Structure-based RNA Design by Step-wise Optimization of Latent Diffusion Model

Qi Si[1*], Xuyang Liu[1*], Penglei Wang[2*], Xin Guo[1†], Yuan Qi[1,3,4], Yuan Cheng[1,3†]

[1]Shanghai Academy of Artificial Intelligence for Science
[2]School of Biomedical Engineering, Shanghai Jiao Tong University
[3]Artificial Intelligence Innovation and Incubation Institute, Fudan University
[4]Zhongshan Hospital, Fudan University
sqwd0616@gmail.com, wangpenglei@sjtu.edu.cn

Abstract

RNA inverse folding, designing sequences to form specific 3D structures, is critical for therapeutics, gene regulation, and synthetic biology. Current methods, focused on sequence recovery, struggle to address structural objectives like secondary structure consistency (SS), minimum free energy (MFE), and local distance difference test (LDDT), leading to suboptimal structural accuracy. To tackle this, we propose a reinforcement learning (RL) framework integrated with a latent diffusion model (LDM). Drawing inspiration from the success of diffusion models in RNA inverse folding, which adeptly model complex sequence-structure interactions, we develop an LDM incorporating pre-trained RNA-FM embeddings from a large-scale RNA model. These embeddings capture co-evolutionary patterns, markedly improving sequence recovery accuracy. However, existing approaches, including diffusion-based methods, cannot effectively handle non-differentiable structural objectives. By contrast, RL excels in this task by using policy-driven reward optimization to navigate complex, non-gradient-based objectives, offering a significant advantage over traditional methods. In summary, we propose the Step-wise Optimization of Latent Diffusion Model (SOLD), a novel RL framework that optimizes single-step noise without sampling the full diffusion trajectory, achieving efficient refinement of multiple structural objectives. Experimental results demonstrate SOLD surpasses its LDM baseline and state-of-the-art methods across all metrics, establishing a robust framework for RNA inverse folding with profound implications for biotechnological and therapeutic applications.

Code — https://github.com/darkflash03/SOLD

Introduction

The RNA inverse folding task involves designing RNA sequences that fold into specific 3D structures, holds significant potential for applications in RNA therapeutics, gene regulation, and synthetic biology (Damase et al. 2021). For instance, rationally designed riboswitches can enable precise control of mRNA translation, paving the way for targeted therapies (Mustafina, Fukunaga, and Yokobayashi 2019).

Existing RNA inverse folding methods primarily include physics-based and deep learning-based methods. Physics-based approaches like Rosetta (Leman et al. 2020). generate sequences via Monte Carlo optimization but are computationally expensive and struggle with polymorphic conformations Algorithms based on 2D structures, such as ViennaRNA (Churkin et al. 2018), achieve computational efficiency but neglect critical 3D geometric information, which restricts their overall applicability. In recent years, deep learning-based methods have gained traction, with generative models showing promise. Unlike variational autoencoders, as RhoDesign (Wong et al. 2024) and RDesign (Tan et al. 2024), which struggle with complex distributions and long-range dependencies, diffusion models excel at capturing intricate sequence-structure interactions through iterative denoising. RiboDiffusion (Huang et al. 2024a) proposed a diffusion-based framework for RNA inverse folding, but it depends on direct sequence modeling, which may not fully capture the rich co-evolutionary context inherent in RNA sequences.

RL has recently been widely applied to optimize the generation objectives of diffusion models. For instance, DDPO (Black et al. 2023) and DPOK (Fan et al. 2023) combined diffusion models with policy optimization to optimize decision-making in continuous action spaces. These studies demonstrate that RL can guide diffusion models to optimize objectives that are difficult to address directly with traditional methods by designing specific reward functions. Therefore, RL is particularly well-suited for diffusion models in RNA inverse folding, as it effectively optimizes non-differentiable structural metrics. Despite its potential, RL remains underexplored in this area.

To address the above limitations, we propose SOLD, a LDM based framework for RNA inverse folding. SOLD leverages the embedding space of the RNA-FM (Chen et al. 2022) to capture co-evolutionary information. Furthermore, SOLD introduces a novel step-wise RL optimization strategy that directly optimizes structural metrics (SS, MFE, LDDT), where step-wise refers to starting from any noise time step and independently optimizing each reverse sampling step. This approach ensures both computational ef-

[*]These authors contributed equally.
[†]Corresponding author.

ficiency and superior performance in generating RNA sequences that align with target structures. Our main contributions are as follows:

- A LDM that leverages pre-trained LLM embedding to improve sequence recovery in RNA design.
- The first integration of RL into LDM for RNA inverse folding, enabling effective optimization of complex structural objectives.
- A step-wise RL optimization algorithm that achieves faster convergence and superior performance.

Related Work

RNA Inverse Folding Methods

RNA inverse folding designs sequences for specified structures, categorized into physics-based, heuristic, and deep learning approaches. Physics-based Rosetta (Das, Karanicolas, and Baker 2010; Leman et al. 2020) uses Monte Carlo sampling with energy functions but is computationally intensive. Early 2D structure-based approaches such as RNAinverse (Hofacker et al. 1994) and ViennaRNA (Lorenz et al. 2011) were dynamically programmed directly on Rfam (Griffiths-Jones et al. 2003), thus ignoring 3D structural information. Among the heuristics, INFO-RNA (Busch et al. 2006) performs a direct random search of the structure, whereas methods such as RNASSD (Andronescu et al. 2004), antaRNA (Kleinkauf, Mann, and Backofen 2015), aRNAque (Merleau and Smerlak 2022), eM2dRNAs (Álvaro Rubio-Largo et al. 2023), and MCTS-RNA (Yang, Yoshizoe, and Tsuda 2017) use global search or evolutionary approaches.

