The Effect of Base Sequence on the Stability of RNA and DNA Single Base Bulges

Jian Zhu and Roger M. Wartell*

School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332

Received July 16, 1999; Revised Manuscript Received October 1, 1999

ABSTRACT: Forty-eight RNA duplexes were constructed that contained all common single base bulges at six different locations. The stabilities of the RNAs were determined by temperature gradient gel electrophoresis (TGGE). The relative stability of a single base bulge was dependent on both base identity and the nearest neighbor context. The single base bulges were placed into two categories. A bulged base with no identical neighboring base was defined as a Group I base bulge. Group II-bulged bases had at least one neighboring base identical to it. Group II bulges were generally more stable than Group I bulges in the same nearest neighbor environments. This indicates that position degeneracy of an unpaired base enhances stability. Differences in the mobility transition temperatures between the RNA fragments with bulges and the completely base-paired reference RNAs were related to free energy differences. Simple models for estimating the free energy contribution of single base bulges were evaluated from the free energy difference data. The contribution of a Group I bulge 5′-(X*N*Z)-3′•5′-(Z′−X′)-3′ where *N* is the unpaired base and X•X′ and Z•Z′ the neighboring base pairs, could be well-represented (±0.34 kcal/mol) by the equation, $\Delta G_{(XNZ)\cdot(Z'-X')} = 3.11 + 0.40\Delta G^s_{(XZ)\cdot(Z'X')}$. $\Delta G^s_{(XZ)\cdot(Z'X')}$ is the stacking energy of the closing base pair doublet. By adding a constant term, $\delta = -0.3$ kcal/mol, to the right side of the above equation, free energies of Group II bulges could also be predicted with the same accuracy. The term δ represents the stabilizing effect due to position degeneracy. A similar equation/model was applied to previous data from 32 DNA fragments with single base bulges. It predicted the free energy differences with a similar standard deviation.

RNA duplex regions often contain "bulge loops" in which there are one or more consecutive bases on one strand with no paired bases on the other strand. The smallest bulge loop has one base and is termed a single base bulge. Structural studies have shown that single base bulges can assume two types of conformations; a "stacked-in" structure in which the unpaired base is stacked between neighboring bases (*1*−*8*), and a "looped-out" structure in which the unpaired base faces away from the helical axis (*9, 10*). RNA bulge motifs have been implicated to be functionally important for several RNA molecules (*11, 12*).

Algorithms for predicting RNA secondary structure from the primary sequence based on free energy minimization are frequently employed in the structure−function analysis of RNA sequences (*13, 14*). A current algorithm predicts on average 73% of known base pairs in the lowest energy structure (*15*). Accuracy of the prediction depend on free energies assigned to structural motifs. The free energy contribution of a single base bulge has been estimated using the formula;

$$\Delta G_{(XNZ)\cdot(Z'-X')} = 3.9 + \Delta G^s_{(XZ)\cdot(Z'X')} \qquad (1)$$

for an unpaired base *N* enclosed by base pairs, X•X′ and Z•Z′ (*16*). $\Delta G^s_{(XZ)\cdot(Z'X')}$ is the stacking energy of the base pair doublet (XZ)•(Z′X′) surrounding the bulge (*17*). The influence of the identity of the bulged base was not included due to insufficient information.

Knowledge of how base sequence influences the free energy of single base bulges in RNA can be expected to improve the prediction of RNA secondary structure and increase understanding of how sequence changes may alter conformation. A previous study on single base bulges in DNA showed that the identity of the bulged base influenced stability (*18*). If a bulged base was identical to a neighboring base, then it tended to be more stable than other similar bulges. In the current investigation temperature gradient gel electrophoresis[1] (*19*) was employed to determine the influence of single base bulges on the stability of a 345 bp RNA duplex. All base bulges were examined in a number of different nearest neighbor contexts.

MATERIALS AND METHODS

DNA Templates. The plasmid pUC8−31 (*20*) was used as the initial template for polymerase chain reactions (PCR). The target sequence was the 339 bp duplex region shown in Figure 1a. Primers were designed to generate fragments with single base pair substitutions and deletions at six positions indicated by the down arrows in Figure 1a: −36, −35, −34, −33, −28, and −27. Two rounds of PCR were performed to produce the DNA templates for in vitro transcription reactions. In the first round, the pUC8−31 plasmid was used

* To whom correspondence should be addressed. Telephone: 404-894-3735. Fax: 404-894-0519. E-mail: roger.wartell@biology.gatech.edu.

[1] Abbreviations: nt, nucleotide; bp, base pairs; TGGE, temperature gradient gel electrophoresis.

