Energy Flow and Long-Range Correlations in Guanine-Binding Riboswitch: A Nonequilibrium Molecular Dynamics Study

Phuong H. Nguyen,[†] Philippe Derreumaux,[‡] and Gerhard Stock*[,†]

Institute of Physical and Theoretical Chemistry, Goethe University, Max-von-Laue-Strasse 7, D-60438 Frankfurt, Germany, and Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC, 13 rue Pierre et Marie Curie, 75005, Paris, France

Received: March 5, 2009; Revised Manuscript Received: April 28, 2009

A nonequilibrium molecular dynamics (MD) study of the temperature-induced energy flow in a RNA−ligand complex is presented, which employs extensive all-atom explicit solvent MD simulations of the aptamer domain of the guanine-sensing riboswitch (GRA). Since the few existing MD investigations of biomolecular energy flow have used quite different computational approaches, the applicability and performance of the various methods are compared first. In particular, a nonequilibrium correlation function $C_{ij}(\tau)$ is introduced that describes the cumulative response of residue j at delay time τ to the energy source at residue i. Employing this analysis, the anisotropic energy flow and long-range correlations in GRA are studied, which can be monitored over distances up to ≈4 nm. To test whether these long-range correlations are relevant for molecular function, the unbinding-induced conformational changes of GRA are calculated using the linear-response theory, assuming that the unbinding of the guanine ligand represents the first step responsible for the function of GRA. Interestingly, it is found that the same residues that are of functional importance are also prominently involved in the energy transfer. In particular, significant correlations between the guanine ligand and the distant "kissing" loops of GRA are found. This finding is in line with recent experiments which indicate that these long-range interactions may be important for the induced-fit binding of the ligand.

I. Introduction

Driven by experimental and theoretical progress, the energy transport through biomolecular systems has recently received considerable interest.[1] On the experimental side, transient infrared experiments have facilitated the real-time observation of the energy flow through protein secondary structure elements such as 3_{10}-helices.[2] On the theoretical side, biomolecular energy flow has been described by harmonic theories[3−7] and equilibrium and nonequilibrium molecular dynamics (MD) simulations.[8−18] These calculations revealed that the transport of vibrational energy in proteins and nucleic acids may be highly specific and anisotropic. The finding raises the question to what extent energy flow is important for the function of biological systems. In particular, it has been discussed whether efficient energy flow or communication between distant sites of a biomolecule is a prerequisite for allosteric regulation,[19] i.e., the regulation of the properties of an active site of a protein after binding of a substrate at a distant second site.[20] In some PDZ domain family proteins, for example, a chain-like network was identified through sequence analysis methods which links an active-site residue with distant sites.[21−23] Subsequent nonequilibrium MD simulations[11−13] identified the very same sequence of amino acids as a preferred energy transport channel.

The significance of functional dynamics is expected to be even more important for the functioning of nucleic acids, which provide only four types of residues (instead of 20) to encode information.[24,25] A prime example are riboswitches which represent a novel class of genetic control elements that function through direct interaction of small metabolite molecules with structured RNA elements.[26−28] The small ligand is bound with high specificity and affinity to its RNA target and induces upon binding conformational changes to the RNA's secondary and tertiary structure. One the best-studied systems is the guanine-sensing riboswitch (Figure 1), where a combination of X-ray, NMR, fluorescence spectroscopy, and MD simulation strongly suggests that conformational dynamics plays an important role in ligand binding and molecular recognition.[29−38] While the free and guanine-bound end states of this riboswitch are well amenable to experimental investigations, only little is known on the multistep hierarchical structural dynamics and conformational rearrangements to get from one state to the other, the whole process taking seconds.[39] Hence, many aspects of the initial mechanisms of binding or unbinding of the ligand and how this information is propagated from the binding pocket to distant sites of the RNA are so far largely unclear.

Moreover, it is not apparent whether this flow of information or energy, occurring typically in a 10 ps time scale, is at all related to the functional dynamics of the RNA, occurring in a time scale of microseconds and longer. Although the time scale separation speaks against it, we nevertheless know that the existence of fast fluctuations is often a prerequisite of a slow process. The most prominent example is the transition-state theory, where the pre-exponential factor can be interpreted as a fast "search rate", whereas the actual reaction rate is significantly slower, dictated by a barrier that has to be surmounted. Despite the time scale separation, the energy flow may, therefore, be of importance for the functional dynamics.[40]

Based on these considerations, in this work we perform extensive nonequilibrium MD simulations of the temperature-induced energy flow in the guanine-sensing riboswitch aptamer domain (GRA). To elucidate the connection of energy flow and function in this system, we also calculate the initial response

* Corresponding author. E-mail: stock@theochem.uni-frankfurt.de.
[†] Goethe University.
[‡] Laboratoire de Biochimie Theorique.

10.1021/jp902013s CCC: $40.75
Published on Web 06/10/2009

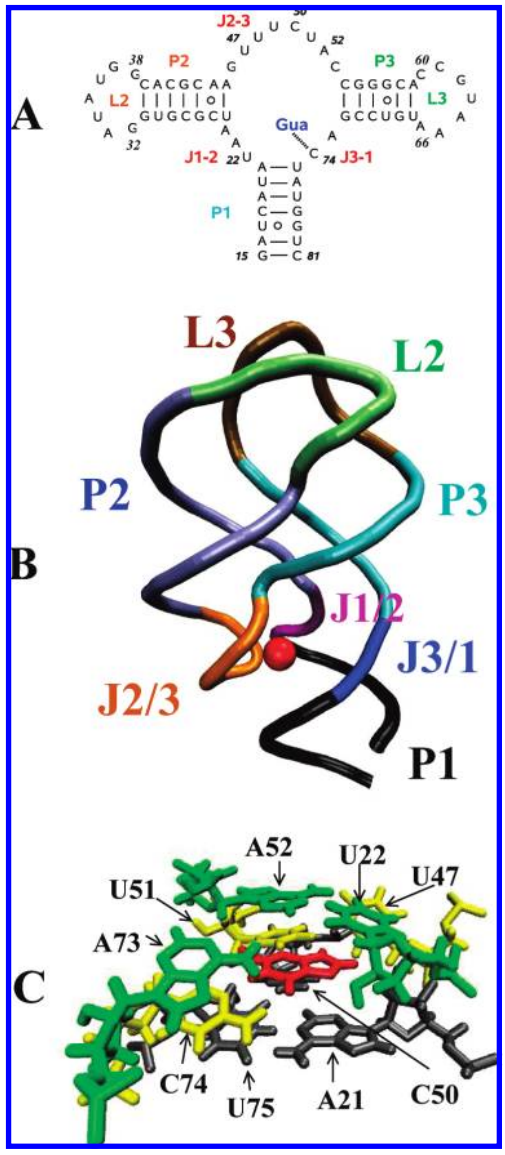

Figure 1. (A) Secondary and (B) tertiary structures of the riboswitch−guanine complex showing the two "kissing hairpins" P2 (ice-blue) and P3 (cyan) and the stem P1 (black) connected by junction segments J1/2 (pink), J2/3 (orange), and J3/1 (blue), which accommodate the ligand (red). The two terminal loops L2 and L3 are shown in green and ochre, respectively. (C) Close up of the binding pocket showing the upper (green) and lower (gray) base triple as well as the neighboring nucleotides (yellow) that are in the plane with the ligand (red).

