sCM-RNA: Preference-Aligned Consistency Models for RNA Inverse Folding

Anonymous authors
Paper under double-blind review

Abstract

RNA inverse folding aims to design sequences that fold into prescribed backbone structures, supporting a wide range of applications in RNA engineering. Sequence recovery measures agreement with native sequences, but provides an incomplete assessment of structural fidelity and sequence diversity. Designing diverse sequences with high structural fidelity therefore calls for objectives that evaluate their predicted folds, together with efficient generation. We introduce **sCM-RNA**, an RNA inverse folding framework that combines continuous-time consistency modeling with preference-based alignment. The consistency model enables one-step sequence generation conditioned on target backbone structures, while the alignment stage uses preferences based on predicted structural fidelity to guide generation and explicitly promote sequence diversity. Together, these components support efficient sampling of diverse sequences whose predicted structures closely match the targets. Across six random seeds, sCM-RNA improves TM-score by **2.3%** and sequence diversity by **10.9%** relative to the strongest evaluated baseline for each metric, while achieving the highest sequence recovery. With one-step inference, sCM-RNA also achieves up to **35**$\times$ faster sequence generation, supporting large-scale computational RNA design.

1 Introduction

RNA inverse folding aims to design sequences that fold into prescribed backbone structures, supporting applications in therapeutics (Thavarajah et al., 2021; Zhu et al., 2022), synthetic biology (Pfeifer et al., 2023), and molecular sensing (Ahi & Khorshid, 2025). The task is inherently *one-to-many*: multiple sequences can be compatible with the same target structure. This motivates generating diverse candidate sequences with high structural fidelity, providing a pool for experimental screening and downstream optimization. Efficient generation is also essential for exploring this sequence space at scale.

Sequence recovery measures how closely a designed sequence matches the native one and provides a useful measure of agreement with observed sequences. However, recovery alone provides an incomplete assessment of inverse folding: it does not establish whether alternative designs preserve the target structure or span a diverse sequence space. For a one-to-many design task, disagreement with the native sequence need not imply an unsuccessful design, while recovery alone does not establish the structural fidelity of generated candidates. Evaluating predicted folds therefore provides a complementary signal for assessing structural fidelity, alongside sequence recovery and diversity. This distinction also motivates training objectives that use predicted structural quality to guide generation beyond matching observed sequences.

Recent advances in deep generative modeling have substantially improved RNA inverse folding. Geometric neural networks (Jing et al., 2021; Joshi et al., 2025), diffusion models (Huang et al., 2024), and flow matching models (Nori & Jin, 2024) learn to generate sequences conditioned on target structures. Nevertheless, fitting observed structure–sequence pairs does not explicitly reward generated candidates according to their predicted folds or encourage diversity among them. In addition, sampling from the diffusion and flow matching models considered here requires multiple model evaluations, increasing the cost of generating large candidate pools. These challenges motivate combining objectives based on predicted structural fidelity with efficient sampling and explicit encouragement of sequence diversity.

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.

In this work, we propose **sCM-RNA**, an RNA inverse folding framework that combines *continuous-time consistency modeling* with *preference-based alignment*. We first train a conditional flow matching model and distill it into a continuous-time consistency model, enabling *one-step sequence generation* conditioned on target backbone structures. We then introduce a preference-based alignment stage that uses predicted structural fidelity to guide the one-step generator. Specifically, an external structure predictor folds generated sequences, and their structural similarity to the target backbone determines preference pairs. An entropy-regularized DPO-style objective favors structurally preferred candidates while explicitly encouraging sequence diversity. Together, these stages support efficient sampling of diverse sequences whose predicted structures closely match the targets.

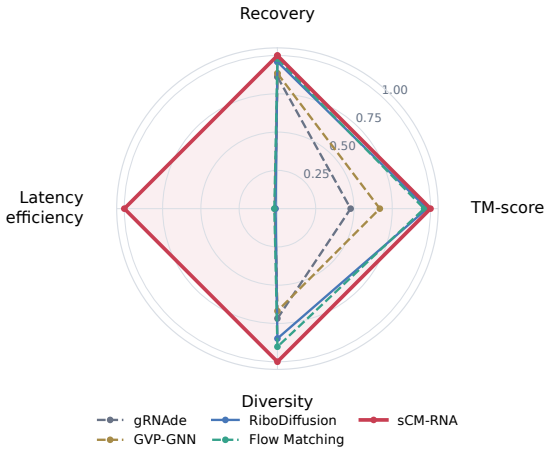


Figure 1: **Normalized performance overview.** Recovery, TM-score and diversity use value/best; latency efficiency uses minimum/value. Larger is better on all four axes. Appendix F reports raw values and evaluation protocols.

We evaluate sCM-RNA in terms of predicted structural fidelity, sequence diversity, sequence recovery, and generation efficiency. Figure 1 summarizes these dimensions in comparison with geometric, diffusion, and flow matching baselines. Across six random seeds, sCM-RNA improves TM-score by **2.3%** and sequence diversity by **10.9%** relative to the strongest evaluated baseline for each metric, while achieving the highest sequence recovery. One-step inference also enables up to **35**$\times$ faster sequence generation. These results support the use of sCM-RNA for large-scale computational RNA design, where both candidate quality and sampling cost matter.

Our contributions are summarized as follows:

- We motivate evaluating RNA inverse folding jointly in terms of predicted structural fidelity, sequence diversity, and generation efficiency, with sequence recovery providing a complementary measure.
- We propose **sCM-RNA**, which integrates continuous-time consistency distillation for one-step conditional generation with structure-driven preference alignment and entropy regularization to encourage sequence diversity.
- Across six random seeds, we demonstrate improvements in TM-score and sequence diversity over the strongest evaluated baseline for each metric, together with the highest sequence recovery and up to **35**$\times$ faster sequence generation.

2 Related Works

RNA Inverse Folding and Generative Modeling. Early RNA inverse folding methods relied on combinatorial search and thermodynamic optimization (Hofacker et al., 1994; Busch & Backofen, 2006; Taneda, 2010; Garcia-Martin et al., 2013), using handcrafted energy functions and heuristic updates to recover native sequences or minimize predefined folding energies. This implicitly treats the reference sequence as unique, despite the biological one-to-many nature of RNA folding, where diverse sequences can realize the same structure; only a few works (Park et al., 2024; Sun et al., 2025; Gasser et al., 2025; Wang et al., 2025) explicitly optimize post folding structural fidelity or diversity. With deep learning, geometric architectures (Gao et al., 2023; Kubaney et al., 2025) such as GVP-GNN (Jing et al., 2021) and gRNAde (Joshi et al., 2025) encode RNA backbones as spatial graphs, improving structural awareness and generalization. Generative models further advanced inverse folding: RiboDiffusion (Huang et al., 2024) formulates it as conditional denoising over sequence representations, while flow matching methods (Nori & Jin, 2024) replace stochastic diffusion with probability flow ODEs for straighter trajectories and improved samples. However, many methods still rely primarily on recovery rate, and both diffusion (Song et al., 2020; Ho et al.,

2020; Nichol & Dhariwal, 2021) and flow matching (Lipman et al., 2023; Gat et al., 2024) require multi step numerical solvers, limiting large scale RNA design and virtual screening.

Consistency Models and One-Step Generation. Consistency Models (CMs) reduce diffusion sampling cost by learning mappings from arbitrary points on probability flow trajectories to clean data, enabling single step or few step generation (Song et al., 2023; Luo et al., 2023). Original CMs use discrete time consistency constraints, which can introduce discretization error and training instability on complex structured manifolds. Recent works have applied discrete time consistency distillation to scientific and inverse problem settings, including the Inverse Consistency Model (Zhang & Zhou, 2025). In contrast, sCM (Lu & Song, 2025) enforces consistency directly in continuous time through invariance along probability flow trajectories, avoiding hand crafted discretization schedules and better matching continuous dynamics. This makes sCM well suited for distilling flow matching teachers into high fidelity one step generators; our work first adapts continuous time consistency models to RNA inverse folding and integrates them into a biologically grounded framework beyond recovery centric objectives.

Preference Optimization and Biological Alignment. Likelihood based objectives often misalign with downstream biological viability, including folding stability, structural accuracy, and functional activity, motivating reinforcement learning and preference based optimization for biological generative models. RLHF style policy gradient methods (Schulman et al., 2017; Shao et al., 2024) can optimize black box rewards but are unstable and hyperparameter sensitive. Direct Preference Optimization (DPO) offers a simpler alternative by learning from pairwise preferences without an explicit reward model. DiffusionDPO (Wallace et al., 2024) adapts DPO to diffusion models, and preference based learning has been applied to protein design and small molecule optimization (Park et al., 2024; Sun et al., 2025). These methods usually operate within diffusion frameworks and optimize surrogate likelihoods from discretized generative processes. We extend preference optimization beyond diffusion to flow matching and sCM-based generators through an entropy-regularized DPO-style formulation for continuous time consistency models, directly aligning one-step RNA generators with structural objectives while encouraging sequence diversity.

