RNADYN: A BENCHMARK FOR GENERATING AND UNDERSTANDING RNA DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Ribonucleic acid (RNA) functions through conformational changes that are not fully captured by static structures. However, large-scale standardized RNA dynamics data remain limited, and existing approaches typically treat trajectory generation and dynamics understanding as separate objectives. Here, we introduce **RNADynBench**, a standardized RNA molecular dynamics (MD) benchmark with 2585 quality-controlled 100-ns all-atom trajectories and leakage-controlled splits. Building on RNADynBench, we develop **RNADynNet**, a unified model for RNA dynamics learning that uses a shared backbone for both trajectory generation and dynamics fingerprint extraction from a single conformer. It combines coordinate denoising, single-frame-to-trajectory alignment, and physical grounding to connect all-atom trajectory generation with dynamics representation learning. Physical grounding improves both generated dynamics and the physical information recoverable from these fingerprints. Across both test sets, including the high-flexibility challenge set, the generated trajectories achieve RMSF correlations of 0.875 and 0.766, while single-conformer predictions show comparable agreement with MD-derived dynamics. RNADynBench and RNADynNet together establish a benchmark and unified modeling framework for generating and understanding RNA dynamics.

1 INTRODUCTION

AI-driven structure prediction has transformed molecular biology, enabling accurate prediction of biomolecular structures directly from sequence (Jumper et al., 2021). Yet biological function depends not only on static structures but also on the conformational dynamics that connect multiple molecular states. Ribonucleic acid (RNA) exemplifies this challenge due to its essential roles in regulating gene expression, recognizing molecular partners, and catalyzing biochemical reactions (Ganser et al., 2019; Joshi et al., 2025), and accurate modeling remains challenging.

On one hand, progress in RNA modeling is constrained by the limited availability and diversity of experimental structural data. Experimentally resolved RNA structures are both less abundant and less diverse than protein structures, and only a small fraction of RNA families have been structurally characterized (Szikszai et al., 2024; Kagaya et al., 2025). Moreover, transformer-based architectures such as AlphaFold (Jumper et al., 2021) rely on multiple-sequence alignments (MSAs), for which identifying and aligning sufficiently diverse RNA homologs pose challenges distinct from those in protein modeling (Magnus et al., 2025; Kagaya et al., 2025). While recent advances in RNA and RNA-containing complex prediction demonstrate the potential of learning-based approaches, they still leave substantial challenges regarding conformational coverage and dynamics-aware modeling (Kagaya et al., 2025; Abramson et al., 2024). Together, these advances and limitations underscore a need for more RNA-specific data and evaluation (Szikszai et al., 2024).

A further challenge is dynamics. Functionally relevant RNA states can be transient, and interactions with ions, ligands, and proteins can reshape their structural conformations (Dethoff et al., 2012; Yadav et al., 2022). In our size-matched comparison with the ATLAS and Dynamic PDB protein MD collections (Vander Meersche et al., 2024; Liu et al., 2024), RNA trajectories show greater lag-dependent displacement and generally larger local fluctuations and pairwise RMSD over the sampled timescales (see Fig. 1**c**–**e** and Fig. 5). A study of bound and unbound RNA–protein complexes likewise found that RNA often undergoes greater conformational rearrangement than proteins (Barik

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

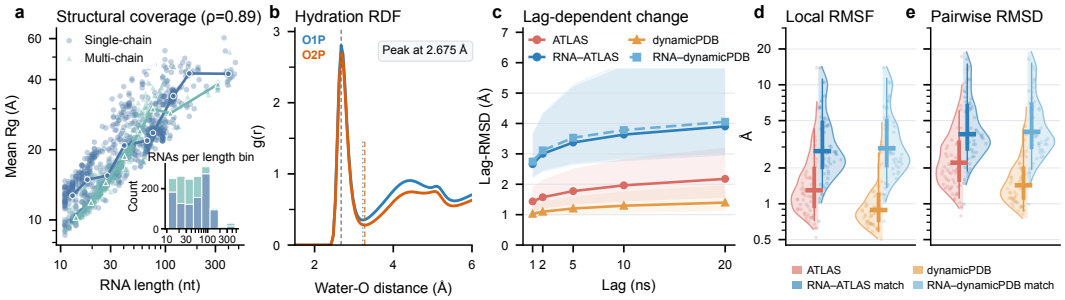


Figure 1: **RNADynBench visualized.** **a**, Mean R_g versus RNA residue count. Points show individual RNAs, lines connect length-binned medians, and the inset shows counts; ρ is Spearman correlation between RNA length and mean R_g. **b**, Median phosphate–water radial distribution functions (RDFs); shading shows the interquartile range, and vertical lines mark the first peak and minima. **c**, Weighted median lag-RMSD for size-matched RNA–protein comparisons; shading shows the weighted interquartile range (IQR). **d**,**e**, Weighted distributions of mean local RMSF (**d**) and median pairwise RMSD (**e**). Split violins show protein left and matched RNA right, with median, IQR, and 5th–95th percentiles. RNA uses C4′ atoms and protein uses Cα atoms.

et al., 2012). These observations suggest that inductive biases for modeling RNA can benefit from conditioning on molecular motion rather than on static structure alone.

All-atom molecular dynamics (MD) simulates atomic motion using forces computed from physics-based molecular potentials. The resulting trajectories resolve conformational changes over time and allow quantification of fluctuations and correlated motions over sampled timescales (Šponer et al., 2018). Learning-based studies have begun to exploit such trajectories, suggesting that MD simulations can provide supervision for learning molecular dynamics. Studies in proteins already show that incorporating dynamic and ensemble information can improve downstream prediction, motivating analogous approaches for RNA (Guo et al., 2025; Kalifa et al., 2026). Yet recent developments have primarily focused on proteins, leaving the molecular motion and folding of nucleic acids largely unexplored. This is because **standardized RNA MD datasets remain scarce** (see Tab. 1), and their construction requires both computational resources and expertise in molecular simulation. Generating trajectories consumes substantial GPU resources, while preparing RNA systems, selecting force fields and solvent conditions, and establishing equilibration protocols requires domain expertise and careful physical modeling (Šponer et al., 2018).

To address this gap, we introduce **RNADynBench**, a standardized RNA MD benchmark, and **RNA-DynNet**, a unified model that demonstrates the capabilities enabled by such data. RNADynBench provides the training foundation for RNADynNet, supplying coordinate targets for generation, trajectory-derived targets for representation alignment, and MD-derived statistics. Motivated by evidence that representation guidance can improve generation (Yu et al., 2025; Huang et al., 2026) and RNA's pronounced conformational dynamics, we introduce physical grounding (PG) to train intermediate representations to predict MD-derived covariance and inter-residue motion coupling. Our experiments show that physical grounding improves both generated dynamics and learned dynamics representations. Our **contributions** are fourfold:

- **RNADynBench.** A standardized RNA MD benchmark with 2,585 quality-controlled 100-ns all-atom trajectories, leakage-controlled splits, and protocols for generation and dynamics prediction.
- **RNADynNet.** A unified framework coupling RNA trajectory generation and single-conformer dynamics learning through alignment and shared physical supervision.
- **All-atom RNA trajectory generation benchmark.** We establish an RNA MD generation benchmark and show that RNADynNet captures geometry, fluctuations, and conformational distributions across structural and high-flexibility tests.
- **Transferable dynamics representations.** Single-conformer fingerprints recover MD-derived fluctuations and motion couplings and transfer to RNA–ligand binding prediction.

We will make our benchmark and implementation publicly available upon publication.

2 RNADynBench: An RNA Molecular Dynamics Benchmark

We introduce RNADynBench, a standardized benchmark for RNA trajectory generation and single-conformer dynamics prediction. It comprises 2,585 all-atom trajectories across 1,469 RNAs, totaling

Table 1: Comparison of RNA MD resources.

Resource	Scale	Simulation time	Task
BIGNASim (2016) (Hospital et al., 2016)	14 traj	15.6 μs total	Database and analysis
CASP15 MD (Nithin et al., 2025)	9 RNAs / 61 traj	300 ns each (18.3 μs total)	Structure refinement
DeepRMSF (Feng et al., 2026c)	371 RNAs	200 ns each (74.2 μs total)	RMSF prediction
RNADynBench	**1,469 RNAs / 2,585 traj**	100 ns each (**258.5** μs total)	**MD generation / prediction**

258.5 μs of simulation. The trajectories are accompanied by molecular mappings, simulation metadata, and quality-control records, and are partitioned into leakage-controlled splits. RNADynBench defines common, architecture-agnostic evaluation protocols for both tasks. Tab. 1 positions its scale and intended tasks alongside existing RNA MD resources; discussed further in Sec. 5.

Data Curation. RNADynBench is constructed from experimentally resolved RNA structures retrieved from RNA-Solo, a repository of cleaned PDB-derived RNA structures (Adamczyk et al., 2022). For each entry, we retrieve the original PDB/mmCIF records by PDB ID to recover its molecular context before curation. We retain the RNA-Solo target chain or chains. Additional RNA chains are retained only if their chain assignments are reliable, they contact the target within 5 Å (heavy atoms) after non-RNA components are removed, and they meet the partner-size limit. When partners are retained, the combined RNA length is limited to 512 nucleotides. Other RNA chains, proteins, DNA, and organic ligands are removed, with component-selection decisions recorded in the metadata. We intentionally exclude structures with unresolved backbone gaps, unsupported chemistry, or untraceable topology changes. Missing atoms and modified nucleotides are thus only handled when residue identity and topology remain unambiguous. Canonical atom and residue mappings link source structures, simulation topologies, and released coordinates while preserving chain identifiers, residue indices, nucleotide identities, and atom names.

Molecular dynamics simulation. Each curated static structure is simulated using all-atom molecular dynamics (MD) to generate a trajectory. MD evolves atomic coordinates by computing interatomic forces from a molecular force field and integrating Newton's equations of motion; the molecular topology remains fixed, so the simulation captures conformational motion rather than chemical reactions. We simulate each RNA in explicit OPC water using GROMACS and the Amber RNA OL3 force field, with charge neutralization and 150 mM NaCl. Potential-energy minimization removes unfavorable initial contacts, followed by staged equilibration under controlled temperature and pressure. A final 5-ns NPT equilibration at 300 K and 1 bar is excluded from the released trajectories. Each system then undergoes a 100-ns production simulation at 300 K and 1 bar using the Nosé-Hoover thermostat (Hoover, 1985) and Parrinello-Rahman barostat (Parrinello & Rahman, 1981), with coordinates saved every 100 ps (1,001 frames). Water and ions used in simulations are omitted from the released trajectories. Further simulation details are given in Sec. A.1.

Quality control and benchmark splits. We validate each simulation output by checking that the full trajectory completed successfully and by screening for coordinate jumps, backbone discontinuities, periodic-boundary artifacts, and chain separation in multi-chain systems. Simulations showing sustained chain separation are excluded, and the separated RNA components are treated as independent systems and re-simulated. A central challenge for RNA dynamics learning is avoiding evaluation leakage from closely related RNA structures and trajectories. Therefore, we group related systems by RNA identity, source structure, replicate lineage, known trajectory aliases, and sequence and structural similarity, and assign each group entirely to one partition. The `test_struct` split evaluates structural generalization, while `test_flex` evaluates generalization to highly flexible systems. The release contains 2,248 training trajectories, 102 validation trajectories, 117 trajectories in `test_struct`, and 100 in `test_flex`, plus 18 homology-guard trajectories reserved for auditing. Detailed quality-control and split-construction procedures are provided in Sec. A.2. These splits are shared by two benchmark tasks: generating 100-ns trajectories from an initial conformer and predicting MD-derived residue covariance, RMSF, and normalized motion coupling (NMC) from a single conformer.

Dataset characteristics. Fig. 1**a** shows broad structural coverage across RNA sizes and chain compositions, with mean R_g increasing strongly with RNA length ($\rho = 0.89$). Radial distribution functions (RDFs) between the non-bridging phosphate oxygens O1P/O2P and water oxygen (Ow) had the same median first-peak position of 2.675 Å and median first-minimum positions of 3.225 and 3.275 Å, respectively (Fig. 1**b**). Both RDFs showed a broad secondary maximum near 4.4 Å. Median first-shell coordination numbers were 2.95 and 2.85 water molecules per non-bridging phos-

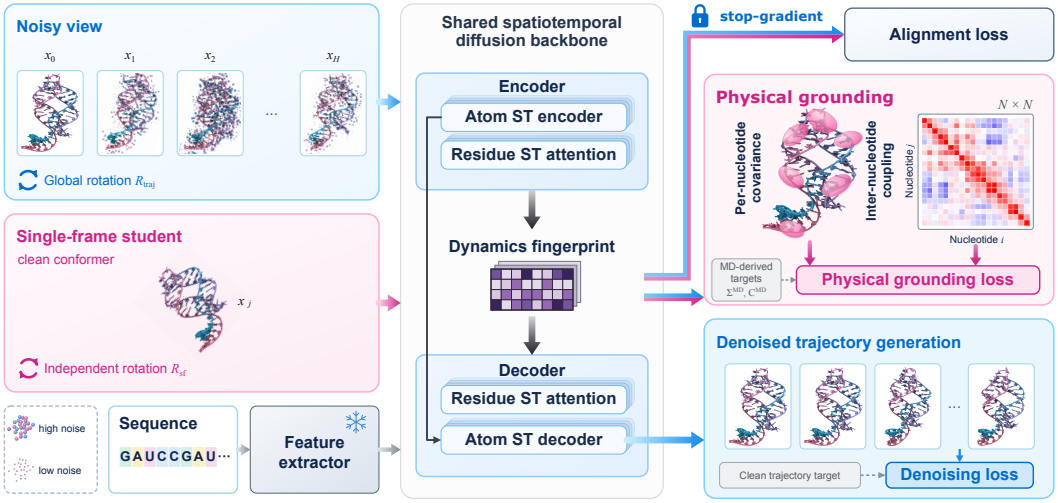


Figure 2: **Overview of RNADynNet.** RNADynNet jointly learns trajectory generation and single-conformer dynamics representations from MD trajectories. During training, the trajectory view (blue) provides temporal context for denoising and alignment, while the single-conformer view (pink) learns a dynamics representation from one conformer. Physical grounding supervises both views with MD-derived dynamical statistics. At inference, a single conformer supports two modes: the denoising path generates a trajectory, whereas the encoder readout directly yields a dynamics fingerprint. ST denotes spatiotemporal.

phate oxygen for O1P and O2P, respectively. The first peak and minimum, together with the broad secondary maximum, are consistent with RDF patterns reported in prior nucleic-acid simulations (Kundu et al., 2020; Perepelytsya et al., 2021), supporting the physical fidelity of RNADynBench.