Deep learning methods such as gRNAde (Joshi et al. 2025) is a concurrent graph-based RNA refolding method, while RhoDesign (Wong et al. 2024), RDesign (Tan et al. 2024), PiFold (Gao, Tan, and Li 2023), StructGNN (Chou et al. 2024), and GVP-GNN (Jing et al. 2021) are representative deep learning methods for inverse folding, which are modified here for RNA compatibility. Recently, RISoTTo (Bibekar, Krapp, and Dal Peraro 2025) proposed a context-aware geometric deep learning framework that integrates structural context into RNA sequence design. While these deep learning approaches enhance the quality of generated sequences, they are unable to directly optimize structure metrics, such as MFE and LDDT.

Diffusion Models in Molecular Generation

RiboDiffusion (Huang et al. 2024a) applies 3D diffusion on PDB dataset for RNA inverse folding. DRAKES (Wang et al. 2024) fine-tunes discrete diffusion with reinforcement learning, optimizing DNA enhancer activity and protein stability. GradeIF (Yi et al. 2023) leverages graph denoising with BLOSUM matrices, enhancing protein sequence diversity and recovery. RNAdiffusion (Huang et al. 2024b) uses latent diffusion to generate RNA sequences, optimizing the translation efficiency. Structured DDPM (Austin et al. 2021) and Dirichlet Flow (Stark et al. 2024) explored discrete diffusion on molecular data. RNAFlow (Chen et al. 2024) integrates 2D/3D diffusion in RNA structures. These models address diverse molecular design challenges, from structure prediction to functional optimization.

Reinforcement Learning for Generative Model Optimization

RL has emerged as a powerful approach for optimizing generative models, particularly diffusion models, by aligning them with diverse objectives. Online RL methods like PPO (Schulman et al. 2017) and RAFT (Dong et al. 2023) leverage direct reward optimization, while offline methods such as DPO (Rafailov et al. 2023) utilize preference datasets for policy alignment.

Meanwhile, alignment of difffusion models to preferences by reinforcement learning has also been widely explored. DDPO (Black et al. 2023) and DPOK (Fan et al. 2023) use the PPO algorithm to enhance image quality in diffusion models; Diffusion-DPO (Wallace et al. 2023) adapts DPO to fine-tune Stable Diffusion XL with human preferences. These studies demonstrate the ability of RL to guide models toward optimizing objectives that traditional methods struggle to address directly.

Comparison with Our Work

SOLD integrates latent diffusion and RL to advance RNA inverse folding, offering some significant advantages over existing methods:

- Unlike RNAdiffusion, which generates sequences without structural constraints and optimizes functional metrics using separately trained reward models, SOLD focuses on structure-based RNA design, directly optimizing structural metrics. By leveraging ViennaRNA for direct evaluation, SOLD eliminates the need for additional reward models, reducing computational overhead and mitigating inaccuracies from reward model predictions.
- DRAKES utilizes discrete diffusion with RL to optimize a single objective for RNA design, relying on differentiable reward models that introduce additional training costs and the potential risk of over-optimization. However, SOLD operates in a continuous latent space for diffusion, enabling efficient optimization of multiple RNA-specific metrics without the need for separately trained reward models.
- In contrast to DDPO and DPOK, which are designed for image and text generation and rely on sampling complete trajectories, incurring high computational costs, SOLD is tailored for RNA inverse folding, and its step-wise RL optimization iteratively refines strategies, achieving greater efficiency compared to trajectory-based methods.

Methods

Latent Diffusion Model

The SOLD employs a latent diffusion model (LDM) to generate RNA sequences, integrating pre-trained RNA-FM to extract embeddings with geometric information from the backbone, producing sequences that conform to specific structural and functional properties as Figure 1. We map RNA sequences directly into variable-length embeddings of

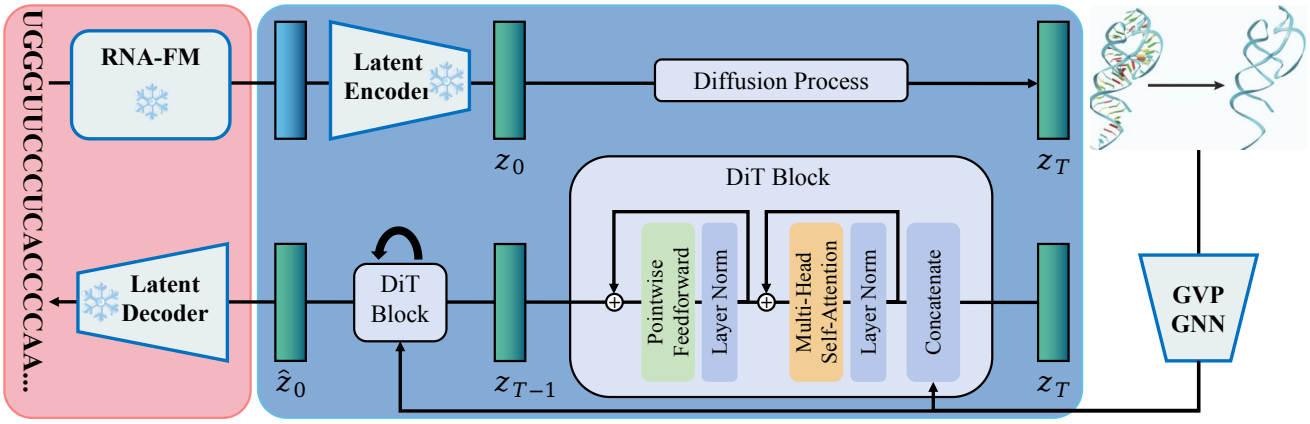

Figure 1: The LDM encodes RNA-FM embeddings into a latent representation, performs denoising with GVP-GNN + DiT blocks, and decodes the refined latent into RNA sequences consistent with structural constraints.

shape $(L, 640)$, where L represents the sequence length. These embeddings are compressed into a latent representation of shape (L, D) via an MLP encoder to improve the efficiency of the diffusion process.

