

Effects of Base Substitutions in an RNA Hairpin from Molecular Dynamics and Free Energy Simulations

Joanna Sarzynska,* Lennart Nilsson,[†] and Tadeusz Kulinski*

*Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; and [†]Center for Structural Biochemistry, Department of Biosciences at NOVUM, Karolinska Institutet, Huddinge, Sweden

ABSTRACT Contributions of individual interactions in the GCGCAAGCC hairpin containing a GCAA tetraloop were studied by computer simulations using base substitutions. The G in the first tetraloop position was replaced by inosine (I) or adenosine (A), and the G in the C-G basepair closing the tetraloop was replaced by I. These substitutions eliminate particular hydrogen bonds proposed in the nuclear magnetic resonance model of the GCAA tetraloop. Molecular dynamics simulations of the GCAA tetraloop in aqueous solvent displayed a well-defined hydrogen pattern between the first and last loop nucleotides (G and A) stabilized by a bridging water molecule. Substitution of G→I in the basepair closing the tetraloop did not significantly influence the loop structure and dynamics. The ICAA loop maintained the overall structure, but displayed variation in the hydrogen-bond network within the tetraloop itself. Molecular dynamics simulations of the ACAA loop led to conformational heterogeneity of the resulting structures. Changes of hairpin formation free energy associated with substitutions of individual bases were calculated by the free energy perturbation method. The calculated decrease of the hairpin stability upon G→I substitution in the C-G basepair closing the tetraloop was in good agreement with experimental thermodynamic data. Our theoretical estimates for G→I and G→A mutations located in the tetraloop suggest larger loop destabilization than corresponding experimental results. The extent of conformational sampling of the structures resulting from base substitutions and its impact on the calculated free energy was discussed.

INTRODUCTION

RNA can fold in huge numbers of tertiary structures, which are responsible for its diverse functions in cells. In the attempts to understand how RNA folds considerable attention is given to the determination of structural motifs associated with specific sequences and the basis of their thermodynamic stability. It was shown that loops with GNRA and UNCG sequences (where *N* is any ribonucleotide, *R* is purine) are more stable than random tetraloops (Antao and Tinoco, Jr., 1992; Dale et al., 2000).

In this work we focus on RNA hairpins with GNRA tetraloops. These hairpins are abundant structure elements in natural RNAs, and ~50% of rRNA tetraloops have the GNRA motif (Proctor et al., 2002; Woese et al., 1990). They are often involved in tertiary contacts in larger RNA structures, serve as recognition sites for protein binding, and likely initiate folding events (Cate et al., 1996; Pley et al., 1994a,b; Correll et al., 1999). The solution structures of GNRA tetraloops characterized by nuclear magnetic resonance (NMR) show that they have well-defined interactions which include hydrogen bonds between the G and A bases, between the G base and the phosphate oxygen of the A nucleotide, and stacking between the third and fourth loop bases (Jucker et al., 1996).

The free energy contributions of various interactions within a GCAA tetraloop have been studied experimentally using substitutions of functional groups (SantaLucia, Jr.

et al., 1992): guanosine was replaced by 2'-deoxyguanosine, 2-aminopurine, inosine (I), or adenosine, and individual adenines by purine. These substitutions eliminate particular hydrogen bonds proposed in NMR models of the GNRA loop. Thermodynamic measurements showed that all substitutions made in the tetraloop destabilize the hairpin by a relatively small amount (<1 kcal/mol). Experimental data for loop structures containing these substitutions are not available and the thermodynamic data was rationalized based on conformational analogy between modified and reference GCAA tetraloops. In particular, substitution of G to I in the loop destabilized the hairpin less than the same substitution in the C-G basepair closing the tetraloop, although the deleted amino group of G was involved in two hydrogen bonds in the tetraloop and in a single hydrogen bond in the closing basepair. SantaLucia and co-workers suggested that the hydrogen-bond free energy depends on the structural context (SantaLucia, Jr. et al., 1992). Jucker and co-workers, on the basis of NMR data, proposed that heterogeneity in the hydrogen-bond network within the GCAA tetraloop, local structural rearrangements, and development of water-mediated interactions may be responsible for the small energy changes observed upon deletion of individual hydrogen bonds in the loop (Jucker et al., 1996).

To obtain additional insight concerning the origin of experimentally observed free energy differences between the GCAA tetraloop and tetraloops with substitutions we have performed theoretical calculations. Molecular dynamics in conjunction with free energy calculations provide a link between details of structure and dynamics at atomic level and difference in thermodynamic stability upon substitutions in an RNA hairpin. Free energy calculations

Submitted January 24, 2003, and accepted for publication July 3, 2003.

Address reprint requests to Tadeusz Kulinski, E-mail: tadkul@ibch.poznan.pl.

© 2003 by the Biophysical Society

0006-3495/03/12/3445/15 \$2.00

employing molecular dynamics have been successfully applied to the study of the relative stability of proteins (Sun et al., 1996), DNA duplexes (Florián J. et al., 2000; Cubero et al., 2001), and RNA tetraloops (Singh and Kollman, 1996; Williams and Hall, 2000a).

In this article we focused on modifications altering the guanosine located at the first position in the loop, which is involved in the largest number of hydrogen bonds in the native GCAA tetraloop, and the G from the C-G basepair closing the tetraloop. The first part of our work addresses the question how G→I and G→A substitutions affect the conformation and dynamics of the GGCGCAAGCC hairpin. In the second part, we have evaluated differences in the hairpin formation free energy due to mutations using the free energy perturbation method.

METHODS

Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed using the CHARMM simulation program, versions 27 and 28 (Brooks et al., 1983; MacKerell et al., 1998), with the CHARMM27 all-atom force field for nucleic acids (Foloppe and MacKerell, 2000) and inosine (I) (Pastor et al., 1999).

Starting coordinates were taken from the NMR structure of a hairpin with a GCAA tetraloop (PDB code 1zih, model 1; Jucker et al., 1996). Substitutions G→I and G→A were done by changing only functional groups of the G and keeping coordinates of the ring atoms. Structures were gently minimized with restraints in vacuum (three rounds of 50 steepest-descent steps using harmonic restraints with 15, 10, and 5 kcal × mol⁻¹ × Å⁻² force constant respectively on all atoms). Then the hairpins were solvated in a 40 × 35 × 35 Å³ water box containing TIP3P water (Jorgensen et al., 1983). Water molecules closer than 2.0 Å from any RNA atom were removed. To neutralize the system, nine Na⁺ ions were added, which corresponds to an Na⁺ concentration of 0.3 M. Ions were placed on the basis of the electrostatic energy of the water oxygens. Finally, the whole system was minimized, usually for 200 steps of steepest descent (SD) with harmonic restraints on RNA (force constant 20 kcal × mol⁻¹ × Å⁻²) and 200 steps of SD without any restraints.

MD simulations started with heating the system from 50 to 300 K in 6 ps followed by 4 ps of equilibration, after which trajectories were collected. The initial 200 ps were considered as further equilibration time and analysis was done on the following parts of trajectories, unless a specific period is mentioned. MD simulations were run in the NVE ensemble, using periodic boundary conditions, a 2-fs timestep, and SHAKE algorithm (Ryckaert et al., 1977) to constrain all bonds to hydrogens. The nonbonded interactions were smoothly shifted to zero at a 12.0 Å cutoff by the atom-based truncation method with a force-shifting function (Steinbach and Brooks, 1994). The nonbonded pair list was generated using a 13 Å cutoff and updated when any atom had moved 0.5 Å or more. Extensive simulations of DNA duplexes have shown this protocol to yield trajectories that are very similar to trajectories obtained using Ewald summation, in terms of both structural stability and flexibility (Norberg and Nilsson, 2000). For each tetraloop sequence, several independent simulations were performed. These repeated simulations differed in initial velocities for all atoms and initial positions of the sodium ions.

MD simulations of guanosine and single-stranded RNA trimer and pentamer fragments as models of unfolded RNA were performed using similar protocol as for hairpins, with initial structures that were either extracted from the hairpin or built as a canonical A-RNA form.

Hydrogen-bonding during MD simulations was characterized by the

average number of hydrogen bonds (N_H), calculated by dividing the total number of hydrogen bonds in all trajectory frames by the number of frames using the following criteria: distance hydrogen-acceptor <2.5 Å; or angle donor-hydrogen-acceptor >135°; only those hydrogen bonds with $N_H > 0.1$ were included in the analysis. The average number of contacts with sodium ions (N_c) was defined as the total number of contacts, defined by the criterion that the sodium ion should be <2.7 Å from the RNA atom, formed between sodium ions and a given RNA atom in all trajectory frames divided by the number of frames.

Free energy calculations

Free energies were calculated by applying free energy perturbation (FEP) method (Kollman, 1993; Pearlman and Govinda, 1998; Mark, 1998). To calculate free energy differences between the reference GCAA hairpin and the mutated hairpins we used a thermodynamic cycle and carried out free energy simulations for alchemical transformation in which one nucleic acid base was transformed to another, either in the context of the hairpin or in an unfolded state represented by a short nucleotide fragment (Straatsma and McCammon, 1992; Jorgensen, 1998).

The alchemical transformations were realized by running the PERT module (B. R. Brooks, unpublished results) in CHARMM that uses the single topology approach, in which every atom in the initial state has a counterpart in the final state. To accomplish this, dummy atoms are introduced for atoms that exist in one state and have no counterpart in the other state. In our calculations dummy atoms were identical to real atoms except that they have no nonbonded (Coulombic or van der Waals) interactions. Covalent terms associated with these dummy atoms were not changed during the course of the mutation.

The FEP simulations were carried out by defining a series of nonphysical intermediate points between the initial and final state (reference and mutated hairpin) and then calculating the sum of free energy change computed for each step. Each PERT simulation was run for 2 ps of equilibration + 2 ps of production using double-wide sampling in 101 windows unless different equilibration or production time was specified.

For each type of mutation we performed several independent free energy simulations that differed in initial velocities or starting conformation, taken from different snapshots of the same trajectory or from independent trajectories from previously performed dynamics of the GCAA tetraloop and models of unfolded RNA. Several models of a single-strand RNA were examined in the calculation of free energy changes due to mutation in unfolded RNA. For each mutation at least one simulation in the backward direction was performed. As a starting structure for the backward simulation, the final structure of the forward simulation after additional 10 ps of equilibration was used. The FEP simulations were run in the NPT ensemble at 1 atm and 300 K using the extended system constant pressure (Feller et al., 1995) and temperature algorithm, timestep 1 fs, and the SHAKE algorithm applied only to the water hydrogens. The temperature was controlled by the Hoover method with the formal mass of the thermal piston 500 kcal × mol⁻¹ × ps² (Hoover, 1985).

