BIOLOGY-GUIDED SYNERGISTIC LEARNING OF RNA FOUNDATION MODELS FOR SECONDARY STRUCTURE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

RNA secondary structure prediction aims to identify base-pairing patterns from primary sequences and requires rich biological priors for accurate modeling. Recent foundation models trained with large-scale pretraining encode rich biological priors into high-dimensional representations, yet their effective exploitation for downstream folding remains challenging. To address this issue, we revisit RNA secondary structure prediction from a representation learning perspective, focusing on a more effective exploitation of existing prior-aware representations rather than developing new deep learning modeling paradigms. We propose **BioSynFold**, a **bio**logy-guided **syn**ergistic learning framework for RNA secondary structure prediction that leverages foundation models from both model and data perspectives. From the model perspective, we propose a unified framework that integrates multiple RNA foundation models and introduces a progressive integration strategy to better leverage heterogeneous representations. From the data perspective, we exploit weak signals derived from sequence-level statistical and compositional properties to guide a mixture-of-experts routing mechanism, enabling condition-aware specialization and improving representation quality. Extensive experiments on multiple datasets indicate that BioSynFold significantly outperforms the state-of-the-art methods. Our code is available at https://anonymous.4open.science/r/BioSynFold.

1 INTRODUCTION

RNA is a single-stranded molecule whose biological functions are determined by its secondary structure (Wan et al., 2014; Wang et al., 2021), which is defined by base-pairing interactions along the sequence (Mortimer et al., 2014). Accurate prediction of these patterns is a fundamental task for understanding their functions (Disney, 2019), with implications for elucidating RNA mechanisms and accelerating the development of RNA-targeted therapeutics and nucleic acid drugs (Roberts et al., 2020; Wang et al., 2020; Mortimer et al., 2014). However, RNA secondary structure prediction requires global inference within a highly constrained discrete space (Rivas et al., 2017; Zhang et al., 2023). This setting entails modeling complex long-range interactions under strict structural constraints, making it difficult to achieve both prediction accuracy and generalization performance.

RNA secondary structure has traditionally been studied using experimental techniques like NMR (Barnwal et al., 2017), X-ray crystallography (Jackson et al., 2023), and cryo-EM (Ma et al., 2022). These methods primarily determine RNA tertiary structures, from which secondary structures can be inferred (Merino et al., 2005; Zhang et al., 2022). However, such approaches are costly, time-consuming, and not scalable for large genomic datasets. To overcome these limitations, computational approaches have been developed. Early methods, including thermodynamic models, dynamic programming, and partition functions, aim to predict base-pairing probabilities (Zuker, 2003; Lorenz et al., 2011; Reuter & Mathews, 2010). Other approaches, such as stochastic context-free grammars and shallow discriminative models, use handcrafted heuristics to score candidate structures (Knudsen & Hein, 2003; Do et al., 2006). These methods are limited by fixed assumptions and handcrafted features, which fail to model long-range dependencies and may overlook biologically significant structures, such as those where base pairs form non-crossing stacks.

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.

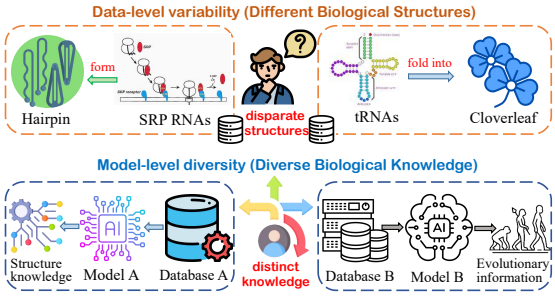


Figure 1: **Motivation**. The upper half illustrates ***data-level variability***: RNA families exhibit structures (e.g., SRP RNAs form hairpin structures, whereas tRNAs fold into cloverleaf structures). The lower half shows ***model-level diversity***: RNA foundation models trained with different paradigms and biological priors specialize in capturing knowledge.

Driven by the availability of high-quality structural annotations, deep learning has emerged as the dominant paradigm for RNA secondary structure prediction, enabling task-specific representation learning directly from sequences. Early efforts such as SPOT-RNA (Singh et al., 2019) and E2Efold (Chen et al., 2020) focused on integrating evolutionary profiles and unrolled constraints. Subsequent studies have explored modeling paradigms, including the thermodynamic-integrated learning of MXfold2 (Sato et al., 2021), the computer-vision-inspired image-to-image translation of UFold (Fu et al., 2022), and the NLP-driven dependency parsing of DEPFold (Wang & Cohen, 2025). Despite their architectural diversity, these methods predominantly rely on simple, context-independent encodings such as one-hot representations and are trained under a fully supervised learning paradigm. As a result, the learned representations are constrained by the scale and distribution of available labeled data, limiting their ability to capture rich biological priors and long-range contextual dependencies. This limitation is particularly critical for RNA secondary structure prediction, which requires modeling both local structural motifs and long-range base-pairing interactions.

Recent RNA pre-trained models represent a paradigm shift from static encodings such as one-hot representations. Instead of treating nucleotides as isolated symbols with fixed representations, these models learn context-aware embeddings over the RNA alphabet (A, U, C, G), where each representation is dynamically modulated by its sequence context and implicitly encodes rich biological priors. For instance, RNA-MSM (Zhang et al., 2024) leverages multi-sequence alignments to capture evolutionary conservation, whereas ERNIE-RNA (Yin et al., 2025) uses base-pair-aware attention to enforce structural constraints, and RiNALMo (Penić et al., 2025) relies on massive sequence corpora for high-level representations. Collectively, these models demonstrate strong potential in enriching RNA representations beyond conventional discrete encodings. Nevertheless, the performance gains remain limited (Wang & Cohen, 2025; Zablocki et al., 2025). This suggests that, although these models encode rich biological priors, their representations are not yet fully aligned with the requirements of downstream structure prediction, leaving substantial room for improvement.

In this work, rather than focusing on designing new pre-training paradigms or task-specific prediction architectures, we instead ask a more fundamental question: ***how can we better utilize and integrate these rich yet under-exploited representations for RNA secondary structure prediction?*** To answer this question, as illustrated in Figure 1, we make two key observations. **(1) Model-level diversity.** Different pre-trained RNA models encode distinct biological priors due to differences in training objectives, architectures, and data sources. This diversity results in heterogeneous representations, suggesting the benefit of considering multiple sources of knowledge rather than relying on a single model. **(2) Data-level variability.** RNA secondary structures vary substantially across different families. This variability is rooted in differences in sequence-level statistical and compositional properties. Such heterogeneity suggests that uniform representations may be suboptimal, motivating the need for adaptive representation learning. To this end, we introduce **BioSynFold**, a **bio**logy-guided **syn**ergistic learning of RNA foundation models for secondary structure prediction:

- We revisit RNA secondary structure prediction by identifying representation quality as a critical bottleneck, and investigate how to more effectively encode discrete nucleotide sequences into structure-aware representations.
- From the model perspective, we make the first attempt to integrate multiple RNA foundation models within a unified synergistic learning framework, and propose a progressive integration strategy to effectively exploit their heterogeneous representations.
- From the data perspective, we leverage weak signals derived from sequence-level statistical and compositional properties to guide a mixture-of-experts routing mechanism, enabling condition-aware specialization and improving representation quality.
- Extensive experimental results on multiple datasets demonstrate that the proposed BioSynFold outperforms the state-of-the-art methods by a large margin.

