Non-Nearest-Neighbor Dependence of the Stability for RNA Bulge Loops Based on the Complete Set of Group I Single-Nucleotide Bulge Loops†

Joshua M. Blose, Michelle L. Manni, Kelly A. Klapec, Yukiko Stranger-Jones, Allison C. Zyra, Vasiliy Sim, Chad A. Griffith, Jason D. Long, and Martin J. Serra*

Department of Chemistry, Allegheny College, 520 North Main Street, Meadville, Pennsylvania 16335

Received April 18, 2007; Revised Manuscript Received October 4, 2007

ABSTRACT: Fifty-nine RNA duplexes containing single-nucleotide bulge loops were optically melted in 1 M NaCl, and the thermodynamic parameters $\Delta H°$, $\Delta S°$, $\Delta G°_{37}$, and T_M for each sequence were determined. Sequences from this study were combined with sequences from previous studies [Longfellow, C. E., et al. (1990) *Biochemistry 29*, 278−285; Znosko, B. M., et al. (2002) *Biochemistry 41*, 10406−10417], thus examining all possible group I single-nucleotide bulge loop and nearest-neighbor sequence combinations. The free energy increments at 37 °C for the introduction of a group I single-nucleotide bulge loop range between 1.3 and 5.2 kcal/mol. The combined data were used to develop a model for predicting the free energy of a RNA duplex containing a single-nucleotide bulge. For bulge loops with adjacent Watson−Crick base pairs, neither the identity of the bulge nor the nearest-neighbor base pairs had an effect on the influence of the bulge loop on duplex stability. The proposed model for prediction of the stability of a duplex containing a bulged nucleotide was primarily affected by non-nearest-neighbor interactions. The destabilization of the duplex by the bulge was related to the stability of the stems adjacent to the bulge. Specifically, there was a direct correlation between the destabilization of the duplex and the stability of the less stable duplex stem. The stability of a duplex containing a bulged nucleotide adjacent to a wobble base pair also was primarily affected by non-nearest-neighbor interactions. Again, there was a direct correlation between the destabilization of the duplex and the stability of the less stable duplex stem. However, when one or both of the bulge nearest neighbors was a wobble base pair, the free energy increment for insertion of a bulge loop is dependent upon the position and orientation of the wobble base pair relative the bulged nucleotide. Bulge sequences of the type $\binom{5'\text{UBX}}{3'\text{G Y}}$, $\binom{5'\text{GBG}}{3'\text{U U}}$, and $\binom{5'\text{UBU}}{3'\text{G G}}$ are less destabilizing by 0.6 kcal/mol, and bulge sequences of the type $\binom{5'\text{GBX}}{3'\text{U Y}}$ and $\binom{5'\text{XBU}}{3'\text{Y G}}$ are more destabilizing by 0.4 kcal/mol than bulge loops adjacent to Watson−Crick base pairs.

RNA fulfills essential cellular roles, including storage of information, protein and small molecule binding, and chemical catalysis (*1*−*10*). The functional diversity of RNA is often predicated by hierarchical folding of complex tertiary structures with secondary structure formation preceding that of the native, functional fold (*11*, *12*). Since the tertiary structure arises from the preformed secondary structure, accurately determining the secondary structure of a RNA molecule would provide clues to its function as well as facilitate prediction of the RNA tertiary fold. Current structure prediction algorithms using updated thermodynamic parameters correctly predict ∼73% of known base pairs (*13*, *14*). Therefore, it follows that refinement of the input thermodynamic parameters for multiple secondary structure motifs would improve predictions of secondary structures. One secondary structure motif for which there remain few thermodynamic analyses is the single-nucleotide bulge (*15*, *16*). A single-nucleotide bulge occurs when an unpaired nucleotide disrupts regions of contiguous base pairing. These unpaired nucleotides can participate in a variety of biological events, including protein binding and tertiary structure formation (*17*−*20*). More recently, single-nucleotide bulges have been implicated in reverse transcriptase-mediated RNA displacement synthesis as single-nucleotide bulges enhance synthesis through stable secondary structures (*22*).

Previously, single-nucleotide bulges were subdivided into groups according to the identity of the bulge and adjacent base pairs (*16*). Group I single-nucleotide bulge loops were defined as bulge loops where the position of the bulge was unambiguous with a bulged nucleotide that is not identical to either of the neighboring nucleotides, and group II sequences were defined as bulge loops where the position of the bulge is ambiguous since the bulge nucleotide is identical to at least one of the neighboring nucleotides (Table 1). With a limited amount of data, a model was presented that provided the parameters for predicting the thermodynamics of single-nucleotide bulges for group I and group II bulges (*16*). The model for group I single-nucleotide bulges, however, examined only two bulges with nearest-neighbor wobble pairs. In considering the single-nucleotide bulge loops with wobble base pairs, it became obvious that additional

† This work was supported by the Camille and Henry Dreyfus Foundation, National Science Foundation Grant MCB-0340958, and National Institutes of Health Grant RGM-068426.

* To whom correspondence should be addressed. Phone: (814) 332-5356. Fax: (814) 332-2789. E-mail: mserra@allegheny.edu.

10.1021/bi700736f CCC: $37.00
Published on Web 11/30/2007

Table 1: Examples of Different Classes of Single-Nucleotide Bulge Loops

sequence	potential bulge positions		duplex sequences ($-\Delta G^\circ_{37}$)	
		Group I		
CGC**A**GCC GCG CGG	CGC**A**GCC GCG CGG		CGCGCC (10.7) GCGCGG	
		Group II		
CG**C**GCC GCG CGG	CGC**C**GCC GCG CGG	CG**C**CGCC GC GCGG	CGCGCC (10.7) GCGCGG	CGCGCC (10.7) GCGCGG
		Group III		
CG**CU**GCC GCG CGG	CGC**U**GCC GCG CGG	CG**C**UGCC GC GCGG	CGCGCC (10.7) GCGCGG	CGUGCC (8.9) GCGCGG
		Group IV		
CG**CCU**GG GCG GCC	CGC**C**UGG GCG GCC CGCC**U**GG GCGG CC	CG**C**CUGG GC GGCC	CGCUGG (8.5) GCGGCC CGCCGG (10.6) GCGGCC	CGCUGG (8.5) GCGGCC

types of bulge ambiguity can arise. For example, with the oligomer 5′CG**CU**GCC/3′GC**G** CGG, the bold G on the bottom strand can form a base pair with either the bold C or the bold U. Depending upon which pair forms, either the C or the U would be the bulged nucleotide and the parent duplex would contain either a Watson−Crick or wobble base pair having different thermodynamic stabilities (Table 1). Sequences with this type of ambiguity are now defined as group III bulge loops. Group IV single-nucleotide bulge loops are defined as those sequences which have characteristics of both group II and group III sequences (Table 1). Some of the sequences originally (*16*) considered group I or II have now been reassigned as group III or IV.

Group I single-nucleotide bulge loops are now defined as bulges with either Watson−Crick or wobble base pair nearest neighbors, and no bulge ambiguity is possible. There are 26 possible group I bulges with only Watson−Crick nearest-neighbor base pairs (Table 3) and 24 possible group I bulges when one or more of the nearest neighbors is a wobble base pair (Table 5). Group II single-nucleotide bulge loops are defined as bulges with either Watson−Crick or wobble nearest neighbors where the bulge is identical to one or more of the adjacent bases. While there are a large number of group II bugles, if more than two identical bases occur; for the case where there are only two identical bases, as in the example in Table 1, there are 68 possible sequence combinations. The possible sequence combinations are given in Table S1 of the Supporting Information. Group III single-nucleotide bulge loops are defined as bulges with either an adjacent AG (GA) or CU (UC) and the possibility of forming either a Watson−Crick or wobble base pair. There are 48 possible group III sequence combinations; they are listed in the Supporting Information (Table S2). There are a large number of group IV single-nucleotide bulge loops that have characteristics of both group II and group III single-nucleotide bulge loops.

In this study, the complete set of group I single-nucleotide bulge loops with either Watson−Crick or wobble base pair nearest neighbors has been thermodynamically characterized to improve our ability to predict the stability of RNA duplexes with bulge loops. The free energy increment for the insertion of a bulge loop into a duplex was found to be primarily influenced by non-nearest-neighbor interactions. The stability of the stem adjacent to the bulge has a direct effect on the duplex destabilization caused by the insertion of a bulge loop. The free energy increment for group I bulges adjacent to Watson−Crick nearest neighbors is not dependent on nearest-neighbor interactions. For bulges adjacent to wobble base pair nearest neighbors, nearest-neighbor interactions influence the free energy increment for insertion of the bulge. The enthalpic contribution for insertion of a group I single-nucleotide bulge loop is independent of both nearest-neighbor and non-nearest-neighbor interactions.