10.1021/bi9916372 CCC: $18.00
Published on Web 11/09/1999

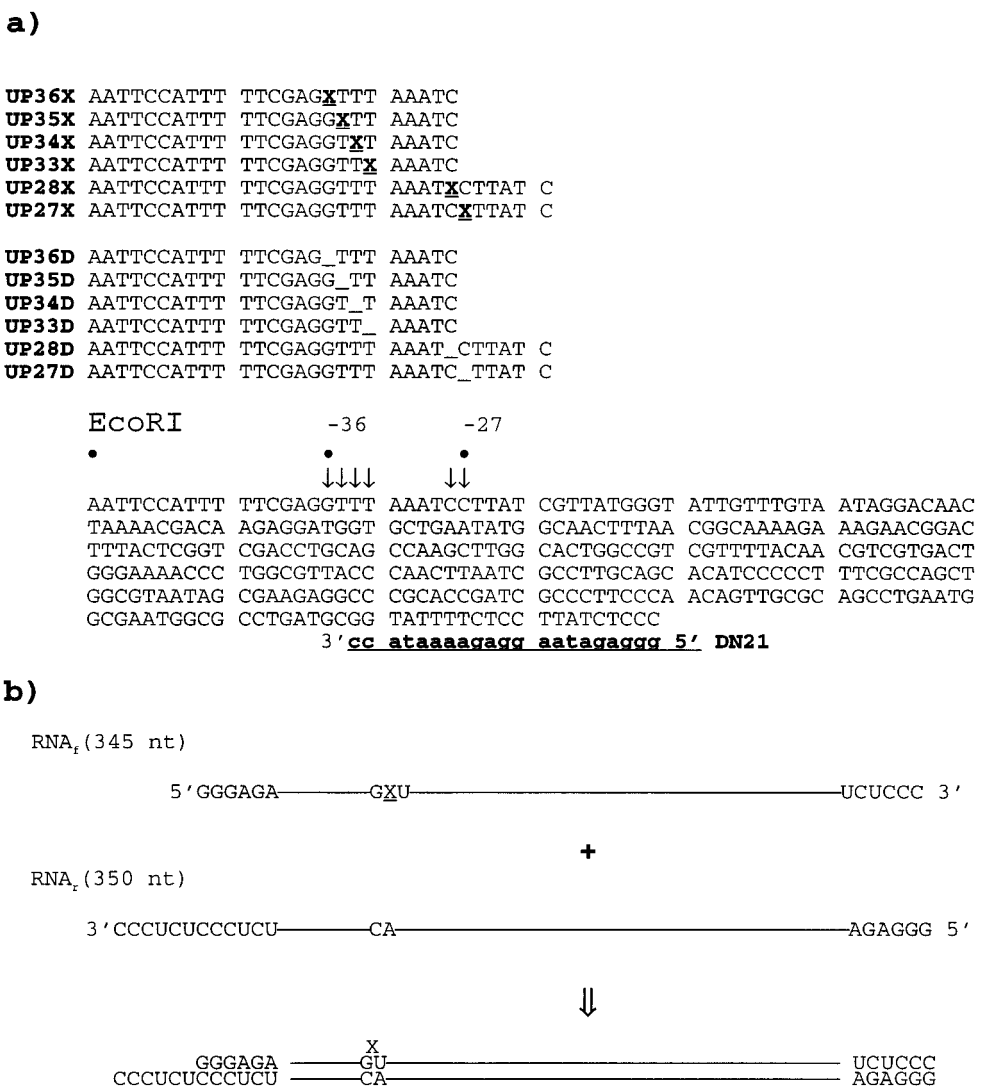

FIGURE 1: Experimental procedure for generating RNA duplex with a single base bulge. (a) 339 bp DNA sequence and primers used to generate DNA templates with single base pair deletions at position 36, 35, 34, 33, 28, and 27. The locations of interest were labeled with down arrows (↓), positions 36 and 27 are denoted by solid circles (●). The sequence starts from an *EcoR*I site. DNA fragments with base pair variations at the above sites were generated using primer sets UP36X, UP35X, UP34X, UP33X, UP28X, and UP27X. X represents A, T, G, or C. The DNA fragments with base pair deletions at designated sites are produced using primer sets UP36D, UP35D, UP34D, UP33D, UP28D, and UP27D. (b) The DNA templates are used to produce single stranded RNAs via in vitro transcription reaction. A duplex RNA carrying a single base bulge on the forward strand was generated by mixed 345 nt forward transcript and 350 nt reverse transcript. RNA_f had a base substitution at the designated site and RNA_r had a base deletion at the location.

as the template. The upstream primers were UP36D, UP35D, UP34D, UP33D, UP28D, and UP27D, which could generate base pair deletions at designated sites, −36, −35, −34, −33, −28, and −27, respectively. UP35D, UP34D, and UP33D are in fact the same primer, but they are named differently in Figure 1 to illustrate that different target locations could be examined. A similar situation occurs with UP28D and UP27D. A common downstream primer (DN21, 5′-GGGAGATAAGGAGAAAATACC-3′) was used. The sequences for the upstream primers employed are given in Figure 1a.

The 338 bp DNAs resulting from the upstream deletion primers were used to template a second round of reactions. In the second PCR, two sets of primers were employed; UP36T7 (5′-TAATACGACTCACTATAGGGAGA*AATTCCATTTTTC*-3′) and DN21. UP36T7 contained a promoter site of T7 RNA polymerase (23 bases) adjacent to the target complimentary sequence that is underlined. The above primer pair were used to generate fragments carrying a T7 promoter at the upstream side of the 338 bp DNA. The resulting fragments were called forward transcription templates (361 bp). Another primer pair, UP25 (5′-GGGAGAGGGAGA*AATTCCAT TTTTC*-3′) and DN38T7 (5′-TAATACGACTCACTATAGGGAGA*TAAGGAGAAAATACC*-3′) produced a fragment with a T7 promoter sequence attached to the down stream side of the original sequence (The italicized sequences are complimentary to the target). These fragments were named reverse transcription templates (367 bp). Both rounds of PCR used the same reaction conditions described in a previous study (*21*). In brief, 35 cycles with denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 1 min.

The DNA fragments with all possible base pairs (A·T, T·A, G·C, and C·G) at a given site were generated through two similar rounds of PCR (*21*). The first round used the upstream primers UP36X, UP35X, ... UP27X shown in Figure 1 with the downstream primer DN21. "X" designates A, T, G, or C. The products of the first round reactions were

339 bp DNAs with base pair substitutions at designated sites. In the second round, the resultant forward templates were 362 bp long and reverse templates were 368 bp and contained the T7 promoter sites as described above.

Transcription. The forward and reverse RNA transcripts were synthesized using the forward and reverse DNA transcription template, respectively. The reaction conditions were previously described (*21*). The forward transcripts with single base deletion were 344 bases in length and the ones with base substitutions were 345 nt long. The reverse transcripts with single base deletions were 350 bases long and the ones with single base substitutions were 351 bases long.

Duplex RNA. The duplex RNAs carrying a single base bulge at a designated location were generated by mixing the forward RNA containing a single base deletion (344 nt) with the reverse RNA carrying base substitution (351 nt), or by mixing forward 345 nt RNA with reverse 350 nt RNA (see Figure 1b). The former carried a bulged base on the reverse strand and the latter had a bulged base on the forward strand. The wild-type duplex RNA was generated by mixing wild-type forward and reverse transcripts. The duplex RNAs with one base pair deletions were produced by mixing forward and reverse single-stranded RNAs carrying base deletions at the same site. All duplex RNAs used in this study carried a 6 nt tail on the 3′ end of the reverse strand. This single stranded tail was important for the TGGE experiments (*21*).

