Supporting Information

Predicting Monovalent Ion Correlation Effects in Nucleic Acids

Li-Zhen Sun\textsuperscript{1,2}, Yuanzhe HOU\textsuperscript{2} Shi-Jie CHEN\textsuperscript{2,*}

\textsuperscript{1}Department of Applied Physics, Zhejiang University of Technology, Hangzhou 310023, China
\textsuperscript{2}Department of Physics, Department of Biochemistry, and Informatics Institute, University of Missouri, Columbia, MO 65211

*Corresponding author; E-mail: chenshi@missouri.edu

1. The original MCTBI model

The gMCTBI model is developed based on the original MCTBI model.\textsuperscript{1} Therefore, we first briefly outline the key points in the original MCTBI model. The original MCTBI model was developed for computing the ion binding properties and the electrostatic free energies for a given nucleic acid structure in an ionic solution. The computation involves three main steps, as shown below.

1. NLPB is applied to roughly estimate the ion concentration distribution and the correlation strength around the nucleic acid. According to the ion correlation strength estimated from the ion concentration, the space around the nucleic acid is classified into a tightly bound (TB) region for TB (strongly correlated) ions and a diffusely bound (DB) region for DB (weakly correlated) ions.

2. The MCID algorithm\textsuperscript{1,2} is used to sample the TB ion distributions in the TB region. In this step, the TB ions are inserted one by one through Monte Carlo simulation until the TB region reaches saturated state with the number of the TB ions equal to the number of phosphates in nucleic acids. Then, the TB ions in the TB region are removed one after another through Monte Carlo simulation such that low energy TB ion distributions are sampled.

3. According to the TB ion distributions, we use NLPB to calculate the electrostatic free energy of DB ions. The electrostatic free energy includes interaction between the TB ions and DB ions.

The details of the original MCTBI model have been reported in previous works.\textsuperscript{1,2} In the previous studies,\textsuperscript{1,2} only divalent ions such as Mg\textsuperscript{2+} are considered for the TB ions while monovalent ions are all regarded as DB ions, regardless of their locations. In other words, the previous MCTBI model and the NLPB model are the same in the treatments for the monovalent ions.

2. The separate statistical weight $w(n)$

The separate statistical weight $w(n)$ for the $n$-th inserted ion is equal to:

$$w(n) = l^3 \sum_{k=1}^{m_k} e^{-\Delta U_n(k)/k_BT} \quad (S1)$$

Here $l$ is the lattice length and $m_k$ is the available (vacant) sites in the TB region. For an ion placed at site $k$, the
interaction energy $\Delta U_n(k)$ is dependent on the pre-existing TB ions as well as the nucleic acid charge distribution:

$$
\Delta U_n(k) = \sum_m Z_n Z_m e^2 \frac{1}{\varepsilon R_{nm}} + \sum_m u_o \left[ \left( \frac{\sigma_{nm}}{r_{nm}} \right)^{12} - \left( \frac{\sigma_{nm}}{r_{nm}} \right)^6 \right] + \left( \frac{1}{\varepsilon_W} - \frac{1}{\varepsilon_R} \right) \sum_m \frac{Z_n Z_m e^2}{\sqrt{r_{nm}^2 + B_n B_m \exp(-r_{nm}^2/4B_n B_m)}}
$$

where the other particles in the TB region (ions inserted in the previous steps and nucleic acid charges).

The first term in the above equation represents the Coulombic interaction energy between the $n$-th inserted ion and the rest charged particles in the TB region. $\varepsilon_R (= 20)$ is the dielectric constant of the nucleic acid and $r_{nm}$ is the distance between two particles. The Lennard-Jones (LJ) potential (the second term) is used to account for the volume exclusion effect. In the second term, $u_o (=0.35)$ is the LJ constant and $\sigma_{nm}$ is the equilibrium distance between particles $n$ and $m$. In our calculations, we set $\sigma_{nm}$ equals the addition of the radii of two particles. In the above equation, the mutual- and self-polarization energies (the third and the forth terms, respectively) are computed from the generalized Born model.$^3$–$^8$ Here $B_n$ and $B_m$ denote the Born radius for particles $n$ and $m$.

3. The electrostatic free energy of the DB ions

The electrostatic free energy for the DB ions $\Delta G_d$ can be determined from NLPB$^9,10$ for a given distribution of the charged particles in the TB region (including the TB ions and charged phosphates in nucleic acid):

$$
\Delta G_d = \frac{1}{2} \int \sum_{\alpha} c_\alpha(x) z_\alpha e \left\{ \psi(x) + \psi'(x) \right\} d^3x + \int \sum_{\alpha} \left\{ c_\alpha(x) \ln \frac{c_\alpha(x)}{c_\alpha^0} - c_\alpha(x) + c_\alpha^0 \right\} d^3x.
$$

where $\alpha$ denotes the ion species including the various cation species (such as $i$ and $j$ in Eqs. 2-4 in main paper) and anion species. $\psi(x)$ and $\psi'(x)$ are the electrical potentials at position $x$ around the nucleic acid with and without ions in the solution, respectively.

