

# Unrestrained Stochastic Dynamics Simulations of the UUCG Tetraloop Using an Implicit Solvation Model

D. Jeremy Williams and Kathleen B. Hall

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110 USA

**ABSTRACT** Three unrestrained stochastic dynamics simulations have been carried out on the RNA hairpin GGAC[UUCG] GUCC, using the AMBER94 force field (Cornell et al., 1995. *J. Am. Chem. Soc.* 117:5179–5197) in MacroModel 5.5 (Mohamadi et al., 1990. *J. Comp. Chem.* 11:440–467) and either the GB/SA continuum solvation model (Still et al., 1990. *J. Am. Chem. Soc.* 112:6127–6129) or a linear distance-dependent dielectric ( $1/R$ ) treatment. The linear distance-dependent treatment results in severe distortion of the nucleic acid structure, restriction of all hydroxyl dihedrals, and collapse of the counterion atmosphere over the course of a 5-ns simulation. An additional vacuum simulation without counterions shows somewhat improved behavior. In contrast, the two GB/SA simulations (1.149 and 3.060 ns in length) give average structures within 1.2 Å of the initial NMR structure and in excellent agreement with results of an earlier explicit solvent simulation (Miller and Kollman, 1997. *J. Mol. Biol.* 270:436–450). In a 3-ns GB/SA simulation starting with the incorrect UUCG tetraloop structure (Cheong et al., 1990. *Nature*. 346:680–682), this loop conformation converts to the correct loop geometry (Allain and Varani, 1995. *J. Mol. Biol.* 250:333–353), suggesting enhanced sampling relative to the previous explicit solvent simulation. Thermodynamic effects of 2'-deoxyribose substitutions of loop nucleotides were experimentally determined and are found to correlate with the fraction of time the ribose 2'-OH is hydrogen bonded and the distribution of the hydroxyl dihedral is observed in the GB/SA simulations. The GB/SA simulations thus appear to faithfully represent structural features of the RNA without the computational expense of explicit solvent.

## INTRODUCTION

In attempts to provide a comprehensive molecular understanding of nucleic acid systems, theoretical methods, particularly molecular dynamics, are applied to the growing body of thermodynamic, kinetic, and structural information (Beveridge and Ravishanker, 1994; Briki et al., 1991; Nowakowski et al., 1996). Molecular dynamics (MD) provides a microscopic description of the structural and dynamic properties of a molecule, which should facilitate the interpretation of experimental observations and allow the extrapolation to conditions not measured (McCammon and Harvey, 1987). However, unrestrained molecular dynamics studies of nucleic acids and other highly charged macromolecules in which both solvent and solute atoms are treated in molecular detail, have until recently been limited by the treatment of electrostatic interactions, which resulted in severe structural distortion during the simulation (Miskiewicz et al., 1993). In the case of nucleic acids, it has been necessary to employ artificial hydrogen bonding or distance restraints, or to modify the phosphate charges to preserve the structural integrity of the molecule. Recent improvements in the treatment of electrostatic interactions allowing stable nanosecond-length simulations of highly charged systems (Cheatham et al., 1995) have led to new

molecular dynamics studies on a variety of nucleic acid structures (Auffinger and Westhof, 1997; Cheatham and Kollman, 1996; Young et al., 1997; Zichi, 1995). These simulations used an efficient [ $N \log(N)$ ] implementation of the Ewald summation technique (particle mesh Ewald, PME), previously applied mainly to molecular liquids, ionic melts, and small peptides, to calculate electrostatic interactions (Darden et al., 1993; Schreiber and Steinhauser, 1992). They demonstrated a remarkable level of agreement with experiment, in many cases reproducing the effects of different environmental conditions on nucleic acid structure. The dependence of the results on simulation protocol was also examined, and important areas for methodological improvement were delineated.

Tetraloops (RNA hairpins with four loop nucleotides) are one of the popular model systems for nucleic acid molecular dynamics studies. They continue to receive considerable attention because of their small sizes, the availability of high-resolution NMR structures (Cheong et al., 1990; Heus and Pardi, 1991), and their ubiquity (and presumed importance) in cellular RNA (Tuerk et al., 1988; Woese et al., 1990). Notably, explicit solvent simulations have been carried out on the UUCG (Miller and Kollman, 1997a,b) and GCAA (Zichi, 1995) tetraloops, the two most frequently occurring tetraloop motifs, which have yielded average structures consistent with NMR and mutagenesis data for these two systems.

Improvements notwithstanding, explicit solvent molecular dynamics simulations remain computationally expensive, even with the use of particle mesh Ewald (Cheatham et al., 1995) or other efficient methods to treat electrostatic effects. Consequently, alternative simulation methodologies

---

Received for publication 9 November 1998 and in final form 25 March 1999.

Address reprint requests to Dr. Kathleen Hall, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110. Tel.: 314-362-4196; Fax: 314-362-7183; E-mail: hall@bionmr3.wustl.edu.

© 1999 by the Biophysical Society

0006-3495/99/06/3192/14 \$2.00

that reduce computational overhead by neglecting solvent degrees of freedom are of great interest. An implicit (continuum) solvation model that can accurately describe the mean effects of solvent on nucleic acid structure is desirable for applications such as NMR refinement and *ab initio* structure generation. In principle, the lack of molecular solvent allows much more efficient searching of solute degrees of freedom, allowing, for example, torsional Monte Carlo (Guarnieri and Still, 1994) or potential smoothing (Pappu et al., 1999; Piela et al., 1989) to be easily coupled with dynamics-based methods. Furthermore, the reduced computational load makes submicrosecond-length simulations feasible and will be of great utility in the interpretation of fluorescence and NMR relaxation experiments (Palmer and Case, 1992; Soman et al., 1993).

A significant effect of solvents such as water is the ability to screen charges; one common way of then accounting for this mean effect is to modify the dielectric constant in Coulomb's law from its gas phase value. Some of these treatments and their limitations are discussed in a review by Harvey (1989). The sigmoidal dielectric model is one particularly successful formulation (Hingerty et al., 1985) that has been used in the simulation of peptides (Daggett et al., 1991) and nucleic acids in solution (Fritsch et al., 1993; Mazur and Jernigan, 1991; Ramstein and Lavery, 1988). This model has been shown to be superior to the simpler linear distance-dependent dielectric model ( $1/R$ ) and in some cases to explicit solvent simulations with simple cutoffs. Nevertheless, the linear distance-dependent dielectric model continues to be used for many applications, notably in electrostatic refinement from NMR data, using X-PLOR (Brunger, 1992).

More advanced continuum solvation treatments, such as the generalized Born/surface area solvation model (GB/SA) (Still et al., 1990) and analytical continuum solvent potential (ACS) (Schaefer, et. al., 1998) based on the Born solvation model or models based on Poisson-Boltzmann theory (Smart et al., 1997), are increasingly used in implicit solvent simulations. These approaches have the advantage that in addition to capturing the mean structural effects of the solvent on the solute, the free energy of solvation can also be accurately predicted, often as precisely as can be done with explicit solvent simulations (Jean-Charles et al., 1991). The GB/SA model in particular has been used successfully in the simulation of peptides (Guarnieri and Weinstein, 1996) and organic molecules (Smith and Vijayakumar, 1991; Williams and Hall, 1996), although it has only recently been applied to nucleic acid systems in which charge-charge interactions are of critical importance (Jayaram et al., 1998; Srinivasan et al., 1998). These recent studies used the GB/SA model to calculate the solvation energy for various nucleic acid conformers generated from explicit solvent molecular dynamics simulations.

Here we present the results of molecular mechanics studies on the GGAC[UUCG]GUCC hairpin to test the efficacy of the GB/SA model for nucleic acid simulations. The GB/SA model should ideally yield results superior to those

of the linear and sigmoidal dielectric models. This model is particularly attractive because it gives solvation energies with an accuracy comparable to that of Poisson-Boltzmann and explicit solvent methods at a fraction of the computational expense (Srinivasan et al., 1998; Still et al., 1990). Stochastic dynamics is used to approximate the dynamic effect of the solvent on hairpin structure (Guarnieri and Still, 1994; van Gunsteren and Berendsen, 1988), and the performance of the GB/SA model is compared to that of a linear distance-dependent ( $1/R$ ) dielectric treatment, which is often used in refinement of NMR data. The implicit solvation simulations are compared to each other, the NMR structure, and to a previous explicit solvent simulation on this system (Miller and Kollman, 1997b). The behavior of the loop hydroxyls in the simulations is examined and related to the thermodynamic results of 2'-substitution studies of these nucleotides. The three simulations reported here differ only in the initial velocities of the two GB/SA (a and b) simulations and the use of the  $1/R$  dielectric model for the vacuum (v) simulation. Consequently, differences in average properties should mainly reflect differences in solvation treatment.

Solvent molecules are the major contributors to the degrees of freedom in liquid simulations of small biomolecules. Implicit solvation models eliminate these degrees of freedom, obviating the need for short nonbonded cutoffs and thus eliminating a major source of artifacts in molecular dynamics simulations—those caused by truncation of long-range electrostatic interactions. Our study demonstrates that accurate molecular dynamics using the GB/SA model itself is possible, even for highly charged systems of nucleic acids with explicit counterions. It is not necessary to modify the phosphate charges or to omit the counterions to produce a simulation of a quality comparable to that of the most recent explicit solvent simulations using PME treatment of electrostatic effects (Miller and Kollman, 1997b).