TGGE. A 6.5% polyacrylamide gel (acrylamide:bis-acrylamide = 37.5:1) containing 58% denaturants was used for temperature gradient gel electrophoresis (TGGE) (*18*). In parallel TGGE experiments, 3 μL of duplex RNA sample ($\sim$0.5 μg) were loaded in each lane with 1 μL loading buffer (30% ficol). The temperature gradient was 3 to 6 °C wide, and between 34 and 45 °C. The running time was 12 to 14 h. Voltage ranged from 80 to 120 V. In perpendicular TGGE in which the temperature gradient is perpendicular to the electrical field, 15 μL of the RNA duplex ($\sim$5 μg) was loaded. The gradient was from a minimum of 30 °C to a maximum of 50 °C. The running time was 12−16 h, depending on the voltage applied. The temperature gradient of a perpendicular gel was determined by inserting a temperature probe vertically into the gel at two or more positions after electrophoresis with the water baths running. The gel was stained with EtBr and visualized under UV light and photographed.

RESULTS

Figure 1 illustrates the experimental procedure for generating the RNA duplexes carrying a single base bulge at designated sites. The 339 bp DNA region shown was amplified by PCR and transcribed as described in Materials and Methods. Previous studies (*21*, *22*) have shown this DNA region has three melting domains. The least stable or first melting domain has approximately 52 base pairs and extends inward from the EcoRI end. The second domain is adjacent to the first and is about 60 bp in length. The third domain contains the remaining 227 bp. Duplex RNAs were generated by hybridizing forward and reverse transcripts (Figure 1b). A previous study showed that three melting domains occur in the RNA duplex (*21*) and that the size of the first melting domain was approximately 54 bp. All positions investigated in this study were located in the first melting domain. It was assumed that the introduction of a single base bulge had no significant effect on the size of the first melting domain or melting cooperativity within this domain. The mobility transition data supported this assumption (see below).

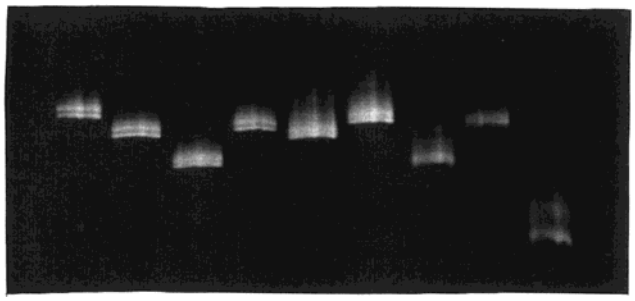

FIGURE 2: Parallel TGGE experiment of RNA duplexes with single base bulges at position 34. The temperature gradient was 36 to 41°C from left to right. Lanes from left to right correspond to fragment with C/−, G/−, U/−, A/−, −/G, −/C, −/A, −/U, and −/− at 34.

Table 1: Stability Orders of Single Base Bulges

position	stability order
36 GXU(AYC)	−/−> −/A > U/−> G/−> −/G > A/− = C/−> −/U = −/C
35 GXU(AYC)	−/−> −/A > U/−> G/−> −/G > A/− = C/−> −/U = −/C
34 UXU(AYA)	−/−> U/− = −/A > A/− = G/−> −/U = −/G = −/C > C/−
33 UXA(UYA)	−/−> A/−> −/A = −/U >U/− = G/−> C/− = −/G = −/C
28 UXC(GYC)	−/−> −/G > −/A > U/− = C/−> −/U = −/C >A/− = G/−
27 CXU(AYG)	−/−> −/A > U/→ −/G > A/− = G/−> C/−> −/U = −/C

The Identity of a Base Bulge and Its Nearest Neighbors Effect Its Stability. Parallel TGGE was used to determine the relative stability of the RNA duplexes carrying all possible single base bulges (8 total) at a given site. For each parallel gel, eight RNA fragments with single bulges were loaded along with a control of the RNA duplex with the base pair deletion at the corresponding site. Figure 2 shows the parallel TGGE result of RNA fragments with a bulge at position 34. The temperature gradient was 36 to 41 °C from top to bottom. RNA fragments moved into the gel as a duplex until they reached a temperature at which their first melting domain unwound. This drastically reduced their mobility. The further a RNA migrated, the greater was the stability of its first melting domain. We note that the RNA bands are observed as closely spaced doublets. This appears to be due to end heterogeneity and its effect on melting. All RNAs, including the fully duplexed RNAs, moved as single bands with indistinguishable mobilities in nondenaturing gels.

Figure 2 shows that the eight RNA fragments with base bulges at position 34 were well-separated from each other. The identity of the bulged base influences its stability. A similar observation was made at the other sites examined. The order of stability of single base bulges at each individual site is summarized in Table 1. Sites with different neighboring base pairs had different orders of stability. For example, at position 36, the order was −/A > U/−> G/> −/G > A/− = C/> −/U = −/C, and at 34 the order was U/− = A/> A/− = C/−> U/− = −/G = −/C > C/−.

Positions 36 and 35 have the same nearest neighbor base pairs but different second neighbors (see Figure 1). Their nearest neighbors were an upstream G·C and a downstream U·A. The second neighbor base pairs for site 36 were an upstream A·U and a downstream U·A, and for site 35 an upstream G·C and a downstream U·A. The order of stability of the single base bulges was exactly the same. Thus, nearest neighbor base pairs appear to dominant the relative order of base bulge stability.

