# RNAtive to recognize native-like structure in a set of RNA 3D models

Jan Pielesiak[1], Maciej Antczak[1,2], Marta Szachniuk[1,2], and Tomasz Zok[1]

[1]Institute of Computing Science, Poznan University of Technology, Poznan, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

## Abstract

**Motivation:** Most widely used methods for evaluating RNA 3D structure models depend on the availability of experimental reference structures, which restricts their use for novel RNAs. Moreover, they often overlook recurrent structural features shared across multiple predictions of the same sequence. Although consensus approaches have proven effective in RNA sequence analysis and evolutionary studies, no existing tool applies these principles to evaluate ensembles of 3D models. This gap hampers the identification of native-like folds in computational predictions, particularly as AI-driven methods become increasingly prevalent.
**Results:** This paper presents RNAtive, the first computational tool to apply consensus-derived secondary structures for reference-free evaluation of RNA 3D models. It aggregates recurrent base-pairing and stacking interactions across ensembles of predicted 3D structures to construct a consensus secondary structure. Models are then ranked using two metrics derived from binary classification. Benchmark results show that models aligning with the consensus tend to exhibit native-like structural features, as validated against experimental data from CASP15. The RNAtive web server offers an intuitive platform for comparing and prioritizing RNA 3D predictions, providing a scalable solution to address the variability inherent in deep learning and fragment-assembly methods. By bridging consensus principles with 3D structural analysis, RNAtive advances the exploration of RNA conformational landscapes and has potential applications in fields like therapeutic RNA design.
**Availability:** RNAtive is a freely accessible web server with a modern, user-friendly interface, available for scientific, educational, and commercial use at `https://rnative.cs.put.poznan.pl/`.
**Contact:** tomasz.zok@cs.put.poznan.pl

## 1 Introduction

RNA molecules exhibit remarkable structural versatility, adopting diverse three-dimensional (3D) conformations influenced by environmental factors and molecular interactions [1, 2, 3, 4, 5]. While a single sequence can yield multiple stable states, as in riboswitches and aptamers, evolutionary pressure preserves a functional structural core [6]. These conserved core motifs – such as catalytic sites or ligand-binding pockets – remain invariant across diverse conformations, while peripheral regions may vary. Identifying these recurrent motifs is central to the concept of consensus, which enables the extraction of biologically meaningful patterns from structural diversity [1, 4].

In bioinformatics, consensus approaches are robust strategies for extracting signals from heterogeneous data. They are used to identify functional motifs in sequence alignments, infer conserved base-pairing patterns from multiple sequence alignments to improve RNA secondary structure prediction, and define invariant cores from ensembles of experimental 3D

models [7, 8, 9, 10, 11, 12].

Consensus principles, however, remain underutilized in the assessment of computational RNA 3D structure predictions. Current evaluation strategies predominantly rely on reference-dependent metrics (e.g., RMSD, INF) that require an experimental structure, limiting their applicability to novel RNAs [13, 14, 15, 16, 17]. Moreover, they fail to capture recurrent structural features across predicted conformations – a critical shortcoming given the high variability of models from emerging deep-learning methods. While other approaches can flag outliers or predict model quality, they do not systematically extract consensus signals from ensembles of predicted structures [18, 19, 20].

This work harnesses consensus principles to identify native-like RNA 3D structures within ensembles of *in silico* models. We present RNAtive, a web-based tool identifying common secondary structure patterns across RNA 3D folds. RNAtive constructs a consensus by aggregating recurrent base-pairing and stacking interactions, then evaluates each model's agreement with this consensus using two metrics: the F1-score and Interaction Network Fidelity (INF). Models aligning closely with the consensus are prioritized, as recurrent interactions are more likely to reflect stable, native-like folds – a principle validated by benchmarks like CASP15 [21]. The RNAtive web server provides a reference-free framework to evaluate computational tertiary predictions and guide their experimental validation.

# 2 System and methods

## 2.1 Input and output

RNAtive accepts tertiary structures in both PDB and PDBx/mmCIF formats. Users can upload multi-model structures or ZIP/TAR.GZ archives, which are automatically unpacked and divided into individual models. Because RNAtive relies on nucleotide interactions, all input models must be consistent regarding nucleotide composition. If needed, RNAtive remaps chain names and renumbers residues internally to achieve the expected consistency. However, all other differences in nucleotide composition cannot be automatically resolved. Thus, they are detected and prevented from submitting the inconsistent model.

An embedded MolProbity [18] filtering mechanism can analyze each input model for stereochemical quality. RNAtive utilizes MolProbity's classification of structural features, assigning *good*, *caution*, or *warning* labels based on factors such as clashscore (the number of clashing atoms per 1000 atoms in the structure), bad bonds, and bad angles. Users can exclude models with high clashscore only or improper stereochemistry from the analysis.

