Biophysical Journal
Article

Biophysical Society

Modeling Structure, Stability, and Flexibility of Double-Stranded RNAs in Salt Solutions

Lei Jin,[1] Ya-Zhou Shi,[2] Chen-Jie Feng,[1] Ya-Lan Tan,[1] and Zhi-Jie Tan[1,*]
[1]Center for Theoretical Physics and Key Laboratory of Artificial Micro- & Nanostructures of Ministry of Education, School of Physics and Technology, Wuhan University, Wuhan, China and [2]Research Center of Nonlinear Science, School of Mathematics and Computer Science, Wuhan Textile University, Wuhan, China

ABSTRACT Double-stranded (ds) RNAs play essential roles in many processes of cell metabolism. The knowledge of three-dimensional (3D) structure, stability, and flexibility of dsRNAs in salt solutions is important for understanding their biological functions. In this work, we further developed our previously proposed coarse-grained model to predict 3D structure, stability, and flexibility for dsRNAs in monovalent and divalent ion solutions through involving an implicit structure-based electrostatic potential. The model can make reliable predictions for 3D structures of extensive dsRNAs with/without bulge/internal loops from their sequences, and the involvement of the structure-based electrostatic potential and corresponding ion condition can improve the predictions for 3D structures of dsRNAs in ion solutions. Furthermore, the model can make good predictions for thermal stability for extensive dsRNAs over the wide range of monovalent/divalent ion concentrations, and our analyses show that the thermally unfolding pathway of dsRNA is generally dependent on its length as well as its sequence. In addition, the model was employed to examine the salt-dependent flexibility of a dsRNA helix, and the calculated salt-dependent persistence lengths are in good accordance with experiments.

INTRODUCTION

RNAs play a pervasive role in gene regulation and expression. In addition to single-stranded (ss) RNAs such as mRNAs and tRNAs, double-stranded (ds) RNAs are widespread in cells and are involved in a variety of biological functions (1–3). For examples, small noncoding dsRNAs can play a critical role in mediating neuronal differentiation (4), dsRNA segments of special lengths can inhibit the translation of mRNA molecules into proteins through attaching to mRNAs (5,6), and dsRNAs of more than 30 basepair (bp) length can be key activators of the innate immune response against viral infections (7). Generally, dsRNAs realize their biological functions through becoming partially melted or changing their conformations (2–9). Furthermore, the interchain interactions in stabilizing structures of dsRNAs are very sensitive to the environment (e.g., temperature and ion conditions) (10–14). Thus, a full understanding of dsRNA-mediated biology would require the knowledge of three-dimensional (3D) structures, structural stability, and flexibility of dsRNAs in ion solutions.

The 3D structures of RNAs, including dsRNAs, can be measured by several experimental methods such as x-ray crystallography, NMR spectroscopy, and cryo-electron microscopy. However, it is still technically challenging and expensive to experimentally derive 3D structures of RNAs at high resolution, and the RNA structures deposited in the Protein Data Bank (PDB) are still limited (15). Therefore, as complementary methods, some computational models have been developed in recent years, aiming to predict RNA 3D structures in silico (16–23). The fragment assembly models (24–31) such as MC-Fold/MC-Sym pipeline (24), 3dRNA (25–27), RNAComposer (28), and Vfold3D (29,30) can successfully predict 3D structures of RNAs, including even large RNAs, at fast speed; however, these methods are generally based on given secondary structures and the limited known RNA 3D structures deposited in the PDB. Although the fragment assembly method of fragment assembly of RNA (FARNA) (31) can predict 3D structures for RNAs from sequences, it can only be efficient for small RNAs because of its full-atomic resolution. In parallel ways, some coarse-grained (CG) models (32–41) such as iFold (42), SimRNA (43), HiRE-RNA (44), and RACER (45,46) have been proposed to predict 3D structures for RNAs with medium length from their sequences based on knowledge-based statistical potentials and/or experiential

Submitted June 4, 2018, and accepted for publication August 24, 2018.
*Correspondence: zjtan@whu.edu.cn
Lei Jin and Ya-Zhou Shi contributed equally to this work.
Editor: Tamar Schlick.
https://doi.org/10.1016/j.bpj.2018.08.030
© 2018 Biophysical Society.

parameters. However, these existing 3D-structure prediction models seldom make quantitative predictions for thermodynamic stability and flexibility of RNAs.

Simultaneously, some models have been employed to predict the thermodynamics of RNAs. Vfold2D/VfoldThermal (29,30) with involving thermodynamic parameters can make reliable predictions on the free-energy landscape of RNAs, including pseudoknots, at the secondary structure level. The model proposed by Denesyuk and Thirumalai (47,48) can predict the thermodynamics of small RNAs well, whereas such a structure-based (Gō-like) model could not predict 3D structures of RNAs solely from the sequences. Although other models such as iFold (42), HiRE-RNA (44), oxRNA (49), and NARES-2P (50,51) may give melting curves of RNAs, there is still a lack of extensive experimental validation for these models.

Furthermore, RNAs are highly charged polyanionic molecules, and RNA structure and stability are generally sensitive to solution ion conditions, especially multivalent ions such as Mg^{2+} (8,10–14). The role of ions in RNA structure and stability—especially the role of Mg^{2+}, which is generally beyond the mean field descriptions (52,53)—is seldom involved in the existing 3D-structure prediction models. To predict the 3D structures and stability of RNAs in ion solutions, we have developed a CG model with implicit electrostatic potential (54,55), and the model has been validated through making reliable predictions on 3D structures and stability of RNA hairpins and pseudoknots as well as the ion effect on their stability. However, the model of the previous version was developed to predict the 3D structure and stability of ssRNA and cannot be directly employed to simulate the folding of RNAs with multiple strands because the translational entropy change of RNA strands during structure formation was not taken into account in the model. Unlike ssRNAs, the structure folding and stability of RNAs with multiple strands would be strongly dependent on strand concentration, which can make predictions on folding of dsRNAs very inefficient, especially for predicting dsRNA stability at (low) experimental strand concentration.

In this work, we further developed our previous three-bead CG model for ssRNAs to predict the 3D structure and stability of dsRNAs from their sequences by involving strand translations associated with strand concentration. For predicting stability of dsRNAs, we involved a correction of finite size effect (56) for dsRNA with two strands in a simulation box and introduced a technique of transforming melting curves of dsRNAs at high strand concentrations to those at low strand concentrations (57,58) to promote the simulation efficiency of the model. Furthermore, an implicit structure-based electrostatic potential is introduced to capture the effect of ions such as Mg^{2+} on 3D structures and stability of dsRNAs. As compared with the extensive experimental data, this model can predict the 3D structures, stability, and flexibility of various dsRNAs with high accuracy, and the effects of monovalent/divalent ions on the stability and flexibility of dsRNAs can be well captured by this model. Additionally, our further analyses show that the thermally unfolding pathway of dsRNA is dependent on not only its length but also its sequence.

METHODS

CG structure model and energy function

To reduce the complexity of nucleotides, in our CG model, one nucleotide is represented by three beads: the phosphate bead (P), sugar ring bead (C), and base bead (N) (54,55). The P and C beads are placed at the P and C4′ atom positions, and the base bead (N) is placed at N9 atom position for purine or N1 for pyrimidine; see Fig. 1. The three beads are treated as van der Waals spheres with the radii of 1.9, 1.7, and 2.2 Å, respectively (54,55).

The potential energy of a CG dsRNA is composed of two parts, bonded potential U_{bonded} and nonbonded potential $U_{nonbonded}$ (54,55):

$$U_{total} = U_{bonded} + U_{nonbonded}. \quad (1)$$

The bonded potential U_{bonded} represents the energy associated with pseudocovalent bonds between contiguous CG beads within any single chain, which includes bond-length energy U_b, bond-angle energy U_a, and dihedral-angle energy U_d:

$$U_{bonded} = U_b + U_a + U_d. \quad (2)$$

The initial parameters of these potentials were derived from the statistical analysis on the available 3D structures of RNA molecules in the PDB (http://www.rcsb.org/pdb/home/home.do), and two sets of parameters $Para_{helical}$ and $Para_{nonhelical}$ were provided for stems and single strands/loops, respectively. Note that only $Para_{nonhelical}$ is used in the folding process and both $Para_{helical}$ and $Para_{nonhelical}$ are used in structure refinement (54,55). The nonbonded potential $U_{nonbonded}$ in Eq. 1 includes the following five components:

$$U_{nonbonded} = U_{bp} + U_{bs} + U_{cs} + U_{exc} + U_{el}. \quad (3)$$

U_{bp} is the base-pairing interaction between Watson-Crick (G-C and A-U) and wobble (G-U) basepairs. U_{bs} and U_{cs} are sequence-dependent base stacking and coaxial stacking interactions between two neighbor basepairs and between two neighbor stems, respectively. The strengths of U_{bs} and U_{cs} were derived from the combined analysis of available thermodynamic parameters and Monte Carlo (MC) simulations (54,55). U_{exc} represents the excluded volume interaction between two CG beads, and it is modeled by a purely repulsive Lennard-Jones potential.

The last term U_{el} in Eq. 3 is a structure-based electrostatic energy for an RNA, which is newly refined in the model to better capture the contribution of monovalent and divalent ions to RNA 3D structures. The electrostatic potential is treated as a combination of Debye-Hückel approximation and the counterion condensation theory (52–55),

$$U_{el} = \frac{Q_i Q_j e^2}{4\pi\epsilon_0 \epsilon r_{ij}} e^{-\frac{r_{ij}}{l_D}}, \quad (4)$$

where r_{ij} is the distance between the i-th and j-th P beads, each of which carries a unit negative charge ($-e$). l_D is the Debye length of ion solution. ϵ_0 is the permittivity vacuum, and ϵ is the effective temperature-dependent dielectric constant of water (54,55). The reduced negative charge Q_i on the i-th P bead is given by

$$Q_i = 1 - f_i, \quad (5)$$

where f_i is the fraction of ion neutralization. In this model, beyond the assumption of uniform distribution of binding ions along RNA chain in

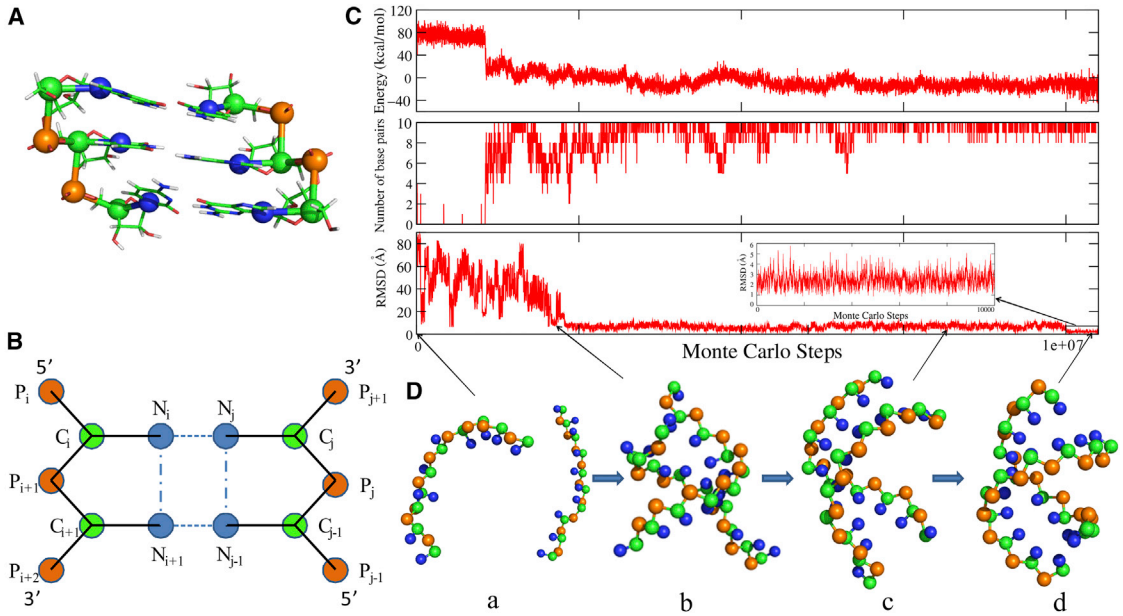

FIGURE 1 (*A*) The CG representation for one fragment of dsRNA superposed on an all-atom representation. The beads of P (*orange*) and C (*green*) are located at the P atom in the phosphate group and the C4′ atom in the sugar ring, respectively. The beads of N (*blue*) are located at the N9 atom position for purine or N1 atom position for pyrimidine. (*B*) A schematic representation for pseudobonds (*solid lines*), base pairing (*dashed lines*), and base stacking (*dash-dotted line*) in our model. (*C* and *D*) An illustration for the folding process of a typical dsRNA (PDB: 2JXQ) in our model. (*C*) The system energy (*top*), number of basepairs (*middle*), and the RMSD (*bottom*) along the simulated annealing MC simulation. The inset is the zoomed RMSD of the structure in the refinement procedure at the end of the simulation. (*D*) The four typical conformation states in the folding procedure. The structures are shown with PyMol (http://www.pymol.org). To see this figure in color, go online.

our previous model, f_i is dependent on RNA 3D structure and includes the contributions of monovalent and divalent ions

$$f_i = xf_i^1 + (1-x)f_i^2. \tag{6}$$

Here, x and $1-x$ represent the contribution fractions from monovalent and divalent ions, which can be derived from the tightly bound ion model (55,59–63). f_i^{ν} ($\nu = 1, 2$) is the binding fraction of ν-valent ions and is given by

$$f_i^{\nu} = \frac{N\bar{f}_i^{\nu}}{\sum_N e^{-\beta\nu\phi_i}} e^{-\beta\nu\phi_i}, \tag{7}$$

where N is the number of P beads. $\bar{f}_i^{\nu}$ represents the average charge-neutralization fraction of ions, and the counterion condensation theory gives that (52–55) $\bar{f}_i^{\nu} = 1 - (b/\nu l_B)$, where b is the average charge spacing on RNA backbone and l_B is the Bjerrum length. ϕ_i in Eq. 7 is the electrostatic potential at the i-th P bead and can be approximately calculated by

$$\phi_i = \sum_{j\neq i}^{N} \frac{l_B Q_j}{r_{ij}} e^{-\frac{r_{ij}}{l_D}}. \tag{8}$$

Equations 5, 6, 7, and 8 show that the structure-based reduced charge fraction Q_i needs to be obtained through an iteration process: 1) use ion-neutralization fraction f_i^{ν} to calculate f_i through Eq. 6, where the initial $f_i^{\nu} = \bar{f}_i^{\nu}$; 2) calculate Q_i through Eq. 5 and substitute Q_i in Eq. 8 to calculate ϕ_i; 3) calculate (renew) f_i^{ν} of ν-valent ions through Eq. 7; 4) repeat 1–3 until f_i^{ν} becomes converged. See more details in the Supporting Materials and Methods.

