JCTC Journal of Chemical Theory and Computation

Conformational Transition Map of an RNA GCAA Tetraloop Explored by Replica-Exchange Molecular Dynamics Simulation

Yufen Zhang,[†,‡] Xian Zhao,*[,‡] and Yuguang Mu*[,†]

School of Biological Sciences, Nanyang Technological University, Singapore 637551, and State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, P.R. China

Received October 9, 2008

Abstract: A 120 ns replica-exchange molecular dynamics simulation in explicit solvent is performed to probe the conformational transitions in 5′-GGGCGCAAGCCU-3′ RNA GCAA tetraloop. The ample structural transition information of the loop is detected on the basis of extensive clustering analysis. The resultant loop structural transition map nicely agrees with the recent ultrafast fluorescence measurement, which confirms the dynamical properties of this tetraloop. Moreover, a new transition pattern that was not disclosed previously is predicted. Meanwhile, the folding free energy landscapes were characterized: the global folding dynamics is coupled mainly with the stem rather than the loop part.

Introduction

Hairpins are elementary structural units responsible for RNA folding.[1,2] Hairpin loop contains a base-paired stem structure and a loop sequence with unpaired nucleotides. Its most obvious property is to function as a "bender" to reverse the direction of backbone. Due to steric hindrance, there exists a minimum of three nucleotides to make a loop structure. However, loops with four nucleotides, known as tetraloops, are found to be much populated.[1,3] Among the four-base tetraloop motif, the family of GNRA (where N is any nucleotide and R is a purine) tetraloops is well structured with unusual stability.[4−6]

The most powerful tools to explore the structure features of RNA tetraloops are crystallography and nuclear magnetic resonance measurements.[5,7] These structural biology tools usually provide a well-defined structure or a structural ensemble with limited fluctuations. With the newly developed techniques, the dynamical features of RNA molecules are attracting attention.[8−10] For example, ^{13}C NMR relaxation measurements discovered substantial dynamic fluctuations in the loop regions of several tetraloops.[11−13] Dynamical properties or structural heterogeneity of the RNA loop can also be resolved by fluorescence spectroscopy.[14,15] For instance, a GAAA tetraloop that was substituted with 2-aminopurine residues and followed by fluorescence-detected temperature-jump relaxation analysis demonstrated the existence of more than a single conformation state with different base stacking patterns in the loop.[14] By incorporating both 2-aminopurine and 7-deazaguanine residues into similar GNRA tetraloops, another group studied the heterogeneity of loop conformation by femtosecond time-resolved fluorescence.[16] What they found not only confirmed the previous observation,[14] but with more position-specific fluorescence decay data, a more detailed dynamic multiconformation model for the tetraloop was proposed.

Molecular dynamics simulations are another powerful tool to explore the conformational dynamics of RNA tetraloops;[17−27] for example, the influence of base substitutions on the stability of GCAA tetraloop has been studied by the free energy perturbation method.[28,29] Most of the theoretical studies were targeted at the folding/unfolding dynamics of short RNA loops. The common features resolved by the modeling studies affirmed the hierarchical properties of folding free energy landscapes and general heterogeneity of loop conformation. However, detailed analysis of the conformational dynamics near the native-structure local mini-

* Corresponding author e-mail: YGMu@ntu.edu.sg (Y.M.) or zhaoxian@icm.sdu.edu.cn (X.Z.).

[†] Nanyang Technological University.

[‡] Shandong University.

10.1021/ct8004276 CCC: $40.75
Published on Web 03/26/2009

mum and direct comparison with available fluorescence experimental data are scarce.

In this paper, the structural transition mechanism of RNA tetraloop near the native-structure minimum at atomic detail was interrogated by a replica exchange molecular dynamics (REMD) simulation in explicit solvent for a GCAA RNA tetraloop. In order to directly compare with experimental data, the alternative stacking patterns of loop residues were monitored. The REMD simulation can overcome the sampling limitations of standard MD methods.[30−32] During REMD simulation, several replicas of a system are simulated at different temperatures in parallel, allowing for exchanges between neighboring replicas at frequent intervals.[30−33] So the REMD simulation can significantly enhance the conformational sampling. This technique has been successfully used for simulations of hairpin loop structure.[22,25,26] The key component in replica exchange simulations is the exchange of configurations between different replicas/temperatures by rescaling the velocities.[34] Such an algorithm helps to overcome large energy barriers and allows large conformational space to be sampled. Meanwhile, it maintains the continuous transformation of structures.[30,35] Together with extensive clustering analysis, the ample structural transition information of the tetraloop was detected. The resultant structural evolution map was able to directly compare with available experimental data, which confirms the predictive power of the current theoretical model.

Methods

The GROMACS program suite[36] and the full atomic Amber parm98 force field[37] were used. With Amber parm98 force field, the dynamics of a UUUU tetraloop RNA was studied by one of us[19] and a good agreement between MD result and NMR relaxation measurement was obtained. To test the reliability of parm98 further, a new Amber force field, parmbsc0,[38] was also employed to perform a constant temperature simulation. All bonds involving hydrogen atoms were constrained in length according to LINCS protocol.[39] This allowed the use of an integration step of 0.002 ps in simulations. Nonbonded pair lists were updated every 5 integration steps.[40] The RNA and the water were separately coupled to an external heat bath with a relaxation time of 0.1 ps. Replica exchange was attempted every 1000 integration steps (2 ps). The trajectories were output every 1 ps. Electrostatic interactions were treated with particle mesh Ewald method with a cutoff of 9 Å, and a cutoff of 14 Å was used in the calculation of van der Waals interactions.

We studied the GCAA tetraloop structure with the sequence of 12-mer single-strand RNA: 5′-GGGCGCAAGCCU-3′. A related NMR hairpin structure (model 1 of PDB accession code 1ZIH)[4] was used as initial structure in this REMD simulation. REMD simulation was carried out with an explicit TIP3P water model,[41] under periodic boundary conditions. The structure was solvated in a cubic box of 42 Å containing 2285 water molecules, 11 Na^+ ions to neutralize the system, and 7255 atoms in total. The REMD simulation was conducted under constant volume with 48 replicas. An exponentially increasing temperature series along the replicas from 300 to 575.5 K was used, which gives approximately uniform acceptance ratios for exchanges between neighboring replicas.[30−32] Exchange probabilities between neighboring replicas were observed of ∼25%. The REMD simulation was continued for 120 ns for each replica. In total, an accumulated simulation time of 5.76 μs was obtained.