2 Related Work

Traditional Methods. Energy-based dynamic programming is a foundational paradigm for RNA secondary structure prediction, optimizing thermodynamic or probabilistic objectives over canonical nested structures. Methods such as Mfold (Zuker, 2003) estimate minimum free energy structures using empirical parameters, RNAfold (Lorenz et al., 2011) extends this by computing base-pairing probabilities, and RNAstructure (Reuter & Mathews, 2010) refines energy models to better capture motifs like hairpins and internal loops. To alleviate the scalability bottleneck of dynamic programming, LinearFold (Huang et al., 2019) reformulates RNA folding as an incremental parsing process, enabling linear-time approximate inference while preserving thermodynamic scoring. Beyond physical modeling, Pfold (Knudsen & Hein, 2003) employs stochastic context-free grammars to incorporate evolutionary information, CONTRAfold (Do et al., 2006) adopts discriminative learning via conditional random fields, and MXfold2 (Sato et al., 2021) integrates neural networks to learn data-driven energy parameters within dynamic programming. Despite these advances, these methods are limited by nested structural assumptions and escalating computational costs for long sequences.

Supervised Deep Learning–based Methods. Driven by the expansion of structural databases, deep learning has shifted RNA secondary structure prediction from energy-based modeling to end-to-end representation learning. Early frameworks such as SPOT-RNA (Singh et al., 2019) and E2Efold (Chen et al., 2020) use convolutional and recurrent architectures for sequence modeling, while later methods shift to 2D representations, with UFold (Fu et al., 2022) adopting U-Net-based image modeling and sincFold (Bugnon et al., 2024) using end-to-end residual CNNs to capture long-range interactions. Recent advances explore more expressive modeling strategies. DSRNAFold (Wang et al., 2025a) integrates structural context. Generative methods such as RNADiffFold (Wang et al., 2025b) refine RNA contact maps via diffusion models. Other approaches include PriFold (Yang et al., 2025), DEPfold (Wang & Cohen, 2025), and RFold (Tan et al., 2024), which incorporate biological priors, dependency parsing, and combinatorial optimization for decoding, respectively. However, these methods are limited by scarce supervised labels and rely on simple encodings such as one-hot representations, limiting representational expressiveness and biological priors.

Pretrained Large Language Model–based Methods. Inspired by protein language models (Rives et al., 2021; Lin et al., 2023), large-scale self-supervised learning has become a key strategy for learning RNA structural representations (Zablocki et al., 2025). Pretrained RNA language models fall into three types. Sequence-only models learn global context and base-pairing patterns from single RNA sequences. RNA-FM (Chen et al., 2022) uses masked language modeling. RiNALMo (Penić et al., 2025) is the largest RNA language model, employing efficient attention and clustered sampling to ensure training data diversity. Structure-augmented models incorporate explicit structural information. RNABERT (Akiyama & Sakakibara, 2022) combines masked language modeling with structure alignment learning to embed base-pairing patterns. ERNIE-RNA (Yin et al., 2025) uses base-pair-aware attention to capture structural dependencies. RNAErnie (Wang et al., 2024) employs motif-aware pretraining to integrate sequence motifs and sub-sequence relationships into the representation. Evolution-aware models, including RNA-MSM (Zhang et al., 2024), leverage multiple sequence alignments to extract evolutionary covariation signals correlated with conserved base pairs. Even so, differences in training data and modeling strategies lead to heterogeneous representations, and how to effectively utilize them for RNA structure prediction remains underexplored.

3 Proposed Method

3.1 Preliminary

Problem Definition. An RNA primary structure is an ordered sequence of nucleotides $\mathbf{x} = (x_1, x_2, \ldots, x_L)$ over the alphabet $\mathcal{V} = \{\text{A}, \text{U}, \text{C}, \text{G}\}$, where L is the sequence length. The secondary structure is represented as a symmetric contact matrix $\mathbf{Y} \in \{0,1\}^{L\times L}$, where $Y_{ij} = 1$ if nucleotides x_i and x_j form a base pair, and $Y_{ij} = 0$ otherwise. The objective of RNA secondary structure prediction is to learn a mapping function $f : \mathcal{V}^L \to [0,1]^{L\times L}$ that yields a contact map.

Representation Perspective. RNA secondary structure formation is governed by biological priors beyond local sequence patterns. Therefore, we revisit RNA secondary structure prediction from a representation learning perspective, asking how to exploit pretrained RNA representations encoding structural and biological knowledge. Specifically, we investigate how to derive informative high-dimensional embeddings $\mathbf{Z} \in \mathbb{R}^{L\times d}$ from RNA foundation models for contact matrix prediction.

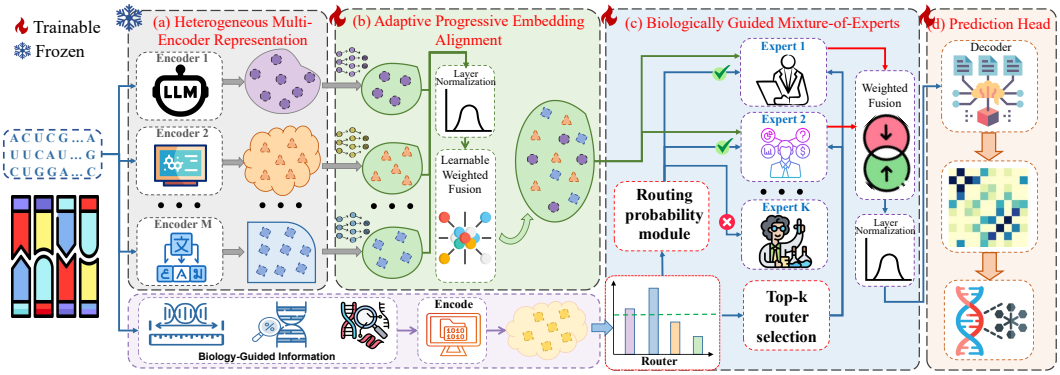


Figure 2: **Schematization of BioSynFold.** The overall architecture consists of four components. *(a)* Multiple RNA foundation models are frozen to extract heterogeneous embeddings that capture diverse structural characteristics. *(b)* The extracted embeddings are projected into a unified latent space and adaptively fused through a learnable weighting mechanism to resolve dimension and distribution mismatches. *(c)* The aligned representations are routed into a mixture-of-experts module guided by biological priors to enable structure-adaptive expert routing. *(d)* The fused expert outputs are finally fed into a prediction head to produce RNA secondary structure predictions.

3.2 Heterogeneous Multi-Encoder Representation

Despite the remarkable progress of recent RNA foundation models in representation learning, different pretraining objectives and architectures lead to diverse biological and structural information encoded in their representations. To obtain more informative representations enriched with diverse biological and structural knowledge, we adopt a multi-encoder architecture that leverages multiple RNA foundation models and integrates the information encoded by their pretrained representations, as illustrated in Figure 2. Importantly, the proposed framework is not restricted to a fixed set of encoders and can be flexibly extended to incorporate other RNA foundation models with different representation characteristics. In this work, we instantiate the framework using three representative RNA foundation models. RNA-MSM (Zhang et al., 2024) leverages multiple sequence alignments to encode evolutionary covariation signals. ERNIE-RNA (Yin et al., 2025) incorporates base-pair-aware attention mechanisms to model local structural dependencies. RiNALMo (Penić et al., 2025) is pretrained on large-scale RNA corpora to learn contextual sequence representations associated with long-range contextual information. Implementation details are provided in Appendix C.

