lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures

Sumit Tarafder[1] and Debswapna Bhattacharya[1*]

[1]Department of Computer Science, Virginia Tech, Blacksburg, Virginia, 24061, USA

Abstract

A scoring function that can reliably assess the accuracy of a 3D RNA structural model in the absence of experimental structure is not only important for model evaluation and selection but also useful for scoring-guided conformational sampling. However, high-fidelity RNA scoring has proven to be difficult using conventional knowledge-based statistical potentials and currently-available machine learning-based approaches. Here we present lociPARSE, a locality-aware invariant point attention architecture for scoring RNA 3D structures. Unlike existing machine learning methods that estimate superposition-based root mean square deviation (RMSD), lociPARSE estimates Local Distance Difference Test (lDDT) scores capturing the accuracy of each nucleotide and its surrounding local atomic environment in a superposition-free manner, before aggregating information to predict global structural accuracy. Tested on multiple datasets including CASP15, lociPARSE significantly outperforms existing statistical potentials (rsRNASP, cgRNASP, DFIRE-RNA, and RASP) and machine learning methods (ARES and RNA3DCNN) across complementary assessment metrics. lociPARSE is freely available at `https://github.com/Bhattacharya-Lab/lociPARSE`.

Keywords: RNA scoring function, RNA structure modeling, attention neural networks, deep learning

1 Introduction

Computational prediction of RNA 3-dimensional structures from nucleotide sequence has garnered considerable research effort over the past decade [1–8] and deep learning-enabled RNA 3D modeling has gained significant attention in the recent past [9–13]. To facilitate practical applicability of predicted 3D models, it is critical to have a scoring function that can reliably assess their global topology and local quality in the absence of experimental structures [14, 15]. Moreover, the ability of a scoring function to distinguish accurate 3D models of previously unseen RNAs from misfolded alternatives plays an important role in guiding conformation sampling towards the native state [16].

Existing methods for scoring RNA structures roughly belong to two categories: knowledge-based statistical potentials and supervised machine learning. Various knowledge-based statistical potentials have been developed, both at all-atom and coarse-grained levels [17–21], using different simulated reference states [22–27]. However, reliably distinguishing accurate structural models of RNA from less accurate ones has proven to be difficult, because the characteristics of energetically favorable RNA structures are not sufficiently well understood and thus the reference states may deviate largely from the ideal one. Machine learning-based methods [28, 29] aim to overcome such limitation by learning to predict the accuracy of an RNA structural model through supervised learning. Indeed, machine learning-based RNA scoring functions, trained to estimate the unfitness score either at the nucleotide level or at the structural level by learning to predict the root mean square deviation (RMSD) from the unknown true structure, have been shown to be effective in RNA-Puzzles blind structure prediction challenges [8].

*Corresponding Author. Email: dbhattacharya@vt.edu

Despite the effectiveness, the existing machine learning methods do not consider some key factors that can significantly improve the sensitivity of RNA scoring functions. First, global superposition-dependent RMSD metric is not length normalized, affected by superposition, dominated by outliers in poorly modeled structural regions, and does not take into account the accuracy of local atomic environment. RNA is a flexible molecule in which irregular loops may affect RMSD measures and global superposition may not be optimal, leading to scoring anomalies. Yet, virtually all existing machine learning-based RNA scoring functions use RMSD as the ground truth during supervised training. Second, similar to other macromolecules, RNA structures have no natural canonical orientation. As such, machine learning methods that are not invariant to global Euclidean transformations such as rotation must account for this aspect of variation by tweaking model architecture and/or parameters, which may affect their expressiveness and generalizability. Third, in consideration of RNA as a flexible molecule in which interplay between various local structural motifs define the global topology, an effective scoring function should not be strongly influenced by the relative motions between the tertiary motifs. That is, the effects of relative movement between the motifs should not lead to artificially unfavorable scores.

Using the Local Distance Difference Test (lDDT) [30] as the ground truth during supervised training is an attractive alternative to the popular RMSD metric. lDDT compares distances between atoms that are nearby (within 15 Å) in the experimental structure to the distances between those atoms in the predicted structure and offers several advantages over RMSD. First, being superposition-free and based on rotation-invariant properties of a structure, lDDT naturally preserves invariance with respect to the global Euclidean transformations of the input RNA structure such as global rotations and translations. Second, lDDT measures the accuracy of local environment of the model in atomic detail, without being affected by superposition or dominated by outliers in poorly modeled structural regions. Third, lDDT exhibits robustness to movements between tertiary structural units such as domains in proteins that can generalize to RNA tertiary motifs, provided a way can be found that ensures rigid motion between a set of local structural units is invariant under global Euclidean transformations on the said units. A solution to this problem comes from Invariant Point Attention (IPA) proposed in AlphaFold2 as part of the structural module [31]. IPA is a form of attention that acts on a set of 3D point clouds and is invariant under global Euclidean transformations on said points, where 3D point clouds are represented using local frames.

How can we capture the aforementioned benefits of lDDT in a neural network architecture for RNA scoring, while maintaining invariance under global Euclidean transformations? Here, we provide such a solution by developing a new attention-based architecture, called lociPARSE (locality-aware invariant Point Attention-based RNA ScorEr), for scoring RNA 3D structures. Different from previous supervised learning approaches that estimate the RMSD metric, our method estimates local nucleotide-wise lDDT scores that are then aggregated over all nucleotides to predict global structural accuracy. Inspired by AlphaFold2, we define nucleotide-wise frames parameterized by rotation matrices and translation vectors operating on predefined RNA conformation at the local level. To model the local atomic environment of each nucleotide as captured by lDDT, the IPA implementation used in the original AlphaFold2 has been modified to incorporate locality information derived from the RNA atomic coordinates. By so doing, we are able to effectively capture the accuracy of each nucleotide while considering the effect of its local atomic environment.

