RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers
Supplementary material

Authors
Zhichao Miao*, Ryszard W. Adamiak, Maciej Antczak, Shi-Jie Chen, Clarence Yu Cheng, Yi Cheng, Fang-Chieh Chou, Rhiju Das, Nikolay V. Dokholyan, Feng Ding, Caleb Geniesse, Yangwei Jiang, Astha Joshi, Andrey Krokhotin, Marcin Magnus, Olivier Mailhot, Francois Major, Thomas H. Mann, Paweł Piątkowski, Radoslaw Pluta, Mariusz Popenda, Joanna Sarzynska, Lizhen Sun, Marta Szachniuk, Jian Wang, Jun Wang, Andrew M. Watkins, Jakub Wiedemann, Yi Xiao, Xiaojun Xu, Joseph D. Yesselman, Dong Zhang, Yi Zhang, Zhenzhen Zhang, Chenhan Zhao, Peinan Zhao, Yuanzhe Zhou, Tomasz Zok, Adriana Żyła, Ren Aiming, Robert T. Batey, Barbara L. Golden, Lin Huang, David M. Lilley, Yijin Liu, Dinshaw J. Patel, Eric Westhof*

*To whom correspondence should be addressed:
Zhichao Miao: zmiao@ebi.ac.uk
Eric Westhof: e.westhof@ibmc-cnrs.unistra.fr

Abstract

RNA-Puzzles is a collective endeavor dedicated to the advancement and improvement of RNA 3D structure prediction. With agreement from crystallographers, the RNA structures are predicted by various groups before the publication of the crystal structures. We now report the prediction of 3D structures for six RNA sequences: four nucleolytic ribozymes and two riboswitches. Systematic protocols for comparing models and crystal structures are described and analyzed. In these six puzzles, we discuss a) the comparison between the automated web servers and human experts; b) the prediction of coaxial stacking; c) the prediction of structural details and ligand binding; d) the development of novel prediction methods; and e) the potential improvements to be made. We show that correct prediction of coaxial stacking and tertiary contacts is essential for the prediction of RNA architecture, while ligand binding modes can be only predicted with low resolution and simultaneous prediction of RNA structure with accurate ligand binding still remains out of reach. All the predicted models are available for the future development of force field parameters and the improvement of comparison and assessment tools.

Supplementary Notes

The prediction methods

Adamiak-Szachniuk (human expert group) and RNAComposer (webserver)

The Adamiak-Szachniuk group has participated in RNA-Puzzles in both webserver and human categories. Both predictions used RNAComposer (Antczak et al. 2016; Purzycka et al. 2015; Popenda et al. 2012). According to the general processing scheme, this system shreds RNA secondary structure, searches the corresponding 3D structure elements in the database being a dedicated, modified version of RNA FRABASE (Popenda et al. 2010), processes and assembles them in a 3D model, and finally, it applies RNA 3D structure refinement by the energy minimization.

For the RNA-Puzzles web server prediction category, the Adamiak-Szachniuk group developed a fully-automated RNAComposer-based service that predicts RNA 3D structure either from the sequence or secondary structure. The service, designed as a four-layer application, continuously listens to the tasks upcoming in electronic messages. Its layers are responsible for (i) preliminary data processing, (ii) RNA 3D structure prediction, (iii) evaluation and ranking of the generated 3D models, and (iv) preparation of output files that are emailed to the task owner. The first layer procedures filter non-canonical base-pairs in RNA secondary structure, process multi-stranded RNAs and decode multi-line representations of pseudoknotted structures (Antczak et al. 2018). The second layer integrates ten RNA secondary structure prediction tools (RNAstructure, RNAfold, CONTRAfold, ContextFold, CentroidFold, IPknot, RNAshapes, HotKnots, CentroidAlifold, and RNAalifold) (Puton et al. 2013). All of them are run to generate a wide set of unique RNA secondary structures that are next fragmented into predefined motifs (loops, double- and single-stranded fragments). Every motif constitutes a search pattern in the exploration of the enlarged database (unpublished) of RNA 3D structure elements. The extended repository of data ensures higher structural diversity of output models. For each secondary structure, up to ten 3D models are predicted. The third layer of the system engages mechanisms of RNApdbee and RNAQUA (Zok et al. 2018; Szachniuk 2019; Magnus et al. 2020; Antczak et al. 2014) to evaluate predicted 3D models based on total energy obtained from XPLOR, gyration radius, $INF_{wc,}$ and the Clash score. All scores are aggregated to rank the models and select up to five most promising ones.

The breadth of the RNAComposer-based system is reflected in the prediction results. For example, an integration of HotKnots allowed us to anticipate pseudoknots in the fully automated prediction of PZ21. The new procedure of handling multi-chain structures was applied in PZ15, PZ19, PZ20. For PZ15, the models with the lowest energy were obtained when the three-way junction with the isolated base-pair C16-G21 (preliminary not encoded in the secondary structure) was automatically included. Structure elements for the junction

were derived from two hammerhead ribozyme structures (PDB IDs: 2GOZ and 2QUS). The best RNAComposer models submitted to Puzzle 15 had RMSD values below 1.6 Å when the fragment corrupted by a crystal contact in the target structure 5DI4 was omitted in the comparison (i.e., RMSD was computed for the 5DI4_delta structure which contained residues A:11-48 and B:1-12 only). For PZ19 and PZ20, automated predictions of RNAComposer were the best in the webserver category with RMSD 9.60 Å and 11.0 Å, respectively. For PZ19, all RNAComposer models had correctly predicted the stacking between P4 and P5.

In the human prediction category, the Adamiak-Szachniuk group used an additional database of RNA loops (unpublished), which was updated before each RNA-Puzzle challenge. A new search algorithm, named IMMG (Imperfect Matching Mixed Graph algorithm) (Zok 2017), was applied to find 3D elements with a secondary structure similar to the pattern which might not exactly match it (*e.g.*, loops with isolated pairs that have not been encoded in the pattern). The algorithm significantly increases the number of 3D structure elements found and overcomes the problem of missing elements, which is the common limitation of fragment assembly methods. IMMG algorithm was run in human predictions PZ19, PZ20, PZ21. The resulting motifs were clustered based on RMSD values using the OC cluster analysis program (Barton 1993). Cluster centroids were introduced as user-own 3D structure elements to the new version of RNAComposer (Antczak et al. 2016; Purzycka et al. 2015). Such an approach led to a diverse conformational space, from which 5 models of different coaxial stacking of the helices in the junctions (PZ19, PZ20) were selected. In the case of PZ21, the structures were selected from the family with the coaxial alignment of P1 and P2. All selected models were validated using the RCSB Maxit Suite (https://sw-tools.rcsb.org/apps/MAXIT/index.html) concerning close atom contacts, bond lengths, angles, planarity, chirality, and oligoribonucleotide linkage.

Bujnicki (human expert group) and SimRNAweb (webserver)

The Bujnicki group submitted RNA 3D structure models constructed with extensive human intervention, as well as models generated by an automated approach.

The Bujnicki group of human experts used a hybrid modeling strategy (Piatkowski et al. 2016) based on the approach tested in the previous editions of the RNA-Puzzles experiment (Cruz et al. 2012; Miao et al. 2015, 2017). If the target sequence exhibited detectable similarity to an RNA with known experimentally determined structure (as happened in the case of Puzzles 9 & 15), the Bujnicki group generated models of the whole molecule or its parts using a template-based (comparative) modeling method ModeRNA (Rother, Rother, et al. 2011) or its server version (Rother, Milanowska, et al. 2011). Remodeling of uncertain regions and modeling of RNA molecules that lacked suitable templates relied mostly on template-free folding using the coarse-grained method SimRNA (Boniecki et al. 2016). Wherever available, template-free folding was aided by spatial restraints obtained from computational predictions (*e.g.* information on the secondary structure or orientation of helices injunctions) and from data found in the literature and in databases, or made available in the course of the RNA Puzzles experiment by the Das group. In particular, for secondary structure prediction, Bujnicki group used a consensus of information taken from various

sources, including predictions calculated by several methods that topped the CompaRNA benchmark (Puton et al. 2013). Following the simulations, structures were clustered and medoids of the largest clusters were taken as candidates for submission if they passed the subjective criterion of visual inspection. For the high-resolution refinement of models, the Bujnicki group used the QRNAS method (Stasiewicz et al. 2019) that extends the AMBER (Junmei Wang et al. 2004) force field with energy terms explicitly modeling hydrogen bonds, idealizes base-pair planarity and regularizes the backbone conformation.

For fully automated predictions, the SimRNAweb server (run with default parameters) (Magnus et al. 2016) was used to generate models, with the exception that restraints on predicted secondary structure was used.

Starting from Puzzle 17, the group expanded the methodology to include a new approach for modeling, EvoClustRNA (Magnus et al. 2019). Briefly, the Bujnicki group modeled not only the target sequence, but also a few selected homologs, and attempted to find a common structure shared by the different sequences. EvoClustRNA was used to generate additional models and to influence the selection of the final models to be submitted.

Puzzle 9

To solve this Puzzle, the group carried out comparative modeling with ModeRNA, based on the guanine riboswitch aptamer structure (PDB ID: 4FEJ). All the submitted RNA structure models were the same (as it turned out: RMSD of 6.19 Å to the reference, the second-best score for this Puzzle), and the group focused on predicting the position of the ligand. The ligand was modeled manually, based on the small molecule ligand structure pose in the template structure. The modeled ligand RMSD ranged from 2.68 Å to 3.2 Å.

Puzzle 15

This Puzzle was modeled by the Bujnicki group only in the fully automated way by the SimRNAweb server. There were two rounds: (1) only sequence information and (2) sequence and secondary structure information. Interestingly, SimRNAweb run based on sequence generated a relatively accurate model (the 3rd in the final ranking in terms of RMSD, 7.12 Å), so the additional information on secondary structure not only did not help SimRNAweb but actually decreased the prediction accuracy. Since there were structures available in the PDB for some of the hammerhead ribozymes, it is not surprising that SimRNAweb as a template-free method was outperformed by RNAComposer, a fragment assembly method. This suggests that the future SimRNAweb server could benefit from jump-starting the predictions from template-based models generated by other methods, such as ModeRNA or RNAComposer.

Puzzle 17

SimRNAweb predictions without any additional information generated models with RMSD to the reference of 5.16 Å, which was the model ranked as the best among all submitted models in terms of RMSD.

Puzzle 19

Puzzle 19 was modeled using the standard protocol, with the use of additional restraints on coaxial stacking according to predictions made with the Cartaj (Lamiable et al. 2012) web server. The best model from the group was ranked as 7th in the final ranking in terms of RMSD (10.18 Å).

Puzzle 20

The Puzzle was modeled using the standard protocol using the SimRNAweb server. The 1st and 4th models submitted by the group demonstrated agreement with the reference structure. The 1st model was ranked as the best among all submitted models in terms of INF_{all} of 0.90 and the 4th ranked the best with a full-atom RMSD of 4.71 Å.

Puzzles 21

The Puzzle was modeled using the standard protocol with the SimRNAweb server using manually curated secondary structures based on the literature. The 4th model submitted by the human group was ranked as the 5th in the final ranking in terms of INF_{all} of 0.72 and with a full-atom RMSD of 6.54 Å.

