REFRACT: GENERATING RNA CONFORMATIONAL ENSEMBLES WITH CONTEXT-AWARE FLOWS

Anonymous authors
Paper under double-blind review

ABSTRACT

Generating RNA conformational ensembles that reflect their molecular environment remains challenging. We introduce REFRACT, an $\mathrm{SE}(3)$-equivariant framework that couples flow-based reconstruction with context-aware transport. During training, the model learns structural regularities around observed conformations by reconstructing individual experimental structures from corrupted inputs, without requiring paired alternative conformations or molecular dynamics (MD) trajectories. CrossIPA incorporates fixed protein geometry through geometric cross-attention, conditioning RNA updates on the evolving RNA–protein spatial relationships. At inference, context-aware transport uses these learned updates to guide independently corrupted RNA conformations toward ensembles compatible with the supplied protein context. On protein-context ensemble recovery and conformational transition across binding conditions, REFRACT achieves average matched TM-scores of 0.669 and 0.648, respectively, exceeding the evaluated baselines. It also more closely matches nucleotide-level fluctuations and conformational distributions in MD reference ensembles. An ablation on the protein-bound benchmark shows that supplying protein context improves conformational recovery and reduces the proportion of generated structures with steric clashes.

1 INTRODUCTION

Many RNAs populate ensembles of distinct three-dimensional conformations. Their biological activities, including catalysis, molecular recognition, and regulation, can depend on both the conformations they adopt and the proportion of molecules in each conformation (Ganser et al., 2019; Ken et al., 2023). Experiments have revealed multiple conformations of the same RNA (Ding et al., 2023), but conformations that are short-lived or adopted by only a small fraction of molecules can be difficult to characterize (Dethoff et al., 2012). Molecular dynamics (MD) simulations complement experiments by modeling atomic motion over time, but sampling infrequent transitions between conformations can require long, computationally expensive simulations (Li et al., 2025). The conformations accessible to an RNA also depend on its molecular environment: bound proteins can stabilize particular conformations and restrict others (Ganser et al., 2019; Garg et al., 2024). We therefore study context-aware RNA ensemble generation: exploring alternative RNA conformations while accounting for how protein binding shapes the accessible conformational space.

Recent deep learning methods have improved the accuracy of RNA three-dimensional structure prediction. Examples include AlphaFold 3 (AF3) (Abramson et al., 2024) and RhoFold+ (Shen et al., 2024), with AF3 also predicting protein–RNA complex structures. However, accurate structure prediction alone does not show whether a method can recover multiple conformations of the same RNA under a given binding condition. RNA ensemble generators aim to capture this structural diversity. RNAbpFlow generates three-dimensional RNA structures conditioned on sequence and base-pair information (Tarafder & Bhattacharya, 2026). DynaRNA samples alternative conformations from an input RNA structure through partial noising and denoising (Li et al., 2025). Neither RNAbpFlow nor DynaRNA explicitly conditions exploration on bound protein structures. This motivates methods that recover alternative RNA conformations compatible with a given protein context and respond when that context changes.

The structures of bound proteins provide information about potential RNA–protein contacts and the space available to the RNA (Ganser et al., 2019; Garg et al., 2024). For example, an RNA conformation that appears plausible on its own may still place atoms in space occupied by a bound

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.

protein. Although an observed RNA structure captures one conformation, its coordinates alone do not fully specify how the surrounding protein may constrain alternative conformations. As candidate RNA conformations change, their distances and contacts with the protein can also change, even when the protein structure remains fixed. These observations motivate a model that learns local structural regularities through corruption and reconstruction and uses the evolving RNA–protein geometry to guide ensemble exploration.

We introduce REFRACT (RNA Ensemble generation via Flow-based Reconstruction And Context-aware Transport), an SE(3)-equivariant framework for context-aware RNA ensemble generation. During training, flow-based reconstruction uses RNA frame representations and flow matching (Anand et al., 2025) to recover individual experimental structures from corrupted inputs. Sharing this reconstruction model across training structures and corruption levels allows REFRACT to learn structural regularities around observed conformations without paired alternative states or MD trajectories. CrossIPA incorporates the supplied protein geometry through geometric cross-attention, so RNA updates depend on the evolving RNA–protein spatial relationships while protein coordinates remain fixed. At inference, independently corrupting an input RNA conformation produces multiple starting points for context-aware transport. The same reconstruction model then guides these inputs toward candidate ensembles compatible with the supplied protein context. Figure 1 illustrates how protein context redirects transport from the same corrupted inputs toward different candidate regions.

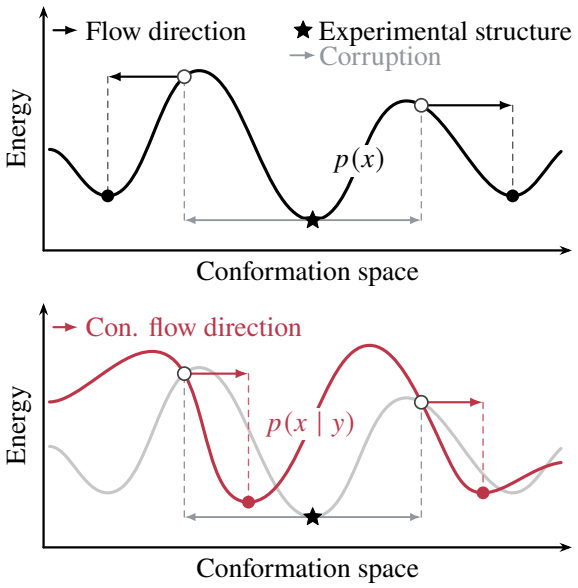


Figure 1: **Context-aware transport guides RNA ensemble generation.** Independent corruptions of an observed RNA conformation initialize transport. **Top:** The model generates candidate conformations without protein input. **Bottom:** Fixed protein geometry redirects transport from the same corrupted inputs toward different candidate regions according to the RNA–protein spatial relationships. The landscapes, transport paths, and energy axes are schematic.

We evaluate three complementary capabilities of context-aware RNA ensemble generation. For ensemble recovery under protein context, REFRACT achieves an average matched TM-score of 0.669 across 41 RNA clusters. For conformational transition across binding conditions, it reaches an average matched TM-score of 0.648 across 34 directed source–target pairs, supporting recovery of conformations associated with the changed context. Both results exceed the evaluated baselines. The generated ensembles also more closely match MD references than those from RNAbpFlow and DynaRNA, both in nucleotide-level fluctuations and conformational distributions. An ablation with the same trained model shows that protein context improves recovery on the protein-bound benchmark and reduces the fraction of structures with steric clashes from 32.7% to 14.1%.

2 METHODS

REFRACT couples flow-based reconstruction during training with context-aware transport during inference (Figure 2). Reconstruction learns structural regularities around observed RNA conformations from corrupted experimental structures. CrossIPA conditions the predicted RNA updates on the evolving RNA–protein geometry. At inference, these predictions define transport from independently corrupted RNA conformations toward candidate ensembles compatible with the supplied protein context.

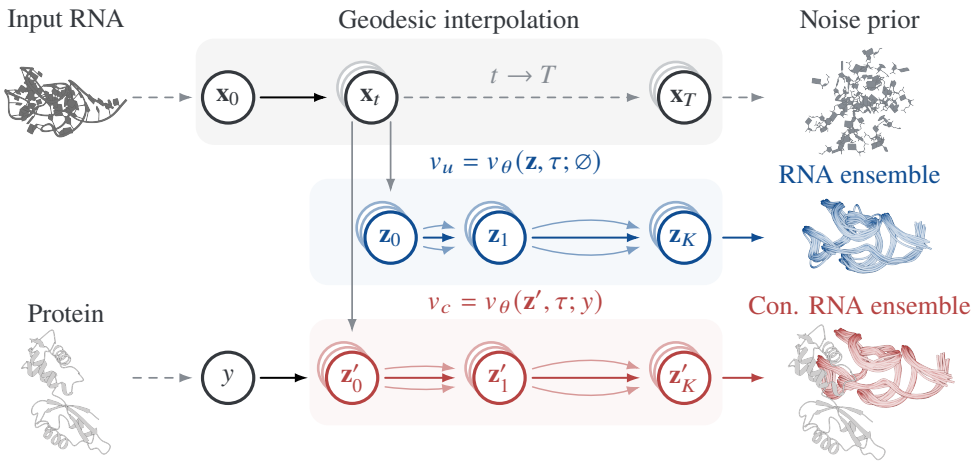


Figure 2: **REFRACT ensemble generation through context-aware transport.** Corruption interpolates the input RNA $\mathbf{x}_0$ toward noise $\mathbf{x}_T$, producing $\mathbf{x}_t$ for $0 < t < T$. Positions interpolate linearly and rotations follow geodesics. The same corrupted inputs initialize transport trajectories $\mathbf{z}_0$ and $\mathbf{z}'_0$ without and with fixed protein context y, respectively. Learned velocity fields v_u and v_c evolve these trajectories in generation time τ. Subscripts $0, \ldots, K$ index integration steps; independent corruptions produce ensembles of terminal conformations.

