Approach to the unfolding and folding dynamics of *add* A-riboswitch upon adenine dissociation using a coarse-grained elastic network model

Chunhua Li,[1,2] Dashuai Lv,[1] Lei Zhang,[1] Feng Yang,[1] Cunxin Wang,[1] Jiguo Su,[3,a)] and Yang Zhang[2,a)]

[1]*College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China*
[2]*Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 45108, USA*
[3]*College of Science, Yanshan University, Qinhuangdao 066004, China*

(Received 13 April 2016; accepted 17 June 2016; published online 6 July 2016)

Riboswitches are noncoding mRNA segments that can regulate the gene expression via altering their structures in response to specific metabolite binding. We proposed a coarse-grained Gaussian network model (GNM) to examine the unfolding and folding dynamics of adenosine deaminase (*add*) A-riboswitch upon the adenine dissociation, in which the RNA is modeled by a nucleotide chain with interaction networks formed by connecting adjoining atomic contacts. It was shown that the adenine binding is critical to the folding of the *add* A-riboswitch while the removal of the ligand can result in drastic increase of the thermodynamic fluctuations especially in the junction regions between helix domains. Under the assumption that the native contacts with the highest thermodynamic fluctuations break first, the iterative GNM simulations showed that the unfolding process of the adenine-free *add* A-riboswitch starts with the denature of the terminal helix stem, followed by the loops and junctions involving ligand binding pocket, and then the central helix domains. Despite the simplified coarse-grained modeling, the unfolding dynamics and pathways are shown in close agreement with the results from atomic-level MD simulations and the NMR and single-molecule force spectroscopy experiments. Overall, the study demonstrates a new avenue to investigate the binding and folding dynamics of *add* A-riboswitch molecule which can be readily extended for other RNA molecules. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4954992]

I. INTRODUCTION

In recent years, the studies of functional noncoding RNAs have received unprecedented attention.[1–3] Riboswitches are the natural genetic control segments found in the 5′ untranslated regions (UTRs) of many mRNAs of bacteria and fungi. They can regulate gene expression at the level of transcription or translation that is highly sensitive to the cellular concentrations of specific metabolites.[2–4] So far, various riboswitches have been discovered which can specifically recognize a large variety of ligands, such as adenine, glycine, lysine, and glucosamine-6-phosphate.[5–8] Riboswitches are generally composed of two domains: the aptamer domain and the downstream expression platform. The former can fold into an intricate three-dimensional structure upon binding to a specific metabolite when the concentration of the metabolite exceeds a threshold. The latter switches a gene on/off through its structural changes triggered by metabolite-induced large-scale conformational changes in the aptamer domain.[8] The adenosine deaminase (*add*) A-riboswitch is one of the structurally simplest riboswitches and has been widely studied.[9–12] The folded aptamer domain of the *add* A-riboswitch bound with the ligand adenine was structurally characterized by X-ray crystallography with Protein Data Bank (PDB) code 1Y26 (Fig. 1).[13] It forms a tuning-fork-like architecture assembled by three helical stems, two of which are hairpins (P2 and P3) interacting via kissing loops (L2 and L3) and aligned on top of the third helix P1. The ligand-binding pocket is formed through stacked bases interconnecting the junctions J12, J13, and J23, with two base pairs (U75-A21 and A76-U20) at the end of P1 stem. The adenine ligand forms a Watson-Crick (WC) type of interaction with U74. In the crystal structure, there are five Mg^{2+} ions binding to the *add* A-riboswitch, one of which is on the surface, and the other four are positioned deep within the grooves primarily formed by the junction-connecting segments. The binding of adenine ligand to the aptamer domain results in the change of its folding pattern, which alters the conformation of the expression platform and activates the gene expression.[14] Thus, the functional performance of the *add* A-riboswitch depends on the conformational transformation of the aptamer between the folded and unfolded states. Therefore, characterizing and elucidating its folding/unfolding process is important for the understanding of its regulatory mechanism.

In the past few years, several experimental and theoretical investigations have been conducted on the folding/unfolding process of the *add* A-riboswitch system. Rieder *et al.*[15] reported that the folding process of the aptamer is induced by adenine ligands and metal ions. Neupane *et al.*[11] have explored its unfolding pathway using single molecule force

a)Authors to whom correspondence should be addressed. Electronic addresses: jiguosu@ysu.edu.cn and zhng@umich.edu

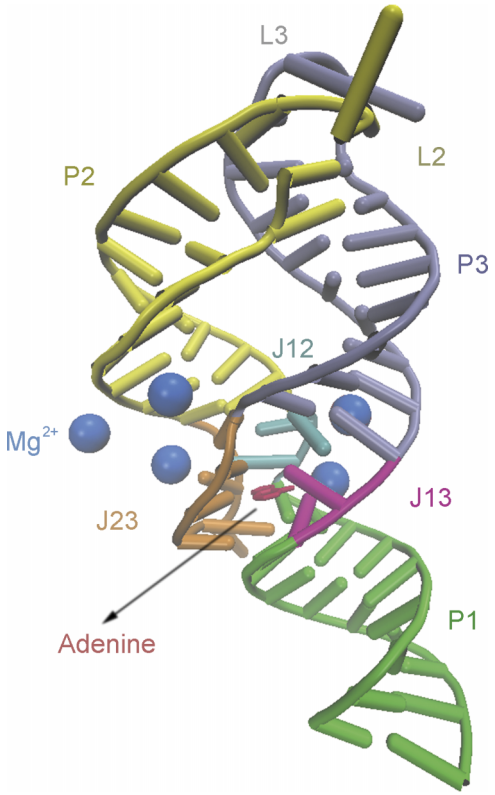

FIG. 1. Three-dimensional structure of *add* A-riboswitch with five Mg^{2+} ions and a ligand adenine (PDB code 1Y26). The ligand and Mg^{2+} ions are colored with red and blue, respectively. The secondary structures of the system are labelled correspondingly with P1, P2, and P3 for stems, L2 and L3 for loops, and J12, J13, and J23 for junctions.

measurements and observed a series of intermediates during its unfolding process. Using a self-organized polymer model with the Langevin dynamics, Lin and Thirumalai[14] described force-triggered unfolding of this system and obtained the same unfolding sequence as the experimental observation. Priyakumar and MacKerell[16] performed a 40 ns all-atom molecular dynamics (MD) simulation to investigate the unfolding behavior in both presence and absence of ligand, but the complete unfolding process was not produced and only structural deviations were observed in the J23 junction and P1 stem. In addition, some investigations have concentrated on the roles of adenine ligand on the structural stability of *add* A-riboswitch. Researchers found that the ligand binding plays a critical role in stabilizing the binding pocket, especially for the folded P1 stem.[11,16,17] While the all-atom MD simulation can provide useful information on the fluctuations and conformational changes of biomacromolecules, it is generally difficult to obtain a complete folding/unfolding picture through MD due to the prohibitive computational costs.