Chen (human expert group)

Briefly, the Chen group used a hierarchical approach to model the 3D structures for Puzzle challenges. From the sequence, the 2D structure was predicted by a free energy-based model named Vfold2D (Cao and Chen 2005; X. Xu, Zhao, and Chen 2014; Cao and Chen 2006). Base-pairs derived from Rfam, if available, were integrated into the 2D structure prediction. If available, structural information from experiments was incorporated in the structure prediction. 3D structures were modeled using the template-based Vfold3D model (X. Xu, Zhao, and Chen 2014; Cao and Chen 2011; X. Xu and Chen 2015).

To alleviate the problem of limited availability in structural templates, a new template search and assembly algorithm through hierarchical loop template assembly, namely VfoldLA (X. Xu and Chen 2018; X. Xu, Zhao, and Chen 2019), was developed and applied to the 3D structure prediction. Unlike the previous Vfold3D models, where the templates were found based on the whole motif or piecewise fragments, the VfoldLA model searches for 3D templates based on single strands of loops/junctions. Therefore, the VfoldLA model has a much higher success rate of finding the appropriate 3D templates than the previous Vfold3D approach and has a much higher computational efficiency than fragment-based methods. The predicted 3D candidate structures were ranked according to the sequence similarities between the target sequence and that of the adopted templates, and the one with the highest sequence similarity was selected as the top-ranked model.

Additionally, the Chen group employed the IsRNA (D. Zhang and Chen 2018) Molecular Dynamics simulation to predict RNA tertiary structure from the 2D structure constraints. Compared with previous coarse-grained models, IsRNA uses an improved force field derived from a novel iterative simulated reference state method. A unique advantage of IsRNA is the consideration of correlations between the different degrees of freedom in a structure (different energy terms in the force field) as well as the inherent chain connectivity and excluded volume. Moreover, since the simulated reference states include both the native

and the non-native interactions, the IsRNA force field can account for the effect of the full energy landscape, including both the native and nonnative conformations.

In the IsRNA-based structure prediction, conformational clustering was used to select the structure candidates. Specifically, from the simulation trajectories, structure snapshots of the top 10% lowest potential energies were collected and clustered on the basis of pairwise RMSDs. The centroid structure of the largest cluster was selected as the model 1 and the centroid structure of the second-largest cluster was model 2, and so forth. The selected models were further refined by all-atom energy minimization in NAMD (M. T. Nelson et al. 1996) with the CHARMM force field (Brooks et al. 2009).

For Puzzle 9, the 3D structures were predicted by the Vfold3D model. Because the algorithm found an ideal template for the structure, the Vfold3D model gave the best prediction (lowest RMSD) among all the submissions. For Puzzles 15 and 17, both the Vfold3D and VfoldLA models were used to obtain the 3D predictions. For the remaining Puzzles 19, 20, and 21, initial 3D structures were generated by the Vfold3D model, then IsRNA coarse-grained Molecular Dynamics simulations were used to refine and rank the predicted models. For Puzzle 21, the ligand guanidine was docked to the candidate 3D structures through an in-house program. For high ligand concentration (“HighLig”), two potential binding poses were predicted, while for the low ligand concentration (“LowLig”), a single optimal binding pose was found.

Das group (Rosetta-based methods: FARFAR, LORES, Stepwise and RW3D webserver)

Predictions from the Das group for the RNA-Puzzles competition used the Rosetta modeling software (Das, Karanicolas, and Baker 2010), particularly the FARFAR algorithm. This method starts from manually identified local template structures and idealized helix conformations for each helix in the assumed secondary structure, builds in remaining nucleotides from a fragment library in numerous Monte Carlo trajectories, minimizes a large pool of 1,000-10,000 fragment-assembled models in a high-resolution, all-atom force field, and then selects models based on energy, with some re-ordering based on visual inspection. See references (Cheng, Chou, and Das 2015; Watkins, Rangan, and Das 2019) for further information. FARFAR also formed the basis for a pilot server called “RNA works 3D” (RW3D), developed for Puzzle 15 and not previously described.

Automated predictions with RW3D relied on the secondary structure as given by ViennaRNA and did not include any attempt to identify templates from the previously solved structure. For the other five targets (Puzzles 9, 17, 19, 20, and 21), predictions used manually constructed secondary structures based on information available in the literature and from Rfam. These secondary structures were broadly accurate, as reflected in excellent INF_{wc} metrics for the predictions, whereas secondary structures from ViennaRNA predicted one or more incorrect helices for puzzles 9, 17, and 21. In addition, manual predictions were informed by. some template identifications, as follows. For PZ9, the Das group used a kissing loop interaction from the purine riboswitch deposited as 2XNW, as well as a T-loop from the tRNA Phe structure 3L0U. For PZ17, the Das group used only secondary structure

information to generate predictions. For PZ19, the Das group used an intercalated T-loop from the tRNA Cys structure 1B23, and considered intercalation from adenosines at nucleotides A8, A9, and B15; the simulations using an intercalating A8 were preferred as they had better energies. For PZ20, since it was identifiable as another twister sister ribozyme, the Das lab used the exact intercalated T-loop from the PZ19 crystal structure 5T5A, renumbered to correspond to the PZ20 context (i.e., since the P1 stem was one base-pair longer, assuming that A7 in PZ20 corresponded to A8 in PZ19). All input template structures used in modeling of these targets may be found at https://github.com/DasLab/rna_benchmark, under input_files/all_rna_puzzles. For the guanidine-III riboswitch (PZ21), the Das group did not use templates but did test updates to Rosetta that allowed the inclusion of the guanidine ligand through an assumed planar 'pseudo-base-pair' to a conserved guanosine in the riboswitch, which turned out to be correct. In addition, PZ21 offered tests of the stepwise Monte Carlo algorithm, an alternative to FARFAR (Watkins et al. 2018). This method, which offered best-submitted models for another puzzle (PZ18; see (Watkins et al. 2018)) avoids the use of fragments or low-resolution scoring. For PZ21, the Das group submitted stepwise Monte Carlo models as "Das" and models predicted using FARFAR as "LORES." In fact, the LORES (FARFAR) models provided the best overall prediction, while the stepwise Monte Carlo simulation over-converged on models with a register shift in the placement of a pseudoknot loop.

On the whole, FARFAR modeling on these targets confirmed the utility of Rosetta approaches across previously unseen folds, with submitted models typically placing among the top 5 amongst all submissions and giving visually correct folds in all puzzles, similar to the previous three rounds of RNA-Puzzles. Modeling still required manual recognition of secondary structures, ligand binding sites, and potential tertiary contacts (such as intercalating T-loops) as well as the curation of models. In terms of improving automation, secondary structure predictions from automated methods like Vienna (used in RW3D) remain worse than literature secondary structure models that incorporate phylogenetic information. Identification of tertiary contacts between regions distal in the secondary structure, such as the intercalated T-loops in PZ19 and PZ20, accelerate 3D structure prediction using the FARFAR algorithm, leading to top models in PZ20; automated prediction of these contacts is an important area of future development for Rosetta RNA modeling methods. Motivated by these puzzles, RNA-ligand modeling is now possible (albeit cumbersome) in Rosetta. As in prior rounds of RNA-Puzzles, selection of the most accurate models from generated models remains difficult; in many cases, the most accurate Rosetta models were not the first-submitted model. The modeling experience and targets of Round IV has motivated several improvements – better automation of the FARFAR pipeline than the RW3D server, automated model selection without expert input, alternative approaches to modeling helices that allow them to flex during modeling (Watkins and Das, n.d.), more selective use of stepwise Monte Carlo – that has led to better accuracy in subsequent RNA-Puzzles.

Ding (human expert group)

The Ding group joined the prediction exercise of Puzzles 9, 17 and 19 using a similar multi-scale discrete molecular dynamics (DMD) approach as described previously (Miao et

al. 2017). Briefly, Coarse-Grained (CG) RNA simulations using replica exchange DMD were performed for each sequence, and the representative low-energy states obtained by cluster analysis were selected for all-atom reconstruction. For large RNAs with more than 60 nucleotides, consensus secondary structures obtained from secondary structure predictions (from RNAstructure (Z. Z. Xu and Mathews 2016b, [a] 2016), Rfam or SHAPE data) and bioinformatics analysis were imposed as constraints in CG simulations. Considering many of these RNAs are ribozymes and riboswitches having compact 3D structures, the Ding group noticed that the predicted structures were often not as compact as their native states and, thus, resulting in large RMSD values. Since it is well known that ions play an important role in stabilizing the RNA tertiary structures (Misra and Draper 1998), the Ding group postulate that the lack of divalent ions in their simulations might lead to suboptimal predictions and propose to evaluate whether explicit modeling of Mg2+ in our simulations will improve our predictions. Only the screened electrostatic interactions for the charged Mg2+ and phosphate groups were included in DMD simulations without considering the coordination bond interactions.

First, the Ding group performed all-atom replica exchange DMD simulations by adding different amounts of divalent magnesium ions, ranging from 1%, 5%, 10%, 15% to 20% of the total number of nucleotides. Indeed, the Ding group found that more native-like conformations with lower RMSDs were increasingly sampled with more Mg2+ included in simulations (**Fig S9a**). In order to efficiently select the native-like conformations out of a large number of snapshots sampled by replica exchange simulations, the Ding group used potential energy and radius of gyration (RG) to filter out presumably non-native conformations with high potential energies and large RG values. The Ding group also excluded similar conformations along the simulation trajectories by clustering consecutive conformations with RMSD less than 4 Å and choosing the lowest energy conformation as the representative. Clustering analysis based on pairwise RMSD of the final structure ensemble (~8,000 snapshots) was performed to select the top clusters, from where the lowest energy structures were then selected. The new results were summarized in **Table S7**. Indeed, the new results were significantly improved (*e.g.*, old and new predictions with respect to the native state of Puzzle 17 in **Fig S9b**). Magnesium ions were found to mostly bind within RNA molecules (**Fig S9c**). Hence, our results suggest that including divalent magnesium ions in all-atom DMD simulations improved structural prediction.

For Puzzle 19, the Ding group was able to sample the native-like structures with RMSD values as low as 5.4 Å, even though the probability to sample these conformations was very low (~0.05%). Potential energies for these native-like conformations were high without long-range stacking interactions as in the native structure (**Fig S9d**), suggesting that those conformations were randomly sampled. Further improvement of the all-atom DMD force field for RNAs is necessary. Additionally, the effect of incorporating divalent ions in CG simulations will be further investigated to increase the predictive power of the multi-scale modeling approach with DMD simulations.

