Conformational Determinants of Tandem GU Mismatches in RNA: Insights from Molecular Dynamics Simulations and Quantum Mechanical Calculations†

Yongping Pan, U. Deva Priyakumar, and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201

Received September 23, 2004; Revised Manuscript Received November 10, 2004

ABSTRACT: Structure and energetic properties of base pair mismatches in duplex RNA have been the focus of numerous investigations due to their role in many important biological functions. Such efforts have contributed to the development of models for secondary structure prediction of RNA, including the nearest-neighbor model. In RNA duplexes containing GU mismatches, 5′-GU-3′ tandem mismatches have a different thermodynamic stability than 5′-UG-3′ mismatches. In addition, 5′-GU-3′ mismatches in some sequence contexts do not follow the nearest-neighbor model for stability. To characterize the underlying atomic forces that determine the structural and thermodynamic properties of GU tandem mismatches, molecular dynamics (MD) simulations were performed on a series of 5′-GU-3′ and 5′-UG-3′ duplexes in different sequence contexts. Overall, the MD-derived structural models agree well with experimental data, including local deviations in base step helicoidal parameters in the region of the GU mismatches and the model where duplex stability is associated with the pattern of GU hydrogen bonding. Further analysis of the simulations, validated by data from quantum mechanical calculations, suggests that the experimentally observed differences in thermodynamic stability are dominated by GG interstrand followed by GU intrastrand base stacking interactions that dictate the one versus two hydrogen bonding scenarios for the GU pairs. In addition, the inability of 5′-GU-3′ mismatches in different sequence contexts to all fit into the nearest-neighbor model is indicated to be associated with interactions of the central four base pairs with the surrounding base pairs. The results emphasize the role of GG and GU stacking interactions on the structure and thermodynamics of GU mismatches in RNA.

GU mismatches are an integral part of RNA structures. These mismatches comprise about 50% of all mismatches in RNA (*1*−*3*) and up to 15% of all the base pairs in the RNA of some species. Accordingly, GU mismatch motifs are indispensable to the proper biological function of many RNAs (*4*−*9*). Thus, an understanding of the structural and thermodynamic properties of mismatched GU base pairs in RNA is as important as such an understanding of Watson−Crick (WC)[1] base pairing interactions.

A large body of experimental work has been performed on mismatch-containing RNA duplexes using techniques such as NMR and X-ray structural determination and measurements of thermodynamic stability (*10*−*16*), among others. These experimental studies have shown that the A-form conformation typical for RNA (*17*) is largely intact upon insertion of a few mismatched base pairs with conformational distortions localized to the mismatch and the nucleotides directly adjacent to the mismatch (*13*, *14*). Accordingly, the thermodynamics and structures of mismatches are believed to be determined largely by the interactions within the mismatched pairs and between the mismatched pairs and its nearest neighbors. Thus, the secondary structures of the mismatches can be easily predicted using the nearest-neighbor model (*11*), which has been widely used for both DNA and RNA. Central to the nearest-neighbor model is the assumption that the secondary structure is not significantly perturbed. However, this model, although relatively successful, has been shown to not be valid for selected single mismatches in different contexts (*18*, *19*). Other techniques for secondary structure prediction, such as free energy minimization (*20*−*22*), have also only achieved limited success for mismatches (*22*, *23*).

Thermodynamic parameters of GU mismatches have revealed some interesting phenomena. 5′-UG-3′ tandem mismatches make a more favorable contribution to the stability of RNA duplexes than the 5′-GU-3′ mismatches in the same sequence context (*10*). A GU mismatch is more stabilizing in GGUC than in CGUG, AGUU or UGUA in RNA octamers. Furthermore, thermodynamic data for all duplexes containing the 5′-GU-3′ motif do not fit the nearest-neighbor model (*24*). In combination, these results suggest that tandem GU mismatches have unique structural properties leading to the observed thermodynamic stabilities.

The structural basis for the different thermodynamic behavior of tandem GU mismatches in different sequence contexts was proposed by Turner and co-workers on the basis of NMR-based structures and thermodynamic data (*10*). They

† This work was supported by the NIH (Grant GM-51501), DOD ACS Major Shared Resource Computing, and the PSC Pittsburgh Supercomputing Center.

* Corresponding author. Phone: (410) 706-7442. Fax: (410) 706-5017. E-mail: alex@outerbanks.umaryland.edu.

[1] Abbreviations: MD, molecular dynamics; QM, quantum mechanics; RMS, root-mean-square; WC, Watson−Crick.

10.1021/bi047932q CCC: $30.25
Published on Web 01/08/2005

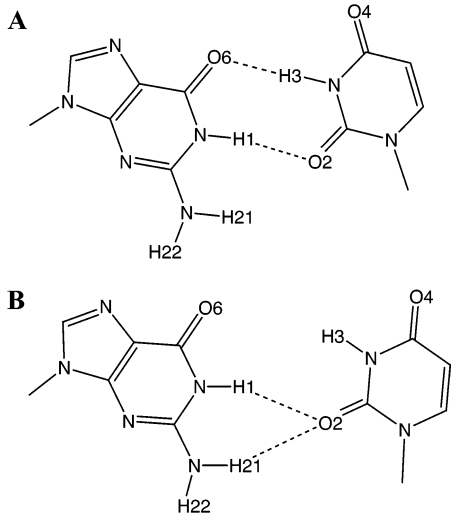

FIGURE 1: Schematic diagram of potential hydrogen bonds between the GU mismatched pairs: (A) two hydrogen bond model with (G)-O6−(U)H3 and (G)H1−(U)O2 hydrogen bonds; (B) one hydrogen bond model with a (G)H1−(U)O2 hydrogen bond and a possible contribution from a (G)H21−(U)O2 interaction leading to a three-centered hydrogen bond.

proposed two models with one or two hydrogen bonds between each GU mismatch (Figure 1). It was suggested that a one hydrogen bond model is associated with a small free energy change, while a two hydrogen bond conformation corresponds to a larger free energy change. These experiments have greatly enhanced our understanding of the possible conformations of the mismatches and their relationship to the stability of RNA. However, the underlying forces that determine the conformational preferences are still largely unknown; interstrand hydrogen bonding, complementary electrostatics of base steps, and stacking interactions have been suggested to make important contributions to the stability of the duplexes (*14*), and solvation contributions, given the essential role of hydration in nucleic acid structure (*25*, *26*), must also be considered. In addition, structural models obtained from NMR data can be equivocal in certain cases (*10*, *13*) where local conformational features cannot be unambiguously determined, hindering the further elucidation of structure−stability relationships.

Empirical force field based molecular modeling, including molecular dynamics (MD) simulations, is a valuable alternative to resolve such issues (*27*). To gain a clearer picture of the conformational preferences of the tandem GU mismatches in different sequence contexts and relate them to their thermodynamic stability, MD simulations were performed on a series of RNA duplexes with tandem GU mismatches that have been subjected to experimental studies (*10*). Resulting structural models from the simulations are in good agreement with published NMR-derived structural data. Detailed analysis of the simulations indicate that interstrand GG and intrastrand GU stacking interactions dominate the orientations of the mismatched base pairs and suggest a model of how these interactions dictate the observed experimental stabilities.

METHODS

MD simulations were performed using the program CHARMM (*28*) with the CHARMM27 nucleic acid force field (*29*−*31*) and the modified TIP3P water model (*32*). All starting structures were the canonical A-form (*33*) and were prepared with QUANTA (*34*). Starting structures were overlaid with a preequilibrated solvent box consisting of water and sodium ions or NaCl, as required (see below). The solvent box was extended 8 Å beyond the nucleic acid non-hydrogen atoms. Solvent molecules with non-hydrogen atoms that were within 1.8 Å of nucleic acid non-hydrogen atoms were deleted. The number of sodium ions in the solvated systems was adjusted to yield NaCl concentrations of approximately 0.3 and 1 M by deleting sodium ions that were furthest from the oligonucleotide or adding sodium and chloride ions at random positions. The number of chloride ions was adjusted to neutralize the systems for the high salt concentration (1 M). The salt concentrations of 0.3 and 1 M were selected to correspond approximately to the concentrations used in the NMR and thermodynamic stability experiments, respectively (*10*, *24*). Periodic boundary conditions were simulated using the CRYSTAL module (*35*) in CHARMM. Each system was minimized for 500 adopted basis Newton−Raphson (ABNR) steps with mass-weighted harmonic restraints of 2.0 kcal/(mol·Å) on the non-hydrogen atoms of the oligonucleotide. The minimized systems were then subjected to a 20 ps MD simulation in the constant volume, isothermal (*NVT*) ensemble in the presence of the harmonic restraints on the RNA. The integration time step was 2 fs, and SHAKE was used to constrain all covalent bonds involving hydrogens (*36*). In all calculations, electrostatics were treated via the particle mesh Ewald method (*37*, *38*), Lennard-Jones interactions were truncated at 12 Å with a force switch smoothing function (*39*) from 10 to 12 Å, and the nonbond atom lists were updated heuristically. Production simulations were performed for 5 ns in the *NPT* ensemble at 300 K with the Leapfrog integrator. The *NPT* ensemble was achieved using Hoover chains (*40*) for temperature control with a thermal piston mass of 1000 kcal/(mol·ps^2) at 300 K with pressure control to 1 atm via the Langevin piston method (*41*) with a piston mass of 600 amu and the piston collision frequency set to 0. No restraints were used in the production simulations. The final 4 ns from each trajectory were used for analysis on time frames saved every 2 ps. Analysis was performed by calculating the respective properties from the individual time frames and then averaging over the values from all the time frames in the production portion of the trajectory.

