Isolation and Characterization of a Family of Stable RNA Tetraloops with the Motif YNMG That Participate in Tertiary Interactions†

David J. Proctor, Janell E. Schaak, Joanne M. Bevilacqua, Christopher J. Falzone, and Philip C. Bevilacqua*

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

Received May 28, 2002; Revised Manuscript Received July 26, 2002

ABSTRACT: RNA is known to fold into a variety of structural elements, many of which have sufficient sequence complexity to make the thermodynamic study of each possible variant impractical. We previously reported a method for isolating stable and unstable RNA sequences from combinatorial libraries using temperature gradient gel electrophoresis (TGGE). This method was used herein to analyze a six-nucleotide RNA hairpin loop library. Three rounds of in vitro selection were performed using TGGE, and unusually stable RNAs were identified by cloning and sequencing. Known stable tetraloops were found, including sequences belonging to the UNCG motif closed by a CG base pair, and the CUUG motif closed by a GC base pair. In addition, unknown tetraloops were found that were nearly as stable as cUNCGg, including sequences related through substitution of the U with a C (Y), the C with an A (M), or both. These substitutions allow hydrogen bonding and stacking interactions in the UNCG loop to be maintained. Thermodynamic analysis of YNMG and variant loops confirmed optimal stability with Y at position 1 and M at position 3. Similarity in structure and stability among YNMG loops was further supported by deoxyribose substitution, CD, and NMR experiments. A conserved tertiary interaction in 16S rRNA exists between a YAMG loop at position 343 and two adenines in the loop at position 159 (*Escherichia coli* numbering). NMR and functional group substitution experiments suggest that YNAG loops in particular have enhanced flexibility, which allows the tertiary interaction to be maintained with diverse loop sequences at position 159. Taken together, these results support the existence of an extended family of UNCG-like tetraloops with the motif cYNMGg that are thermodynamically stable and structurally similar and can engage in tertiary interactions in large RNA molecules.

RNA is essential to diverse cellular processes, including transcription, translation, and splicing. RNA alone can be catalytic and appears to be the catalytic component of some ribonucleoprotein complexes, including the ribosome and spliceosome (*1*, *2*). Large RNAs involved in these processes often fold in a hierarchical fashion using smaller independently stable building blocks to form complex tertiary structures (*3*, *4*). For example, loop−receptor tertiary interactions are built up from a small stable hairpin and an internal loop (*5*, *6*), and the P1 duplex of the *Tetrahymena* ribozyme docks into the catalytic core as a preformed duplex (*7*, *8*). As such, identification and characterization of small structural elements that comprise the foundation of complex RNAs may lead to a deeper understanding of RNA folding.

Hairpins are the dominant secondary structure in RNA and are known to function in nucleating RNA folding, RNA−RNA tertiary interactions, and RNA−protein interactions (*4*, *9*). The tetraloop is the most common hairpin loop size in RNA (*10*−*13*). In particular, in a database of 214 different 16S and 23S rRNAs from 133 organisms, 7484 hairpin loops of three to six nucleotides were found, and 51% of these were tetraloops (D. J. Proctor, Z. Shu, E. Cho, and P. C. Bevilacqua, unpublished data). Also, a study of 624 hairpins in variable regions of 16S-like rRNA that are less likely to interact with proteins or engage in tertiary interactions revealed that approximately 50% of the loops were tetraloops, being somewhat greater for archaea and bacteria and less for eukarya (*12*). Phylogenetic studies of rRNA have revealed that certain tetraloop motifs are especially common, including cUNCGg,[1] cGNRAg, and to a lesser extent gCUUGc (*11*).

The stability and structure of RNA tetraloops have been extensively studied. Thermodynamic studies indicated that the majority of the phylogenetically conserved tetraloop-closing base pair combinations are exceptionally stable (*10*, *14*, *15*). Apparently, thermodynamically stable motifs can serve important biological roles. Atomic-resolution structures have been determined for many tetraloops, including cGAAAg, cGAGAg, cGCAAg (*16*−*18*), gCUUGc (*19*), cUUCGg, and cUACGg (*1*, *20*−*24*), and have revealed extensive stacking, hydrogen bonding, and shape complementarity within the loop.

The UNCG motif has been extensively studied by NMR spectroscopy and X-ray crystallography (*1*, *20*−*24*), and several distinct features have been identified. For cUNCGg, hydrogen bonding interactions occur between U_{L1} O2 and G_{L4} H1 and H2, between U_{L1} H2′ and G_{L4} O6, and between C_{L3} H4 and the *pro-R* phosphoryl oxygen of N_{L2} (Figure 1). Distinctive non-hydrogen bonding features include nonstandard torsion angles for N_{L2} and the closing base pair G, a syn

† Supported by National Science Foundation Career Grant MCB-9984129 and a Camille Dreyfus Teacher-Scholar Award and a Sloan Fellowship to P.C.B.

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

10.1021/bi026201s CCC: $22.00
Published on Web 09/14/2002

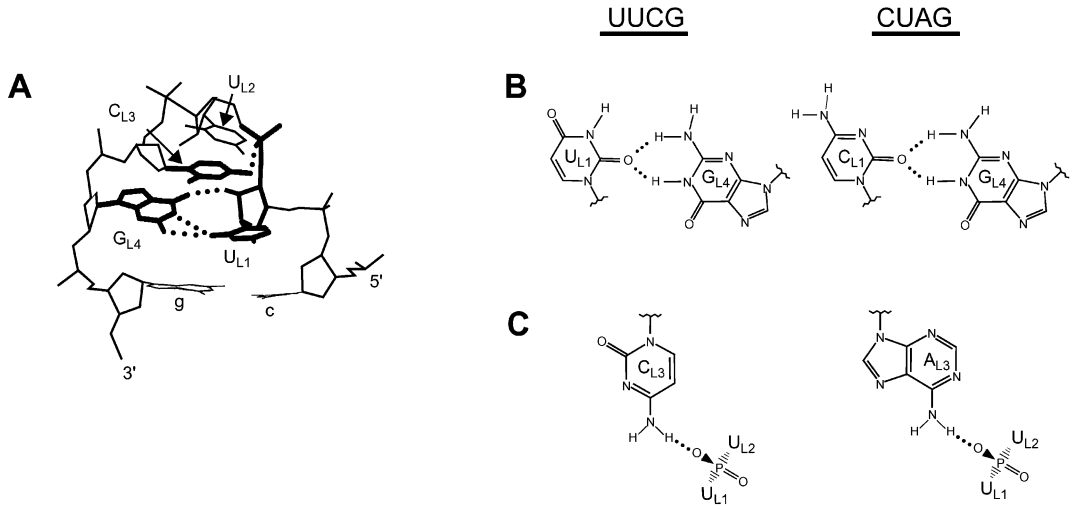

FIGURE 1: Comparison of UUCG and its double variant CUAG reveals potential for similar hydrogen bonding interactions. (A) Crystal structure of the cUUCGg tetraloop [PDB entry 1F7Y (*24*)]. The numbering is for loop nucleotides, with L*n* denoting loop position *n*. Bases, sugars, and phosphates involved in hydrogen bonding are bold. Important hydrogen bonding interactions are shown with dotted lines. (B and C) Isomorphous hydrogen bonding interactions that are possible for UUCG and CUAG. These features are also visible in the crystal structure. (B) U_{L1} and G_{L4} of UUCG form bifurcated hydrogen bonds using the two lone pairs of U_{L1} O2 to G_{L4} H1 and H2 (*20*−*22*); C_{L1} has the potential for an identical interaction with G_{L4}. (C) C_{L3} and the *pro-R* O of U_{L2} form a hydrogen bond using the exocyclic amine of C_{L3}; A_{L3} has the potential for an identical interaction with U_{L2} (*20*−*22*).

conformation for G_{L4}, and C2′-endo sugar puckers for N_{L2} and C_{L3}.

The UUCG tetraloop with a CG closing base pair was originally identified as exceptionally stable from sequence alignment and its ability to terminate reverse transcription reactions (*10*). The CG closing base pair makes a significant contribution to stability. In particular, a CG to GC change destabilizes tetraloop formation, with a $\Delta\Delta G_{37}°$ of 2.3 kcal/mol (*14*). More recently, NMR and computational studies revealed similar structures for UUCG tetraloops closed by CG and GC base pairs at low temperatures, but more dynamic behavior for gUUCGc (*15*). In addition, 2′-hydroxyl groups make important contributions to stability, as revealed by both experimental and computational studies (*15*, *25*−*27*).

RNA molecules with enhanced secondary and tertiary structure stability have been identified by in vitro selection (SELEX) methods (*28*, *29*). We have used the combination of temperature gradient gel electrophoresis (TGGE) and in vitro selection to identify unusually stable and unstable RNA secondary structure motifs (*28*, *30*). The feasibility of this approach was demonstrated by TGGE separation of a simple combinatorial hairpin library according to the thermodynamic stability of its members (*30*). The method was then applied to a triloop-closing base pair combinatorial library comprised of 384 different sequences (*28*). Unusually stable and unstable triloop-closing base pair combinations were isolated and characterized, and several have since been found to occur naturally (*31*−*34*).

In the study presented here, we prepared a tetraloop-closing base pair library by randomization of six nucleotides (4096 different sequences) at the top of a long stem. Stable sequences were isolated by in vitro selection using TGGE, and identified by cloning and sequencing. Families of stable tetraloops related to the UNCG motif were identified by sequence alignment, and characterized by functional group substitution, UV melting, and CD and NMR spectroscopies.

MATERIALS AND METHODS

Preparation of Nucleic Acids. DNA templates were prepared and gel purified as previously described, and RNA was prepared as described previously (*28*, *30*). The various RNA pools were purified and radiolabeled as described previously (*28*, *30*). RNA oligonucleotides were stored in TE [10 mM Tris and 1 mM Na_2EDTA (pH 7.5)] for selection experiments or $P_{10}E_{0.1}$ [10 mM sodium phosphate and 0.1 mM Na_2EDTA (pH 7.0)] for UV melting, CD, and NMR experiments. Concentrations of NaCl of ≥0.1 M resulted in concentration-dependent UV melts for some of the sequences (*14*, *21*, *35*). For thermodynamic studies involving 2′-OH substitutions and NMR experiments, synthetic RNA molecules and RNA−DNA chimeras were chemically synthesized (Dharmacon Research, Inc.). Oligonucleotides were deprotected as per the manufacturer's instructions and dialyzed as described below.

[1] When the loop and the closing base pair are mentioned, the loop is uppercase and the closing pair is lowercase. For example, cUUCGg is a UUCG tetraloop closed by a CG base pair. If not shown explicitly, tetraloops are closed by a CG base pair. RNA mixtures after *n* rounds of selection are termed pool *n*, and the starting library is pool 0. A tetraloop is defined as a loop of four nucleotides closed by either a Watson−Crick or wobble pair. Loop 159 and loop 343 refer to hairpin loops in any 16S rRNA corresponding to loops beginning with residue 159 and 343 in *E. coli*. Loop 159 closes helix 8 (H8) and loop 343 helix 14 (H14), where H8 and H14 refer to the nomenclature used in the crystal structure of the 30S subunit from *Tetrahymena thermophilus*. Abbreviations: CD, circular dichroism; C_T, total strand concentration; d, duplex; DQF-COSY, double-quantum-filtered correlated spectroscopy; EDTA, ethylenediaminetetraacetic acid; D = A, G, or U; H = A, C, or U; M = A or C; R = A or G; Y = C or U; FID, free induction decay; h, hairpin; L_n, loop position *n*; minihairpin, shorter RNA used for biophysical study with the general sequence 5′-ggac$L_1L_2L_3L_4$gucc; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; $P_{10}E_{0.1}$, 10 mM sodium phosphate and 0.1 mM Na_2EDTA; RT, reverse transcription; $TBEK_{50}$, 100 mM Tris, 83 mM boric acid, 1 mM Na_2EDTA, and 50 mM KCl (pH 8.7 and 25 °C); TE, 10 mM Tris and 1 mM Na_2EDTA (pH 7.5); TGGE, temperature gradient gel electrophoresis; T_M, melting temperature; TPPI, time-proportional phase incrementation; Tris, tris(hydroxymethyl)aminomethane; WATERGATE, water suppression by gradient-tailored excitation.

