Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Phys Chem B. Author manuscript; available in PMC 2022 December 08.

Published in final edited form as:

J Phys Chem B. 2021 November 04; 125(43): 11907–11915. doi:10.1021/acs.jpcb.1c07288.

Modeling noncanonical RNA base pairs by coarse-grained IsRNA2 model

Dong Zhang[a], **Shi-Jie Chen**[b,*], **Ruhong Zhou**[a,*]

[a]College of Life Sciences and Institute of Quantitative Biology, Zhejiang University, Hangzhou 310058, China

[b]Department of Physics, Department of Biochemistry, and Institute of Data Science and Informatics, University of Missouri, Columbia, Missouri, 65211, USA

Abstract

Noncanonical base pairs contribute crucially to the three-dimensional architecture of large RNA molecules; however, how to accurate model them remains an open challenge in RNA 3D structure prediction. Here we reported a promising coarse-grained IsRNA2 model to predict noncanonical base pairs in large RNAs through molecular dynamics simulations. By introducing a five-bead per nucleotide coarse-grained representation to reserve the three interacting edges of nucleobases, IsRNA2 accurately models various base pairing interactions, including both canonical and noncanonical base pairs. A benchmark test indicated that IsRNA2 achieves a comparable performance to the atomic model in *de novo* modeling of noncanonical RNA structures. In addition, IsRNA2 was able to refine the 3D structure predictions for large RNAs in RNA-Puzzles challenges. Finally, the graphics processing unit (GPU) acceleration was introduced to speed up the sampling efficiency in IsRNA2 for very large RNA molecules. Therefore, the coarse-grained IsRNA2 model reported here offers a reliable approach to predict structures and dynamics of large RNAs.

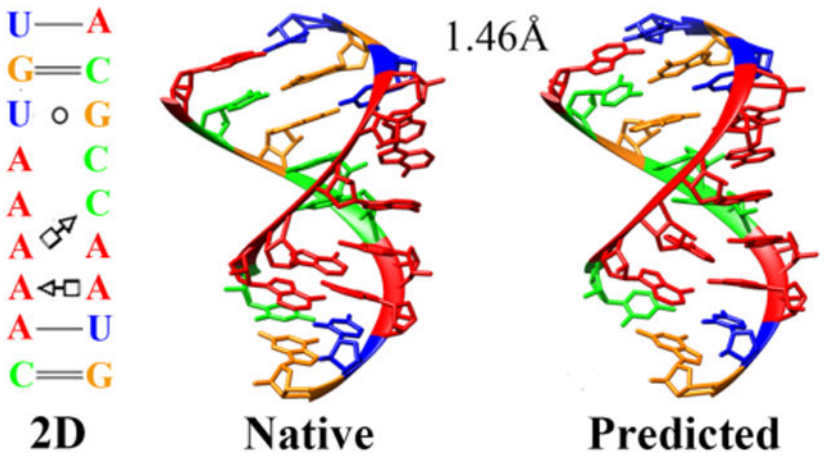
Graphical Abstract

*To whom correspondence should be addressed: rhzhou@zju.edu.cn (R.Z.) ; chenshi@missouri.edu (S.C.).

Supporting Information

The formulas for bond stretching energy $E_{bond}(b)$, bond angle bending energy $E_{angle}(\theta)$, and torsion angle energy $E_{torsion}(\varphi)$; all the relevant energy parameters for the energy functions in IsRNA2 model.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Introduction

The accurate determination of the three-dimensional (3D) structure of RNA molecules is crucial for a better understanding of their various biological functions[1-3], including carrying genetic information, regulating gene expression, performing enzymatic activity and so on. However, experimental determination of the atomic structures of RNA molecules through X-ray crystallography, NMR, or cryo-electron microscopy remains a very challenging task, requiring a substantial amount of time and technical resources. Thus, there is a huge gap between the number of known RNA sequences[4,5] and the number of atomic RNA 3D structures available in the Protein Data Bank (PDB)[6]. As a result, various computational approaches have been developed to complement experiments for RNA 3D structure determination[7-10].

Similar to computational protein structure prediction, different strategies have been adopted to predict the RNA 3D structures[9], such as homologous modeling[11], templated-based approach[12-15], fragment assembly[16-18], and physics-based methodology[19-23]. Among them, the coarse-grained (CG) model is particularly attractive for RNA structure prediction and folding simulations[24], as it can largely improve the efficiency of conformational space sampling, especially for large RNA molecules. In recent years, different CG models of various representations of the nucleotides have been developed to predict RNA 3D structures and to study RNA folding behaviors, with or without the aid of experimental data[9,24]. For instance, through simplifying the nucleotides into three beads, iFoldRNA[19] and its variations[25-26] used discrete molecular dynamics (MD) simulations to predict 3D structures for small to medium-sized RNA molecules, with the aid of hydroxyl radical probing data[25] and sparse NMR constraints[26]. The multilevel representation CG model, SimRNA[23], employed a statistical potential and Monte Carlo method to study the structural and dynamical properties of RNAs up to 190 nucleotides (nts), with the secondary structure and/or additional long-range contact information combined in particular. More recently, we developed an iterative simulated reference state approach to model correlated interactions in RNA folding (IsRNA) and to accurately parameterize the energy functions in the CG model[27]. Subsequently, through MD simulations, a large-scale benchmark test on RNA 3D structure prediction indicated that the updated IsRNA1 (version 1; as compared to version 0

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

IsRNA) model can provide improved performance for relatively large RNAs of complicated topologies, such as large stem-loop structures and structures containing long-range tertiary interactions[28]. Moreover, combined with experimental data, the IsRNA/IsRNA1 model was able to elucidate the folding pathway of an RNA pseudoknot[29], to model the loop composition effect in RNA folding stability[30], and to characterize binding features of an RNA aptamer to its targeted protein[31].

As demonstrated by the RNA-Puzzles[32-35], a Critical Assessment of Protein Structure Prediction (CASP)[36]-like collective evaluation of RNA 3D structure predictions, an accurate description of noncanonical base pairing interactions constitutes the important bottleneck in RNA 3D structure modeling and still remains as an open challenge. Though many methods have enabled predictions for Watson-Crick (WC) base pairs and native-like global folds, the true positive rate for the prediction of noncanonical base pairs is only ~20% or lower, aside from previously solved templates that happen to recur in new challenge[34-35]. Compared to WC base pairing interactions, the noncanonical base pairs are more variable and have abundant covariation rules as they also involve the base's sugar and Hoogsteen edges[37-38]. These interactions have caused significant challenge for *ab initio* predictions of noncanonical base pairing interactions, even with the additional constraints from chemical probing data[34,39]. However, noncanonical base pairs contribute crucially to RNA 3D structures and are central to the 3D architecture of folded RNA molecules[40]. Without realization of these noncanonical interactions, it is hard to explain evolutionary data, difficult to predict molecular partners, or almost impossible to be prospectively tested by compensatory mutagenesis *via* RNA computational modeling. In some pioneering works, Das and coworkers have proposed an all-atom refinement based on atomic energy functions[41] (FARFAR) and enhanced conformational sampling methods, including enumerative stepwise assembly[42] and stepwise Monte Carlo method with a unique add-and-delete move set[43], to predict noncanonical RNA motifs at atomic resolution. Recently, deep learning-based methods have been developed to better describe the noncanonical base pairs for structure refinement[44] and assessment[45]. Despite their success, all these methods demand huge computational resources and are thus less suitable for modeling large RNA molecules. Therefore, it is desired to balance the efficiency and accuracy in the CG model to appropriately describe noncanonical base pairs in RNA molecules.

The previous benchmark test[28] has shown that IsRNA1 CG model failed to recover some noncanonical base pairs in large RNAs, especially for the multi-way junctions. One reason is that the 2-bead CG representation for pyrimidine bases in previous IsRNA1 model could not fully capture noncanonical base pairing interactions such as the difference among base's WC, sugar, and Hoogsteen edges[37]. Hence, through systematic analyses of various noncanonical base pairs, we here developed an updated CG model, named IsRNA2 (version 2), to better account for noncanonical base pairing interactions in large RNAs. In this updated IsRNA2 model, both the purine and pyrimidine bases are represented by three CG beads and the WC, Hoogsten, and sugar edges of bases are sufficiently preserved. After reparameterization of the energy functions through the iterative simulated reference state approach[27], the ability of the updated IsRNA2 model in simulating noncanonical motifs was tested. Then, the IsRNA2 model was used to refine our previous predictions in the RNA-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Puzzles challenges. Moreover, apart from the multi-thread central processing unit (CPU) parallel calculations, we also developed the graphics processing unit (GPU) acceleration for IsRNA2, which significantly improved the simulation speed. Overall, the updated IsRNA2 is a promising CG model to study the folding dynamics and predict the 3D structures of large RNA molecules.

