doi:10.1006/jmbi.2002.5447 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2002) **317**, 493–506

JMB

RNA Simulations: Probing Hairpin Unfolding and the Dynamics of a GNRA Tetraloop

Eric J. Sorin[1], Mark A. Engelhardt[2], Daniel Herschlag[1,2] and Vijay S. Pande[1,3,4,5]*

[1]*Department of Chemistry*

[2]*Department of Biochemistry*

[3]*Department of Biophysics*

[4]*Department of Structural Biology and*

[5]*Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford CA 94305-5080*

**Corresponding author*

Simulations of an RNA hairpin containing a GNRA tetraloop were conducted to allow the characterization of its secondary structure formation and dynamics. Ten 10 ns trajectories of the folded hairpin 5′-GGGC[GCAA]GCCU-3′ were generated using stochastic dynamics and the GB/SA implicit solvent model at 300 K. Overall, we find the stem to be a very stable subunit of this molecule, whereas multiple loop conformations and transitions between them were observed. These trajectories strongly suggest that extension of the C6 base away from the loop occurs cooperatively with an N-type → S-type sugar pucker conversion in that residue and that similar pucker transitions are necessary to stabilize other looped-out bases. In addition, a short-lived conformer with an extended fourth loop residue (A8) lacking this stabilizing 2′-*endo* pucker mode was observed. Results of thermal perturbation at 400 K support this model of loop dynamics. Unfolding trajectories were produced using this same methodology at temperatures of 500 to 700 K. The observed unfolding events display three-state behavior kinetically (including native, globular, and unfolded populations) and, based on these observations, we propose a folding mechanism that consists of three distinct events: (i) collapse of the random unfolded structure and sampling of the globular state; (ii) passage into the folded region of configurational space as stem base-pairs form and gain helicity; and (iii) attainment of proper loop geometry and organization of loop pairing and stacking interactions. These results are considered in the context of current experimental knowledge of this and similar nucleic acid hairpins.

Keywords: RNA folding; GCAA; implicit solvent; stochastic dynamics; sugar pucker

Introduction

Like proteins, RNA molecules adopt specific secondary and tertiary structure in order to act as biocatalysts.[1] Unlike proteins, RNA molecules are thought to form stable secondary structures early in the folding process. To accomplish the 180° backbone inversions necessary for secondary structure formation between nearby hairpin strands, a variety of turn and loop structures exists. Among these, tetraloops are commonplace, with the GNRA, CUUG, and UNCG sequences comprising more than 70% of known tetramer loops while accounting for less than 4% of possible tetraloop sequences.[2,3] Structural and phylogenetic comparisons indicate that these tetraloops are frequently used in the assembly of tertiary structure from secondary structural units.[4-8] Understanding how tetraloop hairpins form and behave is therefore a basic component of understanding the formation and dynamics of larger RNA systems.

GNRA tetraloops, common in ribosomal and signal-recognition particle RNAs, are known to form tertiary contacts in larger RNA structures[2,8,9] and to serve as protein binding recognition sites.[10, 11] Of these, GCAA and GAAA loops are most preva-

Abbreviations used: RMSD, root-mean-squared deviation; GB, generalized Born; PB, Poisson-Boltzmann; SA, surface area; COM, center-of-mass; NC, native contact; F, folded/native configuration; G, compact globular unfolded conformation; U, extended unfolded conformation; N, any ribonucleotide; R, ribonucleotide with a purine base, adenine or guanine; L2/L4, the 2nd/4th nucleotide position in a loop structure.

E-mail address of the corresponding author: pande@standford.edu

0022-2836/02/040493–14 $35.00/0

lent, and share similar structural characteristics.[12] For instance, the stacked third and fourth loop A residues in such tetraloops interact with specific minor groove receptor sites and can engage in loop-to-receptor base stacking within larger RNA structures.[7,13–16]

Still, little is known about either the dynamics taking place in these loop regions or the mechanism of formation of nucleic acid hairpins in general. To investigate these dynamics, we have studied a 12-mer RNA tetraloop hairpin previously characterized by NMR.[12] The sequence 5′-GGG C[GCAA]GCCU-3′ is native to *Escherichia coli* 16 S ribosomal RNA with an additional terminal base-pair. Structural features of this hairpin in solution include an often-frayed terminal G·U *cis* wobble pair, the expected *A*-form helical stem conformation, and a favorable C4·G9 closing base-pair at the stem-loop interface.[3,"17] Early molecular dynamics simulations of this species by Zichi[18] suggested that the stem and loop regions of this hairpin act as distinct structural subunits. Also, the experimentally derived thermodynamic properties for this hairpin at 300 K are well known: the melting temperatures for this fragment are ~68 °C and ~71 °C in the presence of low (~5 mM) and moderate (~100 mM) sodium ion concentrations, respectively,[5, "19] and folding of the hairpin at 300 K occurs with $\Delta G = -3.9$ kcal/mol, $\Delta H = -32.7$ kcal/mol, and $\Delta S = -96.0$ cal/K mol.

We report multiple atomistic Langevin dynamics simulations using implicit solvation energy calculations to model loop dynamics specific to the GCAA tetraloop on the ten-nanosecond timescale. The results of these simulations provide experimentally unattainable insight into conformational changes within the tetraloop and offer a model of the folding of the tetraloop hairpin molecule as a whole.

Unfortunately, standard atomistic dynamic simulations fall short of the micro- to millisecond timescales often necessary to simulate folding events of even the smallest biomolecules, such as nucleic acid and protein secondary structural units.[20] We have therefore employed high temperature unfolding simulations to provide potential insights into pathways between the native and denatured states of this system and to assess dynamics within the hairpin stem. Thermal unfolding has served as a powerful tool for investigating dynamics involving large conformational changes on computationally tractable timescales.[20-24] It has been previously suggested that the unfolding process for proteins can in general reflect the main attributes of the folding event,[20–24] and it has been shown that construction of folding pathways, including transition state ensembles, is possible using high temperature unfolding.[21] A recent report of the direct, atomistic folding of a small protein *in silico*"[25] has shown that previously published thermal unfolding trajectories[21] served as a good tool for predicting the folding rate and mechanism. Furthermore, by maintaining low-temperature properties of the continuum solvent model employed we increase the likelihood that the observed dynamics are representative of low temperature unfolding and folding dynamics.

Results and Discussion

We first introduce general structural features of the GCAA tetraloop hairpin, then discuss dynamics within the loop, and finish with a description of thermal unfolding and its relevance to the folding pathway.

Structural overview and stability

The primary tetraloop hairpin structures observed in the 300 K ensemble are introduced in Figure 1. Stick diagrams illustrate the prevalent loop conformations observed, while schematic diagrams show base positions and interactions including the fairly static stem structure. Blue bars between bases represent the native contacts (NC) considered in our analysis.† Color-coding in the stick diagrams follows that in the schematics. Shown are (a) the dominant ''closed-loop'' conformer that displays a well-aligned C6-A7-A8 stacking triplet, (b) the less favored ''open-loop'' structure in which the C6 residue has left the loop to become fully solvated, and (c) the far less stable ''A8-extended'' loop conformer, which retains C6-A7 stacking.

A comparison of the backbone dihedral angles from the ensemble of ten NMR structures with the relaxed starting structure and the average over 100 ns of simulation time at 300 K is shown in Table 1 and Scheme 1. As the loop in this hairpin undergoes significant conformational fluctuations below the melting temperature, the table includes dihedral angles for a central stem base-pair (G3·C10), the closing base-pair (C4·G9), and the mismatch G5·A8 pair that compromises the loop-stem interface. The greatest deviations from the range suggested by the refined NMR structures are shaded. While a small amount of loop distortion is observed, this results from both our relatively short sampling time and the fluctuations observed in this region at 300 K. Overall, the stem retains its structure: only the G9 α-torsion is significantly outside of the NMR range, a result of slight distortion in the A8 backbone region (Table 1 and data not shown).

† We define a native contact (NC) exist between two stacked or paired bases if their center-of-mass separation is within 1.5 Å of that seen in NMR work; paris with larger separations are considered as broken, solvent-exposed base-pairs.