Our method significantly outperforms traditional knowledge-based statistical potentials as well as state-of-the-art machine learning-based RNA scoring functions such as ARES [29] on multiple independent test datasets including CASP15 blind test targets across a wide-range of performance measures. In particular, lociPARSE exhibits superior ability to reproduce the ground truth lDDT scores both at the global and local levels, rank predictions for a given target with high fidelity, recognize the best predictions consistently, and better discriminate between 'good' and 'bad' predictions. An open-source software implementation of lociPARSE, licensed under the GNU General Public License v3, is freely available at `https://github.com/Bhattacharya-Lab/lociPARSE`.

2 Results

2.1 lociPARSE: locality-aware invariant point attention for RNA scoring

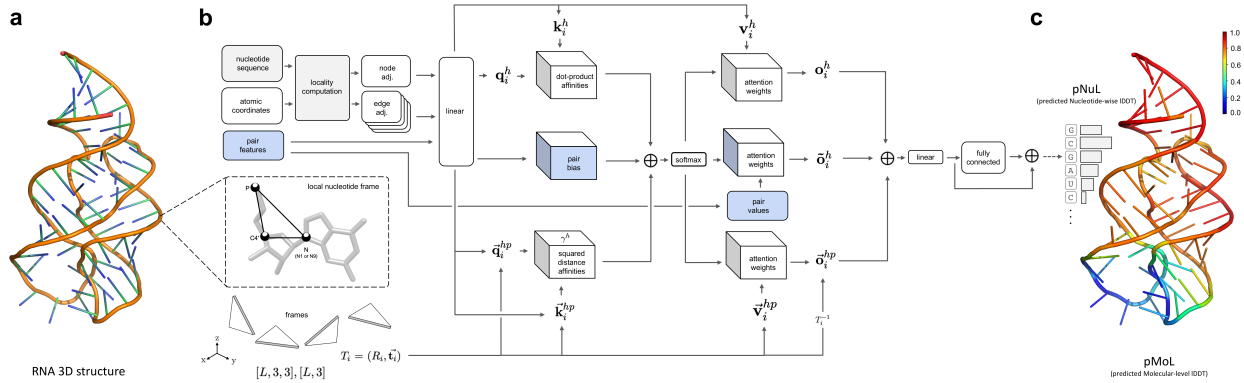

Figure 1: Overview of lociPARSE. Given a 3D RNA structure, we estimate the local nucleotide-wise lDDT scores that are then aggregated over all nucleotides to predict global structural accuracy. (a) An input RNA 3D structure. (b) The architecture of our locality-aware invariant point attention module to capture the accuracy of each nucleotide and the effect of its local atomic environment. (c) By aggregating information at the level of nucleotide, we output predicted nucleotide-wise lDDT (pNuL) scores before making a prediction at the level of the entire RNA structure to output predicted molecular-level lDDT (pMoL).

An overview of our method, lociPARSE, is illustrated in Figure 1. The core component of our architecture, outlined in Figure 1b, is an invariant point attention (IPA) module which utilizes the geometry of the input RNA 3D structure to revise the nucleotide and pair features. This component is similar to the AlphaFold2's IPA formulation used in the structure module, but modified herein to incorporate locality information derived from the RNA atomic coordinates. To do this, we introduce locality-aware geometry and edge-biased attention (see Section 4.2.2) based on nucleotide pair adjacencies to capture the local atomic environment of each nucleotide considering the Euclidean distances of the C4′ - C4′ atoms between nucleotide pairs. In our setting, we define local nucleotide frames (see Section 4.2.1) from the Cartesian coordinates of C4′, P, and glycosidic N atoms. The IPA partitions the nucleotide query and value features into 3D vectors and transforms them from the target nucleotide's local frame into a global reference frame before computing both attention weights and the output of the attention mechanism. Further, we augment nucleotide-nucleotide atomic distances between all pair of 3 atoms P, C4′ and N, encoded with Gaussian radial basis functions as pair features (Section 4.1), and make further use of the pair features to bias attention weights and update scalar features. Our network architecture consists of 4 IPA layers, with the IPA hyperparameters $(N_{heads}, c, N_{\text{query points}}, N_{\text{value points}})$ set to $(4, 128, 8, 4)$ and we use 20 nearest neighbors for the locality computation, determined through ablation experiments using an independent validation set (Section 2.6). The output of the attention layer is invariant to the global Euclidean transformations such as global rotations and translations of the input RNA. Finally, a linear layer followed by a 2 layer fully-connected network are used to estimate the predicted nucleotide-wise lDDT scores (pNuL) before making a prediction at the level of the entire molecule by aggregating nucleotide-level information, leading to predicted molecular-level lDDT (pMoL), thus enabling our method to estimate both local and global quality of the input RNA 3D structure.

