Effects of GA Mismatches on the Structure and Thermodynamics of RNA Internal Loops†

John SantaLucia, Jr.,‡ Ryszard Kierzek,§ and Douglas H. Turner*,‡

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

Received March 22, 1990; Revised Manuscript Received June 4, 1990

ABSTRACT: UV melting, CD and NMR studies indicate rGCGAGCG and rGCAGGCG from unusually stable duplexes of type a and b. The observed $\Delta G°_{37}$'s in 1 M NaCl are −6.7 and −6.3 kcal/mol, respectively. For the related duplex, c, $\Delta G°_{37}$ is −4.2 kcal/mol. The predicted $\Delta G°_{37}$ from nearest-neighbor parameters

```
5'  GC GA GCG  3'        5'  GC AG GCG  3'        5'  GC AA GCG  3'
3'  GCG AG CG  5'        3'  GCG GA CG  5'        3'  GCG AA CG  5'
       a                        b                        c
```

for all three duplexes is −4.7 kcal/mol (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). The results suggest a special interaction in the duplexes containing GA mismatches. Presumably, this is hydrogen bonding between G and A. While the thermodynamics for $(rGCGAGCG)_2$ and $(rGCAGGCG)_2$ are similar, CD and the imino region of the proton NMR spectra indicate their structures are different. In particular, $(rGCAGGCG)_2$ exhibits a CD spectrum typical of A-form geometry with a weak negative band at 280 nm. In contrast, the CD spectrum for $(rGCGAGCG)_2$ has an intense positive band at 285 nm. The NMR spectrum of $(rGCAGGCG)_2$ has a resonance corresponding to a hydrogen-bonded GA mismatch, while for $(rGCGAGCG)_2$ no hydrogen-bonded imino proton is observed for the mismatch. The glycosidic torsion angles of the bases in the GA mismatches of $(rGCAGGCG)_2$ and $(rCGCAGGCG)_2$ are anti. Duplexes of type d, where X is A, G, or U, are more stable than e, and the stability differences are similar to those

```
5'  GC GA GCX  3'        5'  GC GA GC  3'        5'  GCGCX  3'        5'  GCGC  3'
3'  XCG AG CG  5'        3'  CG AG CG  5'        3'  XCGCG  5'        3'  CGCG  5'
       d                        e                        f                  g
```

observed for f versus g. Thus, 3′-dangling ends in this system make contributions to duplex stability that are similar to contributions observed with fully paired duplexes.

In the past, only interactions between Watson–Crick and GU base pairs had been considered as contributing to the stability of RNA folding (Salser, 1977). Recent work has indicated non-Watson–Crick interactions are also important (Freier et al., 1986a,b; Turner et al., 1988; Tuerk et al., 1988). We present two examples of RNA molecules where stability is enhanced by non-Watson–Crick interactions:

```
5'  GC GA GCG  3'        5'  GC AG GCG  3'
3'  GCG AG CG  5'  and   3'  GCG GA CG  5'
```

Both contain an internal loop; i.e., the helix is interrupted by unpaired nucleotides on both strands. Henceforth, internal loop nucleotides are underlined. Preliminary measurements indicated rGCGAGCG self-associates to form a complex with a melting temperature near 45 °C and an unusual CD[1] spectrum (Longfellow et al., 1990). We show that both rGCGAGCG and rGCAGGCG form duplexes that are stabilized by stacking of unpaired terminal nucleotides and by interactions in the internal loops. Moreover, the structure of the duplex depends on the sequence in the internal loop.

A structure very similar to rGCGAGCG is found in the conserved region of *Escherichia coli* 23S rRNA between nucleotides 2348 and 2369 (Gutell & Fox, 1988):

```
5'  UGC GA GCG  3'
3'  ACG AG CGC  5'
```

Moreover, G-A internal loops are often found in rRNAs, whereas A-G internal loops are not. Our results suggest the reason for this preference is structural rather than thermodynamic.

Predictions of RNA structure from sequence are sensitive to parameters for internal loops (Turner et al., 1987a). Moreover, internal loops with terminal GA mismatches are particularly common (Traub & Sussman, 1982; Noller, 1984; Woese, 1983). Nevertheless, few model systems containing internal loops have been studied (Gralla & Crothers, 1973; Varani et al., 1989). Current models for internal loops consider only the number of residues on each side of the internal loop (Jaeger et al., 1989; Papanicolaou et al., 1984), the base pairs closing the loop (Gralla & Crothers, 1973; Turner et al., 1987a), and/or stacking of bases on the closing base pairs (Jaeger et al., 1989). Potential hydrogen bonding interactions within internal loops are not included. Our results suggest

† This work was supported by National Institutes of Health Grant GM22939 to D.H.T. and by NIH Grant RR 03317 and NSF Grant DBM 8611927 for purchase of the NMR spectrometer.

* Author to whom correspondence should be addressed.

‡ University of Rochester.

§ Polish Academy of Sciences.

[1] Abbreviations: CD, circular dichroism; Na_2EDTA, disodium ethylenediaminetetraacetate; HPLC, high-performance liquid chromatography; NOE, nuclear Overhauser enhancement; rRNA, ribosomal ribonucleic acid.

0006-2960/90/0429-8813$02.50/0

these models are oversimplified.