3 METHODS

RNA inverse folding and framework overview. Given a backbone $X \in \mathbb{R}^{L\times 3\times 3}$ containing the $(P, C4', N1/N9)$ coordinates of L residues, we generate a sequence $S = (s_1, \ldots, s_L)$ over $\{A, U, C, G\}$ whose predicted fold matches X. We represent a sequence by $x_1(S) = \textsc{OneHot}(S) \in \mathbb{R}^{L\times 4}$ and model a conditional distribution to accommodate multiple sequences compatible with the same backbone. Figure 2 summarizes our three stages: structure-conditioned flow matching learns sequence generation, continuous-time consistency distillation enables one-step sampling, and structure-guided preference alignment improves predicted folding fidelity while encouraging diversity.

3.1 STAGE I: STRUCTURE-CONDITIONED FLOW MATCHING

Backbone-conditioned sequence modeling. Following RiboDiffusion (Huang et al., 2024), a GVP-GNN encoder (Jing et al., 2021) extracts geometric features h_X from the RNA backbone. These features condition the teacher velocity field $v_\theta(x_t, t \mid X)$ and the student introduced below. Both operate on continuous states in $\mathbb{R}^{L\times 4}$, retaining one four-channel representation per nucleotide position. For a training pair (X, S), we interpolate between Gaussian noise and its nucleotide encoding:

$$x_t = (1-t)x_0 + tx_1(S), \qquad x_0 \sim \mathcal{N}(0, I), \quad t \sim \mathcal{U}[0, 1]. \tag{1}$$

The flow-matching objective regresses the conditional velocity toward the displacement $x_1(S) - x_0$:

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{(X,S),x_0,t}\left[\|v_\theta(x_t, t \mid X) - (x_1(S) - x_0)\|_2^2\right]. \tag{2}$$

The expectation uses training pairs and the noise/time distributions in Equation (1). At inference, the teacher transports noise toward sequence representations by integrating $dx_t/dt = v_\theta(x_t, t \mid X)$ from $t = 0$ to 1. This learns backbone-conditioned generation from observed RNA sequences, but requires multiple network evaluations and does not directly evaluate the folds of generated candidates.

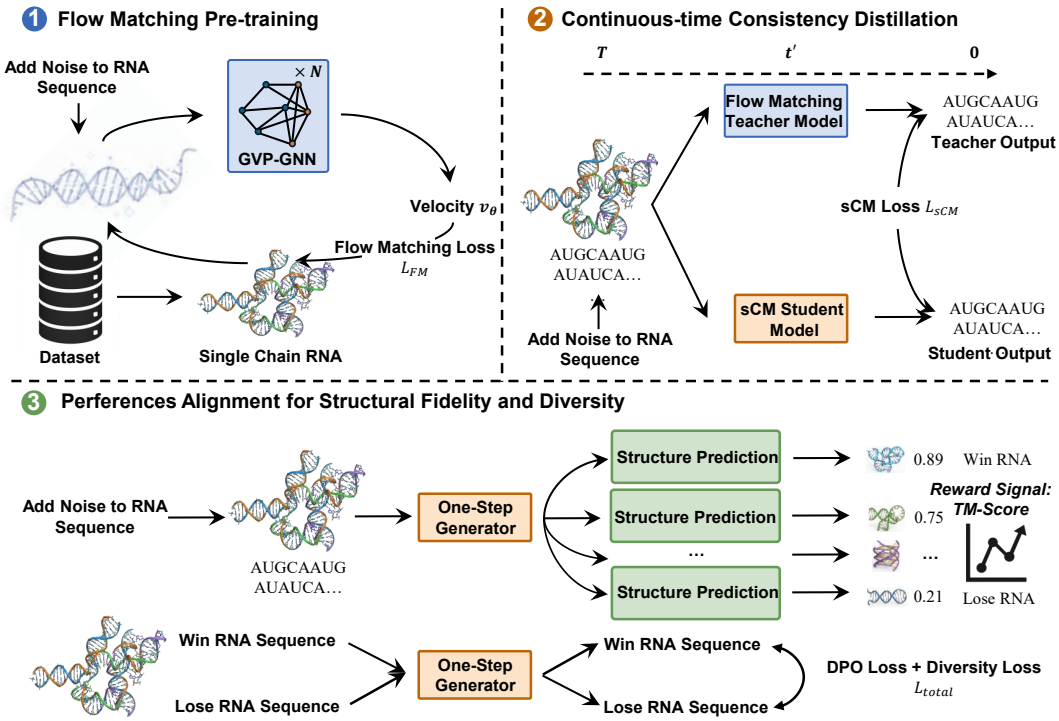


Figure 2: **Overview of sCM-RNA.** We decouple RNA inverse folding into three stages. **(I) Flow matching pre-training:** learn a structure-conditioned flow matching model over RNA sequence. **(II) Continuous-time consistency distillation:** distill the multi-step flow matching model into a one-step sCM for fast conditional generation. **(III) Preference alignment:** sample multiple RNA sequences from the sCM, fold each candidate with RhoFold (Shen et al., 2024), compute structural fidelity by TM-score against the target backbone, construct win/lose pairs, and optimize an sCM-DPO with entropy regularization to improve structural fidelity while preserving diversity.

3.2 Stage II: Consistency Distillation for One-Step RNA Generation

We distill the frozen FM teacher into a structure-conditioned continuous-time consistency model (sCM) (Lu & Song, 2025). The student predicts the terminal sequence representation from a noisy state and its target backbone. We distinguish the residual network u_ϕ from the endpoint predictor:

$$f_\phi(x_t, t \mid X) = x_t + (1-t)u_\phi(x_t, t \mid X). \tag{3}$$

This parameterization enforces $f_\phi(x, 1 \mid X) = x$. For a fixed backbone, endpoint predictions should remain consistent as the input moves along the teacher flow. Writing $D_\theta f_\phi$ for the directional derivative along that flow, we use

$$\mathcal{L}_{\text{sCM}}(\phi) = \mathbb{E}_{(X,S),x_0,t}\left[\frac{w(t)}{L}\,\|D_\theta f_\phi(x_t, t \mid X)\|_2^2\right]. \tag{4}$$

The squared norm sums over the four nucleotide channels and valid residues, with L the number of valid residues. Here x_t follows Equation (1) and $w(t)$ weights the training times. Appendix A expands the derivative and gives the JVP implementation and masking details. The boundary condition and consistency objective aim to approximate the teacher's conditional endpoint map. Generation then requires one student evaluation:

$$\hat{x}_1 = f_\phi(z, 0 \mid X) = z + u_\phi(z, 0 \mid X), \qquad z \sim \mathcal{N}(0, I). \tag{5}$$

The endpoint is decoded into the four-letter RNA alphabet. Independent noise draws produce a candidate pool for the same backbone without repeatedly solving the teacher ODE.

3.3 Stage III: Structure-Guided sCM Preference Alignment

Distillation targets the teacher's generation behavior; it does not directly favor RNA sequences according to their predicted folds. We therefore align the distilled student using structural preferences and a frozen reference $f_{\phi_{\text{ref}}}$ initialized before alignment.

Preferences from predicted RNA folds. For each training backbone X, we generate K candidates from the distilled sCM and fold them with RhoFold (Shen et al., 2024). We score each candidate by $m^{(k)} = \text{TM}(\hat{X}^{(k)}, X)$, select the highest- and lowest-scoring sequences (S_w, S_l), and retain the pair only if $m_w - m_l \geq \delta$. The resulting offline dataset $\mathcal{D}_{\text{pref}} = \{(X, S_w, S_l)\}$ is reused during alignment. Thus the preference signal measures predicted structural compatibility rather than native-sequence agreement. The margin removes weak comparisons but does not guarantee high absolute quality for every winner.

Endpoint compatibility and sCM-DPO. For a candidate S, we construct $x_t(S)$ using Equation (1) and define

$$E_\phi(S; X, t) = \|f_\phi(x_t(S), t \mid X) - x_1(S)\|_2^2, \qquad \mathcal{S}_\phi(S; X, t) = -E_\phi(S; X, t). \tag{6}$$

Energy measures endpoint reconstruction compatibility, not normalized sequence likelihood. The endpoint predictor in Equation (3), rather than the raw residual network, supplies this score. Winner and loser share the sampled time and Gaussian noise while retaining their candidate-specific interpolated states. We optimize their relative score margin against the frozen reference:

$$\mathcal{L}_{\text{sCM-DPO}}(\phi) = -\mathbb{E}_{(X,S_w,S_l),t,x_0}\left[\log \sigma(\beta(\Delta_\phi - \Delta_{\text{ref}}))\right], \tag{7}$$

where $\Delta_\phi = \mathcal{S}_\phi(S_w; X, t) - \mathcal{S}_\phi(S_l; X, t)$ and Δ_{ref} is the same difference under $f_{\phi_{\text{ref}}}$. The expectation samples preference pairs from $\mathcal{D}_{\text{pref}}$ and noise/time as in Equation (1); $\beta > 0$ scales the relative margin. This DPO-style surrogate favors oracle-preferred sequences relative to the reference. Its relationship to diffusion and flow preference objectives (Wallace et al., 2024; Ziv et al., 2025) and its ranking assumptions are discussed in Appendix B.