Dokholyan (human expert group)

For Puzzles 9, 17 and 19, the Dokholyan group first used ViennaRNA (Lorenz et al. 2011) to computationally predict the secondary structures of the sequences. Base-pairing information in the secondary structure was implemented as an additional potential to promote tertiary contacts between corresponding nucleotides. Then, the Dokholyan group utilized iFoldRNA to build an initial ring-like structure and performed a short molecular dynamics simulation to obtain a preliminary structure as a seed for further refinement. Next, they performed Replica Exchange Molecular Dynamics (REMD) simulation using iFoldRNA (Sharma, Ding, and Dokholyan 2008; Krokhotin, Houlihan, and Dokholyan 2015; Feng Ding et al. 2012) to thoroughly sample the conformational space. iFoldRNA (Sharma, Ding, and Dokholyan 2008; Krokhotin, Houlihan, and Dokholyan 2015; Feng Ding et al. 2012) utilizes the Discrete Molecular Dynamics (DMD) simulation engine (Dokholyan et al. 1998; F. Ding et al. 2008; Proctor, Ding, and Dokholyan 2011; Feng Ding and Dokholyan 2005) and Medusa force field (Yin et al. 2008; Feng Ding, Yin, and Dokholyan 2010; Jian Wang and Dokholyan 2019). The REMD simulation was conducted in 8 threads with the temperatures distributed from 0.2 kT to 0.4 kT. Subsequently, the top 1% lowest-energy structures were selected from the conformation ensemble sampled by REMD and subject to clustering by RMSD with a cutoff that is approximately equal to one-tenth of the number of nucleotides for small to medium RNA (< 200 nt) and 20 Å for large RNAs (> 200 nt). Finally, the centroid structures of all clusters were subject to an all-atom reconstruction process, which replaces 3-bead nucleotide with randomly selected rotamers of corresponding nucleotides in an all-atom representation. The P, C, and O atoms in the 3-bead model were aligned with the P, C1', and N1/N3 atoms in the all-atom model through the Kabsch algorithm (Kabsch 1978). Finally, a short all-atom DMD simulation was performed to facilitate bond formations and remove clashes to obtain the final predictions.

Major (human expert group)

The Major group predicted 10 structure models, which come from three approaches, for Puzzle 17. For models 1-5, the Major group input the sequence and the following structural mask '((((********))))********((((**************(((((****)))))**))))' to MC-Fold version 1.6, which returns the minimum free energy (MFE) secondary (2D) structure shown in **Fig S10a**. The Major group submitted this 2D structure to MC-Sym (default parameters 'merge RMSD 1.5 Å', 'clash threshold 1.5 Å', 'model diversity 3 Å', 'model limit 100', and 'time limit 30 minutes') to generate 100 all-atom three-dimensional (3D) solution models. For all the models in PZ17, the runs were stopped after either a CPU time limit of 30 minutes or a model limit of 100, the first criterion that applied. They refined the models using the minimization steps available at the MC-Sym web site: 'Relieve', which stops when the force's root-mean-square is below 100 Kcal/mol/Å; 'Refine', which goes down to 5 Kcal/mol/Å. These two minimization steps are done using the Tinker molecular modeling package with a steepest-descent search on the AMBER (Junmei Wang et al. 2004) '99 force-field. The Major group then applied 'Brush-up' and 'Anneal, which run unrestrained minimization using low-temperature molecular dynamics. The five lowest energy models were submitted (Major models 1-5).

For models 6-8, the Major group submitted to MC-Sym the best free energy MC-Fold solution with a pseudoknot, which is the one that ranked 6th in free energy (**Fig S10b**). To extend the conformational sampling, they used 'merge RMSD 2.0 Å' and 'clash threshold 1.0'. The merge RMSD criterion is a threshold on the RMSD reached by the optimal superimposition of the common base-pair between two adjacent nucleotide cyclic motifs (NCM) (see the original paper on MC-Fold (Parisien and Major 2008) for the definition of NCM and to see how the merge RMSD criterion works). The Major group applied the same minimization protocol as above to the three all-atom resulting models (time limit of 30 minutes reached) and resulted in models 6-8.

For models 9-10, the Major group used the same 2D structure as for models 6-8 (**Fig S10b**), but relaxed further: 'merge RMSD 3.0 Å'; 'clash threshold 0.5'. This generated 25 models. After applying the same minimization protocol as above, two lowest energy models were submitted (models 9-10).

The RMSD of the Major models improves with increasing the quality of the input 2D structure. This is not a surprise since the goal of the pipeline is to produce 3D models that correspond to 2D structure predictions. The MFE 2D structure used satisfies about 60% of the experimental target (MCC = 0.63). The best RMSD is model 1 with 16.8 Å to the experimental target (INF_{all} = 0.57). The second 2D input structure used, the one with the pseudoknot, satisfies about 70% of the experimental target (MCC = 0.70). The best RMSD is model 10 with 13.5 Å to the experimental target (INF_{all} = 0.32).

The models are improved by increasing the flexibility of the 3D conformational sampling by MC-Sym. This allows for more models, which can, however, deviate from the input 2D structure, to be generated. Indeed, derogating from an incorrect 2D structure gives freedom to move closer to the 3D of a different 2D structure. This is a feature and, perhaps, a drawback of the MC-Fold and MC-Sym pipeline, whose goal is to produce 3D models that respect strictly the interactions present in the input 2D structure. If the Major team is to perform better in the puzzles, it needs to improve its 2D structure prediction algorithm, which is currently a difficult task to accomplish from the sequence. In RNA-Puzzles, many factors that are not present in the sequence information need to be considered.

Xiao (human expert group) and 3dRNA (webserver)

First, Mfold (Zuker 2003), CentroidFold (Sato et al. 2009) and ContextFold (Zakov et al. 2011) were used to predict candidate secondary structures of the given sequence. The disadvantage of these programs is that they cannot automatically predict possible pseudoknots. Therefore, PKiss (Janssen and Giegerich 2010) and an in-house RNA secondary structure prediction method were used to predict the secondary structure with pseudoknots. The secondary structure prediction results were compared. Then, the centroid structure (the structure with the greatest similarity to the other structures) and an additional secondary structure were selected based on human expert experience. For these two secondary structures, 3dRNA (Jian Wang and Xiao 2017; Zhao et al. 2012; Jian Wang et al. 2017) was used to predict the tertiary structures. Secondary structure was decomposed into secondary structure elements (SSEs) (Zhao et al. 2012). Then for each SSE, the most

suitable template was searched from the tertiary structure template library (Zhao et al. 2012). If no template was found, the distance geometry algorithm (Mucherino et al. 2012) was used to build a 3D template *ab initio*. These templates were then assembled together and subjected to optimization by the recently developed tool (Zhao et al. 2012) with further energy minimization using AMBER (Junmei Wang et al. 2004) Amber99SB force field (Salomon-Ferrer, Case, and Walker 2013). For each RNA secondary structure, 10 structures were predicted and 3dRNAscore (Jian Wang et al. 2015) was applied to select 5 best structures from the 20 predicted models as the final predictions.

The predictions of 3dRNA in PZ15 ranked poorly (PZ15: 55/64). It was discovered that the bad prediction ability was caused by the lack of necessary 3D templates, which is the common limitation of all fragment-based methods. To solve this problem, the Xiao group developed a Monte Carlo simulation-based program (Jian Wang et al. 2017) to optimize the RNA 3D structure after the initial assembly step. Notably, this optimization improves 3dRNA. According to **Fig S11,** PZ17 and PZ19, which were predicted with the optimization, shows better RMSD ranking than PZ9. Additionally, model 3 and model 2 of 3dRNA in PZ19 demonstrate top ranking in INF_{nwc} and $INF_{stacking}$, respectively.

Although the optimization program brings notable improvement to the prediction accuracy of 3dRNA, the predicted models still greatly deviate the reference structure. It is possible due to the inaccuracy of the energy function used by the optimization program. The optimization program is essentially a *de novo* method.

Supplementary Tables

Table S1. The assessment results of Puzzle 9.

Table S2. The assessment results of Puzzle 15.

Table S3. The assessment results of Puzzle 17.

Table S4. The assessment results of Puzzle 19.

Table S5. The assessment results of Puzzle 20.

Table S6. The assessment results of Puzzle 21.

Table S7. RMSD values before and after reconstruction for PZ9, PZ17, PZ19 (from low to high), and RMSD values for PZ20 and PZ21.

Table S8. The MAXIT report of Puzzle 9.

Table S9. The MAXIT report of Puzzle 15.

Table S10. The MAXIT report of Puzzle 17.

Table S11. The MAXIT report of Puzzle 19.

Table S12. The MAXIT report of Puzzle 20.

Table S13. The MAXIT report of Puzzle 21.

Table S14. The assessment results of Puzzle 15 without the six base pairs between the 5’ and 3’ ends.

Supplementary Figures

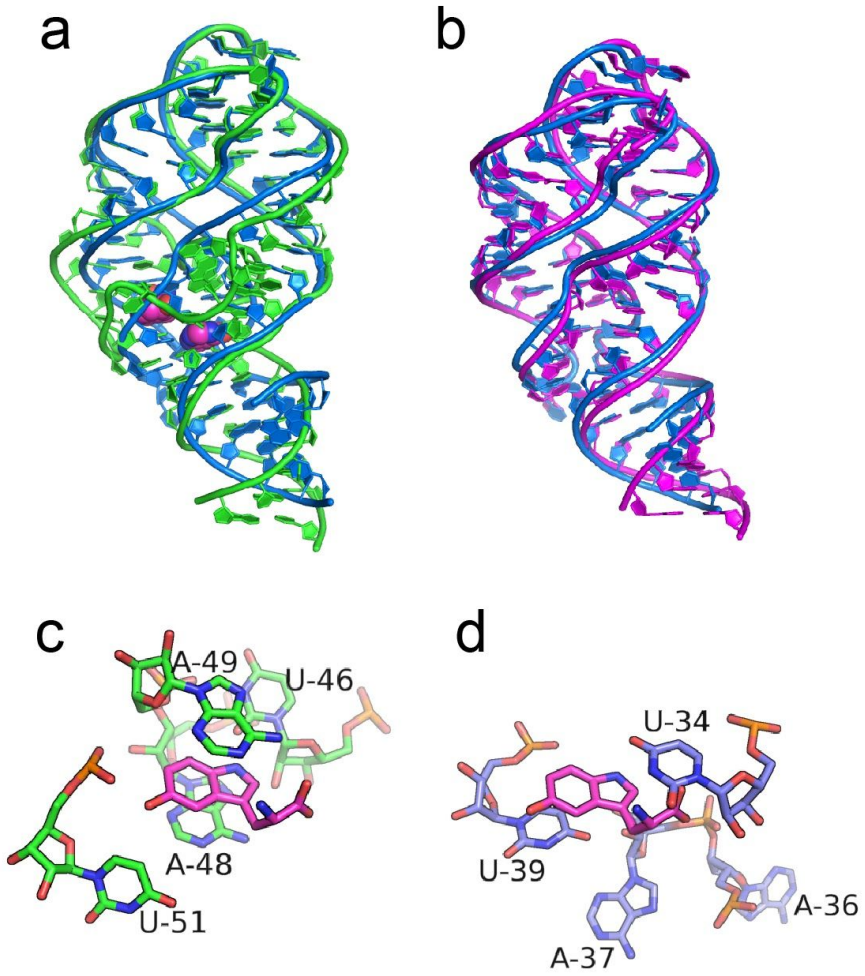

Figure S1. Structural superpositions between the guanine and 5-hydroxytryptophan riboswitches

a) The superposition between the 5HTP riboswitch (green, PDB id 5kpy) and the guanine riboswitch (blue, PDB id 4fe5).
b) The superposition between the guanine riboswitch (blue, PDB id 4fe5) and the best predicted model (magenta) from Chen group model 7.
c) The 5-hydroxy-L-tryptophan (magenta) and its binding pocket (green) in the 5HTP riboswitch,
d) The 5-hydroxy-L-tryptophan (magenta) and the nucleotides aligned to the binding pocket (purple) in Chen group model 7.

