RNA-FRAMEFLOW for *de novo* 3D RNA Backbone Design

Anonymous Authors[1]

Abstract

We introduce RNA-FRAMEFLOW, the first generative model for 3D RNA backbone design. We build upon $SE(3)$ flow matching for protein backbone generation and focus on establishing RNA-specific data augmentations and evaluation protocols. Our formulation of rigid-body *frames* and loss functions account for larger, more conformationally flexible RNA backbones (13 atoms) vs. proteins (4 atoms). Towards tackling the lack of diversity in 3D RNA datasets, we explore training with structural clustering and cropping augmentations. Additionally, we define a suite of *in silico* evaluation metrics to measure whether designed RNAs are globally self-consistent (via inverse folding followed by forward folding) and locally recover RNA-specific structural descriptors. The most performant version of RNA-FRAMEFLOW generates locally realistic backbone structures of 40-150 nucleotides that are 41% globally self-consistent on average ($\texttt{scTM} \geq 0.45$), with fast sampling speeds of $\sim$4 seconds per backbone.

1. Introduction

Why RNA design? Proteins, and the diverse structures they can adopt, drive all essential biological functions in cells. Deep learning has led to breakthroughs in structural modeling and design of proteins (Jumper et al., 2021; Dauparas et al., 2022; Watson et al., 2023; Abramson et al., 2024), driven by the abundance of 3D data from the Protein Data Bank (PDB) (Khakzad et al., 2023). Concurrently, we have witnessed a surge in interest in *Ribonucleic Acids* (RNA) and RNA-based therapeutics like CRISPR and mRNA vaccines (Doudna and Charpentier, 2014; Metkar et al., 2024). RNAs play a dual role as carriers of genetic information coding for proteins (mRNAs) as well as performing functions driven by tertiary structural interactions (riboswitches and ribozymes). While there is growing interest in designing structured RNAs for biotechnology and medicine (Damase et al., 2021), the current toolkit for 3D RNA design uses classical algorithms and heuristics to assemble RNA motifs as building blocks (Han et al., 2017; Yesselman et al., 2019). The use of hand-crafted heuristics and motifs may not fully capture the geometric interactions and conformational dynamics that govern RNA functionality. This presents an exciting opportunity for generative models to go beyond human intuition by learning from existing 3D RNA structures in the Protein Data Bank (PDB).

Challenges of RNA modeling. The primary challenge for deep learning on RNA structure is the paucity of raw 3D structural data, underscoring the absence of ML-ready datasets for model development. Protein structure is primarily driven by hydrogen bonding along the backbone, and current geometric deep learning models incorporate this inductive bias through backbone *frames* to represent residues (Yim et al., 2023a; Jumper et al., 2021). RNA structure, however, is more conformationally flexible and is driven by *base pairing* and *base stacking* interactions across strands (Vicens and Kieft, 2022). Additionally, the RNA equivalent of amino acids – nucleotides – include significantly more atoms (13 compared to 4) which necessitates a generalization of backbone frames where the placement of most atoms is parameterized by torsion angles. These complexities manifest as poor performance in RNA representation learning pipelines like AlphaFold3 (Abramson et al., 2024) and earlier deep learning methods (Kretsch et al., 2023).

Contributions. We introduce RNA-FRAMEFLOW, the first geometric generative model for 3D RNA backbone design. We adapt FrameFlow (Yim et al., 2023b), an $SE(3)$ equivariant flow matching model for proteins, and represent RNA nucleotides as 3D rigid-body *frames* that parameterize all 13 atoms. Alongside RNA-specific modifications to the data preparation and loss formulations for FrameFlow, we develop an evaluation pipeline to benchmark RNA backbone design models' capabilities at recovering local and global structure. Our best model is trained on RNA of lengths 40-150 from the PDB and can unconditionally sample locally plausible backbones with $\sim$41% self-consistency. We hope our engineering contributions will make deep learning for 3D RNA design and its evaluation more broadly accessible.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review at the ICML 2024 *AI for Science* workshop. Do not distribute.

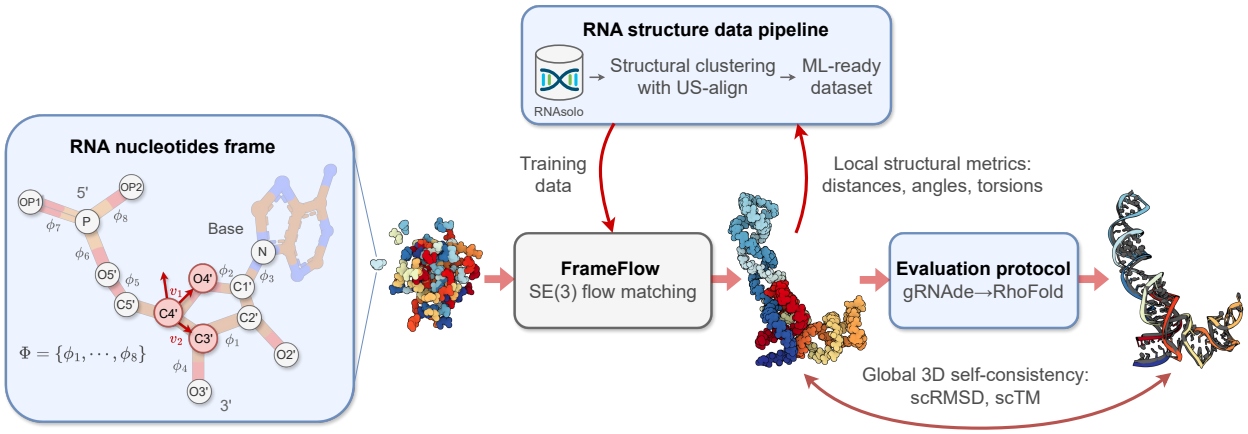

Figure 1. **The RNA-FRAMEFLOW pipeline for 3D backbone generation.** Our implementation builds upon FrameFlow (Yim et al., 2023b) and establish RNA-specific data preparation and evaluation protocols. (1) Each nucleotide in the RNA backbone is converted into a *frame* to parameterise the placement of $C4'$ by a translation, $C3' - C4' - O4'$ by a rotation, and the rest of the atoms via 8 torsion angles Φ. (2) We train generative models on all RNA structures in the PDB of length 40-150 nucleotides. We also explore training with structural clustering and cropping augmentations. (3) We establish evaluation protocols for measuring the recovery of local structural metrics as well as global self-consistency of designed structures via inverse folding followed by forward folding.

2. Method

We are concerned with building a generative model that unconditionally outputs all-atom RNA backbone samples. For a target sequence length of N_{res} nucleotides, we aim to generate a real-valued tensor $\mathbf{X}$ of shape $N_{\text{res}} \times 13 \times 3$ representing 3D atomic coordinates for each of the 13 backbone atoms per nucleotide.

2.1. Representing RNA Backbones

This work introduces a frame analog for nucleic acids that deals with the underlying complexity of working with their backbones. Unlike protein residues with just 4 atoms in the backbone, nucleic acid residues contain 12 atoms along the backbone. As shown in Figure 1, we use the $C4'$, $C3'$, and $O4'$ atoms as the reference frame as done by Morehead et al. (2023). All other backbone atoms are associated with 8 torsions $\Phi = \{\phi_1 \rightarrow \phi_8\}, \phi_i \in SO(2)$ that are predicted post-hoc after frame generation; these atoms are $C1', C2', N9$ (or $N1$), $O3', O5', P, OP1$, and $OP2$.

The Gram-Schmidt process is used on v_1, v_2 defined by the vectors along the $C4' - O4'$ and $C4' - C3'$ bonds; the $C5'$ plays the role of C_β and is imputed after the frames have been created. The 8 torsions, in order, are $C3' - C2'$, $O4' - C1'$, $C1' - N9$, $C3' - O3'$, $C5' - O5'$, $O5' - P$, $P - OP1$, and $P - OP2$. During sampling, *Adenine* (A) is the default nucleic acid base whose idealized geometry is obtained from Gelbin et al. (1996). Geometric imputation is not tractable given the complexity of torsion angles. Given the torsion angles, we autoregressively place non-frame atoms in order of the torsions Φ in Figure 1, constructing the final all-atom RNA backbone structure.

