Isolation and Characterization of Thermodynamically Stable and Unstable RNA Hairpins from a Triloop Combinatorial Library

Zhanyong Shu and Philip C. Bevilacqua*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

Received July 29, 1999; Revised Manuscript Received September 22, 1999

ABSTRACT: Hairpins are the most common elements of RNA secondary structure, playing important roles in RNA tertiary architecture and forming protein binding sites.Triloops are common in a variety of naturally occurring RNA hairpins, but little is known about their thermodynamic stability. Reported here are the sequences and thermodynamic parameters for a variety of stable and unstable triloop hairpins. Temperature gradient gel electrophoresis (TGGE) can be used to separate a simple RNA combinatorial library based on thermal stability [Bevilacqua, J. M., and Bevilacqua, P. C. (1998) *Biochemistry 45*, 15877−15884]. Here we introduce the application of TGGE to separating and analyzing a complex RNA combinatorial library based on thermal stability, using an RNA triloop library. Several rounds of in vitro selection of an RNA triloop library were carried out using TGGE, and preferences for exceptionally stable and unstable closing base pairs and loop sequences were identified. For stable hairpins, the most common closing base pair is CG, and U-rich loop sequences are preferred. Closing base pairs of GC and UA result in moderately stable hairpins when combined with a stable loop sequence. For unstable hairpins, the most common closing base pairs are AU and UG, and U-rich loop sequences are no longer preferred. In general, the contributions of the closing base pair and loop sequence to overall hairpin stability appear to be additive. Thermodynamic parameters for individual hairpins determined by UV melting are generally consistent with outcomes from selection experiments, with hairpins containing a CG closing base pair having a $\Delta\Delta G^\circ_{37}$ 2.1−2.5 kcal/mol more favorable than hairpins with other closing base pairs. Sequences and thermodynamic rules for triloop hairpins should aid in RNA structure prediction and determination of whether naturally occurring triloop hairpins are thermodynamically stable.

Hairpins are the most abundant secondary structural elements in RNA. For example, nearly 70% of *Escherichia coli* ribosomal RNA nucleotides are found in hairpin structures (*1*), and essentially all in vitro transcripts have stem−loops. Hairpin loops provide nucleation sites for RNA folding and transcription termination sites, interact with proteins, and stabilize RNA tertiary structure (*2*).

In the hairpin structures of 16S-like rRNA variable regions, considered unlikely to be protein binding sites, four is the most common loop size and accounts for 60% of the loops in bacteria and 46% of loops in eukaryotes (*3*). Several classes of thermodynamically stable tetraloops have been studied by phylogenetic comparison (*4*), UV melting (*5, 6*) and three-dimensional structure determination (*7, 8*).

Triloops occur less frequently than tetraloops, but are still common in a variety of naturally occurring RNAs, and have been demonstrated in DNA (*9, 10*). For example, in the 16S-like rRNA variable regions, triloops account for 7% of the loops in bacteria and 16% in eukaryotes (*3*). A survey of comparatively derived 16S and 16S-like rRNA structures revealed a small percentage of triloops in bacteria and up to 10% in eukaryotes (*11*). These observations suggest that triloops may possess particularly important functions. Unfortunately, the three-dimensional structures of only a few authentic RNA triloop hairpins have been reported (*12*−*14*), and to our knowledge, only a limited number have been thermodynamically characterized (*1, 15, 16*). Initial studies comparing A3/A4, C3/C4, and U3/U4 loops revealed similar values for ΔG°_{37} for hairpin formation (*16*), suggesting that triloops could potentially be stable. In addition, with the exception of CCC, which is penalized 1.4 kcal/mol for ΔG°_{37}, triloops are currently being modeled as sequence-independent in their contributions to RNA thermodynamic stability (*17*). These observations suggest that triloop hairpins deserve more attention.

A major limitation to the thermodynamic investigation of complex nucleic acid motifs has been handling the large number of possible sequence combinations. For example, a motif with *n*-different positions has 4^n sequence combinations, and flanking nucleotides need to be considered in the motif since they often influence overall stability. In principle, combinatorial chemistry provides a way to approach this type of problem. For nucleic acids, in vitro selection, or SELEX, is a combinatorial method that has been developed to isolate RNA and DNA molecules that specifically bind ligands and proteins (*18, 19*), catalyze chemical reactions (*20*), and have

* To whom correspondence should be addressed. Phone: (814)863-3812. Fax: (814)863-8403. E-mail: pcb@chem.psu.edu.

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; M1, method 1; M2, method 2; N = A, C, G, or U; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; R = A or G; RT, reverse transcription; $TBEK_{50}$, Tris−boric acid−EDTA−50 mM KCl buffer; TE, Tris−EDTA; TGGE, temperature gradient gel electrophoresis; T_M, melting temperature; Tris, tris(hydroxymethyl)aminomethane; UV, ultraviolet.

10.1021/bi991774z CCC: $18.00
Published on Web 10/23/1999

enhanced tertiary stability (*21*). An effective experimental method for separating nucleic acids based on thermodynamic stability is temperature gradient gel electrophoresis (TGGE). This method has been shown to be capable of resolving RNA and DNA molecules with small differences in free energy (*22*−*25*). Recently, we described using TGGE to separate a simple RNA combinatorial library with four members into stable and unstable molecules (*25*). Here we report applying TGGE to in vitro selection of RNA. A complex combinatorial library with 384 members has been used in this study, and exceptionally stable and unstable hairpins have been isolated and characterized.

MATERIALS AND METHODS

TGGE. Experiments were carried out on perpendicular and parallel TGGE apparatuses as previously described (*25*). The temperature gradient was measured by inserting an Omega HH21 microprocessor thermometer into the gel. The gradient was linear with position and did not fluctuate over the course of a 2-h electrophoresis run.

Standard gel conditions were 8% polyacrylamide (acrylamide:bisacrylamide = 29:1) and 2 M urea. The buffer in the gel and reservoirs was 1× $TBEK_{50}$ (= 100 mM Tris−83 mM boric acid−1 mM Na_2EDTA−50 mM KCl, pH 8.8, at 20 °C). Gels were pre-electrophoresed for 30 min at 350 V with both water baths running prior to the loading of samples. At the start of each experiment, the appropriate mixture of RNAs was renatured in 1× TE (= 10 mM Tris−1 mM Na_2EDTA, pH 7.5) by heating for 3 min at 90 °C, followed by cooling on the bench for at least 10 min to minimize any dimerization. Samples were then mixed with TGGE loading buffer (= 1× $TBEK_{50}$, 10% glycerol, 0.05% xylene cyanol, and 0.05% bromophenol blue), with dyes omitted from parallel TGGE samples. For parallel TGGE, samples were loaded into 10-mm lanes, using 2 μL for analytical gels and 7 μL for preparative gels. Gels were run at 350 V for 1.5−2 h depending on the temperature gradient. Analytical gels were dried on Whatman 3M paper and analyzed using a PhosphorImager (Molecular Dynamics).

Preparation of RNA. The DNA templates for synthesis of RNA hairpins were prepared and purified as previously described (*25*). Positions of randomization for selection experiments were prepared at the DNA level. (i) Positions 20−22 of the RNA hairpins (Figure 1A) were completely randomized by mixing the four phosphoramidites for a single coupling. (ii) Positions 19 and 23 (Figure 1A) were partially randomized and contained complementary bases once transcribed. Partial randomization was achieved by preparing four independent triloop sublibraries, one each for the closing base pairs of CG, GC, UR, or RU (R is a mixture of A and G), each with complete randomization at positions 20−22. Each of the triloop sublibraries was transcribed as described below, and its concentration was determined spectrophotometrically. Sublibraries of CG, GC, UR, and RU were mixed in a 1:1:2:2 ratio to prepare a triloop hairpin library having approximately equal populations of the six closing base pairs, referred to as pool 0. The RNA mixture resulting after *n* rounds of selection is referred to as pool *n*.

