

SI Appendix

Multiscale autoregressive modeling generates ordered protein conformational trajectories

Table of Contents

Part I: Representation, Methods, and Evaluation Protocols

S1. Protein Structure Representation	2
S2. Model Architecture, Training, and Task-Specific Generation	3
S3. Evaluation Metrics and Analysis Protocols	8

Part II: Extended Experimental Results

S4. Multiscale versus Single-Scale Ablation	12
S5. Adenylate Kinase Transition Analysis	13
S6. Additional Endpoint-Conditioned Transitions	18
S7. Cross-State Q-Score Patterns	19
S8. Exploratory GPCR Transfer Cases	19

Part I: Representation, Methods, and Evaluation Protocols

S1 Protein Structure Representation

We represent protein structures using a combination of local reference frames and torsion angles, resulting in a rotation and translation equivariant representation. For a protein with L amino acids, our tokenization procedure is as follows:

S1.1 SE(3) Frame Representation

For each residue i , we construct a local reference frame using the backbone atoms (N, C α , C). Specifically:

- The origin O_i is placed at the C α atom position
- The x-axis $\hat{x}_i$ is aligned with the normalized C α -N bond vector
- The temporary vector $\vec{v}_i$ is the normalized C α -C bond vector
- The z-axis $\hat{z}_i$ is computed as $\hat{z}_i = \frac{\hat{x}_i \times \vec{v}_i}{|\hat{x}_i \times \vec{v}_i|}$
- The y-axis $\hat{y}_i$ completes the right-handed coordinate system: $\hat{y}_i = \hat{z}_i \times \hat{x}_i$

Each SE(3) frame consists of a rotation matrix $R_i \in \text{SO}(3)$ and a translation vector $t_i \in \mathbb{R}^3$:

$$R_i = [\hat{x}_i, \hat{y}_i, \hat{z}_i], \quad t_i = O_i \quad (\text{S1})$$

For computational efficiency, we represent rotations using unit quaternions:

$$q_i = \begin{bmatrix} q_{w,i} \\ q_{x,i} \\ q_{y,i} \\ q_{z,i} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \sqrt{1 + R_i[0,0] + R_i[1,1] + R_i[2,2]} \\ \frac{R_i[2,1] - R_i[1,2]}{4q_{w,i}} \\ \frac{R_i[0,2] - R_i[2,0]}{4q_{w,i}} \\ \frac{R_i[1,0] - R_i[0,1]}{4q_{w,i}} \end{bmatrix} \quad (\text{S2})$$

Combined with the translation vector, this yields a 7-dimensional vector $[q_{w,i}, q_{x,i}, q_{y,i}, q_{z,i}, t_{x,i}, t_{y,i}, t_{z,i}]$ for each residue, resulting in a tensor of shape $[L, 7]$ for the entire protein.

S1.2 Torsion Angle Representation

We complement the SE(3) frames with the backbone torsion angles (ϕ , ψ , ω), which define the protein's conformation. For residue i , these angles are defined as:

$$\phi_i = \text{dihedral}(C_{i-1}, N_i, C\alpha_i, C_i) \quad (\text{S3})$$

$$\psi_i = \text{dihedral}(N_i, C\alpha_i, C_i, N_{i+1}) \quad (\text{S4})$$

$$\omega_i = \text{dihedral}(C\alpha_i, C_i, N_{i+1}, C\alpha_{i+1}) \quad (\text{S5})$$

Rather than using the raw angles, we represent each angle as a 2D vector $[\sin(\theta), \cos(\theta)]$ to avoid discontinuities at $\pm\pi$. This results in a 6-dimensional vector per residue:

$$v_{\text{torsion},i} = [\sin(\phi_i), \cos(\phi_i), \sin(\psi_i), \cos(\psi_i), \sin(\omega_i), \cos(\omega_i)] \quad (\text{S6})$$

S1.3 Final Representation

We concatenate the SE(3) frame and torsion angle representations to obtain a comprehensive protein structure encoding of shape [L, 13], where a 13-dimensional vector represents each residue:

$$v_i = [q_{w,i}, q_{x,i}, q_{y,i}, q_{z,i}, t_{x,i}, t_{y,i}, t_{z,i}, \sin(\phi_i), \cos(\phi_i), \sin(\psi_i), \cos(\psi_i), \sin(\omega_i), \cos(\omega_i)] \quad (\text{S7})$$

S1.4 Rotation and Translation Equivariance

Our final protein representation is equivariant to global rigid-body motions. Specifically, a global rotation $R_g \in \text{SO}(3)$ and translation $t_g \in \mathbb{R}^3$ transform each local frame (R_i, t_i) as $(R_g R_i, R_g t_i + t_g)$. When using unit quaternions q_i to represent R_i , this corresponds to a left quaternion multiplication:

$$q'_i = q_g \otimes q_i, \quad t'_i = R_g t_i + t_g \quad (\text{S8})$$

where q_g is the unit quaternion corresponding to R_g , and $\otimes$ denotes quaternion multiplication.

The torsion angles (ϕ, ψ, ω) , represented as $[\sin(\theta), \cos(\theta)]$ pairs, are internal degrees of freedom and remain invariant under global SE(3) transformations.

S2 Model Architecture, Training, and Task-Specific Generation

S2.1 Multiscale training

The slow- and fast-scale models use the same training procedure but differ in how the trajectory frames are sampled:

Slow-scale Training. For the slow-scale model ($\Delta t_s = 20$ ns), we sample frames from the full trajectory at 20-ns intervals so that the prediction targets span the 20-ns training stride.

Fast-scale Training. For the fast-scale model ($\Delta t_f = 1$ ns), we randomly sample continuous 20-ns trajectory segments at 1-ns intervals. Each segment is treated as an independent training sequence at the 1-ns training stride.

S2.2 Notations

- $g_t \in \text{SE}(3)^L$: The frame sequence at time t , where each frame consists of rotation and translation components $(R, \mathbf{t})$ in SE(3) space
- $X_t \in \mathbb{R}^{L \times 13}$: Protein conformation at time t , where each residue includes 13 features as described in Section S1
- T : Number of frames used for both input and prediction windows
- K : Total number of frames in the training trajectory (differs between scales)
- σ_{noise} : Noise scale in our SDE formulation

Each model update maps a window of T input snapshots to a window of T predicted snapshots. The conditional generator therefore uses finite structural history and is not treated as a first-order Markov process in the current conformation X_t alone.

S2.3 Training Algorithm

Algorithm S1 describes our autoregressive training procedure with teacher forcing.

Algorithm S1: Autoregressive Training with Multiple-Frame Windows

Input: ground truth frame sequences $\{g_t\}_{t=0}^{K-1} \in (\text{SE}(3)^L)^K$,
ground truth trajectories $\{X_t\}_{t=0}^{K-1} \in (\mathbb{R}^{L \times 13})^K$,
amino acid identities $A \in \{1, \dots, 20\}^L$,
teaching force probability p_{tf} ,
number of frames T (for both input and prediction)
Output: trained model parameters θ

```
1 for each training iteration at epoch  $e$  of  $E$  do
    // Initialize input buffers with first  $T$  frames
2   Initialize  $X_{buffer} \leftarrow \{X_i\}_{i=0}^{T-1}$ ;
3   Initialize  $g_{buffer} \leftarrow \{g_i\}_{i=0}^{T-1}$ ;
4    $\mathcal{L}_{total} \leftarrow 0$ ;
5   for  $t = 0$  to  $\lfloor K/T \rfloor - 2$  do
6        $z \sim \text{Beta}(1, 5(1 - e/E) + 1)$ ;  $\sigma_{noise} \leftarrow 0.01 + 0.99z$ ;
7        $\epsilon \sim \mathcal{N}(0, I)$ ;
8        $h_{latent} \leftarrow W_{proj}(X_{buffer} + \sigma_{noise}\epsilon)$ ;
9        $s \leftarrow \text{Embed}(A)$ ;
10      for  $i = 1$  to  $n_{ipa}$  do
11           $h_{frame} \leftarrow \text{InvariantPointAttention}(s, g_{buffer}, \sigma_{noise})$ ;
          // spatial processing
12           $h_{spatial} \leftarrow h_{latent} + h_{frame} + \text{Embed}(\sigma_{noise})$ ;
13          for  $j = 1$  to  $n_{att}$  do
14               $h_{spatial} \leftarrow \text{MultiHeadAttention}(h_{spatial})$ ;
          // temporal processing
15           $h_{temporal} \leftarrow h_{spatial}$ ;
16          for  $k = 1$  to  $n_{GRU}$  do
17               $h_{temporal} \leftarrow \text{GRU}(h_{temporal})$ ;
18           $X_{pred} \leftarrow W_{proj}(h_{temporal})$ ;
19           $g_{pred} \leftarrow \text{RigidTransformDecode}(X_{pred})$ ;
20           $\mathcal{L}_{mse} \leftarrow \|X_{pred} - X_{(t+1)T:(t+2)T}\|^2$ ;
21           $\mathcal{L}_{clash} \leftarrow \text{ComputeClashScore}(X_{pred})$ ;
22           $\mathcal{L}_{total} \leftarrow \mathcal{L}_{total} + \mathcal{L}_{mse} + \lambda \mathcal{L}_{clash}$ ;
          // Update buffers
23           $r \sim \mathcal{U}(0, 1)$ ;
24          if  $r > p_{tf}$  then
25               $X_{buffer} \leftarrow X_{pred}$ ;
26               $g_{buffer} \leftarrow g_{pred}$ ;
27          else
28               $X_{buffer} \leftarrow X_{(t+1)T:(t+2)T}$ ;
29               $g_{buffer} \leftarrow g_{(t+1)T:(t+2)T}$ ;
30  Update model parameters  $\theta$  using  $\nabla_{\theta} \mathcal{L}_{total}$ ;
```