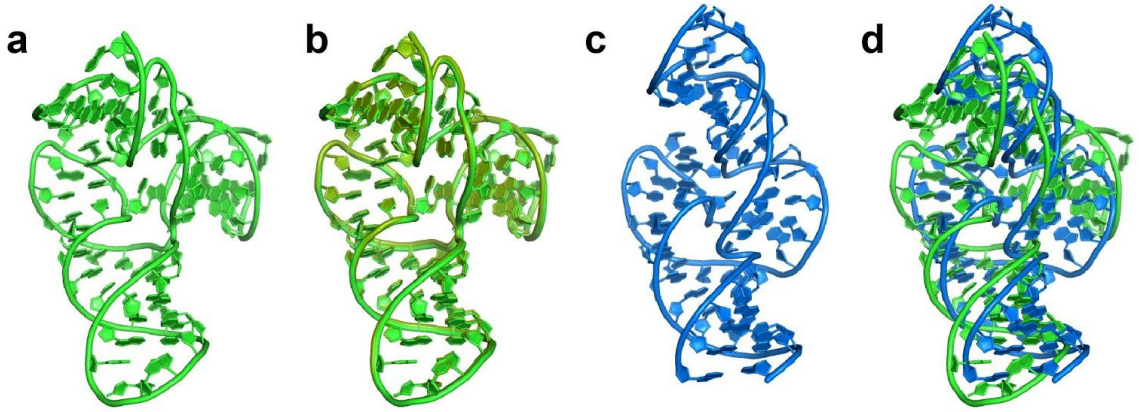

Figure S2. Structural comparisons between the twister-sister ribozymes.

a) The twister-sister ribozyme (PDB 5y85).
b) Superposition between two variant structures solved (PDBs 5y85 and 5y87).
c) The twister-sister ribozyme (PDB 5t5a).
d) Superposition between two other structures (PDBs 5y85 and 5t5a).

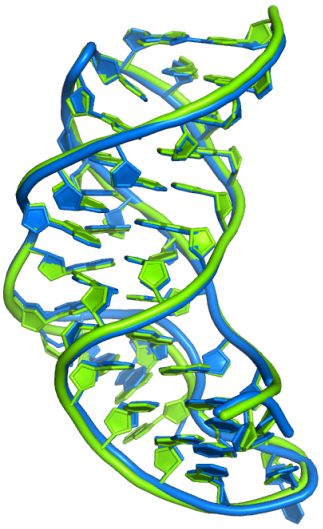

Figure S3. Superposition between the two structures solved in Puzzle 21.

5nwq is shown in green while 5nz6 is in blue.

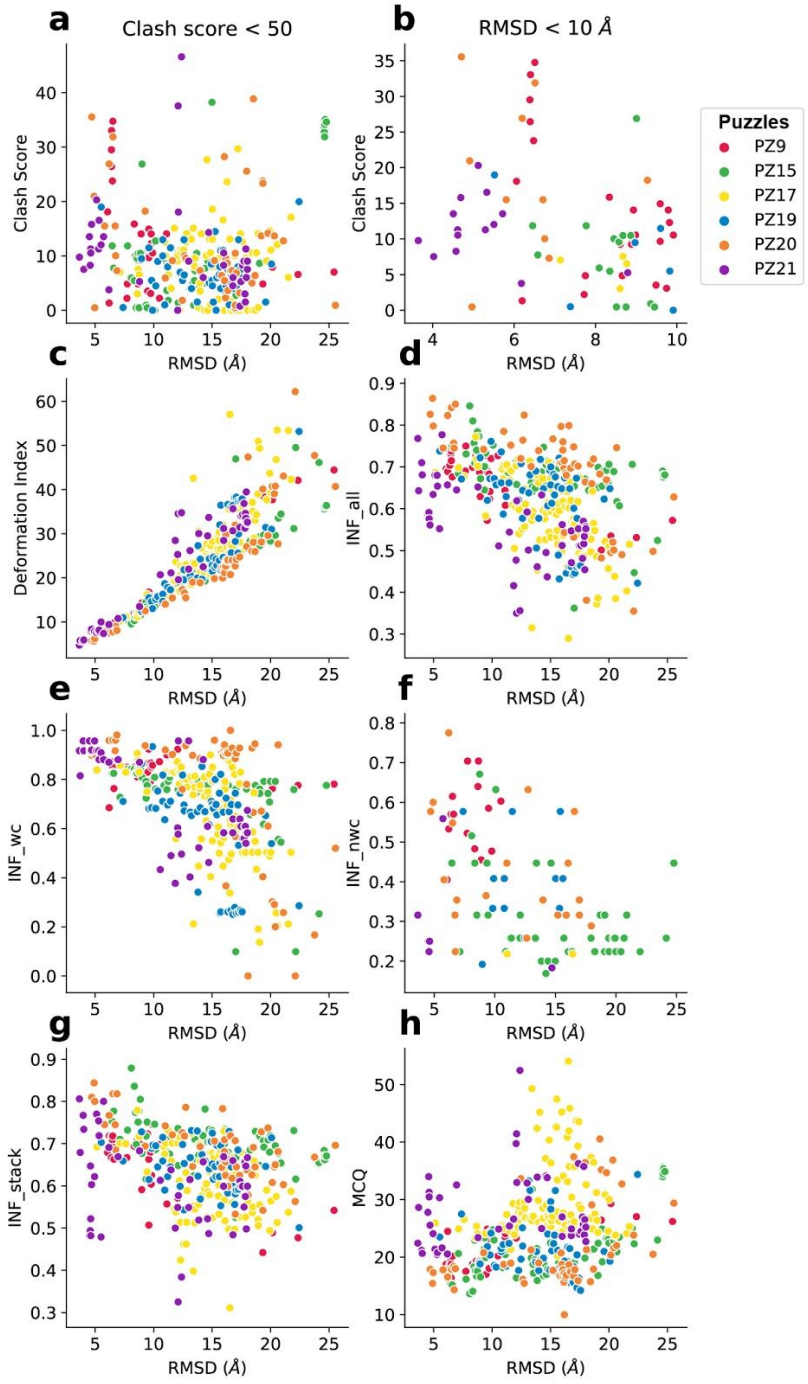

Figure S4. Metrics distribution compared with RMSD.

The clash score(**a** and **b, b** is a zoom-in at the region of RMSD <10Å), Deformation Index(**c**), Interaction Network fidelity (**d**, **e**, **f**, **g**, corresponding to all types of interactions, Watson-Crick interactions, non-Watson-Crick interactions, Stacking) and Mean of Circular Quantities(**h**) are plotted against the RMSD distribution.

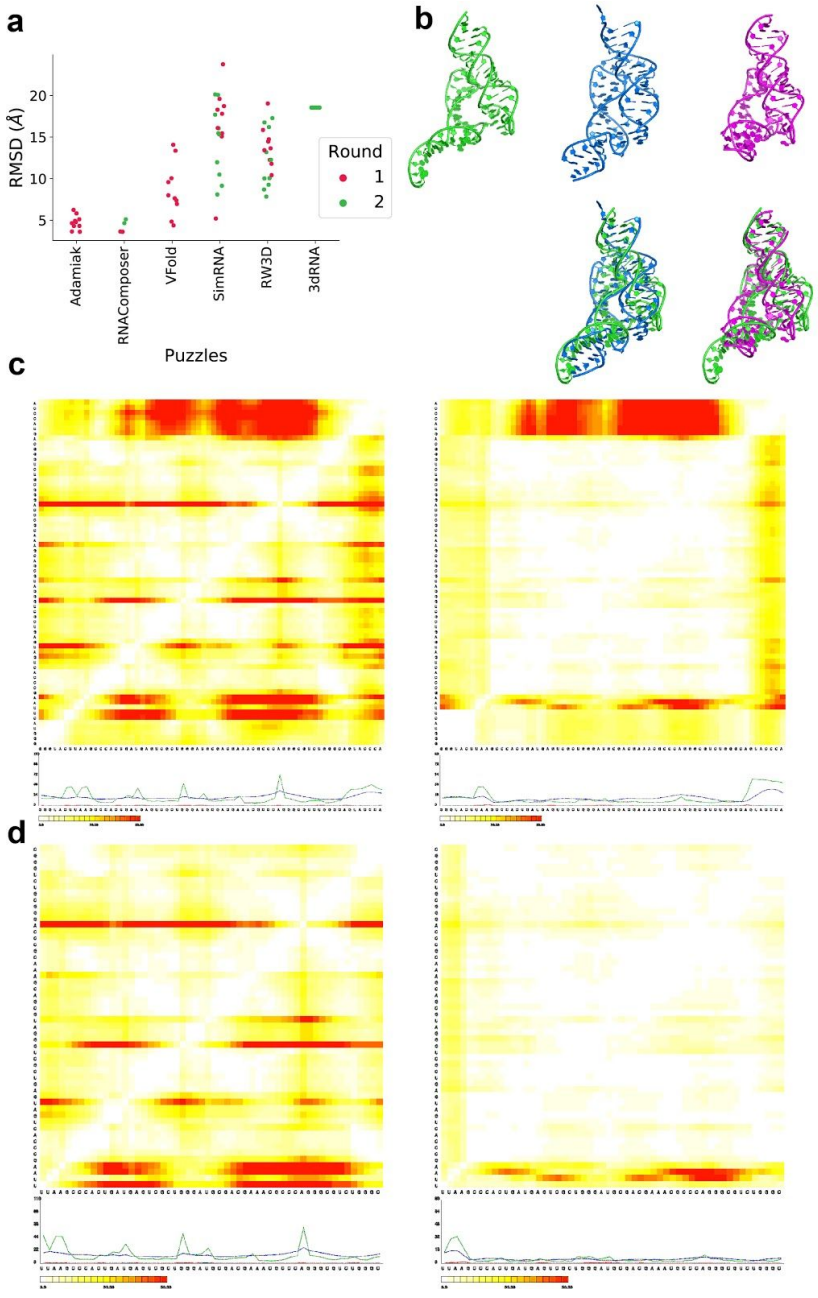

Figure S5. Structural comparisons for the prediction of the hammerhead ribozyme.

a) Shows the RMSD distributions for the two rounds of prediction.
b) Shows the structure comparison between the reference structure (green) and the two structures with the lowest RMSDs, the SimRNA (blue) and RNAComposer (magenta).
c) Shows the deformation profiles for SimRNA (left) and RNAComposer (right) as heatmaps.
d) The heatmaps show the deformation profiles of SimRNA (left) and RNAComposer (right) excluding the 5'- and 3'-ends.

