRES trajectory forward from each selected frame until the first arrival to basin A or B. This procedure does not launch new trajectories, but it uses the forward continuation of the equilibrium DESRES trajectory as an ensemble of first-arrival realizations from configurations lying on the model-predicted separatrix. For the $n = 2028$ resolved frames in this window, the mean predicted committor is

$$\langle q_{\text{pred}} \rangle = 0.4998 \quad \text{with 95\% CI } [0.4985, 0.5010], \quad (\text{S22})$$

while the empirical first-arrival committor is

$$q_{\text{sample}} = \frac{N_B}{N_A + N_B} = 0.5335 \quad \text{with 95\% Wilson CI } [0.5118, 0.5552], \quad (\text{S23})$$

with $N_A = 946$ and $N_B = 1082$. Thus, configurations predicted by Gen-COMPAS to lie on the $q = 0.5$ separatrix commit to the two basins with nearly equal probability in an entirely independent unbiased trajectory. The absolute deviation from the ideal value is only 0.0335, which is minimal given the temporal correlations in the DESRES trajectory, the finite number of statistically independent transition-region samples, and the difference in force-field parameters.

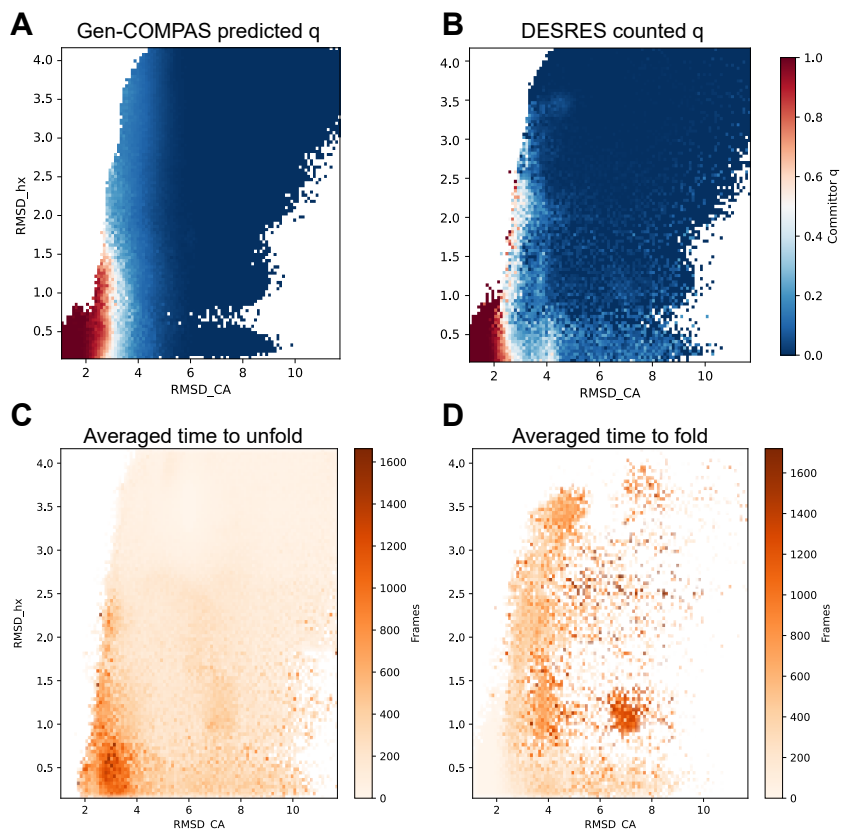


Fig. S12 Independent consistency check of the Gen-COMPAS committor prediction of Trp-cage using the DESRES unbiased trajectory. (A) Gen-COMPAS-predicted committor values, q_{pred} , evaluated on DESRES configurations and projected onto the CV space defined by RMSD_{CA} and RMSD_{HX} . (B) DESRES first-arrival committor estimated in the same CV space by assigning each frame according to the first basin reached by its future trajectory. (C,D) Mean first-arrival times, in saved frames, from each bin to the unfolded basin (C) and folded basin (D), respectively. The Gen-COMPAS separatrix agrees closely with the DESRES intermediate-committor region, and empirical shooting from the model-predicted separatrix yields nearly equal commitment probabilities to the two basins.

Quantity	Estimate	95% CI low	95% CI high
$\langle q_{\text{pred}} \rangle$	0.4998	0.4985	0.5010
$\sigma_{q_{\text{pred}}}$	0.0290	—	—
q_{sample}	0.5335	0.5118	0.5552

$n = 2028$ resolved frames; $N_A = 946$, $N_B = 1082$

Table S2 Empirical separatrix shooting consistent check for the Gen-COMPAS-predicted Trp-cage transition-state slice, defined by $q_{\text{pred}} \in [0.45, 0.55]$. The test uses DESRES configurations that are completely independent of the Gen-COMPAS training data.