Table 1: Sequence Alignment and Occurrence at Loop 343 in 16S rRNA

motif[a]	sequence[b]	occurrence in selection[c]	occurrence in rRNA[d]
cUNCGg	cUACGg	12	72%
	cUUCGg	5	
	cUCCGg	3	
cUNAGg	cUAAGg	2	4%
	cUCAGg	2	
	cUUAGg	2	
cCNCGg	cCACGg	6	3%
cCNAGg	cCUAGg	5	
	cCAAGg	4	16%
	cCGAGg	1	
others	cCUUGg[e]	3	
	gCUUGc[e]	1	
	cAACUg	1	
	cGUAUg	1	

[a] The motif was assigned on the basis of selection results and outcomes of melting, CD, and NMR experiments. The first four motifs are cYNMGg loops. [b] The sequences of positions 19−24 of the library (Figure 2A) are given with the presumed closing base pair in lowercase. [c] The occurrence is based upon sequencing 48 clones without mutations in the stem. Listed in order of occurrences within each subfamily. [d] The occurrence in 16S and 16S-like rRNA sequences at loop 343. Percentages are based upon an analysis of 2290 sequences from archaea, bacteria, and eukarya (*13*). [e] Despite similar sequences and thermodynamic parameters, these RNAs have different CD spectra and phylogenetic conservation rates.

Temperature Gradient Gel Electrophoresis. TGGE was carried out on perpendicular and parallel apparatuses using 0.5 (analytical gels) or 1.5 mm thick spacers (preparative gels) as described previously (*28*, *30*). Standard gel conditions were 8% polyacrylamide (29:1 acrylamide:bisacrylamide ratio) and 4 M urea, and the running buffer was $TBEK_{50}$ [100 mM Tris, 83 mM boric acid, 1 mM Na_2EDTA, and 50 mM KCl (pH 8.7) at 25 °C]. Before loading was carried out, RNA was renatured and gels were pre-electrophoresed to allow thermal equilibration, as described previously (*28*, *30*, *36*). The temperature gradient was found to be linear and constant during a 3 h electrophoresis run.

Selection Procedure. Parallel TGGE was carried out during each round of selection, with temperatures provided in the Results. Bands were visualized by autoradiography at 4 °C without drying the gel, and the desired RNA species was excised from preparative gels with a sterile razor blade. The RNA was recovered by a crush and soak procedure, and RT-PCR reactions were performed using 10 rounds of thermal cycling for selection round 1,and 20 rounds for subsequent rounds. The primers used in RT-PCR were the same as previously described (*28*). After three rounds of selection, the PCR fragment was digested with *Eco*RI and *Bam*HI, cloned into pUC19, and transformed into DH5α *Escherichia coli*. Isolated clones were sequenced by the dideoxy method; the fixed-sequence stem was also examined, and only sequences lacking mutations in this region (which was the majority of sequences) were used in the final alignment.

UV Melting. RNA oligonucleotides used in UV melting experiments were prepared by transcription; in selected cases, chemically synthesized oligonucleotides (Dharmacon Research, Inc.) were also melted and gave similar results (see Table 2). To favor the hairpin conformation, the RNA was renatured prior to each melt by heating to 90 °C for 3 min, followed by slow cooling at ambient temperature for 10 min. Failure to renature resulted in distorted melting profiles with multiple transitions, presumably due to misfolding or dimerization of the RNA. UV absorbance melting profiles were obtained at 260 nm in 1, 5, and 10 mm path length cuvettes with a heating rate of ≈1 °C/min using a Gilford Response II spectrophotometer equipped with a temperature controller. In general, melts were performed by increasing the temperature from 10 to 99 °C. In certain cases, melts were also performed by decreasing the temperature, and overlapping curves were obtained, consistent with reversibility of the transition. Each sample was melted over an ≈250-fold range of strand concentration (C_T), extending from 1 to 250 μM. The similarity of melting profiles for C_T values from 1 to 100 μM for all sequences and for C_T values from 1 to 250 μM for most sequences (see the Results) was consistent with the unfolding transition being due to the hairpin species. If oligonucleotides were forming duplexes, calculations using nearest-neighbor thermodynamic values (*37*−*39*) predict that the T_M should increase by 9−12 °C between C_T values of 1 and 100 μM and by 11−14 °C between C_T values of 1 and 250 μM, well outside the error limits in T_M (±1 °C). Thermodynamic parameters were obtained by fitting to a two-state model with sloping baselines using a nonlinear least-squares program (*40*). Concentrations were determined using absorbance values of the unfolded state at 90 °C, and extinction coefficients were from a nearest-neighbor analysis (*41*, *42*).

Circular Dichroism. Circular dichroism data were obtained using an AVIV 62DS spectrometer at 25 °C in $P_{10}E_{0.1}$ (pH 7.0) and C_T values of 3−30 μM, conditions under which the major species is a folded hairpin for all sequences. Data are the average of three scans from 220 to 320 nm in 1 nm increments using an integration time of 2 s. A buffer blank was subtracted from all spectra, and small separations of spectra at λ of ≥320 nm were corrected using an offset.

CD spectra were converted from ellipticity in degrees (θ) to molar residue ellipticity ($\Delta\epsilon$ in units of M^{-1} cm^{-1}) according to eq 1:

$$\Delta\epsilon = \theta/(33Cnb) \quad (1)$$

where θ is the measured ellipticity in degrees, C is the molar RNA concentration, n is the number of nucleotides, and b is the cuvette path length of 0.5 cm (*43*, *44*).

NMR Sample Preparation and Spectroscopy. Synthetic RNA oligonucleotides (Dharmacon Research, Inc.) for representative YNMG sequences were deprotected, dried in vacuo, and resuspended in 100 mM NaCl and 2 mM Na_2-EDTA (pH 8). For ^{31}P and exchangeable ^1H experiments, oligonucleotides were dialyzed for 8−12 h against each of the following: 100 mM NaCl and 2 mM Na_2EDTA (pH 8), water, and 10 mM sodium phosphate and 0.1 mM Na_2EDTA (pH 5.1) (*45*). Final NMR samples were 95% H_2O and 5% ^2H_2O (v/v) with 10 mM sodium phosphate and 0.1 mM Na_2-EDTA (pH 5.1). For nonexchangeable ^1H experiments, oligonucleotides were dialyzed through the water step described above, and then lyophilized several times in 99.9% ^2H_2O, with a final resuspension in 99.96% ^2H_2O (Isotec) with 10 mM sodium phosphate and 0.1 mM Na_2EDTA (pD 5.5), calculated by adding 0.4 to the pH meter reading (*46*). A pH of 5.1 was used to slow chemical exchange processes (*47*); melts at pH 4.9 did not show significant changes in

Table 2: Thermodynamic Parameters for Minihairpin Formation by All-Ribose Sequences[a]

sequence[b]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G_{37}°$ (kcal/mol)	$\Delta\Delta G_{37}°$ (kcal/mol)	$\Delta G_{55}°$ (kcal/mol)	$\Delta\Delta G_{55}°$ (kcal/mol)	T_M (°C)	ΔT_M (°C)
CNAG								
CAAG	−44.4 ± 0.8	−130.2 ± 2.2	−4.00 ± 0.07	0.82 ± 0.26	−1.69 ± 0.08	0.72 ± 0.17	67.7	−5.2
CCAG	−43.0 ± 1.8	−126.3 ± 5.6	−3.83 ± 0.10	0.99 ± 0.27	−1.60 ± 0.11	0.80 ± 0.18	67.4	−5.6
CGAG	−40.7 ± 1.3	−119.4 ± 3.8	−3.72 ± 0.13	1.10 ± 0.28	−1.57 ± 0.10	0.84 ± 0.18	68.1	−4.8
CUAG	−37.6 ± 1.3	−109.5 ± 4.0	−3.59 ± 0.09	1.23 ± 0.27	−1.62 ± 0.05	0.79 ± 0.15	69.8	−3.1
CNCG								
CACG	−37.4 ± 0.6	−109.1 ± 2.0	−3.56 ± 0.05	1.26 ± 0.26	−1.60 ± 0.03	0.81 ± 0.15	69.6	−3.3
CCCG	−37.3 ± 2.3	−109.5 ± 6.5	−3.38 ± 0.30	1.44 ± 0.39	−1.51 ± 0.27	0.90 ± 0.31	67.8	−5.1
CGCG	−41.6 ± 0.6	−122.3 ± 2.1	−3.67 ± 0.01	1.15 ± 0.25	−1.42 ± 0.27	0.99 ± 0.31	67.0	−5.9
CUCG	−48.0 ± 1.2	−139.4 ± 3.4	−4.78 ± 0.16	0.04 ± 0.30	−2.27 ± 0.11	0.13 ± 0.18	71.3	−1.6
UNAG								
UAAG	−41.7 ± 0.5	−122.8 ± 1.4	−3.62 ± 0.11	1.20 ± 0.28	−1.40 ± 0.10	1.00 ± 0.17	66.4	−6.5
UCAG	−40.7 ± 0.3	−119.6 ± 0.7	−3.56 ± 0.06	1.26 ± 0.26	−1.41 ± 0.05	0.99 ± 0.15	66.8	−6.1
UGAG	−37.7 ± 1.0	−110.9 ± 3.0	−3.33 ± 0.10	1.50 ± 0.27	−1.28 ± 0.11	1.12 ± 0.18	67.0	−5.9
UUAG	−40.3 ± 1.6	−117.8 ± 4.7	−3.71 ± 0.13	1.11 ± 0.29	−1.61 ± 0.08	0.80 ± 0.17	68.5	−4.4
UNCG								
UACG	−49.7 ± 0.3	−144.5 ± 0.7	−4.90 ± 0.06	−0.08 ± 0.26	−2.26 ± 0.10	0.14 ± 0.17	70.9	−2.0
UCCG	−45.3 ± 1.3	−131.7 ± 3.9	−4.46 ± 0.11	0.36 ± 0.28	−2.10 ± 0.05	0.30 ± 0.15	70.8	−2.1
UGCG	−44.8 ± 0.6	−130.2 ± 1.7	−4.44 ± 0.07	0.38 ± 0.26	−2.07 ± 0.08	0.34 ± 0.17	71.1	−1.8
UUCG	−46.4 ± 2.5	−134.2 ± 7.3	−4.82 ± 0.25	−	−2.40 ± 0.14	−	72.9	−
position 1 controls (non-U/C)								
AAAG	−30.3 ± 2.1	−90.0 ± 6.3	−2.42 ± 0.24	2.40 ± 0.35	−0.81 ± 0.19	1.60 ± 0.24	65.0	−7.9
GAAG	−38.6 ± 0.9	−113.8 ± 2.9	−3.33 ± 0.07	1.49 ± 0.26	−1.27 ± 0.05	1.13 ± 0.15	66.3	−6.6
AACG	−27.2 ± 0.8	−82.0 ± 2.8	−1.75 ± 0.12	3.07 ± 0.28	−0.22 ± 0.15	2.18 ± 0.21	58.4	−14.6
GACG	−38.4 ± 1.2	−114.5 ± 3.6	−2.88 ± 0.08	1.94 ± 0.27	−0.81 ± 0.03	1.60 ± 0.15	62.1	−10.8
position 3 controls (non-A/C)								
CAGG	−38.4 ± 2.5	−114.0 ± 7.2	−3.02 ± 0.27	1.80 ± 0.37	−0.97 ± 0.15	1.44 ± 0.21	63.5	−9.5
CAUG	−29.9 ± 3.2	−89.1 ± 9.3	−2.24 ± 0.35	2.58 ± 0.43	−0.64 ± 0.18	1.77 ± 0.23	62.0	−10.9
UAGG	−42.0 ± 1.2	−125.9 ± 3.9	−2.92 ± 0.02	1.90 ± 0.26	−0.63 ± 0.06	1.77 ± 0.16	60.2	−12.7
UAUG	−39.5 ± 1.2	−117.0 ± 3.7	−3.17 ± 0.15	1.65 ± 0.29	−1.06 ± 0.12	1.34 ± 0.19	64.1	−8.9
closing base pair controls (cg to gc)[a,c]								
gCAAGc	−35.8 ± 1.4	−107.3 ± 3.9	−2.52 ± 0.22	1.48 ± 0.35	−0.59 ± 0.17	1.10 ± 0.25	60.5	−7.2
gCACGc	−37.4 ± 2.2	−111.8 ± 6.4	−2.69 ± 0.20	0.87 ± 0.20	−0.68 ± 0.12	0.92 ± 0.13	61.1	−8.5
gUAAGc	−34.3 ± 3.9	−105.1 ± 12.5	−1.76 ± 0.13	1.86 ± 0.17	0.13 ± 0.17	1.53 ± 0.20	53.9	−12.5
gUUCGc[d]				2.3				−11.6

