Supporting Material for

Predicting 3D Structure, Flexibility and Stability of RNA Hairpins in Monovalent and Divalent Ion Solutions

Ya-Zhou Shi[1], Lei Jin[1], Feng-Hua Wang[2], Xiao-Long Zhu[3], and Zhi-Jie Tan[1*]

[1]*Department of Physics and Key Laboratory of Artificial Micro- and Nano-structures of the Ministry of Education, School of Physics and Technology, Wuhan University, Wuhan, China*

[2]*Engineering Training Center, Jianghan University, Wuhan, China*

[3]*Department of Physics, School of Physics and Information Engineering, Jianghan University, Wuhan, China*

[*] To whom correspondence should be addressed. Email: zjtan@whu.edu.cn

The force-field of the coarse-grained model

1. Energy functions

The total energy *U* of the system in the present model includes eight distinct contributions (1):

$$U = U_b + U_a + U_d + U_{exc} + U_{bp} + U_{bs} + U_{el} + U_{cs}, \tag{S1}$$

where

$$U_b = \sum_{bonds} K_b(r - r_0)^2, \tag{S2}$$

$$U_a = \sum_{angles} K_\theta(\theta - \theta_0)^2, \tag{S3}$$

$$U_d = \sum_{dihedrals} \{K_\varphi[1 - cos(\varphi - \varphi_0)] + \frac{1}{2}K_\varphi[1 - cos3(\varphi - \varphi_0)]\}, \tag{S4}$$

$$U_{exc} = \sum_{i<j}^{N} \begin{cases} 4\varepsilon\left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6}\right] & if\ r_{ij} \le \sigma_0 \\ 0 & if\ r_{ij} > \sigma_0 \end{cases}, \tag{S5}$$

$$U_{bp} = \sum_{i<j-3}^{N_{bp}} \frac{\varepsilon_{bp}}{1 + k_{NN}\left(r_{N_iN_j} - r_{NN}\right)^2 + k_{CN}\sum_{i(j)}\left(r_{C_iN_j} - r_{CN}\right)^2 + k_{PN}\sum_{i(j)}\left(r_{P_iN_j} - r_{PN}\right)^2}, \tag{S6}$$

$$U_{bs} = \frac{1}{2}\sum_{i,j}^{N_{st}} |G_{i,i+1,j-1,j}| \left\{\left[5\left(\frac{\sigma_{st}}{r_{i,i+1}}\right)^{12} - 6\left(\frac{\sigma_{st}}{r_{i,i+1}}\right)^{10}\right] + \left[5\left(\frac{\sigma_{st}}{r_{j,j-1}}\right)^{12} - 6\left(\frac{\sigma_{st}}{r_{j,j-1}}\right)^{10}\right]\right\}, \tag{S7}$$

$$U_{el} = \sum_{i<j}^{N_P} \frac{(Qe)^2}{4\pi\varepsilon_0\varepsilon(T)r_{ij}} \mathrm{e}^{-\frac{r_{ij}}{l_D}}. \tag{S8}$$

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

The first three terms in Eq. S1 are typical expressions for virtual bonds U_b, bond angles U_a and dihedrals U_d. The initial parameters including equilibrium distances/angles and corresponding energy strength in these terms are derived from the Boltzmann inversion of the corresponding atomistic distribution functions obtained by the statistical analysis on the experimental structures in the Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do) (1). For these bonded potentials, two sets

of parameters $Para_{helical}$ and $Para_{nonhelical}$ are provided for stems and single-strands/loops (1), respectively, due to the different structural features of stems (helical) and single-strands/loops (nonhelical) in RNAs. The $Para_{nonhelical}$ are used in the folding process to describe the folding of free RNA chains (1), while during the structure refinement based on the 3D structure predicted from the initial folding process, the $Para_{helical}$ and $Para_{nonhelical}$ are used for the stems and single-strands (1), respectively.

