JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Ribose 2′-Hydroxyl Groups Stabilize RNA Hairpin Structures Containing GCUAA Pentaloop

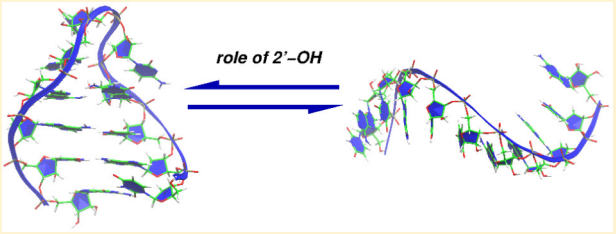
Antonella Paladino[†] and Ronen Zangi*[,†,‡]

[†]Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain

[‡]IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain

Supporting Information

ABSTRACT: The chemical structure of RNA and DNA is very similar; however, the three-dimensional conformation of these two nucleic acids is very different. Whereas the DNA adopts a repetitive structure of a double-stranded helix, RNA is primarily single stranded with a complex three-dimensional structure in which the hairpin is the most common secondary structure. Apart from the difference between uracil and thymine, the difference in the chemical structure between RNA and DNA is the presence of a hydroxyl group at position 2′ of the sugar (ribose) instead of a hydrogen (deoxyribose). In this paper, we present molecular dynamics simulations addressing the contribution of 2′-hydroxyls to the stability of a GCUAA pentaloop motif. The results indicate that the 2′-hydroxyls stabilize the hairpin conformation of the GCUAA pentaloop relative to an analogous oligonucleotide in which the ribose sugars in the loop region were substituted with deoxyriboses. The magnitude of the stabilization was found to be 23.8 ± 4.1 kJ/mol using an alchemical mutations free energy method and 4.2 ± 6.5 kJ/mol using potential of mean force calculations. The latter indicates that in addition to its larger thermodynamic stability the RNA hairpin is also kinetically more stable. We find that the excess stability is a result of intrahairpin hydrogen bonds in the loop region between the 2′-hydroxyls and sugars, bases, and phosphates. The hydrogen bonds with the sugars and phosphates involve predominantly interactions with adjacent nucleotides. However, the hydrogen bonds with the bases involve also interactions between groups on opposite sides of the loop or with the middle base of the loop and are therefore likely to contribute significantly to the stability of the loop. Of these hydrogen bonds, the most frequent is observed between the 2′-hydroxyl at the first position of the pentaloop with N6/N7 of adenine at the forth position, as well as between the 2′-hydroxyl at position −1 with N6 of adenine at the fifth position. Our results contribute to the notion that one of the important roles of the ribose sugars in RNA is to facilitate hairpin formation.



I. INTRODUCTION

The chemical structure of RNA is very similar to that of DNA.[1] In RNA, a hydroxyl group at position 2′ of the sugar ring is present instead of a hydrogen, i.e., ribose instead of deoxyribose. Furthermore, a hydrogen is present instead of a methyl group at position 5 of one of the pyrimidine bases, i.e., uracil instead of thymine. Despite these small differences in the chemical structure, the three-dimensional conformations of these two nucleic acids are completely different. Whereas the DNA adopts a repetitive structure of a double-stranded helix (most commonly in the B-form), RNA is primarily single stranded with a complex three-dimensional structure. Similar to the structure of proteins, the nonrepetitive structure of RNA contains secondary structures. The most common is the hairpin structure; it is composed of a stem, which is a Watson−Crick base-paired helical region (mostly in the A-form), and a loop of few unpaired nucleotides. The loop is the structural element that allows the single stranded RNA to fold back onto itself and, thereby, is necessary for pairing the bases.[2] Note that depending on the external conditions, RNA can also adopt a double stranded helical form.[3] Due to their small size, hairpin structures were the subject of several computational studies exploring stability, dynamics, and folding pathways.[4−11]

The diverse globular folds exhibited by RNAs are probably related to the large number of functions they perform in the cell. In particular, these tertiary folds are recognized by proteins, ligands, and other RNA molecules.[12] What is the role of the ribose 2′-hydroxyl groups on the structure and function of RNAs? In the A-form helical duplex, these hydroxyls are exposed to the solvent,[13] and their role in forming this secondary structure is not yet clear. However, 2′-hydroxyls are known to stabilize many tertiary interactions.[14,15] For example, the ribose-zipper motif is formed by the interaction of four 2′-hydroxyl groups, in which two consecutive 2′-hydroxyl groups on one strand interact with two consecutive 2′-hydroxyl groups of an antiparallel interacting strand.[16,17] Another example is the A-minor motif. It is formed by the insertion of adenines into the minor groove of neighboring helices where they form hydrogen bonds with

Received: July 20, 2012
Published: January 15, 2013

one or two of the helix 2′-hydroxyls.[18] Via these and other similar types of interactions, ribose 2′-hydroxyls were found to play a role in RNA cleavage by ribozymes,[19−23] the polymerization of RNA,[24] and the binding of tRNA to ribosomes.[25] The 2′-hydroxyls are also responsible for the formation of base-triple structures[26] and can serve as sites for RNA−protein[27] and RNA−peptide[28] interactions.

An intriguing possibility for the effect of 2′-hydroxyls characterizing the RNA backbone is their participation in stabilizing loop structures. Results from an X-ray study of a tRNA molecule indicated the existence of several hydrogen bonds involving these hydroxyl groups in the nonhelical conformations of the molecule.[29] Other studies, using NMR spectroscopy[30] and X-ray crystallography[31] of a hairpin containing the stable UUCG tetraloop, reported the existence of hydrogen bonds involving the 2′-hydroxyls in the loop region. Consequently, a DNA analogue of this tetraloop exhibited more flexibility with less structure and no specific internucleotide interactions.[32] From a thermodynamic point of view, the melting temperature of a RNA hairpin that contains the UUCG tetraloop was shown to be 8 °C higher than that of the same hairpin but with deoxyribose sugars in the tetraloop backbone.[33] Follow up studies showed that the ribose → deoxyribose substitution is position dependent in which the substitution of the first position in the tetraloop exhibits the largest destabilization (of about 1 kcal/mol).[34,35] The position dependent behavior has also been inferred by implicit solvent molecular dynamics simulations.