For comparison, the free energy change due to G4→A4 mutation was also calculated using the particle mesh Ewald (PME) method for treatment of electrostatic interactions (Darden et al., 1993). PME calculations were performed using a real space cutoff of 12 Å. The number of grid points was set to 32, 32, and 64 in the x-, y-, and z-directions, respectively. A sixth-order smoothing spline was used and the screening parameter (κ) was optimized to 0.32. Additionally, free energy calculations were performed in a water sphere of radius 19 Å and 13 Å for the hairpin and unfolded state respectively, with stochastic boundary conditions (Brooks III and Karplus, 1983) applying the Langevin dynamics algorithm with friction coefficient 50.0 ps⁻¹ for the water oxygen atoms in the buffer region outside the inner sphere of radius 2 Å shorter than outer sphere. From our earlier experience, simulations of the RNA hairpin in a water sphere give similar results, in terms of structural dynamics, to simulations in a water box (Sarzynska et al.,

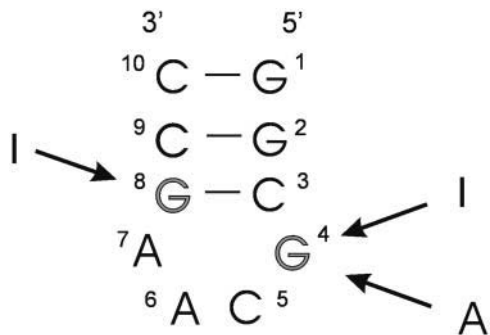


FIGURE 1 Secondary structure of the simulated RNA hairpin with GCAA tetraloop. The substitutions made are indicated by arrows.

2000), but since the system in the water sphere contains fewer atoms (here 4803 atoms in water box and 2815 atoms in water sphere for the RNA hairpin) it can be run faster.

The entropy of the RNA hairpin was estimated by quasi-harmonic analysis (Andricioaei and Karplus, 2001) performed on MD trajectories.

RESULTS

Molecular dynamics simulations

The effect of base substitutions in a GCAA tetraloop on its conformation and dynamics has been examined using MD simulations. Our studies were performed on the GGCGCAAGCC hairpin and hairpins where the G in the tetraloop was replaced by inosine (I) or adenosine (A), and the G in the C-G basepair closing the tetraloop was replaced by I (Fig. 1). The nucleotide sequence of these hairpins was the same as in thermodynamic experiment (SantaLucia, Jr. et al., 1992) and did not include the first basepair in the stem

TABLE 1 Performed simulations of RNA hairpins and the average RMS deviations from initial structures

Tetraloop sequence	Trajectory name	RMSD* [†] [Å]	RMSD loop [Å]
GCAA	R_1 [‡]	1.50 ± 0.16	0.94 ± 0.12
	R_2 [§]	1.65 ± 0.27	1.04 ± 0.15
C(GCAA)I	I_8_1 [¶]	1.56 ± 0.18	0.91 ± 0.12
	I_4_1 [¶]	1.78 ± 0.27	1.37 ± 0.25
ICAA	I_4_2 [¶]	1.57 ± 0.19	0.95 ± 0.11
	A_4_1 [¶]	1.64 ± 0.17	0.90 ± 0.12
ACAA	A_4_2 [§]	2.40 ± 0.16	2.31 ± 0.14
	A_4_3 [§]	2.12 ± 0.27	1.98 ± 0.49
	A_4_4	2.22 ± 0.29	2.24 ± 0.45
	A_4_5 ^{**}	1.77 ± 0.18	1.28 ± 0.35

*All simulations are 1.4-ns long.

[†]RMSD was calculated with 1-ps resolution and for nonhydrogen atoms.

[‡]Initial positions of sodium ions inserted in whole water box.

[§]Sodium ions inserted in the layer 3.9–5.2 Å from RNA.

[¶]Sodium ions were inserted in whole water box, but position of two water molecules with the highest electrostatic energy omitted.

^{||}Sodium ions inserted at a distance from RNA >5.2 Å.

^{**}Initial coordinates for whole system were taken from the final coordinates of G4→A4 alchemical mutation.

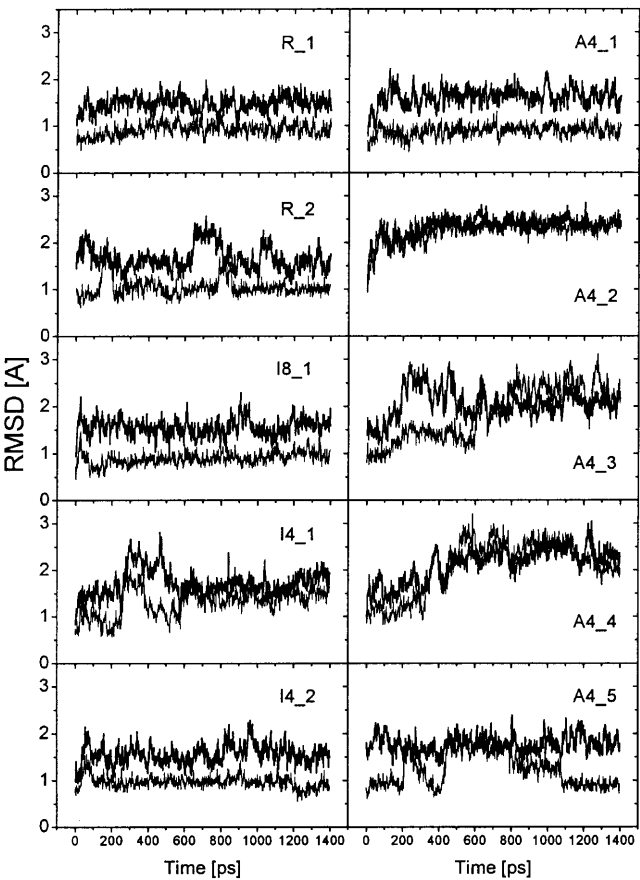


FIGURE 2 Time-course of the RMSD compared to the initial NMR-derived structure for the simulations presented in this work; whole molecule (bold lines) and tetraloop (nucleotides 4–7; thin lines).

of the NMR structure (Jucker et al., 1996). Simulations presented in this work are summarized in Table 1 and Fig. 2.

Hydrogen-bond network within tetraloops

GCAA and C(GCAA)I tetraloops

During MD simulations of hairpins with a GCAA tetraloop (R_1, R_2, and I_8_1), the root-mean-square deviation (RMSD) in reference to the initial NMR structure calculated for just the tetraloop was 0.9–1.0 Å (Table 1). In all trajectories the GCAA tetraloop converged toward a similar conformation with a characteristic pattern of hydrogen bonds between G4 and A7 (Fig. 3). In this structural motif the G4 base made three direct hydrogen bonds, (G4)N2-H...N7(A7), (G4)N2-H...O1P(A7), and (G4)N1-H...O1P(A7), as well as water-mediated hydrogen-bond interactions between (G4)N3 and (A7)N6-H (Fig. 4 A and Table 2 A). This pattern of hydrogen bonds required minor rearrangements of the initial NMR-derived structure. To make room to accommodate a bridging water molecule the (G4)N3...N6(A7) distance increased to the average value of

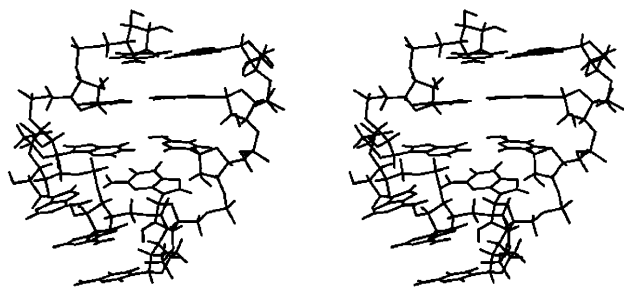


FIGURE 3 A stereoview of the structure of C(GCAA)I loop (structure averaged over the last 20 ps of I₈_1 trajectory and gently minimized).

5.6 Å. At the same time the distance and angle of the (G4)N1-H...O1P(A7) hydrogen bond came closer to the ideal values than in the initial structure. Water molecules involved in water-mediated interactions within the tetraloop displayed relatively long residence times (Table 3). The water molecule present within GCAA tetraloop was preferentially hydrogen-bonded to (A7)N6-H, (G4)N3, and (G4)O2'. Occasionally it could also make hydrogen bonds to (A6)N7 and (G8)N3. Fig. 5 A displays time series of selected distances within the GCAA tetraloop during simulation R_1. One can notice that since the water bridge between G4 and A7 was established, the (G4)N3...N6(A7) distance remained stable. We have observed a few exchanges of bridging water molecules but without reduction of the (G4)N3...N6(A7) distance. The rearrangement back to a conformation, where this distance was too short to accommodate the water, was not observed. The (G4)2'OH...N7(A6) hydrogen bond, proposed by Jucker et al. (1996), was rarely observed in the conformations with a water-mediated G4-A7 contact, even though it was present in fragments of trajectories R_2 and I₈_1 before the water bridge was established. After correlated transitions of backbone torsion angles ϵ_4 , ζ_4 , and α_5 from the ($-ac$, $-sc$, and $+ac$) range to the ($+ap$, $+sc$, and $+sc$) range (Saenger, 1984), (G4)2'OH served as a proton donor to (C5)O1P, and the amino group of the A6 base made dynamic hydrogen bonds to (G4)2'O and to (C5)O1P. The 2'OH groups of C5 and A7 were preferentially hydrogen-bonded to the phosphate oxygen of the next nucleotide from the 3' side.

In the course of simulations, transitions of the torsion angles were observed, mainly correlated rotations of ϵ_n , ζ_n , and α_{n+1} , which change the position of the P_{n+1} atom. This differs from the GCAA tetraloop simulation with the OPLS force field performed by Zichi (1995) where mainly

transitions of α , β , and γ were observed. Torsion angles ϵ_7 , ζ_7 , and α_8 were the most dynamic, with several correlated rotations from the ($+ac$, $-sc$, and $-sc$) range to the ($+ap$, $+sc$, and $-ac$) range. This, however, did not influence the hydrogen-bond network between nucleotides G4 and A7.