Problem Formulation and Structural Representation

Given an RNA sequence s, an experimentally determined starting structure A, and protein context $\mathcal{C}$, our goal is to generate an ensemble of N candidate RNA conformations $\mathcal{E} = \{G^{(1)}, \ldots, G^{(N)}\}$. We seek plausible alternatives to A that are compatible with the supplied protein geometry. Each candidate is generated by applying context-aware transport to an independently corrupted version of A. We write $\mathcal{C} = \varnothing$ for generation without protein input.

Following RNA-FrameFlow (Anand et al., 2025), we represent an RNA of length L by nucleotide frames $G \in \mathrm{SE}(3)^L$. Each frame $g = (R, x)$ specifies a nucleotide's position $x \in \mathbb{R}^3$ and orientation $R \in \mathrm{SO}(3)$. The origin x is located at $\mathrm{C4}'$, and the orientation R is defined by the $\mathrm{C3}'$–$\mathrm{C4}'$–$\mathrm{O4}'$ atoms. Together with these frames, torsion angles u, encoded as sine-cosine pairs, are used to reconstruct the modeled RNA atoms from nucleotide-specific ideal geometry. We omit nucleotide indices and apply frame-level expressions to each nucleotide.

Protein context $\mathcal{C}$ comprises amino-acid identities and backbone frames P, with origins y at $\mathrm{C}\alpha$ atoms. The protein structure is supplied with a known position and orientation relative to the starting RNA. We subtract the mean position of valid RNA $\mathrm{C4}'$ atoms from both RNA and protein coordinates, preserving their relative placement. Protein coordinates remain fixed throughout reconstruction and transport. Context selection and sampling settings are provided in Appendix A.

Flow-Based Reconstruction

We train a shared reconstruction model across individual experimental RNA structures and corruption levels to learn structural regularities around observed conformations. Conditioned on the RNA sequence and protein context, this training requires neither paired alternative conformations nor MD trajectories.

Following Figure 2, corruption time runs from a clean structure G_0 at $t = 0$ to noise G_T at $t = T$. We sample noise frames with centered Gaussian translations x_T and independent uniform rotations R_T on $\mathrm{SO}(3)$; the noise construction is detailed in Appendix A. For each training structure, we sample $t/T \sim \mathcal{U}(0.01, 0.5)$ and construct

$$\begin{aligned} x_t &= (1 - t/T)x_0 + (t/T)x_T, \\ R_t &= R_0 \exp\left[(t/T)\log(R_0^\top R_T)\right]. \end{aligned} \tag{1}$$

Positions interpolate linearly and orientations follow rotational geodesics. The model receives the corrupted frames G_t, sequence s, corruption time t, and unchanged protein context $\mathcal{C}$, and predicts the clean frames and torsions: $(\widehat{G}_0, \widehat{u}) = D_\theta(G_t, t, s, \mathcal{C})$.

Flow-matching objective. The clean-frame predictions parameterize the transport velocity: their displacement from the current state is divided by the remaining normalized corruption t/T. With an uncapped denominator t/T, the endpoint loss below matches the velocities induced by predicted and target clean frames. We cap the weight near the clean endpoint using $c(t) = \max(t/T, 0.1)$:

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}\left[\frac{\lambda\|\widehat{x}_0 - x_0\|_2^2 + \|\log_{R_t}(\widehat{R}_0) - \log_{R_t}(R_0)\|_2^2}{c(t)^2}\right]. \tag{2}$$

Here $\log_R(S) = \log(R^\top S)^\vee$ is the rotation vector from R to S expressed in local coordinates; $\vee$ converts a skew-symmetric matrix into a three-dimensional vector. Squared norms are averaged over valid nucleotides and coordinate components. With coordinates expressed in Å, $\lambda = 0.02$ weights translation errors relative to rotation errors. The time-dependent weight $1/c(t)^2$ is bounded above by 100.

Protein–RNA interface supervision. For protein–RNA complexes, a distance reconstruction loss encourages recovery of the observed interface geometry. Let $d_{ij} = \|x_{0,i} - y_j\|_2$ and $\widehat{d}_{ij} = \|\widehat{x}_{0,i} - y_j\|_2$, where $x_{0,i}$ and $\widehat{x}_{0,i}$ are target and predicted clean RNA C4′ coordinates, and y_j is a fixed protein Cα coordinate. For valid RNA–protein pairs $\mathcal{I}$ with $d_{ij} < 12$ Å, we define

$$\mathcal{L}_{\mathrm{protein}} = \mathbb{E}\left[\frac{1}{|\mathcal{I}|}\sum_{(i,j)\in\mathcal{I}} \rho(\widehat{d}_{ij} - d_{ij})\right]. \tag{3}$$

The inner average is zero when $\mathcal{I}$ is empty, including examples without protein input. With distances expressed in Å, ρ is the Smooth L_1 penalty: $\rho(e) = e^2/2$ for $|e| < 1$ and $\rho(e) = |e| - 1/2$ otherwise.

The full objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{FM}} + \mathcal{L}_{\mathrm{aux}} + \eta\mathcal{L}_{\mathrm{protein}}, \qquad \eta = 0.1, \tag{4}$$

where $\mathcal{L}_{\mathrm{aux}}$ combines RNA frame-atom, local-distance, and torsion reconstruction losses, detailed in Appendix A. Expectations average over training examples, noise, and corruption times, using sequence-balanced training weights.

CrossIPA: Protein-Aware Geometric Attention

CrossIPA conditions RNA reconstruction on the spatial relationships between the evolving RNA conformation and the fixed protein environment (Figure 3). It formulates joint RNA–protein geometric attention using the invariant point attention (IPA) construction (Jumper et al., 2021). Our reconstruction network adapts DenoisingIPA from Str2Str (Lu et al., 2024) to RNA, refining single representations $S = \{s_i\}$ and pair representations $Z = \{z_{ij}\}$ to update nucleotide frames $\mathcal{T} = \{T_i \in \mathrm{SE}(3)\}$.

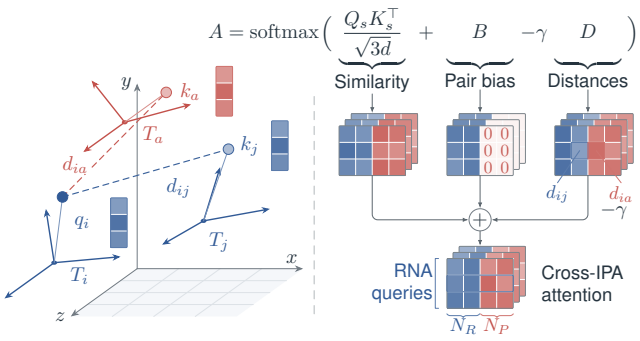


Figure 3: **CrossIPA incorporates protein context into RNA representations.** RNA queries attend jointly to RNA and protein keys using feature similarity, pair bias, and learned point distances. Protein keys have zero pair bias. Blue and red denote RNA and protein, respectively.

RNA nucleotides provide queries, while both RNA nucleotides and protein residues provide keys and values. CrossIPA combines feature similarity with distances between learned points in a shared coordinate system, allowing each nucleotide to attend jointly to RNA and protein geometry. Attention is invariant to a joint rotation or translation of

the complex. Protein features and frames remain fixed, while successive RNA updates change the relative geometry used by subsequent attention blocks.

For one refinement block, let $\mathbf{A}$ denote the CrossIPA attention weights, $\mathbf{V}$ its scalar, point, and RNA pair values, and F the combined RNA and protein frames. Point values $\mathbf{V}_{\mathrm{p}}$ are aggregated as $\mathcal{T}^{-1}[\mathbf{A}F(\mathbf{V}_{\mathrm{p}})]$: they are mapped into shared coordinates, averaged by attention, and expressed in each query RNA frame. Together with scalar and pair messages, they update the single representations and RNA frames:

$$\begin{aligned} S' &= f_{\mathrm{single}}(S, \mathrm{Aggregate}(\mathbf{A}, \mathbf{V}; \mathcal{T}, F)), \\ \mathcal{T}' &= \mathcal{T} \circ f_{\mathrm{frame}}(S'). \end{aligned} \tag{5}$$

Primes denote updates within the reconstruction network. The aggregation follows IPA, f_{single} contains the DenoisingIPA representation updates, and f_{frame} predicts local rigid transformations. Composing these transformations with the RNA frames updates nucleotide positions and orientations equivariantly.