S2.4 Hierarchical Inference

Algorithm S2 details our hierarchical inference strategy. During inference, the slow-scale model f_{θ}^s predicts anchor conformations at the 20-ns sampling stride ($\Delta t_s = 20$ ns). These predictions provide slow-scale context for the fast-scale model f_{ϕ}^f , which generates intermediate frames at the 1-ns sampling stride ($\Delta t_f = 1$ ns). This hierarchy is designed to reduce drift during long autoregressive rollouts while retaining local variation between successive anchors.

Algorithm S2: Multiscale Inference for Conformational Sequence Generation

Input: initial frames $\{X_i\}_{i=0}^{T-1}, \{g_i\}_{i=0}^{T-1}, ;$ // T frames as initial buffer
1 slow-scale sampling stride $\Delta t_s = 20$ ns,
2 fast-scale sampling stride $\Delta t_f = 1$ ns,
3 requested rollout length N_{total} (frames),
4 trained slow-scale model f_{θ}^s ,
5 trained fast-scale model f_{ϕ}^f
Output: complete trajectory $\mathcal{X} = \{X_t\}_{t=0}^{N_{\text{total}}-1}$
6 $X_{\text{buffer}}^s \leftarrow \{X_i\}_{i=0}^{T-1};$
7 $g_{\text{buffer}}^s \leftarrow \{g_i\}_{i=0}^{T-1};$
8 $\mathcal{X} \leftarrow \{X_i\}_{i=0}^{T-1};$
9 **for** $t = 0$ **to** n_s **do**
10 $X_{\text{pred}}^s, g_{\text{pred}}^s \leftarrow f_{\theta}^s(X_{\text{buffer}}^s, g_{\text{buffer}}^s);$
11 $X_{\text{buffer}}^f \leftarrow X_{\text{pred}}^s;$
12 $g_{\text{buffer}}^f \leftarrow g_{\text{pred}}^s;$
13 Append X_{pred}^s to $\mathcal{X};$
14 **for** $k = 0$ **to** n_f **do**
15 $X_{\text{pred}}^f, g_{\text{pred}}^f \leftarrow f_{\phi}^f(X_{\text{buffer}}^f, g_{\text{buffer}}^f)$ // Model forward pass
16 Append X_{pred}^f to $\mathcal{X};$
17 Update $X_{\text{buffer}}^f, g_{\text{buffer}}^f$ with prediction;
18 Update $X_{\text{buffer}}^s, g_{\text{buffer}}^s$ with prediction;
19 **return** $\mathcal{X}$

S2.5 Model Architecture Parameters

Table S1 summarizes the architectural hyperparameters of TEMPO. Both the slow-scale and fast-scale models share the same architecture but are trained independently on trajectories sampled at their respective temporal resolutions.

The spatial encoding module employs Invariant Point Attention (IPA), which operates on SE(3) frames to extract rotation- and translation-equivariant geometric features. Each IPA layer combines scalar attention over residue embeddings with point attention in 3D space, enabling the model to reason about both sequence-level and spatial relationships. The multi-head self-attention layers that follow capture long-range inter-residue interactions beyond local neighborhoods.

The temporal encoding module uses gated recurrent units (GRUs) to model frame-to-frame dependencies. Hidden states provide finite-memory temporal conditioning across autoregressive prediction steps, distinguishing the output from independently sampled conformations.

The output layer projects the combined spatial-temporal representations back to the 13-dimensional per-residue state space (7D for SE(3) frame + 6D for torsion angles). During training, we employ teacher forcing to balance stability with the model’s ability to recover from its own prediction errors during inference.

Table S1: TEMPO model architecture and implementation parameters.

Component	Parameter	Value
<i>Spatial Encoding (IPA)</i>		
	Hidden dimension d	256
	Number of IPA layers	4
	Number of attention heads	8
	Query/key/value dimension	32
	Number of query points	8
	Number of value points	8
<i>Temporal Encoding (GRU)</i>		
	GRU hidden dimension	256
	Number of GRU layers	2
	Dropout rate	0.1
<i>Output Layer</i>		
	MLP hidden layers	[256, 128]
	Output dimension	13 (per residue)
	Activation	GELU
<i>Input Features</i>		
	Sequence embedding dim	128
	Maximum training sequence length	240 (full-length inference)
<i>Training and Inference</i>		
	Noise scales (training; fast inference)	[0.01, 1.0]; 5.0

Algorithm S1 samples noise over [0.01, 1.0] with beta shape $5(1 - e/E) + 1$, shifting from lower values toward uniform sampling as training proceeds. At inference, slow-scale noise remains in this range and the fast-scale model uses $\sigma_{\text{noise}} = 5.0$ for local diversity; these are nonphysical sampling hyperparameters. The 240-residue cap applies only to training, whereas inference uses full-length sequences.

S2.6 Endpoint-conditioned generation

For targeted transition analyses, generation is conditioned on a user-specified endpoint $\mathbf{X}^*$. A learnable encoder E_ψ maps the endpoint to a per-residue embedding $\mathbf{e}^* = E_\psi(\mathbf{X}^*) \in \mathbb{R}^{L \times d}$, which is added to the latent state before spatial and temporal encoding:

$$\mathbf{h}_{\text{latent}}^{\text{cond}} = \mathbf{h}_{\text{latent}} + \mathbf{e}^*. \quad (\text{S9})$$

The endpoint embedding is held fixed throughout inference. Thus, these rollouts are directed toward a specified structure and are analyzed as endpoint-conditioned conformational sequences rather than unbiased equilibrium trajectories.

S2.7 Inference-time steering of DFG-local geometry

For the DFG analysis, a local refinement step is applied after each autoregressive update without changing model parameters. The optimized region $\mathcal{R} = \{i_D - 1, i_D, i_F, i_G, i_G + 1\}$ contains the DFG triplet and one flanking residue on each side. Within $\mathcal{R}$, C α translations and sine-cosine representations of ϕ and ψ are optimized while quaternion orientations and all other residues remain fixed. The objective is

$$\mathcal{L} = \lambda_{\text{tors}}\mathcal{L}_{\text{tors}} + \lambda_{\text{pos}}\mathcal{L}_{\text{pos}} + \lambda_{\text{bond}}\mathcal{L}_{\text{bond}} + \lambda_{\text{ang}}\mathcal{L}_{\text{ang}} + \lambda_{\text{bnd}}\mathcal{L}_{\text{bnd}} + \lambda_{\text{clash}}\mathcal{L}_{\text{clash}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}}. \quad (\text{S10})$$

Torsion guidance. The torsion term biases the DFG residues toward the prespecified in-like or out-like Ramachandran targets (ϕ_r^*, ψ_r^*) :

$$\mathcal{L}_{\text{tors}} = \sum_{r \in \{D, F, G\}} w_r \left[\left(\frac{\Delta\phi_r}{180^\circ} \right)^2 + \left(\frac{\Delta\psi_r}{180^\circ} \right)^2 \right], \quad (\text{S11})$$

where $\Delta\phi_r$ and $\Delta\psi_r$ are periodic angular distances to the target values and $w_F = 2w_D = 2w_G$ gives greater weight to DFG-Phe.

Position displacement. For out-like steering, the DFG C α positions are additionally biased toward target displacements δ_r^* relative to the conformation entering the refinement step:

$$\mathcal{L}_{\text{pos}} = \sum_{r \in \{D, F, G\}} w_r \left\| \mathbf{t}_r - \mathbf{t}_r^{(0)} - \delta_r^* \right\|_2^2, \quad (\text{S12})$$

where $\mathbf{t}_r^{(0)}$ is the initial C α position for that refinement step. The target for DFG-Phe includes an approximately 4–5 Å C α displacement.

