HARMONY: A MULTI-REPRESENTATION FRAMEWORK FOR RNA PROPERTY PREDICTION

Junjie Xu[1,2] *, **Artem Moskalev**[1] , **Tommaso Mansi**[1], **Mangal Prakash**[1] †, **Rui Liao**[1] †
[1]Johnson & Johnson Innovative Medicine, [2]The Pennsylvania State University
junjiexu@psu.edu
{amoskal2, tmansi, mpraka12, rliao2}@its.jnj.com

ABSTRACT

The biological functions of RNA arise from the interplay of sequence (1D), secondary structure (2D), and tertiary structure (3D). While existing machine learning models typically rely on sequence-based representations, recent studies suggest that integrating structural information can improve predictive performance, especially in low-data regimes. However, different representations have trade-offs—3D models are sensitive to noise, whereas sequence-based models are more robust to sequencing noise but lack structural insights. To address this, we introduce HARMONY, a framework that dynamically integrates 1D, 2D, and 3D representations, and seamlessly adapts to diverse real-world scenarios. Our experiments demonstrate that HARMONY consistently outperforms existing baselines across multiple RNA property prediction tasks on established benchmarks, offering a robust and generalizable approach to RNA modeling.

1 INTRODUCTION

RNA is a structurally and functionally versatile biomolecule, playing essential role in gene regulation, catalysis, and cellular processes (Sharp, 2009). Its properties emerge from the interplay of linear nucleotide sequences (1D), base-pair interactions forming secondary structures (2D), and intricate three-dimensional conformations (3D). Each of these representations captures different aspects of RNA function, and comprehensive modeling of RNA molecule requires combining the strengths of all three.

Existing machine learning approaches for RNA property prediction typically rely on sequence-based representations (Chen et al., 2022; Yang et al., 2023; Penić et al., 2024; Wang et al., 2022; Yazdani-Jahromi et al., 2025; Prakash et al., 2024). However, recent work has shown that incorporating structural information, such as 2D base-pairing graphs or 3D atomic geometry, can enhance predictive performance in certain scenarios (Xu et al., 2024; 2025). While 2D and 3D models outperform sequence-based approaches when high-quality structural data is available, these families of models are highly sensitive to sequencing errors and structural inaccuracies, whereas sequence-based models are more robust to such noise (Xu et al., 2024; 2025). With this, if we had prior knowledge of the data quality and labeling conditions, we could select the most suitable representation and a model accordingly. In practice, though, such information is often lacking, highlighting a need for adaptive approaches that can dynamically leverage different RNA representations rather than committing to just one upfront.

To address this limitation, we introduce HARMONY, a framework that dynamically integrates 1D, 2D, and 3D RNA representations to handle varied real-world conditions. Instead of relying on a single representation, HARMONY leverages RNA's natural organization to systematically combine atomic, nucleotide, and molecular-scale information. We conduct experiments on multiple RNA datasets and demonstrate that HARMONY consistently achieves strong performance across diverse RNA property prediction tasks, including those with sequencing noise, limited training data, and out-of-distribution conditions. By eliminating the need for manual selection of the optimal RNA representation, HARMONY offers a robust and generalizable approach to RNA modeling.

*This work was done while the author was an intern at Johnson & Johnson.
†Equal contribution as last authors

2 METHODOLOGY: HARMONY

We introduce HARMONY, a multi-representational model that integrates 1D, 2D and 3D RNA representations (Figure 1). Each RNA is represented as:

$$\mathcal{G} = \{\mathcal{G}_{3D}, \mathcal{G}_{2D}, \mathcal{G}_{1D}\} \tag{1}$$

where $\mathcal{G}_{3D} = \{X_{at}, R\}$ represents RNA 3D structure with atom features $X_{at} \in \mathbb{R}^{N_{at} \times F_{at}}$ and atom positions $R \in \mathbb{R}^{N_{at} \times 3}$ with N_{at} being the number of atoms. RNA 2D structure $\mathcal{G}_{2D} = \{X_{nt}, A\}$ consists of nucleotide features $X_{nt} \in \mathbb{R}^{N_{nt} \times F_{nt}}$ and an adjacency matrix $A \in \mathbb{R}^{N_{nt} \times N_{nt}}$. The 1D RNA sequence $\mathcal{G}_{1D} = \{X_{seq}\}$ encodes RNA's nucleotide sequence.

First, we apply a 3D encoder f_{3D} to extract atomic-level representations:

$$H_{3D} = f_{3D}(X_{at}, R). \tag{2}$$

A nucleotide pooling layer maps atom-wise features to nucleotide-level representations mean pooling representations of atoms corresponding to each nucleotide:

$$H'_{3D} = \text{Nuc.Pooling}(H_{3D}). \tag{3}$$

We concatenate the pooled 3D features H'_{3D} with nucleotide features X_{nt} and apply a 2D encoder f_{2D}:

$$H_{2D} = f_{2D}([X_{nt}, H'_{3D}], A). \tag{4}$$

Since sequence tokens X_{seq} align with nucleotide positions, we concatenate them with H_{2D} and apply a 1D encoder f_{1D}:

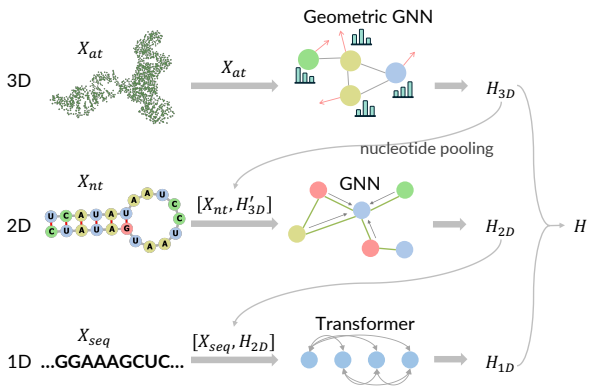


Figure 1: The framework of proposed HARMONY combines 1D, 2D and 3D RNA representations in one model.

$$H_{1D} = f_{1D}([X_{seq}, H_{2D}]). \tag{5}$$

Finally, the representations from all dimensions are combined using a differentiable function σ, in our case an MLP:

$$H = \sigma([H_{3D}, H_{2D}, H_{1D}]). \tag{6}$$

We use EGNN (Satorras et al., 2021) for f_{3D}, the spectral 2D GNN ChebNet (Defferrard et al., 2016) for f_{2D}, and Transformer (Vaswani et al., 2017) for f_{1D}. This selection is based on prior findings (Xu et al., 2024; 2025), which identified these models as the strongest 3D, 2D, and 1D baselines for RNA property prediction, respectively.