3.3 Adaptive Progressive Embedding Alignment

Motivation. The integration of embeddings from multiple RNA foundation models is challenging because different pretraining strategies often produce heterogeneous feature representations with inconsistent distributions and representation patterns. Naively summing or concatenating these heterogeneous embeddings often results in representation imbalance, causing training instability and limiting the effective integration of heterogeneous representations.

Feature Alignment. To reduce representation discrepancies across RNA foundation models, we first introduce a feature alignment module that maps heterogeneous encoder representations into a shared latent space. Given encoder embeddings $\{\mathbf{E}_i \in \mathbb{R}^{B\times L\times H_i}\}_{i=1}^{N}$, where N denotes the number of encoders, B is the batch size, L is the RNA sequence length, and H_i represents the hidden dimension of the i-th encoder, each representation is normalized and projected into a shared latent space $\mathbb{R}^H$:

$$\mathbf{Z}_i = \phi_i(\mathrm{LN}(\mathbf{E}_i)), \quad \mathbf{Z}_i \in \mathbb{R}^{B\times L\times H}, \tag{1}$$

where $\mathrm{LN}(\cdot)$ denotes layer normalization and $\phi_i : \mathbb{R}^{H_i} \to \mathbb{R}^H$ is a learnable linear projection.

Weighted Fusion. We then learn adaptive weights to combine the aligned representations. Specifically, the fusion weights $\mathbf{w}$ are obtained by applying Softmax to learnable parameters $\boldsymbol{\alpha} \in \mathbb{R}^N$:

$$w_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^{N}\exp(\alpha_j)}, \quad \mathbf{F} = \sum_{i=1}^{N} w_i \mathbf{Z}_i. \tag{2}$$

3.4 BIOLOGICALLY GUIDED MIXTURE-OF-EXPERTS

Motivation. RNA sequences with different statistical and compositional properties often exhibit distinct secondary structure formation patterns, suggesting that different inputs may require adaptive representation pathways rather than a uniform modeling strategy. To incorporate such variability into model design, we employ several lightweight biological descriptors, including sequence length, nucleotide composition statistics, and local k-mer distributions, as an initial instantiation of weak structural priors. From a statistical perspective, variations in sequence length, nucleotide composition, and local k-mer distributions often correspond to different families, as they affect base-pairing patterns and global folding constraints. Importantly, the proposed framework is not restricted to these descriptors and can be naturally extended to incorporate other biologically relevant signals. To incorporate such biological priors into representation learning, we introduce a biologically guided Mixture-of-Experts (BioMoE) module. Specifically, these descriptors are used as auxiliary routing signals to guide adaptive expert selection, enabling different experts to specialize in distinct representation patterns under weak biological supervision.

Overall Design of the BioMoE. Unlike conventional Mixture-of-Experts architectures that perform token-level routing, BioMoE adopts a sequence-level routing strategy, allowing each RNA sequence to be processed by a consistent subset of experts that better aligns with its global structural characteristics. Each expert is implemented as a convolutional neural network (CNN) to capture local structural patterns and base-pairing interactions. Given the integrated representation $\mathbf{F}$, the routing network predicts expert routing scores $\mathbf{g} \in \mathbb{R}^E$ based on biologically guided sequence descriptors, where E denotes the number of experts. The construction of these routing signals is described in the next subsection. We then select the top-K experts according to $\mathbf{g}$ and compute the refined representation as:

$$\mathbf{F}' = \sum_{e \in \mathcal{P}} \hat{g}_e \, \psi_e(\mathbf{F}), \qquad \hat{g}_e = \frac{\exp(g_e)}{\sum_{j \in \mathcal{P}} \exp(g_j)}, \tag{3}$$

where $\mathcal{P}$ denotes the index set of the selected experts and $\psi_e(\cdot)$ represents the CNN-based transformation associated with the e-th expert.

Load Balancing Regularization. To prevent routing collapse and encourage balanced expert utilization, we introduce an auxiliary load-balancing regularization during training. Specifically, given routing weights $\mathbf{g}^{(b)} \in \mathbb{R}^E$ for the b-th sequence, we first compute the average routing distribution across a mini-batch of size B:

$$\bar{\mathbf{g}} = \frac{1}{B} \sum_{b=1}^{B} \mathrm{Softmax}\Big(\mathbf{g}^{(b)}\Big). \quad \bar{\mathbf{g}} \in \mathbb{R}^E, \tag{4}$$

where E denotes the number of experts. To encourage balanced expert allocation, we regularize the routing distribution toward a uniform prior $\mathbf{u} \in \mathbb{R}^E$, where $u_e = 1/E$. The load-balancing objective is defined as:

$$\mathcal{L}_{\mathrm{bal}} = D_{\mathrm{KL}} \left(\bar{\mathbf{g}} \parallel \mathbf{u} \right). \tag{5}$$

Minimizing $\mathcal{L}_{\mathrm{bal}}$ encourages more uniform expert utilization and alleviates expert dominance during training.

Biologically Guided Routing. To guide different RNA sequences toward distinct expert parameter subspaces, we introduce biologically informed routing signals for expert selection. Rather than relying solely on high-dimensional encoder embeddings, we construct a sequence-level statistical descriptor $\mathbf{s} \in \mathbb{R}^{21}$ from the input RNA sequence, including sequence length, nucleotide composition, and 2-mer distribution statistics. These descriptors provide coarse-grained biological cues related to RNA structural formation and serve as weak supervisory signals for routing decisions. Importantly, this statistical descriptor is not intended as a fixed handcrafted prior, but rather as an initial and easily replaceable routing signal. To enable adaptive refinement, the descriptor is further transformed through learnable linear projections, allowing the routing network to automatically adjust and extract routing-relevant features:

$$\mathbf{h} = \delta(\phi_1(\mathrm{LN}(\mathbf{s}))), \tag{6}$$

$$\mathbf{g} = \phi_2(\mathrm{LN}(\mathbf{h})), \quad \mathbf{g} \in \mathbb{R}^E, \tag{7}$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ denote learnable linear projections, $\delta(\cdot)$ represents the ReLU activation function, and E is the number of experts. The resulting routing scores $\mathbf{g} = [g_1, \ldots, g_E]^\top$ are used to select the top-K experts for each RNA sequence.

Table 1: Performance comparison on standard RNA secondary structure benchmarks. RNAStrAlign is used for within-family evaluation and is split into training, validation, and test sets. Models trained on RNAStrAlign are directly evaluated on ArchiveII for generalization evaluation without further fine-tuning. bpRNA-TS0 serves as a large-scale benchmark for comprehensive evaluation.

