

<https://doi.org/10.1038/s42003-026-09582-2>

Quality assessment of RNA 3D structure models using deep learning and intermediate 2D maps

Check for updates

Xiaocheng Liu^{1,4}, Wenkai Wang^{1,4}✉, Zongyang Du^{2,3}, Ziyi Wang¹, Zhenling Peng¹ & Jianyi Yang¹✉

Accurate quality assessment is critical for computational prediction and design of RNA three-dimensional (3D) structures, yet it remains a significant challenge. In this work, we introduce RNArank, a deep learning-based approach to both local and global quality assessment of predicted RNA 3D structure models. For a given structure model, RNArank extracts a comprehensive set of multi-modal features and processes them with a Y-shaped residual neural network. This network is trained to predict two intermediate 2D maps, including the inter-nucleotide contact map and the distance deviation map. These maps are then used to estimate the local and global accuracy. Extensive benchmark tests indicate that RNArank consistently outperforms traditional methods and other deep learning-based methods. Moreover, RNArank demonstrates promising performance in identifying high-quality structure models for targets from the recent CASP15 and CASP16 experiments. We anticipate that RNArank will serve as a valuable tool for the RNA biology community, improving the reliability of RNA structure modeling and thereby contributing to a deeper understanding of RNA function.

RNA (ribonucleic acid) plays an essential role in gene expression regulation and numerous cellular processes. Its biological function largely depends on its three-dimensional (3D) structure. Unlike proteins, RNA molecules are more structurally flexible and exhibit a higher degree of conformational diversity. The complexity of RNA structure is increased by the weak and long-range interactions that govern RNA folding, making experimental determination of RNA structure more difficult than protein structure¹.

The inherent challenges of experimental RNA structure determination have long motivated the development of computational RNA structure prediction approaches^{2–11}. In recent years, deep learning-based approaches, in particular, have shown promise in enhancing the accuracy of automated predictions^{12–20}. Nevertheless, unlike the protein field, which has been revolutionized by AlphaFold2²¹, RNA structure prediction currently lacks a comparably high-confidence solution²². Community-wide assessments like CASP and RNA-Puzzles consistently show that while automated servers demonstrate competitive results on specific targets, they still generally lag behind human expert groups^{23–25}. Consequently, a hybrid approach combining human intervention with a diverse mix of modeling methods

remains essential for modeling the most challenging targets²⁵. This context highlights a critical need for RNA structure quality assessment (QA) to rank models generated by both computational and human-led efforts.

Existing QA methods can be broadly classified into two categories: knowledge-based statistical potentials^{26–28} and deep learning-based approaches^{29–31}. Knowledge-based methods, such as cgRNASP²⁶, rsRNASP²⁷, and DFIRE-RNA²⁸, derive scoring potentials from simulated reference states at either all-atom or coarse-grained levels. However, the accuracy of these methods is often constrained by the incomplete understanding of the principles governing RNA energetic stability, which can lead to the use of inaccurate potentials and reference states that diverge from native conformations.

To address the shortcomings of knowledge-based potentials, deep learning-based methods were introduced, leveraging supervised learning to directly predict the accuracy of RNA structure models^{29–31}. Early deep learning approaches like RNA3DCNN³⁰ employed 3D convolutional neural network on a voxelized representation of the 3D structure, while ARES²⁹ utilized an Equivariant Graph Neural Network (EGNN). However, the performance of these methods was often hampered by their reliance on

¹MOE Frontiers Science Center for Nonlinear Expectations, State Key Laboratory of Cryptography and Digital Economy Security, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China. ²Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China. ³Department of Biomedical Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. ⁴These authors contributed equally: Xiaocheng Liu, Wenkai Wang. ✉e-mail: wenkaiwang@sdu.edu.cn; yangjy@sdu.edu.cn

limited training models (e.g., from molecular dynamics simulations or Rosetta fragment assembly) and the use of RMSD as a supervision signal. The superposition- and length-dependent nature of RMSD increased the difficulty of learning and limited their generalization. More recently, lociPARSE³¹, inspired by AlphaFold²¹, used a modified Invariant Point Attention (IPA) module to directly predict the IDDT score³².

In this work, we present RNArank, a deep learning-based approach to both local and global QA. From a multi-modal description of the input structure, RNArank employs a Y-shaped stack of residual neural network (ResNet³³) blocks to predict two intermediate 2D maps, which are then used to derive the local and global accuracy in terms of IDDT. Benchmark tests demonstrate that RNArank excels at both assessing overall structure quality and selecting high-quality models.

Results

Overview of RNArank

The architecture of the RNArank is shown in Fig. 1. For a given RNA structure model, RNArank extracts a comprehensive set of multi-modal features, including 1D features (nucleotide sequence, backbone orientations, ultrafast shape recognition (USR)³⁴), 2D features (inter-nucleotide distances, Rosetta two-body energies, clash probabilities), and 3D features (nucleotide-level voxelization). A detailed description of these features is provided in Supplementary Table 1. These features are processed by a multi-

modal encoder to obtain a unified representation (Fig. 1b). Inspired by protein structure assessment methods^{35,36}, rather than predicting a single scalar value, we utilize a Y-shaped residual neural network (Y-ResNet) to predict two intermediate 2D maps, i.e., a contact map and a distance deviation map, which are the core components of the IDDT calculation. These maps are then used to predict the per-residue and global IDDT score (denoted by pIDDT). A detailed description of the model architecture and training procedures can be found in the “Methods” section.

Furthermore, IDDT was originally defined for protein structure, which is not necessarily appropriate for measuring RNA structure quality. To better reflect the structural properties of RNA, we adapted the standard IDDT (denoted as IDDT_{TAW}) formulation by using larger distance thresholds (see “Methods”, Eq. (1)). This RNA-adapted IDDT (denoted as IDDT_{RNA}) shows a significantly improved correlation with three other well-established global metrics (RMSD, TM-score_{RNA}, and GDT_TS, see Supplementary Figs. 1 and 2). Moreover, IDDT_{RNA} offers superior discriminatory power for near-native models and effectively detects local structural inaccuracies (Supplementary Fig. 3). These features establish it as a robust metric capable of providing both global, molecule-wise and local, residue-wise assessments. A more detailed discussion of the rationale for IDDT_{RNA} can be found in the “Methods” section. Unless otherwise noted, RMSD (calculated via the program RNAalign³⁷) and IDDT_{RNA} (Eqs. (3) and (4)) values are both based on C4' atoms.