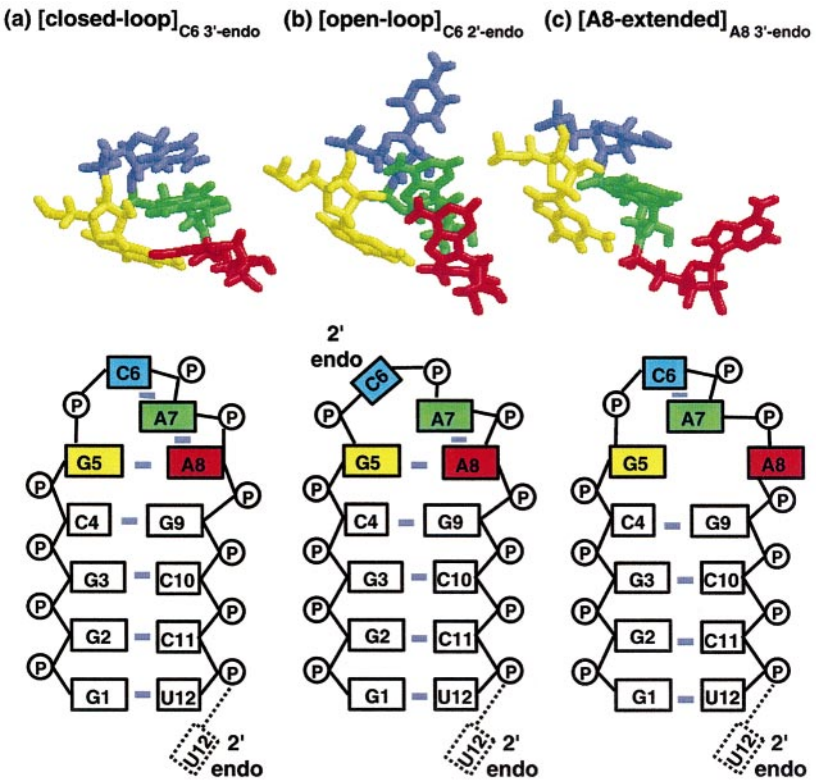
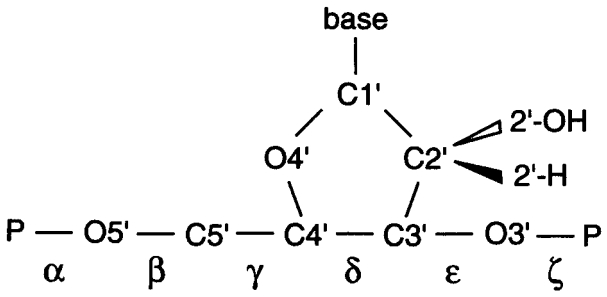


Figure 1. The three GCAA tetraloop hairpin conformations observed at 300 K. Taken directly from our simulations, stick diagrams (top) illustrate the prevalent loop conformations observed. Schematic diagrams (bottom) show base and phosphate group positions and interactions. Blue bars between bases represent the native contacts (NC) considered in our analysis and, for each diagram, the four loop nucleotides are color coded as follows: G5 in yellow, C6 in blue, A7 in green, and A8 in red. All residues have 3′-*endo* pucker modes unless specified and U12 undergoes frequent fraying motions, as noted by broken lines. (a) The dominant closed-loop conformer displays the well-aligned A7-A8 stacking pair. The C6 base is only slightly out of plane from the double-A stack, and maintains a 3′-*endo* pucker. Base-pairing is also occurring between the G5 and A8 bases (yellow-red alignment). (b) In the less common open-loop structure the C6 residue has left the loop to become fully solvated. C6 has a 2′-*endo* sugar pucker only in this conformer, as calculated from NMR data.[12] (c) The far less stable A8-extended loop conformer conserves C6-A7 stacking.

Table 1. Comparison of NMR and simuation dihedral angles in the stem and stem-loop interface

Backbone torsion angle	Central stem		Stem closing		Loop mismatch	
	G3	C10	C4	G9	G5	A8
α	85.29±2.64 85.29 97.88±24.52	89.22±3.63 86.63 87.14±12.30	117.74±25.48 79.99 86.85±8.02	60.83±2.32 110.56 106.74±23.31	87.96±1.80 135.05 139.31±17.69	105.45±27.03 141.108 122.95±26.19
β	170.24±1.33 171.08 166.38±5.40	167.16±2.69 162.80 164.08±12.05	164.04±5.47 172.71 167.10±4.81	135.03±0.07 148.89 138.87±27.10	167.19±1.99 159.789 159.82±17.08	168.95±3.19 161.38 163.13±6.42
γ	70.78±1.57 81.94 97.26±36.16	70.41±1.39 79.40 82.87±19.65	134.89±54.69 72.52 75.53±6.80	79.21±0.68 76.36 111.50±41.29	78.78±30.27 169.98 161.73±20.74	101.02±48.32 176.419 136.36±45.07
δ	95.54±0.66 87.81 90.07±6.02	95.51±0.42 83.06 89.18±5.40	97.43±0.92 83.42 88.60±5.45	95.10±0.56 93.19 90.65±6.44	98.11±0.49 103.72 90.09±7.36	98.95±4.72 90.32 94.156±7.64
ε	157.30±1.34 151.96 155.28±8.67	158.81±1.01 153.30 158.90±7.76	140.60±5.98 170.75 163.09±7.82	163.81±1.27 165.61 155.75±11.03	129.44±11.27 120.30 130.58±17.44	120.32±5.14 161.34 150.73±21.38
ξ	82.19±3.34 76.54 78.30±6.12	70.01±0.91 74.09 79.18±6.96	81.62±4.20 81.45 72.97±8.45	79.29±2.15 77.10 78.93±6.53	74.48±7.14 84.57 94.09±15.09	77.55±1.77 73.88 66.10±8.09

All dihedral angles are in degrees and were calculated from 5′ to 3′ according to Scheme 1.
The top entries give the mean and standard deviation in the ten refined NMR structures.
Below these are the starting structure values following 1 ns relaxation in implicit solvent.
The third entry is the 100 ns average and standard deviation for each torsion.
Values that show the greatest deviation from the refined NMR structural range are shaded.



Scheme 1.

Tetraloop dynamics

Kinetic plots for the ten 300 K trajectories are presented in Figure 2, with each run spanning 10 ns. The number of native contacts NC and the radius of gyration R_g are shown in Figure 2(a). The loop and stem root-mean-squared atomic deviations (RMSDs) are plotted in Figure 2(b), showing the stability of the stem during all conformational changes in the loop subunit. The stem RMSD, which includes the fraying terminal base-pair, rarely reaches 1.5 Å. Figure 2(c) shows the change in base center-of-mass (COM) separation from the starting conformation (ΔD_{NC}) between native contact pairs. Loop contacts (top panel) show considerable structural variation, as discussed below. Sugar puckers for each of the 12 residues are shown in Figure 2(d), with residues 6 and 12 labeled for clarity, as no other nucleotides within the 12-mer are observed to undergo pucker conversion at 300 K.

C6 extensions and ribose pucker

Frequent transitions between closed-loop and open-loop conformations in the low temperature trajectories are observed. The stable closed-loop conformer is the dominant species in nearly all runs at 300 K. In runs 3, 7, 8 and 10 there is a loss of the C6-A7 contact, which defines the open-loop conformer. This conformer is seen in two of the ten refined NMR structures determined by Jucker *et al.*,[12] consistent with the closed-loop structure being more stable.

All runs in which the open-loop conformer persists included simultaneous transitions to the 2′-*endo* pucker mode (Figure 2(d)).† At numerous points in these trajectories, an open-loop conformer fails to persist and quickly returns to the closed-loop conformation (runs 1, 4, 6, and late in run 7). These failures appear in conjunction with a stable 3′-*endo* pucker that never undergoes conversion into the 2′-*endo* mode.

The correlation between the occurrence of 2′-*endo* conformations in C6 and structures with a solvent-exposed C6 residue was statistically evaluated. In both of the open-loop NMR structures proposed by Jucker *et al.*"[12] the base COM separation between C6 and A7 is over 5 Å greater than in corresponding closed-loop structures. Using this change in base separation, $\Delta D_{C6\text{-}A7} \geqslant 5$ Å, to define C6 as being completely looped-out results in a perfect correlation ($R^2 = 1.0$, $n = 53$) between these two structural features. Using significantly smaller base separation cutoffs of 4 Å and 3 Å continues to result in high correlations of 0.99 and 0.93, respectively. This suggests that the open-loop conformer requires a conversion to the 2′-*endo* mode prior to reaching the fully looped-out structure as suggested by two in ten refined NMR structures."[12]

When considering only structures in which the base separation is $\Delta D_{C6\text{-}A7} < 1$ Å from the starting structure, less than 3 % of closed-loop structures have a 2′-*endo* pucker and less than 1 % of structures with a base separation equal to or less than the starting structure are in 2′-*endo* conformations. These rare and extremely brief conversions to 2′ modes in closed-loop structures are not surprising in light of random solvent forces imparted upon loop nucleotides and the small energy barriers for sugar pucker conversion.[27]

The above correlations suggest a highly cooperative relationship between 3′-*endo* ⇌ 2′-endo and closed-loop ⇌ open-loop transitions in the GNRA tetraloop. Within our 300 K ensemble, $[\text{closed-loop}]_{C63'\text{-Endo}}$ structures compromise approximately 87 % of the collected structures (3483 in 4000, spanning 100 ns), consistent with the NMR results.[12] These cooperative base extension/pucker mode transitions may be mechanically driven (Scheme 2). The 2′-*endo* pucker modes expand the loop backbone,[27] and from simple geometric considerations it is reasonable to assume that these 2′-*endo* puckers also allow extension of the base away from the loop. Without this change in sugar pucker or more dramatic changes throughout the backbone, an extension of the base would require substantial deviation from tetrahedral (T_d) geometry around the C1′ carbon (Scheme 2).