Optionally, the entire analysis can be subjected to 2D structure constraints. Users can provide a canonical 2D structure in dot-bracket format, using brackets `()`, `[]`, etc. for expected base pairs, dots `.` for unconstrained nucleotides, and `x` to indicate that a nucleotide should be unpaired (i.e., not forming either canonical or non-canonical base pairs). The inclusion of `x` allows dots to be interpreted as nucleotides that may form non-canonical interactions, consistent with how canonical 2D structures are typically encoded in dot-bracket format.

Users can also configure the analysis by selecting one of six available tools for RNA interaction annotation: RNApolis Annotator [22], BPNet [23], FR3D [24], MC-Annotate [25], RNAView [26], or Barnaba [27]. Additionally, RNAtive supports two modes of operation. If conditionally weighted consensus is enabled, RNAtive treats annotated interactions as a fuzzy set, where an interaction's membership is determined by its frequency of occurrence across all analyzed models. Alternatively, RNAtive can be parameterized with a confidence level, which defines a minimum frequency threshold for an interaction to be considered.

The results page provides an overview of the input parameter values and constraints, the derived consensus (including visualization and interaction tables), final model rankings, and detailed information for each model, including its MolProbity quality scores, if computed.

## 2.2 Workflow

RNAtive starts with a unification protocol to ensure consistent chain names and residue numbering (see Figure 2). All inputs are converted to PDB format since some external software, including MolProbity, only supports this format. It is essential to note that the conversion from PDBx/mmCIF to PDB can fail under certain conditions: 1) if the atom count exceeds 99,999; 2) if the residue count exceeds 9,999; or 3) if the chain count exceeds 62, which includes a combination of uppercase letters, lowercase letters, and digits. We recognize these limitations of RNAtive. However, for its primary purpose – analyzing computationally derived RNA 3D models – we believe the system can effectively handle any realistic dataset.

Once the model consistency is ensured, MolProbity (if selected by the user) filters out models that do not meet the user's quality expectations. The current processing will stop if only one or zero models pass this stage.

All valid models proceed to interaction analysis using the chosen tool. This analysis produces a set of unique interactions (such as base pairs and stacking), with each interaction accompanied by metadata detailing its occurrence frequency among the models and flags indicating compliance with or violations of 2D constraints given. We treat this multiset with accompanying metadata as the virtual consensus structure representing the ensemble of input RNA 3D models (Figure 1). A similar multiset of interactions and metadata is also computed for each model individually for scoring.

The interaction sets are used to compute RNAtive outputs. Two primary metrics are reported for each RNA 3D model: Interaction Network Fidelity (INF) and F1-score. When the confidence level parameter is set to a value $n$ (ranging from 1 to the number of input models), a consensus subset of nucleotide interactions is constructed, including only those interactions with a frequency of at least $n$. For each model, True Positives ($TP$) are interactions found in both the model and the consensus subset; False Positives ($FP$) are interactions present only in the model; and False Negatives ($FN$) are those found only in the consensus. INF and F1-score are then computed as

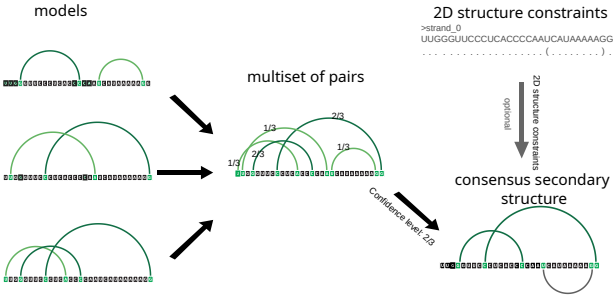


Figure 1: Example inputs, the multiset, and the consensus secondary structure. The additional constraint will benefit the first model and disadvantage the other two, despite the base pair not being included in the consensus, given the selected confidence level.

follows:

$$INF = \sqrt{\frac{|TP|}{|TP| + |FP|} \times \frac{|TP|}{|TP| + |FN|}}$$

$$F1 = \frac{2|TP|}{2|TP| + |FP| + |FN|}$$

When the conditionally weighted consensus mode is active, interaction sets are treated as fuzzy sets; that is, all interactions are considered, and their frequencies are directly incorporated into the computations. This nuanced approach weights each interaction by frequency, rewarding models for including common interactions and penalizing them more for omitting them. Specifically, if an interaction with consensus membership $p$ is common to both the consensus and a given model's set, $p$ is added to $TP_{sum}$, and $1-p$ is added to $FP_{sum}$. Otherwise (if the interaction is in the consensus but not the model's set), the value $p$ is added to $FN_{sum}$. With this approach, an uncommon consensus interaction (membership $p < 0.5$), if present in a model, significantly contributes to its $FP_{sum}$. If absent from the model (but present in the consensus), it contributes its small $p$ value to the model's $FN_{sum}$. Conversely, a common consensus interaction (membership $p > 0.5$) tends to increase the similarity score for models that include it and decrease the score for models that lack it. We believe this is a relevant

approach for consensus-based methods, offering a nuanced way to utilize all interaction data from RNA 3D models without requiring a predefined confidence level. The formulas for INF and F1-score remain the same, with the absolute set sizes $|TP|$, $|FP|$, and $|FN|$ being replaced by their respective sums $TP_{sum}$, $FP_{sum}$, and $FN_{sum}$.