How general are the results from the UUCG sequence? Studies on a hairpin that contains the GCAA tetraloop indicated that the ribose → deoxyribose substitution of the first base of this loop resulted in a very small change in the melting temperature of the hairpin.[36] Accordingly, it was concluded that no change in the thermodynamic stability of the hairpin accompanied this substitution. Thus, contradictory reports are found in the literature for different loops, and it well may be that the contribution of the 2′-hydroxyls to the stability of a loop depends on the sequence of that loop.

In addition to tetraloops, loop motifs with a different number of nucleotides are also known to be abundant in the structure of RNAs. A GCUAA pentaloop with a distinct fold has been reported as well using a NMR technique.[37] This GCU(A/C)A pentaloop is conserved in mammals and birds[38] and is stabilized by a complex interplay of hydrogen bonds and stacking interactions.

In this paper, we present results from molecular dynamics simulations of a hairpin containing the GCUAA pentaloop. We find that the 2′-hydroxyls in the loop region stabilize the hairpin compared with an analogue oligonucleotide in which these groups were substituted with hydrogens. The magnitude of the stabilization is found to be 23.8 ± 4.1 kJ/mol using the alchemical mutations free energy method and 4.2 ± 6.5 kJ/mol using potential of mean force calculations. Although the 2′-OH groups form hydrogen bonds with sugars, bases, and phosphates in the loop region, it is likely that the hydrogen bonds with the bases contribute the most to the stability of the loop because they involve interactions favoring the bending of the loop. In particular, the 2′-hydroxyl at the first position of the penta-loop makes extensive hydrogen bonds with the adenine at the fourth position. Furthermore, the 2′-hydroxyl at position −1 of the penta-loop also makes extensive hydrogen bonds with adenine at position 5. Thus, our results support the role of the 2′-hydroxyls in stabilizing loop structures.

II. METHODS

We study the effect of the ribose 2′-hydroxyls on the stability of a RNA hairpin containing the GCUAA pentaloop motif. To this end, we perform two series of simulations for, both, an all-ribose RNA backbone and a chimaeric oligonucleotide containing deoxyriboses around the loop region. In the first series of simulations, we calculate the relative free energy change of unfolding the hairpin to an extended structure, whereas in the second, we construct the potential of mean force (PMF) of this process.

The model RNA hairpin taken for the simulations is based on the NMR solution structure of the central region of the human R/G stem-loop pre-mRNA (PDB access number 1YSV).[37] We choose the first conformer, out of the 13, deposited. In order to minimize the size of the simulation box, and thereby allow longer simulation times, we consider a shortened hairpin in which eight base-pairs in the stem termini were removed. The resulting 11-nucleotide RNA has the sequence 5′-UAUGCUAAAUG-3′ in which the central five bases constitute the loop region flanked by a three base-paired stem. Note that this shortened version of the hairpin closes with the, wobble, G−U base pair, because the closest G−C base pair to the loop is three base pairs farther in the stem. To include also the effect of the neighboring 2′-hydroxyl of the stem present on each side of the pentaloop, the chimaeric oligonucleotide (5′-rUrAdUdGdCdUdAdAdArUrG-3′ is a hybrid of RNA and DNA backbones and hereafter referred to as "hybrid") contained seven deoxyriboses at the center of the molecule. Thus, the difference between the RNA and the hybrid oligomers is the substitution of seven 2′-OH's with 2′-H. This modification was performed using the Maestro package.[39] To neutralize the system from the negative charge of the 10 phosphate groups, we added 10 sodium ions at random positions. We did not add additional salt because the NMR data were obtained at a very low concentration of 0.01 M of sodium phosphate.

The molecular dynamics package GROMACS version 4.5.5[40] was used to perform all of the computer simulations with a time step of 0.002 ps and periodic boundary conditions applied in all three dimensions. The electrostatic forces were evaluated with the Particle-Mesh Ewald method[41] (with real-space cutoff of 1.0 nm, grid spacing of 0.12 nm, and quadratic interpolation) and the Lennard-Jones forces by a cutoff of 1.0 nm (with long-range dispersion corrections for the energy and pressure). The system was maintained at a constant temperature of 300 K by the velocity rescaling thermostat[42] (with a coupling time of 0.1 ps), and at a pressure of 1.0 bar by the Berendsen thermostat[43] (with a compressibility of 5×10^{-5} 1/bar and a coupling time of 1.0 ps). Water bond distances and angles were constrained using the SETTLE algorithm.[44] The oligonucleotides covalent bond distances were restrained by harmonic potentials in the PMF simulations and constrained using the LINCS algorithm[45] in the mutation free energy simulations. The RNA and hybrid were represented by the AMBER99 force field[46] and the water molecules by the TIP3P model.[47] The sodium counterions were also represented by the AMBER99 force field, i.e., σ_{Na^+} = 0.333 nm, and ε_{Na^+} = 0.0116 kJ/mol. The size of the rectangular simulation box was determined by imposing a minimum distance of 1.4 nm between the oligonucleotide atoms and the box walls. The system was first energy minimized using the steepest descent approach, followed by a 4 ns simulation in which the positions of the hairpin heavy atoms were restrained

by a harmonic potential with a force constant of 1000 kJ/(mol·nm^2).