A Bulged Base with an Identical Neighboring Base Has Enhanced Stability. Table 1 shows that a bulged base identical to one of its neighboring bases was generally more stable than bulged bases with no identical neighboring base in the same sequence context. At position 36 on the forward strand, for example, the G bulge (GGU)·(A−C), and the U bulge (GUU)·(A−C), were more stable than the A bulge (GAU)•(A−C) and the C bulge (GCU)•(A−C). At site 34, the U bulge (UUU)•(A−A) was the most stable fragment on the forward strand, and on the reverse strand, the A bulge (U−U)•(AAA) was the most stable fragment. The enhanced stability of these sequences may be explained by the ability of the bulged base to exchange positions with its identical neighbors, thus increasing entropy and lowering the free energy. However, this tendency was not always observed. At position 36 on the reverse strand, the C bulge (G−U)•(ACC) was less stable than the G bulge (G−U)•(AGC) and the same stability as the U bulge (G−U)•(AUC). Apparently, factors in addition to position degeneracy influence the relative order of stability for a given sequence context.

The RNA fragments with single base bulges were divided into two groups. Group I contained RNAs with base bulges having no identical neighboring bases. Group II consisted of those RNAs with base bulges with one or two identical neighboring bases. Among the 48 RNA fragments investigated, 26 were in Group I and 22 were in Group II. Only 3 of the 22 fragments in Group II were less stable than Group I bulges in the same sequence context. The exceptions were C bulges adjacent to a C·G base pair, at sites 36, 35, and 27. This may indicate a special conformation for this sequence, perhaps a looped-out conformation that inhibits position exchange.

Another observation that may be related to exchanging bulge positions concerns Group I G bulges with an adjacent A base. This type of bulge was always greater or equal to the stability of other Group I base bulges in the same environment. For example, at positions 36 and 35 on the reverse strand, the G bulge (G−U)·(AGC) was more stable than the U bulge (G−U)·(AUC). At position 33, (UGA)·(U−A) was more stable than the bulge (UCA)·(U−A). At position 34, the bulge represented by (U−U)•(AGA) was equal in stability to the C or U bulges at this site. The G base bulge may be able to intermittently form a G·U pair displacing the adjacent A as a bulge. A G•U pair is the most stable of the mismatched base pairs.

Transition Temperature Differences. Perpendicular TGGE was used to measure the differences between the mobility transition temperatures of the RNAs with single base bulges and the corresponding fully duplexed RNAs. RNA fragments carrying A, U, G, and C bulges on either the forward or reverse strand at a given site were loaded on a gel along with the duplex RNA with a deletion at that site. A typical experiment is shown in Figure 3. Only the first melting domains are observed. The transition curves of different fragments were parallel to each other. This supports the assumption that a single base bulge has no significant effect on the size and cooperativity of the first melting domain.

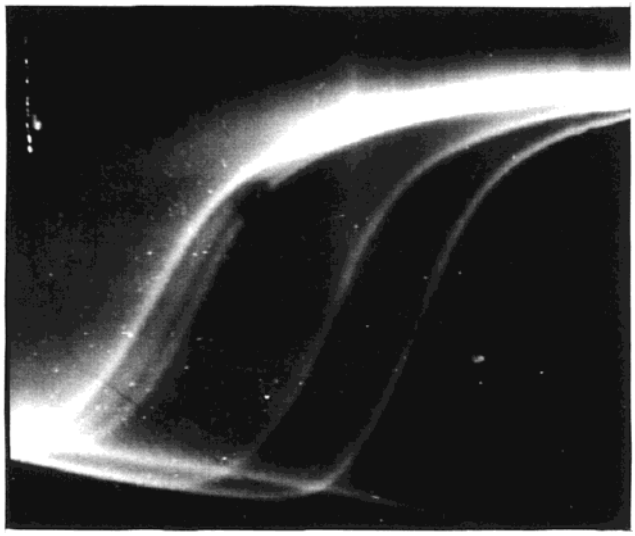

FIGURE 3: Perpendicular TGGE experiment of RNA duplex with single base bulges at position 27. The temperature gradient was 34 to 45 °C from left to right. Five curves from left to right correspond to fragments with −/U, −/G, −/A, −/−, and C·G at 27.

The transition midpoint temperature of the first melting domains, T_μ, was measured as described in previous studies (*18, 21*). T_μ values for the duplex RNAs with base pair deletions at each site ranged from 39 to 41 °C with reproducibility of ±0.2−0.3 °C (Table 2). Using these duplex RNAs as references, transition temperature differences (δT_μ) were determined for each single bulge RNA fragment. The results are shown in Table 2. The temperature decrease due to a single base bulge ranged from 2.0 to 4.0 °C. The maximum variation of δT_μ at a single site was 1.1 °C. The error in measuring temperature differences was ±0.1 °C.

Previous studies (*21*) showed that the free energy difference between an RNA fragment with a "defect" (e.g., base bulge) in the lowest melting domain, and the reference RNA duplex could be calculated using the following equation:

$$\delta\Delta G = -N \cdot \langle \Delta S^\circ \rangle \cdot \delta T_\mu \qquad (2)$$

N is the size of the melting domain (54 bp in this study), $\langle \Delta S^\circ \rangle$ is the average entropy change for opening a base pair in the domain, and δT_μ is the transition temperature difference between the two fragments. The value of $\langle \Delta S^\circ \rangle$ for the solvent employed is not known. We assumed a value obtained by averaging the entropy values of the 10 base pair doublets determined by Xia et al. (*17*). This value, 27.85 cal/K·mol, was determined for conditions of 1 M NaCl and 37 °C. $\delta\Delta G$ values based on eq 2 are listed in Table 3. The average of the free energy differences for all fragments was 3.93 kcal/mol. For a given site, the free energy difference among bulges ranged up to 1.4 kcal/mol.

