



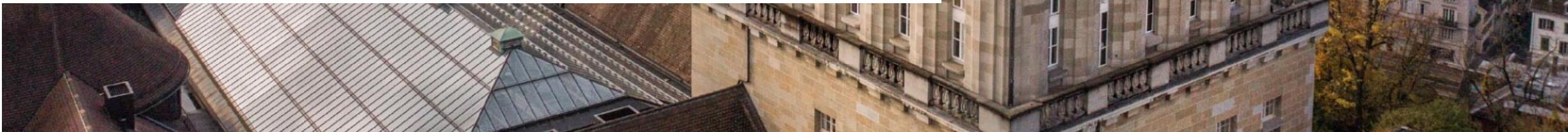
# Enhancing RNA 3D Structure Prediction

Benchmarking DDPM vs. DDIM

Diyar Taskiran, Niklas Schmidt, Elias Müller

AML: Semester Project Presentation

16. July 2025



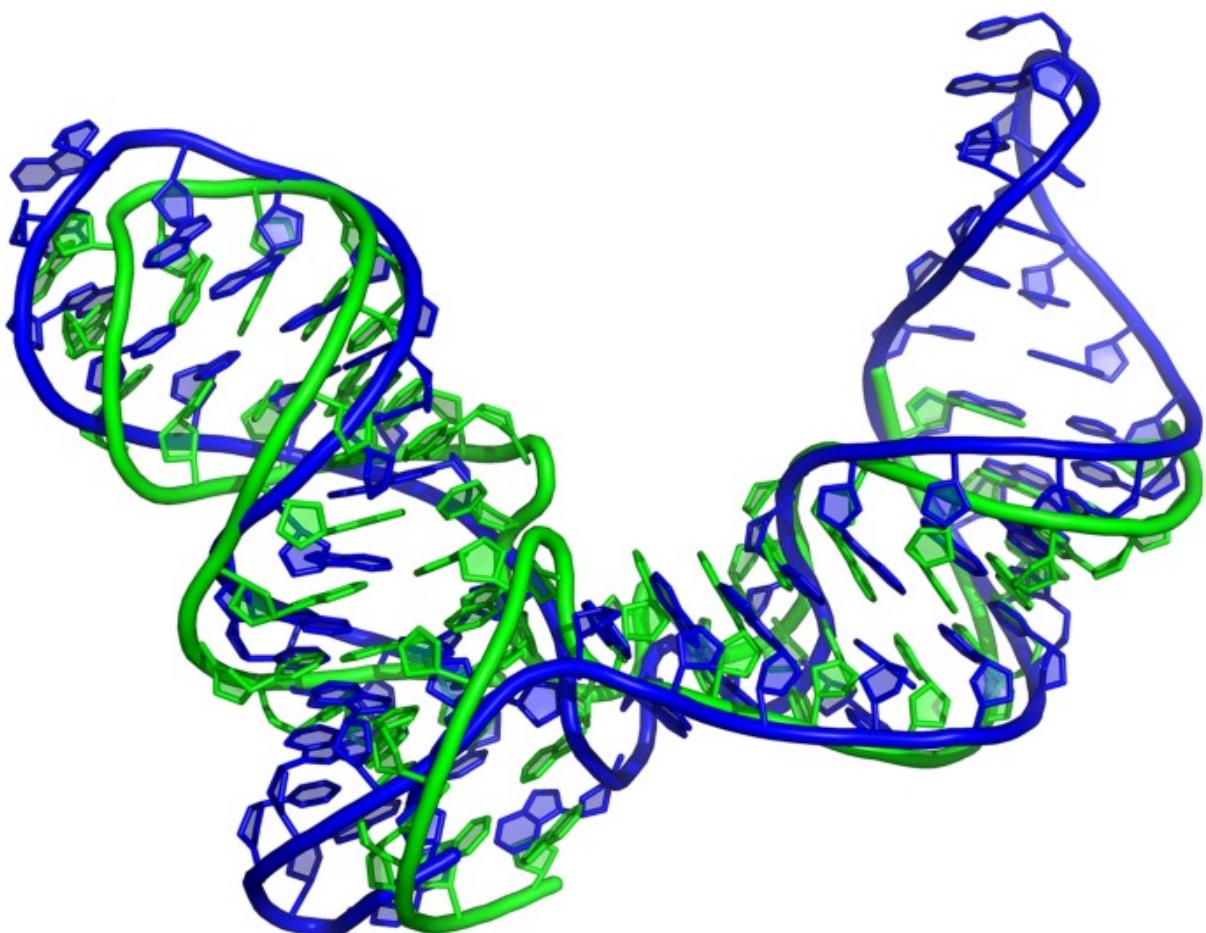
# **Agenda**

- 1. Introduction and Project Overview**
- 2. Problem Setting / Evaluation**
- 3. Theoretical Foundation: DDPM vs. DDIM**
- 4. Approach and Self-Contributition**
- 5. Results**
- 6. Discussion**

# 1

## Introduction and Project Overview

# Introduction



## Background

- Biological function of RNA is determined by their 3D structure – which in turn is determined by the nucleotide interaction [3]
- How a given chain of RNA nucleotides folds and behaves in 3D space remains an unsolved issue at scale [3]
- AlphaFold's success with protein-folding has brought translated to a limited extent to RNA folding [4]

## Recent Advances

- Graph Neural Networks (GNNs) + Denoising Diffusion Probabilistic Models (DDPMs) set new benchmarks [1]
- Focus on predicting local substructures and assembling them into full RNA molecules

## (Some) Open Problems

- Data Scarcity, Imbalances, and Generalizability
- Computational Requirements both during training and at inference

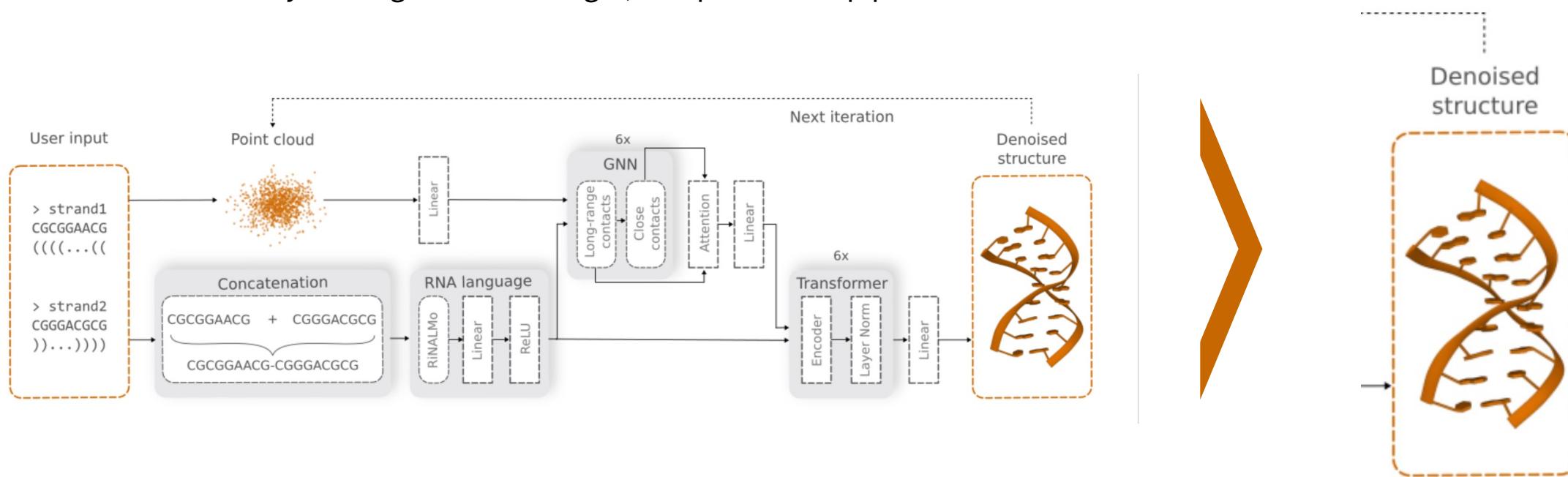
<https://www.chemistryworld.com/news/machine-learning-accurately-predicts-rna-structures-using-tiny-dataset/4014347.article>

