Thermodynamics of Single Mismatches in RNA Duplexes†

Ryszard Kierzek,‡ Mark E. Burkard,§ and Douglas H. Turner*,§

Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-704 Poznan, Noskowskiego, 12/14 Poland, and Department of Chemistry, RC Box 270216, University of Rochester, Rochester, New York 14627-0216

Received May 24, 1999; Revised Manuscript Received August 25, 1999

ABSTRACT: The thermodynamic properties and structures of single mismatches in short RNA duplexes were studied in optical melting and imino proton NMR experiments. The free energy increments at 37 °C measured for non-GU single mismatches range from −2.6 to 1.7 kcal/mol. These increments depend on the identity of the mismatch, adjacent base pairs, and the position in the helix. UU and AA mismatches are more stable close to a helix end, but GG mismatch stability is essentially unaffected by the position in the helix. Approximations are suggested for predicting stabilities of single mismatches in short RNA duplexes.

Ribonucleic acids form complicated three-dimensional structures that control activity. The structures are dominated by helixes of Watson−Crick pairs with intervening non-Watson−Crick paired regions. One non-Watson−Crick motif is the internal loop, which contains unpaired nucleotides on both strands of a region flanked on each side by a Watson−Crick helix. The most common internal loop in ribosomal RNA is the smallest, a single mismatch (*1*).

A great deal of thermodynamic and structural information has been obtained for Watson−Crick paired RNA helixes, but less is known about regions without Watson−Crick pairs. Studies of tandem mismatches have shown that the energetics and structures of these internal loops are strongly dependent on the sequence in the loop and on the adjacent base pair (*2*−*4*). The sheer number of possible tandem mismatches makes it difficult to survey all possible combinations with the current state of technology. In contrast, there are only eight single mismatches so a relatively comprehensive study is possible.

Thermodynamics for certain single mismatches in certain contexts have been reported. For single AA mismatches, Peritz et al. (*1*) reported that the free energy increment, $\Delta G^{o}_{37-\text{loop}}$, is unfavorable by about 1 kcal/mol for $\begin{smallmatrix}5'\text{GAG}3'\\3'\text{CAC}5'\end{smallmatrix}$ and $\begin{smallmatrix}5'\text{CAG}3'\\3'\text{GAC}5'\end{smallmatrix}$. A UU mismatch in $\begin{smallmatrix}5'\text{CUG}3'\\3'\text{GUC}5'\end{smallmatrix}$ was reported to destabilize the duplex by 0.8 kcal/mol at 37 °C (*2*). Gralla and Crothers (*5*) reported the effect on duplex stability of different sizes of internal loops containing cytosines. They found that a single CC mismatch destabilizes the duplex and that the free energy increment depends on the adjacent base pairs; with GC adjacent base pairs, the loop destabilizes the duplex by 0.1 kcal/mol, whereas with AU adjacent base pairs, it destabilizes the duplex by 1.8 kcal/mol. Morse and Draper (*6*) measured free energies for purine-purine single mismatches. On the basis of their data, purine-purine single mismatches have a wide range of loop free energies, from −3 to 5 kcal/mol at 37 °C. Several factors in addition to mismatch identity may contribute to this range because the mismatches have a variety of adjacent base pairs and mismatch positions in the helix, and because some duplexes have two single mismatches per duplex whereas others have one. On the basis of the optical melting of hairpins, Bevilacqua and Bevilacqua (*7*) found that a GG mismatch is 1.7 kcal/mol more stable than GA at 37 °C. On the basis of similar experiments with hairpins containing CA, GA, and UU mismatches in the stem, Meroueh and Chow (*8*) found that mismatch stabilities differed by as much as 1.4 kcal/mol. A somewhat smaller range of free energies is observed in temperature gradient gel electrophoresis studies of single mismatches placed at various positions in an RNA with 345 base pairs (*9*).

There are few structural studies of single mismatches in RNA. In an NMR study of a single UU mismatch within a conserved ribosomal RNA sequence, Wang et al. (*10*) observed one N3−H3···O4 hydrogen bond and suggested that water bridges a second hydrogen bond between O2 and N3−H3 from the other uracil residue. They conclude that UU single mismatches "adopt a range of conformations, varying in the extents of propeller twist, imino proton hydrogen bonding and backbone distortion."

The work reported here provides a survey of the sequence dependence of single mismatch thermodynamics in RNA duplexes. Seven mismatches are compared in the same context. Then, some effects of adjacent base pairs are determined for UU and AA mismatches. The effect of the UU, AA, and GG mismatch position in the duplex is also explored. Finally, some structural information is derived from one-dimensional NMR spectra of selected duplexes.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Purification. RNA samples were synthesized and purified as described previously (*11*). Monomers were obtained as phosphoramidites from Glen Research Inc. (Baltimore, MD). After thin-layer chromatography, the least mobile product band, almost always the most intense, was visualized with ultraviolet light, cut out, and

† This work was supported by NIH Grants GM22939 and TW1068. M.E.B. is a trainee in the Medical Scientist Training Program funded by NIH Grant T32 GM 07356.

* To whom correspondence should be addressed. Telephone: (716) 275-3207. Fax: (716) 473-6889. E-mail: Turner@chem.rochester.edu.

‡ Polish Academy of Sciences.

§ University of Rochester.

10.1021/bi991186l CCC: $18.00
Published on Web 10/05/1999

eluted with water. Then the samples were desalted with a Sep-pak C-18 cartridge. The purity of each oligomer was checked by HPLC[1] on a C-8 analytical reverse phase column (Hamilton), and was greater than 95%.

UV Melting and Thermodynamic Parameters. Thermodynamic parameters were measured in 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA (pH 7.0). Oligoribonucleotide single-strand concentrations, C_T, were calculated from high-temperature absorbances and single-strand extinction coefficients (*12*, *13*). For non-self-complementary duplexes, oligomers were mixed at a 1:1 ratio. Small errors in mixing ratios are not expected to affect thermodynamic measurements (*1*). Melting curves of absorbance versus temperature were recorded at 280 nm with a heating rate of 1 °C/min from 0 to 90 °C on a Gilford 250 spectrometer controlled by a Gilford 2527 thermoprogrammer. Thermodynamic parameters for duplex formation were derived by fitting the shape of each curve to the two-state model with sloping baselines using a nonlinear least-squares program (*14*, *15*), and by plotting the reciprocal of the melting temperature, T_M^{-1}, versus $\ln(C_T/4)$ for non-self-complementary sequences (*16*):

$$T_M^{-1} = \frac{R}{\Delta H^\circ} \ln(C_T/4) + \frac{\Delta S^\circ}{\Delta H^\circ} \quad (1)$$

where R is the gas constant, 1.987 cal K^{-1} mol^{-1}.

NMR Spectra. For NMR spectroscopy, non-self-complementary samples were mixed at a 1:1 ratio. Samples were dried down and redissolved in a 9:1 v/v H_2O/D_2O solution with 80 mM NaCl, 0.5 mM EDTA, and 10 mM phosphate buffered at pH 7. A small amount of TSP was added to some samples as a reference; for others, the third peak of the free EDTA quartet was used as a reference at 3.11 ppm. One-dimensional spectra were recorded on a Varian INOVA 500 MHz spectrometer with a binomial 1:3:3:1 pulse sequence to suppress the water signal (*17*). The frequency offset was selected to maximize the signal near ~13 ppm with first nodes near 5 and 21 ppm. The free induction decay was multiplied by a 2−4 Hz line-broadening function, Fourier transformed, and visualized with the Varian VNMR software package. NOE spectra were obtained by selective saturation of imino proton resonances for 2−3 s, followed by the binomial 1:3:3:1 pulse sequence. Difference spectra for NOEs were calculated by subtracting the spectrum from a reference spectrum, obtained by decoupling at least 250 Hz downfield of the imino resonance with the largest chemical shift.

