Molecular Dynamics Simulation of the Structure, Dynamics, and Thermostability of the RNA Hairpins uCACGg and cUUCGg

Alessandra Villa,* Elisabeth Widjajakusuma, and Gerhard Stock

J.W. Goethe University, Institute of Physical and Theoretical Chemistry, Max-von-Laue-Str. 7, Frankfurt am Main, Germany 60438

Received: August 10, 2007; In Final Form: October 16, 2007

Classical replica-exchange molecular dynamics simulations are performed to study structure, dynamics and thermostability of the 14-mer RNA hairpins uCACGg and cUUCGg. Despite of the different sequence and closing base pair of the two systems, recent NMR studies have shown that the tetraloop CACG is strikingly similar in overall geometry and hydrogen bonding to the canonical UUCG tetraloop. On the other hand, the two systems differ significantly in their functionality and thermostability. The simulations confirm the structural similarities of the two RNA hairpins at room temperature but also reveal that the UUCG loop is more flexible than the CACG loop. Concerning the functionality, the CACG loop shows a stronger attitude to donate hydrogens than the UUCG loop, although their global solvent accessible surface is quite similar. The simulations qualitatively reproduce the experimentally found difference in melting temperatures (20 K). In the case of the uCACGg hairpin, the thermal unfolding occurs cooperatively in an all-or-none fashion, while the cUUCGg hairpin shows less cooperativity but exhibits intermediate states during the unfolding process.

1. Introduction

Coxsackieviruses are a genus of RNA viruses associated with several acute and chronic human diseases. These viruses carry the genetic information on a positive-sense single-strand RNA that can be immediately translated by the host cell. The essential step in the assembly of the viral replication process is the formation of a complex between the viral protease and 5′-nontranslated region of the virus.[1,2] In this domain, the stem-loop D constitutes the major recognition site for the viral protease $3CD^{pro}$. In vitro study showed that stem-loop D alone is sufficient to bind the protein $3CD^{pro}$,[3] and it was suggested[4,5] that the RNA-protein recognition process might be based on structural properties instead of on the specific sequence.

In the last years, nuclear magnetic resonance (NMR) techniques have been used to investigate the structure of the stem-loop D of coxsackievirus B3.[4,6] The NMR structure of its apical tetraloop uCACGg is strikingly similar in overall geometry and hydrogen bonding to the canonical cUUCGg tetraloop,[7,8] despite the different sequence and closing base pair of the two loops. In particular, the CACG loop was found to exhibit the interactions which were thought to explain the unusual stability of the UNCG loop family[9,10] (where N is any nucleotide). This led to the suggestion[11] to extend the UNCG tetraloop family to the motif YNMG (where Y is a pyrimidine and M is an adenine or a cytosine).

Despite their considerable structural similarity, the uCACGg and cUUCGg tetraloops were found to differ in their functionality and thermostability. Mutation studies of the stem-loop D of coxsackievirus B3 showed that the replacement of uCACGg by cUUCGg did not lead to functional binding.[6] Furthermore, the melting temperature of the uCACGg loop is about 20 K lower than the one of the cUUCGg loop.[12] While this difference in stability is in accordance with empirical rules using the base-pair sequence,[13,14] a microscopic picture of the unfolding of the RNA hairpins has not yet been established.

In this work, we employ classical replica-exchange molecular dynamics (REMD) simulations[15−17] to study the effects of the loop sequence and the closing base pair on the conformational distribution and on the thermostability of RNA hairpins. The REMD method performs a number of simulations in parallel at various temperatures, and on the basis of a Metropolis criterion, the conformations are periodically switched between neighboring temperatures. Hence, even if a trajectory is temporarily trapped in a local minimum, it may escape via a transition to a higher temperature. Furthermore, the method naturally yields thermodynamic quantities as a function of temperature. Representing the 14-mer hairpins gguauCACGguacc and ggcacUUCGgugcc (see Figure 1) by the AMBER all-atom force field[18,19] in explicit water, 30 ns REMD simulations using 48 replicas with temperatures between 297 and 495 K were performed. Combining a microscopic description of RNA hairpins[5,10,20−31] with the enhanced sampling qualities of REMD,[32−40] the approach allows us to study the conformational structure and dynamics as well as the thermal folding and unfolding of RNA hairpins in great detail.[27,41−43]

The paper is organized as follows. First, the structural features of both hairpins at 300 K are discussed and compared with the NMR and crystallographic results.[4,6−8,44] Then, conformation distributions of the uCACGg and the cUUCGg hairpin are compared. Finally, the REMD trajectories are used to describe at the microscopic level the thermostability of the hairpins and to give an insight into the thermal folding−unfolding energy landscape.

2. Methods

2.1. Simulation Conditions. The AMBER force field (parm98)[18,19] was employed to describe the 14-mer RNA hairpins. The hairpin was placed in a rhombic dodecahedron

* Corresponding author. Phone: +49-69-79829711. Fax: +49-69-79829709. E-mail: villa@theochem.uni-frankfurt.de.

