# Exploring Electrostatic Energy Landscape for Tetraloop-Receptor Docking

## Supplemental Information

**Zhaojian HE, Yuhong ZHU and Shi-Jie CHEN**

## TBI model[1-6]

The polyvalent ions (such as $Mg^{2+}$) near the RNA surface can produce strong correlation between them, which will significantly affect the properties of the RNA folding. To precisely predict the RNA folding stability, such strong interaction should be taken into account for the calculation of the thermodynamic parameters of RNA in a polyvalent ion solution. The TBI model is developed to account for the ion correlation and fluctuation effect in ion-nucleic acid interaction.

The TBI model distinguishes the strongly correlated ions (called the "tightly bound ions") from the weakly correlated ions (called the "diffusive ions"). The spatial region for the tightly bound ions is called the tightly bound region. The tightly bound region is a thin layer around the RNA. Monovalent ions are usually weakly correlated and polyvalent ions can be strongly correlated around the nucleic acid. The diffusive ions are treated with the Poisson-Boltzmann equation (PB) and the strongly correlated ions (tightly bound ions) are treated based on the discrete ion distributions.

In order to enumerate the ion distributions for the tightly bound ions, for an N-nt RNA, the whole tightly bound region is divided into N cells, each around a phosphate. In the calculation, the space is repartitioned into many grids. Each grid inside the tightly bound region is assigned to a closest phosphate. In this way, a set of grids that are in close proximity of each phosphate can be uniquely identified. These grids constitute the tightly bound cell for each phosphate.

The total partition function $Z$ for the TB ions is given by the summation over all the possible binding modes $M$

$$Z = \sum_M Z_M, \tag{S1}$$

where $Z_M$ is the partition function for a given binding mode $M$.

$$Z_M = Z^{(id)} \left(\frac{N_Z}{V}\right)^{N_b} \left(\int \prod_{i=1}^{N_b} dR_i\right) e^{-(\Delta G_b + \Delta G_d + \Delta G_b^{pol})/k_B T}, \tag{S2}$$

where $Z^{(id)}$ is the partition function for the uniform ion solution. $N_z$ is the total number of z-valent counterions and $V$ is the volume of the solution. $N_b$ and $\int \prod_{i=1}^{N_b} dR_i$ are the tightly bound ions number and its volume integral. $\Delta G_{\mathrm{b}}$ is the mean Coulomb interaction energy between all the discrete charge-charge pairs (ion-ion, ion-phosphate and phosphate-phosphate) in the tightly bound region; $\Delta G_{\mathrm{d}}$ includes three parts, which are the free energy for the electrostatic interactions between the diffusive ions, the free energy of the electrostatic interaction between the diffusive ions and the discrete charges in the tightly bound region and the entropic free energy of the diffusive ions. $\Delta G_{\mathrm{b}}^{\mathrm{pol}}$ is the (Born) self-polarization energy for the discrete charges in the tightly bound region.

Therefore, the electrostatic free energy for a given RNA structure can be determined as

$$\Delta G^{e} = -k_B T ln \sum_M \left[ Z_M / Z^{(id)} \right]. \quad \text{(S3)}$$

For a given state $M$ (the docked or undocked state) of the tetraloop-receptor system, the electrostatic free energy is calculated as the average over the conformational ensemble. As explained in the main text, for a given state (docked or undocked), we classify the conformational ensemble into clusters. For the undocked state, there are 30 clusters. For the (more rigid) docked state, there is one cluster. We use the following equation to calculate its free energy $\Delta G_T(M)$ from the conformational ensemble:

$$\Delta G_T(M) = -k_B T \; ln \left[ \sum_i \left( \frac{N_i^{tot}(M)}{N_i(M)} \cdot \sum_j e^{-\Delta G_T^j(M)} \right) \right], \quad \text{(S4)}$$

where $\Delta G_T^j(M)$ is the electrostatic free energy for the $j^{th}$ conformation in the $i^{th}$ cluster for state $M$. The free energy $\Delta G_T^j(M)$ for a given conformation is calculated from Eq. (S3).$N_i^{tot}(M)$ in the above equation is the total number of the conformations in the $i^{th}$ cluster and $N_i(M)$ is the number of the conformations sampled (in the $i^{th}$ cluster). In our calculation, for each cluster, we sample $N_i(M) = 30$ conformations. The cluster index $i$ runs from 1 to 30 for the (flexible) undocked state and has a fixed value $i = 1$ (single fixed cluster) for the docked state.