# Project Overview: High-Throughput RNA 3D Folding via Fast Diffusion Sampling

**Goal of the Project:** Enhance the RNAGrail framework by *M. Justyna et. al (2024) [1]* for RNA 3D structure prediction by integrating a faster diffusion sampling methods (DDIM) in place of the model's standard implementation using DDPM, aiming to reduce computational cost at inference time while preserving predictive accuracy.

**Project Impact:** Accelerated sampling will enable:

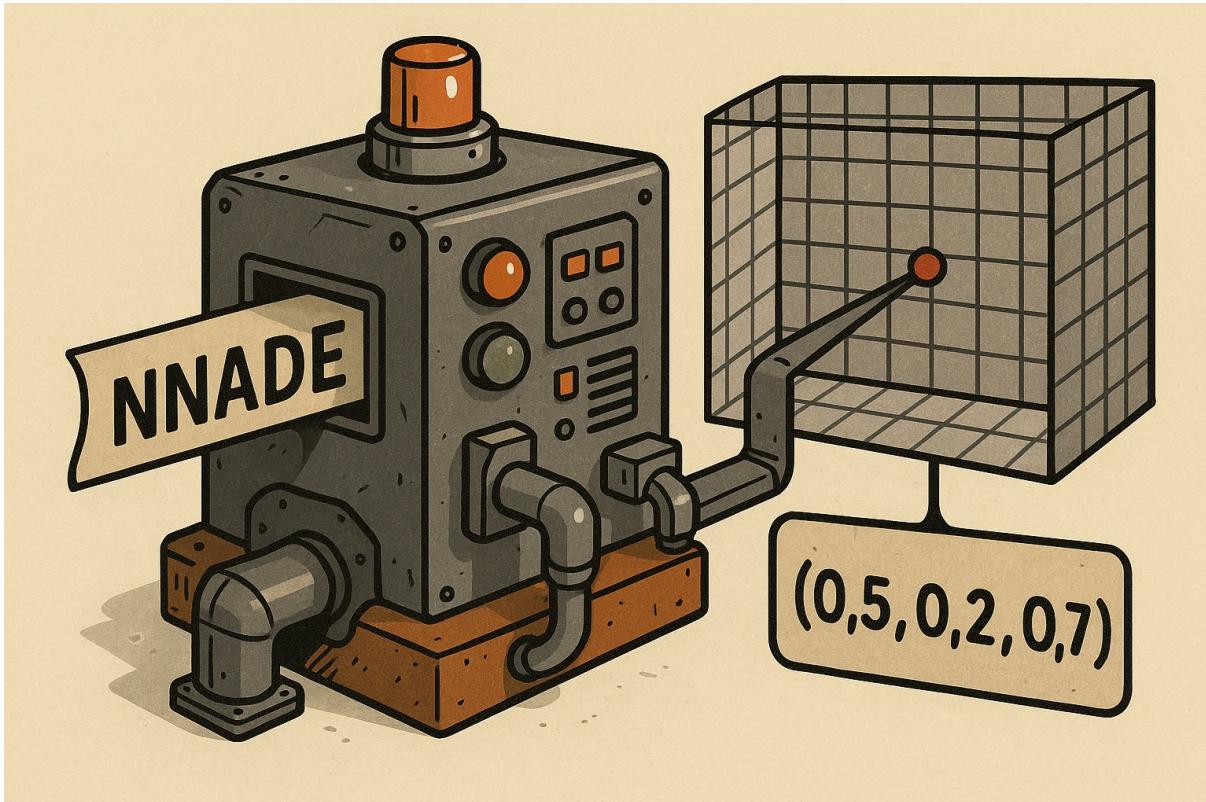
- High-throughput screening of RNA sequences
- Ensemble predictions at scale
- Enhanced feasibility of integration into larger, comprehensive pipelines



# **Problem Setting / Evaluation**



# Problem Setting / Evaluation



## Prediction Task

- *Input:* RNA sequence  $S = \langle n_1, n_2, \dots, n_N \rangle$
- *Output:* 3D coordinates  $P = \langle p_1, p_2, \dots, p_N \rangle$   $p_i$  in  $\mathbb{R}^3$   
→ Each nucleotide's spatial position must be predicted

## Evaluation Criteria

- Accuracy Metric: Template Modeling Score (TM-score) [2]

$$\text{TM-score} = \max \left( \frac{1}{L_{\text{ref}}} \sum_{i=1}^{L_{\text{align}}} \frac{1}{1 + \left( \frac{d_i}{d_0} \right)^2} \right)$$

- Statistical Baseline: Sequential placement of nucleotides with a random angle and a fixed bond length

## Focus:

- Quantify the trade-off between sampling speed and structural accuracy → Focus on *Inference*
- Model training time per epoch (note: computational limitations)

# **Denoising Diffusion Probabilistic Models vs. Denoising Diffusion Implicit Models**

3

# Theoretical Foundation

## Denoising Diffusion Probabilistic Models (DDPM) [5]

- A generative model that learns to reverse a gradual noising process to generate data from pure noise.

$$p_{\theta}(x_{t-1}|x_t) \propto \mathcal{N}\left(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)\right) \text{ s.t. } x_{t-1} = \mu_{\theta}(x_t, t) + \sqrt{\Sigma_{\theta}(x_t, t)}\epsilon_{\theta}$$

$$\mu_{\theta}(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(x_t, t) \right) \quad ; \quad \Sigma_{\theta}(x_t, t) = \tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t)$$

- High-quality and diverse outputs due to the stochastic nature of the process
- **But:** Slow sampling due to the requirement of many timesteps  $T$  and serial nature of the process caused by the Markovian Property of the denoising chain

