A Periodic Table of Symmetric Tandem Mismatches in RNA†

Ming Wu, Jeffrey A. McDowell, and Douglas H. Turner*

Department of Chemistry, University of Rochester, Rochester, New York 14627-0216

Received September 26, 1994; Revised Manuscript Received December 30, 1994®

ABSTRACT: The stabilities and structures of a series of RNA octamers containing symmetric tandem mismatches were studied by UV melting and imino proton NMR. The free energy increments for tandem mismatch formation are found to depend upon both mismatch sequence and adjacent base pairs. The observed sequence dependence of tandem mismatch stability is $\frac{UG}{GU} > \frac{GU}{UG} > \frac{GA}{AG} \geq \frac{AG}{GA} > \frac{UU}{UU} > \frac{CA}{AC} \geq \frac{CU}{UC} \sim \frac{UC}{CU} \sim \frac{CC}{CC} \sim \frac{AC}{CA} \sim \frac{AA}{AA}$, and the closing base pair dependence is $\frac{5'G}{3'C} > \frac{5'C}{3'G} > \frac{5'U}{3'A} \sim \frac{5'A}{3'U}$. These results differ from expectations based on models used in RNA folding algorithms and from the sequence dependence observed for folding of RNA hairpins. Imino proton NMR results indicate the sequence dependence is partially due to hydrogen bonding within mismatches.

Predicting RNA secondary structure from primary sequence is a crucial step toward three-dimensional structure prediction. The most successful method for RNA secondary structure prediction is free energy minimization (Tinoco et al., 1971; Zuker & Stiegler, 1981; Ninio, 1979; Nussinov & Tinoco, 1982; Jacobson et al., 1984; Turner et al., 1988; Jaeger et al., 1989; Walter et al., 1994a) which uses the free energy increments of various RNA motifs as measured in model systems. While free energy increments for Watson−Crick base pairs are relatively well established (Freier et al., 1986), free energy increments for motifs with mismatches are mostly unknown. One such motif is the tandem mismatch. Existing data (SantaLucia et al., 1991; He et al., 1991; Ebel et al., 1994; Walter et al., 1994b) indicate that the sequence dependence of tandem mismatch stability differs from previously proposed models (Gralla & Crothers, 1973a; Jaeger et al., 1989). Here, the dependence of symmetric tandem mismatch stability on mismatch and adjacent base pair is investigated with a series of 18 self-complementary RNA octamers. When these results are combined with previous data, a periodicity of stability as a function of mismatch and adjacent base pair is found. The $\Delta G°_{37}$ of tandem mismatches ranges from −4.8 kcal/mol (stabilizing the helix) to +3.0 kcal/mol (destabilizing the helix), depending upon sequence. The accuracy of RNA secondary structure prediction should be improved by incorporation of this sequence dependence.

MATERIAL AND METHODS

Oligonucleotide Synthesis and Purification. RNA oligomers were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer. RNA monomers were from Applied Biosystems, Inc., with the 2′-hydroxyl protected as the *tert*-butyldimethylsilyl ether and the 5′-hydroxyl protected as the dimethoxytrityl group. Upon completion of coupling on the synthesizer, the oligomer was removed from the solid support and deprotected by treatment with concentrated ammonia in ethanol (3:1, v/v). The 2′-silyl protection was removed by treatment with freshly made 1 M triethylammonium hydrogen fluoride (50 equiv) in pyridine at 55 °C for 48 h. The crude sample was then dried and partitioned between water and diethyl ether. Then the sample was dissolved in 10 mL of 5 mM ammonium acetate (pH 7) and passed through a Sep-pak C-18 column (Waters) to desalt. The sample was purified on an Si500F thin-layer chromatography plate (Baker) and eluted with 1-propanol−ammonia−water (55:35:10, v/v). The least mobile main band was visualized with ultraviolet light, cut out, and eluted with water. The sample was desalted again with a Sep-pak C-18 cartridge. The purity of the oligomers was checked by HPLC on a C-8 analytical column (Hamilton) and was greater than 95%.

UV Melting and Thermodynamic Parameters. Thermodynamic parameters were measured in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. Oligoribonucleotide single-strand concentrations (C_T) were calculated from high-temperature absorbances and single-strand extinction coefficients calculated as described previously (Borer, 1975; Richards, 1975). Absorbance versus temperature melting curves were measured at 280 nm with a heating rate of 1 °C/min on a Gilford 250 spectrometer. Thermodynamic parameters for duplex formation were derived by fitting the shape of each curve to the two-state model (Petersheim & Turner, 1983) and by plotting the reciprocal of the melting temperature, T_M^{-1}, versus log C_T (Borer et al., 1974):

$$T_M^{-1} = \frac{2.303R}{\Delta H°} \log C_T + \frac{\Delta S°}{\Delta H°} \quad (1)$$

Melting curves for GGUGGACC were biphasic as has been observed previously for UGCGGGCA (SantaLucia et al., 1991a). Thus GG tandem mismatches were not studied further.

NMR Spectroscopy. After purification, RNA oligomers were dialyzed against 0.1 mM Na_2EDTA and double-distilled water for 24 h each, dried down, and then dissolved in 0.1 M (unless otherwise stated) NaCl, 0.5 mM Na_2EDTA, and 10 mM sodium phosphates in 90% H_2O and 10% D_2O, pH 7. Imino proton NMR spectra were recorded by a Varian VXR 500S spectrometer at 500 MHz. A binomial $1\bar{3}3\bar{1}$ pulse sequence was used to suppress the solvent peak (Hore,

† This work was supported by NIH Grant GM 22939.
* Author to whom correspondence should be addressed.
® Abstract published in *Advance ACS Abstracts,* March 1, 1995.

0006-2960/95/0434-3204$09.00/0

1983). The frequency offset was set to maximize the signal to noise ratio at about 13 ppm with the first nodes at 21 and 5 ppm. Spectra were collected with 12K points over a sweep width of 12 kHz, multiplied by a 5.0-Hz line-broadening exponential function, and Fourier-transformed on a Sun 4/260 computer running Varian VNMR software. One-dimensional NOE (nuclear Overhauser enhancement) spectra were collected by irradiating for 2−3 s with low decoupler power to result in 50−90% saturation of desired resonances. To minimize spillover artifacts, the off-resonance decoupler frequency was offset by the difference between the on-resonance decoupler frequency and the resonance frequency for which spillover was a concern. On-resonance and off-resonance spectra were interleaved in blocks of 16 scans to reduce the effect of long-term instrumental instability.

