nature protocols

https://doi.org/10.1038/s41596-026-01356-8

Protocol

The trRosettaRNA server for RNA structure prediction

Wenkai Wang, Xiaocheng Liu, Zhenling Peng & Jianyi Yang ✉

Abstract

Similar to proteins, many RNAs fold into three-dimensional (3D) structures to perform biological functions. Here we present the trRosettaRNA server, a web-based platform for automated RNA 3D structure prediction using deep learning. The primary input is the nucleotide sequence of a target RNA, with the option to upload custom multiple sequence alignments and secondary structures. The server uses an end-to-end neural network for automated 3D structure prediction, followed by an energy optimization step to resolve structural violations. As an automated server, trRosettaRNA is distinguished by its state-of-the-art modeling accuracy, flexible input options and comprehensive visualization of prediction results. trRosettaRNA has been successfully applied in various contexts, including predicting structures for Rfam families lacking known 3D structures, where representative cases of high-confidence structure predictions were found to align well with subsequent experimental observations. Utilizing up to 5 central processing unit (CPU) cores in parallel on our computer cluster, the server takes a median time of about 1 h to predict structures for RNA sequences with about 200 nucleotides. The standalone package for trRosettaRNA offers distinct advantages such as enhanced data privacy for sensitive sequences, the ability to bypass server queues and integration into high-throughput automated pipelines. Importantly, the open-source nature of the package empowers researchers to directly modify the codebase for specialized research needs or to develop derivative tools by fine-tuning the underlying neural network. The web server and standalone package of trRosettaRNA are available at https://yanglab.qd.sdu.edu.cn/trRosettaRNA/ and https://github.com/YangLab-SDU/trRosettaRNA2, respectively.

Key points

- trRosettaRNA is an automated, deep learning-based platform for accurate RNA 3D structure prediction characterised by state-of-the-art performance, input flexibility and comprehensive visualizations. As 3D structures generated by the neural network may contain steric violations, clashscores are calculated and used to determine whether to perform PyRosetta-powered energy minimization.
- An open-source standalone package is also provided, enabling large-scale modeling tasks and facilitating customized development for specialized and high-throughput research workflows.

Key references

Wang, W. et al. *Nat. Commun.* **14**, 7266 (2023): https://doi.org/10.1038/s41467-023-42528-4

Wang, W. et al. *Proteins* **94**: 142–153 (2026): https://doi.org/10.1002/prot.70030

Wang, W. et al. Preprint at *bioRxiv* (2025): https://doi.org/10.1101/2025.04.09.647915

MOE Frontiers Science Center for Nonlinear Expectations, State Key Laboratory of Cryptography and Digital Economy Security, Shandong Province Key Laboratory of Financial Risk, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China. ✉e-mail: yangjy@sdu.edu.cn

Introduction

RNA is a fundamental macromolecule involved in diverse essential biological activities, including regulating gene expression, directing protein synthesis and catalyzing reactions as a ribozyme. Similar to proteins, many RNAs fold into three-dimensional (3D) structures to perform biological functions. However, unlike protein structures, RNA 3D structures are notoriously challenging to determine, which is largely due to their intrinsic conformational flexibility and susceptibility to degradation. This disparity is evident in the Protein Data Bank (PDB)[1], which holds only about 8,000 RNA structures (including about 3,000 ribosomal and about 800 nonredundant entries)[2], far fewer than the approximately 200,000 protein structures available. Despite decades of efforts dedicated to developing computational methods for accelerating RNA 3D structure determination[3–11], their practical application was hampered by limitations in accuracy and computational efficiency.

In contrast to the slow advancements in RNA structure prediction, protein structure prediction has experienced substantial breakthroughs driven by the application of deep learning techniques. For example, our laboratory developed trRosetta[12,13], which uses residual convolutional neural networks[14,15] to predict the 2D internucleotide geometry and then builds the 3D structure using energy minimization powered by pyRosetta[16]. Another prominent example is AlphaFold 2 (ref. 17), whose developers received the 2024 Nobel Prize in Chemistry for achieving unprecedented accuracy in protein structure prediction. This system uses a transformer-based neural network to directly predict 3D structures in an end-to-end manner.

Inspired by the success of deep learning in protein structure prediction, our laboratory initially developed RNAcontact[18], a deep learning-based model to predict internucleotide contacts in RNA. While RNAcontact could accurately predict internucleotide contacts, its application to the RNA 3D structure modeling was limited, probably owing to the insufficient restraints from the contact maps for capturing the intricate fold of RNA molecules.

Building upon these observations, we developed trRosettaRNA[19], a fully automated method for RNA 3D structure prediction, which integrates a transformer neural network with physics-based energy minimization for structure folding. Rigorous benchmarks and blind tests have demonstrated the outstanding performance of trRosettaRNA as an automated method[19–22]. Since its launch in 2023, the trRosettaRNA server has processed >24,000 submissions by >4,700 users from >75 countries/regions. This widespread adoption reflects the broad utility of the trRosettaRNA method, which is now extensively used in fields such as structural biology, biochemistry and computational biology. In this protocol, we provide detailed guidance on the applications and features of the trRosettaRNA server, aiming to facilitate its effective utilization by the research community. A glossary of key terms used in this protocol is listed in Table 1.

Development of the protocol

The development of trRosettaRNA was primarily inspired by AlphaFold 2 (ref. 17) and trRosetta[12,13], two deep learning algorithms that achieved success in protein structure prediction. Figure 1 shows the modeling pipeline of trRosettaRNA. The primary input is the nucleotide sequence of a target RNA. The server then automatically constructs a multiple sequence alignment (MSA) by searching for homologous sequences. A secondary structure (SS) is subsequently predicted using a transformer-based neural network. The MSA and SS, which can also be provided by users, are then fed into an end-to-end neural network to predict 2D internucleotide geometry and the 3D structure. Finally, an energy minimization module, implemented within the PyRosetta framework[16], is used to reduce potential structural violations in the predicted 3D structure. To enhance accessibility for biologists and expedite RNA research, we developed a user-friendly web server and released the open-source code as a standalone package.

Updates made since the first release

A few updates were made after the initial release to improve the performance and user experience, which are listed below.

Table 1 | Glossary: key terms

Term/abbreviation	Definition
MSA	Multiple sequence alignment. An alignment of three or more biological sequences used to extract co-evolutionary information to guide structure prediction
SS	Secondary structure. The local topology of an RNA molecule formed by intra-molecular base pairing (e.g., stems and loops)
RMSD[a]	Root mean square deviation. The standard measure of the average distance between the atoms of a predicted model and a known experimental reference. A lower value indicates a more accurate 3D model
eRMSD	Estimated RMSD. The predicted RMSD value between the model and the native structure[19]. Lower scores reflect greater model confidence
LDDT[a]	Local distance difference test. A superposition-independent score (0–1) that evaluates the local atomic environment by comparing interresidue distance networks between the model and the experimental structure
pLDDT	Predicted local distance difference test. A per-residue confidence metric ranging from 0 to 100. Higher scores reflect greater model confidence
TM-score[a]	Template modeling score. A metric (0–1) for assessing global topological similarity. For RNA, a score >0.45 typically indicates a similar global fold
GDT-TS[a]	Global distance test total score. A metric (0–1) used to assess the global similarity between a model and a reference structure
INF[a]	Interaction network fidelity. A measure used to assess the precision of the model's predicted base-pairing interactions. A higher score (0–1) indicates a more accurate base-pairing network
Clashscore[a]	The number of serious atomic overlaps (≥0.4 Å) per 1,000 atoms. A lower score indicates a more physically plausible model with fewer steric violations
CASP	Critical assessment of structure prediction. A premier biennial competition and blind assessment platform that has served as the gold standard for benchmarking protein and RNA structure prediction methods since 1994
BLASTN	A sequence alignment program used to search nucleotide databases for sequences with high local similarity to a query, often used in building MSAs
Infernal	A tool suite used for searching RNA sequence databases for RNA structure and sequence similarities using covariance models, often used in building MSAs

[a]Detailed descriptions of these evaluation metrics used in Fig. 2 are provided in the 'Anticipated results' section.

MSA searching was accelerated. In the initial trRosettaRNA protocol, MSA searching relied on rMSA[23], a heuristic algorithm that iteratively searches against RNAcentral[24] and NT databases[25]. Although this strategy can retrieve more homologous sequences, the runtime was extremely long, and the contribution to final accuracy was found to be negligible. Therefore, we simplified the MSA search pipeline to now exclusively use the RNAcentral database, substantially improving speed without sacrificing accuracy (see the 'Overview of the procedures' section for more details).

SS optimization and custom input support were added. The original trRosettaRNA used SPOT-RNA[26], a third-party method, to predict the SS. To improve the performance and reduce external dependencies, we developed an in-house SS predictor, trRNA-SS (see Supplementary Text 1 and our recent report[20] for details). By integrating MSA information with an advanced transformer architecture, trRNA-SS outperforms SPOT-RNA, which relies on single-sequence inputs and conventional ResNet[14] frameworks. Furthermore, to increase flexibility for specialized applications, the server now supports custom inputs for both MSA and SS. Supported MSA formats include A3M, A2M, FASTA and STO, and supported SS formats are dot-bracket and connectivity table (CT).

End-to-end prediction and a new confidence score were added. Initially, trRosettaRNA adopted a two-step approach: first predicting 2D geometry with a neural network, then folding the 3D structure via energy minimization. We subsequently expanded trRosettaRNA to include an end-to-end prediction capability, directly generating 3D coordinates from MSA and SS inputs (see Supplementary Text 1 and our recent report[20] for details), which improves both accuracy and speed. The original two-step approach involving energy minimization is still available, and is particularly useful when the end-to-end prediction yields structures with substantial

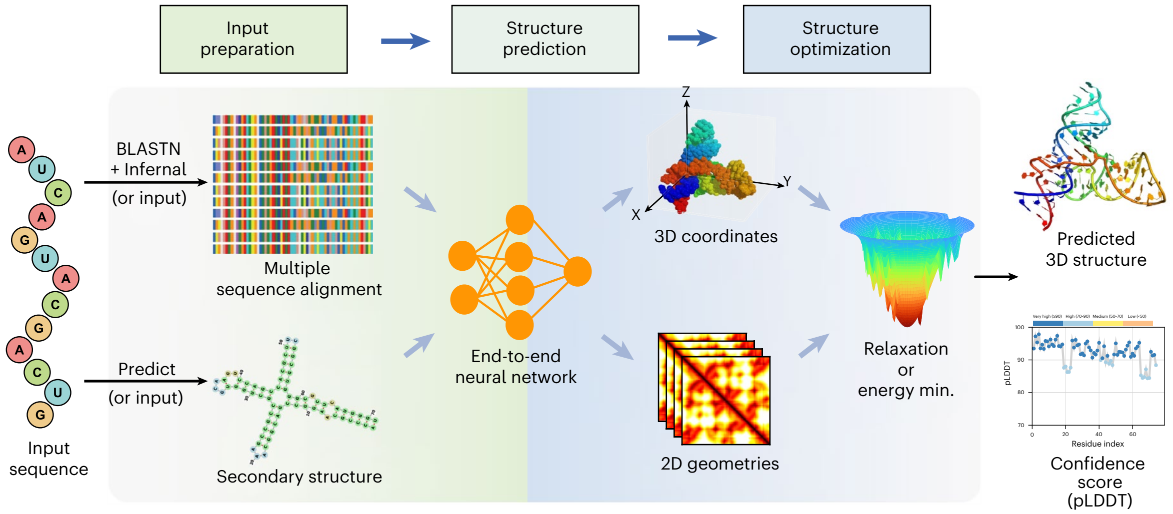


Fig. 1 | Overview of trRosettaRNA protocol. Starting with the sequence of a target RNA, the protocol generates a MSA and a SS. Optionally, users can also input their custom MSA and/or SS. The generated/input MSA and SS are fed into an end-to-end neural network to predict initial 3D atomic coordinates and 2D internucleotide distances. Subsequently, the predicted coordinates are then optimized through fast relaxation or energy minimization (min.) (depending on the severity of steric clash) to produce a physically realistic structure. The final outputs include the predicted 3D structure and an associated confidence score (pLDDT). Clashscores can be independently checked using programs such as Molprobity[54] (Molprobity was used to generate the results shown in Fig. 2b).

steric clashes. To ensure optimal structural quality, the server automatically assesses the clash level of the end-to-end model and, if necessary, re-runs the folding process using the energy minimization pipeline (see the 'Overview of the procedures' for detailed decision criteria). In addition, the confidence score metric was updated from the previously used estimated root mean square deviation (eRMSD) to the predicted local distance difference test (pLDDT). This change is essential because, unlike eRMSD, which relies on multiple decoys generated during energy minimization, pLDDT is an intrinsic score that directly evaluates structural quality from the end-to-end network output. Therefore, pLDDT is better suited for the current single-model prediction workflow and demonstrates a strong correlation with actual model accuracy (see 'Overview of the procedures' for more details).