PairTransition passes the protein information in the updated singles into RNA–RNA pair representations:

$$z'_{ij} = f_{\mathrm{pair}}\big(z_{ij} \| s'_i \| s'_j\big), \tag{6}$$

where $\|$ denotes concatenation and f_{pair} is the learned pair update. Repeated refinement produces the clean-frame prediction $\widehat{G}_0$. Without protein input, CrossIPA reduces to standard IPA over RNA queries, keys, and values.

Context-Aware Transport for Ensemble Generation

At inference, independent corruptions of the input conformation A initialize transport:

$$Z_0^{(n)} = \mathrm{Interp}(A, G_T^{(n)}; t_s/T). \tag{7}$$

Here $G_T^{(n)}$ are independent noise draws, and Interp follows the linear and geodesic rules in Eq. (1). The starting corruption t_s/T controls how much input information is retained: larger values produce stronger corruption.

For a chosen t_s, let τ measure elapsed transport time in units of T, so the remaining corruption is $t(\tau) = t_s - T\tau$. For the current frames $Z_\tau = (R_\tau, z_\tau)$, the same reconstruction model predicts $(\widehat{G}_0, \widehat{u}) = D_\theta(Z_\tau, t(\tau), s, \mathcal{C})$ at every step. The input A determines initialization; subsequent predictions depend on the current state and context and are not fixed to A. CrossIPA reevaluates the RNA–protein geometry throughout transport, defining the context-aware velocity field

$$\frac{dz_\tau}{d\tau} = \frac{T}{t(\tau)}(\widehat{x}_0 - z_\tau), \qquad \frac{dR_\tau}{d\tau} = R_\tau \left[\frac{T}{t(\tau)} \log_{R_\tau}(\widehat{R}_0)\right]_\times, \tag{8}$$

where $[\cdot]_\times$ maps a three-dimensional rotation vector into a skew-symmetric matrix. We integrate from $\tau = 0$ until $t(\tau) = t_f = 0.01T$, using Euler updates for translations and exponential-map updates for rotations. The discrete states $Z_0, \ldots, Z_K$ index integration steps, as in Figure 2. Each candidate is decoded from the final predicted clean frames and torsion angles.

Unless otherwise specified, we generate $N/4$ candidates at each corruption level $t_s/T \in \{0.2, 0.3, 0.4, 0.5\}$ and pool them into an ensemble of N conformations. Independent corruptions provide different starting states, while protein context guides their transport. These samples do not estimate physical transition times or provide calibrated equilibrium probabilities.

3 Experiments

We evaluate the three capabilities introduced in Section 1: recovering RNA ensembles compatible with a supplied protein context (Benchmark I), guiding conformational transition across binding conditions (Benchmark II), and capturing nucleotide-level fluctuations and conformational distributions in MD references (Benchmark III).

3.1 EXPERIMENTAL SETUP

Dataset curation. We collect RNA-only training structures from RNAsolo 2.0 (Adamczyk et al., 2026) and RNA–protein complexes from PRI30k (Han et al., 2025), using January 2024 as the cutoff for experimental structure release dates. The experimental evaluation sets, RBP, RBPTrans, and RNA-only, are curated from post-cutoff structures. CD-HIT filtering excludes evaluation RNAs with sequence identity of at least 80% to a training sequence. All conformations and binding conditions of each retained RNA remain together in the evaluation set.

RBP contains 41 clusters, each with distinct experimental conformations of the same RNA bound to the same protein partner. **RBPTrans** contains 34 directed source–target pairs from 9 RNAs, spanning protein binding, protein removal, and changes in protein partners. The **MD** dataset comprises 23 reference ensembles from 11 RNA systems, covering free and protein-bound conditions and alternative force fields. These publicly deposited trajectories (Sabei et al., 2024; Baltrukevich & Bartos, 2023; Choi et al., 2024) are reserved for evaluation; retained systems and conditions are listed in Appendix D.

Baselines. We compare with AF3 (Abramson et al., 2024) and RHOFOLD+ (Shen et al., 2024) with MSA subsampling, together with RNABPFLOW (Tarafder & Bhattacharya, 2026) and DYNARNA (Li et al., 2025). MSA subsampling follows the conformational sampling strategy studied with AF2 (del Alamo et al., 2022). RNABPFLOW uses sequence and base-pair information, while DYNARNA and REFRACT receive the same experimental starting RNA structure on the experimental-state benchmarks. We use released baseline checkpoints without retraining. MD comparisons include RNABPFLOW and DYNARNA.

Sampling protocol. Unless otherwise specified, REFRACT generates N conformations by pooling $N/4$ samples from each corruption level $t_s/T \in \{0.2, 0.3, 0.4, 0.5\}$, using independent noise draws and the same trained model. Integration and decoding follow Section 2; protein coordinates remain fixed, and protein removal is represented by $\mathcal{C} = \varnothing$. We use $N = 100$ for RBP, RNA-only recovery, and each directed RBPTrans pair. MD comparisons use $1{,}000$ generated samples per method and the same $1{,}000$ sampled reference conformations per RNA–condition case. When multiple registered structure–context pairs are supplied, each corruption level's sample budget is distributed approximately equally across them.

3.2 EVALUATION METRICS

Metric definitions and computation details are provided in Appendix B.

Structural recovery and diversity. Recall-oriented average matched TM-score (AMR-R) measures how closely an ensemble matches the observed states, while recall coverage (COV-R) measures the fraction recovered across similarity thresholds. TM-score and RMSD assess individual structural matches, and pairwise RMSD measures ensemble diversity.

Steric compatibility. Clash measures the percentage of structures with severe RNA–protein overlaps; lower values indicate better compatibility with the supplied protein geometry.

Agreement with MD ensembles. Root-mean-square fluctuation (RMSF) error and correlation assess fluctuation amplitudes and flexibility patterns; supplementary plots also report mean absolute error (MAE). Ensemble RMSD assesses structural similarity, while coordinate-space and MD-PCA Wasserstein-2 distances compare conformational distributions. Principal-component similarity (PC-sim) assesses agreement between dominant modes of variation. Lower errors and distances, and higher correlations and PC-sim, indicate closer agreement with MD.

3.3 BENCHMARK I: RNA ENSEMBLE RECOVERY UNDER PROTEIN CONTEXT

A context-aware RNA ensemble should recover alternative conformations while respecting the geometry of the bound protein. We assess these requirements jointly through experimental-state recovery, steric compatibility, and ensemble diversity, testing exploration within a shared binding environment. We evaluate this capability on RBP.

Table 1 shows that REFRACT achieves the highest AMR-R and COV-R at all thresholds on RBP. AMR-R reaches 0.669, compared with 0.385 for RNABPFLOW, the strongest baseline by this metric.

Table 1: **Experimental-state recovery on RBP (41 clusters).** Clash is a percentage; diversity is in Å. Best AMR-R, COV-R, and Clash values are **bold**.

Method	AMR-R ↑	COV-R ↑				Clash (%) ↓	Diversity (Å)
		@0.50	@0.60	@0.70	@0.80		
AF3 subMSA	0.381	0.195	0.144	0.077	0.000	78.9	2.47
RhoFold+ subMSA	0.349	0.232	0.163	0.087	0.033	76.0	0.38
RNAbpFlow	0.385	0.213	0.154	0.081	0.016	81.3	3.09
DynaRNA	0.328	0.098	0.085	0.006	0.000	81.2	4.53
REFRACT	**0.669**	**0.817**	**0.642**	**0.491**	**0.280**	**14.1**	2.52
REFRACT w/o protein	0.643	0.734	0.574	0.417	0.268	32.7	2.59

Table 2: **RNA-only ensemble recovery (17 clusters).** Metrics follow Table 1; best recovery values are **bold**.

Method	AMR-R ↑	COV-R ↑				Diversity (Å)
		@0.50	@0.60	@0.70	@0.80	
AF3 subMSA	0.629	0.735	0.549	0.353	0.185	2.94
RhoFold+ subMSA	0.365	0.294	0.147	0.088	0.000	1.59
RNAbpFlow	0.397	0.176	0.098	0.000	0.000	4.76
DynaRNA	0.305	0.000	0.000	0.000	0.000	6.08
REFRACT	**0.815**	**0.951**	**0.843**	**0.755**	**0.755**	2.73

Coverage reaches 0.817 at a TM-score threshold of 0.50, supporting recovery of multiple experimental states. RNAbpFlow and DynaRNA produce more dispersed ensembles but lower coverage, showing that greater structural spread alone does not ensure recovery of observed conformations.

The protein ablation isolates the effect of protein input by omitting it only at inference, while retaining the trained model, RNA inputs, and sample budget. Adding protein input improves AMR-R from 0.643 to 0.669 and reduces clashes from 32.7% to 14.1%, while diversity remains similar (2.59 versus 2.52 Å). Protein geometry therefore improves both recovery and steric compatibility without substantially reducing ensemble diversity, supporting its role in constraining exploration within the binding environment.

