LA-RIBO: RNA CO-DESIGN VIA GEOMETRY–LATENT FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

RNA function arises from the coupling of nucleotide sequence and three-dimensional structure, motivating their joint design. Coordinating global folding with nucleotide-level detail remains challenging under limited structural supervision. We introduce La-Ribo, a generative framework for RNA sequence–structure co-design via geometry–latent flow matching. La-Ribo retains a sparse phosphate–sugar–base scaffold and encodes nucleotide identity and local conformation in residue-wise latents. A shared flow network generates both jointly, and an RNA-specific decoder then reconstructs all heavy atoms. To expand supervision, we construct a quality-controlled corpus of 168,561 RNA structures, integrating experimental data with predictions from three folding models, including 10,631 MSA-supported structures generated in this work. La-Ribo improves designability and codesignability over the evaluated baselines across sampling budgets and two refolding models, and the same prior supports scaffold-conditioned inverse folding without additional training. Our code is available at https://anonymous.4open.science/r/la-ribo-7DDA.

1 INTRODUCTION

Ribonucleic acid (RNA) performs diverse functions, including catalysis, gene regulation, and molecular recognition, and these functions depend jointly on nucleotide sequence and three-dimensional structure (Wilson & Lilley, 2021; Cao et al., 2024; Childs-Disney et al., 2022). Computational design has traditionally approached these modalities separately. Sequence generators propose nucleotide sequences for downstream structural assessment, while backbone generators typically obtain sequences through inverse folding and assess compatibility by refolding (Anand et al., 2025). Recent co-design methods instead generate sequence and structure together, enabling direct exploration of RNA sequence–structure space (Rubin et al., 2025; Ma et al., 2025; Li et al., 2026b). Coordinating nucleotide identity with global folding nevertheless remains difficult, particularly under the limited structural supervision available for RNA.

RNA co-design faces two challenges: representation and structural supervision. A useful representation must preserve long-range tertiary organization while capturing nucleotide-specific local conformations. Partially latent protein models offer a complementary strategy, retaining coarse geometry explicitly while absorbing sequence and local detail into continuous per-residue latents (Geffner et al., 2026). Transferring this strategy to RNA is not direct. In proteins, the backbone anchors that define a residue frame are themselves a rigid unit; in RNA, anchors that span the phosphate, sugar, and base groups are not, so informative geometric anchors cannot serve directly as the frame for atom construction. Structural supervision is the second constraint. Experimental RNA structures number only in the thousands, far fewer than those available for proteins, so recent work supplements them with predicted structures. Predictions from different folding models differ in confidence, coverage, and length distribution, however, and they can transfer the biases of the folding models used to produce them, which makes consistent quality control necessary.

We introduce La-Ribo, a generative framework for RNA sequence–structure co-design via geometry–latent flow matching. La-Ribo retains P, $C4'$, and glycosidic N as sparse geometric anchors, while a scaffold-conditioned autoencoder learns residue-wise latents encoding nucleotide identity and local conformation. Nucleotide identity is absorbed into these continuous latents, so generation proceeds entirely in a continuous space and requires no discrete–continuous hybrid flow. The

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.

decoder predicts a residual rotation from the scaffold-defined frame to a canonical sugar frame and reconstructs heavy atoms using nucleotide templates and torsions; this learned interface separates the choice of informative anchors from the rigid-frame requirements of atom construction. A shared geometry–latent flow subsequently learns the joint distribution of scaffold coordinates and latents, from which the frozen decoder recovers sequences and structures. With geometry kept explicit, fixing the observed scaffold and integrating only the latent modality turns the same trained model into a scaffold-conditioned inverse-folding model, without conditional training.

To expand structural supervision, we construct a quality-controlled corpus of 168,561 RNA structures, combining experimental entries with predictions from OpenFold3, NuFold, and trRosettaRNA2 (Adamczyk et al., 2022; Shen et al., 2024; Ahdritz et al., 2024; Kagaya et al., 2025; Wang et al., 2026b). Beyond curating existing resources, we generate 10,631 trRosettaRNA2 structures with multiple sequence alignments, confidence filtering, and structural relaxation. As predicted structures vary in reliability across sources, we apply source-specific confidence screening and common eligibility criteria before unified preprocessing.

Our main contributions are as follows. (i) **Representation:** We introduce an RNA-specific geometry–latent representation combining sparse P, $C4'$, and glycosidic N anchors with residue-wise latents, linked by a learned scaffold-to-sugar frame transformation. (ii) **Generation:** A shared flow network generates geometry and latents jointly, and with geometry kept explicit, the same trained prior serves as a scaffold-conditioned inverse-folding model without conditional training. (iii) **Dataset:** We construct a quality-controlled corpus of 168,561 RNA structures integrating experimental and predicted sources under source-specific screening, including 10,631 MSA-supported and relaxed trRosettaRNA2 structures generated in this work. (iv) **Results:** La-Ribo improves designability and codesignability over the evaluated baselines, across sampling budgets and refolding models.

2 Related work

2.1 RNA generation, inverse folding, and co-design

Computational RNA design spans sequence generation, inverse folding, structure generation, and joint sequence–structure generation. Sequence-centric models learn distributions over RNA sequences with language modeling or diffusion, extending *de novo* generation toward function-, family-, and structure-conditioned design (Zhang et al., 2024; Patel et al., 2025; Wang et al., 2026a). Structure-centric methods instead generate RNA geometry: RNA-FrameFlow applies flow matching to unconditional backbone generation, with sequences recovered through inverse folding (Anand et al., 2025), while RNAbpFlow generates sequence- and base-pair-conditioned all-atom ensembles for conformational sampling rather than *de novo* design (Tarafder & Bhattacharya, 2026). Inverse-folding models generate sequences for a target structure (Joshi et al., 2025; Kubaney et al., 2025; Huang et al., 2024).

Co-design methods generate both modalities within one generative process. RNAFlow integrates inverse folding with a pre-trained structure model for protein-conditioned design (Nori & Jin, 2024), and RiboFlow pretrains an unconditional co-design model before fine-tuning it for target-molecule conditioning (Ma et al., 2025). Among unconditional methods, RiboGen combines discrete and continuous flows over nucleotide identities and all-atom geometry (Rubin et al., 2025), Joint Diffusion couples sequence and structure through a pre-trained RoseTTAFold2NA denoiser (Li et al., 2026b), and DuetRNA introduces a base–sugar dual-frame representation (Li et al., 2026a). La-Ribo instead pairs a sparse geometric state with residue-wise continuous latents. As nucleotide identity is carried by the latents rather than modeled as a discrete variable, generation proceeds as a single continuous flow rather than a discrete–continuous hybrid, and the same prior is reusable for scaffold-conditioned inverse folding without separate conditional training.

2.2 Explicit and latent biomolecular representations

Representations for biomolecular generation must capture global structure and local chemical detail. Geometric models commonly represent structures using Cartesian coordinates, rigid frames, torsions, or combinations thereof, while multimodal approaches couple these geometric states with sequence information (Yim et al., 2023; Bose et al., 2024; Campbell et al., 2024; Huguet et al., 2024). Latent generative models provide a complementary strategy by compressing high-dimensional observations before applying diffusion or flow matching (Kingma & Welling, 2014; Rombach et al., 2022; Esser

et al., 2024). Recent biomolecular models increasingly incorporate learned representations into geometric modeling. La-Proteina introduces a partially latent representation for atomistic protein generation, retaining coarse coordinates explicitly while absorbing sequence and local atomic detail into continuous per-residue latents (Geffner et al., 2026). This strategy does not transfer directly to RNA, where informative geometric anchors span the phosphate, sugar, and base groups and therefore do not form the rigid unit that atom construction requires. For RNA, RiboSphere learns discrete structural tokens through geometric quantization and transfers the learned representation to reconstruction and downstream tasks such as inverse folding (Zhang et al., 2026). La-Ribo instead pairs sparse RNA geometry with continuous residue-wise latents, connected by a learned scaffold-to-sugar frame transformation that resolves this mismatch.

3 METHOD

La-Ribo first learns an RNA-specific geometry–latent representation through a geometry-aware autoencoder, and then models the resulting scaffold and latent states through geometry–latent flow matching. Figure 1 summarizes the representation, generative process, and heavy-atom reconstruction.

3.1 GEOMETRY–LATENT REPRESENTATION

We consider a canonical RNA chain of length N, with sequence $S = (s_1, \dots, s_N)$, where $s_i \in \mathcal{A} = \{A, C, G, U\}$. Following the commonly used atom37 convention in biomolecular modeling (Jumper et al., 2021; Geffner et al., 2026), we represent its structure by heavy-atom coordinates $X \in \mathbb{R}^{N\times 37\times 3}$. Let $x_i^a \in \mathbb{R}^3$ denote the coordinate of atom a in residue i. Rather than directly generating the complete atomistic state, we retain a sparse geometric scaffold

$$b_i = \left[x_i^{\mathrm{P}}; x_i^{\mathrm{C4'}}; x_i^{\nu_i}\right] \in \mathbb{R}^{d_B}, \qquad d_B = 9, \tag{1}$$

where $\nu_i = \mathrm{N9}$ for $s_i \in \{A, G\}$ and $\nu_i = \mathrm{N1}$ for $s_i \in \{C, U\}$. Thus, each scaffold state contains the coordinates of P, C4′, and the glycosidic nitrogen. The scaffold $B = (b_1, \dots, b_N)$ captures phosphate, ribose, and base-attachment geometry. The P–C4′ trace provides a reduced description of RNA backbone conformation through standard pseudotorsional geometry (Wadley et al., 2007), while the glycosidic nitrogen provides an explicit anchor for the base attachment. Appendix C.2 compares one-, two-, and three-anchor scaffolds and latent widths, supporting this choice.

Nucleotide identity and the local atomic detail not determined by the scaffold, including non-anchor heavy-atom positions, ribose conformation, and base orientation, are represented by residue-wise latent variables $z_i \in \mathbb{R}^{d_Z}$, with $Z = (z_1, \dots, z_N)$ and $d_Z = 16$ unless stated otherwise. La-Ribo therefore represents each residue by a $(d_B + d_Z)$-dimensional geometry–latent state, comprising 9 explicit scaffold coordinates and 16 learned latent dimensions.

The latent representation is learned by an encoder E_ϕ, and the sequence and atomistic structure are reconstructed by a structured decoder D_ψ, where ϕ and ψ denote their learnable parameters. Section 3.2 describes this autoencoding process, while Section 3.3 describes learning generative dynamics over the joint state (B, Z).

3.2 SCAFFOLD-CONDITIONED AUTOENCODING

Latent encoding. We learn the residue-wise latent representation using an encoder E_ϕ. Given the sequence S, structure X, and scaffold B, the encoder predicts

$$(\mu, \sigma) = E_\phi(S, X, B), \qquad z_i = \mu_i + \sigma_i \odot \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, I), \tag{2}$$

where $\mu, \sigma \in \mathbb{R}^{N\times d_Z}$ are the latent means and standard deviations. Although $Z = (z_1, \dots, z_N)$ is residue-wise, the context-aware encoder allows each latent to depend on the full RNA chain.

Frame-aware decoding. The geometric scaffold and atom builder have different representational requirements. The P/C4′/N scaffold captures backbone and base-attachment geometry, but its three anchors belong to different chemical groups rather than a single rigid unit. Atomistic reconstruction instead uses a canonical frame within the ribose, constructed from C4′/O4′/C3′ following RNA-FrameFlow (Anand et al., 2025). These two requirements select different atoms: O4′ and C3′ are

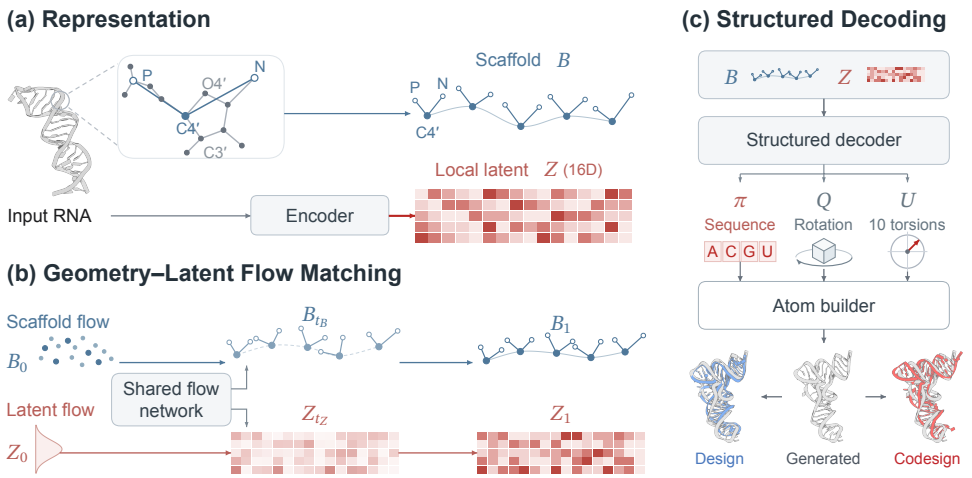


Figure 1: Overview of La-Ribo. (a) Each residue is represented by an explicit P/C4'/glycosidic-N scaffold (blue) and an encoded 16-dimensional latent (red). (b) A shared flow network jointly generates states with modality-specific schedules. (c) The frozen decoder predicts nucleotide identity, a scaffold-to-sugar rotation, and torsions, from which the atom builder builds all heavy atoms.

confined to the ribose ring and carry little information about backbone trace or base placement, whereas P and the glycosidic nitrogen do. We therefore keep the informative anchors in the generated state and learn the transformation between the two frames. Writing $\mathcal{F}(o, u, v) \in \mathrm{SO}(3)$ for the orthonormal local frame centered at o, we define

$$R_i^p = \mathcal{F}(x_i^{\mathrm{C4'}}, x_i^{\nu_i}, x_i^{\mathrm{P}}), \qquad R_i^c = \mathcal{F}(x_i^{\mathrm{C4'}}, x_i^{\mathrm{O4'}}, x_i^{\mathrm{C3'}}), \tag{3}$$

where R_i^p is the scaffold pseudo-frame and R_i^c is the canonical sugar frame. Given (B, Z), the structured decoder predicts

$$(\pi_i, Q_i, U_i) = D_\psi(B, Z)_i, \qquad R_i^{\mathrm{dec}} = R_i^p Q_i. \tag{4}$$

Here π_i is a probability distribution over $\mathcal{A} = \{A, C, G, U\}$, $Q_i \in \mathrm{SO}(3)$ is the predicted rotation from the scaffold pseudo-frame to the sugar frame, and $U_i = \{u_{ik}\}_{k=1}^{10}$ contains 10 normalized sine–cosine torsion representations. The decoded frame R_i^{dec}, torsions U_i, and nucleotide identity specified by π_i together parameterize the structured atom builder for heavy-atom reconstruction. The construction of $\mathcal{F}$ and the torsion and atom-construction conventions are provided in Appendix B.2.

Training objective. The autoencoder jointly supervises nucleotide identity, the scaffold-to-sugar frame transformation, torsional geometry, and atomistic reconstruction. We write the geometric reconstruction objective under template view v as

$$\mathcal{L}_{\mathrm{geom}}^v = \mathcal{L}_{\mathrm{atom}}^v + \mathcal{L}_{\mathrm{scaf}}^v + \mathcal{L}_{\mathrm{pseudo}}^v, \qquad v \in \{\mathrm{gt}, \mathrm{pred}\}, \tag{5}$$

and optimize

$$\mathcal{L}_{\mathrm{AE}} = \mathcal{L}_{\mathrm{seq}} + \mathcal{L}_{\mathrm{rot}} + \mathcal{L}_{\mathrm{tors}} + \lambda_{\mathrm{gt}} \mathcal{L}_{\mathrm{geom}}^{\mathrm{gt}} + \lambda_{\mathrm{pred}} \mathcal{L}_{\mathrm{geom}}^{\mathrm{pred}} + \lambda_{\mathrm{KL}} \mathcal{L}_{\mathrm{KL}}. \tag{6}$$

The sequence, rotation, and torsion terms supervise nucleotide identity, sugar-frame orientation, and local conformation; the geometric terms compare heavy atoms, scaffold coordinates, and backbone pseudotorsions. The ground-truth template provides identity-consistent supervision, while the predicted-template term evaluates reconstruction with the inference-time template-selection rule. The KL term regularizes the latent distribution. Exact losses, weights, and masks are in Appendix B.2.

3.3 Geometry–Latent Flow Matching

With the encoder E_ϕ and decoder D_ψ fixed, we learn generative dynamics over (B, Z) through flow matching (Lipman et al., 2023; Liu et al., 2023). Scaffold coordinates are centered at the valid C4′ centroid and globally scaled, while posterior-mean latents are standardized dimension-wise, yielding

$$x_1^B \in \mathbb{R}^{N \times d_B}, \qquad x_1^Z \in \mathbb{R}^{N \times d_Z}. \tag{7}$$

Let $a \in \{B, Z\}$ index the scaffold and latent modalities. For each modality, we pair the clean state x_1^a with an independently sampled Gaussian base state x_0^a, centered in the same way for the scaffold modality, and define the linear path

$$x_{t_a}^a = (1 - t_a)x_0^a + t_a x_1^a, \qquad a \in \{B, Z\}. \tag{8}$$

The scaffold and latent times t_B and t_Z are sampled independently, allowing the two modalities to occupy different noise levels during training.