Mutation Free Energy Calculations. The free energy change of unfolding the RNA hairpin (to an extended conformation) relative to that of unfolding the hybrid hairpin was computed by the concept of a thermodynamic cycle (see Figure 1). Alchemical mutations of atom types, bonds, angles, and dihedrals were performed to transform riboses to deoxyriboses (i.e., 2′-OH → 2′-H). To circumvent instabilities arising from the process of annihilating atoms, we used soft-core potentials with $p = 2$ and $\alpha = 1.5$. The Coulomb and LJ parameters were modified in a single step. These transformations were performed in the folded hairpin state and in the extended state of the oligonucleotides. Because we reduced the length of the stem, the hairpins can potentially exhibit some degree of unfolding (especially at the termini) when simulated without any restraints. In order to avoid this possibility, we imposed position restraints on the C3′ atom of the second nucleotide and C3′ atom of the tenth nucleotide. The force constant of the restraining harmonic potentials was taken to be 2000 kJ/(mol·nm^2) along the x, y, and z axes. The same restraints were applied to the extended state. On average, the distance between the restrained atoms was 1.46 and 4.13 nm for the folded and extended conformations, respectively, for both oligonucleotides. The free energy changes associated with the RNA to hybrid transformations were computed using the thermodynamic integration technique. For each transformation, 13 equally spaced λ points from $\lambda = 0$ to $\lambda = 1$ were constructed. At locations where the plot of $\partial\mathcal{H}/\partial\lambda$ as a function of λ did not exhibit a smooth behavior, we added eight extra λ points[48] (see Figure S1 in the Supporting Information). At each λ point the system was simulated for 40 ns. The value of $\partial\mathcal{H}/\partial\lambda$ was averaged over all simulation times, discarding the first 5 ns. The estimation of the error in calculating the value of $\Delta\Delta G$ was obtained by integrating the errors obtained for the different λ points. The error at each λ is calculated by dividing the standard deviation of $\partial\mathcal{H}/\partial\lambda$ by the square root of the number of independent data points. The latter was estimated from the total time of the simulation at each λ divided by the corresponding autocorrelation time.[49] In the 2′-OH → 2′-H transformation, a dummy atom is bonded to the 2′-H in the hybrid state. All bonded interactions necessary to describe this dummy atom were kept the same as for the ribose sugar. The number of water molecules in the system of the folded hairpin was 3579, whereas it was 3866 for the extended state.

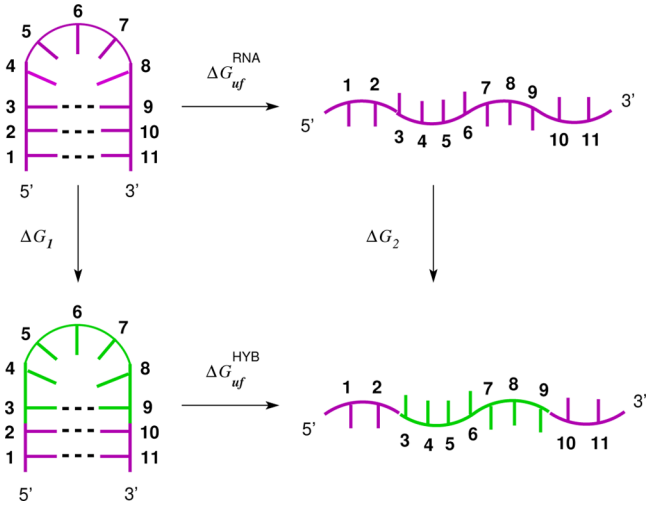


Figure 1. The thermodynamic cycle used to calculate the Gibbs energy change of unfolding the RNA hairpin, 5′-UAUGCUAAAUG-3′ (top), relative to that of unfolding the hybrid hairpin, 5′-rUrAdUdGdCdU-dAdAdArUrG-3′ (bottom), in which seven riboses around the pentaloop were substituted with deoxyriboses. In this cycle, $\Delta G_{uf}^{RNA} - \Delta G_{uf}^{HYB} = \Delta G_1 - \Delta G_2$. Note that the pentaloop is composed of nucleotides 4−8.

Potential of Mean Force Calculations. The reaction coordinate for constructing the PMF ($w(d)$) is defined by the distance, d, between the C3′ atoms of the third and ninth nucleotides. The positions of these two atoms were held fixed in space for the simulations at each value of d. If $\vec{r}_1$ and $\vec{r}_2$ are the (fixed) position vectors of these two atoms and $\vec{F}_1$ and $\vec{F}_2$ are the corresponding total forces acting on the atoms, then, $-\partial w(r_{12})/\partial r_{12} = 1/2\langle \hat{r}_{12}\cdot(\vec{F}_1 - \vec{F}_2)\rangle_{\vec{r}_1,\vec{r}_2}$. The unit vector, $\hat{r}_{12}$, is along the line connecting particles 1 and 2, and the average is over all possible configurations of all other particles. Then, the mean force acting between these atoms along their axis of separation was integrated as a function of their interatom distance, $r_{12} = d$, to yield the PMF (or Gibbs energy profile).[50,51] As the PMF represents only relative values, it was shifted such that the Gibbs energy of the state at the largest

Table 1. The Time Average of the Number of Hydrogen Bonds between Groups A and B, in Which Atoms of Both Groups Can Act As Donors and Acceptors, for the RNA and Hybrid 11-Nucleotide Hairpins[a]

