

## RESEARCH ARTICLE

## IsRNAcirc: 3D structure prediction of circular RNAs based on coarse-grained molecular dynamics simulation

Haolin Jiang<sup>1</sup>, Yulian Xu<sup>2,3</sup>, Yunguang Tong<sup>2,4\*</sup>, Dong Zhang<sup>1\*</sup>, Ruhong Zhou<sup>1,5</sup>

**1** College of Life Sciences and Institute of Quantitative Biology, Zhejiang University, Hangzhou, Zhejiang, China, **2** College of Life Sciences, China Jiliang University, Hangzhou, China, **3** China Jiliang University—Aoming (Hangzhou) Biomedical Co., Ltd. Joint Laboratory, Hangzhou, China, **4** Aoming (Hangzhou) Biomedical Co., Ltd., Hangzhou, China, **5** The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, China

☯ These authors contributed equally to this work.

\* [tony.tong@omigen.cn](mailto:tony.tong@omigen.cn) (Y.T); [zhangd\\_iqb@zju.edu.cn](mailto:zhangd_iqb@zju.edu.cn) (D.Z)



## OPEN ACCESS

**Citation:** Jiang H, Xu Y, Tong Y, Zhang D, Zhou R (2024) IsRNAcirc: 3D structure prediction of circular RNAs based on coarse-grained molecular dynamics simulation. PLoS Comput Biol 20(10): e1012293. <https://doi.org/10.1371/journal.pcbi.1012293>

**Editor:** Nir Ben-Tal, Tel Aviv University, ISRAEL

**Received:** July 1, 2024

**Accepted:** October 12, 2024

**Published:** October 28, 2024

**Copyright:** © 2024 Jiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The source code of IsRNAcirc and the data present in this study are available at <https://github.com/DongZhangRNA/IsRNAcirc>.

**Funding:** The National Natural Science Foundation of China (Nos. 12104396 to D.Z., U1967217 to R.Z.), the National Key R&D Program of China (2021YFF1200404 to R.Z., 2021YFA1201200 to R.Z.), National Independent Innovation Demonstration Zone Shanghai Zhangjiang Major Projects (ZJZX2020014 to R.Z.), the Starry Night Science Fund at Shanghai Institute for Advanced

## Abstract

As an emerging class of RNA molecules, circular RNAs play pivotal roles in various biological processes, thereby determining their three-dimensional (3D) structure is crucial for a deep understanding of their biological significances. Similar to linear RNAs, the development of computational methods for circular RNA 3D structure prediction is challenging, especially considering the inherent flexibility and potentially long length of circular RNAs. Here, we introduce an extension of our previous IsRNA2 model, named IsRNAcirc, to enable circular RNA 3D structure predictions through coarse-grained molecular dynamics simulations. The workflow of IsRNAcirc consists of four main steps, including input preparation, end closure, structure prediction, and model refinement. Our results demonstrate that IsRNAcirc can provide reasonable 3D structure predictions for circular RNAs, which significantly reduce the locally irrational elements contained in the initial input. Moreover, for a validation test set comprising 34 circular RNAs, our IsRNAcirc can generate 3D models with better scores than the template-based 3dRNA method. These findings demonstrate that our IsRNAcirc method is a promising tool to explore the structural details along with intricate interactions of circular RNAs.

## Author summary

Knowledge of the 3D structure of circular RNA molecules can help us better understand their roles in eukaryotic cells. However, experimental determination of circular RNA 3D structure remains challenging, with currently only one complete circular RNA 3D structure available in the Protein Data Bank. Thus, extending from their linear counterparts, computational methods that can predict circular RNA 3D structure are urgently needed. In particular, physics-based *de novo* 3D structure prediction methods are promising due to their powerful sampling capability in the conformational space, which avoid the problem of limited available templates in template-based approaches. We have developed

Study of Zhejiang University (SN-ZJU-SIAS-003 to R.Z., SN-ZJU-SIAS-009 to D. Z.), the National Center of Technology Innovation for Biopharmaceuticals (NCTIB2022HS02010 to R.Z.), Shanghai Artificial Intelligence Lab (P22KN00272 to R.Z.), Aoming Biomedical Research (AO-ZJU-SIAS-001 to R.Z.), the R&D Program of China Jiliang University - Aoming (Hangzhou) Biomedical Co., Ltd. Joint Laboratory (20211008 to Y.T.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

IsRNAcirc, a new physics-based method that employs an accurate coarse-grained force field and replica-exchange molecular dynamics simulations to predict the 3D structure of circular RNAs. IsRNAcirc is shown to significantly reduce the number of locally irrational structures in the input trials, which are caused by the lack of suitable templates and the improper assembly of fragments. Furthermore, in the absence of experimental reference structures, the predicted 3D models were validated using the popular scoring functions, such as DFIRE-RNA and rsRNASP, demonstrating that IsRNAcirc outperforms the template-based 3dRNA method in generating low-energy conformations.

## Introduction

Circular RNA molecules are a unique class of RNA that do not have a cap structure at the 5' end and a poly(A) tail at the 3' end, making them different from traditional linear RNAs. These circular RNAs are formed through a process called back splicing, in which segments of mRNA precursors are covalently joined together to form a circular structure. The initial and terminal junctures within a circular RNA sequence, known as the circularization sites, are specifically referred to as back-splicing junction (BSJ) regions. The discovery of circular RNAs dates back more than four decades, when the first circular RNA molecule was identified in plant viroids [1]. To date, thanks to advances in high-throughput RNA sequencing (RNA-seq) and specialized bioinformatics tools, thousands of circular RNAs have been discovered in various eukaryotic organisms, including fungi, protists, plants, worms, fish, insects, mammals, etc. [2–5]. Interestingly, many of these circular RNAs originate from protein-coding genes, suggesting that they participate in diverse biological processes and perform multiple functions within cells. For example, a prominent function of circular RNAs is their ability to act as microRNA sponges, sequestering miRNAs and preventing them from targeting and degrading specific mRNAs [6,7]. Other discovered functions of circular RNAs include interacting with ribosomes to increase protein synthesis, interacting with RNA-binding proteins (RBPs) to modulate their activity [8,9], directing proteins to specific cellular locations [2], and translating into unique peptides *via* internal ribosome entry sites (IRES) [10–13]. These rich functions of circular RNAs suggest that they should adopt specific three-dimensional (3D) structures.

However, experimental determination of circular RNA 3D structures remains very challenging, mainly due to the inherent flexibility and potentially considerable length of these RNA molecules. As a result, there is currently only one complete circular RNA 3D structure (PDB ID: 2OIU [14]) deposited in the Protein Data Bank (PDB). Therefore, to facilitate our understanding of the mechanisms of their biological functions, developing computational methods to predict the 3D structure of circular RNAs is a much-needed but lacking task. For their linear counterparts, various computational approaches have been developed over the past few decades to predict RNA 3D structure. In general, these computational approaches can be divided into two different categories: template-based methods and *ab initio* predictions. Template-based methods, such as ModeRNA [15], Vfold [16,17], RNAcomposer [18], and 3dRNA [19,20], rely on known RNA structures of similar sequences as templates to predict structures. The main bottleneck of template-based methods is that the available templates deposited in the PDB are limited, so they can easily fail to provide reasonable predictions for unknown RNA sequences. On the other hand, *ab initio* prediction methods, including NAST [21], iFoldRNA [22], SimRNA [23], and our IsRNA [24–26], utilize molecular dynamics (MD) or Monte Carlo (MC) simulations to search the conformational space and a particular scoring scheme (usually the energy functions) to select the lowest energy or most probable structure

from the conformation ensemble as the prediction(s). Due to their template-free nature, *ab initio* methods can in principle provide predictions for any RNA sequences, but at a relatively high computational cost. Additionally, it is worth noting that most of the current template-based methods are combined with MD or MC methods to overcome the bottleneck of limited templates. This combined approach enables 3D structure prediction for any RNA molecule when one or more templates are not available. Recently, artificial intelligence (AI)-based approaches have also exhibited great promise in predicting linear RNA 3D structures [27–29]. Nevertheless, coming back to circular RNAs, there is currently only one computational method that can predict their 3D structures. That is, Xiao and coworkers [19] extended their 3dRNA model to enable the prediction of the 3D structures of circular RNAs. Considering the template-based nature of the 3dRNA model and the limited number of available templates, it is necessary to develop other computational methods to predict circular RNA structures, such as *ab initio* prediction methods.

Here, we introduced an extension of our previous IsRNA2 model, termed IsRNAcirc, to enable circular RNA 3D structure predictions *via* coarse-grained (CG) MD simulations. To evaluate the performance of IsRNAcirc, we collected a test set consisting of 34 circular RNAs of different lengths and topologies from the previous study [30]. The results demonstrate that IsRNAcirc can provide reasonable 3D structure predictions for circular RNAs and significantly reduce the locally irrational elements contained in the initial input. Furthermore, since the true structures are unavailable, a comparison of IsRNAcirc and 3dRNA was conducted by model assessments with two knowledge-based scoring functions. Notably, IsRNAcirc outperforms 3dRNA because the former generates conformations with lower energy in different scoring schemes. The presented IsRNAcirc model provides a feasible approach to capture the intricate features of circular RNA structures and facilitates our understanding of their various functions.

