The Stability and Structure of Tandem GA Mismatches in RNA Depend on Closing Base Pairs†

Amy E. Walter, Ming Wu, and Douglas H. Turner*

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

Received April 28, 1994; Revised Manuscript Received July 18, 1994®

ABSTRACT: UV melting and imino proton NMR studies show that the stabilities and structures of tandem GA mismatches in RNA are dependent upon the closing base pairs around these mismatches. Internal loops of sequence $^{5'XGAY3'}_{3'YAGX5'}$ and $^{5'XAGY3'}_{3'YGAX5'}$ in the middle of octanucleotides have a range of stabilities over 5 kcal/mol when XY is a Watson–Crick or GU pair. The order of stabilities for these internal loops is 5′-GGAC-3′ > UGAG, CGAG > AGAU > UGAA > GGAU. The motifs GGAC, UGAG, and CGAG are stabilizing, while the other GA motifs are destabilizing. The GAGC motif is more stable than CAGG and CGAG, but less stable than GGAC. Chemical shifts for imino protons suggest that the G imino proton of each GA mismatch in 5′-GGAC-3′, 5′-GAGC-3′, and 5′-CAGG-3′ [SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29,* 8813–8819] is involved in a hydrogen bond to the base A, whereas in other 5′-XGAY-3′ sequences, it is not involved in a hydrogen bond to the base A.

RNA structure is of increasing interest as more functions of RNA are discovered, including antisense-based regulation and catalysis (Euguchi et al., 1991; Cech & Bass, 1986). Insight into the structures of various RNA molecules will help to establish structure–function relationships for RNA. One popular method of predicting secondary structure is free energy minimization (Tinoco et al., 1971; DeLisi & Crothers, 1971; Zuker & Stiegler, 1981; Turner et al., 1988). Predictions from free energy minimizations have continuously improved as thermodynamic parameters have been measured for various motifs (Turner et al., 1987; Jaeger et al., 1989; Walter et al., 1994). One motif for which relatively little data exist is the internal loop (Gralla & Crothers, 1973; Uhlenbeck et al., 1971; Peritz et al., 1991; Santa Lucia et al., 1991a; Morse, 1991; Weeks & Crothers, 1993). In particular, the effect of different closing base pairs on internal loop stability has not been systematically investigated. There is evidence that internal loops of tandem GA mismatches in RNA are more stable when closed by GC than by CG (Morse, 1991) and that internal loops of C_n are more stable when closed by GC than by AU (Gralla & Crothers, 1973; Uhlenbeck et al., 1971). Several studies of DNA oligonucleotides containing tandem GA mismatches also suggest a dependence of internal loop stability on the closing base pair (Li et al., 1991a; Ebel et al., 1992; Cheng et al., 1992). Presumably, detailed knowledge of this sequence dependence will improve predictions of RNA folding.

An internal loop sequence commonly found in natural RNA is $^{5'GA3'}_{3'AG5'}$ (Gutell et al., 1994; Gautheret et al., 1994). The thermodynamics and structures of tandem GA mismatches closed by CG in RNA have been studied previously (SantaLucia et al., 1990; SantaLucia & Turner, 1993). Here we report that free energy changes for the internal loop motif $^{5'XGAY3'}_{3'YAGX5'}$ in the middle of octanucleotides range over almost 5 kcal/mol, depending on whether XY is GC, CG, AU, UA, GU, or UG. There also is a difference in the stability of $^{5'XAGY3'}_{3'YGAX5'}$ loops when XY is GC or CG. The observed sequence dependence differs significantly from the approximations currently used in algorithms for folding RNA. Imino proton NMR spectra indicate that the structures of GA mismatches studied here are also dependent on the closing base pair. The sequence dependence of both structure and stability in RNA differs from that previously reported in DNA (Li et al., 1991a,b; Cheng et al., 1992; Ebel et al., 1992).

MATERIALS AND METHODS

RNA Synthesis and Purification. RNA oligomers were synthesized on an Applied Biosystems DNA/RNA 392 synthesizer. RNA monomers were from Applied Biosystems, Inc. (Vinayak et al., 1992). Monomers for the sequence rGCGAGCGC were synthesized in our lab and had the base-protecting groups phenoxyacetyl for A and G and acetyl for C. Oligoribonucleotides were deblocked in 3:1 (v/v) ammonia/ethanol for 3 h at 55 °C and then overnight at room temperature. The 2′-*tert*-butyldimethylsilyl groups were removed by adding freshly made 1 M triethylammonium hydrogen fluoride in pyridine and incubating for 48 h at 55 °C. Oligonucleotides were desalted with Sep-Pak C18 cartridges (Waters). Purification was by thin-layer chromatography (Baker Si500F plates) in 55:35:10 (v/v/v) *n*-propanol/ammonium hydroxide/water (Chou et al., 1989). Final purification and desalting of the oligomer was by a Sep-Pak cartridge. Purity was checked by HPLC on a C8 RP column (Beckman) and was >95% for all oligomers.