A [nucleotides]	B [nucleotides]	RNA(NMR-1)	RNA(NMR-av)	RNA	hybrid	RNA-27
1−11	1−11	12	12.4	11.3 ± 1.7	10.8 ± 0.5	13.4
3−9	3−9	7	7.3	5.9 ± 1.6	4.6 ± 0.7	8.6
3−9 only 2′OH	3−9 excl. 2′-OH	1	1.9	3.3 ± 0.6		3.4
3−9 only 2′OH	3−9 only sugars	0	0.54	2.1 ± 1.1		2.6
3−9 only 2′OH	3−9 only bases	0	0.15	0.7 ± 0.2		0.1
3−9 only 2′OH	3−9 only phosphates	1	1.2	0.6 ± 0.2		0.8
3−9 only 2′OH	3−9 only 2′-OH	0	0	<0.1		<0.1
3−9 only 2′OH	1−2 and 10−11	0	0	0.7 ± 0.4		0.3
1−2 and 10−11	1−2 and 10−11	5	5.1	3.9 ± 0.9	4.8 ± 0.8	4.2

[a]The seven nucleotides, 3−9, that underwent the mutation ribose → deoxyribose include the pentaloop and one nucleotide on each side. The other nucleotides, 1−2 and 10−11, are part of the stem. Also indicated are the corresponding values of the untruncated NMR structure of the RNA hairpin (27-mer oligonucleotide) for the first NMR structure, RNA(NMR-1), and the average over the 13 NMR structures, RNA(NMR-av). For comparison, in the last column we present results from 40 ns simulations of 27-mer untruncated RNA without imposing any restraints or constraints, RNA-27. A hydrogen bond is defined by a donor−acceptor cutoff distance of 0.35 nm and a donor−hydrogen−acceptor angle larger than 150°.

separation (d = 3.2 nm) corresponds to zero. To obtain the PMF of opening the RNA and hybrid hairpins, we performed 21 and 23 simulations, respectively, with different values of extension length d. These extension distances ranged from 1.27 and 1.35 nm for the RNA and hybrid, respectively, to 3.21 nm. At each distance, the system was equilibrated for 5.0 ns and data collected for an additional 60.0 ns. The starting configuration for the simulation at each value of d was obtained from an analogous free energy slow-growth simulation in which the equilibrated NMR folded hairpin was extended to d = 3.2 nm by 10^5 MD steps. The number of water molecules in the RNA and hybrid oligomer systems were 3866 and 4204, respectively. The calculations of the errors at each point of the PMF were obtained by integrating the error estimate of $1/2\langle\hat{r}_{12}\cdot(\vec{F}_1 - \vec{F}_2)\rangle_{\vec{r}_1,\vec{r}_2}$ at each d starting from the largest separation. The estimation of the error at each d followed a similar procedure to that applied in the previous section. Here, however, the standard deviation and the correlation time correspond to the value of $1/2\langle\hat{r}_{12}\cdot(\vec{F}_1 - \vec{F}_2)\rangle_{\vec{r}_1,\vec{r}_2}$.

III. RESULTS AND DISCUSSION

The results of the changes in the Gibbs energy for the alchemical mutations shown in Figure 1 are ΔG_1 = +271.6 ± 2.7 and ΔG_2 = +247.8 ± 1.4 kJ/mol. From this thermodynamic cycle, we obtain that the Gibbs energy change of unfolding the RNA hairpin to an extended structure relative to that of unfolding the hybrid is $\Delta G_{uf}^{RNA} - \Delta G_{uf}^{HYB}$ = +23.8 ± 4.1 kJ/mol. This means that it is easier to form a hairpin structure from the (all-ribose) RNA chain than from the hybrid chain (in which the riboses of the loop region were substituted with deoxyriboses). Because the difference between the RNA and hybrid is only the presence of the 2′-hydroxyl groups in the seven residues around the hairpin loop, the greater stability we observe for forming the RNA hairpin structure must be due to these hydroxyls.

What is the reason the RNA hairpin is more stable? In Table 1, we show the average number of hydrogen bonds between different groups of nucleotides for the RNA and hybrid hairpins. Considering the entire hairpin there are 0.5 more intrachain hydrogen bonds in the RNA oligomer. Larger numbers for the RNA are expected because the 2′-hydroxyls present in the loop region (3−9) of the RNA, but not present in the hybrid, can participate in hydrogen bondings. Indeed, a larger excess of hydrogen bonds, 1.3, are found within the nucleotides of the loop region of the RNA (which are partially compensated by a smaller number in the stem).

We decomposed the excess hydrogen bonds found in the loop region for the RNA hairpin due to the participation of the 2′-hydroxyls into contributions from different groups. The results shown in Table 1 indicate that the number of hydrogen bonds between the 2′-hydroxyls within the loop region is negligible (this is also true for hydrogen bonds between 2′-hydroxyls within the tails). However, the 2′-hydroxyls do form hydrogen bonds with other groups in the loop region; the largest number is with the sugar groups (with the O5′ and O4′), then with the bases, and the least with the phosphates (oxygens). The 2′-hydroxyls of the loop region also form, to some extent, hydrogen bonds with the tail nucleotides. In many cases, the hydrogen bonds formed between the 2′-hydroxyls and the sugar or the phosphate groups involve interactions between adjacent nucleotides. However, hydrogen bonds between a 2′-hydroxyl and a base on the opposite side of the loop also occur and are likely to contribute significantly to the stability of the loop. A very frequent hydrogen bond is between the 2′-hydroxyl of the fourth residue (first position of the pentaloop) and N6 or N7 of adenine 7, as well as between the 2′-OH of adenine 7 and N2 or N7 of the fourth residue. The former pattern of interaction is reminiscent of that formed within the UUCG tetraloop in which the 2′-hydroxyl group at position 1 (of the tetraloop) hydrogen bonds to the guanine base at position 4.[30,31,33,34] Additional frequent hydrogen bonding is established between the 2′-hydroxyl of the third nucleotide (position −1 of the pentaloop) and N6 of adenine 8 (fifth position of the pentaloop). These hydrogen bond patterns within the loop region are likely to render the stabilizing effect of the 2′-hydroxyls position dependent.