Hairpins were synthesized by T7 transcription reactions using a top-strand promoter and phage T7 RNA polymerase (*26*) and were purified by denaturing PAGE as previously

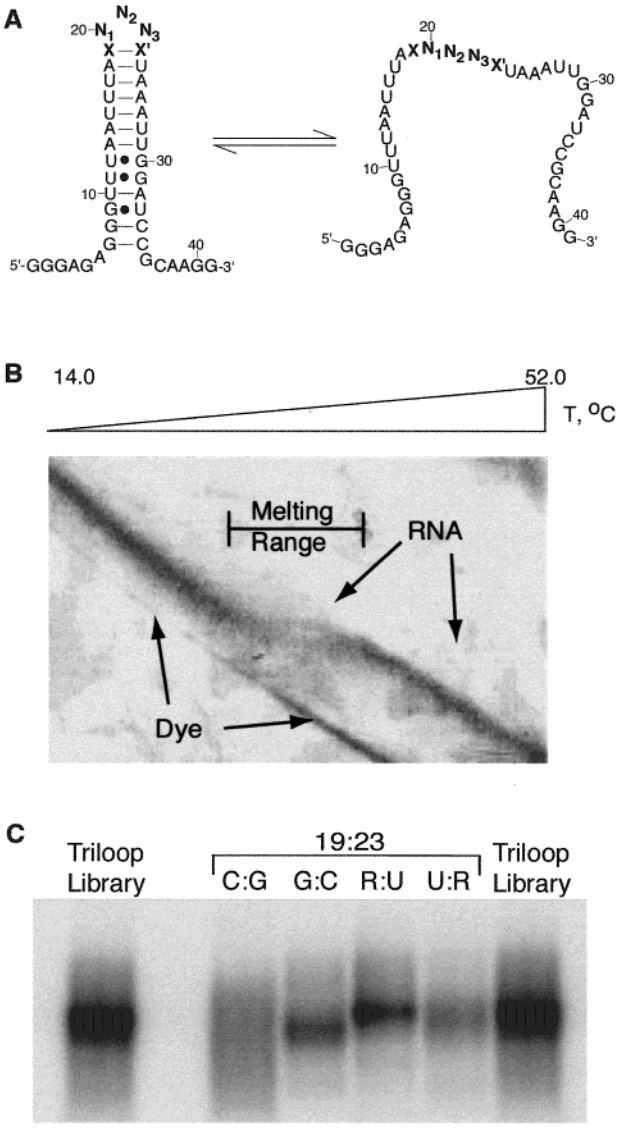

FIGURE 1: Melting of the triloop library. The triloop library is a mixture of 384 different sequences. (A) Equilibrium between folded and unfolded forms of RNA hairpins. Watson−Crick base pairs are denoted with a dash, and wobble pairs are denoted with a dot. The letter N denotes a mix of A, C, G, and U; 1, 2, and 3 are the three positions of randomization. The letters X and X′ denote complementary bases. (B) Perpendicular TGGE of the RNA triloop library containing equal amounts of hairpins with CG, GC, UA, AU, UG, and GU closing base pairs. The temperature in the gel ranged from 14 to 52 °C. This gel was visualized by staining in Stains-all. "Dye" indicates the presence of bromophenol blue and is included to demonstrate the linearity of the temperature gradient. "RNA" indicates the presence of stained RNA. The melting transition is between ≈27 and 37 °C and is indicated with the label "Melting Range". (C) Parallel TGGE of the pool 0 triloop library and sublibraries. The temperature baths were set at 27 and 37 °C. Triloop sublibraries have complete randomization at positions 20−22, and positions 19 and 23 are complementary and have bases as indicated in the figure. R is a mixture of purines.

described (*25*). The RNA oligomers were stored in TE if used for selection or in standard melting buffer $P_{10}E_{0.1}$ (= 10 mM phosphate−0.1 mM Na_2EDTA, pH 7.0) for UV melting. Prior to TGGE, the pool 0 RNA triloop library was dephosphorylated and 5′-end labeled with polynucleotide kinase and [γ-^{32}P]ATP. Radiolabeled RNAs were purified and quantitated as previously described (*25*). During the three rounds of selection, RNA samples were obtained by RT-

PCR and T7 transcription and body-labeled by including [α-^{32}P]ATP in the transcription reaction.

Selection Procedure. RT-PCR reactions were performed similarly to those previously described (*27*). Briefly, selected RNAs were annealed with the bottom strand primer (5′-CCTTGCGGATCCAATTT-3′), and cDNAs were obtained by reverse transcription for 30 min at 60 °C using AMV reverse transcriptase. To this solution, 50 μL of a master mix was added containing the top-strand primer (5′-GGCCGAATTCTAATACGACTCACTATAGGGAGA-GGGTTTAATTT-3′). This mixture was heated at 95 °C for 3 min prior to addition of a *Taq* DNA polymerase plus a proofreading fragment (Clontech), and PCR was performed for 20 cycles.

For selection, preparative parallel TGGE gels were exposed to film at 4 °C without drying, and bands were visualized by autoradiography. The desired RNA species was excised with a razor blade and recovered by crush and soak. After completing the necessary rounds of selection, PCR fragments were digested with *Eco*RI and *Bam*HI and cloned into pUC19. Sequences of isolated clones were determined by the dideoxy method. The fixed-sequence portion of the library upstream of the *Bam*HI site (Figure 1A) was also examined, and only sequences without mutations in this region were used.

UV Melting Experiments. The RNAs used in UV melting studies were minihairpins with the general sequence $GGAXL_1L_2L_3X'UCC$, where X and X′ are complementary closing bases and L_1-L_3 are the loop bases. The RNAs were prepared by T7 transcription, gel purified, concentrated by ethanol precipitation, and dissolved in $P_{10}E_{0.1}$. High concentrations of NaCl, 0.1 or 1 M, resulted in severe problems with concentration-dependent melts and were thus avoided. UV absorbance melting profiles were obtained at 260 nm using a Gilford Response II spectrophotometer and 5- or 10-mm path length cuvettes, with a heating rate of approximately 1 °C/min. At the start of each experiment, the desired RNA was renatured in melting buffer by heating for 3 min at 90 °C followed by cooling to room temperature on the bench for at least 15 min. Generally, melting was from low to high temperature. Representative minihairpins and long hairpins from selection were also melted from high to low temperature, and similar thermodynamic parameters were obtained, suggesting that melting is reversible. Thermodynamic parameters were obtained by fitting to a two-state model with sloping baselines using a nonlinear least-squares program, which provided good fits to the data (*28*). Melting temperatures for pool 0, stable, and unstable RNA mixtures were determined from the maximum in the first derivative curves. Concentrations were calculated using absorbance values collected at 90 °C and extinction coefficients from a nearest-neighbor analysis (*29, 30*).

Surveying of rRNA Database. A database of 95 small subunit and 112 large subunit rRNA secondary structures was surveyed (*31*−*33*). 703 loops designated as triloops were found and sorted based on loop sequence and closing base pair.

RESULTS

Design of the RNA Triloop Library. An RNA triloop hairpin will be formally defined herein as having three nucleotides in a loop closed by a Watson−Crick or wobble pair. Since the relative stabilities of tri-and pentaloop hairpins are unknown, we wanted to limit the possible number of pentaloops by preventing mismatches between positions 19 and 23 (Figure 1A). This was achieved by mixing four triloop sublibraries containing CG, GC, UR, and RU pairs at positions 19 and 23 in a 1:1:2:2 ratio. Positions 20−22 in the loop were completely randomized to obtain all 64 possible triloops, thus creating a library with a total of 384 (= 6 × 64) different sequences. Another advantage of this method of preparation is that it allowed us to preview the average stabilities of hairpins containing different closing base pairs as described below. The RNA triloop library (Figure 1A) was designed to minimize alternative hairpin pairings upon folding and to facilitate RT-PCR reactions (*25*). Binding sites for primers extend to within one nucleotide of the randomized region to minimize potential problems with reverse transcriptase extending through strong secondary structures (*34*). In addition, reverse transcriptase reactions were at 60 °C to help melt any residual RNA secondary structure.

TGGE Analysis of the RNA Triloop Library and Its Sublibraries. The RNA triloop library was electrophoresed by perpendicular TGGE under standard conditions (Figure 1B). The gel was run in 2 M urea to favor melting in the center of the 14.0−52.0 °C range. Melting temperatures for a series of model hairpins have been shown to have very similar linear dependencies on the concentration of urea in the gel (*25*). The transition occurred between roughly 27 and 37 °C and is broad in nature due to the mixture of different hairpins (Figure 1B). Sloping baselines are due to a general increase in electrophoretic mobility with temperature, as evidenced by similar mobilities for a tracking dye (Figure 1B) and observed for fixed-sequence hairpins (*25*).