In addition to RBP, we evaluate RNA-only ensembles whose alternative conformations are experimentally observed without bound protein. This tests recovery of conformational diversity present in protein-free RNA. Using the same trained model without protein input, REFRACT achieves AMR-R 0.815, exceeding the strongest baseline, AF3 subMSA (0.629), with the highest coverage at all thresholds (Table 2). Per-cluster comparisons are provided in Appendix C.1.

3.4 Benchmark II: Protein-Guided Conformational Transition

Changes in protein binding can alter the conformations accessible to an RNA. We test whether changing the supplied context directs generation from a source conformation toward experimentally observed states associated with the target binding condition. We evaluate this capability on RBPTrans through target-state recovery.

Figure 4 reports a mean AMR-R of 0.648 for REFRACT, compared with 0.437 for RhoFold+ subMSA, the strongest baseline by this metric. With the same trained model, source RNA, and sample budget, supplying

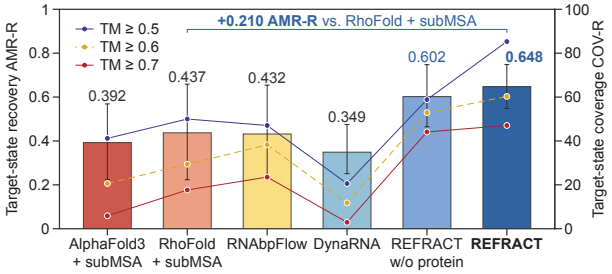


Figure 4: **RBPTrans recovery.** Bars show mean AMR-R (left axis) over 34 directed pairs (9 RNAs), with 95% RNA-cluster bootstrap confidence intervals (20,000 resamples). Curves show COV-R (%, right axis) at TM-score thresholds of 0.5, 0.6, and 0.7.

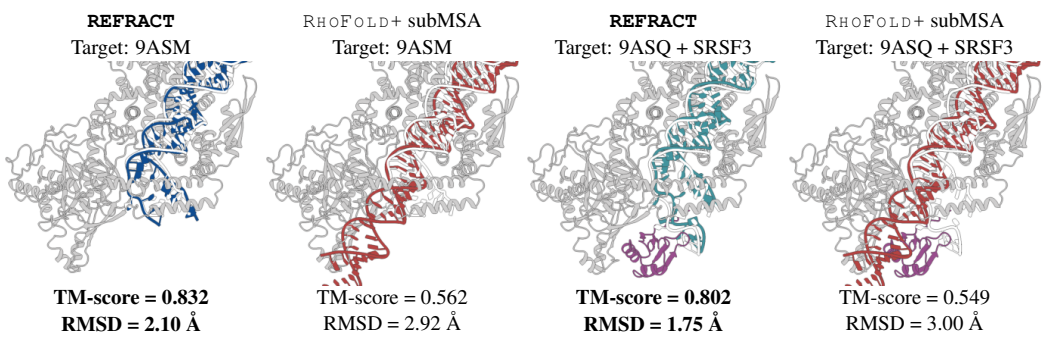


Figure 5: **Conformational transition case study.** The targets are pri-let-7f1 bound to Drosha–DGCR8 in 9ASM on the left and to Drosha–DGCR8 with additional SRSF3 in 9ASQ on the right (Garg et al., 2024). The displayed conformations from REFRACT and RhoFold+ subMSA are evaluated against the corresponding experimental targets using TM-score and RMSD.

Table 3: **Agreement with MD reference ensembles.** RMSF error, ensemble RMSD, and coordinate-space W_2 are in Å. Best values are **bold**; N/A marks undefined metrics for the input structure.

Method	RMSF error ↓	RMSF r ↑	Ens. RMSD ↓	W_2 ↓	MD-PCA W_2 ↓	PC-sim ↑
Input structure	2.36	N/A	2.67	4.06	20.27	N/A
Local perturb.	2.15	−0.274	4.34	5.40	18.65	0.069
RNAbpFlow	3.54	0.538	4.63	7.11	23.20	0.331
DynaRNA	6.96	0.255	8.15	11.49	18.45	0.260
REFRACT	**1.31**	**0.674**	**2.47**	**3.72**	**17.21**	**0.507**

the target protein context increases AMR-R from 0.602 to 0.648 and COV-R@0.50 from 0.588 to 0.853. This supports the role of context in guiding exploration toward states associated with a different binding condition. Coverage gains are largest at the 0.50 threshold.

The pri-let-7f1 case illustrates context-guided conformational transition when the binding partners change (Figure 5). The two targets share the Drosha–DGCR8 complex but differ in the presence of SRSF3. REFRACT recovers distinct conformations matching the respective targets, reaching TM-scores of 0.832 (RMSD 2.10 Å) for 9ASM and 0.802 (1.75 Å) for 9ASQ. The displayed RhoFold+ subMSA conformations remain similar across the two conditions, with lower TM-scores of 0.562 and 0.549. In this case, protein-guided generation captures the conformational change associated with the binding environment, which is absent from the RhoFold+ inputs.

3.5 Benchmark III: Agreement with MD Reference Ensembles

RNA ensembles should capture local flexibility and collective conformational variation under a given binding condition. We test this capability using the MD reference dataset, without MD training. Structure-initialized methods start from the initial RNA structure of the corresponding MD simulation, so the comparison assesses the distribution and fluctuations associated with that starting state and condition. Input structure measures agreement already provided by the initial conformation, while Local perturb. tests whether local variation alone is sufficient.

Table 3 shows that REFRACT achieves the best aggregate values across all six metrics. RMSF error is 1.31 Å, compared with 3.54 Å for RNAbpFlow and 6.96 Å for DynaRNA, supporting more accurate fluctuation amplitudes. RMSF correlation reaches 0.674, versus −0.274 for Local perturb., showing that learned exploration better captures where flexibility occurs along the RNA. The lowest coordinate-space and MD-PCA W_2 and highest PC-sim indicate closer agreement in conformational distributions and dominant modes of variation. Gains over Input structure are smaller for ensemble RMSD and coordinate-space W_2.

Condition-specific results support agreement both with and without protein constraints (Appendix E, Table 5). REFRACT leads on five of six metrics under free conditions and all six against external baselines under bound conditions; DynaRNA has the lowest free-condition MD-PCA W_2. For bound ensembles, supplying protein input improves five metrics over the same model without protein,

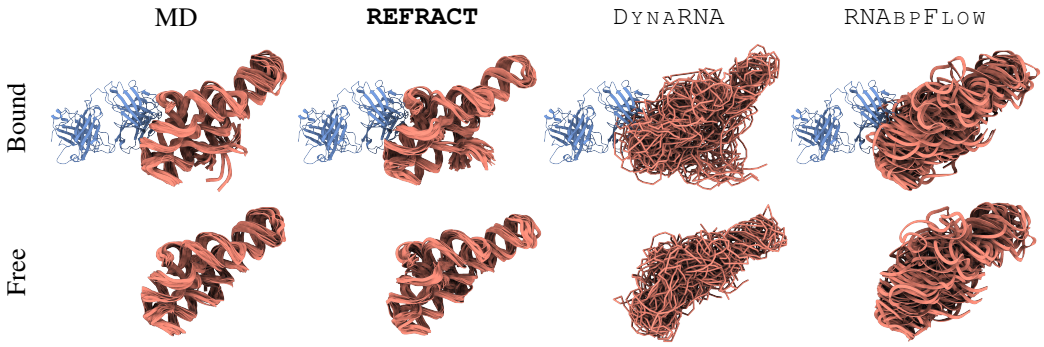


Figure 6: **MD and generated ensembles for the P4–P6 RNA domain.** The top row shows Fab-bound ensembles for PDB 2R8S, RNA chain R; the bottom row shows protein-free ensembles for PDB 1HR2. Columns show MD, REFRACT, DynaRNA, and RNAbpFlow. Overlaid RNA conformations are shown in salmon, with the fixed Fab reference in blue for the bound condition. RNA is displayed as cartoons where full backbone atoms are available and as C4′ traces for DynaRNA.

supporting its contribution to flexibility patterns and conformational distributions. RMSF error is slightly lower in the ablation (0.79 versus 0.81 Å).

Figure 6 shows that REFRACT preserves the P4–P6 fold observed in MD under both free and Fab-bound conditions, while DynaRNA and RNAbpFlow show greater dispersion. The L11-binding RNA example in Appendix E shows similar behavior (Figure 9). At the nucleotide level, Figure 8 presents the eight lowest-MAE RMSF profiles among the 20 available cases, illustrating agreement in local flexibility. For example, 1RKJ reaches correlations of 0.804 and 0.859 under free and bound conditions, respectively. These profiles complement the structural overlays by showing how flexibility varies along the RNA; all available profiles are summarized in Table 6.

