Observation of an A-DNA to B-DNA Transition in a Nonhelical Nucleic Acid Hairpin Molecule Using Molecular Dynamics

Jennifer L. Miller and Peter A. Kollman
Department of Pharmaceutical Chemistry, School of Pharmacy, University of California San Francisco, San Francisco, California 94143-0446 USA

ABSTRACT One of the truly challenging problems for molecular dynamics (MD) simulations is demonstrating that the trajectories can sample not only in the vicinity of an experimentally determined structure, but also that the trajectories can find the correct experimental structure starting from some other structure. Frequently these transitions to the correct structure require that the simulations overcome energetic barriers to conformational change. Here we present unrestrained molecular dynamics simulations of the DNA analogs of the RNA 5′-GGACUUCGGUCC-3′ hairpin tetraloop. In one simulation we have used deoxyuracil residues, and in the other we have used the native DNA deoxythymine residues. We demonstrate that, on a nanosecond time scale, MD is able to simulate the transitions of both of the A-DNA stems to B-DNA stems within the constraints imposed by the four-base loop that caps the helix. These results suggest that we are now in a position to use MD to address the nature of sequence-dependent structural effects in nonduplex DNA structures.

INTRODUCTION

Proteins interact with nucleic acids in many important cellular processes. Some of these interactions are nonspecific in nature, such as the packaging of DNA into nucleosomes and the interaction of DNA with nucleases and polymerase. Many others require the specific recognition of DNA or RNA sequences by the protein. Among these are DNA repressors, activators, and restriction endonucleases. Some of these protein-nucleic acid complexes require a conformational change, or distortion, of the DNA or RNA solution structure. A striking example of this can be seen in the interaction of DNA with CAP (catabolite activating protein), where the complex shows a 90° bend in the DNA duplex (Schultz et al., 1991). An atomic-level understanding of both the sequence-dependent fine structure and the intrinsic flexibility of DNA and RNA is an active area of research of many groups. Such an understanding will be helpful in designing therapeutics that interfere with the protein-nucleic acid interactions required in certain disease processes.

Among the tools used to determine and assess the structure and flexibility of DNA and RNA in solution are x-ray crystallography, NMR, and theoretical methods such as molecular dynamics (MD). Each of these methods has provided insights, but each suffers from some intrinsic limitations. Many nucleic acid sequences are difficult to crystallize, and those that do are subject to packing effects that may or may not represent the solution-phase structure that is recognized by proteins (Dickerson et al., 1987, 1994; Lipanov et al., 1993). NMR-derived structures are becoming increasingly available, but the NOEs observed are limited in range to distances less than 5 Å (Schmitz and James, 1995). This limitation hinders efforts to determine the structure of nonhelical nucleic acid structures in which regions that contain loops and bulges—regions important in many RNA-protein interactions—are frequently underdetermined by the experimental data. MD has traditionally been used both to refine x-ray and NMR structures and to investigate the structure and dynamics of nucleic acid structure in the vicinity of the experimentally determined structure. However, efforts to use MD to study larger-scale motions that contain intrinsic energetic barriers have been limited by the amount of sampling achievable in the simulations (Miller and Kollman, 1996).

There have been a few reports of the successful application of MD methods to simulate conformational transitions in nucleic acid structures. Yang and Pettitt reported on the B-to-A transition of a dodecamer of DNA (Yang and Pettitt, 1996). Their simulation, which was run for 3.5 ns, utilized the CHARMM (Brooks et al., 1983) force field and included a solvent of 0.45 M salt water. Cheatham and Kollman have reported on two different transitions (Cheatham and Kollman, 1996, 1997), one unrestrained and the other restrained; both studies used the Cornell et al. force field and standard solvent conditions (i.e., a neutralizing number of counterions). In the first, they showed that the DNA duplex d(CCAACGTTGG)$_2$ underwent an A-form to B-form transition on a nanosecond time scale in unrestrained MD simulations. In the second study, they performed restrained simulations of an RNA duplex r(CCAACGUUGG)2 where, by forcing a concerted flip in the sugar puckers from C2′-endo to C3′-endo, the B-RNA was converted into A-RNA. Miller and Kollman, using the same force field and solvent conditions as the studies of Cheatham, reported on the convergence of an RNA tetraloop from an incorrect to the correct NMR structure (Miller and Kollman, 1996). This convergence required the removal of the 2′ hydroxyls in the loop residues to improve the

Received for publication 7 April 1997 and in final form 1 August 1997.
Address reprint requests to Dr. Peter A. Kollman, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143. Tel.: 415-476-4637; Fax: 415-476-0688; E-mail: pak@cgl.ucsf.edu.

0006-3495/97/11/2702/09 $2.00

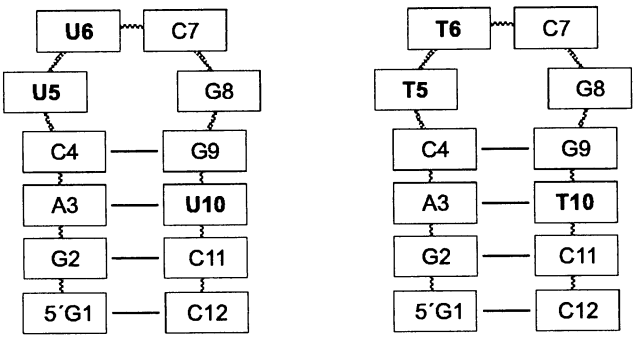

FIGURE 1 Schematic diagram of the two sequences used in the MD simulations presented here. The residues that differed between the S1 and S2 simulations are shown in bold.

sampling and overcome the energetic barriers inherent in simulating RNA structures.