## MATERIALS AND METHODS

### Simulations

The P1 Helix NMR structure containing the C[UUCG]G motif (Allain and Varani, 1995) served as the starting point for the stochastic dynamics simulations. The stem was shortened in SYBYL (version 6.2, molecular modeling software, 1995; Tripos Associates), and the sequence was changed to match that used in an earlier structure determined by Tinoco and co-workers (Cheong et al., 1990; Varani et al., 1991). The original UUCG NMR model was obtained from Ignacio Tinoco, Jr. Eleven counterions were placed using MacroModel V5.5, the structures were minimized for 500 steps, and simulations were carried out using the Batchmin V5.5 molecular modeling program (Guarnieri and Still, 1994; Mohamadi et al., 1990). Each simulation was preequilibrated for 250 ps, followed by either a 1.1- or 3.2-ns production run with GB/SA or a 5-ns vacuum simulation. In addition, a 2-ns vacuum simulation without counterions and 3 ns ( $\gamma$  of  $5 \text{ ps}^{-1}$ ) and 1.75 ns ( $\gamma$  of  $2.5 \text{ ps}^{-1}$ ) unequilibrated GB/SA simulations (original UUCG model; Cheong et al., 1990) were also carried out. Except where noted, all simulations were carried out at a nominal temperature of 300 K, with a 1-fs time step, a 0.1-ps time constant (frictional coefficient  $\gamma$  of  $5 \text{ ps}^{-1}$ ), and SHAKE (Ryckaert et al., 1977) applied to all bonds involving hydrogen atoms (X-H). Nonbonded cutoff

distances were set at 1000 Å (infinite cutoff), and the pair list was split into short-range and long-range components at 10 Å with the long-range nonbonded derivatives updated every 10 time steps. The AMBER94 force field (Cornell et al., 1995) was used in all calculations with either the GB/SA continuum model (Still et al., 1990) or a simple distance-dependent dielectric (1/*R*) implicit solvation treatment. All bonded and nonbonded terms are identical to the original force field of Cornell et al. For the GB/SA model, solvent polarization and surface area derivatives were updated every 10 and 50 time steps, respectively. Analytical approximations for the Born radius and surface area implemented in Batchmin allow calculation of all relevant first and second derivatives for the solvent polarization and surface area contributions (Qiu et al., 1997). The average and instantaneous temperature, kinetic energy, and potential energy were monitored over the course of the simulation, and molecular snapshots were saved every picosecond. All further simulation properties were derived from analyses of these snapshots. Pairwise mass-weighted root mean square deviations of structures were calculated using the SUPERPOSE routine in TINKER (Dudek and Ponder, 1995). Cross-rmsd refers to the mass-weighted root mean square deviation between unminimized average structures from different simulations. The rmsd value from the NMR structure was calculated by averaging the mass-weighted root mean square deviation of each snapshot from the initial NMR structure. The self rmsd was calculated in a similar fashion, except that in place of the NMR structure the average structure of the given simulation was used as a reference. Rmsd comparisons for the GB/SA simulation of the original (incorrect) structure are computed for RNA heavy atoms (hydrogens and counterions not included) and are reported for residues A<sub>3</sub> through U<sub>10</sub> only (G<sub>1</sub>G<sub>2</sub>A<sub>3</sub>C<sub>4</sub>[U<sub>5</sub>U<sub>6</sub>C<sub>7</sub>G<sub>8</sub>][G<sub>9</sub>U<sub>10</sub>C<sub>11</sub>C<sub>12</sub>). Simulations were run on a 12 processor R10000 SGI at the Institute for Biomedical Computing and a R4000 Challenge SGI at the Center for Molecular Design at Washington University.

## RNA synthesis, purification, and characterization

Oligonucleotides were synthesized chemically, using phosphoramidites from Glen Research (Sterling, VA) and purified as previously described (Hall, 1994), except that a Dionex anion exchange column at 65°C was used for high-performance liquid chromatography (HPLC) purification. RNAs were desalted on a Sephadex G-10 column before and after HPLC purification. Extinction coefficients were calculated using a nearest-neighbor model (Freier et al., 1983), assuming that the 2'-substitutions did not affect the absorbance. Thermodynamic studies were carried out in a low-salt buffer (10 mM sodium cacodylate, 0.5 mM Na<sub>3</sub>EDTA, pH 7) to suppress dimerization. UV/Vis thermal denaturation (melting) studies were carried out on a Gilford 260 spectrophotometer equipped with a 2527 Gilford thermoprogrammer. Absorbance measurements were taken at 260 nm with a heating rate of 1°C/min. Before experiments samples were heated to 95°C and allowed to cool slowly to room temperature. Concentrations were determined from upper baselines at 95°C. Melts at different concentrations showed no detectable variation in the midpoint of the thermal transition. Curves from melts were fit to a single intramolecular transition with sloping baselines (Freier et al., 1983), using in-house software (Conlin). Melting temperatures, equilibrium constants, and free energies were calculated from the corresponding van't Hoff enthalpies and entropies determined from global nonlinear least-squares regression of three experiments using the relation

$$K = \frac{\alpha}{1 - \alpha} = \exp(-\Delta H^\circ/RT + \Delta S^\circ/R)$$

where  $\alpha$  is the fraction of strands in the hairpin state,  $R$  is the gas constant, and  $T$  is the temperature in Kelvins. Monte Carlo simulations using 1000 synthetic data sets were also carried out in Conlin to determine parameter confidence intervals (Press et al., 1992).

## RESULTS AND DISCUSSION

### Simulation summary

In this study stochastic dynamics is used to simulate the weakly coupled, random, and dissipative effects of solvent molecules on the GGAC[UUCG]GUCC RNA hairpin. A constant-temperature ensemble is simulated in which there are significant short-term fluctuations in both the kinetic and potential energies of the RNA hairpin. By the end of the 250-ps equilibration period, the trajectories have all converged to stable average thermodynamic values, with an average temperature of 299 K in each case. Despite differences in the absolute average energy, the GB/SA and 1/*R* simulations give almost identical standard deviations for the total energy (data not shown).

The simulations do not show significant differences in average thermodynamic properties or in the fluctuation of these properties; the GB/SA and vacuum simulations have comparable internal energies (excluding solvation contributions) and average temperatures and show similar standard deviations for these properties. However, the vacuum and GB/SA simulations have considerably different effects on the hairpin structure. The 5 ns vacuum simulation (linear distance-dependent dielectric treatment) shows a much greater root mean square deviation from the initial NMR structure (3.3 Å) compared to the 1.1- and 3.1-ns GB/SA simulations (1.2 Å) (counterions not included; summarized in Table 1 and Fig. 1 *A*). It is gratifying to find that the rmsds of the two GB/SA simulations from the initial NMR structure are very similar to those reported for explicit solvent simulations on this system (Miller and Kollman, 1997b). The two GB/SA simulations give virtually identical average structures (0.4 Å rmsd), despite much greater pairwise root mean square deviation seen for individual snapshots. These results are consistent with the findings of previous studies on nucleic acids and proteins that the use of multiple trajectories can significantly enhance local phase

**TABLE 1** Root mean square deviation (Å)

Cross*	GB/SA(a) and GB/SA(b)	GB/SA(a) and vacuum	GB/SA(b) and vacuum
(−) Counterions	0.42	3.07	3.11
(+) Counterions	12.72	9.14	10.52
NMR <sup>#</sup>	GB/SA(a)	GB/SA(b)	Vacuum
(−) Counterions	1.16	1.22	3.32
Self <sup>§</sup>	GB/SA(a)	GB/SA(b)	Vacuum
(−) Counterions	0.82	0.93	0.79

\*Mass-weighted rmsd of average unminimized structures from GB/SA(a) (1.149 ns), GB/SA(b) (3.060 ns), and vacuum simulations (5.011 ns).

<sup>#</sup>Mean pairwise rmsd of unminimized simulation snapshots from NMR structure.

<sup>§</sup>Mean pairwise rmsd of unminimized simulation snapshots from average simulation structure.



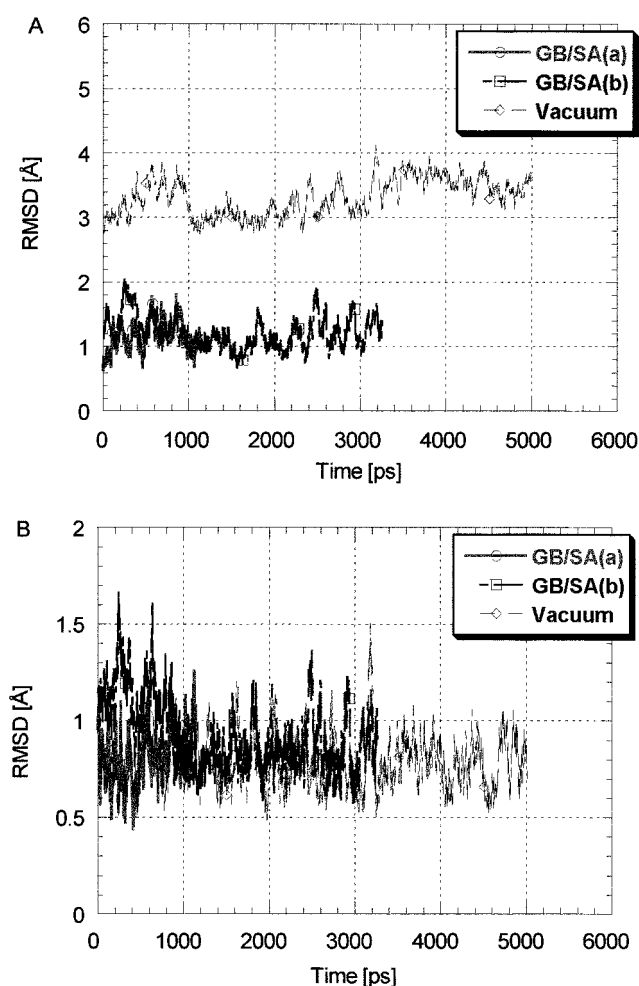


FIGURE 1 All-atom mass-weighted root mean square deviation (not including  $K^+$  counterions) from initial NMR structure (A) or from average simulation structure (B) as a function of time for GB/SA(a) (1.149 ns), GB/SA(b) (3.062 ns) and vacuum simulations (5.011 ns).

space sampling during molecular dynamics simulations (Auffinger et al., 1995; Caves et al., 1998). In contrast, the average vacuum simulation structure has a rmsd of over 3 Å from either GB/SA average.

The vacuum and GB/SA simulations show fluctuations about their mean structures of similar magnitudes (Table 1 and Fig. 1 B), although the detailed dynamic behavior is different. Charged groups (counterions) and flexible hydrogen bonding groups (hydroxyls) in particular are much more restricted in the vacuum simulation in comparison to the GB/SA simulations. This is likely a consequence of the exaggeration of electrostatic interactions at short distances by the linear dielectric model.