4 Related Work

RNA structure prediction. Early RNA structure predictors use motif and fragment assembly, as in MC-Fold/MC-Sym, RNAComposer, and 3dRNA (Parisien & Major, 2008; Popenda et al., 2012; Zhao et al., 2012), or conformational sampling and refinement, as in SimRNA and FARFAR2 (Boniecki et al., 2016; Watkins et al., 2020). Learning-based methods, including trRosettaRNA, RhoFold+, NuFold, and AlphaFold 3, predict structures from sequence, evolutionary, and geometric information (Wang et al., 2023; Shen et al., 2024; Kagaya et al., 2025; Abramson et al., 2024). AlphaFold 3 additionally models protein–RNA complexes.

Flow matching. Flow matching learns transport by regressing velocity fields along probability paths (Lipman et al., 2022), with rectified flow and minibatch optimal transport offering different trajectory and coupling choices (Liu et al., 2022; Tong et al., 2023a). Stochastic interpolants connect flows and diffusions (Albergo et al., 2025); related approaches include score-based diffusion (Song et al., 2020) and Schrödinger bridges (Shi et al., 2023; Tong et al., 2023b). Geometric diffusion and flow models generate protein backbones using rigid-body frames (Yim et al., 2023b; Bose et al., 2023; Yim et al., 2023a), and flow matching extends more generally to Lie groups (Sherry & Smets, 2025). RNA-FrameFlow generates RNA backbones (Anand et al., 2025); RNAbpFlow conditions generation on sequence and base pairs (Tarafder & Bhattacharya, 2026), while DynaRNA samples alternative conformations through partial noising and denoising (Li et al., 2025).

5 Conclusion

We presented REFRACT, a flow-based framework for RNA conformational exploration with optional protein context. CrossIPA couples evolving RNA conformations to fixed protein geometry, enabling generation to respond to the binding environment. Trained on individual experimental structures, REFRACT improves experimental-state recovery, supports conformational transition across protein contexts, and produces ensembles with closer agreement to MD references. The protein ablation supports the contribution of context to recovery and steric compatibility. Protein coordinates remain fixed, and samples do not provide calibrated state populations or transition rates.

AI USE STATEMENT

In this work, we used generative AI tools to assist with writing auxiliary code and data-processing scripts. We have **NOT** used generative AI tools to generate synthetic datasets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation, clean and reformat datasets, support qualitative and thematic data analysis, or interpret results. We have reviewed all AI-assisted work and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work is a computational study of RNA conformational ensembles using publicly available experimental structures and MD reference data. It does not involve new experiments with human participants or animals. Generated conformations are computational hypotheses; agreement with structural or MD references does not establish biological function or experimental validity.

REPRODUCIBILITY STATEMENT

To support reproducibility, we document the model formulation, reconstruction losses, and transport procedure in Section 2 and Appendix A. We identify the public data sources and report release-date and sequence-identity filters, baseline inputs, and sampling budgets in Section 3.1, with the MD reference systems and retained conditions listed in Appendix D. To make the reported comparisons verifiable, we provide explicit metric definitions, alignment and aggregation rules, and bootstrap procedures with fixed random seeds in Appendix B. We also report all available residue-level RMSF results in Appendix E, so the selected examples can be assessed alongside the complete profile set.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

Bartosz Adamczyk, Pawel Boinski, Marta Szachniuk, and Maciej Antczak. RNAsolo 2.0: Multimodal database to study RNAs, their structural families and intermolecular interfaces. *J. Mol. Biol.*, 438(18):169570, 2026.

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *J. Mach. Learn. Res.*, 26(209):1–80, 2025.

Rishabh Anand, Chaitanya K. Joshi, Alex Morehead, Arian Rokkum Jamasb, Charles Harris, Simon V. Mathis, Kieran Didi, Rex Ying, Bryan Hooi, and Pietro Liò. RNA-FrameFlow: Flow matching for de novo 3D RNA backbone design. *Transact. Mach. Learn. Res.*, 2025.

Hanna Baltrukevich and Piia Bartos. RNA-protein complexes and force field polarizability. *Front. Chem.*, 11:1217506, 2023.

Michal J. Boniecki, Grzegorz Lach, Wayne K. Dawson, Konrad Tomala, Pawel Lukasz, Tomasz Soltysinski, Kristian M. Rother, and Janusz M. Bujnicki. SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, 44(7):e63, 2016.

Avishek Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. *CoRR*, abs/2310.02391, 2023.

Taeyoung Choi, Zhengxin Li, Ge Song, and Hai-Feng Chen. Comprehensive comparison and critical assessment of RNA-specific force fields. *J. Chem. Theory Comput.*, 20(6):2676–2688, 2024.

Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, 2022.

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.

Jienyu Ding, Yun-Tzai Lee, Yuba Bhandari, Charles D. Schwieters, Lixin Fan, Ping Yu, Sergey G. Tarosov, Jason R. Stagno, Buyong Ma, Ruth Nussinov, Alan Rein, Jinwei Zhang, and Yun-Xing Wang. Visualizing RNA conformational and architectural heterogeneity in solution. *Nat. Commun.*, 14(1):714, 2023.

Laura R. Ganser, Megan L. Kelly, Daniel Herschlag, and Hashim M. Al-Hashimi. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.*, 20(8):474–489, 2019.

Ankur Garg, Renfu Shang, Todor Cvetanovic, Eric C. Lai, and Leemor Joshua-Tor. The structural landscape of Microprocessor-mediated processing of pri-let-7 miRNAs. *Mol. Cell*, 84(21):4175–4190.e6, 2024.

Rong Han, Xiaohong Liu, Tong Pan, Jing Xu, Xiaoyu Wang, Wuyang Lan, Zhenyu Li, Zixuan Wang, Jiangning Song, Guangyu Wang, and Ting Chen. CoPRA: Bridging cross-domain pre-trained sequence models with complex structures for protein–RNA binding affinity prediction. In *AAAI'25: Proc. of the 2025 AAAI Conference on Artificial Intelligence*, pp. 246–254. AAAI Press, 2025.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

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. *Nat. Commun.*, 16(1):881, 2025.

Megan L. Ken, Rohit Roy, Ainan Geng, Laura R. Ganser, Akanksha Manghrani, Bryan R. Cullen, Ursula Schulze-Gahmen, Daniel Herschlag, and Hashim M. Al-Hashimi. RNA conformational propensities determine cellular activity. *Nature*, 617(7962):835–841, 2023.

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. *Commun. Biol.*, 8(1):1472, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *CoRR*, abs/2210.02747, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *CoRR*, abs/2209.03003, 2022.

Jiarui Lu, Bozitao Zhong, Zuobai Zhang, and Jian Tang. Str2Str: A score-based framework for zero-shot protein conformation sampling. In *ICLR'24: Proc. of the 2024 International Conference on Learning Representations*, pp. 10663–10694, 2024.

Marc Parisien and François Major. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, 452(7183):51–55, 2008.

Mariusz Popenda, Marta Szachniuk, Maciej Antczak, Katarzyna J. Purzycka, Piotr Lukasiak, Natalia Bartol, Jacek Blazewicz, and Ryszard W. Adamiak. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, 40(14):e112, 2012.

Afra Sabei, Cécilia Hognon, Juliette Martin, and Elisa Frezza. Dynamics of protein–RNA interfaces using all-atom molecular dynamics simulations. *J. Phys. Chem. B*, 128(20):4865–4886, 2024.

Tao Shen, Zhihang Hu, Siqi Sun, Di Liu, Felix Wong, Jiuming Wang, Jiayang Chen, Yixuan Wang, Liang Hong, Jin Xiao, Liangzhen Zheng, Tejas Krishnamoorthi, Irwin King, Sheng Wang, Peng Yin, James J. Collins, and Yu Li. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat. Methods*, 21(12):2287–2298, 2024.

Finn M. Sherry and Bart M. N. Smets. Flow matching on Lie groups. *CoRR*, abs/2504.00494, 2025.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *NeurIPS'23: Proc. of the 2023 Conference on Neural Information Processing Systems*, pp. 62183–62223. Curran Associates, Inc., 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *CoRR*, abs/2011.13456, 2020.

Sumit Tarafder and Debswapna Bhattacharya. RNAbpFlow: Base pair-augmented SE(3) flow matching for conditional RNA 3D structure generation. *Nat. Methods*, 23(7):1349–1358, 2026.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *CoRR*, abs/2302.00482, 2023a.

Alexander Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. *CoRR*, abs/2307.03672, 2023b.

Wenkai Wang, Chenjie Feng, Renmin Han, Ziyi Wang, Lisha Ye, Zongyang Du, Hong Wei, Fa Zhang, Zhenling Peng, and Jianyi Yang. trRosettaRNA: Automated prediction of RNA 3D structure with transformer network. *Nat. Commun.*, 14(1):7266, 2023.