A denoising network π_θ, integrated with GVP-GNN (Jing et al. 2021) and Diffusion Transformer (DiT) (Peebles and Xie 2023), is used to predict the denoised latent embedding $\hat{z}_0$ at time $t = 0$. The MLP decoder reconstructs the latent embedding $\hat{z}_0$ into a sequence probability distribution $(L, 4)$, generating sequences of length L that align with the input backbone structure.

The forward process of diffusion introduces Gaussian noise to the latent embedding $z_0 \in \mathbf{R}^{L\times D}$:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t}z_{t-1}, \beta_t I) \quad (1)$$

where β_t is the noise schedule, and $t = 1, \ldots, T$. The reverse process reconstructs the embedding by predicting $\hat{z}_0 = \pi_\theta(z_t, t, c)$:

$$p_\theta(z_{t-1}|z_t, c) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t, c), \sigma_t^2 I) \quad (2)$$

with the mean defined as:

$$\mu_\theta(z_t, t, c) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \hat{z}_0 \quad (3)$$

where c represents the geometric information from backbone, $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, and the variance is defined as $\sigma_t^2 = \frac{(1-\bar{\alpha}_{t-1})(1-\alpha_t)}{1-\bar{\alpha}_t}$.

The training objective combines MSE and cross-entropy:

$$\mathcal{L} = \mathrm{E}_{z_0,t}\left[\|z_0 - \hat{z}_0\|^2\right] - \mathrm{E}_s\left[\sum_{i=1}^{L} \log p(s_i \mid \mathrm{Dec}(\hat{z}_0)_i)\right] \quad (4)$$

The first term of the loss function, denoted as the mean squared error (MSE) term, quantifies the difference between the true latent embedding z_0 and the single-step predicted embedding $\hat{z}_0$, enhancing the precision of denoise network in predicting the clean embedding. The second term referred to the cross-entropy term, ensures that the sequence decoded from the final denoised embedding $\hat{z}_0$ closely aligns with the target sequence s, guiding the frozen decoder to generate accurate nucleotide probability distributions.

The encoder efficiently compresses RNA-FM embeddings into the latent space, thereby facilitating the diffusion process.

$$z_0 = \mathrm{Enc}_\psi(h) = \mathrm{MLP}_\psi(h) \quad (5)$$

where $h \in \mathbf{R}^{L\times 640}$ represents the input embedding, and $z_0 \in \mathbf{R}^{L\times D}$ is the compressed latent representation. The decoder generates sequence probabilities by:

$$p(s) = \mathrm{Dec}_\phi(\hat{z_0}) = \mathrm{softmax}(\mathrm{MLP}_\phi(\hat{z_0})) \quad (6)$$

where $\hat{z_0} \in \mathbf{R}^{L\times D}$ is the final denoised latent embedding, and $p(s) \in \mathbf{R}^{L\times 4}$ represents the probability distribution over the four nucleotide bases (A, U, C, G).

Step-wise Optimization of Latent Diffusion Model

The SOLD algorithm refines the pre-trained LDM through a RL framework, targeting complex objectives critical for RNA inverse folding, such as SS, MFE, and LDDT. The overall architecture is depicted in Figure 2. SOLD adopts a single-step sampling strategy inspired by Denoising Diffusion Implicit Models (DDIM) (Song, Meng, and Ermon 2020) to enhance computational efficiency. In DDIM, the reverse process is approximated with a deterministic mapping, enabling single-step predictions of the target latent embedding $\hat{z}_0$ from a noisy sample z_t.

SOLD's single-step prediction leverages the Denoising Diffusion Implicit Model (DDIM) framework to efficiently generate denoised latent embeddings. The sampling process is defined as:

$$z'_{t-k} = \sqrt{\bar{\alpha}_{t-k}}\,\hat{z}_0 + \gamma_{t-k}\,\epsilon_\theta(z_t, t, c) + \sigma_{t-k}\,\epsilon \quad (7)$$

where $\hat{z}_0 = \pi_\theta(z_t, t, c)$ is the predicted target embedding at $t = 0$, $\gamma_{t-k} = \sqrt{1-\bar{\alpha}_{t-k}-\sigma_{t-k}^2}$ is a shorthand