A shared velocity network, parameterized by θ, predicts both components $v_\theta = (v_\theta^B, v_\theta^Z)$ from $x_t = (x_{t_B}^B, x_{t_Z}^Z)$. Here v_θ^B includes scaffold recentering. Throughout the main text we write $v_\theta(x_t, t_B, t_Z)$, omitting the self-conditioning input x_{sc} from the argument list; Appendix B.3 gives the full signature. We optimize

$$\mathcal{L}_{\text{FM}} = \sum_{a \in \{B,Z\}} \mathbb{E}\left[\frac{1}{d_a|\mathcal{R}|} \sum_{i \in \mathcal{R}} \|v_\theta^a(x_t, t_B, t_Z)_i - (x_1^a - x_0^a)_i\|_2^2\right], \tag{9}$$

where $\mathcal{R}$ contains valid residues. Training uses only geometry–latent targets, without decoded-coordinate or sequence losses. The decoder is frozen at this stage, so the flow is fit directly in the representation space rather than through a reconstruction pathway, which keeps the two stages independent and avoids backpropagating through the non-differentiable template selection of Section 3.2. Exact normalization, scaffold recentering, time sampling, self-conditioning, and endpoint stabilization are described in Appendix B.3.

3.4 Sampling

For unconditional generation, scaffold and latent states are initialized from their base distributions and integrated jointly with forward Euler using modality-specific schedules,

$$t_k^a = \left(\frac{k}{K}\right)^{\rho_a}, \qquad a \in \{B, Z\}, \qquad k = 0, \ldots, K, \tag{10}$$

where K is the number of Euler steps and ρ_B, ρ_Z control the scaffold and latent schedules, respectively. After reversing normalization, the terminal state (B, Z) is decoded by D_ψ into a nucleotide sequence and full heavy-atom structure. For scaffold-conditioned inverse folding, the observed scaffold is fixed at $t_B = 1$ and only the latent modality is integrated. The resulting latent state is decoded with the supplied scaffold. Training times are restricted to $[0.01, 0.99]$ (Appendix B.3), so $t_B = 1$ slightly extrapolates beyond the training range; the procedure is nevertheless approximate, as holding one modality fixed does not exactly recover the conditional of the learned joint distribution.

4 Experiments

4.1 Experimental setup

Dataset. We construct a corpus of 168,561 RNA structures from RNASolo, experimental PDB chains included in the RhoFold release, predicted structures from OpenFold3 and NuFold, and a trRosettaRNA2 set generated in this work (Adamczyk et al., 2022; Shen et al., 2024; Ahdritz et al., 2024; Kagaya et al., 2025; Wang et al., 2026b). Entries span 16–256 nt with at least 90% valid-frame coverage. Sequences are globally clustered with MMseqs2 at 90% identity and 95% coverage, and whole clusters are assigned approximately 95/5 to training and validation, yielding 159,955 and 8,606 structures, respectively. All learned-representation analyses and the inverse-folding benchmark (Section 4.5) use only the cluster-held-out validation set, whose sequence clusters are disjoint from all training data. Appendix A provides the complete data-construction workflow.

Baselines and model variants. We compare against RNA-FrameFlow, evaluated with its released weights (Anand et al., 2025), and RiboFlow (Ma et al., 2025), whose unconditional co-design stage we retrain in two regimes: RiboFlow-Mix on the full La-Ribo training partition, and RiboFlow-Exp on its experimental-only subset, matching the data regime of RNA-FrameFlow. RiboGen (Rubin et al., 2025), Joint Diffusion (Li et al., 2026b), and DuetRNA (Li et al., 2026a) had not released code at the time of evaluation, preventing matched reruns under our protocol. Our two main variants share the same pretrained scaffold-conditioned autoencoder: La-Ribo-Base keeps the pair representation fixed throughout the flow network, whereas La-Ribo-Tri refines it every two blocks using token-to-pair projections and triangle multiplicative updates. Appendix B provides the training details.

Table 1: Unconditional RNA generation at 50–400 steps. Mean ± standard deviation over three 600-RNA campaigns; bold marks the best mean at each step count.

Method	Steps	Designability			Codesignability		
		both@4.0 ↑	TM@0.45 ↑	#Clust.↑	both@4.0 ↑	TM@0.45 ↑	#Clust.↑
RNA-FrameFlow	50	25.17 ± 1.09	36.61 ± 0.19	24.00 ± 1.73	–	–	–
	100	26.61 ± 1.18	39.67 ± 0.33	27.00 ± 2.65	–	–	–
	200	26.17 ± 1.76	38.44 ± 1.84	26.00 ± 1.00	–	–	–
	400	26.89 ± 1.69	38.89 ± 1.69	27.00 ± 3.61	–	–	–
RiboFlow-Exp	50	22.28 ± 0.63	30.44 ± 0.54	19.33 ± 4.04	19.22 ± 0.35	24.94 ± 0.10	15.33 ± 2.52
	100	22.22 ± 0.54	31.22 ± 0.98	18.67 ± 1.53	19.28 ± 0.42	25.78 ± 1.25	15.67 ± 2.08
	200	20.50 ± 0.73	30.06 ± 0.69	18.00 ± 2.65	19.33 ± 1.61	26.44 ± 2.55	15.67 ± 1.15
	400	21.39 ± 0.54	29.94 ± 0.77	17.00 ± 2.00	20.06 ± 0.19	27.61 ± 0.69	15.67 ± 1.53
RiboFlow-Mix	50	29.00 ± 0.17	37.28 ± 1.09	23.00 ± 3.46	26.61 ± 1.34	31.83 ± 0.58	19.00 ± 2.65
	100	29.06 ± 0.54	38.44 ± 0.70	22.33 ± 2.31	26.67 ± 0.67	32.50 ± 0.71	19.33 ± 2.08
	200	27.61 ± 1.07	36.89 ± 0.92	21.67 ± 1.53	26.00 ± 1.38	33.06 ± 1.13	19.00 ± 1.73
	400	28.39 ± 0.10	37.39 ± 0.91	20.67 ± 2.31	26.61 ± 1.57	34.72 ± 1.34	19.33 ± 1.53
La-Ribo-Base	50	36.11 ± 2.07	50.78 ± 2.00	22.67 ± 1.53	**30.00** ± 2.77	**52.78** ± 2.18	24.00 ± 1.00
	100	37.11 ± 0.98	51.83 ± 1.20	22.67 ± 1.15	**32.11** ± 2.14	55.28 ± 1.55	23.33 ± 3.06
	200	38.22 ± 1.27	53.00 ± 0.67	23.00 ± 1.73	**32.56** ± 2.25	56.72 ± 1.78	25.33 ± 2.31
	400	37.72 ± 1.13	53.11 ± 1.55	24.33 ± 1.15	**32.56** ± 1.58	57.17 ± 0.29	25.67 ± 3.51
La-Ribo-Tri	50	**37.11** ± 0.25	**54.06** ± 0.35	**29.00** ± 2.65	28.61 ± 1.13	50.83 ± 2.09	**28.00** ± 1.00
	100	**37.78** ± 1.92	**56.89** ± 2.22	**28.33** ± 5.51	30.78 ± 1.26	**56.11** ± 1.95	**25.33** ± 3.21
	200	**39.11** ± 2.12	**58.83** ± 1.48	**27.67** ± 1.15	32.06 ± 0.75	**57.89** ± 1.71	**27.33** ± 0.58
	400	**38.61** ± 2.46	**59.33** ± 0.93	**29.33** ± 3.79	32.33 ± 0.58	**58.11** ± 0.84	**28.00** ± 1.00

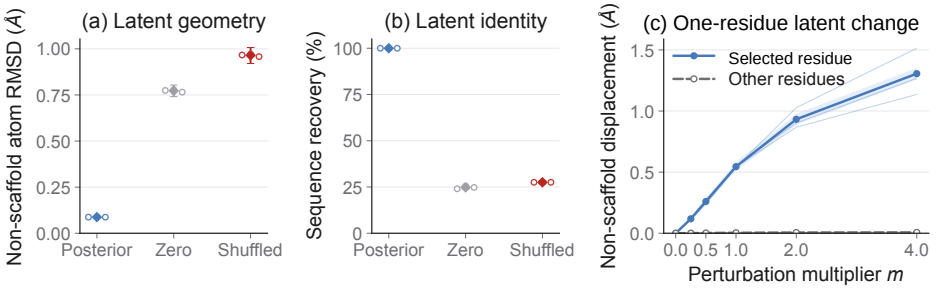


Figure 2: Latent interventions with a fixed scaffold. (a,b) compare posterior-mean, zero, and shuffled latents; (c) shows atom displacement from a separate one-residue latent perturbation.

4.2 Representation and Latent Analysis

Setup. To test what Z contributes beyond the scaffold, we study two scaffold-conditioned autoencoders: VAE, trained on the training partition, and E-VAE, the same architecture trained only on experimental structures from the training partition. Both are evaluated on the same cluster-held-out validation set. With scaffold B fixed, we decode three latent conditions: posterior ($z_i = \mu_i$), zero ($z_i = 0$), and shuffle (a within-chain permutation of the μ_i). A separate locality test changes one residue's latent while keeping B and all other latents fixed. **Evaluation metrics.** Sequence recovery uses the decoder's nucleotide probabilities π_i. We report canonical sugar-frame angular error and non-scaffold heavy-atom RMSD; the latter excludes the scaffold anchors x_i^{P}, $x_i^{\mathrm{C4'}}$, and $x_i^{\nu_i}$. Table 10 also reports reconstructed P and glycosidic-N errors. Locality compares atom displacement at the edited residue with all others as the perturbation magnitude m grows. **Results.** Posterior latents reconstruct almost exactly: 100% sequence recovery, sugar-frame error of 1.78°/2.67°, and non-scaffold RMSD of 0.087/0.122 Å for VAE and E-VAE, whose closeness indicates the corpus adds little at this stage. Zeroing the latents drops recovery to 24.77%, chance level for four nucleotides, so the scaffold alone carries no identity information; RMSD rises to 0.773 Å. Shuffling is worse still

Table 2: Autoencoder corpus × flow architecture at 200 steps. VAE is trained on the full corpus, E-VAE on its experimental subset; all four flows use experimental structures only. Mean ± standard deviation over three 600-RNA campaigns; RNA-FrameFlow and RiboFlow-Exp are references.

Autoencoder	Flow	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑
RNA-FrameFlow		26.17 ± 1.76	31.56 ± 2.10	38.44 ± 1.84	11.47 ± 0.29	0.413 ± 0.008
RiboFlow-Exp		20.50 ± 0.73	25.11 ± 0.42	30.06 ± 0.69	12.73 ± 0.22	0.408 ± 0.003
VAE	Base	25.78 ± 0.92	29.83 ± 2.18	33.67 ± 1.88	13.51 ± 0.38	0.374 ± 0.004
VAE	Tri	**30.72** ± 1.39	**36.17** ± 0.76	43.44 ± 0.54	11.74 ± 0.17	0.433 ± 0.004
E-VAE	Base	25.50 ± 1.20	29.67 ± 1.33	34.89 ± 1.13	12.66 ± 0.14	0.400 ± 0.002
E-VAE	Tri	30.17 ± 1.36	35.22 ± 1.11	**44.06** ± 1.49	**10.82** ± 0.15	**0.444** ± 0.006

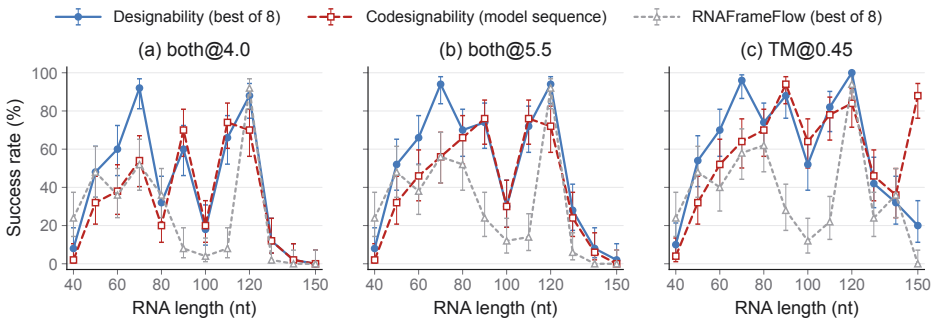


Figure 3: Length-resolved La-Ribo-Tri success at 200 steps for one 600-RNA campaign (50 RNAs per length). Error bars show 95% Wilson intervals.

(0.965 Å), showing the decoder acts on the supplied latents rather than an average conformation. Single-residue perturbations displace the edited residue's atoms far more than all others (Figure 2), so the latents are residue-local. One- and two-anchor scaffolds reach 0.159 and 0.130 Å against 0.122 Å for three anchors, at the same latent width and training regime (Appendix C.2). Appendix C.1 gives the complete metrics.

4.3 Unconditional RNA generation

Setup. Following RNA-FrameFlow and RiboFlow (Anand et al., 2025; Ma et al., 2025), each campaign generates 600 RNAs, 50 at each length from 40 to 150 nt in 10-nt increments. We compare RNA-FrameFlow, RiboFlow-Exp and RiboFlow-Mix, La-Ribo-Base, and La-Ribo-Tri at 50, 100, 200, and 400 steps, with three independently seeded campaigns per setting to estimate sampling variation. The La-Ribo variants use fixed modality schedules. Figure 4 compares local-geometry distributions from 200-step La-Ribo-Tri campaigns with experimental RNA structures.

Evaluation metrics. Designability refolds eight gRNAde sequences per generated structure and succeeds if any candidate passes; codesignability refolds the emitted sequence. Main-text scores use RhoFold; Appendix G.2 repeats the evaluation on the same structures and candidate sequences with OpenDDE. We align each refold with its generated structure using C4′ atoms to compute self-consistency RMSD (scRMSD) and TM-score (scTM). both@4.0 requires the same refold to have scRMSD$\leq$ 4.0 Å and scTM$\geq$ 0.45; both@5.5 changes the RMSD cutoff to 5.5 Å, and TM@0.45 requires scTM$\geq$ 0.45. #Clust. counts qTMclust (Zhang et al., 2022) clusters among generated backbones whose refolds meet TM@0.45.

Results. La-Ribo-Tri leads strict designability (both@4.0) at all four step counts (Table 1; Table 12 reports expanded metrics). At 200 steps it reaches 39.11 ± 2.12%, over eleven points above the strongest baseline (RiboFlow-Mix at 27.61 ± 1.07%; RNA-FrameFlow reaches 26.17 ± 1.76%); its strict codesignability is 32.06±0.75% against 26.00±1.38% for RiboFlow-Mix. RiboFlow-Exp falls below RNA-FrameFlow under matched experimental-only training, while RiboFlow-Mix exceeds it. The OpenDDE evaluation favors both variants over their baselines (Appendix G.2). The two variants

Table 3: Triangle-update interval at 200 steps. Interval counts blocks between pair updates; Base keeps pair features fixed. Mean ± standard deviation over three 600-RNA campaigns.

		Designability			Codesignability		
Model	Interval	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑
La-Ribo-Base	–	38.22 ± 1.27	43.44 ± 0.48	53.00 ± 0.67	**32.56** ± 2.25	**41.28** ± 1.84	56.72 ± 1.78
La-Ribo-Tri4	4	37.33 ± 2.03	45.00 ± 1.92	55.33 ± 1.36	25.17 ± 0.59	36.67 ± 1.21	52.67 ± 1.59
La-Ribo-Tri3	3	38.50 ± 2.09	44.00 ± 1.89	56.00 ± 1.41	26.83 ± 0.60	36.50 ± 1.15	52.00 ± 1.53
La-Ribo-Tri	2	**39.11** ± 2.12	**48.39** ± 2.08	**58.83** ± 1.48	32.06 ± 0.75	39.94 ± 1.26	**57.89** ± 1.71

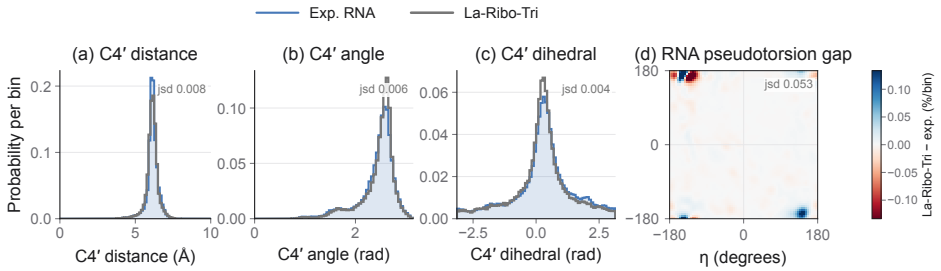


Figure 4: Local geometry of La-Ribo-Tri (gray) and experimental RNA (blue; RNASolo and RhoFold PDB). (a–c) compare C4′ distance, angle, and dihedral; (d) shows generated-minus-experimental pseudotorsion density. JSD is in bits.

are indistinguishable in codesignability, while La-Ribo-Tri leads designability topology success and cluster coverage at every budget; La-Ribo-Base samples faster (Appendix F.2). Strict success degrades beyond 120 nt (Figure 3), while local-geometry distributions track the experimental reference closely (JSD ≤ 0.008 bits for C4′ distance, angle, and dihedral; Figure 4). **Additional analyses.** Beyond self-consistency, the appendix evaluates secondary-structure agreement (Appendix E.1), atomistic geometry including clashes and ribose pucker, where both La-Ribo variants improve on the baselines (Appendix E.2), and proximity to the training corpus (Appendix E.3).