Interestingly, Menger *et al.*[28] briefly considered the possibility that pucker conversions may be essential for transitions between loop conformers. Their discussion focused on the difference in timescales between the pucker conversions previously reported[26, "29] (1 to 10 ns, as observed herein) and the loop transitions that they monitor in microseconds. This timescale difference is most easily reconciled by considering the nature of the transitions being studied: simple stacking changes occur on sub-microsecond timescales,[28] below the limits of detection in their study.

A recent Raman spectroscopy study by Leulliot *et al.*[30] offers insight into our results. In our trajec-

† The timescale of observed pucker conformational shifts within our simulations agrees well with a previous computational study[26] that discusses similar 3′-*endo*⇌2′-*endo* transitions in the UUCG tetraloop. This correlation is particularly noteworthy when considering that an explicit solvent model is used in that study.

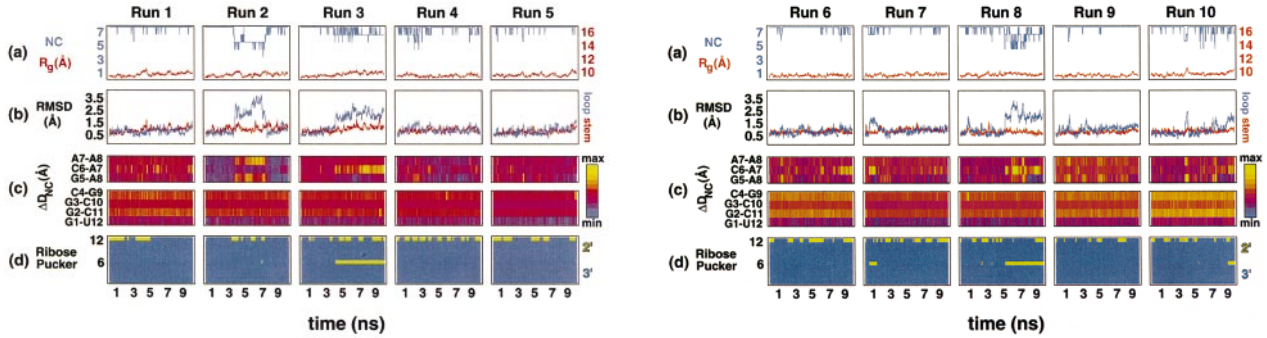
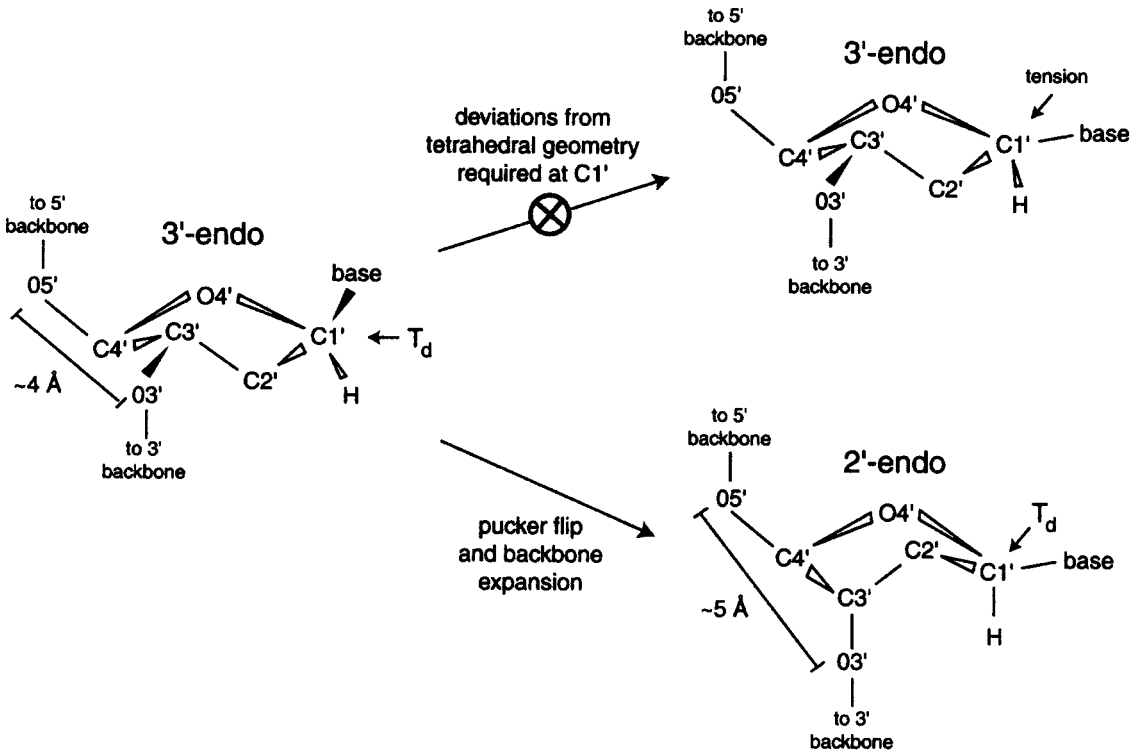


Figure 2. Characterization of the 300 K ensemble. (a) The number of native contacts (blue) and the radius of gyration (red). (b) Loop and stem RMSD values are shown in blue and red, respectively. (c) Changes in contact distances throughout the trajectories (ΔD_{NC}), where blue represents closest contact distance and yellow represents farthest. The values in these panels are scaled between the minimum and maximum values for each set of contacts and serve for comparative purposes only. Because of this scaling, identical colors in different panels cannot be assumed to represent the same absolute change in distance. The top panel shows the three loop contacts, and the bottom shows the four stem contacts. (d) The ribose pucker for each residue, with blue and yellow specifying the 3′-*endo* and 2′-*endo* pucker modes, respectively.

tories, all loop and stem G and C residues remained exclusively or predominantly 3′-*endo* and no glycosidic anti ⇌*syn* conformational changes were observed; this is in agreement with their results. They suggest, based on IR spectra, that at least one loop residue adopts a 2′-*endo* conformation and speculate that this may occur at A8, in disagreement with the refined NMR structures[12] and the mechanism of [closed-loop]$_{C63'\text{-endo}}$ ⇌ [open-loop]$_{C62'\text{-endo}}$ transitions observed in our 300 K data. A definitive determination of loop nucleotide pucker modes will offer insight into all of the methods previously used to examine this GCAA structural feature.

An important aspect of the Leulliot study is the 2′-*endo* shoulder seen in the UUCG tetraloop spectrum, which is assigned to the second loop U residue. This residue is in an extended, open-loop conformation analogous to the C6-extended open-loop structure we observe (Figure 1(b)). This supports our suggestion that transitions to looped-out



Scheme 2.

base conformations may generally occur cooperatively with transitions to 2′-*endo* pucker modes.

Within the family of GNRA loops, sequence-specific tertiary contact behavior is observed. For such loop structures, the position of highest sequence variability, position L2, determines receptor subtype preference,[31] and changes in secondary structure induced by tertiary structure formation have been reported.[16, 32] We speculate that the spatial variability at L2, the only location of significant conformational change within the loop in our simulations at 300 K, serves to facilitate such structural adaptations and interactions.

A8 Extensions and hydrogen bonding

In addition to the C6 extensions described above, residue A8 also leaves the loop briefly, breaking contacts with both G5 and A7 simultaneously (signaled by increased A7-A8 and G5·A8 contact distances in Figure 2(c)). Run 2 in Figure 2 represents the longest series of fluctuations in and out of this conformational state, with very brief occurrences in runs 6, 8, 9, and 10. If we define the A8-extended ensemble to include all structures in which both of these contacts are broken ($\Delta D_{NC} > 1.5$ Å), only ~5% of the 300 K conformations meet this criteria. Moreover, only 0.4% of the structures are included if a conservative contact distance increase of 4 Å is assumed. This is consistent with the inability to observe this conformer by NMR.[12] The ribose pucker of the A8 residue remains in a 3′-*endo* mode throughout all such transitions, and the C6-A7 stacking interaction at the top of the loop is conserved in this rare conformer (Figure 1(c)).

