

Villafranca, J. J., Rhee, S. G., & Chock, P. B. (1978) *Proc. Natl. Acad. Sci. U.S.A.* 75, 1255-1259.
 Williams, K. R., & Konigsberg, W. H. (1978) *J. Biol. Chem.* 253, 2463-2470.

Williams, K. R., LoPresti, M. B., & Setoguchi, M. (1981) *J. Biol. Chem.* 256, 1754-1762.
 Wu, F. Y.-H., & Tyagi, S. C. (1987) *J. Biol. Chem.* 262, 13147-13154.

Stabilities of Consecutive A·C, C·C, G·G, U·C, and U·U Mismatches in RNA Internal Loops: Evidence for Stable Hydrogen-Bonded U·U and C·C+ Pairs[†]

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 12, 1991; Revised Manuscript Received May 24, 1991

ABSTRACT: The stability and structure of RNA duplexes with consecutive A·C, C·A, C·C, G·G, U·C, C·U, and U·U mismatches were studied by UV melting, CD, and NMR. The results are compared to previous results for GA and AA internal loops [SantaLucia, J., Kierzek, R., & Turner, D. H. (1990) *Biochemistry* 29, 8813-8819; Peritz, A., Kierzek, R., & Turner, D. H. (1991) *Biochemistry* 30, 6428-6436]. The observed order for stability increments of internal loop formation at pH 7 is $AG = GA \approx UU > GG \geq CA \geq AA = CU = UC \geq CC \geq AC$. The results suggest two classes for internal loops with consecutive mismatches: (1) loops that stabilize duplexes and have strong hydrogen bonding and (2) loops that destabilize duplexes and may not have strong hydrogen bonding. Surprisingly, rCGCUUGCG forms a very stable duplex at pH 7 in 1 M NaCl with a T_M of 44.8 °C at 1×10^{-4} M and a ΔG°_{37} of -7.2 kcal/mol. NOE studies of the imino protons indicate hydrogen bonding within the U·U mismatches in a wobble-type structure. Resonances corresponding to the hydrogen-bonded uridines are located at 11.3 and 10.4 ppm. At neutral pH, rCGCCCGCG is one of the least stable duplexes with a T_M of 33.2 °C and ΔG°_{37} of -5.1 kcal/mol. Upon lowering the pH to 5.5, however, the T_M increases by 12 °C, and ΔG°_{37} becomes more favorable by 2.5 kcal/mol. The pH dependence of rCGCCCGCG may be due to protonation of the internal loop C's, since no changes in thermodynamic parameters are observed for rCGCUUGCG between pH 7 and 5.5. Furthermore, two broad imino proton resonances are observed at 10.85 and 10.05 ppm for rCGCCCGCG at pH 5.3, but not at pH 6.5. This is also consistent with C·C+ base pairs forming at pH 5.5. rCGCCAGCG and rGGCACGCC have a small pH dependence, with T_M increases of 5 and 3 °C, respectively, upon lowering the pH from 7 to 5.5. rCGCCUGCG and rCGCUCGCG also show little pH dependence, with T_M increases of 0.8 and 1.4 °C, respectively, upon lowering the pH to 5.5. CD spectra of sequences with CC, CU, UC, and UU internal loops are typical of A-form conformation. CD spectra of AC, CA, and GG have a positive band at 280 nm, similar to that observed for GA and AA internal loops (SantaLucia et al., 1990). CD spectra of all sequences studied, except rCGCCCGCG, are independent of pH. For rCGCCCGCG, a weak negative band at 300 nm is observed at pH 7, but at pH 5.5 a weak positive band is observed. Current algorithms for the prediction of RNA secondary structure assume that the stability of internal loops is not sequence-dependent or depends only on stacking. This study indicates these approximations are wrong. The measured internal loop free energy increments range from -0.6 to +2.3 kcal/mol, and do not correlate with known stacking parameters.

Prediction of RNA structure from sequence has become an increasingly important goal due to the rapid accumulation of nucleic acid sequence information. Applications include identification of consensus structures (Konings & Hogeweg, 1989; Le et al., 1988), understanding of structure-function relationships (de Smit & van Duin, 1990; Yager & von Hippel, 1991; Le et al., 1989), and design of antisense agents (Wickstrom et al., 1988). The most successful methods for prediction of secondary structure have been based on energy

minimization using free energy increments for various RNA motifs (Tinoco et al., 1971; Papanicolaou et al., 1984; Turner et al., 1988; Zuker, 1989; Jaeger et al., 1989). The sequence dependence of free energy increments for Watson-Crick base pairs has been determined experimentally (Borer et al., 1974; Freier et al., 1986a). Little is known, however, about the sequence dependence of loop motifs (Gralla & Crothers, 1973a,b; Groebe & Uhlenbeck, 1988; Tuerk et al., 1988; Turner et al., 1988; Varani et al., 1989; SantaLucia et al., 1990).

A common loop motif is the internal loop in which a helix is interrupted by non-Watson-Crick-paired nucleotides on both strands. It has recently been shown that an internal loop containing two consecutive G·A mismatches is more than 2 kcal/mol more stable than one containing two consecutive A·A mismatches (SantaLucia et al., 1990) and that this additional

[†]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. J.S.L. is an E. H. Hooker Fellow.

* Author to whom correspondence should be addressed.

[‡]University of Rochester.

[§]Polish Academy of Sciences.

stability is due to hydrogen bonding (SantaLucia et al., 1991). While most predictions of known RNA secondary structures are only about 70% correct, a structure that is 90% correct is often found within about 5 kcal/mol of the predicted structure (Jaeger et al., 1989; Zuker et al., 1991). Thus, it is important to know the sequence dependence of internal loop stability.

In this paper, we report the sequence dependence of internal loop stability for symmetrical loops with two consecutive non-G·U mismatches. The results suggest two classes of these internal loops. The internal loop stabilizes the duplex when the mismatches are G·A, U·U, or C·C+, and destabilizes it when the mismatches are G·G, C·A, C·U, A·A, or C·C. Imino proton nuclear magnetic resonance (NMR)¹ spectra suggest the stability of U·U mismatches is due to hydrogen bonding involving imino protons. No simple model predicts all the relative stabilities observed.

MATERIALS AND METHODS

RNA Synthesis and Purification. Oligoribonucleotides were synthesized on solid support with a modification of the method of Usman et al. (1987) with the 2'-hydroxyl protected as the *tert*-butyl dimethylsilyl ether. Upon completion of the coupling reactions, the phosphate methyl groups were removed by treatment with thiophenol. The oligomer was removed from solid support, and base blocking groups were removed by treatment with concentrated ammonia in ethanol (3:1 v/v) at 50 °C overnight. The silyl protection was removed by treatment with 1.0 M triethylammonium hydrogen fluoride (50 equiv) in pyridine at 50 °C for 48 h. The crude reaction mixture was then dried and partitioned between water and diethyl ether. The water layer was passed over a Sep-pak C-18 cartridge (Waters) to desalt. The oligomer was further 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). Dimethoxytrityl was removed by treating with 80% acetic acid for 30 min. Oligomers were desalted and further purified with a Sep-pak C-18 cartridge. 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 Na₂EDTA, pH 7. Solutions at pH 5.5 were 1.0 M NaCl, 50 mM MES, and 0.5 mM Na₂EDTA. Single-strand extinction coefficients were calculated from extinction coefficients for dinucleoside monophosphates and nucleotides, as described previously (Borer, 1975; Richards, 1975). In units of 10⁴ M⁻¹ cm⁻¹, the calculated extinction coefficients at 280 nm are as follows: rCGCCAGCG, 3.94; rCGCCCGCG, 4.00; rCGCCUGCG, 3.85; rCGCUCGCG, 3.77; rCGCUUGCG, 3.66; rGGCACGCC, 4.02; rUGCGGGCA, 3.82. Strand concentrations were determined from the high-temperature absorbance at 280 nm. Absorbance vs 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).

Data Analysis. Absorbance vs temperature profiles were fit to a two-state model with sloping base lines 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}) vs the logarithm of the strand concentration ($\log C_T$) were fit to (Borer et al., 1974)

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

Errors in thermodynamic parameters are standard deviations and were calculated as described by SantaLucia et al. (1991). We note that a factor of 2 was omitted preceding σ_{bm}^2 in eq 4 of SantaLucia et al. (1991).

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 and Schimmel (1980).