Melting Experiments. The buffer used for melting experiments was 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM EDTA. Absorbance versus temperature melting curves were measured at 280 nm on a Gilford 250 spectrophotometer with a heating rate of 1 °C/min controlled by a Gilford 2527 thermoprogrammer (Petersheim & Turner, 1983). Curves were analyzed by fitting to a two-state model with sloping base lines using a nonlinear least-squares program (Petersheim & Turner, 1983). Two methods were used to obtain thermodynamic parameters. In one method, enthalpy and entropy changes obtained from the fitted curves were averaged. In the second method, plots of the inverse melting temperature versus log (total olignucleotide concentration) were fit to a

† This work was supported by NIH Grant GM22939. D.H.T. is a Guggenheim Fellow and an American Cancer Society Scholar.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* September 1, 1994.

0006-2960/94/0433-11349$04.50/0

Table 1: Thermodynamic Parameters of Duplex Formation[a]

RNA duplex	T_M^{-1} versus log C_T				curve fit			
	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol·K)	$-\Delta G°_{37}$ (kcal/mol)	T_M at 10^{-4} M (°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol·K)	$-\Delta G°_{37}$ (kcal/mol)	T_M at 10^{-4} M (°C)
CAG^{GA}CUG GUC$_{AG}$GAC	67.69 ± 2	198.47 ± 7	6.13 ± 0.03	39.1	60.75 ± 3	175.95 ± 11	6.18 ± 0.12	39.6
GAG^{GA}CUC CUC$_{AG}$GAG	63.33 ± 1	181.63 ± 4	7.00 ± 0.03	43.6	66.30 ± 4	190.89 ± 13	7.10 ± 0.08	43.8
GCG^{GA}CGC CGC$_{AG}$GCG	79.46 ± 2	225.04 ± 5	9.67 ± 0.08	53.4	71.80 ± 3	201.33 ± 10	9.35 ± 0.07	53.7
GCU^{GA}GGC CGG$_{AG}$UCG	65.12 ± 2	188.36 ± 5	6.70 ± 0.03	42.0	58.42 ± 6	166.71 ± 20	6.71 ± 0.06	42.6
GGC^{GA}GCC[b] CCG$_{AG}$CGG	66.08 ± 2	181.83 ± 7	9.67 ± 0.24	57.0	73.93 ± 4	205.64 ± 12	10.16 ± 0.30	57.0
GGA^{GA}UCC CCU$_{AG}$AGG	62.22 ± 1	179.95 ± 3	6.41 ± 0.02	40.7	55.26 ± 4	157.57 ± 11	6.39 ± 0.09	41.0
GCA^{GA}UGC CGU$_{AG}$ACG	63.92 ± 2	187.99 ± 7	5.62 ± 0.03	36.7	55.89 ± 6	161.69 ± 20	5.74 ± 0.16	37.4
GCU^{GA}AGC CGA$_{AG}$UCG	64.86 ± 3	190.17 ± 11	5.88 ± 0.05	38.0	56.20 ± 6	161.96 ± 19	5.97 ± 0.17	38.6
GCG^{GA}UGC CGU$_{AG}$GCG	51.48 ± 1	151.05 ± 4	4.64 ± 0.03	30.9	50.44 ± 9	147.18 ± 31	4.79 ± 0.27	31.6
GCG^{AG}CGC CGC$_{GA}$GCG	66.66 ± 1	187.66 ± 4	8.45 ± 0.05	50.5	63.46 ± 3	177.62 ± 9	8.37 ± 0.12	50.7
CGC^{AG}GCG[b] GCG$_{GA}$CGC	60.36 ± 3	169.61 ± 8	7.75 ± 0.22	48.1	65.45 ± 3	185.57 ± 9	7.90 ± 0.13	47.9
reference duplexes								
GAGCUC CUCGAG	54.85 ± 1	151.94 ± 3	7.73 ± 0.04	49.1	55.76 ± 4	154.58 ± 12	7.82 ± 0.13	49.4
GCGUGC CGUGCG	46.18 ± 1	132.43 ± 2	5.11 ± 0.01	33.2	49.69 ± 3	143.61 ± 11	5.15 ± 0.06	33.7

[a] Melting buffer is 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. Significant figures are given beyond error estimates to allow accurate calculation of T_M and other parameters. [b] SantaLucia et al., 1990.

straight line (Borer et al., 1974):

$$T_M^{-1} = (2.3R/\Delta H°) \log C_T + \Delta S°/\Delta H° \quad (1)$$

NMR Spectroscopy. RNA oligomers were dialyzed against 0.1 mM Na_2EDTA and then double-distilled water for 24 h each, dried down, and then dissolved in 0.5 mM Na_2EDTA, 10 mM sodium phosphate, and 0.5 mM TSP [sodium 3-(trimethylsilyl)tetradeuteriopropionate, as an internal standard, 0 ppm], with 0.1, 0.2, or 1.0 M NaCl, pH 7, in 90% H_2O and 10% D_2O. Most sequences were studied in 0.1 M NaCl to optimize the signal-to-noise ratio. GCUGAAGC was studied at 1 M NaCl because peaks were overlapped at 0.1 M NaCl. GCUGAGGC was studied at 0.2 M NaCl because spectra at 0.1 M NaCl had additional small peaks suggesting small concentrations of another conformer. Imino proton NMR spectra were recorded by a Varian VXR 500S spectrometer at 500 MHz. Binomial pulse sequence $1\bar{3}3\bar{1}$ was used to suppress the HDO peak (Hore, 1983). Spectra were collected with 12 000 points over a sweep width of 12 kHz, multiplied by a 5.0 Hz line-broadening exponential function, and transformed on a Sun 3/160 computer running Varian VNMR software. One-dimensional NOE (nuclear Overhauser effect) spectra were collected by irradiation for 3–5 s with low power decoupler to result in 70–90% saturation of desired resonances. Off-resonance spectra in blocks of 16 scans were interleaved to reduce the effect of long term instrumental instability.