Choice of RNA frame. We choose $C3'$, $C4'$, and $O4'$ as they spatially shift the least in naturally occurring RNA (Harvey and Prabhakaran, 1986). The non-frame backbones - such as the remaining atoms in the ribose sugar ring ($C1'$, $C2'$) and the farther away Phosphorous (P) and two Oxygens ($OP1, OP2$) - are parameterized by torsion angles to account for their relative conformational flexibility. *Ring puckering* refers to the planar rotation of the ribose sugar ring about the $C4' - C5'$ bond. It affects how the RNA interacts with interaction partners to form complexes (e.g., protein-RNA) with high binding affinities (Clay et al., 2017). To reduce the sensitivity of our generative model to highly mobile atoms, we settle on the current frame composition.

2.2. $SE(3)$ Flow Matching on RNA Frames

A simultaneous rotation and translation (r, x) forms an orientation-preserving rigid-body transformation. The set of all such transformations in 3 dimensions is the Special Euclidean group $SE(3)$. For RNA and proteins, a frame $T = (r, x)$ exists in this $SE(3)$ group as they too can be decomposed into an absolute rotation and translation from the global origin. Formally, $SE(3) \cong SO(3) \ltimes \mathbb{R}^3$. Compared to expensively generating raw Cartesian coordinates, generating new RNA backbones using the frame representation entails performing flow matching on the space of $SE(3)$. Fortunately, Chen and Lipman (2024) and Yim et al. (2023b) provide the necessary ingredients to perform flow matching on $SE(3)$. Specifically, we can decompose a frame $T \in SE(3)$ into a rotation $r \in SO(3)$, the group of orientation-preserving rigid-body rotations, while a translation x trivially belongs to $\mathbb{R}^3$.

Training. Given a random frame $T_0 \sim p_0(T_0)$ and ground truth frame $T_1 \sim p_1(T_1)$ from the target distribution, we construct the conditional flow T_t by following the *geodesic* between T_0 and T_1; this geodesic generalizes the linear interpolation on manifolds (like the $SE(3)$ group) as shown by Chen and Lipman (2024):

$$T_t = \exp_{T_0}(t \cdot \log_{T_0}(T_1)). \tag{1}$$

Here, $\exp(\cdot)$ and $\log(\cdot)$ are the *exponential* and *logarithmic* maps that enable taking random walks on a manifold (like $SE(3)$). Conveniently, by decomposing a frame $T = (r, x)$ into separate rotation and translation terms, we can obtain closed-form geodesics for $SO(3)$ and $\mathbb{R}^3$. Sampling $t \sim \mathcal{U}(0, 1)$, this gives us two independent flows for rotations and translations:

$$\text{Translations: } x_t = tx_1 + (1-t)x_0 \tag{2}$$

$$\text{Rotations: } r_t = \exp_{r_0}(t \cdot \log_{r_0}(r_1)). \tag{3}$$

The random translation x_0 is sampled from $\mathbb{R}^3$ and random rotation r_0 is sampled from $\mathcal{U}(SO(3))$, a generalization of the uniform distribution for the group of rotations, $SO(3)$. The logarithmic and exponential maps for r_t help perform random walks for rotations in the $SO(3)$ space and are trivial to compute using Rodrigues' formula. We can do this parallelly for a set of RNA frames $\mathbf{T} = \{T_1, \cdots, T_N\}$ to get the conditional flow $\mathbf{T}_t$.

Suppose $\{(\hat{r}_t, \hat{x}_t)\}_{n=1}^N = v_\theta(\mathbf{T}_t, t)$ are the predicted frames in $\mathbf{T}_t$. The ground truth vector field u_t can also be decomposed into rotation and translation:

$$\text{Translations: } u_t(x^{(n)} | x_0^{(n)}, x_1^{(n)}) = x_1^{(n)} \tag{4}$$

$$\text{Rotations: } u_t(r^{(n)} | r_0^{(n)}, r_1^{(n)}) = \log_{r_t^{(n)}}(r_1^{(n)}). \tag{5}$$

We can now compute separate losses on $SO(3)$ and $\mathbb{R}^3$. Optimizing the following objective allows us to train the flow matching model on $SE(3)$:

$$\mathcal{L}_{SE(3)} = \mathbb{E}_{t, p_0(\mathbf{T}_0), p_1(\mathbf{T}_1)} \left[\frac{1}{(1-t)^2} \sum_{n=1}^{N} \left\{ \left\| \hat{x}_t^{(n)} - x_1^{(n)} \right\|_{\mathbb{R}^3}^2 + \left\| \log_{r_t^{(n)}}(\hat{r}_1^{(n)}) - \log_{r_t^{(n)}}(r_1^{(n)}) \right\|_{SO(3)}^2 \right\} \right]. \tag{6}$$

Sampling. For an RNA sequence of length N, we initialize a random point cloud of frames $\{T_i\} \in SE(3)^N$ oriented and spatially placed by the constituent random rotations and translations. We integrate from $t = 0.0$ to $t = 1.0$ using an ODE solver for N_T steps. Here, we use an Euler solver to predict $\mathbf{T}_1 = \{(r_1^{(n)}, x_1^{(n)})\}_{n=1}^N$ as $\tilde{\mathbf{T}}_1 = (\tilde{r}_1, \tilde{x}_1) = v_\theta(\mathbf{T}_t, t)$. Given a specific frame, we compute the next translation $x_{t+\Delta t} = x_t + \Delta t \cdot (\tilde{x}_1 - x_t)$. Likewise, we use the geodesic to compute the next rotation $r_{t+\Delta t} = \exp_{r_t}(ct \cdot \log_{r_t}(\tilde{r}_1))$, where c is a tunable hyperparameter governing the exponential sampling schedule.

2.3. Architecture

Following Yim et al. (2023b;a), we use the structure module from AlphaFold2 (Jumper et al., 2021) comprising Invariant Point Attention (IPA). We also use an auxiliary MLP head to predict torsion angles Φ. We provide hyperparameters, objective functions, and additional experimental setup details in Appendix A.1.

2.4. Objective

To accommodate several moving parts in $SE(3)$ flow matching, we use the following multi-loss:

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{SO(3)} + \mathcal{L}_{\mathbb{R}^3} + \mathcal{L}_{\text{aux}}, \tag{7}$$

$$\text{where } \mathcal{L}_{\text{aux}} = \mathcal{L}_{\text{bb}} + \mathcal{L}_{\text{dist}} + \mathcal{L}_{\text{tors}}. \tag{8}$$

Each loss term is weighted by a tunable scalar hyperparameter, influencing its contribution to the total loss. $\mathcal{L}_{SO(3)}$ and $\mathcal{L}_{\mathbb{R}^3}$ are the primary losses coming from $SE(3)$ Flow Matching as described in Section 2.2. Additionally, $\mathcal{L}_{\text{aux}}$ is an auxiliary multi-part loss that operates on the all-atom structure extracted from the generated frames. We summarise all our losses as follows. Suppose $S = \{C4', C3', O4'\}$ is the set of frame atoms and $a, b \in S$ are ground truth atomic coordinates,

- $\mathcal{L}_{\text{bb}}$: A direct all-atom MSE computed between generated and ground truth coordinates of atoms in the frame. Later, in Appendix B.1, we ablate whether including additional *anchor* atoms to account for the larger size of RNA nucleotides. Given a generated coordinate $\hat{a}$,

$$\mathcal{L}_{\text{bb}} = \frac{1}{kN} \sum_{i=1}^{N} \sum_{a \in S} \|a_i^{(0)} - \hat{a}_i^{(0)}\|^2. \tag{9}$$

- $\mathcal{L}_{\text{dist}}$: A pairwise distance loss computed between ground truth and generated coordinates. First, *all-to-all* coordinate differences are computed between ground truth and generated structures before taking another difference between the two pairwise difference tensors. Let $d_{ab}^{(ij)} = \|a_n^{(1)} - b_m^{(1)}\|$ represents this ground truth distance between a, b. Given a generated distogram inter-residue loss $\hat{d}_{ab}^{ij}$,

$$\mathcal{L}_{\text{bb}} = \frac{1}{|S| \cdot N} \sum_{i=1, j=1}^{N} \sum_{a,b \in S} \|d_{ab}^{(ij)} - \hat{d}_{ab}^{(ij)}\|^2. \tag{10}$$

