Published in final edited form as:
J Chem Theory Comput. 2012 June 12; 8(6): 2095–2101. doi:10.1021/ct300227a.

Predicting ion-nucleic acid interactions by energy landscape-guided sampling

Zhaojian He and **Shi-Jie Chen***
Department of Physics, Department of Biochemistry, and Informatics Institute University of Missouri, Columbia, MO 65211

Abstract

The recently developed Tightly Bound Ion (TBI) model offers improved predictions for ion effect in nucleic acid systems by accounting for ion correlation and fluctuation effects. However, further application of the model to larger systems is limited by the low computational efficiency of the model. Here, we develop a new computational efficient TBI model using free energy landscape-guided sampling method. The method leads to drastic reduction in the computer time by a factor of 50 for RNAs of 50-100 nucleotides long. The improvement in the computational efficiency would be more significant for larger structures. To test the new method, we apply the model to predict the free energies and the number of bound ions for a series of RNA folding systems. The validity of this new model is supported by the nearly exact agreement with the results from the original TBI model and the agreement with the experimental data. The method may pave the way for further applications of the TBI model to treat a broad range of biologically significant systems such as tetraloop-receptor and riboswitches.

Keywords

ion-nucleic acid interaction; Tightly Bound Ion theory; RNA folding

1. Introduction

Nucleic acids (RNAs and DNAs) are negatively charged chain molecules. The negative backbone charges on nucleic acids attract cations in the solution. As a result, nucleic acid molecules are effectively dressed with ions. Ions play critical roles in stabilizing nucleic acid structures.[1–18] Understanding nucleic acid folding requires quantitative predictions of ion-mediated nucleic acid energetics. Most available thermodynamic parameters for nucleic acids are based upon thermal melting data for secondary structural motifs such as helix stems and loops under a fixed standard salt condition, namely, 1M Na^+.[19–23] These thermodynamic parameters have led to many successful predictions for nucleic acid secondary structures and folding stabilities.[19–23] However, the problem about accurate quantitative predictions for RNA folding stabilities under the different (non-standard) ionic conditions of the solution and furthermore, the ion effects in tertiary structural folding, remain unsolved. One of the key issues is about how to model the Mg^{2+} ion effect.

* Author to whom correspondence should be addressed; chenshi@missouri.edu.

Supporting Information
The free energy landscape $\Delta G(N_b, \Delta G_M)$ for the BWYV pseudoknot in the different ionic solutions. This information is available free of charge via the Internet at http://pubs.acs.org/.

Mg^{2+} ions can interact with nucleic acids through diffusive nonspecific and site-specific binding.[12] A variety of theories have been developed to treat Mg^{2+} ion binding. For example, for the nonspecific binding, the classical Debye-Huckel theory and the nonlinear Poisson-Boltzmann theory have been used to study the ion binding thermodynamics.[24–30] These theories can efficiently treat the large nucleic acids system. However, these theories do not explicitly consider the potential effects of ion fluctuation and correlation and ion dehydration in ion-RNA binding.

Site-specific ion effect can also play an important role for many RNA and DNA structures. Theoretical studies based on molecular dynamics (MD) simulations have provided invaluable insights into the atomic details for the energetics and dynamic ionic contributions.[18,31–38] For example, recently Chen et al used MD simulation to investigate the ion-mediated interactions in a RNA kissing loop and demonstrated the important roles of ion-specific dehydration effects in counterion binding to a kissing loops.[18] In another application of the MD simulation, Auffinger et al studied the Mg^{2+} binding to the 5S rRNA loop E motif and illustrated two distinct binding modes for pentahydrated Mg^{2+} in the system.[31] The above and many other studies showed the unique advantages of using MD simulations to investigate the ion-specific effects in ion-RNA interactions and the dependence on RNA structure and sequence.[18,31–39] However, the accuracy and efficiency of MD simulations are restricted by the force field and the computational expense of adequately exploring the configurations of nucleic acids and the surrounding water molecules and ions.[33] The problem is more pronounced for the study of the folding of large nucleic acids. A coarse-grained model, on the other hand, may provide more reliable (low-resolution) sampling for ion distribution and RNA structures. Such a model may provide a useful complement to MD simulations.

A compact tertiary structure involves significant charge build-up and strong binding of cations to the nucleic acid. The high local concentration of the counterions can potentially cause strong coupling (correlation) between the ions, i.e., ions form a network due to the strong Coulomb interactions. As a result, the electric field on a particle is a function of not only the coordinate of the particle but also the coordinates of all the other particles (ions). A correlated ion distribution can only be accounted for by discrete ion distributions (instead of the mean-field distribution). For multivalent ions, because of the strong Coulomb interactions, the correlation effect can be significant. Indeed, recent experimental findings suggested that ion correlation effect may be important for multivalent ions around nucleic acids.[6,40–42]

Motivated by the potential importance of ion correlation and fluctuation effects, we recently developed the tightly bound ion (called TBI) model.[43,48] The key idea in the TBI model is to account for the ion correlation effect by using many-ion distribution function instead of the mean-field single-ion distribution. In the TBI model we generate an ensemble of ion distributions by enumerating all the possible discrete modes for the many-ion distribution. We then calculate the many-ion interaction energy. In this way, the model can account for the many-ion correlation effect. Comparisons with experimental data showed that, by accounting for ion correlation and fluctuations, the model can offer improved predictions for the ion-dependence of nucleic acid folding stability.[44–48]

In practical calculation, for a given nucleic acid structure, we first run the Poisson-Boltzmann (PB) calculation to have a crude estimate for the ion concentration. Based on the distribution of the ion concentration, we calculate the spatial distribution of the Coulomb correlation strength. The correlation strength establishes a demarcation of space into the different regions: the tightly bound (TB) and the diffusively bound (DB) regions for the the correlation strength above and below a critical value, respectively. For monovalent ion

solution, the correlation effect is weak and nearly the whole solution is the DB region. For a solution containing multivalent ions, however, the TB region can become significant. The TB region usually corresponds to a thin layer surrounding the nucleic acid structure. For a divalent ion solution, the width of the TB region (layer) can vary from 0 to 3 Angstroms (depending on the multivalent ion concentration). Extensive tests have shown that the theoretical predictions are not very sensitive to the specific boundary between the TB and the DB regions.[50,51,53,54] Corresponding to the two regions, we classify two types of ions: the tightly bound (TB) ions and the diffusively bound (DB) ions, for ions in the TB and the DB regions respectively. For the DB ions, the correlation is weak and the fluid model (PB theory) would be able to provide an accurate description. For the TB ions, the correlation is strong, thus, the mean-field model would fail and we need a model that can account for discrete many-ion distributions instead of a mean single-ion distribution.