The remaining terms of Eq. S1 describe various pairwise, nonbonded interactions (1). U_{exc} represents the excluded volume interaction between the nonbonded CG beads to avoid their overlap. The interaction strength ε = 0.26 kcal/mol, and σ_0 is the sum of the radii of two interactional beads. U_{bp} is the orientation-dependent base-pairing interaction between bases in the canonical Watson-Crick base pairs (G-C and A-U) and the wobble base pairs (G-U), and the interaction strength $\varepsilon_{AU} = \varepsilon_{GU} = 0.5\varepsilon_{GC}$ (1). U_{bs} is the base-stacking interaction between two nearest neighbour base pairs, and the sequence-dependent strength of base-stacking energy $G_{i,i+1,j-1,j}$ can be estimated from the combination of the experimental thermodynamics parameters (2,3) and Monte Carlo algorithm; more details can be found in Ref. 1.

U_{el} is the electrostatic interaction between phosphate groups given by the Debye-Hückel approximation, and based on the counterion condensation theory (4) and the tightly bound ion model (5-7), the reduced charge fraction Q could be written as $Q = f_{\mathrm{Na^+}}\left(\frac{b}{l_B}\right) + (1 - f_{\mathrm{Na^+}})\left(\frac{b}{2l_B}\right)$ for mixed Na^+/Mg^{2+} solutions, where f_{Na^+} and 1-f_{Na^+} are the contribution fraction from Na^+ and Mg^{2+}, respectively. b is the phosphate-phosphate spacing of an RNA and is taken as 5.5Å (1). The empirical formula

$$f_{\mathrm{Na^+}} = \frac{[\mathrm{Na^+}]}{[\mathrm{Na^+}] + x[\mathrm{Mg^{2+}}]} \tag{S10}$$

derived by the tightly bound ion model is used for mixed divalent/monovalent ion solutions, and $x = (8.1 - 64.8/N)(5.2 - \ln[\mathrm{Na^+}])$, where $[Na^+]$ and $[Mg^{2+}]$ are the corresponding concentrations in molar (M) and N is the chain length (5-7).

Similarly to U_{bs}, U_{cs} is also the base-stacking interactions but between the bases which are the interfaces of two discrete stems, and the thermodynamics parameters from UV melting studies have shown that the coaxial stacking between two stems interrupted by one or more nucleotides is

approximately the same stable as the corresponding nearest-neighbour interaction in an uninterrupted helix of equivalent sequence (8,9). Since the noncanonical base pairs neglected by the present model are common when more than one base exists in each single-stranded chain between two discrete stems, we only consider the coaxial stacking for the cases that two stems are interrupted by less than two nucleotides in at least one side (see Fig. S1a) (8,9). To obtain the corresponding geometric constraint a and r_{cs}, we have calculated the distance distribution (Fig. S1b) between interfaced bases in coaxial stacked stems of known large structures including rRNAs, tRNAs and ribonuclease P RNAs.

2. Parameters

The initial parameters of the above described potentials are directly obtained from the statistical analysis on the known structures. The final parameters listed in Table SI and Table SII are derived through the comparisons between the predictions by the model and the experimental data (see more details in Ref. 1). All the parameters with the exception of ones newly introduced for U_{el} and U_{cs} are the same as the previous version of our model (1).

Table SI. The parameters of bonded potentials in Eqs. S2-S4.

Bond U_b				
	K_b (kcal/mol)		r_0 (Å)	
	$Para_{helical}$ [a]	$Para_{nonhelical}$ [b]	$Para_{helical}$	$Para_{nonhelical}$
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)		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_φ (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

[a] The $Para_{helical}$ only are used in the processes of folded structure refinement for the base-pairing regions (stems) in the initially folded structure. [b] The $Para_{nonhelical}$ are used in RNA folding processes to possibly describe RNAs as free chains.

Table SII. The parameters of nonbonded potentials in Eqs. S5-S9.