In the C(GCAA)I hairpin we did not see any significant effect of the substitution G→I made in the CG basepair

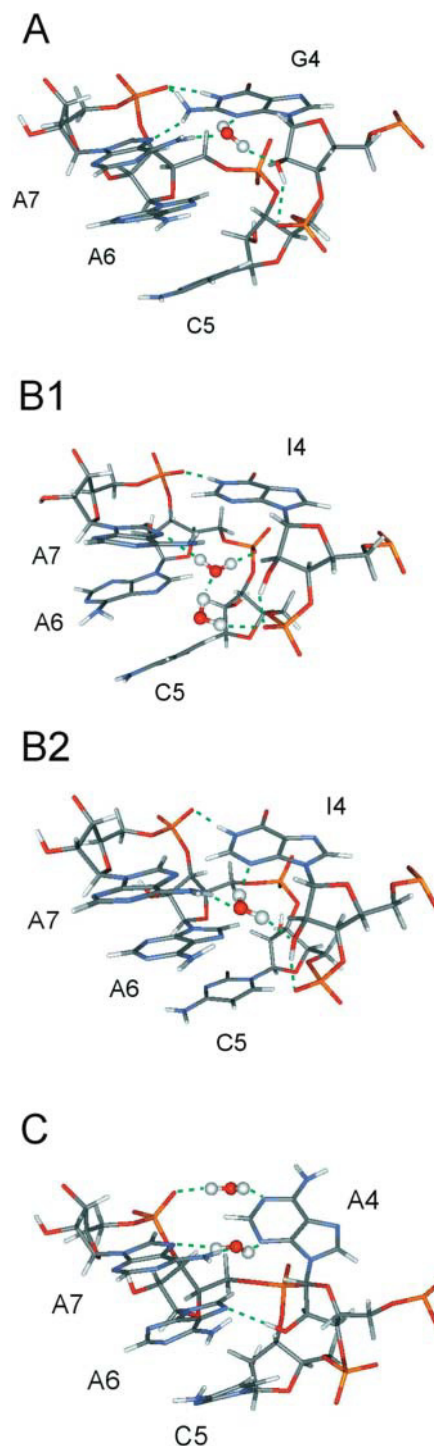


FIGURE 4 Hydrogen-bond network within tetraloops. Bridging water molecules in the loop are shown. (A) GCAA tetraloop (snapshot from simulation R_1 at 1400 ps); (B1, B2) ICAA tetraloop (snapshots from simulation I₄_1 at 1000 ps and I₄_2 at 720 ps, respectively); and (C) ACAA tetraloops (snapshot from simulation A₄_5 at 100 ps).

TABLE 2 Average number of hydrogen bonds within the GCAA and C(GCAA)I (A), ICAA (B), and ACAA (C) tetraloops

A.					
H bond donor...acceptor	R_1	R_2	I I_8_1		
(G4)N1-H...O1P(A7)	0.95	0.92	0.87		
(G4)N2-H...N7(A7)	0.70	0.68	0.72		
(G4)N2-H...O1P(A7)	0.88	0.83	0.86		
(G4)2'OH...O1P(C5)	0.82	0.67	0.63		
(C5)2'OH...O2P(A6)	0.95	0.71	0.96		
(A6)N6-H...O2'(G4)	0.25	0.25	0.27		
(A6)N6-H...O1P(C5)	0.29	0.19	0.11		
(A6)2'OH...O4', O5'(A7)	0.36	0.30	0.25		
(A7)2'OH...O1P(G8)	0.44	0.57	0.44		
B.					
	I_4_1				
H bond donor...acceptor	0–260 ps	260–670 ps	670–1400 ps	I_4_2	
(I4)N1-H...O1P(A7)	0.94		0.98	0.93	
(I4)2'OH...O1P(C5)		0.19	0.61	0.77	
(I4)2'OH...N7(A6)		0.21			
(C5)2'OH...O2P(A6)		0.27		0.94	
(C5)2'OH...O5'(A6)			0.19		
(A6)N6-H...O2'(I4)				0.34	
(A6)N6-H...O1P(C5)				0.20	
(A6)2'OH...O4', O5'(A7)				0.20	
(A6)2'OH...O2P(A7)		0.84			
(A7)N6-H...N3(I4)		0.73			
(A7)N6-H...O2'(I4)		0.11			
(A7)2'OH...O1P(G8)	0.38		0.25	0.41	
C.					
H bond donor...acceptor	A_4_1	A_4_2	A_4_3	A_4_4	A_4_5
(A4)N6-H...O1P(A6)			0.10		
(A4)N6-H...O1P(A7)		0.94	0.54		
(A4)N6-H...N7(G8)		0.53			
(A4)2'OH...O1P(C5)		0.83	0.68	0.10	
(A4)2'OH...O5'(C5)				0.14	
(A4)2'OH...N7(A6)	0.86				0.69
(C5)2'OH...O2P(A6)		0.67	0.71	0.10	
(C5)2'OH...O4', O5'(A6)				0.73	
(A6)N6-H...O2'(A4)		0.10	0.21		
(A6)N6-H...N3(A4)		0.11	0.12		
(A6)N6-H...N3(C5)				0.12	
(A6)N6-H...O1P(C5)			0.27		
(A6)2'OH...O4', O5'(A7)	0.21	0.97	0.85		
(A6)2'OH...O2P(A7)				0.74	0.45
(A7)N6-H...N3(A4)			0.25	0.20	
(A7)N6-H...O2'(A4)				0.23	
(A7)N6-H...O1P(A6)				0.41	
(A7)2'OH...O1P(G8)	0.32	0.29	0.61	0.63	

closing the tetraloop on the hydrogen-bond network within the tetraloop (Table 2 A and Fig. 3).

ICAA tetraloop

Substitution of G by I in the GCAA tetraloop eliminated two hydrogen bonds made by the amino group of the G base. Overall, the hairpin geometry remained stable during two 1.4-ns-long simulations, but the tetraloop itself displayed more variation in the hydrogen-bonding pattern than the GCAA tetraloop.

The conformation with the (I4)N1-H...O1P(A7) hydrogen bond and a bridging water molecule within the tetraloop was observed during the first 1220 ps of trajectory I₄₋₂ (Fig. 4 B2). The (I4)N3...N6(A7) distance fluctuated about the average value of 5.4 Å, but in contrast to the GCAA tetraloop, distance reduction events occurred. We observed distance-reduction events both with and without exchange of bridging water molecules after reestablishment of the water bridge. The bridging water molecule within the tetraloop was less often hydrogen-bonded to (I4)N3 than to (G4)N3 in the GCAA tetraloop (Table 3). Similarly as in the reference GCAA loop, (A6)N6-H made weak hydrogen bonds to O1P and O2' of I4 (Table 2 B). In simulation I₄₋₂ the tetraloop stayed close to the initial structure with RMSD ~ 0.95 Å.

A conformation with a direct (A7)N6-H...N3(I4) hydrogen bond was visible in the time range 260–670 ps of trajectory I₄₋₁ (Table 2). Formation of the (A7)N6-H...N3(I4) hydrogen bond was correlated with breaking the (I4)N1-H...O1P(A7) hydrogen bond (Fig. 5 B). The loss of the (I4)N1-H...O1P(A7) hydrogen bond was associated with a rotation of ϵ_6 , ζ_6 , and α_7 torsion angles from (–ac, –sc, and –sc) to (+ac, +sc, and +sc) that changed the position of (A7)O1P.

Another conformation was seen in the time range 670–1400 ps of trajectory I₄₋₁ (Fig. 4 B1). The I4 base formed a stable (I4)N1-H...O1P(A7) hydrogen bond and the average distance between (I4)N3 and (A7)N6 was 4.2 Å, which is too short to maintain a water-mediated contact between (I4)N3 and (A7)N6-H. For this conformation, water bridges involving two water molecules were found. The first water bridge was created between (A7)N7 and (A6)O1P. It is interesting that this bridge was established at ~600 ps, although the conformation with a direct (A7)N6-H...N3(I4) hydrogen bond was still present. This single water molecule stayed in its bridging position until the end of the simulation (i.e., for 800 ps). The second water bridge was made between the first bridging water and (C5)O1P or (A6)N7. Water molecules involved in the second bridge exchanged more quickly than in the first bridge. The fluctuations of the distance between the (I4)N3 and (A7)N7 atoms were larger than fluctuations of the same distance in GCAA tetraloops. The A6 did not participate in hydrogen-bonding to other nucleotides in any conformation during trajectory I₄₋₁.

ACAA tetraloop

Fig. 2 presents time series of RMSD for five simulations of the hairpin with the ACAA tetraloop. The simulations started from coordinates for GCAA tetraloop where G was replaced by A; only simulation A₄₋₅ started from coordinates taken from the final structure after performing the G→A alchemical mutation. Simulations A₄₋₁ and A₄₋₅ produced stable 1.4-ns trajectories with average RMSD from the GCAA-derived initial structure, calculated for the tetraloop,

TABLE 3 Hydrogen bonds made by long-lived water molecules with the tetraloop

				Acceptors/Donors												
Tetraloop/ simulation	Water molecule	Period time [ps]	Residence time* [ps]	G4/I4/A4				A6			A7			G8		
				O2'	N3	N1	N7	N6-H	O1P	N7	N6-H	O1P	N7	N6-H	N3	N7
GCAA																
R_1	W96	150–680	530	0.38	0.82										0.93	
	W1406	680–970	290	0.44	0.80					0.21					0.90	
R_2	W688	1040–1400	360		0.82										0.96	0.16
	W996	210–700	490	0.21	0.81					0.40			0.14		0.87	
	W533	900–1400	500	0.45	0.78					0.12					0.95	0.13
C(GCAA)I																
I_8_1	W1333	620–930	310	0.37	0.85					0.22					0.94	
	W1151	930–1250	320	0.37	0.79								0.10		0.93	
ICAA																
I_4_1	W29	20–220	200	0.36	0.27					0.41	0.27		0.14	0.77		
	W412	600–1400	800						0.96				0.70			
I_4_2	W1189	80–510	430	0.47	0.54								0.27	0.93		
	W51	510–930	420	0.56	0.60								0.18	0.86		
	W1172	980–1210	230	0.41	0.56								0.27	0.93	0.20	
ACAA																
A_4_1	W136	250–1400	1150		0.93								0.93	0.69		
A_4_2	W144	90–1400	1310			0.89						0.13	0.72	0.76		
	W657	10–270	260				0.75	0.51	0.14							
A_4_3	W2	270–790	520				0.86					0.78				
	W390	800–1400	600				0.86					0.84				
	W17	0–1400	1400			0.61						0.89	0.22			
	W794	150–590	440			0.66		0.20				0.90				0.24
A_4_4	W754	480–830	350		0.25				0.61				0.46	0.30		0.29
	W1361	720–1160	440	0.37					0.91				0.31	0.42		
A_4_5	W1476	940–1400	460		0.5				0.62				0.61			
	W1026	0–1400	1400	0.22	0.85								0.62	0.43		