Simulation model and method

CG representation of RNA nucleotide.

Same as the previous IsRNA, IsRNA1 versions[27-28], the backbone of RNA is represented by two CG beads P and S located at atoms P and C4', which define the phosphate group and the ribose sugar ring, respectively. However, for the base moiety, the updated IsRNA2 model uses three CG beads for both purines and pyrimidines (see Figure 1). To reduce the overall degree of freedoms, the nucleotides adenine and guanine share a common CG bead R1, and the cytosine and uracil share two common CG beads Y1 and Y2. All the base's CG beads are positioned at the center-of-mass of the grouped heavy atoms. Compared to the previous IsRNA/IsRNA1 models (ten CG beads in total), the current IsRNA2 model introduces eleven unique types of CG beads to better describe both the canonical and noncanonical base pairing interactions in RNA molecules. Then, the topology file of RNA is more complicated in IsRNA2 and it requires moderately more computational resources relative to IsRNA1. The properties of those eleven CG beads are summarized in Table 1.

Updated CG force field in IsRNA2.

Generally, the force field in the updated IsRNA2 model can be written as

$$E_{total} = E_{bond}(b) + E_{angle}(\theta) + E_{torsion}(\phi) + E_{bp}(r, \theta, \phi) + E_{pair}(r) \quad (1)$$

Same as our previous models[27-28], the bond stretching energy $E_{bond}(b)$ and bond angle bending energy $E_{angle}(\theta)$ have a form of harmonic function plus a Gaussian term, and the torsion angle energy $E_{torsion}(\varphi)$ is in a quadruple Fourier form; see Supporting Information (SI) for details. The base pair energy $E_{bp}(r, \theta, \varphi)$ is used to restrain the predefined secondary structure and is defined as

$$E_{bp}(r, \theta, \phi) = E_{bond}(r_1) + E_{bond}(r_2) + E_{angle}(\theta) + \sum_{i=1}^{5} E_{torsion}(\phi_i) \quad (2)$$

Here E_{bond}, E_{angle}, and $E_{torsion}$ share the identical formulas for those in Eq. 1 and the definitions of the structural parameters (r_1, r_2, θ, and φ_i) for all the three canonical base pairs are listed in Table 2. Only the canonical base pairs (GC, AU, and GU) are restrained by the energy $E_{bp}(r, \theta, \varphi)$ through the whole work.

To enable GPU acceleration in the LAMMPS platform[46], the pairwise interaction $E_{pair}(r)$, which describes the base-base stacking, noncanonical base pairing, base-backbone and backbone-backbone interactions, is modified as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$E_{pair}(r) = \varepsilon\left(\frac{\sigma}{r}\right)^9 + D_0\left[e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}\right] + \frac{H_1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(r-r_1)^2}{2\sigma_1^2}} + \frac{H_2}{\sqrt{2\pi}\sigma_2}e^{-\frac{(r-r_2)^2}{2\sigma_2^2}}, r < r_{cut} \quad (3)$$

The first term with $\varepsilon = 0.5\ kcal/mol$ and $\sigma = (\sigma_i + \sigma_j)/2$ (σ_i is the diameter of CG bead i) accounts for excluded volume interaction, which is similar with the energy term $E_{LJ}(r)$ in the previous IsRNA/IsRNA1 models[27-28]. The second Morse term describes the overall profile of the pairwise interaction and the following two Gaussian terms are used to fix the major local minimums. For different CG bead pairs, the cutoff distance r_{cut} varies from 6.9 to 13.5Å.

The iterative simulated reference state approach[27] was used to parameterize the CG force field for IsRNA2. In our previous studies[27], this iterative simulated approach has been demonstrated to account for the correlated interactions between different structural degrees of freedoms as well as the effects from inherent chain connectivity and excluded volume, which has resulted in a reasonably accurate CG force field. Here, a simulated dataset contained 70 structures with size 26~188 nts and considerable noncanonical base pairs was used. In each step of the iterative simulated reference state calculation, after relaxation of the native structures, a total of 35,000 snapshots were generated from MD simulation trajectories to deduce the energy parameters. We followed the identical procedure given in our previous work[28] to sequentially parameterize energy functions in the new IsRNA2 model. We noted that no explicit artificial bias toward the native structures was introduced in the determination of energy functions of IsRNA2 through iterative simulated reference state approach[27]. Thus, the potential bias problem in the following test tests can be neglected. See SI Table S1-S4 for all the energy parameters obtained for IsRNA2 model.

Simulation details.

The MD simulations with the new IsRNA2 model were implemented in the modified LAMMPS software[46], and the Langevin dynamics (NVT ensemble) with integration timestep $\Delta t = 1 fs$ was performed. For RNA 3D structure prediction or refinement, replica-exchange MD (REMD) simulations with ten replicas possessed temperatures from 200K to 425K were run to enhance the sampling efficiency in the 3D conformational space. The simulation time for each replica is 50ns and three duplicated runs with different initial structures (if available) were performed. Thus, the total simulation time for a prediction/refinement is 1.5μs (3 duplicated run*10 replica*50 ns). After sufficient relaxation, the structure snapshots were collected from the last 25ns simulations in the interval of 50ps. To obtain the predicted structures, the top 10% structures with lowest potential energies from the collected snapshots (5,000 snapshots in total) were clustered based on pairwise root-mean-square deviations (RMSDs). The detailed process and choice of the cutoff RMSD threshold for the clustering can be found in our previous study[28]. The centroid structures of the top clusters (ranked by their sizes) provide the top predicted 3D structures.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Similar to that in the IsRNA, IsRNA1 versions[27-28], a built-in single-nucleotide fragment matching algorithm was developed to recover the all-atom model from the five-bead CG representation. Finally, an atomic energy minimization was employed to reduce the possible atomic distortion and clash in the predictions.

Results and Discussions

Representation of noncanonical base pairs.

Noncanonical base pairs mediate specific interactions to stabilize the 3D architecture of various RNA motifs[40], including junction topology, kink turn, tetraloop-receptor motif, triple-stranded structure, quadruplex structure, and so on. Apart from the WC edge presented in the canonical interactions, noncanonical base pairs also involve one or two of the other two sides of the nucleobase: Hoogsteen and sugar edges[37-38]. Thus, to accurately predict noncanonical base pairing interactions using coarse-grained representations, a fundamental step is to preserve the essential properties of those three edges of nucleobase. However, this pivotal point was missed due to the two-bead representation of pyrimidines in previous IsRNA and IsRNA1 models. And benchmark test for RNA 3D structure prediction by IsRNA1 indicated that some loop segments or even the full configuration, such as for the multi-way junctions, misfolded in the simulations[28]. Therefore, following the previous classification of RNA base pairs[37], we here introduced a new three-bead CG representation for the base moieties of four types of nucleotides. As shown in Figure 1, in this three-bead representation, the newly updated IsRNA2 model preserves the essential features of the base's WC, Hoogsteen, and sugar edges, which enables appropriate descriptions of various noncanonical base pairing interactions. For the top 20 most occurred base pairs in the Representative Sets of RNA 3D Structures[47] (release 3.115), which contain both the canonical and noncanonical base pairs and cover nearly all the types of interactive base edges and cis-/trans-conformations, IsRNA2 could provide a specific set of distances between particular bead pairs for each case (see Figure 2). If necessary, a set of particular angles and torsions can also be introduced to describe certain base pairing interactions. Overall, the new five-bead CG representation for each nucleotide in IsRNA2 model provides a good starting point to accurately predict noncanonical base pairing interactions in RNA motifs.

De novo prediction of noncanonical RNA structures.

To test the capability of modeling noncanonical base pairing interactions, IsRNA2 was used to predict 3D structure on a benchmark set of 23 RNA noncanonical motifs derived from a previous study[41]. Those noncanonical structures were observed in high-resolution crystallographic models of important RNA molecules and contain various common RNA motifs, such as tetraloop, kink-turn, hook-turn, 3-way junction, pseudoknot, and so on. As shown in Table 3, those RNA motifs involve one or more segments and their sizes vary from 6 to 50 nts. Here, we mainly focused on the conformations of noncanonical regions. Thus, to improve the efficiency and accuracy of conformation sampling, two canonical base pairs immediately adjacent to the motifs at each end are constrained as boundary conditions through Eq. 2. Similar to our previous study, the template-based algorithms Vfold3D[12,48]/

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

VfoldLA[18,49] were used to generate three initial 3D structures, if available, for REMD simulations in IsRNA2 and finally five candidate predictions were provided.

