Ragan, C. I., Galante, Y. M., Hatefi, Y., & Ohnishi, T. (1982) *J. Biol. Chem. 21*, 590-594.
Raitio, M., Jalli, T., & Saraste, M. (1987) *EMBO J. 6*, 2825-2833.
Runswick, M. J., Gennis, R. B., Fearnley, I. M., & Walker, J. E. (1989) *Biochemistry 28*, 9452-9459.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368-379.
Scrutton, N. S., Berry, A., & Perham, R. N. (1990) *Nature 343*, 38-43.
Sharma, P. M., Reddy, G. R., Babior, B. M., & McLachlan, A. (1990) *J. Biol. Chem. 265*, 9006-9010.
Shine, J., & Dalgarno, L. (1975) *Nature 254*, 34-38.
Stouthamer, A. H. (1980) *Trends Biochem. Sci. 5*, 164-166.
Tran-Betcke, A., Warnecke, U., Bocker, C., Zaborosch, C., & Friedrich, B. (1990) *J. Bacteriol. 172*, 2920-2929.
Tuschen, G., Sackmann, U., Nehls, U., Haiker, H., Buse, G., & Weiss, H. (1990) *J. Mol. Biol. 213*, 845-857.
Tzagoloff, A., Capitanio, N., Nobrega, M. P., & Gatti, D. (1990) *EMBO J., 9*, 2759-2764.
Van Spanning, R. J. M., Wansell, C., Harms, N., Oltmann, L. F., & Stouthamer, A. H. (1990a) *J. Bacteriol. 172*, 986-996.
Van Spanning, R. J. M., Warnsell, C. W., Reijnders, W. N. M., Oltmann, L. F., & Stouthamer, A. H. (1990b) *FEBS Lett. 275*, 217-220.
Walker, J. E., Saraste, M., & Gay, N. J. (1984) *Biochim. Biophys. Acta 768*, 164-200.
Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M. S., Elsas, L. J., & Nikoskelainen, E. K. (1988) *Science 242*, 1427-1430.
Wierenga, R. K., Terpstra, P., & Hol, W. G. (1986) *J. Mol. Biol. 187*, 101-107.
Xu, X., & Yagi, T. (1991) *Biochem. Biophys. Res. Commun. 174*, 667-672.
Xu, X., Koyama, N., Cui, M., Yamagishi, A., Nosoh, Y., & Oshima, T. (1991) *J. Biochem.* (in press).
Yagi, T. (1986) *Arch. Biochem. Biophys. 250*, 302-311.
Yagi, T. (1989) *Protein, Nucleic Acid, Enzyme 34*, 351-363.
Yagi, T. (1991) *J. Bioenerg. Biomembr. 23*, 211-225.
Yagi, T., & Dinh, T. M. (1990) *Biochemistry 29*, 5515-5520.
Yagi, T., & Hatefi, Y. (1988) *J. Biol. Chem. 263*, 16150-16155.
Yagi, T., Hon-nami, K., & Ohnishi, T. (1988) *Biochemistry 27*, 2008-2013.
Yasunobu, K. T., & Tanaka, M. (1980) *Methods Enzymol. 69*, 228-238.
Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A., & Shaw, D. C. (1981) *Eur. J. Biochem. 116*, 165-170.

Thermodynamic Study of Internal Loops in Oligoribonucleotides: Symmetric Loops Are More Stable Than Asymmetric Loops†

Adam E. Peritz,‡ Ryszard Kierzek,§ Naoki Sugimoto,‡,∥ 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 November 8, 1990; Revised Manuscript Received April 1, 1991

ABSTRACT: Thermodynamic parameters for internal loops of unpaired adenosines in oligoribonucleotides have been measured by optical melting studies. Comparisons are made between helices containing symmetric and asymmetric loops. Asymmetric loops destabilize a helix more than symmetric loops. The differences in free energy between symmetric and asymmetric loops are roughly half the magnitude suggested from a study of parameters required to give accurate predictions of RNA secondary structure [Papanicolaou, C., Gouy, M., & Ninio, J. (1984) *Nucleic Acids Res. 12*, 31-44]. Circular dichroism spectra indicate no major structural difference between helices containing symmetric and asymmetric loops. The measured sequence dependence of internal loop stability is not consistent with approximations used in current algorithms for predicting RNA secondary structure.

Secondary structures of RNA contain motifs other than double helices. Single-stranded regions such as hairpin, bulge, internal, and bifurcated loops are prominent features. Despite their importance, little data exists on how single-strand motifs contribute to the overall stability of RNA. Incorporation of improved nearest-neighbor parameters (Freier et al., 1986a; Turner et al., 1988) into computer algorithms for structure prediction have recently increased the accuracy of RNA secondary structure predictions (Turner et al., 1987; Jaeger et al., 1989). Additional knowledge of thermodynamic parameters for single-strand motifs should further improve such predictions.

Internal loops form in double-helical RNA when the helix is interrupted by nucleotides on both strands that are not Watson–Crick paired. Symmetric or asymmetric loops can form depending on whether an equal or an unequal number of nucleotides are on opposing strands, respectively. Internal loops are thought to play important roles in biological functions

† This work was supported by National Institutes of Health Grant GM22939.
‡ University of Rochester.
§ Polish Academy of Sciences.
∥ Current address: Department of Chemistry, Konan University, Kobe 658, Japan.

0006-2960/91/0430-6428$02.50/0

including protein recognition (Gregory et al., 1988), assembly of ribosomal subunits (Stern et al., 1988), and splicing (Rodel et al., 1983).

Despite the importance of internal loops, there is little data describing their thermodynamic consequences for RNA folding. The length dependence of stability for symmetric, internal loops of two, four, and six cytidines was measured by Gralla and Crothers (1973). Varani et al. (1989) measured thermodynamic parameters for a duplex containing a symmetric internal loop of six unpaired nucleotides. SantaLucia et al. (1990) reported parameters for internal loops with two GA or AA mismatches. It has been shown, however, that predictions of RNA secondary structure from sequence are sensitive to parameters for internal loops (Papanicolaou et al., 1984; Turner et al., 1987, 1988). For example, inclusion of penalties for asymmetric loops and for closing internal loops by AU pairs increases the percentage of helixes predicted correctly for a set of RNAs from 55 to 70% (Turner et al., 1987). About half of this improvement is due to the asymmetry penalty (J. A. Jaeger and D.H.T., unpublished observations). In this paper, we report thermodynamic parameters for internal loop formation in 20 oligoribonucleotide duplexes. The internal loops are composed of adenosines since most naturally occurring internal loops are rich in purines. It is found that asymmetric loops are more destabilizing than symmetric loops with the same number of nucleotides. This provides the first experimental evidence for the asymmetric loop penalty inferred from comparisons of known and predicted RNA secondary structures (Papanicolaou et al., 1984; Turner et al., 1987).

Materials and Methods

Synthesis and Purification of RNA. Most oligomers were synthesized on solid support with the use of previously described phosphoramidite chemistry (Kierzek et al., 1986). UGACACUCA and $UGAGA_3GUCA$ were synthesized from purified phosphoramidites, using bis(diisopropylamino)methoxyphosphine as the phosphitylating agent.

Oligomers were removed from support, and base labile protecting groups were deblocked with concentrated ammonia. Oligomers synthesized with methoxyphosphoramidites were treated with thiophenol prior to cleavage from support. Oligomers were purified by reverse-phase chromatography on a Hamilton PRP-1 150 mm × 4.1 mm column. Buffer A consisted of 0.01 M ammonium acetate, pH 6.8 (Mallinkrodt, analytical reagent). Buffer B consisted of 1:1 v/v acetonitrile (JT Baker, HPLC grade) and 0.01 M ammonium acetate, pH 6.8. Oligomers were eluted with a gradient from 0 to 70% buffer B in 35 min with a flow rate of 2 mL/min. Oligomers usually eluted at 50–60% buffer B. The partially blocked oligomers were then treated with 0.01 M hydrochloric acid, pH 2, for 24 h to remove acid-labile protecting groups. A Sep-Pak C-18 cartridge was used to further purify and desalt the molecules. Oligomer purity was checked on a Beckman Ultrasphere C-8, 150 mm × 4.6 mm column with a gradient from 100% buffer A (0.01 M aqueous sodium phosphate, pH 7.0) to 30% buffer B [1:1 v/v = 0.01 M aqueous sodium phosphate, pH 7.0; methanol (JT Baker, HPLC grade)] in 30 min with a flow rate of 1 mL/min. Oligomers typically eluted between 15 and 20% buffer B.