Applications

trRosettaRNA has found widespread usage in various applications, which can be broadly categorized into two areas: RNA structure determination and the development of related computational methods.

RNA structure determination

trRosettaRNA is frequently used to aid the experimental determination of RNA structures. For instance, trRosettaRNA was used to predict the upstream noncoding RNA (ncRNA) of prokaryotic defense-associated reverse transcriptases, helping to reveal the mechanisms by which the defense-associated reverse transcriptases–ncRNA system mediates antiphage defense[27]. In addition, trRosettaRNA was used to assist in the construction of the 3D structure model of pre-tRNA, which served as a high-quality template for subsequent manual fitting into cryo-EM density maps of the ELAC2–pre-tRNA complex, thereby contributing to understanding pre-tRNA cleavage[28]. trRosettaRNA also provided helpful structural references when studying an *IscB*-associated ωRNA variant exhibiting substantially enhanced gene editing efficiency[29].

Development of related computational methods

The predictive capabilities of trRosettaRNA, along with its comprehensive training and benchmarking datasets, have facilitated the development of related computational methods.

Table 2 | Comparison between trRosettaRNA and other representative protocols for RNA 3D structure prediction

	Protocol name	Year	Web	Pkg	Conf	SS	URL
Deep learning based	trRosettaRNA[19]	2022	✓	✓	✓	✓	https://yanglab.qd.sdu.edu.cn/trRosettaRNA/
	AlphaFold 3 (ref. 33)	2024	✓	✓	✓	×	https://alphafoldserver.com/
	RoseTTAFoldNA[55]	2022	×	✓	✓	×	https://github.com/uw-ipd/RoseTTAFold2NA
	DeepFoldRNA[56]	2022	✓	✓	×	×	https://zhanglab.comp.nus.edu.sg/DeepFoldRNA/
	RhoFold+[57]	2022	✓	✓	✓	×	https://proj.cse.cuhk.edu.hk/aihlab/RhoFold/
	NuFold[58]	2024	✓	✓	✓	×	https://colab.research.google.com/github/kiharalab/nufold/blob/master/ColabNuFold.ipynb
	DRfold[59]	2023	✓	✓	×	×	https://zhanglab.comp.nus.edu.sg/DRfold/
Traditional	FARFAR2 (ref. 11)	2020	✓	✓	✓	✓	https://rosie.rosettacommons.org/farfar2
	3dRNA[9]	2012	✓	✓	✓	✓	http://biophy.hust.edu.cn/new/3dRNA
	RNAComposer[7]	2012	✓	×	×	✓	https://rnacomposer.cs.put.poznan.pl/
	SimRNA[6]	2016	✓	✓	×	✓	https://genesilico.pl/SimRNAweb
	Vfold[10]	2014	✓	✓	×	✓	https://rna.physics.missouri.edu/vfoldPipeline/

Features compared include: release year (year), availability of a web server (web), availability of a downloadable standalone package (pkg), provision of a confidence score (conf) and support for custom SS input in the web server (SS).

For example, trRosettaRNA was used to generate 3D models based on SSs predicted by RibonanzaNet[30], providing a downstream test of whether improved SS predictions translate to more accurate 3D models. Furthermore, the trRosettaRNA training sets have been instrumental in developing methods for tasks such as RNA quality assessment[31] and conditional RNA 3D structure generation[32].

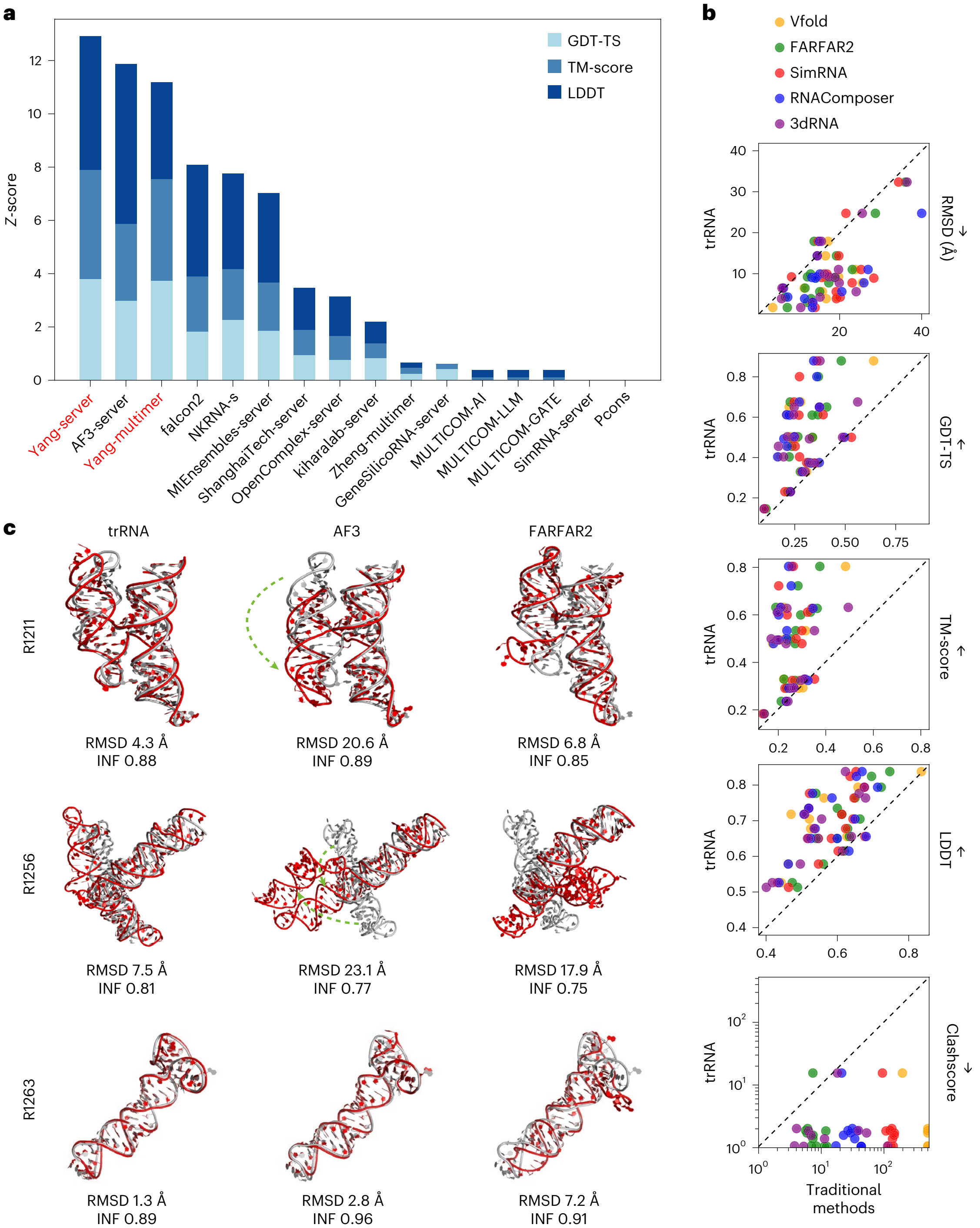
Comparison with other methods

Several other deep learning approaches have also been released concurrently with or after trRosettaRNA. The current RNA 3D structure modeling approaches can be broadly classified into two main types: deep learning-based and traditional fragment- or physics-based approaches. Table 2 summarizes a comparison between trRosettaRNA and 11 other protocols, comprising 6 deep learning-based and 5 traditional approaches. Among these 12 methods overall, 10 provide both web servers and standalone packages, 7 provide confidence scores of predictions and 6 support custom SS in the web server (trRosettaRNA provides both features). Despite the availability of numerous methods, trRosettaRNA demonstrates distinct advantages in several key aspects, as detailed below.

Modeling accuracy

Previous benchmark tests on 30 independent RNAs and blind tests in the 15th critical assessment of structure prediction (CASP15) and RNA-Puzzles have demonstrated the leading accuracy of trRosettaRNA as an automated method[19]. Subsequently, we participated in the recent CASP16 experiments, which involved 36 'RNA monomer' targets[22]. Specifically, we submitted predictions through two automated server groups: Yang-Server (TS052), which used an enhanced trRosettaRNA version guided by diverse SS and template-based modeling[21], and Yang-Multimer (TS456) relying on the default trRosettaRNA method.

According to the official ranking (https://predictioncenter.org/casp16/zscores_rna.cgi)[22], our automated server group (Yang-Server) ranked first among the 16 server groups (fourth among all 64 groups) participating in RNA modeling, outperforming AlphaFold 3 (ref. 33) (AF3-server), which ranked ninth overall. Note that the CASP16 RNA targets included several large RNAs forming complicated homomers (for example, R1251: 833 nucleotides per chain, forming a 14-mer), exceeding the capability of our automated approach. Therefore, for these RNAs, the Yang-Server submission incorporated AF3 predictions (results on these RNAs are presented in Supplementary Table 1). However, as shown in Fig. 2a, for the remaining 23 targets with <400 nucleotides or with available templates (see Supplementary Table 2 for detailed information),



Yang-Server maintained its top-ranking position among server groups. Meanwhile, the default trRosettaRNA (Yang-Multimer) also achieves competitive results with a summed *Z* score (>0) of 11.2 close to Yang-Server (12.9) and AF3-server (11.9), and substantially higher than other servers (for example, the next-best server, falcon2, scored 8.1). These results confirm the state-of-the-art performance of the trRosettaRNA server.