Materials and Methods

RNA Synthesis and Purification. Oligoribonucleotides were synthesized on solid support by using a phosphoramidite approach with the 2′-hydroxyl protected as the tetrahydropyranyl acetal (Markiewicz et al., 1984; Kierzek et al., 1986). After deblocking with ammonia, the crude mixture was purified by HPLC on a PRP-1 semipreparative column (Hamilton) with a gradient of 0–50% acetonitrile buffered with 10 mM ammonium acetate, pH 7 (Ikuta et al., 1984; Longfellow et al., 1989). After removal of the acid-labile protecting groups, the RNA was desalted and further purified with a Sep-pak C-18 cartridge (Waters). Purities were checked by analytical C-8 HPLC (Beckman) and were greater than 95%.

Melting Curves. The buffer for thermodynamic studies was 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2-EDTA, pH 7. Single-strand extinction coefficients were calculated from extinction coefficients for dinucleoside monophosphates and nucleotides, as described previously (Borer, 1975; Richards, 1975). In units of 10^4 M^{-1} cm^{-1}, the calculated extinction coefficients at 280 nm are as follows: rCGCAGGCG, 4.16; rGCAAGCG, 3.36; rGCAGGCG, 3.73; rGCGAGC, 3.05; rGCGAGCA, 3.21; rGCGAGCG, 3.61; rGCGAGCU, 3.34; rGGCGAGCC, 4.29. Strand concentrations were determined from high-temperature absorbance at 280 nm. Absorbance versus temperature melting curves were measured at 280 nm with a heating rate of 1.0 °C min^{-1} on a Gilford 250 spectrophotometer as described previously (Freier et al., 1983; Petersheim & Turner, 1983). Oligomer concentration was varied over an 80–140-fold range.

Data Analysis. Absorbance versus temperature profiles were fit to a two-state model with sloping baselines by using a nonlinear least-squares program (Petersheim & Turner, 1983; Freier et al., 1983). Thermodynamic parameters for duplex formation were obtained by two methods: (1) enthalpy and entropy changes from fits of individual melting curves were averaged and (2) plots of reciprocal melting temperature (T_M^{-1}) versus the logarithm of the strand concentration (log C_T) were fit to eq 1 (Borer et al., 1974):

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

CD Spectroscopy. CD spectra were measured with a Jasco J-40 spectropolarimeter. The buffer was the same as for the UV melting studies. The measured CD was converted to $\Delta\epsilon$ as described by Cantor & Schimmel (1980).

NMR Spectroscopy. NMR spectra were measured with a 500-MHz Varian VXR-500S spectrometer. Oligomers were dissolved in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA, at pH 7 or 6.4. Higher NaCl concentrations were not used due to limited solubility of oligomers at higher salt. The solvent for imino proton studies was 10% D_2O and 90% H_2O. Spectra were collected with 16 000 points over a sweep width of 10 kHz, multiplied by a 4.0-Hz line-broadening exponential function, and transformed by a Sun 3/160 computer running Varian VNMR software. Exchangeable proton spectra were recorded by using the solvent suppression 1:$\bar{3}$:3:$\bar{1}$ pulse sequence (Hore, 1983). The frequency offset was set to maximize the signal to noise ratio at 12.5 ppm. One-dimensional NOE difference spectroscopy of the imino protons was performed by using the 1:$\bar{3}$:3:$\bar{1}$ pulse sequence as the read pulse. Experiments with irradiation on and off resonance were collected in blocks of 16 scans and interleaved to correct for long-term instrumental drift. A presaturation period of 3 s was used to achieve a steady state. For each NOE experiment, the off-resonance frequency was set to minimize spillover artifacts. Three experiments with different decoupler powers were used to distinguish true NOE's from spillover artifacts (Varani et al., 1989).

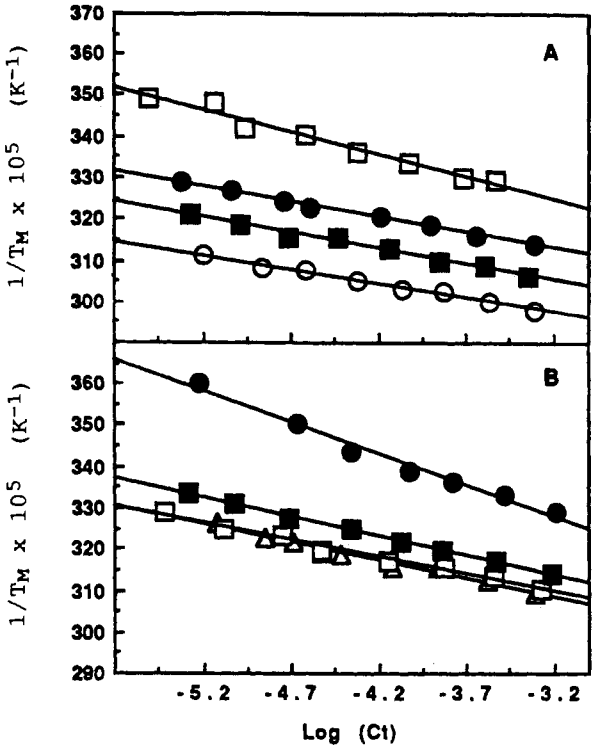

FIGURE 1: Reciprocal melting temperature versus log concentration plots for (A) rGCAAGCG (□), rGCAGGCG (●), rCGCAGGCG (■), and rGGCGAGCC (○) and for (B) rGCGAGC (●), rGCGAGCA (△), rGCGAGCG (□), and rGCGAGCU (■). Solutions are 1.0 M NaCl, 0.01 M sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7.