The detailed descriptions of the CG energy function as well as the parameters for the potentials in Eqs. 1, 2, and 3 can be found in the Supporting Materials and Methods.

Simulation algorithm

To effectively avoid the traps in local energy minima, the MC simulated annealing algorithm is used to sample conformations for dsRNA at given monovalent/divalent ion conditions. Based on the sequence of dsRNA, two initial random CG chains can be generated and be separately placed in a cubic box, the size of which is determined by concentration of ssRNA. Generally, the simulation of a dsRNA system with a given ion condition is performed from a high temperature (e.g., 110°C) to the target temperature (e.g., room/body temperature). At each temperature, the conformations of the dsRNA are sampled by intrastrand pivot moves and interstrand translation/rotation through the Metropolis algorithm until the system reaches enough equilibrium. In this process, the newly refined electrostatic potential U_{el} is involved (see Eq. 4), and U_{el} can only be obtained after an iterative process for Q_i. In practice, U_{el} is renewed over every 20 MC steps, and generally, U_{el} can be obtained through ~4 times iterations for converged Q_i. Thus, the increase in computation cost due to involving the newly refined U_{el} is negligible compared to the whole simulation cost. The equilibrium conformations of the system at each temperature can be saved to obtain 3D structures and structural properties of the dsRNA at each temperature.

Calculation of melting temperature

The stability of dsRNAs generally depends on strand concentration due to the contribution of translation entropy of melted ssRNA chains (64). However, for dsRNA with low strand concentrations (e.g., 0.1 mM in experiments), a very long simulation time is generally required to reach equilibrium for the dsRNA system. To make our calculation efficient, the simulations for dsRNAs can be performed at relatively high strand concentrations C_s^h (44,65). To avoid the boundary effect, we generally took a suitable high strand concentration C_s^h (e.g., 10 mM for dsRNA $\leq$6 bp and

1 mM for dsRNA >6 bp). Correspondingly, the volume of dsRNAs can become negligible compared to the simulation boxes, e.g., dsRNAs of 6 and 14 bp only occupy the simulation boxes with side length of ~70 and ~150 Å by ~2 and ~0.4% in volume, respectively. Based on the equilibrium conformations at each temperature T, the fraction $\Phi(T)$ of unfolded state characterized as a completely dissociated ss chain can be obtained at T. Because the small system of the simulation (two strands in a simulational box) can lead to a significant finite size effect (56), the predicted $\Phi(T)$ needs to be further corrected to the fraction $\theta_h(T)$ of unfolded state at the high bulk strand concentration C_s^h (56):

$$\theta_h(T) = 1 - \left(1 + \frac{\Phi(T)}{2a(1-\Phi(T))}\right) + \sqrt{\left(1 + \frac{\Phi(T)}{2a(1-\Phi(T))}\right)^2 - 1}, \quad (9)$$

where $a = 1$ and 2 for non-self-complementary and self-complementary sequences, respectively (56). Afterwards, based on $\theta_h(T)$ at the high strand concentration, the fraction $\theta(T)$ of unfolded state at an experimental strand concentration C_s (e.g., ~0.1 mM) can be calculated by (57,58)

$$\theta(T) = \frac{\lambda\theta_h(T)}{1 + \lambda\theta_h(T) - \theta_h(T)}, \quad (10)$$

where $\lambda = C_s^h/C_s$. Finally, the fractions $\theta(T)$ of unfolded state can be fitted to a two-state model to obtain the melting temperature T_m (54,55,57),

$$\theta(T) = 1 - \frac{1}{1 + e^{(T-T_m)/dT}}, \quad (11)$$

where dT is an adjustable parameter. More details about the calculation of melting temperature are given in the Supporting Materials and Methods. For long dsRNAs whose unfolding can be a non-two-state transition, we still used the above formulas to estimate their melting temperatures, in analogy to related experiment (66).

RESULTS AND DISCUSSION

In this section, our model was first employed to predict 3D structures of extensive dsRNAs in monovalent/divalent ion solutions. Afterwards, the model was used to predict stability of extensive dsRNAs and the effects of monovalent/divalent ions and further to analyze the thermally unfolding pathway of various dsRNAs. Finally, the model was employed to examine the salt-dependent flexibility of a dsRNA helix. Our predictions and analyses were extensively compared with the available experiments and existing models.

Structure predictions for dsRNAs in ion solutions

Two sets of available dsRNAs were used in this work on 3D structure prediction. One set includes 16 dsRNAs whose structures were determined by x-ray experiments (defined as the x-ray set), and the other set contains 10 dsRNAs whose structures were determined by NMR experiments in ion solutions (defined as the NMR set). The PDB codes as well as the descriptions of the dsRNAs in two sets are shown in Tables 1 and S3, respectively. In the following, because of the lack of ionic conditions for dsRNAs in the x-ray set, we first made the predictions on 3D structures for 26 dsRNAs in the x-ray set and NMR set at high salt (e.g., 1 M Na^+), i.e., assuming dsRNAs are nearly fully neutralized and consequently ignoring electrostatic contribution to the predicted structures. Afterwards, we further employed our model to predict 3D structures for dsRNAs in NMR set at their respective experimental ion conditions to examine whether the model with structure-based electrostatic potential can improve 3D structure predictions for dsRNAs in ion solutions.

Structure predictions for dsRNAs at 1M [Na^+]

For 26 dsRNAs in the x-ray set and NMR set, the 3D structures were predicted from sequences with strand concentration of 1 mM at high salt concentration (e.g., 1 M Na^+), regardless of possible ion effects. In the following, we used a paradigm dsRNA (PDB: 2JXQ; shown in Table 1) to show the structure predicting process of dsRNA with our model, which is shown in Fig. 1 *C*. First, the energy of the system reduces with the decrease of temperature (from 100°C to room temperature), and the dsRNA folds into native-like structures (e.g., structure *c* in Fig. 1 *D*) from an initial random configuration (e.g., structure *a* in Fig. 1 *D*). Second, a further structure refinement (~1.2×10^7 MC steps) is performed at the target temperature (e.g., room temperature), in which the last predicted structure from the annealing process is taken as input and the parameters $Para_{nonhelical}$ of bonded potentials are replaced by $Para_{helical}$ for the base-paired regions to better capture the geometry of helical stems (54,55). Finally, an ensemble of refined 3D structures (~10,000 structures) can be obtained over the last ~1×10^6 MC steps, and these structures can

TABLE 1 10 dsRNAs in the NMR Set for Structure Prediction at Respective Salt Conditions

PDB	Description[a]	Length (nt)	[Na+]/[Mg2+] mM/mM[b]	Mean/Minimal RMSDs (Å)[c]	
				In 1M [Na+]	In Experimental Solution
2GM0	I	70	250/0.1	7.8/3.5	6.1/3.1
2M1O	H	14	80/0.1	2.4/1.1	2.0/0.7
1TUT	I	22	80/3	3.9/2.2	3.2/1.7
2KYD	H	40	150/10	3.8/1.2	3.7/1.2
2D1A	I	78	50/0	7.4/4.1	6.9/3.8
2DD1	I	20	90/0	3.7/1.5	3.3/1.5
2K7E	I	24	110/0	2.8/1.2	2.4/1.3
2LX1	I	22	90/0	5.6/3.3	4.8/2.4
2JXQ	H	20	60/0	2.1/0.8	2.0/0.8
1F5G	I	20	80/0	3.2/1.9	2.8/1.7

[a]H stands for dsRNAs of complementary duplex, and I stands for dsRNAs with internal loop.
[b]The experimental ion conditions for structure determination by NMR method.
[c]The RMSDs are calculated over all three CG beads of predicted structures by our model from the corresponding atoms of the native structures.

be evaluated by the root mean-square deviation (RMSD) values calculated over all the beads in predicted structures from the corresponding atoms in the native structures in the PDB (67). As shown in Figs. 1 *C* and 2, for the dsRNA of PDB: 2JXQ, the mean RMSD (the averaged value over the refined structures) and the minimal RMSD (from the structure closest to the native one) are 2.1 and 0.8 Å, respectively.

Following the above process, the 3D structures of 26 dsRNAs, including 15 dsRNAs with bulge/internal loops, were predicted by this model with an overall mean RMSD of ~3.3 Å and an overall minimal RMSD of ~1.8 Å; see Figs. 2, 3, and Table S1. This shows that our model with coaxial/base stacking can reliably capture the 3D shapes of various dsRNAs, including those with bulge/internal loops. For 13 dsRNAs with internal loops, the overall mean RMSD is ~4.2 Å, which is slightly larger than that (~3.3 Å) of all the 26 dsRNAs. This is because large internal loops generally contain noncanonical basepairs, which is ignored in this model. For example, the dsRNA of PDB: 3WBM contains two internal loops with several noncanonical basepairs to keep the helix more continuous than the predicted one; see Fig. 2 *C*.

Structure predictions for dsRNAs in respective ion solutions

Because RNA structures can be strongly influenced by ions (10–14), we introduced the structure-based electrostatic potential in our model to improve the 3D structure prediction for dsRNAs at the respective ion conditions. In the following, we first examined the structure-based electrostatic potential through the charge-neutralization fractions f^{Na} and f^{Mg} of Na^+ and Mg^{2+} along with an example dsRNA (PDB: 2GM0) in mixed Na^+/Mg^{2+} solutions. As shown in Fig. 3 *B*, f^{Na} and f^{Mg} appear dependent on Na^+/Mg^{2+} concentrations as well as dsRNA structures: 1) as Mg^{2+} concentration increases, f^{Mg} increases and f^{Na} decreases because of the competition between Mg^{2+} and Na^+ in binding to an RNA and the lower binding-entropy penalty for Mg^{2+} at higher Mg^{2+} concentration (10–14,59–63); 2) f^{Na} and f^{Mg} are larger at bent regions and appear less at two ends, which is attributed to the higher charge density of P beads at bending regions and lower charge density of P beads at two ends of the dsRNA. Therefore, the newly refined electrostatic potential (Eqs. 4, 5, 6, 7, and 8) can capture the structure-based ion binding and the competitive binding between Na^+ ions and Mg^{2+} ions to dsRNAs.

To examine whether the involvement of the implicit structure-based electrostatic potential (Eqs. 4, 5, 6, 7, and 8) and corresponding ion conditions can improve 3D structure prediction for dsRNAs, we further predicted the 3D structures for 10 dsRNAs in the NMR set at their respective experimental ion conditions. For the 10 dsRNAs, the overall mean RMSD between predicted structures in ion solutions and the native structures is ~3.7 Å, which is visibly smaller than that (~4.3 Å) of the predictions at 1 M [Na^+]; see Table 1. This suggests that the inclusion of structure-based electrostatic potential can account for ionic conditions for RNAs and the implementation of the corresponding experimental ion conditions can improve the predictions on 3D

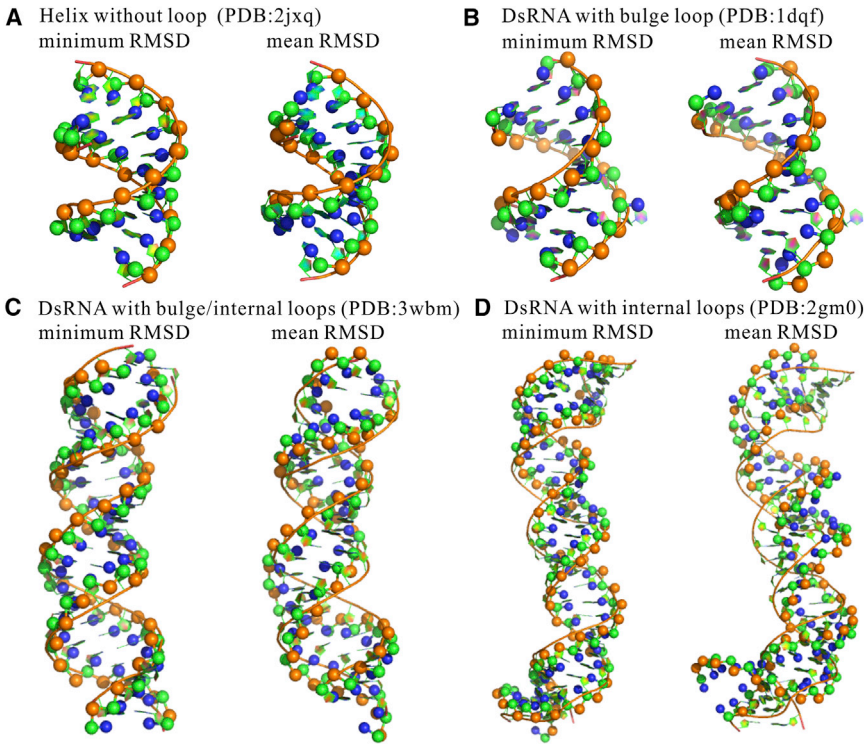

FIGURE 2 The predicted 3D structures (*ball-stick*) with mean/minimal RMSDs in comparison with the corresponding native structures (*cartoon*) for four typical dsRNAs. (*A*) dsRNA helix without loop (PDB: 2JXQ) with mean/minimal RMSDs of 2.0 Å/0.8 Å. (*B*) dsRNA containing a bulge loop (PDB: 1DQF) with mean/minimal RMSDs of 2.5 Å/1.6 Å. (*C*) dsRNA containing two bulge loops and two internal loops (PDB: 3WBM) with mean/minimal RMSDs of 5.4 Å/2.3 Å. (*D*) dsRNA containing four internal loops (PDB: 2GM0) with mean/minimal RMSDs of 6.1 Å/3.1 Å. The structures are shown with PyMol (http://www.pymol.org). To see this figure in color, go online.

