

Nüwa RNA Foundation Model: A Unified Representation of RNA Sequence, Structure, and Function

Nüwa Team¹

¹Shanghai Academy of Artificial Intelligence for Science, Shanghai 200003, China.

Abstract

Ribonucleic acid (RNA) plays a central role in biological regulation and therapeutics, yet deciphering the complex interplay between its primary sequence, secondary structure, and biological function remains a formidable computational challenge. Existing RNA language models often struggle to explicitly integrate structural modalities or lack the capacity to capture microscopic physical properties effectively. Here we introduce the Nüwa RNA Foundation Model, a large-scale generalist model that establishes a unified representation of RNA sequence, structure, and function. Nüwa is trained on a massive corpus of diverse RNA types using a novel multi-level masked self-supervised learning framework. This approach synergizes synchronized single-token masking with span-based masking strategies and explicitly incorporates secondary structure information within an optimized architecture. This multi-modal training regime enables the emergence of advanced cognitive capabilities regarding RNA structural features and chemical modifications. We demonstrate that Nüwa achieves state-of-the-art performance across 43 evaluation metrics spanning sequence generation, structure prediction, and functional inference, ranking first in comprehensive benchmarks. Notably, Nüwa attains optimal results in 42 of these metrics, surpassing leading models such as RNA-FM, RNAGenesis, and AIDO.RNA. Beyond *in silico* benchmarking, we validate Nüwa’s practical utility through a “lab-in-the-loop” system for nucleic acid drug design. In experimental validation targeting five distinct targets across aptamer and siRNA modalities, Nüwa reduced wet-laboratory validation costs by over 90%. As a highly capable and modality-complete RNA foundation model, Nüwa offers a robust, validated framework for accelerating the discovery and optimization of RNA therapeutics.

Keywords: RNA, Foundation Model, Representation Learning, Drug Discovery

1 Introduction

Ribonucleic acid (RNA) has transcended its traditional characterization as a mere intermediary between DNA and proteins to emerge as a central orchestrator of cellular complexity. From non-coding RNAs (ncRNAs) that fine-tune gene expression networks [1] to messenger RNAs (mRNAs) that serve as the blueprint for protein synthesis, RNA molecules exhibit a staggering diversity of functions governed by their ability to fold into intricate secondary and tertiary structures. This structural versatility underpins critical biological processes, including catalysis, ligand sensing, and the regulation of chromatin states, making RNA a prime target for next-generation therapeutics and synthetic biology applications [2]. However, the effective design and analysis of RNA therapeutics—such as aptamers, small interfering RNAs (siRNAs), and mRNA vaccines—require a deep, predictive understanding of the complex mapping from primary sequence to functional structure and biological activity.

Recent advances in deep learning, particularly the advent of large language models (LLMs) adapted for biological sequences, have begun to address these challenges. Models such as RNA-FM [3] and others have demonstrated the utility of self-supervised learning on large-scale sequence databases. Yet, despite these strides, existing RNA foundation models face significant limitations. Many rely predominantly on sequence information alone, failing to explicitly integrate the structural modalities that are deterministic of RNA function. Furthermore, most current models lack the scale and architectural sophistication necessary to capture fine-grained “microscopic” physical properties, such as the thermodynamic stability of secondary structural motifs or the propensity for chemical modifications, which are crucial for the developability of nucleic acid drugs. The absence of a unified framework that can simultaneously model sequence semantics, structural constraints, and functional outcomes at scale remains a critical bottleneck in the field.

To bridge this gap, we introduce the Nüwa RNA Foundation Model, a large-scale generalist model designed to provide a comprehensive, multi-dimensional representation of the RNA universe. Nüwa represents a paradigm shift in RNA modeling, leveraging a massive corpus of diverse RNA types trained on high-performance computing infrastructure. Unlike predecessors that treat structure as a downstream prediction task, Nüwa integrates structural information directly into the pre-training phase. We employ a novel multi-level masked self-supervised learning framework that synergizes synchronized single-token masking with span-based masking strategies, while simultaneously encoding secondary structure information. This multi-modal approach allows the model to learn rich, structure-aware representations that naturally capture the physical and chemical nuances of RNA molecules.