Diversity within the RNA candidate pool. To encourage alternatives compatible with the same backbone, we regularize the per-position nucleotide probabilities with token entropy:

$$\mathcal{L}_{\text{total}}(\phi) = \mathcal{L}_{\text{sCM-DPO}}(\phi) - \lambda \sum_{i=1}^{L} H(p_\phi(s_i \mid X)). \tag{8}$$

We use $\lambda = 0.05$. Token entropy encourages local variation, while structural preferences favor fidelity; diversity of the generated pool is evaluated separately by pairwise Hamming distance. Appendix C.5 gives the alignment hyperparameters, and Algorithm 3 summarizes the procedure.

4 Experiments

4.1 Experimental Setup

Benchmarks. We evaluated RNA inverse folding on standard benchmarks with experimentally determined 3D structures, following prior work. Unless otherwise specified, all methods take the same RNA backbone representation X as input and generate a fixed size candidate pool of sequences per target. For the in domain benchmark, we follow the official split protocol of RiboDiffusion, yielding 6506/773/1370 training/validation/test targets.

To assess robustness under distribution shift, we further construct an out-of-distribution (OOD) benchmark from recently released RNA structures collected from the CAMEO 3D server [1]. Targets are filtered to lengths between 50 and 200 nucleotides, matching the regime used in RiboDiffusion and our in domain experiments. The OOD set contains 60 RNA chains from 47 unique PDB entries released between 2025-01-11 and 2026-01-10. We remove any PDB-ID overlap with training or validation targets and preprocess all structures to ensure a consistent backbone representation. Compared with the training set, the OOD set is length shifted: median/mean chain lengths are 71.5/96.5 nt for OOD versus 45.0/61.0 nt for training, with a two sample KS statistic of 0.50.

[1] https://cameo3d.org/

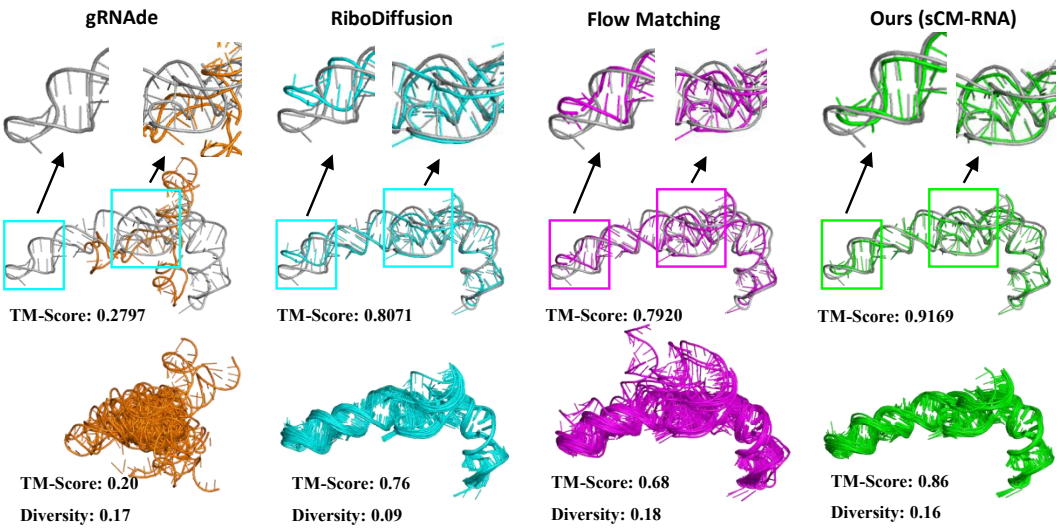


Figure 3: **Qualitative comparison of RNA backbone generation across methods on target *6wru_2*.** Predicted structures are aligned to the ground truth. sCM-RNA achieves the highest structural fidelity while maintaining non-collapsed sample diversity, as evidenced by both TM-score and visual variation across 16 samples.

Baselines. We compare against representative inverse folding approaches, including: (i) GNN-based models, **gRNAde** (Joshi et al., 2025) and **GVP-GNN** (Jing et al., 2021); (ii) diffusion based inverse folding, **RiboDiffusion** (Huang et al., 2024); and (iii) flow matching based variants built on the same GVP-GNN encoder backbone. To isolate the contribution of each stage in our pipeline, we additionally report **Flow Matching** (teacher) and **sCM** (one step distilled student) prior to preference alignment.

Evaluation Setting. For each target backbone, we sample $N = 16$ sequences from every method and report the average metric over the test set under the same candidate budget. For the main structure based evaluation, we fold each generated sequence using **RhoFold** (Shen et al., 2024) and compare the predicted structure with the ground truth backbone. All RhoFold runs use the single sequence setting without MSA or relaxation. During preference construction, we discard ambiguous pairs whose TM-score margin is smaller than $\delta = 0.1$. To assess the generality of structural improvements, we further evaluate generated sequences with multiple folding oracles, including Chai lab folding/evaluation and RhoFold aligned Chai lab evaluation in Appendix Section C.2.

Metrics. We report four complementary metrics that capture fidelity, stability, and one-to-many behavior.

- **Sequence recovery** (↑): fraction of positions that match the native sequence.
- **TM-score** (↑): structural similarity between the RhoFold predicted structure from a generated sequence and the ground truth backbone.
- **MFE** (↓): the minimum free energy calculated uses linear ViennaRNA (Lorenz et al., 2011); lower values indicate higher thermodynamic stability.
- **Diversity** (↑): average pairwise normalized Hamming distance within the candidate pool $\mathcal{S} = \{s_1, \ldots, s_N\}$ for each backbone:

$$\mathrm{Div}(\mathcal{S}) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} \frac{d_H(s_i, s_j)}{L}. \tag{9}$$

4.2 Main Results and OOD Generalization

On the in domain benchmark, sCM-RNA consistently achieves the highest structural fidelity and sequence diversity while achieving the best recovery rate across six random seeds. Compared with

Table 1: **In-domain RhoFold folding/evaluation over six random seeds.** Unless marked otherwise, values are mean ± standard deviation over six sampling seeds.

Method	Recovery↑	TM-score↑	MFE↓	Diversity↑
gRNAde (Joshi et al., 2025)	0.4494 ± 0.004	0.5032 ± 0.017	-17.14 ± 2.10	0.2417 ± 0.018
GVP-GNN (Jing et al., 2021)	0.4609 ± 0.006	0.5601 ± 0.015	-15.91 ± 1.80	0.2553 ± 0.016
RiboDiffusion (Huang et al., 2024)	0.5004 ± 0.020	0.5993 ± 0.011	$\mathbf{-20.04} \pm 1.20$	0.2859 ± 0.015
Flow Matching (ours, teacher)	0.5113 ± 0.011	0.5875 ± 0.012	-18.58 ± 1.25	0.3040 ± 0.016
sCM-RNA	$\mathbf{0.5217} \pm 0.006$	$\mathbf{0.6131} \pm 0.005$	-18.64 ± 0.55	$\mathbf{0.3370} \pm 0.007$

Table 2: **Chai-lab evaluation on generated sequences.** Generated sequences are folded and evaluated with Chai-lab under the same candidate budget.

	In Domain				OOD			
Method	Rec.↑	TM↑	MFE↓	Div.↑	Rec.↑	TM↑	MFE↓	Div.↑
RiboDiffusion	0.5235	0.5961	**-18.88**	0.3032	0.5095	0.4291	-36.86	0.3098
Flow Matching	0.5255	0.5838	-17.59	0.3225	0.5429	0.4218	-32.68	0.3018
sCM-RNA	**0.5286**	**0.6076**	-18.18	**0.3451**	**0.5436**	**0.4376**	**-37.12**	**0.3294**

the strongest baseline for each metric, our method improves the TM-score from 0.5993 ± 0.011 for RiboDiffusion (Huang et al., 2024) to 0.6131 ± 0.005, and improves diversity from 0.3040 ± 0.016 for Flow Matching to 0.3370 ± 0.007. It also improves recovery from 0.5113 ± 0.011 for Flow Matching to 0.5217 ± 0.006, indicating improved coverage of the one-to-many solution space without sacrificing native-sequence consistency. While some baselines achieve lower minimum free energy (MFE), prior studies have shown that MFE alone is an insufficient proxy for correct or robust folding, as energetically favorable sequences may still adopt incorrect structures or lack robustness (Reidys et al., 1997; Garcia-Martin et al., 2013; Sun et al., 2025). In contrast, sCM-RNA consistently improves structure level metrics while preserving thermodynamic stability, suggesting a more biologically meaningful trade off. Notably, although all methods experience performance degradation under distribution shift, the OOD results in Table 6 show that sCM-RNA remains strongest on average under an independent folding oracle, indicating stronger robustness to unseen RNA structures. Because the absolute TM-score gains are modest, we report six-seed statistics in Table 1. sCM-RNA consistently improves the two metrics most central to our claim, TM-score and diversity, while maintaining comparable recovery and smaller variance. We therefore frame the result as a stronger fidelity, diversity and efficiency trade off.