Model	RNAStrAlign			ArchiveII			bpRNA-TS0		
	Recall	Precision	F_1	Recall	Precision	F_1	Recall	Precision	F_1
ContraFold [Bioinfo'06]	0.647	0.600	0.621	0.649	0.593	0.617	0.655	0.482	0.541
RNAstructure [BMC'10]	0.569	0.537	0.551	0.601	0.554	0.573	0.622	0.446	0.505
RNAfold [Mol Biol'11]	0.568	0.515	0.539	0.550	0.611	0.577	0.631	0.446	0.508
LinearFold [Bioinfo'19]	0.576	0.523	0.547	0.609	0.549	0.575	0.633	0.447	0.510
E2Efold [ICLR'20]	0.788	0.866	0.821	0.660	0.734	0.686	0.129	0.140	0.130
UFold [NAR'22]	0.965	0.959	0.962	0.890	0.876	0.881	0.711	0.587	0.630
RFold [ICML'24]	0.973	0.981	0.977	0.910	0.938	0.921	0.635	0.692	0.644
sincFold [BIB'24]	0.984	0.990	0.987	0.875	0.936	0.895	0.619	0.674	0.623
PriFold [AAAI'25]	0.974	0.974	0.973	0.903	0.911	0.904	0.710	**0.810**	0.737
DEPFOLD [ICLR'25]	0.983	0.988	0.985	0.907	0.941	0.920	0.656	0.730	0.676
BioSynFold	**0.992**	**0.992**	**0.992**	**0.934**	**0.954**	**0.941**	**0.764**	0.765	**0.753**

3.5 Prediction Head

The proposed framework is decoder-agnostic and can be naturally integrated with different RNA secondary structure decoders. In this work, we adopt the structural decoder of sincFold (Bugnon et al., 2024) to transform the refined representation $\mathbf{F}'$ into RNA contact maps.

The sincFold decoder employs a two-stage residual architecture to model both local and long-range base-pairing interactions, producing an $L \times L$ contact probability matrix. To satisfy RNA structural constraints, several regularization and post-processing strategies are applied following sincFold. Matrix symmetry is enforced by combining the predicted contact matrix with its transpose, while sparsity is encouraged through an $\mathcal{L}_1$ regularization term. In addition, cross-entropy losses are imposed on both intermediate and final predictions for deep supervision. During post-processing, only the maximum activation in each row and column is retained to ensure that each nucleotide pairs with at most one partner. The prediction objective consists of a final contact-map cross-entropy loss $\mathcal{L}_\alpha$, an intermediate supervision loss $\mathcal{L}_\beta$, and a sparsity regularization term $\mathcal{L}_1$:

$$\mathcal{L}_{\text{pred}} = \mathcal{L}_\alpha + \lambda_\beta \mathcal{L}_\beta + \lambda_1 \mathcal{L}_1, \tag{8}$$

where λ_β and λ_1 are balancing coefficients.

To further stabilize expert utilization, we additionally incorporate the load-balancing regularization term $\mathcal{L}_{\text{bal}}$ introduced in the previous subsection. The overall training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda \mathcal{L}_{\text{bal}}, \tag{9}$$

where λ controls the contribution of the routing regularization term. The entire framework is trained end-to-end through joint optimization of the fusion, routing, and prediction modules.

4 Experiments

4.1 Experimental Setup

Dataset Description. Following the experimental protocols of prior work (Wang & Cohen, 2025), we adopt a unified evaluation pipeline across four benchmark datasets, each serving a distinct purpose. RNAStrAlign (Tan et al., 2017) is used for within-family evaluation and split into training, validation, and test sets. Models trained on RNAStrAlign are then directly evaluated on ArchiveII (Sloma & Mathews, 2016), where the entire dataset is used as a test set to assess cross-dataset generalization without further fine-tuning. For large-scale benchmark evaluation on the bpRNA dataset (Singh et al., 2019), we follow the established benchmark splits by training on TR0, validating on VL0, and testing on TS0. To further evaluate cross-family generalization, we construct an augmented training set TR1 by applying commonly used mutation-based augmentation to TR0 as in prior work (Lorenz et al., 2011). The model is trained on TR1 and evaluated on bpRNA-new (Kalvari et al., 2018), which contains sequences from 1,500 RNA families unseen during training.

Table 2: Evaluation on two challenging RNA structure prediction settings: long-range base-pair prediction on bpRNA-TS0 and cross-family generalization on bpRNA-new.

Model	bpRNA-TS0 (Long-range)			bpRNA-new (Cross-family)		
	Recall	Precision	F_1	Recall	Precision	F_1
ContraFold [Bioinfo'06]	0.614	0.505	0.536	0.736	0.578	0.639
RNAstructure [BMC'10]	0.603	0.490	0.523	0.703	0.542	0.604
RNAfold [Mol Biol'11]	0.627	0.499	0.539	0.720	0.552	0.617
LinearFold [Bioinfo'19]	0.629	0.502	0.542	0.719	0.542	0.615
UFold [NAR'22]	0.711	0.630	0.653	0.695	0.527	0.590
RFold [ICML'24]	0.647	0.666	0.643	0.306	0.528	0.369
sincFold [BIB'24]	0.594	0.683	0.634	0.663	0.581	0.603
DEPFOLD [ICLR'25]	0.689	0.715	0.689	0.624	**0.650**	0.621
BioSynFold	**0.743**	**0.785**	**0.764**	**0.744**	0.619	**0.663**

Compared Methods. To comprehensively evaluate our method, we select representative RNA secondary structure prediction approaches spanning diverse modeling paradigms for comparison. Energy-driven folding methods include CONTRAfold (Do et al., 2006), RNAstructure (Reuter & Mathews, 2010), RNAfold (Lorenz et al., 2011), LinearFold (Huang et al., 2019) and MXfold2 (Sato et al., 2021). Deep learning–based methods include E2Efold (Chen et al., 2020), UFold (Fu et al., 2022), RFold (Tan et al., 2024), sincFold (Bugnon et al., 2024), PriFold (Yang et al., 2025) and DEPFOLD (Wang & Cohen, 2025).

Evaluation Metrics. Following established benchmarks (Fu et al., 2022; Wang & Cohen, 2025), we evaluate model performance by comparing the predicted base-pair set with the ground-truth set. Precision measures the fraction of predicted base pairs that are correct, while Recall quantifies the fraction of true base pairs that are successfully recovered by the model. To provide a single balanced metric that accounts for both accuracy and coverage, we report the F_1 score. All metrics, including the F_1 score, are computed at the sample level and then macro-averaged across sequences, ensuring that each RNA sequence contributes equally to the final evaluation regardless of its length.

4.2 Comparison with State-of-the-Art Methods

We evaluate our method against existing approaches on multiple RNA secondary structure prediction benchmarks. We organize the experiments into two categories: (1) standard benchmark evaluations on well-studied settings; (2) challenging scenarios, including long-range base-pair prediction and cross-family generalization. Not all existing methods report results on these challenging benchmarks. As a result, the number of competing methods is limited in these settings. Our method is highlighted in blue. The best results for each metric are shown in bold.

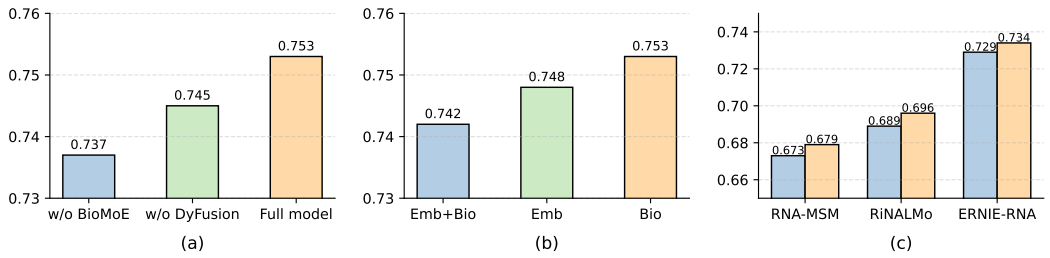


Figure 3: Ablation studies of different modules on bpRNA-TS0. **(a)** Effect of removing key components: "w/o BioMoE" indicates the removal of the biologically guided Mixture-of-Experts module; "w/o DyFusion" indicates the removal of the adaptive fusion module. **(b)** Effect of different information guiding the router: "Emb+Bio" uses both RNA embeddings and biological priors; "Emb" uses only embeddings; "Bio" uses only biological priors. **(c)** Effect of BioMoE with weak biological signals on different single RNA foundation models.