### Structural details

A summary of the dihedral angle distributions for loop ( $U_5U_6C_7G_8$ ) and closing base pair ( $C_4G_9$ ) nucleotides is given in Table 2. Analysis of the NMR structure and results of the explicit solvent simulation are taken from the work of

Miller and Kollman and are shown for comparison (Miller and Kollman, 1997b). The GB/SA simulations are virtually indistinguishable from the explicit solvent simulation, giving similar averages and standard deviations for most backbone dihedrals. The only clear case of disagreement between the GB/SA and explicit solvent simulation involves the  $\alpha$  and  $\beta$  torsions of residue  $G_9$ . At this position, the NMR structure has two distinct families of dihedrals; differences in starting structures may be responsible for the disagreement, although both simulations are consistent with the NMR data.

A few dihedrals in one or both GB/SA simulations show larger standard deviations as a consequence of either single large dihedral transitions, probably due to equilibration of the starting structure, or multiple concerted motions of two dihedrals (crankshaft transitions) (Fig. 2). The GB/SA simulations show more dihedral transitions than the explicit solvent simulation, and these are almost entirely confined to the dihedrals around the  $G_8pG_9$  phosphate at the loop/stem junction. Analogous differences between continuum and explicit solvent systems were observed previously in peptide simulations, where it was suggested that phase space sampling was enhanced for the implicit solvent simulation (Daggett et al., 1991). However, the reasons for this enhancement are not clear; it may be due both to differences between the stochastic representation of the solvent and the kinetic properties of bulk water and to solvation-related differences in the effective barrier heights. Previous studies by McCammon and co-workers have shown that a Poisson-Boltzmann continuum solvation model can give potentials of mean force around  $\phi$  and  $\psi$  torsions in alanine dipeptide in remarkable agreement with explicit solvent simulations (Marrone et al., 1996). Interestingly, the agreement between the two models is better for the minima than for the barrier heights. Sampling differences between explicit and continuum solvent simulations may also be related to the choice of frictional coefficient, which for this and many other studies is smaller than that predicted by Stokes' law for small solutes in water (Smart et al., 1997).

The GB/SA simulations and those by Miller and Kollman do a significantly better job of maintaining the NMR average values of backbone dihedral angles and sugar puckers vis-à-vis the vacuum simulation (Miller and Kollman, 1997b). The riboses of the two central loop nucleotides ( $U_6$  and  $C_7$ ) remain C2'-endo over the course of the simulation, with some variability (also seen in the Miller and Kollman simulation) in the  $U_6$  pucker in the shorter simulation (GB/SA(a)) (Fig. 3). All other sugars are C3'-endo throughout the simulation, except for the terminal 3'-residue  $C_{12}$ , which shows several repuckering events that were also seen in the explicit solvent simulation (Fig. 3). (N.B.: The spikes in Fig. 3 are artifacts of the way the data are represented.) The  $U_6$   $\chi$  angle has a large standard deviation in both simulations, as expected for an unconstrained base exposed to solution. In contrast, several dihedrals have the wrong sign in the vacuum simulation, and some sugars have distorted

**TABLE 2** Backbone dihedral angles for the C(UUCG)G residues from simulation and experiment: average plus standard deviation

	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\chi$	P
C <sub>4</sub> <sup>*</sup>	-75 ± 10	172 ± 9	65 ± 9	79 ± 6	-152 ± 10	-60 ± 9	-158 ± 9	20 ± 12
C <sub>4</sub> <sup>#</sup>	-74 ± 10	172 ± 9	66 ± 8	79 ± 6	-152 ± 10	-60 ± 9	-158 ± 9	20 ± 11
C <sub>4</sub> <sup>  </sup>	-73 ± 10	172 ± 10	62 ± 9	77 ± 7	-155 ± 10	-63 ± 9	-159 ± 8	15 ± 9
C <sub>4</sub> <sup>**</sup>	-71 ± 4	163 ± 3	63 ± 5	92 ± 1	-148 ± 9	-67 ± 10	-140 ± 3	20 ± 4
C <sub>4</sub> <sup>##</sup>	-63 ± 10	133 ± 16	74 ± 10	69 ± 5	124 ± 14	51 ± 11	-165 ± 8	6 ± 8
U <sub>5</sub> <sup>*</sup>	-76 ± 10	175 ± 9	63 ± 8	75 ± 7	-173 ± 9	-94 ± 10	-150 ± 11	22 ± 14
U <sub>5</sub> <sup>#</sup>	-75 ± 10	175 ± 8	64 ± 9	79 ± 12 <sup>†</sup>	-165 ± 32 <sup>†</sup>	-100 ± 24 <sup>†</sup>	-149 ± 11	28 ± 28
U <sub>5</sub> <sup>  </sup>	-73 ± 9	174 ± 8	64 ± 8	77 ± 6	-177 ± 9	-91 ± 9	-152 ± 10	19 ± 12
U <sub>5</sub> <sup>**</sup>	-80 ± 11	-179 ± 15	49 ± 3	92 ± 1	-165 ± 3	-92 ± 4	-140 ± 3	19 ± 3
U <sub>5</sub> <sup>##</sup>	-136 ± 10	-174 ± 8	39 ± 9	64 ± 6	177 ± 12	-98 ± 8	-145 ± 10	50 ± 9
U <sub>6</sub> <sup>*</sup>	-167 ± 14	157 ± 15	59 ± 10	142 ± 9	-89 ± 11	-69 ± 10	-120 ± 32	159 ± 21
U <sub>6</sub> <sup>#</sup>	-167 ± 41 <sup>†</sup>	162 ± 17	60 ± 11	141 ± 9	-95 ± 13	-68 ± 11	-118 ± 64	157 ± 18
U <sub>6</sub> <sup>  </sup>	-156 ± 17	158 ± 17	60 ± 10	135 ± 13	-92 ± 12	-71 ± 11	-32 ± 60	150 ± 23
U <sub>6</sub> <sup>**</sup>	-142 ± 5	112 ± 3	63 ± 5	133 ± 1	-101 ± 6	-58 ± 4	-149 ± 6	143 ± 2
U <sub>6</sub> <sup>##</sup>	-120 ± 17	108 ± 22	41 ± 9	134 ± 11	-82 ± 8	-61 ± 8	-145 ± 11	146 ± 18
C <sub>7</sub> <sup>*</sup>	-64 ± 10	176 ± 11	61 ± 9	140 ± 9	-91 ± 10	73 ± 10	-135 ± 14	153 ± 16
C <sub>7</sub> <sup>#</sup>	-63 ± 10	173 ± 13	63 ± 9	140 ± 9	-90 ± 9	74 ± 9	-133 ± 13	152 ± 15
C <sub>7</sub> <sup>  </sup>	-63 ± 10	177 ± 13	60 ± 9	141 ± 11	-91 ± 10	70 ± 9	-134 ± 17	159 ± 18
C <sub>7</sub> <sup>**</sup>	-55 ± 5	174 ± 4	45 ± 2	136 ± 3	-114 ± 8	105 ± 5	-126 ± 1	149 ± 6
C <sub>7</sub> <sup>##</sup>	-67 ± 9	175 ± 10	61 ± 8	94 ± 11	51 ± 8	-83 ± 7	-130 ± 9	90 ± 25
G <sub>8</sub> <sup>*</sup>	81 ± 42 <sup>‡</sup>	-177 ± 10	176 ± 42 <sup>‡</sup>	80 ± 8	-156 ± 21 <sup>§</sup>	-58 ± 32 <sup>§</sup>	39 ± 18	33 ± 14
G <sub>8</sub> <sup>#</sup>	79 ± 47 <sup>‡</sup>	-177 ± 9	176 ± 42 <sup>‡</sup>	80 ± 7	-153 ± 23 <sup>§</sup>	-62 ± 31 <sup>§</sup>	39 ± 13	33 ± 15
G <sub>8</sub> <sup>  </sup>	74 ± 10	-175 ± 8	-175 ± 8	79 ± 7	-176 ± 8	57 ± 10	34 ± 12	47 ± 17
G <sub>8</sub> <sup>**</sup>	34 ± 6	-145 ± 3	-171 ± 4	91 ± 3	-147 ± 11	-41 ± 46	61 ± 14	50 ± 10
G <sub>8</sub> <sup>##</sup>	-58 ± 7	136 ± 4	-24 ± 11					
	69 ± 11	-149 ± 11	-71 ± 9	69 ± 6	-175 ± 8	-82 ± 8	48 ± 8	34 ± 9
G <sub>9</sub> <sup>*</sup>	134 ± 22 <sup>§</sup>	-126 ± 24 <sup>§</sup>	-175 ± 10	81 ± 8	-152 ± 11	-64 ± 9	-171 ± 8	12 ± 16
G <sub>9</sub> <sup>#</sup>	133 ± 21 <sup>§</sup>	-122 ± 27 <sup>§</sup>	-172 ± 11	80 ± 8	-153 ± 10	-64 ± 8	-172 ± 8	13 ± 16
G <sub>9</sub> <sup>  </sup>	-161 ± 9	77 ± 9	176 ± 9	79 ± 8	-154 ± 9	-93 ± 9	-168 ± 9	15 ± 14
G <sub>9</sub> <sup>**</sup>	-118 ± 25	92 ± 19	-168 ± 25	96 ± 2	-151 ± 2	-68 ± 3	-165 ± 2	-5 ± 3
G <sub>9</sub> <sup>##</sup>	112 ± 9	-124 ± 22	99 ± 0					
	69 ± 11	-149 ± 11	-71 ± 9	69 ± 6	-175 ± 8	-82 ± 8	-148 ± 8	61 ± 7

\*1.149-ns GB/SA simulation.

#3.062-ns GB/SA simulation.

§Crankshaft motions (i.e., concerted flip of more than one dihedral angle).

†Equilibration motion (i.e., crankshaft-like motion that only occurs once during simulation).

||Explicit solvent simulation by Miller and Kollman (1997).

\*\*NMR structure by Allain and Varani (1995).

##Vacuum simulation using 1/R distance-dependent dielectric.

puckers. Furthermore, the U<sub>6</sub>  $\chi$  angle is considerably restricted with respect to either the GB/SA or explicit solvent simulation, and no repuckering is seen for the terminal 3'-residue (C<sub>12</sub>).