10.1021/jp0764337 CCC: $40.75
Published on Web 12/11/2007

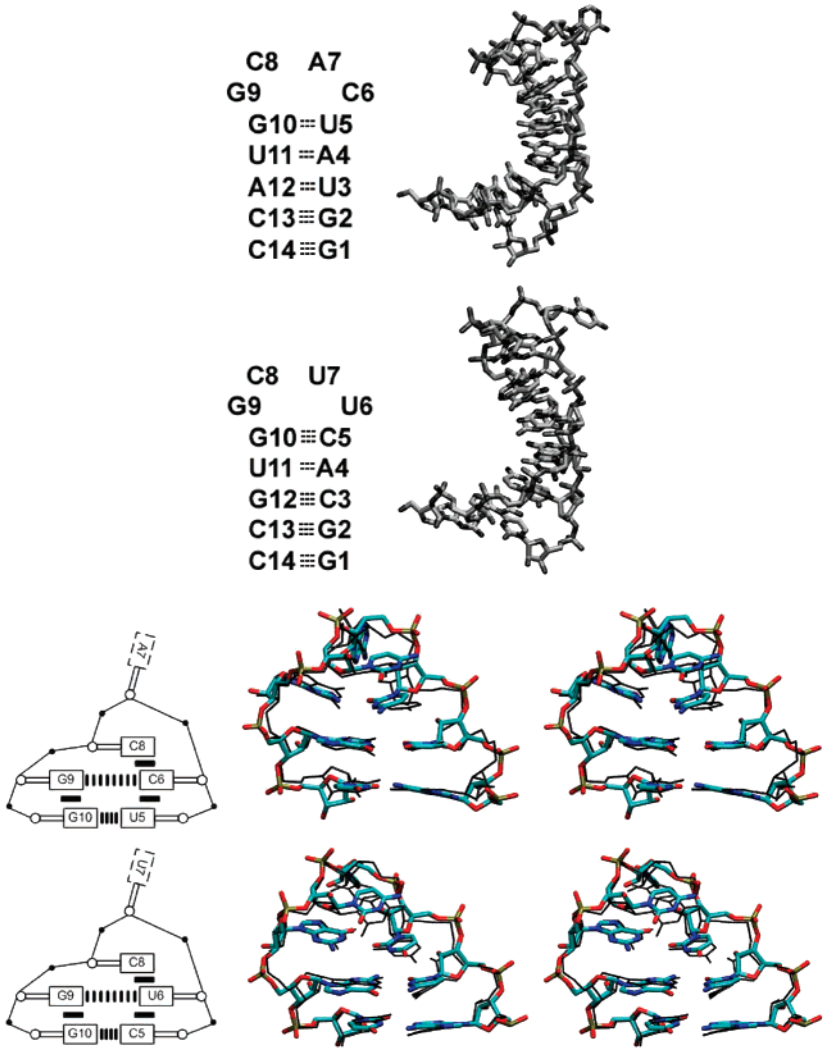

Figure 1. Top: The 14-mer RNA hairpins uCACGg (top) and cUUCGg (bottom) at 300 K, showing (left) secondary structures including stem base-pair hydrogen bonds and residue numbering and (right) representative MD snapshots at 20 ns. Bottom: Secondary structure and stereoview of the loop region of both hairpins, comparing average MD (color) to average NMR[4,7] (black) structure. In the picture of NMR structure, the phosphate oxygens have been omitted. The pictures were performed using the graphical package VMD.[76]

box (edge length approximately 5 nm), which was subsequently filled with about 2700 TIP3P water molecules.[45] To neutralize the system, 13 sodium ions were placed randomly in the simulation box. A twin range cutoff was used for the Lennard-Jones interactions; that is, interactions between atoms within 1.0 nm were evaluated every step, while interactions between atoms within 1.4 nm were evaluated every 5 steps. The particle mesh Ewald method[46] was employed to treat Coulomb interactions, using a switching distance of 1.0 nm, a grid of 0.12 nm, and a beta value of 3.1 nm^{-1}. Constant pressure p and temperature T were maintained by weakly coupling the system to an external bath at 1 bar and at the chosen temperature, using the Berendsen barostat and thermostat, respectively.[47] The RNA, the ions, and the solvent were independently coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 0.5 ps, and the isothermal compressibility was 4.5×10^{-5} bar^{-1}. The bond distances and the bond angle of the solvent water were constrained using the SETTLE algorithm.[48] All other bond distances were constrained using the LINCS algorithm.[49] A leapfrog integrator with a integration time step of 2 fs was used. The starting structure of the 14-mer uCACGg hairpin was taken from the Brookhaven Protein Data Bank (PDB structure 1RFR).[4] The cUUCGg hairpin was modeled using the crystallographic structure of Ennifar et al.[8] for the loop and a structure built by AMBER6 tools[50] for the stem. The structure was then equilibrated at 300 K for 20 ns. To study the effect of the water model on the unfolding behavior of RNA hairpins, we repeated the cUUCGg simulation using the TIP4P-Ew model.[51]

2.2. Replica-Exchange Molecular Dynamics. The replica-exchange molecular dynamics (REMD)[15,16] simulations were performed using the GROMACS suite of programs (version 3.3).[52−54] We choose 48 replicas with temperatures between 297 and 495 K and performed a 30 ns REMD simulation for each hairpin. We used the extension of the REMD to the isobaric−isothermical ensemble, recently proposed by Okabe and co-workers.[16] In this formulation, the exchange probability between replicas i and j is estimated as

$$P(i \leftrightarrow j) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_j} - \frac{1}{k_B T_i}\right)(U_i - U_j) + \Delta_V\right]\right) \quad (1)$$

where T_i and T_j are the replica's temperatures, k_B is Boltzmann's constant, U_i and U_j denote the potential energies of the replicas, and Δ_V takes into account the volume fluctuations; see ref 16 for details. To ensure a uniform exchange probability, the temperatures where chosen according to $T_j = (1 + 1/\sqrt{N_{df}})T_i$, where N_{df} denotes the total number of degrees of freedom.

Exchange between replicas was attempted every 20 ps. The temperatures have been selected to obtain exchange probabilities between 15 and 33% . On average, 1.05 folding−unfolding events are observed in 1 ns simulation time. We define the hairpin as folded when all native base-pair hydrogen bonds and base-stackings are present and as unfolded when none of the native interactions are observed. (Because of the size of the simulation box, no completely extended conformations were sampled.) The maximum number of events per nanosecond in one replica is 6.5 and 3.5 for the uCACGg and the cUUCGg hairpin, respectively. In 13 replicas over 48, the systems do not experience a complete folding−unfolding process during 30 ns REMD simulations. Those replicas performed at temperatures lower than 390 K. No water vaporization was observed up to 495 K, since common fixed-charge water models fail to capture vapor−liquid transitions.[55,56]

2.3. Trajectory Analysis. The analysis of the trajectories was performed with tools from the GROMACS package and with modified versions of them. To define the presence of an hydrogen bond, an acceptor−donor distance smaller than 0.35 nm was requested. Additionally, the number of observed hydrogen bonds did not change when a 150° cutoff for the acceptor−hydrogen−donor angle was applied. The solvent accessible surface was computed numerically.[57] The atomic radii used were 0.16 nm for carbon, 0.13 nm for oxygen, 0.14 nm for nitrogen, 0.20 nm for phosphor, and 0.10 nm for hydrogen.

To compare the conformational fluctuations of both hairpins, we performed a principal component analysis[58−60] of the merged trajectory of both hairpins at 300 K. Only backbone atoms were considered. Before performing the analysis, each conformation was translated and rotated to give the best fit to the averaged structure. The first three (of 249 in total) eigenvectors describe 64% of the total backbone fluctuations.

To define reaction coordinates for the thermal unfolding of the hairpins, we used stem base−base hydrogen bonds and base stacking interactions. In the first case, the presence of the N−H···N hydrogen bond between Watson−Crick base-pair and N−H···O hydrogen bonds between wobble base-pair is used as an indicator. A total number of five hydrogen bonds was used for uCACGg and cUUCGg hairpin (one each stem base-pair). In the second approach, two bases were considered as stacked when their center of mass separation is within 0.15 nm of that seen in the folded structure (simulation at 300 K). Pairs with larger separations are considered as broken.[27] We took into account eight stem base stackings, one stem-loop (5:6), and one inside-loop (6:8) base stacking interactions. Employing these coordinates, the Gibbs free energy is given by

$$\Delta G(n_H, n_S) = -k_B T[\ln P(n_H, n_S) - \ln P_{min}] \qquad (2)$$

where n_H and n_S are the number of selected hydrogen bonds and base stackings, respectively. $P(n_H, n_S)$ denotes the probability of finding a conformation with interactions (n_H, n_S), and P_{min} is the population probability of the global minimum with $\Delta G \equiv 0$.