*Residence time is a time when the water molecule stays in the bridged position; only waters with residence time >200 ps are shown.

itself equal to 0.9 Å and 1.28 Å, respectively (Table 1). During these simulations the A4 base did not make any direct hydrogen bonds within the tetraloop (Fig. 4 C); instead, a direct hydrogen bond was found between (A4)2'OH and (A6)N7 with 70–90% occupancy, different than in GCAA loops where the (G4)2'OH group was preferentially bound to (C5)O1P. The amino group of A6, in contrast to the GCAA tetraloop, was not involved in any hydrogen bonds (Table 2 C). In both simulations a very stable water bridge was formed among (A7)N2-H, (A7)N7, and (A4)N3 (Table 3). The same water molecule was present in the bridging position through the whole simulations (1150 ps for simulation A_{4_1} and 1400 ps for simulation A_{4_5}). For simulation A_{4_1} this conformation appeared to be stabilized by a sodium ion associated to the (A7)O1P and (A4)N1, in contrast to the (G4)N1-H...O1P(A7) hydrogen bond in the GCAA tetraloop. During simulation A_{4_5}, some conformational changes within the tetraloop were evident from the RMSD (Fig. 2) and in the time-course of the (A4)N3...N6(A7) distance (Fig. 5 C). At ~250 ps the A4 base moved toward the major groove for a while and then A7 was slightly looped out, making space for a double water bridge between (A4)N3 and H-N6(A7). A water molecule engaged in the water bridge between the first bridging water

and the N6-H, N7 atoms of A7 were found from 470 ps to 840 ps of simulation. An additional water bridge, present during 53% of the simulation, was made by rapidly exchanging waters between (A4)N1 and O1P(A7).

Simulations A_{4_2} and A_{4_4} led to conformations displaying the largest difference from the initial structure. The RMSD for the tetraloop itself was >2.0 Å (Table 1). In simulation A_{4_2} the A4 base moved toward the minor groove, and the A7 base was somewhat looped out and stacked on the A6 base. The A4 base also developed a hydrogen bond between the amino group of A4 and O1P of A7, with the second hydrogen of the NH₂ inside the loop. The C3-G8 basepair was opened with only one (C3)N4-H...O6(G8) hydrogen bond present. In simulation A_{4_4} the amino group of A7 formed hydrogen bonds first with (A4)N3 and (A4)O2', and finally with (A6)O1P and (C5)O1P (Table 2 C). The final structures from simulations A_{4_2} and A_{4_4} also displayed the largest dissimilarity among structures obtained for ACAA tetraloops (*RMSD* = 3.77 Å for tetraloops).

During simulation A_{4_3}, the A4 and A7 bases made direct (A7)N6-H...N3(A4) hydrogen bonds over the first 600 ps; after that, A7 moved out into the solvent, losing stacking with A6. The A4 base developed a hydrogen bond with

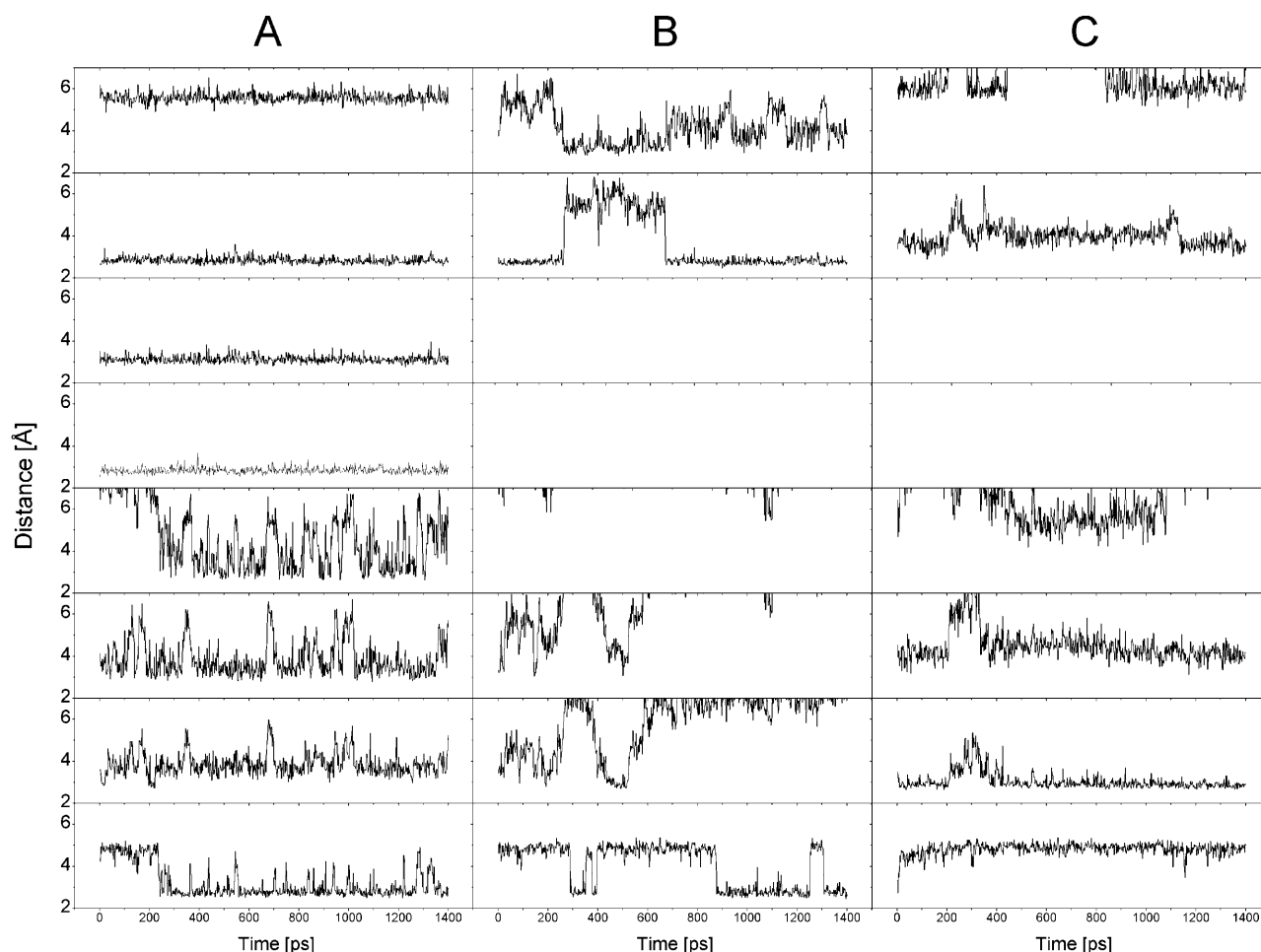


FIGURE 5 Time-course of hydrogen-bond distances within GCAA tetraloop and corresponding distances within ICAA and ACAA tetraloops. (A) GCAA tetraloop (simulation R_1). From the top, the distances correspond to (A7)N6...N3(G4), (G4)N1...O1P(A7), (G4)N2...N7(A7), (G4)N2...O1P(A7), (A6)N6...O1P(C5), (A6)N6...O2'(G4), (G4)O2'...N7(A6), and (G4)O2'...O1P(C5). (B) ICAA tetraloop (simulation I4_1). From the top, the distances correspond to (A7)N6...N3(I4), (I4)N1...O1P(A7), (A6)N6...O1P(C5), (A6)N6...O2'(I4), (I4)O2'...N7(A6), and (I4)O2'...O1P(C5). (C) ACAA tetraloop (simulation A4_5). From the top, the distances correspond to (A7)N6...N3(A4), (A4)N1...O1P(A7), (A6)N6...O1P(C5), (A6)N6...O2'(A4), (A4)O2'...N7(A6), and (A4)O2'...O1P(C5).

(A7)O1P having the second hydrogen of the amino group toward the major groove.

Sugar conformation

The initial conformation of all riboses was C3'-endo and we have not seen any transitions during the course of our simulations, except for the last residue, C10, where for trajectory R_2 a reversible transformation to C2'-endo was observed. However, according to NMR data (Jucker et al., 1996), three loop nucleotides, C5, A6, and A7, displayed conformational equilibrium between C3'-endo and C2'-endo sugar pucker but with dominating C3'-endo conformation. Some transitions between sugar conformations were also seen during the 0.2-ns GCAA tetraloop simulation performed by Zichi (1995), but in that simulation the ribose of the second nucleotide in the loop (C) remained in its initial C2'-endo conformation.

Stacking interactions

Average stacking interaction energies for the bases in the loop from the MD simulations are presented in Table 4. The van der Waals component to a large extent reflects geometrical overlap of the bases. The electrostatic energies were calculated with a dielectric constant of 1.0 and do not take into account solvent effects. The presence of the solvent in general reduces electrostatic interactions but solvent screening effect depends on the RNA architecture and is different for bases buried within RNA and for bases exposed to the solvent.