Melting Curves. Samples were dissolved in 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA,[1] pH 7.0. Single-strand extinction coefficients were calculated from a nearest-neighbor model (Borer, 1975; Richards, 1975). They are listed in Table I. Single-strand concentrations were determined by measuring the absorbance at 280 nm at 80 °C. Oligomers were mixed in 1:1 concentration ratios for studies of non-self-complementary duplexes. All melting curves were measured at 280 nm with a Gilford 250 spectrophotometer connected to a Gilford 2527 thermoprogrammer. Data were collected and fit with a Digital PDP 11/34 or a Zenith Z-248 computer. Thermodynamic parameters were obtained from two methods. (1) Melting curves were fit with the Marquardt non-linear least-squares method. Either the self-complementary (Petersheim & Turner, 1983) or non-self-complementary (Longfellow et al., 1990) algorithm was used, as appropriate. (2) For self-complementary sequences, plots of reciprocal melting temperature, T_M^{-1}, vs the logarithm of total strand concentration, log C_T, were fit to (Borer et al., 1974) eq 1. For

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

non-self-complementary sequences, C_T was replaced by $C_T/4$.

The nearest-neighbor approximation for extinction coefficients of the single strands could lead to mixing of strands in ratios that are not exactly 1:1. To test the effect of this on the derived thermodynamic parameters, simulations were performed for mixing UGAGAAGUCA and UGACAAAACUCA in 1.2:1 and 1.5:1 ratios. In these simulations, the measured absorbance vs temperature for the excess single-stranded UGAGAAGUCA was added to the duplex melts and the resulting curves were refit. T_M^{-1} vs log C_T plots were also generated. For the 1.2:1 ratio, the ΔG°'s from fits and the T_M^{-1} plot changed by 0.6 and 2%, respectively. For the 1.5:1 ratio, the ΔG°'s changed by 3 and 5%, respectively.

CD Spectra. CD spectra were collected on a JASCO J-40 spectrophotometer interfaced to a Digital PDP 11/23 computer. Temperature was controlled by using jacketed cylindrical cells.

NMR Spectra. NMR spectra were measured with a 500-MHz Varian VXR-500S spectrometer. Oligomers were dissolved in 1 M NaCl/10 mM sodium phosphate/0.5 mM EDTA at pH 7. The solvent was 10% D_2O and 90% H_2O. Spectra were collected with 24000 points over a sweep width of 12 kHz, multiplied by a 2.0-Hz line-broadening exponential function, and transformed by a SUN 3/160 computer running Varian VNMR software. 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.

Results

Figure 1 shows normalized absorbance vs temperature melting curves for $UGACA_3CUCA$, $UGAGA_3GUCA$, a 1:1 mixture of the two strands, and a 1:1 mixture of UGACACUCA with $UGAGA_5GUCA$. The individual strands have broad transitions. All the individual, non-self-complementary strands used in this work were melted at two or more concentrations differing by at least a factor of 10. In general, the melts were concentration independent and broad, as expected for single-strand melting. These data were fit to a one-dimensionsl Ising model with a cooperativity parameter, σ, of 1 or 0.5 (Zimm & Bragg, 1959; Applequist, 1963; Dewey & Turner, 1979). The values derived for ΔH° and ΔS° are listed in the supplementary material (see paragraph at end of paper regarding supplementary material). When $\sigma = 1$, values for ΔH° range from −8 to −13 kcal/mol, which is typical for single-strand stacking (Leng & Felsenfeld, 1966; Brahms et al., 1967; Pörschke, 1976; Dewey & Turner, 1979; Freier et

[1] Abbreviations: CD, circular dichroism; Na_2EDTA, disodium ethylenediaminetetraacetate.

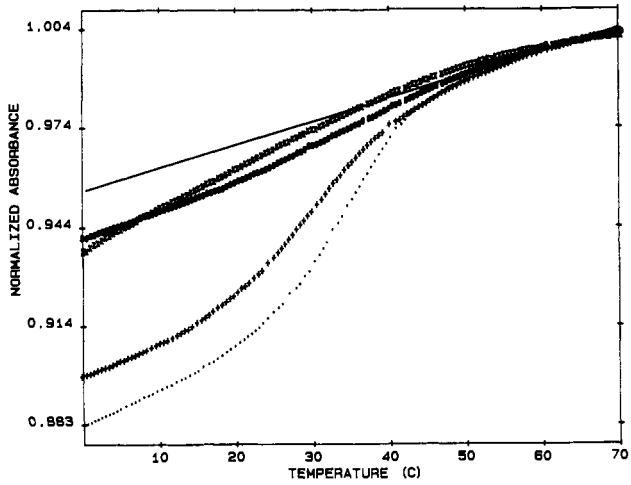


FIGURE 1: Normalized absorbance vs temperature curves for 1.6×10^{-5} M each strand UGAGAAAGUCA + UGACAAACUCA (●), 1.7×10^{-5} M each strand UGACACUCA + UGAGAAAAAGUCA (+), 2.0×10^{-5} M UGACAAACUCA (×), 2.1×10^{-5} M UGAGAAAGUCA (○), and linear approximation of single-strand extinction coefficient vs temperature derived from fitting curve for 1.6×10^{-5} M UGAGAAAGUCA + UGACAAACUCA. Curves are normalized at 64 °C. Solutions are 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7.

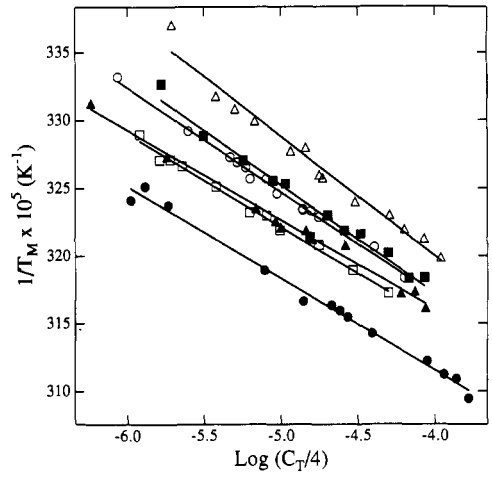


FIGURE 2: Reciprocal melting temperature vs log concentration plots for UGACACUCA + UGAGAGUCA (●), UGACAACUCA + UGAGAAGUCA (▲), UGACAAACUCA + UGAGAAAGUCA (○), UGACCAACUCA + UGAGAAAGUCA (□), UGACAACUCA + UGAGAAAAGUCA (△), and UGACACUCA + UGAGAAAGUCA (■). Solutions are 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7.