4. The prediction quantities from gMCTBI model

We apply the gMCTBI model to investigate the ion effects for nucleic acids in various salt conditions. To compare the theoretical predictions with the experimental data, we employ the gMCTBI model to compute the excess number of the bound cations and the depletion number of excluded anions $\Gamma_\alpha$:

$$
\Gamma_\alpha = \frac{\sum_{N^i_{0}=0}^{N^i_{\text{max}}} \sum_{N^j_{0}=0}^{N^j_{\text{max}}} \Gamma_\alpha(N^i_{b}, N^j_{b}) \times Z(N^i_{b}, N^j_{b})}{Z}
$$

Here $\alpha$ denotes the ion species, including cation $i$, cation $j$, and anion. $\Gamma_\alpha(N^i_{b}, N^j_{b})$ represents the excess number (or depletion number) of ion $\alpha$ in “ion atmosphere” (including the TB and DB regions) corresponding to the cases that
there are $N_i^b$ and $N_j^b$ TB ions in the TB region:

$$\Gamma_\alpha(N_i^b, N_j^b) = \begin{cases} N_\alpha^0 + \int (c_\alpha - c_\alpha^0)d^3r, & \text{if } \alpha \text{ is cation}, \\ \int (c_\alpha^0 - c_\alpha)d^3r, & \text{if } \alpha \text{ is anion}. \end{cases} \quad (S5)$$

Here $c_\alpha$ denotes the local concentration for DB ion species $\alpha$ in DB region. Since all anions are treated as DB ions according to our definitions, the depletion number of anions is computed by integrating the anion density (relative to the bulk) in the DB region. The excess number is an important physics quantity which establish a bridge between the theoretical predictions and the experimental measurements.\textsuperscript{11}

The gMCTBI model can predict not only the number but also the distribution of bound ions. Specifically, the model can give the linear density $\rho_I(x)\textsuperscript{12}$ (spatial distribution function\textsuperscript{13,14}) of the ion excess number:

$$\rho_I(x) = \frac{d\Gamma(x)}{dx} \quad (S6)$$

where $\Gamma(x)$ is the ion excess number within radial distance $x$. For a helix structure, $x$ denotes the distance to the central axis of the cylindrical structure,\textsuperscript{13,14} and for other more complicated structures, $x$ is the distance to the surface of the atoms in RNA.\textsuperscript{12}

Furthermore, the gMCTBI model can predict the average and the most probable distributions for the TB ions.\textsuperscript{1,2} For the average distribution of the TB ions, the probability of finding a TB ion of species $i$ at site $k$ in the TB region is given by:

$$p(k) = \sum_{N_i^b=0}^{N_i^b_{\text{max}}} \left[ \frac{n(N_i^b, k)}{N_{\text{sample}}} \times \left( \sum_{N_j^b=0}^{N_j^b_{\text{max}}} Z(N_i^b, N_j^b) \right) \right] \quad (S7)$$

where $N_{\text{sample}}$ is the number of the sampled ion distributions, $n(N_i^b, k)$ out of the $N_{\text{sample}}$ samples have $N_i^b$ TB ions with site $k$ occupied by an ion of species $i$. Moreover, the most probable distribution of TB ions is determined by the lowest interaction energy of many-particle system of the bound ions (including the TB ions and DB ions), as $\Delta U_{\text{min}} = (\Delta U_{TB} + \Delta G_d)_{\text{min}}$. Here $\Delta U_{TB}$ is the interaction energy of all TB ions, which can be calculated according to Eq. S2, and $\Delta G_d$ is the electrostatic free energy for the DB ions (see Eq. S3).
Figure S1: The linear densities of the DB ions (blue line), the TB ions (red line), and all bound ions (dashed line) around the DNA helix in the solution with single salt. Here [Na$^+$] = 200 mM.
References

[1] Sun, L. Z.; Chen, S. J. Monte Carlo Tightly Bound Ion Model: Predicting Ion-Binding Properties of RNA with Ion Correlations and Fluctuations. J. Chem. Theory Comput. 2016, 12, 3370–3381.

[2] Sun, L. Z.; Zhang, J. X.; Chen, S. J. MCTBI: a web server for predicting metal ion effects in RNA structures. RNA 2017, 23, 1155–1165.

[3] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J. Am. Chem. Soc. 1990, 112, 6127–6129.

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

[5] Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Solute descreening of solute charges from a dielectric medium. Chem. Phys. Lett. 1995, 246, 122–129.

[6] Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the Generalized Born Model. J. Am. Chem. Soc. 1999, 121, 8033–8043.

[7] Liu, H. Y.; Kuntz, I. D.; Zou, X. GB/SA scoring function for structure-based drug design. J. Phys. Chem. B. 2004, 108, 5453–5462.

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

[9] Overbeek, J. Th. G. The Role of Energy and Entropy in the Electrical Double Layer. Colloids Surf. 1990, 51, 61–75.

[10] Stigter, D. Valuation of the Counterion Condensation Theory of Polyelectrolytes. Biophys. J. 1995, 69, 380–388.

[11] Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic acid-Ion Interaction. Annu. Rev. Biochem. 2014, 83, 813–841.

[12] Hayes, R. L.; Noel, J. K.; Mohanty, U.; Whitford, P. C.; Hennelly, S. P.; Onuchic, J. N.; Sanbonmatsu, K. Y. Magnesium fluctuations modulate RNA dynamics in the SAM-I riboswitch. J. Am. Chem. Soc. 2012, 134, 12043–12053.

[13] Giambasu, G. M.; Luchko, T.; Herschlag, D.; York, D. M.; Case, D. A. Ion counting from explicit-solvent simulations and 3D-RISM. Biophys. J. 2014, 106, 883–894.

[14] Giambasu, G. M.; Gebala, M. K.; Panteva, M. T.; Luchko, T.; Case, D. A.; York, D. M. Competitive interaction of monovalent cations with DNA from 3D-RISM. Nucleic Acids Res. 2015, 43, 8405–8415.