RESULTS

Thermodynamic Parameters. Eighteen RNA octamers and two hexamers were melted, and the plots of T_M^{-1} versus log C_T are shown in Figure 1. Thermodynamic parameters for duplex formation derived from these plots and curve fitting are listed in Table 1. For most of the oligomers, the parameters derived from the two methods agree within 15%. This suggests that the two-state model is reasonable for these molecules. GCUAAAGC, GGUUCACC, GGACAUCC, GGAACUCC, and CUGCUCAG, however, have greater than 15% differences in thermodynamic parameters derived from the two methods. These oligomers have melting temperatures below 24 °C at 10^{-4} M strand concentration, which makes it difficult to fit lower base lines. Thus discrepancies between the two methods do not necessarily mean these molecules are non-two-state.

Thermodynamic parameters for internal loop formation are listed in Table 2 and were calculated following Gralla and Crothers (1973). For example, $\Delta G°_{37,\text{loop}}$ of the internal loop $^{CA}_{AC}$ in duplex $^{GGU\,CA\,ACC}_{CCA\,AC\,UGG}$ is calculated as

$$\Delta G°_{37,\text{loop}}\left(^{CA}_{AC}\right) = \Delta G°_{37}\left(^{GGU\,CA\,ACC}_{CCA\,AC\,UGG}\right) - \Delta G°_{37}\left(^{GGUACC}_{CCAUGG}\right) + \Delta G°_{37}\left(^{UA}_{AU}\right)$$

In this equation, $\Delta G°_{37}\left(^{GGU\,CA\,ACC}_{CCA\,AC\,UGG}\right)$ and $\Delta G°_{37}\left(^{GGUACC}_{CCAUGG}\right)$ are the free energy changes of duplex formation derived from T_M^{-1} versus log C_T plots. $\Delta G°_{37}\left(^{UA}_{AU}\right)$ is the free energy increment for the nearest-neighbor $\left(^{UA}_{AU}\right)$ interaction (Freier et al., 1986) that is interrupted by the internal loop $^{CA}_{AC}$. (Note that the top strand is written in a 5′ to 3′ direction throughout this paper.)

In Table 2, all sequences have destabilizing internal loops except GAGGUCUC and GCGUUCGC. Internal loop stability changes with both the loop sequence and the base pairs closing the loop. $\Delta G°_{37,\text{loop}}$ ranges from −4.2 to 3.0 kcal/mol.

Imino Proton NMR. Imino proton NMR spectra (9−15 ppm) of 16 oligoribonucleotides are shown in Figure 2. Spectra of duplexes are listed in order of base pairs adjacent to the tandem mismatch: $^{5'G}_{3'C}$, $^{5'U}_{3'A}$, $^{5'A}_{3'U}$, and $^{5'U}_{3'G}$. Spectra for seven sequences with $^{5'C}_{3'G}$ adjacent to the mismatches have been published previously (SantaLucia et al., 1991a). Fourteen of 16 oligomers in Figure 2 are self-complementary octamers that can form duplexes with a tandem mismatch in the center and three flanking Watson−Crick base pairs

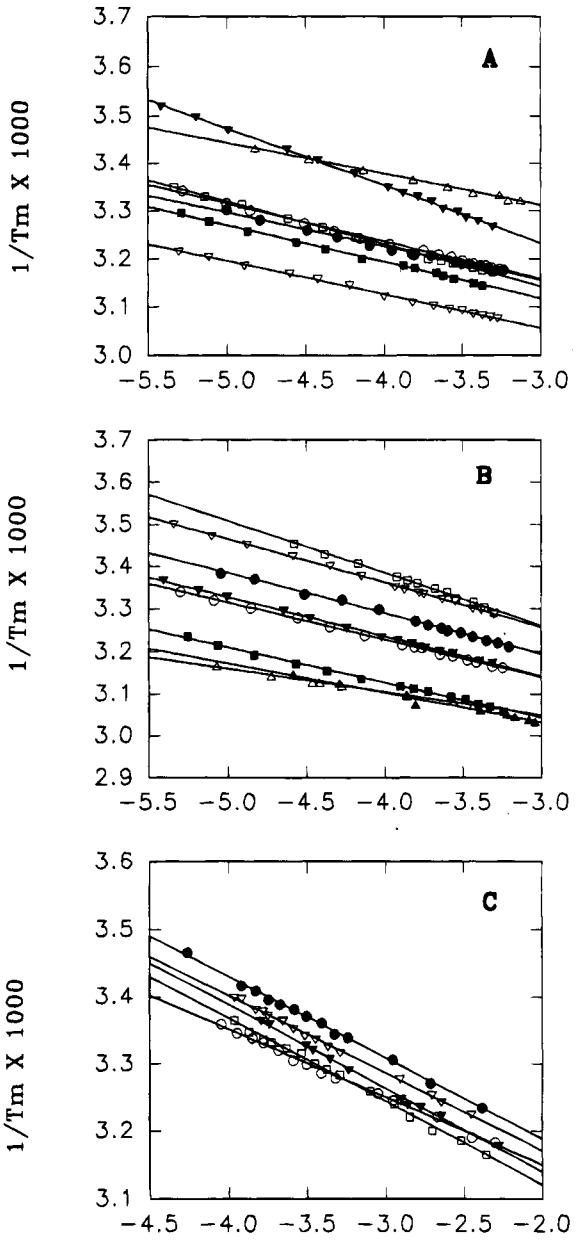

FIGURE 1: Reciprocal of melting temperature versus log concentration plots for (A) GGUCAACC(▼), CUGCUCAG(△), CGGCUCCG(□), GCUUUAGC(○), GCAUUUGC(●), GCGACCGC(■), GCGUUCGC(▽), for (B) GGAACUCC(□) GGUCUACC(▽), GCUAAGGC(●), CGGCACCG(▼), GCGAACGC(○), GGUACC(■), GAGCUC(▲), GAGGUCUC(△), and for (C) GGUUCACC(●), GCAAAUGC(▽), GCUAAAGC(▼), GGACAUCC(□), GCUUUGGC(○).

on each side. GCUUUGGC and GCUAAGGC form duplexes with UG mismatches flanking the center tandem mismatches. There is a center of symmetry in each duplex, so each resonance is attributed to two imino protons. Guanine and uracil have one imino proton each, while adenine and cytosine have none. Therefore each GC or AU pair has one hydrogen-bonded imino proton, and each UG mismatch has two. The assignments on the spectra are based on NMR melting and 1-D NOE experiments.