4.4 Ablation Studies

Autoencoder data and flow architecture We cross the corpus-trained VAE and experimental-only E-VAE with Base and Tri flows, all trained on the experimental partition and evaluated at 200 steps. Tri outperforms Base on all five metrics under both representations, with same-seed strict-designability gains of 4.94 ± 0.54 and 4.67 ± 1.88 points (Table 2). Representation and corpus each contribute. Trained only on experimental structures, E-VAE/Tri already reaches 30.17%, above RNA-FrameFlow (26.17%) and RiboFlow-Exp (20.50%), so the representation helps before any corpus expansion. With the autoencoder fixed, expanding the flow data adds 8.39 points for La-Ribo (39.11% versus 30.72%, Table 1) and 7.11 for RiboFlow (27.61%), so the gain is not representation-specific.

Triangle-update frequency Holding the representation, corpus, and evaluation fixed, we compare triangle updates every two, three, or four blocks with fixed-pair Base. At 200 steps, strict designability is indistinguishable across the four variants (37.33–39.11%), but topology success rises monotonically with update frequency (53.00% for Base to 58.83% for the two-block variant), and the two-block variant exceeds the sparser ones by more than five points in strict codesignability; Base remains comparable there at 32.56% (Table 3). Appendix D.2 reports the 50–400-step results.

Sampling sensitivity We vary solver steps K and schedule powers ρ_B, ρ_Z with 600 RNAs per setting. Averaged over the remaining factors, strict designability varies by under two points between 100 and 400 steps and changes by under 0.4 points across $\rho_B \in [0.70, 1.25]$ (Figure 5), so the reported results do not depend on a narrowly tuned schedule. Performance increases weakly with ρ_Z and our default sits at the top of the swept range; Appendix F.1 shows the full surfaces.

Table 4: Inverse folding on 1,002 experimental validation scaffolds, mean ± sample standard deviation over three independently seeded evaluations. Mean-of-eight averages candidates; best-of-eight uses one candidate for strict success. Rates are percentages; Unique counts distinct candidates (of eight); scRMSD is in Å.

	Sequence statistics		Mean of eight			Best of eight		
Method	Recovery ↑	Unique ↑	both@4.0 ↑	scRMSD ↓	scTM ↑	both@4.0 ↑	scRMSD ↓	scTM ↑
gRNAde	41.69 ± 0.13	6.14 ± 0.02	7.83 ± 0.31	16.39 ± 0.03	$\mathbf{0.258} \pm 0.0011$	11.18 ± 0.70	13.60 ± 0.16	0.289 ± 0.0029
NA-MPNN	$\mathbf{42.65} \pm 0.02$	7.22 ± 0.06	6.68 ± 0.16	16.68 ± 0.05	0.248 ± 0.0007	11.78 ± 0.35	13.65 ± 0.17	0.288 ± 0.0006
RiboFlow-Mix	32.14 ± 0.20	6.96 ± 0.01	4.12 ± 0.09	18.62 ± 0.05	0.192 ± 0.0011	9.02 ± 0.35	13.75 ± 0.24	0.219 ± 0.0008
La-Ribo-Base	34.43 ± 0.06	$\mathbf{7.67} \pm 0.02$	7.95 ± 0.09	16.36 ± 0.15	0.244 ± 0.0011	14.24 ± 0.35	$\mathbf{11.91} \pm 0.21$	0.309 ± 0.0015
La-Ribo-Tri	34.82 ± 0.10	7.62 ± 0.02	$\mathbf{8.84} \pm 0.13$	$\mathbf{15.90} \pm 0.17$	0.252 ± 0.0004	$\mathbf{14.84} \pm 0.70$	12.11 ± 0.13	$\mathbf{0.313} \pm 0.0020$

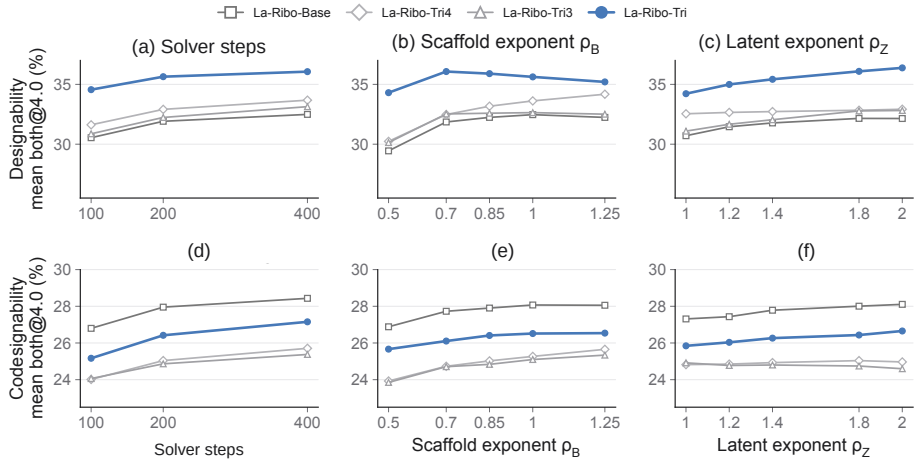


Figure 5: Sampling sensitivity. Columns show solver steps, scaffold power ρ_B, and latent power ρ_Z; top/bottom rows show marginal strict both@4.0 designability/codesignability.

4.5 Scaffold-conditioned inverse folding

Setup. Inverse-folding targets are drawn from the validation partition, whose sequence clusters are disjoint from all La-Ribo training data (Appendix A): its complete experimental subset (1,002 records spanning 107 clusters), with eight sequences per target. La-Ribo fixes B and samples Z without the native sequence. **Results.** Without any conditional training, the same prior directly performs scaffold-conditioned inverse folding: La-Ribo-Tri attains the highest best-of-eight strict success ($14.84 \pm 0.70\%$) and scTM (0.313), and La-Ribo-Base the lowest scRMSD (11.91 Å), versus $11.18 \pm 0.70\%$ for gRNAde, $11.78 \pm 0.35\%$ for NA-MPNN. Recovery orders the methods differently: gRNAde and NA-MPNN recover above 40%, while RiboFlow-Mix and both La-Ribo variants recover 32–35%. The comparison is conservative in training data and inputs: gRNAde and NA-MPNN were trained on PDB-derived corpora that may overlap these targets and receive native structures, whereas La-Ribo is cluster-held-out and conditions on three anchors per residue.

5 Conclusion

La-Ribo couples a sparse RNA scaffold with residue-wise latents in a shared geometry–latent flow for RNA sequence–structure co-design; with nucleotide identity carried by the latents, the same prior also serves as a scaffold-conditioned inverse folder without further training. To ease the structural supervision bottleneck, we built a quality-controlled corpus of 168,561 RNA structures, combining experimental data with predictions from three folding models. La-Ribo improves designability and codesignability over the evaluated baselines across sampling budgets and refolding models. Future work can condition generation on function and molecular interactions, evaluate designs experimentally, extend the corpus pipeline as new structures become available, and scale to longer RNAs, toward functional RNAs of therapeutic and synthetic-biology interest.

Ethics Statement

This work uses experimental and predicted molecular structures from public resources and involves no human subjects or personal data. RNA design may have unintended biological effects, and no sequence in this work was synthesised or assayed. Functional use should be preceded by biosafety review and experimental validation. Appendix H discusses broader impact.

Reproducibility Statement

Section 3 defines the method and Sections 4.1 and 4.3 the data and evaluation protocol. Appendix A documents corpus construction and the cluster-disjoint data split. Training hyperparameters for both stages are given in Tables 7 and 8, the exact objectives in Appendix B.2, and the training and sampling procedures as pseudocode in Algorithms 1 and 2. Appendix G records the evaluation and baseline settings. Code is available at https://anonymous.4open.science/r/la-ribo-7DDA.

AI Use Statement

We used large language models to polish the manuscript and improve the quality of the writing. Synthetic training structures were produced with dedicated biological structure prediction models (OpenFold3, NuFold, and trRosettaRNA2), as described in Section 4.1 and Appendix A. All AI-assisted text was reviewed by the authors, and we take responsibility for the content of this work.

References

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

Gustaf Ahdritz, Nazim Bouatta, Christina Floristean, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J O'Donnell, Daniel Berenberg, Ian Fisk, Niccolò Zanichelli, et al. Openfold: retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature methods*, 21(8):1514–1524, 2024.

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. *Transactions on Machine Learning Research*, 2025.

Aureka AI OpenDDE project. Folding, reasoning, and scaling with open-source drug discovery engine, 2026.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-hao Liu, Andrei Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. Se(3)-stochastic flow matching for protein backbone generation. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 22590–22621, 2024.

Andrew Campbell, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5453–5512. PMLR, 21–27 Jul 2024.

Xinang Cao, Yueying Zhang, Yiliang Ding, and Yue Wan. Identification of RNA structures and their roles in RNA functions. *Nature Reviews Molecular Cell Biology*, 25:784–801, 2024. doi: 10.1038/s41580-024-00748-6.

Jessica L. Childs-Disney, Xueyi Yang, Quentin MR Gibaut, Yuquan Tong, Robert T Batey, and Matthew D Disney. Targeting RNA structures with small molecules. *Nature Reviews Drug Discovery*, 21(10):736–762, 2022. doi: 10.1038/s41573-022-00521-4.

Padideh Danaee, Mason Rouches, Michelle Wiley, Dezhong Deng, Liang Huang, and David Hendrix. bpRNA: Large-Scale Automated Annotation and Analysis of RNA Secondary Structure. *Nucleic Acids Research*, 46(11):5381–5394, 2018. doi: 10.1093/nar/gky285.

Carlos M. Duarte and Anna Marie Pyle. Stepping through an RNA structure: a novel approach to conformational analysis. *Journal of Molecular Biology*, 284(5):1465–1478, 1998. doi: 10.1006/jmbi.1998.2233.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12606–12633. PMLR, 21–27 Jul 2024.

Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 67350–67395, 2026.

Han Huang, Ziqian Lin, Dongchen He, Liang Hong, and Yu Li. RiboDiffusion: tertiary structure-based RNA inverse folding with generative diffusion models. *Bioinformatics*, 40(Supplement_1): i347–i356, 2024.

Guillaume Huguet, James Vuckovic, Kilian Fatras, Eric Thibodeau-Laufer, Pablo Lemos, Riashat Islam, Cheng-Hao Liu, Jarrid Rector-Brooks, Tara Akhound-Sadegh, Michael Bronstein, Alexander Tong, and Avishek Joey Bose. Sequence-augmented se(3)-flow matching for conditional protein generation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 33007–33036. Curran Associates, Inc., 2024. doi: 10.52202/079017-1039.

Chaitanya Joshi, Arian Jamasb, Ramon Viñas, Charles Harris, Simon Mathis, Alex Morehead, Rishabh Anand, and Pietro Lio. grnade: Geometric deep learning for 3d rna inverse design. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 10166–10191, 2025.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):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*, 16:881, 2025. doi: 10.1038/s41467-025-56261-7.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Andrew Kubaney, Andrew Favor, Lilian McHugh, Raktim Mitra, Robert Pecoraro, Justas Dauparas, Cameron Glasscock, and David Baker. RNA Sequence Design and Protein–DNA Specificity Prediction with NA-MPNN. *bioRxiv*, 2025. doi: 10.1101/2025.10.03.679414.

Junzhe Li, Lijian Peng, Yuhao Li, Yize Zhou, Hanqun Cao, Cheng Tan, and Shengchao Liu. Base-and-Sugar Dual-Frame Flow Matching for RNA Co-Design. In *The 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)*, 2026a.

Xiner Li, Masatoshi Uehara, Xingyu Su, Gabriele Scalia, and Shuiwang Ji. A joint diffusion model with pre-trained priors for rna sequence-structure co-design. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 105615–105639, 2026b.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *The Eleventh International Conference on Learning Representations*, 2023.

Ronny Lorenz, Stephan H. Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6:26, 2011. doi: 10.1186/1748-7188-6-26.

Xiang-Jun Lu, Harmen J. Bussemaker, and Wilma K. Olson. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Research*, 43(21):e142, 2015. doi: 10.1093/nar/gkv716.

Runze Ma, Zhongyue Zhang, Zichen Wang, Chenqing Hua, Jiahua Rao, Zhuomin Zhou, and Shuangjia Zheng. Riboflow: Conditional de novo rna co-design via synergistic flow matching. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 69799–69831. Curran Associates, Inc., 2025. doi: 10.52202/085713-2348.

Divya Nori and Wengong Jin. RNAFlow: RNA structure & sequence design via inverse folding-based flow matching. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 38395–38408. PMLR, 21–27 Jul 2024.

Sawan Patel, Fred Zhangzhi Peng, Keith Fraser, Adam D. Friedman, Pranam Chatterjee, and Sherwood Yao. EvoFlow-RNA: Generating and Representing Non-Coding RNA with a Language Model. *bioRxiv*, 2025. doi: 10.1101/2025.02.25.639942.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.

Dana Rubin, Allan Dos Santos Costa, Manvitha Ponnapati, and Joseph Jacobson. RiboGen: RNA Sequence and Structure Co-Generation with Equivariant MultiFlow. In *ICLR 2025 Workshop on AI for Nucleic Acids*, 2025.

Tao Shen, Zhihang Hu, Siqi Sun, Di Liu, Felix Wong, Jiuming Wang, Jiayang Chen, Yixuan Wang, Liang Hong, Jin Xiao, et al. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nature Methods*, 21(12):2287–2298, 2024. doi: 10.1038/s41592-024-02487-0.

Zhen Tan, Yinghan Fu, Gaurav Sharma, and David H. Mathews. TurboFold II: RNA Structural Alignment and Secondary Structure Prediction Informed by Multiple Homologs. *Nucleic Acids Research*, 45(20):11570–11581, 2017. doi: 10.1093/nar/gkx815.

Sumit Tarafder and Debswapna Bhattacharya. RNAbpFlow: Base Pair-Augmented SE(3) Flow Matching for Conditional RNA 3D Structure Generation. *Nature Methods*, 23:1349–1358, 2026. doi: 10.1038/s41592-026-03128-4.

Leven M. Wadley, Kevin S. Keating, Carlos M. Duarte, and Anna Marie Pyle. Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. *Journal of Molecular Biology*, 372(4):942–957, 2007. doi: 10.1016/j.jmb.2007.06.058.

Jue Wang, Jintong Dong, Lan Yang, Tianhao Li, Jianwei Yin, Jintao Chen, Ying Dong, Jia Li, and Cheng Tan. Unlocking Programmable and Creative RNA Sequence Design with RDiffusion. *bioRxiv*, 2026a. doi: 10.64898/2026.06.13.732023.

Wenkai Wang, Zhenling Peng, and Jianyi Yang. Predicting RNA 3D Structure and Conformers Using a Pre-Trained Secondary Structure Model and Structure-Aware Attention. *Nature Machine Intelligence*, 8:722–734, 2026b. doi: 10.1038/s42256-026-01223-x.

Timothy J Wilson and David MJ Lilley. The potential versatility of RNA catalysis. *Wiley Interdisciplinary Reviews: RNA*, 12(5):e1651, 2021.

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. *arXiv preprint arXiv:2310.05297*, 2023.

Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods*, 19(9): 1109–1115, 2022. doi: 10.1038/s41592-022-01585-1.

Zaixi Zhang, Ruofan Jin, Linlin Chao, Guangxue Xu, Yikun Zhang, Guowei Zhou, Di Yin, Yingqing Guo, Yaqi Fu, Yukang Yang, Kaixuan Huang, Xiaotong Wang, Junze Zhang, Yujie Yang, Qirong Yang, Ziyao Xu, Weinan E, Ruhong Zhou, Xiaoming Zhang, Mengdi Wang, and Le Cong. RNAGenesis: A Generalist Foundation Model for Functional RNA Therapeutics. *bioRxiv*, 2024. doi: 10.1101/2024.12.30.630826.

Zhou Zhang, Hanqun Cao, Cheng Tan, Fang Wu, Pheng Ann Heng, and Tianfan Fu. RiboSphere: Learning Unified and Efficient Representations of RNA Structures. In *International Conference on Machine Learning*, 2026.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5745–5753, June 2019.