Two-dimensional Watergate NOESY spectra were acquired on a 500 or 600 MHz Varian INOVA spectrometer. Spectra were collected with mixing times of 2, 10, 100, and 150 ms. For each spectrum, 256 t_2 intervals were used with 64 free induction decays (fids) per interval. For each fid, 4K complex points were collected with a spectral width of 14 000 Hz on the 600 MHz instrument. The recycle delay was 2 s. The Watergate pulse sequence interposed a delay of ~2 ms between the readout pulse and the beginning of acquisition, but did not affect the NOE mixing time. Spectra were apodized with a sine-bell squared function, transformed, and analyzed with the Felix 95.0 software package (Molecular Simulations, Inc.).

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; HPLC, high-pressure liquid chromatography; NOE, nuclear Overhauser effect; T_M, melting temperature.

RESULTS

Thermal Stability of Seven Single Mismatches in the Context $^{5'GAGMGAG3'}_{3'CUCNCUC5'}$. The effect of single mismatches on RNA duplex stability was quantified with the sequences $^{5'GAGMGAG3'}_{3'CUCNCUC5'}$, where M represents A, C, or U and N represents A, C, G, or U. The sequence 5′GAGGGAG3′ was difficult to purify, and melting curves with this sequence appeared to be non-two-state so this sequence was not used. Problems with three consecutive guanines have been observed before (*2*), and may be due to tetraplex formation or aggregation. Therefore, GG mismatches were studied in different contexts (see below). The duplex sequences and measured thermodynamic parameters are listed in Table 1.

The effect of single mismatches was evaluated as the free energy of loop formation, $\Delta G^\circ_{37-loop}$, calculated from (*5*)

$$\Delta G^\circ_{37-loop} = \Delta G^\circ_{37-duplex} - \Delta G^\circ_{37-duplex-w/o-mm} + \Delta G^\circ_{37-NN} \quad (2)$$

where $\Delta G^\circ_{37-duplex}$ is the free energy of formation of the duplex containing the mismatch, $\Delta G^\circ_{37-duplex-w/o-mm}$ is the free energy of the duplex, $^{5'GAGGAG3'}_{3'CUCCUC5'}$, without the mismatch (*4*), and ΔG°_{37-NN} is the free energy for the nearest neighbor $^{5'GG3'}_{3'CC5'}$, −3.26 kcal/mol (*11*), which was interrupted by the mismatch. Similar calculations provided values for the ΔH° and ΔS° of loop formation. The loop thermodynamic parameters are listed in Table 2. The free energy of loop formation by single mismatches other than GU in the context $^{5'GAGMGAG3'}_{3'CUCNCUC5'}$ ranges between −0.4 and 0.6 kcal/mol. The GU mismatch motif, $^{5'GUG3'}_{3'CGC5'}$, is considerably more stable, with a $\Delta G^\circ_{37-loop}$ of −2.5 kcal/mol.

Effect of Adjacent Base Pairs of Single Mismatches on Duplex Stability. Adjacent base pairs affect the free energies of internal loops (*3*, *5*). Moreover, a search of RNA secondary structures derived from comparative sequence analysis (*18*−*20*) reveals that the UU-containing motif $^{5'GUC3'}_{3'CUG5'}$ is about 15-fold more abundant than $^{5'CUC3'}_{3'GUC5'}$ (Table 3). In contrast, AA mismatches appear to have less preference for any particular adjacent Watson−Crick base pairs and GG has small preferences (Table 3). For this reason, the thermodynamics of UU, AA, and GG mismatches were studied with a variety of adjacent base pairs.

The effects of adjacent base pairs on UU mismatch stability were measured in several duplexes (Tables 1 and 2). The results show that UU mismatch stability exhibits a strong dependence on adjacent base pairs (Table 2). When a UU mismatch with adjacent GC pairs is located three base pairs from a helix end, $\Delta G^\circ_{37-loop}$ ranges between −0.8 and 0.5 kcal/mol, with $^{5'GUC3'}_{3'CUG5'}$ and $^{5'CUG3'}_{3'GUC5'}$ being the most and least stable, respectively. This correlates with the phylogenetic prevalence of $^{5'GUC3'}_{3'CUG5'}$ in rRNA (Table 3). On average, the loop $^{5'GUG3'}_{3'CUC5'}$ stabilizes duplex formation by about −0.1 kcal/mol, close to the average of $^{5'GUC3'}_{3'CUG5'}$ and $^{5'CUG3'}_{3'GUC5'}$ loops.

Two of the sequences with UU mismatches adjacent to only AU pairs did not melt in a two-state manner. Extrapolations from the melting temperatures to 37 °C are relatively

Table 1: Thermodynamic Parameters of Duplex Formation in 1 M NaCl[a]