In this work, we demonstrate that Nüwa establishes a new state-of-the-art across a broad spectrum of computational RNA tasks. We evaluate the model on 43 distinct metrics spanning sequence generation, structure prediction, and functional inference, utilizing comprehensive benchmarks such as BEACON [4] and AIDO. Nüwa achieves optimal performance in 42 of these metrics, consistently outperforming leading baselines including RNA-FM, RNAGenesis [5], and AIDO.RNA [6]. Beyond benchmarking, we validate the practical utility of Nüwa in a real-world drug discovery context. By integrating the model into a “lab-in-the-loop” system, we successfully designed and

optimized nucleic acid drugs against five distinct targets. Our experimental results confirm that Nüwa-driven design can reduce wet-laboratory validation costs by over 90%, highlighting its potential to accelerate the development of novel RNA therapeutics.

2 Methods

2.1 Model Architecture

The Nüwa RNA Foundation Model is built upon a modernized encoder-only Transformer architecture, optimized for efficient processing of long biological sequences. Figure 1 illustrates the high-level architecture of the model. We employ a hybrid attention approach that combines sliding window attention with global attention mechanisms. This strategy effectively conserves memory and mitigates the computationally intensive quadratic complexity of standard full attention, which is particularly sensitive to sequence length. While retaining the core capabilities for capturing global sequence dependencies, we incorporate several architectural enhancements to improve training stability and inference efficiency. These include the use of Rotary Positional Embeddings (RoPE) to better capture relative positioning between nucleotides, and the replacement of standard activation functions with GeGLU variants. Furthermore, to accommodate the multi-modal nature of our training objective, the architecture is extended with specialized heads for both masked language modeling and structural constraint prediction. This design allows the model to simultaneously learn from primary sequence data and secondary structure information without increasing inference latency.

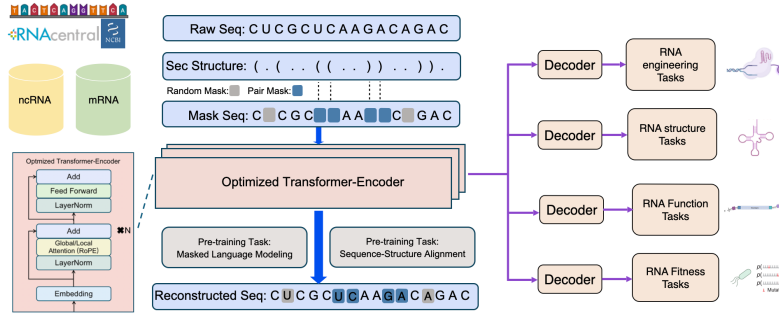


Fig. 1 Nüwa RNA Foundation Model architecture. The model employs a modernized encoder-only Transformer with Rotary Positional Embeddings (RoPE) and GeGLU activation functions. It features specialized heads for masked language modeling and structural constraint prediction to enable multi-modal learning from both sequence and structure data.

2.2 Pre-training Data

To ensure the model captures a comprehensive representation of the RNA universe, we curated a massive-scale pre-training corpus comprising diverse RNA types. The

dataset aggregates sequences from public repositories such as RNACentral, encompassing messenger RNAs (mRNAs), long non-coding RNAs (lncRNAs), ribosomal RNAs (rRNAs), and transfer RNAs (tRNAs). This broad spectrum of data ensures that the model learns generalizable features applicable across different biological contexts. Additionally, secondary structure annotations were generated for a significant subset of these sequences using high-throughput folding algorithms, providing the necessary ground truth for our structure-aware training objectives.