The coarse-grained method has proved to be an efficient tool to explore the large-scale functional motions of biomolecular systems. Recently, we proposed a Gaussian network model (GNM) method, which was successfully applied to protein folding/unfolding studies.[18,19] In this method, the unfolding process of proteins is modeled through gradually breaking the non-covalent native contacts according to the distance fluctuations of residue pairs. This process can be considered to be composed of a series of quasi-equilibrium states corresponding to a slowly increasing temperature, during which the molecule has sufficient time to search for the state with a maximum entropy increase at each unfolding step.

In this work, we extended the iterative GNM approach to explore the unfolding process of the *add* A-riboswitch, with the changes of cross-correlations between nucleotide fluctuations and the fluctuations in the fast modes analyzed in detail. Additionally, we tried to interpret the roles of the adenine ligand in stabilizing the folded aptamer from a topological point of view.

II. METHODS

A. Gaussian network model

In the Gaussian network model, the three-dimensional structure of a molecule is described as an elastic network of nodes that are connected by harmonic springs with a certain cutoff distance.[20,21] In this work, the RNA chain is represented by a set of linked nodes, with each node standing for a nucleotide, where the interactions of backbone-backbone, base-base, and backbone-base, and the effects of Mg^{2+} ions and adenine are considered in the construction of the nucleotide networks. In addition to the backbone links, the nucleotide nodes are connected in the model when one of the following requirements is satisfied:

I. The distance between P atoms ≤8.5 Å.
II. The distance between mark atoms ≤8.0 Å.
III. The distance between the P and mark atoms from different nucleotides ≤8.0 Å.
IV. Both distances of two P atoms from the same Mg^{2+} ion ≤8.5 Å.
V. Both distances of two mark atoms from the mark atom of the ligand adenine ≤8.0 Å.

The mark atoms refer to the four atoms (N6 for adenine, O6 for guanine, N4 for cytosine and O4 for uracil), representing the four kinds of bases, respectively, all of which participate in the formation of hydrogen bonds in canonical base pairs.[22] The requirements I, II, and III are designed to take into account the interactions of backbone-backbone, base-base, and backbone-base, respectively, and V for considering the mediating role of the ligand.

It should be pointed out that as Mg^{2+} ions are important for structure stabilization and folding of the system, all the five Mg^{2+} ions are considered in the network construction. Mg^{2+} ions mainly play a role in mediating the interactions among the negative charges of the highly acidic phosphate backbone, permitting the system to form and retain the folded structures. Thus, in the network construction we consider the mediating roles of Mg^{2+} ions (not treated as separated nodes) by connecting the nucleotide nodes when both distances of their P atoms from the same Mg^{2+} ion ≤8.5 Å, as shown in requirement IV.

Different from the conventional GNM in which the force constants are identical for the springs of covalent and noncovalent pairs, the spring constants of non-bonded nucleotides are set as γ while those of the covalently bonded

pairs along the chain backbone are chosen as $c\gamma$. The values of c and γ are determined by fitting the predicted fluctuations against the crystallographic B-factors. This setting allows us to consider the heterogeneity of RNA chain interactions, since the covalent bonding is generally much stronger than the non-covalent ones.

The internal Hamiltonian of the RNA molecule system can be written as[23]

$$H = \frac{1}{2}\gamma\left[\Delta R^{\mathrm{T}}(\Gamma \otimes E)\Delta R\right], \tag{1}$$

where $\Delta R = \{\Delta R_1, \Delta R_2, \ldots \Delta R_N\}$ represents the $3N$-dimensional column vectors of fluctuations of the nodes in the network, where N is the number of nucleotides; the superscript T denotes the transpose; E is the third-order identity matrix; $\otimes$ is the direct product; and Γ is the $N \times N$ symmetric matrix in which the elements are given by

$$\Gamma_{ij} = \begin{cases} -c & if\ |i-j| = 1 \\ -1 & if\ |i-j| > 1 \text{ and nucleotides } i \text{ and } j \text{ are connected} \\ 0 & if\ |i-j| > 1 \text{ and nucleotides } i \text{ and } j \text{ are not connected} \cdot \\ -\sum_{i,j\neq i}\Gamma_{ij} & if\ i = j \end{cases} \tag{2}$$

The mean-square fluctuation (MSF) of each nucleotide node and the cross-correlation fluctuations between different nucleotide nodes are in proportion to the diagonal and off-diagonal elements of the pseudo-inverse of the Γ matrix. The inverse of the matrix can be decomposed as

$$\Gamma^{-1} = U\Lambda^{-1}U^{\mathrm{T}}, \tag{3}$$

where U is an orthogonal matrix whose columns u_i ($1 < i \leq N$) are the eigenvectors of Γ, and Λ is a diagonal matrix of eigenvalues λ_i of Γ. The cross-correlation fluctuations between the ith and jth nucleotides are written as

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3k_BT}{\gamma}\left[\Gamma^{-1}\right]_{ij}, \tag{4}$$

where k_B is the Boltzmann constant and T is the absolute temperature. When $i = j$, the MSF of the ith nucleotide can be obtained. The B-factor, also called temperature factor, which is related to the MSF, can be computed with the expression

$$B_i = \frac{8\pi^2}{3}\langle \Delta R_i \cdot \Delta R_i \rangle = \frac{8\pi^2 k_BT}{\gamma}\left[\Gamma^{-1}\right]_{ii}. \tag{5}$$