Results

Plots of T_M^{-1} versus log C_T are shown in Figure 1. All sequences show concentration-dependent T_M's, indicating complexes with molecularity greater than 1. Thermodynamic parameters derived from fits of melting curves and from T_M^{-1} versus log C_T plots are listed in Table I. In all cases, the parameters derived from the two methods agree within 10%. This suggests the bimolecular, two-state model is a reasonable approximation for these transitions (Freier et al., 1983; Petersheim & Turner, 1983). For example, if triplexes were forming in a two-state transition, the fitted ΔH° would be expected to be about 50% larger than the ΔH° obtained from eq 1 (Marky & Breslauer, 1987). If significant concentrations of intermediates were present, then the ΔH° from fits is expected to be smaller than the ΔH° obtained from eq 1 (Hickey & Turner, 1985). Thus, the thermodynamic data are consistent with all the sequences studied forming duplexes in a two-state manner.

CD spectra for the duplexes in 1 M Na^+ at 0 °C are shown in Figure 2. Two classes of spectra are observed. Oligomers with an A-G sequence have a small negative band near 280 nm and spectra that are similar to those expected for A-form RNA (Tunis-Schneider & Maestre, 1970). Oligomers with a G-A sequence have positive ellipticity in the 285-nm region that is not observed for fully paired A-form RNA. Thus a single inversion of sequence appears to have a large effect on conformation. The CD spectrum for rGCGAGCG is quantitatively independent of pH between 5.8 and 8.0 (data not shown). This rules out the pH-dependent structure observed for GA mismatches by Gao and Patel (1988).

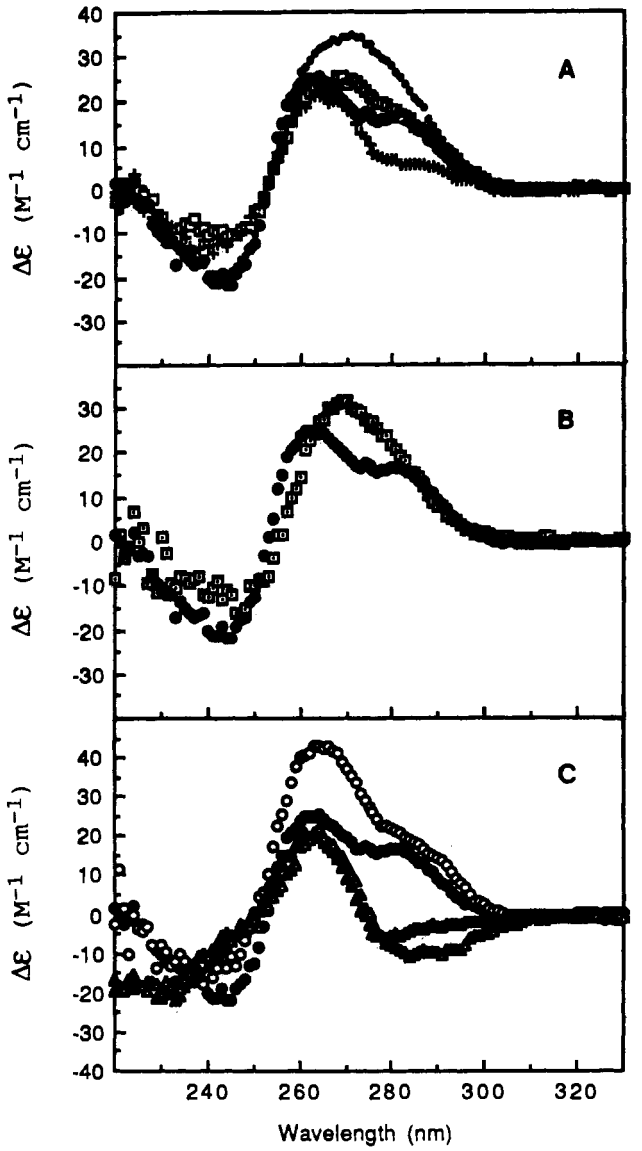

FIGURE 2: CD spectra at 0 °C for (A) 5.7×10^{-5} M rGCGAGC (□), 5.3×10^{-5} M rGCGAGCA (+), 6.1×10^{-5} M rGCGAGCG (●), and 7.5×10^{-5} M rGCGAGCU (•), for (B) 2.7×10^{-5} M rGCAAGCG (⊡) and 6.1×10^{-5} M rGCGAGCG (●), and for (C) 6.1×10^{-5} M rGCGAGCG (●), 6.1×10^{-5} M rGGCGAGCC (○), 9.3×10^{-5} M rGCAGGCG (▲), and 4.9×10^{-5} M rCGCAGGCG (△). Solutions are 1.0 M NaCl, 0.01 M sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7.

NMR spectra for the imino protons of the duplexes in 0.1 M NaCl are shown in Figures 3 and 4. Again, two classes of spectra are observed. In the 11–14 ppm range, heptamers with a G-A sequence have two resonances with equal areas. [Note that the two resonances for rGCGAGCU are overlapped (Figure 3B), but in the presence of 10 mM $MgCl_2$ they are resolved and a third resonance appears at 11.0 ppm (Figure 3A).] In contrast, rGCAGGCG has three resonances with equal areas. Similarly, rGGCGAGCC and rCGCAGGCG have three and four resonances, respectively. These molecules are self-complementary and thus have two-fold symmetry. Therefore, each resonance arises from two protons. Since only hydrogen-bonded imino protons are normally observed in the 11–14 ppm region, this suggests sequences with A-G form duplexes with two more hydrogen bonds than sequences with G-A. Alternatively, exchange for the G-A sequence may be catalyzed unusually efficiently by an intrinsic catalyst (Guéron et al., 1987).