Beyond structural diversity, RNA in RNADynBench exhibits greater conformational motion than proteins. To control for the size dependence of RMSD and RMSF, we reweighted RNA and protein systems to a shared heavy-atom-size distribution before comparison. Across both references, RNA showed consistently higher lag-RMSD and broader local RMSF and pairwise RMSD distributions over 1–20 ns (Fig. 1**c–e**). These observations highlight the need for both standardized RNA dynamics data and dynamics-aware inductive biases for RNA modeling.

3 RNADynNet: Unifying RNA MD Understanding and Generation

Building on RNADynBench, we introduce RNADynNet, a unified model for RNA trajectory generation and single-conformer dynamics representation learning. A shared spatiotemporal diffusion backbone supports both trajectory generation from an initial conformer and dynamics fingerprint extraction from a single conformer. Training couples the two paths through trajectory-informed alignment and MD-derived physical supervision. Fig. 2 summarizes our framework.

Let $\mathbf{X} = (\mathbf{x}_0, \ldots, \mathbf{x}_{L-1}) \in \mathbb{R}^{L\times A\times 3}$ be a trajectory window of L frames, A atoms, and N RNA residues. The frames are separated by Δt and cover a temporal horizon $H = (L-1)\Delta t$. RNADynNet supports trajectory generation from an initial conformer $\mathbf{x}_0$ and deterministic encoding of a single conformer $\mathbf{x}$:

$$\widehat{\mathbf{X}}_{1:L-1} \sim p_{\bar\theta}(\mathbf{X}_{1:L-1} \mid \mathbf{x}_0, H, \Delta t), \qquad \mathbf{H}^{\mathrm{FP}}_{\bar\theta} = \mathcal{E}_{\bar\theta}(\mathbf{x}). \tag{1}$$

Here θ denotes the online parameters and $\bar\theta$ their exponential moving average (EMA). The distribution $p_{\bar\theta}$ is induced by the EMA denoiser and sampler, while $\mathcal{E}_{\bar\theta}$ denotes the shared backbone truncated at the intermediate block ℓ_{fp} and applied to a single conformer. Training uses θ for the single-frame pass, whereas fingerprint extraction uses $\bar\theta$. The EMA fingerprint $\mathbf{H}^{\mathrm{FP}}_{\bar\theta} \in \mathbb{R}^{N\times d}$ is read from ℓ_{fp}; each row $\mathbf{h}^{\mathrm{FP}}_{\bar\theta,i} \in \mathbb{R}^d$ is the dynamics fingerprint of residue i, with d the feature dimension.

Training objective. RNADynNet is trained by multi-objective optimization with the losses:

$$\mathcal{L}_{\mathrm{RNADynNet}} = \mathcal{L}_{\mathrm{gen}} + \lambda_{\mathrm{align}}\mathcal{L}_{\mathrm{align}} + \lambda_{\Sigma}\mathcal{L}_{\Sigma} + \lambda_{C}\mathcal{L}_{C}. \tag{2}$$

where $\mathcal{L}_{\mathrm{gen}}$, $\mathcal{L}_{\mathrm{align}}$, $\mathcal{L}_{\Sigma}$, and $\mathcal{L}_{C}$ denote the losses for trajectory generation, single-frame-to-trajectory alignment, residue covariance, and inter-residue motion coupling, respectively. The coef-

ficients λ_{align}, λ_{Σ}, and λ_C weight the corresponding losses; the last two weighted terms constitute physical grounding (PG). Next, we provide details on the individual loss terms.

3.1 Trajectory and Single-Frame Training

Trajectory learning with heterogeneous noise. We apply heterogeneous coordinate noise as continuous partial masking to encourage contextual learning across RNA residues and frames (Chen et al., 2024). Heavily perturbed coordinates carry less information about the underlying conformer, whereas less perturbed positions retain more reliable structural cues. This information asymmetry encourages the denoiser to use cleaner spatial and temporal context to reconstruct noisier regions, thereby exploiting dependencies across residues and frames.

Noise scales are assigned at frame–residue positions and shared by the atoms of each residue, forming the tensor $\boldsymbol{\sigma}$, such that $\widetilde{\mathbf{X}} = \mathbf{R}_{\text{traj}}\left(\mathbf{X} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}\right)$ where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Each window is aligned and centered before corruption, and one global rotation $\mathbf{R}_{\text{traj}} \in \text{SO}(3)$ acts on coordinates and noise across all frames. Training windows vary in start, duration, and sampling interval; observed frames use a near-clean noise scale for conditioning. The denoiser reconstructs the clean trajectory in the same rotated coordinate frame with the *coordinate-denoising loss*:

$$\mathcal{L}_{\text{denoise}} = \mathbb{E}\left\|\mathcal{D}_\theta(\widetilde{\mathbf{X}}, \boldsymbol{\sigma}, H, \Delta t) - \mathbf{R}_{\text{traj}}\mathbf{X}\right\|_w^2. \tag{3}$$

The weighted coordinate MSE $\|\cdot\|_w^2$ covers valid atoms in all sampled frames, including observed frames, and incorporates the EDM and generative-group weights. The generation objective augments this loss with auxiliary supervision for local structure, atomic and pairwise-distance fluctuations, and short-lag temporal correlations. Its terms and weights are specified in Sec. B.3. These auxiliary losses act on generated coordinates, whereas PG supervises intermediate representations.

Single-frame-to-trajectory alignment. The goal is to infer trajectory-level information from a single experimentally available conformer, rather than merely encode its static geometry. MD trajectories provide temporal context during training, but this context is unavailable when only one conformer is observed. We therefore train the single-conformer encoder to predict residue features from the trajectory pass, transferring information about RNA motion into its latent states.

For each forward window, we sample an anchor frame $J \sim \text{Uniform}\{0, \cdots, L-1\}$ to pair the trajectory and single-frame views at the same conformer $\mathbf{x}_J$. Both training states below correspond to this anchor; we omit J from their notation. From the online trajectory pass, we read the residue state $\mathbf{h}_{\theta,i}^{\text{traj}}$ at intermediate block ℓ_{fp} for frame J. Motivated by layer-dependent representation transferability (Skean et al., 2025), we use $\ell_{\text{fp}} = 17$ in the reported model and evaluate an alternative Block 24 readout and supervision depth in Sec. D.1. This state incorporates information from the full trajectory window through temporal attention. After normalization and stop-gradient, it gives the alignment target $\mathbf{z}_i^{\text{traj}}$. A separate pass through the shared online encoder receives the clean conformer $\mathbf{R}_{\text{sf}}\mathbf{x}_J$, producing $\mathbf{H}_\theta^{\text{FP}} = \mathcal{E}_\theta(\mathbf{R}_{\text{sf}}\mathbf{x}_J)$, whose ith row is $\mathbf{h}_{\theta,i}^{\text{FP}}$. The rotations $\mathbf{R}_{\text{traj}}$ and $\mathbf{R}_{\text{sf}}$ are sampled independently and uniformly from $\text{SO}(3)$. Although the backbone is not explicitly SE(3)-equivariant, centering and rotation augmentation encourage equivariant trajectory generation and orientation-invariant dynamics fingerprints.

We map the single-frame states to the trajectory targets using a predictor q_ψ conditioned on H and Δt:

$$\mathcal{L}_{\text{align}} = 1 - \mathbb{E}_{J,i} \cos\left(q_\psi(\mathbf{h}_{\theta,i}^{\text{FP}}, H, \Delta t), \mathbf{z}_i^{\text{traj}}\right). \tag{4}$$

Here cos denotes cosine similarity, and the expectation averages over sampled anchors and valid residues. Alignment gradients update the single-frame encoder and predictor, while the trajectory target is detached. The target reuses the existing online trajectory pass, requiring no separate target network or additional trajectory pass. Fingerprint extraction with $\mathcal{E}_{\bar{\theta}}$ requires neither H nor Δt.

Backbone implementation. The backbone uses an AF3-style spatial denoiser (Abramson et al., 2024) implemented and initialized with Protenix (Team et al., 2025), with the upstream feature extractor (Input Embedder and Pairformer) frozen. Static features from the frozen Input Embedder and Pairformer are provided to the backbone but omitted from network arguments for notational simplicity. Each spatiotemporal block applies spatial attention followed by temporal self-attention with a Langevin-inspired temporal decay bias (Feng et al., 2026a).

Table 2: Trajectory generation on RNADynBench. Bold marks the best.

Method	test_struct							test_flex						
	Bond MAE↓	O3P max↓	Chiral flip↓	Clash /1k↓	RMSF r ↑	PCA W_2 ↓	GPU time	Bond MAE↓	O3P max↓	Chiral flip↓	Clash /1k↓	RMSF r ↑	PCA W_2 ↓	GPU time
Reference	0.03308	0.1938	0	0	0.9819	0.7295	8.1h	0.03307	0.1966	0	0	0.9633	3.1137	13.9h
ConfRover	0.19315	120.4747	0.03103	581.4520	0.0666	9.3105	340.4s	0.16506	133.0420	0	419.7848	0.2288	11.1922	501.5s
MDGen	0.08024	15.1488	0.03103	95.9398	0.3737	2.4200	7.2s	0.10786	18.9945	$< \mathbf{10^{-5}}$	88.9058	0.2994	8.9282	11.0s
BioKinema	0.02612	4.4682	0.03836	2.0746	**0.8770**	1.9276	40.2s	0.02950	8.3586	0.02012	5.8241	0.7536	5.9406	47.8s
w/o PG	**0.02418**	2.4455	0.03097	1.1884	0.8594	1.9085	/	**0.02555**	9.1171	0.00160	**2.8416**	0.7440	6.1744	/
RNADynNet	0.02424	**1.8858**	**0.03085**	**0.8225**	0.8746	**1.8739**	24.5s	0.02567	**4.2400**	0.00093	3.0404	**0.7658**	**5.9169**	32.9s

3.2 PHYSICAL GROUNDING

Alignment transfers trajectory-contextualized information, but does not explicitly tie the learned representations to physical observables. Motivated by evidence that representation guidance can improve generation (Yu et al., 2025) and RNA's pronounced conformational dynamics, we use PG to explicitly shape intermediate representations with MD-derived residue covariance and inter-residue motion coupling. PG losses update the shared backbone through both trajectory-anchor and single-conformer representations.

MD-derived physical targets. We use residue covariance and inter-residue motion coupling as complementary descriptors of RNA dynamics (Bafna et al., 2026). Targets are computed from a fixed 101-frame, 100 ns reference trajectory, independently of the sampled training-window duration. Each residue is represented by its C1′ position, and all reference frames are rigidly aligned to the initial conformer. Writing $\delta\mathbf{r}_i = \mathbf{r}_i - \langle \mathbf{r}_i \rangle$, we define

$$\mathbf{\Sigma}_i = \langle \delta\mathbf{r}_i \delta\mathbf{r}_i^\top \rangle, \quad \mathrm{RMSF}_i = \sqrt{\mathrm{tr}(\mathbf{\Sigma}_i)}, \qquad C_{ij} = \langle \delta\mathbf{r}_i^\top \delta\mathbf{r}_j \rangle, \quad \rho_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}. \tag{5}$$

Here $\langle\cdot\rangle$ denotes the average over reference conformations, $\mathbf{\Sigma}_i$ captures residue-level fluctuation magnitude and direction, while C_{ij} and ρ_{ij} describe pairwise motion coupling and its normalized form (NMC), respectively. The labels satisfy $C_{ii} = \mathrm{tr}(\mathbf{\Sigma}_i)$. These MD statistics serve as supervision, not encoder inputs.