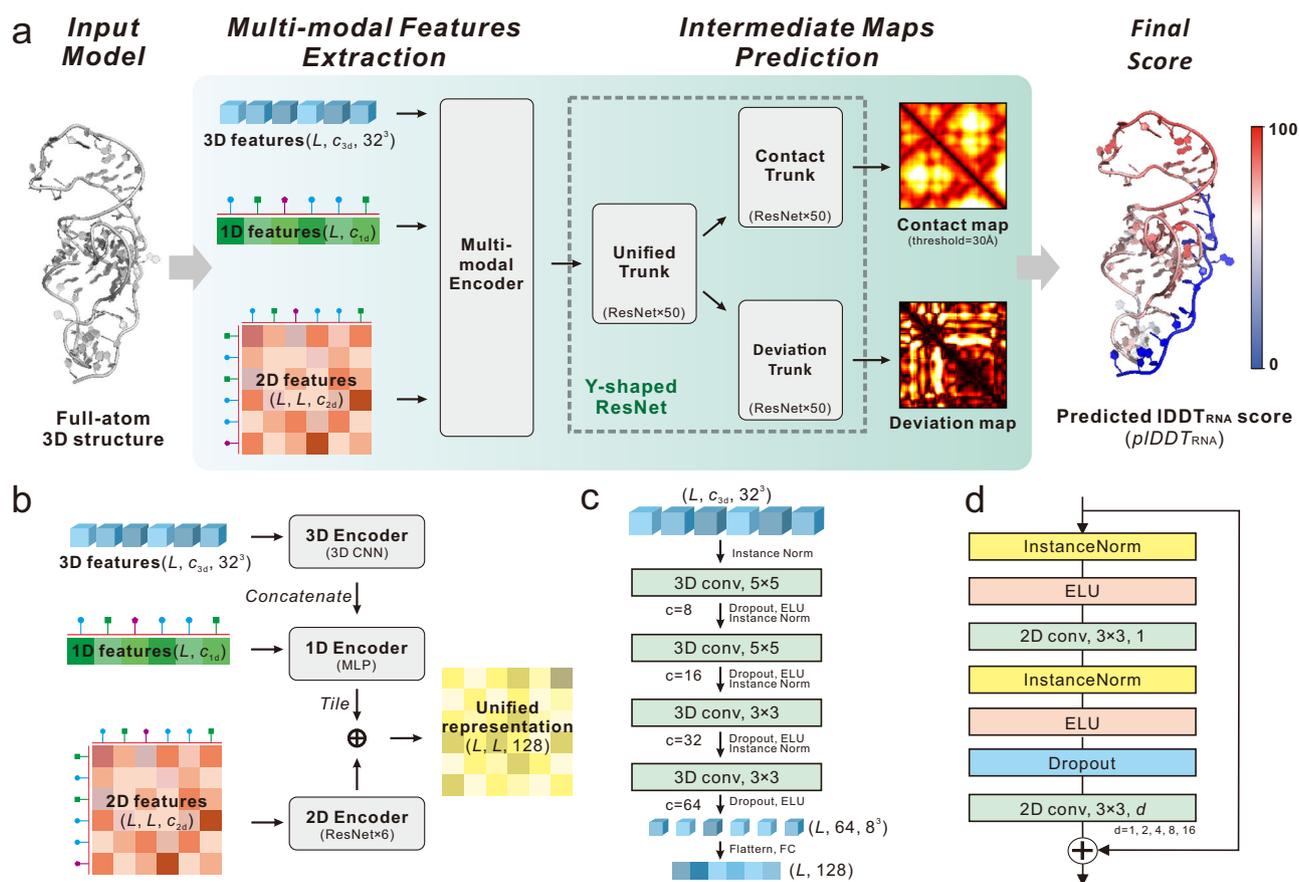


Fig. 1 | Overall architecture of RNArank. **a** The quality assessment pipeline of RNArank. An input RNA structure model is first processed to extract multi-modal features (1D, 2D, and 3D; with dimensions of $c_{1d} = 67$, $c_{2d} = 58$, and $c_{3d} = 3$, respectively). After encoding and aggregating, these features are fed into a Y-shaped ResNet that predicts two intermediate outputs (a contact map and a deviation map) to derive the final per-nucleotide pIDDT score. The global pIDDT score is calculated by averaging the scores over all nucleotides. **b** The multi-modal encoder. The 3D features are first processed by a 3D CNN block, after which they are flattened and concatenated with the 1D features, then updated by an MLP. In parallel, the 2D

features are processed by six ResNet blocks. Finally, the updated 1D representation is tiled to a 2D pairwise map and combined with the processed 2D features via element-wise addition to obtain a unified 2D representation. **c** The 3D CNN block. The 3D features are processed by a stack of 3D convolutional layers (where c denotes the number of output channels), flattened, and then mapped to a 1D representation via a fully connected (FC) layer. **d** The ResNet block used in RNArank. We employ a “full pre-activation” architecture⁴⁹ and incorporate dilated convolutions (with a variable rate, d) to enhance performance.

Performance on 24 independent RNA targets

To evaluate the performance of RNArank, we collected a non-redundant dataset of 24 RNAs from PDB (named T24), all released after the date of the training RNAs (January 2022). Consistent with our training set preparation (see “Methods”), for each RNA in T24, we generated a comprehensive decoy set using a combination of structure prediction methods and in-house decoy generation protocols. The former included state-of-the-art deep learning-based, physics-based, and knowledge-based approaches^{8,9,12–17,38}, while the latter encompassed protocols such as structure perturbation, molecular dynamics simulations, and Rosetta-based fragment assembly (see “Methods” and Supplementary Table 2).

We compare RNArank with five representative QA methods, including three traditional knowledge-based methods (cgRNASP²⁶, rsRNASP²⁷, DFIRE²⁸), and two deep learning-based methods (RNA3DCNN³⁰, lociPARSE³¹). The knowledge-based methods provide global QA scores only, while the deep learning-based methods can give both global and local QA scores. Note that lociPARSE is designed to predict IDDT_{raw}.

Multiple metrics are used for performance evaluation, including Spearman correlation (ρ), accuracy of top-ranked model (IDDT_{RNA} and RMSD), and receiver operating characteristic (ROC). The Spearman

correlation ($\rho_{\text{IDDT}(\text{Global})}$, $\rho_{\text{IDDT}(\text{Local})}$, and ρ_{RMSD}) is defined between the predicted QA scores and the true values at both global (molecule; for IDDT_{RNA} and RMSD) and local (nucleotide; only for IDDT_{RNA}) levels. These correlations are analyzed at two different levels: on an *overall* basis, by pooling all decoys together; and on a *per-target* basis, by first calculating the correlation for each target and then averaging the results. The accuracy of the top-ranked model (IDDT_{RNA} and RMSD) is defined on the *per-target* basis; while the ROC and the area under the curve (AUC) are calculated on the *overall* basis, to distinguish high-quality models from the rest (see “Methods” section).

As shown in Supplementary Table 3, RNArank consistently outperforms existing methods across all evaluation metrics. Notably, despite being trained solely on IDDT_{RNA}, RNArank generalizes robustly to RMSD-based evaluations (Fig. 2a). Regarding global molecule-wise assessment, it achieves the highest overall $\rho_{\text{IDDT}(\text{Global})}$ (0.949 vs. 0.423 for the second-best method) and ρ_{RMSD} (0.832 vs. 0.237 for the second-best). This lead is maintained on a *per-target* basis for both IDDT_{RNA} (0.936 vs. 0.675 for the second-best) and RMSD (0.821 vs. 0.550 for the second-best). This strong performance extends to local residue-wise assessment, where RNArank’s overall (0.879) and per-target average (0.865) correlations significantly exceed the other