Oracle robustness. The main results in Table 1 use RhoFold for both preference construction and structural evaluation. We additionally report Chai-lab team et al. (2024) folding in Table 2 and Chai-lab evaluation after RhoFold alignment in Tables 3 and 6 to assess transfer beyond the alignment oracle. In both settings, sCM-RNA preserves the best overall fidelity diversity trade-off, providing evidence of transfer while shared predictor biases remain possible. For the cross oracle TM-score gap over Flow Matching, the approximate 95% confidence interval is $[0.0109, 0.0507]$ in domain and $[-0.0017, 0.0475]$ on OOD, indicating stronger evidence in domain and positive but less conclusive OOD gains. Pass@16 is reported in Section C.4, showing that the gain is most pronounced at stricter TM-score thresholds.

4.3 Inference Efficiency

As shown in Figure 4, sCM-RNA achieves a substantial efficiency gain over diffusion and flow matching based baselines. sCM-RNA reaches 0.0326s per sample. By distilling multi step probability flow sampling into a one step generator, sCM-RNA reduces inference latency by over **35**× while lowering FLOPs. Notably, the efficiency improvement does not come from reducing model capacity: sCM-RNA uses a similar number of parameters as the flow matching baseline, but achieves lower latency due to one step inference. Such efficiency gains are critical in practical biomolecular design pipelines, where inverse folding models are used to generate large candidate pools for downstream screening. Recent work in de novo protein and antibody design, as well as virtual RNA target screening routinely operates at scales of 10^4–10^6 candidates, making per-sample generation cost a key bottleneck (Bennett et al., 2023; Li et al., 2023; Carlsson & Luttens, 2024; Carvajal-Patiño et al., 2025).

Table 3: **In-domain Chai-lab evaluation after RhoFold alignment.** Preferences are constructed with RhoFold, while test-time structures are folded and scored with Chai-lab.

Method	Rec.↑	TM↑	MFE↓	Div.↑
RiboDiffusion	0.5057 ± 0.018	0.5851 ± 0.012	$\mathbf{-19.98} \pm 1.10$	0.2902 ± 0.014
Flow Matching	0.5146 ± 0.011	0.5718 ± 0.013	-18.79 ± 1.20	0.3085 ± 0.015
sCM-RNA	$\mathbf{0.5217} \pm 0.007$	$\mathbf{0.6026} \pm 0.006$	-18.83 ± 0.65	$\mathbf{0.3373} \pm 0.008$

Table 4: **Auxiliary UTR-LM scores on in-domain and OOD candidates.** Higher scores are UTR-specific model predictions, not experimental functional measurements.

	In Domain			**OOD**		
Method	MRL	TE	EL	MRL	TE	EL
gRNAde (Joshi et al., 2025)	-0.1478	-0.3291	-0.0172	0.0154	0.0774	-0.2749
GVP-GNN (Jing et al., 2021)	-0.0834	-0.2417	-0.0069	0.0412	0.1036	-0.2148
RiboDiffusion (Huang et al., 2024)	-0.4480	-0.1203	0.0583	0.1245	0.1247	-0.1802
Flow Matching	0.3968	0.3706	-0.1208	-0.1652	0.2239	-0.0885
Ours (sCM-RNA)	**0.4285**	**0.4012**	**0.0769**	**0.1897**	**0.2683**	**0.0192**

4.4 Ablation Study

We analyze the contribution of preference alignment and diversity regularization in Table 5. *Base* denotes the distilled sCM before preference learning.

Effect of preference alignment. DPO driven by the TM-score consistently improves structural fidelity from 0.5956 to 0.6112, confirming that post training can directly optimize structure level objectives beyond teacher distillation.

Diversity control. Figure 5 reports the entropy-weight sweep; the final setting is $\lambda = 0.05$. Optimizing diversity alone increases candidate variation but degrades structural fidelity, highlighting a non trivial trade off. Combining TM-score preference with **token-level entropy regularization** yields the best balance: it achieves the highest TM-score while preserving strong diversity, outperforming explicit *TM-Score + Diversity* pairing in structural fidelity. The TER comparison also separates objective level diversity control from decoding heuristics: the one step sCM samples directly from Gaussian noise through $\hat{x}_1 = f_\phi(z, 0 \mid X)$, without a separate autoregressive sampling temperature.

4.5 Auxiliary Sequence-Level Evaluation

Structural fidelity alone is a necessary but insufficient criterion for practical RNA design. In realistic downstream applications (Wu et al., 2019; Tabuchi & Yokobayashi, 2022), multiple variants that realize the same target structure are typically screened for robustness, expression, or regulatory potential (Cao et al., 2021; Castillo-Hair et al., 2024). We include **UTR-LM** (Chu et al., 2024) as an auxiliary sequence-level evaluator of structurally selected candidates. Specifically, after generating $N = 16$ candidates, we first select the sequence with the highest TM-score for each target and then

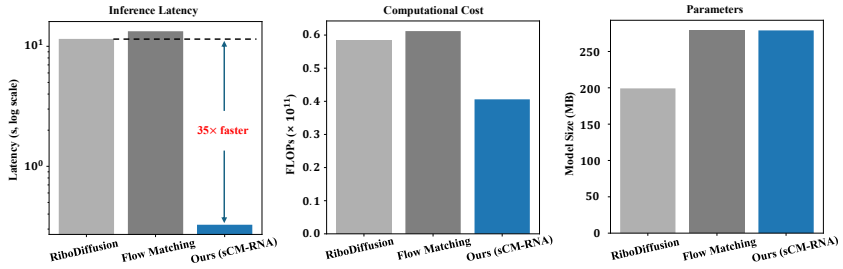


Figure 4: **Inference efficiency comparison.** We compare (a) inference latency, (b) computational cost measured by FLOPs, and (c) model size across methods on the same A6000 GPU. Diffusion and flow matching baselines require 50 step sampling, whereas sCM-RNA performs one step generation, resulting in an over **35× reduction in inference latency** with lower computational cost.

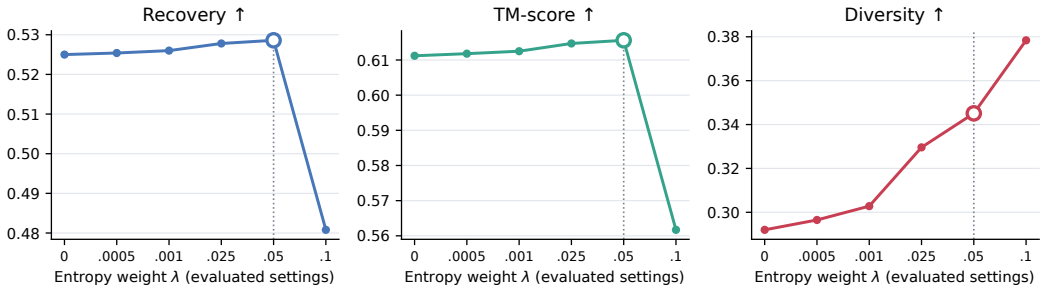


Figure 5: **Effect of entropy regularization.** Sweep over six values of λ, with other settings fixed. The outlined marker denotes the final $\lambda = 0.05$ setting. Evaluated settings are equally spaced for readability. Increasing λ to 0.05 improves the reported fidelity and diversity together; 0.1 increases diversity further at a cost in recovery and TM-score. These point estimates do not establish a global Pareto frontier.

Table 5: **Ablations on alignment objectives.** We vary the preference signal (TM-score / MFE / Diversity) and compare explicit diversity pairing vs. token level entropy regularization(TER). *Base*: distilled sCM without DPO.

Configuration	Recovery ↑	TM-score ↑	MFE ↓	Diversity ↑
sCM (*Base*)	**0.5382**	0.5956	-17.73	0.2818
DPO variations				
+ TM-score only	0.5250	0.6112	-16.83	0.2920
+ Diversity only	0.5224	0.5808	-13.30	**0.3737**
+ MFE only	0.5245	0.5955	**-19.97**	0.3147
+ TM-score + Diversity	0.5247	0.6038	-17.97	0.3467
+ TM-score + TER	0.5286	**0.6156**	-18.00	0.3451

evaluate it with **UTR-LM** (Chu et al., 2024), a pretrained language model for 5′ untranslated regions. UTR-LM provides three complementary proxy metrics: (i) mean ribosome load (MRL), reflecting translation initiation; (ii) translation efficiency (TE), capturing protein output per transcript; and (iii) expression level (EL), integrating transcriptional and translational signals. These scores are model derived and domain specific, not direct experimental measurements or general ncRNA functional validation. This selection does not equalize structural quality across methods or isolate a causal effect of diversity. We interpret the scores as auxiliary descriptions of the selected candidates.