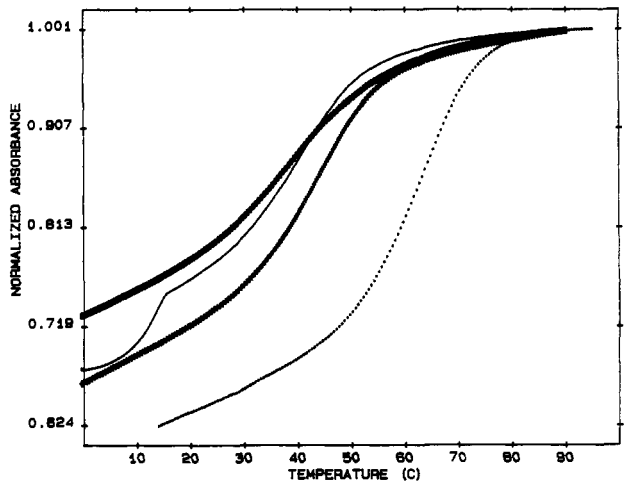


FIGURE 3: Normalized absorbance vs temperature curves for 3.2×10^{-4} M CGCGCG (●), 3.4×10^{-4} M CGCAGCG (+), 2.9×10^{-4} M CGCAAGCG (—), and 3.2×10^{-4} M CGCAAAGCG (○). Curves are normalized at 90 °C. Solutions are 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7.

al., 1981). The melts for the individual strands were also fit to a two-state hairpin model with sloping lower and upper base lines. The $\Delta G°$'s for these fits ranged from −5 to −17 kcal/mol with an average of −12.5 kcal/mol for the CA_nC series and from −11 to −16 kcal/mol with an average of −13.5 kcal/mol for the GA_nG series. These apparent $\Delta H°$'s are considerably less than the predicted values of −24 and −18 kcal/mol from the base pairs in the CA_nC and GA_nG series, respectively (Freier et al., 1986a).

The melting curves for UGACCUCA, UGAGGUCA, and UGAGAGUCA differed from those of the other strands. Both UGACCUCA and UGAGGUCA melted with sharper transitions than the other sequences, and the T_M for UGAGGUCA was concentration dependent. Even at the highest concentration melted (>1 × 10^{-4} M), however, the T_M's for both sequences were less than 20 °C. This is more than 40 °C lower than for the non-self-complementary duplex, UGACCUCA + UGAGGUCA (see below). Thus, the self-associated species are much less stable than the UGACCUCA + UGAGGUCA duplex, and this competition can be neglected (Longfellow et al., 1990). UGAGAGUCA melted with a sharp transition above 1.6×10^{-4} M with T_m's less than 22 °C. Below 3×10^{-5} M, it melted with a broad transition. Thus, the self-associated species of UGAGAGUCA is also too unstable to compete with the non-self-complementary duplexes studied (Longfellow et al., 1990).

Melting curves were cooperative for mixtures designed to give duplexes with internal loops (see Figure 1). Melting temperatures for these mixtures were concentration dependent, and Figure 2 shows typical T_M^{-1} vs log C_T plots. Thermodynamic parameters derived from fits of melting curves and from T_M^{-1} vs log C_T plots are listed in Table I. The measured $\Delta G°_{37}$ for the fully paired duplex, UGACCUCA + UGAGGUCA, is −12.3 kcal/mol, close to the value of −11.5 kcal/mol predicted by the nearest-neighbor model (Freier et al., 1986a; Kierzek et al., 1986). This is a further indication that the self-associated species do not affect the duplex equilibrium.

Figure 3 shows normalized absorbance vs temperature melting curves for the series $CGCA_nGCG$ where n = 1, 2, and 3. When n = 2, an unusual transition is observed at low temperature for strand concentrations above 1×10^{-4} M. Gralla and Crothers (1973) also observed unusual behavior for A_4GCCU_4. The equilibria associated with these unusual transitions are not known. For $CGCA_2GCG$, only data above the low-temperature transition were fit to the two-state model for duplex formation. This assumes that only a duplex to single strand transition occurs at the higher temperatures. The T_M^{-1} vs log C_T plots for the $CGCA_nGCG$ series are shown in supplementary material, and the thermodynamic parameters derived are listed in Table I.

The thermodynamic parameters listed in Table I are derived from a two-state model for the transition with the assumption that extinction coefficients for single strands depend linearly on temperature (Petersheim & Turner, 1983). As shown in Figure 1, this assumption is not exact. If the model was exact, then the $\Delta H°$ and $\Delta S°$ values derived from fits of melting curves and from T_M^{-1} vs log C_T plots should be equal. In practice, agreement within 15% is usually considered a necessary condition to indicate approximate two-state behavior (Freier et al., 1986; Turner et al., 1988). This is based on comparisons of $\Delta H°$'s measured by optical melting and ca-

Table I: Thermodynamic Parameters of Duplex Formation[a]

		$1/T_m$ vs log C_T				CURVE FITS			
	ε (cm^{-1} M^{-1}) (280 nm)	$-\Delta G^0_{37°C}$ (kcal/mole)	$-\Delta H^0$ (kcal/mole)	$-\Delta S^0$ (eu)	T_m (°C) (1×10^{-4} M)	$-\Delta G^0_{37°C}$ (kcal/mole)	$-\Delta H^0$ (kcal/mole)	$-\Delta S^0$ (eu)	T_m (°C) (1×10^{-4} M)
5'UGACCUCA3' 3'ACUGGAGU5'	32,340 36,500	12.34	76.09	205.55	62.6	11.90	69.13	184.52	63.1
5'UGAC^{A}CUCA3' 3'ACUG$_{A}$GAGU5'	35,040 39,600	7.94 (7.73)	67.15 (62.98)	190.91 (178.15)	43.7 (43.0)	7.74 (7.59)	53.77 (55.37)	148.40 (154.07)	44.1 (43.0)
5'UGAC ^{A} CUCA3' 3'ACUG$_{AA}$GAGU5'	43,120	6.42 (6.10)	57.98 (50.26)	166.22 (142.37)	36.4 (34.4)	6.61 (5.73)	46.11 (60.10)	127.34 (175.29)	37.6 (32.9)
5'UGAC^{AA}CUCA3' 3'ACUG $_{A}$ GAGU5'	38,560	6.59 (6.44)	55.76 (54.36)	158.51 (154.50)	37.4 (36.5)	6.71 (6.51)	45.07 (47.02)	123.66 (130.64)	38.3 (36.8)
5'UGAC^{AA}CUCA3' 3'ACUG$_{AA}$GAGU5'		7.04 (6.53)	69.09 (56.51)	200.06 (161.15)	39.3 (37.0)	7.06 (6.49)	51.89 (60.68)	144.54 (174.71)	40.2 (36.8)
5'UGAC ^{A} CUCA3' 3'ACUG$_{A3}$GAGU5'	46,640	6.55 (6.33)	56.77 (59.64)	161.92 (171.89)	37.1 (35.9)	6.64 (6.39)	45.85 (51.85)	126.44 (146.60)	37.7 (36.1)
5'UGAC^{A3}CUCA3' 3'ACUG $_{A}$ GAGU5'	42,080	6.16 (5.98)	49.77 (46.19)	140.60 (129.65)	34.7 (33.3)	6.44 (6.13)	40.96 (42.50)	111.31 (117.25)	36.3 (34.1)
5'UGAC^{A3}CUCA3' 3'ACUG$_{AA}$GAGU5'		6.59 (6.31)	60.45 (56.77)	173.65 (162.68)	37.3 (35.8)	6.69 (6.36)	49.70 (52.68)	138.69 (149.36)	38.0 (36.0)
5'UGAC ^{A} CUCA3' 3'ACUG$_{A4}$GAGU5'	50,160	6.11 (5.83)	51.40 (52.64)	146.00 (150.92)	34.5 (32.9)	6.28 (5.85)	47.43 (52.31)	132.68 (149.78)	35.4 (33.0)
5'UGAC^{A4}CUCA3' 3'ACUG $_{A}$ GAGU5'	45,600	5.56 (5.28)	47.39 (51.40)	134.86 (148.71)	30.8 (29.6)	5.85 (5.55)	39.91 (40.73)	103.36 (113.41)	31.5 (29.7)
5'UGAC^{AAA}CUCA3' 3'ACUG$_{AAA}$GAGU5'		6.67 (6.53)	59.43 (55.01)	170.12 (156.31)	37.7 (37.0)	6.74 (6.45)	57.63 (63.49)	164.09 (183.90)	38.1 (36.6)
5'UGAC^{AA}CUCA3' 3'ACUG$_{A4}$GAGU5'		6.07 (5.80)	51.61 (59.30)	146.84 (172.48)	34.3 (33.3)	6.18 (5.84)	48.52 (55.61)	136.51 (160.46)	34.8 (33.2)
5'UGAC^{A4}CUCA3' 3'ACUG$_{AA}$GAGU5'		6.00 (5.68)	56.12 (54.20)	161.59 (156.45)	34.1 (32.2)	6.21 (5.60)	46.58 (56.59)	130.15 (164.43)	34.9 (32.0)
5'UGAC ^{A} CUCA3' 3'ACUG$_{A5}$GAGU5'	53,680	5.71 (5.51)	51.97 (43.64)	149.16 (122.92)	32.2 (29.9)	6.18 (5.60)	39.63 (40.80)	107.84 (113.51)	34.3 (30.1)
5'UGAC^{A5}CUCA3' 3'ACUG $_{A}$ GAGU5'	49,120	5.30 (4.83)	49.04 (52.55)	141.02 (153.85)	29.4 (27.3)	5.80 (5.25)	37.93 (43.71)	103.60 (124.00)	31.2 (28.1)
5'UGAC^{CAA}CUCA3' 3'ACUG$_{AAA}$GAGU5'	43,590	7.14 (6.78)	67.49 (68.99)	194.59 (200.57)	39.8 (38.1)	7.22 (6.72)	55.11 (72.76)	154.42 (212.95)	40.9 (37.8)
5'UGAC^{AAA}CUCA3' 3'ACUG$_{AAC}$GAGU5'	46,030	7.17 (6.92)	62.90 (56.86)	179.68 (161.01)	40.2 (39.2)	7.24 (6.90)	61.57 (70.30)	175.17 (204.44)	40.6 (38.6)
5'CGCGCG3' 3'GCGCGC5'	29,590	9.21	54.74	146.77	58.4	9.38	56.93	153.29	58.6
5'CGC^{A}GCG3' 3'GCG$_{A}$CGC5'	34,320	6.09	49.98	141.53	39.6	6.16	47.69	133.39	40.1
5'CGC^{AA}GCG3' 3'GCG$_{AA}$CGC5'	37,840	5.44	48.02	137.30	35.5	5.59	40.11	111.30	36.4
5'CGC^{AAA}GCG3' 3'GCG$_{AAA}$CGC5'	41,360	4.88	45.92	132.33	31.7	5.19	30.09	80.26	32.1
5' CGGCCGp3'[b] 3'pGCCGGC 5'	33,340	9.90	54.12	142.57	63.3	10.10	56.63	150.01	63.3
5'CGG^{AAA}CCG3' 3'GCC$_{AAA}$GGC5'	45,200	4.64	38.41	108.87	28.9	4.79	36.88	103.48	29.7