Local geometry and boundary continuity. Soft quadratic penalties constrain consecutive C α distances to 3.5–4.1 Å and C α –C α –C α angles to 90°–150°:

$$\mathcal{L}_{\text{bond}} = \sum_i [\text{ReLU}(3.5 - d_{i,i+1})^2 + \text{ReLU}(d_{i,i+1} - 4.1)^2], \quad (\text{S13})$$

$$\mathcal{L}_{\text{ang}} = \sum_i [\text{ReLU}(90^\circ - \theta_i)^2 + \text{ReLU}(\theta_i - 150^\circ)^2], \quad (\text{S14})$$

where the sums include consecutive bonds or angles within $\mathcal{R}$. The boundary term $\mathcal{L}_{\text{bnd}}$ applies the same distance limits to the connections between $\mathcal{R}$ and its fixed flanking residues.

Clash avoidance and temporal smoothness. Nonlocal C α pairs are penalized when separated by less than 3.0 Å:

$$\mathcal{L}_{\text{clash}} = \sum_{|i-j| \geq 3} \text{ReLU}(3.0 - d_{ij})^2. \quad (\text{S15})$$

The temporal term $\mathcal{L}_{\text{smooth}}$ penalizes large changes in the optimized translations and torsions relative to the preceding generated frame.

We minimized Eq. S10 with Adam for 300 iterations per generated step at a learning rate of 0.3. The weights were $\lambda_{\text{tors}} = 0.95$, $\lambda_{\text{pos}} = 5.0$, $\lambda_{\text{bond}} = 50.0$, $\lambda_{\text{ang}} = 10.0$, $\lambda_{\text{bnd}} = 30.0$, $\lambda_{\text{clash}} = 5.0$, and $\lambda_{\text{smooth}} = 0.01$; the position weight was increased progressively over three phases, and torsion sine-cosine pairs were renormalized after each update. The relatively large bond weight was used to limit local structural collapse during refinement. Accepted frames were identified using the prespecified DFG-Phe Ramachandran-basin and local-geometric criteria used throughout the analysis.

S3 Evaluation Metrics and Analysis Protocols

This section defines the quantitative metrics used throughout the manuscript. For the benchmark comparisons in Fig. 2B–C, we follow the ensemble-evaluation protocol used in AlphaFlow/ESMFlow, with metrics grouped into three categories: (i) flexibility, (ii) distributional agreement in Cartesian and principal-component spaces, and (iii) higher-order ensemble observables such as weak and transient contacts. We additionally define the analysis-specific metrics used in subsequent applications, including the contact-based Q-score for kinase cross-state comparisons and the B-factor agreement metrics for crystallographic flexibility. These metrics assess agreement with MD-derived ensemble statistics and provide descriptive comparisons of contact similarity and residue-level variability. Relative temporal ordering is evaluated separately for the ADK case using tICA (Section S5).

S3.1 Notation and preprocessing

For a protein with N evaluated atoms or residues, let

$$\mathcal{M} = \{M_t\}_{t=1}^{T_M}, \quad \mathcal{G} = \{G_s\}_{s=1}^{T_G} \quad (\text{S16})$$

denote the MD reference ensemble and generated ensemble, respectively. Each conformation is represented by Cartesian coordinates $X = \{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^3$. Unless otherwise specified, benchmark calculations use C α coordinates so that all compared methods are evaluated on the same atom set. If a metric is computed on backbone or all-atom coordinates, the same equations apply after replacing the evaluated atom set.

Before computing ensemble metrics, each frame is aligned to the same reference structure using the Kabsch algorithm. For ATLAS, the reference is the experimental structure used to initialize the MD simulation. For mdCATH, we use the first MD frame as the reference when no separate experimental initializer is provided. Residue indices are matched by sequence alignment when needed, and residues absent from any compared structure are excluded from all methods for that target.

The rigid-body-aligned RMSD between two conformations X and Y is

$$\text{RMSD}(X, Y) = \min_{R \in SO(3), \mathbf{t} \in \mathbb{R}^3} \left[\frac{1}{N} \sum_{i=1}^N \|R\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i\|_2^2 \right]^{1/2}. \quad (\text{S17})$$

For a collection of test proteins indexed by p , Pearson correlation between two target-level quantities a_p and b_p is computed as

$$r(a, b) = \frac{\sum_p (a_p - \bar{a})(b_p - \bar{b})}{\sqrt{\sum_p (a_p - \bar{a})^2} \sqrt{\sum_p (b_p - \bar{b})^2}}. \quad (\text{S18})$$

Unless otherwise noted, the aggregate number reported in the heatmaps is the mean across test proteins for distance metrics and the Pearson correlation or Jaccard similarity defined below for agreement metrics.

S3.2 Flexibility metrics

Pairwise RMSD diversity. For an ensemble $\mathcal{E} = \{E_t\}_{t=1}^T$, the overall conformational diversity is the average pairwise C α RMSD:

$$D_{\text{pRMSD}}(\mathcal{E}) = \frac{2}{T(T-1)} \sum_{1 \leq a < b \leq T} \text{RMSD}(E_a, E_b). \quad (\text{S19})$$

For each protein p , we compute $D_{\text{pRMSD}}(\mathcal{M}_p)$ and $D_{\text{pRMSD}}(\mathcal{G}_p)$. The heatmap entry labeled “RMSD r ” is the Pearson correlation across test proteins:

$$r_{\text{pRMSD}} = r(\{D_{\text{pRMSD}}(\mathcal{G}_p)\}_p, \{D_{\text{pRMSD}}(\mathcal{M}_p)\}_p). \quad (\text{S20})$$

This metric asks whether proteins that are more flexible in MD are also more flexible in the generated ensembles.

Root-mean-square fluctuation. For each residue or atom i , RMSF measures the spatial fluctuation around its ensemble-mean position after alignment:

$$\text{RMSF}_i(\mathcal{E}) = \left[\frac{1}{T} \sum_{t=1}^T \|\mathbf{x}_{t,i} - \bar{\mathbf{x}}_i\|_2^2 \right]^{1/2}, \quad \bar{\mathbf{x}}_i = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{t,i}. \quad (\text{S21})$$

We report two RMSF agreement summaries. First, the global RMSF correlation pools residue-level RMSF values across all proteins:

$$r_{\text{RMSF}}^{\text{global}} = r(\{\text{RMSF}_i(\mathcal{G}_p) : p, i\}, \{\text{RMSF}_i(\mathcal{M}_p) : p, i\}). \quad (\text{S22})$$

Second, the per-target RMSF correlation is computed within each protein and then summarized across proteins:

$$r_{\text{RMSF}}^{\text{target}}(p) = r(\{\text{RMSF}_i(\mathcal{G}_p)\}_{i=1}^{N_p}, \{\text{RMSF}_i(\mathcal{M}_p)\}_{i=1}^{N_p}). \quad (\text{S23})$$

The heatmap entry labeled “RMSF r ” reports the global RMSF correlation unless explicitly marked as per-target. This metric evaluates whether the generated ensemble places residue-level mobility in the same regions as MD.

S3.3 Distributional accuracy metrics

Root mean Wasserstein distance. Following AlphaFlow’s ensemble evaluation, we generalize RMSD to ensembles by fitting a three-dimensional Gaussian positional distribution to each evaluated atom or residue. For atom i in ensemble $\mathcal{E}$, let

$$\mathcal{N}_i^{\mathcal{E}} = \mathcal{N}(\boldsymbol{\mu}_i^{\mathcal{E}}, \boldsymbol{\Sigma}_i^{\mathcal{E}}) \quad (\text{S24})$$

be the Gaussian fitted to $\{\mathbf{x}_{t,i}\}_{t=1}^T$. The squared 2-Wasserstein distance between two three-dimensional Gaussians is

$$W_2^2(\mathcal{N}_1, \mathcal{N}_2) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 + \text{Tr} \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2 \left(\boldsymbol{\Sigma}_2^{1/2} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{1/2} \right)^{1/2} \right). \quad (\text{S25})$$

The root mean Wasserstein distance (RMWD) between generated and MD ensembles is

$$\text{RMWD}(\mathcal{G}, \mathcal{M}) = \left[\frac{1}{N} \sum_{i=1}^N W_2^2(\mathcal{N}_i^{\mathcal{G}}, \mathcal{N}_i^{\mathcal{M}}) \right]^{1/2}. \quad (\text{S26})$$

This quantity reduces to conventional RMSD when each ensemble contains a single structure. When reported, we also decompose RMWD into a mean-position contribution and a covariance contribution:

$$\text{RMWD}_{\text{mean}} = \left[\frac{1}{N} \sum_{i=1}^N \|\boldsymbol{\mu}_i^{\mathcal{G}} - \boldsymbol{\mu}_i^{\mathcal{M}}\|_2^2 \right]^{1/2}, \quad (\text{S27})$$

$$\text{RMWD}_{\text{var}} = \left[\frac{1}{N} \sum_{i=1}^N \text{Tr} \left(\boldsymbol{\Sigma}_i^{\mathcal{G}} + \boldsymbol{\Sigma}_i^{\mathcal{M}} - 2 \left((\boldsymbol{\Sigma}_i^{\mathcal{M}})^{1/2} \boldsymbol{\Sigma}_i^{\mathcal{G}} (\boldsymbol{\Sigma}_i^{\mathcal{M}})^{1/2} \right)^{1/2} \right) \right]^{1/2}. \quad (\text{S28})$$

Lower RMWD indicates closer agreement with the MD positional distribution.