NMR Spectroscopy. ¹H and ³¹P NMR spectra were measured with a Varian VXR-500S spectrometer at 500 and 202.3 MHz, respectively. ¹H and ³¹P chemical shifts are referenced to internal TSP and phosphate buffer, respectively. Oligomers were dissolved in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na₂EDTA at pH 6.5, unless otherwise noted. Higher NaCl concentrations were not used due to limited solubility of oligomers at higher salt. Lower pH samples were obtained by titrating with 1 M HCl. The solvent for imino proton studies was 10% D₂O and 90% H₂O. 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 1:3:3:1 pulse sequence (Hore, 1983). Spectral base lines were corrected with the standard routine in the VNMR software. 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 with the 1:3:3:1 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 1 s was used to achieve a steady state.

RESULTS

Thermodynamic Parameters at pH 7. For most internal loops, stability was determined in the context rCGCXYGCG, where the internal loop is underlined. For GG and AC, the contexts were rUGCGGGCA and rGGCACGCC, respectively, to minimize competition with alternate duplexes. Except for rUGCGGGCA, UV melting profiles are monophasic. Normalized melting curves for rUGCGGGCA and rCGCCCGCG are shown in Figure 1. Plots of T_M^{-1} versus $\log C_T$ are shown in Figure 2. Thermodynamic parameters, derived by assuming two-state duplex formation, are listed in Table I. The results for AC, CA, CC, and UU sequences at pH 7 fit this model at all concentrations. In particular, the melting curves are fit well, the T_M^{-1} versus $\log C_T$ plots are linear, and the ΔH° 's derived from fits and $\log C_T$ plots agree within 10%. Deviations from the model, however, are observed for GG, UC, and CU loops. These are discussed below.

For rUGCGGGCA above 4.8×10^{-5} M, a second transition near 70 °C is observed (see Figure 1). Thus, (rUGCGGGCA)₂ does not melt to a random coil above 4.8×10^{-5} M. For curves exhibiting the second transition, the data were truncated at the start of the second transition and fit with the two-state algorithm. Nevertheless, the T_M^{-1} versus $\log C_T$ plot for rUGCGGGCA is also unusual at high concentrations because the T_M becomes concentration-independent (see Figure 2). To avoid this complication, the results listed

¹ Abbreviations: CD, circular dichroism; MES, 2-(*N*-morpholino)-ethanesulfonic acid; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; Na₂EDTA, disodium ethylenediaminetetraacetate.

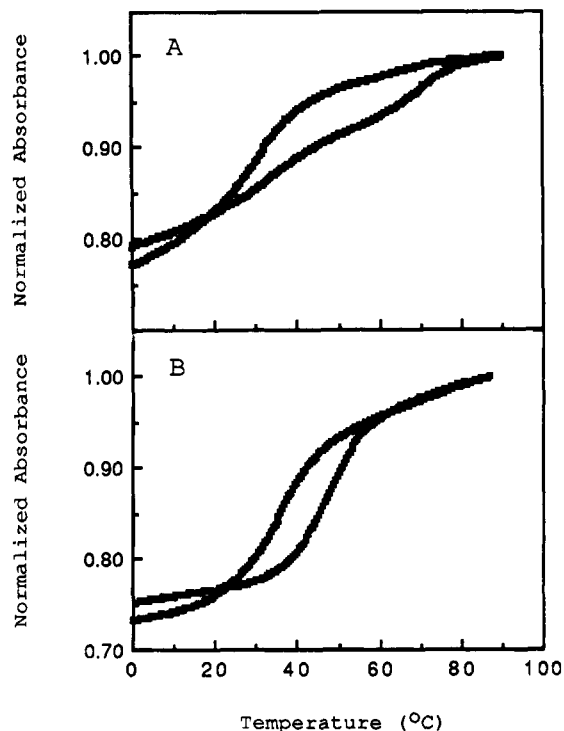


FIGURE 1: Normalized absorbance curves for (A) 2.7×10^{-5} M (monophasic) and 4.8×10^{-4} M (biphasic) rUGCGGGCA and (B) 1.1×10^{-4} M rCGCCCGCG at pH 7 (lower T_M) and 1.3×10^{-4} M rCGCCCGCG at pH 5.5 (higher T_M).

for rUGCGGGCA in Table I are derived only from data at concentrations less than 4.8×10^{-5} M. The transition at 70 °C is, presumably, not due to hairpin formation since such an intramolecular species should also be observed at low concentrations. The results suggest rUGCGGGCA forms a complex with more than two strands at high concentration. Since rUGCGGGCA contains three consecutive internal G residues, it is possible that the higher melting transition corresponds to "G-quartet" formation (Williamson et al., 1989; Sundquist & Klug, 1989).

An apparent flattening of the T_M^{-1} versus $\log C_T$ plot is also observed for rCGCUCGCG above 3.9×10^{-5} M (see Figure 2). The deviations from a straight line, however, are close to experimental error. To assure no interference from a potential higher order structure, the results in Table I are derived only from data at concentrations less than 4×10^{-5} M. Thermodynamic parameters are the same within experimental error when all points are used (see footnote to Table I).

For rCGCCUGCG, deviations from a straight line in T_M^{-1} versus $\log C_T$ plots are also observed. If only concentrations below 5×10^{-5} M are included, the results for ΔH° from the T_M^{-1} versus $\log C_T$ plot and fitted data agree within 15% (see Table I), consistent with a two-state, bimolecular transition (SantaLucia et al., 1990; Marky & Breslauer, 1987). For concentrations above 5×10^{-5} M, however, the T_M^{-1} versus $\log C_T$ plot is nonlinear, suggesting a breakdown of the bimolecular two-state approximation at high concentrations (SantaLucia et al., 1990; Marky & Breslauer, 1987).

CD Spectra at pH 7. CD spectra for several sequences at pH 7 are shown in Figure 3. The spectra for CC, CU, UC, and UU are similar to those observed for AG sequences by SantaLucia et al. (1990) and are typical of the A-form conformation (Tunis-Schneider & Maestre, 1970). On the other hand, AC, CA, and GG have a positive band at 280 nm, similar to that observed by SantaLucia et al. (1990) for GA

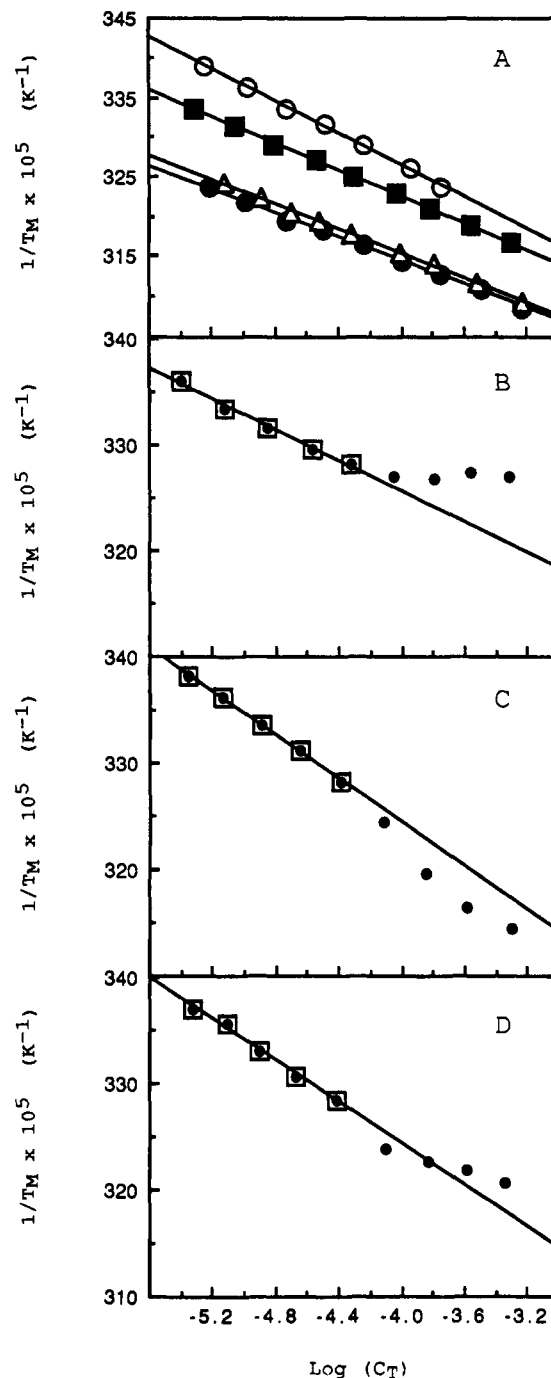


FIGURE 2: Reciprocal melting temperature vs log concentration plots for (A) rCGCCAGCG (■), rCGCCCGCG (○), rCGCUUGCG (●), and rGGCAGGCC (Δ), (B) rUGCGGGCA (●), (C) rCGCCUGCG (●), and (D) rCGCUCGCG (●). For panels B–D, boxed points are included in linear regression. Solutions are 1.0 M NaCl, 0.01 M sodium cacodylate, and 0.5 mM Na₂EDTA pH 7.

and AA sequences.