[a] Although estimated errors in $\Delta G°$, $\Delta H°$, and $\Delta S°$ are ±3%, ±10%, and ±10%, respectively, additional significant figures are given to allow accurate calculation of T_M and other parameters. Solutions are in 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7. Parameters in parentheses are corrected for single-strand extinction coefficients. [b] From Freier et al. (1986).

lorimetry (Breslauer et al., 1975; Albergo et al., 1981; Hickey & Turner, 1985). Only eight of the duplexes listed in Table I fit this 15% criterion. Thus, most of the $\Delta H°$ and $\Delta S°$ values in Table I must be treated cautiously.

For non-self-complementary duplexes other than UGACCUCA + UGAGGUCA, the linear approximation for single-strand extinction coefficients can be replaced with the measured temperature dependence of single-strand extinction coefficients. (For single-stranded UGAGAGUCA, this requires using data for concentrations less than 3×10^{-5} M.) Results from this analysis are listed in parentheses in Table I. Nine sequences that do not appear two-state with the linear approximation do appear two-state when measured extinctions are used. Only one sequence appears two-state with the linear approximation but not two-state when measured extinctions are used. This suggests that much of the apparent non-two-state behavior is due to the linear approximation for single-strand extinctions.

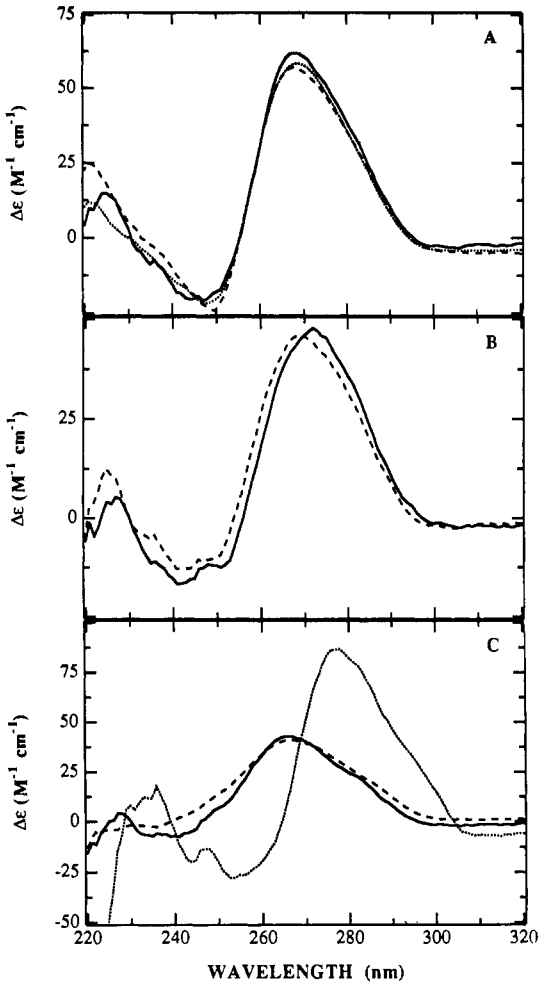


FIGURE 4: CD spectra at 0 °C for (A) 4.1×10^{-5} M UGACAAACUCA + UGAGAAAGUCA (---), 1.7×10^{-5} M UGACACUCA + UGAGAAAAAGUCA (···), and 1.2×10^{-5} M UGACAACUCA + UGAGAAAAGUCA (—); (B) 1.0×10^{-5} M UGACAACUCA + UGAGAAGUCA (—) and 1.4×10^{-5} M UGACACUCA + UGAGAAAGUCA (---); and (C) 1.1×10^{-5} M UGACCUCA + UGAGGUCA (---), 6.3×10^{-6} M UGACAAAACUCA + UGAGAGUCA (···), and 1.4×10^{-5} M UGACAACUCA + UGAGAGUCA (—). Solutions are 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7. Oligomer concentrations are concentrations of each strand.

The most important parameter for prediction of RNA structure is $\Delta G°_{37}$. For all sequences in Table I, the $\Delta G°_{37}$ values derived from fits and T_M^{-1} plots agree within 10%, with an average difference of 3%. The difference between $\Delta G°_{37}$ values derived with and without the measured temperature dependence of single-strand extinctions is at most 13% and on average 5%. This partially reflects compensating errors in $\Delta H°$ and $\Delta S°$ (Petersheim & Turner, 1983). In addition, the most certain thermodynamic parameter derived from optical melting is $\Delta G°$ at T_M (Freier et al., 1984). The sequences with internal loops were designed to melt near 37 °C so that only short extrapolations from T_M are required.

CD spectra for eight duplexes in the $UGACA_mCUCA$ + $UGAGA_nGUCA$ series are shown in Figure 4. Spectra for both symmetric and asymmetric loops are similar to those found for A-form RNA helixes (Tunis-Schneider & Maestre, 1970), though no definitive structural conclusion can be derived from CD spectra in this wavelength range (Sutherland et al., 1983, 1986).

