Understanding the Thermodynamic Stability of an RNA Hairpin and Its Mutant

Suresh B. Singh* and Peter A. Kollman‡
*Department of Structural Biology, Wyeth-Ayerst Research, CN 8000, Princeton, New Jersey 08543, and ‡Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143 USA

ABSTRACT The thermodynamic stability of RNA hairpin loops has been a subject of considerable interest in the recent past (Wimberly et al., 1991). There have been experimental reports indicating that the hairpins with a C(UUCG)G loop sequence are thermodynamically very stable (Wimberly et al., 1991). We used the solution structure of GGAC(UUCG)GUCC (Cheong et al., 1990; Varani et al., 1991) as the starting conformation in our attempt to understand its thermodynamic stability. We carried out molecular dynamics/free energy simulations to understand the basis for the destabilization of the C(UUCG)G loop by mutating cytosine (C_7)→uracil. Because of the limited length of simulation and the presence of kinetic barriers (solvent intervention) to the uracil→cytosine mutation, all of our computed free energy differences are based on multiple forward simulations. Based on these calculations we find that the cytosine→uracil mutation in the loop destabilizes it by ~1.5 kcal/mol relative to that of the reference state, an A-form RNA but with cytosine (C_7) looped out. This is of the same sign and magnitude as that observed in the thermodynamic studies carried out by Varani et al. (1991). We have carried out free energy component analysis to understand the effect of mutating the cytosine residue to uracil on the thermodynamic stability of the C(UUCG)G hairpin loops. Our calculations show that the most significant contribution to the stability is from the phosphate group linking U_5 and U_6, which favors the cytosine residue over uracil by about 6.0 kcal/mol. The residues U_5, U_6, and G_8 in the loop region also contribute significantly to the stability. The contributions from the salt and solvent compensate each other, indicating the dynamic nature of interactions of the environment with the nucleic acid system and the coupling between these two components.

INTRODUCTION

RNA is known to have a number of biological functions, including the transfer of the genetic code for proteins (Stryer, 1988), carrying out catalytic activity (Zaug and Cech, 1986), and serving as the primary genomic material in some organisms (Stryer, 1988). Because there has been only a limited amount of high-resolution structural data available until recently, there have been very few molecular dynamics studies carried out on RNA. However, recently there has been an increase in the number of solution structures available on short RNA segments (Wimberly et al., 1991). The first series of molecular dynamics simulations were carried out by Harvey and co-workers on $tRNA^{phe}$ based on the crystal structure of yeast phenylalanine tRNA (McCammon and Harvey, 1987). Nilsson and co-workers have presented molecular dynamics studies on the self-splicing intervening sequence of the t4 NRDB mRNA (Nilsson et al., 1990). More recently, Zichi reported a 200-ps simulation of the NMR structure of the RNA hairpin GGGCGCAAGCU (Zichi, 1995). Very recently, Cheatham et al. (1995) presented results of molecular dynamics simulations of the RNA hairpin GGAC(UUCG)GUCC with the use of a particle mesh Ewald method for treating long-range electrostatics. To our knowledge the work presented here is the first attempt to understand the structure and stability of a RNA hairpin system by free energy simulations.

RNA hairpins are thought to play an important structural role in providing potential nucleation sites for RNA folding, serving as a termination site for the transcriptional process, and interacting with proteins and nucleic acids (Wimberly et al., 1991; Cheong et al., 1990). In our current studies we have investigated the structure, stability, and dynamics of the RNA hairpin $G_1G_2A_3C_4(U_5U_6C_7G_8)G_9U_{10}C_{11}C_{12}$. The solution structure of the hairpin GGAC(UUCG)GUCC (see Fig. 1) were reported by Varani et al. (1991). They observed that this hairpin is very stable and the mutation of cytosine (C_7)→uracil destabilizes the hairpin by ~1.5 kcal/mol. They ascribed this effect to the loss of the electrostatic interactions between the phosphate group linking U_5-U_6 and the N^4 amino group of C_7 and the stacking interactions between C_7 and the neighboring bases. James and Tinoco presented the structure of the deoxy ribose analog of the above RNA hairpin (d[GGAC(TTCG)GTCC]) and showed that the deoxy ribose analog is less compact than the ribose analog and hence less stable (James and Tinoco, 1993). We have carried out molecular dynamics/free energy simulations (FES) on the hairpin to investigate the nature of the stability of the ribosyl C(UUCG)G loop and we report our findings here.

The hairpin has residues G_1-C_4 and G_9-C_{12} in the stem, and the loop residues U_5 and G_8 have a Hoogsteen-type hydrogen bonding pattern, with the guanine in the *syn* conformation. The other two loop residues are U_6 and C_7.

Received for publication 5 September 1995 and in final form 31 October 1995.

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.

0006-3495/96/04/1940/09 $2.00

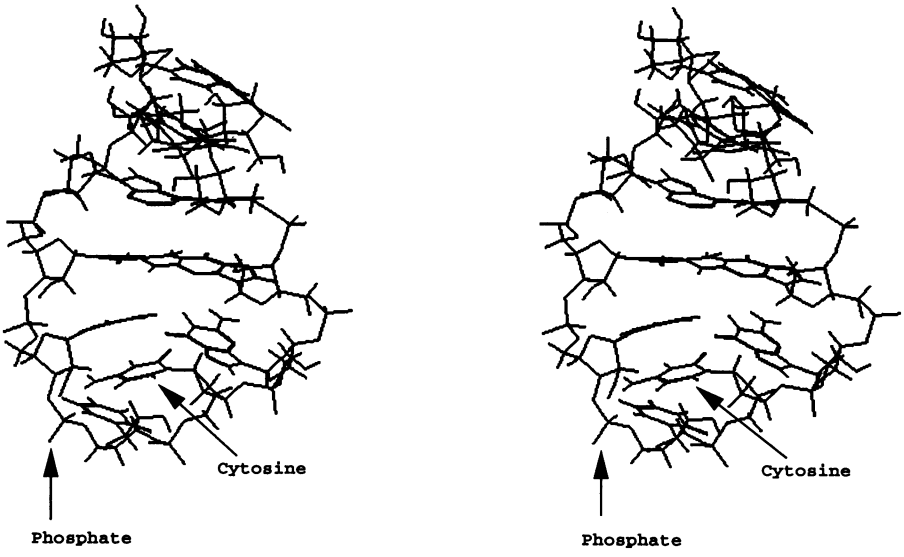

FIGURE 1 The structure of the RNA hairpin GGAC(UUCG)GUCC after 10 ps of molecular dynamics equilibration is shown in stereo. The phosphate residue linking U_5-U_6 and the cytosine (C_7) are labeled.