In this manuscript, we extend the work of both Cheatham and Miller and report the simulation of an A-form to B-form transition in a nonhelical nucleic acid structure. We have used the same RNA tetraloop structure as in our previous work. However, instead of using the native RNA sequence, we have replaced all of the riboses with deoxyribose. This gave us an A-form DNA tetraloop as the starting point for our simulations. We have performed two simulations of the DNA analog, the first using deoxyuracil residues and the second using deoxythymine residues (Fig. 1). NMR studies of this tetraloop have shown that the RNA sequence r(GGACUUCGGUCC) contains an A-form stem with a rigid four-base loop (Allain and Varani, 1995; Varani et al., 1991). Other experimental studies have shown that the DNA analog of this sequence, d(GGACTTCGGTCC), contains a B-form stem and a flexible four-base loop (James and Tinoco, 1993; Sakata et al., 1990). These studies, as well as our previous simulations of this system, provide excellent benchmarks for the evaluation of our current simulations.

METHODS

We utilized the force field of Cornell et al. (1993) and the AMBER 4.1 (Pearlman et al., 1995) suite of programs in all of the simulations presented in this work. Each simulation system consisted of a 12-residue solute (Table 1) with full charges on the phosphates and a neutralizing number of sodium counterions (11 counterions). The counterions were initially placed by the EDIT module of AMBER 4.1, and then the system was surrounded by a periodic box of TIP3P waters. All simulations were carried out with periodic boundary conditions, a constant temperature of 300 K, and a constant pressure of 1 atm. The temperature was maintained by the Berendsen (Berendsen et al., 1984) coupling algorithm, with separate solute-solvent and solvent-solvent coupling constants of 0.2 ps each. 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 time step was 2.0 fs, and the nonbonded pair list was updated every 10 steps. The Coulomb interactions were treated with the particle mesh Ewald (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}. As in our previous work (Miller and Kollman, 1996), we removed the net center of velocity at each restart (every 25 ps) of the simulation to correct for the small energy drains due to the use of SHAKE, a nonbonded pair list update every 10 steps, and constant pressure conditions. Analysis of the trajectories was done with AMBER 4.1, MOIL-View (Simmerling et al., 1995), UCSF MidasPlus (Ferrin et al., 1988), and Dials and Windows (Ravishanker et al., 1989). All simulations were run on the Cray T3D at the Pittsburgh Supercomputing Center.

To generate the deoxyribose analogs of the r(GGACUUCGGACC) tetraloop, we began with the coordinates from the NMR structure of the P1 helix as reported by Allain and Varani (1995). Because our interest was in comparing these simulations to our previous work on the RNA UUCG tetraloop, we retained the coordinates of the loop atoms but altered the stem sequence to match the one from the earlier structural studies (Varani et al., 1991). The new coordinates were obtained by shortening the stem and replacing the residue names to correspond to the previous sequence. The EDIT module of AMBER replaced the appropriate atoms, giving starting structures with the same sequence as the earlier NMR structure, but with the loop conformation of the more recent structure. Both starting structures contained deoxyribose sugars instead of the native riboses. The S1 system contained deoxyuracil in place of the native uridine residues, and the S2 system contained thymidine in place of the uridines. To create the deoxyuracil (dU) residues, we followed a procedure similar to the one used to obtain the

TABLE 1 Simulations presented in this work

Name	Sequence	No. of water molecules	Box size (Å^3)	Simulation length (ps)	RMSD (Å)	
S1	d(GGACUUCGGUCC)	2189	48 × 44 × 39	2547	All	2.3
					Stem	1.6
					Loop	2.1
S2	d(GGACTTCGGTCC)	2218	47 × 44 × 39	2627	All	2.4
					Stem	1.5
					Loop	2.3

nucleoside charges in the force field of Cornell et al. (1995): the charges from the uridine base atoms (rU) were retained and combined with the sugar charges from deoxythymine (dT). The C1′ and H1′ atom charges were adjusted slightly to maintain the neutrality of the residue.

The two simulation 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. The production MD runs started from this point.

RESULTS AND DISCUSSION

We have run two separate nanosecond-length simulations of the deoxyribose analogs of the RNA tetraloop sequence r(GGACUUCGGACC). These studies were undertaken to investigate whether unrestrained MD could be used to observe an A-form to B-form transition in a nucleic acid system that contained both helical and nonhelical tertiary structure. The structure used in this work is a hairpin molecule that contains a 4-bp helical stem capped by a 4-base loop. We have previously established that the MD simulation of the native RNA tetraloop, in which the stem is A-form, is very stable and stays very close to the experimental NMR structure (Miller and Kollman, 1996). The two starting structures for the simulations presented here differed in sequence in that only the first one (S1) contained all deoxyribose residues and the same base sequence as the native RNA molecule. The second one (S2) contained all deoxyribose sugars as well, but had thymine residues instead of uracil. We begin our discussion with a comparison of these two simulations to three recent MD studies of nucleic acid structures, using the same force field of Cornell et al. and the PME method. These studies include the unrestrained MD simulations of the native RNA tetraloop (nRNA) (Miller and Kollman, 1996), the DNA duplex d(CCAACGTTGG)$_2$ (Cheatham and Kollman, 1996), and the RNA duplex r(CCAACGUUGG)$_2$ (Cheatham and Kollman, 1997). This comparison to theoretical studies is followed by a comparison to the experimental NMR studies of the all-DNA analog of the RNA tetraloop (James and Tinoco, 1993; Sakata et al., 1990).

Both simulations converge to the same B-form stem structure

In Fig. 2 we show the all of heavy atom root mean square deviation (RMSD) plots for the S1, S2, and RNA simulations. All three of these simulations started in the same

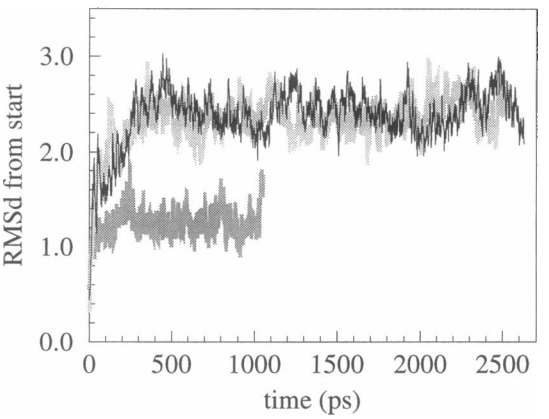

FIGURE 2 Time course of the all-atom RMSDs for the S1 (*light gray*), S2 (*black*), and nRNA (*dark gray*) simulations.