2.2 Experimental setup

For training and performance evaluation, we use existing and publicly available benchmark datasets. Our training dataset contains 1,399 RNA targets collected from the training set used in the recent RNA 3D structure prediction method trRosettaRNA [10]. We extracted the sequences from the 1,399 experimental

structures and generated a total of about 52,000 structural models using a combination of different RNA 3D structure prediction tools including recent deep learning-enabled RNA structure prediction methods [9–13], physics-based RNA folding [32], and experimental structure perturbation using PyRosetta [33] (see Section 4.3). Details of our training procedure are provided in Section 4.4. Our test data includes 30 independent RNAs, also collected from trRosettaRNA following the train and test splits of the original work. We generated 3D structural models for each of these 30 RNAs using the deep learning-enabled RNA structure prediction methods [9–13]. We also use 12 RNA targets from CASP15 as an additional independent test set containing targets cleared for public access as of December 20, 2022, where the corresponding 3D structural models are collected directly from the CASP15 website `https://predictioncenter.org/casp15/` based on the blind predictions submitted by various participating groups in CASP15 RNA 3D structure prediction challenge. We compare our method lociPARSE with traditional knowledge-based statistical potentials including rsRNASP [17], cgRNASP [18], RASP [19] and DFIRE-RNA [20] as well as state-of-the-art machine learning-based RNA scoring functions RNA3DCNN [28] and ARES [29]. To assess the accuracy of different aspects of quality estimation, we use a wide-range of performance measures including the ability to reproduce the ground truth lDDT scores both at the global and local levels, rank predictions for a given target, recognize the best predictions, and discriminate between 'good' and 'bad' predictions. Details of competing methods and evaluation metrics can be found in Section 4.5.

2.3 Performance on 30 independent RNA targets

Table 1 reports the performance of our new method lociPARSE and the other competing methods on 30 independent RNA targets. lociPARSE consistently outperforms all other tested methods across almost all performance criteria. For instance, lociPARSE attains the highest global Pearson's r of 0.67 which is much better than the second-best ARES (0.54). The same trend continues for global Spearman's ρ (lociPARSE: 0.72 vs. the second-best ARES: 0.63) and global Kendall's τ (lociPARSE: 0.54 vs. the second-best ARES: 0.46). Additionally, lociPARSE attains the lowest diff of 0.09, which is lower than the second-best rsRNASP (0.11). Furthermore, lociPARSE always delivers the highest per-target average correlations. In terms of average lDDT loss, however, DFIRE-RNA attains the lowest average loss (0.05). Meanwhile, lociPARSE, ARES, and rsRNASP are tied at the second spot with a comparably low loss of 0.06. Of note, ARES, the second-best performing method after lociPARSE in terms of global correlations, exhibits poor diff. DFIRE-RNA, the method attaining the lowest average lDDT loss, does not deliver top performance in terms of global correlations. That is, there are complementary aspects of scoring and model quality estimation that can lead to performance trade-offs. Our new method lociPARSE strikes an ideal balance to deliver top-notch RNA scoring performance across a wide range of assessment metrics simultaneously.

Table 1: Performance on 30 independent RNA targets, sorted in decreasing order of global Pearson's r. Values in bold indicate the best performance.

Method	**Global**				**Per-target average**			
	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Diff $\downarrow$	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Loss $\downarrow$
lociPARSE	**0.67**	**0.72**	**0.54**	**0.09**	**0.77**	**0.74**	**0.61**	0.06
ARES	0.54	0.63	0.46	0.19	0.73	0.7	0.58	0.06
rsRNASP	0.5	0.5	0.36	0.11	0.75	0.69	0.55	0.06
RASP	0.46	0.5	0.36	0.16	0.72	0.68	0.58	0.08
DFIRE-RNA	0.33	0.32	0.22	0.19	0.75	0.7	0.56	**0.05**
cgRNASP	0.27	0.19	0.13	0.15	0.12	0.07	0.07	0.07
RNA3DCNN	-0.09	-0.06	-0.04	0.25	0.62	0.57	0.41	0.07

It is interesting to note that among the other tested methods, the two machine learning-based scoring function ARES and RNA3DCNN show dramatically different performance. While ARES is consistently better than the traditional knowledge-based statistical potentials in terms of global correlations and comparable in terms of per-target average correlations, RNA3DCNN exhibits poor global and per-target average

correlations, which are much lower than most knowledge-based statistical potentials. A similar trend can be observed between rsRNASP and its coarse-grained counterpart cgRNASP, where rsRNASP consistently attains good global and per-target average correlations but cgRNASP falls short. That is, subtle methodological differences such as the granularity of RNA conformational space representation or the choice of the neural network architecture can lead to dramatic difference in performance. Meanwhile, the novel use of locality-aware invariant point attention in lociPARSE substantially improves RNA scoring performance, surpassing both machine learning-based scoring functions and knowledge-based statistical potentials.

2.4 Performance on CASP15 RNA targets

To investigate the ability of lociPARSE to distinguish 'good' and 'bad' models in comparison with the other tested methods, where a threshold of lDDT > 0.75 is used to differentiate 'good' and 'bad' models following CASP15 official assessment [7], we performed ROC analysis using all structural models for all targets in CASP15. Figure 2a shows the ROC curves with AUC values. Once again, lociPARSE achieves the highest AUC value of 0.96, which is much higher than the second-best method rsRNASP (0.83), demonstrating its better performance in separating good and bad models compared to the others. Furthermore, as shown in Figure 2b, lociPARSE attains the lowest average lDDT loss of 0.07, which is noticeably lower than the second-best rsRNASP and cgRNASP (0.11). It is interesting to note that DFIRE-RNA, the method attaining the lowest loss in 30 independent RNA targets, yields a poor loss (0.16) in CASP15. By contrast, lociPARSE consistently attains low loss in both test sets, indicating its ability to select the best model that generalizes across different datasets. Supplementary table S1 reporting the full set of assessment metrics for lociPARSE against all other tested methods in the CASP15 set further demonstrates the performance generalizability of lociPARSE, which consistently outperforms all other tested methods across almost all metrics.