For 15 of the 23 noncanonical motifs, IsRNA2 provided at least one of five predictions with less than 3.0 Å all-heavy-atom RMSD to the experimentally determined structure (see Table 3). Among them, the best predicted 3D structures for 9 cases reached atomic accuracy (RMSD < 2.0 Å) and nearly all the native noncanonical base pairs were recovered, including not only the nucleotides but also the base edges in the noncanonical contact, when the fluctuations in base pairing interactions are neglected. Our current study cases incorporated those widely adopted RNA motifs, such as fragments with A-C, G-G, and G-A base pairs[50], *Eschericia coli* SRP domain[51], GAGA tetraloop from sarcin/ricin domain[52], J4/5 from P4-P6 domain, *Tetrahymena thermophila* ribozyme[53], and so on (see Fig. 3). However, IsRNA2 failed to predict the near-native 3D structures (RMSD < 5.0 Å) for 4 motifs (see Table 3), including the three-way junctions in the active site[56] and pre-catalytic conformation of hammerhead ribozymes[57], which may be caused by the insufficient sampling of conformational space under the current simulation conditions. Another possibility is that the current energy functions of IsRNA2 could not well capture the particularly sharp turns of the backbone, such as for the kink-turn motif[58] and J5-5a hinge in the P4-P6 domain of *Tetrahymena* ribozyme[53]. To fix those problems in future, incorporation of experimental information, such as the NMR data[26,39], into the simulation as constraints is promising.

Additionally, we compared the performance of IsRNA2 on 3D structure prediction with two other models and the results were summarized in Table 3. Compared with the previous IsRNA1 model[28], because of the clear definition of three base interacting edges in the updated CG representation, IsRNA2 provided better predictions (smaller RMSDs) for 18 of 23 noncanonical structures. Moreover, the average RMSD over all the tested structures decreased from 4.29 Å (IsRNA1) to 3.17 Å (IsRNA2) and the average interaction network fidelity[59] (INF) increased from 0.68 (IsRNA1) to 0.75 (IsRNA2). Ideally, an INF of 1.0 means that the predicted structure perfectly reproduces the interaction networks in the native structure. These results demonstrated the importance of definition of base interacting edges (WC, Hoogsteen, and sugar edges) in CG representation for accurate modeling of noncanonical base pairs. When compared to the atomic FARFAR model[41], which uses fragment assembly drawn from a crystallographic model and a full-atom energy function, the CG IsRNA2 model could still obtain better predictions (smaller RMSDs) for 13 noncanonical structures and slightly lower the average RMSD from 3.54 Å (FARFAR) to 3.17 Å (IsRNA2); See Table 3. Additionally, for the benchmark set of 23 noncanonical RNA structures, the numbers of best predictions (lowest RMSD) provided by IsRNA2, FARFAR, and IsRNA1 are 11, 8, and 5 (a case shares the identical RMSD with FARFAR), respectively. Overall, the updated IsRNA2 CG model demonstrated comparable or better performance than the previous atomic model in *de novo* modeling of noncanonical RNA motifs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Refinement of RNA 3D structure predictions in *RNA-Puzzles*.

As a CASP-like blind assessment of RNA 3D structure prediction, the *RNA-Puzzles* (http://www.rnapuzzles.org/) provides a platform for the evaluation of cutting-edge RNA structure prediction algorithms and the results of four rounds of challenges have been published[32-35]. To further verify the improvement of IsRNA2 on 3D structure prediction for large RNA molecules, we used IsRNA2 model to refine the predictions submitted by our group (Chen group) in the previous RNA-Puzzles challenges. With the top 3 models (if available) with lowest RMSDs submitted by Chen group as the initial structures and the secondary structure (canonical base pairs) extracted from the native structure as constraints (Eq. 2), 50 ns (each replica) REMD simulations were run to refine the 3D predictions. However, same as the previous work[28], only a short 1.5 ns REMD simulation was run for the multi-way junctions to avoid large conformation changes. Finally, top five predictions from the centroid structures of clusters were obtained.

Here, the predictions for 13 challenges in RNA-Puzzles have been refined by IsRNA2 model. Those 13 challenges include ribozyme, riboswitch, virus-associated RNA, and RNA aptamer, with their sizes varying from 37 to 112 nts. Moreover, the structural topologies of those challenges cover stem-loop, multi-way junction, pseudoknot, and structure contained tertiary interaction. As shown in Fig. 4, out of the 13 selected challenges, the lowest RMSDs of five predictions for 11 cases decreased after refinement using IsRNA2, indicating that the IsRNA2 model could indeed further refine the initial submissions from Chen group. For instance, for the glycine riboswitch[60] (PDB id: 3owz, puzzle #3), Mango-III fluorogenic aptamer[61] (PDB id: 6e8u, puzzle #23), and an adenovirus virus-associated RNA[62] (PDB id: 6ol3, puzzle #24), the lowest RMSDs of the best predictions decreased from 7.24, 10.59, and 11.03 Å (initial predictions by Chen group) to 5.34, 7.85, and 7.30 Å (refined by IsRNA2), respectively. Only one of the remaining two challenges had a larger RMSD than the previous prediction (PDB id: 5k7c, puzzle #17), while the other case remained nearly unchanged (PDB id: 5nz6, puzzle #21). The average RMSD over the total 13 challenges decreased from 8.40 Å (Chen group) to 7.10 Å (IsRNA2). Furthermore, for 6 of 13 challenges, refinement by IsRNA2 model can obtain even better predictions than the best models from the submissions of all groups (see Fig. 4), such as for the regulatory motif from the thymidylate synthase mRNA[63] (PDB id: 3mei, puzzle #1), the glycine riboswitch[60] (PDB id: 3owz, puzzle #3), and the twister sister ribozyme[64] (PDB id: 5t5a, puzzle #19). Therefore, IsRNA2 model might be a powerful tool to refine the 3D structures predicted by other programs, for instance, the template-based Vfold3D[12]/VfoldLA[18] model.

Furthermore, we compared the performance of IsRNA2 on model refinement with two recent methods, namely FARFAR2[65] and RNA-BRiQ[44] Based on the previous Rosetta's FARFAR algorithm, FARFAR2 integrated RNA-Puzzle-inspired innovations with updated fragment libraries and helix modeling and could recover native-like structures more accurate than models submitted during the RNA-Puzzles trials[65]. RNA-BRiQ combined a high-resolution knowledge-based potential (BRiQ) with a nucleobase-centric sampling algorithm to provide a robust improvement in refining near-native RNA models[44]. As shown in Table 4, for 11 announced RNA-Puzzles challenges, IsRNA2, FARFAR2, and RNA-BRiQ provide best predictions (over those three methods) for 5, 1, and 6 (one case is identical to IsRNA2)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cases, respectively. We noted that the sizes of the candidate pool to select the optimal prediction model for IsRNA2, FARFAR2, and RNA-BRiQ are 5, 10, and 20, respectively. Overall, despite the recent advancements by FARFAR2 and RNA-BRiQ algorithms, IsRNA2 is a competitive method in RNA model refinement.

GPU acceleration for large RNA molecules.

For large RNA molecules with size > 100 nts, the sufficient sampling in conformational space is challenging even at CG representation due to their huge number of available conformations. On the other hand, GPUs have become popular as accelerators in high-performance computing due to their impressive floating-point capabilities, high memory bandwidth, and low electrical power requirement. For instance, numerous MD codes have been developed to utilize GPUs to gain impressive speedups[66-71]. Thus, based on the "GPU" package[72] in the LAMMPS MD software[46], we have embedded and compiled the related source codes for the modified pairwise energy function in Eq. 3 to enable GPU acceleration in the IsRNA2 model. In this way, we expected the sampling speed for large RNAs to be largely improved.

The performance of GPU acceleration was tested on 10 RNAs with sizes from 50 to 490 nts. With the native 3D structure as initial state and the secondary structure extracted from the native structure as constraints, a 1 ns MD simulation (10^6 steps with integration timestep $\Delta t = 1 fs$) was run on the Intel(R) Core(TM) i9-9900K 3.6GHz CPU and the GeForce RTX 2080 Ti GPU. The wall-clock time was collected and recorded. For comparison, the clock time of one single CPU thread run and two CPU threads parallel run were both reported. As shown in Fig. 5, for a bacterial ribonuclease P RNA[73] (PDB id: 2a64, 298 nts) and the human Ribonuclease P Holoenzyme[74] (PDB id: 6ahu, 413 nts), the clock time on a single CPU thread are 21.4 and 28.3 minutes, respectively, and they decrease to 6.5 and 8.8 minutes when the GPU accelerator (one CPU thread plus a GPU card) was employed, which indicates an about 3.2-fold speedup for GPU acceleration. When compared to the parallel run on two CPU threads, the GPU acceleration could also gain an about 2.0-fold speedup for the those two RNAs. Moreover, we noted a perfect linear relationship between the clock time and the size of RNA molecule for the GPU acceleration. Overall, these results demonstrated that the CG IsRNA2 model plus GPU accelerators can be a powerful platform to study large RNA molecules.