Table 1 displays also the corresponding number of hydrogen bonds in the untruncated 27-mer RNA hairpin obtained by NMR spectroscopy for the first structure deposited in the PDB file and used as the starting conformation for the simulations, as well as for the average over the 13 structures deposited. The agreement of the simulations results is better with the average over all NMR structures than with the first structure. The main difference is in the number of hydrogen bonds the 2′-hydroxyls in the loop region form with other groups. In the first NMR structure (starting conformation for the simulations), the 2′-hydroxyls hydrogen bond only to the phosphate group, whereas in some of the other NMR conformers the 2′-OH's also form hydrogen bonds with the sugars (up to two hydrogen bonds per conformer with the O5′) and the bases (up to one hydrogen bond per conformer with N6 of adenine 7) in agreement with what is obtained in the simulations. Some discrepancies with the NMR structures are expected because the oligonucleotide of the NMR experiment is a 27-mer, whereas in the simulations we considered an 11-mer truncated model. To assess the extent of the discrepancies arising from the use of our truncated model, we performed simulations of the 27-mer, without any restraints or constraints, for 40 ns. The results of the hydrogen bonds patterns are given in Table 1. A better agreement with the NMR structures is obtained, especially in the number of hydrogen bonds the 2′-hydroxyls in the loop region make with the bases and the phosphates. Nevertheless, there is always the possibility that some of the discrepancies that exist are also due to artifacts in the force field.

In order to gain a better insight of the loop opening processes, we also constructed the corresponding potential of mean forces. The PMFs for the RNA and hybrid hairpins are shown in Figure 2 (see Figure S2 in the Supporting Information for the convergence properties of these curves). For both curves, the energy values are calculated relative to that of the largest extension studied, d = 3.2 nm. The steep increase in the PMF, common to both curves, starting at $d \simeq 2.9$ nm is likely to reflect the entropic penalty associated with extending the nucleotide chains toward a fully extended conformation. However, it might also include some nonspecific reduction in the interactions between the nucleotide bases. In the range $1.7 \lesssim d \lesssim 2.9$, the curve of the RNA exhibits a plateau whereas that of the hybrid decreases gradually with decreasing d and displays a small local minimum at around d = 2.0 nm. From the analysis of the trajectories, in this range of extensions, we find that the hybrid chain tends to be more compact than that of the RNA chain and displays a smaller degree of fluctuations. This is shown in Figure 3, which exhibits the radius of gyration for both chains at an extension of $d \simeq 2.0$ nm. Such behavior can be expected because the absence of the 2′-hydroxyls makes the

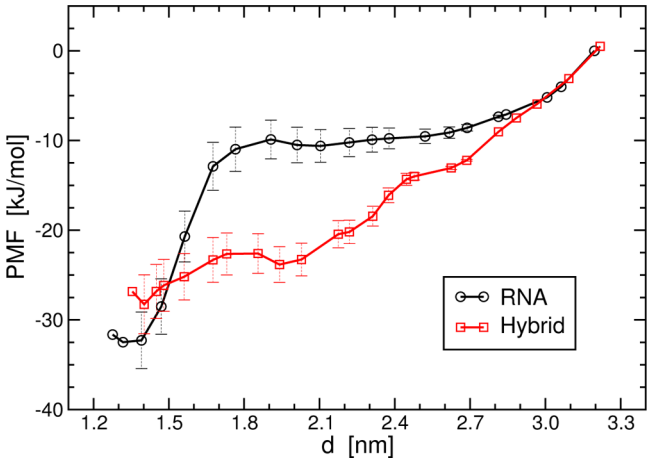


Figure 2. The potential of mean force of extending the RNA and hybrid 11-mer hairpins. The reaction coordinate, *d*, is defined as the distance between the C3′ atom of the third nucleotide and the C3′ atom of the ninth nucleotide. In both cases, the energy level of the curve is relative to that of the largest extension. The estimation of error bars at each point of the PMF is described in the Methods section.

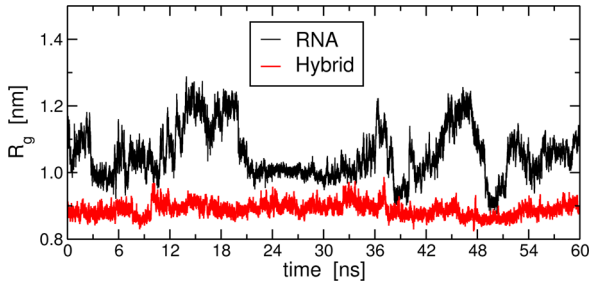


Figure 3. The radius of gyration of the RNA and hybrid oligonucleotides at an extension length of d = 2.01 and 2.03 nm, respectively, indicating that the conformations of the hybrid are more compact and with a smaller degree of fluctuations.

chain more hydrophobic and results in stronger, nonspecific, intranucleotide interactions. The decrease of the extension length of the oligonucleotides from around d = 1.7 nm toward the equilibrium state of the hairpin is characterized by a sharp decrease in the PMF for the RNA of approximately 20 kJ/mol. In contrast, for the hybrid only a 5 kJ/mol decrease of its PMF value is observed in the same range of d. This sharp decrease in the PMF for the RNA chain, or alternatively the large barrier encountered for extending the equilibrium hairpin conformation, is a manifestation of the kinetic stability of the RNA hairpin. At smaller extension lengths, d < 1.3−1.4 nm, both curves experience an increase in the PMF due to compression below the equilibrium state distance. We identify the equilibrium states of the hairpin conformation to be 1.39 and 1.40 nm for the RNA and hybrid, respectively. The difference in the Gibbs energy change of extending the RNA hairpin to any distance larger than 3.0 nm is +4.2 ± 6.5 kJ/mol relative to the hybrid. This is significantly lower than the value of 23.8 ± 4.1 kJ/mol obtained via the thermodynamic cycle shown in Figure 1. We are not exactly sure about the reason responsible for this discrepancy, although, it is likely that in one of these methods alternative conformations of the oligonucleotide were sampled not in their equilibrium probabilities. Nevertheless, the excess stability of the RNA hairpin structure relative to the hybrid can be considered to be inside the range of these two values. Note that the reaction coordinate chosen for the PMF may not be a natural folding/unfolding coordinate of the system. Instead, it represents the unfolding process via extending the hairpin from both sides of its tails.[52]