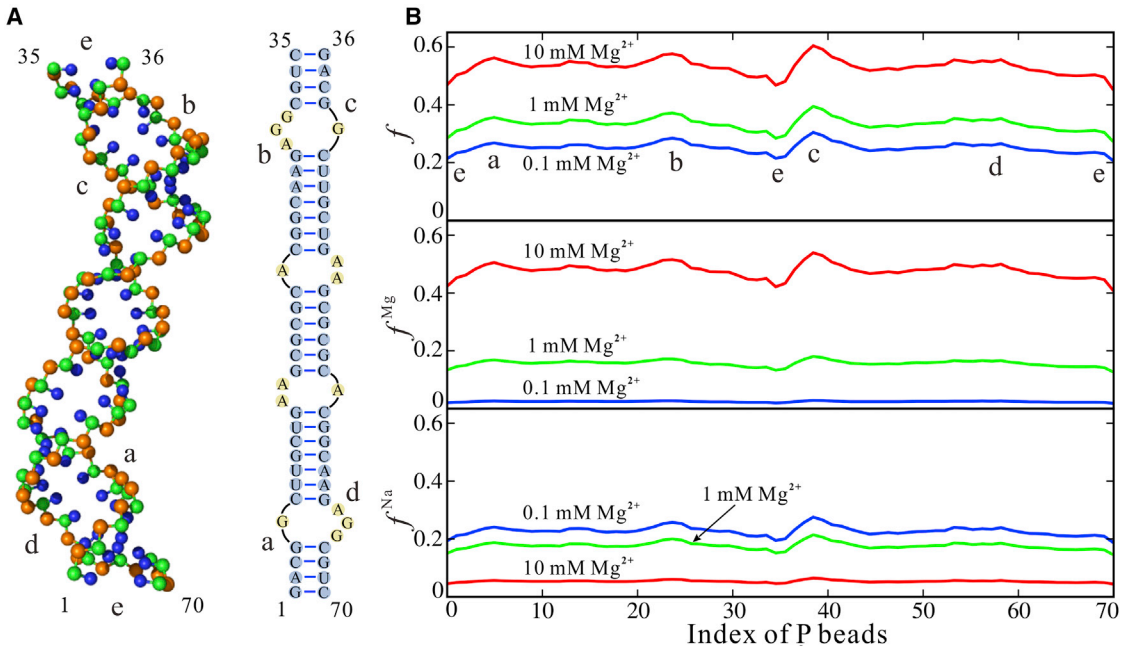

FIGURE 3 (*A*) The predicted 3D structure (*left*, shown with PyMol) and secondary structure (*right*) of dsRNA (PDB: 2GM0). (*B*) The calculated ion-charge-neutralization fractions along P beads of the predicted structure of the dsRNA at different ion conditions; see Eq. 6. f, the total ion-neutralization fractions (*top*); f^{Mg}, neutralization fractions of Mg^{2+} (*middle*); f^{Na}, neutralization fractions of Na^{+} (*bottom*). The blue, green, and red lines denote the cases of the dsRNA in 150 mM Na^{+} solutions mixed with 0.1, 1, and 10 mM Mg^{2+}, respectively. The two bent regions labeled with (*a*, *d*) and (*b*, *c*) induced by the internal loops correspond to the ion-neutralization-fraction peaks in (*B*), and two helical ends of the dsRNA labeled with (*e*) correspond to the ion-neutralization-fraction troughs in (*B*). To see this figure in color, go online.

structures for dsRNAs in ion solutions. Furthermore, Table 1 also shows that such improvement appears more pronounced for the dsRNAs with internal loops and for the ion conditions containing Mg^{2+}, e.g., the mean RMSDs of the dsRNA of 2GM0 and 1TUT decrease from 7.8 and 3.9 Å to 6.1 and 3.2 Å, respectively. This indicates that the newly refined electrostatic potential can effectively involve the RNA structure information and the effect of ions such as Mg^{2+}. In addition, we have further predicted 3D structures for dsRNAs in NMR set using our model in the previous version and for several ssRNAs including hairpins and pseudoknots involved in (55,68) using our model; see Table S4. The comparisons between predictions from our model and its previous version show a visible improvement (~0.4 Å in mean RMSD) on predicting RNA structures, including both dsRNAs and ssRNAs (55,69). This indicates that the inclusion of the newly refined structure-based electrostatic potential can better capture RNA structures at their respective ion conditions.

Comparisons with other models

To further examine our model, we made extensive comparisons with three existing RNA-structure-prediction models: FARNA (31), RACER (45,46), and MC-Fold/MC-Sym pipeline (24). FARNA is a fragment-assembly model with high resolution for small RNAs (31). As shown in Fig. 4 *A*, the average mean RMSD (~3.9 Å) from our model is very slightly smaller than that (~4.1 Å) from FARNA. Afterwards, we made the comparison with a newly developed CG model, RACER (45). As shown in Fig. 4 *B*, the mean RMSD of the predictions from our model (~2.6 Å) is slightly smaller than that (~3.2 Å) from RACER. Furthermore, we made the extensive comparison with MC-Fold/MC-Sym pipeline (24), a well-established RNA 2D/3D structure prediction model with a web server (http://www.major.iric.ca/MC-Pipeline). We used the web server of MC-Fold/MC-Sym pipeline to predict the structures of all the dsRNAs involved in our structure prediction and chose the top predicted structure to make comparisons with our model. As shown in Fig. 4 *C*, the average mean RMSD (~3.3 Å) of the predictions from our model is slightly smaller than that (~3.8 Å) of the top structure from MC-Fold/MC-Sym pipeline. Therefore, the above comparisons show that our model can be reliable in predicting the 3D structures of dsRNAs. Beyond 3D structure predictions, our model can also predict the stability and flexibility of dsRNAs in ion solutions. It should be noted that the computational cost of our model is comparable with that of RACER (46) but much higher than that of the fragment-assembly methods such as MC-Fold/MC-Sym (24) because our model and RACER both involved conformation sampling, unlike fragment-assembly methods. For example, for dsRNA with 20 nt, it takes ~3 h for our model to predict its 3D structure with one thread on the early generation Intel i7 central processing unit and takes ~10 min by MC-Fold/MC-Sym on its server. However, compared with the existing

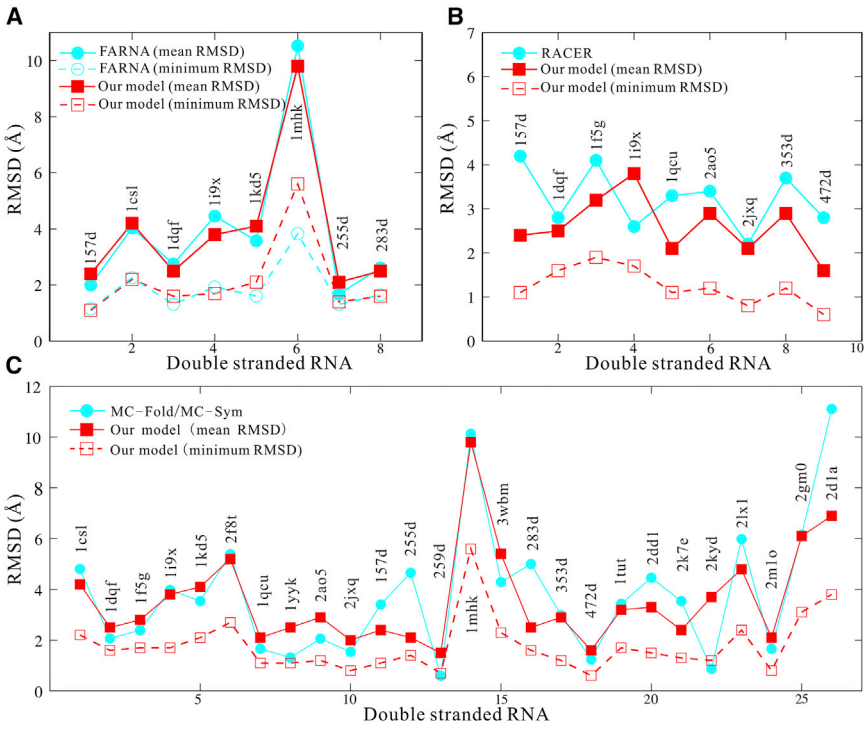

FIGURE 4 The comparisons of the predicted 3D structures between our model and the existing models: (*A*) FARNA (31), (*B*) RACER (45,46), and (*C*) MC-Fold/MC-Sym pipeline (24). The RMSDs of structures predicted by FARNA are calculated over the C4′ atom (31). The RMSDs of structures predicted by RACER, MC-Fold/MC-Sym pipeline, and our model are calculated over all CG beads (24,45). The data of FARNA and RACER are taken from (31) to (45), respectively. To see this figure in color, go online.

models, our model can also predict the 3D structures at different temperatures and can estimate thermodynamic stability for dsRNAs simultaneously.

Stability of dsRNAs in ion solutions

Stability of dsRNAs with various sequences

As described in Methods, for dsRNA with a given strand concentration, the melting curve as well as the melting temperature T_m can be calculated with our model. For example, for the sequence $(CGCG)_2$, the melting curve of the dsRNA with a high strand concentration of 10 mM can be predicted based on the fractions of unfolded state at different temperatures, and the melting curve as well as the melting temperature T_m of the dsRNA at low experimental strand concentration (0.1 mM) can be calculated through Eqs. 9, 10, and 11; see Fig. 5, *A* and *B*. As shown in Fig. 5, *A* and *B*, the predicted T_m of the sample sequence $(CGCG)_2$ with experimental strand concentration of 0.1 mM

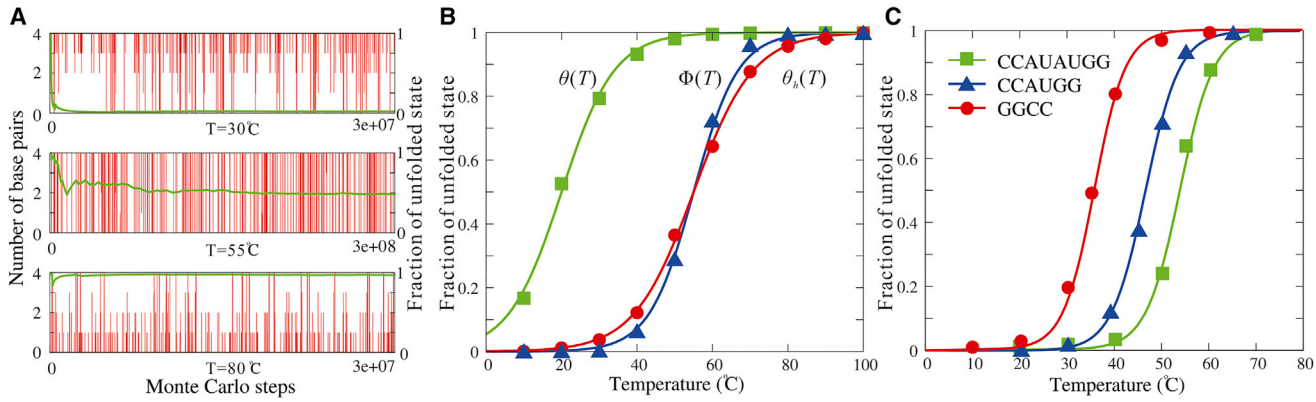

FIGURE 5 (*A*) The time evolution of the number of basepairs (*vertical* or *red lines*) and the average fractions of unfolded state (*transverse* or *green lines*) for the sample dsRNA of $(CGCG)_2$ at 30°C (*top*), 55°C (*middle*), and 80°C (*bottom*). (*B*) The fractions of unfolded state as functions of temperature for the dsRNA of $(CGCG)_2$. Symbols show the predicted data at different temperatures: blue triangle, at high strand concentration (10 mM); red circle, corrected value of $\theta_h(T)$ by Eq. 9; and green square, at experimental strand concentration (0.1 mM) derived by Eq. 10. Lines are the fitted melting curve to the predicted data through Eq. 11. More details can be found in the Supporting Materials and Methods. (*C*) The fractions of unfolded states for three dsRNAs with 0.1 mM strand concentration as functions of temperature. Symbols are the predicted data and lines are the fitted curves for the three dsRNAs through Eq. 11. To see this figure in color, go online.

is ~19.5°C, which agrees well with the corresponding experimental value (~19.3°C). Furthermore, we further predicted the thermodynamic stability of 22 dsRNAs (from 4-bp to 14 bp) with various complementary sequences; see Table 2. Here, dsRNAs are assumed in solutions of 1 M [Na^+] to solely examine the stabilities of dsRNAs of various sequence and make comparisons with extensive experimental data (58,66,70–72). As shown in Table 2, the T_m values of extensive dsRNAs from our model are in good agreement with the corresponding experimental data with the mean deviation ~1.3°C and maximal deviations <2.5°C. Such agreement indicates that the sequence-dependent base-pairing and base-stacking interactions in our model can capture the stability of dsRNAs of extensive sequences and different lengths well (58,66,70–72).

Thermally unfolding pathways of dsRNAs

Because intermediate states of RNAs can be important to their functions (1–3,10,68,69), we made further analyses of thermally unfolding pathways for different dsRNAs. To distinguish the possible different states of dsRNAs at different temperatures in our simulations, all the states for dsRNA were more detailedly divided into unfolded state (U, two disassociated single strands), possible hairpin state (H, with at least one hairpin), folded helix state (F, with the formation of all basepairs except for the two end ones), and partially folded helix state (P, other conformations besides U, F, and H states).

As shown in Fig. 6, the unfolding pathways of dsRNAs are dependent on their length as well as sequences. For short sequences ($\leq$6 bp), dsRNAs undergo the standard two-state melting transitions and there are almost no intermediate states such as P and H states (see Fig. 6, *A* and *B*), which is consistent with the previous experiments (66). As chain length increases to ~8 bp, P states begin to appear and can become visible at ~T_m; see Fig. 6, *C* and *D*. Figs. 6, *C* and *D* also show that the unfolding pathways of dsRNAs with the same chain length but different sequences would be slightly different, e.g., the fraction of P state of $(AACUAGUU)_2$ with end A-U basepairs is slightly higher than that of $(CCAUAUGG)_2$ with end G-C basepairs. This is because the unstable end A-U basepairs can induce more notable P states than stable end G-C basepairs; see Fig. S1.