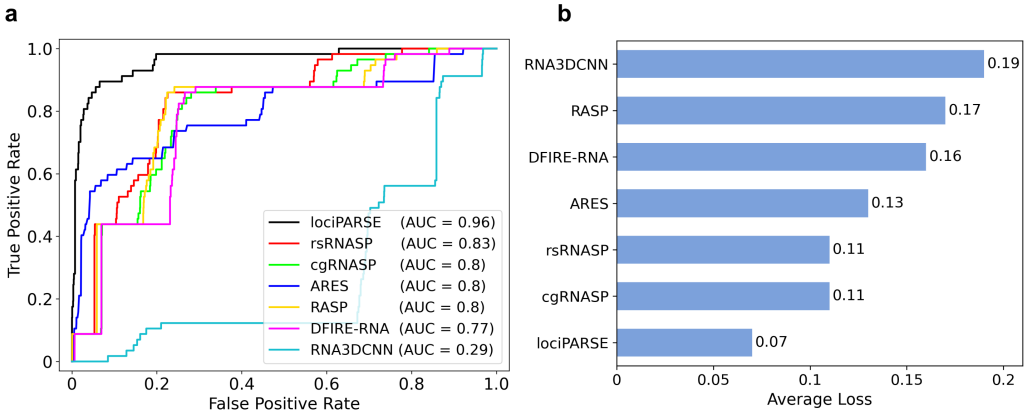

Figure 2: Distinguishability between 'good' and 'bad' models and best model selection performance on CASP15 set. (a) Receiver operating characteristic (ROC) curves with the area under the curve (AUC) values reported, where a threshold of lDDT > 0.75 is used to separate 'good' and 'bad' models. (b) Average top-1 lDDT loss.

Table 2: Local nucleotide-wise scoring performance on CASP15 set. Values in bold indicate the best performance.

Method	**Global**			
	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Diff $\downarrow$
lociPARSE	**0.73**	**0.74**	**0.53**	**0.14**
RNA3DCNN	0.17	0.15	0.1	0.43

When local nucleotide-wise quality is evaluated, lociPARSE is orders of magnitude better than RNA3DCNN, the only other method except lociPARSE that can estimate per-nucleotide score for local quality assessment.

For example, lociPARSE attains more than four times higher global Pearson's r, Spearman's ρ, and Kendall's τ than RNA3DCNN, and achieves less than one-third of the diff attained by RNA3DCNN (Table 2). In summary, lociPARSE represents a leap forward in terms of local nucleotide-wise scoring performance.

2.5 Case study

Figure 3 shows a representative example of local nucleotide-wise quality estimation using lociPARSE for a top-ranked structural model submitted by the winning group AIchemy_RNA2 (group 232) for the CASP15 target R1108 having length of 69. The predicted nucleotide-wise lDDT (pNuL) scores are in close agreement with the ground truth lDDT with a high Pearson's r of 0.89 (Figure 3a). Two local problematic regions are estimated by lociPARSE in nucleotide positions $(19-27)$ and $(59-63)$. These two local problematic regions are visually noticeable when the predicted structural model is aligned to the experimental structure. The poorly modeled structural regions around the hairpin loop in nucleotide positions $(19-27)$ and part of the helix strand in positions $(59-63)$ are obvious even with simple visual inspection (Figure 3b). By contrast, virtually all nucleotides with high pNuL values are structurally well modeled.

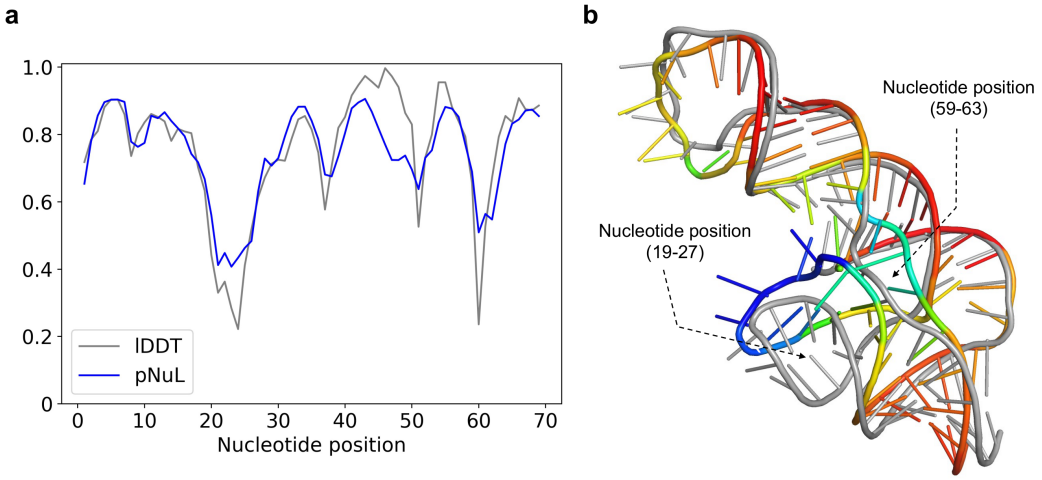

Figure 3: Example of lociPARSE local nucleotide-wise quality estimation for the CASP15 target R1108. (a) Predicted nucleotide-wise lDDT (pNuL) vs. the ground truth lDDT for the top-ranked structural model submitted by AIchemy_RNA2 (group 232). (b) The predicted structural model in rainbow colored with color code ramping from blue to red for low to high pNuL values superimposed on the experimental structure in gray, and two local problematic regions highlighted.