There are three classes of spectra in Figure 2. Most of the spectra have only the number of resonances expected from the base pairs in the helixes flanking the center tandem mismatches (see Figure 2b−e,g,h,m,n,p). For most of these cases, identical sequences in the flanking helixes give similar spectra, suggesting that any perturbation introduced by the

Table 1: Thermodynamic Parameters of Duplex Formation

RNA[a] duplex	$1/T_M$ vs ln C_T parameters				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)
GAG^{GU}CUC CUC$_{UG}$GAG	8.78 ± 0.07	82.8 ± 2.0	238.7 ± 6.3	49.1	8.61 ± 0.08	77.9 ± 1.7	223.3 ± 5.4	49.2
GCG^{UU}CGC CGC$_{UU}$GCG	7.69 ± 0.03	65.5 ± 0.9	186.5 ± 2.9	46.8	7.88 ± 0.17	70.1 ± 3.8	200.7 ± 11.9	47.1
GCA^{UU}UGC CGU$_{UU}$ACG	5.64 ± 0.02	65.1± 1.4	191.8 ± 4.4	36.8	5.67 ± 0.06	65.3 ± 3.4	192.2 ± 11.0	37.0
CGG^{CU}CCG GCC$_{UC}$GGC	5.58 ± 0.01	51.6 ± 0.6	148.4 ± 2.1	36.4	5.56 ± 0.07	52.8 ± 2.0	152.2 ± 6.5	36.3
GCG^{AC}CGC CGC$_{CA}$GCG	6.25 ± 0.01	59.7 ± 0.8	172.4 ± 2.4	40.0	6.33 ± 0.07	61.7 ± 4.6	178.7 ± 14.8	40.3
CGG^{CA}CCG GCC$_{AC}$GGC	5.53 ± 0.01	49.5 ± 0.4	141.9 ± 1.3	36.1	5.53 ± 0.06	54.7 ± 2.7	158.6 ± 8.4	36.2
CUG^{CU}CAG GAC$_{UC}$GUC	2.36 ± 0.18	69.9 ± 5.1	217.9 ± 17.2	23.0	3.18 ± 0.11	46.8 ± 2.5	140.7 ± 8.3	21.3
GCU^{UU}AGC CGA$_{UU}$UCG	5.47 ± 0.01	58.1 ± 0.8	169.6 ± 2.5	35.9	5.56 ± 0.08	64.2 ± 7.8	188.9 ± 25.3	36.4
GCG^{AA}CGC CGC$_{AA}$GCG	5.64 ± 0.01	52.4 ± 0.5	150.7 ± 1.5	36.8	5.75 ± 0.11	53.5 ± 6.6	153.9 ± 21.4	37.4
GCU^{AA}GGC CGG$_{AA}$UCG	4.71 ± 0.02	48.1 ± 0.9	139.9 ± 2.9	30.9	5.00 ± 0.05	42.9 ± 12.7	122.3 ± 42.4	32.2
GGU^{CA}ACC CCA$_{AC}$UGG	4.16 ± 0.02	38.1 ± 0.4	109.5 ± 1.2	25.2	4.03 ± 0.11	43.6 ± 4.8	127.6 ± 15.7	25.7
GGU^{CU}ACC CCA$_{UC}$UGG	3.83 ± 0.02	43.9 ± 0.4	129.1 ± 1.3	24.5	3.69 ± 0.18	49.3 ± 5.5	146.9 ± 18.4	25.0
GCU^{UU}GGC CGG$_{UU}$UCG	3.90 ± 0.02	45.3 ± 0.7	133.4 ± 2.4	25.3	4.06 ± 0.19	47.1 ± 12.2	138.8 ± 39.5	26.7
GGA^{CA}UCC CCU$_{AC}$AGG	4.03 ± 0.02	37.1 ± 0.6	106.7 ± 2.1	23.8	3.92 ± 0.34	55.4 ± 10.4	167.0 ± 32.6	27.4
GGA^{AC}UCC CCU$_{CA}$AGG	3.86 ± 0.04	36.8 ± 0.8	106.1 ± 2.7	22.4	3.63 ± 0.43	45.8 ± 10.6	136.0 ± 35.6	23.7
GGU^{UC}ACC CCA$_{CU}$UGG	3.27 ± 0.02	37.7 ± 0.5	111.2 ± 1.6	18.4	2.66 ± 0.21	55.0 ± 6.8	168.6 ± 21.6	20.9
GCA^{AA}UGC CGU$_{AA}$ACG	3.50 ± 0.02	39.5 ± 0.5	116.0 ± 1.6	20.8	3.36 ± 0.10	45.5 ± 4.1	135.7 ± 13.2	22.0
GCU^{AA}AGC CGA$_{AA}$UCG	3.81 ± 0.02	36.7 ± 0.7	106.0 ± 2.4	22.0	3.61 ± 0.24	53.7 ± 5.6	161.5 ± 17.5	25.5
GAGCUC CUCGAG	7.98 ± 0.16	62.3 ± 2.9	175.3 ± 8.8	48.9	7.75 ± 0.11	58.3 ± 2.0	163.0 ± 6.0	48.4
GGUACC CCAUGG	7.35 ± 0.03	54.9 ± 0.8	153.4 ± 2.6	46.8	7.48 ± 0.14	57.8 ± 2.9	162.3 ± 9.0	47.0

[a] Listed in order of increasing loop free energy (see Table 2). [b] Calculated for 10^{-4} M oligomer concentration.

tandem mismatches is not very dependent on mismatch sequence (compare Figure 2, spectra b and c, d and e, and m and n). The lack of a resonance from the UC mismatches (see Figure 2e,g,h) indicates that the imino proton of the U is exchanging rapidly with water and therefore probably not hydrogen bonded.

Spectra for GCUAAAGC and GCAAAUGC have one less resonance than expected for the base pairs flanking the mismatches (Figure 2j,l). This indicates that the imino proton in the AU pair in both sequences is able to exchange rapidly with water. Thus these AU pairs adjacent to AA mismatches either do not form Watson–Crick pairs or are in rapid equilibrium with a non-Watson–Crick conformation. This has also been observed for other internal loops closed by AU pairs (Varani et al., 1989).