Conclusion

Noncanonical base pairs play a pivotal role in stabilizing RNA 3D structures[40], especially for large RNAs, and pose one of the most challenging bottlenecks for current RNA 3D structure prediction[34-35]. On the other hand, CG models are promising and sometimes more suitable approaches, compared to atomistic models, in studying the dynamics of large RNAs due to reduced degrees of freedom and smoother free energy landscape and hence more efficient conformational sampling[75]. Here, based on our previous efforts in RNA CG modeling[27-28], we developed an updated IsRNA2 (version 2) model to study the noncanonical base pairing interactions in large RNA molecules. By introducing a five-bead CG representation for both purine and pyrimidine nucleotides, the updated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

IsRNA2 model preserves the definition of three interacting edges of bases, namely WC, Hoogsteen and sugar edges, and captures the fundamental elements to accurately describe various base pairing interactions, including both canonical and noncanonical base pairs. After re-parameterizing the energy functions through the iterative simulated reference state approach, the IsRNA2 was used to *de novo* model noncanonical RNA motifs and refine 3D structure predictions in RNA-Puzzles challenges. For 15 out of 23 tested noncanonical RNA structures, IsRNA2 achieved a near atomic-resolution (RMSD < 3.0 Å) prediction and recovered most of the native noncanonical base pairs. With significantly improved accuracy, benchmarks also indicated that IsRNA2 is able to achieve a comparable performance to the atomic model in *de novo* modeling of noncanonical RNA structures. To further confirm the ability of IsRNA2 in modeling noncanonical base pairs, an additional benchmark test on more noncanonical motifs, such as those from Rosetta-SWM method[43], is needed in future. Furthermore, out of 13 selected challenges in RNA-Puzzles, 3D structure predictions for 11 cases were obviously refined by simulations in IsRNA2. For some challenges, IsRNA2 can provide even better models than the previous best submissions. Finally, the GPU acceleration was introduced in IsRNA2 model to boost the sampling speed, with a ~ 3.2-fold speedup for very large RNA molecules. In all, the reported IsRNA2 is a promising coarse-grained model to study the noncanonical base pairs in large RNA molecules.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Kevin Chan, Zhi He and Yangwei Jiang for helpful discussions. This work was partially supported by the National Institutes of Health grants (R01-GM117059 and R35-GM134919 to S. Chen), by the National Natural Science Foundation of China (Grants U1967217, 11574224 to R. Zhou), National Independent Innovation Demonstration Zone Shanghai Zhangjiang Major Projects (ZJZX2020014 to R. Zhou) and the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (SN-ZJU-SIAS-003 to R. Zhou). R.Z. also acknowledges the financial support from W. M. Keck Foundation (Grant award 2019-2022).

References

1. Atkins JF; Gesteland RF; Cech TR RNA Worlds: From Life's Origins to Diversity in Gene Regulation; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2011.
2. Scott WG Ribozymes. Curr. Opin. Struct. Biol 2007, 17, 280–286. [PubMed: 17572081]
3. Serganov A; Patel DJ Molecular recognition and function of riboswitches. Curr. Opin. Struct. Biol 2012, 22, 279–286. [PubMed: 22579413]
4. The RNAcentral Consortium. RNAcentral: a comprehensive database of non-coding RNA sequences. Nucleic Acids Res. 2017, 45, D128–D134. [PubMed: 27794554]
5. RNAcentral Consortium. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. Nucleic Acids Res. 2021, 49, D212–D220. [PubMed: 33106848]
6. Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE The Protein Data Bank. Nucleic Acids Res. 2000, 28, 235–242. [PubMed: 10592235]
7. Laing C; Schlick T Computational approaches to 3D modeling of RNA. J. Phys.: Condens. Matter 2010, 22, 283101. [PubMed: 21399271]
8. Dawson WK; Bujnicki JM Computational modeling of RNA3D structures and interactions. Curr. Opin. Struct. Biol 2016, 37, 22–28. [PubMed: 26689764]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

9. Sun LZ; Zhang D; Chen SJ Theory and modeling of RNA structure and interactions with metal ions and small molecules. Annu. Rev. Biophys 2017, 46, 227–246. [PubMed: 28301768]
10. Dans PD; Gallego D; Balaceanu A; Darré L; Gómez H; Orozco M Modeling, simulations, and bioinformatics at the service of RNA structure. Chem 2019, 5, 51–73.
11. Rother M; Rother K; Puton T; Bujnicki JM ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res. 2011, 39, 4007–22. [PubMed: 21300639]
12. Cao S; Chen S-J Physics-based de novo prediction of RNA 3D structures. J. Phys. Chem. B 2011, 115, 4216–4226. [PubMed: 21413701]
13. Popenda M; Szachniuk M; Antczak M; Purzycka KJ; Lukasiak P; Bartol N; Blazewicz J; Adamiak RW Automated 3D structure composition for large RNAs. Nucleic Acids Res. 2012, 40, e112. [PubMed: 22539264]
14. Zhao Y; Huang Y; Gong Z; Wang Y; Man J; Xiao Y Automated and fast building of three-dimensional RNA structures. Sci. Rep 2012, 2, 734. [PubMed: 23071898]
15. Xu XJ; Chen SJ Topological constraints of RNA pseudoknotted and loop-kissing motifs: applications to three-dimensional structure prediction. Nucleic Acids Res. 2020, 48, 6503–6512. [PubMed: 32491164]
16. Das R; Baker D Automated de novo prediction of native-like RNA tertiary structures. *Proc.* Natl. Acad. Sci. U. S. A. 2007, 104, 14664–14669. [PubMed: 17726102]
17. Parisien M; Major F The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 2008, 452, 51–55. [PubMed: 18322526]
18. Xu X; Chen SJ Hierarchical assembly of RNA three-dimensional structures based on loop templates. J. Phys. Chem. B 2018, 122, 5327–5335. [PubMed: 29258305]
19. Ding F, Sharma S, Chalasani P; Demidov VV; Broude NE; Dokholyan NV Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. RNA 2008, 14, 1164–1173. [PubMed: 18456842]
20. Xia Z; Gardner DP; Gutell RR; Ren P Coarse-grained model for simulation of RNA three-dimensional structures. J. Phys. Chem. B 2010, 114, 13497–13506. [PubMed: 20883011]
21. Chen A A; Garca A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. Proc. Natl. Acad. Sci. U. S. A 2013, 110, 16820–16825. [PubMed: 24043821]
22. Cragnolini T; Laurin Y; Derreumaux P; Pasquali S Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base-pairings. J. Chem. Theory Comput 2015, 11, 3510–3522. [PubMed: 26575783]
23. Boniecki MJ; Lach G; Dawson WK; Tomala K; Lukasz P; Soltysinski T; Rother KM; Bujnicki JM SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res. 2016, 44, e63. [PubMed: 26687716]
24. Dawson WK; Maciejczyk M; Jankowska EJ; Bujnicki JM Coarse-grained modeling of RNA 3D structure. Methods 2016, 103, 138–156. [PubMed: 27125734]
25. Ding F; Lavender CA; Weeks KM; Dokholyan NV Three-dimensional RNA structure refinement by hydroxyl radical probing. Nat. Methods 2012, 9, 603–608. [PubMed: 22504587]
26. Williams B; Zhao B; Tandon A; Ding F; Weeks KM; Zhang Q; Dokholyan NV Structure modeling of RNA using sparse NMR constraints. Nucleic Acids Res. 2017, 45, 12638–12647. [PubMed: 29165648]
27. Zhang D; Chen SJ IsRNA: An iterative simulated reference state approach to modeling correlated interactions in RNA folding. J. Chem. Theory Comput 2018, 14, 2230–2239. [PubMed: 29499114]
28. Zhang D; Li J; Chen SJ IsRNA1: de novo prediction and blind screening of RNA 3D structures. J. Chem. Theory Comput 2021, 17, 1842–1857. [PubMed: 33560836]
29. Zhang X; Zhang D; Zhao C; Tian K; Shi R; Du X; Burcke AJ; Wang J; Chen SJ; Gu LQ Nanopore electric snapshots of an RNA tertiary folding pathway. Nat. Commun 2017, 8, 1458. [PubMed: 29133841]
30. Zhao CH; Zhang D; Jiang YW; Chen SJ Modeling loop composition and ion concentration effects in RNA hairpin folding stability. Biophys. J 2020, 119, 1439–1455. [PubMed: 32949490]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

