

Structure folding of RNA kissing complexes in salt solutions: predicting 3D structure, stability, and folding pathway

LEI JIN,¹ YA-LAN TAN,¹ YAO WU,¹ XUNXUN WANG,¹ YA-ZHOU SHI,² and ZHI-JIE TAN¹

¹Center for Theoretical Physics and Key Laboratory of Artificial Micro and Nano-structures of Ministry of Education, School of Physics and Technology, Wuhan University, Wuhan 430072, China

²Research Center of Nonlinear Science, School of Mathematics and Computer Science, Wuhan Textile University, Wuhan 430073, China

ABSTRACT

RNA kissing complexes are essential for genomic RNA dimerization and regulation of gene expression, and their structures and stability are critical to their biological functions. In this work, we used our previously developed coarse-grained model with an implicit structure-based electrostatic potential to predict three-dimensional (3D) structures and stability of RNA kissing complexes in salt solutions. For extensive RNA kissing complexes, our model shows great reliability in predicting 3D structures from their sequences, and our additional predictions indicate that the model can capture the dependence of 3D structures of RNA kissing complexes on monovalent/divalent ion concentrations. Moreover, the comparisons with extensive experimental data show that the model can make reliable predictions on the stability for various RNA kissing complexes over wide ranges of monovalent/divalent ion concentrations. Notably, for RNA kissing complexes, our further analyses show the important contribution of coaxial stacking to the 3D structures and stronger stability than the corresponding kissing-interface duplexes at high salts. Furthermore, our comprehensive analyses for RNA kissing complexes reveal that the thermally folding pathway for a complex sequence is mainly determined by the relative stability of two possible folded states of kissing complex and extended duplex, which can be significantly modulated by its sequence.

Keywords: RNA kissing complex; 3D structure prediction; thermal stability; ion effect; folding pathway

INTRODUCTION

RNA complexes are functionally important in gene regulation and expression (Baker et al. 2003; Atkins et al. 2011). An RNA kissing complex is formed through base-pairing between the complementary sequences of two loops (e.g., hairpin loop), and can play critical roles in RNA dimerization and packaging processes in retroviruses (Chen and Varani 2001; Miller and White 2006; Dubois et al. 2018). For example, the dimerization of two RNA copies induced by kissing complex has been proposed to be the essential initiation of HIV-1 viral replication (Laughrea and Jetté 1996; Paillart et al. 1996). Moreover, the formation of RNA kissing complex between untranslated mRNA and antisense RNA executes regulatory functions in both prokaryotes and eukaryotes (Kolb et al. 2001; Gamache et al. 2017). For example, the kissing complex between the ColE1 plasmid-specific transcripts RNA I and antisense regulators RNA II can mediate plasmid replication regula-

tion (Marino et al. 1995; Kolb et al. 2000). In addition, kissing complexes are a kind of fundamental tertiary motifs in RNAs such as ribozymes and riboswitches, which can stabilize RNA spatial tertiary structures and provide ligand-binding/protein-recognition sites for activation (Andersen and Collins 2001; Piganeau et al. 2006; Bouchard and Legault 2014b). Generally, the biological functions of RNA kissing complexes could strongly depend on their three-dimensional (3D) structures and stability (Baker et al. 2003; Atkins et al. 2011). Furthermore, as RNA tertiary structure motifs with high charge density, the formation of RNA kissing complexes induced by inter-strand base-pairing/stacking is very sensitive to solution temperature and ion conditions (Draper et al. 2005; Woodson 2005; Koculi et al. 2007; Chen 2008; Wong and Pollack 2010; Lipfert et al. 2014; Mustoe et al. 2014). Therefore, the knowledge of 3D structures and stability of RNA kissing

Corresponding authors: zjtan@whu.edu.cn, yzshi@wtu.edu.cn
Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.071662.119>.

© 2019 Jin et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://majournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

complexes in ion solutions is highly required to fully understand their biological functions.

Experimental methods such as X-ray crystallography, NMR spectroscopy, and newly developed cryo-electron microscopy have been used to obtain RNA 3D structures (Rose et al. 2017). However, it is still time-consuming and expensive to experimentally derive high-resolution 3D structures of RNAs (Hajdin et al. 2010; Pyle and Schlick 2016; Schlick and Pyle 2017). As a result, there are still limited available structures of RNA kissing complexes deposited in the Protein Data Bank (PDB) (Rose et al. 2017). Simultaneously, as complementary methods, some computational models have been developed in recent years to predict 3D structures of RNAs (Shapiro et al. 2007; Sim et al. 2012; Cragolini et al. 2015; Somarowthu 2016; Miao and Westhof 2017; Miao et al. 2017; Sun et al. 2017). The fragment-assembly methods such as MC-Fold/MC-Sym pipeline (Parisien and Major 2008), FARNAs (Das and Baker 2007), Vfold3D (Cao and Chen 2011a), RNAComposer (Popenda et al. 2012), and 3dRNA (Zhao et al. 2012; Wang et al. 2015, 2017) can efficiently predict 3D structures of RNAs, based on given secondary structures. However, due to the lack of 3D structure templates and inexact predicted secondary structures, it is still difficult for them to make reliable predictions on the 3D structures of kissing complexes from sequences (Zuker 2003; Dimitrov and Zuker 2004; Rehmsmeier et al. 2004; Andronescu et al. 2005; Bernhart et al. 2008; Huang et al. 2009; Shi et al. 2014b; Yamasaki et al. 2014). Although RNA2D3D can be used to manually construct the 3D structure of a kissing complex (Martinez et al. 2008), its reliability strongly depends on the expert knowledge of users. In parallel, some coarse-grained (CG) models (Hyeon and Thirumalai 2011; Zhang et al. 2012; He et al. 2013; Kim et al. 2014; Bian et al. 2015; Dawson et al. 2016; Li et al. 2016; Boudard et al. 2017; Jain and Schlick 2017; Sieradzan et al. 2017; Uusitalo et al. 2017) such as iFold (Ding et al. 2008), NAST (Jonikas et al. 2009), SimRNA (Boniecki et al. 2015), HiRE-RNA (Cragolini et al. 2013), RACER (Xia et al. 2013), and oxRNA (Šulc et al. 2014) have been developed to predict RNA 3D structures by involving experimental thermodynamic parameters (Xia et al. 1998) or/and knowledge-based statistical potentials (Tan et al. 2019). However, the structures of kissing complexes have not been involved in these 3D structure prediction models.

Since RNA-mediated functions are often coupled to the RNA structural stability (Chen and Varani 2001; Baker et al. 2003; Atkins et al. 2011), some computational models have been used to predict thermodynamics of RNAs. The VfoldCPX/VfoldThermal with involving thermodynamic parameters and special loop entropy can make reliable predictions on thermodynamics of RNAs including kissing complexes at a secondary structure level (Cao and Chen 2011b; Cao et al. 2014). The structure-based (Gö-like)

TIS model can well predict the thermodynamics of small RNAs including pseudoknots (Denesyuk and Thirumalai 2013; Hori et al. 2016; Denesyuk et al. 2018), but the model could not predict 3D structures of RNAs solely from the sequences. Although some other models such as iFold (Ding et al. 2008), HiRE-RNA (Cragolini et al. 2013), oxRNA (Šulc et al. 2014), and NARES-2P (He et al. 2013) can give melting curves of RNAs, there is still a lack of experimental validation for the predictions and these models are still required to be tested for predicting the stability of RNA kissing complexes.

Very importantly, metal ions, especially divalent ions such as Mg^{2+} , can strongly affect the 3D structures and stability of RNAs due to the polyanionic nature of RNAs (Gregorian and Crothers 1995; Weixlbaumer et al. 2004; Lorenz et al. 2006; Vo et al. 2009; Singh et al. 2011; Sehdev et al. 2012; Kim and Shapiro 2013; Li 2013). Thus, it is also necessary for a predictive model to involve the effect of monovalent/divalent ions to predict the 3D structures and thermodynamics of RNAs in physiological environments containing monovalent and divalent ions (Chen 2008; Sun et al. 2017). However, the effect of ions on the structures and stability of RNAs, especially the effect of divalent ions, is seldom involved in the existing 3D structure prediction models. The effect of ions has not been involved in the fragment-assembly methods due to the limitations of templates at various ion conditions, and the existing physics-based models such as iFold (Ding et al. 2008), HiRE-RNA (Cragolini et al. 2013), NARES-2P (He et al. 2013), and RACER (Xia et al. 2013) also did not involve the effect of ions on 3D structures, although the effect of ions may be incorporated in such kinds of models through certain implicit or explicit modeling. Especially, the involvement of divalent ions in 3D structure predictions of RNAs is a challenge due to the requirement of structure prediction efficiency and the intrinsic treatment complexity of strong correlations of divalent ions (Tan and Chen 2007; Sun et al. 2017; Xi et al. 2018; Hori et al. 2019). To predict 3D structures and stability of RNAs in ion solutions, we have recently developed a CG model with involving an implicit electrostatic potential and sequence-based thermodynamic parameters (Shi et al. 2014a). The model has been validated by making reliable predictions on 3D structures and stability for RNA hairpins, double-stranded RNAs and RNA pseudoknots, and the involvement of the electrostatic potential in the model can visibly improve the prediction accuracy for the 3D structure and stability of the RNAs in extensive monovalent/divalent ion solutions (Shi et al. 2014a, 2015, 2018; Jin et al. 2018). However, the model should be further validated for predicting 3D structures and stability for the very important tertiary motifs of RNA kissing complexes with tertiary contact of inter-strand base-pairing/stacking in monovalent/divalent ion solutions. Moreover, although the coaxial stacking interactions have been found

in the folded structures of RNA kissing complexes (Chen and Varani 2001), it is still lacking a deep understanding on its contribution to the 3D structures and stability of RNA kissing complexes. Furthermore, there can be two possible folded structures—kissing complex and extended duplex for certain complex sequences, while the assembly mechanism for RNA kissing complexes of different sequences is still not fully clear. Therefore, for RNA kissing complexes, to predict the 3D structures/stability in monovalent/divalent salt solutions, and to analyze the contribution of coaxial stacking to 3D structures/stability and thermally folding pathways are still challenging and significant for understanding their structure-related biology.

In this work, we used our previously developed CG model with a new modification on coaxial stacking potential to predict the 3D structures and thermal stability of RNA kissing complexes in monovalent/divalent ion solutions and to analyze the contribution of coaxial stacking to the 3D structure/stability and thermally folding pathway of RNA kissing complexes. As compared with the extensive experimental data, the present model can make reliable predictions on 3D structures and thermal stability for extensive RNA kissing complexes, and can capture the effects of monovalent/divalent salts with high accuracy. Our further analyses also revealed the important contribution of coaxial stacking to the 3D structure and stronger stability of kissing complexes than the kissing-interface duplexes at high salts. Furthermore, our comprehensive analyses on the thermally folding pathway showed that, for a complex sequence, the folding pathway is mainly determined by the relative stability of the two possible folded states of kissing complex and extended duplex, which can be significantly modulated by its sequence.

RESULTS AND DISCUSSION

In this section, we first made predictions on 3D structures for extensive RNA kissing complexes in monovalent/divalent ion solutions. Afterward, we used the present model to predict the stability of RNA kissing complexes with extensive sequences, and further investigated the effect of monovalent/divalent ions on the stability for the HIV DIS type RNA kissing complex. Furthermore, we made comprehensive analyses on the contribution of coaxial stacking to the 3D structures and stability for RNA kissing complexes. Finally, the thermally folding pathways of RNA kissing complexes of typical sequences were analyzed thoroughly to reveal the assembly mechanism of RNA kissing complexes.