The third class of spectra have two resonances more than expected for the base pairs in the helixes flanking the center tandem mismatches. Four of these spectra are for sequences with UU tandem mismatches (see Figure 2a,i,k,o). The 1-D NOE difference spectra for GCAUUUGC and GCUUUAGC are shown in Figures 3 and 4. Large NOEs are observed between the resonances assigned to the imino protons in the UU mismatches. This is also observed for GCGUUCGC (supplementary material; see the paragraph at the end of this paper regarding supplementary material). Large NOEs indicate that the imino protons are close in space, implying that the UU mismatches involve hydrogen bonding through the imino protons. Such hydrogen bonding has also been proposed for tandem UU mismatches in a $^{CUUG}_{GUUC}$ motif

Table 2: Thermodynamic Parameters of Loop Formation[a]

RNA duplex	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)
GAG^{GU}CUC CUC$_{UG}$GAG	-4.20 ± 0.17	-34.7 ± 3.5	-98.4 ± 10.8
GCG^{UU}CGC CGC$_{UU}$GCG	-0.49 ± 0.15	-13.7 ± 2.2	-42.9 ± 6.7
GCA^{UU}UGC CGU$_{UU}$ACG	0.84 ± 0.12	-8.5 ± 4.2	-30.1 ± 13.2
CGG^{CU}CCG GCC$_{UC}$GGC	0.92 ± 0.06	-11.7 ± 1.0	-40.7 ± 3.1
GCG^{AC}CGC CGC$_{CA}$GCG	0.95 ± 0.14	-7.9 ± 2.1	-28.8 ± 6.5
CGG^{CA}CCG GCC$_{AC}$GGC	0.97 ± 0.06	-9.6 ± 0.9	-34.2 ± 2.7
CUG^{CU}CAG GAC$_{UC}$GUC	1.30 ± 0.19	-29.4 ± 5.7	-98.9 ± 18.9
GCU^{UU}AGC CGA$_{UU}$UCG	1.35 ± 0.09	-7.1 ± 2.2	-27.1 ± 7.0
GCG^{AA}CGC CGC$_{AA}$GCG	1.56 ± 0.14	-0.6 ± 2.1	2.9 ± 6.2
GCU^{AA}GGC CGG$_{AA}$UCG	1.64 ± 0.05	-0.1 ± 1.9	-5.1 ± 6.0
GGU^{CA}ACC CCA$_{AC}$UGG	2.09 ± 0.04	8.7 ± 0.9	21.3 ± 2.9
GGU^{CU}ACC CCA$_{UC}$UGG	2.42 ± 0.04	2.9 ± 0.9	1.7 ± 2.9
GCU^{UU}GGC CGG$_{UU}$UCG	2.45 ± 0.05	2.7 ± 1.8	1.4 ± 5.5
GGA^{CA}UCC CCU$_{AC}$AGG	2.51 ± 0.09	10.9 ± 2.2	26.9 ± 6.8
GGA^{AC}UCC CCU$_{CA}$AGG	2.68 ± 0.10	11.2 ± 2.2	27.5 ± 7.0
GGU^{UC}ACC CCA$_{CU}$UGG	2.94 ± 0.04	9.1 ± 0.9	19.6 ± 3.1
GCA^{AA}UGC CGU$_{AA}$ACG	2.98 ± 0.11	17.1 ± 4.0	45.7 ± 12.6
GCU^{AA}AGC CGA$_{AA}$UCG	3.01 ± 0.09	14.3 ± 2.2	36.5 ± 6.9

[a] Calculated from T_M^{-1} vs log C_T parameters.

(SantaLucia et al., 1991a; Nikonowicz & Pardi, 1992, 1993).

The spectrum for GGUCAACC also has two resonances more than expected. Integration of the peaks suggests the resonance at 12 ppm may arise from two protons. Observation of resonances for five or six protons is surprising since neither C nor A normally has an imino proton. The protonated forms of C and A, however, do have imino protons, and protonation of A with a pK_a of 6.5 has been observed in an internal loop (Legault & Pardi, 1994). Preliminary melting experiments on GGUCAACC as a function of pH suggest a small change in $\Delta G°_{37}$ for duplex formation between pH 4.5 and 8.5. The T_M's, however, are too low to reliably quantify the effect. It appears likely that (GGUCAACC)$_2$ either has an unusual conformation or is in chemical exchange with a different conformation.

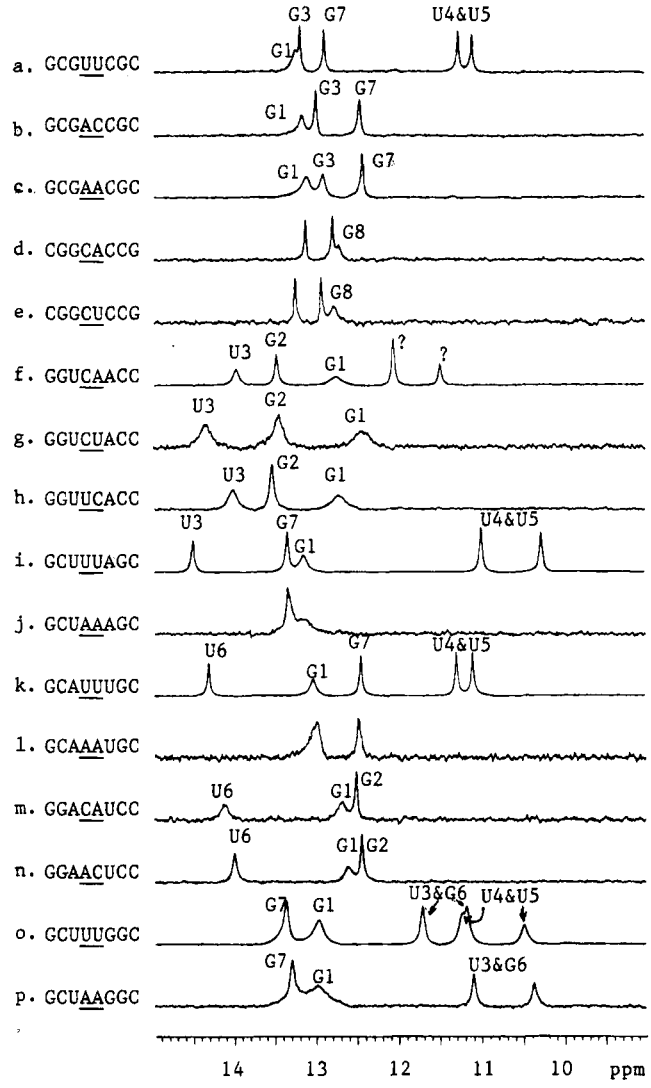

FIGURE 2: 500-MHz imino proton NMR spectra (9−15 ppm) in 0.1 M NaCl, 10 mM sodium phosphates, and 0.5 mM Na_2EDTA (unless otherwise stated) at pH 7 for (a) 0.4 mM rGCGUUCGC at 20 °C, (b) 0.4 mM rGCGACCGC at 10 °C, (c) 1.0 mM rGCGAACGC at 10 °C, (d) 0.3 mM rCGGCACCG at 10 °C, (e) 0.3 mM rCGGCUCCG at 10 °C, (f) 0.5 mM rGGUCAACC at 5 °C, (g) 0.5 mM rGGUCUACC in 1.0 M NaCl at 10 °C, (h) 0.5 mM rGGUUCACC at 5 °C, (i) 3.0 mM rGCUUUAGC at 15 °C, (j) 0.7 mM rGCUAAAGC at 0 °C, (k) 3.1 mM rGCAUUUGC at 20 °C, (l) 0.7 mM rGCAAAUGC at 10 °C, (m) 0.3 mM rGGACAUCC at 10 °C, (n) 0.5 mM rGGAACUCC at 10 °C, (o) 0.6 mM rGCUUUGGC in 1 M NaCl at 13 °C, and (p) 0.4 mM rGCUAAGGC in 1 M NaCl at 10 °C. Most sequences were measured at 0.1 M NaCl since this generally gives sharper resonances than at higher salt. Some sequences, however, had additional resonances at 0.1 M NaCl suggesting multiple conformations. These were remeasured at a higher salt concentration (1 M NaCl).