The N^4-amino group of C_7 has a hydrogen bond with the phosphate group between U_5 and U_6 (see Fig. 1). In our free energy simulations we mutated cytosine (C_7)→uracil in the RNA hairpin structure. For the single-strand denatured model we generated an A-form RNA with the cytosine (C_7) looped out. The thermodynamic cycle used for the FES studies is given in Fig. 2. The processes represented by the free energy changes ΔG_1 and ΔG_2 are the experimentally observed phenomena (Varani et al., 1991) that cannot be simulated in our approach. Instead, in our computer experiments we carried out the processes represented by the free energy changes ΔG_3 and ΔG_4, where the cytosine residue is mutated to uracil via computer alchemy. The experimental free energy difference for the cytosine→uracil change is given by

$$\Delta\Delta G_1 = \Delta G_1 - \Delta G_2, \quad (1)$$

and the computed free energy difference is given by

$$\Delta\Delta G_2 = \Delta G_3 - \Delta G_4. \quad (2)$$

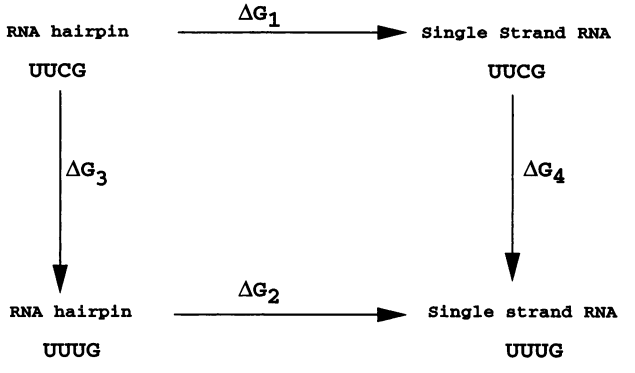

FIGURE 2 Thermodynamic cycle for the calculation of the free energy difference for cytosine→uracil mutation in the RNA hairpin GGAC(UUCG)GUCC.

Below we will show that the computed free energy difference agrees very well with the experimentally determined value (Varani et al., 1991), ($\Delta\Delta G_1 \approx \Delta\Delta G_2$) ≈ 1 kcal/mol.

We carried out free energy component analysis to decompose the contributions from the surrounding bases, solvent, and salt to understand the basis for the destabilization of the hairpin when cytosine is mutated to uracil in the loop.

Recently some serious concerns about the utility of free energy components were expressed (Yunyu et al., 1993; Mark and van Gunsteren, 1994). Some questions were raised about the validity of the approach and the use of components to understand molecular interactions (Mark and van Gunsteren, 1994). It was stated that the free energy is a state function and hence it is path independent, whereas the free energy components are not state functions and hence are path dependent. This has always been our position (Kollman, 1993). It was also stated that it is not possible to get a meaningful decomposition of the free energy in terms of specific inter-residue or inter-atom interactions (Mark and van Gunsteren, 1994). On the other hand, our laboratory has shown in a number of cases that free energy components give a useful insight into the nature of intermolecular interactions that take place during the course of a thermodynamic process (Kollman, 1993). Our opinions and experiences are similar to those recently expressed by Boresch et al. (1994, 1995) and Brady and Sharp (1995).

In our recent efforts we have shown that the A · T base pairs are energetically less favorable in a Z-form than in a B-form DNA (Singh et al., 1993). We were able to attribute this effect to the difference in stacking interactions between the neighboring residues in the B- and Z-forms based on their free energy components. This study and the free energy component analysis helped us understand the nature of inter-residue interactions that contribute to the "Z-phobicity" of A · T base pairs (Singh et al., 1993). Thus, our experience shows that this methodology offers a powerful way to extract very useful information that is otherwise left

unexplored. Hence, we carried out free energy component analysis to understand the basis for the destabilization of the hairpin when cytosine is mutated to uracil in the loop in the present work.

In the next section we present the simulation methodology and the theoretical background for the methods involved in the derivation of free energy components. In the third section we present the issues related to the methodology employed in our current free energy calculations. Finally, we present the free energy data and their components in the third section and discuss the implications of the results and the effect of simulation conditions on them.

METHODS

The starting conformation for our free energy simulations was the NMR structure of $G_1G_2A_3C_4(U_5U_6C_7G_8)G_9U_{10}C_{11}C_{12}$ (Varani et al., 1991). We choose the A-form RNA with the cytosine (C_7) looped out as the solvent-exposed single-strand reference state. The single-strand RNA was first generated based on Arnott's fiber diffraction data for the A-form double-strand RNA (Arnott et al., 1973), and then the backbone torsion angles of C_7 were adjusted such that the cytosine residue was maximally destacked and exposed to the solvent. The backbone torsion angles used for the residues surrounding the cytosine in the starting structure are given in Table 1. To maintain electroneutrality in our calculations we placed sodium ions at 3.6 Å from the phosphate bisector in the initial configuration. The starting structures were solvated by placing three site TIP3P water molecules (Jorgensen et al., 1983) around the solute such that the solvent molecules were placed up to 11 Å away in simulations with a 10 Å cutoff (Runs 1, 2, and 4–6, Table 2) and up to 13 Å away in simulations with a 12 Å cutoff (Run 3, Table 2) along each of the rectangular coordinate axes and as close as 3.0 Å from any given solute atom. The total number of water molecules and the box dimensions employed in the calculations are given in Table 2.