As proposed by Jucker *et al.*,[12] changes in hydrogen bonding pattern within the loop during conformational shifts are observed in our trajectories. An example of the dynamic nature of loop hydrogen bonding is shown in Figure 3, which details the fluctuations seen in run 2. Interatomic distances between the N7 of residue A7 ($A7_{N7}$) and its possible bonding partners are shown (Figure 3(b)). A hydrogen bond between the 2′-H of G5 ($G5_{2'\text{-}H}$) and $A7_{N7}$ was not previously described. However, in both our simulations and the NMR models[12] this proton directly attached to the 2′-carbon (Scheme 1) is near N7 of A7. Hydrogen bonding involving 2′-hydrogen atoms has been previously proposed based on analysis of crystal structures and prior simulations.[33] In our simulations, however, the average $A7_{N7}$ to $G5_{2'\text{-}H}$ separation of over 3 Å suggests that no hydrogen bonding activity is occurring.

Based on their refined NMR structures, Jucker *et al.*[12] have speculated that a hydrogen bond may be possible between the 2′-OH of G5 ($G5_{2'\text{-}OH}$) and $A7_{N7}$, and we observe a significant decrease in the average spacing between these potential bonding partners during both A8 extensions (Figure 3(b)) and C6 extensions (data not shown). Nevertheless, the mean separation (and fluctuation) for this pair

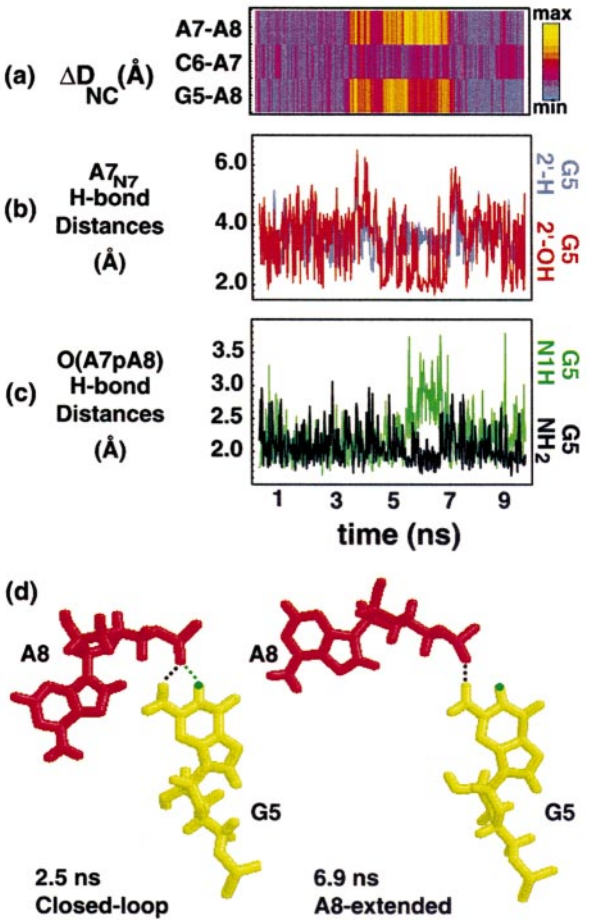


Figure 3. Changes in hydrogen bonding pattern during A8 extensions observed in run 2. (a) Changes in loop contact distances are shown to allow direct comparison to loop rearrangement events. (b) Interatomic distances between $A7_{N7}$ and its possible bonding partners: $G5_{2'\text{-}H}$ (blue) and $G5_{2'\text{-}OH}$ (red). (c) Hydrogen bonding distances for possible contacts formed between the $A7_PA8$ *pro*-S_P phosphoryl oxygen and both the nearest G5 amino proton (black trace) and the N1 hydrogen in that same base (green trace). (d) The two structures shown occurred at 2.5 ns (closed-loop) and 6.9 ns (A8-extended), respectively, and follow the coloring scheme in Figure 1. Hydrogen bond colors correspond to the traces in (c): the black dotted line indicates a contact between the *pro*-S_P phosphoryl oxygen (red) and the amine proton of G5; the green dotted line indicates contact between this same phosphoryl oxygen and the N1 proton of G5 (green-tipped).

in our most static closed-loop trajectory (run 1) is 3.27(±1.05) Å, exceeding the accepted hydrogen bond range. This result is consistent with the thermodynamic results reported by Santa Lucia *et al.*,[19] who found that H substitution at the 2′-OH position in G5 results in a $\Delta\Delta G^\circ$ of 0.3 kcal/mol. Although a direct comparison of the results regarding the proposed $A7_{N7}$ to $G5_{2'\text{-}OH}$ hydrogen bond is not possible, the simulations suggest that loss of this contact would not greatly destabilize the hair-

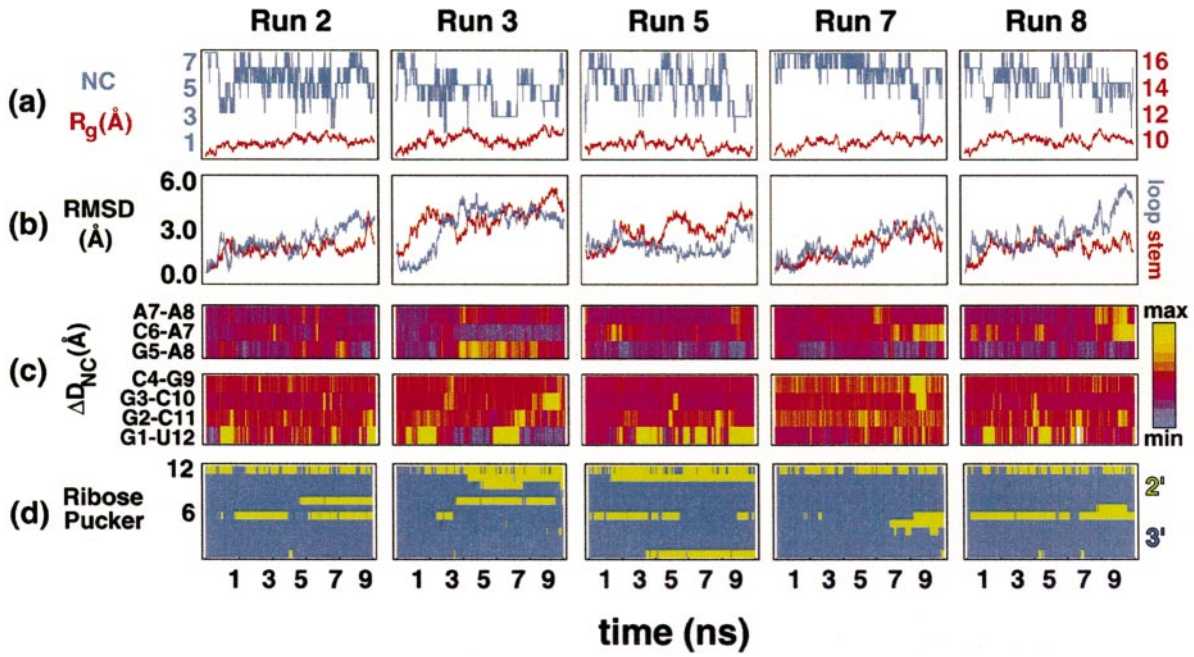


Figure 4. Characterization of the 400 K ensemble. Five of our ten 400 K trajectories are shown (following the format of Figure 2). (a) The number of native contacts (blue) and the radius of gyration (red). (b) Loop and stem RMSD values are shown in blue and red, respectively. (c) The change in contact distances throughout the trajectories (ΔD_{NC}), where blue represents closest contact distance and yellow represents farthest. The top panel shows the three loop contacts, with stem contacts in the lower panel. (d) Pucker modes for these five trajectories, with blue representing a 3′-*endo* (N-type) pucker and yellow representing a 2′-*endo* (*S*-type) mode. For clarity, the sixth and twelfth ribose puckers have been numbered to the left of the figure.

pin. Still, while the 2′-OH of G5 may not form a hydrogen bond in the most stable structure, it is interesting to consider that a hydrogen bond involving this 2′-OH could stabilize conformational excursions.

Figure 3(c) shows the hydrogen bonding distances for possible contacts formed between the $A7_PA8$ *pro*-S_P oxygen (i.e. the phosphoryl oxygen pointing toward the center of the loop) and both a proton of the amino group and the N1 proton of G5. Within the 300 K trajectories, many native closed-loop include bifurcated hydrogen bonding with nearly identical distances of 2.0-2.2 Å between these two donor-acceptor pairs. Interestingly, bifurcated mismatches at the stem-loop interface of tRNA anticodon hairpins have been previously reported for C·A base-pairs with very different overall structure."[34] Figure 3(d) shows the change in relative positioning of the G5 and A8 residues in going from the closed-loop conformer to the A8-extended conformation.

As the A8 residue extends, the C6-A7 stacking couple drops partially into the newly created void and the loop backbone flattens out significantly (Figure 1(c)). This allows only the amino proton to maintain contact with the phosphate oxygen, removing an interaction within the loop. This base extension and loss of the contact between the phosphoryl oxygen and the N1 proton, however, is accompanied by a decreased bond distance between the phosphate oxygen and the amino proton. This is in qualitative agreement with a previous report that substitution of an H in place of the G5 amino group, which removes the $A7_PA8$ phosphate oxygen to G5 amino proton hydrogen bond from the loop, results in a $\Delta\Delta G^\circ$ of ~0.7 kcal/mol.[19] In contrast, similar substitution in the UUCG loop results in a $\Delta\Delta G^\circ$ of ~1.3 kcal/mol. By removing the hydrogen bond to the amino proton, a strengthening of the bond to the N1 proton could potentially result in the small effect they report.