Physical grounding losses. For residue covariance, a shared MLP head f_Σ predicts the six independent entries of a symmetric 3×3 log-covariance matrix. For residue i, we construct an orthonormal frame $\mathbf{R}_i^{(J)}$ from the clean anchor conformer $\mathbf{x}_J$ and express the $\mathbf{\Sigma}_i$ in this frame: $\mathbf{\Sigma}_{i,\mathrm{loc}}^{(J)} = (\mathbf{R}_i^{(J)})^\top \mathbf{\Sigma}_i \mathbf{R}_i^{(J)}$. For each view $v \in \{\mathrm{FP}, \mathrm{traj}\}$, let $\mathbf{h}_{\theta,i}^v$ denote the online single-conformer or corresponding trajectory-anchor state. Both views use the same head and clean-anchor target:

$$\mathcal{L}_\Sigma = \frac{1}{2|\mathcal{V}|} \sum_{v\in\{\mathrm{FP},\mathrm{traj}\}} \sum_{i\in\mathcal{V}} \left\| f_\Sigma(\mathbf{h}_{\theta,i}^v) - \log\left(\mathbf{\Sigma}_{i,\mathrm{loc}}^{(J)}\right) \right\|_{\mathrm{F,mean}}^2 . \tag{6}$$

Here $\mathcal{V}$ contains the valid residues and $\|\cdot\|_{\mathrm{F,mean}}^2$ is the mean squared matrix-entry error. Matrix logarithms act on covariance values expressed numerically in Å^2. At inference, the matrix exponential yields a positive-definite covariance, recovered in global coordinates as $\widehat{\mathbf{\Sigma}}_i = \mathbf{R}_i \exp\big(f_\Sigma(\mathbf{h}_{\theta,i}^{\mathrm{FP}})\big)\mathbf{R}_i^\top$. Local-frame construction and numerical details are given in Sec. B.4.

For inter-residue motion coupling, the head f_C predicts a symmetric positive-definite motion-coupling matrix $\widehat{\mathbf{C}}$ from intermediate residue states. Following Viliuga et al. (2026), we supervise its matrix pattern, magnitude, and normalized off-diagonal coupling with

$$\mathcal{L}_C = 1 - \frac{\langle \widehat{\mathbf{C}}, \mathbf{C} \rangle_F}{\|\widehat{\mathbf{C}}\|_F \|\mathbf{C}\|_F} + \frac{\alpha}{|\mathcal{V}|^2} \sum_{i,j} |\widehat{C}_{ij} - C_{ij}| + \frac{\beta}{|\mathcal{V}|(|\mathcal{V}|-1)} \sum_{i\neq j} (\widehat{\rho}_{ij} - \rho_{ij})^2. \tag{7}$$

Here α and β weight the magnitude and normalized off-diagonal coupling terms, respectively. The construction of f_C and numerical details are given in Sec. B.4.

At inference, trajectory generation uses the EMA denoiser with a hierarchical forecasting-and-interpolation sampler, with alignment and PG heads omitted. Single-conformer inference uses the encoder alone for fingerprint extraction, with PG heads used only for direct descriptor prediction.

4 EXPERIMENTS

We evaluate RNADynNet on three tasks: trajectory generation, single-conformer dynamics prediction, and RNA–ligand binding transfer. Generation and dynamics prediction use the standard `test_struct` and `test_flex` splits from RNADynBench. Experimental protocols and supplementary results are provided in Secs. C and D.

4.1 RNA TRAJECTORY GENERATION UNDER DYNAMICS-AWARE EVALUATION

Existing molecular dynamics generation methods have primarily been developed for proteins rather than RNA. We therefore adapt ConfRover (Shen et al., 2025), MDGen (Jing et al., 2024b), and BioKinema (Feng et al., 2026a) to RNA and retrain them on RNADynBench under the same 12h training limit. The adaptations are described in Sec. C.3. Each method receives only the initial conformer and generates 101-frame trajectories spanning 100 ns. For each test initial conformer, we generate four trajectories, except ConfRover ($K = 1$) due to its slow sampling. The Reference row reports MD-derived metric values and MD simulation time. Tab. 2 depicts local geometric quality, agreement with MD dynamics, and sampling efficiency. We report bond and O3′–P errors, sugar-chirality flips, steric clashes, and MD-reference PCA-W_2 distance. We also report Pearson correlation across residues between generated-sample-mean and MD C1′ RMSF profiles. Bond, O3′–P, and chirality metrics use the initial conformer as reference. Distances are in Å, flips in percent, clashes per 1,000 heavy atoms, and runtimes are measured per trajectory on H200 GPUs. Metric details are given in Sec. C.4.

RNADynNet improves all four reported geometric metrics over BioKinema on both test sets, while achieving comparable or better RMSF correlation, lower PCA-W_2, and faster sampling. In particular, the maximum O3′–P deviation and clash rate are reduced by roughly half on both test sets. MDGen samples faster but shows substantially larger backbone deviations, more clashes, and weaker dynamic fidelity. ConfRover performs worse on most metrics and is much slower to sample; its lower training throughput under the common 12h budget may also have limited convergence.

Physical grounding also improves generation. Relative to w/o PG, the full model improves RMSF correlation, O3′–P continuity, and PCA-W_2 on both panels, with similar bond MAE. Clash rates decrease on `test_struct` and change modestly on `test_flex`. Overall, MD-derived physical supervision improves local backbone geometry and conformational dynamics.

4.2 LEARNING RNA DYNAMICS FROM SINGLE CONFORMERS

A central goal of RNADynBench is to enable learning of molecular dynamics properties from experimentally available static conformers. We therefore evaluate whether RNADynNet can recover MD-derived dynamics without trajectory generation at inference.

Models receive a single RNA conformer and predict MD-derived residue covariance, RMSF, and NMC. Native heads measure direct prediction accuracy, whereas independently trained probes on frozen encoders measure the dynamics information contained in learned representations. Tab. 3 compares RNADynNet and its ablations with RNA-FM (Chen et al., 2022), Geometry-GNN, RNA-FM+2D GCN, BackFlip (Viliuga et al., 2026), and DynaProt (Bafna et al., 2026). RMWD and symmetric KL quantify overall covariance mismatch, while ellip-

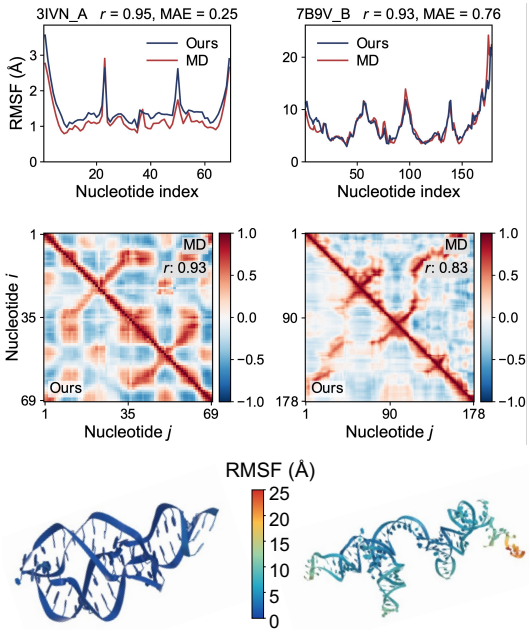


Figure 3: Dynamics inference from single conformers. Predicted per-nucleotide RMSF and pairwise nucleotide motion couplings from RNADynNet are compared with MD references for two RNAs. Motion-coupling matrices show RNADynNet below the diagonal and MD above; structures are colored by RMSF.

Table 3: Single-conformer prediction of MD-derived RNA dynamics. The MD reference compares two non-overlapping 50-ns MD blocks; model predictions are evaluated against full 100-ns MD statistics. † denotes probes on frozen representations. Bold denotes the best result.

	test_struct							test_flex						
	RMSF		Covariance			NMC		RMSF		Covariance			NMC	
Method	r ↑	MAE↓	RMWD↓	Sym. KL↓	IoU↑	r ↑	MAE↓	r ↑	MAE↓	RMWD↓	Sym. KL↓	IoU↑	r ↑	MAE↓
MD reference	0.8755	0.3770	0.7960	0.5525	0.5497	0.8532	0.1462	0.7584	1.6415	3.5505	1.3471	0.4185	0.7520	0.2145
RNA-FM†	0.2939	1.6188	2.1173	4.9868	0.1731	0.3690	0.2644	0.2085	3.8148	5.0390	4.2367	0.2318	0.3635	0.2952
Geometry-GNN	0.2801	1.1241	1.5796	2.8662	0.2561	0.4821	0.2591	0.3385	3.4276	4.6870	3.1959	0.3054	0.4442	0.2910
RNA-FM+2D GCN	0.1257	2.0117	2.6303	8.5914	0.1831	0.4797	0.2569	0.1536	3.5511	4.8377	4.9398	0.2625	0.5333	0.2609
BackFlip	0.8098	0.4608	0.9449	0.7599	0.4919	0.7948	0.1707	0.7311	1.7974	3.7118	4.4011	0.4058	0.7399	0.2084
DynaProt	0.7726	0.5266	0.9786	0.8288	0.4502	0.2454	0.2775	0.5518	2.5861	3.9757	1.9398	0.3692	0.1889	0.3132
w/o align. & PG†	0.7473	0.6056	1.0756	0.9827	0.3966	0.7924	0.1717	0.6232	2.1515	3.8137	1.3343	0.3811	0.7506	0.2088
w/o PG†	0.7797	0.5160	0.9839	0.7542	0.4540	0.7834	0.1744	0.6920	1.8983	3.5810	1.1055	0.4069	0.7443	0.2101
RNADynNet†	0.8626	0.4076	0.8741	0.5823	**0.5276**	0.7958	0.1694	0.7782	1.7560	3.3359	0.8775	0.4522	0.7577	0.2035
RNADynNet	**0.8665**	**0.3983**	**0.8651**	**0.5743**	0.5260	**0.8115**	**0.1629**	**0.7827**	**1.5800**	**3.2700**	**0.8533**	**0.4528**	**0.7614**	**0.2027**

soid IoU measures agreement in fluctuation anisotropy and orientation. RMSF and NMC capture per-residue fluctuation magnitude and pairwise motion coupling, respectively, and are evaluated by Pearson r and MAE. Details are given in Sec. C.4.

RNADynNet's native heads outperform BackFlip and DynaProt on all seven reported metrics in both test panels. RMSF correlations reach 0.8665 and 0.7827, while NMC correlations reach 0.8115 and 0.7614 on `test_struct` and `test_flex`, respectively. Fig. 3 illustrates predicted RMSF and motion couplings for two RNAs. Frozen-probe ablations show that alignment improves all RMSF and covariance metrics on both test sets, although NMC is slightly worse. Adding physical grounding improves all seven metrics on both sets. These results suggest that alignment transfers trajectory-derived information to the single-conformer representation, while physical grounding strengthens its recovery of MD-derived dynamics.

Scaling RNA dynamics learning with MD supervision. To examine how trajectory data contribute to dynamics learning, we train RNADynNet on five fractions of the training data spanning 10–100% and evaluate checkpoints from 1k to 6k steps (Fig. 4). At later checkpoints, larger subsets generally reduce NMC MAE and increase ellipsoid IoU, suggesting that additional MD supervision improves the recovery of RNA dynamics properties.

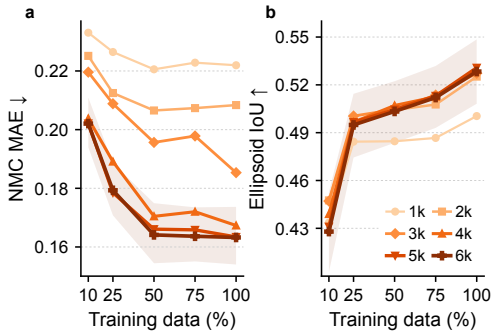


Figure 4: **Predictive data scaling** on `test_struct`. NMC MAE (**a**) and covariance-ellipsoid IoU (**b**) across 10–100% of the training data and 1k–6k training steps.

4.3 RNA–LIGAND BINDING

If learned dynamics fingerprints capture biologically relevant conformational information, they should improve tasks where RNA recognition depends on structural adaptation. We therefore assess downstream transfer to RNA–ligand binding using frozen RNADynNet fingerprints in the GerNA-Bind framework (Xia et al., 2025).

Tab. 4 shows that the binding model using frozen RNADynNet fingerprints improves AUROC in five of eight settings. Notably, performance improves on both RNA-homology splits and both combined homology-and-fingerprint splits, with the latter increasing from 0.7279 to 0.7673 on Biosensor and from 0.6176 to 0.6619 on Robin. Overall, the gains are concentrated on settings involving RNA distribution shift.

5 RELATED WORK

Biomolecular MD Data. Protein dynamics learning benefits from larger MD resources, including BioEmu for protein equilibrium ensembles, mdCATH for protein trajectories, and MISATO for protein–ligand MD data (Lewis et al., 2025; Mirarchi et al., 2024; Siebenmorgen et al., 2024). ATLAS and Dynamic PDB provide protein MD resources with flexibility descriptors and additional physical properties, respectively (Vander Meersche et al., 2024; Liu et al., 2024). RNA trajectory

Table 4: RNA–ligand binding AUROC (**best**, second best).