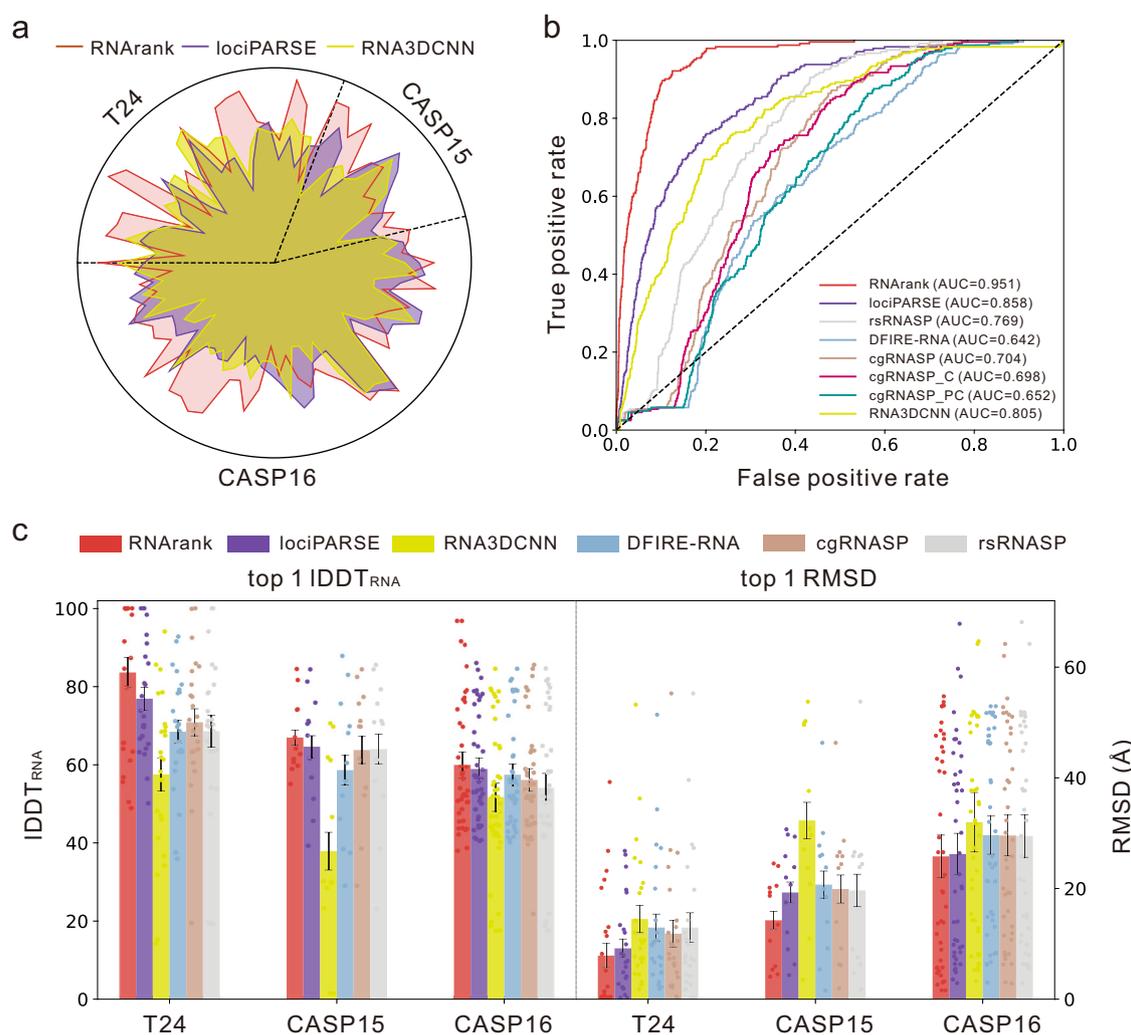


Fig. 2 | Performance of RNArank and comparison with other methods.

a Comparison of the RMSD Spearman correlation (ρ_{RMSD}) against lociPARSE (IDDT-based) and RNA3DCNN (RMSD-based) for all tested targets ($n = 78$ RNAs). The radial axis is normalized such that a point further from the center always indicates a better score. **b** ROC curves of predicted scores for structure models from the T24 dataset ($n = 2837$ decoys). The threshold for the positive class is set at a true

IDDT_{RNA} score of >75. Note that the scores from knowledge-based methods and RNA3DCNN were inverted before calculating the ROC curves. **c** Comparisons of the top 1 IDDT_{RNA} and RMSD on the three test sets. Error bars represent 20% of the standard deviation ($n = 24$ for T24, 12 for CASP15, and 42 for CASP16). Each data point corresponds to an individual target.

methods (0.314 and 0.334 for the second-best, respectively), with the per-target improvement being statistically significant ($P = 3.7 \times 10^{-26}$, two-sided *t*-test). These results demonstrate the state-of-the-art performance of RNArank for both global and local QA.

A crucial aspect of QA is the ability to identify high-quality models. We assess this capability in two ways. First, in the ROC analysis to distinguish high-quality models ($\text{IDDT}_{\text{RNA}} > 75$ for positives; Fig. 2b and Supplementary Table 3), RNArank achieves a leading AUC of 0.951, substantially higher than other methods (0.858 for the second best). Second, we further examine the capacity to identify the best model (i.e., top 1 selection) for each target. As shown in Supplementary Table 3 and Fig. 2c, RNArank yields the average top 1 IDDT_{RNA} of 83.6 and RMSD of 7.9 Å, better than the second-

best method (76.9 and 9.2 Å, respectively). These results demonstrate RNArank's strengths in model selection.

Figure 3 highlights the reliable quality assessment by RNArank on a representative example (PDB ID: 7TZS; Fig. 3a). On this example, the scores predicted by RNArank correlate strongly ($\rho > 0.9$) with the true IDDT_{RNA} values at both the global (Fig. 3b) and local levels (Fig. 3c). This high accuracy is consistently observed across a diverse set of models with varying quality, generated by methods including AlphaFold3, trRosettaRNA, and SimRNA (Fig. 3d–f). In summary, RNArank achieves satisfactory performance in RNA 3D structure QA on the T24 benchmark. This performance marks the advancement in RNA model evaluation made by the RNArank framework.

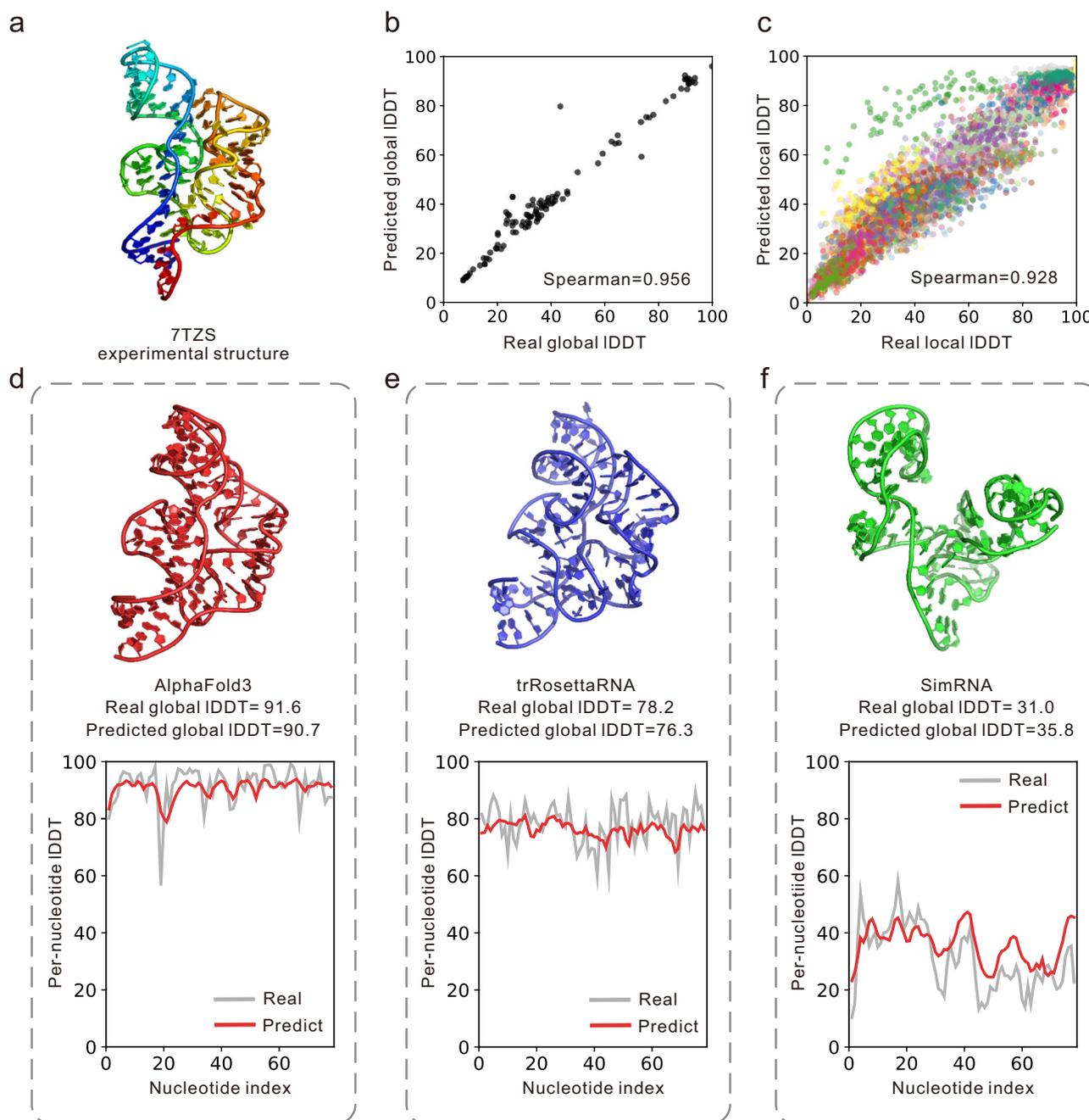


Fig. 3 | Quality assessment by RNArank for an example target (PDB ID: 7TZS). a Experimental structure of 7TZS. Correlation between predicted and true IDDT_{RNA} scores at the **b** global and **c** per-nucleotide (local) levels. Each point in **c** corresponds

to a nucleotide and all nucleotides from the same model are shown with the same color. **d–f** The local IDDT_{RNA} curves for three representative structure models from AlphaFold3, trRosettaRNA, and SimRNA.