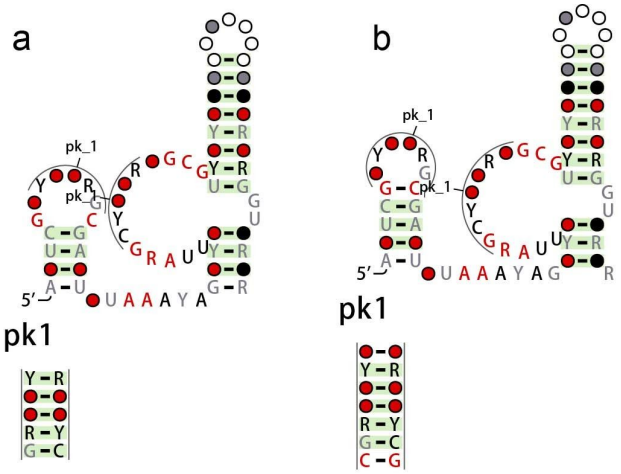

Figure S6. Comparison between the Pistol ribozyme secondary structures from Rfam and deduced from R-scape.

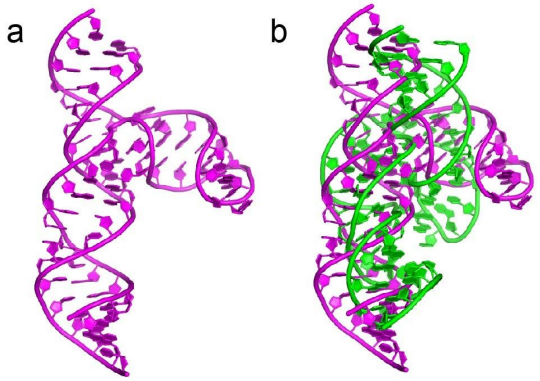

Figure S7. The prediction of Adamiak model 3 in Puzzle 19 twister sister ribozyme 1.

a) The Adamiak model 3 (magenta);
b) The superposition between Adamiak model 3 and reference structure (green).

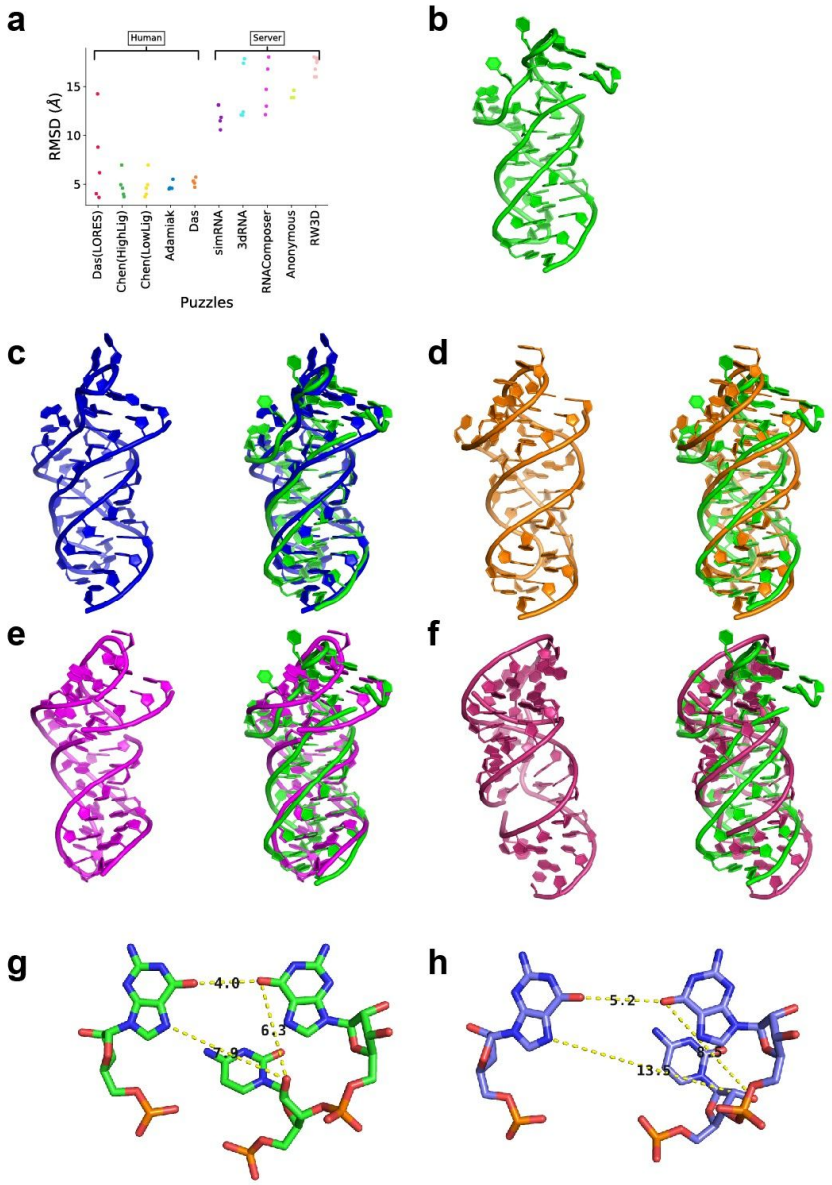

Figure S8. Structure comparisons for Puzzle 21 models (guanidine riboswitch).

a) The RMSD distribution comparison between human expert prediction and automated web servers.
b) The reference structure (green).
c) The structure predicted by LORES (from Das group) model 1 is shown in blue.
d) The structure predicted by Adamiak group model 4 is colored in orange.
e) The structure predicted by the Chen group (high ligand concentration model 2) is shown in magenta.
f) The model 4 predicted by SimRNA is shown in warm pink.
g) The structure of the guanidine binding core (C6, G7, and G17) in the reference structure.
h) The same contacts predicted in the best RMSD model (LORES model 1).

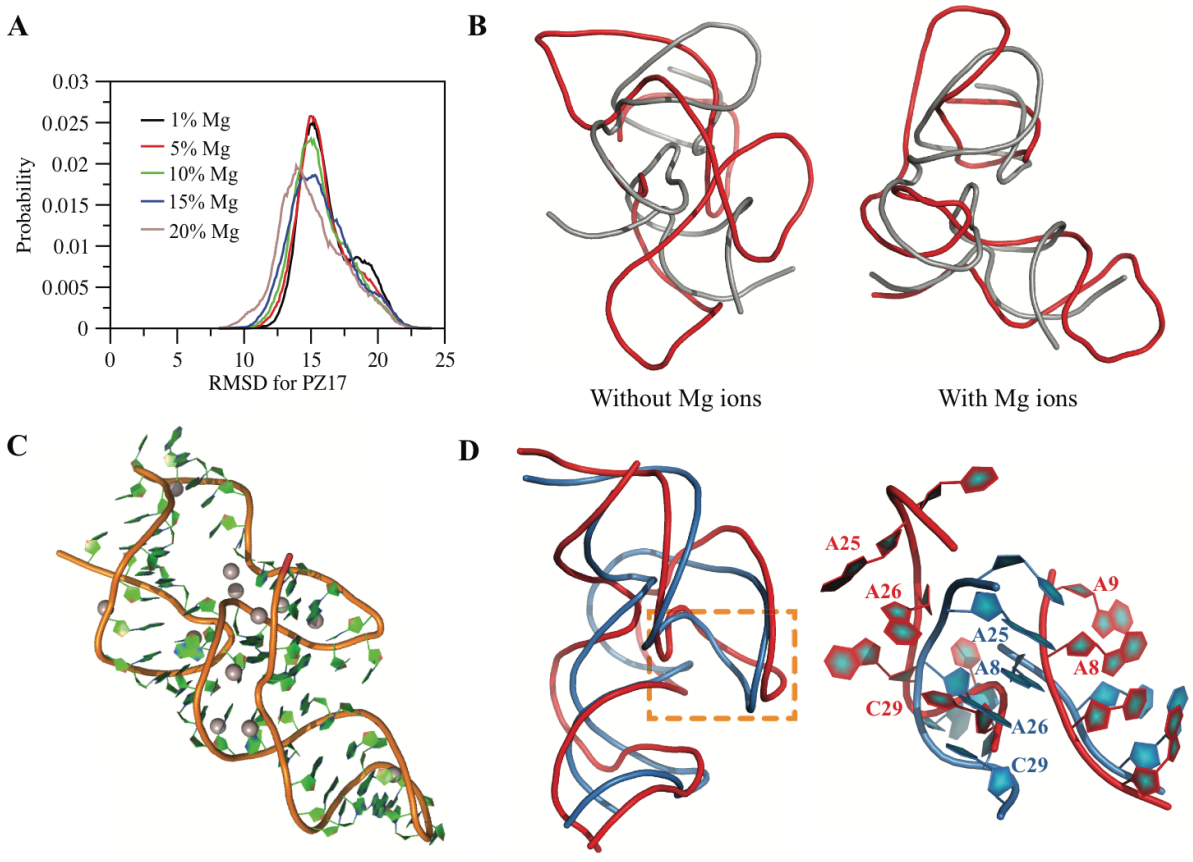

Figure S9. Effects of Mg2+ in all-atom DMD simulations (from Ding lab).

(A) The probability distributions of RMSD values in all-atom DMD simulations of Puzzle 17 after adding 1%, 5%, 10%, 15%, 20% magnesium ions with respect to the RNA length. (B) Top predicted structures before (left, RMSD=16.9 Å) and after (right, RMSD=9.2 Å) adding 20% magnesium ions. The backbone trace of predictions and native states are colored red and grey, respectively. (C) Magnesium ions (gray spheres) were found to bind with the RNA as expected. (D) Left is alignment between the native-like structure (red backbone trace) and standard structure (blue backbone trace) of PZ19. The tertiary stacking interaction structure was highlighted in orange dashed lines and zoomed in to the right. Involved nucleotides were marked out in red for the predicted result and blue for standard.

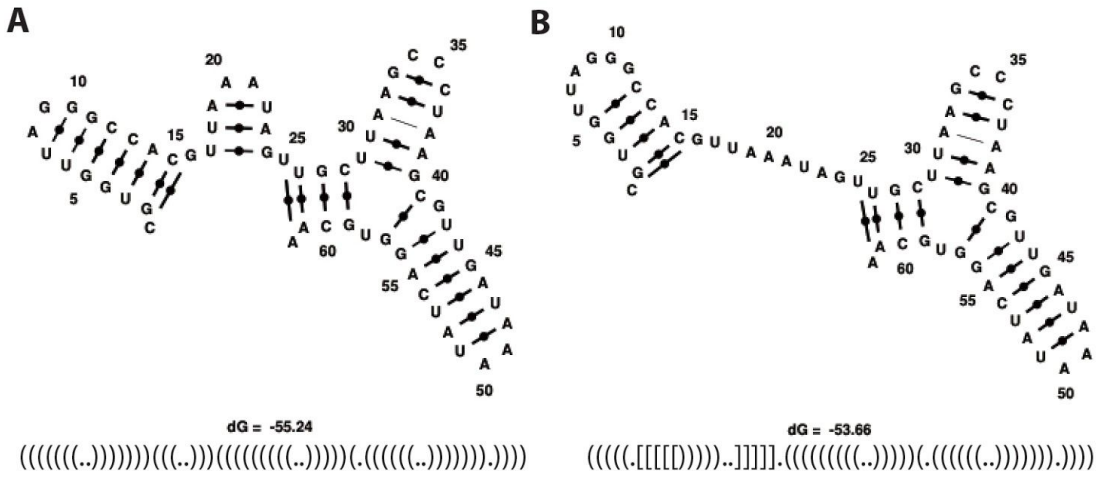

Figure S10. Secondary structures used in the prediction for the Major models in Puzzle 17.