Method	AUROC-Biosensor				AUROC-Robin			
	Random	RNA homology	Ligand fingerprint	Homol. & fingerpr.	Random	RNA homology	Ligand fingerprint	Homol. & fingerpr.
RSAPred (Krishnan et al., 2024)	0.8764	0.7550	0.6707	0.6019	0.6327	0.5392	0.6320	0.4938
DeepDTIs (Wen et al., 2017)	0.9300	0.8399	0.6813	0.6118	0.6302	0.5503	0.6290	0.4987
DeepConv-DTI (Lee et al., 2019)	0.9249	0.8427	0.6894	0.6213	0.6301	0.5625	0.6390	0.5104
GraphDTA (Nguyen et al., 2021)	0.8992	0.8284	0.7014	0.6370	0.6590	0.5528	0.6481	0.5510
GerNA-Bind (Xia et al., 2025)	**0.9755**	0.9014	**0.7723**	0.7279	0.7094	0.6188	**0.6814**	0.6176
RNADynNet	0.9582	**0.9066**	0.7041	**0.7673**	**0.7119**	**0.6899**	0.6760	**0.6619**

resources remain smaller and more specialized (Tab. 1). BIGNASim is a database and analysis portal for nucleic-acid MD trajectories, initially including 14 RNA simulations (Hospital et al., 2016). CASP15 MD evaluates refinement of predicted RNA structures (Nithin et al., 2025), while the simulations assembled for DeepRMSF support flexibility prediction (Feng et al., 2026c). RNADynBench instead contributes a larger all-atom RNA trajectory collection with RNA-group partitions and common protocols for generating and predicting MD-derived descriptors.

Ensemble and Trajectory Generation. Protein ensemble generators such as AlphaFlow (Jing et al., 2024a) and BioEmu (Lewis et al., 2025), and RNA generators such as DynaRNA (Li et al., 2025) and RNAbpFlow (Tarafder & Bhattacharya, 2026), sample conformations without temporal order. For time-resolved dynamics, ConfRover models protein dynamics autoregressively (Shen et al., 2025), whereas MDGen jointly models full trajectories in an SE(3)-invariant representation of residue poses and torsions (Jing et al., 2024b). BioMD and BioKinema instead operate directly on all-atom Cartesian trajectories using spatiotemporal models and hierarchical forecasting and interpolation, with BioKinema additionally introducing a Langevin-inspired temporal decay bias (Feng et al., 2026b;a). Nevertheless, none of these trajectory generators was originally developed for RNA MD generation.

Dynamics Prediction from Structure. RNA elastic networks and DeepRMSF estimate flexibility from a single structure (Pinamonti et al., 2015; Feng et al., 2026c). In proteins, DynaProt predicts directional residue fluctuations and pairwise motion couplings (Bafna et al., 2026), while Ensembits distills information from MD ensembles into representations accessible from a single frame (Shi & Oliver, 2026). RNADynNet extends RNA structure-based dynamics prediction beyond scalar flexibility by learning MD-derived residue covariance and inter-residue motion coupling from a single conformer, within the same backbone used for trajectory generation.

Guided Generative and Representation Learning. Recent work has explored incorporating auxiliary signals into generative and representation learning. REPA aligns generative features with pretrained representations (Yu et al., 2025), HoG-Diff guides graph diffusion with higher-order topology (Huang & Birdal, 2026), and PSG-JEPA grounds predictive latents with physical state and multi-horizon transition targets (Yan et al., 2026). We instead use MD trajectories to jointly supervise trajectory generation and single-conformer dynamics representations through trajectory-informed alignment and MD-derived physical targets.

6 Conclusion

RNADynBench establishes the data foundation for jointly learning RNA trajectory generation and single-conformer dynamics prediction. Building on its trajectories and MD-derived statistics, RNADynNet integrates these capabilities within a shared backbone. Physical grounding improves generated dynamics and the physical information recoverable from single-conformer fingerprints, while an initial RNA–ligand binding evaluation illustrates their downstream transfer. RNADynBench provides a foundation for moving RNA modeling beyond static structure toward models that learn conformational dynamics as a bridge between molecular structure, recognition, and function.

Limitations and future work. Our simulations span 100 ns and may not capture slower conformational dynamics. The study focuses primarily on isolated RNA and does not explicitly model complete RNA–protein or RNA–small-molecule complexes. In the future, RNADynBench could support dynamics-aware RNA design, perturbation-response prediction, and integration with experimental observables. It could also enable studies of how mutations, molecular partners, and environmental conditions reshape RNA dynamics.

AI USE STATEMENT

In this work, we used generative AI tools for language editing and for limited code editing, debugging, and visualization. All AI-assisted text and code were reviewed by the authors. The authors take full responsibility for the final content and artifacts of this work.

ETHICS AND REPRODUCIBILITY STATEMENT

RNA dynamics modeling has many promising applications benefiting humanity, including understanding RNA function, molecular recognition, and RNA-targeting therapeutics. More capable models could also support the engineering of functional RNAs or RNA–molecule interactions, raising potential dual-use concerns. RNADynNet does not perform de novo RNA design or establish biological activity. Therefore, RNA dynamics models generally need to be applied with caution when used in molecular design or therapeutic pipelines.

Sec. A-Sec. C describe data curation, MD simulation and quality control, model architecture and training, baseline adaptation, evaluation metrics, and downstream experiments. Sec. D reports additional scaling, ablation, and statistical results. Our data and implementation will be released upon publication.

REFERENCES

Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

Josh Abramson, Jonas Adler, Jack Dunger, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630:493–500, 2024. doi: 10.1038/s41586-024-07487-w.

Bartosz Adamczyk, Maciej Antczak, and Marta Szachniuk. RNAsolo: a repository of cleaned PDB-derived RNA 3D structures. *Bioinformatics*, 38(14):3668–3670, 2022. doi: 10.1093/bioinformatics/btac386.

Mihir Bafna, Bowen Jing, and Bonnie Berger. Learning residue level protein dynamics with multiscale Gaussians. In *International Conference on Learning Representations*, volume 2026, pp. 20481–20500, 2026.

Amita Barik, C. Nithin, P. Manasa, and Ranjit Prasad Bahadur. A protein–RNA docking benchmark (i): Nonredundant cases. *Proteins: Structure, Function, and Bioinformatics*, 80(7):1866–1871, 2012. doi: 10.1002/prot.24083.

H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 10 1984. ISSN 0021-9606. doi: 10.1063/1.448118.

Boyuan Chen, Diego Martí Monsó, Yilun Du, Max Simchowitz, Russ Tedrake, and Vincent Sitzmann. Diffusion forcing: Next-token prediction meets full-sequence diffusion. *Advances in Neural Information Processing Systems*, 37:24081–24125, 2024.

Jiayang Chen, Zhihang Hu, Siqi Sun, Qingxiong Tan, Yixuan Wang, Qinze Yu, Licheng Zong, Liang Hong, Jin Xiao, Tao Shen, et al. Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *BioRxiv*, pp. 2022–08, 2022.

Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $N \cdot \log(N)$ method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

Elizabeth A. Dethoff, Katja Petzold, Jeetender Chugh, Anette Casiano-Negroni, and Hashim M. Al-Hashimi. Visualizing transient low-populated structures of RNA. *Nature*, 491(7426):724–728, 2012. doi: 10.1038/nature11498.

K. Anton Feenstra, Berk Hess, and Herman J. C. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999. doi: 10.1002/(SICI)1096-987X(199906)20:8⟨786::AID-JCC5⟩3.0.CO;2-B.

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

Bin Feng, Jiying Zhang, Xinni Zhang, Zijing Liu, and Yu Li. BioMD: All-atom generative model for biomolecular dynamics simulation. In *International Conference on Learning Representations*, pp. 73466–73491, 2026b.

Chenjie Feng, Xiaowen Sun, Xintao Song, Lei Bao, Weikang Gong, and Renmin Han. DeepRMSF: a deep learning-based automated approach for predicting atomic-level flexibility in RNA structure. *Briefings in Bioinformatics*, 27(1):bbaf720, 2026c. doi: 10.1093/bib/bbaf720.

Laura R. Ganser, Megan L. Kelly, Daniel Herschlag, and Hashim M. Al-Hashimi. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology*, 20(8):474–489, 2019. doi: 10.1038/s41580-019-0136-0.

Pengkang Guo, Bruno Correia, Pierre Vandergheynst, and Daniel Probst. Boosting protein graph representations through static-dynamic fusion. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 20777–20792, 2025.

William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31: 1695–1697, Mar 1985. doi: 10.1103/PhysRevA.31.1695.

Adam Hospital, Pau Andrio, Cesare Cugnasco, Laia Codo, Yolanda Becerra, Pablo D. Dans, Federica Battistini, Jordi Torres, Ramón Goñi, Modesto Orozco, and Josep Ll. Gelpí. BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Research*, 44(D1):D272–D278, 2016. doi: 10.1093/nar/gkv1301.

Yiming Huang and Tolga Birdal. Hog-diff: Higher-order guided diffusion for graph generation. In *International Conference on Learning Representations*, volume 2026, pp. 111572–111614, 2026.

Yiming Huang, Yujie Zeng, Vijay Prakash Dwivedi, Simone Foti, Jianmin Wang, Jure Leskovec, and Tolga Birdal. Higher-order grammar representations for molecular generation and learning. In *Learning Meaningful Representations of Life (LMRL) Workshop at ICLR*, 2026.

Saeed Izadi, Ramu Anandakrishnan, and Alexey V Onufriev. Building water models: a different approach. *The journal of physical chemistry letters*, 5(21):3863–3871, 2014. doi: 10.1021/jz501780a.

Bowen Jing, Bonnie Berger, and Tommi Jaakkola. AlphaFold meets flow matching for generating protein ensembles. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 22277–22303. PMLR, 2024a.

Bowen Jing, Hannes Stärk, Tommi Jaakkola, and Bonnie Berger. Generative modeling of molecular dynamics trajectories. In *Advances in Neural Information Processing Systems*, volume 37, pp. 40534–40564, 2024b. doi: 10.52202/079017-1282.

Chaitanya K. Joshi, Arian R. Jamasb, Ramon Viñas, Charles Harris, Simon V. Mathis, Alex Morehead, Rishabh Anand, and Pietro Liò. gRNAde: Geometric deep learning for 3D RNA inverse design. In *International Conference on Learning Representations*, volume 2025, pp. 10166–10191, 2025.

John Jumper, Richard Evans, Alexander Pritzel, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Yuki Kagaya, Zicong Zhang, Nabil Ibtehaz, Xiao Wang, Tsukasa Nakamura, Pranav Deep Punuru, and Daisuke Kihara. NuFold: end-to-end approach for RNA tertiary structure prediction with flexible nucleobase center representation. *Nature communications*, 16(1):881, 2025. doi: 10.1038/s41467-025-56261-7.

Dan Kalifa, Eric Horvitz, and Kira Radinsky. Learning protein representations with conformational dynamics. *Bioinformatics*, 42(5):btag254, 2026. doi: 10.1093/bioinformatics/btag254.

Sowmya Ramaswamy Krishnan, Arijit Roy, and M. Michael Gromiha. Reliable method for predicting the binding affinity of RNA–small molecule interactions using machine learning. *Briefings in Bioinformatics*, 25(2):bbae002, 2024.

Achintya Kundu, Jakob Schauss, Benjamin P. Fingerhut, and Thomas Elsaesser. Change of hydration patterns upon rna melting probed by excitations of phosphate backbone vibrations. *The Journal of Physical Chemistry B*, 124(11):2132–2138, 2020. doi: 10.1021/acs.jpcb.0c01474.

Ingoo Lee, Jongsoo Keum, and Hojung Nam. DeepConv-DTI: prediction of drug–target interactions via deep learning with convolution on protein sequences. *PLoS Computational Biology*, 15(6): e1007129, 2019.

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

Zhengxin Li, Junjie Zhu, Xiaokun Hong, Junxi Mu, Zhuoqi Zheng, Taeyoung Cui, Yutong Sun, Ting Wei, and Hai-Feng Chen. DynaRNA: accurate dynamic RNA conformation ensemble generation with diffusion model. *Communications Biology*, 8:1472, 2025. doi: 10.1038/s42003-025-08875-2.

Ce Liu, Jun Wang, Zhiqiang Cai, Yingxu Wang, Huizhen Kuang, Kaihui Cheng, Liwei Zhang, Qingkun Su, Yining Tang, Fenglei Cao, et al. Dynamic pdb: a new dataset and a se (3) model extension by integrating dynamic behaviors and physical properties in protein structures. *arXiv preprint arXiv:2408.12413*, 2024.

Marcin Magnus, William Gao, Nivedita Dutta, Quentin Vicens, and Elena Rivas. RNAhub—an automated pipeline to search and align RNA homologs with secondary structure assessment. *Nucleic Acids Research*, 53(W1):W496–W502, 2025. doi: 10.1093/nar/gkaf342.

Antonio Mirarchi, Toni Giorgino, and Gianni De Fabritiis. mdCATH: A large-scale MD dataset for data-driven computational biophysics. *Scientific Data*, 11:1299, 2024. doi: 10.1038/s41597-024-04140-z.

Thin Nguyen et al. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021.

Chandran Nithin, Smita P. Pilla, and Sebastian Kmiecik. When does molecular dynamics improve RNA models? insights from CASP15 and practical guidelines. *Computational and Structural Biotechnology Journal*, 27:4201–4211, 2025. doi: 10.1016/j.csbj.2025.10.003.

Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

Sergiy Perepelytsya, Jozef Uličný, and Sergey N. Volkov. Molecular dynamics study of the competitive binding of hydrogen peroxide and water molecules with dna phosphate groups. *European Biophysics Journal*, 50(5):759–770, 2021. doi: 10.1007/s00249-021-01522-8.

Giovanni Pinamonti, Sandro Bottaro, Cristian Micheletti, and Giovanni Bussi. Elastic network models for RNA: a comparative assessment with molecular dynamics and SHAPE experiments. *Nucleic Acids Research*, 43(15):7260–7269, 2015. doi: 10.1093/nar/gkv708.