To further enumerate the different distributions of the TB ions, we discretize the TB region. Specifically, we divide the TB region into small cells (called TB cells), each around the phosphate of a nucleotide. For an N-nt nucleic acid structure, the TB region consists of N TB cells. To enumerate the ion distributions for an ensemble of N_b TB ions, we partition the N_b ions into N TB cells. Each such distribution is called an ion distribution mode or ion binding mode (M). It is important to note that each such mode is an N_b-particle distribution instead of a single particle mean-field distribution. Mathematically, an ion distribution mode M is an N_b-ion distribution as defined by the ion occupation numbers m_i (= 0 or 1) in each cell i: $M = \{m_1; m_2; \cdots; m_N\}$. Here m_i is the number of the TB ions in the ith cell. There exist totally 2^N such modes for a N-nt molecule. Previous applications and tests of the TBI model suggest that the model may provide improved predictions for ion-nucleic acid interactions than other models.[44–48] However, because of the rapidly increasing computational time for long chain lengths N, the application of the original TBI model has been mainly restricted to only small systems, which may have limited biological significance.

In the present study, on the basis of the recently developed all-atom TBI model,[48] we develop an energy landscape-based method to effectively reduce the effective number of modes. We sample the low-energy modes (in the basins of the free energy landscape) exhaustively and calculate the exact electrostatic free energy for each of the low-energy modes. For all the other (higher energy) ion distribution modes, we use approximate free energy functions. Tests for (long-chain) RNAs of length 50-90 nts show that the new method leads to 50-fold increase in the computational efficiency over the original TBI model. The improvement is even more dramatic for longer chains. Ion electrostatic effects for RNAs of length longer than 100-nts, which cannot be treated by the original TBI model due to the exceedingly long computational time, can now be readily predicted with only 2% of the original computer time. The new method thus leads to a significant step for TBI-based prediction of the ion effects in nucleic acid folding.

2. Method

2A. The original TBI model and its low computational efficiency

Central to the free energy calculation is the partition function, which involves the sum over all the possible ion distributions. In the TBI model,[43,48] the total partition function Z of the system is the sum of the partition function Z_M for all the possible ion distribution modes M:

$$Z = \sum_M Z_M \quad (1)$$

An ion distribution mode defines only the number (instead of the coordinates) of the TB ions in each TB cell. For a given mode, the TB ions can move around in the respective cells. We evaluate the partition function Z_M for a given mode M by averaging over all the possible positions of the TB ions:

$$Z_M = Z^{id}\left(c^0\right)^{N_b}\left(\int\prod_{i=1}^{N_b} d\mathbf{R_i}\right)e^{-\left(\Delta G_b+\Delta G_d+\Delta G_b^{pol}\right)/k_BT} \quad (2)$$

Here, Z^{id} is the partition function for the uniform ion solution (without the nucleic acids), N_b is the total number of the TB ions for the given mode, and c^0 is the bulk concentration of the counterion, $\mathbf{R_i}$ denotes the coordinate of the i_th TB ion. $\left(\int\prod_{i=1}^{N_b} d\mathbf{R_i}\right)$ is the volume integral over the TB region for the TB ions. The value ΔG_b and ΔG_b^{pol} are the mean Coulombic free energy (including the phosphate groups and the TB ions) and the (Born) self-polarization energy for the charges in the TB region, ΔG_d is the sum of the free energy for the DB ions (including the interactions between the DB ions and the TB ions).[48] It is important to note that for monovalent ions, unless the ion concentration reaches several M's or higher, the correlation effect can be neglected. In contrast, for divalent (such as Mg^{2+}) and higher valent ions, correlation could be important for ion concentrations at mM level. For a mixed monovalent and multivalent ion solution, because the correlation effect for the monovalent ions is negligible, we treat all the monovalent ions as DB ions which form a background for the multivalent ions.

Based on the partition function of the system (Eq. 1), we calculate the electrostatic free energy $G = -k_BT\ln(Z/Z^{id})$. Using 1M NaCl as the reference salt solution, we can compute the excess electrostatic energy: $\Delta G = G - G_0$, where G_0 is the free energy of the system in the 1M NaCl solution.

The number of ion distribution modes increases exponentially with the number of the phosphates (= the number of the nucleotides). The exceedingly long computational time prohibits the application of the TBI model to large nucleic acid molecules (see Fig. 1). In an attempt to reduce the computational time, we previously developed a hybrid method where the low-energy modes are exhaustively enumerated while the high-energy modes are randomly sampled.[44] Because the method avoids exhaustive enumeration for a large portion of the modes, namely, the high energy modes, it led to notable improvement in the computational efficiency. However, the improvement of the efficiency is quite limited because a significant portion of the modes still requires exhaustive enumeration. Indeed, the method cannot treat larger structures such as those with the number of nucleotides larger than 80 (see Fig. 1). In the present paper, we report a new method that can lead to drastic improvement in the computational efficiency.

2B. A new method: energy landscape-guided sampling

The new method is based on the free energy landscape of the system, i.e., the electrostatic free energy as a function of the ion distribution. By focusing on the most important modes, namely, those in the low-energy basins on the energy landscape, we here develop a new method that can significantly enhance the sampling efficiency in the partition function calculation. In the following, we first discuss several important features of the free energy landscape.

2B-1. Free energy landscape—We consider an N-nt RNA structure immersed in a Mg^{2+} solution. The number of the TB ions N_b can vary from 0 to N. For each given N_b,

there exist $\binom{N}{N_b}$ modes. We classify all the modes according to two parameters: N_b and the free energy $G_M = -k_BT\ln(Z_M/Z^{id})$, the free energy of the system for the given mode M of the TB ion distribution. We then evaluate the number of modes (NOM) $\Omega(N_b, \Delta G_M)$ for each given $(N_b, \Delta G_M)$, where $\Delta G_M = G_M - G_0$ is the difference between G_M and the free energy of the system in the 1M NaCl solution. From the NOM $\Omega(N_b, \Delta G_M)$, we calculate the free energy landscape of the system as a function of the above two parameters:

$$\Delta G(N_b, \Delta G_M) = \Delta G_M - k_B T \ln \Omega(N_b, \Delta G_M).$$

Physically, $\Delta G(N_b, \Delta G_M)$ is the free energy for the macrostate consisting of all the modes with the same free energy ΔG_M and the number of the TB ions N_b.