Equilibration

The equilibration protocol adopted in our current simulations is similar to the one presented previously (Singh et al., 1994). The rationale for adopting this particular protocol was to relieve bad contacts between the solute and the solvent, allowing the solvent molecules to reorient and make favorable contacts with the solute. In our first step we restrained the starting structure with 25 kcal/mol residue harmonic restraint and carried out potential energy minimization for about 1000 iterations followed by molecular dynamics at 300 K for about 3 ps. Then we released the restraints on the solute in five steps by reducing the restraint by 5 kcal/mol residue per step followed by 600 iterations of conjugate gradient minimization. Thus in the fifth step the whole system was minimized without harmonic restraints. Then the system was heated up to 300 K with a temperature coupling of 0.2 ps. In all of our simulations with the A-form RNA we restrained the terminal three bases on each end with a 5.0 kcal/mol-residue harmonic restraint. All of the covalent bonds to the hydrogens were held constant to a tolerance of 0.0005 Å by applying the SHAKE routine (van Gunsteren and Berendsen, 1977). We have used a 1-fs time step in all of our calculations. The temperature of the system was allowed to fluctuate around 300 K with a temperature coupling time of 0.2 ps, and the pressure was allowed to fluctuate around 1 bar with a pressure coupling time of 0.6 ps. We implemented cubic periodic boundary conditions with a 10-Å cutoff for the electrostatic and Lennard-Jones potential evaluations for solute-water and water-water interactions. All solute-solute interactions including sodium ions were evaluated under the assumption that this will avoid the cutoff artifacts on electrostatic interactions of sodium ions with phosphates without serious consequences for the solute-solvent interactions. The evaluation of all solute-solute interactions in our periodic simulations is possible because we do not generate images for the solute atoms. This is done to avoid double counting of the nonbonded interactions. All of our calculations were carried out using AMBER 4.0 (Pearlman et al., 1991).

TABLE 1 The backbone torsion angles of U_6-P-C_7 in the A-form RNA

Torsion	Symbol	ϕ (degrees)
C4′-C3′-O3′-P	ϵ	208.4
C3′-O3′-P-O5′	ζ	82.8
O3′-P-O5′-C5′	α	297.9
P-O5′-C5′-C4′	β	150.2
O5′-C5′-C4′-C3′	γ	46.0
C5′-C4′-C3′-O3′	δ	83.4
C4′-C3′-O3′-P	ϵ	200.6

TABLE 2 Water molecules and box dimensions in molecular dynamics and free energy simulations

Conformer	No. H_2O	Box dimensions, Å^3
Hairpin		
Runs 1,* 2,‡ and 4–6	2485	46.761 × 43.527 × 38.694
Run 3§	3809	52.795 × 47.524 × 45.438
Single-strand A-form		
Runs 1, 2, and 4¶ 5,‖ 6**	3957	63.590 × 44.267 × 34.630
Run 3	4592	71.968 × 51.923 × 41.925

* A 100-ps simulation with a 10-Å uniform cutoff.
‡ A 100-ps simulation with a 10-Å cutoff for solute-water and water-water interactions.
§ A 100-ps simulation with a 12-Å cutoff for solute-water and water-water interactions.
¶ A 50-ps simulation deriving free energy components with a 10-Å cutoff for solute-water and water-water interactions.
‖ A 100-ps simulation deriving free energy components with a 10-Å cutoff for solute-water and water-water interactions.
** A 100-ps simulation deriving free energy components with a 10-Å uniform cutoff.

Free energy perturbation simulations

In our FEP calculations we used the thermodynamic windows method with fixed widths. The free energy perturbation was carried out by mutating cytosine→uracil in both the hairpin and in the A-form RNA in the forward ($\lambda = 1 \rightarrow 0$) direction only. The force-field parameters describing the cytosine were changed smoothly and uniformly to the ones for uracil employing the methods described elsewhere (Singh et al., 1987). The change in free energy, ΔG_i, is calculated after each step of the transformation and, because the free energy is a state function, the total free energy change, ΔG, is the sum of the intermediate ΔG_i values. The free energy change in the FEP method is given by the following expression:

$$\Delta G_i = kT \ln\left\langle \exp\left(\frac{-\Delta V(\lambda \rightarrow \lambda')}{kT}\right)\right\rangle_\lambda, \qquad (3)$$

where $\langle\ \rangle_\lambda$ is the ensemble average defined by parameter λ, and λ' is a neighboring state of λ, as has been described elsewhere (Singh et al., 1987).

The total free energy is given by

$$\Delta G = \sum_{i=1}^{n} G_i, \qquad (4)$$

where n is the total number of windows in the simulation.

Free energy component analysis

The free energy function cannot be strictly broken down into its constituent components because of the correlation between the nonbonded energy terms (Smith and van Gunsteren, 1994). However, the free energy component analysis can be carried out to understand the physical mechanism underlying the process and the path dependence of the mechanism (Boresch et al., 1994). In a case where a molecular system traverses similar paths, free energy decomposition can yield useful insight into the molecular interactions responsible for the thermodynamic properties. The free energy component analysis is carried out by decomposing the contributions from various residues to the mutating residue as described by Singh et al. (1993). The free energy component analysis gives us the breakdown of the inter-residue and intra-residue contributions and not the enthalpic and entropic contributions to the total free energy change. We present here a brief overview of the methods employed in deriving the free energy components. The free energy component analysis can be carried out by the thermodynamic integration (TI) technique or the slow growth procedure. The free energy difference between two states ($\lambda = 0$ and $\lambda = 1$) can be expressed as

$$\Delta G = G(\lambda = 1) - G(\lambda = 0) = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(\mathbf{r}, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda, \quad (5)$$

where $V(\mathbf{r}, \lambda)$ is the potential energy function, $\mathbf{r}$ is the set of coordinates for all of the atoms, and λ is the coupling parameter between state 0 and state 1. This expression is employed in the thermodynamic integration method, where the integral is approximated by a summation over discrete intervals of λ:

$$\Delta G = \sum_{k=1}^{m} \left\langle \frac{\partial V(\mathbf{r}, \lambda_k)}{\partial \lambda} \right\rangle_{\lambda} \Delta\lambda_k. \quad (6)$$

In the slow growth procedure the change in free energy is obtained by making the increments $\Delta\lambda$ fairly small; then the above integral (Eq. 5) can be approximated by the sum of m increments:

$$\Delta G \approx \sum_{k=1}^{m} \Delta V(\mathbf{r}, \lambda_k), \quad (7)$$

where m is the total number of time steps to be taken in the simulation.