**Thermodynamic parameters calculation**

The mean Coulomb energy $\Delta G_{\mathrm{b}}$ in TBI model can be defined as

$$\Delta G_b \simeq \sum_i \Phi_1(i) + \sum_{ij} \Phi_2(i,j), \quad \text{(S5)}$$

where $\Phi_1(i)$ and $\Phi_2(i,j)$ are the potential of mean force (PMF) for cell *i* and between cell *i* and cell *j*, respectively. $\Phi_1(i)$ and $\Phi_2(i,j)$ are given as

$$\Phi_1(i) = -k_B T ln\langle e^{-u_{ii}(R_i)/k_B T}\rangle;$$

$$\Phi_2(i) = -k_B T ln\langle e^{-u_{ij}(R_i,R_j)/k_B T}\rangle \quad \text{(S6)}$$

where $u_{ii}$ is the Coulomb interactions between the charges in cell *i* and $u_{ij}$ is the Coulomb interactions between the charges in cell *i* and in cell *j*.

To calculate $\Delta G_{\mathrm{d}}$, we use the results of the mean-field PB theory for the diffusive ions

$$\Delta G_d = \frac{1}{2}\int \sum_\alpha c_\alpha(r)\, z_\alpha e[\psi(r) + \psi'(r)]d^3r$$

$$+k_B T \times \int \sum_\alpha \left[c_\alpha(r) ln \frac{c_\alpha(r)}{c_\alpha^0} - c_\alpha(r) + c_\alpha^0\right] d^3, \quad \text{(S7)}$$

where the first and second integrals correspond to the enthalpic and entropic parts of the free energy, respectively. The value $\psi(r)$ and $\psi'(r)$ are the electrostatic potentials for the system with and without the diffusive salt ions, respectively. The value of $\psi'(r)$ is introduced because that the $\psi(r) - \psi'(r)$ gives the contribution of the diffusive ions total electrostatic potential.

The term $\Delta G_{\mathrm{b}}^{pol}$ is the change of the Born energies for the charges transferred from the bulk solvent to the tightly bound region. It can be described as

$$\Delta G_{\mathrm{b}}^{pol} = \sum_i \Phi_0(i), \quad \text{(S8)}$$

where $\Phi_0(i)$is the Born energy for charges inside the $i^{th}$ tightly bound cell. It is given as

$$\Phi_0(i) = -k_B T ln\,\langle e^{-(\Delta U_{\mathrm{P}}^{pol} + \Delta U_{\mathrm{I}}^{pol})/k_B T}\rangle, \quad \text{(S9)}$$

where $\Delta U_{\mathrm{P}}^{pol}$ and $\Delta U_{\mathrm{I}}^{pol}$ are the self-energies of the phosphate *i* and of the ion at

position $R_i$, respectively. The notation $\langle ... \rangle$ designates the averaging over all possible ion position $R_i$ within the cell.

**Comparison for the receptor structure between the docked and undocked states**

The structures of receptor are different in docked and undocked motif. In our calculation, the different PDB structures are applied for constructing the structure of two states. Fig. S1 shows the 3D-structures of the receptor in docked (green) and undocked (orange) states. These two structures are the fragments cut from the PDB structures (PDB code: 1GID for docked state and 1TLR for undocked state).[7,8]

**Effect of nonpolar hydration**

We estimate the change of the nonpolar solvation free energy $\Delta G_{np}$ in the docking process from the change of the solvent accessible surface area (SAS) with a uniform surface tension coefficient.[9-11]

$$\Delta G_{np} = \gamma \cdot SAS + c, \quad \text{(S10)}$$

where $\gamma$ is the surface tension coefficient, which represents the contribution to the nonpolar solvation free energy per unit surface area, $c$ is a constant. The coefficient $\gamma$ varies with the different molecules. Here we use a medium value 0.019kcal/mol·Å$^2$ to estimate the nonpolar solvation free energy of two states.