3 RNA PROPERTY PREDICTION TASKS

We first provide a brief overview of RNA property prediction tasks introduced in prior work (Xu et al., 2024; 2025). These tasks reflect real-world challenges in computational RNA modeling and assess how different representations impact performance:

1. **Task 1 (Effect of Structural Information)** examines the impact of incorporating structural information (1D, 2D, and 3D) on predictive performance.
2. **Task 2 (Limited Training Data)** Since acquiring large labeled RNA datasets is challenging, we analyze model performance in data-scarce settings by progressively reducing training samples.
3. **Task 3 (Partial Sequence Labeling)** Real-world datasets often contain labels for only part of the sequence. We test model generalization when only partial annotations are available.
4. **Task 4 (Robustness to Sequencing Noise)** assesses robustness to sequencing noise by introducing controlled mutations during training and testing, simulating realistic sequencing error rates while keeping property labels clean.

5. **Task 5 (Generalization to Out-of-Distribution (OOD) Data)** evaluates generalization to out-of-distribution data by training models on clean, high-quality sequences and testing them on noisy data, reflecting deployment in different sequencing conditions or batch effects.

By considering these established tasks, we systematically analyze HARMONY's ability to dynamically integrate 1D, 2D, and 3D representations without requiring prior knowledge of task scenarios.

4 EXPERIMENTS

In this paper, we adopt the COVID, Ribonanaza-2k and Tc-Ribo datasets introduced in (Xu et al., 2024; 2025) where for each dataset, the task is to regress a property of interest (nucleotide level degradation and stability for COVID and Ribonanza-2k respectively and switching factor for Tc-Ribo). We use the same baseline models and hyperparameters used in Xu et al. (2024; 2025). We evaluate all models on five random data splits (70:15:15 train:val:test) and report the mean performance with standard deviation. Following (Wayment-Steele et al., 2022; Xu et al., 2024; 2025), we use mean column-wise root mean squared error (MCRMSE) as the evaluation metric, defined as: $\text{MCRMSE}(f, D) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$ where f is the model, D is the dataset, and $\hat{y}_i$, y_i are the predicted and true values for data point i.

Table 1 presents the results for Task 1, which evaluates the impact of structural information on prediction performance. HARMONY outperforms all baselines relying on single-representation inputs, demonstrating that integrating 1D, 2D, and 3D representations leads to superior predictive performance.

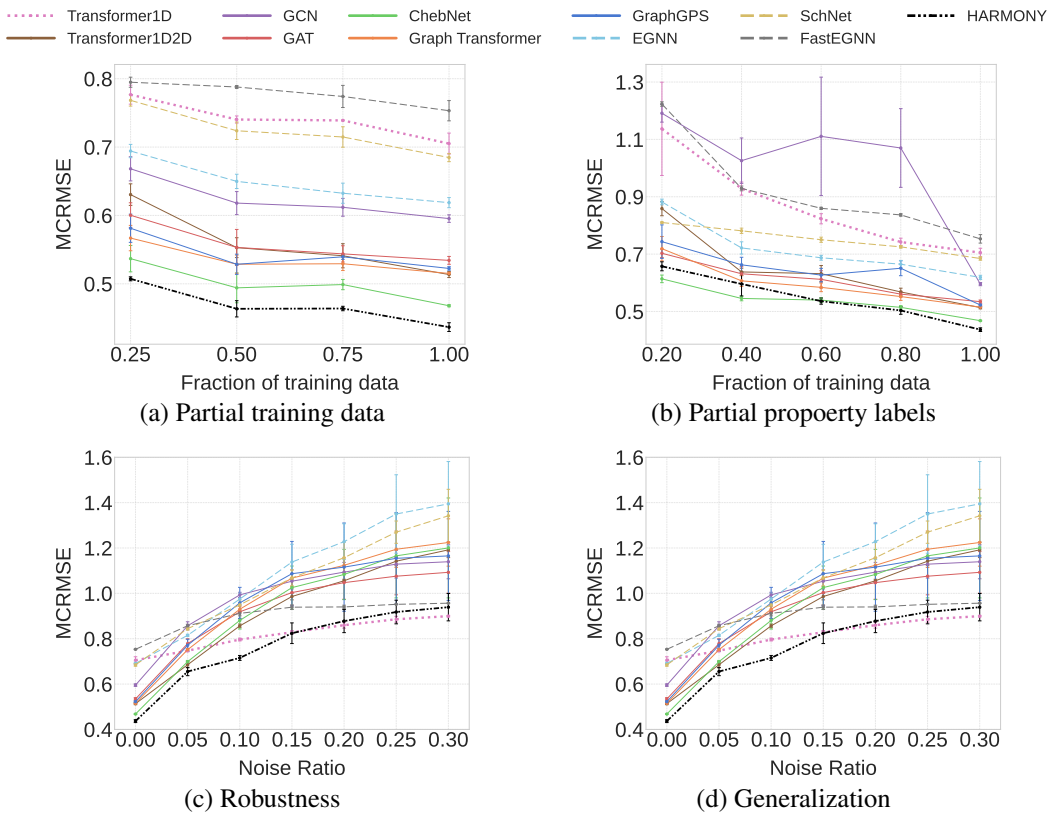
Table 1: **Comparison of 1D, 2D, and 3D models across datasets (MCRMSE ± std.).** The best results are bold. HARMONY consistently achieves state-of-the-art performance.