PCA-space Wasserstein distance. To compare collective conformational distributions, each aligned conformation is flattened into an RMSD-normalized coordinate vector

$$\mathbf{z}_t = N^{-1/2} \text{vec}(E_t) \in \mathbb{R}^{3N}. \quad (\text{S29})$$

For the metric reported as “PCA W_2 ” or “PCA Diff”, we compute principal components from the MD ensemble alone. Let $U_2^{\mathcal{M}} \in \mathbb{R}^{3N \times 2}$ contain the first two MD principal directions. MD and generated conformations are projected onto the same two-dimensional coordinate system:

$$\mathbf{u}_t^{\mathcal{M}} = (U_2^{\mathcal{M}})^\top (\mathbf{z}_t^{\mathcal{M}} - \bar{\mathbf{z}}^{\mathcal{M}}), \quad \mathbf{u}_s^{\mathcal{G}} = (U_2^{\mathcal{M}})^\top (\mathbf{z}_s^{\mathcal{G}} - \bar{\mathbf{z}}^{\mathcal{M}}). \quad (\text{S30})$$

The empirical 2-Wasserstein distance between the two projected ensembles is

$$\widehat{W}_{2,\text{PCA}}(\mathcal{G}, \mathcal{M}) = \left[\min_{\boldsymbol{\pi} \in \Pi(\mathbf{a}, \mathbf{b})} \sum_{s=1}^{T_G} \sum_{t=1}^{T_M} \pi_{st} \|\mathbf{u}_s^{\mathcal{G}} - \mathbf{u}_t^{\mathcal{M}}\|_2^2 \right]^{1/2}, \quad (\text{S31})$$

where $\mathbf{a}$ and $\mathbf{b}$ are uniform weights over generated and MD frames, and $\Pi(\mathbf{a}, \mathbf{b})$ is the set of couplings with these marginals. Lower values indicate that the generated ensemble occupies the same dominant low-dimensional conformational basins as MD. A joint-PCA variant is computed analogously after fitting U_2 to the pooled MD and generated ensembles with equal weighting; unless specified, Fig. 2 reports the MD-PCA version.

Dominant principal-component similarity. The leading principal component describes the dominant direction of collective motion. For each protein, we compute the first principal component of the MD ensemble, $\mathbf{v}_1^{\mathcal{M}}$, and the first principal component of the generated ensemble, $\mathbf{v}_1^{\mathcal{G}}$, in the same aligned coordinate space. Because principal-component signs are arbitrary, we use unsigned cosine similarity:

$$s_{\text{PC}}(\mathcal{G}, \mathcal{M}) = \left| \frac{(\mathbf{v}_1^{\mathcal{G}})^\top \mathbf{v}_1^{\mathcal{M}}}{\|\mathbf{v}_1^{\mathcal{G}}\|_2 \|\mathbf{v}_1^{\mathcal{M}}\|_2} \right|. \quad (\text{S32})$$

A target is counted as successfully recovering the dominant collective motion if $s_{\text{PC}} > 0.5$. The heatmap entry “PC-sim” reports

$$\% \text{PC-sim} = 100 \times \frac{1}{P} \sum_{p=1}^P \mathbf{1}[s_{\text{PC}}(\mathcal{G}_p, \mathcal{M}_p) > 0.5], \quad (\text{S33})$$

where P is the number of test proteins. Higher values indicate better recovery of the MD dominant motion direction.

S3.4 Contact-dynamics metrics

Contact-based observables test whether the generated ensemble reproduces higher-order rearrangements of residue-residue interactions. We use C α –C α contacts with an 8 Å distance threshold, following the AlphaFlow MD-ensemble evaluation. Local sequence neighbors are excluded by requiring $|i - j| > 2$.

For each residue pair (i, j) and ensemble $\mathcal{E}$, the contact probability is

$$p_{ij}(\mathcal{E}) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}[\|\mathbf{x}_{t,i} - \mathbf{x}_{t,j}\|_2 < 8 \text{ Å}]. \quad (\text{S34})$$

Let $c_{ij}^0 = 1$ if residues i and j are in contact in the reference structure and $c_{ij}^0 = 0$ otherwise. With a fluctuation threshold $\tau = 0.10$, weak contacts are native/reference contacts that dissociate in more than 10% of ensemble frames:

$$\mathcal{W}(\mathcal{E}) = \{(i, j) : |i - j| > 2, c_{ij}^0 = 1, 1 - p_{ij}(\mathcal{E}) > \tau\}. \quad (\text{S35})$$

Transient contacts are non-native/reference contacts that form in more than 10% of ensemble frames:

$$\mathcal{T}(\mathcal{E}) = \{(i, j) : |i - j| > 2, c_{ij}^0 = 0, p_{ij}(\mathcal{E}) > \tau\}. \quad (\text{S36})$$

Agreement between generated and MD contact sets is measured with the Jaccard similarity

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (\text{S37})$$

Thus, the weak-contact and transient-contact metrics reported in the heatmaps are

$$J_{\text{weak}} = J(\mathcal{W}(\mathcal{G}), \mathcal{W}(\mathcal{M})), \quad J_{\text{transient}} = J(\mathcal{T}(\mathcal{G}), \mathcal{T}(\mathcal{M})). \quad (\text{S38})$$

If both sets are empty for a target, the target is omitted from the dataset average for that contact metric. Higher values indicate better recovery of contact dissociation or formation events observed in MD.

S3.5 Q-Score (Native Contact Fraction)

Q-score quantifies structural similarity based on the preservation of native inter-residue distances. For a structure $\mathbf{X}$ relative to a native reference $\mathbf{X}^{\text{ref}}$ with n residues:

$$Q(\mathbf{X}, \mathbf{X}^{\text{ref}}) = \frac{2}{(n-2)(n-3)} \sum_{j>i+2} \exp\left(-\frac{(d_{ij} - d_{ij}^{\text{ref}})^2}{2\sigma_{ij}^2}\right) \quad (\text{S39})$$

where d_{ij} and d_{ij}^{ref} are the $\text{C}\alpha$ - $\text{C}\alpha$ distances between residues i and j in the query and reference structures, respectively. The summation runs over all residue pairs with sequence separation $|j-i| > 2$.

The tolerance parameter σ_{ij} depends on sequence separation:

$$\sigma_{ij} = (1 + |i - j|)^{0.15} \text{ \AA} \quad (\text{S40})$$

$Q = 1$ indicates perfect agreement with the reference contact geometry. In the kinase cross-state analysis, we report the continuous score because active and inactive references may share many contacts; no universal Q-score threshold is used to infer a state transition.

S3.6 B-Factor Calculation and Comparison

B-factor from RMSF. We convert RMSF values from generated trajectories to B-factors using the isotropic approximation:

$$B_i = \frac{8\pi^2}{3} \cdot \text{RMSF}_i^2 \quad (\text{S41})$$

where RMSF_i is in $\text{\AA}$ and B_i is in $\text{\AA}^2$.

Z-score normalization. To compare profiles with different absolute magnitudes, we normalize using Z-scores:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (\text{S42})$$

where μ and σ are the mean and standard deviation of the B-factor profile across all residues.

Comparison metrics:

- *Spearman correlation* (ρ): Rank-based correlation between generated and experimental B-factor profiles.
- *Z-score RMSE*: Root-mean-square error between Z-score normalized profiles:

$$\text{RMSE}_z = \sqrt{\frac{1}{L} \sum_{l=1}^L (z_l^{\text{gen}} - z_l^{\text{exp}})^2} \quad (\text{S43})$$

For the crystallographic comparison, C α RMSF is computed after least-squares C α superposition, converted with Eq. S41, and compared with experimental C α B-factors. Spearman correlation evaluates residue-rank agreement, whereas Z-score RMSE evaluates the shape of the normalized profiles; neither metric equates crystallographic disorder with solution-state fluctuations.

S3.7 Kinase cross-state comparison protocol

For each of 104 kinases with paired active and inactive structures, we generated separate 400-frame rollouts from the two endpoints. Every active-initiated frame was scored against the inactive reference, and every inactive-initiated frame was scored against the active reference using Eq. S39. We report the mean and maximum opposite-state Q-score for each rollout.

To summarize starting-state dependence, let $\bar{Q}_{\text{active} \rightarrow \text{inactive}}$ and $\bar{Q}_{\text{inactive} \rightarrow \text{active}}$ denote the two mean opposite-state scores and define

$$|\Delta Q| = |\bar{Q}_{\text{active} \rightarrow \text{inactive}} - \bar{Q}_{\text{inactive} \rightarrow \text{active}}|. \quad (\text{S44})$$

For descriptive aggregation, we use three bins: consistent ($|\Delta Q| \leq 0.05$), moderate ($0.05 < |\Delta Q| \leq 0.10$), and directional ($|\Delta Q| > 0.10$). These bins quantify asymmetry between generated score distributions and are not interpreted as physical transition probabilities or barrier estimates.