RESULTS

Thermodynamic parameters from T_M^{-1} versus log C_T plots and from fitted melting curves are recorded in Table 1. Typical T_M^{-1} versus log C_T plots are shown in Figure 1. Enthalpy changes obtained from these plots and curve fitting agree within 15%, consistent with the two-state model (Freier et al., 1983; Petersheim & Turner, 1983). Thermodynamic parameters of loop formation are listed in Table 2, as calculated from equations analogous to the following (Gralla & Crothers, 1973): $\Delta G°_{37,loop} = \Delta G°_{37}(rGCXGAYGC) - \Delta G°_{37}(rGCXYGC) + \Delta G°_{37}(^{XY}_{YX})$. The free energy change, $\Delta G°_{37}(^{XY}_{YX})$, is taken from Freier et al. (1986) for Watson–Crick or He et al. (1991) for GU nearest-neighbor interactions interrupted by the internal loops. This yields the thermodynamic parameters for various closing base pairs around the tandem mismatches $^{5'XGAY3'}_{3'YAGX5'}$ and $^{5'XAGY3'}_{3'YGAX5'}$. The error for $\Delta G°_{37,loop}$, $\sigma\Delta G°_{37,loop}$, was obtained from (SantaLucia et al., 1991) $\sigma\Delta G°_{37,loop} = [\sigma^2\Delta G°_{37}(GCXGAYGC) + \sigma^2\Delta G°_{37}(GCXYGC) + \sigma^2\Delta G°_{37}(^{XY}_{YX})]^{1/2}$. The loop free energy parameters for the three sequences containing 5′-GGAC-3′ and for the two sequences containing 5′-AGAU-3′ are in agreement within the limits expected for the nearest-neighbor model, given that fully base-paired octamer duplexes with identical nearest neighbors can have free energies that differ by about 10% (Kierzek et al., 1986; Freier et al., 1986).

Imino proton NMR spectra are shown in Figure 2. Duplexes with the same base pairs closing the GA mismatch are placed together. All of the duplexes are self-complementary octamers and have a center of symmetry. Thus, each resonance is attributed to two imino protons. $(GCGAGCGC)_2$ and $(GCGGACGC)_2$ have six Watson–Crick base pairs, so we expect to see three resonances in the imino proton region (usually from 9 to 15 ppm). Since there are four resonances in that region (Figure 2A,B), the fourth one is presumably

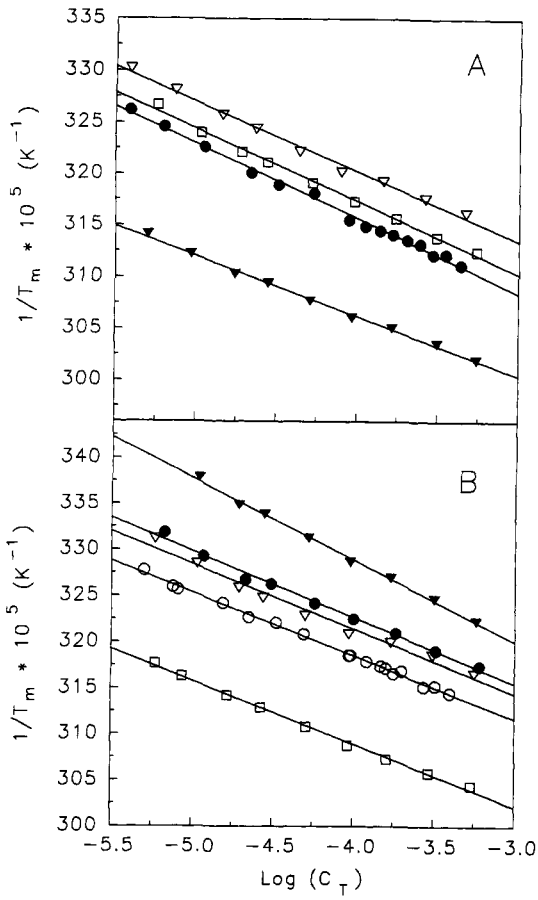


FIGURE 1: Plots of inverse melting temperature versus log concentration for (A) rGAGGACUC (●), rCAGGACUG (▽), rGCGGACGC (▼), and GCUGAGGC (□) and for (B) rGGAGAUCC (○), rGCAGAUGC (●), rGCUGAAGC (▽), rGCGGAUGC, (▼), and rGCGAGCGC (□).