The melting of the uCACGg and cUUCGg hairpins was monitored via the fraction $P_H = \langle n_H \rangle / n_H^{max}$ of hydrogen bonds between the stem base-pairs and the fraction $P_S = \langle n_S \rangle / n_S^{max}$ of base stacking interactions. To describe the correlation of two hydrogen bonds n and m, we calculated the normalized covariance matrix

$$\sigma_{nm} = \frac{\langle (P_{Hn} - \langle P_{Hn} \rangle) \cdot (P_{Hm} - \langle P_{Hm} \rangle) \rangle}{\sqrt{\langle (P_{Hn} - \langle P_{Hn} \rangle)^2 \rangle \cdot \langle (P_{Hm} - \langle P_{Hm} \rangle)^2 \rangle}} \qquad (3)$$

where P_{Hn} is 0 or 1, depending on if the nth hydrogen bond is broken or closed.

Furthermore, the temperature dependence of the folding process can be described by the folding free energy of the hairpin, which is given by

$$\Delta G_f = -k_B T[\ln P_f - \ln(1 - P_f)] \qquad (4)$$

where P_f is the probability of finding a conformation in the fold state. P_f was estimated by the percentage of structures that have more than seven stacking interactions which occur in the folded state at 300 K.

3. Results and Discussion

3.1. Characterization at 300 K. The structures sampled by the 30 ns REMD simulation at 300 K for uCACGg and cUUCGg hairpins are in overall agreement with the corresponding experimental structures;[4,6−8] see Figure 1. The root-mean-square deviation (rmsd) has been calculated for all backbone atoms of residues 4−11 after performing a least-squares fit of the REMD and NMR average structures. The simulated uCACGg loop shows a deviation of 0.12 and 0.10 nm with respect to the NMR structures of Ohlenschläger et al.[4] and of Du et al.,[6] respectively. A low rmsd value (0.06 nm) with respect to the NMR structures of Allain and Varani[7] is likewise observed for the cUUCGg tetraloop.

Similar to experiment, we find a large similarity of the structures of the two hairpins. As an illustration, Figure 1 shows representative snapshots of each hairpin at 20 ns. The residues forming the stem are all involved in Watson−Crick base-pairs and stacking interactions. In the uCACGg hairpin, the closing residues form a wobble base pair. Both loops are mainly stabilized by hydrogen bonds between residues C6/U6 and G9 and residues U7/A7 and C8, as well as by the stacking between residue 6 and residues 5 and/or 8. The residue in position 7 is looped out, and the residue C8 is unpaired. All bases are in anti conformation, except for G9 and A7 which show a syn conformation. Experimentally, the syn conformation is observed only for the residue G9. Furthermore, C3′-endo state is observed for all residues, except for the residues in positions 7 and 8 which are in C2′-endo, as also observed in experiment.

It is instructive to compare the hydrogen bond network observed for both loops to the experimental evidence.[4,6−8] In the uCACGg simulation, the C6 2′-hydroxyl, the C8 amino, and the G9 imino hydrogens are all involved in hydrogen bonding, in agreement with NMR experiments. The experiments indeed show that these protons are protected from rapid exchange.[6] In particular, residues C6 and G9 are involved in base−base hydrogen bonds, and there is a hydrogen bond between the C8 base and the A7 phosphate oxygen as observed in the experimental structures. The calculated and NMR refined structures differ only in the C6 2′-hydroxyl hydrogen. This hydrogen is involved in an intra-residue hydrogen bond with other hydroxyl groups in the MD trajectory, while it is hydrogen-bonded to the residue G9 in NMR refinement. In the UUCG loop, the residues U6 and G9 are involved in base−base and in base−sugar hydrogen bonds. The U7 phosphate oxygen is hydrogen bonded as observed in the experimental structures.[7,8] Moreover, a weak interaction is observed between the 2′-OH group of U7 and the base G9. In the crystallographic structure,[8] the U7 sugar oxygen is hydrogen bonded to the G9 base oxygen, while this is not the case in the NMR structures. Calculating the average probabilities of all hydrogen bonds in the loops,

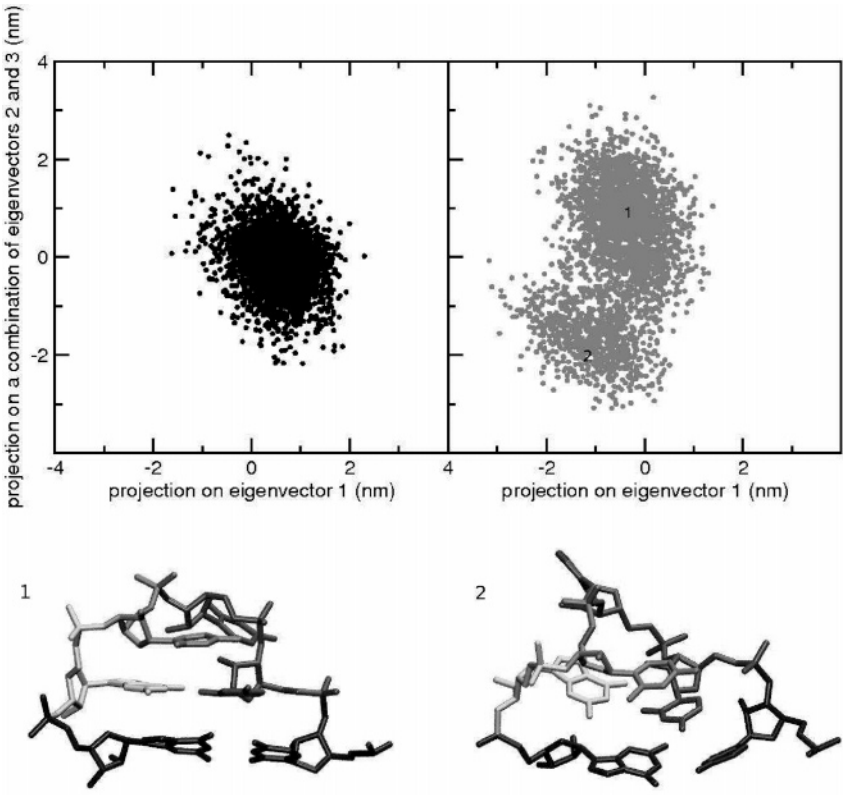

Figure 2. Top: Conformations of the uCACGg (left, black) and cUUCGg (right, gray) hairpins at 300 K, projected on the first eigenvectors of a principal component analysis. Bottom: Representative MD snap shots of the two conformational states of the UUCG loop.

we find 2.5 versus 3.5 hydrogen bonds for the CACG and the UUCG loop, respectively.