Hydrogen bonds in base pairs in the stem are maintained over the course of the GB/SA trajectories, with somewhat greater fraying (lower hydrogen bond frequencies) observed in the major groove of the helix (Table 3). The unusual bifurcated hydrogen bond of the U<sub>5</sub>G<sub>8</sub> base pair is also preserved, as is the C<sub>7</sub>-phosphate interaction. The hydrogen bonding interaction involving the U<sub>5</sub> ribose 2'-hydroxyl group is particularly noteworthy. The proton resonance assigned to this hydroxyl is observed in the NMR spectrum in water, presumably because of its involvement in a stable

hydrogen bond (Allain and Varani, 1995; Tinoco, 1996). After refinement against the NOE data, this hydroxyl appeared to hydrogen bond to the O6 position of the G<sub>8</sub> base. However, the explicit solvent simulation of Miller and Kollman shows this group to be involved in a stable hydrogen bond with a backbone oxygen (O5' of residue U<sub>6</sub>). In our GB/SA simulations, this ribose hydroxyl is observed to jump back and forth between the G<sub>8</sub> O6 and U<sub>6</sub> O5' positions via rotations about the O<sub>2</sub>'-C2' dihedral (Fig. 4).

The vacuum experiment shows significantly less fluctuation of hydrogen bonding groups relative to the GB/SA simulations. Both stem and loop hydrogen bonds are rigidly maintained despite gross deformations in the overall structure, and hydroxyls remain hydrogen bonded to a single

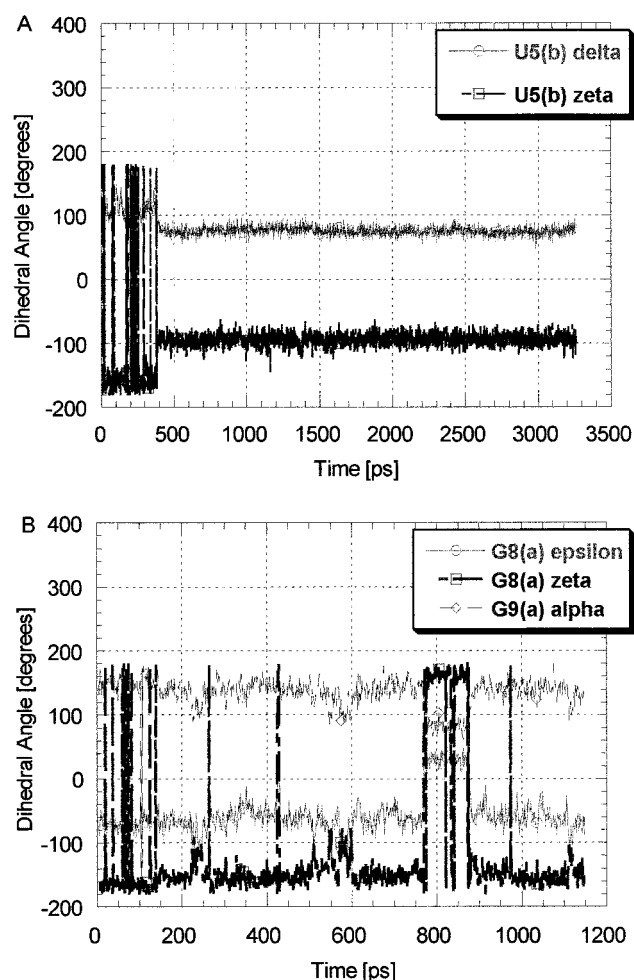


FIGURE 2 Examples of backbone dihedral angles that undergo single (A) and multiple (B) crankshaft motions.

acceptor (the O3' backbone oxygen in the case of stem nucleotides) throughout the simulation. The counterion atmosphere is severely collapsed in the vacuum simulation, where it is almost entirely restricted to the neighborhood of the phosphates, in contrast to the more diffused atmosphere seen for the GB/SA simulations. To test whether this artifactual situation was responsible for the poor average structures, we carried out an additional 2-ns vacuum simulation under identical condition but without counterions. This simulation shows improved behavior with the average rmsd of the snapshots from the NMR structure reduced to 1.8 Å after 1 ns, and the loop geometry is better maintained. However, hydroxyl groups still show severe restriction relative to the GB/SA model, sugar puckers are farther from their NMR-determined geometries (G2 and C11 have incorrect C2'-endo puckers), and hydrogen bonds are rigidly maintained. As in the case of the vacuum simulation with counterions, these kinetic effects are not apparent from the global value of the rms fluctuation about the average structure (1.2 Å). Moreover, the value of the average rmsd from the starting (NMR) structure is somewhat deceptive, as the rmsds for

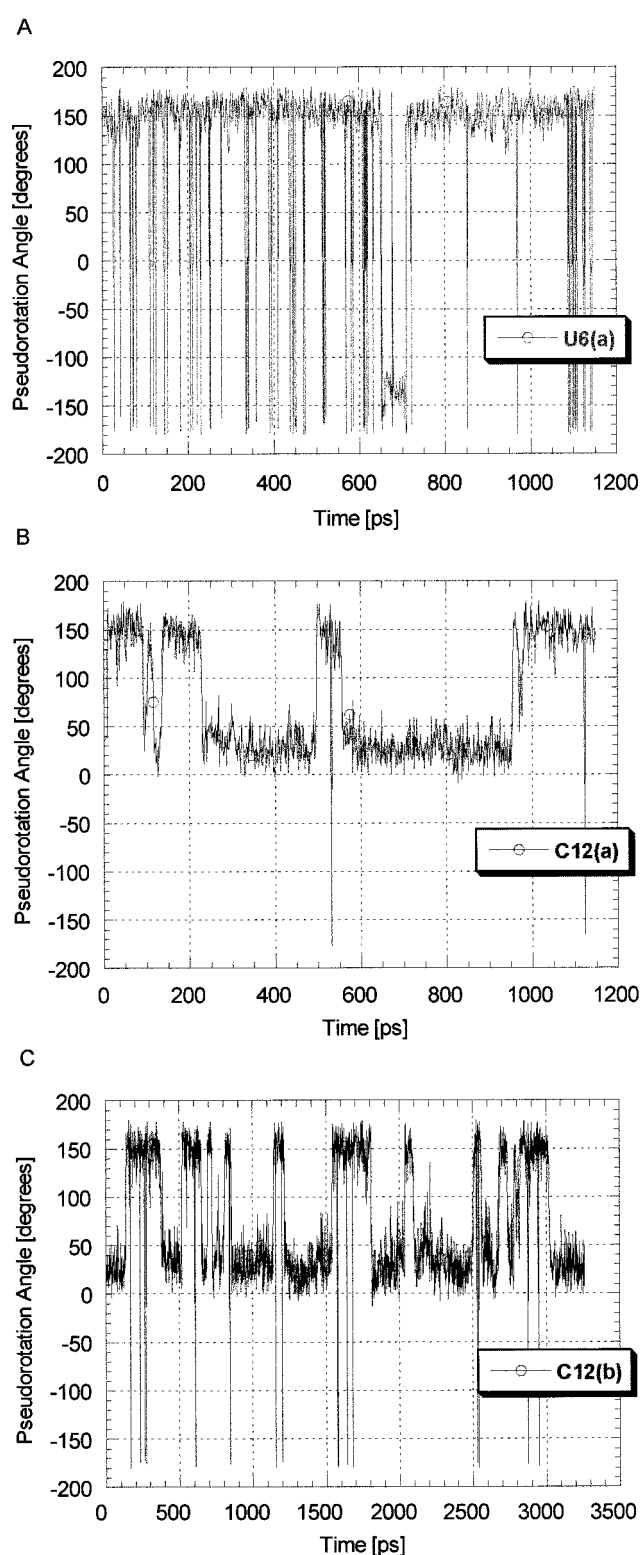


FIGURE 3 Residues that show repuckering of their riboses during the GB/SA (A) (1.149 ns) and GB/SA (B) (3.062 ns) simulations.

structures at the end of the simulation are closer to 2.0 Å from the NMR structure, and it is not clear that the simulation has converged (data not shown).

**TABLE 3** Intramolecular hydrogen bond geometry\*

Hydrogen Bond	$\langle \text{Freq.} \rangle^{\#}$	$\langle R_{\text{DA}} \rangle$ (Å)	$\langle \delta R_{\text{DA}}^2 \rangle^{1/2}$ (Å)	$\langle \theta_{\text{DHA}} \rangle$ (deg)	$\langle \delta \theta_{\text{DHA}}^2 \rangle^{1/2}$ (deg)
G <sub>1</sub> :N <sub>2</sub> –H···O <sub>2</sub> :C <sub>12</sub>	0.958	3.022	0.228	161.109	10.679
	0.982	2.970	0.183	161.364	10.370
G <sub>1</sub> :N <sub>1</sub> –H···N <sub>3</sub> :C <sub>12</sub>	0.960	3.095	0.192	163.359	9.149
	0.953	3.097	0.251	162.553	8.958
G <sub>1</sub> :O <sub>6</sub> –H···N <sub>4</sub> :C <sub>12</sub>	0.883	3.131	0.350	156.725	16.314
	0.874	3.178	0.478	158.777	13.442
G <sub>2</sub> :N <sub>2</sub> –H···O <sub>2</sub> :C <sub>11</sub>	0.997	2.897	0.143	162.174	9.405
	0.990	2.917	0.157	162.531	9.590
G <sub>2</sub> :N <sub>1</sub> –H···N <sub>3</sub> :C <sub>11</sub>	0.936	3.113	0.232	164.041	8.547
	0.932	3.116	0.361	164.237	8.516
G <sub>2</sub> :O <sub>6</sub> –H···N <sub>4</sub> :C <sub>11</sub>	0.805	3.281	0.450	160.345	11.906
	0.857	3.285	0.713	159.991	12.658
A <sub>3</sub> :N <sub>1</sub> –H···N <sub>3</sub> :U <sub>10</sub>	0.981	3.020	0.160	161.511	10.500
	0.974	3.047	0.185	162.481	9.692
A <sub>3</sub> :N <sub>6</sub> –H···O <sub>4</sub> :U <sub>10</sub>	0.724	3.338	0.362	160.734	11.121
	0.753	3.317	0.397	160.820	10.856
C <sub>4</sub> :O <sub>2</sub> –H···N <sub>2</sub> :G <sub>9</sub>	0.995	2.907	0.138	163.542	8.610
	0.997	2.906	0.143	163.106	8.726
C <sub>4</sub> :N <sub>3</sub> –H···N <sub>1</sub> :G <sub>9</sub>	0.997	3.008	0.122	164.178	7.987
	0.990	3.018	0.138	163.704	8.459
C <sub>4</sub> :N <sub>4</sub> –H···O <sub>6</sub> :G <sub>9</sub>	0.950	3.048	0.220	160.923	11.139
	0.941	3.068	0.261	160.358	11.801
U <sub>5</sub> :O <sub>2</sub> –H···N <sub>2</sub> :G <sub>8</sub>	0.447	3.532	0.492	133.913	15.841
	0.530	3.455	0.462	136.459	15.935
U <sub>5</sub> :O <sub>2</sub> –H···N <sub>1</sub> :G <sub>8</sub>	0.808	3.193	0.563	150.352	17.845
	0.883	3.059	0.382	153.311	15.146
U <sub>5</sub> :O <sub>2</sub> –H···O <sub>6</sub> :G <sub>8</sub>	0.574	3.252	0.702	116.200	55.655
	0.563	3.153	0.526	116.751	52.615
U <sub>5</sub> :O <sub>2</sub> –H···O <sub>5</sub> :U <sub>6</sub>	0.360	3.227	0.260	83.680	55.714
	0.274	3.392	0.461	84.193	55.609
C <sub>7</sub> :N <sub>4</sub> –H···O <sub>2p</sub> :U <sub>6</sub>	0.433	3.695	0.666	141.266	16.006
	0.485	3.840	1.123	138.863	18.683

\*1.149-ns GB/SA simulation (clear background); 3.060-ns GB/SA simulation (shaded background).