Moreover, trRosettaRNA surpasses the traditional automated methods (FARFAR2 (ref. 11), 3dRNA[9], SimRNA[6], RNAComposer[7], Vfold[10]) on the CASP16 RNAs across all evaluated metrics

Fig. 2 | Performance comparison on CASP16 RNAs. a, The ranking for 16 server groups participating in the CASP16 RNA experiments, evaluated on 23 RNA targets with <400 nucleotides or with available templates ($n = 23$). As per CASP16 guidelines, the ranking uses a composite Z-score ($Z_{\mathrm{NA}}^{\mathrm{CASP16}}$)[22]. This single value represents the sum of Z-scores for GDT-TS, TM-score and LDDT, calculated per target and summed over the entire benchmark set. trRosettaRNA-based groups are highlighted in bold red in the x-axis ticks. **b**, A head-to-head comparison between the default trRosettaRNA (trRNA; y axes) and five traditional approaches (x axes) on the 16 CASP16 targets for which both experimental structures and Yang-Multimer submissions were available ($n = 16$). All methods used identical inputs generated by the trRosettaRNA server. The inputs included the target sequence (for traditional methods), an MSA (for trRosettaRNA) and/or an SS (if supported). All other parameters were kept at their default settings. Arrows next to the metric labels on the right indicate the preferred direction for each metric (↓, lower is better; ↑, higher is better). **c**, Three examples to illustrate the comparison between trRNA, AlphaFold 3 (AF3) and FARFAR2. The predicted structures (shown as red cartoons) are superimposed with their corresponding experimental structures (gray cartoons) for comparison. Green arrows highlight torsional deviations in specific fragments of the AF3-predicted structures relative to the experimental references. Detailed definitions for the metrics are provided in the 'Anticipated results'. AI, artificial intelligence; LLM, large language model.

(Fig. 2b). For example, in terms of the global distance test-total score (GDT-TS)[34], a primary metric used in CASP, the default trRosettaRNA outperforms FARFAR2, 3dRNA, RNAComposer and Vfold on all targets. It also surpassed SimRNA on almost all targets, except for R1288, where SimRNA achieved a slightly higher score (GDT-TS: 0.53 versus 0.50). These results are consistent with the benchmark results reported in the original methodology paper[19].

Figure 2c shows three examples: R1211, R1256 and R1263, to illustrate the comparison between trRosettaRNA and two state-of-the-art approaches AF3 and FARFAR2 (top-performing in the RNA-Puzzles competition). For R1211 and R1256, AF3 struggled with the correct twist of certain stem-loop regions (highlighted by the green arrows), yielding poor models with root mean square deviation (RMSD) >20 Å. By contrast, trRosettaRNA can accurately capture the correct topology, producing accurate models with RMSDs of 4.3 Å and 7.5 Å, respectively. These trRosettaRNA models also surpassed those from FARFAR2 (RMSDs 6.8 Å and 17.9 Å). For R1263, both AF3 (RMSD 2.8 Å) and FARFAR2 (RMSD 7.5 Å) generated accurate structures, yet trRosettaRNA achieved an even higher level of accuracy, producing a model with an RMSD of just 1.3 Å, a near-experimental resolution.

Support for custom SS inputs

To our knowledge, trRosettaRNA is the only deep learning-based approach that enables custom SS inputs in both its web server and standalone package. This feature is beneficial in two aspects. First, exploring diverse SS inputs can further enhance trRosettaRNA performance, as evidenced by the superior performance of the Yang-Server compared to the Yang-Multimer and AF3-server (Fig. 2). Second, it allows users to effectively leverage the predictive power of deep learning to meet specific research needs. Although traditional approaches typically support custom SS inputs, their overall modeling accuracy generally falls behind trRosettaRNA, potentially reducing the reliability of their predictions. For example, for the three cases shown in Fig. 2c, although FARFAR2 yields accurate base-pair interactions in its predicted 3D structures (as measured by the interaction network fidelity (INF)[35]), its 3D structural accuracy still lags behind trRosettaRNA. This comparison illustrates the better ability of trRosettaRNA to capture high-order interactions beyond simple base pairs, ultimately contributing to more accurate 3D structures. In addition, trRosettaRNA facilitates the incorporation of higher-order base pairing information, such as triple helices, via a matrix input in its standalone package. Supplementary Figure 1 illustrates this capability with an RNA with a triple helix motif (PDB ID: 8HB8), where trRosettaRNA successfully captures this motif and produces a model with near-experimental accuracy (RMSD of about 1 Å).

Structure optimization

Through physics-based optimization, the trRosettaRNA protocol effectively minimizes steric clashes in its predicted structures. As shown in Fig. 2b,c and Supplementary Table 3, trRosettaRNA models exhibit a lower level of steric clashes compared to those generated by other methods. Specifically, trRosettaRNA yields structures with a clashscore <6 for over 90% of targets, a substantially higher proportion than other deep learning approaches such as AF3 (18%) and traditional methods such as FARFAR2 (25%). This result highlights the advantage of the trRosettaRNA pipeline in both modeling accuracy and structural validity.

Response time

The response time is a critical factor for the utility of a web server. To assess this, we benchmarked the available web servers on the 21 CASP16 targets (<400 nucleotides). As shown in Supplementary Fig. 2, the trRosettaRNA server outperforms the majority of other servers in inference speed. It is important to note that the total response time can include a variable queueing period, depending on server load and available computational resources. For example, for a 124-nucleotide RNA (R1255), our server returned an accurate model (4.5 Å RMSD) in 16 min from an input nucleotide sequence, which comprised 6 min of queueing, 7 min for MSA generation and 3 min for the final prediction. This performance demonstrates the practical effectiveness of the trRosettaRNA server for the research community.

In summary, trRosettaRNA combines competitive modeling accuracy with support for custom SS inputs. Performance is enhanced by exploration of diverse SSs, as well as the generation of specific structures with desired topologies. Furthermore, its structure optimization procedure guarantees the physicochemical plausibility of the predicted structures. However, physics-based methods retain unique advantages in modeling RNA dynamics through molecular dynamics simulations, a capability that current deep learning architectures largely struggle to replicate.

Overview of the procedures

We developed two procedures for the trRosettaRNA protocol. Procedure 1 offers an easy-to-use, web-based platform suitable for general users requiring no coding knowledge. Procedure 2 provides the standalone package containing the source code of trRosettaRNA, enabling large-scale applications or further development.

MSA generation

Input can be provided either as a single sequence or a precomputed MSA for both procedures. When a single sequence is provided, the protocol will automatically search for homologous sequences and construct an MSA from the RNAcentral database[24]. This search uses a two-round strategy inspired by RNAcmap[36]. Specifically, the server initially uses BLASTN[37] to perform a fast preliminary search with an *E*-value cutoff of 0.001. The resulting initial MSA, along with the SS predicted by RNAfold[38], is then used to construct a covariance model for a more sensitive Infernal[39] search (cmsearch), for which we use an *E*-value cutoff of 10 (the effectiveness of this cutoff is demonstrated in Supplementary Fig. 3). The final MSA from this process is fed into the structure modeling neural network. Compared to the previous version, which used rMSA and the NT+RNAcentral databases, the current MSA search pipeline is substantially faster without sacrificing quality (Supplementary Figs. 4 and 5). For example, for an RNA with about 200–300 nucleotides, the trRosettaRNA server can now generate an MSA with a median time of around 20 min, compared to a median of 14.1 h for rMSA. In addition, our pipeline performs comparably to, and in some cases better than, using the established Rfam MSAs[40] for 3D prediction (Supplementary Fig. 5c), which may be attributed to our more targeted and comprehensive homolog search for each specific target.

SS prediction

The default SS prediction in trRosettaRNA is handled by an in-house transformer neural network (trRNA-SS). This predictor uses eight stacks of RNAformer blocks to compute base-pairing probabilities from the input MSA. As demonstrated on the 16 CASP16 RNA targets (Supplementary Fig. 6a) and 39 independent RNAs (Supplementary Fig. 6b), trRNA-SS achieves the area under the precision-recall curve values of 0.796 and 0.854, respectively, substantially outperforming SPOT-RNA (0.697 and 0.745), the default SS predictor used in the original trRosettaRNA server. This improvement is further supported by Supplementary Fig. 6c, which shows that predictions improved for 72.7% (40/55) targets when using trRNA-SS. Further analysis reveals that this improvement is more pronounced for pseudoknots (Supplementary Fig. 6d), which can be attributed to the combined power of the MSA and the transformer network to resolve more complicated long-range interactions. These results establish trRNA-SS as a reliable default SS predictor for the trRosettaRNA pipeline.

Alternatively, both procedures accommodate the use of custom SS restraints for specialized applications (discussed in the 'Experimental design' section).

3D structure prediction and optimization

As shown in Fig. 1, the 3D structure is predicted using an end-to-end neural network. Specifically, the input features, MSA and SS, are initially converted into two representations of target RNA: an MSA representation and a pair representation. These two representations are then iteratively updated by a transformer network named RNAformer. During this process, the RNAformer also predicts 2D geometric information, including internucleotide distances and contacts. Finally, a structure module is used to predict the all-atom 3D structure from the updated representations.

As 3D structures generated by the neural network may contain steric violations, we leverage energy minimization to improve physicochemical plausibility. Specifically, in the web server (Procedure 1) structures with clashscore <200 are refined using fast relaxation. More severe clashes necessitate a full energy minimization guided by predicted 2D geometries, which was detailed in the 'Methods' section ('Step 3. Generation of full-atom structure models') of the trRosettaRNA methodological paper[19]. This strategy effectively reduces steric clashes, as demonstrated by an analysis of 3,000 recent server submissions (Supplementary Fig. 7). For the standalone package (Procedure 2), users can optimize the structure using more advanced or customizable approaches (discussed in the 'Experimental design' section).

Confidence score of the predicted structures

In the initially released server, the confidence score was defined as the eRMSD. This score was estimated based on two factors: the quality of the predicted distance maps and the convergence of the top decoys generated during the energy minimization. Although eRMSD correlated well with the actual RMSD, this score is no longer suitable for the end-to-end prediction workflow, which does not necessarily generate multiple decoys. Therefore, we now use an RNA quality assessment model[41] to estimate the LDDT[42] score (termed pLDDT) of the predicted 3D structure. This estimation is more flexible as it does not rely on the predicted 2D geometries or multiple decoys. Benchmark tests on CASP16 decoys show that pLDDT correlates well with the ground-truth RMSD value, yielding a Spearman correlation coefficient (ρ) of −0.797 (P value of 7.31×10^{-69}; Fig. 3a).

Examples of structure prediction

As shown in the final output in Fig. 5, for a four-way junctional twister-sister ribozyme (PDB UD: 5Y85), trRosettaRNA achieves a high-confidence 3D prediction (pLDDT of 91.99) and successfully captures the complex tertiary architecture. Characteristic structural features, including long-range base-pairing, are clearly recovered and depicted in the auxiliary results (Fig. 6).

In addition, deep learning inference accelerated by graphics processing units (GPUs) enables high-throughput prediction across numerous RNA targets. Meanwhile, reliable pLDDT scores facilitate quality assessment, crucial for exploring the vast number of RNAs with unknown 3D structures. Indeed, among the 3,938 Rfam families (version 14.4)[40], only about 3% (123) have solved 3D structures, indicating the insufficient exploration of the RNA 3D structure landscape. As an application, trRosettaRNA was used to predict the structures for 1,752 Rfam families with unknown 3D structures, which have a consensus length ranging from 50 to 200 nucleotides and have more than 30 members.