NMR spectra at 15 °C for the imino protons of $UGAGA_3GUCA$ and its duplex with UGACACUCA are shown in Figure 5. Spectra were also taken for 2.4×10^{-4} M UGACACUCA, but no resonances were observed between 9 and 14 ppm. The spectrum for the duplex has six resonances. Four have chemical shifts less than 13.5 ppm, where GC pairs typically resonate. Two have chemical shifts above 13.5 ppm, where AU pairs typically resonate (Hilbers, 1979; Robillard & Reid, 1979; Pardi et al., 1981; Weiss et al., 1984; Ulrich et al., 1983; Chou et al., 1983). No new resonances are seen at 0 °C, although the lines broaden considerably. If all eight base pairs had a slowly exchanging imino proton, then eight resonances would be observed. The results suggest that the imino protons of the terminal AU base pairs exchange rapidly with water. This is commonly observed and attributed to fraying of the terminal base pair (Patel & Hilbers, 1975; Kan et al., 1975; Hilbers, 1979; Pardi et al., 1981; Chou et al., 1983; Weiss et al., 1984).

The rapid exchange of the terminal imino protons even at 0 °C is consistent with expectations based on thermodynamic parameters measured for terminal base pairs and unpaired terminal nucleotides (Freier et al., 1986b; Turner et al., 1988). In particular, the $\Delta G°$'s of duplex formation at 0 °C for GCGCA and UGCGCA are −13.0 and −13.3 kcal/mol, respectively (Freier et al., 1986b). Thus, providing a terminal 5′ U to hydrogen bond to the terminal 3′ A adds little to duplex stability. This suggests that an equilibrium will exist between hydrogen- and non-hydrogen-bonded conformations of the terminal base pairs. Presumably, the imino proton in the non-hydrogen-bonded conformation can exchange rapidly with water. For example, in 0.1 M Na^+, imino protons are not observed for the unpaired terminal nucleotides in

$$\begin{matrix} GCGAGCU \\ \bullet\bullet\ \ \bullet\bullet \\ UCGAGCG \end{matrix} \quad \text{and} \quad \begin{matrix} GCGAGCG \\ \bullet\bullet\ \ \bullet\bullet \\ GCGAGCG \end{matrix}$$

(SantaLucia et al., 1990). Thus, it is not surprising that no imino resonance is observed for the terminal pair of the $UGUAGA_3GUCA$ + UGACACUCA duplex.

Surprisingly, the NMR spectrum for $UGAGA_3GUCA$ alone has a single resonance at 13.8 ppm, where AU pairs are known to resonate. The resonance is weak, however, and disappears entirely by 22.5 °C. Comparison of its area with the area of the strongest duplex resonance, after normalization for strand concentration and acquisition parameters, suggests that it is at least 10 times weaker. Nevertheless, its presence suggests that single-stranded $UGAGA_3GUCA$ can have a structure involving a slowly exchanging, presumably hydrogen-bonded imino proton, at least some of the time.

Discussion

The main goal of this work is to test the suggestion by Papanicolaou et al. (1984) that asymmetric internal loops are less stable than symmetric internal loops. The free energy change for internal loop formation at 37 °C can be calculated from (Gralla & Crothers, 1973): $\Delta G°_{37}$(loop formation) = $\Delta G°_{37}$(duplex with loop) − $\Delta G°_{37}$(duplex without loop) + $\Delta G°_{37}$(interrupted nearest neighbor interaction. For example,

$$\Delta G°_{37}\left(\begin{matrix} CA_3C \\ \bullet\ \ \bullet \\ GA_3G \end{matrix}\right) = \Delta G°_{37}\left(\begin{matrix} UGAC^{A_3}CUCA \\ \bullet\bullet\bullet\bullet\ \ \bullet\bullet\bullet\bullet \\ ACUG_{A_3}GAGU \end{matrix}\right) -$$

$$\Delta G°_{37}\left(\begin{matrix} UGACCUCA \\ \bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet \\ ACUGGAGU \end{matrix}\right) + \Delta G°_{37}\left(\begin{matrix} CC \\ \bullet\bullet \\ GG \end{matrix}\right) = -6.67 - (-12.34) + (-2.93) = 2.7 \text{ kcal/mol}$$

The $\Delta H°$ and $\Delta S°$ values for internal loop formation can be calculated in analogous ways. The values derived are listed in Table II. Parameters from T_M^{-1} vs log C_T plots and from fits are listed. The T_M^{-1} vs log C_T results are considered the

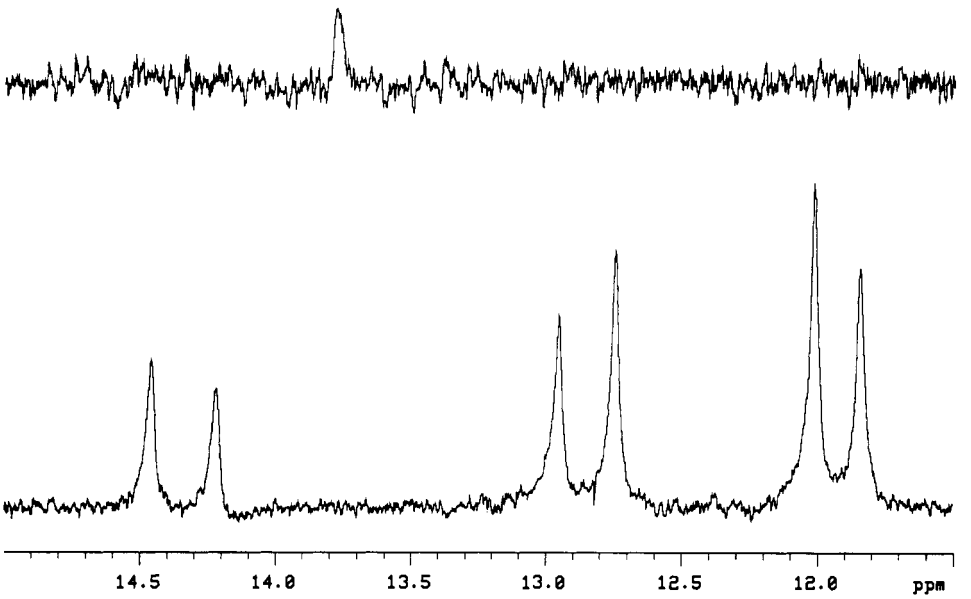


FIGURE 5: 500-MHz proton NMR spectra (11.5–15 ppm) at 15 °C in 1 M NaCl/10 mM sodium phosphate/0.5 mM Na_2EDTA, pH 7 for (top) 0.35 mM UGAGAAAGUCA and (bottom) 0.14 mM each strand UGACACUCA + UCAGAAAGUCA. For both sequences, spectra were also measured at 0.5 °C. Resonances broaden at lower temperatures, but no new resonances are observed for either of the sequences.

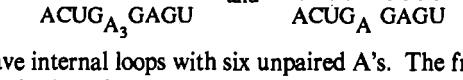
Table II: Thermodynamic Parameters of Internal Loops[a]