of GRA upon unbinding using the linear-response theory.[41] The great majority of computational studies on RNA so far have been equilibrium MD simulations which studied the structure, dynamics, and binding of RNA.[15,16,42−49] Since the few existing MD investigations of biomolecular energy flow have used quite different computational approaches,[8−18] we first compare the applicability and sensitivity of various equilibrium and nonequilibrium methods to correctly predict the energy transport. In particular, we derive a nonequilibrium correlation function $C_{ij}(\tau)$ [eq 8] that describes the cumulative response of residue j at delay time τ to the energy flow in residue i. Based on this analysis, we discuss in some detail the specific and anisotropic energy flow in GRA, which is monitored over distances up to ≈4 nm. By comparing the energy-transfer results upon thermal perturbation to the linear-response results upon unbinding, we observe evident similarities between the two processes. This finding indicates that the same nucleotides that are prominently involved in the energy transfer also may be the ones of functional importance.

II. Theory and Methods

A. Equilibrium Simulations. We used the AMBER99 force field[50] to model the GRA molecule and the TIP3P water model[51] to describe the solvent. The crystal structure of the GRA−guanine complex (PDB code 1Y27)[31] was solvated in an octahedron box containing 13 390 water molecules and 66 Na^+ ions. Using either sodium or magnesium counterions, a recent MD study revealed that the overall structure and binding mode of GRA depend only weakly on the ion environment.[38] The system was simulated by using the GROMACS simulation package.[52,53] The equations of motion were integrated by using a leapfrog algorithm with a time step of 2 fs. We used the particle mesh Ewald method to treat the long-range electrostatic interactions.[54] The nonbonded interaction pair list was updated every 5 fs, using a cutoff of 1.2 nm. The system was first minimized using the steepest descent method and subsequently relaxed for 100 ps at constant volume and temperature ($T = 200$ K), respectively, using the Berendsen coupling procedure with a coupling time of 0.1 ps.[55] Then a 10 ns equilibrium simulation was carried out at constant temperature (200 K) and pressure (1 atm). As a control study, an additional 10 ns equilibrium simulation at low temperature (10 K) was performed.

B. Nonequilibrium Simulations. In a series of papers, we have recently developed several protocols of nonequilibrium MD simulations that are designed to mimic the experimental preparation by either a UV laser pulse triggering a molecular photoswitch[18] or an infrared pumping of a local vibrational mode.[17,56] Following subpicosecond intramolecular vibrational relaxation processes, in both cases the photoinduced energy is converted into randomized (but not necessarily thermalized) vibrational energy in the vicinity of the photoexcited moiety. In the absence of a concrete experiment, one may, therefore, approximately assume an initial temperature jump of some atom group in order to initiate the nonstationary energy flow. In order to mimic experimental conditions, a natural approach is to heat a certain residue (say, up to 1000 K), while the remaining atoms of the molecule are at room temperature (300 K). To achieve a better signal-to-noise ratio, on the other hand, several workers suggested to consider the molecule at low temperatures (e.g., 10 K) and heat the local residue only moderately (e.g., 300 K).[7,11] While this strategy results in reduced thermal fluctuations and, therefore, requires a smaller number of nonequilibrium trajectories, it may give rise to various problems. First, if the starting structures are obtained from an equilibrium simulation at low temperature, then the initial sampling of conformational space may be not adequate. Moreover, since the fluctuations of the system are directly related to its temperature, low-temperature MD simulations cannot correctly reproduce room-temperature quantities that depend on these fluctuations.[57,58]

To clarify these issues, we have performed three sets of nonequilibrium MD simulations. In one case (to be discussed in detail below), we adopted the equilibrium trajectory at 200 K to select 400 statistically independent structures. To cool down the ligand to 10 K, for each structure a short (≈5 ps) simulation was carried out where the ligand was coupled to a Berendsen thermostat at 10 K while the structure of the RNA was constrained. Using the resulting final geometries and velocities, nonequilibrium MD simulations were performed for 100 ps during which the ligand atoms are coupled to a heat bath in order to maintain the temperature at 10 K. For all other atoms, the temperature coupling is turned off. This way heat in the

form of kinetic energy will propagate from the nucleotides to the ligand. Alternatively, the same procedure is applied, but the ligand is kept at high (350 K) temperature which causes energy flow from the ligand to the molecule. Finally, we employed the low-temperature (10 K) equilibrium trajectory to generate starting positions for a third set of nonequilibrium MD simulations, where the ligand is kept at 200 K. A list of all nonequilibrium MD simulations is provided in the Supporting Information, Table S1.

C. Analysis of Nonequilibrium Data. As a straightforward quantity to study biomolecular energy flow, we consider the time evolution of the kinetic energy per atom of the *i*th residue:

$$E_i(t) = \frac{1}{N_a}\sum_{a=1}^{N_a} E_{ia}(t) \tag{1}$$

where N_a denotes the number of atoms of the residue and $E_{ia}(t)$ represents the kinetic energy of atom a in residue i. Performing nonequilibrium simulations, we are concerned with the ensemble average:

$$\langle E_i(t)\rangle = \frac{1}{N}\sum_{r=1}^{N} E_i^{(r)}(t) \tag{2}$$

where $E_i^{(r)}(t)$ is the residue energy along trajectory r and $N = 400$ denotes the number of nonequilibrium trajectories.

To describe the energy transfer from residue i to residue j, we introduce the two-time energy correlation function

$$K_{ij}^{(2)}(t,\tau) = \langle \delta E_i(t)\delta E_j(t+\tau)\rangle \tag{3}$$

where $\delta E_i(t) = E_i(t) - \langle E_i(t)\rangle$. As a consequence of the nonequilibrium excitation of the system for $t \geq 0$, the correlation function depends on two times t and τ, and the averages $\langle E_i(t)\rangle$ are time-dependent. In contrast, at equilibrium, we find time-independent averages, $\langle E_i(t)\rangle = \langle E_i^{eq}\rangle$ and stationary correlation functions satisfying $\langle \delta E_i(t+\tau)\delta E_j(t)\rangle = \langle \delta E_i(\tau)\delta E_j(0)\rangle$.