MATERIALS AND METHODS

RNA Synthesis and Purification. Most oligomers were synthesized on CPG solid supports (Applied Biosystems 392 DNA/RNA synthesizer) utilizing phosphoramidites with the 2′-hydroxyl protected as the *tert*-butyldimethylsilyl ether from Glen Reseach (Sterling, VA) (*23*, *24*). Oligomers underwent ammonia and fluoride deprotection, and crude sample was purified using preparative TLC (55:35:10 *n*-propanol/ammonium hydroxide/water) and Sep-Pak C18 (Waters) chromatography. Some oligomers were ordered from Dharmacon, and deprotection of the oligomers was carried out using the manufacturer's instructions. The oligomers were then purified as described above. Sample purity was determined through analytical TLC or HPLC (C-8) and was greater than 95%.

Melting Curve and Data Analysis. For non-self-complementary sequences, individual strand concentrations were calculated from the high-temperature single-strand absorbance at 260 and 280 nm using nearest-neighbor extinction coefficients (*25*, *26*). Single strands were then annealed in a 1:1 molar ratio. Optical melting experiments were performed using a Beckman DU 640 spectrophotometer and a high-performance temperature controller at 260 or 280 nm. Absorbance changes for oligomers in 1 M NaCl melt buffer [1 M NaCl, 0.01 M cacodylic acid, and 0.5 mM EDTA (pH 7.0)] were recorded as a function of temperature from 95 to 10 °C at a rate of 1 °C/min as described previously (*27*). The experiment was repeated at 10 varying sample concentrations to give an at least 50-fold concentration range (from 10 μM to 1 mM) for each sample. Absorbance versus temperature profiles were fit to a two-state model with sloping baselines using a nonlinear least-squares program (*28*). Thermodynamic parameters for duplex formation were obtained by two methods: (1) enthalpy and entropy changes from the fits of the individual melting curves were averaged, and (2) plots of the reciprocal melting temperature, T_M^{-1}, versus log $C_t/4$ gave enthalpy and entropy changes (*29*):

$$T_M^{-1} = (2.3R/\Delta H^\circ) \log C_t/4 + \Delta S^\circ/\Delta H^\circ \qquad (1)$$

where C_t is the total concentration of oligomer. Parameters derived from the two methods generally agreed within 15%, consistent with the two-state model (*30*, *31*). The Gibbs free energy change at 37 °C was calculated as

$$\Delta G^\circ_{37} = \Delta H^\circ - (310.15\ \mathrm{K})\Delta S^\circ \qquad (2)$$

Determination of the Contribution of Bulge Loops to Duplex Thermodynamics. The free energy of duplex formation can be approximated by the nearest-neighbor model (*32*). The free energy contribution of each bulged nucleotide was calculated from the experimental data and the nearest-neighbor model according to eq 3

$$\Delta G^\circ_{37(\text{bulge})} = \Delta G^\circ_{37(\text{measured})} - \Delta G^\circ_{37(\text{duplex})} \quad (3)$$

where $\Delta G^\circ_{37(\text{measured})}$ is the experimentally determined value from the melts and $\Delta G^\circ_{37(\text{duplex})}$ is calculated from the nearest-neighbor model for the duplex as if it did not contain the bulge. ΔH° values are calculated in a similar fashion.

Phylogenetic Analysis. A database of phylogenetically determined RNA secondary structures (*33*) of 305 SSU rRNAs, 169 LSU rRNAs, 16 group I intron RNAs, and 7 group II intron RNAs was searched for group I single-nucleotide bulge loops. The loops were characterized by nearest-neighbor base pairs and bulge identity.

Statistical Analysis. Statistical analysis of the data was carried out using the statistical software available with GraphPad Prism and GraphPad Instat.

RESULTS

From the study of a set of 23 oligomers containing a single-nucleotide bulge loop with adjacent Watson−Crick base pairs, we have previously determined a model for predicting the parameters for the thermodynamic stability of group I bulge loops (*16*). To more fully determine the role of the bulge identity and the nearest neighbors on the stability of bulge loops, 23 additional (12 purine and 11 pyrimidine bulges) duplexes containing a group I single-nucleotide bulge loop with Watson−Crick nearest-neighbor base pairs were prepared and the thermodynamics of duplex formation measured by optical melting. These oligomers when combined with the previously measured oligomers give measurements for the complete set of group I single-nucleotide bulge loops with Watson−Crick nearest neighbors.

Thermodynamic parameters for duplex formation by these oligonucleotides are listed in Table 2. The oligonucleotides are listed in order of decreasing free energy for the respective duplex. Sequences were divided into two groups: duplexes with purine bulges and duplexes with pyrimidine bulges. Residues in bold are the bulge nucleotides. The parameters are the average values derived from fits of the melt curves and from T_M^{-1} versus $\log(C_t/4)$ plots. The parameters from the two methods agree within 15% for all duplexes in Table 2, suggesting that the two-state model is a reasonable approximation for these transitions (*30*, *31*). The average deviations in thermodynamic parameter values are 7.3, 7.9, and 1.5% for ΔH°, ΔS°, and ΔG°_{37}, respectively.

The free energy and enthalpy contributions of each bulged nucleotide were calculated from the experimental data according to eq 3 and presented in Table 3. The complete set of all single-nucleotide bulge loop duplexes used to provide the results presented in Table 3 is listed in the Supporting Information (Table S3) along with the thermodynamic values used to determine the values presented. As previously observed (*16*), all of bulges destabilize the duplex.

In our initial analysis, we focused on nearest-neighbor effects. The extent of destabilization ranges between 1.3 and 5.2 kcal/mol. Also, as noted previously (*16*), on average, the pyrimidine single bulges are slightly more stable than the purine bulges by 0.3 kcal/mol. The destabilization caused by the introduction of a single group I pyrimidine bulge into a duplex is 3.9 ± 0.8 kcal/mol and for a purine bulge is 4.2 ± 0.8 kcal/mol. Since these values are not statistically different, they can be combined to provide a simple model for predicting the stability of a duplex containing a single group I bulge nucleotide (purine or pyrimidine) which is just the average of all of the measured values (4.0 ± 0.8 kcal/mol). The bold values in Table 3 represent bulge loops which appear to be less destabilizing. These values led to an understanding of the importance of non-nearest-neighbor interactions and will be discussed more fully below (see the Discussion).

The enthalpic increments for single-nucleotide bulge loops are also presented in Tables 3. In all cases except three, the introduction of a bulge lowers the enthalpy of formation of the duplex. As with the free energy, the enthalpy contribution to duplex formation is nearly identical for pyrimidine and purine bulge nucleotides. Therefore, the enthalpy contribution for group I single-nucleotide bulge loops is just the average of the measured values (13.4 ± 9.4 kcal/mol). As noted previously, the standard error for enthalpy measurements is much larger than for free energy measurements (*34*).

Previously, only two group I single-nucleotide bulge loops adjacent to a wobble base pair had been measured (*16*). To determine the influence of single-nucleotide bulge loops adjacent to a GU base pair, 36 additional oligonucleotide duplexes (21 purine and 15 pyrimidine bulges) were prepared and the thermodynamics of duplex formation measured by optical melting. These oligomers when combined with the previously measured oligomers give measurements for the complete set of group I single-nucleotide bulge loops with wobble nearest-neighbor base pairs.

Thermodynamic parameters for duplex formation for oligonucleotide duplexes containing a group I single-nucleotide bulge loop adjacent to a wobble base pair are listed in Table 4. The oligonucleotides are listed in order of decreasing free energy for the respective duplex. Sequences were divided into groups based upon the position and orientation of the wobble base pair relative to the position of the bulged nucleotide. Bulged nucleotides are shown in bold. The parameters are the average values derived from fits of the melt curves and from T_M^{-1} versus $\log(C_t/4)$ plots. Again, the parameters from the two methods agree within 15% for all duplexes in Table 4, suggesting that the two-state model is a reasonable approximation for these transitions. The average deviations in thermodynamic parameter values are 8.8, 10.3, and 2.3% for ΔH°, ΔS°, and ΔG°_{37}, respectively.

The free energy contribution of each bulged nucleotide was calculated from the experimental data according to eq 3 and presented in Table 6. As observed for the bulges with Watson−Crick base pairs, all of bulges with wobble nearest-neighbor base pairs also destabilize the duplex. The extent of destabilization ranges between 0.5 and 5.8 kcal/mol. Given the lack of influence of the Watson−Crick nearest-neighbor base pairs to influence the thermodynamic contribution of the bulge loop to duplex formation, it seems reasonable to treat a bulge adjacent to a wobble base pair in a similar fashion.