2.6 Ablation study and hyperparameter selection

To examine the relative importance of the features and architectural hyperparameters adopted in lociPARSE, we conduct ablation experiments by systematically varying individual parameter during model training using the reduced training set and evaluating the accuracy on the independent validation set (see Section 4.3). Table 3 reports the composite quality score (Q_c) defined in section 4.5 of the full-fledged version of lociPARSE serving as a baseline and its ablated variants. The results demonstrate that all the parameters adopted in the full-fledged version of lociPARSE positively contribute to the overall accuracy achieved by lociPARSE. For example, we notice performance decline when we vary the value of K used in the nearest neighbors for the locality computation from K = 20 used in the baseline to K $\in \{5, 10, 30\}$. Furthermore, to bias the attention weights as well as to update scalar features, we make use of the pair features in the form of nucleotide-nucleotide atomic distances between all pair of 3 atoms P, C4′ and N, encoded with Gaussian radial basis functions (hereafter called $\text{pair}_{\varphi(d)}$). We notice a significant performance drop when $\text{pair}_{\varphi(d)}$

features are isolated. Similarly, we notice consistent performance decline from the baseline configuration whenever we vary the network architecture such as the number of IPA layers (N_{layers} = L) or various IPA hyperparameters (N_{heads} = H, $N_{\text{query points}}$ = Q, $N_{\text{value points}}$ = V), justifying our choice of the parameters adopted in the full-fledged version of lociPARSE.

Table 3: Validation set performance in terms of composite quality score (Q_c) with various settings of features and hyperparameters compared to the full-fledged version of lociPARSE serving as a baseline. Values in bold indicate the best performance.

Settings	Q_c	Hyperparamters	Q_c	Hyperparamters	Q_c
Baseline (K = 20 w/ $\text{pair}_{\varphi(d)}$)	**0.824**	Baseline (L = 4, H = 4)	**0.824**	Baseline (Q = 8, V = 4)	**0.824**
K = 5 w/ $\text{pair}_{\varphi(d)}$	0.806	L = 2	0.814	Q = 2	0.81
K = 10 w/ $\text{pair}_{\varphi(d)}$	0.822	L = 6	0.808	Q = 4	0.81
K = 30 w/ $\text{pair}_{\varphi(d)}$	0.812	H = 2	0.808	V = 2	0.808
K = 20 w/o $\text{pair}_{\varphi(d)}$	0.768	H = 8	0.814	V = 8	0.818

3 Discussion

In this work, we developed lociPARSE, a locality-aware invariant point attention model for scoring RNA 3D structures. lociPARSE uses locality information derived from the RNA atomic coordinates to define nucleotide-wise frames together with its local atomic environment. This, coupled with the invariant point attention architecture, allows for the simultaneous estimation of local quality in the form of predicted nucleotide-wise lDDT (pNuL) scores which are then aggregated over all nucleotides to estimate global structural correctness in the form of predicted molecular-level lDDT (pMoL). Our empirical results demonstrate the superiority of our method in scoring RNA 3D structures compared to existing approaches.

Our locality-aware attention-based architecture can be extended in several ways, including estimating other local quality measures such as the Interaction Network Fidelity (INF) score [34], which is a local interaction metric that captures various types of base–base interactions in RNA. In fact, INF and lDDT have been shown to correlate well in a near-linear and size-independent relationship [7], suggesting that lDDT may capture the subset of interactions measured in INF whereas INF focuses on a selection of RNA-specific interactions. A model with a very similar architecture as lociPARSE would make an excellent candidate for jointly estimating INF and lDDT, thereby capturing complementary aspects of local quality. Further, a promising direction for future work is to investigate the potential benefits of capturing multi-state conformational landscape of RNA, since many RNA targets exhibit conformational flexibility [7]. The lDDT score can be computed simultaneously against multiple reference structures of the same RNA at the same time, without arbitrarily selecting one reference structure for the target or removing parts that show variability. Training our model using multi-reference lDDT to capture different classes of conformations will allow our scoring function to account for conformational flexibility and pave the way to evaluate predictions of conformational ensembles instead of just a single structure.

One limitation of our method is that it does not account for stereochemical quality and physical plausibility of the model being evaluated. This is because unlike for proteins, the currently available implementation of lDDT for RNA do not penalize for stereochemical violations. Using a customized version of lDDT that incorporates stereochemical quality checks in its calculation can address such limitation, and this aspect remains an important future direction.

4 Methods

4.1 Model input

Our model uses only input features derived directly from nucleotide sequence and RNA 3D structural coordinates. We use just the basic nucleotide-level encodings for our input. These include one-hot encoding of the nucleotide (i.e., a binary vector of 5 entries indicating each of the 4 nucleotide types and one for non standard nucleotide) and the relative position of the nucleotide in its sequence calculated as i/L (where i is the nucleotide index and L is the sequence length). For our pair features, we use sequential separation of a nucleotide pair and their spatial proximity information. The sequence separation i.e., the absolute difference between the two nucleotide indices, is discretized into 5 bins and represented by one-hot encoding where the first two bins correspond to self-loops and adjacent bonds respectively. Rest of the three bins are defined based on three types of interactions depending on the sequence separation: short-range (2-5), medium-range (6-24) and long-range (>24), similar to [35, 36]. The other component of our pair features includes nucleotide-nucleotide atomic distances between all pair of P, C4′ and glycosidic N atoms, encoded with Gaussian radial basis functions. It is important to note that all of our nucleotide and pair features are invariant in nature, consistent with the invariant layers of the IPA module.