Bidirectional TMD consistency for Trp-cage.

As an additional diagnostic, we evaluated bidirectional TMD consistency using both the learned committor and independent aimless-shooting estimates. As shown in Fig. S13A, the distribution of δq_{pred} becomes sharply centered near zero at the end of the trajectories, indicating that the folded-to-target and unfolded-to-target TMD paths converge to configurations with consistent predicted committor values.

To test whether this endpoint consistency is also supported by direct dynamics, we further performed aimless shooting validation. Specifically, we randomly selected 10 targets and their corresponding TMD endpoint structures (therefore 20 endpoint structures) and launched 20 unbiased shooting trajectories from each structure, with initial velocities independently drawn from the Maxwell–Boltzmann distribution using different random seeds. Each trajectory was propagated until it reached basin *A* or basin *B*, or until the finite simulation-time cutoff was reached (50 ns). For each endpoint structure, we estimated the sampled committor and computed the δq_{sample} . As shown in Fig. S13B, the endpoint shooting estimates q_{sample} are consistent with the corresponding Gen-COMPAS predictions q_{pred} , and both distributions have the median of near zero. This indicates that the TMD endpoint structures retain consistent committor values not only according to the learned model but also according to independent shooting dynamics. Together with the DESRES map comparison and Trp-cage aimless-shooting validation, this supports the dynamical consistency of the Gen-COMPAS Trp-cage transition-state ensemble and indicates that the TMD protocol does not introduce a detectable systematic committor bias.

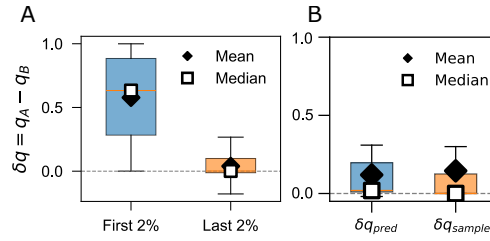


Fig. S13 Bidirectional TMD consistency and endpoint shooting validation for Trp-cage. (A) Distribution of $\delta q_{\text{pred}}(t)$ along paired TMD trajectories initialized from the folded and unfolded basins and driven toward the same diffusion-generated target. The distribution becomes centered near zero at the trajectory endpoints. (B) $\delta q_{\text{pred}}(t)$ and $\delta q_{\text{sample}}(t)$ distributions for ten randomly selected pairs of TMD endpoints structures (each pair with the same target starting from the folded and unfolded basins).

2.7 Ensemble-based uncertainty estimation of the committor

As demonstrated in the preceding section, while validating the committor function via unbiased shooting trajectories is numerically robust, it is also computationally prohibitive and logistically demanding, rendering it infeasible for complex biomolecular systems. To assess the reliability of the learned committor function at an appreciably reduced computational cost, we estimated the predictive uncertainty using an ensemble of independently trained neural networks. Ensemble methods provide a practical approximation of Bayesian model uncertainty, and have been widely adopted for uncertainty quantification in deep learning models^{24,25}, most notably in machine-learning frameworks for atomistic simulations^{26–28}.

An ensemble of $N = 8$ NN-models was trained using identical architectures and training protocols, but with distinct random seeds and independent training-validation splits. Consequently, each model converges to a distinct set of network parameters due to stochastic elements inherent to the optimization procedure, including random weight initialization and stochastic minibatch ordering. The ensemble, therefore, defines a set of committor predictors

$$\{q_1(\mathbf{x}), q_2(\mathbf{x}), \dots, q_N(\mathbf{x})\}, \quad (\text{S24})$$

where $\mathbf{x}$ denotes the molecular configuration represented by the chosen input features (CVs or internal coordinates).

For each configuration, $\mathbf{x}$, the committor value is predicted using all models in the ensemble. The committor estimate is taken as the ensemble mean,

$$\overline{q(\mathbf{x})} = \frac{1}{N} \sum_{i=1}^N q_i(\mathbf{x}), \quad (\text{S25})$$

and the uncertainty of the committor prediction is quantified by the standard deviation of the ensemble predictions,

$$\sigma_q(\mathbf{x}) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(q_i(\mathbf{x}) - \overline{q(\mathbf{x})} \right)^2}. \quad (\text{S26})$$

This quantity reflects the epistemic uncertainty arising from limited training data and model flexibility. Regions where the ensemble members produce inconsistent predictions correspond to configurations where the committor function is poorly constrained by the training data.