Using the BWYV system as an example, in Fig. 2 we show the NOM function of the system. The different panels in the figure correspond to the different $[Mg^{2+}]$ with a fixed 0.054M of Na^+ background. The state of full charge neutralization ($N_b = N/2$) has the largest combinatorial number of partitioning the N_b ions into the N TB cells. Thus, the largest number of the modes (the peaks in Fig. 2) occurs at $N_b = N/2$.

Mg^{2+} binding causes a decrease in the translational entropy of the ion (see, e.g., the factor $(c^0)^{Nb}$ in Eq. 2) and a resultant free energy penalty. For a low $[Mg^{2+}]$, Mg^{2+} binding results in a large free energy penalty, which, as shown in Fig. 2a, causes (a) a sharp, well separated energy distribution (NOM) and (b) higher free energies (see the x-axis values of the figures) at a low $[Mg^{2+}]$ such as 10μM. In contrast, for a high $[Mg^{2+}]$, the entropy decrease and free energy change for ion binding are much smaller. As a result, the free energy is much lower and the distribution is much broader, resulting in a significant overlapping in the NOM profile for the different N_b values (see Fig. 2f). For both the low and high Mg^{2+} concentrations, the different modes for a given N_b are distributed in a Gaussian fashion. However, for $[Mg^{2+}]$ in the intermediate regime, the energy distribution shows a non-Gaussian profile (see Figs. 2b-e). Such a complex energy distribution is due to the competition between the Na^+ and the Mg^{2+} binding and the energy variation for the different modes.

To conveniently visualize the free energy landscape, we plot a one-dimensional free energy profile: the free energy as a function of N_b, the number of the TB ions. For a given N_b (= 0, 1, ⋯, N for an N-nt RNA), we define the partition function $Z(N_b)$ by summing over all the possible modes that have N_b TB ions. The electrostatic free energy for a given N_b can be calculated from $Z(N_b)$ as $G(N_b) = -k_BT\ln Z(N_b)$ and $\Delta G(N_b) = G(N_b) - G_0$. Fig. 3 shows our predicted $\Delta G(N_b)$ with varying $[Mg^{2+}]$ in a fixed background of $[Na^+] = 0.054$M at 25°C. The energy landscape shows several important features.

1. As shown in Fig. 3a, the electrostatic free energy $\Delta G(N_b)$ is a smooth function of N_b. This is because a given N_b corresponds to an ensemble of modes and the ensemble average over all the modes may smooth out the energy change between the different modes. A smooth free energy landscape suggests the possibility of fitting the free energy using an analytical function.
2. For a given $[Mg^{2+}]$, the free energy landscape has a single minimum, corresponding to the (single) most stable ion distribution.
3. The state of the free energy minimum is sensitive to the ionic condition. As $[Mg^{2+}]$ is increased, the electro-static free energy of the system is lowered and N_b^*, the number of the TB ions for the most stable ion distribution (at the free energy

bottom), increases from zero at very low [Mg^{2+}] (no Mg^{2+}-induced charge neutralization) to about $N/2$ at higher [Mg^{2+}] (full charge neutralization from Mg^{2+}). The larger number of the TB ions $\left(N_b^*\right)$ at a higher [Mg^{2+}] shows that the TB ions play a more important role for higher Mg^{2+} concentrations. Furthermore, the (nearly) full charge neutralization at high [Mg^{2+}] results in a less sensitive dependence of the electrostatic energy on $N_b \left(\sim N_b^* \sim N/2\right)$. On the free energy landscape, this is manifested as the wider and shallower basin around the minima.

4. The partition function is overwhelmingly dominated by the ion distribution modes around N_b^* (the most stable state). Fig. 3b shows the ratio between $Z(N_b)$ for an arbitrary N_b and $Z\left(N_b^*\right)$ for the lowest free energy mode. It is clear that $Z(N_b)$ falls drastically as N_b deviates away from N_b^*. As N_b deviates away from N_b^* by 5-6 ions, the partition function $Z(N_b)$ falls below 10^{-4} of $Z\left(N_b^*\right)$.

2B-2. Energy landscape-guided mode reduction—The above features of the energy landscape suggest that in the partition function calculation, we can safely leave out the modes whose N_b's are far away from N_b^*. How to determine the effective range of N_b for the partition function calculation? The effective range of N_b is determined by the width of the free energy basin at the global minimum, which, due to the aforementioned correlation between the size of the TB region and the width of the free energy basin, can be loosely related to the thickness t of the TB region.

In order to derive a general rule for the selection of the modes, we have performed calculations for a series of structures: DNA and RNA duplexes, RNA hairpin, BWYV pseudoknot(PDB : 437D[49]), rRNA fragments(PDB code: 1HC8[50]), tRNA(PDB code: 1TRA[51]) and tetraloop-receptor RNA (PDB code: 2I7Z and 2ADT[52]) in the various Mg^{2+}/Na^+ solution with [Mg^{2+}] ranging from 10μM to 0.1M. For each structure in a given ionic condition, we compute the partition function for the different ionic conditions using (a) the exhaustive enumeration of all the possible modes and (b) local sampling around the free energy minimum. Through systematic calibration, we find the following relationship between the range of the effective $N_b \in \left[N_b^{(l)}, N_b^{(u)}\right]$ for the partition function and the average thickness (t) of the TB region:

$$N_b^{(l)} = max\{0, At - B\}, N_b^{(u)} = N \times (Ct + D) \quad (3)$$

The parameters A, B, C and D are tunable in accordance with the specific nucleic acids structures. In the tested DNAs and RNAs, we find A=C=0.23, B=0.32 and D=0.08. For a given structure in a specific ionic solution, if the $N_b^{(u)}$ (or $N_b^{(l)}$) value given by Eq. 3 is too close to or too far away from the most stable state N_b^*, from the partition function $Z(N_b)$ curve (see Fig. 3b), we can determine the $N_b^{(u)}\left(N_b^{(l)}\right)$ value by setting the partition function at $N_b^{(u)}\left(N_b^{(l)}\right)$ to be equal to $10^{-4}\, Z\left(N_b^*\right)$ (see Fig. 3(b)). In practice, since the partition functions for N_b near the boundaries $N_b^{(u)}$ and $N_b^{(l)}$ are very small, the total partition function for all the different modes is not very sensitive to the specific values of the A, B, C and D coefficients in the above equation.

