PNAS

www.pnas.org

Supplementary Information for

Folding RNA in mixtures of monovalent and divalent cations: Theory and simulations

Hung T. Nguyen, Naoto Hori and D. Thirumalai

Corresponding Author D. Thirumalai
E-mail: dave.thirumalai@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 to S10
Table S1
References for SI reference citations

Supporting Information Text

The Supplementary Information is divided into the following sections. The details of the theory used to numerically solve the RISM integral equation, needed to calculate the potential of mean force between divalent ions and the phosphate groups are given in Section 1. Section 2 describes the development of the RNA force field, which treats divalent cations explicitly and the monovalent ions implicitly. We also describe the methods used to determine the values of the force field parameters. The simulation details and methods used to analyze the data are given in Section 3. Additional tests of the validity of the theory-based construction of the RNA model and simulations are contained in Section 4.

1. Divalent ion–phosphate potential of mean force

Reference Interaction Site Model (RISM). Accurate simulations using coarse-grained models of even modestly sized RNA molecules in explicit monovalent and divalent cations is computationally demanding (1). In order to simplify the problem, while still retaining high level of accuracy achieved previously (1), we treat the electrostatic effects due to monovalent ions implicitly. This leaves us with the task of calculating the effective interactions between the divalent cations and RNA. Our primary goal is to calculate the potential of mean force (PMF), $W(r)$, between Mg^{2+}/Ca^{2+} and phosphate (Eq. 5 in the main text), which can be used in simulations of divalent cation-induced folding of RNA. In order to calculate $W(r)$, we resort to the well-known RISM theory, which was developed to calculate the equilibrium site-site distributions of polyatomic liquids and their associated thermodynamic properties (2–8). The theory has two versions: one is a 1-dimensional RISM (or 1D-RISM) and the other is a 3-dimensional RISM (3D-RISM). The former provides the radial distribution functions, $g_{ij}(r)$, between every interaction site in the system. The latter couples the 1D radial information and the 3D structure of the biomolecule to yield the solvent structure around the biomolecule in the form of a 3D site distribution function, $g_i(\mathbf{r})$, for each solvent site. Because the theory and implementation are widely known, here we only give a very brief summary of the 1D-RISM that is most directly relevant to our work.

We begin with the Ornstein–Zernike (OZ) equation:

$$h(r_{12}) = c(r_{12}) + \rho \int dr_3 c(r_{13}) h(r_{32}),$$
[1]

where r_{ij} is the distance between particles i and j, c is the direct correlation function, and h–the total correlation function–is related to the pair distribution function, $h_{ij}(r_{ij}) \equiv g_{ij}(r_{ij}) - 1$. In order to solve the OZ equation, it is necessary to use an appropriate closure relation connecting h and c, which we write as:

$$g(r_{12}) = \exp\left[-\beta u(r_{12}) + h(r_{12}) - c(r_{12}) + b(r_{12})\right],$$
[2]

or in a short form $g = \exp\left[-\beta u + h - c + b\right]$. In the above equation, u is the potential energy function, $\beta = \frac{1}{k_B T}$ (k_B is the Boltzmann constant and T is the temperature), and b is an unknown "bridge function". In the hypernetted-chain approximation (HNC), b is zero, giving:

$$g_{HNC} = \exp\left[-\beta u + h - c\right].$$
[3]

The HNC closure gives good results for ionic and polar systems, but not for neutral systems. Moreover, it is difficult to find converged solutions (6, 9, 10). To resolve these problems, Kovalenko–Hirata (KH) introduced the following closure relation: (11)

$$g_{KH} = \begin{cases} \exp\left[-\beta u + h - c\right] & \text{if } g \leq 1 \\ 1 - \beta u + h - c & \text{if } g > 1 \end{cases}.$$
[4]

The partial series expansion of order-n (PSE-n) offers a way to interpolate between Eqs. 3 and 4, which improves the results of the KH closure while circumventing the convergence issues associated with the HNC closure: (12)

$$g_{PSE-n} = \begin{cases} \exp\left[-\beta u + h - c\right] & \text{if } g \leq 1 \\ \sum_{i=0}^{n} \frac{[-\beta u + h - c]^i}{i!} & \text{if } g > 1 \end{cases}.$$

Hence, KH is the special case of PSE closure when $n = 1$. In the limit $n \to \infty$, HNC is obtained.

In RISM (as implemented in the Amber force field), the standard Coulomb and Lennard–Jones interaction are used for the pair-wise non-bonded potential: (13)

$$u_{12}(r) = \frac{q_1 q_2}{r} + \varepsilon_{12}\left[\left(\frac{R_{min,12}}{r}\right)^{12} - 2\left(\frac{R_{min,12}}{r}\right)^6\right].$$
[5]

Recently, a 12-6-4 LJ potential was proposed to account for the ion-induced dipole moment interaction, which proved to be important for divalent ions (14). With this scheme, the LJ interaction (the second term in Eq. 5) is modified as:

Hung T. Nguyen, Naoto Hori and D. Thirumalai

$$U_{LJ}(r) = \varepsilon_{12}\left[\left(\frac{R_{min,12}}{r}\right)^{12} - 2\left(\frac{R_{min,12}}{r}\right)^6 - 2\kappa R_{min,12}^2\left(\frac{R_{min,12}}{r}\right)^4\right]$$

$$= \frac{A_{12}}{r^{12}} - \frac{B_{12}}{r^6} - \frac{C_{12}}{r^4}, \tag{6}$$

where the attractive term $\frac{C_{12}}{r^4}$ approximately accounts for the charge-induced dipole interaction. For highly charged systems, the potential in Eq. 6 yields accurate values of hydration free energies, ion-oxygen distances in the first hydration shell, and coordination numbers for divalent ions (and later, extending to trivalent and tetravalent ions) (15). A modified parameter set for Mg^{2+} was subsequently developed to balance the interaction between Mg^{2+} and water, Mg^{2+} and specific sites on nucleic acids (16). Here, we adopt these modifications in RISM and treat only the interaction involving Mg^{2+} (namely, Mg^{2+}–water and Mg^{2+}–P) using the 12-6-4 potential (Eq. 6), while the rest (water–water, water–P, P–P and Mg^{2+}–Mg^{2+}) are modeled using the standard 12-6 potential, $U_{LJ}(r) = \frac{A_{12}}{r^{12}} - \frac{B_{12}}{r^6}$. (The small polarizability of Mg^{2+} allows us to neglect Mg^{2+}–Mg^{2+} interactions.) A similar procedure is used to calculate Ca^{2+}–P interactions.

