

PARSEbp: Pairwise Agreement-based RNA Scoring with Emphasis on Base Pairings

Sumit Tarafder¹ and Debswapna Bhattacharya^{1,*}

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

*Corresponding author. dbhattacharya@vt.edu

Abstract

Motivation: High-fidelity scoring of RNA 3D structures remains a major challenge in RNA structure prediction and conformational sampling. While single-model methods for scoring RNA structures can capture individual structural features, they fail to capture the broader structural consensus within a conformational ensemble, limiting their effectiveness in ranking and model selection.

Results: We present PARSEbp, a fast and effective multi-model RNA scoring method that integrates pairwise structural agreement across the conformational ensemble with base pairing consistency. By leveraging both alignment-based global structural agreement at the 3D level and base pairing consistency at the 2D level, PARSEbp efficiently constructs a consensus similarity matrix from which per-structure accuracy scores are computed. Tested on 16th Critical Assessment of Structure Prediction (CASP16) RNA targets, PARSEbp significantly outperforms existing single- and multi-model RNA scoring functions, including traditional statistical potentials, state-of-the-art deep learning methods, and consensus-based approaches, as well as a baseline variant of PARSEbp without the emphasis on base pairings, across a wide range of complementary assessment metrics.

Availability: PARSEbp is freely available at <https://github.com/Bhattacharya-Lab/PARSEbp>.

Key words: RNA scoring function, base pairing, structural ensemble, RNA 3D structure prediction

1 Introduction

The three-dimensional (3D) structure of RNA plays a central role in determining its catalytic activity, regulatory capacity, and molecular interactions [1, 2]. However, unlike proteins, RNAs are structurally flexible and can adopt diverse conformations stabilized by canonical and non-canonical base pairs, long-range interactions, and interactions with other molecules [3, 4, 5, 6]. Due to the inherent flexibility of RNA molecules, ensemble-based approaches for RNA 3D structure prediction are crucial to capture the range of conformations plausible for a given sequence [7]. Recent advances in computational modeling of RNA 3D structures involving both deep learning-based methods [8, 9, 10, 11] and stochastic energy- or template-based approaches [12, 13] enable the generation of a large pool of candidate 3D structures for a given sequence. Despite the progress, automated prediction of RNA 3D structures still lags behind the human-expert modeling groups in the recent community-wide blind assessments [14, 15, 16], largely due to data scarcity, structural complexity, and weak evolutionary signal of RNA sequences [17]. To render the computational ensembles practically useful, RNA scoring methods are essential to identify high-quality models from misfolded alternatives present in the ensemble to improve the accuracy of RNA 3D structure prediction [18, 19].

Existing approaches for 3D RNA scoring can be broadly divided into two categories: single-model and multi-model methods. Single-model scoring methods evaluate each candidate structure independently without considering the alternative conformations available in the ensemble. These methods rely on

either knowledge-based statistical potentials [20, 21, 22, 23] or deep learning architectures [24, 25, 26, 27] to estimate specific structural quality scores. However, all single-model scoring methods rely on experimental structures, either as supervised signal to train their architecture or fine-tune the statistical potential parameters, thus limited to the paucity of the available RNA 3D structures in Protein Data Bank (PDB) [28]. Additionally, such methods often fail to capture the broader consensus within a structure ensemble, limiting their effectiveness in ranking and model selection. In contrast, multi-model scoring methods can exploit the agreement across the ensemble of candidate structures to infer individual model quality in an unsupervised manner, without the need for experimental structures to train or fine-tune the parameters. A recent method RNActive [29], the only existing publicly available multi-model RNA scoring method in the literature, derives consensus base pairs by identifying recurrent patterns across RNA 3D folds to assign individual quality scores. However, RNActive does not perform an explicit all-vs-all pairwise comparison across the ensemble, limiting the ability to capture the full spectrum of structural agreement. This motivates the need to develop an accurate and computationally efficient multi-model RNA scoring method that can integrate both pairwise global structural similarity and base pairing fidelity across the entire ensemble to deliver robust quality estimates.

In this work, we present PARSEbp, a pairwise multi-model RNA scoring method that integrates global 3D structural similarity with 2D base pairing consistency. PARSEbp computes pairwise TM-scores [30] across the structural ensemble to capture global fold similarity with additional emphasis from pairwise

Interaction Network Fidelity (INF) scores [31], which quantifies the agreement of canonical and non-canonical base pairs between structures. These measures are combined into a consensus similarity matrix, from which per-structure quality scores are computed by averaging the agreement with all other models in the structural ensemble. Benchmarking on 25 RNA targets from the recently concluded CASP16 challenge, PARSEbp achieves the highest global Spearman’s correlation (0.92) while attaining the lowest top-1 loss of 0.05, demonstrating its effectiveness in top model selection from the diverse structural ensemble in CASP16. By unifying pairwise agreement at the 3D and 2D levels, PARSEbp provides a fast, accurate, and consensus-driven quality estimation tool, requiring only a few minutes to score an entire structural ensemble for a typical RNA of 100 nucleotides.