The G4 base in GCAA tetraloop (R_1) was involved in stacking interactions with both bases from the C3-G8 basepair closing the tetraloop and to some extent also with the A6 base. Bases C5, A6, and A7 stacked sequentially, with similar stacking interaction energies. The stacking energy between the A7 base from the loop and the G8 base

TABLE 4 Average base-base interaction energies [kcal/mol]

Trajectory	Total energy	van der Waals term	Electrostatic term
Sequential stacking			
base 3–base 4			
R_1	−8.0 (1.2)*	−4.6 (0.6)	−3.4 (1.2)
I ₈ _1	−8.1 (1.1)	−4.6 (0.7)	−3.5 (1.2)
I ₄ _1	−8.2 (1.2)	−4.3 (0.7)	−3.8 (1.2)
A ₄ _1	−5.6 (1.2)	−4.3 (0.8)	−1.3 (0.9)
base 5–base 6			
R_1	−6.4 (1.3)	−5.2 (0.8)	−1.2 (0.8)
I ₈ _1	−6.5 (1.4)	−5.3 (0.8)	−1.3 (0.9)
I ₄ _1	−4.7 (1.6)	−4.1 (1.1)	−0.5 (1.1)
A ₄ _1	−6.3 (1.2)	−5.2 (0.7)	−1.1 (0.9)
base 6–base 7			
R_1	−6.5 (1.2)	−6.2 (0.8)	−0.3 (0.8)
I ₈ _1	−6.9 (1.2)	−6.5 (0.8)	−0.4 (0.8)
I ₄ _1	−5.0 (1.7)	−4.8 (1.5)	−0.2 (1.0)
A ₄ _1	−7.2 (1.1)	−6.8 (0.7)	−0.4 (0.8)
base 7–base 8			
R_1	−2.2 (0.9)	−2.6 (0.7)	0.4 (0.5)
I ₈ _1	−2.8 (0.8)	−3.2 (0.9)	0.4 (0.6)
I ₄ _1	−2.6 (1.0)	−3.5 (1.3)	0.9 (0.9)
A ₄ _1	−3.8 (1.2)	−4.0 (0.8)	0.3 (1.0)
Cross stacking			
base 4–base 6			
R_1	−1.0 (0.5)	−2.6 (0.5)	1.6 (0.6)
I ₈ _1	−1.1 (0.6)	−2.6 (0.6)	1.4 (0.6)
I ₄ _1	0.3 (0.4)	−0.5 (0.5)	0.7 (0.4)
A ₄ _1	−1.4 (0.5)	−1.6 (0.3)	0.2 (0.5)
base 4–base 8			
R_1	−10.3 (1.2)	−5.8 (1.1)	−4.5 (1.6)
I ₈ _1	−9.5 (1.3)	−4.5 (1.1)	−5.0 (1.5)
I ₄ _1	−8.5 (1.3)	−5.1 (1.2)	−3.4 (1.5)
A ₄ _1	−0.8 (0.9)	−3.0 (0.7)	2.1 (0.8)

*Energy fluctuations are given in parentheses.

from the stem was less favorable than between bases in the CAA loop fragment.

Replacing G by I in position 8 in the stem part (I₈_1) did not significantly change stacking interaction energies within the loop, whereas for the ICAA tetraloop stacking energies were less favorable for the A6 base. In this case the A6 base was somewhat outside the CAA stacking track (I₄_1). The stacking interaction of I4 with G8 from the C3-G8 basepair was slightly weaker than the corresponding stacking interactions of the G4 base.

For the ACAA tetraloop, the results in Table 4 are from a trajectory that represents a GNRA-like conformation (A₄_1). Replacing G4 by A4 in this structure changed the electrostatic contribution to the stacking interactions with the C3-G8 basepair, such that the electrostatic component of stacking interaction energy between A4 and G8 became repulsive. During these simulations a sodium ion was associated to A4, A7, and G8, and reduced repulsive electrostatic interaction between closely situated electronegative atoms. In the ACAA tetraloop stacking interactions of the C5-A6-A7 fragment were quite well maintained.

C5 remained in its initial stacking conformation in all our MD simulations, whereas, in two out of 10 NMR models, the

second nucleotide in the GCAA loop is looped out into the solvent.

Sodium ions

Table 5 shows sites of sodium ion associations to the RNA hairpins. In most simulations one or two (R_2, A₄_1) sodium ions were directly bound to the pocket at the turn of the loop from the major groove side. For all loops, except ACAA, sodium ions were associated to the N7 and O6 atoms of the base at position 4 and to the phosphate oxygen of A6. For ACAA tetraloops sodium ions were associated only to (A4)N7, and for simulation A₄_1 an additional sodium ion coordinated four negative atoms from three residues: N1(A4), O1P(A7), and O6, N7(G8). Phosphate oxygens were another site of sodium ion association, but the average number of contacts ($N_c = 0.2$ – 0.5) was usually lower than in the above pocket.

Free energy calculations

To evaluate the changes in the hairpin folding free energy upon substitutions of individual bases, an approach using FEP in a thermodynamic cycle was applied (Fig. 6).

Dynamics of unfolded RNA models

The relative stabilities of hairpins of different sequence are given by the free energy differences $\Delta\Delta G = \Delta G_u - \Delta G_h$, which were computed from free energy changes for the alchemical mutations G→I and G→A in the hairpin context (ΔG_h) and in models of unfolded RNA (ΔG_u). The calculated $\Delta\Delta G$ may thus depend on the choice of model for the unfolded state, and to assess the magnitude of this dependence we modeled the unfolded RNA using a single nucleoside, a trinucleotide, and a pentanucleotide fragment (Table 6), with nucleotide sequence and numbering

TABLE 5 Sodium ions associated to RNA

Trajectory	Residue atom name	G4/I4/A4			A6		A7	G8		
		O6	N7	N1	O1P	O2P	O1P	O6	N7	O2P
R_1	1	0.84	0.43		1.00					
R_2	1	0.49			0.53					
	2		0.80		0.74					
I ₈ _1	1	0.48	0.69		0.93					
I ₄ _1	1	0.31	0.65			0.70				
I ₄ _2	1	0.52	0.73		1.00					
A ₄ _1	1		0.91							
	2			0.74			0.99	0.48	0.32	
A ₄ _3	1		0.70		0.59					
A ₄ _4	1									0.51
A ₄ _5	1		0.51							

The values given are average numbers of contacts (N_c). Only sodium ions with at least $N_c > 0.50$ to the one RNA atom are shown.

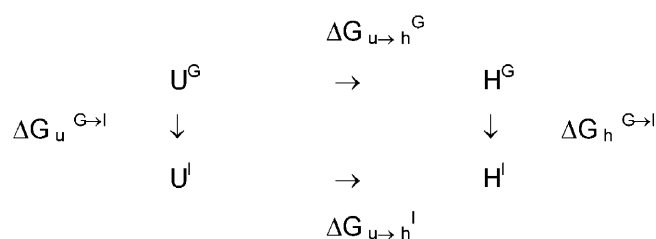


FIGURE 6 Thermodynamic cycle used to calculate relative hairpin formation free energy of mutation $G \rightarrow I$. Subscripts h and u refer to the hairpin and unfolded state; superscripts G and I to reference and mutated hairpin (here G into I); $\Delta G_{u \rightarrow h}^G$ and $\Delta G_{u \rightarrow h}^I$ correspond to the experimental free energy of the hairpins formation; and $\Delta G_h^{G \rightarrow I}$ and $\Delta G_u^{G \rightarrow I}$ correspond to the free energy of alchemical mutations of one molecule fragment into another in hairpin and unfolded state, respectively.

corresponding to the hairpin where the transformed G was in the middle of the sequence.

Nucleoside (guanosine)

During the 2-ns simulation of guanosine the sugar adopted both C3'-endo and C2'-endo pucker, in nearly equal distribution. The base was found in *anti* and *syn* orientation, with a preference for the *syn* conformation, in which the 5'OH was hydrogen-bonded to N3 of guanosine. We did not observe any correlation between sugar pucker and *anti-syn* conformation.

Trinucleotide models

MD simulations of CGC trinucleotide models started from two different initial conformations: one conformation was generated as a canonical A-RNA, and the second was extracted from nucleotides 3–5 of the hairpin. Simulations were carried out for 2 ns. One additional simulation, started from the conformation extracted from the hairpin, was run at 400 K for 1 ns.

The bases in CGC trinucleotide models were in stacking conformation most of the simulation time, but transitions to the unstacked conformation were also observed for C either at the 5'-end or at 3'-end. In the initial structure taken from the hairpin, the bases at the GC step were not stacked. The CG step in this model corresponds to the G4–C5 fragment in the hairpin where the backbone changes direction and C5 does not stack with G4. During the simulation the bases at the GC step adopted a stacked conformation, but showing antiparallel arrangement of the bases, different than in right-handed helices. In the part of the trajectory started from A-RNA, the 3'C-NH₂ made a hydrogen bond with O1P(G).

The MD simulation of the AGC model started from the conformation of nucleotides 7–9 taken from the hairpin fragment, and was run for 1 ns. During the simulation the backbone conformation switched between the A-RNA conformation with ϵ , ξ , and α torsion angles in the regions ($-ac$, $-sc$, and $-sc$), and the conformation described by ϵ , ξ ,

and α in orientations (ap , $+sp/+sc$, and $-ac$). The corresponding AG fragment in the hairpin displayed similar dynamics of torsion angles.

The sugar of the cytosine at the 3' end in trinucleotides was switching between C3'-endo and C2'-endo conformations, whereas the other two sugars remained C3'-endo.

Pentanucleotide model

For the pentanucleotide GCGCA fragment the simulation started from the A-RNA conformation and was run for 0.5 ns. The GC fragment, including the G in the middle of structure, displayed the most dynamic behavior of the backbone, with torsion angles ϵ , ξ , and α switching conformation between regions ($-ac$, $-sc$, and $-sc$) and (ap , $+sp/+sc$, and $-ac$). During the simulation we found intramolecular hydrogen bonds involving the second cytosine base (C5). At an early stage of the simulation the NH₂ group of C5 formed a stable hydrogen bond to O1P(G4) and a second more dynamic to O1P(C3). The sugars maintained the C3'-endo form.

During simulations of trinucleotides and pentanucleotides, torsion angles β defining rotation around C5'–O5' were largely limited to the ap range, and γ defining rotation around C4'–C5' to the $+sc$ region.

Free energy changes in models of unfolded RNA

The free energy change associated with the alchemical mutation $G \rightarrow I$ in a single-stranded RNA was calculated using these different models of unfolded RNA with similar average results (within 0.4 kcal/mol) for all five models (Table 7 A). Calculations of free energy changes for the $G \rightarrow A$ transformation were done on three models (guanosine and two CGC trinucleotide models), giving average values with a spread of ~ 1.0 kcal/mol (Table 7 B). The free energy change is bigger for the trinucleotide CGC model derived from the hairpin fragment than for CGC in A-RNA form. These two models differed in stacking pattern at the GC step, which may influence the calculated energy difference. The free energy changes during alchemical mutations in guanosine alone are close to those obtained for the A-RNA-derived models.