Being interested in the response of GRA to the (cool or hot) guanine ligand, in this study we focus on the correlation $K_{Lj}^{(2)}(t,\tau)$ between the energy flow of the ligand L, playing the role as a signal source, and the energy flow of residue j, receiving the signal. In our numerical cooling experiments, the ligand is kept at constant temperature for $t \geq 0$, that is, its mean kinetic energy is constant

$$\begin{aligned}\langle E_L(t)\rangle &= \langle E_L^{eq}\rangle \\ &= \frac{1}{NM}\sum_{r=1}^{N}\sum_{t=0}^{M-1} E_L^{(r)}(t)\end{aligned} \tag{4}$$

where the second sum represents a time average over M trajectory points. While in principle a single average (either in time or over the ensemble) is sufficient, the double average in eq 4 gives a better signal-to-noise ratio in practice. Assuming that the cooling of the ligand represents only a small perturbation from equilibrium (see Figure 2 below):

$$|\langle E_j(t)\rangle - \langle E_j^{eq}\rangle|/\langle E_j^{eq}\rangle \ll 1 \tag{5}$$

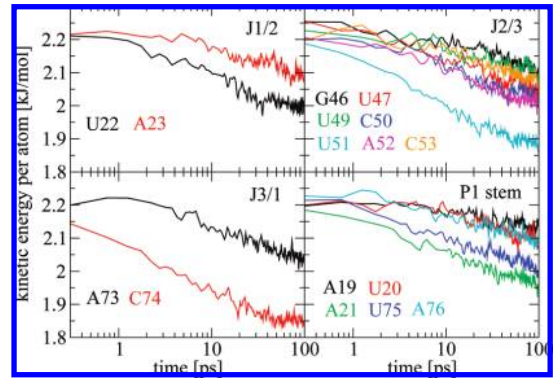

Figure 2. Time evolution of the kinetic energy $\langle E_i(t)\rangle$ of the 16 nucleotides of GRA that respond most strongly to the thermal perturbation.

we may also approximate $\langle E_j(t+\tau)\rangle \approx \langle E_j^{eq}\rangle$, which becomes exact at long times. As a further simplification, we recall that the response of residue j is caused by the *continuous* cooling of the ligand. This suggests to consider the time-integrated energy correlation:

$$K_{Lj}(\tau) = \frac{1}{M}\sum_{t=0}^{M-1} K_{Lj}^{(2)}(t,\tau) \tag{6}$$

We note in passing that, in the case of an impulsive temperature change (mimicking, e.g., the excitation by an ultrashort laser pulse at $t = 0$), one would rather consider $K_{Lj}^{(2)}(0,\tau)$ instead of $K_{Lj}(\tau)$. Put together, we obtain

$$K_{Lj}(\tau) = \frac{1}{NM}\sum_{r=1}^{N}\sum_{t=0}^{M-1}[E_L^{(r)}(t) - \langle E_L^{eq}\rangle][E_j^{(r)}(t+\tau) - \langle E_j^{eq}\rangle] \tag{7}$$

describing the cumulative response of residue j at delay time τ to the cooling of the ligand L. It is interesting to note that this definition of the correlation is given as a time and ensemble double average, which is similar to that of a stationary correlation function. This double average warrants an improved signal-to-noise ratio compared to the direct evaluation of the residue energy in eq 1. Hence, the energy correlation matrix $\{K_{ij}(\tau)\}$ allows us to identify the network of dominant communication pathways, i.e., the pathways which exhibit large correlations within the observed time scales.

We have also considered the normalized correlation:

$$C_{Lj}(\tau) = \frac{K_{Lj}(\tau)}{\sigma_L\sigma_j} \tag{8}$$

where $\sigma_j = \langle E_j^2(\tau)\rangle^{1/2}$ denotes the variance of the fluctuations of E_j (and similar for σ_L). Assuming again only small perturbations from equilibrium, the explicit time dependence of the variance can be neglected to a good approximation, $\sigma_j \approx \langle E_j^{2eq}\rangle^{1/2}$. In practice, we found that $C_{Lj}(\tau) \propto K_{Lj}(\tau)$. Since it is the commonly used definition, we employ the normalized energy correlation $C_{Lj}(\tau)$ in the discussion below.

D. Linear Response Theory. Recently, Kidera and coworkers[41] have shown that the static linear-response theory[59] is well suited to predict structural changes of a protein upon ligand binding when the equilibrium fluctuations of the protein in its

free state and a simple model for the force are employed. Here, we apply this formulation to study the reverse process, i.e., the conformational changes of GRA upon removing the guanine ligand. To this end, we used the crystal structure of the riboswitch−guanine complex[31] to calculate from the MD force field the forces $\mathbf{f}_j^{\text{bound}}$ acting on each atom j of the RNA. Then the ligand was removed, and the resulting forces $\mathbf{f}_j^{\text{free}}$ were calculated. The change of the forces $\delta\mathbf{f}_j = \mathbf{f}_j^{\text{bound}} - \mathbf{f}_j^{\text{free}}$ causes molecular conformational changes $\Delta\mathbf{r}_i = \mathbf{r}_i^{\text{bound}} - \mathbf{r}_i^{\text{free}}$, which in linear-response theory are given by[41]

$$\Delta\mathbf{r}_i = \frac{1}{k_B T}\sum_j \langle \delta\mathbf{r}_i \delta\mathbf{r}_j\rangle_{\text{eq}} \delta\mathbf{f}_j \tag{9}$$

where k_B is the Boltzmann's constant, T is the temperature, and $\langle \delta\mathbf{r}_i \delta\mathbf{r}_j\rangle_{\text{eq}}$ represents the covariance matrix of the atomic fluctuations $\delta r_i = \mathbf{r}_i - \langle \mathbf{r}_i\rangle_{\text{eq}}$ obtained from the equilibrium trajectory of the riboswitch−guanine complex. Assuming that residue i contains N_a atoms, we may define the average displacement d_i of this residue as

$$d_i = \frac{1}{N_a}\sum_{a\in \text{residue}i} |\Delta\mathbf{r}_a| \tag{10}$$

For interpretational purposes, we may also consider the covariance between two residues i and j defined by

$$\sigma_{ij} = \sum_{a\in i, b\in j} \langle \delta\mathbf{r}_a \delta\mathbf{r}_b\rangle_{\text{eq}} \tag{11}$$

III. Results and Discussion

A. General Characterization of Simulations. Let us first consider the structural features of the GRA−guanine complex. Figure 1 shows the tertiary structure of the complex characterized by two "kissing hairpins" (labeled P2 and P3) and a stem (P1) connected by junction segments (J1/2, J2/3, and J3/1) which accommodate the ligand. Stem P1 consists of beginning and end segments of the RNA (residues 15−21 and 75−81, respectively), which are involved in base pairs interactions. Hairpins P2 and P3 are formed by residues 25−45 and 54−72, respectively, and are capped by seven-residue loops L2 and L3. These hairpins are anchored together through the formation of five base pairs. Two of the base pairs are aligned through standard Watson−Crick pairing, while the others form noncanonical base pairs in agreement with X-ray and NMR studies.[30,31,36] As observed in the experiment,[31,35] the residues in the junction area are involved in intramolecular triple (residues A21:U75:C50 and A23:G46:C53) and double (residues U22:A52 and U20:A76) base interactions. Moreover, we observed two base pair interactions between the ligand and RNA, that is, an intermolecular Watson−Crick G−C base pair interactions between the ligand and residues C74 and between the N3/N9 edge of the guanine ligand and residue U51. A hydrogen bond is also observed between the U22 2′-hydroxyl group and N7 of the guanine.[31] A more detailed MD characterization of the structural features of GRA is given in ref 38.