DISCUSSION

Previous work has shown that stabilities of tandem mismatches flanked by $^{5'C}_{3'G}$ depend on the mismatch sequence (SantaLucia et al., 1991a) and that stabilities of $^{GA}_{AG}$ tandem mismatches depend on the flanking Watson−Crick base pair (Walter et al., 1994b). The results reported here allow generalizations of these trends. Table 3 summarizes the free energy increments for tandem mismatches in a periodic table that reveals the major patterns. Loop stability tends to be less favorable as you go down a column or across a row. Thus, for a given tandem mismatch, stability depends

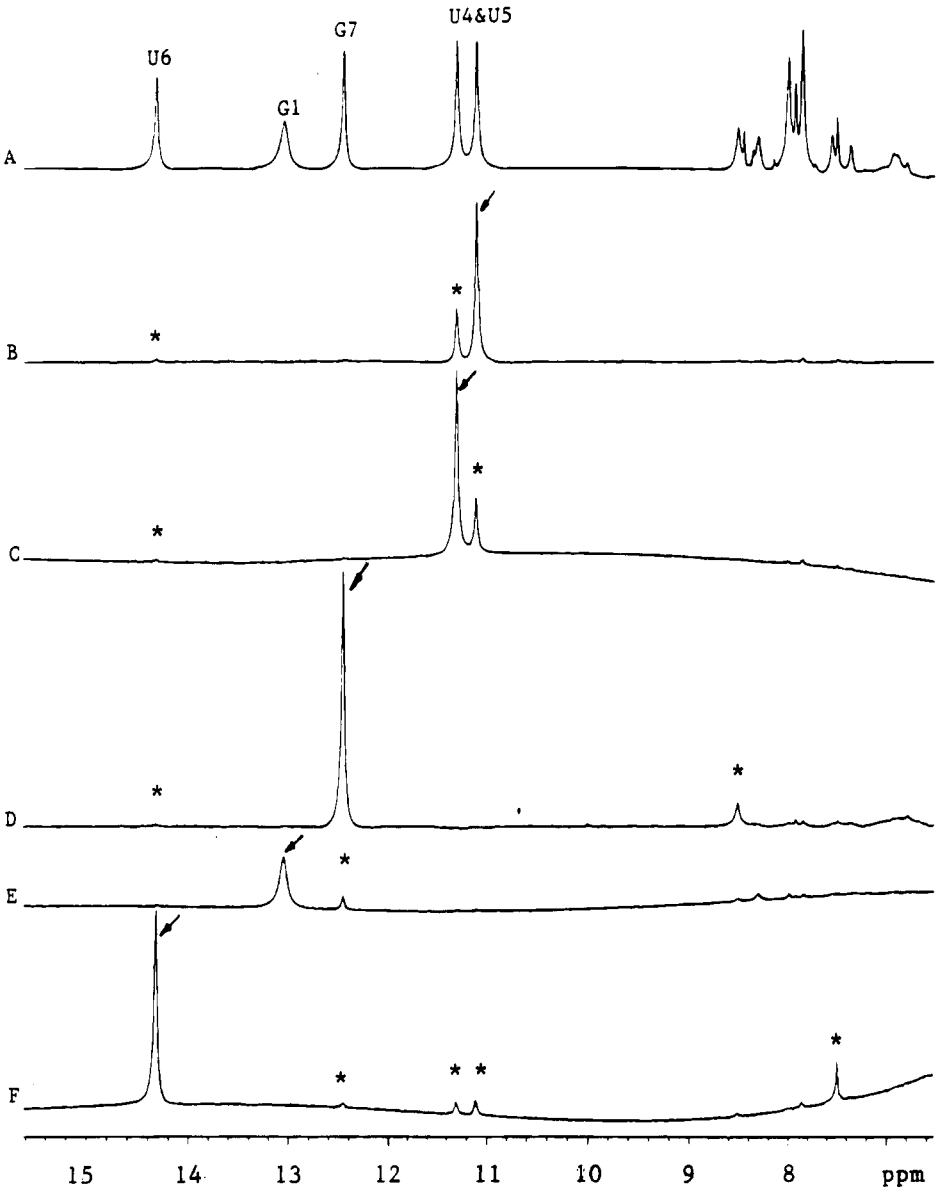

FIGURE 3: 500-MHz proton NOE difference spectra (6.5−15.6 ppm) of 3.1 mM rGCAUUUGC at 20 °C in 0.1 M NaCl, 10 mM sodium phosphates, and 0.5 mM Na_2EDTA in 90% H_2O and 10% D_2O, at pH 7: (A) off-resonance spectrum; (B−F) difference spectra between the off-resonance spectra and on-resonance spectra acquired with a 3-s saturation at 11.11, 11.30, 12.46, 13.05, and 14.32 ppm, respectively. The saturated resonances are indicated by arrows, while the observed NOEs are designated by asterisks. The assignments are shown in spectrum A. Note that in A-form RNA, G7H1 is closer in space to the $C2NH_2$ than to U6H3, consistent with the relative intensities in spectrum D. The large peak at 7.50 ppm in spectrum F is assigned to A3H2, which is expected to be close in space to U6H3 for a Watson−Crick base pair.

on flanking base pair in the order $\frac{5'G}{3'C} > \frac{5'C}{3'G} > \frac{5'U}{3'A} \sim \frac{5'A}{3'U}$. For a given flanking base pair, tandem mismatch stability essentially decreases in the order $\frac{UG}{GU} > \frac{GU}{UG} > \frac{GA}{AG} \geq \frac{AG}{GA} > \frac{UU}{UU} > \frac{CA}{AC} \geq \frac{CU}{UC} \sim \frac{UC}{CU} \sim \frac{CC}{CC} \sim \frac{AC}{CA} \sim \frac{AA}{AA}$. This order is not maintained, however, when the flanking base pair is GU. $(GCUAAGGC)_2$ is more stable than $(GCUUUGGC)_2$. For tandem mismatches measured in more than one oligomer sequence, all the individual values are listed in the appropriate cell of Table 3. The values are within the range expected for the nearest neighbor model (Kierzek et al., 1986; Freier et al., 1986) when the tandem mismatch and adjacent base pairs are considered as a unit.