Models for Estimating Single Base Bulge Free Energies. A comparison was made between the measured free energy differences and predictions based on simple models of the interactions governing single base bulges. The free energy difference between an RNA fragment with a single base bulge and the reference RNA duplex can be expressed in

Table 2: Transition Temperature Differences of RNA Single Base Bulges and T_μ Values of Reference RNA Duplexes

	36 GXU/AYC	35 GXU/AYC	34 UXU/AYA	33 UXA/UYA	28 UXC/GYA	27 CXU/AYG
−/−	0.0	0.0	0.0	0.0	0.0	0.0
A/−	3.0	3.1	2.6	2.0	3.2	2.6
U/−	2.0	2.2	2.2	2.2	2.8	2.4
G/−	2.6	2.5	2.6	2.2	3.2	2.6
C/−	3.0	3.1	2.9	2.4	2.8	2.7
−/A	1.9	2.1	2.2	2.1	2.6	2.2
−/U	3.1	3.2	2.7	2.1	3.0	2.9
−/G	2.8	2.9	2.7	2.4	2.3	2.5
−/C	3.1	3.2	2.7	2.4	3.0	2.9
[a] T_μ	39.0 ± 0.1	40.6 ± 0.3	40.7 ± 0.2	40.5 ± 0.1	39.1 ± 0.2	38.8 ± 0.3

[a] The average mobility transition temperatures in degrees Celsius of the reference RNA with the deleted base pair at the site specified.

terms of local nearest neighbor interactions (*16*). For a base "N" inserted in the top strand of the site (XZ)·(Z′X′), the free energy difference is given by

$$\delta\Delta G_{(XNZ)\cdot(Z'-X')} = \Delta G_{(XNZ)\cdot(Z'-X')} - \Delta G^{s}_{(XZ)\cdot(Z'X')} \quad (3)$$

The term on the left represents the free energy difference between the RNA fragment with the single base bulge and the reference RNA. The terms on the right side of eq 3 are local free energy terms. The first represents the bulge and surrounding base pairs. The second represents the duplex stacking interaction of base pairs at the bulge site. Duplex initiation free energies and all other stacking free energies of the two RNAs cancel.

Equation 1 described in the Introduction provides a simple first-order model for the free energy term $\Delta G_{(XNZ)\cdot(Z'-X')}$. Substituting eq 1 into eq 3 predicts that all measured free energy differences should be the same and equal to 3.9 kcal/mol. While this is clearly not indicated by the data in Table 3, the average of the measured free energy differences is 3.93 kcal/mol in excellent agreement with the prediction. This is unanticipated given the difference between the experimental conditions of TGGE, and those used to establish eq 1 (1 M NaCl). However, previous studies with both DNA and RNA fragments also showed that free energy differences measured by TGGE for base pair stacking interactions were closely correlated to values calculated from parameters evaluated in 1 M NaCl (*21*, *22*, *27*). The average standard deviation between the predictions made with eq 1 and eq 3, and the experimental $\delta\Delta G$ data of Table 3 was ±0.57 kcal/mol.

The interaction model represented by eq 1 assumes that the free energy of a single base bulge is the sum of two contributions, the stacking interaction of the surrounding base pair doublet as it exists in a duplex, and the unfavorable effect of the bulged base. A bulged base is known to distort the stacking interactions of the surrounding base pairs (*23*−*25*). Thus, the stacking interaction of the base pairs surrounding the base bulge is likely to be weakened compared to the duplex case. Unlike eq 1, the coefficient of $\Delta G^{s}_{(XZ)\cdot(Z'X')}$ can be expected to be less than one. In addition, eq 1 ignores the identity of the bulged base. As described earlier, base bulges that were identical to an adjacent base (Group II) were generally more stable than base bulges without identical neighbors (Group I).

A second model for $\Delta G_{(XNZ)\cdot(Z'-X')}$ was developed based on an empirical analysis of the experimental data. At first, only Group I bulges were considered. They were assumed to have definite structures, while Group II bulges can have alternative structures due to position degeneracy of the bulged base. For example, the G bulge on the forward strand at position 36 could alternate with a structure that has the G bulge at position 37. Each of the six sites examined has two or four Group II bulges that can have alternative structures.

The average $\delta\Delta G$ value ($<\delta\Delta G>$) of the Group I bulges at each site was determined. They are listed in the third column of Table 3. These data was plotted against the duplex stacking energy of each site's base pair doublet and is shown in Figure 4. The doublet stacking energies were from Xia et al. (*17*). Although the $\delta\Delta G$ values were determined in a very different solvent from the stacking energies, it is assumed that the difference measurements are valid in 1 M NaCl (*21*, *22*). Both the $\delta\Delta G$ values and the stacking energies were evaluated at a similar temperature, ∼37 °C. A linear correlation was observed between the $<\delta\Delta G>$ values and the stacking energies at the sites. The equation that best fit the data was

$$<\delta\Delta G>_{(XNZ)\cdot(Z'-X')} = 3.11 - 0.60\ \Delta G^{s}_{(XZ)\cdot(Z'X')} \quad (4)$$

The correlation coefficient was $R^2 = 0.62$. Equation 4 indicates that the more negative the stacking free energy of the site (XZ)·(Z′X′), the more unlikely it is to have a base bulge inserted at that site. Combining eqs 3 and 4 one obtains

$$\Delta G_{(XNZ)\cdot(Z'-X')} = 3.11 + 0.40\Delta G^{s}_{(XZ)\cdot(Z'X')} \quad (5)$$

Equation 5 represents a model for the free energy contribution of Group I single base bulges. The two terms on the right side of the equation can be viewed as two components of the free energy. The first term, 3.11 kcal/mol, is the positive free energy due to inserting a base bulge into the duplex. The second term is the free energy contribution of the base pairs surrounding the bulge. The stacking interaction stabilizes the base bulge, but at 40% of the value of a duplex stacking interaction.

A slight modification was made to eq 5 to predict the free energies of Group II single base bulges. The modification required the determination of the dominant structure among the alternatives available for each Group II bulge. The dominant bulge structure was assumed to be the one closed by the least stable base pair doublet allowed, i.e., the structure producing the lowest free energy difference in eq 4. The following examples illustrate the approach.