Because the potential energy function is a sum of the bonding (local) and nonbonding (nonlocal) terms, mathematically the total free energy function in all of the formulations can be broken down into a sum of the individual terms. For example, Eq. 7 can be rewritten as

$$\Delta G \approx \sum_{k=1}^{m} \left[\sum_{\text{bonds}} \Delta V(r_{ij}, \lambda_k) + \sum_{\text{angles}} \Delta V(\theta, \lambda_k) + \sum_{\text{dihedrals}} \Delta V(\phi, \lambda_k) \right] + \sum_{k=1}^{m} \sum_{\text{nbpairs}} \Delta V(r_{ij}, \lambda_k), \quad (8)$$

where nbpairs is the nonbonded pairs, which includes the Lennard-Jones, Coulomb, and the hydrogen bond terms.

In the above formulation only the nonbonded interactions between the perturbed residue and unperturbed residues are accounted. Thus, the unperturbed residues make only nonbonded contributions. Because the nonbonded term in Eq. 8 is in turn a sum of pairwise interactions, these terms can be further decomposed into a sum of individual pairwise contributions. These pairwise interactions can be summed up over each residue in the system, and the free energy contributions made by each of unperturbed residues can be extracted thereby. Thus the nonbonded term in Eq. 8 can be rewritten as

$$\sum_{k=1}^{m} \sum_{\text{nbpairs}} \Delta V(r_{ij}, \lambda_k) = \sum_{k=1}^{m} \sum_{\text{nbpairs}} [\Delta V_{uu}(r_{ij}, \lambda_k) + \Delta V_{pu}(r_{ij}, \lambda_k) + \Delta V_{pp}(r_{ij}, \lambda_k)], \quad (9)$$

where V_{uu} is the potential energy term for the nonbonded interactions between the unperturbed residues, V_{pu} is the potential energy term for the nonbonded interactions between the perturbed residue(s) and the unperturbed residues, and V_{pp} is the potential energy term for the nonbonded interactions between the atoms of the perturbed residue. It should be noted that in our current implementation contributions from the unperturbed residues to more than one perturbed residue cannot be separated. Thus the free energy contribution (G_j) from each unperturbed residue (u_j) in the system toward the total free energy change is given by

$$\Delta G_j \approx \sum_{k=1}^{m} \sum_{\text{nbpairs}} \Delta V_{puj}(r_{ij}, \lambda_k). \quad (10)$$

The free energy components are path dependent and hence are not a state function, unlike the total free energy change. However, within the limits of similar paths traversed by a molecular system, the free energy components are similar and thus can be used to obtain contributions from the surroundings toward the perturbed residue. In our free energy component analysis we have used the slow growth procedure to collect free energy statistics and derive the components from various groups contributing to the mutating residue. In all of our free energy perturbation calculations we have used a $\Delta\lambda$ of 0.005 per window. In the slow growth method, $\Delta\lambda$ is given by

$$\Delta\lambda = \frac{1}{m},$$

where m is the total number of time steps to be taken during the course of the simulation.

All of our calculations were carried out on the CRAY-YMP at the San Diego Supercomputer Center, on the CRAY-C90 at the Pittsburgh Supercomputer Center, on the HP-735s at the University of California at San Francisco, and on the SGI Challenge XL at Wyeth-Ayerst Research.

RESULTS AND DISCUSSION

Free energy simulations: methodological issues

In our initial attempts to carry out free energy perturbation calculations on the hairpin we calculated the free energy difference for the cytosine→uracil mutation both for the forward and backward directions. We observed that we could not obtain structural reversibility for the hairpin. It became obvious to us that when cytosine was mutated to uracil the strong attractive electrostatic interactions between the amino group and the phosphate group were lost, and the cytosine/uracil residue drifted away from the phosphate residue. The space cleared by the cytosine/uracil residue was filled with the solvent. Within our finite simulation time (100 ps) the backward simulation did not facilitate the reformation of a hydrogen bond between the phosphate group and the amino group. The process of returning to the free energy minimum pertaining to the solution structure is kinetically hindered by intervening water(s). Here we do not attempt to understand the kinetics of the hydrogen bond

reformation process. Hence, because of the difficulty of reversing the formation of the hydrogen bond between the N^4 of cytosine and the phosphate oxygen during our finite simulations (50–100 ps), all of our free energy differences presented here are based on calculations performed for the forward direction only (see Table 3).

We also observed that our calculated numbers in the RNA hairpin system are extremely sensitive to the nonbonded cutoffs. From our experience with nucleic acid and nucleic acid-drug system simulations (Singh et al., 1994) we have found out that including all of the solute-solute interactions (including sodium ions) preserves the structure better and prevents sodium ions from drifting away from the phosphate groups. We have run simulations with and without all solute-solute interactions to compare the difference in the computed free energy differences and the structural stability of the molecule.

Because we include all solute-solute interactions in our calculations the interactions between the solute and the solvent are certainly affected. We observe that the structural stability of the solute and solvent is affected when the thickness of the solvent layer around the molecule is less than the cutoff radius employed in the calculation. This in turn appears to affect the computed free energy data. Assume that the length of the RNA segment is 30 Å and its width is 20 Å. The solvent layer is 10 Å thick around the solute. In our typical simulation with cubic periodic boundary conditions we have the images of the solvent surrounding the "real" box. We do not generate images of the solute in our periodic simulations. In a case where the nonbonded cutoff radius is 12 Å, for example, the atoms on both the termini of the nucleic acid system will "see" the same set of solvent molecules. Because we include all solute-solute interactions the electrostatic interactions between the two ends of the chain and the loop region with the same set of solvent molecules around the box could potentially lead to structural instability of both the solute and the solvent, as well as unreliable free energy statistics. Hence our free energy simulations on our first series of calculations with the solvent thickness less than that of the cutoff radius lead to unintuitive and irreproducible free energy data (not presented here). Thus based on our experience we ensured that the solvent layer thickness is greater than that of the cutoff radius in all of our calculations.

TABLE 3 Free energy changes in kcal/mol of cytosine→uracil mutation in the RNA hairpin and the A-form RNA with the looped-out cytosine (C_7)

Conditions	Hairpin ΔG_3*	A-form ΔG_4	$\Delta\Delta G$	RMS‡ (A)
Experimental	—	—	1.5§	—
Eqb¶	—	—	—	1.1
Free energy perturbation				
Run 1‖	19.00	16.90	2.10	3.0
Run 2**	17.85	16.88	0.97	3.7
Run 3‡‡	18.50	17.56	0.94	1.8
Slow growth method				
Run 4§§	17.93	16.91	1.02	1.6
Run 5¶¶	18.75	17.44	1.31	2.8
Run 6‖‖	19.84	16.96	2.88	1.5

* All values presented here are from forward simulations.