It is interesting to study to what extent the above-discussed structural features are reflected in the conformational fluctuations of both hairpins. To this end, we have performed a principal component analysis[58−60] of the merged trajectory of both hairpins at 300 K (see Methods) and considered the projection of each individual hairpin trajectory on the first three eigenvectors of the principal component analysis. Figure 2 shows this projection as a function of the first and the sum of the second and third eigenvectors. Despite the differences in the sequence, the trajectories of the two hairpins show large overlapping regions. However, for the cUUCGg hairpin, we also observe a separated smaller region which indicates a second conformation state with a population of 23% . In agreement with previous MD studies,[21,29] an analysis of the calculated backbone dihedral angles indeed reveals a conformational rearrangement in the UUCG loop region. As an illustration, Figure 2 at the bottom shows the loop structure of these two conformations of the cUUCGg hairpin.

To identify the origin of the difference between the conformational distributions of the hairpins, we considered the atomic root-mean-square fluctuation (rmsf) of the two RNAs with respect to their average structures. Figure 3 shows that the main difference of the atomic fluctuations are observed in the loop region; that is, the UUCG loop is more flexible than the CACG loop. In particular, the atoms around the sixth phosphorus atom deviate significantly from the average structure. This is mainly related to the backbone dihedral angle transitions and the C3′-endo/C2′-endo equilibrium of the sugar ring, observed for residues 6 and 7 in cUUCGg hairpin, while the uCACGg always shows single values for these quantities. Recalling that the UUCG loop contains more stabilizing hydrogen bonds, we have the intriguing situation that the loop with more hydrogen bonds is also the one with higher flexibility.

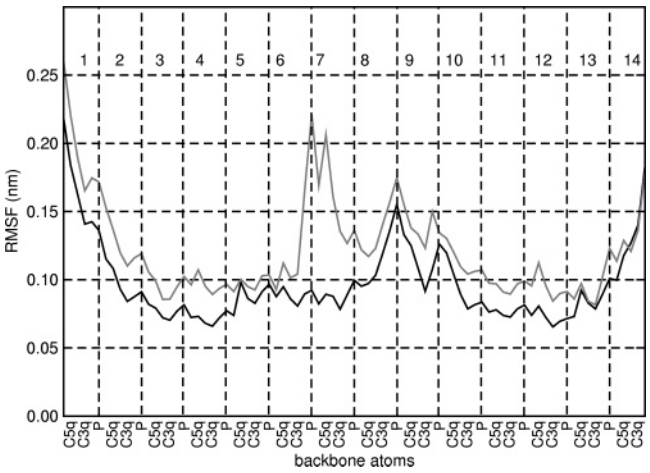

Figure 3. Root-mean-square fluctuation, rmsf, of the backbone atoms of the uCACGg (black) and cUUCGg (gray) hairpins at 300 K.

The presence of looped-out residues and potential hydrogen bond donor and acceptor groups are usually regarded as structural features that affect the functionality of an RNA hairpin, such as its attitude to bind other RNA or protein molecules. Indeed, both investigated hairpins have a looped-out residue in position 7, but the base of this residue adopts a different orientation relative to the sugar moiety in the two hairpins. A7 is in syn conformation, where the dihedral angle χ between the sugar and the base moiety holds values of $-51°$ (72%) and 29° (28%) during the simulation. This allows the amino group to be fully accessible to the solvent. On the other hand, U7 is in an equilibrium between anti (80%) and syn (20%) conformations and has two hydrogen bond acceptors on the surface. As there are no syn conformations observed in experimental structures, it is at present unclear if this finding is correct or an artifact of the used force field. The total solvent

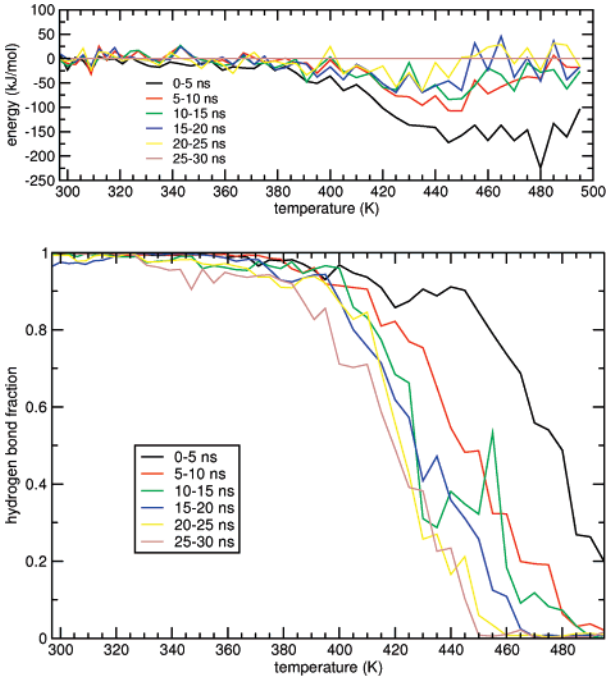

Figure 4. Convergence of the REMD simulations of the cUUCGg hairpin, monitored for six different time intervals. Shown are (above) the mean potential energy (in kJ/mol, relative to the energy of the last 5 ns of the simulation) and (below) the fraction of base−base hydrogen bonds.

accessible surface of the two RNAs is quite similar, 27.9 nm^2 and 27.5 nm^2 for uCACGg and cUUCGg, respectively. In uCACGg, residues 5, 7, and 9 are slightly more exposed and residues 6 and 8 are slightly less exposed than in cUUCGg structures. While the global solvent accessible surface shows only a minor difference, the type of functional groups on the surface are quite different. In the simulation of uCACGg, the amino group of A7, the carbonyl groups of C8 and G9, and a C−H group of G9 were fully accessible to the solvent. In the simulation of cUUCGg, the carbonyl groups of U6, U7, and C8, as well as the C−H groups of U7 and G9, were exposed to the water. Hence, the analysis indicates a stronger attitude to donate hydrogens for the uCACGg hairpin than for the cUUCGg hairpin.

3.2. Thermal Unfolding. Providing the free energy landscape at all temperatures of interest, REMD simulations are an ideal means to study the folding and unfolding of biomolecules.[35,37,41,61] Let us first study the convergence of the REMD calculations with respect to the simulation time. As a representative example, Figure 4 shows the mean potential energy as well as the fraction of base−base hydrogen bonds at six different time intervals of the cUUCGg simulation. Since all replicas start with the equilibrium structure at 300 K, the first few nanoseconds of the high-temperature replicas clearly exhibit too low energies (by up to ≈100 kJ/mol) compared with later times, when all replicas are equilibrated according to their temperature. Similarly, the fraction of base−base hydrogen bonds is found to shift significantly during the first 10 ns of the simulation. In the discussion below, we therefore use only the last 20 ns of the REMD simulations of both hairpins. On average, we observe one folding−unfolding event per nanosecond simulation time for each replica, resulting in a total number of 10^3 such events. Those include both folding and unfolding events. Monitoring the diffusion of the replicas in temperature space, we found that most (73%) replicas travel at least one time through the complete temperature range (as an example, see Figure 7).