The trajectories output from REMD simulation are usually sorted according to temperature. They are discontinuous regarding the structural transformation due to the repeated replica exchange events. These trajectories form canonical ensembles under different temperatures so they are named as ensemble trajectories. According to the exchange information, ensemble trajectories can be resorted in such a way that each trajectory contains continuous structure transitions of the simulated system beginning from each replica. Such trajectories are known as replica trajectories. Evidently, in a replica trajectory, due to the rescaling of velocities during the exchange process, the temperature has no defined meaning.

Root-mean-square deviation of the whole hairpin with respect to the NMR structure[4] (wRMSDnmr) was calculated based on all atoms. Cluster analysis was performed for the corresponding ensembles using the algorithm proposed by Daura et al.[42] The conformations were clustered by their structural similarity, measured by the pairwise rmsd. The related rmsd cutoff was chosen as 2.0 Å. Curves method[43] was used to find the sugar pucker pattern for different conformations in the GCAA tetraloop.

To further study the structural heterogeneity of RNA in direct comparison with NMR measurement, distances of interresidue hydrogen atoms that were employed as nuclear Overhauser effect (NOE) constraints in the work of Jucker et al.[4] were calculated. Forty-seven interresidue distances were selected, of which 22 are within the stem residues and 25 are within the loop and between stem and loop residues. The complete atom pairs and distances are listed in the Supporting Information. For each snapshot, the fraction of NOE distances satisfaction, Q, was calculated:

$$Q_i = N_s/N_{tot}$$

where N_s is the number of interhydrogen distances whose average values are satisfied by the NOE constraints and N_{tot} is the number of total NOE distances. Q_{stem} and Q_{loop} were calculated separately for stem and loop residues. The average interhydrogen distance is calculated by

$$r_{ij}^{avg} = \left(\frac{1}{K}\sum_{k=1} r_{ij}^{-6}\right)^{-1/6}$$

where r_{ij} is the distance and K is the number of snapshots used in the calculation.

Free energy (G, in units of kilojoules per mole) was evaluated by the following equation:

$$G = -k_B T \ln (N_i/N_{ref})$$

where k_B is the Boltzmann constant, T is temperature, N_i refers to the number of members in state i, and N_{ref} refers to the number of members in a reference state.

Results and Discussion

Configurational Space Sampled. In order to demonstrate the sampled configuration space in the REMD simulation,

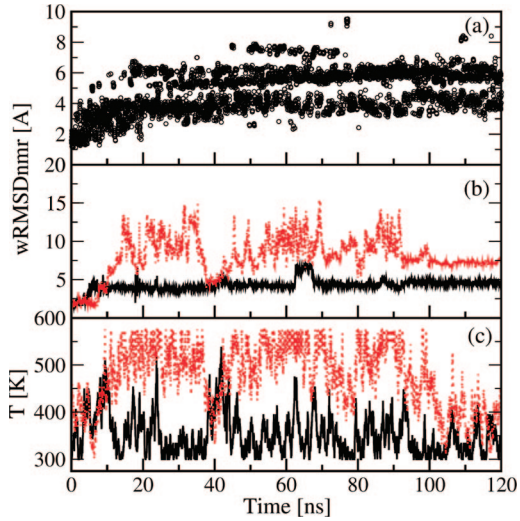

Figure 1. (a) wRMSDnmr evolution of two ensemble trajectories (300 K, black line; 344.6 K, red dotted line). (b) wRMSDnmr evolution of three replica trajectories (replica 1 in black and replica 21 in red). (c) Temperature evolution of the same replica trajectories (replicas 1 and 21).

Figure 1a shows the wRMSDnmr of ensemble trajectories at lowest temperatures (300 K). At the beginning of the simulation the wRMSDnmrs are around 2 Å. After 20 ns, most of the structures have wRMSDnmr of 4 Å. Meanwhile, a group of structures with wRMSDnmr of 6 Å is accumulating. From 40 ns on, structures with wRMSDnmr more than 7 Å enter the low-temperature ensembles with small population.

Figure 1b,c shows the wRMSDnmr and temperature evolution of two replica trajectories with replica 1 (solid black line) and replica 21 (dashed red line). Replica 21 experiences unfolding quickly after the simulation begins. Its wRMSDnmr changes from 2 to 12 Å during the first 20 ns simulation. Around $t = 40$ ns it refolds with wRMSDnmr decreased below 5 Å. Then it unfolds again. Its temperature changes from low values (300 K) to high values (around 550 K) when it unfolds and decreases to low values (350 K) when it refolds. Such unfolding/refolding events happen in other replica trajectories as well.

There are a few replica trajectories whose wRMSDnmr maintains below 5 Å during the whole simulation such as replica 1 (solid line in Figure 1b). Unlike the wRMSDnmr value, its temperature undergoes frequent changes (solid line in Figure 1c) from 300 to 450 K. The invariance of wRMSDnmr against temperature indicates the stability of the configuration in the temperature range.

Figure 2 is the pseudo-free-energy surface as a function of wRMSDnmr and temperature to illustrate the sampled conformation during simulation. There are two local minima at low temperature range ($T < 400$ K). One of the local minima is located at wRMSDnmr < 5 Å centered at wRMSDnmr = 3.5 Å, corresponding to the ensemble of native folded structures and the wRMSDnmr of the other minimum is around 5.7 Å, which is a kind of intermediate state. In the high-temperature region ($T > 450$ K) there is a broad valley on the surface, which represents a heterogeneous ensemble of unfolded structures (wRMSDnmr > 7 Å). The inset in Figure 2 plots the average wRMSDnmr as a function

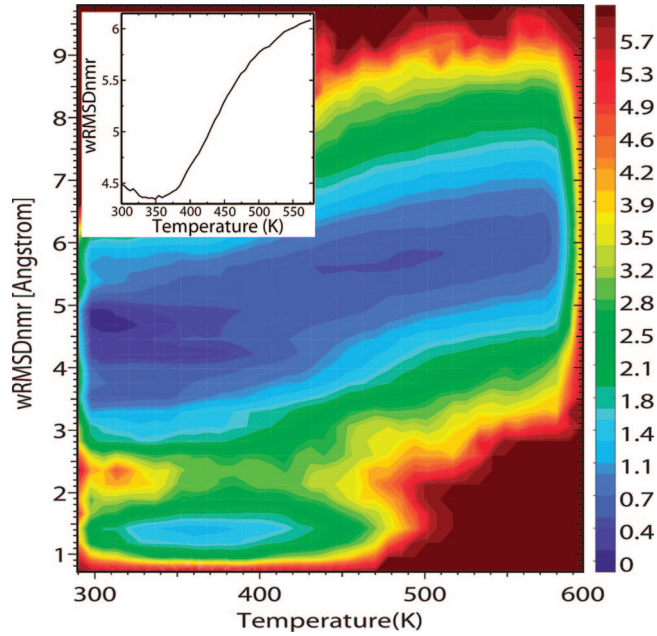

Figure 2. Free energy of sampled conformations as a function of wRMSDnmr and temperature. (Inset) Average wRMSDnmr as a function of temperature.