Predicting 3D structures of RNA kissing complexes

As described in Materials and Methods, for each RNA kissing complex, two random chains were generated from the sequence and separately placed in a simulation box with the size determined by the strand concentration (e.g., 1 mM). Afterward, the MC simulation with a simulated annealing algorithm was performed from a high temperature to the target temperature (e.g., 25°C). As an example, for the HIV-2 TAR-TAR* kissing complex (PDB id: 1KIS), as shown in Figure 1C,D, the total energy of the simulation system decreases gradually along the annealing process, and simultaneously, the total number of the formed base pairs increases. Since the melting temperatures (~60°C–80°C) of individual RNA hairpins are generally much higher than those of their kissing complexes (~30°C–50°C at experimental strand concentrations) (Gregorian and Crothers 1995; Huang et al. 2009; Li and Tinoco 2009;

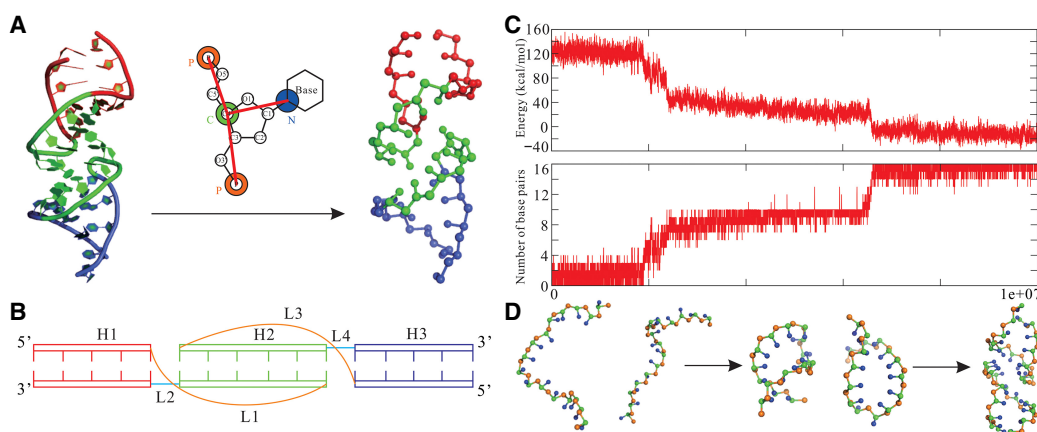


FIGURE 1. (A) The representation of all-atomistic (*left*) and CG (*right*) for the HIV-2 TAR-TAR* RNA kissing complex (PDB id: 1KIS), and the schematic representation for one nucleotide in the present CG model (*middle*). (B) A general secondary structure representation of RNA kissing complex including three helices (H1, H2, and H3) and four loops (L1, L2, L3, and L4), and the helix H2 is formed by base-pairing between two individual RNA hairpins. The corresponding secondary structure elements in A and B are in the same colors: H1 (red), H2 (green), H3 (blue), L1 and L3 (orange), and L2 and L4 (cyan). (C,D) The time-evolution of the energy (*top panel in C*), the number of base pairs (*bottom panel in C*), and the typical 3D structures (*D*) during the Monte Carlo simulated annealing process. All the 3D structures are shown with PyMOL (<http://www.pymol.org>).

Cao and Chen 2011b), the folding process of the HIV-2 TAR-TAR* kissing complex includes two steps: the formations of two hairpins and native kissing complex, respectively (see Fig. 1C,D).

To better capture the geometry of helical parts, additional MC simulation of further structure refinement was performed based on the conformation at the end of the annealing process at the target temperature (e.g., 25°C), in which the parameters of bonded potentials were used for the base-pairing regions to more sufficiently depict the helical geometry of the stems (see also *Para_{helical}* in Supplemental Table S1; Shi et al. 2014a, 2015, 2018; Jin et al. 2018). The predicted structure ensemble from the structure refinement can be evaluated by their root-mean-square deviation (RMSD) values calculated over all the CG beads to the corresponding atoms in the native structures in PDB (Rose et al. 2017), since there is still no reliable scoring function to identify the nearest-native structure from the predicted CG structure ensemble for the all-atom structure conversion (Li et al. 2018; Tan et al. 2019). Since the CG beads were simplified from the structurally fundamental atom groups (phosphate, sugar, and base groups) and are located at the coordinates of key atoms of the groups (P, C4' and N1/N9 atoms), the predicted CG structures of RNAs keep the important structure features of the RNAs, and consequently, the RMSDs from predicted CG structures can serve as an examination quantity for prediction reliability. For predicted structures, the mean RMSD (the averaged value over the structure ensemble in the refinement process) and minimum RMSD (corresponding to the structure closest to the native one in the refinement process) were used to evaluate the reliability of the predictions (see more details in Shi et al. 2014a, 2015, 2018; Jin et al. 2018). As shown in Figure 2, for 1KIS, the mean and minimum RMSDs are 4.8 and 2.5 Å, respectively, suggesting a reliable precision of the present model.

3D structure predictions for RNA kissing complexes

According to the above process, we used the present model to predict the 3D structures of 18 RNA kissing complexes, each of which has been determined by experimental methods (e.g., NMR and X-ray crystallography); see Supplemental Table S3. For the

RNAs measured by NMR, we made predictions at their respective experimental ion conditions; while for the RNAs from X-ray crystallography, we made predictions at 1 M NaCl (i.e., RNAs are nearly full-neutralized) due to the absence of experimental ion conditions.

As shown in Figure 2, for the 18 RNAs with different sequences as well as different lengths of loops, the overall mean (minimum) RMSD between the predicted structures and the experimental structures is ~5.4 Å (~2.7 Å), which suggests that the present model can make reliable predictions for 3D structures of RNA kissing complexes with three coaxial stacked helical stems. For visual convenience, the predicted secondary structures for the 18 RNAs were also shown in Supplemental Figure S2. Generally, in RNA kissing complexes, the junction loops L1 and L3 across the deep major groove of the helix, which can cause a visible bending of the helix toward its major groove (Chen and Varani 2001). We further calculated the stem–stem (H1–H3) angles for the 18 RNA kissing complexes using the DSSR software (Lu et al. 2015). As shown in Figure 3, the calculated stem–stem angles are in good agreement with the experimental data (Chang and Tinoco 1997; Lee and Crothers 1998; Mujeeb et al. 1998; Kim and Tinoco

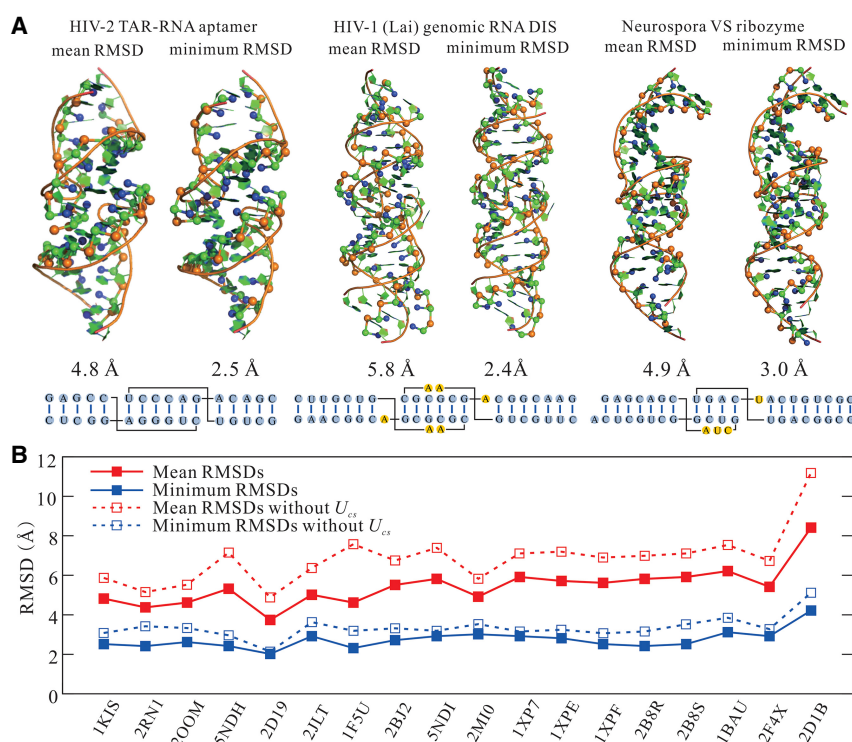


FIGURE 2. (A) The predicted 3D structures (ball-stick) with the mean RMSDs and the minimum RMSDs for three sample RNA kissing complexes (PDB ids: 1KIS [left], 2B8R [middle] and 2M10 [right]) and their corresponding predicted secondary structures (bottom). The mean (minimum) RMSDs for the three kissing complexes are 4.8 Å (2.5 Å), 5.8 Å (2.4 Å), and 4.9 Å (3.0 Å), respectively. (B) The mean (minimum) RMSDs for 18 RNA kissing complexes (from 32-nt to 78-nt) predicted by the present model with/without involving coaxial stacking potential (solid/dashed lines). The 3D structures are shown with PyMOL (<http://www.pymol.org>).

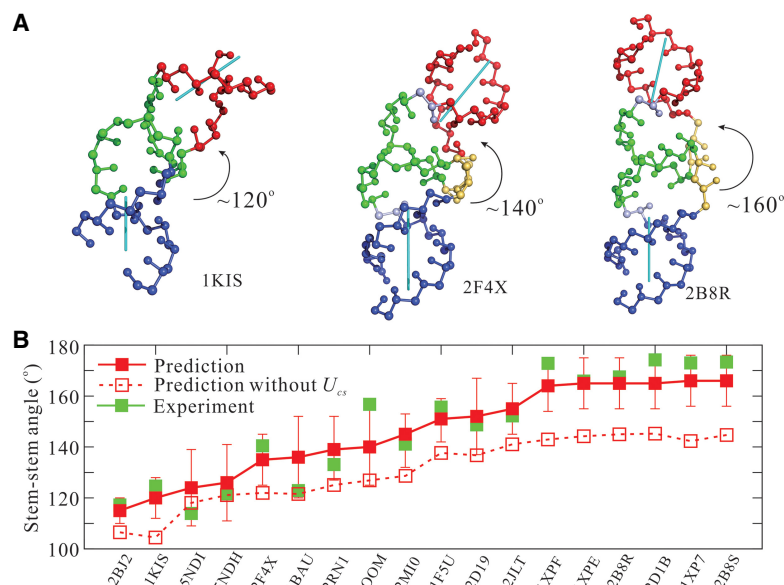


FIGURE 3. (A) The predicted 3D structures (ball-stick) of three RNA kissing complexes with different stem–stem (H1–H3) angles ($\sim 120^\circ$, $\sim 140^\circ$, and $\sim 160^\circ$). The three helical stems of H1, H2, and H3 are represented in red, green, and blue, respectively. Junction loops L1 and L3 are represented in orange. The long cyan sticks represent the helical axes of the helical stems of H1 and H3, which are calculated by DSSR software (Lu et al. 2015). (B) The stem–stem angles predicted by the present model with/without involving the coaxial stacking potential for the 18 RNA kissing complexes in comparisons with the experimental data. The 3D structures are shown with PyMOL (<http://www.pymol.org>).

2000; Ennifar et al. 2001, 2006; Baba et al. 2005; Ennifar and Dumas 2006; Kieken et al. 2006; Lebars et al. 2007, 2008; Van Melckebeke et al. 2008; Bouchard and Legault 2014a; Huang et al. 2017), and the mean deviation of the stem–stem angle is $\sim 6.5^\circ$ between our predictions and experimental data, suggesting that the present model is also effective in capturing the helix orientations of RNA kissing complexes. For example, the HIV-2 TAR-TAR* kissing complex (PDB id: 1KIS) with the junction loops of L1 = L3 = 0-nt has a sharp bending (stem–stem angle $\sim 120^\circ$) to the major groove of H2, while the HIV-1_{Lai} genomic DIS kissing complex (PDB id: 2B8R) with longer junction loops of L1 = L3 = 3-nt has a slight bending angle (stem–stem angle $\sim 160^\circ$; see Fig. 3A). The sharp bending for the HIV-2 TAR-TAR* complex is attributed to the strong mechanical pulling effect from the very short loops of L1 = L3 = 0-nt, and the slight bending for the HIV-1_{Lai} DIS complex results from the coaxial stacking between H2 and H1/H3 stems and the slight pulling effect of long loops of L1 = L3 = 3-nt.