## 2.3 Implementation

The RNAtive application consists of multiple components orchestrated using Docker Compose. The frontend is built with the React framework and the Ant Design UI library, while the backend is implemented in Java 17 using the Spring framework. The system also integrates a PostgreSQL database engine and Nginx, the gateway and ingress controller. RNAtive incorporates the RNApdbee 3 adapters container [28], which provides base pair annotation and visualization tools. It also uses the cli2rest-rnapolis, cli2rest-varna-tz and cli2rest-rchie Docker containers for structural unification and customized visualization via VARNA and R-Chie [29, 30]. The web application is served over HTTPS with HAProxy – a high-performance load-balancer – and secured using Let's Encrypt certificates managed by Certbot automation [31].

## 2.4 Performance evaluation

The primary method used to evaluate RNAtive involved comparing it to other state-of-the-art approaches. These methods can be broadly categorized into knowledge-based and machine learning (ML)-based methods. The knowledge-based techniques, which are more numerous, operate on the fundamental principle of calculating the statistics-driven score of a system of atoms and assessing models based on that calculation. The key differentiators among the various systems are how this is accomplished and the factors considered in the evaluation. In contrast, each solution within the ML-based category employs a different architecture to rank RNA 3D models. Further details about these methods can be found in Table 1.

We utilized two RNA 3D structure datasets for our comparison: CASP15 [42, 21] and the randstr decoys set [32]. Both datasets follow a similar principle:

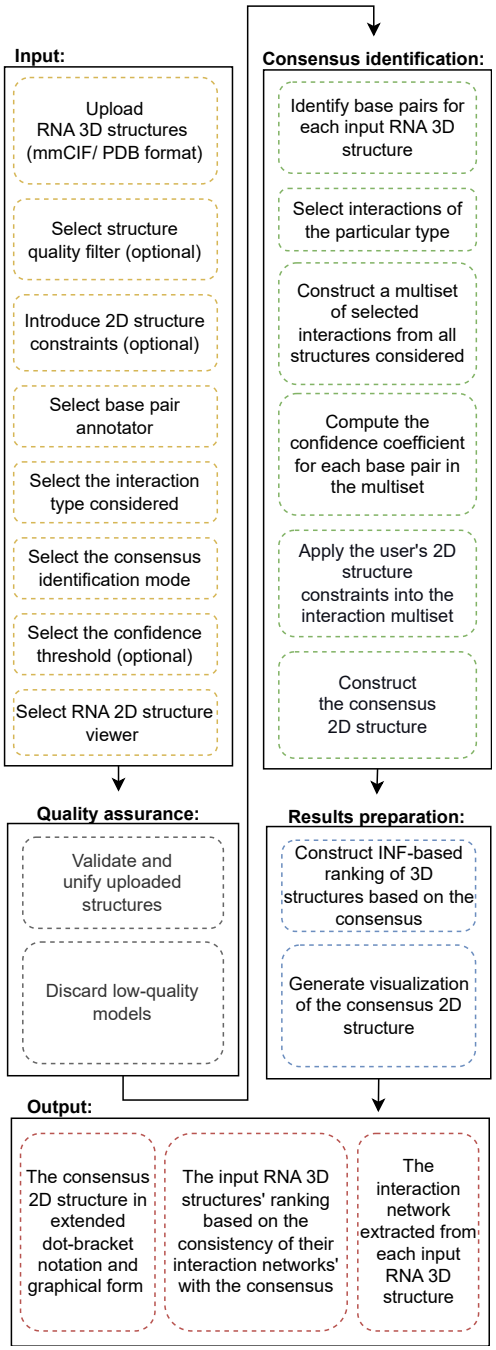


Figure 2: RNAtive workflow.

for each experimentally determined target structure, they provide an ensemble of 3D models. However, the origin of these models differs significantly. The CASP15 set is composed of diverse predictions submitted by independent contestants attempting to model the structure from its sequence. In contrast, the randstr set consists of more uniform decoys generated by computationally perturbing the native target structure. The CASP15 dataset contains 12 targets (1,672 models), while the randstr dataset includes 85 targets (42,585 models).

Each target receives a score, establishing a global ranking of all 3D models. The datasets include pre-computed scores for each model. For randstr files, models are ranked using RMSD-C3', GDT-C3', RMSD-ALL, and GDT-ALL, whereas CASP15 targets are scored with RMSD, lDDT [43], TM-score [44], and GDT-TS [45]. We selected RMSD-ALL for CASP15 decoys from these precomputed values and RMSD for randstr decoys. For additional evaluation, we calculated an INF score for each file [13] using FR3D [24] to extract nucleotide interactions, considering only canonical base pairs, including G-U wobble pairs.