On the basis of the perpendicular TGGE results, the RNA triloop library and its closing base pair sublibraries were electrophoresed by parallel TGGE under standard conditions with temperature baths set at 27 and 37 °C (Figure 1C). On parallel TGGE gels, stable hairpins generally have greater mobility than unstable hairpins (*25*). The bulk of the RNA triloop library ran as a distinct, albeit wide, band, suggesting that most hairpins have fairly similar thermal stability. A small amount of smearing below and above the main band also occurred, suggesting that the library may contain a small population of exceptionally stable and unstable members. Comparison of the electrophoretic mobilities of the CG, GC, RU, and UR sublibraries suggests that the CG closing base pair sublibrary, due to its smear near the bottom of the gel, may contain many of the exceptionally stable RNA triloop hairpins. A substantial portion of the CG sublibrary also runs equal to or above the main band for the RNA triloop library, indicating that this population is very heterogeneous and that not all hairpins closed by a CG base pair are stable. Hairpins with GC and RU closing base pairs appear more homogeneous in terms of their thermal stability; hairpins with UR closing base pairs appear intermediate in homogeneity; and, on average, GC pairs appear to be somewhat more stable than those with RU and UR pairs.

Selection of the RNA Triloop Library. Two different selection methods were carried out, both based upon separations by parallel TGGE. During each round, the top and bottom parts of the gel smear were excised as the unstable

and stable species, respectively. For method 1, parallel gels for each round were run over the same 10° temperature gradient, with baths set at 27 and 37 °C, and the gel was excised toward the very outer parts of the smear, away from the main band.

Running of parallel gels over a 10° temperature window with baths set at higher temperatures gave better resolution of high mobility (i.e., stable) bands than gels with baths set at lower temperatures. Likewise, baths set at the lower temperatures gave better resolution of slower mobility (i.e., unstable) bands than gels with baths set at higher temperatures. A second method was developed to take advantage of these observations. In method 2, for selection of stable species, the gels were run with a 10° temperature window with baths set at increasing temperatures of 27−37, 30−40, and 32−42 °C for rounds 1−3, respectively. Likewise, unstable species were selected by running gels with baths set at decreasing temperatures of 27−37, 24−34, and 22−32 °C for rounds 1−3, respectively. Also, for method 2, the gel was excised somewhat closer to the center of the smear than method 1. After three rounds of selection, distinct bands of fast and slow mobility were found for stable and unstable populations from both selection methods (Figure 2A), and individual clones were obtained and sequenced.

Unstable-M1 (= unstable hairpins collected from method 1) ran similar to unstable-M2, and both consisted of one band, suggesting they may contain species with similar thermal stabilities (Figure 2A). Stable-M2, on the other hand, consisted of two bands of similar intensity, while stable-M1 had only 1 band. By another round of selection, stable-M2 was separated into an upper band (stable-M2A) and a lower band (stable-M2B) with approximately 70% efficiencies (Figure 2A). When the collection of all stable and unstable species in Figure 2A was electrophoresed under the denaturing conditions, 2 M urea and 65 °C, all bands were sharp with identical mobilities (data not shown). This observation is consistent with single and identical lengths in the different selected populations and indicates that separation is dependent on the temperature gradient.

To test whether the collection of fast and slow mobility species had appropriate thermal stabilities under equilibrium conditions, UV melting experiments were carried out on the selected mixtures (Figure 2B). Melting profiles resulted in average T_M values of 42.5, 37.0, and 35.5 °C for stable-M1, pool 0, and unstable-M1 populations, respectively, consistent with predicted stabilities based upon mobilities in parallel TGGE. Moreover, transitions for stable- and unstable-M1 were steeper than for pool 0, consistent with isolation of a limited set of hairpins with similar thermodynamic behavior. Differences in T_M values relative to pool 0 suggest that stable hairpins are substantially more stable than the total population of hairpins (42.5 vs 37.0 °C) but that unstable hairpins are only modestly less stable than the total population (35.5 vs 37.0 °C).

Identification of Stable and Unstable RNA Hairpins. In total, 141 different clones from various populations were isolated and sequenced, and sequence distributions are summarized in Tables 1−5. Table 1 provides a distribution of the closing base pairs found in different hairpins. Among the clones in stable-M1, 65/67 (= 97%) have a CG closing base pair, while the remaining two have a GC closing base pair. Interestingly, the two clones with a GC closing base pair also have UUU loop sequences. Stable-M2B also consists primarily of CG closing base pairs with 16/19 CG, 1/19 GC, and 2/19 UA closing base pairs (Table 1). As expected based on its slower mobility (Figure 2A), stable-M2A appears to be a collection of moderately stable hairpins

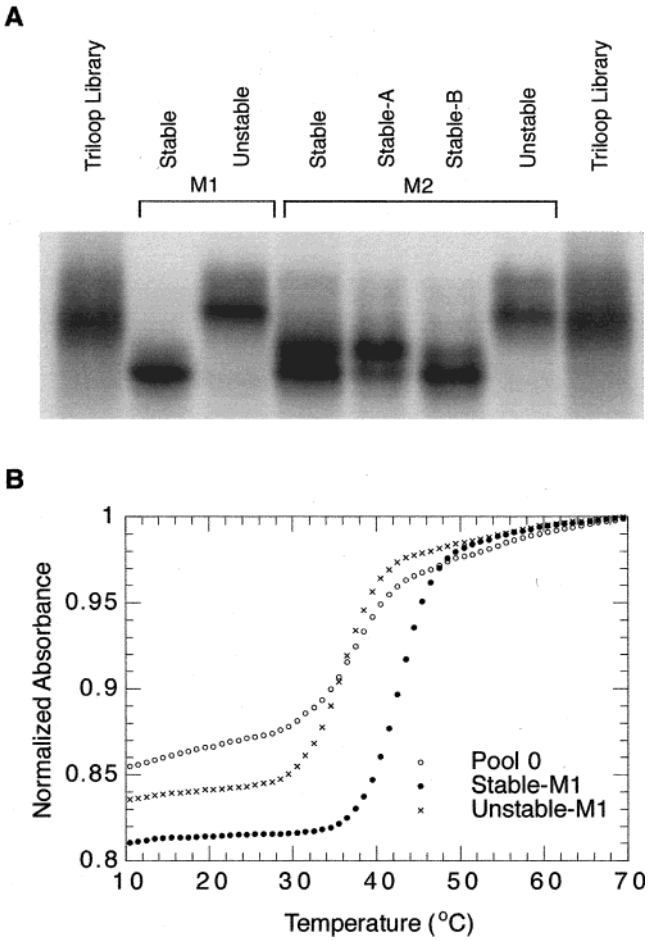

FIGURE 2: Selection results in stable and unstable populations. Stable and unstable samples were obtained through 3−4 rounds of selection. M1 and M2 signify samples obtained via selection methods 1 and 2, respectively. (A) Parallel TGGE gel under standard conditions with temperature baths set at 30 and 37 °C. "Stable" and "Unstable" signify high and low mobility bands, respectively. Stable-M2 has two bands that were separated into stable-M2A and -M2B by one further round of selection. (B) Plot of normalized absorbance versus temperature. UV melting of RNA mixtures is in standard melting buffer $P_{10}E_{0.1}$ (10 mM phosphate, 0.1 mM Na_2EDTA, pH 7.0). The concentration of hairpins was approximately 2 μM. Apparent T_M values for pool 0, stable-M1, and unstable-M1 were 37.0, 42.5, and 35.5 °C, respectively. Only every other point is plotted to aid visualization of symbols.

Table 1: Closing Base Pair Distribution in Hairpins from Selection[a]

X−X′ [b]	stable-M1	stable-M2B	stable-M2A	unstable-M1 and -M2
C−G	65	16	7	0
G−C	2	1	7	2
U−A	0	2	5	1
G−U	0	0	0	5
A−U	0	0	0	13
U−G	0	0	0	15

[a] The total number of clones sequenced was 67, 19, 19, and 36 for stable-M1, stable-M2B, stable-M2A, and unstable-M1 and -M2, respectively, where M*i* denotes method *i*. Closing base pairs are listed in order of stabilities based upon occurrences in stable and unstable libraries. [b] X and X′ denote complementary bases.