Results and interpretation. As shown in Table 4, sCM-RNA consistently achieves superior UTR-LM proxy scores across all three metrics in both in domain and OOD settings. Notably, sCM-RNA is the only method to obtain a positive expression level (EL) on the OOD benchmark. These proxy-score differences are descriptive and do not establish general RNA function or a causal effect of candidate diversity. Additional qualitative analysis and metric details are provided in Appendix Section D.

5 Conclusion

We presented **sCM-RNA**, a biologically aligned framework for RNA inverse folding that jointly optimizes structural fidelity, diversity, and efficiency. By combining sCM, sCM-RNA enables conditional generation in one step that is up to 35× faster than diffusion and flow matching based baselines. Beyond efficiency, sCM-RNA improves structural fidelity and diversity compared to the strongest baselines under six-seed evaluation, achieving a **2.3%** gain in TM-score and a **10.9%** improvement in sequence diversity while achieving the best recovery rate. Cross-oracle evaluation supports transfer beyond the alignment predictor, although OOD gains remain less conclusive. The framework supports efficient computational candidate generation; general biological function requires task-specific validation.

AI USE STATEMENT

We used generative AI tools to assist with language editing and translation of the manuscript and with implementing portions of the research code. AI use was limited to these tasks. The authors reviewed all AI-assisted text and code and take responsibility for the final content of this work, including its claims, results, and implementation.

ETHICS STATEMENT

This work develops a computational method for RNA inverse folding, with potential applications in RNA engineering and molecular design. More efficient sequence generation may also lower barriers to misuse in biological design. sCM-RNA optimizes predicted structural fidelity and sequence diversity; these objectives do not establish biological function, therapeutic efficacy, or safety. In particular, agreement with a folding predictor is not a substitute for experimental validation. Any downstream synthesis or biological application requires task-specific risk assessment, appropriate biosafety oversight, and experimental validation by qualified researchers. We therefore interpret the reported results as evidence of computational design performance and avoid drawing conclusions about the safety or biological activity of generated sequences.

REPRODUCIBILITY STATEMENT

The three stages of sCM-RNA and their optimization objectives are described in Section 3. Appendix A provides training pseudocode and details of the consistency distillation, while Appendix B presents the preference optimization derivation, assumptions, and supporting proofs. Section 4 specifies the benchmarks, baselines, evaluation metrics, and folding settings. Appendix C.1 documents the dataset splits, candidate budget, preference filtering threshold, and training schedule. We report results across six random seeds in Table 1, cross-oracle evaluations in Appendix C.2, and hyperparameter sensitivity analyses in Appendix C.5 to support assessment of the stability of the reported findings.

REFERENCES

Ehsan Pashay Ahi and Mehran Khorshid. Potentials of rna biosensors in developmental biology. *Developmental Biology*, 526:173–188, 2025. ISSN 0012-1606.

Nathaniel R Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, et al. Improving de novo protein binder design with deep learning. *Nature Communications*, 14(1):2625, 2023.

Anke Busch and Rolf Backofen. Info-rna—a fast approach to inverse rna folding. *Bioinformatics*, 22 (15):1823–1831, 2006.

Jicong Cao, Eva Maria Novoa, Zhizhuo Zhang, William CW Chen, Dianbo Liu, Gigi CG Choi, Alan SL Wong, Claudia Wehrspaun, Manolis Kellis, and Timothy K Lu. High-throughput 5' utr engineering for enhanced protein production in non-viral gene therapies. *Nature communications*, 12(1):4138, 2021.

Jens Carlsson and Andreas Luttens. Structure-based virtual screening of vast chemical space as a starting point for drug discovery. *Current Opinion in Structural Biology*, 87:102829, 2024. ISSN 0959-440X.

Juan G Carvajal-Patiño, Vincent Mallet, David Becerra, Luis Fernando Niño Vasquez, Carlos Oliver, and Jérôme Waldispühl. Rnamigos2: accelerated structure-based rna virtual screening with deep graph learning. *Nature Communications*, 16(1):1–12, 2025.

Sebastian Castillo-Hair, Stephen Fedak, Ban Wang, Johannes Linder, Kyle Havens, Michael Certo, and Georg Seelig. Optimizing 5'utrs for mrna-delivered gene editing using deep learning. *Nature Communications*, 15(1):5284, 2024.

Yanyi Chu, Dan Yu, Yupeng Li, Kaixuan Huang, Yue Shen, Le Cong, Jason Zhang, and Mengdi Wang. A 5' utr language model for decoding untranslated regions of mrna and function predictions. *Nature Machine Intelligence*, 6(4):449–460, 2024.

Zhangyang Gao, Cheng Tan, and Stan Z. Li. Pifold: Toward effective and efficient protein inverse folding. In *The Eleventh International Conference on Learning Representations*, 2023.

Juan Antonio Garcia-Martin, Peter Clote, and Ivan Dotu. Rnaifold: a constraint programming algorithm for rna inverse folding and molecular design. *Journal of bioinformatics and computational biology*, 11(02):1350001, 2013.

Hans-Christof Gasser, Diego A Oyarzún, Javier Antonio Alfaro, and Ajitha Rajan. Tuning protein-mpnn to reduce protein visibility via mhc class i through direct preference optimization. *Protein Engineering, Design and Selection*, 38:gzaf003, 2025.

Itai Gat, Tal Remez, Neta Shaul, Felix Kreuk, Ricky TQ Chen, Gabriel Synnaeve, Yossi Adi, and Yaron Lipman. Discrete flow matching. *Advances in Neural Information Processing Systems*, 37: 133345–133385, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in neural information processing systems*, volume 33, pp. 6840–6851, 2020.

Ivo L Hofacker, Walter Fontana, Peter F Stadler, L Sebastian Bonhoeffer, Manfred Tacker, and Peter Schuster. Fast folding and comparison of rna secondary structures. *Monatshefte für Chemie/Chemical Monthly*, 125(2):167–188, 1994.

Han Huang, Ziqian Lin, Dongchen He, Liang Hong, and Yu Li. Ribodiffusion: tertiary structure-based rna inverse folding with generative diffusion models. *Bioinformatics*, 40:i347–i356, 06 2024. ISSN 1367-4811.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021.

Chaitanya K. Joshi, Arian Rokkum Jamasb, Ramon Viñas Torné, Charles Harris, Simon V Mathis, Alex Morehead, Rishabh Anand, and Pietro Lio. gRNAde: Geometric deep learning for 3d RNA inverse design. In *The Thirteenth International Conference on Learning Representations*, 2025.

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*, pp. 2025–10, 2025.

Lin Li, Esther Gupta, John Spaeth, Leslie Shing, Rafael Jaimes, Emily Engelhart, Randolph Lopez, Rajmonda S Caceres, Tristan Bepler, and Matthew E Walsh. Machine learning optimization of candidate antibody yields highly diverse sub-nanomolar affinity antibody libraries. *Nature communications*, 14(1):3454, 2023.

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.

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 (1):26, 2011.

Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Divya Nori and Wengong Jin. Rnaflow: Rna structure & sequence design via inverse folding-based flow matching. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Ryan Park, Darren J Hsu, C Brian Roland, Maria Korshunova, Chen Tessler, Shie Mannor, Olivia Viessmann, and Bruno Trentini. Improving inverse folding for peptide design with diversity-regularized direct preference optimization. *arXiv preprint arXiv:2410.19471*, 2024.

Blaine A Pfeifer, Marie Beitelshees, Andrew Hill, Justin Bassett, and Charles H Jones. Harnessing synthetic biology for advancing rna therapeutics and vaccine design. *NPJ Systems Biology and Applications*, 9(1):60, 2023.

Christian Reidys, Peter F. Stadler, and Peter Schuster. Generic properties of combinatory maps: Neutral networks of rna secondary structures. *Bulletin of Mathematical Biology*, 59(2):339–397, 1997. ISSN 0092-8240.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

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*, pp. 1–12, 2024.

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

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Minghao Sun, Hanqun Cao, Zhou Zhang, Chen Wei, Liang Wang, Tianrui Jia, Zhiyuan Liu, Tianfan Fu, Xiangru Tang, Yejin Choi, et al. Ribopo: Preference optimization for structure-and stability-aware rna design. *arXiv preprint arXiv:2510.21161*, 2025.

Takeshi Tabuchi and Yohei Yokobayashi. High-throughput screening of cell-free riboswitches by fluorescence-activated droplet sorting. *Nucleic Acids Research*, 50(6):3535–3550, 03 2022. ISSN 0305-1048.

Akito Taneda. Modena: a multi-objective rna inverse folding. *Advances and Applications in Bioinformatics and Chemistry*, pp. 1–12, 2010.

Chai Discovery team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *BioRxiv*, pp. 2024–10, 2024.