31. Nguyen PDM; Zheng J; Gremminger TJ; Qiu LM; Zhang D; Tuske S; Lange MJ; Griffin PR; Arnold E; Chen SJ; et al. Binding interface and impact on protease cleavage for an RNA aptamer to HIV-1 reverse transcriptase. Nucleic Acids Res. 2020, 48, 2709–2722. [PubMed: 31943114]
32. Cruz JA; Blanchet MF; Boniecki M; Bujnicki JM; Chen SJ; Cao S; Das R; Ding F; Dokholyan NV; Flores SC; et al. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. RNA 2012, 18, 610–625. [PubMed: 22361291]
33. Miao Z; Adamiak RW; Blanchet MF; Boniecki M; Bujnicki JM; Chen SJ; Cheng C; Chojnowski G; Chou FC; Cordero P; et al. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. RNA 2015, 21, 1066–1084. [PubMed: 25883046]
34. Miao Z; Adamiak RW; Antczak M; Batey RT; Becka AJ; Biesiada M; Boniecki M; Bujnicki JM; Chen SJ; Cheng CY; et al. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. RNA 2017, 23, 655–672. [PubMed: 28138060]
35. Miao Z; Adamiak RW; Antczak M; Boniecki MJ; Bujnicki JM; Chen SJ; Cheng CY; Cheng Y; Chou FC; Das R; et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. RNA 2020, 26, 982–995. [PubMed: 32371455]
36. Venclovas C; Zemla A; Fidelis K; Moult J Comparison of performance in successive CASP experiments. Proteins 2001, Suppl 5, 163–170.
37. Leontis NB; Westhof E Geometric nomenclature and classification of RNA base pairs. RNA 2001, 7, 499–512. [PubMed: 11345429]
38. Lemieux S; Major F RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. Nucleic Acids Res. 2002, 30, 4250–4263. [PubMed: 12364604]
39. Sripakdeevong P; Cevec M; Chang AT; Erat MC; Ziegeler M; Zhao Q; Fox GE; Gao X; Kennedy SD; Kierzek R; et al. Structure determination of noncanonical RNA motifs guided by ^{1}H NMR chemical shifts. Nat Methods 2014, 11, 413–416. [PubMed: 24584194]
40. Butcher SE; Pyle AM The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. Acc. Chem. Res 2011, 44, 1302–1311. [PubMed: 21899297]
41. Das R; Karanicolas J; Baker D Atomic accuracy in predicting and designing noncanonical RNA structure. Nat. Methods 2010, 7, 291–294. [PubMed: 20190761]
42. Sripakdeevong P; Kladwang W; Das R An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. Proc. Natl. Acad. Sci. U.S.A 2011, 108, 20573–20578. [PubMed: 22143768]
43. Watkins AM; Geniesse C; Kladwang W; Zakrevsky P; Jaeger L; Das R Blind prediction of noncanonical RNA structure at atomic accuracy. Sci. Adv 2018, 4, eaar5316. [PubMed: 29806027]
44. Xiong P; Wu R; Zhan J; Zhou Y Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement. Nat. Commun 2021, 12, 2777. [PubMed: 33986288]
45. Townshend RJL; Eismann S; Watkins AM; Rangan R; Karelina M; Das R; Dror RO Geometric deep learning of RNA structure. Science 2021, 373, 1047–1051. [PubMed: 34446608]
46. Plimpton S Fast parallel algorithms for short-range molecular dynamics. J. Comput. Phys 1995, 117, 1–19.
47. Leontis NB; Zirbel CL In RNA 3D Structure Analysis and Prediction; Leontis N, Westhof E, Eds.; Springer Berlin Heidelberg: Berlin, 2012; pp 281–298.
48. Xu X; Zhao P; Chen SJ Vfold: a web server for RNA structure and folding thermodynamics prediction. PLoS One 2014, 9, e107504. [PubMed: 25215508]
49. Xu X; Zhao C; Chen SJ VfoldLA: a web server for loop assembly-based prediction of putative 3D RNA structures. J. Struct. Biol 2019, 207, 235–240. [PubMed: 31173857]
50. Wild K; Weichenrieder O; Leonard GA; Cusack S The 2 Å structure of helix 6 of the human signal recognition particle RNA. Structure 1999, 7, 1345–1352. [PubMed: 10574798]
51. Deng J; Xiong Y; Pan B; Sundaralingam M Structure of an RNA dodecamer containing a fragment from SRP domain IV of Escherichia coli. Acta Crystallogr. D Biol. Crystallogr 2003, 59, 1004–1011. [PubMed: 12777762]
52. Correll CC; Beneken J; Plantinga MJ; Lubbers M; Chan YL The common and distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. Nucleic Acids Res. 2003, 31, 6806–6818. [PubMed: 14627814]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

53. Ye JD; Tereshko V; Frederiksen JK; Koide A; Fellouse FA; Sidhu SS; Koide S; Kossiakoff AA; Piccirilli JA Synthetic antibodies for specific recognition and crystallization of structured RNA. Proc. Natl. Acad. Sci. U. S. A 2008, 105, 82–87. [PubMed: 18162543]
54. Dann III CE; Wakeman CA; Sieling CL; Baker SC; Irnov I; Winkler WC Structure and mechanism of a metal-sensing regulatory RNA. Cell 2007, 130, 878–892. [PubMed: 17803910]
55. Lu XJ; Bussemaker HJ; Olson WK DSSR: an integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Res. 2015, 43, No. e142. [PubMed: 26184874]
56. Martick M; Lee TS; York DM; Scott WG Solvent structure and hammerhead ribozyme catalysis. Chem. Biol 2008, 15, 332–342. [PubMed: 18420140]
57. Feig AL; Scott WG; Uhlenbeck OC Inhibition of the hammerhead ribozyme cleavage reaction by site-specific binding of Tb. Science 1998, 279, 81–84. [PubMed: 9417029]
58. Klein DJ; Schmeing TM; Moore PB; Steitz TA The kink-turn: a new RNA secondary structure motif. EMBO J. 2001, 20, 4214–4221. [PubMed: 11483524]
59. Parisien M; Cruz JA; Westhof E; Major F New metrics for comparing and assessing discrepancies between RNA 3D structures and models. RNA 2009, 15, 1875–1885. [PubMed: 19710185]
60. Huang L; Serganov A; Patel DJ Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. Mol. Cell 2010, 40, 774–786. [PubMed: 21145485]
61. Trachman RJ 3rd.; Autour A; Jeng SCY; Abdolahzadeh A; Andreoni A; Cojocaru R; Garipov R; Dolgosheina EV; Knutson JR; Ryckelynck M; et al. Structure and functional reselection of the Mango-III fluorogenic RNA aptamer. Nat. Chem. Biol 2019, 15, 472–479. [PubMed: 30992561]
62. Hood IV; Gordon JM; Bou-Nader C; Henderson FE; Bahmanjah S; Zhang J Crystal structure of an adenovirus virus-associated RNA. Nat. Commun 2019, 10, 2871–2871. [PubMed: 31253805]
63. Dibrov S; McLean J; Hermann T Structure of an RNA dimer of a regulatory element from human thymidylate synthase mRNA. Acta Crystallogr. D Biol. Crystallogr 2011, 67, 97–104. [PubMed: 21245530]
64. Liu Y; Wilson TJ; Lilley DM The structure of a nucleolytic ribozyme that employs a catalytic metal ion. Nat. Chem. Biol 2017, 13, 508–513. [PubMed: 28263963]
65. Watkins AM; Rangan R; Das R FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. Structure 2020, 28, 963–976. [PubMed: 32531203]
66. Anderson JA; Lorenz CD; Travesset A General purpose molecular dynamics simulations fully implemented on graphics processing units. J. Comput. Phys 2008, 227, 5342–5359.
67. Liu WG; Schmidt B; Voss G; Muller-Wittig W Accelerating molecular dynamics simulations using Graphics Processing Units with CUDA. Comput. Phys. Commun 2008, 179, 634–641.
68. van Meel JA; Arnold A; Frenkel D; Zwart SFP Harvesting graphics power for MD simulations. Mol. Simul 2008, 34, 259–266.
69. Friedrichs MS; Eastman P; Vaidyanathan V; Houston M; Legrand S; Beberg AL; Ensign DL; Bruns CM; Pande VS Accelerating molecular dynamic simulation on graphics processing units. J. Comput. Chem 2009, 30, 864–872. [PubMed: 19191337]
70. Harvey MJ; Giupponi G; De Fabritiis G ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. J. Chem. Theory Comput 2009, 5, 1632–1639. [PubMed: 26609855]
71. Schmid N; Botschi M; Van Gunsteren WF A GPU solvent–solvent interaction calculation accelerator for biomolecular simulations using the GROMOS software. J. Comput. Chem 2010, 31, 1636–1643. [PubMed: 20127715]
72. Brown WM; Wang P; Plimpton SJ; Tharrington AN Implementing molecular dynamics on hybrid high performance computers–short range forces. Comput. Phys. Commun 2011, 182, 898–911.
73. Kazantsev AV; Krivenko AA; Harrington DJ; Holbrook SR; Adams PD; Pace NR Crystal structure of a bacterial ribonuclease P RNA. Proc. Natl. Acad. Sci. U. S. A 2005, 102, 13392–13397. [PubMed: 16157868]
74. Wu J; Niu S; Tan M; Huang C; Li M; Song Y; Wang Q; Chen J; Shi S; Lan P; et al. Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. Cell 2018, 175, 1393–1404.e11. [PubMed: 30454648]
75. Noid WG Perspective: Coarse-grained models for biomolecular systems. J. Chem. Phys 2013, 139, 090901. [PubMed: 24028092]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