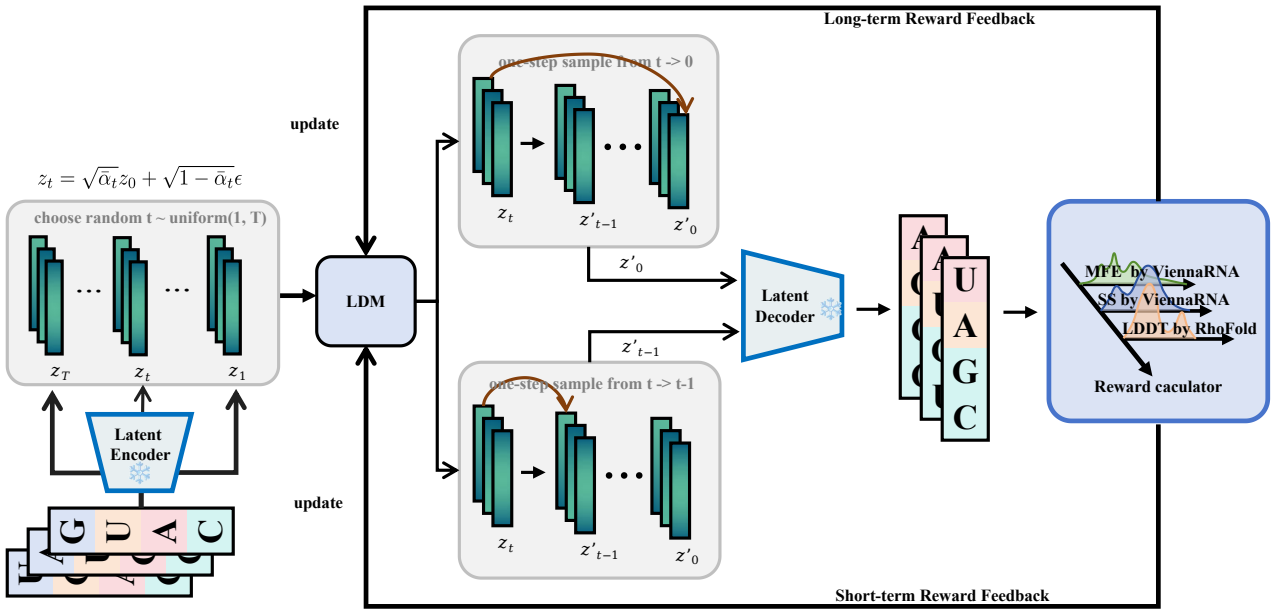

Figure 2: An overview of the SOLD. SOLD utilizes long-term and short-term reward feedback to directly optimize trained latent diffusion model in a random denoising step.

coefficient, $\epsilon_\theta(z_t, t, c) = \frac{z_t - \hat{z}_0\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}}$ is the predicted noise conditioned on structural constraints c, and $\epsilon \sim \mathcal{N}(0, I)$ is standard Gaussian noise. The stochasticity is controlled by $\sigma^2_{t-k} = \eta \cdot \frac{(1-\bar{\alpha}_{t-k})(1-\alpha_t)}{1-\bar{\alpha}_t}$, with hyperparameter η. When $k = t$, SOLD can sample z'_0 in a single step, significantly enhancing computational efficiency. By decoding z'_0 into a sequence s'_0, SOLD evaluates a long-term reward $r_0(t) = R_i(s'_0)$, where R_i assesses RNA-specific objectives.

By randomly sampling a timestep t and directly predicting z'_0 from z_t, SOLD eliminates the need to generate complete trajectories during training, reducing computational complexity. Notably, while SOLD uses single-step sampling for training efficiency, it retains full trajectory denoising during inference to ensure high-quality sequence generation.

In early denoising steps (large t), the smaller $\bar{\alpha}_t$ increases the variance of the noise term, reducing the accuracy of direct z'_0 predictions and the reliability of long-term rewards $r_0(t)$. To address this, SOLD computes short-term rewards by equation 7 with sampling strategy ($k = 1$) to predict the intermediate latent embedding z'_{t-1} from z_t. By decoding z'_{t-1} into a sequence s'_{t-1}, SOLD computes a short-term reward $r_t(t) = R_i(s'_{t-1})$, using the same reward function R_i as the long-term reward. This short-term rewards effectively guide learning during early denoising steps where high noise levels diminish the reliability of long-term rewards.

To balance the strengths of both rewards, SOLD integrates long-term and short-term rewards through a piecewise reward function:

$$r_{\text{total}}(t) = w(t) \cdot r_t(t) + u(t) \cdot r_0(t) \tag{8}$$

where $w(t)$ and $u(t)$ are time-dependent weighting functions. In early denoising steps (large t), $w(t) = 1$ and $u(t) = 0$, prioritizing the short-term reward $r_t(t)$ to guide learning. In later denoising steps (small t), $w(t) = 0$ and $u(t) = 1$, emphasizing the long-term reward $r_0(t)$. This piecewise strategy enhances SOLD's flexibility in optimizing RNA-specific objectives across different denoising stages.

SOLD optimizes the model parameters θ using Proximal Policy Optimization (PPO) (Schulman et al. 2017) based on the total reward $r_{\text{total}}(t)$, with the objective:

$$\mathcal{J}_{\text{SOLD}}(\theta) = \mathrm{E}_{t\sim\mathcal{U}[1,T],c,z_t,z'_0\sim p_\theta(z'_0|z_t,c)}\left[r_{\text{total}}(t)\right] \tag{9}$$

where the policy gradient is computed via the REINFORCE algorithm with importance sampling:

$$\nabla_\theta \mathcal{J}_{\text{SOLD}} = \mathrm{E}_{p_{\text{old}}(t)}\left[w \nabla_\theta \log p_\theta(z'_0 \mid z_t, t, c) r(z_t)\right] \tag{10}$$

where the importance weight is:

$$w = \frac{p_\theta(x'_0 \mid x_t, t, c)}{p_{\text{ref}}(x'_0 \mid x_t, t, c)} \tag{11}$$

with $p_{\text{ref}}(z'_0 \mid z_t, t, c)$ representing the conditional probability, under the previous iteration's model parameters θ_{ref}.

SOLD adopts a clipped surrogate objective, with a clip range of $\epsilon = 0.0001$, to ensure stable policy updates. Additionally, SOLD incorporates a constraint to limit divergence between the current policy p_θ and the previous policy p_{ref}:

$$\mathcal{J}_{\text{ref}}(\theta) = -D_{\text{KL}}(p_\theta \| p_{\text{ref}}) \tag{12}$$

Finally, the total optimization objective is:

$$\mathcal{J}^{\text{reg}}_{\text{SOLD}}(\theta) = \mathcal{J}_{\text{SOLD}}(\theta) + \lambda_{\text{ref}} \mathcal{J}_{\text{ref}}(\theta) \tag{13}$$

where λ_{ref} is the regularization weight for the reference policy. This integration of single-step sampling, piecewise reward framework, and KL-constrained optimization enables SOLD to effectively tackle complex RNA inverse folding design.