Free energy changes in the hairpin

The free energy change due to a $G \rightarrow I$ mutation in the hairpin was computed in two structural contexts: in the loop ($\Delta G_h^{G4 \rightarrow I4}$) and in the stem ($\Delta G_h^{G8 \rightarrow I8}$). In the stem the $G \rightarrow I$ mutation reduces the hairpin stability by 1.7 to 1.8 kcal/mol and in the loop by 3.1 to 3.5 kcal/mol (Table 7 A). Free energy changes in the loop obtained from independent simulations are within a range of 0.8 kcal/mol, suggesting that they are reasonably converged. The hysteresis, which is defined as the difference between energies from the forward

TABLE 6 Simulated models of unfolded RNA

Models of unfolded RNA	Initial structure	Box size [Å]	Time of MD [ns]
Nucleoside			
guanosine	G4 from hairpin	28 × 22 × 22	2.0
Trinucleotide			
CGC	Hairpin fragment	33 × 27 × 27	2.0
	Hairpin fragment	33 × 27 × 27	1.0 (400 K)
	A-RNA	33 × 27 × 27	2.0
AGC	Hairpin fragment	33 × 27 × 27	1.0
Pentanucleotide			
GCGCA	A-RNA	40 × 29 × 29	0.5

and backward perturbations, was also relatively small, 0.1, 0.5, or 1.2 kcal/mol in all G→I simulations. The conformation of the hairpin was not significantly changed by the G→I mutations, and the final structures after the G4→I4 mutation displayed hydrogen-bonding patterns similar to that presented in Fig. 4 B2; in two out of three 404-ps-long forward simulations the bridging water molecule within the tetraloop did not exchange.

The FEP results for the G→A alchemical mutations ($\Delta G_h^{G4 \rightarrow A4}$), performed with the same protocol as for the G→I mutation or with time extended to 808 ps, differed

when they were conducted in forward or backward direction (Table 7 B). The free energies from all backward simulations were systematically >3.5 kcal/mol lower than from forward simulations. In some cases, better reversibility (1.6 kcal/mol hysteresis) was obtained from calculations using the PME method for treatment of electrostatic interactions and additionally prolonging FEP simulation time to 808 ps (0.2 kcal/mol hysteresis), but the destabilizing effect of the G4→A4 mutation was still very large (5.7 kcal/mol).

The final structures from the G4→A4 FEP simulations stayed close to the GNRA fold, although they displayed some diversity. In the most common structure obtained from the simulation of the G4→A4 mutation (Fig. 4 C) the A4 base did not make any direct hydrogen bonds; however, there was a bridging water molecule making hydrogen bonds between (A4)N3, (A7)N7, and (A7)N6-H. In other structures a new (A4)N6-H...O1P(A7) hydrogen bond developed during the gradual, alchemical change of O6 to an amino group, when a sodium ion was initially associated to (G4)O6 in the loop pocket, or (A7)N6-H was directly bonded to N3(A4). We did not see that this particular difference was directly reflected in the free energy change. During the backward simulations the initial hydrogen-bond network specific for the GCAA motif was rebuilt in most cases, and

TABLE 7 Free energy perturbation results for G→I (A) and G→A (B) mutations [kcal/mol]

A.	Hairpin (forward/backward)	Unfolded state models (forward/backward)			
	$\Delta G_h^{G4 \rightarrow I4}$	$\Delta G_u^{G4 \rightarrow I4}$			
		CGC hairpin fragment	CGC A-RNA	GCGCA A-RNA	Guanosine
	76.8/−76.9	74.2/−74.2	73.8/−73.9	73.7/−73.6	73.7/−73.4
	77.2	73.9	74.1	73.6	73.6
	77.6/−77.1*	73.8*			73.5*
Average	77.1 ± 0.3	74.4 ± 0.2	73.9 ± 0.2	73.6 ± 0.1	73.6 ± 0.1
$\Delta \Delta G^{CALC}$		−3.1	−3.2	−3.5	−3.5
	$\Delta G_h^{G8 \rightarrow I8}$	$\Delta G_u^{G8 \rightarrow I8}$			
		AGC hairpin fragment			
	74.5/−74.3	73.7/−73.7			
	75.4/−75.5	73.8			
	76.2*				
Average	75.4 ± 0.7	73.7 ± 0.1			
$\Delta \Delta G^{CALC}$		−1.7			−1.8
B.	$\Delta G_h^{G4 \rightarrow A4}$	$\Delta G_u^{G4 \rightarrow A4}$			
		CGC hairpin fragment	CGC A-RNA	Guanosine	
	80.9/−76.2*	75.2/−75.3	73.8/−74.5	74.3/−75.0	
	82.5/−78.9	75.3	74.6	74.8/−74.0	
	81.1/−75.7†				
Average	79.2 ± 2.8	75.3 ± 0.1	74.3 ± 0.4	74.5 ± 0.5	
$\Delta \Delta G^{CALC}$		−3.9	−4.9	−4.7	
PME method	128.7/−127.1			122.5/−122.4	
	130.6/−125.3			122.2/−122.4	
	128.4/−128.2†				
Average	128.1 ± 1.8			122.4 ± 0.1	
$\Delta \Delta G^{CALC}$				−5.7	

*Different initial configuration.

†808 ps, 4 ps of equilibration + 4 ps of production run.

only these results are included in Table 7 *B*. Additionally, the final conformations obtained during five simulations of the ACAA tetraloop were used as reference structures to perform FEP calculations for the A4→G4 mutation. Only FEP simulations started from the final structure of the A4_1 and A4_5 runs lead to structures reproducing the GNRA motif. The average ΔG obtained from several FEP simulations started from these two structures was -75.2 ± 1.4 and -79.9 ± 0.3 , respectively, although the RMSD between the initial structures was only 0.58 Å (0.54 Å for the tetraloop). However, the structures displayed differences in association of the sodium ions (Table 5).

We also investigated the extent to which the calculated free energy values were affected by simulation methods. In simulations in a water sphere, and with the simulation time increased to 808 ps (Table 8) for the G4→I4 mutation in hairpin, we obtained $\Delta\Delta G = 2.9$ kcal/mol—although perhaps accidentally, taking into account that, still, there was large hysteresis (4.9 kcal/mol).

In the next approach, the G→A transformation was divided into two steps: G→I, followed by the I→A transformation. Results from the first step were similar to the results from the G4→I4 mutations obtained with periodic boundary condition in a water box (Table 7 *A*). The FEP calculations from the next step (I4→A4 mutation) displayed asymmetry. The final $\Delta\Delta G$ results varied from 4.2 kcal/mol to 6.2 kcal/mol.

DISCUSSION

The strategy of nucleotide analog substitutions and measuring an effect on the structure and folding thermodynamics has been used in several experimental (SantaLucia et al., 1991; SantaLucia, Jr., et al., 1992; Chen et al., 2001) and theoretical (Singh and Kollman, 1996; Williams and Hall,

2000a; Williams et al., 2001) attempts to determine the contribution of individual interactions to molecular properties of RNA. We have employed the same approach in our theoretical study of the nature of the stability of GNRA tetraloops. To examine the effect of base substitutions on conformation and dynamics of an RNA hairpin we conducted MD simulations of the GGCGCAAGCC hairpin with a wild-type GCAA tetraloop and modified hairpins with selective substitutions located within the loop and stem parts. Applied substitutions were designated to eliminate particular hydrogen bonds in the loop thought to stabilize the GNRA motif.

The GCAA tetraloop simulations showed a well-defined hydrogen-bonding pattern between the first and fourth loop nucleotides (G4 and A7). We have observed development of water-mediated hydrogen bonds between G4 and A7, which are not possible to see in NMR models obtained from calculations in vacuum. The establishment of a water bridge in our simulation improves and stabilizes the geometry of hydrogen bonds between the G4 and A7 in comparison to the NMR model. The third loop nucleotide (A6) participated in the intermittently formed hydrogen bonds between its amino proton and the O2'(G4) as well as the O1P(C5). The ribose-base hydrogen bond between 2'OH of G4 and (A6)N7 proposed for GCAA tetraloop was rarely observed during simulation.

GNRA tetraloops have previously been simulated in hairpins in the presence of counterions and explicit water molecules with the CHARMM22 force field (Sarzynska et al., 2000), the OPLS force field (Zichi, 1995), and as a part of hammerhead RNA with the AMBER force field (Hermann et al., 1998). The water bridge between the first and last loop bases was reported by Zichi during simulation of a GCAA tetraloop and by Sarzynska et al. during simulation of a GUGA tetraloop. Dynamic interactions between the 2'-OH group of the G and the base of the third purine in the loop was observed in all mentioned simulations, but its occupancy can depend on the tetraloop sequence and force field; in the present study the preferred orientation of the 2'OH group in the helix is toward the phosphate compared to the orientation toward O3' found in other cases (Sarzynska et al., 2000; Hermann et al., 1998).

When the tetraloop sequence was changed from GCAA to ICAA, the overall loop structure was maintained during molecular dynamics simulations, but we have noticed several possible hydrogen-bonding arrangements and increased fluctuations. Compared to the GCAA tetraloop, the relative positions of I4 and A7 as measured by the (I4)N3...N6(A7) distance displayed larger fluctuations as well as conformational transitions.

Several independent MD simulations of the ACAA tetraloop led to conformational heterogeneity of the resulting structures. Some of the trajectories sampled structures close to the initial conformation derived from the GCAA tetraloop but with increased flexibility, as was monitored in large fluctuations of the (A4)N3...N6(A7) distance. However, we

TABLE 8 Free energy perturbation results from simulations in water sphere [kcal/mol]

Mutation	ΔG^* (forward/backward)	ΔG_{AVG}	$\Delta\Delta G_{CALC}$
Unfolded [†] G4→A4 [‡]	74.4/−74.2	74.3	
Hairpin G4→A4 [§]	79.6/−74.7	77.2	2.9
Hairpin G4→I4	77.0/−78.0	77.5	
Hairpin I4→A4	4.2/−1.8	3.0	
Hairpin G4→A4 via I4	81.2/−79.8	80.5	6.2
Hairpin I4→G4 [¶]	−74.0		
Hairpin A4→G4 via I4 [¶]	−75.8		4.2

*404 ps FEP simulation, 2 ps equilibration + 2 ps production in each window, in water spheres of radius 19 Å and 13 Å for the hairpin and unfolded models, respectively.

[†]Guanosine model.

[‡]202 ps FEP simulation, 1 ps equilibration + 1 ps production in each window.

[§]808 ps FEP simulation, 6 ps equilibration + 2ps production in each window.

[¶]I4→G4 starting from final state of the A4→I4 simulation.

have also obtained alternative structures with hydrogen-bond patterns and stacking interactions in the tetraloop different than in the canonical GNRA motif. These structures covered a conformational space too divergent to expect that they will eventually converge to a new family of structures different from the original hairpin in reasonable simulation time.