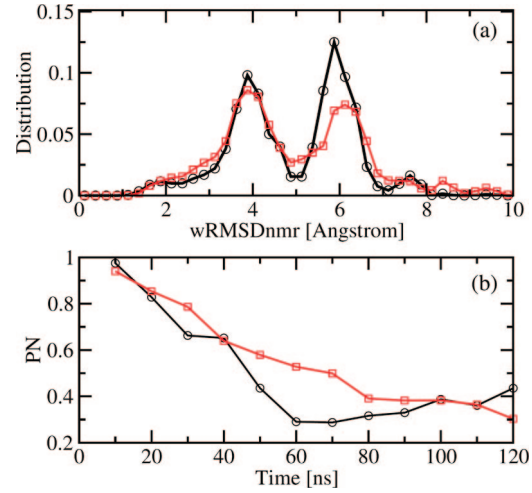

Figure 3. (a) Distribution of wRMSDnmr at $T = 300$ K (black line and circles) and $T = 350$ K (red line and squares). (b) Evolution of the population of near-native state as a function of simulation time; the population is calculated in an interval of 10 ns.

as temperature. The average wRMSDnmr in the temperature range of 300∼350 K is around 5 Å, and the value of wRMSDnmr quickly increases as the temperature rises.

Figure 3a shows the distribution of wRMSDnmr at two temperatures, $T = 300$ K (black circles) and $T = 350$ K (red squares). From these wRMSDnmr distributions it is clearly seen that the structural ensembles at these temperatures can be classified into two states, one with wRMSDnmr smaller than 5 Å, which is called native/near-native state, and the other with wRMSDnmr larger than 5 Å, which is denatured state. To further check the convergence of the simulation, the population of near-native state (PN) as a function of simulation time is monitored and shown in Figure 3b. The near-native state is defined by wRMSDnmr ≤ 5 Å. The PN is calculated in an interval of 10 ns at two temperature, $T =$

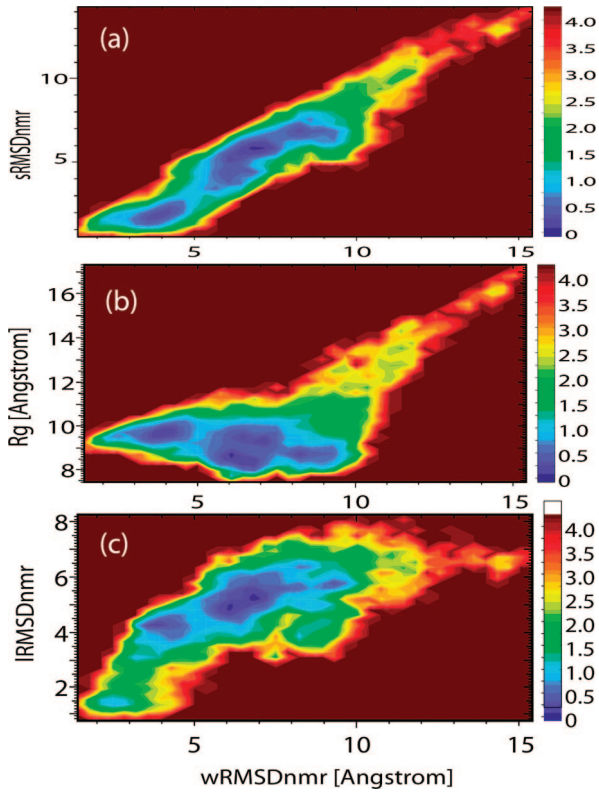

Figure 4. Free energy surfaces at 418 K as functions of (a) sRMSDnmr and wRMSDnmr, (b) R_g and wRMSDnmr, and (c) lRMSDnmr and wRMSDnmr.

300 K (black circles) and $T = 350$ K (red squares). Because the initial configurations are native, PN is close to 1 at the beginning of simulation. PN decreases as the simulation goes on until $t = 60$ ns when PN drops to 0.3 at $T = 300$ K, which indicates a general unfolding behavior during the first 60 ns. After that, the PN value is maintained between 0.3 and 0.4 during the simulation at $T = 300$K. Thus a rough equilibrium between unfolding and refolding processes is being established during the last 60 ns simulation.

Folding/Unfolding of GCAA Tetraloop. Simulation of the folding and unfolding of biomolecules has been one of the major computational challenges of molecular biology.[23] REMD simulations have been successfully applied in folding and unfolding study for small nucleic acids.[22,26] Conformational dynamic transitions related to unfolding and refolding in the GCAA tetraloop were also sampled in the current REMD simulation.

Four parameters were employed here to monitor the structural transitions: wRMSDnmr, stem rmsd (sRMSDnmr, rmsd calculated only for stem residues G1, G2, G3, C4, G9, C10, C11, and U12), loop rmsd (lRMSDnmr, rmsd calculated only for loop residues C4, G5, C6, A7, A8, and G9), and radius of gyration of the whole RNA, R_g. Figure 4 shows the free energy surfaces as a function of (a) wRMSDnmr and sRMSDnmr, (b) wRMSDnmr and R_g, and (c) wRMSDnmr and lRMSDnmr at 418 K. The reason to choose 418 K is that around this temperature the folded and unfolded conformations have nearly equal populations. Figure 4a displays high correlation between sRMSDnmr and wRMSDnmr, which indicates that the folding/unfolding of the tetraloop RNA is mainly determined by the formation/disturbance of the stem part. This observation is consistent with a recent MD study on short RNA loop sequences.[24] On the contrary, the loop dynamics seems uncoupled from the whole RNA (Figure 4c): when wRMSDnmr decreases from 10 to 5 Å, the lRMSDnmr changes only from 6 to 5 Å; in the native basin where wRMSDnmr < 5 Å, there is still a large amount of loop dynamics happening with lRMSDnmr changes between 5 and 1 Å and such loop dynamics in the native basin is the main focus of the next section. The free energy surface of wRMSDnmr and R_g (Figure 4b) clearly indicates that a structural collapse happens at the very early stage of folding and some intermediate states with wRMSDnmr around 7 Å are more compact than native structures.