4.2 Network architecture

4.2.1 Construction of local nucleotide frames

To perform invariant point attention on a set of 3D points, we represent each nucleotide in a geometric abstraction using the concept of frames. Each nucleotide frame in the form of a tuple is defined as an Euclidean transform $T = (R, \vec{\mathbf{t}})$, where $R \in \mathbb{R}^{3\times 3}$ is a rotation matrix and $\vec{\mathbf{t}} \in \mathbb{R}^3$ is the translation vector that can be applied to transform a position in local coordinates ($\vec{\mathbf{x}}_{local} \in \mathbb{R}^3$) to a position in global coordinates ($\vec{\mathbf{x}}_{global} \in \mathbb{R}^3$) as:

$$\begin{aligned} \vec{\mathbf{x}}_{global} &= T \circ \vec{\mathbf{x}}_{local} \\ &= (R, \vec{\mathbf{t}}) \circ \vec{\mathbf{x}}_{local} \\ &= R\vec{\mathbf{x}}_{local} + \vec{\mathbf{t}} \end{aligned} \tag{1}$$

In our setting, we define local nucleotide frames from the Cartesian coordinates of P, C4′, and glycosidic N atoms of the input RNA 3D structure and construct 3-bead coordinate frame using a Gram–Schmidt process specified in Alphafold2 (Algorithm 21) that takes the input coordinates (scaled by 0.1) of P as $\vec{\mathbf{x}_1}$, C4′ as $\vec{\mathbf{x}_2}$, and N as $\vec{\mathbf{x}_3}$. Note that the translation vector $\vec{\mathbf{t}}$ is assigned to the centre atom $\vec{\mathbf{x}_2}$.

4.2.2 Locality-aware invariant point attention

The formulation of locality-aware IPA used in our work combines sequence representation, $\mathbf{s}_i$, from each nucleotide i of the input RNA, pair representation $\mathbf{e}_{ij}$ of nucleotide i with other nucleotides j based on nucleotide pair adjacencies capturing the local atomic environment $\mathcal{N}$ of nucleotide i, where $j \in \mathcal{N}_i$ is the locality information derived from the RNA atomic coordinates. Consequently, the update function of the IPA layer is as follows:

$$\mathbf{s}'_i = \mathbf{s}_i + IPA\Big(\mathbf{s}_i, \{\mathbf{s}_j\}_{j\in\mathcal{N}_i}, \{\mathbf{e}_{ij}\}_{j\in\mathcal{N}_i}\Big) \tag{2}$$

To perform attention on 3D point clouds, IPA derives query ($\mathbf{q}_i^h$), key ($\mathbf{k}_i^h$) and value ($\mathbf{v}_i^h$) embeddings from a linear projection of $\mathbf{s}_i$ to a latent representation of dimension c for each nucleotide i, where $\mathbf{q}_i^h, \mathbf{k}_i^h, \mathbf{v}_i^h \in \mathbb{R}^c$ and $h \in \{1, \dots, N_{head}\}$ which represents number of attention heads in the IPA module. 3D query, key and value points are also generated considering the local frame T_i of each nucleotide i, where $\vec{\mathbf{q}}_i^{hp}, \vec{\mathbf{k}}_i^{hp} \in$

$\mathbb{R}^3, p \in \{1, \ldots, N_{\text{query points}}\}$ and $\vec{\mathbf{v}}_i^{hp} \in \mathbb{R}^3, p \in \{1, \ldots, N_{\text{value points}}\}$. The IPA module acts on a set of frames (parameterized as Euclidean transforms of the local frame T_i) and is invariant under global Euclidean transformations T_{global} on said frames. By performing locality-aware geometry and edge-biased attention, the IPA module transforms the 3D points from the target nucleotide's local frame into a global reference frame for computing the attention weights as follows:

$$a_{ij}^h = \text{softmax}_j \left(w_L \left(\frac{1}{\sqrt{c}} \mathbf{q}_i^{h^\top} \mathbf{k}_j^h + b_{ij}^h - \frac{\gamma^h w_C}{2} \sum\nolimits_p \left\| T_i \circ \vec{\mathbf{q}}_i^{hp} - T_j \circ \vec{\mathbf{k}}_j^{hp} \right\|^2 \right) \right) \tag{3}$$

where, b_{ij}^h is the attention bias derived from the linear projection of $\mathbf{e}_{ij}$ to hidden dimension c, weighting factors w_L and w_C are taken from the IPA formulation specified in AlphaFold2 and $\gamma^h \in \mathbb{R}$ is a learned scalar value. The attention mechanism acting on a set of local frames ensures invariance under global Euclidean transformations such as global rotations and translations of the input RNA due to the invariant nature of ℓ_2-norm of a vector under such rigid transformations.

The attention weights are used to compute the outputs of the attention mechanism, while mapping them back to the local frame and preserving invariance, as follows:

$$\tilde{\mathbf{o}}_i^h = \sum\nolimits_j a_{ij}^h \mathbf{e}_{ij} \tag{4}$$

$$\mathbf{o}_i^h = \sum\nolimits_j a_{ij}^h \mathbf{v}_j^h \tag{5}$$

$$\vec{\mathbf{o}}_i^{hp} = T_i^{-1} \circ \sum\nolimits_j a_{ij}^h \left(T_j \circ \vec{\mathbf{v}}_j^{hp} \right) \tag{6}$$

The outputs of the attention mechanism are then concatenated and passed through a linear layer to compute the updated sequence representations $\mathbf{s}_i'$ of each nucleotide as follows:

$$\mathbf{s}_i' = \mathbf{s}_i + \text{Linear}\left(\text{concat}_{h,p}(\tilde{\mathbf{o}}_i^h, \mathbf{o}_i^h, \vec{\mathbf{o}}_i^{hp}, \left\| \vec{\mathbf{o}}_i^{hp} \right\|) \right) \tag{7}$$