‡ Root-mean-square deviation between the NMR structure of the hairpin and the structure at the end of the free energy simulation.

§ Experimental data (Varani et al., 1991).

¶ After 10 ps of equilibration.

‖ A 100-ps simulation with a 10-Å uniform cutoff.

** A 100-ps simulation with a 10-Å cutoff for solute-water and water-water interactions.

‡‡ A 100-ps simulation with a 12-Å cutoff for solute-water and water-water interactions.

§§ A 50-ps simulation deriving free energy components with a 10-Å cutoff for solute-water and water-water interactions.

¶¶ A 100-ps simulation deriving free energy components with a 10-Å cutoff for solute-water and water-water interactions.

‖‖ A 100-ps simulation deriving free energy components with a 10-Å uniform cutoff.

Because of a lack of understanding of the effects of periodic boundary conditions and nonbonded cutoff radii on the calculated free energies of nucleic acid systems, we decided to run FEP simulations with different cutoff radii and box dimensions. We present here results of six forward simulations (see Table 3). In the first simulation we used a 10-Å uniform cutoff radius for all of the nonbonded interactions and ran the simulation up to 100 ps. In the second simulation we used a 10-Å cutoff radius for solute-water and water-water interactions but have included all of the solute-solute interactions (including the sodium ions) and ran up to 100 ps. In the third simulation we used a 12-Å cutoff radius for solute-water and water-water interactions and ran up to 100 ps. In the fourth and fifth simulations we carried out free energy component analysis with the slow growth method up to 50 and 100 ps, respectively, with a 10-Å cutoff radius for solute-water and water-water interactions. In the sixth simulation we employed a 10-Å uniform cutoff radius and computed free energy components in a 100-ps simulation. The root-mean-square deviation of the structures at the end of free energy simulations from the NMR structure of the hairpin for each simulation are given in Table 3. The structures at the end of the free energy simulations are in the vicinity of the experimental structure. In most cases the deviation of the structure from the equilibrated structure during free energy simulations is dominated by the disruption of the hydrogen bond between N^4 of C_7 and the phosphate group interacting with it (see Table 4). In all of our simulations the backbone of the RNA hairpin remained close to the experimental structure. This could be largely due to the attractive electrostatic interactions between all the 2′ hydroxyl hydrogens and the 3′ oxygens, which is consistent with our observation that these interactions are maintained throughout all of our simulations. The 2′ hydroxyl hydrogens in the NMR structure were model built because they were not observed experimentally (Wimberly et al., 1991; Varani et al., 1991). The resultant structure had all 12 HO2′···O3′ distances between 2.62 and 2.74 Å. In all of our simulations these 12 HO2′···O3′ distances

TABLE 4 The distances in angstroms between N^4 of cytosine (C_7) and oxygen atom of the phosphate group between U_5 and U_6

Conformer	$N^4/O^4 \cdots OB^*$
NMR	2.67
Eqb‡	2.91
Run 1	5.58
Run 2	6.50
Run 3	6.13
Run 4	3.87
Run 5	10.87
Run 6	6.40

* The phosphate oxygen closest to the N^4 amino group.
‡ After 10 ps of equilibration.

were maintained, and the distances in simulation 5 were between 2.4 and 2.9. The strong electrostatic interactions between HO2′ and O3′ were also observed to occur in the work presented by Rao and Kollman (1986). They stated that the strong interactions between the HO2′ and O3′ could mainly be due to the lack of explicit salt and solvent. However, we see that these interactions also occur in explicit salt and solvent conditions.

Free energy data

The computed free energy changes for the cytosine→uracil mutation in the hairpin and in the A-form RNA with the looped-out cytosine are given in Table 3. The experimental free energy change given in Table 3 is from the thermal denaturation studies of Varani et al. (1991). The differential free energy change ($\Delta\Delta G$) in the simulations where all solute-solute interactions are included is about the same (Runs 2–5), whereas in the simulation where a 10-Å uniform cutoff radius was employed (Runs 1 and 6) the differential free energy change is about twice that of all the other runs. The average differential free energy ($\Delta\Delta G_{av}$) change for all the simulations is 1.54 ± 0.79 kcal/mol, which is in good agreement with the free energy change estimated from thermal denaturation studies carried out by Varani et al. (1991). The results from the simulations with a 10-Å cutoff for solute-solvent and solvent-solvent interactions (Runs 2, 4, and 5) are similar to those from a simulation with a 12-Å cutoff (Run 3) with a similar nonbonded interaction treatment.

Free energy decomposition

We carried out free energy component analysis to understand the nature of interactions in the loop area of the hairpin, using free energy simulations with the slow growth method (Runs 4, 5, and 6 given in Table 3). During these simulations we extracted the free energy contributions from both the local and nonlocal elements interacting with the cytosine being mutated. The free energy components (van der Waals + electrostatic) from all the interacting bases, sodium ions, and water for the hairpin are given in Fig. 3,