[a] Minihairpins have the sequence 5′-ggac$L_1L_2L_3L_4$gucc, except for closing base pair controls which have the sequence 5′-ggag$L_1L_2L_3L_4$cucc. Parameters are from UV melts in $P_{10}E_{0.1}$ and fitting to a two-state model with sloping baselines (*40*). Experiments were performed over a C_T range of 1−250 μM. T_M values were independent of concentration between 1 and 100 μM for all oligomers and between 1 and 250 μM for most. Parameters are the average of at least five independently prepared samples at varying concentrations over the concentration-independent range. Values for $\Delta G°$ are also provided at 55 °C since this is closer to the T_M and may be somewhat more accurate than $\Delta G_{37}°$ values; similar trends in $\Delta G°$ and T_M were observed at 37 and 55 °C. Errors for $\Delta H°$, $\Delta S°$, $\Delta G_{37}°$, and $\Delta G_{55}°$ are standard deviations from repeated experiments. Errors in $\Delta\Delta G_{37}°$ and $\Delta\Delta G_{55}°$ were propagated from $\Delta G°$ values by standard methods (*70*). Three figures are provided for $\Delta G°$ values to avoid round-off errors in calculations. Maximal errors in T_M are estimated to be ±1 °C. Values for $\Delta\Delta G_{37}°$, $\Delta\Delta G_{55}°$, and ΔT_M are relative to UUCG. [b] Sequences are for L_1−L_4 and are listed alphabetically by YNMG subfamily or as a control. Sequences prepared by chemical synthesis, which lacked the 5′-triphosphate, possessed thermodynamic parameters which were slightly less stable, as expected on the basis of other studies (*71*). For example, synthesized UUCG resulted in the following values: $\Delta H° = -45.8$ kcal/mol, $\Delta S° = -132.3$ eu, $\Delta G_{37}° = -4.72$ kcal/mol, and $T_M = 72.7$ °C. [c] Differences in free energy and melting temperature are relative to the same loop sequence with a CG closing base pair, and errors were propagated by standard methods (*70*). [d] From ref *14*, under similar solution conditions.

$\Delta H°$ or $\Delta S°$, suggesting that pH 5.1 does not affect the structural integrity of the YNMG loop. RNA concentrations ranged from approximately 0.5 to 1.6 mM, and the RNA was renatured prior to the start of each experiment, as described below.

NMR data were collected on either a Bruker AMX2-500 spectrometer using a 5 mm broadband probe or a Bruker DRX-600 spectrometer using a fixed-frequency triple-resonance (^{1}H, ^{13}C, and ^{15}N) probe. Solvent suppression for exchangeable ^{1}H spectra was achieved using a 3-9-19 pulse sequence with gradients (WATERGATE) (*48*, *49*) and a spectral width of 12 500 Hz, and for nonexchangeable ^{1}H spectra with a low-power presaturation pulse of 3−4 s and a spectral width of 4596 Hz. Proton spectra were collected with a recycle delay of 3−4 s, and referenced to the residual water signal. ^{31}P spectra were acquired with a spectral width of 3049 Hz, a recycle delay of 1.5 s, and proton decoupling (WALTZ-16) (*50*). Spectra were referenced to 85% H_3PO_4 at 0.0 ppm contained in an internal coaxial tube. Data sets were processed using NMRPipe (*51*) or XWINNMR (Bruker). FIDs were zero-filled and apodized with phase-shifted sine-squared bell functions for ^{1}H spectra, or 1 Hz line broadening for ^{31}P spectra.

Homonuclear NOESY and DQF-COSY data were collected in a 95% H_2O/5% ^2H_2O mixture and in 99.96% ^2H_2O, respectively. Quadrature detection in the indirect dimension was achieved using the TPPI−States method (*52*). NOESY data in H_2O were acquired using a mixing time of 400 ms. For all multidimensional experiments, 400−450 FIDs and 2−4K complex points were collected. For each FID, 48−64 scans were acquired with a recycle delay of 3−4 s; data were zero-filled along both the direct and indirect dimensions, and apodized with phase-shifted sine-squared bell functions.

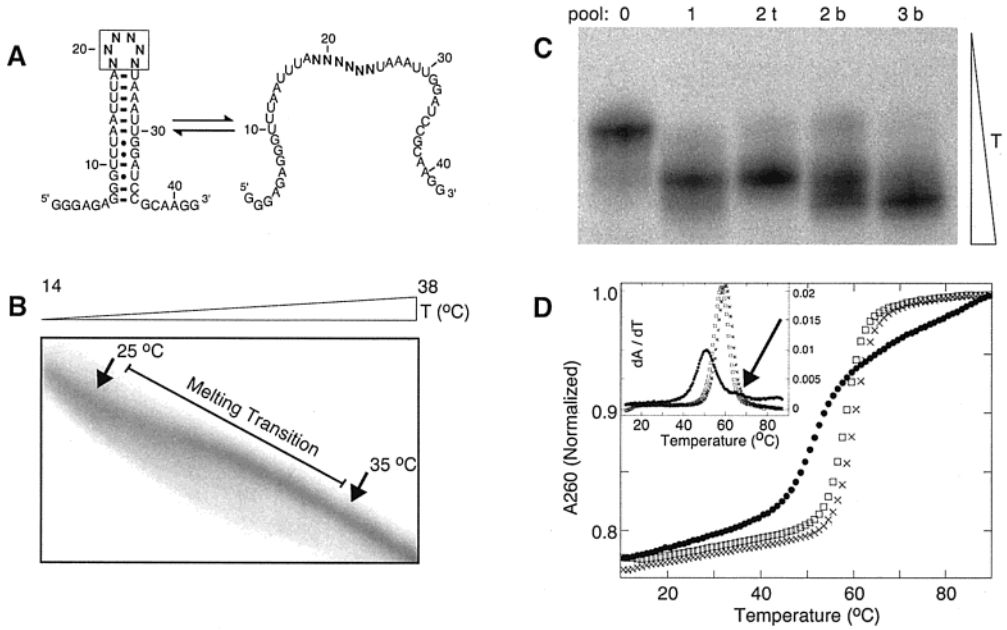

FIGURE 2: Melting and selection of the tetraloop library. (A) Equilibrium between folded and unfolded forms of the RNA library. Watson−Crick base pairs are denoted with a dash and wobble pairs with a dot. N denotes a mixture of A, C, G, and U; 19−24 are the positions of randomization. (B) Perpendicular TGGE of pool 0. Gel temperatures ranged linearly from 14 to 57 °C, with the region between 14 and 38 °C shown. The major melting transition is between 25 and 35 °C, as indicated by arrows, and is broad due to the sequence complexity of pool 0. Collected in $TBEK_{50}$. (C) Parallel TGGE selection showing pools 0−3; t and b denote the top and bottom fractions, respectively, which were excised from the preceding pool. Temperatures in the water baths were 23 and 37 °C. Collected in $TBEK_{50}$. (D) UV melting on mixtures of full-length transcripts for pool 0 (●), 2t (□), or 3b (×); every second data point is omitted for clarity. The normalized absorbance at 260 nm (A_{260}) was obtained by dividing each curve by the maximum absorbance. Pool 3b (×, T_M = 60 °C) is stable relative to pool 2t (□, T_M = 58 °C) and significantly more stable than pool 0 (●, T_M = 51 °C). Collected in $P_{10}E_{0.1}$. The inset shows first-derivative curves calculated using 11-point (5.5 °C window) smoothing for pool 0 (●), 2t (□), and 3b (×). The maximum in the curve gives a qualitative T_M. The larger d*A*/d*T* maxima for the transitions of pools 2t and 3b support a larger $\Delta H°$ for unfolding. The unfolding transition for pool 0 is broader, with a shoulder near the T_M of pools 2t and 3b (see arrow).

Monomer−Dimer Equilibria. At NMR concentrations, stabilities of hairpins and duplexes were similar for several sequences. Solution conditions were chosen to favor the hairpin conformation, although the duplex species could not be entirely eliminated for some samples; these conditions included low RNA (≈0.5 mM) and salt concentrations (<20 mM Na^+), and a four-step renaturation similar to that previously described (*53*, *54*). As an example, the UUCG minihairpin had renaturation temperatures of 60 °C for 1 min (duplex denaturation), 75 °C for 30 s (hairpin denaturation), 60 °C for 1 min (duplex denaturation), and rapid cooling on ice.

Phylogenetic Database Survey. A database of 7484 hairpin loops was constructed using 102 small subunit and 112 large subunit rRNA secondary structures obtained from http://www.rna.icmb.utexas.edu/ (*55*). Within this database, loops were sorted on the basis of sequence and closing base pair. The database contained 734/7484 (10%) triloops, 3805/7484 (51%) tetraloops, 1541/7484 (21%) pentaloops, and 1404/7484 (19%) hexaloops.

RESULTS

Design of the Tetraloop Library. A loop with six randomized nucleotides was chosen for the tetraloop library, which allowed the closing base pair to be probed (Figure 2A). The expectation in creating the library in this fashion was that hexaloop sequences closed by an AU base pair would be significantly less stable than tetraloop sequences with an additional, optimized base pair; this prediction was borne out in our study (Table 1). Diloops were not expected to be as stable as tetraloops due to strain in the loop, although it is possible that a few were present in the final pool.

The fixed sequence portion of the library hairpin was designed to incorporate several features (Figure 2A) (*30*). Primer binding sites for RT-PCR are present at the 5′- and 3′-ends of the hairpin; the stem sequence is designed to avoid slipped pairings, and the distance of the 3′-end of the transcript from the base of the stem helps minimize any potential spurious effect of the random nucleotide added by T7 RNA polymerase to the 3′-end of transcripts. The hairpin library was designed to undergo a large structural change upon unfolding, leading to a significant change in electrophoretic mobility.

Selection of Stable Tetraloop Sequences. The tetraloop library was subjected to gel electrophoresis with temperature gradients perpendicular or parallel to the electric field. Moderate concentrations of urea were added to all gels to move the T_M of the library into a convenient range (*30*). Perpendicular TGGE of pool 0 on a 14 to 57 °C gradient/4 M urea gel resulted in a smeary melting transition between ≈25 and 35 °C (Figure 2B). This transition is consistent with a mixed population of stabilities. Upper and lower baselines were nearly coincident for all hairpins, consistent with similar folded and unfolded structures (Figure 2B). This suggests that differences in mobility on parallel TGGE gels should be associated primarily with differences in stability.

Next, parallel TGGE experiments (Figure 2C) were run between the temperatures of the melting transition in perpendicular TGGE. Parallel TGGE gels were designed to enrich for with stable oligonucleotides, which have faster

electrophoretic mobilities (*30*). Parallel TGGE of pool 0 revealed one major band with a smear below it (Figure 2C), which suggested that the initial library contained some unusually stable sequences. The bottom part of the smear was excised, and the RNA was amplified by RT-PCR followed by forward transcription to afford pool 1. Enrichment of stable sequences was confirmed by faster mobility for pool 1 (Figure 2C). In a similar fashion, pool 2-top (2t) and pool 2-bottom (2b) were obtained from the top and bottom of pool 1, respectively. Pool 3-bottom (3b) was from further enrichment of the bottom of pool 2b, and was chosen for sequencing and characterization. Electrophoresis of all selected pools under completely denaturing conditions revealed identical mobilities, as expected (data not shown).