Nonbonded				
U_{exc}	ε (kcal/mol)	0.26	σ_{st} (Å)	$R_i^a+R_j$
U_{bp}	$\varepsilon_{bp(\mathrm{GC})}$ (kcal/mol)	-3.5	$\varepsilon_{bp(\mathrm{AU/GU})}$ (kcal/mol)	-1.75
	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[b]	σ_{st} (Å)	4.8
U_{el}	Q	Salt-dependent	b (Å)	5.5
U_{cs}	$G_{i\text{-}j,k\text{-}l}$ (kcal/mol)	Sequence-dependent[b]	a (Å^{-1})	0.4
			r_{cs} (Å)	5.0

[a] The radius of bead i, and the radii of P, C and N beads are 1.9 Å, 1.7 Å and 2.2 Å, respectively. [b] For different combinations of bases, $G_{sequence} = \Delta H - T(\Delta S - \Delta S_c)$, where ΔH and ΔS are the RNA thermodynamics parameters derived from the corresponding experiments for base-pair stacking (2,3) or coaxial stacking (8,9), and ΔS_c is equal to -9.7eu (see Ref. 1).

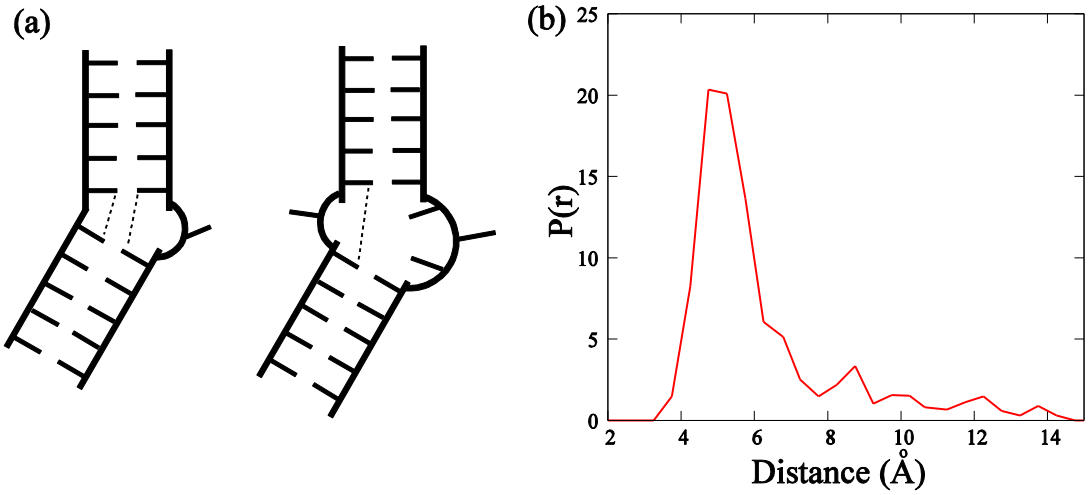

Figure S1. (a) A schematic representation of the coaxial stacking potential used in the model. (b) The normalized probability distribution $P(r)$ of the distances between two interfaced bases in coaxially stacked stems (dash line in (a)), which is obtained by the statistical analysis over the known structures of 18 large RNAs (PDB code: 1c2w, 1njn, 1p9x, 1y0q, 1ffk, 1fg0, 2a64, 2o44, 3g78, 3ize, 3j2h, 3j3c, 3j3d, 3j5z, 3l0u, 3rg5, 4c4q, 4p5j).