Performance on CASP15 RNAs

The CASP15 in 2022 introduced the RNA structure prediction track, reflecting growing interest in RNA structure modeling. The 12 RNA targets in CASP15 (8 natural, 4 synthetic) presented significantly greater challenges compared to the T24 set, evidenced by a higher and wider RMSD range (Supplementary Fig. 4). The corresponding 1622 decoys were generated by a diverse range of structure prediction methods, from new deep learning techniques like trRosettaRNA¹⁴ (Yang-Server), RhoFold¹⁷ (AIchemy_RNA), RoseTTAFoldNA¹⁵ (BAKER), and DeepFoldRNA¹² (DF_RNA) to traditional approaches such as BRIQ³⁹ (AIchemy_RNA2), Vfold¹¹ (Chen), RNAComposer³⁸ (RNAPolis), and SimRNA⁹ (GeneSilico), as well as the interventions from human experts. This methodological diversity ensures a more comprehensive evaluation of RNArank's performance. Furthermore, these targets were released after the date of training RNAs and are non-identical with our training dataset (see "Methods"), making them a rigorous benchmark for RNArank's performance.

As shown in Supplementary Table 4, RNArank consistently outperforms other methods across 8 of the 9 evaluated metrics in the CASP15 test dataset. On the *overall* basis, it achieves Spearman correlation of 0.706 (with global IDDT_{RNA}) and 0.686 (with RMSD), significantly outperforming the second-best method (0.511 and 0.534, respectively). RNArank also achieves a substantially higher AUC of 0.915 compared to others (<0.7), indicating a robust capacity for distinguishing high-quality decoys (Supplementary Fig. 5a). This high performance extended to the local nucleotide-level assessment (local IDDT_{RNA}), where RNArank reaches the highest Spearman correlation among all the methods.

On the *per-target* basis, RNArank still outperforms other methods on the RMSD and local IDDT_{RNA} correlation; however, its global IDDT_{RNA} correlation is slightly lower than lociPARSE, which is primarily due to the inaccurate assessment for four targets: R1149, R1156, R1189, and R1190. Among these challenging targets, R1149 and R1156 are known to be highly dynamic and adopt various conformations. The other two targets, R1189 and R1190, are involved in the complicated interaction with more than 4 protein chains, which may distort their assessments as the sole RNA. Analysis of these four targets (Supplementary Fig. 6) reveals that RNArank tends to give high scores for the models submitted by TS081 (RNAPolis group in CASP15, marked by red stars in Supplementary Fig. 6), which uses RNAComposer, a method designed for RNA-only modeling that performed imperfectly on these four targets. This suggests that RNArank is challenged by the conformational diversity of dynamic RNAs (R1149, R1156), and may also favor unbound conformations over native bound states for targets like R1189 and R1190.

Nevertheless, when focusing on the critical task of identifying the single best decoy (Fig. 2c), RNArank again achieves the highest average top 1 IDDT_{RNA} (67.0) and RMSD (14.3 Å), outperforming the second-best method (64.6 and 19.3 Å, respectively). These findings confirm that despite the challenges posed by complex or dynamic targets, RNArank's robust scoring mechanism makes it effective at selecting the high-confidence model from a large and diverse pool of models.

Performance on CASP16 RNAs

The recent CASP16 competition provided a new and richer RNA benchmark set containing 42 targets and almost 6500 models, allowing for further rigorous testing of RNArank. This set features a wide range of target types (both monomers and multimers) and complexity levels (58–833 nucleotides). With decoys submitted by 48 human and 16 server groups, the CASP16 set displays even greater diversity than previous benchmarks (Supplementary Fig. 4). As shown in Fig. 2c and Supplementary Table 5, RNArank continues to demonstrate strong performance across various metrics (7 of 9) on the CASP16 dataset, further highlighting its robustness.

Notably, RNArank excels in identifying an ensemble of high-quality decoys. This is a particularly relevant capability, as blind assessment experiments like CASP and RNA-Puzzles require participants to submit 5–10 models for each target. To rigorously evaluate this capability, we introduce a top-*k* success rate analysis (Supplementary Fig. 7a), where a

"success" is defined as identifying at least one of the 5 best decoys (measured by the ground-truth IDDT_{RNA}) within the top-*k* ranked models. On the challenging CASP16 set, RNArank (red line) achieves success rates of 31.0% for *k* = 5 and 42.9% for *k* = 10, outperforming the next-best method, lociPARSE (green line; 21.4% and 35.7%, respectively). This performance lead is even more pronounced on the T24 benchmark, while RNArank remains highly competitive on the smaller, compositionally biased CASP15 set. This result suggests the potential of RNArank for application in blind test experiments such as CASP and RNA-Puzzles. In fact, its integration into our automated server, Yang-Server, contributed to our group's top performance in the automated RNA structure prediction track of the CASP16 blind test⁴⁰.

However, we found that RNArank's overall Spearman correlation for RMSD ($\rho_{\text{RMSD}} = 0.722$) was slightly below that of the energy-based method DFIRE-RNA (0.751), a trend not observed in the previous two test sets. This performance gap stems from two factors: (1) RMSD's intrinsic strong correlation with sequence length, and (2) the different sensitivities of scoring methods to sequence length. Firstly, RMSD values naturally increase for larger structures. E.g., for CASP16 targets, RMSD shows a clear correlation with sequence length (Supplementary Fig. 8a; Spearman correlation: 0.835). Similarly, energy-based methods (such as DFIRE-RNA, rsRNASP, and cgRNASP) display strong negative correlations between their scores and sequence length (Supplementary Fig. 8d–f; Spearman correlations: -0.979 , -0.944 , -0.953 , respectively), which is consistent with the RMSD-length relationship. In contrast, the IDDT score predicted by RNArank is less influenced by sequence length, with a lower Spearman correlation of 0.606 (Supplementary Fig. 8b). This difference in length dependency influences the *overall* evaluation, which mixes all targets of different lengths together for the correlation calculation. The significantly greater target length variation in CASP16 (357.7 ± 264.8) than the other two sets (T24: 106.5 ± 100.3 ; CASP15: 208.5 ± 186.5) further magnified this effect, explaining the distinct trend observed for RMSD correlation in the CASP16 set. Nevertheless, when evaluated on the more rigorous *per-target* basis, RNArank maintains its leading performance in terms of ρ_{RMSD} , though no method, whether trained on IDDT or RMSD, achieves a value > 0.4.

Generalization analysis and ablation study

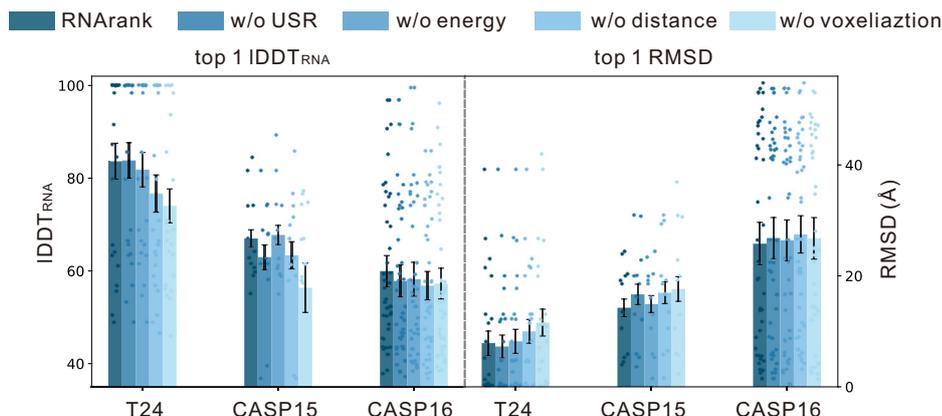
To rigorously assess the generalization capabilities of RNArank, we conducted evaluations on strictly non-redundant subsets of the three benchmarks. We constructed these subsets by excluding all targets with over 80% sequence identity to any entry in our training data, which yielded final test sets of 17 (T24), 10 (CASP15), and 28 (CASP16) targets, respectively. As detailed in Supplementary Tables 6–8, RNArank's state-of-the-art performance holds even on these sequence-dissimilar targets. This strongly indicates that the model has learned fundamental principles of RNA structure quality rather than merely memorizing features specific to the training data.