A) the minimum free energy structure without pseudoknot. B) the best free energy structure with pseudoknot.

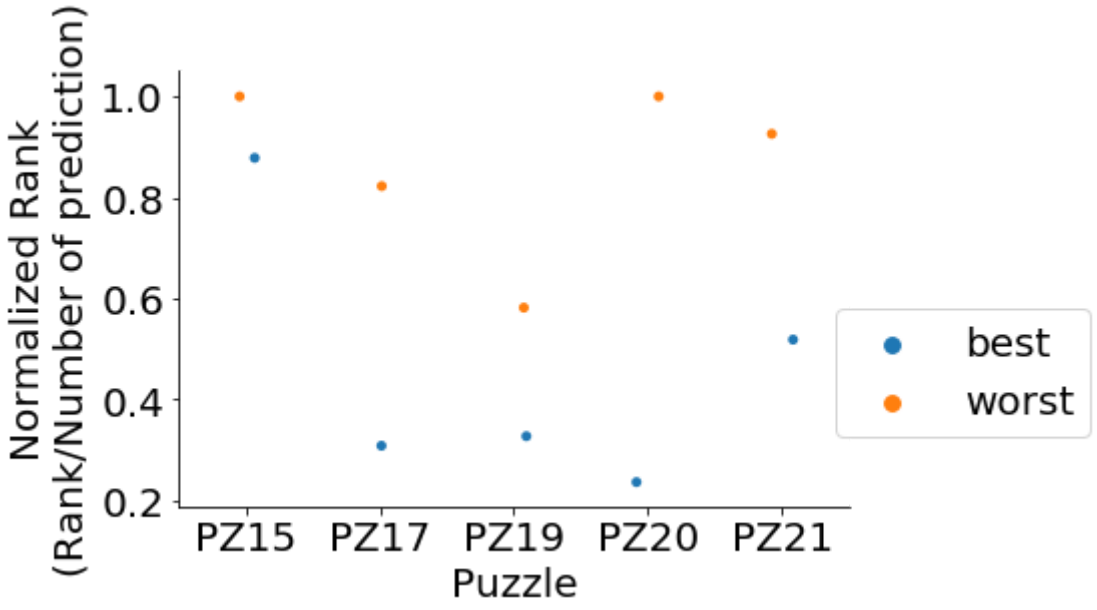

Figure S11. Normalized rank of 3dRNA in PZ15, PZ17, PZ18 and PZ19.

The y-axis is the normalized rank calculated by dividing the rank of the model by the total number of submitted models. The best and worst ranked models are shown in blue and orange, respectively.

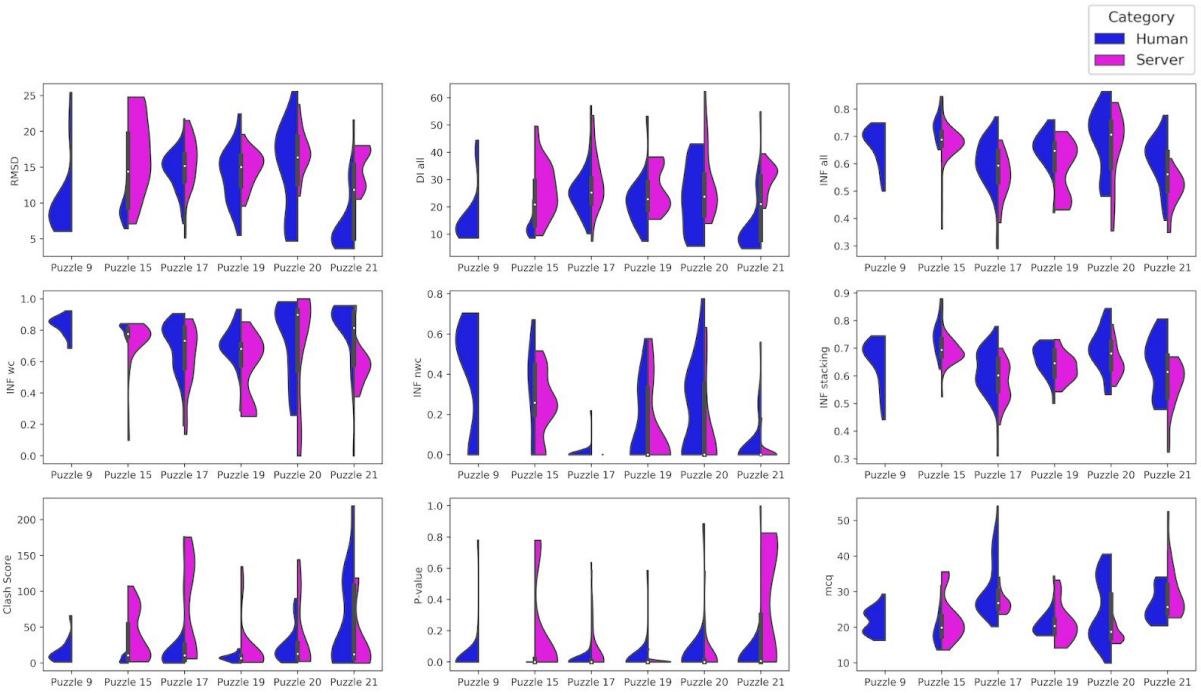

Figure S12. Comparison between human experts prediction and automatic server prediction.

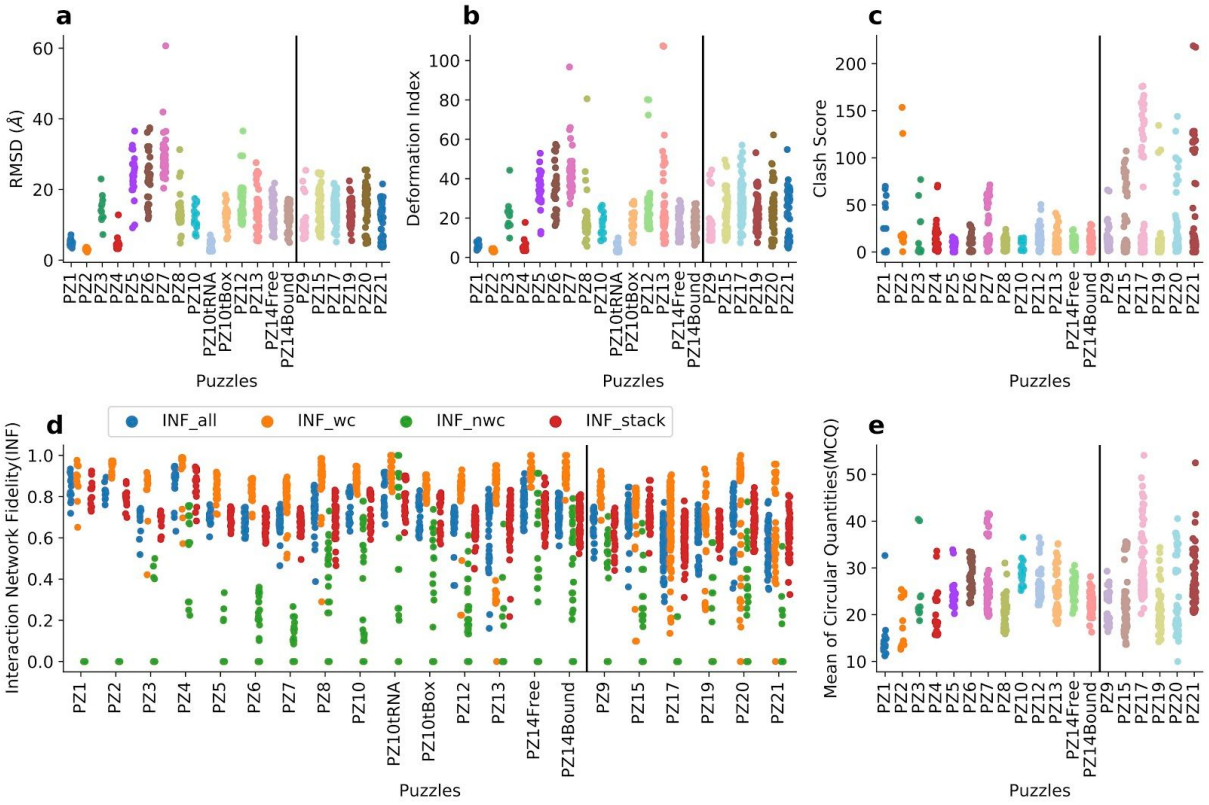

Figure S13. The prediction metrics distribution in comparison with puzzles reported before.

This plot demonstrates the metrics of (a) RMSD, (b) Deformation Index, (c) clash score, (d) Interaction network fidelity (total, Watson-Crick, non-Watson-Crick, stacking) and Mean of Circular Quantities(MCQ) (e).

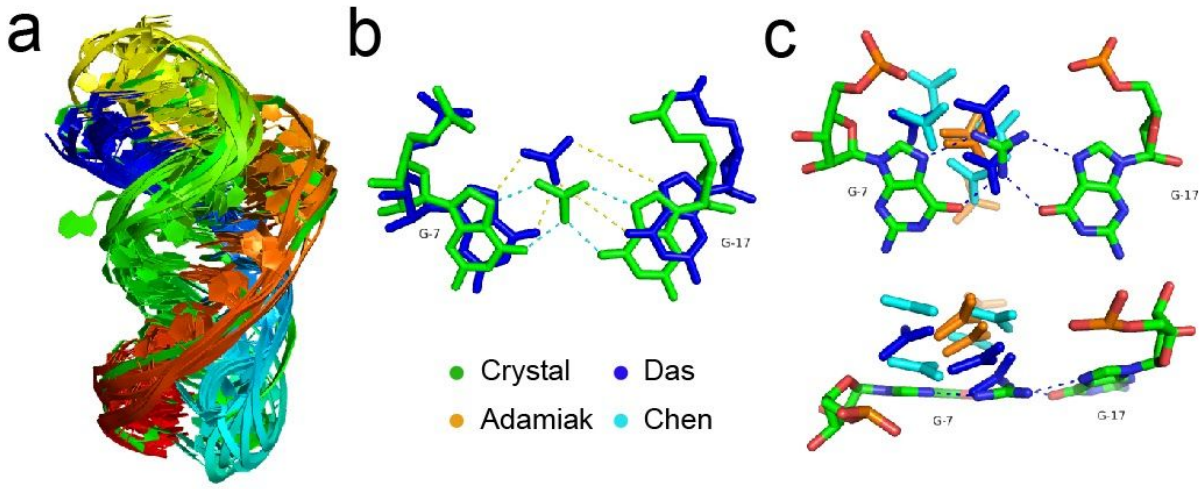

Figure S14. The prediction of guanidine ligand position in guanidine-III riboswitch aptamer.

a) The predicted structures within 5 Å RMSD from the crystal structure, while the crystal structure is shown as green;
b) The ligand-binding position compared between crystal structure (green) and LORES model 1 from Das group (blue);
c) All the predicted ligand positions compared with the crystal structure: Das models (blue), Adamiak (orange) and Chen (cyan).