Table 2: Loop Sequence Distribution in Stable Hairpins from Selection

X−X′ [a]	loop sequence [b]	occurrences	X−X′ [a]	loop sequence [b]	occurrences
C−G [c]	UUA	11	CG [d]	AAC	1
	UUU	9		AAU	1
	UGU	8		CGU	1
	GUU	6		UGC	1
	UUG	6		UGU	1
	AUU	4		UUC	1
	CUG	4		UUU	1
	AGU	3	GC [c]	UUU	2
	AUC	2		CUA	1
	GAG	2	GC [d]	UUU	2
	GCU	2		AUG	1
	GUC	2		CUG	1
	GUG	2		GUU	1
	UAU	2		UCC	1
	UCG	2		UUG	1
	UCU	2	UA [d]	UUU	2
	UGC	2		UCG	1
	AAU	1		UUA	1
	ACA	1		UUG	1
	ACU	1	UA [e]	UUA	1
	AUA	1		UUG	1
	CAU	1			
	CGA	1			
	CUA	1			
	GCA	1			
	GCC	1			
	GGU	1			
	UCC	1			
	UUC	1			

[a] X and X′ denote complementary bases. [b] Listed by occurrences, then alphabetically. [c] From stable-M1 and -M2B. [d] From stable-M2A. [e] From stable-M2B only.

containing 7/19 CG, 7/19 GC, and 5/19 UA closing base pairs (Table 1). The two UA sequences found in stable-M2B may, in fact, be due in part to 30% contamination by the moderately stable-M2A stable band (Figure 2A). From these data, it appears that the most stable triloop hairpins have a CG closing base pair and that GC and UA closing base pairs afford moderately stable triloop hairpins.

The unstable species from methods 1 and 2 have similar mobilities on a 27−37 °C gel, suggesting they have similar stabilities and will therefore be treated together. In total, 36 clones of unstable triloop hairpins were sequenced, and closing base pairs found to be stable above were underrepresented in this pool (Table 1). For example, 0/36 CG and only 2/36 GC and 1/36 UA closing base pairs were found, further supporting the stability of triloop hairpins with these closing base pairs. In contrast, UG and AU closing base pairs, which were not in the stable species, occurred 15/36 and 13/36 times, respectively, suggesting that triloop hairpins with UG and AU closing base pairs are unstable.

The distribution of loop sequences for stable hairpins was examined (Table 2). In the 81 clones from stable-M1 and -M2B with a CG closing base pair, 29 different kinds of loop sequences were found. Of these loop sequences, 8/29 occur three or more times and make up 63% of the population, indicating that certain loop sequences are overrepresented. The most common loop sequences are U-rich (Table 2). At each position in the loop, U is the most common base, occurring 44/81, 49/81, and 40/81 at positions 1−3, respectively (Table 3). The eight different stable loop sequences with a GC closing base pair from stable-M1 and -M2 are U-rich, as are the six different stable loop sequences with a UA closing base pair from stable-M2A and -M2B (Table 2). These results suggest that U-rich loop sequences afford stable hairpins when combined with CG, GC, or UA closing base pairs.

Table 3: Nucleotide Distribution in Stable Hairpins from Selection with a CG Closing Base Pair [a]

position	A	C	G	U
1	13	7	17	44
2	6	11	15	49
3	16	9	16	40

[a] The total number of sequences is 81. Includes stable sequences with a CG closing base pair from stable-M1 and -M2B.

Table 4: Loop Sequence Distribution in Unstable Hairpins from Selection [a]

loop sequence [b]	occurrences
AAC	5 [c]
GGC	3
UGC	3 [c]
GAC	2
GGU	2 [c]
GUU	2
AAG	1
AAU	1 [c]
ACA	1 [c]
ACG	1
ACU	1 [c]
AGC	1
AUC	1 [c]
CAG	1
CAU	1 [c]
CGC	1
CGG	1
CGU	1
GAU	1
GGA	1
GUA	1
GUG	1 [c]
UAC	1
UCC	1 [c]
UGG	1

[a] Summary of loop sequences isolated from unstable-M1 and -M2. The total number of clones sequenced was 36. Closing base pair distributions are 15/36 UG, 13/36 AU, 5/36 GU, 2/36 GC, 1/36 UA, and 0/36 CG. [b] Listed by occurrences and then alphabetically. [c] Also appears in stable loop sequences 1−2 times with a CG or GC closing base pair.

The distribution of loop sequences for unstable hairpins was also examined (Table 4). In the 36 unstable clones, 25 different kinds of loop sequences were found, suggesting less consensus than for stable loop sequences. Nevertheless, several trends are suggested by the data. First, with only one exception, none of the selected loop sequences contain more than a single U (Table 4), and U residues are no longer overrepresented in the loop (Table 5). This suggests that U-rich loops are unique to stable hairpins. In addition, there appears to be a preference for purines at positions 1 (26/36) and 2 (27/36) and for pyrimidines at position 3 (27/36) (Table 5). In particular, position 3 appears to display a preference for C as the pyrimidine since triloops that occur three or more times have a C here and because C occurs 17/36 times versus U at 10/36 at this position (Tables 4 and 5). Accordingly, AAC and GGC appear 5/36 and 3/36 times (Table 4), respectively, suggesting that these loop sequences

Table 5: Nucleotide Distribution in Unstable Hairpins from Selection[a]

position[b]	A	C	G	U
X	13	0	7	16
1	13	5	13	5
2	13	4	14	5
3	3	17	6	10
X′	1	2	15	18

[a] The total number of sequences is 36. [b] X and X′ denote complementary bases.

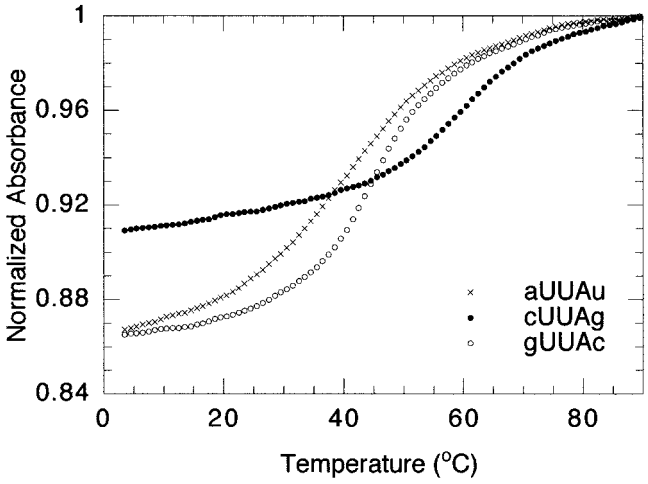

FIGURE 3: UV melting analysis of triloop minihairpins. Plot of normalized absorbance versus temperature. Melting is of minihairpins with the sequence GGAXUUAX′UCC, where X and X′ denote the closing bases. Substitutions of X and X′ are AU (×), CG (●), and GC (○). The buffer was the standard melting buffer, $P_{10}E_{0.1}$ (10 mM phosphate, 0.1 mM Na_2EDTA, pH 7.0), and the concentration of hairpins was approximately 6 μM. Only every other point is plotted to aid visualization of symbols.

may be exceptionally unstable. The few loop sequences found in both stable and unstable populations noted in Table 4 have CG or GC closing base pairs when stable and are found only one or two times in the stable loop sequences.

Stable-M2A appears to contain moderately stable hairpins, and the loop sequences for hairpins with a CG closing base pair are somewhat different from those in stable-M1 and -M2B (Table 2). The preference for U is not as strong, and the distribution of nucleotides at positions 1−3 is more like that for the unstable loops with AAC also appearing. This observation suggests that non-U-rich loop sequences are destabilizing when combined with a CG closing base pair as well.