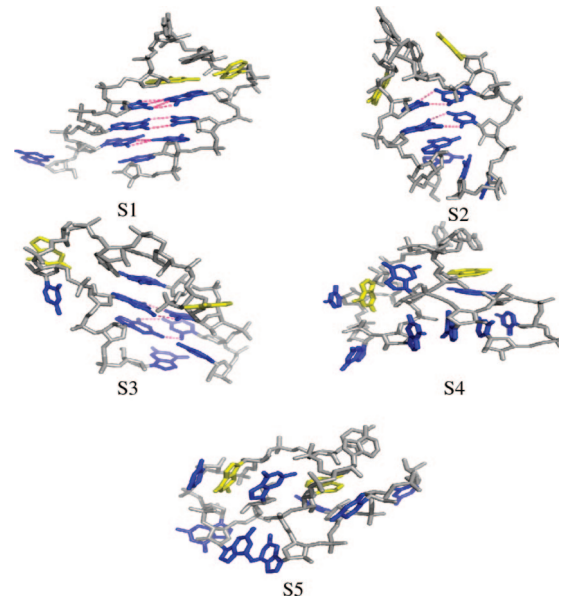

Figure 5. Three-dimensional conformations of the five representative structures (the eight bases of the stem are shown in blue, the G5 and A8 bases of the loop are shown in yellow, and the polar contacts between base pairs are shown in pink).

Based on the information provided by the three free energy surfaces, five representative structures (denoted as S1−S5) corresponding to the minima on the surfaces were selected and are shown in Figure 5. Their structural characterization is summarized in Table 1. In Figure 5 the eight bases of the stem are shown in blue, the G5 and A8 bases of the loop are shown in yellow, and the polar contacts between base pairs are indicated with pink dotted lines. S1 represents the folded conformation with $R_g = 9.6$ Å, wRMSDnmr = 4.1 Å, sRMSDnmr = 2.0 Å, and lRMSDnmr = 4.3 Å. It contains three hydrogen-bonded stem base pairs (C4-G9, G3-C10, and G2-C11) and the bases are stacking properly in the stem. S2 and S3 have two stem paired bases with wRMSDnmr larger than 6 Å, which represent the intermediate states. Both S4 and S5 do not have paired stem residues and the base stacking pattern in the stem region is lost, which represent the unfolded conformations with small R_g and large wRMSDnmr values.

From the free energy surfaces and the representative structures, some conclusions about the folding/unfolding of RNA GCAA tetraloop can be drawn. First, an obvious

Table 1. Structural Characteristics of Representative Structures (Denoted as S1−S5) Sampled in the RNA GCAA Tetraloop at 418 K

structure (rmsd for total, stem, loop, and R_g; Å)	stem base pairs	base stacking patterns in stem
S1 (4.1, 2.0, 4.3, 9.6)	C4-G9 G3-C10 G2-C11	C4/G3/G2/U12 G5/G9/C10/C11
S2 (6.1, 5.0, 4.7, 8.8)	C4-G9 G3-C10	C4/G3/G2/G1 G9/C10 C11/U12
S3 (7.1, 5.9, 5.2, 9.4)	G3-C10 G2-C11	G9/G3/G2
S4 (8.2, 6.9, 5.6, 8.7)		G1/G5/C4 A8/C10 G2/G9/U12/C11
S5 (9.4, 6.7, 5.8, 8.9)		G1/G5/C4 G2/G9

multiple-state (including folded, intermediate, and unfolded states) folding landscape of the RNA GCAA tetraloop is disclosed, which is in good agreement with the previous RNA simulations.[17,21] Second, the folded-to-unfolded conformational transitions happen beginning with breaking of the terminal stem base pairs and decreasing of R_g. The decrease of R_g is mainly related to the loss of stacking pattern of bases on both strands of the stem. This is well consistent with the previous simulations,[17] which demonstrated that the lost of terminal base pair is the first step of unfolding mechanism in the GNRA tetraloop and proposed that a folding pathway includes the collapse of the unfolded state. Third, some intermediate conformations are observed with the central stem base pairs (G3-C10, G2-C11) intact and with wRMSDnmr of 6−7 Å.

Loop Dynamics near the Native Basin. As mentioned in the Introduction, fluorescence spectroscopic measurements indicated that ample conformational transitions of the loop residues exist within the native state.[14,15] In order to study the loop dynamics in more detail, an extensive clustering was performed based on the ensembles; in total 2 880 000 structural snapshots from the last 60 ns simulation. First 72 000 structural snapshots were chosen by taking one structure out of every 40. A clustering analysis was performed on the 72 000 structures with pairwise loop RMSDs (lRMSDp) cutoff of 1.5 Å. The first largest 100 clusters were chosen. The center of each cluster identified by clustering analysis tool was taken as the representative structure. Then the lRMSDp of all 2 880 000 structures with respect to the 100 representative structures was calculated. Each structure was assigned a cluster label if the lRMSDp with respect to the cluster center is smaller than 1.3 Å. Structures whose lRMSDps with respect to the 100 cluster centers are larger than 1.3 Å are called outliers, labeled as 0.

To check the near-native nature of the representative structures that are related to the cluster centers, Figure 6a shows the wRMSDnmr of the cluster center of the largest 34 clusters. Clusters 1, 4, 7, and 26 represents the native folded structure with wRMSDnmr ∼3 Å, and the other clusters correspond to the near-native ensemble of structures with wRMSDnmr ≤ 5 Å, with the exception of clusters 14 and 22 whose wRMSDnmrs are larger than 7 Å. It is not surprising to find that most clusters with many members belong to near-native ensembles because the structural heterogeneity of the unfolded ensembles is greater than that of near-native ones.