2.3 Multi-level Masked Self-Supervised Learning

We introduce a novel multi-level masked self-supervised learning framework to train Nüwa. Unlike traditional approaches that rely solely on random token masking, our strategy synergizes distinct masking paradigms to capture both local and global dependencies:

- **Synchronized Single-Token Masking:** Our framework simultaneously masks single tokens in both sequences and structures. This dual-masking approach incorporates structural information directly into the learning process. Specifically, when a token is randomly selected for masking, its paired tokens in the secondary structure are also masked, encouraging the model to learn local nucleotide dependencies and sequence motifs in the context of their structural environment.
- **Span-Based Masking:** To capture longer-range dependencies and functional motifs, we implement a span masking strategy applied to sequences. By masking contiguous segments of nucleotides rather than individual tokens, we introduce motif-level information, forcing the model to predict entire functional units based on the surrounding context.

Furthermore, we explicitly incorporate secondary structure information into the pre-training process. This is achieved by introducing a structural consistency objective, where the model must predict the pairing status of nucleotides (paired vs. unpaired) and recover masked structural motifs. This multi-modal training regime enables the model to internalize the physical constraints governing RNA folding, leading to the emergence of advanced cognitive capabilities regarding RNA structural features.

2.4 Lab-in-the-Loop Validation System

To validate the practical utility of Nüwa in therapeutic discovery, we developed a “Lab-in-the-loop” system that tightly integrates computational predictions with experimental verification. The workflow consists of the following steps:

1. **Generative Design:** The Nüwa model is used to generate candidate RNA sequences (e.g., aptamers or siRNAs) optimized for specific binding affinities or silencing efficiencies against a target protein.
2. **In Silico Screening:** Generated candidates are filtered and ranked based on the model’s predicted structural stability and functional scores.
3. **Experimental Verification:** Top-ranked candidates are synthesized and tested in wet-lab assays to measure their actual biological activity (e.g., K_D values for aptamers).

4. **Feedback Loop:** Experimental results are fed back into the system to fine-tune the model and refine subsequent generations of candidates.

This closed-loop system allows for the rapid iteration and optimization of nucleic acid drugs, significantly reducing the time and cost associated with traditional screening methods.

3 Results

3.1 Comprehensive Evaluation on RNA Understanding and Generation Benchmarks

To systematically assess the capabilities of the Nüwa RNA Foundation Model, we conducted extensive evaluations using the BEACON and AIDO benchmarks, which collectively encompass 43 metrics across core RNA tasks including sequence generation, structure prediction, expression abundance prediction, and splice site identification. Nüwa demonstrated superior performance, ranking first in comprehensive scores across these benchmarks and achieving state-of-the-art results in 42 out of the 43 evaluated metrics. Figure 2 illustrates the comparative performance of Nüwa against other leading models across various metrics in the benchmark.