Thermodynamic Characterization of Minihairpins. Selection experiments suggested closing base pairs and loop sequences for exceptionally stable and unstable triloop hairpins, some of which were selected for thermodynamic analysis by UV melting. The RNAs used for melting, referred to as minihairpins, have short stems of four base pairs to facilitate two-state behavior, minimize the tendency for dimerization, and maximize the percent contribution of the triloop to overall hairpin stability (Figure 3). For hairpins with a CG closing base pair, the loop sequences UUA, UUG, UUU, AUU, and GUU, which were common in the stable selection, have the lowest $\Delta G°_{37}$ and highest melting temperature, T_M (Table 6). In particular, UUA, the most common stable loop from selection, is also the most stable, having a $\Delta G°_{37}$ of −2.9 kcal/mol and a T_M of 61.3 °C. Other loop sequences, such as AUA, GCA, and CUA, which are common in rRNA, are also stable to moderately stable by the melting results. The loops AAC and GGC, which were found as unstable by selection (Table 4), are unstable when combined with a CG closing base pair as well. Apparently, unstable loop sequences selected with unstable closing base pairs are also unstable in the context of a CG closing base pair. In particular, AAC, the most common unstable loop from selection, is also the least stable loop, having a $\Delta G°_{37}$ of −1.1 kcal/mol and a T_M of 55.4 °C.

In the second part of Table 6, results of varying the closing base pair with the most stable loop sequence, UUA, are shown. With a UUA loop sequence, the most stable closing base pair is CG, having a $\Delta\Delta G°_{37}$ of −2.1 to −2.5 kcal/mol and a ΔT_M of 15.4−19.5 °C as compared to other closing base pairs. Melting experiments with an AAC loop resulted in concentration-dependent melts for all closing base pairs except CG, suggesting that oligomers were forming dimer complexes. While a quantitative analysis of this data set is not possible, the observation of concentration independence for a CG closing base pair-only is nevertheless consistent with extra stability of a CG for unstable loop sequences as well.

DISCUSSION

RNA sequence can be influenced by tertiary interactions and protein binding, the kinetics and uniqueness of folding, and the thermal stability of a particular fold (*17, 25, 35*−*39*). Evolution can result in RNA sequences that favor or disfavor one or more of these factors, and compromises between various factors may be needed. Examples are known in which thermodynamically stable RNA sequences and structures are preferred in nature (*5, 34, 40*). Stable structures could provide thermal stability, make a fold resistant to nucleolytic digestion, or promote transcription pausing or termination. Alternatively, an unstable structure may be selected to prevent misfolds or to expose bases for recognition by an RNA or protein (*40*). As such, identifying stable and unstable sequences may help reveal the roles thermal stability plays in vivo and aid in efforts to predict RNA structure from sequence.

Parallel TGGE Reveals Stable and Unstable Triloop Hairpins. Early thermodynamic studies on RNA tetraloops showed that hairpin stability is influenced by the loop sequence and the identity of the closing base pair (*5*). To account for this, the library used here contained complete randomization of the loop nucleotides and partial randomization of the nucleotides immediately 5′ and 3′ of the loop. Only nucleotides that had the ability to form Watson−Crick or wobble closing pairs were introduced at the closing positions (Figure 1A). Complete randomization of these two positions was not done since we wanted to study triloops and do not know the relative stabilities of triloops versus pentaloops. Despite this precaution, it is possible that some of the selected hairpins may involve either noncanonical pairing or no pairing of these two positions. This conclusion is based in part on NMR structures of the loop sequences GGGGC and GUUUC, which were found to have noncanonical pairing of first G and last C (*41, 42*). In addition,

Table 6: Thermodynamic Parameters for Minihairpin Formation Measured by UV Melting[a]

	sequence[b]	$\Delta G°_{37}$ (kcal mol^{-1})	$\Delta H°$ (kcal mol^{-1})	$\Delta S°$ (eu)	T_M (°C)	stable[c]	unstable[d]
$CL_1L_2L_3G$[e]	UUA	−2.9 ± 0.1	−40 ± 2	−121 ± 5	61.3	11	0
	AUU	−2.6 ± 0.1	−39 ± 1	−117 ± 2	59.5	4	0
	GUU	−2.6 ± 0.1	−38 ± 2	−115 ± 4	59.5	6	0
	AUA	−2.5 ± 0.1	−36 ± 1	−108 ± 3	59.7	1	0
	CUA	−2.5 ± 0.1	−37 ± 2	−110 ± 7	59.4	1	0
	UUU	−2.5 ± 0.2	−37 ± 4	−112 ± 12	59.2	9	0
	GUC	−2.4 ± 0.1	−37 ± 1	−113 ± 4	58.8	2	0
	UUG	−2.4 ± 0.2	−33 ± 3	−100 ± 8	60.8	6	0
	AAA	−2.3 ± 0.2	−35 ± 2	−105 ± 5	58.9	0	0
	UGU	−2.2 ± 0.1	−35 ± 1	−104 ± 3	57.8	8	0
	GCA	−2.1 ± 0.1	−31 ± 1	−94 ± 3	59.4	1	0
	CCC	−1.7 ± 0.1	−33 ± 1	−101 ± 3	53.5	0	0
	CUC	−1.6 ± 0.1	−31 ± 1	−97 ± 2	53.3	0	0
	GGC	−1.5 ± 0.1	−26 ± 1	−80 ± 4	55.2	0	3
	AAC[f]	−1.1 ± 0.1	−19 ± 3	−57 ± 10	55.4	0	5
$XUUAX'$[g]	CG	−2.9 ± 0.1	−40 ± 2	−121 ± 5	61.3	11	0
	GC	−0.8 ± 0.1	−31 ± 3	−97 ± 10	45.4	0	0
	UA	−0.7 ± 0.1	−26 ± 4	−80 ± 11	45.9	1	0
	UG	−0.5 ± 0.1	−24 ± 3	−74 ± 10	43.2	0	0
	AU	−0.4 ± 0.1	−23 ± 1	−74 ± 4	42.8	0	0
	GU	−0.4 ± 0.2	−24 ± 3	−75 ± 7	41.8	0	0

[a] Solutions were 10 mM sodium phosphate and 0.1 mM Na_2EDTA, pH 7.0. Melting is of minihairpins with the sequence $GGAXL_1L_2L_3X'UCC$, where X and X′ denote the closing bases and L_1, L_2, and L_3 denote the three loop bases. Melting experiments were performed over a range of 1−20 μM oligomer. Parameters are the average of at least three independently prepared samples between 2 and 20 μM, and errors are standard deviations. Maximal errors in T_M values are estimated at ±1 °C based on comparison of multiple melts. For the minihairpins in this table, the T_M was constant between ≈1 and 20 μM. If oligomers were forming self-complementary duplexes, calculations using nearest-neighbor thermodynamic values determined at 1 M NaCl predict that the T_M should change by 9 °C for this concentration range (*53*), well outside the error limits of the measured T_M values. [b] Listed by $\Delta G°_{37}$ and then alphabetically. [c] Number of stable sequences with specified closing base pair from stable-M1 and -M2B. [d] Number of unstable sequences with specified closing base pair from unstable-M1 and -M2. [e] Upper portion of table is for hairpins closed by a CG closing base pair. [f] For the unstable loop, AAC, different closing base pairs X and X′ were examined. CG was the only closing base pair that gave concentration-independent melts between ≈2 and 20 μM. Other closing base pairs resulted in ΔT_M values of ≈8 °C, indicating that they were no longer forming exclusively hairpins. [g] The lower portion of the table is for hairpins with a stable loop, UUA, and different closing base pairs X and X′.

structure mapping experiments on the genomic ribozyme from hepatitis delta virus reveals that ribonuclease U2 cleaves the A of a hairpin with the sequence 5′-GGGACCGUCCC-3′, consistent with little or no pairing between the A and U (ref *43*; D. Chadalavada and P. C. Bevilacqua, unpublished results). Based on nomenclature in use with tetraloops in which noncanonical pairing between positions 1 and 4 of the tetraloop is not a basis for defining these as diloops, we consider noncanonical or no pairing of nucleotides immediately 5′ and 3′ of the fully randomized nucleotides to comprise part of a pentaloop. The sequences found by selection will be discussed herein as triloops, although it remains possible that future studies may reveal that some hairpins, in particular selected unstable hairpins, do in fact contain pentaloops.

Selection was performed by two different methods as described above. Stable hairpins from method 2 consisted of two distinct bands that were separated by a further round of selection. By increasing the stringency of the selection gradually with each round through temperature increases and excising the smear somewhat closer to the main band, method 2 appears to offer the advantage that hairpins of moderate stability may also be isolated.