2 Materials & Methods

Given a set of N RNA 3D structures $S = \{S_1, S_2, \dots, S_N\}$ as input, we evaluate the structural quality of the decoys by combining pairwise 3D structural similarity with base pair (2D) consistency among the structures in the pool. Specifically, for each ordered pair of structures (S_i, S_j) with $i \neq j$, we compute the pairwise similarity score T_{ij} as follows:

$$T_{ij} = \text{TM-score}(S_i, S_j) \quad \forall i, j \quad T_{ij} \in [0, 1] \quad (1)$$

This results in a complete $N \times N$ pairwise similarity matrix $T = [T_{ij}]$ which encodes the global structural similarity in terms of TM-score [30] across the ensemble. To guide this global pairwise similarity, we further assess the base pairing consistency among the structures by extracting 2D base pair map S_i^{bp} from each 3D structure using MC-annotate [32], a widely used method for RNA base pair annotation. For every pair of structures (S_i, S_j) , we compute the pairwise Interaction Network Fidelity (INF) score [31], which quantitatively measures the overlap of base pairing patterns as a degree of agreement between the two structures. INF is defined as the geometric mean between precision and recall and is calculated in the following way:

$$\text{INF} = \sqrt{\frac{|TP|}{|TP| + |FP|} \times \frac{|TP|}{|TP| + |FN|}} \quad (2)$$

Here, true positives (TP) denote overlapping interaction pairs between structures (S_i) and (S_j) while false positives (FP) and false negatives (FN) denote the set of interactions only found in either (S_i) or (S_j) , respectively. The pairwise base pair consistency I_{ij} between two maps S_i^{bp} and S_j^{bp} , yielding the pairwise base pair similarity matrix $I = [I_{ij}]$ is defined as:

$$I_{ij} = \text{INF}(S_i^{bp}, S_j^{bp}) \quad \forall i, j \quad I_{ij} \in [0, 1] \quad (3)$$

We combine the pairwise 3D and 2D agreements by taking an element-wise multiplication between the two similarity matrices T and I as follows:

$$M = T \circ I, \quad M_{ij} = T_{ij} I_{ij} \quad (4)$$

This yields the consensus matrix $M = [M_{ij}]$, which captures both the global 3D similarity between structures and the extent to which this similarity is reinforced by consistent base pairing interactions. The quality score Q_i for each structure S_i is defined as the mean agreement with all other structures in the ensemble:

$$Q_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N M_{ij} \quad (5)$$

The resulting set of scores $Q = \{Q_1, Q_2, \dots, Q_N\}$ provides a quantitative self-assessment of the structures, where higher

scores correspond to structures consistently supported by both 3D structural similarity and 2D base pairing patterns across the ensemble. To ensure scalability for long RNA sequences and large conformational ensembles, all pairwise computations are parallelized and executed in batches.

We evaluate our method PARSEbp against a broad set of both single-model and multi-model RNA scoring methods. Among the single-model methods, we compare against traditional knowledge-based statistical potentials such as rsRNASP [20], rsRNASP1 [21], cgRNAsp [22], DFIRE-RNA [23] and state-of-the-art deep learning-based approaches, including RNA3DCNN [24], ARES [27], lociPARSE [25] and RNArank [26]. For multi-model methods, we compare PARSEbp against the only available web-based tool RNative [29] along with a baseline variant of PARSEbp without the emphasis on base pairings. Our benchmark set contains 4,750 decoy 3D structures across 25 RNA targets with sequence length ≤ 500 , as listed in **Supplementary Table S1**. Targets with length > 500 are omitted from evaluation, as CASP16 results indicate that they predominantly yielded low-quality decoys in the ensemble [16]. Ground truth scores of all the structures are obtained from the official CASP16 assessment, following the CASP16 composite evaluation formula [16] proposed by the assessors, defined as the weighted combination of three structural similarity measures: TM-score [30], Local Distance Difference Test (IDDT) [33] with stereochemical checks enabled, and GDT_TS [34], as follows:

$$G = 0.3 \times \text{TM-score} + 0.3 \times \text{GDT_TS} + 0.4 \times \text{IDDT} \quad (6)$$

The scoring performance of various methods is evaluated using global Pearson’s correlation (r), Spearman’s rank correlation (ρ), and root mean squared error (RMSE) between predicted and ground truth scores, along with per-target average r , ρ , and loss (top-1 loss), which is defined as the absolute difference between the ground truth of the top-ranked structure and that of the best structure for each target.

