The Essential Role of Stacking Adenines in a Two-Base-Pair RNA Kissing Complex

William Stephenson,[†,‡] Papa Nii Asare-Okai,[§] Alan A. Chen,[#] Sean Keller,[∥] Rachel Santiago,[∥] Scott A. Tenenbaum,[‡,⊥] Angel E. Garcia,[⊥,#] Daniele Fabris,[§,∥,⊥] and Pan T. X. Li*[,§,⊥]

[†]Nanoscale Engineering Graduate Program, [‡]College of Nanoscale Science and Engineering, [§]Department of Chemistry, [∥]Department of Biological Sciences, and [⊥]The RNA Institute, University at Albany, State University of New York, 1400 Washington Avenue, Albany, New York 12222, United States

[#]Department of Physics and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, New York 12180, United States

Ⓢ *Supporting Information*

ABSTRACT: In minimal RNA kissing complexes formed between hairpins with cognate GA<u>C</u>G tetraloops, the two tertiary GC pairs are likely stabilized by the stacking of 5′-unpaired adenines at each end of the short helix. To test this hypothesis, we mutated the flanking adenines to various nucleosides and examined their effects on the kissing interaction. Electrospray ionization mass spectrometry was used to detect kissing dimers in a multiequilibria mixture, whereas optical tweezers were applied to monitor the (un)folding trajectories of single RNA molecules. The experimental findings were rationalized by molecular dynamics simulations. Together, the results showed that the stacked adenines are indispensable for the tertiary interaction. By shielding the tertiary base pairs from solvent and reducing their fraying, the stacked adenines made terminal pairs act more like interior base pairs. The purine double-ring of adenine was essential for effective stacking, whereas additional functional groups modulated the stabilizing effects through varying hydrophobic and electrostatic forces. Furthermore, formation of the kissing complex was dominated by base pairing, whereas its dissociation was significantly influenced by the flanking bases. Together, these findings indicate that unpaired flanking nucleotides play essential roles in the formation of otherwise unstable two-base-pair RNA tertiary interactions.
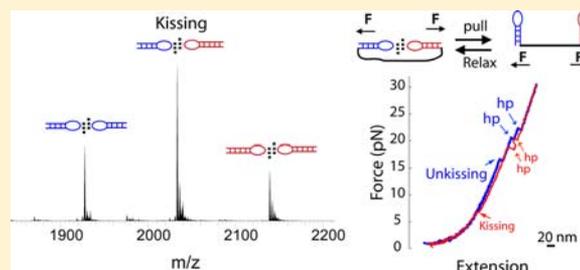
■ INTRODUCTION

Base pairing plays a predominant role in RNA folding, particularly in tertiary structures.[1,2] Seminal studies showed that stable helical duplexes require at least three Watson−Crick pairs.[3−5] However, a two-base-pair duplex formed between self-complementary GpCpA triribonucleotides was stabilized by stacking of the 3′-end unpaired adenines onto the adjacent base pairs.[6] More recently, a hairpin with GACG loop was shown to form a two-base-pair intermolecular complex that exhibited excellent stability even in the absence of divalent ions under mild ionic conditions (Figure 1a).[7,8] The NMR structure of this so-called minimal kissing complex exhibited two GC pairs with the 5′-adenines stacked onto the intermolecular base pairs.[7] The possibility that the stacked adenines could contribute to stabilize the kissing complex, however, was not consistent with the observation that 5′-unpaired nucleotides provided little or no stabilization to RNA secondary structure.[9−13] The apparent contradiction led us to question whether the two GC base pairs may be truly sufficient to stabilize the tertiary interaction, or whether the 5′-unpaired adenines may play a greater role in the kissing complex than generally thought.

A recent molecular dynamics (MDS) study suggested that adenine stacking may be critical for the stability of the kissing complex.[14] In simulated dissociation of the complex under mechanical tension, a highly stable intermediate was predicted in which both base pairs turned perpendicular to the helical axis. Interestingly, the flanking adenines rotated with the kissing pairs and kept the same stacked positions observed in the corresponding high-resolution NMR structure (Figure 1b).[7] The GC base pairs remained intact until both adenine bases flipped out of their stacked positions, at which point the kissing complex quickly dissociated. The simulations further predicted that dissociation of the kissing complex would be greatly enhanced by replacing the flanking adenines by uridines, due to their inferior stacking proficiency.

The relatively weak minimal kissing complex provides no distinctive signals in UV melting (data not shown) or gel mobility shift assay,[15] thus precluding its thermodynamic characterization by established methods. In this study, we employed a three-pronged approach to explore the structural determinants of the loop−loop interaction by systematically mutating the unpaired adenine (underlined) in GA<u>C</u>G tetraloop hairpins. The effects of substituting adenine with different pyrimidine and purine analogues were evaluated experimentally by electrospray ionization mass spectrometry
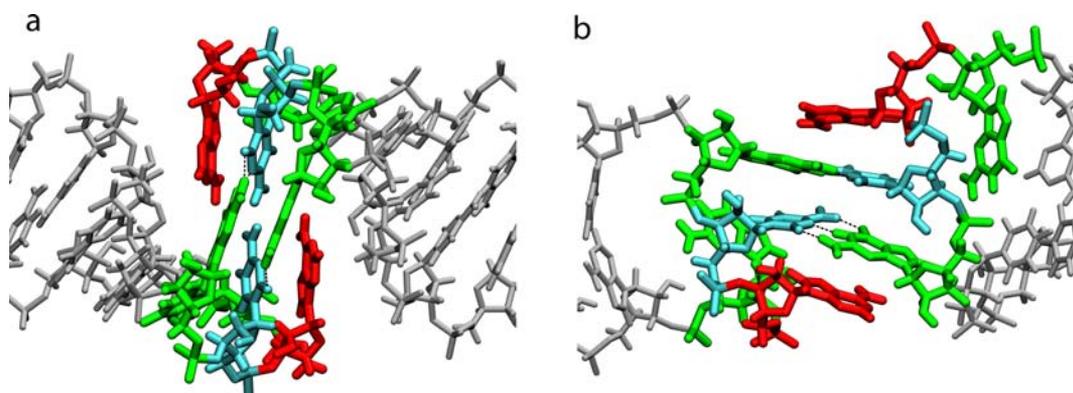
Figure 1. Adenosine stacking in the GACG-GACG kissing complex: (a) side view of the NMR structure;[7] (b) a stable intermediate during mechanical unfolding observed in MDS.[14] The hairpin stems are in gray, and the GACG tetraloops in color, with the unpaired adenosines in red. The dashed lines indicate hydrogen bonds within the interstrand GC base pairs.

(ESI-MS)[16,17] and single-molecule force spectroscopy.[18] It is well-established that, under proper conditions, ESI-MS is capable of taking accurate snapshots of binding equilibria without perturbation of the equilibrium positions in solution.[19,20] This favorable feature allows not only to monitor multiple simultaneous equilibria in solution,[21,22] but also to determine the partitioning between free and bound species at equilibrium, which are comparable to dissociation constants obtained by traditional methods.[23,24] Optical tweezers-based single-molecule force spectroscopy can be used to measure RNA unfolding/refolding kinetics,[8,25] and to observe specific behaviors of individual molecules, which may be masked by the ensemble averaging in bulk techniques.[26,27] Finally, molecular dynamic simulations (MDS) were employed to analyze energetic contributions from hydrophobic and electrostatic forces. The concerted utilization of these approaches offered the opportunity to investigate the structural determinants of the minimal kissing complex and to obtain new insights into the role of flanking 5′-nucleotides.

