J. Mol. Biol. (1997) **270**, 436–450

JMB

Theoretical Studies of an Exceptionally Stable RNA Tetraloop: Observation of Convergence from an Incorrect NMR Structure to the Correct One Using Unrestrained Molecular Dynamics

Jennifer L. Miller and Peter A. Kollman*

Department of Pharmaceutical Chemistry, School of Pharmacy University of California San Francisco, San Francisco CA 94143-0446, USA

**Corresponding author*

We report on the results of five independent and unrestrained molecular dynamics simulations of an RNA tetraloop, r(GGACUUCGGUCC), and its related structures with the loop UUCG sugars changed to deoxyribose. Two separate NMR structures have been reported for the loop portion of this molecule, with the second refinement resulting in a slightly different and more accurate conformation for the loop. The root-mean-square deviation (RMSd) between the two NMR structures, for the loop portions only, is 2.5 Å. Our simulations, starting from the two NMR structures, demonstrate that this tetraloop is a very stable and rigid structure with both nanosecond length simulations staying very close to the initial structures. Additionally, both simulations preserved most, if not all, of the NMR-derived interactions and violated very few of the nuclear Overhauser effect (NOE)-derived distances used in the structure refinements. However, when the two NMR structures were simulated with deoxyriboses in the loops instead of the native riboses, the flexibility of the systems increased and we observed a conversion from the incorrect to the correct loop conformation in the simulation which started in the incorrect loop conformation. When the riboses were subsequently re-introduced back into the structure which underwent the conversion, the agreement between this simulation and the one starting from the correct NMR structure was a remarkably low 0.5 Å, demonstrating an almost complete convergence from the incorrect to the correct structure using unrestrained molecular dynamics.

Keywords: molecular dynamics; nucleic acids; RNA; tetraloop; conformational transition

Introduction

RNA is an essential component in a wide array of cellular processes. It plays a central role as the primer in DNA synthesis, as the carrier of the genetic message during transcription and as the cellular shuttle for the gene-encoded amino acids during translation. This ultimately results in the production of proteins. RNA has both structural and functional roles in ribosomes and spliceosomes, and has been found to possess catalytic activity, for example, in certain group I and group II introns and in RNase P. In retroviruses such as RNA tumor viruses and HIV, RNA is the primary genetic material. The structure of RNA is as varied as its functions, with the single-stranded nucleic acid folding into complex structures not unlike that of proteins. Hairpins are the most abundant element of RNA secondary structure (Woese *et al.*, 1990). They form when the single-stranded RNA molecule folds back on itself to form a double-helical stem capped by a loop of unpaired nucleotides of various sizes. In addition to playing a part in many of the cellular processes listed above, RNA hairpins are thought to provide nucleation sites for RNA folding (Tuerk *et al.*, 1998) and tertiary recognition sites for both proteins and nucleic acids (Varani, 1995). Because of their importance to both RNA structure and function, hairpins have been

Abbreviations used: RMSd, root-mean-square deviation; NOE, nuclear Overhauser effect; MD, molecular dynamics; 2D, two-dimensional; ID, one-dimensional.

0022–2836/97/280436–15 $25.00/0/mb971113

the focus of recent structural studies (Wimberly *et al.*, 1991). These studies have resulted in high-resolution solution structures for three of the most frequently occurring RNA hairpin tetraloops: GCAA (Heus & Pardi, 1991), UUCG (Allain & Varani, 1995; Varani *et al.*, 1991), and CUUG (Jucker & Pardi, 1995).

A tetraloop is an RNA hairpin which contains four unpaired bases in the loop. These motifs are most prevalent in rRNA, accounting for 55% of all hairpins in the small 16 S subunit and 38% in the large 23 S subunit. Sequence comparison studies have shown that these tetraloops are highly constrained in sequence, with most falling into three families: UNCG, GNRA, and CUUG (N is any nucleotide, R is either adenosine or guanosine; Woese *et al.*, 1990). Thermodynamic studies have classified these families as ''unusually stable'' because their melting temperatures are much higher than similar RNA sequences possessing normal thermodynamic stability (Antao *et al.*, 1991; Antao & Tinoco, 1992; Tuerk *et al.*, 1998). Besides their obvious biological importance, the remarkable stability and ubiquity of these tetraloops has led to their use in many experimental studies. For example, they are used as markers for NMR structure determinations (Allain & Varani, 1995; Molinaro & Tinoco, 1995), as caps to shorten helical regions in X-ray studies (Pley *et al.*, 1994) and to ensure proper folding in anti-sense studies (Noonberg *et al.*, 1994).

Despite structural studies to date, an atomic level understanding of the remarkable stability of the UNCG tetraloop is still unresolved. Thermodynamic and spectroscopic studies of Sakata *et al.* (1990) indicated that the ribose 2′-hydroxyl groups in the loop are primarily responsible for the unusual stability. Early NMR studies (Cheong *et al.*, 1990; Varani *et al.*, 1991) were not able to determine the nature of the 2′-OH interactions in the loop, although they identified that at least two of the first three hydroxyl groups show a reduced rate of exchange with solvent. In a subsequent NMR study of the P1 helix from group I self-splicing introns, this tetraloop was used to cap the helix and assist in the NMR structure determination (Allain & Varani, 1995). During their refinement, Allain & Varani (1995) discovered that in the previous structure determination of the smaller UNCG stem-loop system a single hydroxyl group resonance had been misassigned as an amino resonance. The corrected assignment led to a slightly different loop structure and provides some evidence that the hydroxyl group of the first loop residue (U) is engaging in a stabilizing hydrogen bond to the closing base of the loop (G). There have been two reported attempts to solve the crystal structure of this system (Cruse *et al.*, 1994; Holbrook *et al.*, 1991), but the conditions and/or packing effects unfold the tetraloop and the resulting structure is an *A*-form double helix with mismatched base-pairs in the middle of the helix.

The two NMR solution structures, while differing in the loop conformations, share many structural features (Figure 1(a)). The main difference between the two structures is in the configuration of the U·G base-pair (Figure 1(b)). The first NMR structure determination gave this as a reverse wobble base-pair. In contrast, the second determination resulted in a U·G configuration which is very different from a reverse wobble. There is a bifurcated hydrogen bond (Fritsch *et al.*, 1993) between one of the uracil carboxyl oxygen atoms and the imino and amino groups of the guanine. In addition to the different hydrogen-bond patterns, the U·G base-pairs also differ in their degree of planar-

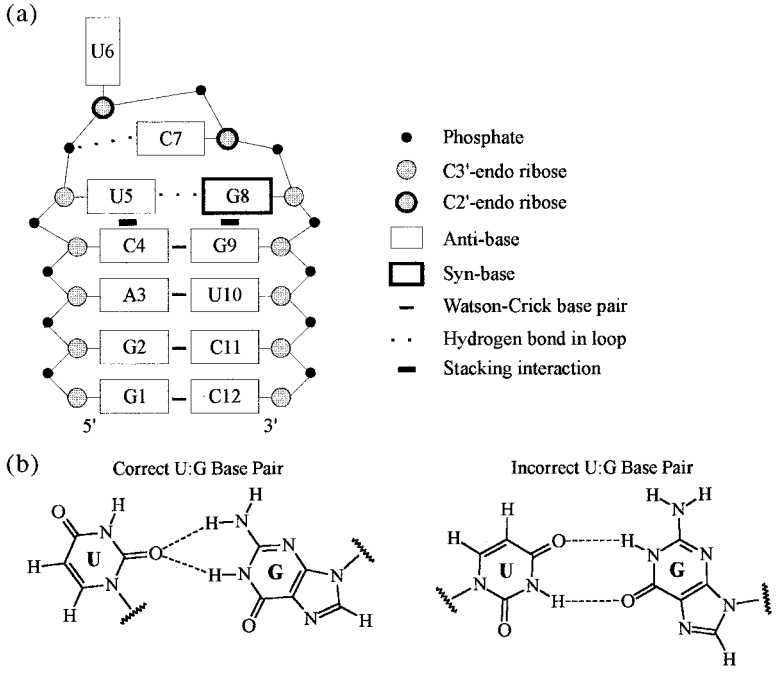

Figure 1. (a) Illustration of the secondary structure of the UUCG tetraloop including the tertiary interactions in the loop domain; (b) NMR-derived hydrogen-bonding patterns for the correct and incorrect geometry of the U5·G8 base-pair.

ity. The old structure contains a slightly buckled base-pair with non-ideal geometry for the hydrogen bonds while the base-pair from the new structure is planar with more ideal hydrogen bond angles.