The melting of the uCACGg and cUUCGg hairpins was monitored via the fraction P_H of hydrogen bonds between the stem base-pairs and by the fraction P_S of base stacking interactions; see Methods. Plotted as a function of temperature, Figure 5 shows these observables which run from 1 (folded state) to 0 (unfolded state). Defining the melting temperature T_m by $P_H(T_m) \equiv 0.5$, we obtain $T_m \approx 392$ and 425 K for the uCACGg and the cUUCGg hairpin, respectively. Employing the criterion $P_S \equiv 0.5$, the melting temperature increases by about 14 (uCACGg) and 4 K(cUUCGg). Averaging over both criteria, we obtain a computational melting temperature difference of ≈28 K. Although there are no experimental data available for the two 14-mer hairpins under consideration, various experimental studies on similar RNA hairpin suggest that the melting temperatures are ≈350 and 330 K for the uCACGg and the cUUCGg hairpin, respectively.[12] While the experimental difference in melting temperatures $\Delta T_m \approx 20$ K is reproduced by the REMD simulations within statistical and experimental uncertainties, the absolute melting temperatures are about 20% too high.

This significant discrepancy between measured and simulated transition temperatures is well-known from numerous peptide folding studies.[35,37,41,61] Besides sampling problems (see above), it is most likely related to deficiencies of the force field, since standard biomolecular force fields have been parametrized to reproduce properties at room temperature. Apart from the modeling of the solute, the high-temperature description of the aqueous solvent is a well-known problem. For example, it is known that the isobaric thermal expansion of TIP3P water is a factor of 3 higher than in experiment, and its temperature derivative is a factor of 2 lower.[62] This quantity is a measure of the enthalpic contribution to the reorganization of the solvent and is important in the solvation of hydrophobic residues. To study the effect of the water model on the unfolding behavior of RNA hairpins, we repeated the cUUCGg simulation using the TIP4P-Ew model.[51] This water model was parametrized to reproduce both experimental density and enthalpies of vaporization at different temperatures ranking from 235.5 to 400 K, and therefore describes well both kinetic and thermodynamic properties, such as thermal expansion, heat capacity, self-diffusion coefficient, and compressibility. Nonetheless, Figure 5 reveals that the melting curves obtained for the TIP4P-Ew model are within statistical errors equivalent to the results for the TIP3P model. Apparently, the discrepancy between measured and simulated transition temperatures cannot directly be attributed to the mentioned thermodynamic properties of the water model. Another possibility that could increase the calculated melting temperature concerns the description of the sodium concentration in the simulation, which is a factor of 20 higher than in experiment.[63,64]

As a further means to estimate the quality of our simulations, we calculated the folding free energy ΔG_f from the probability that the hairpin is in the folded state, using a native stacking criterion (see Methods). We obtained ΔG_f of −23 and −21 kJ/mol at 300 K, −12 and −19 kJ/mol at 328 K, and 0 and −8 kJ/mol at 393 K for the uCACGg and the cUUCGg hairpin, respectively. Although there are no experimental data available for the two 14-mer hairpins under consideration, experiments on similar cUUCGg hairpin suggest values around −20 kJ/mol at 310 K and −10 kJ/mol at 328 K for cUUCGg.[11,63−65] When changing the loop sequence to CACG, the free energy increases by 3 kJ/mol at 328 K.[11] As discussed above, the simulation temperature needs to be scaled by a factor 0.2 to be comparable with the experimental temperature; that is, the experimental

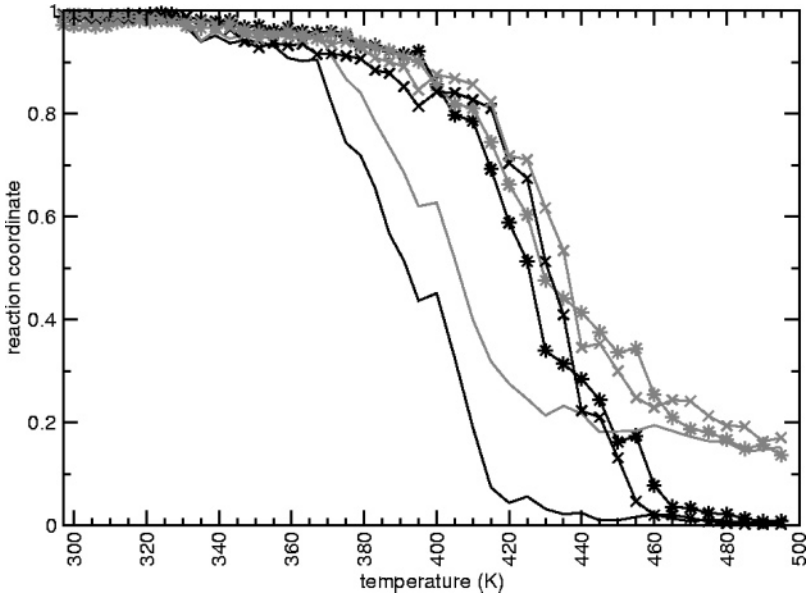

Figure 5. Melting curves of the uCACGg (no symbol) and the cUUCGg (star) hairpin, using the fraction of hydrogen bonds between the stem base-pairs (black) and the fraction of base stacking interactions (gray) as reaction coordinates. Also shown are results for cUUCGg obtained for the TIP4P-Ew water model (X).

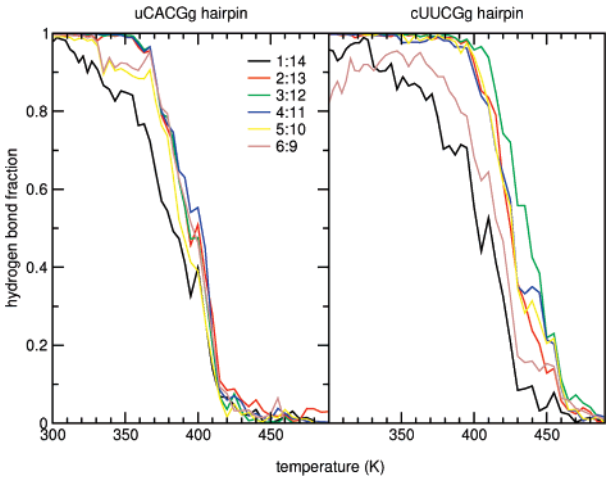

Figure 6. Melting curves of the selected hydrogen bonds between individual base pairs of the two 14-mer RNA hairpins. Shown are temperature-dependent hydrogen-bond probabilities of the stem and loop base-pairs.