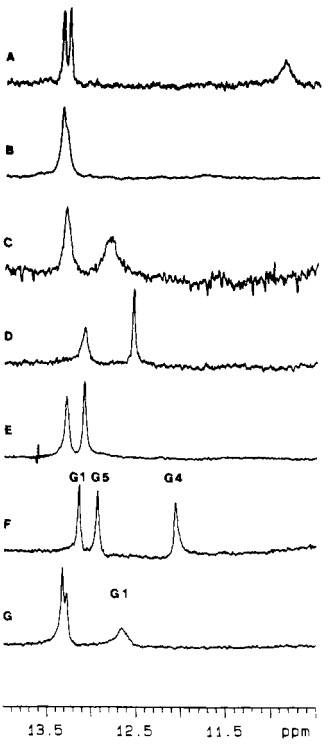

FIGURE 3: 500-MHz proton NMR spectra (10.5–14 ppm) in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA for (A) 0.8 mM rGCGAGCU at 9 °C, pH 6.4, with addition of 10 mM $MgCl_2$, (B) 1.3 mM rGCGAGCU at 5 °C, pH 6.4, without magnesium, (C) 0.2 mM rGCGAGC at 10 °C, pH 7, (D) 0.5 mM rGCGAGCA at 10 °C, pH 7, (E) 0.7 mM rGCGAGCG at 10 °C, pH 7, (F) 0.7 mM rGCAGGCG at 10 °C, pH 7, and (G) 2.0 mM rGGCGAGCC at 23 °C, pH 6.4. For all sequences, spectra were also measured at 1 °C or lower. For some cases, resonances broaden at lower temperatures, but no new resonances are observed for any of the sequences.

The missing imino proton resonance for sequences containing G-A internal loops could be rationalized if these duplexes were much less stable in the 0.1 M NaCl used for NMR studies relative to in the 1 M NaCl used for thermodynamic and CD studies. To test this possibility, thermodynamic parameters were measured for duplex formation by rGCGAGCG, rGCGAGCU, and rGGCGAGCC in 0.1 M NaCl. These results are listed in Table I. For each case, the melting temperatures are about 10 °C lower in 0.1 M NaCl relative to in 1.0 M NaCl. Nevertheless, at the strand concentrations used for the NMR experiments, the T_M's for rGCGAGCG, rGCGAGCU, and rGGCGAGCC are 39, 42, and 57 °C, respectively. These are about 30 °C higher than the temperatures of the NMR experiments shown in Figure 1. Spectra for all sequences were also measured at 1 °C or lower, and no new resonances were observed. Thus the missing imino proton resonance is not exchanging due to a rapid helix–coil equilibrium. CD spectra were also measured for rGCGAGCG, rGCGAGCU, and rGGCGAGCC in 0.1 M NaCl at 0 °C (data not shown) and were not significantly different than those in 1.0 M NaCl. Evidently, the structures of these molecules also do not change significantly between 0.1 and 1.0 M NaCl.

Assignments of the imino protons of rCGCAGGCG were made from 1-D NOE experiments (Figure 4). The resonance at 12.92 ppm was assigned to the end base pair (G8) because it is the first resonance to broaden when the temperature is raised above 5 °C. Irradiation of this resonance gives a weak NOE at 13.17 ppm. The resonance at 13.17 ppm is therefore

Table I: Thermodynamic Parameters of Duplex Formation[a]

RNA duplex	$1/T_M$ vs log C_T parameters				curve fit parameters			
	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M (°C)[d]	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M (°C)[d]
			1 M NaCl[b]					
$GC^{GA}GC$ $CG_{AG}CG$	4.01	30.37	85.00	20.1	4.17	28.57	78.68	21.5
$GC^{GA}GCA$ $ACG_{AG}CG$	6.79	51.51	144.22	43.8	6.87	54.79	154.50	43.9
$GC^{GA}GCG$ $GCG_{AG}CG$	6.68	55.08	156.05	42.8	6.73	55.91	158.56	43.0
$GC^{GA}GCU$ $UCG_{AG}CG$	5.85	48.48	137.47	38.1	5.85	52.76	151.25	38.0
$GC^{AA}GCG$ $GCG_{AA}CG$	4.20	42.83	124.55	26.7	4.28	41.38	119.61	26.9
$GC^{AG}GCG$ $GCG_{GA}CG$	6.33	61.27	177.13	40.3	6.34	60.43	174.40	40.4
$GGC^{GA}GCC$ $CCG_{AG}CGG$	9.68	66.08	181.83	57.0	10.16	73.93	205.64	57.0
$CGC^{AA}GCG$[e] $GCG_{AA}CGC$	5.44	48.02	137.30	35.5	5.59	40.11	111.30	36.4
$CGC^{AG}GCG$ $GCG_{GA}CGC$	7.75	60.36	169.61	48.1	7.90	65.45	185.57	47.9
			Reference Duplexes in 1 M NaCl[b]					
GCGC[f] CGCG	4.63	30.5	83.4	26.5	4.8	34.1	94.6	28.3
GCGCAp[f] pACGCG	7.92	50.6	137.6	51.8	7.90	49.4	133.8	52.0
GCGCGp[f] pGCGCG	7.67	46.9	126.5	50.7	7.8	48.9	132.6	50.7
GCGCUp[f] pUCGCG	6.93	46.6	127.9	45.4	6.78	44.4	121.3	45.6
GGCGCCp[f] pCCGCGG	11.35	67.8	182.0	65.2	11.0	63.4	169.0	65.4
CGCGCGp[f] pGCGCGC	9.09	54.5	146.4	57.9	9.1	53.1	142.0	58.1
			0.1 M NaCl[c]					
$GC^{GA}GCG$ $GCG_{AG}CG$	4.85	56.94	167.97	32.5	4.93	54.05	158.37	32.8
$GC^{GA}GCU$ $UCG_{AG}CG$	4.76	48.26	140.25	31.2	4.63	53.79	158.51	31.1
$GGC^{GA}GCC$ $CCG_{AG}CGG$	7.70	62.48	176.61	47.4	7.91	69.38	198.21	47.3