	SOLD TEST		CASP15 TEST	
Method	Sequence Recovery	NT Recovery	Sequence Recovery	NT Recovery
RhoDesign	0.2734	0.2859	0.2606	0.2575
RDesign	0.4457	0.3966	0.3264	0.3251
gRNAde	0.5108	0.4890	0.5097	0.5149
RiboDiffusion	0.5125	0.4416	0.5388	0.3871
DRAKES-Pretrain	0.4524	0.4088	0.3357	0.3374
LDM	**0.5728**	**0.5034**	**0.5462**	**0.5473**

Table 1: Sequence Recovery and NT Recovery Comparison

Experiment

To evaluate the performance of SOLD for RNA inverse folding, we conducted a series of experiments. Our experiments leverage a high-quality dataset constructed from the RCSB Protein Data Bank, RNAsolo, and CASP15 RNA dataset, as detailed in Appendix A. The dataset was pre-processed by clustering RNA structural data using PSI-CD-HIT (sequence threshold: 0.3) and US-align (structural threshold: 0.45). After pre-processing, we obtained 8222 structures, split into pre-training (7067 structures), RL fine-tuning (389 structures), and SOLD TEST (766 structures) datasets, with deduplication to prevent information leakage. The CASP15 TEST dataset served as an independent test benchmark.

We compared SOLD against state-of-the-art (SOTA) methods, including RhoDesign (Wong et al. 2024), RDesign (Tan et al. 2024), gRNAde (Joshi et al. 2025), RiboDiffusion (Huang et al. 2024a), and DRAKES (Wang et al. 2024). Notably, DRAKES-Pretrain refers to the pre-trained DRAKES model without RL finetuning, while DRAKES incorporates optimization for a single metric (MFE), using a discrete diffusion framework with differentiable reward models. In contrast, SOLD employs LDM with stepwise RL optimization, targeting multiple structurally relevant metrics without any differentiable reward model.

The experiments are structured in three parts: (1) evaluating the LDM's superiority in 1D metrics (Sequence Recovery, Nucleotide (NT) Recovery) to test its foundational generative capability; (2) assessing SOLD's RL finetuning for single-objective optimization, evaluating its speed and effectiveness in optimizing 2D (SS, MFE) and 3D (LDDT) metrics; and (3) validating the effectiveness of SOLD's weighted multi-objective optimization across 1D (Sequence Recovery), 2D (SS, MFE), and 3D (RMSD, LDDT) metrics, with a practical case study to examine SOLD's real-world utility in RNA design. All experiments were conducted on a single A100 GPU.

LDM Performance on Sequence Generation

To assess the effectiveness of the LDM backbone in SOLD, we evaluated its sequence generation performance using 1D metrics, specifically Sequence Recovery and NT Recovery, which measure the similarity of generated sequences to target sequences and the accuracy of nucleotide composition, respectively. These metrics focus on the model's ability to predict the correct nucleotide at each position, with structural information serving as input features that implicitly influence sequence generation rather than as explicit optimization targets. As the pre-training phase of LDM focuses on optimizing sequence prediction rather than directly targeting 2D or 3D structural metrics, we evaluate performance only using 1D metrics. Table 1 presents the results on the SOLD TEST and CASP15 TEST datasets, comparing SOLD's LDM against RhoDesign, RDesign, gRNAde, RiboDiffusion, and DRAKES-Pretrain. Since this evaluation precedes RL finetuning, we compare against DRAKES-Pretrain to ensure a fair assessment of pre-trained models. LDM consistently outperforms competitors, showing higher Sequence Recovery and NT Recovery across both test datasets. Compared with RiboDiffusion, which performs SDE-based diffusion directly in the one-hot RNA sequence space, our LDM operates in the RNA-FM embedding space while keeping the backbone architecture unchanged. This latent representation captures richer structural and co-evolutionary features, leading to markedly improved sequence recovery. This method appears to enhance sequence prediction accuracy while maintaining computational efficiency, establishing a strong foundation for subsequent RL finetuning.

RL Finetuning for Single-Objective Optimization

We evaluated SOLD's RL finetuning stage for single-objective optimization, targeting on 2D (SS, MFE) and 3D (LDDT) metrics, which are critical for RNA functionality but challenging to optimize directly in SOTA RNA design algorithms. Table 2 presents the results of single-objective optimization for MFE, SS, and LDDT on the SOLD TEST and CASP15 TEST datasets, comparing SOLD against its LDM baseline, DDPO, and DPOK. For MFE optimization, we additionally compared the DRAKES-Pretrain and DRAKES methods to elucidate the impact of reinforcement learning. As DRAKES exclusively optimizes MFE, it was not included in comparisons for SS and LDDT.

For MFE, SOLD demonstrates a better improvement over its LDM baseline compared to the improvement of DRAKES over DRAKES-Pretrain, while also achieving superior final performance. Across all objectives (MFE, SS, LDDT), SOLD consistently improves upon LDM, with varying degrees of enhancement depending on the metric. Figure 3 illustrates the rapid convergence of SOLD compared to DDPO and DPOK, highlighting its efficiency in optimizing these structurally relevant metrics. Table 3 further quantifies this efficiency, revealing that SOLD completes a single training round for each objective much faster than