Walter Thavarajah, Laura M Hertz, David Z Bushhouse, Chloé M Archuleta, and Julius B Lucks. Rna engineering for public health: innovations in rna-based diagnostics and therapeutics. *Annual review of chemical and biomolecular engineering*, 12(1):263–286, 2021.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8228–8238, 2024.

Ziwen Wang, Jiajun Fan, Ruihan Guo, Thao Nguyen, Heng Ji, and Ge Liu. Proteinzero: Self-improving protein generation via online reinforcement learning. *arXiv preprint arXiv:2506.07459*, 2025.

Michelle J. Wu, Johan O. L. Andreasson, Wipapat Kladwang, William Greenleaf, and Rhiju Das. Automated design of diverse stand-alone riboswitches. *ACS Synthetic Biology*, 8(8):1838–1846, 2019.

Yuchen Zhang and Jian Zhou. Inverse flow and consistency models. In *Forty-second International Conference on Machine Learning*. JMLR.org, 2025.

Yiran Zhu, Liyuan Zhu, Xian Wang, and Hongchuan Jin. Rna-based therapeutics: an overview and prospectus. *Cell death & disease*, 13(7):644, 2022.

Alon Ziv, Sanyuan Chen, Andros Tjandra, Yossi Adi, Wei-Ning Hsu, and Bowen Shi. Mr-flowdpo: Multi-reward direct preference optimization for flow-matching text-to-music generation. *arXiv preprint arXiv:2512.10264*, 2025.

Algorithm 1 Flow Matching Pre-training

Input: dataset $\mathcal{D} = \{(X, S)\}$, conditional velocity model v_θ
Hyperparameters: batch size, optimizer, time horizon T with default $T = 1$
repeat
  Sample minibatch $(X, S) \sim \mathcal{D}$
  Set $x_1 \leftarrow \textsc{OneHot}(S) \in \mathbb{R}^{L\times 4}$
  Sample $x_0 \sim \mathcal{N}(0, I)$ and $t \sim \mathcal{U}(0, T)$
  Normalize time $\tilde{t} \leftarrow t/T$
  Interpolate $x_{\tilde{t}} \leftarrow (1 - \tilde{t})x_0 + \tilde{t}x_1$
  Set target velocity $u_{\tilde{t}} \leftarrow x_1 - x_0$
  Compute FM loss:

$$\mathcal{L}_{\mathrm{FM}} \leftarrow \left\| v_\theta(x_{\tilde{t}}, \tilde{t} \mid X) - u_{\tilde{t}} \right\|_2^2$$

  Update θ using $\nabla_\theta \mathcal{L}_{\mathrm{FM}}$
until convergence
Output: trained teacher v_θ

A Algorithmic Details

Stage I: Flow Matching Pre-training. Algorithm 1 describes FM training used in Stage I (Section 3.1).

Teacher flow and tangent consistency. The structure-conditioned teacher defines

$$\frac{dx_t}{dt} = v_\theta(x_t, t \mid X), \qquad t \in [0, 1]. \tag{10}$$

For a fixed RNA backbone X, the chain rule gives the derivative used in Equation (4):

$$D_\theta f_\phi(x_t, t \mid X) = \nabla_x f_\phi(x_t, t \mid X)\, v_\theta(x_t, t \mid X) + \partial_t f_\phi(x_t, t \mid X). \tag{11}$$

Consistency encourages this derivative to vanish along teacher trajectories. Together with $f_\phi(x, 1 \mid X) = x$, this specifies the terminal prediction along a trajectory in the ideal exact-consistency setting. Finite training only approximates this condition. We evaluate the derivative on interpolated training states, using the frozen teacher velocity as its direction. The spatial term is computed by a Jacobian–vector product without materializing the full Jacobian.

Stage II: Continuous-time Consistency Distillation. We use a teacher-tangent residual formulation motivated by continuous-time consistency (Lu & Song, 2025). Equation 4 specifies the squared-residual objective used in this manuscript; it should not be identified with the stop-gradient training objective or full stabilization recipe of that work. Algorithm 2 instantiates the distillation objective in Equation (4) and uses a padding mask to ignore non-residue positions.

Stage III: Preference Alignment for Structural Fidelity and Diversity.

B From DiffusionDPO to Preference Optimization for Probability Flow Models

DiffusionDPO (Wallace et al., 2024) develops a diffusion preference objective using denoising errors. MR-FlowDPO (Ziv et al., 2025) adapts a reference-relative logistic objective to differences in flow-matching velocity errors. These constructions motivate our use of a differentiable compatibility score; they do not establish that an sCM endpoint error is a normalized sequence log-likelihood.

Our setting differs in that one-step sCMs do not provide tractable transition densities, and sCM distillation enforces *tangent consistency* rather than likelihood maximization. To enable preference optimization, we therefore follow the common practice in diffusion or flow matching based preference learning and define a *model-induced surrogate score* from the training semantics of sCMs. Specifically, since $f_\phi(x_t(S), t \mid X)$ is trained to predict the terminal clean state x_1, we define the sCM energy

$$E_\phi(S; X, t) = \| f_\phi(x_t(S), t \mid X) - x_1(S) \|_2^2, \tag{12}$$

Algorithm 2 JVP-based Continuous-time Consistency Distillation (sCM)

Input: dataset $\mathcal{D} = \{(X, S, M)\}$, teacher velocity v_θ, student predictor f_ϕ
Hyperparameters: EMA rate μ, time weighting $w(t)$
Initialize $\bar{\phi} \leftarrow \phi$; freeze teacher parameters θ
Use endpoint predictor $f_\phi(x, t \mid X) = x + (1-t)u_\phi(x, t \mid X)$
repeat
  Sample minibatch $(X, S, M) \sim \mathcal{D}$
  Set $x_1 \leftarrow \textsc{OneHot}(S)$
  Sample $x_0 \sim \mathcal{N}(0, I)$ and $t \sim \mathcal{U}(0, 1)$
  Construct OT state $x_t \leftarrow (1-t)x_0 + tx_1$
  Compute teacher tangent $v \leftarrow v_\theta(x_t, t \mid X)$
  Compute $\partial_t f_\phi(x_t, t \mid X)$ and JVP $\nabla_x f_\phi(x_t, t \mid X)\, v$
  Form masked sCM loss on valid residues:

$$\mathcal{L}_{\text{sCM}} \leftarrow w(t)\, \frac{\left\| M \odot \left(\nabla_x f_\phi(x_t, t \mid X)\, v + \partial_t f_\phi(x_t, t \mid X) \right) \right\|_2^2}{\sum_i M_i}$$

  Update ϕ using $\nabla_\phi \mathcal{L}_{\text{sCM}}$
  EMA update $\bar{\phi} \leftarrow \mu\bar{\phi} + (1-\mu)\phi$
until convergence
Output: one-step model f_ϕ

where $x_t(S)$ is constructed by perturbing $x_1(S)$ using the OT-style path in Stage I. This energy is low when the one-step predictor maps perturbed states back to the target sequence representation and provides a practical surrogate score $\mathcal{S}_\phi = -E_\phi$ for preference learning, analogous in spirit to the surrogate scores used in DiffusionDPO and MR-FlowDPO.

Following DPO and MR-FlowDPO, we introduce a frozen reference sCM $f_{\phi_{\text{ref}}}$ and define the relative preference advantage

$$\Delta_\phi = -E_\phi(S_w; X, t) + E_\phi(S_l; X, t), \qquad \Delta_{\text{ref}} = -E_{\phi_{\text{ref}}}(S_w; X, t) + E_{\phi_{\text{ref}}}(S_l; X, t). \tag{13}$$

Substituting this relative score into the standard logistic objective of DPO yields the loss of sCM-DPO in Equation (7).

B.1 When Does Terminal-State Error Preserve Preference Ranking?

We next formalize when the L2 terminal-state prediction error used by sCM-DPO preserves the ranking required by DPO-style optimization. The statement is intentionally local to the candidate pool used for preference construction: it does not assume that the surrogate globally represents biological quality over all possible RNA sequences.

Let $m(S; X)$ denote the oracle structural score used to construct preferences, e.g., TM-score after folding. For a fixed target X, noise draw x_0 and time $t \in [0, 1)$, let $x_t(S) = (1-t)x_0 + tx_1(S)$ and define the sCM surrogate score (suppressing the fixed noise argument)

$$q_\phi(S; X, t) = -E_\phi(S; X, t) = -\| f_\phi(x_t(S), t \mid X) - x_1(S) \|_2^2. \tag{14}$$

Assumption 1 (Local monotone calibration)**.** *For each target X and sampled candidate pool $\mathcal{C}_X$, there exists a strictly increasing function $g_X : \mathbb{R} \to \mathbb{R}$ and constants $\gamma > 0$, $\epsilon \geq 0$ such that, at the fixed noise and time under consideration, for all $S \in \mathcal{C}_X$,*

$$|q_\phi(S; X, t) - g_X(m(S; X))| \leq \epsilon, \tag{15}$$

and $g_X(a) - g_X(b) \geq \gamma(a - b)$ whenever $a \geq b$.