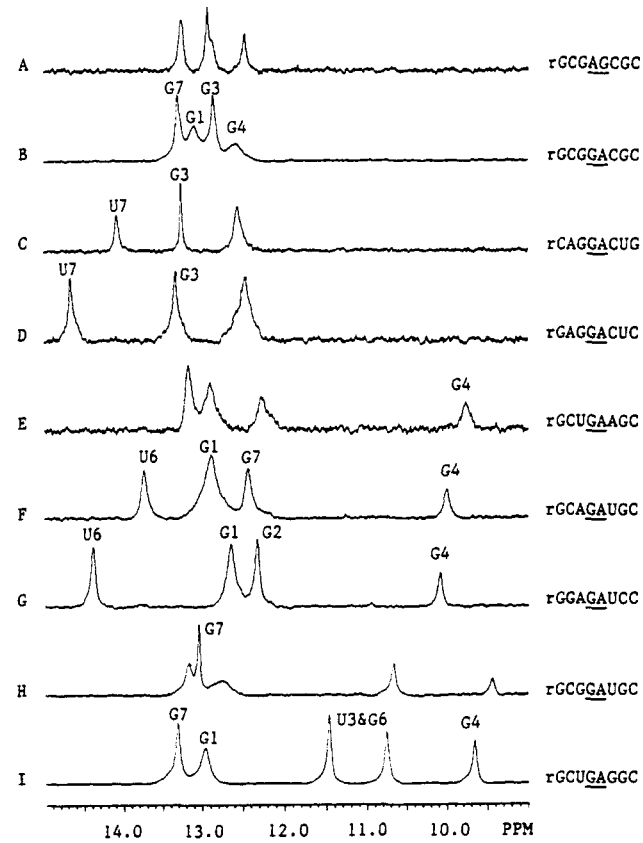


FIGURE 2: 500 MHz imino proton NMR spectra (9–15 ppm) in 0.1 M NaCl, 10 mM sodium phosphate, 0.5 mM Na_2EDTA, and 0.5 mM TSP (unless otherwise stated) for (A) rGCGAGCGC at 20 °C, (B) rGCGGACGC at 13 °C, (C) rCAGGACUG at 20 °C, (D) rGAGGACUC at 20 °C, (E) rGCUGAAGC in 1.0 M NaCl at 20 °C, (F) rGCAGAUGC at 0 °C, (G) rGGAGAUCC at 10 °C, (H) rGCGGAUGC at 20 °C, and (I) rGCUGAGGC in 0.2 M NaCl at 15 °C. Preliminary assignments are based on the expected positions for AU and GC resonances and on the temperature dependence of line widths. Assignments for rGCGGACGC and rGCUGAGGC are based on NOE difference spectra (e.g., Figure 3). Qualitatively similar spectra were observed in 1 M NaCl for sequences A–D, F, and H and in 0.1 M NaCl for I. For sequence E, the two peaks near 13 ppm in 1 M NaCl apparently overlap in 0.1 M NaCl.

from the G4 imino proton of the GA mismatch. For $(GCGGACGC)_2$, this was confirmed by the observation of an NOE between the imino resonance at 12.51 ppm and the A5H2 resonance at 7.76 ppm (supplementary material). $(CAGGACUG)_2$ and $(GAGGACUC)_2$ have three resonances each (Figure 2C,D). For both sequences, however, the resonance at approximately 12.5 ppm covers almost twice as much area as either of the other two resonances and is presumably the superposition of the resonances of the GA mismatch and the terminal GC pair. $(GCUGAAGC)_2$ has four resonances (Figure 2E). The resonance at 9.78 ppm is assigned to the G4 imino proton of the GA mismatch on the basis of a comparison with the imino proton NMR spectrum of $(GCUAAAGC)_2$ (data not shown), which shows three resonances similar, in both chemical shift and line width, to the three resonances of $(GCUGAAGC)_2$ between 12 and 13.5 ppm. $(GCAGAUGC)_2$ and $(GGAGAUCC)_2$ (Figure 2F,G) have a resonance at about 10 ppm, which is assigned to the G4 imino proton of the GA mismatch since both GC and AU pairs typically resonate above 12 ppm. $(GCGGAUGC)_2$ and $(GCUGAGGC)_2$ have five resonances as expected because GU mismatches give rise to two resonances (Figure 2H,I). Evidently, the G imino proton of each GA mismatch is protected from exchange with water since each spectrum has one more resonance than expected from the other base pairs. The preliminary assignments, however, suggest that at least two different structures are responsible for protection since some GAs resonate near 10 ppm, while others resonate between 12 and 13 ppm.

Assignments for $(GCUGAGGC)_2$ are based on 1D NOE difference spectra (Figure 3). The resonance at 12.96 ppm is assigned to the terminal GC base pair (G1) because it broadens first when the temperature is raised, due to terminal base pair fraying. There is an NOE to this G1 imino proton only upon irradiation at 13.31 ppm (Figure 3E). The resonance at 13.31 ppm is therefore assigned to G7. Irradiation of G7 (13.31 ppm) gives rise to NOEs at 10.75 and 11.45 ppm in addition to 12.96 ppm (G1) (Figure 3E). Therefore, resonances at 10.75 and 11.45 ppm are assigned to U3 and G6 of the GU mismatch. Strong NOEs from U3 to G6 and from G6 to U3 (Figure 3C,D) are characteristic of GU mismatches, and the chemical shifts of U3 and G6 are consistent with those previously reported for GU mismatches (He et al., 1991; Hare & Reid, 1982). Irradiations at 10.75 and 11.45 ppm both give rise to NOEs at 13.31 (G7) and 9.66 ppm. The latter is therefore assigned to the G4 imino proton of the GA mismatch. NOEs from G4 to U3 and G6 (Figure 3B) confirm the previous assignments.