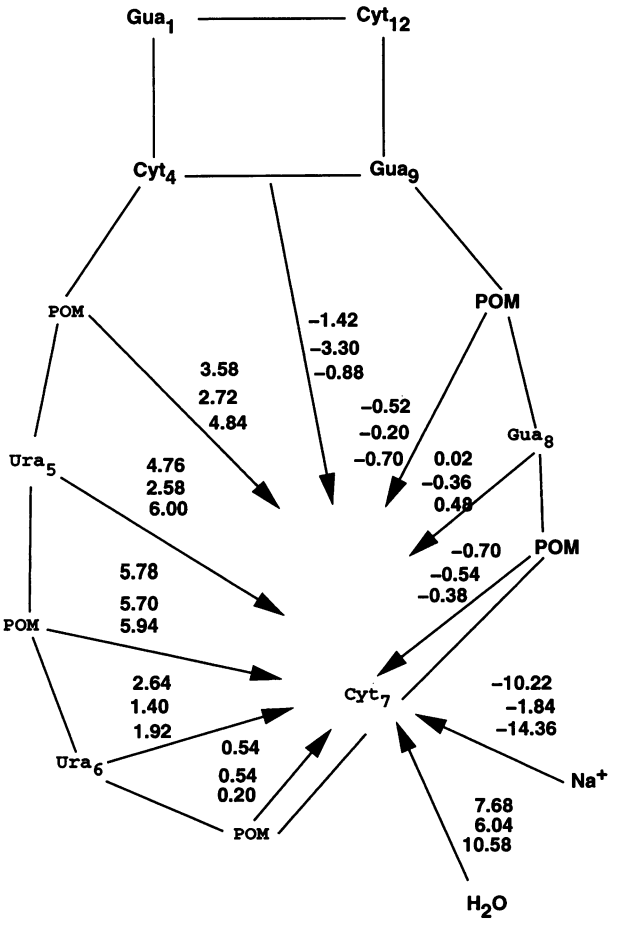

FIGURE 3 The free energy components in kcal/mol of $C_7 \rightarrow U_7$ mutation in the hairpin GGAC(UUCG)GUCC. The components shown here are a sum of the contributions from the base and the sugar atoms. There are three values shown above each arrow, indicating the contributions of each group toward the cytosine→uracil mutation for Runs 4, 5, and 6, respectively. The values shown at the top are from a 50-ps forward run (Run 4), and the values shown in the middle (Run 5) are from a 100-ps forward run with a 10-Å cutoff radius for solute-water and water-water interactions only. The values shown at the bottom (Run 6) are from a 100-ps forward run with 10-Å uniform cutoff radius. The interlinking phosphate group is labeled POM.

and those for the A-form RNA with looped-out cytosine are given in Fig. 4. The free energy components of the differential destabilization of the RNA hairpin with uracil relative to cytosine are given in Fig. 5. The components given in Fig. 5 are the differences of the components of the hairpin (Fig. 3) from those of the single-strand A-form (Fig. 4) ($\Delta\Delta G = \Delta G_{hairpin} - \Delta G_{A\text{-}form}$). The components from the nucleic acid segments are similar in both the 50-ps and 100-ps simulations. However, the contributions from sodium ions and water are very different in these three simulations, indicating that these components are coupled and hence are sensitive to the simulation conditions. In a mixed molecular weight system (macromolecule + solvent molecules + salt atoms) salt and solvent molecules undergo large-scale motions relative to the macromolecule in subnanosecond time scales and hence the greater sensitivity of

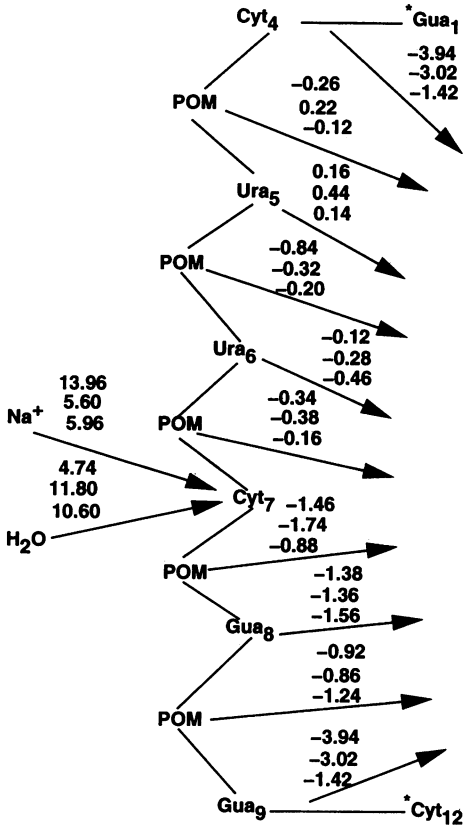

FIGURE 4 The free energy components in kcal/mol of $C_7 \rightarrow U_7$ mutation in the looped-out cytosine in the A-form single-strand GGACUUCG-GUCC. The components shown here are a sum of the contributions from the base and the sugar atoms. There are three values shown above each arrow, indicating the contributions of each group toward the cytosine→uracil mutation for Runs 4, 5, and 6, respectively. The values shown at the top are from a 50-ps forward run (Run 4), and the values shown in the middle are from a 100-ps forward run (Run 5) with a 10-Å cutoff radius for solute-water and water-water interactions only. The values shown at the bottom are from a 100-ps forward run (Run 6) with 10-Å uniform cutoff radius. The interlinking phosphate group is labeled POM. *Gua_1-Cyt_4 and Gua_9-Cyt_{12} are treated as one group in our calculations and hence their contributions are identical.

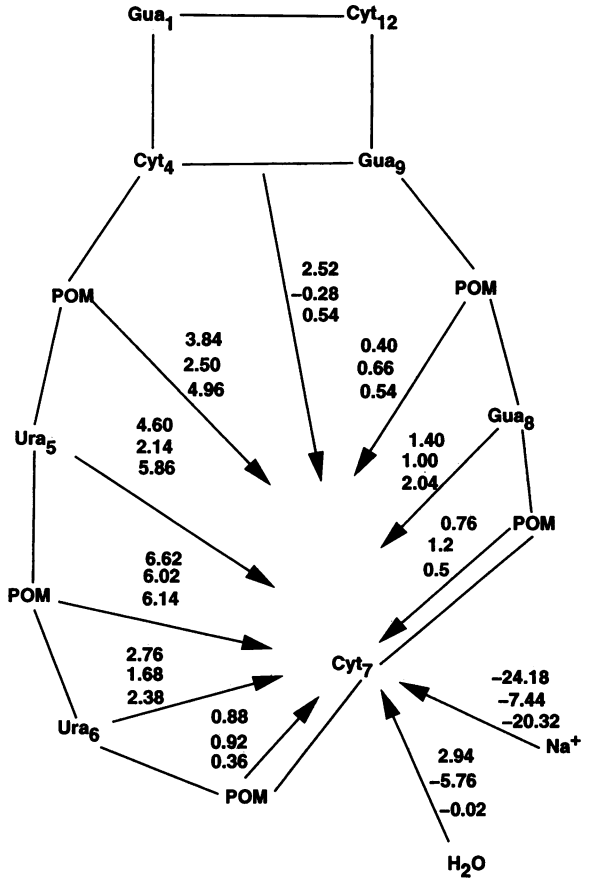

FIGURE 5 The free energy components in kcal/mol of the differential destabilization ($\Delta\Delta G$) of $C_7 \rightarrow U_7$ mutation in the RNA hairpin relative to the single-strand A-form. The components shown here are differences of the components of the hairpin loop (Fig. 3) from those of the single-strand A-form (Fig. 4) ($\Delta\Delta G = \Delta G_{\text{hairpin}} - \Delta G_{\text{A-form}}$). Three values are shown above each arrow, indicating the contributions of each group toward the cytosine→uracil mutation for Runs 4, 5, and 6, respectively. The values shown at the top are from a 50-ps forward run (Run 4), and the values shown in the middle are from a 100-ps forward run (Run 5) with a 10-Å cutoff radius for solute-water and water-water interactions only. The values shown at the bottom are from a 100-ps forward run (Run 6) with 10-Å uniform cutoff radius. The interlinking phosphate group is labeled POM.

the free energy components for the salt and solvent components relative to the macromolecular components. This shows that the free energy components of rapidly fluctuating parts (salt and solvent) of a system exhibit path dependence, because a particular path represents only a part of the phase space accessible to the whole system.