<sup>#</sup>Normalized hydrogen bond frequencies (i.e., donor-acceptor distance less than 3.5 Å and donor-hydrogen-acceptor angle greater than 120°).

## Ribose 2'-hydroxyl groups

One may anticipate that the greatest differences between the explicit and implicit solvation treatments for nucleic acid simulations will involve the properties of flexible moieties that are available for hydrogen bonding interactions, including the relatively unhindered 2'-hydroxyl groups on the ribose sugar and NH<sub>2</sub> groups on unpaired bases. Unfortunately, the rotameric states of these groups are severely underdetermined in structural studies, and much of what is known has been inferred from structural, theoretical, and mutagenic analyses. In the case of the UUCG tetraloop, several lines of evidence strongly suggest that the ribose hydroxyls of loop nucleotides are important to the unusual stability of the motif. In particular, several sharp <sup>1</sup>H resonances in the 6.5–6.7-ppm range have been attributed to slowly exchanging hydroxyls, suggesting their involvement in stable hydrogen bonding interactions (Varani et al., 1991). Experimentally, replacement of the four loop residues with deoxynucleotides resulted in a significant loss of stability—a seven degree decrease in the *T*<sub>m</sub> (Sakata et al., 1990).

Preliminary results from more detailed 2'-substitution studies on this tetraloop, summarized in Table 4, suggest a

very strong position dependence of the effects of 2'-substitutions. Deoxyribose substitutions are very destabilizing (by ~1 kcal/mol at 37°C) at the U<sub>5</sub> position, are less so at the C<sub>7</sub> and G<sub>8</sub> positions, and have no effect when made at U<sub>6</sub>. A 2'-deoxy substitution in the middle of the A-form helix has an effect similar in magnitude to that of the C<sub>7</sub> and G<sub>8</sub> substitutions, although the interactions responsible for the stability are not necessarily the same.

If one assumes that positional differences in the effects of 2'-hydroxyl mutations can be largely attributed to differences in the native state, an attempt can be made to correlate dynamic and structural differences observed for these groups in the hairpin simulations with the thermodynamic results. It should not be inferred that there are no thermodynamic consequences of the substitution on the single-stranded state. Instead, we make the weaker assumption that individual hydroxyls will experience similar mean environments in the single stranded state because of the loss of secondary and tertiary constraints and consequently that the effects of deleting these groups will be relatively sequence independent.

If we examine the 2'-hydroxyl dihedral (H-O-C2'-C3') distributions summarized in Table 5 and Fig. 5 with these



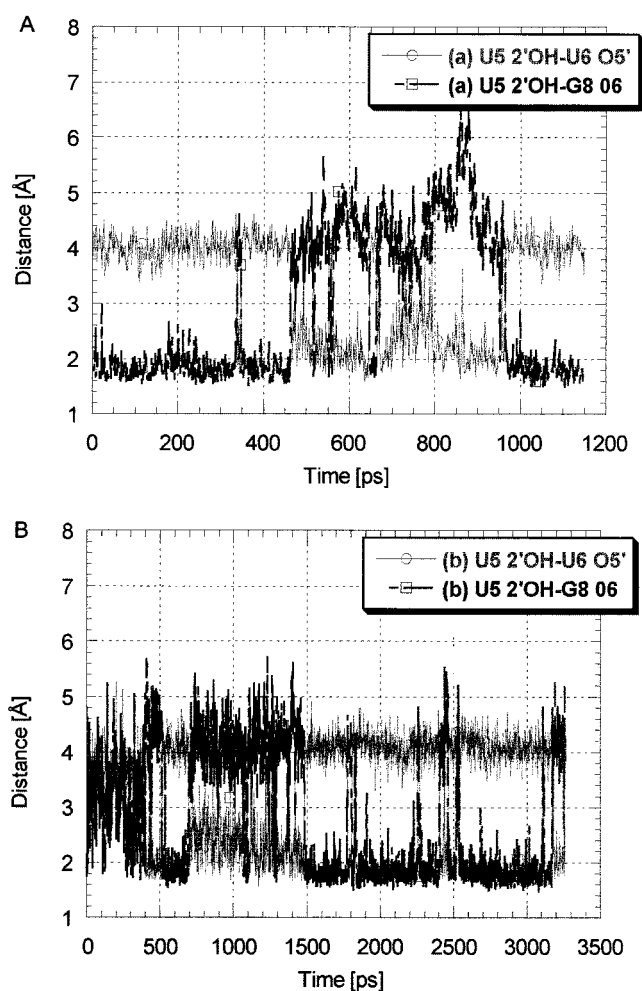


FIGURE 4 Distance from  $U_5$  2'-OH proton to  $U_6$  O5' and  $G_8$  O6 hydrogen bond acceptors for the GB/SA simulations. (A) 1.149 ns. (B) 3.062 ns.

caveats in mind, we see a possible correlation with the mutagenesis and simulation data. With a single exception, all hydroxyl groups show well-defined dihedral distributions, and deoxy substitutions at these positions are destabilizing. The  $U_6$  hydroxyl is anomalous and displays a broad dihedral distribution as a result of frequent transitions between several positions and is less constrained than even the stem hydroxyls. Strikingly, it is the only position in the molecule where the 2'-deoxy substitution has no net thermodynamic effect.

The loop hydroxyls show potentially interesting sequence dependence in their behavior. The  $U_5$  hydroxyl is observed to alternate every few hundred picoseconds between the  $G_8$  O6 position seen in the NMR structure and the  $U_6$  O5' position found in the explicit solvent simulation. From the standard deviations of the fluctuations and the frequency of the transitions one can infer that the  $U_5$  hydroxyl jumps between two narrow energy wells corresponding to two accessible hydrogen bonding positions. In contrast, the  $C_7$  and  $G_8$  hydroxyls have single, fairly narrow conformational minima available to them, but unlike the  $U_5$  2'-OH group,

these two hydroxyls are involved in poor hydrogen bonding interactions with backbone oxygens, as determined from the O-H...O distances and angles. Except for the  $U_6$  residue, all hydroxyls appear to have significant intramolecular interactions in the native state, as is seen from their restricted orientations and loss of stability upon substitution with hydrogen.

Our discussion of the energetic contributions of the hydroxyl groups is necessarily qualitative because of the neglect of single strand effects. Moreover, the complex flexible nature of the unfolded state makes simplified harmonic treatments problematic. Consequently, we have initiated a more quantitative free energy perturbation analysis of this system using a Monte Carlo/Stochastic Dynamics protocol (Guarnieri and Still, 1994) to increase sampling (data not shown). Other 2'-substituents (fluorine, amine) will also be examined as part of this study.

From the point of view of previous explicit solvent simulations, the dihedral distributions of the ribose hydroxyls in the GB/SA simulations seem more reasonable than those from the vacuum simulations (Auffinger and Westhof, 1997). In the GB/SA treatment, all hydroxyls show significant dihedral fluctuations, and as expected, the  $C_{12}$  position shows enhanced flexibility relative to other stem hydroxyls (Table 5). In contrast, 2'-OH groups in the vacuum simulation, including the 3'-terminal hydroxyl, are completely restricted, with all stem hydroxyls monopolized in hydrogen bonding interactions with the O3' position, whether or not counterions are included.

The preferred hydrogen bond acceptor for stem hydroxyls in the GB/SA simulations is the O3'(n), in agreement with previous explicit solvent simulations (Auffinger and Westhof, 1997; Singh and Kollman, 1996). In addition, the stem residues show broad bimodal loop hydroxyl distributions due to rapid transitions between hydrogen bonding with the O3'(n) groups and with the O4'(n + 1) position. In contrast, Westhof and co-workers find that the 2'-OH groups preferentially hydrogen bond with base and O4'(n) positions over the O4'(n + 1) positions (Auffinger and Westhof, 1997). Although it has been inferred from a previous NMR study by Lane and co-workers that ribose hydroxyls hydrogen bond to internal groups, their NMR study does not allow discrimination among the several possible hydrogen bond acceptors (Conte et al., 1996). The 2'-OH hydrogen bonding patterns of the GB/SA and explicit solvent simulations are both consistent with the results of this NMR study of RNA hydration.