conformation. Whereas the RNA simulation is very stable and only moves ~1.3 Å away from the starting structure, both DNA trajectories move much farther away (~2.5 Å). This divergence from the starting structure is notable in light of our previous simulations of the native and chimeric (r[GGAC]d[UUCG]r[GUCC]) sequences of this tetraloop (Miller and Kollman, 1996). In that work, we demonstrated that even though the loop flexibility was increased with the removal of the 2′ hydroxyls, the simulation stayed very close to its starting structure (~1.3 Å RMSD), and the average structure maintained an A-form stem. Moreover, the all-atom RMSD between the average structures from the native and chimeric sequences was only 1.5 Å. This low RMSD is remarkable because it includes the effects of relaxation away from the experimentally determined structure and the removal of the loop hydroxyls. Clearly, whereas the removal of the loop 2′ hydroxyls did not significantly alter the structure, the removal of the stem 2′ hydroxyls did.

A one-dimensional RMSD plot is able to show only the extent of deviation from a static structure and does not contain any information about whether the simulation is converging to another conformation. To investigate whether the simulations were converging to another conformation, we calculated the 2D-RMSD map for the two new trajectories. We show this map for the S1 simulation in Fig. 3, where the RMS fit and the RMSD calculation were performed on all of the heavy atoms in the stem. It is clear from this figure that the simulation has converged to another conformation by ~300 ps. The S2 map demonstrates the same behavior, although the time to convergence is slightly longer (~350 ps, data not shown). A comparison of the average stem structures (after convergence) to the average stem structures from the native RNA simulation and to canonical A and B forms for this same sequence is presented in Table 2. This table shows that the S1 and S2 stems converge to essentially the same structure, with an RMSD between them of only 0.4 Å. This result is not unexpected, because the two stem sequences differed by only one residue. Not surprisingly, as indicated from the 1D-RMSD

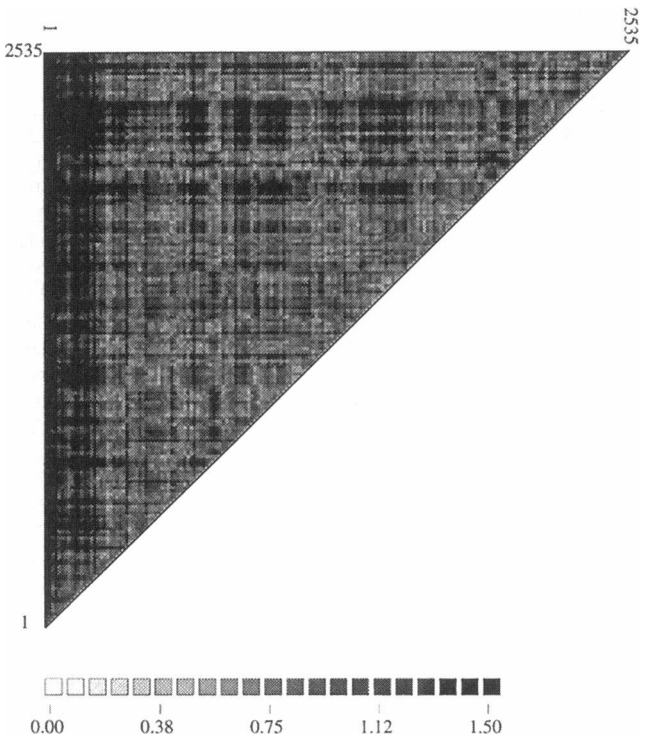

FIGURE 3 2D-RMSD map for the S1 simulation, showing the transition to a new conformation at ~300 ps. The labels indicate the time during the trajectory. As indicated in the legend, darker squares correspond to higher RMSD values.

plots, the converged structures for the DNA stems are 1.9 (S1) and 1.8 (S2) Å away from the converged stem of the native RNA sequence. The comparison of S1 and S2 with the canonical forms shows a more dramatic difference with RMSDs of 2.4 and 2.3 Å, respectively, from the A-form, and only 1.5 and 1.4 Å from the B-form.

Although the RMSD results provide some evidence that the simulations have converged to B-form stems, a more accurate and sensitive measure of nucleic conformation comes from an examination of the backbone and glycosidic torsion angles, the sugar pucker pseudorotation angles, the helicoidal parameters, and the minor groove width. Table 3 contains these values for the converged parts of the S1 and S2 simulations, as well as for the native RNA simulation. We also include in Table 3 these values calculated from simulations of a DNA duplex (Cheatham and Kollman, 1996) and an RNA duplex (Cheatham and Kollman, 1997). The average values for the S1 and S2 trajectories are, in general, closer to the average B-DNA values than to either the native RNA tetraloop or the A-RNA values. This is certainly true for the torsional angles and sugar puckers, where the S1 and S2 simulations, which started in the nRNA conformation, have converged to within a few degrees of the B-DNA values. The one exception is the backbone angle ζ, where the converged angle (~270°) is midway between the B-DNA (~258°) and RNA averages (~292°). This difference could be due to the simulations not having fully converged, although the average angle is within the experimentally determined *gauche* range (James and Tinoco, 1993). As in the B-DNA simulation, we are slightly underestimating the expected χ and δ angles based on NMR solution data of DNA duplexes (χ = 247°, δ = 138°) (Ulyanov and James, 1995). However, we feel that this agreement between the simulation averages of these two hairpin structures and the duplex DNA—which is not subject to the constraints of a 4-base cap—is rather remarkable and shows that the simulations are able to overcome barriers and sample a wide range of torsional states.

TABLE 2 Pairwise RMSDs between stem sequences (Å)

	A-DNA	B-DNA	S1	S2	nRNA
A-DNA	—	2.3	2.4	2.3	1.1
B-DNA		—	1.5	1.4	2.2
S1			—	0.4	1.9
S2				—	1.8
nRNA					—

The RMS fits were done on the middle two base pairs of the stem, and the RMSDs were calculated using all atoms, except for the thymine C7 hydrogens and the 5′ and 3′ terminal hydrogens.