[a] Although estimated errors in $\Delta G°$, $\Delta H°$, and $\Delta S°$ are ±2%, ±10%, and ±10%, respectively, additional significant figures are given to allow accurate calculation of T_M and other parameters. [b] Solutions are 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. [c] Solutions are 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA, pH 7. [d] Calculated for 10^{-4} M oligomer concentration. [e] N. Sugimoto, R. Kierzek, and D. H. Turner, unpublished experiments. [f] Freier et al., 1986b.

assigned to G2 (Figure 4D). Irradiation of G2 (Figure 4E) gives rise to an NOE at 12.77 ppm, which is assigned to G6. Note that an NOE from G2 to G8 is not observed, probably because of chemical exchange with water. Irradiation of G6 (Figure 4C) shows an NOE to G2 and an NOE to a resonance at 11.98 ppm, which is assigned to G5. This assignment is confirmed by the NOE observed to G6 when G5 is irradiated (Figure 4B). Thus, the resonance at 11.98 ppm arises from the AG mismatches in the internal loop. Interestingly, when the G5 imino proton is irradiated, a strong NOE is observed to the non-exchangeable A4-H2 proton (see supplementary material). This confirms the assignments given above and indicates that the glycosidic torsion angles of both A4 and G5 are anti. The assignments for rGCAGGCG have also been determined by NOE difference spectroscopy (data not shown) and are shown in Figure 3F.

Resonances were not assigned for the G-A sequences. For the heptamers, both imino resonances broaden similarly as temperature is raised. For rGGCGAGCC, the resonance at 12.6 ppm broadens first as temperature is raised, suggesting it corresponds to G1. Unfortunately, there is no temperature where this resonance is sharp and the other two resonances are not overlapped.

Discussion

Formation of Watson–Crick and GU base pairs is considered to provide the dominant driving force for double-helix formation in RNA. Nearest-neighbor interactions between Watson–Crick base pairs, however, do not completely account for the observed stabilities of $(rGCGAGCG)_2$ and $(rGCAGGCG)_2$. With the two most commonly used sets of parameters for secondary structure prediction, $(rGCGAGCG)_2$ is predicted to form with a free energy change of −2.0 (Salser, 1977) or −4.7 kcal/mol (Freier et al., 1986a). The observed $\Delta G°_{37}$ is −6.7 kcal/mol. A similar disparity is observed for $(rGCAGGCG)_2$. Within 5 kcal/mol, secondary structures

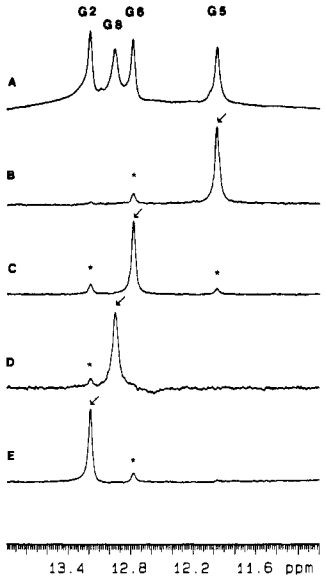

FIGURE 4: (A) 500-MHz proton NMR spectrum (11–14 ppm) of 1.5 mM rCGCAGGCG at 0.5 °C in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA in 90% H_2O, pH 6.4. Difference spectra following 3-s saturation of (B) the resonance at 11.98 ppm, (C) the resonance at 12.77 ppm, (D) the resonance at 12.92 ppm, and (E) the resonance at 13.17 ppm. Saturation power levels resulted in ~70% saturation of the desired resonance, except spectrum D, which resulted in only a ~30% saturation. The saturated resonance is designated by an arrow while the observed NOE's are designated by asterisks. Assignments are shown above spectrum A.

predicted for sequences of about 400 nucleotides often improve from 70 to 90% homology with known structures (Jaeger et al., 1989). Thus it is important to understand the origin of the stabilities of $(rGCGAGCG)_2$ and $(rGCAGGCG)_2$.