The trends observed in Table 3 differ from those expected on the basis of models of internal loops currently used in algorithms for predicting RNA secondary structure (Jaeger et al., 1989; Walter et al., 1994a). For example, the model of Jaeger et al. (1989) assumes that the stability of an internal loop will depend on the stacking of the first mismatch with the adjacent base pair as measured at the ends of model duplexes (Turner et al., 1988). In this model, internal loops closed by $\frac{5'C}{3'G}$ are more stable than those closed by $\frac{5'G}{3'C}$, opposite the trend displayed in Table 3. The dependence of stability on first mismatch predicted by the model also differs from that displayed in Table 3. For example, $\frac{AA}{AA}$ is predicted to be more stable than $\frac{CA}{AC}$, whereas the opposite is observed. The recent model of Walter et al. (1994a) is more consistent with the dependence of stability on mismatch identity in Table 3, but does not include the enhanced stability associated with $\frac{5'G}{3'C}$ closure.

Some trends observed in Table 3 also differ from those observed for hairpins of six nucleotides (Serra et al., 1993, 1994). For these hairpins, stability depends on closing base pair in the order $\frac{5'G}{3'C} \sim \frac{5'C}{3'G} > \frac{5'U}{3'A} \sim \frac{5'A}{3'U}$. Thus $\frac{5'G}{3'C}$ closure does

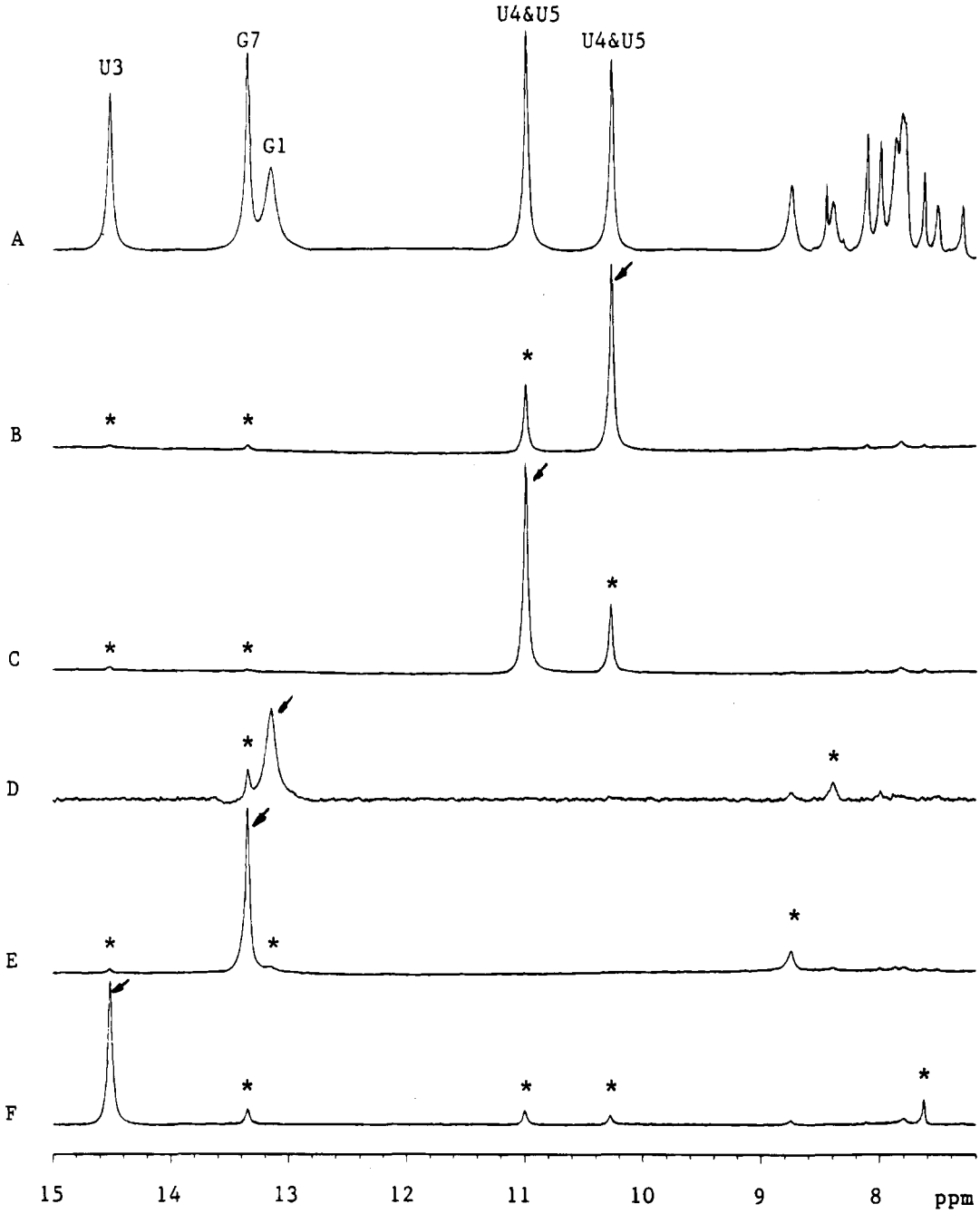

FIGURE 4: 500-MHz proton NOE difference spectra (7.2−15 ppm) of 3.0 mM rGCUUUAGC at 15 °C in 0.1 M NaCl, 10 mM sodium phosphates, and 0.5 mM Na_2EDTA in 90% H_2O and 10% D_2O at pH 7: (A) off-resonance spectrum; (B−F) difference spectra between the off-resonance spectra and on-resonance spectra acquired with a 3-s saturation at 10.29, 11.00, 13.15, 13.35, and 14.15 ppm, respectively. The saturated resonances are indicated by arrows, while the observed NOEs are designated by asterisks. The assignments are shown in spectrum A. Note that in A-form RNA, G7H1 is closer in space to the $C2NH_2$ than to U3H3, consistent with the relative intensities in spectrum E. The peak at 7.61 ppm in spectrum F is assigned to A6H2 which is expected to be close in space to U3H3 for a Watson−Crick base pair.

not enhance stability for hairpins as it does for tandem mismatches. Unlike tandem mismatches, however, hairpin stabilities trend with the stabilities of terminal mismatches on model duplexes. The exceptions are GA and UU mismatches which are extra stable in both hairpins and tandem mismatches. Evidently, the constraints imposed by the base-paired helixes flanking tandem mismatches lead to rules for stability that differ from those for hairpins or duplex termini. Presumably, larger internal loops will have fewer constraints. Thus it is possible that larger internal loops have rules for stability that are more like those for large hairpins than for tandem mismatches.