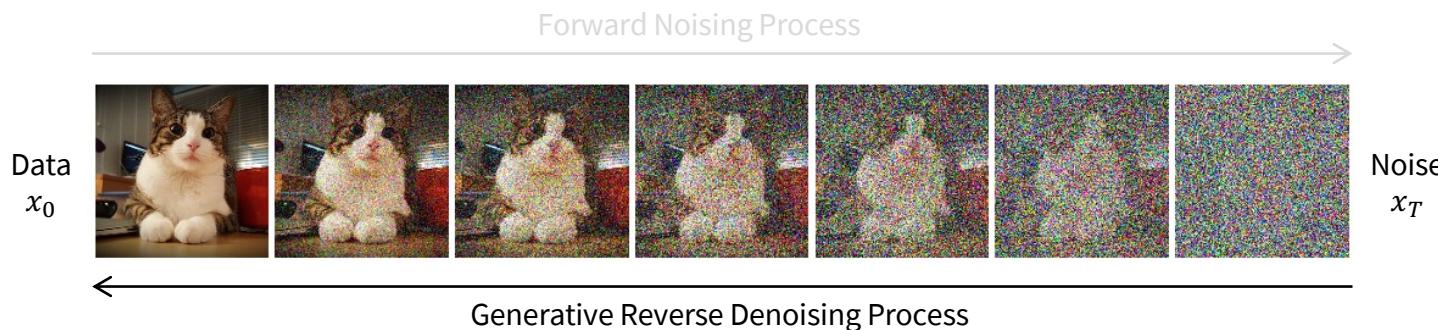


Image: <https://sushant-kumar.com/blog/ddpm-denoising-diffusion-probabilistic-models>

# Theoretical Foundation

## Denoising Diffusion Implicit Models (DDIM) [6]

- A non-Markovian, *deterministic* (for  $\eta = 0$ ) sampler that re-uses the same network  $\epsilon_\theta$  trained under the DDPM objective.
- A variation of the reverse denoising process *at Inference*
- **But:** Trades off some sample diversity (controlled by  $\eta$ ) for speed
- The update rule allows the application of the same DDIM update at each chosen step  $t$ , since there is no (or limited and controllable) randomness to accumulate – dependence of  $x_{t-1}$  on  $x_t$  **and the current estimate of  $x_0$**

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}} x_0^{\text{predicted}} + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \cdot \epsilon_\theta(x_t, t) + \sigma_t z$$

$$x_0^{\text{predicted}} = \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} ; \quad \sigma_t = \eta \sqrt{\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}} \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}} ; \quad z \sim \mathcal{N}(0, \mathbf{I}) ; \quad \eta \in [0, 1]$$

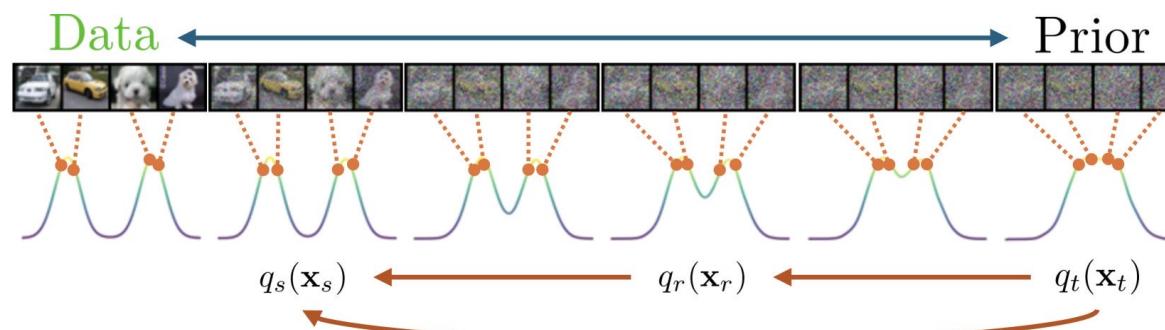
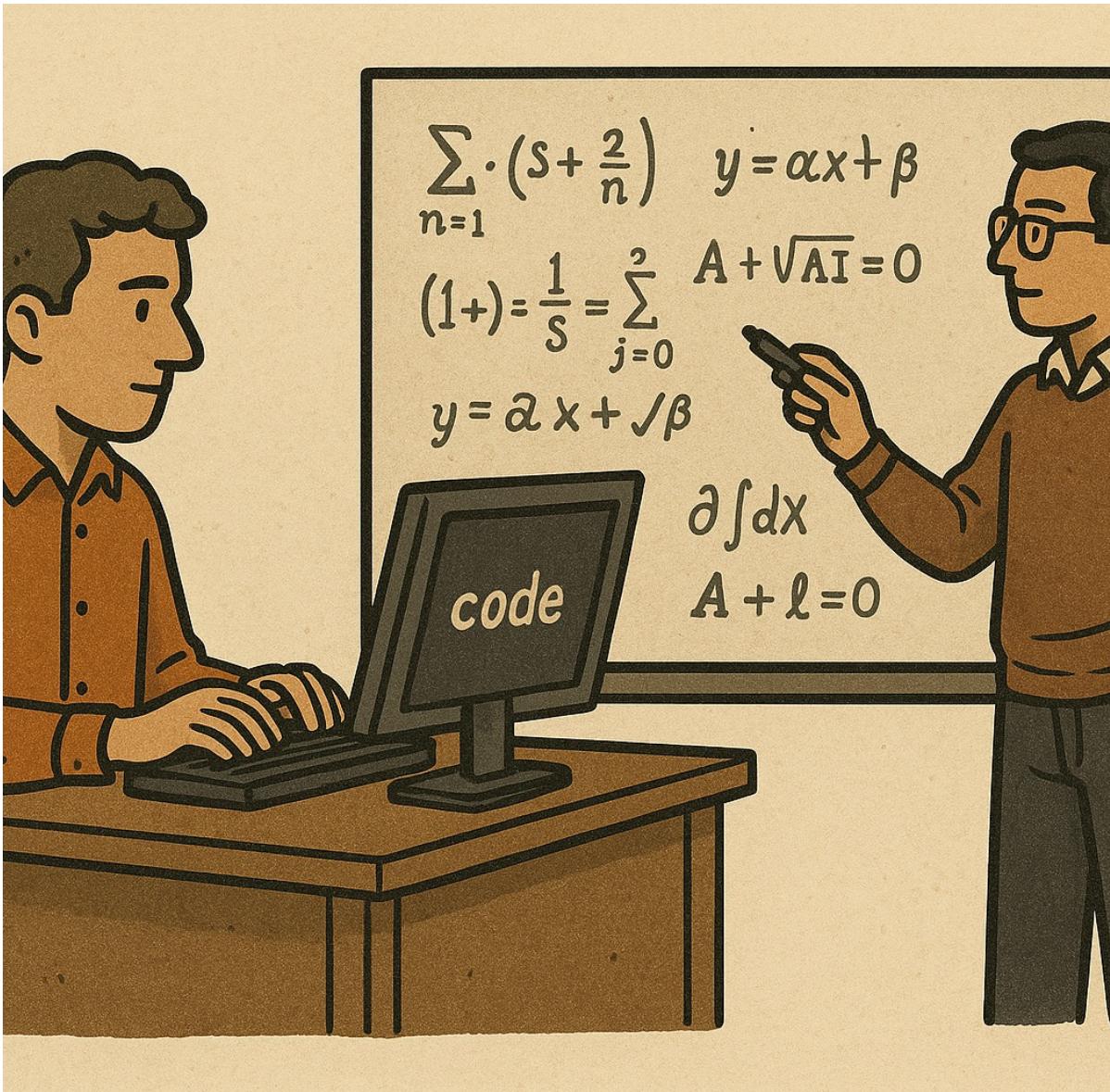


Image: <https://sushant-kumar.com/blog/ddpm-denoising-diffusion-probabilistic-models>