The use of only the modes with N_b between $N_b^{(l)}$ and $N_b^{(u)}$ leads to a drastic reduction in the effective number of modes from 2^N to a much smaller number $\sum_{N_b=N_b^{(l)}}^{N_b^{(u)}} \begin{pmatrix} N \\ N_b \end{pmatrix}$. The reduction is more striking for low ion concentrations in which case the average thickness t of the TB region is small and hence the free energy basin is narrow. For example, for 58-nt rRNA fragment (PDB code: 1HC8[50]) in a Mg^{2+}/K^+ mixed solution with $[K^+]$ fixed at 0.02M, thickness t is below 1 Angstrom for $[Mg^{2+}]$ below 0.15mM. Correspondingly, the mode-reduction and the computer time-reduction are more significant. For the rRNA fragment in the Mg^{2+}/K^+ mixed solution, the computational time is reduced from about 43.0 hours for the original TBI model to 24 minutes for the current new TBI model.

For higher (multivalent) ion concentration, the average thickness t of the TB region is larger. For example, for the 58-nt rRNA fragment in Mg^{2+}/K^+ solution, t can exceed 1.5 Angstrom for $[Mg^{2+}]$ above 0.35mM. The larger t value leads to the larger number of the selected N_b's. For example, for t ~1.5 Angstrom, the range of the N_b's to be used in the calculation can reach $0.4N$. For a long sequence (large N), the reduced number of the modes can still be large and the computational time can still be demanding. In such a case, we use further approximations for the free energy landscape to enhance the computational efficiency. The smooth shape of the landscape $\Delta G(N_b)$ (Fig. 3) suggests a reliable fitting of the electrostatic free energy $\Delta G(N_b)$ as a polynomial of N_b. We randomly select 24 N_b's in the range between $N_b^{(l)}$ and $N_b^{(u)}$ and compute the electrostatic energy $\Delta G(N_b)$ for each selected N_b. From the 24 data points, we then fit $\Delta G(N_b)$ with an 8^{th}-degree polynomial of N_b. Our tests for 96-nt DNA duplex, tRNA, tetraloop-receptor RNA, show that the 24-point fitting causes relative error less than 1% and hence is reliable.

In the practical calculation, from the computed TB region, we estimate the mean thickness t and calculate $N_b^{(l)}$ and $N_b^{(u)}$ from Eq. 3. If there exist less than 24 N_b values between $N_b^{(l)}$ and $N_b^{(u)}$, we calculate the partition function $Z(N_b)$ and the electrostatic energy $\Delta G(N_b)$ for every N_b located in the range. If the N_b range exceeds 24, we randomly select 24 different N_b's in the range and calculate the $Z(N_b)$ and $\Delta G(N_b)$ for each selected N_b. From the 24 data points, we then fit the $\Delta G(N_b)$ curve with an 8^{th}-degree polynomial of N_b. From the $\Delta G(N_b)$ curve, we predict the binding fraction and electrostatic free energy of the system.

3. Application and Discussion

The purpose here is to test the new method through comparisons with the experimental data. In addition, through comparisons with the previous TBI method, we show the remarkable improvement in the computational efficiency (see Fig. 1). In order to make direct comparison with the experiments, we calculate the fractional number of bound ions. The fractional number of the bound ions $f_{Mg}{}^{2+}$ per nucleotide consists of the diffusively bound (DB) ions and the tightly bound (TB) ions:[46]

$$f_{Mg^{2+}} = \frac{\sum_{N_b} N_b Z(N_b)}{N \sum_{N_b} Z} + \frac{1}{N} \int \left[c_{Mg^{2+}}(\vec{r}) - c^0_{Mg^{2+}} \right] d^3\vec{r}, \quad f_{Na^+} = \frac{1}{N} \int \left[c_{Na^+}(\vec{r}) - c_{Na^+} \right] d^3\vec{r} \quad (4)$$

Here, the term involving $\sum_{N_b}$ is the number of the TB ions averaged over all the possible ion binding modes, the integral terms give the numbers of bound ions from the excess ion concentration,[43] and $c_{Mg}{}^{2+}$ and $c_{Na}{}^+$ are the local concentrations of the diffusively bound Mg^{2+} and Na^+ ions, respectively. In Fig. 4a, we show the predicted electrostatic free energy for BWYV pseudoknot. For comparison, the binding fractions determined in the

experimental measurements are also shown by black symbols in the figure. The theoretical predictions from the current mode-reduction method show agreements with the experimental data. Moreover, the current new method gives nearly exactly the same results as the original TBI predictions which are based on the brute force mode enumeration.[48]

To further test the improvement in the computational efficiency for the mode-reduction method developed here, we calculate the ion electrostatics for RNAs having long sequences (longer than 50-nts). For example, for the 58-nt fragment of ribosomal RNA (PDB code: 1HC8;[50] see Fig.4b) immersed in a KCl/$MgCl_2$ solution with a fixed [KCl] = 0.02M, the mode reduction method leads to a dramatic reduction in the computational time from about 43 hours (average time for the different Mg^{2+} concentrations) for the previous TBI model to 48 minutes for the current new model. For larger RNAs in higher Mg^{2+} concentration, we employ the aforementioned 24-point fitting method for the free energy profile $\Delta G(N_b)$. For example, in Fig. 4c, we predict the thermodynamics of the 76-nt yeast $tRNA^{Phe}$ (PDB code: 1TRA;[51]) by using of the mode-reduction method in combination with the 24-point fitting method. The average time for the different Mg^{2+} concentration is about 2 hours, which is about 50 times shorter than the 93 hours required by the previous TBI model. In Fig. 4d, we calculate the thermodynamic properties for an 80-nt RNA duplex using the current energy landscape-guided sampling method. The average time for the different Mg^{2+} concentration is about 3 hours, which is about 52 times shorter than the 156 hours required by the previous TBI model based on the enumeration of the ion binding modes.