Arkajyoti Sengupta, Zhen Li, Lin Frank Song, Pengfei Li, and Kenneth M. Merz, Jr. Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *Journal of Chemical Information and Modeling*, 61(2):869–880, 2021. doi: 10.1021/acs.jcim.0c01390.

Yuning Shen, Lihao Wang, Huizhuo Yuan, Yan Wang, Bangji Yang, and Quanquan Gu. Simultaneous modeling of protein conformation and dynamics via autoregression. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Kaiwen Shi and Carlos Oliver. Ensembits: an alphabet of protein conformational ensembles. *arXiv preprint arXiv:2605.13789*, 2026.

Till Siebenmorgen, Filipe Menezes, Sabrina Benassou, Erinc Merdivan, Kieran Didi, André Santos Dias Mourão, Radosław Kitel, Pietro Liò, Stefan Kesselheim, Marie Piraud, et al. MISATO: machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature computational science*, 4(5):367–378, 2024. doi: 10.1038/s43588-024-00627-2.

Oscar Skean, Md Rifat Arefin, Dan Zhao, Niket Nikul Patel, Jalal Naghiyev, Yann Lecun, and Ravid Shwartz-Ziv. Layer by layer: Uncovering hidden representations in language models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55854–55875. PMLR, 13–19 Jul 2025.

Jiří Šponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A. Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, Nils G. Walter, and Michal Otyepka. RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chemical Reviews*, 118(8):4177–4338, 2018. doi: 10.1021/acs.chemrev.7b00427.

Marcell Szikszai, Marcin Magnus, Siddhant Sanghi, Sachin Kadyan, Nazim Bouatta, and Elena Rivas. RNA3DB: A structurally-dissimilar dataset split for training and benchmarking deep learning models for RNA structure prediction. *Journal of Molecular Biology*, 436(17):168552, 2024.

Sumit Tarafder and Debswapna Bhattacharya. RNAbpFlow: base pair-augmented SE(3) flow matching for conditional RNA 3D structure generation. *Nature Methods*, 23(7):1349–1358, 2026. doi: 10.1038/s41592-026-03128-4.

ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, et al. Protenix - advancing structure prediction through a comprehensive AlphaFold3 reproduction. *bioRxiv*, 2025. doi: 10.1101/2025.01.08.631967.

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

Vsevolod Viliuga, Leif Seute, Matteo Tadiello, Nicolas Wolf, Frauke Gräter, and Arne Elofsson. Predicting directional flexibility in proteins. *arXiv preprint arXiv:2609.08474*, 2026.

Ming Wen et al. Deep-learning-based drug–target interaction prediction. *Journal of Proteome Research*, 16(4):1401–1409, 2017.

Yunpeng Xia, Jiayi Li, Yi-Ting Chu, Jiahua Rao, Jing Chen, Chenqing Hua, Dong-Jun Yu, Xiu-Cai Chen, and Shuangjia Zheng. Deciphering RNA–ligand binding specificity with GerNA-Bind. *Nature Machine Intelligence*, 7(12):1996–2008, 2025. doi: 10.1038/s42256-025-01154-z.

Rajeev Yadav, Julia R. Widom, Adrien Chauvier, and Nils G. Walter. An anionic ligand snap-locks a long-range interaction in a magnesium-folded riboswitch. *Nature Communications*, 13:207, 2022. doi: 10.1038/s41467-021-27827-y.

Haodong Yan, Jiaguan Zhu, Mingyuan Jia, Ruiqing Yin, Junjie He, Zhide Zhong, Junfeng Li, Jinxuan Lu, Hengtao Li, Tianran Zhang, et al. Is forward prediction enough? physical state grounding for jepa world models. *arXiv preprint arXiv:2608.06799*, 2026.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. In *International Conference on Learning Representations*, 2025.

Marie Zgarbová, Michal Otyepka, Jiri Sponer, Arnost Mladek, Pavel Banas, Thomas E Cheatham III, and Petr Jurecka. Refinement of the cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of chemical theory and computation*, 7(9):2886–2902, 2011.

Appendix

A RNADynBench construction and characterization **15**
A.1 Structure preparation and MD simulation 15
A.2 Quality control and dataset splits 16
A.3 Size-matched RNA–protein comparisons 16

B RNADynNet architecture and objectives **17**
B.1 Backbone and temporal conditioning 17
B.2 Training views and fingerprint extraction 18
B.3 Trajectory generation objective 19
B.4 Physical grounding targets and prediction heads 19

C Experimental protocols **20**
C.1 Training configuration 20
C.2 Trajectory sampling 20
C.3 Baselines and ablation settings 20
C.4 Evaluation metrics and MD references 21
C.5 RNA–ligand binding protocol 22

D Supplementary results **22**
D.1 Grounding targets and supervision depth 22
D.2 Predictive data scaling 23
D.3 Qualitative trajectory generation 23

A RNADynBench construction and characterization

A.1 Structure preparation and MD simulation

Tab. 5 summarizes the release-level molecular dynamics protocol. The default solvent box was a rhombic dodecahedron with 12 Å padding. Systems showing periodic-image contacts during equilibration were rebuilt with 18 or 24 Å padding. The released files contain the RNA-only post-processed coordinates.

Table 5: RNADynBench molecular dynamics protocol.

Setting	Value
Simulation engine	GROMACS (Abraham et al., 2015)
RNA force field	Amber RNA OL3 (Zgarbová et al., 2011)
Solvent and ions	OPC water (Izadi et al., 2014); Li–Merz 12-6 OPC Na^+/Cl^- parameters (Sengupta et al., 2021); neutralizing ions plus 150 mM NaCl
Periodic box	Rhombic dodecahedron; 12 Å default padding, increased to 18 or 24 Å for periodic-image contacts
Electrostatics	Particle-mesh Ewald (Darden et al., 1993)
Integration	4 fs time step with hydrogen mass repartitioning (Feenstra et al., 1999)
Energy minimization	Steepest-descent minimization to the configured force/energy convergence criterion
NVT equilibration	200 ps at 300 K; Berendsen thermostat (Berendsen et al., 1984)
NPT equilibration	1 ns at 300 K and 1 bar; Nosé–Hoover thermostat (Hoover, 1985) and Berendsen barostat
Final NPT equilibration	5 ns; Nosé–Hoover thermostat and Parrinello–Rahman barostat (Parrinello & Rahman, 1981)
Production	100 ns; Nosé–Hoover thermostat and Parrinello–Rahman barostat
Coordinate output	100 ps stride; 1,001 frames from 0 to 100 ns

Computational resources. The accepted trajectories were generated on NVIDIA H200, GH200, and A100 GPUs, requiring approximately **23,000 H200-equivalent GPU hours** including equilibration and production.

A.2 Quality control and dataset splits

We checked residue and atom identities across source structures, prepared RNA systems, GROMACS topologies, and the first released trajectory frames. Source residue numbers and insertion codes were preserved alongside a sample-level residue index; component-selection and repair decisions were recorded in the metadata. Structures with missing or ambiguous chain mappings, unresolved internal backbone gaps, unsupported modified residues, or failed topology reconstruction were excluded.

Run-level checks covered energy minimization, equilibration stability, production completion, coordinate spikes, covalent-backbone continuity, and periodic-boundary behavior. For multi-chain systems, the full 100 ns window was additionally assessed for sustained loss of inter-chain contacts and chain separation. Release checks verified the monotonic 0–100 ns time axis, 1,001 readable frames, the presence and checksums of all public files, complete metadata, and the absence of exact duplicate trajectory aliases. Additional geometric and chain-separation findings were recorded as quality-control flags for retained trajectories.

Leakage-controlled splits. Tab. 6 reports the fixed RNADynBench partitions and their roles. The homology-guard trajectories are included in the public release but excluded from training, model selection, and formal evaluation. The structural test set starts from 113 RNA entries mapped from the gRNAde Das test set (Joshi et al., 2025) and expands by RNA identity and source membership to 117 trajectories. Corrected source identity is used for grouping while historical RNA-Solo identifiers are retained for provenance. Direct sequence screening requires at least 80% identity, defined as the number of matching residues divided by the shorter full RNA sequence length, and at least 80% coverage on each sequence, measured using actually paired residues. Multi-chain sequences retain the concatenated representation. Guards include direct sequence or structural matches and samples sharing a sequence or structural cluster with the structural test set. Direct structural screening was limited to test–candidate pairs from sequence- or structural-cluster neighborhoods and previously observed direct matches, using a TM-score threshold of 0.45 normalized by the longer RNA.

For `test_flex`, each trajectory is scored by its median pairwise aligned C1′ RMSD over 101 frames sampled at 1-ns intervals. After excluding `test_struct` and guards, we group trajectories sharing an exact sequence, corrected RNA identity, source structure, sequence cluster, or structural cluster. Each group receives the median member score. Eligible groups (at most 25 trajectories) are ranked by score within strata of rounded mean length (≤ 40, 41–100, or > 100 nucleotides), chain multiplicity, predominant source class, and quality-control warnings. Seeded round-robin selection adds whole groups only when they bring `test_flex` closer to 100 trajectories. A seeded selection from the remaining groups yields 102 validation trajectories; all other groups form the training set.

Table 6: RNADynBench dataset splits.

Partition	Trajectories	Primary role
Training	2,248	Model optimization and downstream probe training
Validation	102	Checkpoint and hyperparameter selection
`test_struct`	117	Structural and molecular generalization
`test_flex`	100	High-flexibility generalization
Homology guard	18	Excluded homology/cluster guards
Total	2,585	

A.3 Size-matched RNA–protein comparisons

For each protein reference, we retained the overlapping heavy-atom-count range and partitioned log heavy-atom count into five bins. RNA and protein systems then received common bin weights. ATLAS and dynamicPDB were matched separately because their size distributions differed.

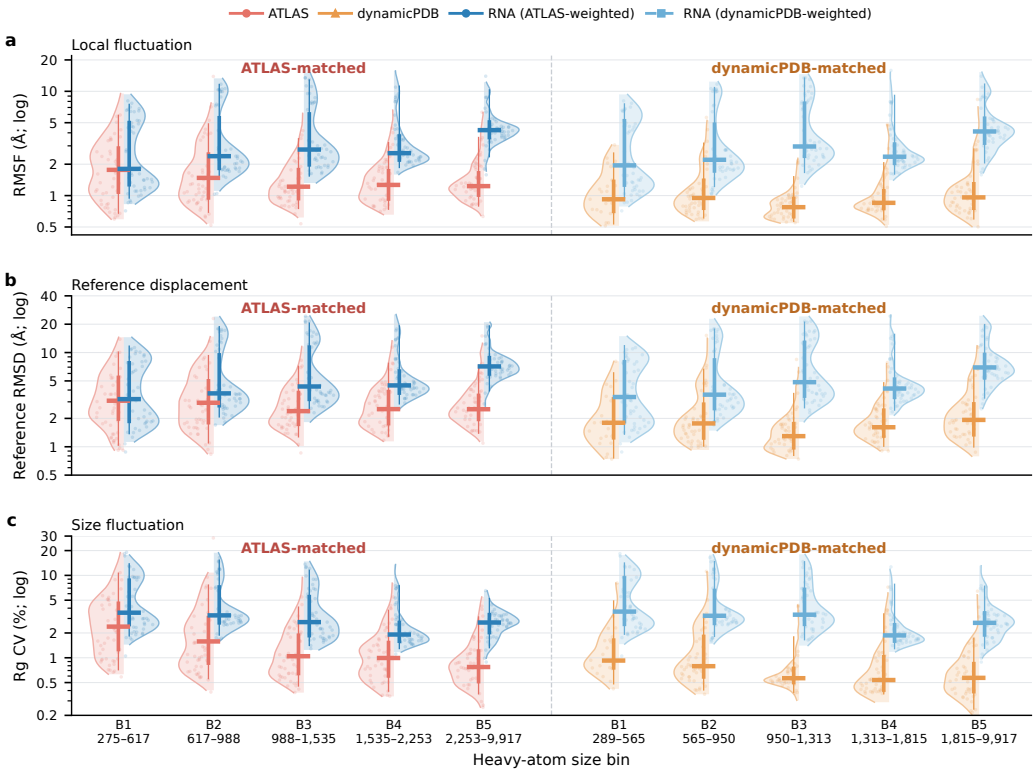


Figure 5: **Size-stratified RNA–protein dynamics.** Distributions of (**a**) mean marker-atom RMSF, (**b**) mean RMSD to the first sampled frame, and (**c**) heavy-atom R_g coefficient of variation across five size bins. RNA uses C4′ markers and protein uses Cα; vertical axes are logarithmic. Split violins show proteins on the left and matched RNAs on the right; bars indicate the median, interquartile range, and 5th–95th percentiles.

B RNADynNet architecture and objectives

Bold symbols denote vectors, matrices, or tensors. Subscript f indexes frames, and i, j index residues. Tab. 7 summarizes the principal notation for RNADynNet. Quantities used only in the downstream binding model are defined in its own section.

B.1 Backbone and temporal conditioning

RNADynNet initializes its diffusion backbone from Protenix (Team et al., 2025). Following Protenix's AF3-style tokenization, standard RNA nucleotides form residue tokens, and the frozen Input Embedder and Pairformer in Fig. 2 provide single and pair features. An atom encoder maps EDM-preconditioned noisy coordinates to residue tokens before spatiotemporal attention. Each spatiotemporal block applies spatial attention followed by temporal self-attention. The diffusion backbone is optimized jointly with the representation heads. The latent objectives use an intermediate token representation, whereas coordinate prediction continues through the denoising path.