- $\mathcal{L}_{\text{tors}}$: A angular loss on the 8 predicted torsions by the auxiliary MLP head. This enables supervision on

the angles from the ground truth computed from the all-atom structure. Suppose $\phi_j \in \Phi_i$ is the ground truth torsion angles for a residue i, we wish to ensure the angles are close to the unit circle while being close to the ground truth angles. Given a set of generated angles $\hat{\Phi}_i$,

$$\mathcal{L}_{\text{tors}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{\phi_j\in\Phi_i}\Big(\|\phi_{ij}-\hat{\phi}_{ij}\|^2 + \|\|\phi_{ij}\|-1\|\Big). \tag{11}$$

3. Experiments

3D RNA Dataset. RNAsolo (Adamczyk et al., 2022) is a recent dataset containing a diverse range of RNA sequences of varying lengths and their corresponding 3D structures extracted from isolated RNAs, protein-RNA complexes, and DNA-RNA hybrids from the Protein Data Bank (PDB), an online repository of proteins and related biomolecules. The dataset contains 14,366 samples (structure and sequence) available at a resolution $\leq$4Å (1Å = 0.1nm). From RNAsolo, we select sequences of lengths between 40 and 150 nucleotides (5,319 in total) as they have prominent, well-folded tertiary structures than smaller sequences with relatively disordered folds (Boivin et al., 2019).

Evaluation. Following the protein backbone design literature (Yim et al., 2023a;b; Lin and AlQuraishi, 2023), we generate 50 backbones for target lengths uniformly sampled between 40 and 150. We then compute three indicators of quality for these backbones:

- **Validity** ($\texttt{scTM} \geq 0.45$): We inverse fold each generated backbone using gRNAde (Joshi et al., 2023) and pass $N_{\text{seq}} = 8$ generated sequences into RhoFold (Shen et al., 2022). We then compute the self-consistency TM-score (scTM) between the predicted RhoFold structure and our backbone at the $C4'$ level. We say a backbone is *valid* if $\texttt{scTM} \geq 0.45$; this threshold corresponds to roughly the same fold between two RNA strands (Zhang et al., 2022). We expand on this framework in Figure 4.
- **Diversity**: Among the valid samples, we compute the number of unique structural clusters formed using qTMclust (Zhang et al., 2022) and take the ratio to the total number of generated samples. Two structures are considered similar if their TM-score ≥ 0.45. This metric shows how much each generated sample varies from others across various sequence lengths.
- **Novelty**: Among the valid samples, we use US-align (Zhang et al., 2022) to compute how structurally dissimilar the generated backbones are from the training distribution. For every generated backbone, we compute the TM score to every training sample, and then take the highest average - a metric we call pdbTM, aligning with the protein design literature.
- **Structural Measurements**: We measure bond distances, bond angles, and torsion (dihedral) angles from the generated samples, and then match this to empirical distributions from our training dataset and idealized geometries from Gelbin et al. (1996).

On 3D self-consistency. The protein design community has broadly used self-consistency as a proxy for experimental success; popularly, this metric is called *designability* with a sample being designable (i.e., there exists a string of amino acids that yield that fold) if self-consistency RMSD (scRMSD) is below a threshold, typically 2 Å. We move away from scRMSD as used by Yim et al. (2023b;a); Watson et al. (2023); Lin and AlQuraishi (2023) to accommodate the high conformational complexity of RNA, where strict global alignments may penalize otherwise realistic backbone samples.

Training. Our filtered training dataset with sequences of lengths between 40 and 150 consists of 5,319 samples. For the denoiser, we use 6 IPA blocks and an additional torsion predictor head, the latter being a 3-layer MLP that takes in node embeddings from the IPA module to predict 8 torsion angles. Our final model contains 16.8M trainable parameters. We use the Adam optimizer with a learning rate of 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$. We train for $120K$ gradient update steps on four NVIDIA GeForce RTX 3090 GPUs for 15 hours with a batch size of $B = 20$. Each batch comprises padded samples from randomly selected structural clusters across sequence lengths.

4. Results

4.1. Global Evaluation of Generated RNA Backbones

We begin by analyzing RNA-FRAMEFLOW's samples using the aforementioned evaluation metrics. For validity, we report percentage of samples with $\texttt{scTM} \geq 0.45$; for diversity, we report the ratio of unique structural clusters to total **valid** samples; and for novelty, we report the highest pdbTM to a match from the PDB. For each sequence length between 40 and 150, at intervals of 10, we generate 50 backbones. Table 1 reports these metrics across different variants for the number of denoising steps N_T. We compare our model to protein-RNA-DNA complex co-design model MMDIFF (Morehead et al., 2023). As the original version of MMDIFF was trained on shorted RNA sequences, we retrain it on our sequence length split of RNAsolo. Additionally, we inverse folded MMDIFF's backbones using gRNAde.

We identify $N_T = 50$ as the best-performing model that balances validity, diversity, and novelty; furthermore, it takes 4.74 seconds (averaged over 5 runs) to sample a backbone of length 100, as opposed to 27.3 seconds for MMDIFF with

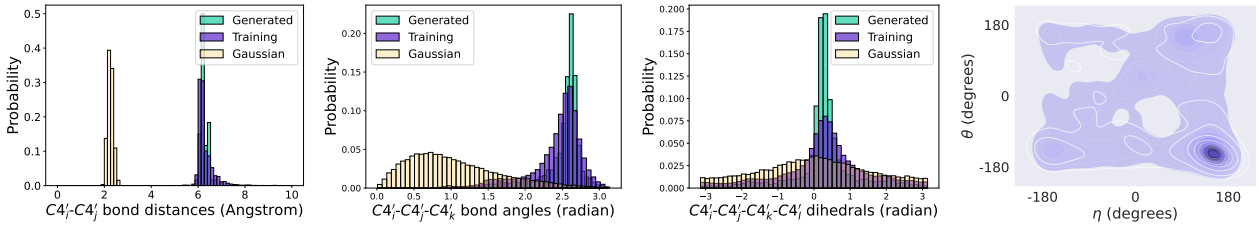

Figure 2. **Structural measurements** from 600 generated backbone samples compared to random Gaussian point cloud as a sanity check. Our model can recapitulate local structural descriptors. **(Subplots 1-3)** Histograms of inter-nucleotide bond distances, bond angles between nucleotide triplets, and torsion angles between every four nucleotides. **(Subplot 4)**: RNA-centric Ramachandran plot of structures from RNAsolo (purple) and generated backbones (white).

MODEL	% VALID ↑	DIVERSITY ↑	NOVELTY ↓
$N_T = 50$	**41.0**	0.606	**0.541**
$N_T = 10$	16.7	**0.62**	0.70
$N_T = 100$	20.0	0.61	0.69
$N_T = 500$	20.0	0.57	0.67
MMDIFF	0.0	-	-

Table 1. **RNA backbone generation results**. The best performing model uses $N_T = 50$ timesteps for denoising.

MODEL	EMD ↓ (DIST)	EMD ↓ (ANGLES)	EMD ↓ (TORSIONS)
$N_T = 50$	**0.167**	**0.11**	**2.36**
MMDIFF (ORIGINAL)	1.38	0.43	3.06
MMDIFF (RNASOLO)	0.396	0.21	3.23
GAUSSIAN	29.00	6.35	4.37

Table 2. Ablations of loss terms on Earth Mover's Distance scores for structural measurements compared to ground truth measurements from RNAsolo. The first row corresponds to the baseline model. Our model can recapitulate local structural descriptors and achieves the best EMD scores.

100 diffusion steps. We note that increasing N_T does not improve validity despite allowing the model to perform more updates to atomic coordinate placements. Our model also out-performs MMDIFF. On manual inspection, samples from MMDIFF had significant chain breaks and disconnected floating strands; see Appendix C.1.