Despite the absence of stabilizing Mg^{2+} ions, the equilibrium simulation of the complex was found to be quite stable, showing an average root-mean-square deviation (rmsd) of 0.20 nm with respect to the X-ray structure.[31] Interestingly, the rmsd as well as the radius of gyration $R_g = (\sum_i m_i d_i^2)^{1/2}/(\sum_i m_i)^{1/2}$ hardly ($\lesssim$1%) changes during the nonequilibrium simulations, thus, revealing that the riboswitch does not undergo conformational changes within the 100 ps of nonequilibrium simulation time.

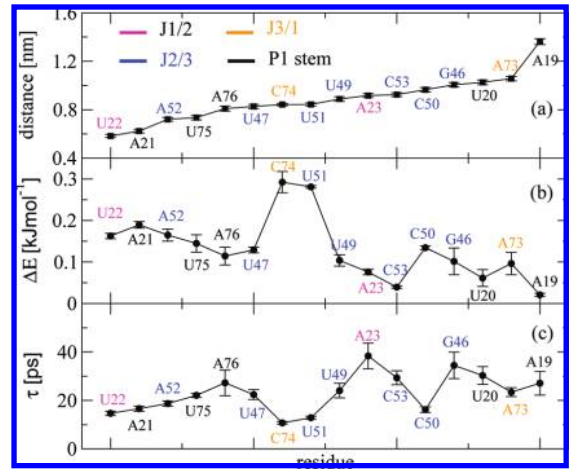

Figure 3. Analysis of the energy flow in GRA as shown in Figure 2. Shown are (a) the average distance between the centers of mass of the ligand and the nucleotides, respectively, (b) the kinetic energy change ΔE_i of each nucleotide, and (c) the time scale of the energy transfer as obtained from a single exponential fit of the energy $\langle E_i(t)\rangle$ in Figure 2. Error bars represent standard deviations due to time and ensemble averages.

B. Nonequilibrium Energy Flow. Following the thermal perturbation of the system at time $t = 0$ (by cooling the guanine ligand to 10 K), the nucleotides of GRA sense the low temperature of the ligand and consequently change their kinetic energy. Adopting the 16 nucleotides that respond most strongly to the perturbation, Figure 2 shows the time evolution of the nucleotide's kinetic energy $\langle E_i(t)\rangle$ defined in eq 2. As expected, these residues are mostly located in direct vicinity of the ligand, including nucleotides of the junction segments J1/2 (U22, A23), J2/3 (G46, U47, U49, C50, U51, A52, C53), and J3/1 (A73, C74), as well as of the P1 stem (A19, U20, A21, U75, A76). The results clearly show that residues C74 and U51 respond most strongly to the perturbation followed by the somewhat weaker response of residues A21 and U22. Hence, these nucleotides are the cornerstones mediating the communication between the ligand and the riboswitch.

Interestingly, though, the energy transport in GRA does not occur in a simple isotropic diffusional manner. This is demonstrated in Figure 3, which clearly reveals that the kinetic energy change $\Delta E_i = \langle E_i(0)\rangle - \langle E_i(100\text{ps})\rangle$ of a nucleotide does not correlate with the corresponding ligand−nucleotide distance. Furthermore, the time scale τ_i of the response of each nucleotide is shown. The energy-transfer rate $1/\tau_i$ roughly correlates with the energy ΔE_i but, again, not with the distance. Moreover, the energy flow does not seem to depend on the nucleotide type. For example, the uracils U49 and U51 are in similar distance to the ligand but respond quite differently.

A closer analysis suggests that the energy flow may follow the hydrogen-bonding network rather than the distance. For example, residues C74 and U51 are on average bound by three and two hydrogen bonds to the ligand, respectively. These nucleotides get about twice as much energy as residue U22, which is at half the distance but only connected through a single hydrogen bond. Apart from hydrogen bonding, van der Waals and electrostatic interactions also may become important for the ligand−nucleotide energy transfer, in particular, in the presence of a large common border surface. For example, A21 and U22 exhibit a similar distance and energy change, which in the case of U22 is mostly due to one hydrogen bond to the

ligand, while in the case of A21, the absence of such a hydrogen bond is mostly caused by its weak stacking interaction with the ligand. We also note that residues U47, U51, and C74 lie in the same plane and have the same distances to the ligand (see Figure 1C), but U47 responds quite weakly compared to U51 and C74. This indicates that U47 is quite stable, which is in line with recent experimental findings that U47 forms an extensive set of hydrogen-bonding interactions with its surrounding environment.[32] As a consequence of the dense packing of the guanine ligand in the binding pocket, the details of the ligand−nucleotide energy transfer might be quite complex. Nevertheless, we may conclude from Figure 3 that residues C74, U51, and A21/U22 represent the three main locations where the RNA−ligand energy transfer takes place. As such, these nucleotides are primarily responsible for the subsequent communication between the ligand and the riboswitch.

As explained in Section II.B., several nonequilibrium simulation protocols have been suggested to calculate the energy transfer in biomolecules. This includes rather realistic simulations of the experiments[2,17] (photoinduced heating of a certain residue with remaining atoms at room temperature) as well as low-temperature simulations,[7,11] which require fewer trajectories to achieve a sufficient signal-to-noise ratio, and the methods pursued here of watching the energy flow of a 200 K molecule to a 10 K ligand. Hence, it is interesting to study to what extent the energy transfer depends on these simulation conditions.