One of the goals of our ongoing studies of this UUCG tetraloop has been to identify the specific stabilizing interactions of the loop 2′-hydroxyl groups. However, we felt that the existence of the two separate solution structures, with their slightly different loop conformations, provided us with a unique opportunity to test the ability of our methods to adequately sample phase space in this very compact RNA hairpin loop. We have asked whether it is possible to use molecule dynamics (MD) to simulate a transition from the first, or incorrect, conformation (Varani *et al*., 1991) to the second, or correct, conformation (Allain & Varani, 1995). In principle, MD simulations provide us with a tool to examine the atomic-level dynamic behavior of biomolecular systems. It is not subject to the crystal packing forces of X-ray structures and it adds time-dependent dynamics to the system (Allen & Tildesley, 1987; McCammon & Harvey, 1987). Also, MD is able to illustrate atomic motions which may be hidden by the time-averaged nature of the NMR experiment (i.e. sugar pucker averaging in nucleic acids). A recent improvement in the treatment of electrostatic interactions has allowed unrestrained MD simulations of nucleic acids (including hairpin loops) to reach into the nanosecond regime (Auffinger & Westhof, 1996; Cheatham *et al*., 1995; Zichi, 1995). Moreover, an *A*-DNA to *B*-DNA transition was observed during an unrestrained MD simulation (Cheatham & Kollman, 1996). Here, we present the results of five nanosecond-length simulations which were run in an attempt to observe a similar transition in the RNA UUCG tetraloop.

Results and Discussion

Simulations of the two NMR structures: standard MD analysis

One of the most widely used measures to judge the performance of a simulation is the atomic root-mean-square deviation (RMSd). As indicated by the average RMSd values listed in Table 1, both of these simulations stay very close to their starting structure. Following the initial equilibration period of 35 ps, the snapshots from each simulation stay within 2.0 Å of the final minimized structure (i.e. before any heating of the solute). These values are among the lowest ever obtained for nanosecond length, unrestrained MD simulations of a non-helical nucleic acid structure. As we have previously shown (Cheatham *et al*., 1995), this remarkable stability is attributable to the treatment of the electrostatic interactions with the PME method.

While the all-atom RMSds indicate that the simulations stayed very close to their starting structures, the values themselves do not provide any further insight into the nature of the atomic motions. An examination of the movies of these trajectories coupled with the domain motions reported by Zichi in a simulation of the RNA GCAA tetraloop (Zichi, 1995) led us to examine whether or not this same type of domain behavior was occurring in these systems. Domain behavior can be indicated from a matrix of the 2D cross-correlation coefficients of atomic displacements (Brünger *et al*., 1985; Swaminathan *et al*., 1991a; Zichi, 1995). We calculated this matrix for consecutive windows of 80 ps each for the entire oRNA trajectory. The resulting matrix is shown in Figure 2 with positive coefficients shown above the matrix diagonal and negative ones below it. The domain motions are obvious with the stem atoms highly correlated and both the 5′-stem and the 3′-stem atoms showing anti-correlated motions with respect to the loop atoms. This behavior was also indicated from the matrices calculated for 20 and 40 ps windows of the oRNA simulation and the same three matrices calculated for the nRNA simulation (data not shown).

Another way to obtain insight into the atomic motions is to examine to what extent the individual atoms deviate from their average structure. This type of ''fluctuation'' analysis is able to identify regions of high and low mobility in a structure. We did not find any discernible difference between the oRNA and nRNA simulations. Figure 3(a) shows the atomic RMS fluctuations from each simulation's average structure. The stem portions of the tetraloop (residues 1 to 4 and 9 to 12) show a similar level of motion for all of the trajectories. The phosphate atoms fluctuate slightly more than the other stem atoms, but overall the stem atoms move very little from their average positions. The loop atoms (residues 5 to 8) are clearly fluctuating

Table 1. Average RMSds of the simulations presented in this work

Simulation	Length (ps)	Group	RMSd (Å)[a]
oRNA	2025	Stem	1.3 ± 0.3
		Loop	1.3 ± 0.3
		All	1.7 ± 0.3
oDNA	1020	Stem	1.1 ± 0.3
		Loop	1.9 ± 0.3
		All	1.7 ± 0.3
nRNA	1055	Stem	1.2 ± 0.2
		Loop	1.1 ± 0.2
		All	1.3 ± 0.2
nDNA	570	Stem	1.2 ± 0.2
		Loop	1.1 ± 0.3
		All	1.3 ± 0.2
oD(R)NA	1015	Stem	0.7 ± 0.1
		Loop	1.0 ± 0.2
		All	1.0 ± 0.2

[a] RMSd from final minimized structure including standard deviations.

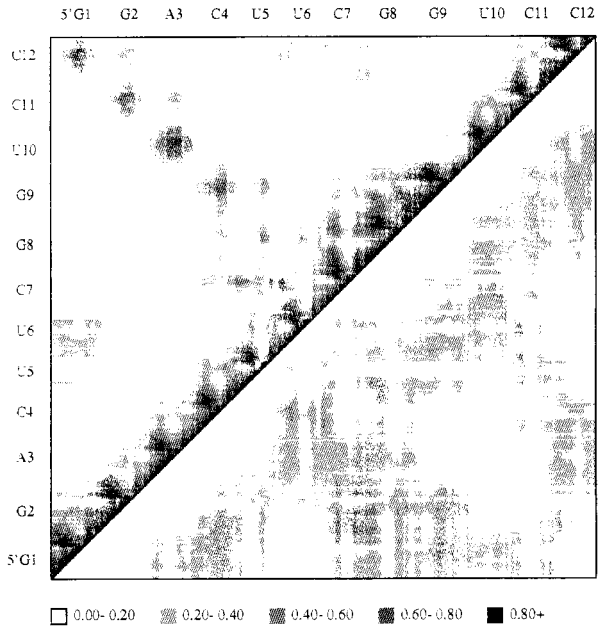

Figure 2. Cross-correlation of atomic displacements for the oRNA simulation. Coefficients were calculated from consecutive 80 ps windows of the trajectory following the equilibration period. Positive correlations appear above the diagonal and negative ones below it. The key at the bottom of the Figure corresponds to the absolute value of the coefficients.

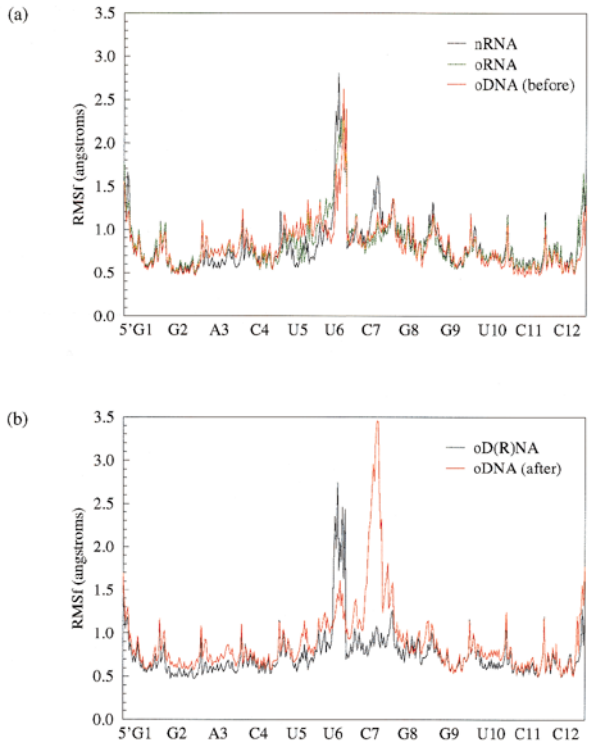

Figure 3. Atomic RMS fluctuations for all of the atoms in the tetraloop relative to their average positions. The peaks correspond to the backbone phosphate groups. (a) RMSfs for nRNA, oRNA and oDNA (before the conformational transition) simulations; (b) RMSfs for oD(R)NA and oDNA (after the conformational transition). The 5′ and 3′-stem sequences flank the intervening loop sequence.

more than the stem atoms. Residue U6 is identifiable as it shows greater fluctuations than the rest of the structure due to its lack of specific atomic interactions with the loop.

Given these domain motions, the greater mobility of the loop atoms, and our interest in the loop region of the hairpins, we analyzed the stem and loop behaviors separately. We performed a Dials and Windows analysis on the stems from both simulations to measure how well the simulations maintained the *A*-form helicoidal parameters. We do not discuss these results here other than to point out that the stems stay very close to their starting conformations and remain *A*-like throughout the simulations. Deviations were observed for the terminal base-pair (G1·C12) and the base-pair adjacent to the loop (C4·G9). All subsequent RMSd analyses on the simulations presented herein involve a fit on the loop atoms (or a subset of loop atoms) and do not include the stem atoms. A plot of the loop atoms' RMSd *versus* time is shown in Figure 4(a). Both simulations show very little movement from their initial structures with the RMSd values ranging between 0.5 and 1.0 Å.

Simulations of the two NMR structures: comparison to the NMR data

To provide a more direct comparison of these simulations to their experimental (NMR) structures, we examined whether the loop hydrogen bonds in the NMR structure are maintained, whether the sugar puckers and torsion values stay close to the NMR values and, most importantly, whether the structure during the simulation is violating any of the NMR-derived distances used to refine the three-dimensional structure. The results of these types of analysis can be seen in Tables 2 to 4 and are discussed in this section.