Time conditioning and the single-frame predictor.

The alignment target of Eq. (10) depends on the horizon H and interval Δt of the window, so the single-frame predictor is conditioned on both. The predictor embeds standardized logarithms of H and Δt:

$$\mathbf{e}_{\text{time}} = \text{MLP}_{\text{time}}\left(\left[\frac{\log H - \mu_H}{s_H}, \frac{\log \Delta t - \mu_{\Delta t}}{s_{\Delta t}}\right]\right), \tag{8}$$

Table 7: Notation used throughout our manuscript.

Symbol	Meaning
L, A, N, d	Numbers of frames, atoms, residues, and feature dimensions.
$\mathbf{X}, \mathbf{x}_f$	All-atom trajectory window and its frame-f conformer.
$\theta, \bar{\theta}$	Online parameters and their EMA; fingerprints use the latter.
$\mathcal{D}_\theta, \mathcal{E}_\theta, \mathcal{E}_{\bar{\theta}}, \ell_{\mathrm{fp}}$	Trajectory denoiser, online single-conformer encoder, EMA fingerprint encoder, and readout block.
$\widetilde{\mathbf{X}}, \boldsymbol{\sigma}, \boldsymbol{\epsilon}$	Noised trajectory, coordinate-noise scale tensor, and Gaussian perturbation.
$\mathbf{R}_{\mathrm{traj}}, \mathbf{R}_{\mathrm{sf}}$	Independent global rotations of the trajectory and single-frame views.
$J, \mathcal{V}$	Sampled anchor-frame index and set of valid residues.
$\mathbf{H}^{\mathrm{FP}}_{\bar{\theta}}, \mathbf{h}^{\mathrm{FP}}_{\bar{\theta},i}$	EMA single-conformer fingerprint matrix and its residue-i vector.
$\mathbf{h}^{\mathrm{FP}}_{\theta,i}, \mathbf{h}^{\mathrm{traj}}_{\theta,i}$	Online single-conformer and trajectory-anchor states for residue i at sampled anchor J.
$q_\psi, \mathbf{z}^{\mathrm{traj}}_i$	Alignment predictor conditioned on H and Δt, and detached trajectory-anchor target.
$\boldsymbol{\Sigma}_i, \mathbf{R}^{(J)}_i$	Residue covariance and orthonormal residue frame of clean anchor J.
$\mathbf{C}, \boldsymbol{\rho}$	Inter-residue motion-coupling matrix and its normalized form (NMC).
f_Σ, f_C	Covariance and motion-coupling heads for physical grounding.
$\mathcal{L}_{\mathrm{gen}}, \mathcal{L}_{\mathrm{align}}, \mathcal{L}_{\mathrm{PG}}$	Generation, single-frame-to-trajectory alignment, and physical-grounding objectives.

and the shared residue-wise predictor applies FiLM modulation,

$$\begin{aligned}[\boldsymbol{\gamma}_{\mathrm{time}}, \boldsymbol{\beta}_{\mathrm{time}}] &= W_{\mathrm{FiLM}} \mathbf{e}_{\mathrm{time}}, \\ q_\psi\left(\mathbf{h}, H, \Delta t\right) &= \mathbf{W}_2\, \mathrm{GELU}\left[(1 + \boldsymbol{\gamma}_{\mathrm{time}}) \odot (\mathbf{W}_1 \mathbf{h} + \mathbf{b}_1) + \boldsymbol{\beta}_{\mathrm{time}}\right] + \mathbf{b}_2.\end{aligned} \tag{9}$$

The FiLM projection is zero-initialized, so the predictor initially applies no time-dependent modulation. The statistics $(\mu_H, s_H, \mu_{\Delta t}, s_{\Delta t})$ are computed once from the training-window distribution and held fixed during training.

B.2 Training views and fingerprint extraction

Each sampled trajectory is rigidly aligned to its first frame and centered before view construction. Let $\mathcal{V}$ denote the RNA residues with at least one resolved atom.

View construction. The trajectory view uses one global rotation $\mathbf{R}_{\mathrm{traj}}$ across all frames, applied jointly to coordinates and Gaussian noise. Its reconstruction target is $\mathbf{R}_{\mathrm{traj}}\mathbf{X}$. The single-frame pass receives the corresponding clean anchor $\mathbf{R}_{\mathrm{sf}}\mathbf{x}_J$, without coordinate noise. The two rotations are sampled independently and uniformly from $\mathrm{SO}(3)$.

Readouts and alignment target. Let $\mathcal{H}^{\ell_{\mathrm{fp}}}_\theta$ denote the online backbone states at the fingerprint readout block. For the sampled anchor J, the trajectory readout and alignment target are

$$\begin{aligned}\mathbf{h}^{\mathrm{traj}}_{\theta,i} &= \mathrm{LN}\left[\mathcal{H}^{\ell_{\mathrm{fp}}}_\theta\left(\widetilde{\mathbf{X}}, \boldsymbol{\sigma}, H, \Delta t\right)_{J,i}\right], \\ \mathbf{z}^{\mathrm{traj}}_i &= \mathrm{sg}\left[\mathrm{norm}\left(\mathbf{h}^{\mathrm{traj}}_{\theta,i}\right)\right].\end{aligned} \tag{10}$$

Here LN is affine-free LayerNorm, $\mathrm{norm}(\mathbf{z}) = \mathbf{z}/(\|\mathbf{z}\|_2 + \varepsilon)$, and sg stops gradients. The normalized, detached target is read at one anchor after temporal attention over the full window. The single-frame readout uses affine-free LayerNorm at the same depth, before the predictor. Eq. (4) averages the alignment loss over valid residues; its gradients update the single-frame path.

Reported single-conformer fingerprints use the EMA encoder in evaluation mode on one centered conformer, without random rotation or coordinate noise. Any probe-specific normalization is applied after this readout.

B.3 TRAJECTORY GENERATION OBJECTIVE

The generation objective combines coordinate denoising, local structural agreement, fluctuation matching, and short-lag autocorrelation losses:

$$\begin{aligned}\mathcal{L}_{\text{gen}} = \mathcal{L}_{\text{denoise}} + \lambda_{\text{sLD}}\mathcal{L}_{\text{smooth-lDDT}} + \lambda_{\text{RMSF}}\mathcal{L}_{\text{RMSF}} \\ + \lambda_{\text{rel}}\mathcal{L}_{\text{rel-RMSF}} + \lambda_{\text{local}}\mathcal{L}_{\text{local-RMSF}} + \lambda_{\text{ACF}}\mathcal{L}_{\text{ACF}}.\end{aligned} \tag{11}$$

For the reported model, the effective multipliers are $\lambda_{\text{sLD}} = 4$, $\lambda_{\text{RMSF}} = 0.4$, $\lambda_{\text{rel}} = \lambda_{\text{local}} = \lambda_{\text{ACF}} = 3.2$. The denoising loss in Eq. (3) already includes its generative-group multiplier.

The coordinate objective follows BioKinema's global-frame reconstruction convention, without fitting a new rigid alignment between the predicted and target coordinates. For each sampled frame, squared Euclidean errors are weighted by atom validity, molecular type, and the residue-specific EDM factor. Weighted errors are summed over atoms, divided by the number of resolved atoms and by three spatial dimensions, and averaged over all sampled frames. Thus the coordinate MSE includes near-clean conditioning frames; AF3-style smooth-lDDT (Abramson et al., 2024) instead uses the explicit target-frame mask to supervise local structural agreement. The RMSF-based terms follow BioKinema's fluctuation supervision (Feng et al., 2026a). The RMSF term compares per-atom positional fluctuations over the window. Relative and local RMSF compare fluctuations of selected inter-token and within-token atom-pair distances, respectively. The ACF term compares normalized, time-centered positional autocorrelations over short frame lags for selected atoms. The RMSF, relative-RMSF, local-RMSF, and ACF terms apply only to windows with at least three frames.

B.4 PHYSICAL GROUNDING TARGETS AND PREDICTION HEADS

Reference statistics and local coordinates. PG uses the same 101-frame, 100-ns reference trajectory for covariance and motion-coupling labels, with population normalization over all 101 frames. We align all frames to frame zero by a C1′-based rigid fit and remove the temporal mean separately for each residue. The labels satisfy $C_{ii} = \operatorname{tr}(\mathbf{\Sigma}_i)$, while $\mathbf{C}$ is invariant to a common global rotation. For each clean anchor, the first local axis points from C1′ to C3′; the second is the C1′-to-C4′ vector orthogonalized against the first, and the third is their cross product. The three normalized axes form the columns of $\mathbf{R}_i^{(J)}$. Full-trajectory statistics are expressed in this clean-anchor frame, even when the supervised hidden state comes from the noisy trajectory pass. Labels are matched by residue identity and only valid residues enter the reductions.

The main text uses numerical covariance values in Å^2. With $s_0 = 1\,\text{Å}^2$, the matrix logarithm acts on the dimensionless local covariance $\mathbf{\Sigma}_{i,\text{loc}}^{(J)}/s_0$ after eigenvalue flooring. The physical prediction is recovered as $\widehat{\mathbf{\Sigma}}_i = s_0\mathbf{R}_i \exp\big(f_\Sigma(\mathbf{h}_{\theta,i}^{\text{FP}})\big)\mathbf{R}_i^\top$.

Shared covariance head. The single-conformer and trajectory-anchor representations are read at the same intermediate block. The covariance head is a normalized multilayer perceptron with six outputs. Its six outputs fill the independent entries of a symmetric 3×3 matrix $\mathbf{S}$, directly parameterizing the log-covariance. Training compares $\mathbf{S}$ with detached matrix-log targets, without differentiating through an eigendecomposition of the prediction. The matrix-entry loss in Eq. (6) averages squared errors over all nine entries, valid residues, and the two views. Each symmetric off-diagonal entry contributes twice. The matrix exponential recovers covariance descriptors at inference and is absent from the training loss.

Motion-coupling head and reductions. The motion-coupling head reads intermediate residue states. A normalized residue-level multilayer perceptron produces vectors $\mathbf{u}_i$; symmetric pair features $[\mathbf{u}_i + \mathbf{u}_j; |\mathbf{u}_i - \mathbf{u}_j|]$ are mapped to a scalar matrix $\mathbf{A}_C$. Its lower-triangular factor is

$$\mathbf{L}_C = \operatorname{tril}(\mathbf{A}_C, -1) + \operatorname{diag}(\operatorname{softplus}(\operatorname{diag}\mathbf{A}_C) + \varepsilon_C)\,, \tag{12}$$

where $\varepsilon_C > 0$ stabilizes the diagonal and $\widehat{\mathbf{C}} = \mathbf{L}_C\mathbf{L}_C^\top$ is positive definite. The lower-triangular factor is constructed in the input residue order. The readout applies an MLP to RNADynNet's residue states in place of BackFlip's IPA edge encoder.

In Eq. (7), the matrix-pattern and magnitude terms include diagonal entries, while NMC MSE uses off-diagonal entries only. Small positive floors stabilize the Frobenius-norm denominator in Eq. (7)

and NMC normalization. NMC is derived from the same motion-coupling matrix, with no separate prediction head. The per-sample losses are averaged over the batch. The covariance and coupling heads are supervised separately, without an explicit cross-head consistency constraint. PG supervises covariance and zero-lag motion coupling; the generation objective separately supervises short-lag autocorrelation. The combined objective is $\mathcal{L}_{\mathrm{PG}} = \lambda_{\Sigma}\mathcal{L}_{\Sigma} + \lambda_{C}\mathcal{L}_{C}$.

C EXPERIMENTAL PROTOCOLS

The code and configurations specify remaining masks and reductions, head dimensions, numerical constants, objective weights, and training, sampling, and downstream optimization settings.

C.1 TRAINING CONFIGURATION

Training uses the official RNADynBench training partition, with variable window starts, spacings, and lengths. Stored-frame spacing is converted to the physical interval Δt for temporal conditioning. The sampler mixes unconditional, initial-frame-conditioned, and endpoint-conditioned windows while varying their temporal extent and resolution.

Training starts from the Protenix (Team et al., 2025) initialization with a fresh optimizer state. We use validation-guided optimization, learning-rate scheduling, gradient clipping, mixed-precision training, and an EMA copy for trajectory generation. PG is active from the first update. Single-frame-to-trajectory alignment is activated after an initial warm-up and smoothly increased to its final weight, using one sampled anchor per eligible forward window. The alignment target is obtained from the online trajectory pass, while EMA weights are retained for generation.

Evaluation sets. RNADynBench defines `test_struct` as 117 trajectories from 62 RNAs and `test_flex` as 100 trajectories from 65 RNAs. Generation results for RNADynNet and frozen-probe evaluations for RNADynNet, w/o PG, and w/o alignment and PG use these panels. BioKinema and RNADynNet generation use fixed EMA checkpoints.

C.2 TRAJECTORY SAMPLING

Both test panels use the same EMA denoiser and hierarchical forecasting-and-interpolation protocol. The coarse stage generates temporally spaced anchors from the supplied initial conformer. The fine stage uses the denoiser conditioned on adjacent anchors to generate intermediate frames on the 101-frame, 100-ns evaluation grid. Each of the four samples is a complete trajectory.