Bands labeled as "Stable" in Figure 2A were identified as such based on their greater mobility by parallel TGGE, their higher average T_M versus pool 0 (42.5 vs 37 °C) (Figure 2B), and their higher T_M and lower $\Delta G°_{37}$ values measured on individual minihairpins (Figure 3, Table 6). The percent intensity of a spot on the triloop library lane equal in size and position to the stable-M1 band (Figure 2A) is 6.2%, suggesting that approximately 24 (= 6.2% × 384) different sequences should be present in the stable population. As described in Table 2, sequencing of 81 different clones resulted in 29 different hairpins, eight of which occurred three or more times. Both the repeat occurrences of stable sequences and the similarity of the number of sequences found and expected based on the above calculation suggested that enough clones had been sequenced to sample the population adequately.

Sequence preferences for stable triloop hairpins include a U-rich loop and CG as the closing base pair (discussed below) (Tables 1−3). Several lines of evidence strongly suggest that this preference is not due to error or bias in the preparation of the triloop library. (i) Unstable hairpins, selected from the *same* triloop library, do not contain CG closing base pairs and are not U-rich (Tables 1, 4, and 5). (ii) Melting of minihairpins demonstrates directly that U-rich loop sequences and CG closing base pairs are exceptionally stable as compared to other triloop hairpins. In particular, cUUAg, the most common loop in the selection for stable hairpins (Table 2), is also the most stable loop (Table 6); AAC, the most common loop in the selection for unstable hairpins (Table 4), is also the least stable loop as judged by UV melting (Table 6). In addition, several results provide evidence against the possibility that hairpins were selected for nonthermodynamic reasons such as abnormal electrophoretic mobility. (i) Some purine-rich loop sequences occur uniquely in the stable band, such as GAG two times (Table 2). (ii) Some pyrimidine-rich loop sequences occur uniquely in the unstable band, such as GUU two times (Table 4). (iii) Stable hairpins run as a sharp band of distinctly greater mobility than pool 0 (Figure 2A). (iv) Melting of minihair-

pins demonstrates directly that selected sequences are indeed stable and unstable as expected.

Bands labeled as "Unstable" in Figure 2A were identified as such based on their slower mobility by parallel TGGE, their lower average T_M versus pool 0 (35.5 vs 37 °C) (Figure 2B), and their lower T_M and higher $\Delta G°_{37}$ values measured on individual minihairpins (Figure 3, Table 6). The percent intensity of a spot on the triloop library lane equal in size and position to the unstable-M1 band on a parallel TGGE with baths set at 27−37 °C is 23%, suggesting that approximately 88 (= 23% × 384) different sequences should be present in the unstable population. As described in Table 4, sequencing of 36 clones resulted in 25 different hairpins, only three of which occurred three or more times. The lack of recurrent sequences in the unstable population (with a few exceptions discussed below) is consistent with the large number of sequences in the unstable band, and further sequencing of this population was not pursued. Furthermore, the unstable-M1 population was only modestly less stable than pool 0 (35.5 vs 37.0 °C) (Figure 2B), consistent with the absence of a substantial number of exceptionally unstable hairpins. Nevertheless, a few of the unstable hairpins appear to be exceptionally unstable. In particular, the most common unstable loop, AAC, also had the least stable $\Delta G°_{37}$ of −1.1 kcal/mol (Table 6). Moreover, AAC and UUA, the most common stable loop, have similar frequencies of 13.9% and 13.6%, respectively. These results suggest that the selection methods used here are capable of identifying both exceptionally stable and unstable RNA hairpins.

Free Energy Rules for Triloop Stability. Sequence distributions from selection experiments suggest that a CG closing base pair contributes the most to overall hairpin stability since it occurs almost exclusively in the stable hairpins from methods 1 and 2B and is absent from the unstable hairpins (Table 1). The next most frequent closing base pairs are GC and UA, with UG and AU dominating the unstable sequences and absent from the stable sequences. Melting results with a stable loop sequence, UUA, were consistent with these sequence distributions, with a CG closing base pair having a $\Delta G°_{37}$ of −2.9 kcal/mol and a T_M of 61.3 °C; GC and UA having $\Delta G°_{37}$ values of −0.7 to −0.8 kcal/mol and T_M values of 45−46 °C; and AU, UG, and GU having $\Delta G°_{37}$ values of −0.4 to −0.5 kcal/mol and T_M values of 42−43 °C (Table 6).

Other studies indicate that a CG closing base pair can contribute substantially to overall stability. Phylogenetic comparisons pointed out that CG closing base pairs are preferred for the stable tetraloops UNCG (*4*). Studies on tetraloop hairpins revealed that a CG closing base pair is 2.3 kcal/mol more stable than GC in $\Delta G°_{37}$ at 10 mM NaCl for the exceptionally stable tetraloop UUCG (*5*). On the basis of the 1 M NaCl nearest-neighbor data, the term $^{5'AC3'}_{3'UG5'}$ is predicted to be only ≈0.2 kcal/mol more favorable in $\Delta G°_{37}$ than $^{5'AG3'}_{3'UC5'}$ (*44*), considerably smaller than the observed effect. The $\Delta\Delta G°_{37}$ for overall stability between CG and GC closing base pairs for the UUA triloop sequence found here is −2.1 kcal/mol in 10 mM NaCl, similar to the value of −2.3 kcal/mol found for tetraloops at this salt concentration. This observation suggests that CG closing base pairs can make unusually large contributions to the stability of certain triloop hairpins as well.

The bonus free energy for a CG closing base pair may not be general however. For example, studies on the hexaloop AUAAUA revealed that loops with CG and GC closing base pairs have the same $\Delta G°_{37}$ for loop formation (*45*). Studies on the triloop AUA reached a similar conclusion (*15*). The triloop AUA occurred one time in the selected stable loop sequences and was found to be moderately stable by UV melting (Tables 1 and 6). In addition, the $\Delta\Delta G°_{37}$ for loop formation for tetraloop hairpins comparing CG and GC closing base pairs is very dependent on the loop sequence and close to zero for tetraloops that are not unusually stable (*5*). Whether specific triloop hairpins have an extra contribution for the CG closing base pair to overall stability, as for example with UUA, may have to be determined directly for individual hairpins.

Studies on model RNA hairpins suggest that including C residues in the loop destabilizes the hairpin (*16*), and RNA structure prediction parameters assign a free energy penalty to hairpins with poly-C loops (*17*). At positions 1 and 3 in the stable loop sequences, C is the least common nucleotide occurring 7/81 and 9/81 times, respectively (Table 3), suggesting that C may be excluded from stable loops at these positions. However, C is not particularly common in the unstable hairpin loops, except for position 3 (Table 5). Moreover, hairpins with CCC or CUC loop sequences and a CG closing base pair are 0.5−0.6 kcal/mol more stable than a hairpin with an AAC loop sequence (Table 6). Thus, while C-rich loops do appear to be incompatible with stable hairpins and cause hairpins to be generally unstable, other specific loop sequences such as AAC are even more destabilizing than C alone. Occurrences of UUA and AAC as exceptionally stable and unstable loop sequences suggest that, while general stability trends for loop sequences are useful (Tables 3 and 5), the interactions within any given loop are intricate and occasionally an exceptionally stable or unstable loop sequence may arise in a fashion that is not fully predicted from frequencies of nucleotides in the loop.

Likewise, the stability of a hairpin is affected to differing degrees by single nucleotide changes in its loop sequence. For example, the most stable triloop, UUA, is only modestly destabilized by changing position 1 to an A or C, with $\Delta\Delta G°_{37}$ values of +0.4 kcal/mol (Table 6). Also, GUU is destabilized by only +0.2 kcal/mol upon changing position 3 to a C (Table 6). However, the C at position 3 of the least stable loop AAC, is critical for destabilizing the loop since changing it to an A results in a $\Delta\Delta G°_{37}$ of −1.2 kcal/mol. If loop bases are unable to stack effectively in the constrained environment of a triloop, as observed in the known triloop structures (*12*−*14*), then this destabilizing effect may reflect penalties for unstacking favorable base−base interactions from the unfolded state. Consistent with this notion, model studies revealed that single-strand stacking is favorable in poly-C and unfavorable in poly-U (*46*, *47*), correlating with their observed unfavorable and favorable $\Delta G°_{37}$ values for minihairpin formation, respectively (Table 6).