In our initial analysis, we again focused on nearest-neighbor effects. The influence of a group I single-nucleotide bulge loop adjacent to a wobble base pair depends upon both

Table 2: Thermodynamic Parameters for Duplex Formation for Oligonucleotides Containing a Group I Single-Nucleotide Bulge Adjacent to Watson−Crick Base Pairs[a]

	T_M^{-1} vs log C_T plots				average of curve fits			
oligomers[b]	-ΔH° (kcal/mol)	-ΔS° (eu)	-ΔG°$_{37}$ (kcal/mol)	T_M[c] (°C)	-ΔH° (kcal/mol)	-ΔS° (eu)	-ΔG°$_{37}$ (kcal/mol)	T_M[c] (°C)
Group I Purine		T_M^{-1} vs log C_T plots				average of curve fits		
GGCG**A**CUCG CCGC GAGC	60.2	162.3	9.9	55.4	60.5	163.0	10.0	55.8
GAGC**A**GGUC CUCG CCAG	65.2	180.3	9.3	50.8	69.5	193.2	9.6	51.2
GGCG**A**UUCC CCGC AAGG	66.6	185.0	9.2	49.9	70.9	198.6	9.3	49.6
GGCG**A**UUCG CCGC AAGC	67.6	191.3	8.3	45.4	66.0	186.0	8.3	45.6
GACC**G**UAGC CUGG AUCG	54.5	149.9	8.0	45.4	56.1	155.2	8.0	45.4
GUGC**A**UGAG CACG ACUC	64.8	183.5	7.9	43.7	65.3	184.8	8.0	43.9
CAGU**G**CAGC GUCA GUCG	48.5	133.1	7.3	41.8	52.3	144.7	7.4	42.5
CAGU**A**GAGC GUCA CUCG	59.1	167.8	7.1	39.9	63.5	181.7	7.1	39.8
CAGU**A**CAGC GUCA GUCG	52.8	148.0	6.9	39.4	49.2	136.4	6.9	39.6
GUGU**G**UGUG CACA ACAC	54.6	155.2	6.4	36.4	58.4	167.1	6.5	37.0
GAG**A**CAC CUC GUG	45.0	127.9	5.3	28.6	50.7	146.8	5.2	29.1
GAU**A**GAC CUA CUG	37.4	108.5	3.8	15.7	34.8	99.1	4.1	16.6
Group I Pyrimidine								
GGCA**U**GACG CCGU CUGC	53.7	145.3	8.7	49.9	60.4	166.5	8.8	49.0
GACA**U**GUGC CUGU CACG	67.7	181.3	8.5	46.7	57.9	159.7	8.4	47.1
CAGG**U**AAGC GUCC UUCG	72.6	207.0	8.4	45.1	78.5	225.7	8.5	45.0
CACA**U**GCAC GUGU CGUG	61.3	171.2	8.2	45.6	64.8	182.4	8.3	45.6
CACA**C**GCAC GUGU CGUG	47.8	127.9	8.1	47.6	58.8	162.9	8.3	46.8
CAGG**C**AAGC GUCC UUCG	62.3	175.9	7.8	43.3	69.5	198.4	7.9	43.3
GACG**C**UAGC CUGC AUCG	37.8	97.6	7.6	45.8	35.9	91.1	7.6	46.5
CAGU**C**GAGC GUCA CUCG	56.7	158.8	7.5	42.4	57.0	159.4	7.6	43.0
GCAU**C**UGUG CGUA ACAC	65.7	192.6	6.0	34.4	66.2	194.1	6.0	34.7
CAGA**C**UAGC GUCU AUCG	43.6	122.6	5.6	30.4	35.9	97.6	5.7	29.6
GUG**C**UUC CAC AAG	44.5	129.1	4.4	23.0	44.9	130.0	4.6	24.0

[a] Solutions contained 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM EDTA (pH 7). [b] The nucleotide in bold is the bulge residue. The top sequence is written 5′ → 3′. [c] Calculated at an oligomer concentration of 10^{-4} M.

Table 3: Thermodynamic Parameters and Natural Occurrence for Single-Nucleotide Group I Bulges Closed by Watson−Crick Base Pairs

bulge sequence[a]	$\Delta G°_{37bulge}$[b] (kcal/mol)	$\Delta H°_{bulge}$[b] (kcal/mol)	percentage of naturally occurring group I sequences[e]	bulge sequence[a]	$\Delta G°_{37bulge}$[b] (kcal/mol)	$\Delta H°_{bulge}$[b] (kcal/mol)	percentage of naturally occurring group I sequences[e]
	A Bulges				C Bulges		
CAC	3.9[d]	4.7[d]		UCA			
G G	4.3[c]	13.1[c]	12.8	A U	3.9[c]	8.5[c]	1.8
GAG	3.8[d]	8.3[d]		GCG			
C C	3.7[c]	9.4[c]	11.4	C C	3.0[c]	−4.5[c]	1.5
CAG	4.5	14.7		UCU	3.9	4.2	
G C	4.4[c]	10.0[c]	10.6	A A	2.9[c]	19.7[c]	1.0
CAU	4.4	11.9		ACA			
G A	4.4[c]	13.9[c]	10.4	U U	4.3[c]	−3.7[c]	0.9
UAC	5.4	26.0		ACG	4.0	22.6	
A G	4.7[c]	8.5[c]	9.9	U C	4.1[c]	12.4[c]	0.7
GAC	2.9	8.2		UCG	4.8	20.1	
C G	5.2	23.4	8.2	A C	4.1[c]	8.7[c]	0.7
GAU	4.5	9.8		ACU			
C A	4.5	10.6	2.8	U A	4.8	32.4	0.4
UAU				GCA			
A A	4.0[c]	20.8[c]	1.5	C U	4.3	9.4	0.4
UAG	**2.2**	**16.4**		GCU	**1.3**	**4.2**	
A C	5.2	15.6	1.2	C A	4.3	38.5	0.3
	G Bulges				U Bulges		
CGC				GUG	3.9[d]	13.0[d]	
G G	4.0[c]	8.8[c]	1.9	C C	3.3[c]	2.3[c]	12.3
UGC				AUG	4.0, 4.0	15.1, 12.8	
A G	5.0	26.6	1.1	U C	5.0	23.0	3.2
CGU				AUA			
G A	4.7	21.8	0.3	U U	3.7[c]	3.7[c]	3.0
UGU	3.4	12.2		GUA			
A A	2.9[c]	28.8[c]	0.2	C U	3.7	−0.3	1.6

[a] The top strand is written 5′ → 3′. [b] Values calculated as described in the text. [c] From ref *16*. [d] From ref *15*. [e] The total number of single-nucleotide group I bulge sequences closed by Watson−Crick base pairs is 3520.

the position of the wobble pair relative to the bulge (5′ or 3′) and the orientation of the wobble (G or U in the strand with the bulged nucleotide). For bulges of the type $\binom{5'GBX}{3'U\ \ Y}$ and $\binom{5'XBU}{3'Y\ \ G}$, where XY is a Watson−Crick base pair, the average destabilization caused by the bulge is not significantly different from that due to bulges adjacent to Watson−Crick base pairs, 3.2 ± 1.4 and 3.8 ± 1.7 kcal/mol, respectively. For bulges of the type $\binom{5'UBX}{3'G\ \ Y}$ and bulges with two wobble nearest neighbors, the bulge is significantly less destabilizing than a bulge adjacent to Watson−Crick base pairs. The average free energy increments for $\binom{5'UBX}{3'G\ \ Y}$ and bulges with two wobble nearest neighbors are 3.2 ± 0.7 and 2.8 ± 0.6 kcal/mol, respectively. For bulges of the type $\binom{5'XBG}{3'Y\ \ U}$, the bulge is significantly more destabilizing than a bulge adjacent to Watson−Crick base pairs. The average free energy increment is 4.5 ± 0.3 kcal/mol. The six bold values in Table 5 represent bulge loops which are less destabilizing. These values will be discussed more fully below in the context of non-nearest-neighbor interactions (see the Discussion).

The enthalpic increments for single-nucleotide bulge loops adjacent to wobble base pairs are also presented in Tables 5. In all cases except one, the introduction of a bulge lowers the enthalpy of formation of the duplex. The enthalpy increments for insertion of a bulge loop into a duplex do not vary depending upon the position or orientation of the wobble base pair relative to the bulge. Therefore, the enthalpy contribution for group I single-nucleotide bulge loops adjacent to a wobble base pair is just the average of the measured values (18.0 ± 13.6 kcal/mol).

DISCUSSION

There has been a recent explosion in nucleic acid sequence information derived from the various genome and other sequencing projects. This information provided a wealth of information about both coding and noncoding RNAs. To make full use of this information, particularly for noncoding RNAs, it is necessary to understand how the RNA sequence is related to the functionally relevant three-dimensional structure of the molecule. For this reason, there has been an increased level of interest in predicing both the secondary and tertiary structure of RNA from sequence. The most commonly used programs for secondary structure prediction, mfold and RNAstructure (*37*), use thermodynamic parameters to predict the most stable (and suboptimal) secondary structural folds. The ability of these programs to predict the secondary structure depends upon the availability of accurate models to provide the thermodynamic parameters for the various RNA secondary structural motifs. Bugle loops are a common motif, yet relatively few studies have investigated the influence of sequence on the thermodynamics of secondary structure formation (*15*, *16*).