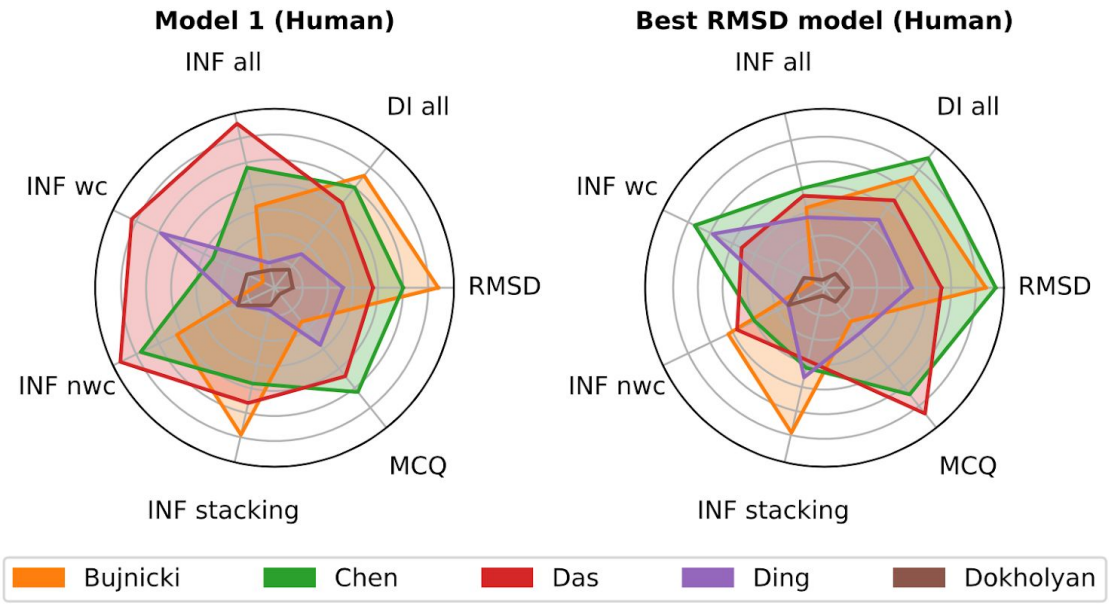

Figure S15. A radar plot of model ranking in Puzzle 9.

This plot shows the radar plot of the ranking of model 1 or the best predicted model for all the assessment metrics. Each assessment metric is normalized to the rank number between 1 to N, where N is the total number of submitted predictions. And a larger number indicates more similar to the reference structure. This also applies to **Figures S16-S20**.

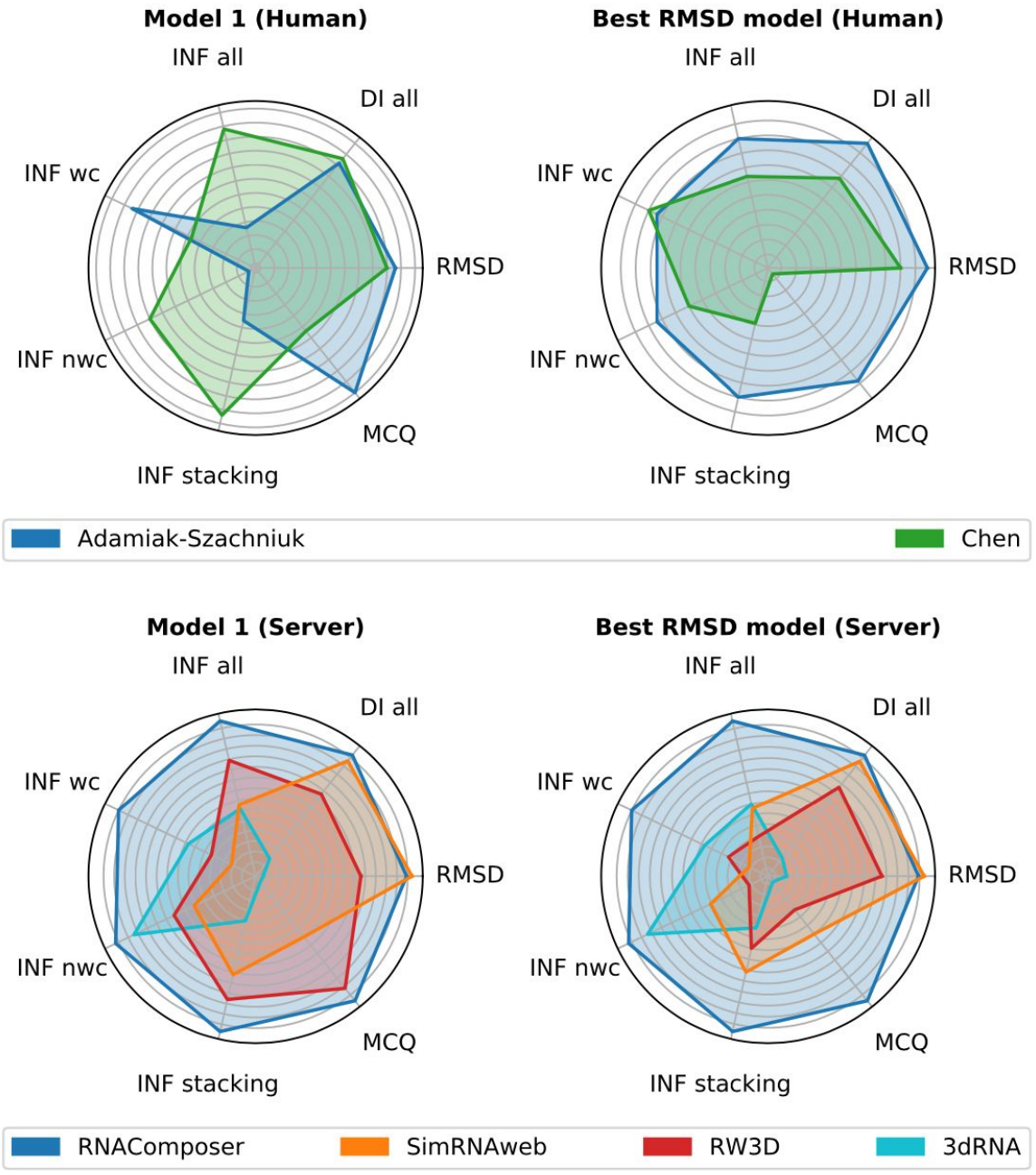

Figure S16. A radar plot of model ranking in Puzzle 15.

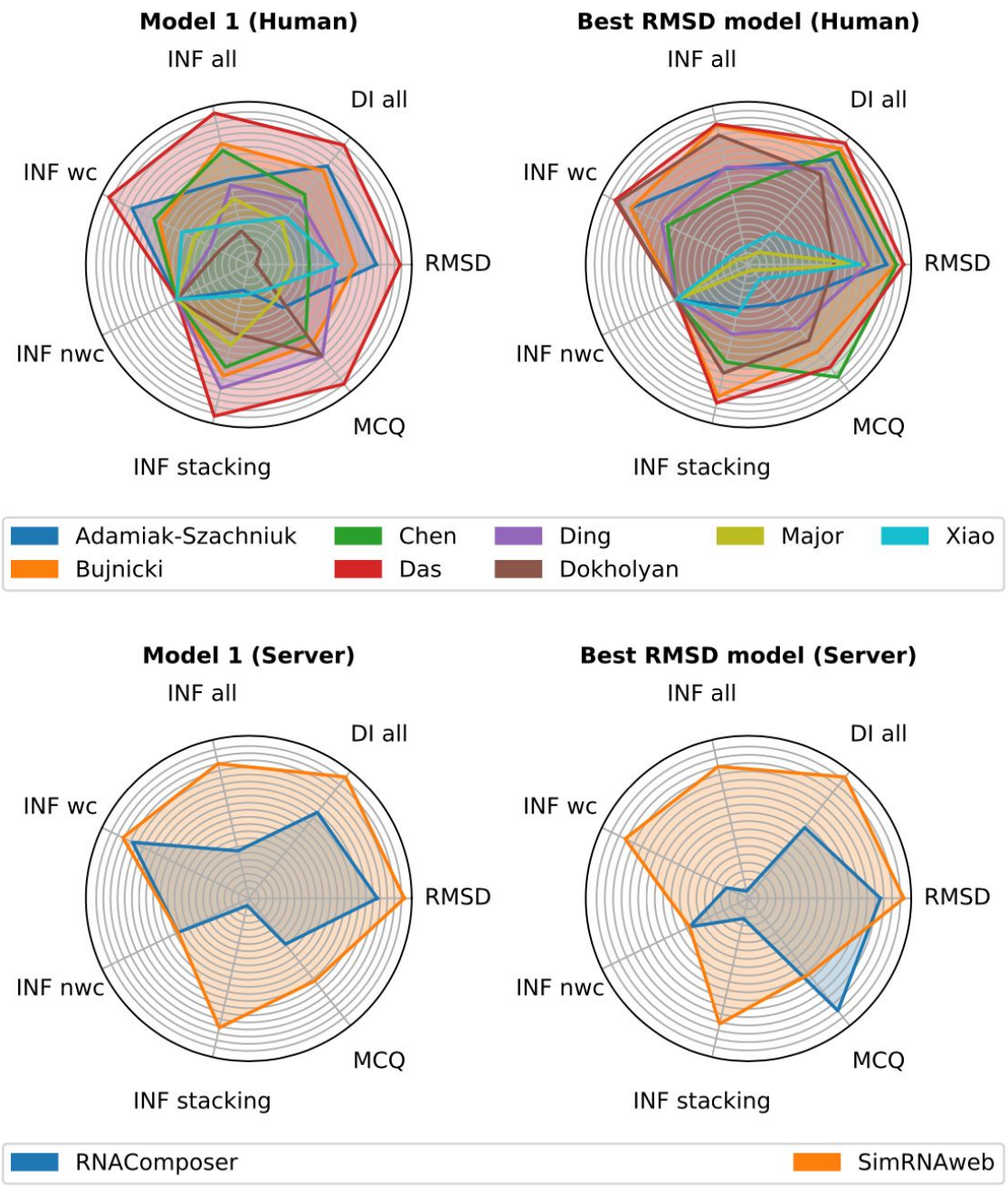

Figure S17. A radar plot of model ranking in Puzzle 17.