The hydrogen bonds in the U·G base-pair of the oRNA structure are not maintained during the simulation. While the two distances appear reasonable, the U5 : N3–G8 : O6 bond angle is clearly too small. This is due to a severe buckling in the U·G base-pair which develops early in the simulation (Figure 5). Conversely, in the nRNA simulation, the hydrogen bonding pattern is well-maintained with both bonds remaining close to ideal distances and acceptable angles. Also, the U·G base-pair remains close to an ideal planar geometry with an average value of 156(±10)° for the entire ~1 ns simulation. Both of the simulations maintained the C7 : N4–U6 : O2P hydrogen bond throughout the simulations, although the average length from the nRNA simulation was a little long.

It is interesting to note that in their re-refinement of this tetraloop structure, Varani and co-workers

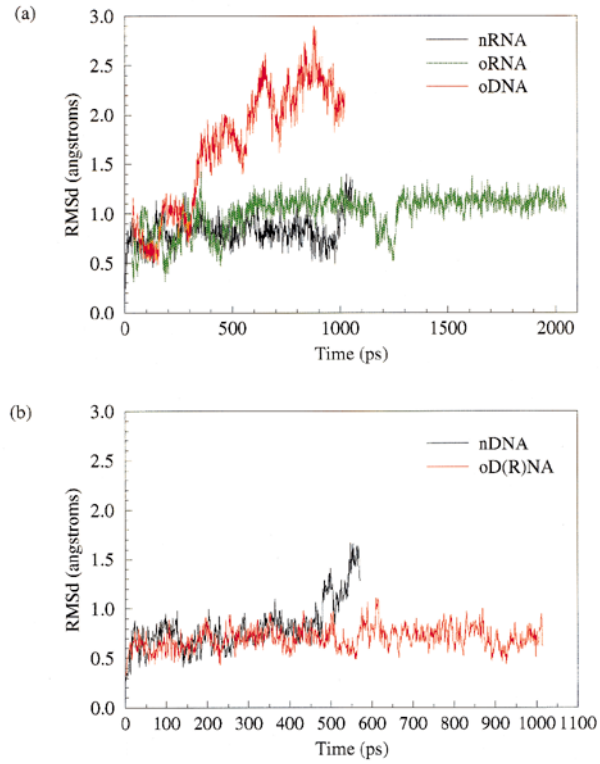

Figure 4. Time-course of the loop atoms' RMSd for the simulations presented in this work. These values are the deviations from the minimized structures of each simulation. (a) RMSds for nRNA, oRNA and oDNA; (b) RMSds for nDNA and oD(R)NA. The RMSd fits and calculated values included all of the atoms in the loop except for the base atoms of the highly flexible residue U6.

found a sugar-base hydrogen bond between the 2′-hydroxyl group of U5 and the carbonyl oxygen of G8. They postulated that this hydrogen bond was the reason for the extra thermodynamic stability of the tetraloop. We find that this bond is not maintained once the nRNA structure is solvated. Our results show that a sugar-backbone hydrogen bond forms at about 90 ps and remains throughout the entire simulation. The NMR refinement was done with a different force field and under *in vacuo* conditions. Our own *in vacuo* minimizations result in the formation of the same hydrogen bond, indicating no force field dependence in the NMR refinement, but once the system is placed in a water box and dynamics are begun, the preferred hydrogen bond acceptor for the U5 2′OH group is the backbone U5pU6 oxygen atom (O5′) and not the guanine O6 atom. In our simulations, the solvent becomes the donor to the G8 O6 atom. While this particular ribose may be the key to the stability of the system, we feel that it is not due to a sugar-base contact.

The backbone torsions and sugar puckers in both simulations generally stayed close to their experimental values (Table 3, data shown for nRNA only). For the sugars in both the oRNA and nRNA simulations, we only saw repuckering in C12 (data not shown). All other sugars maintained their experimental conformations. We did observe some variability in the pucker of U6 in nRNA, where twice it briefly sampled between O4′-*endo* and C4′-*endo*, both times returning to the correct C2′-*endo* conformation. In both simulations, the χ angle of U6 was highly variable with standard deviations over the simulations of 41 Å and 60 Å for the

Table 2. Loop hydrogen bond geometry

		Simulation		NMR[a]	
Simulation	Hydrogen bond	⟨Length⟩ (Å)	⟨Angle⟩ (deg.)	Length (Å)	Angle (deg.)
oRNA	U5:N3–G8:O6	3.2 ± 0.3	87.2 ± 34.2	2.7	147.7
	U5:O4–G8:N1	3.0 ± 0.2	158.1 ± 10.7	2.8	160.8
	C7:N4–U6:O2P	2.9 ± 0.2	127.5 ± 7.1	2.7	126.5
oDNA	U5:N3–G8:O6	5.0 ± 1.3	78.1 ± 27.4		
	U5:O4–G8:N1	6.0 ± 1.9	156.3 ± 10.0		
	U5:O2–G8:N1	3.6 ± 1.2	139.8 ± 25.2		
	U5:O2–G8:N2	4.5 ± 1.9	123.6 ± 17.9		
	C7:N4–U6:O2P	5.3 ± 2.4	128.8 ± 19.6		
nRNA	U5:O2–G8:N1	3.0 ± 0.3	146.6 ± 13.6	2.9 ± 0.1	153.7 ± 7.0
	U5:O2–G8:N2	3.1 ± 0.3	145.5 ± 14.1	3.1 ± 0.2	145.5 ± 5.9
	C7:N4–U6:O2P	3.5 ± 1.0	135.4 ± 13.8	4.5 ± 1.1	131.5 ± 5.7
	U5:O2′–G8:O6	3.7 ± 0.5	51.9 ± 30.0	2.6 ± 0.0	72.5 ± 19.5
nDNA	U5:O2–G8:N1	2.9 ± 0.1	155.4 ± 11.5		
	U5:O2–G8:N2	3.2 ± 0.3	137.7 ± 10.8		
	C7:N4–U6:O2P	4.4 ± 1.4	145.7 ± 14.6		
oD(R)NA	U5:O2–G8:N1	3.0 ± 0.3	149.0 ± 13.7		
	U5:O2–G8:N2	3.1 ± 0.4	143.1 ± 15.4		
	C7:N4–U6:O2P	3.2 ± 0.6	131.9 ± 10.4		
	U5:O2′–G8:O6	3.6 ± 0.6	64.2 ± 42.9		

Values include standard deviations. Brackets indicate simulation averages.
[a] For nRNA, values are from the family of NMR structures.

Table 3. Comparison of simulated and experimental torsion values for the C(UUCG)G residues in the nRNA simulation

	α	β	γ	δ	ε	ζ	χ	P
C4	−73 ± 10	172 ± 10	62 ± 9	77 ± 7	−155 ± 10	−63 ± 9	−159 ± 8	15 ± 9
	−71 ± 4	163 ± 3	63 ± 5	92 ± 1	−148 ± 9	−67 ± 10	−140 ± 3	20 ± 4
U5	−73 ± 9	174 ± 8	64 ± 8	77 ± 6	−177 ± 9	−91 ± 9	−152 ± 10	19 ± 12
	−80 ± 11	−179 ± 15	49 ± 3	92 ± 1	−165 ± 3	−92 ± 4	−140 ± 3	19 ± 3
U6	−156 ± 17	158 ± 17	60 ± 10	135 ± 13	−92 ± 12	−71 ± 11	−32 ± 60	150 ± 23
	−142 ± 5	112 ± 3	63 ± 5	133 ± 1	−101 ± 6	−58 ± 4	−149 ± 6	143 ± 2
C7	−63 ± 10	177 ± 13	60 ± 9	141 ± 11	−91 ± 10	70 ± 9	−134 ± 17	159 ± 18
	−55 ± 5	174 ± 4	45 ± 2	136 ± 3	−114 ± 8	105 ± 5	−126 ± 1	149 ± 6
G8	74 ± 10	−175 ± 8	−175 ± 8	79 ± 7	−176 ± 8	57 ± 10	34 ± 12	47 ± 17
	34 ± 6	−145 ± 3	−171 ± 4	91 ± 3	−147 ± 11	−41 ± 46	61 ± 14	50 ± 10
	−58 ± 7	136 ± 4	−24 ± 11					
G9	−161 ± 9	77 ± 9	176 ± 9	79 ± 8	−154 ± 9	−93 ± 9	−168 ± 9	15 ± 14
	−118 ± 25	92 ± 19	−168 ± 25	96 ± 2	−151 ± 2	−68 ± 3	−165 ± 2	−5 ± 3
	112 ± 9	−124 ± 22	99 ± 0					

Top value listed is the simulation average plus standard deviation. Bottom value is the average value from the family of NMR structures. Among the NMR structures, six torsions were found to cluster into two sub-families. The G8:ζ torsion showed no preference with the NMR structures varying between −116 and 32 degrees.

oRNA and nRNA simulations, respectively. This behavior was expected due to the lack of specific interactions of this base with the rest of the loop and because this is the variable position found in the UNCG class of tetraloops.