The MD simulations revealed the important role of water molecules located in the GCAA loop as well as in the modified ICAA and ACAA loops. These long-lived bridging water molecules, considered as structural waters, have in their environment >2 potential hydrogen acceptors, and, in consequence, we have observed various possibilities of forming hydrogen bonds. Substitutions of the first loop nucleotide (G) and small structural rearrangements within the loop were accommodated by just changing the hydrogen-bonding partners for the bridging water molecule.

Thermodynamic effects of base substitution were studied by free energy perturbation method. Results of our FEP calculations show that the G→I and G→A substitutions in the tetraloop and stem parts destabilize the GGCGCAAGCC hairpin, which is in agreement with experimental results. The magnitude of the free energy difference for G→I substitution in the C-G basepair closing the tetraloop (1.7–1.8 kcal/mol) also corresponds well to the experimental data (1.3 kcal/mol). A similar magnitude of destabilization was estimated experimentally and from calculations upon replacement of G by I in a Watson-Crick G-C basepair in a DNA duplex (1.4 and 1.2 kcal/mol, respectively; Cubero et al., 2001). From our calculations, a G→I substitution in the tetraloop destabilizes the hairpin more (3.1–3.5 kcal/mol) than the same substitution in the basepair closing the tetraloop, whereas experiment showed only minor destabilization of the hairpin with substitution made in the tetraloop (0.65 kcal/mol; SantaLucia, Jr. et al., 1992). The destabilization of the hairpin upon changing the tetraloop sequence from GCAA to ACAA estimated by calculations (2.9–6.2 kcal/mol) is not precise and often significantly higher than found experimentally (0.75 kcal/mol) (SantaLucia, Jr. et al., 1992).

To date, two examples of free energy calculations in nonhelical RNA regions were reported. Both calculations were performed for an UUCG tetraloop and concerned base or deoxyribose substitutions (Singh and Kollman, 1996; Williams and Hall, 2000a,b). Although we have obtained adequate results for hairpin destabilization due to mutation in the stem part, we met some difficulties in estimating the free energy difference when the mutation was performed in the loop. Several reasons for these difficulties can be indicated.

One possible explanation of the results obtained for substitutions in the tetraloop is that the modified loops did not converge to the correct structure, and the calculated $\Delta\Delta G$ values do not reflect the stability of the experimentally observed hairpin conformers with ICAA or ACAA tetraloops, for which the atomic structures have not been solved yet. The final structures obtained in the G→I and G→A FEP simulations were close to the initial conformation of the

GCAA tetraloop, and our calculated free energy changes thus reflect the stability of sequence modifications in structures, which were not significantly different from the initial NMR-derived structures. If the mutation leads to a different conformation separated by a high barrier, it may be difficult to obtain proper sampling during standard MD of limited length. Similar problems have been reported by other authors. The convergence from the incorrect to the correct form of a UUCG tetraloop was not observed during standard MD simulation, but only when a locally enhanced sampling procedure was used (Simmerling et al., 1998) or if the loop riboses were changed to deoxyriboses (Miller and Kollman, 1997).

In regard to G→A mutation, several arguments support the suggestion that the ACAA tetraloop does not maintain the GNRA fold. The diversity of structures found in our independent MD simulations may suggest that the ACAA tetraloop is characterized by a broad conformational heterogeneity and several conformations exist at the experimental conditions. The ACAA loop does not belong to any known class of tetraloops with a defined structural motif like GNRA, UNCG, CUYG, or (U/A)GNN, and CD spectra of a GCAA tetraloop and loops with several modifications incorporated in the first and fourth loop position indicated differences in base stacking (Worner et al., 1999). An example where minor thermodynamic effects of mutations was connected with a large conformational change is provided by the GGAA tetraloop, where replacing the first G by m²G and the two As by m²A resulted in a small destabilizing effect, but NMR studies showed that this methylated GGAA tetraloop has a different conformation than the GNRA motif (Rife et al., 1998; Rife and Moore, 1998). In the methylated tetraloop the backbone turns toward the minor groove, whereas in the tetraloop of GNRA type the backbone turns toward the major groove and the NRA bases are on the minor groove side.

One indicator that our MD simulations sampled only limited conformational space is that during several independent MD runs we did not reach all conformations reported by NMR studies. We have not seen the second nucleotide C in an unstacked conformation, looped out into the solvent. In a recent simulation of an RNA hairpin containing a GCAA tetraloop using the GB/SA implicit solvent model (Sorin et al., 2002), three GCAA loop conformations were found: the dominant conformation with the C nucleotide in a stacked conformation, a less favored with C in an unstacked conformation, and a far less stable with the last A nucleotide moved out from the loop. A correlation between transformation from stacked to the unstacked conformation and change of the sugar pucker was also observed in that study. A more flexible behavior of the second nucleotide in a GAAG tetraloop, which displays the GNRA motif, was observed during a simulation with GB/SA model compared to the simulation in explicit solvent (Li et al., 2001).

The calculated destabilization effect of a G→I mutation in the tetraloop is in accordance with rational expectation, when taking into account the structural similarity between the original and mutated tetraloops and the number of deleted hydrogen bonds. The G→I substitution in the stem part destabilized the hairpin by 1.7–1.8 kcal/mol, which can be explained by the loss of one hydrogen bond between the amino group of G and the carbonyl oxygen of C. In the tetraloop, the G→I substitution eliminates two hydrogen bonds buried inside the loop, and the calculated decrease of hairpin stability due to this G→I substitution is $\sim 2\times$ bigger (3.1 kcal/mol–3.5 kcal/mol) than for the substitution in the flanking C-G basepair.

The hairpin stability change upon G→A substitution in the tetraloop evaluated by our calculations seems to be too high to be explained only as $\Delta\Delta G$ between GCAA and ACAA hairpins having similar structures. Additionally, results obtained from applying several approaches to calculate the free energy difference due to G→A mutation are spread from 2.9 kcal/mol to 6.2 kcal/mol, and these differences cannot originate only from small structural differences observed during FEP simulations. We suggest that another source of the difficulties in accurately estimating the free energy change during G→A transformations can be nonreproducible interactions of the sodium ions located close to the mutation site, where the amino group is changed to a carbonyl.

The ion distribution found in the MD simulations (Table 5) indicated that the pocket at the turn of the loop from the major groove is the preferential site for sodium ion binding. During most of the simulations, even where initially there was no sodium ion in the vicinity of the loop pocket, a sodium ion finally diffused into this pocket.

Our simulations indicate that sodium ions directly associated to RNA did not significantly influence the structure of CGAA tetraloop. In test simulations (data not shown), where sodium ions initially were placed at a distance longer than 5.2 Å from RNA and did not diffuse to the first solvation shell of the loop pocket during 0.5 ns of simulation, the structure of the GCAA tetraloop was similar (RMSD of 1.59 Å from starting structure) to the structure with bound sodium ions. On the other hand, the ACAA loop simulation with the tetraloop maintaining its initial structure (A_{4-1}) was probably stabilized by a directly bound sodium ion, which replaced the (G4)N1-H...O1P(A7) hydrogen bond present in the GCAA tetraloop and additionally reduced the repulsive electrostatic component of the stacking interaction energy between A4 and G8. In another trajectory of the ACAA tetraloop (A_{4-5}), with a conformation close to GNRA but without this bound ion, the A4 and A7 nucleotides displayed larger fluctuations. In the G4→A4 FEP simulations the presence of a sodium ion having a preference to interact with the mutated chemical group can influence the calculated energy.

Most FEP results for G4→A4 mutation in the loop showed asymmetry. It is suggested that, in such cases, results from

the direction going from the higher to the lower entropy of the system rather than average values from forward and backward simulations should be used (Lu and Kofke, 2001a,b). The entropies calculated from all simulations show some differences between trajectories but without apparent tendency to be higher for ACAA tetraloop compared to GCAA or ICAA tetraloops (Table 9). We think rather that differences in sodium location in GCAA and ACAA tetraloops could be the source of observed asymmetry.

The GNRA loop has been shown to be the binding site for different multivalent ions (Rudisser and Tinoco, Jr., 2000; Maderia et al., 2000; Mundoma and Greenbaum, 2002). Brownian dynamics simulations also predicted the binding of divalent metal ions to the GNRA motif (Hermann and Westhof, 1998). In the phosphate backbone charge neutralization required for RNA folding, monovalent cations dominate. However, information about specific monovalent metal ion binding sites within RNA is limited due to difficulty in their identification in crystal structures, although examples of a potassium ion being bound to the AA platform of the GAAA tetraloop receptor (Basu et al., 1998) or completely buried within the tertiary structure of rRNA were reported (Conn et al., 2002). The solution structure of the GNRA loop (Jucker et al., 1996) and the thermodynamic data for substitutions in the GCAA tetraloop (SantaLucia, Jr. et al., 1992) refer to experimental conditions without multivalent ions, similar to our simulation.

The application of a thermodynamic cycle requires the structure of an unfolded single RNA strand. There is little experimental data concerning single-stranded RNA, but there are indications of the existence of a preorganized A-like backbone conformation (Lindqvist et al., 2000). We examined several models of single-stranded RNA which differ in size, initial structure, or sequence. MD simulations of CGC RNA fragments showed that, during 2 ns, only a limited conformational space is searched, and that the accessible conformations depend on the initial structure. In our studies, the free energy change associated with the G→I substitution in single-stranded RNA did not depend significantly on the model. For the G→A substitution, our results depended, to some extent, on the structure of the

TABLE 9 Entropy of RNA hairpins at 300 K [kcal/mol/K]

Tetraloop sequence	Trajectory name	Whole hairpin	Backbone	Base
GCAA	R_1	0.87	0.65	0.39
	R_2	0.94	0.66	0.45
C(GCAA)I	I ₈₋₁	0.89	0.66	0.39
ICAA	I ₄₋₁	0.95	0.72	0.42
	I ₄₋₂	0.91	0.68	0.40
ACAA	A ₄₋₁	0.88	0.66	0.38
	A ₄₋₂	0.91	0.68	0.41
	A ₄₋₃	0.94	0.70	0.42
	A ₄₋₄	0.95	0.72	0.42
	A ₄₋₅	0.96	0.72	0.42

trinucleotide model used in the calculations. We concluded that, for the base mutations performed in this work, a single nucleoside seems to be a sufficient model for unfolded RNA.

Other authors have also considered the relevance of the complexity of unfolded state models used in FEP calculations. The free energy change associated with substitution of a base in a single-stranded DNA was well-approximated by the corresponding change in a single nucleoside (Florián J. et al., 2000). Williams studied several models of single-strand RNA and reported that free energy change for deoxyribose substitutions depends on the nucleotide identity rather than on structural context (Williams and Hall, 2000a).