Imino Proton NMR Spectra at pH 6.5. NMR spectra for the imino protons of the duplexes studied and for the related duplexes (rGCAGGCG)₂ and (rGGCGAGCC)₂ (SantaLucia et al., 1990) are shown in Figure 4. Since the duplexes are self-complementary with three different Watson–Crick base pairs, three resonances are expected in the base-pairing region from 12 to 15 ppm (Blommers et al., 1989). At pH 6.5, three separate resonances are observed for the sequences containing CA, AC, UC, and GA internal loops. For CC, two resonances are observed, but the resonance at 13.3 ppm has twice the integrated area of the resonance at

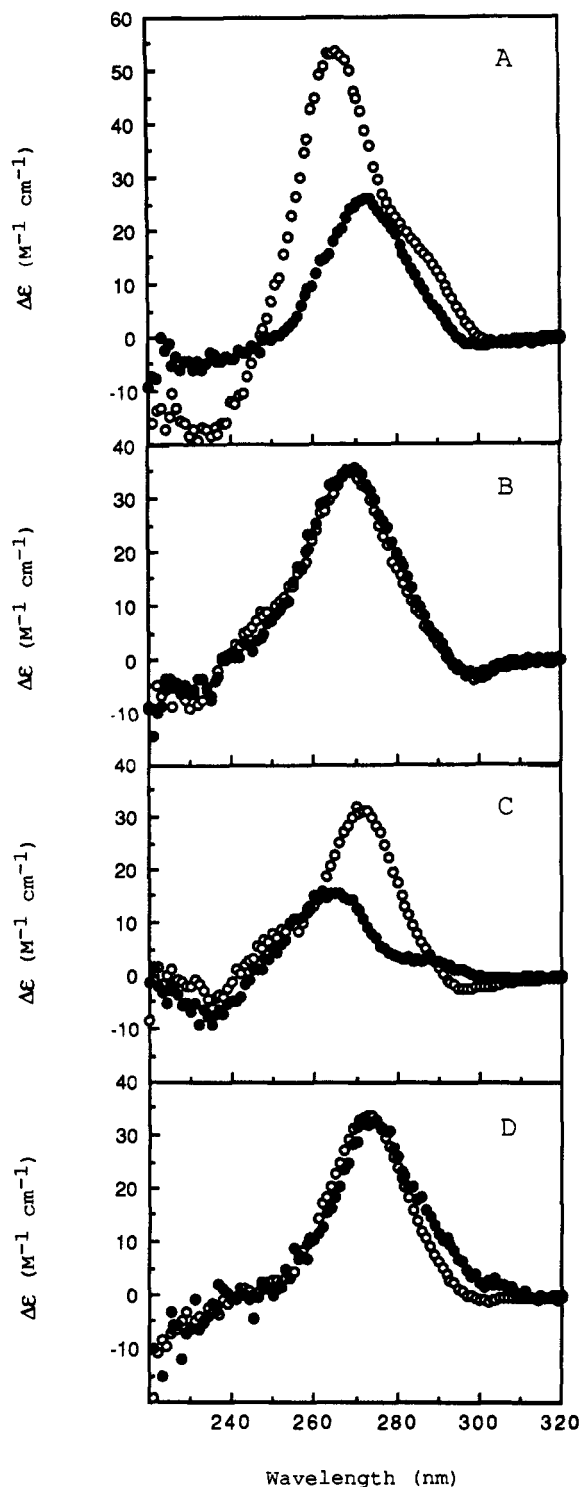


FIGURE 3: CD spectra at 0 °C for (A) 6.3×10^{-5} M rGGCACGCC (○) and 7.1×10^{-5} M rCGCCAGCG (●), (B) 5.3×10^{-5} M rCGCCUUGCG (○) and rCGCUUGCG (●), (C) 8.4×10^{-5} M rCGCUUGCG (○) and 6.8×10^{-5} M rUGCGGGCA (●), and (D) 7.5×10^{-5} M rCGCCCGCG at pH 7 (○) and 3.8×10^{-5} M rCGCCCGCG at pH 5.5 (●) (solution is 1.0 M NaCl, 50 mM MES, and 0.5 mM Na₂EDTA). All solutions except in (D) are 1.0 M NaCl, 0.01 M sodium cacodylate, and 0.5 mM Na₂EDTA pH 7.

12.9 ppm, suggesting a total of three protons. The GA sequence has an additional broad resonance with a line width of 130 Hz at 10.1 ppm (SantaLucia et al., 1991), normally outside the region for hydrogen-bonded imino protons (Blommers et al., 1989). In contrast, rCGCACGCC has four resonances with the resonance at 12.0 ppm assigned to a hydrogen-bonded imino proton in the G-A mismatch (SantaLucia

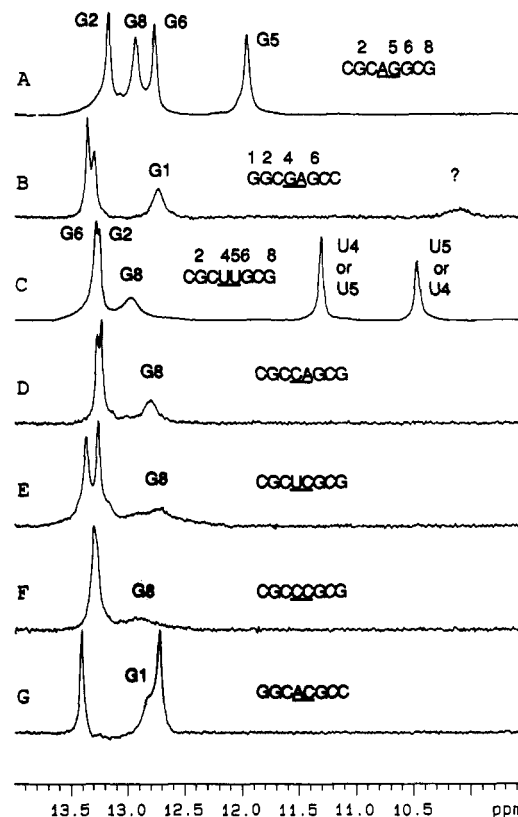


FIGURE 4: 500-MHz proton NMR spectra (9.5–14.0 ppm) in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na₂EDTA in 90% H₂O/10% D₂O, pH 6.5, for (A) 1.5 mM rCGCACGCC at 0.5 °C (SantaLucia et al., 1990), (B) 0.7 mM rGGCAGGCC at 20 °C (at lower temperatures, resonances broaden due to aggregation), (C) 1.0 mM rCGCUUGCG at 5 °C, (D) 1.0 mM rCGCCAGCG at 10 °C, (E) 1.7 mM rCGCUUGCG at 10 °C, (F) 0.8 mM rCGCCCGCG at 10 °C, and (G) 0.2 mM rGGCAGGCC at 10 °C.

et al., 1990). Only rCGCUUGCG has five sharp resonances (line width = 17–19 Hz) between 10 and 14 ppm. This suggests the U·U mismatches are hydrogen-bonded.