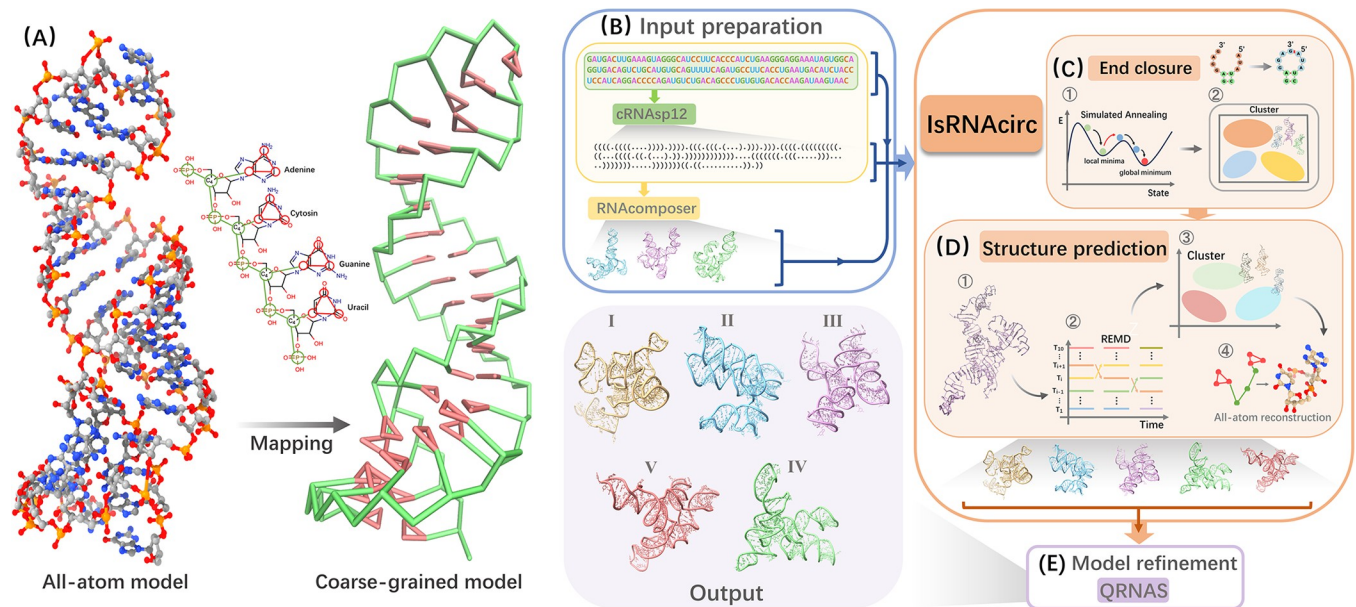
## Results and discussions

### Overview of IsRNA2 coarse-grained model

To predict the 3D structures of linear RNAs, IsRNA2 [26] incorporates three essential features (S1 Fig): a five-bead per nucleotide CG representation to preserve the three interacting edges of nucleobases, an accurate CG force field derived from the iterative simulated reference state approach, and the utilization of replica-exchange molecular dynamics (REMD) simulations to effectively sample the conformational space. Using RNA sequence, secondary (2D) structure, and initial 3D structures (optionally) as input, the workflow of IsRNA2 contains sampling the conformational space through REMD simulations with 10 replicas possessing different temperatures, constructing a conformational ensemble (containing 5,000 structure snapshots) from simulation trajectories, clustering the top 10% structures with the lowest potential energies to generate the most probable predictions, and finally recovering all-atom structures and energy minimization (S1B Fig). Our benchmark test showed that IsRNA2 achieves comparable performance to the atomic model in *de novo* modeling of noncanonical RNA motifs and can refine large RNA 3D models predicted by other programs. More details on IsRNA2 can be found in our previous work [26].

### The framework of IsRNAcirc for circular RNA 3D structure prediction

IsRNAcirc is an extension of our previous CG IsRNA2 [26] model to enable prediction of circular RNA 3D structure. The main difference in circular RNAs compared to linear RNAs is the presence of covalently bound ends, such that the connection between the two "end nucleotides" (including bonds, bond angles, and torsion angles) is the same as the connection



**Fig 1. The framework for predicting circular RNA 3D structures using IsRNAcirc.** (A) Converting an all-atom model (PDB ID:1Z5C, left) into the coarse-grained representation (right) through predefined mapping relationships in IsRNAcirc. IsRNAcirc predicts circular RNA 3D structure using four main steps: (B) Input preparation. From sequence information, cRNAasp12 [31] was used to predict the 2D structure of circular RNAs. Based on the sequence and predicted 2D structure, RNAComposer [18] was used to generate initial 3D structures of the corresponding linear counterpart. Sequence information, predicted 2D structure, and those obtained initial 3D structures are used as inputs for IsRNAcirc for circular RNA 3D structure predictions; (C) End closure. The 5'- and 3'-end of circular RNA are circularized using a harmonic restraint between the two ends of linear structure and the simulated annealing technique. A clustering procedure was performed based on the last 10% MD trajectories to obtain starting structures for the subsequent step; (D) Structure prediction. Possible 3D structures of circular RNA were predicted following a similar process to IsRNA2, including conformational sampling through REMD simulations, clustering the lowest energy conformations to generate the most probable structures, and all-atom reconstruction. By default, five predicted 3D models are provided; (E) Model refinement. QRNAS [32] was employed to refine the predicted 3D models by eliminating problems such as spatial steric resistance and bond breakage.

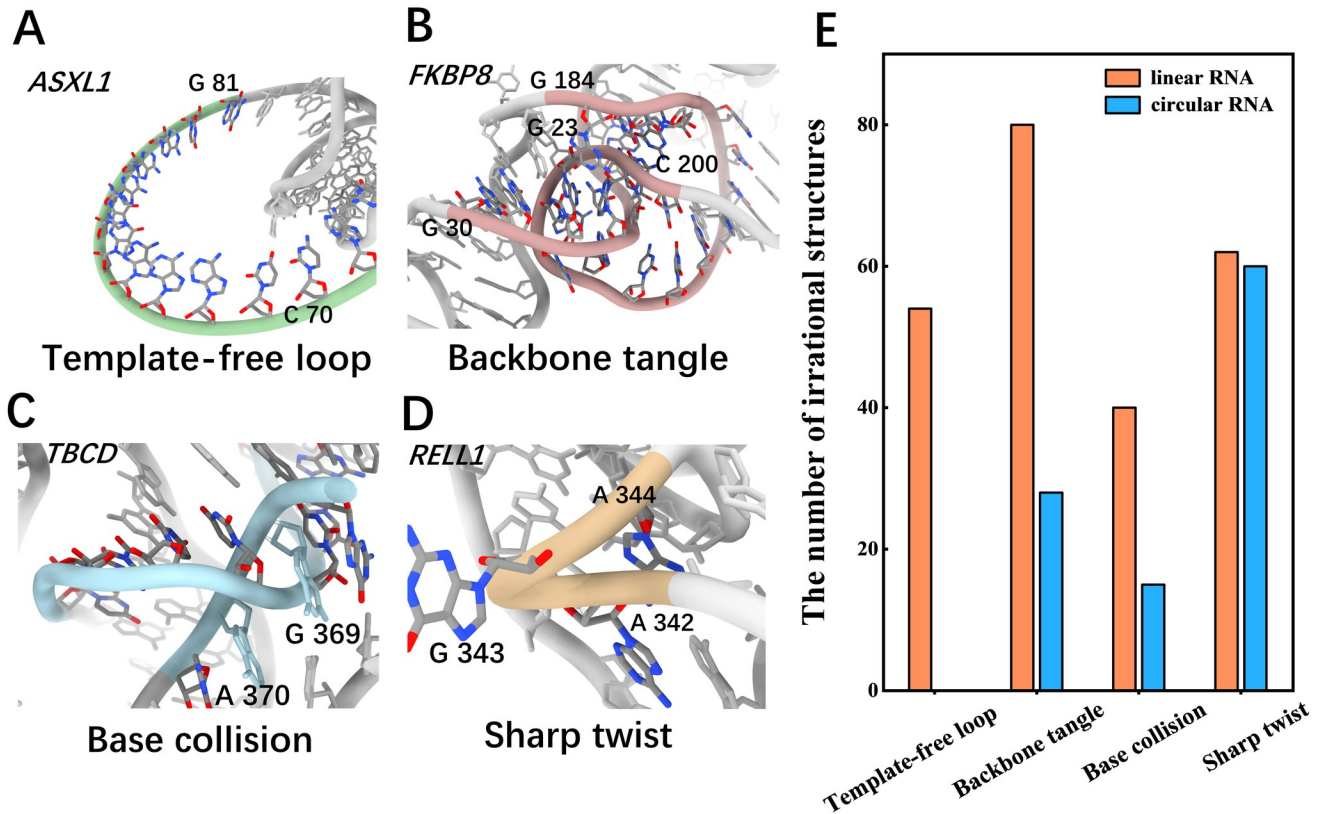
<https://doi.org/10.1371/journal.pcbi.1012293.g001>

between any two adjacent nucleotides in the sequence. Thus, circular RNAs may adopt different 2D and 3D structures compared to their linear counterparts. There are four main steps to predict circular RNA 3D structure using IsRNAcirc model, including input preparation, end closure, structure prediction, and model refinement (Fig 1). Firstly, from the sequence information, cRNAasp12 tool [31] was used to predict the 2D structure of circular RNA molecules. And based on the sequence information and predicted 2D structure, RNAComposer [18] was utilized to generate three (if available) initial 3D structures for the corresponding linear counterpart. Collectively, the sequence information, the predicted 2D structure, and those obtained initial 3D structures serve as input to IsRNAcirc (Fig 1B). Secondly, in the end closure step, we combined a harmonic potential constraint with a gradually increased force constant and the simulated annealing simulation technique to close the 5'- and 3'-end of the initial 3D structures (Fig 1C). Meanwhile, in order to reduce unreasonable topologies caused by improper fragment assembly in the initial 3D structures, the base-pairing interactions were also weakened in the early stages of simulated annealing simulations. That is, the end closure step went through a two-step process to try to repair the unreasonable elements contained in the initial structures: first unfolding and then refolding. Starting from the initial 3D structure, the unfolding process (high temperature and weak base-pairing interactions) first opened the base pairs to allow the RNA chain to reach a (partially) extend conformation; thereafter, the refolding process (low temperature and strong base-pairing interactions) guided the RNA molecule to form a more reasonable folded conformation based on physical interactions. See [S1 Movie](#)

for an illustrative example of the unfolding-refolding process. The most probable conformation (obtained by clustering analysis) in the later stage of the refolding simulation was selected as the starting structure for subsequent simulations. This end closure step was specifically designed for circular RNA 3D structure prediction, which is the primary difference between IsRNAcirc and the original IsRNA2 model. Thirdly, once the two “terminal nucleotides” were circularized, standard bonded energy parameters (including bond, bond angles, and torsion angles) were introduced to describe their connection. And a similar process to IsRNA2 was performed to predict circular RNA 3D structures. That is, sampling the conformational space via REMD simulation, constructing the conformation ensemble and clustering the top 10% lowest energy structures, and all-atom reconstruction (Figs 1D and S1B). Considering the longer size of circular RNAs, the simulation time used in IsRNAcirc (100 ns) is longer than the original IsRNA2 model (50 ns). We chose the top five predicted models (centroid structures of the top five largest clusters) after all-atom recovery as candidate 3D structures of certain circular RNA. Finally, since the CG nature of IsRNAcirc may lead to the loss of fine structural details, a model refinement step was performed based on QRNAS [32] (Fig 1E), which can effectively eliminate the imperfect stereochemical features, unnatural bond lengths, and spatial steric resistance present in the predicted 3D models. Overall, through the above four steps, users can use IsRNAcirc to generate five predicted 3D structure models of circular RNAs in PDB format.