Melts of various RNA pools were conducted (Figure 2D). The average T_M for pool 3B was more than 10 °C higher than that for pool 0, and the transitions for pools 2t and 3b were much steeper than for pool 0 (Figure 2D and inset). In addition, the average melting temperature for pool 3b was slightly greater than that for pool 2t. A plot of d*A*/d*T* versus temperature revealed a shoulder near 60 °C for pool 0 (Figure 2D inset), consistent with a small population of stable sequences. These results suggest that the selections were successful in isolating hairpins with enhanced stability.

Identification of Stable RNA Hairpins. Among the 48 isolated clones (Table 1), 47 have a CG closing base pair and one has a GC. The loop with the GC closing base pair is CUUG; this tetraloop-closing base pair combination is common in rRNA and is unusually stable (*11*, *14*). Of the 47 loop sequences closed by a CG base pair, 20 (42%) have a sequence belonging to UHCG (where H is A, C, or U), which is a subset of the UNCG motif; this tetraloop-closing base pair combination is very common in rRNA and is extremely stable (*11*, *14*). Recovery of known stable tetraloop-closing base pair combinations suggests that unknown stable tetraloops may be present. Overall, inspection of the selected sequences reveals that 42 (88%) can be described by the motif cYNMGg (M is A or C; Y is C or U), although some representatives of this motif were absent. Curiously, there is a paucity of guanine at position 2 of the loop, with only one of the 42 YNMG sequences having a G at this position. This could be due to dimerization and the unusual electrophoretic mobility of these RNA molecules. Importantly, the lack of G at position 2 does not appear to be due to the decreased thermodynamic stability of the hairpin conformation of these sequences (see the next section). Other loops that were common in the selection include 10 (21%) CDAG (where D is A, G, or U), six (13%) CACG, six (13%) UHAG, and three (6%) cCUUGg sequences.

Thermodynamic Characterization of Minihairpins by UV Melting. To determine the stability of YNMG loops, UV melting experiments were conducted on minihairpins. Minihairpins are stem−loop structures containing the selected loop/closing base pair combination, but with stems of only four base pairs (*28*). These smaller hairpins were used to facilitate two-state behavior and to suppress alternative folds and kinetically trapped dimers. Minihairpins were constructed and melted for each of the 16 sequences in the YNMG motif and for control sequences. Examination of representative melting profiles reveals that YNMG loops are especially stable, having T_M values that are on average 6.2 °C higher than position 1 and 3 controls, and transitions that are significantly sharper (Figure 3 and Table 2).

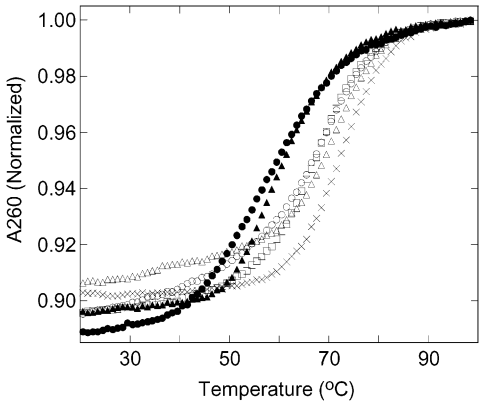

FIGURE 3: Representative UV melting curves for minihairpins from each of the four subfamilies and controls. Loop sequences are as follows: cUUCGg (×), cCUAGg (○), cCUCGg (△), cUUAGg (□), cAACGg (●), and cUAGGg (▲); thermodynamic parameters are provided in Table 2. Collected in $P_{10}E_{0.1}$ (pH 7.0). Every second data point is omitted for clarity.

In all cases, melts were conducted at C_T values between ≈1 and 250 μM. Normalized melting profiles and $\Delta H°$ and $\Delta S°$ parameters were independent of strand concentration between 1 and 100 μM for all oligonucleotides, and between 100 and 250 μM for most oligonucleotides with the exception of CUAG, CUCG, CACG, all position 3 controls, and AACG. These exceptions may occur because CUAG, CUCG, and CACG can form CG base pairs in the duplex conformation; position 3 controls maintain the potential for base pairing between loop positions 1 and 4 in the duplex state, and lack the extra stability in the hairpin state associated with the YNMG motif, and AACG has the least favorable thermodynamic parameters for hairpin formation (Table 2). The YNMG sequences were organized into four subfamilies: CNAG, CNCG, UNAG, and UNCG (Tables 1 and 2). Among these, the UNCG subfamily showed the most favorable $\Delta G_{37}°$ and T_M values, ranging from −4.4 to −4.9 kcal/mol and from 70.8 to 72.9 °C, respectively, similar to those from previous studies (*14*). Note that the UGCG sequence is similar in stability to other UNCG sequences (Table 2), indicating that the absence of G at position 2 in the selection is not for a thermodynamic reason. The CNAG, CNCG, and UNAG subfamilies were only slightly less stable than the UNCG sequences, with $\Delta\Delta G_{37}°$ and ΔT_M values relative to UUCG ranging from 0.04 to 1.5 kcal/mol and from −1.6 to −6.5 °C, respectively. The CUCG sequence was exceptionally stable among the non-UNCG sequences, with a stability similar to those of UNCG sequences. Similar thermodynamic stability among the 16 YNMG sequences suggested that their structures may be related.

To test the importance of Y at position 1 and M at position 3, we changed position 1 to A or G and position 3 to G or U. For all controls, position 4 was a G and position 2 was an A since adenines have a low tendency to form alternative folds, and A was common at position 2 in the selections (Table 1) and in rRNA YNMG loops (*13*). Position 1 and 3 control sequences showed thermodynamic values significantly less favorable than those of YNMG sequences, with $\Delta\Delta G_{37}°$ and ΔT_M values relative to UUCG ranging from 1.5 to 3.1 kcal/mol and from −6.6 to −14.6 °C, respectively.

Table 3: Thermodynamic Parameters for Minihairpin Formation by Chimeric Sequences[a]

sequence[b]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G_{37}°$ (kcal/mol)	$\Delta\Delta G_{37}°$ (kcal/mol)	$\Delta G_{55}°$ (kcal/mol)	$\Delta\Delta G_{55}°$ (kcal/mol)	T_M (°C)	ΔT_M (°C)
CAAG	−44.4 ± 0.8	−130.2 ± 2.2	−4.00 ± 0.07	–	−1.69 ± 0.08	–	67.7	–
CAAG	−38.6 ± 3.3	−115.0 ± 10.0	−2.99 ± 0.21	1.01 ± 0.22	−0.92 ± 0.07	0.68 ± 0.12	63.0	−4.7
CAAG	−41.5 ± 1.6	−122.5 ± 5.0	−3.54 ± 0.08	0.46 ± 0.11	−1.34 ± 0.04	0.26 ± 0.11	65.9	−1.8
CACG	−37.4 ± 0.6	−109.5 ± 2.0	−3.56 ± 0.05	–	−1.60 ± 0.03	–	69.6	–
CACG	−39.3 ± 2.0	−117.5 ± 6.1	−2.84 ± 0.15	0.72 ± 0.15	−0.73 ± 0.08	0.97 ± 0.13	61.2	−8.4
CACG	−45.6 ± 2.4	−132.9 ± 7.0	−4.35 ± 0.20	−0.79 ± 0.21	−1.95 ± 0.08	0.25 ± 0.13	69.7	0.1
UAAG	−41.7 ± 0.5	−122.8 ± 1.4	−3.62 ± 0.11	–	−1.40 ± 0.10	–	66.4	–
UAAG	−38.6 ± 3.0	−115.4 ± 8.9	−2.81 ± 0.29	0.81 ± 0.31	−0.73 ± 0.15	0.67 ± 0.18	61.3	−5.1
UAAG	−42.2 ± 1.4	−125.2 ± 4.4	−3.34 ± 0.07	0.28 ± 0.13	−1.08 ± 0.03	0.32 ± 0.11	63.7	−2.7
UUCG	−46.4 ± 2.5	−134.2 ± 7.3	−4.82 ± 0.25	–	−2.40 ± 0.14	–	72.7	–
UUCG	−42.2 ± 2.3	−124.9 ± 7.0	−3.48 ± 0.17	1.31 ± 0.29	−1.23 ± 0.07	1.17 ± 0.12	64.9	−8.1
UUCG	−45.6 ± 1.9	−131.4 ± 5.8	−4.79 ± 0.11	0.00 ± 0.25	−2.43 ± 0.09	0.03 ± 0.13	73.5	0.5

[a] Minihairpins have the sequence 5′-ggac$L_1L_2L_3L_4$gucc. Parameters were obtained as described in Table 2. Experiments were performed over a C_T range of 3−60 μM. T_M values were independent of concentration over this range. Parameters are the average of at least six independently prepared samples at varying concentrations, and errors are standard deviations from repeated experiments. Differences in free energy and melting temperature are relative to those of the unsubstituted sequence, and errors were propagated by standard methods (*70*). Values for $\Delta G°$ are also provided at 55 °C since this is closer to the T_M and may be somewhat more accurate than $\Delta G_{37}°$ values. [b] Sequences are for L_1−L_4 and are listed alphabetically by YNMG subfamily. Underlined nucleotides denote the positions of deoxy substitution.

One exception was GAAG; this sequence was the least destabilizing, with $\Delta\Delta G_{37}°$ and ΔT_M values similar to those of the least stable YNMG members (Table 2). This may be due in part to the resemblance between GAAG and the GNRA motif (see CD Results).

The majority of the loops (47 of 48) had a CG closing base pair (Table 1), including all sequences that conform to the YNMG motif. Previous experiments indicated that the closing base pair can make an exceptional contribution to stability. For example, a CG to GC change of the closing base pair for the loop sequence UUCG has a $\Delta\Delta G_{37}°$ of 2.3 kcal/mol and a ΔT_M of 11.6 °C (*14*), and the same closing base pair change for the stable triloop sequence UUA has a $\Delta\Delta G_{37}°$ of 2.1 kcal/mol and a ΔT_M of 15.9 °C (*28*). CG to GC closing base pair changes for CAAG, CACG, and UAAG gave $\Delta\Delta G_{37}°$ values of 1.5, 0.9, and 1.9 kcal/mol and ΔT_M values of −7.2, −8.5, and −12.5 °C, respectively (Table 2), consistent with the importance of a CG closing base pair. In summary, melting experiments on YNMG members and variants indicate that Y and M at loop positions 1 and 3, respectively, and the CG closing base pair make important contributions to stability.

Thermodynamic Characterization of Single 2′-Deoxy-Substituted Minihairpins. The contributions of 2′-hydroxyl groups to the stability of the UUCG tetraloop have been probed experimentally and computationally (*25*−*27*). In particular, 2′-H substitution at U_{L1} resulted in a $\Delta\Delta G_{37}°$ of 1.1 kcal/mol, consistent with hydrogen bonding of this 2′-OH (Figure 1A), whereas substitution at U_{L2} resulted in a $\Delta\Delta G_{37}°$ of −0.08 kcal/mol, consistent with the absence of hydrogen bonding (*25*). Theoretical considerations indicated the basis for these effects is complex and involves an interplay of hydrogen bonding, solvation, intrinsic sugar pucker preferences, and sequence-dependent effects in the single-stranded state (*26*).