Menger *et al*. have proposed that a 5′-stacked loop structure exists in equilibrium on the microsecond timescale.[28] The A8-extended conformation seen herein could be a precursor to such 5′-stacked structures. As a stabilized A8-extended conformer has not been witnessed on the 100 ns timescale, this could rationalize the ''relatively low rate of conformational transition'' reported in their temperature-jump experiments.[28] Future experimental studies including 2-aminopurine substitution at the L2 position might provide a test of this model, as a dramatic decrease in fluorescence would be expected from stacking between G5 and A7.

At the L4 position within this tetraloop, A8 acts as part of the interface between the rigid stem and more flexible loop region, and we postulate that the favored C4·G9 closing base-pair imposes rigidity at the A8 position that the ribose ring cannot easily overcome on the nanosecond timescale at

low temperatures. This is supported by the relatively small torsion angle fluctuations shown in Table 1 for the closing base-pair. This, in concert with the partial loss of hydrogen bonding described above, could explain the relatively short-lived nature of this loop conformer.

Thermal perturbation and a model of loop dynamics

As all but one of the 400 K trajectories maintained stable or globular stem structures over the course of 10 ns, the sugar puckers and base orientations throughout the ten trajectories at this higher temperature have been examined for comparison to the 300 K ensemble. As in the 300 K data, no *anti* ⇌ *syn* glycosidic transitions were witnessed. Five of the ten 400 K trajectories are shown in Figure 4. The results of these simulations offer insight into, and support for, the arguments made above based on our 300 K ensemble. As predicted by the 300 K runs, all closed-loop ⇌ open-loop transitions occur cooperatively with C6 3′-*endo* ⇌ 2′-*endo* pucker conversions (early in run 5, late in run 7, and most of run 8). Furthermore, closed-loop ⇌ A8-extended transitions are seen at this higher temperature that clearly correlate with simultaneous 3′-*endo* to 2′-*endo* pucker conversions in the A8 ribose (runs 2 and 3; also runs 4 and 9, not shown).

This thermal perturbation has allowed us to examine an ensemble in which looped-out bases become much more prevalent. However, at some point in most of the 400 K trajectories the loop backbone loses native-like geometry and once the loop opens fully to solvent a variety of pucker and base orientations is easily accessible. Considering only the initial appearance of these A8-extended structures prior to loop dissociation, we again see an absolute correlation with the 2′-*endo* pucker, meaning that the initial loop rearrangements to A8-extended conformers all adopted 2′-*endo* pucker modes during the transition.

An extension of residue A7 away from the loop late in run 8 is also observed and, as is the case for other loop residues, a 3′-*endo*⇌2′-*endo* conversion occurs in tandem. Run 8 featured one of the more slowly degenerating hairpin structures, with the stem RMSD remaining under ~ 3 Å, thus allowing this comparison late in the simulation. This again supports the observation that a 2′*endo* pucker mode is essential in stabilizing extended, solvent-exposed loop bases.

Based on the above results, we propose that an equilibrium of at least three loop structures exists in the GCAA loop on sub-microsecond timescales (see Figure 1). Our model consists of a [closed-loop]$_{\text{C6 3′-endo}}$⇌[open-loop]$_{\text{C6 2′-endo}}$ transition and a less frequent [closed-loop]$_{\text{A8 3′-endo}}$⇌[A8-extended]$_{\text{A8 3′-endo}}$ transition, with the A8-extended conformer existing only for bursts that are short on current experimental timescales. This model will serve as both a starting point and an element of comparison for modeling other GNRA tetraloops, as well as the previously noted UNCG loops, which are similar in structure.[6]

Thermal unfolding dynamics

To enhance our sampling of configurations of the hairpin, we have examined unfolding at a variety of temperatures. Although it is experimentally observed that folding of this hairpin takes place as a two-state system event,[17, 30] a globular intermediate (described below) that occurs in transitions between the folded and unfolded regions of configurational space was detected.† The relative conformational probabilities observed in the ensemble of 500 K unfolding trajectories as a function of the total hairpin RMSD and the number of native contacts within each given structure is shown in Figure 5.‡ In the context of hairpin folding and unfolding dynamics, three probability maxima are easily identified, and are labeled as the unfolded (U), globular intermediate (G), and folded (F) states of the tetraloop hairpin: F consists of structures with high NC and low RMSD; G consists of conformations with RMSD values between ~3 Å and ~6 Å and two to four native contacts; U consists of structures with less than two native contacts and RMSD values of greater than 6 Å. Because the

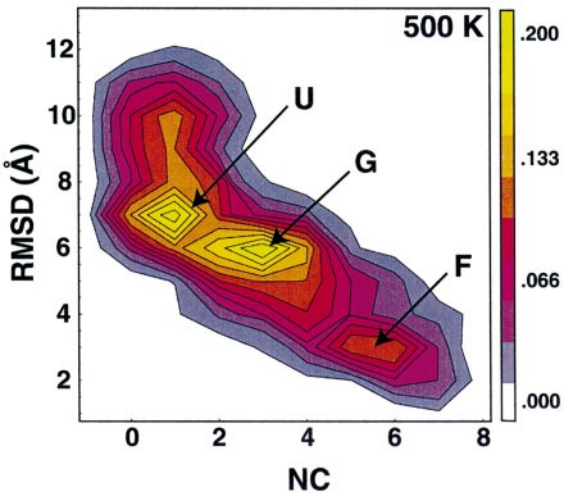


Figure 5. Observed relative conformational probabilities at 500 K. The unfolded (U), globular (G), and folded (F) states are labeled on the probability surface created by the RMSD and NC degrees of freedom. Data from ten 500 K trajectories are included in the plot, giving a total of 4330 structures over a total sampling time of 65 ns.

† Two-state kinetics found experimentally means that there is a single rate-limiting step, but does not imply that there are not additional metastable states along the folding pathway.

‡ As these trajectories have not crossed all relevant free energy barriers numerous times, this plot does not represent equilibrium conformational probabilities.

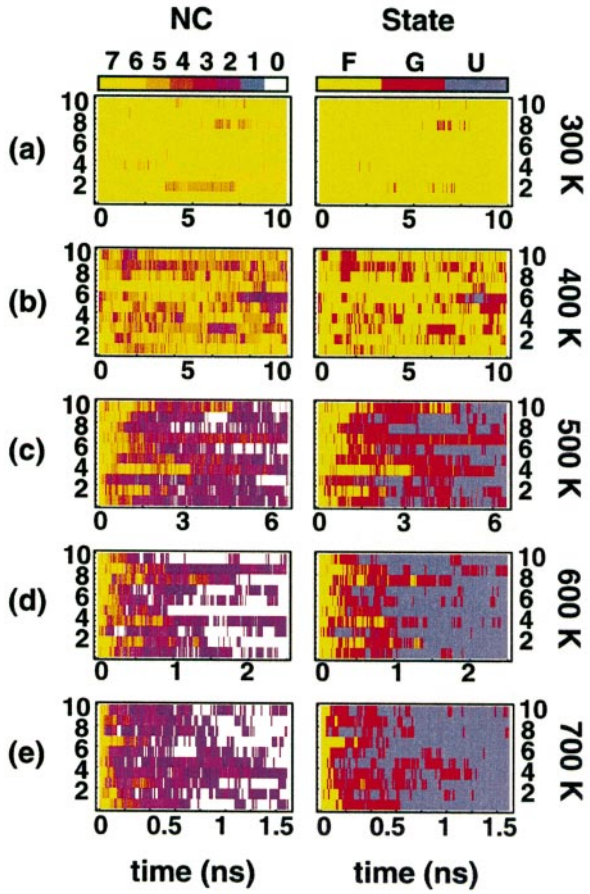


Figure 6. A summary of the 50 simulations analyzed. All ten runs at each of the five temperatures studied are shown (run number *versus* time in ns for each temperature block). The 300 K trajectories are shown in the top row (a), with T increasing in 100 K increments down each row to 700 K (e). Left: the number of native contacts for each trajectory ranging from yellow (NC = 7) to white (NC = 0). Right: simplified trajectory representations based on the three-state model shown in Figure 5.

400 K trajectories did not unfold completely and the higher temperature trajectories unfolded rapidly, the 500 K data set best represents an ensemble of unfolding events.

A total of 50 trajectories were collected at temperatures from 300 to 700 K, as outlined in Figure 6, which shows the number of native contacts and the simplified trajectory representations throughout each trajectory. The globular state was sampled very briefly in only four of our 300 K trajectories, with this classification being the result of terminal fraying and the A8 extensions described above. Of the 400 K ensemble, only one brief period of complete unfolding was witnessed. At 500 K full unfolding was observed in nearly all of our simulations.