For dsRNAs with more than 10 bp, their thermally unfolding pathways become more complex and interesting because their ss chain may fold to hairpin structures. As shown in Fig. 6, *E* and *F* for the dsRNA of $(CCUUGAUAUCAAGG)_2$ and $(AAAAAAAUUUUUUU)_2$, the fractions of F state are near unity at low temperature. As temperature increases, the dsRNAs begin to melt and H states would form from the melted ssRNAs with the maximal fractions of ~0.2 and ~0.5 for the two dsRNAs, respectively. At higher temperatures, the dsRNAs become almost completely melted as the U state. Notably, as shown in Fig. 6 *F*, the unfolding pathway of $(AAAAAAAUUUUUUU)_2$ predicted by our model is very close to the corresponding experiments (66,71), suggesting that the melting pathways of dsRNAs can be well captured by our analyses with this model. The difference on unfolding pathways between the two dsRNAs is attributed to the different sequences, i.e., G-C content, especially at two ends. Specifically, the fractions of states follow the order of $F > U \gtrsim H > P$ for $(CCUUGAUAUCAAGG)_2$ at ~70°C, whereas for $(AAAAAAAUUUUUUU)_2$ at ~45°C, such an order becomes $H > F \gtrsim U > P$. To reveal what determines the order, we calculated the stability for the states with Mfold (73). We found that the order of state fractions is in agreement with that of state stability. For example, the formation free energies for F, H, and P states are ~−0.5, ~0.2, and ~1.5 kcal/mol for $(CCUUGAUAUCAAGG)_2$ at ~70°C, respectively. For $(AAAAAAAUUUUUUU)_2$ at ~45°C, the formation free energies for H, F, and P states are ~−0.8, ~−0.1, and ~0.6 kcal/mol. This indicates that the unfolding pathway of dsRNA is dependent on the stability of possible states.

Although unfolding of long dsRNAs can be non-two-state transitions, in analogy to experiments (66), we can still estimate their melting temperatures by assuming strands completely disassociated state as U state (66); see Eq. 11 in Methods.

TABLE 2 The Melting Temperature T_m Values for 20 dsRNAs in 1 M [Na^+] Solution

	Sequence[a]	Length (bp)	Expt. (°C)[d]	Pred. (°C)	Deviation (°C)
1	CGCG	4	19.9	19.5	−0.4
2	GGCC	4	34.3	35.1	+0.8
3	GCGC	4	26.6	28.1	+1.5
4	CCGG	4	27.2	26.6	−0.6
5	CGCGCG	6	57.8	58.5	+0.7
6	CCGCGG	6	59.8	61.2	+1.4
7	CCAUGG	6	46.4	45.3	−1.1
8	GUCGAC	6	45.3	46.8	+1.5
9	AGCGCU	6	50.2	51.1	+0.9
10	UCAUGA	6	27.2	25.7	−1.5
11	GAGGAG	6	50.9	52.0	+1.1
12	AACUAGUU	8	45.7	45.5	−0.2
13	ACUUAAGU	8	40.3	41.3	+1.0
14	ACCUUUGG	8	56.3	54.8	−1.5
15	GCCAUGGC	8	71.4	69.5	−1.9
16	GCUGCGAC	8	67.9	65.8	−2.1
17	CCAUAUGG	8	57.0	59.1	+2.1
18	GAACGUUC	8	52.3	53.6	+1.3
19	AAGGUUGGAA	10	66.5	64.1	−2.4
20	AUUGGAUACAAA[b]	12	55.4	58.1	+2.7
21	AAAAAAAUUUUUUU[c]	14	33.2	35.0	+1.8
22	CCUUGAUAUCAAGG	14	76.7	74.3	−2.4

[a]There is only one sequence for dsRNA, and the other is complementary to the shown one. The strand concentrations are 0.1 mM.
[b]The strand concentration of AUUGGAUACAAA is 8 μM.
[c]The strand concentration of AAAAAAAUUUUUUU is 3.9 μM.
[d]The experimental data are from (66,70–72).

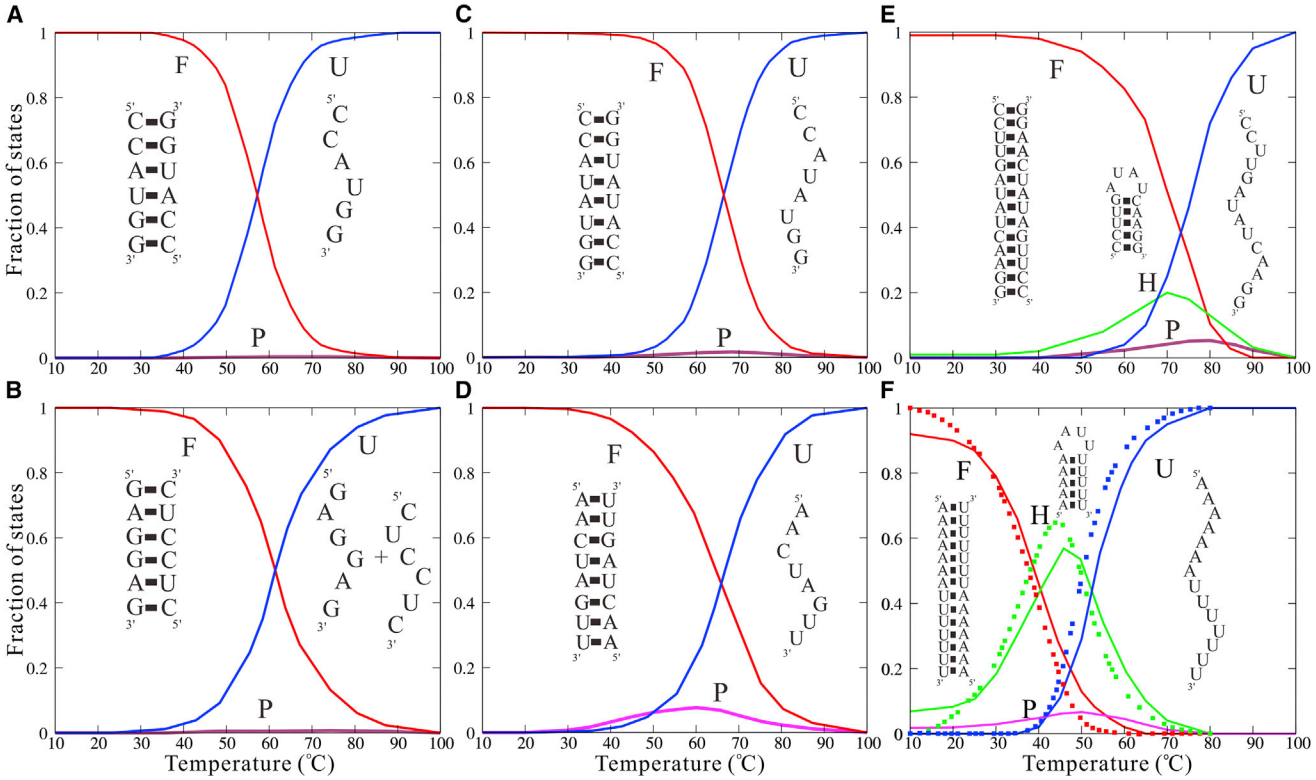

FIGURE 6 The fractions of F (*folded*, *red lines*), P (*partially folded*, *pink lines*), H (*hairpin*, *green lines*), and U (*unfolded*, *blue lines*) states as functions of temperature for the unfolding of $(CCAUGG)_2$ (*A*), (GAGGAG, CUCCUC) (*B*), $(CCAUAUGG)_2$ (*C*), $(AACUAGUU)_2$ (*D*), $(CCUUGAUAUCAAGG)_2$ (*E*), and $(AAAAAAAUUUUUUU)_2$ (*F*). The full symbols in (*F*) are the corresponding experimental data from (71). The insets illustrate the typical predicted secondary structures for F, U, and H states. To see this figure in color, go online.

Stability of dsRNAs with bulge/internal loops

Beyond the dsRNAs with complementary sequences shown above, the stability of other eight dsRNAs with bulge/internal loops was examined by our model. As shown in Table 3, for the dsRNAs with single/double-bulge loops of different loop lengths (sequences 1–6) and the dsRNAs with internal loops (sequences seven and eight), the mean deviation between the predicted T_m values and the experimental data (74–80) is ~2.6°C, which indicates that our model with the coaxial stacking potential can roughly estimate the stability of dsRNAs with bulge/internal loops. However, such predictions, especially for dsRNAs with long bulge/internal loops, are not as precise as those for dsRNAs without loops. The detailed comparisons with experimental data show that the predicted T_m values for the dsRNAs with a one-nucleotide (nt) bulge loop are slightly higher than experimental data, whereas our model underestimates the stability of dsRNAs with longer bulge loops, which may suggest that the coaxial stacking potential U_{cs} involved in our model may slightly overestimate the coaxial interaction strength while underestimating the coaxial interaction range. For the dsRNAs with an internal loop (e.g., AA/AA

TABLE 3 The Melting Temperatures T_m for Eight dsRNAs with Bulge/Internal Loop in 1 M Na^+ Solution

	Sequence[a]	Description[b]	Experiment (°C)[c]	Prediction (°C)	Deviation (°C)
1	UGACGCUCA, ACUG GAGU	B	42.2	45.0	+2.8
2	GACUAUGUC, CUGA ACAG	B	34.2	36.1	+1.9
3	GACUAGUGUC, CUGA ACAG	B	31.9	29.5	−2.4
4	GACUCAUCCUG, CUGG GGAC	B	40.7	37.9	−2.8
5	GCAGUUCACG, CG CAAG GC	B	35.5	36.8	+1.3
6	CGAGUAC CG, GC CAUGAGC	B	43.6	41.5	−2.1
7	CGC AA GCG, GCG AA CGC	I	35.5	33.2	−2.3
8	CGCAAAGGC, GCGAAACCG	I	38.5	33.1	−5.4

[a]The strand concentrations are 0.1 mM.
[b]B stands for dsRNA with bulge loop, and I stands for dsRNA with internal loop.
[c]The experimental data are from (74–80).

or AAA/AAA), our model underestimates their stability, which may be attributed to the ignorance of noncanonical basepairs in this model (80).

Effects of monovalent and divalent ions

The thermal stability of RNA molecules is generally sensitive to the ionic conditions (10–14,59–63). Particularly, Mg^{2+} is efficient in neutralizing the negative charges on the RNA molecule and generally plays an important role in RNA folding (14,59–63,81–84). However, most of the existing structure prediction models cannot quantitatively predict the stability of dsRNAs in ion solutions, especially in the presence of Mg^{2+}. Here, we employed our model to examine the stability for dsRNAs over a wide range of monovalent and divalent ion concentrations.

First, we examined the effect of monovalent ions on the stability of dsRNAs. As shown in Fig. 7 *A*, for five dsRNAs with different sequences and lengths, the predicted melting temperature T_m values from our model agree well with the experimental data (58,66,70–72,84), with a mean deviation <2°C over the wide range of $[Na^+]$. As $[Na^+]$ increases from 10 mM to 1 M, T_m values of the dsRNAs obviously increase, which is attributed to lower ion-binding entropy penalty and stronger ion neutralization for basepair formation at higher $[Na^+]$; see also Fig. S1. Furthermore, Fig. 7 *A* shows that the $[Na^+]$-dependence of T_m is stronger for longer dsRNAs. This is because basepair formation of longer dsRNAs causes larger buildup of negative charges and consequently causes stronger $[Na^+]$-dependent ion binding.

Second, we examined the stability of dsRNAs in mixed monovalent and divalent ion solutions. As shown in Fig. 7 *B*, for three different dsRNAs, the predicted T_m values are in good accordance with the experimental data over the wide range of $[Mg^{2+}]$ (84). Fig. 7 *B* also shows that there are three ranges in T_m-$[Mg^{2+}]$ curves: 1) at low $[Mg^{2+}]$ (relatively to $[Na^+]$), the stability of the dsRNAs is dominated by the background $[Na^+]$ and T_m values of the dsRNAs are almost the same as the corresponding pure $[Na^+]$; 2) with the increase of $[Mg^{2+}]$, Mg^{2+} ions begin to play a role and T_m increases correspondingly; and 3) when $[Mg^{2+}]$ becomes very high (relatively to $[Na^+]$), the stability is dominated by Mg^{2+}. Furthermore, it is shown that Mg^{2+} is very efficient in stabilizing dsRNAs. Even in the background of 110 mM Na^+, ~1 mM Mg^{2+} begins to enhance the stability of dsRNAs, and 10 mM Mg^{2+} (+110 mM background Na^+) can achieve the similar stability to 1 M Na^+ for dsRNAs; see sequences of CCUUGAUAUCAAGG and CCAUAUGG in Fig. 7, *A* and *B*. This is attributed to the high ionic charge of Mg^{2+} and the consequent efficient role in stabilizing dsRNAs (58–63,81–84).