To analyze the spatial distribution of the uncertainty, $\sigma_q(\mathbf{x})$ was projected onto a two-dimensional CV space used for visualization. The CV space was discretized into a regular grid, and the average uncertainty in each bin $\mathcal{B}$ was computed as,

$$\langle \sigma_q \rangle_{\mathcal{B}} = \frac{\sum_{x \in \mathcal{B}} \sigma_q(x)}{\sum_{x \in \mathcal{B}} 1}. \quad (\text{S27})$$

The resulting uncertainty field highlights regions of configuration space where the committor model is less reliable. Such regions often correspond to poorly sampled states, or to transition regions where the committor function is sensitive to structural fluctuations (see Fig. S14).

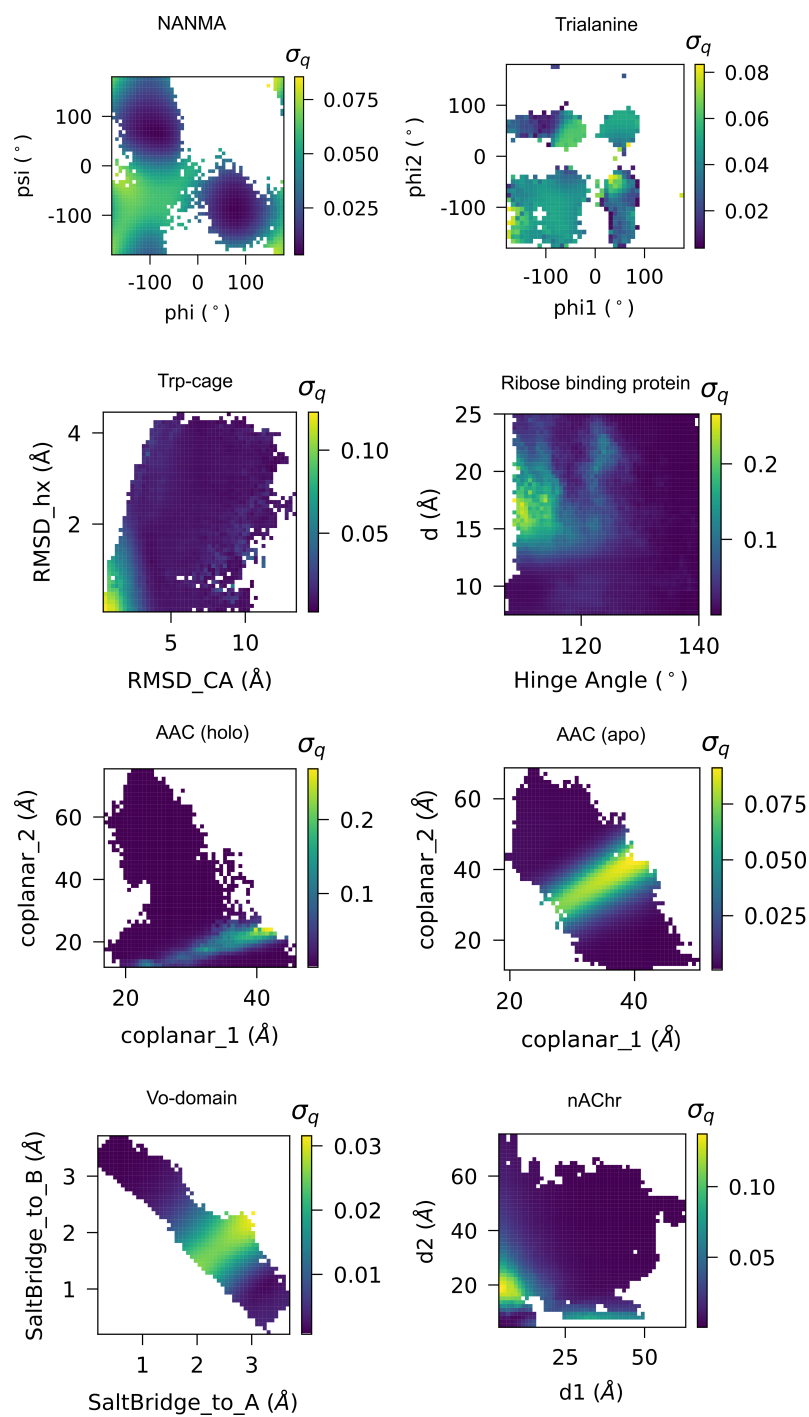


Fig. S14 Uncertainty estimation of committor function The uncertainty estimation of the committor function from Gen-COMPAS by checking the standard deviation of the ensemble predictions ($N = 8$) projected on two dimensions.