■ RESULTS AND DISCUSSION

Detecting Intermolecular Kissing Complexes by ESI-MS. Nucleic acid duplexes were among the first noncovalent complexes detected intact by ESI-MS,[28,29] which demonstrated the possibility of preserving relatively weak interactions during the desorption of analytes into the gas phase. In our hands, this platform enabled the investigation of intact duplexes formed by the 20-mer DNA/RNA oligonucleotides that mimicked the HIV-1 polypurine tract,[22,30] or by the 6-base-pair kissing interaction of the dimerization initiation site domain of the HIV-1 genome.[31,32] This approach was employed here to assess the formation of the minimal kissing complexes involving far fewer intermolecular pairing interactions.

Depending on experimental conditions, each individual construct may be capable of forming homodimeric complexes in either loop–loop kissing or extended duplex forms (Figure 2a). In order to minimize mutual hybridization, an additional GACG construct was designed with different stem sequences that would enable the formation of heterodimers exclusively linked by a kissing complex (Figure 2b). When the two cognate constructs were mixed in a 1:1 molar ratio (see Experimental Methods), a mixture of homo- and heterodimers was generated in solution by simultaneous binding equilibria, which were readily resolved by ESI-MS (Figure 3a). The validity of the analytical approach was confirmed by analyzing a control
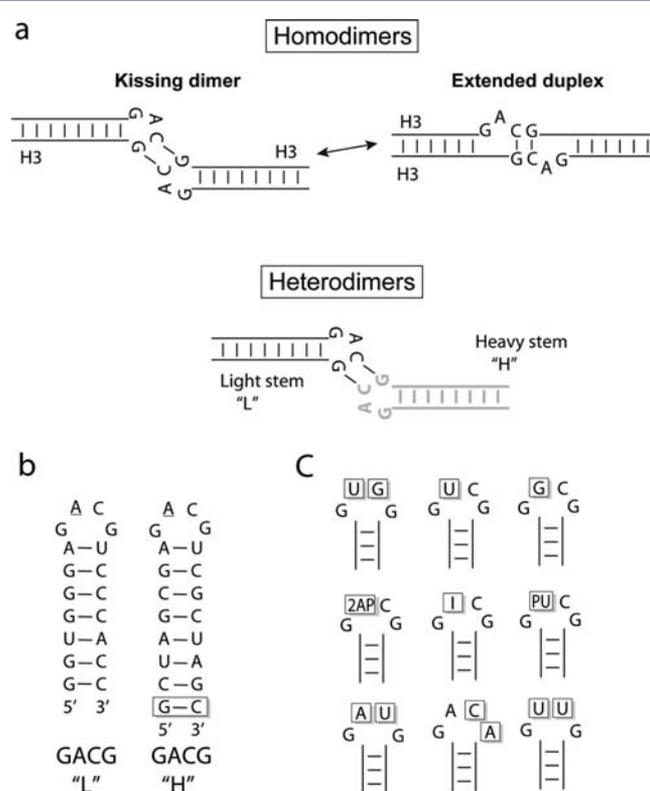
Figure 2. Design of the ESI-MS experiments. (a) Homodimers of the constructs employed in the study can readily adopt either loop–loop kissing or extended duplex conformations. In contrast, a heterodimer can only exist in the kissing conformation. (b) GACG tetraloop with either a light or a heavy stem. (c) Various loops used in mass spectrometry. The mutated bases are highlighted with a box. "2AP" stands for 2-amino purine, "I" for inosine, and "Pu" for purine.

construct with a GUGG loop devoid of self-complementarity (Figure S1a), which exhibited the expected monomeric hairpin and absence of dimeric complexes. Therefore, when the data provided by the GACG sample were examined (Figure 3a), it was possible to assign unambiguously the heterodimer with mass of 12 206.76 Da (12 206.52 Da calculated) to the kissing form, and the two homodimers with masses of 11 571.66 and 12 841.92 Da (11 571.14 and 12 841.9 Da calculated) (Figure 2a). The partitioning between free and bound species at equilibrium in solution was determined from the respective signal intensities according to accepted procedures.[23,24] This

5603

dx.doi.org/10.1021/ja310820h | *J. Am. Chem. Soc.* 2013, 135, 5602–5611
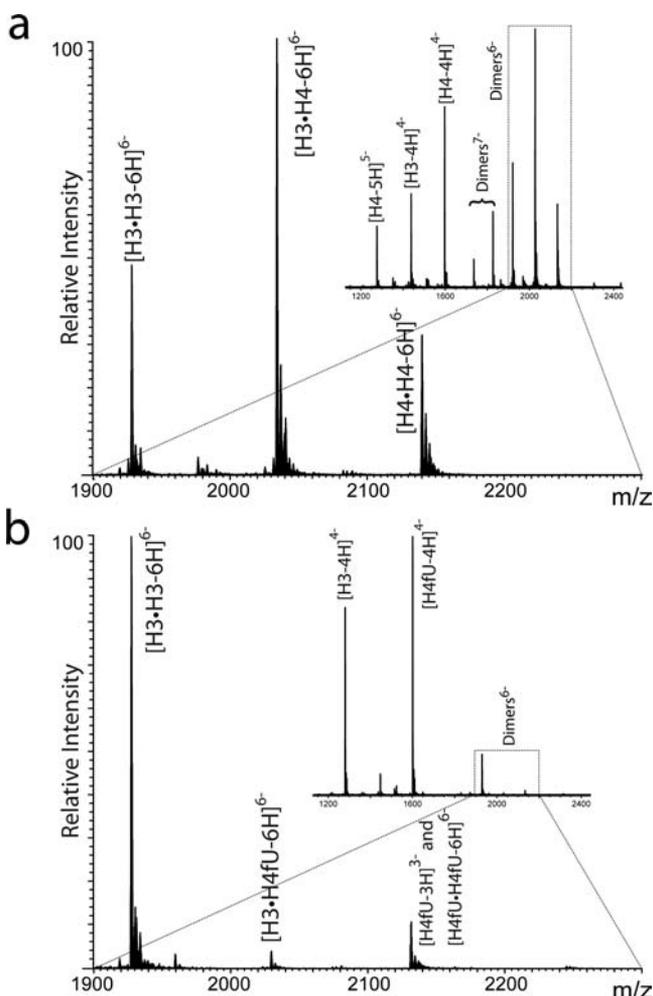
Figure 3. ESI-MS spectra of kissing interaction. Spectra obtained from samples containing (a) 1:1 GACG loops with light (H3) and heavy (H4) stems or (b) 1:1 GACG (H3) and GUCG (H4fU) loops in 150 M ammonium acetate (see Experimental Methods for conditions). Expanded views show the regions containing the 6− charge states of the dimeric species. Insets show full range spectra containing the corresponding monomeric species also. All signals were isotopically resolved to enable unambiguous assignment of charge and association states (see Figure 1S for details). The following dimeric species were detected: H3·H3 (GACG·GACG, 11 571.66 Da exp., 11 571.14 calculated from sequence); H3·H4 (GACG·GACG, 12 206.76 Da exp., 12 206.52 Da from seq.); H4·H4 (GACG·GACG, 12 841.92 Da exp., 12 841.90 Da from seq.); H3·H4fU (GACG·GUCG, 12 183.72 Da exp., 12 183.48 Da from seq.). Heterodimers were indicated by arrows.

treatment enabled us to estimate the abundance of GACG heterodimer—the kissing form—at 27 ± 5% of total RNA in the sample (Table 1).