Across this set, trRosettaRNA achieved a median pLDDT of 62.2 (Fig. 3b) and generated high-confidence predictions (pLDDT >75) for 467 families. Two high-confidence examples are shown in Fig. 3c. For RF01684, a *MALAT1*-associated small cytoplasmic RNA, trRosettaRNA yielded a confident prediction with pLDDT of 93.5. Notably, this RNA was recently resolved by researchers (PDB ID: 8K0Y; released after the training data cutoff of trRosettaRNA). The comparison reveals that our prediction closely resembles the experimental structure, with an RMSD of 1.9 Å. Another example is RF02966, a DUF3268 RNA motif, for which trRosettaRNA generated a confident prediction with a pLDDT of 76.4. According to its Rfam description,

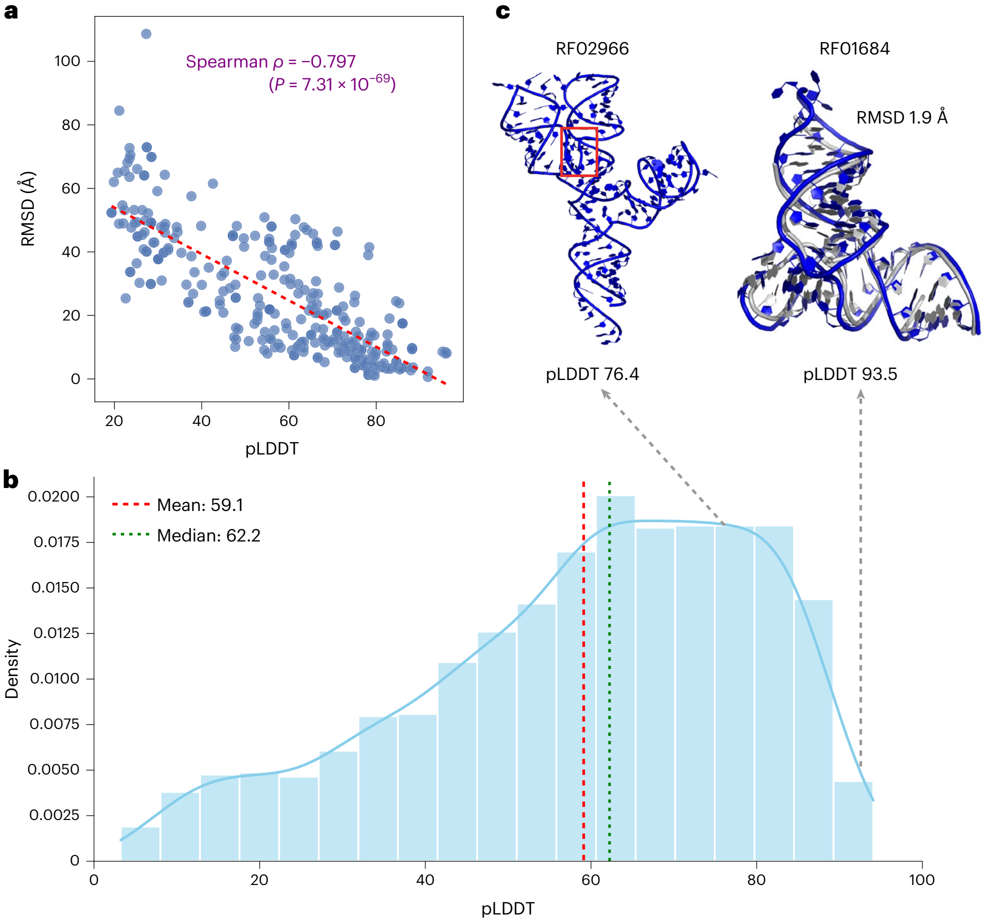


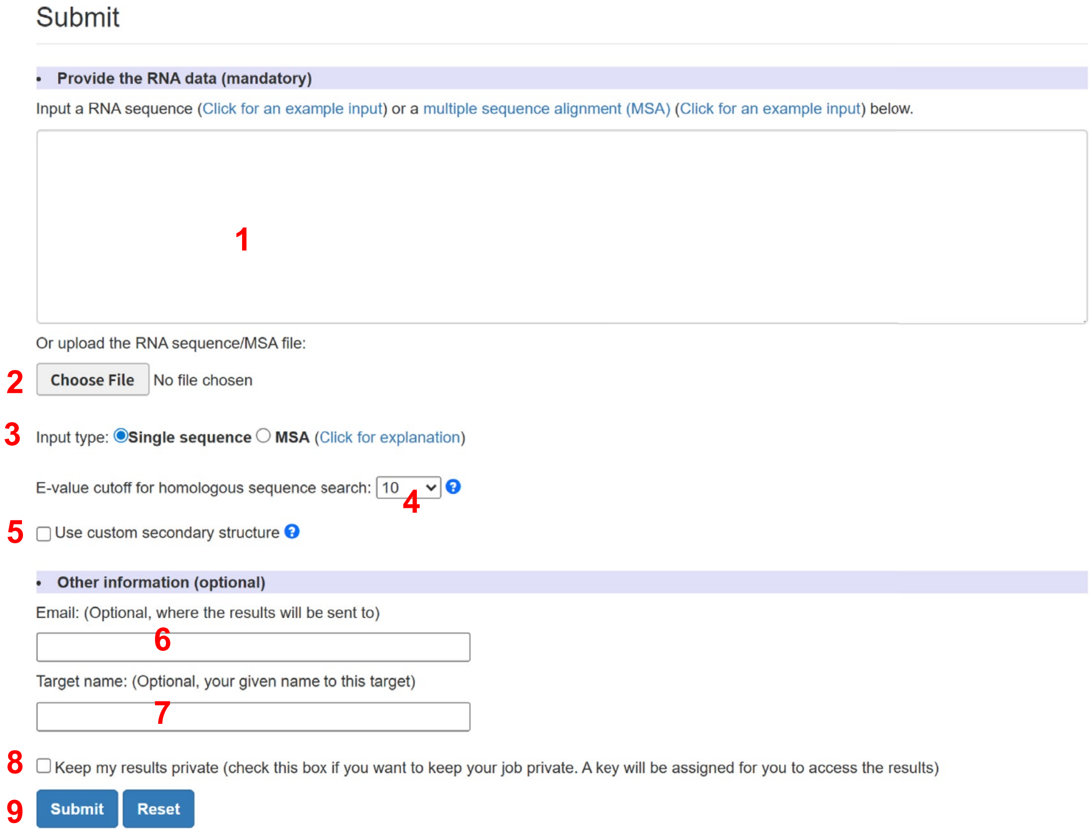
Fig. 3 | Confidence score (pLDDT) of trRosettaRNA and its application to Rfam families without known 3D structures. **a**, The relationship between pLDDT and actual RMSD for Yang-Server and Yang-Multimer models submitted in CASP16 ($n = 311$). **b**, Distribution of pLDDT of trRosettaRNA models for 1,752 Rfam families without known 3D structures ($n = 1{,}752$). **c**, Two examples (RF01684 and RF02966) of high-confidence trRosettaRNA predictions, shown as blue cartoons. For RF01684, the trRosettaRNA prediction is superimposed onto the recently solved experimental structure (PDB ID: 8K0Y; gray cartoon). The dashed gray lines indicate the pLDDT scores of these two examples within the distribution shown in **b**.

this RNA contains an experimentally validated kink turn, which was accurately captured by trRosettaRNA (highlighted by a red frame in Fig. 3c). These cases demonstrate trRosettaRNA's ability to generate predictions consistent with experimental findings. The 467 high-confidence predictions are accessible at https://yanglab.qd.sdu.edu.cn/trRosettaRNA/rfam/.

Experimental design

Custom MSA strategies

The default homology search against the RNAcentral database is optimized for speed and provides sufficient evolutionary information for most well-characterized RNA targets. However, providing a preconstructed, custom MSA is a critical design choice for orphan or novel RNAs that lack representation in public databases, as users can leverage internal sequencing data or sensitive offline searches against larger databases such as NT to bypass the limitations of online tools. This approach also substantially enhances computational efficiency during iterative testing of the same sequence and ensures rigorous benchmarking by maintaining structural consistency across different prediction platforms. Furthermore, although the default *E*-value cutoff of 10 offers the best overall performance, users can also explore stricter thresholds (for example, 0.01) for specific targets where a higher signal-to-noise ratio in homology detection is required (Supplementary Fig. 3).



Submit

- Provide the RNA data (mandatory)

Input a RNA sequence (Click for an example input) or a multiple sequence alignment (MSA) (Click for an example input) below.

Or upload the RNA sequence/MSA file:

No file chosen

Input type: ● Single sequence ○ MSA (Click for explanation)

E-value cutoff for homologous sequence search: 10

☐ Use custom secondary structure

- Other information (optional)

Email: (Optional, where the results will be sent to)

Target name: (Optional, your given name to this target)

☐ Keep my results private (check this box if you want to keep your job private. A key will be assigned for you to access the results)

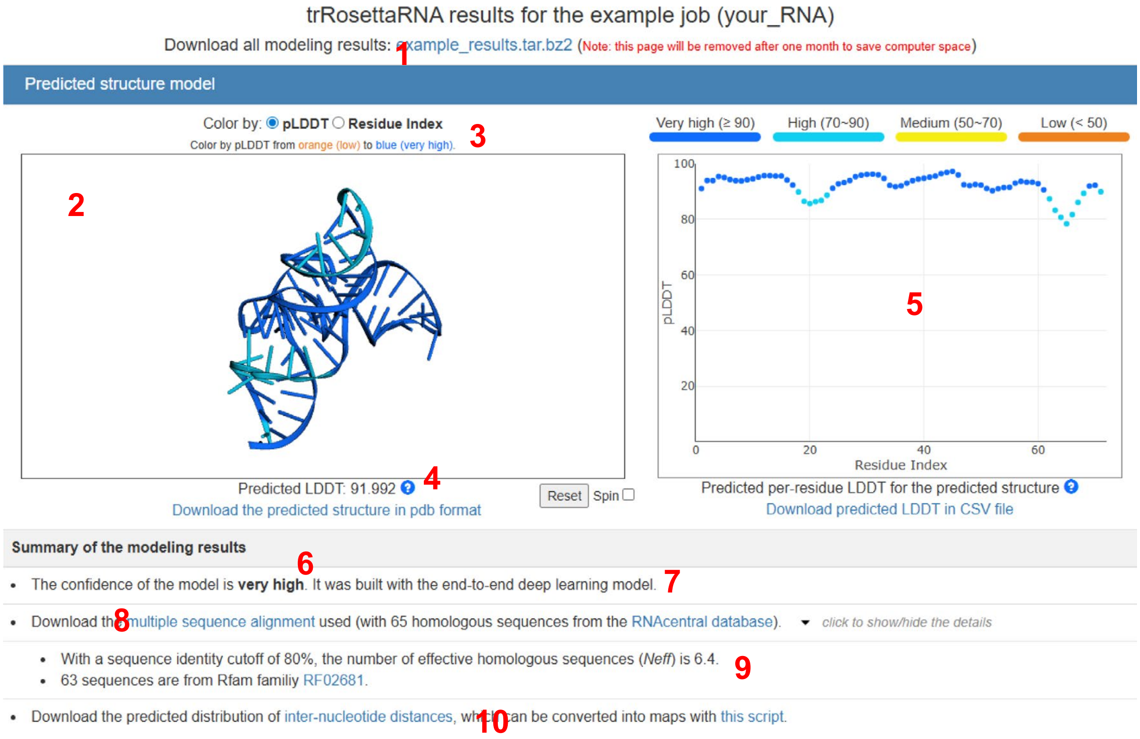
Fig. 4 | The trRosettaRNA homepage for job submission. Users should provide the nucleotide sequence or an MSA of the target RNA via string pasting (label 1) or file upload (label 2), then specify the input type (single sequence or MSA; label 3). For single sequence inputs, users can also choose an *E*-value cutoff for homolog search (label 4). Optionally, a custom SS can be provided via label 5. Entering an email address (label 6) allows users to receive job status notifications, and specifying a target name (label 7) aids in better tracking of results. By default, results are publicly accessible; check the box (label 8) if you want to keep the results private, which will assign a password upon submission. Finally, users can click the 'Submit' button (label 9) to submit and monitor the job.