Proposition 1 (Margin-preserving ranking)**.** *Consider a preference pair (S_w, S_l) satisfying $m(S_w; X) \geq m(S_l; X) + \delta$. Under assumption 1, if $2\epsilon < \gamma\delta$, then*

$$q_\phi(S_w; X, t) > q_\phi(S_l; X, t). \tag{16}$$

Equivalently, the L2 terminal-state energy ranks the oracle-preferred sequence lower: $E_\phi(S_w; X, t) < E_\phi(S_l; X, t)$.

Algorithm 3 Preference Alignment for One-step sCM

Input: one-step generator f_ϕ, backbone encoder, folding oracle $\mathcal{F}$, e.g., RhoFold (Shen et al., 2024)
Hyperparameters: candidates K, margin δ, inverse temperature β, entropy weight λ
Freeze reference $f_{\phi_{\text{ref}}} \leftarrow f_\phi$ and initialize $\mathcal{D}_{\text{pref}} \leftarrow \varnothing$
for each training backbone X **do**
  Generate candidates $\{S^{(k)}\}_{k=1}^{K} \sim f_\phi(\cdot \mid X)$
  Fold $\hat{X}^{(k)} \leftarrow \mathcal{F}(S^{(k)})$ and score $m^{(k)} \leftarrow \text{TM}(\hat{X}^{(k)}, X)$
  Select

$$k_w \leftarrow \arg\max_k m^{(k)}, \qquad k_l \leftarrow \arg\min_k m^{(k)}$$

  Set $S_w \leftarrow S^{(k_w)}$, $S_l \leftarrow S^{(k_l)}$, $m_w \leftarrow m^{(k_w)}$, and $m_l \leftarrow m^{(k_l)}$
  if $m_w - m_l < \delta$ **then**
    continue
  end if
  Add (X, S_w, S_l) to $\mathcal{D}_{\text{pref}}$
end for
repeat
  Sample $(X, S_w, S_l) \sim \mathcal{D}_{\text{pref}}$
  Sample $t \sim \mathcal{U}(0, 1)$ and construct (x_t, t) for S_w and S_l using the same noise
  Compute the sCM-DPO loss:

$$\mathcal{L}_{\text{sCM-DPO}} \leftarrow -\log \sigma(\beta(-\Delta_{\text{ref}} + \Delta_\phi))$$

  using Equation (7)
  Compute the entropy regularizer:

$$\mathcal{L}_{\text{ent}} \leftarrow -\sum_{i=1}^{L} H(p_\phi(s_i \mid X))$$

  Compute total loss:

$$\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{sCM-DPO}} + \lambda \mathcal{L}_{\text{ent}}$$

  Update parameters:

$$\phi \leftarrow \phi - \eta \nabla_\phi \mathcal{L}_{\text{total}}(\phi)$$

until convergence
Output: preference-aligned one-step generator f_ϕ

Proof. By assumption 1,

$$\begin{aligned} q_\phi(S_w; X, t) - q_\phi(S_l; X, t) &\geq g_X(m(S_w; X)) - g_X(m(S_l; X)) - 2\epsilon \\ &\geq \gamma\,(m(S_w; X) - m(S_l; X)) - 2\epsilon \\ &\geq \gamma\delta - 2\epsilon. \end{aligned} \tag{17}$$

When $2\epsilon < \gamma\delta$, the right-hand side is positive, hence $q_\phi(S_w; X, t) > q_\phi(S_l; X, t)$. Since $q_\phi = -E_\phi$, this is equivalent to $E_\phi(S_w; X, t) < E_\phi(S_l; X, t)$. □

Corollary 1 (Monotone DPO signal)**.** *For a fixed reference model, the sCM-DPO loss for one pair is*

$$\ell(\phi) = -\log \sigma\left(\beta\left[q_\phi(S_w; X, t) - q_\phi(S_l; X, t) - \Delta_{\text{ref}}\right]\right), \tag{18}$$

where Δ_{ref} is constant with respect to ϕ. Its derivative with respect to the current score margin $q_\phi(S_w; X, t) - q_\phi(S_l; X, t)$ is

$$\frac{\partial \ell}{\partial\left(q_\phi(S_w; X, t) - q_\phi(S_l; X, t)\right)} = -\beta\,\sigma\left(-\beta\left[q_\phi(S_w; X, t) - q_\phi(S_l; X, t) - \Delta_{\text{ref}}\right]\right) < 0. \tag{19}$$

The scalar loss decreases as this pair's score margin increases. This is a statement about the loss as a function of its margin, not a guarantee that a neural-network parameter update increases every pair's margin or improves generated-sequence quality.

Remark 1. Proposition 1 and Corollary 1 give a sufficient characterization, not a claim of global convergence for a non-convex neural network. They justify the design choice used in sCM-DPO: after removing ambiguous pairs with TM-score gap below δ, the terminal-state L2 surrogate preserves preference ranking whenever its local calibration error is smaller than half of the transformed oracle margin. This is also why we discard low-margin pairs during preference construction. The calibration condition is an assumption, not a consequence of the endpoint loss. At $t = 1$, all candidate energies are zero by the boundary condition, so strict ranking cannot hold there. Cross-oracle evaluation and Appendix G.1 provide empirical diagnostics without establishing calibration globally.

C Additional Experimental Details and Robustness Analysis

C.1 Dataset, Sampling Budget, and Training Details

For the in-domain benchmark, we follow the official RiboDiffusion split, which contains 6506 training targets, 773 validation targets, and 1370 test targets. For OOD evaluation, we collect 60 RNA chains from 47 unique CAMEO PDB entries released between 2025-01-11 and 2026-01-10, with zero PDB-ID overlap with the training pool. All methods are evaluated with the same fixed candidate budget, $N = 16$ generated sequences per target. During preference alignment, we discard ambiguous preference pairs whose TM-score margin is smaller than $\delta = 0.1$. Unless otherwise specified, the reported main results use 50 epochs for FM/sCM training and 5 epochs for DPO alignment. The sCM distillation uses EMA and the time weighting $w(t) = 0.5 + 0.5 \sin(\pi t/2)$.

C.2 Oracle Robustness: RhoFold and Chai-lab

The main evaluation uses RhoFold for preference construction and test-time folding. Tables 2, 3, and 6 evaluate RhoFold-aligned generation with Chai-lab under the original single-run and repeated-sampling settings. Replacing the test-time folding oracle and evaluating with Chai-lab are the same intervention, not two independent protocols. A separate experiment changes the preference oracle to Chai-lab and retrains Stage III (Appendix G). All settings use 16 candidates per target.

C.3 Additional Multi-Seed Stability

The main text reports the six-seed statistical evaluation in Table 1 and the folding-oracle validation in Tables 2, 3 and 6. We keep the appendix here for training and evaluation details only.

C.4 Fixed-Budget Pass@16

Mean TM-score alone does not fully capture whether a method reliably produces high-quality candidates under a fixed screening budget. We therefore report pass@16, defined as the fraction of targets for which at least one of 16 generated candidates exceeds each TM-score threshold.

C.5 Hyperparameter Sensitivity

The main ablation in Table 5 covers the effect of the final entropy-regularized setting and its relation to the TM-score-only configuration. The final alignment configuration uses $\lambda = 0.05$ and $\beta = 0.4$. Figure 5 reports the entropy-weight sweep; $\lambda = 0.1$ is a separate setting.

C.6 Recovery Drift and OOD Structural Gain

Preference alignment optimizes structural fidelity and diversity rather than native-sequence recovery, so a natural question is whether OOD improvements simply arise from drifting away from native sequences. We test this with a target-matched delta analysis on the OOD set, comparing sCM-RNA against the vanilla distilled sCM. For each target, we compute $\Delta\mathrm{Recovery} = \mathrm{Recovery}_{\text{sCM-RNA}} - \mathrm{Recovery}_{\text{sCM}}$ and $\Delta\mathrm{TM} = \mathrm{TM}_{\text{sCM-RNA}} - \mathrm{TM}_{\text{sCM}}$. If improved OOD robustness were mainly caused by lower recovery, larger recovery drops would correlate negatively with TM-score gains. Instead, we observe a positive Pearson correlation of $r = 0.3486$ ($p = 0.0073$) and a positive Spearman correlation of $\rho = 0.3663$ ($p = 0.0047$), while the mean recovery change is slightly negative and the mean OOD TM change is positive. These positive correlations do not support a

Table 6: **OOD Chai-lab evaluation after RhoFold alignment.** Preferences are constructed with RhoFold, while test-time structures are folded and scored with Chai-lab.