To test for similar patterns in YNMG loops, chimeric oligonucleotides were prepared in which independent deoxy substitutions were made at loop positions 1 and 2, and thermodynamic stabilities were determined by UV melting (Table 3). For these experiments, position 2 was fixed as an A for non-UNCG sequences as discussed above, and as a U for UNCG sequences to allow comparisons with literature values. Substitution of a deoxy at position 1 resulted in significant destabilization. For CAAG, CACG, UAAG, and UUCG, the $\Delta\Delta G_{37}°$ ($\Delta\Delta G_{55}°$) values were 1.0 (0.7), 0.7 (1.0), 0.8 (0.7), and 1.3 (1.2) kcal/mol and the ΔT_M values were −4.7, −8.4, −5.1, and −8.1 °C, respectively (Table 3). In each case, comparison is made to the all-ribose loops of the same sequence. The $\Delta\Delta G_{55}°$ and ΔT_M values for UUCG are similar to published values under similar conditions (*25*−*27*). Substitutions at position 2 resulted in smaller $\Delta\Delta G°$ and ΔT_M effects. For CAAG, CACG, UAAG, and UUCG, $\Delta\Delta G_{37}°$ ($\Delta\Delta G_{55}°$) values were 0.5 (0.3), −0.8 (0.25), 0.3 (0.3), and 0.0 (0.0) kcal/mol and ΔT_M values were −1.8, 0.1, −2.7, and 0.5 °C, respectively. Overall, the position dependence of deoxy substitutions in the YNMG loops parallels the published trends for UNCG (*25*−*27*), with position 1 having a destabilizing effect and position 2 substitutions having little or no effect.

Interestingly, 2′-deoxy substitution at position 1 of the loop is less penalizing for YNAG sequences than for YNCG sequences; compare average $\Delta\Delta G_{55}°$ values of 0.7 versus 1.1 kcal/mol, and average ΔT_M values of −4.9 versus −8.3 °C for YNAG and YNCG sequences, respectively. This difference may be because YNAG loops have enhanced flexibility, which better tolerates these changes (see NMR Results and Discussion). A similar effect of enhanced tolerance of ribose substitutions was found when the closing base pair of the UUCG loop was changed from a CG pair to a GC pair, which was attributed to increased loop flexibility with a GC closing base pair (*27*).

Characterization of Minihairpins by CD. Circular dichroism is effective at reporting certain changes in nucleic acid conformation. For example, A-form dsDNA and B-form dsDNA have differing behavior at 260 and 280 nm (*56*). Importantly, secondary structure defects can lead to deviations from this spectrum (*57*). CD spectra were similar for YNMG sequences, whereas significant differences were found for some controls (Figure 4, and Figure 1 of the Supporting Information). For YNMG sequences, the spectra had a negative minimum between 235 and 240 nm, a maximum between 260 and 263 nm, and a shoulder near 280 nm with positive ellipticity. The first two features resemble standard A-form RNA, whereas the third feature

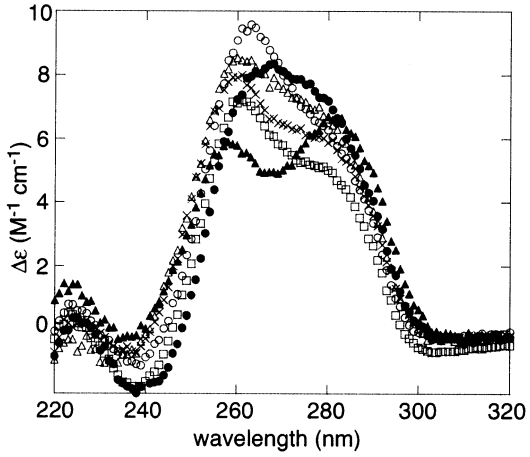

FIGURE 4: Representative CD spectra for minihairpins from each of the four YNMG subfamilies and controls. Loop sequences are as follows: UCAG (□), UCCG (○), CCAG (×), CCCG (△), UUUU (●), and GGAA (▲). Collected at 25 °C in $P_{10}E_{0.1}$ (pH 7.0). Additional spectra are available in the Supporting Information.

is unique. Spectra for the UNCG and CNAG sequences were identical within their respective subfamilies and similar to each other (Figure 1 of the Supporting Information). The UNAG subfamily also had similar spectra that were related in shape to those of UNCG and CNAG, except for UAAG, which had more intense positive and negative peaks (Figure 1C of the Supporting Information). The unique spectrum of UAAG could be due to stacking of position 2 in the loop (see the Discussion). Nevertheless, no evidence was found for unusual thermodynamic behavior (Tables 2 and 3) or anomalous structures of UAAG at low temperatures (see NMR Results). The CNCG subfamily of sequences had spectra that exhibited a less distinct shoulder at 280 nm; however, the 260 nm/280 nm ratios were similar among all 16 YNMG loops and ranged from 1.29 to 1.86 (Figure 1 of the Supporting Information).

Spectra of several non-YNMG loops were dissimilar to those of the YNMG loops in both shape and relative signal intensity at 260 and 280 nm (Figure 4, and Figure 1 of the Supporting Information). The UUUU sequence did not show a distinct maximum near 260 nm or a shoulder near 280 nm. Strikingly, the GGAA tetraloop possessed a bimodal spectrum with a maximum near 280 nm slightly greater in intensity than a maximum near 260 nm (Figure 4). The difference between the GNRA and UNCG spectra is consistent with their different structures (*16*, *22*). Position 1 control sequences showed low 260 nm/280 nm ellipticity ratio values in the range of 1.16−1.35 (Figure 1 of the Supporting Information). Among these, GAAG had the smallest 260 nm/280 nm ellipticity ratio (1.16), which is lower than the values for any of the YNMG loops. This ratio tends toward that for GGAA (0.87), suggesting that the GAAG RNA may adopt a fold with GNRA-like stacking interactions, which may account for its unusual stability among the controls. The CD spectra of position 3 control sequences were similar to each other and to those of YNMG loops (Figure 1 of the Supporting Information). Since circular dichroism is sensitive to stacking, these similarities may reflect a fold with some similarity to that of the YNMG family at the temperature of the CD experiments (25 °C) but without extensive hydrogen bonding and concomitant stability. In summary, CD spectra for YNMG sequences were similar and differed from some of the controls, supporting similarity among the YNMG loops.

Structural and Dynamic Characterization of Minihairpins by Exchangeable Proton NMR. To test the structural relatedness of the YNMG loops in greater detail, we carried out one-dimensional (1D) ^{1}H and ^{31}P, NOESY, and DQF-COSY NMR experiments on representative sequences from each subfamily. As described, concentrations of RNA required for NMR resulted in varying degrees of dimerization, which complicated some of the spectra. Nevertheless, a number of spectral features were found for YNMG loops that were similar to features for the UUCG loop.

1D imino ^{1}H spectra were collected for CAAG, CACG, UAAG, and UUCG (Figure 5). The resonance near 14.3 ppm, choosing the slightly more downfield-shifted resonance in the case of overlapping peaks (*54*), was tentatively assigned to U10hH3 (h refers to the hairpin conformation) (Figure 5) (*58*). This was confirmed by NOE connectivities in the NOESY data to neighboring G9hH1 and G2hH1 (Figure 6), and by strong NOEs to the aromatic region (7−8.5 ppm). Likewise, the peak at 12.5 ppm was assigned to G2hH1, and was confirmed on the basis of NOE connectivity to neighboring U10hH3 (Figure 6). The broad resonance near 12.5 ppm is tentatively assigned to the terminal G1hH1 in exchange with water (*4*). Since Watson−Crick GC base pairs that close internal or hairpin loops are typically shifted to ≈13.5 ppm (*59*, *60*), the resonance near 13.5 ppm was tentatively assigned to G9hH1. For G9H1 of CAAG and CACG, overlap between hairpin and duplex resonances was not as severe as for other stem residues, most likely because G9 samples the different chemical environments in a hairpin loop and an internal loop. Assignment of G9hH1 was confirmed by NOE connectivities to U10hH3 and in some cases to G8hH1 of the loop (Figure 6). Overall, NOEs for the stem imino protons were observed between U10 and G2 and between U10 and G9, but not between G2 and G9 (Figure 6). Similarity of chemical shifts for G8hH1 and G9hH1 in the four YNMG loops (Figure 5) is consistent with similar folds.

G8hH1 of the UUCG tetraloop has an intense resonance near 10 ppm that serves as an indicator of the UNCG tetraloop motif (Figure 5) (*20*, *21*, *61*). An unusually upfield-shifted resonance was also observed for UAAG, CAAG, and CACG between 9.6 and 10.2 ppm, and was tentatively assigned to G8hH1. This assignment was confirmed by a weak cross-peak in the NOESY spectrum to G9hH1 (Figure 6) for UAAG, and by comparison to UAAG and process of elimination for CAAG and CACG. The inability to observe the G8hH1−G9hH1 cross-peak in the NOESY spectrum for CAAG and CACG may be due to the large interproton separation of 5.15 Å seen in the crystal structure of the UUCG loop (*24*). Other structural features known to lead to chemical shifts in this region include GdH1 and UdH3 protons located in GU wobble base pairs (*59*). This alternative was dismissed on the basis of the assignment of the GU wobble of the duplex in the NOESY spectrum (see intense cross-peaks in Figure 6), and the absence of duplex-like concentration dependence for the resonance assigned to G8hH1.

The G8hH1 assignment accounts for all hairpin loop imino protons in CAAG and CACG. For UAAG, U5hH3 was assigned to the remaining hairpin resonance at ≈11.2 ppm

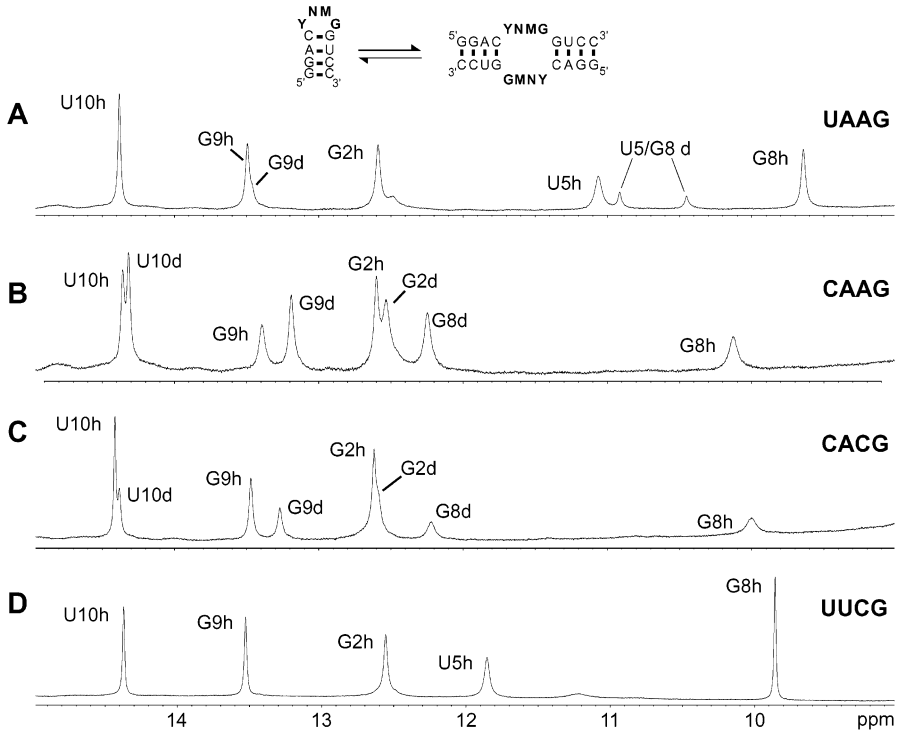

FIGURE 5: 1D ^{1}H NMR spectra for minihairpins showing the imino proton spectral region. Loop sequences and concentrations are as follows: (A) 0.7 mM cUAAGg, (B) 0.6 mM cCAAGg, (C) 0.6 mM cCACGg, and (D) 1.1 mM cUUCGg. "h" and "d" denote hairpin and duplex conformations, respectively. The equilibrium between the motif in the hairpin and duplex conformations is depicted at the top. Assignments are based upon standard NOESY methods (see Figure 6) as well as comparison of spectra at varying strand concentrations; U5d and G8d of UAAG and several other resonances remain unassigned. Spectra were collected at 600 MHz in a 95% H_2O/5% ^2H_2O mixture (v/v) in $P_{10}E_{0.1}$ (pH 5.1) at 1 °C, and referenced to internal $^1H^2HO$.