	$1/T_M$ plots				curve fit parameters			
	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M[b] (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M[b] (°C)
5′GAGMGAG3′/3′CUCNCUC5′ series								
GAGUGAG3′/CUCGCUC5′	7.78 ± 0.05	64.2 ± 1.9	181.8 ± 6.1	43.1	8.00 ± 0.28	70.3 ± 4.9	200.8 ± 14.9	43.6
GAGCGAG3′/CUCACUC5′	5.59 ± 0.03	55.6 ± 2.1	161.1 ± 6.7	31.9	5.64 ± 0.11	54.2 ± 3.2	156.6 ± 10.4	32.0
GAGAGAG3′/CUCGCUC5′	5.34 ± 0.03	53.7 ± 2.2	156.0 ± 7.1	30.3	5.40 ± 0.12	47.5 ± 1.0	135.8 ± 3.2	29.8
GAGCGAG3′/CUCUCUC5′	5.27 ± 0.10	52.7 ± 3.9	153.1 ± 12.8	29.8	5.30 ± 0.10	50.7 ± 4.5	146.4 ± 14.8	29.7
GAGUGAG3′/CUCUCUC5′	5.21 ± 0.02	54.9 ± 1.4	160.1 ± 4.7	29.7	5.22 ± 0.20	63.3 ± 6.2	187.3 ± 19.5	30.7
GAGAGAG3′/CUCCCUC5′	4.93 ± 0.07	40.9 ± 2.1	115.9 ± 7.1	25.3	5.04 ± 0.24	42.3 ± 7.6	120.0 ± 25.0	26.5
GAGAGAG3′/CUCACUC5′	(4.83)	(54.6)	(160.4)	27.6	(4.97)	(45.0)	(129.1)	26.6
GAGCGAG3′/CUCCCUC5′	4.72 ± 0.09	39.3 ± 2.9	111.4 ± 9.4	23.3	4.84 ± 0.14	38.1 ± 7.1	107.4 ± 23.0	23.9
GAGUGAG3′/CUCCCUC5′	4.69 ± 0.12	50.0 ± 3.1	145.9 ± 10.5	26.0	4.44 ± 0.15	57.0 ± 3.1	169.4 ± 10.2	26.0
UU mismatches with different surrounding base pairs								
GCGUCCG3′/CGCUGGC5′	8.24 ± 0.07	62.6 ± 1.7	175.3 ± 5.1	45.7	8.45 ± 0.21	66.5 ± 4.2	187.2 ± 12.8	46.2
(GCGUCGC3′)$_2$	7.89 ± 0.03	66.8 ± 1.3	189.9 ± 4.0	47.6	7.98 ± 0.16	68.6 ± 4.0	195.5 ± 12.5	47.8
GCGUGCG3′/CGCUCGC5′	7.82 ± 0.05	61.1 ± 1.7	171.9 ± 5.3	43.7	7.89 ± 0.14	62.4 ± 2.7	175.8 ± 8.4	43.9
GCCUGCG3′/CGGUCGC5′	8.13 ± 0.06	61. 4 ±1.5	171.7 ± 4.6	45.3	8.32 ± 0.18	65.3 ± 3.3	183.7 ± 10.3	45.7
(CGCUGCG3′)$_2$	6.31[c]	56.8[c]	162.8[c]					
GCGUUCG3′/CGCUAGC5′	(6.03)	(57.2)	(165.1)	34.3	(6.05)	(31.3)	(81.5)	32.3
GUGUUCG3′/CACUAGC5′	3.82 ± 0.10	44.0 ± 2.8	129.6 ± 9.3	19.0	3.91 ± 0.34	46.6 ± 11.7	137.6 ± 38.7	20.5
GCAUUCG3′/CGUUAGC5′	(4.94)	(49.3)	(142.9)	27.3	(4.86)	(60.4)	(179.2)	28.7
GCAUACG3′/CGUUUGC5′	(4.58)	(48.1)	(140.4)	24.9	(4.37)	(60.5)	(181.1)	26.3
GCUUACG3′/CGAUUGC5′	4.50 ± 0.12	52.2 ± 3.1	153.9 ± 10.2	25.4	4.32 ± 0.19	60.7 ± 6.7	181.7 ± 21.5	26.1
UU mismatches one or two base pairs from helix end								
GUCCGCG3′/CUGGCGC5′	9.25 ± 0.05	56.9 ± 0.8	153.8 ± 2.5	52.6	9.26 ± 0.29	57.6 ± 4.0	155.9 ± 12.0	52.4
CGUCCGG3′/GCUGGCC5′	8.47 ± 0.05	61.8 ± 1.5	171.9 ± 4.6	47.0	8.47 ± 0.11	61.5 ± 3.1	170.9 ± 9.7	47.1
CGUCGUCG3′/GCUGCUGC5′	4.93 ± 0.05	56.5 ± 1.5	166.3 ± 5.0	33.0	4.91 ± 0.11	57.8 ± 3.1	170.6 ± 10.3	32.9
GCUGCUGC3′/CGUCGUCG5′	5.10±0.09	41.7±2.3	118.1±7.8	32.8	5.11±0.22	43.9±6.4	125.0±21.2	33.0
AA mismatches with different surrounding base pairs								
(CGCAGCG3′)$_2$	6.09[d]	50.0[d]	141.5[d]	39.6[d]	6.16[d]	47.7[d]	133.4[d]	40.1[d]
GCGACCG3′/CGCAGGC5′	6.71 ± 0.04	57.5 ± 2.8	163.9 ± 9.0	38.0	6.74 ± 0.08	57.5 ± 4.1	163.5 ± 13.0	38.1
(GGCAGCC3′)$_2$	7.84 ± 0.07	62.2 ± 2.1	175.3 ± 6.6	48.2	7.84 ± 0.17	61.0 ± 3.3	171.3 ± 10.0	48.4
UGAGAGUCA3′/ACUCACAGU5′	(7.94)[d]	(67.2)[d]	(190.9)[d]	43.7[d]	(7.74)[d]	(53.8)[d]	(148.4)[d]	(44.1)[d]
AA mismatches one or two base pairs from helix end								
GACCGCG3′/CAGGCGC5′	8.76 ± 0.04	48.6 ± 1.8	128.4 ± 5.6	51.9	8.79 ± 0.11	48.2 ± 3.5	127.0 ± 10.9	52.2
CGACCGC3′/GCAGGCG5′	7.13 ± 0.01	55.9 ± 1.4	157.2 ± 4.7	40.4	7.14 ± 0.07	51.8 ± 2.0	144.1 ± 6.3	40.7
GG mismatches								
GGCUGAG3′/CGGACUC5′	6.88 ± 0.02	49.7 ± 1.5	138.1 ± 4.9	39.2	6.92 ± 0.13	53.9 ± 3.2	151.4 ± 10.1	39.3
CGGCAUG3′/GCGGUAC5′	6.07 ± 0.02	58.8 ± 1.9	170.0 ± 5.9	34.6	6.10 ± 0.09	58.7 ± 2.6	169.5 ± 8.2	34.7
GUGGCAG3′/CACul;1GGUC5′	6.66 ± 0.01	61.4 ± 1.2	176.5 ± 3.9	37.7	6.72 ± 0.04	62.1 ± 4.2	178.5 ± 13.5	37.9
(GCGGCGC3′)$_2$	9.30 ± 0.06	69.2 ± 1.4	193.2 ± 4.4	54.1	9.36 ± 0.17	69.9 ± 3.3	195.1 ± 10.0	54.3
(CGCGGCG3′)$_2$	8.16 ± 0.07	65.8 ± 2.8	185.9 ± 9.0	49.2	7.97 ± 0.10	57.8 ± 1.3	160.7 ± 4.0	49.8
reference duplexes								
GCGGCG3′/CGCCGC5′	10.92 ± 0.02	61.5 ± 0.3	163.1 ± 0.9	60.8	11.01 ± 0.19	63.3 ± 3.2	168.5 ± 9.6	60.6
	10.4[e]	58.5[e]	155[e]	58.7[e]	10.73[e]	60.82[e]	161.4[e]	60.2[e]
(GCGCGC)$_2$	10.83 ± 0.22	68.2 ± 3.1	185.0 ± 9.2	62.4	10.36 ± 0.18	61.2 ± 2.9	163.8 ± 8.9	62.7
GAGGAG3′/CUCCUC5′	8.50[f]	55.7[f]	152.2[f]	50.9[f]	8.66[f]	58.8[f]	161.8[f]	48.6[f]

[a] Measured in 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA (pH 7.0). Values in parentheses indicate that the $\Delta H°$ values determined from T_M^{-1} vs ln(C_T/4) plots and from curve fitting differ by more than 15%, thus indicating non-two-state melting. The errors are based on the standard deviations of the thermodynamic parameters and were calculated as described previously (*2*, *4*). When all sources of error for T_M^{-1} vs ln(C_T/4) parameters are considered, the typical error estimate for $\Delta G°_{37}$ is 4%, for $\Delta H°$ is 12%, and for $\Delta S°$ is 13.5%. Significant figures are given beyond error estimates to allow accurate calculation of T_M and other parameters. The duplex 5′GCGGCG3′/3′CGCCGC5′ was synthesized, and thermodynamic parameters were measured and compared with those reported in ref *48*. The values agree within the expected error limits.[b] T_M for a total oligonucleotide strand concentration of 10^{-4} M. [c] From ref *2*. [d] From ref *1*. [e] From ref *48*. [f] From ref *4*.

short, however, so the $\Delta G°_{37}$ values are expected to be reasonably reliable (*21*). The range of loop free energies is only 0.5 kcal/mol when UU has two adjacent AU base pairs, roughly a third of the range with adjacent GC base pairs. Evidently, for a single UU mismatch, the orientation of closing GC base pairs has more effect on the stability of the duplex than the orientation of adjacent AU base pairs. This may result from the higher dipole moment and rigidity of GC pairs. The higher dipole moment implies a stronger electrostatic interaction with the mismatch, and the rigidity precludes optimization of electrostatics by a change in conformation. On average, a UU mismatch adjacent to two AU base pairs is 1.6 kcal/mol less stable than a UU mismatch adjacent to two GC pairs when each mismatch is three base pairs from a helix end. The reduced stability observed with two adjacent AU pairs is similar to that observed with tandem mismatches (*3*) and with single CC mismatches (*5*).

With GG mismatches three base pairs from a helix end, the motif $\begin{smallmatrix}5'GGC3'\\3'CGG5'\end{smallmatrix}$ is on average 0.8 kcal/mol more stable

Table 2: Thermodynamic Parameters of Loop Formation in 1 M NaCl[a]