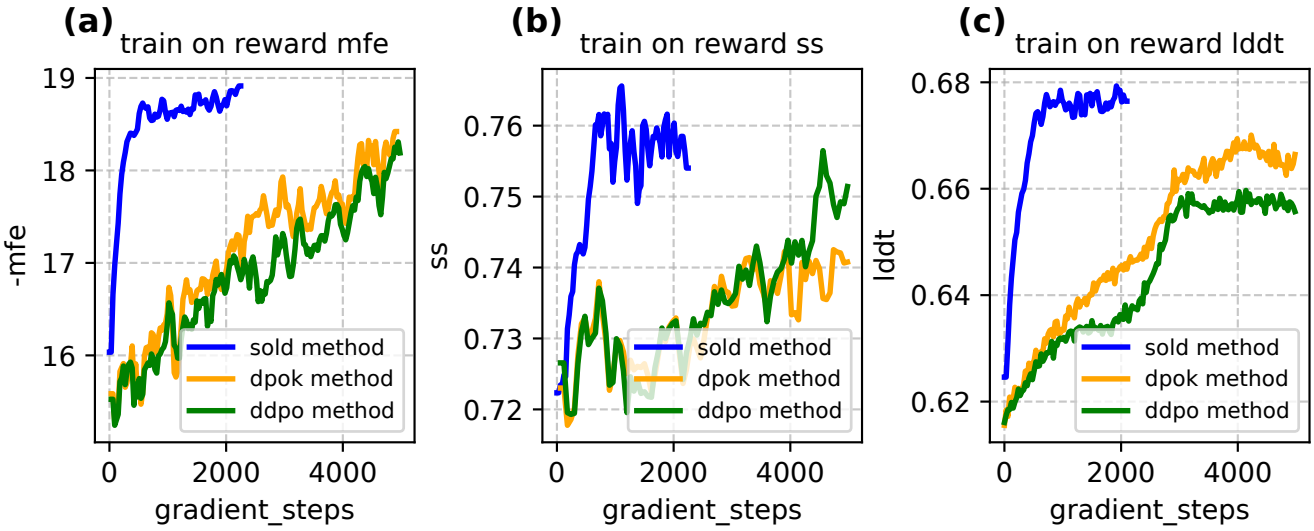

Figure 3: rewards for SOLD, DPOK, and DDPO with MFE, SS, and LDDT as reward objectives: (a) MFE, (b) SS, (c) LDDT.

Method	SOLD TEST	CASP15 TEST
MFE Training (MFE reward) ↓		
DRAKES-Pretrain	-12.5123	-52.6827
DRAKES	-14.2374	-61.0354
LDM	-13.1519	-52.7387
DDPO	-18.7498	-63.9567
DPOK	-17.4660	-67.7949
SOLD	**-19.7428**	**-68.2100**
SS Training (SS reward) ↑		
LDM	0.7274	0.5543
DDPO	**0.7595**	0.6649
DPOK	0.7511	0.6303
SOLD	0.7551	**0.7010**
LDDT Training (LDDT reward) ↑		
LDM	0.6184	0.3237
DDPO	0.6286	0.3406
DPOK	0.6329	0.3351
SOLD	**0.6384**	**0.3548**

Table 2: Single-Objective Reward Comparison

DDPO and DPOK. For LDDT, SOLD's training speed is slightly slower due to the computational bottleneck of structure prediction, but it still remains competitive. Table 2 further confirms that the performance of SOLD on MFE, SS, and LDDT is generally on par with or superior to DDPO and DPOK, demonstrating its effectiveness alongside its speed advantage.

SOLD's step-wise RL optimization, utilizing single-step sampling, accelerates convergence compared to trajectory-based methods like DDPO and DPOK. Additionally, the integration of ViennaRNA for direct reward evaluation eliminates the need for separate reward models, enhancing both efficiency and effectiveness in optimizing these challenging metrics.

Method	MFE (s)	SS (s)	LDDT (s)
DDPO	5953	6190	14000
DPOK	7677	7330	14200
SOLD	**256**	**263**	**6900**

Table 3: Average training time per epoch for DDPO, DPOK, and SOLD across different objectives

Multi-Objective Optimization Performance

To validate SOLD's capability for weighted multi-objective optimization, we conducted experiments targeting the simultaneous improvement of 1D, 2D, and 3D metrics. To ensure generality, we applied this weighting without tuning hyperparameters. We optimize SS, MFE, and LDDT using an equal weighting scheme, addressing the differing scales by mapping MFE to the (0,1) interval with the function:

$$\text{reward}_{\text{MFE}} = \exp\left(\frac{1}{\text{MFE} - \frac{1}{4}}\right) \tag{14}$$

Table 4 compares SOLD against SOTA RNA design methods on the SOLD TEST and CASP15 TEST datasets. SOLD consistently outperforms existing methods across nearly all metrics, demonstrating a superior balance between sequence naturalness and structural fidelity. Compared to its LDM baseline, SOLD maintains or improves sequence recovery with notable improvements in SS, MFE, LDDT, and RMSD. This balanced enhancement underscores SOLD's ability to optimize multiple objectives without sacrificing any single metric. The practical efficacy of SOLD is further demonstrated with a TPP-specific riboswitch case study (Figure 4). Here, SOLD successfully designed a sequence that folds into the precise target structure, whereas other methods failed, producing conformations distant from the goal. This case confirms SOLD's capacity to satisfy stringent structural and functional constraints.

Method	Sequence Recovery ↑	MFE ↓	SS ↑	RMSD ↓	LDDT ↑
SOLD TEST					
RhoDesign	0.2734	-11.9212	0.6499	16.3577	0.5031
RDesign	0.4457	-10.6990	0.6135	16.1315	0.5238
gRNAde	0.5108	-10.5409	0.5624	17.9960	0.4848
RiboDiffusion	0.5125	-15.2128	0.7632	12.3170	0.6102
DRAKES	0.4400	-14.2374	**0.7691**	11.9077	0.6191
LDM	0.5728	-13.3275	0.7269	12.5732	0.6178
SOLD	**0.5732**	**-16.8611**	0.7601	**11.8612**	**0.6360**
CASP15 TEST					
RhoDesign	0.2606	-45.9620	0.4708	30.2765	0.3306
RDesign	0.3264	-37.0679	0.4239	30.0221	0.3165
gRNAde	0.5097	-40.8491	0.4186	30.8809	0.3121
RiboDiffusion	0.5388	-41.1593	0.4699	29.5904	0.3210
DRAKES	0.3484	-61.0354	0.6412	27.1322	0.3587
LDM	0.5462	-52.7387	0.5543	29.7038	0.3237
SOLD	**0.5888**	**-64.0375**	**0.6957**	**26.8422**	**0.3680**