We note that the theoretical predictions (Fig. 4a) are not perfect. One of the possible sources of the error may come from the monovalent ions as explained below. The charge density of some discrete local regions of a a compact tertiary RNA structure may become so high that even monovalent ions in these regions can become correlated. In the current form of the theory, all the monovalent ions are assumed to be (weakly correlated) diffusive ions. This approximation may require further examination for compact RNA tertiary structures.

4. Conclusion

By accounting for ion correlation and fluctuation effects, the tightly bound ion (TBI) model can provide reliable predictions for the ion effects in a variety of nucleic acid systems. However, applications of the model to large nucleic structures have been limited by the low computational efficiency of the model. In the present study, we develop a new method that shows remarkable improvement in the computational efficiency for the TBI model. The new method is based on the free energy landscape: a single-valley, smooth energy landscape as a function of the the number of the tightly bound ions. The steep single valley in the landscape allows us to reduce the complete ensemble of the ion distribution modes down to a set of low-lying modes around the energy minimum. In addition, the smooth landscape allows us to fit the landscape through random sampling of a finite number of ion distribution modes. Tests for a series of RNA systems indicate that the new method can indeed lead to remarkable increase in the computational efficiency by a factor of 50. The enhancement in the computational efficiency would be more drastic for larger structures. This new method may open up a new door for the TBI-based predictions for ion effects in biologically significant systems such as the Mg^{2+}-induced riboswitch,[56] tetraloop-receptor structure,[57] and the different types of ribozymes.[58]

Future development of the model should address several key issues. Besides the problem of the possible correlation effects for monovalent ions in RNA tertiary structures, the accuracy of the model is also limited by the inability to treat site-specific ion effect such as the sequence-specific binding of the structurally bound and partially dehydratedions. For many nucleic acid structures and sequences, divalent ion binding can be affected by non-

electrostatic effects and site-specific ion binding can be critical[31–33] for nucleic acid stability. For such systems, MD simulations would be particularly useful.[31–33]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NSF grants MCB0920067 and MCB0920411 and NIH grant GM063732. Most of the numerical calculations involved in this research were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).

References

1. Brion P, Westhof E. Hierarchy and dynamics of RNA folding. Annu. Rev. Biophys. Biomol. Struct. 1997; 26:113–137. [PubMed: 9241415]
2. Tinoco I Jr, Bustamante C. How RNA folds. J. Mol. Biol. 1999; 293:271–281. [PubMed: 10550208]
3. Sosnick TR, Pan T. RNA folding: models and perspectives. Curr. Opin. Struct. Biol. 2003; 13:309–316. [PubMed: 12831881]
4. Woodson SA. Metal ions and RNA folding: a highly charged topic with a dynamic future. Curr. Opin. Chem. Biol. 2005; 9:104–109. [PubMed: 15811793]
5. Chu VB, Herschlag D. Unwinding RNAs secrets: advances in the biology, physics, and modeling of complex RNAs. Curr. Opin. Struct. Biol. 2008; 18:305–314. [PubMed: 18555681]
6. Draper DE. RNA folding: thermodynamic and molecular descriptions of the roles of ions. Biophys. J. 2008; 95:5489–5495. [PubMed: 18835912]
7. Chen SJ. RNA folding: conformational statistics, folding kinetics, and ion electrostatics. Annu. Rev. Biophys. 2008; 37:197–214. [PubMed: 18573079]
8. Rook MS, Treiber DK, Williamson JR. An optimal Mg2+ concentration for kinetic folding of the Tetrahymena ribozyme. Proc. Natl. Acad. Sci. USA. 1999; 96:12471–12476. [PubMed: 10535946]
9. Takamoto K, He Q, Brenowitz M. Monovalent cations mediate formation of native tertiary structure of the Tetrahymena thermophila ribozyme. Nat. Struct. Biol. 2002; 9:928–933. [PubMed: 12434149]
10. Thirumalai D, Hyeon C. RNA and Protein Folding: Common Themes and Variations. Biochemistry. 2005; 44:49574970.
11. Koculi E, Hyeon C, Woodson SA. Charge density of divalent metal cations determines RNA stability. J. Am. Chem. Soc. 2007; 129:2676–2682. [PubMed: 17295487]
12. Soto AM, Misra V, Draper DE. Tertiary structure of an RNA pseudoknot is stabilized by diffuse Mg^{2+} ions. Biochemistry. 2007; 46:2973–2983. [PubMed: 17315982]
13. Stellwagen E, Dong Q, Stellwagen NC. Quantitative analysis of monovalent counterion binding to random-sequence, double-stranded DNA using the replacement ion method. Biochemistry. 2007; 46:2050–2058. [PubMed: 17253778]
14. Qiu X, Andresen K, Pollack L. Abrupt transition from a free, repulsive to a condensed, attractive DNA phase, induced by multivalent polyamine cations. Phys. Rev. Lett. 2008; 101:228101. [PubMed: 19113524]
15. Schlatterer JC, Kwok LW, Pollack L. Hinge stiffness is a barrier to RNA folding. J. Mol. Biol. 2008; 379:859–870. [PubMed: 18471829]
16. Lipfert J, Sim AY, Doniach S. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. RNA. 2010; 16:708–719. [PubMed: 20194520]
17. Chen AA, Pappu RV. Quantitative characterization of ion pairing and cluster formation in strong 1:1 electrolytes. J. Phys. Chem. B. 2007; 111:64696478.
18. Chen AA, Draper DE, Pappu RV. Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. J. Mol. Biol. 2009; 390:805819.