	$1/T_m$ vs log C_T			CURVE FITS		
	ΔG^0_{loop} (37 °C) (kcal/mole)	ΔH^0_{loop} (kcal/mole)	ΔS^0_{loop} (eu)	ΔG^0_{loop} (37 °C) (kcal/mole)	ΔH^0_{loop} (kcal/mole)	ΔS^0_{loop} (eu)
5'UGAC^{A}CUCA3' 3'ACUG$_{A}$GAGU5'	1.5 (1.7)	-3.2 (1.0)	-15.1 (-2.4)	1.2 (1.4)	3.2 (1.6)	6.4 (0.7)
5'UGAC ^{A} CUCA3' 3'ACUG$_{AA}$GAGU5'	3.0 (3.3)	6.0 (13.7)	9.6 (33.4)	2.4 (3.2)	10.9 (-3.1)	27.4 (-20.5)
5'UGAC^{AA}CUCA3' 3'ACUG $_{A}$GAGU5'	2.8 (3.0)	8.2 (9.6)	17.3 (21.3)	2.3 (2.5)	11.9 (10.0)	31.1 (24.1)
5'UGAC^{AA}CUCA3' 3'ACUG$_{AA}$GACU5'	2.4 (2.9)	-5.2 (7.4)	-24.3 (14.7)	1.9 (2.5)	5.1 (-3.7)	10.2 (-19.9)
5'UGAC ^{A} CUCA3' 3'ACUG$_{A3}$GAGU5'	2.9 (3.1)	7.2 (4.3)	13.9 (4.0)	2.3 (2.6)	11.1 (5.1)	28.3 (8.2)
5'UGAC^{A3}CUCA3' 3'ACUG $_{A}$GAGU5'	3.3 (3.4)	14.2 (17.7)	35.2 (46.2)	2.5 (2.8)	16.0 (14.5)	43.5 (37.5)
5'UGAC^{A3}CUCA3' 3'ACUG$_{AA}$GAGU5'	2.8 (3.1)	3.5 (7.2)	2.2 (13.1)	2.3 (2.6)	7.3 (4.3)	16.1 (5.4)
5'UGAC ^{A} CUCA3' 3'ACUG$_{A4}$GAGU5'	3.3 (3.6)	12.5 (11.3)	29.8 (24.9)	2.7 (3.1)	9.5 (4.7)	22.1 (5.0)
5'UGAC^{A4}CUCA3' 3'ACUG $_{A}$GAGU5'	3.9 (4.1)	16.5 (12.5)	40.9 (27.1)	3.1 (3.4)	17.1 (16.2)	51.4 (41.4)
5'UGAC^{AAA}CUCA3' 3'ACUG$_{AAA}$GAGU5'	2.7 (2.9)	4.5 (8.9)	5.7 (19.5)	2.2 (2.5)	-0.7 (-6.5)	-9.3 (-29.1)
5'UGAC^{AA}CUCA3' 3'ACUG$_{A4}$GAGU5'	3.3 (3.6)	12.3 (4.6)	29.0 (3.3)	2.8 (3.1)	8.5 (1.4)	18.3 (-5.7)
5'UGAC^{A4}CUCA3' 3'ACUG$_{AA}$GAGU5'	3.4 (3.7)	7.8 (9.7)	14.2 (19.4)	2.8 (3.4)	10.4 (0.4)	24.6 (-9.7)
5'UGAC ^{A} CUCA3' 3'ACUG$_{A5}$GAGU5'	3.7 (3.9)	12.0 (20.3)	26.6 (52.9)	2.8 (3.4)	17.3 (16.2)	46.9 (41.3)
5'UGAC^{A5}CUCA3' 3'ACUG $_{A}$GAGU5'	4.1 (4.6)	14.9 (11.4)	34.8 (22.0)	3.2 (3.7)	19.0 (13.3)	51.2 (30.8)
5'UGAC^{CAA}CUCA3' 3'ACUG$_{AAA}$GAGU5'	2.3 (2.6)	-3.6 (-5.1)	-18.8 (-24.8)	1.8 (2.3)	1.9 (-15.8)	0.4 (-58.2)
5'UGAC^{AAA}CUCA3' 3'ACUG$_{AAC}$GAGU5'	2.2 (2.5)	1.0 (7.1)	-3.9 (14.8)	1.7 (2.1)	-4.6 (-13.3)	-20.4 (-49.7)
5'CGC^{A}GCG3' 3'GCG$_{A}$CGC5'	1.1	-3.3	-14.1	1.2	1.2	0.5
5'CGC^{AA}GCG3' 3'GCG$_{AA}$CGC5'	1.8	-1.3	-9.9	1.8	8.8	22.6
5'CGC^{AAA}GCG3' 3'GCG$_{AAA}$CGC5'	2.3	0.8	-4.9	2.2	18.8	53.7
5'CGG^{AAA}CCG3' 3'GCC$_{AAA}$GGC5'	1.9	1.5	-1.2	1.9	5.5	11.6

[a] Solutions are in 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7. Parameters in parentheses are corrected for single-strand extinction coefficients.

most reliable (Hickey & Turner, 1985).

Symmetric Internal Loops Destabilize the Helix Less Than Asymmetric Loops. All internal loops studied here destabilize duplexes. The results in Table II indicate that symmetric loops destabilize a helix less than asymmetric loops. For example,



both have internal loops with six unpaired A's. The free energy changes for loop formation at 37 °C, however, are 2.7 and 4.1 kcal/mol, respectively.

The measured penalty for loop asymmetry is defined as $\Delta G°$(asymmetry penalty) = $\Delta G°$(measured for asymmetric loop of l) − $\Delta G°$(measured for symmetric loop of l). When the number of nucleotides in the loop, l, is odd, $\Delta G°$ (measured for symmetric loop) is taken as the average measured for the symmetric loops with $l - 1$ and $l + 1$ nucleotides. These penalties are listed in Table III and plotted in Figure 6.

Papanicolaou et al. (1984) suggested that asymmetric internal loops with branches of N_1 and N_2 nucleotides should be penalized by the minimum of 6.0 or $N f(M)$ kcal/mol, where $N = |N_1 - N_2|$, M is the minimum of 5, N_1, or N_2, and

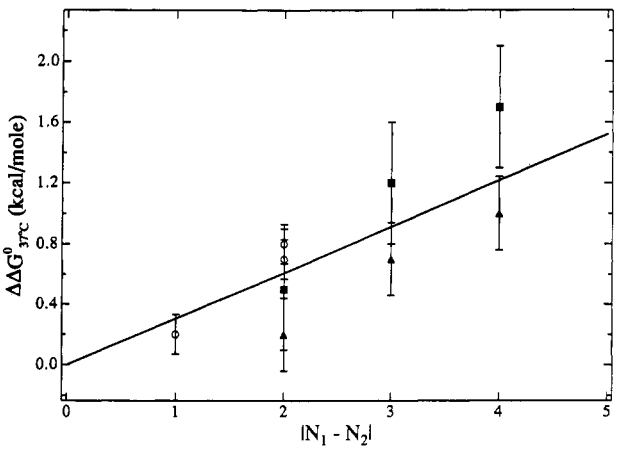


FIGURE 6: Asymmetry penalty vs difference between number of non-base-paired nucleotides on each strand for duplexes with UGAGAGUCA (■), for duplexes with UGACACUCA (▲), and for duplexes with either UGACAACUCA or UGAGAAGUCA (○). The line is the best fit of points with intercept forced to 0. The slope is 0.3. Values are calculated from duplex parameters derived with measured temperature dependence for single-strand extinction coefficients.

Table III: Penalties for Loop Asymmetry[a]

Molecule	N_1+N_2	$\lvert N_1-N_2\rvert$	Linear single strand baseline $\Delta\Delta G^\circ_{37}$ (kcal/mole)	Measured single strand baseline $\Delta\Delta G^\circ_{37}$ (kcal/mole)
5'UGAC A CUCA3' 3'ACUG$_{AA}$GAGU5'	3	1	1.1	1.0
5'UGAC^{AA}CUCA3' 3'ACUG $_{A}$ GAGU5'	3	1	0.9	0.7
5'UGAC A CUCA3' 3'ACUG$_{A3}$GAGU5'	4	2	0.5	0.2
5'UGAC^{A3}CUCA3' 3'ACUG $_{A}$ GAGU5'	4	2	0.9	0.5
5'UGAC^{A3}CUCA3' 3'ACUG$_{AA}$GAGU5'	5	1	0.3	0.2
5'UGAC A CUCA3' 3'ACUG$_{A4}$GAGU5'	5	3	0.8	0.7
5'UGAC^{A4}CUCA3' 3'ACUG $_{A}$ GAGU5'	5	3	1.4	1.2
5'UGAC^{AA}CUCA3' 3'ACUG$_{A4}$GAGU5'	6	2	0.6	0.7
5'UGAC^{A4}CUCA3' 3'ACUG$_{AA}$GAGU5'	6	2	0.7	0.8
5'UGAC A CUCA3' 3'ACUG$_{A5}$GAGU5'	6	4	1.0	1.0
5'UGAC^{A5}CUCA3' 3'ACUG $_{A}$ GAGU5'	6	4	1.4	1.7

[a] Solutions are in 1 M NaCl/10 mM sodium phosphate/0.1 mM Na_2EDTA, pH 7. Values are calculated using parameters from $1/T_M$ vs log C_T plots.