Snapshots of the oligonucleotides conformations for different extension lengths along the PMF reaction coordinate are shown in Figure 4. These extensions correspond to the equilibrium hairpin states (d = 1.4 nm), d = 2.0 nm, and the largest extension studied (d = 3.2 nm). The application of the distance restraints in the mutation free energy calculations or the distance constraints in the PMF calculations was not entirely enough to keep the Watson−Crick hydrogen bonds pattern between the base pairs A2−U10 and U3−A9 (the interaction

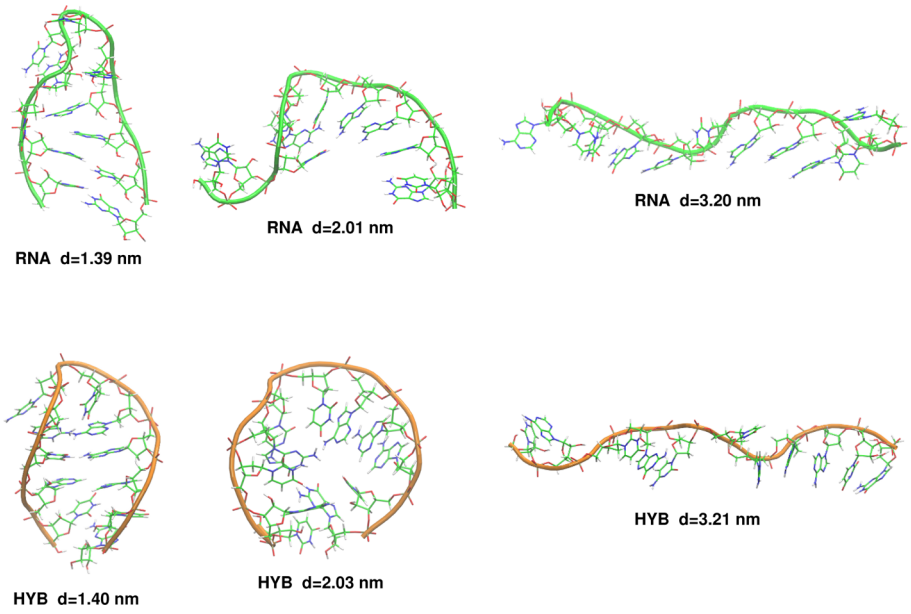


Figure 4. Instantaneous conformation of the RNA (top panel) and hybrid (lower panel) nucleotide chains at different extension lengths along the unfolding reaction coordinate.