DISCUSSION

The closing base pairs for internal loops of tandem GA mismatches have dramatic effects on the thermodynamic parameters of loop formation. For 5′-XGAY-3′, the range

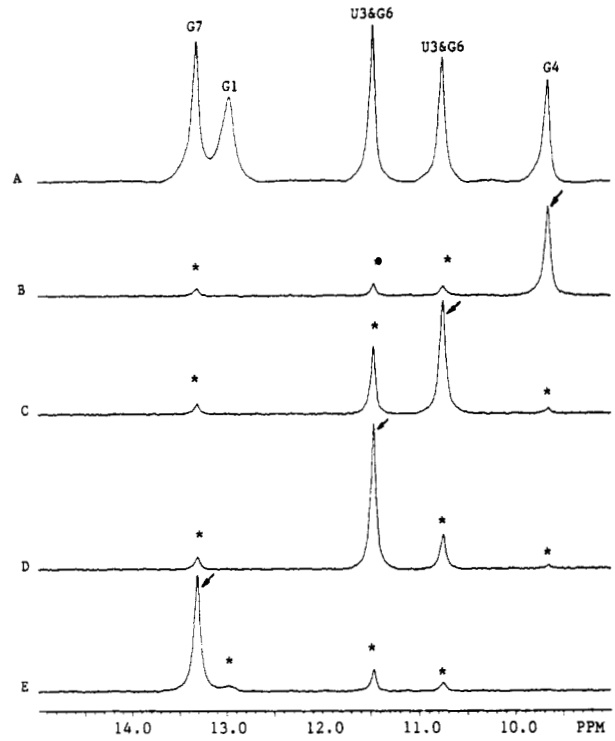
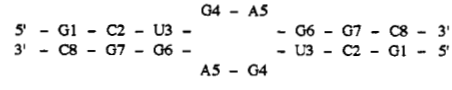


FIGURE 3: Imino proton NOE difference spectra (9–15 ppm) of 1.6 mM rGCUGAGGC at 15 °C in 0.2 M NaCl, 20 mM sodium phosphate, 1.0 mM Na_2EDTA, and 1 mM TSP in 90% H_2O and 10% D_2O: (A) off-resonance spectrum; (B–E) difference spectra between the off-resonance spectra and on-resonance spectra acquired with 5 s saturation at 9.66, 10.75, 11.45, and 13.31 ppm, respectively. The saturated resonances are designated by arrows, while the observed NOEs are designated by asterisks. This is the way each nucleotide is numbered, and the assignments are shown in spectrum A.



in loop free energies is –2.9 to +2 kcal/mol (Table 2). The loop is most stable when the closing base pair XY is GC (average $\Delta G°_{37} = -2.7$ kcal/mol). When XY is UG or CG, loop formation is slightly stabilizing (average $\Delta G°_{37} = -0.4$ kcal/mol). The closing base pairs AU, UA, and GU destabilize loop formation, with $\Delta G°_{37,loop}$ ranging from +0.1 to +2.0 kcal/mol. When XY is GC for the loop 5′-XAGY-3′, the loop is about 0.6 kcal/mol more stable than when XY is CG.

There is also a large difference in free energy for the loops 5′-GGAC-3′ and 5′-GAGC-3′. Duplex formation with rGCGGACGC is 1.2 kcal/mol more favorable than with rGCGAGCGC. This contrasts with the result for CG closure, where 5′-CGAG-3′ is 0.3 kcal/mol less favorable than 5′-CAGG-3′ (SantaLucia et al., 1990). These results indicate that the base closure of the loop and the sequence of the loop are both important for determining the overall stability of the loop.

The stability of double GA mismatches with various closing base pairs in DNA has been studied by thermodynamic measurements and NMR. From thermodynamic measurements, Li et al. (1991a) concluded GC closing base pairs provide more stability than AT closing base pairs. The results shown here for RNA are similar. Li et al. (1991a) also propose that a 5′-pyrimidine-GA-purine-3′ sequence is more stable than a 5′-purine-GA-pyrimidine-3′ sequence when comparing a 5′-C to a 5′-G or a 5′-T to a 5′-A. This is not seen with RNA where closure by a 5′-G in a GC pair is far more stable than a 5′-C in a CG pair. By imino proton NMR studies on DNA oligomers, Cheng et al. (1992) found that the GGAC motif is much less stable than AGAT, TGAA, or CGAG. Interestingly, this is opposite the result in RNA, where GGAC is the most stable motif.