The RISM equations (Eqs. 1 and 2) are solved iteratively until converged results are obtained. We are interested only in the probability distribution between Mg^{2+} and phosphate, from which the PMF is computed using $W(r) = -k_B T \ln g(r)$.

Numerical solution of the RISM calculation. The PMFs between the divalent cations and phosphates were calculated using the 1D-RISM implemented in Amber (13) by modifying the `rism1d` code to include the potential in Eq. 6. The theory requires bulk concentrations and topologies (bond lengths and bond angles) of every molecule and ion in the system as well as the pairwise interaction potentials between them. In our case, the system was comprised of Mg^{2+} (or Ca^{2+}), phosphate and water. The concentration of XP_2 (where $X = Mg^{2+}$ or Ca^{2+}) is 1 mM. We used a 1-dimensional grid with a grid spacing of 0.025 Å and 131,072 grid points. Parameters for Mg^{2+} and Ca^{2+} were taken from the Amber force field that takes into account the charge-induced dipole interactions (17). We used the cSPC/E model for water, which introduces van der Waals terms for the hydrogen atoms of the SPC/E water model to prevent them to collapse in RISM (13). For phosphate, we used a single-site representation as in the coarse-grained model rather than an all-atom representation to overcome the convergence problems in RISM. To derive the phosphate parameters, we started with the Cl⁻ parameters and tuned $\epsilon = 0.027$ kcal/mol, and $R_{min} = 2.6$ Å (Eq. 6) to obtain the location of the first peak in the $g_{Mg^{2+}-P}(r)$ at around 2.5 Å (see Fig. 7B in the main text), which is somewhat longer than Mg^{2+}–O distance 2.06 Å in the first hydration shell. Note that the position of the coarse-grained P site is located at the center of geometry of the phosphate group. We found that $g_{Mg^{2+}-P}(r)$ is not sensitive to the parameters provided the first peak is around 2.5 Å. We iteratively solved the RISM equations using the PSE-3 closure to a residual tolerance of 10^{-12} at 25°C. We emphasize that no simulation was performed at this stage. The pair distribution function between divalent ion–phosphate $g_{X^{2+}-P}(r)$ (X^{2+} is Mg^{2+} or Ca^{2+}) is one of the direct outputs of the RISM program, from which we obtain $W(r) = -k_B T \ln g_{X^{2+}-P}(r)$.

2. Development of coarse-grained TIS model

Coarse-grained RNA force field. Our goal is to produce an accurate coarse-grained model, which treats all the key interactions in a manner that can lead to quantitative predictions of thermodynamics and kinetics of RNA with arbitrary length. To this end, we build on the TIS model (18) in order to develop an RNA force field that treats divalent cations explicitly while describing the monovalent effects implicitly. Following our previous study (1, 19), each nucleotide is represented by three interaction sites, corresponding to phosphate (P), ribose (S) and base (B), where P and S represent the backbone; the B site depends on the nature of the nucleotide, and therefore carries the sequence information. The energy function has the following form:

$$U = U_{BA} + U_{EV} + U_{ST} + U_{HB} + U_{EL},$$

where U_{BA} is the *bonded term*, comprising of bond and angle restraints between connected beads. These constraints use harmonic potentials to keep the bonds and the angles close to the A-form helix. The parameters are the same as in our previous work (1, 19).

Excluded volume interactions U_{EV}***.*** We model U_{EV} using a modified LJ potential (1), which is evaluated using:

$$U_{EV} = \sqrt{\varepsilon_i \varepsilon_j}\left[\left(\frac{1.6}{r + 1.6 - D_{ij}}\right)^{12} - 2\left(\frac{1.6}{r + 1.6 - D_{ij}}\right)^6 + 1\right]$$

if $r \leq D_{ij} = R_i + R_j$. If $r > D_{ij}$ then we set $U_{EV} = 0$. To allow for favorable base stacking, we set $D_{BB} = 3.2$ Å. It is not necessary to include excluded volume interactions between the divalent cation and P as it is taken into account in the effective potential displayed in Figs. 1 and 7A in the main text. The full set of parameters used in the simulations is given in Table S1.

	Mass, Da	R_i, Å	ε_i, kcal/mol	q_i
P	62.974	1.89	0.200	(varied, see main text)
S	131.106	2.61	0.200	0
A	134.132	2.52	0.200	0
G	150.132	2.70	0.200	0
C	110.102	2.43	0.200	0
U	111.084	2.43	0.200	0
Mg^{2+}	24.305	2.00	0.895	+2
Ca^{2+}	40.078	2.80	1.000	+2

Table S1. Parameters for excluded volume and Debye–Huckel interactions.

100 ***Justification for the divalent ion excluded volume parameters.*** The radius of the divalent cations used in this work to compute the
101 excluded volume interactions are large (2.00 Å for Mg^{2+} and 2.80 Å for Ca^{2+}) compared to the values used in atomistic
102 simulations. In general, the size of various interaction sites has to be larger in coarse-grained models. We justify the value
103 in Table S1 by arguing that the divalent ion radius in this work represents the fully hydrated form of the ion. For Mg^{2+},
104 the relevant size is the radius of the hexahydrated form $Mg(H_2O)_6^{2+}$, which coincides with the distance between Mg^{2+} and
105 the oxygen atom in the first hydration shell $d_{Mg^{2+}-Ow} \approx 2.02 - 2.10$ Å. Therefore, in our model, we assume that Mg^{2+} ions
106 do not dehydrate in order to interact with base or sugar moieties and other Mg^{2+} ions. Although the most frequent inner
107 sphere coordination of Mg^{2+} occurs with the phosphate groups, it has been documented that Mg^{2+} also coordinates with
108 nucleobases and less frequently with sugars (20). However, the neglect of such interactions, which likely do not contribute to
109 charge neutralization of the phosphate groups, should not considerably affect the folding of RNA. From the perspective of
110 folding thermodynamics, it is crucial to treat Mg^{2+}–P interactions accurately, which we do rigorously using the RISM theory.
111 We note that in our model, the divalent ion radius does not play any role in the divalent ion–phosphate interactions since these
112 interactions are calculated based on the PMF (Eq. 5, main text). We do, therefore, allow the Mg^{2+} dehydration once they
113 are near phosphate groups, as shown in Fig. 1 in the main text. The excellent predictions of the free energies for a variety
114 of systems reported here show that it is crucial to account for the physics of divalent ion–phosphate interactions. This is
115 accomplished using the RISM theory.