We then employed state-of-the-art RNA 3D model evaluation tools to assess each PDB file for every target. A ranking was established based on each of these assessments. These rankings were then compared with those derived from the RMSD and Interaction Network Fidelity for canonical interactions ($\mathrm{INF_{WC}}$) scores using Kendall's $\tau$ correlation coefficient, Spearman's $\rho$ correlation coefficient, and the Enrichment Score (ES) [46]. This procedure was repeated for RNAtive across multiple configurations to ensure a thorough assessment. Next, we calculated mean correlation coefficients (using Fisher's z-transformation) and mean ES values individually for the CASP15 and the randstr decoys datasets. These values were normalized to the 0–1 range: the sign of correlation coefficients for RMSD-based comparisons was inverted, and ES values originally in the 0–10 range were scaled to the 0–1 range. The methods were ranked by the cumulative value of these three normalized components (i.e., $\rho$, $\tau$, and ES). The results are presented in Section 3 and illustrated in Fig. 3 for the CASP15 dataset and Fig. 4 for the randstr dataset.

# 3 Results and Discussion

We compared rankings generated by RNAtive (using different parameter sets) to official rankings based on RMSD and $\mathrm{INF_{WC}}$. Focusing on only the canonical base pairs of input models and the $\mathrm{INF_{WC}}$ official ranking (Figure 3B), RNAtive emerges as the top contender, achieving correlation coefficients up to 0.92 (Spearman's $\rho$) and 0.78 (Kendall's $\tau$), surpassing all remaining methods by a large margin. The best alternative among state-of-the-art methods, lociPARSE, achieves 0.67 (Spearman's $\rho$) and 0.49 (Kendall's $\tau$). RNAtive's enrichment score (ES) is also impressive in this case, often ranging from 4 to 5 (although the figure displays values normalized to the 0-1 range, the actual values are in the 0-10 range). The sum of normalized scores (ranging from 0 to 3, with 3 being optimal) is above 2 for most RNAtive configurations, while it remains below 1.6 for all other state-of-the-art methods.

Our method remains superior for the CASP15 dataset and RMSD ranking (Figure 3A) even though its correlation coefficients and ES values are lower. This is understandable, as RNAtive is based on nucleotide interactions, whereas RMSD is a metric that captures structural similarity differently. Nevertheless, it is impressive that RNAtive-generated rankings are more similar to the official ones than those from other state-of-the-art methods, some of which (e.g., ARES) are specifically designed to predict RMSD. In contrast, some competing methods achieve mean correlation coefficients below 0.2.

For the randstr decoys (Figure 4), all methods generally achieve higher values; for example, Spearman's $\rho$ values are typically above 0.5. When the $\mathrm{INF_{WC}}$ ranking is considered, the best RNAtive configuration (using the RNApolis annotator with all base-base interactions) achieves a cumulative normalized score of 2.05, while RASP – the top performer in this particular comparison – scores 2.17 (out of a maximum of 3), indicating a small margin. A comparison with the RMSD ranking shows that RNAtive is surpassed by other methods: the cumulative normalized score of its best configuration is 1.91, versus 2.47 for RASP. However, RNAtive's performance is notably stable across datasets. Its Spearman's $\rho$ correlation coefficients,

typically ranging from 0.6 to 0.8 in both datasets, demonstrate its capability to build a consensus structure and rank models against it effectively. While results for the randstr decoys dataset might suggest RNAtive is easily surpassed, the other state-of-the-art methods were heavily trained on this commonly used dataset. We believe their poor performance on the CASP15 dataset highlights that these knowledge-based and ML-based approaches have limitations in generalizing to new datasets, particularly those highly versatile and differentiated.

This analysis highlights RNAtive's main advantages: its robust, interpretable, and highly flexible performance. While many state-of-the-art methods operate as rigid, pre-trained models that struggle to generalize, RNAtive maintains stable performance across diverse datasets. Its unique configurability enhances this robustness, allowing users to guide the analysis by applying quality filters, introducing 2D structure constraints, and fine-tuning the consensus definition with confidence thresholds or conditional weighting – features other methods lack. Crucially, RNAtive is not a black box as it presents the consensus structure it derives, providing direct insight into the recurrent interaction patterns that underpin its high-quality results.

# 4 Conclusions

RNAtive represents a novel approach to scoring and ranking an ensemble of RNA 3D models predicted for the particular RNA molecule. The tool is freely available at `https://rnative.cs.put.poznan.pl` with an intuitive user interface, helpful visualizations, and tabular data, which can be easily downloaded and reused. RNAtive can be customized with its rich set of parameters. The consensus-based approach can be easily extended to 2D input structures. The workflow can also incorporate additional constraints or apply weights to specific inputs to guide the ranking generation based on user expertise. In the future, we plan to experiment with an independent weighting of different classes of functionally crucial non-canonical base pairs. We believe that our tool will be helpful, especially to experimentalists, in selecting native-like RNA 3D structures from a diverse set of *in silico* 3D predictions.