Table 4: Multi-Objective Performance Comparison

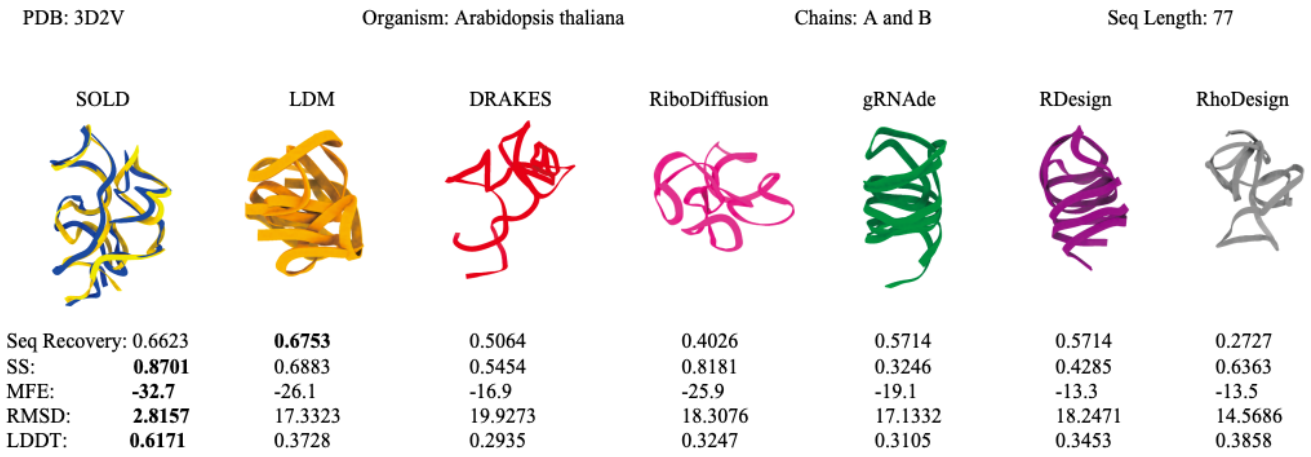

Figure 4: Comparison of rna design methods for example (PDB: 3D2V), ground-truth structure (gold), SOLD (blue)

Conclusion

In this paper, we introduce SOLD, a novel RNA inverse folding framework that integrates a latent diffusion model (LDM) with reinforcement learning (RL) for superior performance. Combining RNA-FM embeddings with 3D backbone geometric features, SOLD's LDM excels in sequence recovery, laying a strong foundation for sequence generation. A step-wise RL algorithm further enhances performance over the LDM baseline, optimizing structurally relevant metrics with improved convergence efficiency compared to RL methods requiring full trajectory sampling. This approach represents a highly competitive algorithm that balances sequence naturalness with structural fidelity, positioning SOLD as a powerful tool for RNA design with potential for applications in therapeutics and biotechnology.

However, SOLD's performance is constrained by the limited availability of high-quality RNA structural data. Additionally, we have not extensively explored how 1D, 2D, and 3D metrics interact and coordinate in RNA design. Moreover, current reward evaluation tools, such as ViennaRNA and RhoFold, inevitably introduce approximation errors that may affect optimization accuracy. Nonetheless, our proposed framework is modular and tool-agnostic — the reward component is fully pluggable, allowing seamless substitution with more accurate structure prediction or energy evaluation models as they become available. In future work, we aim to expand datasets, refine reward evaluation with improved predictors, and explore synergistic optimization of multi-scale metrics, further enhancing SOLD's robustness and generalizability for diverse RNA design challenges.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 82394432 and 92249302), and the Shanghai Municipal Science and Technology Major Project (Grant No. 2023SHZDZX02). The authors acknowledge support from the AI for Science Program, Shanghai Municipal Commission of Economy and Information.

References

Andronescu, M.; Fejes, A. P.; Hutter, F.; Hoos, H. H.; and Condon, A. 2004. A New Algorithm for RNA Secondary Structure Design. *Journal of Molecular Biology*, 336(3): 607–624.

Austin, J.; Johnson, D. D.; Ho, J.; Tarlow, D.; and van den Berg, R. 2021. Structured Denoising Diffusion Models in Discrete State-Spaces. *arXiv preprint arXiv:2107.03006*.

Bibekar, P.; Krapp, L. F.; and Dal Peraro, M. 2025. Context-aware geometric deep learning for RNA sequence design. *bioRxiv*. Preprint.

Black, K.; Janner, M.; Du, Y.; Kostrikov, I.; and Levine, S. 2023. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*.

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

Chen, C.; Zhang, H.; Feng, X.; Shen, K.; Zhang, Y.; and Li, Y. 2024. RNAFlow: RNA Structure and Sequence Design via Inverse Folding-based Flow Matching. *arXiv preprint arXiv:2405.11009*.

Chen, J.; Hu, Z.; Sun, S.; Tan, Q.; Wang, Y.; Yu, Q.; Zong, L.; Hong, L.; Xiao, J.; Shen, T.; et al. 2022. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. *arXiv preprint arXiv:2204.00300*.

Chou, Y.-T.; Chang, W.-T.; Jean, J. G.; Chang, K.-H.; Huang, Y.-N.; and Chen, C.-S. 2024. StructGNN: An efficient graph neural network framework for static structural analysis. *Computers & Structures*, 299: 107385.

Churkin, A.; Retwitzer, M. D.; Reinharz, V.; Ponty, Y.; Waldispühl, J.; and Barash, D. 2018. Design of RNAs: Comparing Programs for Inverse RNA Folding. *Briefings in Bioinformatics*, 19(2): 350–358.

Damase, T. R.; Sukhovershin, R.; Boada, C.; Taraballi, F.; Pettigrew, R. I.; and Cooke, J. P. 2021. The Limitless Future of RNA Therapeutics. *Frontiers in Bioengineering and Biotechnology*, 9: 628137.