Integration of custom SS

trRosettaRNA uses the internal trRNA-SS predictor by default, but users can choose to provide custom SS in dot-bracket or CT format to guide the modeling towards desired 3D structures. This design choice is particularly effective when incorporating the experimental SS restraints derived from SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension), dimethyl sulfate or other chemical probing techniques, which can substantially improve the accuracy and physical plausibility of the resulting 3D models. Beyond experimental guidance, this approach theoretically facilitates the exploration of different functional states of an RNA by modulating the input SSs to reflect alternative folds. Crucially, our results in CASP16 demonstrated that exploring an ensemble of diverse SS inputs increases the likelihood of yielding higher 3D structural accuracy compared to relying on a single, fixed SS prediction[21]. In all cases, users must ensure that the custom SS input corresponds exactly to the length of the query sequence to prevent potential execution errors during the modeling process (see Tables 3 and 4 for details).

Choice between end-to-end prediction and PyRosetta energy minimization

The trRosettaRNA server uses an automated pipeline to determine whether to perform PyRosetta-powered energy minimization based on the steric clash level detected in the raw end-to-end prediction. This dual-track strategy balances computational speed with structural plausibility, though users must consider the increased runtimes of energy minimization, particularly for long sequences exceeding 400 nucleotides. For high-throughput applications, the fast, GPU-accelerated end-to-end mode is generally preferred, whereas the standalone



trRosettaRNA results for the example job (your_RNA)

Download all modeling results: example_results.tar.bz2 (Note: this page will be removed after one month to save computer space)

Predicted structure model

Color by: ● pLDDT ○ Residue Index

Color by pLDDT from orange (low) to blue (very high).

Predicted LDDT: 91.992

Download the predicted structure in pdb format

Predicted per-residue LDDT for the predicted structure

Download predicted LDDT in CSV file

Summary of the modeling results

- The confidence of the model is **very high**. It was built with the end-to-end deep learning model.
- Download the multiple sequence alignment used (with 65 homologous sequences from the RNAcentral database). *click to show/hide the details*
  - With a sequence identity cutoff of 80%, the number of effective homologous sequences (*Neff*) is 6.4.
  - 63 sequences are from Rfam familiy RF02681.
- Download the predicted distribution of inter-nucleotide distances, which can be converted into maps with this script.

Fig. 5 | The main interface on the trRosettaRNA results page for an example RNA (PDB ID: 5Y85). This part provides a download link for the results archive (label 1) and visualizes the predicted 3D structure in a cartoon (label 2), colored by pLDDT value (default) or nucleotide index (optional; label 3). The overall pLDDT (label 4) and nucleotide-wise pLDDT (label 5) are also shown. In addition, a results summary is given, including confidence level (label 6), the modeling method (label 7; end-to-end or energy minimization), the MSA download link (label 8) and statistics (label 9), and a download link for distance predictions (label 10; between five atom pairs: P, C3′, C1′ and two base atoms N1/9, C4/2 for pyrimidine/purine).

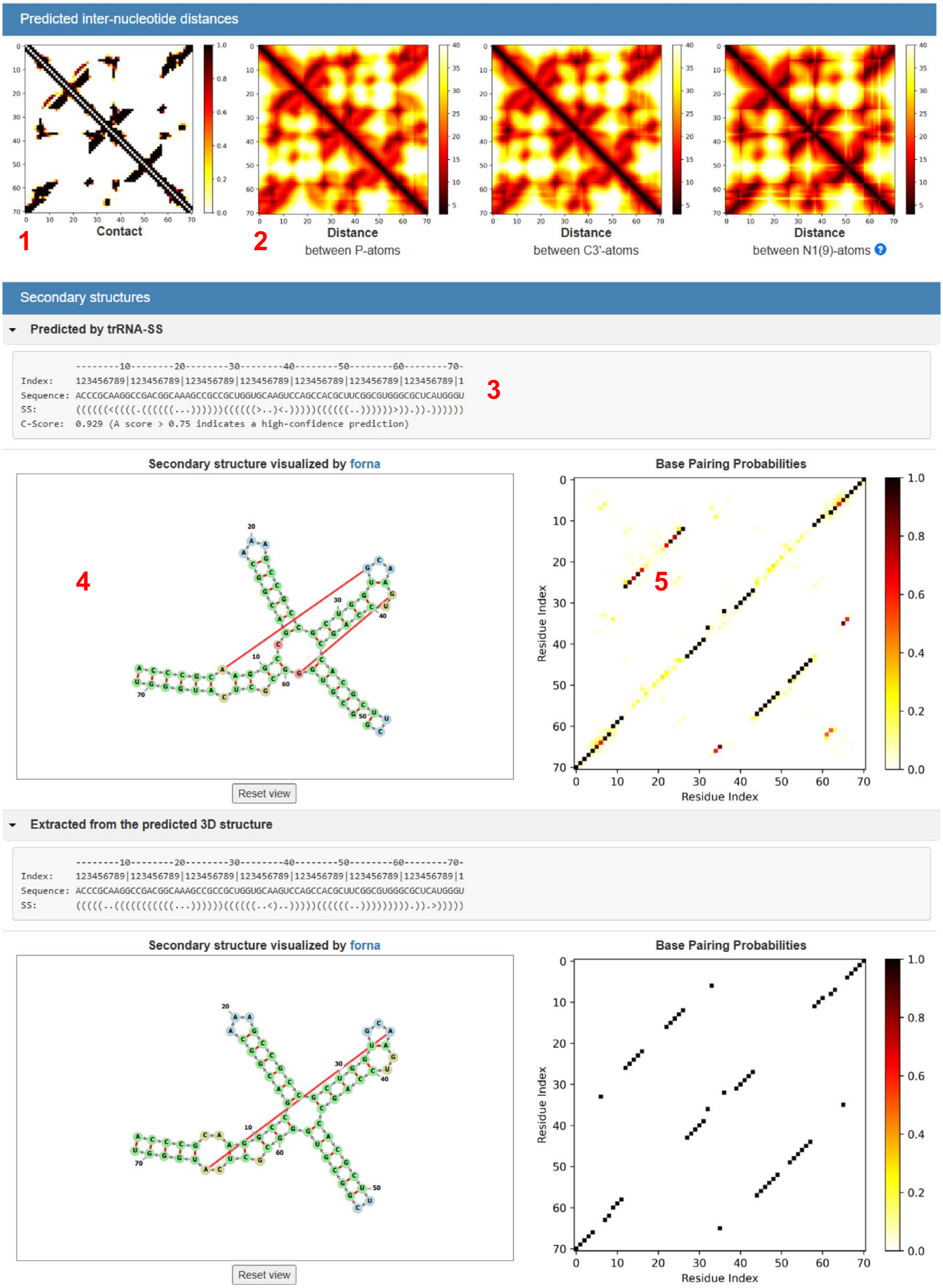
package offers users the flexibility to inspect structural violations and perform optimization on a case-by-case basis. Specialized tasks may further require customizing the number of candidate models or optimization steps to fine-tune the resulting energy landscape, as detailed in the parameter descriptions in Box 1.

Limitations and perspectives

Though trRosettaRNA represents an effective automated RNA 3D structure prediction protocol, it shares limitations common to all automated methods.

First, as highlighted in CASP15, automated approaches struggle with RNAs possessing complex topologies and low sequence homology to known structures, such as the four challenging synthetic targets[43]. According to our earlier analysis[19], this challenge is common to all automated approaches, including both deep learning and traditional methods. Encouragingly, template information seems beneficial for these difficult cases, an observation supported by our CASP16 results, which are detailed in our recent report[21]. However, for the large RNAs without available templates (such as the CASP16 oligomer targets), both automated servers and human experts struggled to consistently generate accurate predictions (Supplementary Table 1), suggesting the substantial challenges in modeling such large targets. This limitation also restricts the rigorous validation of de novo designed sequences, a sophisticated downstream task of RNA structure prediction.

Second, unlike AlphaFold 3 and RoseTTAFoldNA, trRosettaRNA does not explicitly model interactions with binding partners (proteins, ligands), a factor that could influence the accuracy of the predicted RNA structure. It is noteworthy, however, that including binding partners did not always enhance the performance of AlphaFold 3. For instance, in the cases of the ligand-binding RNA R1288 and the protein-binding RNA R1293 from CASP16, modeling the RNA as part of a complex yielded lower accuracy than single-chain predictions[20]. This performance



Predicted by trRNA-SS

```
          --------10--------20--------30--------40--------50--------60--------70-
Index:    123456789|123456789|123456789|123456789|123456789|123456789|123456789|1
Sequence: ACCCGCAAGGCCGACGGCAAAGCCGCCGCUGGUGCAAGUCCAGCCACGCUUCGGCGUGGGCGCUCAUGGGU
SS:       ((((((<((((.((((((...))))))((((((>..)<.)))))((((((..))))))>)).)).))))))
C-Score:  0.929 (A score > 0.75 indicates a high-confidence prediction)
```

Extracted from the predicted 3D structure

```
          --------10--------20--------30--------40--------50--------60--------70-
Index:    123456789|123456789|123456789|123456789|123456789|123456789|123456789|1
Sequence: ACCCGCAAGGCCGACGGCAAAGCCGCCGCUGGUGCAAGUCCAGCCACGCUUCGGCGUGGGCGCUCAUGGGU
SS:       (((((..(((((((((((...))))))((((((..<)..)))))((((((..))))))))).)).>)))))
```

Fig. 6 | Auxiliary prediction outputs from trRosettaRNA for an example RNA (PDB ID: 5Y85). The auxiliary results include the predicted contact (label 1) and distance (label 2) maps, alongside the input/predicted SS and the SS extracted from the predicted 3D structure (labels 3–5).

BOX 1

Representative optional arguments (parameters) of ‘trRNA2.predict’

- `-nrec, --num_recycles`: a non-negative integer controlling the number of network recycles (default: 3)
- `-mid, --return_mid`: whether to save outputs from intermediate cycles (default: False). To activate, use the flag without a value (i.e., ‘`-mid`’ or ‘`--return_mid`’)
- `-nrows, --nrows`: a positive integer controlling the maximum number of sequences included in the MSA representation for the transformer network (default: 500). A higher value may slightly improve performance at the cost of more computational resources. Conversely, this provides an option to save memory for large RNAs with deep MSAs
- `-relax_steps, --relax_steps`: a non-negative integer controlling maximum steps of fast relaxation. The default value is conditional: if the ‘`–pyrosetta`’ flag is used, this is automatically set to 0 (skipping relaxation); otherwise, it defaults to 200

The following options are only applicable when using the ‘`--pyrosetta`’ mode:

- `-nm, --nmodels`: a positive integer controlling the number of decoy models to generate during energy minimization (default: 5). Increasing this value can potentially improve performance by enlarging the decoy set for scoring and selection, but at the cost of increased computational time and resources
- `-dcut, --dcut`: a float between 0 and 1 controlling the probability threshold for including distance restraints in the energy function (default: 0.45). Although the default value yields good overall performance, adjusting this value can be used to generate more diverse decoy sets
- `-tmp, --tmpdir`: specifies a directory for storing intermediate files generated during energy minimization (e.g., decoy models and restraint files). If not provided (default), these files are automatically removed upon task completion
  For a complete list of all optional arguments, please refer to the help message by running ‘`python -m trRNA2.predict -h`’.

degradation probably stems from the fact that the inaccurate modeling of novel intermolecular interfaces, particularly those lacking homologs in the training set, can paradoxically distort the individual RNA fold and lead to lower accuracy than single-chain predictions.