4.2. Local Evaluation with Structural Measurements

For our best-performing model at $N_T = 50$, histograms of bond distance, bond angles, and torsion angles are in Figure 2. We include the *Earth Mover's Distance* (EMD) between measurements from the training and generated distributions as an indicator of local realism (using 30 bins for each quantity). An ideal generative model will score an EMD of 0 across all categories (i.e., consistent with the training set comprising naturally occurring RNA). In Table 2, we observe EMD values from our best-performing $N_T = 50$ model's backbones being significantly closer to 0 compared to MMDIFF and random Gaussian all-atom point clouds (akin to an untrained model) which serve as sanity checks. We include histograms for MMDIFF in Appendix C.1.

For a nucleotide i along the generated chain, we compare the distribution of dihedrals using Ramachandran plots for RNA. (Keating et al., 2011) introduce $\eta - \theta$ plots that track the separate dihedral angles formed by $\{C4'_i, P_{i+1}, C4'_{i+1}, P_{i+2}\}$ and $\{P_i, C4'_i, P_{i+1}, C4'_{i+1}\}$ respectively. We show these RNA Ramachandran plots in Figure 2 and observe that RNA-FRAMEFLOW can recapitulate the underlying torsional distribution.

4.3. Generation Quality Across Sequence Lengths

We next investigate how sequence length affects the global realism of generated samples (`scTM`). Figure 3 **(Left)** shows RNA-FRAMEFLOW performs for different sequence lengths. We observe our model generates samples with high `scTM` for specific sequence lengths like 50, 60, 70, and 120 while generating poorer quality structures for other lengths. We partially attribute the heavy fluctuation of TM-scores to the inherent length bias of RhoFold; see Appendix A.2. With a better structure predictor, we expect an increase in valid samples that meet the 0.45 TM-score threshold.

We also analyze the novelty of our generated samples (`pdbTM`) in Figure 3 **(Middle)**. We are particularly interested in samples that lie in the right half with high `scTM` and low `pdbTM`. We observe that RNAsolo has a high composition of samples with high structural similarity; running `qTMclust` (Zhang et al., 2022) on our filtered training dataset from RNAsolo reveals only 342 unique clusters from 5,319 samples, which indicates that the model does not encounter a diverse set of samples during training. This results in many generated samples looking similar to the training distribution (for instance, the `pdbTM` ≈ 0.9 and `scTM` ≈ 0.9 for samples of length 120, indicating close likeness to existing RNA). We include two such examples in Figure 3 **(Right)**: both yield relatively high `pdbTM` scores and look similar to their respective closest matching chain from RNAsolo. Similarly, we include figures on validity and novelty for MMDIFF's samples in Appendix C.1

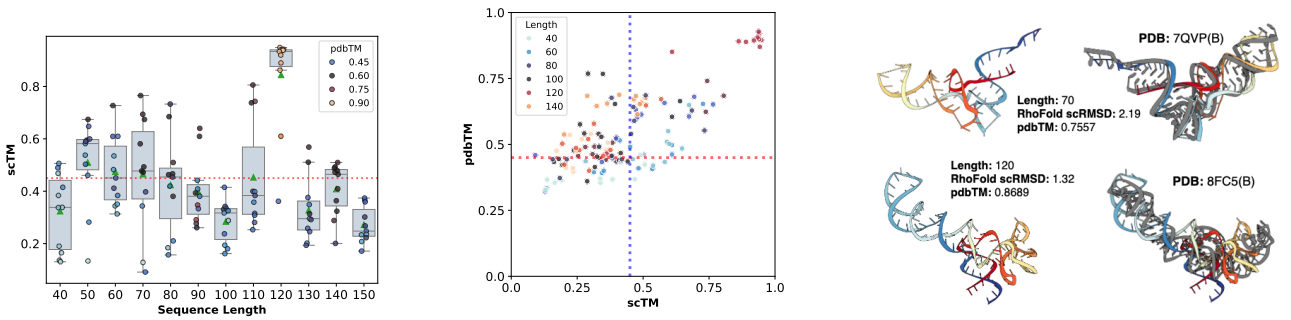

Figure 3. **Validity and novelty of generated backbones. (Left)** `scTM` of backbones of lengths 40-150 with the mean and spread of `scTM` for each length. **(Middle)** Scatter plot of self-consistency TM-score (`scTM`) and novelty (`pdbTM`) across lengths. Vertical and horizontal dotted lines represent TM-score thresholds of 0.45. **(Right)** Selected samples with high pdbTM scores (colored) with the closest, aligned match from the PDB (gray). Our model generates valid backbones for certain sequence lengths and has a tendency to recapitulate the most frequent folds in the PDB (e.g., tRNAs, small rRNAs).

MODEL	% VALID ↑	DIVERSITY ↑	NOVELTY ↓
BASELINE	**41.0**	0.606	0.541
CLUST	2.0	0.775	0.498
CLUST + CROP	11.0	**0.858**	**0.474**

Table 3. **Impact of data preparation strategies**. Increasing the diversity of the training dataset using a combination of strategies improves diversity and novelty of generated structures but also leads to less designs passing the validity threshold.

4.4. Data Preparation Protocols

We observe an overrepresentation of RNA strands of certain lengths (mostly corresponding to tRNA or 5S ribosomal RNA) in RNAsolo, resulting in our models generating close likenesses for those lengths, achieving high self-consistency. To avoid this memorized recapitulation and promote increased diversity in our samples, we sought to develop data preparation protocols to balance RNA folds across sequence lengths. We also train the models on these data splits for 120K gradient steps, with evaluation results reported in Table 3 showing improved diversity and novelty at the cost of validity (full results in Appendix C.2).

- **Structural Clustering:** We cluster RNAsolo using `qTMclust` (Zhang et al., 2022); we consider two structures similar if their TM-score is above 0.45. When creating a batch for training, we sample random clusters and within them, a random structure from each cluster. This ensures a batch does not comprise solely of samples for a single sequence length or is dominated by overrepresented RNA from RNAsolo. There are only 342 structural clusters for the 5,319 samples within sequence lengths 40-150, which highlights the lack of diversity in RNA structural data.
- **Cropping Augmentation:** We expand our training set by cropping longer RNA strands beyond length 150 by sampling a random crop length (in $[40, 150]$) and extracting a contiguous segment from the larger chain(s). As cropped RNA are not standalone molecules and serve only to augment the dataset, we consider a randomly chosen 20% of the training set size (5,319 samples) to balance uncropped and cropped samples; this gives an additional 1,063 cropped samples.

5. Limitations and Discussions

Altogether, our experiments demonstrate that the $SE(3)$ flow matching framework is sufficiently expressive for learning the distribution of 3D RNA structure and generating realistic RNA backbones similar to well-represented RNA folds in the PDB. Some cherry-picked examples are shown in Figure 5. We have also identified notable limitations and avenues for future work, which we highlight below.

Physical violations. While well-trained models usually generate realistic RNA backbones, we do observe some physical violations: generated backbones sometimes have chains that are either too close by or directly clash with one another, are highly coiled and have excessive loops and intertwined helices that are not physically possible, or have chain breaks. We highlight these limitations in Figure 6.

RNA tertiary structure folding is driven by *base pairing* and *base stacking* interactions (Vicens and Kieft, 2022) which influence the formation of helices, loops, and other tertiary motifs. Base pairing refers to nucleotides along adjacent chains forming hydrogen bonds, while base stacking involves interactions between rings of adjacent nucleotide bases along a chain. To the best of our knowledge, all current deep learning models operate on individual nucleotides and only implicitly learn base pairing and stacking. Developing explicit representations of these interactions as part of the architecture may further minimize physical violations and provide a stronger inductive bias for learning complex tertiary RNA motifs.