	Model	COVID	Ribonanza	Tc-Ribo
1D	Transformer1D	0.361±0.017	0.705±0.015	0.705±0.079
	RNA-FM	0.591±0.081	0.990±0.144	0.693±0.001
	SpliceBERT	0.588±0.077	1.022±0.144	0.708±0.003
2D	Transformer1D2D	0.305±0.012	0.514±0.004	0.633±0.001
	GCN	0.359±0.009	0.595±0.006	0.701±0.004
	GAT	0.315±0.006	0.534±0.006	0.685±0.024
	ChebNet	0.279±0.007	0.468±0.002	0.621±0.022
	Graph Transformer	0.318±0.008	0.515±0.001	0.710±0.041
	GraphGPS	0.332±0.013	0.523±0.003	0.715±0.012
3D	EGNN	0.480±0.025	0.808±0.023	0.725±0.002
	SchNet	0.499±0.003	0.843±0.004	0.696±0.008
	FAENet	0.486±0.010	0.834±0.003	0.703±0.011
	DimeNet	0.497±0.012	0.855±0.006	0.712±0.004
	GVP	0.467±0.010	0.797±0.012	0.744±0.004
	FastEGNN	0.477±0.005	0.816±0.014	0.753±0.001
Multi-dim	**HARMONY**	**0.271±0.003**	**0.437±0.006**	**0.554±0.004**

Figure 2 summarizes results for Tasks 2–5 on the Ribonanza-2k dataset, with additional datasets provided in Appendix Section A. In Task 2 (model efficiency under limited training data), HARMONY consistently outperforms all baselines, indicating its strong sample efficiency in data-scarce scenarios. In Task 3 (performance with partial sequence labeling), where models must generalize beyond available labels, HARMONY achieves results comparable to the best single-modality ChebNet model, suggesting its ability to leverage incomplete annotations effectively.

For Task 4 (robustness to sequencing noise), where models are evaluated under stable noise conditions, HARMONY achieves the best overall performance, confirming its resilience to sequencing errors. In Task 5 (generalization to out-of-distribution noise), which assesses adaptability to unseen noise variations, Transformer1D slightly outperforms HARMONY at high noise levels on Ribonanza, but HARMONY remains superior to all other baselines. Complete experimental results for other datasets are available in Appendix A.

In summary, the experimental results demonstrate that HARMONY consistently achieves the best overall performance across diverse real-world scenarios. By integrating information from 1D, 2D and 3D representations, HARMONY effectively combines the strengths of these representations. As a result, it not only excels in settings where high quality structural data is available (Task 1) but also either outperforms or is at par with the best baseline in noisy and data-scarce scenarios. Given the lack of prior knowledge about data quality and labeling conditions in real-world applications, HARMONY delivers a highly robust and reliable solution, maintaining strong performance across diverse conditions.



(a) Partial training data (b) Partial propoerty labels

(c) Robustness (d) Generalization

Figure 2: **Experiment results of 5 tasks on Ribonanza dataset.** Dotted, solid, and dashed lines denote 1D, 2D, and 3D methods, respectively, which apply consistently throughout all figures in this paper.

4.1 Ablation Study

To assess the contribution of 1D, 2D and 3D representations to the overall performance of HARMONY, we conduct an ablation study by systematically removing 1D, 2D, and 3D information. The results, presented in Table 2, demonstrate that the full HARMONY model, which integrates all three types of information, achieves the best performance. Removing any individual component results in a noticeable performance drop, highlighting the importance of incorporating information from sequence, secondary structure, and tertiary structure features for RNA property prediction.

Table 2: Ablation study on the impact of 1D, 2D, and 3D information. The full HARMONY model performs best, while removing any dimension degrades performance. MCRMSE is reported (lower is better) with best highlighted in bold.

	COVID	Ribonanza	Tc-Ribo
HARMONY	**0.271±0.003**	**0.437±0.006**	**0.554±0.004**
HARMONY w/o 1D	0.298±0.003	0.502±0.003	0.597±0.045
HARMONY w/o 2D	0.352±0.005	0.750±0.009	0.674±0.022
HARMONY w/o 3D	0.271±0.009	0.459±0.005	0.600±0.046

5 Conclusion

In this work, we propose HARMONY, a unified framework that integrates 1D, 2D, and 3D RNA representations to address the diverse challenges in RNA property prediction. By leveraging RNA's natural organization spanned by atoms and nucleotides, and combining the strengths of different representations, HARMONY achieves robust and superior performance across varying real-world experimental settings on existing benchmarks. This work highlights the importance of multi-representation integration in RNA modeling and provides a foundation for future advancements in computational RNA research. Our results demonstrate the potential of HARMONY to drive biological discovery and facilitate the development of RNA-based therapeutics.

REFERENCES

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.

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.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, pp. 3844–3852, 2016.

Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Chaitanya K. Joshi, Cristian Bodnar, Simon V. Mathis, Taco Cohen, and Pietro Liò. On the expressive power of geometric graph neural networks. In *International Conference on Machine Learning*, 2023.

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. *arXiv preprint arXiv:2403.00043*, 2024.

Mangal Prakash, Artem Moskalev, Peter DiMaggio Jr., Steven Combs, Tommaso Mansi, Justin Scheer, and Rui Liao. Bridging biomolecular modalities for knowledge transfer in bio-language models. In *Neurips 2024 Workshop Foundation Models for Science: Progress, Opportunities, and Challenges*, 2024. URL https://openreview.net/forum?id=dicOSQVPLm.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Phillip A Sharp. The centrality of rna. *Cell*, 136(4):577–580, 2009.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

Hong Wang, Shuyu Wang, Yong Zhang, Shoudong Bi, and Xiaolei Zhu. A brief review of machine learning methods for rna methylation sites prediction. *Methods*, 203:399–421, 2022.

Hannah K Wayment-Steele, Wipapat Kladwang, Andrew M Watkins, Do Soon Kim, Bojan Tunguz, Walter Reade, Maggie Demkin, Jonathan Romano, Roger Wellington-Oguri, John J Nicol, et al. Deep learning models for predicting rna degradation via dual crowdsourcing. *Nature Machine Intelligence*, 4(12):1174–1184, 2022.