Effect of coaxial stacking on 3D structures

Since coaxial stacking is important for RNA kissing complexes to keep the quasi-straight helix with proper bending (Marino et al. 1995; Chang and Tinoco 1997; Lee and Crothers 1998; Ennifar et al. 2001, 2006; Baba et al. 2005; Ennifar and Dumas 2006; Lebars et al. 2007, 2008;

Van Melckebeke et al. 2008; Li and Tinoco 2009; Mundigala et al. 2014), we further used our model to examine the effect of coaxial stacking on the 3D structures of the 18 RNA kissing complexes through removing the coaxial stacking potential in our predictions.

As shown in Figure 2B and Supplemental Table S3, the overall mean/minimum RMSDs ($6.9 \text{ \AA}/3.4 \text{ \AA}$) of the structures predicted by the model without the coaxial stacking potential are apparently larger than those ($5.4 \text{ \AA}/2.7 \text{ \AA}$) from the model with the coaxial stacking potential. Furthermore, our analyses show that there is significant deflection in the stem–stem (H1–H3) angles between the experimental structures and the structures predicted by the model without involving the coaxial stacking potential (see Fig. 3B). Such apparent deviations due to removing coaxial stacking potential suggest that the coaxial stacking makes a significant contribution to the global 3D structures and helix orientation at the junction

region of RNA kissing complexes, and the inclusion of the coaxial stacking potential could be important for a model to effectively reproduce the 3D structures and helix orientation of RNA kissing complexes.

Monovalent/divalent salt effects on 3D structures

Since ion solutions may strongly influence 3D structure of RNAs (Gregorian and Crothers 1995; Jossinet et al. 1999; Serra et al. 2002; Weixlbaumer et al. 2004; Lorenz et al. 2006; Vo et al. 2009; Leipply and Draper 2011; Singh et al. 2011; Sehdev et al. 2012; Kim and Shapiro 2013; Li 2013; Wu et al. 2015; Zhang et al. 2017; Chen et al. 2019), we have introduced a structure-based implicit electrostatic potential in the model (Jin et al. 2018). Here, to examine monovalent/divalent salt effects on 3D structures, for the 10 RNA kissing complexes determined by NMR, we made two separate structure predictions using the present model at their corresponding experimental ion conditions and at 1 M NaCl, that is, RNAs are nearly full-neutralized (Shi et al. 2015; Jin et al. 2018). As shown in Table 1, the involvement of the experimental ion conditions causes visible improvements for all the sequences in the 3D structure predictions, and such improvement appears even stronger for longer sequences, for example, for the sequence of 2D1B (78-nt), the mean RMSD of the predicted structures decreases by more than 3 Å. It is

TABLE 1. Ten RNA kissing complexes for structure predictions at experimental salt conditions

PDB	Length ^a (bp/nt)	[Na ⁺]/ [Mg ²⁺] ^b (mM/mM)	RMSD (Å) ^c mean/ minimum	RMSD (Å) ^d mean/ minimum
1KIS	16/32	180/5	4.8/2.5	6.1/3.3
2RN1	16/32	60/0	4.5/2.6	5.1/2.9
2OOM	16/32	10/0	4.6/2.6	5.7/3.4
2F4X	20/48	10/0	5.4/2.9	7.3/4.2
1BAU	20/46	100/0	6.2/3.1	7.3/4.1
2D19	14/34	50/0	4.0/2.1	4.8/3.1
2D1B	30/78	50/0	8.4/4.2	11.9/4.7
2MI0	16/43	50/5	4.9/3.0	5.6/3.5
1F5U	16/36	100/0	4.6/2.3	7.9/4.1
2BJ2	20/40	25/2	5.5/2.7	7.9/3.6

^aFor sequence length, bp stands for number of canonical base pairs, and nt stands for total number of nucleotides in structures.

^bThe experimental Na⁺ and Mg²⁺ ion conditions from solution NMR measurements.

^cThe predictions at 1 M NaCl, that is, RNAs are nearly full-neutralized.

^dThe predictions at experimental ion conditions.

reasonable, because RNA would be slightly less compact at lower salt due to the polyanionic nature and such effect would be more pronounced for longer RNAs. For the 10 RNA kissing complexes, the overall improvement is ~ 1.7 Å in mean RMSD and ~ 1.0 Å in minimum RMSD.

Furthermore, we examined the effects of Na⁺ and Mg²⁺ on the 3D structures of RNA kissing complexes of two typical RNA kissing complexes, HIV-2 TAR-TAR* (PDB id: 1KIS) and HIV-1_{Lai} DIS (PDB id: 2F4X), in wide ranges of [Na⁺] and [Mg²⁺]. Due to the available experimental measurements at various salts, we examined the dependence of stem–stem (H1–H3) angles of the complexes on monovalent/divalent salts. As shown in Figure 4A,B, the predicted stem–stem angles of the two RNA kissing complexes at different ion conditions agree well with the available experimental data (Chang and Tinoco 1997; Mujeeb et al. 1998; Baba et al. 2005; Kieken et al. 2006; Lebars et al. 2008; Van Melckebeke et al. 2008). It is interesting that the dependences of stem–stem angles on ion concentration are opposite for the two complexes: (i) The HIV DIS type RNA kissing complex (L1 = L3 = 2-nt, L2 = L4 = 1-nt) bends to a stem–stem angle of $\sim 140^\circ$ toward a major groove at very low Na⁺/Mg²⁺ concentrations and becomes quasi-straight with stem–stem angle $\sim 160^\circ$ at high Na⁺/Mg²⁺ concentrations; (ii) The HIV-2 TAR-TAR* complex (L1 = L2 = L3 = L4 = 0-nt) exhibits bent (stem–stem angle $\sim 140^\circ$) at very low Na⁺/Mg²⁺ concentrations and becomes visibly more bent (stem–stem angle $\sim 100^\circ$) at high Na⁺/Mg²⁺ concentrations. Our analyses revealed that such interesting opposite salt dependences for the two typical complexes are attributed to the competition between coaxial stacking and mechanical pulling of the loop at the

junction regions which are oppositely modulated by ions. Generally, coaxial stacking would make a quasi-continuous helix straight while a loop at junctions would cause it bent (Shi et al. 2015). For the folded complexes, ions can enhance the coaxial stacking interaction due to the neutralization for negative charges on RNAs (Chen et al. 2019), which can overcome the (weak) mechanical pulling by the (long) loops of L1 = L3 = 2-nt and promote the structures of HIV-1_{Lai} DIS to be more straight (see Fig. 4C,D). However, for the HIV-2 TAR-TAR* complex with very short loops of L1 = L3 = 0-nt where the loop pulling effect is strong, the increase of salt concentration could reduce the (inter-helix) electrostatic repulsion induced by the loop-constrained inter-helix bending, which would indirectly strengthen the mechanical pulling effect of the very short loops. Such a salt-enhanced effect can be overwhelming the promotion on the coaxial stacking interaction due to higher salt, causing a stronger bending at high Na⁺/Mg²⁺ concentrations (see Fig. 4C,D; Supplemental Table S3).

Predicting thermal stability of RNA kissing complexes in salt solutions

Beyond 3D structure predictions, we used the present model to predict the thermal stability of extensive kissing complexes and to analyze the Na⁺/Mg²⁺ dependence of thermal stability of a typical kissing complex. Furthermore, the contribution of coaxial stacking to the stability of kissing complexes was explicitly analyzed through the present model.

Thermal stability of RNA kissing complexes

Based on the predicted structure ensemble at each temperature for a kissing complex, the fractions of unfolded state at different temperatures can be obtained, and afterward, the melting curve as well as the melting temperature T_m of the RNA kissing complex at an experimental strand concentration (e.g., 0.01 mM) can be calculated through Equations 2 and 3 in Materials and Methods. As an example shown in Figure 5, the predicted melting temperature of HIV-1 DIS type RNA kissing complex (sequence of kissing-interface duplex H2: CUAAAC/GAUUUG) at a strand concentration of 0.01 mM is $\sim 31.2^\circ\text{C}$, which agrees well with the experimental value ($\sim 32.0^\circ\text{C}$) (Weixlbaumer et al. 2004).

For a more extensive examination on thermal stability, we made predictions on the thermal stability for 15 RNA kissing complexes with different sequences or at different strand/ion conditions. As shown in Table 2, for all complexes, the predicted melting temperatures are very close to the experimental measurements (Gregorian and Crothers 1995; Muriaux et al. 1996; Weixlbaumer et al. 2004; Lorenz et al. 2006; Lebars et al. 2007, 2008; Van Melckebeke et al. 2008; Sehdev et al. 2012) with an overall

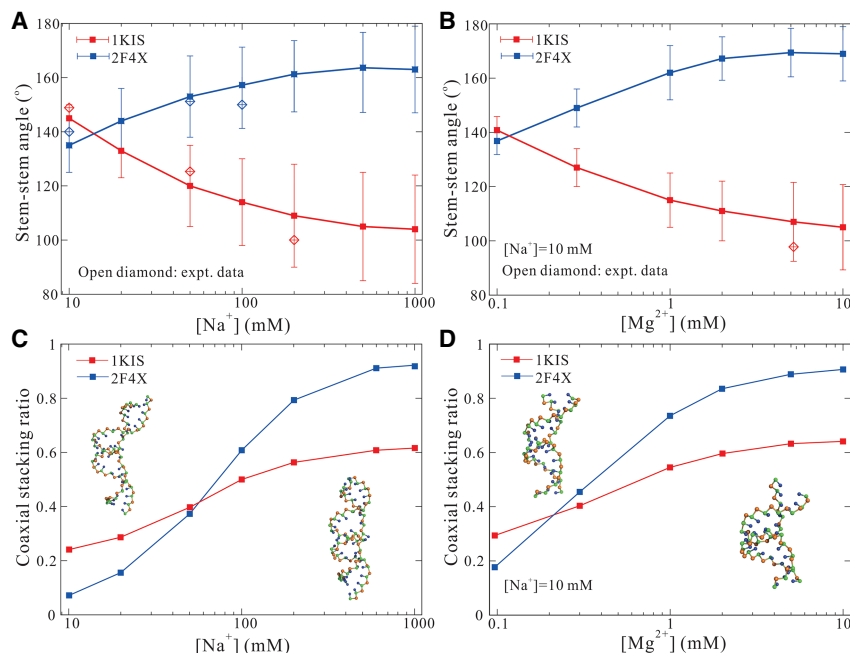


FIGURE 4. The predicted stem-stem (H1-H3) angles and coaxial stacking ratio of two typical kissing complexes as functions of $[\text{Na}^+]$ (A,C) and $[\text{Mg}^{2+}]$ (B,D): HIV-2 TAR-TAR* (PDB id: 1KIS) and HIV-1 DIS (PDB id: 2F4X) complexes. The coaxial stacking was determined through the method of RNAView (Yang et al. 2003), and the coaxial stacking ratio was calculated as the fraction of the conformations in which the coaxial stacking between H1 and H2 stems and H2 and H3 stems are both formed. The experimental data are taken from Chang and Tinoco (1997), Mujeeb et al. (1998), Baba et al. (2005), Lebars et al. (2008), and Van Melckebeke et al. (2008). The 3D structures in panels C (left, 2F4X at 10 mM Na^+ ; right, 2F4X at 1000 mM Na^+) and D (left, 1KIS at 0.1 mM Mg^{2+} ; right, 1KIS at 10 mM Mg^{2+}) are shown with PyMOL (<http://www.pymol.org>).

mean deviation of $\sim 2.4^\circ\text{C}$. This suggests that the present model with the sequence-dependent base-pairing/stacking potential and structure-based electrostatic potential can make successful predictions on the stability for RNA kissing complexes of extensive sequences in ion solutions.