# Data and Code Availability

The source code for RNAtive and its core components is open source and available on GitHub:

- **Main application:** `https://github.com/put-rnative/RNAtive`
- **Bioinformatics tools wrapper:** `https://github.com/tzok/cli2rest-bio`
- **Interaction annotation services:** `https://github.com/rnapdbee/rnapdbee-adapters`

The benchmark datasets, evaluation scripts, and full results discussed in this study are openly available in a Zenodo repository at `https://doi.org/10.5281/zenodo.15773391`, cited as [47]. The state-of-the-art tools used for comparison were run via Docker containers to ensure reproducibility, available from the repository at `https://github.com/put-rnative/rna-tools-dockerized`.

# Competing interests

No competing interest is declared.

# Author contributions statement

MA, MS, and TZ developed the method. JP, MA, and TZ implemented RNAtive and conducted computational experiments. All authors contributed to writing and reviewing the manuscript.

# Acknowledgments

This work is supported by funds from the National Science Center, Poland [2023/51/D/ST6/01207 to TZ] and statutory funds of the Institute of Computing Science, Poznan University of Technology. We thank Bartosz Adamczyk for sharing the Dockerized version of ARES used in our benchmark analysis and Clément Bernard for insightful discussions regarding RNAtive.

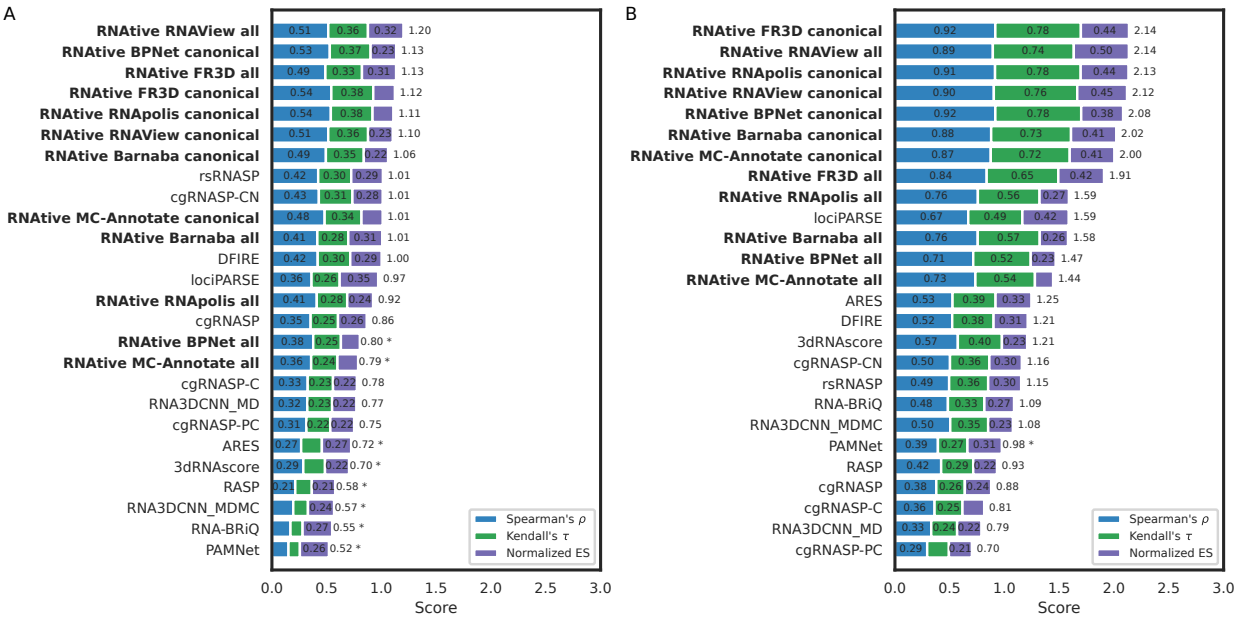


Figure 3: Ranking similarity for the CASP15 dataset based on (A) RMSD and (B) $\mathrm{INF}_{\mathrm{WC}}$. For each scoring method, the corresponding row shows the absolute value of Spearman's $\rho$, Kendall's $\tau$, and the normalized Enrichment Score. Methods are ranked by the sum of these three values. An asterisk (*) denotes that $\geq 20\%$ of the underlying correlation coefficients were not statistically significant (i.e., $\rho$ or $\tau$ close to zero). Each bar is labeled with its corresponding correlation or enrichment value; values $< 0.2$ are omitted for clarity. Results for RNAtive with different parameter settings are shown in bold.

