Supplementary Information

MCTBI: a web server for predicting metal ion effects in RNA structures

Li-Zhen SUN[1,2], Jing-Xiang Zhang[3] and Shi-Jie CHEN[1,‡]

[1]Department of Physics, Department of Biochemistry, and Informatics Institute, University of Missouri, Columbia, MO 65211

[2]Department of Applied Physics, Zhejiang University of Technology, Hangzhou 310023, China

[3]School of Science and Technology, Zhejiang International Studies University, Hangzhou 310012, China

E-mail: chenshi@missouri.edu

1. Electrostatic interaction energies

The electrostatic interaction energy of the RNA $\Delta U_{\rm RNA}$ is determined by the charged atoms in the backbone of RNA, namely the phosphorus atoms of phosphate groups. It includes the Coulombic interaction and dielectric polarization energies. In the MCTBI model, the polarization energy is computed from the Generalized Born (GB) model.[1–6] Thus, $\Delta U_{\rm RNA}$ could be expressed as:

$$\begin{aligned}\Delta U_{\rm RNA} =& \frac{1}{2}\sum_{ij}\frac{Z_iZ_je^2}{\varepsilon_{\rm R}r_{ij}} \\ &+\frac{1}{2}(\frac{1}{\varepsilon_{\rm W}}-\frac{1}{\varepsilon_{\rm R}})\sum_{ij}\frac{Z_iZ_je^2}{\sqrt{r_{ij}^2+B_iB_j\exp(-\frac{r_{ij}^2}{4B_iB_j})}} \\ &+(\frac{1}{\varepsilon_{\rm W}}-\frac{1}{\varepsilon_{\rm R}})\sum_i(\frac{1}{B_i})Z_i^2e^2\end{aligned} \tag{S1}$$

Here, subscripts i and j denote the phosphorus atoms. $Z_{i({\rm or}\,j)}e$ denotes the charge of the phosphorus atom $i({\rm or}\,j)$, $\varepsilon_{\rm R}$ (= 20 in our calculation) and $\varepsilon_{\rm W}$ are the dielectric constants of RNA and water, and r_{ij} is the distance between the phosphorus atoms i and j. $B_{i({\rm or}\,j)}$ is the Born radius for phosphorus atoms $i({\rm or}\,j)$. In the web server, $\varepsilon_{\rm W}$ is assumed to be dependent on the temperature:

$$\varepsilon_{\rm W} = 87.740-(0.4008)T+(9.398\times10^{-4})T^2-(1.41\times10^{-6})T^3 \tag{S2}$$

where T is the temperature in Celsius.

For the i-th TB ion, the interaction energy should include an additional term, the volume exclusion. In the web server, the excluded volume effect is modeled by a Lennard-Jones (LJ) potential. Therefore, the interaction energy of the i-th TB ion placed at site k is given by:

$$\begin{aligned}\Delta U_i(k) =& \sum_j\frac{Z_iZ_je^2}{\varepsilon_{\rm R}r_{ij}}+\sum_j u_o[(\frac{\sigma_{ij}}{r_{ij}})^{12}-\frac{\sigma_{ij}}{r_{ij}})^6] \\ &+(\frac{1}{\varepsilon_{\rm W}}-\frac{1}{\varepsilon_{\rm R}})\sum_j\frac{Z_iZ_je^2}{\sqrt{r_{ij}^2+B_iB_j\exp(-\frac{r_{ij}^2}{4B_iB_j})}} \\ &+(\frac{1}{\varepsilon_{\rm W}}-\frac{1}{\varepsilon_{\rm R}})(\frac{1}{B_i}-\frac{1}{B_i^0})Z_i^2e^2.\end{aligned} \tag{S3}$$

Here, the subscript i represents the i-th TB ions and j denotes all the other charged particles, including the phosphorus atoms in RNA and pre-existing $(i-1)$ TB ions. The second term in the above equation is the LJ potential with $u_o(=0.35)$ as the LJ constant and σ_{ij} as the equilibrium distance between particles i and j. B_i^0 is the Born radius of the ion in an isolated system.