In an earlier investigation, using a limited number of bulge loop nucleotides, we arrived at a model for predicting the parameters for the influence of bulge loops on the stability of RNA duplexes based upon the type of loop sequence (group I or II) and the identity of the bulge nucleotide when the nearest neighbors were Watson−Crick base pairs. As we

Table 4: Thermodynamic Parameters for Duplex Formation for Oligonucleotides Containing a Group I Single-Nucleotide Bulge Adjacent to GU Wobble Base Pairs[a]

	T_M^{-1} vs log C_T plots				average of curve fits			
oligomers[b]	-ΔH° (kcal/mol)	-ΔS° (eu)	-ΔG°$_{37}$ (kcal/mol)	T_M[c] (°C)	-ΔH° (kcal/mol)	-ΔS° (eu)	-ΔG°$_{37}$ (kcal/mol)	T_M[c] (°C)
purine bulges with 5′ adjacent G-U pairs								
CGGU**G**CACG GCCG GUGC	52.7	139.1	9.6	55.8	51.2	134.1	9.6	57.0
CGGU**A**CACG GCCG GUGC	55.3	148.3	9.3	53.2	53.8	143.7	9.3	53.6
CAGU**A**GACG GUCG CUGC	56.4	158.0	7.4	41.7	68.5	197.2	7.4	40.8
GCUU**A**CGAC CGAG GCUG	79.2	231.6	7.4	40.3	75.2	218.8	7.3	40.3
GCUU**A**UUGG CGAG AACC	43.2	119.2	6.3	35.0	52.3	149.0	6.1	34.7
GUCU**G**UCAC CAGG AGUG	58.6	169.1	6.1	34.9	64.2	187.2	6.1	35.0
GCUU**G**UACC CGAG AUGG	53.5	153.6	5.8	33.0	59.5	173.0	5.9	33.5
GCU**A**GAC CGG CUG	49.9	142.1	5.8	32.7	55.6	160.5	5.8	32.9
purine bulges with 3′ adjacent G-U pairs								
GAC**A**UAGUC CUG GUCAG	73.6	207.6	9.2	48.8	63.1	174.4	9.0	49.4
CGGC**A**UACG GCCG GUGC	40.2	103.2	8.2	50.5	36.3	90.0	8.4	53.6
GUCG**A**UCAC CAGC GGUG	61.4	175.5	6.9	39.1	59.6	169.7	6.9	39.1
CAAG**A**UAGUC GUUC GUCAG	76.9	226.4	6.7	37.7	66.1	191.3	6.8	38.1
GGC**G**UAC CCG GUG	34.0	89.6	6.2	34.0	37.0	99.5	6.2	34.0
CAGU**G**UAGUC GUCA GUCAG	28.0	72.0	5.7	27.8	25.1	62.0	5.9	29.5
GCUU**G**UACC CGAA GUGG	57.2	167.1	5.4	31.0	56.1	162.9	5.6	31.9
CAGU**A**UAGUC GUCA GUCAG	49.3	141.8	5.3	29.6	42.7	119.7	5.6	30.4
GCG**A**UAG CGC GUC	34.5	95.3	5.0	23.4	26.9	69.0	5.5	25.5
purine bulges with both 5′and 3′ adjacent G-U pairs								
GCUU**G**UACC CGAG GUGG	49.4	140.9	5.6	31.7	49.8	142.6	5.6	31.5
GUGU**A**UGUG CACG GCAC	50.8	146.2	5.5	30.8	55.8	162.7	5.4	30.7
GAU**G**UUAGC CUG GAUCG	54.0	160.3	4.3	24.9	45.7	131.6	4.8	26.0
pyrimidine bulges with 5′ adjacent G-U pairs								
ACUG**U**GCG UGAU CGC	54.4	148.9	8.2	46.9	55.7	153.2	8.2	46.6
CAGG**C**GAGC GUCU CUCG	68.5	196.6	7.6	41.7	59.9	168.8	7.6	42.5
CAGG**U**GAGC GUCU CUCG	74.3	214.9	7.6	41.5	73.6	212.8	7.6	41.8
CAGG**C**UAGC GUCU AUCG	42.8	119.1	5.8	31.9	41.7	115.3	6.0	32.7
CAGG**C**AAGC GUCU UUCG	44.4	125.6	5.5	30.0	45.4	128.6	5.6	30.4
CAGG**U**AAGC GUCU UUCG	61.1	180.0	5.3	30.9	60.2	176.6	5.4	31.4

Table 4: (Continued)

	T_M^{-1} vs log C_T plots				average of curve fits			
oligomers[b]	-ΔH° (kcal/mol)	-ΔS° (eu)	-ΔG°$_{37}$ (kcal/mol)	T_M[c] (°C)	-ΔH° (kcal/mol)	-ΔS° (eu)	-ΔG°$_{37}$ (kcal/mol)	T_M[c] (°C)
pyrimidine bulges with 3′ adjacent G-U pairs								
CAGG**U**GAGUC GUCC UUCAG	86.5	250.9	8.7	45.0	103.5	305.0	8.9	44.3
CAGG**C**GAGUC GUCC UUCAG	74.8	214.1	8.4	44.8	70.1	199.1	8.3	45.1
CAGG**U**GAGC GUCC UUCG	61.9	176.8	7.0	39.6	61.3	174.8	7.0	39.6
CAGU**C**GAGUC GUCA UUCAG	46.4	127.6	6.8	38.8	50.2	140.4	6.7	37.7
CAGG**C**GAGC GUCC UUCG	58.9	168.0	6.7	38.1	63.4	183.0	6.7	37.8
CAGA**C**GAGUC GUCU UUCAG	40.5	109.2	6.6	27.7	40.4	108.4	6.8	38.7
CAGA**U**GAGUC GUCU UUCAG	37.5	101.4	6.0	32.9	38.8	105.0	6.2	34.5
Pyrimidine bulges with both 5′and 3′ adjacent G-U pairs								
CAGG**C**GAGC GUCU UUCG	52.5	153.3	4.9	27.8	54.6	160.8	4.7	27.0
CAGG**U**GAGC GUCU UUCG	50.6	148.4	4.6	25.9	54.7	161.8	4.5	26.0

[a] Solutions contained 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM EDTA (pH 7). [b] The nucleotide in bold is the bulge residue. Top sequence is written 5′→3′. [c] Calculated at 10^{-4} M oligomer concentration.

Table 5: Thermodynamic Parameters and Natural Occurrence for Single-Nucleotide Group I Bulges Closed by GU Base Pairs

bulge sequence[a]	ΔG°$_{37bulge}$[b] (kcal/mol)	ΔH°$_{bulge}$[b] (kcal/mol)	percentage of naturally occurring group I sequences[c]	bulge sequence[a]	ΔG°$_{37bulge}$[b] (kcal/mol)	ΔH°$_{bulge}$[b] (kcal/mol)	percentage of naturally occurring group I sequences[c]
	Bulges with 5′G/U				Bulges with 3′G/U		
GUG	**0.5**	**4.2**		GUG	4.7	15.6	
U C	4.0	1.6	2.1	C U	4.0	−8.8	2.2
GUA				AUG			
U U	4.0	10.0	0.6	U U	4.8	42.0	1.1
GCA				GCG	5.0	16.1	
U U	3.8	25.8	0	C U	4.5	13.7	1.0
GCU				UCG			
U A	3.9	25.2	0.0	A U	4.5	34.6	0.6
GCG				ACG			
U C	4.0	11.4	0.0	U U	4.2	39.7	0.1
	Bulges with 5′U/G				Bulges with 3′U/G		
UAC	3.0,[d] 3.0	6.2,[d] 19.3		GAU	**2.0**, 4.4	**21.3**, 13.1	
G G	3.8	0.2	37.9	C G	4.5	6.5	10.1
UAG	**1.6**, 3.6	**0.1**, 7.5		CAU	4.4	38.2	
G C	3.3[d]	8.1[d]	23.0	G G	**1.2**	**5.3**	2.6
UGC				UAU			
G G	2.7	21.8	8.5	A G	5.8	36.8	0.7
UAU				CGU			
G A	3.3	20.7	4.8	G G	**1.7**	**19.7**	0.6
UGU	4.2	13.8		UGU	5.5	56.3	
G A	3.8	8.4	2.0	A G	4.6	16.5	0.6
	Bulges with both 5′- and 3′G/U				Bulges with both 5′- and 3′U/G		
GCG				UAU			
U U	3.1	23.8	1.0	G G	2.6	18.4	0.4
GUG				UGU	**1.9**	**17.5**	
U U	3.4	24.6	0.0	G G	3.1	30.2	0.0

[a] The top strand is written 5′ → 3′. [b] Values calculated as described in the text. [c] The total number of single-nucleotide group I bulge sequences with adjacent GU base pairs is 810. [d] From ref *16*.

began to investigate the influence of bulge loops adjacent to wobble base pairs, we came to recognize additional types of bulge loop sequences (group III and IV). In our initial investigation, some of the sequences originally identified as group I are now classified as either group III or IV. In this study, the complete set of group I single-nucleotide bulge loops with either Watson−Crick or wobble base pair nearest neighbors has been thermodynamically characterized to