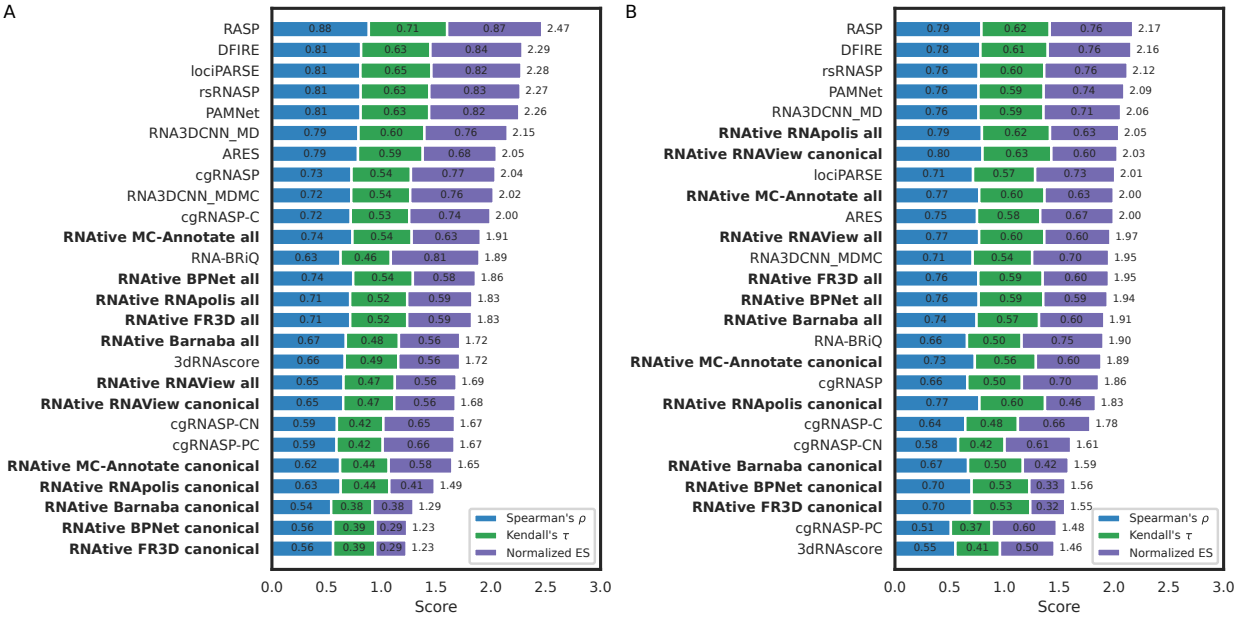


Figure 4: Ranking similarity for the randstr decoys dataset based on (A) RMSD and (B) $\mathrm{INF}_{\mathrm{WC}}$. For each scoring method, the corresponding row shows the absolute value of Spearman's $\rho$, Kendall's $\tau$, and the normalized Enrichment Score. Methods are ranked by the sum of these three values. All correlation coefficients are statistically significant. Each bar is labeled with its corresponding correlation or enrichment value. Results for RNAtive with different parameter settings are shown in bold.

## References


[1] Farzin Haque, Fengmei Pi, Zhengyi Zhao, Shanqing Gu, Haibo Hu, Hang Yu, and Peixuan Guo. RNA versatility, flexibility, and thermostability for practice in RNA nanotechnology and biomedical applications. *WIREs RNA*, 9(1):e1452, 2018.

[2] Renée Schroeder, Andrea Barta, and Katharina Semrad. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol*, 5(11):908–919, 2004.

[3] Quentin Vicens and Jeffrey S. Kieft. Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad Sci USA*, 119(17):e2112677119, 2022.

[4] Kathleen A. Leamy, Sarah M. Assmann, David H. Mathews, and Philip C. Bevilacqua. Bridging the gap between *in vitro* and *in vivo* RNA folding. *Q Rev Biophys*, 49:e10, 2016.

[5] Athma A. Pai and Francesca Luca. Environmental influences on RNA processing: Biochemical, molecular and genetic regulators of cellular response. *WIREs RNA*, 10(1):e1503, 2019.

[6] Philip C. Bevilacqua and Blanton S. Tolbert. Regulatory Mechanisms through RNA Conformational Switching and Dynamics. *J Mol Biol*, 434(18):167794, 2022.

[7] Ivo L. Hofacker. RNA Consensus Structure Prediction With RNAalifold. In John M. Walker and Nicholas H. Bergman, editors, *Comparative Genomics*, volume 395, pages 527–543. 2007.

[8] S. H. Bernhart and I. L. Hofacker. From consensus structure prediction to RNA gene finding. *Brief Funct Genomic Proteomic*, 8(6):461–471, 2009.

[9] Michiaki Hamada. RNA Secondary Structure Prediction from Multi-Aligned Sequences. *Methods Mol Biol*, 1269:17–38, 2015.

[10] Tsukasa Fukunaga and Michiaki Hamada. LinAliFold and CentroidLinAliFold: Fast RNA consensus secondary structure prediction for aligned sequences using beam search methods. *Bioinform Adv*, 2(1):vbac078, 2022.