Das, R.; Karanicolas, J.; and Baker, D. 2010. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nature Methods*, 7(4): 291–294.

Dong, H.; Xiong, W.; Goyal, D.; Zhang, Y.; Chow, W.; Pan, R.; Diao, S.; Zhang, J.; Shum, K.; and Zhang, T. 2023. Raft: Reward ranked finetuning for generative foundation model alignment. *arXiv preprint arXiv:2304.06767*.

Fan, Y.; Watkins, O.; Du, Y.; Liu, H.; Ryu, M.; Boutilier, C.; Abbeel, P.; Ghavamzadeh, M.; Lee, K.; and Lee, K. 2023. DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models. *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*. ArXiv:2305.16381.

Gao, Z.; Tan, C.; and Li, S. Z. 2023. PiFold: Toward effective and efficient protein inverse folding. In *International Conference on Learning Representations*.

Griffiths-Jones, S.; Bateman, A.; Marshall, M.; Khanna, A.; and Eddy, S. R. 2003. Rfam: an RNA family database. *Nucleic acids research*, 31(1): 439–441.

Hofacker, I. L.; Fontana, W.; Stadler, P. F.; Bonhoeffer, S.; Tacker, M.; and Schuster, P. 1994. Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte für Chemie / Chemical Monthly*, 125(9): 167–188.

Huang, H.; Lin, Z.; He, D.; Hong, L.; and Li, Y. 2024a. RiboDiffusion: Tertiary Structure-based RNA Inverse Folding with Generative Diffusion Models. *Bioinformatics*, 40(Suppl 1): i347–i356. Presented at ISMB 2024.

Huang, K.; Yang, Y.; Fu, K.; Chu, Y.; Cong, L.; and Wang, M. 2024b. Latent Diffusion Models for Controllable RNA Sequence Generation. *arXiv preprint arXiv:2409.09828*.

Jing, B.; Eismann, S.; Suriana, P.; Townshend, R. J. L.; and Dror, R. 2021. Learning from Protein Structure with Geometric Vector Perceptrons. In *International Conference on Learning Representations*.

Joshi, C. K.; Jamasb, A. R.; Viñas, R.; Harris, C.; Mathis, S. V.; Morehead, A.; Anand, R.; and Liò, P. 2025. gRNAde: Geometric Deep Learning for 3D RNA inverse design. In *International Conference on Learning Representations (ICLR)*.

Kleinkauf, R.; Mann, M.; and Backofen, R. 2015. antaRNA: ant colony-based RNA sequence design. *Bioinformatics*, 31(19): 3114–3121.

Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; et al. 2020. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature methods*, 17(7): 665–680.

Lorenz, R.; Bernhart, S. H.; zu Siederdissen, C. H.; Tafer, H.; Flamm, C.; and Hofacker, P. F. S. . I. L. 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1): 1–14.

Merleau, N. S. C.; and Smerlak, M. 2022. aRNAque: an evolutionary algorithm for inverse pseudoknotted RNA folding inspired by Lévy flights. *BMC Bioinformatics*, 23.

Mustafina, K.; Fukunaga, K.; and Yokobayashi, Y. 2019. Design of mammalian ON-riboswitches based on tandemly fused aptamer and ribozyme. *ACS Synthetic Biology*, 9(1): 19–25.

Peebles, W.; and Xie, S. 2023. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4195–4205.

Rafailov, R.; Sharma, A.; Mitchell, E.; Ermon, S.; Manning, C. D.; and Finn, C. 2023. Direct preference optimization: Your language model is secretly a reward model. *arXiv preprint arXiv:2305.18290*.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising Diffusion Implicit Models. *arXiv:2010.02502*.

Stark, H.; Jing, B.; Wang, C.; Corso, G.; Berger, B.; Barzilay, R.; and Jaakkola, T. 2024. Dirichlet Flow Matching

with Applications to DNA Sequence Design. *arXiv preprint arXiv:2402.05841*.

Tan, C.; Zhang, Y.; Gao, Z.; Hu, B.; Li, S.; Liu, Z.; and Li, S. Z. 2024. RDesign: Hierarchical Data-efficient Representation Learning for Tertiary Structure-based RNA Design. In *The Twelfth International Conference on Learning Representations*.

Wallace, B.; Dang, M.; Rafailov, R.; Zhou, L.; Lou, A.; Purushwalkam, S.; Ermon, S.; Xiong, C.; Joty, S.; and Naik, N. 2023. Diffusion Model Alignment Using Direct Preference Optimization. *arXiv preprint arXiv:2311.12908*.

Wang, C.; Uehara, M.; He, Y.; Wang, A.; Biancalani, T.; Lal, A.; Jaakkola, T.; Levine, S.; Wang, H.; and Regev, A. 2024. Fine-Tuning Discrete Diffusion Models via Reward Optimization with Applications to DNA and Protein Design. *arXiv preprint arXiv:2410.13643*.

Wong, F.; He, D.; Krishnan, A.; Hong, L.; Wang, A. Z.; Wang, J.; Hu, Z.; Omori, S.; Li, A.; Rao, J.; et al. 2024. Deep generative design of RNA aptamers using structural predictions. *Nature Computational Science*, 1–11.

Yang, X.; Yoshizoe, K.; and Tsuda, A. T. . K. 2017. RNA inverse folding using Monte Carlo tree search. *BMC Bioinformatics*, 18.

Yi, K.; Zhou, B.; Shen, Y.; Lio, P.; and Wang, Y. G. 2023. Graph Denoising Diffusion for Inverse Protein Folding. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Álvaro Rubio-Largo; Lozano-García, N.; Granado-Criado, J. M.; and Vega-Rodríguez, M. A. 2023. Solving the RNA inverse folding problem through target structure decomposition and Multiobjective Evolutionary Computation. *Applied Soft Computing*, 147: 110779.