Table 6: Group I Single-Nucleotide Bulge Loop Thermodynamic Parameters

free energy increment for the insertion of a group I single-nucleotide bulge loop into a RNA duplex
$\Delta G^\circ_{37\text{bulge loop}} = -0.62\Delta G^\circ_{37\text{less stable stem}} + 0.25$ (kcal/mol)
nearest-neighbor bonus and penalty values
$\binom{5'\text{UBX}}{3'\text{G Y}}$, $\binom{5'\text{GBG}}{3'\text{U U}}$, and $\binom{5'\text{UBU}}{3'\text{G G}} - 0.6$ kcal/mol
$\binom{5'\text{GBX}}{3'\text{U Y}}$ and $\binom{5'\text{XBU}}{3'\text{Y G}} +0.4$ kcal/mol
enthalpic increment for the insertion of a group I single-nucleotide bulge loop into a RNA duplex
$H^\circ_{\text{bulge loop}} = 16.5$ kcal/mol (independent of stem or nearest-neighbor effects)

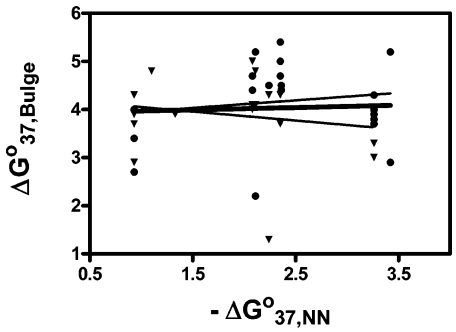

FIGURE 1: Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the free energy increment, $\Delta G^\circ_{37\text{nn}}$, for the Watson−Crick nearest-neighbor interaction at the site of the bulge. Energy differences between purine (●) and pyrimidine (▼) bulge loops. Thin solid lines are the linear regression lines for the purine and pyrimidine data. The dark solid line is the linear regression for all data points.

improve our ability to predict the stability of RNA duplexes with bulge loops.

Nearest-Neighbor Influences on the Thermodynamics of Bulge Loops Adjacent to Watson−Crick Base Pairs. Since the range of thermodynamic contributions of bulge loops adjacent to Watson−Crick base pairs to duplex formation (Table 3) is relatively large, 1.3−5.4 kcal/mol, we reinvestigated the influence of the bulge nearest neighbors on the thermodynamic contribution of bulge loops on duplex formation. Figure 1 displays the plot of the free energy of the nearest-neighbor base pairs versus the free energy of the bulge. Initially, the purine and pyrimidine bulge data were plotted separately and the linear regression for each set determined. The two linear regression lines were found not to be statistically different for each other. Therefore, the two data sets were combined and the linear regression of the entire set of group I single nucleotide bulge loop data determined. The slope of the regression line (0.05 ± 0.16) was not significantly different from zero, further indicating that neither the identity of the bulge nor its nearest neighbors influence the thermodynamics for bulge loop insertion on duplex formation.

Non-Nearest-Neighbor Influences on the Thermodynamics for Bulge Loops Adjacent to Watson−Crick Base Pairs. The data highlighted in bold in Table 3 were originally singled out because they seemed to represent single-nucleotide bulge loops that were much less destabilizing than the remainder of the bulges examined. For example, in Table 3, the bulge sequence $\binom{5'\text{GCU}}{3'\text{C A}}$, imbedded within two different duplex sequences, influences the stability of the duplex by either

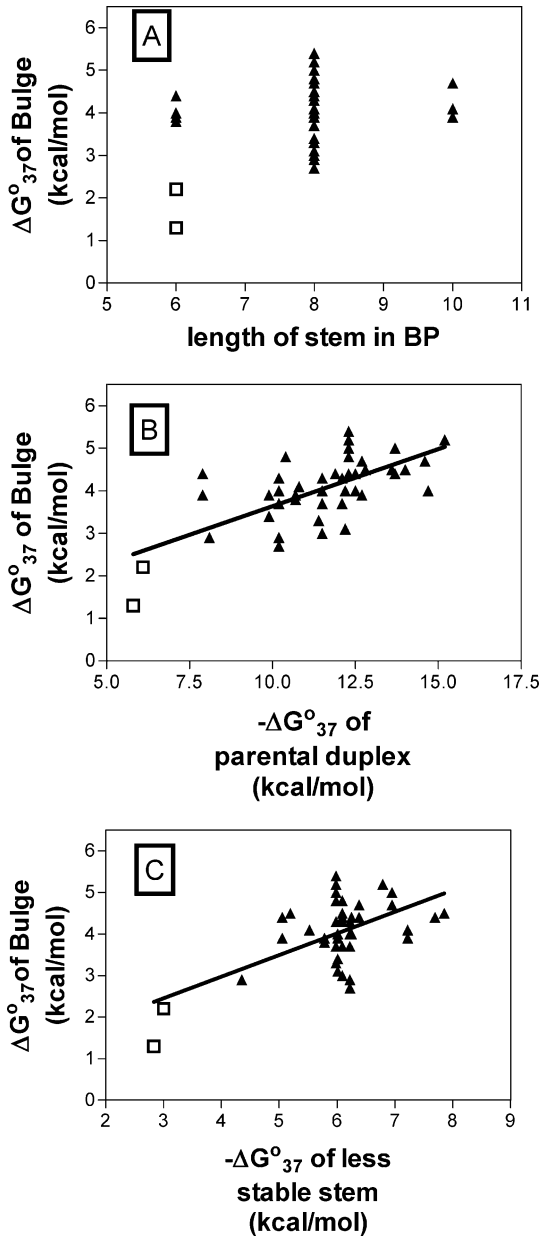

FIGURE 2: (A) Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the length of the duplex. (B) Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the free energy of the parental duplex. (C) Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the free energy of the less stable duplex stem (stem free energy was calculated without the inclusion of the free energy of duplex initiation.). Bulge loops adjacent to Watson−Crick nearest neighbors (▲) and "bold" bulge loops from Table 3 (□). Lines are the least-squares fits of the data points.

1.3 or 4.3 kcal/mol. In an attempt to understand the origin of the difference in the thermodynamic contribution of these bulges to duplex formation, we examined a number of non-nearest-neighbor factors which might contribute to the difference. The first difference we noted was that the bold data points were derived from bulge loops imbedded within relatively short helix stems. Figure 2a examines the influence of helix size upon the thermodynamic contribution of the bulge to duplex formation. Although the bold data points do indeed represent short helices, many of the small helices (hexamer) had bulges which were not unusually destabilizing. Next we investigated whether the stability of the parental

helix was related to the thermodynamic contribution of the bulge to helix formation. This analysis is shown in Figure 2b. There was a clear relationship between the stability of the duplex and the thermodynamic contribution of the bulge to helix formation. A linear fit of the data in Figure 2b gives the relation $\Delta G^\circ_{37\text{bulge loop}} = -0.27\Delta G^\circ_{37\text{duplex}} + 1.0$ ($r^2 = 0.46$), where $\Delta G^\circ_{37\text{bulge loop}}$ is the free energy increment (in kilocalories per mole) for insertion of a bulge loop into a duplex and $\Delta G^\circ_{37\text{duplex}}$ is the free energy of the duplex without the bulge. Finally, we compared the stability of the less stable duplex stem with the thermodynamic contribution of the bulge to helix formation (Figure 2c). A linear fit of the data in Figure 2c gives the following: $\Delta G^\circ_{37\text{bulge loop}} = -0.51\Delta G^\circ_{37\text{duplex}} + 1.0$ ($r^2 = 0.32$). The bulge loops which are less destabilizing to duplex formation (in bold in Table 3) are found inserted into duplexes where the duplex or one of its stems is relatively unstable. While the fit to the data in Figure 2c is not as good as in Figure 2b, inclusion of the data for bulge loops adjacent to wobble base pairs gives a more meaningful relationship (see below).