C.3 BASELINES AND ABLATION SETTINGS

Trajectory-generation baselines. The baselines use distinct RNA input representations. BioKinema uses precomputed Protenix single and pair features for all-atom coordinate denoising and is retrained on RNADynBench. Our MDGen adaptation is trained from scratch and retains the original flow-matching architecture and sampler. It represents each nucleotide by a rigid-frame offset and ten sine–cosine torsion pairs (27 dimensions), with RNA-specific residue identities and coordinate decoding. Validation-selected EMA weights are used for evaluation. At evaluation, decoded frame zero is replaced by the input conformer; generated future frames are left unchanged. ConfRover uses RNA frames and ten torsion pairs, $\mathrm{C1}'$ pair geometry, and frozen Protenix features from sequence and topology. Its causal temporal and SE(3) diffusion modules and RNA torsion head are trained from scratch; an RNA decoder reconstructs atomic coordinates. ConfRover generates one trajectory per target ($K = 1$); its runtime is per trajectory, and it is excluded from the $K = 4$ ranking.

Descriptor baselines. All descriptor models receive the experimental initial structure. Geometry-GNN builds a distance-based graph from a single RNA conformer and jointly trains its graph network and descriptor heads on MD-derived targets. RNA-FM+2D GCN freezes RNA-FM but trains its GCN and prediction heads. Other unmarked model rows use native descriptor heads. DynaProt combines validation-selected marginal and joint native heads. Rows marked with † use independently trained descriptor probes on frozen representations.

Representation-learning ablations. The *w/o align. & PG* variant uses only $\mathcal{L}_{\text{gen}}$; *w/o PG* adds $\mathcal{L}_{\text{align}}$, and the full model also includes PG. All variants retain the coordinate, structural, fluctuation, and short-lag correlation losses in $\mathcal{L}_{\text{gen}}$ (Sec. B.3). For their † rows, we freeze the EMA encoders and train separate probes on MD-derived descriptors.

C.4 Evaluation metrics and MD references

Trajectory generation. Geometry metrics use the 100 future frames; frame zero is the supplied initial conformer. Bond MAE averages absolute RNA heavy-atom bond-length deviations from that conformer. O3′–P max is the largest absolute initial-linkage deviation over future frames and consecutive within-chain residues. The maximum is computed per trajectory before averaging generated samples. Chiral flip is the percentage of sugar-stereocenter observations with a signed-volume sign different from frame zero, including zero volume. Clashes count nonbonded heavy-atom pairs closer than 1.5 Å, excluding directly bonded pairs and pairs sharing a bonded neighbor. Counts are normalized per 1,000 heavy atoms. For dynamics metrics, C1′ coordinates in every frame are rigidly aligned to the MD initial conformer without reflection. RMSF r is the Pearson correlation across residues between MD and the generated-sample-mean RMSF profiles, each computed over all 101 frames. PCA-W_2 compares equal-weight distributions of the 100 future frames in a two-dimensional PCA space fitted only to MD. It uses optimal assignment and normalization by the square root of the C1′ atom count. Framewise geometry metrics are averaged over frames and generated samples, then within RNA systems and across RNAs. Per-target RMSF correlations and sample-mean PCA-W_2 values are likewise averaged within RNAs and then across RNAs.

Single-conformer dynamics prediction. Model predictions are compared with descriptors from the full 100-ns MD trajectory. The MD reference instead compares two non-overlapping 50-ns blocks from that same trajectory to indicate finite-sampling variability. As in prior covariance-only evaluations (Bafna et al., 2026; Viliuga et al., 2026), we compare predicted and MD residue fluctuations as zero-mean Gaussians. Both covariance matrices are symmetrized and projected to positive definiteness before computing RMWD, symmetric KL, and IoU; the equations below use these projected matrices. For N evaluated residues, root-mean 2-Wasserstein distance is

$$\text{RMWD} = \left[\frac{1}{N} \sum_{i=1}^{N} \text{tr} \left(\widehat{\mathbf{\Sigma}}_i + \mathbf{\Sigma}_i - 2 \left(\mathbf{\Sigma}_i^{1/2} \widehat{\mathbf{\Sigma}}_i \mathbf{\Sigma}_i^{1/2} \right)^{1/2} \right) \right]^{1/2}. \tag{13}$$

The symmetric Kullback–Leibler divergence is

$$\text{KL}_{\text{sym}} = \frac{1}{4N} \sum_{i=1}^{N} \left[\text{tr}(\mathbf{\Sigma}_i^{-1} \widehat{\mathbf{\Sigma}}_i) + \text{tr}(\widehat{\mathbf{\Sigma}}_i^{-1} \mathbf{\Sigma}_i) - 2d \right], \qquad d = 3. \tag{14}$$

Because both Gaussians are centered, neither metric includes mean-coordinate error. Ellipsoid IoU also evaluates the shape and orientation of the covariance:

$$\mathcal{E}(\mathbf{\Sigma}_i) = \{ \mathbf{x} \in \mathbb{R}^3 : \mathbf{x}^\top \mathbf{\Sigma}_i^{-1} \mathbf{x} \leq 1 \}. \tag{15}$$

Its eigenvectors define principal directions, and square-root eigenvalues define semi-axis lengths. IoU averages each residue's ellipsoid-intersection volume divided by its union volume, estimated by spherical quadrature. Thus, IoU reflects fluctuation scale as well as anisotropy and orientation.

RMSF and NMC are derived as in Eq. (5). For RMSF, Pearson r compares residue-wise flexibility patterns, while MAE measures absolute fluctuation-amplitude error. For NMC, r compares coupling patterns and MAE measures absolute coupling error over strictly upper-triangular residue pairs ($i < j$). RMSF uses the raw covariance trace rather than the projected matrices used for RMWD, KL, and IoU. Metrics are computed per trajectory, averaged within each RNA group, and reported as split-level RNA-group macro means. RMWD and RMSF MAE are in Å; the other descriptor scores are dimensionless.

Generation reference and aggregation. BioKinema, MDGen, and RNADynNet generate four trajectories per target. Reference-MD RNA-level intervals use cluster resampling. Geometry reference values are empirical RNA-level means from MD. Trajectory-matching reference values are medians of RNA-level statistics from conditional moving-block bootstrap samples of individual MD trajectories. These intervals describe resampling variability conditional on the observed trajectories.

C.5 RNA–LIGAND BINDING PROTOCOL

This subsection details the binding pipeline summarized in Sec. 4.3. RNA residues are represented by combining the frozen residue-level fingerprint $\mathbf{h}_{\bar{\theta},i}^{\mathrm{FP}}$ exported by RNADynNet with frozen sequence and secondary-structure features:

$$\widetilde{\mathbf{h}}_i^{\mathrm{FP}} = \mathrm{MLP}_{\mathrm{FP}}\left(\mathbf{h}_{\bar{\theta},i}^{\mathrm{FP}}\right), \quad \mathbf{h}_{\mathrm{RNA},i} = \mathrm{MLP}_{\mathrm{fuse}}\left(\widetilde{\mathbf{h}}_i^{\mathrm{FP}} \,\|\, \mathbf{e}_i^{\mathrm{1D}} \,\|\, \mathbf{e}_i^{\mathrm{2D}}\right). \tag{16}$$

The ligand encoder combines atom representations from a 2D graph convolutional network and a 3D Equiformer. The GerNA-Bind interaction stack (Xia et al., 2025) constructs residue–atom pair features, refines them with triangle multiplication and row-wise attention using intramolecular RNA and ligand distances, and pools the pair tensor with dual mutual attention. A prediction head maps the pooled representation to a binding score. Training uses the binary evidential objective from the GerNA-Bind implementation. No RNA–ligand complex pose or cross-molecule distance is required. Downstream optimization follows the released implementation. The fingerprint encoder and the pretrained RNA and ligand encoders from GerNA-Bind were frozen. The new RNA and ligand fusion layers, the retained interaction stack, and the prediction head were optimized from the released initialization. This setup evaluates the downstream utility of frozen RNADynNet fingerprints within the GerNA-Bind interaction and prediction framework.

D SUPPLEMENTARY RESULTS

D.1 GROUNDING TARGETS AND SUPERVISION DEPTH

Tabs. 8 and 9 compare the full model with variants retaining only covariance or only motion-coupling grounding. Block 24 moves the fingerprint readout and its alignment and PG supervision from Block 17 to Block 24; Block 17 supplies the main-text results. All generation variants use the same coarse and fine sampling schedule. Block 17 has the highest generation RMSF correlation on both test panels, whereas Block 24 has lower predictive RMWD on both panels.

Table 8: Trajectory-generation ablations. RNA-group macro means; distances are in Å and chiral flips in percent.

	test_struct						test_flex					
Model	Bond MAE↓	O3P max↓	Chiral flip↓	Clash /1k↓	RMSF r ↑	PCA W_2 ↓	Bond MAE↓	O3P max↓	Chiral flip↓	Clash /1k↓	RMSF r ↑	PCA W_2 ↓
Cov-only	0.02423	1.7091	0.03092	0.9583	0.8712	1.8699	0.02561	3.8573	0.00126	3.2085	0.7609	5.8609
Coupling-only	0.02414	1.0845	0.03088	0.9891	0.8631	1.8983	0.02537	5.4953	0.00093	2.8008	0.7623	5.8490
Full model (Block 24)	0.02430	1.2558	0.03091	0.9256	0.8720	1.8896	0.02586	6.2126	0.00043	3.1947	0.7592	5.8033
Full model (Block 17)	0.02424	1.8858	0.03085	0.8225	0.8746	1.8739	0.02567	4.2400	0.00093	3.0404	0.7658	5.9169

Table 9: Single-conformer dynamics ablations using independent probes on frozen EMA backbones. RNA-group macro means; RMWD and RMSF MAE are in Å.

	test_struct							test_flex						
	RMSF		Covariance			NMC		RMSF		Covariance			NMC	
Method	r ↑	MAE↓	RMWD↓	Sym. KL↓	IoU↑	r ↑	MAE↓	r ↑	MAE↓	RMWD↓	Sym. KL↓	IoU↑	r ↑	MAE↓
Cov-only	0.8540	0.4307	0.8856	0.6007	0.5309	0.7936	0.1722	0.7711	1.7128	3.3259	0.8238	0.4547	0.7537	0.2090
Coupling-only	0.8448	0.4535	0.9297	0.6690	0.4806	0.8160	0.1626	0.7468	1.7138	3.3845	0.9596	0.4322	0.7646	0.2012
Full model (Block 24)	0.8622	0.4124	0.8716	0.5892	0.5276	0.8016	0.1659	0.7696	1.5995	3.2786	0.8498	0.4476	0.7592	0.2023
Full model (Block 17)	0.8626	0.4076	0.8741	0.5823	0.5276	0.7958	0.1694	0.7782	1.7560	3.3359	0.8775	0.4522	0.7577	0.2035

D.2 PREDICTIVE DATA SCALING

Fig. 6 shows the seven metrics reported in Tab. 3 across both held-out splits and all training-data fractions and checkpoints.

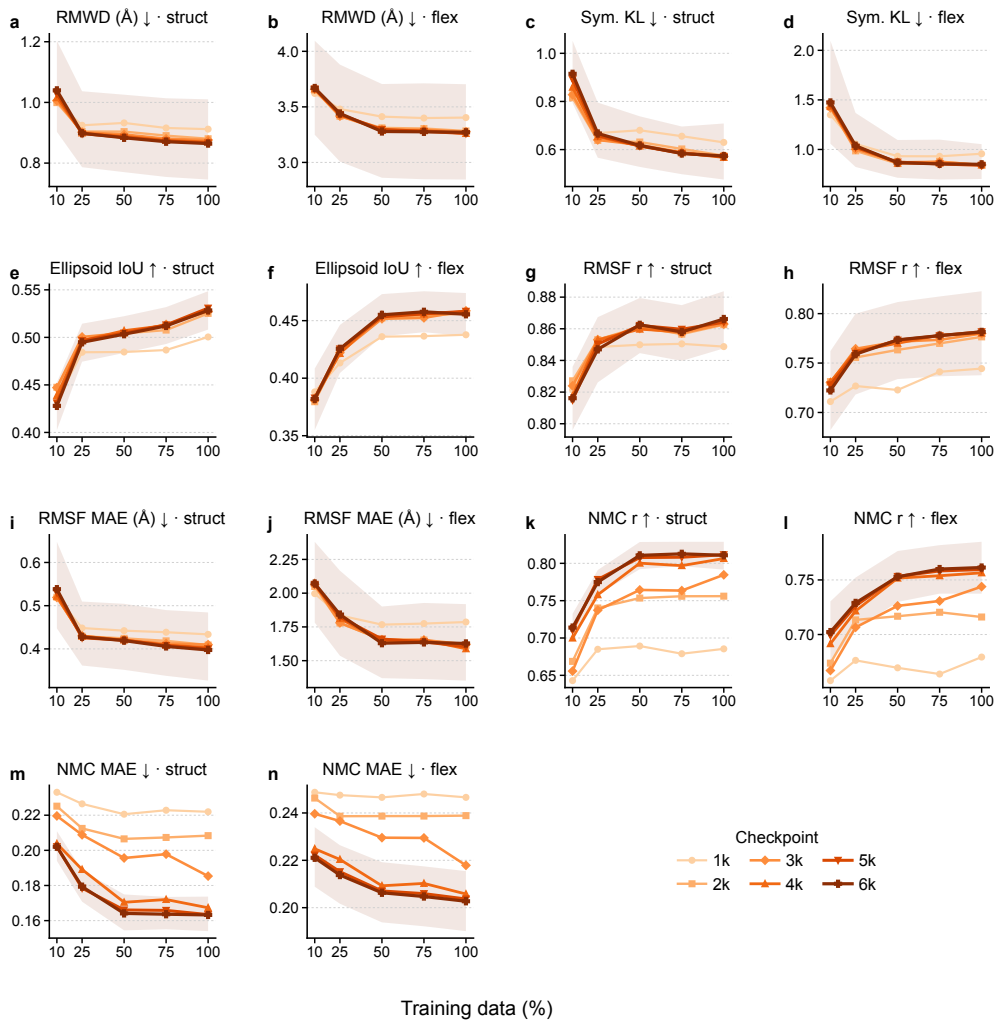


Figure 6: **Predictive data scaling.** Orange curves darken from 1k to 6k training updates across 10–100% of the training data. Shading shows pointwise 95% source-RNA-group bootstrap intervals for 6k only. Each checkpoint uses its EMA backbone and EMA descriptor heads.

