Biophysical Journal, Volume *115*

Supplemental Information

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

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

The energy functions of the coarse-grained model

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Calculating melting temperature at low experimental strand concentration

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

Table S1. The parameters of bonded potential

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

Table S2. The parameters of nonbonded potential

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

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

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

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

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

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

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

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

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

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

[e] Using electrostatic potential described in this work.

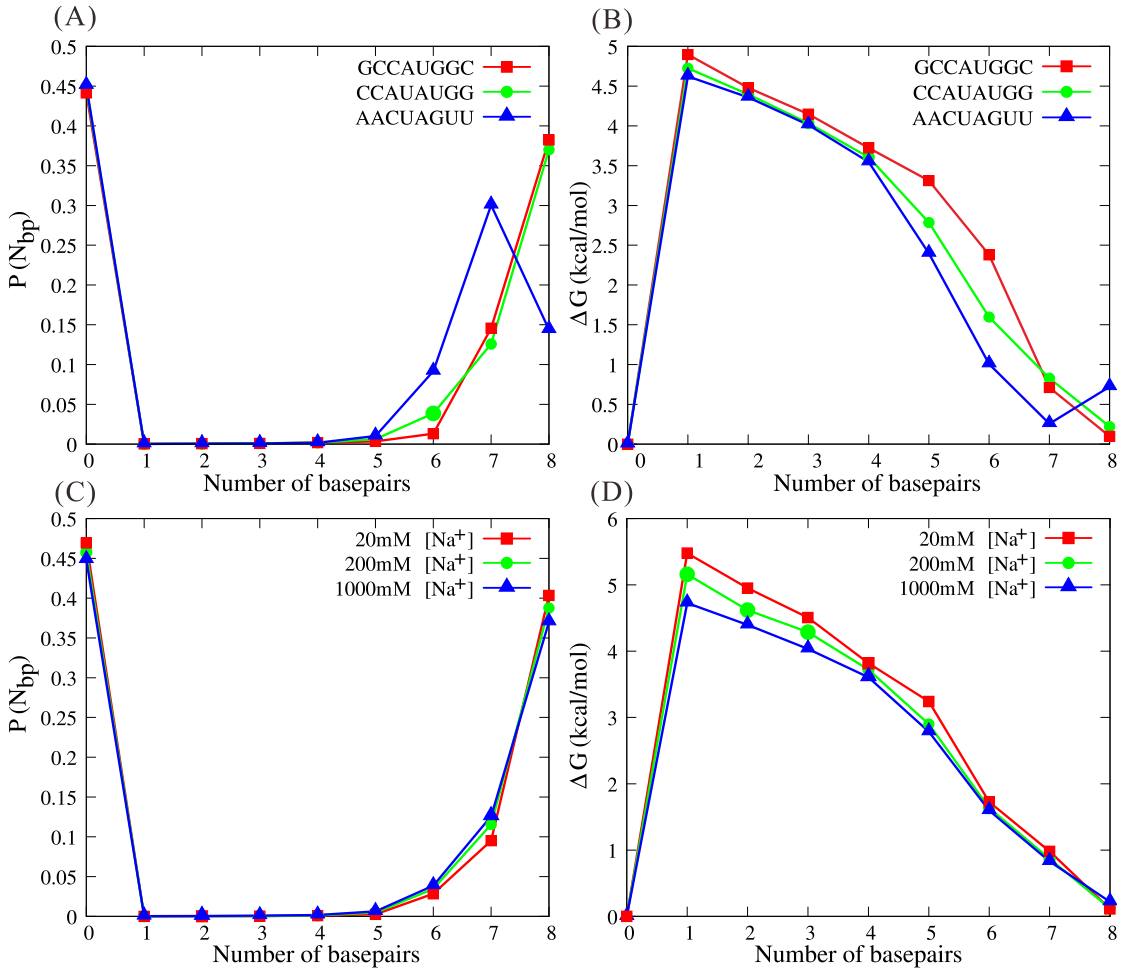

FIGURE S1. (A, B) The normalized populations of formed base pair number $P(N_{bp})$ (A) and the free energy barrier ΔG (B) as functions of the number of base pairs of three sequences at T_m. (C, D) The normalized populations of formed base pair number $P(N_{bp})$ (C) and the free energy barrier ΔG (D)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