Andrew Martin Watkins, Ramya Rangan, and Rhiju Das. FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. *Structure*, 28(8):963–976.e6, 2020.

Jason Yim, Andrew Campbell, Andrew Y. K. Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S. Veeling, Regina Barzilay, Tommi Jaakkola, and Frank Noé. Fast protein backbone generation with SE(3) flow matching. In *NeurIPS'23: Machine Learning for Structural Biology Workshop*, 2023a.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *ICML'23: Proc. of the 2023 International Conference on Machine Learning*, pp. 40001–40039. PMLR, 2023b.

Yunjie Zhao, Yangyu Huang, Zhou Gong, Yanjie Wang, Jianfen Man, and Yi Xiao. Automated and fast building of three-dimensional RNA structures. *Sci. Rep.*, 2(1):734, 2012.

A Auxiliary Reconstruction Loss and Sampling Settings

The auxiliary loss complements frame supervision with local geometry and torsion supervision. Predicted frames $\widehat{G}_0$ and torsions $\widehat{u}$ determine atom coordinates through a differentiable reconstruction map $\mathcal{B}$ adapted from RNA-FrameFlow (Anand et al., 2025), using the nucleotide-specific ideal geometry described in Section 2:

$$\widehat{Y} = \mathcal{B}(\widehat{G}_0, \widehat{u}; s), \qquad Y = \mathcal{B}(G_0, u; s). \tag{9}$$

Here s specifies the nucleotide identities, and both coordinate sets use nucleotide-specific ideal geometry. Let $\mathcal{A}$ contain the $\mathrm{C3}'$, $\mathrm{C4}'$, and $\mathrm{O4}'$ atoms of valid nucleotides. The coordinate and distance losses act on these frame atoms, while torsion supervision constrains the remaining internal geometry. With coordinates in Å, the per-structure atom loss is

$$\ell_{\mathrm{atom}} = \frac{0.01}{c(t)^2\,|\mathcal{A}|} \sum_{a \in \mathcal{A}} \|\widehat{Y}_a - Y_a\|_2^2, \tag{10}$$

where $c(t) = \max(t/T, 0.1)$ is the same capped time factor used in the frame objective. The factor 0.01 is the square of the coordinate scaling factor 0.1.

To supervise local spatial relationships, let $\mathcal{P}$ contain distinct ordered pairs of atoms in $\mathcal{A}$ whose separation is below 6 Å in either the target or predicted structure. The distance loss is

$$\ell_{\mathrm{dist}} = \frac{0.01}{c(t)^2\,|\mathcal{P}|} \sum_{(a,b) \in \mathcal{P}} \Big(\|\widehat{Y}_a - \widehat{Y}_b\|_2 - \|Y_a - Y_b\|_2 \Big)^2 . \tag{11}$$

Including neighbors from both structures penalizes disrupted target contacts and artificially close predicted pairs. Self-pairs are excluded, and the cutoff is applied before coordinate and time scaling.

For torsions, u and $\widehat{u}$ denote target and predicted sine–cosine pairs. We average their squared Euclidean errors over observed torsions within each valid nucleotide, then over valid nucleotides:

$$\ell_{\mathrm{tors}} = \operatorname*{mean}_{\text{nucleotide}} \left[\operatorname*{mean}_{\text{observed torsions}} \|\widehat{u} - u\|_2^2 \right]. \tag{12}$$

Only torsions with all defining atoms resolved are supervised; a nucleotide with no observed torsions contributes zero. Empty atom or pair selections likewise contribute zero.

The three terms have unit relative weights and are active throughout the training interval. Their sequence-balanced expectation defines the auxiliary term in equation 4:

$$\mathcal{L}_{\mathrm{aux}} = \mathbb{E}[\ell_{\mathrm{atom}} + \ell_{\mathrm{dist}} + \ell_{\mathrm{tors}}]\,. \tag{13}$$

Sampling settings. Independent standard Gaussian translation vectors are centered over valid nucleotides and scaled by 10 Å. Protein context retains up to 128 nearby residues, with unavailable frame orientations masked.

B Evaluation Metric Definitions

Let $\mathcal{G}$ and $\mathcal{Y}$ denote generated and reference ensembles. All structural comparisons use RNA $\mathrm{C4}'$ coordinates, with distances in Å.

Structural recovery and diversity. Let $\mathcal{A}$ be the residue pairs aligned by US-align, d_{ij} their distances after superposition, L_Y the reference length, and $d_0(L_Y)$ the nucleic-acid distance scale.

$$\mathrm{TM}(G, Y) = \frac{1}{L_Y} \sum_{(i,j)\in\mathcal{A}} \frac{1}{1 + \big(d_{ij}/d_0(L_Y)\big)^2}. \tag{14}$$

$$\mathrm{RMSD}_{\mathrm{US}}(G, Y) = \sqrt{\frac{1}{|\mathcal{A}|} \sum_{(i,j)\in\mathcal{A}} d_{ij}^2}. \tag{15}$$

$$\text{AMR-R} = \frac{1}{|\mathcal{Y}|} \sum_{Y\in\mathcal{Y}} \max_{G\in\mathcal{G}} \mathrm{TM}(G, Y). \tag{16}$$

$$\text{COV-R@}\tau = \frac{1}{|\mathcal{Y}|} \sum_{Y\in\mathcal{Y}} \mathbf{1}\left[\max_{G\in\mathcal{G}} \mathrm{TM}(G, Y) \geq \tau\right]. \tag{17}$$

We use $\tau \in \{0.5, 0.6, 0.7, 0.8\}$; RBPTrans uses references from the target binding condition. For the set $\mathcal{V}$ of successfully aligned unordered pairs within $\mathcal{G}$,

$$\text{Diversity} = \frac{1}{|\mathcal{V}|} \sum_{(g,h)\in\mathcal{V}} \mathrm{RMSD}_{\mathrm{US}}\big(G^{(g)}, G^{(h)}\big). \tag{18}$$

Singleton ensembles have zero diversity. AMR-R and COV-R lie in $[0, 1]$; higher values indicate better recovery. COV-R is also displayed as a percentage in Figure 4. Scores are averaged equally over RNA clusters, or directed pairs for RBPTrans. These recall metrics reward recovery of reference states; diversity alone does not indicate whether generated conformations resemble those states.

Steric compatibility. Let $\mathcal{G}'$ contain valid conformations registered to the input complex, $\tilde{\mathbf{g}}_i$ their matched RNA coordinates, and $\mathcal{P}$ the protein heavy-atom coordinates. With $\delta = 1$ Å,

$$\text{Clash}(\%) = \frac{100}{|\mathcal{G}'|} \sum_{G\in\mathcal{G}'} \mathbf{1}\left[\min_{i,\,\mathbf{p}\in\mathcal{P}} \|\tilde{\mathbf{g}}_i - \mathbf{p}\|_2 < \delta\right]. \tag{19}$$

Clash is averaged equally over protein-containing clusters, using the same protein geometry for all methods.

Aggregation and uncertainty. For RBP and RNA-only, we compute the arithmetic mean of per-cluster scores, giving each of the 41 and 17 clusters equal weight. The 95% confidence intervals in Figure 7 are percentile intervals from 20,000 bootstrap resamples of whole clusters (NumPy generator seed 20260921). All methods use the same resampled cluster indices within a dataset.

RBPTrans contains 34 directed pairs from 9 RNAs. Its point estimates give each directed pair equal weight. To preserve dependence between pairs from the same RNA, each bootstrap replicate samples nine RNAs with replacement, retains all directed pairs for each selected RNA, and recomputes the pair-weighted mean. The reported intervals use the 2.5th and 97.5th percentiles of 20,000 replicates (seed 20260917). RNA clusters with more directed pairs retain their larger contribution to the point estimate.

These intervals describe uncertainty across the evaluated RNAs or clusters; they do not measure variability across independently trained models. The per-cluster win counts and mean-difference annotations are descriptive comparisons and are not interpreted as hypothesis tests. In the paired recovery plots, the comparison value for each cluster is the highest AMR-R among AlphaFold3 + subMSA, RhoFold + subMSA, RNAbpFlow, and DynaRNA. Neither the protein ablation nor the input-structure control enters this baseline maximum.