Other simple 2'-substituents may be analyzed within the framework of the GB/SA simulations, although additional complex effects such as steric or hydrophobic contributions in the case of 2'-OMe substitutions may modulate the energetics (Table 4). For example, the O-methyl substitution is significantly more destabilizing at the  $U_5$  position than the deoxy substitution (Table 4), which we attribute to the difficulty of inserting the methyl group into the tight turn structure of this loop. This interpretation is supported by NMR imino proton data indicating that the  $U_5$  O-methyl



**TABLE 4** Thermodynamics of 2'-OH substitutions in the UUCG tetraloop

RNA sequence	$\Delta H^\circ$ (kcal/mol)	$\Delta S^\circ$ (kcal/mol)	$T_M$ (°C)	$\Delta G^\circ(55)$ (kcal/mol)	$\Delta G^\circ(37)$ (kcal/mol)
CGCUUCGGCG*	40.6	118.8	68.7	1.6	3.8
	37.1	108.8	68.1	1.4	3.4
CGCdUUCGGCG*	39.4	118.0	60.6	0.7	2.8
CGCUdUCGGCG*	40.1	117.0	69.1	1.6	3.8
	38.8	112.9	70.5	1.7	3.8
CGCUUdCGGGCG*	35.9	106.6	63.7	0.9	2.8
	38.3	113.7	63.8	1.0	3.0
CGCUUCdGGCG*	35.7	105.7	64.1	1.0	2.9
	33.3	98.5	64.9	1.0	2.7
CdGCUUCGGCG*	32.9 <sup>§</sup>	97.8	63.4	0.8	2.6
	32.3 <sup>§</sup>	95.9	63.7	0.8	2.6
-ACUUCGGU- <sup>#</sup>	48.5	140.4	72.0	2.4	4.9
	40.3 <sup>§,¶</sup>	116.9	71.8	2.0	4.1
	50.9	147.3	72.3	2.6	5.2
-ACdUUCGGU- <sup>#</sup>	47.3	140.1	64.6	1.3	3.9
-ACmUUCGGU- <sup>#</sup>	41.7 <sup>¶</sup>	126.0	57.7	0.3	2.6
-ACUUmCGGU- <sup>#</sup>	39.2 <sup>¶</sup>	113.6	71.9	1.9	4.0

Series A	Series B
XY	UY
X Z	X G
CG	CG
GC	AU
5'-CG-3'	GC
	5'-GC-3'
X = U or dU	X = U, dU or 2'-OMeU
Y = C or dC	Y = C, dC or 2'-OMeC
Z = G or dG	

Results from global analysis of three or more UV/Vis melts (15–95°C) collected at 260 nm with a heating rate of 1°/min. Buffer: 10 mM sodium cacodylate, 0.5 mM EDTA, pH 7. Multiple entries for a given molecule show the typical reproducibility of these experiments.

\*RNA molecules from series A.

<sup>#</sup>Molecules from series B.

<sup>§</sup>Single experiment.

<sup>¶</sup>Measured by Dr. Lance Laing.

substitution completely disrupts the loop structure (data not shown). Experimental results show that a 2'-OMe substitution at C<sub>7</sub> is isoenergetic, in contrast to the destabilizing 2'-deoxy substitution; we hypothesize that the O-methyl group may engage in favorable hydrophobic contacts at this

position, compensating for the loss of the hydrogen bond. Similar interactions may be responsible for O-methyl substitutions being slightly stabilizing in the context of A-form helices (Lesnik and Freier, 1998).

In the simulation of Miller and Kollman, the C<sub>7</sub> 2'-OH group is solvent exposed and appears to interact strictly with water molecules, whereas this hydroxyl interacts stably with its own O3' oxygen in the GB/SA simulation (2.3 ± 0.4 Å average distance), although the angle is suboptimal for hydrogen bonding (107 ± 20°). The discrepancy between the two simulations may be due to kinetic barriers in the explicit solvent simulation, as a result of the C<sub>7</sub> 2'-OH being involved in stable solvent-mediated hydrogen bonds. Unless the solvent molecule is observed to exchange more than once during the simulation, this possibility cannot be ruled out (the authors do not state whether this is so). Kinetic barriers due to the slow diffusion of water molecules have previously been implicated in the irreversibility seen in certain molecular dynamics simulations (Singh and Kollman, 1996). Although the hydrogen bonding pattern may be an artifact of the GB/SA solvation treatment, the significant positional differences seen in the behavior of hydroxyl

**TABLE 5** 2'OH dihedral angle (H-O2'-C2'-C3'): mean and standard deviation

	GB/SA*	Vacuum (1/R) <sup>#</sup>
G <sub>1</sub>	-78 ± 64	-12 ± 13
G <sub>2</sub>	-108 ± 62	-20 ± 14
A <sub>3</sub>	-101 ± 57	-70 ± 23
C <sub>4</sub>	-55 ± 63	-48 ± 10
U <sub>5</sub>	-167 ± 91	-70 ± 38
U <sub>6</sub>	9 ± 91	24 ± 24
C <sub>7</sub>	25 ± 41	40 ± 12
G <sub>8</sub>	-44 ± 26	-51 ± 10
G <sub>9</sub>	-73 ± 67	-66 ± 24
U <sub>10</sub>	-73 ± 43	-61 ± 45
C <sub>11</sub>	-74 ± 64	-36 ± 29
C <sub>12</sub>	-149 ± 75	173 ± 26

\*Combined 1.149-ns and 3.060-ns GB/SA simulations.

<sup>#</sup>5.011-ns simulation with distance-dependent dielectric (1/R).

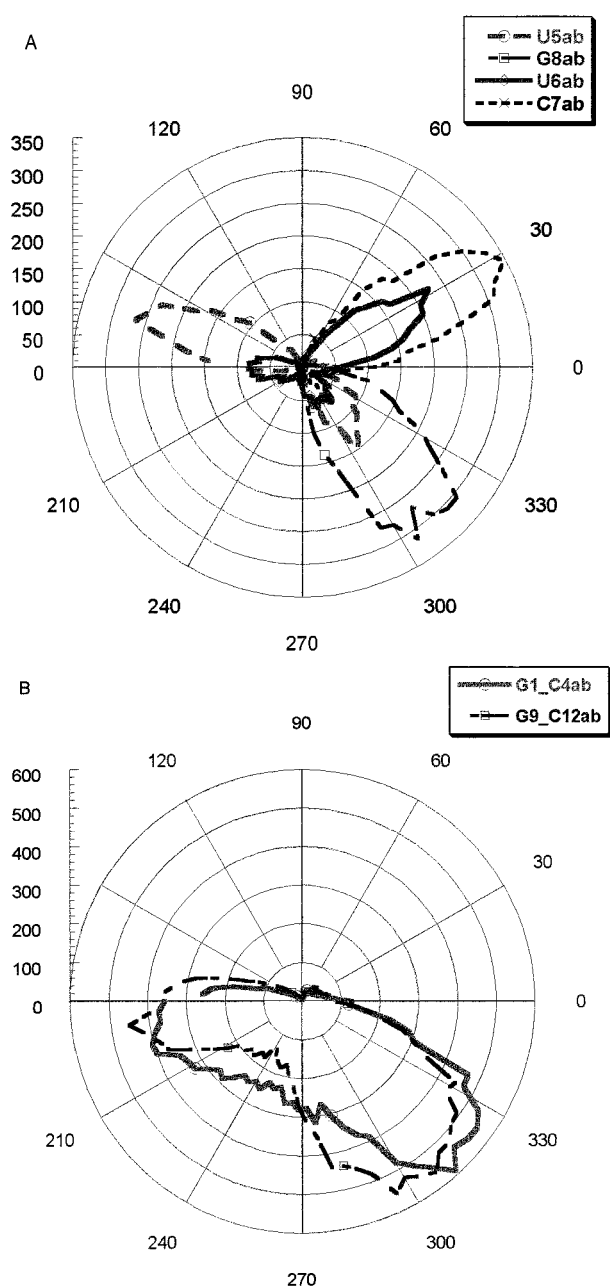


FIGURE 5 Polar plot of 2'-OH dihedral angle distributions (H-O2'-C2'-C3') for loop (A) and stem (B) residues: GB/SA(a) and GB/SA(b) combined.

groups in the GB/SA simulations and the differences from the vacuum simulations suggest that the hydrogen bonding patterns are not simply determined by the absence of explicit solvent. Experimental evidence to test these possibilities can be obtained from NMR studies to identify those hydroxyls protected from solvent exchange and to determine rates of exchange as a function of pH to determine whether these groups are involved in stable intramolecular hydrogen bonds.

The U<sub>5</sub> hydroxyl proton is observed to interact with the G<sub>8</sub> O6 oxygen in the refined NMR structure (Allain and

Varani, 1995), whereas it rotates to hydrogen bond with the U<sub>5</sub>pU<sub>6</sub> O5' backbone oxygen in the explicit solvent simulation (Miller and Kollman, 1997b). In the two GB/SA simulations, the hydroxyl alternates between two dihedral orientations corresponding to hydrogen bonding with the O6 and O5' positions (Table 3 and Fig. 5). The hydroxyl spends more time on average hydrogen bonded to the G<sub>8</sub> O6 group, giving an overall *trans* orientation with respect to the C3' position (Table 5), in agreement with the interpretation of recent NMR measurements of long-range <sup>13</sup>C coupling constants in the UUCG tetraloop (Lynch et al., 1996).

## CONCLUSIONS

The spectacular performance of the GB/SA continuum model in reproducing the structural properties of the UUCG tetraloop suggests that this implicit solvation model is not limited in utility to studies of organic or polar molecules (Fig. 6 A). In contrast, the 1/*R* distance-dependent dielectric model performed poorly under identical simulation conditions, resulting in converged structures more than 3 Å from the starting NMR structure. As seen in Fig. 6 B, these structures show significant distortion of the overall structure, despite the fact that the hydrogen bonding patterns in the loop and stem are maintained. An additional vacuum simulation without counterions shows somewhat improved behavior, but remains inferior to the GB/SA simulations. Despite their limitations, simple distance-dependent dielectric models with their low computational overhead continue to be used in a variety of applications in which some degree of solvation is necessary. Performance of the linear dielectric model has historically been improved with a judicious choice of phosphate charge (e.g., decreasing the magnitude of the phosphate charge to account for screening due to the counterions) or by the use of a 1/4*R* potential. However, it is not known a priori what the effective phosphate charge should be for a short oligomer, and the behavior of this simple dielectric model is highly system dependent. There seems to be little reason to prefer a linear dielectric model, given the superior performance of the GB/SA model at a cost of at most five times the CPU requirement for a comparable vacuum simulation (212 (GB/SA) versus 43 (1/*R*) cpu seconds per picosecond simulation). An explicit solvent simulation on a system of this size (hairpin and counterions plus 2000 water molecules) with a 9-Å cutoff and a pair list updated every 10 time steps would be approximately five times as expensive as a GB/SA simulation. PME treatment of electrostatics would cost an additional 40–50% over a simple 9-Å cutoff (Cheatham et al., 1995).