2. The free energy of the weakly correlated ions

We call the weakly correlated ion as diffusely bound (DB) ions. The DB ions consist of weakly correlated Mg^{2+}, all monovalent cations Na^+ (or K^+) and anions Cl^-. Because the ion-ion correlation can be neglected for the DB ions, the free energy of the DB ions $\Delta G_{\rm d}$ could be calculated from the effective single-particle ion distribution solved from NLPB calculation:[7,8]

$$\begin{aligned}\Delta G_{\rm d} =& \frac{1}{2}\int \sum_{\alpha} c_\alpha(r) z_\alpha e[\psi(r)+\psi'(r)]d^3r \\ &+ \int \sum_{\alpha}[c_\alpha(r)\ln\frac{c_\alpha(r)}{c_\alpha^0} - c_\alpha(r) + c_\alpha^0]d^3r.\end{aligned} \tag{S4}$$

Here α denotes the ion species. $\psi(r)$ and $\psi'(r)$ are the electric potentials at position r with and without the DB ions in the solution. $c_\alpha(r)$ and c_α^0 represent the local and bulk concentrations of ion α. The local electric potentials ($\psi(r)$ and $\psi'(r)$) and the concentration $c_\alpha(r)$ could be solved from the NLPB calculation according to not only the charges in RNA but also the average distributions of the $N_{\rm b}$ TB ions. Therefore, in the above equation the first integral includes the free energy of the interaction between the DB ions and TB ions and the enthalpic part of the free energy for DB ions. The second integral is the entropic part of free energy for the DB ions.

3. The parallel MCID algorithm

To further enhance the computational efficiency, we tested the parallelization of the MCTBI algorithm using the simple open multi-processing (OpenMP) method. The OpenMP method is an implementation of multithreading,[9] a method based on Fork-Join model[10] with shared-memory. Because the implementation of MCID algorithm in the MCTBI model is the most time consuming part, we employ the Fork-Join model to the MCID procedure, as shown in the flowchart in Figure S1A. In the original MCTBI model, we run the loop part (see Figure 1B in main text) $M_{\rm f}$ times for the MCID algorithm. In the parallel MCTBI (P-MCTBI) model, the MCID calculation is partitioned into N threads, and each thread independently runs the loop part $M_{\rm f}/N$ times. To further accelerate the computational speed, the server employs the OpenMP (Open Multi-Processing) programming interface to solve NLPB (for the weakly correlated ions).

We tested the computational efficiency for the P-MCTBI model with the different numbers of the threads for a 40-base pair (80-nt) and 80-base pair (160-nt) RNA helix, respectively. As shown in Fig. S1B, with 1-24 threads, the computer time would be 24%-1.8% and 10%-0.6% that of the original TBI model for the 80-nt and 160-nt helices, respectively. We also tested the effective utilization (EU) for each of the N threads used in P-MCTBI model: EU $= T(1)/[N*T(N)]$, where $T(N)$ is the computational time. For example, EU$=1$ for an ideal parallel computation such that each thread has 100% utilization. As shown in Fig. S1C in the SI, with the increase in the thread number,

the EU in our computation first increases and then decreases. For the optimal balance between the computational time and the utilization, we use the 4-threading computation for RNAs shorter than 60-nt, 8-threading for RNA between 60 and 120-nt, and 12-threading for RNA longer than 120-nt.

4. The binding fractions

For each $N_{\rm b}$ TB Mg^{2+} ions in the TB region, the number of excess ions $\Gamma_\alpha(N_{\rm b})$, including the TB ions and DB ions, can be calculated as

$$\Gamma_\alpha(N_{\rm b}) = \begin{cases} N_{\rm b} + \int (c_\alpha - c_\alpha^0) d^3r, & \text{if } \alpha \text{ is } \mathrm{Mg}^{2+}, \\ \int (c_\alpha - c_\alpha^0) d^3r, & \text{if } \alpha \text{ is monovalention}. \end{cases} \tag{S5}$$