Flexibility of dsRNA in ion solutions

DsRNAs generally are rather flexible in ion solutions because of their polymeric nature, and the flexibility is extremely important for their biological functions. Additionally, dsRNA flexibility is highly dependent on solution ion conditions (85–93). In this section, we further employed our model to examine the flexibility of a 40-bp RNA helix in ion solutions. The sequence of the dsRNA helix, which is selected according to a previous study (91), is 5′-CGA CUCUACGGAAGGGCAUCCUUCGGGCAUCACUACG CGC-3′, with 57% CG content in its central 30-bp segment and the other chain being fully complementary to it. First, we predicted the 3D structures for the dsRNA helix from the sequence at 25°C, and afterwards, we performed further simulations for the dsRNA helix at various ion conditions based on the predicted structures. Enough conformations at equilibrium at each ion condition were used to analyze

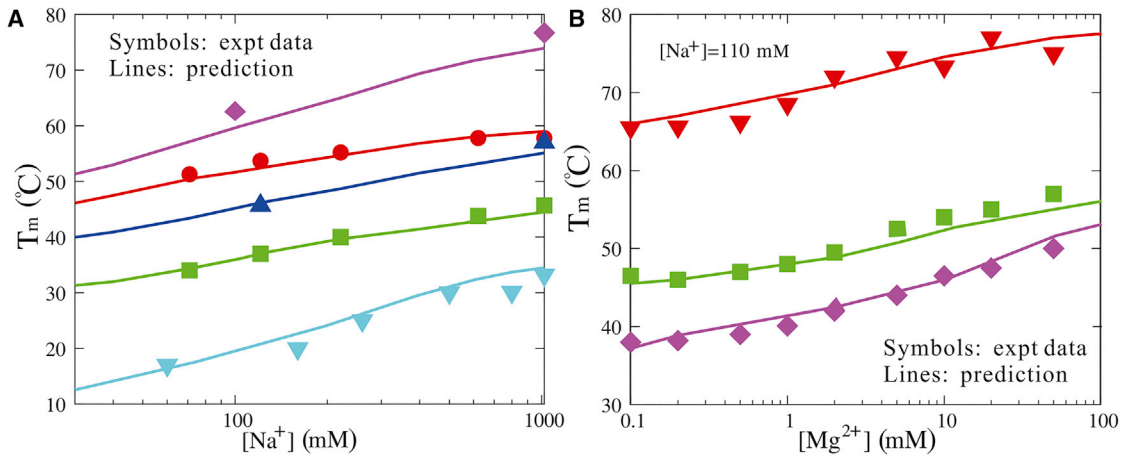

FIGURE 7 (*A*) The predicted melting temperatures T_m and corresponding experimental data (58,66,70–72,84) as functions of $[Na^+]$ for sequences CCUUGAUAUCAAGG, CGCGCG, CCAUAUGG, AACUAGUU, and AAAAAAAUUUUUUU (from *top* to *bottom*). (*B*) The predicted melting temperatures T_m and corresponding experimental data (84) as functions of $[Mg^{2+}]$ for sequences CCUUGAUAUCAAGG, CCAUAUGG, and CCAUGG (from *top* to *bottom*). Note that the solutions contain 110 mM Na^+ as background (84). To see this figure in color, go online.

the salt-dependent flexibility of the dsRNA helix; see Fig. S2.

Structure fluctuation of dsRNA in ion solutions

In the following, we first examined the structure fluctuation of the dsRNA helix through calculating end-to-end distance, RMSD variance, and root mean-square fluctuation (RMSF) at different [Na^+] values (85). As [Na^+] increases, the end-to-end distance of the dsRNA helix decreases, e.g., from ~125 Å at 10 mM [Na^+] to ~90 Å at 1 M [Na^+], and simultaneously, the variance of end-to-end distance increases; see Fig. 8, *A* and *B*. This indicates the stronger bending conformations and the higher bending fluctuation for the dsRNA helix at higher ion concentrations, which are attributed to the stronger ion neutralization on P bead charges and consequently the reduced electrostatic repulsion due to bending (86–94). The RMSD variance of the dsRNA helix at different [Na^+] values calculated based on the conformation-averaged reference structure also indicates that the dsRNA helix would become more flexible with the increase of [Na^+]; see Fig. 8 *C*. To examine local structure fluctuation, we further calculated the RMSF of the centers of each basepair of the dsRNA helix at different [Na^+] values. As shown in Fig. 8 *D*, the RMSF increases as [Na^+] increases from 0.01 to 1 M, which is because the stronger ion binding and charge neutralization on P beads enable the larger fluctuation of basepairs along the helix (94). Additionally, end effect contributes to an extra increase of RMSF at the two helical ends (94).

Persistence length of dsRNA helix in ion solutions

Generally, the flexibility of a polymer can be described by its persistence length l_p (95), and l_p can be calculated by the following (96):

$$\cos\theta_i = \exp\left(-\frac{ib}{l_p}\right), \tag{12}$$

where $\cos\theta_i = \hat{r}_i \cdot \hat{r}_0$ and $\hat{r}_0$ and $\hat{r}_i$ are the first and i-th bond-direction vectors, respectively. b in Eq. 12 is the average bond length. According to Eq. 12, l_p of the dsRNA helix can be obtained through modeling the dsRNA helix as a bead chain composed of the central beads of basepairs; see Fig. 8 *E*. To avoid the end effect (94), the first and last five basepairs were excluded in our calculations and the bond vectors are selected as those over every five continuous basepairs (91).

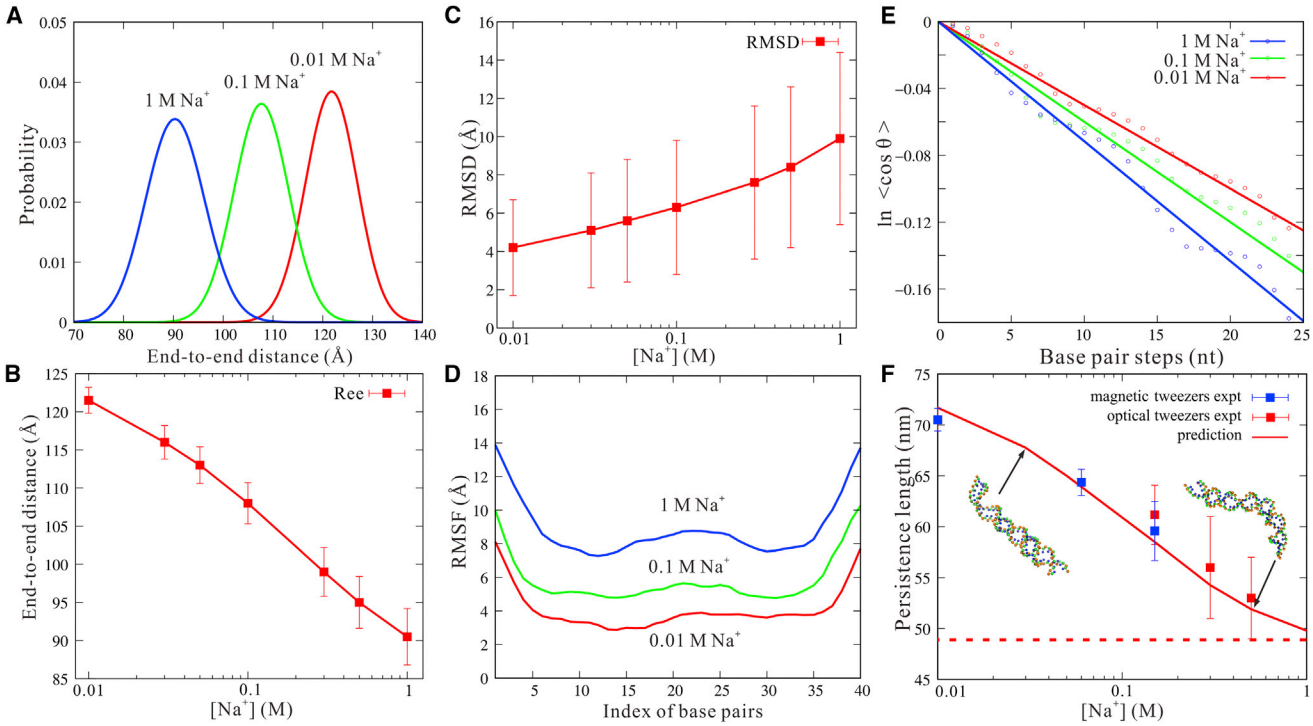

FIGURE 8 (*A*) The distributions of end-to-end distance for the 40-bp dsRNA helix at 1, 0.1, and 0.01 M [Na^+], respectively. (*B*) The mean end-to-end distances of the 40-bp dsRNA helix as a function of [Na^+]. Here, the error bars denote the variance for the end-to-end distances. (*C*) The RMSD of the 40-bp dsRNA helix calculated based on the conformation-averaged reference structure of the respective simulation as a function of [Na^+]. Here, the error bars denote the variances for the RMSDs. (*D*) The RMSF of basepairs along the 40-bp dsRNA helix from simulated ensembles at 1, 0.1, and 0.01 M [Na^+] values. (*E*) The fitting curves for l_p through Eq. 12 for the 40-bp dsRNA helix at 1, 0.1, and 0.01 M [Na^+] values. (*F*) The predicted persistence length l_p (*lines*) for the 40-bp dsRNA helix as a function of [Na^+]. Blue squares, experimental data from magnetic tweezers method (93); red squares, experimental data from optical tweezers method (93). The dashed line in (*F*) shows the predicted l_p assuming all P beads are electrically neutral. To see this figure in color, go online.

As shown in Fig. 8 *F*, the persistence lengths of the 40-bp dsRNA helix at different [Na^+] values predicted by our model are in quantitative agreement with the corresponding experimental data (93). For example, the deviation of l_p between prediction and experiments is less than ~2 nm over the wide range of [Na^+]. As [Na^+] increases from 0.01 to 1 M, l_p of the dsRNA helix decreases from ~70 to ~50 nm. This is because more binding ions neutralize the negative P bead charges on the dsRNA helix more strongly and can reduce the electrostatic bending repulsion along the strands more strongly, causing stronger bending flexibility at high [Na^+]. Additionally, we have taken out the representative structures at 0.1 mM [Na^+] and have calculated the ion-charge-neutralization fractions along the structures. As shown in Fig. S3, the ion-charge-neutralization fraction changes as RNA structure is changed, especially in the bending regions.

CONCLUSIONS

Knowledge of the 3D structures and thermodynamic properties of dsRNAs are crucial for understanding their biological functions. In this work, we have further developed our previous CG model by introducing a structure-based electrostatic potential and employed the model to predict 3D structures, stability, and flexibility of dsRNAs in monovalent/divalent ion solutions. Our predictions were extensively compared with experimental data, and the following conclusions have been obtained:

1) Our model can predict 3D structures from sequences for extensive dsRNAs with/without bulge/internal loops in monovalent/divalent ion solutions well with overall mean RMSD <3.5 Å, and the involvement of the structure-based electrostatic potential and corresponding experimental ion conditions generally improves the structure predictions with smaller RMSDs for dsRNAs in ion solutions.
2) Our model can make good predictions for the stability for dsRNAs with extensive sequences over wide ranges of monovalent/divalent ion concentrations with mean deviation <2°C, and our analyses show that the thermally unfolding pathway of dsRNA is dependent on its length as well as its sequence.
3) Our model can well capture the salt-dependent flexibility of dsRNAs, and the predicted salt-dependent persistence lengths are in good accordance with experiments.

Although our predictions agree well with the extensive experimental data on 3D structure, stability and flexibility of dsRNAs, there are still several limitations in our model. First, despite the fact that the structure-based electrostatic potential can efficiently capture the effects of monovalent/divalent ions on the structure, stability, and flexibility of dsRNAs, the model was not examined for RNAs with more complex structures, and the model cannot consider concrete ion distribution and specific ion binding around an RNA. Further development of this model may need to involve the effect of ions through an implicit-explicit combined treatment for ions (54). Second, the model only involves canonical and wobble basepairs (A-U, G-C, and G-U) and ignores noncanonical basepairs because over ~90% basepairs in the ds stems of natural RNAs are canonical ones (97). However, noncanonical basepairs are usually found in RNAs of complex structures, e.g., in interhelical junction loops (98), and for those dsRNAs with bulge/internal loops, the ignorance of noncanonical basepairs may (only) slightly affect the predictions on the structure and stability (97). The noncanonical basepairs can also be possibly involved in the model for predicting RNAs of complex structures at the level of three CG beads through two treatments: 1) derive the (local) bond energy parameters for noncanonical basepairs based on the corresponding structures in the PDB and involve them in Eqs. S3–S5 (54); or 2) given Turner's nearest-neighbor thermodynamic parameters for noncanonical base stacking (99), derive the basepair/stacking parameters for noncanonical basepairs based on the corresponding structures in the PDB (54) and involve these parameters in Eqs. S7–S9. Finally, our model is a CG model, and it is still necessary to rebuild all-atom structures based on predicted CG ones. Nevertheless, our model can predict the 3D structures, stability, and flexibility of dsRNAs over the wide ranges of monovalent/divalent ion concentrations well and can be a good basis for further development for a predictive model with higher accuracy.

SUPPORTING MATERIAL

Supporting Materials and Methods, three figures, and four tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)31006-3.

AUTHOR CONTRIBUTIONS

Z.-J.T., L.J., and Y.-Z. S. designed the research. L.J., Y.-Z.S., and C.-J.F. performed the simulation. Z.-J.T., L.J., and Y.-L.T. analyzed the data. L.J., Y.-Z.S., and Z.-J.T. wrote the manuscript. All authors discussed the result and reviewed the manuscript.

ACKNOWLEDGMENTS

We are grateful to Professors Shi-Jie Chen (University of Missouri), Xiangyun Qiu (George Washington University), Jian Zhang (Nanjing University) and Wenbing Zhang (Wuhan University) for valuable discussions.

This work was supported by grants from the National Science Foundation of China (11575128, 11605125, and 11774272). Parts of the numerical calculation in this work were performed on the supercomputing system in the Supercomputing Center of Wuhan University.