$f(1) = 0.7, f(2) = 0.6, f(3) = 0.4, f(4) = 0.2$, and $f(5) = 0.1$. Comparison with the results in Table II indicates that these penalties are too large. Thus, Jaeger et al. (1989) reduced the penalties by roughly a factor of 2 for a recent study of the accuracy of RNA secondary structure predictions. For loops with four to six nucleotides, the approximation of Jaeger et al. (1989) differs from the experimental penalty by an average of 0.2 kcal/mol. Slightly better agreement is obtained if the asymmetry penalty is simply set to $|N_1 - N_2| \times 0.3$ kcal/mol (see Figure 6). The plot in Figure 6, however, suggests that this is oversimplified. The asymmetry penalty depends on sequence as well as length: for loops larger than three, the asymmetry penalty for the GAG series is always larger than for the CAC series. The loop

CA C
• •
GA$_2$G

seems unusually unstable. Clearly, too little information is available to include such effects in a model for loop stability. A reasonable approximation of the data for loops larger than three, however, is to penalize asymmetry by the minimum of 3.0 or 0.3 $|N_1 - N_2|$ kcal/mol.

The results for loops of three nucleotides support a somewhat larger asymmetry penalty of about 0.9 kcal/mol. This suggests that existing free energy increments for loops of three (Jaeger et al., 1989; Turner et al., 1988) should be made less favorable, with the asymmetry penalty of $|N_1 - N_2| \times 0.3$ kcal/mol applied only to loops with more than three nucleotides. Of course, these effects may also depend on the sequence in and around the loop. Most internal loops, however, are purine rich so that the A loops studied here are reasonable models for loops closed by GC pairs.

The thermodynamic trends discussed above can be compared with known secondary structures of RNA to see if there is any correlation between stability and occurrence of internal loop motifs. Table IV lists the occurrence of various internal loops in a set of RNA structures determined by phylogenetic comparisons (Gutell et al., 1985; Gutell & Fox, 1988). Several of the trends in thermodynamic parameters are paralleled by trends in prevalence. The relatively unstable internal loop of three occurs less often than loops of two or four. Loops with higher asymmetry occur less often than loops with lower asymmetry. The comparison suggests that the improved thermodynamic parameters for internal loops may lead to improved predictions of RNA secondary structure.

This work has measured the effect of asymmetry in internal loops of A's in oligoribonucleotides. In large RNAs, however, there may be cases where asymmetric internal loops are stabilized by tertiary interactions. Such interactions cannot be included in a dynamic programming algorithm for free energy minimization to predict RNA secondary structure (Zuker, 1986). Fortunately, however, dynamic programming algorithms are able to generate an ensemble of structures within a free energy window (Steger et al., 1984; Williams & Tinoco, 1986; Zuker, 1989). Free energies for these "suboptimal" structures can then be recomputed with inclusion of tertiary interactions and other complex effects. Thus, it is important that energy parameters used in dynamic programming algorithms eliminate as many unfavorable structures as possible without eliminating the correct structure, which may be stabilized by tertiary interactions. A modest asymmetry penalty similar to the one reported here appears to perform this function well (Jaeger et al., 1989; Zuker et al., 1991). In particular, for a set of RNAs with known secondary structure, a structure that is on average 86% correct is found within 5% of the free energy of the predicted lowest free energy structure (Jaeger et al., 1989). The number of structures within this free energy window, however, is many orders of magnitude smaller than the total number of possible structures (Zuker et al., 1991).

The melting data indicate that the asymmetry effect on thermodynamic parameters is modest. The CD data indicate that the asymmetry effect on structure is also modest. Spectra for sequences with symmetric and asymmetric loops are similar and typical of A-form RNA spectra (Tunis-Schneider & Maestre, 1970). Thus, there appears to be no large structural perturbation induced by asymmetric internal loops. This suggests that the thermodynamic parameters measured should work well in the nearest-neighbor model for structure prediction (Tinoco et al., 1971; Turner et al., 1988).

Loop Stability Does Not Depend Simply on Loop Size or Stacking. Algorithms for predicting RNA structure require

Table IV: Frequency of Occurrence for Small Internal Loops in 16S- and 23S-like RNAs

type of RNA	no. of sequences	no. of nucleotides in loop and distribution													
				4		5		6			7			8	
		2	3	2 vs 2	3 vs 1	3 vs 2	4 vs 1	3 vs 3	4 vs 2	5 vs 1	4 vs 3	5 vs 2	6 vs 1	4 vs 4	other
16S[a]	21	175	56	48	41	46	5	23	1	0	10	6	1	3	17
23S[b]	38	488	166	151	54	97	22	87	21	6	26	21	2	39	23

[a] From Gutell et al. (1985). [b] From Gutell and Fox (1988).

a model for the sequence dependence of internal loop stability. Due to lack of experimental data, current models assume that stability depends on loop size (Salser, 1977; Freier et al., 1986a) or on loop size and stacking between the first mismatch and the base pair closing the loop (Jaeger et al., 1989). Studies of internal loops containing two GA mismatches have shown that these models are oversimplified because they neglect hydrogen-bonding interactions in GA mismatches (SantaLucia et al., 1990). The results in Tables I and II provide further evidence against these models. For symmetric loops of six, the measured loop free energy increments range from 1.9 to 2.7 kcal/mol, indicating that length does not completely determine stability. The loops

$$\begin{array}{cc} \text{CCAAC} & \text{CAAAC} \\ \bullet\ \ \bullet & \bullet\ \ \bullet \\ \text{GAAAG}, & \text{GAACG}, \end{array} \quad \text{and} \quad \begin{array}{c} \text{CAAAC} \\ \bullet\ \ \bullet \\ \text{GAAAG} \end{array}$$

all have the same stability within 0.5 kcal/mol, with the loop shown in the center being most stable. If the stacking effects suggested by Jaeger et al. (1989) dominated, then the loop shown on the right would be most stable by about 0.9 kcal/mol. Evidently, none of the current models for internal loops satisfactorily describe the new data.

Hairpins with Short Stems Closed by AU Pairs Are Less Stable Than Expected. The single strands, $UGACA_nCUCA$ and $UGAGA_mGUCA$, both have the potential to form hairpin structures. Current parameters for base pair and hairpin loop formation (Freier et al., 1986a; Groebe & Uhlenbeck, 1988; Jaeger et al., 1989) predict that most of these hairpins should have T_M's between 30 and 40 °C. If base pairs provide the only $\Delta H°$ for hairpin formation, then the CA_nC and GA_mG series are predicted to have $\Delta H°$'s of −24 and −18 kcal/mol, respectively. The experimental results, however, provide no evidence for such stable hairpins. When the melting curves for the individual strands are fit to a hairpin model, the $\Delta H°$'s average −12.5 and −13.5 kcal/mol for the CA_nC and GA_mG series, respectively. When the curves are fit to a single-strand stacking model, the derived $\Delta H°$'s range from −8 to −13 kcal/mol as expected for single-strand stacking. NMR spectra in the imino proton region for UGACACUCA and $UGAGA_3GUCA$ show no and one weak resonance, respectively, even though these sequences are predicted to form hairpins with T_M's of 33 and 36 °C from the parameters collected by Jaeger et al. (1989). The hairpin loop parameters are derived from studies of hairpins with long stems closed by GC pairs (Groebe & Uhlenbeck, 1988). The results suggest that hairpins with short stems closed by AU pairs are less stable. Previous work has also suggested that hairpins closed by AU pairs are less stable than those closed by GC pairs (Uhlenbeck et al., 1973; Gralla & Crothers, 1973; Tinoco et al., 1973).