the strong electrostatic neutralization at very high $[\text{Mg}^{2+}]$; see Figure 6B. This is because the bindings of Na^+ and Mg^{2+} are generally anti-cooperative and Mg^{2+} -binding is more efficient in stabilizing RNA kissing complexes with compact tertiary structures (Draper et al. 2005; Woodson

Effects of monovalent and divalent ions on stability

Since ions can play important roles in stabilizing RNA tertiary structures, we examined the monovalent/divalent ion effects on the thermal stability for a typical RNA kissing complex, HIV-1 DIS type RNA kissing complex (sequence of kissing-interface duplex H2: CCGACC/GGCUGG). As shown in Figure 6A,B, the predicted melting temperatures as functions of $[\text{Na}^+]$ and $[\text{Mg}^{2+}]$ for the RNA complex are in good accordance with the experimental data (Weixlbaumer et al. 2004), suggesting that the present model can almost give the quantitative predictions on the melting temperatures of RNA kissing complexes over broad ranges of $\text{Na}^+/\text{Mg}^{2+}$ concentrations. It should be noted that with the combination with the TBI model (Tan and Chen 2006), the present model can capture the competition between monovalent and divalent ions on the RNA kissing complex stability. For example, at very low $[\text{Mg}^{2+}]$, Na^+ dominates the stability of the HIV-1 DIS type RNA kissing complex, while the increase of $[\text{Mg}^{2+}]$ enhances the stability significantly until

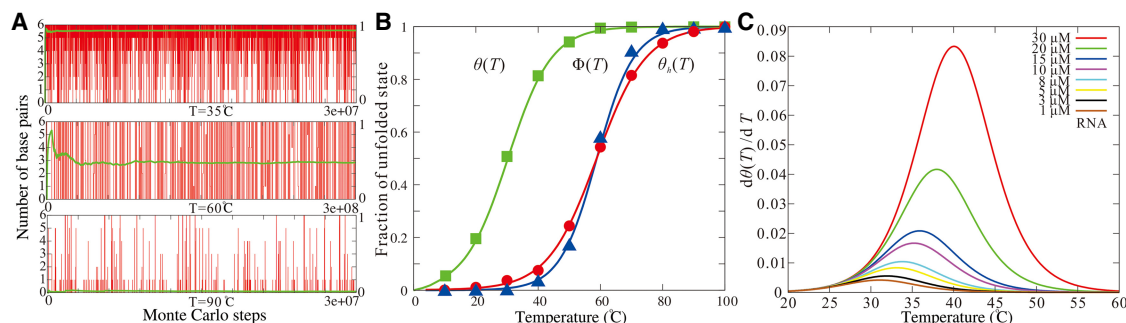


FIGURE 5. (A) The time-evolution of the number of base pairs (vertical red lines) and the average fractions of unfolded state (transverse green lines) for HIV-1 DIS type RNA kissing complex (sequence 4 in Table 3) at different temperatures. (B) The fractions of the unfolded state as functions of temperature for the HIV-1 DIS type RNA kissing complex. (Symbols) The predicted data at different temperatures; blue triangle: $\Phi(T)$ at high strand concentration of 1 mM; red circle: corrected value is $\theta_h(T)$; green square: $\theta(T)$ at experimental strand concentration (0.01 mM) derived through Equation 2 (Jin et al. 2018). (Lines) The fitted melting curve to the predicted data through Equation 3. More details can be found in the Supplemental Material. (C) The first derivative of $\theta(T)$ with the temperature at different strand concentrations.

TABLE 2. Melting temperature prediction for RNA kissing complexes

	Origin ^a	Kissing-interface sequences ^b	RNA/[Na ⁺]/[Mg ²⁺] ^c (μ M/mM/mM)	Kissing complex T_m		
				Prediction (°C)	Expt. ^d (°C)	Deviation (°C)
1	HIV-1 DIS type	AACCUGCCA AGGACGGAA	10/1000/0	65.7	67.3	−1.6
2	HIV-1 DIS type	AACCGACCA AGGCUUGAA	10/1000/0	60.9	64.7	−3.8
3	HIV-1 DIS type	AAGAGAGGA ACUCUCCAA	10/1000/0	57.6	57.0	+0.6
4	HIV-1 DIS type	AACUAAACA AGAUUUGAA	10/1000/0	31.2	32.0	−0.8
5	HIV-1 DIS type	ACCGACCA AGGCUUGA	2/1000/0	55.1	57.4	−2.3
6	HIV-1 DIS type	CCGACC GGCUGG	2/1000/0	54.6	58.0	−3.4
7	HIV-1 DIS _{Lai} dimer	AACGCGCGA AGCGCGCAA	0.8/100/0	40.2	37	+3.2
8	HIV-1 TAR-aptamer	UCCCAG CUGGGA	1/160/0.3	33.5	31.3	+2.2
9	HIV-1 TAR-aptamer	UCCCAG CUGGGA	1/160/3	49.8	47.3	+2.5
10	HIV-2 TAR-aptamer	UCCCAG CUGGGA	2/50/0	25.8	27.2	−1.4
11	HIV-2 TAR-aptamer	UCCCAG CUGGGA	2/50/0.3	29.7	32.5	−2.6
12	<i>E. coli</i> R1w-R2w	UUGGUAG AACCAUC	2/50/5	38.6	40	−1.4
13	<i>E. coli</i> R1i-R2i	GAUGGUU CUACCAA	2/50/5	55.8	59	−3.2
14	<i>E. coli</i> R1i-R2i	GAUGGUU CUACCAA	4/850/0	71.5	75.3	−3.8
15	<i>E. coli</i> R1i-R2i	GAUGGUU CUACCAA	4/10/0.7	56.1	53	+3.1

^aThe biological origin of the RNA kissing complexes.^bTwo sequences of the two hairpin loops at the kissing interface in complementary alignment.^cThe RNA strand concentrations, Na⁺ concentrations, and Mg²⁺ concentrations for the predictions and the corresponding experiments.^dThe experimental data are from Gregorian and Crothers (1995), Muriaux et al. (1996), Weixlbaumer et al. (2004), Lorenz et al. (2006), Lebars et al. (2007, 2008), Van Melckebeke et al. (2008), and Sehdev et al. (2012).

2005; Koculi et al. 2007; Chen 2008; Wong and Pollack 2010; Lipfert et al. 2014; Mustoe et al. 2014).

It is interesting to make comparisons between kissing complex and the corresponding kissing-interface duplex H2 on the salt-dependence of stability. As shown in Figure 6A,B, the kissing complex has higher stability than the duplex H2 at high salts, while it has apparently lower stability than the duplex H2 at low salts. This causes that the Na⁺/Mg²⁺-dependence of the kissing complex stability is apparently stronger than that of the duplex. The predicted salt (Na⁺ and Mg²⁺)-dependence of T_m calculated as $\Delta T_m/\Delta \log[\text{Na}^+]$ in the range of 50–500 mM and $\Delta T_m/\Delta \log[\text{Mg}^{2+}]$ in the range of 0.01–1 mM are $\sim 24^\circ\text{C}$ and $\sim 15^\circ\text{C}$ for the kissing complex, and are $\sim 10^\circ\text{C}$ and $\sim 5^\circ\text{C}$ for the corresponding kissing-interface duplex, respectively. The stronger salt (Na⁺ and Mg²⁺)-dependence of T_m and

the apparently lower T_m at low salts for the kissing complex than those for the corresponding kissing-interface duplex are because the folding of the kissing complex involves a much higher charge buildup than the duplex H2 (Tan and Chen 2011; Xi et al. 2018).

Contribution of coaxial stacking to stability

For a deep understanding on the relative stability of kissing complex (to the corresponding kissing-interface duplex), we would analyze the contribution of coaxial stacking in the following. For four complexes, we made additional calculations through the present model without involving the coaxial stacking potential. As shown in Table 3, the melting temperatures of the kissing complexes decrease by a mean value of $\sim 2.4^\circ\text{C}$ due to removing the coaxial

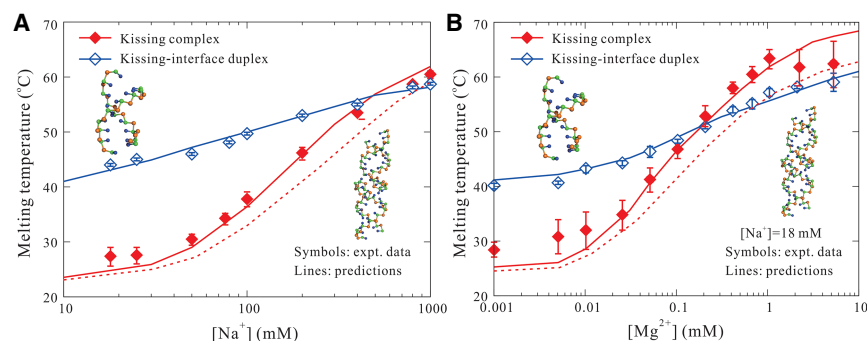


FIGURE 6. The predicted melting temperatures T_m (solid lines) and corresponding experimental data (symbols) (Weixlbaumer et al. 2004) as functions of $[Na^+]$ (A) and $[Mg^{2+}]$ (B) for the HIV-1 DIS type kissing complex (sequence of helical stem H2: CCGACC/GGCUGG) and its corresponding kissing-interface duplex. The dashed lines in panels A and B are the predictions from the present model without involving the coaxial stacking potential. The 3D structures in panels A and B demonstrate the kissing-interface duplex (left in A and B) and kissing complex (right in A and B), which are shown with PyMol (<http://www.pymol.org>).

stacking, suggesting the important role of coaxial stacking in stabilizing RNA kissing complexes. Furthermore, the melting temperatures from the calculations with removing the coaxial stacking potential are still slightly higher than those of the corresponding kissing-interface duplexes by a mean value of $\sim 1.3^\circ\text{C}$. This indicates that coaxial stacking makes the major contribution to the higher stability of kissing complex than the kissing-interface duplex (at high salts). However, the slightly higher melting temperatures without coaxial stacking than the corresponding kissing-interface duplexes may be attributed to that the formation of the kissing-interface duplex H2 from two folded hairpins loses less entropy than the isolated duplex from two unfolded single strands (Tan and Chen 2011). Therefore, the higher stability of kissing complex than the corresponding kissing-interface duplex is mainly attributed to the coaxial stacking between H1–H2 and between H2–H3 stems.

Since salt can modulate the contribution of coaxial stacking to the 3D structures of RNA kissing complexes, it is also necessary to examine the effect of coaxial stacking on the Na^+/Mg^{2+} -dependence of the kissing complex stability. We performed additional predictions by our model with removing the coaxial stacking potential for the HIV-1 DIS type RNA kissing complexes at various Na^+/Mg^{2+} concentrations. As shown in Figure 6, at low Na^+/Mg^{2+} concentrations, the predicted melting temperatures from the model without coaxial stacking potential are very close to those with involving the coaxial stacking potential. It is because the coaxial stacking can be disrupted by the (inter-helix)

electrostatic repulsion due to low Na^+/Mg^{2+} concentrations as discussed above. As Na^+/Mg^{2+} concentrations increase, the calculated melting temperatures for the kissing complex by the present model with/without coaxial stacking both increase rapidly, but the Na^+/Mg^{2+} -dependence for the melting temperature without coaxial stacking is not as strong as that with the coaxial stacking. This is reasonable. Higher ion concentrations can enhance the charge neutralization to phosphates and consequently can promote the coaxial stacking in the kissing complex, which causes that the kissing complex has visibly higher stability than that without involving coaxial stacking at high salts.