We did observe dihedral transitions during both simulations, most in the form of crankshaft motions. In RNA double helices, these types of correlated transitions in the backbone angles have been found to have little effect on the stacking interactions of the base atoms (Portmann *et al*., 1995). Similar crankshaft motions have been observed in other theoretical studies of both RNA and DNA systems (Auffinger & Westhof, 1996; Beveridge & Ravishanker, 1994; Cheatham & Kollman, 1996a,b; Srinivasan *et al*., 1990; Swaminathan *et al*., 1991b; Weerasinghe *et al*., 1995; Zichi, 1995). During the oRNA simulation, we observed a total of fifteen dihedral transitions, all of them occurring after 500 ps of MD. Most of these transitions were crankshaft motions in the backbone torsions at the G8pG9 step. This region, along with the χ angle of C7, also displayed variability in the family of structures found during the NMR refinement (Varani *et al*., 1991). We only observed two crankshaft motions (four transitions) in the nRNA simulation. This brief event occurred just after 100 ps in the U6(α,β) torsions. Even though the NMR data indicated some variability in the α, β, and γ angles of C7, and the α and β angles of G8 (Allain & Varani, 1995), we did not observe any transitions in this part of the backbone. Table 3 compares the loop torsions from the nRNA simulation to the family of NMR structures resulting from the NMR studies. The simulation maintains the experimentally determined rotamers for most loop torsions but does not sample all of the subfamilies, as noted above.

Perhaps the best way to judge how well the simulations maintained their respective NMR structures is through a direct comparison of the atomic distances derived from the nuclear Over-

Table 4. Violations of NMR distances from the oRNA and nRNA simulations

Simulation	Atom pair (residue: atom name)		NMR restraint bounds (Å) Lower	Upper	Simulation average (Å)[a]	NMR structure (Å)[b]
oRNA	G2:H1′	A3:H4′	2.5	5.0	6.3 ± 0.3	6.4
	C4:H2′1	U5:H5	4.0	10.0	3.2 ± 0.7	3.3
	C4:H6	U5:H6	2.5	5.0	6.3 ± 0.9	5.2
	U5:H2′1	U6:H6	4.0	10.0	3.1 ± 0.5	4.1
	U5:H3′	U6:H5	2.5	5.0	6.1 ± 0.5	5.6
	U5:H4′	U6:H5	2.5	5.0	5.9 ± 0.7	3.6
	U5:H2′1	U6:H3′	1.9	3.0	3.8 ± 0.4	3.4
	U5:H2′1	U6:H2′1	2.5	5.0	2.2 ± 0.2	2.5
	G9:P	G9:C4′	3.2	3.6	3.9 ± 0.1	3.8
nRNA	C7:H2′1	G8:H1	1.8	6.0	6.7 ± 0.5	6.0 ± 0.0
	C7:H4′	G8:H8	1.8	5.0	6.0 ± 0.3	5.0 ± 0.0

[a] Average value including standard deviation.
[b] For nRNA, values are from the family of NMR structures.

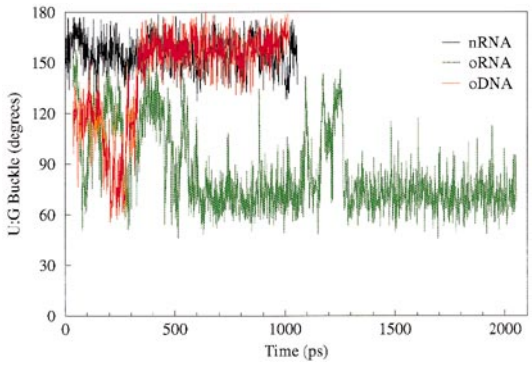

Figure 5. Time-course of the buckle of the U·G base-pair in the tetraloop for the nRNA, oRNA and oDNA simulations. The buckle was defined as the angle between the two planes described by: U6:N1,N3,C5 and G8:N1,N3,C8. Using this definition, an ideal angle for the buckle is 180°.

hauser effects (NOEs) with the same distances from the simulations themselves. To do this, we calculated the $\langle r_{ij}^{-6}\rangle^{-1/6}$ values from the trajectories, then compared them with the restraint bounds used in the respective structure refinements. In our comparisons, we have used all of the distances from the oRNA refinement (275) and the distances used to refine the loop portion of the nRNA structure (227). Our comparison to the nRNA restraint bounds is limited to the AC(UUCG)GU residues because we simulated a slightly different stem sequence than was used in the experiment. It is well known that a single structure may not be able to satisfy all of the NMR restraints. Accordingly, we have included the fluctuations of the distances as part of our analysis. In Table 4 we list only those distances from the trajectories where the range of values (average ± standard deviation) was outside of the restraint bounds. For each of the violated restraints, we have also included the distance measured from the NMR structure(s).

In the oRNA simulation, there are quite a few distance violations, particularly in the region of the loop. We also find that the NMR structure violates six of the same distance restraints. With the exception of the first two distance violations listed, all of the violations are due to the incorrect geometry (buckling and hydrogen bonding pattern) of the U·G base-pair. Conversely, for nRNA, we see that the simulation does a much better job of maintaining the NMR distance restraints. Only two violations are found and both are within an angstrom of the upper restraint bound.

The oRNA simulation does not converge to the correct loop conformation

As we have shown, both of these simulations are very stable and both do an excellent job of maintaining their respective experimental (starting) structures. Even though the oRNA simulation was very long (~2 ns), we did not observe a conformational change from the starting conformation to the correct conformation as determined from the structure determination of Allain & Varani (1995). The most telling measure that this convergence does not take place is the value of the RMSd between the average loop conformations of the oRNA and nRNA systems. Table 5 contains the pairwise RMSds between the average loop structures for the simulations presented in this work. The oRNA and nRNA structures start 2.5 Å apart and, after ~2 ns of simulation, the oRNA structure is still 1.7 Å from the nRNA structure. This is not a surprising result given the limited sampling available with standard MD. Interestingly, the two loop conformations have almost identical internal energies (Miller, 1996). The similarity in internal energies indicates that the preferred loop conformation is most likely due to more favorable solvation. Using a simple Arrenhius equation,

$$k = \frac{k_B T}{h} e^{-\Delta G^{\neq}/RT},$$

as a rough rule-of-thumb, MD at 300 K should be able to overcome barriers of ~5 kcal/mol every 1000 ps. Apparently, the incorrect loop conformation, including explicit solvation, is a metastable state and the barrier between the two conformations is higher than this value. Recently we have investigated the ability of an RNA double helix to convert between *B* and *A* forms during MD (Cheatham & Kollman, 1997). His results indicate that the barrier to conversion is high as *B*-RNA seems to be a metastable state. This rigidity observed in RNA systems during unrestrained MD is not seen for simulations of DNA helices where McConnell *et al*. (1994) have observed substates of DNA and Cheatham & Kollman (1996) observed an *A*- to *B*-DNA transition in the double helical system $d(CCAACGTTGG)_2$.

The barriers to conformational change in this RNA tetraloop may well involve the hydrogen bonds formed by the 2′OH groups of the riboses. In the two simulations, the four loop riboses have similar torsional values for the C3′-C2′-O2′-2′OH dihedral angles. The U6:2′-OH engages in different interactions in the separate simulations because of the different backbone conformations. In both simulations, U5:2′OH donates to a backbone oxy-

Table 5. RMSd values between the average loop structures (Å)[a]

	nRNA	oRNA	oDNA[b]	nDNA	oD(R)NA
nRNA	–	1.7	1.0	0.6	0.5
oRNA	2.5[c]	–	1.4	1.7	1.6
oDNA			–	0.7	0.6
nDNA				–	0.6

[a] RMS fits and values were done using all loop atoms except the base atoms of U6.
[b] Average structure after conversion (~700 ps).
[c] RMSd between the two NMR structures.

gen (U6:O5′), C7:2′OH is solvent exposed and appears to interact strictly with water molecules throughout the simulation in water-mediated hydrogen bonds to G8:O1P and O2P, and G8:2′OH donates to another backbone oxygen atom (G9:O2P). The hydrogen bond acceptor for U6:2′OH is different in the two systems. For the oRNA simulation, there is a pocket in the molecular surface of the loop which is usually occupied by a water molecule. The U6 hydroxyl group donates to this water molecule, when present, and participates in a two-centered hydrogen bond with U6:O3′ and C7:O4′ when not. In the nRNA loop conformation, where there is not a pocket in the surface, the U6 hydroxyl group appears to show restricted rotation for two reasons. The first is a weak two-centered hydrogen bond as in the oRNA simulation. The second reason is because of a van der Waals clash between the U6:2′OH atom and the C7:H5′2 atom. Because both tetraloop conformations are very compact, these hydrogen bonds are stable and do not show a lot of variability (i.e. solvent does not compete as the hydrogen bond acceptor).

Multiple simulations with slightly different conditions can lead to very different system behavior (Auffinger *et al*., 1995, 1996; Auffinger & Westhof, 1996; Louise-May *et al*., 1995). To investigate the dependence of our results on the simulation conditions, we have also run a separate ~1 ns simulation of the oRNA system using a different time step and non-bonded pair list update. We observed the same overall stable behavior with the same buckling of the U·G base-pair in this simulation as well (data not shown).