The binding fraction of the excess number of ions per nucleotide is given by:

$$f_\alpha = \frac{1}{N_{\rm p}} \sum_{N_{\rm b}=0}^{N_{\rm p}} \frac{\Gamma_\alpha(N_{\rm b}) Z_{\rm b} Z_{\rm d} e^{-\Delta U_{\rm RNA}/k_{\rm B}T}}{Z}. \tag{S6}$$

with $Z_{\rm b}$ and $Z_{\rm d}$ as the partition function of TB ions and DB ions, and Z is the partition function of the whole system. For a mixed salt solution, the total binding fraction of all ions is given by:

$$f_{\rm tot} = 2 \times f_{\rm Mg^{2+}} + f_{\rm Na^+(or\ K^+)} - f_{\rm Cl^-}. \tag{S7}$$

Here $f_{\rm Mg^{2+}}$, $f_{\rm Na^+(or\ K^+)}$, and $f_{\rm Cl^-}$ are the binding fractions of Mg^{2+}, Na^+ (or K^+), and Cl^-.

5. The free energy and its components

The free energy of the whole system can be expressed as:

$$\Delta G_{\rm tot} = -k_{\rm B}T \ln(Z). \tag{S8}$$

We divided the free energy into two parts. The first part is the electrostatic free energy of the RNA, which is given by:

$$\Delta G_{\rm RNA} = -k_{\rm B}T \ln(e^{-\Delta U_{\rm RNA}/k_{\rm B}T}). \tag{S9}$$

For a given RNA structure, the electrostatic free energy of the RNA $\Delta G_{\rm RNA}$ equals the electrostatic interaction energy $\Delta U_{\rm RNA}$. The second part is the ion-induced free energy:

$$\Delta G_{\rm ion} = \Delta G_{\rm tot} - \Delta G_{\rm RNA}. \tag{S10}$$

6. The average binding fraction of each nucleotide

In the MCTBI model, the TB region is divided into $N_{\rm s}$ sites. The distance between a site k and a nucleotide N of RNA is shorter than the distance between this site and other nucleotide, the site k is assumed to belong to nucleotide N. We first calculate the probability of finding a TB ion at site k

$$p(k) = \sum_{N_{\rm b}=0}^{N_{\rm p}} \frac{n(N_{\rm b}, k)}{M_{\rm f} \times M_{\rm b}} \times \frac{Z(N_{\rm b})}{Z} \tag{S11}$$

Here $n(N_b, k)$ is the number of N_b-ion distributions (out of the totally $M_{\rm f}M_{\rm b}$ sampled distributions) with site k occupied by a TB ion. Then the average binding fraction of each nucleotide $f(N)$ with N the index of a nucleotide is given by

$$f(N) = \sum_{k=1}^{N_{\rm s}} p(k) \text{if k belongs to N} \tag{S12}$$

7. The probability distribution of the TB Mg^{2+} ions

The web server also provides the probability distribution for the number of TB Mg^{2+} ions (instead of total excess ions). The probability for $N_{\rm b}$ TB ions is given by:

$$p(N_{\rm b}) = \frac{Z(N_{\rm b})}{Z}. \tag{S13}$$

References

[1] Still WC, Tempczyk A, Hawley RC, Hendrickson T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**: 6127-6129.

[2] Nymeyer H, Garcia AE. 2003. Simulation of the folding equilibrium of a-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA*. **100**: 13934-13939.

[3] Hawkins GD, Cramer C J, Truhlar DG. 1995. Solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **246**: 122-129.

[4] Zou X, Sun Y, Kuntz ID. 1999. Inclusion of solvation in ligand binding free energy calculations using the Generalized Born Model. *J. Am. Chem. Soc.* **121**: 8033-8043.

[5] Liu HY, Kuntz ID, Zou X. 2004. GB/SA scoring function for structure-based drug design. *J. Phys. Chem. B.* **108**: 5453-5462.

[6] Liu HY, Zou X. 2006. Electrostatics of ligand binding: parameterization of the generalized Born model and comparison with the Poisson-Boltzmann approach. *J. Phys. Chem. B.* **110**: 9304-9313.