REFERENCES

1. Watson, J. D. 2008. Molecular Biology of the Gene. Pearson/Benjamin Cummings, San Francisco, CA.

2. Tinoco, I., Jr., and C. Bustamante. 1999. How RNA folds. *J. Mol. Biol.* 293:271–281.
3. Li, P. T., J. Vieregg, and I. Tinoco, Jr. 2008. How RNA unfolds and refolds. *Annu. Rev. Biochem.* 77:77–100.
4. Kuwabara, T., J. Hsieh, …, F. H. Gage. 2004. A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell.* 116:779–793.
5. Hannon, G. J. 2002. RNA interference. *Nature.* 418:244–251.
6. Meister, G., and T. Tuschl. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature.* 431:343–349.
7. Akira, S., and K. Takeda. 2004. Toll-like receptor signalling. *Nat. Rev. Immunol.* 4:499–511.
8. Chen, S. J. 2008. RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* 37:197–214.
9. Mustoe, A. M., C. L. Brooks, and H. M. Al-Hashimi. 2014. Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* 83:441–466.
10. Draper, D. E., D. Grilley, and A. M. Soto. 2005. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* 34:221–243.
11. Lipfert, J., S. Doniach, …, D. Herschlag. 2014. Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* 83:813–841.
12. Woodson, S. A. 2005. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* 9:104–109.
13. Draper, D. E. 2013. Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water. *Biopolymers.* 99:1105–1113.
14. Koculi, E., C. Hyeon, …, S. A. Woodson. 2007. Charge density of divalent metal cations determines RNA stability. *J. Am. Chem. Soc.* 129:2676–2682.
15. Rose, P. W., A. Prlić, …, S. K. Burley. 2017. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45:D271–D281.
16. Sim, A. Y., P. Minary, and M. Levitt. 2012. Modeling nucleic acids. *Curr. Opin. Struct. Biol.* 22:273–278.
17. Miao, Z., and E. Westhof. 2017. RNA Structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* 46:483–503.
18. Schlick, T., and A. M. Pyle. 2017. Opportunities and challenges in RNA structural modeling and design. *Biophys. J.* 113:225–234.
19. Sun, L. Z., D. Zhang, and S. J. Chen. 2017. Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annu. Rev. Biophys.* 46:227–246.
20. Somarowthu, S. 2016. Progress and current challenges in modeling large RNAs. *J. Mol. Biol.* 428:736–747.
21. Shi, Y. Z., Y. Y. Wu, …, Z. J. Tan. 2014. RNA structure prediction: progress and perspective. *Chin. Phys. B.* 23:078701.
22. Cragnolini, T., P. Derreumaux, and S. Pasquali. 2015. Ab initio RNA folding. *J. Phys. Condens. Matter.* 27:233102.
23. Zhou, H. X. 2014. Theoretical frameworks for multiscale modeling and simulation. *Curr. Opin. Struct. Biol.* 25:67–76.
24. Parisien, M., and F. Major. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* 452:51–55.
25. Zhao, Y., Y. Huang, …, Y. Xiao. 2012. Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* 2:734.
26. Wang, J., Y. Zhao, …, Y. Xiao. 2015. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res.* 43:e63.
27. Wang, J., K. Mao, …, Y. Xiao. 2017. Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res.* 45:6299–6309.
28. Popenda, M., M. Szachniuk, …, R. W. Adamiak. 2012. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 40:e112.
29. Cao, S., and S. J. Chen. 2011. Physics-based de novo prediction of RNA 3D structures. *J. Phys. Chem. B.* 115:4216–4226.
30. Xu, X., P. Zhao, and S. J. Chen. 2014. Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One.* 9:e107504.
31. Das, R., and D. Baker. 2007. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. USA.* 104:14664–14669.
32. Hyeon, C., and D. Thirumalai. 2011. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* 2:487.
33. Jonikas, M. A., R. J. Radmer, …, R. B. Altman. 2009. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA.* 15:189–199.
34. Boudard, M., D. Barth, …, J. Cohen. 2017. GARN2: coarse-grained prediction of 3D structure of large RNA molecules by regret minimization. *Bioinformatics.* 33:2479–2486.
35. Kim, N., C. Laing, …, T. Schlick. 2014. Graph-based sampling for approximating global helical topologies of RNA. *Proc. Natl. Acad. Sci. USA.* 111:4079–4084.
36. Jain, S., and T. Schlick. 2017. F-RAG: generating atomic coordinates from RNA graphs by fragment assembly. *J. Mol. Biol.* 429:3587–3605.
37. Zhang, J., Y. Bian, …, W. Wang. 2012. RNA fragment modeling with a nucleobase discrete-state model. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85:021909.
38. Bian, Y., J. Zhang, …, W. Wang. 2015. Free energy landscape and multiple folding pathways of an H-type RNA pseudoknot. *PLoS One.* 10:e0129089.
39. Li, J., J. Zhang, …, W. Wang. 2016. Structure prediction of RNA loops with a probabilistic approach. *PLoS Comput. Biol.* 12:e1005032.
40. Uusitalo, J. J., H. I. Ingólfsson, …, I. Faustino. 2017. Martini coarse-grained force field: extension to RNA. *Biophys. J.* 113:246–256.
41. Sieradzan, A. K., M. Makowski, …, A. Liwo. 2017. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. I. Backbone potentials of coarse-grained polypeptide chains. *J. Chem. Phys.* 146:124106.
42. Ding, F., S. Sharma, …, N. V. Dokholyan. 2008. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA.* 14:1164–1173.
43. Boniecki, M. J., G. Lach, …, J. M. Bujnicki. 2016. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* 44:e63.
44. Cragnolini, T., P. Derreumaux, and S. Pasquali. 2013. Coarse-grained simulations of RNA and DNA duplexes. *J. Phys. Chem. B.* 117: 8047–8060.
45. Xia, Z., D. R. Bell, …, P. Ren. 2013. RNA 3D structure prediction by using a coarse-grained model and experimental data. *J. Phys. Chem. B.* 117:3135–3144.
46. Bell, D. R., S. Y. Cheng, …, P. Ren. 2017. Capturing RNA folding free energy with coarse-grained molecular dynamics simulations. *Sci. Rep.* 7:45812.
47. Denesyuk, N. A., and D. Thirumalai. 2013. Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B.* 117: 4901–4911.
48. Hori, N., N. A. Denesyuk, and D. Thirumalai. 2016. Salt effects on the thermodynamics of a frameshifting RNA pseudoknot under tension. *J. Mol. Biol.* 428:2847–2859.
49. Šulc, P., F. Romano, …, A. A. Louis. 2014. A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.* 140:235102.
50. He, Y., M. Maciejczyk, …, A. Liwo. 2013. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys. Rev. Lett.* 110:098101.
51. He, Y., A. Liwo, and H. A. Scheraga. 2015. Optimization of a nucleic acids united-RESidue 2-point model (NARES-2P) with a maximum-likelihood approach. *J. Chem. Phys.* 143:243111.

52. Manning, G. S. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* 11:179–246.

53. Hayes, R. L., J. K. Noel, …, J. N. Onuchic. 2015. Generalized Manning condensation model captures the RNA ion atmosphere. *Phys. Rev. Lett.* 114:258105.

54. Shi, Y. Z., F. H. Wang, …, Z. J. Tan. 2014. A coarse-grained model with implicit salt for RNAs: predicting 3D structure, stability and salt effect. *J. Chem. Phys.* 141:105102.

55. Shi, Y. Z., L. Jin, …, Z. J. Tan. 2015. Predicting 3D structure, flexibility, and stability of RNA hairpins in monovalent and divalent ion solutions. *Biophys. J.* 109:2654–2665.

56. Ouldridge, T. E., A. A. Louis, and J. P. Doye. 2010. Extracting bulk properties of self-assembling systems from small simulations. *J. Phys. Condens. Matter.* 22:104102.

57. Borer, P. N., B. Dengler, …, O. C. Uhlenbeck. 1974. Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* 86:843–853.

58. Tan, Z. J., and S. J. Chen. 2007. RNA helix stability in mixed Na^+/Mg^{2+} solution. *Biophys. J.* 92:3615–3632.

59. Tan, Z. J., and S. J. Chen. 2005. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J. Chem. Phys.* 122:44903.

60. Tan, Z. J., and S. J. Chen. 2011. Salt contribution to RNA tertiary structure folding stability. *Biophys. J.* 101:176–187.

61. Tan, Z. J., and S. J. Chen. 2010. Predicting ion binding properties for RNA tertiary structures. *Biophys. J.* 99:1565–1576.

62. Wang, F. H., Y. Y. Wu, and Z. J. Tan. 2013. Salt contribution to the flexibility of single-stranded nucleic acid offinite length. *Biopolymers.* 99:370–381.

63. Xi, K., F. H. Wang, …, Z. J. Tan. 2018. Competitive binding of Mg^{2+} and Na^+ ions to nucleic acids: from helices to tertiary structures. *Biophys. J.* 114:1776–1790.

64. Privalov, P. L., and C. Crane-Robinson. 2018. Translational entropy and DNA duplex stability. *Biophys. J.* 114:15–20.

65. Cao, S., and S. J. Chen. 2006. Free energy landscapes of RNA/RNA complexes: with applications to snRNA complexes in spliceosomes. *J. Mol. Biol.* 357:292–312.

66. Xia, T., J. SantaLucia, Jr., …, D. H. Turner. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry.* 37: 14719–14735.

67. Parisien, M., J. A. Cruz, …, F. Major. 2009. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA.* 15:1875–1885.

68. Wang, Y., S. Gong, …, W. Zhang. 2016. The thermodynamics and kinetics of a nucleotide base pair. *J. Chem. Phys.* 144:115101.

69. Shi, Y. Z., L. Jin, …, Z. J. Tan. 2018. Predicting 3D structure and stability of RNA pseudoknots in monovalent and divalent ion solutions. *PLoS Comput. Biol.* 14:e1006222.

70. Nakano, S., M. Fujimoto, …, N. Sugimoto. 1999. Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res.* 27:2957–2965.

71. Hickey, D. R., and D. H. Turner. 1985. Solvent effects on the stability of A7U7p. *Biochemistry.* 24:2086–2094.

72. Chen, Z., and B. M. Znosko. 2013. Effect of sodium ions on RNA duplex stability. *Biochemistry.* 52:7477–7485.

73. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.

74. Tomcho, J. C., M. R. Tillman, and B. M. Znosko. 2015. Improved model for predicting the free energy contribution of dinucleotide bulges to RNA duplex stability. *Biochemistry.* 54:5290–5296.

75. Crowther, C. V., L. E. Jones, …, M. J. Serra. 2017. Influence of two bulge loops on the stability of RNA duplexes. *RNA.* 23:217–228.

76. Znosko, B. M., S. B. Silvestri, …, M. J. Serra. 2002. Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges. *Biochemistry.* 41:10406–10417.

77. Murray, M. H., J. A. Hard, and B. M. Znosko. 2014. Improved model to predict the free energy contribution of trinucleotide bulges to RNA duplex stability. *Biochemistry.* 53:3502–3508.

78. Chen, G., B. M. Znosko, …, D. H. Turner. 2004. Factors affecting thermodynamic stabilities of RNA 3 x 3 internal loops. *Biochemistry.* 43:12865–12876.

79. Longfellow, C. E., R. Kierzek, and D. H. Turner. 1990. Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry.* 29:278–285.

80. SantaLucia, J., Jr., R. Kierzek, and D. H. Turner. 1991. Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U.U and C.C.+ pairs. *Biochemistry.* 30:8242–8251.

81. Zhang, Z. L., Y. Y. Wu, …, Z. J. Tan. 2017. Divalent ion-mediated DNA-DNA interactions: a comparative study of triplex and duplex. *Biophys. J.* 113:517–528.

82. Denesyuk, N. A., and D. Thirumalai. 2015. How do metal ions direct ribozyme folding? *Nat. Chem.* 7:793–801.

83. Leipply, D., and D. E. Draper. 2011. Effects of Mg^{2+} on the free energy landscape for folding a purine riboswitch RNA. *Biochemistry.* 50:2790–2799.

84. Serra, M. J., J. D. Baird, …, E. Westhof. 2002. Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA.* 8:307–323.

85. Hagerman, P. J. 1997. Flexibility of RNA. *Annu. Rev. Biophys. Biomol. Struct.* 26:139–156.

86. Bao, L., X. Zhang, …, Z. J. Tan. 2016. Flexibility of nucleic acids: from DNA to RNA. *Chin. Phys. B.* 25:018703.

87. Li, J., S. S. Wijeratne, …, C. H. Kiang. 2015. DNA under force: mechanics, electrostatics, and hydration. *Nanomaterials (Basel).* 5:246–267.

88. Sutton, J. L., and L. Pollack. 2015. Tuning RNA flexibility with helix length and junction sequence. *Biophys. J.* 109:2644–2653.

89. Chen, H., S. P. Meisburger, …, L. Pollack. 2012. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. USA.* 109:799–804.

90. Drozdetski, A. V., I. S. Tolokh, …, A. V. Onufriev. 2016. Opposing effects of multivalent ions on the flexibility of DNA and RNA. *Phys. Rev. Lett.* 117:028101.

91. Bao, L., X. Zhang, …, Z. J. Tan. 2017. Understanding the relative flexibility of RNA and DNA duplexes: stretching and twist-stretch coupling. *Biophys. J.* 112:1094–1104.

92. Zhang, X., L. Bao, …, Z. J. Tan. 2017. Radial distribution function of semiflexible oligomers with stretching flexibility. *J. Chem. Phys.* 147:054901.

93. Herrero-Galán, E., M. E. Fuentes-Perez, …, J. R. Arias-Gonzalez. 2013. Mechanical identities of RNA and DNA double helices unveiled at the single-molecule level. *J. Am. Chem. Soc.* 135:122–131.

94. Wu, Y. Y., L. Bao, …, Z. J. Tan. 2015. Flexibility of short DNA helices with finite-length effect: from base pairs to tens of base pairs. *J. Chem. Phys.* 142:125103.

95. Kebbekus, P., D. E. Draper, and P. Hagerman. 1995. Persistence length of RNA. *Biochemistry.* 34:4354–4357.

96. Ullner, M., and C. E. Woodward. 2002. Orientational correlation function and persistence lengths of flexible polyelectrolytes. *Macromolecules.* 35:1437–1445.

97. Roy, A., S. Panigrahi, …, D. Bhattacharyya. 2008. Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies. *J. Phys. Chem. B.* 112:3786–3796.