Junjie Xu, Artem Moskalev, Tommaso Mansi, Mangal Prakash, and Rui Liao. Beyond sequence: Impact of geometric context for RNA property prediction. In *NeurIPS 2024 Workshop on AI for New Drug Modalities*, 2024. URL https://openreview.net/forum?id=otpO8k1fua.

Junjie Xu, Artem Moskalev, Tommaso Mansi, Mangal Prakash, and Rui Liao. Beyond sequence: Impact of geometric context for RNA property prediction. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=9htTvHkUhh.

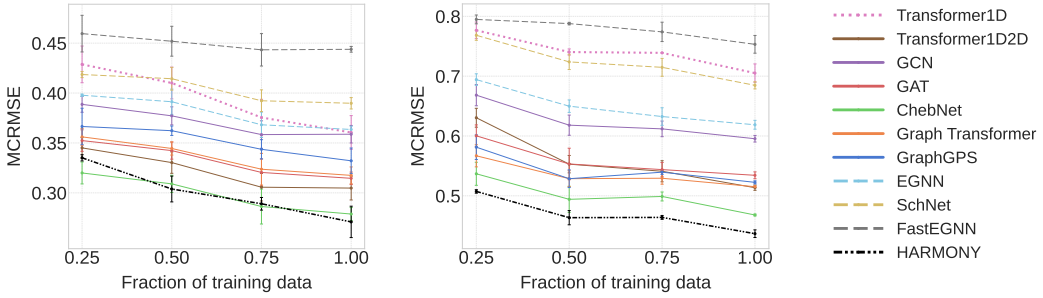
Yuning Yang, Gen Li, Kuan Pang, Wuxinhao Cao, Xiangtao Li, and Zhaolei Zhang. Deciphering 3'utr mediated gene regulation using interpretable deep representation learning. *bioRxiv*, pp. 2023–09, 2023.

Mehdi Yazdani-Jahromi, Mangal Prakash, Tommaso Mansi, Artem Moskalev, and Rui Liao. HELM: Hierarchical encoding for mRNA language modeling. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=MMHqnUOnl0.

A More Experiments

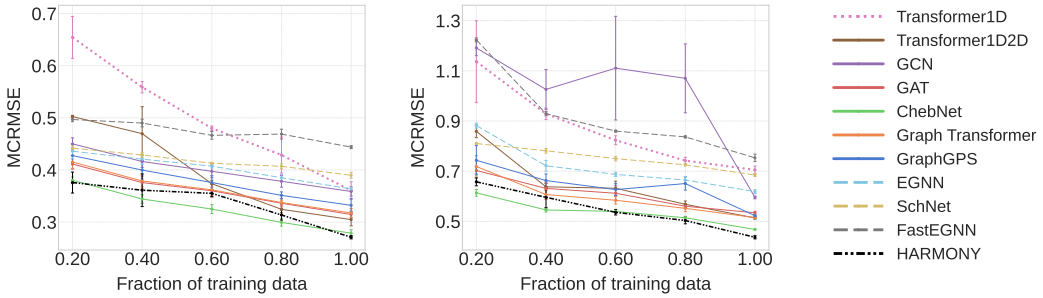
A.1 Results on other datasets

In this section, we provide the complete experiment results of Tasks 2 - 5 on each dataset. The results of Task 2 and Task 3 are presented in Fig.3 and Fig.4, respectively. In the limited data experiment (Task 2), HARMONY consistently outperforms all baselines, while in the partial sequence labeling task (Task 3), it achieves performance comparable to the best ChebNet model. The Tc-Ribo dataset is excluded from Task 2 since varying the training ratio from 0.25 to 1 requires a larger dataset. Task 3 is conducted only on nucleotide-level prediction datasets (COVID and Ribonanza), where the labeled nucleotide ratio per sequence is varied from 0.2 to 1.



(a) COVID (b) Ribonanza

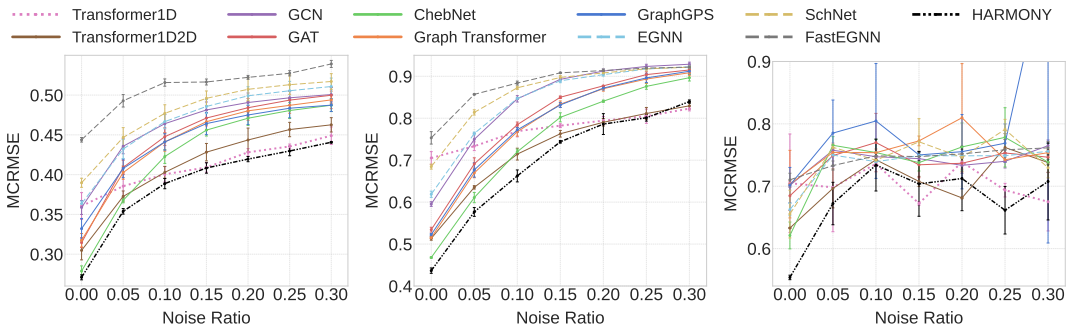
Figure 3: Performance vs. fraction of training data across various datasets.



(a) COVID (b) Ribonanza

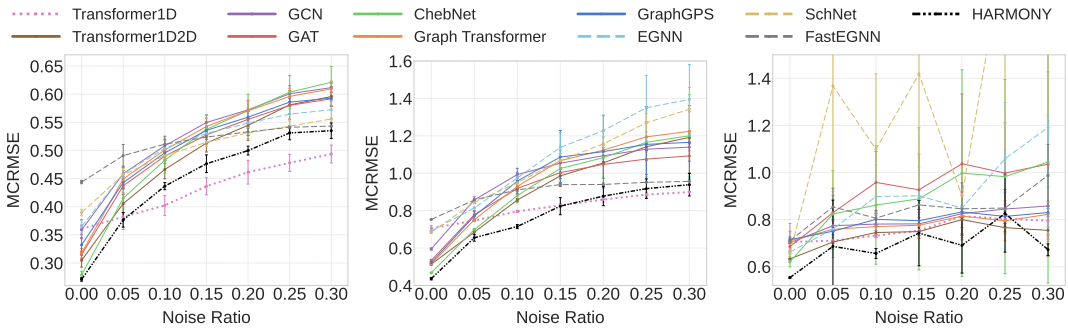
Figure 4: Performance vs. partial property labels on COVID and Ribonanza datasets.