Part II: Extended Experimental Results

S4 Multiscale versus Single-Scale Ablation

Autoregressive generation can accumulate error over long rollouts because every prediction becomes part of the context for subsequent updates. TEMPO addresses this problem with a hierarchy in which slow-scale predictions provide structural anchors and the fast-scale model generates intermediate variation conditioned on those anchors. To determine whether this decomposition provides a measurable benefit beyond the shared network architecture, we compared the full framework with a single-scale baseline that directly generates all 400 frames at the 1-ns training stride without hierarchical guidance.

Table S2 summarizes the comparison on the mdCATH test set. The multiscale model consistently outperforms the single-scale baseline across all evaluated metrics. Most notably, the long-horizon RMSD error is reduced from 8.62 Å to 1.78 Å, indicating substantially improved trajectory stability over extended autoregressive rollout. The multiscale design also yields much better agreement with MD-derived flexibility statistics, improving the Pearson correlation of pairwise RMSD from 0.14 to 0.77 and the global RMSF correlation from 0.15 to 0.67. In addition, conformational diversity and ensemble agreement are better preserved, with pairwise RMSD closer to the MD reference (2.78 versus 7.46, compared with ground truth 3.26) and lower root mean Wasserstein

distance (4.21 versus 8.27). Contact-dynamics agreement also improves, with weak-contact Jaccard similarity increasing from 0.23 to 0.43 and transient-contact similarity increasing from 0.09 to 0.20.

These results show that the benefit of the multiscale design is empirical rather than only conceptual: relative to direct single-scale generation, the hierarchy substantially limits long-rollout error while improving agreement with MD-derived ensemble geometry, flexibility, and contact observables.

Table S2: Ablation study comparing multiscale TEMPO with a single-scale baseline on mdCATH. Ground truth values are shown in parentheses where applicable. The single-scale model directly generates 400 frames without hierarchical guidance.

Metric	TEMPO (Multiscale)	TEMPO (Single-scale)
Pairwise RMSD (= 3.26)	2.78	7.46
Pairwise RMSD $r \uparrow$	0.77	0.14
All-atom RMSF (= 1.64)	1.60	4.27
Global RMSF $r \uparrow$	0.67	0.15
Root mean $\mathcal{W}_2 \downarrow$	4.21	8.27
MD PCA $\mathcal{W}_2 \downarrow$	2.33	2.53
Weak contacts $J \uparrow$	0.43	0.23
Transient contacts $J \uparrow$	0.20	0.09
RMSD Error $\downarrow$	1.78	8.62

S5 Adenylate Kinase Transition Analysis

S5.1 Time-lagged independent component analysis

We applied tICA to the same 10 independently generated ADK trajectories used in main-text Fig. 3, with 500 frames per trajectory. Frames were aligned to closed 1AKE using C α atoms in the ADK CORE, and the aligned C α coordinates of all residues formed the feature vector $\mathbf{x}_t$. Each seed was treated as a separate trajectory so that lagged pairs never crossed seed boundaries.

For a lag τ , let $\delta\mathbf{x}_t = \mathbf{x}_t - \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the pair-weighted mean. The symmetrized covariance matrices are

$$\mathbf{C}_0 = \frac{1}{2} \langle \delta\mathbf{x}_t \delta\mathbf{x}_t^T + \delta\mathbf{x}_{t+\tau} \delta\mathbf{x}_{t+\tau}^T \rangle_{\text{pairs}}, \quad (\text{S45})$$

$$\mathbf{C}_\tau = \frac{1}{2} \langle \delta\mathbf{x}_t \delta\mathbf{x}_{t+\tau}^T + \delta\mathbf{x}_{t+\tau} \delta\mathbf{x}_t^T \rangle_{\text{pairs}}. \quad (\text{S46})$$

The tICA modes satisfy the generalized eigenvalue problem

$$\mathbf{C}_\tau \mathbf{v}_i = \lambda_i \mathbf{C}_0 \mathbf{v}_i, \quad \xi_i(t) = \lambda_i \mathbf{v}_i^T \delta\mathbf{x}_t, \quad (\text{S47})$$

where ξ_i is the kinetic-map-scaled tIC coordinate. We used a lag of $\tau = 5$ generated frames, corresponding to five 1-ns intervals under the training-data indexing but not to a calibrated kinetic lag; the qualitative ordering of the leading mode was unchanged over lags of 1–10 frames.

The all-seed tICA fit identifies tIC1 as the common closed-to-open coordinate, whereas tIC2 mainly separates replicate-specific transverse routes. For clarity, main-text Fig. 3C displays the two complete trajectories closest to the all-seed median tIC2 path (seeds 4 and 7), but the tICA basis

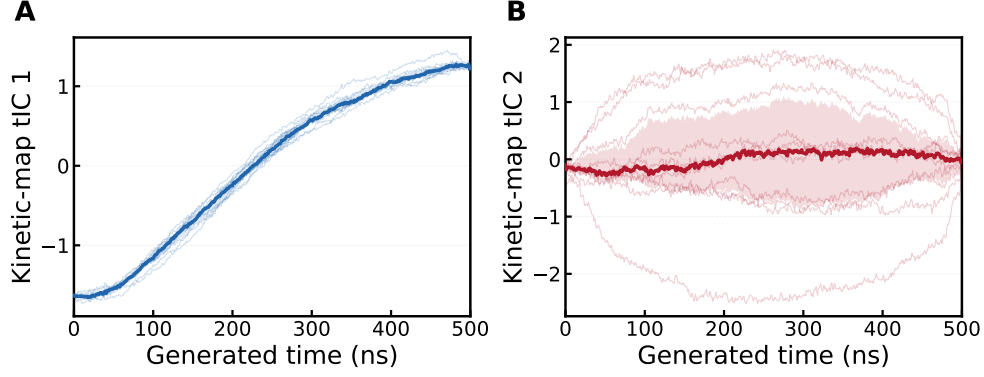


Figure S1: **Frame-wise progression of the leading ADK tICA coordinates across all 10 generated trajectories.** Thin lines denote individual trajectory seeds, dark lines denote the pointwise median, and shaded bands denote the interquartile range. tIC1 shows a reproducible closed-to-open progression with generated frame index, whereas tIC2 captures route-to-route variation around the common transition coordinate.

was estimated from all 10 trajectories. States 1–3 were identified by clustering in the endpoint-RMSD projection in main-text Fig. 3A; one representative frame selected from each cluster was then projected and marked in the tICA space, without redefining the states by clustering in tICA coordinates. The contour reports the relative occupancy

$$U_{\text{occ}}(\xi_1, \xi_2) = -\ln \left[\frac{P(\xi_1, \xi_2)}{P_{\text{max}}} \right], \quad (\text{S48})$$

rather than an equilibrium free energy.

Across all 10 trajectories, tIC1 progresses consistently with generated frame index (pooled Spearman $\rho_s = 0.994$), while tIC2 varies among seeds (Fig. S1). Together with the progressive contact exchange in main-text Fig. 3D, this result supports an ordered conformational sequence rather than an unordered collection of closed-, intermediate-, and open-like structures. Because the trajectories are endpoint-conditioned, we do not interpret the tICA eigenvalues as equilibrium relaxation times or as recovery of absolute kinetic rates.

S5.2 Residue-level flexibility

As a complementary local analysis, we calculated $C\alpha$ RMSF after structural superposition and converted it to an isotropic fluctuation-derived B-factor using Eqs. S21 and S41. The representative trajectory localizes the largest fluctuations to the NMP and LID domains while retaining a comparatively rigid CORE (Fig. S2). Its B-factor profile also places several flexible regions at positions similar to the crystallographic profile, although absolute magnitudes differ. We use this comparison only as a qualitative residue-level consistency check because crystallographic B-factors include crystal-packing, static-disorder, and refinement contributions.