At position 36, the sequence context is (AG*X*UUUA)·(UAAA*Y*CU) with position 36 underlined. A G bulge at 36 has two alternative structures (AG*G*UUUA)·(UAAA−CU) and (A*G*GUUUA)·(UAAAC−U). The second structure has

Table 3: Estimating Stability of RNA Single Base Bulges Using Different Models

site	sequence	[a]$\delta\Delta G$	[b]avge	[c]ΔG^s	[d]$\delta\Delta G_1$	[e]d_1	[f]$\delta\Delta G_2$	[g]d_2
36	GAU(A−C)	4.5*	4.45	−2.24	3.9	−0.6	4.45	−0.05
	GUU(A−C)	3.0				0.9	3.37	0.37
	GGU(A−C)	3.9				0.0	4.07	0.27
	GCU(A−C)	4.5*				−0.6	4.45	−0.05
	G−U(AAC)	2.85				1.05	3.37	0.52
	G−U(AUC)	4.65*				−0.75	4.45	−0.20
	G−U(AGC)	4.2*				−0.3	4.45	0.25
	G−U(ACC)	4.65				−0.75	4.07	−0.58
35	GAU(A−C)	4.65*	4.6	−2.24	3.9	−0.75	4.45	−0.20
	GUU(A−C)	3.3				0.6	3.37	0.07
	GGU(A−C)	3.75				0.15	4.07	0.32
	GCU(A−C)	4.65*				−0.75	4.45	−0.20
	G−U(AAC)	3.15				0.75	3.37	0.22
	G−U(AUC)	4.8*				−0.9	4.45	−0.35
	G−U(AGC)	4.35*				−0.45	4.45	0.10
	G−U(ACC)	4.8				−0.9	4.07	−0.73
34	UAU(A−A)	3.9*	4.05	−0.93	3.9	0.0	3.67	−0.23
	UUU(A−A)	3.3				0.6	3.37	0.07
	UGU(A−A)	3.9*				0.0	3.67	−0.23
	UCU(A−A)	4.35*				−0.45	3.67	−0.68
	U−U(AAA)	3.3				0.6	3.37	0.07
	U−U(AUA)	4.05*				−0.15	3.67	−0.38
	U−U(AGA)	4.05*				−0.15	3.67	−0.38
	U−U(ACA)	4.05*				−0.15	3.67	−0.38
33	UAA(U−A)	3.0	3.5	−1.3	3.9	0.9	3.37	0.37
	UUA(U−A)	3.3				0.6	3.37	0.07
	UGA(U−A)	3.3*				0.6	3.91	0.61
	UCA(U−A)	3.6*				0.3	3.91	0.31
	U−A(UAA)	3.15				0.75	3.37	0.22
	U−A(UUA)	3.15				0.75	3.37	0.22
	U−A(UGA)	3.6*				0.3	3.91	0.31
	U−A(UCA)	3.6*				0.3	3.91	0.31
28	UAC(G−A)	4.8*	4.65	−2.35	3.9	−0.9	4.52	−0.28
	UUC(G−A)	4.2				−0.3	3.47	−0.73
	UGC(G−A)	4.8*				−0.9	4.52	−0.28
	UCC(G−A)	4.2				−0.3	4.07	−0.13
	U−C(GAA)	3.9				0.0	3.47	−0.43
	U−C(GUA)	4.5*				−0.6	4.52	0.02
	U−C(GGA)	3.45				0.45	4.07	0.62
	U−C(GCA)	4.5*				−0.6	4.52	0.02
27	CAU(A−G)	3.9*	4.1	−1.7	3.9	0.0	4.37	0.47
	CUU(A−G)	3.6				0.3	3.37	−0.23
	CGU(A−G)	3.9*				0.0	4.37	0.47
	CCU(A−G)	4.05				−0.15	4.07	0.02
	C−U(AAG)	3.3				0.6	3.37	0.07
	C−U(AUG)	4.35*				−0.45	4.37	0.02
	C−U(AGG)	3.85				0.05	4.07	0.22
	C−U(ACG)	4.35*				−0.45	4.07	−0.28
	average	3.93			3.9		3.93	
	[h]STDEV					0.57		0.34

[a] Experimental free energy differences using data in Table 2 and eq 2. [b] Average free energy difference at each site using Group I bulges (*). [c] The stacking energy of the base pair doublet that closes Group I bulges at the site. From Xia et al. (*16*). [d] Predicted free energy differences using model from Serra and Turner (*15*), eq 1. [e] d_1 equals $\delta\Delta G_1 - \delta\Delta G$; differences between the elements of columns *d* and *a*. [f] Predicted free energy differences using model of eq 7. [g] d_2 equals $\delta\Delta G_2 - \delta\Delta G$; the differences between the elements of columns *f* and *a*. [h] The average standard deviation.

the G bulge moved to position 37. Using the RNA fragment with a deletion at 36 as reference, the free energy difference for the first structure is $3.1 - 0.60\Delta G^s_{(GU)\cdot(AC)} = 4.45$ kcal/mol, while the free energy difference for the second structure is $3.11 - 0.60\Delta G^s_{(AG)\cdot(CU)} = 4.37$ kcal/mol. The second structure has a lower energy and is considered the dominant structure. Similarly, a U bulge at position 36 had four alternative structures; (AG*U*UUUA)·(UAAA−CU), (AGU*U*UUA)·(UAA−ACU), (AGUU*U*UA)·(UA−AAACU), and (AGUUU*U*A)·(U−AAAACU). For these alternative U bulge structures, the free energy differences from eq 4 are

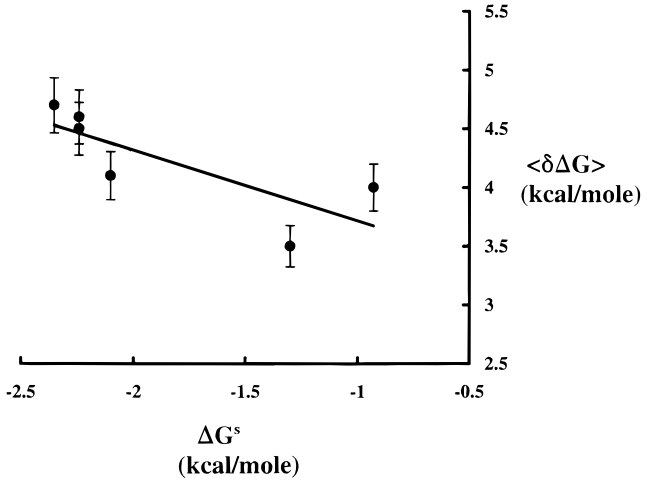

FIGURE 4: Plot of average free energy differences of the Group I RNA bulge structures for each site against the stacking energy of the closing base pair doublet.