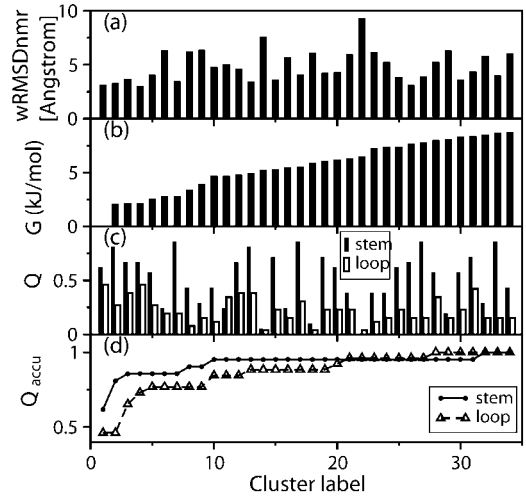

Figure 6. (a) Global wRMSDnmr of the largest 34 clusters with respect to NMR structure. The clustering is based on loop C(GCAA)G structure. (b) Relative free energy of the 34 clusters. (c) Degree of NOE constraint satisfaction, Q, for each cluster. (d) Accumulated Q values (Q_{accu}) as a function of n largest clusters

Figure 6b shows the free energy of different structural clusters relative to class 1 at a simulation temperature of 300 K. The free energy differences range from 0 to 8 kJ/mol. These relatively small free energy differences demonstrate that the structures of GCAA tetraloop are quite flexible and multiple structural ensembles coexist with moderate free energy difference. Our finding is quite consistent with the previous experimental data which indicated the flexibility of the tetraloop.[4,14,16]

NOE Distances Comparison. NOE constraints provide direct structural information obtained from NMR measurement and can be used to gauge the simulated results directly. In Figure 6c, the degree of NOE constraint satisfaction, Q, for each cluster is shown. On average the Q value is less than 0.5, which means in each individual structural cluster less than half of the NOE constraints are satisfied simultaneously. For clusters 1 and 4, which have the lowest wRMSDnmr values, Q values turn out to be larger than 0.5. Generally the Q values are larger for stem than for loop. The relatively low value of Q for each structural cluster indicates that, due to the structural heterogeneity of the RNA tetraloop, it is not the case that an individual structural snapshot may satisfy all of the NOE constraints, which should be an ensemble property intrinsic to the NMR measurement.[44]

To test the ensemble properties of NOE constraints, the accumulated Q values (Q_{accu}) were applied. In calculation of Q_{accu}, the interhydrogen distance was not averaged within each structural cluster. More and more clusters were included to evaluate the Q value instead. Thus Q_{accu} values as a function of n largest clusters included were obtained and are shown in Figure 6d. Q_{accu} values for both stem and loop residues increase quickly as more clusters are included. When

Table 2. Base Stacking Patterns for Nine Groups of Conformations in RNA G1C2A3A4 Tetraloop[a]

conformation type	base stacking patterns in G1C2A3A4 loop[b]
A	G1×C2/A3/A4
B	G1×C2×A3/A4 (C2 out of loop)
C	G1×C2/A3×A4, G1/A3
D	G1×C2×A3×A4, G1/A3
E	G1/C2/A3×A4
F	G1/C2×A3×A4
G	G1/C2×A3/A4
H	G1×C2/A3×A4
I	G1×C2×A3/A4, G1/A3

[a] Based on the distance distribution, the cutoff values are 5.1, 5.6, 6.7, 5.6, and 6.1 Å for the maximum distances between bases 1 and 2, 1 and 3, 2 and 3, 2 and 4, and 3 and 4, respectively. [b] × refers to unstacking; / refers to stacking.

the first five clusters are included in the calculation, the value of Q_{accu} for loop is 75% and that for stem is 85%. When the first 13 clusters are taken as an ensemble, both Q_{accu} values approach 0.90.

Combing Q and Q_{accu} values, our simulation results indicate the heterogeneity of the individual configuration sampled, and on the other hand the NMR NOE constrains are able to be satisfied as an ensemble.

Base Stacking Pattern Identification. In order to have a direct comparison of loop dynamics found in the simulation with those indicated by fluorescence measurements,[14,16] the structures sampled by the REMD simulation were grouped according to base stacking pattern. Two bases in the loop were set to be stacked only when the *maximum* distance between six heavy atoms of a six-membered ring on the two bases is smaller than a cutoff value, which is determined on the basis of the distance distribution. Due to the distinct stacking patterns, nine groups with different loop structure were identified (labeled as A−I); their loop base stacking patterns are explained in Table 2. Six out of these nine structure groups (A−F) were found to be analogues of the experimental model.[16] Three-dimensional representative structures of the nine conformations are shown in Figure 7. The four bases in the loop part, G1C2A3A4, are shown in yellow.

Distinct Loop Stacking Patterns Found in the Simulation. From Figure 7 and Table 2, it is found that group A represents the 3′-stacked structure, in which the last three bases are stacked on the 3′ side. This kind of structure has been revealed to be the main conformation both for the isolated tetraloop and for the tetraloop−receptor complexes.[4−6,16,45] In group B, A3 stacks on A4 on the 3′ side and C2 is looped out. Such a conformation has been illustrated by NMR experiment.[4] Group C has the base stacking pattern of G1, C2, and A3 stacking on the 5′ side. In group D, bases A3 and G1 are stacking on the 5′ side. Group E is a typical 5′-stacking conformation, with the first three bases stacking on the 5′ side. The conformations found in the simulation support the earlier proposal that a 5′-stacking form coexists with a 3′-stacking form in the GNRA tetraloop.[14,16] Group F has the base stacking pattern of C2 on G1 in the 5′ stack. In group G, the stacking pattern is C2, G1 stacking on the 5′ side and A3, A4 on the 3′ side.

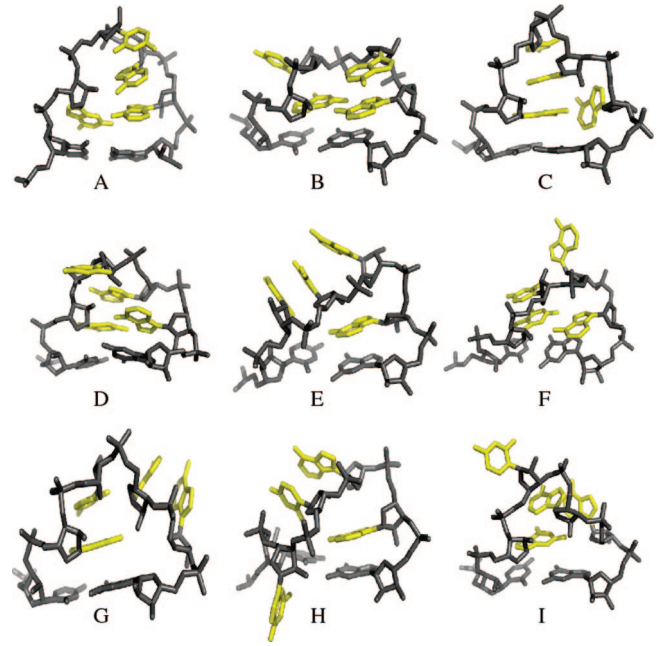

Figure 7. Representative structures of nine structural groups of G(GCAA)C loop, with the four bases of the GCAA tetraloop shown in yellow.