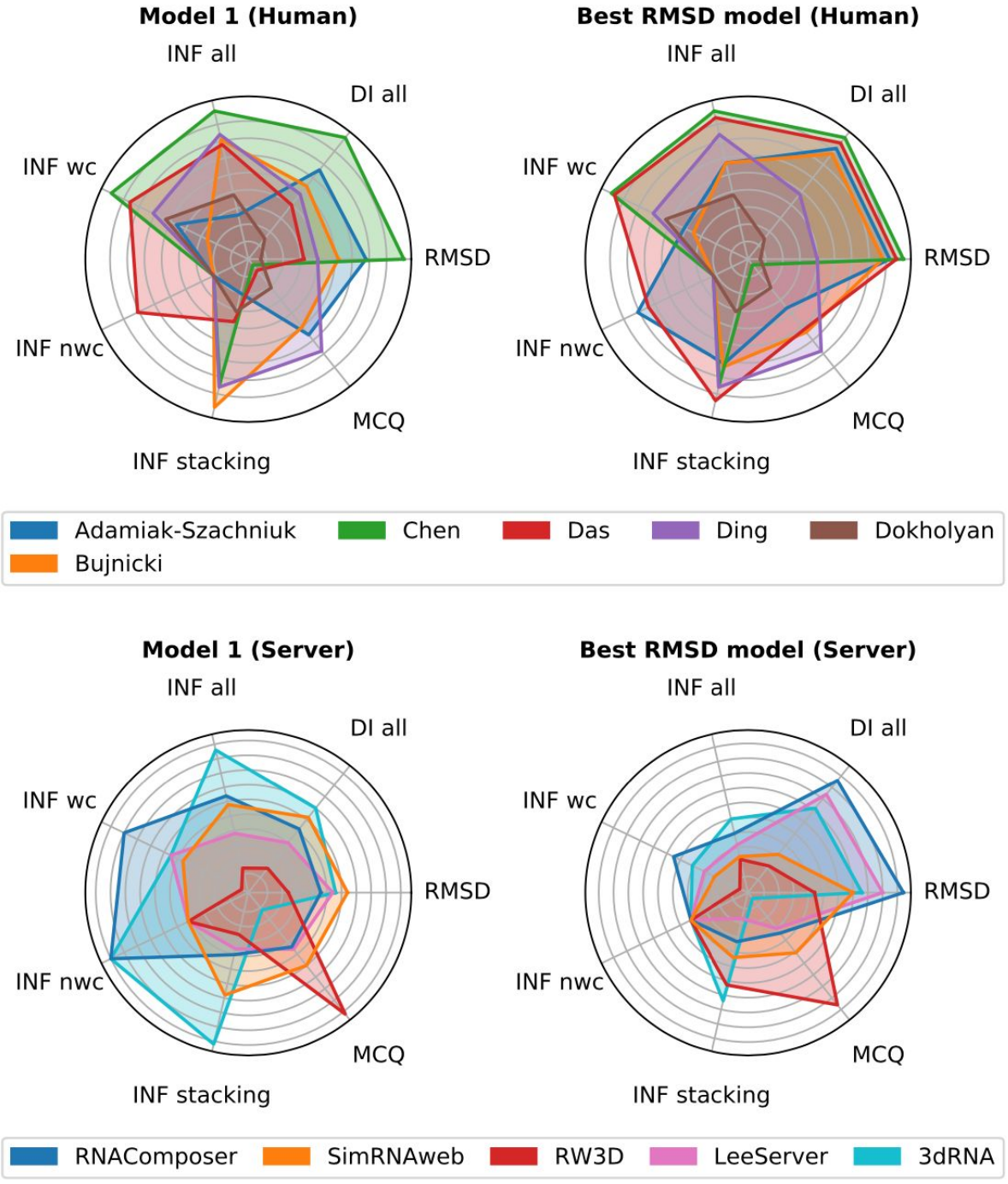

Figure S18. A radar plot of model ranking in Puzzle 19.

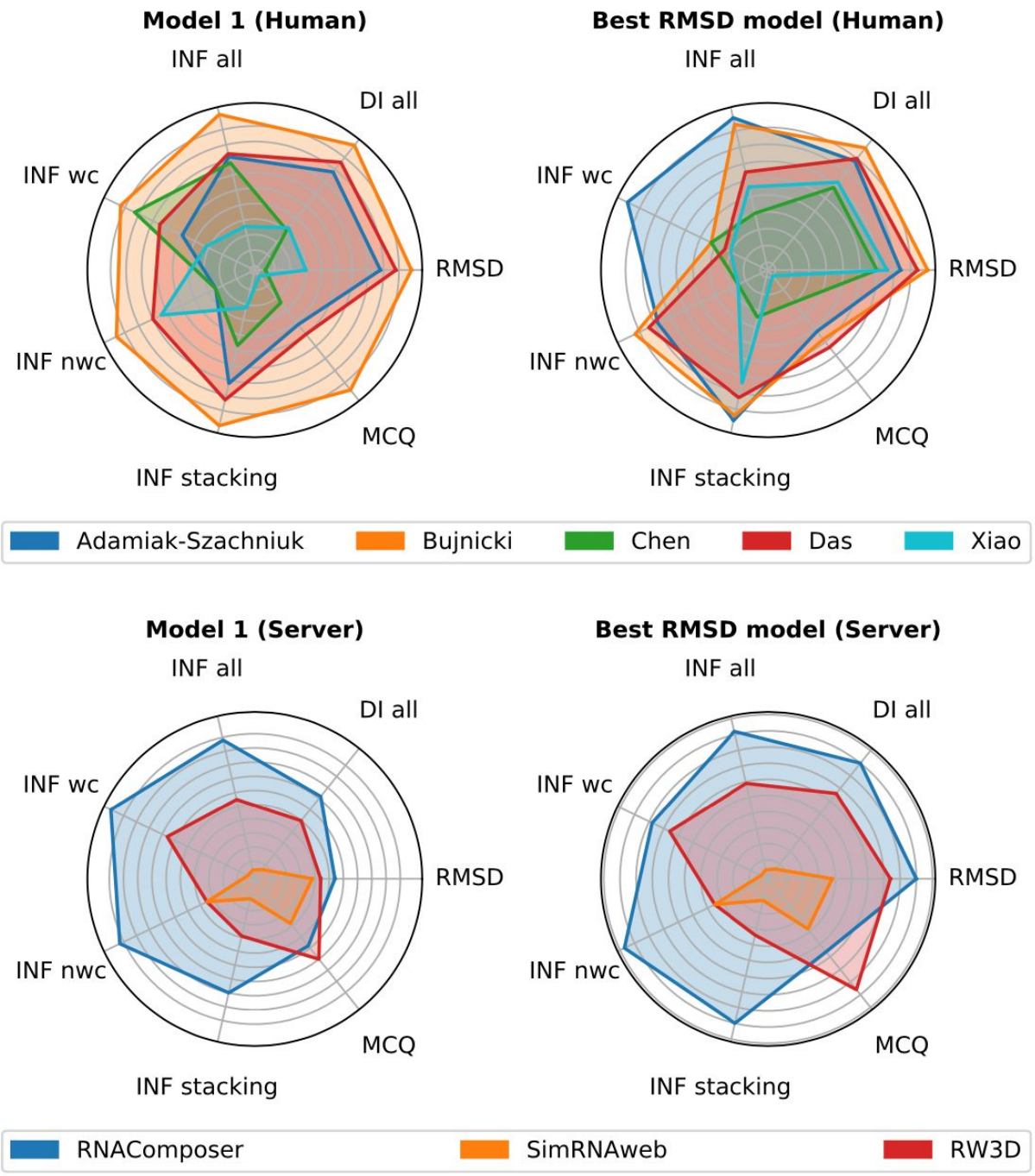

Figure S19. A radar plot of model ranking in Puzzle 20.

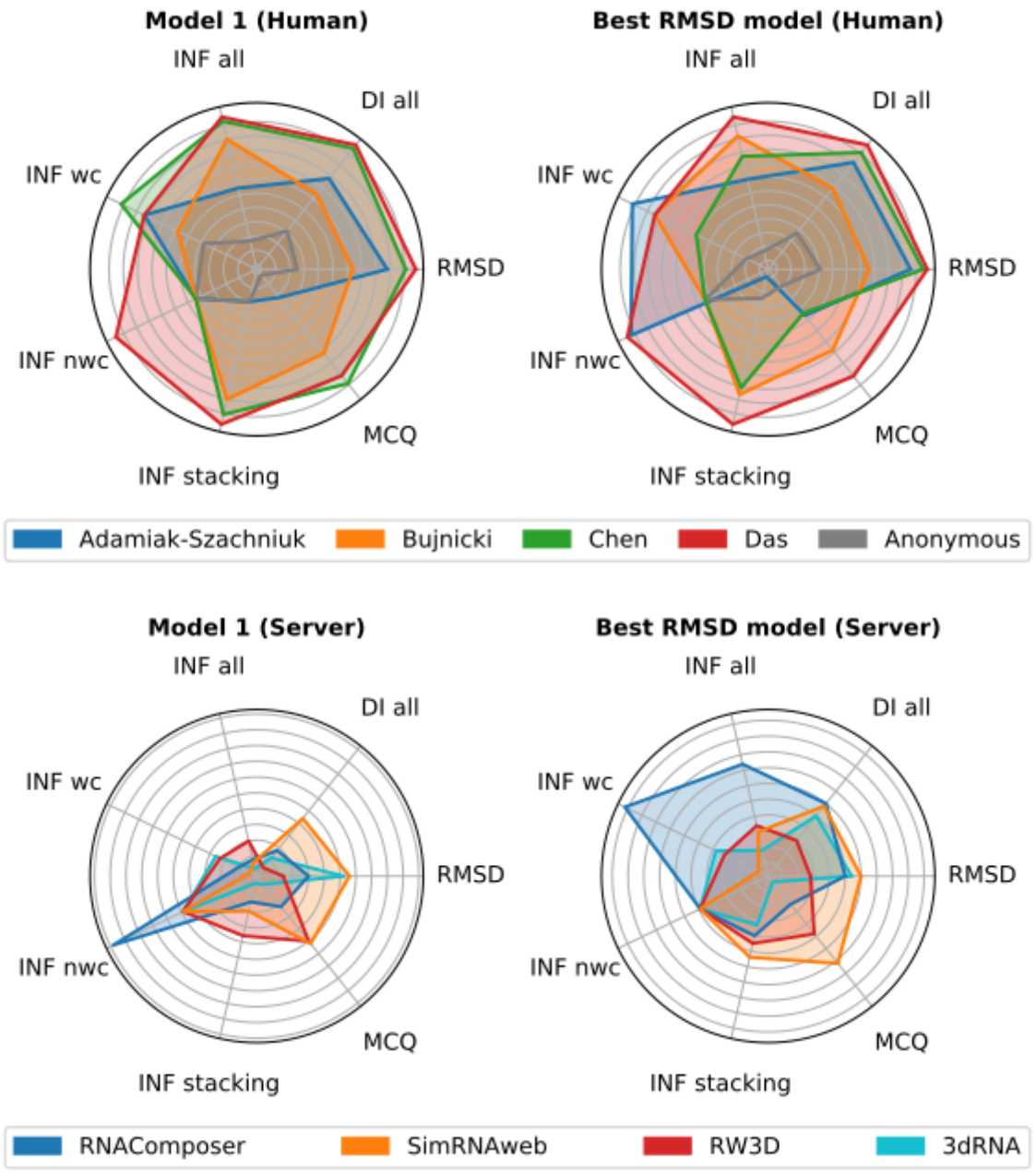

Figure S20. A radar plot of model ranking in Puzzle 21.

References

Parisien, Marc, José Almeida Cruz, Eric Westhof, and François Major. 2009. “New Metrics for Comparing and Assessing Discrepancies between RNA 3D Structures and Models.” *RNA* 15 (10): 1875–85.

Zok, Tomasz, Mariusz Popenda, and Marta Szachniuk. 2014. “MCQ4Structures to Compute Similarity of Molecule Structures.” *Central European Journal of Operations Research*. https://doi.org/10.1007/s10100-013-0296-5.