The SAS is calculated from the NASSESS software (http://www.bioionf.manchester.ac.uk/naccess). We calculate the average values of the SAS from 900 conformations for the undocked state and 30 conformations for the docked state. We find the average SAS is reduced from around 17150Å$^2$ for the undocked state to around 16540Å$^2$ for the docked state, corresponding to a free energy decrease of $\Delta G_{np}$ ~11kcal/mol. The value of $\Delta G_{np}$ is close to the theory-experiment difference for the entropic component of the docking free energy. We note that our electrostatic calculation does not account for the nonpolar solvation energy. The nonpolar solvation energy may be responsible for the theory-experiment difference in the entropic free energy.

To further investigate the distribution of SAS over the RNA structure, we choose

undocked conformations (Fig. 10*a*) and docked conformations (Figs. 2*b, c* and *d*) to calculate the SAS for each nucleotide in the tetraloop-receptor system. We find that the SAS change mainly comes from the tetraloop-receptor complex and the SAS increase due to the bending of the linker conformations in the docked state (see Fig. S2).

## Reference:


1 Tan, Z. J.; Chen, S. J. *J. chem. phys.* 2005, **122**, 44903.

2 Tan, Z. J.; Chen, S. J. *Nucleic Acids Res.* 2006, **34**, 6629.

3 Tan, Z. J.; Chen, S. J. *Biophys. J* . 2008, **94**, 3137.

4 Tan, Z. J.; Chen, S. J. *Biophys. J* . 2010, **99**, 1565.

5 Tan, Z. J.; Chen, S. J. *Biophys. J* . 2011, **101**, 176.

6 He, Z.; Chen, S.-J. *J. Chem. Theory Comput.* 2012, **8**, 2095.

7 Costa, M.; Michel, F. o. *EMBO J.* 1995, **14**, 1276.

8 Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* 1996, **273**, 1678.

9 Simonson, T.; Bringer, A. T. *J. Phys. Chem.* 1994, **98**, 4683.

10 Vallone, B.; E.Miele, A.; Vecchini, P.; Chiancone, E.; Brunori, M. *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95**, 6103.