The Supporting Information, Figure S1, shows the kinetic energy $\langle E_i(t)\rangle$ of the same nucleotides as in Figure 2 for the cases (a) ligand at 350 K and RNA at 200 K (L350/R200) and (b) ligand at 200 K and RNA at 10 K (L200/R10). (For a list of all simulations, see Supporting Information, Table S1.) While case (L350/R200) looks similar to the results of our "standard" (L10/R200) simulation, it is obvious that high temperature and, therefore, large fluctuations of ligand *and* RNA require significantly more sampling in order to achieve a reasonable signal-to-noise ratio. In contrast, the signal-to-noise ratio is much better for the (L200/R10) results because the majority of atoms of the total system are at a low temperature. We note, however, that the energy transport properties may change with temperature.[57,58] Showing the energy change ΔE_i and transfer times τ_i for the cases (L350/R200) and (L200/R10), Supporting Information, Figure S2 reveals that these results look similar but not identical to the (L10/R200) case shown in Figure 3. This is less so for the high-temperature results of (L350/R200) where the C74 and U51 nucleotides still respond most strongly, but certainly for the low-temperature (L200/R10) case where the response times differ particularly from the (L200/R10) results.

Finally, we have also employed the rmsd of a residue with respect to its initial structure as a descriptor of energy transfer in the (L10/R200) simulation. As shown in Supporting Information, Figure S3, the nucleotide's rmsd remained virtually constant throughout the nonequilibrium simulation time of 100 ps and is, therefore, not suited to describe the energy flow in GRA. This is in contrast to the findings in ref 11 where it suggests that the rmsd monitors well the energy transfer in a PDZ domain family protein. The different findings may be caused by the fact that the protein studied in ref 11 is significantly more flexible than the GRA system under consideration here. Recalling the quite different definitions, moreover, it is a priori not clear to what extent the rmsd may reflect the energy flow.

C. Long-Range Correlations. Figure 2 has demonstrated that the change of kinetic energy becomes small for sites that are distant to the ligand. Just as in the experiment, the energy flow is hard to observe due to the decreasing signal-to-noise ratio. As a quantity that is less affected by the noise of the complete system, we, therefore, consider the time-delayed energy correlation $C_{Li}(\tau)$ defined in eq 8. It reflects the energy flow of the ligand L, playing the role as a signal source, to residue i, receiving the signal. Figure 4 shows the time evolution of the nonequilibrium correlation for two close residues, U51 and C74, as well as for two distant residues, U34 and C60. The correlation functions of U51 and C74 start at their maximum values, reflecting the fact that these nucleotides feel the cold ligand almost immediately (see Figure 2). The correlation decays on a time scale of 10 ps, which roughly coincides with the risetime of $\langle E_i(t)\rangle$ shown in Figure 2.

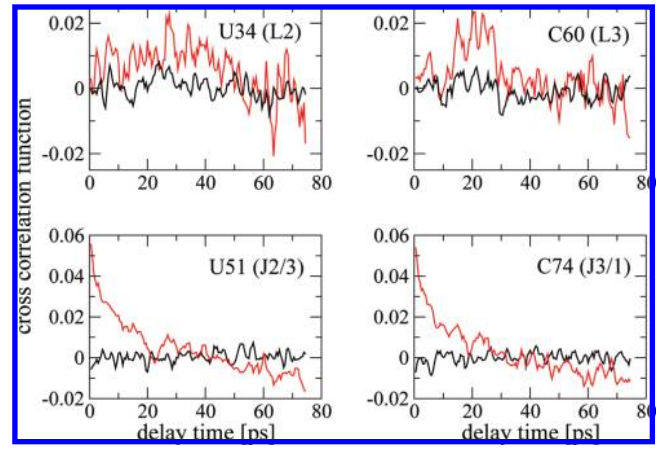

Figure 4. Time-delayed energy correlation $C_{Li}(\tau)$ between the ligand and selected nucleotides of GRA obtained from nonequilibrium (red) and equilibrium (black) MD simulations.

To estimate the noise of the data, the correlation $C_{Li}(\tau)$ was also calculated from a equilibrium trajectory. Since the kinetic energy $(1/2)\sum_i p_i^2/m_i$ is given as a sum of independent variables, we obtain under equilibrium conditions $\langle f(p_i)g(p_j)\rangle = \langle f(p_i)\rangle\langle g(p_j)\rangle$ for arbitrary functions f and g and $i \neq j$. Hence, in equilibrium, the correlation of the kinetic energies of two different atoms vanishes, i.e., $C_{ij}(\tau) = 0$. The fluctuations of the equilibrium correlation around zero, thus, define the noise level of the data. By comparing equilibrium and nonequilibrium results for $C_{Li}(\tau)$, Figure 4, thus, reveals that the nonequilibrium correlation is significantly over the noise level. We note in passing that we have also considered the case of a hot (300 K) ligand in a cold (10 K) RNA (L300/R10). As shown in the Supporting Information, Figure S4, the resulting time evolution of $C_{Li}(\tau)$ is quite similar to the (L10/R200) case discussed in Figure 4 except that the signal-to-noise ratio of the (L300/R10) data is inferior due to the higher temperature (300 K vs 200 K). We note that the situation is actually the other way around for the energy changes $\langle E_i(t)\rangle$ shown in Figures 2 and Supporting Information, Figure S1, i.e., the signal-to-noise ratio of the (L300/R10) results is better than for the (L10/R200). This is because $\langle E_i(t)\rangle$ reflects the fluctuations of the entire system (which are much lower if the protein is cold), while $C_{Li}(\tau)$ depends only on the fluctuations of the ligand and a single residue.

We are now in a position to study the correlation between the ligand and distant sites of GRA. Considering the kissing-loop residues C60 (L3) and U34 (L2), Figure 4 reveals that $C_{Li}(\tau)$ builds up within 20 and 30 ps. This long-range correlation is relatively strong (about one-third of the value obtained for the direct neighbors U51 and C74) and clearly over the noise level. This finding is interesting in light of the fact that several

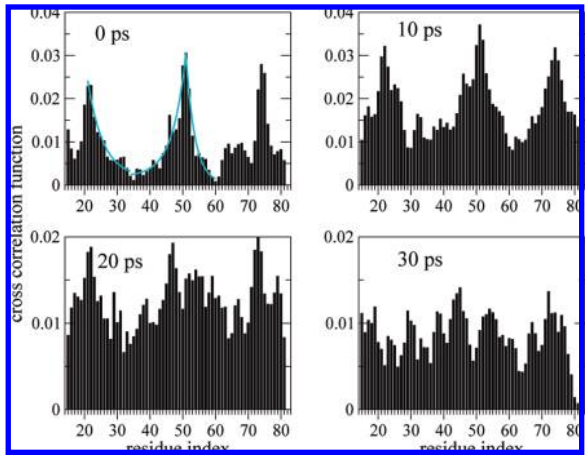

Figure 5. Time evolution of correlation $C_{Lj}(\tau)$ between the guanine ligand and the jth nucleotide of the RNA. For better representation, each point of $C_{Lj}(\tau)$ has been averaged over three nucleotides ($j-1$, j, and $j+1$) as well as over a time window of ± 4 ps. Colored lines in the ($\tau = 0$) panel are exponential fits of the correlation decay, yielding correlation lengths of $n = 4.6$, 4.1, and 2.7 residues for fitting ranges $22 \leq j \leq 35$, $35 \leq j \leq 51$, and $51 \leq j \leq 63$, respectively.