In turning to the helicoidal parameters presented in Table 3, we see that the effects of the four loop bases and relatively short stem sequence are more pronounced. Although the reported values from the two duplex simulations do include the terminal base pairs, the averages are dominated by the eight internal base pairs. For the hairpin simulations there are only two internal base pairs, and the calculated averages include both the terminal base pair and the base pair adjacent to the loop. Hence our analysis of the convergence of the helicoidal parameters focuses on the trends in the data between the A and B forms of the two duplexes and the three hairpin molecules. From this perspective, we see that the differences between the A and B forms is reproduced for the hairpin simulations. In all cases, where there is a increase or decrease in the average value of a parameter in going from A- to B-form in the duplex simulations, there is a similar change in the hairpin simulations. This is trend is evident in the propeller twist, buckle (S2 only), opening, rise, tilt, roll, tip, x displacement, and inclination. In addition, whereas the duplex simulations did not demonstrate the expected increase in the base-pair twist, the hairpin simulations did. For the other parameters, either we do not see any significant difference between the two forms, or the averages for the hairpins do not reflect the differences observed in the duplexes. For example, the base-pair stretch is one parameter for which the differences are not reproduced. However, when we take the average of this parameter over the two internal base pairs only, we do see the same change as in the duplex simulations (data not shown). We also see an increase in the end-to-end length of the short helix from the 9.7 Å in the RNA molecule to ~10.7 Å for the DNA molecules, a change that is expected in going from A-form to B-form. Moreover, the minor groove width shows a significant decrease, mirroring the change observed in the duplex simulations.

TABLE 3 Average values for the standard angle and helicoidal parameters

	B-form helices			A-form helices	
	S1	S2	B-DNA	nRNA	A-RNA
α (°)	292.5 (10.9)	290.8 (18.3)	290.4 (11.6)	277.9 (16.9)	277.0 (10.5)
β (°)	171.4 (15.3)	170.8 (14.4)	168.4 (12.6)	160.5 (17.1)	175.6 (9.5)
γ (°)	55.8 (10.5)	56.2 (11.3)	54.3 (10.6)	78.8 (10.4)	69.8 (8.9)
δ (°)	120.6 (20.0)	120.4 (21.4)	116.6 (18.0)	79.7 (12.6)	79.3 (8.2)
ε (°)	188.6 (25.3)	190.7 (20.2)	197.0 (18.5)	203.3 (12.0)	201.6 (10.1)
ζ (°)	269.7 (33.2)	270.7 (27.1)	258.0 (25.5)	292.8 (11.1)	291.4 (8.6)
χ (°)	234.5 (22.5)	236.8 (41.6)	234.2 (16.7)	199.7 (59.8)	201.7 (9.1)
Pucker (°)	132.2 (31.2)	131.0 (37.1)	122.8 (28.1)	18.0 (23.4)	22.6 (16.5)
Propeller (°)	−5.8 (13.5)	−5.1 (15.1)	−10.4 (12.3)	−6.2 (12.3)	−12.6 (12.0)
Buckle (°)	−3.8 (10.8)	−1.7 (14.7)	0.4 (11.4)	−4.0 (11.8)	−0.8 (10.6)
Opening (°)	2.1 (7.3)	2.3 (7.2)	2.0 (5.6)	2.6 (7.2)	3.3 (6.0)
Rise (Å)	3.6 (0.6)	3.6 (0.6)	3.3 (0.5)	3.3 (0.8)	2.7 (0.6)
Tilt (°)	2.2 (6.7)	0.9 (7.0)	0.0 (5.4)	−1.2 (6.1)	−0.5 (5.4)
Roll (°)	7.8 (8.4)	8.0 (8.8)	1.3 (8.7)	11.1 (10.2)	2.4 (7.7)
Twist (°)	30.2 (5.7)	30.2 (5.9)	30.9 (5.1)	28.6 (5.5)	30.9 (4.3)
x Disp (Å)	−1.6 (1.2)	−1.3 (1.2)	−3.0 (0.7)	−2.8 (1.8)	−5.2 (0.8)
y Disp (Å)	0.1 (0.9)	0.3 (1.0)	0.0 (0.5)	0.2 (1.1)	0.1 (0.7)
Inclination (°)	−8.9 (9.8)	−7.7 (11.6)	4.9 (7.3)	−6.6 (15.6)	15.0 (9.3)
Tip (°)	2.0 (8.4)	1.6 (8.9)	0.4 (5.8)	−1.5 (10.3)	−1.9 (6.7)
Shear (Å)	0.0 (0.5)	0.0 (0.5)	0.0 (0.4)	0.1 (0.4)	0.0 (0.5)
Stretch (Å)	0.1 (0.3)	0.1 (0.3)	0.1 (0.2)	0.2 (0.4)	0.4 (0.4)
Stagger (Å)	0.0 (0.6)	0.0 (0.6)	−0.2 (0.5)	−0.3 (0.6)	−0.2 (0.5)
Shift (Å)	0.1 (0.6)	0.0 (0.6)	0.0 (0.6)	0.0 (0.6)	0.0 (0.6)
Slide (Å)	−0.1 (0.4)	−0.2 (0.5)	−0.1 (0.4)	−0.3 (0.4)	−0.1 (0.4)
Width (Å)	13.4 (1.6)	13.9 (1.6)	12.4 (1.4)	17.2 (0.7)	17.0 (0.5)
Length (Å)	10.8 (0.9)	10.7 (0.9)	N/A	9.7 (0.5)	N/A

Backbone torsions and helicoidal parameters are the average over all stem residues (S1, S2, nRNA) and over all residues (B-DNA, A-RNA). The width was determined by averaging over the P5–P12 phosphate distance in the tetraloop simulations and over selected base pairs in the double-helix molecules. Standard deviations for all values are in parentheses. The deviations shown for S1, S2, and nRNA are the maximum standard deviations calculated from 1-ps snapshots.