The MSF of the ith nucleotide associating with the kth mode is obtained by

$$\langle \Delta R_i \cdot \Delta R_i \rangle_k = \frac{3k_BT}{\gamma}\lambda_k^{-1}[u_k]_i[u_k]_i. \tag{6}$$

The MSF in the distance vector R_{ij} between the nucleotides i and j can be calculated with the expression[24]

$$\begin{aligned} \left\langle (\Delta R_{ij})^2 \right\rangle &= \left\langle \left(R_{ij} - R_{ij}^0\right)^2 \right\rangle = \left\langle (\Delta R_i - \Delta R_j)^2 \right\rangle \\ &= \langle \Delta R_i \cdot \Delta R_i \rangle + \langle \Delta R_j \cdot \Delta R_j \rangle - 2\langle \Delta R_i \cdot \Delta R_j \rangle \\ &= \frac{3k_BT}{\gamma}\left(\left[\Gamma^{-1}\right]_{ii} + \left[\Gamma^{-1}\right]_{jj} - 2\left[\Gamma^{-1}\right]_{ij}\right), \end{aligned} \tag{7}$$

where R_{ij} and R_{ij}^0 are the instantaneous and equilibrium separation vectors between nucleotides i and j. In the GNM, the cross-correlation is normalized as

$$C_{ij} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{\left[\left\langle \Delta R_i^2 \right\rangle \cdot \left\langle \Delta R_j^2 \right\rangle\right]^{1/2}}. \tag{8}$$

B. Iterative Gaussian network model

The unfolding process of the riboswitch molecules is modeled based on the thermodynamic fluctuations. Generally, with a higher fluctuation the native contacts of riboswitches between nucleotides are expected to break in a higher possibility, which results in unfolded structures. The fluctuations in the distance between all nucleotides are calculated by the GNM. The nonlinear elasticity during RNA unfolding is considered via iterative normal mode calculations, i.e.,[18]

I. The MSFs of the distances in all nucleotide pairs are calculated according to the native structure topology with Eq. (7).
II. The contact in the nucleotide pair with the largest distance fluctuation is broken. Then, a new matrix Γ is constructed, which represents a new topology during RNA unfolding.
III. The MSFs of the distances in all nucleotide pairs are recalculated based on the new matrix Γ with Eq. (7).
IV. The above steps II and III are repeated until all the noncovalent contacts are broken.
V. All the topologies of different conformations during RNA unfolding are obtained and the unfolding pathway can be derived from the above data.

It should be noted that many previous studies have been conducted on examining the roles of Mg^{2+} ions on the *add* A-riboswitch, indicating that they mainly affect the pre-organization and stabilization of the folded conformation rather than the intervention of the unfolding process.[12,15] Therefore, we did not specifically examine the unfolding behavior induced by Mg^{2+} release. Here, the contacts mediated

by Mg^{2+} are not treated specially and the unfolding process is simulated purely in a fluctuation-dependent manner.

III. RESULTS AND DISCUSSION

A. Force constants for the covalent and noncovalent interactions

The *add* A-riboswitch is modeled by the Gaussian network model system with each node representing a nucleotide and the interaction network formed by the adjoining atomic contacts (see Sec. II). The spring constant values of covalent and noncovalent interactions, i.e., $c\gamma$ and γ, are determined by matching the calculated and experimental B-factor values. As the absolute value of γ does not affect the relative size of nucleotide fluctuations, it has no influence on the correlation between the calculated and experimental B-factors. Therefore, we first set $\gamma = 1$ to determine the value of c via maximizing the correlation between the computed and experimental B-factors. The value of γ was then determined through normalizing the calculated fluctuation with the experimental B-factors. According to this method, we obtained $c = 8.0$ and $\gamma = 0.125\ k_BT/\text{Å}^2$, where the correlation coefficient between the theoretical and experimental B-factors is 0.728.

Compared to the traditional GNM with uniform spring contacts, the introduction of heterogeneity in the covalent and noncovalent spring contacts in our model improves the correlation coefficient of B-factors from 0.606 to 0.728. Part of the effect is due to the consideration of the mediating roles of Mg^{2+} ions in our network construction. Without considering the Mg^{2+} ions, the correlation coefficient of B-factors will reduce to 0.717, which partly highlights the roles of Mg^{2+} ions in the stabilization and folding of the molecule.

We also computed this correlation coefficient using the conventional RNA single-point GNM, in which one node represents one nucleotide and the connections between them are determined only based on the distances between P atoms;[25] this gives an even lower correlation coefficient 0.587. The results indicate that the new iterative GNM approach, which considers the RNA backbone-backbone, backbone-base, base-base interactions, and Mg^{2+} ion binding, and meanwhile distinguishes the covalent from noncovalent interactions, can help generate a better accordance with experimental fluctuation data.

B. Roles of ligand adenine and Mg^{2+} ions on the dynamics of *add* A-riboswitch

The ligand adenine was found to be important for the structure and function of *add* A-riboswitch.[26] To examine the specific role of the ligand adenine on the dynamics of the system, we computed the MSF of each nucleotide in the GNM with and without adenine. Fig. 2 shows that the nucleotides have a relatively higher MSF in the adenine-free state than in the adenine-binding state, although the difference in MSFs varies along the sequence. The most significant increase in MSFs occurs in the three junctions (J12, J13, and J23) associated with the binding pockets as well as the nucleotides

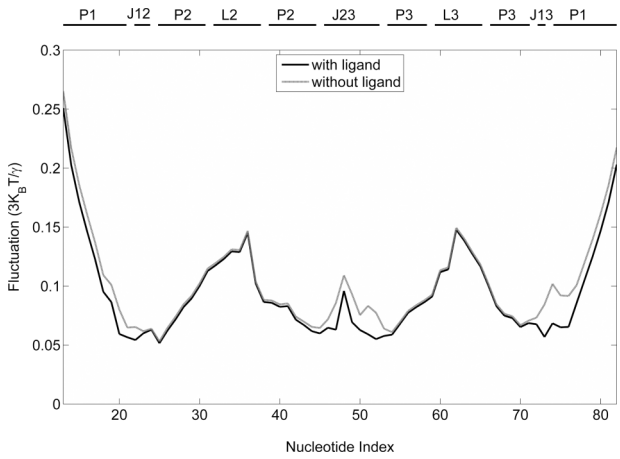

FIG. 2. Fluctuations of nucleotides of *add* A-riboswitch in ligand-bound state (solid line) and ligand-free one (shaded line), respectively.