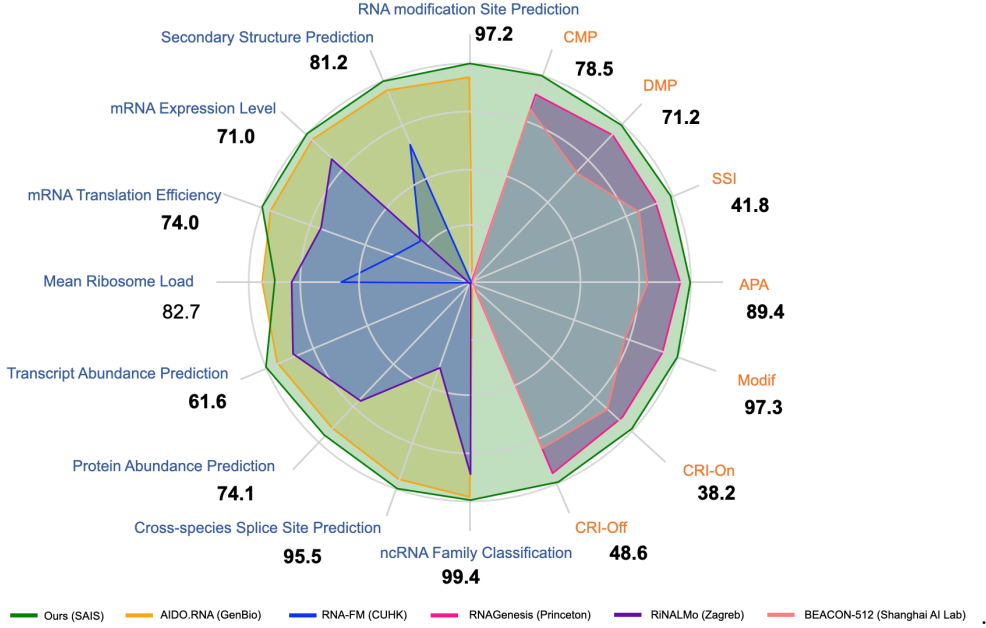


Fig. 2 Performance comparison on the benchmark. The radar chart displays the scores of Nüwa (green line) against other models such as AIDO.RNA, RNA-FM, RNAGenesis, RiNALMo, and BEACON-512 across representative tasks including Secondary Structure Prediction, RNA modification Site Prediction, and others. Nüwa consistently achieves the highest scores across most categories.

In structural modeling tasks, Nüwa exhibited precise cognitive capabilities regarding microscopic physical properties. For secondary structure prediction, the model achieved a score of 81.2, significantly outperforming generalist baselines such as RNA-FM and AIDO.RNA. Similarly, in functional classification and site prediction, Nüwa demonstrated exceptional accuracy, attaining 99.4 in non-coding RNA (ncRNA) family classification and 95.5 in cross-species splice site prediction. High-precision performance was also observed in epitranscriptomic tasks, with an RNA modification site prediction score of 97.2 and a modification (Modif) score of 97.3.

Furthermore, Nüwa showed robust predictive power in quantitative gene regulation tasks. It achieved an mRNA translation efficiency score of 74.0 and a mean ribosome load score of 82.7, indicating a deep understanding of the regulatory syntax governing protein synthesis. In expression profiling, the model attained scores of 71.0 for mRNA expression level and 74.1 for protein abundance prediction. These results highlight Nüwa’s capacity to unify sequence semantics with functional outputs across diverse biological contexts, consistently surpassing leading domain-specific and generalist models including RNAGenesis and RiNALMo.

3.2 Experimental Validation in Nucleic Acid Drug Design

Beyond *in silico* benchmarking, we validated the practical utility of Nüwa in therapeutic development through a self-developed “Lab-in-the-loop” (dry-wet closed-loop) system. This system integrates the model’s generative and predictive capabilities directly with experimental verification to accelerate nucleic acid drug design.

We applied Nüwa to the design and optimization of two distinct drug modalities: aptamers and small interfering RNAs (siRNAs). In collaboration with biopharmaceutical partners, we conducted in-depth evaluations on over five specific therapeutic targets. The model successfully generated high-affinity and high-specificity candidates that were subsequently validated *in vitro*. Analysis of the experimental data indicates that the Nüwa-driven closed-loop approach reduced the average costs associated with *in vitro* wet laboratory validation by over 90%. To our knowledge, this represents the first instance of an RNA foundation model enabling a complete, validated solution for nucleic acid drug design, bridging the gap between computational representation learning and clinical translation.

References

- [1] Mercer, T.R., Dinger, M.E., Mattick, J.S.: Long non-coding rnas: insights into functions. *Nature Reviews Genetics* (2009)
- [2] Zhu, Y., Zhu, L., Wang, X., Jin, H.: Rna-based therapeutics: an overview and prospectus. *Cell Death Dis.* (2022)
- [3] Shen, T., et al.: Accurate rna 3d structure prediction using a language model-based deep learning approach. *Nature Methods* (2024)

- [4] Ren, Y., et al.: Beacon: Benchmark for comprehensive rna tasks and language models. NeurIPS (2024)
- [5] Zhang, Z., et al.: Rnagenesis: A generalist foundation model for functional rna therapeutics. bioRxiv (2025)
- [6] Zou, S., et al.: A large-scale foundation model for rna function and structure prediction. bioRxiv (2024)