2.8 TMD validation and reproducibility

2.8.1 Practical guidance for parameter tuning

The hyperparameters reported in Tables S3–S5 should be treated as empirically validated settings for the present systems rather than universal constants. In practice, however, most machine-learning hyperparameters were not tuned on a system-by-system basis. For the complex systems studied here, we used the same generative-model architecture, optimizer settings, training protocol, and diffusion-related hyperparameters, with the batch size being adjusted only when required by GPU memory limitations. The exact values used in each application are reported in Tables S3–S5.

For the VCN, the lag time τ is weakly sensitive within a broad physically meaningful range. As discussed in our previous work, training is mainly affected when τ is chosen at pathological limits: if it is too short, e.g., on the order of ~ 1 fs, the data retain non-Markovian vibrational correlations; if it is too long, e.g., on the order of ~ 1 ns, transition-region correlations are largely lost and the training statistics deteriorate. Between these limits, the learned committor is only weakly affected by the precise choice of τ . The values used in the present applications are reported in Table S5.

The parameters that require the most practical attention are the TMD force constant and the noise level used during generative sampling. The TMD force constant should be large enough to reduce the RMSD to the generated target over the chosen TMD duration, but not so large that it produces abrupt structural distortions or unstable forces. We therefore recommend choosing this parameter by inspecting the first and final 2% RMSD distributions, as shown in Figs. S15 and S16. A useful setting produces a clear final-RMSD decrease while maintaining structurally plausible trajectories.

For generative sampling, we recommend noise values in the approximate range 0–50. Larger values can be useful during the initial iterations, especially when the available trajectories are short or sparse, because they increase the diversity of generated intermediates and help initialize subsequent sampling. After additional transition data have been accumulated, smaller values, typically 0–2, are usually sufficient. Increasing the noise moves generated structures farther from the current data distribution and improves diversity, whereas decreasing it keeps generated structures closer to previously sampled configurations but reduces exploration. In all cases, the final acceptance criterion is not the generative loss alone, but the downstream MD diagnostics: TMD convergence, bidirectional committor consistency, shooting validation where feasible, and reproducibility across independent Gen-COMPAS runs.

2.8.2 Validation of TMD parameters

Because the TMD stage is used to guide the system toward generative intermediate structures before launching unbiased shooting trajectories, it is important to ensure that the TMD parameters produce physically meaningful configurations without introducing excessive structural distortions. To validate the choice of TMD parameters, we performed a systematic parameter scan for the Trp-cage system during the 0th iteration of the Gen-COMPAS workflow. Specifically, we explored a grid of 4 TMD durations and 5 force constants, resulting in 20 independent TMD simulations. For each trajectory, we computed RMSD between the instantaneous structure and the target configuration.

To assess the stability and convergence behavior of the TMD steering, we compared the distributions of RMSD values during the first 2% and the final 2% of each TMD trajectory (Fig. S15). This analysis allows us to evaluate whether the system approaches the target configuration smoothly while avoiding abrupt structural distortions. Based on these results, we selected parameter combinations that achieve reliable convergence toward the target while maintaining realistic structural transitions. The same diagnostic analysis was subsequently performed for all systems studied in this work (Fig. S16) to verify that the chosen TMD parameters lead to stable refinement of generated intermediates prior to launching unbiased molecular dynamics trajectories.

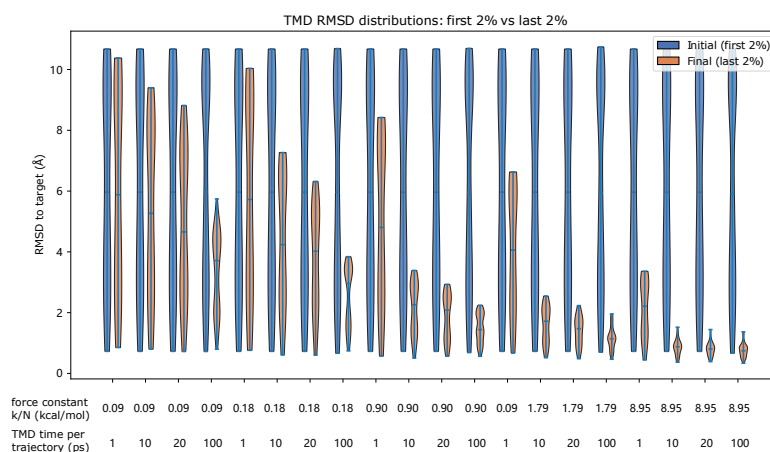


Fig. S15 Dependency on TMD parameters tested on Trp-cage. The TMD is replicated for the initial iteration of the Gen-COMPAS.