at the terminal P1 stem that are near the binding pocket. In particular, the region from U74-A76 pairing with the adenine has the highest MSF increase. Similar MSF changes have been observed in the former atomic MD and experimental studies,[12,16,17,27] suggesting that the adenine ligand-binding indeed plays an important role in stabilizing the binding pocket, especially for the folded P1 stem and J23 junction regions. Functionally, binding with the adenine ligand leads to a change in the stability of P1 stem, which in effect, constitutes a structural switch in the expression platform, since the 3′ strand of P1 stem is involved in alternative base pairing with the Shine-Dalgarno sequence of the expression platform in the absence of adenine binding, which masks the Shine-Dalgarno sequence and causes the stop of translation.[8,28,29]

The ligand dissociation leads to a marked rise in MSF at J23 U51 (base pairing with ligand), which causes further instability of its neighbors C50 and U49 that are involved in the formation of base triplets with the top base pairs of P1 stem: U75-A21 and A76-U20, respectively.[12] This explains the causal relationship between the fluctuations of P1 stem and J23 junction, observed by Sharma *et al.*[12] Additionally, the data imply that the transmission pathway of structural changes caused by ligand dissociation may start from ligand-bound U51, and then to its neighbors, finally to the P1 stem. Also, the high flexibility of J23 in the absence of the adenine ligand suggests that the ligand should associate and dissociate from the riboswitch in the vicinity of J23.

Recently, the experimental and theoretical researches have reported that Mg^{2+} ions appear to be important in preorganizing and stabilizing the aptamer fold.[12,15] In order to detect the roles of Mg^{2+} ions on the dynamics of the system, we calculated the MSF of each nucleotide in the GNM of the system with and without Mg^{2+} ions (see Figure S1 of the supplementary material[38]). Overall, the nucleotides have relatively higher MSFs in the Mg^{2+} ion free state than in the Mg^{2+} ion binding state, especially for the regions around the ligand-binding pocket: J12, J23, and the junction proximal ends of three stems P1 (U20-A21), P2 (G42-A45), and P3 (C69-U71), which was also observed in the previous MD simulation study.[12]

In conclusion, both ligand adenine and Mg^{2+} ions show their importance in stabilizing the folded conformation of the riboswitch. Ligand adenine mainly contributes to the stabilization of P1 and J23, and the effect of Mg^{2+} ions is mainly reflected in stabilizing the ligand binding pocket.

C. Pathway of *add* A-riboswitch unfolding

The GNM based calculations showed that the riboswitch molecule becomes unstable and tends to unfold in the absence of adenine. To examine the unfolding pathway of the riboswitch, we monitored the loss of the native contacts between nucleotides in the course of unfolding simulation based on the iterative GNM, which starts from the native riboswitch structure but removing the adenine ligand. Fig. 3 shows the snapshots of the native contact maps with each native contact marked by a dot in the maps, where Figs. 3(a)-3(e) run from the loss-number-of-native-contact (LNNC) equal to 0, 50, 100, 175, and 250, respectively. In Figure 4, we present 3D conformation for each of the snapshots, with the connections and nodes of the networks mapped on the molecule (see Fig. 4).

The first unfolding event is the loss of the native contacts within P1 stem, as shown in Figs. 3(b) and 4(b). Similar observations were also made in the NMR experiment[9,30] and MD simulation,[12,16] i.e., the main conformational change upon ligand dissociation is the unfolding of P1 helix. This result means that P1 is probably easier to unfold compared to the other parts of the riboswitch.

Next, the native tertiary contacts between L2 and L3 loops are lost (Figs. 3(c) and 4(c)), indicating that the two loops separate from each other. Therefore, the removal of the ligand not only destabilizes P1 directly but also causes a slight collapse of the kissing loops (L2-L3) as indicated by the MD studies.[9,12] Allner *et al.*,[9] further suggested that along with the L2-L3 opening process, a key event is that the stacking interactions between A24 and G72/A73 are lost and A24 becomes exposed to the solvent, losing its function as a hook between helices P2 and P3. This insight was also observed in our modeling in which the native contacts between junction J12 A23/A24 and J13 A73 disappear, causing the structural instability and exposure of the hydrophobic core. After the disruption of L2-L3 interactions, the native contacts within P2 stem, as well as several contacts between J12 and J23 are lost (Figs. 3(d) and 4(d)), which is followed by the unfolding of P3 stem (Figs. 3(e) and 4(e)). Finally, the *add* A-riboswitch is fully unwound and the hydrophobic binding pocket is completely exposed. There are several contacts among the three junctions (J12, J23, and J13) remaining till the unfolding event is almost over. These retained parts which unfold last may fold first in a refolding process due to their lower free-energy basins, which was also proposed by Allnér *et al.*[9] for *add* A-riboswitch with all-atom MD simulations. Overall, despite the simplified modeling, the unfolding pathway of the *add* A-riboswitch calculated here is in close agreement with the experimental data obtained by the single-molecule force spectroscopy.[11]

It should be mentioned that the detailed interactions among nucleotides have been replaced by constant springs in our coarse-grained GNM and the topological structure is therefore the determinant factor for the results. The close agreement between the modeling results and the experimental data implies that the topology plays an important role in the folding and unfolding process of *add* A-riboswitch.