116 To ascertain that the radius of the divalent ion used here is physically reasonable, we compute the distance between the
117 Mg^{2+} and sugars/bases in the simulations and compare them to the values in the crystal structure. The idea is to see if Mg^{2+}
118 approaches the RNA in the simulations at distances that are too small or large. For the crystal structures, we took all RNA
119 structures in the PDB that have Mg^{2+} and the resolution is at least 2.5 Å. With this criterion, we obtained 147 structures. We
120 coarse-grained the RNAs and generated the histogram of the distances between Mg^{2+}–base and Mg^{2+}–sugar. Fig. S1 compares
121 the radial distribution function $g_{Mg^{2+}-\alpha}$ computed in the simulations of BWYV and the histogram generated from PDB. In
122 the histogram, signals in the region $r < 4.0$ Å arise due to interactions with partially dehydrated Mg^{2+} ions, and therefore
123 are not present in $g_{Mg^{2+}-\alpha}$. The key point here is that the closest Mg^{2+} could approach either moieties in the simulations is
124 around 4.0 Å, which is in good agreement with the histogram, and therefore justifies the choice of the divalent ion radius used
125 in our coarse-grained simulations. In addition, the agreement between our simulations and experimental data for a wide variety
126 of thermodynamic properties furthermore justifies our choice of the divalent ion radius.
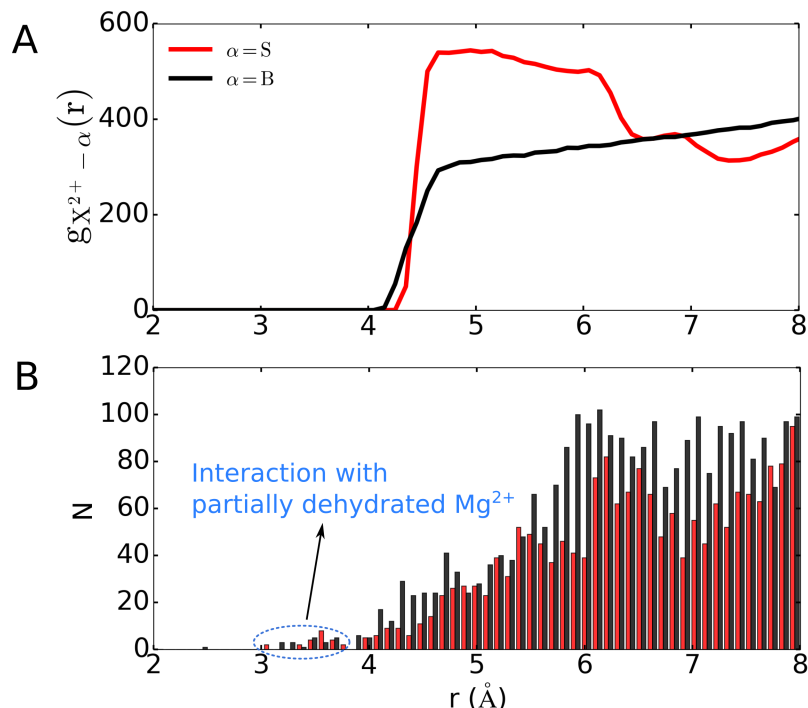
Fig. S1. Comparison of the distances between Mg²⁺ and sugars (red) and bases (black). (A) Radial distribution function of Mg²⁺ and sugar (red) or base (black) calculated in simulations of BWYV. (B) Histogram of the distances between Mg²⁺ and sugar (red) or base (black) in 147 RNA structures. There is only a small occurrence of interaction between partially dehydrated Mg²⁺ with both sugar and base moieties in RNA at $r < 4.0$ Å.

Stacking interactions U_{ST}. Interactions between two consecutive bases, or secondary stacking, are modeled using $U_{ST} = \frac{U_{st}^o}{1+u_1}$ where u_1 is a linear combination of harmonic constraints, which biases the stacking topology to the A-form helix (19), and $U_{st}^o = -h + k_B (T - T_m) s$ where h and s are independently obtained for the 16 nucleotide dimers by reproducing their experimental stacking thermodynamics (19). In the simulations, we computed the stability of the stacked dimers using:

$$\Delta G = -k_B T \ln p + k_B T \ln (1 - p) + \Delta G_o, \qquad [7]$$

where p is the fraction of all sampled conformations for which $U_{ST} < -k_B T$. The parameters h (but not s) in U_{st}^o, thus, are functions of a single free energy correction term, ΔG_o. The value of ΔG_o, assumed to be a constant for all dimers, is used to adjust the balance between stacking and hydrogen bonding, which is essential to accurately reproduce RNA thermodynamics (see below).

Stacking between non-consecutive bases is detected from the input structure using the geometric criteria reported elsewhere (21). We evaluated tertiary stacking using $U_{ST} = \frac{U_s}{1+u_2}$, where $U_s = -5.0$ kcal/mol, and u_2 is also a linear combination of harmonic constraints, similarly to u_1, but instead is chosen to bias the stacking topology to the crystal structure. Since a base could stack with others on both sides, we keep track of both the number of stacking each side participates in during the simulations. Each side is allowed to stack with a maximum of two other bases, and tertiary stacking is given a higher priority over secondary stacking. In other words, once tertiary stacks are formed (the two bases are closer than 10.0 Å) and the side reaches the maximum stacking capacity then the secondary stacking is disallowed.

Hydrogen bond potential U_{HB}. We used $U_{HB} = N_b U_{hb}^o \exp(-u_2)$ where u_2 has the same form as tertiary stacking, biasing the structure towards an A-form RNA for canonical bonds (G-C, A-U and G-U) or the experimental structure for non-canonical bonds; N_b is the number of hydrogen bonds between the beads. For Watson–Crick base pairing, N_b is 2 for A-U and G-U, and 3 for G-C. For non-canonical bonds, N_b is computed from the experimental structure of the RNA. Hydrogen bonds involving S and P beads are also considered. We use $U_{hb}^o = -2.70$ kcal/mol, which is fit in order to reproduce heat capacities of RNA hairpins and pseudoknots (Fig. S2). Our model also permits non-native base pairing formed between G and C, A and U, G and U separated by at least 4 nucleotides along the chain. Thus, the folded structures can be disrupted, allowing for non-native structures to be populated in the simulations. Each bead has a maximum number of hydrogen bonds it could potentially form, and one base is involved in only one canonical base-pair.