Simulations of the chimeric sequence r(GGAC)d(UUCG)r(GUCC): oDNA and nDNA

In order to observe a conformational change in this RNA tetraloop, clearly the barriers needed to be reduced. Among the commonly used strategies to overcome barriers in an MD simulation are increased temperature, locally enhanced sampling, and torsional driving. While any of these strategies would have been an acceptable choice for this system, we felt that the best choice was to take direction from previously published experimental work. It is a well-established fact that the DNA duplex is more flexible than the RNA duplex (Saenger, 1984). Portmann *et al*. (1995) have recently reported on crystal structures of analogous sequences of RNA and DNA and found that while the DNA sequence shows a considerable amount of variability, the RNA sequence does not, even though it crystallizes in two lattice forms. This enhanced flexibility in RNA duplexes is also seen in ^{31}P NMR experiments (Fujiwara & Shindo, 1985; Shindo *et al*., 1985). Specifically, we used the results of Sakata *et al*. (1990) on the relative stabilities of the r(GGACUUCGGUCC) and r(GGAC)d(UUCG)r(GUCC) molecules. The system containing the deoxyriboses in the loop residues has a melting temperature ~10° less than the native RNA sequence indicating that the chimeric sequence is not as stable. Accordingly, our first attempt to lower the barriers in the RNA simulations was to run MD simulations with the loop riboses replaced by deoxyriboses. We ran two simulations of this chimeric sequence (oDNA, nDNA) where, again, we have used the two NMR structures as starting points. Another possibility was to alter the force field parameters for the ribose moieties. By lowering the energetic barriers to sugar repuckering, it might be possible to decouple the hydrogen bonding contributions of the 2′OH groups from the reduced sampling due to them. However, the loop riboses were already in the correct pucker states, a condition which indicated that the barrier between the two conformations was a result of the hydrogen bonding, and not the reduced conformational sampling of the loop sugars.

In the oDNA simulation, we observed that the loop moved farther from its starting conformation than in the oRNA simulation. The average loop RMSd of 1.9 Å shown in Table 1 does not adequately describe the amount of deviation from the starting conformation. The oDNA plot in Figure 4(a) clearly shows that the RMSd from the starting conformation increases greatly (~1.0 to 2.5 Å) during the 300 to 800 ps segment of the simulation but appears to be leveling off by the end. While Figure 4(a) does not give any indication as to what is happening to the loop conformation, we have clearly increased the flexibility of the system by removing the loop hydroxyl groups. In contrast, as Table 1 and Figure 4(b) show, the RMSds for the nDNA simulation indicate that both the stem and the loop domains stay very close to their starting conformations through the >500 ps of this simulation. From Figure 4(b), we see that the loop conformation is indeed moving more than in either of the RNA simulations.

An interesting comparison between the RNA and chimeric simulations comes from an examination of the number of sugar repuckering events and dihedral transitions in the loop domain. As in the RNA simulations, we observed a repuckering of the C12 ribose from C3′-*endo* to C2′-*endo* during both of the simulations. Unlike the RNA simulations, in nDNA there was a brief repuckering of residue U5 (C3′-*endo* to C2′-*endo*) and in oDNA there was both a brief repuckering of residue G8 (C3′-*endo* to O4′-*endo*) and an increased sampling of the pseudorotation cycle for residue C7, which varied between C2′-*endo* and O4′-*endo*. The other riboses maintained their initial conformations. In the nDNA simulation, the only glycosidic torsion which displayed significant rotation was that of U6, just as in the RNA simulations. A different behavior was seen in the oDNA simulation where five of the residues, C4 to G8, showed rotation about this torsion, with the largest motion observed in U5 and U6. While we would expect U6 to be highly variable, the rotation about U5:χ was unexpected due to its stacking on C4 at the top of the stem and its participation in the U·G base-pair.

In both of these chimeric simulations, there was much more variability in the backbone torsions of the loop domain compared to either of the RNA simulations, another example of the increased flexibility due to the lack of 2′ hydroxyl groups. In nDNA, this consisted of twenty-seven dihedral transitions in U5(α,β,γ), U5(ε,ζ), and U6(α,β,γ). All of these torsion flips were transitory crankshaft motions, but one of them lasted for ~200 ps. In contrast, in the oDNA simulation, we observed both temporary crankshaft motions and distinct changes in some of the backbone torsions.

As with the RNA structures, we observed differences in the U·G buckle in the chimeric simulations. In nDNA, this base-pair maintains a planar geometry throughout the entire simulation, with an average value of 157(±11)°. However, in the oDNA simulation this base-pair becomes very buckled early in the simulation, then flattens out to a more ideal geometry (Figure 5). The range of the buckle observed in this simulation is very broad with a minimum value of 56° and a maximum of 180°.

The oDNA simulation does converge to the correct loop conformation

The nDNA system, although clearly more flexible than the nRNA system, demonstrated the same stable behavior: it stayed very close to the experimental structure of the native sequence, maintained a planar U·G base-pair, and maintained the correct hydrogen bond geometry for this pair (Table 2). On the other hand, the oDNA simulation was not as stable as the loop domain moved very far from its starting conformation. As we discuss in this section, this movement away from the starting structure was actually a convergence from the incorrect to the correct loop conformation; a conversion made possible by lowering the barriers to the conformational change and one not seen in multiple 1 ns simulations of the native all-RNA sequence.

Another way to investigate the stability and convergence behavior of a simulation is through the examination of a two-dimensional (2D) RMSd map (McConnell *et al*., 1994). This kind of two-dimensional analysis allows for the identification of substates in the trajectory. While the one-dimensional (1D) RMSd plot of the oDNA trajectory indicates that the tetraloop conformation is moving away from the starting structure, the 2D map (Figure 6) clearly identifies that the loop atoms have converged to a different conformation by ~300 ps and stay there throughout the remainder of the simulation. This conformational change is also reflected in the plots of both the loop backbone dihedral angles (data not shown) and the U·G hydrogen bonding pattern. The time-course of the hydrogen bond lengths corresponding to the two U·G base-pairs is shown in Figure 7. The hydrogen bonds, which start in the incorrect loop conformation, undergo a sharp transition to the correct pattern just after 300 ps. This change in pattern is not reversed during the remainder of the ~1 ns simulation. The variability in the U5:O2–G8:N2 hydrogen bond is also observed in both the nRNA and nDNA simulations (data not shown). This concerted change is also reflected in the plot of the U·G buckle (Figure 5), where the base-pair becomes progressively more buckled until approximately 300 ps where it flattens out, eventually converging to the same planarity observed in the nRNA simulation.

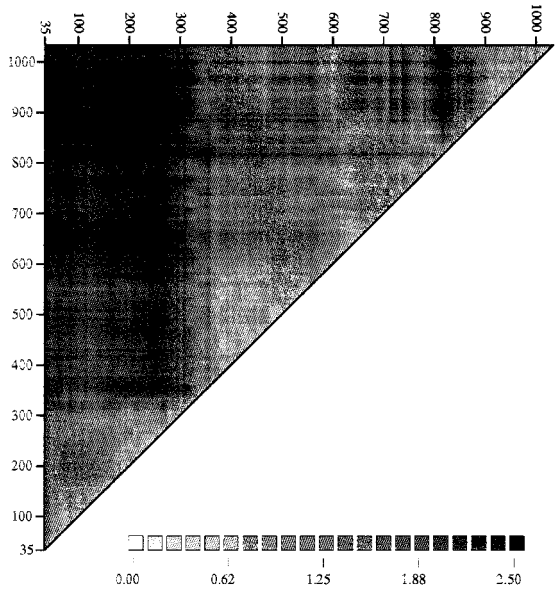

Figure 6. 2D-RMSd map for the loop atoms in the oDNA simulation showing the transition to a new conformation at ~300 ps. The labels indicate the time during the trajectory and begin at 35 ps (after the equilibration period). As indicated in the legend, darker squares correspond to higher RMSd values.

The best evidence that a conformational change to the correct loop conformation took place comes from a comparison of the RMSds between the average loop conformations from the nRNA simulation

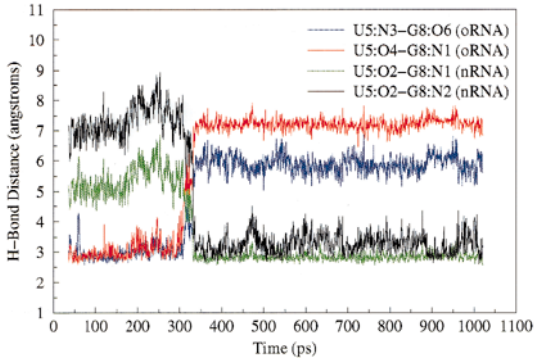

Figure 7. Time-course of the two hydrogen bonding patterns of the U·G base-pair in the oDNA simulation. The sharp transition from the incorrect to the correct structure is indicated by the change in hydrogen bonding pattern just after 300 ps.

and the oDNA simulation after this concerted change. The third column in Table 5 shows that the oDNA loop, after the conformational change, is 1.4 Å away from the average oRNA conformation, but only 1.0 Å away from the average nRNA conformation. Additionally, in the fourth column we show the values between the nDNA simulation and the other three simulations. As expected, the average nDNA structure is very close to the nRNA structure (0.6 Å) and far from the oRNA structure (1.7 Å). What is unexpected is that the two chimeric simulations are, on average, only 0.7 Å apart. The RMSds in Table 5 are calculated using all of the loop atoms except for the base atoms of residue U6. We feel that excluding these base atoms gives the most accurate indication of the (dis)similarity between the loops because this residue displays highly variable motion throughout all of the trajectories and including its average position in the RMSd calculation would skew the best fit of the more stable parts of the loop. Perhaps the most illustrative way to demonstrate the convergence is by looking at the structures of the average loop conformations. Figure 8 displays the average loop structures from the oRNA and oDNA simulations, the latter after its conformational change, overlaid on the average nRNA structure. The RMSd values in Table 5 correspond to the RMS fits shown in this Figure.