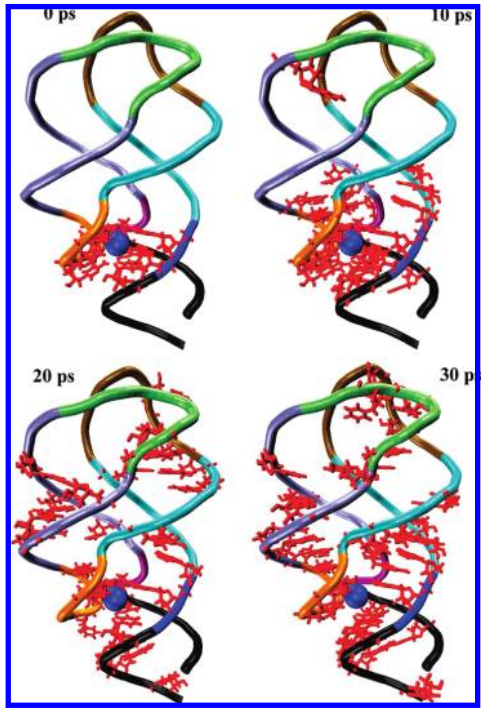

Figure 6. Structural illustration of the buildup of the energy flow in GRA as monitored by the time-delayed energy correlation function $C_{Li}(\tau)$ for $\tau = 0$, 10, 20, and 30 ps. Nucleotides that exhibit strong correlation (defined by $C_{Li}(\tau) \geq 1/2 \max\{C_{Lk}(\tau)$, for all $k\}$) are drawn in red. The color code of the RNA backbone is the same as in Figure 1, and the ligand is in blue.

experiments have indicated that these long-range ligand−loops interactions may be important for the induced-fit binding of the ligand.[34,36,37]

To illustrate the overall energy transfer in GRA, Figures 5 and 6 show two complementary representations of correlations $C_{Lj}(\tau)$ for delay times $\tau = 0$, 10, 20, and 30 ps. Initially ($\tau = 0$), we find direct and strong correlations between the ligand and its neighboring nucleotides A21/U22, U51, and C74. These are precisely the nucleotides whose kinetic energy shows the fastest and strongest response in Figure 2. The bar representation of $C_{Li}(\tau)$ in Figure 5 reveals the correlation decays from the nucleotides in the ligand region to the distant loop regions L2 (residues 35 ± 3) and L3 (residues 63 ± 3). For example, the decay of $C_{Lj}(\tau)$ from $j = 22$ to 35 is well approximated by a function $\propto e^{-(j\ -\ 22)/n}$ with a correlation length of $n = 4.6$ residues. As an exception of this exponential behavior, we find that the distant L3 region exhibits significant "instantaneous" correlation with the ligand. This means that during the averaging window of 4 ps already some communication between the ligand and L3 loop takes place.

Within the next 10 ps, the energy spreads out to the remaining nucleotides of the junction region as well as to the hairpins P2 and P3. Interestingly, we find at these early times a relatively large correlation with some residues in the distant L2 region, while the initial correlation with the L3 loop has vanished. At $\tau = 20$, the temperature-induced perturbation has propagated through the entire RNA backbone. The differences between highly and weakly correlated regions by now are a factor of 2 at the most (compared to a factor of 10 at $\tau = 0$). We note a relatively high correlation of the L3 region, while the correlation with the L2 region has reached a minimum. At longer times ($\tau = 30$), the correlation pattern of $C_{Li}(\tau)$ becomes quite complicated and hampers a simple interpretation.

To further investigate the mechanism by which long-range correlations are established, we may also consider the time evolution of the entire energy correlation matrix $\{C_{ij}(\tau)\}$ shown in Figure 7. At $\tau = 0$, the correlation matrix nicely reveals the connectivities of the riboswitch. We find strong correlations between residues A21/U22, U51/A52, and C74 which represent the three cornerstones mediating the communication between the ligand and riboswitch. Furthermore, the base pairs along hairpin P2 (i.e., 25:45,..., 31:39) result in weak correlations (but less so for hairpin P3). For short delay times ($\tau \lesssim 10$, data not shown), the correlation of residues A21/U22, U51/A52, and C74 is seen to spread out to the neighboring nucleotides. Furthermore, one observes that the initial symmetry of the correlation matrix $C_{ij}(\tau)$ gets lost with an increase in time delay τ. The symmetry is completely lost at $\tau = 20$ (Figure 7B). Most interestingly, we find that within that time the correlation has increased from the main nucleotides A21/U22, U51/A52, and C74 all along the backbone (horizontal features in Figure 7B). That is, A21/U22 and C74 correlate mostly with residues 40−80, while U51/A52 correlate with almost all of the residues. These findings suggest that the long-range correlation is mediated via through-bond interactions along the backbone.

It is interesting to contrast the nonequilibrium correlation matrix $C_{ij}(\tau)$ with results for the equilibrium covariance matrix σ_{ij} [eq 11] because this quantity also has been employed as an indicator for communication networks.[15,16] As shown in Figure 7C, however, σ_{ij} essentially reflects the connectivity of the RNA, i.e., we obtain large covariance along all three of the hairpins and the interacting kissing loops. Hence, there is some similarity of the equilibrium covariance matrix σ_{ij} with the nonequilibrium correlation matrix $C_{ij}(\tau)$ at $\tau = 0$ (Figure 7A), but σ_{ij} cannot capture the energy transfer shown by $C_{ij}(\tau)$ at $\tau = 20$ (Figure 7B).