4.45, 3.67, 3.67, and 3.9 kcal/mol, respectively. The dominant structure for the U bulge designated at position 36 has the U bulge shifted to position 34 or 35, closed by a (UU)•-(AA).

For each Group II base bulge a dominant structure was determined using the above approach. $\Delta G_{(XNZ)\cdot(Z'-X')}$ was then calculated from a modified eq 5. The stacking energy term, $\Delta G^s_{(XZ)\cdot(Z'X')}$, of the dominant structure was employed and a constant term δg was subtracted from the right side of eq 5. The latter term represents the influence of positional degeneracy on a Group II bulge. Position degeneracy is expected to increase entropy and lower the free energy. Thus, the modified version of eq 5 for predicting the free energy contribution for single base bulges is

$$\Delta G_{(XNZ)\cdot(Z'-X')} = 3.11 + 0.40\Delta G^s_{(XZ)\cdot(Z'X')} - \delta g \quad (6)$$

The analogous version of eq 4 is

$$\delta\Delta G_{(XNZ)\cdot(Z'-X')} = 3.11 - 0.60\Delta G^s_{(XZ)\cdot(Z'X')} - \delta g \quad (7)$$

δg is zero for a Group I bulge, and a positive constant if *N* is identical to either X or Z. δg was determined by minimizing the average standard deviation between all experimental and calculated free energy differences. Predicted free energy differences using the value of $\delta g = 0.3$ kcal/mol are listed in Table 3 under $\delta\Delta G_2$. Differences between the predictions and the experimental values are listed under d_2. The average standard deviation, ±0.34 kcal/mol, was significantly lower than the value obtained with the first model (±0.57 kcal/mol). Further improvement of the predictions can be made if the model is made more sequence specific, but at the loss of equation simplicity.

DISCUSSION

Results from this work show that the free energy of RNA single base bulges depend on the base as well as the surrounding base pairs. At a given site, the free energy varied up to 1.4 kcal/mol, depending on the base that was unpaired. Position degeneracy appears to play a significant role in determining the stability of single base bulges. The strong correlation between the classification of the base bulges as

Table 4: Transition Temperature Differences of DNA Single Base Bulges[a]

	43 TXT/AYA	39 CXA/TYG	38 GXG/CYC	36 GXT/AYC
−/−	0.0	0.0	0.0	0.0
A/−	2.9	2.4	2.9	3.3
T/−	2.4	2.8	3.1	2.8
G/−	2.9	2.5	2.0	2.4
C/−	2.9	2.7	3.3	3.3
−/A	2.2	2.7	3.4	2.6
−/T	2.5	2.8	3.6	3.1
−/G	2.5	2.2	3.3	3.1
−/C	2.5	3.0	2.9	3.1

[a] From ref *17*.

Group I or Group II and their relative stability supports this conclusion. Consideration of position degeneracy produced a simple equation, eq 6, for predicting the sequence dependent free energy contribution of single base bulges with good accuracy (±0.34 kcal/mol). Including the sequence dependence to the free energies of single base bulges should enhance the accuracy of RNA secondary structure predictions.

The ability of eq 7 to predict free energy differences with good accuracy, and the similar relative stability of base bulges at sites 36 and 35 indicates that nearest neighbor interactions are the major determinant of base bulge stability. However, additional studies are needed to examine this issue. A few base bulge sequences (e.g., C bulge next to a C·G) did not follow the trend expected from the majority of single base bulges. Other studies have shown that cations and nonneighbor effects can influence the structure and relative stability of local RNA defects (*25*).

It was of interest to determine if a model similar to the one derived for the RNA bulges would provide a good estimate for the stability of DNA single base bulges. A previous study has shown that DNA base bulges in a duplex has properties similar to the RNA molecules (*18*). The measured T_μ differences of DNA fragments with bulges at position 36, 38, 39, and 43 and reference duplex DNAs are listed in Table 4. As observed for the RNA bulges, Group II base bulges were generally more stable than Group I bulges for the same base pair context. Following an analysis similar to that described by eqs 3−7, the free energy differences between the DNA fragments with single base bulges and reference duplex DNA fragments were estimated using the following equation.

$$\delta\Delta G_{(\mathrm{XNZ})\cdot(\mathrm{X'-Z'})} = 2.72 - 0.52\Delta G^{s}_{(\mathrm{XZ})\cdot(\mathrm{Z'X'})} - \delta\mathrm{g} \quad (8)$$

It was assumed that the experimental free energy differences were valid at 37 °C in 1 M NaCl, and thus the stacking free energies employed in deriving eq 8 were those determined in 1 M NaCl by Allawi and Santa Lucia (*27*). The local average free energy contribution of a DNA bulge can be expressed as:

$$\Delta G_{(\mathrm{XNZ})\cdot(\mathrm{X'-Z'})} = 2.72 + 0.48\Delta G^{s}_{(\mathrm{XZ})\cdot(\mathrm{X'Z'})} - \delta\mathrm{g} \quad (9)$$

For Group I bulges, δg is equal to 0, and for Group II bulges, δg is 0.4 kcal/mol.