The similarities of $\Delta H°$'s derived from fitting melting curves and from T_M^{-1} versus log C_T plots indicate rGCGAGCG and the other sequences in Table I form duplexes. rGCGAGCG can form 2 possible duplexes:

$$\begin{array}{llll} 5' & GC^{GA}GCG & 3' & \\ 3' & GCG_{AG}CG & 5' & \end{array} \quad \text{or} \quad \begin{array}{llll} 5' & GCG^{AG}CG & 3' & \\ 3' & GC_{GA}GCG & 5' & \end{array}$$

From nearest-neighbor thermodynamic parameters (Freier et al., 1986a), the former is predicted to be 5.2 kcal/mol more stable than the latter at 37 °C. Thus the concentration of

$$\begin{array}{llll} 5' & GC^{GA}GCG & 3' \\ 3' & GCG_{AG}CG & 5' \end{array}$$

is expected to be almost 5000 times that of

$$\begin{array}{llll} 5' & GCG^{AG}CG & 3' \\ 3' & GC_{GA}GCG & 5' \end{array}$$

at 37 °C. Observations on sequence variants are consistent with the predicted dominance of

$$\begin{array}{llll} 5' & GC^{GA}GCG & 3' \\ 3' & GCG_{AG}CG & 5' \end{array}$$

For example, rGCGAGCA and rGCGAGCU can both form similar duplexes with 3′-dangling ends but cannot form alternate duplexes with 5′-dangling ends. CD and NMR spectra of the imino protons indicate these sequences form duplexes similar to rGCGAGCG. The CD spectra all have positive bands around 285 nm. The NMR spectra all have two imino proton resonances. The sequence rGGCGAGCC also has the characteristic CD spectrum (Figure 2C). Its NMR spectrum has one additional imino proton resonance, as expected from the additional GC pairs. Thus rGCGAGCG appears to form a duplex with two 3′-dangling end G's.

Table II: Thermodynamic Parameters of Loop and End Formation[a]

RNA duplex	loop parameters $\Delta G°_{37}$ (kcal/mol)	loop parameters $\Delta H°$ (kcal/mol)	loop parameters $\Delta S°$ (eu)	end parameters $\Delta G°_{37}$ (kcal/mol)	end parameters $\Delta H°$ (kcal/mol)	end parameters $\Delta S°$ (eu)
$GC^{GA}GC$ $CG_{AG}CG$	−1.4	−7.9	−21.0			
$GC^{GA}GCA$ $ACG_{AG}CG$	−0.9	−8.9	−26.0	−1.4	−10.6	−29.6
GCGCAp[b] pACGCG				−1.7	−10.1	−27.1
$GC^{GA}GCG$ $GCG_{AG}CG$	−1.0	−16.2	−49.0	−1.3	−12.4	−35.5
GCGCGp[b] pGCGCG				−1.5	−8.2	−21.6
$GC^{GA}GCU$ $UCG_{AG}CG$	−0.9	−9.9	−29.0	−0.9	−9.1	−26.2
GCGCUp[b] pUCGCG				−1.2	−8.1	−22.3
$GGC^{GA}GCC$ $CCG_{AG}CGG$	−0.3	−6.3	−19.2	−2.8	−17.9	−48.4
GGCGCC[b] CCGCGG				−3.4	−18.7	−49.3
$GC^{AA}GCG$ $GCG_{AA}CG$	1.5	−3.9	−17.5			
$GC^{AG}GCG$ $GCG_{GA}CG$	−0.7	−22.4	−70.0			
$CGC^{AA}GCG$[c] $GCG_{AA}CGC$	1.7	−1.5	−10.0			
$CGC^{AG}GCG$ $GCG_{GA}CGC$	−0.7	−13.9	−42.6			

[a] Solutions are 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. Parameters are based on results from T_M^{-1} versus log C_T plots (see Table I). [b] Freier et al., 1986b. [c] N. Sugimoto, R. Kierzek, & D. H. Turner, unpublished experiments.

The contribution of 3′-dangling end nucleotides to duplex stability in the presence of the G-A internal loop can be derived by comparing the stabilities of rGCGAGCX duplexes with $(rGCGAGC)_2$: $\Delta G°_{37,end} = 0.5[\Delta G°_{37}(rGCGAGCX) - \Delta G°_{37}(rGCGAGC)]$. These stability increments are listed in Table II along with those previously measured for corresponding rGCGCX duplexes (Freier et al., 1986b). For $(rGCGAGCX)_2$ where X = A, G, or U, the stability increments, $\Delta G°_{37,end}$, are −1.4, −1.3, and −0.9 kcal/mol of dangling end, respectively. These are similar to the values of −1.7, −1.5, and −1.2 kcal/mol, respectively, measured for the rGCGCX series (Freier et al., 1986b). Enthalpy and entropy increments for 3′-dangling ends can be calculated in a similar way and are also similar for duplexes with and without the G-A internal loop (see Table II). Thus, an internal loop with two GA mismatches does not have a strong effect on the interactions of the dangling ends. An opposite conclusion was recently reached for bulge loops (Longfellow et al., 1990). For example, a bulge loop of three nucleotides can reduce the free energy increment for 3′-dangling ends by 85%.

The dangling end effect explains −2.7 kcal/mol of the difference between the measured stability of $(rGCGAGCG)_2$ and that predicted from the parameters collected by Salser (1977). Salser's parameters include no favorable interactions from non-base-paired regions. On the other hand, the measured dangling end effect for the duplex is about 0.6 kcal/mol less favorable than that suggested by Freier et al. (1986a).

Thus the unusual stability of (rGCGAGCG)$_2$ relative to predictions based on these parameters must be due to the G-A internal loop.