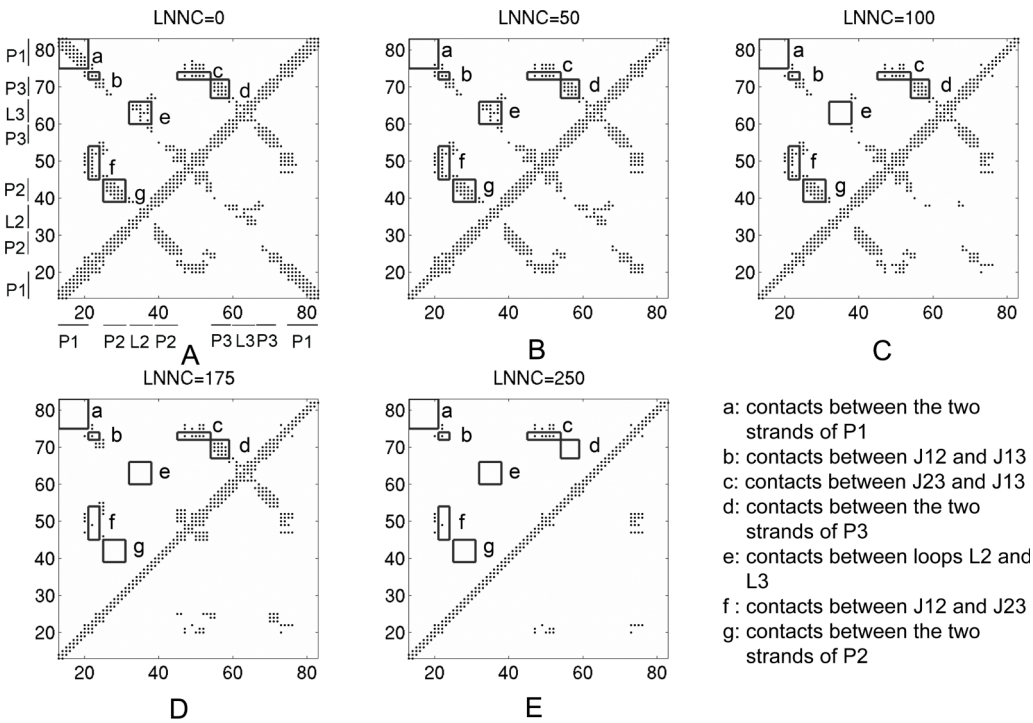

FIG. 3. The contact maps of the native conformation (a) and the conformations with the LNNC to be 50 (b), 100 (c), 175 (d), and 250 (e) for *add* A-riboswitch, respectively. Each native contact is marked by a dot in the maps. The rectangles labelled with letters (a)–(g) represent the contacts between different secondary structures that are indicated in the lower right corner.

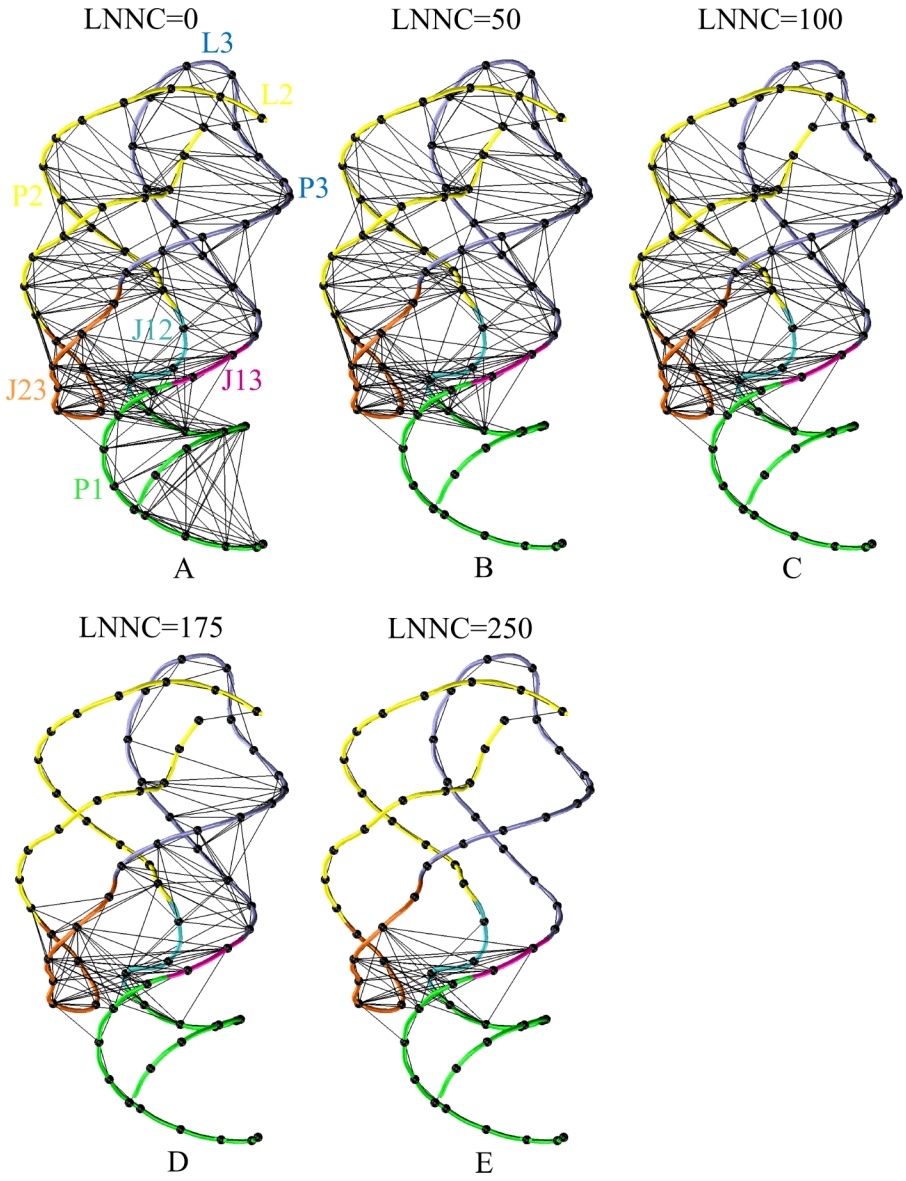

FIG. 4. The connections (black lines) and nodes (black balls) of the networks mapped on the three-dimensional structures with the LNNC to be 0 (a), 50 (b), 100 (c), 175 (d), and 250 (e) for *add* A-riboswitch, respectively.