The effects of purine to pyrimidine mutation at the second tetraloop position were evaluated in a similar fashion. A pair of hairpins with GUCG loops (Figure 2c) provided only negligible signals for the corresponding heterodimer (Table 1). Furthermore, when the GUCG hairpin was mixed with GACG, very weak signal was observed for the corresponding heterodimer (Figure 3b) even when ionic strength was progressively increased from 150 mM to 1 M ammonium acetate to favor dimer formation (Table 1). These results clearly showed that the substitution of even a single adenine with uridine had a significant destabilizing impact on the kissing

Table 1. Percentage of Heterodimer over Total RNA Detected by ESI-MS[a]

Hairpins	% Heterodimer	
GACG·GACG	27 ± 5	Adenine
GUCG·GUCG	0.3 ± 0.1	Uridine
GACG·GUCG	0.2 ± 0.1	
GACG·GUCG , 1M salt	0.7 ± 0.1	
GPuCG·GACG	6.0 ± 0.5	Purine
G2ApCG·GACG	11 ± 0.5	2-aminopurine
GICG·GACG	2.0 ± 0.2	Inosine
GGCG·GACG	2.8 ± 0.2	
GAUG·GACA	1.6 ± 0.3	*
GUUG·GACA	0.9 ± 0.3	*
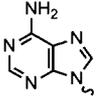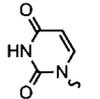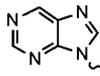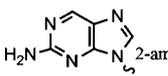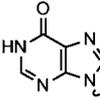
[a]In each sample, the two hairpins were mixed in 1:1 ratio. The sequences of the different constructs are shown in Figure 2. The chemical structures of the different nucleobases are shown in the last column. * These loop pairs can form UG/AC base pairs.

complex. The inhibitory effects by this mutation can be directly traced to the ability of flanking nucleobases to stack onto the adjacent base pairs. In the stacked conformation, the fused ring system of adenine extends on top of the underlying hydrogen bonds to effectively establish interstrand stacking (Figure 1a).[7] In contrast, the single ring of pyrimidine cannot reach far enough to protect the underlying hydrogen bonds from the solvent.[14]

We next explored whether the stabilizing effects of stacked adenines were limited to GC pairs by designing two hairpins with GAUG and GACA loops, which would be capable of establishing intermolecular AU/GC pairs. When mixed together, these constructs provided only 1.6 ± 0.3% of kissing heterodimer (Table 1), consistent with the intrinsically lower stability of AU versus GC base pairs. However, when a GUUG construct was mixed with GACA, the A-to-U mutation further reduced heterodimer formation to only 0.9 ± 0.3% of total RNA (Table 1). Such stabilizing effect of flanking adenines on the adjacent AU pair is evidence of a more general role for the 5′-unpaired adenines in loop−loop tertiary interactions.

Additional purine analogues were investigated to elucidate the influence of ring substituents on the properties of stacked adenines. A mixture of GPuCG and GACG constructs, which contained a purine (Pu) nucleotide devoid of the 6-amino
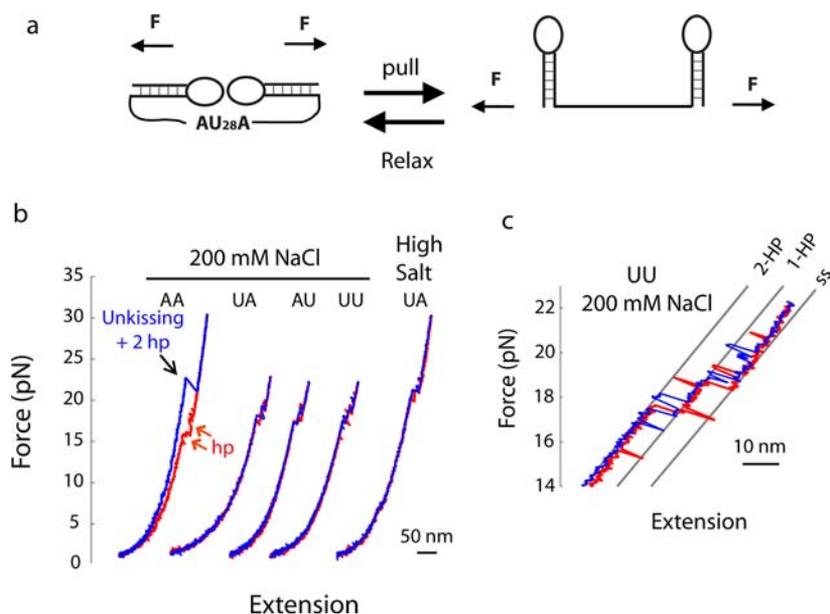
Figure 4. Effect of adenosine to uridine substitutions examined by single-molecule mechanical unfolding. (a) A single kissing RNA complex was mechanically unfolded into a two-hairpin conformation. (b) Force−extension curve of the AA, AU, UA, and UU constructs at (un)loading rates of 4 pN/s. The unfolding curve is shown in blue and refolding in red. The high salt solution contained 1 M NaCl and 20 mM MgCl₂ at pH 7.4. (c) An enlarged view of the unfolding and refolding of the UU RNA. Gray lines represented three states corresponding to the 2-hairpin, 1-hairpin, and single-stranded structures.
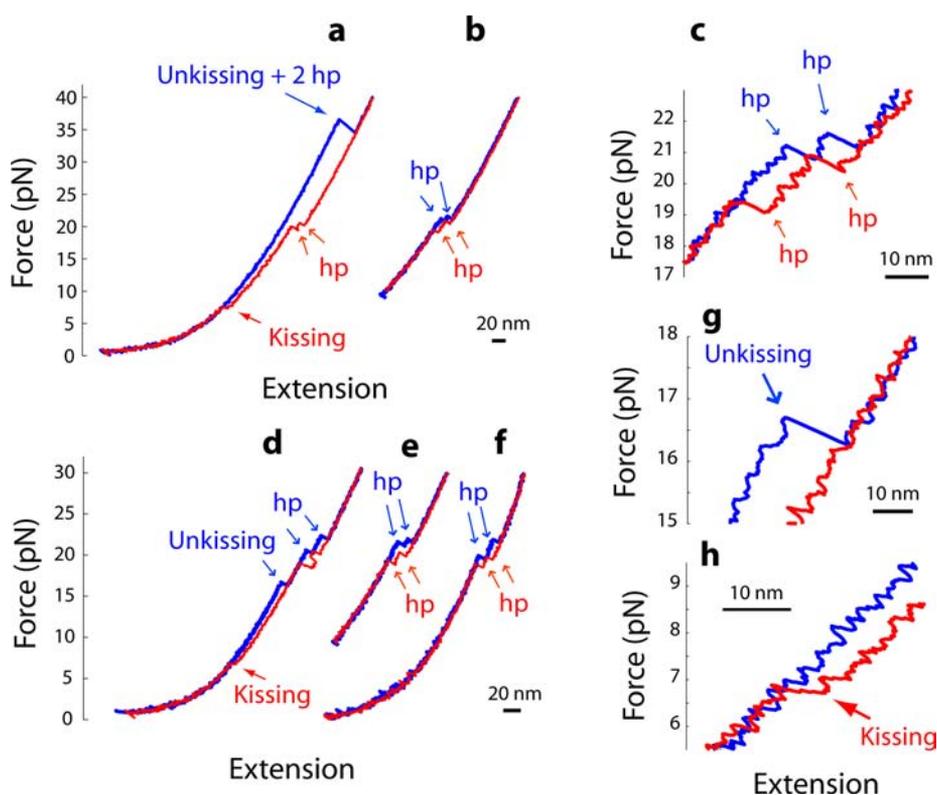
Figure 5. Effect of adenosine to guanosine substitutions in the tetraloop. Force−extension curves of the AA (a and b), GA (d and e), and GG (f) RNAs in 1 M NaCl at pH 7.4. In b and e, the lowest relaxation force was set at 10 pN to prevent kissing formation. (c) An enlarged view of the hairpin folding in trajectory b. (g and h) Enlarged views of the unkiss and kissing transitions in d, respectively.