Folding pathway from simulated unfolding

Figure 7 details one trajectory from each of the data sets at 500 K (run 4), 600 K (run 8), and 700 K (run 1). The globular intermediate is detectable in nearly all of the 500 K trajectories and most of the 600 K trajectories, whereas the 700 K trajectories appear much more continuous with respect to all degrees of freedom monitored, suggesting that at higher temperatures the barrier separating the U and G states is not present to any appreciable extent. Run 4 of the 500 K series (first row in Figure 7, see also Figure 8) shows a very discrete jump to a globular conformation that persists until near the end of the run when the fully unfolded state is reached. A similar but less discrete event is observed in the 600 K run, with both trajectories shown manifesting fairly constant RMSD values during this period, regardless of the lifetime of the intermediate.

Analysis of these 30 trajectories has led to a clear unfolding scheme: the general unfolding pathway of the GCAA tetraloop hairpin witnessed in this study is shown in Figure 8. From the folded state (Figure 8(a)), the loss of terminal and loop interactions and flattening of the loop backbone occurs first. The mismatched terminal base-pair exhibits some fraying at equilibrium and the slight instability inherent to this pair makes it one of the first structural locations to degenerate. As described

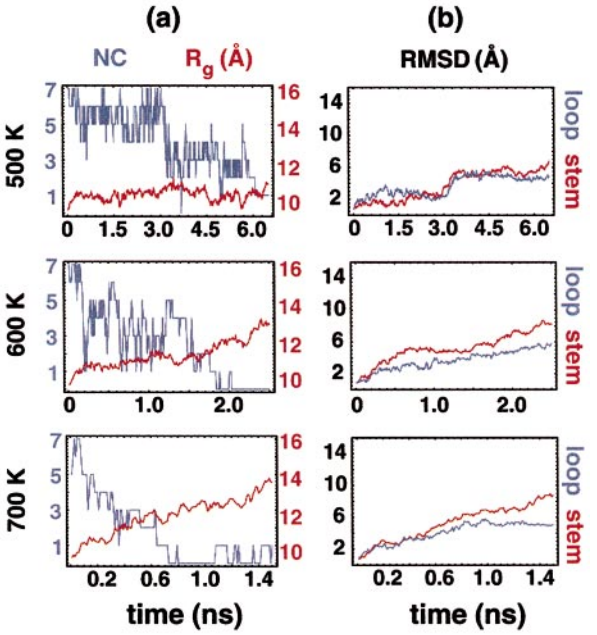


Figure 7. Characterization of high temperature unfolding trajectories. Shown in descending order is one representative run from the 500 K (run 4, see also Figure 8), 600 K (run 8), and 700 K (run 1) unfolding ensembles. (a) The number of native contacts (blue) and radius of gyration (red). (b) The loop (blue) and stem (red) RMSD values are equally scaled to show relative changes between simulations.

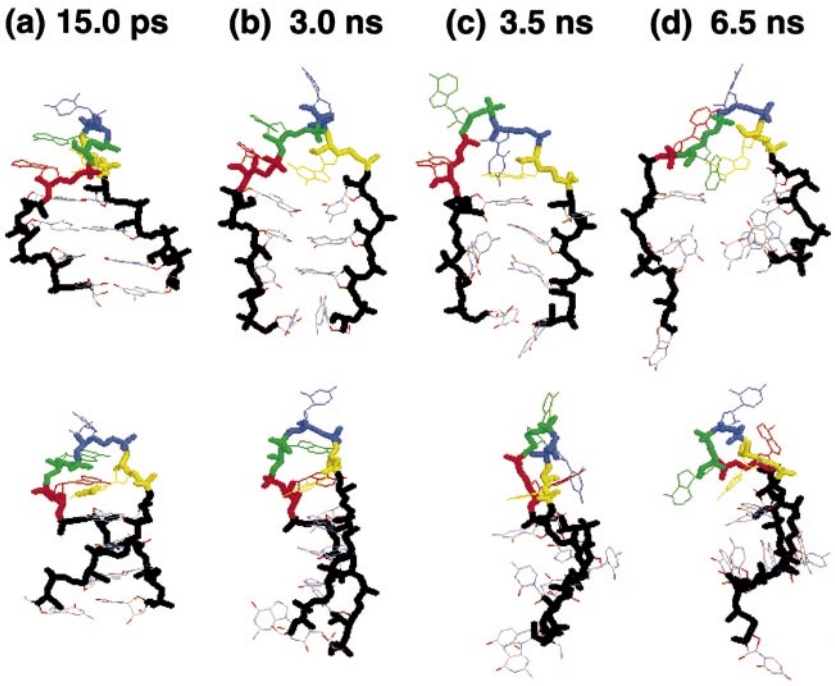


Figure 8. Observed unfolding pathway. The pathway followed by run 4 in the 500 K series (as outlined quantitatively in Figure 7, top row) is shown. The phosphate-ribose backbones of the stem are displayed as black sticks for clarity, with the loop backbone and bases colored according to the same scheme used in Figure 1 (from 5′ to 3′: G5 in red, C6 in blue, A7 in green, and A8 in red). The lower panel of images shows each structure rotated ~90° to display helicity (or lack thereof) within each structure.

above, the tetraloop is itself relatively flexible and undergoes two significant, if brief, conformational changes on the nanosecond timescale at lower temperatures. This flexibility contributes to the early loss of stacking and/or base-pairing interactions in the loop. Initial contact loss occurs both at the terminal base-pair and in the loop region, suggesting that both ends of this hairpin are roughly equally susceptible to perturbation. Helicity is maintained only near the center of the hairpin in this initial step, a result of the remaining stem base-pair contacts (Figure 8(b)). These central base-pair contacts then break while the stem backbone flattens out, resulting in a mostly planar, symmetrically bent structure (Figure 8(c)). Fraying of the terminal ends and loop region are highly variable between trajectories. Finally, the structure relaxes slowly (relative to the initial stages of unfolding) and bases begin to randomly sample various solvated orientations (Figure 8(d)). The mostly linear strands of the hairpin then begin diffusing away from each other.

If we assume that the folding of this fragment is the reverse of the unfolding trajectories,[24] the folding mechanism is straightforward. Initial collapse of the unfolded structures would precede sampling of the globular state and would take place rather quickly, stabilized by random stacking and base-pairing interactions. Next, from these globular intermediates, the hairpin crosses over the barrier into the correctly folded region of configurational space, which could only occur after the "core" of the hairpin (i.e. the central base-pairs of the stem) had properly formed. Upon formation of the central base-pairs, stem helicity would be partially complete. Proper stacking and base pairing in the loop in tandem with finalization of *A*-form helicity would constitute the end of the event.

Kinetics of unfolding and folding

At each temperature in the 400 to 700 K range, the fraction of the population still folded (NC $\geqslant$ 5) was fit to a single exponential function of simulation time (not shown). Only data between the starting point ($t_T = 0$) and the point at which all of the simulations at a given temperature had entered the globular state ($t_{400\ K} \approx 10$ ns; $t_{500\ K} \approx 5$ ns; $t_{600\ K} \approx 1$ ns; $t_{700\ K} \approx 0.25$ ns) were included in these fits. The resulting temperature-dependent unfolding rates $k_u(T)$ are shown in Figure 9. From an Arrhenius fit (Figure 9, inset), the unfolding rate constant at 300 K was approximated as k_u (300 K) $\approx 0.1\ \mu s^{-1}$.

To approximate upper and lower bounds on this unfolding rate and to estimate the folding rate at 300 K, the extreme linear fits produced by the small data set were calculated (Figure 9, inset: broken lines). This results in lower and upper bounds on k_u(300 K) of $\sim 0.03\ \mu s^{-1}$ and $\sim 1\ \mu s^{-1}$. From the known free energy change of folding for this hairpin, an equilibrium constant of $K_{eq} \sim 700$ requires a difference between the folding and unfolding rates of approximately three orders of magnitude. The resulting bounds on the folding rate k_f(300 K) based on these simulations are then on the order of 30 μs^{-1} and 1000 μs^{-1}. Considering the estimated rate of base-pair formation in nucleic acid helices[35] of $\sim$ 4 μs^{-1}, a reasonable approximation of the folding of this small hairpin might be on the 10 μs timescale, as predicted for slightly larger DNA hairpins.[36,37]

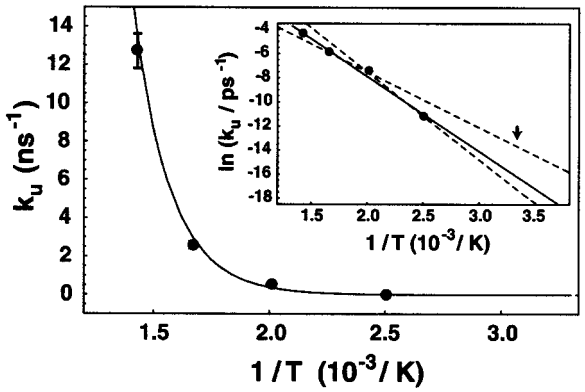


Figure 9. Approximation of the unfolding rate at 300 K. The unfolding rate k_u at each temperature is plotted against inverse temperature and fit to a single exponential. Only the 700 K rate approximation had an error beyond the point size shown. Inset: approximation of the temperature-dependent unfolding rate of the 12-mer hairpin by a linear fitting (continuous line) of the natural logarithm of the rate constant *versus* inverse temperature for the 400 to 700 K series. The error inherent to each data point falls within the employed point size. Broken lines represent the upper and lower bounds on our approximation and the arrow indicates the range of approximate extrapolated values at 300 K.