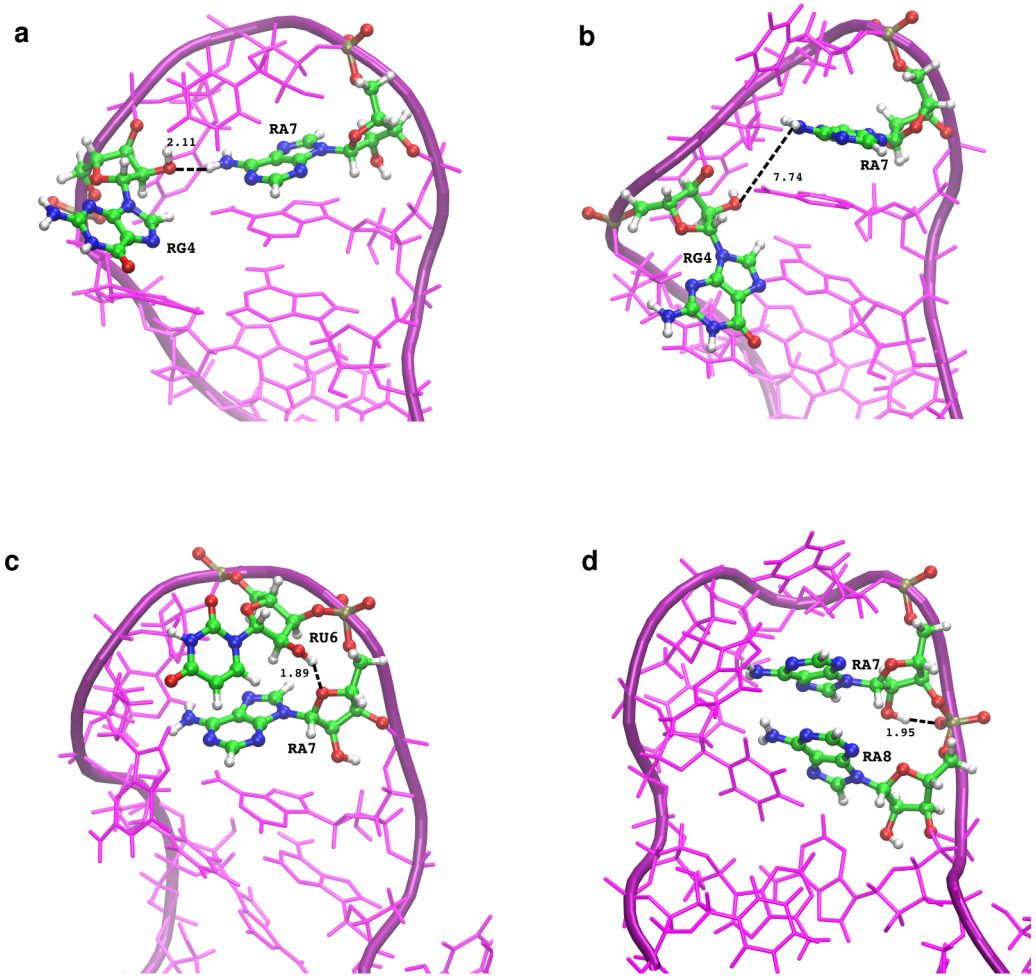


Figure 5. Snapshots from the simulations demonstrating the participation of the ribose 2′-hydroxyls of the loop region in forming intrahairpin hydrogen bonds. (a) A hydrogen bond between 2′-hydroxyl of guanine 4 (first position in the pentaloop) and N6 of adenine 7. (b) The conformation of the oligonucleotide when this hydrogen bond is broken. (c) A hydrogen bond between 2′-hydroxyl of uracil 6 and the ribose (O4′) of adenine 7. (d) A hydrogen bond between 2′-hydroxyl of adenine 7 and the phosphate oxygen of adenine 8.

between U1 and G11 is a wobble base pair) in both systems. Nevertheless, the conformation of the loop persisted throughout the trajectories. In Figure S3 in the Supporting Information we provide plots of the RMSD of the pentaloop backbone heavy atoms, after least-squares fitting to these group of atoms, as a function of time for the RNA (d = 1.39 nm) and hybrid (d = 1.40 nm). The least-squares fit was performed with respect to the first conformer deposited in the Protein Data Bank (NMR-1) and which was taken as the starting structure for the simulations, as well as with respect to three other conformers. Note that for both the RNA and hybrid, the lowest value of the RMSD is not when the fit is with respect to the first conformer but when it is with respect to the ninth conformer, suggesting that the choice of the first NMR conformer as the starting conformation for the simulations is so not important. The average values of the RMSD for the RNA relative to the different NMR conformers are in the range of 0.29−0.31 nm. Smaller RMSD values are obtained for the hybrid, 0.25−0.28 nm. Interestingly, although it is harder to unfold the RNA, the structure of the hybrid at $d \simeq$1.40 nm is slightly more similar (as judged by RMSD values) to the NMR structure.

Obviously, the hairpin is not static in solution and can adopt slightly different conformations in which the 2′-hydroxyls in the loop region are hydrogen bonded differently. We calculated the root mean squared deviations of the other 12 structures (deposited with the same PDB accession code) with respect to the first structure and found that the uracil at the middle of the pentaloop displays the largest value which decreases, almost symmetrically, toward the stem nucleotides in both (5′ and 3′) directions. Snapshots from the simulations of the hydrogen bonds formed in the loop region between a 2′-hydroxyl and a base, sugar, and phosphate are plotted in Figure 5. A hydrogen bond with a sugar or a phosphate is mostly observed between adjacent nucleotides because the interaction is between atoms of the backbone. It is not clear to what extent this will stabilize the hairpin. However, hydrogen bonds between the 2′-hydroxyls and the bases are often established between groups on opposite sides of the loop as well as with the middle nucleotide of the pentaloop and are likely to contribute directly to the stability of the loop. In Figure 5, we display an example of such a hydrogen bond, as well as a snapshot from the same trajectory in which this hydrogen bond is broken and accompanied by a deformation in the hairpin structure. Note that Table 1 indicates there are about 0.7 such hydrogen bonds in the RNA hairpin. Nevertheless, the difference in stability between the RNA and hybrid hairpins we found is in the range of 4.2−23.8 kJ/mol. Considering the strength of a hydrogen bond between a hydroxyl group and amines to be around 26

dx.doi.org/10.1021/ct3006216 | J. Chem. Theory Comput. 2013, 9, 1214−1221

kJ/mol,[53] the range of the hairpin stabilization by the 2′-OH is 0.2−0.9 times the strength of this hydrogen bond.

IV. CONCLUSIONS

In this paper, we showed that the ribose 2′-hydroxyls stabilize the hairpin conformation of an oligonucleotide containing GCUAA pentaloop. The stability is a result of hydrogen bonds the 2′-hydroxyls make with sugars, bases, and phosphates in the loop region and are likely to be position dependent. The sugar (through O5′ and O4′) and phosphate (oxygens) groups act as acceptors and are located in an adjacent nucleotide to the donating 2′-hydroxyl. Interactions between nonadjacent nucleotides involve hydrogen bonds with the bases and are likely to contribute substantially to the stability of the hairpin conformation. In many of these cases, the 2′-hydroxyls hydrogen bond with N6 or N7 of adenines located farther (toward the 3′ end) along the nucleotide chain. For example, the 2′-hydroxyl at the first position of the pentaloop makes a hydrogen bond with adenine at the fourth position. Similar interaction is also observed between the 2′-hydroxyl located one nucleotide before (toward the 5′ end) the pentaloop and adenine at the fifth position. We calculate, using the alchemical mutations free energy method, the magnitude of the hairpin excess stability to be around 23.8 kJ/mol compared with an analogous hairpin in which the ribose sugars in the loop region were substituted with deoxyribose sugars. We also constructed, independently, PMFs for opening these two hairpin structures to an extended conformation. These plots indicate that the 2′-hydroxyls also stabilize kinetically the folded hairpin structure. However, from these PMFs the thermodynamic excess stability of the RNA hairpin is calculated to be only 4.2 kJ/mol. Thus, a value inside the range of these two numbers is likely to represent the excess stability for a hairpin conformation of the RNA oligonucleotide.