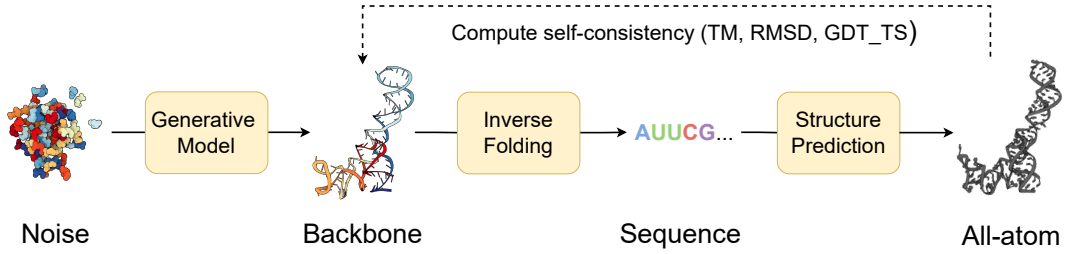

Figure 4. **Structural self-consistency evaluation.** We sample a backbone from our model and pass it through an inverse folding model (gRNAde) to obtain N_{seq} sequences. Each sequence is fed into a structure prediction model (RhoFold) to get the predicted all-atom backbone. Self-consistency between each predicted backbone and the generated sample is measured with TM-score (we also report RMSD and GDT_TS). For a given generated sample, we thus have N_{seq} TM-scores of which we take the maximum as the `scTM` score.

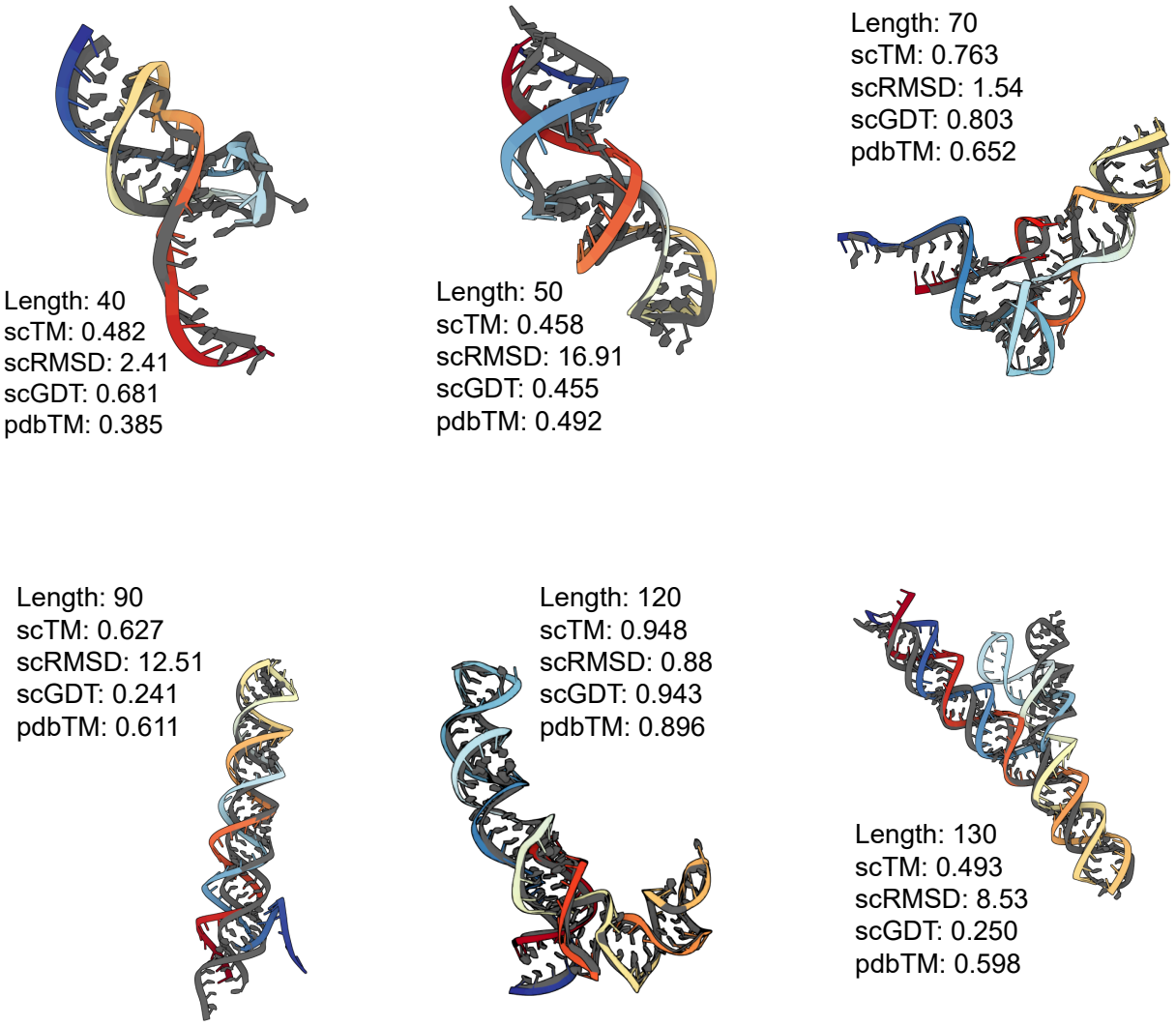

Figure 5. Some generated RNA backbones (colored) of varying lengths aligned with their RhoFold predicted structure (gray).

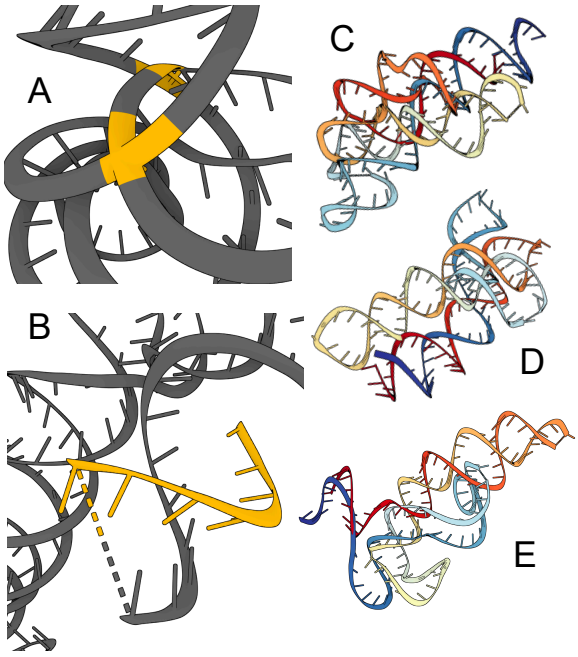

Figure 6. **Physical violations in generated samples.** (A) Steric clashes between generated chains (highlighted in yellow). (B) Chain breaks and stray strands (highlighted in yellow). (C)-(E) Excessive loops and intertwined helices.

Generalisation and novelty. We observed that the best designs from our models (as measured by `scTM` score) are sampled at lengths 70-80 and 120-130, and often have closely matching structures in the PDB (high TM-scores). This suggests that models can recapitulate well-represented RNA folds in their training distribution (e.g., both tRNAs at length 70-90 and small 5S ribosomal RNAs at length 120 are very frequent in RNAsolo). However, self-consistency metrics were relatively poorer for less frequent lengths, suggesting that the model is not designing novel folds at present.

We would also like to note that the models we use for structure prediction and inverse folding may be similarly biased to perform well for certain sequence lengths, leading to the overall pipeline being reliable for commonly occurring lengths and unreliable for less frequent ones (see Appendix A.2 for an analysis on RhoFold). We evaluated preliminary strategies for structural clustering and cropping augmentations during training, which improved the novelty of designed structures but led to fewer designs passing the validity filter. The relative scarcity of RNA structural data compared to proteins necessitates greater care in preparing data pipelines for generative models, which we hope to continue improving upon.

6. Related Work

Here, we summarize recent developments in deep learning for 3D RNA modeling and design. Recent RNA structure prediction tools include RhoFold (Shen et al., 2022), RoseTTAFold2NA (Baek et al., 2022), DRFold (Li et al., 2023), and AlphaFold3 (Abramson et al., 2024), each with varying performance that is yet to match the current state-of-the-art for proteins. However, structure prediction tools are not directly capable of designing new structures, which this work aims to address by adapting an $SE(3)$ flow matching framework for proteins (Yim et al., 2023b). MMDIFF (Morehead et al., 2023), a diffusion model for protein-nucleic acid complex generation, is also capable of designing RNA structures. Our evaluation shows that our flow matching model significantly outperforms both the original and RNA-only versions of MMDIFF.