Sometimes the thermodynamic results of a particular motif can be related to the occurrence of the motif in nature (Tuerk et al., 1988; He et al., 1991; Peritz et al., 1991). From the phylogenetic structures of 51 large subunit (Gutell et al., 1992, 1993) and 21 small subunit (Gutell et al., 1985) ribosomal RNAs, there are 71 cases of tandem GA internal loops closed by Watson–Crick or GU base pairs. The number of occurrences for each closing base pair are listed in Table 3. By far the most common closing base pairs are CG, UG, and GC. These all have favorable free energies in the self-complementary oligomers studied here. Since these are the most common in phylogenetic structures and the most stable, these base pairs may have been selected by nature partially due to their stability. Of the self-complementary motifs listed in Table 2, however, only CGAG is actually found in the database of ribosomal RNA structures (D. Gautheret, personal communication). Thus, it will be necessary to determine whether the dependence of stability on the closing base pair observed in Table 2 is also found when the closing base pairs are different on the two sides of the tandem GA mismatch. Surely stability is not the only determinant of selection since tandem GA mismatches with sequence 5′-AG-3′ are never observed even though they are thermodynamically favorable (SantaLucia et al., 1990).

For the RNA loops studied here, loops with GC closure are always the most stable. Morse (1991) also saw that GC closure is much more stable than CG closure around GA and AG internal loops. The same pattern is observed for GU and UG tandem mismatches where a GC closing pair is dramatically more stable than CG, UA, and AU closing pairs (He et al., 1991). He et al. (1991) pointed out that the most stable nearest-neighbor combinations of Watson–Crick base pairs and of a Watson–Crick base pair and a GU mismatch all contain a 5′-G. Evidently, tandem GA mismatches fit this trend of stabilization by 5′-G, with the exception that $\begin{matrix}5'GGAU3' \\ 3'UAGG5'\end{matrix}$ is the least stable tandem. It is possible, however, that the latter result may be an artifact of the $\Delta G°_{37}$ used for the GU/GU nearest-neighbor interaction interrupted by the tandem GA mismatch. The $\Delta G°_{37}$ for GU/GU varies from –1.0 to +1.5 kcal/mol, depending on context, so that non-nearest-neighbor effects are important for this sequence (He et al., 1991).

This work illustrates the importance of the base pairs closing RNA internal loops. The measured sequence dependence differs from that used in structure prediction algorithms primarily because of the unexpected enhancement of stability by GC closure. This is illustrated in Table 2, where approximations of free energies used in folding algorithms (Gralla & Crothers, 1973; Tinoco et al., 1973; Jaeger et al., 1989; Walter et al., 1994) are compared with the measured loop free energies. The enhancement from GC closure also differs from observations on the hairpin loop AUAAUA, where Serra et al. (1993) found equal stabilities for GC and CG closure. These findings and other information about RNA loops should help improve RNA structure predictions.

The imino proton NMR spectra indicate that the structures of the tandem GA mismatches in RNA are also dependent on the closing base pair. The imino protons of the GA mismatches in sequences with a GGAC motif appear to resonate between 12 and 13 ppm, whereas they resonate near

Table 2: Thermodynamic Parameters of Loop Formation[a]

RNA duplex	measured ΔH°_{loop} (kcal/mol)	measured ΔS°_{loop} (eu)	measured $\Delta G^\circ_{37,loop}$ (kcal/mol)	approximations in folding algorithms,[c] $\Delta G^\circ_{37,loop}$ (kcal/mol) Jaeger et al., 1989	Walter et al., 1994
CAGGACUG	−30.3 ± 3.4	−88.7 ± 11	−2.85 ± 0.1	+1.7	−0.5
GAGGACUC	−22.7 ± 1.8	−64.6 ± 5.6	−2.67 ± 0.1	+1.7	−0.5
GCGGACGC	−27.7 ± 2.7	−81.4 ± 8.2	−2.45 ± 0.2	+1.7	−0.5
GCUGAGGC	−17.1 ± 2.6	−53.6 ± 8.4	−0.39 ± 0.1	+2.4	0
GGCGAGCC[b]	−6.4 ± 2.6	−19.4 ± 7.7	−0.36 ± 0.3	+2.1[d]	−0.5
GGAGAUCC	−14.2 ± 2.1	−46.4 ± 6.7	+0.13 ± 0.1	+3.3	+0.5
GCAGAUGC	−7.3 ± 2.8	−26.3 ± 9.3	+0.86 ± 0.2	+3.3	+0.5
GCUGAAGC	−13.8 ± 4.1	−47.7 ± 13	+0.94 ± 0.1	+2.7	+0.5
GCGGAUGC	−14.6 ± 2.1	−53.7 ± 6.7	+1.97 ± 0.1	+3.3	+0.5
GCGAGCGC	−14.9 ± 2.6	−44.1 ± 7.8	−1.23 ± 0.2	+2.3	−0.5
CGCAGGCG[b]	−13.9 ± 2.8	−42.6 ± 8.4	−0.64 ± 0.2	+2.1[d]	−0.5

[a] Calculations are based on results from T_M^{-1} versus log C_T plots. Loop parameters are calculated as follows: $\Delta G^\circ_{37,loop} = \Delta G^\circ_{37}(rGCXGAYGC) - \Delta G^\circ_{37}(GCXYGC) + \Delta G^\circ_{37}(^{XY}_{YX})$. The reference duplexes are listed in Table 1 or Freier et al. (1986), except for GCUGGC and GCUAGC (Sugimoto et al., 1986) and GGAUCC (He et al., 1991). [b] SantaLucia et al., 1990. [c] Initially all internal loops of four were given a ΔG°_{25} of +2 kcal/mol (Tinoco et al., 1973; Gralla & Crothers, 1973). Listed in the table are subsequent approximations that have been used for the prediction of RNA secondary structures. [d] Uses ΔG°_{37} = −1.4 kcal/mol for $^{CG}_{GA}$ and $^{CA}_{GG}$ mismatches (SantaLucia et al., 1991b).