Considering the limited availability of 3D structural data for RNA and its related complexes, coupled with the inherent conformational flexibility of these systems, a promising future direction lies in the closer integration of deep learning with experimental restraints and structural templates. For example, trRosettaRNA generates heterogeneous conformational ensembles by incorporating diverse SS restraints[20]. Similarly, deep learning has been successfully applied to determine RNA 3D structures and conformers from cryo-EM and atomic force microscopy data[44,45]. Furthermore, both the CASP16 assessment[22] and a recent blind Kaggle challenge[46] underscore the vital role of structural templates when available. We anticipate that trRosettaRNA can be further enhanced through the effective incorporation of template and experimental information.

Materials

Equipment

Procedure 1

- Computer: a personal computer with internet connectivity and a web browser such as Google Chrome, Microsoft Edge, Firefox or Safari
- Input data: the nucleotide sequence or a precomputed MSA for a target RNA of interest and, optionally, the SS in dot-bracket or CT format

Procedure 2

- Hardware/OS: a personal computer or computer cluster with a supported Linux distribution (for example, CentOS, Red Hat, Ubuntu)
- GPU (optional but recommended): a GPU is not mandatory, but it will accelerate the neural network computations considerably compared to CPU-only execution

Software for Procedure 2

Local execution of trRosettaRNA requires the prior installation of Mamba (https://mamba.readthedocs.io/en/latest/installation/mamba-installation.html). Although Conda (https://www.anaconda.com/docs/getting-started/miniconda/install) can also be used, Mamba is recommended as it is typically much faster. If you already have Conda installed, please install Mamba into the base environment with ‘`conda install -n base mamba -c conda-forge`’ to avoid potential conflicts, rather than performing a separate Mamba installation.

Subsequently, the trRosettaRNA package can be downloaded from https://github.com/YangLab-SDU/trRosettaRNA2.

Procedure 1: web-based platform

Job submission

● **TIMING** 3 min

1. Navigate to the trRosettaRNA server website at https://yanglab.qd.sdu.edu.cn/trRosettaRNA/.
2. In the section ‘Provide the RNA data (mandatory)’ (label 1 in Fig. 4), either paste your target sequence (in FASTA format) or a precomputed MSA (in A3M, FASTA, A2M or STO format) or upload the corresponding data file by clicking ‘Choose File’ (label 2 in Fig. 4).
    - Specify the input type for your RNA data (select either ‘Single sequence’ or ‘MSA’; label 3 in Fig. 4).
    - For single sequence inputs, you can choose an *E*-value cutoff for homolog detection (label 4 in Fig. 4).

    ▲ **CRITICAL STEP** trRosettaRNA provides an option to input your custom MSA. Selecting the correct input type is crucial. Choose ‘Single sequence’ if you are providing only the nucleotide sequence of the target RNA. The server will then automatically generate the required MSA. Choose ‘MSA ’ if you are providing a pre-computed MSA. If ‘Single sequence’ is wrongly selected when the MSA is provided, the input will be misinterpreted as a multichain file and only the first sequence will be processed.

    ▲ **CAUTION** The trRosettaRNA server supports RNAs with lengths between 10 and 1,000 nucleotides. Only canonical nucleobases (A, U, C, G, T) are supported as input. The noncanonical nucleobases in the query sequence will be automatically removed. If they appear at the homologous sequences in the submitted MSA, they will be replaced with gaps (‘-’). Note: if the provided MSA exceeds 20,000 sequences, it will be pruned to retain the first 20,000. As shown in Supplementary Table 4, this pruning does not negatively affect performance.

    ◆ **TROUBLESHOOTING**
3. (Optional) Provide your custom SS. Supported formats include dot-bracket notation and CT (label 5 in Fig. 4).

    ▲ **CRITICAL STEP** Although trRosettaRNA’s internal SS predictor is highly accurate, providing experimentally determined SS can further improve 3D structure prediction accuracy. Furthermore, using diverse SS inputs enables the generation of heterogeneous 3D conformations, which is essential for studying RNA conformational landscapes.

    ▲ **CAUTION** The server automatically checks the validity and consistency of the provided SS format. For dot-bracket notation, the server supports various base-pair symbols, including standard paired brackets ‘()’, ‘[]’, ‘{}’, ‘<>’ and alphabet-based pairs such as ‘Aa’. Unpaired nucleotides can be represented by ‘.’, ‘,’, ‘_’, ‘:’ or ‘-’. Input containing unsupported characters or improper nesting will result in an error, requiring the user to correct and resubmit the job.

For the CT file, the first line must contain the sequence length, followed by any title information. Subsequent lines describe base pairing, starting from the second line, and must use 1-based nucleotide indexing (that is, begin with 1). Each data line should contain the nucleotide index, the base identity, the previous index, the next index, the pairing partner index and historical numbering (often the same as the nucleotide index). A detailed description of the above two formats is provided at https://yanglab.qd.sdu.edu.cn/trRosettaRNA/ss_format.html.

◆ **TROUBLESHOOTING**

4. (Optional) Provide your email address (label 6 in Fig. 4) to receive notifications. You will receive a confirmation email upon job submission and a notification email once the job is complete.

   ◆ **TROUBLESHOOTING**

5. (Optional) Specify a target name (label 7 in Fig. 4) for display on the results page (defaults to 'your_RNA' if unspecified).
6. (Optional) Configure results privacy (label 8 in Fig. 4). Results are public by default. If you require confidentiality, select the 'Keep my results private' checkbox. A unique password will then be generated upon submission, which is necessary to view private results.
7. Click on the 'Submit' button to complete the submission (label 9 in Fig. 4). A confirmation email will be sent to the email address specified in Step 4 (if provided).

   ◆ **TROUBLESHOOTING**

Job monitoring

● **TIMING depends on RNA size (~1 h for an RNA with ~200 nucleotides)**

8. Upon submission, a confirmation page will appear, displaying your job ID and a URL link to track modeling status and view results. Private jobs (Step 6 box checked) will provide two links: one requiring the password and one direct link (password included). Public jobs provide a single direct link. Bookmark the direct results link generated upon submission or provide an email address (Step 4) to receive the access details automatically.

   ◆ **TROUBLESHOOTING**

9. Use the URL from Step 8 to access the job-monitoring page, which refreshes automatically every minute. Track your job's progress via the status indicator (current stage highlighted in red) and note the estimated completion time. This estimate references historical data correlating sequence length and runtime for recent submissions (see Supplementary Fig. 8, data up to November 2025). For reference, an RNA of about 200 nucleotides typically takes about 1 h to finish prediction. The entire pipeline consists of six stages, starting with queuing and ending with job completion. You may leave the monitoring page open or bookmark the URL to view results later. When the job is completed, the results will appear automatically on this page. An email notification containing the results link will also be sent if you provided an email address in Step 4.

   ◆ **TROUBLESHOOTING**

Results analysis

● **TIMING 5 min**

10. After completion, you can download the compressed results archive (.tar.bz2) provided at the top of the results page (label 1 in Fig. 5). Save this file to your local storage as the server results will be deleted after 1 month to save storage space.
11. View the 'Predicted structure model' section.
    - On the left-hand side (label 2 in Fig. 5), the predicted 3D structure is visualized as a cartoon using 3Dmol[47], colored by pLDDT confidence scores, which range from orange (low) to blue (very high). Alternatively, users can click the 'Residue Index' option (label 3 in Fig. 5) to switch the coloring to a standard rainbow spectrum (purple, 5′ terminus to red, 3′ terminus). Users can download this structure in PDB format by clicking on the links below the predicted lDDT.
    - A confidence score (pLDDT) is displayed below the 3D structure visualization (label 4 in Fig. 5). This score ranges from 0 to 100; higher values indicate greater confidence in the prediction.

  - The right-hand side shows the per-nucleotide pLDDT plot (label 5 in Fig. 5). Nucleotides are color coded according to confidence levels: very high (pLDDT ≥90), high ($70 \leq$ pLDDT < 90), medium ($50 \leq$ pLDDT < 70) and low (pLDDT <50). Hovering the mouse pointer over a point on this plot will display the pLDDT value for that nucleotide and highlight the corresponding nucleotide in the 3D structure view. Users can download the CSV file storing the pre-nucleotide pLDDT score by clicking on the links provided below this plot.
  - The bottom of this section provides a summary of the modeling results, including the confidence level of prediction (label 6 in Fig. 5), the modeling approach (end-to-end or energy minimization; label 7 in Fig. 5), the MSA information (labels 8–9 in Fig. 5) and a download link for the predicted internucleotide distances (label 10 in Fig. 5). For MSAs generated directly by the web server, key statistics such as the effective number of homologous sequences (N_{eff}) and the most frequent Rfam family hit are also presented (label 9 in Fig. 5).

  ◆ **TROUBLESHOOTING**

12. View the 'Predicted internucleotide distances' section.
  - This section visualizes the predicted 2D contact and distance maps.
  - The contact map (label 1 in Fig. 6) presents the probability of each nucleotide pair being in contact. A contact is defined when the distance between any heavy atoms of the two nucleotides is below 8 Å.
  - The distance maps (label 2 in Fig. 6) visualize the predicted distances between specific atom pairs (P–P, C3′–C3′ and the glycosidic nitrogen atoms; N1 for pyrimidines or N9 for purines). The distance values are calculated as weighted averages based on the predicted probability distributions among 38 bins, capped within the 4 Å to 40 Å range.
13. View the 'Secondary structures' section.
  - This section visualizes the input and output SSs. Each SS is visualized in three ways. First, users can click on the triangle icon at the top of each panel to see the dot-bracket notation of the corresponding SSs (label 3 in Fig. 6). Second, a 2D force-directed graph is generated using the software forna[48] (label 4 in Fig. 6). Third, a 2D map is plotted displaying the base-pairing probabilities for all nucleotide pairs (label 5 in Fig. 6).
  - The source of the input SS can be one of two types: 'Predicted by trRNA-SS' (the default internal prediction) or 'Input by user' (if a custom SS was provided in Step 3). For the 'Predicted by trRNA-SS' input, a confidence score is presented (where a value > 0.75 signifies a high-confidence prediction), and the corresponding 2D map displays predicted probabilities (ranging from 0 to 1). For the 'Input by user' SS, the 2D map shows a binary representation (values are either 0 or 1, indicating the presence or absence of a base pair).
  - The output SS is extracted from the predicted 3D structure using the DSSR software[49]. Its corresponding 2D map is also binary.

Procedure 2: standalone package containing the source code of trRosettaRNA

Installation

● **TIMING** 10 min

1. Clone the trRosettaRNA repository from GitHub and change into the directory

```
» git clone https://github.com/YangLab-SDU/trRosettaRNA2.git
» cd trRosettaRNA2
```

▲ **CAUTION** The installation steps (Steps 1–4) only need to be performed once. If you have already completed this setup, you can skip directly to Step 5.