Agreement with MD ensembles. Both ensembles contain $N = 1{,}000$ conformations with L corresponding nucleotides, aligned to the same simulation starting structure. Let $\tilde{\mathbf{x}}_i$ be aligned coordinates and $\boldsymbol{\mu}_i^{\mathcal{E}}$ their ensemble mean for $\mathcal{E} \in \{\mathcal{G}, \mathcal{Y}\}$. Per-nucleotide RMSF is

$$f_i^{\mathcal{E}} = \sqrt{\frac{1}{N} \sum_{X \in \mathcal{E}} \|\tilde{\mathbf{x}}_i - \boldsymbol{\mu}_i^{\mathcal{E}}\|_2^2}. \tag{20}$$

$$\text{RMSF error} = \sqrt{\frac{1}{L} \sum_{i=1}^{L} (f_i^{\mathcal{G}} - f_i^{\mathcal{Y}})^2}. \tag{21}$$

$$\text{MAE} = \frac{1}{L} \sum_{i=1}^{L} |f_i^{\mathcal{G}} - f_i^{\mathcal{Y}}|. \tag{22}$$

With $\bar{f}^{\mathcal{E}}$ denoting the mean RMSF over nucleotides,

$$r = \frac{\sum_{i=1}^{L} (f_i^{\mathcal{G}} - \bar{f}^{\mathcal{G}})(f_i^{\mathcal{Y}} - \bar{f}^{\mathcal{Y}})}{\sqrt{\sum_{i=1}^{L} (f_i^{\mathcal{G}} - \bar{f}^{\mathcal{G}})^2} \sqrt{\sum_{i=1}^{L} (f_i^{\mathcal{Y}} - \bar{f}^{\mathcal{Y}})^2}}. \tag{23}$$

Let $d(G, Y)$ denote coordinate RMSD in the common frame without further superposition, and $\mathfrak{S}_N$ the set of permutations of N samples.

$$\text{Ens. RMSD} = \frac{1}{2N} \left[\sum_{k=1}^{N} \min_{\ell} d(G^{(k)}, Y^{(\ell)}) + \sum_{\ell=1}^{N} \min_{k} d(G^{(k)}, Y^{(\ell)}) \right]. \tag{24}$$

$$W_2 = \left[\min_{\pi \in \mathfrak{S}_N} \frac{1}{N} \sum_{k=1}^{N} d(G^{(k)}, Y^{(\pi(k))})^2 \right]^{1/2}. \tag{25}$$

Let $\mathbf{q}(X)$ be the projection of flattened aligned coordinates onto the leading two MD principal axes, centered at the MD mean, without whitening or length scaling.

$$W_2^{\text{MD-PCA}} = \left[\min_{\pi \in \mathfrak{S}_N} \frac{1}{N} \sum_{k=1}^{N} \|\mathbf{q}(G^{(k)}) - \mathbf{q}(Y^{(\pi(k))})\|_2^2 \right]^{1/2}. \tag{26}$$

For leading unit principal axes $\mathbf{v}_1^{\mathcal{G}}$ and $\mathbf{v}_1^{\mathcal{Y}}$ fitted separately to each ensemble,

$$\text{PC-sim} = \left| (\mathbf{v}_1^{\mathcal{G}})^\top \mathbf{v}_1^{\mathcal{Y}} \right|. \tag{27}$$

MD scores are averaged over conditions within each RNA, then equally over RNAs. Repeating one structure gives zero RMSF and undefined r and PC-sim; the input-structure control is therefore marked N/A for these two metrics.

RMSF error is a root-mean-square error across nucleotides and is distinct from the MAE shown in profile plots. Correlation describes the spatial pattern of flexibility and can remain high when fluctuation amplitudes are inaccurate. Coordinate-space W_2 uses an RMSD cost, whereas MD-PCA W_2 uses unnormalized projected coordinates; the latter therefore has a different scale and depends on RNA length. PC-sim is insensitive to the arbitrary sign of a principal axis. These quantities compare conformational ensembles, without measuring transition rates, kinetic pathways, or the physical timescale of generated samples.

C Additional Experimental-State Recovery Results

C.1 RNA-only Ensemble Recovery

The RNA-only set contains 17 clusters with multiple experimental conformations observed without protein. We evaluate the same trained REFRACT model without protein input, using $N = 100$ samples per case.

Table 2 reports aggregate RNA-only recovery in the main text. Figure 7 provides uncertainty intervals and per-cluster comparisons for both RBP and RNA-only.

REFRACT exceeds the best external baseline in $38/41$ RBP and $17/17$ RNA-only clusters, with mean paired AMR-R gains of 0.230 and 0.174 (Figure 7). The two REFRACT variants coincide on RNA-only and are shown once.

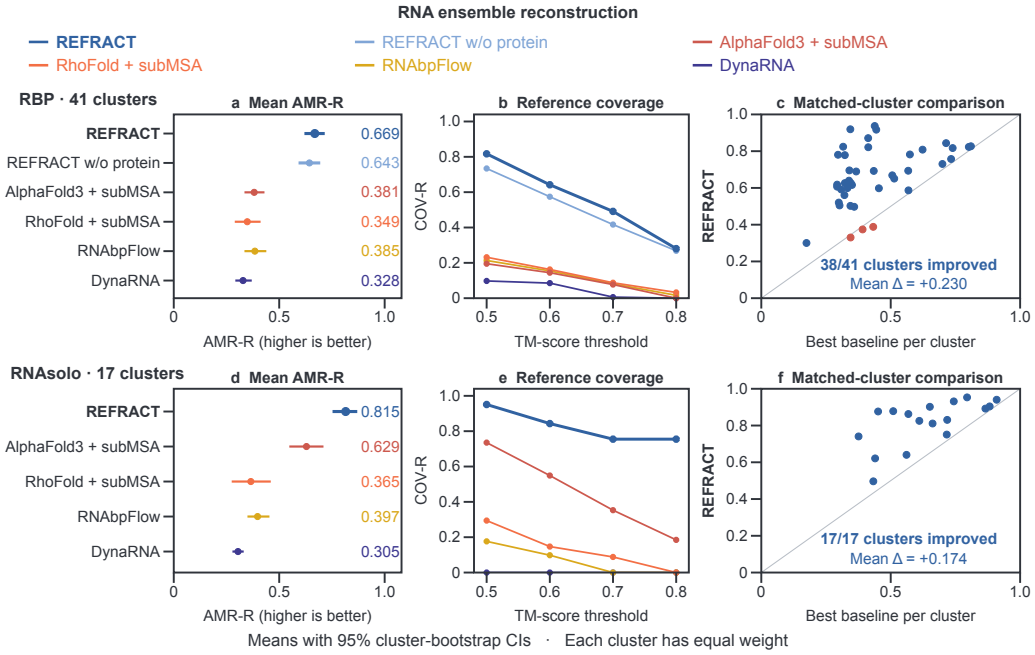


Figure 7: **Experimental-state recovery on RBP and RNA-only (RNAsolo).** (a, d) Mean AMR-R with 95% cluster-bootstrap intervals. (b, e) Mean COV-R at four thresholds. (c, f) Per-cluster AMR-R of REFRACT versus the best external baseline; the diagonal denotes equality. Rows contain 41 RBP and 17 RNA-only clusters, with equal weight per cluster.

D MD REFERENCE ENSEMBLES

The retained subset comprises 23 reference ensembles from 11 RNA systems (Table 4): free and protein-bound ensembles for nine systems from Sabei et al. (2024), one Cas12j-bound RNA ensemble from Baltrukevich & Bartos (2023), and four force-field-specific ensembles of 1R2P from Choi et al. (2024). We sample 1,000 RNA conformations from each reference ensemble and reserve all MD data exclusively for evaluation.

The retained Cas12j ensemble uses ff19SB/OL3. For 1R2P, the four retained conditions correspond to ff99$_{\mathrm{bsc0}}\chi$OL3, AMBER$_{\mathrm{DES}}$, ff99OL3_CMAP1, and AMBER$_{\mathrm{MaxEnt}}$. Each force-field condition is evaluated separately.

Table 4: **MD reference systems and retained conditions.** Each reference ensemble contributes 1,000 sampled RNA conformations.

RNA system	RNA type	Protein partner	Retained conditions	Ensembles
1ASY	tRNA(Asp)	Aspartyl-tRNA synthetase	Free / bound	2
1JBS	rRNA sarcin–ricin analog	Restrictocin	Free / bound	2
1MMS	23S rRNA fragment	Ribosomal protein L11	Free / bound	2
1OOA	RNA aptamer	NF-κB p50	Free / bound	2
1RKJ	Pre-rRNA fragment	Nucleolin	Free / bound	2
2R8S	P4–P6 ribozyme domain	Fab antibody fragment	Free / bound	2
2VPL	L1-operon mRNA fragment	Ribosomal L1 (domain I)	Free / bound	2
2ZM5	tRNA(Phe)	MiaA	Free / bound	2
3IEV	16S rRNA fragment	Era GTPase	Free / bound	2
Cas12j (7M5O)	CRISPR RNA	CasΦ-2 (Cas12j)	Bound	1
1R2P	Group II intron domain 5	None	Four force fields	4
			Total reference ensembles	23

Alignment and interpretation. Each reference ensemble is compared with 1,000 generated conformations using corresponding RNA $\mathrm{C4}'$ atoms. Structure-initialized methods use the initial RNA structure of the corresponding MD simulation. After alignment to that structure, the same coordinate frame is retained for RMSF, ensemble RMSD, and distributional comparisons (Appendix B). Conditions are first averaged within an RNA, followed by equal weighting of RNA systems. The four 1R2P force fields therefore contribute through one RNA-level average to the overall benchmark.