The resonances for CGCUUGCG were assigned with the 1-D NOE experiments shown in Figure 5. The resonance assignments consistent with all the data are shown in Figure 5A. The resonance at 12.9 ppm was assigned to the terminal base pair (G8) because it is the first resonance to broaden when the temperature is raised above 5 °C. For the related sequence, rCGCACGCC, the resonance at 12.9 ppm is also first to broaden, consistent with this assignment (SantaLucia et al., 1990). Irradiation of the resonance at 11.3 ppm gives rise to a strong NOE (17%) to the resonance at 10.4 ppm and a weak NOE (6%) to the resonance at 13.3 ppm. Irradiation of the resonance at 10.4 ppm gives a strong NOE (28%) to the resonance at 11.3 ppm and a weak NOE (7%) to the resonance at 13.3 ppm. The strong NOE's indicate the protons giving rise to the resonances at 11.3 and 10.4 ppm are less than 0.3 nm apart. This suggests hydrogen bonding within each U·U mismatch as shown in the potential structures of Figure 6. The NMR results are consistent with formation of either structure or a rapid interconversion between forms. The weak NOE's indicate the resonance at 13.3 ppm is from G6. The resonance at 13.2 ppm is assigned to G2 by elimination. The proton-decoupled ³¹P NMR spectrum of rCGCUUGCG is shown in Figure 7. All phosphorous nuclei resonate within a 1 ppm range; this suggests no large backbone distortion is required to accommodate the internal loop (Gorenstein, 1984). Similar observations and conclusions have been made for a

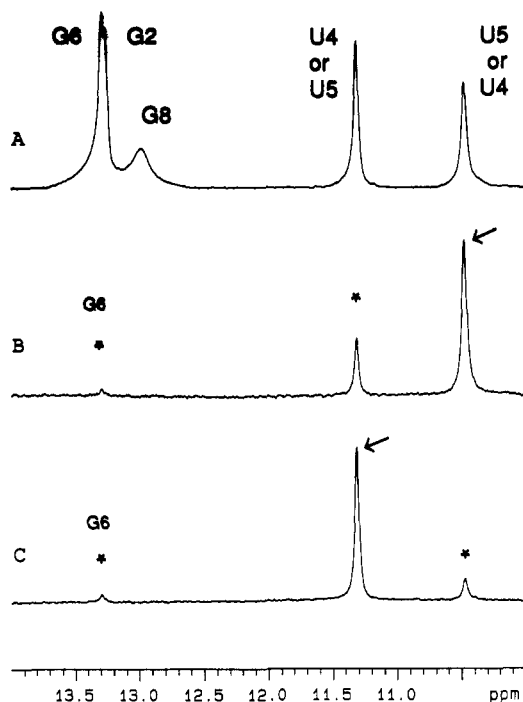


FIGURE 5: (A) 500-MHz proton NMR spectrum (10–14 ppm) of 1.0 mM rCGCUUGCG at 5 °C in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na₂EDTA in 90% H₂O, pH 6.5. Difference spectra following 1.0-s saturation of (B) the resonance at 10.4 ppm and (C) the resonance at 11.3 ppm. Decoupler power levels resulted in 80% saturation of the desired resonance. The saturated resonance is designated by an arrow while the observed NOE's are designated by asterisks. Assignments are shown above spectrum A.

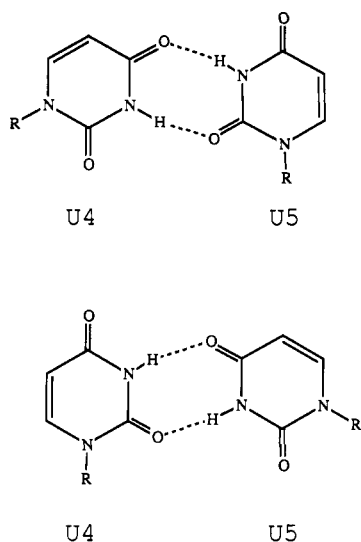


FIGURE 6: U-U wobble mismatch structures consistent with NMR data for rCGCUUGCG.

DNA oligonucleotide with a T·T pair (Kouchakdjian et al., 1988).

Thermodynamic Parameters at pH 5.5. Thermodynamic parameters for oligomers at pH 5.5 are listed in parentheses in Table I. Thermodynamic parameters for the sequences with GA, UU, GG, and UC internal loops change little between pH 7.0 and 5.5. For CU, the T_M^{-1} versus $\log C_T$ plot for CU at pH 5.5 is linear over a larger concentration range than at pH 7, and the ΔH° 's from these plots are different at pH 7.0 and 5.5. There is little change in other parameters, however. Large changes in thermodynamic parameters are observed for rCGCCCGCG, upon changing pH. The T_M increases by 12

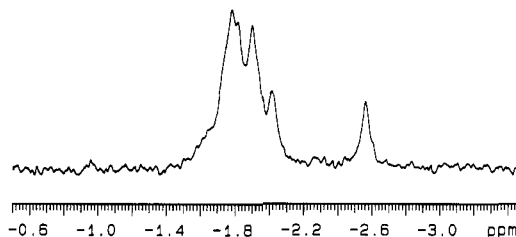


FIGURE 7: Proton-decoupled ³¹P NMR spectrum (−0.5 to −3.5 ppm) of 1.0 mM rCGCUUGCG at 20 °C. 3.0-Hz line broadening was applied before Fourier transformation.

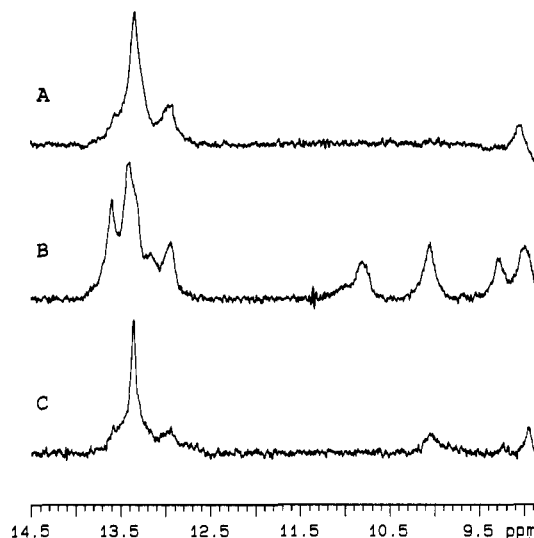


FIGURE 8: 500-MHz proton NMR spectra (8.8–14.5 ppm) of 0.8 mM rCGCCCGCG in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na₂EDTA at (A) 0.5 °C, pH 6.5, (B) 0.5 °C, pH 5.3, and (C) 10 °C, pH 5.3. The spectrum at 10 °C, pH 6.5, is shown in Figure 4 (F).

°C, and the ΔG°_{37} becomes more favorable by almost 2.5 kcal/mol when the pH is lowered from 7.0 to 5.5. rGGCACGCC and rCGCCAGCG show a small dependence on pH. Between pH 7.0 and 5.5, the T_M 's increase by 3 and 5 °C, and ΔG°_{37} is more favorable by 0.8 and 0.9 kcal/mol, respectively.

CD Spectra at pH 5.5. No changes in the CD spectra of AC, CA, CU, UC, GG, and UU were observed upon lowering the pH to 5.5 (data not shown). For rCGCCCGCG, a weak negative band at 300 nm is observed at pH 7, but at pH 5.5, a weak positive band is observed (see Figure 3). This is consistent with the formation of C·C+ base pairs in DNA (Gray et al., 1984; Edwards et al., 1990) and RNA (Brahms et al., 1967; Guschlbauer 1975).

Imino Proton NMR Spectra at pH 5–5.5. NMR spectra for some sequences are also pH-dependent. For rCGCCCGCG at pH 6.5 and 0.5 °C, no resonances are observed from 9 to 11 ppm. Upon lowering the pH to 5.3, however, new resonances are observed at 10.85, 10.05, and 9.2 ppm (Figure 8). The circa 80-Hz line widths for these resonances indicate chemical exchange in the intermediate time regime. The presence of more than three resonances between 12 and 14 ppm is also consistent with chemical exchange. Upon raising the temperature to 10 °C, the resonances below 11 ppm broaden significantly or disappear, suggesting exchange with solvent.