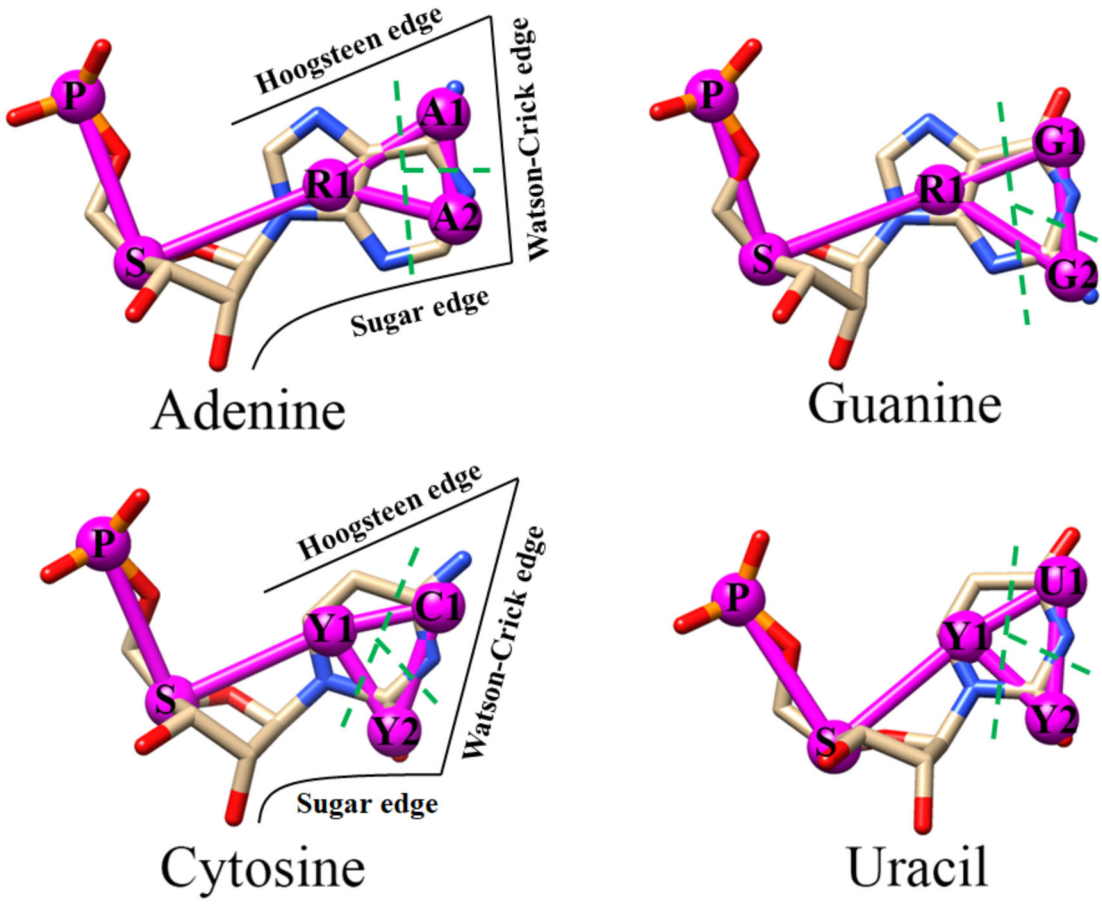


Figure 1.

Atomistic and coarse-grained representation of the four nucleotides in RNA molecule. The atomistic models are displayed as sticks with heavy atoms phosphorus, oxygen, carbon, and nitrogen colored by orange, red, gray, and blue, respectively. The coarse-grained representations in IsRNA2 are indicated by magenta bead-spring model, wherein the phosphate group and sugar ring are represented by bead P and S, respectively, and each base is coarse-graining into three different beads. The related base moiety for each coarse-grained bead is divided by the green dashed line. Base's Watson-Crick, sugar, and Hoogsteen edges[37] for purine and pyrimidine are identified in the left plane.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

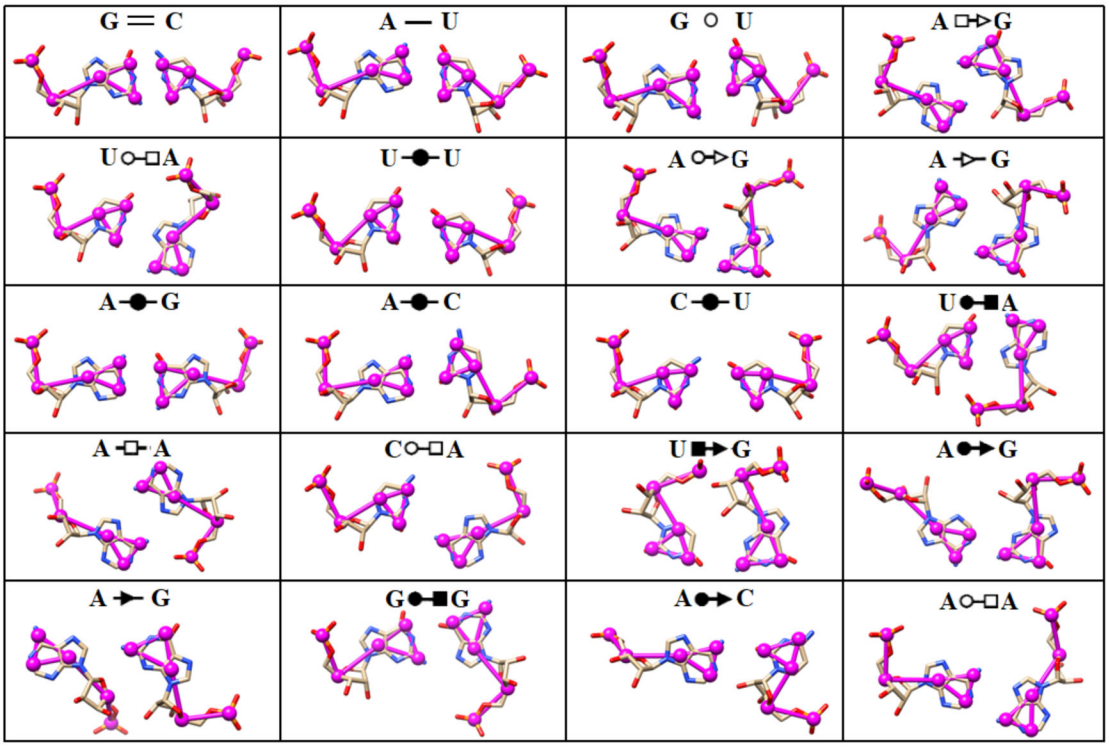


Figure 2.

Representations of 20 most occurred base pairing interactions in the IsRNA2 model. Both the canonical (the first three types) and noncanonical (the remaining) base pairs are annotated in the LW form[37]. : GC cis Watson-Crick, : AU cis Watson-Crick, : GU wobble, : trans Hoogsteen/Sugar edge, : trans Watson-Crick/Hoogsteen, : cis Watson-Crick/Watson-Crick, : trans Watson-Crick/Sugar edge, : trans Sugar edge/Sugar edge, : cis Watson-Crick/Hoogsteen, : trans Hoogsteen/Hoogsteen, : cis Hoogsteen/Sugar edge, : cis Watson-Crick/Sugar edge, : cis Sugar edge/Sugar edge.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