The updated sequence embeddings $\mathbf{s}_i'$ for each nucleotide i are subsequently stacked together to obtain the embedding $\mathbf{s}'$ for all nucleotides in the RNA. Finally, a linear layer followed by a 2 layer fully-connected network implemented as a multilayer perceptron (MLP) are used to obtain the final representation $\mathbf{s}^f$ before estimating nucleotide-wise lDDT scores as follows:

$$\mathbf{s}^f = \mathbf{s}' + \text{MLP}(\mathbf{s}') \tag{8}$$

4.3 Training and validation datasets

To curate our training datatset, we first obtained the training dataset used in trRosettaRNA [10] containing $3,632$ RNA targets. We then filtered this set by removing duplicate chains and discontinuous structures, separating monomers from complexes, splitting multiple chains into single chains, and correcting formatting issues in the coordinates files. We removed sequences with length > 200 nucleotides and ensured that our training and test sets are non-redundant by running CD-HIT-est [37] with default parameter settings, which reduced the training set to $1,399$ RNA targets. We generated a total of about $52,000$ structural models for the $1,399$ targets using a combination of different RNA 3D structure prediction tools including recent deep learning-enabled RNA structure prediction methods [9–13], physics-based RNA folding [32], and experimental structure perturbation using PyRosetta [33]. In addition, we separately curated a validation set for ablation study and hyperparameter selection from the Protein Data Back (PDB) [38] with experimental structures released between January 1, 2022 and July 6, 2023. Such a date range was chosen to avoid any overlap with our training dataset collected from trRosettaRNA which used structures released before January 1, 2022.

Once again, we used CD-HIT-est with default parameter settings to ensure non-redundancy of the validation dataset, resulting in 60 RNA targets. We generated 3D structural models for each of these 60 RNAs using the recent deep learning-based RNA structure prediction methods [9–13]. We created a reduced training subset consisting of $6,872$ structural models for $1,399$ RNA targets through clustering [39] for ablation study and hyperparameter selection.

4.4 Training details

To train our model, lociPARSE, we obtained nuculeotide-wise ground truth lDDT scores by comparing the predicted structural models in our training dataset against the corresponding experimental structures using the docker version of OpenStructure [40] available at `https://git.scicore.unibas.ch/schwede/openstructure/-/tree/master/docker`. During the ground truth lDDT computation, we enabled the option '–lddt-no-stereochecks' to skip stereochemical quality checks in its calculation following the recent CASP assessment in [7]. lociPARSE was implemented in PyTorch [41] with $\mathcal{L}1$ loss function to learn the mean absolute error between ground truth lDDT and predictions on nucleotide level, thereby formulating the local nuculeotide-wise quality estimation as a regression task. We trained our model using the Adam optimizer [42] having parameters $\beta 1 = 0.9$ and $\beta 2 = 0.999$ with a learning rate of 0.001 and dropout rate of 0.1. The training process consists of 50 epochs on an 80-GB NVIDIA A100 GPU.

4.5 Competing methods and evaluation metrics

lociPARSE is compared against both traditional knowledge-based statistical potentials (rsRNASP [17], RASP [19], DFIRE-RNA [20], and cgRNASP [18]) and recent machine learning-based RNA scoring functions (RNA3DCNN [28] and ARES [29]). rsRNASP is an all-atom distance-dependent potential considering short and long-ranged interactions present in RNA based on sequence separation aiming to capture the hierarchical nature of RNA folding [43]. Ribonucleic Acids Statistical Potential (RASP) is another all-atom statistical potential based on the averaging reference state. Similar to rsRNASP, RASP also separates interaction pairs into local and non-local categories and takes into account the base stacking and base pairing interactions present in RNA. DFIRE-RNA is yet another distance-scaled statistical potential designed using finite-ideal-gas reference state. Finally cgRNASP, a coarse-grained counterpart of rsRNASP potential introduces three different variants of coarse-grained potentials for RNA scoring. We have used the 3-bead representation of cgRNASP in this work which takes into account P, C4′ and N atoms. The Atomic Rotationally Equivariant Scorer (ARES) is a equivariant graph neural network which scores RNA structures by identifying complex structural motifs through equivariant convolutions. ARES employs E3NN [44] to predict the global RMSD of the structure. Finally, RNA3DCNN uses 3D convolutional neural network to predict the RMSD-like unfitness score of a nucleotide to its surroundings by considering RNA 3D structure as a 3D image and representing each nucleotide as an array of voxels. For prediction, we used the model that was trained on samples generated from both molecular dynamics (MD) and Monte Carlo (MC) simulations. It is worth noting that except lociPARSE, RNA3DCNN is the only other method that estimates both local and global quality.

Our assessment metrics include global and average per-target Pearson (r), Spearman rank (ρ) and Kendall's Tau rank (τ) correlation coefficients between the estimated score and the ground truth lDDT, computed both between the predicted molecular-level lDDT (pMoL) and ground truth lDDT for the overall structure as well as between the predicted nucleotide-wise lDDT scores (pNuL) and ground truth lDDT for individual nucleotides. Meanwhile, a higher correlation indicates better performance. Diff, another assessment metric, is calculated at the global level as the mean absolute difference between pMoL and ground truth lDDT. Loss or top-1 lDDT loss is calculated as the absolute difference between the ground truth lDDT of the structural model ranked at the top by pMoL and the ground truth lDDT of the most accurate structural model for each target averaged over all targets. Lower values of diff and loss, therefore, indicate better performance. We additionally perform receiver operating characteristics (ROC) analysis using a lDDT threshold of 0.75 to separate 'good' and 'bad' structural models, following [7]. Consequently, the area under the ROC curve

(AUC) quantifies the ability of a scoring function to distinguish good and bad models. Finally, an average of all assessment metrics is taken to combine the results of all the different metrics into a single composite quality score, called Q_c, defined as:

$$Q_c = \frac{1}{5}\Big(r_g + (1 - D) + r_a + (1 - L) + AUC\Big) \tag{9}$$

where, r_g = global Pearson's r, D = global diff, r_a = per-target average Pearson's r, L = average loss, and AUC = area under the ROC curve. We use the composite quality score for ablation study and hyperparameter selection, where higher values of Q_c indicate better performance.