This is slower than our lower bound and this difference may be accounted for by the absence of explicit counterions in simulation, since they would likely slow unfolding and thus lower the predicted rate of folding; a temperature dependence of unfolding that is not single-exponential in form, as is assumed above; the use of the course-grained NC reaction coordinate to predict rates; or lack of sufficient data points in our small sample to accurately predict the 300 K unfolding rate. With these possible sources of error in mind, our apparent lack of accuracy is understandable, and we speculate that the calculated rates would be in much better agreement with experiment were such factors accounted for.

Conclusion

We have used stochastic dynamics simulations to sample a total of 100 ns of GCAA tetraloop hairpin configurations at 300 K. These trajectories have led to the proposal of an equilibrium of three loop conformations consistent with current experimentally determined structural properties: closed-loop structures (C6 stacked above the 3′ strand with A7 and A8) are dominant, open-loop conformers (C6 extended into solvent) make up ~15-20 % of the ensemble, and transient A8-extended conformations occur rarely and lack stability. In addition, the second loop nucleotide (C6 in GCAA) undergoes cooperative 3′-*endo* ⇌ 2′-*endo* sugar pucker conversions and closed-loop ⇌ open-loop transitions consistent with the NMR-derived conformers.[12] An equivalent simulated ensemble at 400 K supports this model of cooperativity between transitions to solvent-exposed base conformers and transitions to *S*-type pucker conformations, which is seen in other loop nucleotides (both A7 and A8). Experimental data on UUCG tetraloops, which are similar in structure, are consistent with this model.[30]

In regards to the folding dynamics of nucleic acid hairpins, we have analyzed 30 trajectories which span a temperature range from 500 to 700 K and observed a single, generic ''ends-first'' unfolding mechanism which occurs in three recognizable steps: (a) terminal base-pair and loop contacts are lost; (b) the breaking of central base-pairs occurs cooperatively with a loss of stem helicity; and (c) random orientation/extension of the disconnected strands is gained *via* diffusion. From the observed unfolding pathway, a glimpse of the folding event has been obtained. We propose a folding pathway that includes (a) collapse of the unfolded 12-mer and sampling of the globular state, (b) a passage into the folded region of configurational space as the central base-pairs of the stem form properly and gain helicity, and (c) pairing of the terminal base pair and proper stacking and base-pairing in the loop in tandem with achievement of full *A*-form helicity. This model of hairpin formation is in good agreement with a recently reported statistical mechanical model of DNA hairpin folding.[36]

The stability of this RNA hairpin in our simulations and the agreement observed between our trajectories and experimental data on this and other relevant systems lends credence to this method of RNA simulation. However, addressing the inclusion of an explicitly ionic solvent and the accuracy of the rate calculations should be a future focus of atomistic RNA simulation. In particular, the timescales investigated here suggest that it may be possible to conduct (a) atomistic studies of the changes in secondary structure upon tertiary contact formation, (b) simulations of the direct folding of this and similar sequences *in silico*, and (c) studies of the equilibrium thermodynamic properties of loop structures at longer (μs) timescales.

Materials and Methods

All simulations discussed herein employed the Amber95/AmberN force field for nucleic acids"[38] and the GB/SA continuum solvent model."[39] For solvent energy and force calculations a surface area prefactor of 7.0 cal/mol Å^2 and a probe sphere radius of 1.4 Å for surface area calculations were used, with the solvent polarization and surface area derivatives (forces) being calculated at every dynamic timestep. As simulation studies do not require electroneutrality, our calculations included 389 atoms with no explicit water molecules, ions, or periodic boundaries. We assume that our ion-free environment can approximate the low sodium behavior of this fragment, noting that previous molecular dynamics simulations of the $tRNA^{Asp}$ anticodon hairpin have strongly suggested that sodium ions lack the bind-

ing specificity seen in inorganic ions such as ammonium."[40] A 2.0 fs integration step was used and the Rattle algorithm[41] was employed to constrain only H-X covalent bonds (where X represents heavy atoms). No electrostatic cutoffs or tapers were used.

The production runs presented herein were started from the same configuration, which was the result of a 1 ns molecular dynamics relaxation period performed on the first (closed-loop) model of the most recent NMR work (PDB ID no. 1ZIH[12]). This relaxation run employed Beeman integration (zero solvent-induced friction) at 300 K. The resulting relaxed structure had an RMSD = 1.62 (±0.16) Å from the ensemble of ten NMR models and maintained all structural and interactive characteristics of the refined NMR structures.

Production runs were conducted on the SGI Origin 2000 supercomputer at Stanford, CA using Allen's stochastic integrator[42] within the Tinker 3.8.† package. To represent the physics of real solvents, the force on a Newtonian particle can be considered as the sum of random thermal fluctuations (***R***) and viscous drag (with friction coefficient γ), resulting in the three-dimensional, many-particle Langevin equation:

$$m \times \ddot{\vec{r}}(i) + \gamma \dot{\vec{r}}(i) + \vec{R}(i) = 0. \quad (1)$$

Numerical integration of this differential equation results in the following equations of motion, with a time interval δt

$$\vec{r}(t_{i+1}) = \vec{r}(t_i) + c_1 \delta t \dot{\vec{r}}(t_i) + c_2 \delta t^2 \ddot{\vec{r}}(t_i) + \vec{R}_r, \quad (2)$$

$$\dot{\vec{r}}(t_{i+1}) = c_0 \dot{\vec{r}}(t_i) + c_1 \delta t \ddot{\vec{r}}(t_i) + \vec{R}_v. \quad (3)$$

Here R_r and R_v are randomly chosen from a temperature-dependent bivariate Gaussian distribution at each timestep, and the integration coefficients are given by:

$$c_0 = e^{-\gamma \cdot \delta t}, \quad c_1 = \frac{1 - c_0}{\gamma \delta t}, \quad \text{and} \quad c_2 = \frac{1 - c_1}{\gamma \delta t}. \quad (4)$$

Assuming that the proper friction coefficient is used to characterize the desired solvent, the addition of a stochastic force term in this algorithm does not induce deviations from the simulation times expected of analogous molecular dynamics simulations. Although early simulations[43–45] of aqueous systems included friction coefficient (γ) values over the wide range of 50-200 ps^{-1}, proper modeling of solvent effects requires consideration of the relaxation time of the solvent as given by the velocity autocorrelation function:

$$\langle \dot{r}(t) \dot{r}(0) \rangle \propto e^{-\gamma \cdot t}. \quad (5)$$

For water, the accepted decorrelation time of ~10 fs gives $\gamma \cong 100$ ps^{-1}, which is in agreement with a recent review of implicit solvent models by Cramer and Truhlar.[46] Here we have used $\gamma = 91$ ps^{-1}, which has previously yielded quantitatively correct rates in simulations of (protein) β-hairpin folding.[25] These production runs thus included both energetic (GB/SA) and viscous (Langevin) representations of the aqueous environment.

† http://www.dasher.wustl.edu/tinker/

‡ http://www.biochem.ucl.ac.uk/~martin/swreg.html

Ten trajectories were collected at each of five temperatures and frames were saved every 25 ps at 300 K, 20 ps at 400 K, 15 ps at 500 K, and 10 ps for both 600 K and 700 K. The final data set consists of: 10 ns for each 300 K run, 10 ns for each 400 K run, 6.5 ns for each 500 K run, 2.5 ns for each 600 K run, and 1.5 ns for each 700 K run, giving composite simulation times of 100 ns at 300 K and over 0.3 μs in all. Each of these 50 trajectories was started with a unique random number seed to initiate atomic velocities according to the relevant Boltzmann distribution and was therefore started from a unique location in phase space $P(\boldsymbol{r}^N, \boldsymbol{p}^N)$.