Overall Performance on Standard Benchmarks. Table 1 summarizes results on RNAStrAlign (within-family), ArchiveII (cross-dataset), and bpRNA-TS0 (large-scale benchmark). Classical energy-based methods achieve limited performance across all settings, while learning-based approaches provide substantial improvements. BioSynFold consistently achieves the best overall performance, achieving the highest F_1 scores across all benchmarks, which reflects a superior balance between precision and recall. On RNAStrAlign, BioSynFold reaches near-saturated performance, demonstrating its effectiveness in modeling within-family structural patterns. On ArchiveII, most methods experience noticeable degradation due to distribution shift, whereas BioSynFold maintains strong performance without any adaptation, indicating robust cross-dataset generalization. On the large-scale bpRNA-TS0 benchmark, BioSynFold achieves the highest F_1 score with a well-balanced trade-off between precision and recall, showing stable performance across diverse RNA structures. Overall, these results demonstrate that BioSynFold not only excels in standard within-family settings but also generalizes effectively to cross-dataset and large-scale scenarios.

Challenging Evaluation on Long-range and Cross-family Settings. Table 2 reports results on long-range base-pair prediction and cross-family generalization. For long-range base pairs, defined as nucleotide pairs separated by more than half of the sequence length, BioSynFold achieves the best precision, recall, and F_1 scores among all compared methods, demonstrating its effectiveness in capturing long-range dependencies and global structural patterns. For cross-family generalization on bpRNA-new, this setting introduces substantial distribution shift, under which several learning-based methods show noticeable performance degradation. In contrast, energy-based methods become more competitive, reflecting their robustness in out-of-distribution scenarios. Despite this challenge, BioSynFold achieves the highest F_1 score, demonstrating a strong balance between precision and recall, even when individual metrics are not always the best. These results suggest that BioSynFold can effectively handle cross-family variability and maintain reliable structure prediction on unseen RNA families.

Table 3: Ablation of different fusion strategies after feature alignment on bpRNA-TS0.

Methods	Recall	Precision	F_1
Weighted Average	0.749	0.750	0.737
Concatenation	0.748	0.759	0.741
Cross-Attention	0.759	0.757	0.745
BioMoE (Ours)	**0.764**	**0.765**	**0.753**

4.3 Ablation Studies and Analysis

In this section, we aim to explore three key questions. First, whether each module in our progressive integration strategy contributes effectively to the overall performance. Second, whether model-level heterogeneity is fully leveraged. Third, whether the BioMoE module effectively captures data-level variability.

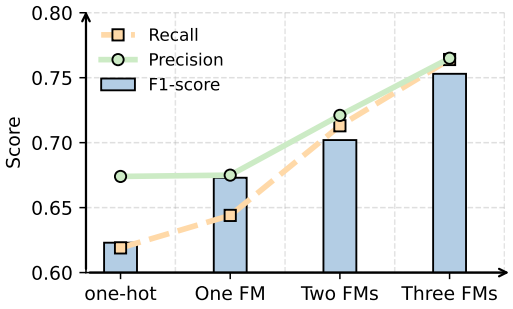


Figure 4: Ablation study of the multi-model fusion framework on bpRNA-TS0.

Effectiveness of Each Module in Progressive Integration Strategy. Figure 3(a) and (b) present the ablation studies on the progressive fusion module. In (a), removing any key component leads to a performance drop, demonstrating that both adaptive fusion and BioMoE are essential. This suggests that progressively combining multi-source embeddings effectively exploits diverse biological knowledge captured by different models. In (b), we evaluate different guiding signals for routing. Sequence-level statistical features achieve the best F_1 score, indicating their effectiveness in capturing data-level variability. In contrast, using sequence embeddings, either alone or combined with statistical features, does not improve performance, suggesting that high-dimensional embeddings may introduce redundancy and weaken routing effectiveness.

Table 3 compares BioMoE with several alternative static fusion strategies, including concatenation, weighted averaging, and cross-attention. To ensure a fair comparison, additional CNN layers are introduced for concatenation and weighted averaging, while stacked cross-attention layers are used for

the cross-attention baseline to match model capacity. Despite this, these static fusion strategies consistently underperform our method, indicating that simply increasing fusion capacity is insufficient to handle the inherent data-level variability across RNA sequences. In contrast, BioMoE performs adaptive fusion through a biologically guided routing mechanism, enabling condition-dependent expert selection and more effective integration of heterogeneous representations.

Exploiting Model-Level Diversity. As illustrated in Figure 4, "One FM" corresponds to using RNA-MSM alone; "Two FMs" combines RNA-MSM and RiNALMo; and "Three FMs" further integrates ERNIE-RNA. The results show that performance consistently improves as more foundation models are incorporated, indicating that richer pretrained representations lead to better RNA secondary structure prediction. This demonstrates the effectiveness of our fusion framework in leveraging model-level diversity.

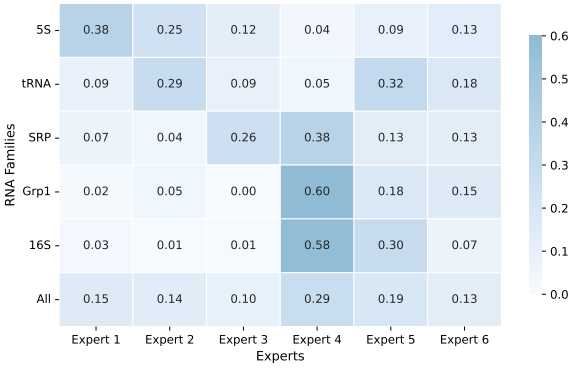


Figure 5: Expert utilization across RNA families (5S, tRNA, SRP, Grp1, 16S, and All).

Appendix B presents more detailed experimental results. As shown, incorporating additional RNA foundation models consistently improves performance, demonstrating that different pretrained representations encode non-redundant structural information. Notably, even models with relatively weak standalone performance still provide substantial gains when integrated, suggesting that they capture distinct inductive biases that are not fully reflected in stronger models. This indicates that the improvement is not merely driven by model quality, but by the diversity of representation spaces induced by different pretraining paradigms.

Capturing Data-Level Variability with BioMoE. Figure 3(c) shows the effect of BioMoE guided by weak biological signals on single RNA foundation models. The results demonstrate that even when applied to a single foundation model, BioMoE consistently improves performance under weak supervision from biological signals. Figure 5 demonstrates that the routing mechanism guided by weak signals derived from sequence-level statistical and compositional properties leads to clear family-dependent expert preferences. This indicates that BioMoE enables condition-aware specialization of experts under data-level variability across sequences, resulting in more adaptive utilization of heterogeneous representations. In addition, the relatively balanced activation in the All condition suggests that the BioMoE is well-optimized without severe expert collapse or under-utilization.