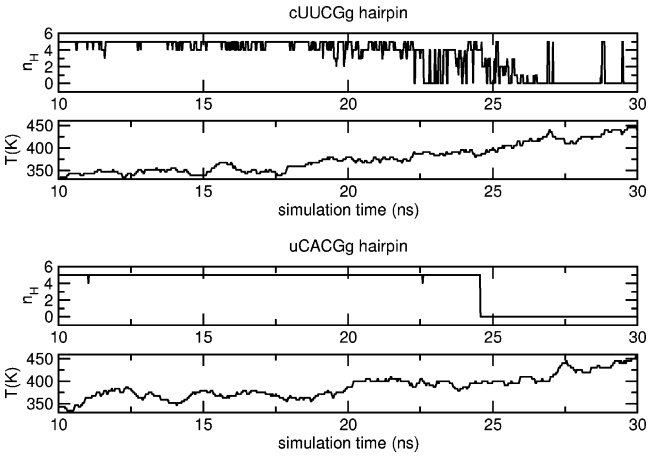

Figure 7. Time evolution of the number $n_H(t)$ of selected hydrogen bonds together with the instantaneous temperature, T, of a representative replica, obtained from the REMD simulation of the uCACGg and the cUUCGg hairpin, respectively.

temperature of 328 K corresponds to a simulation temperature of 393 K. Under this assumption, the calculated free energy for cUUCGg nicely agrees with experimental data, while the calculated values for uCACGg are 7 kJ/mol higher compared with those of experiment on a cCACGg hairpin. The latter finding seems to indicate that changes in the loop sequence are not enough to explain the free energy difference between the two investigated hairpins.[11,65−67]

After checking the quality of the force field and the sampling, we are now in a position to investigate the unfolding mechanism of the two RNA hairpins. Apart from the lower transition temperature, Figure 5 reveals that the width of the melting curve of the uCACGg hairpin is smaller than for the cUUCGg hairpin. Defining the width as $\Delta T = T(P_H = 0.9) - T(P_H = 0.1)$, we obtain $\Delta T = 50$ and 65 K for the uCACGg and the cUUCGg hairpin, respectively. This finding indicates that the thermal folding−unfolding process of the uCACGg hairpin is more cooperative than for the cUUCGg hairpin.

The phenomenon is studied in more detail in Figure 6 which shows the temperature-dependent probability of the hydrogen bonds between the individual base pairs of the stem. Apart from the first base pair G1:C14, which partially opens already at room temperature, the remaining stem hydrogen bonds are found to open in a concerted manner. A notable exception is the G12:C3 hydrogen bond of cUUCGg, which clearly remains more stable than the other base pairs of the stem. A further difference between the two hairpins is found in the melting curve of the hydrogen bond between the loop residues 6 and 9. While the base pair C6:G9 of the uCACGg loop appears to open along with the stem, the U6:G9 hydrogen bond in the UUCG loop is less cooperative and also less stable than the stem hydrogen bonds. Moreover, some premelting involving the closing (U5:G10) and the loop (C6:G9) base pairs was observed for the uCACGg hairpin.

We also considered the possibility of non-native base-pair interactions in both hairpins. Although we found several hairpin conformations with a single non-native interaction, their populations were too small to be seen in the free energy landscape. This observation might support the idea of nonspecific intrachain interactions prior to nucleation, shown in previous kinetic studies.[68−70]

TABLE 1: Correlation σ_{nm} between Two Hydrogen Bonds *n* and *m*, as Defined in Eq 3[a]

	σ_{nm} (cUUCGg)					σ_{nm} (uCACGg)				
n/*m*	(1:14)	(2:13)	(3:12)	(4:11)	(5:10)	(1:14)	(2:13)	(3:12)	(4:11)	(5:10)
(1:14)	1.00	0.56	0.44	0.51	0.51	1.00	0.69	0.64	0.60	0.58
(2:13)		1.00	0.75	0.68	0.62		1.00	0.87	0.81	0.79
(3:12)			1.00	0.76	0.70			1.00	0.84	0.82
(4:11)				1.00	0.76				1.00	0.78
(5:10)					1.00					1.00

[a] Considered are the hydrogen bonds associated with the base pair (i:j) of the stem of the cUUCGg and the uCACGg hairpin, respectively. Since $\sigma_{nm} = \sigma_{mn}$, only the upper part of the matrix is shown.

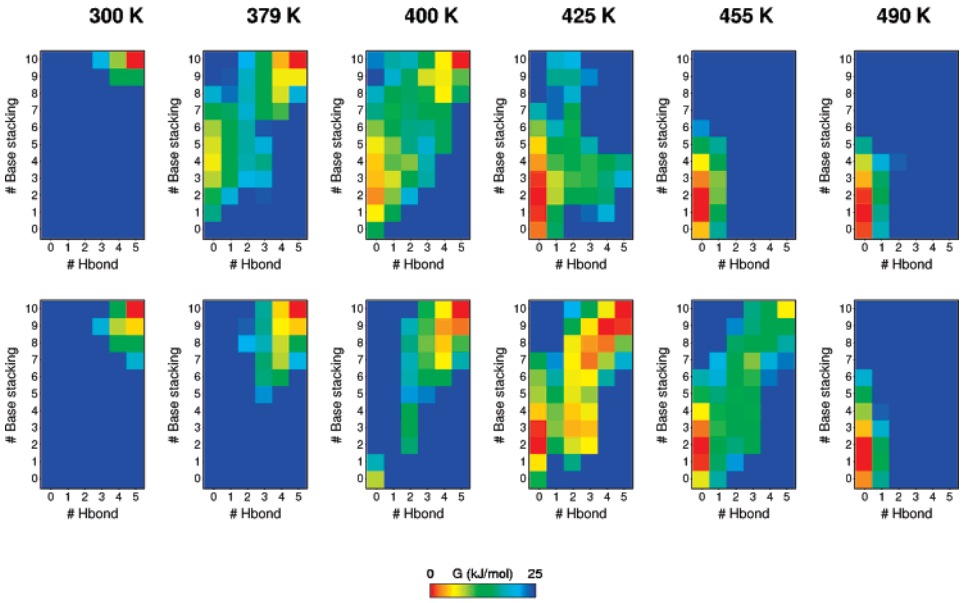

Figure 8. Free energy landscape $\Delta G(n_H, n_S)$ (in kiloJoules per mole) of the uCACGg (top) and the cUUCGg (bottom) hairpin at various temperatures, shown as a function of the number of base−base hydrogen bonds n_H and base stacking interactions n_S.

To assess the cooperativity in the making and breaking of base-pair hydrogen bonds, we have calculated the covariance matrix σ_{nm} (eq 3), which describes the correlation between two hydrogen bonds *n* and *m*. We have restricted the calculation to temperatures in the melting region (370−420 K and 400−465 K for uCACGg and cUUCGg, respectively) in order to focus on the melting process. In agreement with the above observations, Table 1 reveals (i) that the hydrogen-bonding correlations are generally higher for the uCACGg than for the cUUCGg hairpin, (ii) that the lowest correlation is found between the first (G1:C14) and the remaining base pairs and that the making or breaking of the G12:C3 hydrogen bond of cUUCGg is somewhat less correlated with the dynamics of the other base pairs of the stem.