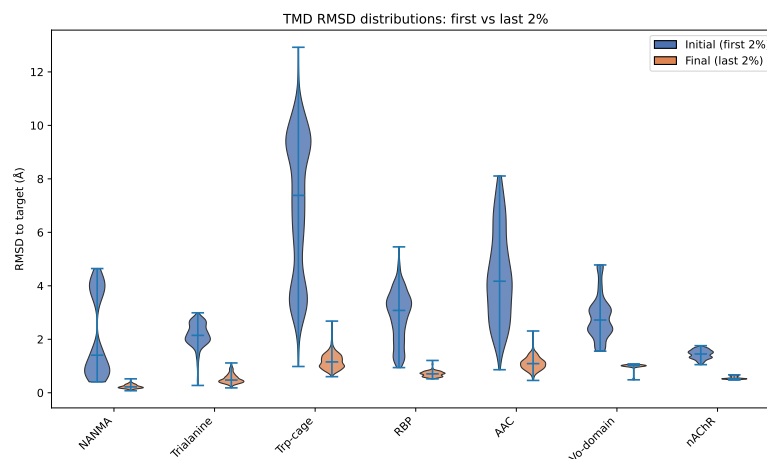


Fig. S16 TMD convergence check on all systems for all iterations of the Gen-COMPAS.

2.8.3 Reproducibility tests

To evaluate the robustness and reproducibility of the Gen-COMPAS framework, we performed multiple independent runs of the complete workflow for the Trp-cage folding system. In total, five independent Gen-COMPAS simulations were carried out. Three runs used the same parameters as reported in the main manuscript, while two additional runs were performed using modified TMD force constants in order to assess the sensitivity of the method to suboptimal steering parameters.

Across these independent runs, the reconstructed FELs and committor predictions were found to be consistent, demonstrating that the method reliably identifies the relevant transition region even when moderate variations in TMD parameters are introduced (see Fig. S17). Even if the TMD parameters are chosen too poorly (in the case of **Run 1**, where the force constant is set to be too small), when the sampling of the transition states is much less ideal, the results of Gen-COMPAS are still qualitatively consistent. The issue can also be easily avoided as long as users have checked the TMD parameters as shown in Figs. S15 and S16. These tests confirm that the Gen-COMPAS workflow is robust with respect to stochastic initialization and parameter perturbations, and that the mechanistic conclusions reported in this work are reproducible across independent simulations.

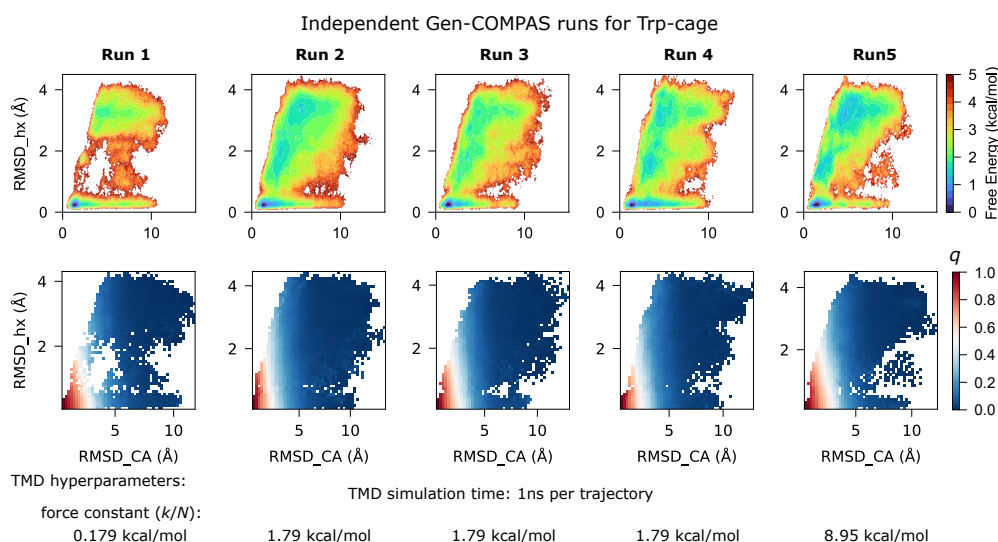


Fig. S17 Independent runs of Gen-COMPAS on Trp-cage

2.9 Molecular systems and the Gen-COMPAS setup

In this section, we describe the setup of the systems shown in the SI as well as the initialization parameters for all systems of the Gen-COMPAS framework. All of the MD simulations described below were performed using the popular NAMD MD engine²⁹.