Similarly, the topology-driven view of folding has been put forth for protein folding.[31] A variety of experimental and theoretical evidences has suggested that proteins, especially small fast-folding proteins, have selected the sequences with minimal energetic frustration under evolutionary pressure, leaving the topology as the main source of the frustration in protein folding/unfolding.[18,32–35] Yet an analogous definite statement has not been made for RNA molecules, most likely due to the strong coupling between the ionic environment and conformational energetics that make RNA folding more complex than protein folding. Recently, Sorin *et al.*[36] have found that the native tRNA topology serves as a dominant predictor of the bulk folding mechanism. The detailed comparison between protein and RNA hairpin structures indicated that the native topology defines the folding landscape for nucleic acids in a manner analogous to that suggested for proteins, with the specific interactions within the polymer determining the relative weighting between parallel pathways.[37] These insights are again consistent with our results here that the structural topology is a critical factor for determining the folding/unfolding dynamics of the *add* A-riboswitch.

D. Robustness of unfolding process against noise

In our iterative GNM approach, we identify and remove the nucleotide pairs of the highest MSF at each step. To examine the robustness of the unfolding process, we introduced a stochastic noise to the model by randomly selecting and removing a contact from those with the top-three highest MSFs. The unfolding process with such noise was simulated 500 times and the average result of the contact losing process is shown in Fig. 5(a). As a comparison, the unfolding process without noise is displayed in Fig. 5(b). It can be seen that the two figures are very similar except for a small difference in the unfolding order of P2 and P3 domains. In both simulations, the stem P1 unfolds first, followed by the separation between L2 and L3. But in the simulation without noise, P3 starts to unfold after P2 completes the unfolding, while in the model with noise the unfolding of

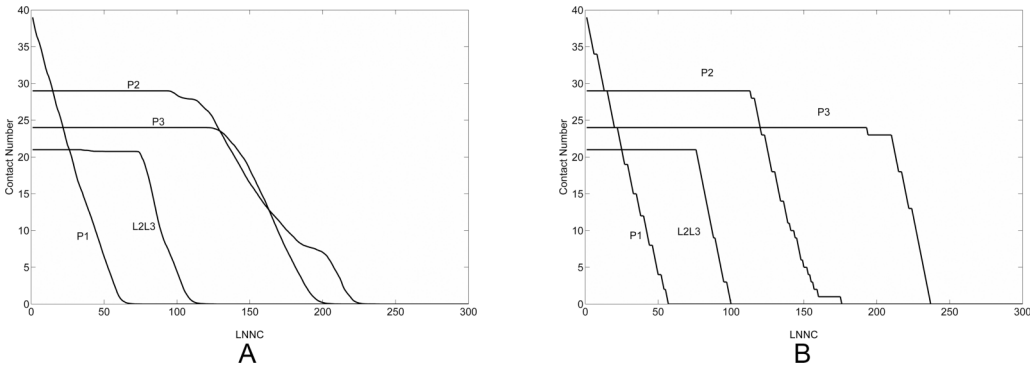

FIG. 5. The process of losing of native contacts during unfolding of *add* A-riboswitch. The unfolding process with some perturbation is simulated 500 times and the average result is shown in panel A. In contrast, the unfolding process without perturbation is represented in panel B.

P3 completes slightly prior to P2 although P2 breaks earlier than P3. Indeed, the P2 stem is not as stable as P3, but the free energy difference between them is merely 1 kcal/mol.[9,14] This small free energy difference can explain why introducing some randomness into the thermal denaturation can result in a light effect on the unfolding order of P2 and P3. But the overall unfolding pathway of *add* A-riboswitch is quite robust against external random disturbances (Fig. 5).

E. Change of MSFs in the fast modes during *add* A-riboswitch unfolding process

It has been well-established that the fast motion modes in protein system correspond to geometric irregularity in the local structure, and the residues acting in the fast modes are considered as kinetically hot residues that are crucially important for the stability of the tertiary fold.[18] In RNA, however, the problem whether nucleotides acting in fast modes play the same role remains unclear. To attack this problem, we select multiple conformations from the unfolding process and calculate the fluctuations of all nucleotides in these conformations using the iterative GNM approach (Eq. (6)). Fig. 6 gives the fluctuations of nucleotides in the eight fastest modes for the conformations during the unfolding of the system. It is shown that the peaks of the fluctuations are mainly situated in junctions J12 (U22-A24), J23 (U49-A50, A52-C53), and stems P1 (U20-A21, A76), P2 (U25, G42-A45), and P3 (A56, A66-C67, C69-U70) which are close to the junctions. Most of these nucleotides are considered to constitute the hydrophobic binding pocket.[11,12,15] Interestingly, the peaks of the fluctuations largely overlap for the conformations selected from different stages in the unfolding process, which indicates that these nucleotides retain most of their contacts until the end of the unfolding process. Thus, similar to protein system, these relative high fluctuation nucleotides acting in fast modes are critical to maintain the stability of the tertiary fold of *add* A-riboswitch.

F. Change of cross-correlation between the fluctuations of nucleotides during unfolding process

The correlation between fluctuations can reveal information related to the spatial motion and interaction. The pair-wise cross-correlations between nucleotides can be calculated using Eq. (8), with the values ranging from −1 to 1. The positive values represent the motions of nucleotides are in the same direction, and the negatives represent that they move in the opposite direction. The higher the absolute value is, the more the two nucleotides are correlated. The value $C_{ij} = 0$ means that the motions of nucleotides are completely uncorrelated.

Fig. 7 presents the change of the cross-correlation maps of the nucleotide fluctuations along the course of unfolding. Prior to the unfolding (Fig. 7(a)), the opposite strands in the helical stems (P1, P2, P3) form strong positive correlations, due to the stable folded structure and strong base pairing interactions. Similarly, the contacts between loops (L2 and L3) and junctions (J12, J13, and J23) also dictate the positive correlations in these regions, consistent with the structural feature of the binding pocket.[11,12,15]

After the removal of the ligand, the system becomes unstable and the structure starts to unfold. Accordingly, the strong positive correlations, which were formed in the helical stems, loops, and junctions, are gradually converted to weak and then strong negative, as shown from Figs. 7(b)-7(d) which corresponds to LNNC = 50, 100, and 175, respectively. The results imply that these structural segments move away from