The periodicity of Table 3 allows predictions to be made for stabilities of tandem mismatches that have not been measured. These are listed in parentheses in Table 3. The predictions for tandem CC mismatches can be compared with a measurement made by Weeks and Crothers (1993) on a duplex with an $^{ACCA}_{UCCU}$ motif flanked on each side by nine additional base pairs. When calculated in the same way as the values in Table 3, their results give a free energy increment of 3.0 kcal/mol. This is within experimental error of the predicted values of 2.6 and 2.9 kcal/mol, respectively, for the motifs $^{ACCU}_{UCCA}$ and $^{UCCA}_{ACCU}$ (Table 3). In a similar way,

Table 3: A Periodic Table of Tandem Mismatches: Free Energy Changes of Tandem Mismatch Formation $\Delta G°_{37}(^{XY}_{YX})$ (kcal/mol) with Different Closing Base Pairs $^{5'P}_{3'Q}$ in RNA Octamers of Sequence $^{5'\text{--}pXYQ\text{--}3'}_{3'\text{--}QYXP\text{--}5'}$, Where * Represents a Nucleotide in a Watson–Crick Base Pair[a]

Mismatches ($^{XY}_{YX}$)→	UG GU	GU UG	GA AG	AG GA	UU UU	CA AC	CU UC	UC CU	CC CC	AC CA	AA AA
Closing base pairs ($^{P}_{Q}$)↓											
5'G 3'C	-4.8[c]	-4.7[c]	-2.8[e]	-1.2[e]	-0.5	1.0	0.8	*(1.0)*[b]	*(1.0)*[b]	1.0	1.6
		-4.2	-2.7[e]				0.9				
		-3.4[c]	-2.5[e]				1.3				
5'C 3'G	-3.9[c]	-1.0[c]	-0.4[d]	-0.6[d]	-0.1[d]	1.4[d]	1.7[d]	1.7[d]	2.0[d]	2.3[d]	1.7[d]
		-0.4[c]									
5'U 3'A	-2.7[c]	-0.1[c]	0.9[e]	*(0.9)*[b]	1.4	[2.1][f]	2.4	2.9	*(2.9)*[b]	*(2.9)*[b]	3.0
5'A 3'U	-1.9[c]	0.1[c]	0.1[e]	*(0.5)*[b]	0.8	2.5	*(2.6)*[b]	*(2.6)*[b]	*(2.6)*[b]	2.7	3.0
			0.9[e]								

[a] $\Delta G°_{37}$ values of the same mismatch with the same closing base pairs but different stem base pairs are listed together. Underlined sequences have NMR spectra suggesting hydrogen bonding of imino protons in the mismatches. [b] Data in italics and parentheses are predicted free energy changes. For predictions, it was assumed that $^{AG}_{GA} \sim {}^{GA}_{AG}$ or that $\Delta G°$ was the average of measured values in the nearest flanking cells of the same row. [c] He et al., 1991. [d] SantaLucia et al., 1991. [e] Walter et al., 1994. [f] This sequence has either an unusual conformation or a mixture of conformations.

Table 4: Frequency of Occurrence of Tandem Mismatches[a] in Phylogenetic Structures of 24 Small Subunit and 51 Large Subunit rRNAs[b]

Mismatches:	UG GU	GA AG	AA AG	GU UG	AA GG	UU UU	GA AA	AA AA	UG UA	CU CU	UG CA	UA UA	CA CG	GA AC	AC AA	CA CG	CC CC	GC AA	AU CU	UC UU
Frequency of Occurrence:	221	71	54	32	24	10	7	5	5	5	4	4	2	2	2	2	2	2	2	2
Percentage of F.O.(%):	45	15	11	6.6	4.9	2.0	1.4	1.0	1.0	1.0	0.8	0.8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4

[a] The top of each tandem mismatch is written in a 5′ to 3′ direction; e.g., $^{GA}_{AA}$ stands for $^{5'GA}_{3'AA}$. This table only shows the tandem mismatches that occur more than once. The total number of tandem mismatches is 487. The frequencies of occurrence for the flanking base pairs in this set are $^{5'C}_{3'G}$ (38%), $^{5'G}_{3'C}$ (23%), $^{5'U}_{3'A}$ (20%), $^{5'A}_{3'U}$ (9%), $^{5'U}_{3'G}$ (8%), and $^{5'G}_{3'U}$ (2%). [b] Gutell et al., 1992, 1993. $^{GG}_{UU}$ has been omitted.

the predicted value of 1.0 kcal/mol at 37 °C for the $^{GCCC}_{CCCG}$ motif is reasonably close to an estimate of 1.7 kcal/mol based on the results of Gralla and Crothers (1973a) for $(A_4\text{-}G\underline{CC}CU_4)_2$.

The enhanced stability associated with $^{5'G}_{3'C}$ relative to $^{5'C}_{3'G}$ closure of a tandem mismatch is similar to the enhanced stability associated with Watson–Crick and GU nearest neighbors that have a 5′G (He et al., 1991). For example $^{GC}_{CG}$ is 1.4 kcal/mol more stable than $^{CG}_{GC}$. This trend is not observed for hairpins (Serra et al., 1993, 1994) and is reversed for mismatches and 3′ unpaired nucleotides at the ends of duplexes (Turner et al., 1988; Serra et al., 1994). Evidently, enhanced stability for a $^{5'G}_{3'C}$ is observed when the sugar–phosphate backbone is constrained.

Hydrogen Bonding within Mismatches. In Table 3, mismatches to the left of $^{CA}_{AC}$ are generally significantly more stable than the other mismatches. The imino proton NMR spectra provide one rationale for this trend. All measured imino proton spectra for sequences to the left of $^{CA}_{AC}$ have resonances attributed to at least one imino proton in the mismatch (Figure 2; SantaLucia et al., 1991a; He et al., 1991; Walter et al., 1994b). Spectra for the other sequences, except GGU<u>CA</u>ACC, only have resonances attributed to the Watson–Crick base pairs. With the exceptions of $^{CGAG}_{GAGC}$, $^{UGAA}_{AAGU}$, and $^{AGAU}_{UAGA}$, the resonance positions and NOE difference spectra for imino protons in the stable mismatches indicate hydrogen bonding of the imino protons. For $^{CGAG}_{GAGC}$, functional group substitution studies (SantaLucia et al., 1991b) and a structure determination by 2-D NMR (SantaLucia & Turner, 1993) indicate hydrogen bonding of the G 2-amino and N3 with the A N7 and 6-amino, respectively [also see Ebel et al. (1994)]. On the basis of the resonance frequency for the G imino protons, $^{UGAA}_{AAGU}$ and $^{AGAU}_{UAGA}$ probably have the same hydrogen bonds in the mismatch. Evidently, hydrogen bonding within mismatches is one reason mismatches to the left of $^{CA}_{AC}$ are more stable than other mismatches.