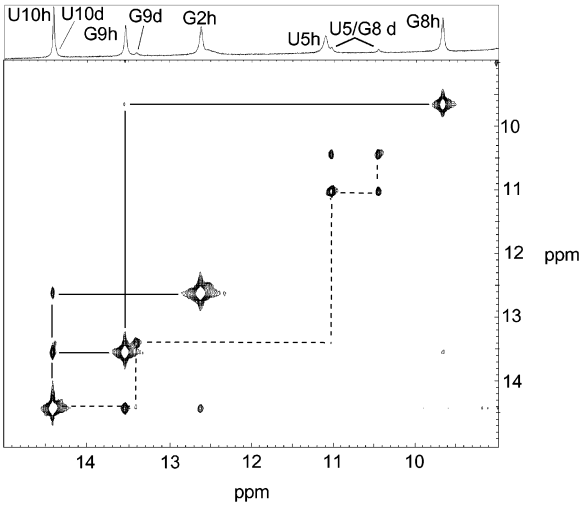

FIGURE 6: Detail of a NOESY map of the cUAAGg minihairpin for the imino proton spectral region (τ_m = 400 ms). Assignments are indicated in the 1D spectrum at the top; U5d and G8d assignments remain ambiguous. A mixture of hairpin (h) and duplex (d) conformations was observed. Internucleotide cross-peaks are indicated by a solid line for hairpin resonances (above diagonal) and a dashed line for duplex resonances (below diagonal). Spectra were collected in a 95% H_2O/5% ^2H_2O mixture (v/v) with $P_{10}E_{0.1}$ (pH 5.1) at 1 °C, and referenced to internal $^1H^2HO$.

(Figure 5), similar to the published chemical shift in UUCG (*21*). This resonance broadens out with temperature (Figure 2 of the Supporting Information) and was not observed in NOESY data at long mixing times (400 ms) (Figure 6). The broad resonance at 11.2 ppm in the UUCG spectrum may be from U6, and it had a temperature dependence similar to that previously reported (*62*).

The resonance for G8hH1 at 1 °C was sharp for UUCG, consistent with previous studies (*21*, *61*), but was broader for UAAG and especially for CAAG and CACG (Figure 5). This suggested that the non-UNCG YNMG loops might be more dynamic than the UUCG loop. To investigate this, the temperature dependence of the imino proton region was examined. Comparison of the intensities and line widths of loop resonances relative to stem resonances can provide insight into loop dynamics (*15*, *21*). The temperature dependence of the stem resonances for the non-UNCG YNMGs was complicated by spectral overlap with duplex resonances, especially at elevated temperatures. Nevertheless, we were able to compare G8hH1 to G2hH1 clearly in most instances (Figure 2 of the Supporting Information). G2hH1 was chosen as a reference since it is located in a GC base pair in the internal portion of the stem, and its exchange rate is expected to be relatively unchanged between 5 and 35 °C (*15*, *21*). The UUCG G8hH1 resonance did not change appreciably over this temperature range (data not shown), consistent with previous reports (*15*, *21*). In contrast, the G8hH1 resonances for UAAG and CAAG broadened out considerably between 5 and 35 °C relative to G2hH1 (Figure 2 of the Supporting Information). Interestingly, the G8hH1 resonance of CACG, while most broad at 1 °C (Figure 5), did not broaden further even at temperatures as high as 35 °C (Figure 2 of the Supporting Information). In addition, the closing base pair G9hH1 resonance of CACG did not broaden significantly with temperature, whereas it did for UAAG and CAAG (Figure 2 of the Supporting Information). These data suggest that YNAG loops may be more dynamic than YNCG loops. Similar effects were reported for the

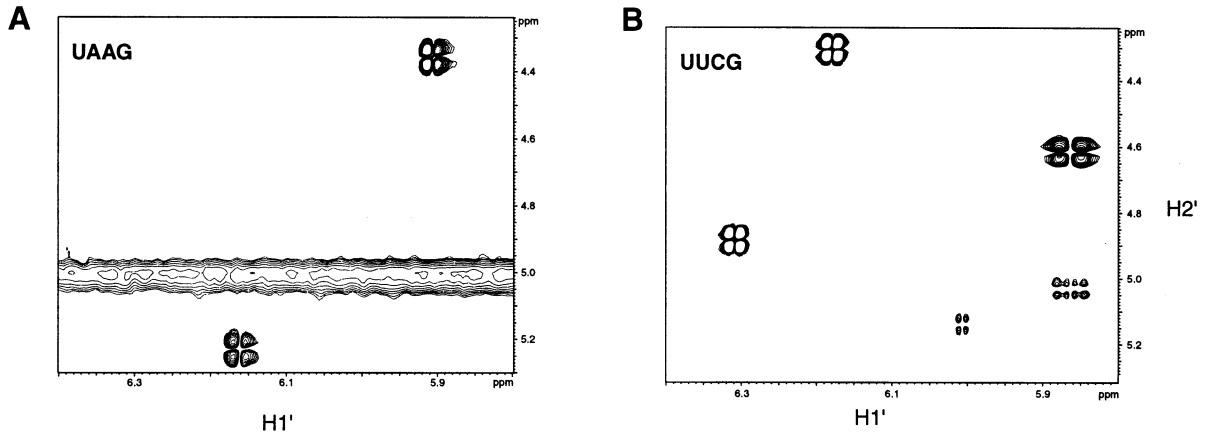

FIGURE 7: Details of DQF-COSY maps of cUAAGg and cUUCGg minihairpins for the H1′−H2′ spectral region: (A) cUAAGg (0.7 mM) at 1 °C and (B) cUUCGg (1.1 mM) at 20 °C. Two intense cross-peaks were observed in each spectrum, consistent with a C2′-endo conformation for positions 2 and 3. A reduced temperature was used for UAAG in an effort to reduce the flexibility found for YNAG sequences. Spectra were collected in 99.96% ^2H_2O in $P_{10}E_{0.1}$ (pD 5.5).

UUCG loop upon changing the closing base pair from CG to GC (*15*).

Although the shape and intensity of the G8hH1 resonance of CACG were not particularly sensitive to temperature, it was broad at all temperatures. This could be because G8hH1 has less optimal hydrogen bonding to O2 of C_{L1}, or because interaction with the closing base pair is less optimal. Either possibility might allow for enhanced chemical exchange with water, also consistent with the slightly less favorable thermodynamic parameters for CACG (Table 2).

Structural Characterization of Minihairpins by Nonexchangeable Proton NMR. Base and sugar H1′ protons were tentatively assigned on the basis of sequential connectivities from 400 ms NOESY experiments for UUCG and UAAG. These sequences were chosen since they exhibit the least propensity to form duplexes (Figure 5). Assignments of H5−H6 cross-peaks were confirmed by DQF-COSY data. Internucleotide NOEs were missing between U5 and U6 for UUCG and between A6 and A7 for UAAG, likely resulting from motions of the loop nucleotides. Cross-peak patterns and intensities in the base−H1′ region were consistent with the formation of an A-form stem containing anti glycosidic torsion angles (*47*). Significantly, the base−H1′ region of the NOESY spectra of both sequences had one additional intense cross-peak (data not shown). This cross-peak was more intense than the pyrimidine H5−H6 cross-peaks, consistent with a syn base in both RNAs.

Sugar conformations for the UUCG and UAAG loops were also qualitatively characterized using DQF-COSY experiments (Figure 7). For UUCG, two intense cross-peaks were observed in the H1′−H2′ region, consistent with the existence of two residues possessing DNA-like C2′-endo sugar puckers (Figure 7). These cross-peaks were assigned to U_{L2} and C_{L3} on the basis of comparison to previous NMR studies (*20*, *21*). Similarly, two intense cross-peaks were observed in the H1′−H2′ region of the DQF-COSY spectrum for UAAG, consistent with the presence of the same number of residues possessing C2′-endo sugar puckers (Figure 7). These cross-peaks were tentatively assigned to A_{L2} and A_{L3} by comparison to the UUCG tetraloop (*20*, *21*). In addition, weak H1′−H2′ cross-peaks were observed for UUCG, suggesting the presence of residues sampling both C2′-endo and C3′-endo sugar puckers. These features were also present for UAAG at slightly upfield values. In summary, observation of the nonexchangeable spectral features characteristic of two C2′-endo sugar puckers and a syn base in representative YNMG loops strongly supports structural similarities among the different RNAs. NOESY experiments were also performed on CACG and CAAG (data not shown). Although the data were complicated by the duplex conformation, an intense NOE was observed at the position expected for a syn base, supporting the structural similarity of these two sequences to the UNCG loop as well.

Structural Characterization of Minihairpins by Phosphorus NMR. 1D ^{31}P spectra were collected for representative sequences (Figure 8). Because of complications from the duplex conformation, especially problematic for CAAG and CACG, spectra were collected at approximately 2-fold different concentrations to assign resonances as hairpin or duplex (Figure 8). Calculations using free energies measured for hairpins (Table 2) and nearest-neighbor free energies predicted for duplexes (*37*−*39*) indicated that a 2-fold decrease in RNA concentration should result in an approximately 30−40% increase in hairpin population. Moreover, the calculations suggested that the fraction of duplex should decrease at elevated temperatures; thus, spectra were acquired at 45 °C, which is still in the folded baseline (Figure 3). After accounting for phosphate buffer, which has a pH-dependent chemical shift, we observed 11 resonances for UUCG (Figure 8A), and for UAAG at both strand concentrations (Figure 8B), consistent with these sequences being nearly 100% in the hairpin conformation. Importantly, an unusually downfield shifted peak was found in each case, which has been previously assigned to G9P (*21*). Observation of this feature for UAAG is consistent with a similar backbone conformation.

Spectra for CAAG and CACG exhibited more than 11 resonances (Figure 8C,D). Certain resonances have an increase in intensity with concentration (red line in Figure 8), and were assigned to the duplex state. Three upfield-shifted resonances, common to all four sequences, and the downfield-shifted resonance did not change in intensity with concentration and were thus assigned to the hairpin. Confirmation of hairpin assignments was obtained from the effect of temperature upon ^{31}P resonances (data not shown). Substantial increases in hairpin/duplex ratios were observed

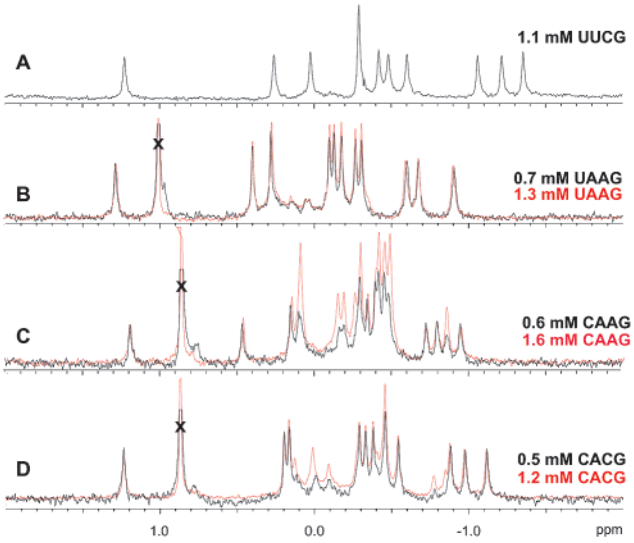

FIGURE 8: 1D ^{1}H-decoupled ^{31}P spectra for minihairpins. Loop sequences and concentrations were as follows: (A) 1.1 mM cUUCGg, (B) 0.7 (black) and 1.3 mM cUAAGg (red), (C) 0.6 (black) and 1.6 mM (red) cCAAGg, and (D) 0.5 (black) and 1.6 mM (red) cCACGg. Eleven resonances were expected and were observed for the hairpin conformation for UUCG (two resonances near −0.29 ppm overlap significantly) and UAAG. The intense resonance near 1 ppm (denoted by **X**) was pH-dependent and was assigned to the phosphate buffer. For panels B−D, spectra are normalized using the three well-resolved upfield peaks observed in the hairpin conformations of UUCG and UAAG. Significant increases in intensity were observed for some CAAG and CACG resonances at different concentrations, and these were assigned to the duplex conformation. The downfield-shifted resonance at approximately 1.2 ppm is from the hairpin conformation and is tentatively assigned to G9P based upon comparison to published spectra for UUCG (*21*). Spectra were collected in $P_{10}E_{0.1}$ [(A) pH 6.5 and (B−D) pH 5.1] at 45 °C, and referenced to 85% phosphoric acid in an internal coaxial tube.

with increasing temperature between 25 and 45 °C, as expected. Similarity of the chemical shift for the downfield-shifted hairpin ^{31}P resonance for CAAG, CACG, UAAG, and UUCG is consistent with all four tetraloops adopting similar conformations. The exchangeable ^{1}H spectra data supported enhanced flexibility for YNAG loops relative to YNCG loops (see above). In contrast, the phosphorus data were not especially temperature dependent for the YNCG or YNAG loops, even to 45 °C. This suggests that motion in the YNAG loops up to the physiological temperature is local and does not involve complete unfolding of the loop.