The results of Task 4 (Robustness to Noise) and Task 5 (Generalization to Noisy Data) are presented in Fig.5 and Fig.6, respectively. In the robustness experiments, HARMONY achieves the best overall performance. For generalization, while Transformer1D slightly outperforms HARMONY at high noise levels on the COVID and Ribonanza datasets, HARMONY remains superior to all other baselines. HARMONY remains the top-performing model on Tc-Riboswithces dataset.



(a) COVID (b) Ribonanza (d) Tc-Riboswitches

Figure 5: Robustness experiments that train and test on noisy datasets.



(a) COVID (b) Ribonanza (d) Tc-Riboswitches

Figure 6: Generalization experiments train on clean datasets while test on noisy datasets.

A.2 Training details

All experimental protocols including dataset preprocessing, baselines and hyperparameter tuning are taken from Xu et al. (2024; 2025). All experiments were conducted on a single NVIDIA A100 GPU. For each baseline, hyperparameters were optimized using Optuna (Akiba et al., 2019), restricting the search to models with fewer than 10 million parameters that fit within the memory constraints of an 80GB NVIDIA A100 GPU. Most baseline implementations were sourced from PyTorch Geometric (Fey & Lenssen, 2019). The Transformer1D model was adapted to Transformer1D2D as detailed in the paper. For EGNN, we utilized the authors' implementation (Satorras et al., 2021), and for SchNet, the implementation from (Joshi et al., 2023) was used.

B Detailed Results

In this section, we present the additional results supporting Figures 3, 4, 5, and 6.

B.1 Impact of data availability

The detailed results of partial training data from Figure 3 are shown in Tables 3 and 4.

Table 3: Performance (MCRMSE) of different models across various training data (mean $\pm$ standard deviation) fractions on COVID dataset.

COVID	**0.25**	**0.50**	**0.75**	**1.00**
Transformer1D	0.429 ± 0.018	0.410 ± 0.016	0.375 ± 0.018	0.361 ± 0.017
Transformer1D2D	0.345 ± 0.010	0.330 ± 0.011	0.306 ± 0.014	0.305 ± 0.012
GCN	0.389 ± 0.008	0.377 ± 0.009	0.358 ± 0.014	0.359 ± 0.009
GAT	0.352 ± 0.011	0.342 ± 0.009	0.320 ± 0.014	0.315 ± 0.006
ChebNet	0.320 ± 0.011	0.309 ± 0.009	0.286 ± 0.018	0.279 ± 0.007
Graph Transformer	0.356 ± 0.008	0.344 ± 0.007	0.324 ± 0.016	0.318 ± 0.008
GraphGPS	0.367 ± 0.018	0.362 ± 0.005	0.344 ± 0.010	0.332 ± 0.013
EGNN	0.398 ± 0.001	0.391 ± 0.013	0.368 ± 0.013	0.364 ± 0.003
SchNet	0.419 ± 0.003	0.414 ± 0.011	0.392 ± 0.011	0.390 ± 0.006
FastEGNN	0.460 ± 0.018	0.452 ± 0.015	0.443 ± 0.016	0.444 ± 0.003
MultiDimGNN	0.335 ± 0.003	0.304 ± 0.013	0.289 ± 0.006	0.271 ± 0.016

B.2 Impact of partial labeling

The detailed results of the partial labeling sequence from Figure 4 are shown in Tables 5 and 6.

B.3 Robustness to sequencing noise

The results of the robustness experiment from Figure 5 are shown in Tables 7, 8, and 9.

Table 4: Performance (MCRMSE) of different models across various fractions of training data (mean ± standard deviation) on Ribonanza dataset.

Ribonanza	**0.25**	**0.50**	**0.75**	**1.00**
Transformer1D	0.777 ± 0.014	0.740 ± 0.005	0.739 ± 0.001	0.705 ± 0.015
Transformer1D2D	0.630 ± 0.016	0.553 ± 0.015	0.541 ± 0.018	0.514 ± 0.004
GCN	0.668 ± 0.018	0.618 ± 0.017	0.612 ± 0.013	0.595 ± 0.006
GAT	0.600 ± 0.018	0.553 ± 0.026	0.544 ± 0.012	0.534 ± 0.006
ChebNet	0.537 ± 0.019	0.494 ± 0.022	0.499 ± 0.007	0.468 ± 0.002
Graph Transformer	0.567 ± 0.019	0.529 ± 0.013	0.529 ± 0.010	0.515 ± 0.001
GraphGPS	0.581 ± 0.021	0.529 ± 0.015	0.540 ± 0.004	0.523 ± 0.003
EGNN	0.694 ± 0.010	0.650 ± 0.010	0.632 ± 0.015	0.619 ± 0.007
SchNet	0.768 ± 0.008	0.724 ± 0.013	0.715 ± 0.015	0.685 ± 0.006
FastEGNN	0.795 ± 0.007	0.788 ± 0.002	0.774 ± 0.016	0.753 ± 0.015
MultiDimGNN	0.507 ± 0.003	0.463 ± 0.012	0.464 ± 0.003	0.437 ± 0.006

Table 5: Performance (MCRMSE) of different models across various fractions of sequence labeling (mean ± standard deviation) on COVID dataset.