Explicit solvent simulations of this tetraloop using the AMBER94 force field have shown that RNA structures can be metastable over the range of simulation times employed in this study (Miller and Kollman, 1997b). In the work by Miller and Kollman, an earlier model of the UUCG tetraloop having a somewhat different loop hydrogen bonding geometry did not convert to the correct loop structure during

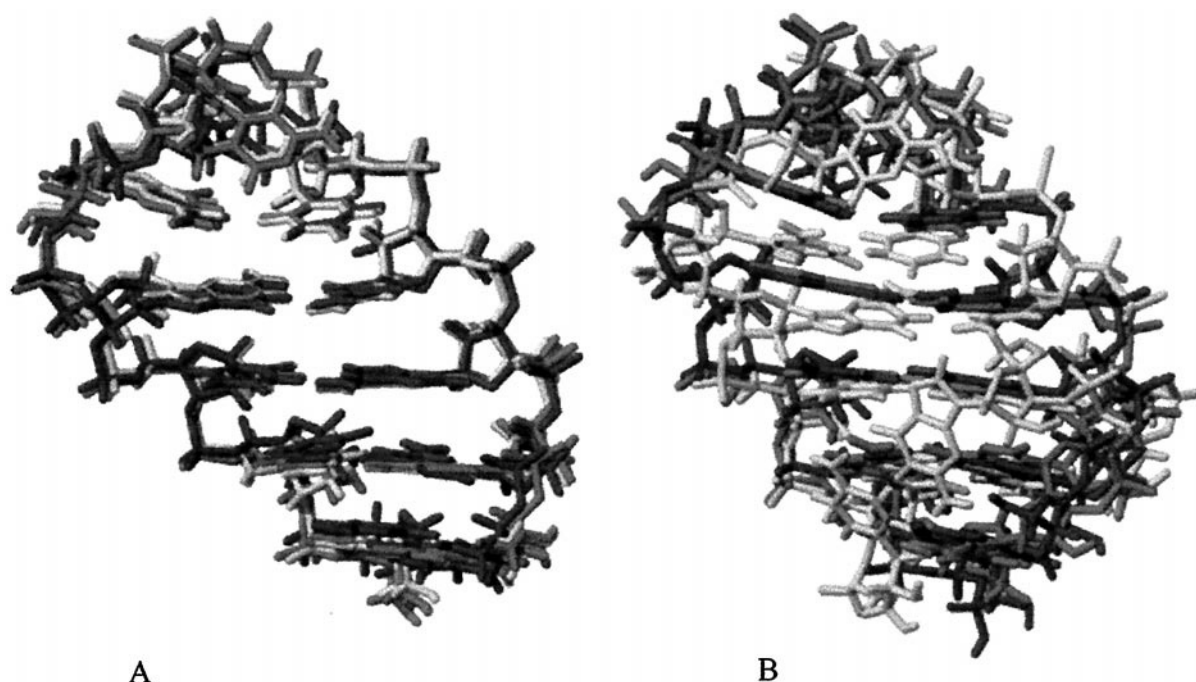
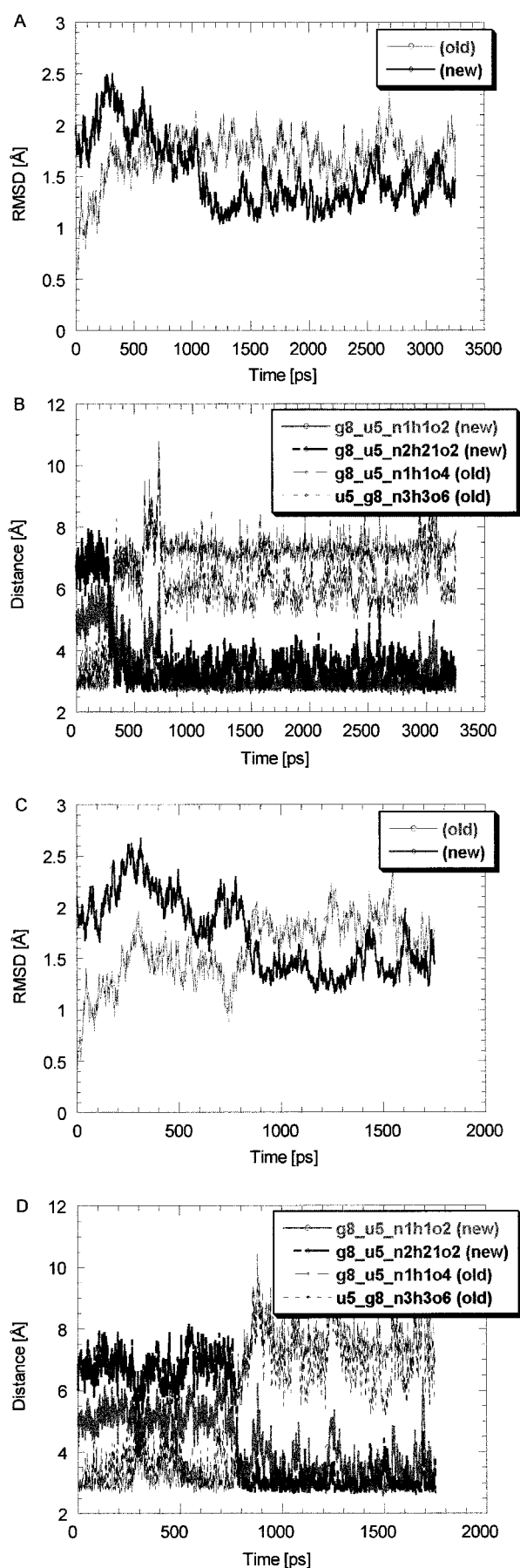


FIGURE 6 Average simulation structure for GB/SA(a), GB/SA (b), and initial NMR structure on the left (A). Average simulation structure for vacuum simulation plus counterions, vacuum simulation without counterions, and the initial NMR structure on the right (B).

a 2-ns explicit solvent simulation. Although the simulation using the earlier model was somewhat less stable than the one starting from the correct loop geometry, the critical loop hydrogen bonds did not rearrange—the structures did not interconvert. A recent molecular dynamics study by Kollman and co-workers using locally enhanced sampling techniques (LES) to increase exploration of the potential surface was able to observe this conversion between structures (Simmerling et al., 1998). We performed a similar test of our simulation methodology by carrying out a 3-ns GB/SA simulation starting from the old NMR model. As can be seen in Fig. 7, after  $\sim 1200$  ps the GB/SA simulation converted from the old to the new geometry loop. The hydrogen bond geometry in the loop rearranges in less than 1000 ps, but the complete conversion, as monitored by the rmsd of the simulation (residues  $A_3$  through  $U_{10}$ ) from either starting structure, is not complete until later. The terminal base pair shows considerable fraying in this simulation, and so the simulation has not completely converged to the “new” NMR structure at the end of the 3-ns period. Excluding the first two base pairs, however, the rmsd fit between the simulation and the new NMR structure is only 1.35 Å, compared with an initial value of 1.99 Å. The enhanced sampling seen in the GB/SA simulations relative to the analogous explicit solvent system is particularly noteworthy, because both studies employed the AMBER94 force field. However, the potentially complex interplay between the implicit solvation model and the stochastic dynamics treatment makes detailed kinetic analysis and comparison of the two systems difficult on the basis of a single set of simulations. As shown in Fig. 7, an almost identical second

simulation, differing only in the frictional coefficient ( $2.5$  instead of  $5 \text{ ps}^{-1}$ ) showed both a slightly different time scale and pathway for the transition. Clearly the kinetics of the simulation can be influenced by the stochastic model, but it is difficult to draw quantitative conclusions from two short simulations.

Methodological advances in molecular dynamics protocols have resulted in explicit solvent simulations of nucleic acids increasingly able to reproduce experimental data. One application is the use of MD to model the properties of DNA, as shown in the analysis of the Dickerson dodecamer by Beveridge and co-workers (Young et al., 1997). This seminal study described the DNA over the course of a 5-ns simulation, during which several sugar repuckering events and sequence-dependent bending motions were observed. The authors also studied in detail the sodium counterion and solvent distributions around the B-form DNA duplex. The extensive exploration of simulation parameters and conditions reported by Young et al. provides an example of the rigorous methodological examinations that must be applied to GB/SA simulations of nucleic acid systems. In another application, Cheatham and Kollman (1996, 1997) used explicit solvent simulations with PME treatment of electrostatics in a thorough analysis of the A-to-B transition of DNA, and a comparison of A- and B-form structures. This simulation was remarkably successful in modeling this solvent-driven transition, suggesting that this methodology can be applied to other conformational changes such as the DNA B-to-Z and RNA A-to-Z transitions. If the GB/SA continuum model is to be accepted as a successful substitute for explicit solvent simulations, then it will be important to



test its ability to reproduce these conformational transitions. In other applications, explicit simulations have been successfully combined with GB/SA and other continuum solvent treatments to perform free energy analyses of solvent-mediated conformational preferences (Jayaram et al., 1998; Srinivasan et al., 1998). Curiously, the explicit solvent/PME simulation was unsuccessful in reporting on the transition of the UUCG tetraloop from the incorrect to correct tetraloop conformation (Miller and Kollman 1997b); this transition was observed in a more recent simulation using LES to enhance sampling (Simmerling et al., 1998). In contrast, the structural interconversion occurs relatively quickly for both GB/SA simulations started in the incorrect conformation. The reasons for this success are possibly complex and may involve both solvation-dependent differences in effective barrier heights and the influence of the stochastic model.

The GB/SA solvation treatment allows nucleic acid simulations to be carried out at a fraction of the expense of an explicit solvent model, apparently without significant loss of accuracy and notably without the need for reparameterization (the model was parameterized for the OPLS charge set). The lack of explicit solvent probably allows greater sampling of internal degrees of freedom during the simulation and may be particularly advantageous in *ab initio* structure prediction or NMR refinement. The reduced computational overhead of these simulations relative to explicit ones allows much longer simulations to be carried out in a reasonable time. We believe these simulations can be used in complementary fashion with shorter, more detailed explicit solvent simulations to study dynamic and structural properties of nucleic acids. Undoubtedly, the neglect of solvent degrees of freedom may be a significant limitation for certain applications. Highly solvated, flexible hydrogen bonding groups are particularly sensitive to the solvation treatment. Furthermore, the simple stochastic model of the solvent (white noise) may not adequately represent the details of the dynamic interactions of certain groups with the solvent. Nevertheless, the agreement between simulation and experiment shown here encourages future studies with more detailed analysis and greater comparison with experiments. Our results indicate that the GB/SA model will continue to prove useful for simulations of nucleic acids in solution both independently and in conjunction with explicit solvent studies.