5 Acknowledgements

This work was partially supported by the National Institute of General Medical Sciences (R35GM138146 to D.B.) and the National Science Foundation (DBI2208679 to D.B.).

References

[1] Kieft, J. S. Experiences gathered and lessons learned from 20 years of RNA structure. *RNA* **21**, 661–663 (2015).

[2] Vicens, Q. & Kieft, J. S. Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences* **119**, e2112677119 (2022).

[3] Miao, Z. & Westhof, E. RNA structure: advances and assessment of 3D structure prediction. *Annual review of biophysics* **46**, 483–503 (2017).

[4] Keating, K. S., Humphris, E. L. & Pyle, A. M. A new way to see RNA. *Quarterly reviews of biophysics* **44**, 433–466 (2011).

[5] Cable, J. *et al.* Noncoding RNAs: biology and applications—a Keystone Symposia report. *Annals of the new York Academy of Sciences* **1506**, 118–141 (2021).

[6] Zhang, J., Lang, M., Zhou, Y. & Zhang, Y. Predicting RNA structures and functions by artificial intelligence. *Trends in Genetics* (2023).

[7] Das, R. *et al.* Assessment of three-dimensional RNA structure prediction in CASP15. *bioRxiv* (2023).

[8] Miao, Z. *et al.* RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* **26**, 982–995 (2020).

[9] Pearce, R., Omenn, G. S. & Zhang, Y. De novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. *bioRxiv* 2022–05 (2022).

[10] Feng, C. *et al.* Accurate de novo prediction of RNA 3D structure with transformer network. *bioRxiv* 2022–10 (2022).

[11] Li, Y. *et al.* Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nature Communications* **14**, 5745 (2023).

[12] Baek, M., McHugh, R., Anishchenko, I., Baker, D. & DiMaio, F. Accurate prediction of nucleic acid and protein-nucleic acid complexes using RoseTTAFoldNA. *bioRxiv* 2022–09 (2022).

[13] Shen, T. *et al.* E2Efold-3D: end-to-end deep learning method for accurate de novo RNA 3D structure prediction. *arXiv preprint arXiv:2207.01586* (2022).

[14] Lange, S. J. *et al.* Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic acids research* **40**, 5215–5226 (2012).

[15] Micheletti, C., Di Stefano, M. & Orland, H. Absence of knots in known RNA structures. *Proceedings of the National Academy of Sciences* **112**, 2052–2057 (2015).

[16] Das, R. & Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proceedings of the National Academy of Sciences* **104**, 14664–14669 (2007).

[17] Tan, Y.-L., Wang, X., Shi, Y.-Z., Zhang, W. & Tan, Z.-J. rsRNASP: A residue-separation-based statistical potential for RNA 3D structure evaluation. *Biophysical Journal* **121**, 142–156 (2022).

[18] Tan, Y.-L., Wang, X., Yu, S., Zhang, B. & Tan, Z.-J. cgRNASP: coarse-grained statistical potentials with residue separation for RNA structure evaluation. *NAR Genomics and Bioinformatics* **5**, lqad016 (2023).

[19] Capriotti, E., Norambuena, T., Marti-Renom, M. A. & Melo, F. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* **27**, 1086–1093 (2011).

[20] Zhang, T., Hu, G., Yang, Y., Wang, J. & Zhou, Y. All-atom knowledge-based potential for RNA structure discrimination based on the distance-scaled finite ideal-gas reference state. *Journal of computational biology* **27**, 856–867 (2020).

[21] Wang, J., Zhao, Y., Zhu, C. & Xiao, Y. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic acids research* **43**, e63–e63 (2015).

[22] Samudrala, R. & Moult, J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *Journal of molecular biology* **275**, 895–916 (1998).

[23] Lu, H. & Skolnick, J. A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins: Structure, Function, and Bioinformatics* **44**, 223–232 (2001).

[24] Rykunov, D. & Fiser, A. New statistical potential for quality assessment of protein models and a survey of energy functions. *BMC bioinformatics* **11**, 1–11 (2010).

[25] Zhou, H. & Zhou, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein science* **11**, 2714–2726 (2002).

[26] Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein science* **15**, 2507–2524 (2006).

[27] Zhang, J. & Zhang, Y. A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. *PloS one* **5**, e15386 (2010).

[28] Li, J. *et al.* RNA3DCNN: Local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLoS computational biology* **14**, e1006514 (2018).

[29] Townshend, R. J. *et al.* Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).

[30] Mariani, V., Biasini, M., Barbato, A. & Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).

[31] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

[32] Boniecki, M. J. *et al.* SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic acids research* **44**, e63–e63 (2016).

[33] Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

[34] Parisien, M., Cruz, J. A., Westhof, É. & Major, F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* **15**, 1875–1885 (2009).

[35] Jian, Y. *et al.* DIRECT: RNA contact predictions by integrating structural patterns. *BMC bioinformatics* **20**, 1–12 (2019).

[36] Weinreb, C. *et al.* 3D RNA and functional interactions from evolutionary couplings. *Cell* **165**, 963–975 (2016).

[37] Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

[38] Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Research* **28**, 235–242 (2000).

[39] Zhang, Y. & Skolnick, J. SPICKER: a clustering approach to identify near-native protein folds. *Journal of computational chemistry* **25**, 865–871 (2004).

[40] Biasini, M. *et al.* OpenStructure: an integrated software framework for computational structural biology. *Acta Crystallographica Section D: Biological Crystallography* **69**, 701–709 (2013).

[41] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32** (2019).

[42] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[43] Brion, P. & Westhof, E. Hierarchy and dynamics of RNA folding. *Annual review of biophysics and biomolecular structure* **26**, 113–137 (1997).

[44] Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453* (2022).