Group H has the base stacking pattern of C2 on A3, and these two bases are in a looped-out configuration. In group I, A3 stacks on both A4 and G1.

Stacking Patterns in NMR Structures. Table 3 shows stacking patterns resolved in NMR-determined structures of GCAA tetraloop as well as of other tetraloops (including GNRA, UNCG and RNYA, where Y = pyrimidine).[46] In the canonical conformation of GNRA tetraloop, PDB entry 1hmh,[47] the last three bases are stacked on the 3′ side, corresponding to group A. For other GNRA loop conformations, PDB entries 1cn8,[48] 1etf, [49] and 1f1t[50] show that the second base N is looped out and the last two bases are stacked on the 3′ side, similar to group B. In the standard conformation of UNCG tetraloop, PDB entry 1dk1,[51] the first and third bases are posed in the 5′ stack and the second base N is looped out, similar to group D. In other structures, PDB entries 1bgz,[7] 1d6k,[52] and 1tlr,[53] the second and third bases are looped out, corresponding to group H. In one conformation of RNYA tetraloop, PDB entry 5msf,[54] the first and third bases come close on the 5′ side and the second base is looped out, similar to group D. In another structure of RNYA tetraloops, PDB entry 1tfn,[55] the second and third bases are looped out, corresponding to group H. In other structures of RNYA tetraloops, from PDB entries 1d0t and 1d0u,[56] the first two bases are in the 5′ stacking and the third is looped out, similar to group F.

The loop stacking patterns in groups A, B, D, F, and H can find their analogues in the NMR-determined structures. Such matching is an encouraging sign, which indicates that sampling results from the REMD simulation are testable by the experimental models.

Structure Transition Map. Combined with the "kinetic" information available from continuous replica trajectories, a structural transition map for the GCAA tetraloop between distinct groups was constructed, which is shown in Figure

Table 3. Comparison of Base Stacking Patterns of GCAA Tetraloops Sampled in Our Simulation with NMR Models (Including GNRA, UNCG, and RNYA)

tetraloop	PDB entry	12[a]	23[a]	34[a]	13[a]	conformation
GNRA	1hmh		3′ stack	3′ stack		A
GNRA	1cn8, 1etf, 1f1t	loop-out of N2		3′ stack		B
UNCG	1dk1	loop-out of N2			5′ stack	D
UNCG	1bgz, 1d6k, 1tlr		loop-out of N2 and C3			H
RNYA	5msf	loop-out of N2		loop-out of A4	5′ stack	D
RNYA	1tfn		loop-out of N2 and C3			H
RNYA	1d0t, 1d0u	5′ stack	loop-out of Y3			F

[a] Numbers 1, 2, 3, and 4 refer to the first, second, third, and last base in the tetraloop, respectively.

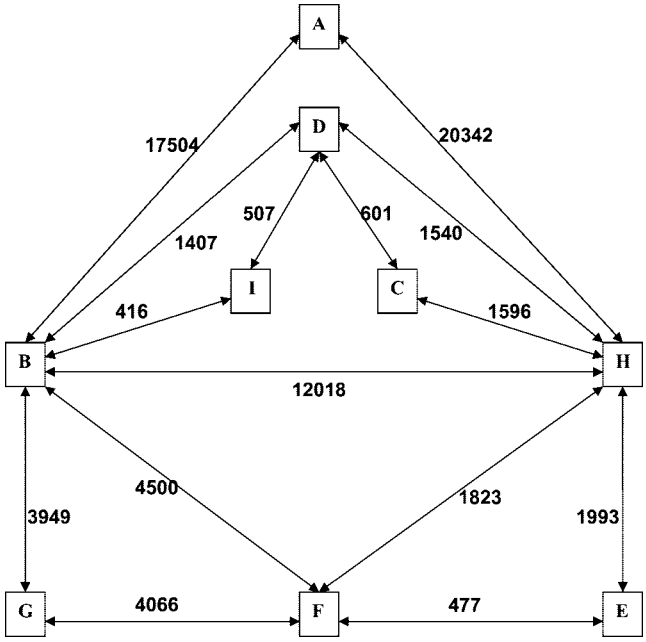

Figure 8. Conformational transition map for the GCAA tetraloop, where A−I are nine groups of conformations that have distinct base stacking patterns. Arrows between groups indicates conformational transitions, with the number of transitions nearby.

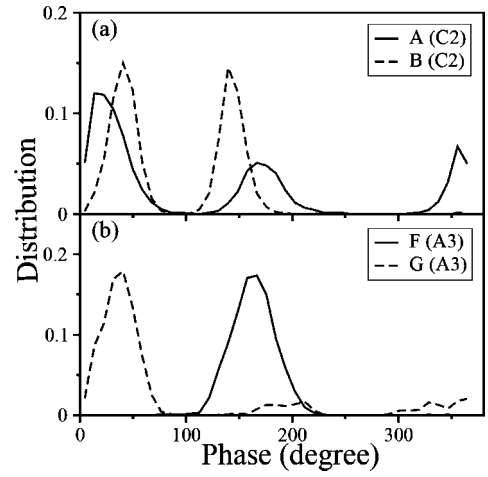

Figure 9. Distributions of the pucker phase: (a) C2 in groups A and B; (b) A3 in groups F and G.

8. Such a transition map can be directly compared with recent fluorescence measurements in which a detailed dynamic multiconformation model is suggested.[16] Encouragingly, the transition map shown in Figure 8 supports the multiconformation model, and moreover, detailed transition pathways are consistent with each other.

The A−B−F−E transition pathway is in good agreement with the experimental model,[16] which proposed an A−B−D−F−E transition pathway and predicted that the equilibrium between groups A and B exists in GNRA tetraloops and could be an important pathway for transition from group A to other conformations.

The transition between B and D groups is also consistent with the proposal of the experiment.[16] Such a transition indicates that the third base R3 can slide over the G1-A4 sheared base pair, switching from stacking on A4 to stacking on G1 and vice versa.