APPENDIX CONTENTS

A Dataset construction and corpus statistics . . . 16

A.1 Dataset construction . . . 16

A.2 Prepared representation . . . 17

A.3 Corpus distributions and sampling . . . 17

B Model and training details . . . 20

B.1 Architecture and optimization . . . 20

B.2 Scaffold-conditioned autoencoding: exact objective . . . 21

B.3 Geometry–latent flow: training and sampling . . . 23

C Additional representation analyses . . . 25

C.1 Latent interventions . . . 25

C.2 Scaffold choice and latent dimensionality . . . 25

D Extended generation results and controlled ablations . . . 26

D.1 Complete benchmark metrics . . . 26

D.2 Triangle-update frequency across budgets . . . 28

E Structural quality and novelty of generated structures . . . 28

E.1 Secondary-structure agreement . . . 28

E.2 Atomistic geometry . . . 29

E.3 Novelty relative to the training corpus . . . 30

F Sampling sensitivity and computational cost . . . 31

F.1 Modality-schedule sensitivity . . . 31

F.2 Sampling-factor response and runtime . . . 33

G Evaluation protocols, robustness, and inverse folding . . . 34

G.1 Primary evaluation settings and baseline implementation . . . 34

G.2 Unconditional generation with an alternative refolding model . . . 34

G.3 Scaffold-conditioned inverse-folding protocol . . . 34

H Limitations and broader impact . . . 35

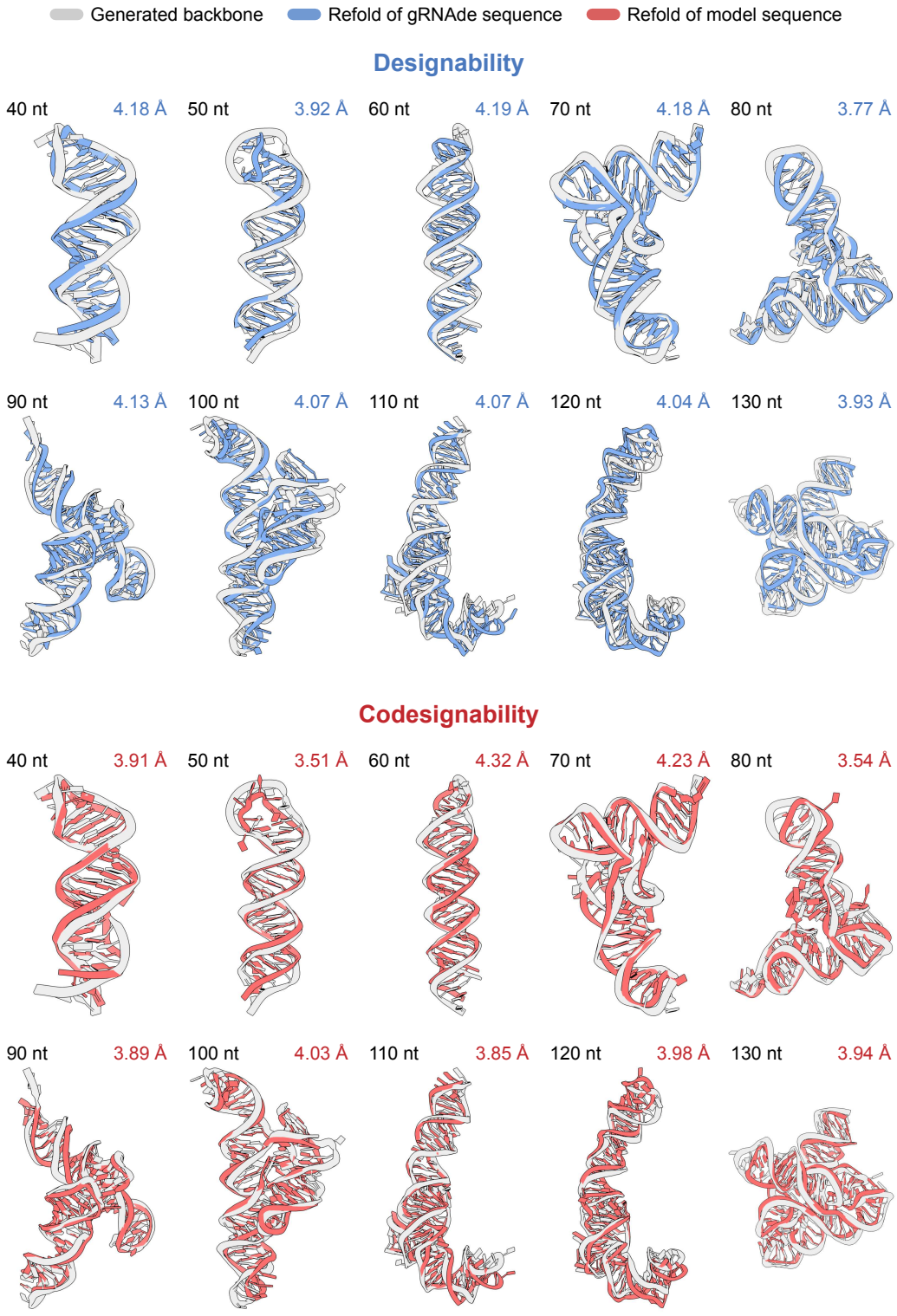


Figure 6: La-Ribo backbones (40–130 nt, grey) overlaid with refolds of the best of eight gRNAde sequences (top) and the co-designed sequence (bottom). Numbers: C4′ scRMSD (Å).

A Dataset construction and corpus statistics

We construct the RNA corpus by combining curated experimental structures, filtered synthetic structures from existing resources, and newly generated trRosettaRNA2 structures. After source-specific screening and common eligibility checks, 168,561 structures satisfy the 16–256-nucleotide length range and minimum 90% valid-frame coverage (Table 5). Source-specific preparation is summarized below; common preprocessing is detailed in Appendix A.1.

Table 5: Eligible RNA structures by source after length and valid-frame filtering. RNASolo and the RhoFold subset are experimental; other sources are synthetic.

Source	Type	Structures
RNASolo	Experimental	5,772
RhoFold subset	Experimental	3,268
OpenFold3 Distillation	Synthetic	125,833
NuFold Distillation	Synthetic	23,057
trRosettaRNA2 Distillation	Synthetic	10,631
Total	Experimental + Synthetic	**168,561**

Experimental structures. RNASolo contributes 5,772 eligible PDB-derived RNA chains (Adamczyk et al., 2022). A further 3,268 experimental PDB chains come from the RhoFold data collection (Shen et al., 2024).

Existing synthetic resources. OpenFold3 Distillation uses the public RNA self-distillation release[1], comprising synthetic structures for Rfam 15.1 cluster representatives after source confidence filtering. The common eligibility criteria retain 125,833 structures (Ahdritz et al., 2024). NuFold Distillation uses the public NuFold Structure Database[2]. We retain structures with mean pLDDT ≥ 80, R2DT consistency ≥ 0.70, and MSA depth ≥ 100, yielding 23,057 eligible structures (Kagaya et al., 2025).

New trRosettaRNA2 synthetic set. We merge sequences from bpRNA-1m(90) and RNAStralign (Danaee et al., 2018; Tan et al., 2017), restrict lengths to 16–256 nucleotides, and remove exact duplicates. We then generate MSA-supported structures with trRosettaRNA2, retain entries with mean pLDDT above 80, and use the relaxed top-ranked structure. This newly generated subset contributes 10,631 eligible structures (Wang et al., 2026b).

A.1 Dataset construction

The common pipeline aligns sequence and structure, extracts the geometric representation used by La-Ribo, and defines the dataset partition.

Sequence–structure alignment. Each entry pairs a canonical sequence $S = (s_1, \ldots, s_N)$, $s_i \in \mathcal{A} = \{A, C, G, U\}$, with a single-chain atomistic structure $X \in \mathbb{R}^{N\times 37\times 3}$. Residues are aligned by chain order or, for internally unresolved positions, by residue number. Ambiguous orderings, duplicate residue identifiers, unsupported residue names, and sequence–structure identity mismatches are removed.

Geometric feature extraction. Resolved heavy atoms are mapped to the atom37 representation with presence masks. For each residue, we extract the scaffold $b_i = [x_i^{\mathrm{P}}; x_i^{\mathrm{C4'}}; x_i^{\nu_i}]$, where $\nu_i = \mathrm{N9}$ for purines and $\nu_i = \mathrm{N1}$ for pyrimidines, with the canonical sugar frame $R_i^c = \mathcal{F}(x_i^{\mathrm{C4'}}, x_i^{\mathrm{O4'}}, x_i^{\mathrm{C3'}})$ and ten torsion targets. Valid-frame coverage determines structure-level eligibility; residue, atom, and torsion masks define the entries that contribute to autoencoder supervision.

[1] OpenFold3 datasets: https://portal.openfold.omsf.io/datasets.
[2] NuFold downloads: https://nufold.kiharalab.org/downloads.

Sequence-cluster partitioning. Sequences are clustered across all five sources with MMseqs2 at 90% sequence identity and 95% coverage. Entire clusters are assigned to a single partition, approximately 95/5, so no MMseqs2 cluster is shared across the two partitions. The resulting split contains 159,955 training and 8,606 validation records. Both scaffold-conditioned autoencoding and geometry–latent flow training use only the training partition; the validation partition is reserved exclusively for evaluation.

A.2 Prepared representation

Table 6 summarizes the prepared data and geometric quantities used by the scaffold-conditioned autoencoder. The scaffold B is extracted from X, while the latent representation Z is inferred by E_ϕ. Residue, atom, and torsion masks select valid supervision targets. Neither autoencoding nor geometry–latent flow uses MSAs or homologous structural templates as model inputs; MSAs are used only upstream to generate part of the synthetic corpus, while fixed nucleotide templates support local atom construction in the decoder.

Table 6: Prepared geometry and atom features for an RNA of length N; coordinates are in Å.

Quantity	**Shape**	**Definition and role**
Sequence S	N	Canonical nucleotide identities $s_i \in \{A, C, G, U\}$.
Heavy-atom coordinates X	$N \times 37 \times 3$	Atom37-style coordinates; $x_i^a \in \mathbb{R}^3$ denotes atom a in residue i.
Scaffold $B = (b_i)_{i=1}^N$	$N \times 9$	$b_i = [x_i^{\mathrm{P}}; x_i^{\mathrm{C4'}}; x_i^{\nu_i}]$, with $\nu_i = \mathrm{N9}$ for A/G and N1 for C/U.
Latent representation Z	$N \times 16$	Residue-wise latent inferred by $E_\phi(S, X, B)$; it complements B with nucleotide identity and local atomistic information.
Canonical sugar frame R_i^c	$N \times 3 \times 3$	$\mathcal{F}(x_i^{\mathrm{C4'}}, x_i^{\mathrm{O4'}}, x_i^{\mathrm{C3'}})$, centered at $x_i^{\mathrm{C4'}}$.
Torsion targets	$N \times 10 \times 2$	Ten RNA dihedrals represented as normalized sine–cosine pairs and used to supervise U_i.
Validity masks	$N \times 37$, $N \times 10$, N	Atom, torsion, and residue masks selecting valid supervision targets.

A.3 Corpus distributions and sampling

Experimental and synthetic structures share the dominant local-geometry modes (Figure 7). One-dimensional Jensen–Shannon divergences are below 0.055 bits, and median consecutive-$\mathrm{C4'}$ distances differ by less than 0.014 Å. The joint (η, θ) divergence is 0.1007 bits; among synthetic sources, OpenFold3 and trRosettaRNA2 differ most (0.1697 bits). The length distribution peaks near 73 nucleotides because of the large OpenFold3 collection, while the experimental, NuFold, and trRosettaRNA2 sources provide complementary coverage (Figure 8). These differences motivate retaining both source and length information during training. Distance plots stop at 10 Å, excluding fewer than 0.004% of observations; summary statistics retain all values. Restricting the analysis to the training partition leaves the reported medians and circular means unchanged.

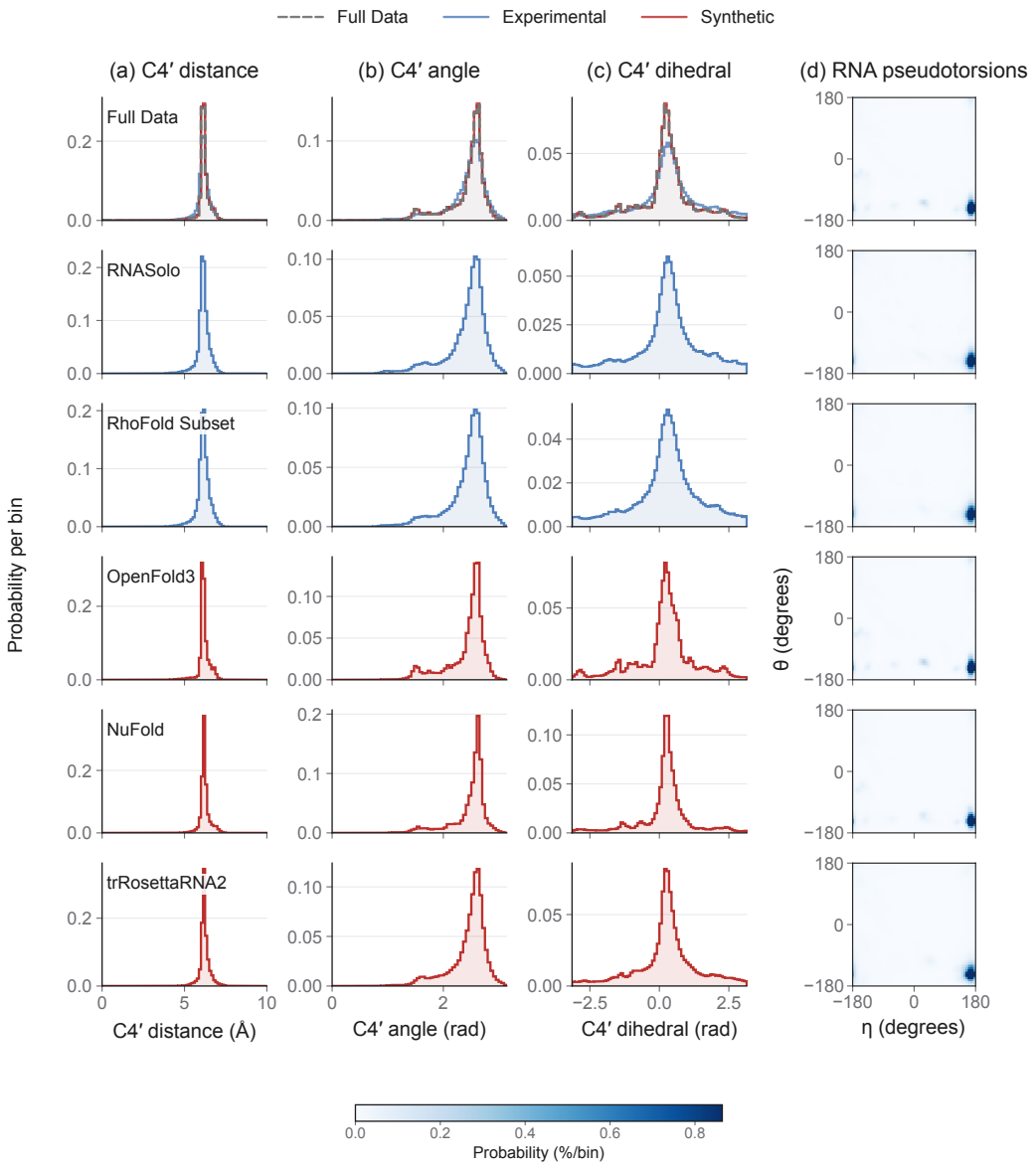


Figure 7: Local geometry by corpus source. Rows show the pooled corpus and individual sources; columns show C4′ distance, angle, dihedral, and joint (η, θ) density. In the pooled row, blue and red outlines distinguish experimental and synthetic structures. Histograms are normalized within each cohort; density panels share a color scale.

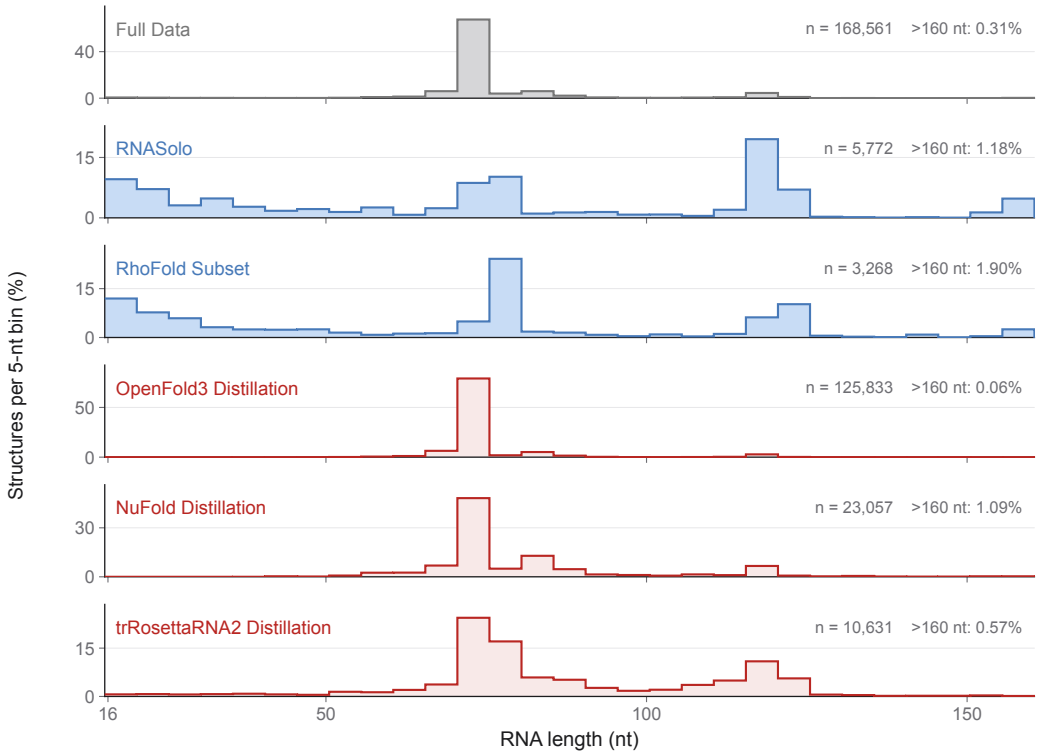


Figure 8: Length distributions of 168,561 eligible RNAs, pooled (top) and by source. Histograms are row-normalized; labels report the fraction above 160 nt. Gray, blue, and red denote pooled, experimental, and synthetic structures.