QM calculations were performed with the Gaussian 98 and 03 packages (*42*) with the 6-31G*, 6-311+G**, and the aug-cc-pVDZ (*43*) basis sets at the MP2 level of theory (*44*). LMP2 calculations (*45*) were performed using version 4.1 of the program Jaguar (Schrödinger Inc.). In addition to the above basis sets, the 6-31G basis set with diffuse polarization functions on all non-hydrogen atoms with an exponent of 0.25 was also used (6-31G*(0.25)). The 6-31G*(0.25) basis set has been shown to provide a better description of dispersion energy in stacking interactions between nucleic acid bases (*46*, *47*) over standard basis sets. Unless noted, optimizations were performed to the default tolerances at the MP2/6-31G* level. Base pairing calculations were performed with the base planarity enforced. For stacking interactions, geometries for the respective base non-hydrogen atoms were obtained from the MD simulations with the stacking interaction energies including both electrostatic and

Table 1: RNA Duplexes Included in This Study[a]

ID	duplex	[NaCl] (M)	T_{GU}[b]	T_{WC}[c]	T_m	NMR[d]	no. of H-bonds[e]
Ia	GGC**UG**GCC	0.3, 1	66	68	65		2
Ib	GGC**GU**GCC	0.3, 1	35	50	56	√	1
IIa	GAG**UG**CUC	0.3, 1	50	50	50	√	2
IIb	GAG**GU**CUC	0.3, 1	45	43	47	√	2
IIIa	GGA**UG**UCC	0.3, 1	50	50	52	√	2
IIIb	GGA**GU**UCC	0.3, 1	15−30	30	39	√	1
IVa	GGC**UA**GCC	0.3, 1					
IVb	GGC**AU**GCC	0.3, 1			69		
Va	GAG**UA**CUC	0.3, 1			56		
Vb	GAG**AU**CUC	0.3, 1					

[a] Experimental data as reported in Table 3 of ref *10*. Melting experiments were performed in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM sodium EDTA, pH 7.0. [b] Temperature at which the GU imino proton resonances are no longer observed. [c] Temperature at which the WC imino proton resonances are no longer observed. [d] Duplexes with available 3D experimental NMR structures are marked with a √. NMR experiments were performed in 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM sodium EDTA, pH 7.0 (*10*). [e] Number of GU intermolecular hydrogen bonds suggested from NMR models or NMR spectra.

van der Waals (VDW) contributions. Hydrogen atoms were added to the N1 or N9 atom of the pyrmidines and purines, respectively, and the structures were optimized at the MP2/6-31G* level with the nuclear coordinates of all non-hydrogen atoms frozen; similar constraints were used for all of the monomers to calculate the interaction energies. Single-point energy calculations at the MP2 level were done with the 6-31G*(0.25), 6-311+G**, and aug-cc-pVDZ basis sets. The interaction energies were corrected for basis set superposition error (BSSE) using the counterpoise correction methodology (*48*).

RESULTS

Three pairs of duplexes were studied (Table 1, duplexes **Ia** and **Ib**, **IIa** and **IIb**, and **IIIa** and **IIIb**) with the only difference being the order of the GU mismatches in each a,b pair. In addition, control sequences (Table 1, duplexes **IVa, IVb, Va**, and **Vb**) corresponding to **I** and **II** were studied. Several of the duplexes with mismatches have been subjected to NMR and thermodynamic experiment, and therefore, direct comparison between simulations and experiment can be made to validate the presented MD simulations. The results from MD simulations at salt concentrations of 0.3 and 1 M were similar based on structural analysis. For example, the backbone dihedral angles for the central four nucleotides were almost identical (see Tables S1 and S2 of the Supporting Information). Accordingly, only the simulations at 1 M NaCl are presented in detail for brevity, unless otherwise noted.

Overall Structure. Average root-mean-square (RMS) differences for the central six base pairs over the final 4 ns of the MD trajectories with respect to the canonical A-form or experimental structures are presented in Table 2. As is evident, the structures stay close to the experimental and canonical A-form structures. Additional structural analysis was performed by obtaining the average dihedral angles for the central four nucleotides (Tables S1 and S2). In all cases, the average values are similar to those from canonical A-form RNA (Tables S1 and S2, bottom). Combined, the average RMS differences and dihedral angles show that the simulated structures were similar to the available NMR or canonical A-form (*49*) structures. These results are consistent with the notion that the A-form structures of duplex RNA are mostly intact despite insertion of the mismatched base pairs (*13*, *14*, *50*).

Table 2: RMS Differences (Å) of the Central Six Base Pairs over the Last 4 ns of the 1 M NaCl Trajectories[a]

	RMSD	reference state
Ia	2.1 ± 0.6	A-form
Ib	2.7 ± 0.5	NMR
IIa	1.4 ± 0.5	NMR
IIb	2.3 ± 0.6	NMR
IIIa	2.1 ± 0.5	NMR
IIIb	1.8 ± 0.4	NMR
IVa	1.4 ± 0.4	A-form
IVb	1.4 ± 0.3	A-form
Va	2.5 ± 0.7	A-form
Vb	1.4 ± 0.4	A-form

[a] Errors represent the RMS fluctuations.

More detailed analysis was performed by calculating the average values of slide, rise, and twist for the central five base steps in the simulations. The resulting data is in Table 3 for the structures with GU mismatches and in Table 4 for the control sequences. In general, the values are similar to those in A-form structures, consistent with the RMSD and dihedral data discussed above. However, with the mismatches, there are local deviations from the standard A-form values. This is clearly seen from comparison of the mismatches (Table 3) with the respective control sequences (Table 4). For example, values of twist for the central three base steps show alternating values versus similar values for the central three base steps in the control duplexes. Specifically, the central base step had a small twist and the two neighboring steps showed large twist values when the mismatch order is 5′-UG-3′ (**Ia**, **IIa**, and **IIIa**), while the reverse pattern was observed for the central three base steps for the 5′-GU-3′ mismatches (**Ib**, **IIb**, and **IIIb**). Similar patterns are present for slide; that is, when the twist is small, the slide is large for the central three base steps, or vice versa. Other helical parameters changed accordingly, although the pattern is less obvious (not shown). Such variations for both types of 5′-GU-3′ and 5′-UG-3′ mismatches are in agreement with values from NMR-derived structures (Tables 3 and 4) (*13*, *14*, *50*−*52*), although some differences between the MD and NMR values are present. Such differences may be due potential limitations in the NMR restraint data as well as force field contributions associated with use of the AMBER force field (*53*) for the NMR structure determination versus use of CHARMM in the present, unrestrained MD simulations. However, in combination, data from the present MD simulations and the NMR structures point toward a unique sequence-specific structural pattern present in the tandem GU mismatches.