	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)
5′GAGMGAG3′/3′CUCNCUC5′ series			
GAGUGAG3′/CUCGCUC5′	−2.54 ± 0.10	−21.9 ± 2.8	−62.3 ± 8.9
GAGCGAG3′/CUCACUC5′	−0.35 ± 0.09	−13.4 ± 3.0	−41.6 ± 9.3
GAGAGAG3′/CUCGCUC5′	−0.10 ± 0.10	−11.4 ± 3.0	−36.5 ± 9.6
GAGCGAG3′/CUCUCUC5′	−0.03 ± 0.13	−10.4 ± 4.4	−33.6 ± 14.3
GAGUGAG3′/CUCUCUC5′	0.03 ± 0.09	−12.6 ± 2.5	−40.6 ± 8.0
GAGAGAG3′/CUCCCUC5′	0.31 ± 0.11	1.4 ± 3.0	3.6 ± 9.6
GAGAGAG3′/CUCACUC5′	(0.41)	(−12.3)	(−40.9)
GAGCGAG3′/CUCCCUC5′	0.52 ± 0.12	3.0 ± 3.6	8.1 ± 11.4
GAGUGAG3′/CUCCCUC5′	0.55 ± 0.15	−7.7 ± 3.7	−26.4 ± 12.3
UU mismatches with different surrounding base pairs			
GCGUCCG3′/CGCUGGC5′	−0.75 ± 0.24	−19.6 ± 3.7	−60.9 ± 10.4
(GCGUCGC3′)$_2$	−0.48 ± 0.24	−13.5 ± 3.7	−41.8 ± 11.2
GCGUGCG3′/CGCUCGC5′	−0.16 ± 0.09	−13.0 ± 2.1	−41.5 ± 6.6
GCCUGCG3′/CGGUCGC5′	0.39 ± 0.24	−12.3 ± 3.7	−41.0 ± 10.4
(CGCUGCG3′)$_2$	0.45	−12.9	−43.1
GCGUUCG3′/CGCUAGC5′	(0.49)	(−16.2)	(−54.0)
GUGUUCG3′/CACUAGC5′	1.12 ± 0.12	−4.5 ± 3.5	−18.2 ± 11.5
GCAUUCG3′/CGUUAGC5′	(1.22)	(−6.8)	(−25.7)
GCAUACG3′/CGUUUGC5′	(1.50)	(−4.3)	(−18.9)
GCUUACG3′/CGAUUGC5′	1.73 ± 0.16	−1.9 ± 4.5	−11.7 ± 14.3
UU mismatches one or two base pairs from helix end			
GUCCGCG3′/CUGGCGC5′	−1.79 ± 0.24	−12.1 ± 3.5	−33.3 ± 9.6
CGUCCGG3′/GCUGGCC5′	−1.30 ± 0.12	−18.0 ± 2.8	−53.3 ± 8.5
CGUCGUCG3′/GCUGCUGC5′	−1.33 ± 0.11	−15.9 ± 1.9	−46.9 ± 5.6
GCUGCUGC3′/CGUCGUCG5′	0.51 ± 0.13	2.7 ± 2.3	6.8 ± 7.0
AA mismatches with different surrounding base pairs			
(CGCAGCG3′)$_2$	0.67	−6.1	−21.8
GCGACCG3′/CGCAGGC5′	0.78 ± 0.11	−14.5 ± 2.5	−49.5 ± 8.1
(GGCAGCC3′)$_2$	1.13 ± 0.25	−5.0 ± 4.3	−20.0 ± 12.3
UGAGAGUCA3′/ACUCACAGU5′	(1.14)	(−4.5)	(−18.0)
AA mismatches one or two base pairs from helix end			
GACCGCG3′/CAGGCGC5′	−1.30 ± 0.24	−3.8 ± 3.8	−7.9 ± 10.4
CGACCGC3′/GCAGGCG5′	0.37 ± 0.08	−9.3 ± 2.1	−31.0 ± 6.8
GG mismatches			
GGCUGAG3′/CGGACUC5′	−2.58 ± 0.08	−8.7 ± 2.3	−19.8 ± 7.4
CGGCAUG3′/GCGGUAC5′	−2.49 ± 0.09	−25.1 ± 3.0	−72.9 ± 9.4
GUGGCAG3′/CACGGUC5′	−2.41 ± 0.08	−20.4 ± 2.1	−57.8 ± 6.8
(GCGGCGC3′)$_2$	−1.89 ± 0.24	−15.9 ± 3.8	−45.1 ± 11.1
(CGCGGCG3′)$_2$	−1.40 ± 0.22	−21.9 ± 4.2	−66.2 ± 12.6

[a] Calculations are values from T_M^{-1} vs ln(C_T/4) plots. Values in parentheses are derived from non-two-state transitions.

Table 3: Frequencies of Base Pairs Adjacent to AA, GG, and UU Mismatches in Large and Small Subunit rRNA and Group I Intron Secondary Structures[a]

	AA mismatch						GG mismatch						UU mismatch					
W→ / Z	C G	G C	U A	A U	U G	G U	C G	G C	U A	A U	U G	G U	C G	G C	U A	A U	U G	G U
X / V ↓																		
G C	14	9	10	25	3	9	12	19	7	18	2	8	175	31	57	21	17	36
C G		4	2	13	6	7		4	48	26	1	6		11	33	25	10	28
A U			17	13	3	4			6	18	3	1			28	59	7	24
U A				33	2	1				5	6	3				7	4	20
G U					1	6					1	0					1	1
U G						1						2						3

[a] The table gives the number of occurrences of $^{5'XNW3'}_{3'VNZ5'}$, where $^{N}_{N}$ represents the mismatch.

than $^{5'CGG3'}_{3'GGC5'}$, similar to the sequence dependence with UU mismatches. With AA mismatches, however, the $^{5'GAC3'}_{3'CAG5'}$ and $^{5'CAG3'}_{3'GAC5'}$ motifs appear to have equal stability within experimental error.

Effect of Position of the Single Mismatch within the Duplex. An analysis of a database of rRNA secondary structures (*19*, *20*) suggests that some single mismatches are preferentially located close to helix ends. For example, a UU mismatch is separated by only one base pair from a helix end in 357 of the 598 times that it appears (60%), whereas if helix lengths are chosen randomly according to their distribution in this same database, chance would place only 130 of the single UU mismatches one pair from the end (22%).

Thermodynamic parameters of heptamers containing a single UU mismatch between one and three Watson−Crick pairs from the end of the helix are listed in Table 1. Analysis of the parameters shows that moving a UU mismatch toward the end of a helix makes the duplex more stable. The loop free energies for the motif $^{5'GUC3'}_{3'CUG5'}$ are −0.6, −1.3, and −1.8 kcal/mol for the single UU mismatch three, two, and one base pair(s) from the helix end, respectively (Table 2). Each base pair closer to the helix end stabilizes the duplex by ∼0.5 kcal/mol.

Shifting the AA mismatch in a $^{5'GAC3'}_{3'CAG5'}$ motif toward the helix end also has an effect (Table 2). The loop free energies for an AA mismatch placed three, two, and one base pair(s) from the helix end are 0.8, 0.4, and −1.3 kcal/mol, respectively. In contrast, the free energy increments for the $^{5'GGC3'}_{3'CGG5'}$ motif are about −2.3 kcal/mol, essentially independent of the position in the helix (Table 2).

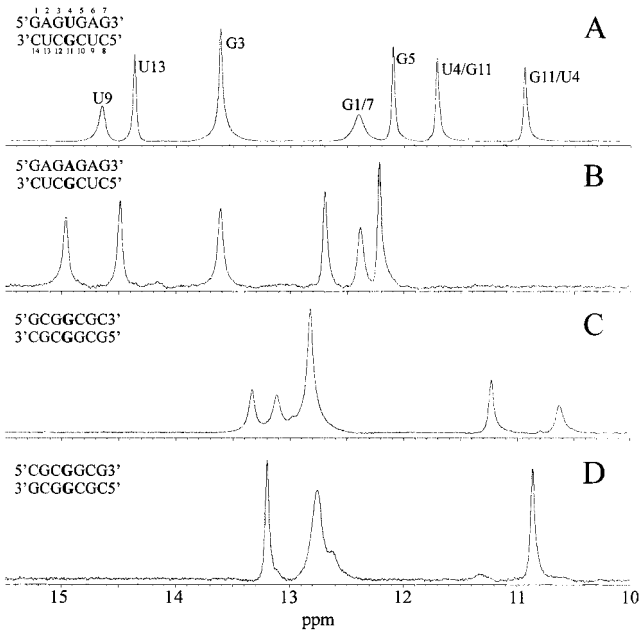

FIGURE 1: Imino proton region of the one-dimensional NMR spectra for single mismatch duplexes at 10 °C. From top to bottom, the mismatches are UG, AG, GG, and GG. Samples were dissolved in 9:1 v/v H_2O/D_2O with 80 mM NaCl, 0.5 mM EDTA, and 10 mM phosphate buffer at pH 7.