When we examined the extent of torsional sampling in the loop residues before and after the conformational change, we found different behaviors. A few of the torsions in the loop, both those that underwent transitions and one that did not, displayed reduced fluctuations (i.e. did not sample as widely) after the conversion from the incorrect to the correct loop conformation. These include U5:ε and χ, both of which also underwent sharp transitions; and U6:α and ε, where α showed a sharp transition and ε settled down into the upper part of the torsional space which it had sampled before the loop conversion. One loop torsion, U6:β, actually displayed increased sampling after it underwent the brief crankshaft motion. In general, after the conversion took place, the loop torsions in the oDNA simulation were very similar to those in nDNA in terms of which residues and torsions sampled more widely.

Even though we continued the oDNA simulation for another ~700 ps following the conformational change, there were five loop torsions which did not converge to those values found in the correct NMR structure. These five torsions all occur in the G8pG9 step (G8:ε,ζ and G9:α,β,γ) on the 3′-side of the loop. For nRNA, these five angles have the values −176°, 57°, −131°, 77° and 176°. For oDNA they are −134°, −55°, −65°, 158° and 70°. Figures 8 and 9 show that these differences are largely compensatory in nature as they do not influence the interactions in the loop or the stacking interactions between the loop and stem domains. It should be noted that the backbone conformation is the least well-defined by the NMR experiment. There are not a lot of NOEs in this region and the restraints used to refine the backbone are semi-quantitative torsions derived from weak coupling constants. It is possible that there is some flexibility in this region. We observed, in both the oRNA and oDNA simulations, periods where these torsions sampled the nRNA values, some for as long as 300 ps.

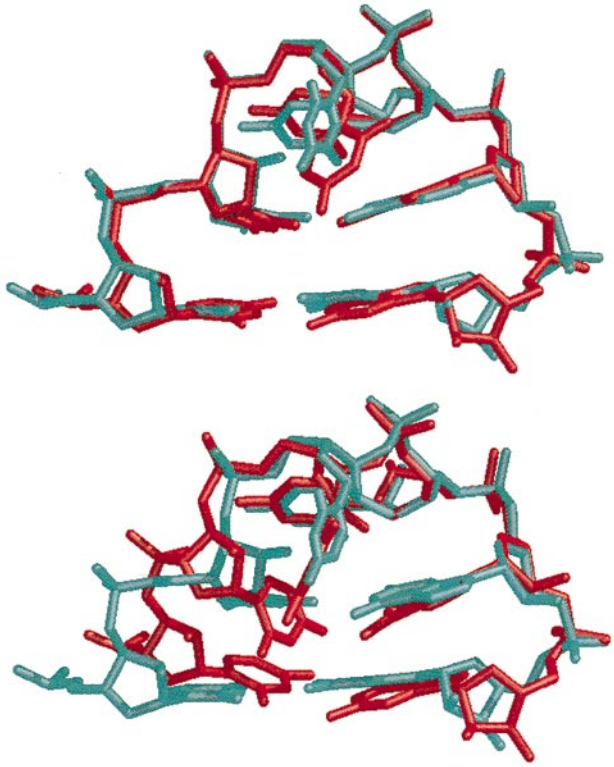

Figure 8. Comparison of the average loop conformations (including residues C4 and G9) from the nRNA and oDNA simulations (top) and the nRNA and oRNA simulations (bottom). The nRNA loop is colored cyan in both the top and bottom. The average conformation for the oDNA simulation was calculated after the conformational transition took place. The RMSd values in Table 5 correspond to the fits shown in this Figure.

In summary, we see a few changes prior to the transition event, mostly to accommodate the buckling of residue U5, and a general re-equilibration of the loop conformation in response to the rapid conformational change in the U·G base-pair. There are a couple of general trends in the transition pathway which are interesting and worth pointing out. Most of the changes took place from the 5′-side of the loop where there are four pyrimidine residues (C4 to C7) preceding the two purine residues (G8 to G9). It is known that the allowed χ angle for purine nucleosides covers a broader range (180° to −60°) than that for pyrimidine nucleosides, which are restricted to a more classical *anti* conformation (180° to −90°; Saenger, 1984). It is possible that the pyrimidine-rich 5′-side of the loop, rather than the purine-rich 3′-side, more easily accommodated the drastic changes in glycosidic angles which were part of the transition. Additionally, the stack-

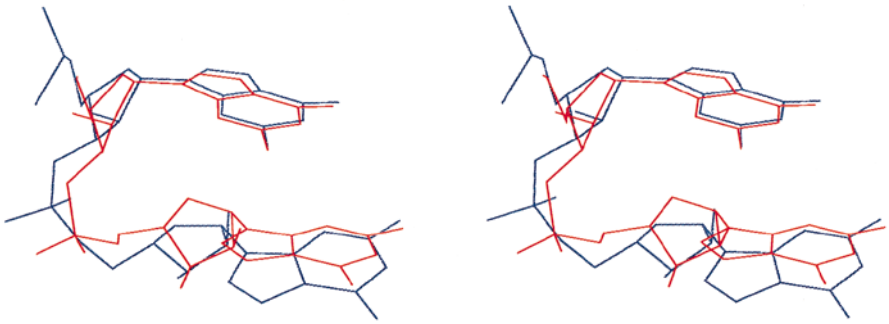

Figure 9. Stereo view of the G8pG9 step for the average structures from the nRNA (blue) and oD(R)NA (red) simulations showing the remaining torsional differences between the experimental structure and the converged MD structure. The RMS fit was done on all loop atoms including the closing C4·G9 base-pair. The view is from the major-groove side with the G8 residue on top and stacking on the G9 residue.

ing interactions continue from both sides of the stem into the U·G base-pair of the loop. This gives a pyrimidine–pyrimidine stacking interaction on the 5′-side of the loop (C4-U5) and a purine–purine interaction on the 3′-side (G8-G9). Experimental studies have found that purine–purine stacks are the most stable (Saenger, 1984), a factor which may also influence the transition occurring from the 5′-side of the loop. These trends are also reflected in the nDNA simulation where the only dihedral transitions we observed occurred in the 5′-side of the loop.

Adding the loop hydroxyl groups to the oDNA system locks down the conformation

A fifth simulation (oD(R)NA) was run to address the role of the loop 2′-hydroxyl groups which, although they do not appear to determine the loop conformation in our simulations, certainly provide the barriers to conformational change. To do this, we took the average structure from the oDNA simulation after the conformational change took place, added the 2′-hydroxyl groups, then equilibrated and ran this system under the same protocol as detailed in Methods. All of the loop hydroxyls were placed into ''default'' positions: C3′-C2′-O2′-HO′2 is *trans* (180°). Here, we briefly describe the results of this simulation.

In Figure 4(b), we compare the time-dependent RMSd of the oD(R)NA simulation with that of the nDNA simulation. While we see some movement away from the starting structure of the chimeric tetraloop, adding the hydroxyl groups to the converged oDNA loop indeed keeps it very close to the starting structure. This simulation turned out to be the most stable of all five presented in this work, with an average loop RMSd over the 1 ns simulation of 0.7(±0.1) Å (Table 1). The U·G base-pair stayed planar throughout the simulation (155(±10)°) and the correct hydrogen-bond geometry is maintained with a little more variability in the U5:O2-G8:N2 bond, the same behavior as we observed in the nRNA and nDNA simulations. The C7:N4-U6:O2P hydrogen bond has also become more stable with the addition of the loop hydroxyl groups (Table 2). In the oDNA and nDNA simulations, the increased flexibility of the backbone allowed the C7 residue to fluctuate more than in the RNA simulations (Figure 3(b)). A consequence of this higher mobility is that the hydrogen bond between the C7 amino and the U6 phosphate oxygen is not maintained. In Figure 3(b), we see that one of the effects of the hydroxyl groups was to reduce the fluctuations in the loop and essentially ''lock'' the loop into a single conformation. Finally, by locking the loop into a single conformation, we find that the RMSd between the average loop structure from this simulation (oD(R)NA) and the nRNA simulation is the smallest (best match) between any two of the simulations presented here (Table 5). This RMSd is a remarkably small 0.5 Å, which is even closer than the two simulations which started from the correct NMR structure (nRNA and nDNA).