Interestingly, a new transition pathway linking groups A and E through group H is found, which has not been resolved by experimental measurement. The structural characteristic of group H is that the middle two bases (2 and 3) are stacking with each other, however, in a looped-out configuration. Although group H has not been found in NMR structure models of GCAA, for other kinds of tetraloop RNA (UNCG and RNYA), NMR structural models do show configurations similar to group H.[7,55]

The transitions between looped-in and looped-out base conformations may generally occur cooperatively with transitions of puckering of furanose rings from 3′-endo to 2′-endo in the RNA GCAA tetraloop.[17] Experimental support came from the Raman spectroscopy study by Leulliot et al.[57] In their study a 2′-endo shoulder was seen in the UUCG tetraloop spectrum, which was assigned to the second base U2 within a looped-out conformation. The sugar pucker phase was calculated here to investigate the pucker mode of the residues in different structural groups. Distributions of pucker phase angles are shown in Figure 9. Distributions of the pucker phase of C2 for groups A and B (shown in Figure 9a) demonstrate that the occupancy located at ∼150° (corresponding to 2′-endo pucker) increases from structure group A to B. The pucker phase of A3 shows a similar trend (Figure 9b): when it is in the looped-out configuration (structure group F), A3 takes the 2′-endo conformation predominantly. From the structural point of view, the 2′-endo pucker mode expands the backbone of the nucleotide.[58] Our simulation confirms that when the nucleotide takes the looped-out conformation, its sugar pucker phase has more probability to be in the 2′-endo state.

The results of our simulation show that the GCAA tetraloop is not rigid and may undergo multiple conformational transitions, which provide theoretical support of observations of previous fluorescence studies.[4,14,16] And these structural fluctuations may lead to a variable set of hydrogen bonds.[4] These heterogeneous hydrogen bonds can signifi-

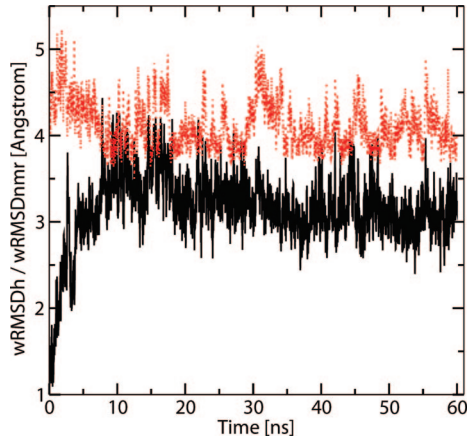

Figure 10. Evolution of rmsd values simulated by use of the new Amber force field parmbsc0, initiated from a loop configuration of H type with respect to initial structure (black solid line) and NMR structure (red dotted line).

cantly stabilize these multiple conformations. This may be one of the possible explanations for the multiple conformational transitions in the RNA GCAA tetraloop. A recent NMR ^{13}C relaxation study on the same GCAA tetraloop also revealed from the fast dynamics that this tetraloop has a rather rigid stem and a significantly more flexible loop.[13]

New Amber parmbsc0 Force Field. All the simulations so far were run with the force field Amber parm98. Recently a new modification of Amber force field was proposed that corrected the fault of artificial α/γ crankshaft motion of the nucleic acid backbone.[38] To test to what extent our simulation results depend on the force field, a test simulation was performed. An initial structure with loop pattern of H type (only bases C2 and A3 are stacked in the loop) was chosen. A 60 ns simulation with the new parmbsc0 force field was performed at 300 K. The reason that the H-type configuration was tested is that this configuration is the prediction of the current study that was not mentioned in earlier experimental interpretations. The rmsd values for the whole hairpin with respect to the initial H-type configuration (wRMSDh) and with respect to the NMR structural model (wRMSDnmr) are monitored as shown in Figure 10. During the last 40 ns of the simulation, wRMSDh fluctuates between 2.5 and 4 Å (black solid line) and wRMSDnmr fluctuates between 3.5 and 5 Å (red dotted line). Three dominant loop stacking patterns, B, H, and A type, were found with populations of 42%, 12%, and 6%, respectively. Frequent transitions between these three loop stacking patterns were observed, which is consistent with the transition map shown in Figure 8.

The detailed differences between the parm98 and parmbsc0 simulations are disclosed in the distributions of the backbone dihedral angles, α and γ, on which the force field parameters have been refined in parmbsc0. Both simulations produce similar α and γ distributions for stem residues; for example, see the distributions of α and γ for residue 3 in Figure 11. For residues on the loop, the parm98 simulation gives a broader distribution, probably due to the REMD simulation protocol, which can generate enhanced sampling: for example, the α and γ dihedral angles of residue 7, which is the third residue in the loop, are mainly located around

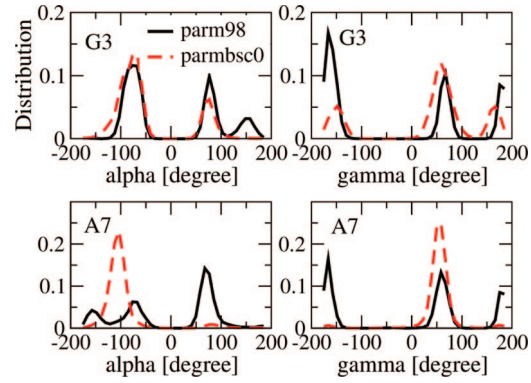

Figure 11. Distributions of backbone dihedral angles, α and γ, sampled with Amber parm98 force field (black solid line) in comparison with those sampled with the new force field parmbsc0 (red dashed line) for residues 3 and 7 of the tetraloop RNA.

$-100°$ and 60°, respectively, in the parmbsc0 simulation; in the parm98 REMD simulation, the distributions of both α and γ angles are multiple-peaked with one of the maxima positions coinciding with or close to the maximum position found in the parmbsc0 simulation. Thus we think that the main results of our simulation do not depend on the force field in a sensitive way.

Summary

A REMD simulation in explicit solvent was performed to probe the multiple conformations coexisting in RNA GCAA tetraloop. An obvious multiple-state (including folded, intermediate, and unfolded states) folding landscape of the RNA GCAA tetraloop is disclosed, which is in good agreement with previous RNA simulations.[17,21] On the basis of extensive cluster analysis of ample simulation trajectories, a dynamic structure transition map for the GCAA tetraloop is constructed, which is well consistent with the model from fluorescence measurements. The structure transition map of the GCAA tetraloop presented here should lead to a deeper understanding of the dynamic transition mechanism of the GNRA tetraloop family.