Hydrogen Bonding between GU Mismatches. GU base pair mismatches typically form either one or two hydrogen bonds. In the two hydrogen bond scenario (Figure 1A), both the (G)O6−(U)H3 and (G)H1−(U)O2 interactions occur. The one hydrogen bond scenario (Figure 1B) involves a single (G)H1−(U)O2 hydrogen bond, possibly being supplemented with a (G)H21−(U)O2 interaction, leading to a three-centered hydrogen bond. In both cases, longer hydrogen bond-like interactions can occur (e.g., (G)H21−(U)O2 for

Table 3: Base Step Parameters in Duplexes with Mismatches at 1 M NaCl[a]

	slide (Å)	rise (Å)	twist (deg)
	Ia		
G2C3	−1.9 ± 0.6	3.4 ± 0.3	31.3 ± 5.8
C3U4	−1.9 ± 0.5	3.5 ± 0.3	38.0 ± 5.4
U4G5	−2.8 ± 0.9	3.7 ± 0.9	6.4 ± 8.7
G5G6	−1.8 ± 0.5	3.7 ± 0.7	39.1 ± 5.5
G6C7	−1.6 ± 0.7	3.8 ± 0.8	52.5 ± 16.7
avg	−2.0	3.6	33.5
	Ib		
G2C3	−0.4 ± 0.7 (−2.1)	3.5 ± 0.4 (3.1)	47.4 ± 6.7 (31.6)
C3G4	−5.0 ± 0.7 (−2.2)	3.6 ± 0.9 (3.3)	11.0 ± 6.2 (20.6)
G4U5	−0.4 ± 0.6 (−1.2)	3.1 ± 0.3 (3.3)	45.8 ± 5.5 (45.3)
U5G6	−3.6 ± 1.3 (−2.2)	3.9 ± 0.7 (3.4)	3.3 ± 14.7 (20.4)
G6C7	−1.3 ± 1.0 (−2.0)	3.2 ± 0.3 (3.1)	34.5 ± 9.7 (31.6)
avg	−2.1 (−1.9)	3.5 (3.2)	24 (29.9)
	IIa		
A2G3	−2.0 ± 0.4 (−2.1)	3.5 ± 0.4 (3.2)	28.6 ± 5.6 (29.7)
G3U4	−1.5 ± 0.7 (−2.1)	3.4 ± 0.3 (3.2)	39.5 ± 4.7 (37.3)
U4G5	−2.6 ± 0.8 (−2.4)	3.3 ± 0.6 (3.6)	9.4 ± 7.5 (16.4)
G5C6	−1.7 ± 0.7 (−2.1)	3.5 ± 0.3 (3.2)	40.0 ± 4.0 (37.5)
C6U7	−2.0 ± 0.5 (−2.0)	3.6 ± 0.5 (3.2)	29.3 ± 5.3 (29.7)
avg	−2.0 (−2.1)	3.5 (3.3)	29.4 (30.1)
	IIb		
A2G3	−2.1 ± 0.5 (−2.6)	3.7 ± 0.5 (3.2)	49.1 ± 10.3 (28.9)
G3G4	−2.7 ± 0.5 (−2.3)	3.0 ± 0.5 (3.2)	12.0 ± 7.1 (25.8)
G4U5	−0.3 ± 0.6 (−2.1)	3.6 ± 0.4 (3.1)	54.1 ± 7.2 (47.9)
U5C6	−2.9 ± 0.6 (−2.3)	3.0 ± 0.5 (3.2)	12.2 ± 7.0 (26.6)
C6U7	−6.4 ± 2.5 (−2.6)	4.7 ± 3.9 (3.2)	23.3 ± 73.6 (27.1)
avg	−2.9 (−2.4)	3.6 (3.2)	30.1 (31.3)
	IIIa		
G2A3	−2.6 ± 0.9 (−2.7)	4.1 ± 1.0 (3.5)	10.1 ± 19.2 (28.2)
A3U4	−0.2 ± 1.1 (−1.9)	3.4 ± 0.5 (3.2)	54.5 ± 14.6 (40.7)
U4G5	−4.0 ± 1.1 (−2.7)	3.5 ± 0.8 (2.3)	−2.6 ± 11.4 (14.4)
G5U6	−1.3 ± 0.6 (−1.9)	3.3 ± 0.3 (3.3)	37.9 ± 4.6 (40.7)
U6C7	−2.0 ± 0.5 (−2.7)	3.6 ± 0.4 (3.5)	27.8 ± 7.0 (29.8)
avg	−2.0 (−2.2)	3.6 (3.2)	25.5 (30.8)
	IIIb		
G2A3	−1.6 ± 0.7 (−2.3)	3.6 ± 0.5 (2.9)	34.8 ± 9.0 (29.4)
A3G4	−2.3 ± 0.4 (−2.4)	2.5 ± 0.5 (3.0)	18.3 ± 5.9 (25.7)
G4U5	−0.8 ± 0.6 (−1.7)	4.1 ± 0.5 (3.2)	54.5 ± 7.0 (49.5)
U5U6	−1.7 ± 0.5 (−2.3)	1.8 ± 0.5 (2.9)	−7.6 ± 5.4 (24.1)
U6C7	−1.9 ± 0.4 (−2.2)	3.7 ± 0.5 (3.0)	58.0 ± 5.4 (29.8)
avg	−1.7 (−2.2)	3.1 (3.0)	31.6 (31.7)

[a] Data in parentheses were calculated from the NMR structures. Errors represent the RMS fluctuations.

Table 4: Base Step Parameters in Control Duplexes at 1 M NaCl[a]

	slide (Å)	rise (Å)	twist (deg)
	IVa		
G2C3	−1.5 ± 0.8	3.4 ± 0.3	36.1 ± 7.7
C3U4	−2.2 ± 0.6	3.3 ± 0.5	24.0 ± 8.8
U4A5	−1.8 ± 0.4	3.6 ± 0.4	29.0 ± 3.8
A5G6	−1.9 ± 0.5	3.5 ± 0.4	28.9 ± 4.6
G6C7	−1.7 ± 0.6	3.4 ± 0.3	32.8 ± 5.1
avg	−1.8	3.4	30.2
A-form	−2.1	3.5	31.5
	IVb		
G2C3	−1.8 ± 0.5	3.3 ± 0.3	29.9 ± 3.9
C3A4	−1.7 ± 0.4	3.9 ± 0.5	34.9 ± 6.8
A4U5	−1.8 ± 0.4	3.2 ± 0.5	26.5 ± 5.1
U5G6	−1.9 ± 0.4	3.7 ± 0.5	30.4 ± 4.5
G6C7	−1.6 ± 0.5	3.3 ± 0.3	31.8 ± 5.5
avg	−1.8	3.5	30.7
A-form	−2.1	3.2	31.0
	Va		
A2G3	−2.1 ± 0.5	3.3 ± 0.5	26.0 ± 6.1
G3U4	−1.8 ± 0.9	3.4 ± 0.4	31.1 ± 7.8
U4A5	−2.7 ± 1.1	4.3 ± 0.7	23.2 ± 15.9
A5C6	−0.8 ± 0.8	3.2 ± 0.5	45.9 ± 11.1
C6U7	−3.2 ± 0.8	3.0 ± 0.6	3.4 ± 12.6
avg	−2.1	3.4	25.9
A-form	−2.1	3.2	30.9
	Vb		
A2G3	−2.0 ± 0.4	3.5 ± 0.4	29.8 ± 4.2
G3A4	−2.0 ± 0.5	3.7 ± 0.4	31.5 ± 4.1
A4U5	−2.0 ± 0.5	3.4 ± 0.4	29.5 ± 4.8
U5C6	−2.0 ± 0.6	3.5 ± 0.5	28.2 ± 5.3
C6U7	−1.8 ± 0.5	3.3 ± 0.4	27.9 ± 5.3
avg	−2.0	3.5	29.4
A-form	−2.1	3.5	31.5

[a] A-form data were calculated using the corresponding canonical structures (*49*) generated with the QUANTA program (*34*). Errors represent the RMS fluctuations.

the two hydrogen bond scenario) with the interactions defining the one versus two hydrogen bond scenarios being the shortest interactions. To evaluate the hydrogen bond scenarios in the different duplexes, probability distributions of the distances for the three potential hydrogen bonds were calculated (Figure 2). The two hydrogen bond scenario occurs in **Ia** at both salt concentrations, in **IIa** at 1 M salt, and in **IIb** and **IIIa** in 1 M salt, although the contributions for these sequences are minor. For **Ib**, **IIb**, and **IIIb**, the one hydrogen bond scenario dominated, with only a small amount of the two hydrogen bond scenario occurring in **IIb**. Thus, the one hydrogen bond scenario dominates in duplexes **Ib**, **IIb**, and **IIIb**, which all contain tandem 5′-GU-3′ mismatches, while in the duplexes containing 5′-UG-3′ tandem mismatches, **Ia**, **IIa**, and **IIIa**, there is a significant sampling of the two hydrogen bond scenario, especially in 1 M salt.

The calculated hydrogen bonding patterns may be compared with available experimental data. NMR-determined structures for **Ib** and **IIa** show one and two hydrogen bond scenarios, respectively (*10*), consistent with the data in Figure 2. Chen et al. (*10*) also noted that the difference in temperature for imino proton signal disappearance in GU mismatches (T_{GU}) versus WC base pairs (T_{WC}) was consistent with a one versus two hydrogen bond scenario, where a significantly smaller T_{GU} value is indicative of the one hydrogen bond scenario. The T_{GU} and T_{WC} values for the studied duplexes along with the proposed number of hydrogen bonds are included in Table 1. Based on this model, **IIa** and **IIb** would be expected to have similar hydrogen bonding, while differential hydrogen bonding would be expected in **IIIa** versus **IIIb**. Examination of Figure 2 shows some sampling of the two hydrogen bond scenario in both **IIa** and **IIb**, while in **IIIa** there is some sampling of the two hydrogen bond scenario, while none is observed with **IIIb**. Thus, the calculated GU hydrogen bonding patterns are consistent with experimental data. Furthermore, the present results indicate that rather than well-defined hydrogen bond patterns, in some of the duplexes equilibria between the different hydrogen bonding scenarios exist, as previously mentioned (*10*).