Choice of ΔG_o **and** U_{hb}^o**.** Following our previous studies, we calibrated the parameters to reproduce known experimental quantities. Here, we adjusted ΔG_o in Eq. 7 (0.90 kcal/mol) and U_{hb}^o (-2.70 kcal/mol) to fit the heat capacity of human telomerase RNA hairpin (hTR HP) in 200 mM KCl and the Beet Western Yellow Virus pseudoknot (BWYV PK) in 500 mM KCl. In Fig. S2, the values of the melting temperatures for both the hTR HP and BWYV PK predicted theoretically are in good agreement with experiments.
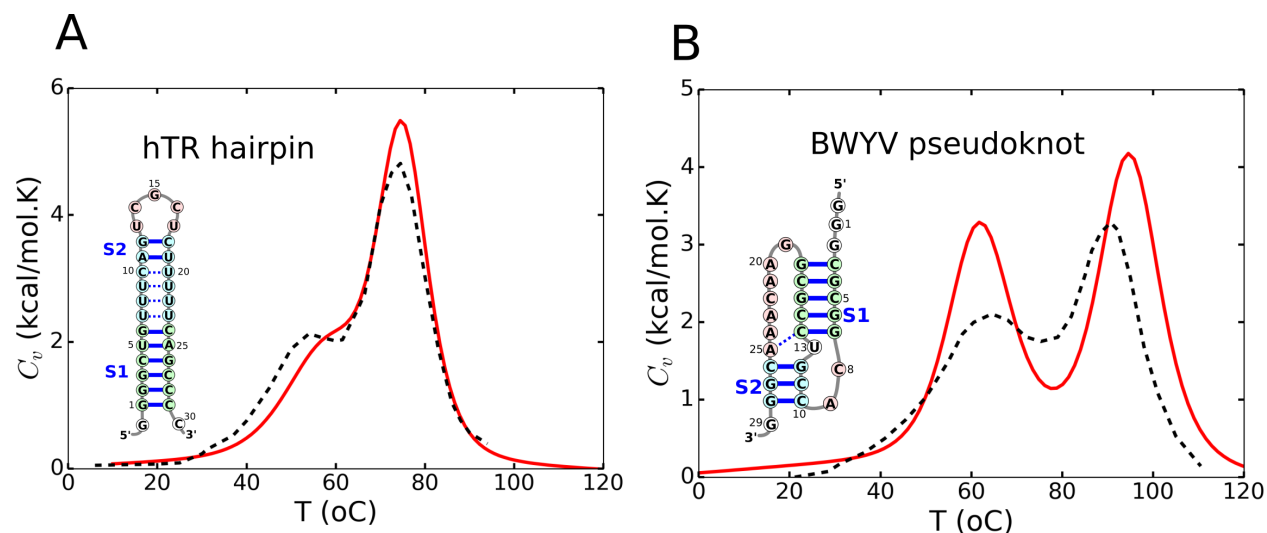
Fig. S2. Heat capacity of hTR hairpin (A) and BWYV pseudoknot (B). Simulations were performed for hTR in 200 mM KCl and for BWYV in 500 mM KCl at pH=7.0 with no divalent cations. Experimental data for hTR and BWYV are black dashed lines, taken from Ref. (22) and (23), respectively. Simulation data are shown as red lines.

Electrostatic interactions U_{EL}**.** We developed a new way to treat the interaction of ions with a charged phosphate group. A typical buffer used in RNA folding experiments (Tris for example) contains known amount of monovalent ions. To this buffer, a solution containing divalent cations is added in order to induce RNA folding. Thus, in such experiments, the solution contains a mixture of divalent and monovalent ions. Treating both the ions on equal footing, which we previously carried out for a few RNA constructs including the *Azoarcus* ribozyme (1), is computationally demanding. As explained here and in the main text, we treated monovalent ions implicitly and Mg^{2+}/Ca^{2+} explicitly. This procedure is justified because divalent cations are indispensable for the folding of most RNAs, with the exception of some small hairpins and pseudoknots. It is, therefore, important to include divalent ions explicitly to take into account the ion size and their specific interactions with the RNA. On the other hand, the majority of monovalent ions interacts with the RNAs via non-specific electrostatic interactions, screening the charge–charge repulsion between phosphate groups. There are few examples where specific interactions between monovalent ions and RNA play an essential role in RNA folding (24–26). Hence, it is reasonable to treat the monovalent ion as a continuum using classical Debye–Huckel theory. In most RNA folding experiments, the concentration of monovalent ions in the buffer solution typically is in far excess of the divalent cation, and therefore they screen the interactions between the divalent ions. Thus, we assume that the Debye–Huckel potential accurately describes the electrostatic interactions between divalent cations and P–P repulsions. With this approximation, U_{EL} is the sum of pairwise interactions between all the divalent cations, repulsions between the P groups, and attractive interactions between the divalent cations and the P groups. The bare charge on P is replaced by an effective charge, $Q(T, C_1, C_2)$, which is calculated using the counter ion condensation theory. The value of $Q(T, C_1, C_2)$ depends on the temperature, T, as well as the concentrations of monovalent (C_1) and divalent cations (C_2). It only remains to determine the interaction between the divalent cations and P, which is given by Eq. 5 in the main text. The PMF, $W(r)$, is calculated using the RISM theory in liquid state physics, described in the previous section.

Validation of the ion condensation theory. The assumption of the Oosawa–Manning counter ion condensation theory is that ions at distances that are larger than the size of the RNA corresponding to the bulk (B), and the ones condensed (C) onto the polyanion are at equilibrium (27). The value of the renormalized charge on the phosphate group is calculated by equating the chemical potentials of the B and C ions. In the main text, we derived an approximate expression of charge neutralization to obtain the effective charge on the phosphates using this physical picture. In order to assess if our estimate of charge renormalization is reasonable, we rely on experimental techniques that probe both the monovalent and divalent ion atmosphere around nucleic acids such as ion counting (BE-AES) or anomalous small-angle X-ray scattering (ASAXS) (28–30). Fig. S3 compares θ_{Na^+} and $\theta_{Mg^{2+}}$, $\theta_i = \frac{\Gamma_i}{N_P}\left(1 - \frac{b}{l_B}\right)$, computed using the theory at 20 mM NaCl and results from ion counting experiment for a 24bp duplex DNA. The experimental θ are calculated using the assumption that the total DNA charge neutralized is similar to the theory, hence the factor $\left(1 - \frac{b}{l_B}\right)$. This is necessary because we set $b = 4.4$ Å in our theory and simulations. We had to choose DNA because the data for RNA is currently not available. Because of the physics of ion condensation likely does not depend greatly on the differences between DNA and RNA, comparison with duplex DNA is sufficient to validate our theory. The calculated values are in good agreement with experiments, indicating that our theory describes ion competition in a buffer containing both monovalent and divalent cations accurately.