5 Conclusion

In this work, we revisit RNA secondary structure prediction from a representation learning perspective, focusing on how to better utilize and integrate representations from RNA foundation models. We identify representation quality as a key bottleneck and propose BioSynFold, a biology-guided synergistic learning framework. From the model perspective, we integrate multiple RNA foundation models within a unified synergistic learning framework with a progressive fusion strategy to exploit heterogeneous representations. From the data perspective, we leverage sequence-level statistical signals to guide a mixture-of-experts routing mechanism for improved representation learning. We acknowledge two key **limitations**. First, the biological signals used to guide MoE routing are relatively simple, and richer biological priors may further improve performance. Second, learning more effective fusion strategies for multi-model representations remains an open direction. Despite these limitations, our approach has great potential beyond secondary structure prediction, inspiring RNA representation learning and enabling application to a broader range of RNA-related tasks.

AI Use Statement

In this work, we used generative AI tools to draft portions of the manuscript, edit and polish the writing, and assist in finding related literature. We did not use generative AI for research ideation or

execution, synthetic data generation, or mathematical proofs; other tasks requiring disclosure were not applicable. The authors reviewed the AI-assisted text and references and take responsibility for all claims, citations, results, and other content in the final manuscript.

REFERENCES

Manato Akiyama and Yasubumi Sakakibara. Informative rna base embedding for rna structural alignment and clustering by deep representation learning. *NAR Genomics and Bioinformatics*, 4 (1):lqac012, 2022.

Ravi P Barnwal, Fan Yang, and Gabriele Varani. Applications of nmr to structure determination of rnas large and small. *Archives of Biochemistry and Biophysics*, 628:42–56, 2017.

Leandro A Bugnon, Leandro Di Persia, Matias Gerard, Jonathan Raad, Santiago Prochetto, Emilio Fenoy, Uciel Chorostecki, Federico Ariel, Georgina Stegmayer, and Diego H Milone. sincfold: end-to-end learning of short-and long-range interactions in rna secondary structure. *Briefings in Bioinformatics*, 25(4), 2024.

Jiayang Chen, Zhihang Hu, Siqi Sun, Qingxiong Tan, Yixuan Wang, Qinze Yu, Licheng Zong, Liang Hong, Jin Xiao, Tao Shen, et al. Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *arXiv preprint arXiv:2204.00300*, 2022.

Xinshi Chen, Yu Li, Ramzan Umarov, Xin Gao, and Le Song. Rna secondary structure prediction by learning unrolled algorithms. In *8th International Conference on Learning Representations, ICLR 2020*. International Conference on Learning Representations, ICLR, 2020.

Matthew D Disney. Targeting rna with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *Journal of the American Chemical Society*, 141(17):6776–6790, 2019.

Chuong B Do, Daniel A Woods, and Serafim Batzoglou. Contrafold: Rna secondary structure prediction without physics-based models. *Bioinformatics*, 22(14):e90–e98, 2006.

Laiyi Fu, Yingxin Cao, Jie Wu, Qinke Peng, Qing Nie, and Xiaohui Xie. Ufold: fast and accurate rna secondary structure prediction with deep learning. *Nucleic Acids Research*, 50(3):e14–e14, 2022.

Liang Huang, He Zhang, Dezhong Deng, Kai Zhao, Kaibo Liu, David A Hendrix, and David H Mathews. Linearfold: linear-time approximate rna folding by 5'-to-3'dynamic programming and beam search. *Bioinformatics*, 35(14):i295–i304, 2019.

Ryland W Jackson, Claire M Smathers, and Aaron R Robart. General strategies for rna x-ray crystallography. *Molecules*, 28(5):2111, 2023.

Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, and Anton I Petrov. Rfam 13.0: shifting to a genome-centric resource for non-coding rna families. *Nucleic acids research*, 46(D1):D335–D342, 2018.

Bjarne Knudsen and Jotun Hein. Pfold: Rna secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Research*, 31(13):3423–3428, 2003.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 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.

Haiyun Ma, Xinyu Jia, Kaiming Zhang, and Zhaoming Su. Cryo-em advances in rna structure determination. *Signal Transduction and Targeted Therapy*, 7(1):58, 2022.

Edward J Merino, Kevin A Wilkinson, Jennifer L Coughlan, and Kevin M Weeks. Rna structure analysis at single nucleotide resolution by selective 2 '-hydroxyl acylation and primer extension (shape). *Journal of the American Chemical Society*, 127(12):4223–4231, 2005.

Stefanie A Mortimer, Mary Anne Kidwell, and Jennifer A Doudna. Insights into rna structure and function from genome-wide studies. *Nature Reviews Genetics*, 15(7):469–479, 2014.

Rafael Josip Penić, Tin Vlašić, Roland G Huber, Yue Wan, and Mile Šikić. Rinalmo: General-purpose rna language models can generalize well on structure prediction tasks. *Nature Communications*, 16(1):5671, 2025.

Jessica S Reuter and David H Mathews. Rnastructure: software for rna secondary structure prediction and analysis. *BMC Bioinformatics*, 11(1):129, 2010.

Elena Rivas, Jody Clements, and Sean R Eddy. A statistical test for conserved rna structure shows lack of evidence for structure in lncrnas. *Nature Methods*, 14(1):45–48, 2017.

Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021.

Thomas C Roberts, Robert Langer, and Matthew JA Wood. Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery*, 19(10):673–694, 2020.

Kengo Sato, Manato Akiyama, and Yasubumi Sakakibara. Rna secondary structure prediction using deep learning with thermodynamic integration. *Nature Communications*, 12(1):941, 2021.

Jaswinder Singh, Jack Hanson, Kuldip Paliwal, and Yaoqi Zhou. Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature Communications*, 10(1):5407, 2019.

Michael F Sloma and David H Mathews. Exact calculation of loop formation probability identifies folding motifs in rna secondary structures. *RNA*, 22(12):1808–1818, 2016.

Cheng Tan, Zhangyang Gao, Xingran Chen, Ge Wang, Lirong Wu, Jun Xia, Jiangbin Zheng, Stan Z Li, et al. Deciphering rna secondary structure prediction: A probabilistic k-rook matching perspective. In *Forty-first International Conference on Machine Learning*, 2024.

Zhen Tan, Yinghan Fu, Gaurav Sharma, and David H Mathews. Turbofold ii: Rna structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic acids research*, 45(20):11570–11581, 2017.

Yue Wan, Kun Qu, Qiangfeng Cliff Zhang, Ryan A Flynn, Ohad Manor, Zhengqing Ouyang, Jiajing Zhang, Robert C Spitale, Michael P Snyder, Eran Segal, et al. Landscape and variation of rna secondary structure across the human transcriptome. *Nature*, 505(7485):706–709, 2014.

Feng Wang, Travis Zuroske, Jonathan K Watts, et al. Rna therapeutics on the rise. *Nat Rev Drug Discov*, 19(7):441–442, 2020.

Ke Wang and Shay B Cohen. Depfold: Rna secondary structure prediction as dependency parsing. In *The Thirteenth International Conference on Learning Representations*, 2025.