Existing parameters for internal loops in RNA (Freier et al., 1986a; Turner et al., 1988) are based on measurements of oligomers containing internal loops with all CC (Gralla & Crothers, 1973) or all AA mismatches (A. Peritz, N. Sugimoto, R. Kierzek, and D. H. Turner, unpublished experiments). To determine if the poor predictions for (rGCGAGCG)$_2$ were a result of this limited data set, the sequence rGCAAGCG was studied. The measured $\Delta G°_{37}$ for (rGCAAGCG)$_2$ is −4.2 kcal/mol. This is close to the predicted value of −4.7 kcal/mol but is much less than the $\Delta G°_{37}$ of −6.7 kcal/mol measured for (rGCGAGCG)$_2$. Thus, it appears the sequence in the internal loop can have a major effect on stability. The CD spectra of rGCAAGCG and rGCGAGCG, however, are similar (Figure 2B), indicating the stacking is similar in both sequences. GA mismatches have many possibilities for formation of hydrogen bonds, whereas AA mismatches do not. This suggests hydrogen bonding may account for the greater stability of (rGCGAGCG)$_2$ compared with that of (rGCAAGCG)$_2$ (Turner et al., 1987b).

The stability increment attributed to the internal loop of (rGCGAGCG)$_2$ can be calculated from (Gralla & Crothers, 1973) $\Delta G°_{37,loop} = \Delta G°_{37}$(rGCGAGCG) − $\Delta G°_{37}$(rGCGCG) + $\Delta G°_{37}(^{CG}_{GC})$. $\Delta G°_{37}(^{CG}_{GC})$ is the free energy increment for the nearest-neighbor interaction (Freier et al., 1986a) interrupted by the internal loop. Other thermodynamic increments for the internal loop of rGCGAGCG and of other sequences can be calculated in an analogous fashion. These values are listed in Table II. The average free energy increments for G-A and A-A internal loops are −0.9 and +1.6 kcal/mol, respectively. Previous data for a symmetric internal loop with four C residues (Gralla & Crothers, 1973) gives a $\Delta G°_{37,loop}$ of +1.6 kcal/mol.

The large differences between stabilities of G-A and A-A internal loops is surprising. It is considerably larger than observed for single mismatches, XY, in the deoxy series

```
5'  CAAAXAAAG  3'
3'  GTTTYTTTC  5'
```

where an AA mismatch is only 0.5 and 0.2 kcal/mol less stable than the corresponding GA mismatches (Aboul-ela et al., 1985). The difference is also larger than observed for single mismatches at the ends of short RNA duplexes. For that case, AA is at most 0.5 kcal/mol less stable than GA for the sequences that have been measured (Turner et al., 1988). One DNA duplex with two consecutive GA mismatches, however, was observed to be unexpectedly stable (Wilson et al., 1988). Interestingly, in the secondary structures for *E. coli* 16S rRNA (Gutell et al., 1985) and 23S rRNA (Gutell & Fox, 1988), internal loops are closed by GA, AA, and CC mismatches 54, 25, and 5 times, respectively. Traub and Sussman (1982) noticed similar trends when all helix ends were counted. Evidently, the bias for GA mismatches may have a thermodynamic basis that is not mimicked by current approximations for internal loops (Freier et al., 1986a; Turner et al., 1988; Jaeger et al., 1989).

In the phylogenetically determined secondary structures of 21 16S rRNAs (Gutell et al., 1985) and 38 23S rRNAs (Gutell & Fox, 1988), the internal loop $^{5'\,GA\,3'}_{3'\,AG\,5'}$ is observed 45 times, $^{5'\,AA\,3'}_{3'\,AA\,5'}$ is observed 5 times, and $^{5'\,AG\,3'}_{3'\,GA\,5'}$ is never observed. Similarly, $^{5'\,GA\,3'}_{3'\,AA\,5'}$ is observed 39 times, but $^{5'\,AG\,3'}_{3'\,AA\,5'}$ is never observed. To determine if the order of the GA mismatches affects duplex stability, the sequences rGCAGGCG and rCGCAGGCG were studied. As shown in Table II, the $\Delta G°_{37}$ increments for the A-G internal loops ($\Delta G°_{37,loop}$) are about 0.4 kcal/mol less stable than those for the G-A internal loops but about 2 kcal/mol more stable than A-A internal loops. Evidently, the main thermodynamic difference between AA and GA mismatches is not dependent on the order of the GA mismatches. It would be surprising if the small thermodynamic difference between G-A and A-G was responsible for the disparity of occurrence in known secondary structures of RNA.

While the thermodynamic differences between G-A and A-G sequences are small, the structural differences are large. The CD spectra shown in Figure 2 suggest G-A and A-G sequences are structurally different. While CD spectra for duplexes with A-G internal loops are similar to those for A-form RNA (Tunis-Schneider & Maestre, 1970), CD spectra for duplexes with G-A internal loops are different from those previously measured for RNA. NMR spectra for duplexes with A-G internal loops have the number of resonances expected if each base pair and GA mismatch contained a slowly exchanging, hydrogen-bonded imino proton. This pattern is expected from previous NMR studies of GA mismatches in deoxyoligonucleotides (Gao & Patel, 1988; Kan et al., 1983; Orbons et al., 1987). For duplexes with G-A internal loops, however, only the number of resonances expected for the GC base pairs are observed in the 11–14 ppm region. Li et al. (1990) have recently observed a similar result for a G-A internal loop in DNA. This suggests the imino proton in the GA mismatch is exchanging rapidly on the NMR time scale. This implies either that the imino proton is not hydrogen bonded or that it is hydrogen bonded but with exchange catalyzed unusually efficiently by an intrinsic catalyst (Guéron et al., 1987). Either way, the results suggest evolution selected the G-A internal loop for structural reasons.