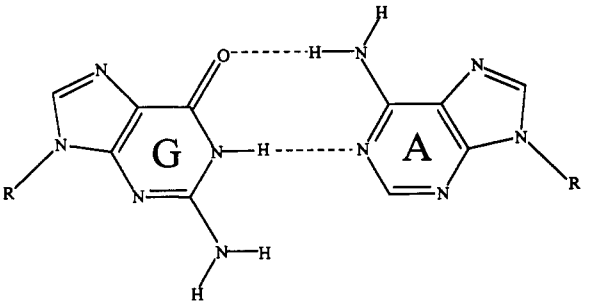
Table 3: Occurrences of 5′ Closing Base Pairs in Phylogenetic Structures of 51 Large Subunit[a] and 21 Small Subunit[b] Ribosomal RNAs

motif	5′CGA GAG	5′UGA GAG	5′GGA CAG	5′AGA UAG	5′UAG AAG	5′GGA UAG
occurrences	54[c]	41	28	10	7	2

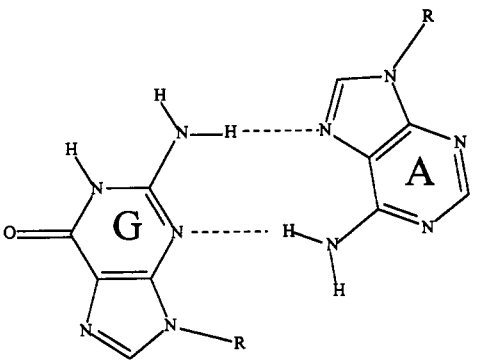
[a] Gutell et al., 1992, 1993. [b] Gutell et al., 1985. [c] 34 of these occurrences are due to 17 occurrences of CGAG.

10 ppm for the 5′-GA-3′ sequence with other Watson–Crick closing pairs [Figure 2 and SantaLucia et al. (1991a)]. Imino proton NMR spectra of tandem GA mismatches in DNA are also known to depend on the closing base pair (Cheng et al., 1992). In both DNA (Kan et al., 1983; Cheng et al., 1992) and RNA (SantaLucia et al., 1990), resonances between 12 and 13 ppm have been attributed to the G imino proton hydrogen bonding to adenine (structure I in Figure 4). For the $(GCGGACGC)_2$ duplex studied here, structure I is confirmed by the observation of an NOE between the G4 imino and A5H2 protons (supplementary material). Similarly, in both DNA (Li et al., 1991b; Cheng et al., 1992; Maskos et al., 1993) and RNA (Heus & Pardi, 1991; SantaLucia & Turner, 1993), resonances near 10 ppm have been attributed to a hydrogen-bonding pattern in which the G imino proton does not hydrogen bond to adenine (structure II in Figure 4). For example, the GA mismatches in $(GGCGAGCC)_2$ are known to exist in structure II on the basis of a structure determination by NMR and restrained molecular dynamics (SantaLucia & Turner, 1993). Thus, in RNA, it appears that GGAC, GAGC (Figure 2 and supplementary material), and CAGG (SantaLucia et al., 1990) motifs adopt structure I (Figure 4), whereas 5′-GA-3′ tandem mismatches closed by other Watson–Crick pairs (Figure 2) adopt structure II (Figure 4). This differs from DNA, where AGAT adopts structure I and CGAG and TGAA adopt structure II (Cheng et al., 1992). The striking sequence dependencies and contrasts between RNA and DNA in both stability and structure suggest that tandem GA mismatches may be an interesting motif for computational studies aimed at understanding the interactions governing these properties.

The dependence of GA structure on the adjacent base pair observed for RNA internal loops may extend to other types of RNA loops. For example, yeast phenylalanine tRNA has a $^{GA}_{CG}$ sequence in the multibranch loop, and the crystal structure has the AG mismatch in structure I of Figure 4



(I)



(II)

FIGURE 4: Two known conformations of GA mismatches in GA tandems.

(Kim et al., 1974; Robertus et al., 1974). This is consistent with the interpretation of the imino proton NMR spectrum of $(GCGAGCGC)_2$. Conversely, the dihydrouridine hairpin loop of *Thermus thermophilus* serine tRNA has a $^{CG}_{GA}$ sequence, and the crystal structure has the GA mismatch in structure II of Figure 4 (Biou et al., 1994). This is consistent with the structure of $(GGCGAGCC)_2$ derived from NMR (SantaLucia & Turner, 1993).

The dependence of GA structure on adjacent base pair may be important for tertiary interactions in RNA. The GA

mismatch in serine tRNA described above is involved in a base triple with another G (Biou et al., 1994) that would be impossible if the GA was in structure I. The NMR results presented here suggest that this tertiary interaction would be less favorable if the sequence was switched from $^{CG}_{GA}$ to $^{GG}_{CA}$. Adjacent to this base triple is another triple involving the recognition of an AA pair by a U (Biou et al., 1994). Gautheret et al. (1994) have pointed out that $^{GA}_{AA}$ tandem mismatches are often replaced by $^{GA}_{AG}$ tandems. The double base triple motif observed with the $^{GA}_{AA}$ tandem is also possible with a $^{GA}_{AG}$ tandem if both GA mismatches have structure II.