Non-Nearest-Neighbor Influences on the Thermodynamics for Bulge Loops Adjacent to Wobble Base Pairs. A similar non-nearest-neighbor analysis of the data for bulge loops adjacent to a wobble base pair is presented in Figure 3. The data for the bulge loops adjacent to Watson−Crick base pairs are included in Figure 3 for comparison. Figure 3a examines the influence of helix length upon the thermodynamic contribution of the bulge to duplex formation. For the bulges adjacent to wobble base pairs, several of the less destabilizing bulges were found within longer duplexes of seven or eight base pairs, and as seen for the bulges adjacent to Watson−Crick base pairs, many of the small helices (hexamer) had bulges which were not unusually less destabilizing. Next we investigated whether the stability of the parental helix was related to the thermodynamic contribution of the bulge to helix formation. This analysis is shown in Figure 3b. As for the bulges adjacent to Watson−Crick base pairs, there was a relationship between the stability of the duplex and the thermodynamic contribution of the bulge to helix formation. A linear fit of the data for bulge loops adjacent to wobble base pairs gives the following: $\Delta G^\circ_{37\text{bulge loop}} = -0.42\Delta G^\circ_{37\text{duplex}} - 0.7$ ($r^2 = 0.37$). This line was not significantly different from the line for the bulge loops adjacent to Watson−Crick base pairs, so both sets of data were combined. The linear fit for the combined data gives the relation $\Delta G^\circ_{37\text{bulge loop}} = -0.33\Delta G^\circ_{37\text{duplex}} - 0.7$ ($r^2 = 0.41$), where $\Delta G^\circ_{37\text{bulge loop}}$ is the free energy increment (in kilocalories per mole) for insertion of a bulge loop into a duplex and $\Delta G^\circ_{37\text{duplex}}$ is the free energy of the duplex without the bulge. Finally, we compared the stability of the less stable duplex stem with the thermodynamic contribution of the bulge adjacent to a wobble base pair to helix formation (Figure 3c). A linear fit of the data for bulge loops adjacent to wobble base pairs in Figure 3c gives the following: $\Delta G^\circ_{37\text{bulge loop}} = -0.66\Delta G^\circ_{37\text{less stable stem}} + 0.0$ ($r^2 = 0.37$). This line was not significantly different from the line for the bulge loops adjacent to Watson−Crick base pairs, so both sets of data were combined. The linear fit for the combined data gives the following: $\Delta G^\circ_{37\text{bulge loop}} = -0.62\Delta G^\circ_{37\text{less stable stem}} - 0.25$ ($r^2 = 0.38$). While the fit of the data in Figure 3c is slightly better than the fit in Figure 3b, the difference is so small that it is meaningless. So in considering the two

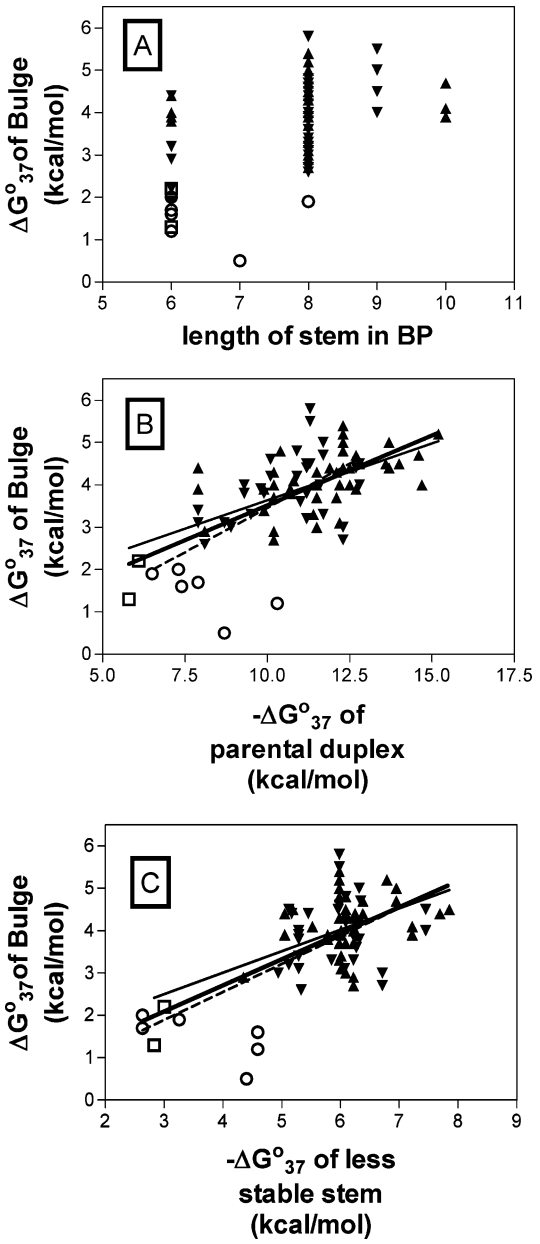

FIGURE 3: (A) Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the length of the duplex. (B) Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the free energy of the parental duplex. (C) Plot of the free energy change for bulge loop formation, $\Delta G^\circ_{37\text{bulge}}$, vs the free energy of the less stable duplex stem (stem free energy was calculated without the inclusion of the free energy of duplex initiation). Bulge loops adjacent to Watson−Crick nearest neighbors (▲), bulge loops adjacent to wobble nearest neighbors (▼), and "bold" bulge loops from Tables 3 (□) and 6 (○). Lines are the least-squares fits of the data points.

relationships, we found using the less stable stem to be more meaningful, because in Figure 3c, all of the less destabilizing bulge data points are now clustered. They all appear to be inserted adjacent to a stem that had a stability of less than 5 kcal/mol. These results are summarized in Table 6.

In an earlier report on bulge loops (*15*), non-nearest-neighbor interactions were observed for the thermodynamics of insertion of a bulge loop into a duplex. Substitution of a CG base pair for an AU base pair, two nucleotides removed from the bulge, caused the bulge to be more destabilizing. It was also observed that the addition of a 3′-dangling end

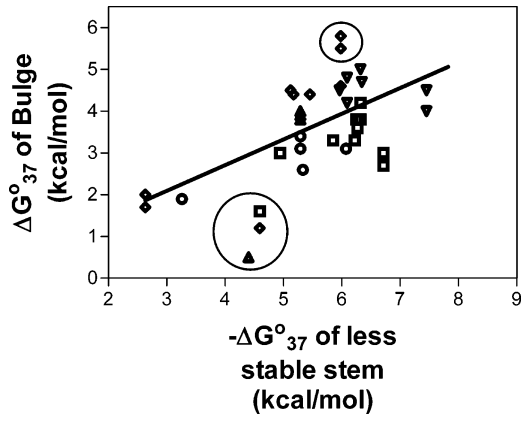

FIGURE 4: (A) Plot of the free energy change for bulge loop formation, $\Delta G^{\circ}_{37bulge}$, vs the free energy of the parental duplex. (B) Plot of the free energy change for bulge loop formation, $\Delta G^{\circ}_{37bulge}$, vs the free energy of the less stable duplex stem (stem free energy was calculated without the inclusion of the free energy of duplex initiation). $\binom{5'UBX}{3'G\ \ Y}$ bulge loops (□), $\binom{5'GBX}{3'U\ \ Y}$ bulge loops (◇), $\binom{5'XBG}{3'Y\ \ U}$ bulge loops (○), $\binom{5'XBU}{3'Y\ \ G}$ bulge loops (△), and double GU bulge loops (▽). Lines are the linear regressions for all Watson−Crick and wobble base pair data points.

also caused the bulge to have a more destabilizing effect. Both of these observations are in concordance with the relationship described above. In both cases, substitution of a GC base pair for an AU base pair and the addition of a 3′-dangling end increase the stability of the stem adjacent to the bulge and therefore should increase the destabilizing effect of the inserted bulge.

Similar non-nearest-neighbor effects have been observed with other structural motifs. In particular, there are several reports where non-nearest-neighbor effects have influenced the thermodynamics of single mismatches in RNA duplexes (*42*, *43*). The position of the single mismatch with a duplex has been shown to influence the stability of the duplex. As the bulge was moved closer to the terminus (decreasing the stability of the stem), the stability of the duplex increased. This effect was somewhat sequence specific as it was observed for UU and AA single mismatches but not for GG mismatches (*42*).

The influence of the adjacent stem on the thermodynamics of insertion of a bulge loop into a RNA helix may be related to the tight packing and electrostatic stain present in the RNA duplex (*43*). The lower the stability of the adjacent stem, the easier the structural perturbation from the insertion of the bulged nucleotide can be dissipated along the neighboring base pairs. Therefore, as the stability of the neighboring stem decreases, it becomes easier (less energetically costly) to insert a bulge loop, or other structural motif (e.g., single mismatch).

Nearest-Neighbor Influences on the Thermodynamics for Bulge Loops Adjacent to Wobble Base Pairs. In our analysis thus far, we have considered the bulge loops adjacent to wobble base pairs as a single group, but in fact, the position and orientation of the nearest-neighbor wobble base pair influence the thermodynamics of bulge loop formation. These influences are examined in Figure 4, where the free energy increment for bulge loop formation is graphed as a function of the position and orientation of the wobble base pair in relation to the bulge. In Figure 4, the free energy increment for bulge loop formation as a function of the less stable stem for each family of bulge loops is graphed relative to the least-squares fit for all of the data points. For example, if we focus on the bulge sequences of the type $\binom{5'UBX}{3'G\ \ Y}$, all 10 of the bulge's data points fall below the line by an average of 0.6 kcal/mol. Therefore, these bulges have a smaller destabilization on duplex formation than other bulge loops. This may be caused by the geometry of the GU base pair which produces a change in the twist angle of the RNA duplex (*35*). For sequences of the type $\binom{5'UX}{3'GY}$, the twist angle between the base pairs is increased to ~40° which leads to unstacking of the wobble base pair with the adjacent base pair in the helix (Figure 3A). $\binom{5'UX}{3'GY}$ nearest-neighbor base pairs are among the less stable. For sequences of the type $\binom{5'XBG}{3'Y\ \ U}$, five of the seven bulges have values that fall above the line, although the average is only 0.2 kcal/mol greater, within the error of the experiment.