Upon lowering the pH to 5.3 for the sequence rCGCCUGCG, new resonances at 11.4, 9.6, and 9.2 ppm are observed (see Figure 9). Upon lowering the pH for rCGCUCGCG, the resonances between 12 and 14 ppm

Table I: Thermodynamic Parameters of Duplex Formation for Helices with Internal Loops^{a,b}

| | 1/T _M vs log C _T parameters | | | | curve fit parameters | | |
|---|---|----------------------------|--------------------------------|----------------------------------|----------------------------------|----------------------------|--------------------------------|
| | -ΔG° ₃₇ (kcal/mol) | -ΔH° (kcal/mol) | -ΔS° (eu) | T _M (°C) ^c | -ΔG° ₃₇ (kcal/mol) | -ΔH° (kcal/mol) | -ΔS° (eu) |
| RNA duplex | | | | | | | |
| GC ^{UO} GCG ^c GCG ^{GU} CG | 9.58 ± 0.12 | 67.6 ± 1.3 | 187.1 ± 3.8 | 56.0 | 9.80 ± 0.13 | 72.1 ± 2.2 | 201.0 ± 6.8 |
| CGC ^{AG} GCG ^d GCG ^{GA} CGC | 7.75 ± 0.22 | 60.3 ± 2.6 | 169.4 ± 7.9 | 48.1 | 7.90 ± 0.13 | 65.5 ± 2.8 | 185.6 ± 8.9 |
| GGC ^{GU} GCC ^e CCG ^{UG} CGG | 9.71 ± 0.11 | 73.8 ± 1.3 | 206.5 ± 3.8 | 55.0 | 9.75 ± 0.09 | 74.3 ± 1.8 | 208.0 ± 5.6 |
| GGC ^{GA} GCC ^d CCG ^{AG} CGG | 9.67 ± 0.24 (9.26 ± 0.27) | 66.0 ± 2.4 (61.0 ± 2.4) | 181.6 ± 7.2 (166.9 ± 7.1) | 57.0 (56.3) | 10.16 ± 0.30 (9.91 ± 0.34) | 73.9 ± 3.9 (72.6 ± 3.7) | 205.6 ± 11.7 (202.2 ± 11.0) |
| CGC ^{UU} GCG GCG ^{UU} CGC | 7.19 ± 0.09 (7.25 ± 0.18) | 61.3 ± 1.1 (64.3 ± 2.1) | 174.5 ± 3.5 (183.8 ± 6.3) | 44.8 (44.8) | 7.26 ± 0.11 (7.25 ± 0.04) | 63.0 ± 3.5 (62.8 ± 2.8) | 179.7 ± 11.1 (179.2 ± 8.9) |
| UGC ^{GO} GCA ^f ACG ^{GO} CGU | 5.04 ± 0.32 (4.93 ± 1.03) | 63.7 ± 3.6 (59.8 ± 9.0) | 189.2 ± 11.5 (176.9 ± 27.9) | 33.9 (33.2) | 5.37 ± 0.12 (5.19 ± 0.08) | 52.3 ± 2.2 (50.6 ± 1.4) | 151.4 ± 7.3 (146.6 ± 4.5) |
| CGC ^{CA} GCG GCG ^{AC} CGC | 5.67 ± 0.03 (6.54 ± 0.38) | 54.6 ± 0.4 (55.3 ± 3.9) | 157.7 ± 1.4 (157.1 ± 12.2) | 36.9 (41.9) | 5.70 ± 0.05 (6.59 ± 0.10) | 54.2 ± 2.5 (53.6 ± 5.5) | 156.4 ± 8.3 (151.5 ± 17.7) |
| CGC ^{AA} GCG ^g GCG ^{AA} CGC | 5.43 ± 0.07 | 48.0 ± 1.1 | 137.1 ± 3.5 | 35.5 | 5.59 ± 0.22 | 40.1 ± 4.1 | 111.3 ± 13.7 |
| CGC ^{CU} GCG ^h GCG ^{UC} CGC | 5.41 ± 0.11 (5.47 ± 0.22) | 44.4 ± 0.9 (60.6 ± 2.7) | 125.6 ± 2.8 (177.7 ± 8.4) | 35.1 (35.9) | 5.20 ± 0.07 (5.55 ± 0.09) | 51.2 ± 3.7 (56.0 ± 1.7) | 148.5 ± 12.1 (162.7 ± 5.6) |
| CGC ^{UC} GCG ^h GCG ^{CU} CGC | 5.41 ± 0.24 (5.51 ± 0.20) | 46.6 ± 2.0 (51.8 ± 2.1) | 132.9 ± 6.2 (149.3 ± 6.7) | 34.6 (36.0) | 5.36 ± 0.09 (5.53 ± 0.04) | 48.9 ± 4.3 (54.3 ± 4.2) | 140.3 ± 14.0 (157.3 ± 13.7) |
| CGC ^{CC} GCG GCG ^{CC} CGC | 5.11 ± 0.07 (7.59 ± 0.54) | 45.6 ± 0.7 (74.6 ± 6.9) | 130.5 ± 2.3 (216.2 ± 21.4) | 33.2 (45.2) | 5.00 ± 0.06 (7.37 ± 0.17) | 50.7 ± 2.5 (61.8 ± 2.7) | 147.4 ± 7.9 (175.5 ± 7.5) |
| GGC ^{AC} GCC CCG ^{CA} CGG | 6.98 ± 0.07 (7.75 ± 0.30) | 58.8 ± 0.9 (64.2 ± 3.3) | 167.1 ± 2.7 (182.1 ± 10.1) | 44.0 (47.3) | 7.08 ± 0.14 (7.74 ± 0.08) | 61.9 ± 3.7 (63.3 ± 2.4) | 176.6 ± 11.7 (179.1 ± 7.5) |
| reference duplexes | | | | | | | |
| CGCGCG ^p pGCGCGC | 9.11 ± 0.05 | 54.4 ± 0.5 | 146.2 ± 1.6 | 57.9 | 9.06 ± 0.14 | 53.1 ± 2.0 | 142.1 ± 6.1 |
| GGCGCC ^p pCCGCGG | 11.31 ± 0.05 | 67.7 ± 0.5 | 181.8 ± 1.4 | 65.2 | 10.99 ± 0.12 | 63.4 ± 1.8 | 169.1 ± 5.4 |

^aListed in order of decreasing loop free energy (see Table II). ^bSolutions are 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na₂EDTA, pH 7. Values in parentheses were measured in 1.0 M NaCl, 50 mM MES, and 0.5 mM Na₂EDTA, pH 5.5. ^cCalculated for 10⁻⁴ M oligomer concentration. L. He, J. SantaLucia, A. Walter, R. Kierzek, and D. Turner, unpublished results. ^dSantaLucia et al. (1990). ^eParameters for concentrations below 4.8 × 10⁻⁵ M (see Figure 2). ^fPeritz et al. (1990). ^gParameters for concentrations below 4.1 × 10⁻⁵ M (see Figure 2). If all points are included, the 1/T_M vs log C_T parameters are ΔG°₃₇ = -5.68 kcal/mol, ΔH° = -37.1 kcal/mol, and ΔS° = -101.5 eu. ^hParameters for concentrations below 4.0 × 10⁻⁵ M. Parameters do not change, within experimental error, when all points are included. ⁱFreier et al. (1986b).

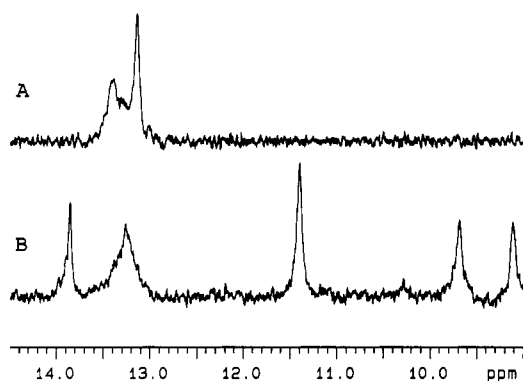


FIGURE 9: 500-MHz proton NMR spectra (9.0–14.5 ppm) of 0.9 mM rCGCCUGCG in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM Na₂EDTA at (A) 10 °C, pH 6.5, and (B) 0.5 °C, pH 5.0.

broaden significantly, and no new resonances are observed below 12 ppm (data not shown). Sequences with AC and CA internal loops do not show extra resonances upon lowering the pH and temperature (data not shown).