Effect of Two Single UU Mismatches in a Duplex. The nearest-neighbor model for approximating duplex stability assumes there is no coupling between single mismatches separated by at least two Watson−Crick pairs. This was tested for UU pairs by measuring the thermodynamics of duplex formation for $(CG\underline{U}CG\underline{U}CG)_2$ and $(GC\underline{U}GC\underline{U}GC)_2$ (Tables 1 and 2). For these cases, the free energy increment for each mismatch was calculated in a manner analogous to

$$\Delta G^\circ_{37-\text{loop}} = {}^1/_2[\Delta G^\circ_{37}(CG\underline{U}CG\underline{U}CG)_2 - \Delta G^\circ_{37}(CGCGCG)_2 + 2\Delta G^\circ_{37}(\begin{smallmatrix}5'GC3'\\3'CG5'\end{smallmatrix})] \quad (3)$$

For the motif $\begin{smallmatrix}5'GUC3'\\3'CUG5'\end{smallmatrix}$, $\Delta G^\circ_{37} = -1.3$ kcal/mol, the same as the value for a single $\begin{smallmatrix}5'GUC3'\\3'CUG5'\end{smallmatrix}$ motif with the UU mismatch placed two base pairs from a helix end (Table 2). For the motif $\begin{smallmatrix}5'CUG3'\\3'GUC5'\end{smallmatrix}$, $\Delta G^\circ_{37} = 0.5$ kcal/mol, similar to the average value of 0.4 kcal/mol measured for a single $\begin{smallmatrix}5'CUG3'\\3'GUC5'\end{smallmatrix}$ motif with the UU mismatch placed three base pairs from a helix end, but somewhat less stable than might be expected for $\begin{smallmatrix}5'CUG3'\\3'GUC5'\end{smallmatrix}$ with the UU mismatch placed two base pairs from a helix end. The results suggest the nearest-neighbor approximation is reasonable for these sequences.

NMR Spectra. Figure 1 shows imino proton spectra for four duplexes with single mismatches. The top two spectra are for mismatches in the context 5′GAGMGAG3′/3′CUCNCUC5′. Two imino peaks are observed between 14 and 15 ppm, as expected for the two AU pairs in each duplex (*4*, *22*−*25*). Several imino peaks are also observed between 12 and 14 ppm, as expected for the GC pairs in each duplex. For the duplex with a UG mismatch (Figure 1A), integration suggests there are two protons producing the peak at 13.6 ppm; thus, a total of eight imino protons are observed. An NOE experiment (Figure 2) verifies that the imino protons between 14 and 15 ppm are each very close to an adenosine H2 (strong NOE around 7.5 ppm), and therefore represent

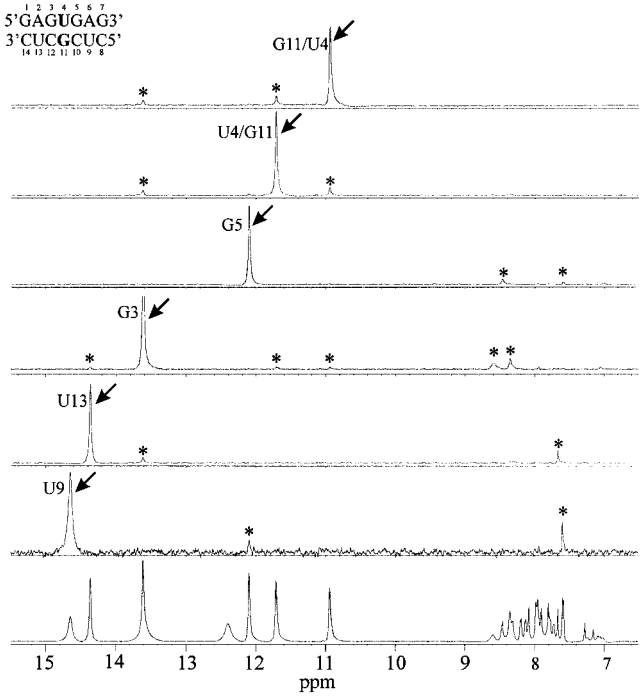

FIGURE 2: One-dimensional NOE difference spectra for the duplex. 5′GAGUGAG3′/3′CUCGCUC5′. Arrows point to the resonance that was saturated, and asterisks represent observable NOEs. The spectra were acquired under the conditions described in the legend of Figure 1.

protons involved in AU pairs. The peaks at 10.9 and 11.7 ppm are assigned to the GU pair on the basis of the NOE between them, an NOE to the same GC imino proton, and the absence of NOEs to adenosine H2 or C amino resonances between 7 and 9 ppm. This is typical of a GU wobble pair. Molecular modeling suggests that in A-form RNA, the G and U imino protons are significantly closer to the G3 imino proton than the G5 imino proton, so one resonance of the peak at 13.6 ppm probably represents G3, based on its NOE to the mismatched pair.

For the GA mismatch duplex (Figure 1B), there are four imino proton resonances below 14 ppm. The spectrum is similar to that of the Watson−Crick paired duplex (*4*), $\begin{smallmatrix}5'GAGGAG3'\\3'CUCCUC5'\end{smallmatrix}$, suggesting the imino proton in the GA mismatch is exchanging with water. It is possible, however, that one terminal base pair produces a broad peak so that one of the resonances may correspond to a hydrogen-bonded G imino proton in the mismatch.

For the self-complementary GG mismatch duplexes (spectra C and D of Figure 1), the NMR spectra show the presence of GC pairs above 12 ppm and mismatched G protons between 10 and 11 ppm. Although the Figure 1C duplex is self-complementary, the two distinct resonances below 12 ppm show that it is asymmetric. This asymmetry is not surprising since a hydrogen-bonded GG mismatch cannot be inserted in the center of a helix in a symmetrical manner (see the Discussion).

Watergate−NOESY spectra further elucidate the nature of the $GCGG\underline{C}GC$ duplex (Figure 3). Whereas cross-peaks between imino protons in adjacent pairs are often small or not apparent in spectra with mixing times of less than 100 ms (*26*), the cross-peaks between 10.6 and 11 ppm and between 13.2 and 13.4 ppm are apparent after a mixing time of only 2 ms (Figure 3A). The buildup of cross-peaks is

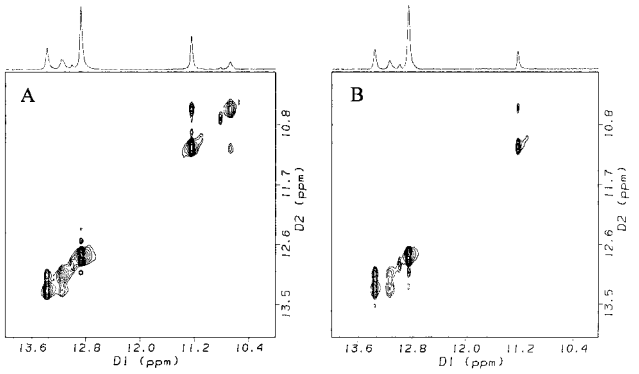

FIGURE 3: Imino proton region of the NOESY spectrum for $(5'GCGGCGC3')_2$ at 10 °C. A Watergate pulse sequence was used to suppress the water resonance (*51*). The one-dimensional cross section of the diagonal is shown above: (A) 2 ms mixing time and (B) 100 ms mixing time.

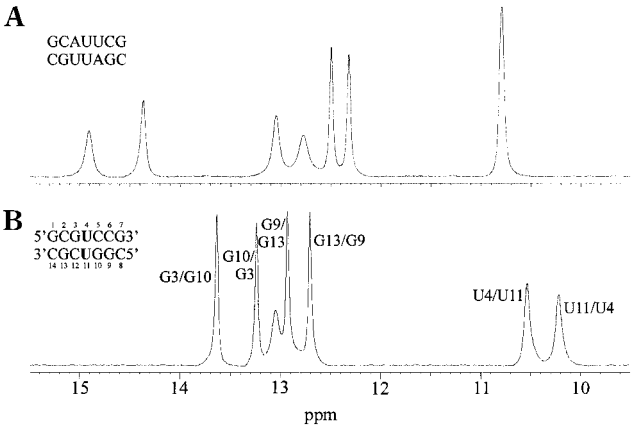

FIGURE 4: Imino proton region of the one-dimensional NMR spectra for UU mismatches with different closing base pairs. Spectra were acquired at 10 °C. The solution conditions were those described in the legend of Figure 1.

nonlinear for the 2, 10, 100, and 150 ms NOE mixing times; the cross-peaks arise at very short mixing times, and their magnitudes increase more slowly as mixing times lengthen. This is consistent with the buildup expected from conformational exchange, which can occur not only during the mixing time but also during the several milliseconds just prior to acquisition during the Watergate suppression pulses and during acquisition.