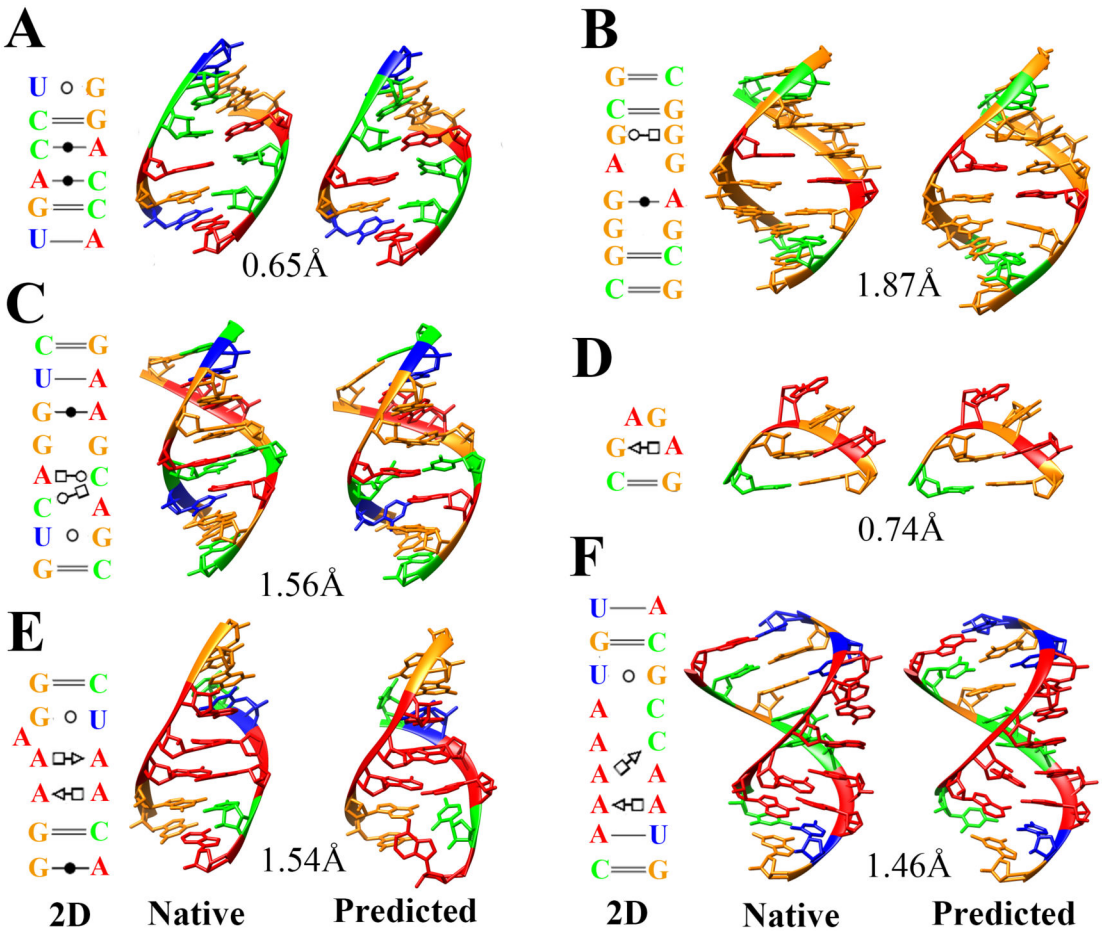


Figure 3.

3D structure modeling for several noncanonical RNA motifs with IsRNA2 model. (A) Fragment with A-C pairs, SRP helix VI[50] (PDB id: 1d4r, row 2 in Table 3), (B) fragment with G-G and G-A base pairs, SRP helix VI[50] (PDB id: 1d4r, row 3 in Table 3), (C) *Escherichia coli* SRP domain IV[51] (PDB id: 1lnt, row 7 in Table 3), (D) GAGA tetraloop from sarcin/ricin domain[52] (PDB id: 1q9a, row 9 in Table 3), (E) J4/5 from P4-P6 domain, *Tetrahymena thermophila* ribozyme[53] (PDB id: 2r8s, row 17 in Table 3), and (F) J4a-4b region, metal-sensing riboswitch[54] (PDB id: 2qbz, row 15 in Table 3). Adenine, guanine, cytosine, and uracil are colored by red, orange, green, and blue, respectively. The 2D structure (left column) is extracted from the native structure (middle column) by the program DSSR[55]. The predicted 3D structure (right column) is from the top five clusters with lowest RMSD. The all heavy-atom RMSDs for those predictions are also given.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

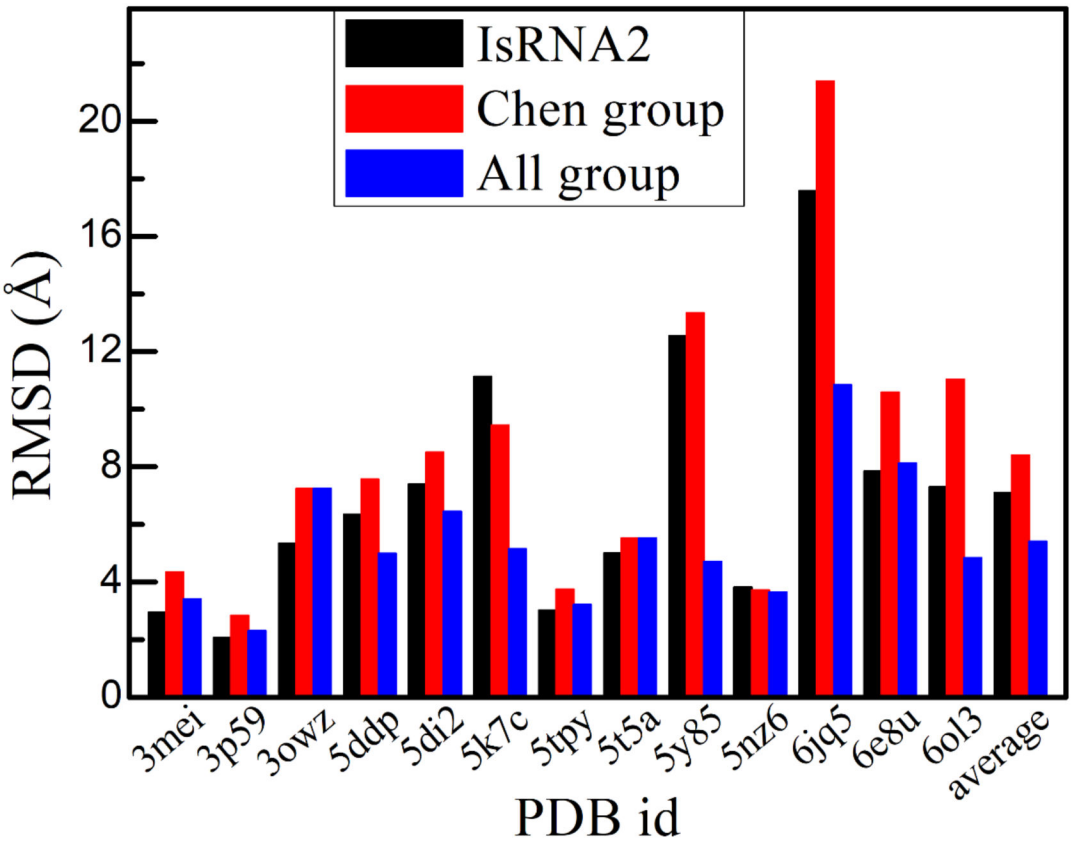


Figure 4.

Refinement of 3D structures predicted by Chen group in the RNA-Puzzles challenges through IsRNA2 model. Lowest RMSDs for the models from the predictions after refinement by IsRNA2 (“IsRNA2”) are shown. For comparison, the best prediction from initial submissions by Chen group (“Chen group”) and that from all group (“All group”, including Chen group) are also given for each case.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

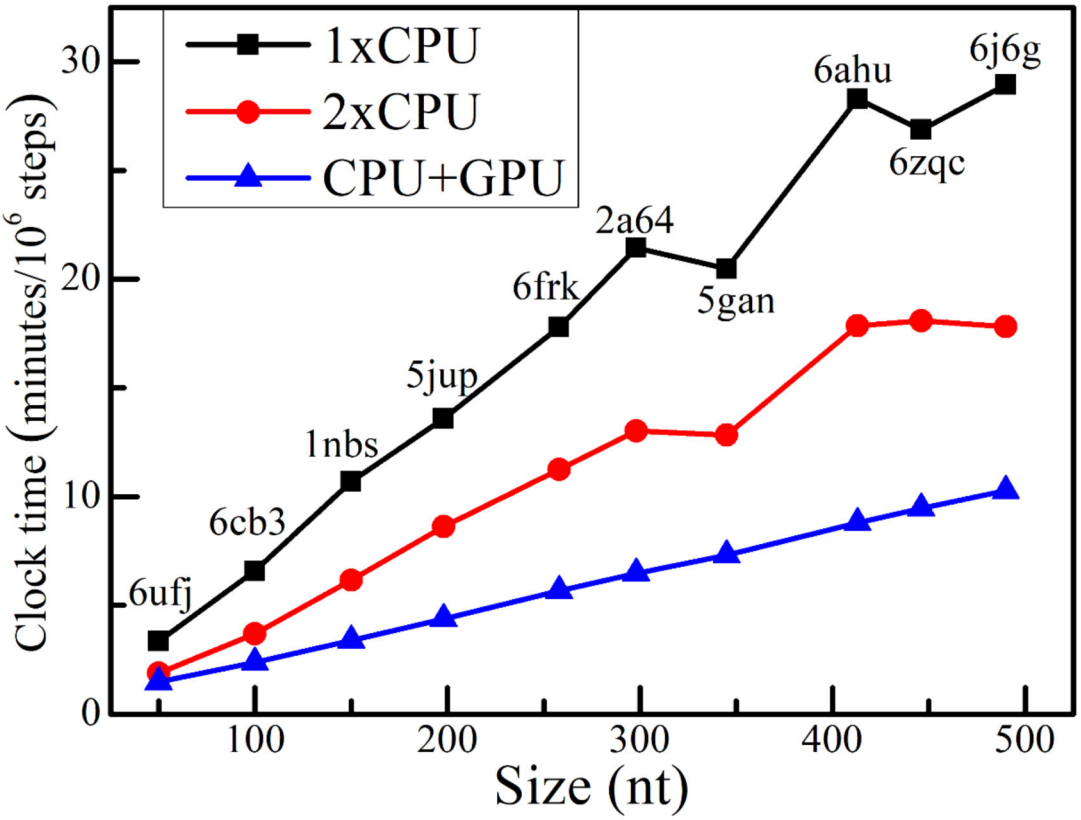


Figure 5.