To elucidate the factors contributing to RNArank's performance, we conduct an ablation study to evaluate the impact of four key features: ultrafast shape recognition (USR), voxelization, Rosetta two-body energy, and inter-nucleotide distances. For each feature, we removed it from the input feature sets and re-trained an ablated model using the same training dataset and configuration. The performance of these ablated models was benchmarked against the fully-featured RNArank on the T24 dataset, with results summarized in Fig. 4 and Supplementary Fig. 7b, Supplementary Tables 9–11.

As anticipated, the voxelization feature proved to be the most critical component. Its removal caused the most significant performance degradation across all three datasets and nearly all metrics (Fig. 4, and Supplementary Tables 9–11). For instance, on the challenging CASP16 test set, removing this feature decreased the per-target average $\rho_{\text{IDDT}(\text{Global})}$ from 0.549 to 0.323, $\rho_{\text{IDDT}(\text{Local})}$ from 0.421 to 0.316 (Supplementary Table 11). Removing the distance feature also consistently decreased performance. The top-*k* success rate analysis (Supplementary Fig. 7b) provides further evidence for the crucial roles of both the voxelization and distance features in achieving robust model selection.

Fig. 4 | Ablation study on the three test sets.

Comparison of the per-target average top 1 $\text{IDDDT}_{\text{RNA}}$ and RMSD between the full RNArank model and its ablated variants on T24 ($n = 24$), CASP15 ($n = 12$), and CASP16 ($n = 42$). Error bars represent 20% of the standard deviations. Each data point corresponds to an individual target.



The contributions of the other feature modalities (energy and USR) were more nuanced. For example, removing USR paradoxically improved the Top-1 selection on the T24 set. However, a more robust metric, the Top- k success rate, reveals the true picture: removing USR actually decreased performance across most of the k -range, confirming its overall positive contribution (Supplementary Fig. 7b). A similar complex effect was observed for the energy features, which appeared to have a negligible impact on the T24 and CASP15 sets but led to a noticeable performance drop when removed from the CASP16 set (Supplementary Fig. 7b).

Overall, the full RNArank model achieved more robust performance than all ablated variants. This validates our multi-modal feature strategy and underscores its importance for reliable structure quality assessment.

Comparison with the AlphaFold3 self-assessment module

Given the recent release and widespread adoption of AlphaFold3 (AF3), a powerful structure prediction tool with a built-in self-assessment module, we conducted a direct comparison on the CASP16 dataset to evaluate RNArank's performance on the AF3-generated decoys. Our analysis of the decoys submitted by the AF3-server (group TS304) confirms that AF3's self-assessment (pIDDT) is exceptionally accurate for its own predictions, achieving Spearman correlations of 0.905 with $\text{IDDDT}_{\text{RNA}}$ and -0.896 with RMSD (Supplementary Fig. 9). On this same set of decoys, RNArank also demonstrates strong and competitive performance, with correlations of 0.799 and -0.832 , respectively. This performance significantly outperforms other methods like lociPARSE (correlations of 0.458 and -0.511 , respectively).

However, this comparison highlights a fundamental and critical distinction in the purpose and applicability of these methods. The outstanding performance of AF3's pIDDT likely stems from its ability to leverage the network's internal states and co-evolutionary signals generated during the prediction process, making it a powerful internal confidence metric. This tight integration is both its primary strength and its main limitation: it cannot function as a stand-alone tool to assess models from any other source (e.g., alternative predictors, physics-based simulations, or human-expert models).

In contrast, RNArank is designed to serve as a universal, method-agnostic quality assessment tool. It achieves this by operating on the final 3D structure alone, agnostic to its origin, using a rich set of multi-modal features to detect universal structural inaccuracies. This capability is indispensable for community-wide blind assessments like CASP and for any research that requires an unbiased comparison across diverse modeling techniques. Thus, while AF3's self-assessment excels for its own predictions, RNArank provides the essential universal framework for objectively comparing models from the entire ecosystem of structure prediction.

Analysis of the computational efficiency

To confirm RNArank's suitability for high-throughput applications, we benchmarked its computational performance on the CASP16 dataset,

which contains targets ranging from ~ 50 to >800 nucleotides. As detailed in Supplementary Table 12, RNArank is highly efficient, capable of scoring a typical 100–300-nucleotide structure in under 6 seconds on average. This total runtime is dominated by the CPU-based feature extraction step, which accounts for over 90% of the processing time, while the GPU-based model inference is completed in seconds even for the largest RNAs (~ 850 nt). Memory consumption is also modest, with peak usage remaining well within the capacity of standard hardware (<3 GB on CPU and <5 GB on GPU).

Overall, these results demonstrate that RNArank is computationally efficient and well-suited for high-throughput applications. Future optimizations could focus on accelerating the feature extraction process, such as by simplifying the voxelization protocol or implementing GPU-based calculations for Rosetta energy terms.

Discussion

RNA 3D structure prediction has drawn increased interest in recent years, and the reliable quality assessment plays a critical role in improving structure prediction accuracy. To achieve this, we have developed an automated approach for RNA 3D structure quality assessment: RNArank, which utilizes comprehensive multi-modal features and a Y-shaped ResNet to estimate the IDDT of a predicted structure model. We rigorously benchmarked RNArank on three datasets, including 24 independent RNAs and two sets from CASP15 and CASP16. The results show that RNArank consistently achieves promising performance in selecting decoys and outperforms other approaches, including both the traditional energy-based methods and the recent deep learning-based methods.

However, benchmarking on CASP datasets demonstrates that it remains challenging to assess models for RNAs that form complicated multimers and/or interact with proteins. Even with promising methods for assessing docking results like DRPScore⁴¹, fully accounting for the conformational changes induced by protein binding remains a significant challenge that may hinder the accurate evaluation of complexes from de novo predictors like AlphaFold3. Given RNA's natural tendency to rely on intermolecular interactions for structural stability, the next-generation RNA QA methods must evolve to effectively incorporate information from binding partners and consider the dynamics of complex formation. In addition, considering RNA dynamics will be crucial for a comprehensive understanding of their conformational landscapes, including folding pathways—an area of increasing focus in the protein field⁴².

Methods

IDDT_{RNA} definition

The definition of IDDT in this work ($\text{IDDDT}_{\text{RNA}}$) introduces two critical modifications to the original IDDT ($\text{IDDDT}_{\text{raw}}$) that was designed for proteins. First, we adjusted the set of deviation cutoffs. While $\text{IDDDT}_{\text{raw}}$ uses thresholds of $\{0.5 \text{ \AA}, 1.0 \text{ \AA}, 2.0 \text{ \AA}, 4.0 \text{ \AA}\}$, $\text{IDDDT}_{\text{RNA}}$ employs a shifted set of

{1.0 Å, 2.0 Å, 4.0 Å, 6.0 Å}. This change accounts for the larger distance scale inherent to RNA structures. Second, we extended the inclusion radius for contact definition from the standard 15 Å (used in IDDT_{raw}) to 30 Å. This expansion is crucial to accommodate the larger size and more extended geometry of nucleotides; for instance, the C4'-C4' distance between base-paired nucleotides can exceed the standard 15 Å cutoff. Specifically, the IDDT_{RNA} score is formulated as:

$$\text{IDDT}_{\text{RNA}} = \frac{100}{4L} \sum_{\epsilon \in \{1,2,4,6\}} \sum_{i=1}^L \frac{\sum_{j: D_{ij} < 30} \mathbb{I}(|d_{ij} - D_{ij}| < \epsilon)}{\sum_{j: D_{ij} < 30} 1} \quad (1)$$

where L is the total number of nucleotides; d_{ij} and D_{ij} are the C4' distance between i th and j th nucleotides in the predicted and experimental structures, respectively; $\mathbb{I}()$ is the indicator function.