Spectra of nonexchangeable protons for $(GCGGCGC)_2$ at 30 °C exhibit sharp base and H1′ peaks for G1, C2, G6, and C7, but broad peaks for C5 (data not shown). Peaks for G3 do not appear, as is possible if the protons on G3 are in an intermediate exchange regime at this temperature.

Figure 4 shows imino proton spectra for two duplexes with a UU mismatch between adjacent Watson−Crick pairs. In the spectrum for 5′GCAUUCG3′/3′CGUUAGC5′ (Figure 4A), the peaks between 14 and 15 ppm represent U imino protons in AU pairs. Around 13 ppm, there is a broad resonance for the terminal GC pairs and there are sharper resonances for the internal GC pairs. Imino protons of both U's in the mismatch resonate just below 11 ppm. For the most stable context, $\begin{smallmatrix}5'GUC3'\\3'CUG5'\end{smallmatrix}$, there are two strong peaks between 10 and 11 ppm (Figure 4B). A one-dimensional NOE experiment (Figure 5) provides evidence that these can be attributed to imino protons of the mismatched U's. These

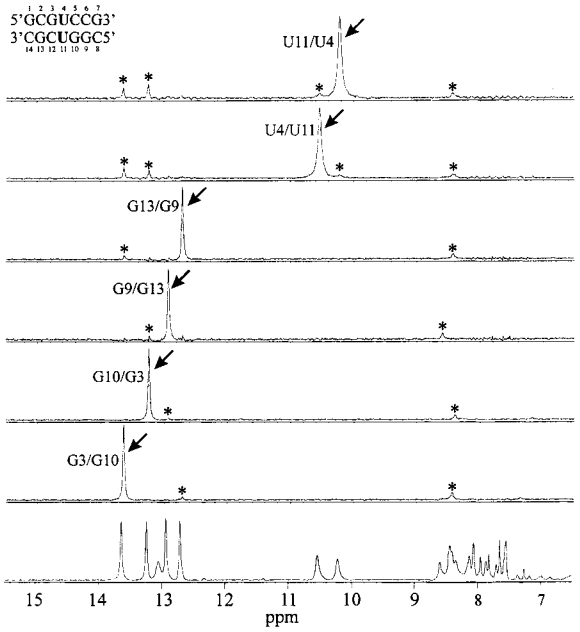

FIGURE 5: One-dimensional NOE difference spectra for the duplex 5′GCGUCCG3′/3′CGCUGGC5′. Arrows point to the resonance that was saturated, and asterisks represent observeable NOEs. The spectra were acquired under the conditions described in the legend of Figure 1.

Table 4: Frequencies of Single Mismatches in Phylogenetic RNA Secondary Structures

mismatch	M N	GMC CNG	CMG GNC	GMG CNC	CMC GNG
G U	6353	1026	462	724	404
A G	878	33	23	14	47
U U	598	175	11	31	
C A	564	26	51	29	35
C U	367	46	13	6	9
A A	207	14	4	9	
G G	196	12	4	19	
C C	90	6	6	12	

two peaks each have NOEs to the same two resonances at 13.2 and 13.6 ppm, which represent the adjacent GC pairs. NOEs between 8 and 9 ppm represent amino protons of cytosine near imino protons of guanines in the respective Watson−Crick pairs. The mismatched U's also have an NOE to a nearby C amino proton.

DISCUSSION

A database of RNA secondary structures (*18*−*20*) containing 101 small subunit rRNAs, 218 large subunit rRNAs, and 75 group I introns contains 9253 single mismatches. As shown in Table 4, some mismatches are present much more often than others. The most prevalent single mismatches are GU, AG, UU, and CA with 6353, 878, 598, and 564 occurrences, respectively. In some cases, the mismatches are conserved across species at particular locations. For example, a CA mismatch is common in group I introns, but not in rRNAs. This suggests that this pair is structurally significant in at least one situation.

The results collected in Table 2 show that single mismatches can stabilize or destabilize a duplex depending on the identity of the mismatch, its adjacent base pairs, and its

position in the helix. Measured free energy increments range from −2.6 to 1.7 kcal/mol. Thus, rules for stabilities of single mismatches are likely to be important for predicting RNA secondary structure and associations.

Effect of Adjacent Base Pairs. Gralla and Crothers (*5*) reported that for $(A_5\underline{C}U_5)_2$, the loop free energy at 25 °C is 1.8 kcal/mol, whereas for $(A_4G\underline{C}CU_4)_2$, it is 0.1 kcal/mol. The results listed in Table 2 show that the average difference in free energy between single UU mismatches adjacent to two GC and two AU base pairs is 1.6 kcal/mol. This free energy difference can be largely explained by the free energy in the Watson−Crick base pairs; an AU pair is less favorable than a GC pair, presumably due primarily to the difference in the number of hydrogen bonds (*11*). Our results show that the orientation of adjacent base pairs can also affect loop free energy. For example, the free energy increment of the $^{5'GUC3'}_{3'CUG5'}$ motif is on average 1 kcal/mol more favorable than that of $^{5'CUG3'}_{3'GUC5'}$ when both are three base pairs from a helix end. The prevalence of these pairs adjacent to UU in a database of RNA secondary structures correlates with this result; $^{5'GUC3'}_{3'CUG5'}$ occurs 175 times, while $^{5'CUG3'}_{3'GUC5'}$ appears only 11 times (Tables 3 and 4). The stability increment for $^{5'GUG3'}_{3'CUC5'}$ is approximately the average of those for $^{5'GUC3'}_{3'CUG5'}$ and $^{5'CUG3'}_{3'GUC5'}$.

Distance from the Helix End. Analysis of rRNA and group I intron secondary structures indicates some single mismatches occur preferentially near the ends of helixes. Therefore, the effect of the mismatch position on duplex stability was studied. UU, AA, and GG mismatches one, two, or three base pairs from the helix end and flanked with two adjacent GC pairs were studied. The results demonstrate that moving the position of the mismatch toward the end of the helix enhances the stability for UU and AA mismatches by ∼0.5 kcal/mol per each position closer to the helix end, except for AA only one base pair from an end (see below). This is a non-nearest-neighbor effect. In contrast, the stabilities of GG mismatches are relatively insensitive to the position within the helix.

It is not clear why loop free energy can become more favorable when the mismatch position is moved toward the helix end. For AA in 5′G$\underline{A}$CCGCG3′/3′C$\underline{A}$GGCGC5′, it is possible that the terminal GC base pair does not form since the predicted stability of 5′$\underline{A}$CCGCG3′/3′$\underline{A}$GGCGC5′ is −8.41 kcal/mol at 37 °C (*11*, *44*), close to the value of −8.76 kcal/mol measured for the sequence with a potential terminal GC pair. The predicted stability for 5′$\underline{U}$CCGCG3′/3′$\underline{U}$GGCGC5′ is 1.2 kcal/mol less favorable than the value measured for 5′G$\underline{U}$CCGCG3′/3′C$\underline{U}$GGCGC5′, however, so the terminal GC may form in 5′G$\underline{U}$CCGCG3′/3′C$\underline{U}$GGCGC5′. Papanicolaou et al. (*27*) observed that prediction of secondary structures of tRNA and 5S rRNA was improved when a mismatch near a helix end was favored. The authors posit that moving a mismatch toward the helix terminus imposes fewer constraints on the mismatch shape. Presumably, it is also less costly to relieve helix distortions caused by mismatches near helix ends. If GG fits well into a Watson−Crick helix, as suggested by its thermodynamic stability, then the mismatch is not expected to be more stable near the helix end.