group, provided 6.0 ± 0.5% of the heterodimer (Table 1). Reintroducing an amino group at the 2′-position of the ring system (2Ap) partially restored heterodimer formation to 11.0 ± 0.5%, suggesting a positional effect by the amino group. The presence of an electron-withdrawing oxygen on the inosine (I)

system reduced heterodimer formation in the corresponding GICG/GACG sample to only ~2.0 ± 0.2% of total RNA. This reduction was only incompletely compensated by the concomitant reintroduction of a 2-amino group in the GGCG/GACG couple, which produced 2.8 ± 0.2% hetero-

5605

dx.doi.org/10.1021/ja310820h | J. Am. Chem. Soc. 2013, 135, 5602−5611
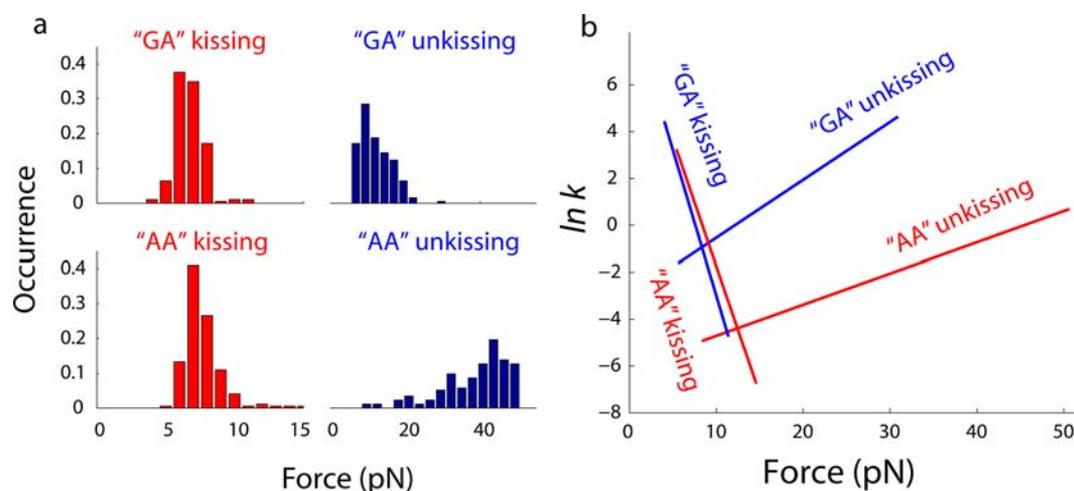
Figure 6. Formation and dissociation of kissing interaction in the GA and AA RNAs. (a) Distributions of unkissing (blue) and kissing forces. Occurrences were normalized by the total number of observations (186 for GA and 173 for AA). The mean kissing force was 6.7 pN for GA and 7.7 pN for AA. The mean unfolding force was 12.2 pN for GA and 38.8 pN for AA. (b) Unkissing and kissing rates for the GA (blue) and AA (red) RNAs.

dimer (Table 1). There was a clear indication that electron-withdrawing groups reduced the stabilizing effects over the Pu mutant devoid of ring substituents. Conversely, electron-donating groups showed increasing stabilizing effects over Pu levels in a position dependent fashion.

Observing Intramolecular Kissing Interactions within a Single RNA Molecule. Thermodynamics and kinetics of kissing interactions were investigated in detail by employing optical tweezers to monitor folding trajectories of single RNA molecules. We designed a series of RNAs capable of folding into two distinct 9-base-pair hairpins spaced by 30 unpaired nucleotides (Figure 4a). The stem sequences were designed to prevent their annealing into a long hairpin structure analogous to an extended duplex dimer. The U-rich linker was chosen to be sufficiently long and elastic to allow the formation of an intramolecular kissing contact between the hairpins. In addition, the linker allows multiple rounds of pulling and relaxing a single RNA molecule.[8] Using optical tweezers, a single RNA molecule is repeatedly stretched and relaxed from its 5′- and 3′-ends, whereas structural transitions are inferred from force−extension patterns.[33] Formation and especially dissociation of kissing interactions display signals that are clearly distinguishable from those generated by hairpin un/refolding.[8,25,34]

We first repeated mechanical unfolding of the GACG•-GACG (AA) loop pair in 20 mM HEPES at pH 7.5 and 21 °C, in the presence of 200 mM NaCl.[8] The ionic strength was comparable to that employed in our ESI-MS analyses and in the original NMR study.[7] The unfolding trajectory (1st blue curve in Figure 4b, AA) provided a characteristic triple-transition "rip" at ~22 pN, which was previously assigned to the dissociation of the kissing complex ("unkissing") followed by immediate unfolding of the hairpin structures.[8] As the hairpins and kissing complex refolded at much lower forces than the unfolding, the RNA structures were unlikely to refold once disrupted. Therefore, the force−extension curve of the irreversible unfolding of a kissing complex displayed characteristic hysteresis (Figure 4b, AA).[8,25,34,35] The change in extension during the triple-transition, ΔX, was >25 nm. This value agreed with the sum of ΔX by the individual unfolding of each hairpin and the unkissing at the rip force (Figure S2).[8] On

force relaxation, the two hairpins refolded at between 16 and 18 pN (1st red curve in Figure 4b, AA), as indicated by multiple back-and-forth transitions that reflected structural bistability between hairpins and single-stranded RNA at such forces.[33,36,37] Formation of the kissing complex occurred at below 5 pN with little change in extension. However, if the lowest relaxation force was set at above 5 pN, the subsequent pulling curve displayed only unfolding of the hairpins (Figure 5b for an example).[8]