S5.3 Strain-based PCA

Strain-based PCA provides a complementary view of the ADK transition by separating local mechanical deformation from rigid-body rotation and translation of the NMP and LID domains. ADK opening involves coordinated rearrangements of the LID and NMP domains and can proceed through multiple intermediate conformations [1, 2]. We applied the analysis to the same $M = 10$

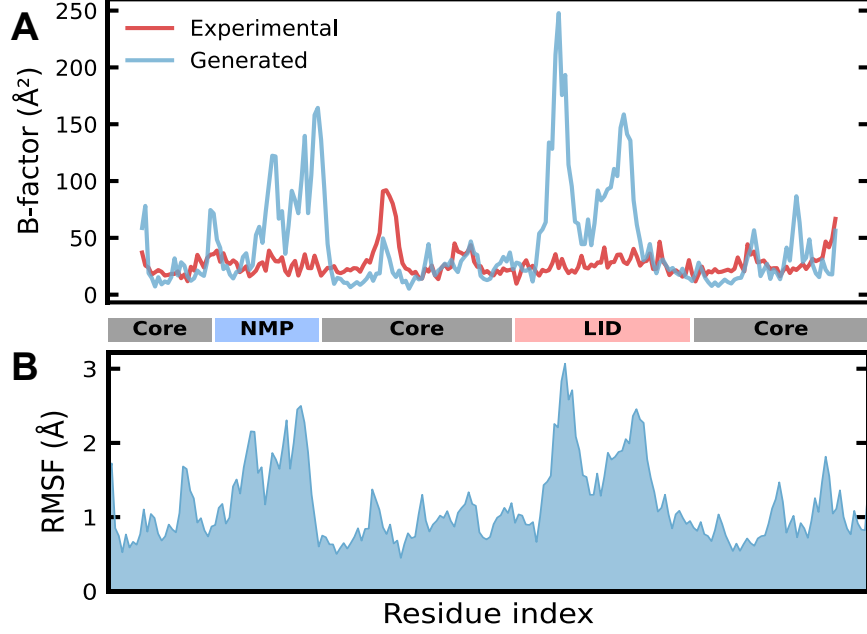


Figure S2: **Residue-level flexibility along a representative TEMPO-generated ADK transition.** (A) Per-residue B-factors derived from the generated trajectory and the crystallographic ADK structure. (B) Per-residue $C\alpha$ RMSF from the same trajectory, with CORE, NMP, and LID annotations.

endpoint-conditioned trajectories used in main-text Fig. 3, with $T = 500$ frames per trajectory. For a protein with L residues, let

$$\mathbf{r}_i^{m,t} \in \mathbb{R}^3 \quad (\text{S49})$$

denote the $C\alpha$ coordinate of residue i in replicate m at frame t , where $i = 1, \dots, L$, $m = 1, \dots, M$, and $t = 0, \dots, T - 1$. Closed 1AKE was used as the reference conformation, with reference coordinates $\mathbf{r}_i^0$.

For each residue, we defined a reference neighborhood from $C\alpha$ atoms within $d_c = 10$ Å:

$$\mathcal{N}_i = \left\{ j \neq i : \|\mathbf{r}_j^0 - \mathbf{r}_i^0\|_2 < d_c \right\}. \quad (\text{S50})$$

The reference and deformed neighbor vectors are

$$\mathbf{x}_{ij}^0 = \mathbf{r}_j^0 - \mathbf{r}_i^0, \quad \mathbf{y}_{ij}^{m,t} = \mathbf{r}_j^{m,t} - \mathbf{r}_i^{m,t}. \quad (\text{S51})$$

For every residue, frame, and replicate, the local deformation gradient was estimated by least squares using a local finite-strain formulation:

$$\mathbf{F}_i^{m,t} = \arg \min_{\mathbf{F} \in \mathbb{R}^{3 \times 3}} \sum_{j \in \mathcal{N}_i} w_{ij} \left\| \mathbf{y}_{ij}^{m,t} - \mathbf{F} \mathbf{x}_{ij}^0 \right\|_2^2, \quad (\text{S52})$$

where uniform weights w_{ij} were used. The corresponding Green–Lagrange strain tensor is

$$\mathbf{E}_i^{m,t} = \frac{1}{2} \left[\left(\mathbf{F}_i^{m,t} \right)^T \mathbf{F}_i^{m,t} - \mathbf{I} \right]. \quad (\text{S53})$$

Because this tensor removes the rotational component of the local affine transformation, it reports deformation of the residue neighborhood rather than overall translation or rigid-body domain rotation.

The six independent components of the symmetric strain tensor form a residue-level vector

$$\mathbf{e}_i^{m,t} = [E_{i,xx}^{m,t}, E_{i,yy}^{m,t}, E_{i,zz}^{m,t}, E_{i,xy}^{m,t}, E_{i,xz}^{m,t}, E_{i,yz}^{m,t}]. \quad (\text{S54})$$

Crucially, the residue vectors were concatenated within each conformation to obtain one frame-level feature vector,

$$\mathbf{z}^{m,t} = [\mathbf{e}_1^{m,t}, \mathbf{e}_2^{m,t}, \dots, \mathbf{e}_L^{m,t}] \in \mathbb{R}^{6L}. \quad (\text{S55})$$

Thus, residues are features of each frame rather than independent PCA samples. Each of the $6L$ features was standardized across the pooled set of $M \times T = 5,000$ conformations, and PCA was fitted to the resulting frame-by-feature matrix. The score of frame (m, t) on strain principal component k is

$$p_k^{m,t} = (\mathbf{z}^{m,t} - \bar{\mathbf{z}})^T \mathbf{v}_k, \quad (\text{S56})$$

where $\bar{\mathbf{z}}$ is the pooled mean and $\mathbf{v}_k \in \mathbb{R}^{6L}$ is the loading vector for component k .

Figure S3A shows the strain-PC scores $(p_1^{m,t}, p_2^{m,t})$ of the generated conformations. Points are colored by normalized transition progress,

$$\tau_t = \frac{t}{T-1}, \quad (\text{S57})$$

where $\tau_t = 0$ and 1 denote the beginning and end of each generated path, respectively. Early and late frames occupy different portions of the strain-PC projection, and the intervening frames form a progression between them. The first two strain PCs explain 31.2% and 13.2% of the variance, indicating that the leading components summarize a substantial, but not exhaustive, fraction of the local-deformation variability.

To identify which residues contribute to the dominant strain mode, we grouped the six PC1 loading components associated with each residue and calculated

$$\ell_i^{(1)} = \left[\sum_{c \in \{xx, yy, zz, xy, xz, yz\}} (v_{1,i,c})^2 \right]^{1/2}, \quad (\text{S58})$$

where $v_{1,i,c}$ is the PC1 loading of strain component c for residue i . The residue-wise profile in Fig. S3B therefore measures the contribution of each residue's local strain components to PC1; it is not a PCA performed independently over pooled residues. Prominent contributions occur in the NMP and LID regions and at NMP-CORE and LID-CORE interfaces, whereas most of the CORE has smaller loading magnitude. This pattern is consistent with domain opening coupled through localized hinge and interface deformation rather than uniform deformation of the CORE scaffold, and with prior evidence that localized conformational fluctuations contribute to ADK allostery [3, 4].

For the residue-time map, scalar local strain was calculated as the Frobenius norm

$$s_i^{m,t} = \|\mathbf{E}_i^{m,t}\|_F = \left[\sum_{\alpha, \beta \in \{x, y, z\}} (E_{i, \alpha\beta}^{m,t})^2 \right]^{1/2}, \quad (\text{S59})$$