For sequences of the type $\binom{5'GBX}{3'U\ \ Y}$ and $\binom{5'XBU}{3'Y\ \ G}$, 11 of the 15 bulges have values that fall above the line, indicating that they are more destabilizing than other bulges. If the very less destabilizing $\binom{5'GUG}{3'U\ \ C}$ bulge is excluded (first value in Table 5), the average value is 0.4 kcal/mol above the line. In a similar analysis of the geometry of a $\binom{5'GBX}{3'U\ \ Y}$ base pair, the wobble base pair decreases the twist angle between the base pairs (Figure 3B). In this instance, the base pair still stacks well with the wobble pair and these nearest-neighbor pairs are among the more stable, so the bulge has a greater potential to disrupt the nearest-neighbor base pairs. For the bulges inserted between two wobble base pairs, all five of the data points fall below the line by an average of 0.6 kcal/mol. Consecutive wobble pairs also lead to unstacking of the bases as seen in Figure 3C. That a bulge between two adjacent wobble base pairs is the least destabilizing may be due to the fact that the tandem GU pairs are among the least stable nearest-neighbor base pairs. In fact, they contribute a positive free energy to duplex formation (*36*). These results are summarized in Table 4. The parameters in Table 4 can be combined with the nearest-neighbor parameters for duplex formation (*32*) to predict the thermodynamic stability of RNA duplexes containing bulge loops.

In Figure 4, five data points are circled, three that correspond to bulge loops that are significantly less destabilizing than predicted from the linear regression and two that are more destabilizing than predicted. The three sequences that are less destabilizing are $\binom{5'ACUGUGCG}{3'UGAU\ \ CGC}$, $\binom{5'GCUAGAC}{3'CGG\ \ CUG}$, and $\binom{5'GACAUAGUC}{3'CUG\ \ GUCAG}$. There does not appear to be any common nearest-neighbor or non-nearest-neighbor features that would explain their free energy increment. The position and orientation of the bulge are different in each case as well as the differences in the length and stability of the stems relative to the position of the bulge. In other sequence contexts, all three of the bulges have approximately the predicted value for the free energy increment of duplex destabilization. The two bulges that were circled as more destabilizing than predicted have the sequences $\binom{5'CAGUAUAGUC}{3'GUCA\ \ GUCAG}$ and $\binom{5'CAGUGUAGUC}{3'GUCA\ \ GUCAG}$ In this case, both bulges are within the same duplex context.

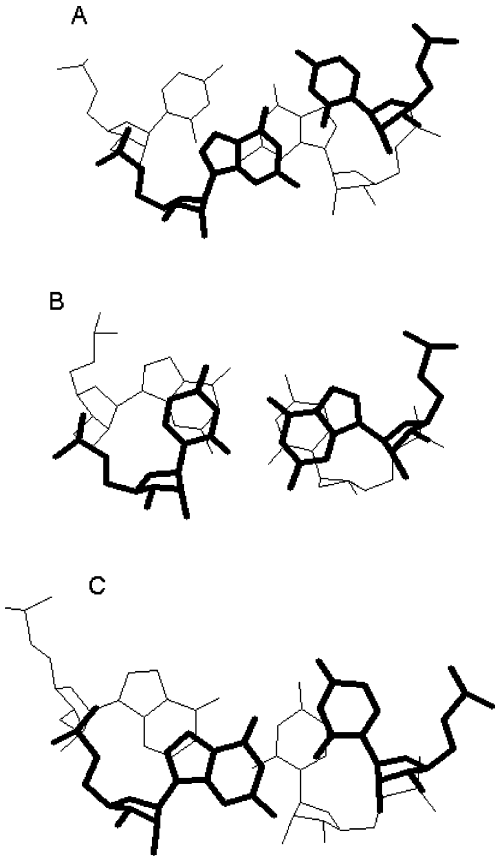

FIGURE 5: Stacking of wobble GU base pairs. (A) View down the helix axis of 5′**U**G/3′**G**C. (B) View down helix axis of 5′G**U**/3′C**G**. (C) View down the helix axis of 5′G**G**/3′U**U**. The wobble base pairs are shown as the nearer base pair and are bold. Examples are taken from the crystal structure of $[r(CGUGAUCG)dC]_2$ (*40*) (A and B) and $(GGUAUUGCGGUACC)_2$ (*41*) (C).

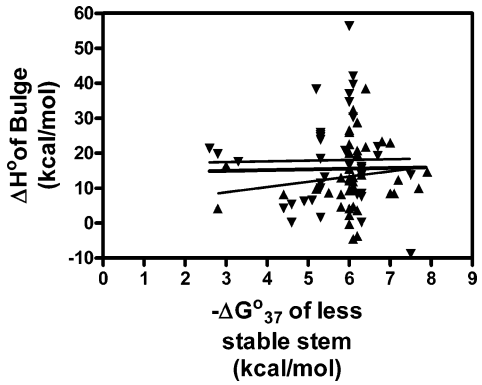

FIGURE 6: Plot of the enthalpy change for bulge loop formation, $\Delta H°_{bulge}$, vs the free energy of the less stable duplex stem (stem free energy was calculated without the inclusion of the free energy of duplex initiation). Bulges adjacent to Watson−Crick nearest neighbors (▲) and wobble nearest neighbors (▼). Lines are the linear regressions for Watson−Crick and wobble base pair data points.

Again, there is no obvious explanation for the larger free energy increment for these bulges.

Enthalpic Contributions of Group 1 Single-Nucleotide Bulge Loops to Duplex Stability. Since the free energy increment of the insertion of a bulge loop into a duplex was dependent upon the stability of the stem adjacent to the insertion, we investigated the influence of stem stability on the enthalpy of bulge loop insertion. Figure 6 displays the enthalpy of bugle loop insertion as a function of the stability of the less stable stem of the duplex for bulges adjacent to both Watson−Crick and wobble base pairs. The least-squares fits for the Watson−Crick and wobble data were first considered separately. Neither line had a slope significantly different from zero so the data sets were combined. The slopes of the line for the combined Watson−Crick and wobble enthalpy values were again not significantly different from zero, suggesting that the average enthalpy value is a reasonable approximation for the enthalpic energy contribution. The average enthalpic value for the combined data is 16.5 ± 11.4 kcal/mol. The enthalpy value can be used in conjunction with the free energy and enthalpy parameters for other nearest-neighbor motifs to determine the stability of RNA structures at temperatures other than 37 °C (*34*).

Phylogenetic Analysis of Single-Nucleotide Bulge Loops. The database examined in this study contained 4330 group I single-nucleotide bulge loops, 3520 closed with Watson−Crick base pairs, and 810 closed wobble base pairs. The frequencies of occurrence for the group I single-nucleotide bulge loops are listed in Tables 3 and 6. For the Watson−Crick closed group I single-nucleotide bulge loops, nearly 70% of the total represent adenosine bulge loops. No other bulge sequence occurs more frequently than 4% except (G**U**G/C C). Since the stability of group I single-nucleotide bulge loops is independent of the identity of the bulge, there is no correlation between the thermodynamic contribution of the bulge and its frequency of occurrence. Therefore, the selection of naturally occurring bulge nucleotide sequences must be related to factors other than stability.

When the bulge has a wobble nearest neighbor, the position and orientation of the bulge do affect its contribution to stability. Nearly 90% of the naturally occurring bulged nucleotides are purine residues, with adenosine representing nearly 80%. The most frequent (76%) type of purine bulges has a 5′U/G, the more stable orientation, while the less stable 3′U/G represents less than 15% of the naturally occurring bulge sequences. However, the most stable bulge loop (with two adjacent wobble base pairs) represents less than 2% of the naturally occurring sequences, and several bulge combinations were not observed in the database.

Single-Nucleotide Bulge Loops in Natural Contexts. The effect of stem stability on the influence of single-nucleotide bulge loop on duplex formation is particularly important to the analysis of naturally occurring RNAs where the average duplex has fewer than seven nucleotides. In fact, of the 20 single-nucleotide bulge loops found in the small and large *E. coli* ribosomal RNA (*33*), 19 have a stem with a stability of <5 kcal/mol. The average stability for the less stable stem is approximately 3 kcal/mol, excluding the influence of terminal mismatches. Therefore, because of the low stability of duplex stems in natural RNAs, the influence of single-nucleotide bulge loops on duplex stability is weaker than previously thought (*16*).

Although the experiments from this and previous bulge loops studies (*15*, *16*) are by agreement performed at 1 M NaCl and show that bulges are generally destabilizing to double-stranded regions of RNA, the conformations of bulged nucleotides can afford them stabilization through binding of other biologically relevant ligands. Bulges can be stabilized via either direct or water-mediated interactions with divalent cations (*42*), often within the context of tertiary

interactions (*43*−*46*). Examples include interactions of the A-rich bulge of P4−P6 of the group I intron (*43*), metal binding in the bulge region of HIV-1 TARN NRA (*44*, *45*), or metal binding within the bulged region of the leadzyme (*46*). In addition, bulges function as critical portions of recognition elements for RNA−protein interactions (*47*−*50*). Thus, the destabilization of double-stranded RNA by bulge loops is often necessary for RNA to properly fold into its functional structures or form RNA−protein complexes which serve to offset the destabilization of the secondary structure.