19. Serra MJ, Turner DH. Predicting thermodynamic properties of RNA. Methods Enzymol. 1995; 259:242–261. [PubMed: 8538457]
20. SantaLucia, John, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA. 1998; 95:1460–1465. [PubMed: 9465037]
21. Mathews DH, Sabina J, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 1999; 288:911–940. [PubMed: 10329189]
22. Chen SJ, Dill KA. RNA folding energy landscapes. Proc. Natl. Acad. Sci. USA. 2000; 97:646–651. [PubMed: 10639133]
23. Zuker M. MFold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003; 31:3406–3415. [PubMed: 12824337]
24. Debye P, Huckel E. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung undverwandte Erscheinungen. Physikalische Zeitschrift. 1923; 24:185–206.
25. Misra V, Draper DE. The interpretation of Mg^{2+} binding isotherms for nucleic acids using Poisson-Boltzmann theory. J. Mol. Biol. 1999; 17:1135–1147. [PubMed: 10600372]
26. Baker NA. Improving implicit solvent simulations: A Poisson-centric view. Curr. Opin. Struct. Biol. 2005; 15:137–143. [PubMed: 15837170]
27. Baker NA, Sept D, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA. 2001; 98:10037–10041. [PubMed: 11517324]
28. Zhou H-X. Macromolecular electrostatic energy within the nonlinear Poisson-Boltzmann equation. J. Chem. Phys. 1994; 100:3152–3162.
29. Tjong H, Zhou H-X. The dependence of electrostatic solvation energy on dielectric constants in Poisson-Boltzmann calculations. J. Chem. Phys. 2006; 125:206101. [PubMed: 17144745]
30. Tjong H, Zhou H-X. GBr6NL: a generalized Born method for accurately reproducing solvation energy of the nonlinear Poisson-Boltzmann equation. J. Chem. Phys. 2007; 126:195102. [PubMed: 17523838]
31. Auffinger P, Bielecki L, Westhof E. The Mg^{2+} Binding Sites of the 5S rRNA Loop E Motif as Investigated by Molecular Dynamics Simulations. Chem. Biol. 2003; 10:551–561. [PubMed: 12837388]
32. Krasovska MV, Sefcikova J, Sponer J. Cations and Hydration in Catalytic RNA: Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. Biophys. J. 2006; 91:626–638. [PubMed: 16617077]
33. Dong F, Olsen B, Baker NA. Computational Methods for Biomolecular Electrostatics. Methods in Cell Biology. 2008; 84:843–870. [PubMed: 17964951]
34. Joung I, Cheatham TE. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model specific ion parameters. J. Phys. Chem. B. 2009; 113:13279–13290. [PubMed: 19757835]
35. Kuczera K, Jas G, Elber R. The kinetics of helix unfolding: Molecular dynamics simulations with Milestoning. J. Phys. Chem. A. 2009; 113:7431–7473. [PubMed: 19552476]
36. Chen A, Marucho M, Baker NA, Pappu R. Simulations of RNA Interactions with Monovalent Ions. Methods in Enzymology. 2009; 469:411–432. [PubMed: 20946801]
37. Sklenovsky P, Florova P, Banas P, Reblova K, Lankas F, Otyepka M, Sponer J. Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. J. Chem. Theory Comput. 2011; 7:2963.
38. Do TN, Ippoliti E, Parrinello M. Counterion Redistribution upon Binding of a Tat-Protein Mimic to HIV-1 TAR RNA. J. Chem. Theory Comput. 2012 Article ASAP.
39. Kirmizialtin S, Pabit SA, Meisburger SP, Pollack L, Elber R. RNA and Its Ionic Cloud: Solution Scattering Experiments and Atomically Detailed Simulations. Biophys. J. 2012; 102:819–828. [PubMed: 22385853]
40. Bai Y, Greenfeld M, Herschlag D. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. J. Am. Chem. Soc. 2007; 129:14981–14988. [PubMed: 17990882]

41. Grochowski P, Trylska J. Continuum molecular electrostatics, salt effects, and counterion binding–a review of the Poisson-Boltzmann theory and its modifications. Biopolymers. 2008; 89:93113.
42. Wang K, Yu YX, Gao GH. Density functional study on the structural and thermodynamic properties of aqueous DNA-electrolyte solution in the framework of cell model. J. Chem. Phys. 2008; 128:185101. [PubMed: 18532848]
43. Tan ZJ, Chen SJ. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. J. Chem. Phys. 2005; 122:44903. [PubMed: 15740294]
44. Tan ZJ, Chen SJ. Predicting ion binding properties for RNA tertiary structures. Biophys. J. 2006; 91:518–536. [PubMed: 16648172]
45. Tan ZJ, Chen SJ. Electrostatic free energy landscape for nucleic acid helix assembly. Nucleic Acids Res. 2006; 34:6629–6639. [PubMed: 17145719]
46. Tan ZJ, Chen SJ. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. Biophys. J. 2007; 92:3615–3632. [PubMed: 17325014]
47. Tan ZJ, Chen SJ. Salt dependence of nucleic acid hairpin stabilit. Biophys. J. 2008; 95:738–752. [PubMed: 18424500]
48. Tan ZJ, Chen SJ. Predicting ion binding properties for RNA tertiary structures. Biophys. J. 2010; 99:1–12. [PubMed: 20655826]
49. Su L, Chen L, Rich A. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. Nat. Struct. Biol. 1999; 6:285–292. [PubMed: 10074948]
50. Conn GL, Gittis AG, Draper DE. A compact RNA tertiary structure contains a buried backbone-K^+ complex. J. Mol. Biol. 2002; 318:963–973. [PubMed: 12054794]
51. Westhof E, Sundaralingam M. Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA. Temperature factors and dynamics, coordinated waters, and base-pair propeller twist angles. Biochemistry. 1986; 25:4868–4878. [PubMed: 3533142]
52. Davis JH, Foster TR, Tonelli M, Butcher SM. Transposition of two amino acids changes a promiscuous RNA binding protein into a sequence-specific RNA binding protein. RNA. 2007; 13:1–11. [PubMed: 17123956]
53. Grilley D, Misra V, Draper DE. Importance of partially unfolded conformations for Mg^{2+}-induced folding of RNA tertiary structure: structural models and free energies of Mg2+ interactions. Biochemistry. 2007; 46:10266–10278. [PubMed: 17705557]
54. Romer R, Hach R. tRNA conformation and magnesium binding. A study of a yeast phenylalanine-specific tRNA by a fluorescent indicator and differential melting curves. Eur. J. Biochem. 1975; 55:271–284. [PubMed: 1100382]
55. Krakauer H. The binding of Mg2+ ions to polyadenylate, polyuridylate, and their complexes. Biopolymers. 1971; 10:2459–2490. [PubMed: 5126519]
56. Cromie M, Shi Y, Croisman EA. An RNA sensor for intracellular Mg^{2+}. Cell. 2006; 125:71–84. [PubMed: 16615891]
57. Hodak JH, Fiore J, Pardi A. Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. Proc. Natl. Acad. Sci. 2005; 102:10505–10510. [PubMed: 16024731]
58. Martha S,R, Daniel K,T, James R,W. An optimal Mg^{2+} concentration for kinetic folding of the Tetrahymena ribozyme. Proc. Natl. Acad. Sci. 1999; 96:12471–12476. [PubMed: 10535946]