◆ **TROUBLESHOOTING**

2. Install and activate the required environment. This step will create a dedicated environment named trRNA2 with all necessary dependencies listed in the environment.yml file.

```
»mamba env create -f environment.yml
»mamba activate trRNA2
```

◆ **TROUBLESHOOTING**

3. Download and extract the network parameters, which are needed for the deep learning models to function correctly.

```
»wget http://yanglab.qd.sdu.edu.cn/trRosettaRNA/download/params_trRNA2.tar.bz2
»tar -jxvf params_trRNA2.tar.bz2
```

◆ **TROUBLESHOOTING**

4. (Optional) Download the sequence database used for MSA generation. This step is only required if you intend for trRosettaRNA to generate MSAs automatically. You can safely skip this if you provide your own precomputed MSA files.

```
»bash scripts/download_database.sh $db_dir/
```

Replace '$db_dir' with the path to your desired storage location for the database. The database files will be uncompressed into the '$db_dir/library/' directory.
▲ **CAUTION** The uncompressed database requires about 32 GB of disk space.
◆ **TROUBLESHOOTING**

Input preparation

● **TIMING 10 min**

5. Prepare your target sequence in FASTA format (for example, 'example/seq.fasta'). Then, run the 'scripts/search_MSA.sh' script to generate the MSA required by the model.

```
»bash scripts/search_MSA.sh example/seq.fasta $msa_out_dir $db_dir/
library/rnacentral_99_rep_seq.fasta 4
```

Replace '$msa_out_dir' with the path to the desired directory to store the generated MSA, which will be saved as an A3M file '$msa_out_dir/seq.a3m'.
▲ **CAUTION** Only the standard nucleobases A, U, C, G, T and N (for unknown) are supported in the input FASTA sequence. Using other characters will cause errors during the structure prediction step (Step 7). Note: similar to Protocol 1, if the provided MSA exceeds 20,000 sequences, it will be pruned to retain the first 20,000.

6. (Optional) Prepare custom SS. The standalone package accepts SS input in several formats: dot-bracket, CT, bpseq or a plain text file representing the base-pairing probability matrix.
▲ **CAUTION** The validation rules described for the web server (Procedure 1) also apply here regarding allowed symbols and proper nesting for dot-bracket, and header/indexing requirements for CT files. Errors require correction and rerunning of the script. In addition, Procedure 2 additionally supports the input in bpseq format and the probability matrix (TXT file) format, which are not available on the web server for stability reasons. We provide the example files for all supported formats in the directory 'example/SS/' within the package.

Structure prediction

● **TIMING 1 min (end-to-end version), 9 min (PyRosetta version)**

There are different options that can be chosen when running the structure prediction script. These are described in Steps 7–10.

7. (option A) Run the structure prediction script for a basic end-to-end prediction:

```
»python -m trRNA2.predict -i $msa_out_dir/seq.a3m -o $output_dir
```

Replace '$output_dir' with the directory where output files will be saved. This directory will be created automatically if it does not exist.

▲ **CAUTION** By default, the prediction script will automatically use the first available GPU (CUDA device 0) if detected. If no GPU is found, it defaults to using 5 CPU threads. You can use the argument '`-gpu <ID>`' to select a specific GPU or '`-cpu <N>`' to set the number of CPU threads.

◆ **TROUBLESHOOTING**

8. (Option B) If you find the basic end-to-end prediction exhibits substantial structural violations (for example, clashscore>20), you can alternatively run the PyRosetta-powered energy minimization using the '`-pyrosetta`' flag. This approach may potentially yield more physically plausible structures with slightly improved accuracy, but it comes at the cost of longer runtimes compared to the default end-to-end method.

```
»python -m trRNA2.predict -i $msa_out_dir/seq.a3m -o $output_dir
-pyrosetta -fas example/seq.fasta
```

▲ **CAUTION** The '`-fas <fasta_file>`' argument, providing the query sequence in FASTA format, is mandatory when using the '`-pyrosetta`' flag.

◆ **TROUBLESHOOTING**

9. (option C) To provide your own SS restraints (prepared in Step 6), use the '`-ss`' argument to specify the file path and the '`-ss_fmt`' argument to indicate its format. Here is an example (assuming your SS file is 'example/seq.dbn' in dot-bracket format):

```
»python -m trRNA2.predict -i $msa_out_dir/seq.a3m -o $output_dir -ss
example/seq.dbn -ss_fmt dot_bracket
```

▲ **CAUTION** The '`-ss_fmt`' argument must be correctly assigned according to the format of input SS. Use the following values:

'`-ss_fmt dot_bracket`' for dot-bracket notation files
'`-ss_fmt ct`' for CT files
'`-ss_fmt bpseq`' for bpseq format files
'`-ss_fmt prob`' for base-pairing probability matrix (TXT) files

Incorrect format specifications will lead to errors or incorrect interpretation of the input SSs.

◆ **TROUBLESHOOTING**

10. (option D) Refer to Box 1, which details several advanced options that are unavailable in Procedure 1.

Results analysis

● **TIMING 2 min**

11. After the structure prediction (Steps 7–10), results are stored in the designated output directory ($output_dir).
    - For the standard end-to-end prediction, the main outputs are the final 3D structure, saved as '$output_dir/model_1_relaxed200.pdb' and the predicted 2D distances in '$output_dir/model_1_2D.npz'
    - If the '-pyrosetta' flag was selected, the predicted 3D structure will be saved to '$output_dir/model_1_pyrosetta.pdb' instead, and the PDB files for the top 5 decoys will be saved into the '$output_dir/pyrosetta_top5/' directory
12. Check the confidence score (pLDDT) of the predicted 3D structure. It will be printed at the standard output from the prediction run (Steps 7–10). Alternatively, per-nucleotide pLDDT scores are saved in a CSV file '$output_dir/plddt.csv'. The global pLDDT value reported in the standard output is the average of these per-nucleotide scores.

Troubleshooting

Troubleshooting advice can be found in Tables 3 and 4 for procedures 1 and 2 respectively.

Table 3 | Troubleshooting table for the web server

Step	Problem	Possible reason	Solution
2, 7	Error: lengths of sequences are not equal. Please check your input MSA and try again	Sequences in the submitted MSA have inconsistent lengths relative to the query sequence	Check the submitted MSA. Verify that all sequences are properly aligned and have the same length as the query. Fix the MSA file and resubmit
3, 7	Error: length mismatch. The provided custom SS has a length which does not match the input sequence length. Please ensure the SS definition corresponds exactly to the sequence length	The length of the submitted SS does not match the length of the target nucleotide sequence	Ensure consistency between your inputs: the SS must have the same number of nucleotides as the target RNA sequence. Adjust the SS or sequence/MSA input accordingly and resubmit
	Error: some brackets are not matched in the provided dot-bracket notation! Please check and resubmit	The submitted dot-bracket notation contains mismatched brackets	Check and correct the nesting in your dot-bracket notation and resubmit the job to ensure prediction uses your intended input
	Error: a format error occurred in the provided CT file. Please check the file and resubmit	The provided CT file contains a format error, which may involve length, nucleotide index, and base-pairing consistency	Check and ensure your CT file: (1) has a length matching the input sequence and correctly declared on the first line; (2) uses 1-based nucleotide indexing; (3) contains valid, self-consistent base pairing information (reciprocal pairs, valid indices, no conflicts)
4, 7, 9	Do not receive the notification email despite providing an email address during submission	Owing to the restriction of Google usage in the mainland of China, our system may fail to send an email notification to Google-hosted email addresses (e.g., Gmail)	Bookmark the results page to access the modeling results later, or use a different email address
	Error messages about the length/SS after submitting with a specified email address	Using your browser's autofill feature for the email field can sometimes unintentionally overwrite or alter the data previously entered in the sequence, MSA or SS input boxes	Beware of browser autofill. After entering your email using autofill, always double-check that your sequence, MSA, and SS inputs have not been accidentally modified. Ensure all data is correct before clicking 'Submit'
8	Lose the password or the results link	Failure to record or save the results link and password upon submission	Email us with your job ID for manual retrieval of the results link
11	Prediction results are based on a single sequence only	Could not find/use an MSA. Either no homologs were detected, or 'MSA' was selected as the input type for a single sequence submission (Step 2)	Try submitting a pre-computed MSA (select 'MSA' input type), or if submitting a single sequence, ensure 'Single sequence' is selected (Step 2) and resubmit. If there are still no homologs upon resubmission, do not worry, as the MSA contribution is often minor for trRosettaRNA compared to the SS input
	The confidence score (pLDDT) is low	The predicted 3D structure is possibly inaccurate. This may stem from inaccuracies in the input MSA or SS, or inherent challenges with the target RNA itself (e.g., complex topology, extensive interaction with proteins)	Try using a custom MSA/SS (from experiment or literature). Alternatively, for large targets, split into domains and model separately
	Input contained multiple sequences/chains, but only the first was modeled	Multi-chain and batch inputs are not currently supported. The server processes only the first sequence found	If modeling an RNA complex (multimer), manually concatenate the chains (e.g., with linkers such as 'AAAAA') and submit the combined sequence. If processing multiple independent RNAs, submit each sequence as a distinct job

Table 4 | Troubleshooting table for the standalone package

Step	Problem	Possible reason	Solution
1	Failed to clone the repository using git clone	This could be due to Git not being installed or network problems accessing GitHub	First, ensure Git is installed and your network connection is working, then retry the git clone command If cloning still fails or you prefer not to use Git, download the code directly: latest code (ZIP) https://github.com/YangLab-SDU/trRosettaRNA2/archive/refs/heads/main.zip or the stable releases https://github.com/YangLab-SDU/trRosettaRNA2/releases, then uncompress
2	Mamba command not found or fails	Mamba is probably not installed correctly, not initialized properly in your shell environment, or its installation path is not included in your system's PATH variable	Reinstall Miniforge from the official source, carefully following their instructions, especially regarding shell initialization (e.g., running '`mamba init`'). If you have pre-installed Conda, you can run '`conda install -n base mamba -c conda-forge`' instead. As a last resort, manually create a new Conda or virtual environment and install the packages listed in the 'environment.yml' file using '`conda install <package>`' or '`pip install <package>`'

Table 4 (continued) | Troubleshooting table for the standalone package

Step	Problem	Possible reason	Solution
3, 4	Timeout when downloading model weights or sequence database	Network issues or regional download restrictions	Download the files from our official Hugging Face mirror: https://huggingface.co/datasets/quailwwk/trRNA2/. The package includes a dedicated script for the sequence database: '`bash scripts/download_database_huggingface.sh`'
7	GPU not used by PyTorch, even if specified	Incorrect NVIDIA driver or CUDA toolkit setup (missing, outdated, or incompatible)	Confirm that an NVIDIA GPU is installed in your system. Update or install the NVIDIA driver (≥460) and CUDA toolkit (≥11.2) from https://developer.nvidia.com/cuda-downloads. If GPU setup fails or is unavailable, run on CPU ('`-gpu -1 -cpu <number_threads>`'), which will be slower than on GPU
	Out of memory	The input MSA is too large to fit into the available GPU memory	Prune your input MSA (e.g., using HHfilter[60]; see Supplementary Table 4 for the performance comparison), reduce the recycling times (adjust the '`-nrec <num_recycle>`' argument), or use CPU instead
8	Error: Unrecognized nucleotide: X	The input sequence includes unsupported symbols	Check and ensure that the input sequence only contains nucleobases in (A, U, C, G, T, N)
9	ValueError: The lengths of the input sequence, SS, and MSA are inconsistent	Length mismatch between the input SS and MSA	Check and ensure that the input SS and MSA share the same number of nucleotides
	ValueError: Unknown notation in dbn: ?	The input dot-bracket notation contains unknown symbols	Check and ensure that the input dot-bracket notation only contains valid symbols described in https://yanglab.qd.sdu.edu.cn/trRosettaRNA/ss_format.html
	ValueError: Provided dot-bracket notation is not completely matched!	The input dot-bracket string contains improperly nested or unbalanced base pairs, e.g., more opening than closing brackets, or pairs that cross like '([...)])'	Carefully review the submitted dot-bracket notation. Ensure that all opening brackets have a corresponding, correctly positioned closing bracket, and that base pairs do not cross incorrectly

Timing

Procedure 1

Steps 1–7, job submission: 3 min
Steps 8–9, job monitoring: depends on RNA size (~1 h for an RNA with ~200 nucleotides)
Steps 10–13, results analysis: 5 min

Procedure 2

Steps 1–4, installation: 10 min
Steps 5–6, input preparation: 10 min for an RNA with ~70 nucleotides
Steps 7–10, structure prediction: 1 min (end-to-end version) and 9 min (PyRosetta version) for an RNA with ~70 nucleotides
Steps 11–12, results analysis: 2 min

Anticipated results

Definitions for metrics used in Fig. 2

The metrics used in Fig. 2 are primarily from the official CASP assessment and are defined as follows. For Fig. 2b, we recalculated these metrics locally to ensure consistency with the traditional approaches, which were run locally rather than assessed in CASP. The specific definitions are as follows:

RMSD

The RMSD between the predicted model and the experimental structure, calculated on all nonhydrogen atoms after an optimal, sequence-dependent superposition using the Kabsch algorithm[50]. For Fig. 2b, RMSD scores were calculated using the RNA_assessment tool[51] based on all heavy atoms. RMSD ranges from 0 (perfect match) to any positive value, with a lower RMSD indicating a better model.