As an illustration of this cooperativity, we may consider the time evolution of the stem hydrogen bonds along some chosen replica of the REMD simulation. By displaying the number $n_H(t)$ of selected stem hydrogen bonds together with the instantaneous temperature of the replica, Figure 7 shows that the rupture of hydrogen bonds occurs collectively in the case of the uCACGg hairpin and via some intermediate states in the case of the cUUCGg hairpin. In both cases, the folded state is characterized by five (or occasionally four) stem hydrogen bonds. With increasing temperature, the uCACGg hairpin unfolds in an all-or-none fashion, while the cUUCGg hairpin exhibits intermediate states characterized by 1−3 base-pair hydrogen bonds.

Let us finally study the energy landscape obtained for the two RNA hairpins. Figure 8 shows the free energy $\Delta G(n_H, n_S)$ of both loops as a function of the number of base−base hydrogen bonds n_H and base stacking interactions n_S; see Methods. The energy landscapes are displayed for six different temperatures, ranging from 300 to 490 K. At 300 K, both hairpins are completely in the folded state (with $n_H \approx 5$, $n_S \approx 10$) while at 490 K the systems are unfolded (with $n_H \approx 0$, $n_S \approx 1$). Along the two coordinates under consideration, the uCACGg hairpin exhibits a simple two-state behavior in the complete temperature range. The energy landscapes of the cUUCGg hairpin, on the other hand, is more complex and exhibits at least one intermediate state at the transition temperature of 425 K. Located along the diagonal of the $\Delta G(n_H, n_S)$ energy surface, the intermediate states are characterized by $n_H = 2-3$ hydrogen bonds and $n_S = 3-4$ stacking interactions. Figure 9 shows a representative snapshot of the intermediate state at 425 K. In all intermediates, the central stem residues are still involved in native base-pair and/or stacking interactions, while the global helicity of the hairpin is lost.

To estimate the role of the different loop sequence of the two hairpins, we have applied the popular model of Zuker which employs the Turner parameters to calculate the melting profiles of the systems.[14,71] For the uCACGg and the cUUCGg hairpins, the model predicts the correct difference in melting temperature of about 25 K. Changing the loop sequence of the uCACGg hairpin to uUUCGg results in an increase of the melting temperature of 15 K, while changing the loop sequence of the cUUCGg hairpin to cCACGg results in an decrease of the melting temperature of 10 K. This indicates that the differences in the loops may as well affect the measured difference of melting temperatures of the two hairpins under consideration.

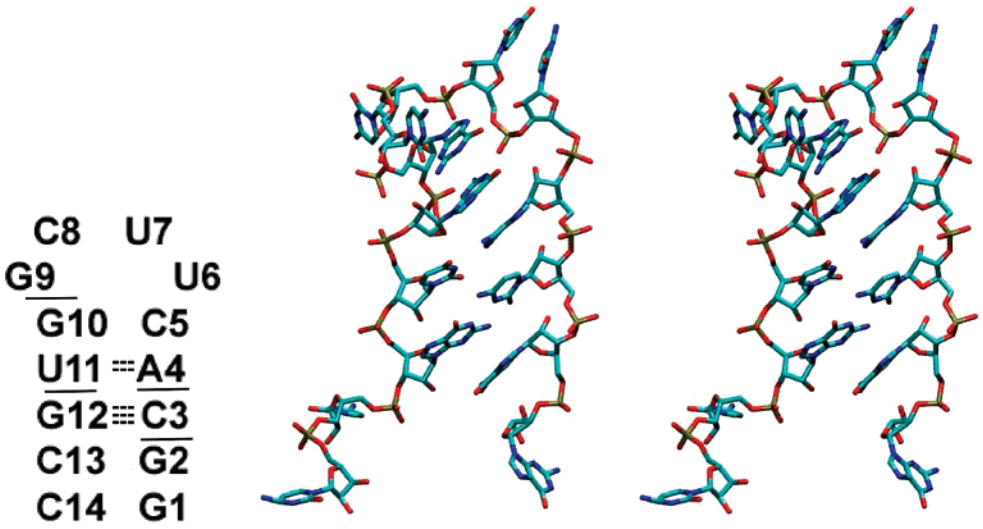

Figure 9. Representative structure of the intermediate state of the cUUCGg hairpin at 425 K. Left: Secondary structures indicating native hydrogen bonds and stacking interactions. Right: Stereoview of the structure.

In summary, we can draw the following picture of the thermal unfolding of the two RNA hairpins. The REMD study qualitatively reproduces the experimentally measured difference of melting temperatures $\Delta T_m \approx 20$ K. In the case of the uCACGg hairpin, the thermal unfolding occurs cooperatively in an all-or-none fashion, thus resulting in a simple two-state behavior.[41,43,67,72−74] The cUUCGg hairpin, on the other hand, shows less cooperativity but exhibits intermediate states in the unfolding process.[27,66,68,75] While the global helicity of the hairpin is lost in these states, base pairs at stem middle are still involved in native interactions. Interestingly, recent temperature-jump experiments on an 8-mer cUUCGg hairpin were interpreted by a four-state energy landscape.[75]

4. Conclusions

Extensive REMD simulations (≈4.3 μs total simulation time) have been performed to study structure, dynamics, and melting of the structurally similar 14-mer RNA hairpins uCACGg and cUUCGg. The simulations have confirmed the experimentally found structural similarities of the two RNA hairpins at room temperature. In particular, the hydrogen bond network and base stacking interactions are quite similar for both systems. Studying the conformational fluctuations, it has been found that the cUUCGg loop is more flexible than the uCACGg loop. Indeed, residues 6 and 7 in cUUCGg hairpin are involved in backbone dihedral angle transitions and C3′-endo/C2′-endo sugar equilibrium. As the cUUCGg loop contains more stabilizing hydrogen bonds, we have the intriguing situation that the loop with more hydrogen bonds is also the one with higher flexibility.

To understand the experimentally found differences in the binding behavior of the RNA hairpins, we have studied the relevance of the looped-out residue in position 7 and of potential hydrogen bond donor and acceptor groups. While the global solvent accessible surface was found to be quite similar for both RNA loops, the type of functional groups on the surface were quite different. In particular, the analysis has indicated a stronger attitude to donate hydrogens for the uCACGg hairpin than for the cUUCGg hairpin.

Providing the free energy landscape at all temperatures of interest, REMD simulations are an ideal means to study the folding and unfolding of biomolecules. Although the calculated absolute melting temperatures are about 20% too high, the REMD simulations reproduces the experimentally found difference in melting temperatures of $\Delta T_m \approx 20$ K within statistical and experimental uncertainties. In the case of the uCACGg hairpin, the thermal unfolding occurs cooperatively in an all-or-none fashion, thus resulting in a simple two-state behavior. The cUUCGg hairpin, on the other hand, shows less cooperativity, but exhibits intermediate states in the unfolding process.[27,68,66,75] The joint computational and experimental identification and characterization of intermediate states in the folding and unfolding of biomolecules has just yet become feasible.