## Approach and Self Contribution



# Approach / Self Contribution

## Kaggle Dataset Adaptation [7]

- Modify RNAGrail's codebase and I/O pipeline to support the Stanford RNA 3D Folding Kaggle Challenge dataset
- Selective sampling of small nucleotide chains to mitigate our computational limitations

## Baseline Setup

- Implement the original RNAGrail architecture locally
- Run without CUDA (MPS-only setup for testing and training)
- Distributed implementation for testing with CUDA systems

## Integration of DDIM Sampler

- Implement DDIM sampler from scratch
- Exchange DDPM sampler with DDIM sampler

## Comparative Evaluation

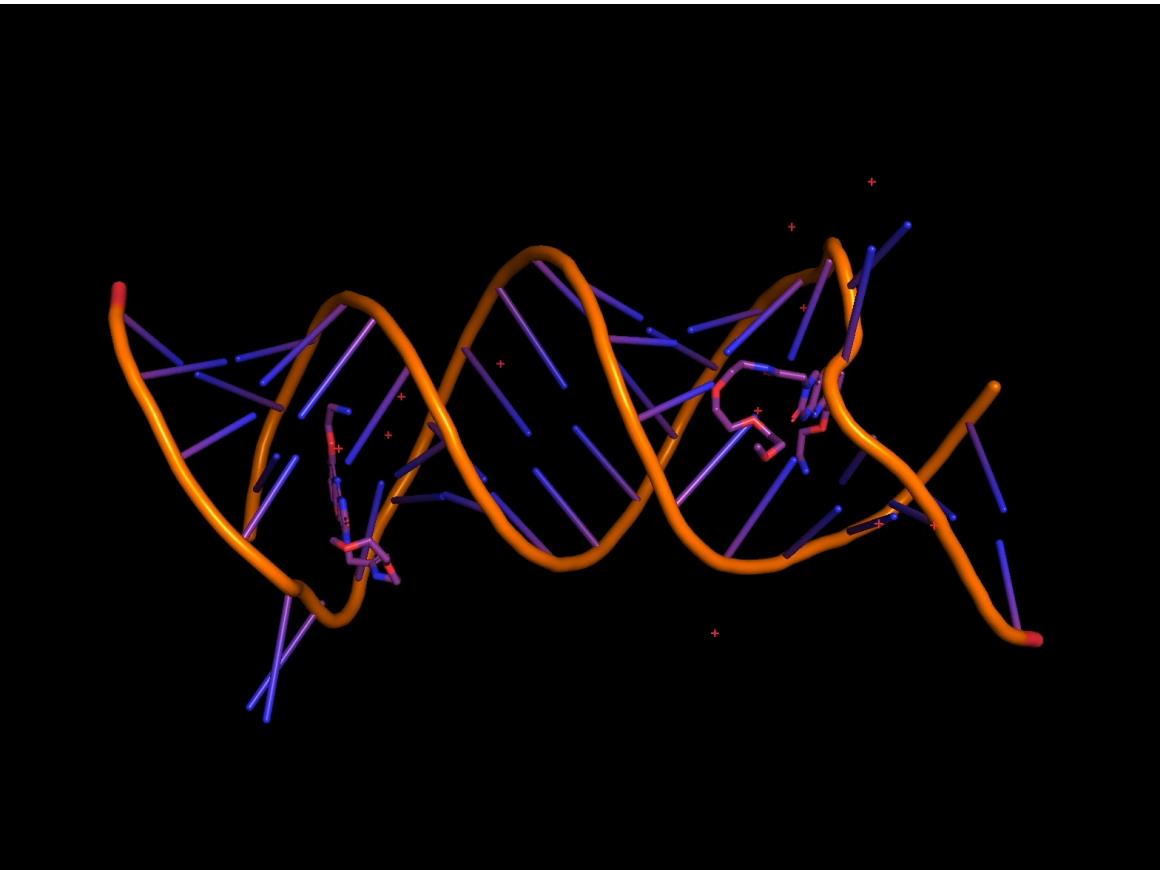
- Accuracy: TM-score of predicted structures
- Efficiency: Sampling speed and training time

# Results



# Results

## Training of a minimal model: (Proof of Concept)



### 1. Constraints

- Trained for 10 epochs on 10% of the data
- Restricted to the smallest 10% of the nucleotides
- Smaller model than the used pre-trained model

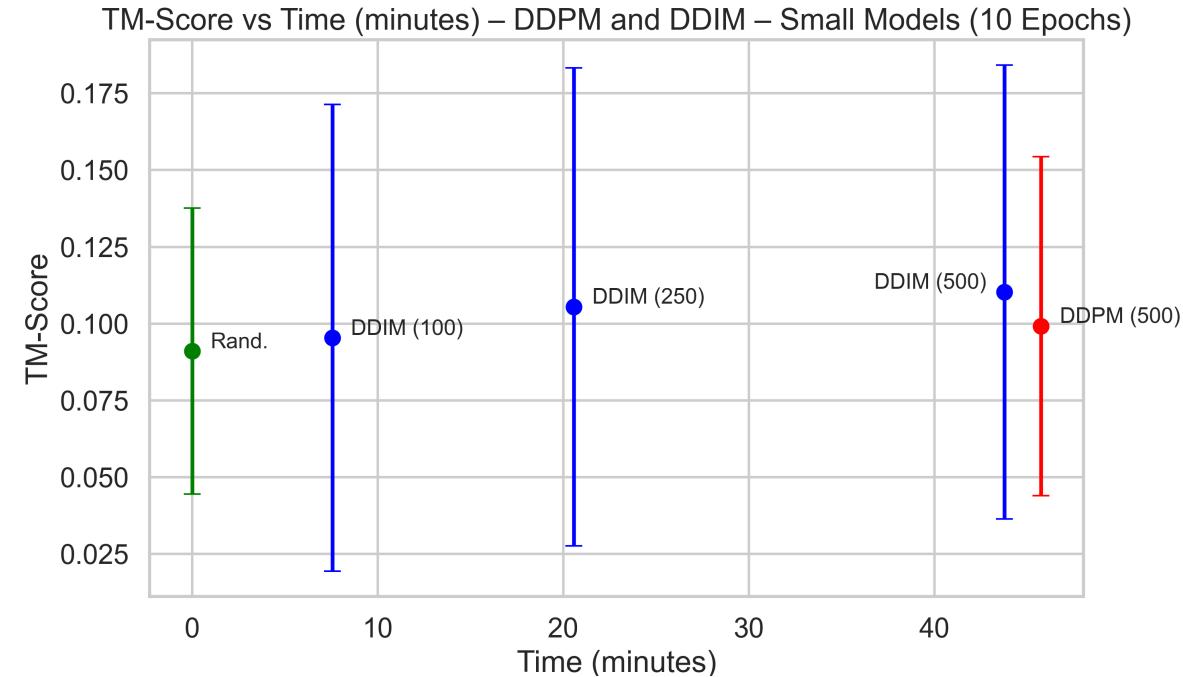
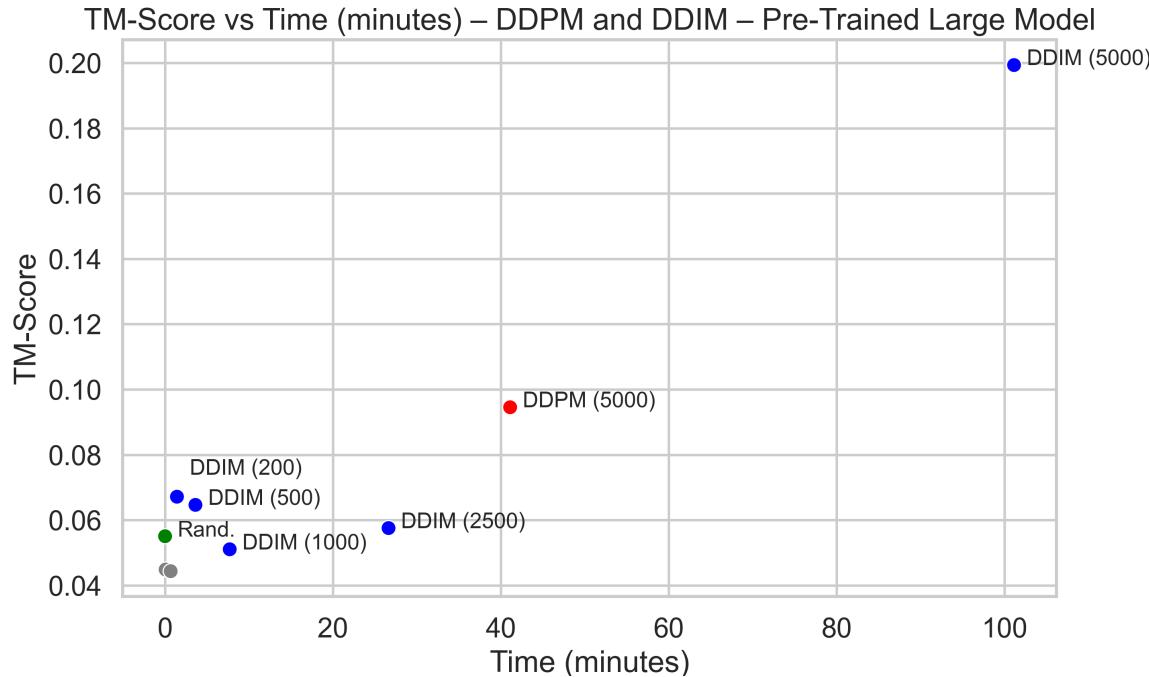
### 2. Outcomes