[7] Theodoor J, Overbeek G. 1990. The role of energy and entropy in the electrical double layer. *Colloids Surf.* **51**: 61-75.

[8] Stigter D. 1995. Evaluation of the counterion condensation theory of polyelectrolytes. *Biophys. J.* **69**: 380-388.

[9] Chandra R, Dagum L, Kohr D, Maydan D, McDonald J, Menon, R. 2001. *Parallel Programming in OpenMP*. Academic Press, San Diego, CA.

[10] Cormen TH, Leiserson CE, Rivest RL, Stein C. 2009 Introduction to Algorithms (Third Edition). MIT Press, Cambridge, MA.

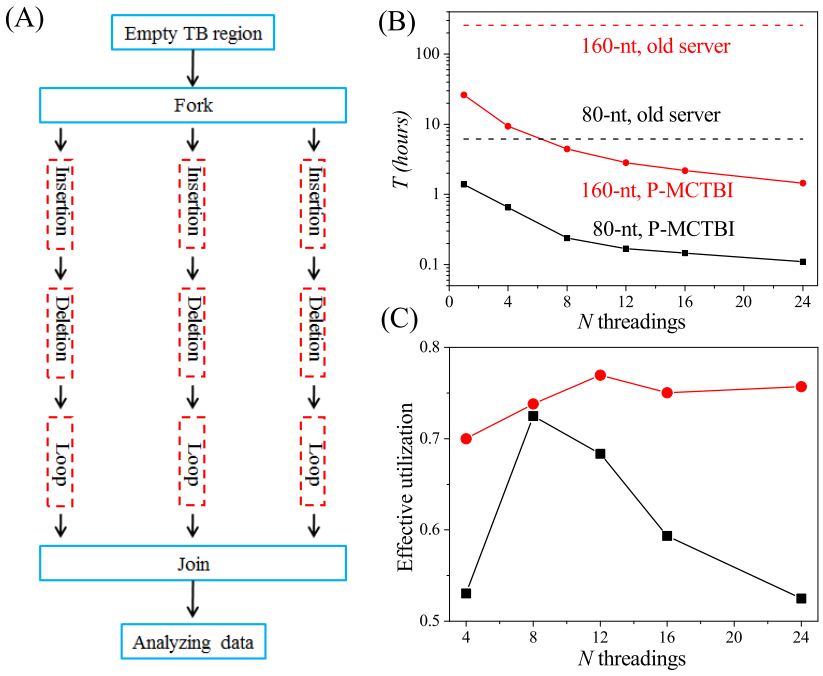

Figure S1: (A) The flowchat of the parallel MCID algorithm using the "Fork-Join" method. (B) The computational time as a function of N, the number of parallel threading, for various RNA helix lengths in a solution of 20mM Na^{+} and 5mM Mg^{2+}. The dashed lines show the computer time for the original TBI model. (C) The effective utilization of each threading as a function of the number of threadings used in the parallel MCTBI model. The red and black lines correspond to 160-nt and 80-nt RNA helix, respectively.