### Benchmarking of IsRNAcirc on 34 circular RNAs

The predictive ability of IsRNAcirc was tested on a dataset from Chen’s group [30]. This dataset comprises 34 circular RNAs with lengths ranging from 161 to 435 nucleotides. Similar to the previous study [19], those 34 circular RNAs were classified into four types according to the location of BSJ in the 2D structure: helix-circular, hairpin-circular, internal-circular, and junction-circular RNAs. For each type of circular RNA, the two “terminal nucleotides” are well circularized in the 3D models predicted by IsRNAcirc (S2 Fig). Moreover, we aligned the predicted circular RNA 3D models with the initial 3D structures of the corresponding linear RNAs (obtained by RNAComposer) and calculated their root mean square deviations (RMSDs). For all four types of circular RNAs, the predicted circular 3D models by IsRNAcirc differ significantly from the initial linear 3D structures generated by RNAComposer, with average RMSD values larger than 30 Å (S3A Fig). This result is different from observations in 3dRNA [19], which showed that when 3dRNA were used to predict both the linear and circular RNAs, helix-circular and hairpin-circular RNAs adopt almost the same folded structures as their linear counterparts, whereas the assembled structures of internal-circular and junction-circular RNAs are very different from those of their linear counterparts. To validate our results, we first compared the circular 3D models predicted by 3dRNA with the initial linear 3D models generated by RNAComposer. As shown in S4A Fig, large RMSD values were also observed between the linear and circular RNA structures, indicating that the circular 3D models predicted by 3dRNA were also significantly different from the linear 3D structures generated by RNAComposer. Then, we noticed that it is still challenging to accurately predict the 3D structure of large-sized RNA ( $\geq 100$  nucleotides). Even given the same sequence and 2D structure as input, different methods (such as RNAComposer, 3dRNA, and IsRNA2) may provide different 3D structure predictions, and it is difficult to determine which approach is better. In this situation, the approach with stronger sampling capacity (e.g., *ab initio* methods) seems to be more promising, as they can provide many reasonable 3D candidate structures for subsequent studies of interest. To this end, considering that IsRNAcirc has a much stronger sampling ability in conformational space than templated-based approaches such as 3dRNA and



**Fig 2. IsRNAcirc can significantly reduce the locally irrational topologies contained in initial 3D structures (generated by RNAComposer).** (A–D) Four types of locally irrational topology were identified in initial 3D structures through view inspection: (A) template-free loop due to a lack of related templates in the library, (B) backbone tangle caused by improper template assembly, (C) base collision, in which a backbone segment passes through the middle of two nucleobases, and (D) sharp twist, in which an apparently unreasonable sharp turn is observed in the backbone. The names of circular RNAs for which the relevant irrational topology was observed in the initial 3D structures were also given in top left. (E) The total numbers of locally irrational topologies observed in all initial 3D structures (three structures for each RNA) and the predicted 3D models by IsRNAcirc (five models for each RNA, if available) were displayed in a type-dependent manner. All 34 circular RNAs were considered.

<https://doi.org/10.1371/journal.pcbi.1012293.g002>

RNAComposer, we believe that our results are acceptable, that is, there are large differences between the 3D structures of all circular RNAs predicted by IsRNAcirc and their linear counterparts generated by RNAComposer, regardless of their circularized types. Moreover, we also constructed two illustrative RNA molecules (a helix-circular RNA and a hairpin-circular RNA) to further compare the differences between the IsRNA2 generated linear 3D structures and the IsRNAcirc predicted circular 3D models. As expected, as shown in S4B and S4C Fig, there are only slight differences between the linear and circular structures. Overall, these results suggested that our IsRNAcirc predictions are reasonable. Furthermore, a positive correlation was observed between the lengths of circular RNAs and the calculated RMSD values between linear and circular structures (see S3B Fig), which implies that longer circular RNAs tend to adopt more different folded conformations relative to the initial linear 3D structures generated by RNAComposer. This point further emphasized the stronger conformational sampling capability of our IsRNAcirc model.

In particular, through view inspection, we observed a lot of locally irrational topologies in the initial 3D structures generated by RNAComposer. Specifically, four main types of locally irrational topologies were identified, including template-free loop, backbone tangle, base collision, and sharp twist; see Methods Section for more details of their definitions. Fig 2A–2D display illustrated examples for each of these four types of locally irrational topologies. Due to the

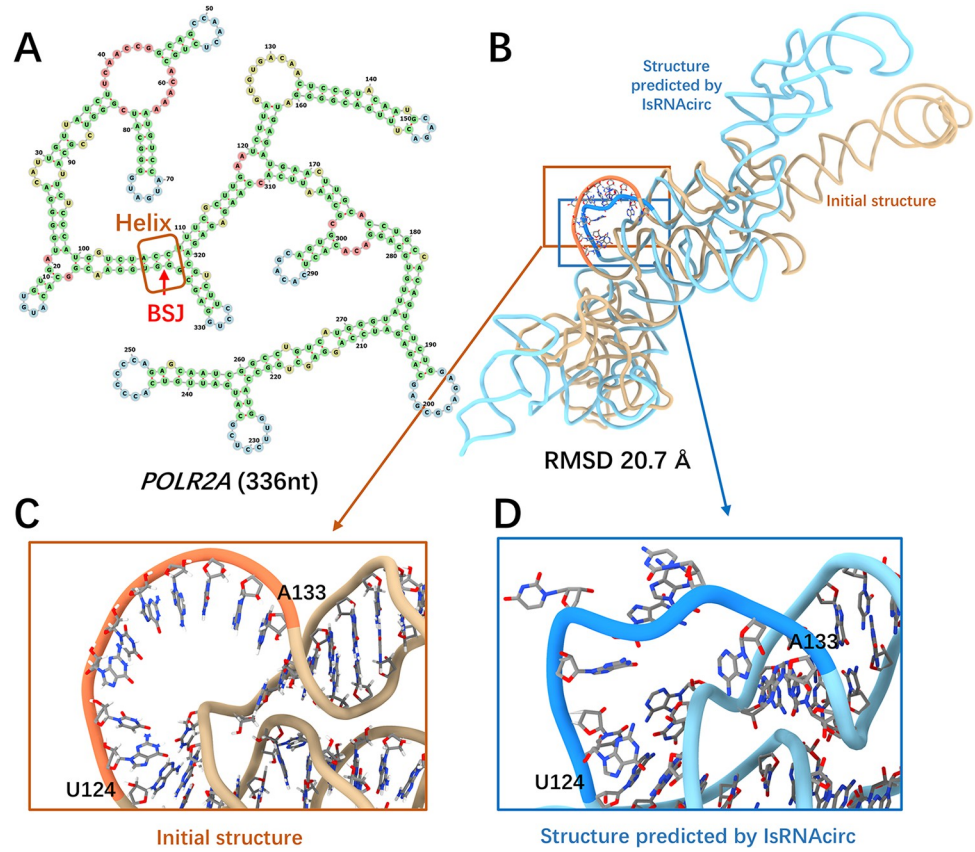
template-based nature of RNAComposer approach and the long length ( $> 160$  nucleotides) of RNA molecules, these locally irrational topologies are thought to be mainly caused by the lack of appropriate templates in the library and improper assembly of the template fragments. For instance, because of the lack of template for long loops ( $\geq 15$  nucleotides in length) in the library, an alternative single strand with predefined geometries was introduced in the initial 3D structures, which results in the second most common type of locally irrational topology (termed template-free loop, Fig 2A). Besides, the most prevalent type of locally irrational topology is base collision, in which a backbone segment traverses through the space between two adjacent nucleobases and causes severe atom clashes (Fig 2C). Failure to consider possible overlaps between different template fragments during assembly process is thought to be responsible for the presence of bases collision.

Notably, our IsRNAcirc can fix most of these aforementioned locally irrational topologies observed in the initial 3D structures through force field-guided simulations. As shown in Figs 2E and S5, the numbers of observed template-free loops (from 54 to 0), backbone tangles (from 80 to 28), base collisions (from 40 to 15), and sharp twists (from 62 to 60) are all largely decreased in the 3D models predicted by IsRNAcirc. It is worth noting that the number of initial 3D structures per RNA molecule is three, but the number of predicted 3D models by IsRNAcirc for each circular RNA is five. Therefore, the occurrence of these locally irrational topologies is significantly decreased in predicted 3D models (S5 Fig). These results demonstrated that our IsRNAcirc is a very promising model for predicting circular RNA 3D structures, especially for those molecules with very long size.

### Representative predictions for each type of circular RNAs

To further illustrate the ability of IsRNAcirc in predicting circular RNA 3D structures, four representative examples were selected for each circular RNA type, and their prediction results were analyzed separately below.