B MODEL AND TRAINING DETAILS

This appendix specifies the architecture, optimization, and exact training procedures underlying the scaffold-conditioned autoencoder and geometry–latent flow. Tables 7 and 8 summarize the stage-specific hyperparameters; subsequent sections expand the two objectives and the sampling procedure.

B.1 ARCHITECTURE AND OPTIMIZATION

Table 7: Scaffold-conditioned autoencoder hyperparameters.

Setting	Value
Optimizer	`AdamW`, $\beta = (0.9, 0.999)$
Learning rate	10^{-4}
Weight decay	10^{-2}
Epochs	100
Batch size	24 max (length-dependent)
Gradient clipping	1.0
Objective weights	$\lambda_{\mathrm{gt}} = 0.25$, $\lambda_{\mathrm{pred}} = 0.10$, $\lambda_{\mathrm{KL}} = 10^{-2}$

Table 8: Geometry–latent flow hyperparameters and default sampling settings.

Setting	Value
Optimizer	`AdamW`, $\beta = (0.9, 0.999)$
Learning rate	10^{-4}
Weight decay	10^{-2}
Epochs	100
Batch size	24 max (length-dependent)
Gradient clipping	1.0
Self-conditioning probability	0.5
EMA decay	0.999
Default sampling parameters	$\rho_B = 1.25$, $\rho_Z = 2.0$

Architecture. The encoder E_ϕ and decoder D_ψ each use six token–pair blocks, token width 256, pair width 128, and eight attention heads. The latent width is 16, and the autoencoder contains 14.91M parameters. The flow network uses 12 blocks, token width 512, pair width 256, and eight heads. Its residue features combine noisy scaffold coordinates, latent states, and modality times. Pair features encode relative sequence separation and distances among the scaffold anchors x_i^{P}, $x_i^{\mathrm{C4'}}$, and $x_i^{\nu_i}$ from the noisy and self-conditioned scaffold states; these features enter residue attention as pair-dependent biases.

La-Ribo-Base keeps the pair representation fixed throughout the flow trunk. La-Ribo-Tri additionally refines the pair stream from evolving residue features using token-to-pair projections and triangle multiplicative updates every two blocks. La-Ribo-Tri3 and La-Ribo-Tri4 use the same update with intervals of three and four blocks. The trainable parameter counts are 58.82M for La-Ribo-Base, 61.46M for La-Ribo-Tri4, 62.77M for La-Ribo-Tri3, and 65.41M for La-Ribo-Tri. These variants therefore compare different pair-update architectures rather than parameter-matched models. Figure 9 gives a compact architectural sketch focused on the Base–Tri difference.

Length-aware batching. To control memory growth with sequence length, training uses length buckets with a length-dependent per-device batch cap of at most 24 sequences. Within each bucket, source and sequence cluster are sampled hierarchically before a cluster member is drawn.

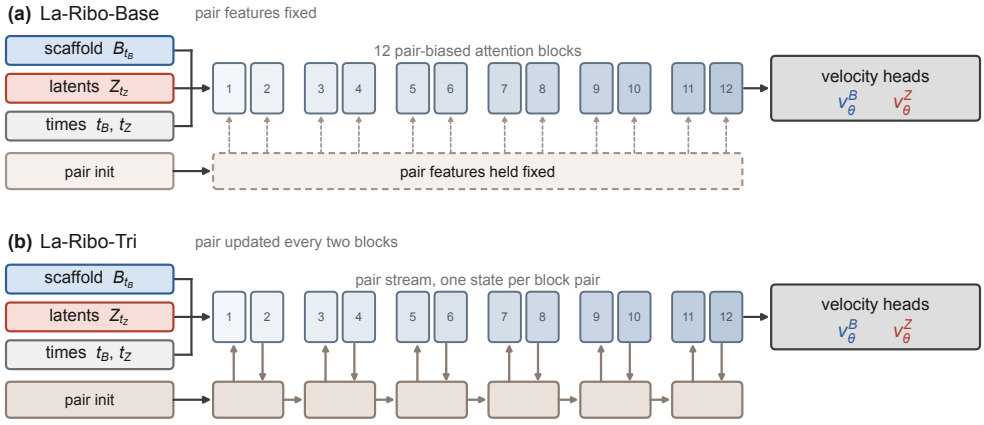


Figure 9: Base and Tri flow architectures. Both use a 12-block trunk to predict scaffold and latent velocities. La-Ribo-Base keeps pair features fixed as attention biases; La-Ribo-Tri updates them every two blocks using token-to-pair projections and triangle multiplicative updates.

B.2 Scaffold-conditioned autoencoding: exact objective

This section expands the scaffold-conditioned autoencoding objective in Section 3.2. We retain the main-text convention that Q_i, U_i, and R_i^{dec} are decoder outputs; a superscript ref identifies a target torsion where a distinction is needed.

Local frames. For noncollinear anchors, the deterministic operator $\mathcal{F}(o, u, v) = [e_1, e_2, e_3]$ constructs a right-handed orientation:

$$e_1 = \frac{u - o}{\|u - o\|_2}, \quad e_2 = \frac{(v - o) - e_1 e_1^\top (v - o)}{\|(v - o) - e_1 e_1^\top (v - o)\|_2}, \quad e_3 = e_1 \times e_2. \tag{11}$$

Basis vectors are matrix columns; the origin o is stored separately. The scaffold pseudo-frame is built from the input scaffold and the target sugar frame from atomistic coordinates:

$$R_i^p = \mathcal{F}(x_i^{\mathrm{C4'}}, x_i^{\nu_i}, x_i^{\mathrm{P}}), \quad R_i^c = \mathcal{F}(x_i^{\mathrm{C4'}}, x_i^{\mathrm{O4'}}, x_i^{\mathrm{C3'}}). \tag{12}$$

Both frames use the same C4′ origin. The target relative rotation is $(R_i^p)^\top R_i^c$. The decoder's six-dimensional rotation output (Zhou et al., 2019) is orthonormalized into $Q_i \in \mathrm{SO}(3)$, giving $R_i^{\mathrm{dec}} = R_i^p Q_i$. Degenerate inputs use the implementation's deterministic numerical guards. The three scaffold anchors need not form a rigid chemical group to define this coordinate frame.

Table 9: Atom-construction torsions and their standard relations. Each quartet defines a reconstruction dihedral; "reversed" reverses its order without changing the angle, and "shifted" replaces terminal atoms. Here ν_i follows Equation (1), and A_{base} is C4 for A/G or C2 for C/U.

Index	Atom quartet	Standard relation
1	$\mathrm{C5'} - \mathrm{C4'} - \mathrm{C3'} - \mathrm{C2'}$	δ shifted (C2′ for O3′)
2	$\mathrm{C3'} - \mathrm{C4'} - \mathrm{O4'} - \mathrm{C1'}$	ν_4
3	$\mathrm{C4'} - \mathrm{O4'} - \mathrm{C1'} - \nu_i$	ν_0 shifted (ν_i for C2′)
4	$\mathrm{C5'} - \mathrm{C4'} - \mathrm{C3'} - \mathrm{O3'}$	δ
5	$\mathrm{C3'} - \mathrm{C4'} - \mathrm{C5'} - \mathrm{O5'}$	γ reversed
6	$\mathrm{C4'} - \mathrm{C5'} - \mathrm{O5'} - \mathrm{P}$	β reversed
7	$\mathrm{C5'} - \mathrm{O5'} - \mathrm{P} - \mathrm{OP1}$	α reversed and shifted (OP1 for $\mathrm{O3'}_{i-1}$)
8	$\mathrm{C5'} - \mathrm{O5'} - \mathrm{P} - \mathrm{OP2}$	α reversed and shifted (OP2 for $\mathrm{O3'}_{i-1}$)
9	$\nu_i - \mathrm{C1'} - \mathrm{C2'} - \mathrm{O2'}$	ν_1 shifted (ν_i, O2′)
10	$\mathrm{C2'} - \mathrm{C1'} - \nu_i - A_{\mathrm{base}}$	χ shifted (C2′ for O4′)

Torsion-based atom construction. Table 9 specifies the ten dihedrals used in extraction and reconstruction. Entries 1 and 4 share the $C4' - C3'$ bond, and entries 7 and 8 share $O5' - P$; the ten entries therefore span eight bonds and allow separate placement of substituent branches. Inter-residue ϵ and ζ are not construction variables because each residue is built from its own sugar frame. For entries 9 and 10, upstream sugar rotations alter the atoms defining the local frame. We rebuild their zero-angle frames from the current $C1'$, $C2'$, and ν_i coordinates.

Template expansion and gradient paths. One decoder forward pass produces nucleotide probabilities π, relative rotations Q, and normalized sine–cosine torsion vectors U. The same frames and torsions are expanded twice: the gt view uses native identities s_i, and the pred view uses $s_i^{\mathrm{pred}} = \arg\max_{a \in \mathcal{A}} \pi_i(a)$. The $C4'$ coordinate is copied as the frame origin; the P and ν_i anchors are reconstructed. Both views are supervised against the reference structure.

Predicted-template selection is non-differentiable. Its geometric losses update the encoder, shared decoder, and geometric output heads, but do not backpropagate through argmax to the nucleotide logits. The nucleotide head is supervised by cross-entropy. Sequence, rotation, and torsion losses are included once; the second template view adds only geometric reconstruction terms.

Template-independent losses. Let $\mathcal{R}$ be the valid residue set and $\mathcal{T}$ the valid residue–torsion pairs. For target angles τ_{ik}, define $u_{ik}^{\mathrm{ref}} = (\sin \tau_{ik}, \cos \tau_{ik})$; u_{ik} is the corresponding normalized decoder output. The losses are

$$\begin{aligned} \mathcal{L}_{\mathrm{seq}} &= -\frac{1}{|\mathcal{R}|} \sum_{i \in \mathcal{R}} \log \pi_i(s_i), \\ \mathcal{L}_{\mathrm{rot}} &= \frac{1}{|\mathcal{R}|} \sum_{i \in \mathcal{R}} d^2_{\mathrm{SO}(3)}(R_i^{\mathrm{dec}}, R_i^{c}), \\ \mathcal{L}_{\mathrm{tors}} &= \frac{1}{2|\mathcal{T}|} \sum_{(i,k) \in \mathcal{T}} \|u_{ik} - u_{ik}^{\mathrm{ref}}\|_2^2. \end{aligned} \tag{13}$$

Here $d_{\mathrm{SO}(3)}$ is the geodesic rotation angle. All sums exclude masked entries. The factor $1/2$ in the torsion loss averages the two sine–cosine components.

Template-dependent geometric losses. For $v \in \{\mathrm{gt}, \mathrm{pred}\}$, let M_i be the valid target atom names and M_i^v those instantiated by template v. Write $x_i^{a,v}$ for a reconstructed atom and b_i^v for the scaffold extracted from that reconstruction using the anchors in Equation (1). The glycosidic anchor is selected according to the instantiated nucleotide, so it remains an attachment-site comparison even when base type changes. We use

$$\begin{aligned} \mathcal{L}^v_{\mathrm{atom}} &= \frac{\sum_{i \in \mathcal{R}} \sum_{a \in M_i \cap M_i^v} \|x_i^{a,v} - x_i^a\|_2^2}{\sum_{i \in \mathcal{R}} |M_i \cap M_i^v|}, \\ \mathcal{L}^v_{\mathrm{scaf}} &= \frac{1}{9|\mathcal{R}|} \sum_{i \in \mathcal{R}} \|b_i^v - b_i\|_2^2, \\ \mathcal{L}^v_{\mathrm{pseudo}} &= \frac{1}{2} \left[\ell_{\mathrm{circ}}(\eta^v, \eta; \Omega_\eta) + \ell_{\mathrm{circ}}(\theta^v, \theta; \Omega_\theta) \right]. \end{aligned} \tag{14}$$

The atom term averages squared Euclidean distances over shared atom names; unmatched atoms are not assigned zero-coordinate targets. The scaffold term averages its nine coordinate components against the reference scaffold. For a periodic angle, the circular loss is

$$\ell_{\mathrm{circ}}(\alpha^v, \alpha; \Omega) = \frac{1}{|\Omega|} \sum_{j \in \Omega} \|c(\alpha_j^v) - c(\alpha_j)\|_2^2, \qquad c(\alpha) = (\sin \alpha, \cos \alpha). \tag{15}$$

Ω_η and Ω_θ contain the measurable trace pseudotorsions. Using coordinate arguments throughout,

$$\begin{aligned} \eta_i &= \mathrm{dih}(x_{i-1}^{C4'}, x_i^{\mathrm{P}}, x_i^{C4'}, x_{i+1}^{\mathrm{P}}), \\ \theta_i &= \mathrm{dih}(x_i^{\mathrm{P}}, x_i^{C4'}, x_{i+1}^{\mathrm{P}}, x_{i+1}^{C4'}). \end{aligned} \tag{16}$$

These follow the RNA pseudotorsional convention of Duarte & Pyle (1998); Wadley et al. (2007). Reconstructed values use X^v; joint distributions pair η and θ at the same residue. Their three geometric losses sum to $\mathcal{L}^v_{\mathrm{geom}}$ in Equation (5).

KL regularization and weights. The residue-wise Gaussian parameters from E_ϕ are regularized by

$$\mathcal{L}_{\mathrm{KL}} = \frac{1}{2|\mathcal{R}|} \sum_{i \in \mathcal{R}} \sum_{d=1}^{16} \left(\mu_{id}^2 + \sigma_{id}^2 - \log \sigma_{id}^2 - 1 \right). \tag{17}$$

Latent dimensions are summed within each residue and then averaged over valid residues. The small KL coefficient trades off latent regularization and reconstruction accuracy.

B.3 Geometry–latent flow: training and sampling

The geometry–latent flow is trained after freezing E_ϕ and D_ψ. Algorithms 1 and 2 summarize training and unconditional sampling procedures using the same states and updates as the main-text formulation. Exact normalization, scaffold recentering, time sampling, and endpoint stabilization are specified immediately afterward; sg denotes stop-gradient.

Algorithm 1 La-Ribo Training

Require: RNA example (S, X, B); frozen encoder E_ϕ; velocity network v_θ

Ensure: Updated flow parameters θ

$\mu \leftarrow E_\phi(S, X, B)$

$(x_1^B, x_1^Z) \leftarrow \mathrm{Normalize}(B, \mu)$

$\epsilon^B, \epsilon^Z \sim \mathcal{N}(0, I)$; $x_0^B \leftarrow \Pi_B \epsilon^B$; $x_0^Z \leftarrow \epsilon^Z$

Sample $t_B, t_Z \sim p_t$ independently

for $a \in \{B, Z\}$ **do**

  $x_{t_a}^a \leftarrow (1 - t_a) x_0^a + t_a x_1^a$ ▷ Eq. 8

end for

$x_{\mathrm{sc}} \leftarrow 0$

$u \sim \mathrm{Bernoulli}(0.5)$

if $u = 1$ **then**

  $(\widetilde{v}^B, \widetilde{v}^Z) \leftarrow v_\theta(x_t, 0, t_B, t_Z)$

  for $a \in \{B, Z\}$ **do**

    $x_{\mathrm{sc}}^a \leftarrow \mathrm{sg}\left[x_{t_a}^a + (1 - t_a)\widetilde{v}^a\right]$

  end for

end if

$(v^B, v^Z) \leftarrow v_\theta(x_t, x_{\mathrm{sc}}, t_B, t_Z)$

for $a \in \{B, Z\}$ **do**

  $x_{1,\theta}^a \leftarrow x_{t_a}^a + (1 - t_a) v^a$ ▷ Eq. 20

end for

$$\mathcal{L}_{\mathrm{FM}} \leftarrow \sum_{a \in \{B,Z\}} \frac{\sum_{i \in \mathcal{R}} \|(x_{1,\theta}^a - x_1^a)_i\|_2^2}{d_a |\mathcal{R}| \left((1 - t_a)^2 + \varepsilon_{\mathrm{stab}}\right)}$$

Update $\theta \leftarrow \mathrm{AdamW}(\theta, \nabla_\theta \mathcal{L}_{\mathrm{FM}})$

Algorithm 2 La-Ribo Unconditional Sampling

Require: Length N; steps K; powers (ρ_B, ρ_Z); trained v_θ; frozen decoder D_ψ

Ensure: Generated nucleotide sequence $\widehat{S}$ and atomistic coordinates $\widehat{X}$

$\epsilon^B, \epsilon^Z \sim \mathcal{N}(0, I)$; $x_0^B \leftarrow \Pi_B \epsilon^B$; $x_0^Z \leftarrow \epsilon^Z$
$x_{\mathrm{sc}} \leftarrow 0$
for $k = 0, \ldots, K-1$ **do**
  $t_k^Z \leftarrow (k/K)^{\rho_Z}$; $t_{k+1}^Z \leftarrow ((k+1)/K)^{\rho_Z}$
  $t_k^B \leftarrow (k/K)^{\rho_B}$; $t_{k+1}^B \leftarrow ((k+1)/K)^{\rho_B}$ ▷ Eq. 10
  $(v_k^B, v_k^Z) \leftarrow v_\theta(x_k, x_{\mathrm{sc}}, t_k^B, t_k^Z)$
  for $a \in \{B, Z\}$ **do**
    $x_{1,k}^a \leftarrow x_{t_k^a}^a + (1 - t_k^a) v_k^a$
  end for
  $x_{t_{k+1}^B}^B \leftarrow x_{t_k^B}^B + (t_{k+1}^B - t_k^B) v_k^B$
  $x_{t_{k+1}^Z}^Z \leftarrow x_{t_k^Z}^Z + (t_{k+1}^Z - t_k^Z) v_k^Z$ ▷ Eq. 22
  $x_{\mathrm{sc}} \leftarrow \mathrm{sg}(x_{1,k})$
end for
$(B, Z) \leftarrow \mathrm{Denormalize}(x_1^B, x_1^Z)$
$(\pi, Q, U) \leftarrow D_\psi(B, Z)$; $\widehat{S}_i \leftarrow \arg\max_{a \in \mathcal{A}} \pi_i(a)$
$\widehat{X} \leftarrow \mathrm{AtomConstruct}(B, \widehat{S}, Q, U)$
return $(\widehat{S}, \widehat{X})$

Flow normalization and recentering. The flow targets are posterior means and scaffold coordinates. Each latent dimension is standardized by its empirical mean and standard deviation; the C4′-centered scaffold is divided by a single global RMS scale. Let Y be any scaffold-shaped coordinate or velocity array. Define

$$[\Pi_B Y]_i^a = Y_i^a - \frac{1}{|\mathcal{R}|} \sum_{j \in \mathcal{R}} Y_j^{\mathrm{C4}'}, \quad a \in \{\mathrm{P}, \mathrm{C4}', \nu_i\}, \qquad \Pi_Z = I. \tag{18}$$

This removes a common translation from all three anchors. Independent standard Gaussian arrays are used for the two base states, with $x_0^a = \Pi_a \epsilon^a$ for modality $a \in \{B, Z\}$. Thus the scaffold base is a centered Gaussian, not an unconstrained standard normal in the full coordinate space.