3 Results

Table 1 summarizes the performance of PARSEbp against competing single-model and multi-model RNA scoring methods on 25 CASP16 RNA targets using the composite ground truth metric G (Eq. 6). At the global level, PARSEbp achieves the highest Pearson’s correlation ($r = 0.91$) and Spearman’s correlation ($\rho = 0.92$) with respect to ground truth scores while yielding the lowest root mean square error (Error = 0.11). On a per-target basis, our method consistently delivers superior correlations and the lowest average loss compared to all other methods. PARSEbp exhibits a remarkable performance advantage compared to single-model approaches. Knowledge-based statistical potentials such as rsRNASP, rsRNASP1, cgRNAsp, and DFIRE-RNA achieve only modest performance, with global correlations not exceeding $r = 0.22$ and substantially higher errors, underscoring the challenge of generalizing from limited reference states. Deep learning-based methods perform comparatively better, with RNArank attaining the highest global Pearson’s correlation among single-model approaches ($r = 0.72$) and lociPARSE achieving the highest per-target average Pearson’s correlation ($r = 0.70$). In terms of per-target average loss, lociPARSE and RNA3DCNN both reach 0.09, representing the best performance within the single-model methods. Nonetheless, PARSEbp surpasses all single-model approaches in both global and per-target evaluations by large

Table 1. Performance on 25 CASP16 RNA targets (4,750 decoys) based on the CASP16 official assessment metric as the ground truth, sorted in increasing order of global Pearson’s r . Values in bold indicate the best performance.

Type	Method	Global			Per-target average		
		$r \uparrow$	$\rho \uparrow$	Error \downarrow	$r \uparrow$	$\rho \uparrow$	Loss \downarrow
Single-Model	DFIRE-RNA	0.12	0.02	0.55	0.57	0.58	0.11
	rsRNAspl	0.13	0.16	0.55	0.61	0.61	0.12
	cgRNAsp	0.18	0.09	0.48	0.56	0.54	0.12
	RNA3DCNN	0.20	0.24	0.54	0.68	0.68	0.09
	rsRNAsp	0.22	0.16	0.46	0.62	0.61	0.15
	ARES	0.44	0.44	0.34	0.57	0.57	0.21
	lociPARSE	0.61	0.62	0.32	0.70	0.67	0.09
	RNArank	0.72	0.80	0.28	0.66	0.59	0.10
Multi-Model	RNative *	0.52	0.60	0.45	0.63	0.63	0.11
	PARSEbp w/o bp	0.89	0.90	0.12	0.83	0.79	0.05
	PARSEbp	0.91	0.92	0.11	0.85	0.82	0.05

* Evaluated on 12 targets as RNative fails to run on the remaining set of CASP16 targets (see **Supplementary Table S3** for performance comparison against PARSEbp on common set of targets)

margins, achieving the highest per-target average Pearson’s correlation ($r = 0.85$) and the lowest top-1 loss (0.05). These results demonstrate the reliability of PARSEbp in consistently delivering top-notch performance across various complementary assessment metrics. Detailed per-target results are available in **Supplementary Table S2**.

When considering multi-model RNA scoring approaches, PARSEbp is compared against RNative, the only other multi-model method currently available in the literature. However, RNative is not only limited by ensemble size but also fails on roughly half of the CASP16 targets, particularly for targets dominated by low-quality structures, necessitating additional filtering based on clash score, bond lengths, or angle geometry to proceed, and ultimately making it unsuitable for evaluation across the full set of targets. On the reduced set of 12 targets where RNative runs successfully, its performance remains modest with a global error of 0.45 and a per-target average loss of 0.11, lower than even the state-of-the-art single-model approaches. To ensure fairness, we benchmark PARSEbp on this reduced set and report the results in **Supplementary Table S3**. On this reduced benchmark of common set of targets, PARSEbp consistently outperforms RNative, achieving a global Pearson’s correlation of $r = 0.87$ and error of 0.13, much better than RNative’s $r = 0.52$ and error of 0.45. At the per-target level, PARSEbp attains a loss of 0.03, far lower than that of RNative (0.11). To further assess the contribution of the emphasis on the base pair, we compare PARSEbp with a baseline variant that relies only on pairwise TM-score similarity, without the emphasis on base pairings. As shown in **Table 1**, removing base pairs leads to a noticeable drop in performance in terms of both global and per-target average correlations. The drops in correlations are statistically significant at 95% confidence level as revealed by the Wilcoxon signed-rank test with P -value = 0.02 for Pearson’s r and P -value = 0.03 for Spearman’s ρ , confirming that the incorporation of base pair consistency provides a meaningful gain in performance for multi-model RNA scoring. In summary, our method significantly outperforms the existing multi-model RNA scoring approach RNative and base pair information meaningfully contributes to the improved performance of PARSEbp.