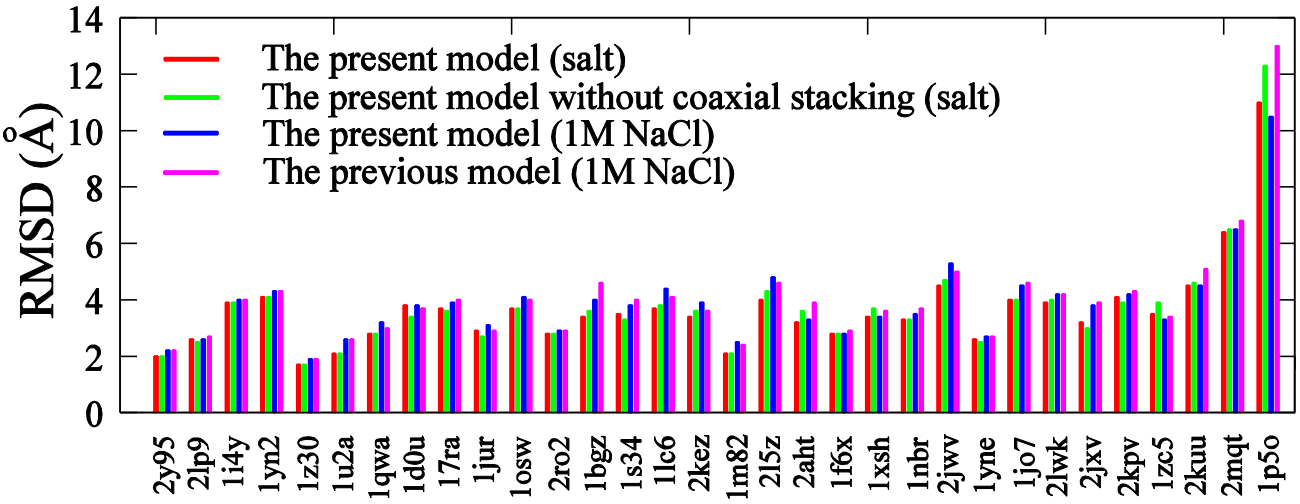

Figure S2. Comparison of the RMSDs of RNA 3D structures predicted by our present and previous models. The RMSDs of predicted structures are calculated over C beads from the corresponding C4' atoms in the native structures. Red: predicted by the present model with coaxial stacking at experimental ion conditions; green: predicted by the present model without coaxial stacking at experimental ion conditions; blue: predicted by the present model with coaxial stacking at 1M NaCl; pink: predicted by our previous model without coaxial stacking at 1M NaCl. For the four independent predictions, the overall mean RMSDs of 32 RNA hairpins are 3.64 Å, 3.71 Å, 3.89 Å and 4.02 Å, respectively.