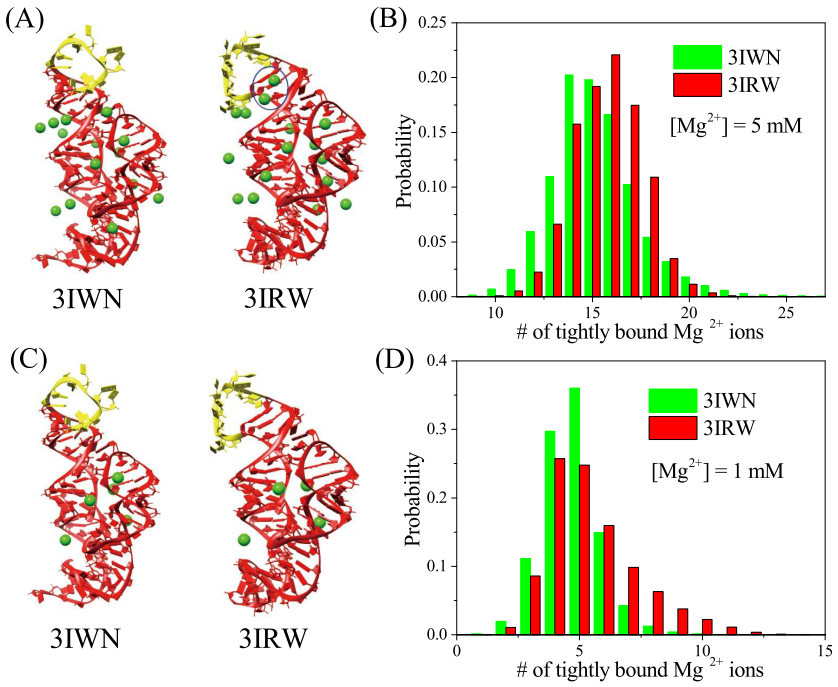

Figure S2: The most probable distribution of the TB ions around the structures of 3IWN and 3IRW at (A) 5mM Mg^{2+} and (C) 1 mM Mg^{2+}. (B) and (D) are the probability distributions for the two structures at the different salt conditions, respectively.

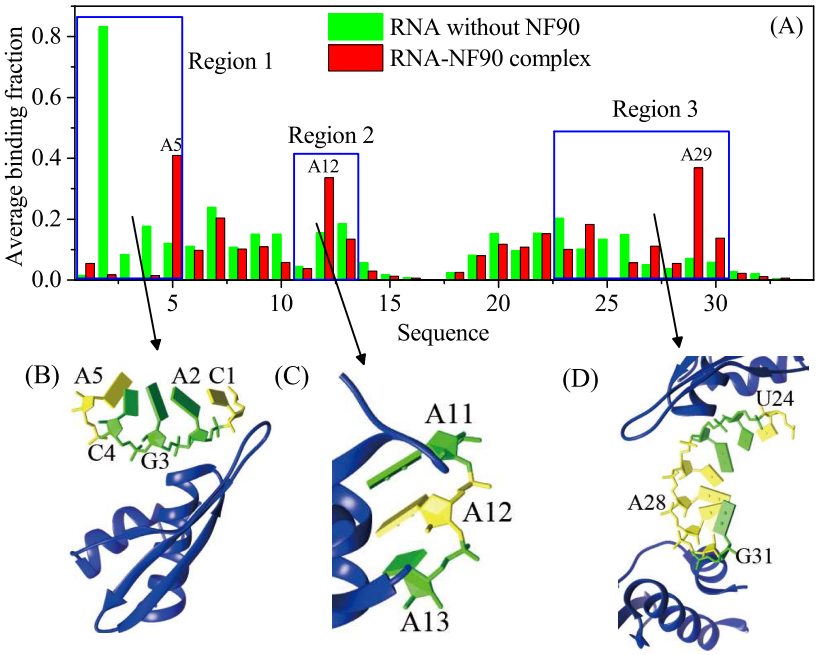

Figure S3: (A) The average binding fraction for each nucleotide for RNA and RNA-protein complex at $[Mg^{2+}]$ = 5 mM. (B), (C) and (D) show the parts of the structure for the RNA-protein complex corresponding to the three regions marked in (A). The (average) binding fraction of the nucleotides labeled with yellow/green color increases/decreases after the protein binding.