An interesting pattern emerges when the hydrogen bond scenario is compared to the temperatures (T_{GU}) at which the proton signals of GU mismatches disappear (Table 1). For all the duplexes with two hydrogen bonds between the GU mismatches, **Ia**, **IIa**, and **IIIa**, both their GU and WC proton signals disappeared at higher temperatures than their 5′-GU-3′ mismatch counterparts, **Ib**, **IIb**, and **IIIb**, which are dominated by the one hydrogen bond scenario. This relation-

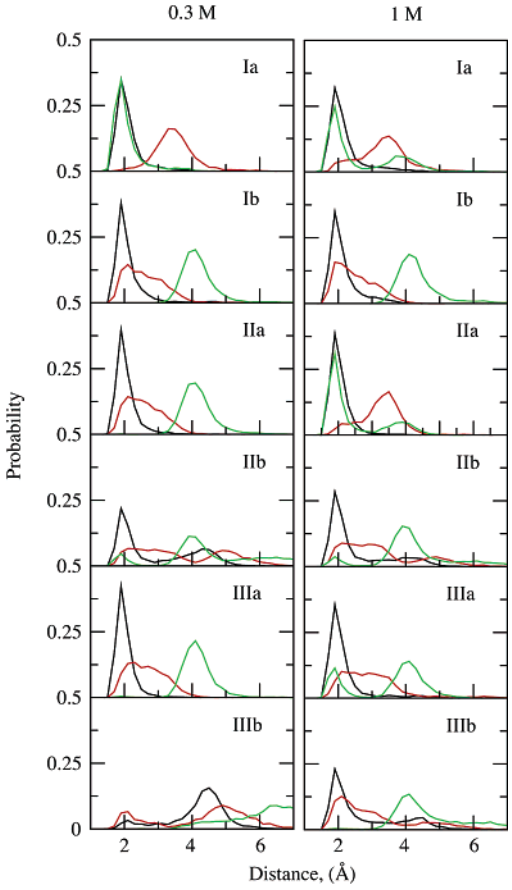

FIGURE 2: Distributions of the potential hydrogen bond distances (G)O6−(U)H3 (green), (G)H1−(U)O2 (black), and (G)H21−(U)O2 (red) for the GU mismatches. Note that the two hydrogen bond scenario corresponds to the green and black lines having maxima in the vicinity of 2 Å and the one hydrogen bond scenario has red and black lines with maxima in the vicinity of 2 Å.

Table 5: Average Base Step Intrastrand Interaction Energies for the Central Five Steps in the Mismatch Containing Duplexes at 1 M NaCl[a]

	B2B3	B3B4	B4B5	B5B6	B6B7
Ia	G2C3	C3U4	U4G5	G5G6	G6C7
	−10.6 ± 5.7	−3.4 ± 1.3	−1.9 ± 1.7	−3.0 ± 1.7	−9.4 ± 6.5
Ib	G2C3	C3G4	G4U5	U5G6	G6C7
	−9.9 ± 5.0	−0.8 ± 4.6	−7.9 ± 1.5	−0.5 ± 0.7	−11.8 ± 3.6
IIa	A2G3	G3U4	U4G5	G5C6	C6U7
	−7.5 ± 1.8	−7.3 ± 1.6	−2.2 ± 1.7	−13.0 ± 2.5	−4.3 ± 1.1
IIb	A2G3	G3G4	G4U5	U5C6	C6U7
	−1.8 ± 1.9	−2.8 ± 1.5	−7.5 ± 1.6	−1.8 ± 0.8	−3.7 ± 1.4
IIIa	G2A3	A3U4	U4G5	G5U6	U6C7
	−7.4 ± 3.6	−3.6 ± 1.1	−1.1 ± 1.7	−6.3 ± 1.6	−2.7 ± 1.0
IIIb	G2A3	A3G4	G4U5	U5U6	U6C7
	−8.3 ± 2.3	−4.5 ± 3.2	−4.3 ± 2.5	−3.1 ± 0.9	−0.6 ± 1.8

[a] Energies in kcal/mol, averages over the final 4 ns of each simulation and over the symmetry-related pairs in the two strands, and errors represent the RMS fluctuations.

Table 6: Average Base Step Interstrand Interaction Energies for the Central Five Steps in the Duplexes with Mismatches at 1 M NaCl[a]

	B3B7	B4B6	B5B5	B6B4	B7B3
Ia	C3C7	U4G6	G5G5	G6U4	C7C3
	4.7 ± 1.8	−2.7 ± 1.6	−6.7 ± 1.9	−1.7 ± 2.0	0.3 ± 2.3
Ib	C3C7	G4G6	U5U5	G6G4	C7C3
	2.3 ± 1.5	−7.6 ± 2.6	0.3 ± 0.5	−7.9 ± 2.1	2.6 ± 3.0
IIa	G3U7	U4C6	G5G5	C6U4	U7G3
	−1.9 ± 1.0	0.8 ± 0.8	−7.1 ± 1.6	0.6 ± 1.0	−1.9 ± 1.1
IIb	G3U7	G4C6	U5U5	C6G4	U7G3
	−2.7 ± 1.6	−4.8 ± 2.1	0.2 ± 0.2	−4.5 ± 2.1	−2.0 ± 1.5
IIIa	A3C7	U4U6	G5G5	U6U4	C7A3
	−2.8 ± 2.2	−0.2 ± 1.0	−7.3 ± 1.7	−0.6 ± 0.9	−1.2 ± 1.3
IIIb	A3C7	G4U6	U5U5	U6G4	C7A3
	−0.9 ± 1.1	−2.1 ± 0.9	−0.1 ± 0.2	−3.0 ± 1.8	−1.6 ± 1.1

[a] Energies in kcal/mol, averages over the final 4 ns of each simulation, and errors represent the RMS fluctuations.

ship extends to the melting temperatures, T_m (Table 1), where the duplexes with 5′-GU-3′ tandem mismatches consistently have lower T_m values.

Closer analysis of the T_m values show the smallest differential between the 5′-UG-3′ and its partner 5′-GU-3′mismatch to occur with duplex **II**. Analysis of Figure 2 shows **IIb** to be the only 5′-GU-3′ tandem mismatch containing species to sample the two hydrogen bond scenario, albeit at a low population. Thus, the stability of the GU tandem mismatches correlates well with the presence of the two hydrogen bond scenario versus the one hydrogen bond scenario with the two hydrogen bond scenario associated with a higher T_m.

The presence of the two hydrogen bond scenario leading to the more stable duplexes is not surprising and has previously been proposed (*10*). However, the question remains as to why this hydrogen-bonding pattern does not occur in all the duplexes that contain GU tandem mismatches. This is especially relevant considering that the two hydrogen bond scenario is approximately 2−6 kcal/mol more favorable than the one-hydrogen bond scenario based on QM calculations at various levels of theory (see below). Clearly, there are other interactions controlling the GU hydrogen bonding pattern. Accordingly, additional analysis was undertaken to identify these interactions.

Local Stacking Interactions. Stacking in oligonucleotides has long been known to have a significant impact on structure and stability (*17*, *54*−*56*). Accordingly, systematic analysis of the simulations was performed to identify possible inter- or intrastrand stacking interactions that could impact the observed structural and thermodynamic data on the tandem GU mismatches. Due to difficulties in geometrically defining stacking interactions, energetic analysis, via base−base interaction energies, was used to quantify these interactions. Accordingly, average stacking interactions for all possible intra- and interstrand base−base interactions involving the central five base steps were calculated for all the duplexes studied. Analysis of these results in Tables S3 and S4 of the Supporting Information show favorable interactions to occur for all five central base pairs for the intrastrand stacking interactions, while with the interstrand stacking interactions, only those involving the B3B7 through B7B3 pairs are generally favorable. The results for these particular interactions for the studied mismatches are presented in Tables 5 and 6 for the intra- and interstrand base steps, respectively.