The GACG·GUCG (AU), GUCG·GACG (UA), and GUCG·GUCG (UU) systems were compared with the wild-type GACG·GACG (AA) to evaluate the effects of purine to pyrimidine mutations. The force−extension curves of these mutants were significantly different from those of the AA construct in that they lacked both the triple-transition and the characteristic hysteresis of loop−loop dissociation (Figure 4b). Instead, these constructs showed multiple transitions at forces between 16 and 20 pN, which are typical of the unfolding/refolding of two separate hairpins. A closer look at these transitions revealed three distinct states corresponding to a 2-hairpin, a 1-hairpin, and a single-stranded conformation (Figure 4c). The ΔX observed for the unfolding of each hairpin was approximately 8 nm, consistent with the values calculated for the corresponding hairpin structures (Figure S2). The unfolding patterns of the three mutants were strikingly similar to those displayed by GAAA tetraloop hairpins incapable of forming a kissing complex,[37] or the AA construct in experiments that precluded kissing formation (Figure 5b). To further favor kissing interactions in mutant RNAs, we performed a much slower relaxation, held the single molecules at zero force for up to one minute to increase refolding time, and raised the salt concentration to 1 M NaCl and 20 mM MgCl$_2$ (Figure 4b, "high salt"). However, we observed no evidence of a kissing formation under all these conditions. These results were consistent with observations from ESI-MS that the presence of uridine in flanking positions severely affected the formation of two-base-pair kissing complexes. The fact that elevated ionic strength could not compensate for an increased exposure of the terminal pairs further reinforced the importance of the protection effects exerted by stacked adenines.
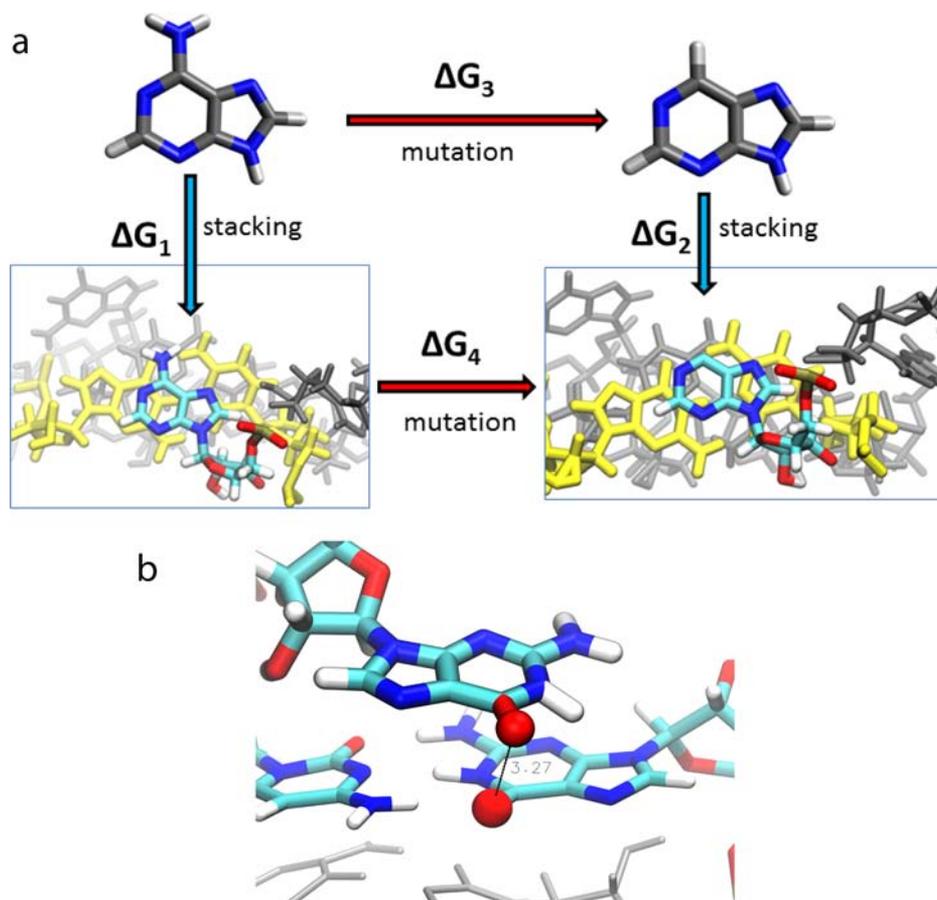
Figure 7. Molecular dynamics simulation of purine mutants. (a) Thermodynamic cycle for computing $\Delta\Delta G_{stack}$ for mutating. The bottom panels show the energy minimized structures. Vertical arrows ΔG_1 and ΔG_2 represent the free-energy contribution of base stacking at the 5′ flanking position of the kissing-loop "bridging complex" described in previous work.[14] Horizontal arrows ΔG_3 and ΔG_4 represent the free energy cost for mutating the base moiety in aqueous solvent or in the context of the kissing complex, respectively. By invoking a thermodynamic cycle, the change in stacking free energy upon mutation $\Delta\Delta G_{stack} = \Delta G_2 - \Delta G_1$ is equivalent to $\Delta G_4 - \Delta G_3$. The latter is calculated using a series of equilibrium molecular dynamics simulations containing "alchemical" intermediates of each mutated base, as reported in Table 2. (b) Side view of GGCG loop shows the close proximity between the 6-carbonyl of the flanking guanine and the carbonyl of the underlying guanine.

The folding of GG̲CG·GA̲CG (GA) and GG̲CG·GG̲CG (GG) constructs was compared to that of wild-type AA in 1 M NaCl to evaluate the effects of guanine at the second loop position. At this relatively high ionic strength, the AA RNA displayed triple-transitions at ∼38 pN on average (Figure 5a). In the refolding curves, these conditions enabled us to recognize discrete signals for hairpin refolding at >20 pN and kissing formation at ∼7 pN. The assignments of AA transitions were validated by setting the lowest relaxation force to 10 pN to prevent folded hairpins from kissing; the subsequent pulling trajectories showed only hairpin unfolding (Figure 5b,c). The GA molecule displayed three unfolding transitions (Figure 5d,g). The first took place mostly below 20 pN, followed by two consecutive events between 20 and 23 pN. The latter two reflected unfolding of the hairpins, whereas the first was suggestive of unkissing. On force relaxation, the hairpins first refolded at >20 pN, followed by a transition at ∼7 pN (Figure 5d,h), which was similar to the kissing formation observed for the AA RNA. These observations suggested that the GA molecule followed the same hierarchical unfolding and refolding pathways of the AA RNA. This hypothesis was tested by allowing the force to relax to only 10 pN to prevent the last event from taking place (Figure 5e). As expected, subsequent pulling showed only two transitions corresponding to hairpin

unfolding. Furthermore, when the force−extension curves were superimposed, all three unfolding transitions in GA matched reasonably well with the triple-transition in the AA RNA (Figure S4). Together, these results confirmed that the first unfolding and last refolding transitions in GA reflected the unkissing and kissing, respectively. In contrast, the GG RNA showed only unfolding and refolding of the hairpins (Figure 5f).

It was surprising that the kissing of the AA and GA RNAs occurred in a similar force range, but their average dissociation forces differed by 26.6 pN (Figure 6a). The lower dissociation forces exhibited by GA indicated that the mutant kissing structure dissociated significantly faster than its AA counterpart. This observation was supported also by dissociation/association kinetics extrapolated from force distributions using methods described previously,[8,26] which showed similar kissing but drastically different unkissing rates for the two RNAs (Figure 6b). Therefore, these results confirmed that the effects of flanking bases are predominant during dissociation of the kissing complex, as previously suggested.[14] In addition, our data clearly demonstrated that formation of the kissing complex is not significantly affected by the flanking bases, but was instead largely dependent on the tertiary base pairs.