[11] Marco Pietrosanto, Eugenio Mattei, Manuela Helmer-Citterich, and Fabrizio Ferrè. A novel method for the identification of conserved structural patterns in RNA: From small scale to high-throughput applications. *Nucleic Acids Res*, 44(18):8600–8609, 2016.

[12] Y Zhao and Z Wang. Consensus RNA Secondary Structure Prediction Based on Support Vector Machine Classification. *Chin J Biotechnol*, 24(7):1140–1148, 2008.

[13] M. Parisien, J. A. Cruz, E. Westhof, and F. Major. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA*, 15(10):1875–1885, 2009.

[14] Tomasz Zok, Mariusz Popenda, and Marta Szachniuk. MCQ4Structures to compute similarity of molecule structures. *Cent Eur J Oper Res*, 22(3):457–473, 2014.

[15] Jakub Wiedemann, Tomasz Zok, Maciej Milostan, and Marta Szachniuk. LCS-TA to identify similar fragments in RNA 3D structures. *BMC Bioinformatics*, 18(1):456, 2017.

[16] Marcin Magnus, Maciej Antczak, Tomasz Zok, Jakub Wiedemann, Piotr Lukasiak, Yang Cao, Janusz M Bujnicki, Eric Westhof, Marta Szachniuk, and Zhichao Miao. RNA-Puzzles toolkit: A computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic Acids Res*, 48(2):576–588, 2020.

[17] Marta Mackowiak, Bartosz Adamczyk, Marta Szachniuk, and Tomasz Zok. RNAtango: Analysing and comparing RNA 3D structures via torsional angles. *PLOS Comput Biol*, 20(10):e1012500, 2024.

[18] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, and D. C. Richardson. MolProbity: All-atom

contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, 35(W1):W375–W383, 2007.

[19] Kamil Luwanski, Vladyslav Hlushchenko, Mariusz Popenda, Tomasz Zok, Joanna Sarzynska, Daniil Martsich, Marta Szachniuk, and Maciej Antczak. RNAspider: A webserver to analyze entanglements in RNA 3D structures. *Nucleic Acids Res*, 50(W1):W663–W669, 2022.

[20] Raphael J. L. Townshend, Stephan Eismann, Andrew M. Watkins, Ramya Rangan, Maria Karelina, Rhiju Das, and Ron O. Dror. Geometric deep learning of RNA structure. *Science*, 373(6558):1047–1051, 2021.

[21] Rhiju Das, Rachael C. Kretsch, Adam J. Simpkin, Thomas Mulvaney, Phillip Pham, Ramya Rangan, Fan Bu, Ronan M. Keegan, Maya Topf, Daniel J. Rigden, Zhichao Miao, and Eric Westhof. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins: Struct, Funct, Bioinf*, 91(12):1747–1770, 2023.

[22] Marta Szachniuk. RNApolis: Computational platform for RNA structure analysis. *Fund Computing Decis Sci*, 44:241–257, 2019.

[23] Parthajit Roy and Dhananjay Bhattacharyya. Contact networks in RNA: A structural bioinformatics study with a new tool. *J Comput Aided Mol Des*, 36(2):131–140, 2022.

[24] Michael Sarver, Craig L. Zirbel, Jesse Stombaugh, Ali Mokdad, and Neocles B. Leontis. FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol*, 56(1-2):215–252, 2007.

[25] Patrick Gendron, Sébastien Lemieux, and François Major. Quantitative analysis of nucleic acid three-dimensional structures. *J Mol Biol*, 308(5):919–936, 2001.

[26] Huanwang Yang, Fabrice Jossinet, Neocles Leontis, Li Chen, John Westbrook, Helen Berman, and Eric Westhof. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res*, 31(13):3450–3460, 2003.

[27] Sandro Bottaro, Giovanni Bussi, Giovanni Pinamonti, Sabine Reißer, Wouter Boomsma, and Kresten Lindorff-Larsen. Barnaba: Software for analysis of nucleic acid structures and trajectories. *RNA*, 25(2):219–231, 2019.

[28] Tomasz Zok, Maciej Antczak, Michal Zurkowski, Mariusz Popenda, Jacek Blazewicz, Ryszard W Adamiak, and Marta Szachniuk. RNApdbee 2.0: Multifunctional tool for RNA structure annotation. *Nucleic Acids Res*, 46(W1):W30–W35, 2018.

[29] Kévin Darty, Alain Denise, and Yann Ponty. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15):1974–1975, 2009.

[30] Volodymyr Tsybulskyi, Mohamed Mounir, and Irmtraud M Meyer. R-chie: A web server and R package for visualizing cis and trans RNA–RNA, RNA–DNA and DNA–DNA interactions. *Nucleic Acids Res*, 48(18):e105–e105, 2020.

[31] Josh Aas, Richard Barnes, Benton Case, Zakir Durumeric, Peter Eckersley, Alan Flores-López, J. Alex Halderman, Jacob Hoffman-Andrews, James Kasten, Eric Rescorla, Seth Schoen, and Brad Warren. Let's Encrypt: An Automated Certificate Authority to Encrypt the Entire Web. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 2473–2487, London United Kingdom, 2019. ACM.