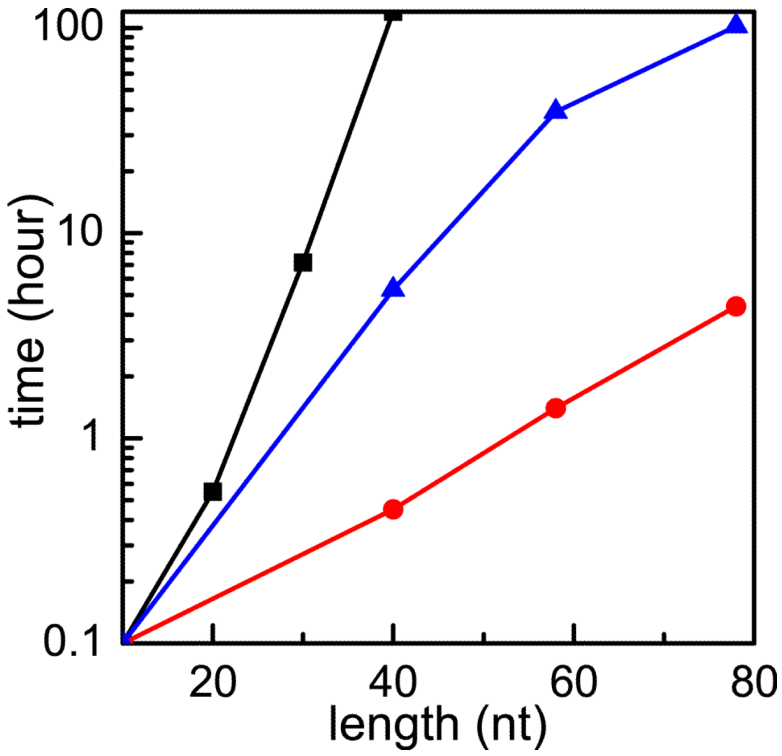

Figure 1.
The computer time as a function of the sequence length for an RNA duplex in 0.02M NaCl and 0.005M $MgCl_2$ mixed solution. The original TBI: square; the previously improved TBI: triangle; the current new TBI model: circle.

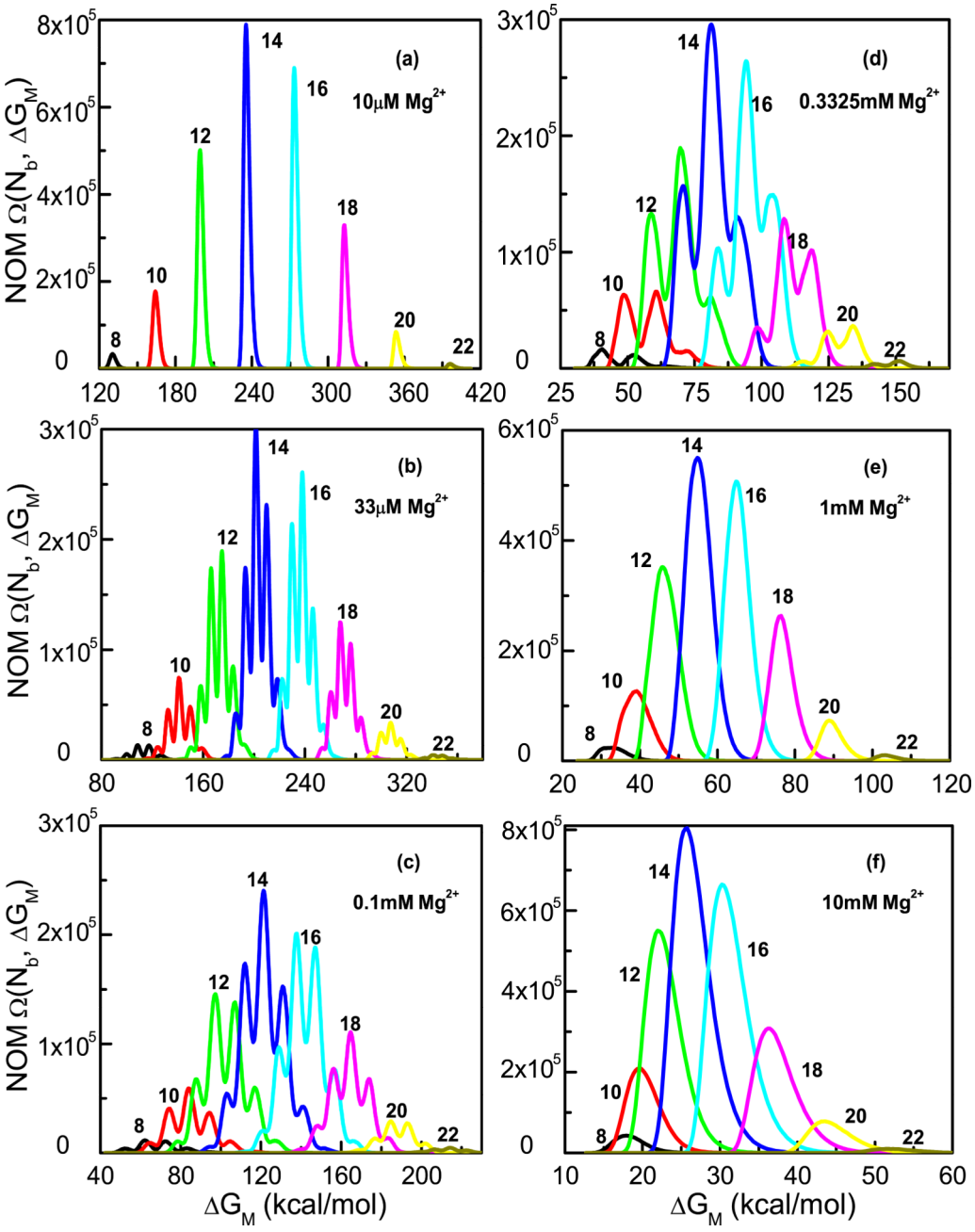

Figure 2.