In terms of computational efficiency, PARSEbp completes ensemble-wide scoring for typical RNA targets of length approximately 100 nucleotides within half a minute. By comparison, RNative requires substantially longer runtime (approximately half an hour) even for shorter targets (≤ 100), highlighting the suitability of PARSEbp for large-scale applications. Runtime analysis on the CASP16 benchmark set, provided in **Supplementary Figure S1**, further confirms that PARSEbp exhibits scalability with increasing sequence length and structural ensemble size without sacrificing accuracy. Collectively, these results demonstrate that PARSEbp is a robust and accurate method for multi-model RNA scoring.

4 Conclusion

We present PARSEbp, a fast, accurate, and reliable multi-model RNA scoring method that unifies pairwise global structural similarity at the 3D level and base pairing consistency at the 2D level. By explicitly performing all-vs-all alignment-based comparisons given a structural ensemble with emphasis on base pairings, PARSEbp provides consensus-driven high-fidelity scoring of RNA 3D structures, significantly outperforming all existing single- and multi-model methods in CASP16 test set. PARSEbp should be broadly useful for scoring RNA 3D structures at scale.

5 Data Availability

The open-source PARSEbp method is freely available at <https://github.com/Bhattacharya-Lab/PARSEbp>.

The RNA 3D decoy structures and corresponding ground truth scores for the CASP16 RNA targets have been downloaded from https://predictioncenter.org/download_area/CASP16/predictions/RNA/ and https://predictioncenter.org/casp16/results.cgi?view=targets&tr_type=rna, respectively.

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

- Jennifer A Doudna and Thomas R Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(6894):222–228, 2002.