Biological Significance of Loop Sequences As Determined by a Phylogenetic Database Survey. To assess the potential biological importance of the YNMG loop motif, we examined its occurrence in nature. We focused on ribosomal RNA since a rich phylogenetic database is available (*55*). Woese and co-workers originally pointed out that certain tetraloops are unusually common in rRNA (*11*). They identified three general tetraloop motifs: GNRA, UNCG, and, less frequently, CUUG. Also, the closing base pairs were found to be largely CG for UNCG loops and GC for CUUG loops. In an effort to update these statistics and assess the potential biological importance of non-UNCG YNMG loops, we constructed a database of rRNA hairpin loops (see Materials and Methods). A survey of 102 small subunit and 112 large subunit rRNAs revealed that 213/1357 (16%) small subunit tetraloops and 340/2448 (14%) large subunit tetraloops have the YNMG motif. By comparison, 754/1357 (56%) small subunit and 1184/2448 (48%) large subunit tetraloops have the GNRA motif, and 19/1357 (1.4%) small subunit and 20/2448 (0.8%) large subunit rRNA tetraloops have the CUUG motif. Thus, YNMG sequences are more common than CUUG sequences, but less common than GNRA sequences. We also examined how the frequencies of YNMG sequences are split between UNCG and non-UNCG sequences. The following frequencies were observed for YNMG subfamilies in small and large subunits, respectively: 161/1357 (12%) and 237/2448 (9.7%) UNCG, 17/1357 (1.3%) and 68/2448 (2.8%) UNAG, 15/1357 (1.1%) and 32/2448 (1.3%) CNCG, and 20/1357 (1.5%) and 3/2448 (0.12%) CNAG. Thus, many of the non-UNCG YNMG sequences are as common or more common than CUUG. The thermodynamic parameters for hairpin formation of ggagCUUGcucc, under solution conditions similar to those used in the study presented here, were as follows: $\Delta H^\circ = -47.5$ kcal/mol, $\Delta S^\circ = -142$ eu, and $\Delta G_{37}^\circ = -3.6$ kcal/mol, with a T_M of 62.4 °C (*14*). These parameters are similar to, or somewhat less stable, than those for non-UNCG YNMG loops (Table 2). Thus, it appears that non-UNCG YNMG loops are similar in stability to gCUUGc loops, and similar in occurrence in rRNA.

A comparison of RNA folds of 16S and 16S-like rRNAs by Konings and Gutell (*13*) revealed that tetraloops corresponding to *E. coli* loop 343 are largely YAMG sequences closed by a CG base pair. (Throughout, "loop 159" and "loop 343" refer to hairpin loops in any 16S rRNA corresponding to loops beginning with positions 159 and 343, respectively, in *E. coli*.) At loop 343, archaea, bacteria, and chloroplasts are 99, 98, and 97% UACG, respectively; mitochondria are 54% CACG, and eukarya are 5% UACG, 19% UAAG, and 74% CAAG (*13*). These pioneering phylogenetic studies established the importance of conservation at loop 343; however, the molecular basis for this conservation was not clear at the time since the relatedness of the YNMG loops was unknown and the structure of the ribosome was unavailable (*13*).

A recent solution of the crystal structures of the 30S and 70S ribosomes at or near atomic resolution has revealed a role for loop 343 in an RNA−RNA tertiary interaction (*23*, *63*). The first two As of the GAAA tetraloop in loop 159 of helix 8 (H8) of 16S rRNA interact with loop 343 of helix 14 (H14) (Figure 9A). The structure of cUACGg at loop 343 is very similar to the structures of isolated UNCG loops (*20*−*22*, *24*), with all of the known hydrogen bonds and stacking interactions maintained and a 0.5 Å rmsd relative to cUUCGg from a crystal structure (*24*).

Direct interactions with loop 343 involve primarily its CG closing base pair, which has been shown herein to be crucial for thermodynamic stability (Table 2). A_{L2} of loop 159 interacts with the minor groove of the closing G of loop 343, with a hydrogen bond from A_{L2} N1 to GH2, and from A_{L2} H6 to GN3 (Figure 9B). In addition, A_{L3} N1 of loop 159 accepts a hydrogen bond from the 2′-OH group of the closing base pair G of loop 343. The interaction involving A_{L3} appears to conform to the A-minor motif, in which the minor groove of an interacting A is positioned in the minor groove of a receptor base pair (*23*, *64*). A-minor motifs prefer CG base pair receptors and often hydrogen bond with 2′-OH groups. Interaction of A_{L2} of loop 159 does not conform to the A-minor motif since N6 is involved in the interaction.

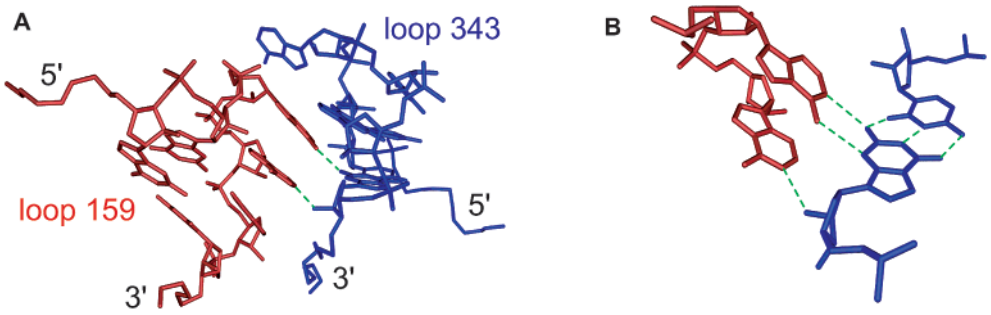

FIGURE 9: Involvement of a cYNMGg tetraloop in a tertiary interaction in 16S rRNA. The interaction is between gGAAAc of loop 159 from H8 (red) and cUACGg of loop 343 from H14 (blue) of 16S rRNA. (A) Overview of the interaction. The GC closing base pair and stacking of the three consecutive loop As are visible for loop 159. Loop 159 is nestled under A_{L2} of loop 343. Hydrogen bonding is depicted with green lines. (B) Details of hydrogen bonding interactions. The three hydrogen bonds of the Watson−Crick CG closing base pair of loop 343 (dashed) are shown. A_{L2} interacts with the minor groove of the G of the CG closing base pair of loop 343, with a hydrogen bond from GH2 to A_{L2} N1, and from GN3 to A_{L2} H6. In addition, A_{L3} N1 accepts a hydrogen bond from the G2′ OH group. Drawings were generated from the crystal structure of 30S rRNA [PDB entry 1FJF (*23*)] using WebLab Viewer (MSI).

DISCUSSION

It is of interest to identify nucleic acid sequences with exceptional stability, and to understand the molecular basis for their stability. Stable structural elements can help build complex nucleic acid tertiary structures and can serve important biological roles, including nucleating folding and regulating biological phenomena such as transcriptional pausing and replication (*3*−*6*). Slight decreases in stability may also favor tertiary structure formation and protein binding (see below). In addition, knowledge of the thermodynamic rules for formation of such sequences might improve RNA structure prediction (*39*, *65*, *66*). However, many secondary and tertiary structural motifs in RNA and DNA have sufficient sequence complexity to make direct study of all possible variants impractical. In an effort to address the issues of sequence complexity and thermodynamic stability, we developed a combinatorial approach for isolating thermodynamically stable and unstable sequences from a combinatorial library by combining TGGE and in vitro selection (or SELEX) (*28*, *30*).

Exhaustiveness of TGGE Selections. In the study presented here, a tetraloop library with 4096 possible sequences was made and stable sequences related to the cUNCGg motif were found. Isolation of known stable motifs confirms that this method is capable of isolating stable sequences. However, cGNRAg sequences, known to be similar in stability to cUNCGg sequences (*14*), were not found. Control experiments verified that these sequences, when embedded in the minihairpins, were similar in stability to the selected sequences. Moreover, other controls in which the GAAA was inserted into the top of the long hairpin used for the selections confirmed that under identical parallel TGGE selection conditions of 4 M urea, GAAA sequences have native gel electrophoretic mobility identical to that of YNMG sequences (data not shown). Thus, the absence of the GAAA sequences cannot be ascribed to unusual gel mobility or sensitivity to urea.

It appears that the loss of the GNRA sequences is due instead to problems with the PCR. Similar problems also occurred during selections of a DNA tetraloop library, and were corrected by redesign of the primer binding sites, which tolerated higher temperatures in PCR (M. Nakano, E. M. Moody, J. Liang, and P. C. Bevilacqua, manuscript submitted for publication). Interestingly, GNRA sequences are stable in both RNA and DNA forms, whereas the UNCG sequences are exceptionally stable only in RNA (*14*). Thus, the loss of GNRA sequences may arise from inefficient PCR. Although it is ultimately desirable that the TGGE-based selections be exhaustive, the subset of sequences selected in these studies was in fact unusually stable and their relatedness to each other allowed a family of sequences to be identified.

An Expanded Family of UNCG-like Tetraloops. The effectiveness of the selection in isolating stable sequences was initially apparent in that known stable sequences were recovered. The closing base pair in the selected sequences was also strongly conserved, with 47 of 48 sequences having a CG closing base pair. The contribution of the closing base pair to stability was evaluated by changing it to a GC pair, and large destabilizations of 0.9−1.9 kcal/mol in $\Delta\Delta G_{37}^{\circ}$ resulted (Table 2). The extra stability of the CG base pairs for tetraloops is unexpected on the basis of nearest-neighbor data, where the CG base pair is expected to be only ≈0.2 kcal/mol more favorable than a GC base pair for the stem that was used (*28*, *38*). A similar bias of stable loops toward CG closing base pairs has been observed for triloops and a less extensive set of tetraloops, with a $\Delta\Delta G_{37}^{\circ}$ of −2.1 to −2.3 kcal/mol relative to GC base pairs (*14*, *28*).

The main outcome of the study presented here was identification of a family of 16 hairpins with the cYNMGg motif. Similarity in the overall fold was provided by similarity in CD spectra (Figure 4, and Figure 1 of the Supporting Information). Hydrogen bonding between L_1 H2′ and G_{L4} O6 was supported by thermodynamic measurements on deoxy substitutions at loop position 1 (Table 3). A syn conformation for G_{L4} and C2′-endo sugar puckers for L_2 and L_3 were supported by DQF-COSY and NOESY experiments (Figure 7). A hydrogen bond from L_1 O2 to G_{L4} H1 and H2 was supported by the similarity of the imino proton spectra to that of UUCG (Figures 5 and 6) (*21*) and by thermodynamic measurements for substitutions at loop position 1 (Table 2). A hydrogen bond from the exocyclic amine of L_3 to the *pro-R* phosphoryl oxygen of N_{L2} was supported by the destabilizing effects of non-A, C substitutions at position 3 of the loop (Table 2). Last, an unusual conformation of the β torsion angle for the linkage between G8 and G9 was supported by the presence of an unusually downfield shifted hairpin ^{31}P resonance (*21*) (Figure 8).

Tertiary Interactions of YNMG Loops. In the Results, we described a tertiary interaction in 16S rRNA between cUACGg of loop 343 and gGAAAc of loop 159. Direct interaction between loop 343 and loop 159 involves the G of the CG closing base pair of a structurally unperturbed UACG at loop 343. Phylogenetic comparisons indicate that this base pair, which is critical for the thermodynamic stability of the YNMG loop (Table 2), is highly conserved (*55*). Thus, there appear to be at least two plausible reasons for conservation of the CG closing base pair for loop 343, tertiary interactions and loop stability.