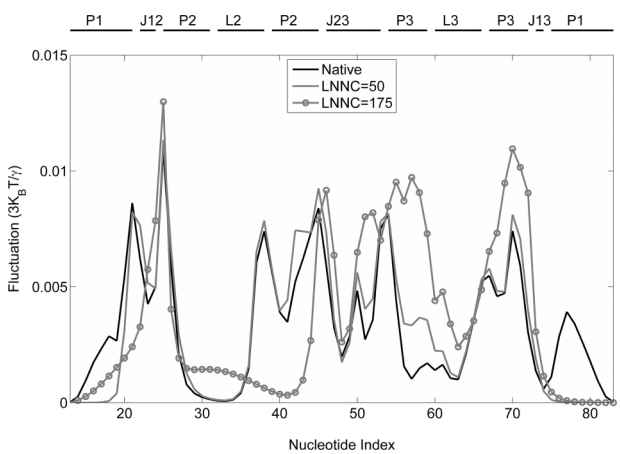

FIG. 6. The fluctuations of the nucleotides in the fast eight modes from GNM analysis for several conformations visited during the unfolding of *add* A-riboswitch. The solid, shaded, and shaded circled lines represent the native conformation, the conformations with LNNC = 50 and LNNC = 175, respectively.

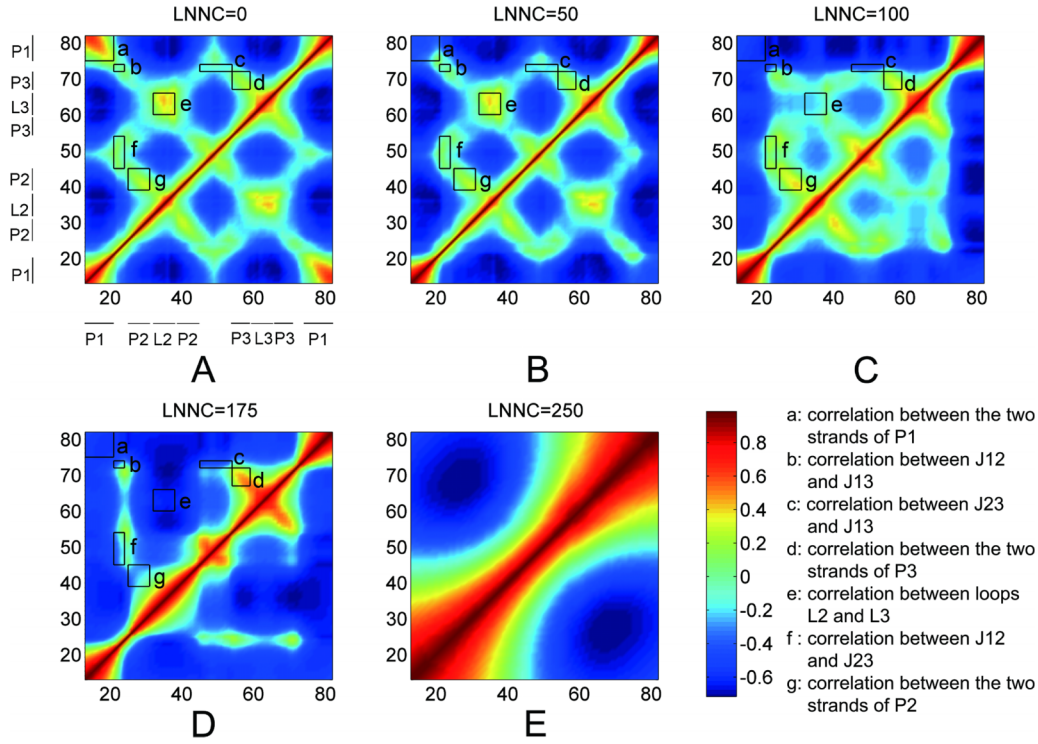

FIG. 7. The cross-correlation maps calculated using all modes for native conformation (a) and several conformations with LNNC to be 50 (b), 100 (c), 175 (d), and 250 (e) during the unfolding process of *add* A-riboswitch. As shown in the color bar, the blue regions in the figure indicate negative correlation and the green-yellow-red regions present positive correlation. The rectangles labelled with letters a–g represent the correlations between different secondary structures that are indicated in the lower right corner.

each other to expose the hydrophobic binding pocket during the unfolding process of the RNA molecule. Interestingly, when the contacts within P2 are disrupted, P2 and L2 form negative correlations with P3 and L3 as shown in Fig. 7(d), and the correlations among the three junctions remain weak, indicating that P2 and L2 move apart from P3 and L3 but some contacts still exist among junctions due to the low free-energy basin involved in the binding pocket region.

When the contacts among the junctions are lost, the structure of *add* A-riboswitch is finally divided into two parts that fluctuate in the opposite directions (Fig. 7(e)). The unfolded RNA molecule does not behave as a random coil, suggesting that long-range correlations still exist in the denatured state probably due to the cooperative motion. Maybe the highly cooperative motion can help and promote the *add* A-riboswitch refold into the native state speedily through reducing the population searched in the conformational space.

IV. CONCLUSIONS

We have proposed a general coarse-grained Gaussian network model to describe the structure and folding dynamics of the functional noncoding RNA molecules. The network construction and introduction of heterogeneous spring constants for covalent and noncovalent interactions resulted in significantly improved correlation of modeled nucleotide fluctuations compared to experimental data.

The model was extended to examine the folding dynamics of the adenosine deaminase (*add*) A-riboswitch. It was found that the ligand adenine and Mg^{2+} ions play a critical role in stabilizing the structural fold of the molecule and the removal of adenine can result in unfolding due to the increased thermodynamic fluctuations in the junction and stem areas. The iterative GNM simulations show that the unfolding process of the *add* A-riboswitch starts with the denature of terminal P1 helix stem, followed by the loop and junction regions that are relevant to hooking the central helix domains and the ligand binding pocket, and then the central helix domains. Although the iterative GNM is an overly simplified model involving only two types of node interactions, the unfolding dynamics and pathway are in close agreement with the results from expensive atomic-level MD simulations[12,16] and NMR and single-molecule force spectroscopy experiments.[9,11] This result suggests that the topology of the RNA structures is responsible for the folding and unfolding dynamics, an analog to the observation made in the topology-driven folding process of proteins.[31] Nevertheless, it should be mentioned that the spring constants in the GNM model are based on B-factors, which are not necessarily the best measure of the conformational flexibility. In particular, many experimental effects, including crystal packing and ligand binding, have not been taken into account, description of which should request for more detailed modeling.