These deposited trajectories provide reference conformational distributions under their respective simulation conditions. Agreement with them does not establish force-field-independent experimental accuracy or reproduce dynamical timescales. The free/bound analysis and the residue-level profile subset below address complementary aspects of agreement and do not redefine the 23-ensemble benchmark.

E Additional MD Ensemble Comparisons

Table 5 reports MD ensemble agreement separately for free and protein-bound conditions, including the bound-condition REFRACT w/o protein variant.

Table 5: **Agreement with MD reference ensembles under free and bound conditions.** Best values within each condition are **bold**, with gray highlights for our methods; N/A denotes metrics not applicable to Input structure.

Condition	Method	RMSF error (Å) ↓	RMSF r ↑	Ens. RMSD (Å) ↓	W_2 (Å) ↓	MD-PCA W_2 ↓	PC-sim ↑
Free	Input structure	2.74	N/A	2.98	4.59	23.85	N/A
	Local perturb.	2.14	−0.267	4.58	5.83	22.17	0.084
	RNABPFLOW	2.93	0.628	3.45	6.02	23.70	0.450
	DYNARNA	6.30	0.297	7.67	11.45	**20.33**	0.256
	REFRACT	**1.75**	**0.633**	**2.80**	**4.36**	21.77	**0.536**
Bound	Input structure	1.87	N/A	2.21	3.26	15.45	N/A
	Local perturb.	2.30	−0.265	4.05	4.84	13.87	0.067
	RNABPFLOW	2.98	0.385	4.44	6.51	17.65	0.211
	DYNARNA	6.93	0.230	7.98	10.75	17.16	0.279
	REFRACT w/o protein	**0.79**	0.644	2.12	2.93	11.37	0.515
	REFRACT	0.81	**0.688**	**2.06**	**2.85**	**11.36**	**0.531**

REFRACT leads on five free-condition metrics; DynaRNA has the lowest MD-PCA W_2. In bound conditions, protein input improves five metrics relative to the ablation, while RMSF error is slightly lower without protein (0.79 versus 0.81 Å). Thus, gains in distributional agreement need not coincide with gains in fluctuation amplitude.

Figure 8 shows the eight lowest-MAE cases among 20 available profiles, an illustrative selection of close agreement. It is not a random sample. Table 6 reports all available profiles.

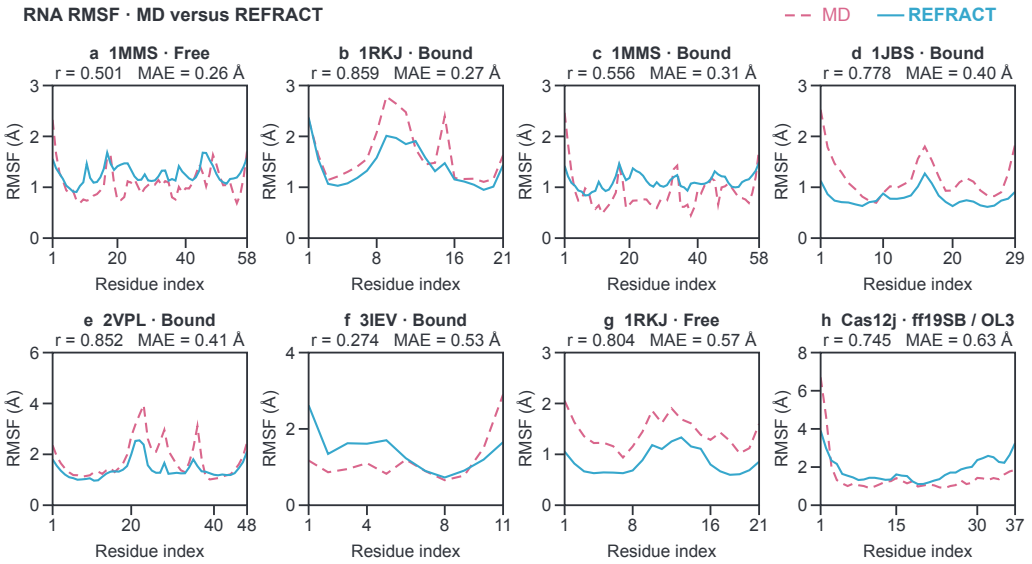


Figure 8: **Eight MD–REFRACT RMSF profiles with the lowest MAE.** Pink dashed curves show MD and cyan solid curves show REFRACT. Cases are ordered by increasing MAE among the 20 available profiles. Each panel reports residue-wise Pearson r and MAE (Å). Axis ranges vary across panels; curves connect the supplied values without smoothing.

Complete residue-level RMSF results. The 20 available profiles cover the 18 free/bound conditions from Sabei et al. (2024), Cas12j ff19SB/OL3, and 1R2P OL3. The other three 1R2P force fields in the 23-ensemble benchmark are absent from this profile table.

Table 6: **All available MD–REFRACT RMSF profiles.** Metrics are recomputed across nucleotides. RMSF error denotes RMSE (equation 21); panels refer to Figure 8.

RNA	Condition	Nucleotides	Pearson r	MAE (Å)	RMSF error (Å)	Panel
1ASY	Bound	73	0.688	0.977	1.180	—
1ASY	Free	73	0.804	0.815	0.898	—
1JBS	Bound	29	0.778	0.397	0.494	d
1JBS	Free	29	0.362	1.322	1.452	—
1MMS	Bound	58	0.556	0.311	0.368	c
1MMS	Free	58	0.501	0.256	0.322	a
1OOA	Bound	27	0.689	1.545	1.678	—
1OOA	Free	27	0.944	1.621	1.723	—
1RKJ	Bound	21	0.859	0.272	0.373	b
1RKJ	Free	21	0.804	0.569	0.595	g
2R8S	Bound	159	0.634	0.735	1.122	—
2R8S	Free	157	0.158	0.684	0.913	—
2VPL	Bound	48	0.852	0.411	0.575	e
2VPL	Free	48	0.319	4.613	4.920	—
2ZM5	Bound	72	0.865	0.711	0.752	—
2ZM5	Free	72	0.880	1.202	1.271	—
3IEV	Bound	11	0.274	0.526	0.706	f
3IEV	Free	11	0.920	3.391	3.655	—
1R2P	OL3	34	0.722	2.002	2.082	—
Cas12j	ff19SB/OL3	37	0.745	0.632	0.826	h

High correlation can coexist with amplitude error: free 1OOA has $r = 0.944$ and MAE 1.621 Å. Free 2VPL has MAE 4.613 Å and $r = 0.319$. Per-residue means do not provide replicate-level uncertainty.

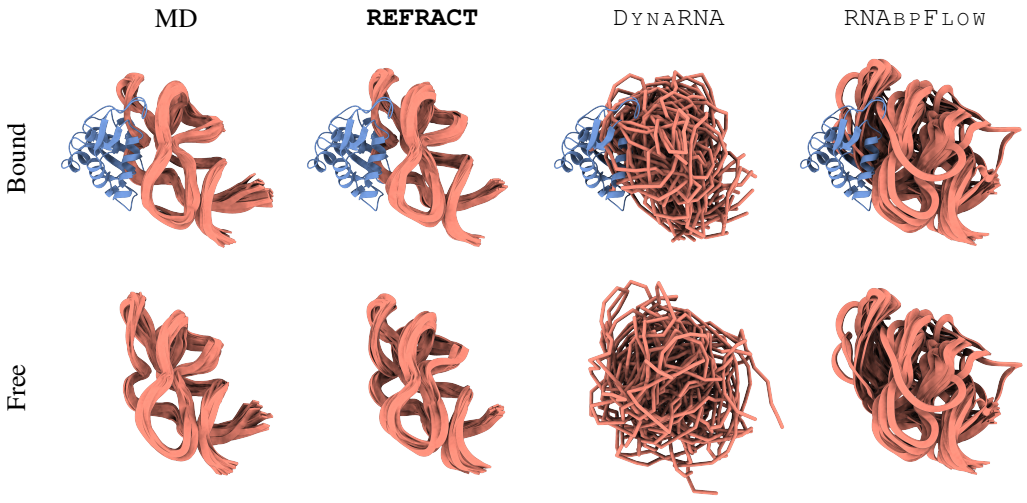


Figure 9: **L11-binding rRNA ensembles.** Top: L11-bound 1MMS, RNA chain C. Bottom: protein-free 3I9C, chain A. Columns show MD, REFRACT, DynaRNA, and RNAbpFlow. RNA is salmon; fixed L11 is blue. Cartoons show full backbones where available; DynaRNA uses C4′ traces. This complements the 2R8S comparison in Figure 6.