The behavior of the previously discussed U5:2′OH is also consistent with the nRNA simulation. The initial *trans* position of this group laces the hydroxyl in a favorable conformation to form a hydrogen bond with the G8:O6 atom, the same bond proposed by Allain & Varani (1995) to account for the thermodynamic stability of the tetraloop. However, after ~160 ps of MD this hydroxyl group rotates to form the same bond found in the nRNA simulation (U5:O2′-U6:O5′) where it stays for the remainder of the simulation.

Conclusion

We have presented the results of five unrestrained molecular dynamics simulations of an RNA tetraloop, describing our efforts to observe the conversion between an incorrect and a correct loop conformation. We have utilized two separately determined and published NMR structures of this tetraloop (Allain & Varani, 1995; Varani *et al.*, 1991) as the starting structures in our simulations. The second NMR determination, which was refined with many more restraints than the first one, resulted in a different conformation for the loop portion of the structure. Because the differences in the two loop conformations were mainly in the geome-

try of the U·G base-pair of the tetraloop, our hope was that we could observe the conformational transition from the old, incorrect loop structure to the new, correct one. We have previously shown (Cheatham *et al*., 1995) that long unrestrained MD simulations are obtainable through the application of an appropriate force field (Cornell *et al*., 1995) and a proper treatment of the long-ranged electrostatic interactions (Essmann *et al*., 1995). In this work, we performed nanosecond length simulations of both of these NMR structures and provide a direct comparison between our calculations and both sets of NMR data.

Our simulations starting from each of the NMR structures not only stay very close to the initial models and preserve the non-bonded interactions in the loops, but they also maintain most of the NMR-derived distances used in their respective structure refinements. These results indicate that our simulations sampled conformational states only in the vicinity of the starting structures, suggesting that the barriers to conformational change are very high in RNA systems. We did not observe the transition between the incorrect and correct loop structures even though one of the simulations was run for >2 ns. We have also run a separate 1 ns simulation of the first NMR structure and found that it, too, stayed very close to the starting structure. Similar results indicating that the barriers to conformational change in RNA systems are significantly higher than in DNA systems have been found in RNA:RNA helices and RNA:DNA hybrid helices (Cheatham & Kollman, 1997).

In an attempt to observe the conformational change, we ran two subsequent simulations of this tetraloop but with the loop riboses replaced with deoxyribose. These simulations were undertaken under the assumption that by removing the 2′OH moieties from the loop domain we would lower the transitional barriers and observe the conformational change. Again, we used each of the NMR structures as starting points for these simulations. Not only were we were able to increase the flexibility of the loops in both simulations, but the simulation which started in the incorrect structure underwent a transition to the correct conformation. The simulation which started in the correct conformation, although it demonstrated increased sampling, remained in the correct conformation. To our knowledge, this is the first example of an observation of a transition in a non-helical nucleic acid system using MD. There have been two reports of transitions observed in DNA helices, one an *A*-DNA to *B*-DNA conversion using the same force field and PME treatment of the electrostatics as in this work (Cheatham & Kollman, 1996). The second was a *B*-DNA to *A*-DNA conversion using the CHARMM23 all-hydrogen parameter set and an Ewald treatment of the electrostatics (Yang & Pettitt, 1996). Based on our results with this small model system, we feel that this method is worthy of consideration during the refinement of nucleic acid structures. We feel that it could be particularly useful when applied to non-helical regions where the structure is often not well defined by the experimental data.

We have also reported on a fifth simulation which was run starting from the average structure calculated from the chimeric simulation after it underwent the conformational change. In this simulation, we added the 2′-hydroxyl groups back to the average structure before the dynamics were run. This structure was the most stable of the five presented here. The effect of the hydroxyl atoms was to lock the loop into a single conformation. Perhaps the most remarkable result presented here is that, after re-introducing the riboses into the loop, the RMSd between the average structure from this simulation and that from the simulation of the correct NMR structure is only 0.5 Å.

This study, as well as the recent work of Cheatham and Kollman, suggest that the multiple molecular dynamics (MMD) method (Auffinger *et al*., 1995, 1996; Auffinger & Westhof, 1996; Louise-May *et al*., 1995) may not be as successful as hoped as a way to increase the sampling in RNA simulations. The MMD method uses different initial velocities for the same starting geometry in order to obtain independent trajectories which sample more of phase space. Our two simulations of the oRNA system were trapped in the incorrect conformation even though the simulations differed in the timestep used and the frequency of the non-bonded pairlist update. Similar to the MMD method, these differences in the simulation conditions generated independent trajectories which should have increased the overall sampling. However, because of the intrinsic rigidity of RNA, we observed increased sampling only when we changed the starting geometry of the system from RNA to DNA.

We have also examined the behavior of the U5:2′OH group in the simulation starting from the correct NMR structure and in the simulation starting from the converged structure. Allain & Varani (1995), in their *in vacuo* refinement of the new loop conformation, had found that this moiety makes hydrogen bonds to the O6 atom of the loop closing guanosine residue (G8). They postulated that this hydrogen bond was responsible for the remarkable thermodynamic stability of the UNCG class of RNA tetraloops. However, our results show that once the system is solvated, the preferred hydrogen-bond acceptor for the U5:2′OH is not the G8:O6 atom, but rather the backbone O5′ atom of the following residue (U6). This hydrogen bond is very stable during the simulation which is consistent with the reduced rate of exchange observed in both of the NMR experiments. Also, it does not violate any of the five distance restraints derived from the NOEs. The interactions of this particular hydroxyl group may indeed provide the stability to this system, but our results from solvated simulations indicate a different interaction than the *in vacuo* results do.

We are continuing our MD investigations into the atomic-level interactions of the loop hydroxyls of this tetraloop in an attempt to understand the remarkable stability of this system. Presently, these investigations are restricted to MD simulations and it may be necessary to perform free energy calculations in order to fully understand which hydroxyl groups provide the stability. However, we feel that the results presented here are very encouraging to not only those engaged in theoretical studies of nucleic acids but also to those researchers interested in the unusual structural characteristics found in many RNA sequences.

Methods

The two NMR structures served as the starting points for the simulations presented herein. Because our investigations were focused on the loop portions of the structure, we shortened the stem from the new NMR structure and changed its sequence to match that used in the first structure determination. The naming conventions for the five MD simulations reported herein are as follows: nRNA and oRNA refer to the simulations starting from the new or old structures, respectively; nDNA and oDNA refer to chimeric simulations where the loop riboses were replaced with deoxyribose; and oD(R)NA refers to a control simulation of an RNA structure and is discussed in more detail below.

AMBER 4.1 (Pearlman *et al*., 1995) and the force field of Cornell *et al*. (1995) were used in all of the calculations presented here. We included full charges on the phosphate groups and a neutralizing number of sodium counterions (11) which were initially placed by the EDIT module of AMBER 4.1. This solute-plus-counterions system was then surrounded by a periodic box of TIP3P water molecules. Each simulation box was approximately $50\ \text{Å} \times 43\ \text{Å} \times 40\ \text{Å}$ and contained about 2200 water molecules. All simulations were carried out in the NPT ensemble with periodic boundary conditions, a temperature of 300 K and at a pressure of 1 atm. The temperature was maintained by the Berendsen coupling algorithm with separate solute-solvent and solvent-solvent coupling constants of 0.2 ps (Berendsen *et al*., 1984). Constant pressure was maintained with isotropic molecule based scaling (Berendsen *et al*., 1984). SHAKE (Ryckaert *et al*., 1977) was applied to all bonds involving hydrogen atoms (X-H). The integration timestep was either 1.5 (oRNA, oDNA) or 2.0 (nRNA, nDNA, oD(R)NA) fs and the non-bonded pair list was updated every ten steps. The reduced timestep of 1.5 fs was required to obtain stable trajectories on the Cray T3D for the two simulations starting in the incorrect conformation. The coulomb interactions were treated with the PME method and the Lennard-Jones interactions were subjected to a 9 Å cutoff. The PME charge grid spacing was ~1.0 Å and was interpolated on a cubic B-spline with the direct sum tolerance set to 10^{-5}. A small amount of energy drain occurs in the simulations due to the use of SHAKE, a non-bonded pairlist update of greater than 1, and constant pressure conditions. An artifact of the Berendsen coupling algorithm, which uniformly scales the velocities to maintain the temperature, is that the center-of-mass velocity of the simulation cell can slowly grow. To overcome this problem, at each restart of the simulation (every 25 ps) the net center-of-velocity was removed. In some of our analysis of the nRNA simulation, we compared the MD results to the family of NMR structures deposited in the Brookhaven Protein Data Bank (1hlx). Analysis of the trajectories was done with the MDANAL, CARNAL, and RDPARM modules of AMBER 4.1, MOIL-View (Simmerling *et al*., 1995), UCSF MidasPlus (Ferrin *et al*., 1988), and Dials & Windows (Ravishanker *et al*., 1989). All simulations were run on the Cray T3D at the Pittsburgh Supercomputing Center, an HP 735, or an IBM RS6000.