COVID	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
Transformer1D	0.654 ± 0.040	0.559 ± 0.011	0.480 ± 0.004	0.429 ± 0.034	0.361 ± 0.017
Transformer1D2D	0.502 ± 0.002	0.470 ± 0.052	0.374 ± 0.007	0.325 ± 0.006	0.305 ± 0.012
GCN	0.450 ± 0.012	0.416 ± 0.012	0.397 ± 0.012	0.378 ± 0.011	0.359 ± 0.009
GAT	0.411 ± 0.010	0.376 ± 0.012	0.360 ± 0.012	0.336 ± 0.009	0.315 ± 0.006
ChebNet	0.380 ± 0.007	0.344 ± 0.008	0.325 ± 0.009	0.299 ± 0.007	0.279 ± 0.007
Graph Transformer	0.415 ± 0.012	0.379 ± 0.011	0.362 ± 0.011	0.338 ± 0.004	0.318 ± 0.008
GraphGPS	0.428 ± 0.015	0.400 ± 0.017	0.376 ± 0.013	0.351 ± 0.007	0.332 ± 0.013
EGNN	0.436 ± 0.014	0.421 ± 0.010	0.407 ± 0.004	0.385 ± 0.006	0.364 ± 0.003
SchNet	0.442 ± 0.004	0.429 ± 0.005	0.413 ± 0.001	0.407 ± 0.005	0.390 ± 0.006
FastEGNN	0.497 ± 0.004	0.490 ± 0.007	0.466 ± 0.007	0.469 ± 0.009	0.444 ± 0.003
MultiDimGNN	0.376 ± 0.020	0.361 ± 0.031	0.355 ± 0.006	0.314 ± 0.010	0.271 ± 0.003

B.4 Generalization to OOD data

The results of the generalization experiment from Figure 6 are shown in Tables 10, 11, and 12.

Table 6: Performance (MCRMSE) of different models across various fractions of sequence labeling (mean ± standard deviation) on Ribonanza dataset.

Ribonanza	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
Transformer1D	1.137 ± 0.163	0.929 ± 0.023	0.823 ± 0.018	0.742 ± 0.013	0.705 ± 0.015
Transformer1D2D	0.859 ± 0.025	0.638 ± 0.013	0.632 ± 0.028	0.568 ± 0.013	0.514 ± 0.004
GCN	1.191 ± 0.031	1.026 ± 0.079	1.111 ± 0.206	1.070 ± 0.137	0.595 ± 0.006
GAT	0.703 ± 0.015	0.632 ± 0.025	0.612 ± 0.030	0.560 ± 0.010	0.534 ± 0.006
ChebNet	0.614 ± 0.013	0.546 ± 0.008	0.540 ± 0.008	0.514 ± 0.006	0.468 ± 0.002
Graph Transformer	0.719 ± 0.043	0.607 ± 0.020	0.584 ± 0.015	0.552 ± 0.013	0.515 ± 0.001
GraphGPS	0.743 ± 0.058	0.663 ± 0.026	0.627 ± 0.024	0.651 ± 0.026	0.523 ± 0.003
EGNN	0.882 ± 0.010	0.722 ± 0.021	0.687 ± 0.008	0.665 ± 0.013	0.619 ± 0.007
SchNet	0.810 ± 0.002	0.781 ± 0.009	0.750 ± 0.009	0.725 ± 0.004	0.685 ± 0.006
FastEGNN	1.223 ± 0.008	0.929 ± 0.008	0.860 ± 0.004	0.837 ± 0.004	0.753 ±0.015
MultiDimGNN	0.658 ± 0.016	0.596 ± 0.042	0.536 ± 0.011	0.504 ± 0.013	0.437 ± 0.006

Table 7: Performance (MCRMSE) of various models in **robustness** experiments on the **COVID** dataset (mean ± standard deviation).

Model	**0.00**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**
Transformer1D	0.361 ± 0.017	0.386 ± 0.015	0.400 ± 0.010	0.409 ± 0.006	0.428 ± 0.005	0.435 ± 0.003	0.449 ± 0.011
Transformer1D2D	0.305 ± 0.012	0.373 ± 0.007	0.403 ± 0.007	0.428 ± 0.011	0.444 ± 0.015	0.457 ± 0.009	0.463 ± 0.009
GCN	0.359 ± 0.009	0.436 ± 0.009	0.464 ± 0.011	0.481 ± 0.010	0.491 ± 0.012	0.497 ± 0.009	0.501 ± 0.009
GAT	0.315 ± 0.006	0.409 ± 0.009	0.448 ± 0.011	0.471 ± 0.010	0.484 ± 0.012	0.494 ± 0.011	0.500 ± 0.010
ChebNet	0.279 ± 0.007	0.368 ± 0.003	0.423 ± 0.009	0.456 ± 0.007	0.471 ± 0.009	0.481 ± 0.010	0.487 ± 0.008
Graph Transformer	0.318 ± 0.008	0.403 ± 0.008	0.441 ± 0.012	0.467 ± 0.011	0.480 ± 0.012	0.487 ± 0.010	0.494 ± 0.011
GraphGPS	0.332 ± 0.013	0.408 ± 0.012	0.441 ± 0.010	0.464 ± 0.014	0.475 ± 0.012	0.484 ± 0.012	0.487 ± 0.008
EGNN	0.364 ± 0.003	0.432 ± 0.012	0.467 ± 0.009	0.486 ± 0.009	0.499 ± 0.011	0.505 ± 0.012	0.511 ± 0.011
SchNet	0.390 ± 0.006	0.447 ± 0.012	0.477 ± 0.011	0.496 ± 0.009	0.507 ± 0.014	0.513 ± 0.012	0.517 ± 0.010
FastEGNN	0.444 ± 0.003	0.49283 ± 0.008	0.516 ± 0.005	0.516 ± 0.004	0.522 ± 0.002	0.527 ± 0.003	0.540 ± 0.004
MultiDimGNN	0.271 ± 0.003	0.354 ± 0.004	0.389 ± 0.006	0.408 ± 0.007	0.420 ± 0.003	0.430 ± 0.006	0.441 ± 0.001

Table 8: Performance (MCRMSE) of various models in **robustness** experiments on the **Ribonanza** dataset (mean ± standard deviation).