We thank Prof. Ignacio Tinoco, Jr., for providing us with the coordinates of the original tetraloop structure. We also acknowledge the Center for Molecular Design and the Institute of Biomedical Computing at Washing-

FIGURE 7 Mass-weighted root mean square deviation (not including  $K^+$  counterions or hydrogens) from original NMR structure (old) and revised NMR structure (new) for residues A<sub>3</sub>-U<sub>10</sub> as a function of time for a GB/SA simulation starting with the original model. (A, C) 3.0-ns production phase plus 0.25-ns equilibration period. (B, D) Hydrogen bond distances for the four hydrogen bonds characteristic of the two loop structures. (A, B)  $\gamma = 0.1 \text{ ps}^{-1}$ . (C, D)  $\gamma = 0.2 \text{ ps}^{-1}$ .



ton University and, in particular, Prof. Garland Marshal and Prof. Michael Zuker for the use of their computer resources. We are also greatly indebted to Dr. Rohit Pappu and Dr. Reece Hart for their assistance and many helpful discussions.

This work was supported in part by a Gerty T. Cori Sigma Chemical Company Predoctoral Fellowship to DJW.

## 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.
- Auffinger, P., S. Louise-May, and E. Westhof. 1995. Multiple molecular dynamics simulations of the anticodon loop of tRNA(Asp) in aqueous solutions with counterions. *J. Am. Chem. Soc.* 117:6720–6726.
- Auffinger, P., and E. Westhof. 1997. Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.* 274:54–63.
- Beveridge, D. L., and G. Ravishanker. 1994. Molecular dynamics studies of DNA. *Curr. Opin. Struct. Biol.* 4:246–255.
- Briki, F., J. Ramstein, R. Lavery, and D. Genest. 1991. Evidence for the stochastic nature of base pair opening in DNA: a Brownian dynamics simulation. *J. Am. Chem. Soc.* 113:2490–2493.
- Brunger, A. T. 1992. X-PLOR 3.1: A System for X-ray Crystallography and NMR. Yale University Press, New Haven, CT.
- Caves, L. S. D., J. D. Evanseck, and M. Karplus. 1998. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* 7:649–666.
- 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 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, J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 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., G. Varani, and I. Tinoco, Jr. 1990. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature*. 346:680–682.
- Conte, M. R., L. Graeme, T. Brown, and A. N. Lane. 1996. Hydration of the RNA duplex r(CGCAAAUUGCG)<sub>2</sub> determined by NMR. *Nucleic Acids Res.* 24:3693–3699.
- 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.
- Daggett, V., P. A. Kollman, and I. D. Kuntz. 1991. Molecular dynamics simulations of small peptides: dependence on dielectric model and pH. *Biopolymers*. 31:285–304.
- Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an  $N \log(N)$  method for Ewald sums in large systems. *J. Phys. Chem.* 98:10089–10092.
- Dudek, M. J., and J. W. Ponder. 1995. Accurate modeling of the intramolecular electrostatic energy of proteins. *J. Comp. Chem.* 16:791–816.
- Freier, S. M., D. D. Albergo, and D. H. Turner. 1983. Solvent effects on the dynamics of (dG-dC)<sub>3</sub>. *Biopolymers*. 22:1107–1131.
- Fritsch, V., G. Ravishanker, D. L. Beveridge, and E. Westhof. 1993. Molecular dynamics simulations of poly(dA)poly(dT): comparisons between implicit and explicit solvent representations. *Biopolymers*. 33:1537–1552.
- Guarnieri, F., and W. C. Still. 1994. A rapidly convergent simulation method: mixed Monte Carlo/stochastic dynamics. *J. Comp. Chem.* 15:1302–1310.
- Guarnieri, F., and H. Weinstein. 1996. Conformational memories and the exploration of biologically relevant peptide conformations: an illustration for the gonadotropin-releasing hormone. *J. Am. Chem. Soc.* 118:5580–5589.
- Hall, K. B. 1994. Interactions of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry*. 33:10076–10088.
- Harvey, S. C. 1989. Treatment of electrostatic effects in macromolecular modeling. *Proteins*. 5:78–92.
- Heus, H. A., and A. Pardi. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*. 253:191–194.
- Hingerty, B. E., R. H. Ritchie, T. L. Ferrell, and J. E. Turner. 1985. Dielectric effects in biopolymers: the theory of ionic saturation revisited. *Biopolymers*. 24:427–439.
- Jayaram, B., D. Sprous, M. A. Young, and D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–10633.
- Jean-Charles, A., A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T. F. Hendrickson, and W. C. Still. 1991. Electrostatic contributions to solvation energies: comparison of free energy perturbation and continuum calculations. *J. Am. Chem. Soc.* 113:1454–1455.
- Lesnik, E. A., and S. M. Freier. 1998. What affects the effect of 2'-alkoxy modifications? 1. Stabilization of 2'-methoxy substitutions in uniformly modified DNA oligonucleotides. *Biochemistry*. 37:6991–6997.
- Lynch, S. R., J. G. Pelton, and I. Tinoco, Jr. 1996. NMR assignment of a 2'-hydroxyl proton from the UUCG tetraloop through long-range correlations with <sup>13</sup>C. *Magn. Reson. Chem.* 34:S11–S17.
- Marrone, T. J., M. K. Gilson, and J. A. McCammon. 1996. Comparison of continuum and explicit models of solvation: potentials of mean force for alanine dipeptide. *J. Phys. Chem.* 100:1439–1441.
- Mazur, J., and R. L. Jernigan. 1991. Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers*. 31:1615–1629.
- McCammon, J. A., and S. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge.
- Miaskiewicz, K., R. Osman, and H. Weinstein. 1993. Molecular dynamics simulations of the hydrated d(CGCGAATTCGCG)<sub>2</sub> dodecamer. *J. Am. Chem. Soc.* 115:1526–1537.
- Miller, J. L., and P. A. Kollman. 1997a. Observation of an A-DNA to B-DNA transition in a non-helical nucleic acid hairpin molecule using molecular dynamics. *Biophys. J.* 73:2702–2710.
- Miller, J. L., and P. A. Kollman. 1997b. 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.* 270:436–450.
- Mohamadi, F., N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, and W. C. Still. 1990. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comp. Chem.* 11:440–467.
- Nowakowski, J., J. L. Miller, P. A. Kollman, and I. Tinoco, Jr. 1996. Time evolution of NMR proton chemical shifts of an RNA hairpin during a molecular dynamics simulation. *J. Am. Chem. Soc.* 118:12812–12820.
- Palmer, A. G., III, and D. A. Case. 1992. Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide. *J. Am. Chem. Soc.* 114:9059–9067.
- Pappu, R. V., G. R. Marshall, and J. W. Ponder. 1999. A potential smoothing algorithm accurately predicts transmembrane helix packing. *Nature Struct. Biol.* 6:50–55.
- Piela, L., J. Kostrowicki, and H. A. Scheraga. 1989. The multiple-minima problem in the conformational analysis of molecules. Deformation of the potential energy hypersurface by the diffusion equation method. *J. Phys. Chem.* 93:3339–3346.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1992. Numerical Recipes: The Art of Scientific Computing, 2nd ed. Cambridge University Press, Cambridge. 689–692.
- Qiu, D., P. S. Shenkin, F. P. Hollinger, and W. C. Still. 1997. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem.* 101:3005–3014.
- Ramstein, J., and R. Lavery. 1988. Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. USA*. 85:7231–7235.

- 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. Comp. Phys.* 23: 327–341.
- Sakata, T., H. Hiroaki, Y. Oda, T. Tanak, 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.
- Schaefer, M., C. Bartels, and M. Karplus. 1998. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* 284:835–848.
- Schreiber, H., and O. Steinhauser. 1992. Cutoff size does influence molecular dynamics results on solvated polypeptides. *Biochemistry.* 31: 5856–5860.
- Simmerling, C., J. L. Miller, and P. A. Kollman. 1998. Combined locally enhanced sampling and particle mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *J. Am. Chem. Soc.* 120:7149–7155.
- Singh, S. B., and P. A. Kollman. 1996. Understanding the thermodynamic stability of an RNA hairpin and its mutant. *Biophys. J.* 70:1940–1948.
- Smart, J. L., T. J. Marrone, and J. A. McCammon. 1997. Conformational sampling with Poisson-Boltzmann forces and a stochastic dynamics/Monte Carlo method: application to alanine dipeptide. *J. Comput. Chem.* 18:1750–1759.
- Smith, D. A., and S. Vijayakumar. 1991. Molecular modeling of intramolecular hydrogen bonding in simple oligoamides 2. GB/SA CH2Cl2. *Tetrahedron Lett.* 32:3617–3620.
- Soman, K. V., A. Karimi, and D. A. Case. 1993. Molecular dynamics analysis of a ribonuclease C-peptide analogue. *Biopolymers.* 33: 1567–1580.
- Srinivasan, J., T. E. Cheatham, III, P. Cieplak, P. A. Kollman, and D. A. Case. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidite-DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.
- Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.
- Tinoco, I., Jr. 1996. Nucleic acid structures, energetics, and dynamics. *J. Phys. Chem.* 100:13311–13322.
- Tuerk, C., P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. D'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco, Jr., E. N. Brody, and L. Gold. 1988. CUUCGG hairpins: extraordinarily stable RNA secondary structure associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA.* 85:1364–1368.
- van Gunsteren, W. F., and H. J. C. Berendsen. 1988. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* 1:173–185.
- Varani, G., G. Cheong, and I. Tinoco, Jr. 1991. Structure of an unusually stable RNA hairpin. *Biochemistry.* 30:3280–3289.
- Williams, D. J., and K. B. Hall. 1996. Conformations of 1,2-dimethoxyethane in gas and solution phase from molecular mechanics and Monte Carlo/stochastic dynamics simulations. *J. Phys. Chem.* 100: 8224–8229.
- Woese, C. R., S. Winker, and R. R. Gutell. 1990. Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops.” *Proc. Natl. Acad. Sci. USA.* 87:8467–8471.
- Young, M. A., G. Ravishanker, and D. L. Beveridge. 1997. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys. J.* 73:2313–2336.
- 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.