D.3 QUALITATIVE TRAJECTORY GENERATION

Figs. 7 to 9 show stepwise trajectory generation for low-, median-, and high-motion held-out RNAs selected by reference-MD C1′ RMSD percentiles.

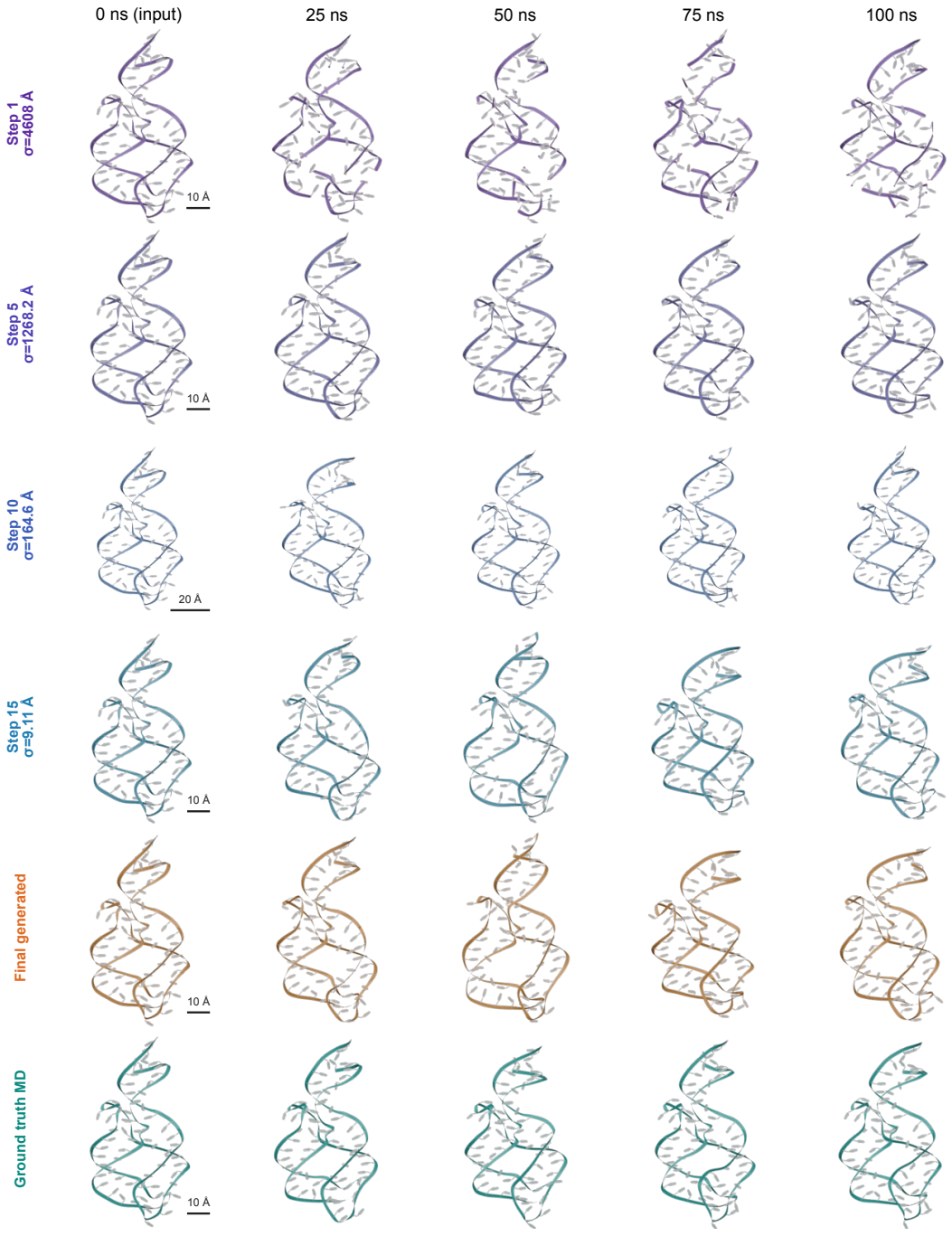


Figure 7: **Stepwise denoising of the low-motion held-out RNA 3IVN_A.** Columns show 0, 25, 50, 75, and 100 ns. Step rows are diagnostic $\hat{x}_0$ estimates at the labeled noise scales; the final two rows show the generated trajectory and reference MD. Ribbon gaps mark O3′–P distances of at least 2.6 Å.

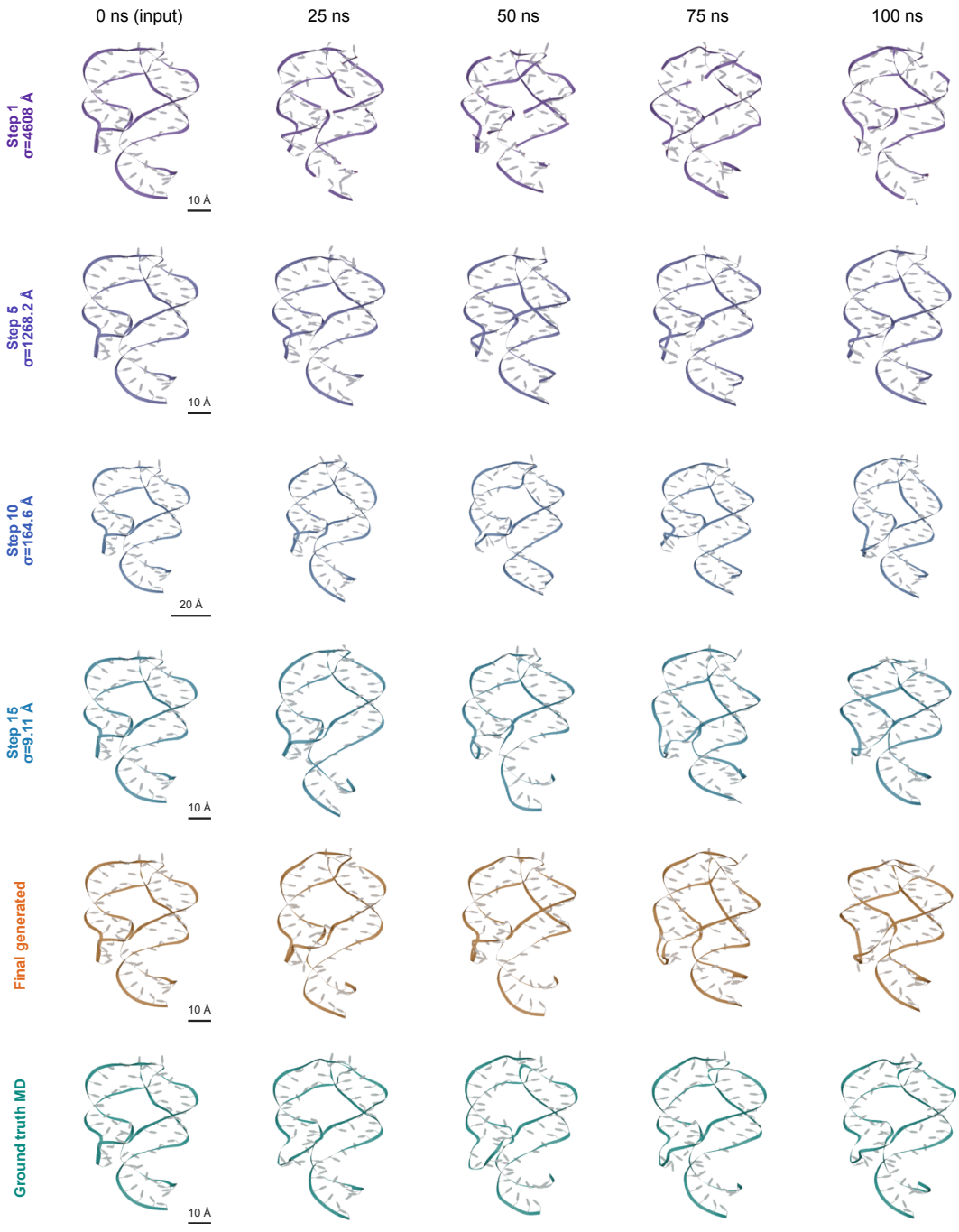


Figure 8: **Stepwise denoising of the median-motion held-out RNA 3SKR_A.** Columns show 0, 25, 50, 75, and 100 ns. Step rows are diagnostic $\hat{\mathbf{x}}_0$ estimates at the labeled noise scales; the final two rows show the generated trajectory and reference MD. Ribbon gaps mark O3′–P distances of at least 2.6 Å.

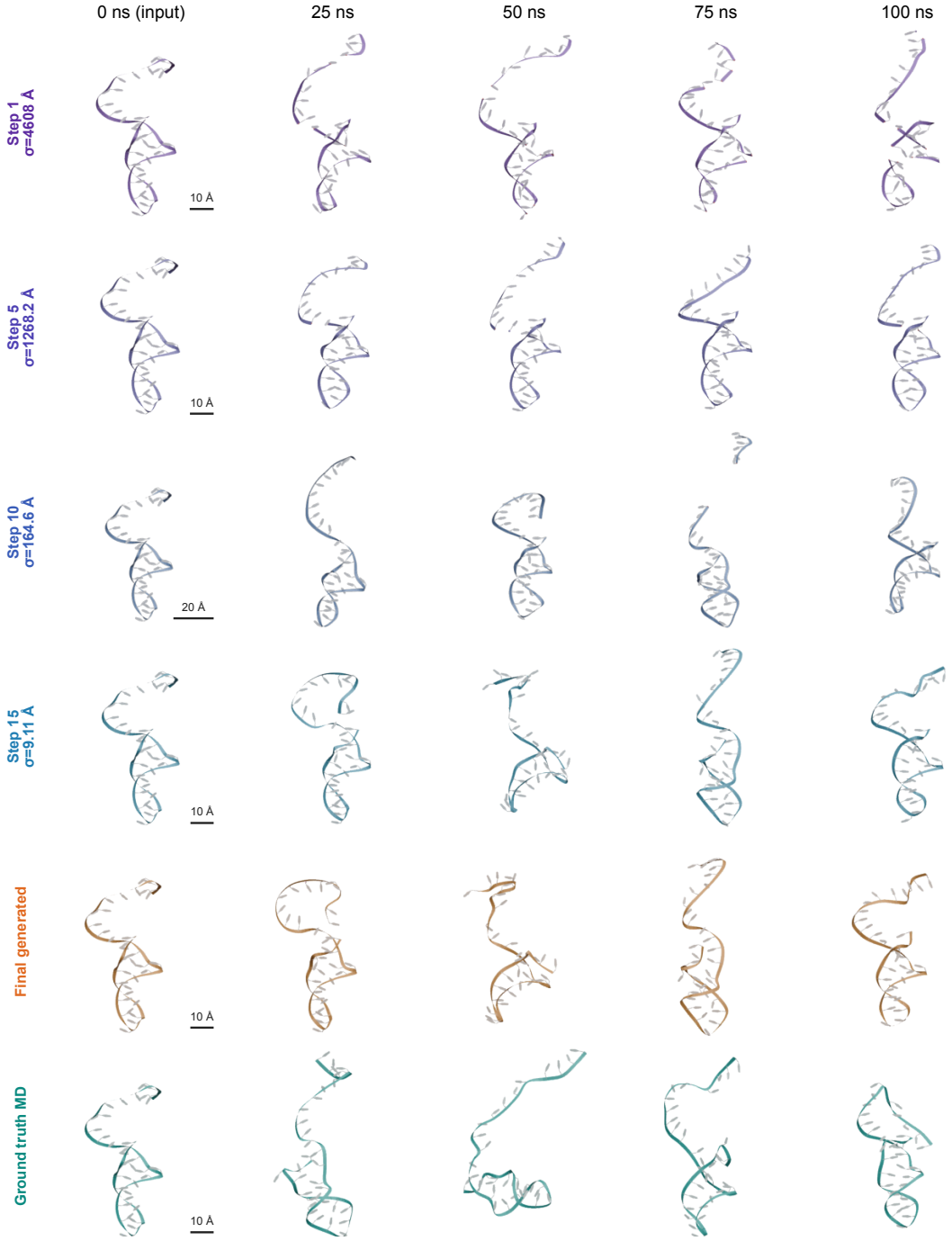


Figure 9: **Stepwise denoising of the high-motion held-out RNA 6AZ4_A.** Columns show 0, 25, 50, 75, and 100 ns. Step rows are diagnostic $\hat{\mathbf{x}}_0$ estimates at the labeled noise scales; the final two rows show the generated trajectory and reference MD. Ribbon gaps mark O3′–P distances of at least 2.6 Å.