While conservation of the closing base pair can be understood, it is less clear why the loop at position 343 is conserved as a YNMG, or why position 2 is conserved as an A. One possibility is that the shape of the YNMG loop is compatible with the close approach of A_{L2} and A_{L3} from loop 159. Examination of the GAAA−UACG interaction suggests that this is the case as loop 159 fits snugly into loop 343, with A_{L2} of loop 343 placed over the top of loop 159. Perhaps the fold of a non-YNMG loop would present a steric block to the close approach of loop 159. Shape compatibility of loops 159 and 343 suggests that the different YNMG loops which substitute at loop 343 have similar shapes. An additional possibility for YNMG conservation at loop 343 is that the enhanced stability of the YNMG loop may act to nucleate RNA folding in this region of 16S rRNA, minimize the population of alternative secondary structures, or both.

The A at position 2 of the YAMG as shown in the 30S crystal structure (*23*) is neither close enough nor at the proper angle to form direct hydrogen bonds with the residues in loop 159, and it is not positioned properly to stack with the bases in loop 159. However, A_{L2} of loop 343 may interact with loop 159 via water-mediated hydrogen bonds or by partial pyramidalization of A_{L2} N6 (*67*). N6 and N1 of A_{L2} of loop 343 are 2.9 and 3.9 Å from 2′-OH and O4′ of A_{L2} of loop 159, respectively. Phylogenetic conservation of A_{L2} of loop 343 suggests that the different YNMG sequences are able to present this A in a similar fashion, supporting similar folds for the different YNMG sequences.

Strikingly, when non-UNCG YNMG loops were found at loop 343, loop 159 was found to vary. In a preliminary survey of 35 eukarya (nuclear) 16S secondary structures, loop 343 was usually a cYAAGg sequence (33 of 35), but loop 159 varied widely. However, in each case (35 of 35), loop 159 was found to have two consecutive As, with the most common sequence being gUAAUUc (16 of 35). Conservation of consecutive As suggests that similar tertiary interactions may be occurring. Association of non-GAAA loops at position 159 with YAAG loops at position 343 suggests that YNAG loops are more plastic in their recognition capabilities than YNCG loops. It should be noted that there was no obvious maintenance of loop 343−loop 159 sequence conservation in mitochondria, and at times, no analogous loop 159 could be found, suggesting that this tertiary interaction may not exist in mitochondrial ribosomes (D. J. Proctor and P. C. Bevilacqua, unpublished observations).

One reason that the YNAG loops may be able to interact with a broader range of hairpin loops is that YNAG loops may be more dynamic than YNCG loops. Several results are consistent with enhanced flexibility of YNAG loops. First, the YNAG loops can tolerate 2′-deoxy substitutions at position 1 of the loop better than YNCG loops (Table 3). It is possible that YNAG loops with their enhanced motion can be deformed into a new shape with a less severe thermodynamic penalty than YNCG loops. Second, the temperature dependence of the imino proton spectra shows that YNAG loops broaden considerably between 5 and 35 °C, whereas YNCG loops do not (Figure 2 of the Supporting Information). The chemical exchange of hydrogen-bonded imino protons occurs on the millisecond time scale (*68*). Thus, one possibility is that YNAG loops have increased motion on the millisecond or faster time scales. Third, the CD spectrum for UAAG shows an unusually large positive displacement at 260 nm and negative displacement near 240 nm (Figure 1C of the Supporting Information). The difference CD spectrum between UAAG and other UNAG sequences looks remarkably similar to the CD spectrum of ribodinucleoside phosphates involving A steps in the loop, including UpA, ApA, and ApG (*69*). This observation suggests that A_{L2} might be able to partially stack on one of the other bases in the loop. The ability of position 2 of the loop to stack would be consistent with enhanced flexibility of the YNAG loop.

Increased motion of YNAG loops relative to YNCG loops may occur because A_{L3} with its additional ring is not as readily accommodated in the loop structure as C_{L3}. Since melting of the stem and loop may be decoupled for YNAG sequences, the thermodynamic parameters reported in Table 2, which are based on a two-state model, should be viewed as qualitative. It is worth noting that UNCG loops do not undergo a significant change in conformation upon forming the tertiary interaction with loop 159 (see the Results), suggesting that their extraordinary stability may provide a large barrier to deformation.

The structural, dynamic, and thermodynamic effects of changing the closing base pair of the UUCG loop from CG to GC have been studied by Williams and Hall (*15*, *25*−*27*). There are strong parallels between the behavior of gUUCGc and cYNAGg loops. Identical structures were found for cUUCGg and gUUCGc loops at low temperatures (*15*); 2′-OH substitutions were tolerated better for the GC than the CG closing base pair (*27*), and the imino proton for G_{L4} of gUUCGc broadened considerably between 4 and 25 °C (*15*). In addition, a survey of the closing base pair frequencies for variable regions revealed that GC closing base pairs are much more common for eukaryotes than bacteria (*12*, *15*), similar to YNAG loops (*13*). Results on the gUUCGc loops were interpreted as being consistent with enhanced motion at temperatures well below their T_M values (*15*). The authors suggested that this increased flexibility may facilitate protein binding by lowering the energetic penalty for changing the structure of the loop. The enhanced flexibility of YNAG loops may also favor interactions with proteins or with diverse RNA elements. It should prove interesting to see if more RNA tertiary interactions are composed of preformed secondary structural elements having enhanced flexibility.

ACKNOWLEDGMENT

We thank the group of Professor Bob Matthews for assistance with CD experiments and Erica Cho and Dr. Zhanyong Shu for help with the phylogenetic database. We also thank Professors Juliette Lecomte and John SantaLucia

for helpful discussions and members of the Bevilacqua laboratory for comments on the manuscript.

SUPPORTING INFORMATION AVAILABLE

CD spectra for all 24 minihairpins listed in Table 2 and temperature dependencies of 1D imino proton spectra for UAAG, CACG, and UAAG. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 920−930.
2. Valadkhan, S., and Manley, J. L. (2001) *Nature 413*, 701−707.
3. Brion, P., and Westhof, E. (1997) *Annu. Rev. Biophys. Biomol. Struct. 26*, 113−137.
4. Tinoco, I., Jr., and Bustamante, C. (1999) *J. Mol. Biol. 293*, 271−281.
5. Costa, M., and Michel, F. (1995) *EMBO J. 14*, 1276−1285.
6. Costa, M., and Michel, F. (1997) *EMBO J. 16*, 3289−3302.
7. Bevilacqua, P. C., Kierzek, R., Johnson, K. A., and Turner, D. H. (1992) *Science 258*, 1355−1358.
8. Herschlag, D. (1992) *Biochemistry 31*, 1386−1399.
9. Varani, G. (1995) *Annu. Rev. Biophys. Biophys. Chem. 24*, 379−404.
10. 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.
11. Woese, C. R., Winker, S., and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467−8471.
12. Wolters, J. (1992) *Nucleic Acids Res. 20*, 1843−1850.
13. Konings, D. A., and Gutell, R. R. (1995) *RNA 1*, 559−574.
14. Antao, V. P., Lai, S. Y., and Tinoco, I., Jr. (1991) *Nucleic Acids Res. 19*, 5901−5905.
15. Williams, D. J., and Hall, K. B. (2000) *J. Mol. Biol. 297*, 1045−1061.
16. Heus, H. A., and Pardi, A. (1991) *Science 253*, 191−194.
17. 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.
18. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−980.
19. Jucker, F. M., and Pardi, A. (1995) *Biochemistry 34*, 14416−14427.
20. Cheong, C., Varani, G., and Tinoco, I., Jr. (1990) *Nature 346*, 680−682.
21. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
22. Allain, F. H., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.
23. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature 407*, 327−339.
24. Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S., and Dumas, P. (2000) *J. Mol. Biol. 304*, 35−42.
25. Williams, D. J., and Hall, K. B. (1999) *Biophys. J. 76*, 3192−3205.
26. Williams, D. J., and Hall, K. B. (2000) *J. Mol. Biol. 297*, 251−265.
27. Williams, D. J., Boots, J. L., and Hall, K. B. (2001) *RNA 7*, 44−53.
28. Shu, Z., and Bevilacqua, P. C. (1999) *Biochemistry 38*, 15369−15379.
29. Juneau, K., and Cech, T. R. (1999) *RNA 5*, 1119−1129.
30. Bevilacqua, J. M., and Bevilacqua, P. C. (1998) *Biochemistry 37*, 15877−15884.
31. Kim, C. H., and Tinoco, I., Jr. (2001) *J. Mol. Biol. 307*, 827−839.
32. Kim, C. H., and Kao, C. C. (2001) *RNA 7*, 1476−1485.
33. Larsson, S. L., and Nygard, O. (2001) *Biochemistry 40*, 3222−3231.
34. Lee, S. G., Kim, D. Y., Hyun, B. H., and Bae, Y. S. (2002) *J. Virol. 76*, 1649−1662.
35. Antao, V. P., and Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 819−824.
36. Wartell, R. M., Hosseini, S. H., and Moran, C. P., Jr. (1990) *Nucleic Acids Res. 18*, 2699−2705.
37. Serra, M. J., and Turner, D. H. (1995) *Methods Enzymol. 259*, 242−261.
38. 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.
39. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
40. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
41. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) pp 597, CRC Press, Cleveland, OH.
42. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) pp 597, CRC Press, Cleveland, OH.
43. Cantor, C. R., and Schimmel, P. R. (1980) in *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure and Function*, W. H. Freeman, San Francisco.
44. Kierzek, R., Li, Y., Turner, D. H., and Bevilacqua, P. C. (1993) *J. Am. Chem. Soc. 115*, 4985−4992.
45. Puglisi, J. D., and Wyatt, J. R. (1995) *Methods Enzymol. 261*, 323−350.
46. Schowen, K. B., and Schowen, R. L. (1982) *Methods Enzymol. 87*, 551−606.
47. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
48. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
49. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) *J. Magn. Reson., Ser. A 102*, 241−245.
50. Shaka, A. J., Keeler, J., and Freeman, R. (1983) *J. Magn. Reson. 53*, 313−340.
51. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
52. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
53. Pappalardo, L., Kerwood, D. J., Pelczer, I., and Borer, P. N. (1998) *J. Mol. Biol. 282*, 801−818.
54. Kirchner, R., Vogtherr, M., Limmer, S., and Sprinzl, M. (1998) *Antisense Nucleic Acid Drug Dev. 8*, 507−516.
55. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müeller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) *BMC Bioinformatics 3*, 2.
56. Tunis-Schneider, M. J., and Maestre, M. F. (1970) *J. Mol. Biol. 52*, 521−541.
57. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1990) *Biochemistry 29*, 8813−8819.
58. Wijmenga, S. S., Heus, H. A., Werten, B., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1994) *J. Magn. Reson., Ser. B 103*, 134−141.
59. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) *Biochemistry 38*, 14214−14223.
60. Schroeder, S. J., and Turner, D. H. (2000) *Biochemistry 39*, 9257−9274.
61. Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M., and Uesugi, S. (1990) *Nucleic Acids Res. 18*, 3831−3839.
62. Molinaro, M., and Tinoco, I., Jr. (1995) *Nucleic Acids Res. 23*, 3056−3063.
63. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) *Science 292*, 883−896.
64. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 4899−4903.
65. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.
66. Zuker, M., Jaeger, J. A., and Turner, D. H. (1991) *Nucleic Acids Res. 19*, 2707−2714.
67. Sponer, J., Leszczynski, J., and Hobza, P. (2001) *Biopolymers 61*, 3−31.
68. Gueron, M., Kochoyan, M., and Leroy, J. L. (1987) *Nature 328*, 89−92.
69. Warshaw, M. M., and Cantor, C. R. (1970) *Biopolymers 9*, 1079−1103.
70. Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.
71. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., and Turner, D. (1985) *Biochemistry 24*, 4533−4539.

BI026201S