The main-text velocity v_θ includes this recentering. If f_θ denotes the raw network outputs, then

$$v_\theta^a(x_t, t_B, t_Z) = \Pi_a f_\theta^a(x_t, x_{\mathrm{sc}}, t_B, t_Z). \tag{19}$$

Here x_{sc} is a detached endpoint estimate supplied by self-conditioning, omitted from the main-text argument list. It is used with probability 0.5 during training. For each modality independently, time is drawn from an equal mixture of uniform and $\mathrm{Beta}(2, 2)$ distributions on $[0, 1]$, then rescaled to $[0.01, 0.99]$.

Stabilized endpoint objective. Define the predicted endpoint, with $\varepsilon_{\mathrm{stab}} = 10^{-5}$, as

$$x_{1,\theta}^a = x_{t_a}^a + (1 - t_a) v_\theta^a(x_t, t_B, t_Z). \tag{20}$$

The implemented objective is

$$\sum_{a \in \{B,Z\}} \mathbb{E}\left[\frac{\sum_{i \in \mathcal{R}} \|(x_{1,\theta}^a - x_1^a)_i\|_2^2}{d_a |\mathcal{R}| \big((1 - t_a)^2 + \varepsilon_{\mathrm{stab}}\big)} \right]. \tag{21}$$

Without the stabilizer this equals Equation (9); the implemented stabilizer multiplies each squared velocity error by $(1 - t_a)^2 / ((1 - t_a)^2 + \varepsilon_{\mathrm{stab}})$. No decoded atom or sequence loss is added during flow training.

Euler updates and conditional sampling. At step k, one shared network evaluation returns both velocities. Each modality uses its own step size,

$$x^a_{t^a_{k+1}} = x^a_{t^a_k} + (t^a_{k+1} - t^a_k) v^a_\theta(x_k, t^B_k, t^Z_k), \qquad x_k = (x^B_{t^B_k}, x^Z_{t^Z_k}). \tag{22}$$

The effective scaffold velocity enforces recentering, and the previous endpoint estimate supplies self-conditioning. The K Euler steps therefore require K shared velocity-network evaluations. Before decoding, coordinate scaling and latent standardization are reversed. For scaffold-conditioned inverse folding, the observed scaffold is fixed at $t_B = 1$ and only the latent modality is integrated. The resulting latent state is decoded together with the supplied scaffold, providing an approximate conditional sampler without a separately trained model.

C Additional representation analyses

C.1 Latent interventions

The intervention conditions are paired by structure on the cluster-held-out validation set. Scaffold B stays fixed. For atom-level reconstruction we use each RNA's reference nucleotide identity in every condition, whereas sequence recovery is computed from the decoder's nucleotide probabilities. The within-chain shuffle is separate from the single-residue locality perturbation. Non-scaffold RMSD excludes x_i^{P}, $x_i^{\mathrm{C4'}}$, and $x_i^{\nu_i}$. The $x_i^{\mathrm{C4'}}$ anchor is omitted from the anchor-error breakdown because it is copied as the frame origin. Under both regimes, zeroing or shuffling lowers sequence recovery and increases reconstruction error, demonstrating that latent values and their residue assignments contribute information beyond the supplied scaffold. Figure 2 visualizes the VAE interventions; Table 10 reports the expanded metrics for both models.

Table 10: Reconstruction under latent interventions.

Model	Latent	Sequence (%) ↑	Frame (°) ↓	Non-scaffold RMSD (Å) ↓	P (Å) ↓	Glycosidic N (Å) ↓
VAE	posterior	100.00 ± 0.00	1.779 ± 0.010	0.08746 ± 0.00021	0.08111 ± 0.00023	0.04709 ± 0.00048
VAE	zero	24.77 ± 0.73	16.447 ± 0.081	0.77254 ± 0.00651	0.64596 ± 0.02262	0.12169 ± 0.02587
VAE	shuffle	27.54 ± 0.02	20.699 ± 0.009	0.96503 ± 0.00650	0.75054 ± 0.00275	0.10439 ± 0.00108
E-VAE	posterior	100.00 ± 0.00	2.673 ± 0.000	0.12248 ± 0.00000	0.09940 ± 0.00000	0.05380 ± 0.00000
E-VAE	zero	24.99 ± 0.00	16.910 ± 0.000	0.87019 ± 0.00000	0.69486 ± 0.00000	0.11057 ± 0.00000
E-VAE	shuffle	27.58 ± 0.21	20.669 ± 0.052	0.98181 ± 0.00219	0.76289 ± 0.00061	0.11122 ± 0.00024

To test locality, we perturb one valid residue's latent while keeping the scaffold and all other latents fixed. We add $\Delta z = m(s_z \odot d)$ at that residue: m controls strength, s_z is the autoencoder-specific empirical standard deviation of posterior-mean latent coordinates, and d is a unit-RMS Gaussian direction held fixed across magnitudes. We compare non-scaffold atom displacement at the edited residue with displacement elsewhere. The much smaller response outside the edited residue indicates a predominantly local effect (Figure 2c).

C.2 Scaffold choice and latent dimensionality

We compare scaffold and latent-width variants on the same validation dataset. All variants are trained on experimental structures and evaluated by posterior-mean decoding.

Atoms are expanded with ground-truth nucleotide templates. For every representation, non-scaffold RMSD excludes P, C4′, and the glycosidic N. The C4′-only variant constructs its frame from the chain tangent and the plane of adjacent C4′ steps. The P/C4′ variant uses the C4′-to-P direction as its first axis and the orthogonalized chain tangent as its second; both use C4′ as the origin.

The three-anchor scaffold reduces all-atom and non-scaffold RMSD relative to the one- and two-anchor alternatives (Table 11). Among the three-anchor variants, width 16 gives the lowest RMSD values (mean non-scaffold RMSD 0.1225 Å, versus 0.1334 and 0.1227 Å at widths 8 and 32); all generation experiments use the three-anchor scaffold with width 16.

Table 11: Reconstruction by scaffold and latent width d_Z using experimental-only training and a common validation subset. Scaffold labels P, $\mathrm{C4}'$, and ν_i abbreviate the anchors x_i^{P}, $x_i^{\mathrm{C4}'}$, and $x_i^{\nu_i}$. AA/NS denote all-atom/non-scaffold RMSD (unaligned mean/median, Å); Frame is angular error and Torsion is MAE (degrees, averaged over valid residues); Seq is sequence recovery in percent. Bold marks the best geometric value per column.

Scaffold	d_Z	AA-RMSD ↓	NS-RMSD ↓	Frame ↓	Torsion ↓	Seq ↑
$\mathrm{C4}'$	16	0.1533 / 0.1444	0.1589 / 0.1498	2.389	2.326	99.999
$\mathrm{P}, \mathrm{C4}'$	16	0.1234 / 0.1173	0.1304 / 0.1240	**2.313**	**2.132**	100.000
$\mathrm{P}, \mathrm{C4}', \nu_i$	8	0.1259 / 0.1198	0.1334 / 0.1267	2.516	2.418	100.000
$\mathrm{P}, \mathrm{C4}', \nu_i$	16	**0.1159 / 0.1107**	**0.1225 / 0.1170**	2.673	2.649	100.000
$\mathrm{P}, \mathrm{C4}', \nu_i$	32	0.1267 / 0.1117	0.1227 / 0.1176	2.886	2.891	100.000

D Extended generation results and controlled ablations

D.1 Complete benchmark metrics

The following tables expand the common protocol in Section 4.1 and Table 1. Success rates are percentages; continuous errors and cluster counts are reported alongside thresholded endpoints. Sampling steps are velocity-network evaluations for the forward-Euler solver.

Table 12: Expanded unconditional-generation metrics under the protocol of Table 1. Rates are percentages and scRMSD is in Å; bold marks the best mean across operating points.

(a) Designability

Method	Steps	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑	#Clust. ↑
RNA-FrameFlow	50	25.17 ± 1.09	29.89 ± 0.42	36.61 ± 0.19	11.58 ± 0.47	0.406 ± 0.006	24.00 ± 1.73
	100	26.61 ± 1.18	32.17 ± 1.48	39.67 ± 0.33	11.36 ± 0.16	0.417 ± 0.003	27.00 ± 2.65
	200	26.17 ± 1.76	31.56 ± 2.10	38.44 ± 1.84	11.47 ± 0.29	0.413 ± 0.008	26.00 ± 1.00
	400	26.89 ± 1.69	32.89 ± 1.84	38.89 ± 1.69	11.36 ± 0.30	0.415 ± 0.005	27.00 ± 3.61
RiboFlow-Exp	50	22.28 ± 0.63	25.11 ± 0.42	30.44 ± 0.54	12.87 ± 0.24	0.409 ± 0.004	19.33 ± 4.04
	100	22.22 ± 0.54	26.06 ± 0.59	31.22 ± 0.98	12.80 ± 0.41	0.412 ± 0.003	18.67 ± 1.53
	200	20.50 ± 0.73	25.11 ± 0.42	30.06 ± 0.69	12.73 ± 0.22	0.408 ± 0.003	18.00 ± 2.65
	400	21.39 ± 0.54	25.61 ± 0.92	29.94 ± 0.77	12.93 ± 0.42	0.405 ± 0.004	17.00 ± 2.00
RiboFlow-Mix	50	29.00 ± 0.17	29.67 ± 0.47	37.28 ± 1.09	11.26 ± 0.21	0.428 ± 0.005	23.00 ± 3.46
	100	29.06 ± 0.54	33.67 ± 0.67	38.44 ± 0.70	11.48 ± 0.36	0.452 ± 0.004	22.33 ± 2.31
	200	27.61 ± 1.07	30.89 ± 0.38	36.89 ± 0.92	11.52 ± 0.19	0.443 ± 0.002	21.67 ± 1.53
	400	28.39 ± 0.10	30.78 ± 0.83	37.39 ± 0.91	10.99 ± 0.47	0.437 ± 0.003	20.67 ± 2.31
La-Ribo-Base	50	36.11 ± 2.07	42.89 ± 1.73	50.78 ± 2.00	9.67 ± 0.60	0.457 ± 0.012	22.67 ± 1.53
	100	37.11 ± 0.98	43.83 ± 0.50	51.83 ± 1.20	9.54 ± 0.43	0.464 ± 0.007	22.67 ± 1.15
	200	38.22 ± 1.27	43.44 ± 0.48	53.00 ± 0.67	9.38 ± 0.29	0.465 ± 0.006	23.00 ± 1.73
	400	37.72 ± 1.13	44.17 ± 0.60	53.11 ± 1.55	9.45 ± 0.36	0.465 ± 0.007	24.33 ± 1.15
La-Ribo-Tri	50	37.11 ± 0.25	45.22 ± 1.21	54.06 ± 0.35	8.80 ± 0.24	0.474 ± 0.004	29.00 ± 2.65
	100	37.78 ± 1.92	46.89 ± 2.41	56.89 ± 2.22	8.56 ± 0.32	0.487 ± 0.008	28.33 ± 5.51
	200	**39.11** ± 2.12	**48.39** ± 2.08	58.83 ± 1.48	8.59 ± 0.08	0.491 ± 0.002	27.67 ± 1.15
	400	38.61 ± 2.46	**48.39** ± 2.75	**59.33** ± 0.93	**8.45** ± 0.02	**0.494** ± 0.004	**29.33** ± 3.79

(b) Codesignability

Method	Steps	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑	#Clust. ↑
RNA-FrameFlow	50	–	–	–	–	–	–
	100	–	–	–	–	–	–
	200	–	–	–	–	–	–
	400	–	–	–	–	–	–
RiboFlow-Exp	50	19.22 ± 0.35	21.83 ± 0.76	24.94 ± 0.10	15.42 ± 0.21	0.380 ± 0.003	15.33 ± 2.52
	100	19.28 ± 0.42	22.83 ± 1.09	25.78 ± 1.25	15.52 ± 0.11	0.386 ± 0.001	15.67 ± 2.08
	200	19.33 ± 1.61	22.83 ± 1.76	26.44 ± 2.55	15.01 ± 0.14	0.388 ± 0.006	15.67 ± 1.15
	400	20.06 ± 0.19	23.56 ± 0.48	27.61 ± 0.69	15.04 ± 0.42	0.392 ± 0.003	15.67 ± 1.53
RiboFlow-Mix	50	26.61 ± 1.34	27.89 ± 0.83	31.83 ± 0.58	14.01 ± 0.18	0.419 ± 0.004	19.00 ± 2.65
	100	26.67 ± 0.67	30.00 ± 1.01	32.50 ± 0.71	13.61 ± 0.14	0.437 ± 0.002	19.33 ± 2.08
	200	26.00 ± 1.38	29.06 ± 1.62	33.06 ± 1.13	14.15 ± 0.17	0.399 ± 0.005	19.00 ± 1.73
	400	26.61 ± 1.57	27.83 ± 0.58	34.72 ± 1.34	13.01 ± 0.36	0.403 ± 0.004	19.33 ± 1.53
La-Ribo-Base	50	30.00 ± 2.77	39.11 ± 2.38	52.78 ± 2.18	10.31 ± 0.29	0.456 ± 0.010	24.00 ± 1.00
	100	32.11 ± 2.14	40.56 ± 1.87	55.28 ± 1.55	10.14 ± 0.13	0.467 ± 0.007	23.33 ± 3.06
	200	**32.56** ± 2.25	41.28 ± 1.84	56.72 ± 1.78	9.85 ± 0.16	0.473 ± 0.007	25.33 ± 2.31
	400	**32.56** ± 1.58	**41.72** ± 0.69	57.17 ± 0.29	9.93 ± 0.04	0.475 ± 0.002	25.67 ± 3.51
La-Ribo-Tri	50	28.61 ± 1.13	36.06 ± 1.93	50.83 ± 2.09	10.53 ± 0.28	0.449 ± 0.009	**28.00** ± 1.00
	100	30.78 ± 1.26	39.00 ± 1.83	56.11 ± 1.95	9.86 ± 0.24	0.471 ± 0.008	25.33 ± 3.21
	200	32.06 ± 0.75	39.94 ± 1.26	57.89 ± 1.71	9.67 ± 0.11	0.477 ± 0.003	27.33 ± 0.58
	400	32.33 ± 0.58	40.17 ± 1.17	**58.11** ± 0.84	**9.55** ± 0.09	**0.479** ± 0.003	**28.00** ± 1.00

D.2 TRIANGLE-UPDATE FREQUENCY ACROSS BUDGETS

We compare pair updates every two, three, and four blocks at the common sampling budgets. The two-block variant is La-Ribo-Tri; La-Ribo-Tri3 and La-Ribo-Tri4 denote the three- and four-block variants. The representation, flow objective, and training corpus are shared; the number of refinement modules and parameter count vary with the update interval.

Table 13: Designability by triangle-update interval. Mean ± standard deviation over three 600-RNA campaigns per setting; rates are percentages and scRMSD is in Å.