Hung T. Nguyen, Naoto Hori and D. Thirumalai
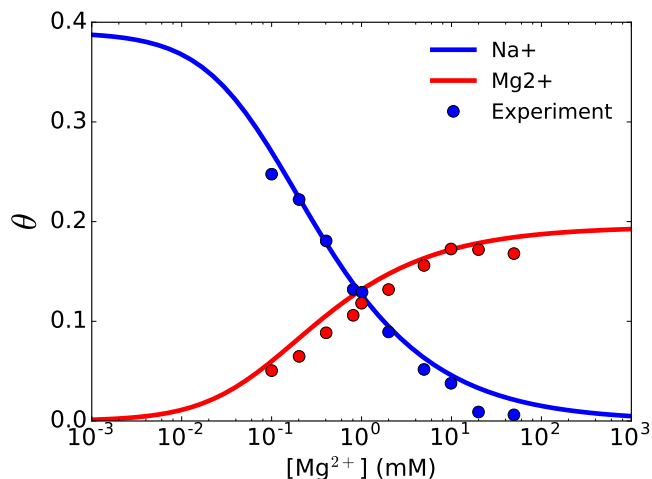
Fig. S3. Competition between Na$^+$ and Mg^{2+} ions around a 24bp DNA duplex. The plot shows the number of condensed ions per phosphate group at a fixed 20 mM NaCl solution as a function of Mg^{2+} concentration. Calculations were done by analytically solving Eqs. 3 and 4 (main text). The blue and red circles are experimental data from Ref. (28) and $\theta_i = \frac{N_i}{N_P}\left(1 - \frac{b}{l_B}\right)$, with $b = 4.4$ Å and $N_P = 46$ is the total number of phosphate groups in the DNA.

3. Simulation details and data analyses

Simulations. We performed simulations using the Langevin dynamics with the CG force field using an in-house code, which is available at https://github.com/tienhungf91/RNA_cg. Divalent cations were randomly added to a cubic box containing an RNA molecule, whose initial coordinates were taken from the structure of the folded state in the PDB. The box size varied from 700-3,000 Å depending on the bulk concentration of divalent cations. We used large boxes to make sure that at least 200 divalent cations were present in the simulations. Enlarging the box size does not introduce more particles into the system, but only dilutes the divalent cation concentration. The performance of our model, therefore, is insensitive to the box size, allowing us to probe the effect of the arbitrarily small concentration of divalent cations on RNA folding. This is another advantage of treating monovalent ions implicitly. We used periodic boundary conditions in the simulations to minimize the effect of finite box size. Numerical integration of the equations of motion was performed using the leap-frog algorithm with the time step $h = 0.05\tau$ where $\tau = a_0 \sqrt{\frac{m_0}{e_0}}$ is the unit of time, $a_0 = 1$ Å, $m_0 = 1$ Da and $e_0 = 1$ kcal/mol. We performed low-friction dynamics to increase the sampling efficiency of the conformations, in which the viscosity of water was reduced 100 times (31). Snapshots were recorded every 10,000 steps, from which only the last two-thirds were used to compute all the quantities of interest.

Calculation of the heat capacity. We performed replica-exchange simulation (REMD) at several temperatures (32). Exchange was attempted every 5,000 steps between neighboring replicas. The system energy was recorded every 10,000 steps, and the heat capacity was computed using $C_v = \frac{\partial U}{\partial T}$ with WHAM. The REMD was found to give converged results after $\sim 5 \times 10^8$ integration steps.

Calculation of the folding free energy, $\Delta G(c)$. For a given monovalent concentration C_1, the folding free energy of the RNA is calculated using: (19)

$$G(T) = G(T^*) + \frac{\partial G}{\partial T}(T^*)(T - T^*) + T\frac{\partial^2 G}{\partial T^2}(T^*)\left(T^* - T + T\ln\frac{T}{T^*}\right),$$

where T^* is the reference temperature. At temperatures that are low compared to the melting temperature, only the folded state of the RNA is predominantly populated. Thus, the free energy of the folded state, $G_f(T)$, can be determined by using, for instance, $T^* = 10$°C as the reference temperature. Similarly, the free energy of the unfolded state, $G_u(T)$, is computed using $T^* = 120$°C. The free energy of the intermediate state, $G_i(T)$, is computed with T^* in between the two melting temperatures. For BWYV PK, $T^* = 70$°C (see Fig. S2). We performed REMD simulations at several temperatures and used WHAM to compute $G(T)$. The free energy for each state is then calculated using the above equation with appropriate reference temperatures. The folding free energy is then evaluated using $\Delta G_{f-u}(T) = G_f(T) - G_u(T)$.

Equilibrium between the bulk and condensed ions and $\Gamma_{X^{2+}}$. Because the divalent ions are attracted strongly to the highly negatively charged RNA, the actual ion bulk concentration differs from the concentration computed by dividing the total number of ions by the volume of the simulation box. Failure to account for the equilibrium between these populations results in an incorrect calculation of $\Gamma_{X^{2+}}$ and $\Delta G_{X^{2+}-RNA}$. One could enlarge the simulation box to alleviate this problem, but it is computational demanding. Instead, we calculated the ion concentration in the bulk C_2 after the concentration profile of

ions, $C_2(r)$, plateaus at large separation from the RNA (Fig. S4). The preferential interaction coefficient is then evaluated as $\Gamma_{X^{2+}} = C_2 \int \left(\frac{C_2(\mathbf{r})}{C_2} - 1 \right) d\mathbf{r}$. In practice, we truncated the integration at the distance where $C_2(r) = C_2$.
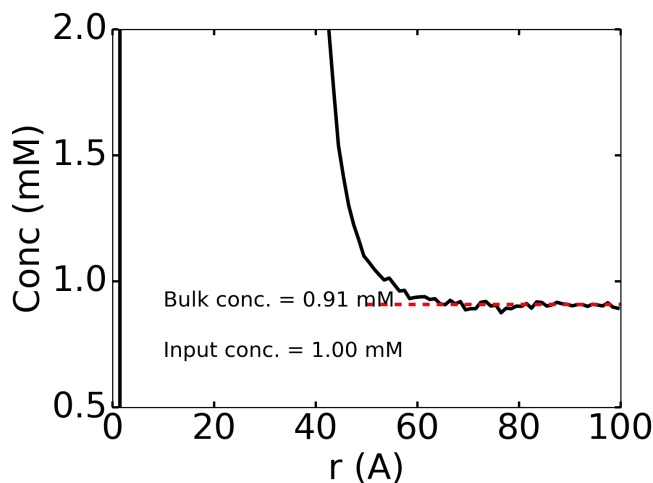
Fig. S4. Concentration profile of Mg^{2+} around adenine riboswitch. The bulk concentration is computed at a large separation from the center-of-mass of the RNA and subsequently used for $\Gamma_{Mg^{2+}}$ determination.