For the intrastrand base step stacking, the majority of interactions are favorable with the most interesting trends occurring with the central three bases. The B4B5 (i.e., base 4 to base 5, Figure 3) stacking energies, which are all GU pairs, are more favorable in the 5′-GU-3′ mismatches, **Ib, IIb**, and **IIIb**, while the B5B6 interactions are more favorable for the 5′-UG-3′ mismatches, **Ia**, **IIa**, and **IIIa**. With the interstrand stacking interactions, there are even more systematic differences (Table 6). All the 5′-UG-3′ mismatches

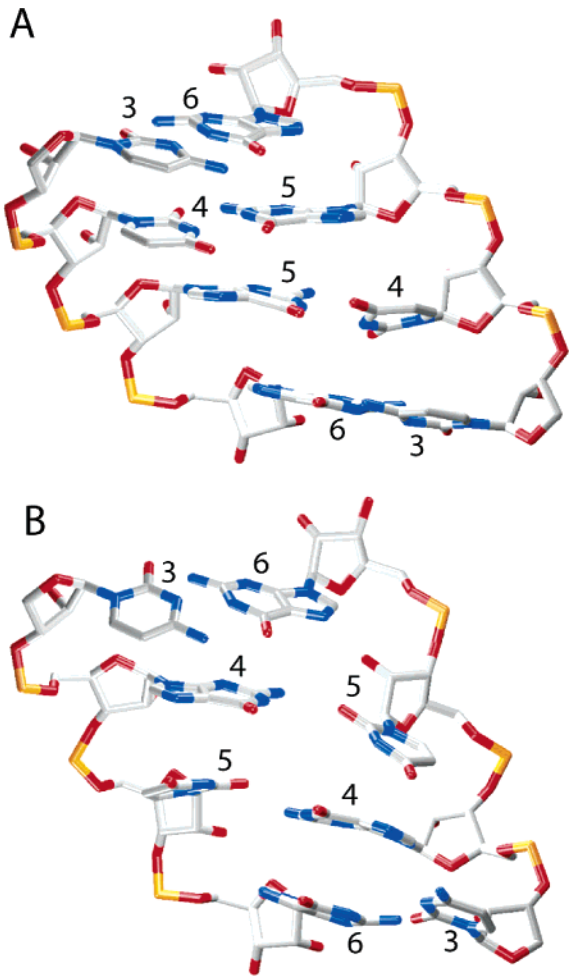

FIGURE 3: Snapshots of the central four base pairs at 3 ns from the duplex **Ia** (A) and **Ib** (B) trajectories showing the difference in interstrand stacking interactions. In duplex **Ia**, two G5 bases from each strand stack together, favoring a two hydrogen bond conformation for the GU mismatches, while in duplex **Ib**, G4 and G6 from each strand interact favorably, leading to a one hydrogen bond structure.

have very favorable B5B5 interactions, which are all GG stacking interactions, while with the 5′-GU-3′ mismatches, the B4B6 and B6B4 interstrand interactions are quite favorable. These latter interactions are GG, GC/CG, or GU/UG stacking with the most favorable such interactions being the GG interaction in **Ib**.

The stacking interactions impact the geometries of the duplexes. Analysis of the helicoidal parameters (Table 3) shows that in the three duplexes with 5′-UG-3′ mismatches (**Ia**, **IIa**, and **IIIa**), the favorable G5G5 interstrand interactions correlate with a large slide and small twist for the central B4 to B5 base steps. Similarly, in **Ib**, **IIb**, and **IIIb**, with the 5′-GU-3′ mismatches, the favorable B4B5 intrastrand and B4B6/B6B4 interstrand interaction energies are correlated with the large slide and small twist for the B3B4 and B5B6 base steps. Furthermore, the slide difference between the base steps was much larger for **Ib** than for **IIb** and **IIIb**; such differences are consistent with the NMR data, in which the slide difference for **Ib** is 1.0 Å while they are 0.2 and 0.7 Å for **IIb** and **IIIb**, respectively (Table 3). The larger slide difference in **Ib** is consistent with the highly favorable B4B6/B6B4 interstrand interaction energies associated with the GG interactions. These results indicate a correlation between stacking interactions and large slide and small twist values for a given base steps with GG interstrand and GU intrastrand stacking interactions playing a dominant role.

The structural impact of the GU and GG stacking interactions may be seen in Figure 3. For **Ia**, a 5′-UG-3′ mismatch, the interstrand stacking overlap of the 5 position bases is evident (Figure 3A). This leads to the large slide and small twist values (Table 3). In contrast, with **Ib**, a 5′-GU-3′ mismatch, intrastrand B4B5 and interstrand B4B6 and B6B4 base stacking can be seen (Figure 3B). These interactions lead to the large slide and small twist values for the B3B4 and B5B6 base steps in **Ib**, **IIb**, and **IIIb** (Table 3). Moreover, the visual inspection suggests that the these stacking interactions tend to pull the GU bases from the central portion of the duplex, leading to the one hydrogen bonding scenario in the 5′-GU-3′ mismatches. Alternatively, with the 5′-UG-3′ mismatches, the strong G5G5 interstrand stacking is suggested to stablize the central region of the helix with additional contributions from the intrastrand B5B6 interactions, favoring the two hydrogen bond scenario (Figure 1) and leading to their greater duplex stability (Table 1) over the 5′-GU-3′ mismatches. These observations strongly indicate a role for stacking interactions, in particular GG interstrand and GU intrastrand interactions, in influencing the stability of GU tandem mismatches.

B4B6/B6B4 interstrand stacking interactions may also contribute to the relative stabilities of the 5′-GU-3′ mismatches. Analysis of Table 6 shows the most favorable interaction energies to occur in **Ib**, associated with the GG interactions; this duplex is the most stable of the 5′-GU-3′ mismatches (Table 1). As the B4B6/B6B4 interstrand interaction energies becomes less favorable upon going from **IIb** to **IIIb**, there is a further destabilization of the duplex. Thus, while B4B6/B6B4 interstrand stacking interactions are suggested to destabilize 5′-GU-3′ versus 5′-UG-3′ mismatches, if those stacking interactions are GG interactions, they are predicted to stabilize the structure relative to other B4B6/B6B4 interstrand stacking pairs.

The ordering of the 5′-UG-3′ mismatch duplexes may also be related, in part, to interstrand stacking interactions. As stated above, all the G5G5 interactions are significantly favorable; however, the B4B6/B6B4 interstrand interactions in the 5′-UG-3′ mismatches show more variability. In **Ia**, these interaction energies are favorable, while in **IIa**, they are unfavorable, and in **IIIa**, they again are favorable with the values being more favorable with **Ia** versus **IIIa**. This is consistent with the greater stability of **Ia** (Table 1) and, notably, with **IIa** being slightly less stable than **IIIa**. The favorable G5G5 interaction in **Ia** and the favorable B4B6 interactions are suggested to lead to the dominance of the two hydrogen bond scenario at both 0.3 and 1 M salt as compared to **IIa** and **IIIa** (Figure 2), reinforcing the role of stacking interactions in stabilizing the hydrogen bonding scenario and thereby impacting duplex stability.

QM Validation of MM Energetics. The energetic data presented above is based on the CHARMM27 nucleic acid force field. While this force field reproduces a variety of experimental and QM energetic data, including data on base pairing, base stacking, and experimental heats of sublimation of bases (*29*), validation of the present observations is warranted. Accordingly, QM calculations were undertaken

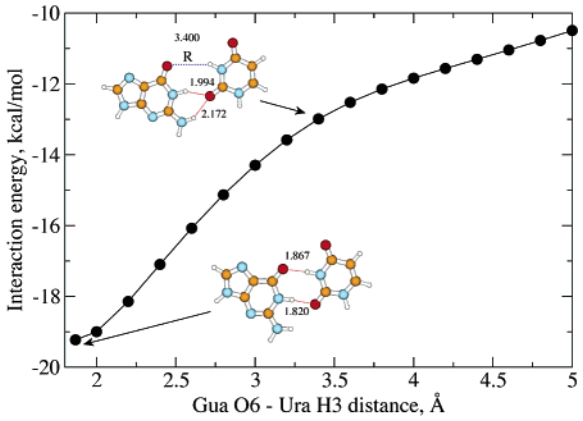

FIGURE 4: Interaction energy between G−U bases as a function of the GuaO6−UraH3 distance. The conformations of the base pair for the fully optimized structure, corresponding to the two hydrogen bond scenario, and with a GuaO6−UraH3 distance of 3.4 Å, approximating the one hydrogen bond scenario, are shown as insets.

on the GU hydrogen bonding interaction and various base stacking interactions. The studied geometries were selected to be consistent with those obtained from the MD simulations.