Model	Steps	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑	#Clust.↑
La-Ribo-Tri	50	37.11 ± 0.25	45.22 ± 1.21	54.06 ± 0.35	8.80 ± 0.24	0.474 ± 0.004	29.00 ± 2.65
	100	37.78 ± 1.92	46.89 ± 2.41	56.89 ± 2.22	8.56 ± 0.32	0.487 ± 0.008	28.33 ± 5.51
	200	39.11 ± 2.12	48.39 ± 2.08	58.83 ± 1.48	8.59 ± 0.08	0.491 ± 0.002	27.67 ± 1.15
	400	38.61 ± 2.46	48.39 ± 2.75	59.33 ± 0.93	8.45 ± 0.02	0.494 ± 0.004	29.33 ± 3.79
La-Ribo-Tri3	50	31.17 ± 0.17	37.67 ± 1.00	49.00 ± 0.29	9.54 ± 0.26	0.459 ± 0.004	26.33 ± 2.31
	100	36.50 ± 1.86	42.00 ± 2.16	52.83 ± 2.06	9.06 ± 0.34	0.479 ± 0.008	26.33 ± 5.13
	200	38.50 ± 2.09	44.00 ± 1.89	56.00 ± 1.41	8.64 ± 0.09	0.487 ± 0.002	26.33 ± 1.15
	400	38.00 ± 2.42	45.50 ± 2.59	55.33 ± 0.87	8.40 ± 0.02	0.489 ± 0.004	27.33 ± 3.51
La-Ribo-Tri4	50	33.33 ± 0.29	40.17 ± 1.09	50.50 ± 0.33	9.74 ± 0.26	0.464 ± 0.004	27.00 ± 2.65
	100	36.17 ± 1.84	42.17 ± 2.17	52.94 ± 1.94	9.55 ± 0.35	0.477 ± 0.008	26.67 ± 5.03
	200	37.33 ± 2.03	45.00 ± 1.92	55.33 ± 1.36	9.32 ± 0.09	0.483 ± 0.002	26.00 ± 1.00
	400	36.00 ± 2.29	44.00 ± 2.50	53.67 ± 0.83	9.57 ± 0.03	0.481 ± 0.004	26.67 ± 3.51

Table 14: Codesignability by triangle-update interval. Mean ± standard deviation over three 600-RNA campaigns per setting; rates are percentages and scRMSD is in Å.

Model	Steps	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑	#Clust.↑
La-Ribo-Tri	50	28.61 ± 1.13	36.06 ± 1.93	50.83 ± 2.09	10.53 ± 0.28	0.449 ± 0.009	28.00 ± 1.00
	100	30.78 ± 1.26	39.00 ± 1.83	56.11 ± 1.95	9.86 ± 0.24	0.471 ± 0.008	25.33 ± 3.21
	200	32.06 ± 0.75	39.94 ± 1.26	57.89 ± 1.71	9.67 ± 0.11	0.477 ± 0.003	27.33 ± 0.58
	400	32.33 ± 0.58	40.17 ± 1.17	58.11 ± 0.84	9.55 ± 0.09	0.479 ± 0.003	28.00 ± 1.00
La-Ribo-Tri3	50	22.83 ± 0.93	31.50 ± 1.69	46.00 ± 1.89	11.32 ± 0.30	0.427 ± 0.009	25.33 ± 1.15
	100	26.67 ± 1.09	35.67 ± 1.69	51.17 ± 1.77	10.11 ± 0.25	0.461 ± 0.008	23.00 ± 3.00
	200	26.83 ± 0.60	36.50 ± 1.15	52.00 ± 1.53	10.30 ± 0.12	0.463 ± 0.003	24.67 ± 0.58
	400	28.67 ± 0.50	36.83 ± 1.09	51.67 ± 0.76	10.35 ± 0.09	0.467 ± 0.003	25.00 ± 1.00
La-Ribo-Tri4	50	22.00 ± 0.87	32.83 ± 1.76	48.83 ± 2.01	11.84 ± 0.32	0.436 ± 0.009	27.00 ± 1.00
	100	23.83 ± 0.98	34.50 ± 1.61	50.33 ± 1.76	11.45 ± 0.28	0.449 ± 0.008	22.67 ± 2.89
	200	25.17 ± 0.59	36.67 ± 1.21	52.67 ± 1.59	11.01 ± 0.13	0.458 ± 0.002	25.00 ± 1.00
	400	25.50 ± 0.44	37.00 ± 1.09	53.50 ± 0.76	10.89 ± 0.10	0.463 ± 0.003	25.67 ± 1.15

Across 50–400 steps, the two-block variant has higher strict codesignability than either sparser triangle-update variant. At 200 steps the gains are 5.23 and 6.89 percentage points. This pattern is stronger than the difference in strict designability. Base, reported in the main benchmark, remains competitive and shows that triangle refinement is not uniformly preferable for every endpoint.

E STRUCTURAL QUALITY AND NOVELTY OF GENERATED STRUCTURES

E.1 SECONDARY-STRUCTURE AGREEMENT

ViennaRNA RNAfold 2.4.14 (Lorenz et al., 2011) predicts a minimum-free-energy base-pair set from each sequence. We compare it with two annotations of the corresponding 3D structure: geometric contacts and DSSR 1.9.10 (Lu et al., 2015). Geometric AU, GC, and GU contacts require sequence separation greater than three residues and both diagnostic atom distances to be at most 4.0 Å: A(N6)–U(O4) and A(N1)–U(N3); G(N1)–C(N3) and G(O6)–C(N4); or G(N1)–U(O2) and G(O6)–U(N3),

respectively. DSSR pairs are restricted to canonical cis Watson–Crick/Watson–Crick AU/GC and GU wobble interactions. This assay uses no learned refolding model.

Table 15: RNAfold agreement with geometric-contact and DSSR annotations on RNASolo dataset and 600 La-Ribo-Tri structures. F1 and MCC are summarized per structure.

3D annotation	Cohort	Mean F1	Median F1	Mean MCC	Median MCC
Geometric contacts	Experimental	0.5815	0.6341	0.5727	0.6320
	La-Ribo-Tri	0.5510	0.6023	0.5465	0.6051
DSSR	Experimental	0.6451	0.7000	0.6367	0.7002
	La-Ribo-Tri	0.6056	0.6563	0.5919	0.6568

DSSR gives higher RNAfold agreement than geometric contacts in both cohorts (Table 15). Under either annotation, La-Ribo-Tri closely approaches the experimental reference, with mean F1 and MCC within 0.05 of the values for experimental structures.

E.2 Atomistic geometry

We next evaluate local atomistic geometry directly from the generated coordinates of 200-step 600-RNA campaigns of La-Ribo-Base and La-Ribo-Tri, with RNA-FrameFlow, RiboFlow-Mix, and the experimental and synthetic subsets of the training corpus as references. La-Ribo values are mean $\pm$ sample standard deviation over three independently seeded campaigns; corpus references are computed over all records of each subset.

Measured quantities. Clash counts steric violations per structure. Ribose pucker is assigned to C3′-endo, C2′-endo, or a remaining category covering all other conformations, and is reported as population fractions over valid residues. The glycosidic torsion χ is classified as *syn* or *anti*, and base-ring planarity is the RMSD of ring atoms from their least-squares plane. Contact distances are averaged over the diagnostic atom pairs of canonical AU, GC, and GU pairs, using the geometric criteria defined above. RNA-FrameFlow and RiboFlow-Mix generate backbone heavy atoms (frames plus torsion-reconstructed phosphate and sugar atoms, including the glycosidic nitrogen but no other base atoms), so χ, planarity, and contact distances, which require base atoms, are undefined for both; their clash counts cover these backbone-level outputs only, whereas the La-Ribo variants are evaluated on all heavy atoms.

Table 16: Atomistic geometry of generated structures at 200 steps against the experimental (Exp) and synthetic (Syn) subsets of the training corpus. La-Ribo values are mean $\pm$ sample standard deviation over three 600-RNA campaigns; pucker and χ are population percentages over valid residues; planarity and contact distances are in Å.

Metric	Exp	Syn	La-Ribo-Base	La-Ribo-Tri	RNA-FrameFlow	RiboFlow-Mix
Clash ↓	0.089	0.646	2.936 ± 0.200	1.846 ± 0.097	4.010 ± 0.261	6.214 ± 0.283
C3′-endo (%)	83.482	86.167	85.696 ± 0.061	84.744 ± 0.067	91.185 ± 0.094	83.890 ± 0.168
C2′-endo (%)	10.327	9.595	8.667 ± 0.058	9.648 ± 0.121	6.466 ± 0.007	2.636 ± 0.088
Other pucker (%)	6.191	4.238	5.636 ± 0.080	5.608 ± 0.179	2.349 ± 0.090	13.474 ± 0.200
χ *syn* (%)	4.358	3.136	3.592 ± 0.031	3.651 ± 0.026	–	–
χ *anti* (%)	95.642	96.864	96.408 ± 0.031	96.349 ± 0.026	–	–
Base-ring planarity RMSD	0.0039	0.0022	0.00113 ± 0.000	0.00113 ± 0.000	–	–
AU contact distance	3.009	2.996	3.100 ± 0.001	3.089 ± 0.001	–	–
GC contact distance	3.060	3.074	3.098 ± 0.003	3.095 ± 0.001	–	–
GU contact distance	3.102	3.162	3.151 ± 0.004	3.147 ± 0.001	–	–

Steric quality and ribose pucker. Both La-Ribo variants produce fewer clashes than either baseline, and La-Ribo-Tri reduces the clash count by 37% relative to La-Ribo-Base (Table 16); this ordering differs from strict codesignability, where Base is slightly ahead (Table 1). The three-way pucker populations of both La-Ribo variants track the experimental distribution in all categories. RNA-FrameFlow concentrates 91.19% of residues in C3′-endo against an experimental 83.48%, consistent

with a collapse toward A-form-like local geometry, while RiboFlow-Mix places almost no residues in C2′-endo (2.64% versus 10.33%) and 13.47% outside the two canonical pucker basins. Pucker is a nucleotide-level property that the scaffold anchors do not determine, so this agreement is attributable to the latent branch and the structured decoder (Appendix C.1).

Glycosidic conformation and base-pair geometry. The *syn* population is slightly below the experimental value (3.59–3.65% versus 4.36%), indicating mild under-sampling of *syn* conformations. Base-ring planarity sits below the corpus references (0.0011 versus 0.0022–0.0039), consistent with the idealized nucleotide templates of the atom builder. Mean AU, GC, and GU contact distances exceed the experimental values by 0.08–0.09, 0.03–0.04, and 0.05 Å, respectively, across the two variants; the offsets are one to two orders of magnitude larger than the across-campaign standard deviations, indicating a systematic tendency toward slightly loose pairing. The synthetic and experimental subsets themselves differ measurably in these quantities (e.g., 0.646 versus 0.089 clashes per structure), and synthetic structures vastly outnumber experimental ones in the corpus (Appendix A); the residual gap between generated and experimental geometry is therefore expected to partly reflect this distribution shift rather than model error alone.

E.3 Novelty relative to the training corpus

Our corpus assembles nearly all currently available high-quality RNA structures, combining experimentally determined entries with synthetic ones (Section 4.1 and Appendix A). Under this coverage, the usual novelty analysis, which asks whether a model generalizes beyond a standard held-out split drawn from the same data sources, is no longer meaningful: such a split would largely reproduce the training distribution rather than probe extrapolation beyond it. We therefore measure directly how close generated structures are to the training partition, and calibrate the measurement against experimental structures that are known to be held out.

Protocol. TM-score is a length-normalized measure of global structural similarity taking values in $[0, 1]$, with higher values indicating closer folds; we compute it with US-align (Zhang et al., 2022). For each generated structure, the maximum TM-score against the full training partition is denoted $\mathrm{TM}_{\mathrm{train}}$, and $\mathrm{TM}_{\mathrm{exp}}$ is the analogous maximum over all experimental structures in the corpus, excluding the query structure itself. Sequence identity and coverage are computed against the same nearest neighbour. A structure is counted as a *near-copy* when its nearest training neighbour satisfies $\mathrm{TM} \geq 0.8$, sequence identity ≥ 0.8, and coverage ≥ 0.8 in both directions. We evaluate 600-RNA campaigns from La-Ribo-Base, La-Ribo-Tri, and RiboFlow-Mix at 200 steps. Because absolute TM values are difficult to interpret in isolation at this length range, we apply the identical procedure to the 1,002 experimentally determined structures in the cluster-held-out validation partition (Appendix A), whose MMseqs2 clusters are disjoint from all training data; these structures bound what legitimate generalization looks like under this measurement.

Table 17: Proximity to the training partition at 200 steps. $\mathrm{TM}_{\mathrm{train}}$ and $\mathrm{TM}_{\mathrm{exp}}$ are maxima over the training partition and over all experimental structures, respectively. Near-copy requires $\mathrm{TM} \geq 0.8$, sequence identity ≥ 0.8, and coverage ≥ 0.8 in both directions against the nearest training neighbour. The validation cohort contains the 1,002 experimentally determined structures held out from all training data and serves as the reference cohort.

Cohort	Mean $\mathrm{TM}_{\mathrm{train}}$	$\mathrm{TM}_{\mathrm{train}} \geq 0.8$ (%)	Mean $\mathrm{TM}_{\mathrm{exp}}$	Near-copy (%)
Validation partition (held-out)	0.785	67.6	0.873	12.7
RiboFlow-Mix	0.541	20.6	0.573	14.7
La-Ribo-Base	0.623	31.1	0.625	14.9
La-Ribo-Tri	0.644	37.8	0.659	16.1

Proximity to the training partition. Held-out experimental structures have mean $\mathrm{TM}_{\mathrm{train}} = 0.785$, and 67.6% of them have a training neighbour at $\mathrm{TM} \geq 0.8$ (Table 17): sequence-level clustering does not produce structural separation at this length range. Both La-Ribo variants are substantially less similar to the training partition than real held-out structures are (31.1% and 37.8%), so the generated distributions are not concentrated on reproductions of training examples. The

generated cohorts also score similarly against experimental and training references, with gaps of $+0.002$, $+0.015$, and $+0.032$ for Base, Tri, and RiboFlow-Mix, and La-Ribo-Tri reaches the highest $\mathrm{TM}_{\mathrm{exp}}$ among them (Table 17). The near-copy rate additionally requires sequence identity of at least 0.8 to a training entry; the cluster-based split excludes cross-split matches above 90% sequence identity at 95% coverage, but not matches in the 80–90% range, so held-out structures can still satisfy this condition. Their near-copy rate (12.7%) is close to those of the generated cohorts (14.7–16.1%), so the near-copy assessment agrees with the structural columns above.

F Sampling sensitivity and computational cost

F.1 Modality-schedule sensitivity

We vary the scaffold and latent schedule powers independently to characterize sensitivity, separately from the fixed main-benchmark configuration. The balanced sensitivity collection covers four architectures, three step budgets (100, 200, 400), $\rho_B \in \{0.50, 0.70, 0.85, 1.00, 1.25\}$, and $\rho_Z \in \{1.00, 1.20, 1.40, 1.80, 2.00\}$. Across the complete balanced collection, every operating point contains 600 generated RNAs. The powers control progress along the two probability paths; they are schedule parameters, not multiplicative scales on the Gaussian noise.

The heatmaps visualize the observed response envelope across the sensitivity collection, while the marginal summaries average over the remaining factors. Neither display defines the main-benchmark protocol.

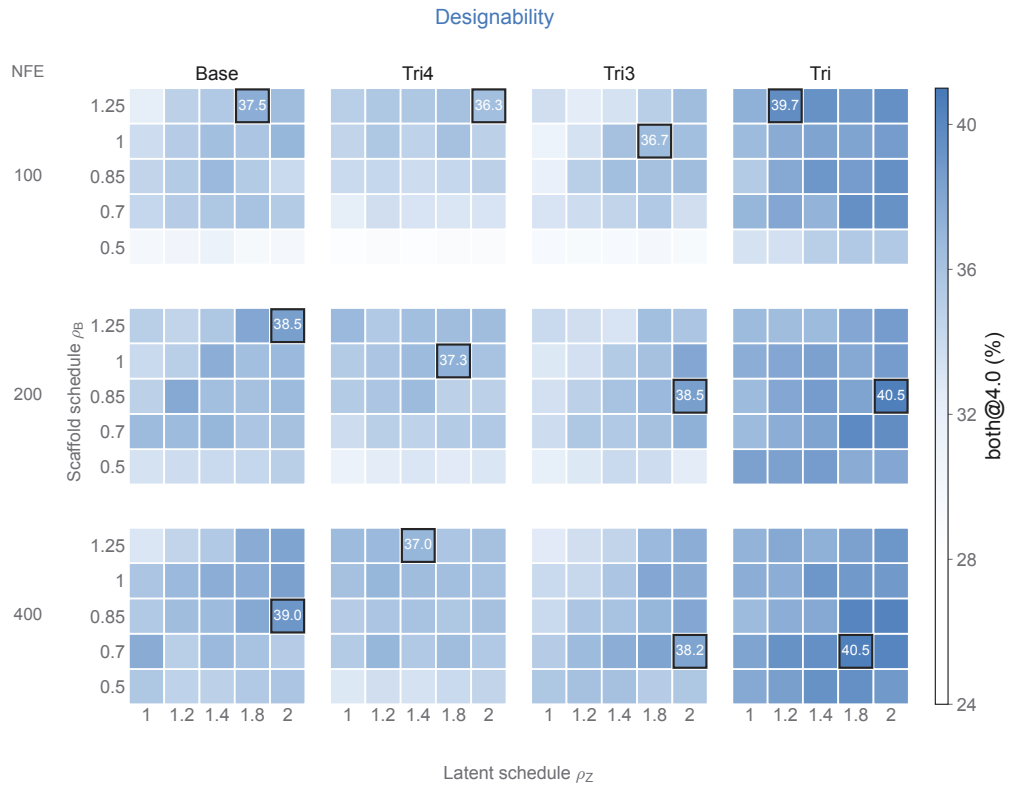


Figure 10: Strict designability across modality schedules and four architectures. Rows show 100, 200, and 400 steps; each schedule-pair cell contains 600 RNAs. Panels share a color scale, and black outlines mark maxima.