Joshi et al. (2023) introduce GRNADE for 3D RNA inverse folding, a closely related task of designing new sequences conditioned on backbone structures. We use GRNADE followed by RhoFold in our evaluation pipeline to forward fold designed backbones and measure structural self-consistency. Independently and concurrent to our work, Nori and Jin (2024) propose RNAFlow, which uses GRNADE combined with RoseTTAFold2NA as a denoiser in the flow matching setup to design RNA sequences conditioned on protein structures. Our work tackles *de novo* 3D RNA backbone generation, an orthogonal RNA design task.

7. Conclusion

We introduce RNA-FRAMEFLOW, a generative model for 3D RNA backbone design. *In silico* evaluations show that our model can design locally realistic and moderately novel backbones of length 40 – 150 nucleotides. We achieve a validity score of 41.0% and relatively strong diversity and novelty scores compared to diffusion model baselines and ablated variants. While generative models can successfully recapitulate well-represented RNA folds (e.g., tRNAs, small rRNAs), the lack of diversity in the training data hinders broad generalization at present. We are actively exploring improved data preparation strategies combined with inductive biases that explicitly incorporate interactions that drive RNA structure: *base pairing* and *base stacking*. We hope RNA-FRAMEFLOW and the associated evaluation framework can serve as foundations for the community to explore 3D RNA design, towards developing conditional generative models for real-world design scenarios (Ingraham et al., 2022; Watson et al., 2023).

References

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 2022.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 2023.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 2024.

Hamed Khakzad, Ilia Igashov, Arne Schneuing, Casper Goverde, Michael Bronstein, and Bruno Correia. A new age in protein design empowered by deep learning. *Cell Systems*, 14(11):925–939, 2023.

Jennifer A Doudna and Emmanuelle Charpentier. The new frontier of genome engineering with crispr-cas9. *Science*, 346(6213):1258096, 2014.

Mihir Metkar, Christopher S Pepin, and Melissa J Moore. Tailor made: the art of therapeutic mrna design. *Nature Reviews Drug Discovery*, 23(1):67–83, 2024.

Tulsi Ram Damase, Roman Sukhovershin, Christian Boada, Francesca Taraballi, Roderic I Pettigrew, and John P Cooke. The limitless future of rna therapeutics. *Frontiers in bioengineering and biotechnology*, 9:628137, 2021.

Dongran Han, Xiaodong Qi, Cameron Myhrvold, Bei Wang, Mingjie Dai, Shuoxing Jiang, Maxwell Bates, Yan Liu, Byoungkwon An, Fei Zhang, et al. Single-stranded dna and rna origami. *Science*, 2017.

Joseph D Yesselman, Daniel Eiler, Erik D Carlson, Michael R Gotrik, Anne E d'Aquino, Alexandra N Ooms, Wipapat Kladwang, Paul D Carlson, Xuesong Shi, David A Costantino, et al. Computational design of three-dimensional rna structure and function. *Nature nanotechnology*, 2019.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se(3) diffusion model with application to protein backbone generation, 2023a.

Quentin Vicens and Jeffrey S Kieft. Thoughts on how to think (and talk) about rna structure. *Proceedings of the National Academy of Sciences*, 2022.

Rachael C Kretsch, Ebbe S Andersen, Janusz M Bujnicki, Wah Chiu, Rhiju Das, Bingnan Luo, Benoît Masquida, Ewan KS McRae, Griffin M Schroeder, Zhaoming Su, et al. Rna target highlights in casp15: Evaluation of predicted models by structure providers. *Proteins: Structure, Function, and Bioinformatics*, 2023.

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, 2023b.

Alex Morehead, Jeffrey Ruffolo, Aadyot Bhatnagar, and Ali Madani. Towards joint sequence-structure generation of nucleic acid and protein complexes with se(3)-discrete diffusion, 2023.

Anke Gelbin, Bohdan Schneider, Lester Clowney, Shu-Hsin Hsieh, Wilma K Olson, and Helen M Berman. Geometric parameters in nucleic acids: sugar and phosphate constituents. *Journal of the American Chemical Society*, 118 (3):519–529, 1996.

Stephen C Harvey and M Prabhakaran. Ribose puckering: structure, dynamics, energetics, and the pseudorotation cycle. *Journal of the American Chemical Society*, 108 (20):6128–6136, 1986.

Mary C Clay, Laura R Ganser, Dawn K Merriman, and Hashim M Al-Hashimi. Resolving sugar puckers in rna excited states exposes slow modes of repuckering dynamics. *Nucleic acids research*, 45(14):e134–e134, 2017.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries, 2024.

Bartosz Adamczyk, Maciej Antczak, and Marta Szachniuk. RNAsolo: a repository of cleaned PDB-derived RNA 3D structures. *Bioinformatics*, 2022.

Vincent Boivin, Laurence Faucher-Giguère, Michelle Scott, and Sherif Abou-Elela. The cellular landscape of mid-size noncoding rna. *Wiley Interdisciplinary Reviews: RNA*, 10 (4):e1530, 2019.

Yeqing Lin and Mohammed AlQuraishi. Generating novel, designable, and diverse protein structures by equivariantly diffusing oriented residue clouds, 2023.

Chaitanya K. Joshi, Arian R. Jamasb, Ramon Viñas, Charles Harris, Simon Mathis, and Pietro Liò. grnade: Geometric deep learning for 3d rna inverse design, 2023.

Tao Shen, Zhihang Hu, Zhangzhi Peng, Jiayang Chen, Peng Xiong, Liang Hong, Liangzhen Zheng, Yixuan Wang, Irwin King, Sheng Wang, Siqi Sun, and Yu Li. E2efold-3d: End-to-end deep learning method for accurate de novo rna 3d structure prediction, 2022.

Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. Us-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature methods*, 2022.

Kevin S Keating, Elisabeth L Humphris, and Anna Marie Pyle. A new way to see rna. *Quarterly reviews of biophysics*, 44(4):433–466, 2011.

Minkyung Baek, Ryan McHugh, Ivan Anishchenko, David Baker, and Frank DiMaio. Accurate prediction of nucleic acid and protein-nucleic acid complexes using rosettafoldna. *bioRxiv*, 2022.

Yang Li, Chengxin Zhang, Chenjie Feng, Robin Pearce, P Lydia Freddolino, and Yang Zhang. Integrating end-to-end learning with deep geometrical potentials for ab initio rna structure prediction. *Nature Communications*, 2023.

Divya Nori and Wengong Jin. Rnaflow: Rna structure & sequence co-design via inverse folding-based flow matching. In *ICLR 2024 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2024.

John Ingraham, Max Baranov, Zak Costello, Vincent Frappier, Ahmed Ismail, Shan Tie, Wujie Wang, Vincent Xue, Fritz Obermeyer, Andrew Beam, et al. Illuminating protein space with a programmable generative model. *bioRxiv*, pages 2022–12, 2022.

A. Additional Experimental Details

A.1. Architectural Details

We list hyperparameters used for our denoiser model in Table 4 below:

Category	Hyperparameter	Value
Invariant Point Attention (IPA)	Atom embedding dimension D_h	256
	Hidden dimension D_z	128
	Number of blocks	6
	Query and key points	8
	Number of heads	8
	Key points	12
Transformer	Number of heads	4
	Number of layers	2
Torsion Prediction MLP	Input dimension	256
	Hidden dimension	128
Schedule	Translations (training)	linear
	Rotations (training)	linear
	Translations (sampling)	linear
	Rotations (sampling)	exponential
	Number of denoising steps N_T	$[10, \mathbf{50}, 100, 500]$

Table 4. Hyperparameters for the baseline model.

A.2. RhoFold Length Bias

We investigate the performance of RhoFold on the RNAsolo training dataset used for our generative model. Figure 7 shows sequence length bias where RhoFold predicts structures with extremely low RMSDs for sequence lengths (like 70, 100, and 120) while predicting poor structures for other lengths with larger RMSDs. The performance across lengths is disparate (like AlphaFold2) and may influence what is considered valid. Furthermore, RhoFold is not optimized for *de novo* designed RNA, only naturally occurring RNA. To compensate for bias, we resort to a *ranking* instead of thresholding done by (Yim et al., 2023a;b) when measuring validity.