Model	**0.00**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**
Transformer1D	0.705 ± 0.015	0.733 ± 0.010	0.769 ± 0.014	0.782 ± 0.005	0.794 ± 0.010	0.805 ± 0.017	0.823 ± 0.005
Transformer1D2D	0.514 ± 0.004	0.635 ± 0.004	0.714 ± 0.014	0.763 ± 0.008	0.790 ± 0.009	0.811 ± 0.014	0.830 ± 0.008
GCN	0.595 ± 0.006	0.750 ± 0.014	0.846 ± 0.008	0.893 ± 0.003	0.912 ± 0.005	0.924 ± 0.005	0.929 ± 0.005
GAT	0.534 ± 0.006	0.691 ± 0.015	0.785 ± 0.006	0.850 ± 0.003	0.877 ± 0.001	0.904 ± 0.007	0.915 ± 0.007
ChebNet	0.468 ± 0.002	0.611 ± 0.012	0.720 ± 0.006	0.802 ± 0.011	0.841 ± 0.003	0.876 ± 0.007	0.897 ± 0.008
Graph Transformer	0.515 ± 0.001	0.670 ± 0.011	0.768 ± 0.011	0.833 ± 0.008	0.870 ± 0.006	0.893 ± 0.010	0.908 ± 0.007
GraphGPS	0.523 ± 0.003	0.677 ± 0.017	0.772 ± 0.006	0.832 ± 0.006	0.872 ± 0.004	0.896 ± 0.011	0.912 ± 0.006
EGNN	0.619 ± 0.007	0.764 ± 0.003	0.847 ± 0.003	0.889 ± 0.005	0.904 ± 0.003	0.917 ± 0.000	0.922 ± 0.002
SchNet	0.685 ± 0.006	0.814 ± 0.006	0.873 ± 0.004	0.897 ± 0.004	0.908 ± 0.004	0.918 ± 0.005	0.922 ± 0.005
FastEGNN	0.753 ± 0.015	0.857 ± 0.001	0.884 ± 0.005	0.908 ± 0.001	0.914 ± 0.004	0.920 ± 0.003	0.922 ± 0.003
MultiDimGNN	0.437 ± 0.006	0.577 ± 0.011	0.662 ± 0.014	0.744 ± 0.003	0.786 ± 0.025	0.801 ± 0.008	0.840 ± 0.004

Table 9: Performance (MCRMSE) of various models in **robustness** experiments on the **Tc-riboswitches** dataset (mean ± standard deviation).

Model	**0.00**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**
Transformer1D	0.705 ± 0.079	0.698 ± 0.071	0.736 ± 0.004	0.672 ± 0.003	0.739 ± 0.008	0.694 ± 0.011	0.675 ± 0.047
Transformer1D2D	0.633 ± 0.001	0.697 ± 0.031	0.742 ± 0.003	0.708 ± 0.008	0.681 ± 0.001	0.762 ± 0.022	0.738 ± 0.016
GCN	0.701 ± 0.004	0.758 ± 0.003	0.747 ± 0.005	0.744 ± 0.004	0.733 ± 0.013	0.740 ± 0.011	0.765 ± 0.009
GAT	0.685 ± 0.024	0.749 ± 0.017	0.770 ± 0.047	0.734 ± 0.021	0.737 ± 0.009	0.753 ± 0.001	0.747 ± 0.011
ChebNet	0.621 ± 0.022	0.766 ± 0.014	0.754 ± 0.021	0.738 ± 0.014	0.763 ± 0.039	0.778 ± 0.048	0.739 ± 0.004
Graph Transformer	0.703 ± 0.054	0.754 ± 0.005	0.754 ± 0.006	0.773 ± 0.008	0.810 ± 0.087	0.742 ± 0.004	0.754 ± 0.005
GraphGPS	0.702 ± 0.028	0.785 ± 0.053	0.805 ± 0.092	0.750 ± 0.006	0.755 ± 0.060	0.769 ± 0.031	1.078 ± 0.469
EGNN	0.663 ± 0.010	0.750 ± 0.001	0.739 ± 0.002	0.749 ± 0.005	0.749 ± 0.001	0.749 ± 0.001	0.756 ± 0.013
SchNet	0.655 ± 0.038	0.762 ± 0.005	0.742 ± 0.002	0.771 ± 0.037	0.746 ± 0.005	0.791 ± 0.016	0.730 ± 0.016
FastEGNN	0.710 ± 0.010	0.733 ± 0.007	0.749 ± 0.006	0.748 ± 0.006	0.752 ± 0.008	0.758 ± 0.017	0.761 ± 0.010
MultiDimGNN	0.554 ± 0.004	0.672 ± 0.034	0.734 ± 0.042	0.704 ± 0.052	0.712 ± 0.052	0.661 ± 0.038	0.708 ± 0.062

Table 10: Performance (MCRMSE) of various models in **generalization** experiments on the **COVID** dataset (mean ± standard deviation).

Model	0.00	0.05	0.10	0.15	0.20	0.25	0.30
Transformer1D	0.361 ± 0.017	0.382 ± 0.022	0.402 ± 0.018	0.436 ± 0.015	0.461 ± 0.021	0.478 ± 0.015	0.494 ± 0.016
Transformer1D2D	0.305 ± 0.012	0.406 ± 0.016	0.466 ± 0.017	0.513 ± 0.016	0.545 ± 0.027	0.581 ± 0.025	0.596 ± 0.018
GCN	0.359 ± 0.009	0.459 ± 0.011	0.508 ± 0.011	0.550 ± 0.014	0.572 ± 0.016	0.601 ± 0.014	0.612 ± 0.008
GAT	0.315 ± 0.006	0.437 ± 0.013	0.490 ± 0.013	0.528 ± 0.008	0.555 ± 0.013	0.580 ± 0.015	0.592 ± 0.012
ChebNet	0.279 ± 0.007	0.415 ± 0.017	0.483 ± 0.023	0.538 ± 0.025	0.571 ± 0.029	0.604 ± 0.030	0.621 ± 0.028
Graph Transformer	0.318 ± 0.008	0.449 ± 0.015	0.501 ± 0.018	0.543 ± 0.015	0.571 ± 0.019	0.596 ± 0.014	0.609 ± 0.014
GraphGPS	0.332 ± 0.013	0.443 ± 0.011	0.496 ± 0.006	0.536 ± 0.005	0.559 ± 0.010	0.586 ± 0.007	0.593 ± 0.005
EGNN	0.365 ± 0.011	0.458 ± 0.014	0.504 ± 0.018	0.530 ± 0.020	0.549 ± 0.021	0.565 ± 0.022	0.572 ± 0.022
SchNet	0.390 ± 0.006	0.457 ± 0.011	0.491 ± 0.008	0.515 ± 0.007	0.531 ± 0.010	0.543 ± 0.009	0.556 ± 0.002
FastEGNN	0.444 ± 0.003	0.491 ± 0.020	0.511 ± 0.014	0.524 ± 0.009	0.533 ± 0.006	0.541 ± 0.003	0.543 ± 0.001
MultiDimGNN	0.271 ± 0.003	0.377 ± 0.013	0.437 ± 0.006	0.476 ± 0.016	0.500 ± 0.008	0.531 ± 0.012	0.535 ± 0.014