Table 2. Energetic Cost for Purine Mutations in 250 mM Monovalent Salt[a]

base mutation	$\Delta\Delta G_{VdW}$	$\Delta\Delta G_{Coul}$	$\Delta\Delta G_{Stack}$	ratio$_{Heterodimer}$	$\Delta\Delta G_{pulling}$
adenine → purine	+1.2 ± 0.2	−0.14 ± 0.01	+1.1 ± 0.2	0.22x (27% → 6.0%)	-
purine → 2-aminopurine	+0.03 ± 0.5	−0.5 ± 0.2	−0.5 ± 0.5	1.83x (6.0% → 11%)	-
2-aminopurine → guanine	−0.31 ± 0.04	+1.2 ± 0.4	+0.9 ± 0.4	0.25x (11% → 2.8%)	-
adenine → guanine	0.9 ± 0.5	0.6 ± 0.4	+1.5 ± 0.5	0.1x (27% → 2.8%)	4 ± 2

[a]In both simulation and experiment, only one out of the two adenines was mutated. $\Delta\Delta G_{Stack}$ is the sum of Van der Waal ($\Delta\Delta G_{VdW}$) and Coulomb ($\Delta\Delta G_{Coul}$) forces computed by molecular dynamics simulations. Experimental results for the mutations from mass spectrometry (Ratio$_{Heteromer}$) and single-molecule force spectroscopy ($\Delta\Delta G_{pulling}$) were also listed. $\Delta\Delta G$ in unit of kcal/mol.

The stochastic nature of unpaired base stacking was clearly evident in the unkissing force distributions, especially in those of rare events (Figure 6a). The distribution of the GA dissociation force, with a mean value of 12.2 pN from a total of 186 trajectories, showed a tail toward high forces (Figure 6a). The highest dissociation force was observed at 31 pN, with five other transitions occurring at above 20 pN. In contrast, the force distribution of AA dissociation displayed a tail toward low forces. Four out of 173 trajectories showed an unkissing transition below 15 pN. These rare events suggested that adenines at the 5′-position favor the stacked conformation; their occasional unstacking leads to unstable structures that generated a low force tail in the force distribution. In contrast, guanines at the same position are relatively more flexible; their occasionally long stay in stacked conformations may result in stable kissing complexes that require relatively high forces to unfold.

The equilibrium force of the kissing interaction, at which unkissing and kissing rates are equal, was 8.3 pN for GA and 12.4 pN for AA. The folding energy of kissing complexes was further estimated from the reversible mechanical work done to unfold the kissing complex at the equilibrium force.[8,26] $\Delta G_{0pN, 1 M NaCl}$ of 8 ± 2 kcal/mol was obtained for GA and 14 ± 2 kcal/mol for AA, thus indicating that a single G to A substitution decreased the stability of the two-base-pair kissing complex by a net 6 kcal/mol in 1 M NaCl. When both flanking adenines were replaced, the $\Delta G_{0pN, 1 M NaCl}$ calculated for the GG kissing interaction turned out to be merely 2 kcal/mol. Such a low value agreed well with the observation that the GG construct displayed no signal corresponding to either kissing formation or dissociation (Figure 5f).

Dissecting Stacking Energetics by Molecular Dynamics Simulation. The mechanism by which the purine system and its functional groups may affect stacking in kissing structures was investigated by all-atom MDS with explicit representations of water and excess salt to compute the free energy cost for mutating the flanking base. Reference calculations of free energy cost for mutating the nucleotides in unstacked positions were also carried out, thus enabling the use of a thermodynamic cycle to discern differential contributions of stacking ($\Delta\Delta G_{stack}$) from the overall free energy cost of each mutation (ΔG_{mut}) (see Experimental Methods). As shown in Figure 7a, the horizontal processes ΔG_3 and ΔG_4 reflected the free energy cost for mutating the isolated base in solution and in the context of the kissing loop, respectively; the vertical processes ΔG_1 and ΔG_2 reflected the free energy of stacking of each base to the kissing complex. The change in stacking free energy upon mutation, $\Delta\Delta G_{stack} = \Delta G_1 - \Delta G_2$, could be then calculated from the mutation free energies, ΔG_4 and ΔG_3. Following this procedure, the $\Delta\Delta G_{stack}$ induced by changing each functional group was determined by sequentially mutating one of the two adenines to purine, 2-amino purine, and guanine. The sign and magnitude of the $\Delta\Delta G_{stack}$ values mirrored closely to changes in the abundance of heterodimers observed in ESI-MS experiments (Table 2). The value of $\Delta\Delta G_{G\to A}$ measured by pulling experiments in 1 M NaCl was 6 kcal/mol (Figure 6b). After adjusting the ionic condition to 250 mM NaCl,[34] we derived a $\Delta\Delta G_{G\to A}$ of 4 ± 2 kcal/mol, which compared favorably with a computed $\Delta\Delta G_{stack}$ of 1.5 ± 0.5 kcal/mol under the same ionic condition (Table 2).

The value of $\Delta\Delta G_{stack}$ associated with each mutation was further dissected to appreciate the contributions of van der Waals ($\Delta\Delta G_{VdW}$) and Coulomb's ($\Delta\Delta G_{Coul}$) interactions. Although molecular mechanics simulations cannot capture all of the intricate details of base-stacking such as $\pi-\pi$ interactions, it should be noted that the reported $\Delta\Delta G_{VdW}$ inherently includes both the hydrophobic and dispersion interactions known to be highly correlated with aqueous base-stacking propensities.[38] The results showed that loss of the 6-amino group (A→Pu) resulted primarily in an unfavorable loss of van der Waals interactions, suggesting greater exposure of the underlying GC pair. Conversely, when the 2-amino group was reintroduced to the purine ring (Pu→2Ap), only a small fraction of the original van der Waals stabilization enjoyed by adenine was recovered. This calculation suggested that substituents at the 2′-position may not be in direct contact with the underlying GC pair. The addition of the 6-carbonyl group (2Ap→G) was unfavorable to stacking primarily due to electrostatic interactions. This effect can be explained by comparing the partial charges assigned to guanine and adenosine within the AMBER fixed-charge framework.[39,40] In guanine, the carbonyl group possesses a negative partial charge of −0.544e, whereas in adenine the proton in the same position exhibits a charge of +0.132e. Therefore, the 6-carbonyl of the stacked guanine can create unfavorable charge repulsion between the carbonyl groups of the unpaired guanine and the underlying cytosine (Figure 7b). Together, the A→G mutation reduced both van der Waals and Coulomb's contributions to the stability of the interaction between the unpaired base and the adjacent kissing base pair (Table 2).

CONCLUSIONS

Our results have clearly shown that the stability of the two-base-pair complexes is strongly influenced by the efficient interstrand stacking of 5′-unpaired adenines. The general observation that stable duplexes require at least three Watson−Crick pairs stems from the ability of the internal pair to hold the two strands together during the transient fraying of the unprotected terminal pairs.[3−5] In one of very few exceptions, a self-complementary GpCpA ribonucleotide was shown to form a two-base-pair duplex in which the 3′-unpaired adenines stacked onto the adjacent base pair.[6] Other exceptions consisted of two-base-pair helices found in larger RNAs, such as

the one formed between the 521−522 and 527−528 regions of the 16S rRNA.[41−44] This type of interaction, however, did not enjoy the additional stabilization afforded by stacked terminal adenines, but benefited greatly from an overall structural context that constrained the position of the interacting regions and limited their conformational freedom. The simpler kissing complexes examined here mirrored more closely the two-base-pairs duplex established by the GpCpA model, with the conspicuous distinction that the unpaired adenines were now present at the 5′-end. In both systems, the flanking adenines shielded the adjacent GC pairs from the solvent, making them analogous to internal base pairs.