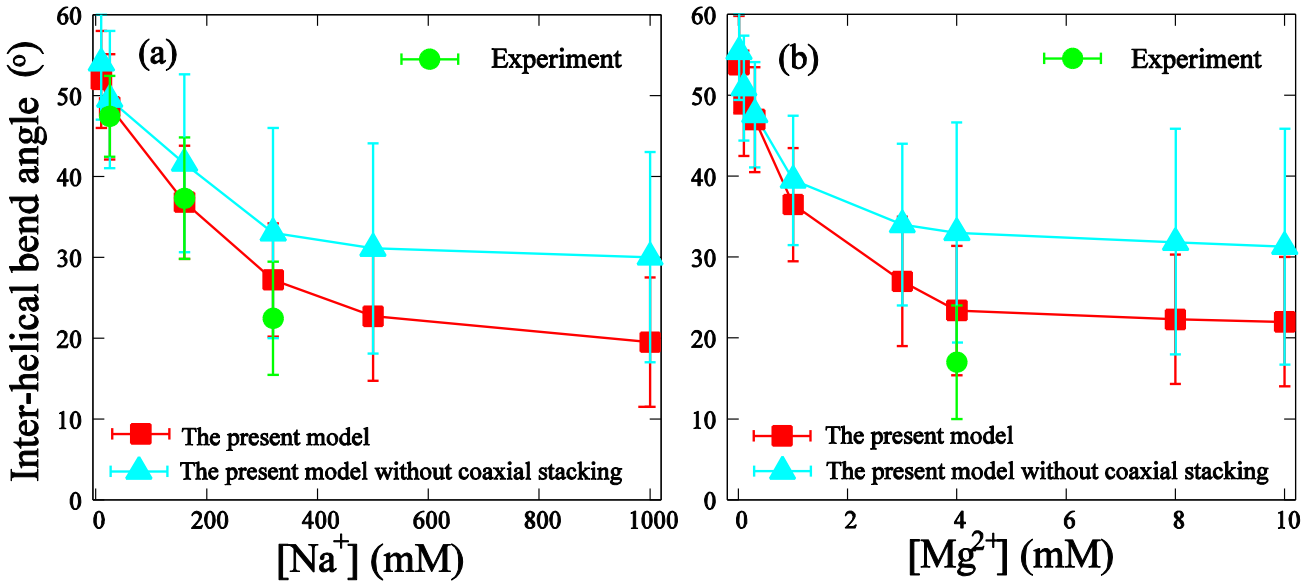

Figure S3. (a, b) The experimental (green; Ref. 10) and predicted inter-helical bend angle as functions of $[Na^+]$ (a) and $[Mg^{2+}]$ (b) for HIV-1 TAR variant (see Fig. 3). Red: predicted by the present model. Cyan: predicted by the present model without coaxial stacking.

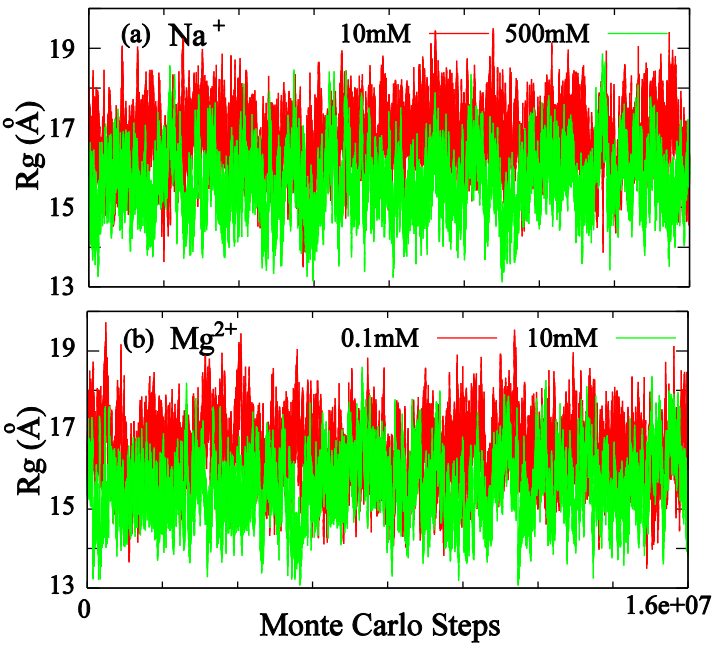

Figure S4. The time-evolution of the radius of gyration for HIV-1 TAR variant (shown in Fig. 3 in main text) at different Na^{+}/Mg^{2+} conditions.

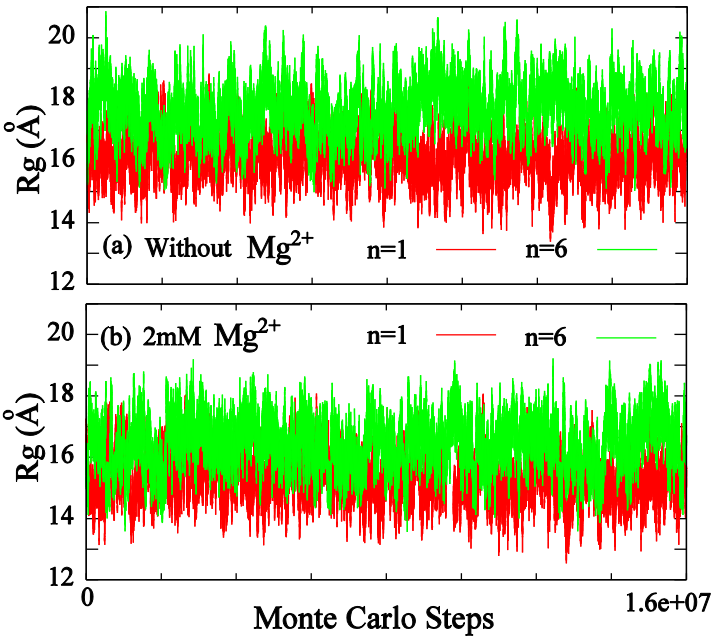

Figure S5. The time-evolution of the radius of gyration for HIV-2 TAR variant (shown in Fig. 3 in main text) with different lengths of bulge at 5 mM $NaPO_4$ without (a) or with (b) 2 mM Mg^{2+}.