B. Ablations

B.1. Composition of Backbone Coordinate Loss

We also analyze how changing the composition of atoms considered in the inter-atom losses affects performance. We increase the number of atoms being supervised in the $\mathcal{L}_{\text{bb}}$ loss described above. Aside from the frame comprising $C3'$, $C4'$, and $O4'$, we try two settings with 3 and 7 additional non-frame atoms included in the loss. For the 3 non-frame atoms, we choose $C1'$, P, and $O3'$, and for the 7 non-frame atoms, we choose a superset $C1'$, P, $O3'$, $C5'$, $OP1$, $OP2$, and $N1/N9$. We posit the additional supervision may increase the local structural realism, which may further improve validity, as shown in Table 5.

FRAME COMPOSITION IN $\mathcal{L}_{\text{BB}}$	% VALID ↑	DIVERSITY ↑	NOVELTY ↓
FRAME ONLY (BASELINE)	41.0	**0.606**	**0.571**
FRAME AND 3 NON-FRAME	45.0	0.281	0.794
FRAME AND 7 NON-FRAME	**46.7**	0.356	0.858

Table 5. Ablating composition of backbone loss $\mathcal{L}_{\text{bb}}$. Supervising more non-frame atoms improves validity but worsens diversity and novelty. Best per-column result is **bolded**.

We indeed observe increasing validity as we increase the frame complexity in the auxiliary backbone loss. The minute RMSD contributions from disordered fragments of the RNA may be minimal, accounting for greater likeness to the RhoFold

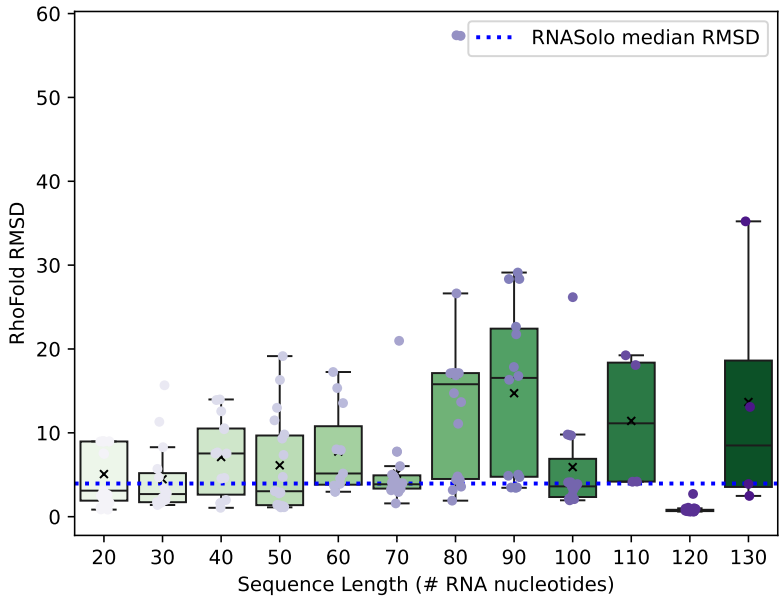

Figure 7. **RhoFold length bias**. RhoFold has a strong bias for certain sequence lengths over others. This affects its efficacy when used to compute the 3D self-consistency of generated backbones. The blue dotted line represents the median RMSD of RhoFold predictions to the samples from RNAsolo. To minimize the influence of this length bias, we use TM-score for self-consistency because it does not penalize flexible regions as much as RMSD.

predicted structures, scoring relatively higher `scTM` scores. However, the original frame-only baseline model has better diversity and novelty which we attribute to high local variation in atomic placements. This variation causes two generated structures for the same sequence length to look very different at an all-atom resolution.

B.2. Composition of Auxiliary Loss

We ablate the inclusion of different auxiliary loss terms that guide our $SE(3)$ flow matching setup; results are in Table 6. Although, there is an increase in EMD for bond distances as we remove distance-based losses like backbone coordinate loss $\mathcal{L}_{\text{bb}}$ and all-to-all pairwise distance loss ($\mathcal{L}_{\text{dist}}$). However, we also observe the model still learns realistic distributions despite removing different loss terms, indicating that each loss makes up for the absence of the other. Moreover, the best model still uses all losses with any removal causing a drop in validity.

$\mathcal{L}_{\text{BB}}$	$\mathcal{L}_{\text{DIST}}$	$\mathcal{L}_{SO(3)}$	EMD (DISTANCE) ↓	EMD (ANGLES) ↓	EMD (TORSIONS) ↓	% VALID ↑
✓	✓	✓	1.5E-6	0.1086	2.360	41.0
✓		✓	3.0E-4	0.1433	3.850	35.0
✓	✓		1.5E-6	0.1180	3.727	13.3
	✓	✓	8.4E-3	0.1891	3.598	16.7

Table 6. Ablations of loss terms on Earth Mover's Distance scores for structural measurements compared to ground truth measurements from RNAsolo. The first row corresponds to the baseline model. Distance-based losses like the backbone coordinate loss ($\mathcal{L}_{\text{bb}}$) and all-to-all pairwise distance loss ($\mathcal{L}_{\text{dist}}$) are necessary to learn geometric properties like bond distances adequately.

Further inspecting the samples from the models without each loss term reveals structural deformities at the all-atom level. Figure 8 shows such artifacts resulting from not enforcing geometric constraints through explicit losses.

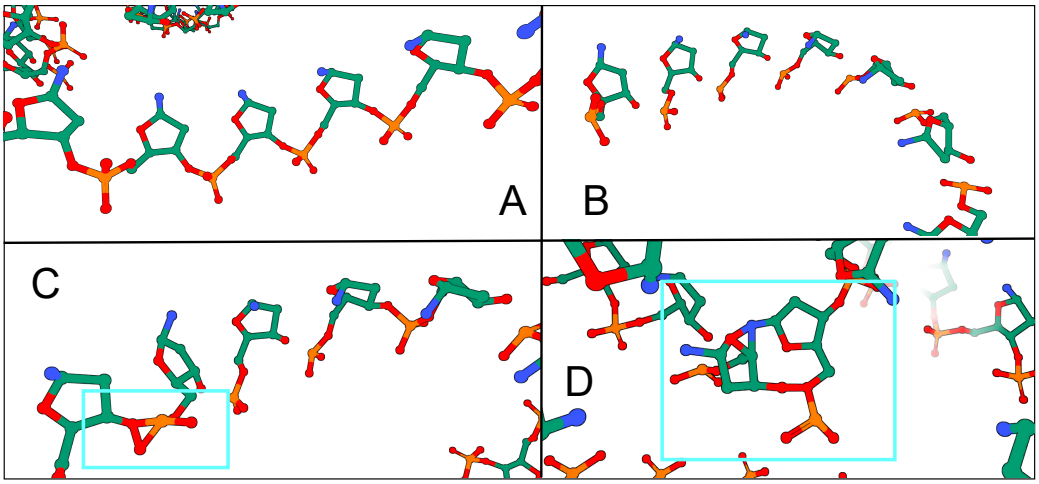

Figure 8. Not including auxiliary losses causes structural issues in generated RNAs. **(A)** RNA backbone from our baseline model with expected adherence to bonding between nucleotides. **(B)** Not including the rotation loss $\mathcal{L}_{SO(3)}$ causes nucleotides to have random orientations, preventing them from connecting contiguously. **(C)** Not including the backbone atom loss $\mathcal{L}_{\text{bb}}$ causes intra-residue atoms to be placed too close to one another resulting in bonds that should not exist. **(D)** Not including the all-to-all pairwise distance loss $\mathcal{L}_{\text{dist}}$ causes deformations and fusing between adjacent frames, and unrealistic nucleotide placements, especially along helices and loops.

C. Additional Results

C.1. Evaluation of MMDIFF Samples

Here, we document global and local metrics from samples generated by MMDIFF. MMDIFF has a validity score of 0.0% as all the samples have a poor `scTM` score below the 0.45 threshold to the RhoFold predicted backbones. Even though none of the samples are valid, we show the average `pdbTM` scores for the samples, which are trivially low as there are no structures from the PDB that match them due to poor quality.