Comparison of the predicted and experimental free energy differences for DNA fragments with and without a base bulge are given in Table 5. The standard deviation of ±0.32 kcal/mol was essentially the same as observed for the RNA molecules. Stacking interactions have been evaluated for the TGGE solvent conditions. If they are employed in deriving eqs 8 and 9, then the constants of these equations differ but the comparison of predicted and experimental free energy differences gives a similar standard deviation. The model for single base bulges produces similar accuracy for both RNA and DNA. It is of interest to note that for both RNA and DNA, bulge position degeneracy produces an average stabilization (δg) of −0.3 to −0.4 kcal/mol. This correlates well with the simplified entropic estimate for a two position degeneracy; $-RT\ln 2 = -0.43$ kcal/mol at 37 °C.

Table 5: Comparison of Experimental and Predicted DNA Single Base Bulge Stabilities

site	sequence	[a]$\delta\Delta G$	[b]average		[c]$\delta\Delta G_3$	[e]d_3
36	GAT(A−C)	3.8*	3.7	−1.44	3.50	−0.30
	GTT(A−C)	3.2			2.79	−0.41
	GGT(A−C)	2.8			3.01	0.21
	GCT(A−C)	3.8*			3.50	−0.30
	G−T(AAC)	3.0			2.79	−0.21
	G−T(ATC)	3.6*			3.50	−0.10
	G−T(AGC)	3.6*			3.50	−0.10
	G−T(ACC)	3.6			3.01	−0.59
38	GAG(C−C)	3.4*	3.8	−1.84	3.72	0.32
	GTG(C−C)	3.6*			3.72	0.12
	GGG(C−C)	2.3			3.10	0.80
	GCG(C−C)	3.8*			3.72	−0.08
	G−G(CAC)	4.1*			3.72	−0.38
	G−G(CTC)	4.3*			3.72	−0.58
	G−G(CGC)	3.8*			3.72	−0.08
	G−G(CCC)	3.4			3.10	−0.30
39	CAA(T−G)	2.8	3.2	−1.45	3.01	0.21
	CTA(T−G)	3.2*			3.51	0.31
	CGA(T−G)	2.9*			3.51	0.61
	CCA(T−G)	3.1			3.03	−0.07
	C−A(TAG)	3.1*			3.51	0.41
	C−A(TTG)	3.2			3.01	−0.19
	C−A(TGG)	2.6			3.03	0.43
	C−A(TCG)	3.5*			3.51	0.01
43	TAT(A−A)	3.4*	3.3	−1.00	3.26	−0.14
	TTT(A−A)	2.9			2.79	−0.11
	TGT(A−A)	3.4*			3.26	−0.14
	TCT(A−A)	3.4*			3.26	−0.14
	T−T(AAA)	2.6			2.79	0.19
	T−T(ATA)	3.1*			3.26	0.16
	T−T(AGA)	3.1*			3.26	0.16
	T−T(ACA)	3.1*			3.26	0.16
	average	3.3			3.3	0.01
	STDEV					0.32

[a] Free energy difference using temperature difference given in Table 4 using eq 2 with $<\Delta S°> = -22.4$ cal·mol·K^{-1}, and $N = 52$ bp. [b] The stacking energy for closing base pair doublet from Allawi and Santa Lucia (1997). [c] Free energy differences calculated from eq 8. [d] The difference between a and c.

REFERENCES

1. Patel, R., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., and Breslauer, K. J. (1982) *Biochemistry 21*, 445−451.
2. Hare, D., Shapiro, L., and Patel, D. J. (1986) *Biochemistry 25*, 7456−7464.
3. Roy, S., Sklenar, V., Appella, E., and Cohen, J. S. (1987) *Biopolymer 26*, 2041−2052.
4. Kalnik, M. W., Norma, D. G., Swann, P. F., and Patel, D. J. (1989) *J. Biol. Chem. 264*, 3702−3712.
5. Kalnik, M. W., Norma, D. G., Zagorski, M. G., Swann, P. F., and Patel, D. J. (1989) *Biochemistry 28*, 294−303.
6. Woodson, S. A., and Crothers, D. M. (1988) *Biochemistry 27*, 436−445.

7. Woodson, S. A., and Crothers, D. M. (1988) *Biochemistry 27*, 3130−3141.
8. Woodson, S. A., and Crothers, D. M. (1989) *Biopolymers 26*, 1149−1177.
9. Morden, K. M., Chu, Y. G., Martin, F. H., and Tinoco, I, Jr. (1983) *Biochemistry 22*, 5557−5563.
10. Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., and Patel, D. J. (1990) *J. Biol. Chem. 265*, 636−647.
11. Ippolito, J. A., and Steitz, T. A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9819−9824.
12. Fourmy, D., Recht, M., Blanchard, S. C., and Puglisi, J. D. (1996) *Science 274*, 1367−1369.
13. Zuker, M. (1989) *Science 244*, 48−52.
14. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) *Annu. Rev. Biophys. Chem. 17*, 167−192.
15. Mathews, D. H, Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
16. Serra, M. J., and Turner, D. H. (1995) *Methods Enzymol. 259*, 242−261.
17. Xia, T., Santa Lucia, J., Jr., Burkard, M., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
18. Ke, S. H., and Wartell, R. M. (1995) *Biochemistry 34*, 4593−4599.
19. Wartell, R. M., Hosseini, S, H., and Moran, J. D. (1990) *Nucleic Acids Res. 18*, 2699−2705.
20. Tatti, K. M., and Moran, C. P., Jr. (1984) *J. Mol. Biol. 175*, 285−297.
21. Zhu, J., and Wartell, R. M. (1997) *Biochemistry 36*, 15326−15335.
22. Johnson, T., Zhu, J., and Wartell, R. M. (1998) *Biochemistry 37*, 12343−12350.
23. Zacharias, M., and Hagerman, P. J. (1995) *J. Mol. Biol. 247*, 486−500.
24. Tang, R. S., and Draper, D. E. (1990) *Biochemistry 29*, 5232−5237.
25. Tang, R. S., and Draper, D. E. (1994) *Biochemistry 33*, 10089−10093.
26. Draper, D. E., and Misra, V. K. (1998) *Nature Struct. Biol. 5*, 927−30.
27. Allawi, H. T., and Santa Lucia, J., Jr. (1997) *Biochemistry 36*, 10581−10594.

BI9916372