For helix-circular RNAs, the POLR2A molecule consisting of 336 nucleotides was selected, and the prediction results are shown in Fig 3. According to the predicted 2D structure by cRNAsp12, the BSJ of POLR2A circular RNA is located in a 6-bp stem (Fig 3A). The initial linear 3D structure generated by RNAComposer and the predicted 3D model by IsRNAcirc adopt different global folds, and the RMSD between these two structures is 20.7 Å (Fig 3B). Most importantly, the total number of locally irrational topologies of this molecule has decreased from 3 template-free loops (over all three initial linear structures) to 0 (over all five models predicted by IsRNAcirc). For example, a template-free loop of 10-nucleotide length (U124-A133) in the initial structure (caused by the lack of an appropriate template in the library, Fig 3C) was changed to a segment with a more reasonable geometry in the predicted model (Fig 3D). Furthermore, similar improvements in local topology were also observed after IsRNAcirc optimization for other helix-circular RNAs, such as ASXL1 (195 nucleotides) and KIAA0368 (435 nucleotides) molecules in S5A Fig. In addition, we also used IsRNAcirc to predict the 3D structure of the L1 ribozyme ligase circular addcut[14] (PDB ID: 2OIU), which is the only experimental circular RNA structure deposited in the PDB and has the BSJ located in the helical region. The predicted 3D structure was displayed in S6 Fig, using the sequence information and native 2D structure extracted from the experimental structure as input. Overall, the predicted 3D structure adopted a global folding similar to the experimental structure, and the heavy-atom RMSD between the predicted and experimental structures is 11.7 Å. We noted that this relatively large RMSD value is mainly caused by the poor prediction of the core three-way junction, while the local RMSDs of the three arms are 1.79 Å, 1.06 Å, and 2.46 Å, respectively (see S6 Fig). These results indicated that it is still challenging to accurately predict



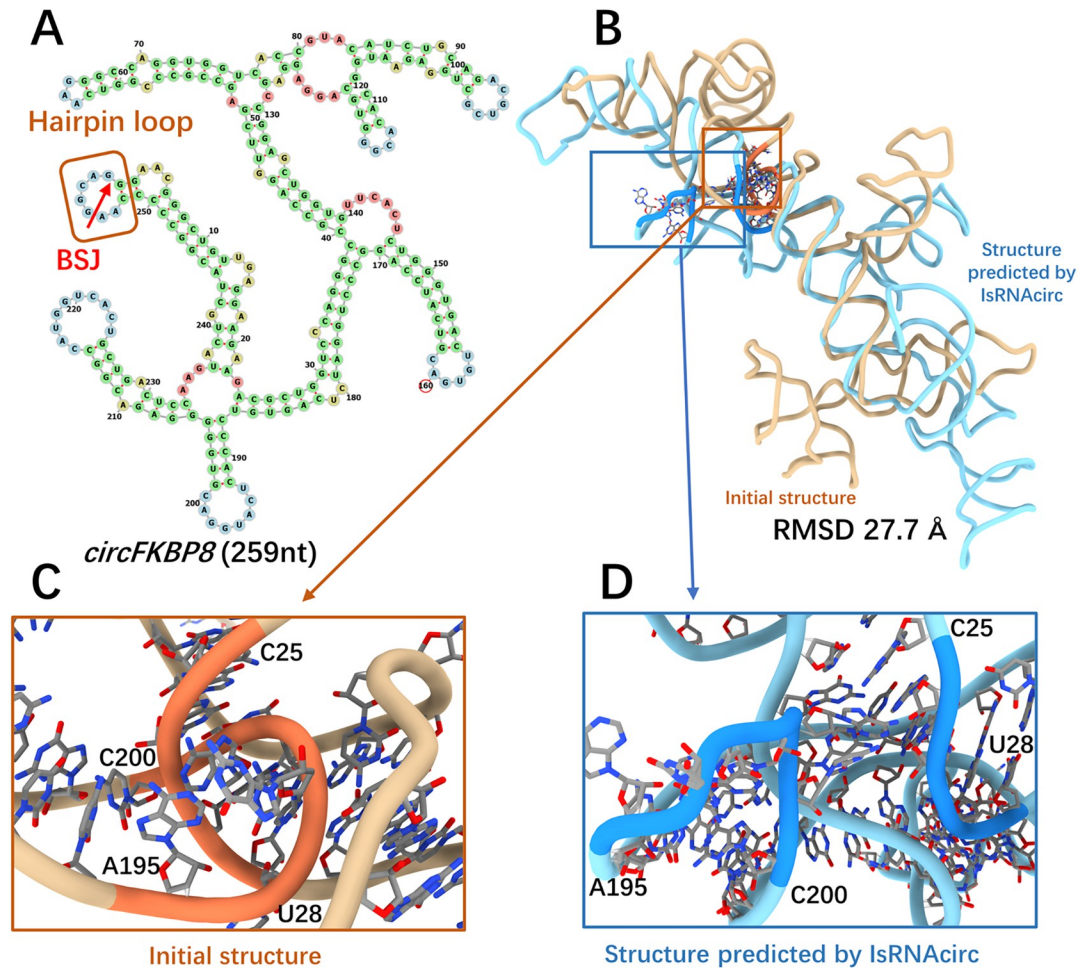
**Fig 3. A representative prediction for helix-circular RNAs.** (A) Predicted 2D structure of POLR2A consisting of 336 nucleotides (circRNA ID: hsa\_circ\_0000741), in which the BSJ is located in the helical region (marked by red arrow). (B) Superposition of the initial linear 3D structure (colored by light brown) and the predicted 3D structure of POLR2A circular RNA by IsRNAcirc (colored by light blue). (C-D) Zoom-in show the local details of 3D models: (C) the presence of template-free loop in the initial 3D structure and (D) the corresponding segment in the predicted 3D model by IsRNAcirc. The atoms carbon, oxygen, and nitrogen are colored by silver, red, and blue, respectively.

<https://doi.org/10.1371/journal.pcbi.1012293.g003>

the 3D structure of RNA containing multi-way junction, which is consistent with our previous observations [25].

The FKBP8 molecule, composed of 259 nucleotides, was chosen as a representative case of hairpin-circular RNAs. Based on the predicted 2D structure shown in Fig 4A, the BSJ of this circular RNA is located in a 7-nucleotide hairpin loop. The RMSD between the initial linear structure and the predicted 3D model by IsRNAcirc is 27.7 Å (Fig 4B). As shown in Fig 4C, a backbone tangle between two segments (C25-U28 and A195-C200) was observed in the initial linear structure. After IsRNAcirc optimization, this backbone tangle was repaired in the predicted 3D model (Fig 4D). Overall, the total number of observed locally irrational topologies for the FKBP8 circular RNA decreases from 7 (over all three initial linear structures) to 5 (over all five models predicted by IsRNAcirc). For other hairpin-circular RNAs, such as ASAP1 (229 nucleotides) and RELL1 (434 nucleotides), a notable reduction in the number of locally irrational topologies was also found in the 3D models that predicted by IsRNAcirc (S5B Fig).

For internal-circular RNAs, the MBOTA2 circular RNA composed of 224 nucleotides was selected as a representative example, and the prediction results are displayed in Fig 5. Based on the 2D structure predicted by cRNAsp12, the BSJ of this molecule is located in a 1x1 internal loop (Fig 5A). The initial linear 3D structure generated by RNAComposer and the predicted

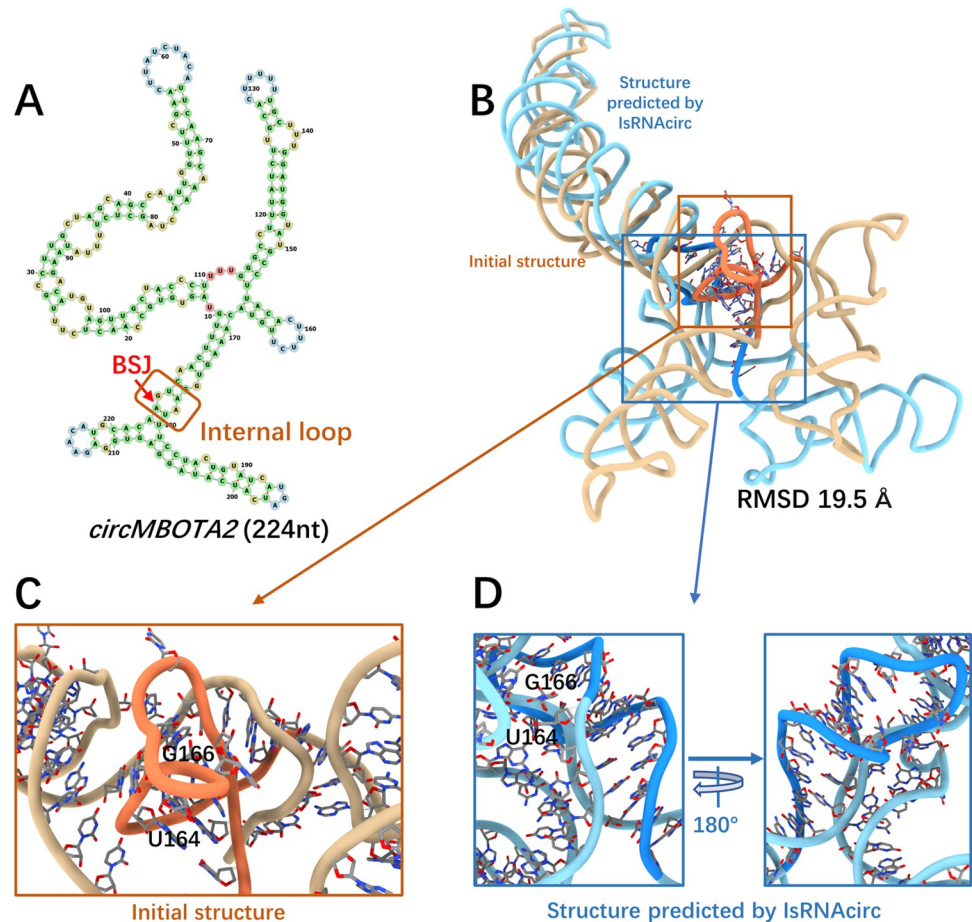


**Fig 4. A representative prediction for hairpin-circular RNAs.** (A) Predicted 2D structure of FKBP8 consisting of 259 nucleotides (circRNA ID: hsa\_circ\_0000915), in which the BSJ is located in the hairpin loop (marked by red arrow). (B) Superposition of the initial linear 3D structure (colored by light brown) and the predicted 3D structure of FKBP8 by IsRNAcirc (colored by light blue). (C-D) Zoom-in show the local details of 3D models: (C) the presence of a backbone tangle in the initial 3D structure and (D) the optimized segments in the 3D model predicted by IsRNAcirc. The atoms carbon, oxygen, and nitrogen are colored by silver, red, and blue, respectively.