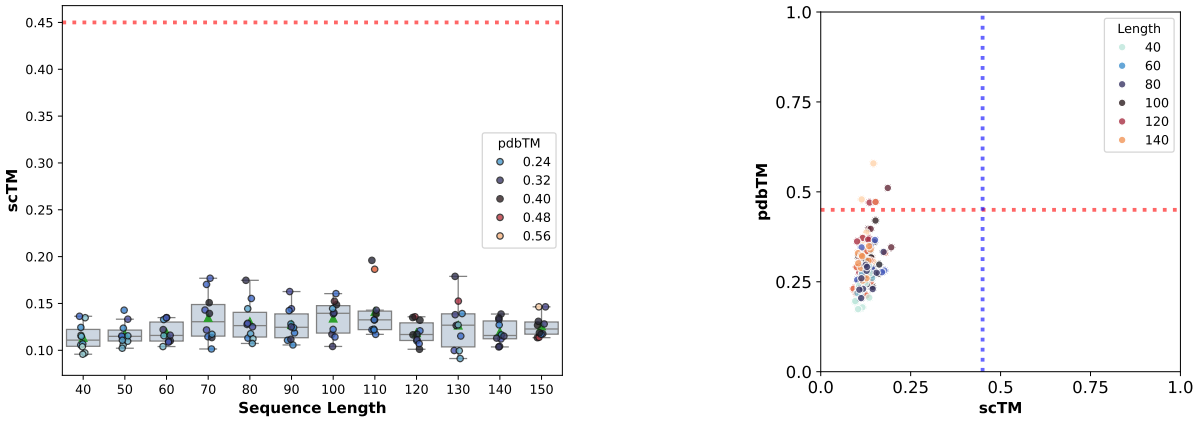

Figure 9. **Validity and novelty of retrained MMDIFF's generated backbones. (Left)** `scTM` of backbones of lengths 40-150 with the mean and spread of `scTM` for each length. **(Middle)** Scatter plot of self-consistency TM-score (`scTM`) and novelty (`pdbTM`) across lengths. Vertical and horizontal dotted lines represent TM-score thresholds of 0.45. Overall, MMDIFF retrained on RNAsolo does not generate realistic RNA structures.

While MMDIFF's samples locally resemble RNA structures given realistic, manual inspection reveals multiple chain breaks and disconnected floating strands, resulting in 0.0% validity. In 10 **(Subplot 1)**, we see inter-residue $C4'$ distances slightly varying, causing the chain breaks. Furthermore, the Ramachandran plot in Figure 10 **(Subplot 4)** reveals a more complex angular distribution than found in RNAsolo, which may be a consequence of excessively folded regions or substructures that may have folded in on themselves.

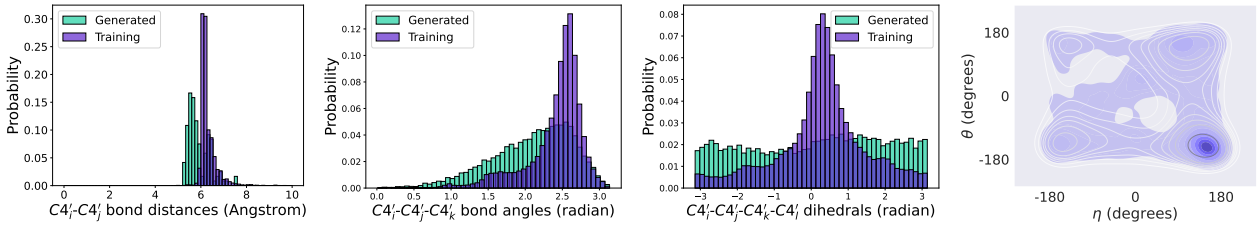

Figure 10. **Structural measurements** from samples generated by MMDIFF. **(Subplots 1-3)** Left: histogram of inter-nucleotide bond distances in Angstrom. Middle: histogram of bond angles between nucleotide triplets. Right: histogram of torsion (dihedral) angles between every four nucleotides. **(Subplot 4)**: RNA-centric Ramachandran plot of structures from RNAsolo (purple) and MMDIFF's generated backbones (white).

C.2. Evaluation of Data Preparation Strategies

We include global evaluation metrics for the two data preparation strategies presented in the main text, namely structural clustering and cropping augmentation.

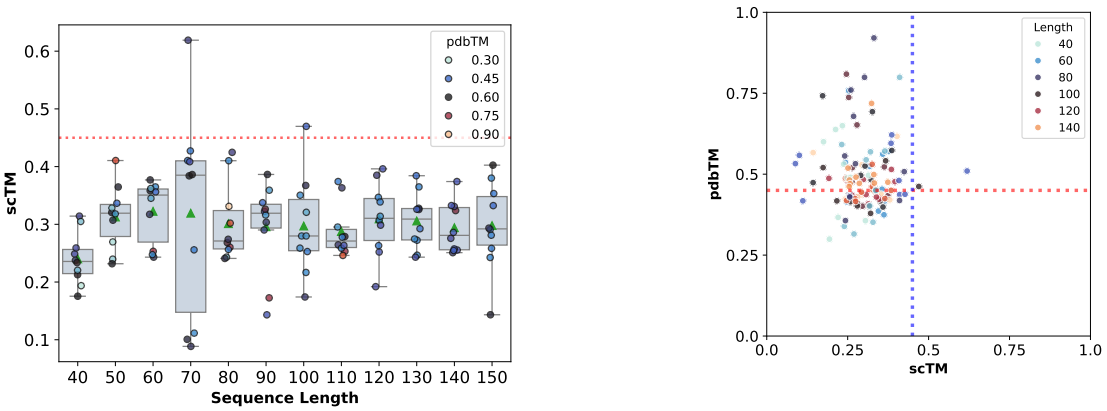

Figure 11. **Validity and novelty of generated backbones from model trained with only structural clustering. (Left)** `scTM` of backbones of lengths 40-150 with the mean and spread of `scTM` for each length. **(Middle)** Scatter plot of self-consistency TM-score (`scTM`) and novelty (`pdbTM`) across lengths. Vertical and horizontal dotted lines represent TM-score thresholds of 0.45.

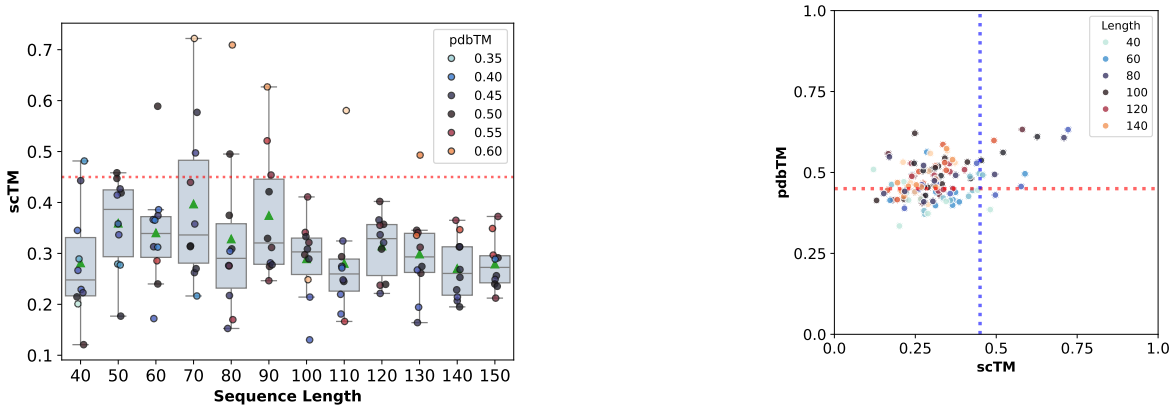

Figure 12. **Validity and novelty of generated backbones from model trained with structural clustering and cropping. (Left)** `scTM` of backbones of lengths 40-150 with the mean and spread of `scTM` for each length. **(Middle)** Scatter plot of self-consistency TM-score (`scTM`) and novelty (`pdbTM`) across lengths. Vertical and horizontal dotted lines represent TM-score thresholds of 0.45.