D. Response upon Unbinding. Whether anisotropic energy transport and long-range correlations are relevant for molecular function remains a key question in the discussion of biomolecular energy flow. To address this issue, we calculated the conformational changes of GRA upon the unbinding of the ligand, which presumably represents the first step responsible for the function of GRA. Employing the linear-response theory as described in Section II.D., Figure 8 displays the average displacements [eq 10] of the nucleotides of GRA upon unbind-

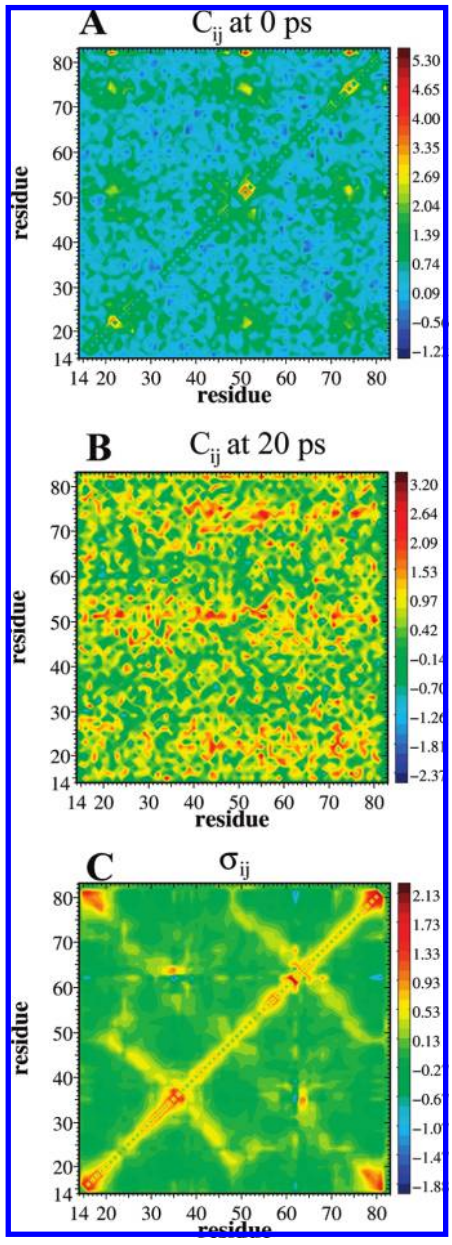

Figure 7. (A) and (B) Time evolution of the energy correlation matrix $\{C_{ij}(\tau)\}$ of GRA, where the time-delayed energies $\delta E_i(\tau)$ are plotted along the horizontal axis. The guanine ligand is assigned the residue number 82. (C) Equilibrium residue−residue covariance matrix σ_{ij}. To facilitate the representation of the long-range off-diagonal correlations, diagonal elements of both matrices are disregarded. The numbers of the color bar refer to 100 × $C_{ij}(\sigma_{ij})$.

ing. As may be expected, we find that the nucleotides of the binding site respond most strongly to the removing of the ligand. The most pronounced changes are seen for the nucleotides U22 (J1/2), U51, and A52 (J2/3) as well as A73 and C74 (J3/1). Interestingly, the upper base triple of the binding site (U22, A52, and A73) is somewhat opened in the free state of GRA (see inset of Figure 8), while the lower base triple (A21, C50, and U75) is largely conserved. This suggests that U22 and A52 may function as a gateway for the ligand to enter or leave the binding pocket. Most importantly, we observe a long-range response of the nucleotides in the loop regions L2 and L3, that is, we find small but well-defined conformational changes around nucleotide A35 in loop L2 as well as around C61 in loop L3.

By comparing the linear-response results of unbinding to the energy-transfer results of thermal perturbation, we observe

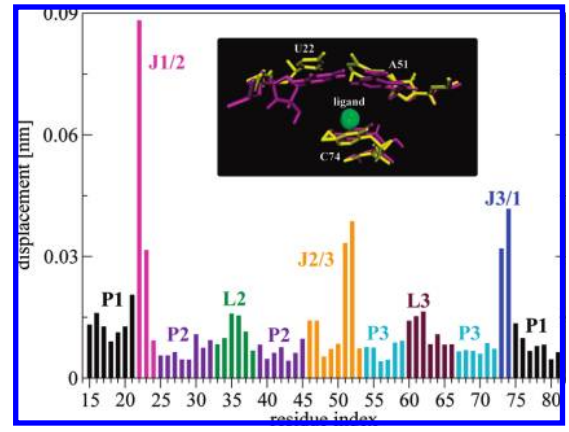

Figure 8. Average displacements of the nucleotides of GRA upon unbinding of the guanine ligand as obtained from the linear-response theory. The color code of the nucleotides is the same as in Figure 1. The inset compares crystal (magenta) and predicted unbinding (yellow) structures of nucleotides U22, A51, and C74.

evident similarities between the two approaches; a strong response of the nucleotides in the binding site, and a weak, but well-defined, correlation of the nucleotides in the ligand region with the distant nucleotides of the kissing loops. Assuming that the linear-response results in Figure 8 indicate the residues of functional importance for the unbinding process, we have found that the same nucleotides that are of functional importance are also prominently involved in the energy transfer.

IV. Concluding Remarks

We have presented comprehensive nonequilibrium MD simulations of the temperature-induced energy flow in a RNA−ligand complex. The simulations have shown that the energy transfer from the guanine ligand to its neighboring residues is anisotropic, that is, the transferred energy may differ by up to a factor of 10 (typically by a factor of 2) for residues in similar distance to the ligand (see Figure 3). Hence, the predicted anisotropy should be observable in transient IR experiments with suitable isotope labeling.[2] The energy transport is mediated by the hydrogen-bonding network between the ligand and its surrounding nucleotides and, to a smaller extent, by stacking interactions. Comparing several nonequilibrium simulation protocols, we have found that it is important to conduct the simulations at the appropriate temperature, since the conformational fluctuations may change the transport properties of the system.[57,58]

To facilitate the study of long-range energy transfer, we have introduced a nonequilibrium correlation function $C_{ij}(\tau)$ [eq 8], which describes the cumulative response of residue j at delay time τ to the energy flow in residue i. The improved signal-to-noise ratio of this method enabled us to identify spatial correlations over distances up to 4 nm (see Figures 4−7). In particular, we found significant long-range correlations between the guanine ligand and the distant kissing loops of GRA. This finding is interesting in the light of recent experimental evidence which suggests that these long-range interactions may be important for the induced-fit binding of the ligand.[34,36,37]

Assuming that the unbinding of the guanine ligand represents the first step responsible for the function of GRA, we have calculated the unbinding-induced conformational changes of GRA using the linear-response theory. Interestingly, it has been found that the same nucleotides that are of functional importance are also prominently involved in the energy transfer. That is, we observed a strong response of the nucleotides in the binding

site and a weaker, but well-defined, correlation of the nucleotides in the ligand region with the distant nucleotides of the kissing loops. While transport of vibrational energy may not be the reason for functional dynamics, it might, nevertheless, indicate possible connectivity in biomolecules, which is important for functional dynamics.

Acknowledgment. We thank Rainer Hegger and Peter Hamm for numerous inspiring and helpful discussions. This work has been supported by the Frankfurt Center for Scientific Computing, the Fonds der Chemischen Industrie, the Deutsche Forschungsgemeinschaft, CNRS, and University of Paris 7 Denis-Diderot.