Finally, we examined the impact of the stochastic noise on the unfolding pathway by allowing nucleotides with slightly lower thermodynamic fluctuations to unfold. It shows that the unfolding pathway is largely robust to the noise except for the switch of the unfolding order between two

central helix domains that involves similar free energy basins. The calculations on the cross-correlation between nucleotide fluctuations further confirmed the unfolding pathway revealed by fluctuation-based simulations. Additionally, the nucleotides near the hydrophobic pocket with higher fluctuations in fast modes remain a high number of contacts throughout the unfolding process and are very important for maintaining the stability of the folded state. Overall, this study demonstrates a new simplified but efficient Gaussian network modeling approach to reveal the binding and folding dynamics of the adenosine deaminase A-riboswitch which can be readily extended for other RNA molecules.

ACKNOWLEDGMENTS

This work was supported in part by grants from the Chinese Natural Science Foundation (Grant Nos. 11474013, 31171267, and 11204267), the Beijing Natural Science Foundation (Grant No. 4152011), the China Scholarship Council (Grant No. 201308110231), and the National Institutes of General Medical Sciences (Grant Nos. GM083107 and GM116960).

[1]D. L. Lafontaine, Nat. Struct. Mol. Biol. **22**, 11 (2015).
[2]A. Serganov and E. Nudler, Cell **152**, 17 (2013).
[3]R. R. Breaker, Cold Spring Harbor Perspect. Biol. **4**, a003566 (2012).
[4]J. R. Mellin and P. Cossart, Trends Genet. **31**, 150 (2015).
[5]F. J. Grundy, S. C. Lehman, and T. M. Henkin, Proc. Natl. Acad. Sci. U. S. A. **100**, 12057 (2003).
[6]M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, and R. R. Breaker, Science **306**, 275 (2004).
[7]W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, and R. R. Breaker, Nature **428**, 281 (2004).
[8]M. Mandal and R. R. Breaker, Nat. Rev. Mol. Cell Biol. **5**, 451 (2004).
[9]O. Allnér, L. Nilsson, and A. Villa, RNA **19**, 916 (2012).
[10]V. Kumar, T. Endoh, K. Murakami, and N. Sugimoto, Chem. Commun. **48**, 9693 (2012).
[11]K. Neupane, H. Yu, D. A. Foster, F. Wang, and M. T. Woodside, Nucleic Acids Res. **39**, 7677 (2011).
[12]M. Sharma, G. Bulusu, and A. Mitra, RNA **15**, 1673 (2009).
[13]A. Serganov, Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, and D. J. Patel, Chem. Biol. **11**, 1729 (2004).
[14]J. C. Lin and D. Thirumalai, J. Am. Chem. Soc. **130**, 14080 (2008).
[15]R. Rieder, K. Lang, D. Graber, and R. Micura, ChemBioChem **8**, 896 (2007).
[16]U. D. Priyakumar and A. J. MacKerell, J. Mol. Biol. **396**, 1422 (2010).
[17]Z. Gong, Y. Zhao, C. Chen, and Y. Xiao, J. Biomol. Struct. Dyn. **29**, 403 (2011).
[18]J. G. Su, C. H. Li, R. Hao, W. Z. Chen, and C. X. Wang, Biophys. J. **94**, 4586 (2008).
[19]M. Liu, S. Wang, T. Sun, J. Su, Y. Zhang, J. Yue, and Z. Sun, PLoS One **7**, e40441 (2012).
[20]I. Bahar, B. Erman, and T. Haliloglu, Phys. Rev. Lett. **79**, 3090 (1997).
[21]L. W. Yang, X. Liu, C. J. Jursa, M. Holliman, A. J. Rader, H. A. Karimi, and I. Bahar, Bioinformatics **21**, 2978 (2005).
[22]W. K. Olson, M. Esguerra, Y. Xin, and X. Lu, Methods **47**, 177 (2009).
[23]R. L. Jernigan, M. C. Demirel, and I. Bahar, Int. J. Quantum Chem. **75**, 301 (1999).
[24]A. Kloczkowski, J. E. Mark, and B. Erman, Macromolecules **22**, 1423 (1989).
[25]L. W. Yang, A. J. Rader, X. Liu, C. J. Jursa, S. C. Chen, H. A. Karimi, and I. Bahar, Nucleic Acids Res. **34**, W24 (2006).
[26]D. Leipply and D. E. Draper, Biochemistry **50**, 2790 (2011).
[27]C. D. Stoddard, S. D. Gilbert, and R. T. Batey, RNA **14**, 675 (2008).
[28]J. K. Soukup and G. A. Soukup, Curr. Opin. Struct. Biol. **14**, 344 (2004).
[29]M. Mandal and R. R. Breaker, Nat. Struct. Mol. Biol. **11**, 29 (2004).
[30]O. M. Ottink, S. M. Rampersad, M. Tessari, G. J. Zaman, H. A. Heus, and S. S. Wijmenga, RNA **13**, 2202 (2007).
[31]D. Baker, Nature **405**, 39 (2000).
[32]N. Koga and S. Takada, J. Mol. Biol. **313**, 171 (2001).
[33]C. Clementi, P. A. Jennings, and J. N. Onuchic, Proc. Natl. Acad. Sci. U. S. A. **97**, 5871 (2000).
[34]O. V. Galzitskaya and A. V. Finkelstein, Proc. Natl. Acad. Sci. U. S. A. **96**, 11299 (1999).
[35]E. Alm and D. Baker, Proc. Natl. Acad. Sci. U. S. A. **96**, 11305 (1999).
[36]E. J. Sorin, B. J. Nakatani, Y. M. Rhee, G. Jayachandran, V. Vishal, and V. S. Pande, J. Mol. Biol. **337**, 789 (2004).
[37]E. J. Sorin, Y. M. Rhee, B. J. Nakatani, and V. S. Pande, Biophys. J. **85**, 790 (2003).
[38]See supplementary material at http://dx.doi.org/10.1063/1.4954992 for the fluctuations of nucleotides of *add* A-riboswitch with and without Mg^{2+} ions.