Comparison with Prevalence in rRNA. Table 4 lists the frequency of occurrence of tandem mismatches in the phylogenetically determined secondary structures of 24 small

subunit and 51 large subunit rRNAs (Gutell et al., 1992, 1993). The symmetric tandem mismatches in order of prevalence are $\frac{UG}{GU} > \frac{GA}{AG} > \frac{GU}{UG} > \frac{UU}{UU} > \frac{AA}{AA} > \frac{CC}{CC}$. Thus the most common symmetric tandem mismatches are thermodynamically stable. As pointed out previously (SantaLucia et al., 1990), however, $\frac{AG}{GA}$ is never observed, even though it is both hydrogen bonded and stable. Also, the trend in frequency of occurrence for the flanking base pair is $\frac{5'C}{3'G} > \frac{5'G}{3'C} \sim \frac{5'U}{3'A} > \frac{5'A}{3'U} \sim \frac{5'U}{3'G} > \frac{5'G}{3'U}$ (Table 4). This differs from the trend observed for the thermodynamic stability. Presumably, prevalence is dependent on both stability and structure. The preference for hydrogen-bonded mismatches is also evident when all tandem mismatches are considered. Of the eight possible mismatches, three are hydrogen bonded (AG, GU, and UU) and four are not (AA, AC, CC, and CU), and GG has not been studied because it has unusual properties at high concentrations (SantaLucia et al., 1991a). Thus, excluding GG, if mismatches were randomly combined, there would be 43% (=3/7) of mismatches that are hydrogen bonded. In the rRNA data set, however, there are 87.1% (=841/966) hydrogen-bonded mismatches. Evidently, nature has a preference for hydrogen-bonded mismatches.

ACKNOWLEDGMENT

We thank Matthew A. Fountain for the help with NMR experiments.

SUPPLEMENTARY MATERIAL AVAILABLE

One figure showing NOE difference spectra for (GCG-UUCGC)$_2$ in H_2O (1 page). Ordering information is given on any current masthead page.

REFERENCES

Blommers, M. J. J., Walters, L. I., Haasnoot, C. A. G., Aelen, M. A., & van der Marel, G. A. (1989) *Biochemistry 28,* 7491−7498.
Borer, P. N., Dengler, B., & Tinoco, I., Jr. (1974) *J. Mol. Biol. 86,* 843−853.
Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. I, p 597, CRC Press, Cleveland, OH.
Cheng, J. W., Chou, S. H., & Reid, B. R. (1992) *J. Mol. Biol. 228,* 1037−1041.
Ebel, S., Brown, T., & Lane, A. N. (1994) *Eur. J. Biochem. 220,* 703−715.
Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., & Turner, D. H. (1986) *Biochemistry 25,* 3214−3219.
Gralla, J., & Crothers, D. M. (1973a) *J. Mol. Biol. 78,* 301−319.
Gralla, J., & Crothers, D. M. (1973b) *J. Mol. Biol. 73,* 497−511.
Gutell, R. R., Schnare, M. N., & Gray, M. W. (1992) *Nucleic Acids Res. 20 (Suppl.),* 2095−2109.
Gutell, R. R., Gray, M. W., & Schnare, M. N. (1993) *Nucleic Acids Res. 21,* 3055−3074.
Hare, D. R., & Reid, B. R. (1982) *Biochemistry 21,* 5129−5134.
He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., & Turner, D. H. (1990) *Biochemistry 30,* 11124−11132.
Hore, P. J. (1983) *J. Magn. Reson. 55,* 283−300.
Jacobson, A. B., Good, L., Simonetti, M., & Zuker, M. (1984) *Nucleic Acids Res. 12,* 45−52.
Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86,* 7706−7710.
Johnson, P. D., & Redfield, A. D. (1981) *Biochemistry 20,* 1147−1156.
Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., & Freier, S. M. (1986) *Biochemistry 25,* 7840−7846.
Legualt, P., & Pardi, A. (1994) *J. Am. Chem. Soc. 116,* 8390−8391.
Li, Y., Zon, G., & Wilson, W. D. (1991) *Biochemistry 30,* 7566−7572.
Nikonowicz, E. P., & Pardi, A. (1992) *Nature 355,* 184−186.
Nikonowicz, E. P., & Pardi, A. (1993) *J. Mol. Biol. 232,* 1141−1156.
Ninio, J. (1979) *Biochimie 61,* 1133−1150.
Nussinov, R., & Tinoco, I., Jr. (1982) *Nucleic Acids Res. 10,* 351−363.
Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Fed. Proc. 43,* 2663−2670.
Peritz, A. E., Kierzek, R., Sugimoto, N., & Turner, D. H. (1991) *Biochemistry 30,* 6428−6436.
Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. I, p 579, CRC Press, Cleveland, OH.
Saenger, W. (1984) *Principles of Nucleic Acid Structure,* Springer-Verlag, New York.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29,* 8813−8819.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991a) *Biochemistry 30,* 8242−8251.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991b) *J. Am. Chem. Soc. 113,* 4313−4322.
SantaLucia, J., Jr., & Turner, D. H. (1993) *Biochemistry 32,* 12612−12623.
Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A., & Turner, D. H. (1993) *Nucleic Acids Res. 21,* 3845−3849.
Serra, M. J., Axenson, T. J., & Turner, D. H. (1994) *Biochemistry* (in press).
Tinoco, I., Jr., Uhlenbeck. O. C., & Levine, M. D. *Nature 230,* 363−367.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17,* 167−192.
Varani, G., Wimberly, B., & Tinoco, I., Jr. (1989) *Biochemistry 28,* 7760−7772.
Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Müller, P., Mathews, D. H., & Zuker, M. (1994a) *Proc. Natl. Acad. Sci. U.S.A. 91,* 9218−9222.
Walter, A. E., Wu, M., & Turner, D. H. (1994b) *Biochemistry 33,* 11349−11354.
Weeks, K. M., & Crothers, D. M. (1993) *Science 261,* 1574−1577.
Zuker, M., & Stiegler, P. (1981) *Nucleic Acids Res. 9,* 133−148.

BI942272I