DISCUSSION

Thermodynamic parameters for formation of internal loops with two consecutive mismatches can be derived from the data

Table II: Thermodynamic Parameters of Loop Formation^{a,b}

| RNA sequence | ΔG° ₃₇ (kcal/mol) | ΔH° (kcal/mol) | ΔS° (eu) |
|------------------------------------|---------------------------------|-------------------|--------------|
| GCUGGCG | -3.94 ± 0.16 | -28.8 ± 1.8 | -80.2 ± 5.3 |
| CGCAGGCG | -0.64 ± 0.23 | -13.9 ± 2.8 | -42.6 ± 8.4 |
| GGCGUGCC | -0.40 ± 0.14 | -14.1 ± 1.6 | -44.1 ± 4.7 |
| GGCGAGCC | -0.36 ± 0.28 | -6.3 ± 2.6 | -19.2 ± 7.7 |
| CGCUUGCG | -0.08 ± 0.12 | -14.9 ± 1.5 | -47.7 ± 4.5 |
| UGC ^{GG} GCA ^c | 1.18 ± 0.36 | -20.2 ± 6.3 | -68.9 ± 18.3 |
| CGCCAGCG | 1.44 ± 0.08 | -8.2 ± 1.0 | -30.9 ± 3.2 |
| CGCAAGCG | 1.69 ± 0.11 | -1.6 ± 1.5 | -10.3 ± 4.5 |
| CGCCUGCG | 1.70 ± 0.13 | 2.0 ± 1.3 | 1.2 ± 4.0 |
| CGCUUGCG | 1.70 ± 0.25 | -0.2 ± 2.2 | -6.1 ± 6.8 |
| CGCCCGCG | 2.00 ± 0.11 | 0.8 ± 1.2 | -3.7 ± 3.7 |
| GGCACGCG | 2.33 ± 0.11 | 0.9 ± 1.3 | -4.7 ± 3.9 |

^aListed in order of decreasing loop free energy. ^bCalculated from T_M⁻¹ vs log C_T parameters. ^cError based on assuming error for rUGCGCA is 10% for ΔH° and ΔS° and 2% for ΔG° (Freier et al., 1986b).

in Table I with equations such as (Gralla & Crothers, 1973a) ΔG°_{37,loop(UU)} = ΔG°_{37(rCGCUUGCG)} - ΔG°_{37(rCGCGCG)} + ΔG°_{37(CG)}. Here G°_{37(CG)} is the free energy increment for the nearest-neighbor interaction (Freier et al., 1986a) interrupted by the internal loop. Results from such calculations are listed in Table II in order of loop stability at

pH 7 at 37 °C. The value of 2.0 kcal/mol for the CC loop at pH 7 is similar to the value of 1.6 kcal/mol obtained from the data of Gralla and Crothers (1973a) for the CC loop in $rA_4GC_2CU_4$. Free energy increments at 37 °C for the loops in Table I range from -0.6 to +2.3 kcal/mol. For RNA's of about 400 nucleotides, significantly different secondary structures are often predicted to be within a few kilocalories per mole of each other (Jaeger et al., 1989; Zuker et al., 1991). Thus, an understanding of the sequence dependence of internal loop stability is important for improving predictions of secondary structure.

The results suggest two classes for internal loops with consecutive mismatches: (1) loops that stabilize duplexes and have strong hydrogen bonds and (2) loops that destabilize duplexes and may not have strong hydrogen bonds. Thus, loops containing two G-A or U-U mismatches stabilize the duplex by -0.1 to -0.6 kcal/mol. Loops with other mismatches destabilize the duplex by 1.1-2.3 kcal/mol. NMR spectra indicate the AG (SantaLucia et al., 1990) and UU sequences have hydrogen bonding of imino protons within the mismatches. Replacing the amino group of A with hydrogen in GA internal loops destabilizes the duplexes by almost 3 kcal/mol, suggesting strong hydrogen bonding involving the A amino group (SantaLucia et al., 1991). Thus, the stabilizing internal loops all have hydrogen bonding. Preliminary experiments with a series of consecutive G-U pairs indicate they are also in this class (L. He, J. SantaLucia, R. Kierzek, A. E. Walter, and D. H. Turner, unpublished experiments).

Conversely, there is no evidence for hydrogen bonding in the destabilizing loops. Removal of the four amino groups in an AA internal loop decreases stability by less than 0.3 kcal/mol per amino group (SantaLucia et al., 1991). The number of imino resonances observed for sequences with destabilizing loops is never more than the number of different Watson-Crick base pairs expected (see Figure 4). Absence of an imino resonance, however, does not rule out the possibility of hydrogen bonding (Gueron et al., 1987). Thus, the NMR data only suggest that destabilizing loops lack strong hydrogen bonding. Finally, the ΔH° 's and ΔS° 's for loop formation show a rough trend, with more stable loops generally having more negative ΔH° 's and ΔS° 's (see Table II). This is also consistent with more stable loops having hydrogen bonding. The GA loop does not clearly fit this trend. Spectroscopic data have suggested GA loops have unusual structures, however (SantaLucia et al., 1990).

The NMR evidence shown in Figures 5 and 7 suggests the observed thermodynamic stability of rCGCUUGCG is due to strong hydrogen bonding, as shown in Figure 6, within a helix which is relatively undistorted. The NMR evidence is not sufficient to deduce which structure is formed or whether the two forms interconvert rapidly. A U-U mismatch has previously been found to stabilize anticodon-anticodon interactions between tRNA^{Asp}, both in solution (Romby et al., 1985) and in crystals (Moras et al., 1980; Westhof et al., 1985).

There have been extensive structural and thermodynamic studies of single mismatches in DNA oligomers (Patel et al., 1984; Aboul-ela et al., 1985; Arnold et al., 1987; Werntges et al., 1986; Brown et al., 1986; Kouchakdjian et al., 1988). At pH 7, the structural studies provide evidence for hydrogen bonding between G-U, G-A (Patel et al., 1984; Brown et al., 1986), and T-T (Kouchakdjian et al., 1988), but not other mismatches. T-T mismatches with wobble-type structure have also been found in the loop of a DNA hairpin (Blommers et al., 1989). These structural studies of DNA are consistent with the hydrogen-bonding hypothesis for the RNA internal

loops. The stabilities of the single mismatches in DNA, however, do not follow the trend observed in Table II. In particular, G-G is often the most stable, and T-T is one of the least stable mismatches (Aboul-ela et al., 1985). This is perhaps not surprising, since trends in the sequence dependence of Watson-Crick base pairs also differ for DNA and RNA (Breslauer et al., 1986; Freier et al., 1986a).

Current algorithms for the prediction of RNA secondary structure from sequence (Zuker, 1989; Zuker & Stiegler, 1981; Papanicolaou et al., 1984) use either a sequence-independent (Gralla & Crothers, 1973b) or a simple stacking model (Jaeger et al., 1989) for the sequence dependence of internal loop stability. These approximations were necessary due to lack of experimental data. The results in Table II show that neither model is adequate. In a sequence-independent model, all internal loops of four nucleotides have the same free energy. In fact, the free energies differ by 3 kcal/mol. In the stacking model, the order of stabilities is predicted to be AC = AG = GA = GG = AA > UC > CC = CA = CU = UU. The measured order of stabilities is AG = GA \approx UU > GG \geq CA \geq AA = CU = UC \geq CC \geq AC. It is interesting that the stacking model (based on stacking increments measured for 3'-dangling ends) would predict AC to be more stable than CA by -0.9 kcal/mol but actually AC is 0.9 kcal/mol less stable than CA. This suggests that either hydrogen bonding within each C-A pair or stacking between adjacent C-A mismatches must have a profound sequence dependence. The results show that an improved model is required for internal loop stability.

While an improved model for internal loop stability is necessary, the best way to develop this model is not clear. For helices with only A-U and G-C base pairs, a nearest-neighbor model is adequate for predicting stability (Borer et al., 1974; Freier et al., 1986a; Kierzek et al., 1986). A nearest-neighbor model for symmetric internal loops, however, requires 126 additional free energy parameters, too many to determine experimentally. Moreover, recent results suggest a separate set of parameters may be required for hairpin loops. For example, hairpin loops containing all U's or all A's have essentially identical stabilities (Groebe & Uhlenbeck, 1988), whereas the internal loops with all U's and all A's in Table II differ in stability by almost 2 kcal/mol. In other cases, however, stabilities of hairpin loops do depend on sequence (Tuerk et al., 1988). Clearly, approximations are required. In principle, computations using free energy perturbation and related methods (McCammon & Harvey, 1987; Bash et al., 1987) might provide reasonable approximations for internal loop sequences not measured. Alternatively, a simple base-mismatch model neglecting nearest-neighbor effects between mismatches might be adequate. In this internal loop model, favorable free energies are used for G-A, U-U, and G-U mismatches, and unfavorable free energies for other mismatches. This and other models can be tested by comparing predictions of secondary structures with known structures (Papanicolaou et al., 1984; Turner et al., 1987; Jaeger et al., 1989). Promising models can then be tested further with oligonucleotide model systems.