Several lines of the data suggest that interactions between a loop and its closing base pair are indirect and, therefore, generally additive. First, stable hairpins appear to be composed almost entirely of CG closing base pairs (Table 1). If a closing base pair interacted directly with its loop, then different closing base pairs might be expected to be stable and, furthermore, might interact in a specific fashion

with certain loop sequences. In contrast, the same loop sequences appear to be stable for a variety of different closing base pairs. For example, stable loop sequences for CG closing base pairs isolated from methods 1 and 2B are U-rich, as are stable loop sequences for GC and UA closing base pairs isolated from methods 1 and 2. Likewise, unstable loop sequences appear to be unstable independent of the closing base pair. For example, the unstable loop sequences isolated from selection (Table 4) did not have CG closing base pairs (Table 1) but in the background of a CG closing base pair caused a $\Delta\Delta G^\circ_{37}$ of up to +1.8 kcal/mol as compared to stable loop sequences (Table 6). On the basis of the clones obtained, stable hairpins (Table 2, methods 1 and 2B) appear to be require a stable loop sequence plus a stable closing base pair; moderately stable hairpins (Table 2, method 2A) require either a stable closing base pair (CG) plus a suboptimal loop sequence or a stable loop sequence plus a suboptimal closing base pair (GC or UA); and unstable hairpins (Table 4) require an unstable loop sequence and an unstable closing base pair. These observations suggest that interactions of the loop and closing base pair are indirect and largely additive; therefore, free energy bonuses and penalties may be applied in a straightforward manner for use in RNA structure prediction.

One exception to the additivity of the loop and closing base pair may be gUUUc. This appears to be the only exceptionally stable hairpin from the library with a GC closing base pair. An NMR structure of a longer hairpin containing the sequence cGUUUCg revealed a pentaloop with a CG closing base pair in which the O6 of the loop guanine hydrogen bonds to the exocyclic amine of the guanine in the CG closing base pair (*42*). It does not appear, however, that an identical interaction is possible with the AU base pair adjacent to the gUUUc in our construct, so it remains unclear at this time whether this is a triloop or a pentaloop. A number of true triloops with closing base pairs having Watson−Crick geometry are known (*12*−*14*), and perhaps most of the remaining hairpins contain true triloops because of their apparent additive behavior.

The ΔG°_{37} for formation of a hairpin containing the stable triloop cUUAg (−2.9 kcal/mol) is not as favorable as for the stable tetraloop cUUCGg (−4.9 kcal/mol), with an identical stem under identical conditions (D. J. Proctor, J. M. Bevilacqua, and P. C. Bevilacqua, unpublished data). The NMR structures of UUCG, GCAA, and GAAA tetraloops have been solved and reveal an intricate network of stacks and tertiary hydrogen bonds that serve to stabilize the tetraloop (*7, 8, 48*). Structures of stable UUU and UCU triloops show that loop nucleotides are splayed out with little or no base-base stacking (*12*−*14*). The constrained environment of a triloop may not permit as extensive a network of hydrogen bonding and stacking in the folded loop, thus leading to overall lesser stability.

One of the practical applications of the data obtained here is that it can provide thermodynamic parameters for use in RNA structure prediction algorithms. Currently, the stability of triloops is modeled as independent of the loop sequence (except CCC, which is penalized 1.4 kcal/mol), and CG base pairs are not given a bonus free energy (*17*). Data obtained here by UV melting indicate that this is an oversimplification. For example, CG closing base pairs can be up to 2.5 kcal/mol more stable than others, and differences in loop sequence can cause a $\Delta\Delta G^\circ_{37}$ of up to 1.8 kcal/mol (Table 6). Sequences such as AAC, GGC, and CUC should be penalized in addition to CCC. These results suggest that the rules for triloop stability should be modified.

The melting experiments done here were at 10 mM NaCl, but most of the thermodynamic rules currently in use were obtained in 1 M NaCl. We attempted melting triloop hairpins in 1 and 0.1 M NaCl, but most gave concentration-dependent melts, often with multiple transitions, indicating that they were no longer exclusively hairpins. Data on tetraloops melted at 10 mM and 1 M NaCl, however, suggest that ordering of hairpins according to ΔG°_{37} remains virtually unchanged and that ΔG°_{37} values become more favorable by only 0.4−1.2 kcal/mol upon increasing salt concentration (*5, 6*). Likewise ΔG°_{37} for formation of a hairpin containing the GCAA tetraloop is almost unchanged upon addition of 0.1 M NaCl (*49*). Thus, the above free energy rules for closing base pairs and triloop sequences may need to be incremented by about −0.8 kcal/mol to account for salt concentration but are unlikely to change in terms of loop sequence and closing base pair preferences.

Biological Importance of Triloops. We surveyed a database of 95 small subunit and 112 large subunit rRNAs (*31*−*33*) and found 703 loops designated as triloops (Tables 7−9). Closing base pairs are distributed as follows: 309 CG, 196 GC, 136 UA, 45 AU, 9 GU, and 8 UG. These frequencies are consistent with selection results discussed above with CG being most stable; GC and UA being moderately stable; and AU, GU, and UG being unstable (Tables 1 and 6). The most common loop sequence in the rRNAs is GGG, with the sequence GGGGC found 106 times in large subunit rRNA. This sequence is part of the peptidyl transferase center and interacts with tRNA in the P site (*50, 51*). A recent NMR structure reveals that the hairpin forms a pentaloop (*41*). Apparently, this loop was selected during evolution for a functional role. If the rRNA database is corrected by subtracting these loop sequences, then CG closing base pairs account for 52% of the remaining putative triloops.

Certain rRNA triloop sequences occur with high frequency. In particular, GGG, UAA, GCA, AUA, UUU, and GUA occur 111, 81, 58, 46, 44, and 32 times, respectively (Table 7). Of these loop sequences, only UUU is common in the stable-M1 and -M2B selected species (Table 2) with GGG, UAA, GCA, AUA, and GUA occurring only 0, 0, 1, 1, and 0 times, respectively. Melting of some of these samples indicates that hairpins with AUA and UUU are relatively stable, with ΔG°_{37} values of −2.5 kcal/mol, and that a hairpin with GCA is moderately stable with a ΔG°_{37} of −2.1 kcal/mol (Table 6). It thus appears likely that some triloop hairpins may have been selected in nature, at least in part, for their thermal stability. The relatively low representation of the loop sequence AUA in in vitro selection is unclear and could result from undersampling or possibly from an unidentified bias in the selection procedure. The loop sequence UUU may have been selected in part for thermodynamic stability since 24/44 UUU loop sequences in rRNAs are closed with a CG base pair, and only 2/44 are closed with the GC base pair, which is known to form a pentaloop in certain contexts (*42*). At position 83 in *E. coli* 16S rRNA, a hairpin loop occurs with the tetraloop sequences CUUG 45% and UUCG 36% but is occasionally substituted with a UUU triloop closed by CG or UA base pairs (*4*). On the

Table 7: Loop Sequence Distribution in rRNA Triloop Hairpins[a]

loop sequence[b]	X:X′ [c]	C:G[d]	loop sequence	X:X′	C:G	loop sequence	X:X′	C:G	loop sequence	X:X′	C:G
AAA	24	10	CAA	9	1	GAA	30	5	UAA	81	39
AAC	7	5	CAC	0	0	GAC	2	1	UAC	1	0
AAG	3	2	CAG	1	1	GAG	5	1	UAG	2	0
AAU	1	0	CAU	3	0	GAU	2	2	UAU	17	5
ACA	18	4	CCA	1	0	GCA	58	45	UCA	4	4
ACC	2	0	CCC	3	1	GCC	1	1	UCC	7	5
ACG	6	2	CCG	1	1	GCG	3	2	UCG	4	4
ACU	7	3	CCU	5	1	GCU	4	2	UCU	21	15
AGA	17	14	CGA	2	1	GGA	4	3	UGA	3	1
AGC	4	1	CGC	2	1	GGC	1	1	UGC	3	1
AGG	0	0	CGG	3	0	GGG	111	2	UGG	2	1
AGU	7	1	CGU	1	0	GGU	2	0	UGU	6	1
AUA	46	36	CUA	12	10	GUA	32	13	UUA	13	4
AUC	5	1	CUC	4	3	GUC	0	0	UUC	13	11
AUG	2	2	CUG	0	0	GUG	6	2	UUG	4	2
AUU	11	5	CUU	4	2	GUU	6	4	UUU	44	24

[a] The total number of triloop hairpins is 703 with 95 small subunit and 112 large subunit sequences surveyed (*31*−*33*). Closing base pair distributions are 309/703 CG, 196/703 GC, 136/703 UA, 45/703 AU, 9/703 GU, and 8/703 UG. [b] Listed alphabetically for all 64 possible triloops. [c] X:X′ denotes total number of hairpins with all possible complementary bases for the given loop sequence. [d] C:G denotes total number of hairpins with a CG closing base pair for the given loop sequence.