Acknowledgment. This work was supported by the University Research Committee, URC RG65/06, grant and academic research fund AcRF Tier 2 by the MOE (T206B3210RS), which are gratefully acknowledged. Financial support from the China Scholarship Council is gratefully acknowledged by Y.Z.

Supporting Information Available: Complete atom pairs and distances for NOE calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Uhlenbeck, O. C. *Nature (London)* **1990**, *346*, 613.

(2) Cheong, C. J.; Varani, G.; Tinoco, I. *Nature (London)* **1990**, *346*, 680.

(3) Wolters, J. *Nucleic Acids Res.* **1992**, *20*, 1843.

(4) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968.

(5) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191.

(6) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K. H.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678.

(7) Kalurachchi, K.; Nikonowicz, E. P. *J. Mol. Biol.* **1998**, *280*, 639.

(8) Zhang, Q.; Al-Hashimi, H. M. *Nat. Methods* **2008**, *5*, 243.

(9) Hodak, J. H.; Fiore, J. L.; Nesbitt, D. J.; Downey, C. D.; Pardi, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10505.

(10) Furtig, B.; Richter, C.; Wohnert, J.; Schwalbe, H. *ChemBioChem* **2003**, *4*, 936.

(11) Legault, P.; Hoogstraten, C. G.; Metlitzky, E.; Pardi, A. *J. Mol. Biol.* **1998**, *284*, 325.

(12) Hoogstraten, C. G.; Wank, J. R.; Pardi, A. *Biochemistry* **2000**, *39*, 9951.

(13) Trantirek, L.; Caha, E.; Kaderavek, P.; Fiala, R. *J. Biomol. Struct. Dyn.* **2007**, *25*, 243.

(14) Menger, M.; Eckstein, F.; Porschke, D. *Biochemistry* **2000**, *39*, 4500.

(15) Rist, M. J.; Marino, J. P. *Curr. Org. Chem.* **2002**, *6*, 775.

(16) Zhao, L.; Xia, T. B. *J. Am. Chem. Soc.* **2007**, *129*, 4118.

(17) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493.

(18) Nivon, L. G.; Shakhnovich, E. I. *J. Mol. Biol.* **2004**, *344*, 29.

(19) Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. *Structure* **2005**, *13*, 1255.

(20) Hyeon, C.; Thirumalai, D. *Biophys. J.* **2006**, *90*, 3410.

(21) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523.

(22) Kannan, S.; Zacharias, M. *Biophys. J.* **2007**, *93*, 3218.

(23) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers* **2007**, *85*, 169.

(24) Nystrom, B.; Nilsson, L. *J. Biomol. Struct. Dyn.* **2007**, *24*, 525.

(25) Zhuang, L.; Jaeger, L.; Shea, J.-E. *Nucleic Acids Res.* **2007**, *35*, 6995.

(26) Garcia, A. E.; Paschek, D. *J. Am. Chem. Soc.* **2008**, *130*, 815.

(27) Villa, A.; Widjajakusuma, E.; Stock, G. *J. Phys. Chem. B* **2008**, *112*, 134.

(28) Sarzynska, J.; Nilsson, L.; Kulinski, T. *Biophys. J.* **2003**, *85*, 3445.

(29) Sarzynska, J.; Kulinski, T. *J. Biomol. Struct. Dyn.* **2005**, *22*, 425.

(30) Sanbonmatsu, K. Y.; Garcıa, A. E. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225.

(31) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.

(32) Zhou, R. H. *J. Mol. Graph.* **2004**, *22*, 451.

(33) Mu, Y.; Nordenskiold, L.; Tam, J. P. *Biophys. J.* **2006**, *90*, 3983.

(34) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(35) Zhang, J.; Qin, M.; Wang, W. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 672.

(36) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.

(37) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.

(38) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

(39) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J *J. Comput. Chem.* **1997**, *18*, 1463.

(40) Daura, X.; van Gunsteren, W. F.; Rigo, D.; Jaun, B.; Seebach, D. *Chem.—Eur. J.* **1997**, *3*, 1410.

(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(42) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236.

(43) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.

(44) Daura, X.; Antes, I.; van Gunsteren, W. F.; Thiel, W.; Mark, A. E. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 542.

(45) Davis, J. H.; Tonelli, M.; Scott, L. G.; Jaeger, L.; Williamson, J. R.; Butcher, S. E. *J. Mol. Biol.* **2005**, *351*, 371.

(46) Klosterman, P. S.; Hendrix, D. K.; Tamura, M.; Holbrook, S. R.; Brenner, S. E. *Nucleic Acids Res.* **2004**, *32*, 2342.

(47) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature (London)* **1994**, *372*, 111.

(48) Mao, H. Y.; White, S. A.; Williamson, J. R. *Nat. Struct. Biol.* **1999**, *6*, 1139.

(49) Battiste, J. L.; Mao, H. Y.; Rao, N. S.; Tan, R. Y.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. *Science* **1996**, *273*, 1547.

(50) Baugh, C.; Grate, D.; Wilson, C. *J. Mol. Biol.* **2000**, *301*, 117.

(51) Nikulin, A.; Serganov, A.; Ennifar, E.; Tishchenko, S.; Nevskaya, N.; Shepard, W.; Portier, C.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *Nat. Struct. Biol.* **2000**, *7*, 273.

(52) Stoldt, M.; Wohnert, J.; Ohlenschlager, O.; Gorlach, M.; Brown, L. R. *EMBO J.* **1999**, *18*, 6508.

(53) Butcher, S. E.; Dieckmann, T.; Feigon, J. *EMBO J.* **1997**, *16*, 7490.

(54) Rowsell, S.; Stonehouse, N. J.; Convery, M. A.; Adams, C. J.; Ellington, A. D.; Hirao, I.; Peabody, D. S.; Stockley, P. G.; Phillips, S. E. V. *Nat. Struct. Biol.* **1998**, *5*, 970.

(55) Kerwood, D. J.; Borer, P. N. *Magn. Reson. Chem.* **1996**, *34*, S136.

(56) Smith, J. S.; Nikonowicz, E. P. *Biochemistry* **2000**, *39*, 5642.

(57) Leulliot, N.; Baumruk, V.; Abdelkafi, M.; Turpin, P. Y.; Namane, A.; Gouyette, C.; Huynh-Dinh, T.; Ghomi, M. *Nucleic Acids Res.* **1999**, *27*, 1398.

(58) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

CT8004276