X^{2+}–RNA free energy. The plot of $\Gamma_{X^{2+}}$ vs. $\ln C_2$ curve (in Fig. 2 in the main text, for example) (X^{2+} is either Mg^{2+} or Ca^{2+}) is fit using a fourth order polynomial, $y = b(x-a)^2 + c(x-a)^3 + d(x-a)^4$ and the fit polynomial is integrated analytically, which allows us to evaluate the integral in Eq. 7 analytically and obtain $\Delta G_{X^{2+},RNA}$.

Fraction of native contacts. The fraction of native contacts, $Q(X_k)$, for an RNA conformation X_k is computed using: (33)

$$Q(X_k) = \frac{1}{N} \sum \frac{1}{1 + \exp\left[\alpha\left(r_{ij}(X_k) - \lambda r_{ij}^o\right)\right]},$$

where the sum runs over N pairs of native contacts (i,j) separated by r_{ij}^o in the crystal structure, $\alpha = 5$ Å^{-1} is a smoothing parameter and $\lambda = 1.5$ accounts for fluctuations when contacts formed. The list of N contacts are determined by native hydrogen bonds and stackings in the PDB structure, including secondary and tertiary interactions.

4. Robustness of the model

BWYV. In order to provide additional evidence of the robustness of the force field, we calculated the free energy difference, $\Delta G_{F-U}(c)$, between the folded and unfolded states, as a function of monovalent salt concentrations C_1 for BWYV pseudoknot. Fig. S5 shows that the simulated values of $\Delta G(c)$ are in very good agreement with experiments.

Hung T. Nguyen, Naoto Hori and D. Thirumalai
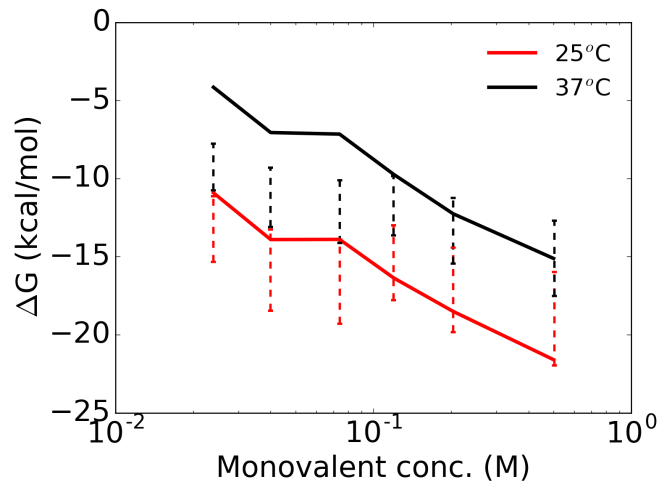
Fig. S5. Free energy difference (ΔG) between the folded and unfolded states of BWYV pseudoknot as a function of monovalent concentration (C_1) evaluated at 25℃ (red) and 37℃ (black). Calculated values are plotted as solid lines. The folding enthalpy and entropy reported in Ref. (34) were used to compute the experimental free energies shown with error bars. The calculated values of ΔG from simulations are in remarkable agreement with experiments, except for an underestimation at 37℃ at low ion concentrations.

58-nt rRNA. As a further validation, we calculated the heat capacity (C_v) of the 58-nt fragment of rRNA (Fig. S6) and compared the results with the UV absorbance data (35). The melting temperature at 20 mM KCl, identified with the maximum in C_v, agrees reasonably with the experimental data. The value of the high temperature peak in C_v at 60 mM KCl and 1 mM MgCl$_2$ is in good agreement with the experimental data. However, the shoulder at the lower temperature obtained in the simulations, if it exists at all, is much less pronounced in experiments. The results in Figs. S2, S5 and S6 show that our RNA force field is sufficiently accurate to reproduce many aspects of the thermodynamics for several RNAs over a wide range of ion concentrations. It is worth remarking that currently there is no other computational model that can calculate ion-dependent folding thermodynamic properties of RNA, such as free energy changes and heat capacities, let alone achieve the level of accuracy reported here.
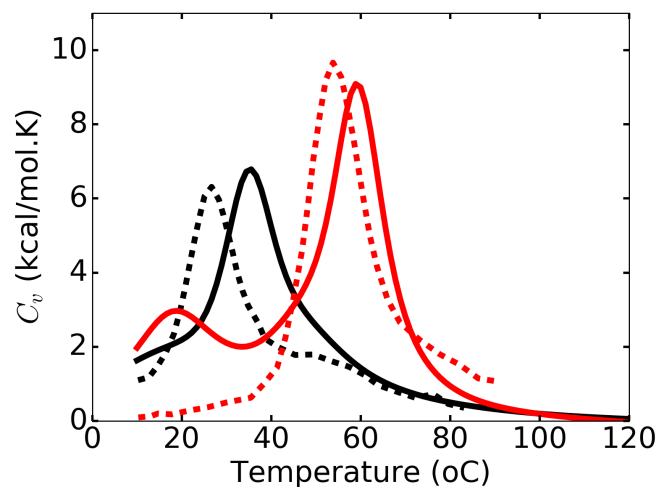