Table 11: Performance (MCRMSE) of various models in **generalization** experiments on the **Ribonanza** dataset (mean ± standard deviation).

Model	0.00	0.05	0.10	0.15	0.20	0.25	0.30
Transformer1D	0.705 ± 0.015	0.747 ± 0.005	0.796 ± 0.006	0.828 ± 0.008	0.860 ± 0.013	0.886 ± 0.013	0.899 ± 0.003
Transformer1D2D	0.514 ± 0.004	0.685 ± 0.014	0.857 ± 0.008	0.986 ± 0.015	1.055 ± 0.007	1.142 ± 0.020	1.192 ± 0.034
GCN	0.595 ± 0.006	0.857 ± 0.018	0.993 ± 0.012	1.054 ± 0.034	1.094 ± 0.043	1.129 ± 0.061	1.139 ± 0.075
GAT	0.534 ± 0.006	0.778 ± 0.021	0.919 ± 0.030	1.003 ± 0.056	1.047 ± 0.073	1.076 ± 0.082	1.093 ± 0.091
ChebNet	0.468 ± 0.002	0.699 ± 0.005	0.881 ± 0.038	1.025 ± 0.095	1.083 ± 0.111	1.165 ± 0.185	1.200 ± 0.220
Graph Transformer	0.515 ± 0.001	0.752 ± 0.005	0.930 ± 0.013	1.067 ± 0.036	1.124 ± 0.033	1.194 ± 0.080	1.224 ± 0.104
GraphGPS	0.523 ± 0.003	0.771 ± 0.026	0.958 ± 0.068	1.087 ± 0.142	1.116 ± 0.195	1.154 ± 0.202	1.165 ± 0.196
EGNN	0.691 ± 0.006	0.815 ± 0.004	0.975 ± 0.026	1.138 ± 0.078	1.228 ± 0.079	1.350 ± 0.173	1.395 ± 0.187
SchNet	0.685 ± 0.006	0.844 ± 0.006	0.949 ± 0.022	1.068 ± 0.035	1.157 ± 0.069	1.270 ± 0.049	1.342 ± 0.117
FastEGNN	0.753 ± 0.015	0.857 ± 0.001	0.912 ± 0.007	0.939 ± 0.011	0.940 ± 0.010	0.952 ± 0.008	0.957 ± 0.001
MultiDimGNN	0.437 ± 0.006	0.655 ± 0.018	0.715 ± 0.011	0.824 ± 0.046	0.878 ± 0.051	0.918 ± 0.052	0.939 ± 0.060

Table 12: Performance (MCRMSE) of various models in **generalization** experiments on the **Tc-riboswitches** dataset (mean ± standard deviation).

Model	0.00	0.05	0.10	0.15	0.20	0.25	0.30
Transformer1D	0.705 ± 0.079	0.711 ± 0.038	0.732 ± 0.007	0.753 ± 0.019	0.815 ± 0.091	0.803 ± 0.062	0.796 ± 0.079
Transformer1D2D	0.633 ± 0.001	0.705 ± 0.007	0.745 ± 0.008	0.749 ± 0.017	0.800 ± 0.034	0.766 ± 0.017	0.754 ± 0.014
GCN	0.701 ± 0.004	0.774 ± 0.026	0.781 ± 0.051	0.782 ± 0.062	0.825 ± 0.093	0.845 ± 0.072	0.858 ± 0.126
GAT	0.685 ± 0.024	0.829 ± 0.074	0.958 ± 0.131	0.926 ± 0.152	1.037 ± 0.297	0.998 ± 0.213	1.036 ± 0.392
ChebNet	0.621 ± 0.022	0.824 ± 0.185	0.862 ± 0.251	0.888 ± 0.302	0.997 ± 0.439	0.983 ± 0.412	1.045 ± 0.514
Graph Transformer	0.710 ± 0.041	0.759 ± 0.035	0.770 ± 0.047	0.776 ± 0.054	0.815 ± 0.101	0.795 ± 0.085	0.822 ± 0.115
GraphGPS	0.715 ± 0.012	0.751 ± 0.016	0.801 ± 0.023	0.795 ± 0.014	0.833 ± 0.025	0.814 ± 0.022	0.830 ± 0.018
EGNN	0.663 ± 0.010	0.760 ± 0.043	0.898 ± 0.134	0.901 ± 0.123	0.849 ± 0.053	1.058 ± 0.239	1.192 ± 0.453
SchNet	0.655 ± 0.038	1.368 ± 0.449	1.099 ± 0.320	1.419 ± 0.491	0.900 ± 0.119	1.957 ± 0.858	2.025 ± 0.779
FastEGNN	0.710 ± 0.011	0.854 ± 0.079	0.806 ± 0.032	0.863 ± 0.020	0.846 ± 0.110	0.849 ± 0.079	0.988 ± 0.131
MultiDimGNN	0.554 ± 0.004	0.686 ± 0.197	0.656 ± 0.021	0.743 ± 0.139	0.691 ± 0.118	0.826 ± 0.164	0.672 ± 0.024