Our results demonstrated also that the formation of kissing complexes is dominated by base pairing, whereas dissociation is clearly influenced by the stacking interactions. The mechanical unfolding data are consistent with the hypothesis that dissociation of kissing complexes may occur only after the flanking adenines flip out of their stacked positions.[14] Therefore, flanking base stabilization is critically dependent on how long such bases may spend in stacked conformations, in analogy to what has been observed in secondary structure overhangs.[9] As demonstrated by the purine analogues, this characteristic is influenced significantly by the nature of the ring substituents, which affects the stacking interaction through both hydrophobic and electrostatic forces. This characteristic may also explain the apparent contradiction with previous reports on the inability of 5′-unpaired adenines to stabilize base pairing in secondary structure, especially in comparison to 3′-end overhangs.[10−12] Indeed, even an ApCpG analogue of the GpCpA model was found to be incapable of producing observable base pairing.[6] Further, unpaired bases at the 5′-end of an A-form helix are shown to be more dynamic than those at the 3′-end and much less inclined to remain stacked on the helical ends.[9] In kissing complexes, however, the conformation assumed by the hairpin loops may restrict the dynamics of 5′-unpaired adenines, thus enabling them to hold a stacking interaction for longer intervals.

In conclusion, the study of mutant hairpins has provided new insights into the structural determinants and mechanism of loop−loop kissing interactions. At the same time, it provided also a glimpse of the potential afforded by the combination of ESI-MS and single-molecule force spectroscopy in the investigation of weak interactions that elude traditional techniques. For these reasons, future work will aim not only at investigating other prototypical interactions that play critical roles in RNA folding, but also at reaching a greater understanding of the synergies and limitations of our concerted approach.

■ EXPERIMENTAL METHODS

Mass Spectrometry. RNA samples were purchased from Dharmacon (Lafayette, CO). The sequences of the various constructs are shown in Figure 2a. Unless otherwise noted, all RNA samples were dissolved in 150 mM ammonium acetate, pH 7.0. Formation of kissing dimers was facilitated by heating each construct to 95 °C for 5 min and quickly cooling it in an ice-slurry to obtain the kinetically favored hairpin structure.[21,32] Immediately prior to analysis, each sample was added with a 10% volume of 2-propanol to facilitate spray optimization. All analyses were carried out on a Thermo Fisher Scientific (West Palm Beach, CA) LTQ Orbitrap Velos mass spectrometer, which was operated in anionic nanospray mode by using quartz capillary emitters pulled in-house. Up to 2 μL samples were typically loaded onto each emitter by using a gel-loader pipet tip. A stainless steel wire was inserted in the back-end of the emitter to

supply an ionizing voltage of approximately 900−1200 V. The percentage of each RNA species detected in the sample was calculated by using the intensities of their respective signals normalized by their charge state to minimize any possible analytical bias, as discussed in refs 23,24 and references therein. For example, in the case of the GACG·GACG heterodimer formed by mixing H3 (light stem) and H4 (heavy stem) hairpins:

$$\%_{H3:H4} = \frac{\sum (I_{H3:H4}/z_{H3:H4})}{\sum (I_{H3:H4}/z_{H3:H4}) + \sum (I_{H3}/z_{H3}) + \sum (I_{H4}/z_{H4})} \times 100 \tag{1}$$

in which I_x is the peak intensity of each signal and z_x is the respective charge state. Only ^{12}C monoisotopic peaks were used in these calculations. When isotopic resolution is achieved, it has been demonstrated that signal intensities provide the same results as peak areas, as discussed previously.[23,24]

Single-Molecule Mechanical Unfolding. The sequence of the AA RNA is 5′-CCCGGGUGA**GACG**UCACCCGGG*AUUUU-UUUUUUUUUUUUUUUUUUUUUUUUUUACC*GGCGUGA**GACG**-UCACGCCGGAC-3′. Tetraloops were bold, whereas the 30-mer linker was in italic. **GACG** and **GUCG** tetraloops were present in AU; **GUCG** and **GACG** in UA; and **GUCG** and **GUCG** in UU. Similarly, **GGCG** and **GACG** were present in GA, and **GGCG** and **GGCG** in GG. DNA sequences corresponding to these RNAs were cloned into plasmids using a pCR 2.1-TOPO vector (Invitrogen). Preparations of the molecules for pulling experiments from the plasmids were described previously.[8,25,33] In the prepared samples, the RNA to be studied was flanked by two DNA/RNA "handles", each of 1.5 kb. The DNA handle upstream of the RNA was modified by a digoxigenin at the 5′-end, whereas the downstream DNA handle had a biotin at its 3′-end. Through noncovalent interactions, the entire molecule was tethered to two microspheres coated with antidigoxigenin antibody and streptavidin, respectively (Spherotech, Lake Forest, IL). The mechanical unfolding reaction was performed in a flow chamber, in which a pair of microspheres tethered by a single molecule were held by a force-measuring optical trap and a micropipet, respectively. The optical trap was steered relative to the stationary micropipet such that the single molecule was stretched or relaxed. Force and extension were acquired at a rate of 200 Hz. All the mechanical unfolding experiments were performed at 21 ± 1 °C. All solutions used in pulling experiments contained 20 mM HEPES at pH 7.5 and various concentrations of NaCl and MgCl₂. Details of the instrument, the "Minitweezers", and its operation, as well as sample preparation procedure, are available on http://tweezerslab.unipr.it.

Molecular Dynamics Simulations. All free energy calculations were performed using the GROMACS 4.5.5 molecular dynamics software,[45] the AMBER-99 force-field for RNA,[46] the TIP3P model for water,[47] and the Äqvist[48] model for KCl ions corrected to eliminate spurious ion aggregation.[49] Initial configurations of the two-base-pair kissing complex in the "bridging" form with parallel intermolecular hydrogen bonds were obtained from nonequilibrium dissociation simulations from previous work.[14] All free-energy calculations contained 21 633 atoms consisting of the 16-nucleotide kissing complex, 7023 waters, 30 K⁺ cations, and 16 Cl⁻ anions, reflecting effective excess salt concentration of ~250 mM in a 5.0 × 5.0 × 9.0 nm³ rectangular simulation cell. Reference free energy calculations of isolated nucleosides contained 2640 atoms consisting of a single purine nucleoside, 864 waters, and 8 KCl ion pairs in a 3.0 nm cubic box. All simulations maintained a constant pressure of 1 bar and a temperature of 298 K using the weak coupling algorithms by Berendsen and co-workers[50] with coupling constants of 1 and 0.2 ps, respectively. The equations of motion were integrated using a 2 fs time step and the leapfrog algorithm.[51] The two bond lengths and one bond angle in each water molecule were constrained to values prescribed by the TIP3P model using the SETTLE algorithm of Miyamoto and Kollman.[52] Snapshots were saved for analysis once every 2 ps. Periodic boundary conditions were employed to mimic the macroscopic setting for electrolytes. Long-range electrostatic interactions between periodic images were treated using the particle mesh