Fig. S6. Heat capacity of the 58-nt rRNA at two salt concentrations. Black–20 mM KCl (no Mg^{2+}), red –60 mM KCl + 1 mM MgCl$_2$. Dashed curves show UV absorbance data from experiments (35), solid lines are from simulations. We should stress that UV absorbance data is not the same physical variable as C_v, which is computed using the fluctuations in energy. Therefore, for the purposes of comparison, only the peak positions are relevant. The differences in the major melting temperature between the simulations and experiments are ~ 9℃ at 20 mM KCl (no Mg^{2+}) and ~ 5℃ at 60 mM KCl + 1 mM MgCl$_2$. The comparison further demonstrates that the agreement between simulations and experiments is excellent.
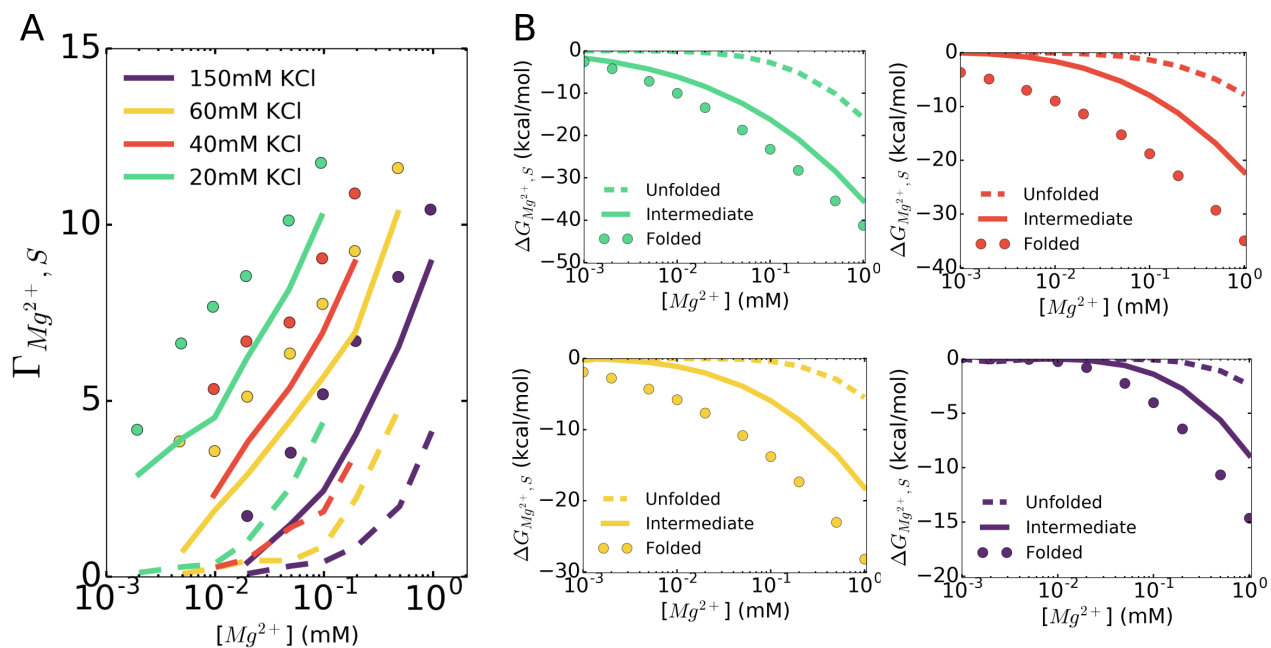
Fig. S7. (A) $\Gamma_{Mg^{2+},S}$ and (B) $\Delta G_{Mg^{2+},S}$ for the folded, intermediate and unfolded states of 58-nt rRNA. Simulations were performed by constraining the ensemble of RNA conformations to specific states. See the main text for details.
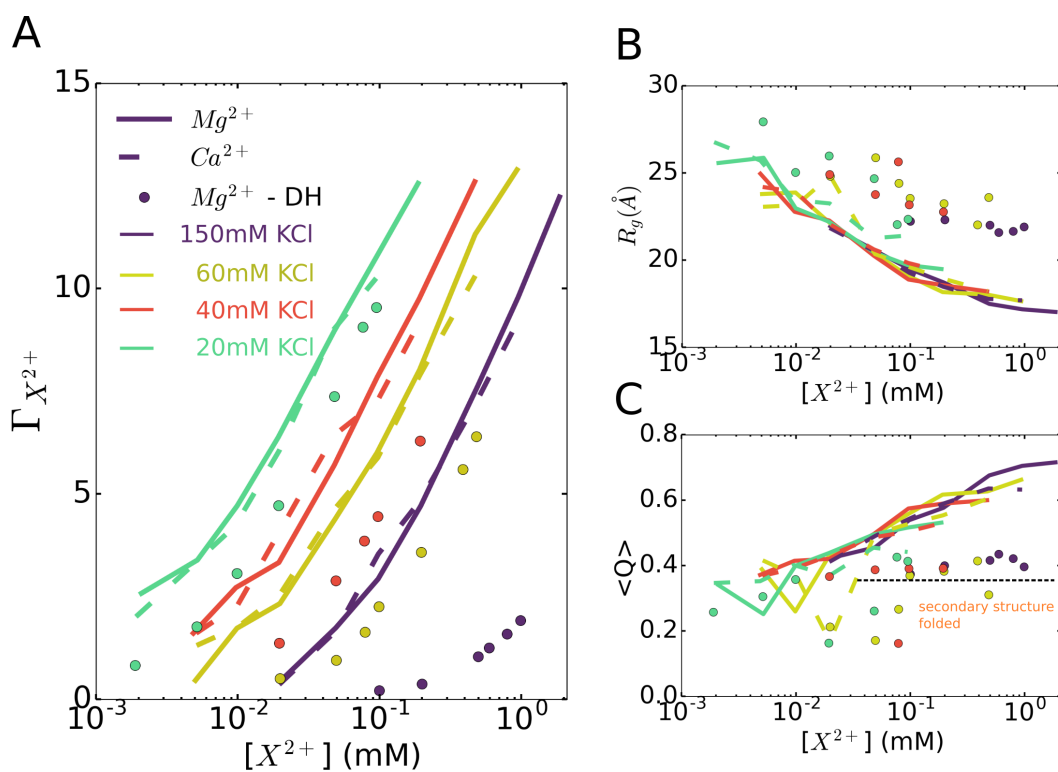
Fig. S8. Shown here are the full data for Fig. 7 in the main text for the 58-nt rRNA.

Hung T. Nguyen, Naoto Hori and D. Thirumalai
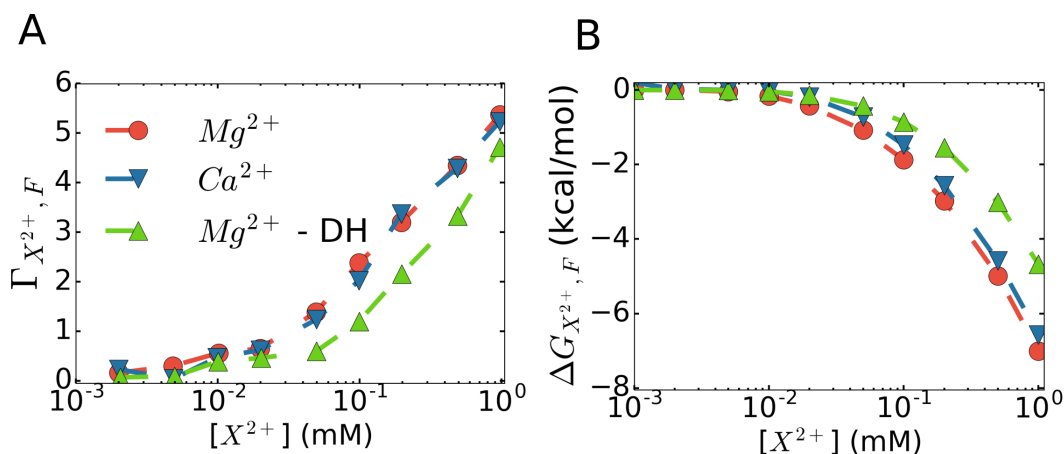
Fig. S9. Ion preferential interaction coefficients (A) and free energies of divalent ion–RNA interactions (B) of Mg^{2+} (red) and Ca^{2+} (blue) computed for BWYV at 54 mM KCl. Calculations were also performed for Mg^{2+}, in which the interactions are treated at the Debye–Huckel level (green).
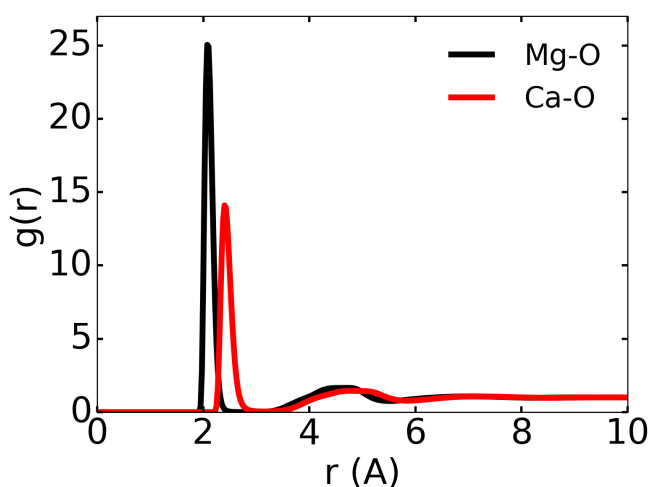
Fig. S10. g(r) between divalent ion and water computed from RISM theory.