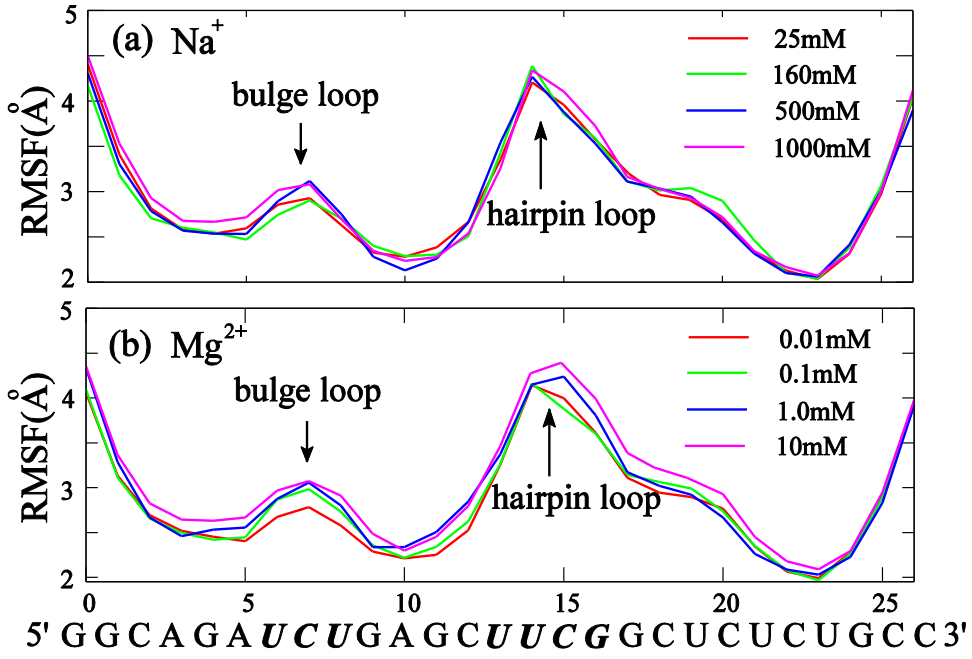

Figure S6. The RMSF for C-beads along HIV-1 TAR variant (shown in Fig. 3 in main text) in different Na^{+}/Mg^{2+} solutions. The sequences of loops are in italics.

SUPPORTING REFERENCES

1. Shi, Y. Z., F. H. Wang, Y. Y. Wu, and 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.
2. Xia, T., J. SantaLucia, M. E. Burkand, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and 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.
3. Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner. 1999. Expended sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288: 911-940.
4. Manning, G. S. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q. Rev. Biophys. 11:179-246.
5. Tan, Z. J., W. Zhang, Y. Z. Shi, and F. H. Wang. 2015. RNA folding: structure prediction, folding kinetics and ion electrostatics. Adv. Exp. Med. Biol. 827:143-183.
6. Tan, Z. J., and S. J. Chen, 2007. RNA helix stability in mixed Na^+/Mg^{2+} solution. Biophys. J. 92:3615–3632.
7. 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.
8. Walter, A. E., D. H. Turner, J. Kim, M. H. Lyttle, P. Müller, D. H. Mathews, and M. Zuker. 1994. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. Proc. Natl. Acad. Sci. 91:9218–9222.
9. Walter, A. E., and D. H. Turner. 1994. Sequence dependence of stability for coaxial stacking of RNA helixes with Watson-Crick base paired interface. Biochemistry. 33:12715–12719.
10. Casiano-Negroni, A., X. Sun, and H. M. Al-Hashimi. 2007. Probing Na^+-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. Biochemistry. 46:6525-6535.