2. Laura R Ganser, Megan L Kelly, Daniel Herschlag, and Hashim M Al-Hashimi. The roles of structural dynamics in the cellular functions of rnas. *Nature reviews Molecular cell biology*, 20(8):474–489, 2019.
3. Renée Schroeder, Andrea Barta, and Katharina Semrad. Strategies for rna folding and assembly. *Nature reviews Molecular cell biology*, 5(11):908–919, 2004.
4. Quentin Vicens and Jeffrey S Kieft. Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences*, 119(17):e2112677119, 2022.
5. Sungchul Hohng, Timothy J Wilson, Elliot Tan, Robert M Clegg, David MJ Lilley, and Taekjip Ha. Conformational flexibility of four-way junctions in rna. *Journal of molecular biology*, 336(1):69–79, 2004.
6. Robert C Spitale and Danny Incarnato. Probing the dynamic RNA structure and its functions. *Nature Reviews Genetics*, 24(3):178–196, 2023.
7. Elodie Laine, Sergei Grudinin, Roman Klypa, and Isaure Chauvot de Beauchêne. Navigating protein–nucleic acid sequence–structure landscapes with deep learning. *Current Opinion in Structural Biology*, 95:103162, 2025.
8. Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, pages 1–3, 2024.
9. Sumit Tarafder and Debswapna Bhattacharya. RNAbpFlow: Base pair-augmented SE (3)-flow matching for conditional RNA 3D structure generation. *bioRxiv*, 2025.
10. Yang Li, Chengxin Zhang, Chenjie Feng, Robin Pearce, P Lydia Freddolino, and Yang Zhang. Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nature Communications*, 14(1):5745, 2023.
11. Wenkai Wang, Chenjie Feng, Renmin Han, Ziyi Wang, Lisha Ye, Zongyang Du, Hong Wei, Fa Zhang, Zhenling Peng, and Jianyi Yang. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nature Communications*, 14(1):7266, 2023.
12. Andrew Martin Watkins, Ramya Rangan, and Rhiju Das. FARFAR2: improved de novo rosetta prediction of complex global RNA folds. *Structure*, 28(8):963–976, 2020.
13. Marcin Biesiada, Katarzyna J Purzycka, Marta Szachniuk, Jacek Blazewicz, and Ryszard W Adamiak. Automated RNA 3D structure prediction with RNAComposer. In *RNA Structure Determination: Methods and Protocols*, pages 199–215. Springer, 2016.
14. Zhichao Miao, Ryszard W Adamiak, Maciej Antczak, Michał J Boniecki, Janusz Bujnicki, Shi-Jie Chen, Clarence Yu Cheng, Yi Cheng, Fang-Chieh Chou, Rhiju Das, et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA*, 26(8):982–995, 2020.
15. Rhiju Das, Rachael C Kretscher, Adam J Simpkin, Thomas Mulvaney, Phillip Pham, Ramya Rangan, Fan Bu, Ronan M Keegan, Maya Topf, Daniel J Rigden, et al. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins: Structure, Function, and Bioinformatics*, 91(12):1747–1770, 2023.
16. Rachael C Kretscher, Alissa M Hummer, Shujun He, Rongqing Yuan, Jing Zhang, Thomas Karagianes, Qian Cong, Andriy Kryshchavych, and Rhiju Das. Assessment of nucleic acid structure prediction in CASP16. *bioRxiv*, pages 2025–05, 2025.
17. Bohdan Schneider, Blake Alexander Sweeney, Alex Bateman, Jiri Cerny, Tomasz Zok, and Marta Szachniuk. When will RNA get its AlphaFold moment? *Nucleic Acids Research*, 51(18):9522–9532, 2023.
18. Sunandan Mukherjee, S Naeim Moafinejad, Nagendar Goud Badepally, Katarzyna Merdas, and Janusz M Bujnicki. Advances in the field of RNA 3D structure prediction and modeling, with purely theoretical approaches, and with the use of experimental data. *Structure*, 32(11):1860–1876, 2024.
19. Ivona Martinović, Tin Vlašić, Yang Li, Bryan Hooi, Yang Zhang, and Mile Sikic. Generalization and Scoring in RNA 3D Structure Prediction: A Benchmarking Study. In *AI4X 2025 International Conference*, 2025.
20. Ya-Lan Tan, Xunxun Wang, Ya-Zhou Shi, Wenbing Zhang, and Zhi-Jie Tan. rsRNASP: A residue-separation-based statistical potential for RNA 3D structure evaluation. *Biophysical Journal*, 121(1):142–156, 2022.
21. En Lou, Chen-Chen Zheng, Shixiong Yu, Ya-Lan Tan, and Zhi-Jie Tan. rsRNASP1: A distance-and dihedral-dependent statistical potential for RNA 3D structure evaluation. *Biophysical Journal*, 124(16):2740–2753, 2025.
22. Ya-Lan Tan, Xunxun Wang, Shixiong Yu, Bengong Zhang, and Zhi-Jie Tan. cgRNASP: coarse-grained statistical potentials with residue separation for RNA structure evaluation. *NAR Genomics and Bioinformatics*, 5(1):lqad016, 2023.
23. Tongchuan Zhang, Guodong Hu, Yuedong Yang, Jihua Wang, and Yaoqi Zhou. All-atom knowledge-based potential for RNA structure discrimination based on the distance-scaled finite ideal-gas reference state. *Journal of computational biology*, 27(6):856–867, 2020.
24. Jun Li, Wei Zhu, Jun Wang, Wenfei Li, Sheng Gong, Jian Zhang, and Wei Wang. RNA3DCNN: Local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLoS computational biology*, 14(11):e1006514, 2018.
25. Sumit Tarafder and Debswapna Bhattacharya. lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures. *Journal of Chemical Information and Modeling*, 64(22):8655–8664, 2024.
26. Xiaocheng Liu, Wenkai Wang, Zongyang Du, Ziyi Wang, Zhenling Peng, and Jianyi Yang. Quality assessment of RNA 3D structure models using deep learning and intermediate 2D maps. *bioRxiv*, pages 2025–07, 2025.
27. Raphael JL Townshend, Stephan Eismann, Andrew M Watkins, Ramya Rangan, Masha Karelina, Rhiju Das, and Ron O Dror. Geometric deep learning of RNA structure. *Science*, 373(6558):1047–1051, 2021.
28. Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.
29. Jan Pieleśiak, Maciej Antczak, Marta Szachniuk, and Tomasz Zok. RNActive to recognize native-like structure in a set of RNA 3D models. *bioRxiv*, pages 2025–07, 2025.
30. Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. Us-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature methods*, 19(9):1109–1115, 2022.
31. Marc Parisien, José Almeida Cruz, Éric Westhof, and François Major. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA*, 15(10):1875–1885, 2009.
32. Patrick Gendron, Sébastien Lemieux, and François Major. Quantitative analysis of nucleic acid three-dimensional structures. *Journal of molecular biology*, 308(5):919–936, 2001.
33. Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.
34. Adam Zemla. LGA: a method for finding 3D similarities in protein structures. *Nucleic acids research*, 31(13):3370–3374, 2003.