11 Raschke, T. M.; Tsai, J.; Levitt, M. *Proc. Natl. Acad. Sci. U. S. A.* 2001, **98**, 5965.

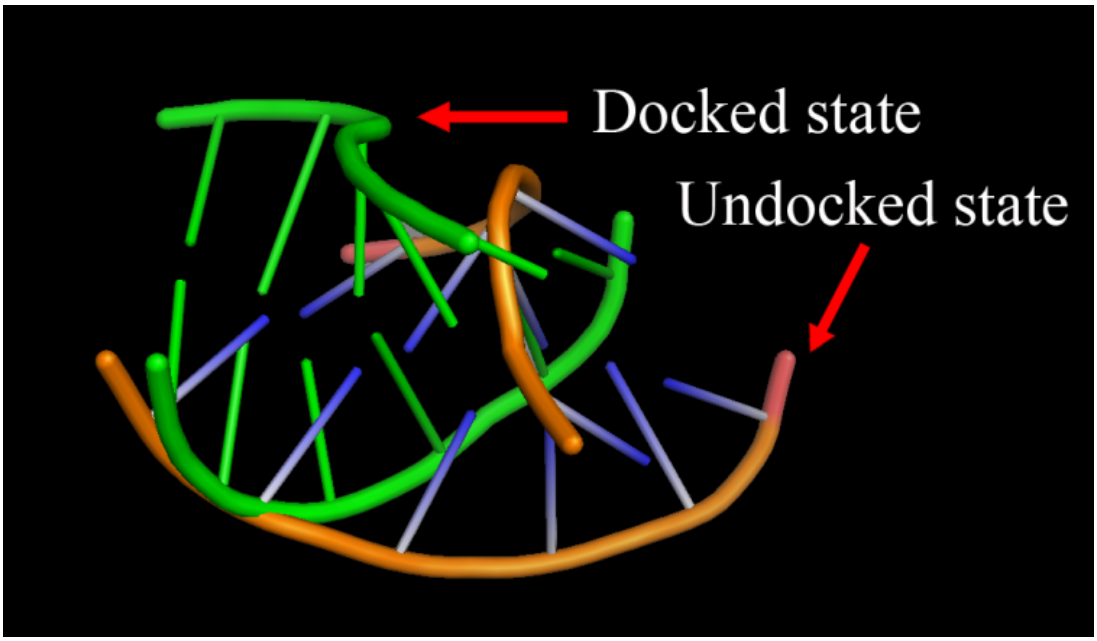


Figure S1**:** Comparison of the receptor structures for docked and undocked

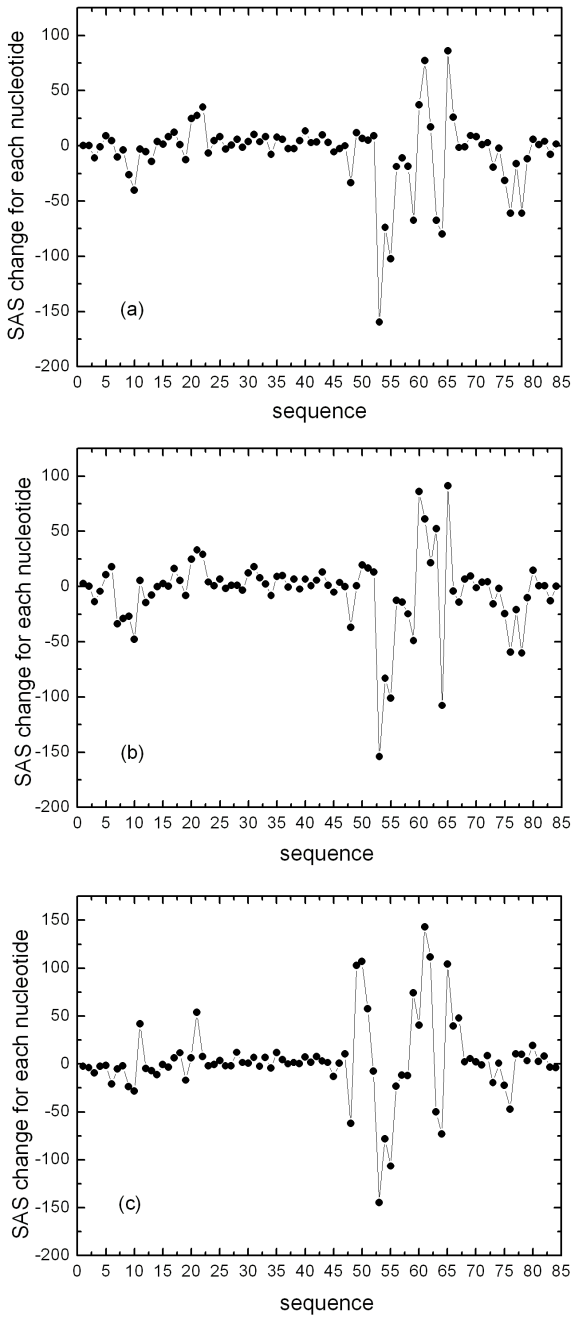


Figure S2: The distributions of the SAS changes for each nulecotide between the undocked and the docked states. The 3D structure of the docked state is shown in Fig. 2b. The conformations of the undocked state are shown Fig. 10b (a), Fig. 10c (b) and Fig. 10d (c), respectively.The figure shows the SAS change in the docking process in the different regions: the receptor (nucleotide number: 6-10, 74-79), the tetraloop (nucleotide number: 52-55), the bulge loop (nucleotide number: 20-21) and the linker (nucleotide number: 59-65) (see also Fig. 1b).