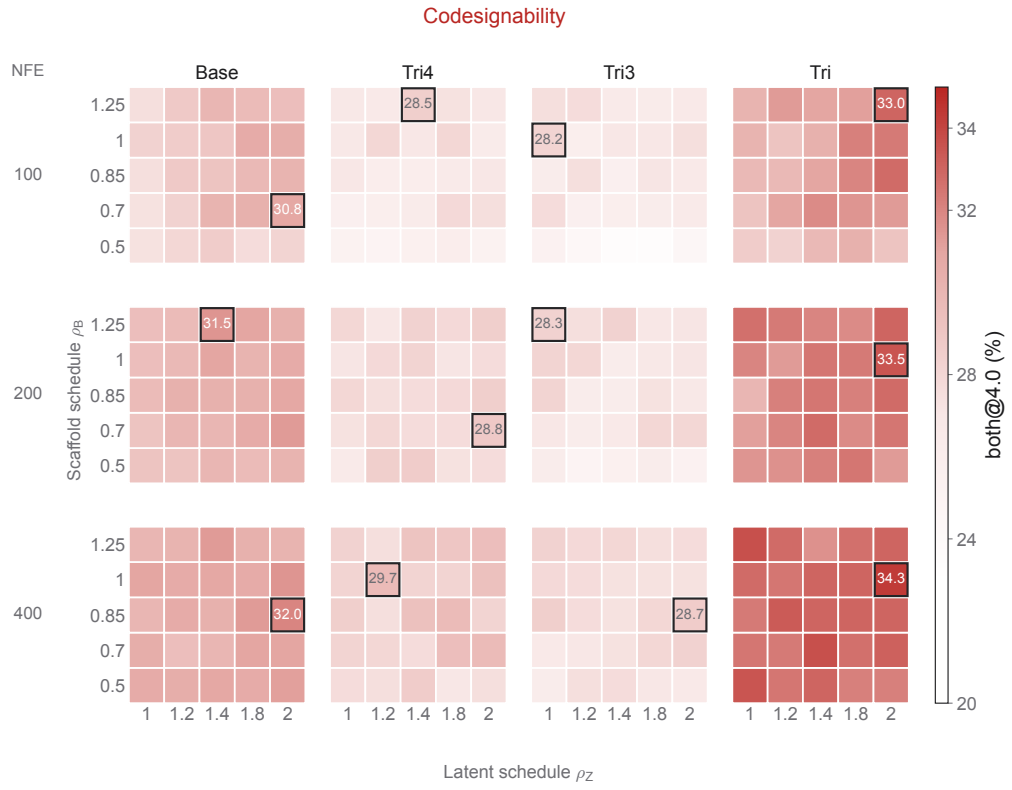


Figure 11: Strict codesignability on the schedule grid of Figure 10. Panels share a color scale, and black outlines mark maxima.

F.2 SAMPLING-FACTOR RESPONSE AND RUNTIME

We average each factor over the full balanced sensitivity collection and the remaining factor levels. Table 18 reports the architecture-averaged marginal responses. The architecture-specific curves are shown in main-text Figure 5, while the heatmaps above retain interactions averaged out in these summaries.

Table 18: Marginal strict both@4.0 success by sampling factor, averaged over the remaining grid settings. Each setting has 600 RNAs; rates are percentages.

Factor	Level	Designability ↑	Codesignability ↑
Solver steps K	100	31.90	25.01
	200	33.18	26.07
	400	33.84	26.67
Scaffold power ρ_B	0.50	31.03	25.08
	0.70	33.23	25.82
	0.85	33.48	26.04
	1.00	33.59	26.24
	1.25	33.53	26.40
Latent power ρ_Z	1.00	32.14	25.72
	1.20	32.69	25.77
	1.40	32.99	25.94
	1.80	33.47	26.06
	2.00	33.57	26.08

Increasing the budget from 100 to 400 steps changes either marginal strict endpoint by less than two percentage points. Sensitivity flattens at scaffold powers 0.70–1.25 and latent powers 1.80–2.00, although some architecture-specific maxima remain at the upper latent-power boundary. The sampled range thus contains a broad high-performing region without establishing an interior optimum.

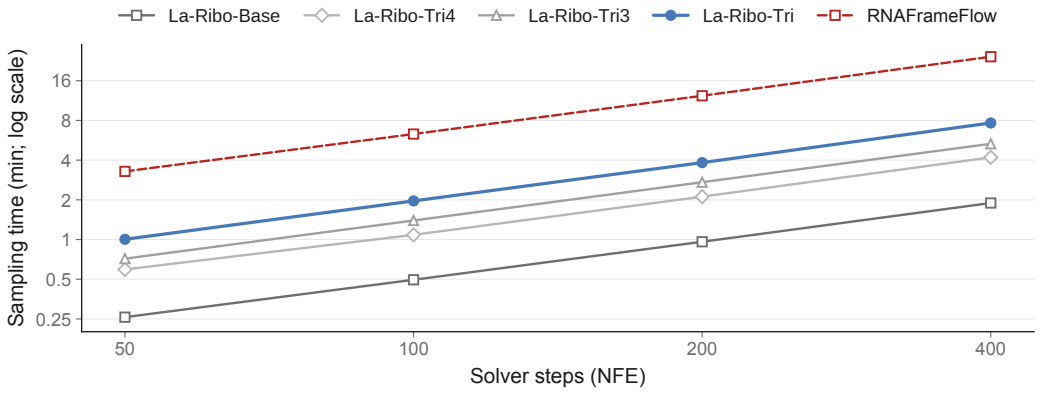


Figure 12: Generation time for 600 RNAs by solver steps, excluding inverse folding, refolding, and scoring. La-Ribo uses eight A800 GPUs; RNA-FrameFlow time is throughput-normalized. Timing scopes differ; steps are categorical and time is logarithmic.

Base has the lowest measured sampling cost; Tri costs approximately four times as much, with La-Ribo-Tri3 and La-Ribo-Tri4 between them. The measured RNA-FrameFlow pipeline costs about three times as much as Tri. Different timing scopes make these implementation-level comparisons rather than intrinsic algorithmic speedups.

G EVALUATION PROTOCOLS, ROBUSTNESS, AND INVERSE FOLDING

G.1 PRIMARY EVALUATION SETTINGS AND BASELINE IMPLEMENTATION

Unconditional generation. Each campaign contains 600 structures, with 50 samples at each length from 40 to 150 nt in increments of ten. Three independently seeded campaigns are evaluated for every main-benchmark model and sampling budget, and reported standard deviations summarize sampling variation of the fixed trained models. Designability uses eight inverse-folded candidate sequences per generated structure, whereas codesignability evaluates the emitted sequence. All reported generated structures are direct sampler outputs.

Inverse folding and refolding. We use the released gRNAde ARv1 single-state all-data weights at temperature 0.1. RhoFold receives each candidate sequence as its sole alignment input, without external MSA search, and its output coordinates are unrelaxed. These settings are shared across methods. Strict designability requires at least one of the eight candidates to satisfy both thresholds; continuous best-candidate metrics and cluster counts follow the same definitions throughout the expanded result tables.

Structural cluster count. For each 600-RNA campaign, the eligible set consists of generated backbones with scTM≥ 0.45 under the relevant evaluation route: the maximum across eight gRNAde sequences for designability, or the emitted sequence for codesignability. We cluster the original generated backbone PDBs, not the RhoFold refolds, with qTMclust (Zhang et al., 2022) (`-TMcut 0.45`) and count the output clusters as #Clust. We cluster the three campaigns independently and report the mean $\pm$ sample standard deviation of their counts. Failed or missing samples do not enter clustering but remain in the 600-sample success-rate denominator.

G.2 UNCONDITIONAL GENERATION WITH AN ALTERNATIVE REFOLDING MODEL

OpenDDE (Aureka AI OpenDDE project, 2026) replaces RhoFold while retaining each generated structure, candidate-sequence route, and score definition. Table 19 reports this evaluation. Both La-Ribo variants retain higher strict designability and codesignability than the corresponding baselines.

G.3 SCAFFOLD-CONDITIONED INVERSE-FOLDING PROTOCOL

Targets. The validation partition is held out from all training stages by sequence cluster (Appendix A); its complete experimental subset comprises 1,002 records (RNASolo and RhoFold-PDB chains) spanning 107 clusters, and every record serves as a scaffold target.

Inputs and scoring. La-Ribo fixes the centered P/C4′/N scaffold and its clean time endpoint, then integrates the latent branch for 200 steps. RiboFlow-Mix fixes native residue frames while evolving sequence states. Neither joint model receives the native sequence. Dedicated inverse folders, gRNAde and NA-MPNN, receive canonicalized native structures. Each method supplies eight candidates per target, for 8,016 candidate sequences. RhoFold refolding and C4′-based scoring use the same thresholds as the unconditional benchmark. Mean-of-eight metrics average candidate scores per target. Best-of-eight strict success requires a single candidate satisfying both thresholds. Best scRMSD and best scTM are optimized separately within a target and then averaged. All 1,002 targets and 8,016 candidate slots remain in success-rate denominators; nonfinite predictions count as failures. Continuous means use the finite predictions. Each method is evaluated with three independently seeded runs, and reported values are the mean $\pm$ sample standard deviation across runs. The complete inverse-folding comparison is reported in Table 4.

Table 19: Unconditional generation with OpenDDE instead of RhoFold under the settings of Table 1. Rates are percentages; scRMSD is in Å, and bold marks the best mean per panel.

(a) Designability

Method	Steps	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑	#Clust. ↑
RNA-FrameFlow	50	25.39 ± 1.64	30.22 ± 1.36	39.39 ± 0.35	12.33 ± 0.49	0.426 ± 0.006	31.67 ± 4.04
	100	27.06 ± 0.51	33.44 ± 0.92	42.83 ± 1.17	12.10 ± 0.21	0.434 ± 0.003	33.33 ± 4.04
	200	26.61 ± 1.58	33.56 ± 1.95	42.11 ± 3.06	12.28 ± 0.40	0.429 ± 0.009	35.00 ± 2.65
	400	27.50 ± 0.29	33.33 ± 0.87	41.56 ± 0.59	12.08 ± 0.31	0.429 ± 0.005	32.67 ± 2.08
RiboFlow-Exp	50	22.67 ± 0.67	27.67 ± 0.50	34.00 ± 1.20	12.29 ± 0.30	0.450 ± 0.002	26.33 ± 2.52
	100	22.94 ± 1.92	28.89 ± 2.20	36.11 ± 0.86	12.04 ± 0.35	0.456 ± 0.005	28.33 ± 2.89
	200	22.78 ± 0.38	28.56 ± 0.38	36.00 ± 0.33	12.29 ± 0.21	0.452 ± 0.003	26.33 ± 4.04
	400	23.33 ± 0.17	28.44 ± 0.42	34.83 ± 0.87	12.41 ± 0.31	0.447 ± 0.002	23.67 ± 0.58
RiboFlow-Mix	50	29.22 ± 0.61	32.28 ± 0.57	41.06 ± 0.93	11.11 ± 0.26	0.489 ± 0.003	30.00 ± 3.00
	100	29.89 ± 1.46	34.78 ± 1.94	42.56 ± 1.21	10.85 ± 0.31	0.478 ± 0.004	32.00 ± 2.65
	200	30.00 ± 1.15	34.17 ± 0.44	43.89 ± 0.82	10.92 ± 0.25	0.499 ± 0.004	30.33 ± 3.21
	400	30.50 ± 0.58	33.56 ± 0.35	41.67 ± 1.18	10.84 ± 0.34	0.483 ± 0.003	27.33 ± 1.15
La-Ribo-Base	50	34.33 ± 1.64	45.78 ± 1.06	58.22 ± 0.42	9.61 ± 0.33	0.492 ± 0.007	32.00 ± 3.61
	100	33.11 ± 2.20	45.11 ± 0.35	56.72 ± 0.54	9.70 ± 0.34	0.490 ± 0.006	29.67 ± 2.08
	200	34.11 ± 0.79	46.89 ± 1.34	59.44 ± 1.84	9.56 ± 0.31	0.495 ± 0.009	31.33 ± 1.53
	400	33.78 ± 1.84	46.44 ± 1.78	60.06 ± 2.11	9.45 ± 0.42	0.495 ± 0.008	33.00 ± 2.00
La-Ribo-Tri	50	37.50 ± 1.74	46.72 ± 0.67	59.67 ± 0.44	8.93 ± 0.37	0.517 ± 0.001	41.33 ± 3.06
	100	39.72 ± 2.56	49.50 ± 2.20	63.50 ± 1.01	**8.56** ± 0.50	0.531 ± 0.008	41.00 ± 2.65
	200	38.33 ± 0.76	49.72 ± 0.42	63.94 ± 0.54	8.82 ± 0.10	0.530 ± 0.004	39.67 ± 1.53
	400	**39.83** ± 1.00	**50.28** ± 0.67	**65.22** ± 0.67	**8.56** ± 0.05	**0.537** ± 0.001	**43.33** ± 8.02

(b) Codesignability

Method	Steps	both@4.0 ↑	both@5.5 ↑	TM@0.45 ↑	scRMSD ↓	scTM ↑	#Clust. ↑
RiboFlow-Exp	50	20.89 ± 0.35	25.72 ± 1.64	30.50 ± 1.59	15.44 ± 0.09	0.431 ± 0.005	18.00 ± 2.00
	100	20.61 ± 0.86	26.11 ± 0.84	32.00 ± 0.76	15.12 ± 0.38	0.437 ± 0.004	20.33 ± 1.15
	200	20.72 ± 1.84	25.89 ± 1.08	32.17 ± 0.29	15.25 ± 0.09	0.433 ± 0.001	19.33 ± 0.58
	400	21.11 ± 1.93	26.89 ± 0.42	32.39 ± 0.67	15.07 ± 0.39	0.436 ± 0.001	18.33 ± 3.79
RiboFlow-Mix	50	27.67 ± 1.66	31.83 ± 1.52	38.11 ± 0.86	14.19 ± 0.12	0.479 ± 0.004	21.67 ± 2.31
	100	27.50 ± 1.36	30.56 ± 0.91	38.72 ± 0.19	14.11 ± 0.33	0.458 ± 0.005	24.00 ± 1.73
	200	27.67 ± 1.34	29.06 ± 1.16	39.22 ± 1.35	13.79 ± 0.11	0.435 ± 0.002	23.00 ± 1.00
	400	27.78 ± 0.79	32.00 ± 0.47	38.89 ± 0.93	13.14 ± 0.42	0.491 ± 0.002	22.00 ± 3.00
La-Ribo-Base	50	31.17 ± 2.89	40.72 ± 2.38	56.28 ± 1.84	10.76 ± 0.31	0.500 ± 0.011	30.33 ± 0.58
	100	32.00 ± 1.42	42.78 ± 2.20	58.22 ± 0.75	10.58 ± 0.51	0.508 ± 0.010	32.00 ± 3.00
	200	32.50 ± 1.26	44.44 ± 2.01	59.78 ± 1.99	10.45 ± 0.52	0.514 ± 0.012	32.33 ± 1.53
	400	**33.72** ± 2.26	**44.78** ± 1.70	58.83 ± 1.73	**10.44** ± 0.25	0.512 ± 0.007	29.00 ± 1.00
La-Ribo-Tri	50	28.33 ± 1.50	36.94 ± 1.29	56.11 ± 2.12	11.73 ± 0.25	0.501 ± 0.007	40.00 ± 2.00
	100	29.39 ± 0.95	38.39 ± 0.86	58.33 ± 1.04	11.15 ± 0.30	0.518 ± 0.006	36.67 ± 3.79
	200	33.33 ± 2.05	41.00 ± 2.20	60.22 ± 1.78	11.17 ± 0.14	0.522 ± 0.007	40.00 ± 2.65
	400	32.06 ± 1.23	41.56 ± 1.42	**61.00** ± 2.42	11.27 ± 0.05	**0.526** ± 0.004	**41.00** ± 4.36

H LIMITATIONS AND BROADER IMPACT

Limitations. Experimentally determined RNA structures are scarce, so most of our corpus consists of predicted or synthetic structures, whose atom-level geometry does not fully match that of experimentally resolved structures (Appendix E.2). This distribution gap limits the atomistic accuracy that models trained on this corpus can attain. We expect it to gradually narrow as structure prediction models continue to improve. Besides, long RNAs remain difficult because high-quality training structures at longer lengths are scarce in both experimental and synthetic sources. Only 0.31% of the corpus exceeds 160 nt: the experimental RNASolo and RhoFold subsets allocate just 1.18% and 1.90% of their entries to this range, and OpenFold3, which dominates the corpus, contributes only 0.06% (Figure 8). Consistent with this thinning coverage, strict success declines toward the upper

end of the 40–150-nt benchmark range (Figure 3). Because the missing supervision is a property of the available data rather than of the model, we expect this gap to persist until larger collections of high-quality long-RNA structures, experimental or predicted, become available.

Broad impact. La-Ribo offers a framework for joint RNA sequence and structure generation. We hope it will support future RNA research by enabling studies of sequence–structure relationships and informing new RNA design methods. Functional applications should undergo experimental validation and biosafety review because generated sequences may have unintended biological effects.