The effectiveness of these modifications was validated by comparison against established metrics. As shown in Supplementary Fig. 10, a contact radius of 30 Å achieves an optimal balance: it is large enough to achieve a plateaued correlation with other global metrics, yet modest enough to preserve sensitivity to local atomic arrangements (Supplementary Fig. 3f). This prevents the performance degradation seen with excessively large radii: for instance, the 50 Å cutoff yields a lower correlation with GDT-TS than the 30 Å cutoff. Furthermore, Supplementary Figs. 1 and 2 reveal a clear and consistent improvement for IDDT_{RNA} in nearly all cases in terms of the correlation with other established metrics, confirming its superior descriptive power over IDDT_{raw}. Overall, IDDT_{RNA} provides a more robust and reliable assessment of RNA model quality than the standard IDDT formulation.

Furthermore, a detailed comparison against TM-score_{RNA} on the CASP16 dataset (Supplementary Fig. 3) highlights the key advantages of IDDT_{RNA}. While TM-score_{RNA} exhibits a weaker dependency on sequence length, this comes at the cost of systematically assigning low scores to short RNAs, a known limitation highlighted in the official CASP15 assessment²³. In contrast, our analysis shows that IDDT_{RNA} does not suffer from this drawback, providing a more meaningful and objective scoring range for these targets (Supplementary Fig. 3a). A clear example is R1288TS481_1: despite being broadly topologically correct (RMSD 5.28 Å, Supplementary Fig. 3c), it receives a misleadingly low TM-score_{RNA} of 0.26. In contrast, IDDT_{RNA} assigns a much more objective score of 65.9. In addition, IDDT_{RNA} offers superior discriminatory power for near-native models (Supplementary Fig. 3d, e) as well as reflecting the local structural accuracies (measured by interaction network fidelity or INF, which measures the base-pairing accuracy; Supplementary Fig. 3f, g). Finally, a key practical advantage is that IDDT_{RNA} can be decomposed into a robust per-residue score—a capability not offered by most established global metrics but essential for detailed local analysis.

RNArank algorithm

As shown in Fig. 1a, the RNArank pipeline involves three key steps: (1) preparation of the multi-modal input features, (2) prediction of contact and deviation maps using a Y-shaped residual neural network, and (3) calculation of the quality assessment score pIDDT. An overview of the network architecture and feature types is presented below, and a more detailed description of the features is available in Supplementary Table 1.

Step 1. Preparation of input features. The input of RNArank is the full-atom structure of the model to be assessed. Three sets of features are calculated from this input as follows:

1D features. This set includes the RNA sequence, local backbone geometry, and global shape information from ultrafast shape recognition (USR). The sequence is represented by one-hot encoding of the four nucleotide types, while local geometry is defined by the seven backbone torsion angles. To capture global topology in per-nucleotide view, we employ the USR method, which has demonstrated its effectiveness in protein QA^{34,36}. In RNArank,

USR features are calculated using the C4' atom as the representative point for each nucleotide, describing the overall structure via per-nucleotide topological moments.

2D features. This set primarily describes the inter-nucleotide relationships, comprising Euclidean distances, Rosetta two-body centroid energy terms, and the steric clashes (Eq. (2)). Inter-nucleotide distances are calculated between the C4', P, and N atoms, encoded using Gaussian radial basis function. The Rosetta energy features consist of various two-body terms, including *fa_atr*, *fa_rep*, *lk_nonpolar*, *rna_torsion*, *fa_stack*, *stack_elec*, *geom_sol_fast*, *hbond_sc*, and *fa_elec_rna_phos_phos*. These energy terms are normalized before being fed into the neural network. The steric clashes are quantified by calculating the frequency of non-hydrogen atom pairs whose distance falls below a threshold determined by their van der Waals radii. The specific formula for this clash metric is as follows:

$$P_{\text{clash}}(M, N) = \frac{\sum_{m \in M} \sum_{n \in N} \mathbb{I}(d_{mn} < r_m + r_n - 0.4)}{\sum_{n \in N} \sum_{m \in N} 1} \quad (2)$$

where d_{mn} is the distance between atoms m and n , which belong to nucleotides M and N , respectively. r_i is the van der Waals radius of atom i .

3D features. This set consists of the voxelized representation of each nucleotide's local atomic environment, using a method similar to RNA3DCNN³⁰. To ensure rotational and translational invariance, we first establish a local coordinate system for each nucleotide based on its P, C4', and key base atoms (N1 for pyrimidines; N9 for purines). The surrounding environment is then mapped onto a 3D grid aligned with this local frame. Each grid box represents a voxel with three channels, which store the summed atomic occupation, mass, and charge, respectively. This process transforms each nucleotide's surrounding environment into a 3D representation that is invariant to global translations and rotations.

The above three sets of input features are encoded through a multi-modal encoder consisting of separate convolutional layers corresponding to their respective dimensionalities (Fig. 1b). Specifically, the 3D voxelization features are passed through 3D convolutional layers, flattened, and then concatenated with the 1D features (Fig. 1c); the resulting tensor is subsequently encoded by 1D convolutions, tiled to match the 2D feature dimensions, and added to the 2D features. Finally, this combined tensor is fed into a 2D ResNet block to generate a unified 2D representation.

Step 2. Prediction of 2D contact and deviation maps. In the second step, the unified representation obtained in step 1 is fed into a Y-shaped ResNet module to predict the contact and deviation maps. The contact map is defined as the inter-nucleotide C4' distance below 30 Å. The deviation map is defined as the one-hot encoding of the distance error falling into five bins corresponding to the deviation cutoffs: <1 Å, 1 ~ 2 Å, 2 ~ 4 Å, 4 ~ 6 Å, and >6 Å. Specifically, this module consists of three stages:

1. **Unified representation updating.** The initial unified representation is updated using a stack of 50 ResNet blocks. Notably, dilated convolutions (Fig. 1d) are incorporated to improve the efficiency of capturing long-range dependencies while maintaining computational efficiency.
2. **Task-specific representation learning.** The output from the first stage is then channeled into two parallel branches: one with 50 ResNet blocks to learn the contact representation, and another with 50 ResNet blocks for the deviation representation.
3. **Output layer.** Finally, each representation is passed through a linear layer followed by either a sigmoid (for contact prediction) or a softmax (for deviation prediction) activation to generate the predicted probability maps.

Step 3. Calculation of the predicted IDDT (pIDDT) score. In the final step, the predicted 2D contact and deviation maps are used to calculate the pIDDT score. The calculation proceeds as follows. For any pair of

nucleotides (i, j), the two predicted maps can derive: (1) contact probability, P_{contact}^{ij} , which is the model's confidence that the $C4'-C4'$ distance between these nucleotides are within 30 Å; (2) deviation probabilities, $P_{\text{dev} < d}^{ij}$, representing the probability that the deviation in the $C4'-C4'$ distance is less than a given threshold $d \in \{1.0 \text{ Å}, 2.0 \text{ Å}, 4.0 \text{ Å}, 6.0 \text{ Å}\}$. Subsequently, the local pLDDT score for the i th nucleotide is calculated following the IDDT_{RNA} definition (Eq. (1)). The global pLDDT score is then computed by averaging all the local scores. The specific formulas are as follows:

$$\text{pLDDT}_i = \frac{100}{4} \cdot \frac{\sum_{i \neq j} (P_{\text{dev} < 1}^{ij} + P_{\text{dev} < 2}^{ij} + P_{\text{dev} < 4}^{ij} + P_{\text{dev} < 6}^{ij})}{\sum_{i \neq j} P_{\text{contact}}^{ij}} \quad (3)$$

$$\text{pLDDT}_{\text{global}} = \frac{1}{L} \sum_{i=1}^L \text{pLDDT}_i \quad (4)$$

where L is the sequence length.