The dependence of GA structure on the adjacent base pair may also have functional consequences. For example, Ruffner et al. (1990) found that switching from a CG to a GC pair adjacent to a tandem GA mismatch in the multibranch loop of the hammerhead ribozyme decreased activity by more than a hundredfold. This could result from a switch from structure II to structure I of Figure 4. This is not the only possibility, however, since an AU pair in the same position also decreased the activity by more than a hundredfold, even though the NMR data presented here suggest that AU should favor structure II.

ACKNOWLEDGMENT

We thank Drs. Art Pardi, Daniel Gautheret, and Kathy Morden for comments on the manuscript.

SUPPLEMENTARY MATERIAL AVAILABLE

One figure showing the NOE difference spectrum for $(GCGGACGC)_2$ upon irradiation of the imino resonance at 12.51 ppm (1 page). Ordering information is given on any current masthead page.

REFERENCES

Biou, V., Yaremchuk, A., Tukalo, M., & Cusack, S. (1994) *Science 263*, 1404–1410.
Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843–853.
Cech, T. R., & Bass, B. L. (1986) *Annu. Rev. Biochem. 55*, 599–629.
Cheng, J.-W., Chou, S.-H., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 1037–1041.
Chou, S.-H., Flynn, P., & Reid, B. R. (1989) *Biochemistry 28*, 2422–2435.
DeLisi, C., & Crothers, D. M. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2682–2685.
Ebel, S., Lane, A. N., & Brown, T. (1992) *Biochemistry 31*, 12083–12086.
Euguchi, Y., Itoh, T., & Tomizawa, J. (1991) *Annu. Rev. Biochem. 60*, 631–652.
Freier, S. M., Burger, B. J., Alkema, D., Neilson, T., & Turner, D. H. (1983) *Biochemistry 22*, 6198–6206.
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373–9377.
Gautheret, D., Konings, D., & Gutell, R. R. (1994) *J. Mol. Biol.* (in press).
Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301–319.
Gutell, R. R., Weiser, B., Woese, C. R., & Noller, H. F. (1985) *Prog. Nucleic Acid Res. Mol. Biol. 23*, 155–195.
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.
Gutell, R. R., Larsen, N., & Woese, C. R. (1994) *Microbiol. Rev. 58*, 10–56.
Hare, D. R., & Reid, B. R. (1982) *Biochemistry 21*, 5129–5135.
He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., & Turner, D. H. (1991) *Biochemistry 30*, 11124–11132.
Heus, H. A., & Pardi, A. (1991) *Science 253*, 191–193.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.
Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706–7710.
Kan, L.-S., Chandrasegarank, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.
Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., & Freier, S. M. (1986) *Biochemistry 25*, 7840–7846.
Kim, S. H., Suddath, F. L., Quigley, G. J., McPhereson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974) *Science 185*, 435–440.
Li, Y., Zon, G., & Wilson, W. D. (1991a) *Biochemistry 30*, 7566–7572.
Li, Y., Zon, G., & Wilson, W. D. (1991b) *Proc. Natl. Acad Sci. U.S.A. 88*, 26–30.
Maskos, K., Gunn, B. M., LeBlanc, D. A., & Morden, K. M. (1993) *Biochemistry 32*, 3583–3595.
Morse, S. (1991) Ph.D. Thesis, Johns Hopkins University, Baltimore, MD.
Peritz, A. E., Kierzek, R., Sugimoto, N., & Turner, D. H. (1991) *Biochemistry 30*, 6428–6436.
Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 256–263.
Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. D., Clark, B. F. C., & Klug, A. (1974) *Nature 250*, 546–551.
Ruffner, D. E., Stormo, G. D., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 10695–10702.
SantaLucia, J., Jr., & Turner, D. H. (1993) *Biochemistry 32*, 12612–12623.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813–8819.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991a) *Biochemistry 308*, 8242–8251.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991b) *J. Am. Chem. Soc. 113*, 4313–4322.
Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A., & Turner, D. H. (1993) *Nucleic Acids Res. 21*, 3845–3849.
Sugimoto, N., Kierzek, R., Freier, S. M., & Turner, D. H. (1986) *Biochemistry 25*, 5755–5759.
Tinoco, I., Jr., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature 230*, 362–367.
Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) *Nature New Biol. 246*, 40–41.
Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N., & Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364–1368.
Turner, D. H., Sugimoto, N., Jaeger, J. A., Longfellow, C. E., Freier, S. M., & Kierzek, R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 123–133.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167–192.
Uhlenbeck, O. C., Martin, F. H., & Doty, P. (1971) *J. Mol. Biol. 57*, 217–229.
Vinayak, R., Anderson, P., McCollum, C., & Hampel, A. (1992) *Nucleic Acids Res. 20*, 1265–1269.
Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., & Zuker, M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Weeks, K. M., & Crothers, D. M. (1993) *Science 261*, 1574–1577.
Zuker, M., & Stiegler, P. (1981) *Nucleic Acids Res. 9*, 133–148.