Ning Wang, Jiang Bian, Yuchen Li, Xuhong Li, Shahid Mumtaz, Linghe Kong, and Haoyi Xiong. Multi-purpose rna language modelling with motif-aware pretraining and type-guided fine-tuning. *Nature Machine Intelligence*, 6(5):548–557, 2024.

Xi-Wen Wang, Chu-Xiao Liu, Ling-Ling Chen, and Qiangfeng Cliff Zhang. Rna structure probing uncovers rna structure-dependent biological functions. *Nature chemical biology*, 17(7):755–766, 2021.

Yongtian Wang, Yewei Shen, Jiahao Li, Tao Wang, Jiajie Peng, and Xuequn Shang. Enhanced rna secondary structure prediction through integrative deep learning and structural context analysis. *Nucleic Acids Research*, 53(11):gkaf533, 2025a.

Zhen Wang, Yizhen Feng, Qingwen Tian, Ziqi Liu, Pengju Yan, and Xiaolin Li. Rnadifffold: generative rna secondary structure prediction using discrete diffusion models. *Briefings in Bioinformatics*, 26(1):bbae618, 2025b.

Chenchen Yang, Hao Wu, Tao Shen, Kai Zou, and Siqi Sun. Prifold: Biological priors improve rna secondary structure predictions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 950–958, 2025.

Weijie Yin, Zhaoyu Zhang, Shuo Zhang, Liang He, Ruiyang Zhang, Rui Jiang, Gan Liu, Jingyi Wang, Xuegong Zhang, Tao Qin, et al. Ernie-rna: an rna language model with structure-enhanced representations. *Nature Communications*, 16(1):10076, 2025.

Luciano I Zablocki, Leandro A Bugnon, Matias Gerard, Leandro Di Persia, Georgina Stegmayer, and Diego H Milone. Comprehensive benchmarking of large language models for rna secondary structure prediction. *Briefings in Bioinformatics*, 26(2):bbaf137, 2025.

He Zhang, Liang Zhang, Ang Lin, Congcong Xu, Ziyu Li, Kaibo Liu, Boxiang Liu, Xiaopin Ma, Fanfan Zhao, Huiling Jiang, et al. Algorithm for optimized mrna design improves stability and immunogenicity. *Nature*, 621(7978):396–403, 2023.

Jinsong Zhang, Yuhan Fei, Lei Sun, and Qiangfeng Cliff Zhang. Advances and opportunities in rna structure experimental determination and computational modeling. *Nature Methods*, 19(10): 1193–1207, 2022.

Yikun Zhang, Mei Lang, Jiuhong Jiang, Zhiqiang Gao, Fan Xu, Thomas Litfin, Ke Chen, Jaswinder Singh, Xiansong Huang, Guoli Song, et al. Multiple sequence alignment-based rna language model and its application to structural inference. *Nucleic Acids Research*, 52(1):e3–e3, 2024.

Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, 2003.

A Implementation Details

All models are implemented in PyTorch. Training is performed on a single RTX 4090 GPU with 24GB memory. The batch size is 4, and models are trained for a maximum of 200 epochs with early stopping patience of 30. The optimizer is Adam with a learning rate of 1×10^{-4}. The balancing coefficient is set to $\lambda = 0.01$. The model uses heterogeneous embeddings fused into a 512-dimensional space. For the MoE module, the number of experts is set to twice the number of underlying RNA foundation models, and a top-2 routing strategy is used for expert selection. The parameters of the prediction head are kept exactly the same as in sincFold (Bugnon et al., 2024).

B Additional Experiments

As a supplement to the main results, we conduct three additional experiments to further validate and analyze our framework. These experiments provide deeper insights into the effectiveness of the proposed modules and the behavior of the model under different settings.

Table 4: Ablation study of different RNA foundation model combinations on bpRNA-TS0.

RNA-MSM	RiNALMo	ERNIE-RNA	Recall	Precision	F_1-score
✓			0.644	0.675	0.637
	✓		0.692	0.726	0.689
		✓	0.758	0.727	0.729
✓	✓		0.713	0.721	0.702
✓		✓	0.760	0.762	0.751
	✓	✓	0.763	0.759	0.750
✓	✓	✓	**0.764**	**0.765**	**0.753**

Additional Analysis of RNA Foundation Model Combinations. As a complement to Figure 4, Table 4 provides a more detailed comparison of different RNA foundation model combinations. We observe that incorporating an additional pretrained model consistently improves performance, even when the added model individually achieves relatively weaker results. This suggests that different RNA foundation models capture diverse structural characteristics that are not fully represented by a single model alone. Moreover, combination involving all three models generally achieve higher and more balanced Recall, Precision, and F_1 scores, indicating improved robustness and predictive consistency. Overall, these findings demonstrate that integrating multiple RNA foundation models enables the framework to effectively exploit the diversity of biological and structural information encoded by different pretrained representations, leading to more comprehensive RNA structure representations and improved prediction performance.

Effect of Expert Number on Performance. Figure 6 shows that performance peaks when using 6 experts. Using fewer experts limits the model's capacity to capture diverse structural patterns, resulting in lower Recall, Precision, and F_1 scores. In contrast, using more than 6 experts introduces redundancy among experts and slightly destabilizes training, which can slightly degrade perfor-

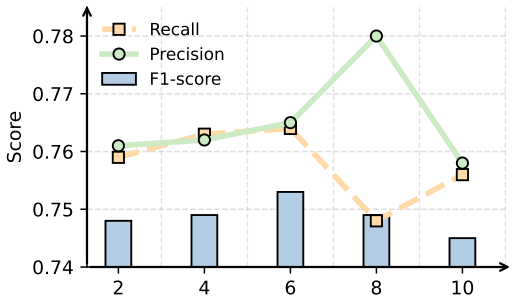


Figure 6: Performance is evaluated with 2, 4, 6, 8, and 10 experts on bpRNA-TS0.

Table 5: Performance across different random seeds on bpRNA-new.

Seed	Recall	Precision	F_1
1	0.698	0.647	0.656
2	0.744	0.619	0.663
3	0.704	0.641	0.656
4	0.703	0.647	0.659
5	0.735	0.612	0.656
42	0.716	0.638	0.661
Mean $\pm$ Std	0.717 ± 0.018	0.634 ± 0.015	0.659 ± 0.003

mance. Importantly, the observed improvements are not solely due to an increase in the number of parameters, but rather reflect the model's ability to adaptively allocate complementary structural knowledge across experts, enabling more effective utilization of the diversity among pretrained RNA representations.

Model Stability under Different Random Seeds. Since previous work does not report evaluation under multiple random seeds, we do not include standard deviation results in the main tables for fair comparison. To further assess the stability of our method, we conduct additional experiments on the bpRNA-new dataset using six different random seeds while keeping all other settings identical. The results show consistent performance across different seeds, with only minor fluctuations, indicating that our method is stable and robust to random initialization.

C Embedding Extraction and Computational Cost

We extract RNA representations from pretrained RNA foundation models using a unified inference pipeline, following the general paradigm of Zablocki et al. (2025). Given an input RNA dataset, sequences are preprocessed and fed into the corresponding models for forward inference. We extract residue-level hidden representations from the final model layers, remove special tokens (e.g., CLS and END), and align the embeddings with the original nucleotide sequences. When necessary, embeddings are truncated or padded to ensure one-to-one correspondence with nucleotides. The extracted embeddings are then stored in HDF5 format and indexed by sequence identifiers for efficient downstream access.