Several parameters were calculated for each trajectory. The total, loop, and stem (including terminal base-pair) RMSD values were calculated using the academic distribution ProFit.‡ The radius of gyration R_g was calculated using only heavy atoms and included the terminal base-pair. Total potential GB energies were calculated using the analyze routine in Tinker 3.8 and the analogous Poisson-Boltzmann based energy was calculated using analyze with an interface to the ZAP algorithm,[47] with both employing a solute inner dielectric of 1. Base center of mass assignments included only the six-membered ring in each nucleotide for symmetry, and we defined a native contact as existing if the distance between two COMs was less than the starting distance (native COM gap) plus 1.5 Å. As is demonstrated in the schematics shown in Figure 1, we considered seven native contacts (shown in blue between the relative bases) in the folded closed-loop structure and six in the native open-loop structure (note that fraying of the terminal pair can remove a NC from the total periodically). By definition, any contact that is ''broken'' allows passage of water molecules between the relevant bases in the broken pair.

An assessment of the application of the GB/SA solvation model to GCAA tetraloop hairpin simulations was conducted prior to collection of production runs. A fit of the GB energies to the expectedly more accurate PB energies[39] was calculated for a preliminary data set consisting of five trajectories comprising a total of 1662 individual configurations, with one run at each of the temperatures employed here. The resulting PB to GB linear regression returned an R^2 value of 0.997. As has been discussed elsewhere,[48–52] it would appear that the relative accuracy of the GB model makes this an acceptable approximation of the solvation energy term for DNA and RNA secondary structure simulations and allows for successful simulation of small nucleic acid molecules in the absence of an explicit solvent model.

Acknowledgments

We thank Stefan Larson, Bojan Zagrovic, Rhiju Das, and the rest of the Pande and Herschlag groups for their helpful discussions regarding this manuscript, and offer appreciation to our referees for their valuable insight. This work was supported in part by ACS PRF grant 36028-AC4 and CPIMA seed funds. D.H. acknowledges NIH grant (GM49243). M.A.E. is a NSERC postgraduate fellow.

References

1. Doherty, E. & Doudna, J. (2001). Ribozyme structures and mechanisms. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 457-475.
2. Uhlenbeck, O. C. (1990). Nucleic-acid structure-tetraloops and RNA folding. *Nature,* **346**, 613-614.
3. Woese, C. R., Winker, S. & Gutell, R. R. (1990). Architecture of ribosomal RNA: constraints on the sequence of ''tetra-loops''. *Proc. Natl Acad. Sci. USA,* **87**, 8467-8471.
4. Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276-1285.
5. Brion, P. & Westhof, E. (1997). Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113-137.
6. Moore, P. B. (1999). Structural motifs in RNA. *Annu. Rev. Biochem.* **68**, 287-300.
7. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E. *et al.* (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science,* **273**, 1678-1685.
8. Ferré-D'Amaré, A. R. & Doudna, J. A. (1999). RNA folds: insights from recent crystal structures. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 57-73.
9. Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature,* **372**, 111-113.
10. Endo, Y., Chan, Y. L., Lin, A., Tsurugi, K. & Wool, I. G. (1988). The cytotoxins α-sarcin and ricin retain their specificity when tested on a synthetic oligoribonucleotide (35-mer) that mimics a region of 28S ribosomal ribonucleotide acids. *J. Biol. Chem.* **263**, 7917-7920.
11. Gluck, A., Endo, Y. & Wool, I. G. (1992). Ribosomal-RNA identity elements for ricin A-chain recognition and catalysis-analysis with tetraloop mutants. *J. Mol. Biol.* **226**, 411-424.
12. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M. & Pardi, A. (1996). A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **264**, 968-980.
13. Jaeger, L., Michel, F. & Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range tertiary interactions. *J. Mol. Biol.* **236**, 1271-1276.
14. Murphy, F. L. & Cech, T. R. (1994). GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.* **236**, 49-63.
15. Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. (2001). A universal mode of helix packing in RNA. *Nature Struct. Biol.* **8**, 339-343.
16. Costa, M. & Michel, F. (1997). Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.* **16**, 3289-3302.
17. Heus, H. A. & Pardi, A. (1990). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science,* **253**, 191-194.
18. Zichi, D. A. (1995). Molecular dynamics of RNA with the OPLS force field: aqueous simulation of a hairpin containing a tetranucleotide loop. *J. Amer. Chem. Soc.* **117**, 2957-2969.
19. Santa Lucia, J., Jr, Kierzek, R. & Turner, D. H. (1992). Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science,* **256**, 217-219.
20. Brooks, C. L., III (1998). Simulations of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **8**, 222-226.
21. Pande, V. S. & Rokhsar, D. S. (1999). Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl Acad. Sci. USA,* **96**, 9062-9067.
22. Daggett, V. & Levitt, M. (1993). Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **232**, 600-619.
23. Lazaridis, T. & Karplus, M. (1997). ''New view'' of protein folding reconciled with the old through multiple unfolding simulations. *Science,* **278**, 1928-1931.
24. Dinner, A. & Karplus, M. (1999). Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* **292**, 403-419.
25. Zagrovic, B., Sorin, E. J. & Pande, V. S. (2001). β-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313**, 151-169.
26. Nowakowski, J., Miller, J. L., Kollman, P. A. & Tinoco, I., Jr. (1996). Time evolution of NMR proton chemical shifts of an RNA hairpin during a molecular dynamics simulation. *J. Am. Chem. Soc.* **118**, 12812-12820.
27. Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag. New York.
28. Menger, M., Eckstein, F. & Porschke, D. (2000). Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry,* **39**, 4500-4507.
29. Hermann, T., Auffinger, P. & Westhof, E. (1998). Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* **27**, 153-165.
30. Leulliot, N., Baumruk, V., Abdelkafi, M., Turpin, P., Namane, A., Gouyette, C. *et al.* (1999). Unusual nucleotide conformations in GNRA and UNCG type tetraloop hairpins: evidence from Raman markers assignments. *Nucl. Acids Research,* **27**, 1398-1404.
31. Thirumalai, D. (1998). Native secondary structure formation in RNA may be a slave to tertiary folding. *Proc. Natl Acad. Sci. USA,* **95**, 11506-11508.
32. Wu, M. & Tinoco, I., Jr. (1998). RNA folding causes secondary structure rearrangement. *Proc. Natl Acad. Sci. USA,* **95**, 11555-11560.
33. Auffinger, P. & Westhof, E. (1997). Rules governing the orientation of the 20-hydroxyl group in RNA. *J. Mol. Biol.* **274**, 54-63.
34. Auffinger, P. & Westhof, E. (1999). Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.* **292**, 467-483.
35. Cohen, R. J. & Crothers, D. M. (1971). Rate of unwinding small DNA. *J. Mol. Biol.* **61**, 525-542.
36. Ansari, A., Kuznetsov, S. V. & Shen, Y. (2001). Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl Acad. Sci. USA,* **98**, 7773-7776.
37. Shen, Y., Kuznetsov, S. V. & Ansari, A. (2001). Loop dependence of the dynamics of DNA hairpins. *J. Phys. Chem. B,* **105**, 12202-12211.
38. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.
39. Qui, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. (1997). The GB/SA continuum model for solvation. A fast analytical method for the calculation for approximate Born radii. *J. Phys. Chem.* **101**, 3005-3014.

40. Auffinger, P. & Westhof, E. (1998). Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* **8**, 227-236.
41. Andersen, H. C. (1983). Rattle: a ''velocity'' version of the Shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24-34.
42. Allen, M. P. (1980). Brownian dynamics simulation of a chemical reaction in solution. *Mol. Phys.* **40**, 1073-1087.
43. Brooks, C. L., III & Karplus, M. (1983). Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* **79**, 6312-6325.
44. Brünger, A., Brooks, C. L., III & Karplus, M. (1984). Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Letters,* **105**, 495-500.
45. Brünger, A., Brooks, C. L., III & Karplus, M. (1985). Active site dynamics of ribonuclease. *Proc. Natl Acad. Sci. USA,* **82**, 8458-8462.
46. Cramer, C. J. & Truhlar, D. G. (1999). Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev,* **99**, 2161-2200.
47. Grant, J. A., Pickup, B. T. & Nicholls, A. (2000). A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comp. Chem.* **22**, 608-640.
48. Zacharias, M. (2000). Simulation of the structure and dynamics of nonhelical RNA motifs. *Curr. Opin. Struct. Biol.* **10**, 311-317.
49. Zacharias, M. (2001). Conformational analysis of DNA-trinucleotide-hairpin-loop structures using a continuum solvent model. *Biophys. J.* **80**, 2350-2363.
50. Srinivasan, J., Miller, J., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dynam.* **16**, 671-682.
51. Williams, D. J. & Hall, K. B. (1999). Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys. J.* **76**, 3192-3205.
52. Williams, D. J. & Hall, K. B. (2000). Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J. Mol. Biol.* **297**, 1045-1061.

Edited by J. Dovdna

(Received 18 September 2001; received in revised form 17 January 2002; accepted 18 January 2002)

http://www.academicpress.com/jmb

Supplementary Material comprising @ 35 unpublished trajectories is available on IDEAL