There are three major and two minor nucleic acid contributions and two major environmental (salt and solvent) contributions to the destabilization of the hairpin when cytosine is replaced by uracil (see Fig. 5). The most significant nucleic acid contribution is from the phosphate group interacting with the cytosine residue, by 6.3 kcal/mol, which is in accordance with the conclusions of Varani et al. (1991). (The average values given here for the free energy components are based on the values from the three simulations (Runs 4–6, Table 3).) The other two major nucleic acid contributions are from the adjacent moieties, U_5 and the phosphate group preceding it, contributing 3.8 and 4.2 kcal/mol, respectively. The contribution from U_5 could be due to the combination of unfavorable electrostatic interactions with U_7 with the conversion of the amino group to the keto group at C^4, due to the loss of favorable interactions caused by the reorientation of the cytosine/uracil moiety, and disruption of the hydrogen bonding interactions with G_8. The two minor nucleic acid contributions are from U_6 and G_8 with 2.3 kcal/mol and 1.5 kcal/mol, respectively. The contribution from G_8 could be due to the disruption of the hydrogen bonding interactions between U_5 and G_8 caused by the reorientation of the cytosine/uracil moiety.

The two major environmental contributions are from sodium ions and water molecules surrounding the hairpin, contributing −17.3 kcal/mol and −1.0 kcal/mol, respectively. It is clear that there is significant cancellation be-

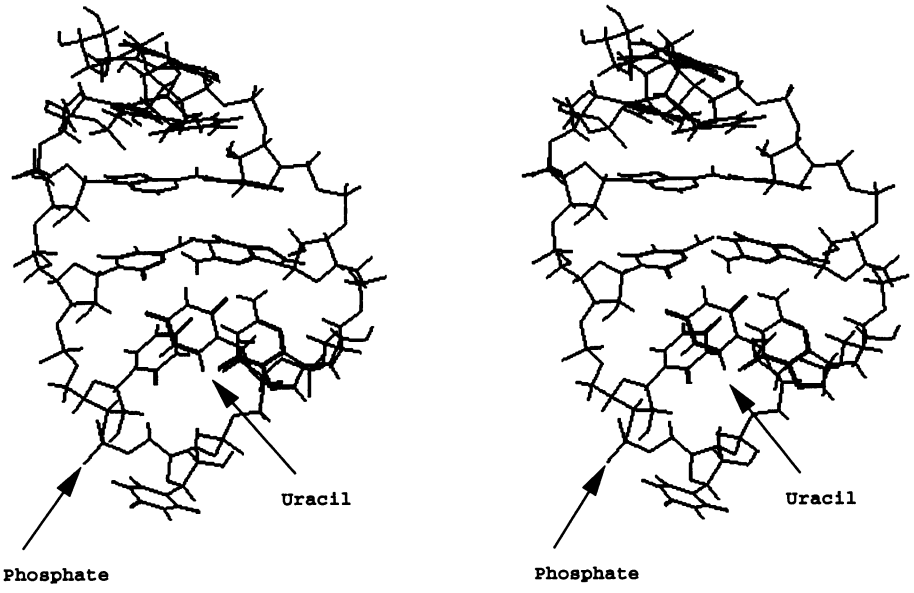

FIGURE 6 The structure of the RNA hairpin GGAC(UUCG)GUCC after 100 ps of free energy simulation with a 10-Å cutoff for the solute-water and water-water interactions is shown in stereo. The phosphate residue linking U_5-U_6 and the uracil (U_7) are labeled.

tween the sodium and the nucleic acid free energy differences for the cytosine→uracil mutation. It is also clear that the salt and solvent components are intricately coupled, hence the sensitivity to the simulation conditions. However, the components from the RNA segments are quite similar in all three simulations (Runs 4–6), indicating that the structural changes associated with cytosine→uracil mutation should have taken similar paths. The experimental evidence shows that the native and its mutant are very similar in structure (Varani et al., 1991). Thus the nucleotide components are useful in gaining insight into the molecular interactions responsible for the thermodynamic stability of the hairpin. Our calculations show that the phosphate moiety is the single largest nucleic acid contribution to the preference for cytosine over uracil loop, consistent with the implication of the experimental structure (Varani et al., 1991). In addition, we show that the residues in the loop region also contribute significantly toward the stability. The positive free energy contributions from many of these residues are largely canceled by the negative free energy contribution from the sodium ions. Despite these large cancellation effects it is encouraging to see that the key contributions from the loop region are intuitively reasonable and the total free energy change is in good agreement with the experiment.

CONCLUSIONS

We presented here the first free-energy calculations for a RNA system with full molecular representation of solvent and counter-ions. We have presented here a methodology to simulate solvated nucleic acid systems. This methodology has allowed us to maintain structural stability and obtain reliable free energy data. Also presented here are our findings about the nature of interactions that are responsible for the stability of rC(UUCG)G loops. Our calculations show that the hairpin with a UUCG loop is more stable than the hairpin with a UUUG loop by 1.5 ± 0.79 kcal/mol. The agreement with the experimental value of 1.5 kcal/mol (Varani et al., 1991), although perhaps only fortuitously quantitative, is encouraging in that it is of the same sign and magnitude as the experimental data. The free energy decomposition analysis reveals that this stability stems from the strong electrostatic interactions between the phosphate group and the C^4 amino group. This is obvious from the free energy contribution of ~6.0 kcal/mol from the phosphate group in stabilizing the loop with a UUCG sequence over that with a UUUG sequence (see Fig. 5). The other significant contributions are from the neighboring residues U_5, U_6, and G_8 and the phosphate moieties linking C_4-U_5 and U_6-C_7/U_7. The salt and solvent components are different under different simulation conditions, indicating that these components are coupled and thus exhibiting path dependence. The components from the RNA segments are quite similar in all three simulations (Runs 4–6), indicating that the structural changes associated with cytosine→uracil mutation are similar, which is consistent with the experimental observation that the native and the mutant hairpin loops are very similar in structure (Varani et al., 1991).