Method	Rec.↑	TM↑	MFE↓	Div.↑
RiboDiffusion	0.4875 ± 0.023	0.4151 ± 0.015	$\mathbf{-38.91} \pm 2.05$	0.2948 ± 0.016
Flow Matching	0.5239 ± 0.020	0.4068 ± 0.016	-34.53 ± 1.85	0.2878 ± 0.015
sCM-RNA	$\mathbf{0.5358} \pm 0.008$	$\mathbf{0.4297} \pm 0.008$	-38.22 ± 1.10	$\mathbf{0.3206} \pm 0.009$

Table 7: **In-domain pass@16 under a fixed candidate budget.**

Method	$\text{pass@16}_{\geq 0.30}$	$\text{pass@16}_{\geq 0.50}$	$\text{pass@16}_{\geq 0.70}$	$\text{pass@16}_{\geq 0.90}$
gRNAde	0.651	0.295	0.101	0.008
RiboDiffusion	0.967	0.834	0.593	0.141
Flow Matching	0.970	0.850	0.609	0.157
sCM-RNA	**0.983**	**0.876**	**0.642**	**0.182**

simple negative association between recovery and structural gains; they do not identify the causal mechanism of the improvement.

D Auxiliary Sequence-Level Evaluation

UTR-LM (Chu et al., 2024) is specialized for 5′-UTR sequence-level prediction. We report its mean ribosome load (MRL), translation efficiency (TE), and expression level (EL) outputs as auxiliary proxy scores. They are neither experimental measurements nor validation of general ncRNA function.

For each backbone, we generate 16 candidates, select the sequence with the highest TM-score, and then apply UTR-LM. This selection does not match structural quality across methods, isolate the causal effect of diversity, or select the candidate with the best UTR-LM score. Table 4 therefore characterizes structurally selected sequences only. Higher scores cannot by themselves establish functional generalization; differences may also reflect the evaluator's domain mismatch.

E Additional Results: TM-Score and Diversity

In addition to the primary results, we provide further comparison for two RNA targets: 6ufk_C and 1osw_A. As shown in Figure 6, our method, sCM-RNA, not only achieves the highest TM-scores but also maintains competitive diversity, reinforcing the advantages of our approach over other SOTA methods.

F Teaser Data and Normalization

Figure 1 retains four separate axes rather than aggregating them into a score. For a quality metric q (recovery, TM-score, or diversity), we use $q_m / \max_j q_j$. For latency c, we use $\min_j c_j / c_m$. Thus every axis has a zero-based scale with a best observed value of one. No minimum subtraction, logarithmic transformation, or axis truncation is applied. Polygon area is not a performance measure.

Quality values are the original main-table means, except for the updated recovery estimates for gRNAde and GVP-GNN. Their original recovery standard deviations are not reused. Efficiency values are from the updated RTX 6000 protocol: 12 targets, three in each of four length bins, batch size one, two warm-up calls, and 15 timed calls. Latency is the equally weighted mean of the four bin means; FLOPs are the reported overall profiler GFLOPs. Unsupported operators are omitted, so profiler counts are lower bounds. The sCM timing measures the one-step generator used by sCM-RNA. Generation timing excludes folding and preference-data construction.

These quality and efficiency measurements use different target populations and are presented as a descriptive overview. The original main-text efficiency figure and its 35× statement retain their

Table 8: **OOD pass@16 under a fixed candidate budget.**

Method	$\text{pass@16}_{\geq 0.30}$	$\text{pass@16}_{\geq 0.50}$	$\text{pass@16}_{\geq 0.70}$
gRNAde	0.477	0.091	0.000
RiboDiffusion	0.750	0.341	0.205
Flow Matching	0.795	0.386	0.295
sCM-RNA	**0.818**	**0.454**	**0.341**

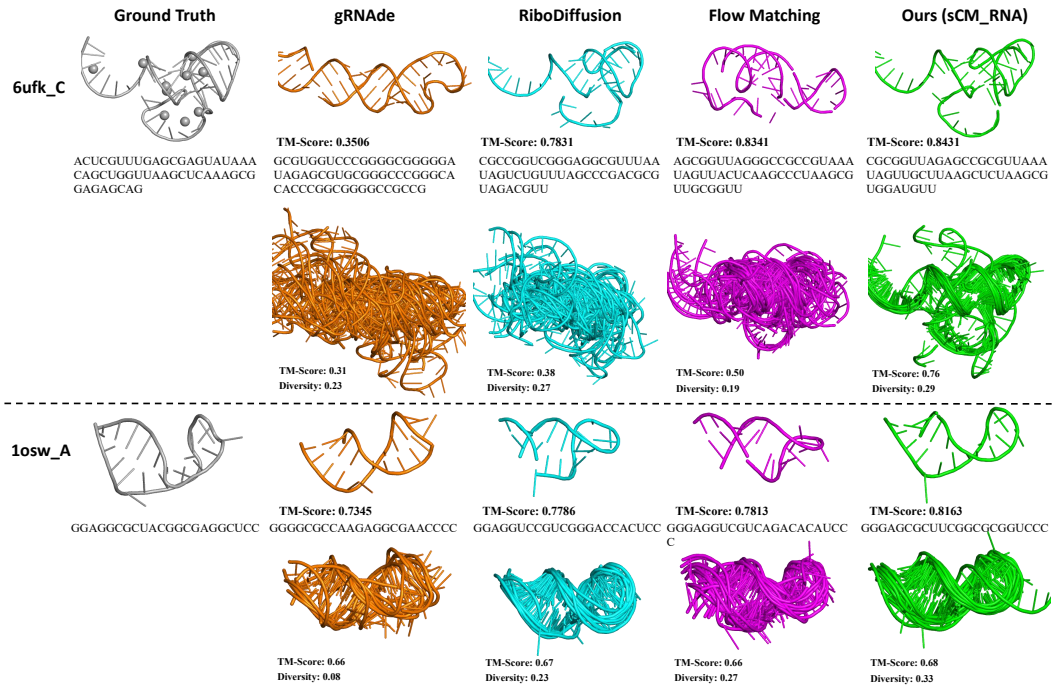


Figure 6: Comparison of RNA folding across multiple methods for two target structures: *6ufk_C* and *1osw_A*. The TM-scores and diversity of generated sequences are shown for each model.

original A6000 measurements; they should not be numerically combined with this updated RTX 6000 protocol.

G Additional Robustness Diagnostics

G.1 Endpoint Energy and Preference Ranking

We evaluate the compatibility surrogate using the same best/worst selection among 16 candidates and TM-score margin of at least 0.1 used for preference construction. For each retained target, we draw one shared Gaussian noise sample z and evaluate $f_\phi(z, 0 \mid X)$ once. The candidate-specific energies are $\|f_\phi(z, 0 \mid X) - x_1(S_w)\|_2^2$ and $\|f_\phi(z, 0 \mid X) - x_1(S_l)\|_2^2$. The fraction with lower energy for the oracle-preferred candidate is 0.760, with target-bootstrap 95% CI $[0.600, 0.920]$.

This diagnostic supports a local pairwise interpretation at $t = 0$ under high-margin best/worst selection. It does not establish calibrated sequence likelihood, agreement over arbitrary candidate pairs, or ranking preservation throughout the training time distribution. The sufficient condition in Appendix B.1 is an assumption-dependent statement, not a convergence guarantee.

Table 9: Raw teaser values and complementary compute measurements. Latency is in milliseconds per generation call; GFLOPs are reported for efficiency analysis and are not plotted in the teaser.

Method	Recovery	TM-score	Diversity	Latency (ms)	GFLOPs
gRNAde	0.4494	0.2932	0.2417	1414.2075	6.61
GVP-GNN	0.4609	0.4101	0.2253	1602.8450	2.95
RiboDiffusion	0.5004	0.5993	0.2859	2040.7800	1415.36
Flow Matching	0.5113	0.5875	0.3040	1226.9825	2488.57
sCM-RNA	0.5217	0.6131	0.3370	24.8300	49.77

G.2 Changing the Preference Oracle

Replacing RhoFold with Chai-lab for preference construction and retraining Stage III gives ID/OOD TM-scores of 0.6072/0.4402 under Chai-lab evaluation, compared with 0.6076/0.4376 for the original RhoFold-aligned setting. The Chai-lab-aligned model has ID/OOD recovery of 0.5301/0.5498, MFE of $-17.98/-37.18$, and diversity of 0.3502/0.3308. These are separate point estimates, not replacements for the original multi-seed results. Changing the alignment oracle provides additional evidence against purely RhoFold-specific gains, but does not eliminate shared predictor biases.

G.3 Non-Neural Secondary-Structure Evaluation

Using PDB-derived target base pairs and ViennaRNA 2.7.2 on 678 targets, sCM-RNA achieves MCC 0.5500 $[0.5354, 0.5646]$, compared with RiboDiffusion's 0.5405 $[0.5195, 0.5614]$ and the FM teacher's 0.5368 $[0.5225, 0.5509]$. Intervals are target-bootstrap 95% confidence intervals. The paired MCC differences are $+0.0095$ $[0.0032, 0.0159]$ versus RiboDiffusion and $+0.0132$ $[0.0037, 0.0230]$ versus FM. This supplies complementary secondary-structure evidence; it does not establish oracle-independent 3D folding correctness.