Ewald approach,[53] with a grid size of 0.12 nm, fourth-order cubic interpolation, and a tolerance of 10^{-5}. Neighbor lists were updated every 10 time steps. A cutoff of 10 Å was used for van der Waals interactions, real space Coulomb interactions, and for updating neighbor lists. Free energy calculations utilized established best practices for establishing converged equilibrium work values;[54,55] Hamiltonians for the mutated base were perturbed in discrete steps through 21 intermediate λ values for van der Waals interactions and 23 λ values for coulomb interactions, with each λ value entailing an independent equilibrium MDS of 10 to 20 ns (for the kissing-loops and isolated nucleosides respectively). Each of the 3 free energy cycles entailed a cumulative ~400 ns of the kissing simulations and ~900 ns of isolated nucleotide simulations. λ values were initially chosen in 10 uniform intervals between 0 and 1 with additional λ values iteratively added to minimize the error in $dH/d\lambda$. A soft-core scaling function with an α value of 0.5, σ value of 0.3, and exponent of 1 is utilized to avoid numerical singularities when λ approaches 0 or 1. Equilibrium work values were calculated using the multistep Bennett-acceptance ratio[56,57] with standard deviations defined by block averaging over 5 blocks. Error estimates for the derived $\Delta\Delta G_{stack}$ values follow standard error propagation rules.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional experimental data and figures. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail address: pli@albany.edu.

Author Contributions
William Stephenson and Papa Nii Asare-Okai contributed equally to this work.

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Butcher, S. E.; Pyle, A. M. *Acc. Chem. Res.* **2011**, *44*, 1302.
(2) Westhof, E.; Masquida, B.; Jossinet, F. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a003632.
(3) Borer, P. N.; Dengler, B.; Tinoco, I. J.; Uhlenbeck, O. C. *J. Mol. Biol.* **1974**, *86*, 843.
(4) Brahms, J.; Aubertin, A. M.; Dirheimer, G.; Grunberg-Manago, M. *Biochemistry* **1969**, *8*, 3269.
(5) Jaskunas, S. R.; Cantor, C. R.; Tinoco, I. J. *Biochemistry* **1968**, *7*, 3164.
(6) Alkema, D.; Bell, R. A.; Hader, P. A.; Neilson, T. *J. Am. Chem. Soc.* **1981**, *103*, 2866.
(7) Kim, C.; Tinoco, I. J. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9396.
(8) Li, P. T. X.; Bustamante, C.; Tinoco, I., Jr. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15847.
(9) Liu, J. D.; Zhao, L.; Xia, T. *Biochemistry* **2008**, *47*, 5962.
(10) Isaksson, J.; Chattopadhyaya. *Biochemistry* **2005**, *44*, 5390.
(11) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 16.
(12) Serra, M. J.; Turner, D. H. *Methods Enzymol.* **1995**, *259*, 242.
(13) Turner, D. H.; Mathews, D. H. *Nucleic Acids Res.* **2009**, *38*, D280.
(14) Chen, A. A.; Garcia, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1530.
(15) D'Souza, V.; Melamed, J.; Habib, D.; Pullen, K.; Wallace, K.; Summers, M. F. *J. Mol. Biol.* **2001**, *314*, 217.
(16) Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4671.
(17) Fabris, D. *Anal. Chem.* **2011**, *83* (15), 5810.
(18) Smith, S. B.; Cui, Y.; Bustamante, C. *Methods Enzymol.* **2003**, *361*, 134.
(19) Hofstadler, S. A.; Griffey, R. H. *Chem. Rev.* **2001**, *101*, 377.
(20) Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom. Ion Proc.* **2002**, *216*, 1.
(21) Turner, K. B.; Kohlway, A. S.; Hagan, N. A.; Fabris, D. *Biopolymers* **2009**, *91*, 283.
(22) Brinson, R. G.; Turner, K. B.; Yi-Brunozzi, H. Y.; Le Grice, S. F.; Fabris, D.; Marino, J. P. *Biochemistry* **2009**, *48* (29), 6988.
(23) Hagan, N. A.; Fabris, D. *Biochemistry* **2003**, *42*, 10736.
(24) Turner, K. B.; Hagan, N. A.; Fabris, D. *Nucleic Acids Res.* **2006**, *34*, 1305.
(25) Li, P. T.; Tinoco, I., Jr. *J. Mol. Biol.* **2009**, *386*, 1343.
(26) Tinoco, I.; Jr, Li, P. T. X.; Bustamante, C. *Q. Rev. Biophys.* **2006**, *39*, 325.
(27) Woodside, M. T.; Garcia-Garcia, C.; Block, S. M. *Curr. Opin. Chem. Biol.* **2008**, *12*, 640.
(28) Ganem, B.; Li, Y.-T.; Henion, J. D. *Tetrahedron Lett.* **1993**, *34*, 1445.
(29) Light-Wahl, K. J.; Springer, D. L.; Winger, B. E.; Edmonds, C. G.; Camp, D. G.; Thrall, B. D.; Smith, R. D. *J. Am. Chem. Soc.* **1993**, *115*, 803.
(30) Turner, K. B.; Brinson, R. G.; Yi-Brunozzi, Y. H.; Rausch, J. W.; Milletr, J. T.; Le Grice, S. F.; Marino, J. P.; Fabris, D. *Nucleic Acids Res.* **2008**, *36*, 2799.
(31) Hagan, N. A.; Fabris, D. *J. Mol. Biol.* **2007**, *365*, 396.
(32) Turner, K. B.; Hagan, N. A.; Fabris, D. *J. Mol. Biol.* **2007**, *369*, 812.
(33) Liphardt, J.; Onoa, B.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2001**, *292*, 733.
(34) Li, P. T. X. submitted.
(35) Li, P. T. X. *Soft Matter* **2013**, *9*, 3246−3254.
(36) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Block, S. M. *Science* **2006**, *314*, 1001.
(37) Li, P. T. X.; Collin, D.; Smith, S. B.; Bustamante, C.; Tinoco, I., Jr. *Biophys. J.* **2006**, *90*, 250.
(38) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213.
(39) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y. *Nucleic Acids Res.* **2011**, *39*, W511.
(40) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(41) Dunkle, J. A.; Cate, J. H. *Annu. Rev. Biophys.* **2010**, *39*, 227.
(42) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. *Science* **2005**, *310*, 827.
(43) Steitz, T. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242.
(44) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F. *Science* **2001**, *292*, 883.
(45) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
(46) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(48) Äqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021.
(49) Chen, A. A.; Pappu, R. V. *J. Phys. Chem.* **2007**, *111*, 11884.
(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(51) Hockney, R. W.; Goel, S. P.; Eastwood, J. *J. Comput. Chem.* **1974**, *14*, 148.

(52) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.

(53) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(54) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(55) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *J. Chem. Phys.* **2011**, *134*, 244107.

(56) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2008**, *129*, 124105.

(57) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.

5611

dx.doi.org/10.1021/ja310820h | *J. Am. Chem. Soc.* 2013, 135, 5602−5611