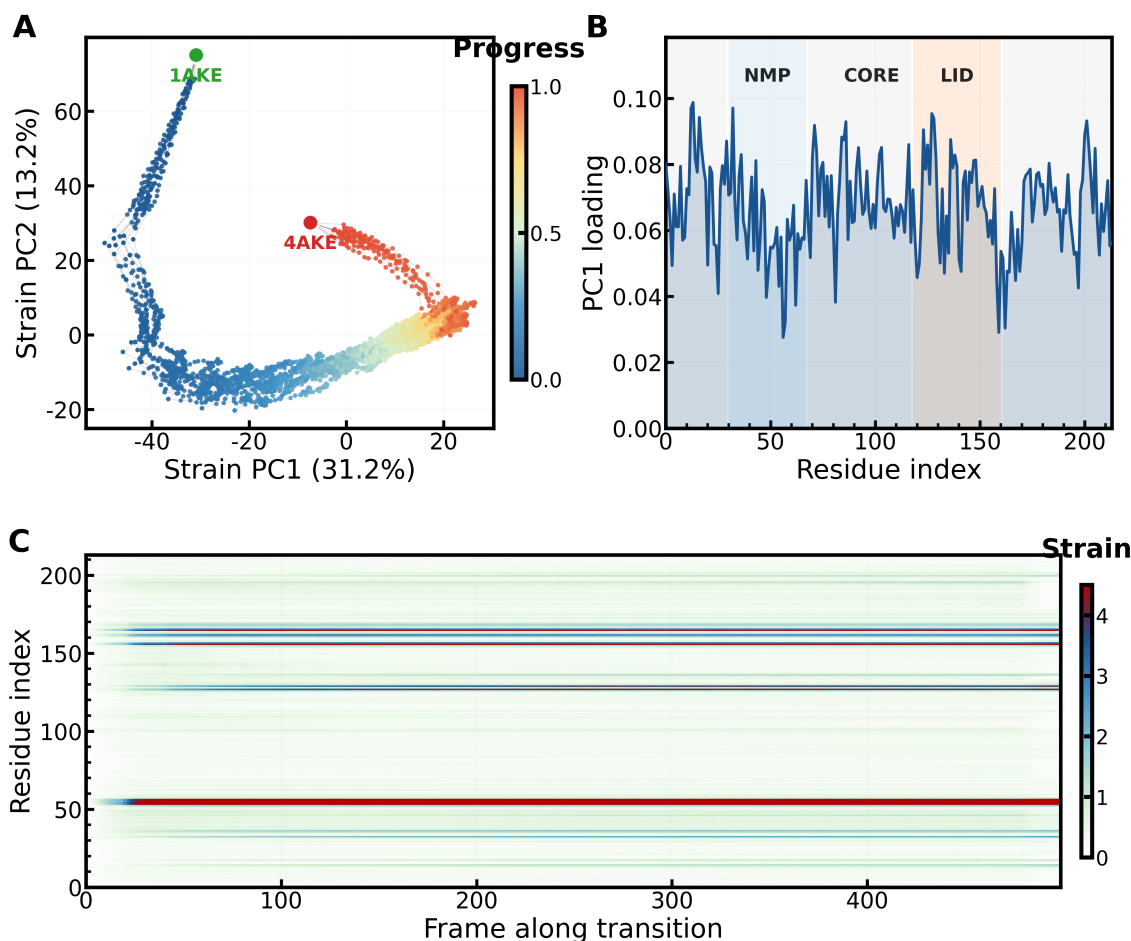


Figure S3: **Strain-based PCA of the TEMPO-generated ADK transition ensemble.** (A) Projection of the 5,000 generated conformations onto the first two strain PCs, colored by normalized transition progress; the components explain 31.2% and 13.2% of the variance. Large green and red markers denote the closed 1AKE and open 4AKE references projected using the same feature transformation. (B) Residue-wise strain-PC1 loading magnitude, obtained by combining the six PC1 loading components assigned to each residue, with CORE, NMP, and LID annotations. (C) Residue-time map of the Frobenius norm of the local Green-Lagrange strain tensor, averaged across the 10 trajectory seeds at each generated-frame index.

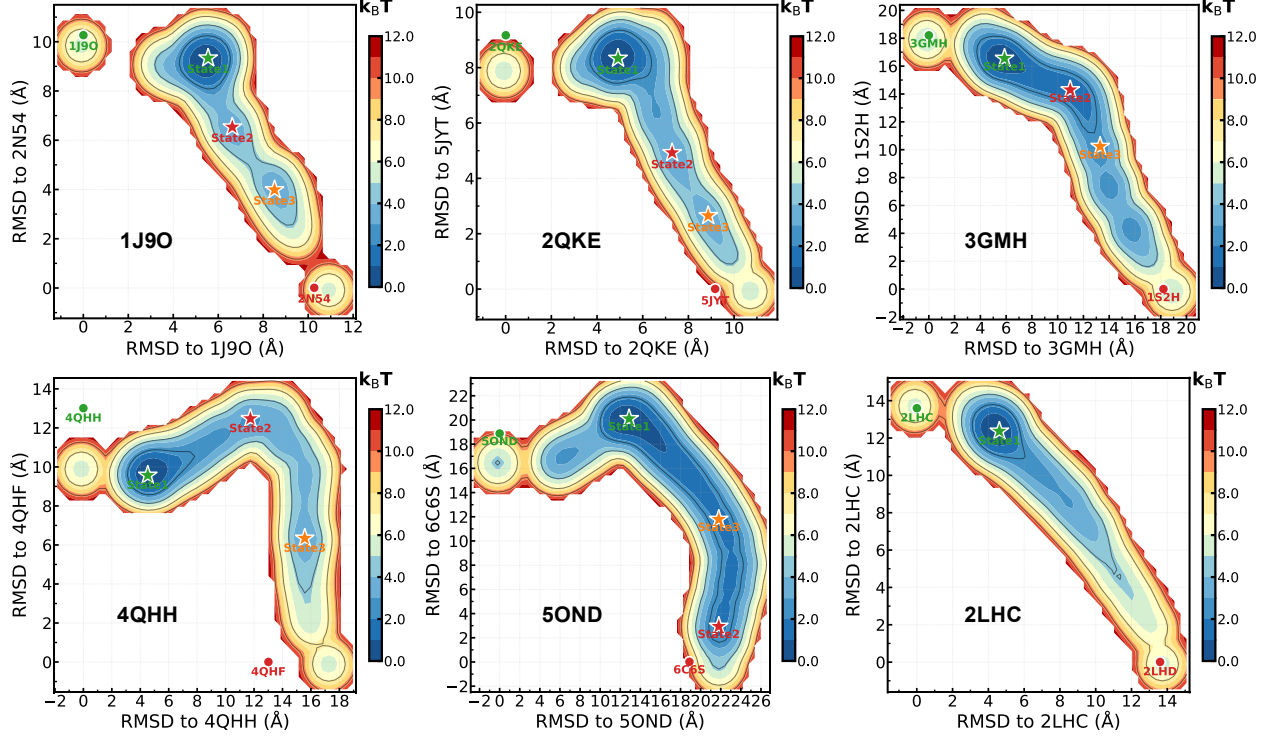


Figure S4: Projected relative occupancy for six additional endpoint-conditioned systems. The axes are RMSDs to the two endpoint structures (green and red circles), and stars mark representative generated conformations at intermediate RMSD combinations. Color denotes the dimensionless occupancy-derived score $-\ln(P/P_{\max})$ and is not interpreted as an equilibrium free energy. Top row: 1J9O $\leftrightarrow$ 2N54, 2QKE $\leftrightarrow$ 5JYT, and 3GMH $\leftrightarrow$ 1S2H. Bottom row: 4QHH $\leftrightarrow$ 4QHF, 5OND $\leftrightarrow$ 6C6S, and 2LHC $\leftrightarrow$ 2LHD.

and averaged across independent trajectories at each frame index:

$$\bar{s}_i^t = \frac{1}{M} \sum_{m=1}^M s_i^{m,t}. \quad (\text{S60})$$

The resulting heat map (Fig. S3C) localizes persistent deformation to specific residue bands, particularly within the NMP/LID regions and their interfaces with the CORE, rather than showing diffuse strain throughout the protein. Together, the strain-PC projection, residue-wise loading profile, and residue-time map provide a local-deformation complement to the global tICA progression and contact exchange in main-text Fig. 3.

S6 Additional Endpoint-Conditioned Transitions

We applied the endpoint-conditioned procedure in Section S2 to six systems beyond ADK, initiating each rollout from one experimental structure and conditioning on its paired endpoint. Figure S4 projects the resulting conformations onto RMSD to the two endpoints. All six cases contain generated frames spanning the endpoint-defined coordinate range and connected regions of intermediate RMSD, including the larger excursions in 3GMH $\leftrightarrow$ 1S2H and 5OND $\leftrightarrow$ 6C6S. The representative structures mark positions along these projected corridors; they are not assigned as metastable states.

1J9O↔2N54. The generated density forms a bent corridor between the two endpoint-defined regions. The representative states follow the turn in this corridor: State 1 lies in the broad high-occupancy region closer to the 1J9O side of the projection, whereas States 2 and 3 occupy progressively more 2N54-like RMSD combinations.

2QKE↔5JYT. This projection also contains a curved, connected corridor. Its width is greatest around State 1 and narrows as the generated conformations progress through States 2 and 3 toward the 5JYT-defined end of the coordinate range. The projected path therefore resolves intermediate combinations of the two endpoint RMSDs rather than placing generated frames only near the endpoints.

3GMH↔1S2H. The two endpoint structures are separated by an RMSD range approaching 20 Å. Generated conformations populate an elongated diagonal-to-curved corridor across this range, and the three representative states mark early, middle, and later portions of the projected progression. Compared with the first two systems, the occupied region is broader and contains greater variation transverse to the endpoint-connecting direction.

4QHH↔4QHF. The projected density follows a pronounced curved path: it first extends mainly along RMSD to 4QHH and then turns toward lower RMSD to 4QHF. States 1–3 sample successive portions of this bend, showing that the two endpoint distances do not change in a simple linear proportion along the generated path.

5OND↔6C6S. This system spans the largest endpoint-RMSD range in the panel. The generated conformations form a broad arc with several regions of elevated occupancy, extending from the 5OND side through States 1 and 3 to the 6C6S-proximal region containing State 2. The non-monotonic geometry of the arc indicates substantial variation in the two endpoint distances across the generated ensemble; it is not used to assign a unique sequence of microscopic structural events.

2LHC↔2LHD. The generated density is concentrated in a comparatively narrow diagonal corridor between the endpoint-defined regions. The marked representative state lies in its highest-occupancy portion, while the remaining generated conformations extend continuously toward the 2LHD side of the projection.

The shape and width of the occupied corridor differ among systems, indicating system-dependent variation in the endpoint-conditioned rollouts. Because these ensembles are generated under fixed endpoint conditioning and are not Boltzmann-weighted samples, the projected densities do not determine equilibrium populations, barrier heights, interconversion rates, or a unique physical mechanism. They provide a qualitative extension of the ADK result by showing that the same procedure generates conformations spanning intermediate RMSD combinations for multiple endpoint pairs.