Thermally folding pathways of RNA kissing complexes

Since the metastable structures of RNAs could be important to their functions (Huthoff and Berkout 2002; Baker

TABLE 3. Melting temperature predictions for RNA kissing complexes and corresponding kissing-interface duplexes

Kissing-interface sequences ^a	RNA/ $[Na^+]$ / $[Mg^{2+}]^b$ ($\mu\text{M}/\text{mM}/\text{mM}$)	Kissing complex T_m			Kissing-interface duplex T_m	
		Prediction ($^\circ\text{C}$)	Expt. ^c ($^\circ\text{C}$)	Prediction ^d ($^\circ\text{C}$)	Expt. ^c ($^\circ\text{C}$)	
CCUGCC	10/1000/0	65.7	67.3	63.5	61.6	
GGACGG						
CCGACC	10/1000/0	60.9	64.7	58.6	58.6	
GGCUGG						
GAGAGG	10/1000/0	57.6	57.0	54.9	53.5	
CUCUCC						
CUAAAC	10/1000/0	31.2	32.0	28.7	26.9	
GAUUUG						

^aTwo sequences of the two hairpin loops at the kissing interface in complementary alignment.

^bThe RNA strand concentration, Na^+ concentration, and Mg^{2+} concentration for the prediction and the corresponding experiments.

^cThe experimental data are from Weixlbaumer et al. (2004).

^dThe predictions from the model without involving the coaxial stacking potential.

et al. 2003; Atkins et al. 2011; Plumridge et al. 2018; Wang et al. 2018), the thermally folding pathways of RNA kissing complexes would be crucial for understanding their functions (Muriaux et al. 1996; Bernacchi et al. 2007; Mujeeb et al. 2007; Tam et al. 2007; Lee et al. 2009; Cao and Chen 2011b; Livshits et al. 2011). To explore thermally folding pathways of RNA kissing complexes, we made further analyses for three typical sequences of RNA kissing complexes, based on the extensive examination of the folding structures in the annealing simulations for the 18 RNA kissing complex sequences used for the structure prediction. According to the base-pairing states of predicted structures at different temperatures, there are mainly four possible states for typical RNA kissing complexes: dimer of kissing complex (K, at least two continuous kissing base pairs between two folded hairpins), dimer of extended duplex (D), monomer of two hairpins (H, at least one hairpin formed), and monomer of two unfolded single strands (S) (Jin et al. 2018). Although RNA helices can assemble to nanoparticles through blunt-end coaxial stacking (Dibrov et al. 2011), the state of two hairpins with coaxial stacking between blunt ends almost never appears in our simulations. The major reason may be that the stability of coaxial stacking of hairpin blunt ends (e.g., ~ -2 kcal/mol for two G-C ends) is much weaker than that of a kissing interface with base-pairing/stacking and coaxial stacking (e.g., ~ -15 kcal/mol for a kissing complex with three kissing base pairs) (Walter and Turner 1994; Walter et al. 1994; Xia et al. 1998).

For three typical RNA kissing complexes (HIV-2 TAR-TAR*, HIV-1_{Lai} DIS and HIV-1_{Mal} DIS), the fractions of four states at different temperatures were calculated as functions of temperatures, and were shown in Figure 7 where the sequences and predicted structures of the four complexes can also be found. For the HIV-2 TAR-TAR* kissing complex, as temperature is decreased from a high temperature (e.g., $\sim 120^\circ\text{C}$), the fraction of S state gradually decreases, and simultaneously the fraction of H state increases gradually and reaches its maximum value at $\sim 50^\circ\text{C}$, indicating that the two hairpins would fold first. With the further decrease of temperature, the fraction of K state increases accompanied with the decrease of the fraction of H state, suggesting the formation of the kissing complex at low temperatures (e.g., $\sim 20^\circ\text{C}$). Thus, the thermally folding pathway of the HIV-2 TAR-TAR* kissing complex follows the order of $S \rightarrow H \rightarrow K$ (see the illustration in Fig. 7D), which is in accordance with the previous experiments (Ennifar et al. 2001; Lebars et al. 2007, 2008; Van Melckebeke et al. 2008). In contrast, for the HIV-1 DIS kissing complexes (e.g., Lai and Mal), the fractions of K and D states both increase accompanied with the decrease of the fraction of H state when temperature is decreased from $\sim 70^\circ\text{C}$ (see Fig. 7B,C). This suggests that there are two major folding pathways ($S \rightarrow H \rightarrow K$ and $S \rightarrow H \rightarrow D$) and one very minor pathway ($S \rightarrow D$) for the two HIV-1 DIS kissing complexes, which has been suggested previously by sever-

al experiments (Muriaux et al. 1995b; Bernacchi et al. 2007; Mujeeb et al. 2007; Tam et al. 2007; Lee et al. 2009; Livshits et al. 2011). Such coexistence of both kissing complex and extended duplex has been proposed to be necessary for the dimerization process of the HIV-1 RNA genome (Laughrea and Jetté 1996; Paillart et al. 1996). In addition, at low temperature, the fraction of D state is comparable to that of K state for HIV-1 DIS_{Lai} complex, while for HIV-1 DIS_{Mal} complex, the fraction of D state is visibly higher than that of K state. In addition, our predictions showed that the thermally folding pathways of the 18 RNA kissing complex sequences used for the structure prediction can be classified into the three typical types described above.

Why do RNA kissing complexes with different sequences have very different folding pathways? To examine what dominates the folding pathway of the complexes, we further calculated the free energies of the states for the three typical complexes at different temperatures and 1 M [Na⁺] using the Vfold2D algorithm (see Supplemental Fig. S4; Cao and Chen 2011b). For the HIV-2 TAR-TAR* kissing complex, the two possible folded states (K and D states) are sufficiently different in stability, and the K state appears significantly more stable than the D state especially at low temperatures with the relative-free energy $\Delta\Delta G = \Delta G_K - \Delta G_D$ between two states of ~ -15 kcal/mol at 25°C . This should be the reason that the D state never appears in the folding process of the complex and there is only one folding pathway ($S \rightarrow H \rightarrow K$) for the HIV-2 TAR-TAR* complex. However, for the two HIV-1 DIS kissing complexes, the two (K and D) states are comparable in stability with the relative free energy $\Delta\Delta G$ of 0.4–1.6 kcal/mol, which would lead to the two major folding pathways in parallel. Due to the difference between the sequences in the kissing-interface for the two complexes (HIV-1_{Lai} DIS: CGCGCG and HIV-1_{Mal} DIS: GUGCAC), the relative free energy $\Delta\Delta G = \Delta G_K - \Delta G_D = \sim 0.4$ kcal/mol for the HIV-1_{Lai} DIS complex is much smaller than that ($\Delta\Delta G = \sim 1.6$ kcal/mol) for the HIV-1_{Mal} DIS complex, which would result in the slightly higher fraction of K state for the HIV-1_{Lai} DIS complex and the apparently higher fraction of K state (than D state) for the HIV-1_{Mal} DIS complex at low temperatures. In addition, since D state can be formed directly from S state and has slightly higher stability than K state, there is a minor pathway of $S \rightarrow D$. Therefore, the above analyses for the three complexes show that the thermally folding pathway of an RNA complex sequence is mainly dependent on the relative stability of the folded states (K state and D state), which can be significantly modulated by the sequence (Cho et al. 2009; Chen et al. 2018; Shi et al. 2018).

Conclusions

The knowledge of 3D structures and thermal stability of RNA kissing complexes in ion solutions is crucial for understanding their biological functions. In this work, we used

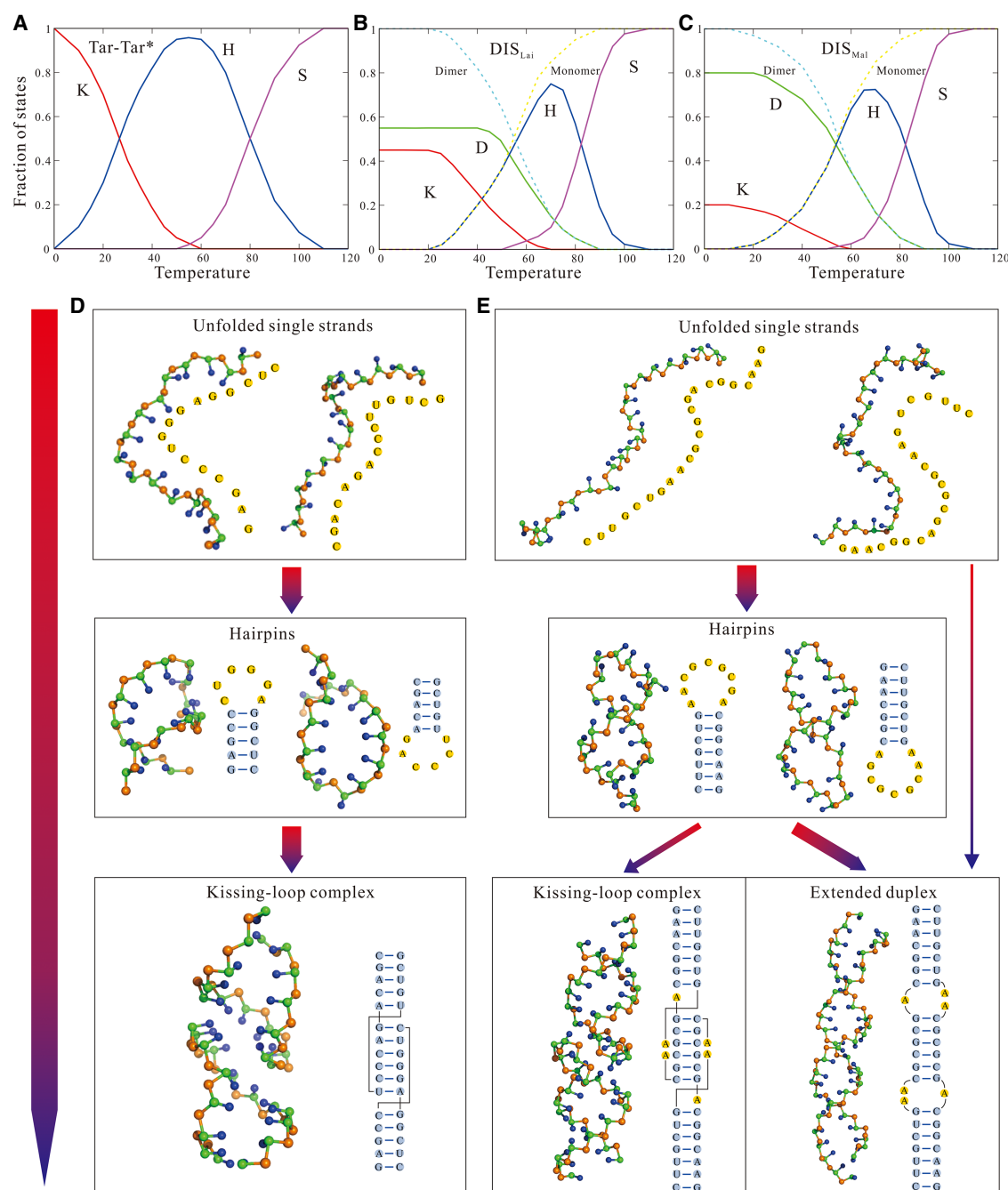


FIGURE 7. (A) The state fractions of kissing complex K, hairpins H, and single-stranded chains S as functions of temperature for the thermal folding of HIV-2 TAR-TAR* (PDB id: 1KIS). (B,C) The state fractions of kissing complex K, extended duplex D, hairpins H, and single-stranded chains S as functions of temperature for the thermal folding of HIV-1_{Lai} DIS (B, PDB id: 2B8R) and HIV-1_{Mal} DIS (C, PDB id: 2B8S) complexes. In panels B and C, the dashed lines in cyan and yellow represent the fractions of dimers ($K + D$) and monomers ($H + S$), respectively. (D,E) The schematic diagrams show the structural transitions of HIV-2 TAR-TAR* (D) and HIV-1_{Lai} kissing complex (E) along the thermally folding pathways. The ball-stick graphs represent the 3D structure and the corresponding secondary structure of each state. The left down arrow denotes the direction of decreasing temperatures.

our CG model with the implicit structure-based electrostatic potential to predict 3D structures, thermal stability, and folding pathways for RNA kissing complexes in monovalent/divalent ion solutions. The following are the major conclusions:

1. For extensive RNA kissing complexes in monovalent/divalent ion solutions, we successfully predicted the native-like 3D structures with overall mean (minimum) RMSD of 5.4 Å (2.7 Å) from their sequences, and found that the coaxial stacking interaction and salt condition

can play important roles in maintaining their 3D structures.