Experimental model compound studies indicate that normal two-center hydrogen bonds (e.g., N−H···O) are energetically favored over one three-centered hydrogen bond (Figure 1B) (*57*). Thus, the two-hydrogen bond scenario (Figure 1A), which contains two two-center hydrogen bonds, should be significantly more stable, consistent with the T_m data. To better quantify the relative energy of the one versus two hydrogen bond scenarios, QM calculations were undertaken on the GU base pair. Initially, the two individual bases, guanine and uracil, and the two complexes (A and B in Figure 1) were optimized at the MP2/6-31G* level of theory. All initial putative conformations chosen for complex B upon optimization collapsed to complex A indicating that the one hydrogen bond scenario is not a stable local minimum. Therefore, complex B was accessed by starting with the complex A optimized structure and gradually increasing the GuaO6−UraH3 distance in 0.2 Å increments with full MP2/6-31G* optimizations at each point. The resulting interaction energy profile is presented in Figure 4. Upon going from the fully optimized orientation, corresponding to the two-hydrogen bond scenario, to a one-hydrogen bond scenario, with a GuaO6−UraH3 distance of 3.4 Å, consistent the experimentally observed distance (*10*), there is a significant increase of energy. At the MP2/6-31G* level, this loss is calculated to be 6.2 kcal/mol, while single-point calculations on the MP2/6-31G* geometries at the MP2/6-311+G** with and without BSSE correction and the LMP2/cc-pVQZ (*58*) levels yielded changes in the interaction energies of 5.1, 3.7, and 1.9 kcal/mol, respectively. Thus, QM calculations verify the energetic favoring of the two over the one hydrogen bond scenario.

As discussed above, the present MD simulation study suggests that base stacking differences in 5′-UG-3′ versus 5′-GU-3′ mismatches allow for the one hydrogen bond scenario to be significantly sampled in the latter, leading to the decreased stability. This is due to the significantly favorable B4B5 intrastrand and B4B6/B6B4 interstrand base stacking interactions (Table 3). To validate this contribution, QM calculations were undertaken to obtain *ab initio* estimates of the interstrand and intrastrand stacking interaction energies. Calculations were performed starting from the interstrand G5G5 and intrastrand G5G6 base pair Cartesian coordinates obtained from snapshots at 1, 2, 3, 4, and 5 ns from the **Ia**, 0.3 M salt MD simulation following which the hydrogen positions were optimized. Calculations following the same protocol were performed on the monomers from the same snapshots with the identical procedure performed using CHARMM27 to allow for a one to one comparison between the QM and empirical results. Data in Table 7 includes the MP2 interaction energies for several basis sets, the BSSE, and the MP2 BSSE-corrected interaction energies along with the CHARMM27 results. The MP2 results show favorable interaction for both the intra- and intermolecular orientations for all the basis sets with intermolecular stacking interactions more favorable than the intramolecular ones for all the basis sets. Comparison of the QM and CHARMM27 results show the agreement to be reasonable. These results support the ability of the empirical force field calculations to satisfactorily treat base stacking interactions.

It should be emphasized that stacking interactions are difficult to treat accurately with QM methods due to the extensive orbital overlap occurring in the stacked orientations (*59*), making more rigorous comparison of the QM and empirical data difficult. This is emphasized by the large BSSE energies, the BSSE being larger for inter- versus intrastrand stacking, and the increased interaction energies upon going to the larger basis sets (Table 7). Indeed, for the benzene dimer, it has been shown that treatment of electron correlation with coupled cluster along with extrapolation to the basis set limit is required for accurate interaction energies (*60*) with similar conclusions being made for the uracil dimer (*61*).

Solvation Contributions. The central role of the solvent environment on oligonucleotide structure (*25*, *26*) requires that changes in solvation, which may impact the relative stabilities of the GU tandem mismatches, be considered. Possible alterations in solvation as a function of mismatch and sequence were investigated via calculation of the solvent accessibilities (*62*) of the central four bases over the MD simulations, with the accessibilities separated into the base and backbone components. In addition, solvent−RNA interaction energies were calculated for the central four bases. Those results are presented in Tables S5 and S6, respectively, of the Supporting Information. Analysis of the data revealed no correlations between the accessibilities with the T_m values (Table 1) across the sequences studied. Thus, the present results indicate that solvation effects are not dominating the experimentally observed stabilities, although, when considering the proposed contribution of stacking interactions in the present study, some contribution from differential solvation on duplex stability cannot be excluded.

DISCUSSION

MD simulations were undertaken on a series of RNA octamers that contain GU tandem mismatches, along with several control duplexes (Table 1). Overall, all the simulated structures maintained an A-form conformation based on both RMS difference (Table 2), average backbone dihedral angles (Tables S1 and S2 of the Supporting Information), and helicoidal parameters (Tables 3 and 4). However, detailed

Table 7: Guanosine−Guanosine Base Stacking Energies from QM Calculations Using Geometries Obtained from MD Simulations of Duplex **Ia**[a]

	6-31G(0.25)		6-311+G**		aug-cc-pVDZ	
	inter	intra	inter	intra	inter	intra
MP2	−15.3 ± 0.8	−10.8 ± 0.7	−10.4 ± 0.5	−7.7 ± 0.4	−12.3 ± 0.5	−8.8 ± 0.4
BSSE correction	11.2	8.2	7.2	5.7	7.3	5.3
MP2 corrected	−4.1 ± 0.4	−2.6 ± 0.3	−3.2 ± 0.5	−2.1 ± 0.4	−5.0 ± 0.4	−3.5 ± 0.3
CHARMM27					−5.8 ± 0.8	−1.3 ± 0.5

[a] Energies in kcal/mol. Averages and standard deviations are over the five snapshots obtained from the MD simulations. "Inter" represents the B5−B5 interstrand base stacking interaction and "intra" represents the B5−B6 intrastrand base stacking interaction (Table 5). MP2 corrected indicates MP2 level energies corrected for BSSE.

analysis of selected helicoidal parameters indicated the presence of significant local distortions in the central four bases. Importantly, these distortions showed trends common to the 5′-UG-3′ and 5′-GU-3′ tandem mismatches, suggesting a relationship between sequence, calculated structural changes, and duplex stability.

Analysis of hydrogen bonding between the bases in the GU mismatches shows the presence of the two and one hydrogen bond scenarios (Figure 1), as previously described (*10*). In the duplexes with 5′-UG-3′ tandem mismatches, the two hydrogen bond scenario dominated, while the one hydrogen bond scenario was significantly sampled in the 5′-GU-3′ tandem mismatches. These results are consistent with experimental observations (*10*, *13*, *14*) and indicate the dynamic nature of the hydrogen bonding pattern. Moreover, the results support the assertion that the stability of 5′-UG-3′ over 5′-GU-3′ tandem mismatches is due to enhanced sampling of the two hydrogen bond scenario (*10*, *13*).

These observations still left open the question of what was leading to the presence of the one hydrogen bond scenario in the 5′-GU-3′ tandem mismatches, which QM calculations show to be 4 or more kcal/mol less favorable than the two hydrogen bond scenario (Figure 4, i.e., each GU pair in the tandem mismatch contributes approximately 2−6 kcal/mol). Turner and co-workers approached the problem based on both the electrostatic potentials of the bases (*10*, *13*) and different stacking patterns (*10*). In their electrostatic model, regions of negative potential on the edge of the bases in the major groove were suggested to overlap significantly in the 5′-GU-3′ mismatch in **IIIb**, corresponding to bases 4 and 5 in the present study, while they were approximately orthogonal in the 5′-UG-3′ mismatch in **IIIa**, leading to the increased stability of the latter. Turner and co-workers further argued that the conformational difference between **IIa** and **Ib** is due to the overlap of the G base with the adjacent C base in the same strand and also with the G base across the strand in **IIa**, leading to the increased stability, while the only significant overlap in **Ib** involves the intrastrand GU interaction (i.e., B4−B5 in Table 5) (*10*).

These previous models can be compared to the data in Tables 5 and 6. In the model based on the differential overlap of electrostatic potentials in the position 4 to 5 stacked base pairs of **IIIa** and **IIIb** (Figure 10 of ref *13*), decreased overlap of **IIIa** is responsible for its greater stability. However, the present results show that the intrastrand B4−B5 interaction in **IIIa** is less favorable than that in **IIIb**, in disagreement with their qualitative observation. Concerning the relative stabilities of **IIa** versus **Ib**, the present data support the proposed role of G to C intrastrand and G to G interstrand stacking in **IIa**, where both interactions in the present work are highly favorable (i.e., B5−B6 intrastrand and B5−B5 interstrand in Tables 5 and 6, respectively). Thus, Turner and co-workers were correct in their hypotheses of the importance of stacking interactions on the relative stabilities of GU tandem mismatches, assuming that the electrostatic potential model is also a stacking interaction, with the present work yielding a more refined picture of the relationship of stacking to structural properties and stability. Interestingly, previous work on GU mismatches suggested that stacking interactions may contribute to the decreased stability of duplexes when the uracil is replaced by 5-fluorouracil (*63*), further supporting the role of stacking interactions in influencing the properties of GU mismatches.