98. Leontis, N. B., and E. Westhof. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA.* 7:499–512.

99. Turner, D. H., and D. H. Mathews. 2010. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* 38:D280–D282.

Biophysical Journal, Volume *115*

Supplemental Information

Modeling Structure, Stability, and Flexibility of Double-Stranded RNAs in Salt Solutions

Lei Jin, Ya-Zhou Shi, Chen-Jie Feng, Ya-Lan Tan, and Zhi-Jie Tan

The energy functions of the coarse-grained model

The force field in the present coarse-grained (CG) model contains two parts, the bonded potential and nonbonded potential:

$$U_{total} = U_{bonded} + U_{nonbonded}. \quad \text{(S1)}$$

The bonded potential U_{bonded} including energies of bond length U_b, bond angle U_a and dihedral angle U_d, is used to describe the local connectivity between CG beads:

$$U_{bonded} = U_b + U_a + U_d, \quad \text{(S2)}$$

where

$$U_b = \sum_{bonds} K_b(r - r_0)^2, \quad \text{(S3)}$$

$$U_a = \sum_{angles} K_\theta(\theta - \theta_0)^2, \quad \text{(S4)}$$

$$U_d = \sum_{dihedrals} \left\{K_\varphi[1 - cos(\varphi - \varphi_0)] + \frac{1}{2}K_\varphi[1 - cos3(\varphi - \varphi_0)]\right\}. \quad \text{(S5)}$$

In Eqs. S3-S5, K_b, K_θ and K_φ represent the energy strength; r_0, θ_0 and φ_0 are the corresponding values at energy minimum. The initial parameters of these three potentials were derived from the Boltzmann inversion of corresponding atomistic distribution functions obtained from the statistical analysis on the experimental structures in the Protein Data Bank (PDB, https://www.rcsb.org/). The bonded potential works only on the CG beads in each single-stranded chain, and two sets of parameters $Para_{\text{helical}}$ and $Para_{\text{nonhelical}}$ are provided for CG beads in base-pairing and non-base-pairing regions, respectively. It should be noted that the $Para_{\text{nonhelical}}$ is used in folding process and the $Para_{\text{helical}}$ is used only for helical parts in structure refinement process; see more details in Refs. 1 and 2. The nonbonded potential $U_{nonbonded}$ in Eq. S1 is used to describe nonbonded interactions between CG beads intra- or inter-chains, and it includes hydrogen bond potential U_{bp}, base pairs stacking potential U_{bc}, coaxial stacking potential U_{cs}, excluded volume potential U_{exc}, and electrostatic potential U_{el}:

$$U_{nonbonded} = U_{bp} + U_{bc} + U_{cs} + U_{exc} + U_{el}. \quad \text{(S6)}$$

The hydrogen bond potential is calculated for every possible base pair (G-C, A-U and G-U) and is given by

$$U_{bp} = \sum_{i<j-3}^{N_{bp}} \frac{\varepsilon_{bp}}{1 + k_{NN}\left(r_{N_iN_j} - r_{NN}\right)^2 + k_{CN}\sum_{i(j)}\left(r_{C_iN_j} - r_{CN}\right)^2 + k_{PN}\sum_{i(j)}\left(r_{P_iN_j} - r_{PN}\right)^2}, \quad \text{(S7)}$$

where ε_{bp} ($\varepsilon_{GC} = 2\varepsilon_{AU} = 2\varepsilon_{GU}$) is the interaction strength. r_{NN}, r_{CN} and r_{PN} are three distances between the corresponding atoms of P, C and N in two paired nucleotides to determine whether the base-pair is well formed. k_{NN}, k_{CN} and k_{PN} are the corresponding energy strength. The base pairs stacking potential is calculated between every two nearest neighbor base pairs and is given by

$$U_{bs} = \frac{1}{2}\sum_{i,j}^{N_{bs}} |G_{i,i+1,j-1,j}| \left\{\left[5\left(\frac{\sigma_{st}}{r_{i,i+1}}\right)^{12} - 6\left(\frac{\sigma_{st}}{r_{i,i+1}}\right)^{10}\right] + \left[5\left(\frac{\sigma_{st}}{r_{j,j-1}}\right)^{12} - 6\left(\frac{\sigma_{st}}{r_{j,j-1}}\right)^{10}\right]\right\}, \quad \text{(S8)}$$

where σ_{st} is the optimum distance of two neighbor bases in the known helix structures and $G_{i,i+1,j-1,j}$ is the strength of base stacking energy and can be estimated from the combination of the experimental thermodynamics parameters (3) and Monte Carlo simulations; see more details in Refs. 1 and 2. The coaxial stacking potential is calculated between two discontinuous neighbor helices and is given by

$$U_{cs} = \frac{1}{2}\sum_{i\text{-}j,k\text{-}l}^{N_{cs}} \left|G_{i\text{-}j,k\text{-}l}\right| \{\left[1 - e^{-a(r_{ik} - r_{cs})}\right]^2 + \left[1 - e^{-a(r_{jl} - r_{cs})}\right]^2 - 2\}, \quad \text{(S9)}$$

where $G_{i\text{-}j,k\text{-}l}$ is the sequence-dependent base stacking strength, which is approximately taken as the stacking strength between the corresponding nearest neighbor base-pairs in an uninterrupted helix (3-5), and r_{cs} is the optimum distance between two coaxially stacked stems, which is directly obtained from the statistical analysis on the known structures in PDB; see more details in Ref. 2. The excluded volume potential represents the excluded volume interaction between the nonbonded CG beads and is given by

$$U_{exc} = \sum_{i<j}^{N} \begin{cases} 4\varepsilon\left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6}\right] & if\ r_{ij} \le \sigma_0 \\ 0 & if\ r_{ij} > \sigma_0 \end{cases}, \quad \text{(S10)}$$

where $\varepsilon = 0.26$ kcal/mol and σ_0 is the sum of the radii of bead *i* and *j*.

The structure based electrostatic potential is newly introduced in the present model to represent

the electrostatic interactions between charged *P* beads, and it is treated as a combination of Debye-Hückel approximation and the counterion condensation (CC) theory (1, 6):

$$U_{el} = \sum_{i<j}^{N} \frac{Q_i Q_j e^2}{4\pi\epsilon_0 \epsilon r_{ij}} e^{-\frac{r_{ij}}{l_D}}. \quad \text{(S11)}$$

Here, r_{ij} is the distance between the *i*-th and *j*-th P beads, and l_D is the Debye length. The reduced charge Q_i on the *i*-th P bead is

$$Q_i = 1 - f_i, \quad \text{(S12)}$$

where f_i is ion neutralization fraction. Here, beyond the assumption of uniform distribution of binding ions along RNA chain, f_i is RNA structure-dependent and includes the contributions of monovalent and divalent ions

$$f_i = x f_i^1 + (1-x) f_i^2, \quad \text{(S13)}$$

where f_i^ν $(\nu = 1,2)$ is the binding fraction of *ν*-valent ions for the *i*-th P bead. *x* and $(1-x)$ represent the contribution fractions of monovalent and divalent ions which can be derived from the Tightly Bound Ion (TBI) model (7-9). If we use Na^+ and Mg^{2+} to represent monovalent and divalent ions respectively, x can be given by the empirical formula

$$x = \frac{[Na^+]}{[Na^+] + \alpha[Mg^{2+}]}, \quad \text{(S14)}$$

where $\alpha = (8.1 - 64.8/N)(5.2 - \ln[Na^+])$, $[Na^+]$ and $[Mg^{2+}]$ are the corresponding bulk concentrations and *N* is the chain length (7-10); see more details in Ref. 2.

To further refine the electrostatic potential based on RNA structure, f_i^ν is given by

$$f_i^\nu = \frac{N \bar{f}_i^\nu}{\sum_N e^{-\beta\nu\phi_i}} e^{-\beta\nu\phi_i}. \quad \text{(S15)}$$

Here, $\bar{f}_i^\nu$ represents the average neutralization fraction for the *i*-th P bead and can be given by the CC theory (6): $\bar{f}_i^\nu = 1 - (\frac{b}{\nu l_B})$, where *b* is the average charge spacing on RNA backbone and l_B is Bjerrum length. ϕ_i in Eq. S15 is the electrostatic potential at the *i*-th P bead and can be approximately calculated by

$$\phi_i = \sum_{i \neq j}^{N} \frac{l_B Q_j}{r_{ij}} e^{-\frac{r_{ij}}{l_D}}. \quad \text{(S16)}$$

Therefore, the structure-dependent Q_i needs to be obtained through an iteration process by Eqs. S12-S16.

Calculating melting temperature at low experimental strand concentration

Since the experimental strand concentration of a dsRNA is generally very low, to improve the computation efficiency, the simulations for dsRNAs are generally performed at high strand concentrations C_s^h (e.g. 10 mM) to calculate melting temperatures. Based on the equilibrium conformations at each temperature *T*, the fraction $\Phi(T)$ of unfolded state characterized as completely dissociated single-stranded chain can be obtained at *T*. Since the small system of the simulation (two strands in a simulational box) can lead to significant finite-size effect (11), the predicted $\Phi(T)$ needs to be further corrected to the fraction $\theta_h(T)$ of unfolded state at the high bulk strand concentration C_s^h (11):

$$\theta_h(T) = 1 - \left(1 + \frac{\Phi(T)}{2a(1-\Phi(T))}\right) + \sqrt{\left(1 + \frac{\Phi(T)}{2a(1-\Phi(T))}\right)^2 - 1}, \quad \text{(S17)}$$

where *a*=1 and 2 for nonself-complementary and self-complementary sequences, respectively (11). Meanwhile, the ratio of folded to unfolded state $\phi(T)$ at each temperature can be written as

$$\phi(T) = \frac{Z_f}{Z_u} = e^{-\beta \Delta G_h}, \quad \text{(S18)}$$

where Z_f and Z_u are the partition functions of the ensembles of folded and unfolded states, respectively. ΔG_h is the Helmholtz free energy change of the system due to the transition from unfolded state to folded state and $\beta = \frac{1}{k_B T}$. Hence, $\theta_h(T)$ can also be written as

$$\theta_h(T) = 1 - \frac{\phi(T)}{1+\phi(T)} = \frac{1}{1+e^{-\beta \Delta G_h}}. \quad \text{(S19)}$$

Since the contribution of the translation entropy determined by the strand concentration can be derived as $k_B \ln C_s^h/\mu$ (12), ΔG_h can be derived as

$$\Delta G_h = \Delta G_0 - k_B T \ln C_s^h/\mu, \quad \text{(S20)}$$

where ΔG_0 is the free energy change only depending on sequence, and $\mu = 1$ for self-complementary sequences and $\mu = 4$ for nonself-complementary sequences (12). Similarly, For a low experimental strand concentration C_s (e.g., 0.1 mM), the fraction $\theta(T)$ of unfolded state can also be written as

$$\theta(T) = \frac{1}{1+e^{-\beta \Delta G}}, \quad \text{(S21)}$$

and

$$\Delta G = \Delta G_0 - k_{\mathrm{B}} T \ln C_s/\mu. \tag{S22}$$

Then, based on the Eqs. S18-S22, $\theta(T)$ at low experimental strand concentration can be calculated by

$$\theta(T) = \frac{\lambda\theta_h(T)}{1 + \lambda\theta_h(T) - \theta_h(T)}, \tag{S23}$$

where $\lambda = C_s^h/C_s$. Furthermore, based on the fraction $\theta(T)$ of unfolded state, the melting curve can be obtained by fitting to

$$\theta(T) = 1 - \frac{1}{1 + e^{(T-T_m)/dT}}, \tag{S24}$$

where *dT* is an adjustable parameter (1,2).

Table S1. The parameters of bonded potential

Bond U_b				
	K_b (kcal/mol/Å^2)		r_0 (Å)	
	$Para_{helical}$	$Para_{nonhelical}$	$Para_{helical}$	$Para_{nonhelical}$
P_iC_i	133.4	98.2	3.95	3.95
C_iP_{i+1}	75.0	42.5	3.93	3.93
C_iN_i	85.6	24.8	3.35	3.45
Angle U_a				
	K_θ (kcal/mol/rad^2)		θ_0 (rad)	
	$Para_{helical}$	$Para_{nonhelical}$	$Para_{helical}$	$Para_{nonhelical}$
$P_iC_iP_{i+1}$	18.3	9.3	1.74	1.75
$C_{i-1}P_iC_i$	43.9	21.3	1.76	1.78
$P_iC_iN_i$	35.5	9.7	1.63	1.64
$N_iC_iP_{i+1}$	99.8	15.2	1.66	1.66
Dihedral U_d				
	$K\varphi$ (kcal/mol)		φ_0 (rad)	
	$Para_{helical}$	$Para_{nonhelical}$	$Para_{helical}$	$Para_{nonhelical}$
P_iC_i $P_{i+1}C_{i+1}$	2.8	1.1	2.56	2.51
$C_{i-1}P_iC_iP_{i+1}$	10.5	4.3	-2.94	-2.92
$C_{i-1}P_iC_iN_i$	3.8	0.8	-1.16	-1.18
$N_{i-1}C_{i-1}P_iC_i$	4.2	0.7	0.88	0.78

Table S2. The parameters of nonbonded potential

U_{exc}	ε (kcal/mol)	0.26	σ_{st} (Å)	R_i+R_j
U_{bp}	ε_{bp}(GC) (kcal/mol)	-3.5		
	k_{NN} (Å^{-2})	3.6	r_{NN} (Å)	8.9
	k_{CN} (Å^{-2})	1.9	r_{CN} (Å)	12.2
	k_{PN} (Å^{-2})	0.7	r_{PN} (Å)	13.9
U_{bs}	$G_{i,i+1,j-1,j}$(kcal/mol)	Sequence-dependent	σ_{st} (Å)	4.8
U_{el}	Q_i	Structure-dependent	b (Å)	5.5
U_{cs}	$G_{i-j,k-l}$ (kcal/mol)	Sequence-dependent	a (Å^{-1})	0.4
			r_{cs} (Å)	5.0

Table S3. 16 dsRNAs in X-ray set for structure prediction at 1 M [Na^+]

PDB code	Description[a]	Length (nt)	Base pairs[b] (bp)	$RMSD_{mean}$ (Å)	$RMSD_{min}$ (Å)
472d	H	16	8	1.6	0.6
259d	H	16	8	1.5	0.7
1dqf	B	19	9	2.5	1.6
2ao5	H	20	10	2.9	1.2
1kd5	I	22	6	4.1	2.1
1qcu	H	22	11	2.1	1.1
1yyk	H	24	12	2.5	1.1
353d	H	24	12	2.9	1.2
157d	I	24	10	2.4	1.1
255d	I	24	10	2.1	1.4
283d	I	24	8	2.5	1.6
1i9x	B	26	12	3.8	1.7
1mhk	T	26	6	1.9 (9.8) [c]	1.2 (5.6) [c]
1csl	B&I	28	9	4.2	2.2
3wbm	B&I	50	20	5.4	2.3
2f8t	T	50	22	5.2	2.7

[a] H stands for dsRNAs of complementary duplex, B stands for dsRNAs with bulge loop, I stands for dsRNAs with internal loop, and T stands for dsRNAs with single-stranded tail.
[b] Only Watson-Crick base pairs (G-C and A-U) and Wobble base pairs (G-U) are concluded.
[c] The RMSD in bracket is calculated with the involvement of dangling tails.