Interaction of Two Single Mismatches. The effect of two single UU mismatches within a duplex was measured to test for possible non-nearest-neighbor interactions. The oligomer, $(GC\underline{U}GC\underline{U}GC)_2$ is a portion of the sequence of triplet CUG repeats; such repeats in DNA are implicated in the etiology of myotonic muscular dystrophy, and they might be transcribed to RNA (*28*−*30*). The loop free energy for each mismatch in this oligomer is 0.5 kcal/mol, similar to that observed for a single UU mismatch adjacent to the same base pairs and placed three base pairs from the helix terminus. For comparison, the $^{5'GUC3'}_{3'CUG5'}$ motifs in $(CG\underline{U}C\text{-}G\underline{U}CG)_2$ have loop free energies of −1.3 kcal/mol, similar to that measured for the same motif in a duplex with one UU mismatch two base pairs from the end. Evidently, helix distortions induced by one UU mismatch do not strongly affect the thermodynamics of another UU mismatch two GC base pairs away.

Structures of Single Mismatches. The NMR data show that imino protons from UG, GG, and UU mismatches often produce strong, sharp resonances; thus, they can be protected from rapid exchange with water. This observation usually indicates the presence of hydrogen bonds. Hydrogen bonding within UU mismatches is likely to distort the helix from the A form (*31*). The ability of RNA helixes to incorporate hydrogen-bonded non-Watson−Crick pairs provides evidence of the flexibility of the phosphodiester−sugar backbone. This idea is reinforced by the presence of sharp imino proton resonances that suggest that adjacent pairs have normal Watson−Crick hydrogen bonding. It is also consistent with the approximate additivity of thermodynamic effects from two UU mismatches in a simple duplex.

Although the tandem GA mismatch motif $^{5'CGAG3'}_{3'GAGC5'}$ appears to be more stable than the tandem GG mismatch motif $^{5'CGGG3'}_{3'GGGC5'}$ (*2*), as a single mismatch GG is more stable than GA. Tandem mismatches provide more flexibility for incorporation of noncanonical pairs. In a tandem context, GA's are in either the sheared or imino conformation (*32*−*34*), but a single mismatch structure shows a G(syn)-A^+(anti) pairing (*35*). Perhaps the more constricted single mismatch environment is not able to accommodate a more stable GA pairing. In contrast, the stable single GG apparently fits well into the single mismatch motif, but is less stable in tandem contexts, perhaps due to unfavorable stacking interactions of adjacent GG mismatches.

Imino proton NMR spectra for GG mismatches indicate that the mismatch is asymmetric. In fact, a single mismatch cannot occupy the center position of a symmetric duplex if the mismatched bases hydrogen bond together. Even the "symmetric" GG pairs require one G to be flipped by glycosidic bond rotation or by backbone reversal (*31*); this destroys symmetry. Since there is no preference for one G over the other in a particular position of the mismatch, the two G's will occupy either position with equal probability. If the rate of the G's interchanging positions is slow, then two resonances are observed in the NMR spectrum. Similarly, conformational heterogeneity of the single GG in the crystal structure of Shah and Brunger may have contributed to the observed disorder in the crystal (*36*).

Conformational exchange of single GG's has also been observed in DNA. Lane and Peck (*37*) propose a syn-anti GG pair where the G's swap positions at a rate of ∼14000 Hz at 303 K. SantaLucia and co-workers observed a similar asymmetry of GG mismatches at the center of self-

complementary DNA duplexes at 283 K, with strong exchange resonances in one-dimensional NOE experiments (*38*).

Comparisons with Previous Work. Zhu and Wartell (*9*) used temperature gradient gel electrophoresis (TGGE) to study single mismatches in an RNA duplex with 345 base pairs. Direct comparisons with this work are difficult because (1) free energy differences were derived from differences in melting temperatures under the assumption that $\Delta S°$ is identical for each base pair, (2) the gel solution contained 0.5× TBE (0.045 M Tris, 0.045 M boric acid, and 2 mM EDTA) at pH 8.3, 4.06 M urea, and 23.3% (v/v) formamide, (3) all mismatches but one were placed 11−35 base pairs away from a helix end, and (4) only one mismatch with two adjacent GC pairs was studied. Nevertheless, some trends are similar to those seen in Table 2. First, GG mismatches were the most stable noncanonical pair, although the difference between free energy increments for GG and other noncanonical pairs appears to be smaller than those shown in Table 2. Second, when GG is omitted, the range in free energy increments is −0.3 to 1.2 kcal/mol when converted to a scale equivalent to eq 2 and Table 2. Third, the mismatch and adjacent base pairs are identical to those in Table 2 for four sequences, and the free energy increments in Table 2 are 0.3−1.2 kcal/mol less favorable with an average difference of 0.7 kcal/mol. Thus, for mismatches other than GU and GG, both studies show a similar range of sequence dependence, although the absolute magnitudes differ.

Bevilacqua and Bevilacqua (*7*) used both TGGE and optical melting to study the relative stabilities of hairpins containing GU, GG, and GA mismatches with adjacent UA base pairs in 0.1 M NaCl (pH 7). The results in Table 2 are consistent with their results in that the order of stabilities is as follows: GU > GG > GA. At 37 °C, the GG and GA pairs are less stable than the GC pair in their context by 2.4 and 4.1 kcal/mol, respectively. On the basis of the results listed in Table 2, the nearest-neighbor parameters of Xia et al. (*11*), and the model presented below for single mismatches, differences of about 4.2 and 5.6 kcal/mol would be expected. Thus, GG and GA in certain contexts may be more stable than predicted from the model systems studied here.

Meroueh and Chow (*8*) used optical melting to measure the relative stabilities of hairpins containing GU, CA, GA, and UU mismatches with adjacent GC pairs in about 40 mM Na^+ (pH 5 and 7). At 37 °C and pH 7, the CA, AC, GA, and UU pairs in one context are less stable than the AU pair by 3.0, 3.5, 3.9, and 4.4 kcal/mol, respectively. On the basis of the results listed in Table 2, the nearest-neighbor parameters of Xia et al. (*11*), and the model presented below, differences of about 4.9, 4.9, 4.9, and 3.5 kcal/mol would be expected. Thus, three of the four mismatches are more stable than those predicted from the model systems studied here, whereas UU is less stable. The studies by Meroueh and Chow (*8*) and by Bevilacqua and Bevilacqua (*7*) used hairpins with melting temperatures that are higher than those of the duplexes reported here, with unpaired nucleotides at the base of the stem, and in Na^+ concentrations of ≤0.1 M. The results suggest these conditions may often give smaller differences between Watson−Crick base pair and mismatch stability than those observed with short duplexes melting near 37 °C in 1 M NaCl (Table 2). It is also possible that single mismatches distort the helix backbone so that stabilities are very context-dependent in a non-nearest-neighbor way.

Xiang et al. (*39*) studied the effect on binding to a group II ribozyme in 100 mM Mg^{2+} when AU and GC base pairs in the recognition sequence were replaced with mismatches. For mismatches flanked by canonical pairs, they report an average loss of binding free energy of 2.7 kcal/mol. The model described below for mismatch thermodynamics predicts an average loss of 4.2 kcal/mol. This comparison also suggests that mismatch effects on duplex formation by short oligonucleotides may be more unfavorable than those observed in more restricted environments.

Morse and Draper (*6*) used optical melting to study single AA, GA, and GG mismatches in duplexes with nine or ten base pairs in both 0.1 and 1 M NaCl (pH 7). Two of their sequences, $(CGCG\underline{A}CGCG)_2$ and $(CGCG\underline{G}CGCG)_2$, contain motifs represented in Table 2 and can be analyzed using eq 2. The free energy increments calculated for the two loop motifs are −0.1 and −3.0 kcal/mol, respectively, compared with values of 0.8 and −2.3 kcal/mol from the results shown in Table 2. Thus, the $^{5'GGC3'}_{3'CGG5'}$ motif is 3 kcal/mol more stable than $^{5'GAC3'}_{3'CAG5'}$ in both studies. Surprisingly, however, Morse and Draper (*6*) find that $^{5'CAG3'}_{3'GAC5'}$ is more stable than $^{5'CGG3'}_{3'GGC5'}$, which is the opposite of the trend in Table 2. It has been pointed out that the sequences used by Morse and Draper have possibilities for forming hairpins (*40*). Such competing structures may affect the measured thermodynamics.