The unusual stability of the hairpin loop could be due to a combination of all of the specific interactions mentioned above and the favorable electrostatic interactions between the 2′ hydroxyl hydrogen and the O3′, which keeps the backbone in the hairpin and the loop region fairly rigid during the course of all of our simulations. The lack of 2′ hydroxyl in the deoxy ribose analog could be one of the factors responsible for the flexibility of the loop observed by James and Tinoco (1993). We wish to investigate the differences in the nature of dynamics between the ribose and deoxy ribose analogs of the hairpin GGAC(UUCG)-GUCC in our future studies to understand further the basis for the unusual stability of the C(UUCG)G hairpin loops.

The results of our calculations actually "illustrate" the opposing arguments of both Mark and van Gunsteren (1994) and Boresch et al. (1994). First, the calculations do show the greater sensitivity of components than the total free energy to simulation protocol/length, particularly those involving the charged counter-ions (an illustration of the path dependence). Second, the component analysis does support its usefulness in gaining insight into the nature of the stability of the hairpin loop. The calculations show that the phosphate group and the uracil (U_5) are the two largest factors that stabilize cytosine over uracil at the seventh position in the hairpin. This result strongly predicts that either neutralizing this key phosphate moiety by substituting a methylphosphonate or by making a stereospecific (pro-R) phosphorothiolate at that position would significantly reduce the stability of the loop. In the case of the methylphosphonate, one would expect that the stereoisomer that replaces the pro-R oxygen with a methyl would have a much larger reduction in stability than the one that replaces the anionic oxygen with a neutral one.

We acknowledge the support for the current research from the National Institutes of Health (CA-25644 to P.A.K.) and the use of the University of California, San Francisco Computer Graphics Lab (supported by RR-1081; R. Langridge, principal investigator). We also acknowledge the computing support provided by the National Science Foundation's Pittsburgh Supercomputer Center through the Grand Challenge award MCA93S017P to PAK and by the National Science Foundation's San Diego Supercomputer Center. The computing support provided by Wyeth-Ayerst Research (a pharmaceutical research division of American Home Products) is also acknowledged.

REFERENCES

Arnott, S., D. W. L. Hukins, S. D. Dover, W. Fuller, and A. R. Hodgson. 1973. Structures of synthetic polynucleotides in the A-RNA and A′-RNA conformations: x-ray diffraction analyses of the molecular conformations of polyadenylic acid · polyuridylic acid and polyinosinic acid · polycytidylic acid. *J. Mol. Biol.* 81:107–122.

Boresch, S., G. Archontis, and M. Karplus. 1994. Free energy simulations: the meaning of the individual contributions from a component analysis. *Proteins Struct. Funct. Genet.* 20:25–33.

Boresch, S., and M. Karplus. 1995. The meaning of component analysis-decomposition of the free energy in terms of specific interactions. *J. Mol. Biol.* 254:801–807.

Brady, G. P., and K. A. Sharp. 1995. Decomposition of interaction free energies in proteins and other complex systems. *J. Mol. Biol.* 254: 77–85.

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

James, J. K., and I. Tinoco, Jr. 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.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Kollman, P. A. 1993. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395–2417.

Mark, A. E., and W. F. van Gunsteren. 1994. Decomposition of the free energy of a system in terms of specific interactions: implications for theoretical and experimental studies. *J. Mol. Biol.* 240:167–176.

McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, England.

Nilsson, L., A. Åhgren-Stalhandske, A. S. Sjögren, S. Hahne, and B-M. Sjöberg. 1990. 3-Dimensional model and molecular dynamics simulation of the active site of the self-splicing intervening sequence of the bacteriophage t4 NRDB messenger RNA. *Biochemistry.* 29: 10317–10322.

Pearlman, D. A., D. A. Case, J. C. Caldwell, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman. 1991. AMBER(4.0). University of California, San Francisco.

Rao, S. N., and P. A. Kollman. 1986. Conformations of the 8-methylated and unmethylated ribohexamer r(CGCGCG)$_2$. *J. Am. Chem. Soc.* 108: 3048–3053.

Singh, S. B., Ajay, D. E. Wemmer, and P. A. Kollman. 1994. Relative binding affinities of distamycin and its analog to d(CGCAAGTTGGC) · d(GCCAACTTGCG): comparison of simulation results with experiment. *Proc. Natl. Acad. Sci. USA.* 91:7673–7677.

Singh, U. C., F. Brown, P. Bash, and P. A. Kollman. 1987. An approach to the application of free energy perturbation methods using molecular dynamics: applications to the transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, glycine→alanine, and alanine→phenylalanine in aqueous solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the gas phase. *J. Am. Chem. Soc.* 109:1607–1614.

Singh, S. B., D. A. Pearlman, and P. A. Kollman. 1993. Free energy component analysis: application to the "Z-phobicity" of A · T base pairs. *J. Biomol. Struct. Dyn.* 11:303–311.

Smith, P. E., and W. F. van Gunsteren. 1994. When are free energy components meaningful? *J. Phys. Chem.* 98:13735–13740.

Stryer, L. 1988. Biochemistry, 3rd Ed. W. H. Freeman and Company, New York.

van Gunsteren, W. F., and H. J. C. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34: 1311–1327.

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

Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

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

Yunyu, S., A. E. Mark, C. X. Wang, F. H. Huang, H. J. C. Berendsen, and W. F. van Gunsteren. 1993. Can the stability of protein mutants be predicted by free energy calculations? *Protein Eng.* 6:289–295.

Zaug, A. J., and T. R. Cech. 1986. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science.* 231:431–475.

Zichi, D. A. 1995. Molecular dynamics of RNA with the OPLS force field: aqueous simulation of a hairpin containing a tetranucleotide loop. *J. Am. Chem. Soc.* 117:2957–2969.