Time course of the transition

Whereas the 2D-RMSD maps (a measure of similarity in Cartesian coordinate space) indicated that the S1 and S2 structures had converged to a different structure, and the average helix parameters indicated that this converged structure contains a B-form stem, a closer examination of the time course of some of the torsion angle and helicoidal parameters reveals some insights into the pathway of this transition. Fig. 4 contains the graphs of five of these parameters over the course of the trajectories. The glycosidic torsion angle (χ) and the sugar pucker demonstrate the transition on the nucleotide level; the x displacement, end-to-end length, and minor groove width capture, in a global sense, all of the changes in the backbone that occur during the transition. Furthermore, all five parameters are sensitive to the differences between the A and B forms and, therefore, are good indicators of conformational change in this small 4-bp stem. To construct the graphs of the χ angle and sugar pucker, averages were taken over all of the stem residues. For the x displacement, averages were taken over the appropriate stem base pairs. Finally, the data have been smoothed in each of the graphs by performing a running average in time over 25 ps before plotting.

Both of the torsion angles shown in Fig. 4, *a* and *b*, indicate a rather fast transition from A-form to B-form. The sugar pucker pseudorotation angle undergoes the most distinct change. This angle displays a very fast rise away from the A-form C3′-endo, eventually converging (~300 ps) to an average value around 130°, which corresponds roughly to a conformation of C1′-exo. This average pucker value is in excellent agreement with a recent survey of NMR solution data from DNA structures, which found that the expected pseudorotation value is 138° (Ulyanov and James, 1995), but disagrees with the NMR data for this particular system (discussed in a later section). As expected from theoretical studies (Olson, 1982), the pathway goes through O4′-endo rather than the higher energy O4′-exo pucker state. It should be noted, however, that the pucker pseudorotation value plotted in Fig. 4 *a* represents an average over both the trajectory and all of the stem residues. Although this average value demonstrates a difference between the A- and B-forms, it does not reflect either the behavior of the individual residues or the two-state behavior of this parameter as the sugars repucker between the C2′ and C3′-endo states. An examination of the individual puckers showed that all went through the same transition pathway (through O4′-endo; data not shown). The averages ranged between 110° and 154° for the S1 simulation and between 116° and 147° for the S2 simulation. The glycosidic torsion angle also undergoes a distinct change during the transition, al-

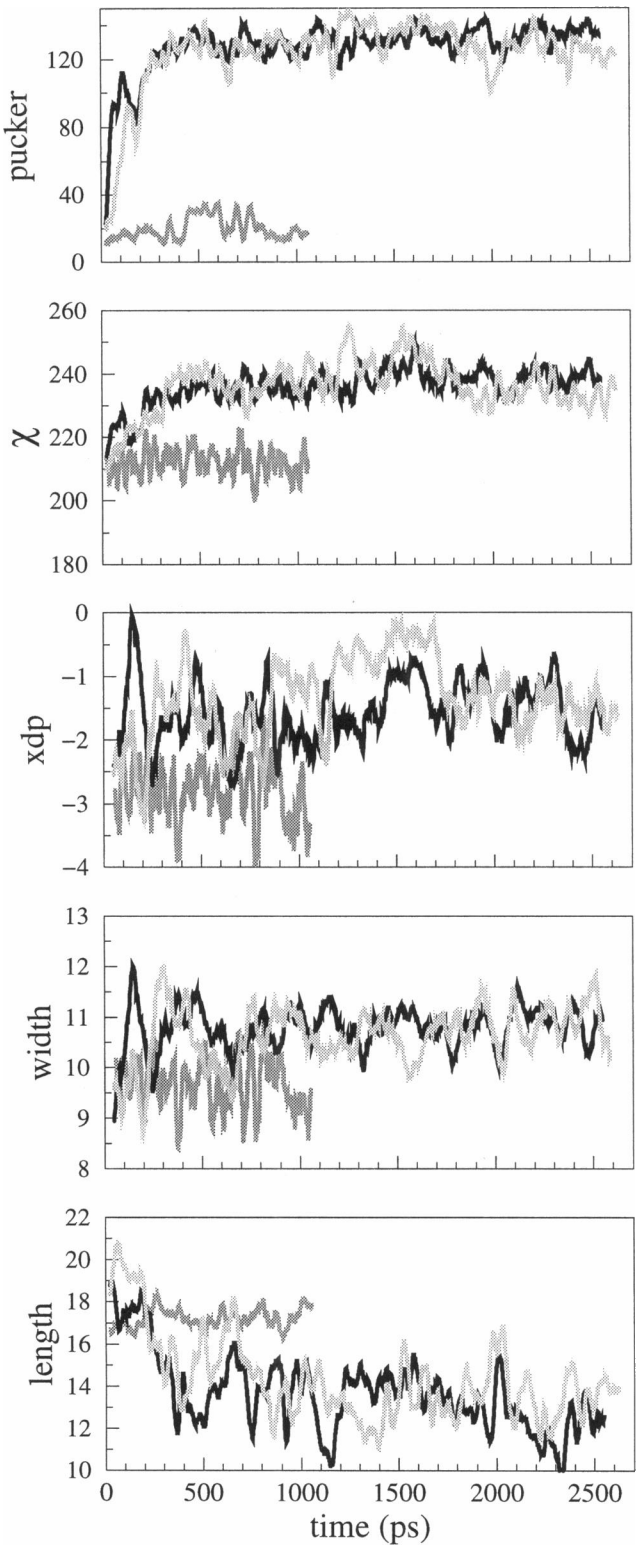

FIGURE 4 Time course of various parameters for the stem portions of the tetraloop in the S1 (*light gray*), S2 (*black*), and nRNA (*dark gray*) simulations. (*a*) Sugar pucker pseudorotation angle. (*b*) Glycosidic torsion angle. (*c*) *x*-Displacement. (*d*) End-to-end length. (*e*) Minor groove width.

though the rise to the B-form value is not as rapid as that found for the sugar puckers. This angle, which is a measure of the positioning of the nucleotide base relative to the furanose, has an A-form value in the range of 180°-210° and a B-form value in the range of 210°-270° (Saenger, 1984). As is seen in the graph, the nRNA simulation stays in the A-form range, whereas the S1 and S2 trajectories both converge to the B-form range.