2.9.1 NANMA isomerization in vacuum

N-acetyl-*N'*-methylalaninamide (NANMA) in vacuum was simulated at 300K using, for comparison purposes with previous studies¹⁻⁷, the CHARMM22 force field³⁰, with a Langevin thermostat with a damping coefficient of 1 ps^{-1} . For the initial sampling, we performed two 500-ps unbiased simulations starting from representative conformations of the FEL, namely, C_{7eq} and C_{7ax} . These trajectories were used to train the 0-th iteration of the diffusion-based generative model, which produced candidate intermediates connecting the two metastable states.

2.9.2 Trialanine conformational equilibrium in vacuum

Trialanine in vacuum was simulated at 300K using the AMBER ff14SB force field³¹ and a Langevin thermostat with a damping coefficient of 1 ps^{-1} . Two initial unbiased simulations of 500 ps were performed starting from two metastable states (defined in the Supplementary Text). These data provided the first training set for the diffusion model of the 0-th iteration of Gen-COMPAS.

2.9.3 Ribose-binding protein binding-upon-folding

The ribose-binding protein (RBP) was simulated in explicit TIP3P water using the CHARMM36m force field³². Simulations were initialized from both the apo (open) state and the holo (closed, ligand-bound) state. The system was equilibrated and sampled in the NPT ensemble at 310 K and 1 atm, using a Langevin thermostat with a damping coefficient of 1 ps^{-1} . Unbiased simulations from the apo and holo states were run respectively (2 ns for each) to train the diffusion model to initialize the Gen-COMPAS framework.

2.9.4 V-ATPase V_o domain rotary transition

The membrane-embedded V_o domain of the vacuolar ATPase was simulated using the same molecular dynamics protocol described as the RBP. All simulations employed the CHARMM36m force field³² with explicit TIP3P water. The system was simulated in the NPT ensemble at 310 K and 1 atm using a Langevin thermostat with a damping coefficient of 1 ps^{-1} .

Initial configurations corresponding to the two endpoint states of the rotary transition were taken from the structures reported¹⁵, which characterize adjacent rotational states of the c-ring relative to the a-subunit in the V_o domain. These structures represent two metastable configurations of the rotor-stator interface and define the reactant (*A*) and product (*B*) states used to initialize the Gen-COMPAS framework.

Short unbiased simulations were performed starting from each endpoint configuration to generate the initial dataset used to train the diffusion model. Specifically, 2 ns of unbiased molecular dynamics were carried out from each state, producing an ensemble of configurations that serve as input for the first iteration of the Gen-COMPAS workflow. Subsequent sampling

followed the iterative generative–committor-guided procedure described above to explore the transition region between the two rotational states of the V_o motor.

2.9.5 Numerical details

A summary of the different parameters used in the Gen-COMPAS framework can be found in Tables S3–S5 for the illustrations reported herein: NANMA isomerization, trialanine conformational equilibrium, Trp-cage folding, RBP binding-upon-folding, AAC conformational transition, V_o domain rotation, and nAChR activation. Table S6 records the estimated computational cost for all the systems investigated in the manuscript other than the two toy systems (NANMA and trialanine).

Hyperparameters	NANMA	Trialanine	Trp-cage	RBP	AAC	Vo-domain	nAChR
Beta schedule	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine
Diffusion steps (t_T)	200	200	200	200	200	200	200
Hidden dimensions	128	128	128	128	128	128	128
Node feature	64	64	64	64	64	64	64
k neighbors	16	16	16	16	16	16	16
SchNet layers	4	4	4	4	4	4	4
Attention heads	4	4	4	4	4	4	4
Time embedding	128	128	128	128	128	128	128
Batch size	1280	1280	64	16	16	16	16
Epochs	50(0th), 20, 15, 10, 10	50(0th), 20, 15, 10	50(0th), 20, 10, 10	50(0th), 20, 10, 10	25(0th), 25, 10, 10	50(0th), 20, 10, 10	50(0th), 20, 10, 10
Learning rate	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}
Weight decay	1.0×10^{-6}	1.0×10^{-6}	1.0×10^{-6}	1.0×10^{-6}	1.0×10^{-6}	1.0×10^{-6}	1.0×10^{-6}
Noise	2.0(0th), 3.0, 3.0, 1.5, 4.0	2.0(0th), 2.0, 2.0, 2.0	2.0(0th), 2.0, 0.5, 0.5	10.0(0th), 10.0, 1.0, 1.0	20.0(0th), 1.5, 1.5, 1.5	0.8(0th), 2.0, 2.0, 1.5	10.0(0th), 10.0, 5.0, 3.0
Generated structures	$10,000 \times 5$	$10,000 \times 4$	$10,000 \times 4$	$10,000 \times 4$	$10,000 \times 4$	$10,000 \times 4$	$10,000 \times 4$

Table S3 Hyperparameters for the diffusion model used in Gen-COMPAS simulations. The cosine beta schedule defines $\bar{\alpha}_t = g(t)/g(0)$, such that $g(t) = \cos\left(\frac{t/t_T + s}{1+s} \frac{\pi}{2}\right)$, where s is an offset (set to $s = 0.008$ ³³) and t_T is the total number of diffusion steps.