The various systems were equilibrated as follows: 25 kcal/mol restraints were placed on all solute atoms including the counterions. The water was minimized for 1000 steps, followed by 3 ps of 300 K MD which allowed the solvent to relax around the solute. This was followed by five rounds of 600 step minimizations on the entire system, reducing the solute restraints by 5 kcal/mol during each round. In the final step, the entire system, with no restraints, was heated to 300 K over 10 ps, then equilibrated for another 25 ps. This gave a total thermalization and equilibration time of 35 ps. All production MD runs started from this point.

Acknowledgments

P. A. K. acknowledges research support from the NIH through grant CA-25644. J. L. M. is grateful for research support from the NIH (Pharmaceutical Chemistry, Pharmacology, Toxicology, GM07175) and the American Foundation for Pharmaceutical Education. We thank the Computer Graphics Laboratory for the use of their facilities (T. Ferrin, director, HIH RR-1081) and the Pittsburgh Supercomputing Center for their resources. Both P. A. K. and J. L. M. thank Ignacio Tinoco, Jr and Gabriele Varani for providing us with the coordinates of the tetraloop structures and the restraints used in their NMR structure refinements. J. L. M. thanks Carlos Simmerling, Thomas Fox, Tom Cheatham, and Jed Pitera for helpful discussions.

References

Allain, F. H. T. & Varani, G. (1995). Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **250**, 333–353.

Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids,* Clarendon Press, London.

Antao, V. P. & Tinoco, I., Jr (1992). Thermodynamic parameters for loop formation in RNA and DNA hairpins. *Nucl. Acids Res.* **20**, 819–824.

Antao, V. P., Lai, S. Y. & Tinoco, I., Jr (1991). A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucl. Acids Res.* **19**, 5901–5905.

Auffinger, P. & Westhof, E. (1996). H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys. J.* **71**, 940–954.

Auffinger, P., Louise-May, S. & Westhof, E. (1995). Multiple molecular dynamics simulations of the anticodon loop of tRNA(Asp) in aqueous solution with counterions. *J. Am. Chem. Soc.* **117**, 6720–6726.

Auffinger, P., Louise-May, S. & Westhof, E. (1996). Molecular dynamics simulations of the anticodon hairpin of tRNA(Asp): structuring effects of C-H:O hydrogen bonds and of long-range hydration forces. *J. Am. Chem. Soc.* **118**, 1181–1189.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.

Beveridge, D. L. & Ravishanker, G. (1994). Molecular dynamics studies of DNA. *Curr. Opin. Struct. Biol.* **4**, 246–255.

Brünger, A. T., Brooks, C. L., III & Karplus, M. (1985). Active site dynamics of ribonuclease. *Proc. Natl Acad. Sci. USA,* **82**, 8458–8462.

Cheatham, T. E., III & Kollman, P. A. (1996). Observation of the *A*-DNA to *B*-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **259**, 434–444.

Cheatham, T. E., III & Kollman, P. A. (1997). Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* **119**, 4805–4825.

Cheatham, T. E., III, Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. (1995). Molecular dynamics simulations on solvated biomolecular systems: the Particle Mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117**, 4193–4194.

Cheong, C., Varani, G. & Tinoco, I., Jr (1990). Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature,* **346**, 680–682.

Cornell, W. D., Cieplak, P., Bayly, C. I. & Kollman, P. A. (1993). Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620–9631.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.

Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prange, T. & Kennard, O. (1994). Structure of a mispaired RNA double helix at 1.6 Angstrom resolution and implications for the prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA,* **91**, 4160–4164.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). A smooth Particle Mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13–27.

Fritsch, V., Ravishanker, G., Beveridge, D. L. & Westhof, E. (1993). Molecular dynamics simulations of Poly(dA)·Poly(dT): comparisons between implicit and explicit solvent representations. *Biopolymers,* **33**, 1537–1552.

Fujiwara, T. & Shindo, H. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Molecular motions in hydrated DNA. *Biochemistry,* **24**, 896–902.

Heus, H. A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science,* **253**, 191–194.

Holbrook, S. R., Cheong, C., Tinoco, I., Jr & Kim, S.-H. (1991). Crystal Structure of an RNA double helix incorporating a track of non-Watson-Crick base-pairs. *Nature,* **353**, 579–581.

Jucker, F. M. & Pardi, A. (1995). Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry,* **34**, 14416–14427.

Louise-May, S., Auffinger, P. & Westhof, E. (1995). *Ninth Conversation in the Discipline of Biomolecular Stereodynamics,* State University of New York, Albany, NY.

McCammon, J. A. & Harvey, S. (1987). *Dynamics of Proteins and Nucleic Acids,* University Press, Cambridge.

McConnell, K. J., Nirmala, R., Young, M. A., Ravishanker, G. & Beveridge, D. L. (1994). A nanosecond molecular dynamics trajectory for a *B*-DNA double helix: evidence for substates. *J. Am. Chem. Soc.* **116**, 4461–4462.

Miller, J. L. (1996). *Solvation, Structure, and Dynamics of Nucleic Acids: Insights from Simulations,* University of California, San Francisco.

Molinaro, M. & Tinoco, I., Jr (1995). Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucl. Acids Res.* **23**, 3056–3063.

Noonberg, S. H., Scott, G. K., Garovoy, M. R., Benz, C. C. & Hunt, C. A. (1994). *In vivo* generation of highly abundant sequence-specific oligonucleotides for antisense and triplex gene regulation. *Nucl. Acids Res.* **22**, 2830–2836.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E. III, Ferguson, D. M., Seibel, G. L., Singh, U. C., Wiener, P. K. & Kollman, P. A. (1995). *AMBER 4.1,* University of California, San Francisco.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature,* **372**, 68–74.

Portmann, S., Usman, N. & Egli, M. (1995). The crystal structure of r(CCCCGGGG) in two distinct lattices. *Biochemistry,* **34**, 7569–7575.

Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. & Sklenar, H. (1989). Conformational and helicoidal analysis of 30 ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: ''curves'', dials and windows. *J. Biomol. Struct. Dynam.* **6**, 669–699.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Phys.* **23**, 327–341.

Saenger, W. (1984). *Principles of Nucleic Acid Structure* (Cantor, C. R., ed.), Springer Advanced Texts in Chemistry, Springer-Verlag, New York.

Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M. & Uesugi, S. (1990). Studies on the structure and stabilizing factor of the CUUCGG hairpin RNA using chemically synthesized oligonucleotides. *Nucl. Acids Res.* **18**, 3831–3839.

Shindo, H., Fujiwara, T., Akutsu, H., Matsumoto, U. & Kyogoku, Y. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Static geometry of DNA double helices. *Biochemistry,* **24**, 887–895.

Simmerling, C., Elber, R. & Zhang, J. (1995). MOILview: a program for visualization of structure and dynamics of biomolecules and STO: a program from computing stochastic paths. In *Modelling of Biomolecular Structures and Mechanisms* (Pullman, *et al*., eds), pp. 241–265, Kluwer Academic Publishers, Amsterdam, Netherlands.

Srinivasan, J., Withka, J. M. & Beveridge, D. L. (1990). Molecular dynamics of an in vacuo model of duplex d(CGCGAATTCGCG) in the *B*-form based on the AMBER 3.0 force field. *Biophys. J.* **58**, 533–547.

Swaminathan, S., Harte, W. E., Jr & Beveridge, D. J. (1991a). Investigation of domain structure in proteins *via* molecular dynamics simulation: application to HIV-I protease dimer. *J. Am. Chem. Soc.* **113**, 2717–2721.

Swaminathan, S., Ravishanker, G. & Beveridge, D. L. (1991b). Molecular dynamics of *B*-DNA including water and counterions: a 140-ps trajectory for d(CGCGAATTCGCG) based on the GROMOS force field. *J. Am. Chem. Soc.* **113**, 5027–5040.

Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I. J., Brody, E. N. & Gold, L. (1998). CUUCGG hairpins: extraordinarily stable RNA secondary structure associated with various biochemical processes. *Proc. Natl Acad. Sci. USA,* **85**, 1364–1368.

Varani, G. (1995). Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379–404.

Varani, G., Cheong, G. & Tinoco, I., Jr (1991). Structure of an unusually stable RNA hairpin. *Biochemistry,* **30**, 3280–3289.

Weerasinghe, S., Smith, P. E., Mohan, V., Cheng, Y. K. & Pettitt, B. M. (1995). Nanosecond dynamics and structure of a model DNA triple helix in saltwater solution. *J. Am. Chem. Soc.* **117**, 2147–2158.

Wimberly, B., Varani, G. & Tinoco, I., Jr (1991). Structural determinants of RNA function. *Curr. Opin Struct. Biol.* **1**, 405–409.

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.

Yang, L. & Pettitt, B. M. (1996). *B* to *A* transition of DNA on the nanosecond time scale. *J. Phys. Chem.* **100**, 2564–2566.

Zichi, D. A. (1995). Molecular Dynamics of RNA with the OPLS Force Field. Aqueous simulations of a hairpin containing a tetranucleotide loop. *J. Am. Chem. Soc.* **117**, 2957–2969.

Edited by I. Tinoco Jr

(Received 6 November 1996; received in revised form 7 April 1997; accepted 7 April 1997)