<https://doi.org/10.1371/journal.pcbi.1012293.g004>

3D model of the MBOTA2 molecule through IsRNAcirc adopt similar global folds in some regions, and the RMSD between these two structures is 19.5 Å (Fig 5B). Notably, no locally irrational topology was found in the 3D models predicted by IsRNAcirc, relative to 3 backbone tangles and 6 sharp twists observed in the initial linear structures. For instance, a sharp twist of the backbone segment (U164-G166) was found in the initial structure (Fig 5C), and this unreasonable topology was fixed by force field-guided MD simulations in IsRNAcirc (Fig 5D). In addition, similar repairs of locally irrational topologies by simulations in IsRNAcirc were also observed in other internal-circular RNAs, such as CNNB1 (378 nucleotides) and PVT1 (410 nucleotides) circular RNAs (S5C Fig).

The SLC22A23 molecule, consisting of 259 nucleotides, was chosen as an illustrated example of junction-circular RNAs. According to the predicted 2D structure shown in Fig 6A, the BSJ of SLC22A23 circular RNA is located in a 3-way junction loop. Due to the great difference in the topological structure of multi-way junctions (circular RNA) and open-loops (linear counterparts), the initial linear structure and the predicted 3D model by IsRNAcirc adopt very



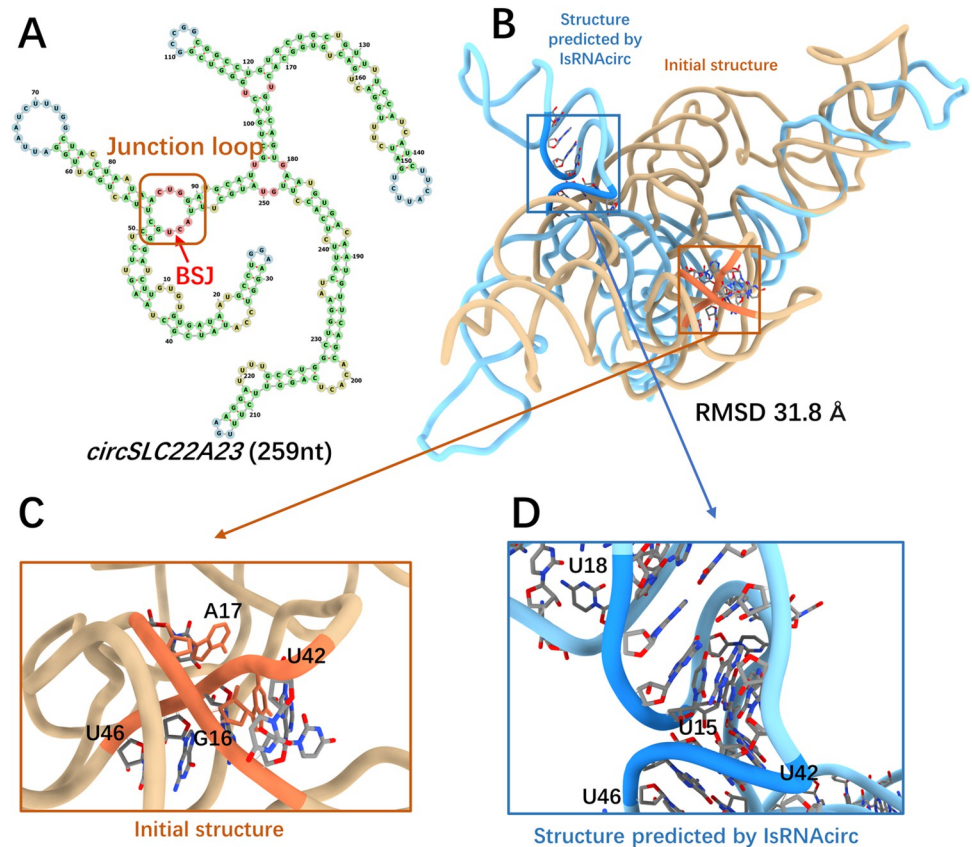
**Fig 5. A representative prediction for internal-circular RNAs.** (A) Predicted 2D structure of MBOTA2 consisting of 224 nucleotides (circRNA ID: hsa\_circ\_0007334), in which the BSJ is located in the 1x1 internal loop (marked by red arrow). (B) Superposition of the initial linear 3D structure (colored by light brown) and the predicted 3D structure of MBOTA2 by IsRNAcirc (colored by light blue). (C-D) Zoom-in show the local details of 3D models: (C) the presence of sharp twist of backbone in the initial 3D structure and (D) the optimized topology in the 3D model predicted by IsRNAcirc. The atoms carbon, oxygen, and nitrogen are colored by silver, red, and blue, respectively.

<https://doi.org/10.1371/journal.pcbi.1012293.g005>

different global folds, which results in a large RMSD value of 31.8 Å (Fig 6B). As shown in Fig 6C, there is a base collision in the local region of the initial 3D structure, where a backbone segment (U42-U46) passes through the middle space of two nucleobases (G16 and A17). After IsRNAcirc optimization, this irrational topology was fixed in the predicted 3D model, and a more reasonable conformation was presented. In total, the number of observed locally irrational topologies for the SLC22A23 molecule decreases from 3 (over all three initial linear structures) to 1 (over all five models predicted by IsRNAcirc). Moreover, for other junction-circular RNAs, such as EPHB4 (362 nucleotides) and PTK2 (394 nucleotides), the number of observed locally irrational topologies is also significantly decreased (S5D Fig).

### Comparison with 3dRNA in circular RNA 3D structure prediction

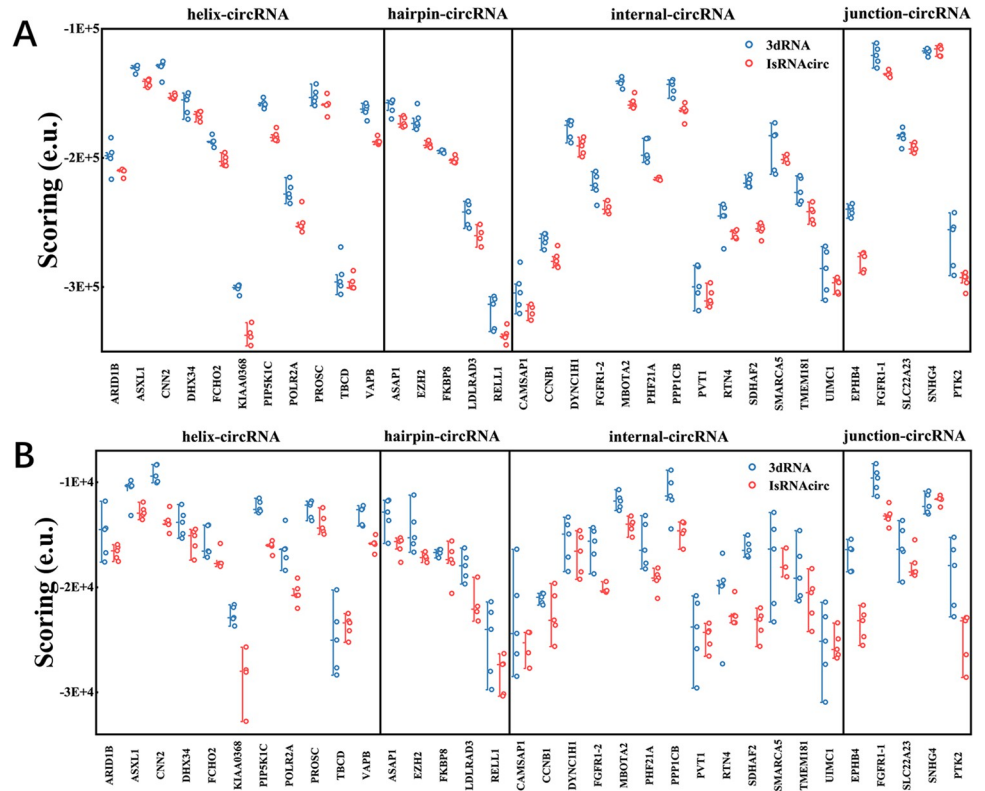
Due to the current absence of experimental 3D structures of circular RNAs, the performance of IsRNAcirc in circular RNA 3D structure prediction cannot be directly validated or compared with the template-based 3dRNA approach. Therefore, apart from the aforementioned plausibility examination of the local topologies, an indirect comparison was also performed



**Fig 6. A representative prediction for junction-circular RNAs.** (A) Predicted 2D structure of SLC22A23 consisting of 259 nucleotides (circRNA ID: hsa\_circ\_0075504), in which the BSJ is located in the 3-way junction loop (marked by red arrow). (B) Superposition of the initial linear 3D structure (colored by light brown) and the predicted 3D structure of SLC22A23 by IsRNAcirc (colored by light blue). (C-D) Zoom-in show the local details of 3D models: (C) the presence of base collision in the initial 3D structure and (D) the corresponding segments in the 3D model predicted by IsRNAcirc. The atoms carbon, oxygen, and nitrogen are colored by silver, red, and blue, respectively.

<https://doi.org/10.1371/journal.pcbi.1012293.g006>

based on scoring functions. Previously, various scoring functions (energy functions) have been developed to select the near-native conformations from decoy structures, including 3dRNA-score [33], RASP [34], rsRNASP [35], DFIRE-RNA [36], and so on. In general, due to the energy function nature of most scoring functions, conformations with lower scores are considered closer to the native state. Consequently, an approach that can generate low-scoring conformations is better than one that can only obtain high-scoring structures. Considering the availability, user experience, and release time of the scoring functions, here we employed two knowledge-based scoring functions, DFIRE-RNA [36] and rsRNASP [35], to assess the predicted 3D structure models generated by IsRNAcirc and 3dRNA. To be fair, the same sequence and 2D structure inputs were used for both IsRNAcirc and 3dRNA, and the predicted 3D models were scored individually by those two scoring functions. Here the default procedure of 3dRNA was used to predict circular 3D structures. As a preliminary insight, we calculated the pairwise RMSD between the IsRNAcirc and 3dRNA predicted 3D structures for all 34 tested circular RNAs, and the results were summarized in S7A Fig. Overall, the 3D structures predicted by IsRNAcirc are significantly different from those predicted by 3dRNA; see S7B and S7C Fig for two representative cases. Then, as shown in Fig 7A, based on the DFIRE-RNA scoring scheme, our IsRNAcirc can always generate lower energy conformations than the



**Fig 7.** Comparison of the performance of circular RNA 3D structure predictions between IsRNAcirc and 3dRNA based on the (A) DFIRE-RNA and (B) rsRNASP scoring function. Each 3D model predicted by IsRNAcirc or 3dRNA method was scored individually and displayed according to the circularized types. e.u. (energy unit).