- For various RNA kissing complexes at different strand/ion conditions, we made reliable predictions on their thermal stability with an overall mean deviation of $\sim 2.4^\circ\text{C}$ in melting temperature from the experimental data, and our analyses showed that ions, especially divalent ions, can promote coaxial stacking and consequently stabilize RNA kissing complexes more strongly at higher salts. Such coaxial stacking makes the major contribution to the higher stability of kissing complex than the corresponding kissing-interface duplex at high salts.
- Our comprehensive analyses on thermally folding pathways of typical RNA kissing complexes revealed that the folding pathways as well as final folded structures (kissing complex or extended duplex) are mainly determined by the relative stability between different states, which could be significantly modulated by their sequences.

Although the present model can predict 3D structure/thermal stability for RNA kissing complexes with relatively high accuracy and the role of coaxial stacking/thermally folding pathways was analyzed thoroughly, there are still several limitations that can be improved in the future. First, this work only involved hairpin–hairpin kissing-loop complexes in the predictions, and other types of kissing-loop complexes such as hairpin–bulge/internal kissing-loops were not covered. Fortunately, the present model as well as the analyses can be conveniently extended to examine other types of kissing-loop complex structures beyond hairpin–hairpin kissing-loop complexes. Second, the present model only includes the interactions between canonical Watson–Crick and wobble base pairs and consequently ignores the noncanonical base pairs (e.g., A–G and A–A base pair) which could appear frequently in the complicated RNA tertiary structures (Lemieux and Major 2002). The noncanonical base-pairing may make an additional contribution to the 3D structure and stability of some RNA kissing complexes. At present, our CG model has been successfully used to make structure prediction and stability analyses for small RNAs including hairpins, duplexes, and pseudoknots and kissing complexes in ion solutions. However, it is still difficult for the model to reliably predict structures and stability of large RNAs with more complicated structures (e.g., riboswitches and ribozymes) in ion solutions, since these RNAs generally involve a lot of complex segments/interactions such as multi-branched loops, noncanonical base pairs, base triple interactions, base–backbone hydrogen-bond interactions and possible specific ion binding. These complicated structure segments/interactions are required to be taken into account in the further development of the model. Finally,

the predicted RNA 3D structures from the model are still at CG level, and it is still very necessary to rebuild all-atom structures based on the predicted CG ones for more strict examination and possible practical use.

Nevertheless, the model can well predict the 3D structures and stabilities of extensive RNA kissing complexes and can well capture the effects of ions on their 3D structure and stability. Furthermore, our comprehensive analyses suggested the important contribution of coaxial stacking to the 3D structure and stability, and revealed the assembly mechanism for the folding of RNA kissing complexes. Therefore, this work can also be very helpful for understanding the roles of ions and coaxial stacking in stabilizing the 3D structure and the structure assembly mechanism for the folding of RNA kissing complexes.

MATERIALS AND METHODS

Coarse-grained RNA structure representation

In our model, to reduce computational complexity and retain the major structure features of RNAs, each nucleotide was simplified into three CG spherical beads: phosphate (P) bead, sugar ring (C) bead, and base (N) bead with van der Waals radii of 1.9, 1.7, and 2.2 Å, respectively (see Fig. 1A; Supplemental Fig. S1; Shi et al. 2014a, 2015, 2018; Jin et al. 2018). The three (P, C, and N) beads were placed at the positions of P, C4', and N1 for pyrimidine (or N9 for purine) atoms, respectively. Each P bead has a charge of $-e$ on its center (Shi et al. 2014a, 2015, 2018; Jin et al. 2018).

Coarse-grained force field

The total potential energy (U_{total}) of an RNA CG conformation is composed of two contributions of bonded potential (U_{bond}) and nonbonded potential (U_{nonbond}) (Shi et al. 2014a, 2015, 2018; Jin et al. 2018):

$$U_{\text{total}} = U_{\text{bond}} + U_{\text{nonbond}} \quad (1)$$

The bond potential in Equation 1 accounts for the energy associated with virtual bonds in the CG structure, and consists of bond-length energy U_b , bond-angle energy U_a , and dihedral energy U_d . The nonbonded potential in Equation 1 describes excluded volume interaction U_{exe} , base-pairing interaction U_{bp} , base stacking interaction U_{bs} , coaxial stacking interaction U_{cs} , and effective electrostatic interaction U_{el} between negatively charged P beads. The coaxial stacking interaction between two discontinued RNA double-helical stems has been found of crucial importance in stabilizing the structures of RNA kissing complexes (Marino et al. 1995; Chang and Tinoco 1997; Lee and Crothers 1998; Ennifar et al. 2001, 2006; Baba et al. 2005; Ennifar and Dumas 2006; Lebars et al. 2007, 2008; Van Melckebeke et al. 2008; Li and Tinoco 2009; Mundigala et al. 2014). As illustrated in Figure 1B, a typical RNA kissing complex contains three helical stems (H1, H2, and H3) and four loops (L1, L2, L3, and L4), in which H2 is formed by the interstrand base-pairing/stacking between two hairpin loops. Generally, H1 and H2 as well as H2 and H3 are coaxially stacked to keep the whole kissing complex as a

quasi-continuous helix (Chen and Varani 2001). In the present model, a slight modification was made to more accurately describe the stacking interaction between the two ends of the helical stems at the kissing interface, since our previous examination has shown that the U_{cs} in our previous model underestimates the coaxial stacking interaction range (Jin et al. 2018). Such slight modification on U_{cs} has been described in detail in the [Supplemental Material](#). In addition, as described previously (Jin et al. 2018), an implicit structure-based electrostatic potential was involved in capturing the contribution of monovalent/divalent salt to RNA 3D structure and stability, based on the combination of counter ion condensation theory (Manning 1978) and tightly bound ion (TBI) model (Tan and Chen 2006, 2010, 2011; Xi et al. 2018). The involvement of the electrostatic potential would enable our model to predict the effect of monovalent/divalent ions and the effect of the competition between monovalent and divalent ions on the 3D structures and stability of RNAs (Shi et al. 2015; Jin et al. 2018). The detailed description of the potential components in Equation 1 can be found in the [Supplemental Material](#).

Simulation procedure

The Monte Carlo (MC) simulated annealing algorithm was used to predict 3D structures of RNA kissing complexes in ion solutions based on the above described CG force field. Following the previously developed simulation method (Jin et al. 2018), two initial unfolded CG chains randomly generated based on sequences were separately placed in a cubic box, the size of which was determined by the strand concentration. The simulation system was gradually cooled from a high temperature (e.g., 120°C) to the target low temperature (e.g., 25°C). At each temperature, the RNA conformations were renovated by intra-strand pivot moves and inter-strand translation/rotation through the Metropolis algorithm until the system reached enough equilibrium (Jin et al. 2018).

For some sequences of RNA kissing complexes, there may be two possible folded states with comparable folding free energies (Huthoff and Berkhout 2002; Cao and Chen 2011b), for example, the HIV-1 DIS sequence can fold to kissing complex and extended duplex which are coexistent in vivo and vitro (Muriaux et al. 1996; Paillart et al. 1996; Bernacchi et al. 2007; Mujeeb et al. 2007; Tam et al. 2007; Lee et al. 2009; Livshits et al. 2011). In this work, we generally implemented multitrajectory simulations to constitute the conformation ensemble at each temperature. By analyzing structure ensemble in the order of interchain/intrachain base pairs, the 3D structures at different temperatures and the thermodynamic properties for a complex sequence can be obtained (Jin et al. 2018).

Calculations for melting temperatures

The stability of RNA complexes generally depends on the strand concentration due to the contribution of translation entropy of RNA strands (Privalov and Crane-Robinson 2018). In order to improve the simulation efficiency for the systems at each temperature, we generally performed the MC simulations for RNA complexes at a relatively high strand concentration C_s^h (e.g., 1 mM) instead of low experimental strand concentrations C_s (e.g., 0.01 mM) (Jin et al. 2018). Based on the equilibrium confor-

mations at each temperature T , the fraction $\theta_h(T)$ of unfolded state at C_s^h can be obtained (Jin et al. 2018). After a correction for finite-box-size effect at the high bulk strand concentration C_s^h (Ouldrige et al. 2010), the fraction $\theta(T)$ of unfolded state at an experimental strand concentration C_s can be calculated by Jin et al. (2018):

$$\theta(T) = \frac{\lambda \theta_h(T)}{1 + \lambda \theta_h(T) - \theta_h(T)}, \quad (2)$$

where $\lambda = C_s^h/C_s$. Afterward, the melting temperature T_m for a complex can be obtained by fitting the fraction $\theta(T)$ to the two-state model (Shi et al. 2014a),

$$\theta(T) = 1 - \frac{1}{1 + e^{(T-T_m)/dT}}, \quad (3)$$

where dT is an adjustable parameter. More details about the calculation of melting temperature were given in the [Supplemental Material](#).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We are grateful to Professors Shi-Jie Chen (University of Missouri), Xiangyun Qiu (The 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 (11774272, 11605125, and 11575128) and the Department of Education of Hubei Province (Q20171607). Parts of the numerical calculation in this work were performed on the super computing system in the Super Computing Center of Wuhan University.

Author contributions: Z.J.T. and L.J. designed the research; L.J. and Y.Z.S. performed the simulations; Z.J.T., L.J., Y.L.T., Y.W., and X.W. analyzed the data; L.J., Y.Z.S., and Z.J.T. wrote the manuscript. All authors discussed the results and reviewed the manuscript.

Received April 30, 2019; accepted August 2, 2019.

REFERENCES

- Andersen AA, Collins RA. 2001. Intramolecular secondary structure rearrangement by the kissing interaction of the *Neurospora* VS ribozyme. *Proc Natl Acad Sci* **98**: 7730–7735. doi:10.1073/pnas.141039198
- Androneanu M, Zhang ZC, Condon A. 2005. Secondary structure prediction of interacting RNA molecules. *J Mol Biol* **345**: 987–1001. doi:10.1016/j.jmb.2004.10.082
- Atkins J, Gesteland RF, Cech TR. 2011. *RNA worlds: from life's origins to diversity in gene regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Baba S, Takahashi KI, Noguchi S, Takaku H, Koyanagi Y, Yamamoto N, Kawai G. 2005. Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J Biochem* **138**: 583–592. doi:10.1093/jb/mvi158