The effects reported here for internal loops containing two mismatches may also be present in larger internal and multibranch loops. For example, Longfellow et al. (1990) recently reported preliminary observations showing the sequence rGCGAAGCG associates to give an unexpectedly stable complex with only two imino proton resonances and with a CD spectrum resembling that of rGCGAGCG. Varani et al. (1989) did not observe imino proton resonances for GA mismatches in the internal loop

```
5'  GGAAU  3'
5'  CAUGA  3'
```

Traub and Sussman (1982) pointed out that helices in RNA secondary structures are often capped by a GA mismatch with A on the 5′-end and G on the 3′-end. The conserved sequences of the "hammerhead" ribozyme include a G-A sequence double GA mismatch capping a helix (Forster & Symons, 1987a,b). The available results suggest the structural features leading to selection of this motif may be found in the G-A internal loops studied here.

Conclusions

The results indicate rGCGAGCG forms a duplex

```
5'   GCGAGCG  3'
3'  GCGAGCG  5'
```

that is stabilized by dangling-end stacking and unusually strong interactions within the G-A internal loop. Surprisingly,

however, hydrogen bonding of the imino proton in the GA mismatch cannot be detected by NMR. An A-G internal loop has similar stability and the expected hydrogen bonding of the imino proton is observed by NMR. CD spectra suggest the G-A internal loop induces an unusual structure in the duplex, whereas A-G does not. The results indicate the sequence of an internal loop is important for determining both stability and structure. Thus, parameters reflecting such effects must be added to algorithms for predicting RNA structure from sequence.

ACKNOWLEDGMENTS

We thank Carl E. Longfellow for stimulating discussions and technical assistance. We also thank Adam E. Peritz for synthesizing blocked adenosine for oligonucleotide synthesis.

SUPPLEMENTARY MATERIAL AVAILABLE

One figure showing (A) the 500-MHz proton NMR spectrum (7–14 ppm) of rCGCAGGCG and (B) the NOE difference spectrum following irradiation of the resonance at 11.98 ppm (1 page). Ordering information is given on any current masthead page.

REFERENCES

Aboul-ela, F., Koh, D., Tinoco, I., Jr., & Martin, F. H. (1985) *Nucleic Acids Res. 13*, 4811–4824.

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.

Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843–853.

Brown, T., Hunter, W. H., Kneale, G., & Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2402–2406.

Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry Part II*: *Techniques for the Study of Biological Structure and Function*, Chapter 8, W. H. Freeman, San Francisco, CA.

Forster, A. C., & Symons, R. H. (1987a) *Cell 49*, 211–220.

Forster, A. C., & Symons, R. H. (1987b) *Cell 50*, 9–16.

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. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373–9377.

Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., & Turner, D. H. (1986b) *Biochemistry 25*, 3214–3219.

Gao, X., & Patel, D. J. (1988) *J. Am. Chem. Soc. 110*, 5178–5182.

Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301–319.

Guêron, M., Kochoyan, M., & Leroy, J.-L. (1987) *Nature 328*, 89–92.

Gutell, R. R., & Fox, G. E. (1988) *Nucleic Acids Res. 16*, (supplement), r175–r269.

Gutell, R. R., Weiser, B., Woese, C. R., & Noller, H. F. (1985) *Prog. Nucleic Acids Res. Mol. Biol. 32*, 155–216.

Hickey, D. R., & Turner, D. H. (1985) *Biochemistry 24*, 2086–2094.

Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.

Ikuta, S., Chattopadhyaya, R., & Dickerson, R. E. (1984) *Anal. Chem. 56*, 2253–2256.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706–7710.

Kan, L.-S., Chandrasegaran, 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.

Li, Y., Wilson, W. D., & Zon, G. (1990) *Biophys. J. 57*, 448a.

Longfellow, C. E., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 278–285.

Markiewicz, W. T., Biala, E., & Kierzek, R. (1984) *Bull. Pol. Acad. Sci. 32*, 433–450.

Marky, L. A. & Breslauer, K. J. (1987) *Biopolymers 26*, 1601–1620.

Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119–162.

Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1987) *Eur. J. Biochem. 170*, 225–239.

Papanicolaou, C., Gouy, M., & Ninio, J. (1984) *Nucleic Acids Res. 13*, 1717–1731.

Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 256–263

Prive, G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.

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.

Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol. 42*, 985–1002.

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.

Tunis-Schneider, M. J. B., & Maestre, M. F. (1970) *J. Mol. Biol. 52*, 521–541.

Turner, D. H., Sugimoto, N., Jaeger, J. A., Longfellow, C. E., Freier, S. M., & Kierzek, R. (1987a) *Cold Spring Harbor Symp. Quant. Biol. 52*, 123–133.

Turner, D. H., Sugimoto, N., Kierzek, R., & Dreiker, S. D. (1987b) *J. Am. Chem. Soc. 109*, 3783–3785.

Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167–192.

Traub, W., & Sussman, J. L. (1982) *Nucleic Acids Res. 10*, 2701–2708.

Varani, G., Wimberly, B., & Tinoco, I., Jr. (1989) *Biochemistry 28*, 7760–7772.

Wilson, W. D., Hoa, D. T. M., Zuo, E. T., & Zon, G. (1988) *Nucleic Acids Res. 16*, 5137–5151.

Woese, C. R. (1983) *Microbiol. Rev. 51*, 221–271.