S7 Cross-State Q-Score Patterns

Figure S5 compares the mean opposite-state Q-scores from active- and inactive-initiated rollouts using the protocol in Section S3. Across 104 kinases, 63% have $|\Delta Q| \leq 0.05$, 22% have $0.05 < |\Delta Q| \leq 0.10$, and 15% have $|\Delta Q| > 0.10$. Thus, most proteins show similar mean cross-state contact similarity in the two rollout directions, whereas a subset shows stronger starting-state dependence.

S8 Exploratory GPCR Transfer Cases

Of the six GPCR systems analyzed (two from each of Classes A–C), the four not displayed in the main text are C-C chemokine receptor type 5 bound to maraviroc, corticotropin-releasing factor receptor 1 bound to CP-376395, apo corticotropin-releasing factor receptor 1, and metabotropic glutamate receptor 5 bound to 2-[2-(3-methoxyphenyl)ethynyl]-6-methyl-pyridine. For each system,

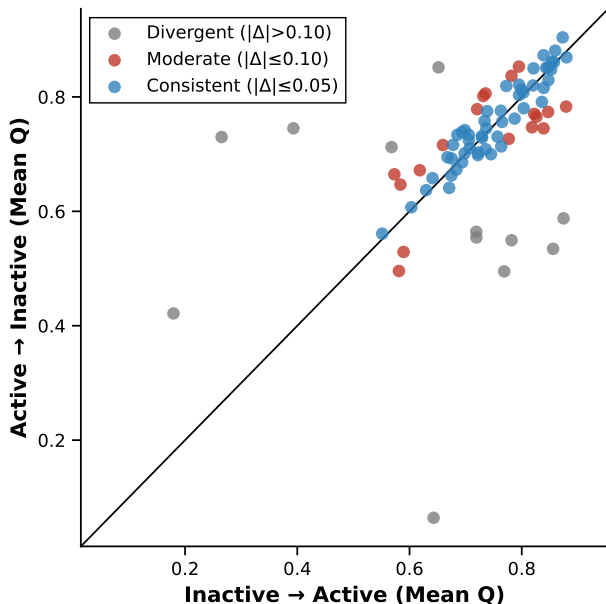


Figure S5: Starting-state dependence of opposite-state Q-scores for 104 human kinases. Each point represents one kinase, with the mean Q-score of inactive-initiated frames against the active reference on the x -axis and the mean Q-score of active-initiated frames against the inactive reference on the y -axis. Colors denote the descriptive $|\Delta Q|$ bins defined in Section S3; the diagonal denotes equal mean scores in the two directions.

nonprotein components were removed from the reference trajectory, 500 receptor-only MD frames spanning 500 ns were retained, and TEMPO generated a 500-frame backbone rollout from the first processed frame without GPCR-specific parameter updates. Principal components were fitted to the corresponding receptor-only MD ensemble, and generated conformations were projected into the same PC1–PC2 coordinate system.

Generated conformations overlap portions of the corresponding reference projection in all four cases (Fig. S6). For C-C chemokine receptor type 5 and CP-376395-bound corticotropin-releasing factor receptor 1, the generated ensembles extend across multiple connected regions of the MD-derived projection rather than collapsing to a single point. The apo corticotropin-releasing factor receptor 1 ensemble covers the dominant central region and part of its surrounding projected density, whereas the metabotropic glutamate receptor 5 ensemble is more localized. Across the representative conformations, the largest structural differences occur mainly in extracellular loops and terminal segments, while the transmembrane backbone remains comparatively stable. Together with the two main-text examples, these results provide an exploratory test of receptor-backbone transfer. The small panel, receptor-only representation, and two-dimensional projection do not establish recovery of ligand-, membrane-, solvent-, or ion-dependent GPCR dynamics.

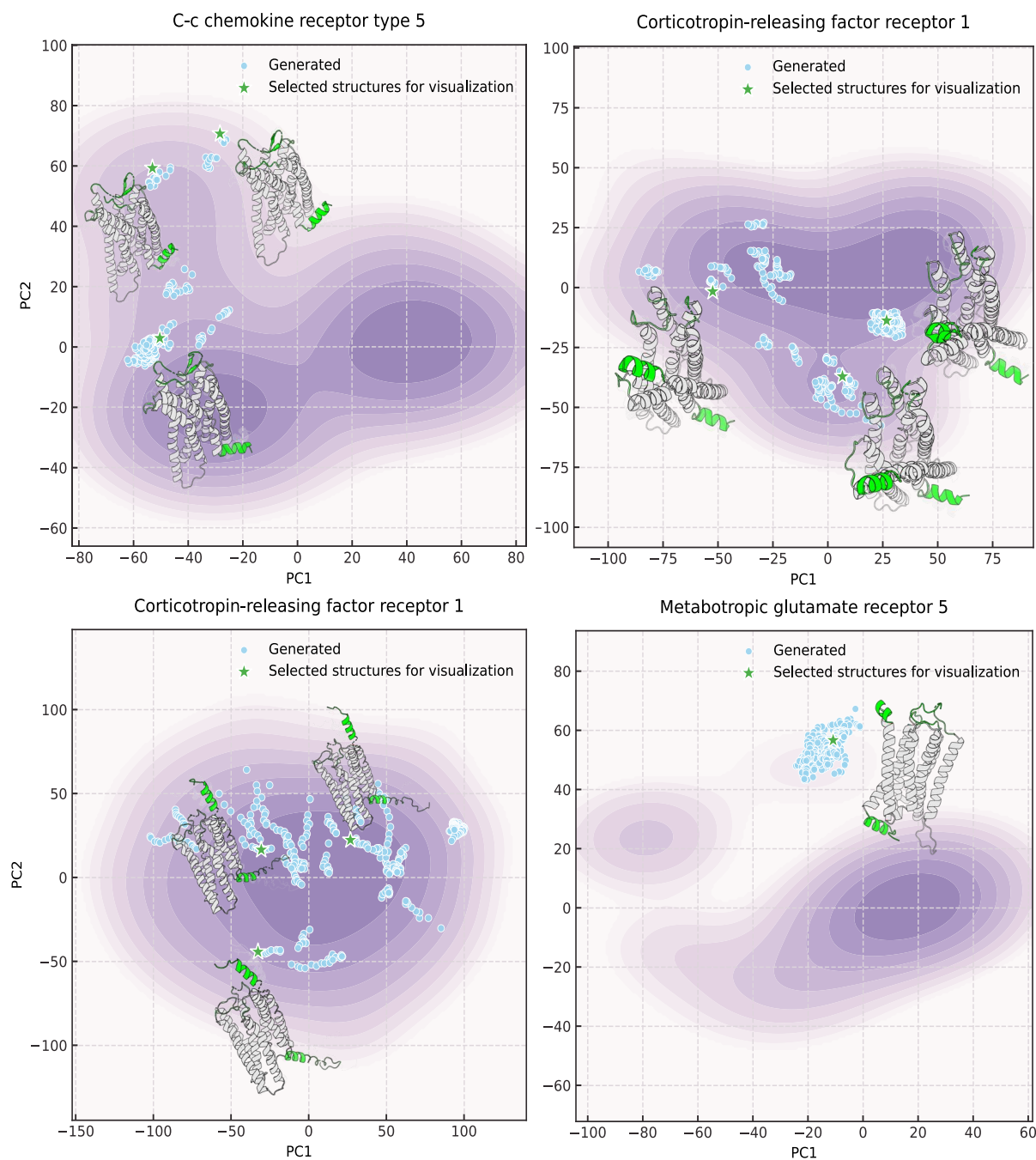


Figure S6: Additional exploratory GPCR transfer cases. Purple contours denote the projected density of each receptor-only MD reference in PC1-PC2 space, blue points denote TEMPO-generated conformations, and green stars mark representative generated conformations. Top left: C-C chemokine receptor type 5 bound to maraviroc. Top right: corticotropin-releasing factor receptor 1 bound to CP-376395. Bottom left: apo corticotropin-releasing factor receptor 1. Bottom right: metabotropic glutamate receptor 5 bound to 2-[2-(3-methoxyphenyl)ethynyl]-6-methyl-pyridine.

References

- [1] Paul C Whitford, Shachi Gosavi, and José N Onuchic. Conformational transitions in adenylate kinase: allosteric communication reduces misligation. *Journal of Biological Chemistry*, 283(4): 2042–2048, 2008.
- [2] Dechang Li, Ming S Liu, and Baohua Ji. Mapping the dynamics landscape of conformational transitions in enzyme: the adenylate kinase case. *Biophysical journal*, 109(3):647–660, 2015.
- [3] Osamu Miyashita, José Nelson Onuchic, and Peter G Wolynes. Nonlinear elasticity, protein-quakes, and the energy landscapes of functional transitions in proteins. *Proceedings of the National Academy of Sciences*, 100(22):12570–12575, 2003.
- [4] Travis P Schrank, D Wayne Bolen, and Vincent J Hilser. Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. *Proceedings of the National Academy of Sciences*, 106(40):16984–16989, 2009.