Benchmark test for accelerating simulations in IsRNA2 through parallel central processing unit (CPU) threads and graphics processing unit (GPU) computing. For RNA molecules with different sizes (PDB ids are labeled), performances were tested on single CPU thread (1xCPU), two CPU threads (2xCPU), and GPU accelerator (CPU+GPU). The tested CPU is Intel(R) Core(TM) i9-9900K 3.6GHz and GPU is GeForce RTX 2080 Ti.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Properties of eleven types of coarse-grained beads in the updated IsRNA2 model.

CG bead	Mass (amu)	Diameter (Å)	Grouped heavy atoms
P	94.97	3.7	P, OP1, OP2, O5', O3'
S	92.05	3.1	C5', C4', O4', C3', C2', O2', C1'
R1	78.05	3.2	N9, C8, N7, C5, C4, N3
A1	26.02	2.5	C6, N6
A2	26.02	2.7	C2, N1
G1	42.02	2.7	C6, O6, N1
G2	26.02	2.7	C2, N2
Y1	38.03	3.1	N1, C5, C6
Y2	28.01	2.9	C2, O2
C1	40.03	2.5	N3, C4, N4
U1	42.02	2.7	N3, C4, O4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Structural parameters for secondary structure constrains used in Eq. 2.

Base pair	r_1	r_2	θ	φ_1	φ_2	φ_3	φ_4	φ_5
GC	G1-C1	G2-Y2	G2-Y2-C1	R1-G1-G2-Y2	G1-G2-Y2-C1	G2-Y2-C1-Y1	R1-G1-C1-Y1	R1-G2-Y2-Y1
AU	A1-U1	A2-Y2	A2-Y2-U1	R1-A1-A2-Y2	A1-A2-Y2-U1	A2-Y2-U1-Y1	R1-A1-U1-Y1	R1-A2-Y2-Y1
GU	G1-U1	G1-Y2	R1-G1-U1	R1-G2-G1-Y2	G2-G1-Y2-U1	G1-Y2-U1-Y1	R1-G1-U1-Y1	---

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Comparison of *de novo* modeling of noncanonical RNA motifs with IsRNA2, IsRNA1, and FARFAR model. The best performance over three models with lowest all heavy-atom RMSD is boldfaced for each case.

No.	Motif name (PDB id)	Size[a]	IsRNA2			IsRNA1			FARFAR[e]	
			Rank[b]	RMSD[c]	INF[d]	rank	RMSD	INF	rank	RMSD
1	Rev response element high-affinity site (1csl)	6+7	1	**2.72**	0.77	2	2.88	0.67	2	3.95
2	Fragment with A-C pairs, SRP helix VI (1d4r)	6+6	3	**0.65**	0.90	1	1.00	0.97	1	1.83
3	Fragment with G-G and G-A base pairs, SRP helix VI (1d4r)	8+8	1	**1.87**	0.78	1	2.41	0.76	3	3.27
4	UUCG tetraloop (1f7y)	8	1	2.60	0.71	1	2.77	0.67	1	**1.12**
5	Kink-turn motif (1jj2)	7+10	2	9.05	0.52	4	**8.55**	0.62	2	8.85
6	Helix with A-C base pairs (1kd5)	8+8	2	2.73	0.79	2	**2.45**	0.73	2	**2.45**
7	SRP domain IV (1lnt)	8+8	4	1.56	0.76	2	2.89	0.75	4	**1.54**
8	Hook-turn motif (1mhk)	5+6	3	3.58	0.64	4	4.99	0.46	5	**2.56**
9	GAGA tetraloop from sarcin-ricin loop (1q9a)	6	2	**0.74**	0.91	3	1.15	0.61	1	0.82
10	Loop 8, A-type RNase P (1u9s)	9	2	3.08	0.80	1	4.45	0.67	5	**1.38**
11	Pentaloop from conserved region of SARS (1xjr)	9	5	2.57	0.59	2	3.70	0.56	3	**1.10**
12	L3, thiamine pyrophosphate riboswitch (2gdi)	9	5	2.09	0.74	1	3.05	0.56	4	**2.00**
13	Active site, hammerhead ribozyme (2oeu)	11+7+5	4	7.67	0.65	2	**6.88**	0.60	4	8.64
14	Stem C internal loop, L1 ligase (2oiu)	8+8	2	**1.87**	0.78	5	5.68	0.43	1	2.24
15	J4a-4b region, metal-sensing riboswitch (2qbz)	9+9	5	**1.46**	0.82	2	3.29	0.77	3	3.71
16	P1-L3, SAM-II riboswitch (2qwy)	50	2	**3.94**	0.62	2	9.82	0.51	5	7.40
17	J4/5 from P4-P6 domain, Tetrahymena thermophila ribozyme (2r8s)	7+6	3	**1.54**	0.90	4	2.98	0.72	1	1.76
18	J5-5a hinge, P4-P6 domain, Tetrahymena ribozyme (2r8s)	10+9	5	**9.35**	0.55	2	9.95	0.65	3	9.99
19	Pseudoknot, domain III, CPV internal ribosome entry site (3b31)	12+8	1	3.77	0.85	5	**3.15**	0.91	4	3.55
20	G-A base pair (157d)	5+5	4	**0.86**	0.92	3	0.88	0.96	1	1.19
21	Helix with U-C base pairs (255d)	6+6	1	1.38	0.87	1	**1.30**	0.85	2	2.10
22	Loop E motif, 5S RNA (354d)	11+11	1	2.57	0.62	5	7.25	0.58	2	**1.64**
23	Pre-catalytic conformation, hammerhead ribozyme (359d)	11+8+6	2	**5.34**	0.66	2	7.14	0.71	5	8.44
	average		2.65	3.17	0.75	2.48	4.29	0.68	2.78	3.54

[a] For RNA motif contained multiple chains, the size (number of nucleotides) for each chain is separated by "+".

[b] The rank of the best prediction from the top five clusters.

[c] Lowest all-heavy-atom root-mean-square deviation (in Å) for the best prediction from the top five clusters.

[d] INF is the interaction network fidelity[59] for all the canonical and noncanonical base-pairing and base-stacking interactions.

[e] Since the original predictions of FARFAR are unavailable, the INFs for FARFAR are absent here.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Comparison of model refinement for RNA-Puzzles challenges with IsRNA2, FARFAR2, and RNA-BRiQ. The best prediction over those three models with lowest all heavy-atom RMSD is boldfaced for each challenge.

Puzzle (PDB)	Length (nt)	RNA	RMSD (Å)		
			IsRNA2[a]	FARFAR2[b]	RNA-BRiQ[c]
1 (3mei)	46	thymidylate synthase motif	2.94	2.50	**1.97**
2 (3p59)	100	nanosquare	**2.06**	2.71	**2.06**
3 (3owz)	84	glycine riboswitch	**5.34**	12.41	6.53
14b(5ddp)	61	Gln riboswitch (bound)	6.35	6.88	**6.14**
15 (5di2)	68	hammerhead ribozyme	7.40	**5.98**	6.78
17 (5k7c)	58	pistol ribozyme	11.13	6.69	**5.68**
18 (5tpy)	71	Zika xrRNA	**3.02**	5.02	3.47
19 (5t5a)	62	twister sister ribozyme	**5.01**	5.16	6.97
20 (5y85)	68	twister sister ribozyme	12.55	4.03	**3.54**
21 (5nz6)	41	guanidinium-In riboswitch	**3.82**	6.04	3.83
24 (6ol3)	112	adenovirus virus-associated RNA	7.30	7.68	**5.53**

[a] Lowest RMSD of top 5 predictions.

[b] Best of 10 low-energy clusters, data collected from Ref 65.

[c] Lowest energy model within 20 refinement models, data collected from Ref 44.