<https://doi.org/10.1371/journal.pcbi.1012293.g007>

3dRNA method for majority of tested circular RNAs with different circularized types. For example, for KIAA0368 (helix-circular RNA, 435 nucleotides) and EPHB4 (junction-circular RNA, 362 nucleotides) molecules, the average scores ( $-3.37 \times 10^5$  and  $-2.80 \times 10^5$  e.u.) of 3D models predicted by IsRNAcirc are obviously lower than the average scores ( $-3.01 \times 10^5$  and  $-2.41 \times 10^5$  e.u.) of models generated by 3dRNA. Furthermore, consistent results were also observed based on the alternative rsRNASP scoring function (Fig 7B). In order to have a more direct overview of the performance of IsRNAcirc relative to 3dRNA, we calculated the energy score reduction ratio between IsRNAcirc and 3dRNA predictions, namely  $(Scoring_{IsRNAcirc} - Scoring_{3dRNA}) / Scoring_{3dRNA}$ , based on the two selected scoring functions. As shown in Table 1, based on the DFIRE-RNA (rsRNASP) scoring scheme, among all 34 tested circular RNAs, there is 73.5% (85.2%) cases where the average energy score of the IsRNAcirc predictions was lower than that of the corresponding 3dRNA predictions by more than 5%. These results indirectly demonstrate that IsRNAcirc has superior performance in circular RNA 3D structure prediction compared to the templated-based 3dRNA method, which mainly benefits from the more powerful sampling capability throughout the conformational space and accurate energy functions involved in the *ab initio* prediction method (e.g., IsRNAcirc here).

Furthermore, an indirect validation was performed in the previous study [19] to examine the performance of 3dRNA in predicting circular RNA structures. They compared the binding ability of circular RNA and its linear cognate RNA to double-stranded RNA (dsRNA)-binding proteins, such as the innate immune dsRNA receptor PKR. PKR is a serine/threonine kinase

**Table 1. Summary of the performance of IsRNAcirc relative to 3dRNA in circular RNA 3D structure prediction.** Lists are the number of cases where IsRNAcirc prediction has a lower energy score in different range based on the DFIRE-RNA and rsRNASP scoring schemes.

Energy Score Ratio <sup>a</sup>	DFIRE-RNA					rsRNASP				
	Circular RNA types:					Circular RNA types:				
	helix	hairpin	internal	junction	total	helix	hairpin	Internal	junction	total
< 5%	2	1	4	2	9	1		3	1	5
5%—10%	4	3	5		12		1	3		4
10%—15%	3	1	3	2	9	4	1	2	1	8
15%—20%	2		1	1	4	1	2	1		4
20%—25%						1	1	1		3
25%—30%						2		1		3
30%—35%						1		1	1	3
35%—50%						1		1	2	4
percentage with $\geq 5\%$ <sup>b</sup>	81.8%	80.0%	69.2%	60.0%	73.5%	90.9%	100.0%	76.9%	80.0%	85.2%

<sup>a</sup>Energy Score Ratio =  $(\overline{\text{Scoring}}_{\text{IsRNAcirc}} - \overline{\text{Scoring}}_{\text{3dRNA}}) / \overline{\text{Scoring}}_{\text{3dRNA}}$ , where  $\overline{\text{Scoring}}_{\text{IsRNAcirc}}$  and  $\overline{\text{Scoring}}_{\text{3dRNA}}$  are the average energy scores over five 3D structures predicted by IsRNAcirc and 3dRNA, respectively.

<sup>b</sup>Percentage of cases where the average energy score of IsRNAcirc predictions falls below more than 5% relative to the corresponding 3dRNA predictions.

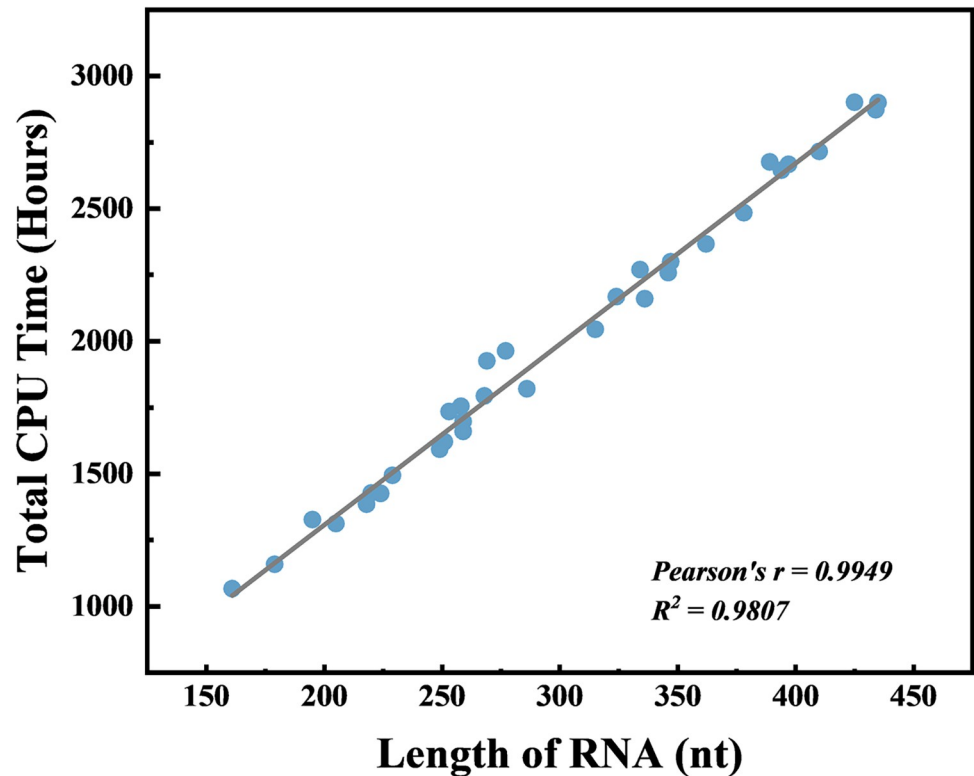
<https://doi.org/10.1371/journal.pcbi.1012293.t001>

that plays a crucial role in the cellular response to viral infections, particularly those involving dsRNA [19,20,30]. Compared with the linear cognate RNA, PKR prefers to bind to the intramolecular imperfect RNA double helix formed in circular RNA. Therefore, we also performed a similar binding ability comparison to further evaluate the performance of IsRNAcirc. For the circPOLR2A molecule, three 3D structures were predicted by IsRNAcirc, and three 3D models for its linear counterpart were also prepared by IsRNA2. We then used HDock [37] for molecular docking to determine the possible binding poses of circular/linear RNA with PKR and the top ten candidates were recorded. As shown in S8 Fig, the mean docking score (binding energy) of the circular RNA-PKR complex is always lower than that of the linear cognate RNA-PKR complex in all three docking experiments, suggesting that the binding between circular RNA and PKR is indeed stronger than that between its linear counterpart and PKR. Overall, these findings illustrate that our IsRNAcirc method is a promising tool to explore the structural details along with intricate interactions of circular RNAs.

## Runtime

Runtime is one of the pivotal criteria for assessing the applicability and efficiency of any computational tool. Because IsRNAcirc relies on MD simulations to sample the huge conformational space, the running time consumed by IsRNAcirc to predict the 3D structures of circular RNAs is relatively long. For instance, IsRNAcirc took about 1,593 total CPU hours (Intel (R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz) to generate the predicted 3D models for PIP5K1C molecule (consisting of 249 nucleotides). In addition, the relationship between runtime of IsRNAcirc and the length of circular RNA was also analyzed and the result was displayed in Fig 8. Notably, the total CPU time for predicting circular RNA 3D structure using IsRNAcirc increases almost linearly with the length of RNA molecules, mainly due to the coarse-grained nature of IsRNAcirc method.

Since IsRNAcirc employed the end closure step to close the 5'- and 3'-end of the initial 3D structures in the prediction process, it is worth exploring whether the runtime will be reduced if we use the circular 3D structures predicted by 3dRNA as the starting states. As shown in S9 Fig, the total runtime was only slightly reduced when we launched the predictions using the



**Fig 8. Runtime (total CPU time) in IsRNAcirc as a function of length of circular RNAs.** Circles represent 34 circular RNAs and the line indicates the fitting result. Predictions were performed on intel(R) Xeon(X) CPU E5-2620 v4 @ 2.10GHz.

<https://doi.org/10.1371/journal.pcbi.1012293.g008>

3dRNA predicted initial circular 3D structures compared to the RNAComposer generated linear 3D structures. This result is predictable if we considered the presence of the unfolding-refolding process in the end closure step to repair the potential unreasonable elements included in the initial structures, indicating that our simulations in IsRNAcirc are relatively insensitive to the starting state.

To further enhance the usability and practicality of IsRNAcirc in predicting long-size circular RNA 3D structures, several approaches will be considered in the future to improve computational efficiency, such as optimizing the source code, utilizing parallel computing technique, and introducing graphics processing unit (GPU)-based acceleration.