On the basis of the present analysis of various base step stacking interactions, it is proposed that favorable interstand stacking interactions between G bases (i.e., the G5−G5 interstrand interactions, Table 6) and intrastrand stacking between GU bases (i.e., the G4−U5 intrastrand interactions, Table 5) lead to the stabilization of the two hydrogen bond scenario in 5′-UG-3′ mismatches. This stabilization of the two hydrogen bond scenario is responsible for the increased stabilities of the 5′-GU-3′ versus the 5′-UG-3′ mismatches (Table 1). Furthermore, GG interactions appear to contribute to the relative ordering of the stabilities of the 5′-GU-3′ containing duplexes, with the least stable duplexes, **IIb** and **IIIb**, lacking GG interstrand stacking interactions (i.e., B4B6/B6B4 interstrand interactions, Table 6). Finally, the relative stabilities of the 5′-UG-3′ mismatches, **Ia** > **IIIa** > **IIa** (Table 1), are suggested to be associated with interstrand B4B6/B6B4 interactions. Thus, the present results support a model in which GG interstrand base stacking interactions, followed by GU intrastrand stacking interactions, dominate the relative stabilities of GU tandem mismatches, with other stacking interactions also having an impact.

Two of the types of stacking interactions that have the most impact in the present study, the GU intrastrand (B4B5 in Table 5) and the GG interstrand interactions (B4B6, B5B5, and B6B4 in Table 6), are not unique to the GU tandem mismatches. Shown in Figure 5A,B are RNA 5′-UG-3′ and 5′-GU-3′ dimers in the canonical A-form conformation (*49*). For the 5′-UG-3′ dimer, the overlap of the G bases, corresponding to the GG interstrand interactions, is evident. Similarly, the overlap of the GU bases, corresponding to the GU intrastand stack, may be seen in Figure 5B for the 5′-GU-3′ dimer. Thus, stacking interactions indicated in the present work to influence the structure and stability of GU mismatches are not specific to the GU tandem mismatch structures but occur in the canonical A-form of RNA.

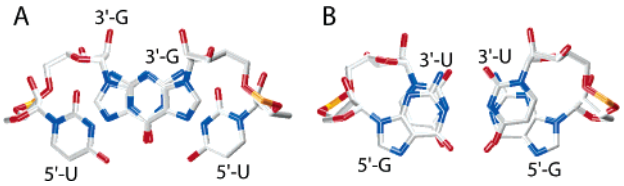

FIGURE 5: Structures of (A) 5′-UG-3′ and (B) 5′-GU-3′ RNA dimers in the canonical A-form (*49*).

To validate the present MD results a series of QM calculations were undertaken on the in-plane base pairing between G and U bases and on both inter- and intrastrand stacking interactions between G bases. The two hydrogen bond scenario is shown to be more favorable than the one hydrogen bond scenario by 2−6 kcal/mol (Figure 4). QM calculations of stacking support the favorable nature of the G−G stacking interactions with interstrand GG stacking being more favorable than intrastrand stacking as is observed in the empirical model. More quantitative agreement between the QM and empirical models for the stacking interactions requires higher level QM calculations that are difficult to perform via current computational resources.

The hydrogen bonding strength of the GU base pairs relative to the normal Watson−Crick GC interaction is also of interest. With the GU mismatches, the two and one hydrogen bond scenarios have interaction energies of −10.0 and −8.1 kcal/mol at the LMP2/cc-pVQZ//MP2/6-31G* level, significantly less than the value of −22.2 kcal/mol for the GC base pair (*58*). Such a decrease in the GU mismatch versus the GC WC pair is expected due to the decreased number of hydrogen bonds. However, the magnitude of the difference suggests that the base stacking interactions indicated to impact the stability and structure of the GU mismatches will not have as large an impact on duplexes in the case of normal GC WC base pairs. This is consistent with the GG interstrand and GU intrastrand interactions being present in canonical A-form RNA, although these interactions do not appear to play as dominate a role in normal RNA as they do in the GU tandem mismatches.

A point of interest with respect to the nearest neighbor model is the inability of the model to fit all 5′-GU-3′ mismatch pairs (*10*, *24*). For example, a model developed to treat 5′-GGUC-3′ (**IIb**) cannot treat 5′-CGUG-3′ (**Ib**), 5′-UGUA-3′, and 5′-AGUU-3′ (**IIIb**) and vice versa. Notably, 5′-GU-3′ in the 5′-GGUC-3′ sequence is stabilizing while in the context of the remaining sequences it is destabilizing. Based on the present results, the stabilizing contribution of the 5′-GGUC-3′ mismatch is suggested to be associated with the ability of the sequence to sample the two hydrogen bond scenario as well as the one hydrogen bond scenario (Figure 2), consistent with previous reports (*10*). This appears to be related to the failure of the nearest neighbor model, but again, what is leading to this effect? Analysis of the base step helicoidal parameters (Table 3) suggest a possible cause. While the patterns for the central three steps (e.g., 3−4, 4−5, and 5−6) do not show any interesting trends, analysis of the 2−3 and 6−7 steps do. In the case of slide, **Ib** and **IIIb** have the least negative values, while with rise, they have the largest values, trends that are also observed in the NMR structures. With twist, the RMS fluctuations from the MD simulations are large, making it difficult to identify trends, although in the NMR structures, the **IIb** twist values for the 2−3 and 6−7 steps are smaller than both the **Ib** and **IIIb** structures. To investigate this further, stacking energies involving bases 2 and 7 bases were analyzed (Tables S3 and S4 of the Supporting Information). From this analysis, it was observed that for the B2B3 intrastrand stack (Table 5), the interaction energy is least favorable for **IIb** while it is most favorable in that sequence for the B3B7 and B7B3 interstrand stacking interactions (Table 6). Such differences are suggested to account for the trends observed in the helicoidal parameters as well as the inability of the nearest neighbor model to simultaneously treat sequences **Ib**, **IIb**, and **IIIb**. These results further emphasize the importance of stacking interactions on the stability of GU tandem mismatches. It should be noted that the significant differences between the 5′-UG-3′ sequences concerning these stacking interactions are also present. However, it is hypothesized that the increased stability of the central region of the duplex, as evidenced by the dominance of the two hydrogen bond scenario, diminishes the impact of these interactions on the thermodynamic stability.

Overall, the present results indicate that the stability of 5′-UG-3′ tandem mismatches over 5′-GU-3′ mismatches is associated with stacking interactions that allow for increased sampling of the two hydrogen bonding scenario for the central GU mismatches. In particular, GG, followed by GU, stacking interactions are predicted to have the largest impact on the stability, although other stacking interactions contribute. Thus, it appears that the higher frequency of 5′-UG-3′ over 5′-GU-3′ tandem mismatches (*3*, *64*) is due, in part, to the optimization of base stacking interactions leading to the increased stability of the former.

SUPPORTING INFORMATION AVAILABLE

Tables providing backbone dihedral angles for the central four nucleotides, intra- and interstrand base−base interactions involving the central five base steps, solvent accessibilities of the central four bases, and solvent−RNA interaction energies for the central four bases. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gautheret, D., Konings, D., and Gutell, R. R. (1995) G.U base pairing motifs in ribosomal RNA, *RNA 1*, 807−814.
2. Damberger, S. H., and Gutell, R. R. (1994) A comparative database of group I intron structures, *Nucleic Acids Res. 22*, 3508−3510.
3. Gutell, R. R. (1994) Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994, *Nucleic Acids Res. 22*, 3502−3507.
4. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase, *Nature 352*, 213−218.
5. McClain, W. H., Chen, Y. M., Foss, K., and Schneider, J. (1988) Association of transfer RNA acceptor identity with a helical irregularity, *Science 242*, 1681−1684.
6. Gabriel, K., Schneider, J., and McClain, W. H. (1996) Functional evidence for indirect recognition of G.U in tRNA(Ala) by alanyl-tRNA synthetase, *Science 271*, 195−197.
7. Doudna, J. A., Cormack, B. P., and Szostak, J. W. (1989) RNA structure, not sequence, determines the 5′ splice-site specificity of a group I intron, *Proc. Natl. Acad. Sci. U.S.A. 86*, 7402−7406.
8. Simpson, L., and Thiemann, O. H. (1995) Sense from nonsense: RNA editing in mitochondria of kinetoplastid protozoa and slime molds, *Cell 81*, 837−840.
9. Wu, M., McDowell, J. A., and Turner, D. H. (1995) A periodic table of symmetric tandem mismatches in RNA, *Biochemistry 34*, 3204−3211.

10. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for G·U pairs: one hydrogen bond for each G·U pair in r(GGCGUGCC)$_2$ and two for each G·U pair in r(GAGUGCUC)$_2$, *Biochemistry 39*, 8970−8982.
11. Allawi, H. T., and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G·T mismatches in DNA, *Biochemistry 36*, 10581−10594.
12. Tanaka, Y., Kojima, C., Yamazaki, T., Kodama, T. S., Yasuno, K., Miyashita, S., Ono, A., Kainosho, M., and Kyogoku, Y. (2000) Solution structure of an RNA duplex including a C−U base pair, *Biochemistry 39*, 7074−7080.
13. McDowell, J. A., He, L., Chen, X., and Turner, D. H. (1997) Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of (rGGAGUUCC)$_2$ and (rGGAUGUCC)$_2$, *Biochemistry 36*, 8030−8038.
14. McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)$_2$ by two-dimensional NMR and simulated annealing, *Biochemistry 35*, 14077−14089.
15. Burkard, M. E., and Turner, D. H. (2000) NMR structures of r(GCAGGCGUGC)$_2$ and determinants of stability for single guanosine−guanosine base pairs, *Biochemistry 39*, 11748−11762.
16. Matthews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure, *J. Mol. Biol. 288*, 911−940.
17. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
18. Zhu, J., and Wartell, R. M. (1997) The relative stabilities of base pair stacking interactions and single mismatches in long RNA measured by temperature gradient gel electrophoresis, *Biochemistry 36*, 15326−15335.
19. Morse, S. E., and Draper, D. E. (1995) Purine-purine mismatches in RNA helices: evidence for protonated G.A pairs and next-nearest neighbor effects, *Nucleic Acids Res. 23*, 302−306.
20. Martinez, H. M. (1990) Detecting pseudoknots and other local base-pairing structures in RNA sequences, *Methods Enzymol. 183*, 306−317.
21. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule, *Science 244*, 48−52.
22. Gultyaev, A. P., van Batenburg, F. H., and Pleij, C. W. (1995) The computer simulation of RNA folding pathways using a genetic algorithm, *J. Mol. Biol. 250*, 37−51.
23. Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., and Zuker, M. (1994) Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding, *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
24. He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., and Turner, D. H. (1991) Nearest-neighbor parameters for G·U mismatches: 5′GU3′ is destabilizing in the contexts CGUG, UGUA, and AGUU but stabilizing in GGUC, *Biochemistry 30*, 11124−11132.
25. Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides, *Q. Rev. Biophys. 11*, 179−246.
26. Record, M. T., Jr., Anderson, C. F., and Lohman, T. M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity, *Q. Rev. Biophys. 11*, 103−178.
27. MacKerell, A. D., Jr., and Nilsson, L. (2001) in *Computational Biochemistry and Biophysics* (Becker, O. M., MacKerell, A. D., Jr., Roux, B., and Watanabe, M., Eds.) pp 441−463, Marcel Dekker, Inc., New York.
28. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
29. Foloppe, N., and MacKerell, A. D., Jr. (2000) All-atom empirical force field for nucleic acids. 1. Parameter optimization based on small molecule and condensed phase macromolecular target data, *J. Comput. Chem. 21*, 86−104.
30. Foloppe, N., Nilsson, L., and MacKerell, A. D., Jr. (2001) Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics, *Biopolymers 61*, 61−76.
31. Beglov, D., and Roux, B. (1997) Finite representation of an infinite bulk system: solvent boundary potential for computer simulations, *J. Chem. Phys. 100*, 9050−9063.
32. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926−935.
33. Arnott, S., and Hukins, D. W. (1973) Refinement of the structure of B-DNA and implications for the analysis of X-ray diffraction data from fibers of biopolymers, *J. Mol. Biol. 81*, 93−105.
34. *Quanta* (2001) Accelrys Inc., San Diego, CA.
35. Field, M. J., and Karplus, M. (1992) *CRSTAL: Program for crystal calculations in CHARMM*, Harvard University, Cambridge, MA.
36. Ryckaert, J., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *J. Comput. Phys. 23*, 327−341.
37. Darden, T., Perera, L., Li, L., and Pedersen, L. (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations, *Struct. Fold Des. 7*, R55−R60.
38. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T. A., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh ewald method, *J. Chem. Phys. 103*, 8577−8593.
39. Steinbach, P. J., and Brooks, B. R. (1994) New Spherical-Cutoff Methods of Long-Range Forces in Macromolecular Simulations, *J. Comput. Chem. 15*, 667−683.
40. Hoover, W. G. (1985) Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A 31*, 1695−1697.
41. Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, R. W. (1995) Constant pressure molecular dynamics simulations: the Langevin piston method, *J. Chem. Phys. 103*, 4613−4621.
42. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Jr., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., and Pople, J. A. (1998) *Gaussian 98*, revision A.6, Gaussian, Inc., Pittsburgh, PA.
43. Woon, D. E., and Dunning, T. H., Jr. (1993) Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon, *J. Chem. Phys. 98*, 1358−1371.
44. Møller, C., and Plesset, M. S. (1934) Note on an Approximation Treatment for Many-Electron Systems, *Phys. Rev. 46*, 618−622.
45. Saebø, S., Tong, W., and Pulay, P. (1993) Efficient elimination of basis set superposition error by the local correlation method: Accurate ab inito calculations of the water dimer, *J. Chem. Phys. 98*, 2170−2175.
46. Hobza, P., and Sponer, J. (2002) Toward True DNA Base-Stacking Energies: MP2, CCSD(T), and Complete Basis Set Calculations, *J. Am. Chem. Soc. 124*, 11802−11808.
47. Hobza, P., and Sponer, J. (1999) Structure, Energetics and Dynamics of the Nucleic Acid Base Pairs: Nonempirical Ab initio Calculations, *Chem. Rev. 99*, 3247−3276.
48. Boys, S. F., and Bernardi, F. (1970) The Calculation of Small Molecular Interaction by the Differences of Separate Total Energies. Some Procedures with Reduced Errors, *Mol. Phys. 19*, 553−566.
49. Arnott, S., Hukins, D. W. L., Dover, S. D., Fuller, W., and Hodgson, A. R. (1973) Structures of Synthetic Polynucleotides in the A-RNA and A′-RNA Conformations: X-ray Diffraction Analyses of the Molecular Conformations of Polyadenylic acid-polyuridic acid and Polyinosinic acid-polycytidylic acid, *J. Mol. Biol. 81*, 102−122.
50. Biswas, R., Wahl, M. C., Ban, C., and Sundaralingam, M. (1997) Crystal structure of an alternating octamer r(GUAUGUA)dC with adjacent G x U wobble pairs, *J. Mol. Biol. 267*, 1149−1156.
51. Kneale, G., Brown, T., Kennard, O., and Rabinovich, D. (1985) G. T base-pairs in a DNA helix: the crystal structure of d(G-G-G-G-T-C-C-C), *J. Mol. Biol. 186*, 805−814.

52. Kieft, J. S., and Tinoco, I., Jr. (1997) Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine, *Structure 5*, 713−721.
53. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
54. Calladine, C. R. (1982) Mechanics of Sequence-dependent Stacking of Bases in B-DNA, *J. Mol. Biol. 161*, 343−352.
55. Petersheim, M., and Turner, D. H. (1983) Base-Stacking and Base-Pairing Contributions to Helix Stability: Thermodynamics of Double-Helix Formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp, *Biochemistry 22*, 256−263.
56. Marky, L. A., and Breslauer, K. J. (1982) Calorimetric Determination of Base-Stacking Enthalpies in Double-Helical DNA Molecules, *Biopolymers 21*, 2185−2194.
57. Yang, J. T., and Gellman, S. H. (1998) Energetic Superiority of Two-Center Hydrogen Bonding Relative to Three-Center Hydrogen Bonding in a Model System, *J. Am. Chem. Soc. 120*, 9090−9091.
58. Huang, N., and MacKerell, A. D., Jr. (2002) An Ab Initio Quantum Mechanical Study of Hydrogen-Bonded Complexes of Biological Interest, *J. Phys. Chem. B 106*, 7820−7827.
59. Rappé, A. K., and Bernstein, E. R. (2000) Ab Initio Calculation of Nonbonded Interactions: Are We There Yet?, *J. Phys. Chem. A 104*, 6117−6128.
60. Sinnokrot, M. O., Valeev, E., and Sherril, C. D. (2002) Estimates of the Ab Initio Limit for $\pi-\pi$ Interactions: The Benzene Dimer, *J. Am. Chem. Soc. 124*, 10887−10893.
61. Leininger, M. L., Nielsen, I. M. B., Colvin, M. E., and Janssen, C. L. (2002) Accurate Structures and Binding Energies for Stacked Uracil Dimers, *J. Phys. Chem. A 106*, 3850−3854.
62. Lee, B., and Richards, F. M. (1971) The Interpretation of Protein Structures: Estimation of Static Accessibility, *J. Mol. Biol. 55*, 379−400.
63. Sahasrabudhe, P. V., and Gmeiner, W. H. (1997) Solution Structures of 5-Fluorouracil-Substituted RNA Duplexes Containing G−U Wobble Base Pairs, *Biochemistry 36*, 5981−5991.
64. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993, *Nucleic Acids Res. 21*, 3055−3074.

BI047932Q