GDT-TS

A distance-based metric calculated as an average fraction of residues where the distance between the model and experimental structure (after superposition) is within various defined cutoffs[34]. For Fig. 2b, GDT-TS scores were calculated using the LGA program[34]. The score ranges from 0 to 1, with a higher score indicating a better model.

TM-score

The template-modeling (TM) score is a length-independent score that measures global structural similarity, originally proposed for proteins[52] and later adapted for RNA[53]. For Fig. 2b, the TM-score was calculated using the RNAalign program[53] with the C3′ as the representative atom for each nucleotide. The score ranges from 0 to 1, with a higher score being better.

LDDT

A superposition-independent score that evaluates the accuracy of the local atomic environment by comparing the network of inter-residue distances in the model to those in the experimental structure. For Fig. 2b, the LDDT scores were calculated using the following equation:

$$\mathrm{LDDT} = \frac{1}{4L}\sum_{\varepsilon\in\{0.5,1,2,4\}}\sum_{i=1}^{L}\frac{\sum_{j:D_{ij}<15} I\left(\left|d_{ij}-D_{ij}\right|<\varepsilon\right)}{\sum_{j:D_{ij}<15}1}$$

where L is the total number of nucleotides; d_{ij} and D_{ij} are the C3′ distance between the ith and jth nucleotides in the predicted and experimental structures, respectively; the calculation is restricted to residue pairs where D_{ij} <15 Å. $I()$ is the indicator function. The score ranges from 0 to 1, with a higher score indicating a more accurate local model.

Clashscore

A score measuring the severity of steric clashes in a structure, defined as the number of serious atomic overlaps (≥0.4 Å) per 1,000 atoms. For Fig. 2b, the clashscores were calculated using the Molprobity program[54]. Molprobity is a a widely recognized structure validation tool that evaluates the stereochemical quality of macromolecular models; although it is not bundled with trRosettaRNA, you can run Molprobity manually on trRosettaRNA outputs. This score ranges from 0 upwards, with a lower score indicating a more physically plausible model.

INF

The Matthews correlation coefficient of the base-pairing interaction network in the predicted structure[35]. The score ranges from 0 to 1, with a higher score indicating a more accurate base-pairing interaction network.

Procedure 1

The prediction results of the trRosettaRNA server are presented on a webpage (Figs. 5 and 6) organized into the following three sections:

1. Predicted structure model. This section visualizes the predicted 3D structure and its confidence score (pLDDT).
2. Predicted internucleotide distances. This section shows the predicted contact probability map and distance maps for specific atom pairs.
3. SS. This part displays visualizations of both the initial SS used as input (either predicted by trRNA-SS or provided by user) and the final SS extracted from the predicted 3D coordinates.

Procedure 2

The standalone trRosettaRNA package primarily generates the following key output files:

1. Predicted internucleotide distances (NPZ format).
2. The predicted 3D structure (PDB format), and the top five decoys if '`-pyrosetta`' flag was set.
3. Per-nucleotide pLDDT confidence scores (CSV file).

Reporting summary

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

Data availability

The example input and output files can be downloaded from https://yanglab.qd.sdu.edu.cn/trRosettaRNA/example/. Source data are provided with this paper.

Code availability

The trRosettaRNA server is available at https://yanglab.qd.sdu.edu.cn/trRosettaRNA/. The standalone package is available via GitHub at https://github.com/YangLab-SDU/trRosettaRNA2/ under the Apache-2.0 license.

Received: 7 May 2025; Accepted: 9 March 2026;
Published online: 08 April 2026

References

1. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
2. Wang, W., Su, B., Peng, Z. & Yang, J. Integrated experimental and AI innovations for RNA structure determination. *Nat. Biotechnol.* **44**, 205–214 (2026).
3. Sharma, S., Ding, F. & Dokholyan, N. V. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* **24**, 1951–1952 (2008).
4. Das, R. & Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl Acad. Sci. USA* **104**, 14664–14669 (2007).
5. Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* **7**, 291–294 (2010).
6. Boniecki, M. J. et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **44**, e63–e63 (2016).
7. Popenda, M. et al. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40**, e112–e112 (2012).
8. Zhao, Y. et al. Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* **2**, 734 (2012).
9. Zhang, Y., Wang, J. & Xiao, Y. 3dRNA: 3D structure prediction from linear to circular RNAs. *J. Mol. Biol.* **434**, 167452 (2022).
10. 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).
11. Watkins, A. M., Rangan, R. & Das, R. FARFAR2: improved de novo Rosetta prediction of complex global RNA folds. *Structure* **28**, 963–976 (2020).
12. Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).
13. Du, Z. et al. The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* **16**, 5634–5651 (2021).
14. He, K., Zhang, X., Ren, S. & Sun, J. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).
15. Gao, S. H. et al. Res2Net: a new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 652–662 (2021).
16. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
17. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
18. Sun, S., Wang, W., Peng, Z. & Yang, J. RNA internucleotide 3D closeness prediction by deep residual neural networks. *Bioinformatics* **37**, 1093–1098 (2021).
19. Wang, W. et al. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nat. Commun.* **14**, 7266 (2023).
20. 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://doi.org/10.1101/2025.04.09.647915 (2025).
21. Wang, W., Luo, Y., Peng, Z. & Yang, J. Accurate biomolecular structure prediction in CASP16 with optimized inputs to state-of-the-art predictors. *Proteins* **94**, 142–153 (2026).
22. Kretsch, R. C. et al. Assessment of nucleic acid structure prediction in CASP16. *Proteins* **94**, 192–217 (2026).
23. Zhang, C., Zhang, Y. & Pyle, A. M. rMSA: a sequence search and alignment algorithm to improve RNA structure modeling. *J. Mol. Biol.* **435**, 167904 (2023).
24. The Rnacentral Consortium RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.* **47**, D221–D229 (2019).
25. Sayers, EricW. et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* **53**, D20–D29 (2025).
26. Singh, J., Hanson, J., Paliwal, K. & Zhou, Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.* **10**, 5407 (2019).
27. Song, X.-Y. et al. Bacterial reverse transcriptase synthesizes long poly-A-rich cDNA for antiphage defense. *Science* **388**, eads4639 (2025).
28. Xue, C., Tian, J., Chen, Y. & Liu, Z. Structural insights into human ELAC2 as a tRNA 3′ processing enzyme. *Nucleic Acids Res.* **52**, 13434–13446 (2024).
29. Guo, R. et al. Engineered IscB-ωRNA system with improved base editing efficiency for disease correction via single AAV delivery in mice. *Cell Rep.* **43** (2024).
30. He, S. et al. Ribonanza: deep learning of RNA structure through dual crowdsourcing. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.24.581671 (2024).
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. Tarafder, S. & Bhattacharya, D. RNAbpFlow: base pair-augmented SE(3)-flow matching for conditional RNA 3D structure generation. Preprint at *bioRxiv* https://doi.org/10.1101/2025.01.24.634669 (2025).
33. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
34. Zemla, A. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31**, 3370–3374 (2003).
35. Parisien, M., Cruz, J. A., Westhof, É. & Major, F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* **15**, 1875–1885 (2009).
36. Zhang, T. et al. RNAcmap: a fully automatic pipeline for predicting contact maps of RNAs by evolutionary coupling analysis. *Bioinformatics* **37**, 3494–3500 (2021).
37. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
38. Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
39. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
40. Kalvari, I. et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).
41. Liu, X. et al. Quality assessment of RNA 3D structure models using deep learning and intermediate 2D maps. *Commun. Biol.* **9**, 293 (2026).
42. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).
43. Das, R. et al. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins* **91**, 1747–1770 (2023).
44. Li, T. et al. All-atom RNA structure determination from cryo-EM maps. *Nat. Biotechnol.* **43**, 97–105 (2025).
45. Degenhardt, M. F. S. et al. Determining structures of RNA conformers using AFM and deep neural networks. *Nature* **637**, 1234–1243 (2025).
46. Lee, Y. et al. Template-based RNA structure prediction advanced through a blind code competition. Preprint at *bioRxiv* https://doi.org/10.64898/2025.12.30.696949 (2025).
47. Rego, N. & Koes, D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324 (2015).
48. Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).
49. Lu, X.-J., Bussemaker, H. J. & Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **43**, e142–e142 (2015).

50. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystall. A* **32**, 922–923 (1976).
51. Bu, F. et al. RNA-Puzzles round V: blind predictions of 23 RNA structures. *Nat. Methods* **22**, 399–411 (2025).
52. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).
53. 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).
54. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
55. Baek, M. et al. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* **21**, 117–121 (2024).
56. 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://doi.org/10.1101/2022.05.15.491755 (2022).
57. Shen, T. et al. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat. Methods* **21**, 2287–2298 (2024).
58. Kagaya, Y. et al. NuFold: end-to-end approach for RNA tertiary structure prediction with flexible nucleobase center representation. *Nat. Commun.* **16**, 881 (2025).
59. Li, Y. et al. Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nat. Commun.* **14**, 5745 (2023).
60. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).

Acknowledgements

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

Author contributions

J.Y. conceived and supervised the project. W.W. designed and performed the experiments. X.L. developed the quality assessment algorithm RNArank. Z.P. analyzed the CASP data. 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/s41596-026-01356-8.

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

Peer review information *Nature Protocols* thanks Jamie Cate, who co-reviewed with Conner Langeberg, Ivo Hofacker and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Jianyi Yang

Last updated by author(s): Jan 30, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The benchmark RNAs are collected from CASP16 competition.
Data analysis	RNA_assessment tool (version 1.0), LGA program (version 1.0), Molprobity program (version 1.0), and RNA-align (version 20220412) were used to evaluate the predicted structures. PyMol (version 3.1.0) was used to display the 3D structures.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The example input and output files can be downloaded from: https://yanglab.qd.sdu.edu.cn/trRosettaRNA/example/.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. The CASP16 targets were collected from the CASP16 website, containing 36 "RNA monomer" targets.
Data exclusions	No data were excluded from the analyses
Replication	We have run the programs (108=(36*3) times with each target running three replications. ALL replications were successful. Please follow our instructions on the web server page, or instructions in the standalone package.
Randomization	The CASP16 targets were grouped according to the RNA length and template availability, resulting in 23 targets with < 400 nt or available templates, and 13 targets that met neither criterion. Within these two groups, 16 and 11 targets respectively have experimental structures available for validation.
Blinding	There was no blinding group or analysis in this manuscript as all experimental structures in our test have the deposited models in the public databases.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A