Acknowledgments

We thank John SantaLucia, Jr., for stimulating discussions.

Supplementary Material Available

One table listing thermodynamic parameters for unimolecular single-strand melting in 1 M NaCl and one figure with T_M^{-1} vs log C_T plots for the $CGCA_nGCG$ series (2 pages). Ordering information is given on any current masthead page.

References

Albergo, D. D., & Turner, D. H. (1981) *Biochemistry 20*, 1413–1418.

Applequist, J. (1963) *J. Chem. Phys. 38*, 934–941.

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., Sturtevant, J. M., & Tinoco, I., Jr. (1975) *J. Mol. Biol. 99*, 549–565.

Chou, S.-H., Hare, D. R., Wemmer, D. E., & Reid, B. R. (1983) *Biochemistry 22*, 3037–3041.

Dewey, T. G., & Turner, D. H. (1979) *Biochemistry 18*, 5757–5762.

Freier, S. M., Hill, K. O., Dewey, T. G., Marky, L. A., Breslauer, K. J., & Turner, D. H. (1981) *Biochemistry 20*, 1419–1426.

Freier, S. M., Petersheim, M., Hickey, D. R., & Turner, D. H. (1984) *J. Biomol. Struct. Dyn. 1*, 1229–1242.

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.

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

Gregory, R. J., Cahill, P. B. F., Thurlow, O. L., & Zimmermann, R. A. (1988) *J. Mol. Biol. 204*, 295–307.

Groebe, D. R., & Uhlenbeck, O. C. (1988) *Nucleic Acids Res. 16*, 11725–11735.

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

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

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

Hilbers, C. W. (1979) in *Biological Applications of Magnetic Resonance* (Shulman, R. G., Ed.) Academic Press, New York.

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

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

Kan, L. S., Borer, P. N., & Ts's, P. O. P. (1975) *Biochemistry 14*, 4864–4869.

Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., & Freier, S. M. (1986) *Biochemistry 25*, 7840–7846.

Leng, M., & Felsenfeld, G. (1966) *J. Mol. Biol. 15*, 455–466.

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

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

Pardi, A., Martin, F. H., & Tinoco, I., Jr. (1981) *Biochemistry 20*, 3986–3996.

Patel, D. J., & Hilbers, C. W. (1975) *Biochemistry 14*, 2651–2656.

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

Pörschke, D. (1976) *Biochemistry 15*, 1495–1499.

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.

Robillard, G. T., & Reid, B. R. (1979) in *Biological Applications of Magnetic Resonance* (Shulman, R. G., Ed.) Academic Press, New York.

Rödel, G., Holl, J., Schmelzer, C., Schmidt, C., Schweyen, R. J., Weiss-Brummer, B., & Kaudewitz, F. (1983) in *Mitochondria 1983* (Schweyen, R., Wolf, K., & Kaudewitz, F., Eds.) pp 191-201, Walter de Gruyter & Co., Berlin.
Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol. 42*, 985-1002.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813-8819.
Steger, G., Hofmann, H., Förtsch, J., Gross, H. J., Randles, J. W., Sänger, H. L., & Riesner, D. (1984) *J. Biomol. Struct. Dyn. 2*, 543-571.
Stern, S., Weiser, B., & Noller, H. F. (1988) *J. Mol. Biol. 204*, 447-481.
Sutherland, J. C., & Griffin, K. P. (1983) *Biopolymers 22*, 1445-1448.
Sutherland, J. C., Lin, B., Mugavero, J., Trunk, J., Tomasz, M., Santella, R., Marky, L., & Breslauer, K. J. (1986) *Photochem. Photobiol. 44*, 295-301.
Tinoco, I., Jr., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature (London) 230*, 362-367.
Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) *Nature (London), New Biol. 246*, 40-41.
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. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 123-133.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167-192.
Uhlenbeck, O. C., Borer, P. N., Dengler, B., & Tinoco, I., Jr. (1973) *J. Mol. Biol. 73*, 483-496.
Ulrich, E. L., John, E.-M. M., Gough, G. R., Brunden, M. J., Gilham, P. T., Westler, W. M., & Markley, J. L. (1983) *Biochemistry 22*, 4362-4365.
Varani, G., Wimberly, B., & Tinoco, I., Jr. (1989) *Biochemistry 28*, 7760-7772.
Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Nucleic Acids Res. 12*. 4035-4047.
Williams, A., Jr., & Tinoco, I., Jr. (1986) *Nucleic Acids Res. 14*, 299-315.
Zimm, B. H., & Bragg, J. K. (1959) *J. Chem. Phys. 31*, 526-535.
Zuker, M. (1986) *Lect. Math. Life Sci. 17*, 87-124.
Zuker, M., Jaeger, J. A., & Turner, D. H. (1991) *Nucleic Acids Res.* (in press).

Photoaffinity Labeling of the Primer Binding Domain in Murine Leukemia Virus Reverse Transcriptase†

Radhakrishna S. Tirumalai and Mukund J. Modak*

Department of Biochemistry and Molecular Biology, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, New Jersey 07103

Received December 3, 1990; Revised Manuscript Received April 10, 1991

ABSTRACT: We have labeled the primer binding domain of murine leukemia virus reverse transcriptase (MuLV RT) by covalently cross-linking 5′ end labeled $d(T)_8$ to MuLV RT, using ultraviolet light energy. The specificity and the functional significance of the primer cross-linking reaction were demonstrated by the fact that (i) other oligomeric primers, tRNAs, and also template-primers readily compete with radiolabeled $d(T)_8$ for the cross-linking reaction, (ii) under similar conditions, the competing primers and template-primer also inhibit the DNA polymerase activity of MuLV RT to a similar extent, (iii) substrate deoxynucleotides have no effect, and (iv) the reaction is sensitive to high ionic strength. In order to identify the primer binding domains/sites in MuLV RT; tryptic digests prepared from the covalently cross-linked MuLV RT and $[^{32}P]d(T)_8$ complexes were resolved on C-18 columns by reverse-phase HPLC. Three distinct radiolabeled peptides were found to contain the majority of the bound primer. Of these, peptide I contained approximately 65% radioactivity, while the remainder was associated with peptides II and III. Amino acid composition and sequence analyses of the individual peptides revealed that peptide I spans amino acid residues 72-80 in the primary amino acid sequence of MuLV RT and is located in the polymerase domain. The primer cross-linking site appears to be at or near Pro-76. Peptides II and III span amino acid residues 602-609 and 615-622, respectively, and are located in the RNase H domain. The probable cross-linking sites in peptides II and III are suggested to be at or near Leu-604 and Leu-618, respectively.

Moloney murine leukemia virus reverse transcriptase (MuLV RT)[1] is a single-subunit protein of M_r 80 000 and catalyzes both RNA- and DNA-directed DNA polymerase activities as well as a ribonuclease H (RNase H) activity (Dickson et al., 1982). The primary amino acid sequence of MuLV RT has been deduced from the nucleotide sequence of a noninfectious proviral DNA (Shinnick et al., 1981), and the exact location of the genome segment coding for RT within the Pol gene has been deciphered and confirmed by NH_2-

† This research was supported in part by a grant from the National Institute of Allergy and Infectious Diseases (AI-26652) and by a postdoctoral fellowship award (to R.S.T.) from the New Jersey Commission on Cancer Research.

* Address correspondence to this author.

[1] Abbreviations: MuLV RT, murine Moloney leukemia virus, reverse transcriptase; AMV RT, avian myeloblastosis virus reverse transcriptase; DNA pol I, *Escherichia coli* DNA polymerase I; dNTP, deoxynucleoside 5′-triphosphate; 8-azido-dATP, 8-azidodeoxyadenosine 5′-triphosphate; DTT, dithiothreitol; HPLC, high-performance liquid chromatography; TPCK, tosylphenylalanine chloromethyl ketone; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; SDS, sodium dodecyl sulfate; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid; tRNA, transfer RNA.

0006-2960/91/0430-6436$02.50/0