	Training Loss Avg.	Validation Loss Avg.
DDPM	0.02256	0.02276
DDIM	0.02256	0.02276

→ **Both samplers lead to identical results after training**

Since the sampling is only modified during inference time, the models have the same parameters, regardless of sampler used.

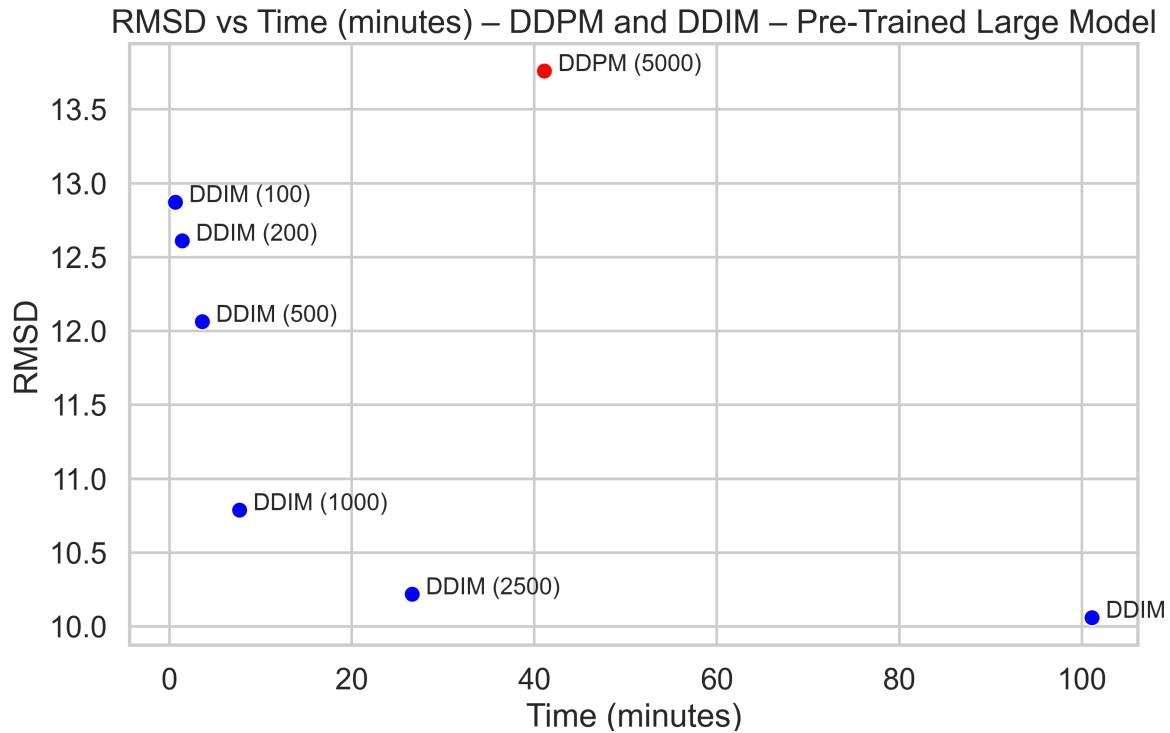
# Results – Template Modeling Scores



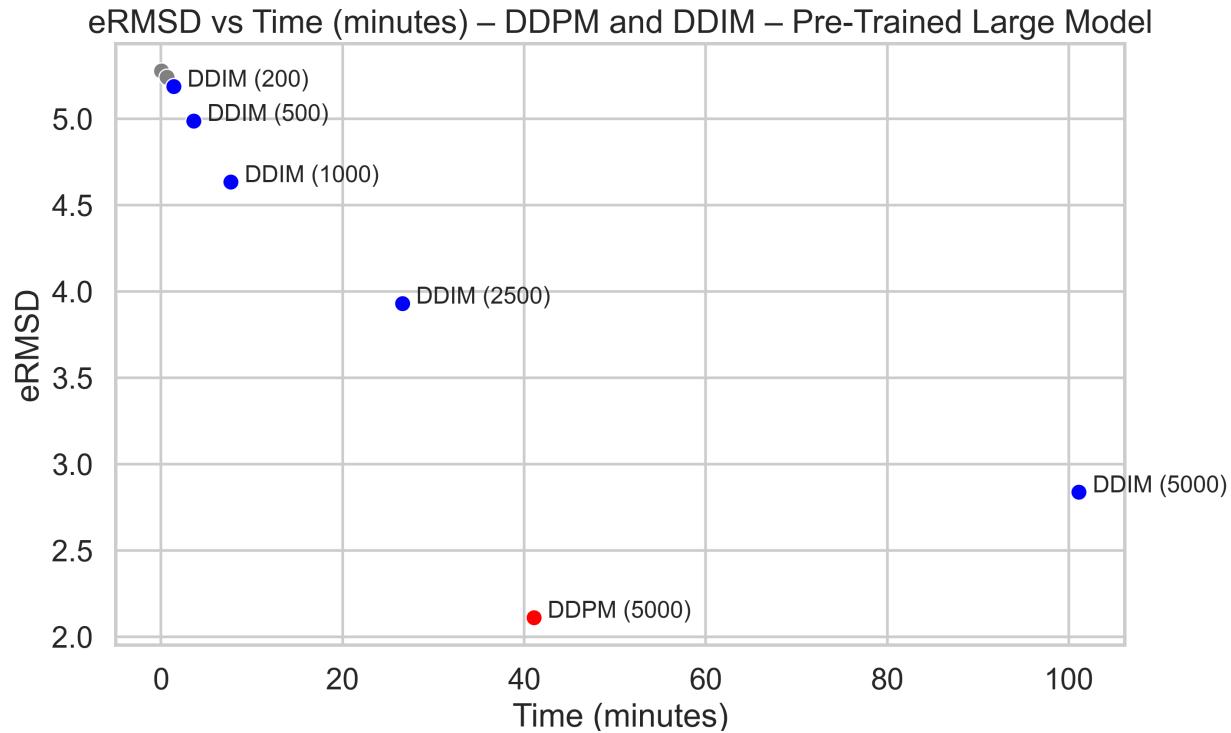
**Template Modeling score:** Topological similarity of the full-length structures of the predicted RNA, indicative of global fold similarity. Higher is better.  
Left: Large, pre-trained model on single molecule (1A1T). Right: Smaller model trained for 10 epochs, applied to validation dataset.