[32] Emidio Capriotti, Tomas Norambuena, Marc A. Marti-Renom, and Francisco Melo. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics*, 27(8):1086–1093, 2011.

[33] Jian Wang, Yunjie Zhao, Chunyan Zhu, and Yi Xiao. 3dRNAscore: A distance and torsion angle dependent evaluation function of 3D RNA

structures. *Nucleic Acids Res*, 43(10):e63–e63, 2015.

[34] 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. *J Comput Biol*, 27(6):856–867, 2020.

[35] Peng Xiong, Ruibo Wu, Jian Zhan, and Yaoqi Zhou. Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement. *Nat Commun*, 12(1):2777, 2021.

[36] 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. *Biophys J*, 121(1):142–156, 2022.

[37] Ling Song, Shixiong Yu, Xunxun Wang, Ya-Lan Tan, and Zhi-Jie Tan. cgRNASP-CN: A minimal coarse-grained representation-based statistical potential for RNA 3D structure evaluation. *Commun Theor Phys*, 74(7):075602, 2022.

[38] 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 Genom Bioinform*, 5(1):lqad016, 2023.

[39] 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 Comput Biol*, 14(11):e1006514, 2018.

[40] Shuo Zhang, Yang Liu, and Lei Xie. A universal framework for accurate and efficient geometric deep learning of molecular systems. *Sci Rep*, 13(1):19171, 2023.

[41] Sumit Tarafder and Debswapna Bhattacharya. lociPARSE: A Locality-aware Invariant Point Attention Model for Scoring RNA 3D Structures. *J Chem Inf Model*, 64(22):8655–8664, 2024.

[42] Andriy Kryshtafovych, Maciej Antczak, Marta Szachniuk, Tomasz Zok, Rachael C. Kretsch, Ramya Rangan, Phillip Pham, Rhiju Das, Xavier Robin, Gabriel Studer, Janani Durairaj, Jerome Eberhardt, Aaron Sweeney, Maya Topf, Torsten Schwede, Krzysztof Fidelis, and John Moult. New prediction categories in CASP15. *Proteins: Struct, Funct, Bioinf*, 91(12):1550–1557, 2023.

[43] Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

[44] Sha Gong, Chengxin Zhang, and Yang Zhang. RNA-align: Quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA. *Bioinformatics*, 35(21):4459–4461, 2019.

[45] A. Zemla. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res*, 31(13):3370–3374, 2003.

[46] Jerry Tsai, Richard Bonneau, Alexandre V. Morozov, Brian Kuhlman, Carol A. Rohl, and David Baker. An improved protein decoy set for testing energy functions for protein structure prediction. *Proteins: Struct, Funct, Bioinf*, 53(1):76–87, 2003.

[47] Jan Pielesiak, Maciej Antczak, Marta Szachniuk, and Tomasz Zok. Benchmark Data for: RNAtive to recognize native-like structure in a set of RNA 3D models. Zenodo, June 2025. Version 1.0.0.

Table 1: State-of-the-art approaches for RNA 3D model quality assessment.

| Method | Reference | Description |
|---|---|---|
| **Knowledge-based approaches** | | |
| RASP | [32] | An all-atom statistical potential to rank models. |
| 3dRNAscore | [33] | Distance- and dihedral-dependent energies to evaluate models. |
| DFIRE | [34] | All-atom distance-dependent energy function based on a reference state (distance-scaled finite ideal-gas reference state). |
| BRiQ | [35] | Statistical energy function with a nucleobase-centric sampling algorithm, capturing full-orientation dependence of base-base, oxygen-base, and oxygen-oxygen interactions with the RNA backbone. Quantum-mechanical calculations ensure proper weights to minimize the possible effects of indirect interaction. |
| rsRNASP | [36] | All-atom distance-dependent statistical potential based on residue separation. |
| cgRNASP, cgRNASP-C, cgRNASP-CN, cgRNASP-PC | [37, 38] | Direct successor of rsRNASP, uses both long- and short-range interactions by residue separation. |
| **Machine learning-based approaches** | | |
| RNA3DCNN_MD, RNA3DCNN_MDMC | [39] | Uses a Convolutional Neural Network to analyze every nucleotide individually with a 3D grid representation without manual structure extraction. |
| ARES | [20] | A deep neural network with an architecture formulated for direct 3D structure input that predicts the model's RMSD using the spatial position of atoms and their chemical element types. |
| PAMNet | [40] | The Physics-Aware Multiplex Graph Neural Network that induces a physics-informed bias that allows for modeling of local and non-local interactions of molecules based on different geometric information. |
| lociPARSE | [41] | A Locality-aware invariant point attention architecture that estimates the Local Distance Difference Test (lDDT) scores to assess the accuracy of each nucleotide and the atomic environment that surrounds it in a superposition-free manner. |