The differences in the *x* displacement between the A-form stem of the nRNA simulation and the stems of the two DNA analogs is not as dramatic as seen in the torsions just discussed. As was seen in the study of the A-to-B transition of a DNA duplex (Cheatham and Kollman, 1996), this parameter varies over a much broader range, even after the transition takes place. It is not clear that this parameter—which is widely used as a measure of distinction between A- and B-forms—is appropriate for these hairpin loop structures with short stems. This is evident from the graph, where the *x* displacement of the A-form stem from the nRNA simulation is quite a bit higher than the canonical A-form value (−5.4 Å). This small displacement is not a force-field effect, as a recent study of an RNA duplex from our group found an average *x* displacement of −5.2 Å (Cheatham and Kollman, 1997). It is more likely that the small values are an artifact of including the terminal and loop-closing base pairs in the determination of a global helical axis required to calculate the displacement. However, the trends observed are consistent with the differences between A- and B-forms, with the base pairs moving closer to the helical axis (smaller displacement) as the two simulations undergo a conformational transition.

The two other parameters plotted in Fig. 4, *d* and *e*, the end-to-end length and minor groove width, are measures of the overall shape of the helix. These values account for the differences in all of the backbone torsions between the A- and B-forms. In general terms, a B-form duplex is longer and narrower than an A-form duplex. Both of these trends are reproduced in the plots in Fig. 4. In the case of end-to-end length, although there is some overlap between the lines, the S1 and S2 traces are generally greater (longer) than the nRNA trace. As was the case for the *x* displacement, this parameter appears to be one that converges very slowly, with a clear distinction between the two forms becoming evident at ~900 ps. Cheatham and Kollman observed a similarly slow convergence in this parameter for the A-DNA to B-DNA transition of a decamer duplex (Cheatham and Kollman, 1996). Because the stem of this hairpin is composed of only 4 bp, the minor groove width values plotted in Fig. 4 *e* reflect the running average in time for only one P–P distance. This distance included the phosphorus atoms of the C11pC12 and the C4pU5 (or T5) steps, the former adjacent to the terminal base pair and the latter part of the hairpin loop, so it is not expected that the calculated values would exactly match the average minor groove widths calculated from simulations of much longer duplex structures. However, there is a clear transition in the values for the S1 and S2 simulations, which actually begin

at a higher value than the nRNA stucture but quickly fall toward a more B-like distance. Both the S1 and S2 trajectories sample a wide range of P–P distances compared to the nRNA trajectory, but seem to have converged to the average of ~13 Å by ~700 ps. Even though the widths of the S1 and S2 trajectories do not converge to the overall average found in the B-DNA simulation (Table 3), the average values reported here do match those found for the terminal P–P distances in that study (14.2 and 14.1 Å) (Cheatham and Kollman, 1997).

Comparison of the average structures to the NMR data

In this section, we briefly compare the average structures, posttransition, of the S1 and S2 simulations to an NMR study that focused on the tertiary structure determination of this system (James and Tinoco, 1993). The base sequence used in the S2 simulation matches the one used in the experiments. Our comparison to this work includes both the stem and loop domains of the hairpin, but it is qualitative because we do not have the atomic coordinates for the family of structures reported in the NMR study.

We begin our comparison with an examination of how well the simulations reproduced the structural constraints determined from the NMR experiments. These constraints are listed in Table 4 and include the backbone and glycosidic torsion angles and the sugar pucker pseudorotation angles. For comparison purposes, we have used the same notation for the values of the backbone angles as was presented in table 3 of the experimental paper (i.e., *g*, *gauche*; *t*, *trans*; etc.). For the α and ζ torsions, the NMR data only indicated a *gauche* conformation, and we have specified whether the simulation contained a *gauche*$^-$ or *gauche*$^+$ conformation. In general, we see excellent agreement between the experimental and theoretical results, especially in the stem region of the molecule. The range of backbone torsional values sampled in the MD simulations agree very well with those determined from the NMR data. We have listed all conformations sampled in the simulations, regardless of the amount of time spent in that conformation. Some of these events were very quick crankshaft motions that transitioned back after only a few picoseconds. The α angle of residues 6 and 7, and the ζ angle of residues 5, 9, and 10 reflect this type of behavior. These short sampling times would most likely not be reflected in the NMR data, as they only account for a small percentage of the conformations. Perhaps more importantly, the simulations were able to capture the multiconformation sampling found in some of the residues, for example in the ϵ angle of residues 5, 6, 9, and 10. The glycosidic torsions are systematically lower than the values determined from the NMR data, but more closely match the values reported after refinement (data not shown, see Table 4 of James and Tinoco, 1993). The one exception is the glycosidic torsion of residue G8, which is clearly in the *anti* conformation in the NMR studies of the DNA analog, but stays very close to the *syn* conformation in which the simulations began. Because the loop domain of this structure is very compact, it is not surprising that this torsion did not change, as doing so would have required a significant amount of structural rearrangement that we are unlikely to observe on a nanosecond time scale. Also from Table 4, we see that the simulations underrepresent the sugar pucker pseudorotation values for most of the residues. These angles are well determined in the NMR studies, and we are still not certain of the reasons for the low values sampled in the MD simulations. The sugar of residue T5 (or U5) does not repucker to the preferred C2′-endo conformation during either of the MD simulations. Again, we feel that this is a deficiency in the amount of sampling and an artifact of the starting conformation, where residues 5 and 8 are hydrogen bonded to each other through the O_2 atom of residue 5 and the N1 and N2 atoms of residue 8.