The conclusion obtained experimentally that the ribose 2′-hydroxyls do not stabilize the GCAA tetraloop is based on the substitution of only the ribose of the first residue in this tetraloop to deoxyribose.[36] If indeed this conclusion will hold for the substitutions of the other riboses, then our results, and the results for the UUCG tetraloop, indicate that the ability of the ribose 2′-hydroxyls in stabilizing hairpin conformations depends on the sequence of the loop.

■ ASSOCIATED CONTENT

Supporting Information

Plots of $\langle \partial \mathcal{H}/\partial \lambda \rangle$ as a function of λ for mutating the 2′-OH to 2′-H, the convergence properties of the PMFs, and the RMSD for the pentaloop backbone atoms. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: r.zangi@ikerbasque.org.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been funded with support from the Spanish Ministry of Science and Innovation, MICINN (grant number CTQ2010-20297), and the Basque Government, the ETORTEK program (BioGUNE2010, Expt. # IE10-275). Technical and human support provided by SGIker (USED SERVICES) (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged. The authors also thankfully acknowledge the computer resources and technical assistance provided by the Barcelona Supercomputing Center—Centro Nacional de Supercomputación.

■ REFERENCES

(1) Stryer, L. *Biochemistry*, 3rd ed.; Freeman: New York, 1988.
(2) Bevilacqua, P. C.; Blose, J. M. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79−103.
(3) Deng, N.-J.; Cieplak, P. *J. Chem. Theory Comput.* **2007**, *3*, 1435−1450.
(4) Miller, J. L.; Kollman, P. A. *J. Mol. Biol.* **1997**, *270*, 436−450.
(5) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493−506.
(6) Garcia, A. E.; Paschek, D. *J. Am. Chem. Soc.* **2008**, *130*, 815−817.
(7) Villa, A.; Widjajakusuma, E.; Stock, G. *J. Phys. Chem. B* **2008**, *112*, 134−142.
(8) Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2008**, *130*, 9676−9678.
(9) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(10) Sarkar, K.; Nguyen, D. A.; Gruebele, M. *RNA* **2010**, *16*, 2427−2434.
(11) Deng, N.-J.; Cieplak, P. *Biophys. J.* **2010**, *98*, 627−636.
(12) Svoboda, P.; Cara, A. D. *Cell. Mol. Life Sci.* **2006**, *63*, 901−908.
(13) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. *J. Mol. Biol.* **1989**, *209*, 459−474.
(14) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 111−113.
(15) Butcher, S. E.; Pyle, A. M. *Acc. Chem. Res.* **2011**, *44*, 1302−1311.
(16) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678−1685.
(17) Tamura, M.; Holbrook, S. R. *J. Mol. Biol.* **2002**, *320*, 455−474.
(18) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4899−4903.
(19) Pyle, A. M.; Cech, T. R. *Nature* **1991**, *350*, 628−631.
(20) Bevilacqua, P. C.; Turner, D. H. *Biochemistry* **1991**, *30*, 10632−10640.
(21) Pyle, A. M.; Murphy, F. L.; Cech, T. R. *Nature* **1992**, *358*, 123−128.
(22) Strobel, S. A.; Cech, T. R. *Nat. Struct. Biol.* **1994**, *1*, 13−17.
(23) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 68−74.
(24) Müller, U. F.; Bartel, D. P. *Chem. Biol.* **2003**, *10*, 799−806.
(25) von Ahsen, U.; Green, R.; Schroeder, R.; Noller, H. F. *RNA* **1997**, *3*, 49−56.
(26) Chastain, M.; Tinoco, I. *Biochemistry* **1992**, *31*, 12733−12741.
(27) Draper, D. E. *J. Mol. Biol.* **1999**, *293*, 255−270.
(28) Landt, S. G.; Tipton, A. R.; Frankel, A. D. *Biochemistry* **2005**, *44*, 6547−6558.
(29) Quigley, G. J.; Rich, A. *Science* **1976**, *194*, 796−806.
(30) Allain, F. H.-T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333−353.
(31) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35−42.
(32) James, J. K.; Tinoco, I. *Nucleic Acids Res.* **1993**, *21*, 3287−3293.
(33) Sakata, T.; Hiroaki, H.; Oda, Y.; Tanaka, T.; Ikehara, M.; Uesugi, S. *Nucleic Acids Res.* **1990**, *18*, 3831−3839.
(34) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 251−265.
(35) Williams, D. J.; Boots, J. L.; Hall, K. B. *RNA* **2001**, *7*, 44−53.
(36) SantaLucia, J.; Kierzek, R.; Turner, D. H. *Science* **1992**, *256*, 217−219.
(37) Stefl, R.; Allain, F. H.-T. *RNA* **2005**, *11*, 592−597.
(38) Aruscavage, P. J.; Bass, B. L. *RNA* **2000**, *6*, 257−269.

(39) *Maestro*, version 9.2; Schrödinger, LLC: New York, 2011.
(40) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(41) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(42) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(43) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(44) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(45) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(46) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(48) Guimarães, C. R. W.; Kopecky, D. J.; Mihalic, J.; Shen, S.; Jeffries, S.; Thibault, S. T.; Chen, X.; Walker, N.; Cardozo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18139−18146.
(49) Hassan, S. A. *J. Phys. Chem. B* **2004**, *108*, 19501−19509.
(50) Pangali, C. S.; Rao, M.; Berne, B. J. In *Computer Modeling of Matter*, Lykos, P., Ed.; ACS: Washington, DC, 1978; ACS Symposium Series No. 86, pp 32−34.
(51) Zangi, R. *J. Phys. Chem. B* **2011**, *115*, 2303−2311.
(52) Colizzi, F.; Bussi, G. *J. Am. Chem. Soc.* **2012**, *134*, 5173−5179.
(53) Kolev, S. K.; Petkov, P. S.; Rangelov, M. A.; Vayssilov, G. N. *J. Phys. Chem. A* **2011**, *115*, 14054−14068.