SUPPORTING INFORMATION AVAILABLE

List of all possible group II single-nucleotide bulge loops (Table S1), list of all possible group III single-nucleotide bulge loops (Table S2), thermodynamic parameters for duplex formation used to generate Table 3 (Table S3), and thermodynamic parameters for duplex formation used to generate Table 5 (Table S4). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) Structures of the bacterial ribosome at 3.5 Å resolution, *Science 310*, 827−834.
2. Allen, G. S., Zavialov, A., Gursky, R., Ehrenberg, M., and Frank, J. (2005) The cryo-EM structure of a translation initiation complex from *Escherichia coli*, *Cell 121*, 703−712.
3. Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H. F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements, *Cell 126*, 1065−1077.
4. Doudna, J. A., and Cech, T. R. (2002) The chemical repertoire of natural ribozymes, *Nature 418*, 222−228.
5. Fedor, M. J., and Williamson, J. R. (2005) The catalytic diversity of RNAs, *Nat. Rev. Mol. Cell Biol. 6*, 399−412.
6. Lilley, D. M. (2005) Structure, folding and mechanisms of ribozymes, *Curr. Opin. Struct. Biol. 15*, 313−323.
7. Bevilacqua, P. C., and Yajima, R. (2006) Nucleobase catalysis in ribozyme mechanism, *Curr. Opin. Chem. Biol. 10*, 455−464.
8. Tucker, B. J., and Breaker, R. R. (2005) Riboswitches as versatile gene control elements, *Curr. Opin. Struct. Biol. 15*, 342−348.
9. Gilbert, S. D., Montange, R. K., Stoddard, C. D., and Batey, R. T. (2006) Structural Studies of the Purine and SAM Binding Riboswitches, *Cold Spring Harbor Symp. Quant. Biol. 71*, 259−268.
10. Sashital, D. G., and Butcher, S. E. (2006) Flipping off the riboswitch: RNA structures that control gene expression, *ACS Chem. Biol. 1*, 341−345.
11. Brion, P., and Westhof, E. (1997) Hierarchy and dynamics of RNA folding, *Annu. Rev. Biophys. Biomol. Struct. 26*, 113−137.
12. Tinoco, I., Jr., and Bustamante, C. (1999) How RNA folds, *J. Mol. Biol. 293*, 271−281.
13. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure, *J. Mol. Biol. 288*, 911−940.
14. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7287−7292.
15. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry 29*, 278−285.
16. Znosko, B. M., Silvestri, S. B., Volkman, H., Boswell, B., and Serra, M. J. (2002) Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for the Formation of RNA Duplexes with Single Nucleotide Bulges, *Biochemistry 41*, 10406−10417.
17. Harper, J. W., and Logsdon, N. J. (1991) Refolded HIV-1 tat protein protects both bulge and loop nucleotides in TAR RNA from ribonucleolytic cleavage, *Biochemistry 30*, 8060−8066.
18. Rounseville, M. P., and Kumar, A. (1992) Binding of a host cell nuclear protein to the stem region of human immunodeficiency virus type 1 trans-activation-responsive RNA, *J. Virol. 66*, 1688−1694.
19. Rounseville, M. P., Lin, H. C., Agbottah, E., Shukla, R. R., Rabson, A. B., and Kumar, A. (1996) Inhibition of HIV-1 replication in viral mutants with altered TAR RNA stem structures, *Virology 216*, 411−417.
20. Klasens, B. I., Thiesen, M., Virtanen, A., and Berkhout, B. (1999) The ability of HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure, *Nucleic Acids Res. 27*, 446−454.
21. Woese, C. R., and Gutell, R. R. (1989) Evidence for several higher order structural elements in ribosomal RNA, *Proc. Natl. Acad. Sci. U.S.A. 86*, 3119−3122.
22. Lanciault, C., and Champoux, J. J. (2005) Effects of unpaired nucleotides within HIV-1 genomic secondary structures on pausing and strand transfer, *J. Biol. Chem. 280*, 2413−2423.
23. Usman, N., Ogilvie, K. K., Jiang, M.-V., and Cedergren, R. (1987) The automated chemical synthesis of long oligoribuncleotides using 2′-O-silylated ribonucleoside 3′-O-phosphoramidites on a controlled-pore glass support: Synthesis of a 43-nucleotide sequence similar to the 3′-half molecule of an *Escherichia coli* formylmethionine tRNA, *J. Am. Chem. Soc. 109*, 7845−7854.
24. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes, *Nucleic Acids Res. 23*, 2677−2684.
25. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 589, CRC Press, Cleveland, OH.
26. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 197, CRC Press, Cleveland, OH.
27. Serra, M. J., Axenson, T. J., and Turner, D. H. (1994) A Model for the Stabilities of RNA Hairpins Based on a Study of the Sequence Dependence of Stability for Hairpins of Six Nucleotides, *Biochemistry 33*, 14289−14296.
28. McDowell, J. A., and Turner, D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAG*GU*CUC)$_2$ by Two-Dimensional NMR and Simulated Annealing, *Biochemistry 35*, 14077−14089.
29. Borer, P., Dengler, B., and Tinoco, I., Jr. (1974) Stability of Ribonucleic acid Double-stranded Helices, *J. Mol. Biol. 86*, 843−853.
30. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability, *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.
31. Allawi, H. T., and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G·T mismatches in DNA, *Biochemistry 36*, 10581−10594.
32. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs, *Biochemistry 37*, 14719−14735.
33. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinf. 3*, 2.
34. Lu, Z. J., Turner, D. H., and Mathews, D. H. (2006) A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation, *Nucleic Acids Res. 34*, 4912−4924.
35. Masquida, B., and Westhof, E. (2000) On the wobble GoU and related pairs, *RNA 6*, 9−15.
36. He, L., Kierzek, R., Santa Lucia, J., Jr., Walter, A. E., and Turner, D. H. (1991) Nearest Neighbor Parameters for GU Mismatches: GU/UG Is Destabilizing in the Context CGUG/GUGC UGUA/AUGA, and AGUU/UUGA but Stabilizing in GGUC/CUGG, *Biochemistry 30*, 11124−11132.
37. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure, *J. Mol. Biol. 288*, 911−940.

38. Wu, M., McDowell, J. A., and Turner, D. H. (1995) A periodic table of symmetric tandem mismatches in RNA, *Biochemistry 34*, 3204−3211.
39. Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, *Biochemistry 39*, 9257−9274.
40. Masquida, B., Sauter, C., and Westhof, E. (1999) A sulfate pocket formed by three GU pairs in the 0.97 Å resolution X-ray of a nonameric RNA, *RNA 5*, 99−112.
41. Trikha, J., Filman, D. J., and Hogle, J. M. (1999) Crystal structure of a 14 bp RNA duplex with non-symmetrical tandem GxU wobble base pairs, *Nucleic Acids Res. 27*, 1728−1739.
42. Xiong, Y., Deng, J., Sudarsanakumar, C., and Sundaralingam, M. (2001) Crystal structure of an RNA duplex r(gugucgcac)$_2$ with uridine bulges, *J. Mol. Biol. 313*, 573−582.
43. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1999) RNA tertiary structure mediation by adenosine platforms, *Science 273*, 1678−1685.
44. Ippolito, J. A., and Steitz, T. A. (1998) A 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation, *Proc. Natl. Acad. Sci. U.S.A. 95*, 9819−9824.
45. Olejniczak, M., Gdaniec, Z., Fischer, A., Grabarkiewicz, T., Bielecki, L., and Adamiak, R. W. (2002) The bulge region of HIV-1 TAR RNA binds metal ions in solution, *Nucleic Acids Res. 30*, 4241−4249.
46. Wedekind, J. E., and McKay, D. B. (2003) Crystal structure of the leadzyme at 1.8 Å resolution: Metal ion binding and the implications for catalytic mechanism and allo site ion regulation, *Biochemistry 42*, 9554−9563.
47. Jaffrey, S. R., Haile, D. J., Klausner, R. D., and Harford, J. B. (1993) The interaction between the iron-responsive element binding protein and its cognate RNA is highly dependent upon both RNA sequence and structure, *Nucleic Acids Res. 21*, 4627−4631.
48. Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G., and Liljas, L. (1997) The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions, *J. Mol. Biol. 270*, 724−738.
49. Lim, F., and Peabody, D. S. (2002) RNA recognition site of PP7 coat protein, *Nucleic Acids Res. 30*, 4138−4144.
50. Seto, A. G., Livengood, A. J., Tzfati, Y., Blackburn, E. H., and Cech, T. R. (2002) A bulged stem tethers Est1p to telomerase RNA in budding yeast, *Genes Dev. 16*, 2800−2812.

BI700736F