Table 8: Nucleotide Distribution in RRNA Triloop Hairpins[a]

position[b]	A	C	G	U
X	45	309	205	144
1	160	51	269	223
2	187	147	168	201
3	355	55	153	140
X′	136	196	317	54

[a] The total number of sequences is 703. [b] X and X′ denote complementary bases.

Table 9: Nucleotide Distribution in rRNA Triloop Hairpins with a CG Closing Base Pair[a]

position	A	C	G	U
1	86	22	84	117
2	72	90	28	119
3	190	32	22	65

[a] The total number of sequences is 309.

basis of known stability of CUUG and UUCG tetraloops (*5*) and the occurrences of cUUUg and uUUUa triloops in the selection (Table 2), this result is consistent with a role for UUU as a stable triloop sequence. A structure of a UUU triloop reveals that the U residues are splayed out (*13*), suggesting that loop nucleotides could also serve as a recognition site for protein binding or tertiary structure formation. It is further possible that certain RNA loop sequences can have multiple roles, such as stability elements in folding intermediates, and recognition elements in the final structure.

The distribution of nucleotides in the rRNA triloops is reported in Table 8. Unlike the results from the in vitro selection, no obvious trends for base preferences in the loop could be discerned. However, this table contains loops that may have been selected for nonthermodynamic reasons as discussed above and has triloops with unstable closing base pairs. For example, triloops can form protein binding sites that interact strongly with the protein (*52*). In Table 9, the distribution of loop nucleotides for hairpins with a CG closing base pair are reported. This comparison is based on the observation that stable hairpins require a stable loop plus a stable closing base pair, and it may therefore be easier to see trends for stable loops by examining CG closing base pair data. In agreement with selection results (Table 3), U is preferred at positions 1 and 2, occurring 117/309 and 119/309, respectively (Table 9). Interestingly, an A is preferred at position 3, occurring 190/309 times. This preference for A was not observed in the nucleotide distributions for stable sequences (Table 3), although A is underrepresented at position 3 in the unstable loop sequences (Table 5), and the most stable loop sequence found was UUA (Table 6).

ACKNOWLEDGMENT

We thank David Proctor for assistance in identifying triloop hairpins in the PDB and for his comments on the manuscript. We also thank Paul Babitzke, Joanne Bevilacqua, Durga Chadalavada, Marty Serra, and Doug Turner for critical readings of the manuscript.

REFERENCES

1. Giese, M. R., Betschart, K., Dale, T., Riley, C. K., Rowan, C., Sprouse, K. J., and Serra, M. J. (1998) *Biochemistry 37*, 1094−1100.
2. Varani, G. (1995) *Annu. Rev. Biophys. Biophys. Chem. 24*, 379−404.
3. Wolters, J. (1992) *Nucleic Acids Res. 20*, 1843−1850.
4. Woese, C. R., Winker, S., and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467−8471.
5. Antao, V. P., Lai, S. Y., and Tinoco, I., Jr. (1991) *Nucleic Acids Res. 19*, 5901−5905.
6. Antao, V. P., and Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 819−824.
7. Cheong, C., Varani, G., and Tinoco, I., Jr. (1990) *Nature 346*, 680−682.
8. Heus, H. A., and Pardi, A. (1991) *Science 253*, 191−194.
9. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K., and Miura, K. (1994) *Nucleic Acids Res. 22*, 576−582.
10. Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K., and Hirao, I. (1997) *Biochemistry 36*, 4761−4767.
11. Konings, D. A., and Gutell, R. R. (1995) *RNA 1*, 559−574.
12. Puglisi, J. D., Wyatt, J. R., and Tinoco, I., Jr. (1990) *Biochemistry 29*, 4215−4226.
13. Davis, P. W., Thurmes, W., and Tinoco, I., Jr. (1993) *Nucleic Acids Res. 21*, 537−545.

14. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
15. Serra, M. J., Barnes, T. W., Betschart, K., Gutierrez, M. J., Sprouse, K. J., Riley, C. K., Stewart, L., and Temel, R. E. (1997) *Biochemistry 36*, 4844−4851.
16. Groebe, D. R., and Uhlenbeck, O. C. (1988) *Nucleic Acids Res. 16*, 11725−11735.
17. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
18. Joyce, G. F. (1994) *Curr. Opin. Struct. Biol. 4*, 331−336.
19. Osborne, S. E., and Ellington, A. D. (1997) *Chem. Rev. 97*, 349−370.
20. Breaker, R. R. (1997) *Chem. Rev. 97*, 371−390.
21. Juneau, K., and Cech, T. R. (1999) *RNA 5*, 1119−1129.
22. Wartell, R. M., Hosseini, S. H., and Moran, C. P., Jr. (1990) *Nucleic Acids Res. 18*, 2699−2705.
23. Ke, S. H., and Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137−5143.
24. Zhu, J., and Wartell, R. M. (1997) *Biochemistry 36*, 15326−15335.
25. Bevilacqua, J. M., and Bevilacqua, P. C. (1998) *Biochemistry 37*, 15877−15884.
26. Milligan, J. F., and Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51−62.
27. Bevilacqua, P. C., George, C. X., Samuel, C. E., and Cech, T. R. (1998) *Biochemistry 37*, 6303−6316.
28. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
29. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 597, CRC Press, Cleveland, OH.
30. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 597, CRC Press, Cleveland, OH.
31. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
32. Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3502−3507.
33. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) *J. Mol. Biol. 256*, 701−719.
34. 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., and Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364−1368.
35. Draper, D. E. (1995) *Annu. Rev. Biochem. 64*, 593−620.
36. Costa, M., and Michel, F. (1995) *EMBO J. 14*, 1276−1285.
37. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S. A. (1998) *Science 279*, 1940−1943.
38. Treiber, D. K., Rook, M. S., Zarrinkar, P. P., and Williamson, J. R. (1998) *Science 279*, 1943−1946.
39. Zuker, M., and Jacobson, A. B. (1998) *RNA 4*, 669−679.
40. Babitzke, P. (1997) *Mol. Microbiol. 26*, 1−9.
41. Puglisi, E. V., Green, R., Noller, H. F., and Puglisi, J. D. (1997) *Nat. Struct. Biol. 4*, 775−778.
42. Sich, C., Ohlenschlager, O., Ramachandran, R., Gorlach, M., and Brown, L. R. (1997) *Biochemistry 36*, 13989−14002.
43. Rosenstein, S. P., and Been, M. D. (1991) *Nucleic Acids Res 19*, 5409−5416.
44. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
45. Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A., and Turner, D. H. (1993) *Nucleic Acids Res. 21*, 3845−3849.
46. Richards, E. G., Flessel, C. P., and Fresco, J. R. (1963) *Biopolymers 1*, 431−446.
47. Freier, S. M., Hill, K. O., Dewey, T. G., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) *Biochemistry 20*, 1419−1426.
48. Allain, F. H., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.
49. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1992) *Science 256*, 217−219.
50. Moazed, D., and Noller, H. F. (1989) *Cell 57*, 585−597.
51. Green, R., Samaha, R. R., and Noller, H. F. (1997) *J. Mol. Biol. 266*, 40−50.
52. Convery, M. A., Rowsell, S., Stonehouse, N. J., Ellington, A. D., Hirao, I., Murray, J. B., Peabody, D. S., Phillips, S. E., and Stockley, P. G. (1998) *Nat. Struct. Biol. 5*, 133−139.
53. Serra, M. J., and Turner, D. H. (1995) *Methods Enzymol. 259*, 242−261.

BI991774Z