The NMR data indicated that the three loop imino protons of T5, T6, and G8 do not participate in any hydrogen bonding interactions with other residues. The simulations are consistent with this for the T5 and T6 (or U5 and U6) residues, where these iminos are in contact with the solvent throughout the trajectories. But, as discussed in the previous

TABLE 4 Comparison of NMR-derived structural parameters to the simulation averages*

	P	α	β	γ	ε	ζ	χ
G1	179/130/136	NA	NA	NA	t, g$^-$/t/t	g/g$^-$/g$^-$, g$^+$	212/214/232
G2	174/128/138	g/g$^-$/g$^-$	t/t/t	g$^+$/g$^+$/g$^+$	t, g$^-$/t/t	g/g$^-$/g$^-$	270/225/232
A3	186/111/117	g/g$^-$/g$^-$	t/t/t	g$^+$/g$^+$/g$^+$	t, g$^-$/t/t	g/g$^-$/g$^-$	260/224/227
C4	193/148/147	g/g$^-$/g$^-$	t/t/t	g$^+$/g$^+$/g$^+$	t, g$^-$/t/t	g/g$^-$/g$^-$	270/255/258
T5	189/54/58	g/g$^-$/g$^-$	t/t/t	g$^+$/g$^+$/g$^+$	t, g$^-$/t, g$^-$/t, g$^-$	g/g$^-$, t/g$^-$, t	290/235/225
T6	182/143/155	g/t, g$^-$, g$^+$/t, g$^-$, g$^+$	t/t, g$^+$/t, g$^+$	g$^+$/g$^+$/g$^+$, t	t, g$^-$/g$^-$/t, g$^-$	g/g/g$^-$, t	280/221/223
C7	193/133/139	g/g$^-$/g$^-$, t, g$^+$	t/t/t	g$^+$/g$^+$/g$^+$, t, g$^-$	t, g$^-$/g$^-$/g$^-$	g/g$^+$/g$^+$	270/215/214
G8	198/154/152	g/g$^+$/g$^+$	—/t/t	—/t/t	t, g$^-$/t/t	g/g$^-$/g$^-$	230/63/61
G9	198/158/145	g/g$^-$/g$^-$	—/t/t	—/g$^+$/g$^+$	t, g$^-$/t, g$^-$/t, g$^-$	g/g$^-$, t/g$^-$, t	270/259/250
T10	187/125/117	g/g$^-$/g$^-$	t/t/t	g$^+$/g$^+$/g$^+$	t, g$^-$/t, g$^-$/t, g$^-$	g/g$^-$, t/g$^-$, t	260/238/236
C11	144/133/121	g/g$^-$/g$^-$	—/t/t	—/g$^+$/g$^+$	t, g$^-$/t/t	g/g$^-$/g$^-$	270/236/231
C12	—/125/127	g/g$^-$/g$^-$	—/t/t	—/g$^+$/g$^+$	NA	NA	anti/224/229

*NMR values from James and Tinoco (1993). All values are in degrees. The NMR values are listed first, followed by the averages from the S1 and S2 trajectories. Dashes indicate values that could not be determined. NA, Not applicable.

paragraph, the G8 residue in both the S1 and S2 simulations is still participating in a hydrogen bond with residue 5. This interaction, which starts out in a bifurcated conformation, moves to a single hydrogen bond interaction between T(U)5:O_2 and G8:N1 during the simulations. Other than this residual interaction, the simulations do not contain any other hydrogen bonds in the loop domain, including any base-phosphate interactions. This behavior is not only consistent with the NMR results, which indicate that there are no hydrogens bonds in the loop domain, but also represents a movement away from the starting structure, where there was a base-phosphate interaction between the amino group of C7 and the phosphate between residues 5 and 6.

In the stem portion of the molecule, the NMR data suggested that the terminal base pair (G1:C12) was somewhat dynamic and that the C11 and C12 sugars underwent frequent repuckering between the C2′-endo and C3′-endo conformations. In the S1 and S2 simulations, we did observe a few very brief periods where the terminal base pair opened, but it quickly returned to a Watson-Crick hydrogen-bonded interaction. The majority of MD snapshots (taken every picosecond) demonstrated a stable G1:C12 base pair, which is reflected in the average value of the "opening" parameter taken after the conformational transition. This value is ~1.4° for the two trajectories. We did observe a significantly greater rate of repuckering in the sugars of the C11 and C12 residues compared to the other stem residues. The NMR data shown in Table 4 indicate that the C12 pucker could not be determined, but has the C11 pucker in the C1′-exo range. The averages calculated from the trajectories put both C11 and C12 in the C1′-exo range, which matches the NMR data. In Table 5, we present the percentages of pseudorotation states sampled during the simulations for residues C11 and C12. Our simulations sampled a broad range of pucker phases between C3′-endo and C2′-endo, with the C1′-exo state sampled most frequently. We did not observe the ~50% N and S for C11 and C12 that was reported in the NMR paper. However, it should be noted that the NMR data for these two sugar puckers is somewhat underdetermined. Only three of five coupling constants—the experimentally measured values used to derive the pucker pseudorotation angle—could be determined for C11, and only one of five could be determined for C12. It is possible that the simulations are providing a more accurate description of the behaviors of these two puckers, where we observe broad and dynamic distributions of pucker angles rather than an equilibrium between C2′ and C3′-endo conformations. However, this could also be a force-field deficiency manifested as an underrepresentation of the C3′-endo phase.

TABLE 5 Distribution of sugar pucker pseudorotation angles

	S1		S2	
	C11	C12	C11	C12
C3′-endo	0.5	0.9	6.4	3.5
C4′-exo	2.6	3.3	6.9	4.9
O4′-endo	13.3	16.8	17.8	12.8
C1′-exo	46.5	59.6	40.0	51.9
C2′-endo	36.3	19.1	28.0	26.4
C3′-exo	0.8	0.3	0.8	0.4
C4′-endo	0.0	0.0	0.0	0.0
O4′-exo	0.0	0.0	0.0	0.0
C1′-endo	0.0	0.0	0.0	0.0
C2′-exo	0.0	0.0	0.1	0.1

All values listed are percentages calculated from the trajectories after the A-to-B transition.

As stated above, although our simulations are consistent with the experimental results for the stem, there are two remaining differences in the loop. Both are in the U:G base pair: the U5 sugar did not repucker to C2′-endo, and the G8 glycosidic torsion did not convert from *syn* to *anti*. We feel that both of these differences are due to insufficient sampling and are not due to deficiencies in the force field. To verify this, we ran a control simulation in which we applied constraints to the system to move it toward the experimental structure. We then removed the constraints and ran dynamics at 300 K to see whether the structure was stable. After 200 ps of unrestained MD, the stem was still B-form and the U5 and G8 residues remained in the correct conformations.