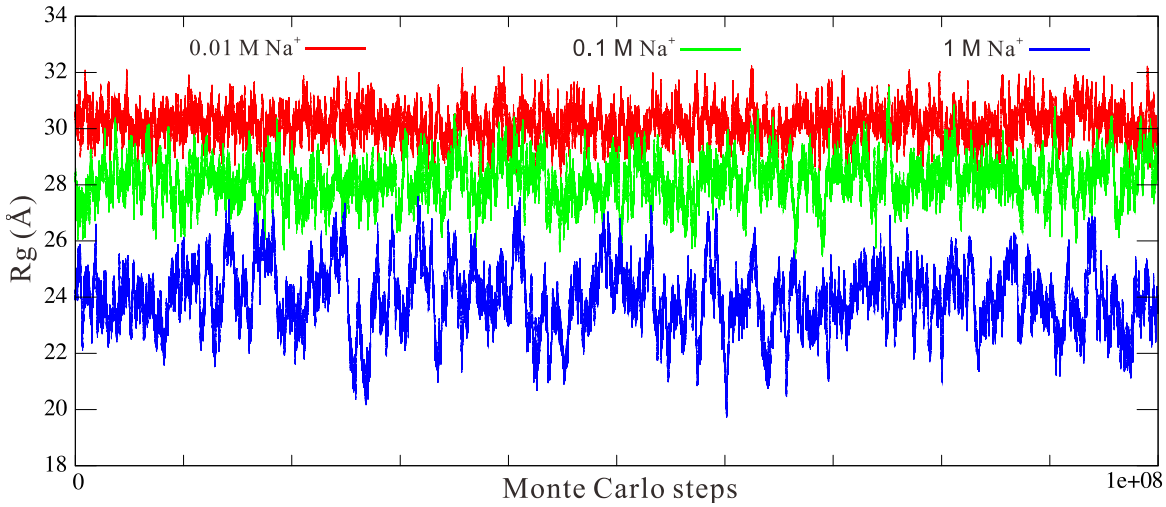

Figure S2. The time-evolution of the radius of gyration for the 40-bp dsRNA helix at different Na^+ concentration.

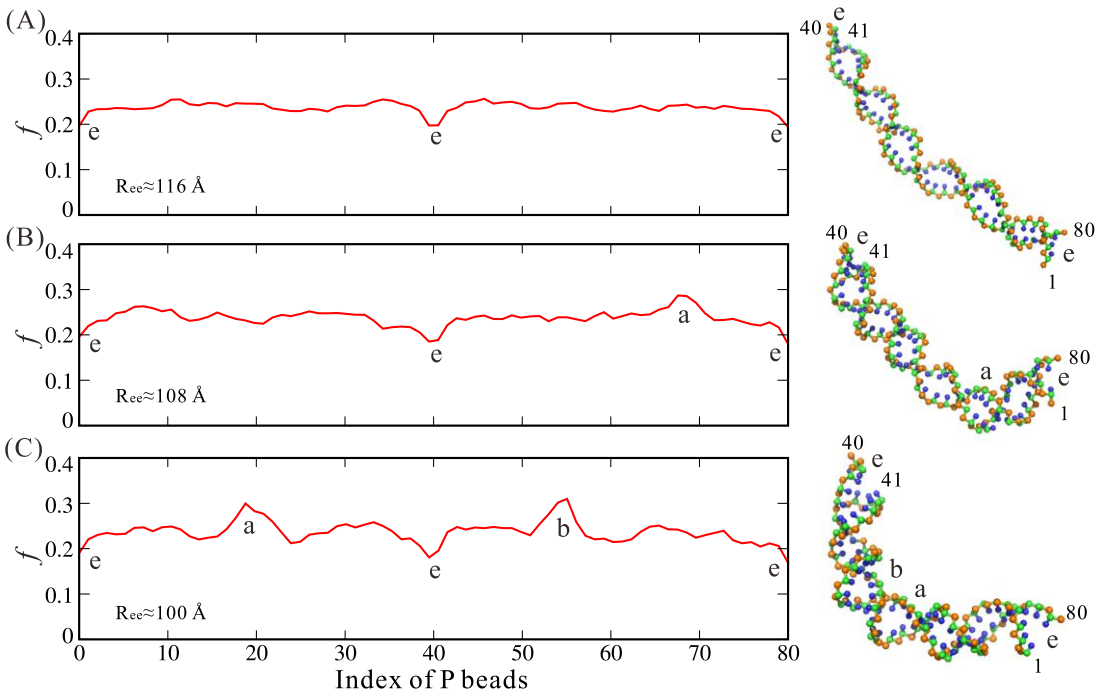

Figure S3. The calculated ion charge neutralization fraction f (left) along P beads of the corresponding structures (right) with the end-to-end distances Ree of ~116 Å (A), ~108 Å (B) and ~100 Å (C) of the 40-bp dsRNA at 0.1 M [Na^+]. The P beads in the bending region of the structures are labeled with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

References:

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

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

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

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

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

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

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

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

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

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

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

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

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