The above discussion is relevant to internal loops of four or more nucleotides. Single mismatches may also contribute to stability. For example, two U-U mismatches have been proposed in important regions of the self-splicing group I intron from *Tetrahymena thermophila* (Michel et al., 1982; Davies et al., 1982; Michel & Westhof, 1990). Alkema et al. (1982), however, found that a U-U mismatch was much less stable than an A-A or C-C mismatch in the duplexes (rAGACU)₂,

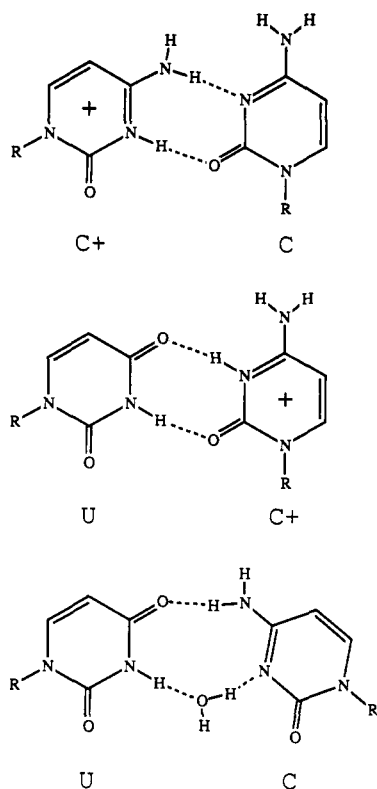


FIGURE 10: Potential mismatch structures for (top) C-C+ (Gray et al., 1984), (middle) U-C+, and (bottom) water-mediated U-C (Patel et al., 1984).

(rAGCCU)₂, and (rAGUCU)₂. In preliminary experiments on rCGCUGCG (S. Dreiker, N. Sugimoto, and D. H. Turner, unpublished experiments), however, ΔG°_{37} , ΔH° , and ΔS° from T_M^{-1} plots are -6.3 kcal/mol, -56.8 kcal/mol, and -162.8 eu, respectively, and the melting appears two-state. This gives a ΔG°_{37} for the U-U mismatch loop of 0.8 kcal/mol, similar to the 1.1 kcal/mol measured for the A-A mismatch loop in rCGCAGCG (Peritz et al., 1991). Apparently, in this context, single U-U and A-A mismatches have similar stabilities, whereas consecutive U-U mismatches are more stable than consecutive A-A mismatches. This could reflect nearest-neighbor interactions between the mismatches or that some local flexibility is required for the enhanced U-U stability. Interestingly, the two U-U mismatches proposed by Michel and Westhof (1990) in a self-splicing group I intron are at the end and one base pair removed from the end of their helices. Thus, they are in regions likely to be more flexible than the middle of a helix. Papanicolaou et al. (1984) have suggested that stability might depend on the distance from the mismatch to the helix end. Data for the series rGCGAGC, rGCGAGCG, and rGGCGAGCC (SantaLucia et al., 1990) support this hypothesis, since the $\Delta G^{\circ}_{37, \text{loop}}$ values for the GA loops are -1.4 , -1.0 , and -0.4 kcal/mol, respectively.

Several studies in DNA (Gao & Patel, 1988; Kouchakdjian et al., 1988, 1989; Edwards et al., 1990) and RNA (Puglisi et al., 1990) have suggested that the structure and stability of some mismatches depend on solution pH. Upon lowering the pH to 5.5, rCGCCGCG is stabilized by almost 2.5 kcal/mol, and the T_M increases by 12 °C. The pH dependence of the stability of rCGCCGCG is most likely due to protonation of internal loop C's, since no changes in thermodynamic parameters are observed for rCGCUUGCG between pH 7 and 5.5. Furthermore, new NMR resonances at 10.05 and 10.85 ppm are observed (Figure 8) upon lowering the pH,

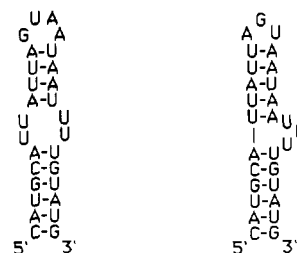


FIGURE 11: Nucleotides 98–125 of the yeast B1 intron. The structure on the left was derived by comparative sequence analysis (phylogeny) (Michel et al., 1989). The structure on the right was derived by energy minimization with the loop parameters of Jaeger et al. (1989).

suggesting a C-C+ base pair is formed that is isomorphous with the U-U mismatch (see Figure 10). In CD spectra, a positive band at 290 nm is indicative of C-C+ base-pair formation (Brahms et al., 1967; Gray et al., 1984). We also observe this for rCGCCGCG at pH 5.5 (see Figure 3D). The CD data suggest that no large conformational change is required to form the C-C+ base pair. Interestingly, Gutell and Woese (1990) have found several examples in 16S and 23S rRNA of tertiary pairing between U-U mismatches, and that these U-U mismatches are most often substituted by C-C mismatches.

In principle, UC mismatches can form a U-C+ pair that is isomorphous with U-U and C-C+ mismatches (see Figure 10). The thermodynamic data and CD spectra for rCGCUGCG and rGCCUGCG, however, show very small pH dependences. For rCGCUGCG, imino proton resonances broaden, but no new resonances are observed upon lowering the pH (data not shown). Large changes in the NMR spectrum for rGCCUGCG, however, suggest the structure is pH-dependent at low temperatures. Specifically, new resonances are observed at 11.4, 9.7, and 9.1 ppm, and there is a downfield shift of a sharp resonance to 13.85 ppm. Patel et al. (1984) have presented indirect evidence for a hydrated U-C structure as shown in Figure 10 (bottom). Further investigation is required to determine which structure is formed.

A-C mismatches can form an A+C wobble pair at low pH (Puglisi et al., 1990) which is isomorphous with the G-U wobble pair (Crick, 1966). Sequences with AC and CA internal loops, however, do not show extra resonances upon lowering the pH (data not shown). The thermodynamics of AC and CA sequences are only slightly pH-dependent between 7.0 and 5.5. This suggests the pK_a 's for protonation of the A-C mismatches in rGCCAGCG and rGGCAGCC are lower than 5.5.

In this report, we suggest that RNA helices with internal loops involve novel pairing schemes that affect stability and structure. Natural RNA molecules may use these motifs. Figure 11 illustrates one example where the results from Table II can improve predictions of RNA structure. Two alternate foldings of nucleotides 98–125 of the yeast B1 intron are shown. The phylogenetic structure contains an internal loop with two consecutive U-U mismatches (Michel et al., 1989). With the base-pair parameters of Freier et al. (1986a) and the loop model of Jaeger et al. (1989), however, the structure containing a bulge of 3 U's is predicted to be 1.2 kcal/mol lower in free energy. In this model, the internal loop, UU, has a free energy of $+3.6$ kcal/mol. If the UU loop closed by two A-U pairs in Figure 11 is similar to the UU loop closed by two C-G pairs in Table II, then the free energy of this loop is -0.1 kcal/mol. This leads to the prediction that the phylogenetic structure is actually 2.5 kcal/mol more stable than

the alternate structure. The phylogenetic structure also contains a potentially unusually stable hairpin loop with the sequence GUAA (Tuerk et al., 1988; Woese et al., 1990; Cheong et al., 1990). The comparison suggests that further understanding of the interactions that determine stabilities of loop regions will improve predictions of RNA structure from sequence.

ACKNOWLEDGMENTS

We thank Scott Dreiker for his data on rCGCUGCG and Liyan He and Adam Peritz for stimulating discussions.