Construction of datasets

Benchmark datasets. Our evaluation is based on three distinct benchmark datasets. The first set (T24) contains 24 RNA-only entries from PDB (released after January 2022) for which we generated a comprehensive model library using a variety of RNA structure prediction methods, including AlphaFold3¹⁶, DeepFoldRNA¹², DRFold¹³, RhoFold¹⁷, RNAComposer^{8,43}, RoseTTAFold2NA¹⁵, SimRNA⁹, and trRosettaRNA¹⁴. Furthermore, to increase the diversity of the test set, we generated additional decoys using methods similar to those for our training set (see below), namely structure perturbation, molecular dynamics simulations, and Rosetta fragment assembly-based modeling. The number of structure models generated by each method is listed in Supplementary Table 2. Two additional test sets were derived from the CASP15 and CASP16 competitions, containing 12 and 42 targets, respectively. For these two sets, we assessed the original models submitted by the official participants, providing a realistic and challenging benchmark. For the homo-multimeric targets in CASP16, our assessment was performed on their single, representative chains.

Training sets. To train our models, we initially collected RNAs from the trRosettaRNA training dataset, which included experimental structures resolved before January 2022. This dataset was then filtered to remove RNAs with non-standard nucleotides, chain discontinuities, or formatting errors in their coordinate files. We also excluded sequences longer than 200 nucleotides. To ensure a strict separation from our test sets, we used CD-HIT-EST⁴⁴ to remove any sequences identical to those in our test sets. This filtering process resulted in a final set of 1635 RNAs.

For each of these 1635 RNAs, we generated a diverse set of structure models, totaling ~200,000 models. These models were generated using a combination of various strategies: (1) deep learning-based prediction, (2) physics-based folding, (3) structure perturbation, (4) molecular dynamics simulations, and (5) fragment assembly-based modeling. The details of each approach are outlined below.

1. *Deep learning prediction.* AlphaFold3, DeepFoldRNA, DRFold, RoseTTAFold2NA, RhoFold, and trRosettaRNA were installed and run locally on our computing cluster.
2. *Physics-based folding.* We also employed SimRNA, a physics-based RNA folding method that simulates the folding process based on the physical interactions between nucleotides, providing a complementary approach to deep learning-based methods.
3. *Structure perturbation.* For each experimental structure, we applied random perturbations to the backbone angles of each nucleotide using a standard normal distribution. These perturbations allow for the generation of diverse decoys while maintaining structural plausibility, providing a more comprehensive dataset for training the model.
4. *Molecular dynamics simulations.* We performed molecular dynamics (MD) simulations using GROMACS⁴⁵ to model RNA structural

dynamics. For each RNA experimental structure, we solvated the system with water and neutralized it by adding ions. The CHARMM36-jul2022 force field was applied, and energy minimization was followed by NVT and NPT equilibration to relax the system. A production MD simulation was then conducted to generate trajectories, with snapshots extracted at regular intervals for analysis.

5. *Rosetta fragment assembly-based modeling.* We employed Rosetta to generate decoys using various templating strategies, including native fragments and templates identified through searches. Additionally, we also used secondary structures predicted by RNAfold^{46,47} to generate structure models.

To ensure structural diversity and reduce redundancy within the generated structure models, we performed a clustering step. Specifically, for each RNA, we applied k-means clustering to group the models according to their pairwise IDDT_{RNA} matrix, and selected representative models from each cluster to form the final training set. On average, we generated ~120 structure models for each RNA, with the explicit number depending on the success of the MD simulations.

The quality distribution of this final, curated training set is shown in Supplementary Fig. 11. Specifically, deep learning predictors provided a rich population of high-quality models (IDDT_{RNA} > 75), which are crucial for learning the fine details of near-native structures. Structure perturbation and molecular dynamics simulations were particularly effective at populating the medium-quality range (IDDT_{RNA} = 50~75), helping the model learn a smooth and discriminative scoring landscape. Finally, physics-based folding and fragment assembly generated a wide array of low-quality models (IDDT_{RNA} < 50), training the model to recognize and penalize common structural flaws. This diverse and well-distributed dataset is fundamental to RNArank's ability to develop a robust understanding of RNA structural features and excel at high-resolution model selection.

Nonredundant test subsets. To rigorously evaluate generalization performance, we further constructed stringent, non-redundant subsets for each of the three benchmarks. In this evaluation, we removed all test targets sharing over 80% sequence identity with any sequence in our training set using CD-HIT-EST. This filtering resulted in final test sets of 17 (T24), 10 (CASP15), and 28 (CASP16) targets.

Loss function and training procedure

The ground-truth labels for training include contact maps, deviation maps, and per-nucleotide IDDT scores of each structure model. Correspondingly, the total loss function guiding the training of RNArank is a weighted sum of three components: a contact loss, a deviation loss, and an IDDT loss. The overall objective is to minimize:

$$L_{\text{total}} = \lambda_1 L_{\text{contact}} + \lambda_2 L_{\text{deviation}} + \lambda_3 L_{\text{IDDT}} \quad (5)$$

where the weighting factors λ_1 , λ_2 , and λ_3 were empirically set to 1.0, 1.0, and 20.0, respectively. The individual loss components are defined as follows:

1. L_{contact} : A binary cross-entropy loss of the contact map prediction, which classifies nucleotide pairs as being in contact (<30 Å) or not.
2. $L_{\text{deviation}}$: A categorical cross-entropy loss of the deviation map prediction, which is framed as a multi-class classification problem over discrete distance error bins. This loss is calculated only for nucleotide pairs that are in contact (i.e., distance <30 Å) in the experimental structure, thereby focusing the model on learning distance errors for interacting pairs.
3. L_{IDDT} : A mean squared error (MSE) loss of the final per-nucleotide pLDDT score, encouraging the network to accurately regress the true IDDT value.

RNArank was trained using the Adam optimizer with an initial learning rate of 5e-4, an exponential decay rate of 0.99 per epoch, and a dropout rate of 0.2 applied to the ResNet blocks. Each training epoch

involved iterating over all RNAs in the training set, with one structure model randomly sampled for each RNA. The model checkpoint with the lowest validation loss (10% of the training samples were retained for validation) was saved for final evaluation. To enhance prediction robustness, we trained two neural network models. The final RNArank score is the average of the pIDDT outputs from these two models, providing a more stable and reliable prediction.

Selected baseline methods

We benchmark RNArank against a comprehensive suite of existing methods, including traditional knowledge-based statistical potentials (cgRNASP²⁶, rsRNASP²⁷, and DFIRE-RNA²⁸) and deep learning-based methods (RNA3DCNN³⁰ and lociPARSE³¹). rsRNASP and cgRNASP are both residue-separation-based statistical potentials. They differ primarily in their granularity; rsRNASP considers all atoms, whereas cgRNASP utilizes three coarse-grained (CG) representations to increase efficiency. In this work, we evaluated all three versions of cgRNASP: the default 3-bead cgRNASP (representing each nucleotide with P, C4', and N1/N9 atoms for pyrimidine/purine), the 2-bead cgRNASP-PC (P and C4' atoms), and the 1-bead cgRNASP-C (C4' atom only). DFIRE-RNA is another all-atom potential, using a distance-scaled finite ideal-gas reference state to derive its energy function.

Among the deep learning-based methods, RNA3DCNN uses a 3D convolutional neural network to predict an RMSD-based “unfitness” score for each nucleotide by treating its local environment as a 3D image. More recently, lociPARSE was developed to estimate the IDDT_{raw} score using the invariant point attention (IPA) architecture, which is inherently invariant to global rotations and translations.

For a fair and rigorous comparison, all baseline methods were installed and run locally on our computing cluster under their default configurations. The sole input for each method was the PDB file of a predicted structure model. For RNA3DCNN, we employed the RNA3DCNN_MD model, which was trained on structure models from molecular dynamics (MD) simulations.

Performance assessment

The performance of all methods was assessed across multiple criteria to provide a comprehensive evaluation, including correlation with true quality scores, the ability to distinguish high-quality models, and top 1 model selection. The specific metrics are defined as follows:

Global correlation. We assess global molecule-wise performance by computing the Spearman’s rank correlation coefficient (ρ) between predicted scores and true quality metrics. To provide a comprehensive view, this analysis is conducted on two bases: (1) *overall* basis, where the correlation is computed across the pooled set of all structure models from different targets, and (2) *per-target* basis, where the correlation is first calculated for each target’s structure models and then averaged across all targets. For both analyzes, we evaluate the correlation against two distinct ground-truth metrics: global IDDT_{RNA} and RMSD. We choose Spearman ranking correlation due to its robustness against non-linear relationships and non-normal data distributions.