Table S4. The comparison between the present model and its previous version in RNA structure predictions

PDB code	Description[a]	Length (nt)	$[Na^+/K^+]/[Mg^{2+}]$[b] mM/mM	mean/minimum RMSDs[c] (Å) Previous version[d]	New version[e]
2gm0	D,I	70	250/0.1	7.3/3.9	6.1/3.1
2m1o	D,H	14	80/0.1	2.2/0.7	2.0/0.7
1tut	D,I	22	80/3	3.5/1.9	3.2/1.7
2kyd	D,H	40	150/10	3.9/1.2	3.7/1.2
2d1a	D,I	78	50/0	7.3/3.9	6.9/3.8
2dd1	D,I	20	90/0	3.5/1.7	3.3/1.5
2k7e	D,I	24	110/0	2.8/1.4	2.4/1.3
2lx1	D,I	22	90/0	5.4/2.8	4.8/2.4
2jxq	D,H	20	60/0	2.1/0.8	2.0/0.8
1f5g	D,I	20	80/0	3.0 /1.8	2.8/1.7
1j4y	S	17	20/0	3.9/1.9	3.8/1.9
1d0u	S,B	21	50/0	3.8/1.5	3.5/1.4
2l5z	S,I	26	50/5	4.0/2.6	3.6/2.2
1p5o	S,B,I	77	100/5	11.0/8.7	9.8/6.6
2g1w	S,P	22	50/0	4.8/3.3	4.1/2.2
2rp1	S,P	27	110/5	4.1/2.7	3.9/2,4
1kpy	S,P	33	100/5	4.2/2.4	3.9/2.2
2ap5	S,P	28	100/5	5.6/3.8	4.9/3.3

[a] D stands for dsRNA, S stands for ssRNA, H stands for complementary duplex, B stands for bulge loop, I stands for internal loop, and P stands for pseudoknot.

[b] The experimental ion conditions for structure determination by NMR method.

[c] The RMSDs are calculated over all three CG beads of predicted structures by the present model from the corresponding atoms of the native structures.

[d] Using electrostatic potential described in Refs. 2 and 13.

[e] Using electrostatic potential described in this work.

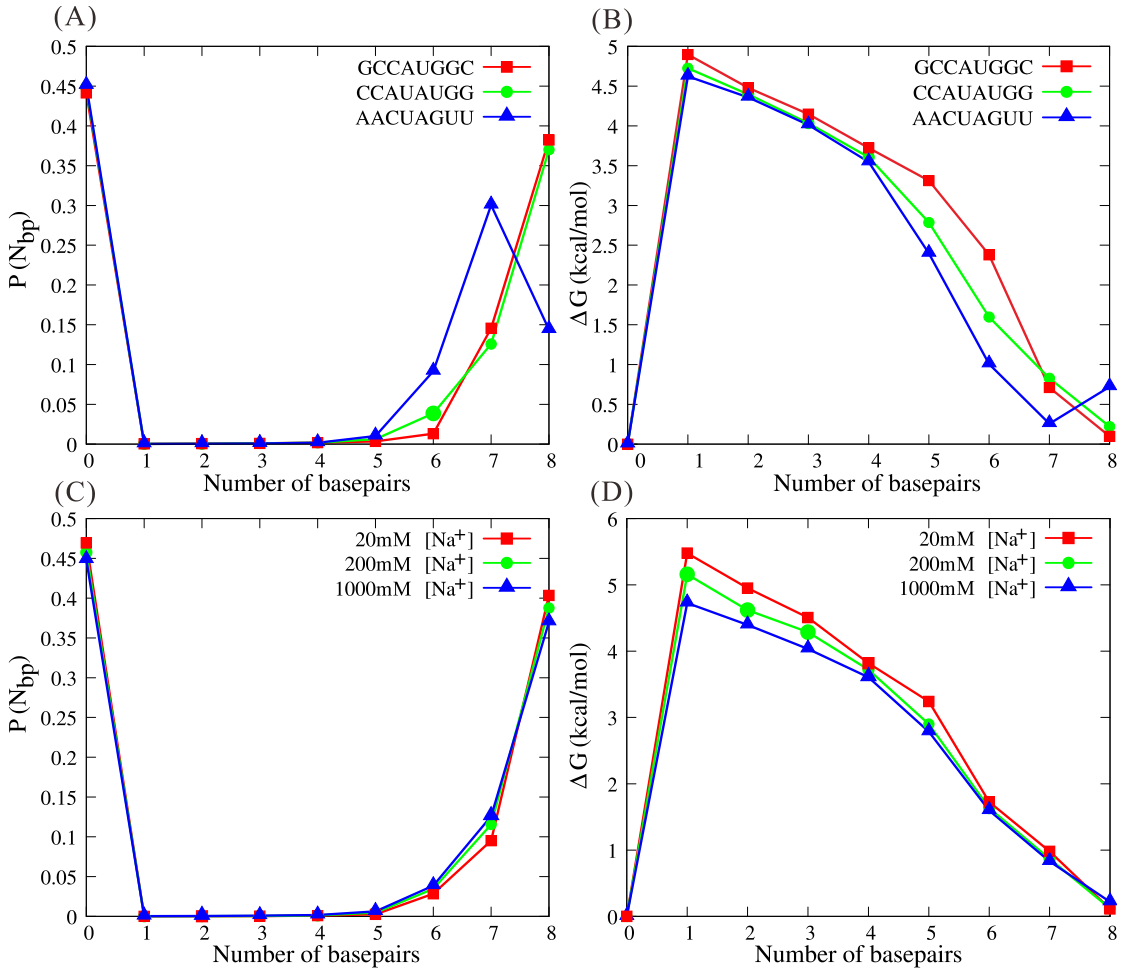

FIGURE S1. (A, B) The normalized populations of formed base pair number $P(N_{bp})$ (A) and the free energy barrier ΔG (B) as functions of the number of base pairs of three sequences at T_m. (C, D) The normalized populations of formed base pair number $P(N_{bp})$ (C) and the free energy barrier ΔG (D) of the sequence CCAUAUGG at T_m at different [Na^+]'s. Here, the free energies are calculated by $\Delta G = -k_B T \ln[P(N_{bp})]$.

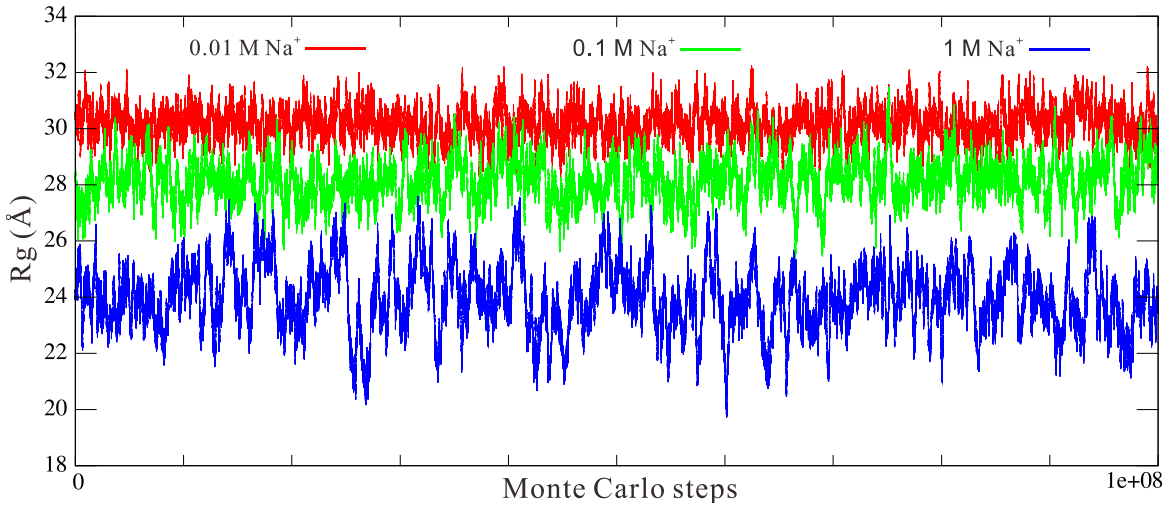

Figure S2. The time-evolution of the radius of gyration for the 40-bp dsRNA helix at different Na^+ concentration.

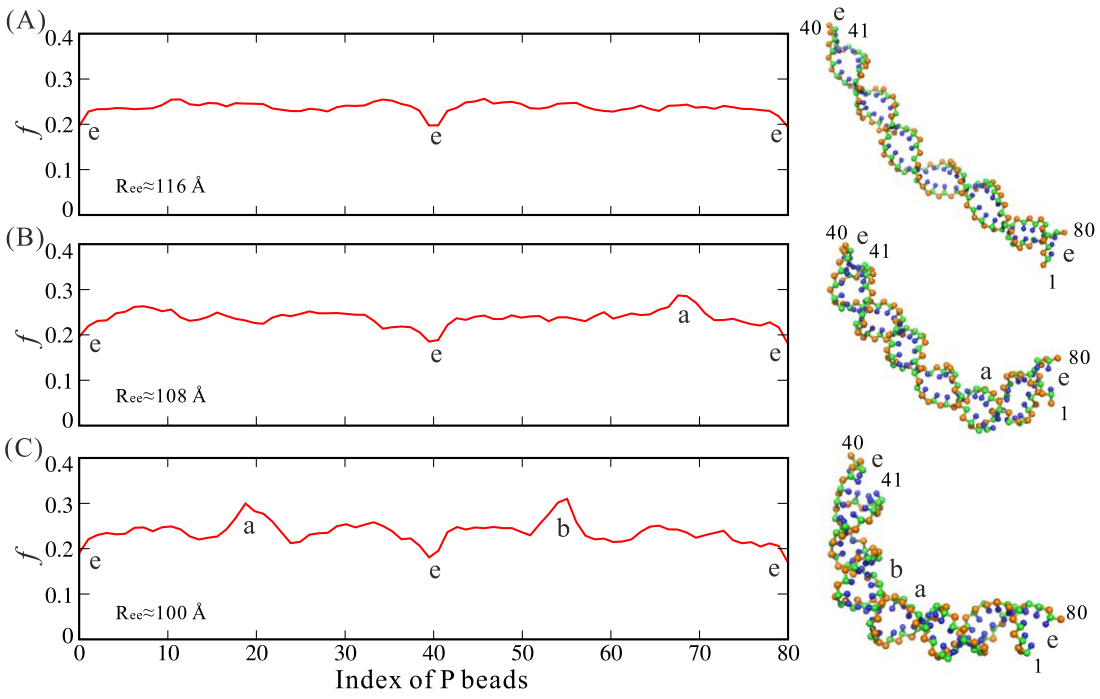

Figure S3. The calculated ion charge neutralization fraction f (left) along P beads of the corresponding structures (right) with the end-to-end distances Ree of ~116 Å (A), ~108 Å (B) and ~100 Å (C) of the 40-bp dsRNA at 0.1 M [Na^+]. The P beads in the bending region of the structures are labeled with a and b corresponding to the peaks of ion charge neutralization fraction, and the ends of the helix are labeled with e corresponding to the ion charge neutralization fraction troughs.

References:

1 Shi, Y. Z., Wang, F. F., Wu, Y. Y., and Tan, Z. J. 2014. A coarse-grained model with implicit salt for RNAs: Predicting 3D structure, stability and salt effect. *J. Chem. Phys.*, 141. 105102.

2 Shi, Y. Z., Jin, L., Wang, F. F., Zhu, X. L., and Tan, Z. J. 2015. Predicting 3D Structure, Flexibility, and Stability of RNA Hairpins in Monovalent and Divalent Ion Solutions. *Biophys. J.*, 109. 2654-2665.

3 Xia, T., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. 1998. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs. *Biochemistry*, 37. 14719-14735.

4 Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., and Zuker, M. 1994. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA*,. 91. 9218-22.

5 Walter, A. E. and Turner, D. H. 1994. Sequence dependence of stability for coaxial stacking of RNA helixes with Watson-Crick base paired interfaces. *Biochemistry*, 33. 12715-9.

6 Manning, G. S. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.*, 11. 179-246.

7 Tan, Z. J. and Chen, S. J. 2005. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J. Chem. Phys.*, 122. 044903.

8 Tan, Z. J. and Chen, S. J. 2011. Salt Contribution to RNA Tertiary Structure Folding Stability. *Biophys. J.*, 101. 176-187.

9 Wang, F. H., Wu, Y. Y. and Tan, Z. J. 2013. Salt Contribution to the Flexibility of Single-Stranded Nucleic Acid of Finite Length. *Biopolymers*, 99. 370-381.

10 Xi, K., Wang, F. H., Xiong, G., Zhang, Z. L. and Tan, Z. J. 2018. Competitive binding of Mg^{2+} and Na^{+} ions to nucleic acids: from helices to tertiary structures. *Biophys. J.*, 114. 1776-1790.

11 Ouldridge, T. E., Louis, A. A. and Doye, J. P. 2010. Extracting bulk properties of self-assembling systems from small simulations. *J. Phys. Condens. Matt.*, 22, 104102.

12 Borer, P. N., Dengler, B., Tinoco, I. J., and Uhlenbeck, O. C. 1974. Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.*, 86. 843-53.

13 Shi, Y. Z, Jin, L., Feng, C. J., Tan, Y. L. and Tan, Z. J. 2018. Predicting 3D structure and stability of RNA pseudoknots in monovalent and divalent ion solutions. *PLoS Comput. Biol.*, 14(6), e1006222.