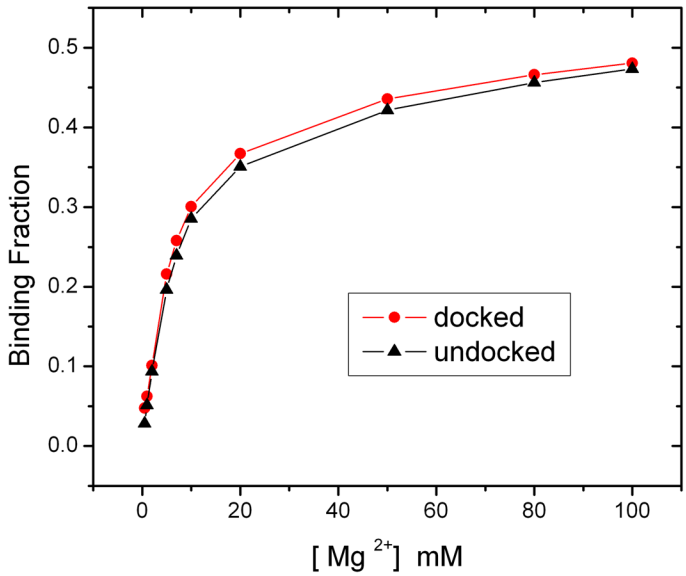


Figure S3: The binding fraction for the docked and the undocked states as a function of the $Mg^{2+}$ concentration in the background of 0.1M NaCl at 37℃. For the docked state, we randomly selected one structure from the ensemble of the docked conformations (Fig. 2d). For the undocked structures, we randomly select 8 out of the total 30 clusters (Fig. 2c), and then randomly select one structure from each of the 8 clusters. The result of the undocked state is the average over the 8 structures.

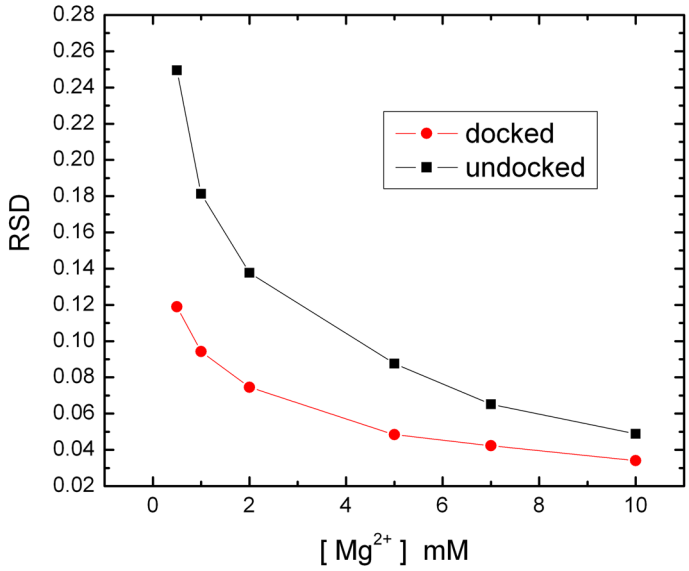


Figure S4: The RSD (relative standard deviation) of the binding fraction for the docked and the undocked states as a function of the $Mg^{2+}$ concentration in the background of 0.1M NaCl at 37℃. The calculation is based on the 4 complete sets of Monte Carlo-generated confromational samples for the two states. Here RSD is calculated from the average of the binding fraction $f_i$ over all the conformation $i$'s: $RSD = \sqrt{\sum_{i=1}^{N} \frac{(f_i - \bar{f})^2}{N-1}} / \bar{f}$, where $\bar{f}$ is the mean binding fraction.