Local correlation. In this case, the correlation is computed between the predicted per-nucleotide scores and the true local IDDT_{RNA} values. Similarly, local-level performance is evaluated on both the *overall* and *per-target* bases.

Ability to distinguish high-quality models. This ability is quantified by the AUC calculated on the *overall* basis, with models having a true IDDT_{RNA} >75 considered positive samples.

Top 1 model selection. To evaluate a QA method’s ability to identify the best model, we report the per-target average accuracy of the top-ranked

model in terms of both IDDT_{RNA} and RMSD. A higher average top 1 IDDT_{RNA} and a lower average top 1 RMSD thus indicate better performance.

Statistics and reproducibility

All data were collected and analyzed using standard statistical methods. Detailed information regarding these analyzes is provided in the relevant figures, figure legends, and the “Results” section. To ensure reproducibility, we performed three independent replications for each target across all datasets (totaling 78 targets), resulting in 234 independent runs. All experiments were completed successfully and yielded consistent results.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The authors declare that the data supporting the results and conclusions of this study are available within the paper and its Supplementary Information. The source data underlying Figs. 2–4 are provided in the Supplementary Data file. The test source data used in this study is available at Zenodo⁴⁸ and our website (<https://yanglab.qd.sdu.edu.cn/RNArank/benchmark/>).

Code availability

The RNArank web server is available at: <https://yanglab.qd.sdu.edu.cn/RNArank/>. The source codes are available at Zenodo⁴⁸ and GitHub (<https://github.com/YangLab-SDU/RNArank/>).

Received: 7 August 2025; Accepted: 12 January 2026;

Published online: 21 January 2026

References

1. Kwon, D. RNA function follows form - why is it so hard to predict?. *Nature* **639**, 1106–1108 (2025).
2. Madison, J. T., Everett, G. A. & Kung, H. Nucleotide sequence of a yeast tyrosine transfer RNA. *Science* **153**, 531–534 (1966).
3. Levitt, M. Detailed molecular model for transfer ribonucleic acid. *Nature* **224**, 759–763 (1969).
4. Sharma, S., Ding, F. & Dokholyan, N. V. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* **24**, 1951–1952 (2008).
5. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55 (2008).
6. Rother, M., Rother, K., Puton, T. & Bujnicki, J. M. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* **39**, 4007–4022 (2011).
7. Zhao, Y. et al. Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* **2**, 734 (2012).
8. Popena, M. et al. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40**, e112 (2012).
9. Boniecki, M. J. et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **44**, e63 (2016).
10. Watkins, A. M., Rangan, R. & Das, R. FARFAR2: improved De Novo Rosetta prediction of complex global RNA folds. *Structure* **28**, 963–976.e966 (2020).
11. Li, J., Zhang, S., Zhang, D. & Chen, S.-J. Vfold-Pipeline: a web server for RNA 3D structure prediction from sequences. *Bioinformatics* **38**, 4042–4043 (2022).
12. Pearce, R., Omenn, G. S. & Zhang, Y. De Novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2022.05.15.491755v1> (2022).

13. Li, Y. et al. Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nat. Commun.* **14**, 5745 (2023).
14. Wang, W. et al. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nat. Commun.* **14**, 7266 (2023).
15. Baek, M. et al. Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* **21**, 117–121 (2024).
16. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
17. Shen, T. et al. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat. Methods* **21**, 2287–2298 (2024).
18. Wang, W., Peng, Z. & Yang, J. Predicting RNA 3D structure and conformers using a pre-trained secondary structure model and structure-aware attention. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2025.04.09.647915v1> (2025).
19. Kagaya, Y. et al. NuFold: end-to-end approach for RNA tertiary structure prediction with flexible nucleobase center representation. *Nat. Commun.* **16**, 881 (2025).
20. Li, Y., Feng, C., Zhang, X. & Zhang, Y. Ab initio RNA structure prediction with composite language model and denoised end-to-end learning. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2025.03.05.641632v1> (2025).
21. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
22. Schneider, B. et al. When will RNA get its AlphaFold moment?. *Nucleic Acids Res.* **51**, 9522–9532 (2023).
23. Das, R. et al. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins* **91**, 1747–1770 (2023).
24. Bu, F. et al. RNA-puzzles round V: blind predictions of 23 RNA structures. *Nat. Methods* **22**, 399–411 (2025).
25. Kretsch, R. C. et al. Assessment of nucleic acid structure prediction in CASP16. *Proteins* <https://doi.org/10.1002/prot.70072> (2025).
26. 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 Genom. Bioinf.* **5**, lqad016 (2023).
27. 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. *Biophys. J.* **121**, 142–156 (2022).
28. 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. *J. Comput. Biol.* **27**, 856–867 (2020).
29. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
30. Li, J. et al. RNA3DCNN: local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLoS Comput. Biol.* **14**, e1006514 (2018).
31. Tarafder, S. & Bhattacharya, D. lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures. *J. Chem. Inf. Model.* **64**, 8655–8664 (2024).
32. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).
33. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778 (IEEE, 2016).
34. Ballester, P. J., Finn, P. W. & Richards, W. G. Ultrafast shape recognition: evaluating a new ligand-based virtual screening technology. *J. Mol. Graph. Model.* **27**, 836–845 (2009).
35. Hiranuma, N. et al. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat. Commun.* **12**, 1340 (2021).
36. Guo, S. S., Liu, J., Zhou, X. G. & Zhang, G. J. DeepUMQA: ultrafast shape recognition-based protein model quality assessment using deep learning. *Bioinformatics* **38**, 1895–1903 (2022).
37. Gong, S., Zhang, C. & Zhang, Y. RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA. *Bioinformatics* **35**, 4459–4461 (2019).
38. Antczak, M. et al. New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim. Polonica* **63**, 737–744 (2016).
39. Xiong, P., Wu, R., Zhan, J. & Zhou, Y. Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement. *Nat. Commun.* **12**, 2777 (2021).
40. Wang, W., Luo, Y., Peng, Z. & Yang, J. Accurate biomolecular structure prediction in CASP16 with optimized inputs to state-of-the-art predictors. *Proteins* <https://doi.org/10.1002/prot.70030> (2025).
41. Zeng, C., Jian, Y., Vosoughi, S., Zeng, C. & Zhao, Y. Evaluating native-like structures of RNA-protein complexes through the deep learning method. *Nat. Commun.* **14**, 1060 (2023).
42. Zhao, K., Zhao, P., Wang, S., Xia, Y. & Zhang, G. FoldPathreader: predicting protein folding pathway using a novel folding force field model derived from known protein universe. *Genome Biol.* **25**, 152 (2024).
43. Sarzynska, J., Popenda, M., Antczak, M. & Szachniuk, M. RNA tertiary structure prediction using RNAComposer in CASP15. *Proteins* **91**, 1790–1799 (2023).
44. 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).
45. Páll, S. et al. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **153**, 134110 (2020).
46. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
47. Lorenz, R. et al. ViennaRNA Package 2.0. *Algor. Mol. Biol.* **6**, 26 (2011).
48. Liu, X. et al. Source code and data for “Quality assessment of RNA 3D structure models using deep learning and intermediate 2D maps”. Zenodo, <https://doi.org/10.5281/zenodo.17889848> (2025).
49. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. *ECCV 2016*, 630–645 (Springer, 2016).

Acknowledgements

This work is supported by the following funding sources: National Natural Science Foundation of China (NSFC T2225007, T2222012, 32430063, 62402075, 62501364), Postdoctoral Fellowship Program and China Postdoctoral Science Foundation (BX20240212, 2025M783122), the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202300639), and Fundamental Research Funds for the Central Universities.

Author contributions

J.Y. conceptualized and administered the study. X.L. designed and implemented the network. X.L., W.W., and J.Y. conducted the formal analysis. X.L., W.W., Z.D., and Z.W. performed data curation and generated decoys. Z.P. and W.W. co-supervised the study. All authors wrote and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-026-09582-2>.

Correspondence and requests for materials should be addressed to Wenkai Wang or Jianyi Yang.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Xiangtao Li and Laura Rodríguez Pérez.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026