References

1. NA Denesyuk, D Thirumalai, How do metal ions direct ribozyme folding? *Nat. Chem.* **7**, 793–801 (2015).
2. D Chandler, HC Andersen, Optimized Cluster Expansions for Classical Fluids. II. Theory of Molecular Liquids. *J. Chem. Phys.* **57**, 1930–1937 (1972).
3. F Hirata, P Rossky, An extended RISM equation for molecular polar fluids. *Chem. Phys. Lett.* **83**, 329–334 (1981).
4. F Hirata, B Pettitt, P Rossky, Application of an extended RISM equation to dipolar and quadrupolar fluids. *J. Chem. Phys.* **77**, 509–520 (1982).
5. BM Pettitt, PJ Rossky, Integral equation predictions of liquid state structure for waterlike intermolecular potentials. *J. Chem. Phys.* **77**, 1451–1457 (1982).
6. T Luchko, IS Joung, DA Case, Chapter 4:Integral Equation Theory of Biomolecules and Electrolytes in *Innovations in Biomolecular Modeling and Simulations: Volume 1*. (The Royal Society of Chemistry), pp. 51–86 (2012).
7. F Hirata, Theory of Molecular Liquids in *Molecular Theory of Solvation*, Understanding Chemical Reactivity, ed. F Hirata. (Springer Netherlands) No. 24, pp. 1–60 (2004).
8. EL Ratkova, DS Palmer, MV Fedorov, Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.* **115**, 6312–6356 (2015).
9. IS Joung, T Luchko, DA Case, Simple electrolyte solutions: Comparison of DRISM and molecular dynamics results for alkali halide solutions. *J. Chem. Phys.* **138**, 044103 (2013).
10. J Perkyns, BM Pettitt, A site-site theory for finite concentration saline solutions. *J. Chem. Phys.* **97**, 7656–7666 (1992).
11. A Kovalenko, F Hirata, Self-consistent description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **110**, 10095–10112 (1999).

12. SM Kast, T Kloss, Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *J. Chem. Phys.* **129**, 236101 (2008).
13. T Luchko, et al., Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **6**, 607–624 (2010).
14. P Li, KM Merz, Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **10**, 289–297 (2014).
15. P Li, KM Merz, Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **117**, 1564–1686 (2017).
16. MT Panteva, GM Giambasu, DM York, Force Field for Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cd^{2+} Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **119**, 15460–15470 (2015).
17. P Li, BP Roberts, DK Chakravorty, KM Merz, Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **9**, 2733–2748 (2013).
18. C Hyeon, D Thirumalai, Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci.* **102**, 6789–6794 (2005).
19. NA Denesyuk, D Thirumalai, Coarse-Grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **117**, 4901–4911 (2013).
20. H Zheng, IG Shabalin, KB Handing, JM Bujnicki, W Minor, Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection. *Nucl. Acids Res.* **43**, 3789–3801 (2015).
21. DE Condon, et al., Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **11**, 2729–2742 (2015).
22. CA Theimer, LD Finger, L Trantirek, J Feigon, Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc. Natl. Acad. Sci.* **100**, 449–454 (2003).
23. PL Nixon, DP Giedroc, Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J. Mol. Biol.* **296**, 659–671 (2000).
24. YX Wang, M Lu, DE Draper, Specific ammonium ion requirement for functional ribosomal RNA tertiary structure. *Biochemistry* **32**, 12279–12282 (1993).
25. S Haider, GN Parkinson, S Neidle, Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *J. Mol. Biol.* **320**, 189–200 (2002).
26. J Li, JJ Correia, L Wang, JO Trent, JB Chaires, Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucl. Acids Res.* **33**, 4649–4659 (2005).
27. GS Manning, The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quart. Rev. Biophys.* **11**, 179–246 (1978).
28. Y Bai, et al., Quantitative and Comprehensive Decomposition of the Ion Atmosphere around Nucleic Acids. *J. Am. Chem. Soc.* **129**, 14981–14988 (2007).
29. SA Pabit, et al., Counting Ions around DNA with Anomalous Small-Angle X-ray Scattering. *J. Am. Chem. Soc.* **132**, 16334–16336 (2010).
30. HT Nguyen, SA Pabit, L Pollack, DA Case, Extracting water and ion distributions from solution X-ray scattering experiments. *J. Chem. Phys.* **144**, 214105 (2016).
31. JD Honeycutt, D Thirumalai, The nature of folded states of globular proteins. *Biopolymers* **32**, 695–709 (1992).
32. Y Sugita, A Kitao, Y Okamoto, Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **113**, 6042–6051 (2000).
33. RB Best, G Hummer, WA Eaton, Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci.* **110**, 17874–17879 (2013).
34. AM Soto, V Misra, DE Draper, Tertiary Structure of an RNA Pseudoknot Is Stabilized by "Diffuse" Mg^{2+} Ions. *Biochemistry* **46**, 2973–2983 (2007).
35. D Grilley, V Misra, G Caliskan, DE Draper, Importance of Partially Unfolded Conformations for Mg^{2+}-Induced Folding of RNA Tertiary Structure: Structural Models and Free Energies of Mg^{2+} Interactions. *Biochemistry* **46**, 10266–10278 (2007).

Hung T. Nguyen, Naoto Hori and D. Thirumalai