The distribution of the number of modes (NOM) $\Omega(N_b, \Delta G_M)$ for the BWYV pseudoknot in the different [Mg^{2+}] ion solutions with a fixed 0.054 [Na^+] background. The numbers (8, 10 · · ·) in the figure denote the N_b values. Not all the N_b results are shown in the figure.

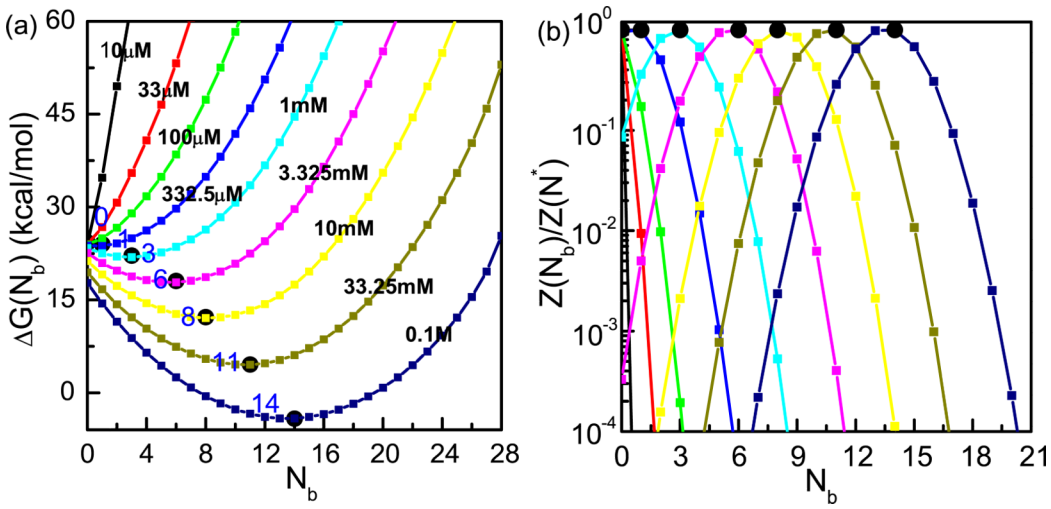

Figure 3.

(a) The reduced (1D) electrostatic energy landscape $\Delta G(N_b)$ for the BWYV pseudoknot in the different ion solutions with the fixed $[Na^+]$=0.054M and varying $[Mg^{2+}]$ (in M) (labeled on the curves). (b)The partition functions as a function of N_b for the different solutions. The filled squares and the labels denote the locations of the free energy minima and the corresponding N_b^*.

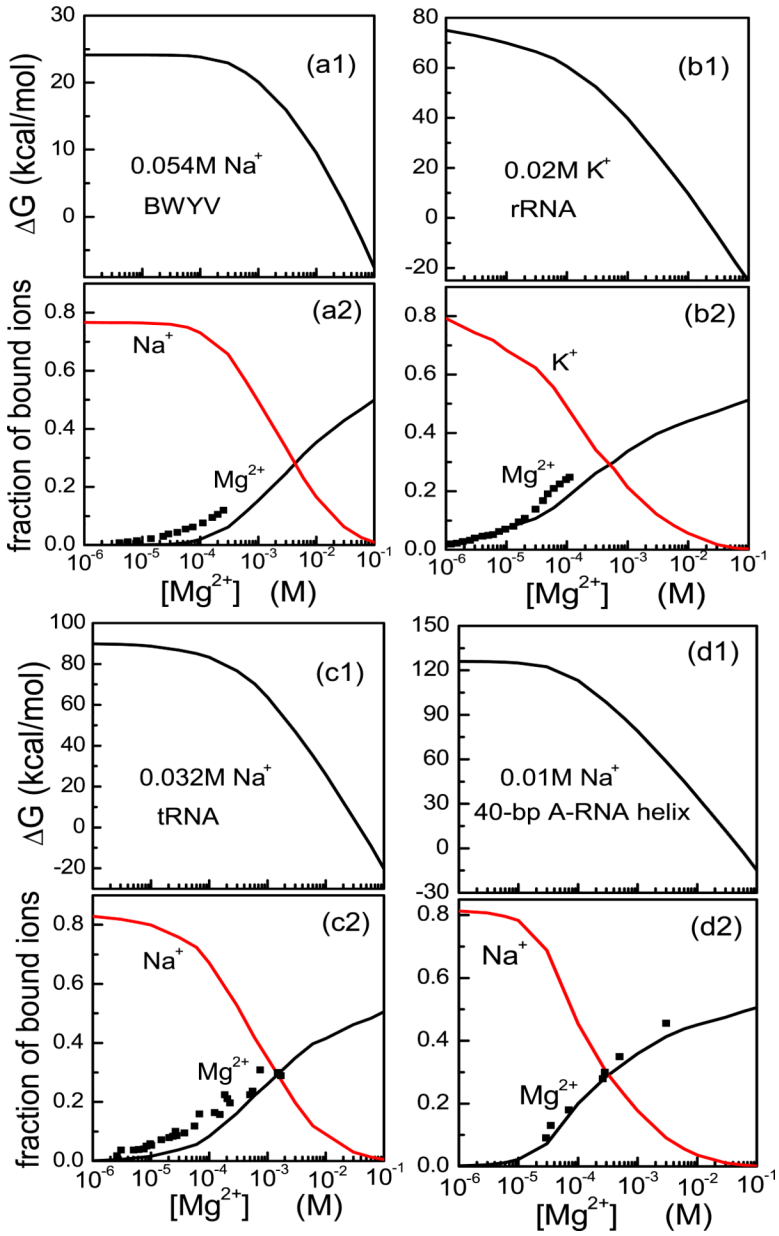

Figure 4.

The electrostatic free energy ΔG (upper panels) and the Mg^{2+} and Na^{+} (or K^{+}) binding fractions per nucleotide (lower panel) for various RNA tertiary structures: (a) BWYV pseudoknot fragment, (b) 58-nt rRNA fragment, (c) Yeast $tRNA^{Phe}$. (d) 40-bp RNA duplex. The corresponding experimental data are from Refs.[13,53–55] respectively. In the lower panels, the black symbols are the experimental data and solid lines are the predicted results.