Supporting Information Available: All nonequilibrium MD simulations are provided, kinetic energies of the nucleotides. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) *Proteins: Energy, Heat and Signal Flow (Computation in Chemistry)*; Leitner, D.; Straub, J., Eds.; Taylor and Francis/CRC Press: London, 2009.
(2) Botan, V.; Backus, E.; Pfister, R.; Moretto, A.; Crisma, M.; Toniolo, C.; Nguyen, P. H.; Stock, G.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12749–12754.
(3) Yu, X.; Leitner, D. M. *J. Phys. Chem. B* **2003**, *107*, 1698–1707.
(4) Yu, X.; Leitner, D. M. *J. Chem. Phys.* **2005**, *122*, 054902.
(5) Fujisaki, H.; Straub, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6726–6731.
(6) Fujisaki, H.; Straub, J. *J. Phys. Chem. B* **2007**, *111*, 12017–12023.
(7) Moritsugu, K.; Miyashita, O.; Kidera, A. *Phys. Rev. Lett.* **2000**, *85*, 3970–3973.
(8) Henry, E. R.; Eaton, W. A.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8982–8986.
(9) Sagnella, D. E.; Straub, J. E. *J. Phys. Chem. B* **2001**, *105*, 7057–7063.
(10) Moritsugu, K.; Miyashita, O.; Kidera, A. *J. Phys. Chem. B* **2003**, *107*, 3309–3317.
(11) Ota, N.; Agard, D. A. *J. Mol. Biol.* **2005**, *351*, 345–354.
(12) Sharp, K.; Skinner, J. J. *Proteins* **2006**, *65*, 347–361.
(13) Kong, Y.; Karplus, M. *Proteins* **2009**, *74*, 145–154.
(14) Ishikura, T.; Yamato, T. *Chem. Phys. Lett.* **2006**, *432*, 533–537.
(15) Ghosh, A.; Vishveshwara, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *104*, 15711–15716.
(16) Seth, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6620–6625.
(17) Nguyen, P. H.; Stock, G. *J. Chem. Phys.* **2003**, *119*, 11350.
(18) Nguyen, P. H.; Stock, G. *Chem. Phys.* **2006**, *323*, 36–44.
(19) Bahar, I.; Chennubhotla, C.; Tobi, D. *Curr. Opin. Struct. Biol* **2007**, *17*, 633–640.
(20) Changeux, J.-P.; Edelstein, S. J. *Science* **200**, *308*, 1424–1428.
(21) Lockless, S. W.; Ranganathan, R. *Science* **1999**, *286*, 295–299.
(22) Shulman, A. I.; Larson, C.; Mangelsdorf, D. J.; Ranganathan, R. *Cell* **2004**, *116*, 417–429.
(23) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. *Nat. Struct. Biol.* **2003**, *10*, 59–69.
(24) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. *Science* **2006**, *311* (5761), 653–656.
(25) Hall, K. B. *Curr. Opin. Chem. Biol.* **2008**, *12*, 612–618.
(26) Edwards, T. E.; Klein, D. J.; Ferre-d'Amare, A. R. *Curr. Opin. Struct. Biol.* **2007**, *17*, 273–279.
(27) Schwalbe, H.; Buck, J.; Fürtig, B.; Noeske, J.; Wöhnert, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 1212–1219.
(28) Montange, R. K.; Batey, R. T. *Annu. Rev. Biochem.* **2008**, *37*, 117–133.
(29) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. *Cell* **2003**, *113*, 577–586.
(30) Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, *432*, 411–415.
(31) Serganov, A.; Yuan, Y.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chem. Biol.* **2004**, *11*, 1729–1741.
(32) Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. *J. Mol. Biol.* **2006**, *359*, 754–768.
(33) Lemay, J.-F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M. J.; Lafontaine, D. A. *Chem. Biol.* **2006**, *13*, 857–868.
(34) Rieder, R.; Lang, K.; Graber, D.; Micura, R. *ChemBioChem* **2007**, *8*, 896–902.
(35) Noeske, J.; Richter, C.; Grundl, M. A.; Nasiri, H. R.; Schwalbe, H.; Wöhnert, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1372–1377.
(36) Noeske, J.; Buck, J.; Fürtig, B.; Schwalbe, H.; Wöhnert, J. *Nucleic Acids Res.* **2007**, *35*, 572–582.
(37) Stoddard, C. D.; Gilbert, S. D.; Batey, R. T. *RNA* **2008**, *14*, 675–684.
(38) Villa, A.; Wöhnert, J.; Stock, G. *Nucl. Acids Res.* **2009**, in press.
(39) Buck, J.; Fürtig, B.; Noeske, J.; Wöhnert, J.; Schwalbe, H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15699–15704.
(40) Nguyen, P. H.; Hamm, P.; Stock, G. In *Proteins: Energy, Heat and Signal Flow (Computation in Chemistry)*; Leitner, D.; Straub, J., Eds., Taylor and Francis/CRC Press: London, 2009.
(41) Ikeguchi, M.; Ueno, J. M. S.; Kidera, A. *Phys. Rev. Lett.* **2005**, *94*, 078102.
(42) Zacharias, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311–317.
(43) Mu, Y.; Stock, G. *Biophys. J.* **2006**, *90*, 391–399.
(44) Rázga, F.; Zacharias, M.; Réblová, K.; Koča, J.; Šponer, J. *Structure* **2006**, *14*, 825–835.
(45) Kormos, B. L.; Baranger, A. M.; Beveridge, D. L. *J. Am. Chem. Soc.* **2006**, *128*, 8992–8993.
(46) Auffinger, P.; Hashem, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325–333.
(47) Hashem, Y.; Westhof, E.; Auffinger, P. In *Computational Structural Biology*; Schwede, T.; Peitsch, M. C., Eds.; World Scientific: New York, 2008.
(48) Chen, S.-J. *Annu. Rev. Biophys.* **2008**, *37*, 197–214.
(49) Villa, A.; Widjajakusuma, E.; Stock, G. *J. Phys. Chem. B* **2008**, *112*, 134–142.
(50) Wang, J.; Cieplak, P.; Kollman, P. *J. Comput. Chem.* **2000**, *21*, 1049.
(51) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
(52) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.
(53) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(54) Darden, T.; York, D.; Petersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(56) Backus, E.; Nguyen, P. H.; Botan, V.; Pfister, R.; Moretto, A.; Crisma, M.; Toniolo, C.; Stock, G.; Hamm, P. *J. Phys. Chem. B* **2008**, *112*, 9091–9099.
(57) Backus, E.; Nguyen, P. H.; Botan, V.; Pfister, R.; Moretto, A.; Crisma, M.; Toniolo, C.; Stock, G.; Hamm, P. *J. Phys. Chem. B* **2008**, *112*, 15487–15492.
(58) Chennubhotla, C.; Bahar, I. *PLoS Comput. Biol.* **2007**, *9*, 1716.
(59) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University: Oxford, 2001.

JP902013S