## Conclusion

In summary, based on our previous IsRNA2 method for linear RNA 3D structure prediction, we have developed an extended IsRNAcirc model to enable 3D structure prediction of circular RNAs. Based on the CG representation of nucleotides and the derived force field, IsRNAcirc employs REMD simulations to sample the huge conformational space and generate the most probable structures as the predicted 3D models of circular RNAs. Compared with the initial 3D structure of the linear counterpart obtained from RNAComposer, the local topologies of circular RNA 3D models predicted by IsRNAcirc show significant improvements. Although the performance of IsRNAcirc in predicting circular RNA 3D structures cannot be directly assessed at present due to the lack of experimental 3D structures, an indirect evaluation was performed utilizing energy-based score functions and then compared with the template-based

3dRNA method. For a test set comprising 34 circular RNAs (spanning in size from 161 to 435 nucleotides), our IsRNACirc can generate 3D models with lower scores than 3dRNA in all the cases, regardless of the circularized type of circular RNAs. These results demonstrate that the proposed IsRNACirc is a promising method to explore the structural details along with intricate interactions of circular RNAs.

## Methods

### Simulation details

IsRNACirc uses the same force field as IsRNA2, and the details of this force field can be found in our previous work [26]. All MD simulations were implemented in the LAMMPS [38] software with modified source code, and the Langevin dynamics (NVT ensemble) with an integration timestep  $\Delta t = 1fs$  was performed. During the end closure stage (closing the 5'- and 3'-end of circular RNAs), a harmonic potential of equilibrium length  $b_0 = 3.8\text{\AA}$  and a gradually increasing force constant (from  $k = 0.001$  to  $5.0 \text{ kcal/mol/\AA}^2$ ) were imposed on bead P of the first nucleotide and bead S of the last nucleotide. Simulated annealing simulations (from temperature  $T = 500$  to  $300$  K) with a length of  $10\text{--}50$  ns (depending on the separated distance between the two ends) were run to sufficiently relax the initial structures and avoid unphysical structures. Meanwhile, to reduce the locally irrational elements involved in the initial structures, weakened base-pairing interactions (multiplied by a factor of  $0.01/0.1/1.0$ ) were also used during the simulated annealing simulations. For each initial structure, the most probable conformation (identified by clustering method) in the last 10% MD trajectory was selected as the starting structure for the subsequent simulation. During the structure prediction stage, REMD simulations with 10 replicas with temperatures ranging from  $280$  to  $460$  K were performed to efficiently sample the 3D conformational space. The simulation time for each replica is  $100$  ns, and three duplicated runs with different starting structures (if available) were run. Here, standard energy parameters were used for the relevant bonds/bond angles/torsion angles between two end nucleotides. After sufficient relaxation, the structure snapshots were collected from the last  $50$  ns simulations in the interval of  $100$  ps to construct the conformational ensemble ( $15,000$  structures in total). Then, following the same steps as in IsRNA2, the top 10% structures with the lowest potential energies were submitted to the clustering procedure to generate the 3D predictions.

### Input and output

IsRNACirc requires the sequence information, 2D structure (in dot-bracket format), and an initial 3D structure (in PDB format) of the corresponding linear RNA as input to predict circular RNA 3D structures. From the sequence information, the 2D structure of circular RNAs was predicted by the cRNAsp12 tool [31] (<http://xxulab.org.cn/crnasp12/>) using the default set. As multiple 2D structure predictions were provided by cRNAsp12, we selected the structure with the lowest folding free energy for further analysis. The 2D structure predicted by cRNAsp12 is a closed chain (in dot-bracket format) whose start and end points are the head-to-tail junction (BSJ) of the circular RNA, and the sequence information is also presented in the same way. When we ignored this BSJ, the linear counterpart of the circular RNA was generated in a similar sequence/dot-bracket 2D representation. It should be noted that when converting the circular RNA sequence to its linear counterpart, the choice of the cleavage point will more or less affect the structural topology of the initial 3D structures and thus the prediction results. However, for simplicity, we chose the start/end point of the circular sequence as the cleavage site here. We anticipated that the MD simulations performed during the end closure step could partially offset the structural disruptions caused by the introduction of this cleavage site. We

will further address this issue in the future to minimize structural disruptions during the conversion process. Then, due to its good performance in predicting linear RNA 3D structures and out of personal preference, we used RNAComposer [18] (<https://rnacomposer.cs.put.poznan.pl/>) to generate three (if available) initial 3D structures for the corresponding linear counterparts based on sequence information and predicted 2D structures. Actually, the 3D structures predicted by other programs (such as 3dRNA) can also be used as input initial structures. Additionally, another input of IsRNAcirc is a configuration file that specifies relevant parameters of the prediction process, which includes the simulation steps (simulation time), output parameters for collection of the conformational ensemble, cutoff threshold for clustering, the desired number of predicted models, and other relevant parameters.

The primary output of IsRNAcirc is the predicted 3D models (in PDB format) for given circular RNAs. By default, five 3D models are generated, which are the centroid structures of the top five largest clusters (if available) and then recovered to all-atom models through a built-in single-nucleotide fragment match algorithm [26]. In addition, users can flexibly obtain different numbers of predicted 3D models by modifying the relevant output parameter in the configuration file. Optionally, QRNAS [32] was employed to refine the predicted models by fixing possible errors in local geometries (such as unphysical bond lengths) and reducing the clash scores between atoms. Moreover, other useful outputs include simulation trajectories and corresponding energy files, log files of the clustering procedure, *etc.*

### Test dataset collection

To evaluate the predictive power of IsRNAcirc, a test dataset was collected from the previous study [30]. This test dataset consists of 34 circular RNAs with varying lengths ranging from 161 to 435 nucleotides, as listed in [S1 Table](#). According to the location of BSJ in the 2D structure, these 34 circular RNAs can be categorized into four groups: 11 helix-circular RNAs, 5 hairpin-circular RNAs, 13 internal-circular RNAs, and 5 junction-circular RNAs.

### Definition of four types of locally irrational elements

Various locally irrational elements were observed in the initial 3D structures generated by RNAComposer. Based on their structural characteristics, we divided them into four different categories: template-free loop, backbone tangle, base collision, and sharp twist. Template-free loops are loop fragments consisting of multiple nucleotides, each of which exhibits a similar pose, which is extremely physically implausible. The backbone tangle is a tangle of one or two backbones (a phosphate group and a ribose ring linked by a 3',5'-phosphodiester bond), which resembles the "sticky knot" formed by silk thread winding. Base collision is a situation in which the backbone of a circular RNA passes between two bases on the backbone of another location, resulting in atomic clashes between the 3D structures. The sharp twist is a sharp bend in the backbone of a circular RNA. See [Fig 2](#) for an illustrative example of each type of locally irrational element.

### Supporting information

**S1 Fig. Overview of the coarse-grained (CG) IsRNA2 model for linear RNA 3D structure prediction.** (A) Mapping relationships between the all-atom model and CG representation in IsRNA2: the backbone (colored by green) is represented by two CG beads (bead P located on the atom P for the phosphate group and bead S located on the atom C4' for the ribose ring) and the nucleobase (colored by pink) is represented by three CG beads (each bead located at the center of mass of the related heavy-atom group). (B) An illustrated example to display the workflow for linear RNA 3D structure prediction by IsRNA2 model. Using sequence, 2D

structure, and initial 3D structure as input, IsRNA2 employs a four-step process (conformation sampling through REMD, conformational ensemble construction, clustering procedure, and all-atom reconstruction and energy minimization) to generate the predicted 3D structures. For reference, the RMSD of each predicted model relative to the native structure (PDB ID: 1Z5C) was also shown.

(TIF)

**S2 Fig.** Illustrative examples depict the structural details of the 5' and 3' terminals of four distinct types of circular RNAs: (A) helical-circular, (B) hairpin-circular, (C) internal-circular, and (D) junction-circular RNA. From left to right, the secondary structure of the back-splice junction (BSJ) region, the 3D structure of the 5' and 3' terminals in the initial structure, and the circularized terminal nucleotides by IsRNAcirc.

(TIF)

**S3 Fig.** (A) Box plot of root-mean-square deviations (RMSDs) between the initial structure generated by RNAComposer and the predicted 3D structure by IsRNAcirc for four types of circular RNAs. The RMSD values were calculated by aligning the initial 3D structure with the 3D structure predicted by IsRNAcirc using PYMOL software. Boxes, interquartile range (IQR); center lines, median; black dots, mean; whiskers, values within  $1.5 \times$  IQR of the top and bottom quartiles. (B) Scatter plot of RMSD value as a function of the length of circular RNA. Values are mean  $\pm$  s.d. ( $n = 12/15$  independent RMSD values).

(TIF)

**S4 Fig.** (A) Box plots of the pairwise RMSDs between the circular 3D models predicted by 3dRNA and their corresponding linear 3D structures generated by RNAComposer for all 34 tested circular RNAs. (B) Superposition of circular 3D structures predicted by IsRNAcirc and their corresponding linear 3D structure predicted by IsRNA2 for two constructed RNA molecules: helix-circular RNA (left, sequence: UCGUAAAAACGAUCGAAAAACGA; 2D structure: “((( (. . . )))((( (. . . )))”) and hairpin-circular RNA (right, sequence: AACGAUCGUA AAAAACGAUCGAAA; 2D structure: “..(((((((. . . ))))))). . .”).

(TIF)

**S5 Fig.** Detailed number of irrational structures in the 3D structure of each tested circular RNA: (A) helix-circular, (B) hairpin-circular, (C) internal-circular, and (D) junction-circular RNAs. (E) Total number of irrational structures for four types of circular RNAs. The initial structure of corresponding linear RNA and the predicted circular RNA structure by IsRNAcirc are shown in orange and blue bars, respectively.

(TIF)

**S6 Fig. Predicted 3D structure of the L1 ribozyme ligase circular addcut (PDB ID: 2OIU) by using IsRNAcirc.** The sequence and native 2D structure extracted from the experimental structure were used as input. The experimental structure and the predicted 3D structure are colored in red and green, respectively.

(TIF)

**S7 Fig.** (A) Box plots of the pairwise RMSDs between the IsRNAcirc and 3dRNA predicted 3D structures for all 34 tested circular RNAs. (B) Superposition of the 3D structures of the circKIAA0368(435 nt) molecule predicted by IsRNAcirc and 3dRNA. (C) Superposition of the 3D structures of the circASXL1(195 nt) circular RNA predicted by IsRNAcirc and 3dRNA.