- Baker TA, Watson JD, Bell SP, Gann A, Losick MA, Levine R. 2003. *Molecular biology of the gene*. Benjamin-Cummings Publishing Company, San Francisco, CA.
- Bernacchi S, Freisz S, Maechling C, Spiess B, Marquet R, Dumas P, Ennifar E. 2007. Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion. *Nucleic Acids Res* **35**: 7128–7139. doi:10.1093/nar/gkm856
- Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**: 474. doi:10.1186/1471-2105-9-474
- Bian Y, Zhang J, Wang J, Wang J, Wang W. 2015. Free energy landscape and multiple folding pathways of an H-type RNA pseudoknot. *PLoS One* **10**: e0129089. doi:10.1371/journal.pone.0129089
- Boniewicz MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, Rother KM, Bujnicki JM. 2015. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res* **44**: e63. doi:10.1093/nar/gkv1479
- Bouchard P, Legault P. 2014a. Structural insights into substrate recognition by the *Neurospora* Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction. *Biochemistry* **53**: 258–269. doi:10.1021/bi401491g
- Bouchard P, Legault P. 2014b. A remarkably stable kissing-loop interaction defines substrate recognition by the *Neurospora* Varkud Satellite ribozyme. *RNA* **20**: 1451–1464. doi:10.1261/ma.046144.114
- Boudard M, Barth D, Bernauer J, Denise A, Cohen J. 2017. GARN2: coarse-grained prediction of 3D structure of large RNA molecules by regret minimization. *Bioinformatics* **33**: 2479–2486. doi:10.1093/bioinformatics/btx175
- Cao S, Chen SJ. 2011a. Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* **115**: 4216–4226. doi:10.1021/jp112059y
- Cao S, Chen SJ. 2011b. Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. *RNA* **17**: 2130–2143. doi:10.1261/ma.026658.111
- Cao S, Xu X, Chen SJ. 2014. Predicting structure and stability for RNA complexes with intermolecular loop-loop base-pairing. *RNA* **20**: 835–845. doi:10.1261/ma.043976.113
- Chang KY, Tinoco I Jr. 1997. The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J Mol Biol* **269**: 52–66. doi:10.1006/jmbi.1997.1021
- Chen SJ. 2008. RNA folding: combinatorial statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* **37**: 197–214. doi:10.1146/annurev.biophys.37.032807.125957
- Chen Y, Varani G. 2001. RNA structure. In *eLS*. Wiley, NY. doi:10.1002/9780470015902.a0001339.pub2
- Chen YL, Sutton JL, Pollack L. 2018. How the conformations of an internal junction contribute to fold an RNA domain. *J Phys Chem B* **122**: 11363–11372. doi:10.1021/acs.jpcc.8b07262
- Chen YL, Lee T, Elber R, Pollack L. 2019. Conformations of an RNA helix-junction-helix construct revealed by SAXS refinement of MD simulations. *Biophys J* **116**: 19–30. doi:10.1016/j.bpj.2018.11.020
- Cho SS, Pincus DL, Thirumalai D. 2009. Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures. *Proc Natl Acad Sci* **106**: 17349–17354. doi:10.1073/pnas.0906625106
- Cragolini T, Derreumaux P, Pasquali S. 2013. Coarse-grained simulations of RNA and DNA duplexes. *J Phys Chem B* **117**: 8047–8060. doi:10.1021/jp400786b
- Cragolini T, Derreumaux P, Pasquali S. 2015. *Ab initio* RNA folding. *J Phys Condens Matter* **27**: 233102. doi:10.1088/0953-8984/27/23/233102
- Das R, Baker D. 2007. Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci* **104**: 14664–14669. doi:10.1073/pnas.0703836104
- Dawson WK, Maciejczyk M, Jankowska EJ, Bujnicki JM. 2016. Coarse-grained modeling of RNA 3D structure. *Methods* **103**: 138–156. doi:10.1016/j.ymeth.2016.04.026
- Denesyuk NA, Thirumalai D. 2013. Coarse-grained model for predicting RNA folding thermodynamics. *J Phys Chem B* **117**: 4901–4911. doi:10.1021/jp401087x
- Denesyuk NA, Hori N, Thirumalai D. 2018. Molecular simulations of ion effects on the thermodynamics of RNA folding. *J Phys Chem B* **122**: 11860–11867. doi:10.1021/acs.jpcc.8b08142
- Dibrov SM, McLean J, Parsons J, Hermann T. 2011. Self-assembling RNA square. *Proc Natl Acad Sci* **108**: 6405–6408. doi:10.1073/pnas.1017999108
- Dimitrov RA, Zuker M. 2004. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J* **87**: 215–226. doi:10.1529/biophysj.103.020743
- Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. 2008. *Ab initio* RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* **14**: 1164–1173. doi:10.1261/ma.894608
- Draper DE, Grilley D, Soto AM. 2005. Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **34**: 221–243. doi:10.1146/annurev.biophys.34.040204.144511
- Dubois N, Marquet R, Paillart JC, Bernacchi S. 2018. Retroviral RNA dimerization: from structure to functions. *Front Microbiol* **9**: 527. doi:10.3389/fmicb.2018.00527
- Ennifar E, Dumas P. 2006. Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J Mol Biol* **356**: 771–782. doi:10.1016/j.jmb.2005.12.022
- Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. 2001. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Mol Biol* **8**: 1064–1068. doi:10.1038/nsb727
- Ennifar E, Paillart JC, Bodlender A, Walter P, Weibel JM, Aubertin AM, Pale P, Dumas P, Marquet R. 2006. Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res* **34**: 2328–2339. doi:10.1093/nar/gkl317
- Gamache ER, Doh JH, Ritz J, Laederach A, Bellaousov S, Mathews DH, Curcio MJ. 2017. Structure-function model for kissing loop interactions that initiate dimerization of Ty1 RNA. *Viruses* **9**: 93. doi:10.3390/v9050093
- Gregorian RS Jr, Crothers DM. 1995. Determinants of RNA hairpin loop-loop complex stability. *J Mol Biol* **248**: 968–984. doi:10.1006/jmbi.1995.0275
- Hajdin CE, Ding F, Dokholyan NV, Weeks KM. 2010. On the significance of an RNA tertiary structure prediction. *RNA* **16**: 1340–1349. doi:10.1261/ma.1837410
- He Y, Maciejczyk M, Ołdziej S, Scheraga HA, Liwo A. 2013. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys Rev Lett* **110**: 098101. doi:10.1103/PhysRevLett.110.098101
- Hori N, Denesyuk NA, Thirumalai D. 2016. Salt effects on the thermodynamics of a frameshifting RNA pseudoknot under tension. *J Mol Biol* **428**: 2847–2859. doi:10.1016/j.jmb.2016.06.002
- Hori N, Denesyuk NA, Thirumalai D. 2019. Ion condensation onto ribozyme is site-specific and fold-dependent. *Biophys J* **116**: 2400–2410. doi:10.1016/j.bpj.2019.04.037
- Huang FW, Qin J, Reidys CM, Stadler PF. 2009. Partition function and base pairing probabilities for RNA–RNA interaction prediction. *Bioinformatics* **25**: 2646–2654. doi:10.1093/bioinformatics/btp481

- Huang L, Wang J, Lilley DMJ. 2017. The structure of the guanidine-II riboswitch. *Cell Chem Biol* **24**: 695–702.e2. doi:10.1016/j.chembiol.2017.05.014
- Huthoff H, Berkhout B. 2002. Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry* **41**: 10439–10445. doi:10.1021/bi025993n
- Hyeon C, Thirumalai D. 2011. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* **2**: 487. doi:10.1038/ncomms1481
- Jain S, Schlick T. 2017. F-RAG: generating atomic coordinates from RNA graphs by fragment assembly. *J Mol Biol* **429**: 3587–3605. doi:10.1016/j.jmb.2017.09.017
- Jin L, Shi YZ, Feng CJ, Tan YL, Tan ZJ. 2018. Modeling structure, stability, and flexibility of double-stranded RNAs in salt solutions. *Biophys J* **115**: 1403–1416. doi:10.1016/j.bpj.2018.08.030
- Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. 2009. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* **15**: 189–199. doi:10.1261/rna.1270809
- Jossinet F, Paillart JC, Westhof E, Hermann T, Skripkin E, Lodmell JS, Ehresmann C, Ehresmann B, Marquet R. 1999. Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *RNA* **5**: 1222–1234. doi:10.1017/S1355838299990982
- Kieken F, Paquet F, Brule F, Paoletti J, Lancelot G. 2006. A new NMR solution structure of the SL1 HIV-1 Lai loop-loop dimer. *Nucleic Acids Res* **34**: 343–352. doi:10.1093/nar/gkj427
- Kim T, Shapiro BA. 2013. The role of salt concentration and magnesium binding in HIV-1 subtype-A and subtype-B kissing loop monomer structures. *J Biomol Struct Dyn* **31**: 495–510. doi:10.1080/07391102.2012.706072
- Kim CH, Tinoco I. 2000. A retroviral RNA kissing complex containing only two G-C base pairs. *Proc Natl Acad Sci* **97**: 9396–9401. doi:10.1073/pnas.170283697
- Kim N, Laing C, Elmetwaly S, Jung S, Curuksu J, Schlick T. 2014. Graph-based sampling for approximating global helical topologies of RNA. *Proc Natl Acad Sci* **111**: 4079–4084. doi:10.1073/pnas.1318893111
- Koculi E, Hyeon C, Thirumalai D, Woodson SA. 2007. Charge density of divalent metal cations determines RNA stability. *J Am Chem Soc* **129**: 2676–2682. doi:10.1021/ja068027r
- Kolb FA, Engdahl HM, Slagter-Jäger JG, Ehresmann B, Ehresmann C, Westhof E, Gerhart E, Wagner H, Romby P. 2000. Progression of a loop-loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA. *EMBO J* **19**: 5905–5915. doi:10.1093/emboj/19.21.5905
- Kolb FA, Westhof E, Ehresmann B, Ehresmann C, Wagner EGH, Romby P. 2001. Four-way junctions in antisense RNA-mRNA complexes involved in plasmid replication control: a common theme? *J Mol Biol* **309**: 605–614. doi:10.1006/jmbi.2001.4677
- Laughrea M, Jetté L. 1996. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248–271 are dispensable for dimer formation. *Biochemistry* **35**: 1589–1598. doi:10.1021/bi951838f
- Lebars I, Richard T, Di Primo C, Toulmé JJ. 2007. NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic Acids Res* **35**: 6103–6114. doi:10.1093/nar/gkm655
- Lebars I, Legrand P, Aimé A, Pinaud N, Fribourg S, Di Primo C. 2008. Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res* **36**: 7146–7156. doi:10.1093/nar/gkn831
- Lee AJ, Crothers DM. 1998. The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. *Structure* **6**: 993–1007. doi:10.1016/S0969-2126(98)00101-4
- Lee HW, Briggs KT, Marino JP. 2009. Dissecting structural transitions in the HIV-1 dimerization initiation site RNA using 2-aminopurine fluorescence. *Methods* **49**: 118–127. doi:10.1016/j.jymeth.2009.05.006
- Leipply D, Draper DE. 2011. Effects of Mg^{2+} on the free energy landscape for folding a purine riboswitch RNA. *Biochemistry* **50**: 2790–2799. doi:10.1021/bi101948k
- Lemieux S, Major F. 2002. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res* **30**: 4250–4263. doi:10.1093/nar/gkf540
- Li PT. 2013. Analysis of diffuse K^{+} and Mg^{2+} ion binding to a two-base-pair kissing complex by single-molecule mechanical unfolding. *Biochemistry* **52**: 4991–5001. doi:10.1021/bi400646x
- Li PT, Tinoco TI Jr. 2009. Mechanical unfolding of two DIS RNA kissing complexes from HIV-1. *J Mol Biol* **386**: 1343–1356. doi:10.1016/j.jmb.2009.01.023
- Li J, Zhang J, Wang J, Li W, Wang W. 2016. Structure prediction of RNA loops with a probabilistic approach. *PLoS Comput Biol* **12**: e1005032. doi:10.1371/journal.pcbi.1005032
- Li J, Zhu W, Wang J, Li W, Gong S, Zhang J, Wang W. 2018. RNA3DCNN: local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLoS Comput Biol* **14**: e1006514. doi:10.1371/journal.pcbi.1006514
- Lipfert J, Doniach S, Das R, Herschlag D. 2014. Understanding nucleic acid-ion interactions. *Annu Rev Biochem* **83**: 813–841. doi:10.1146/annurev-biochem-060409-092720
- Livshits MA, Kaluzhny DN, Minyat EE. 2011. Melting of model HIV-1 stem-loop 1 RNA dimers monitored by 2-aminopurine fluorescence. *J Biomol Struct Dyn* **28**: 939–947. doi:10.1080/07391102.2011.10508620
- Lorenz C, Piganeau N, Schroeder R. 2006. Stabilities of HIV-1 DIS type RNA loop-loop interactions in vitro and in vivo. *Nucleic Acids Res* **34**: 334–342. doi:10.1093/nar/gkj435
- Lu XJ, Bussemaker HJ, Olson WK. 2015. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* **43**: e142.
- Manning GS. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* **11**: 179–246. doi:10.1017/S0033583500002031
- Marino JP, Gregorian RS, Csankovszki G, Crothers DM. 1995. Bent helix formation between RNA hairpins with complementary loops. *Science* **268**: 1448–1454. doi:10.1126/science.7539549
- Martinez HM, Maizel JV Jr, Shapiro BA. 2008. RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn* **25**: 669–683. doi:10.1080/07391102.2008.10531240
- Miao Z, Westhof E. 2017. RNA structure: advances and assessment of 3D structure prediction. *Annu Rev Biophys* **46**: 483–503. doi:10.1146/annurev-biophys-070816-034125
- Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, Boniecki MJ, Bujnicki JM, Chen SJ, Cheng CY, et al. 2017. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23**: 655–672. doi:10.1261/rna.060368.116
- Miller WA, White KA. 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. *Annu Rev Phytopathol* **44**: 447–467. doi:10.1146/annurev.phyto.44.070505.143353
- Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG. 1998. Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Mol Biol* **5**: 432–436. doi:10.1038/nsb0698-432