\*Time references are only applicable to the machine used at inference (MacBook Pro Silicon M4 Max)

# Further Evaluation Metrics – Large Model – Inference for Single Molecule



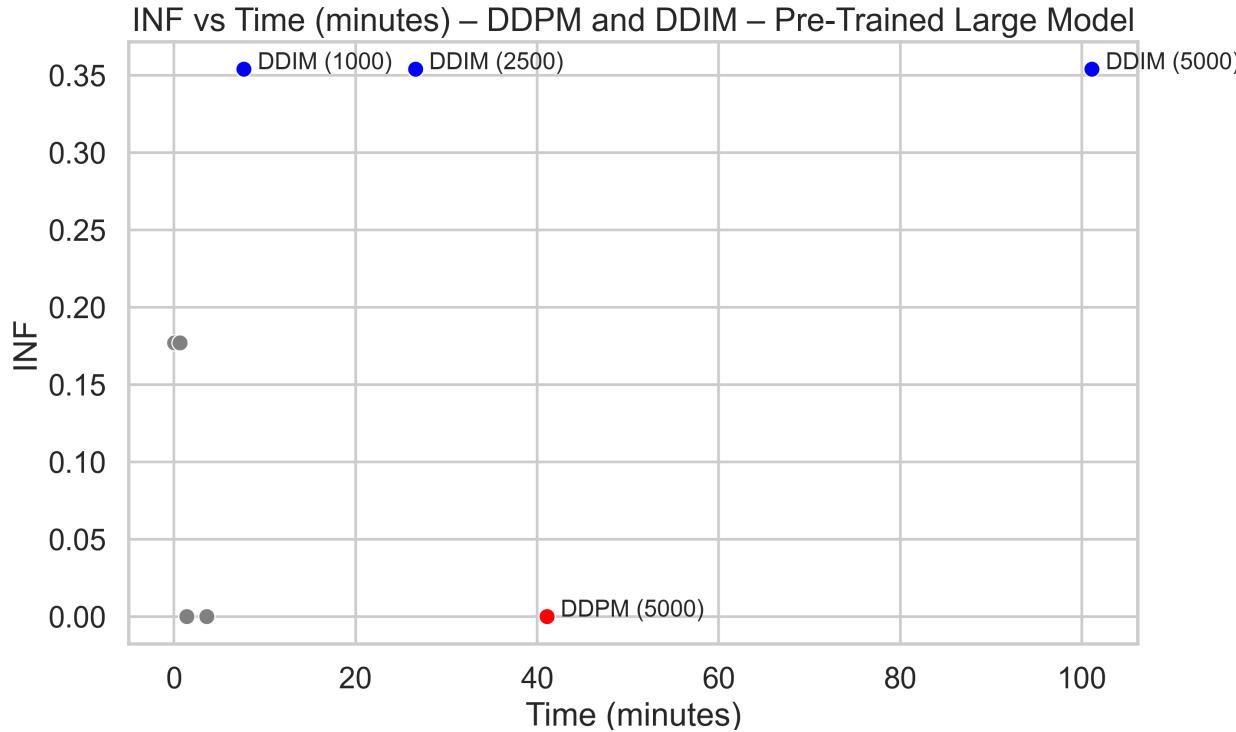
**Root Mean Square Deviation:** Measure of average distance between the backbone atoms of superimposed structures.  
(global structural similarity between two 3D structures).  
Lower is better. Molecule: 1A1T



**Estimated Root Mean Square Deviation:** Considered more sensitive to local structural differences than traditional RMSD, especially for RNA molecules.  
Lower is better. Molecule: 1A1T

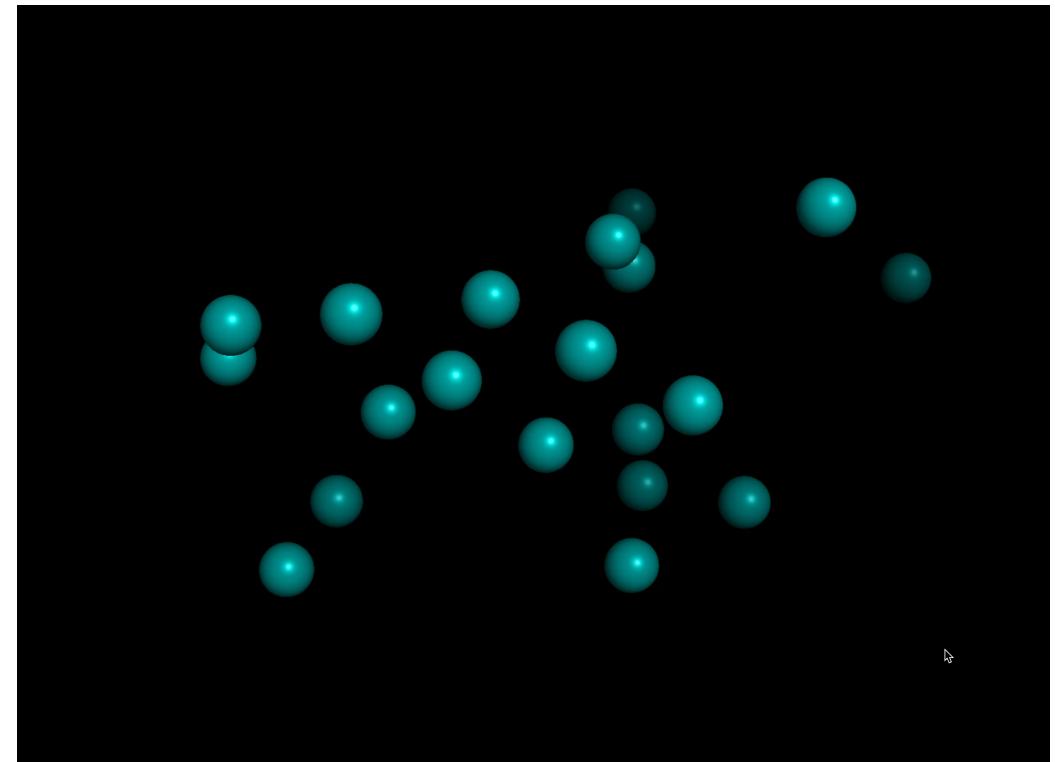
\*Time references are only applicable to the machine used at inference (MacBook Pro Silicon M4 Max)