(TIF)

**S8 Fig. Comparison of docking scores (binding energies) of circular RNA-PKR complexes (blue symbols) and its linear cognate RNA-PKR complexes (orange symbols).** Three 3D models were prepared for the circPOLR2A molecule (predicted by IsRNAcirc) and its linear counterpart (predicted by IsRNA2), respectively. For each RNA 3D model, we used HDock to generate the top ten possible binding poses and recorded the associated docking scores. The mean±s.d. of the docking scores for each RNA 3D model was also shown.

(TIF)

**S9 Fig. Comparison of IsRNAcirc prediction runtimes for four typical circular RNA molecules starting from different initial structures.** The runtimes for the end closure and structure prediction (REMD) stages were recorded separately. Predictions starting from the 3dRNA predicted circular 3D structure and the RNAComposer generated linear 3D structure were shown in orange and blue bars, respectively.

(TIF)

**S1 Table. Summary of 34 circular RNAs in the test set.**

(XLSX)

**S1 Movie. An illustrative example of the unfolding-refolding process in the end closure step.**

(MP4)

## Author Contributions

**Conceptualization:** Yunguang Tong, Dong Zhang, Ruhong Zhou.

**Data curation:** Haolin Jiang, Yulian Xu.

**Formal analysis:** Haolin Jiang, Yulian Xu, Dong Zhang.

**Funding acquisition:** Yunguang Tong, Dong Zhang, Ruhong Zhou.

**Investigation:** Haolin Jiang, Yulian Xu.

**Methodology:** Haolin Jiang, Dong Zhang.

**Project administration:** Dong Zhang.

**Resources:** Haolin Jiang, Yulian Xu.

**Software:** Haolin Jiang, Dong Zhang.

**Supervision:** Yunguang Tong, Dong Zhang, Ruhong Zhou.

**Validation:** Haolin Jiang, Yunguang Tong, Dong Zhang.

**Visualization:** Haolin Jiang, Yulian Xu.

**Writing – original draft:** Haolin Jiang, Yulian Xu.

**Writing – review & editing:** Haolin Jiang, Yulian Xu, Yunguang Tong, Dong Zhang, Ruhong Zhou.

## References

1. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U. S. A.* 1976; 73(11):3852–6. <https://doi.org/10.1073/pnas.73.11.3852> PMID: 1069269

2. Chen N, Zhao G, Yan X, Lv Z, Yin H, Zhang S, et al. A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biol.* 2018; 19(1):218. <https://doi.org/10.1186/s13059-018-1594-y> PMID: 30537986
3. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet.* 2013; 9(9):e1003777. <https://doi.org/10.1371/journal.pgen.1003777> PMID: 24039610
4. Gao Y, Wang J, Zheng Y, Zhang J, Chen S, Zhao F. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.* 2016; 7(1):12060. <https://doi.org/10.1038/ncomms12060> PMID: 27350239
5. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 2010; 28(5):511–515. <https://doi.org/10.1038/nbt.1621> PMID: 20436464
6. Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 2016; 7(1):11215. <https://doi.org/10.1038/ncomms11215> PMID: 27050392
7. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013; 495(7441):333–338. <https://doi.org/10.1038/nature11928> PMID: 23446348
8. Zhang Y, Zhang X, Chen T, Xiang J, Yin Q, Xing Y, et al. Circular Intronic Long Noncoding RNAs. *Mol. Cell* 2013; 51(6):792–806. <https://doi.org/10.1016/j.molcel.2013.08.017> PMID: 24035497
9. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 2015; 22(3):256–264. <https://doi.org/10.1038/nsmb.2959> PMID: 25664725
10. Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, et al. 5' UTR m6A Promotes Cap-Independent Translation. *Cell* 2015; 163(4):999–1010. <https://doi.org/10.1016/j.cell.2015.10.012> PMID: 26593424
11. Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian S. Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* 2015; 526(7574):591–594. <https://doi.org/10.1038/nature15377> PMID: 26458103
12. Wang Y, Wang Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* 2015; 21(2):172–179. <https://doi.org/10.1261/rna.048272.114> PMID: 25449546
13. Abe N, Matsumoto K, Nishihara M, Nakano Y, Shibata A, Maruyama H, et al. Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* 2015; 5(1). <https://doi.org/10.1038/srep16435> PMID: 26553571
14. Robertson MP, Scott WG. The structural basis of ribozyme-catalyzed RNA assembly. *Science* 2007; 315(5818):1549–53. <https://doi.org/10.1126/science.1136231> PMID: 17363667
15. Rother M, Rother K, Puton T, Bujnicki JM. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic. Acids. Res.* 2011; 39(10):4007–22. <https://doi.org/10.1093/nar/gkq1320> PMID: 21300639
16. Xu X, Zhao P, Chen SJ. Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One* 2014; 9(9):e107504. <https://doi.org/10.1371/journal.pone.0107504> PMID: 25215508
17. Li J, Zhang S, Zhang D, Chen S. Vfold-Pipeline: a web server for RNA 3D structure prediction from sequences. *Bioinformatics* 2022; 38(16):4042–4043. <https://doi.org/10.1093/bioinformatics/btac426> PMID: 35758624
18. Biesiada M, Pachulska-Wieczorek K, Adamiak RW, Purzycka KJ. RNAComposer and RNA 3D structure prediction for nanotechnology. *Methods* 2016; 103:120–127. <https://doi.org/10.1016/j.ymeth.2016.03.010> PMID: 27016145
19. Zhang Y, Wang J, Xiao Y. 3dRNA: 3D Structure Prediction from Linear to Circular RNAs. *J. Mol. Biol.* 2022; 434(11):167452–167452. <https://doi.org/10.1016/j.jmb.2022.167452> PMID: 35662453
20. Zhang Y, Wang J, Xiao Y. 3dRNA: Building RNA 3D structure with improved template library. *Comp. Struct. Biotechnol. J.* 2020; 18:2416–2423. <https://doi.org/10.1016/j.csbj.2020.08.017> PMID: 33005304
21. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, et al. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 2009; 15(2):189–199. <https://doi.org/10.1261/rna.1270809> PMID: 19144906
22. Sharma S, Ding F, Dokholyan NV. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* 2008; 24(17):1951–1952. <https://doi.org/10.1093/bioinformatics/btn328> PMID: 18579566
23. Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic. Acids. Res.* 2016; 44(7):e63–e63. <https://doi.org/10.1093/nar/gkv1479> PMID: 26687716

24. Zhang D, Chen S. IsRNA: An Iterative Simulated Reference State Approach to Modeling Correlated Interactions in RNA Folding. *J. Chem. Theory Comput.* 2018; 14(4):2230–2239. <https://doi.org/10.1021/acs.jctc.7b01228> PMID: 29499114
25. Zhang D, Li J, Chen S. IsRNA1: *De Novo* Prediction and Blind Screening of RNA 3D Structures. *J. Chem. Theory Comput.* 2021; 17(3):1842–1857. <https://doi.org/10.1021/acs.jctc.0c01148> PMID: 33560836
26. Zhang D, Chen S, Zhou R. Modeling Noncanonical RNA Base Pairs by a Coarse-Grained IsRNA2 Model. *J. Phys. Chem. B* 2021; 125(43):11907–11915. <https://doi.org/10.1021/acs.jpcc.1c07288> PMID: 34694128
27. Baek M, Mchugh R, Anishchenko I, Jiang H, Baker D, DiMaio F. Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* 2024; 21(1):22–23. <https://doi.org/10.1038/s41592-023-02086-5> PMID: 37996753
28. Shen T, Hu Z, Peng Z, Chen J, Xiong P, Liang H, et al. E2Efold-3D: End-to-End Deep Learning Method for accurate de novo RNA 3D Structure Prediction. In. Ithaca: Cornell University Library, [arXiv.org](https://arxiv.org); 2022.
29. Li Y, Zhang C, Feng C, Pearce R, Freddolino PL, Zhang Y. Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nat. Commun.* 2023; 14(1):5745. <https://doi.org/10.1038/s41467-023-41303-9> PMID: 37717036
30. Liu C, Li X, Nan F, Jiang S, Gao X, Guo S, et al. Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell* 2019; 177(4):865–880.e21. <https://doi.org/10.1016/j.cell.2019.03.046> PMID: 31031002
31. Wang F, Li W, Li B, Xie L, Tong Y, Xu X. cRNAsp12 Web Server for the Prediction of Circular RNA Secondary Structures and Stabilities. *Int. J. Mol. Sci.* 2023; 24(4):3822. <https://doi.org/10.3390/ijms24043822> PMID: 36835231
32. Stasiewicz J, Mukherjee S, Nithin C, Bujnicki JM. QRNAS: software tool for refinement of nucleic acid structures. *BMC Struct. Biol.* 2019; 19(1):5. <https://doi.org/10.1186/s12900-019-0103-1> PMID: 30898165
33. Wang J, Zhao Y, Zhu C, Xiao Y. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic. Acids. Res.* 2015; 43(10):e63. <https://doi.org/10.1093/nar/gkv141> PMID: 25712091
34. Capriotti E, Norambuena T, Marti-Renom MA, Melo F. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* 2011; 27(8):1086–1093. <https://doi.org/10.1093/bioinformatics/btr093> PMID: 21349865
35. Tan Y, Wang X, Shi Y, Zhang W, Tan Z. rsRNASP: A residue-separation-based statistical potential for RNA 3D structure evaluation. *Biophys. J.* 2022; 121(1):142–156. <https://doi.org/10.1016/j.bpj.2021.11.016> PMID: 34798137
36. Zhang T, Hu G, Yang Y, Wang J, Zhou Y. All-Atom Knowledge-Based Potential for RNA Structure Discrimination Based on the Distance-Scaled Finite Ideal-Gas Reference State. *J. Comput. Biol.* 2020; 27(6):856–867. <https://doi.org/10.1089/cmb.2019.0251> PMID: 31638408
37. Yan Y, Di Zhang, Zhou P, Li B, Huang S. HDock: a web server for protein-protein and protein-DNA / RNA docking based on a hybrid strategy. *Nucleic. Acids. Res.* 2017; 45(W1):W365–W373. <https://doi.org/10.1093/nar/gkx407> PMID: 28521030
38. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 1995; 117(1):1–19. <https://doi.org/10.1006/jcph.1995.1039>