Parameters	NANMA	trialanine	Trp-cage	RBP	AAC (apo/holo)	Vo-domain	nAChR
Timestep (fs)	0.5	0.5	2.0	2.0	2.0	2.0	4.0
Temperature (K)	300	300	290	310	310	310	310
TMD time (ps)	10	10	100	100	100	100	100
Selected atoms for TMD	heavy	heavy	side-chain carbon	protein C _α ligand carbon	protein C _α (ligand PB N6)	C _α	protein C _α ligand N1 C10
TMD force constant (k/N , kcal/mol)	50	5.0	1.79	1.28	2.29 (2.28)	1.28	4.99
Each unbiased shooting time (ps)	100	100	1,000	2,000	2,000	2,000	2,000
Total unbiased MD time (ns)	66	47.2	594	480	320 (480)	480	480

Table S4 Parameters for the targeted MD (TMD) and unbiased MD simulations.

Parameters	NANMA	trialanine	Trp-cage	RBP	AAC (apo/holo)	Vo-domain	nAChR
Atom selection	heavy	heavy	side-chain carbon	protein C _α ligand heavy	protein C _α (ligand heavy)	protein C _α	protein C _α and ligands heavy
Time lag (fs)	100	100	2,000	4,000	4,000	4,000	4,000
Epochs	500	500	500	500	500	500	500
Patience	20	20	20	20	20	20	20
Hidden layers	4	4	4	4	4	4	4
Hidden nodes	64	64	128	128	128	128	128
Learning rate	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}
ϵ (i -th iteration)	0.2(1st), 0.2, 0.1, 0.1	0.3(1st), 0.1, 0.1	0.1(1st), 0.1, 0.1	0.2(1st), 0.05, 0.05	0.3(1st), 0.3, 0.3	0.3(1st), 0.1, 0.1	0.3(1st), 0.1, 0.1
Clustering ($q \approx 0.5$)	No	No	No	Yes	No(apo) Yes(holo)	Yes	Yes

Table S5 Hyperparameters for the VCN.

Step (GPU hours)	Trp-cage	RBP	AAC (holo)	AAC (apo)	Vo-domain	nAChR
DDPM training	0.9, 5.25, 14.5	1.42, 10.53, 14.68	1.82, 13.11, 12.17	2.19, 5.55, 13.20	28.00, 42.00, 56.02	33.12, 19.86, 32.67
DDPM generation	0.45, 0.45, 0.45	5.6, 5.6, 5.6	16.7, 16.7, 16.7	11.7, 11.7, 11.7	17.56, 17.56, 17.56	15.56, 15.56, 15.56
TMD simulations	0.64, 8.27, 10.88	3.34, 6.69, 6.69	8.36, 12.55, 12.55	5.89, 8.83, 8.83	38.51, 38.51, 38.51	12.64, 12.64, 12.64
Unbiased shooting	11.31, 66.77, 87.91	29.41, 58.83, 58.83	40.49, 74.87, 74.87	28.07, 42.11, 42.11	64.6, 64.6, 64.6	165.2, 165.2, 165.2
VCN training	0.08, 0.13, 0.12	0.12, 0.11, 0.14	0.25, 0.21, 0.37	0.23, 0.19, 0.19	0.45, 0.43, 0.45	0.22, 0.29, 0.35
Total GPU hours	208.11	207.59	301.72	192.49	489.36	666.71

Table S6 Estimated GPU hours required for each major stage of the Gen-COMPAS workflow for the different biomolecular systems (excluding toy models) studied in this work. GPU hours were measured on $8 \times$ NVIDIA L40s GPUs and included DDPM training, DDPM generation, TMD and unbiased shooting simulations, and committor model (VCN) training. The trainings are parallelized using Distributed Data Parallel. The wall-clock times were estimated by single epoch benchmarking, single batch generation benchmarking, and single MD trajectory benchmarking.