- Mujeeb A, Ulyanov NB, Georgantis S, Smirnov I, Chung J, Parslow TG, James TL. 2007. Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acids Res* **35**: 2026–2034. doi:10.1093/nar/gkm097
- Mundigala H, Michaux JB, Feig AL, Ennifar E, Rueda D. 2014. HIV-1 DIS stem loop forms an obligatory bent kissing intermediate in the dimerization pathway. *Nucleic Acids Res* **42**: 7281–7289. doi:10.1093/nar/gku332
- Muriaux D, Fossé P, Paoletti J. 1996. A kissing complex together with a stable dimer is involved in the HIV-1 RNA dimerization process in vitro. *Biochemistry* **35**: 5075–5082. doi:10.1021/bi952822s
- Mustoe AM, Brooks CL, Al-Hashimi HM. 2014. Hierarchy of RNA functional dynamics. *Annu Rev Biochem* **83**: 441–466. doi:10.1146/annurev-biochem-060713-035524
- Ouldrige TE, Louis AA, Doye JP. 2010. Extracting bulk properties of self-assembling systems from small simulations. *J Phys Condens Matter* **22**: 104102. doi:10.1088/0953-8984/22/10/104102
- Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 1996. A loop–loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci* **93**: 5572–5577. doi:10.1073/pnas.93.11.5572
- Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51–55. doi:10.1038/nature06684
- Piganeau N, Schauer UE, Schroeder R. 2006. A yeast RNA-hybrid system for the detection of RNA–RNA interactions in vivo. *RNA* **12**: 177–184. doi:10.1261/rna.2105506
- Plumridge A, Katz AM, Calvey GD, Elber R, Kirmizialtin S, Pollack L. 2018. Revealing the distinct folding phases of an RNA three-helix junction. *Nucleic Acids Res* **46**: 7354–7365. doi:10.1093/nar/gky363
- Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. 2012. Automated 3D structure composition for large RNAs. *Nucleic Acids Res* **40**: e112. doi:10.1093/nar/gks339
- Privalov PL, Crane-Robinson C. 2018. Translational entropy and DNA duplex stability. *Biophys J* **114**: 15–20. doi:10.1016/j.bpj.2017.11.003
- Pyle AM, Schlick T. 2016. Challenges in RNA structural modeling and design. *J Mol Biol* **428**: 733–735. doi:10.1016/j.jmb.2016.02.012
- Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**: 1507–1517. doi:10.1261/rna.5248604
- Rose PW, Prlic A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, et al. 2017. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res* **45**: D271–D281. doi:10.1093/nar/gkw1042
- Schlick T, Pyle AM. 2017. Opportunities and challenges in RNA structural modeling and design. *Biophys J* **113**: 225–234. doi:10.1016/j.bpj.2016.12.037
- Sehdev P, Crews G, Soto AM. 2012. Effect of helix stability on the formation of loop–loop complexes. *Biochemistry* **51**: 9612–9623. doi:10.1021/bi300481v
- Serra MJ, Baird JD, Dale T, Fey BL, Retatagos K, Westhof E. 2002. Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA* **8**: 307–323. doi:10.1017/S1355838202024226
- Shapiro BA, Yingling YG, Kasprzak W, Bindewald E. 2007. Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol* **17**: 157–165. doi:10.1016/j.sbi.2007.03.001
- Shi YZ, Wang FF, Wu YY, Tan ZJ. 2014a. A coarse-grained model with implicit salt for RNAs: predicting 3D structure, stability and salt effect. *J Chem Phys* **141**: 105102. doi:10.1063/1.4894752
- Shi YZ, Wu YY, Wang FH, Tan ZJ. 2014b. RNA structure prediction: progress and perspective. *Chin Phys B* **23**: 078701. doi:10.1088/1674-1056/23/7/078701
- Shi YZ, Jin L, Wang FH, Zhu XL, Tan ZJ. 2015. Predicting 3D structure, flexibility, and stability of RNA hairpins in monovalent and divalent ion solutions. *Biophys J* **109**: 2654–2665. doi:10.1016/j.bpj.2015.11.006
- Shi YZ, Jin L, Feng CJ, Tan YL, Tan ZJ. 2018. Predicting 3D structure and stability of RNA pseudoknots in monovalent and divalent ion solutions. *PLoS Comput Biol* **14**: e1006222. doi:10.1371/journal.pcbi.1006222
- Sieradzan AK, Makowski M, Augustynowicz A, Liwo A. 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. doi:10.1063/1.4978680
- Sim AY, Minary P, Levitt M. 2012. Modeling nucleic acids. *Curr Opin Struct Biol* **22**: 273–278. doi:10.1016/j.sbi.2012.03.012
- Singh A, Sethaphong L, Yingling YG. 2011. Interactions of cations with RNA loop–loop complexes. *Biophys J* **101**: 727–735. doi:10.1016/j.bpj.2011.06.033
- Somarowthu S. 2016. Progress and current challenges in modeling large RNAs. *J Mol Biol* **428**: 736–747. doi:10.1016/j.jmb.2015.11.011
- Šulc P, Romano F, Ouldrige TE, Doye JPK, Louis AA. 2014. A nucleotide-level coarse-grained model of RNA. *J Chem Phys* **140**: 235102. doi:10.1063/1.4881424
- Sun LZ, Zhang D, Chen SJ. 2017. Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annu Rev Biophys* **46**: 227–246. doi:10.1146/annurev-biophys-070816-033920
- Tan ZJ, Chen SJ. 2006. Electrostatic free energy landscapes for nucleic acid helix assembly. *Nucleic Acids Res* **34**: 6629–6639. doi:10.1093/nar/gkl810
- Tan ZJ, Chen SJ. 2007. RNA helix stability in mixed Na⁺/Mg²⁺ solution. *Biophys J* **92**: 3615–3632. doi:10.1529/biophysj.106.100388
- Tan ZJ, Chen SJ. 2010. Predicting ion binding properties for RNA tertiary structures. *Biophys J* **99**: 1565–1576. doi:10.1016/j.bpj.2010.06.029
- Tan ZJ, Chen SJ. 2011. Salt contribution to RNA tertiary structure folding stability. *Biophys J* **101**: 176–187. doi:10.1016/j.bpj.2011.05.050
- Tan YL, Feng CJ, Jin L, Shi YZ, Zhang W, Tan ZJ. 2019. What is the best reference state for building statistical potentials in RNA 3D structure evaluation? *RNA* **25**: 793–812. doi:10.1261/rna.069872.118
- Tam VK, Kwong D, Tor Y. 2007. Fluorescent HIV-1 dimerization initiation site: design, properties, and use for ligand discovery. *J Am Chem Soc* **129**: 3257–3266. doi:10.1021/ja0675797
- Uusitalo JJ, Ingólfsson HI, Marrink SJ, Faustino I. 2017. Martini coarse-grained force field: extension to RNA. *Biophys J* **113**: 246–256. doi:10.1016/j.bpj.2017.05.043
- Van Melckebeke H, Devany M, Di Primo C, Beaurain F, Toulmé JJ, Bryce DL, Boisbouvier J. 2008. Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex. *Proc Natl Acad Sci* **105**: 9210–9215. doi:10.1073/pnas.0712121105
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2009. Effect of Mg²⁺ and Na⁺ on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. *J Mol Biol* **386**: 773–788. doi:10.1016/j.jmb.2008.12.073
- Walter AE, Turner DH. 1994. Sequence dependence of stability for coaxial stacking of RNA helices with Watson–Crick base paired interfaces. *Biochemistry* **33**: 12715–12719. doi:10.1021/bi00208a024

- Walter AE, Turner DH, Kim J, Lyttle MH, Muller P, Mathews DH, Zuker M. 1994. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci* **91**: 9218–9222. doi:10.1073/pnas.91.20.9218
- Wang J, Zhao Y, Zhu C, Xiao Y. 2015. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res* **43**: e63. doi:10.1093/nar/gkv141
- Wang J, Mao K, Zhao Y, Zeng C, Xiang J, Zhang Y, Xiao Y. 2017. Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide–nucleotide interactions from direct coupling analysis. *Nucleic Acids Res* **45**: 6299–6309. doi:10.1093/nar/gkx386
- Wang Y, Wang Z, Liu T, Gong S, Zhang W. 2018. Effects of flanking regions on HDV cotranscriptional folding kinetics. *RNA* **24**: 1229–1240. doi:10.1261/rna.065961.118
- Weixlbaumer A, Werner A, Flamm C, Westhof E, Schroeder R. 2004. Determination of thermodynamic parameters for HIV DIS type loop–loop kissing complexes. *Nucleic Acids Res* **32**: 5126–5133. doi:10.1093/nar/gkh841
- Wong GC, Pollack L. 2010. Electrostatics of strongly charged biological polymers: ion-mediated interactions and self-organization in nucleic acids and proteins. *Annu Rev Phys Chem* **61**: 171–189. doi:10.1146/annurev.physchem.58.032806.104436
- Woodson SA. 2005. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol* **9**: 104–109. doi:10.1016/j.cbpa.2005.02.004
- Wu YY, Zhang ZL, Zhang JS, Zhu XL, Tan ZJ. 2015. Multivalent ion-mediated nucleic acid helix–helix interactions: RNA versus DNA. *Nucleic Acids Res* **43**: 6156–6165. doi:10.1093/nar/gkv570
- Xi K, Wang F, Xiong G, Zhang Z, Tan ZJ. 2018. Competitive binding of Mg^{2+} and Na^{+} ions to nucleic acids: from helices to tertiary structures. *Biophys J* **114**: 1776–1790. doi:10.1016/j.bpj.2018.03.001
- Xia T, SantaLucia J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37**: 14719–14735. doi:10.1021/bi9809425
- Xia Z, Bell DR, Shi Y, Ren P. 2013. RNA 3D structure prediction by using a coarse-grained model and experimental data. *J Phys Chem B* **117**: 3135–3144. doi:10.1021/jp400751w
- Yamasaki S, Hirokawa T, Asai K, Fukui K. 2014. Tertiary structure prediction of RNA–RNA complexes using a secondary structure and fragment-based method. *J Chem Inf Model* **54**: 672–682. doi:10.1021/ci400525t
- Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. 2003. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* **31**: 3450–3460. doi:10.1093/nar/gkg529
- Zhang J, Bian Y, Lin H, Wang W. 2012. RNA fragment modeling with a nucleobase discrete-state model. *Phys Rev E Stat Nonlin Soft Matter Phys* **85**: 021909. doi:10.1103/PhysRevE.85.021909
- Zhang ZL, Wu YY, Xi K, Sang JP, Tan ZJ. 2017. Divalent ion-mediated DNA–DNA interactions: a comparative study of triplex and duplex. *Biophys J* **113**: 517–528. doi:10.1016/j.bpj.2017.06.021
- Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. 2012. Automated and fast building of three-dimensional RNA structures. *Sci Rep* **2**: 734. doi:10.1038/srep00734
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415. doi:10.1093/nar/gkg595