# Further Evaluation Metrics – Large Model – Inference for Single Molecule



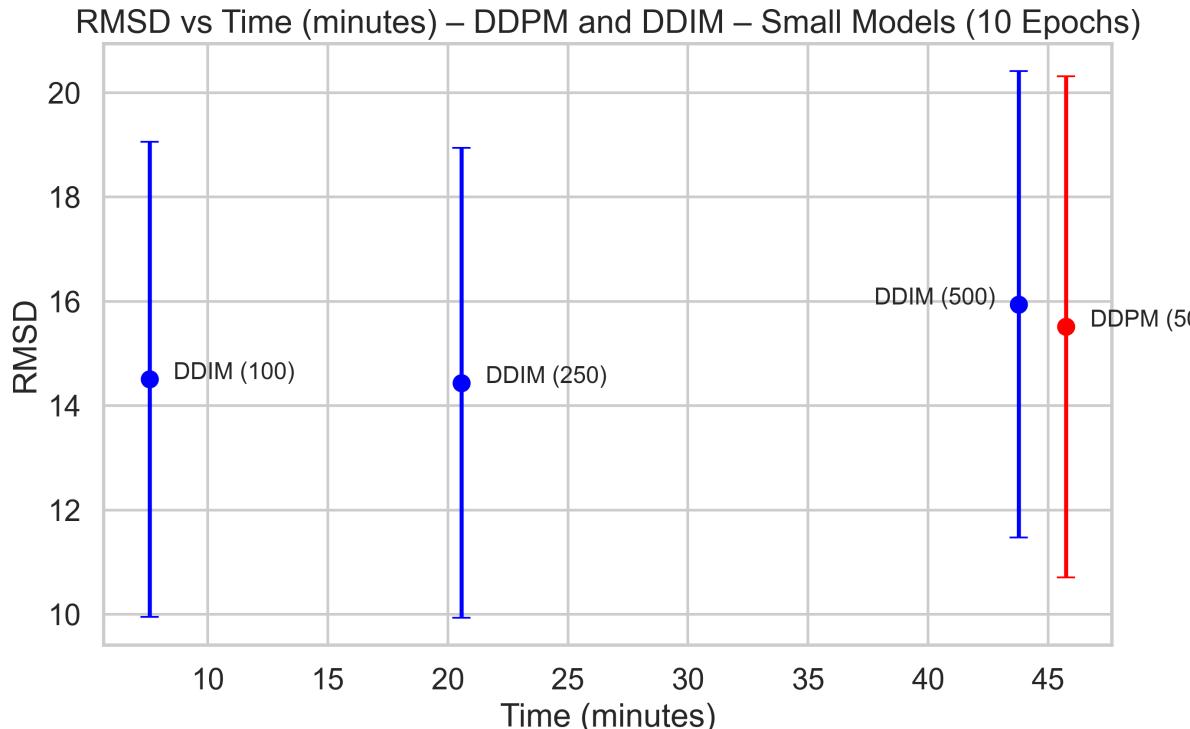
**Interaction Network Fidelity:** Similarity of the secondary structures (2D interactions) between the predicted and target RNA. Calculated based on the confusion matrix.  
Higher is better. Molecule: 1A1T

\*Time references are only applicable to the machine used at inference (MacBook Pro Silicon M4 Max)

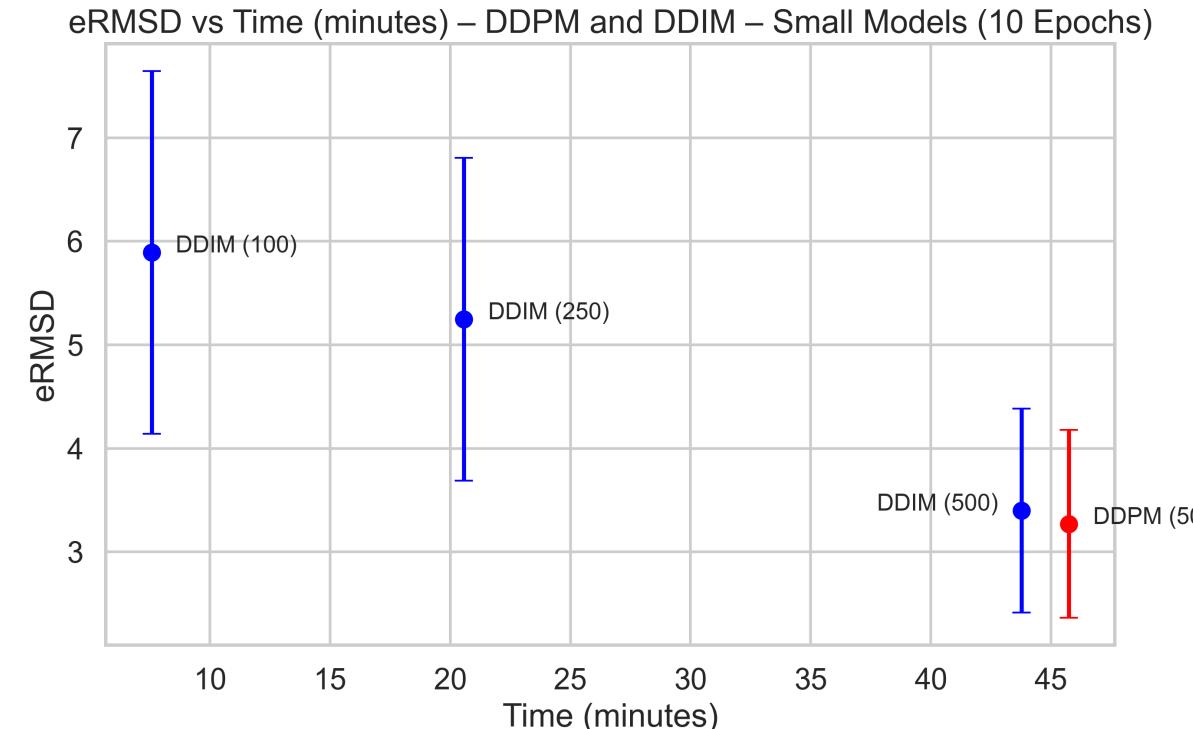


Video Sequence showing various reconstructions, including random baseline, DDIM, DDPM, and ground truth.