The ability to predict the influence of functional groups substitutions on the properties of RNA and its analogs is crucial for the rational design of RNA-derived molecules of therapeutic application. Our study suggests that GNRA tetraloops where G is replaced by I or A may have other low-energy conformations distinct from the GNRA fold, which were not reached during FEP simulations. This observation is consistent with the idea that the phylogenetic preference for GNRA tetraloops, beside their exceptional thermodynamic stability, is due to their specific structure which enables them to form tertiary interaction. Substitution of the GCAA tetraloop sequence to ACAA does not significantly change stability, as was shown experimentally, but the GCAA tetraloop probably cannot be replaced by an ACAA tetraloop because, due to structural differences, they cannot perform the same biological function.

Note added in proof: After this paper was accepted, the structure of an ACAA tetraloop was determined by NMR (Staple, D. W., S. E. Butcher, 2003. Solution structure of the HIV-1 frameshift inducing stem-loop RNA. *Nucl. Acids. Res.* 31:4326–4331) and found to be of the AGNN-tetraloop type, rather than of the GNRA-type. This confirms our suggestion that ACAA tetraloop does not maintain GNRA fold.

This work was supported by a grant from the State Committee for Scientific Research, Republic of Poland (8 T11F002 19), and the Swedish Institute. The generous support of the Poznań Supercomputing and Networking Centre is also acknowledged.

REFERENCES

- Andricioaei, I., and M. Karplus. 2001. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115: 6289–6292.
- Antao, V. P., and I. Tinoco, Jr. 1992. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* 20:819–824.
- Basu, S., R. P. Rambo, J. Strauss-Soukup, J. H. Cate, A. R. Ferre-D'Amare, S. A. Strobel, and J. A. Doudna. 1998. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* 5:986–992.
- Brooks III, C. L., and M. Karplus. 1983. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* 79:6312–6325.
- Cate, J. H., A. R. Gooding, E. Podell, K. H. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*. 273: 1678–1685.
- Chen, X. Y., R. Kierzek, and D. H. Turner. 2001. Stability and structure of RNA duplexes containing isoguanosine and isocytidine. *J. Am. Chem. Soc.* 123:1267–1274.
- Conn, G. L., A. G. Gittis, E. E. Lattman, V. K. Misra, and D. E. Draper. 2002. A compact RNA tertiary structure contains a buried backbone-K⁺ complex. *J. Mol. Biol.* 318:963–973.
- Correll, C. C., I. G. Wool, and A. Munishkin. 1999. The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Ångstrom resolution. *J. Mol. Biol.* 292:275–287.
- Cubero, E., R. Guimil-Garcia, F. J. Luque, R. Eritja, and M. Orozco. 2001. The effect of amino groups on the stability of DNA duplexes and triplexes based on purines derived from inosine. *Nucleic Acids Res.* 29:2522–2534.
- Dale, T., R. Smith, and M. J. Serra. 2000. A test of the model to predict unusually stable RNA hairpin loop stability. *RNA*. 6:608–615.
- Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
- Feller, S. E., Y. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103:4613–4621.
- Florián, J., M. F. Goodman, and A. Warshel. 2000. Free-energy perturbation calculations of DNA destabilization by base substitutions: the effect of neutral guanine-thymine, adenine-cytosine and adenine--difluorotoluene mismatches. *J. Phys. Chem. B*. 104:10092–10099.
- Foloppe, N., and A. D. MacKerell. 2000. All-atom empirical force field for nucleic acids. I. Parameter optimization based on small-molecule and condensed-phase macromolecular target data. *J. Comp. Chem.* 21: 86–104.
- Hermann, T., P. Auffinger, and E. Westhof. 1998. Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* 27: 153–165.
- Hermann, T., and E. Westhof. 1998. Exploration of metal ion binding sites in RNA folds by Brownian dynamics simulations. *Struct. Fold. Des.* 6:1303–1314.
- Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*. 31:1695–1697.
- Jorgensen, W. J. 1998. Free energy changes in solution. In *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. E. Schreiner, editors. John Wiley & Sons, Chichester, UK. 1061–1070.
- 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.* 147:926–935.
- Jucker, F. M., H. A. Heus, P. F. Yip, E. H. Moors, and A. Pardi. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* 264:968–980.
- Kim, C. H., and I. Tinoco, Jr. 2001. Structural and thermodynamic studies on mutant RNA motifs that impair the specificity between a viral replicase and its promoter. *J. Mol. Biol.* 307:827–839.
- Kollman, P. 1993. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395–2417.
- Li, W., B. Y. Ma, and B. A. Shapiro. 2001. Molecular dynamics simulations of the denaturation and refolding of an RNA tetraloop. *J. Biomol. Struct. Dyn.* 19:381–396.
- Lindqvist, M., M. Sarkar, A. Winqvist, E. Rozners, R. Stromberg, and A. Graslund. 2000. Optical spectroscopic study of the effects of a single deoxyribose substitution in a ribose backbone: implications in RNA-RNA interaction. *Biochemistry*. 39:1693–1701.
- Lu, N. D., and D. A. Kofke. 2001a. Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *J. Chem. Phys.* 114: 7303–7311.
- Lu, N. D., and D. A. Kofke. 2001b. Accuracy of free-energy perturbation calculations in molecular simulation. II. Heuristics. *J. Chem. Phys.* 115:6866–6875.

- MacKerell, Jr., D. A., B. Brooks, C. L. Brooks II, L. Nilsson, B. Roux, and M. Karplus. 1998. CHARMM: the energy function and its parameterization with an overview of the program. In *Encyclopedia of Computational Chemistry*. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. E. Schreiner, editors. John Wiley & Sons, Chichester, UK. 271–277.
- Mark, A. E. 1998. Free energy perturbation calculation. In *Encyclopedia of Computational Chemistry*. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. E. Schreiner, editors. John Wiley & Sons, Chichester, UK. 1070–1083.
- Maderia, M., T. E. Horton, and V. J. DeRose. 2000. Metal interactions with a GAAA RNA tetraloop characterized by (31)P NMR and phosphorothioate substitutions. *Biochemistry*. 39:8193–8200.
- Miller, J. L., and P. A. Kollman. 1997. Theoretical studies of an exceptionally stable RNA tetraloop: observation of convergence from an incorrect NMR structure to the correct one using unrestrained molecular dynamics. *J. Mol. Biol.* 270:436–450.
- Mundoma, C., and N. L. Greenbaum. 2002. Sequestering of Eu(III) by a GAAA RNA tetraloop. *J. Am. Chem. Soc.* 124:3525–3532.
- Norberg, J., and L. Nilsson. 2000. On the truncation of long-range electrostatic interactions in DNA. *Biophys. J.* 79:1537–1553.
- Pastor, N., A. D. MacKerell, Jr., and H. Weinstein. 1999. TIT for TAT: the properties of inosine and adenosine in TATA box DNA. *J. Biomol. Struct. Dyn.* 16:787–810.
- Pearlman, D. A., and R. Govinda. 1998. Free energy calculations: methods and applications. In *Encyclopedia of Computational Chemistry*. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. E. Schreiner, editors. John Wiley & Sons, Chichester, UK. 1036–1061.
- Pley, H. W., K. M. Flaherty, and D. B. McKay. 1994a. Three-dimensional structure of a hammerhead ribozyme. *Nature*. 372:68–74.
- Pley, H. W., K. M. Flaherty, and D. B. McKay. 1994b. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*. 372:111–113.
- Proctor, D. J., J. E. Schaak, J. M. Bevilacqua, C. J. Falzone, and P. C. Bevilacqua. 2002. Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry*. 41:12062–12075.
- Rife, J. P., C. S. Cheng, P. B. Moore, and S. A. Strobel. 1998. N-2-methylguanosine is iso-energetic with guanosine in RNA duplexes and GNRA tetraloops. *Nucleic Acids Res.* 26:3640–3644.
- Rife, J. P., and P. B. Moore. 1998. The structure of a methylated tetraloop in 16S ribosomal RNA. *Structure*. 6:747–756.
- Rudisser, S., and I. Tinoco, Jr. 2000. Solution structure of Cobalt^{III} hexammine complexed to the GAAA tetraloop, and metal-ion binding to G:A mismatches. *J. Mol. Biol.* 295:1211–1223.
- Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Phys.* 23: 327–341.
- Saenger, W. 1984. *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- SantaLucia, J., R. Kierzek, and D. H. Turner. 1991. Functional group substitutions as probes of hydrogen bonding between GA mismatches in RNA internal loops. *J. Am. Chem. Soc.* 113:4313–4322.
- SantaLucia, J., Jr., R. Kierzek, and D. H. Turner. 1992. Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*. 256:217–219.
- Sarzynska, J., T. Kulinski, and L. Nilsson. 2000. Conformational dynamics of a 5S rRNA hairpin domain containing loop D and a single nucleotide bulge. *Biophys. J.* 79:1213–1227.
- Simmerling, C., J. Miller, and P. 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.
- Sorin, E. J., M. A. Engelhardt, D. Herschlag, and V. S. Pande. 2002. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* 317:493–506.
- Straatsma, T. P., and J. A. McCammon. 1992. Computational alchemy. *Annu. Rev. Phys. Chem.* 43:407–435.
- Steinbach, P. J., and B. R. Brooks. 1994. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comp. Chem.* 15: 667–683.
- Sun, Y. C., D. L. Veenstra, and P. A. Kollman. 1996. Free energy calculations of the mutation of Ile96→Ala in barnase: contributions to the difference in stability. *Protein Eng.* 9:273–281.
- Williams, D. J., J. L. Boots, and K. B. Hall. 2001. Thermodynamics of 2'-ribose substitutions in UUCG tetraloops. *RNA*. 7:44–53.
- Williams, D. J., and K. B. Hall. 2000a. Experimental and theoretical studies of the effects of deoxyribose substitutions on the stability of the UUCG tetraloop. *J. Mol. Biol.* 297:251–265.
- Williams, D. J., and K. B. Hall. 2000b. Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J. Mol. Biol.* 297:1045–1061.
- 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.
- Worner, K., T. Strube, and J. W. Engels. 1999. Synthesis and stability of GNRA-loop analogs. *Helv. Chim. Acta*. 82:2094–2104.
- Zichi, D. A. 1995. Molecular dynamic of RNA with the OPLS force field. Aqueous simulation of a hairpin containing a tetranucleotide loop. *J. Am. Chem. Soc.* 117:2957–2969.