REFERENCES

- Aboul-ela, F., Koh, D., Tinoco, I., Jr., & Martin, F. H. (1985) *Nucleic Acids Res.* **13**, 4811–4824.
- Alkema, D., Hader, P. A., Bell, R. A., & Neilson, T. (1982) *Biochemistry* **21**, 2109–2117.
- Arnold, F. H., Wolk, S., Cruz, P., & Tinoco, I., Jr. (1987) *Biochemistry* **26**, 4068–4075.
- Bash, P. A., Singh, U. C., Langridge, R., & Kollman, P. A. (1987) *Science* **236**, 564–568.
- Blommers, M. J. J., Walters, J., Haasnoot, C. A. J., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1989) *Biochemistry* **28**, 7491–7498.
- 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.
- Brahms, J., Maurizot, J. C., & Michelson, A. M. (1967) *J. Mol. Biol.* **25**, 465–480.
- Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3746–3750.
- 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.
- Cheong, C., Varani, G., & Tinoco, I., Jr. (1990) *Nature* **346**, 680–682.
- Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548–555.
- Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A., & Scazzocchio, C. (1982) *Nature (London)* **300**, 719–724.
- de Smit, M. H., & van Duin, J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7668–7672.
- Edwards, E. L., Patrick, M. H., Ratliff, R. L., & Gray, D. M. (1990) *Biochemistry* **29**, 828–836.
- 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.
- Gorenstein, D. (1984) in *³¹P NMR, Principles and Applications*, Academic, New York.
- Gralla, J., & Crothers, D. M. (1973a) *J. Mol. Biol.* **73**, 497–511.
- Gralla, J., & Crothers, D. M. (1973b) *J. Mol. Biol.* **78**, 301–319.
- Gray, D., Cui, T., & Ratliff, R. (1984) *Nucleic Acids Res.* **12**, 7565–7580.
- Groebe, D. R., & Uhlenbeck, O. C. (1988) *Nucleic Acids Res.* **16**, 11725–11735.
- Guéron, M., Kochoyan, M., & Leroy, J.-L. (1987) *Nature* **328**, 89–92.
- Guschlbauer, W. (1975) *Nucleic Acids Res.* **2**, 353–360.
- Gutell, R. R., & Woese, C. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 663–667.
- 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.
- Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., & Freier, S. M. (1986) *Biochemistry* **25**, 7840–7846.
- Konings, D. A. M., & Hogeweg, P. (1989) *J. Mol. Biol.* **207**, 597–614.
- Kouchakdjian, M., Li, B. F. L., Swan, P. F., & Patel, D. J. (1988) *J. Mol. Biol.* **202**, 139–155.
- Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. J. (1989) *Biochemistry* **28**, 5647–5657.
- Le, S.-Y., Chen, J.-H., Braun, M. J., Gonda, M. A., & Maizel, J. V. (1988) *Nucleic Acids Res.* **16**, 5153–5168.
- Le, S.-Y., Chen, J.-H., & Maizel, J. V. (1989) *Nucleic Acids Res.* **17**, 6143–6152.
- Marky, L. A., & Breslauer, K. J. (1987) *Biopolymers* **26**, 1601–1620.
- McCammon, J. A., & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, U.K.
- Michel, F., & Westhof, E. (1990) *J. Mol. Biol.* **216**, 585–610.
- Michel, F., Jacquier, A., & Dujon, B. (1982) *Biochimie* **64**, 867–881.
- Michel, F., Umeson, K., & Ozeki, H. (1989) *Gene* **82**, 5–30.
- Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., & Giege, R. (1980) *Nature* **288**, 669–674.
- Papanicolaou, C., Gouy, M., & Ninio, J. (1984) *Nucleic Acids Res.* **13**, 1717–1731.
- Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **43**, 2663–2670.
- 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.
- Puglisi, J. D., Wyatt, J. R., & Tinoco, I., Jr. (1990) *Biochemistry* **29**, 4215–4226.
- 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.
- Romby, P., Giege, R., Houssier, C., & Grosjean, H. (1985) *J. Mol. Biol.* **184**, 107–118.
- SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry* **29**, 8813–8819.
- SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991) *J. Am. Chem. Soc.* **113**, 4313–4322.
- Sundquist, W. I., & Klug, A. (1989) *Nature* **342**, 825–829.
- Tinoco, I., Jr., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature (London)* **230**, 362–367.
- 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., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167–192.

- Usman, N., Ogilvie, K. K., Jiang, M. Y., & Cedergren, R. L. (1987) *J. Am. Chem. Soc.* 109, 7845-7854.
- Varani, G., Wimberly, B., & Tinoco, I., Jr. (1989) *Biochemistry* 28, 7760-7772.
- Werntges, H., Steger, G., Riesner, D., & Fritz, H. J. (1986) *Nucleic Acids Res.* 14, 3773-3790.
- Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol.* 184, 119-145.
- Wickstrom, E. L., Bacon, T. A., Gonzalez, A., Freeman, D. L., Lyman, G. H., & Wickstrom, E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 1028-1032.
- Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell* 59, 871-880.
- Woese, C. R., Winker, S., & Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 8467-8471.
- Yager, T. D., & von Hippel, P. H. (1991) *Biochemistry* 30, 1097-1118.
- Zuker, M. (1989) *Science* 244, 48-52.
- Zuker, M., & Stiegler, P. (1981) *Nucleic Acids Res.* 9, 133-148.
- Zuker, M., Jaeger, J. A., & Turner, D. H. (1991) *Nucleic Acids Res.* 19, 2707-2714.

Cell Adhesion Promoting Peptide GVKGDKGNPGWPGAP from the Collagen Type IV Triple Helix: Cis/Trans Proline-Induced Multiple ^1H NMR Conformations and Evidence for a KG/PG Multiple Turn Repeat Motif in the All-Trans Proline State[†]

Kevin H. Mayo* and Dennisse Parra-Diaz

Program in Molecular Pharmacology and Structural Biology, Department of Pharmacology, Jefferson Cancer Institute, Thomas Jefferson University, Life Sciences Building, Philadelphia, Pennsylvania 19107

James B. McCarthy and Mary Chelberg

Department of Lab Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, Minnesota 55455

Received January 9, 1991; Revised Manuscript Received May 31, 1991

ABSTRACT: Peptide GVKGDKGNPGWPGAPY (called peptide IV-H1), derived from the protein sequence of human collagen type IV, triple-helix domain residues 1263-1277, represents an RGD-independent, cell-specific, adhesion, spreading, and motility promoting domain in type IV collagen. In this study, peptide IV-H1 has been investigated by ^1H NMR (500 MHz) spectroscopy. Cis-trans proline isomerization at each of the three proline residues gives rise to a number of slowly exchanging (500-MHz NMR time scale) conformation states. At least five such states are observed, for example, for the well-resolved A14 βH_3 group, and K3, which is six residues sequentially removed from the nearest proline, i.e., P9, shows two sets. The presence of more than two sets of resonances for residues sequentially proximal to a proline, e.g., A14-cis-P15 and A14-trans-P15, and more than one set for a residue sequentially well-removed from a proline, e.g., K3, indicates long range conformation interactions and the presence of preferred structure in this short linear peptide. Many resonances belonging to these multiple species have been assigned by using mono-proline-substituted analogues. Conformational (isomer) state-specific 2D ^1H NMR assignments for the combination of cis and trans proline states have been made via analysis of COSY-type, HOHAHA, and NOESY spectra. Peptide IV-H1 in the all-trans proline state ttt exists in relatively well-defined conformation populations showing numerous short- and long-range NOEs and long-lived backbone amide protons and reduced backbone NH temperature coefficients, suggesting hydrogen-bonding, and structurally informative $^3J_{\alpha\text{N}}$ coupling constants. The NMR data indicate significant β -turn populations centered at K3-G4, K5-G6, P9-G10, and P12-G13, and a C-terminal γ -turn within the A14-P15-Y16 sequence. These NMR data are supported by circular dichroic studies which indicate the presence of 52% β -turn, 10% helix, and 38% random coil structural populations. Since equally spaced KG and PG residues are found on both sides of peptide IV-H1 in the native collagen type IV sequence, this multiple turn repeat motif may continue through a longer segment of the protein. Synthetic peptide IV-H1 overlapping sequence "walk throughs" indicate that the primary biological activity is localized in the GNPGWPGAP double β -turn domain, which contains the backbone constraining proline residues. This proline-domain conformation may suggest a collagen type IV receptor-specific, metastatic cell adhesion promoting binding domain.

Type IV collagen is a collagenous glycoprotein that forms the major scaffolding of basement membranes (Timpl et al.,

[†]This work was supported by generous research grants from the W. W. Smith Charitable Trust, the Elsa Pardee Foundation, the Leukemia Task Force, and the National Cancer Institute (CA-43924), and benefited from NMR facilities made available to Temple University through Grant RR-04040 (to K.H.M.) from the National Institutes of Health. D.P.-D. is indebted to the Graduate School of Temple University for a Future Faculty Fellows postdoctoral fellowship award. Dr. Carl Burke of Merck Sharp & Dohme Labs, West Point, PA, is gratefully acknowledged for having run and analyzed the CD spectrum of peptide IV-H1.

*Address correspondence to this author.

1981; Yurchenco & Furthmayr, 1983; Tsilibary & Charonis, 1986) and binds to other components like laminin, entactin, heparan sulfate, etc. (Timpl & Dziadek, 1986; Charonis et al., 1985; Laurie et al., 1986; Fugiwara et al., 1984). Type IV collagen functions in part as an adhesive protein since it promotes cell adhesion and specifically interacts with the surfaces of a variety of cell types (Kurkinen, 1984; Aumailley & Timpl, 1986; Sugrue, 1987; Murray et al., 1979). Members of the integrin superfamily of membrane proteins that bind to multiple extracellular matrix components, such as fibronectin and laminin (Ruoslahti & Pierschbacher, 1987; Ignatius & Reichardt, 1988; Gehlsen et al., 1986), have recently been