SantaLucia and co-workers have comprehensively studied single mismatches in duplex DNA (*38*, *40*−*42*). They also found that GG is the most stable noncanonical pair and that GG pairing involves hydrogen bonding as based on imino proton NMR spectra. When GG and GT are omitted, the range of free energy increments for single mismatches with two adjacent GC pairs is about −0.7 to 1.6 kcal/mol (*38*). This is similar to the range of −0.8 to 1.1 kcal/mol observed for RNA mismatches located three base pairs from a helix end (Table 2).

Predicting RNA Secondary Structure. One major application of thermodynamic parameters for RNA motifs is prediction of secondary structure from sequence (*43*, *44*). Due to the lack of experimental data, the most popular algorithm for predicting secondary structure has used a value of 0.8 kcal/mol for all single mismatches other than GU (*45*). The results discussed above indicate there is considerably more sequence dependence. Moreover, the sequence dependence is not restricted to nearest-neighbor effects since the free energy increments for AA and UU mismatches become more favorable as the mismatch approaches the end of a helix (Table 2). Given the large number of possible motifs, it would be very time-consuming to measure all the various aspects of the sequence dependence, until it can be done with oligonucleotide array or other new technology. The results in Table 2, however, suggest some simple approximations that may suffice for current structure prediction algorithms. The approximations are based only on the results in Table 2 since direct comparisons with literature results for RNA are difficult, as discussed above. The approximations also neglect potential sequence-dependent structure in the unpaired single strands. Such structure is difficult to quantify and likely contributes minimally to the observed sequence

dependence for these short oligonucleotides. For example, stacking in dinucleoside monophosphates is not strongly sequence-dependent (*49*) at 37 °C, the average melting temperature of the duplexes listed in Table 1. Moreover, the melting temperature for A_7 is about 35 °C (*50*), so less than half of the nucleotides are structured above 35 °C, even for a sequence with relatively high stacking propensity.

GU and GG mismatches are clearly more stable than others. Due to the prevalence of GU mismatches, nearest-neighbor parameters have been determined for them (*46*, *47*). The average free energy increment for GG mismatches with two adjacent GC pairs is −2.2 kcal/mol, and this appears to be a reasonable approximation irrespective of the position in the helix. The motif $\begin{smallmatrix}5'GUC3'\\3'CUG3'\end{smallmatrix}$ also appears to be unusually stable with an average free energy increment of −0.6 kcal/mol when located three base pairs from a helix end. This motif becomes more stable by about 0.5 kcal/mol for each base pair it moves closer to the helix end.

Aside from the above exceptions, the range in average free energy increments for the mismatches with two adjacent GC pairs and located three base pairs from a helix end is −0.4 to 1.1 kcal/mol (Table 2). When $\begin{smallmatrix}5'GUC3'\\3'CUG3'\end{smallmatrix}$ is omitted, and with AA, AC, AG, CC, CU, and UU weighted equally, the average free energy increment for these mismatches with two adjacent GC pairs is 0.3 kcal/mol. This appears to be a reasonable approximation for this motif, but mismatches closer to a helix end may be more stable, as seen for UU and AA in Table 2.

For UU mismatches three base pairs from helix ends, the average difference in free energy between having two adjacent GC pairs and two adjacent AU pairs is 1.6 kcal/mol (Table 2). This probably results from the fact that GC pairs have three hydrogen bonds whereas AU pairs have only two. Thus, a reasonable approximation for mismatches with adjacent AU or GU pairs is to make the stability increment adopted for the mismatch with two adjacent GC pairs less favorable by 0.8 kcal/mol for each adjacent AU or GU pair. Presumably, these approximations will be superseded when the detailed sequence dependence is revealed by future experiments.

ACKNOWLEDGMENT

We thank David Mathews for calculations with the data of Zhu and Wartell (*9*) and Professor Thomas R. Krugh for helpful discussions about the NMR experiments.

REFERENCES

1. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) *Biochemistry 30*, 6428−6436.
2. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) *Biochemistry 30*, 8242−8251.
3. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
4. Xia, T., McDowell, J. A., and Turner, D. H. (1997) *Biochemistry 36*, 12486−12497.
5. Gralla, J., and Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301−319.
6. Morse, S. E., and Draper, D. E. (1995) *Nucleic Acids Res. 23*, 302−306.
7. Bevilacqua, J. M., and Bevilacqua, P. C. (1998) *Biochemistry 37*, 15877−15884.
8. Meroueh, M., and Chow, C. S. (1999) *Nucleic Acids Res. 27*, 1118−1125.
9. Zhu, J., and Wartell, R. M. (1997) *Biochemistry 36*, 15326−15335.
10. Wang, Y.-X., Huang, S., and Draper, D. E. (1996) *Nucleic Acids Res. 24*, 2666−2672.
11. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
12. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. I, p 589, CRC Press, Cleveland, OH.
13. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. I, p 597, CRC Press, Cleveland, OH.
14. Petersheim, M., and Turner, D. H. (1983) *Biochemistry 22*, 256−263.
15. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
16. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843−853.
17. Hore, P. J. (1983) *J. Magn. Reson. 55*, 283−300.
18. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
19. Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3502−3507.
20. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
21. Freier, S. M., Petersheim, M., Hickey, D. R., and Turner, D. H. (1984) *J. Biomol. Struct. Dyn. 1*, 1229−1242.
22. Hilbers, C. W. (1979) in *Biological Applications of Magnetic Resonance* (Shulman, R. G., Ed.) Academic Press, New York.
23. Pardi, A., Martin, F. H., and Tinoco, I., Jr. (1981) *Biochemistry 20*, 3986−3996.
24. Ulrich, E. L., John, E.-M. M., Gough, G. R., Brunden, M. J., Gilham, P. T., Westler, W. M., and Markley, J. L. (1983) *Biochemistry 22*, 4362−4365.
25. Chou, S.-H., Hare, D. R., Wemmer, D. E., and Reid, B. R. (1983) *Biochemistry 22*, 3037−3041.
26. Wu, M., and Tinoco, I., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11555−11560.
27. Papanicolaou, C., Gouy, M., and Ninio, J. (1984) *Nucleic Acids Res. 12*, 31−44.
28. Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995) *J. Cell Biol. 128*, 995−1002.
29. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. E. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 7388−7393.
30. Hamshere, M. G., Newman, E. E., Alwazzan, M., Athwal, B. S., and Brook, J. D. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 7394−7399.
31. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., Appendix 1, pp 675−680, Cold Spring Harbor Laboratory Press, Plainview, NY.
32. SantaLucia, J., Jr., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
33. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.
34. Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) *Biochemistry 36*, 4449−4460.
35. Pan, B., Mitra, S. N., and Sundaralingam, M. (1999) *Biochemistry 38*, 2826−2831.
36. Shah, S. A., and Brunger, A. T. (1999) *J. Mol. Biol. 285*, 1577−1588.
37. Lane, A. N., and Peck, B. (1995) *Eur. J. Biochem. 230*, 1073−1087.
38. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999) *Biochemistry 38*, 3468−3477.
39. Xiang, Q., Qin, P. Z., Michels, W. J., Freeland, K., and Pyle, A. M. (1998) *Biochemistry 37*, 3839−3849.
40. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry 37*, 2170−2179.
41. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry 37*, 9435−9444.

42. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry 37*, 2694−2701.
43. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) *Nat. New Biol. 246*, 40−41.
44. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
45. Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Müller, P., Mathews, D. H., and Zuker, M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
46. He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., and Turner, D. H. (1991) *Biochemistry 30*, 11124−11132.
47. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
48. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) *Biochemistry 29*, 278−285.
49. Davis, R. C., and Tinoco, I., Jr. (1968) *Biopolymers 6*, 223−242.
50. Dewey, T. G., and Turner, D. H. (1979) *Biochemistry 18*, 5757−5762.
51. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.

BI991186L