References

- [1] Bolhuis, P. G., Dellago, C. & Chandler, D. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5877–5882 (2000).
- [2] Rosso, L., Abrams, J. B. & Tuckerman, M. E. Mapping the backbone dihedral free-energy surfaces in small peptides in solution using adiabatic free-energy dynamics. *J. Phys. Chem. B* **109**, 4162–4167 (2005).
- [3] Hénin, J., Fiorin, G., Chipot, C. & Klein, M. L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.* **6**, 35–47 (2010).
- [4] Chen, H., Roux, B. & Chipot, C. Discovering reaction pathways, slow variables, and committor probabilities with machine learning. *J. Chem. Theory Comput.* **19**, 4414–4426 (2023).
- [5] Kang, P., Trizio, E. & Parrinello, M. Computing the committor with the committor to study the transition state ensemble. *Nat. Comput. Sci.* **4**, 451–460 (2024).
- [6] Trizio, E., Kang, P. & Parrinello, M. Everything everywhere all at once: a probability-based enhanced sampling approach to rare events. *Nat. Comput. Sci.* **5**, 582–591 (2025).
- [7] Megías, A. *et al.* Iterative variational learning of committor-consistent transition pathways using artificial neural networks. *Nat. Comput. Sci.* **5**, 592–602 (2025).
- [8] Tiwary, P. & Berne, B. J. Predicting reaction coordinates in energy landscapes with diffusion anisotropy. *J. Chem. Phys.* **147**, 152701 (2017).
- [9] Fu, H., Shao, X., Cai, W. & Chipot, C. Taming rugged free energy landscapes using an average force. *Acc. Chem. Res.* **52**, 3254–3264 (2019).
- [10] Chen, H. *et al.* A companion guide to the string method with swarms of trajectories: Characterization, performance, and pitfalls. *J. Chem. Theory Comput.* **18**, 1406–1422 (2022).
- [11] Ravindranathan, K. P., Gallicchio, E. & Levy, R. M. Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol.* **353**, 196–210 (2005).
- [12] Li, H. Y., Cao, Z. X., Zhao, L. L. & Wang, J. H. Analysis of conformational motions and residue fluctuations for escherichia coli ribose-binding protein revealed with elastic network models. *Int. J. Mol. Sci.* **14**, 10552–10569 (2013).
- [13] Ren, W. *et al.* Unraveling the coupling between conformational changes and ligand binding in rbp_g134r via enhanced sampling. *J. Phys. Chem. B* **125**, 2898–2909 (2021).

- [14] Li, H. *et al.* Ligand-induced structural changes analysis of rbp as studied by md simulations. *Technol. Health Care* **29**, 103–114 (2021).
- [15] Roh, S.-H. *et al.* Cryo-em and md infer water-mediated proton transport and autoinhibition mechanisms of vo complex. *Sci. Adv.* **6**, eabb9605 (2020).
- [16] Blanc, F. E. C. & Hummer, G. Mechanism of proton-powered c-ring rotation in a mitochondrial atp synthase. *Proc. Natl. Acad. Sci. U.S.A* **121**, e2314199121 (2024).
- [17] Song, J., Meng, C. & Ermon, S. Denoising diffusion implicit models. *Proc. International Conference on Learning Representations (ICLR)* (2021).
- [18] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148** (2018).
- [19] Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process.* **30** (2017).
- [20] Giuseppe Chen, C. *et al.* Following the committor flow: A data-driven discovery of transition pathways. *J. Chem. Theory Comput.* **22**, 1258–1265 (2026).
- [21] Kania, S., Webber, R. J., Simpson, G., Aristoff, D. & Zuckerman, D. M. Randomized iterative trajectory reweighting for steady-state distributions without discretization error. *Proc. Natl. Acad. Sci. U.S.A.* **123**, e2529246123 (2026).
- [22] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **32**, 922–923 (1976).
- [23] Villani, C. *et al.* *Optimal transport: old and new* (Springer, 2008).
- [24] Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. in Neural Information Processing Systems* (2017).
- [25] Abdar, M. & et al. A review of uncertainty quantification in deep learning. *Inf. Fusion* **76**, 243–297 (2021).
- [26] Smith, J. S. *et al.* Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* .
- [27] Schran, C., Brezina, K. & Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. *J. Chem. Phys.* **153**, 104105 (2020).
- [28] Xu, H. *et al.* Evidential deep learning for interatomic potentials. *Nat. Comm.* (2025).
- [29] Phillips, J. C. *et al.* Scalable molecular dynamics on cpu and gpu architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).

- [30] MacKerell, A. D. J. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- [31] Maier, J. A. *et al.* ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
- [32] Huang, J. *et al.* Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* **14**, 71–73 (2017).
- [33] Nichol, A. & Dhariwal, P. Improved denoising diffusion probabilistic models. *arXiv preprint arXiv:2102.09672* (2021).