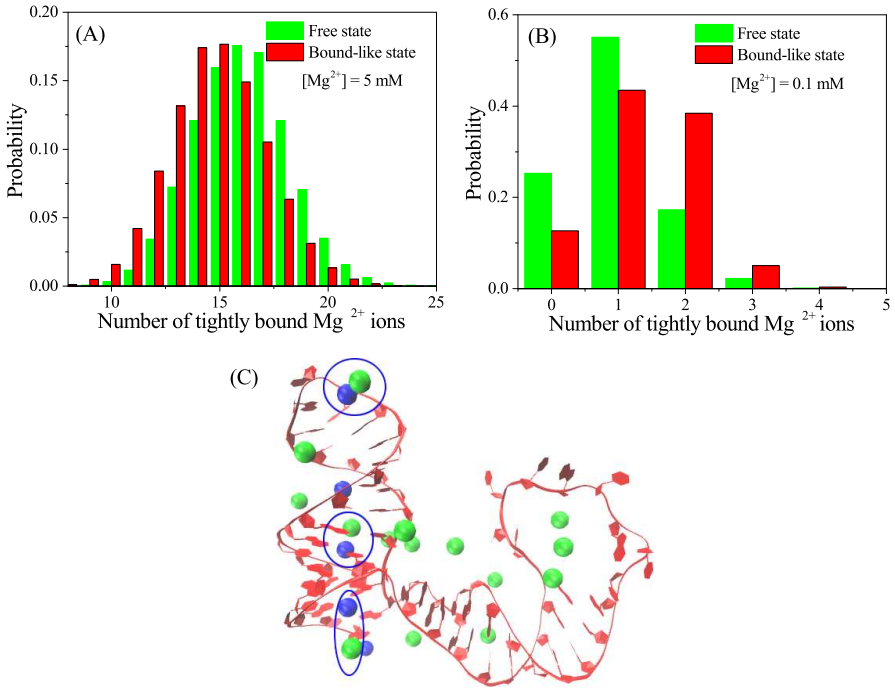

Figure S4: The probability distributions of the number of the TB Mg^{2+} ions for the free state and the bound-like state of the glutamine riboswitch at (A) $[Mg^{2+}]$ = 5 mM and (B) $[Mg^{2+}]$ = 0.1 mM. (C) The predicted most probable distribution for the TB Mg^{2+} ions (labeled by green spheres) around the RNA (labeled red). The experimentally determined specific binding sites are labeled with blue spheres.

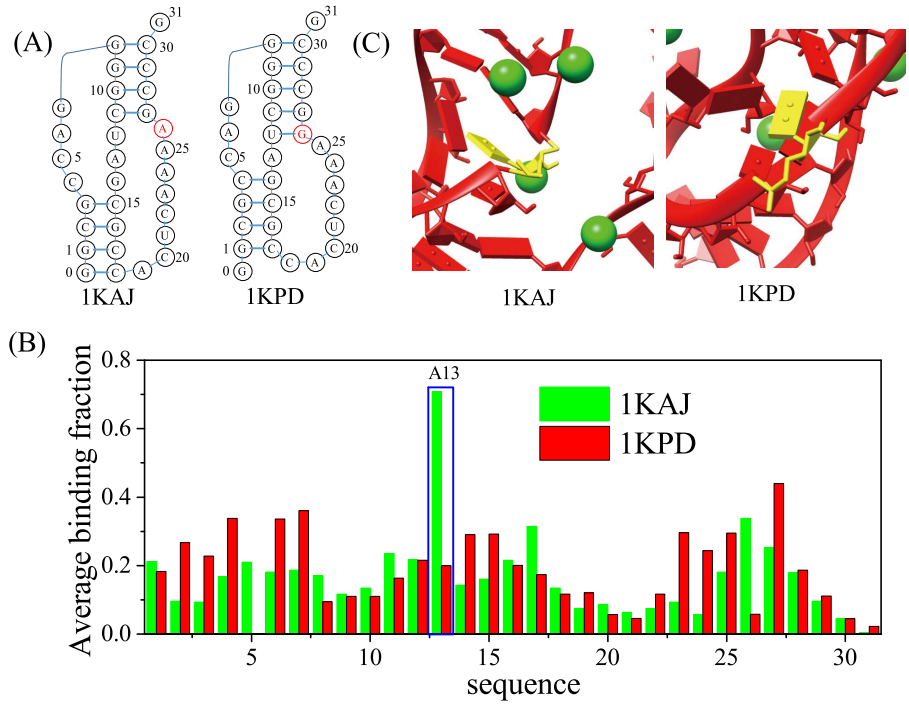

Figure S5: (A) The 2D structures of the RNA pseudoknot and its mutant. The mutant nucleotide is labeled by red color. (B) The comparison of the average binding fraction for each nucleotide for the two structures. (C) The predicted positions of the TB Mg^{2+} ions near nucleotide A13, which is labeled in yellow color.