CONCLUSION

We have presented the results of two independent and unrestrained nanosecond-length MD simulations of the deoxyribose analogs of the RNA UUCG tetraloop, in which we have observed an A-to-B transition in the stem portion of the molecule. The two simulations differed only in base sequence, with one containing deoxyuracils and the other deoxythymines—the sequence of the true DNA analog. We have shown that both of the simulations, which started in the native RNA conformation, did converge to a B-form stem as measured by RMSDs and, more conclusively, by an examination of the backbone and glycosidic angles, helicoidal parameters, sugar puckers, and minor groove width. This work extends the recent studies of Cheatham and Kollman to the realm of nonhelical systems and provides further evidence that, using such a theoretical model, we are now in a position to address sequence-dependent structural effects in DNA in nonduplex forms.

Our results are in very good agreement with two recent MD studies of DNA and RNA duplex structures (Cheatham and Kollman, 1996, 1997), as well as with the experimental NMR data available for the DNA analog of this tetraloop (James and Tinoco, 1993). However, although the unrestrained simulations converged to the experimental data for the stem portion of the hairpin, the loop residues required the use of constraints to overcome the barriers between the starting and the experimental conformations. In addition to the results presented here, we have previously demonstrated the enhanced sampling achieved by replacing ribose with deoxyribose in a portion of the molecule (Miller and Kollman, 1996). We feel that this replacement strategy is worthy of consideration during the refinement of NMR structures.

This could have a great impact on nonhelical regions, where the structure is frequently underdetermined by the experimental data.

This work is also important as a probe of the extent of conformational sampling attainable with MD studies of such highly charged systems. One of the last and most important steps in the refinement of experimentally determined structures involves employing MD techniques to locate a structure that satisfies as many experimental restraints as possible. Whereas the restraints are rather static in the case of crystallographic studies, those from NMR are often time-averaged by the experiment. It is difficult, if not impossible, to satisfy all of these restraints by using a single conformation. In these cases, the ability to extensively sample phase space is crucial in obtaining a correct description of the structure. This work directly addresses this question on a very challenging model system, a nonhelical RNA tetraloop. This model is particularly relevant, given that nonhelical nucleic acids play important roles in protein-nucleic acid interactions, and we expect many more of these structures to be solved in the near future. Moreover, nucleic acid structures are not easy to determine by using x-ray techniques (Dickerson et al., 1987, 1994; Lipanov et al., 1993), and the NMR data from these systems are limited to rather short (~5 Å) distances (Schmitz and James, 1995). It is extremely important during the refinement of such structures that the modeling techniques be both reliable and robust. Because it is not always possible to start from a good structure during a refinement process, the ability of a force field to correct a poor initial guess is invaluable.

We thank the Computer Graphics Lab for the use of their facilities (T. Ferrin, director; HIH RR-1081) and the Pittsburgh Supercomputing Center for their resources. We thank Gabriele Varani for providing us with the coordinates of the tetraloop. JLM thanks Carlos Simmerling and Tom Cheatham for helpful discussions.

PAK is grateful for research support from the National Institutes of Health through grant CA-25644. JLM is grateful for research support from the American Foundation for Pharmaceutical Education.

REFERENCES

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

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

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and J. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculation. *J. Comput. Chem.* 4:187–217.

Cheatham, T. E., III, and P. A. Kollman. 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, and P. A. Kollman. 1997. Molecular dynamics simulations can reasonably represent the structural differences in DNA: DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* 119:4805–4825.

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

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

Dickerson, R. E., D. S. Goodsell, M. L. Kopka, and P. E. Pjura. 1987. The effect of crystal packing on oligonucleotide crystal structures. *J. Biomol. Struct. Dyn.* 5:557–579.

Dickerson, R. E., D. S. Goodsell, and S. Neile. 1994. ". . . the tyranny of the lattice. . . ". *Proc. Natl. Acad. Sci. USA.* 91:3579–3583.

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

James, J. K., and J. I. Tinoco. 1993. The solution structure of a d[C(TTCG)G] DNA hairpin and comparison to the unusually stable RNA analogue. *Nucleic Acids. Res.* 21:3287–3293.

Lipanov, A., M. L. Kopka, M. Kaczor-Grzeskowiak, J. Quintana, and R. E. Dickerson. 1993. Structure of the B-DNA decamer CCAACITTGG in two different space groups: conformational flexibility of B-DNA. *Biochemistry.* 32:1373–1389.

Miller, J. L., and P. A. Kollman. 1997. Theoretical studies of an exceptionally stable RNA tetraloop: observation of convergence from an incorrect NMR structure to the correct one using unrestrained molecular dynamics. *J. Mol. Biol.* In press.

Olson, W. K. 1982. How flexible is the furanose ring? 2. An updated potential energy estimate. *J. Am. Chem. Soc.* 104:278–286.

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

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

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

Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer Advanced Texts in Chemistry. C. R. Cantor, editor. Springer Verlag, New York.

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

Schmitz, U., and T. L. James, editors. 1995. How to generate accurate solution structures of double-helical nucleic acid fragments using nuclear magnetic resonance and restrained molecular dynamics. *Methods Enzymol.* 261:3–44.

Schultz, S. C., G. C. Shields, and T. A. Steitz. 1991. Crystal structure of a CAP-DNA complex—the DNA is bent by 90 degrees. *Science.* 253: 1001–1007.

Simmerling, C., R. Elber, and J. Zhang. 1995. MOIL-View—a program for visualization of structure and dynamics of biomolecules and STO—a program for computing stochastic paths. *In* Modelling of Biomolecular Structures and Mechanisms. A. Pullman, J. Jortner, and B. Pullman, editors. Kluwer Academic Publishers, the Netherlands. 241–265.

Ulyanov, N. B., and T. L. James. 1995. Statistical analysis of DNA duplex structural features. *Methods Enzymol.* 261:90–125.

Varani, G., G. Cheong, and I. J. Tinoco. 1991. Structure of an unusually stable RNA hairpin. *Biochemistry.* 30:3280–3289.

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