To reduce the computational overhead of using multiple RNA foundation models, we adopt an offline embedding extraction strategy, where representations are precomputed and cached before downstream training and inference. Thus, the computational cost of the foundation models is decoupled from the structure prediction stage. Since the encoders operate independently, embedding extraction can also be parallelized across available computational resources.

To quantify the computational cost, we measured the embedding extraction efficiency on the bpRNA dataset using a single RTX 4090 GPU with batch size 4. The average extraction time, GPU memory consumption, and cached embedding storage cost for each RNA foundation model are reported in Table 6. These measurements include model initialization and loading overhead.

Table 6: Computational cost of embedding extraction on the bpRNA dataset using a single RTX 4090 GPU with batch size 4.

Foundation Model	GPU Memory	Time/RNA	Storage/RNA
ERNIE-RNA	329 MB	15.0 ms	0.716 MB
RiNALMo	2483 MB	48.4 ms	0.597 MB
RNA-MSM	366 MB	6.2 ms	0.358 MB

The cached embedding storage cost is manageable, averaging 0.716 MB, 0.597 MB, and 0.358 MB per RNA sequence for ERNIE-RNA, RiNALMo, and RNA-MSM, respectively. In practice, extraction throughput can be further improved by increasing the batch size and processing larger datasets.

Table 7: Computational cost of downstream structure prediction.

Model	Input Representation	GPU Memory	Inference Time/RNA
sincFold	One-hot	372 MB	2.16 ms
BioSynFold	Multi-foundation-model embeddings	377 MB	2.78 ms

After embedding extraction, the downstream BioSynFold predictor introduces only marginal computational overhead. The trainable components require only 11.2 MB of parameter storage. Compared with SincFold using one-hot sequence representations, BioSynFold increases the inference time from 2.16 ms/RNA to 2.78 ms/RNA, while GPU memory consumption increases from 372 MB to 377 MB, as shown in Table 7.

These results show that although BioSynFold leverages multiple pretrained RNA foundation models to obtain richer representations, the additional computational cost during downstream prediction is limited. The main computational overhead is incurred during offline embedding extraction, which can be amortized through caching, parallel extraction, and large-scale batch processing.

D Further Analysis of Biologically Guided Routing

Table 8: Ablation of sequence-level routing descriptors on bpRNA-TS0. All variants use the same model architecture; only the routing descriptors differ.

Routing descriptor	Recall	Precision	F_1
Length only	0.751	0.752	0.741
Base composition only	0.756	0.756	0.746
2-mer only	0.756	0.761	0.748
Shuffled descriptors	0.763	0.757	0.747
Full (BioMoE)	**0.764**	**0.765**	**0.753**

Effect of Routing Descriptors. Table 8 compares the full sequence-level descriptor with individual feature groups and shuffled descriptors. The full descriptor achieves the highest F_1 score (0.753), while length alone performs worst (0.741). Nucleotide composition and 2-mer statistics each improve upon length alone, suggesting that the feature groups provide useful routing information beyond sequence length. Shuffling the descriptors reduces F_1 to 0.747, although the small gap to the full model warrants caution in attributing the improvement specifically to biologically meaningful routing.

E Further Analysis of Representation Similarity Among Foundation Models

To examine whether the three RNA foundation models produce distinct representations, we compute linear centered kernel alignment (CKA) between their cached embeddings on bpRNA. We sample 2,000 sequences shared by all three encoders using random seed 42. For sequence-level CKA, we mean-pool the residue embeddings of each sequence and compute CKA across the 2,000 sequence representations. For residue-level CKA, we compute CKA across nucleotide positions separately for each of the first 200 sampled sequences, then average the resulting scores.

Table 9 shows that representation similarity varies across encoder pairs. In particular, sequence-level CKA ranges from 0.211 to 0.438, indicating that the mean-pooled embeddings are not identical across models. Residue-level similarity is higher for some pairs, notably ERNIE-RNA and RiNALMo (0.855). These observations support the presence of heterogeneous representations and motivate their integration.

Table 9: Linear CKA between RNA foundation-model representations on bpRNA. Sequence-level CKA is computed across 2,000 mean-pooled RNA representations; residue-level CKA is the mean of per-sequence CKA scores for 200 RNAs. The two columns use different sample units and should not be compared directly.

Encoder pair	Sequence-level CKA	Residue-level CKA
ERNIE-RNA / RNA-MSM	0.211	0.633
ERNIE-RNA / RiNALMo	0.438	0.855
RNA-MSM / RiNALMo	0.357	0.711

F VISUALIZATION

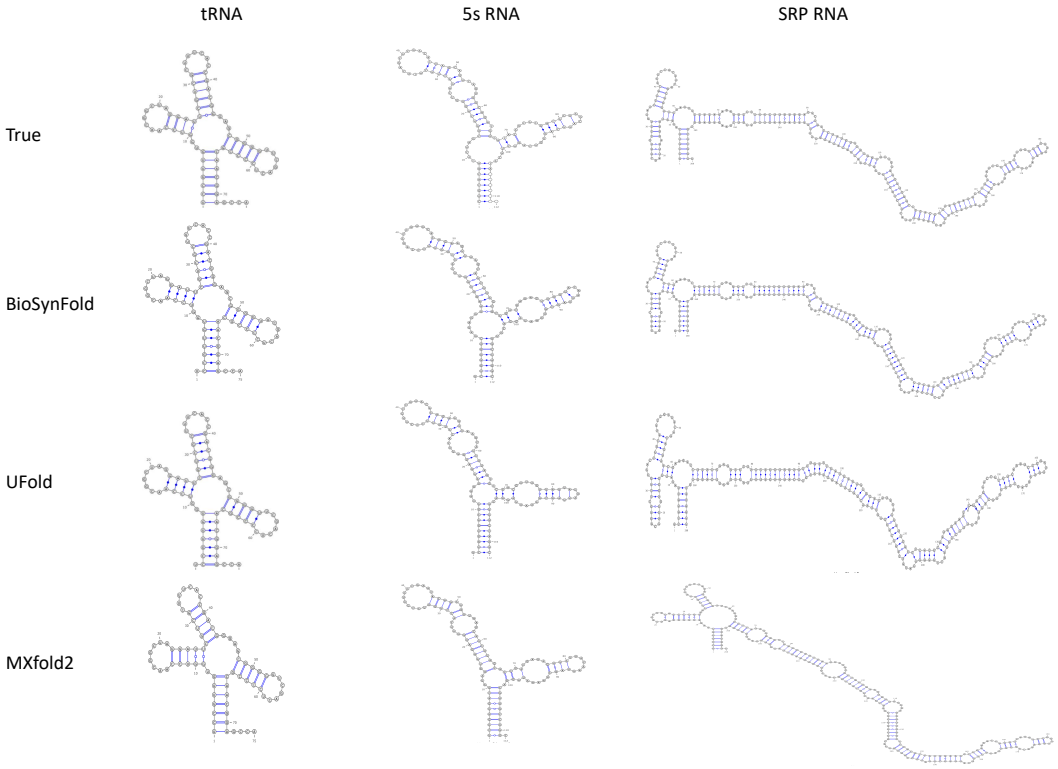


Figure 7: Visualization of the secondary structures of tRNA, 5S RNA, and SRP RNA using BioSynFold, UFold and MXfold2.