# Further Evaluation Metrics – Small Model – Average Scores for Validation Dataset



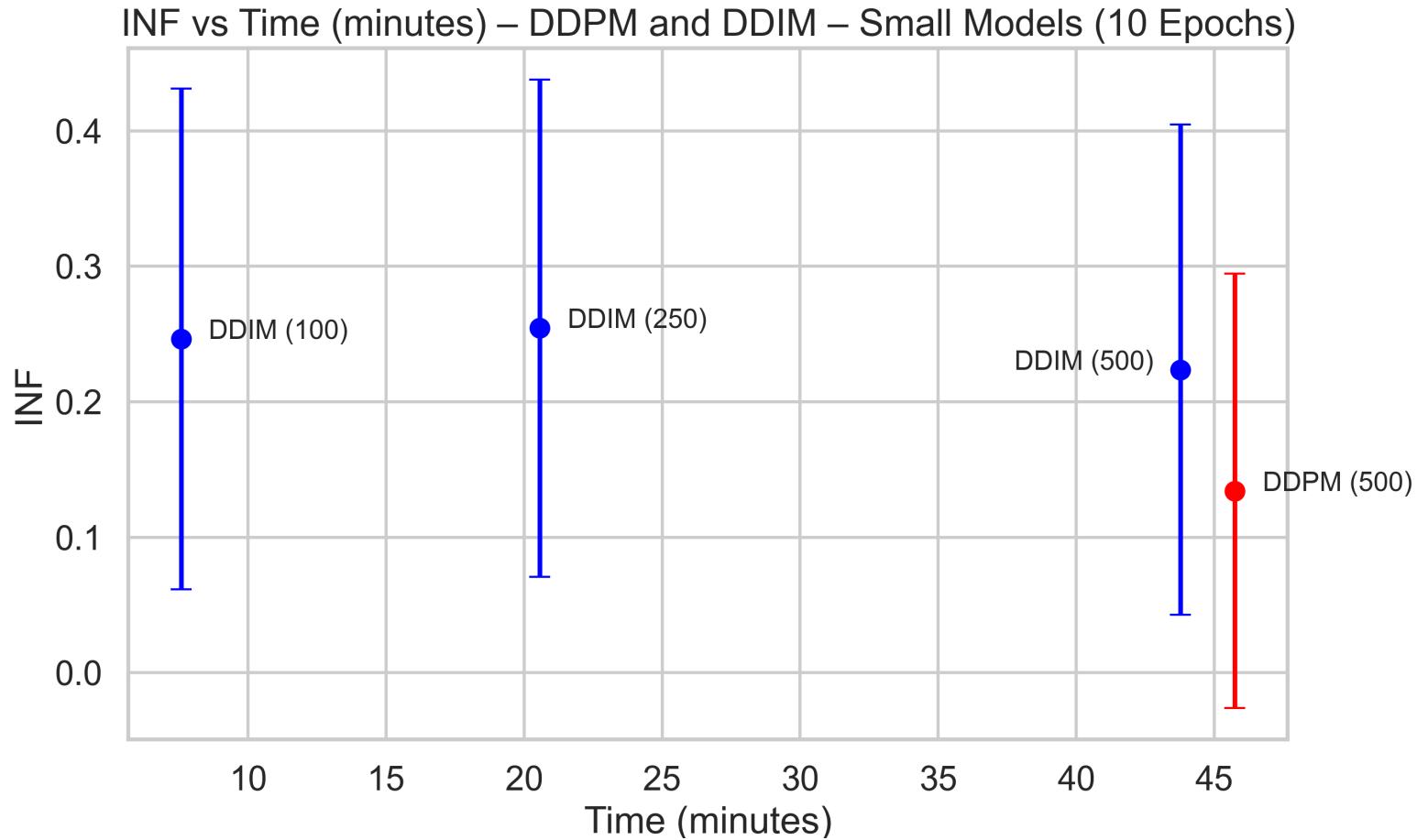
**Root Mean Square Deviation:** Measure of average distance between the backbone atoms of superimposed structures.  
(global structural similarity between two 3D structures).  
Lower is better. Average of validation dataset.



**Estimated Root Mean Square Deviation:** Considered more sensitive to local structural differences than traditional RMSD, especially for RNA molecules.  
Lower is better. Average of validation dataset.

\*Time references are only applicable to the machine used at inference (MacBook Pro Silicon M4 Max)

# Further Evaluation Metrics – Small Model – Average Scores for Validation Dataset



**Interaction Network Fidelity:** Similarity of the secondary structures (2D interactions) between the predicted and target RNA. Calculated based on the confusion matrix. Higher is better. Average of validation dataset.

\*Time references are only applicable to the machine used at inference (MacBook Pro Silicon M4 Max)

# Discussion

6

# Discussion

## Key Outcomes:

- *Creation of a codebase*: extending research regarding DDIM in the context of RNA structure predictions
- *Integrated a modular DDIM sampler*: preliminary experimental trials partially confirm the theoretical assumptions of faster inference at similar accuracy for DDIM vs DDPM
- *Out-of-Distribution limitations*: experimental outcomes hints that OOD sampling has only limited success (based on the limited scope of the data considered)

## Challenges & Limitations

- *Compute constraints*: MPS-only setup prevented full hyperparameter tuning, cross-validation and model convergence
- *Data bias*: Training/testing on only the smallest 10 % of RNA chains limits generalizability to longer, more complex structures
- *Workflow fragility*: Reliance on evolving third-party libraries (rnaPolis, RiNalMo, Arena) required frequent custom patches
- *Statistical Baseline limitations*: RNA structures require a specific nucleotide spatial arrangement to form a physically sound construct; thus, the statistical baseline works on a numerical basis, yet fails to provide a ‘bio-chemical’ reference

## Future Research

- Extend experimental reach by incorporating a more diverse and larger set of nucleotide chains (both at training and inference)
- Implement and benchmark the IMM sampler to assess speed–accuracy trade-offs (might require training of a new model depending on the IMM variant)

## Takeaways

- *Interdisciplinary necessity*: Effective RNA 3D prediction demands close collaboration with domain specialists
- *Theory vs. practice*: Gained solid theoretical insights in diffusion sampling methods, yet practical application demands deeper subject-matter insight

## References

# Reference

- [1] Justyna, Marek, et al. “RNAGrail: graph neural network and diffusion model for RNA 3D structure prediction.” *Machine Learning in Structural Biology*. (2024)
- [2] Xu, Jinrui, and Yang Zhang. “How significant is a protein structure similarity with TM-score = 0.5?” *Bioinformatics (Oxford, England)* vol. 26,7 (2010): 889-95. doi:10.1093/bioinformatics/btq066
- [3] Ignacio Tinoco Jr, and Carlos Bustamante. “How RNA folds”. *Journal of Molecular Biology* vol. 293,7 (1999): 271-81. doi:10.1006/jmbi.1999.3001
- [4] Clément Bernard, Guillaume Postic, Sahar Ghannay and Fariza Tahi. “Has AlphaFold 3 achieved success for RNA?” *Acta Crystallographica* vol. 81, 2 (2025). doi:10.1107/S2059798325000592
- [5] Calvin Luo. “Understanding Diffusion Models: A Unified Perspective.” (2022). <https://arxiv.org/abs/2208.11970>
- [6] Jiaming Song, Chenlin Meng and Stefano Ermon. “Denoising Diffusion Implicit Models.” *CoRR* vol. 2010, 02502 (2020). <https://arxiv.org/abs/2010.02502>
- [7] Kaggle. “Stanford RNA 3D Folding.” (2025). <https://www.kaggle.com/competitions/stanford-rna-3d-folding>