Acknowledgment. The authors thank Harald Schwalbe and Jens Wöhnert for numerous inspiring and helpful discussions. This work has been supported by the Frankfurt Center for Scientific Computing, the Fonds der Chemischen Industrie, and the Deutsche Forschungsgemeinschaft (SFB 579 "RNA-ligand interactions"). E.W. thanks the DAAD for financial support.

References and Notes

(1) Andino, R.; Rieckhof, G. E.; Baltimore, D. *Cell* **1990**, *63*, 369−380.
(2) Rieder, E.; Xiang, W.; Paul, A.; Wimmer, E. *J. Gen. Virol.* **2003**, *84*, 2203−2216.
(3) Zell, R.; Sidigi, K.; Stelzner, A.; Görlach, M. *RNA* **2002**, 8, 188−201.
(4) Ohlenschläger, O.; Wöhnert, J.; Bucci, E.; Seitz, S.; Häfner, S.; Ramachandran, R.; Zell, R.; Görlach, M. *Structure* **2004**, 12, 237−248.
(5) Ihle, Y.; Ohlenschläger, O.; Häfne, S.; Duchardt, E.; Zacharias, M.; Seitz, S.; Zell, R.; Ramachandran, R.; Görlach, M. *Nucleic Acids Res.* **2005**, *33*, 2003−2011.
(6) Du, Z.; Yu, J.; Andino, R.; James, T. L. *Biochemistry* **2003**, 42, 4373−4383.
(7) Allain, F.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333−353.
(8) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35−42.
(9) Woese, C.; Winker, S.; Gutell, R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467−8471.
(10) Williams, J.; Hall, K. *J. Mol. Biol.* **2000**, *297*, 1045−1061.
(11) Proctor, D. J.; Schaak, J. E.; Bevilacqua, J. M.; Falzone, C. J.; Bevilacqua, P. C. *Biochemistry* **2003**, *41*, 12062−12075.
(12) The melting temperature of a 12-nucleotide cUUCGg hairpin was reported as 346 K.[11] A 38-mer uCACGg loop of the coxsackievirus was observed to melt at about 338 K,[6] while the broadening of the NMR imino proton resonance of a 30-mer uCACGg loop was already recorded at 318 K[4]
(13) Saenger, W. *Principles of nucleic acid structure;* Springer-Verlag: New York, 1988.
(14) Walter, A. E.; Turner, D. H.; Kim, J.; Lyttle, M. H.; P. Müller, D. H. M.; Zucker, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9218−9222.

(15) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(16) Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. *Chem. Phys. Lett.* **2001**, *335*, 435−439.
(17) Frenkel, D.; Smit, B. *Understanding molecular simulations: from algorithms to applications;* Academic Press: London, 2002.
(18) Cheatham, T.; Cieplak, P.; Kollman, P. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−861.
(19) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(20) Zichi, D. A. *J. Am. Chem. Soc.* **1995**, *117*, 2957−2969.
(21) Miller, J.; Kollman, P. *J. Mol. Biol.* **1997**, *270*, 436−450.
(22) Hermann, T.; Auffinger, P.; Westhof, E. *Eur. Biophys. J.* **1998**, 27, 153−165.
(23) Williams, J.; Hall, K. *Biophys. J.* **1999**, *76*, 3192−3205.
(24) Zacharias, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311−317.
(25) Sarzynska, J.; Kulinski, T.; Nilsson, L. *Biophys. J.* **2000**, 79, 1213−1227.
(26) Nina, M.; Simonson, T. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.
(27) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493−506.
(28) Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. *Structure* **2005**, *13*, 1255−1267.
(29) Villa, A.; Stock, G. *J. Chem. Theory Comput.* **2006**, *2*, 1228−1236.
(30) Špačková, N.; Šponer, J. *Nucleic Acids Res.* **2006**, *34*, 697−708.
(31) Deng, N.; Cieplak, P. *J. Chem. Theory Comput.* **2007**, *3*, 1435−1450.
(32) García, A. E.; Sanbonmatsu, K. Y. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345−354.
(33) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931−14936.
(34) Sanbonmatsu, K. Y.; García, A. E. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225−234.
(35) García, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898−13903.
(36) Rao, F.; Caflisch, A. *J. Chem. Phys.* **2003**, *119*, 4035−4042.
(37) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.
(38) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 445−786.
(39) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354*, 173−183.
(40) Sanbonmatsu, K. Y. *Biochimie* **2006**, *88*, 1053−1059.
(41) Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. *Biophys. J.* **2003**, *85*, 790−803.
(42) Sorin, E. J.; Nakatani, B. T.; Rhee, Y. M.; Jayachandran, G.; Vishal, V.; Pande, V. S. *J. Mol. Biol.* **2004**, *337*, 789−797.
(43) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2516−2524.
(44) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481−1491.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(46) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(47) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(48) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(49) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(50) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 6;* University of California: San Francisco, CA, 1999.
(51) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(52) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(53) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(54) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(55) Walser, R.; Mark, A. E.; van Gunsteren, W. F. *Biophys. J.* **2000**, *78*, 2752−2760.
(56) Zahn, D. *Phys. Rev. Lett.* **2004**, *93*, 227801.
(57) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. *J. Comput. Chem.* **1995**, *16*, 273−284.
(58) Ichiye, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1991**, 11, 205−217.
(59) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.
(60) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.
(61) Pande, V.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 9062−9067.
(62) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616−17626.
(63) Antao, V. P.; Lai, S. Y.; Tinoco, I., Jr. *Nucleic Acids Res.* **1991**, *19*, 5901−5905.
(64) Antao, V. P.; Tinoco, I., Jr. *Nucleic Acids Res.* **1992**, *20*, 819−824.
(65) Dale, T.; Smith, R.; Serra, M. J. *RNA* **2000**, *6*, 608−615.
(66) Chen, S.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646−651.
(67) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8602−8606.
(68) Ansari, A.; Kuznetsov, S. V.; Shen, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7771−7776.
(69) Ansari, A.; Kuznetsov, S. V. *J. Phys. Chem. B* **2005**, *109*, 12982−12989.
(70) Wallace, M. I.; Ying, L.; Balasubramanian, S.; Klenerman, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5584−5589.
(71) Markham, N. R.; Zucker, M. *Nucleic Acids Res.* **2005**, *33*, W577−W581.
(72) Nivón, L. G.; Shakhnovich, E. I. *J. Mol. Biol.* **2004**, *344*, 29−45.
(73) Wilson, K. S.; von Hippel, P. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8793−8796.
(74) Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 6789−6794.
(75) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530.
(76) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1990**, *14*, 33−38.